Online ISSN : 0975-4172
Print ISSN : 0975-4350
DOI : 10.17406/GJCST

Global Journal

of Computer Science and Technology: H

Information & Technology

Linux Operating Systems

Use of Dynamic Programming

Highlights

Non-Relational Data Models

Penetration Testing Framework

Discovering Thoughts, Inventing Future

VOLUME 23 ISSUE 2 VERSION 1.0

Global Journal of Computer Science and Technology: H
Information & Technology

Global Journal of Computer Science and Technology: H
Information & Technology

Volume 23 Issue 2 (Ver. 1.0)

Open Association of Research Society

Incorporation No.: 0423089
License No.: 42125/022010/1186
Registration No.: 430374
Import-Export Code: 1109007027
Employer Identification Number (EIN):
USA Tax ID: 98-0673427

Global Journals Inc.

*(A Delaware USA Incorporation with "Good Standing"; **Reg. Number: 0423089**)*

Sponsors: Open Association of Research Society
Open Scientific Standards

Publisher's Headquarters office

Global Journals® Headquarters
945th Concord Streets,
Framingham Massachusetts Pin: 01701,
United States of America
USA Toll Free: +001-888-839-7392
USA Toll Free Fax: +001-888-839-7392

Offset Typesetting

Global Journals Incorporated
2nd, Lansdowne, Lansdowne Rd., Croydon-Surrey,
Pin: CR9 2ER, United Kingdom

Packaging & Continental Dispatching

Global Journals Pvt Ltd
E-3130 Sudama Nagar, Near Gopur Square,
Indore, M.P., Pin:452009, India

Find a correspondence nodal officer near you

To find nodal officer of your country, please email us at *local@globaljournals.org*

eContacts

Press Inquiries: *press@globaljournals.org*
Investor Inquiries: *investors@globaljournals.org*
Technical Support: *technology@globaljournals.org*
Media & Releases: *media@globaljournals.org*

Pricing (Excluding Air Parcel Charges):

Yearly Subscription (Personal & Institutional)
250 USD (B/W) & 350 USD (Color)

Editorial Board

Global Journal of Computer Science and Technology

Dr. Stefano Berretti

Ph.D. in Computer Engineering and Telecommunications, University of Firenze Professor Department of Information Engineering, University of Firenze, Italy

Dr. Abdurrahman Arslanyilmaz

Computer Science & Information Systems Department Youngstown State University Ph.D., Texas A&M University University of Missouri, Columbia Gazi University, Turkey

Dr. Lamri Sayad

Ph.d in Computer science, University of BEJAIA, Algeria

Dr. Tauqeer Ahmad Usmani

Ph.D in Computer Science, Oman

Dr. Hazra Imran

Ph.D in Computer Science (Information Retrieval), Athabasca University, Canada

Dr. Magdy Shayboub Ali

Ph.D in Computer Sciences, MSc in Computer Sciences and Engineering, BSc in Electronic Engineering, Suez Canal University, Egypt

Dr. Nurul Akmar Binti Emran

Ph.D in Computer Science, MSc in Computer Science, Universiti Teknikal Malaysia Melaka, Malaysia

Dr. Asim Sinan Yuksel

Ph.D in Computer Engineering, M.Sc., B.Eng., Suleyman Demirel University, Turkey

Dr. Anis Bey

Dept. of Computer Science, Badji Mokhtar-Annaba University, Annaba, Algeria

Alessandra Lumini

Associate Researcher Department of Computer Science and Engineering University of Bologna Italy

Dr. Rajesh Kumar Rolen

Ph.D in Computer Science, MCA & BCA - IGNOU, MCTS & MCP - MIcrosoft, SCJP - Sun Microsystems, Singhania University, India

Dr. Rajneesh Kumar Gujral

Ph.D in Computer Science and Engineering, M.TECH in Information Technology, B. E. in Computer Science and Engineering, CCNA Certified Network Instructor, Diploma Course in Computer Servicing and Maintenance (DCS), Maharishi Markandeshwar University Mullana, India

Dr. Aziz M. Barbar

Ph.D. IEEE Senior Member Chairperson, Department of Computer Science AUST - American University of Science & Technology Alfred Naccash Avenue Ashrafieh, Lebanon

Dr. Federico Tramarin

Ph.D., Computer Engineering and Networks Group, Institute of Electronics, Italy Department of Information Engineering of the University of Padova, Italy

Dr. Chutisant Kerdvibulvech

Dept. of Inf. & Commun. Technol., Rangsit University Pathum Thani, Thailand Chulalongkorn University Ph.D. Thailand Keio University, Tokyo, Japan

Dr. Roheet Bhatnagar

Ph.D in Computer Science, B.Tech in Computer Science, M.Tech in Remote Sensing, Sikkim Manipal University, India

Contents of the Issue

Global Journal of Computer Science and Technology: H Information & Technology
Volume 23 Issue 2 Version 1.0 Year 2023
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Providing in RDBMSs the flexibility to Work with Various Non-Relational Data Models

By Joachim Tankoano
University of Joseph Ki-Zerbo

Abstract- The inability of pure relational DBMSs to meet the new requirements of the applications that have emerged on the web has led to the advent of No SQL DBMSs. In the last few years, significant progress has been made in integrating into relational DBMSs the features that are essential to consider those new requirements that primarily concern flexibility, performance, horizontal scaling, and very high availability. This paper focuses on the features that can enable relational DBMSs to provide applications with the flexibility to work with various non-relational data models while providing the guarantees of independence, integrity, and performance of query evaluation.

Keywords: ANSI/SPARC architecture of the DBMSs' schemata, database design and models, relational data model, non-relational data models, virtual databases.

GJCST-H Classification: LCC Code: QA76.9.D3

Strictly as per the compliance and regulations of:

Providing in RDBMSs the Flexibility to Work with Various Non-Relational Data Models

Joachim Tankoano

Abstract- **The inability of pure relational DBMSs to meet the new requirements of the applications that have emerged on the web has led to the advent of No SQL DBMSs. In the last few years, significant progress has been made in integrating into relational DBMSs the features that are essential to consider those new requirements that primarily concern flexibility, performance, horizontal scaling, and very high availability. This paper focuses on the features that can enable relational DBMSs to provide applications with the flexibility to work with various non-relational data models while providing the guarantees of independence, integrity, and performance of query evaluation.**

Keywords: ANSI/SPARC architecture of the DBMSs' schemata, database design and models, relational data model, non-relational data models, virtual databases.

I. Introduction

Data models are used in database technology: (1) To define the logical structure of the content of a database, (2) To provide DBMSs with the capability to maintain the integrity of that content, and (3) To provide an abstract language for the manipulation of that content.

[1] considers that each data model characterizes, using four sets (T, S, O, C) of its own, a virtual machine with which the users of the database interact, where: T is a set of data types, S is a set of data structure types, O is a set of data operation types, and C is a set of integrity constraint types.

Conceptual models (such as UML [2] and the entity-relationship model [3]) and the pure relational model [4] are those that have been the subject of many theoretical studies about their set C. Consequently, those data models are the ones that offer the most possibilities to guarantee data quality and reliability thanks to the simple and complex types of integrity constraints contained in C [5-6].

In exchange, the pure relational model stores data in tabular relations made of rows where each column contains an atomic value. While such rigidity is very suitable for structured data describing non-complex entities in the real world, such rigidity is inappropriate for structured, semi-structured, unstructured, or hybrid data describing complex entities [5] or for data organized as a graph.

Author: Université Joseph Ki-Zerbo – Bp 10048 Ouaga Zogona, Ouagadougou, Burkina Faso. e-mail: tankoanoj@gmail.com

The other commonly used data models are alternative models of the pure relational model. Among them are [7-10] the nested relational model, the object-oriented model, the document-oriented model (XML [11] and JSON [12], for example), the graph-oriented model, the column-oriented model, and the key/value-oriented model. Those data models have been introduced in database technology with the primary purpose of providing greater flexibility than the pure relational model thanks to the kinds of data types, data structure types, and data operation types contained in T, S, and O, on the one hand, for describing the data structure of the complex entities and, on the other hand, for handling the variability in the data types and the data structure types due to data source diversity [9].

In exchange (see paragraph 3), those data models sacrifice data independence and typically consider only a subset of the simplest integrity constraint types of the pure relational model, such as primary, unique, and referential keys. Additionally, some of those data models induce other integrity constraint types, such as the integrity constraints on the materialization of relationships that most DBMSs ignore. Those data models, therefore, lead the DBMSs to sacrifice the guarantees of independence and integrity.

The guarantees, other than the guarantee of integrity, on which purely relational DBMSs have also focused concern: (1) Data independence, (2) Confidentiality, (3) Simultaneous access to data, (4) Data security after an incident, (5) Performances in terms of possibilities of handling high volumes of data and in terms of data access speed, (6) The adequacy of the access interface for data manipulation according to the relational approach.

After the adoption of those purely relational DBMSs, new types of applications with new requirements that those purely relational DBMSs are unable to satisfy have emerged on the web.

Indeed [9], to guarantee data consistency and to avoid storage anomalies, the primary purpose of the techniques used in the design of a pure relational database is to eliminate, through a normalization process, any possibility of redundancy at the data level. For those new types of applications, it is, on the contrary essential to use redundancy and distribution to guarantee, on the one hand, the availability of the data whenever a failure happens and, on the other hand, the

horizontal scaling of the services that the DBMSs offer to those applications.

In addition, to avoid the occurrence of anomalies generated on the data by the concurrency of the execution of the transactions and to ensure that recovery is possible whenever an incident occurs, pure relational DBMSs provide transactions with a runtime environment whose properties ensure ACIDity, i.e., Atomicity, Consistency, Isolation, and Durability. When the data are duplicated and distributed, DBMSs should, instead, provide a runtime environment with properties to ensure Consistency (C), Availability (A) for reads and writes, and tolerance for Network Partitioning (P). Brewer's CAP theorem states that a NoSQL DBMS can simultaneously provide only two properties [10].

As a result, new types of DBMSs (the object-oriented DBMSs [8] and the NoSQL DBMSs [9]), purely non-relational, have been designed to provide an adequate response to the rigidity of the pure relational model and those new requirements by relying on the alternative data models of the pure relational model and by moving away from its strengths in favor of the improvement of performances, of the scaling, and of the availability.

Simultaneously, an evolutionary approach has also emerged. The ambition of that approach is to integrate new features in the relational DBMSs to consider the requirements relating to the logical structure of databases [13,14] as well as the requirements relating to performances, scaling, and very high resistance to failures [15,16] while preserving the benefits of the pure relational model.

The guarantees pursued in that evolutionary approach concern: (1) Data independence, (2) Flexibility to work with various non-relational data models, (3) Data consistency, (4) Efficiency of data physical access, (4) Application scaling, (6) Availability for read and write operations, (7) Resistance to network partitioning whenever an outage happens.

In this paper, we focus on the ANSI/SPARC architecture [17] on which that evolutionary approach should rely to allow relational DBMSs to offer each application the flexibility to work with its preferred non-relational data model without sacrificing the guarantees of independence, integrity, and efficiency of evaluating queries on the database.

The explicit adoption of the restrictions of this ANSI/SPARC architecture of the schemata by a DBMS leads applications to manipulate a database by relying on a non-relational model, whereas the data model used to model that database internally is the pure relational model.

Take the view that data mapping means: redefining, using another data model, the data representation defined by relying on a given data model. The contribution of this paper is that it shows that with this ANSI/SPARC architecture of the schemata, it is possible to implement a process of data mapping that allows applications to manipulate a database according to the approach they prefer, whereas internally, that database is manipulated at the logical and physical levels as if it was a nested-relational database, with all the resulting benefits on the independence, integrity, and query evaluation efficiency guarantees. That process of data mapping is performed by relying on a relational database organized at the logical and physical levels for fast access to the abstract data that describe the complex real-world entities and by reusing the frameworks developed as part of relational database technology, thus preserving the results of investments made around that technology since its advent in 1970.

In the following, we address successively: (1) For illustration purposes, the rules commonly used to redefine the representation of a database, defined by relying on the UML conceptual model, using, on the one hand, the pure relational model as described in SQL2, and on the other hand, the object-relational and XML models as described in SQL3, while focusing on the consequences on the integrity of relationships, (2) The logic behind the ANSI/SPARC architecture of schemata on which evolutionary approaches must rely, (3) The mapping of the data that this architecture allows to implement at the logical and physical levels, (4) The comparison of that architecture where the database used at the logical level is modeled using the pure relational model to architectures where it is modeled using an extended relational model, XML, JSON, or a data model that leads to a logical implementation using the nested relational model.

II. Rules for Transforming a Conceptual Schema into a Relational Logical Schema

Fig.1 outlines the essential rules commonly used to derive a relational logical schema from the UML conceptual schema of a database [18]. In those derivation rules, the materialization of the relationships of the conceptual schema and the materialization of the deductible integrity constraints of their cardinalities, as well as the materialization of inheritance relationships, are carried out using foreign keys whose semantics allow the DBMSs to ensure the data integrity as defined by the conceptual schema.

In those rules, foreign keys are underlined and marked with the symbol "*" when they must respect the unique integrity constraint or the symbol "+" otherwise.

A relational logical schema derived from those rules describes the perception that enables database manipulation according to the relational approach and the perception required to guarantee data independence and integrity.

III. Rules for Transforming a Conceptual Schema into an Object-Relational or XML Logical Schema

The main rules [18], based on the semantics of the new concepts of SQL3 [13], for transforming a UML conceptual schema into an object-relational schema are defined in Fig.2. Rules that involve logical pointers are surrounded. Those rules are intended to facilitate the manipulation of the database according to the object-oriented approach.

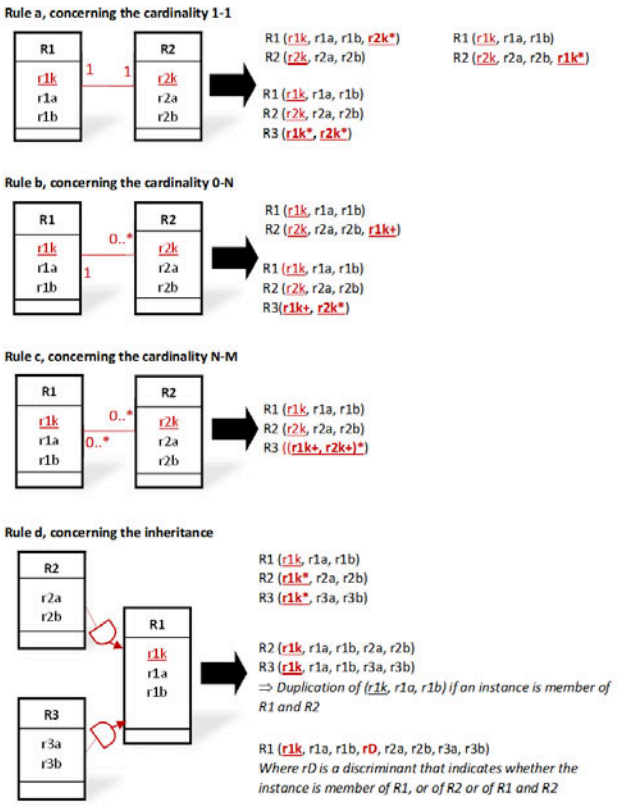


Fig. 1: Main Rules for Transforming a UML Conceptual Schema to a Relational Logical Schema

In those rules of Fig. 2:

1. "Ri (ria, rib, ...)" should be interpreted as denoting an object-relational table named "Ri" characterized by its atomic, composed, or relation-valued attributes named "ria, rib, ... ".
2. "RiTuple (ria, rib, ...)" should be interpreted as denoting a composed attribute characterized by its atomic, composed, or relation-valued attributes named "ria, rib, ... ".
3. "RiTupleSet (ria, rib, ...)" should be interpreted as denoting a relation-valued attribute containing a set of tuples characterized each by its atomic, composed, or relation-valued attributes named "ria, rib, ... ".
4. "RiRef" should be interpreted as denoting an atomic attribute containing a reference to an object of the object-relational table "Ri", i.e., a logical pointer to that object.
5. "RiRefSet (:RiRefType)" should be interpreted as denoting a relation-valued attribute containing a set of references to the objects of the object-relational table "Ri", i.e., a set of logical pointers to those objects, where the type of those logical pointers is denoted by "RiRefType".
6. "rik*" should be interpreted as denoting a foreign key corresponding to the primary key of the table "Ri" that respects the uniqueness integrity constraint.
7. "rik+" should be interpreted as denoting a foreign key corresponding to the primary key of the table "Ri" that does not respect the uniqueness integrity constraint.
8. "RiRefSet (rik)" should be interpreted as denoting a relation-valued attribute containing a set of tuples where each tuple contains a value of the foreign key "rik" corresponding to the primary key of the table "Ri".

The rules based on foreign keys are applicable for deriving an XML logical schema from a conceptual schema, as defined in SQL3 [14], by considering that the tags of the XML elements and attributes have been

omitted for each XML document. In addition, it should also be considered that for an object-relational table named "Ri", the type of an attribute can be XMLTYPE and thus contain an XML document stored (as an atomic value) with all its tags in the binary format of XML and managed using the technology of the document-oriented DBMSs. Therefore, in an object-relational table, the content of an object or attribute can be, respectively, an XML document shredded in several attributes, or an XML document stored in a single attribute.

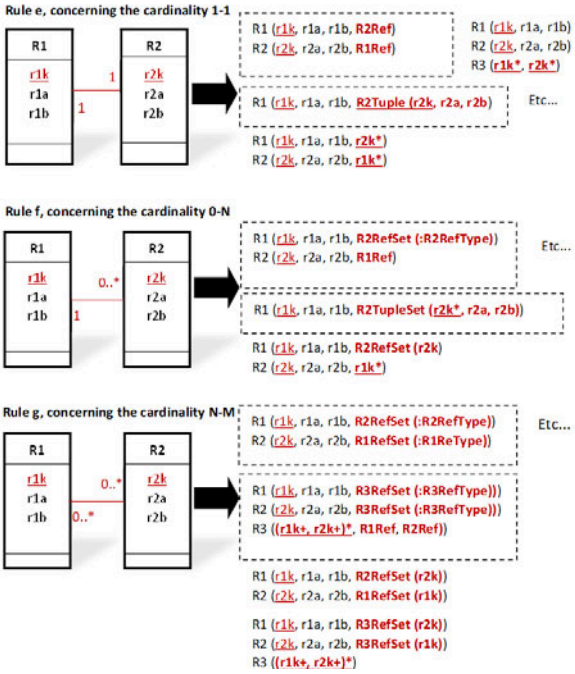


Fig. 2: Main Rules for Transforming a UML Conceptual Schema to an Object-Relational or XML Logical Schema

Those rules depict that the object-relational model and the XML model give the possibility of materializing each relationship by a set of pairs of unidirectional semantic links where in each pair, each link must be the inverse of the other, despite that the semantics of the concepts used, namely the concepts of logical pointer and foreign key, does not say how DBMSs should do to ensure the integrity of those links.

Furthermore, the data's logical structure based on the object-relational and XML models arises from design decisions that depend on both the data's semantics and the ways the user's applications intend to process those data but not only on the data's semantics, as for data's logical structure based on the relational model. After a change in the hierarchical structure of the data in the logical schema, the developer may be forced to modify the application logic accordingly [19].

IV. The Logic Behind the Architecture of the Schemata in the Evolutionary Approach

Fig. 3 schematizes the ANSI/SPARC architecture of the schemata on which evolutionary approaches should rely. One of the key features of that architecture is that it forces to decouple, on the one hand, the description concerning the perception that aims at facilitating data manipulation by applications and, on the other hand, the description of the perception induced by the database used internally for ensuring independence and integrity.

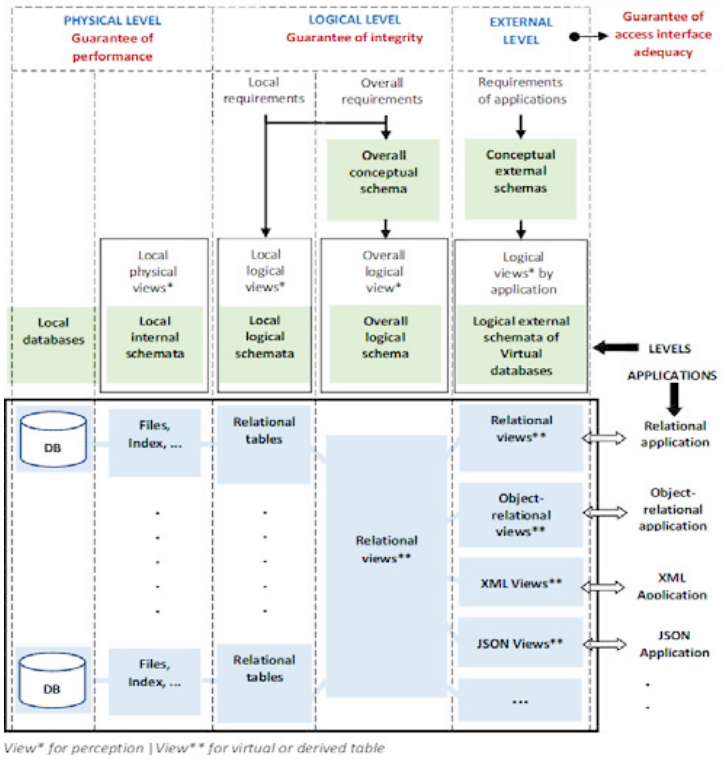


Fig. 3: The ANSI/SPARC Architecture of the Schemata, Which Provides Integrity and Independence Guarantees

Since the pure relational model is the one where T, S, O, and C offer the most possibilities to guarantee independence and integrity, that architecture leads to doing so that the perception induced by the database used internally to ensure independence and integrity can only be described using a pure relational logical schema.

As for the perception that aims to facilitate the manipulation of the database by an application, that architecture makes it only possible to describe that perception by relying on an external logical schema and using a data model where T, S, O, and C offer the required flexibility to model and manipulate the database used internally according to the approach that the application prefers, regardless of the possibilities provided by C.

Depending on the query, data manipulated by that application according to its preferred approach are determined dynamically and efficiently (see paragraphs 6 and 8) from the data stored in the pure relational database used internally to guarantee independence and integrity.

That ANSI/SPARC architecture of the schemata then leads to data mappings at the logical and physical levels. Those data mappings are required: (1) when designing the relational overall logical schema of the database, (2) when designing a physical database, (3) when designing the external logical schema of an application, (4) and when generating and optimizing the logical execution plan of a query.

V. Data Mapping in the Design Process of the Overall Relational Logical Schema

The overall relational logical schema that defines a database used internally intended to be manipulated according to non-relational approaches can be derived simply from an overall conceptual schema by bringing out in that conceptual schema the whole-part relationships arising from the perception of the real-world's complex entities. For each real-world's complex entity denoted by "E1", all the other entities whose existence of their instances depends on an instance of "E1" must be identified by relying, as proposed in [3], on the concept of "existence dependency" of one entity on another and the concepts of "regular" (entity/relationship) and "weak" (entity/relationship).

Let's take the real-world's entities "Employee" and "Child" as examples. Each instance of "Child" depends on the existence of one instance of "Employee". Therefore, the entity "Child" is a weak entity. In an overall relational logical schema derived from an overall conceptual schema, the tabular relations derived from the existence dependency of "Child" on "Employee" must be "Employee (E-no, E-name, E-age, …)' and "Child (E-no+, Child-name, Child-age, …)' where "E-no" in the tabular relation denoted by "Child" is a foreign key. Notice that in Fig.1, this is a variant of the first of the two transformation possibilities defined by rule b, where the

foreign key is part of a primary key. That rule allows us to describe, in the relational logical schema, the data about each complex entity of the real world using a hierarchy of tabular relations linked by foreign keys, where each abstract data that describes one instance of that complex entity must be stored in a hierarchy of rows distributed in those tabular relations.

The purpose of that rule, which says how to map each instance of a complex entity in the relational database used internally, is to result, as described in the next paragraph 6, in a physical organization of that database, which can guarantee fast access to each abstract data that describes an instance of a complex entity.

VI. Data Mapping in the Design Process of the Physical Database

The primary purpose of the design process of a physical database is to make explicit how each abstract data describing the instance of one complex entity should be organized and stored on the physical storage device to make physical access to that abstract data fast. As we saw in paragraph 5, each abstract data is stored in a hierarchy of rows distributed in tabular relations connected by foreign keys.

To achieve this, the two leading families of techniques developed within the framework of the pure relational DBMSs for the data physical storage structures on the disks can be reused and improved, i.e., physical clustering and indexes.

a) Physical Clustering

The main choices, which can make fast physical access to each abstract data describing one instance of a complex entity, are:

1. Creating a table cluster for each regular (entity/relationship) defined in the conceptual schema. The role of such a cluster is to group all the tabular relations connected by foreign keys, where the abstract data about the instances of that regular (entity/relationship) must be stored.
2. Storing each abstract data made of a hierarchy of rows (scattered across those tabular relations) in one or more contiguous pages of that cluster. That storage renders unnecessary the need to perform join operations for grouping those rows, which reduces to a strict minimum the average time to access all or part of each complex abstract data.
3. Implementing logically and physically those table clusters using the nested relational model [7]. The main benefit of that implementation is that this data model expands the sets T, S, and O of the relational model to overcome its limitations. As a result, this data model allows: (i) to describe each complex entity of the real world using a not decomposed complex abstract data (defined as a whole-part), (ii) to define a simple nested expression to recursively apply the selection and projection operators to attributes nested at any level in the structure of that abstract data, (iii) to simplify that nested expression logically, and (iv) to evaluate, at the lowest possible cost, the resulting optimized nested expression, without any join operations.

When evaluating a query, those choices allow us to manipulate logically the database used internally as if it was a nested relational database where each table cluster is represented using a single nested relational table where the content of each row is a logical implementation of an abstract data stored in a hierarchy of rows of the tabular relations of the relational database, with all the benefits resulting in terms of performance.

To physically store in one or more contiguous pages of a cluster one abstract data consisting of a hierarchy of rows distributed in the tabular relations of that cluster and logically implemented in a row of a nested relational table representing that cluster, it is possible to logically group those rows in a data structure corresponding to a tree-like data structure having the same hierarchical organization as that hierarchy of rows, by following the data organization rules of the nested relational model. In doing so, each node of that tree must contain the pointers to its parent, children, and siblings and the corresponding row in this hierarchy of rows. This leads to a physical implementation (based on the nested relational model) of each abstract data about an instance of a complex entity in one or more contiguous pages.

Fig. 4.a contains as an example the conceptual schema, defined using the entity-relationship model [3], of a complex regular entity denoted by "rx" on which depend the weak entities denoted by "ry", "rz", and "rw".

The schemata of the tabular relations of the cluster, required for storing the abstract data about the instances of that entity "rx", are ' rx (A), ry (A+, B), rz ((A, B)+, C), rw (A+, D, E)'. As for the schema of the nested relational table that represents that cluster according to the syntax proposed in [7], it can be defined as follows: rx (A, ry (B, rz (C)), rw (D, E)).

Fig. 4.b schematizes, as an example, in one page of that cluster, the data layout of one abstract data about an instance of that entity "rx". This figure depicts that this instance of that complex entity "rx" is described by a hierarchy of rows consisting of one row from the tabular relation "rx", two rows from the tabular relation "ry", three rows from the tabular relation "rz", and two rows from the tabular relation "rw", connected by foreign keys. This figure also depicts that those rows are stored in a tree-like data structure according to the rules of the nested relational model. Each node contains a row consisting of atomic columns, subsets of pointers to child nodes where each subset corresponds to a relation-valued column, and pointers to the parent node and sibling nodes for easing navigation.

Physical storage of the rows of the relational database used internally inside the pages of the physical storage devices by following that approach is an improvement of the traditional approach. In that approach, rows are not stored only sequentially (ordering them or not). Rows are also logically organized within the pages in terms of tree-like data structures for rows that concern complex abstract data and linear data structures for rows that concern the same tabular relation.

The indexing of a tabular relation as well as access using an index to all or part of one abstract data stored inside a page within a tree-like data structure can be achieved, on the one hand, by creating a dictionary of table clusters containing, for each table cluster, the identifier of each table of that cluster, the type of that table (regular or weak), and its primary key, and on the other hand, by allocating, in the header of each page, an entry for each row within that page containing the identifier of the concerned table, the value of the primary key of that row, and its beginning address within that page.

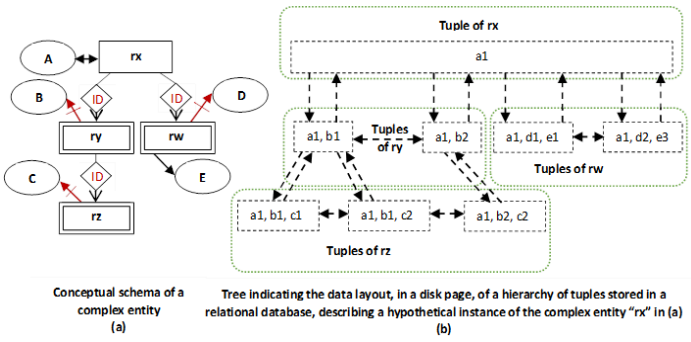


Fig. 4: Illustration of the Storage Structure in a Disk Page of Data about a Complex Entity Stored Logically in a Relational Database

b) Indexes

Depending on a query complexity on a virtual database, its evaluation in an efficient way may require at the physical level to access the tabular relations where data are stored by following combinable different approaches:

1) A sequential approach allowing to traverse the rows of a tabular relation in a predetermined order.
2) An associative approach allowing access in a tabular relation to a set of rows qualified by the values of a subset of columns.
3) A navigational approach, allowing navigation from one row of a tabular relation towards rows in other tabular relations related in some way.
4) An approach, path expression oriented, allowing for each given path "Ti.a1.a2….an. x", derived from a hierarchy of class (or user-defined structured type) attributes, as that hierarchy is defined in [20], to determine either the set of rows of the tabular relation "Ti" or the set of the instances (either partial or not) of that path, related in the two cases to a given value of the attribute "x".

At the physical level, many techniques for creating access paths to tabular relations have been defined to fulfill those needs. The best known are join indexes [21] (essential for dynamically materializing relationships of complex entities or for performing table cluster joins efficiently), pointer chains [22], secondary indexes based on B+ trees [23] and dynamic hashing [24-25], and bitmap indexes [26].

An expression of the nested relational algebra derived from a query (as that derivation is described in paragraph 8) can be executed efficiently thanks to, on the one hand, the physical organization of the database described in this paragraph, and on the other hand, extended frameworks of the relational DBMSs.

VII. Data Mapping in the Design Process of the External Logical Schema of an Application

When a DBMS adheres to the ANSI/SPARC architecture of schemata, which is schematized in Fig.3, the key role of the external logical schema of each application is to define a virtual database to allow that application to work with the data model it prefers. That virtual database is defined by creating, using a user-defined structured type, a custom-typed view for each regular (entity/relationship) defined in the conceptual schema, by following the derivation rules of Fig. 2.

To make it possible to manipulate that virtual database as if it was a nested relational virtual database, each of those typed views must be redefined using a schema that extends the schema of the nested relational table representing the cluster created at the physical level for the corresponding regular (entity/ relationship). The extension of that schema is carried out by adding the necessary attributes for the materialization of relationships.

Each cluster can therefore be considered as being the materialization of a typed view without the materialization of the relationships defined in the conceptual schema.

Consequently, the instances of a typed view in a virtual database do not have real existence as in the typed tables of a non-relational database. Those instances are derived dynamically from the tabular relations of the relational database used internally to ensure independence and integrity by relying on the data physical organization that results from using cluster and index concepts to make required computing fast.

More concretely, the data mapping in the design process of an external logical schema of an application is carried out by the database's administrator through the SELECT statements defining the typed views of the virtual database of that application. The SELECT statement defining a typed view must indicate how each instance of that typed view should be calculated dynamically using the relevant data stored in the relational database used internally for ensuring data independence and integrity. The role of triggers associated with a typed view must be to dynamically support the update operations of the instances of that typed view ("INSERT", "UPDATE", "DELETE") when the means put in place do not allow the DBMS to provide that support because this involves several tabular relations.

To facilitate the work of the databases' administrators, DBMS providers should instead consider offering the following possibility: indicating equivalently using new clauses or annotations in specification statements of the user-defined structured types, how for each user-defined structured type the value of each attribute of its instances must be calculated from the data stored in the relational database used internally for ensuring data independence and integrity, mainly concerning the relationships materialization. When evaluating a query formulated on a virtual database, the information contained in those specifications must provide the DBMS with the same possibilities as the SELECT statements and triggers associated with the typed views of that virtual database. This kind of approach for the definition of data mapping is followed, for example, in the object-relational mapper of Hibernate [27] and for storing shredded XML documents into object-relational tables [30] or relational tables [31].

For short illustrative purposes, by relying on a subset of rules presented in paragraphs 2 and 3, this can be achieved by specifying, for example.

1. For each user-defined structured type "R1" concerned: the type of the corresponding entity or relationship (regular or weak) in the conceptual schema, the name of the main table "T1p" of the overall relational logical schema, which is used for the derivation of its instances.
2. For each atomic-valued attribute "ai" of that user-defined structured type "R1", whose type is a basic scalar type: the name of the column corresponding to it in the main table "T1p".
3. For each atomic-valued attribute "ai" of that user-defined structured type "R1", which is used to materialize a "1-1" relationship between "R1" and "R2" and whose type is a logical pointer type "R2RefType": the type of the relationship concerned, the name of the main table "T2p", which is used for the derivation of the instances of the structured type "R2", the name of the foreign key in "T1p", which refers to the main table "T2p" and which is used to calculate the value of that attribute "ai" corresponding to a logical pointer to an instance of "R2".
4. For each atomic-valued attribute "ai" of that user-defined structured type "R1", which is used to materialize an "N-1" relationship between "R1" and "R2" and whose type is a logical pointer type "R2RefType": the type of relationship concerned, the name of the main table "T2p", which is used for the derivation of instances of the structured type "R2", the name of the foreign key of "T1p", which refers to the main table "T2p" and which is used to calculate the value of that attribute "ai" corresponding to a logical pointer to an instance of " R2".
5. For each relation-valued attribute "ai" of that user-defined structured type "R1", which is used to materialize a "1-N" relationship between "R1" and "R2" and whose type "R2RefSetType" is a collection type of values of logical pointers: the type of relationship concerned, the name of the main table "T2p", which is used for the derivation of the instances of the structured type "R2", the name of the foreign key in "T2p", which refers to the main table "T1p" and which must for each instance of "R1" be used for the derivation of the value of that attribute "ai" corresponding to a collection consisting of logical pointers to the instances of "R2", which point back to that instance of "R1".
6. For each relation-valued attribute "ai" of that user-defined structured type "R1", which is used to materialize a "1-N" relationship corresponding to a composition relationship between "R1" and "R2" and whose type "R2TupleSetType" is a collection type of the instances of the structured type "R2": the type of the relationship concerned, the name of the main table "T2p", which is used for the derivation of the instances of the structured type "R2", the name of the foreign key in "T2p", which is part of its primary key and which refers to the main table "T1p" and which must for each instance of "R1" be used to calculate the value of that attribute "ai" corresponding to a collection consisting of the instances of the structured type "R2" linked to that instance of "R1".

7. For each relation-valued attribute "ai" of that user-defined structured type "R1", which is used to materialize an "N-M" relationship between "R1" and "R2" (using the user-defined structured type "R3") and whose type "R3RefSetType" is a collection type of values of logical pointers: the type of the relationship concerned, the names of the main tables "T2p" and "T3p", which are used for the derivation of the instances of the structured types "R2" and "R3", the name of the foreign key in "T3p" being part of its primary key, which refers to the main table "T1p", as well as that of the foreign key in "T3p" being part of its primary key, which refers to the main table "T2p", which both must be used for each instance of "R1" to calculate the value of that attribute "ai" corresponding to a collection consisting of the logical pointers to the instances of "R3", which point back to that instance of "R1".

This approach gives the DBMS the capability to play its role fully.

1. By ensuring the consistency of those specifications when creating a user-defined structured type in its catalog.
2. By ensuring, when creating each typed view, the automatic generation of the nested expression, which calculates the instances of its structured user-defined type from the tabular relations of the overall relational logical schema (including for each relationship concerned the calculation of its semantic links by ensuring that each link has an inverse link).
3. By ensuring, when executing the update operations of the instances of a typed view ("INSERT", "UPDATE", "DELETE")), the transformation of those operations into processes of updating the tabular relations from which those instances are derived.

VIII. Data Mapping when Generating and Optimizing Query Logical Execution Plan

Paragraphs 6 and 7 showed that it is possible at the logical and physical levels to represent a virtual database (defined by the external logical schema of an application) using the nested relational model. This paragraph shows how that possibility enables fast execution of queries formulated on that virtual database. Formulating a SQL query for dynamic calculation of the instances of a typed view, as described in paragraph 7, may require using the various possibilities that SQL offers for nesting other queries, particularly in the SELECT clause but also in the FROM and WHERE clauses [28].

As [32] shows, DBMSs can represent this kind of SQL query containing nested queries by an expression of the nested relational algebra that defines a logical execution plan that uses tabular relations. Conceptually, each instance of a typed view calculated using this kind of expression of the nested relational algebra can be manipulated as if it was stored in a nested relational table, even though the value of each of its attributes is determined dynamically from values contained in the tabular relations used internally to ensure independence and integrity guarantees.

Depending on its type, by applying the transformation functions provided for this purpose to the value of each attribute determined dynamically, those instances of a typed view can be perceived and manipulated by the developer as structured abstract data corresponding either to objects or to XML documents, for example.

Consider as an example the SQL statements in row 1 of Table 1 about the creation in the catalog of an Oracle DBMS [28] of the tabular relations "CUSTOMERS" and "ORDERS", corresponding to regular entities.

The SQL statements [29] for creating the object-relational typed view "CUSTOMERS_VOR" on the regular entity "CUSTOMERS" can be defined as in row 2 of Table 1.

The expression in row 3 of Table 1 is a nested expression derived for illustrative purposes from the SELECT statement that defines that typed view.

That expression is a logical execution plan for dynamically calculating all objects in the typed view named "CUSTOMERS_VOR". It calculates each instance of that typed view so that that instance can be perceived as if it was stored in a nested relational table. The values of the atomic-valued attributes named "CUSTNO" and "CUSTNAME" of that instance are defined as being the values of the columns having the same name in a row about a customer in the tabular relation named "CUSTOMERS". As for the value of its relation-valued attribute named "CUSTORDERS", it is defined as the result of projecting on column "ORDERNO" all the rows about that customer, selected in the tabular relation "ORDERS".

Each instance calculated using that expression is matched to an object of the typed view "CUSTOMERS_VOR" simply by applying the ORACLE transformation function "MAKE_REF()" to each element of the relation-valued attribute named "CUSTORDERS".

Each query formulated using the data manipulation language of the data model used to describe a virtual database can be translated by the kernel of the DBMS to a logical execution plan based on nested relational algebra, using techniques comparable to those developed for standard SQL.

For example, consider the query in row 4 of Table 1 formulated using the extended SQL language of the object-relational model on the virtual database containing the typed view "CUSTOMERS_VOR".

The expression in row 5 of Table 1 is an optimized expression of the nested relational algebra

derived from that query that uses only the typed views of that virtual database.

The substitution in that expression of "CUSTOMERS_VOR", which denotes the name of a typed view of that virtual database, by the expression in row 3 of Table 1, which represents the SELECT statement of that typed view, leads to the expression in row 6 of Table 1, which uses only the tabular relations of the relational database used internally.

Simplifying that expression results in the expression of row 7 of Table 1, which corresponds to the optimized logical execution plan of the initial query, generated by considering the internal representation in tabular relations of the abstract data perceived and manipulated by the developer in his query as being objects.

IX. Related Work

We exclude from the scope of our analysis the object-relational mapping [27] that the deployment technologies of the web applications perform, outside the DBMS, to ensure, from a program, the persistence of the objects in a relational database. That very popular data mapping is a functionality of those technological offerings and not of the technology of the databases. Non-relational databases are intended to render obsolete the need for this mapping which is detrimental to performance.

The following compares the architecture of the schemata of Fig.3, where the model at the logical level is the pure relational model, to four broad categories of architectures where the model at the logical level is: (i) an extended relational model where in the tables the relation-valued columns contain nested tables, (ii) an extended relational model where in the tables the relation-valued columns contain XML documents or JSON documents, (iii) XML or JSON, (iv) leads to a logical implementation using nested relational tables.

In our comparison, we focus on the following three aspects: (i) the guarantee of independence, (ii) the guarantee of integrity, (iii) and the guarantee of performance when evaluating a query on the database.

a) The Guarantee of Independence

The approaches to designing a database commonly used are based on two major schools of thought.

In the first stage of the oldest school, the database is modeled independently of the individual perception of applications. In this first stage, the designer's primary objective is to define, using a conceptual model, the perception of the actors of the enterprise independently of any technological choice. That school is best suited when integrity is a fundamental requirement and handling the difference between data types and data structure types required by different applications is an uphill task.

The second school developed recently at the same time as NoSQL DBMSs. In that school, the database is modeled at the same time as the design of an application by relying on a flexible data model, such as the document-oriented model, to consider the specific needs of that application in terms of data organization as soon as they appear or are challenged. That school is best suited for an application when handling the variability in data types and data structure types is an uphill task.

Table 1: Examples of Paragraph 8 for Illustrating the Transformation of a Query to a Logical Execution Plan

1	CREATE TABLE CUSTOMERS (CUSTNO NUMBER (5) PRIMARY KEY, CUSTNAME VARCHAR2 (30) NOT NULL) ; CREATE TABLE ORDERS (ORDERNO NUMBER (8) PRIMARY KEY, ORD_CUSTNO NUMBER (5) NOT NULL REFERENCES CUSTOMERS (CUSTNO)) ;
2	CREATE TYPE CUSTOMER_T ; CREATE OR REPLACE TYPE ORDER_T AS OBJECT (ORDERNO NUMBER (8), ORDERCUST REF CUSTOMER_T) ; CREATE OR REPLACE TYPE ORDER_T_LIST AS TABLE OF REF ORDER_T ; CREATE OR REPLACE TYPE CUSTOMER_T AS OBJECT (CUSTNO NUMBER (5), CUSTNAME VARCHAR2 (30), CUSTORDERS ORDER_T_LIST) ; CREATE OR REPLACE FORCE VIEW CUSTOMERS_VOR OF CUSTOMER_T WITH OBJECT IDENTIFIER (CUSTNO) AS SELECT c.CUSTNO, c.CUSTNAME, CAST (MULTISET (SELECT MAKE_REF (ORDERS_VOR, o.ORDERNO) FROM ORDERS o WHERE o.ORD_CUSTNO = c.CUSTNO) AS ORDER_T_LIST) AS CUSTORDERS FROM CUSTOMERS c ;
3	π [CUSTNO, CUSTNAME, π [ORDERNO] (σ[ORD_CUSTNO = CUSTNO] (ORDERS)): CUSTORDERS](CUSTOMERS)
4	SELECT c.CUSTNO, c.CUSTNAME FROM CUSTOMERS_VOR c WHERE c.CUSTNO = 100;
5	π [CUSTNO, CUSTNAME] (σ[CUSTNO =100] (CUSTOMERS_VOR))
6	π [CUSTNO, CUSTNAME] (σ [CUSTNO=100] (π [CUSTNO, CUSTNAME, π [ORDERNO] (σ[ORD_CUSTNO=CUSTNO] (ORDERS)): CUSTORDERS] (CUSTOMERS)))
7	π [CUSTNO, CUSTNAME] (σ[CUSTNO=100] (CUSTOMERS))

The architectures of categories (i), (ii), (iii), and (iv).

Data mapping is typically accomplished in two steps within those architectures.

In the first step, data mapping consists of deriving the perception that facilitates the manipulation of the database according to the preferred non-relational approach from the conceptual schema or the kind of use case concerned. The designer's primary objective is to meet the requirements of developers by eliminating the drawbacks that arise from impedance mismatches and by adapting the data logical structure to how the processing is conducted.

In this first step, the way the database can be perceived and manipulated is captured using user-defined structured types that allow the information system to be perceived as being made of abstract data describing instances of complex entities in the real world, optionally by using flexible schemata, easy to modify.

In the second step, data mapping consists of deriving from the user-defined structured types a storage structure for a logical implementation of the instances of those user-defined structured types that can facilitate a physical implementation on the storage devices.

For architectures where the model at the logical level is an extended relational model, XML or JSON, this amounts in the first step to defining a non-relational logical schema and in the second step to implementing each non-relational table using a set of tabular relations where all columns are atomic-valued, optionally by storing in a binary format the XML and JSON documents. As in the architecture of Fig.3, this amounts to storing the data about each instance of a complex entity in several tabular relations connected by foreign keys. As a result, this makes it possible to reuse the frameworks of the pure relational DBMSs for query optimization and evaluation. When XML and JSON documents are stored in binary format, it is also possible to rely on a hybrid system that integrates the required features for modeling the complex data concerned according to the approach of the second school.

For architectures where the model at the logical level leads to a logical implementation using nested relational tables, this comes down in the first step to defining a non-relational logical schema that meets the developers' requirements and in the second step to derive a nested relational database from that non-relational logical schema.

Therefore, in those architectures of categories (i), (ii), (iii), and (iv), the logical schema of the database always stems from the perception required for a particular application and can make it more complex for another application to manipulate that database.

The architecture where the database is modeled internally using the relational model.

As far as this architecture is concerned, it is instead the derivation of the database used internally that is carried out first from the conceptual schema before the derivation for each application of the perception that facilitates the manipulation of that database according to the required approach. The primary goal is to define, regardless of the individual perception of the applications, the data logical organization, the integrity constraints, and the data physical organization that can guarantee fast access to each abstract data that describes one instance of a complex entity in the real world. It is more in line with the key objective of the ANSI/SPARC architecture of the schemata, which is to allow DBMSs to provide data independence guarantees. As a result, compared to architectures where the data model used internally is an extended relational model, XML, or JSON, that architecture ensures greater data independence. Additionally, that architecture allows partitioning the database into datasets, for which the best school for the design can be considered independently.

b) The Guarantee of Integrity

Among the architectures considered in this paragraph, the most common are those of categories (i), (ii), and (iii) that implement SQL3 and those of category (iii) used by document-oriented NoSQL DBMSs. Those architectures do not allow to guarantee the integrity of relationships defined in conceptual schemata when those relationships are materialized using pairs of semantic links where in each pair, each link must be the inverse of the other. The reason for that significant drawback is that the semantics of the concepts used for materializing those relationships, namely the concepts of foreign key and logical pointer, do not indicate how DBMSs can guarantee the integrity of those relationships. To overcome that shortcoming, the architecture presented in that paper materializes those relationships in a virtual database. The resulting calculation is achieved (using rules that can be predefined and played by the DBMSs) using relevant data from the relational database used internally to ensure independence and integrity. The alternatives are: (i) either to materialize each relationship using in the logical schema two functions, such as each is defined as the inverse function of the other [8], (ii) or to define and implement path constraints [33].

Furthermore, the set of types of simple and complex integrity constraints that may be defined in a pure relational logical schema is a superset of each set of the types of integrity constraints that can be defined in a logical schema of an architecture of the categories (i), (ii), (iii), and (iv). One of the main benefits of the architecture of Fig.3 is that it allows the designer to define the custom view of an application by relying on any data logical model on which the architectures of the categories (i), (ii), (iii), and (iv) rely and by enforcing

types of integrity constraints that this data logical model does not consider thanks to the overall relational logical schema of Fig. 3. For some use cases, this is an alternative to extending the set C of the non-relational data logical models with complex integrity constraint types that are difficult to express and enforce.

c) *The Guarantee of Performance when Evaluating a Query on the Database*

One of the main benefits of the schemata architecture of Fig. 3 is that it allows applications to manipulate the data according to the approach they prefer, whereas internally, that data is manipulated at the logical and physical levels as if it were stored in nested relational tables. In other words, this allows at the logical and physical levels to manipulate as efficiently as possible the data that describe the complex entities in the real world as if they have not been broken down and distributed in tabular relations.

In the architectures where the model at the logical level is an extended relational model, XML, or JSON, when the data describing complex entities in the real world are stored internally in several tabular relations, on the contrary, this makes query evaluation inefficient because of the join operations that can result.

It should be noted, however, that for the architectures where the model at the logical level is the pure relational model, an extended relational model, XML, or JSON, when the XML or JSON documents are stored internally in the binary format, the resulting performance benefits are those recognized for document-oriented DBMSs such as MongoDB.

The architectures where the model at the logical level leads to a logical implementation using nested relational tables make it possible to use nested relational algebra for better logical optimization of queries. Those architectures also lead to an implementation of the nested relational tables on the storage devices by storing each row of those tables (corresponding to one abstract data describing an instance of a complex entity) in one or more contiguous pages using a format close to that of the nested relational model. This ensures efficient query evaluation but requires implementing a new manager of abstract data (i.e., a new complex storage engine) responsible for providing an interface for manipulating those complex abstract data at the logical level and managing their physical storage using clustering techniques.

X. Conclusion

In this paper, we focused on the features that should be built into relational DBMSs so they can provide applications with the flexibility to work with the non-relational data model they prefer without sacrificing guarantees of independence and integrity, as well as the guarantee of query performance. One of the most critical aspects of the features that have been integrated into the relational DBMSs in recent years to meet the requirements of horizontal scaling and very high availability [15, 16] concerns the resulting level of performance when evaluating queries that require a considerable number of join operations. This aspect can be addressed more effectively by dynamically applying the sharding and distribution techniques to table clusters instead of directly applying those techniques to tabular relations. That choice would allow extended relational DBMSs to give applications the flexibility to work with their preferred data storage models without sacrificing any benefit of the technology of the databases.

References Références Referencias

1. Eessaar, E.: Using Meta-modeling in order to Evaluate Data Models. *Proceedings of the 6th WSEAS Int. Conf. on Artificial Intelligence, Knowledge Engineering and Data Bases, Corfu Island, Greece.* February 16-19, 2007.
2. Object Modeling Group: *Unified Modeling Language Specification. Version 2.5*. October 2012.
3. Chen, P., P-S.: The entity-relationship model - toward a unified view of data. *ACM TODS Volume 1, Issue 1*. pp 9–36. March 1976.
4. Codd E. F.: *The relational model for database management. Second Edition*. Addison-Wesley Publishing Company, Inc., 1990.
5. Date, C.J.: *An Introduction to Database Systems, 8th Edition*. Pearson Education, Inc, July 2003.
6. Vidaković, J., & al: Extended Tuple Constraint Type as a Complex Integrity Constraint Type in XML Data Model – Definition and Enforcement. *Computer Science and Information Systems*. Oct. 2018. 15 (3): 821-843. DOI: 10.2298/CSIS123456789X.
7. SCHEK, H.-J., & al: The relational model with relation-valued attributes. *Information Systems, Volume 11, Issue 2. pp 137–147*. 1986. DOI: 10.1016/0306-4379 (86) 90003-7.
8. Cattell, R. G. G., & al: The Object Data Standard: ODMG 3.0. *Edited by R.G.G. Cattell and al Publishers, 1st edition* (February 2, 2000)
9. Wajid, A., & al: Comparison between SQL and NoSQL Databases and Their Relationship with Big Data Analytics. *Asian Journal of Computer Science and Information Technology* · October 2019. DOI: 10.9734/AJRCOS/2019/v4i230108.
10. Changlin, H.: Survey on NoSQL Database Technology. Journal of Applied Science and Engineering Innovation, Vol.2 No.2. 2015.
11. W3C Recommendations: XML TECHNOLOGY. https://www.w3.org/standards/xml/
12. Bourhis, P., & al: JSON: Data model, query languages and schema specification. *The 36th ACM SIGMOD-SIGACT-SIGAI Symposium*, May 2017.

13. ANSI/ISO/IEC International Standard (IS): Database Language SQL-Part 2: Foundation (SQL/Foundation) «Part 2». September 1999.
14. ISO-ANSI Working Draft: *Database languages, SQL, Part 14: XML-Related Specifications (SQL/XML)*. August 2002.
15. ORACLE: Oracle Database Using Oracle Sharding, 21c F32165-09. 2023.
16. MARIADB: *Distributed SQL: the Architecture behind MARIADB XPAND* https://mariadb.com/products/enterprise/xpand/. April 2021.
17. ANSI/X3/SPARC study group: The ANSI/X3/SPARC DBMS framework report of the study group on database management systems. *Information systems, volume 3, issue 3*. pages 173-191. 1978.
18. Soutou, C.: Modeling relationships in object-relational databases. *Data & Knowledge Engineering, Volume 36, Issue 1*. Pages 79-107. January 2001.
19. Codd, E., F.: A relational model of data for large, shared data banks. *CACM 13, No 6*. Jun 1970.
20. Bertino, E., & al: Indexing Techniques for Queries on Nested Objects. *IEEE Transactions on Knowledge and Data Engineering, 1 (2), pp 196 – 214*, July 1989
21. Valduriez, P.: Join Indices. *ACM TODS, Vol. 12, No. 2*, Pages 218-246. June 1987.
22. Astrahan, M., M., & al: System R: Relational Approach to Database Management. *ACM TODS, Vol. 1, No. 2., Pages 97-137*, June 1976.
23. Comer, D.: The Ubiquitous B-Tree. *Computing Surveys, vol. 11, n° 2*, June 1979.
24. Fagin R., & al: Extendible Hashing – A Fast Access Method for Dynamic Files. *ACM TODS, Vol. 4, No. 3, Pages 315-344*. September 1979 .
25. Litwin W.: Linear Hashing – A New Tool for File and Table Addressing. *6th Very Large Data Bases, Montreal, p. 224-232.* October 1980.
26. Chan, C-Y., al.: Bitmap index Design and Evaluation. *ACM SIGMOD Intl. Conf., SIGMOD Record V° 27, n° 2, Seattle, USA*, 1998.
27. Hibernate: Hibernate ORM, Documentation 6.2. https://hibernate.org/orm/documentation/6.2/
28. ORACLE: Oracle Database SQL Language Reference 12c Release 1 (12.1). July 2017.
29. ORACLE: Oracle Database Object-Relational Developer's Guide 12c Release 1 (12.1). August 2014.
30. ORACLE: Oracle XML DB Developer's Guide, 12c Release 2 (12.2). April 2019.
31. IBM Corp.: DB2 10 for z/OS, SQL Reference. October 13, 2017, edition.
32. Hölsch J., & al: Optimization of Nested Queries using the NF2 Algebra. *Proceedings of the 2016 International Conference on Management of Data - SIGMOD '16*. doi:10.1145/2882903.2915241.
33. Peter Buneman, P., & al: Path Constraints in Semistructured Databases. *Journal of Computer and System Sciences*. 2000. doi:10.1006/jcss.2000.1710.

This page is intentionally left blank

Global Journal of Computer Science and Technology: H Information & Technology
Volume 23 Issue 2 Version 1.0 Year 2023
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

A Smart Contract Blockchain Penetration Testing Framework

By Shyam Meshram & Isha Sood
University of Ajeenkya D.Y. Patil

Abstract- Likened to old-style contracts, smart agreements motorized by blockchain ensure that deal processes are real, safe, then well-organized. Without the need aimed at third-party mediators like lawyers, smart contracts enable transparent processes, cost-effectiveness, time efficiency, and trust lessness. While old-style cybersecurity attacks on keen agreement requests can be thwarted by blockchain, new threats and attack vectors are constantly emerging, which affect blockchain in a manner alike toward additional web and application-based systems. Organizations can develop and use the technology securely with connected infrastructure by using effective blockchain testing. However, the authors discovered throughout the sequence of their investigate that Blockchain technology has security issues like permanent dealings, insufficient access, and ineffective plans.

Keywords: smart contracts, attack vectors, cyber-security, blockchain, cyber threats.

GJCST-H Classification: LCC: QA76.9.B56

Strictly as per the compliance and regulations of:

A Smart Contract Blockchain Penetration Testing Framework

Shyam Meshram [α] & Isha Sood [σ]

Abstract- Likened to old-style contracts, smart agreements motorized by blockchain ensure that deal processes are real, safe, then well-organized. Without the need aimed at third-party mediators like lawyers, smart contracts enable transparent processes, cost-effectiveness, time efficiency, and trust lessness. While old-style cybersecurity attacks on keen agreement requests can be thwarted by blockchain, new threats and attack vectors are constantly emerging, which affect blockchain in a manner alike toward additional web and application-based systems. Organizations can develop and use the technology securely with connected infrastructure by using effective blockchain testing. However, the authors discovered throughout the sequence of their investigate that Blockchain technology has security issues like permanent dealings, insufficient access, and ineffective plans. Web portals and other applications do not contain attack vectors like these. This study introduces a brand new penetration testing framework for decentralized apps and clever contracts. Results from the suggested penetration-testing methodology were com-pared by those from automatic diffusion examination scanners by the authors. The findings revealed gaps in vulnerabilities that were not disclosed during routine pen testing.

Keywords: smart contracts, attack vectors, cyber-security, blockchain, cyber threats.

I. Introduction

Research into and adoption of blockchain technology has exploded across a wide range of businesses. Blockchain relies happening peer-to-peer dealings and is dispersed decentralized without any centralized authority or third-party involvement. Digital programmed scripts of codes known as Smart Con-tracts [1] are kept inside a Blockchain. Once sure sections [3] by particular predefined circumstances remain met, these programmed become anger resistant, being self-verifying, self-executing, and self-enforcing [2] numerical contracts. Smart Contracts are able to carry out transactions in real-time, for a small fee, and with a higher level of security [4]. Cryptocurrency nodes on the Blockchain network work toward inform the distributed, see-through ledger. All nodes view this inform, which remains checked [5] before it is accepted by the network.

Consider purchasing a new car as an illustration. The con-ventional process entails visiting a car trader (an intermediate 3rd party) and haggling over the car of your choice. Instead of involving the insurance company and the transportation department in the paperwork, as well as successful toward a bank for a car advance (yet additional 3rd party). Here is a waiting retro before the car is delivered after all formalities and payments have been made. This procedure requires patience and requires communication with numerous additional 3rd parties.

Presumptuous the similar car's information, possession, IDs, then proposal are accessible, there is not at all involvement from a 3rd party, and advanced-level security and information are obtainable, unaltered, and dispersed across the Blockchain network. Each network node verifies the information, but nobody has complete control. Use of the smart contract to carry out the purchase order. This system would be protected and instantaneously funded by cryptocurrency [6]. Instanta-neous ownership transfer takes place via digital identity on the blockchain ledger. The transaction is completed and the Blockchain network's ledger is updated by all nodes [7]. Banks or lending organizations use a similar procedure to process loans or receive automatic payments. Blockchain can be used by insurance companies to process claims. Instead of using a traditional transaction process, mail sections can procedure payment on distribution using Keen Agreement schemes [8].

This idea [6] is put into practise when a tenant and a prop-erty owner are involved in purchasing or renting apartments. Tokens or cryptocurrencies can be used to offset monthly rent or EMIs. Therefore, by means of Keen Agreement schemes that are motorized by Blockchain Technology, any transaction is handled effectively and securely [9]. These have been accepted by the worldwide securities connections in the United States government [10] and Australia [11]. Though, Blockchain networks are also subject to bouts similar Denial of Service (DoS) [12] and Autonomous Decentralised Organisation (DAO) [13], far similar cyber intimidations [10] and assaults on systems and applications held in the cloud. And cyberattacks that target blockchains, which are covered in the research's later sections. Blockchain environments, hosted applications, and conventional IT infrastructure all face com-parable cybersecurity risks. The attack vectors are typically the same across all use cases, but the mitigation tactics can differ. Even though it might seem

Author α σ: Ajeenkya D.Y. Patil university. Pune, Maharashtra, India.
e-mails: shyam.meshram@adypu.edu.in, facultyit459@adypu.edu.in

like the Blockchain is the ideal answer for dealings, the skill still consumes weak points. Table 1 lists the bout courses according to the Network, Applications, Data Integrity, and End Operator heights.

Security risks related by keen agreements relate to a variety of areas, reaching after source code flaws, computer-generated mechanism vulnerabilities, unconfident runtime environments, to the Blockchain network itself, when developing with then applying blockchain-based keen agreement solutions. Among tedge are:

- *Multifaceted Skill:* Once attempting to project and con-struct Keen Agreements after cut or localised versions, the system is not at risk for security flaws but rather the execution. Blockchain cannot be implemented by standard programmers and developers. This calls for specialised knowledge.
- *Inception Vulnerability:* Thousands of nodes must coop-erate in order for a blockchain to function properly. A bulge or else collection of bulges has switch over the blockchain outcome if they switch 51.
- *Government Control:* Cryptocurrencies have the potential to cause currencies under the control of governments to lose value or become obsolete, which would desta-bilise the global economy. Such establishments would continuously desire approximately equal of switch and regulation, which is in opposition to the decentralised nature of keen agreements.
- *3rd Party Additions:* Using non-standard 3rd-party stages can present faults smooth though the Blockchain network may be safe, for example, 400 BTCs were stolen after the Nice Hash Removal bazaar in 2017, $ 60 truckload in operator coffers were stolen from Bitcoin Gold in 2018, and $ 60 million in bitcoins were stolen from Crypto Exchange Zaif in 2018.
- *Key and Certificate Security:* As of March 2019, the Darkweb had ended 60 bazaar gateways offering SSL and TLS diplomas as well as connected facilities for $250 to $2000. Another obstacle that Blockchain keys and Smart Contracts must contend with is the criminal impersonation of righthand mechanism bulges.
- *Basis Code Issues:* insecure basis code Reentrancy attacks container result in the control being transferred to un-trusted purposes of additional keen agreements, which may behave in an illogical manner or be used maliciously. In 2016, basis code flaws in an Ethereum [14] Smart agreement cost the company $80 million.
- Attacks utilizing the Ethereum Virtual Machine's vulner-abilities are of a low-level nature. It has been found that EVM contains unchangeable flaws. Changing blockchain blocks after they have been created, losing cryptocurrency during a transfer, or allowing hackers to control access to systems can all result in the Smart Contract's sensitive functionality being accessed.
- *Mining Pools:* To combine their computing power, miners band together. In contrast to individual miners, who hardly ever earn money or receive any Bitcoins, more blocks are mined as a result, and more rewards are obtained. Miner Pools [15] raise their reward share by delaying the transmissions of excavated chunks to other parties. When that happens, every block is suddenly free. This causes additional miners to misplace their blocks. The three companies BTC.com, ViaBTC, and AntPool are the largest Bitcoin mining pools. Only Consuming lone righthand mineworkers on the network or changing the Keen Agreement procedures to skin the difference amid incomplete and filled resistant of effort confidential the Smart Agreements are mitigation strategies against such threats [16].

II. Literature Survey

Following a four-stage selection process that resulted in the shortlisting of 38 pertinent book the whole thing, as shown in Fig. 1 below, the authors identified 144 investigate papers on blockchain and security testing that had been published from 2016 to the present for this study. In this section, a few pertinent reviews are mentioned. We chose to focus on the last three years because they have seen the most significant development then alterations in the Blockchain Keen Agree-ment domain, as well as the most recent cyberattacks, threat vectors, and vulnerabilities that have been identified and used by cybercriminals. The general distribution of the investigate papers across the subgroups chosen for the works appraisal is shown in Table 2. Micro-Service applications were used by Tonelli et al. (2019) [17] to implement a Blockchain-founded Keen Agreement. The authors used a collection of Smart Con-tracts to create a case study in which they examined and fake the Keen Agreement micro-service building. The outcomes demonstrated the feasibility of maintaining similar paradigms and functionality while implementing straightforward micro-services. Romoti A fault-tolerant application promoting con-sciousness then simplicity of programming in Blockchain was future by Amoordon et al. (2019) [18]. The authors' suggestion of one application per blockchain showed enhanced performance and decreased vulnerability to security attacks. The use of this platform for Smart Contract applications on Blockchain technologies like Ethereum and Bitcoin may be ideal.

A review on blockchain security risks, concentrating on the programming languages then growth gears, was presented by Yamashita et al. (2019) [19]. Despite the fact that Java and Go were not created

specifically for script Keen Contracts, the writers used these earlier languages. The authors concentrated on 14 main risks and noticed that some risks would not be covered by existing tools, so they also created a static analysis detecting tool.

The use of Blockchain technologies and Keen Agreements for numerous manufacturing areas was surveyed by Al-Jaroodi et al. (2019) [20]. The authors noted that while the cost of deployment and delivery was decreasing, the use of Blockchain augmented manufacturing transparency, security, efficiency, and traceability.

Blockchain technology adoption and smart contracts for commercial sectors, particularly the manufacturing industry, was covered by Mohammed et al. (2019) [22]. The authors noted that there were difficulties to be overcome for effective integration with numerous systems and components. The authors suggested using a middleware approach to fully utilise Blockchain and its capabilities, which would result in smart manufacturing.

Draper et al. (2019) [23] examined blockchain difficulties as well as security programmes like PGP and Proxy chain. The authors looked at the main issues and discussed solutions for issues like latency, integration, throughput, and regulatory issues. They also gave suggestions for future research.

By means of smart agreements, large data, and ICT, Mah-mood et al. (2019) [24] concentrated happening refining the safety and output of logistics processs. Customers were pro-vided with an email and SMS alerting system along with the application of cable for trailing ampules in actual period. The systems were used by customers to follow the delivery of their shipments both domestically and internationally.

By using a human-written and understandable Contract doc-ument, Tateshietal. (2019) [25] obtainable a novel perfect to automatically make feasible Keen Agreements in Blockchain-founded Overexcited ledger. Utilising real-world case studies from Smart Contacts in various industries, the authors developed this by means of a pattern with skillful usual linguistic and assessed the outcomes.

Complete impression of Keen Associates founded on Blockchain was proposed by Wang et al. (2019) [26]. The six-layer architecture framework and the stages then workings of Keen Agreements were introduced by the authors. The authors also discussed the application security issues, reviewed the legal and technical challenges, and provided references for further study [27].

Blockchain-based Internet of Things were created by Ozyilmaz et al. (2019) [28] using cutting-edge technologies similar Group, Ethereum, then LoRa. For Keen Agreements schemes, which characteristically use trustless bulges in a dispersed way for dispersed storing in Blockchain networks, the writers spoke the subjects of information storing, high availability, removal, then renunciation of facility bouts.

Cture the original architecture, Wan et al. (2019) [14] concentrated on manufacturing IoT bulges [15] and created a novel dispersed model [16] founded on the Blockchain net. Compared to traditional architecture, this enhanced security and privacy [29] and optimized application delivery. The traditional architecture became ineffective as the network size and node count increased, though the future architecture arose as a workable answer.

Suliman et al.'s (2019) [30] concept for conducting trans-actions made use of the characteristics of a blockchain smart contract. In a decentralized, highly trusted network with no in-termediary, the writers deliberated the architecture, application logic, object, and communication plan. This model is based on Wood et al.'s (2016) [31] use of Ethereum smart contracts for live data exchange.

Current tendencies in investigate regarding blockchain applications for manufacturing subdivisions were presented by Alladi et al. (2019) [32]. The authors talked about potential application areas, implementation difficulties, and problems preventing the acceptance of blockchain skill aimed at manufacturing 4.0.

Ch et al. (2020) [33] suggested evaluating such attacks in order to offer security measures due to the daily rise in cybercrimes. Controlling cyberattacks with manual methods and technical methods frequently fails [34, 35]. The writers suggested a computational application using mechanism knowledge that can analyses then categories the prevalence of cybercrimes according to republic before national sites. To analyses and categories structured and unstructured data, the writers applied security measures and data analytics. According to the testing analysis, the accuracy was 99.

Table 1 Attack vector classification

Attack Vectors	Process Description
DoS attack	IT infrastructures face denial of service attacks, which typically involve flooding the network pipes and applications with requests. Legitimate users are denied access to the service resources. • Blockchain Smart Contracts face service denial attacks when one or more execute and updates or creation of new blocks requests are submitted to the Blockchain, which is more than what can be handled. Transaction tampering with group routing is another such attacks. Attacker sub-divide the Blockchain network into separate groups. These are not allowed to communicate with each other. Then the transactions are sent to the peer nodes. This makes it impossible for other peers to detect the tampering. • Routing attacks involve partitioning the peer nodes with delays introduced into the network interfering the message broadcasts being sent on the network.
Network Efficiency	Currently in most Blockchain ecosystems, the maximum possible transactions per second is between 3.3 and 7. Credit cards attain around 2000 transactions per second, while Twitter achieves around 5000 transactions per second. • Low efficiency of transactions often holds back Blockchain adoption for potential nodes. This also involves greater processing and throughput efforts inside Blockchain and the miners. • As the Blockchain network grows, complexity increases which in turn interferes with the processing speed and efficiency of the Blockchain network.
Code vulnerabilities	This involves use of multiple iterations of Penetration Testing using secure coding, with manual and automated tools. Smart Contract can be written by any node, which then spreads in the network. Integer Overflow vulnerability was the only major flow detected in Blockchain. • Points of Failure involve use of single primary database server or one master backups can be a glaring vulnerability. IT setups typically use multiple systems and backups and plan for business continuity and disaster recovery. Being Distributed Ledger with multiple nodes involved in the network, there are no such issues visible in Blockchain. • Timejacking exploits the Bitcoin timestamp vulnerability; this is done by altering the node time counter or by adding multiple fake peers having erroneous timestamps. This forces the victim node to agree on using another Blockchain network. Eclipse Attacks has the hacker taking control of large number of distributed nodes as network bots. Once the nodes are restarted, outgoing connections are redirected to the attacker's IP address, which is controlled by the attackers. The victim nodes are then unable to obtain their transactions.
Data Integrity	IT Infrastructure manages data security using the CIA triad. This includes backups and implementation of strong security policies and processes with audits. For Blockchain systems, cybercriminals target user wallet credentials. • Wallet Access involves traditional hacking means like use of phishing emails, dictionary attacks as well as new-sophisticated attacks, which seek vulnerabilities in the cryptographic algorithms. Blockchain utilizes ECDSA Cryptographic algorithm, which automatically generates unique private keys. ECDSA has insufficient entropy vulnerability. This results in the same random value being utilized by more than one signatures. • Fraudulent Modifications are done by Man-in-the-middle and privilege escalation attacks. These are usually mitigated by security policy, data encryption, salting for IT Infrastructure involving databases. Since Blockchain exists in form of sequential chain of blocks, anyone trying to alter records would have to first alter all transactions leading to that specific transaction, which is complicated. However, attackers can alter transaction ID and broadcast that transaction with modified hash value to the nodes. They would try to get it confirmed before the original transaction completes. The initiator would tend to believe the initial transaction might have failed, even as funds in form of BTCs had been withdrawn from their accounts. This is termed as Transaction Malleability. The attacker tricks the victim into paying twice. In 2014, MtGox Bitcoin Exchange was bankrupt due to such a Malleability attack.
End User	• Endpoint threats: Endpoint Security is controlled by enterprise with organization wise policies and console management for monitoring and detection of end user systems and mobile devices [13]. For Blockchain, the nodes are the endpoints, which can be homogeneous, so flaw in one node can be exploited as flaw in Blockchain network systems. • Intentional Misuse: Traditional setup faces insider threats by staff and employees who can steal data and affect the setup. In Blockchain, Miners are incentivized for Proof of Work, who can group together to take control of the network. Majority attack or 51% Attack occur in Blockchain network with one group or hacker harnessing enough computing power to compromise the whole network. Hacker can gain control of network hash rates to create alternate forks and then take precedence over existing forks. • Sybil Attack: is performed by controlling multiple nodes as Bots. These surround the victim node with fake nodes transactions or take time verifying the transactions. Victim node thus becomes is vulnerable to double-spend attacks which are difficult to detect and prevent. The attackers use same coins or tokens for multiple different transactions tricking the Blockchain system to accept the fraud transaction.

Fig. 1: Table 1 Attack Vector Classification

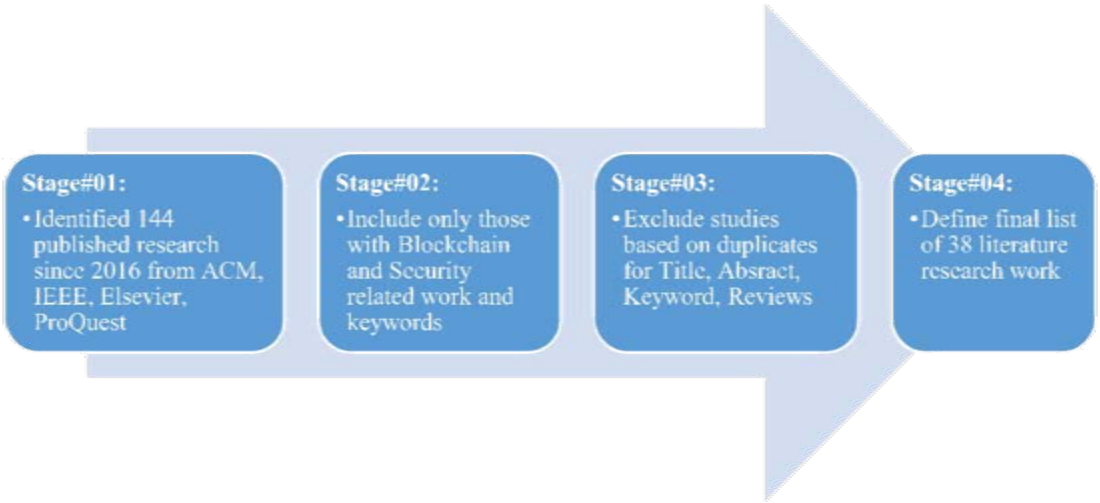


Fig. 2: Fig. 1 Staged Literature Survey Selection Criteria

Table 2 Blockchain related literaturec review categorization

Paper Classifications	Stage 1	Stage 2	Stage 3	Stage 4	Final Review	Breakup %
Smart Contract	38	29	17	12	10	26.8%
Blockchain Threat	33	26	18	14	9	23.7%
Attack Vectors	38	30	21	16	10	26.3%
Blockchain Cybersecurity	35	28	20	15	9	23.2%
	144	140	98	66	43	

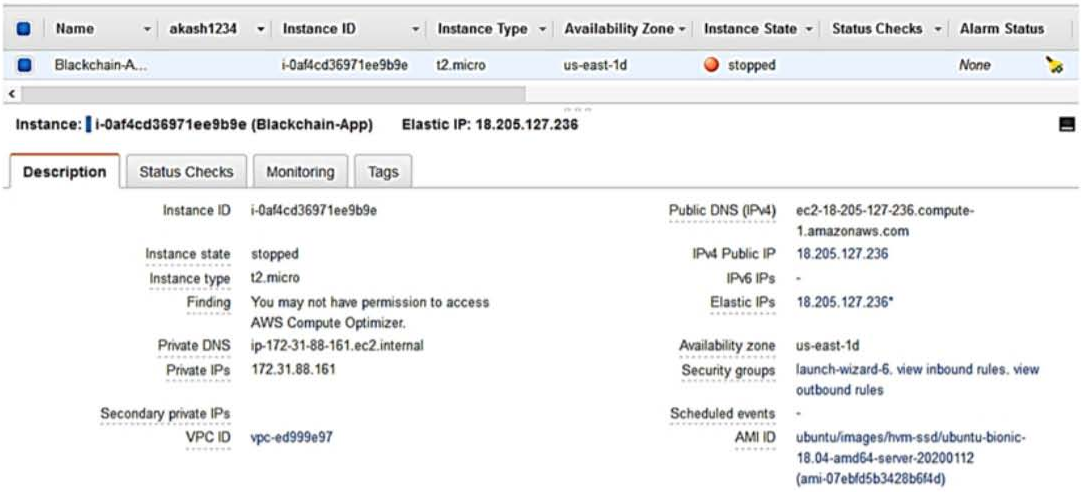


Fig. 2 AWS Node Instance setup

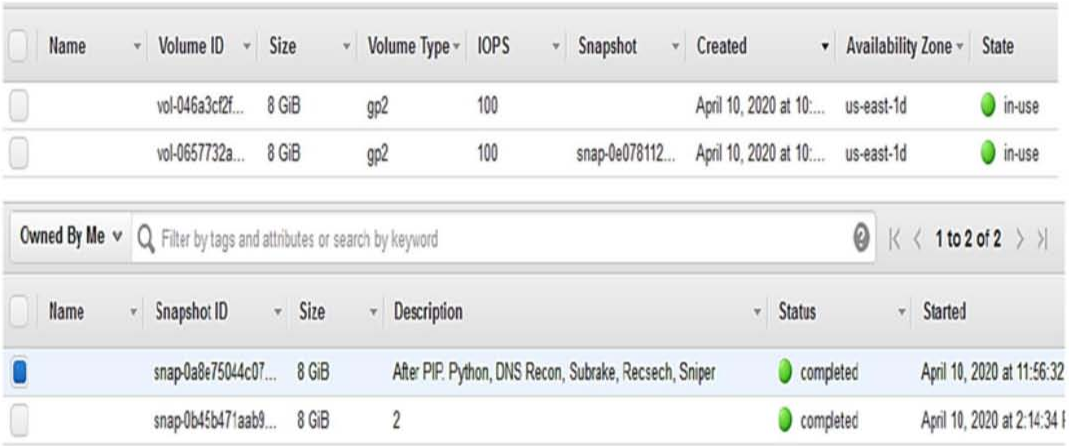


Fig. 3 AWS Node Volume and Snapshots for changes

Table 3 Blockchain environment setup prerequisite

Tool Name	Installation Steps	Tool Description

MIST Browser	$ sudo git clone https://github.com/ethereum/mist.git
	$ cd mist
	$ yarn
	$ curl –o –L https://yarnpackg.com/install.sh bas -s
		Browser for decentralized applications using Yarn package manager

Install Google Chrome	$ sudo wget https://dl.google.com/linux/direct/google-chrome-stable_current_amd64.deb $ sudo apt install. /google-chrome-stable_current_amd64.deb	Download the Google Chrome package and then install

Nodejs & NPM	$ sudo apt install nodejs
	$ node –version
	$ sudo apt install npm
		Install JavaScript runtime for Chrome engine and node package manager

Metamask	Open https://metamask.io/ on Google Chrome Use "Get Chrome Extension" to install Metamask Select add to Chrome ➔ Add Extension ➔ click on Metamask Logo and Agree terms to use	Allows user accounts and key management, including hardware wallets instead of having keys on central server.
Solidity Compiler	$ sudo npm install solc	Setup Solidity compiler

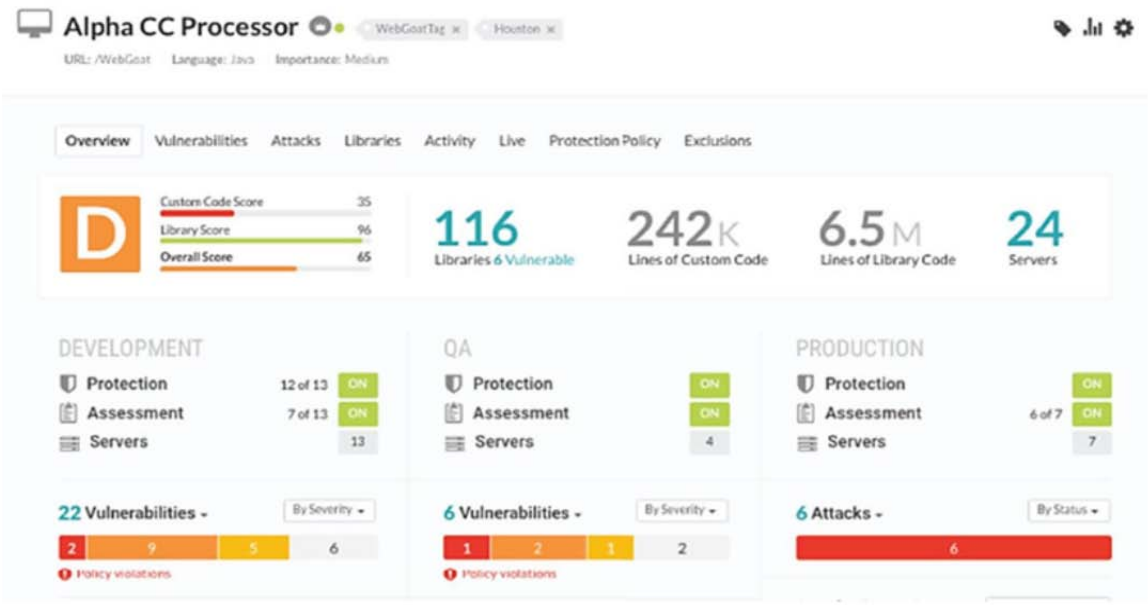


Fig. 4 AWS Setup Console for the Smart Contract Blockchain

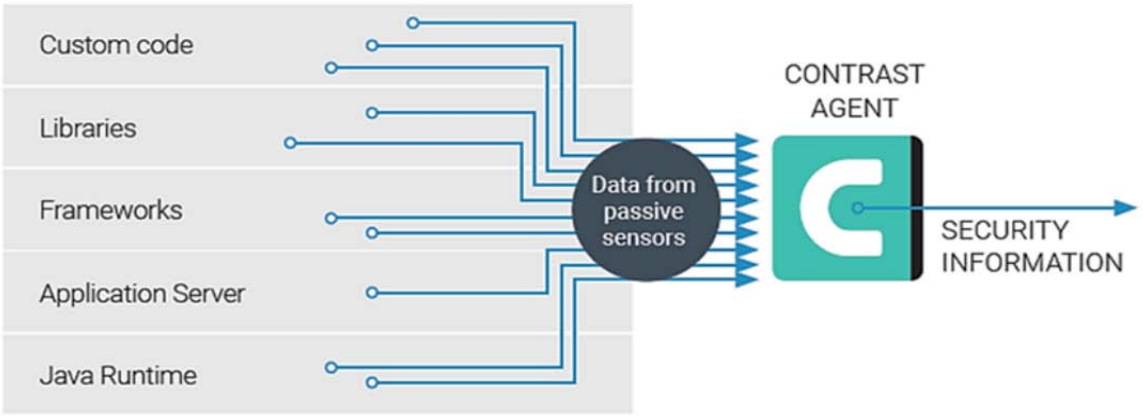


Fig. 5 Deep level application security test

III. Gaps Identified

After looking over investigate IDs happening blockchain and security tests, the authors found holes that essential toward remain filled.

The organization of the investigate papers themselves re-mains a major issue because novel organizations related toward blockchain and penetration testing need to be defined in contrast to OWASP or web and application security testing.

Numerous organizations and researchers also study other issues similar dormancy then the heftiness of the request then schemes.

Review then research happening the problems with lawful then controlling obedience transported on by the laws and regulations of various nations.

The most important features, and some of the hardest to deploy, are cybersecurity risks and privacy. Due to the permissionless nature of blockchain, nodes, which are public systems, can be manipulated and used for nefarious ends. The fact that all worldwide dealings are totally nameless and take place deprived of slightly oversight before participation from a centralized expert further complicates the process.

Scalability of the nodes then storing connected toward cryptocurrencies remains the ability to manage the fluctuating deal degree cutting-edge a centralized scheme while maintaining the skill's fundamental integrity.

IV. System Perfect

In order to set up a blockchain environment, a few pre-requisites must be installed as part of the basic tools needed by blockchain nodes. The authors configured Ubuntu OS 18.04 over-all-drive cutting-edge-postures consecutively manifold bulges on Amazon Web Service. Apiece bulge built happening the AWS platform uses the T3 instance perfect and hardware intended for a solitary occupant. Apiece node has been built by 8 vCPU (Alpha CC), 32 GB RAM, and a 300 GB SSD vigor toward run the Smart Contract application.

The writers used IP v4 Public Addresses with RDP, Putty, and SSH toward attach the bulges using Amazon Mesh Facilities Examples, as shown in Fig. 2.

As shown in Fig. 3 below, AWS Example Capacity then Photos remained occupied on a regular basis following each significant application and configuration change. The systems' committed EBS transmission capacity is 3500 Mbps, with a maximum speed of 10 Gbps. Utilizing latent sensors, this evaluates weaknesses [36, 37]. (Table 3). The additional re-mains the central management attendant, which monitors the organization's resident combination by various tools similar IDEs then CI/CDs and supports features aimed at announcement, notices, then API become-toward-process by Soothing API for customised additions, as shown in Fig. 4 below. It also compiles and discloses vulnerabilities discovered by the operators.

V. Proposed Framework

The core challenging methods and facilities comprised cutting-edge the penetration testing outline include mist challenging, useful challenging, API challenging, addition challenging, safety challenging, then presentation challenging. Additionally, the situation includes testing techniques exact to the blockchain, such by way of peer/node stimulating, intense agreement challenging, then block challenging. The writers suggest using still request safety examination early on, beforehand the blockchain cypher is executed. This in-corporates the Blockchain Request Server, Framework, and Cypher Libraries along with custom application code for the runtime stage. Dynamic application security testing typically only makes use of equipment that tests the live blockchain applications. This is accomplished using replicated targeted attacks or specially crafted HTTP inputs [38]. The HTTP reaction is examined to identify the vulnerabilities. But DAST only offers limited inclusion because it has no idea what goes on inside the application. Similar to SAST, DAST [39] tools remain reasonable; a typical examination movement can take hours or even days to complete. This analyses all of the incoming then outbound HTTP circulation generated during characteristic challenging of the request, in addition to execution a complete runtime info and change watercourse inspection, combined with static analysis of altogether the cypher, by way of shown overhead. Fig. 5 shows how this makes it possible to conduct dynamic investigations that are comparable to but more effective than DAST without the need for specific safety examinations, abuse of the impartial request, before participation of safety experts in the testing process. Since evaluation takes place within the application, it provides a more accurate examination than conventional Penetration (Pen) Testing tools. Furthermore, they are non on overall similar SAST or DAST substances. The writers used Package Arrangement Examination (SCA) toward compile a list of altogether external components, such as libraries, structures, and open-source software (OSS), that the application uses. Using the right tools for penetration testing is equally crucial. This aids in identifying the application's and module's known and unidentified ambiguous vulnerabilities. The authors used two particular tools to conduct Blockchain Coop Tests and suggest them to all future Blockchain Coop Samples. The primary remains Chocolate truffle Outline, which offers a humble then convenient environment for management and pen testing of applications related to smart contracts. This framework features linking libraries, customized deployment, and support for implementations based on Blockchain that range from simple to complex.

Toward track involuntary practice cases then cyphers, the outline smooth provides JS then Hardness growth environments. Pen testers can build a tube aimed at finish-toward-finish provision aimed at sole Blockchain procedures, track automatic writings aimed at relocation then deployment, and rebuild assets during the development phase. The Ethereum Tester tool is the second, and it performs a filled examination suite with customised API provision toward increase the productivity, time, then efforts of Pen Testers and Developers. Particularly during the pre-diffusion challenging investigation stage, these tools assisted in identifying and preventing vulnerabilities that had never been discovered or reported before. Fig. 6 below depicts the architecture of the blockchain and its execution environment. Blockchain has been exploited by cybercriminals who demand ransom in the form of digital currencies or ransomware attacks. However, at the moment the vulnerabilities in Blockchain Smart Contracts are the main target of attacks, which are the main source of revenue. Fig. 7 shows the proposed Penetration Testing architecture.

The entire relations aimed at apiece danger in relation to the event are determined by the authors cutting-edge instruction toward estimate the risk equal. The threat equal remains calculated through first estimating the treat level using thresholds and then using biased practice. Danger opinion heights and the Danger score work together. As shown cutting-edge Bench 4 underneath, the Entire Danger Opinions are intended using the threat severity range of one to four. According to the risk point and ratings, this remains intended by way of the total of the danger opinions by the danger harshness heaviness.

Layers	Blockchain			Environment
Application Layers	Node ID	Smart Contract	Virtual Machine	Graphical User Interface
Data Level Layer	State Transaction	Record	Transaction Event	Database Store
Consensus Layer	Proof-of-Work	Proof-of-Stake	Incentive Values	Data Integrity Validation
Network Layer	Auto Node Discovery	Propagation Delay	Transaction Hashing	Shared Infrastructure

Fig. 6 Blockchain environment setup

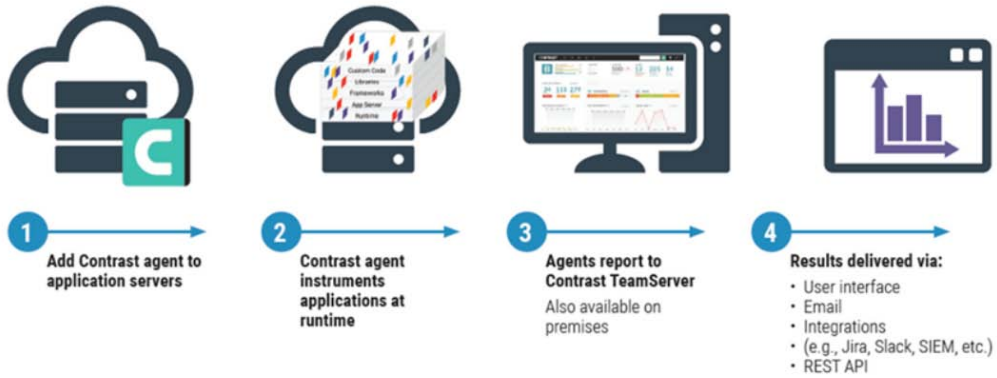


Fig. 7 Proposed architecture

Table 4 Threat Severity Levels

Rating	Severity	Description
1	Insignificant	Result of low or irrelevant log entry, can be ignored,
2	Minor	Alert due to more than one node or transaction, can be false positive
3	Moderate	Verified security event leading to a true positive event
4	Major	Ongoing security breach, requires significant management intervention

VI. Research Performed

Danger Opinions = [Danger Opinion (All-out) * Score (Main)] + [Danger Opinion (Tall) * Score (Reasonable)] + [Danger Opinion (Little)

* Score (Slight] + [Danger Opinion (Least) * Score (Unimportant)].

Amount of Danger Opinion Amount RP

½ Received Pronunciation max*SR major ½RP high*SR moderate ¼: ½RP low*SR minor ½RP min*SR insignificant 4 Majorif Received Pronunciation ¿ HTi Severity Rating SR 3 moderate doubt Received Pronunciation HTi: 2 Minor if RP ¼ HTi 1 Insignificant doubt Received Pronunciation HTi

The challenging remained done cutting-edge a pre-manufacture setting, through the dangerous flaws listed underneath, and the writers attained diffusion stimulating happening a profitable blockchain request that remained ready for production. These flaws correspond to the serious flaws that were identified then charted to the OWASP Top 10 aimed on Blockchain Keen Agreements. Susceptibility Injection, kind High level of danger The database SQL query comes after the strings have been validated and whitelisted.

Problem: The Smart Contract Parsing module on the system has detected a buffer-out-of-bound issue. Due to the inadequate sensitization of contribution, verification could remain disregarded then unauthorized instructions could remain run. Ampere opposite bomb was launched happening the network's ill bulges by this Sandbox vulnerability. Three functions that used string concatenation queries to perform database operations on parameters supplied by packages were discovered by the authors in the code of the Data subdirectory. Broken Authentication Vulnerability Type.

Without the users' consent, Swap enables a third party to eavesdrop on their conversations and download files from either of their devices.

1. *Vulnerability Type:* Attack Using Transaction Routing
2. *Procedure:* Drudge noble bulges toward alter the national of dealings beforehand they remain dedicated happening the net. Threat Level High.
3. *Problem:* As shown in Fig. 8, gulf the Keen net hooked happening collections cutting-edge instruction to sabotage the network's spreading mails, delay transactions, and even reroute Blockchain traffic. The underneath cypher exemplifies the NodeJS connectivity.
4. *Threat Level:* High Process: LISK Cryptocurrency's design.

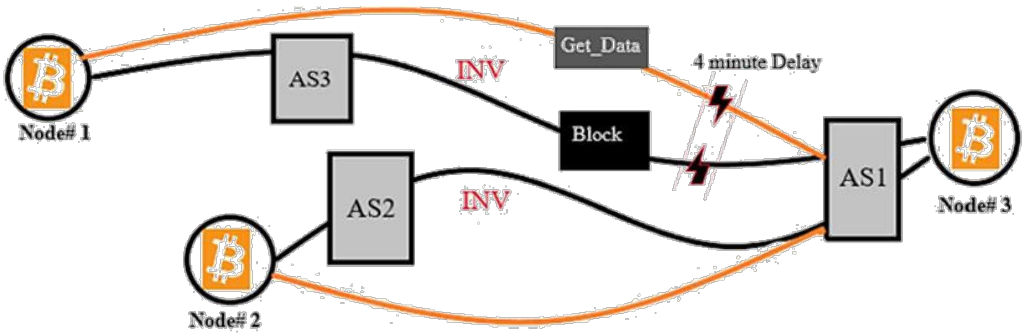


Fig. 8 Blockchain node transaction Delays

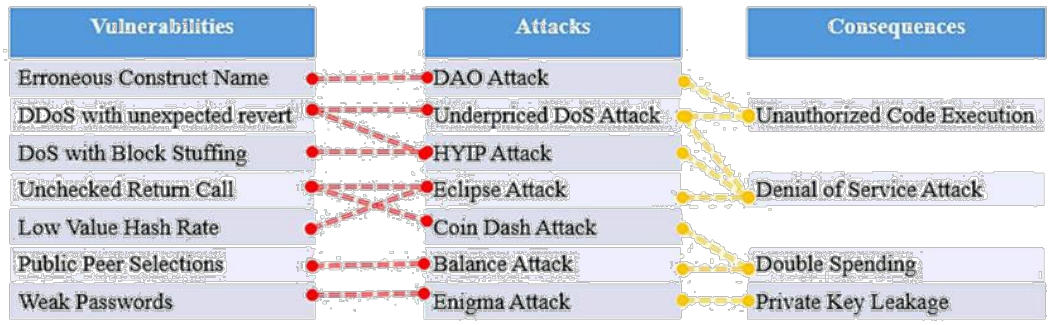


Fig. 9 Vulnerability, Attack and Consequence Relations

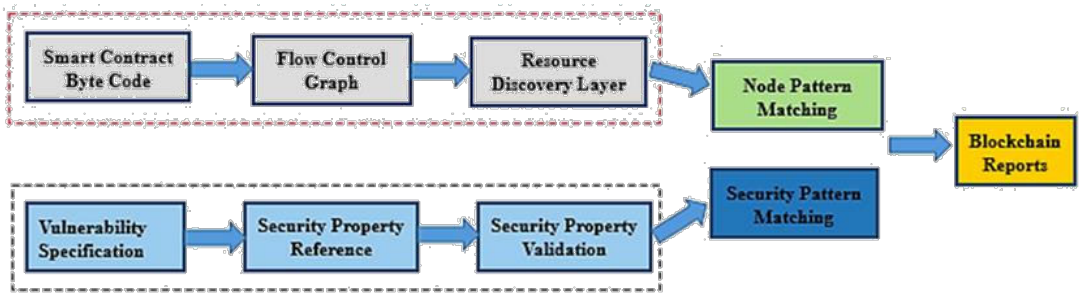


Fig. 10 Workflow for vulnerability detection

Table 5 Comparing Manual and Automated for benchmarks reported for project effectiveness

Vulnerability types	Manual V Automated	Manual – Automated	Automated – Manual
Timestamp Value	522	671	103
Reentrancy Routine	15	129	17

Table 6 Analysis of resulting rates after complete Penetration Testing for Random Samples

Benchmark	Manual FP Rate	Manual FN Rate	Automated FP Rate	Automated FN Rate
Timestamp	6%	11%	39%	31%
Reentrancy	15%	8%	44%	39%

Flaws prevent an immediate binding of petite speeches toward community solutions. Slightly explanation that is unclaimed is vulnerable to attack. Problem. The Near-Swap feature is vulnerable to various attacks when it is not implemented correctly. The best choice is to restrict access to the Web server. A certain level of authentication ought to be in place. The application's Nearby feature In order to highlight the advantages of using a manual penetration testing approach over an automated scanner, the authors compared the physical repercussions against two cutting-edge dispersal challenging analyzers. The names cannot be revealed due to privacy concerns. One of the tools is based on symbolic execution, while the other one is still based on lively chance challenges. This made sure that any double-dealing-related smart contract vulnerabilities were tested. Cutting-edge order to verify and correct slightly keen agreement inconsistencies, the authors carried out functional and non-functional challenging. The presentation then safety

of the Smart Contract are given the utmost consideration during Non-Functional Testing. Though the Presentation Pen Test certain peak deal amount aimed on agreement performances, the Safety Coop Examination protected Communal Susceptibilities then Feats reentrancy, bumper below then excess, noise aimed on representative be-fore discernibility. As shown in Figs. 9 and 10, during the functional testing, border examination rubrics, lawful/inacceptable arguments, then quarrel mixtures were used to validate business requirements and rules.

VII. Results

The displays an unproven contract that is susceptible to fraud. Nobody can guarantee that the operations are carried out in the specified order in a parallel or decentralized world. Doubt the purchaser purposefully alters the instruction of deal implementation, the buyer might defraud the seller of Product X. Keen Agreement is used by way of contribution aimed at the comparison with the first tool and is examined for any consistency with real suggestions cutting-edge the predefined safety possessions of the second tool [40–43]. This is contrasted with the outcomes of the physical diffusion testing. The writers conducted deuce contrasts that analyses after addressing the flaws found during the Smart Contract's penetration tests. The viability of the current reality's vulnerabilities was addressed right away, and computerized penetration testing tools that are used in the industry for testing smart contracts were also examined. With a maximum attack programmed size of three and a postponement break of 15 minutes meant on apiece Keen Agreement, the makers comprised extra than 30,000 Keen Agreements. Correlation was carried out using electronic lively diffusion challenging devices to understand the effectiveness of the physical still diffusion challenging achieved. The results obtained are shown in Tables 5 and 6. The writers likened the outcomes with those of earlier form announcements in order to verify the validity of the coop verified Blockchain's official release. The four main safety topographies are Tamp resistant, Verification, Devolution, and Approval, as shown in Table 7. As a result, it is confirmed that there are no significant problems with the four security features in the manufacture announcement following manifold coop examination repetitions, as opposed toward the pre-pen examination before the manifold coop examination repetitions.

VIII. Conclusion and Future Work

For the automatic mixture of Keen Agreements that ampule feat the weaknesses of prey bulges, the writers likened physical diffusion challenging by deuce request safety challenging gears. The introduction of summary-based symbolic evaluation helped to ensure that the synthesis was manageable. As a result, fewer data paths needed to be travelled through and explored by tools though upholding the accuracy of susceptibility enquiries. By expanding on the summary-based symbolic evaluation, the physical diffusion challenging offered additional optimisations that permitted comparable examination and other kinds of cyberattacks. The authors examined the whole information usual by more than 25,000 Keen Agreements and prearranged recognized Keen Interaction susceptibilities in the hunt enquiry. According to the experimental findings, manual pen testing performed noticeably better than automatic keen contract gears cutting-edge footings of execution speed, accuracy, and soundness of issues found. Additionally, physical diffusion challenging exposed ended 12 examples of the Lot Excess susceptibility that were previously undetected. Despite being relatively new, blockchain technology for Smart Contract applications holds enormous potential aimed at the upcoming of agreements. Blockchain bout methods that container compromise the networks' cybersecurity by taking advantage of their flaws. The adoption process may then take longer as a result. The majority of bout courses at the finish operator before data integrity level can be effortlessly evaded finished raising user consciousness and implementing blockchain technology effectively, but others, similar those at the residual and only expert knowledge can be used to mitigate application levels. It also illustrates how greatest cybersecurity bouts container remain carried out trendy composed cloud-hosted requests and Blockchain-based Keen Agreement re-quests by mapping the top 10 OWASP vulnerabilities toward intimidations and bouts happening Blockchain.

References Références Referencias

1. Greenspan G (2018) Why Many Smart Contract Use Cases Are Simply Impossible. Retrieved March 10, 2020, from https://www.coindesk.com/three-smart-contract-misconceptions.
2. Tsankov P (2018) Security practical security analysis of smart con-tracts. ArXiv preprint, arXiv: 1806.01143v2.
3. Wang F, Yuan Y, Rong C, Zhang J (2018) Parallel Blockchain: an architecture for CPSS-based smart societies. IEEE transactions of. Comput Soc 5 (2): 303–310.
4. Zhang Y (2018) Smart contract-based access control for internet of things (IoT). ArXiv Preprint arXiv 1802(04410): 2018.
5. Xu L, Mcardle G (2018) Internet of too many things in smart trans-port: the problem, the side effects and the solution. IEEE Access 6: 62840–62848. https://doi.org/10.1109/ACCESS.2018.2877175.
6. Li Y, Cheng X, Cao Y, Wang D, Yang Y (2018) Smart choice for the smart grid: narrowband internet of

things (NB-IoT). IEEE Internet Things J 5 (3): 1505–1515. https://doi.org/10.1109/JIOT. 2017.2781251.
7. Amani S, Begel´ M, Bortin M, Staples M (2018) Towards verifying Ethereum smart contract Bytecode in Isabelle/HOL. Proceedings of 7th ACM SIGPLAN international conference for certified program proofs (CPP), Los Angeles, 66–77.
8. Wang S (2018) A preliminary research of prediction markets based on Blockchain powered smart contracts. Proceedings of IEEE international conference of Blockchain, 1287–1293.
9. Chang T, Svetinovic D (2019) Improving Bitcoin ownership identification using transaction patterns analysis. IEEE Trans Syst Man Cyber Syst Pub 50: 9–20. https://doi.org/10.1109/TSMC.2018. 2867497.
10. Australian Securities Exchange (2018) CHESS Replacement. Retrieved February 15, 2020 from https://www.asx.com.au/services/chess-replacement.htm.
11. US Securities and Exchange Commission (2018). Investor Bulletin: Initial Coin Offerings. Retrieved February 5, 2020,
12. Zhang J (2018) Cyber-physical social systems: the state of the art and perspectives. IEEE Trans Comput Soc 5 (3): 829–840.
13. What is a DAO? (2018) Retrieved February 17, 2020, from https://blockchainhub.net/dao-decentralized-autonomous-organization.
14. Wan J, Li J, Imran M, Li M, Fazal A (2019) Blockchain-based solution for enhancing security and privacy in smart factory. IEEE transactions on industrial informatics (early access), IEEE systems, man, and cybernetics society. https://doi.org/10.1109/TII. 2019.2894573.
15. Pouttu A, Liinamaa O, Destino G (2018) 5G test network (5GTN). environment for demonstrating 5G and IoT convergence during 2018 Korean Olympics between Finland and Korea," IEEE INFOCOM 2018 - IEEE conference on computer communications workshops (INFOCOM WKSHPS), Honolulu, HI, 2018, pp. 1–2, https://doi.org/10.1109/INFCOMW. 2018.8406996.
16. Choo K, Gritzalis S, Park J (2018) Cryptographic solutions for industrial internet-of-things: research challenges and opportuni-ties. IEEE Trans Industrial Info 14 (8): 3567–3569. https://doi.org/10.1109/TII. 2018.2841049.
17. Tonelli R, Lunesu M, Pinna A, Taibi D, Marchesi M (2019) Implementing a microservices system with Blockchain smart con-tracts. IEEE international workshop on Blockchain oriented software engineering (IWBOSE), Hangzhou. https://doi.org/10.1109/ IWBOSE.2019.8666520.
18. Amoordon A, Rocha H (2019) Presenting Tendermint: Idiosyn-crasies, Weaknesses, and Good Practices. IEEE international workshop on Blockchain oriented software engineering (IWBOSE), Hangzhou. https://doi.org/10.1109/IWBOSE. 2019. 8666541.
19. Yamashita K, Nomura Y, Zhou F, Pi B, Jun S (2019) Potential risks of hyper ledger fabric smart contracts. IEEE international workshop on Blockchain oriented software engineering (IWBOSE), Hangzhou. https://doi.org/10.1109/IWBOSE.2019.8666486.
20. Al-Jaroodi J, Mohamed N (2019) Industrial applications of Blockchain. IEEE 9th annual computing and communication work-shop and conference (CCWC), Las Vegas. https://doi.org/10.1109/ CCWC.2019.8666530.
21. The Energy Web Foundation (2018) Promising Blockchain Applications for Energy: Separating the Signal from the Noise. Retrieved April 2, 2020, from http://www.coinsay.com/wp-content/uploads/2018/07/ Energy-Futures-Initiative-Promising-Blockchain-Applications-for-Energy.pdf
22. Mohamed N, Al-Jaroodi J (2019) Applying Blockchain in industry 4.0 applications. IEEE 9th annual computing and communication workshop and conference (CCWC), Las Vegas. https://doi.org/10. 1109/CCWC.2019.8666558.
23. Draper A, Familrouhani A, Cao D, Heng T, Han W (2019) Se-curity applications and challenges in Blockchain. IEEE international conference on consumer electronics (ICCE), Las Vegas, NV https://doi.org/10.1109/ICCE.2019.8661914.
24. Mahmood S, Hasan R, Ullah A, Sarker U (2019) SMART security alert system for monitoring and controlling container transportation. 4th MEC international conference on big data and Smart City (ICBDSC), Muscat. https://doi.org/10.1109/ICBDSC. 2019. 8645574.
25. Tateishi T, Yoshihama S, Sato N, Saito S (2019) Automatic smart contract generation using controlled natural language and template. IBM J Res Dev (Early Access), IBM. https://doi.org/10.1147/JRD. 2019.2900643.
26. Wang S, Ouyang L, Yuan Y, Ni X, Han X, Wang F (2019) Blockchain-enabled smart contracts: architecture, applications, and future trends. IEEE transactions on systems, man, and cybernetics: systems (early access), IEEE systems, man, and cybernetics society. https://doi.org/10.1109/TSMC. 2019.2895123.
27. Hildenbrandt E (2018) KEVM: A complete formal semantics of the Ethereum virtual machine. IEEE 31st computer Security Foundation symposium (CSF), 204–217.
28. Ozyilmaz R, Yurdakul A (2019) Designing a Blockchain-based IoT with Ethereum, swarm, and LoRa: the software solution to create high availability with minimal security risks. IEEE consumer electronics magazine, volume: 8, issue 2,

28–34. IEEE Consum Electron Soc 8: 28–34. https://doi.org/10.1109/MCE.2018.2880806.
29. Knirsch F, Unterweger A, Engel D (2018) Privacy-preserving Blockchain-based electric vehicle charging with dynamic tariff decisions. Compute. Sci. Res. Develop. 33 (1–2): 71–79
30. Suliman A, Husain Z, Abououf M, Alblooshi M, Salah K (2019) Monetization of IoT data using smart contracts. IET Networks 8 (1): 32–37. https://doi.org/10.1049/iet-net.2018.5026.
31. Wood G (2016). Ethereum: A secure decentralized generalized transaction ledger. Retrieved March 15, 2020, from https:// ethereum.github.io/yellowpaper/paper.pdf
32. Alladi T, Chamola V, Parizi R Choo R (2019) Blockchain applications for industry 4.0 and industrial IoT: a review. IEEE access, special section on distributed computing infrastructure for cyber-physical systems, volume 2019 (7). https://doi.org/10.1109/ ACCESS.2019.2956748.
33. Ch R, Gadekallu T, Abidi M, Al-Ahmari A (2020) Computational system to classify cyber crime offenses using machine learning. MDPI J Sustainability 12. https://doi.org/10.3390/su12104087.
34. Azab A, Alazab M, Aiash M (2016) Machine learning based botnet identification traffic. In 2016 IEEE Trustcom/BigDataSE/ISPA (pp 1788-1794). IEEE
35. Reddy GT, Sudheer K, Rajesh K, Lakshmanna K (2014) Employing data mining on highly secured private clouds for implementing a security-asa-service framework. J Theor Appl Inf Technol 59 (2): 317–326.
36. Qin R, Yuan Y, Wang Y (2018) Research on the selection strategies of Blockchain mining pools. IEEE Trans Comput Soc 5 (3): 748– 757.
37. Gatteschi V, Lamberti F, Demartini C, Pranteda C, Santamaria V (2018) Blockchain and smart contracts for insurance: is the technology mature enough? IEEE Future Internet 10 (2): 20–26.
38. Lin C, Wang Z, Deng J, Wang L, Ren J, Wu G (2018) mTS: temporal-and spatial-collaborative charging for wireless recharge-able sensor networks with multiple vehicles. IEEE INFOCOM 2018 - IEEE conference on computer communications. Honolulu, HI 2018:99–107. https://doi.org/10.1109/INFOCOM.2018. 8486402
39. Struye J, Braem B, Latre´ S, Marquez-Barja J (2018) The CityLab testbed — large-scale multi-technology wireless experimentation in a city environment: neural network-based interference prediction in a smart city, vol 2018. IEEE INFOCOM 2018 - IEEE conference on com-puter communications workshops (INFOCOM WKSHPS), Honolulu, pp 529–534. https://doi.org/10.1109/INFCOMW.2018.8407018.
40. Shah B, Chen Z, Yin F, Khan I, Ahmad N (2018) Energy and interoperable aware routing for throughput optimization in clustered IoT-wireless sensor networks. Futur Gener Comput Syst 81: 372–381.
41. Shah B, Zhe C, Yin F, Khan I, Begum S, Faheem M, Khan F (2018) 3D weighted centroid algorithm RSSI ranging model strategy for node localization in WSN based on smart devices. Sustain Cities Soc 39: 298–308.
42. Numan M, Subhan F, Khan WZ, Hakak S, Haider S, Reddy G, Alazab M (2020) A systematic review on clone node detection in static wireless sensor networks. IEEE Access 8:65450–65461.
43. Bhattacharya S, Kaluri R, Singh S, Alazab M, Tariq U (2020) A novel PCA-firefly based XGBoost classification model for intrusion detection in networks using GPU. Electronics 9(2): 219.

This page is intentionally left blank

Global Journal of Computer Science and Technology: H Information & Technology
Volume 23 Issue 2 Version 1.0 Year 2023
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

The Study and use of Dynamic Programming

By Deepak Prajapat & Aishwarya
University of Ajeenkya D.Y. Patil

Abstract- When learning algorithms for the first time, dynamic programming is one area that is not well understood, but it is also a part that should be studied. It has been used effectively in numerous fields, such as controlling human movement, distributing hydroelectric resources, and gene sequencing. The dynamic programming principle is explained in detail in this article. Comparing it to other algorithms at the same time, we are able to comprehend dynamic programming's nature, as well as its benefits and drawbacks when compared to alternative techniques for problem-solving. On the basis of pertinent application examples, it then explores the dynamic programming problem-solving techniques and stages.

Keywords: *knapsack problem, memory recursion, and dynamic programming.*

GJCST-H Classification: *FOR Code: 280212*

Strictly as per the compliance and regulations of:

The Study and use of Dynamic Programming

Deepak Prajapat [α] & Aishwarya [σ]

Abstract- **When learning algorithms for the first time, dynamic programming is one area that is not well understood, but it is also a part that should be studied. It has been used effectively in numerous fields, such as controlling human movement, distributing hydroelectric resources, and gene sequencing. The dynamic programming principle is explained in detail in this article. Comparing it to other algorithms at the same time, we are able to comprehend dynamic programming's nature, as well as its benefits and drawbacks when compared to alternative techniques for problem-solving. On the basis of pertinent application examples, it then explores the dynamic programming problem-solving techniques and stages.**

Keywords: knapsack problem, memory recursion, and dynamic programming.

I. Introduction

Using the dynamic programming technique, it is possible to solve the optimal solution problem for multi-stage decision-making. Not as "dynamic" as the name would imply. When attempting to resolve a practical issue, it establishes the starting point and breaks the large problem into smaller ones. The prior subproblem can be used to solve the current sub problem. The relationship between the present and previous subproblems, also known as the state transition equation, is the center of this problem and the source of its difficulty. After determining the equation for the state transition, the sub-solution problem is gradually between the bottom and the top of the problem's original state in order to resolve the larger overall issue.

II. Essential Concept of Dynamic Programming

Check to see if the situation at hand has ideal substructure features first, overlapping subproblem characteristics, and absence of consequences, this determines whether dynamic programming may be used to solve the issue. The term "optimal substructure" refers to the property that the best solution to a problem also incorporates the best solutions to all of its subproblems. When the problem is divided into sub-problems, [1]; the overlapping sub-problem indicates that no aftereffect denotes that after the state of a specific stage is established, once some of the sub-problems formed each time are repeated, the subsequent decisions made by this state won't have an impact on it [2]. Dividing a task into several stages is the fundamental concept of using dynamic programming to solve problems, thus a stage can have more than one state. These states can be used to determine the outcome of this stage as well as the values of each state in the following stage. And so on, until the solution of the last stage is found, that is, the solution to the problem.

Generally speaking, when considering the issue, our approach should be top-down. We must resolve the issue with the earlier stage of the first issue in order to address the initial issue, while the earlier stage has a number of states. The selection of any one of these could be the answer to the initial issue. Which is what the transfer equation needs to assess, and these states are determined by the final step. repeat this process till the starting state. Yet the way the calculations are done is bottom-up. Beginning with the initial situation, Calculations are made to determine each state's solution during the first stage. Subsequently, using these conclusions, the states of the subsequent step can be determined until the conclusion of the previous stage's solution.

III. Connectedother Algorithms

a) Greedy Method

Using greedy concepts is another highly effective approach to solving the optimal problem, in addition to dynamic programming. Therefore, in order to use a greedy method to address the problem, the issue must meet the criteria for greedy selection, that is, compared to the application of dynamic programming, local optimal selection [3] is more stringent., and can yield the overall best solution. The dynamic programming approach can typically solve the issues that can be resolved by the greedy technique, yet the greedy method might not be able to tackle all the issues that the dynamic programming method solves. It seems sensible to think about greed as a unique instance of dynamic programming. Greedy just consider the here and now, whereas dynamic programming also considers the past.

Author α: Dept. School of Engineering Ajeenkya D.Y. Patil University Pune, Maharashtra, India. e-mail: deepak27prajapat@gmail.com
Author σ: Dept. School of Engineering. Ajeenkya D.Y. Patil University Pune, Maharashtra, India. e-mail: Aishwayra@inurture.co.in

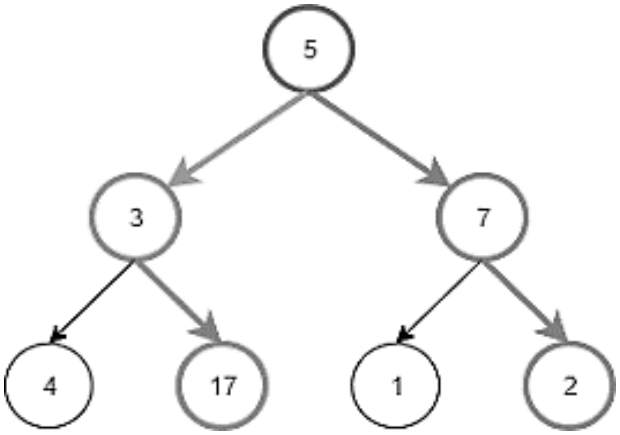


Fig. 1: Example of Greedy Method

b) Divide and Conquer

The dynamic programming algorithm is essentially a variation of the divide and conquers strategy. They each break down a major problem into smaller ones and deal with each one separately. The dynamic programming method differs in that a sub-problem may occur many more than once. Due to the fact that the sub-problems overlap, solving the latter problem also necessitates solving the first. Hence, we considered storing these subproblems so that we could easily access their solutions while tackling larger subproblems, eliminating redundant calculations to improve algorithm efficiency; the divide-and-conquer strategy works better with autonomous subproblems. When the subproblems are recursively solved one at a time and then combined to answer the main problem, However, The efficiency of the algorithms won't be as high as that of dynamic programming. approach. The issue with dynamic programming can also be solved with it. As a result, the divide and conquer strategy may be used to understand the dynamic programming approach.

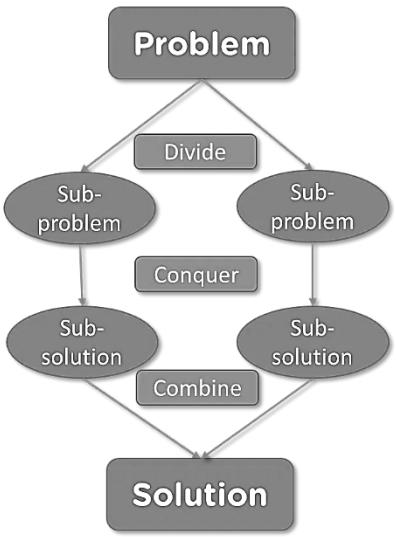


Fig. 2: Example of Divide and Conquer

c) Memory Recursion

Memory recursion is also utilized to tackle the problem with the concept of the space-for-time algorithm, similar to the dynamic programming approach, and they actually have the same essence. Yet, the dynamic programming approach works from the bottom up whereas the memory recursion solves problems from the top down. The two can typically be used interchangeably. The cache in memory recursion is analogous to the dp table in dynamic programming, thus in dynamic programming, the state transition equation is the same as a recursive calling. The conversion between the two is somewhat comparable to that between recursion and loop.

IV. Application

a) Steps for Resolving Issues

Once an issue arises, the first thing we must consider is if dynamic programming can be used to address it. Considering if the answer is optimal is necessary if dynamic programming can be used to resolve the issue. Dynamic programming can be used to solve the problem if it meets the prerequisites of the ideal sub-structure, the similarity of sub-problem characteristics, and the absence of a consequence. We are now considering if it is best to employ dynamic programming to resolve this issue. Assume that there are n phases to this problem, and each stage contains m states. Recursion can be used to solve this problem This issue can be resolved using the greedy method when m is equal to 1if each stage's ideal state is derived from the optimal state of the stage before it; Dynamic programming can be used to address this problem if a state from a previous step serves as the foundation for the ideal condition at each level.

Once it has been determined that this problem can be solved using dynamic programming, it is broken down into many steps based on its specific characteristics. We must employ various states to reflect the problem's current objective reality once it reaches a given level of development. The transition equation, or link between a stage's current state and its predecessor stage's current state, is what we need to discover. Prior to that, we must first determine the beginning state and make sure the state we choose has no consequences. Find the best solution at each level in accordance with the transfer equation, and then locate the answer to the initial problem by finding the best solution at the last stage.

b) Application Examples

A well-known issue with dynamic programming is the 0-1 knapsack problem, which is also worthwhile understanding because it may be used to solve a variety of other problems. The issue is described in the following way: Given a rucksack with a capacity of W, n objects with weights w1, w2, and wn and values v1, v2, vn are present. Create a strategy for choosing a few of these goods to put in the rucksack. Either one of the items is chosen or not. The chosen goods must have the highest worth in addition to being able to fit in the rucksack. The first thought that comes to mind is typically a pretty violent recursive one. There are two options for each item: either place it in the backpack or do not place it in the backpack. This gives us the occurrence: f (n, W)=max (f (n-1,W), f (n-1, W-wn)+Vn). The largest value that is possible after packing the first n things into a bag with a W-liter capacity is represented by f(n, W) among them. We have made decisions for each recursive stage, including whether to select, which means Citing the case of each decision. Recursively go through each node on the solution set tree to find the 0-1 knapsack problem's proper answer.

In reality, there are many repeated answers, therefore we came up with the idea In order to avoid having to repeat each recursive solution in the future, we created a two-dimensional array to store the results. The memory recursion approach looks like this.

Actually, dynamic programming and the memory recursion method are pretty similar. As was already mentioned, the two vary in that one is bottom-up and the other is top-down. In fact, they may also be thought of as the way recursion and looping interact. Theoretically, loop and recursion are interchangeable. Consequently, the dynamic programming method's solution can be reached by turning with a two-dimensional array and the dynamic transfer equation that was previously memorized, turning the recursive process into a loop. Each item's choice can be viewed as a stage in the dynamic programming problem.

V. Conclusions

Using the recursive solution approach, finding the transfer equation, which is the same as the recursive formula, is the main objective of the dynamic programming approach. This article presents the fundamental concepts, steps for solving problems, and examples of applications of the dynamic programming method, and specifically explains how the dynamic programming approach differs from other approaches in terms of conversion relations is discussed. The essence of dynamic programming is evaluated and taken into consideration through comparison with other algorithms: apply the solutions to old issues to solve new ones.

References Références Referencias

1. Rethinking algorithm design and analysis, Ananya Levitin. 2019, 32 (1): 14-20.
2. Ulrich Pferschy and Rosario Scatamacchia. Results of improved dynamic programming and approximation for the setups knapsack problem. 2017, 25 (2): 677-662.
3. D. B. Dereventsov and V. F. Temlyakov. a methodical approach to studying several greedy algorithms 2022, 227 (12): 69-54.
4. Teaching Algorithms. SIGACT News, 36 (December 2015), 58–56. Baeza-Yates, R.
5. Programming Pearls, by J. Bentley, Addison-Wesley, 2016.
6. Fundamentals of algorithms, by G. Brassard and P. Bratley, Prentice-Hall, 2016.
7. Algorithms: An Introduction, T. Cormen et al. MIT, 1992.
8. Computer Algorithms, 2019, Computer Science Press, Horowitz et al.

9. Programming Practice, by B. Kernigan and R. Pike. 2005, Addison-Wesley.
10. Should we teach the correct algorithm design techniques? Levitin, A. 179–183 in Proc. SIGCSE '99 (March 2013).
11. Foundations of Algorithms, by R. Neapolitan and K. Naimipour. Jones and Bartlett, 1997, second edition.
12. How to Solve It, by G. Polya, Princeton University Press, 2015.
13. Compared To What? : an Introduction to the Study of Algorithms, J. Rawlins. 2019, Comp. Sc. Press. [14] Algorithms, R. Sedgewick, 2020, Addison-Wesley.
14. The Algorithm Design Manual by Skiena, S. 1997, Springer Verlag.
15. The maximal rectangle problem. Dr. Dobb's Journal, 23 (April 2021), 28–32. Vandervooerde, D.

Global Journal of Computer Science and Technology: H Information & Technology
Volume 23 Issue 2 Version 1.0 Year 2023
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Performance Evaluation of Linux Operating Systems

By Praval Pratap Singh
University of Ajeenkya Dy Patil

Abstract- Recently, Linux has undergone a significant progress and since it provides many helpful features to companies and home users, it has become one of the most commonly used operating systems in IT industry. Considering the popularity of available Linux distributions, three Linux desktop distributions were selected for further evaluation. The idea of this paper is to evaluate performance and compare three different Linux distributions and their influence on processor, memory, graphics system and disk drive performance. Measurements were performed with a three different benchmarking tools specialized for a specific computer component. All Linux operating system distributions were installed on the same desktop computer and based on the achieved performance measurement results it was concluded that the best results were achieved by using Pop OS 20.04. Linux distribution, which is very surprising because it is a relatively new distribution on the market.

Keywords: *benchmark, linux, operating system, performance evaluation.*

GJCST-H Classification: *LCC Code: QA76.9.O35*

Strictly as per the compliance and regulations of:

Performance Evaluation of Linux Operating Systems

Praval Pratap Singh

Abstract- Recently, Linux has undergone a significant progress and since it provides many helpful features to companies and home users, it has become one of the most commonly used operating systems in IT industry. Considering the popularity of available Linux distributions, three Linux desktop distributions were selected for further evaluation. The idea of this paper is to evaluate performance and compare three different Linux distributions and their influence on processor, memory, graphics system and disk drive performance. Measurements were performed with a three different benchmarking tools specialized for a specific computer component. All Linux operating system distributions were installed on the same desktop computer and based on the achieved performance measurement results it was concluded that the best results were achieved by using Pop OS 20.04. Linux distribution, which is very surprising because it is a relatively new distribution on the market.

Keywords: benchmark, linux, operating system, performance evaluation.

I. Introduction

Nowadays, it is difficult to imagine everyday life without the use of computers. Every computer needs an operating system (OS) to work. An operating system is a set of basic system programs that manage computer hardware to perform basic computer functions, that is, it enables a connection between hardware and user programs. One of the most used systems today that contributed to the computer revolution is the Linux operating system named after its original author Linus Torvalds. Analyzing market data for April 2020, Linux ranks third according to [1]. Among many available distributions, three Linux based distribution were selected for further evaluation. Linux Mint 19.4. Tricia and Ubuntu 20.04. LTS were a logical choices because of their significant popularity and reliability. The third selected distribution was Pop! OS, which in 2020 represents an excellent choice judging by the appearance and performance, according to [2] and [3]. Research of this paper focuses on evaluating desktop computer performance with different Linux distributions as host operating system.

The idea behind this paper is to study in detail, describe and compare the architecture, and evaluate performance results of three selected Linux based distribution while putting emphasis on performance of specific components such as processor (CPU), memory (RAM), graphics (GPU) and disk drive, in this case, solid state disk drive (SSD). Next, it is necessary to explain and describe the used benchmark tools.

In the experimental part, it is necessary to install each version of Linux, perform performance measurements and compare the performance of Linux operating systems on a high Performance desktop computer. Benchmark tools have to be installed on a computer for each operating system separately and tests have to be repeated with each benchmark tool five times.

To ensure consistency and accuracy, at least two benchmark tools must be used. Based on the obtained results, it is necessary to draw conclusions about the performance of Linux operating systems, in particularly, performance of each main computer component such as CPU, RAM, GPU and SSD.

The paper is organized in several sections. In section 2 related work was presented. In section 3 some basic information about selected Linux distributions were given. Section 4 describes benchmark tools for testing and evaluating computer performance. Section 5 explains the way in which the asurements were carried out and it gives the basic performance measurement methodology. Also, it explains impact of computer hardware components on the performance results. In section 6 obtained results were presented and discussed. At the end, section 7 gives some concluding remarks.

II. Related Work

Performance often plays a very important role in the field of software evaluation, especially when it is necessary to guarantee the quality, reliability and performance of an operating system [4]. However, it is a complex and time consuming task because it requires controlled environment and same conditions must be applied for all tested systems, so it remains a bit challenging task, especially in a field of operating systems.

In our previous work, several different aspects of an operating systems performance were studied on personal computers. In [5] a comprehensive performance evaluation based on CPU scheduling, disk drive management, graphic subsystem management, memory management and networking performance of the three different versions of Windows operating system, namely Windows XP, Windows Vista and Windows 7, was performed. Also, measurements

Author: Dept- BCA-CT, Ajeenkya Dy Patil University City- Pune, Country-India. e-mail: praval.singh@adypu.edu.in

were conducted in two different environments. Performance measurement results showed better performance of Windows XP in the majority of tests on the low-end computer system when compared to Windows Vista and Windows 7. Furthermore, on the high-end computer system, newer operating systems showed improved performance in the area of graphics subsystem and memory management. Also, a performance evaluation model for Windows operating systems was developed and the similar will be used in this paper with some modifications to accommodate Linux based operating systems. In [6], three popular virtual private server hosts based on Linux Debian operating system were evaluated. In [7], authors evaluated networking performance of three version of Windows operating systems, namely Windows XP, Windows Vista and Windows. Research in the field of networking performance was continued and expanded in [8] where three latest Windows operating systems, namely Windows 7, Windows 8.1 and Windows 10, were evaluated. Experimental results showed that Windows 10 introduced some improvements in network management, delays were lowered, however, CPU usage was increased.

When considering performance of an operating system, one of the less addressed topics is performance of a virtual machine. There are only a few studies of that topic available in the literature. In [9], performance of some typical virtualization techniques while under denial of service (DoS) attack was presented. It was shown that even a light DoS attack suffer from a performance degradation on virtualized servers in comparison to same services on nonvirtualized servers. In [10], performance evaluation of virtualized Windows XP operating system focused on read performance was conducted. Results show that several factors, such as cache configurations, access modes and request sizes can affect throughput. In [11], Windows XP, Windows Vista and Windows 7 operating systems were used as host operating systems while Windows Vista was used as a virtual machine. Performance measurement of a virtual machine was conducted with five different benchmark applications and also by performing data compression and video encoding which are very resources demanding operations. Based on the performance evaluation results it was concluded that using Windows 7 as a host operating system provides the best performance for a virtual operating system. In [11], performance of three different virtual machines was studied. Linux Ubuntu was used as a host while Windows 7, Windows 8.1 and Windows 10 were used as a virtual machine. Evaluation was done with three different measurement tools and results showed that Windows 7 has the best performance when used as a virtual operating system on Ubuntu host.

When discussing Linux operating systems, in [13], authors analyzed real-time performance of Linux kernel between 2.4 and 2.6. They concluded that performance of 2.6 kernel showed better performance. In [14], goal was to measure interactive task response on Linux as well as Windows operating systems. Conclusion was that Linux CPU scheduler has lower latencies than Windows 10 in most cases. In [15], authors presented novel open-source performance analysis tool that gives detailed information about system and application performance. In [4] a novel performance test method for Linux process scheduling was proposed. CFS scheduler was compared with O (1) scheduler and showed better performance both in scheduling efficiency and interactivity. Evaluation of the performance of IPv4 and IPv6 on Windows and Linux operating systems was done in [16]. The performance was examined on TCP and UDP protocols measuring throughput, jitter, delay and CPU utilization. Results showed better performance on Linux based operating system.

III. Host Operating Systems

Linux was created on October 5, 1991 and the first official version of Linux was 0.02 [17]. In the early days, Linux served as an experimental system used by students, hackers, programmers and generally people very closely working with computers. Linux is still a quite rare occurrence on home and office computers but it is constantly growing, especially in IT sector due to the fact that it is known as a secure and stable operating system. The fact that is open-source is also a big advantage for a small, developing companies. Percentage of Linux based operating systems among all used operating systems is still very low, at 1.63%, for servers is about 74.16% and on mobile phones as the Android operating system about 75.16% [18].

In this paper three popular Linux distributions were selected, namely Ubuntu 20.04 LTS, Linux Mint 19.3 and Pop!_OS 20.4, and their performance will be evaluated.

a) Ubuntu 20.04 LTS

Ubuntu [19] originated as a derivative of the Debian system. It uses Linux as the kernel of the operating system. For personal computers, it contains a graphical user interface (GUI). Ubuntu pays the most attention to ease of installation and use. Currently, Ubuntu is the world's most popular opensource desktop operating system.

b) Linux Mint 19.3 Tricia

Linux Mint [20] is an operating system based on the Ubuntu distribution, which in turn is based on Debian. Linux Mint started in 2006 with a beta version called 1.0 "Ada". In 2020, it is one of the most popular

Linux distributions. The latest release is Linux Mint 19.3. Tricia that originated in December 2019, also known as Cinnamon. Linux Mint is similar in terms of the graphical interface to the Windows 10 operating system. It primarily uses free, open-source software.

c) Pop!_OS 20.04

Pop!_OS [21] is one of the new Ubuntu-based distributions by system76 hardware vendors. The basic edition of Pop!_OS contains the graphical user interface GNOME 3.36. There are two different installations, one AMD drivers and the other includes NVidia drivers which makes it the only Linux distribution currently on the market to offer installed graphics card drivers. Pop!_OS encrypts the installation partition by default for better security and privacy. It comes with firmware that allows you to seamlessly upgrade to a newer version of the software. It also comes with a lot of preinstalled development tools, security features and performance profiles. That makes it very popular and community around it is growing fast.

IV. Performance Measurement Tools – Benchmarks

Tools that are specifically developed for evaluating performance of a specific subsystem or a complete computer system are called benchmark applications. Linux is a very complex operating system and therefore it is necessary to use more than one performance measurement tool to get the accurate and relevant data about each computer sub system. There are benchmark tools that integrate the results of all computer components, but they are not suitable for accurate testing. The aim of this paper is to obtain precise, credible and accurate results. For this reason, three benchmark tools were used to evaluate each computer component, namely Hardinfo, Geekbench and Phoronix Test Suite. Within each benchmark tool, there are various tests for each computer subsystem. All the results are expressed in numerical value and unit of measurement. Hardinfo and Geekbench were used to evaluate the CPU and Phoronix Test Suite to evaluate RAM, SSD and GPU.

a) Hardinfo v0.6

Hardinfo (Hardware Information) [22] is a system profile and benchmark tool for Linux based operating systems. It is available in most Linux distributions. It can collect computer and operating system data, run a variety of reference values, and export data to HTML. Hardinfo can evaluate following subsystems: CPU, HDD/SSD, memory cards and USB drives. In this paper, only CPU tests were used as they provide detailed information about CPU performance.

b) Geekbench v5.1.1

Geekbench [23] is a benchmark tool used to evaluate various operating systems. One of the things that sets it apart from other benchmark tools is the fact that it is available on Windows, all Linux distributions, macOS, Android and iOS. In this paper, Geekbench was used to evaluate the CPU. Geekbench uses some advanced application areas for evaluation a performance such as augmented reality and machine learning. On Linux, there is no graphical user interface (GUI), results can only be obtained via terminal or official web page. GeekBench performs single-threaded and multi-threaded tests, making it a great tool for CPU performance evaluation.

c) Phoronix Test Suite v9.6

Phoronix Test Suite [23] was developed in June 2004 by Michael Larabel, who currently works as the owner and editorin-chief. It is free and open-source software. This benchmark tool is one of the most comprehensive tools that can be found on the market, and it is very simple to use. It can be used to test the entire computer system. Running the tests is very simple and the whole process takes only a few minutes after which the results are displayed in an Internet browser. The results are very detailed and can be compared with the results of other users. Linux named it the best platform for evaluating operating system performance. It supports more than 220 test profiles and over 60 test groups. The Phoronix Test Suite can evaluate RAM, SSD and GPU performance.

V. Experimental Setup

a) Performance Measurement Setup

Performance measurement was performed on a high performance desktop computer with a Linux based operating system. Linux distributions used in this paper are namely, Ubuntu 20.04 LTS, Linux Mint 19.3 Tricia and Pop OS 20.04, as they are currently some of the most widely used Linux distributions. All used operating systems were 64-bit. Each operating system was installed with default settings, afterwards, it was updated to include the latest features. Updating operating system is one of the most important steps since updates include various enhancements that improve security, performance and general stability of an operating system. Next important step is to install the latest device drivers since they are constantly upgrading to remove known errors and to improve general performance. While some additional configuration of an operating system can further enhance performance, that aspect was not considered in scope of this paper. Computer hardware and considered operating systems used in this paper are shown in Table 1.

Table I: Hardware and Considered Operating Systems

Hardware components	
RAM	Crucial 2 x 4 GB DDR3 1600 MHz
CPU	Intel® Core™ i5-9400F, 2.9 GHz, 6 cores
GPU	NVidia GeForce GTX 1660 Ti 6GB
SSD	SanDisk 512GB SATA
MBO	Acer Predator PO3-600
Host operating system	
Ubuntu 20.04 LTS v5.4.0-28	
Inux Mint 19.3 Tricia v5.3.0-46	
Pop!_OS 20.04 v5.4.0-7625	

b) Performance Measurement Methodology

It is necessary to create a performance measurement methodology to ensure reliability, repeatability, consistency and accuracy of the conducted experiments and obtained results. Process shown in Fig. 1. represents defined performance measurement methodology used in this paper. The first three steps must be repeated for each host operating system. Upon completing those steps, next five steps must be repeated for each used benchmark application. During experiments only benchmark application should be installed, all other application should be uninstalled, user activity is not allowed and other user files should be deleted since they are likely to degrade performance of an operating system. All measurements have to be repeated five times in a row. The final result is calculated as the arithmetic mean of those five repetitions, as shown in (1), rounded on two decimal places.

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \rightarrow N = 5 \qquad (1)$$

$$x_{imin} \leq \bar{x} \leq x_{imax}$$

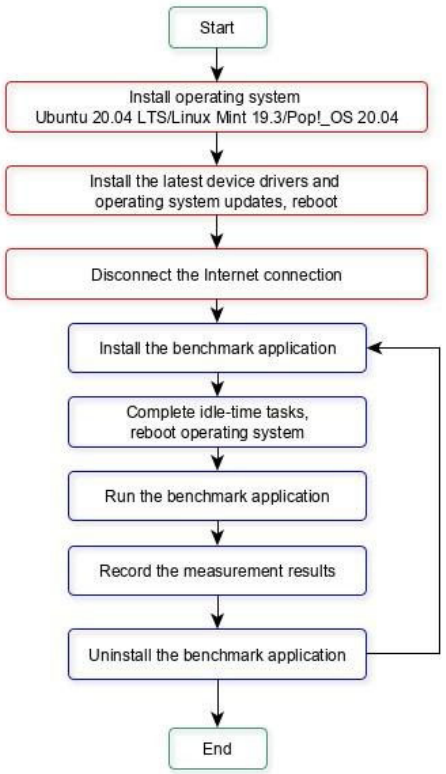


Fig. 1: Performance Measurement Process

Performance results between different operating systems was compared by using Ubuntu results as referent values while percentage difference was calculated for two other tested operating systems with respect to Ubuntu values, as shown in (2).

$$Diff_{\%} = \frac{\text{Mint\_or\_Pop\_val} - \text{Ubuntu\_val}}{\text{Ubuntu\_value}} x100\% \quad (2)$$

The overall performance of the computer and an operating system was evaluated through three main metrics:

- Speed - Represents number of achieved units in period of time, i.e. throughput (Mb/s), more is better,
- Time - Represents required time to perform a certain task, less is better,
- Points - Number of points achieved during a certain task, more is better.

c) Hardware Impact on Performance Measurement Results

Computer systems are very complex and composed of many different interconnected components. All those software and hardware components must work efficiently in order to deliver the best possible performance. With constant advancement in technology, complexity of an operating system is rising and each new version includes a large number of new features and capabilities when compared to the previous ones. It is expected that all those enhancements will deliver some improvement in performance. This is also expected when considering hardware development. Hardware has also a huge impact on computer performance and must be carefully taken into consideration when choosing components for a specific task. Main hardware components that can cause lower performance are mentioned below:

- CPU - The heart of the computer and one of the most important components. It receives and processes data, executes instructions and communicates with input/output units. Insufficient processing power can cause delays in performing tasks if CPU cannot keep up with amount of data that need to be processed,
- GPU - A subsystem that is designed to rapidly manipulate memory to speed up the process of creation images in a frame buffer intended for output to a display device. Similar to CPU, insufficient processing power can cause delays and performance degradation,
- SSD - Is primarily used as storage area. Host operating system as well as used benchmark applications are installed on it. It is important for disk drive to have sufficient throughput to ensure uninterrupted data flow,
- RAM - Random access memory is a form of primary computer memory whose contents can be accessed directly. It contains currently running applications or services run by the user or the operating system. The performance of entire system can degrade significant if current task requires more memory than it is available.

VI. Results and Analysis

In this paper, performance of three Linux based operating systems was evaluated in their default configurations. Three selected operating systems were Ubuntu 20.04 LTS, Linux Mint 19.3 and Pop OS 20.04. Performance measurement was conducted using three different benchmarking tools while the main emphasis was put on CPU, GPU, RAM and SSD performance. When comparing the same performance measurement results, it can be concluded which operating system performs the best in some general tasks. In Tables II-V all performance measurement results are shown.

When comparing CPU test results shown in Table II, it can be seen that both Linux Mint and Pop OS achieved a lot better results than Ubuntu, up to 52%. Only test where Ubuntu showed better result than both other operating systems was calculating Fibonacci order. That result can be disregarded since difference was less than 1%. When comparing Linux Mint to Pop OS it can be seen that in single-threaded tasks they show similar performance while in multi-threaded tasks Pop _OS showed up to 13% better performance. When comparing results of each benchmarking tools, it can be also seen that Hard info tool slightly favorizes Linux Mint while when observing Geek bench results it can be seen that Pop OS achieved better results.

In Table III are presented results of GPU performance evaluation. All tested operating systems showed similar performance. Slight advantage went to Pop _OS which showed up to 3.22% better results than Ubuntu and up to 7% better than Linux Mint. Interesting is the fact that one test, VGT-

Triangle, showed the best performance for Linux Mint, 24.31% better than both Ubuntu and Pop!_OS while all other test were in favor of other two tested operating systems.

When comparing RAM test result, which can be seen in Table IV, again, all operating systems show similar performance, but again, Pop!_OS showed the best performance, up to 1.32% better than Ubuntu and 3.32% better than Linux Mint. Interestingly, one test result, VPT-TriadInteger, showed a lot worse performance for Linux Mint, 20.24% less than Ubuntu which is a way worse than in average.

By evaluating performance of SSD, a lot more diverse results were obtained, as shown in Table V. For example, in sequential disk write test Ubuntu achieved 12.53% better result than Pop!_OS and 29.95% better result than Linux Mint. On the other hand, Linux Mint achieved the best result in VPTUnpacking test where it achieved 4.2% better result than Pop!_OS and 4.9% better than Ubuntu. Pop!_OS achieved the best result in test VPT-1000files-NoSync, where it was 12.57% better than Ubuntu and 11.1% better than Linux Mint. RAM test showed that all operating systems have some advantages which ensured them the best results in certain tests. In average, it can be concluded that the best SSD results achieved Ubuntu.

VII. Conclusion

When considering operating systems, it is obvious that their main purpose is to enable an efficient usage of all resources. In recent time Linux has undergone a significant progress and since it provides many helpful features to companies and home users, it has become one of the most commonly used operating systems in IT industry. Due to the fact that it is gaining popularity rapidly but it is still relatively unknown to most other non-Linux user, the idea of this paper is to take some of the most popular Linux distributions and evaluate their performance in everyday tasks. Three Linux based operating systems were evaluated, namely Ubuntu 20.04 LTS, Linux Mint 19.3 and Pop!_OS 20.04. Evaluation was done by using three benchmark applications, Hardinfo, Geekbench and Phoronix Test Suite. During the evaluation process, emphasis was put on CPU, GPU, RAM and SSD performance.

Performance measurement results showed that relatively new Linux distribution, Pop!_OS, has the best performance while the most widely used distribution, Ubuntu, came second. Ubuntu showed the best results in SSD tests while Pop!_OS showed better results for all other tested components. Linux Mint came last but not far behind Ubuntu. For future work it would be very interesting to compare performance of Linux distributions with most commonly used versions of Windows operating systems since Windows is still most widely used operating system.

Table II: CPU Test Results

Performed test	Ubuntu	Linux Mint	Pop!_OS	Unit	Diff LM to Ubuntu	Diff PopOS to Ubuntu
CPU-Blowfish	1.17	1.19	1.14	sec	-1.30%	2.34%
CPU-CryptoHash	9067.57	9109.80	10385.28	Mb/sec	0.40%	4.50%
CPU-Fibonacci	0.43	0.43	0.43	sec sec	-0.46%	-0.93%
CPU-N-Queens	7.01	6.68	6.88	HIMarks	4.70%	1.85%
CPU-Zlib	1.65	1.78	1.68	Sec	7.76%	1.56%
FPU-FFT	0.61	0.62	0.59		-0.98%	2.60%
Single-thread				score		
NPT-CPU-single-thread	1091.60	1104.20	1102.00	score	1.15%	0.95%
VPT-Encryption	113.60	114.60	113.60	score	0.88%	0.88%
VPT-IntegerMath-	1088.00	1097.40	1096.00	score	0.86%	0.73%
VPT-FloatingPointMath	1261.80	1284.40	1279.80	MB/sec	1.79%	1.42%
VPT-TextCompression	990.80	1036.00	1032.60	Mpixels/sec	4.56%	4.21%
VPT-ImageCompression	1140.00	1150.00	1147.60	Images/sec	0.87%	0.66%
VPT-MachineLearning	1107.00	1196.20	1198.00		7.45%	7.60%
Multi-thread						
NPT-CPU-total	3921.00	4935.80	5292.60	score	25.88%	34.98%
VPT-Encryption	508.60	588.20	654.20	score	15.65%	28.63%
VPT-IntegerMath	3929.60	4986.00	5332.40	score	26.88%	35.70%
VPT-FloatingPointMath	4471.00	5576.00	5980.00	score	24.71%	33.75%
VPT-TextCompression	3066.20	4428.80	4661.40	MB/sec	44.44%	52.03%
VPT-ImageCompression	5481.80	5767.20	6471.60	Mpixels/sec	5.21%	18.06%
VPT-MachineLearning	2707.80	2690.40	2721.00	Images/sec	-0.64%	0.49%

Table III: GPU Test Results

Performed test	Ubuntu	Linux Mint	Pop!_OS	Unit	Diff LM to Ubuntu	Diff PopOS to Ubuntu
NUH-FPS-D3D9	110.79	111.32	114.35	FPS	0.48%	3.22%
NUH-FPS-D3D11	111.12	110.91	114.59	FPS	-0.19%	3.13%
NUH-FPS-OpenGL	111.69	110.85	114.95	FPS	-0.75%	2.92%
NUT-FPS	284.33	282.68	293.24	FPS	-0.58%	3.13%
NUS-FPS	354.11	351.92	360.20	FPS	-0.62%	1.72%
VGT-GiMark	7915.80	7451.00	8008.60	score score	-5.87%	1.17%
VGT-Plot3D	86958.80	87078.00 7268.00	86886.60	score score	0.14%	-0.08%
	6981.80	19963.60	6976.40	score score	4.10%	-0.07%
	21044.60	157746.00	21068.40	score	-5.14%	0.11%

Table V: SSD Test Results

Performed test	Ubuntu	Linux Mint	Pop!_OS	Unit	Diff LM to Ubuntu	Diff PopOS to Ubuntu
VGT-Plot3D	6981.80	19963.60	6976.40	score score	4.10%	-0.07%
VGT-FurMark	21044.60	157746.00	21068.40	score	-5.14%	0.11%
VGT-TessMark	126924.40	2831.00	126842.20		24.28% -	-0.07%
VGT-Triangle	2896.20	5498.00	2893.80		2.25%	-0.08%
VGT-PixmarkPiano	5681.00		5702.60		-3.22%	0.38%
VGT-Volplosion						
VPT-Add-Integer	23212.53	22478.37	23182.48	MB/sec	-3.16%	-0.13%
VPT-Copy-Integer	20374.75	20393.41	20346.04	MB/sec	0.09%	-0.14%
VPT-Scale-Integer	20215.29	20244.68	20281.90	MB/sec	0.15%	0.33%
VPT-Triad-Integer VPT-Add-Float	22964.19	18314.06	23034.80	MB/sec	-20.24%	0.31%
VPT-Copy-Float	23004.66	22495.91	23058.83	MB/sec	-2.21%	0.24%
VPT-Scale-Float VPT-Triad-Float	20281.12	20414.57	20326.60	MB/sec	0.65%	0.22%
VPT-MBW-128MB	20320.05	20393.99	20369.09	MB/sec	0.36%	0.24%
VPT-MBW-Fx128MB VPT-Latency	23024.25	22500.23	23099.86	MB/sec	-2.28%	0.33%
	13461.25	13616.97 8672.61	13483.04	MiB/sec	1.16%	0.16%
	8498.87	15026.10	8611.88	MiB/sec	2.04%	1.33%
	15177.63		14973.98	MB/sec	-0.99%	-1.34%

Table IV: Ram Test Results

Performed test	Ubuntu	Linux Mint	Pop!_OS	Unit	Diff LM to Ubuntu	Diff PopOS to Ubuntu
VPT-DiskSeqRead	556857.80	566100.66	572420.20	score score	1.66%	2.79%
VPT-DiskSeqWrite	35843.60	25108.66	31352.00	MB/sec	-29.95%	-12.53%
VPT-CacheRead	18100.65	12682.67	17947.04	MB/sec sec	-26.93%	-0.85%
VPT-BufferedDiskRead VPT-UnpackingLK	192.18 5.94	180.56	191.98 5.90	files/sec	-6.04%	-0.10%
VPT-1000files,1Mbsize	283.84	5.65	270.66	files/sec	4.90%	0.69%
VPT-5000files,1Mbsize,4Threads	367.92	282.04	370.40	files/sec	-0.63%	-4.63%
VPT-4000files,32SubDirs	205.68	363.57	192.62	files/sec	-1.18%	0.67%
VPT-1000files,1MBSize,No Sync/Fsync	2454.82	170.40	2516.88	MB/sec	-17.15%	-6.35%
VPTRead-RS:1MB,FS:8GB	10545.32	2234.82	11870.39	MB/sec	-8.96%	2.53%
VPTRead-RS:4KB,FS:8GB	4791.72	10708.96	5126.36	MB/sec	1.55%	12.57%
VPTWriteRS:1MB,FS:8GB	816.99	4815.24	748.38	MB/sec	0.49%	6.98%
VPTWrite-RS:4KB,FS:8GB	728.93	718.11	749.67		-12.10%	-8.40%
		675.33			-7.35%	2.85%

References Références Referencias

1. Desktop operating system market share worldwide, https://gs.statcounter.com/os-market-share/desktop/worldwide [Accesed: May 2020]
2. It's FOSS Review, Hands on with System 76's beautiful Linux distro Pop!_OS, https://itsfoss.com/pop-os-linux-review/ [Accesed: May, 2020]
3. J.Evangelho, 3 reasons why Pop!_OS is the best Ubuntu based Linux distribution, https://www.forbes.com/sites/jasonevangelho/2020/01/24/3-reasonswhy pop-os-is-the-perfect-choice-for-linux-gamers/#2b0c4bc16096 [Accesed: May, 2020]
4. L. Yuqing, X. Hao, L. Xiaohui, The research of performance test method for Linux process scheduling, Fourth International Symposium on Information Science and Engineering, Shanghai, 2012. pp. 216–219.
5. G. Martinovic, J. Balen, B. Cukic, Performance evaluation of recent windows operating systems, J Universal Comput Sci (JUCS), 2012, Vol 18, No 2, p. 218–263.
6. J. Balen, D. Vajak, K. Salah, Comparative performance evaluation of popular virtual private servers", Journal of Internet Technology, Vol. 21, No. 2, 2020, pp. 343-356.
7. J. Balen, G. Martinovic, Z. Hocenski, Network performance evaluation of latest Windows operating systems, Proceedings of the 20th International Conference on Software, Telecommunications and Computer Network, SoftCOM, Split, Croatia, 2012, pp. 1–6.
8. K. Vdovjak, J. Balen, K. Nenadic, Experimental Evaluation of Desktop Operating Systems Networking Performance, International Journal of Electrical and Computer Engineering Systems (IJECES), 2020, Vol 11, No 2, pp. 67-76
9. R. Shea, G. Liu G., Performance of virtual machines under networked Denial of service attacks: experiments and analysis, IEEE Syst J., 2013, Vol. 7, No. 2, pp. 335–345.
10. Y. Zhang, G. Su, L. Hong, On virtual-machine-based windows file reads: a performance study,

Proceedings of the IEEE Pacific-Asia View publication stats Workshop on Computational Intelligence and Industrial Application, Wuhan, 2008, pp. 944–948.
11. G. Martinovic, J. Balen, S. Rimac-Drlje, Impact of the host operating systems on virtual machine performance, The 33rd International Convention MIPRO, Opatija, 2010, pp. 613–618.
12. J. Balen, K. Vdovjak, G. Martinovic, Performance evaluation of Windows virtual machines on a Linux host", Automatika, Vol. 61, No. 3, 2020, pp. 425-435.
13. Z. Yanyan, R. Xiangjin, Analysis of Linux kernel's real-time performance, 2018 International Conference on Smart Grid and Electrical Automation, pp. 191-194
14. W.C. Fan, C.S. Wong, W.K. Lee, S.O. Hwang, Comparison of interactivity ierformance of Linux CFS and Windows 10 CPU schedulers, 2020 International Conference on Green and Human Information Technology (ICGHIT), pp. 31-34.
15. T. Ilsche, R. Schöne, M. Bielert, A. Gocht and D. Hackenberg, lo2s - Multi-core system and application performance analysis for Linux, 2017 IEEE International Conference on Cluster Computing (CLUSTER), Honolulu, HI, 2017, pp. 801-804.
16. O. Opeyemi, Performance evaluation of Linux and Microsoft Windows operating systems on configured tunnels of IPv6 and IPv4, Computer Science and Telecommunications, 2016, Vol. 50, No. 4, pp. 93-113.
17. What is Linux, https://sarata.com/what_is_linux.html [Accesed: May, 2020]
18. 111+ Linux Statistics and Facts, https://hosting tribunal.com/blog/linuxstatistics/#gref [Accessed: May 2020].
19. Ubuntu, https://www.ubuntu.com [Accessed: May 2020]
20. Linux Mint, https://www.linuxmint.com/ [Accessed: May 2020]
21. Pop!_OS, https://pop.system76.com/ [Accessed: May 2020]
22. Hardinfo, Hardware Information benchmark, http:// www.hardinfobenchmark.com/ [Accessed: May 2020]
23. Geekbench 5, Geekbench benchmark, http://www. geekbench.com, [Accessed: May 2020]
24. Phoronix Test Suite, Phoronix Test Site benchmark, http://phoronix testsuite.com, [Accessed: May 2020]

Global Journal of Computer Science and Technology: H Information & Technology
Volume 23 Issue 2 Version 1.0 Year 2023
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Hybrid Approach: Project Management for Construction Projects

By Yaiseth Frangakis Cano

Iberoamerican International University

Abstract- The purpose of this research is to innovate in the field of project management by proving the benefits of applying the Hybrid Project Management Methodology to construction projects, an industry that is currently developing in complex environments with high levels of change.

This research evidenced that complementing and integrating some Agile Methods with Traditional Project Management Methodology taking into consideration the specific characteristics of each construction project; improves project execution with flexibility, adaptation, continuous delivery of value, motivation, effective communication, and constant feedback throughout its life cycle; improving the results and fulfillment of objectives.

The need felt by construction professionals to use methodologies focused on guiding, improving, and developing their project management skills was also demonstrated.

Also, it was statistically validated that in the construction sector it is appropriate to apply the Hybrid Project Management Methodology with a result of 3.903, being 4. "Very appropriate".

Keywords: hybrid approach, agile methods, construction projects, delphi technique, hybrid methodology, project management, traditional methodology, agile, construction industry.

GJCST-H Classification: LCC code: TH1-9745

Strictly as per the compliance and regulations of:

Hybrid Approach: Project Management for Construction Projects

Yaiseth Frangakis Cano

Abstract- The purpose of this research is to innovate in the field of project management by proving the benefits of applying the Hybrid Project Management Methodology to construction projects, an industry that is currently developing in complex environments with high levels of change.

This research evidenced that complementing and integrating some Agile Methods with Traditional Project Management Methodology taking into consideration the specific characteristics of each construction project; improves project execution with flexibility, adaptation, continuous delivery of value, motivation, effective communication, and constant feedback throughout its life cycle; improving the results and fulfillment of objectives.

The need felt by construction professionals to use methodologies focused on guiding, improving, and developing their project management skills was also demonstrated.

Also, it was statistically validated that in the construction sector it is appropriate to apply the Hybrid Project Management Methodology with a result of 3.903, being 4. "Very appropriate".

Keywords: hybrid approach, agile methods, construction projects, delphi technique, hybrid methodology, project management, traditional methodology, agile, construction industry.

I. Introduction

Actualmente la industria de la construcción se está convirtiendo en uno de los motores globales más importantes para el crecimiento económico mundial. De acuerdo con Marsh y GuyCarpenter (2021): "Future of Construction. A Global Forecast for Construction to 2030"; se proyecta que para el año 2030 el crecimiento de la construcción en Asia-Pacífico será superior al 50%, en América del Norte crecerá 32%, en Europa Occidental crecerá 23%, en Latinoamérica se proyecta un crecimiento de 35% y a nivel mundial se proyecta un crecimiento de 42%.

Por otro lado, el entorno actual en el que se están desarrollando los proyectos de construcción ha cambiado, es más complejo, con mayores retos, un ambiente muy versátil en el que surgen situaciones y problemas inesperados que desafían la gestión de proyectos afectando potencialmente los resultados.

El sector construcción se está enfrentando a grandes cambios: las consideraciones Ambientales, Sociales y. Gubernamentales (ESG), nuevos y modernos métodos de construcción, la industria de deconstrucción. La dificultad se incrementa ya que la mayoría de los profesionales que dirigen los proyectos no son especialistas de Dirección de Proyectos o no aplican los conocimientos de la forma adecuada; esto se refleja en el incumplimiento de los objetivos. Todos estos factores están cambiando el perfil de riesgos del sector construcción, pero a la vez creando oportunidades.

El crecimiento continuo, cambios y desafíos a los que se está enfrentando el sector construcción, requiere un cambio de mentalidad y la utilización de prácticas innovadoras para lograr una buena ejecución y vencer los obstáculos para la culminación de proyectos de acuerdo con sus restricciones.

Es importante que la Metodología de Dirección de Proyectos aplicada se mantenga en constante actualización, crecimiento gradual y mejoramiento continuo de técnicas y métodos aplicados para generar indicadores de desempeño y evaluaciones para obtener una visión general y oportuna del comportamiento de los proyectos.

La implementación de procesos más eficientes produce ahorros de recursos y costos, cumplimiento con la calidad y tiempo estipulado, permitiendo incrementar la posibilidad de éxito de los proyectos con clientes satisfechos e incremento de márgenes de utilidad para las empresas.

Para enfrentar los retos de la industria de la construcción es imprescindible que exista una orientación enfocada a mejorar las habilidades de los profesionales que se dedican a la construcción, y estos a su vez deben mantener una mentalidad enfocada en la educación y mejora profesional continua.

De acuerdo con los resultados de la presente investigación, los mayores problemas de las construcciones están en el incumplimiento del tiempo de entrega, incremento del presupuesto, insatisfacción de los clientes y calidad deficiente. Incrementándose por la ineficiente resolución de problemas e incidentes, riesgos no detectados oportunamente, baja capacidad para corregir fallas, poca adaptación a los cambios, falta de objetivos claramente identificados por el cliente y su poca participación en el proyecto.

La Metodología Híbrida de Dirección de Proyectos, validada en esta investigación, se fundamenta en la normativa internacional del Project Management Institute UNE-ISO 21500 para dirección de

Author: Universidad Internacional Iberoamericana. Doctorado en Proyectos – Unini- Mx.
e-mail: yaiseth.frangakis@doctorado.unini.edu.mx

proyectos; documento de consenso más actualizado que define el lenguaje universal de la dirección de proyectos tradicional. Para el análisis de las Metodologías Ágiles se utilizó: la Guía Práctica y los Fundamentos de los Métodos Ágiles para determinar aquellos Métodos Ágiles con mayores aportes al integrarse con los grupos de procesos de la Metodología Tradicional. Se utilizó también la Guía de los fundamentos para la dirección de proyectos (PMBOK), que posee conceptos de las Metodologías Tradicional, Ágil e Híbrida y de esta forma complementar y fortalecer con conceptos ágiles e híbridos las prácticas actuales más utilizadas en la dirección para proyectos de construcción.

La presente investigación se centra en el diseño y validación de una Metodología Híbrida para la dirección de proyectos de construcción, ofreciendo a sus profesionales innovadoras prácticas que puedan aplicar durante la ejecución de sus proyectos permitiéndoles lograr los resultados finales esperados.

Los objetivos principales del presente trabajo de investigación son los siguientes:

- Determinar los principales problemas de realizar una dirección de proyectos de construcción deficiente.
- Desarrollar una Metodología Híbrida para dirección de proyectos de construcción en base a los resultados obtenidos del análisis estadístico de los estudios aplicados a los profesionales del sector construcción.
- Analizar las características de los métodos ágiles e identificar los más favorables para aplicarlos en el sector construcción.
- Validar mediante la Técnica Delphi la Metodología Híbrida de Dirección de Proyectos de construcción que integre la Metodología Tradicional con Métodos Ágiles.

II. Literature Review

Los criterios de búsqueda de información utilizados para la confección de la presente investigación fueron: la actualidad del contenido, su contextualización y relevancia. El proceso consistió en recopilar y organizar bibliográficamente artículos de revistas especializadas, libros, bases de datos estadísticas, normativas y estándares actuales de dirección de proyectos.

Para la selección de la información se consideraron criterios epistemológicos y el nivel de representatividad; es decir, la estructura lógica del material, los aspectos más relevantes del contenido y los aportes de los autores más destacados del tema.

Se realizó una revisión de la Norma ISO 21500 y los fundamentos de los métodos ágiles para lograr una Metodología Híbrida, fundamentada en la Metodología Tradicional complementada con técnicas y métodos ágiles que aportan flexibilidad y mayor adaptación a las condiciones específicas de cada proyecto.

Se analizaron las técnicas y herramientas prácticas de dirección de proyectos y su correspondencia con los procesos establecidos en la Norma ISO 21500 y su posible complemento e integración con herramientas y técnicas ágiles.

El proceso de recogida de datos de campo y generación de la información se realizó a partir de cuestionarios, para conocer la problemática real que enfrenta la dirección de proyectos de la construcción, así como sus posibles alternativas de solución.

La información recopilada, se analizó estadísticamente mediante el programa SPSS Statistics.

a) Global Context of the Problem

De acuerdo con la Guía de los Fundamentos para la Dirección de Proyectos PMBOK Guide (2021), la dirección de proyectos es definida como: la aplicación de conocimientos, habilidades, herramientas y técnicas aplicadas a las actividades del proyecto para cumplir con sus requisitos. Según M. De Cos (1997), la dirección de proyectos como sistema facilita la identificación y estudio de los distintos subsistemas que la integran: la planificación, la organización, el control, la información, la tecnología y el ambiente sociocultural, son los subsistemas que permiten la comprensión de la dirección de proyectos.

A pesar de la utilización sistemática de las herramientas y técnicas de dirección de proyectos, la mayoría de las veces no se culminan los proyectos de la forma esperada.

Investigaciones realizadas a partir de 1950 determinaron que, de cada 100 proyectos, únicamente 17 se consideraban exitosos. Según investigaciones actuales de H. Duncan (2012), 32% de los proyectos son culminados dentro de sus restricciones, demostrando que se ha mejorado; pero también que los métodos siguen siendo ineficientes, que debe incrementarse profesionalismo a las estrategias y evitar el empirismo.

Como lo indica G. Montero (2012), a partir de 1965 se inició la creación de organizaciones dedicadas al desarrollo y promoción de la dirección de proyectos, en 1965 se fundó en Europa el International Project Management Association y en 1969 el Project Management Institute en Norte América; con la finalidad de documentar, estandarizar la información y las prácticas generalmente aprobadas en la dirección de proyectos. Dando origen a la Metodología Tradicional de proyectos, cuya primera guía estandarizada fue publicada en 1987. Su aplicación es mediante acciones secuenciales realizadas en cada uno de los procesos establecidos en los grupos de procesos o de materia mediante herramientas, técnicas, habilidades y

tecnologías que permiten su ejecución; teniendo siempre en consideración los objetivos del proyecto.

La dirección de proyectos siguió evolucionando y en el año 2001 surgió el Manifiesto Ágil, debido a que los desarrolladores de software enfrentaban problemas utilizando las prácticas tradicionales existentes; los proyectos se caracterizaban por la incertidumbre, complejidad, cambiantes y fuertemente afectados por el entorno, la tecnología, la innovación y el mercado. Es por esto por lo que la metodología ágil de dirección de proyectos permite adaptar la forma de trabajo a las condiciones del proyecto, consiguiendo flexibilidad e inmediatez en la respuesta para amoldar el proyecto y su desarrollo a las circunstancias específicas del entorno.

Por mucho tiempo los dos enfoques de dirección de proyectos tradicional y ágil se han confrontado, cada uno defendiendo sus conceptos y métodos; indicando por qué uno es mejor que el otro. Actualmente la visión de los expertos internacionales ha cambiado, como indica Teixido J. (2016) "Es hora de una nueva iteración en la reflexión sobre cómo se relacionan los enfoques ágiles con el predictivo. Añade que, los mejores equipos de proyecto no son los que adoptan una sola metodología y luego cierran sus mentes a otras opciones. Se debe adaptar el enfoque al equipo, al proyecto y a la organización, aun cuando eso implique ir más allá de una sola metodología; se deben conocer ambas metodologías para buscar fortalezas y adaptarlas de acuerdo con los objetivos requeridos por el proyecto dando origen a la Metodología Híbrida de dirección de proyectos.

Actualmente los profesionales de la construcción se enfrentan a muchos desafíos: Optimización de recursos y cronogramas, globalización, revolución tecnológica y digital, inteligencia artificial, nuevas formas de construcción verde y sostenible, escasez de adquisiciones, equipos remotos, respuesta rápida al cambio, responsabilidad ambiental y social, alineación de los proyectos con la estrategia. Basado en ese entorno, la Guía práctica de ágil (2017) indica que, no es necesario usar un único enfoque híbrido de dirección de proyecto para el desarrollo de un proyecto, pueden combinarse diferentes enfoques durante su ciclo de vida con la finalidad de lograr los objetivos; considerando los requerimientos y las características específicas del proyecto.

La guía práctica de ágil (2017), sugiere que los enfoques de dirección híbridos pueden ser: Enfoque ágil seguido por un enfoque predictivo. Enfoques ágiles y predictivos combinados al mismo tiempo. Enfoque predictivo con algunos componentes ágiles. O un enfoque ágil con un componente predictivo. La selección del enfoque de dirección más conveniente para el proyecto; depende de sus características propias, importante recordar que ningún proyecto es igual a otro.

III. Metodology

En la presente investigación, se utilizó el método inductivo partiendo de normativas y estándares internacionales existentes, que son las guías internacionales previamente desarrolladas y aceptadas como válidas por la comunidad científica para lograr una adecuada dirección de proyectos mediante su aplicación.

De acuerdo con C. Collado, P. Lucio, R. Sampieri (2003), la presente investigación es de tipo descriptiva, ya que se identificaron y describieron características para estimular nuevo conocimiento. Se recolectaron los datos, y se llevó a la predicción e identificación de las relaciones existentes entre dos o más variables.

Con este tipo de investigación se analiza el estado de una o más variables en un momento dado, o la relación entre un conjunto de variables en un punto en el tiempo.

a) Design of the Investigation

La presente investigación recolectó, analizó y vinculó datos cuantitativos y cualitativos en un mismo estudio, según A. Tashakkori y C. Teddlie C. (2003), se considera un diseño con un enfoque mixto.

El método cuantitativo se aplicó en la recolección y el análisis de los datos; se utilizaron mediciones numéricas, el conteo y la estadística para determinar el comportamiento de la población.

Con el método cualitativo se profundizó en el campo de estudio y permitió realizar las observaciones y descripciones para afinar los datos recolectados cuantitativamente.

En una investigación con enfoque mixto, se usan tanto los métodos cuantitativos como los cualitativos, a la vez que se puede convertir datos cualitativos en cuantitativos y viceversa, según D. Mertens (2015).

Desde el punto de vista operativo, el diseño de la investigación fue no experimental, enfocado en el estudio de la realidad en su dinámica natural para describirla, explicarla y predecirla. Con relación a la dimensión temporal de la investigación y el número de momentos en el tiempo en que se recolectaron los datos, la modalidad del diseño de la investigación fue transversal correlacional, según C. Collado, P. Lucio, R. Sampieri. (2003). Lo que permitió analizar el nivel de una o varias variables en un momento dado, o la relación entre un conjunto de variables, en un punto del tiempo; facilitando la información de la realidad en un momento específico abarcando varios indicadores.

b) Hypothesys

Las hipótesis que se validaron en la presente investigación fueron las siguientes:

H1: El juicio unificado de los expertos consultados valida que estandarizar los procesos de dirección de

proyectos de construcción en las organizaciones; mediante la aplicación de una metodología híbrida de dirección de proyectos (tradicional complementada con métodos ágiles) que cuente con herramientas, técnicas y plantillas ayuda al logro de los objetivos organizacionales y la maduración empresarial.

H2: El criterio consensuado de los expertos consultados corrobora que con la aplicación de una Metodología Híbrida de Dirección de Proyectos (tradicional complementada con ágil) basada en las características propias del proyecto mejoraría los resultados finales del proyecto.

H0: La opinión homologada de los expertos consultados es que no existe diferencia entre el logro de los objetivos organizacionales, la madurez empresarial y los resultados finales de los proyectos de construcción de una organización que estandariza sus procesos de dirección de proyectos con una metodología híbrida (tradicional complementada con métodos ágiles), que cuente con las herramientas, técnicas y plantillas según las características de sus proyectos, y una organización que no utiliza los procesos de dirección de proyectos estandarizados con metodología híbrida de dirección de proyectos para construcciones.

c) Sampling and Sample Subjects

Los sujetos de la muestra de la presente investigación fueron elegidos por tener características esenciales para la investigación, está técnica de muestreo aplicada se denomina no probabilístico intencional.

Los criterios para la selección de los sujetos de la muestra fueron los siguientes:

- Profesional de la construcción o dirección de proyectos.
- Dirigir proyectos de construcción o pertenecer a un sector relacionado.
- Contar con estudios en dirección de proyectos.
- Cinco años o más de experiencia en proyectos.
- Conocimiento de los métodos tradicionales, ágiles o híbridos para la dirección de proyectos.

Para obtener el tamaño ideal de la muestra finita y conocida, se utilizó la fórmula de Spiegel M. y Stephens L. (2009). La población objetivo fue de 158 profesionales del Colegio de Ingenieros Civiles de Panamá y 136 profesionales del *Project Management Institute* capítulo de Panamá; un total de 294 profesionales que cumplieron con los criterios establecidos para selección de la muestra.

A continuación, la ecuación (1) utilizada para obtener el tamaño ideal de la muestra.

Los parámetros establecidos para su determinación fueron los siguientes: nivel de confianza del 95%, con una proporción del 50% (debido a que no se tiene caracterizada esta población) y un error máximo permisible de 5%.

$$n = \frac{N\frac{Z^2}{e^2}\cdot p\cdot q}{N+\frac{Z^2}{e^2}} \quad (1)$$

Donde:

n = Tamaño de la muestra.

N = Tamaño de la población (294 Profesionales de construcción)

p y q = Proporciones complementarias que toman cualquier valor entre 0 y 1

p = 0.5

q = 0.5

Z = Nivel de confianza de 95%. Valor constante 1.96

e = Error estándar o error debido al muestreo 5%

De acuerdo con los cálculos obtenidos y los criterios de selección establecidos, el tamaño ideal de la muestra de la investigación fue de 62 profesionales de dirección de proyectos de construcción.

El cuestionario se aplicó a 86 profesionales de dirección de proyectos, al realizar el análisis estadístico mediante el programa SPSS, 72 cuestionarios resultaron válidos y 14 fueron excluidos por estar incompletos o con errores.

Se cumplió con el cálculo del tamaño de la muestra ideal obtenido mediante la fórmula de Spiegel M. y Stephens L. $n = 62$, en donde la muestra real utilizada en la investigación fue de $n = 72$.

d) Data Collection Instrument

El instrumento de recolección de datos de la presente investigación fue el cuestionario aplicado a profesionales de la construcción y la dirección de proyectos.

Se lograron 72 cuestionarios válidos, con lo que se obtuvieron los datos y la información base para el desarrollo de la metodología híbrida aplicada a la industria de la construcción.

El cuestionario se formuló con 42 ítems pertenecientes a 9 dimensiones relacionadas con los diferentes aspectos investigados. Se utilizaron preguntas cerradas que aportaron datos relacionados al dominio personal y la conducta profesional de los participantes; su experiencia y el conocimiento actual de la metodología de dirección de proyectos.

Para recopilar información acerca de las opiniones, el nivel de información y las expectativas acerca de la dirección de proyectos; Se empleó una escala tipo Likert en la que los profesionales valoraron cada ítem del cuestionario otorgando una puntuación de tres grados: bajo (1), medio (3) y alto (5).

La información base obtenida fue la siguiente: el nivel de cumplimiento de las restricciones principales del proyecto aplicando la metodología de dirección de proyecto que actualmente utiliza, los aspectos de la metodología de la dirección de proyectos que utiliza que pueden mejorarse, la influencia de los aspectos externos en la metodología de dirección de proyectos que afectan la ejecución del proyecto, la afectación de las empresas por fallas en los proyectos, los aspectos de mejora para la metodología para la dirección de proyectos que actualmente utiliza, expectativas para una metodología de dirección de proyectos ideal y aquellas aplicaciones que son consideradas las más importantes al ejecutar un proyecto de construcción.

e) Variables of the Investigation

Las variables independientes son explicativas, cuya asociación o influencia en la variable dependiente se pretende descubrir en la investigación.

Las variables independientes utilizadas en la recopilación de datos de esta investigación fueron: sexo, nivel educativo, industria, utilización de metodología de dirección de proyectos, tipo de metodología de dirección utilizada, método de dirección de proyectos más aplicado, gestión de dirección de proyectos, herramientas, técnicas o plantillas de métodos ágiles que puedan complementar los grupos de procesos de los métodos tradicionales.

Las variables dependientes son aquellas que pueden hipotéticamente ser influidas por una variable independiente, por lo tanto, el comportamiento de la primera depende de la segunda.

Las variables dependientes utilizadas fueron: resultados del proyecto obtenidos con la aplicación de alguna metodología de dirección de proyectos, herramientas o técnicas.

IV. Statistical Criteria for Data Analysis

Los datos y resultados obtenidos de los cuestionarios de investigación fueron analizados estadísticamente mediante el programa SPSS, corroborando la fiabilidad y la validez del constructo del cuestionario; permitió crear la matriz de datos y variables, análisis y representaciones gráficas de los resultados y el resumen de datos.

La fiabilidad del cuestionario de investigación fue determinada mediante los siguientes criterios:

- *Índice De Discriminación:* que indica el grado en que un ítem contribuye a la consistencia interna del cuestionario reforzando su carácter unidimensional, se determinó mediante la correlación total de los elementos corregida.

El índice de discriminación de los ítems fue superior a 0.20, se eliminaron los ítems con un índice de discriminación en el rango de 0.10 a 0.19 que de acuerdo con A. Pantoja (2015) están en el límite.

- *El coeficiente Alfa De Cronbach:* Se utilizó como modelo de consistencia interna para el análisis del cuestionario de la investigación, se basa en las correlaciones de los ítems aportando fiabilidad a las preguntas. Su valor oscila entre 0 y 1 considerando como criterio general un coeficiente aceptable cuando su valor es igual o superior a 0.70.

En la Tabla 1, se muestra el coeficiente de Alfa de Cronbach para el cuestionario de esta investigación que fue de 0.815, calculado para las 20 variables con índices de discriminación mayores de 0.2.

Para verificar la validez del constructo, se comprobó que las relaciones entre las variables del cuestionario de investigación tenían una estructura dimensional invariable; necesaria para interpretar los resultados. Este análisis se realizó mediante el índice Kaiser-Mayer-Olkin (KMO) que varía entre 0 y 1, un valor menor a 0.5 indica que la correlación entre dichas variables no es suficientemente significativa.

Se confirmó el resultado mediante la prueba de esfericidad de Barlett, que contrasta la hipótesis nula, si el nivel de significación del estadístico de Bartlett es mayor que 0.05 no se puede rechazar la hipótesis nula y, por tanto, no se podría realizar análisis factorial del cuestionario.

Los criterios para determinar la validez del constructo fueron: Índice de Kaiser-Mayer-Olkin (KOM) mayor a 0.50, confirmado por la prueba de esfericidad de Barlett menor de 0.050.

En la Tabla 2 se muestran los índices de verificación de la validez del constructo del cuestionario de investigación.

Tabla. 1: Fiabilidad Del Cuestionario De Investigación.

Alfa de Cronbach	Alfa de Cronbach basada en elementos estandarizados	N° de elementos
0.806	0.815	20

Tabla. 2: Validez Del Constructo Del Cuestionario De La Investigación.

Medida Kaiser-Meyer-Olkin de adecuación de muestreo.	0.693
Prueba de esfericidad de Barlett	0.000

Basado en los resultados de los análisis estadísticos, los criterios de fiabilidad y validez del constructo del cuestionario de investigación se cumplen: la consistencia interna del cuestionario es buena ya que las variables tienen el índice de discriminación mayor a 0.2.

La fiabilidad del cuestionario de investigación es aceptable, ya que el índice de Alfa de Cronbach es $0.815 > 0.70$.

La validez del constructo es adecuada, el índice de Kaiser-Mayer-Olkin (KMO) $0.693 > 0.5$ y la prueba de esfericidad de Barlett $0.000 < 0.05$. En base al cumplimiento de los criterios estadísticos se realizó el análisis factorial del cuestionario de investigación para agrupar las variables relacionadas.

V. Discussion and Results

De acuerdo con los resultados obtenidos del cuestionario de investigación, se desarrolló la metodología híbrida de dirección de proyectos para el sector construcción; validada por expertos nacionales e internacionales.

Del análisis estadístico se recopiló la siguiente información: La metodología de dirección de proyectos más utilizada en el sector construcción, es la metodología tradicional, utilizada por 63.41% de los profesionales de la construcción, siendo los estándares más conocidos los del PMI (Project Management Institute).

26.6% de los profesionales de la construcción tiene algún conocimiento o ha escuchado de la metodología híbrida. De este porcentaje, únicamente 8.6% la ha aplicado a la construcción.

Los métodos ágiles más conocidos en el sector de la construcción son: Kanban, Lean y Scrum.

Los resultados del cuestionario de investigación comprobaron que existen factores externos (variables independientes) a la metodología de dirección del proyecto (variable dependiente), que influyen directamente en sus resultados.

Las variables independientes más importantes son: Equipo de trabajo capacitado, experimentado, integrado, colaborativo, motivado, con buen liderazgo.
Conocimiento y uso correcto de los programas, plataformas, herramientas y tecnologías.
Objetivos claros y buena comunicación.

Seguimiento, control y retroalimentación constante durante la ejecución del proyecto.

Se comprobó estadísticamente que la influencia de estas variables o factores es directamente proporcional a los resultados de la metodología de dirección del proyecto; entre más alta la aportación positiva de estas variables, mayores son los beneficios que se obtienen de la aplicación de la metodología de dirección del proyecto.

Al analizar estadísticamente la influencia de la metodología de dirección como (variable independiente) en el cumplimiento de las restricciones principales del proyecto (variables dependientes), se obtuvo que el tiempo de ejecución no se cumple en el 63.41% de los proyectos, seguido por el incumplimiento con el presupuesto en un 15.85% y un 10.98% de incumplimiento con la satisfacción al cliente.

Los resultados obtenidos de la información de los profesionales de la construcción comprobaron que una metodología de dirección de proyectos eficiente debe considerar: integrar todos los factores del proyecto e incrementar la agilidad de ejecución de los procesos complicados, reducir métodos rígidos y técnicas muy estructuradas que producen alto consumo de tiempo y recursos. De acuerdo con el análisis estadístico de estas variables y de acuerdo con la escala de Likert empleada, el nivel de las mejoras requeridas en la aplicación de la metodología tradicional en cuanto a los aspectos anteriores es de medio a alto.

En base a todos los resultados anteriores, se desarrolló la metodología híbrida para el sector construcción: Se enfatizó en técnicas y métodos que ayudarán a obtener resultados favorables con respecto a las restricciones de tiempo, presupuesto y satisfacción al cliente.

Se integraron los factores externos que afectan la dirección de proyectos independientemente de la metodología de dirección utilizada.

También se consideraron las características que según los profesionales de la construcción debe tener una metodología de dirección de proyectos eficiente para el sector.

a) Development of Hybrid Project Management Approach for Construction Projects

La metodología híbrida de dirección para proyectos de construcción desarrollada en esta investigación utiliza un enfoque tradicional a lo largo de todo el ciclo de vida del proyecto, combinado simultáneamente con un enfoque ágil aplicado a los grupos de procesos recomendados por la Norma ISO 21500:2012 de dirección de proyectos; como se muestra en la Fig. 1.

Fig. 1: Enfoque Híbrido Combinando Metodologías Ágiles Y Tradicionales Simultáneamente En Todo El Ciclo De Vida Del Proyecto.

Es importante recordar que cada proyecto es único por naturaleza y se debe realizar la adaptación de la metodología híbrida a utilizar considerando sus características específicas y sus restricciones: alcance, cronograma, costo, recursos, calidad y riesgos, en donde la importancia de cada una cambia con los objetivos de cada proyecto. Se debe considerar también: el entorno del proyecto, la cultura de la organización, las necesidades de los interesados y la gobernanza. En la Fig. 2 se presenta el modelo de incertidumbre y complejidad para determinar el enfoque más adecuado para el proyecto.

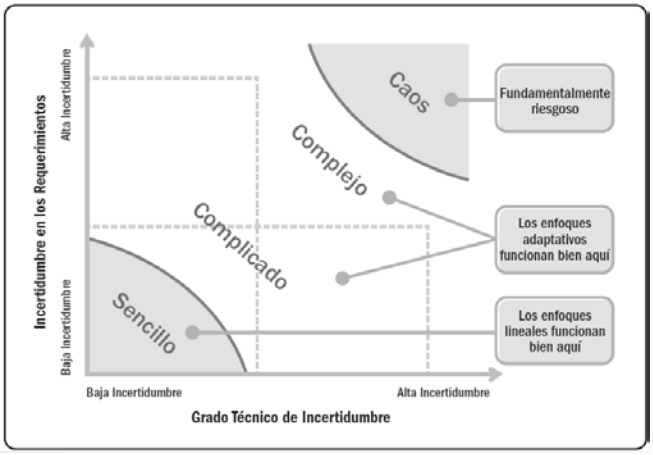


Fig. 2: Modelo De Incertidumbre Y Complejidad. Fuente: Guía Práctica De Ágil (2017).

Los métodos ágiles utilizados como complemento a la metodología tradicional, analizados en esta investigación se seleccionaron de acuerdo con los siguientes criterios: Marcos de trabajos ágiles más holísticos, orientados a un amplio conjunto de actividades del proyecto. Uso común en una variedad de contextos. Popularidad de aplicación adoptado por varias industrias. Características aplicables a la industria de la construcción y los resultados del cuestionario de investigación.

Basado en los criterios anteriores, los métodos ágiles seleccionados fueron: Scrum, Kanban, Lean Construction, Scrumban, Desarrollo impulsado por características y el Método de desarrollo de sistemas dinámicos. De acuerdo con el análisis y la validación de la metodología híbrida, los métodos ágiles compatibles con las características y requerimientos del sector construcción que complementan la metodología tradicional son: Scrumban y Lean Construction.

Lean Construction: Maximiza los resultados a la vez que minimiza los retrasos generados por demoras y desperdicios en el uso de recursos para cumplir con los objetivos. Sus obstáculos más comunes son: falta de claridad en la definición del alcance y objetivos del proyecto, excesiva dedicación de tiempo en reuniones innecesarias, mal uso de los recursos: personas, tiempo y dinero. Se basa en la eficiencia y el flujo de resultados, realización de reuniones eficientes, lograr equipos comprometidos, la administración del riesgo, planificación de reservas, aplicación de procesos estandarizados, la priorización de los proyectos y la calidad de vida del equipo.

Scrumban: Transición entre Scrum y Kanban. El trabajo se debe organizar en pequeñas iteraciones, utilizar tableros para visualizar y monitorear el trabajo, realizar reuniones para mantener la colaboración e integración del equipo y eliminar los impedimentos.

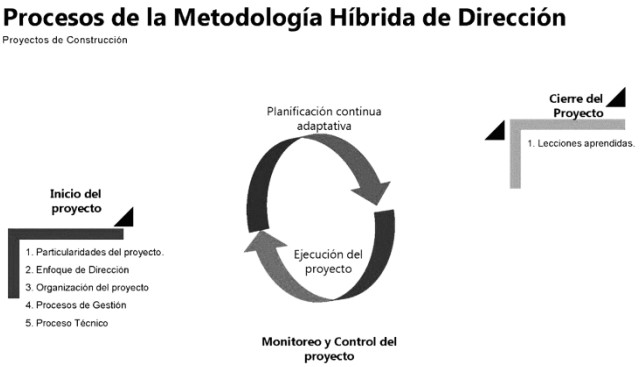
La metodología híbrida de dirección de proyectos para construcciones combina los procesos de la Norma 21500:2012 con los métodos ágiles Scrumban y Lean Construction para integrar el enfoque tradicional y ágil en todo el ciclo de vida del proyecto; buscando innovación rápida, enfoque en los resultados y entregas continuas y de valor, solución a la complejidad actual e integración y comunicación del equipo en los proyectos de construcción.

Los grupos de procesos de la metodología híbrida desarrollada son los siguientes: Inicio. Planificación iterativa-adaptativa y ejecución, ambas monitoreándose y controlándose a lo largo de todo el ciclo de vida del proyecto. Cierre. En la Fig. 3, se muestra el modelo de los procesos de la metodología híbrida para proyectos de construcción.

Los proyectos son por naturaleza integradores, la integración con relaciones iterativas es primordial en la metodología híbrida ya que las tareas incluyen a todos los grupos de procesos. En la Fig. 4, se presenta el modelo de integración de los grupos de procesos durante todo el ciclo de vida del proyecto.

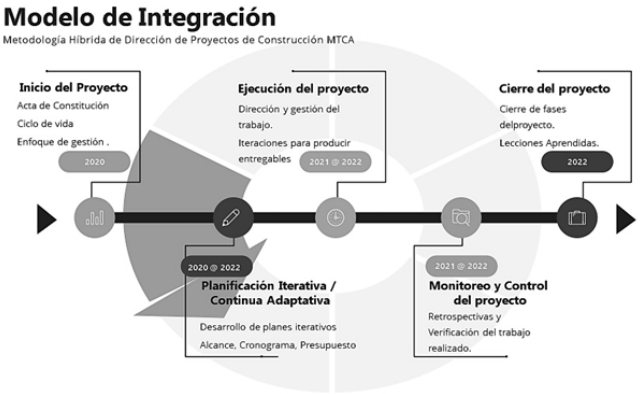
La planificación iterativa/adaptativa, es el fundamento de la metodología híbrida, se aplica mayormente en los proyectos donde el cambio es continuo e imprevisible. Se realiza con anticipación a un alto nivel y se adapta a través de retroalimentación constante, la incorporación de los cambios y

replanificación acorde al avance y a los acontecimientos reales del proyecto. En la Fig.5, se presenta el flujo de actividades que deben ser consideradas para elaborar una planificación iterativa, que se realiza de forma continua y permite la corrección del plan en caso de existir desviaciones para el logro de objetivos.



Fuente: Norma 21500:2012 Y Guía Práctica De Ágil. Elaboración Y. Frangakis.

Fig. 3: Procesos De La Metodología Híbrida De Dirección De Proyectos Para El Sector Construcción.



Fuente: Norma 21500: 2012, Guía Práctica De Ágil. Elaboración Y. Frangakis.

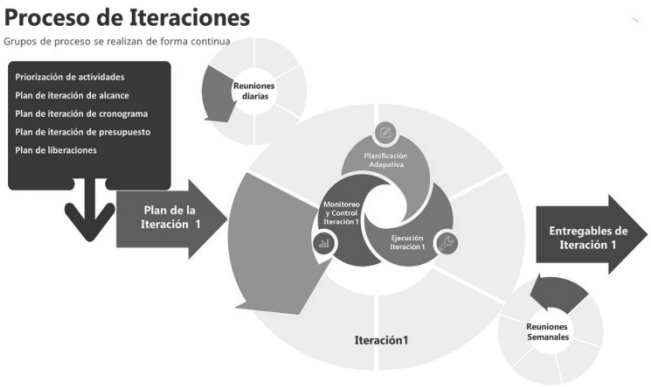
Fig. 4: Modelo Grupo De Materia De Integración.

Fuente: Norma 21500:2012, Guía Práctica De Ágil. Elaboración Y. Frangakis.

Fig 5: Flujo De Actividades Para Elaborar Una Planificación Iterativa.

Utilizar iteraciones en un proyecto de construcción ofrece muchas ventajas: permite recopilar requisitos prioritarios a corto plazo, definir más claramente el alcance, detallar la estructura de desglose del trabajo de manera más precisa, estimar el tiempo de ejecución y el presupuesto más certeramente. Es importante la retroalimentación diaria y realizar la retrospectiva o revisión de la iteración al final del tiempo establecido.

En la Fig. 6, se presenta un modelo de los procesos involucrados en una iteración para facilitar la liberación de entregables prioritarios del proyecto a tiempo. Y en la Fig. 7, se muestran las actividades que deben ser consideradas en cada iteración.



Fuente: Guía Práctica De Ágil. Elaboración Y. Frangakis

Fig. 6: Modelo Del Proceso De Una Iteración.

Fuente: Guía práctica de ágil. Elaboración Y. Frangakis.

Fig. 7: Actividades Al Realizar Una Iteración.

Si el alcance, los objetivos o los riesgos del proyecto no están completamente claros desde el inicio, al aplicar la metodología híbrida, es posible definir de forma general al principio y detallar a medida que avanza el proyecto. Con cada iteración se gana más información, claridad de los requerimientos y necesidades del cliente, se debe incluir los problemas reales que se van descubriendo durante la ejecución.

La metodología híbrida es aplicable mayormente a proyectos complejos con algún grado de incertidumbre, en donde el alcance se va conociendo mejor a medida que se ejecuta el proyecto.

Para cumplir con el tiempo de ejecución, se deben analizar las restricciones que impiden el normal desarrollo de las actividades, aplicar las iteraciones y la planificación semanal enmarcadas dentro de la programación inicial o el plan maestro del proyecto. Para fortalecer el cumplimiento de esta restricción (tiempo), que actualmente solo se cumple en 63.41% de los proyectos de construcción, la metodología híbrida define un plan y el trabajo se ejecuta en ciclos cortos para revisar resultados y adaptar según sea necesario. Las prioridades pueden cambiar y el plan necesitará ser ajustado, se requiere retroalimentación rápida sobre el trabajo ejecutado y la adecuación de los entregables mediante la programación iterativa que permite fácilmente planificar, organizar y priorizar.

En la gestión de tiempo es importante el establecimiento de objetivos reales y alcanzables, la identificación y planificación de las tareas de acuerdo con el tiempo establecido en las iteraciones y la priorización. En la Fig. 8, se muestran los procesos que deben llevarse a cabo en cada iteración para estimar el tiempo de ejecución de los entregables del proyecto.

Fuente: Guía Práctica De Ágil. Elaboración Y. Frangakis.

Fig. 8: Procesos Para Estimar El Tiempo De Ejecución En Un Proyecto Con Enfoque Híbrido.

Una herramienta que contribuye al logro del cumplimiento del tiempo es "*Last Planner*". Su aplicación es sencilla, requiere estricto cumplimiento, se basa en aumentar el desempeño de las actividades de la construcción mediante la disminución de la incertidumbre en la planificación. Como se muestra en la Fig. 9, el plan maestro cubre todas las actividades del proyecto desde su inicio hasta su terminación. El plan intermedio o de iteraciones se obtiene del plan maestro y se puede realizar para un plazo de uno a tres meses. El plan semanal se determina con base en el plan intermedio o de iteraciones, este plan contiene las actividades que se ejecutarán cada semana.

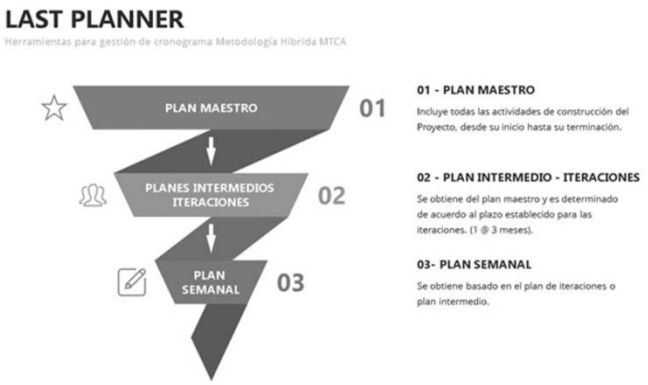


Fig. 9: Aplicación De Last Planner En La Gestión De Tiempo. Elaboración: Y. Frangakis.

La segunda restricción de los proyectos de construcción, con menor porcentaje de cumplimiento es el presupuesto. Muchas veces los proyectos con enfoque híbrido que se caracterizan por su alta variabilidad también pueden estar sujetos a presupuestos estrictos como una de sus principales restricciones; siendo así, el alcance y el cronograma se deben ajustar con frecuencia para permanecer dentro de los límites de los costos. En este tipo de proyectos las estimaciones detalladas se realizan en la planificación a corto plazo en una modalidad justo a tiempo.

b) Delphi Technique: Validation of Hybrid Project Management Methodology

La técnica Delphi fue el instrumento de validación de la presente investigación. Consistió en la utilización sistemática de un juicio intuitivo emitido por un grupo de expertos nacionales e internacionales de la dirección de proyectos quienes mediante rondas de cuestionarios anónimos realizaron evaluaciones y aportaron sus opiniones para lograr el consenso de validación de la metodología híbrida de dirección de proyectos aplicada a la construcción.

La escogencia de los expertos para la validación de la metodología híbrida desarrollada en la presente investigación se basó en: su competencia profesional, tomando en cuenta los años de experiencia, el nivel académico o científico, las publicaciones realizadas acerca de la dirección de proyectos, el nivel de conocimiento y de actualización, así como de los resultados obtenidos del análisis del coeficiente de competencia *k*.

La competencia de los expertos se midió a partir del coeficiente de competencia *k*, calculado como se muestra en la ecuación (2), en donde *kc* es el coeficiente de conocimiento y *ka* es el coeficiente de argumentación de los expertos, en donde influyeron los

análisis teóricos propios, la experiencia, los trabajos de autores nacionales e internacionales, el conocimiento del estado del problema.

$$k = \frac{1}{2}(kc + ka) \quad (2)$$

El proceso de validación mediante Técnica Delphi consistió en la elaboración de tres cuestionarios donde se ordenó, clasificó y jerarquizó cada uno de los puntos importantes del desarrollo de la metodología híbrida de dirección para proyectos de construcción, los criterios en el diseño del cuestionario fueron los siguientes: Orden lógico y claridad en las preguntas. Preguntas cortas. No se realizaron preguntas que implicaran respuestas condicionadas ni preguntas implícitas. Se realizaron las preguntas de los puntos más importantes para validar la metodología desarrollada.

El cuestionario consistió en 27 variables con preguntas cerradas para las cuales se aplicó la escala de Likert en sus respuestas con un rango de 1 al 5.

Al analizar los resultados del cuestionario, se compartió su análisis con los expertos y de acuerdo con sus aportes se reformularon las preguntas de las variables que presentaron mayor dispersión.

Los resultados estadísticos utilizados en la validación de esta investigación fueron: las medidas de tendencia central y dispersión (media, mediana, moda, máximo, mínimo y desviación típica).

En los cuestionarios se aplicó la escala de Likert en donde: No es Apropiado (1), Poco Apropiado (2), Apropiado (3), Bastante Apropiado (4), Muy Apropiado (5).

A continuación, en la Tabla 3. se presentan los resultados estadísticos descriptivos finales de las variables del cuestionario de validación de la metodología híbrida para proyectos de construcción.

Tabla 3: Resultados Estadísticos Descriptivos De Las Variables Del Cuestionario Final De Validación Mediante Técnica Delphi De La Metodología Híbrida De Dirección De Proyectos De Construcción.

Estadísticos Descriptivos			
Var.	Descripción	Media	Desv.
1	¿Al utilizar un Enfoque Híbrido de Dirección de proyectos durante todo su ciclo de vida, se mejoran los resultados en la ejecución y el cumplimiento de objetivos finales?	3.63	1.061
2	¿Utilizar una Metodología Híbrida de Dirección para proyectos de construcción donde la Metodología Tradicional es complementada con Metodologías Ágiles mejorará los resultados finales del proyecto e incrementa las entregas de valor al cliente?	4.38	0.744

Estadísticos Descriptivos			
Var.	Descripción	Media	Desv.
3	Las herramientas y técnicas que brindan los Métodos Ágiles como Scrum para proyectos de construcción variables y de alta incertidumbre; Lean (Last Planner) para proyectos de media a baja incertidumbre; combinados con la Metodología Tradicional son aplicables a los proyectos de construcción.	3.63	0.744
4	¿La Metodología Híbrida de Dirección debe aplicarse considerando las características y restricciones particulares de cada proyecto de construcción?	4.5	0.926
5	En la Metodología Híbrida de Dirección para proyectos de construcción se aplica la "planificación continua adaptativa" y los grupos de proceso se realizan de forma continua durante todo el ciclo de vida del proyecto con relaciones iterativas.	3.75	0.886
6	La Metodología Híbrida de Dirección para proyectos de construcción interrelaciona los grupos de proceso y sus resultados en cada iteración logrando una visión integral y cumplimiento de los objetivos finales.	3.5	0.756
7	¿Considera apropiados los procesos a realizar en cada iteración?	3.63	1.302

Estadísticos Descriptivos			
Var.	Descripción	Media	Desv.
8	La Metodología Híbrida de dirección para proyectos de construcción implementa realizar reuniones diarias de no más de 15 minutos con cada miembro del equipo para revisar el plan de trabajo del día y el cumplimiento del trabajo del día anterior para eliminar obstáculos y mantener la alineación del equipo con plan de la iteración.	4.13	0.835
9	La Metodología Híbrida de dirección para proyectos de construcción implementa realizar reuniones de equipo al final de cada semana para revisar y analizar datos e información del desempeño obtenido que permitan realizar mejoras en la ejecución y cumplir con el plan de iteración.	3.63	1.061
10	La Metodología Híbrida de dirección para proyectos de construcción implementa realizar reuniones retrospectivas al culminar cada iteración para evaluar el desempeño del trabajo completado y buscar posibles mejoras.	3.88	0.835
11	¿Considera apropiado el siguiente contenido del Plan de Dirección para un proyecto de construcción ejecutado mediante un enfoque híbrido de gestión?	3.5	0.535

Estadísticos Descriptivos			
Var.	Descripción	Media	Desv.
12	La Metodología Híbrida de Dirección para proyectos de construcción, implementa la retroalimentación constante y la re -planificación continua mediante reuniones de planificación de iteraciones en donde se detallan los resultados de la iteración y se planifica la siguiente iteración.	3.38	1.061
13	La Metodología Híbrida de Dirección de proyectos para construcción implementa un modelo de Plan General en que el alcance, el cronograma y el presupuesto se realizan a un alto nivel al principio y se detallan posteriormente en cada iteración.	3.75	1.035
14	La Metodología Híbrida de Dirección de proyectos implementa la priorización de actividades de acuerdo con los requisitos del cliente para realizar entregas funcionales en menor tiempo.	4	0.756
15	La Metodología Híbrida de Dirección para proyectos de construcción implementa un modelo para el desarrollo del Plan de alcance, cronograma y presupuesto por iteraciones.	3.63	1.302

Estadísticos Descriptivos			
Var.	Descripción	Media	Desv.
16	En un proyecto de construcción donde existe limitación en los fondos, la Metodología Híbrida de Dirección para proyectos de construcción permite utilizar los fondos disponibles para las actividades prioritarias que generan valor al dueño del proyecto identificadas mediante el cronograma y presupuesto de la iteración.	3.25	1.165
17	La Metodología Híbrida de Dirección para proyectos de construcción, implementa la utilización de un plan de trabajo iterativo basado en ciclos cortos y planificación ágil de liberaciones para obtener entregables de áreas prioritarias y resultados de valor.	3.5	1.414
18	La Metodología Híbrida de Dirección para proyectos de construcción implementa evaluar la calidad del trabajo realizado en las iteraciones y encontrar la causa raíz de los problemas si existen y buscar soluciones para su mejora.	4.25	0.707
19	La Metodología Híbrida de Dirección para proyectos de construcción implementa la participación del equipo del proyecto en la planificación y en la entrega de las actividades completadas.	4.25	0.707

Estadísticos Descriptivos			
Var.	Descripción	Media	Desv.
20	La Metodología Híbrida de Dirección para proyectos de construcción promueve la comunicación constante con el equipo para resolver posibles obstáculos, garantizar claridad y compresión del trabajo a realizar en la iteración.	4.25	0.886
21	La Metodología Híbrida de Dirección para proyectos de construcción implementa la identificación de riesgos en cada iteración.	4.5	0.756
22	En proyectos de construcción con incertidumbre media a alta la Metodología Híbrida de Dirección para proyectos de construcción implementa la utilización de contrataciones que permitan mayor flexibilidad a los cambios; se debe considerar la actividad a contratar y la particularidad de cada situación.	4.38	0.916
23	La Metodología Híbrida de Dirección para proyectos de construcción promueve la participación de los interesados para facilitar toma de decisiones.	4	0.756
24	La Metodología Híbrida de Dirección para proyectos de construcción implementa la revisión de métricas semanales para comparar la cantidad de trabajo ejecutado y la cantidad de trabajo estimado para determinar el estado del cronograma y realizar las correcciones pertinentes.	3.38	0.916

Estadísticos Descriptivos			
Var.	Descripción	Media	Desv.
25	La Metodología Híbrida de Dirección para proyectos de construcción implementa la utilización de Tablero de tareas, reuniones diarias, retrospectivas junto con otras herramientas y técnicas de control usualmente aplicadas en construcción y seleccionadas de acuerdo con las particularidades del proyecto.	4.13	0.835
26	La Metodología Híbrida de Dirección para proyectos de construcción monitorea semanalmente el cumplimiento de las actividades de la iteración mediante un indicador del porcentaje de actividades cumplidas.	4.38	1.061
27	La Metodología Hibrida de Dirección para proyectos de construcción implementa el uso de herramientas tecnológicas de integración, dashboard para análisis y visualización de datos para control eficiente y en tiempo real de la ejecución del proyecto	4.25	0.707

De acuerdo con los resultados estadísticos obtenidos de la validación de la presente investigación mediante la Técnica Delphi; el valor obtenido fue de "3.903", que indica que la aplicación de la metodología híbrida de dirección de proyectos en el sector de la construcción es "bastante apropiado".

En la Tabla 4 se muestra el resumen del análisis estadístico obtenido de las variables del cuestionario.

Tabla 4. Resultados Estadísticos Finales Del Cuestionario De Validación

	Media	Mínimo	Máximo	Varianza	N De Elementos
Medias de elemento	3.903	3.25	4.5	0.149	27
Varianzas de elemento	0.878	0.286	2	0.177	27
Covarianzas entre elementos	0.321	-0.607	1.5	0.111	27
Correlaciones entre elementos	0.392	-0.63	1	0.126	27

El valor del coeficiente Alfa de Cronbach obtenido del cuestionario final de validación de la Metodología Híbrida de Dirección para Proyectos de Construcción fue de 0.946, confirmando que los resultados de la validación son fidedignos.

Los aspectos que permitieron valorar positivamente la fiabilidad de los resultados son los siguientes:

- ✓ Estabilidad y calidad de los expertos que participaron en la validación, el coeficiente de competencia *k* de los expertos fue alto.
- ✓ Tiempo transcurrido entre rondas fue de aproximadamente un mes, el tiempo total empleado en la validación mediante la Técnica Delphi fue de aproximadamente seis meses.
- ✓ El consenso entre las opiniones de los expertos.
- ✓ Estabilidad de los resultados entre rondas.
- ✓ En la Tabla 5 se presentan los valores del Alfa de Cronbach que indica la fiabilidad del cuestionario final de validación mediante Técnica Delphi.

Tabla 5: Valor Obtenido Del Alfa De Cronbach

Estadísticas de fiabilidad		
Alfa de Cronbach	Alfa de Cronbach basada en elementos estandarizados	N° de elementos
.940	.946	27

c) Limitations and Implications

La principal limitación de la aplicación de la metodología de dirección de proyectos híbrida en construcción es la actitud de resistencia al cambio que usualmente se encuentra en los profesionales al tratar de realizar mejoras a los procesos que están acostumbrados a utilizar, aunque no obtengan los resultados esperados.

En los proyectos de construcción el equipo del proyecto es usualmente temporal, por lo que se debe considerar el tiempo de capacitación y adaptación para que se logre aplicar la metodología híbrida de dirección de proyectos de construcción y realmente se logren los beneficios de su aplicación.

La resistencia de las empresas constructoras a la utilización de una metodología híbrida estandarizada para dirección de proyectos de construcción, por asumir que es un costo adicional.

La metodología híbrida para dirección de proyectos de construcción ha sido validada por expertos del tema, como trabajo futuro se propone su implementación en los proyectos de construcción de una organización que quiera mejorar los resultados que obtiene durante la ejecución de sus proyectos.

Los resultados de la presente investigación confirmaron que la metodología híbrida de dirección de proyectos puede ser aplicada en la industria de la construcción aportando grandes beneficios durante el desarrollo del proyecto. A la vez son los cimientos para nuevas preguntas, nuevas ideas y abre nuevas líneas de trabajo.

Las nuevas líneas de investigación podrían basarse específicamente en uno de los Métodos ágiles que tienen mayor afinidad con las características de los proyectos del sector construcción y desarrollar una metodología híbrida fundamentada en la metodología

de dirección de proyectos tradicional complementada con un método ágil especifico, ejemplo Lean Construction o Scrumban.

Otra línea de investigación futura, podría ser el desarrollo de una Guía Práctica de construcción basada en la metodología híbrida enfocada en uno de los grupos de procesos específicos del ciclo de vida de un proyecto de construcción; ejemplo planificación, tiempo, monitoreo y control.

VI. Conclusions

Por medio del Método Delphi se comprobaron las hipótesis de la investigación. Se validó que estandarizar los procesos de dirección de proyectos mediante la aplicación de una metodología híbrida de dirección de proyectos (tradicional complementada con métodos ágiles) que cuente con herramientas, técnicas y plantillas ayuda al logro de los objetivos del proyecto y de la empresa.

El aporte principal de la presente investigación es contribuir con el desarrollo y crecimiento de la dirección de proyectos mediante una metodología integradora de los beneficios de las dos grandes metodologías de dirección de proyectos existentes (Tradicional-Ágil) que permita utilizar recursos eficientemente, mantenerse flexibles ante los cambios y competitivas en el ambiente de negocios actual.

No existe una metodología única para dirigir los proyectos de construcción, la metodología a utilizar dependerá de sus características particulares; entre las que destacan la metodología tradicional y la metodología híbrida (Metodología tradicional combinada con herramientas o métodos ágiles).

Por las características especiales de los proyectos de construcción, no todos los métodos ágiles son aplicables. La utilización de herramientas y técnicas ágiles en un proyecto de construcción depende de su variabilidad, nivel de incertidumbre, características técnicas y del equipo.

Al aplicar la metodología híbrida en cada uno de los grupos de procesos de un proyecto de construcción, seleccionando los métodos ágiles de acuerdo con sus particularidades; se agrega valor al proyecto en cada uno de sus grupos de procesos y por consiguiente en los objetivos del proyecto.

Utilizar la herramienta de planificación continua adaptativa e iteraciones en proyectos de construcción, ayuda a mantener el plan de ejecución de acuerdo con las situaciones reales que van aconteciendo durante la ejecución del proyecto y por consiguiente dan claridad en la finalización real del proyecto.

Un proyecto de construcción también puede incrementar su valor realizando entregas parciales de acuerdo con las prioridades y requerimientos de los interesados. De esta forma se acelera también la aceptación del proyecto.

Una visión integral es esencial para el cumplimiento de los objetivos finales de un proyecto de construcción, esto puede lograrse al aplicar la Metodología Híbrida de Dirección interrelacionando los grupos de proceso y sus resultados en cada iteración. Obteniendo información de calidad y veraz a tiempo permitiendo tomar las correcciones que se requieran para lograr resultados exitosos.

En un proyecto de construcción es importante obtener retroalimentación constante, esto puede lograrse mediante la Metodología Híbrida al implementar reuniones retrospectivas al culminar cada iteración y evaluar el desempeño del trabajo completado con la finalidad de buscar mejoras.

La comunicación y participación del equipo durante todo el ciclo del proyecto es fundamental para mantener la retroalimentación constante, detección oportuna de problemas y de esta forma darles solución temprana, replanificar y no afectar los objetivos de tiempo finales.

En un proyecto de construcción es relevante considerar la capacitación, la motivación, la cohesión e integridad del equipo del proyecto ya que es un factor clave para el éxito o fracaso del proyecto.

Es muy importante el monitoreo semanal del cumplimiento de las actividades de la iteración mediante un indicador del porcentaje de las actividades cumplidas, ya que lo que se mide es lo único que se puede mejorar.

La implementación de herramientas tecnológicas de integración, tableros para análisis y visualización de datos permiten el control eficiente y monitoreo en tiempo real de la ejecución del proyecto, importante que el equipo tenga las calificaciones técnicas apropiadas.

Acknowledgements

A todos los miembros del Colegio de Ingenieros Civiles de Panamá y del Project Management Institute capítulo de Panamá quienes con su participación en el cuestionario de investigación; hicieron posible conocer los problemas reales a los que se enfrentan dirigiendo los proyectos y de esta forma permitir el desarrollo de la presente investigación.

A los expertos en dirección de proyectos que contribuyeron con su tiempo, sus valiosos comentarios y aportes para validar la metodología híbrida de dirección de proyectos de construcción desarrollada en esta investigación.

References Références Referencias

1. Arias C. (2007). Estudio de Mercado - Sector de la Construcción en Panamá. Recuperado el 15 de marzo de 2017, de http://servicios. Pro-comer.go.cr.

2. Blasco, Jaume (2000). Los artefactos y susproyectos. Revista POLITEXT Área de Ingeniería Mecánica. Barcelona-España: Editorial UPC.
3. Blasco, Jaume. (2001). Los proyectos, el proyectar y el proyectado. Revista POLITEXT Área de Ingeniería Mecánica. Barcelona-España: Editorial UPC.
4. Bonnie E. (2015). La Gestión de Proyectos en 2015. Recuperado el 20 de noviembre de 2016, de https://www.wrike.com/es/blog/la-gestion-de-proyectos-en-2015-infografia.
5. Bredillet C. (2010). PMI Research and Education Conference 2010. Project Management Journal. *DOI:* 10.1002.
6. Carrizo A. (2017). Metodologías ágiles, ¿son compatibles con Gestión de la demanda de TI? Recuperado el 19 de abril de 2017, de https://www.xeridia.com/blog/metodologias-agiles-compatibles-con-gestion-demanda-TI.
7. Centro Nacional de Competitividad. (2016). Evolución y Competitividad en el sector de la construcción en Panamá. Recuperado el 15 de marzo de 2017, dehttp://cncpanama.org.
8. Chiavenato, I. (2007). Introducción a la Teoría General de la Administración. Séptima edición. Méjico, DF: editorial Mc Graw Hill. ISBN: 13- 978-970-10-5500-7.
9. Collado C., Lucio P. y Sampieri R. (2003). Metodología de la investigación. Tercera Edición. Méjico, DF: McGraw-Hill Inter-americana. ISBN: 9789701036327.
10. Comité de Proyecto ISO/PC 236. (2012). Norma Internacional ISO 21500:2012: Dirección y Gestión de Proyectos. Primera Edición. Suiza.
11. Dahlbom, Bo; y, Mathiassen, Lars. (1995). Computers in Context. The Philosophical and Practice of Systems Design. Cambridge: Editorial Blackwell.ISBN: 0-89791-926-2.
12. De Cos M. (1997). Teoría General del Proyecto. Volumen II. Vallehermoso, Madrid: Síntesis. ISBN: 978-84-773845-2-6.
13. Duncan H. (s.f). Breve historia sobre la administración de proyectos. Recuperado el 12 de septiembre de 2015, de http://www.liderdeproyecto.com/manual/breve_historia_sobre_la_administracion_de_proyectos.html.
14. Estay-Niculcar, C. (2007). Rigor y relevancia, perspectivas filosóficas y gestión de proyectos de Investigación- Acción en Sistemas de Información. Departamento de humanidades universidad internacional de Catalunya. Barcelona, España. Tesis Doctoral. ISBN: 978-8 46-91-3851.
15. Friedman A. y Miles S. (2006). Stakeholders: Theory and Practice. Oxford, New York: Oxford University Press. ISBN: 978-019-92-6987-7.
16. Giddens, A. 1986. The Constitution of Society. Berkeley: University of California Press. ISBN: 978-052-00-5728-9.
17. Goméz E., Senentt M., Chiner M., Capuz S., Aragones P., y Santamaría, J.L. (1996 a). Teoría de las Dimensiones del Proyecto. III Congreso International de Ingeniería de Proyectos. Barcelona, Terrasa. Departamento de Proyectos de Ingeniería. Universidad Politéc-nica de Catalunya.
18. Goméz E., Senentt M., Chiner M., Capuz S., Aragones P., y Santamaría, J. L. (1996b). ¿Es el proyecto un Sistema?IIICongreso International de Ingeniería de Proyectos. Barcelona, Terrasa. Departamento de Proyectos de Ingeniería. Universidad Politécnica de Catalunya.
19. Gómez E. y SenenttM. (2001). El Proyecto Diseño en Ingeniería. Méjico, DF: Alfaomega-Universidad Politécnica de Valencia. ISBN: 970-15-0622-7.
20. Jiménez E. y Orantes S. (2012). Metodología Híbrida para Desarrollo de Software en Méjico. Tecnología de Software y Bases de Datos, Centro de Investigación en Computación (CIC), IPN. Ciudad de México, C.P. 07738, México. [Versión Electrónica]. Recuperada el 15 de abril de 2017 dewww.revista.unam. mx/vol.13/ num1 /art16.
21. Kerzner, H. (2000). Applied project management best practices on implementation. New York-USA: WILEY. ISBN: 0-471-36352-9.
22. Klee J. (2016, junio 16). Gestión de los Interesados del Proyecto: Planeación y comunicación. Conferencia impartida por el Project Management Institute. Capítulo de Panamá.
23. Langley M. (2016). PMI's Pulse of the profession. VIII Encuesta mundial sobre dirección de proyectos: El alto costo de un bajo desempeño. ¿Cómo mejorara los resultados de los negocios? [Versión electrónica]. Revista de PMI. Recuperado el 18 de septiembre de 2016, de http://www.pmi.org/learning/thought-leader-ship/pulse/pulse-of-the-profession-2016.
24. Lavagnon, I. (2009). Project Success as a Topic in Project Management Journals. Project Management Journal, Volumen 40. Número 4.
25. Marsh, E. (1975). The Harmonogramof Karol Adamiecki. The Academy of Management Journal, 18 (2), 358–364.DOI: 10.2307/255537.
26. Mertens, D. (2015). Research methods in education and psychology: Integrating diversity with quantitative and qualitative approaches. Cuarta Edición. Thousand Oaks, California: Sage Publications. ISBN: 978-1452240275.
27. Ministerio de Economía y Finanzas. (2015). Informe económico y social - 2015. Dirección de Análisis económico y social.
28. Montero G. (2015). La historia de la gestión de proyectos. Recuperado el 10 de septiembre de

2015, de http://www.ideassencillas.com/2012/ 05/la-historia-de-la-gestion-de-proyectos.html
29. Morres R. (2016). La Evolución Ágil. Revista. PM Network. Diciembre 2016. Volumen 30. Número 12.
30. Mulcay R. y Diethelm L. (2011). Preparación para el examen del PMP. RMC Publications Inc. ISBN: 978-1-932735-49-9.
31. Oxford Economics Ltd (2021). Future of Construction. A Global Forecast for construction to 2030.
32. Palacio J. y Ruata C. (2011). Scrum Máster Gestión de Proyectos. Revisión 1.4.0. Universidad de Chile. Recuperado el 20 de junio de 2016, de http://www.scrummanager.net/files/sm_proyecto.ISBN 1012268137397.
33. Pantoja A. (Coord.) (2015). Manual básico para la elaboración de tesinas, tesis, y trabajos de investigación. Madrid: EOS.ISBN:978-84-9727-611-5.
34. Parker D. y Stacey R. (1996). Chaos, Management and Economics. Great Britain: London. Institute of economic affairs. ISBN: 13: 978-0-255-36333-4.
35. Parsi N. (2017). PMI's Pulse of the Profession 2017. Revista. PM Network. Marzo 2017. Volumen 31. Número 3.
36. Project Management Institute. (2013). Guía de los Fundamentos para la Dirección de Proyectos (5taed.), PMBOK GUIDE. Quinta Edición. Newtown Square, Pennsylvania: PMI Publications. ISBN: 978-1-62825-009-1.
37. Sáez A. (2012). El éxito de la gestión de proyectos. Un nuevo enfoque entre lo tradicional y lo dinámico. ESADE. Tesis Doctoral.
38. Santamaría J. L., Gómez E., Senentt M., y Chiner M. (1996). Tendencias y enunciados para una teoría del proyecto. III Congreso International de Ingeniería de Proyectos. Barcelona, Terrasa. Departamento de Proyectos de Ingeniería. Universidad Politécnica de Catalunya.
39. Shenhar, A. y Dvir, D. (2007). Reinventing Project Management. Boston, Massachusetts: Harvard Business School Press. ISBN: 13: 978-1-59139-800-4.
40. Singh H. y Singh A. (2002). Principles of Complexity and Chaos Theory in Project Execution: A New Approach to Management. Revista de Association for the Advancement of Cost Engineering. Volumen 44. Número 12.
41. Spiegel M. y Stephens L. (2009). Estadística. Cuarta Edición. Méjico, D.F: Mc Graw- Hill. ISBN: 970-10-6887-8.
42. Tashakkori A. y Teddlie C. (2003). Handbook of mixed methods in social and behavioral research. Thousand Oaks, California: Sage Publications. ISBN: 0-7619-2073-0.
43. Terry G. y Franklin L. (1988). Principios de la Administración, Méjico, DF: editorial Continental. ISBN: 968-26-0586-5
44. Teixido J. (2016). La Evolución Ágil. Revista. PM Network. Diciembre 2016. Volumen 30. Número 12.
45. Wysocki R. (2011). Effective project manage-ment traditional, agile, extreme.Sexta Edición. Editorial: Wiley. ISBN: 978-1-118-01619-0.

This page is intentionally left blank

Global Journal of Computer Science and Technology: H Information & Technology
Volume 23 Issue 2 Version 1.0 Year 2023
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Using Artificial Intelligence to Evaluate the Effective Efficiency of Internal Combustion Engines

By Federico Ricci, Luca Petrucci & Francesco Mariani

University of Perugia

Abstract- Internal Combustion Engines play a crucial role in various sectors of modern transportation and industry. However, the ever-stringent regulations ask for enhanced efficiency and reduced pollutant emissions. To address these challenges, the adoption of advanced technologies, such as Machine Learning (ML) approaches, has emerged as a promising solution.

In this study, the authors extend previous research by simultaneously predicting the torque and fuel consumption of a three-cylinder spark ignition engine to determine overall engine efficiency. Specifically, they compare the performance of a NARX algorithm with the traditional Backpropagation (BP) structure. They conducted a preliminary sensitivity analysis to optimize prediction accuracy, eliminating less influential parameters from the dataset, and significantly reducing computational effort and operational time.

GJCST-H Classification: LCC Code: TH1-9745

Strictly as per the compliance and regulations of:

Using Artificial Intelligence to Evaluate the Effective Efficiency of Internal Combustion Engines

Federico Ricci [α], Luca Petrucci [σ] & Francesco Mariani [ρ]

Abstract- Internal Combustion Engines play a crucial role in various sectors of modern transportation and industry. However, the ever-stringent regulations ask for enhanced efficiency and reduced pollutant emissions. To address these challenges, the adoption of advanced technologies, such as Machine Learning (ML) approaches, has emerged as a promising solution.

In this study, the authors extend previous research by simultaneously predicting the torque and fuel consumption of a three-cylinder spark ignition engine to determine overall engine efficiency. Specifically, they compare the performance of a NARX algorithm with the traditional Backpropagation (BP) structure. They conducted a preliminary sensitivity analysis to optimize prediction accuracy, eliminating less influential parameters from the dataset, and significantly reducing computational effort and operational time.

The experimental dataset used for training and testing the algorithms was obtained from transient cycle experiments. The grid search method performed hyperparametric optimizations for both BP and NARX architectures. The results demonstrated that the NARX approach consistently outperformed the BP model. The BP model achieved an average error of under6%, while NARX yielded average errors up to3.8%. NARX effectively captured fast signal oscillations and accurately followed torque trends. Overall, the NARX model showed superior performance in accurately predicting torque behavior and fuel consumption compared to the traditional BP architecture. The application of ML algorithms like NARX has the potential to significantly enhance the efficiency of ICEs, reduce fuel consumption, and mitigate pollutant emissions. This research provides valuable insights into the effective use of ML approaches for optimizing engine design and performance in the pursuit of a sustainable future for the transportation and industrial sectors. The study enables the estimation of fundamental parameters such as instantaneous torque and fuel consumption, that are not provided on any engine control unit of real vehicles.

I. Introduction

Internal combustion Engines (ICE) play a crucial rolein modern transportation and industry, being employed in diverse sectors like planes, cars, and power generation [1-3]. The ever-strict regulations on pollutant emissions of harmful gases (NOx, CO, HC and soot) and $CO_2$are forcing the ICE to become more efficient to reduce their impact on the environment [4, 5]. Therefore, the research field is called to make significant efforts to enhance the overall efficiency of ICEs [6, 7]. In the industry, achieving a 0.05% improvement in efficiency is a challenging and exceptional endeavor, directly impacting both fuel consumption and emissions. The provided Table exemplifies this; for instance, NOx emissions decrease from 0.08 to 0.06 g/km. Even minor differences between Euro standards require substantial efforts, motivating advanced research.

Table 1: European Emission Standards for Gasoline Engines [2]

	Starting from	CO_2 (g/km)	HC (g/km)	NOx (g/km)	HC + NOx (g/km)	PM (g/km)
Euro 1	12/1992	2.72	/	/	0.97	/
Euro 2	01/1997	2.30	/	/	0.5	/
Euro 3	01/2000	2.20	0.20	0.15	/	/
Euro 4	01/2005	1.00	0.10	0.08	/	/
Euro 5	09/2009	1.00	0.10	0.06	/	0.005
Euro 6	08/2014	1.00	0.10	0.06	/	0.005

Original Equipment Manufacturers (OEMs) and the engine research community agree in considering an engine thermal efficiency, of the Spark - Ignition (SI) engines, above 40% as a feasible goal [8, 9]. Advanced combustion techniques such as cooled external exhaust gas recirculation (EGR) [10], engine boosting in conjunction with downsizing [11], water injection [12], and lean mixture operations [13-15] proved to be effective ways to meet these demands. EGR reduces combustion temperatures, mitigating NOx emissions

Author α σ ρ: Engineering Department, University of Perugia, Via Goffredo Duranti, 93, 06125 Perugia, Italy.
e-mail: francesco.mariani@unipg.it (F.M.)

while improving efficiency by reducing pumping losses and improving combustion stability. Combining boosting with downsizing results in smaller engines that maintain or even enhance power output. This reduces fuel consumption and emissions due to the smaller engine size. Water injection improves efficiency by cooling the intake charge, reducing knock, and allowing for higher compression ratios. It also reduces emissions by lowering combustion temperatures. Operating with a lean air-to-fuel mixture reduces fuel consumption and CO_2 emissions. It can, however, increase NOx emissions, which is where technologies like EGR and water injection can offset these drawbacks. However, the implementation of new technologies increases the engine complexity and magnifies the amount of data to be collected from the different physical sensors both during engine calibration and run-time operations [16]. As a result, significant computational efforts are required, resulting in longer operating times and increased costs [17]. Advanced technologies, such as the Machine Learning (ML) approach [18], are currently being investigated to effectively monitor the vehicle's parameters of spark ignition (SI) engine with the aim of overcoming the abovementioned issues. The optimization of the engine design passes through effective efficiency, whose knowledge can be of pivotal importance to improve the engine's performance while reducing fuel consumption and pollutant emissions, at the same time.

Incorporating insights from advanced technologies like Machine Learning, the optimization of engine design can be further refined, leveraging data-driven approaches to enhance combustion strategies and ultimately achieve greater efficiency and emission reduction. Efficient performance can be assessed by measuring the engine's delivered torque and fuel consumption. However, implementing these measurements directly on-board is challenging using physical instruments. Preliminary characterizations on specific test benches are required to calibrate the engine properly [19]. In the present study, an unconventional methodology, primarily leveraging a NARX ML model, is being introduced to forecast engine parameters that would typically pose significant challenges for experimental measurement. This model is intended to function as a virtual sensor for quantities not typically available in today's commercial vehicles. Extending this concept, this approach might potentially substitute the measurement of certain parameters currently ascertained using conventional sensors and actuators, thereby resulting in cost reduction.

In a study conducted by Togun and Baysec [20], they used an artificial neural network (ANN) to predict the specific fuel consumption (SFC) and torque of a gasoline engine. The input parameters considered were ignition advance, throttling status, and engine speed, while the outputs were predicted separately using different network architectures. To predict engine torque, a network architecture consisting of 3 input parameters, one hidden layer with 13 neurons, and one output layer was used. For predicting brake-specific fuel consumption (BSFC), a 3-15-1 network architecture was used. The results for torque prediction showed correlation coefficients of approximately 0.99 for both training and testing. For BSFC prediction, the correlation coefficients were 0.9971 for training and 0.98331 for testing. The mean absolute percentage error (MAPE) for torque measurement was 0.2912 for training and 1.74 for testing. As for BSFC prediction, the MAPE values were 1.0186 for training and 2.7588 for testing. Cay [21] developed three separate artificial neural network (ANN) models to predict brake-specific fuel consumption (BSFC), exhaust gas temperature (EGT), and effective power. They wanted to study the impact of three input parameters on these output parameters. The inputs to the ANN included fuel flow, speed, inlet manifold temperature, torque, and water temperature. During testing, the mean error percentage was found to be less than 2.7%, and root mean square error (RMSE) values were less than 0.02, indicating the accuracy of the predictions. For both training and testing data, the R^2 value, which represents the goodness of fit, was close to 0.99, further demonstrating the effectiveness of the ANN models. Golcu et al. [22] investigated how variable valve timing affects the performance and fuel economy of an engine. They experimented by changing the crank angle by 10 to 30 degrees and created an artificial neural network (ANN) model using valve-timing and speed as inputs. The output data chosen were fuel flow and torque. The testing results for torque and fuel flow showed root mean square errors (RMSE) of approximately 0.9% and 0.28%, respectively. All the mentioned references used the Back Propagation (BA) approach for the forecasting activities.

A previous work of the same research group [23], showed the performance of a NARX (nonlinear autoregressive network with exogenous inputs) structure in reproducing with remarkable accuracy across various transient cycles the torque delivered by a SI engine. The optimized NARX architecture exhibited an average error up to 70% less than a critical threshold of 10%.

Within this context, the present work aims to extend the results of the previous analysis by simultaneously computing the torque and the fuel consumption of a three-cylinder SI engine in order to define a methodology to determine the engine's overall efficiency through artificial intelligence (AI). The performance of a NARX algorithm was compared with the ones of a BP structure. A preliminary sensitivity analysis was performed to identify and eliminate the less influential parameters for the prediction of the outputs, i.e. torque and fuel consumption, from the initial dataset. As a result, the amount of data to be processed was reduced, significantly decreasing computational effort

and operational working time. Hyper parametric optimizations of the tested architectures, i.e. BP and NARX, were realized through the grid search method based on a transient cycle experimentally realized. The results show that BP reproduces physical trends well with an average error of 5.9% for torque and 4.7% for fuel consumption, both below a critical threshold of 10%. NARX outperforms BP by 35% in both cases, with a lower average error under 4% for torque and fuel consumption.

II. Materials and Methods

a) Experimental Setup

Tests were performed on a 1L three-cylinder turbocharged engine of 62 kW, chosen to represent a common and efficient configuration in modern automobiles and in all test benches, enabling a realistic and relevant assessment of engine parameters for widespread applicability and industry relevance. The engine operates in Port Fuel Injection (PFI) with European market gasoline, and it is controlled by an electric motor both in motored and firing conditions. More detailed information about the engine and experimental setup can be found in [23]. The torque delivered by the engine is measured using a torque meter located near the engine crankshaft whereas the fuel consumption is computed by a dynamic AVL 733Sfuel meter. An AdaMo Hyper software records all parameters during engine operations, enabling the simultaneous control of engine speed, torque, and valve throttle position. AdaMo acquired data from various sources, including the Engine Control Unit (ECU), pressure sensors, thermocouples, torque meter, and dynamic fuel meter, at a sampling frequency of 10 Hz. From this data, a total of 15 variables, that were deemed most characteristic, were selected as inputs (red in Figure 1) for the neural structure to predict torque and fuel consumption (blue in Figure 1). The signals coming from thermocouples TCK and pressure sensors PTX 1000 are acquired by data acquisition systems of National Instrument. All the above quantities are recorded by the AdaMo Hyper software during engine operations. Table 2 displays some values of the variables acquired during the experimental activities.

Parameter	Unit	Handled by
F_TorquemeterAI	Nm	TorqueMeter
AVL733SGetMeas	kg/h	Dynamic Fuel Meter 733S
EngineSpeed	rpm	ECU
F_Throttle Position	%	ECU
InjectionTime	us	ECU
Fuel_Pressure	bar	ECU
SparkAdvance	CAD bTDC	ECU
TC_H2O	°C	ECU
F_Engine Oil T	°C	PT100 sensor
F_TC_Turbine IN	°C	Thermocouple TCK
F_TC_Turbine OUT	°C	Thermocouple TCK
F_TC_Air_Intake	°C	Thermocouple TCK
FC_TC_Intercooler_OUT	°C	Thermocouple TCK
FC_TC_Compressor_OUT	°C	Thermocouple TCK
F_P_Compressor_OUT	mbar	PTX1000 sensor
F_P_Turbine_Out	°C	Thermocouple TCK
Lambda	-	Lamba Sensor

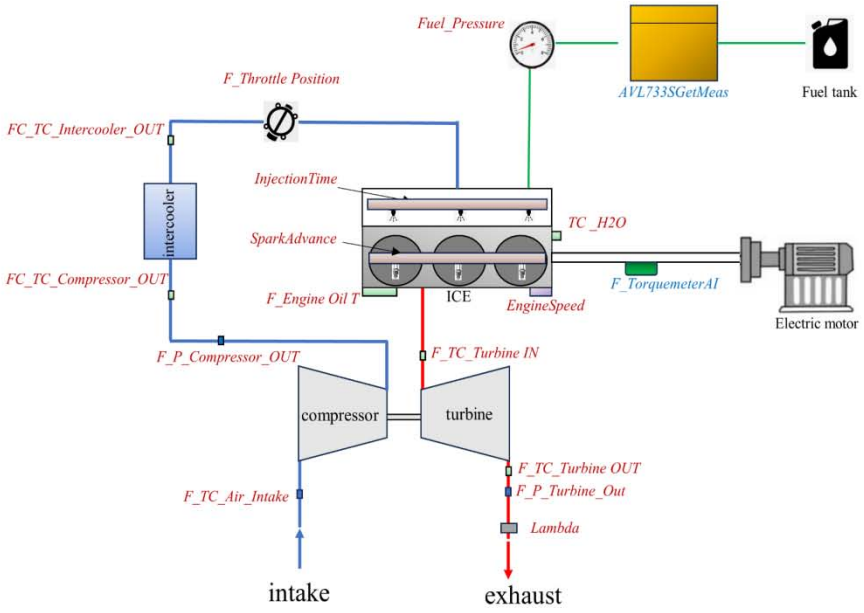


Figure 1: Schematic Representation of the Engine Setup and Corresponding Features Considered to Predict Torque and Fuel Consumption

Table 2: Some Values of the Input and Output Variables Acquired During the Experimental Activities

Timestamp	s	0.1	0.2	.....	200	200.1	.....	349.9	350	.....	525.3	525.4
F_TorquemeterAI	Nm	14.9	14.9	.....	4.8	5.3	.....	7.2	6.3	.....	6.8	6.3
AVL733SGetMeas	kg/h	1.908	1.827	.....	0.587	0.57	.....	0.945	0.945	.....	0.25	0.511
EngineSpeed	rpm	1201	1201	.....	1369	1361	.....	2311	2348	.....	2403	2439
F_Throttle Position	%	8.2	8.2	.....	6.6	6.2	.....	10.9	11.7	.....	10.2	10.2
F_Engine Oil T	°C	57.2	57.2	.....	60.1	60.7	.....	63.2	63.2	.....	65.8	65.8
F_TC_Turbine IN	°C	151.9	179.2	.....	476.3	476.3	.....	559.6	559.6	.....	567.9	567.9
F_TC_Turbine OUT	°C	112.5	112.5	.....	295	295	.....	387.6	388.1	.....	402.9	402.9
F_TC_Air_Intake	°C	34.2	34.2	.....	35.9	35.9	.....	36.2	36.2	.....	35.9	35.9
FC_TC_Intercooler_OUT	°C	33.8	33.8	.....	34.9	34.9	.....	35.9	35.9	.....	35.6	35.6
F_TC_Compressor_OUT	°C	12.01	12.01	.....	10.13	10.13	.....	11.12	11.12	.....	14.09	14.09

F_P_Compressor_OUT	mbar	987	987	.....	985.1	985.1	.....	986.1	986.1	.....	989.1	989.1
F_P_Turbine_Out	mbar	982.29	982.29	.....	982.83	982.83	.....	983.22	983.22	.....	981.69	981.69
Fuel_Pressure	bar	3215	3215	.....	3215	3215	.....	3204	3204	.....	3215	3215
InjectionTime	us	3404	3404	.....	2361	2440	.....	2361	2361	.....	2167	2010
SparkAdvance	CAD	24.5	24.5	.....	26.75	27	.....	32.5	32.5	.....	32.75	32.75
TC_H2O	°C	49.4	49.4	.....	46.9	46.9	.....	49.4	49.4	.....	50.6	50.6
Lambda	-	1.008	1.008	.....	0.969	0.938	.....	0.945	0.945	.....	0.883	0.898

b) Data Correlation Analysis

Starting from the initial dataset, it is important to understand the relationship between the numerous input parameters recorded during the experimental activities and their impact on the engine-delivered torque and fuel consumption.

It is essential to evaluate each parameter individually to understand the influence of such parameters on the output prediction to identify if they can be removed in order to reduce the model's complexity by improving its accuracy at the same time.

It is worth mentioning that the variables for which you want to assess the correlation should be sortable and, ideally, continuous. To this aim, three different correlation parameters have been employed: Pearson correlation coefficient R [24], Spearman correlation coefficient ρ [25] and Shapley value φ [26].

The Pearson correlation coefficient *R* evaluates the degree of linear correlation between parameters by assessing the direction and strength of their relationship based on their actual values, as follows (eq.1):

$$R = \frac{\mathrm{CoV}\ (X,Y)}{\sigma_x \sigma_y} = \frac{n\ \sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{\sqrt{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2}\sqrt{n\sum_{i=1}^{n} y_i^2 - (\sum_{i=1}^{n} y_i)^2}} \quad (1)$$

where n is the number of input parameters x_i, y_i the variable to be predicted, CoV(X,Y) is the covariance of two sets of data X and Y, and σ_x and σ_y the standard deviations of such sets. The Pearson correlation coefficient also ranges from -1 to +1. A positive value indicates a positive linear relationship, while a negative value indicates a negative linear relationship. A coefficient close to 0 means there is little to no linear relationship between the variables.

The *Spearman correlation* coefficient ρ is a statistical method used to determine the strength and direction of the relationship between two variables. It is a non-parametric measure, meaning it doesn't assume a specific distribution for the data. Instead, it focuses on assessing the monotonic relationship between the variables. Monotonicity implies that as one variable increases, the other variable consistently either increases or decreases, though not necessarily at a constant rate. The *Spearman correlation* (eq. 2) evaluates the similarity of the rank orders of data points, disregarding their actual numerical values.

$$\rho = 1 - \frac{6\ \sum_{i=1}^{n} d_i^2}{n(n^2-1)} \quad (2)$$

where d_i^2 is the rank difference of two variables after sorting. The Spearman correlation coefficient ranges from -1 to +1. A positive value indicates a direct (increasing) monotonic relationship, while a negative value indicates an inverse (decreasing) monotonic relationship. A coefficient of 0 means there is no monotonic relationship between the variables and indicates a weaker correlation and a smaller influence relationship.

The *Shapley value* φ [26] (eq. 3) attempts to explain an instance's prediction by assessing the contribution of each attribute to the forecast. The Shapley Value calculates the average contribution of each player across all possible coalitions to quantify the impact of the single measured quantities on the objective function.

$$\varphi_i(v) = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(n-|S|-1)!}{n!} (v(S \cup \{i\}) - v(S)) \quad (3)$$

where φ_i is the *Shapley value* for player *i*, *v*(*S*) is the worth coalition *S*, *n* is the number of total players and *m* is the number of players in the coalition *S* before player *i* joins ($0 \leq m \leq n-1$).Positive values imply a positive correlation between input and target while negative values imply that the player detracts from the coalition's value when they join. It is worth highlighting that, as dealing with two output variables, two separate sensitivity analyses were conducted, one for torque and one for fuel consumption. This allowed to identify how the input variables can independently influence the outputs.

To explain the choice of the input parameters used to build the tested neural structures, a step-by-step explanation is reported below. Table 3 shows positive and negative values to highlight the trend

correlation between input and output. To quantify the influence, the absolute value (ABS) of such quantity must be taken into account.

Since the goal of the study is to build a single ANN capable of simultaneously predicting both variables under examination, some considerations were made regarding the indicators, in order to include all influential variables while avoiding the exclusion of others that might positively impact one of the two outputs. First, all variables that simultaneously exhibit negative Shapley values have been removed from the input dataset, i.e., *SparkAdvance, F_P_Turbine _Out, FC_TC_Air_Intake,Lambda, F_TC_Turbine OUT,FC_TC_Compressor_OUT*. Even if *Fuel_pressure* positively influences the FC prediction, the high negative value showed for the torque forecasting led to excluding it as well. *F_Engine Oil T* presents a shapley close to zero for one output and negative for the other: therefore, also sucha variable has been excluded from the initial dataset. *Engine Speed* was included in the initial dataset because, despite the negative Shapley value for torque, the Spearman and Pearson correlations of both output variables with *Engine Speed* were among the highest values.

In other words, 8 of the initial 15 variables have been excluded resulting in roughly over 50% of the initial input variables.

Table 3: Correlation Analysis Results between Input Parameters and Output, I.E. Engine Delivered Torque

Variable Name	Unit	Torque (*F_TorquemeterAI*)			Fuel Consumption (*AVL733SGetMeas*)		
		Shapley	Spearman	Pearson	Shapley	Spearman	Pearson
FC_TC_Compressor_OUT	°C	-8.40	0.55	0.53	-2.56	0.82	0.83
EngineSpeed	rpm	-0.20	0.59	0.58	0.05	0.82	0.73
F_Throttle Position	%	1.03	0.81	0.83	0.01	0.84	0.79
F_Engine Oil T	°C	0.19	-0.01	-0.09	-0.13	-0.17	-0.19
F_TC_Turbine IN	°C	2.19	0.42	0.35	0.03	0.53	0.43
F_TC_Turbine OUT	°C	-1.95	0.36	0.26	-0.08	0.43	0.37
F_TC_Air_Intake	°C	-9.95	-0.11	-0.14	-1.50	-0.27	-0.24
FC_TC_Intercooler_OUT	°C	16.37	0.19	0.12	0.73	0.13	0.14
F_P_Compressor_OUT	mbar	55.37	0.55	0.53	61.43	0.82	0.83
F_P_Turbine_Out	mbar	-54.43	0.12	0.06	24.40	0.06	0.09
Fuel_Pressure	bar	-7.49	-0.40	-0.41	1.17	-0.35	-0.33
InjectionTime	µs	0.96	0.62	0.64	0.04	0.22	0.35
SparkAdvance	CAD	-0.62	0.43	0.40	-0.02	0.70	0.55
TC_H2O	°C	4.22	0.18	0.10	0.11	0.10	0.09
Lambda	-	-1.00	-0.20	-0.34	-0.01	-0.32	-0.37

Since, previous work of the same research group [27] certified increments in the prediction performance when operating with a reduced dataset, the artificial architectures presented have been built by considering a total of 7 input variables. As mentioned before, by reducing the number of input variables, the model becomes more compact, requiring less storage space and this leads to faster training times, quicker model convergence, and shorter inference times during predictions.

c) Preparing the Initial Dataset

After conducting the initial analysis to identify the relevant input parameters, the data (Table 3) is normalized to reduce potential prediction errors and enable the model to converge more quickly. Normalization is necessary to handle variations in input and output parameters effectively. During normalization, the values are scaled to fit within the range [0, 1]. Once the prediction process is complete, the predicted data is de-normalized to facilitate a direct comparison with the actual target values obtained from experimental measurements. Figure 2illustrates the complete dataset used in this study, including the separation of input and output. The training session was realized on 80% of the entire dataset, while the test session, concerning the prediction was realized on the remaining 20%. It is worth reminding that AdaMo acquires data with a sampling frequency of 10 Hz, therefore 5254samples correspond to 525.4 seconds.

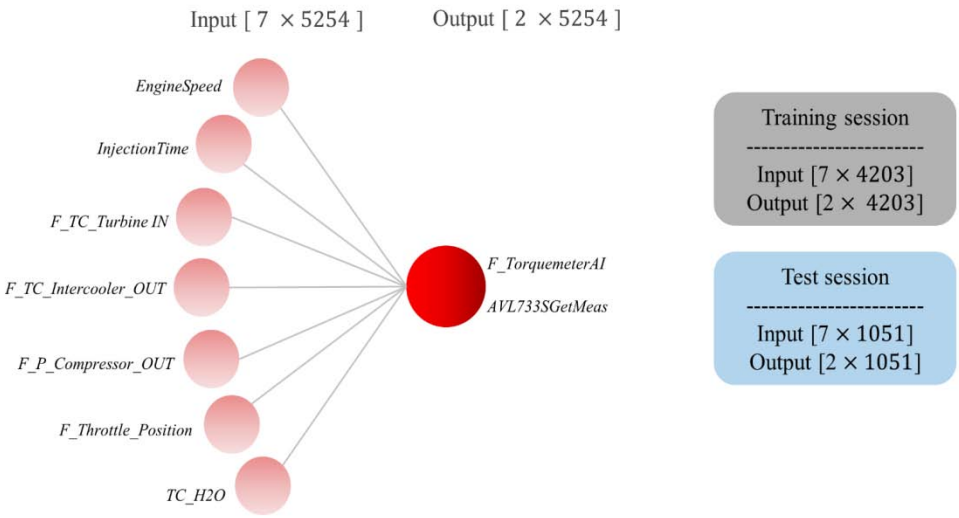


Figure 2: Representation of Input and Output Used to Construct the Artificial Neural Networks (Left) and Subdivision of the Initial Dataset into Training and Test Sessions (Right)

III. Construction of The Neural Architectures

This section details the design and refinement of neural network architectures to predict SI engine torque and fuel consumption. It includes considerations for optimal structure, training techniques like Back Propagation and the NARX approach, and evaluation metrics for performance assessment.

a) Preliminary Considerations

This section describes the different neural network architectures employed in this study to predict the torque delivered and fuel consumption of the SI engine (see *2.1 Experimental setup*) and all the operations the authors made to optimize the internal structure of each of them. In the process of optimization, dealing with numerous variables can be quite challenging. To address this issue, certain considerations must be made to streamline the task. By incorporating these considerations [28-30]:

– The number of neurons N in each hidden layer should be between the number of inputs and the number of outputs (i.e., $2<N<7$).
– The number of neurons in each hidden layer should be 2/3 the number of inputs, plus the number of outputs (i.e., $N=6$).
– The number of neurons in each hidden layer should be less than twice the number of inputs (i.e., $N<14$).

The optimization process can be more efficient. A maximum of 4 hidden units, equal to the double number of variables to be predicted, has been chosen. Each hidden layer is composed of diverse numbers of neurons, i.e. from 4 to13, by following the considerations listed. The number of the epoch was set equal to 1000, since this value has been observed to achieve good performance with reduced computational efforts. To avoid over fitting issues, *Early Stopping* technique [31] is also employed to stop the training process if no significant improvement in performance (MSE eq.4) is recorded above a certain number of epochs. The initial learning rate is set to 0.0001 and the bias and weights of each neuron are continuously optimized by Adaptive Moment Estimation (Adam) which combines the benefits of both RMSprop and momentum optimization to automatically adjusts the learning rate, based on historical gradients and past updates [32]. In fact, it is worth highlighting that, if the learning rate is too high, the model can become unstable and therefore unable to find the best solution. On the other hand, if the learning rate is too small, the model may take a long time to learn without finding a good solution. So, finding the right learning rate is important to train the model effectively [33]. To sum up:

- The number of neurons varies from 4 to 13.
- The hidden layers vary from 1 to 4.
- The number of the epoch is set to 1000.
- The initial learning rate is set at 0.0001.

The definition of the optimal neural structures is determined through preliminary analysis considering the training sessions' performance. To evaluate the training performance of model parameters, the loss function is created, and the mean square error (MSE) is chosen as the loss function (eq.4):

$$MSE = \frac{1}{n}\sum_{i=o}^{n}(Y^{i}_{target} - Y^{i}_{predicted})^{2} \quad (4)$$

with n the number of samples. Set the number of network epoch iterations to 1000, to calculate the final value of the loss function for each prediction model once the network training reaches the maximum learning iteration.

To make a comparison over the entire predicted range, for each forecast i, the deviation of the prediction $Y^i_{predicted}$ from the target Y^i_{target} throughout the range is computed (eq.5):

$$\text{Err}_i = \frac{\left|Y^i_{target} - Y^i_{predicted}\right|}{Y^i_{target}} \times 100 \quad (5)$$

where N is the number of samples considered for the test case and i the i^{th} sample. The average percentual error, i.e. Err_{avg}, is computed as well to draw attention to the global prediction quality. For this kind of application, a maximum critical threshold of 10 is established for the abovementioned errors. Moreover, other two evaluation metrics are used to compare the test performance of the architectures, i.e. RMSE (eq.6) and R^2 (eq.7).

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=o}^{n}(Y^i_{target} - Y^i_{predicted})^2} \quad (6)$$

$$R^2 = 1 - \frac{\sum_{i=o}^{n}(Y^i_{target} - Y^i_{predicted})^2}{\sum_{i=o}^{n}(Y^i_{target} - Y^i_{avg})^2} \quad (7)$$

where Y^i_{avg} is the average value of the prediction.

Due to the large number of combinations analyzed for each structure, it is not feasible to illustrate the performance of every single one. As a result, the authors focused on presenting the features of the best-performing structure for each ANN.

b) Back Propagation

The BP structure has been optimized through the grid search method by testing the following hyper parameter combinations:

- One input layer with size equals the number of features in the input data.
- From 1 to 4 hidden layers H_L each of which with 4 to 13 neurons N. After each hidden layer, a fully connected layer is applied, followed by a ReLU activation function [35].
- An output layer with two output units for the two target variables which directly provides the predicted values without any further processing.
- A regression layer computes the mean squared error (MSE) loss between the predicted output and the target output during the training process.
- The training process presents an Adam optimizer, as previously mentioned, a maximum of 1000 epochs, and the data is processed in mini-batches M_b of 8, 16, 32, 64 or 128 samples per iteration.

c) NARX

NARX approach is a type of recurrent dynamic neural network employed for modeling nonlinear dynamic systems and time-series forecasting [36]. The NARX network can have either a series-parallel (i.e., open-loop) or a parallel architecture (i.e., close-loop) [37], as illustrated in Figure 3.

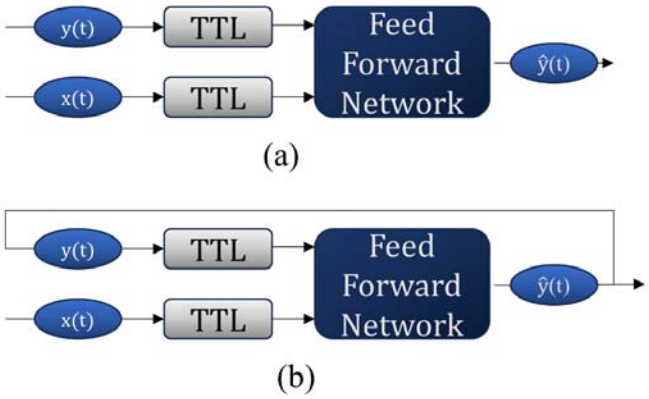


Figure 3: Design of the Type of NARX Architectures: (A) Open-Loop Configuration and (B) Close-Loop One

In the series-parallel architecture (Figure 3 (a)), the network predicts the desired output value, denoted as $\hat{y}$(t), based on the current and past values of the input x(t) and the actual past value of the time series, y (t) (eq.8).

$$\hat{y}(t) = \hat{f}(y_0(t-1); x(t-1)) = \hat{f}(y(t-1), y(t-2)) \ldots \\ \ldots, y(t-n_y), x(t-1), x(t-1) \ldots, x(t-n_y)) \quad (8)$$

with n_x e n_y input memory order and output memory order, respectively.

On the other hand, in the parallel architecture (Figure 3 (b)), the prediction relies on the current and past values of x(t) and the predicted value of $\hat{y}$ (t) (eq.9).

$$\hat{y}(t+1) = \hat{f}(y_0(t-1); x(t-1)) = \hat{f}(y(t-1), y(t-2))\ldots \\ \ldots, y(t-n_y), x(t-1), x(t-1) \ldots, x(t-n_y)) \quad (9)$$

During the training phase, a series-parallel architecture is utilized because it has access to the true past value of the time series. Subsequently, this architecture is transformed into a parallel one, which is more suitable for multi-step-ahead forecasting.

The NARX structure has been optimized through the grid search method by testing the following hyper parameter combinations.

- From 1 to 4 hidden layers H_L each of which with 4 to 13 neurons N.
- An output layer with two output units for the two target variables which directly provides the predicted values without any further processing.
- A regression layer computes the mean squared error (MSE) loss between the predicted output and the target output during the training process.
- The training process presents an Adam optimizer, as previously mentioned, and a maximum of 1000 epochs.

IV. Results and Discussions

a) Results on Training

The first comparison between the proposed algorithms is performed via training_loss function as described in the previous paragraph. The structure of the architectures presenting the best performance on training are composed as follows.

- Back Propagation is composed of three hidden layers H_L, respectively with 11, 9, and 9 neurons N, mini-batch size M_b equals 128, and learning rate equals 0.01.
- NARX is composed of two hidden layers H_L, respectively with 4 and 9 neurons N (Figure 4). Also, the sum of neurons of the two hidden layers (i.e 13) does not exceed twice the number of inputs (see the previous considerations listed regarding the definition of the number of neurons.

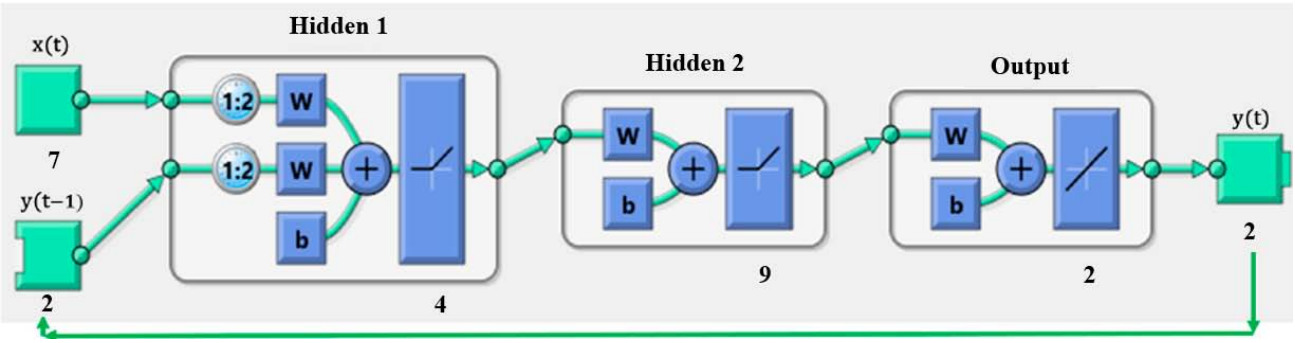


Figure 4: NARX Optimized Structure used to Predict Torque and Fuel Consumption

Figure 5 illustrates a comparison of the training loss among the different architectures. In all cases, there is a consistent decreasing trend in training loss as the number of epochs increases. These architectures tend to reach a point of stability around 600 epochs, where the training loss falls below 0.001 by the time the 1000th epoch is reached. This indicates that the models converge effectively without experiencing over fitting issues. Notably, the NARX architecture demonstrates the quickest convergence, achieving a training loss below 0.005 in approximately at 100 epochs. Furthermore, once it reaches this stabilized state, it exhibits minimal oscillations in the training loss, suggesting that it may be more robust compared to the other architectures.

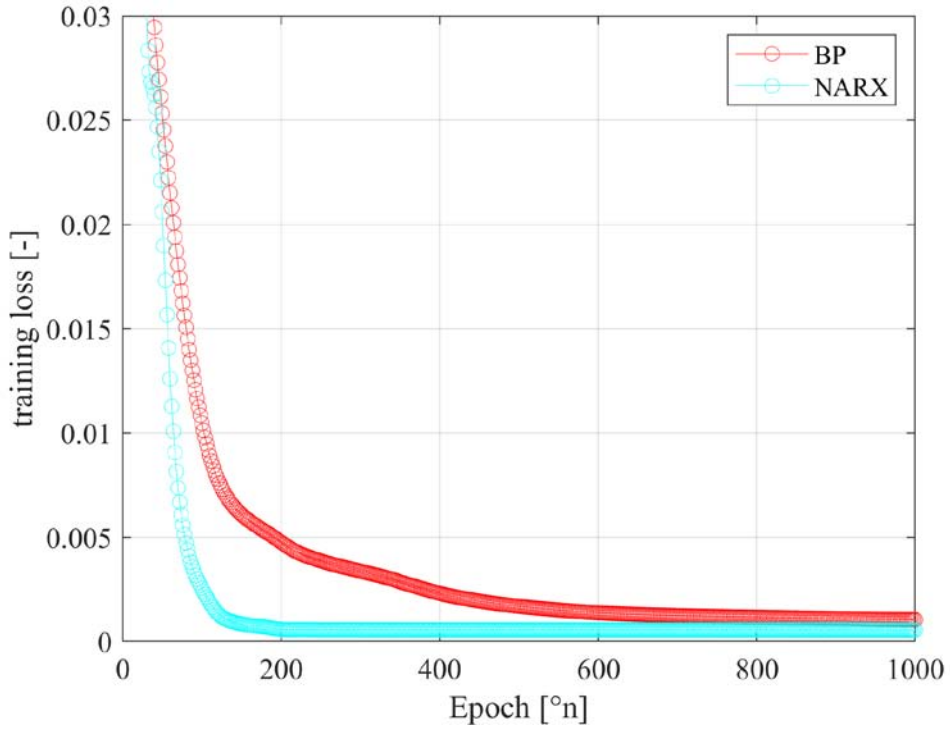


Figure 5: Training Loss for the Tested Architecture during the Training Session

b) Results on Test

Starting from the Back Propagation, Figure 6 displays the comparison between the experimental traces experimentally obtained (black color) and the ones predicted by the optimized BP structure (red color). The prediction error distribution generated by the BP architecture is also presented to enhance the assessment of forecasting accuracy. This approach aims to provide a more precise evaluation of the predictive performance in quantitative terms. From a qualitative point of view, BP is capable of reproducing the physical trend of both analyzed targets. In particular, it well-reproduces the fluctuation of the FC signal, which is characterized by lower oscillations if compared to the torque signal. It is worth noting that, BP model reproduces greater signal fluctuation, particularly noticeable from 465 seconds to 495, taking the torque signal as an example. However, it underestimates the highest peak around 490 seconds. The architecture performs an average error Err_{avg} of about 5.9% regarding the torque prediction and of about 4.7% on the fuel consumption, both less than the critical threshold of 10. Concerning the torque prediction, the architecture exceeds 169 times the Err=10%, corresponding to about 16% of the total samples predicted while, for the fuel consumption, it is able to enhance the prediction quality. In fact, even if the number of samples predicted over the 10% threshold is equal to213, the structure is capable of well-reproducing the target trend. However, the maximum error computed in the FC prediction is 28%, corresponding to about 0.025 kg/h in the initial part of the test (around 420 seconds), and therefore negligible.

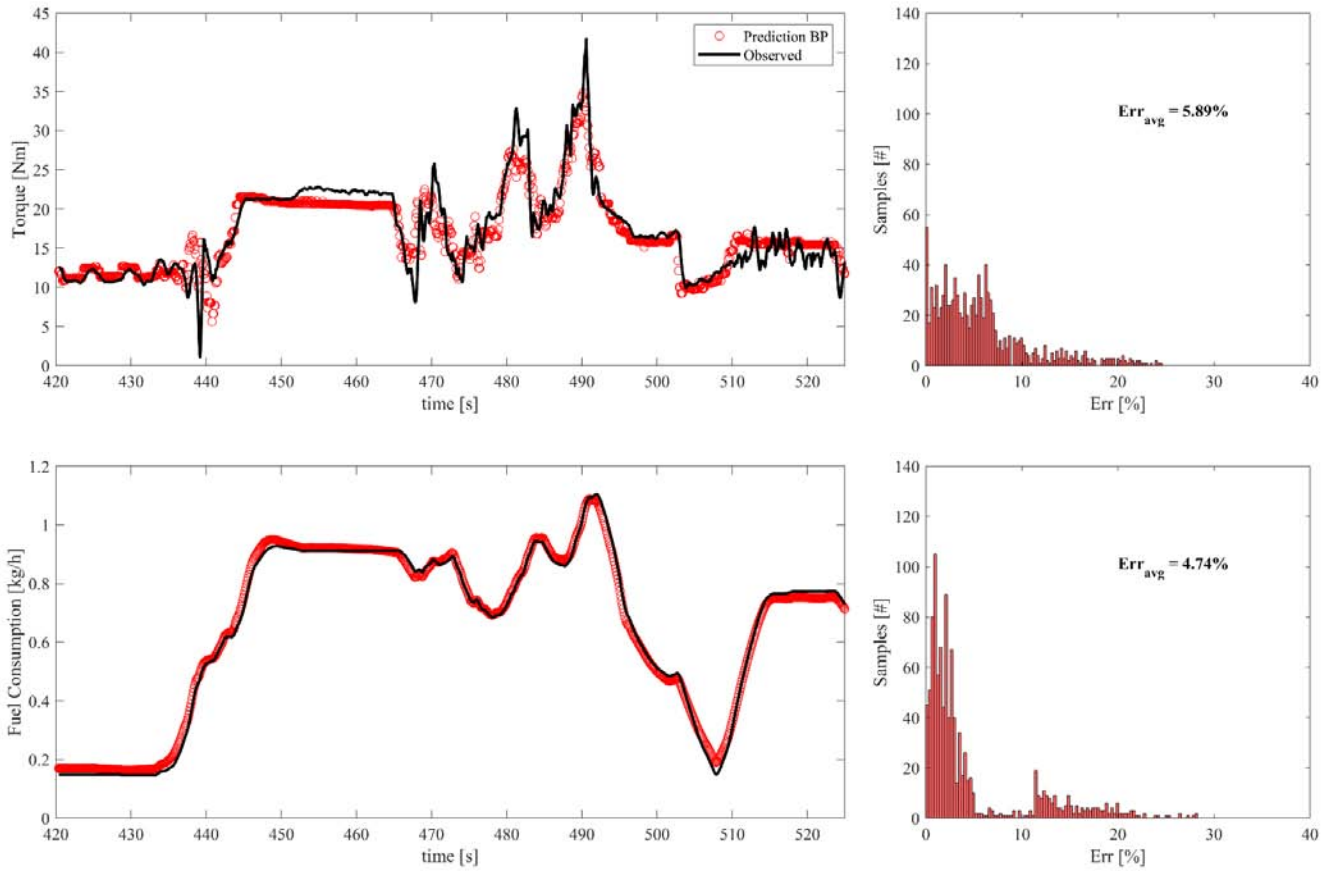


Figure 6: Forecasting Results of the BP Structure. (Left) Comparison between the Performance of the BP Algorithm (Red Curve) and the target Signals (Black Curve)

Figure 7 illustrates the outcomes of the NARX architecture, following the same format as Figure 6 did for the BP architecture. From a qualitative point of view, such a structure is able to mitigate the error in the prediction of BP where the experimental signals are approximatively constants, i.e. in the range from 142 to 152 seconds, and conversely to BP, is capable of reproducing the three highest torque peaks in the interval 450÷500 seconds. NARX outperforms BP with Err_{avg} respectively with about 3.8% and 3.1% of Err_{avg} in torque and fuel consumption prediction, corresponding to improvements of about 35% in both cases. It shows only 6%of the predicted samples exceeding the 10% of Err, in the torque and FC predictions. In the latter case, the Err on prediction never exceeds 16%, conversely to BP which shows errors up to 28%.

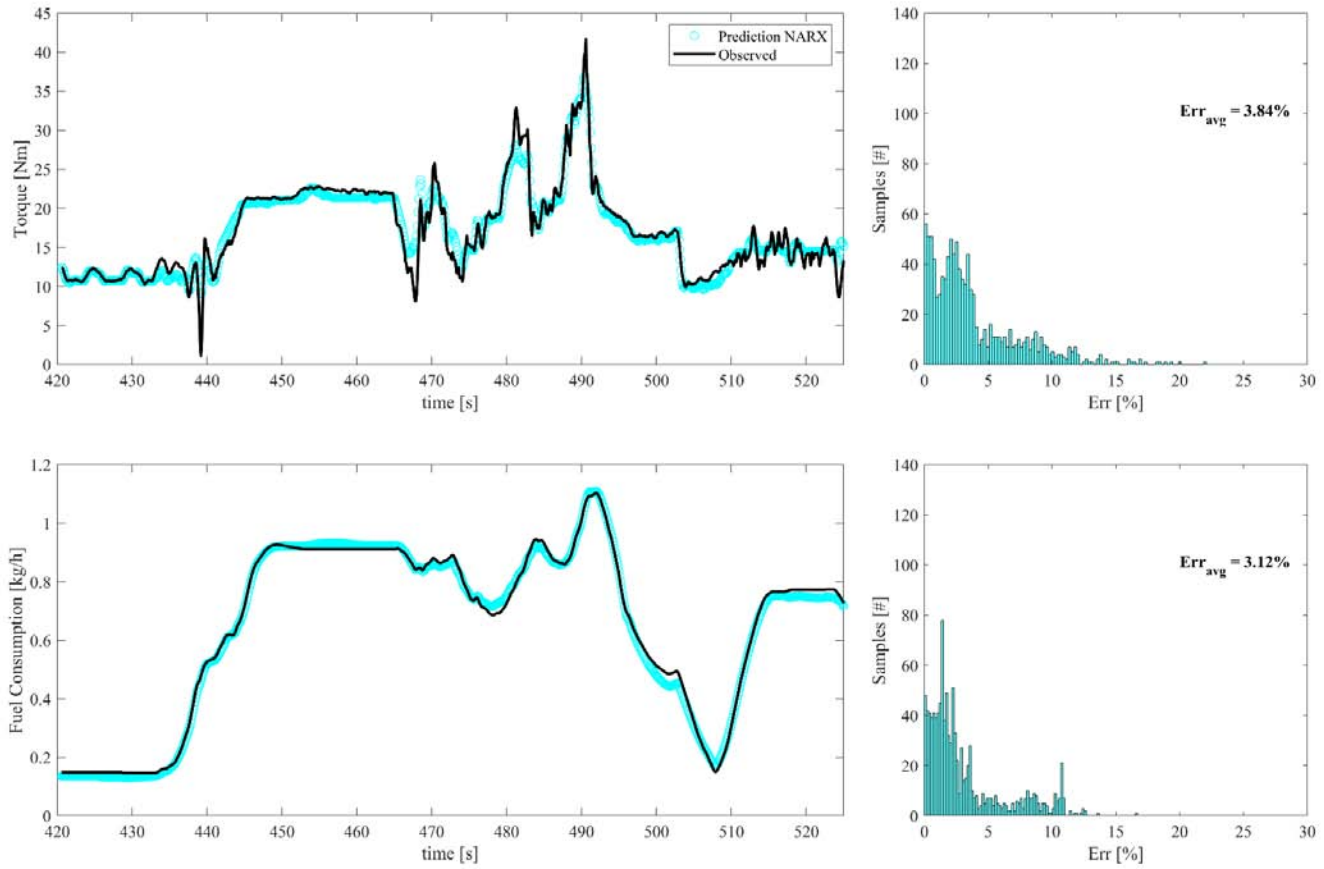
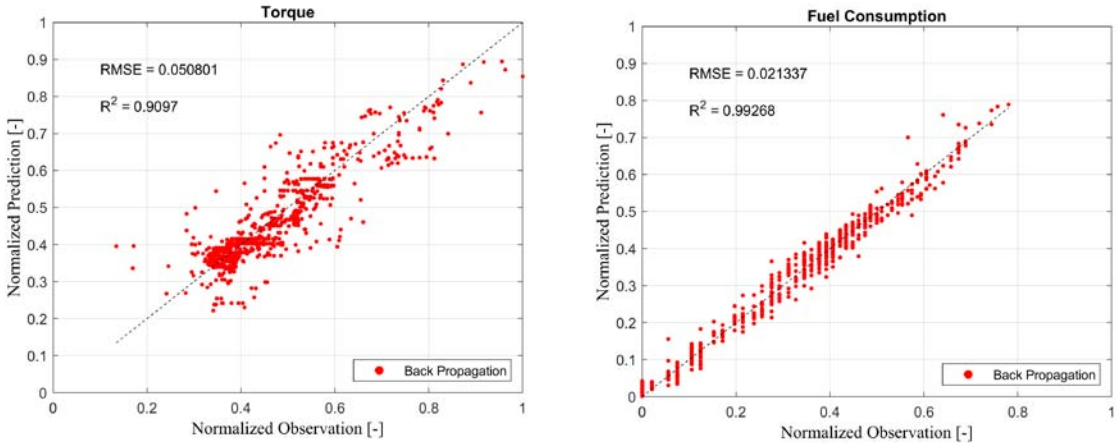


Figure 7: Forecasting Results of the NARX Structure. (Left) Comparison between the Performance of the NARX Algorithm (Green Curve) and the Target Signals (Black Curve)

Figure 8 illustrates the regression accuracy of each model, as defined by eq.7, along with the corresponding Root Mean Square Error. Concerning the Torque Prediction, the NARX model stands out with superior accuracy and smaller prediction errors, as previously demonstrated. In contrast, the BP structure exhibits deviations in the lowest and highest ranges tested. Specifically, it tends to overestimate predictions for low values and underestimate them for high torque levels. NARX improves BP's R-squared values of about 6%, and significantly reduces RMSE of about a 40%. These improvements mitigate the error associated with BP in all the analyzed range. Concerning the fuel consumption, the R^2 and RMSE certify again the capability of both structures to well-predicted such an output feature. Bot structures show R2 close to the unit and RMSE approximately equals to 0.02. These results confirm the higher ability of NARX to effectively addresses prediction of the analyzed targets and its superior nonlinear fitting capabilities.



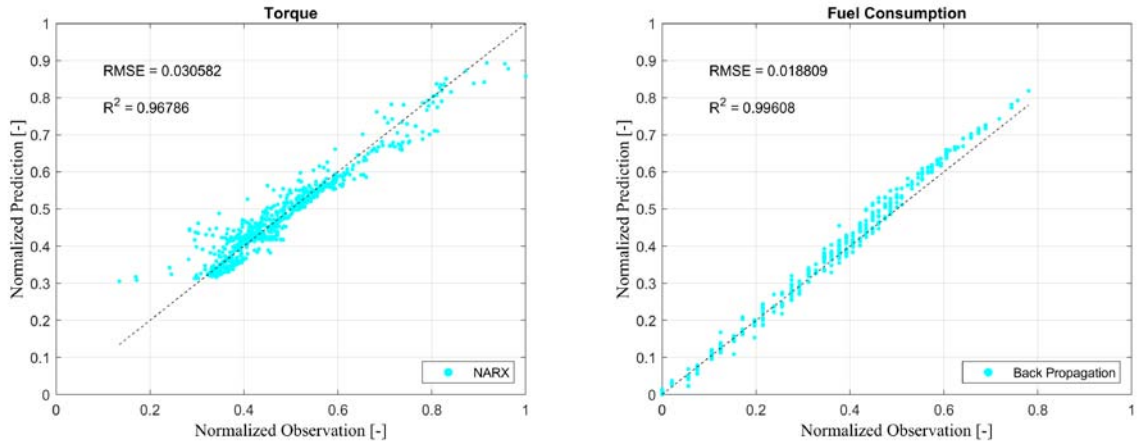


Figure 8: Comparison between the Predictions Performed by the Tested Structures in Terms of R^2 and RMSE

Based on the obtained results, from the torque and FC signals, it's possible to calculate the effective efficiency η_{eff} of the engineby usingeq.10, and then compare it with the experimental efficiency.

$$\eta_{eff} = \frac{\text{Torque} \times w}{FC \times LHV} \times 100 \quad (10)$$

with Torque expressed in Nm, w the engine speed in rad/s, FC the fuel consumption in kg/s and Lower Heating Value (LHV) of the gasoline equals 44 MJ/kg. Figure 9reports the preliminary attempt to compare the efficiency computed from the signals experimentally obtained and the ones predicted by the optimized NARX structure and certifies effectiveness in accurately estimating efficiency across transient conditions.

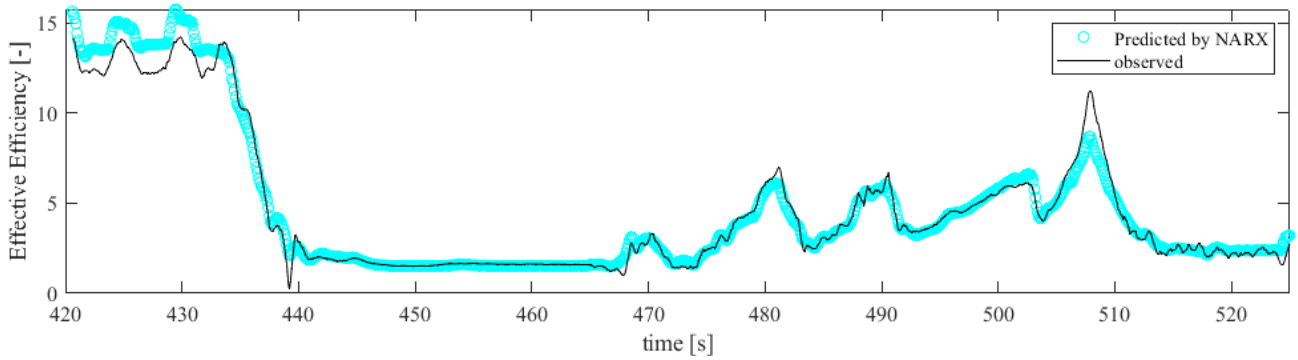


Figure 9: Effective Efficiency Computed by Following Eq.10

V. Conclusions

The goal of this study was to develop an AI-based methodology for evaluating the overall efficiency of a three-cylinder SI engine by simultaneously predicting its torque and fuel consumption. It compared the performance of a Back Propagation (BP) structure with a Nonlinear Autoregressive Network with Exogenous Inputs (NARX) algorithm. In an effort to optimize prediction accuracy, a sensitivity analysis was conducted to select influential parameters while reducing computational complexity and operational time. The experimental dataset was acquired through transient cycle experiments, and hyper parameter optimizations were performed for both BP and NARX architectures using a grid search method.

The results clearly demonstrated that.

- All models effectively converged without over fitting, with training loss consistently decreasing as epochs increase. Notably, the NARX architecture exhibited the quickest convergence, reaching a training loss below 0.005 in around 100 epochs. This suggests its potential robustness if compared to BP.
- NARX outperformed BP in the test phases with an average error of about 3.8% for torque prediction and 3.1% for fuel consumption, representing a 35% improvement in both cases. The NARX model had fewer samples exceeding the 10% error threshold compared to BP.

- NARX demonstrated superior accuracy and smaller prediction errors in torque prediction compared to BP.BP exhibited deviations in the lowest and highest ranges, overestimating for low values and underestimating for high torque levels. NARX improved BP's R^2values by approximately 6% and significantly reduced RMSE by about 40%, mitigating errors across all analyzed ranges. Both models performed well in fuel consumption prediction, with high R-squared values close to the unit and low RMSE.

In summary, the NARX model consistently outperformed the BP architecture in terms of accuracy, prediction error, and robustness, particularly in torque prediction and efficiency estimation. Its ability to capture complex patterns and adapt to changing conditions underscores its superiority in nonlinear fitting. These compelling results underscore the NARX model's critical role in revolutionizing automotive design by significantly enhancing engine efficiency assessment. The model's remarkable predictive accuracy and adaptability hold immense promise for fostering innovation in the automotive industry, ultimately driving the development of greener and more sustainable automotive technologies. The NARX model stands as a valuable tool for engineers and researchers seeking to optimize internal combustion engines, shaping a future where energy-efficient and environmentally friendly vehicles play a central role in transportation.

Nomeclature

Err: Percentage Errors.
Erravg: Average Percentage Errors.
ABSV: Absolute Shapley Values.
BP: Back Propagation
ECU: Engine Control Unit.
ICE: Internal Combustion Engine.
ML: Machine Learning
PFI: Port Fuel Injection.
SI: Spark Ignition.

Author Contributions: Conceptualization, L.P. and F.R.; methodology, F.R.; software, F.R.; validation, L.P., F.R.; formal analysis, L.P. and F.R.; investigation, F.R.; resources, F.M.; data curation, F.R.; writing original draft preparation, F.R.; writing re-view and editing, L.P.; visualization, L.P.; supervision, F.M.; project administration, F.M.; funding acquisition, F.M. All authors have read and agreed to the published version of the manuscript.

Funding: "This research received no external funding"

Data Availability Statement: The data presented in this study are available from the corresponding author. The data are not publicly available due to privacy-related choices.

Conflicts of Interest: the authors declare no conflict of interest.

Bibliography

1. Khan, Muhammed Zafar Ali, et al. "Potential of clean liquid fuels in decarbonizing transportation–An over-looked net-zero pathway?." Renewable and Sustainable Energy Reviews 183 (2023): 113483.
2. Santos, Nathalia Duarte Souza Alvarenga, et al. "Internal combustion engines and biofuels: Examining why this robust combination should not be ignored for future sustainable transportation." Renewable and Sustainable Energy Reviews 148 (2021): 111292.
3. Ricci, Federico, etal. "Using a Machine Learning Approach to Evaluate the NOx Emissions in a Spark-Ignition Optical Engine." Information 14.4 (2023): 224.
4. Sheikh, Shabbir, et al. "Electric Camshaft Phasing System to Meet Euro 6/BS-VI Emission Norms for Gasoline Engine." SAE International Journal of Advances and Current Practices in Mobility 1.2019-26-0055 (2019): 67-75.
5. Huang, Yuhan, et al. "Fuel consumption and emissions performance under real driving: Comparison between hybrid and conventional vehicles." Science of the Total Environment 659 (2019): 275-282.
6. Leach, Felix, et al. "The scope for improving the efficiency and environmental impact of internal combustion engines." Transportation Engineering 1 (2020): 100005.
7. Zembi, Jacopo, et al. Numerical Simulation of the Early Flame Development Produced by a Barrier Discharge Igniter in an Optical Access Engine. No. 2021-24-0011. SAE Technical Paper, 2021.
8. K. Nakata et al."Engine Technologies for Achieving 45% Thermal Efficiency of S.I. Engine", SAE Int. J. Engines 9, 2015-01–1896 (2015).
9. Sellnau, M., Foster, M., Moore, W., Sinnamon, J. et al., "Pathway to 50% Brake Thermal Efficiency Using Gasoline Direct Injection Compression Ignition," SAE Int. J. Adv. & Curr. Prac. in Mobility 1(4): 1581-1603, 2019, https://doi.org/10.4271/2019-01-1154.
10. Xiumin Yu et.al, "Experimental study on the effects of EGR on combustion and emission of an SI engine with gasoline port injection plus ethanol direct injection", Fuel, Volume 305, 2021, 121421, ISSN 0016-2361, https://doi.org/10.1016/j.fuel. 2021.121421.
11. Zsiga, N.; Voser, C.; Onder, C.; Guzzella, L. Intake Manifold Boosting of Turbocharged Spark-Ignited Engines. Energies 2013, 6, 1746–1763, doi: 10.3390/en6031746.
12. Qian, Lijun, et al. "Experimental investigation of water injection and spark timing effects on combustion and emissions of a hybrid hydrogen-gasoline engine." Fuel 322 (2022): 124051

13. Martinelli, Roberto, et al. Lean Combustion Analysis of a Plasma-Assisted Ignition System in a Single Cylinder Engine fueled with E85. No. 2022-24-0034. SAE Technical Paper, 2022.
14. Petrucci Luca, et al. "A Development of a New Image Analysis Technique for Detecting the Flame Front Evolution in Spark Ignition Engine under Lean Condition." Vehicles 4.1 (2022): 145-166.
15. Petrucci, Luca, et al. "Detecting the Flame Front Evolution in Spark-Ignition Engine under Lean Condition Using the Mask R-CNN Approach." Vehicles 4.4 (2022): 978-995.
16. Cervantes-Bobadilla, M., et al. "Multiple fault detection and isolation using artificial neural networks in sensors of an internal combustion engine." Engineering Applications of Artificial Intelligence 117 (2023): 105524.
17. Petrucci, L., et al. "Performance analysis of artificial neural networks for control in internal combustion engines." AIP Conference Proceedings. Vol. 2191. No. 1. AIP Publishing, 2019.
18. Goyal, Somya. "Software measurements using machine learning techniques-a review." Recent Advances in Computer Science and Communications (Formerly: Recent Patents on Computer Science) 16.1 (2023): 38-55.
19. Petrucci Luca, et al. Engine knock evaluation using a machine learning approach. No. 2020-24-0005. SAE Technical Paper, 2020.
20. Togun, Necla Kara, and Sedat Baysec. "Prediction of torque and specific fuel consumption of a gasoline engine by using artificial neural networks." Applied Energy 87.1 (2010): 349-355.
21. Cay, Yusuf. "Prediction of a gasoline engine performance with artificial neural network." Fuel 111 (2013): 324-331.
22. Gölcü, Mustafa, et al. "Artificial neural-network based modeling of variable valve-timing in a spark-ignition engine." Applied Energy 81.2 (2005): 187-197.
23. RicciFederico, et al. "NARX Technique to Predict Torque in Internal Combustion Engines." Information 14.7 (2023): 417.
24. Saadatmorad, Morteza, et al. "Pearson correlation and discrete wavelet transform for crack identification in steel beams." Mathematics 10.15 (2022): 2689.
25. van den Heuvel, Edwin, and Zhuozhao Zhan. "Myths about linear and monotonic associations: Pearson'sr, Spearman's ρ, and Kendall's τ." The American Statistician 76.1 (2022): 44-52.
26. Merrick, Luke, and Ankur Taly. "The explanation game: Explaining machine learning models using shapley values." Machine Learning and Knowledge Extraction: 4th IFIP TC 5, TC 12, WG 8.4, WG 8.9, WG 12.9 International Cross-Domain Conference, CD-MAKE 2020, Dublin, Ireland, August 25–28, 2020, Proceedings 4. Springer International Publishing, 2020.
27. Petrucci, Luca, et al. "From real to virtual sensors, an artificial intelligence approach for the industrial phase of end-of-line quality control of GDI pumps." Measurement 199 (2022): 111583.
28. Charu, C. Aggarwal. Neural networks and deep learning: a textbook. Spinger, 2018.
29. Prusa, Joseph D., and Taghi M. Khoshgoftaar. "Improving deep neural network design with new text data representations." Journal of Big Data 4 (2017): 1-16.
30. Medium.com/geekculture/introduction-to-neural-network-2f8b8221fbd3#:~:text=The%20number%20of%20hidden%20neurons,size%20of%20the%20input%20layer.
31. Bentoumi, Mohamed, et al. "Improvement of emotion recognition from facial images using deep learning and early stopping cross validation." Multimedia Tools and applications 81.21 (2022): 29887-29917.
32. Jais, Imran Khan Mohd, Amelia Ritahani Ismail, and Syed Qamrun Nisa. "Adam optimization algorithm for wide and deep neural network." Knowledge Engineering and Data Science 2.1 (2019): 41-46.
33. Wilson, D. Randall, and Tony R. Martinez. "The need for small learning rates on large problems." IJCNN'01. International Joint Conference on Neural Networks. Proceedings (Cat. No. 01CH37222). Vol. 1. IEEE, 2001.
34. Chicco, Davide, Matthijs J. Warrens, and Giuseppe Jurman. "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation." PeerJ Computer Science 7 (2021): e623.
35. Bai, Yuhan. "RELU-function and derived function review." SHS Web of Conferences. Vol. 144. EDP Sciences, 2022.
36. M. S. H. Lipu, M. A. Hannan, A. Hussain, M. H. M. Saad, A. Ayob, and F. Blaabjerg, ''State of charge estimation for lithium-ion battery using recurrent NARX neural network model based lighting search algorithm,'' IEEE Access, vol. 6, pp. 28150–28161, 2018.
37. Aquize, Rubén, et al. "System Identification Methodology of a Gas Turbine Based on Artificial Recurrent Neural Networks." Sensors 23.4 (2023): 2231.

Global Journal of Computer Science and Technology: H Information & Technology
Volume 23 Issue 2 Version 1.0 Year 2023
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Scalability and Performance of Microservices Architectures

By Amirali Kerimovs

Annotation- The inevitability of continuous evolution and seamless integration of dynamic alterations remains a paramount consideration in the realm of software engineering. This concern is particularly pronounced within the context of contemporary microservices architectures embedded in heterogeneous and decentralized systems composed of numerous interdependent components. A pivotal focal point within such a software design paradigm is to sustain optimal performance quality by ensuring harmonious collaboration among autonomous facets within an intricate framework. The challenge of microservices evolution has predominantly revolved around upholding the harmonization of diverse microservices versions during updates, all while curbing the computational overhead associated with such validation. This study leverages previous research outcomes and tackles the evolution predicament by introducing an innovative formal model, coupled with a fresh exposition of microservices RESTful APIs. The amalgamation of Formal Concept Analysis and the Liskov Substitution Principle plays a pivotal role in this proposed solution. A series of compatibility constraints is delineated and subjected to validation through a controlled experiment employing a representative microservices system. The suggested approach is poised to enhance the development of more sustainable microservices applications and elevate the efficacy of DevOps practices engaged in the creation and upkeep of microservices architectures.

Keywords: *software engineering, project management, microservices; API compatibility; service evolution; concept lattices; substitution principle.*

GJCST-H Classification: *ACM Code: D.2.11*

Strictly as per the compliance and regulations of:

Scalability and Performance of Microservices Architectures

Amirali Kerimovs

Annotation- The inevitability of continuous evolution and seamless integration of dynamic alterations remains a paramount consideration in the realm of software engineering. This concern is particularly pronounced within the context of contemporary microservices architectures embedded in heterogeneous and decentralized systems composed of numerous interdependent components. A pivotal focal point within such a software design paradigm is to sustain optimal performance quality by ensuring harmonious collaboration among autonomous facets within an intricate framework. The challenge of microservices evolution has predominantly revolved around upholding the harmonization of diverse microservices versions during updates, all while curbing the computational overhead associated with such validation. This study leverages previous research outcomes and tackles the evolution predicament by introducing an innovative formal model, coupled with a fresh exposition of microservices RESTful APIs. The amalgamation of Formal Concept Analysis and the Liskov Substitution Principle plays a pivotal role in this proposed solution. A series of compatibility constraints is delineated and subjected to validation through a controlled experiment employing a representative microservices system. The suggested approach is poised to enhance the development of more sustainable microservices applications and elevate the efficacy of DevOps practices engaged in the creation and upkeep of microservices architectures.

Keywords: software engineering, project management, microservices; API compatibility; service evolution; concept lattices; substitution principle.

I. Introduction

In an era where the intricacies and ingenuity of software systems' business logic continue to evolve, microservices have garnered a heightened degree of contemporary attention. This architectural paradigm configures systems as constellations of autonomously deployable services, each strategically aligned to fulfill specific business requisites within well-defined bounds of a domain model. While the merits of this approach abound – allowing for the development, implementation, and deployment of individualized microservices with novel functionalities and bug fixes sans the need to overhaul the entire system – it is not without its inherent constraints.

Amidst the array of benefits, complexities emerge, exemplified by the elevated overall intricacy of the system and the amplified level of inter-service communication. Furthermore, the introduction of each successive iteration of a deployed microservice holds the potential to influence its dependent counterparts, engendering challenges of compatibility both retroactively and proactively, unless meticulous design is observed.

The preceding research encompassed an analysis aimed at juxtaposing two distinct software architectural paradigms – microservices and monolithic architectures. The scope of this analysis encompasses the arrangement of architectural components, the dynamics of their interactions, and the constraints governing these intricate interplays [1]. The findings underscored the advantages inherent to microservices. Among these merits, heightened module isolation emerges as a pivotal facet, facilitating autonomous deployment tracks, embracing heterogeneous technological selections, and fostering augmented productivity [5]. Furthermore, the prevalence of loosely-coupled constructs, exemplified by microservices, engenders an instantaneous impact upon deploying new iterations, thus catering seamlessly to the demands of swiftly-evolving systems. This concurrent flexibility is further underscored by the absence of constraints on the quantity of versions amenable to deployment – an essential feature, considering the perpetual need to incorporate novel functionalities or enhance existing ones.

However, this effusion of microservices, accompanied by their intricate interconnections, begets a palpable challenge concerning the management of contracts and seamless service integration. Consequently, the imperative to anticipate and adeptly manage such transformations at every stage of software development looms large. In essence, the predicament accentuates the exigency for a novel approach, one that can be seamlessly incorporated within the context of this emergent architectural paradigm known as micro-services. This approach ought to draw inspiration from prior research that addresses the crucial facet of ensuring service compatibility within the realm of service-oriented architectures.

II. Literature Review

The intricate challenge of ensuring seamless compatibility throughout the evolutionary journey of services has garnered notable attention in recent years, prompting diverse avenues of research exploration. In our pursuit of comprehensively addressing this intricacy,

Author: Independent Researcher, Riga, Latvia.
e-mail: kerimovsoftdev@gmail.com

we have undertaken a methodical systematic literature review [2], a scholarly endeavor geared towards dissecting the present landscape of this compelling issue.

Within the confines of this research endeavor, we have meticulously articulated a set of stringent prerequisites that guide the selection and evaluation of pertinent studies. These stipulations delineate a clear trajectory for research inclusion: the chosen studies must be firmly rooted within the expansive domain of services/microservices evolution; they must actively encapsulate a meticulously formulated perspective of the intricate conundrum surrounding services /microservices compatibility; and most notably, the studies must introduce novel methodologies that bear their own consequential implications, encapsulating the potential to unravel the complexities of this dynamic arena.

A pivotal cornerstone within our examination is the insightful contribution presented within the paper [10]. This scholarly exposition delves intently into the realm of microservices versioning, meticulously probing their compatibility while concurrently orchestrating a blueprint for the orchestrated deployment of harmonious configurations. At the heart of this endeavor lies the formulation of a version dependency model, an ingeniously crafted framework geared towards facilitating deployment management. The underpinning mechanics of this model are deeply rooted in the construct of a service dependency graph, augmented by a judiciously devised greedy-based optimization algorithm. Significantly, the discourse within [10] does not merely conclude with this innovative framework; rather, it resonates with the clarion call for the expansion and adaptation of existing algorithms to deftly accommodate the evolving landscape of user requisites. A compelling crescendo within this scholarly odyssey is the articulation of an overarching objective – the dual quest to minimize the average response time while concurrently curtailing the formidable evolution cost.

In the grand tapestry of research inquiries, our efforts converge at the crossroads of innovation, grappling with the multifaceted challenge of services evolution and compatibility. Through our systematic review and scholarly analysis, we endeavor to chart a course towards more streamlined and effective methodologies that illuminate the path forward in the intricate realm of services/microservices evolution.

The primary objective of the study documented in [7] revolved around the formulation of an intricate theoretical framework, one that orchestrates the management of service evolution with a nuanced focus encompassing the realms of structure, behavior, and Quality of Service (QoS) induced alterations. A cardinal aspiration was to engineer a paradigm that would seamlessly accommodate these changes while preserving the integrity of type-safe operations. To this end, the research introduced a methodical formalization of the notion of "service compatibility." This endeavor harnessed the potent construct of Abstract Service Descriptions (ASD), meticulously stratified into three distinct layers: structural, behavioral, and non-functional attributes. Notably, the subtyping relation within ASD records was harnessed as a linchpin mechanism for scrutinizing the compatibility interplay among different versions of services.

As the intellectual terrain was traversed, a panorama of diverse methodologies emerged within the purview of microservices evolution. These scholarly offerings unveiled a spectrum of approaches, ranging from those that demonstrated commendable efficacy within real-world contexts to those necessitating recalibration and adaptation to more practical projects. A preponderance of these endeavors pivoted around grappling with the evolutionary challenge at the nexus of software testing and deployment. However, a salient revelation surfaced, spotlighting the paucity of approaches attuned to tackling this predicament at the architectural echelon. Within the crucible of this review, a pivotal quandary crystallized – one that revolves around harnessing the intrinsic features of microservice architecture to fortify compatibility amidst the relentless march of microservices evolution.

Within the contours of this very paper, we embark upon an exploration that envisions an API as an intricate set of meticulously defined regulations that underpin seamless communication between disparate applications. Notably, our discerning focus zeroes in on the contours and intricacies of the RESTful API, an architectural facet meticulously aligned with the design tenets of the REST architectural style. Through we need to unravel the interplay of intricacies inherent within this realm, with the overarching aim of ushering forth a more harmonious coexistence and evolution of microservices.

a) The Purpose of the Article

In this work, the definition of application programming interfaces (APIs) of microservices is formalized, which highlights the specific features of this type of architecture that can be used to maintain compatibility within the framework of updating their versions.

b) Presentation of the Main Material

Effective communication stands as a cornerstone of optimal performance within the realm of microservices architecture. However, it is pivotal to recognize that the microservices responsible for exposing an interface, and the corresponding microservices that leverage said interface, manifest as distinct and autonomously deployable entities [3]. In the event of modifications that disrupt backward compatibility, developers are confronted with a critical choice: either synchronize the deployment of consuming

microservices alongside the introduction of the altered interface, ensuring their seamless transition to the new iteration, or devise a strategy to stagger the introduction of the updated microservice contract. This underscores the imperative of effecting changes in microservices with utmost caution, vigilantly safeguarding compatibility with the intricate web of consuming microservices. An inherent tenet of successful implementation involves the seamless incorporation of backward-compatible alterations, underscored by the imperative to render interfaces conspicuously explicit. In this pursuit, schemas emerge as invaluable tools, assuming the role of custodians to ensure compatibility is sustained.

Furthermore, the underlying technology facilitating this evolution must be engineered with an unwavering commitment to agnosticism, shunning integration technologies that dictate the technology stacks mandated for microservices implementation. An exemplar technology poised to support the seamless evolution of microservices is the REpresentational State Transfer (REST). At the core of this architectural paradigm is the cardinal concept of resources and their multifaceted representations. A pivotal facet is the decoupling of external resource representations from their internal storage mechanisms. This structural underpinning empowers clients to orchestrate requests for alterations to resources, with the server endowed with the prerogative to accede or abstain from compliance.

In the realm of consumption, the OpenAPI specification assumes paramount significance. This specification offers a potent avenue to meticulously outline an array of essential information pertaining to REST endpoints, thereby furnishing the groundwork necessary for the generation of client-side codes across a gamut of programming languages. In essence, the article delves into the intricate fabric of microservices communication, emphasizing compatibility, technology-agnosticism, and the transformative potential of REST and the OpenAPI specification as pivotal instruments in the orchestrated evolution of microservices architectures.

Polymorphism is another important concept that enables procedures or data abstractions to work for multiple types. For example, a sorting procedure should work for any type of element in the array as long as it is possible to compare them. This is called polymorphism. When related types exist in a program, polymorphism is likely to be used, especially when a polymorphic module is needed. In such cases, the supertype is often virtual and serves as a placeholder in the hierarchy for the related types [7].

Let sets be defined in the information system: users U = {u1, u2, u3, u4, u5} and powers P = {p1, p2, p3, p4}. The mapping UP: U → 2P is given by the matrix UP (Fig. 1a). Consider the formal context PU = (P, U, PU), whose matrix PU is determined by the rule: PU = UPT (Fig. 1). The hierarchy of roles generated by the lattice of PU context concepts is shown in Fig. 1c. Roles have the following scope and content.

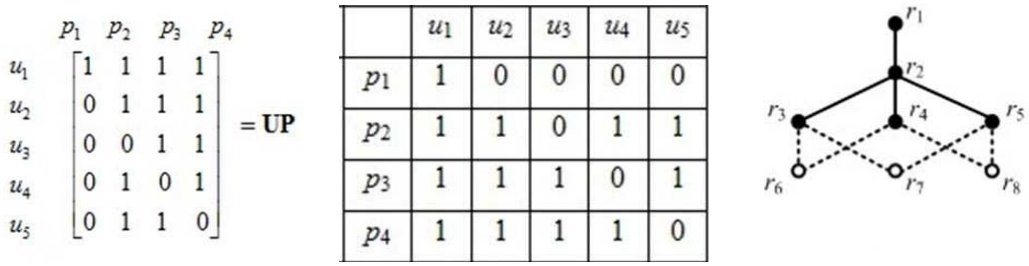


Fig. 1: The process of building a role graph: a) mapping *UP*; b) formal context; c) hierarchy of roles (orientation of arcs: "top-down").

Thus, the main mappings *RP: R → 2 P*, *UR: U → 2 R* and *RR: R → 2 R* will be described by the matrices RP, UR and RR respectively. The proposed algorithm for building a role-based access control model can be used to optimize the existing role hierarchy in the system. Let's demonstrate this approach with an example. This section will show how to move from interfaces and microservices systems to concept grids (FCA approvals).

Information visualization is an important part of data mining (DMA). Most people perceive graphical information best. The visual representation enables information to be visualized and often reveals patterns at a glance that would otherwise only be found through time-consuming analysis. When solving many problems, the task of visualizing partially ordered sets, as well as their particular case, lattices, arises. This problem is especially relevant when using one of the most powerful methods of IAD-the analysis of formal concepts. Formal Concept Analysis (FCA) was proposed by Wille in 1981 [9] and is still actively developed today. In this thesis, the research methodology employed is design science research, which aims to create and evaluate IT artifacts designed to address specific organizational issues. This approach was deemed appropriate for the current project, as the

end goal is to develop and implement a DevOps pipeline as an artifact. However, several challenges must be tackled in order to achieve this, including identifying and motivating the problems, defining solution objectives, designing and developing the artifact, demonstrating its efficacy, evaluating its impact, and communicating the findings. Figure 9 depicts the process model for the design science research methodology (DSRM).

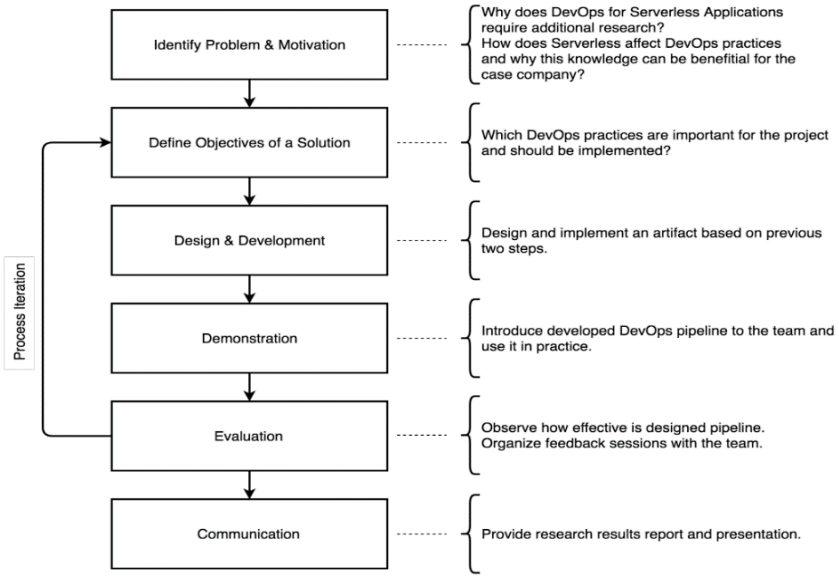


Fig. 2: DSRM Process Model

A formal concept (*concept*) consists of scope and content. Content is a set of properties that describe a concept. The scope includes all objects from the context that have all the properties from the content. In this case, the content should be maximum, i.e., include all the general properties of objects from the scope of the concept. A mathematically formal concept is given using the Galois correspondences [3].

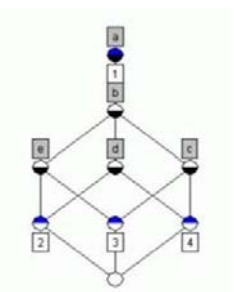
A precedence relation is established between concepts: concept $(A_1, B_1) \leq (A_2, B_2)$ if $A_1 \subseteq A_2$ and $B_1 \supseteq B_2$, where A_1, A_2 are the volumes of concepts 1 and 2, and B_1, B_2 - respectively, their content. Traditionally, Hasse diagrams are used to visualize partially ordered sets. In FCA, a variation of them is used, which uses an abbreviated label - each object and attribute is shown on the diagram only once [4]. The name of an object is assigned to the intersection of all concepts that contain this object, and the name of a property is assigned to the union of all concepts whose contents include this property. Thus, the name of an object is assigned to the smallest of the concepts in which the given object occurs, and the name of the property is assigned to the largest of the concepts in which the property occurs. Such charts are called line charts.

A good line chart should be "transparent, easy to read and easy to interpret the data presented" [8]. However, how this is achieved depends on the goals of the interpretation and on the structure of the lattice.

When drawing a line diagram, it is mandatory that all subconcepts of a concept must be located below it.

	a	b	c	d	e
1	×				
2	×	×		×	×
3	×	×	×		×
4	×	×	×	×	

a)



b)

Fig. 3: a) Formal Context b) Line Diagram of the Lattice of Concepts

It is desirable that the following conditions be met [6]:

- Edges should be drawn as straight lines;
- Two vertices must not be located at the same point;
- The number of intersections between lattice edges should be as small as possible;
- An edge must not intersect a vertex that is not its end;
- The lattice structure should be visually represented;
- Using as few different directions as possible;
- Maximization of the number of parallel lines.

The Liskov Substitution Principle (LSP) is a fundamental principle of object-oriented programming that states that if a program is using a base class, it should be able to use any of its derived classes without knowing it. In the context of APIs evolution processes, this principle can be used to ensure that changes made to the system do not cause unexpected side effects or break existing functionality [11]. Here are some examples of how LSP can be applied in APIs evolution processes:

- Adding a new microservice to a system should not break any existing interfaces or dependencies between services. If the new microservice follows the same interface and dependency rules as the existing services, it can be safely added to the system without any negative impacts.
- Upgrading a software library or tool should not require any changes to the existing codebase that uses it. If the new version of the library or tool is backward-compatible and does not introduce any breaking changes, it can be safely upgraded without causing any issues.
- Adding new features to an application should not break any existing features or workflows. If the new features follow the same patterns and conventions as the existing ones, they can be safely added without causing any issues for users.

In all of these examples, following the Liskov Substitution Principle leads to permissible changes that do not break existing functionality or introduce unexpected side effects. By ensuring that changes are made in a way that adheres to this principle, teams can maintain the stability and reliability of the system over time. To ensure the necessary formalization of microservices, the method of formal concepts analysis (FCA) [14] is used to classify datasets describing microservices APIs into conceptual structures. Then a concept lattice is constructed, which can be used to extract the hierarchical order of concepts, as well as meanings and associations between concepts.

It is expected that this presentation of the API description as a concept and consideration of the evolution of the API as a lattice of concepts will help determine which changes can be compatible. To represent an API as a concept lattice within FCA, it is possible to use the following mathematical equations [14]:

1) A context C = (G, M, I) is defined as a binary relation between objects G (functions or methods) and attributes M (input parameters or output types). The incidence matrix I represents the presence or absence of attributes for each object.
2) The set of all concepts in the lattice L can be obtained using the Galois connection demonstrated in equations (1) and (2):

$$f: 2^G \to 2^M, f(A) = \{m \in M | A \subseteq \{g \in G | (g, m) \in I\}\} \quad (1)$$

$$g: 2^M \to 2^G, g(B) = \{g \in G | B \subseteq \{m \in M | (g, m) \in I\}\} \quad (2)$$

3) The set of all formal concepts in the lattice can be represented as equation (3):

$$L = \{(A, B) | A \subseteq G, B \subseteq M, f(A) = B, g(B) = A\} \quad (3)$$

4) The lattice is formed by pairs of concepts (A, B), where A is a subset of B. The bottom concept represents the set of all objects in the API, while the top concept represents the empty set.

The evolution of an API can be represented as a sequence of concept lattices. Each lattice can be represented as a node in the graph, and the changes made between versions of the API can be represented as edges between the corresponding nodes. The sublattice hierarchy can also be represented as edges between the nodes in the graph.

API compatibility is a measure of how well different versions of an API can interoperate with each other. In general, API compatibility means that changes made to an API should not break existing applications that rely on that API. Let API_v1 be the set of functions and methods provided by version v1 of the API. Let API_v2 be the set of functions and methods provided by version v2 of the API. API compatibility can be expressed as a relation between API_v1 and API_v2, denoted as API_v1 is compatible with API_v2, or symbolically, API_v1 ⊆ API_v2.

This means that all the functions and methods provided by v1 are also provided by v2. It also means that the behavior of these functions and methods in v1 is preserved in v2, i.e., for any f in API_v1, f (args) in v1 should produce the same result as f (args) in v2. Finally, new functionality can be added to v2 without breaking existing applications that rely on v1, i.e., API_v2 \ API_v1 contains only new functions and methods that do not affect the behavior of the existing ones. At the same time, as a binary relation defining a semilattice of concepts, the type hierarchy rule is used to preserve compatibility, or in another way, the substitution principle of Barbara Liskov (Liskov Substitution

Principle) [7] with the following is accepted: φ(x), φ(y) ∈ M - multiple context attributes, x, y ∈ G - a set of context objects, where x has the type T, and y has the type S, in this case , subsets A, B are set to define concepts so that A ⊆ G и B ⊆ M. Definition of concepts: (A1, B1) - the concept of type S, (A2, B2) - the concept of type T. Then, according to the hierarchy of concepts, the concept of type S (A1, B1) is under the concept of a type concept T (A2, B2): (A1, B1) ≤ (A2, B2) when A1 ⊆ A2. Equivalent to (A1, B1) ≤ (A2, B2), B1 ⊇ B2, where in turn is the type S is a subtype of T. Let C be a concept lattice that represents an API, with a set of objects O and a set of attributes A. Let C1 and C2 be two concept lattices that represent different versions of the API, where C2 is a subclass of C1, denoted as C2 ≤ C1. This means that all the objects and attributes in C1 are also in C2, and C2 may have additional objects and attributes.

To apply the Liskov Substitution Principle to concept lattices, it need to ensure that replacing an object from C1 with an object from C2 does not affect the correctness of the program that uses the API. This can be mathematically expressed as [4]: for any object o ∈ O and attribute a ∈ A, if o has the attribute a in C1, then o also has the attribute a in C2. Therefore, by applying the Liskov Substitution Principle to concept lattices and using FCA algorithms to analyze the sublattice structure of the API, ensure that changes made to the API preserve the structure and behavior of the API and do not break backward compatibility.

To represent the API changes using concept lattices, we can first define the lattice structure in terms of the attributes and operations that are supported by the API. We can represent these attributes and operations as concepts in a lattice structure, as shown below:

```graphql
        Add
       /   \
Vector     Scalar
       \   /
      Multiply
```

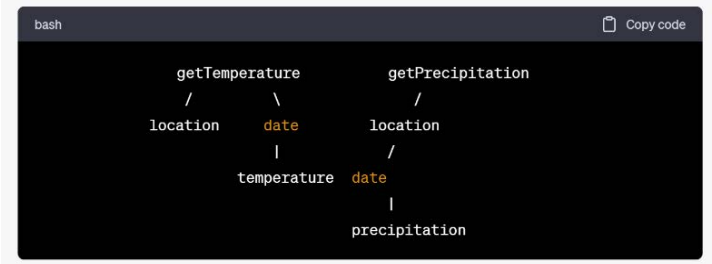
Now, let's say we want to update the API to support matrices. We can add a new attribute "Matrix" to the lattice, and update the "Multiply" operation to support both vectors and matrices, as shown below:

```graphql
        Add
       /   \
Vector     Scalar
       \   /
      Multiply
       /   \
   Matrix   Vector
```

Let's consider another example in the context of a specific microservice, where the objects are API methods and the attributes are variables and data objects (this is the context of a specific microservice). Suppose we have a microservice that provides weather information to users. The microservice has several API methods that provide information about temperature, wind speed, humidity, etc. Each method takes a set of parameters and returns a result. We can represent the API methods and their parameters as objects, and the variables and data objects used in the methods as attributes. For example, the "get Temperature" API method might have the following objects and attributes:

```bash
          getTemperature
           /         \
      location      date
                     |
                temperature
```

Suppose we want to add a new API method that provides information about precipitation. We can add a new object "get Precipitation" to the lattice, along with the relevant attributes. The lattice would now look like this:



We can use the Liskov substitution principle to check the compatibility of the updated API, as before. If the new API method and its parameters can be substituted for the existing API methods without causing unexpected behavior, then they are compatible.

III. Conclusion

The challenge encompassing the evolution of microservices is decisively delineated, underpinned by the compelling necessity to curtail resource expenditure entailed in the verification of microservices compatibility during the intricate realm of DevOps processes and versioning. The clarion call for this endeavor emanates from the imperative to streamline and optimize this vital facet of software development. To illuminate the path forward, a comprehensive literature review is meticulously undertaken, both to glean insights from extant solutions and to discern the contours of unsolved quandaries that persist within this intricate arena.

In the subsequent phases of this scholarly odyssey, a bespoke methodology emerges as the central fulcrum. This strategic blueprint is meticulously crafted, hinging upon the synthesis and harmonization of established theories – notably, the edifice of formal concept analysis and the venerable Liskov Substitution Principle. The crux of this methodology envisages the crystallization of the API's description into intricate concept lattices, an ingeniously conceived framework aimed at elucidating the contours of compatibility. Concurrently, the safeguarding of adherence to the Liskov Substitution Principle assumes paramount significance, unfurling a vista wherein alterations are meticulously scrutinized to discern their compatibility quotient.

The culmination of this holistic approach is a rigorous evaluation, manifested through an empirical analysis of a case study conducted within controlled laboratory environs. This qualitative inquiry discerningly appraises the efficacy, resilience, and practical viability of the constructed artifact. As the curtain falls on this endeavor, its resonating impact reverberates across the expansive landscape of microservices evolution. The research unfurls a pioneering formal methodology, one poised to transcend the realms of academia and seamlessly integrate into the crucible of industrial exigencies. With an overarching commitment to fortify and expedite DevOps processes, this research is imbued with the potential to reshape and optimize the trajectory of microservices evolution within contemporary software engineering paradigms.

The work itself consists of the introductory part, which presents the relevance of the problem; the review and setting part, which demonstrates the thorough review of the existing research on the problem and formulation of the problem. Also, the work includes software implementation for the analysis of applying changes and visualization of the system's evolution, as well as guidelines for a proper usage of the proposed method.

References Références Referencias

1. Andrikopoulos, V., Benbernou, S., & Papazoglou, M. P. On the evolution of services//IEEE Transactions on Software Engineering, 2011, Vol. 38, Issue 3, pp. 609-628.
2. Bernhard Ganter, Rudolf Wille. Formal Concept Analysis: Mathematical Foundations. Springer Berlin, Heidelberg, 1999. 284 pp.
3. Building Microservices. by Sam Newman. Released February 2015. Publisher(s): O'Reilly Media, Inc. ISBN: 9781491950357.
4. Fish, J. (2021). Preface: Revolutionizing Engineering Practice Bymultiscale Methods. International Journal for Multiscale Computational Engineering.
5. He, X., Tu, Z., Liu, L., Xu, X., & Wang, Z. Optimal evolution planning and execution for multi-version coexisting microservice systems//18th International Conference, ICSOC 2020, Dubai, United Arab Emirates, December 14–17, 2020, pp. 3-18.
6. Jonkman, J. M., Wright, A. D., Hayman, G., & Robertson, A. N. (2018). Full-System Linearization for Floating Offshore Wind Turbines in OpenFAST: Preprint.
7. Liskov, B. H., & Wing, J. M. Behavioral subtyping using invariants and constraints//Formal methods for distributed processing: a survey of object-oriented approaches, 2001, pp. 254-280.

8. Taran T. A., Gatin A. R., Ushinkina E. S. Study of the Cognitive Space of the Personality//4th International School-Seminar on Artificial Intelligence for Students and Postgraduates (Braslav School - 2000)/ Sat. scientific _ works. - Minsk: BSUIR, 2000.
9. Wille R. Lattices in data analysis: how to draw them with computer//Algorithms and Order (I. Rival, ed.) Kluwer , Dordrecht -Boston. 1989.
10. X. He, Z. Tu, L. Liu, X. Xu, Z. Wang. "Optimal evolution planning and execution for multi-version coexisting microservice systems," In International Conference on Service-Oriented Computing, vol. 3, pp. 3–18, 2020 Springer, Cham.
11. Yu, S., & Gang, Z. (2020). The Construction of Resource Discovery System Platform Based on Microservices Architecture.

Global Journals Guidelines Handbook 2023

www.GlobalJournals.org

Memberships

Fellows/Associates of Computer Science Research Council

FCSRC/ACSRC Memberships

Introduction

FCSRC/ACSRC is the most prestigious membership of Global Journals accredited by Open Association of Research Society, U.S.A (OARS). The credentials of Fellow and Associate designations signify that the researcher has gained the knowledge of the fundamental and high-level concepts, and is a subject matter expert, proficient in an expertise course covering the professional code of conduct, and follows recognized standards of practice. The credentials are designated only to the researchers, scientists, and professionals that have been selected by a rigorous process by our Editorial Board and Management Board.

Associates of FCSRC/ACSRC are scientists and researchers from around the world are working on projects/researches that have huge potentials. Members support Global Journals' mission to advance technology for humanity and the profession.

FCSRC

Fellow of Computer Science Research Council

FELLOW OF COMPUTER SCIENCE RESEARCH COUNCIL is the most prestigious membership of Global Journals. It is an award and membership granted to individuals that the Open Association of Research Society judges to have made a 'substantial contribution to the improvement of computer science, technology, and electronics engineering.

The primary objective is to recognize the leaders in research and scientific fields of the current era with a global perspective and to create a channel between them and other researchers for better exposure and knowledge sharing. Members are most eminent scientists, engineers, and technologists from all across the world. Fellows are elected for life through a peer review process on the basis of excellence in the respective domain. There is no limit on the number of new nominations made in any year. Each year, the Open Association of Research Society elect up to 12 new Fellow Members.

Benefit

To the institution

Get letter of appreciation

Global Journals sends a letter of appreciation of author to the Dean or CEO of the University or Company of which author is a part, signed by editor in chief or chief author.

Exclusive Network

Get access to a closed network

A FCSRC member gets access to a closed network of Tier 1 researchers and scientists with direct communication channel through our website. Fellows can reach out to other members or researchers directly.They should also be open to reaching out by other.

Career | Credibility | Exclusive | Reputation

Certificate

Certificate, LoR and Laser-Momento

Fellows receive a printed copy of a certificate signed by our Chief Author that may be used for academic purposes and a personal recommendation letter to the dean of member's university.

Career | Credibility | Exclusive | Reputation

Designation

Get honored title of membership

Fellows can use the honored title of membership. The “FCSRC” is an honored title which is accorded to a person’s name viz. Dr. John E. Hall, Ph.D., FCSRC or William Walldroff, M.S., FCSRC.

Career | Credibility | Exclusive | Reputation

Recognition on the Platform

Better visibility and citation

All the Fellow members of FCSRC get a badge of "Leading Member of Global Journals" on the Research Community that distinguishes them from others. Additionally, the profile is also partially maintained by our team for better visibility and citation. All fellows get a dedicated page on the website with their biography.

Career | Credibility | Reputation

Future Work

Get discounts on the future publications

Fellows receive discounts on future publications with Global Journals up to 60%. Through our recommendation programs, members also receive discounts on publications made with OARS affiliated organizations.

Career | Financial

GJ Account

Unlimited forward of Emails

Fellows get secure and fast GJ work emails with unlimited forward of emails that they may use them as their primary email. For example, john [AT] globaljournals [DOT] org.

Career | Credibility | Reputation

Premium Tools

Access to all the premium tools

To take future researches to the zenith, fellows receive access to all the premium tools that Global Journals have to offer along with the partnership with some of the best marketing leading tools out there.

Financial

Conferences & Events

Organize seminar/conference

Fellows are authorized to organize symposium/seminar/conference on behalf of Global Journal Incorporation (USA). They can also participate in the same organized by another institution as representative of Global Journal. In both the cases, it is mandatory for him to discuss with us and obtain our consent. Additionally, they get free research conferences (and others) alerts.

Career | Credibility | Financial

Early Invitations

Early invitations to all the symposiums, seminars, conferences

All fellows receive the early invitations to all the symposiums, seminars, conferences and webinars hosted by Global Journals in their subject.

Exclusive

Publishing Articles & Books

Earn 60% of sales proceeds

Fellows can publish articles (limited) without any fees. Also, they can earn up to 70% of sales proceeds from the sale of reference/review books/literature/publishing of research paper. The FCSRC member can decide its price and we can help in making the right decision.

Exclusive | Financial

Reviewers

Get a remuneration of 15% of author fees

Fellow members are eligible to join as a paid peer reviewer at Global Journals Incorporation (USA) and can get a remuneration of 15% of author fees, taken from the author of a respective paper.

Financial

Access to Editorial Board

Become a member of the Editorial Board

Fellows may join as a member of the Editorial Board of Global Journals Incorporation (USA) after successful completion of three years as Fellow and as Peer Reviewer. Additionally, Fellows get a chance to nominate other members for Editorial Board.

Career | Credibility | Exclusive | Reputation

And Much More

Get access to scientific museums and observatories across the globe

All members get access to 5 selected scientific museums and observatories across the globe. All researches published with Global Journals will be kept under deep archival facilities across regions for future protections and disaster recovery. They get 10 GB free secure cloud access for storing research files.

ASSOCIATE OF COMPUTER SCIENCE RESEARCH COUNCIL

ASSOCIATE OF COMPUTER SCIENCE RESEARCH COUNCIL is the membership of Global Journals awarded to individuals that the Open Association of Research Society judges to have made a 'substantial contribution to the improvement of computer science, technology, and electronics engineering.

The primary objective is to recognize the leaders in research and scientific fields of the current era with a global perspective and to create a channel between them and other researchers for better exposure and knowledge sharing. Members are most eminent scientists, engineers, and technologists from all across the world. Associate membership can later be promoted to Fellow Membership. Associates are elected for life through a peer review process on the basis of excellence in the respective domain. There is no limit on the number of new nominations made in any year. Each year, the Open Association of Research Society elect up to 12 new Associate Members.

To the institution

Get letter of appreciation

Global Journals sends a letter of appreciation of author to the Dean or CEO of the University or Company of which author is a part, signed by editor in chief or chief author.

Exclusive Network

Get access to a closed network

A ACSRC member gets access to a closed network of Tier 2 researchers and scientists with direct communication channel through our website. Associates can reach out to other members or researchers directly.They should also be open to reaching out by other.

Career | Credibility | Exclusive | Reputation

Certificate

Certificate, LoR and Laser-Momento

Associates receive a printed copy of a certificate signed by our Chief Author that may be used for academic purposes and a personal recommendation letter to the dean of member's university.

Career | Credibility | Exclusive | Reputation

Designation

Get honored title of membership

Associates can use the honored title of membership. The “ACSRC” is an honored title which is accorded to a person’s name viz. Dr. John E. Hall, Ph.D., ACSRC or William Walldroff, M.S., ACSRC.

Career | Credibility | Exclusive | Reputation

Recognition on the Platform

Better visibility and citation

All the Associate members of ACSRC get a badge of "Leading Member of Global Journals" on the Research Community that distinguishes them from others. Additionally, the profile is also partially maintained by our team for better visibility and citation.

Career | Credibility | Reputation

Future Work

Get discounts on the future publications

Associates receive discounts on future publications with Global Journals up to 30%. Through our recommendation programs, members also receive discounts on publications made with OARS affiliated organizations.

Career | Financial

GJ Account

Unlimited forward of Emails

Associates get secure and fast GJ work emails with 5GB forward of emails that they may use them as their primary email. For example, john [AT] globaljournals [DOT] org.

Career | Credibility | Reputation

Premium Tools

Access to all the premium tools

To take future researches to the zenith, associates receive access to all the premium tools that Global Journals have to offer along with the partnership with some of the best marketing leading tools out there.

Financial

Conferences & Events

Organize seminar/conference

Associates are authorized to organize symposium/seminar/conference on behalf of Global Journal Incorporation (USA). They can also participate in the same organized by another institution as representative of Global Journal. In both the cases, it is mandatory for him to discuss with us and obtain our consent. Additionally, they get free research conferences (and others) alerts.

Career | Credibility | Financial

Early Invitations

Early invitations to all the symposiums, seminars, conferences

All associates receive the early invitations to all the symposiums, seminars, conferences and webinars hosted by Global Journals in their subject.

Exclusive

Publishing Articles & Books

Earn 30-40% of sales proceeds

Associates can publish articles (limited) without any fees. Also, they can earn up to 30-40% of sales proceeds from the sale of reference/review books/literature/publishing of research paper.

Exclusive | Financial

Reviewers

Get a remuneration of 15% of author fees

Associate members are eligible to join as a paid peer reviewer at Global Journals Incorporation (USA) and can get a remuneration of 15% of author fees, taken from the author of a respective paper.

Financial

And Much More

Get access to scientific museums and observatories across the globe

All members get access to 2 selected scientific museums and observatories across the globe. All researches published with Global Journals will be kept under deep archival facilities across regions for future protections and disaster recovery. They get 5 GB free secure cloud access for storing research files.

Associate	Fellow	Research Group	Basic
$4800 **lifetime designation**	$6800 **lifetime designation**	$12500.00 **organizational**	APC **per article**
Certificate, LoR and Momento **2** discounted publishing/year **Gradation** of Research **10** research contacts/day **1 GB** Cloud Storage **GJ** Community Access	**Certificate,** LoR and Momento **Unlimited** discounted publishing/year **Gradation** of Research **Unlimited** research contacts/day **5 GB** Cloud Storage **Online Presense** Assistance **GJ** Community Access	**Certificates,** LoRs and Momentos **Unlimited** free publishing/year **Gradation** of Research **Unlimited** research contacts/day **Unlimited** Cloud Storage **Online Presense** Assistance **GJ** Community Access	**GJ** Community Access

Preferred Author Guidelines

We accept the manuscript submissions in any standard (generic) format.

We typeset manuscripts using advanced typesetting tools like Adobe In Design, CorelDraw, TeXnicCenter, and TeXStudio. We usually recommend authors submit their research using any standard format they are comfortable with, and let Global Journals do the rest.

Alternatively, you can download our basic template from https://globaljournals.org/Template.zip

Authors should submit their complete paper/article, including text illustrations, graphics, conclusions, artwork, and tables. Authors who are not able to submit manuscript using the form above can email the manuscript department at submit@globaljournals.org or get in touch with chiefeditor@globaljournals.org if they wish to send the abstract before submission.

Before and during Submission

Authors must ensure the information provided during the submission of a paper is authentic. Please go through the following checklist before submitting:

1. Authors must go through the complete author guideline and understand and *agree to Global Journals' ethics and code of conduct,* along with author responsibilities.
2. Authors must accept the privacy policy, terms, and conditions of Global Journals.
3. Ensure corresponding author's email address and postal address are accurate and reachable.
4. Manuscript to be submitted must include keywords, an abstract, a paper title, co-author(s') names and details (email address, name, phone number, and institution), figures and illustrations in vector format including appropriate captions, tables, including titles and footnotes, a conclusion, results, acknowledgments and references.
5. Authors should submit paper in a ZIP archive if any supplementary files are required along with the paper.
6. Proper permissions must be acquired for the use of any copyrighted material.
7. Manuscript submitted *must not have been submitted or published elsewhere* and all authors must be aware of the submission.

Declaration of Conflicts of Interest

It is required for authors to declare all financial, institutional, and personal relationships with other individuals and organizations that could influence (bias) their research.

Policy on Plagiarism

Plagiarism is not acceptable in Global Journals submissions at all.

Plagiarized content will not be considered for publication. We reserve the right to inform authors' institutions about plagiarism detected either before or after publication. If plagiarism is identified, we will follow COPE guidelines:

Authors are solely responsible for all the plagiarism that is found. The author must not fabricate, falsify or plagiarize existing research data. The following, if copied, will be considered plagiarism:

- Words (language)
- Ideas
- Findings
- Writings
- Diagrams
- Graphs
- Illustrations
- Lectures

- Printed material
- Graphic representations
- Computer programs
- Electronic material
- Any other original work

Authorship Policies

Global Journals follows the definition of authorship set up by the Open Association of Research Society, USA. According to its guidelines, authorship criteria must be based on:

1. Substantial contributions to the conception and acquisition of data, analysis, and interpretation of findings.
2. Drafting the paper and revising it critically regarding important academic content.
3. Final approval of the version of the paper to be published.

Changes in Authorship

The corresponding author should mention the name and complete details of all co-authors during submission and in manuscript. We support addition, rearrangement, manipulation, and deletions in authors list till the early view publication of the journal. We expect that corresponding author will notify all co-authors of submission. We follow COPE guidelines for changes in authorship.

Copyright

During submission of the manuscript, the author is confirming an exclusive license agreement with Global Journals which gives Global Journals the authority to reproduce, reuse, and republish authors' research. We also believe in flexible copyright terms where copyright may remain with authors/employers/institutions as well. Contact your editor after acceptance to choose your copyright policy. You may follow this form for copyright transfers.

Appealing Decisions

Unless specified in the notification, the Editorial Board's decision on publication of the paper is final and cannot be appealed before making the major change in the manuscript.

Acknowledgments

Contributors to the research other than authors credited should be mentioned in Acknowledgments. The source of funding for the research can be included. Suppliers of resources may be mentioned along with their addresses.

Declaration of funding sources

Global Journals is in partnership with various universities, laboratories, and other institutions worldwide in the research domain. Authors are requested to disclose their source of funding during every stage of their research, such as making analysis, performing laboratory operations, computing data, and using institutional resources, from writing an article to its submission. This will also help authors to get reimbursements by requesting an open access publication letter from Global Journals and submitting to the respective funding source.

Preparing your Manuscript

Authors can submit papers and articles in an acceptable file format: MS Word (doc, docx), LaTeX (.tex, .zip or .rar including all of your files), Adobe PDF (.pdf), rich text format (.rtf), simple text document (.txt), Open Document Text (.odt), and Apple Pages (.pages). Our professional layout editors will format the entire paper according to our official guidelines. This is one of the highlights of publishing with Global Journals—authors should not be concerned about the formatting of their paper. Global Journals accepts articles and manuscripts in every major language, be it Spanish, Chinese, Japanese, Portuguese, Russian, French, German, Dutch, Italian, Greek, or any other national language, but the title, subtitle, and abstract should be in English. This will facilitate indexing and the pre-peer review process.

The following is the official style and template developed for publication of a research paper. Authors are not required to follow this style during the submission of the paper. It is just for reference purposes.

Manuscript Style Instruction (Optional)

- Microsoft Word Document Setting Instructions.
- Font type of all text should be Swis721 Lt BT.
- Page size: 8.27" x 11'", left margin: 0.65, right margin: 0.65, bottom margin: 0.75.
- Paper title should be in one column of font size 24.
- Author name in font size of 11 in one column.
- Abstract: font size 9 with the word "Abstract" in bold italics.
- Main text: font size 10 with two justified columns.
- Two columns with equal column width of 3.38 and spacing of 0.2.
- First character must be three lines drop-capped.
- The paragraph before spacing of 1 pt and after of 0 pt.
- Line spacing of 1 pt.
- Large images must be in one column.
- The names of first main headings (Heading 1) must be in Roman font, capital letters, and font size of 10.
- The names of second main headings (Heading 2) must not include numbers and must be in italics with a font size of 10.

Structure and Format of Manuscript

The recommended size of an original research paper is under 15,000 words and review papers under 7,000 words. Research articles should be less than 10,000 words. Research papers are usually longer than review papers. Review papers are reports of significant research (typically less than 7,000 words, including tables, figures, and references)

A research paper must include:

a) A title which should be relevant to the theme of the paper.
b) A summary, known as an abstract (less than 150 words), containing the major results and conclusions.
c) Up to 10 keywords that precisely identify the paper's subject, purpose, and focus.
d) An introduction, giving fundamental background objectives.
e) Resources and techniques with sufficient complete experimental details (wherever possible by reference) to permit repetition, sources of information must be given, and numerical methods must be specified by reference.
f) Results which should be presented concisely by well-designed tables and figures.
g) Suitable statistical data should also be given.
h) All data must have been gathered with attention to numerical detail in the planning stage.

Design has been recognized to be essential to experiments for a considerable time, and the editor has decided that any paper that appears not to have adequate numerical treatments of the data will be returned unrefereed.

i) Discussion should cover implications and consequences and not just recapitulate the results; conclusions should also be summarized.
j) There should be brief acknowledgments.
k) There ought to be references in the conventional format. Global Journals recommends APA format.

Authors should carefully consider the preparation of papers to ensure that they communicate effectively. Papers are much more likely to be accepted if they are carefully designed and laid out, contain few or no errors, are summarizing, and follow instructions. They will also be published with much fewer delays than those that require much technical and editorial correction.

The Editorial Board reserves the right to make literary corrections and suggestions to improve brevity.

Format Structure

It is necessary that authors take care in submitting a manuscript that is written in simple language and adheres to published guidelines.

All manuscripts submitted to Global Journals should include:

Title

The title page must carry an informative title that reflects the content, a running title (less than 45 characters together with spaces), names of the authors and co-authors, and the place(s) where the work was carried out.

Author details

The full postal address of any related author(s) must be specified.

Abstract

The abstract is the foundation of the research paper. It should be clear and concise and must contain the objective of the paper and inferences drawn. It is advised to not include big mathematical equations or complicated jargon.

Many researchers searching for information online will use search engines such as Google, Yahoo or others. By optimizing your paper for search engines, you will amplify the chance of someone finding it. In turn, this will make it more likely to be viewed and cited in further works. Global Journals has compiled these guidelines to facilitate you to maximize the web-friendliness of the most public part of your paper.

Keywords

A major lynchpin of research work for the writing of research papers is the keyword search, which one will employ to find both library and internet resources. Up to eleven keywords or very brief phrases have to be given to help data retrieval, mining, and indexing.

One must be persistent and creative in using keywords. An effective keyword search requires a strategy: planning of a list of possible keywords and phrases to try.

Choice of the main keywords is the first tool of writing a research paper. Research paper writing is an art. Keyword search should be as strategic as possible.

One should start brainstorming lists of potential keywords before even beginning searching. Think about the most important concepts related to research work. Ask, “What words would a source have to include to be truly valuable in a research paper?” Then consider synonyms for the important words.

It may take the discovery of only one important paper to steer in the right keyword direction because, in most databases, the keywords under which a research paper is abstracted are listed with the paper.

Numerical Methods

Numerical methods used should be transparent and, where appropriate, supported by references.

Abbreviations

Authors must list all the abbreviations used in the paper at the end of the paper or in a separate table before using them.

Formulas and equations

Authors are advised to submit any mathematical equation using either MathJax, KaTeX, or LaTeX, or in a very high-quality image.

Tables, Figures, and Figure Legends

Tables: Tables should be cautiously designed, uncrowned, and include only essential data. Each must have an Arabic number, e.g., Table 4, a self-explanatory caption, and be on a separate sheet. Authors must submit tables in an editable format and not as images. References to these tables (if any) must be mentioned accurately.

Figures

Figures are supposed to be submitted as separate files. Always include a citation in the text for each figure using Arabic numbers, e.g., Fig. 4. Artwork must be submitted online in vector electronic form or by emailing it.

Preparation of Eletronic Figures for Publication

Although low-quality images are sufficient for review purposes, print publication requires high-quality images to prevent the final product being blurred or fuzzy. Submit (possibly by e-mail) EPS (line art) or TIFF (halftone/ photographs) files only. MS PowerPoint and Word Graphics are unsuitable for printed pictures. Avoid using pixel-oriented software. Scans (TIFF only) should have a resolution of at least 350 dpi (halftone) or 700 to 1100 dpi (line drawings). Please give the data for figures in black and white or submit a Color Work Agreement form. EPS files must be saved with fonts embedded (and with a TIFF preview, if possible).

For scanned images, the scanning resolution at final image size ought to be as follows to ensure good reproduction: line art: >650 dpi; halftones (including gel photographs): >350 dpi; figures containing both halftone and line images: >650 dpi.

Color charges: Authors are advised to pay the full cost for the reproduction of their color artwork. Hence, please note that if there is color artwork in your manuscript when it is accepted for publication, we would require you to complete and return a Color Work Agreement form before your paper can be published. Also, you can email your editor to remove the color fee after acceptance of the paper.

Tips for writing a good quality Computer Science Research Paper

Techniques for writing a good quality computer science research paper:

1. Choosing the topic: In most cases, the topic is selected by the interests of the author, but it can also be suggested by the guides. You can have several topics, and then judge which you are most comfortable with. This may be done by asking several questions of yourself, like "Will I be able to carry out a search in this area? Will I find all necessary resources to accomplish the search? Will I be able to find all information in this field area?" If the answer to this type of question is "yes," then you ought to choose that topic. In most cases, you may have to conduct surveys and visit several places. Also, you might have to do a lot of work to find all the rises and falls of the various data on that subject. Sometimes, detailed information plays a vital role, instead of short information. Evaluators are human: The first thing to remember is that evaluators are also human beings. They are not only meant for rejecting a paper. They are here to evaluate your paper. So present your best aspect.

2. Think like evaluators: If you are in confusion or getting demotivated because your paper may not be accepted by the evaluators, then think, and try to evaluate your paper like an evaluator. Try to understand what an evaluator wants in your research paper, and you will automatically have your answer. Make blueprints of paper: The outline is the plan or framework that will help you to arrange your thoughts. It will make your paper logical. But remember that all points of your outline must be related to the topic you have chosen.

3. Ask your guides: If you are having any difficulty with your research, then do not hesitate to share your difficulty with your guide (if you have one). They will surely help you out and resolve your doubts. If you can't clarify what exactly you require for your work, then ask your supervisor to help you with an alternative. He or she might also provide you with a list of essential readings.

4. Use of computer is recommended: As you are doing research in the field of computer science then this point is quite obvious. Use right software: Always use good quality software packages. If you are not capable of judging good software, then you can lose the quality of your paper unknowingly. There are various programs available to help you which you can get through the internet.

5. Use the internet for help: An excellent start for your paper is using Google. It is a wondrous search engine, where you can have your doubts resolved. You may also read some answers for the frequent question of how to write your research paper or find a model research paper. You can download books from the internet. If you have all the required books, place importance on reading, selecting, and analyzing the specified information. Then sketch out your research paper. Use big pictures: You may use encyclopedias like Wikipedia to get pictures with the best resolution. At Global Journals, you should strictly follow here.

6. Bookmarks are useful: When you read any book or magazine, you generally use bookmarks, right? It is a good habit which helps to not lose your continuity. You should always use bookmarks while searching on the internet also, which will make your search easier.

7. Revise what you wrote: When you write anything, always read it, summarize it, and then finalize it.

8. Make every effort: Make every effort to mention what you are going to write in your paper. That means always have a good start. Try to mention everything in the introduction—what is the need for a particular research paper. Polish your work with good writing skills and always give an evaluator what he wants. Make backups: When you are going to do any important thing like making a research paper, you should always have backup copies of it either on your computer or on paper. This protects you from losing any portion of your important data.

9. Produce good diagrams of your own: Always try to include good charts or diagrams in your paper to improve quality. Using several unnecessary diagrams will degrade the quality of your paper by creating a hodgepodge. So always try to include diagrams which were made by you to improve the readability of your paper. Use of direct quotes: When you do research relevant to literature, history, or current affairs, then use of quotes becomes essential, but if the study is relevant to science, use of quotes is not preferable.

10.Use proper verb tense: Use proper verb tenses in your paper. Use past tense to present those events that have happened. Use present tense to indicate events that are going on. Use future tense to indicate events that will happen in the future. Use of wrong tenses will confuse the evaluator. Avoid sentences that are incomplete.

11. Pick a good study spot: Always try to pick a spot for your research which is quiet. Not every spot is good for studying.

12. Know what you know: Always try to know what you know by making objectives, otherwise you will be confused and unable to achieve your target.

13. Use good grammar: Always use good grammar and words that will have a positive impact on the evaluator; use of good vocabulary does not mean using tough words which the evaluator has to find in a dictionary. Do not fragment sentences. Eliminate one-word sentences. Do not ever use a big word when a smaller one would suffice.

Verbs have to be in agreement with their subjects. In a research paper, do not start sentences with conjunctions or finish them with prepositions. When writing formally, it is advisable to never split an infinitive because someone will (wrongly) complain. Avoid clichés like a disease. Always shun irritating alliteration. Use language which is simple and straightforward. Put together a neat summary.

14. Arrangement of information: Each section of the main body should start with an opening sentence, and there should be a changeover at the end of the section. Give only valid and powerful arguments for your topic. You may also maintain your arguments with records.

15. Never start at the last minute: Always allow enough time for research work. Leaving everything to the last minute will degrade your paper and spoil your work.

16. Multitasking in research is not good: Doing several things at the same time is a bad habit in the case of research activity. Research is an area where everything has a particular time slot. Divide your research work into parts, and do a particular part in a particular time slot.

17. Never copy others' work: Never copy others' work and give it your name because if the evaluator has seen it anywhere, you will be in trouble. Take proper rest and food: No matter how many hours you spend on your research activity, if you are not taking care of your health, then all your efforts will have been in vain. For quality research, take proper rest and food.

18. Go to seminars: Attend seminars if the topic is relevant to your research area. Utilize all your resources.

19. Refresh your mind after intervals: Try to give your mind a rest by listening to soft music or sleeping in intervals. This will also improve your memory. Acquire colleagues: Always try to acquire colleagues. No matter how sharp you are, if you acquire colleagues, they can give you ideas which will be helpful to your research.

20. Think technically: Always think technically. If anything happens, search for its reasons, benefits, and demerits. Think and then print: When you go to print your paper, check that tables are not split, headings are not detached from their descriptions, and page sequence is maintained.

21. Adding unnecessary information: Do not add unnecessary information like "I have used MS Excel to draw graphs." Irrelevant and inappropriate material is superfluous. Foreign terminology and phrases are not apropos. One should never take a broad view. Analogy is like feathers on a snake. Use words properly, regardless of how others use them. Remove quotations. Puns are for kids, not grunt readers. Never oversimplify: When adding material to your research paper, never go for oversimplification; this will definitely irritate the evaluator. Be specific. Never use rhythmic redundancies. Contractions shouldn't be used in a research paper. Comparisons are as terrible as clichés. Give up ampersands, abbreviations, and so on. Remove commas that are not necessary. Parenthetical words should be between brackets or commas. Understatement is always the best way to put forward earth-shaking thoughts. Give a detailed literary review.

22. Report concluded results: Use concluded results. From raw data, filter the results, and then conclude your studies based on measurements and observations taken. An appropriate number of decimal places should be used. Parenthetical remarks are prohibited here. Proofread carefully at the final stage. At the end, give an outline to your arguments. Spot perspectives of further study of the subject. Justify your conclusion at the bottom sufficiently, which will probably include examples.

23. Upon conclusion: Once you have concluded your research, the next most important step is to present your findings. Presentation is extremely important as it is the definite medium though which your research is going to be in print for the rest of the crowd. Care should be taken to categorize your thoughts well and present them in a logical and neat manner. A good quality research paper format is essential because it serves to highlight your research paper and bring to light all necessary aspects of your research.

Informal Guidelines of Research Paper Writing

Key points to remember:

- Submit all work in its final form.
- Write your paper in the form which is presented in the guidelines using the template.
- Please note the criteria peer reviewers will use for grading the final paper.

Final points:

One purpose of organizing a research paper is to let people interpret your efforts selectively. The journal requires the following sections, submitted in the order listed, with each section starting on a new page:

The introduction: This will be compiled from reference matter and reflect the design processes or outline of basis that directed you to make a study. As you carry out the process of study, the method and process section will be constructed like that. The results segment will show related statistics in nearly sequential order and direct reviewers to similar intellectual paths throughout the data that you gathered to carry out your study.

The discussion section:

This will provide understanding of the data and projections as to the implications of the results. The use of good quality references throughout the paper will give the effort trustworthiness by representing an alertness to prior workings.

Writing a research paper is not an easy job, no matter how trouble-free the actual research or concept. Practice, excellent preparation, and controlled record-keeping are the only means to make straightforward progression.

General style:

Specific editorial column necessities for compliance of a manuscript will always take over from directions in these general guidelines.

To make a paper clear: Adhere to recommended page limits.

Mistakes to avoid:

- Insertion of a title at the foot of a page with subsequent text on the next page.
- Separating a table, chart, or figure—confine each to a single page.
- Submitting a manuscript with pages out of sequence.
- In every section of your document, use standard writing style, including articles ("a" and "the").
- Keep paying attention to the topic of the paper.
- Use paragraphs to split each significant point (excluding the abstract).
- Align the primary line of each section.
- Present your points in sound order.
- Use present tense to report well-accepted matters.
- Use past tense to describe specific results.
- Do not use familiar wording; don't address the reviewer directly. Don't use slang or superlatives.
- Avoid use of extra pictures—include only those figures essential to presenting results.

Title page:

Choose a revealing title. It should be short and include the name(s) and address(es) of all authors. It should not have acronyms or abbreviations or exceed two printed lines.

Abstract: This summary should be two hundred words or less. It should clearly and briefly explain the key findings reported in the manuscript and must have precise statistics. It should not have acronyms or abbreviations. It should be logical in itself. Do not cite references at this point.

An abstract is a brief, distinct paragraph summary of finished work or work in development. In a minute or less, a reviewer can be taught the foundation behind the study, common approaches to the problem, relevant results, and significant conclusions or new questions.

Write your summary when your paper is completed because how can you write the summary of anything which is not yet written? Wealth of terminology is very essential in abstract. Use comprehensive sentences, and do not sacrifice readability for brevity; you can maintain it succinctly by phrasing sentences so that they provide more than a lone rationale. The author can at this moment go straight to shortening the outcome. Sum up the study with the subsequent elements in any summary. Try to limit the initial two items to no more than one line each.

Reason for writing the article—theory, overall issue, purpose.

- Fundamental goal.
- To-the-point depiction of the research.
- Consequences, including definite statistics—if the consequences are quantitative in nature, account for this; results of any numerical analysis should be reported. Significant conclusions or questions that emerge from the research.

Approach:

- Single section and succinct.
- An outline of the job done is always written in past tense.
- Concentrate on shortening results—limit background information to a verdict or two.
- Exact spelling, clarity of sentences and phrases, and appropriate reporting of quantities (proper units, important statistics) are just as significant in an abstract as they are anywhere else.

Introduction:

The introduction should "introduce" the manuscript. The reviewer should be presented with sufficient background information to be capable of comprehending and calculating the purpose of your study without having to refer to other works. The basis for the study should be offered. Give the most important references, but avoid making a comprehensive appraisal of the topic. Describe the problem visibly. If the problem is not acknowledged in a logical, reasonable way, the reviewer will give no attention to your results. Speak in common terms about techniques used to explain the problem, if needed, but do not present any particulars about the protocols here.

The following approach can create a valuable beginning:

- Explain the value (significance) of the study.
- Defend the model—why did you employ this particular system or method? What is its compensation? Remark upon its appropriateness from an abstract point of view as well as pointing out sensible reasons for using it.
- Present a justification. State your particular theory(-ies) or aim(s), and describe the logic that led you to choose them.
- Briefly explain the study's tentative purpose and how it meets the declared objectives.

Approach:

Use past tense except for when referring to recognized facts. After all, the manuscript will be submitted after the entire job is done. Sort out your thoughts; manufacture one key point for every section. If you make the four points listed above, you will need at least four paragraphs. Present surrounding information only when it is necessary to support a situation. The reviewer does not desire to read everything you know about a topic. Shape the theory specifically—do not take a broad view.

As always, give awareness to spelling, simplicity, and correctness of sentences and phrases.

Procedures (methods and materials):

This part is supposed to be the easiest to carve if you have good skills. A soundly written procedures segment allows a capable scientist to replicate your results. Present precise information about your supplies. The suppliers and clarity of reagents can be helpful bits of information. Present methods in sequential order, but linked methodologies can be grouped as a segment. Be concise when relating the protocols. Attempt to give the least amount of information that would permit another capable scientist to replicate your outcome, but be cautious that vital information is integrated. The use of subheadings is suggested and ought to be synchronized with the results section.

When a technique is used that has been well-described in another section, mention the specific item describing the way, but draw the basic principle while stating the situation. The purpose is to show all particular resources and broad procedures so that another person may use some or all of the methods in one more study or referee the scientific value of your work. It is not to be a step-by-step report of the whole thing you did, nor is a methods section a set of orders.

Materials:

Materials may be reported in part of a section or else they may be recognized along with your measures.

Methods:

- Report the method and not the particulars of each process that engaged the same methodology.
- Describe the method entirely.
- To be succinct, present methods under headings dedicated to specific dealings or groups of measures.
- Simplify—detail how procedures were completed, not how they were performed on a particular day.
- If well-known procedures were used, account for the procedure by name, possibly with a reference, and that's all.

Approach:

It is embarrassing to use vigorous voice when documenting methods without using first person, which would focus the reviewer's interest on the researcher rather than the job. As a result, when writing up the methods, most authors use third person passive voice.

Use standard style in this and every other part of the paper—avoid familiar lists, and use full sentences.

What to keep away from:

- Resources and methods are not a set of information.
- Skip all descriptive information and surroundings—save it for the argument.
- Leave out information that is immaterial to a third party.

Results:

The principle of a results segment is to present and demonstrate your conclusion. Create this part as entirely objective details of the outcome, and save all understanding for the discussion.

The page length of this segment is set by the sum and types of data to be reported. Use statistics and tables, if suitable, to present consequences most efficiently.

You must clearly differentiate material which would usually be incorporated in a study editorial from any unprocessed data or additional appendix matter that would not be available. In fact, such matters should not be submitted at all except if requested by the instructor.

Content:

- Sum up your conclusions in text and demonstrate them, if suitable, with figures and tables.
- In the manuscript, explain each of your consequences, and point the reader to remarks that are most appropriate.
- Present a background, such as by describing the question that was addressed by creation of an exacting study.
- Explain results of control experiments and give remarks that are not accessible in a prescribed figure or table, if appropriate.
- Examine your data, then prepare the analyzed (transformed) data in the form of a figure (graph), table, or manuscript.

What to stay away from:

- Do not discuss or infer your outcome, report surrounding information, or try to explain anything.
- Do not include raw data or intermediate calculations in a research manuscript.
- Do not present similar data more than once.
- A manuscript should complement any figures or tables, not duplicate information.
- Never confuse figures with tables—there is a difference.

Approach:

As always, use past tense when you submit your results, and put the whole thing in a reasonable order.

Put figures and tables, appropriately numbered, in order at the end of the report.

If you desire, you may place your figures and tables properly within the text of your results section.

Figures and tables:

If you put figures and tables at the end of some details, make certain that they are visibly distinguished from any attached appendix materials, such as raw facts. Whatever the position, each table must be titled, numbered one after the other, and include a heading. All figures and tables must be divided from the text.

Discussion:

The discussion is expected to be the trickiest segment to write. A lot of papers submitted to the journal are discarded based on problems with the discussion. There is no rule for how long an argument should be.

Position your understanding of the outcome visibly to lead the reviewer through your conclusions, and then finish the paper with a summing up of the implications of the study. The purpose here is to offer an understanding of your results and support all of your conclusions, using facts from your research and generally accepted information, if suitable. The implication of results should be fully described.

Infer your data in the conversation in suitable depth. This means that when you clarify an observable fact, you must explain mechanisms that may account for the observation. If your results vary from your prospect, make clear why that may have happened. If your results agree, then explain the theory that the proof supported. It is never suitable to just state that the data approved the prospect, and let it drop at that. Make a decision as to whether each premise is supported or discarded or if you cannot make a conclusion with assurance. Do not just dismiss a study or part of a study as "uncertain."

Research papers are not acknowledged if the work is imperfect. Draw what conclusions you can based upon the results that you have, and take care of the study as a finished work.

- You may propose future guidelines, such as how an experiment might be personalized to accomplish a new idea.
- Give details of all of your remarks as much as possible, focusing on mechanisms.
- Make a decision as to whether the tentative design sufficiently addressed the theory and whether or not it was correctly restricted. Try to present substitute explanations if they are sensible alternatives.
- One piece of research will not counter an overall question, so maintain the large picture in mind. Where do you go next? The best studies unlock new avenues of study. What questions remain?
- Recommendations for detailed papers will offer supplementary suggestions.

Approach:

When you refer to information, differentiate data generated by your own studies from other available information. Present work done by specific persons (including you) in past tense.

Describe generally acknowledged facts and main beliefs in present tense.

The Administration Rules

Administration Rules to Be Strictly Followed before Submitting Your Research Paper to Global Journals Inc.

Please read the following rules and regulations carefully before submitting your research paper to Global Journals Inc. to avoid rejection.

Segment draft and final research paper: You have to strictly follow the template of a research paper, failing which your paper may get rejected. You are expected to write each part of the paper wholly on your own. The peer reviewers need to identify your own perspective of the concepts in your own terms. Please do not extract straight from any other source, and do not rephrase someone else's analysis. Do not allow anyone else to proofread your manuscript.

Written material: You may discuss this with your guides and key sources. Do not copy anyone else's paper, even if this is only imitation, otherwise it will be rejected on the grounds of plagiarism, which is illegal. Various methods to avoid plagiarism are strictly applied by us to every paper, and, if found guilty, you may be blacklisted, which could affect your career adversely. To guard yourself and others from possible illegal use, please do not permit anyone to use or even read your paper and file.

Criterion for Grading a Research Paper (Compilation) by Global Journals Inc. (US)

Please note that following table is only a Grading of "Paper Compilation" and not on "Performed/Stated Research" whose grading solely depends on Individual Assigned Peer Reviewer and Editorial Board Member. These can be available only on request and after decision of Paper. This report will be the property of Global Journals Inc. (US).

Topics	Grades		
	A-B	**C-D**	**E-F**
Abstract	Clear and concise with appropriate content, Correct format. 200 words or below	Unclear summary and no specific data, Incorrect form Above 200 words	No specific data with ambiguous information Above 250 words
Introduction	Containing all background details with clear goal and appropriate details, flow specification, no grammar and spelling mistake, well organized sentence and paragraph, reference cited	Unclear and confusing data, appropriate format, grammar and spelling errors with unorganized matter	Out of place depth and content, hazy format
Methods and Procedures	Clear and to the point with well arranged paragraph, precision and accuracy of facts and figures, well organized subheads	Difficult to comprehend with embarrassed text, too much explanation but completed	Incorrect and unorganized structure with hazy meaning
Result	Well organized, Clear and specific, Correct units with precision, correct data, well structuring of paragraph, no grammar and spelling mistake	Complete and embarrassed text, difficult to comprehend	Irregular format with wrong facts and figures
Discussion	Well organized, meaningful specification, sound conclusion, logical and concise explanation, highly structured paragraph reference cited	Wordy, unclear conclusion, spurious	Conclusion is not cited, unorganized, difficult to comprehend
References	Complete and correct format, well organized	Beside the point, Incomplete	Wrong format and structuring

Index

save our planet

Global Journal of Computer Science and Technology

Visit us on the Web at www.GlobalJournals.org | www.ComputerResearch.org
or email us at helpdesk@globaljournals.org

ISSN 9754350