

Next Generation Data Warehouse Design with Big Data for Big Analytics and Better Insights

P Renjith Kumar¹ and Lt. Dr. Santhosh Baboo²

¹ Manonmaniam Sundaranar University, Tirunelveli - 627012

Received: 10 December 2012 Accepted: 2 January 2013 Published: 15 January 2013

Abstract

Traditionally organizations invested more in decision support systems. With the evolution of business intelligence tools many organizations were able to get analytical reports based on OLAP systems. Now with the frequently changing trends in customer behaviour and customer markets there is a huge necessity for enterprises to get analytical reports beyond OLAP system based analysis. There is huge innovation in the area of hardware and software which helps enterprises to gain advantage of all available formats of data and help enterprise to get business insights based on that data. Big data is one of the key factors to be focused which can help to get real time analytics on all available formats of data. This document presents the overview of the next generation data warehouse architecture based on Big data for better business insights.

Index terms— big data, big data analytics, business intelligence, real time analytics.

1 Introduction

Organizations are now generating and recording huge volume of data. There are many business intelligence tools available which could extract these data and provide analytical reports for better decision making. But the business is changing now, consumers buying behaviours are changing rapidly, the market is unpredictable which all makes a necessity to further make use of all forms of data that is available. The organizations in future will succeed if its executives can reliably forecast the future demand and based on that effectively assess the alternative business strategies and implement them with optimal technology and right business solutions. Many enterprises now record terabytes, peta bytes of data. These data are left unanalyzed. With the huge growth in the internet based enterprises and availability of many social networks there is huge volume of data that is available for analysis. Though there are enough business intelligence and reporting tools available for analytical reporting purpose these data are not considered for the analytics purpose.

When we think about these unused data now it creates a Question.

Author ¹ : Reader, P.G. and Research, Dept. of Computer Science, D.G. Vaishnav College, Chennai - 600106. E-mail : santhos2001@sify.com Author ² : Research scholar, Computer Science Manonmaniam Sundaranar University, Tirunelveli -627012, SAP Business Intelligence Consultant, SAP Labs India. E-mail : Renjith.sriram@gmail.com "Will it be possible to have analytical reports based on these Big data, is there any scope for Big data analytics which could provide further insights into the business and help organizations to utilize the full potential of data that is ;abundantly available?" If it is possible to achieve this type of analytical reporting based on the Big data as a business intelligence tool it will be a step towards next generation data warehouse design and this will enable business and users to react to business events more quickly through the insights provided by big data and with availability of real-time analysis over big data will help management to take analytical decision with innovative ideas and critical business decisions in real time.

2 II.

3 Introduction to Big Data

Your Gartner defines big data as below [1] "Big data is high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making".

When termed about variety companies are digging out amazing insights from text, locations or log files. Elevator logs help to predict vacated real estate, shoplifters tweet about stolen goods right next to the store, emails contain communication patterns of successful projects. Most of this data already belongs to organizations, but it is sitting there unused -that's why Gartner calls it dark data. Similar to dark matter in physics, dark data cannot be seen directly, yet it is the bulk of the organizational universe. [2] When termed about velocity it is frequently equated to real time analytics. Yet, velocity is also about the rate of changes, about linking data sets that are coming with different speeds and about bursts of activities [2] Volume is about the number of big data mentions in the social media [2] Every day, enterprisers create 2.5 quintillion bytes of data so much that 90% of the data in the world today has been created in the last two years alone. This data comes from everywhere: sensors used to gather climate information, posts to social media sites, digital pictures and videos, purchase transaction records, and cell phone GPS signals to name a few. This data is big With the evolution of business intelligence tools many organizations were able to get analytical reports based on OLAP systems. Now with the frequently changing trends in customer behaviour and customer markets there is a huge necessity for enterprises to get analytical reports beyond OLAP system based analysis. There is huge innovation in the area of hardware and software which helps enterprises to gain advantage of all available formats of data and help enterprise to get business insights based on that data. Big data is one of the key factors to be focused which can help to get real time analytics on all available formats of data. This document presents the overview of the next generation data warehouse architecture based on Big data for better business insights.

data [3]. Hence Big data is more than simply a matter of size; it is an opportunity to find insights in new and emerging types of data and content, to make your business more agile and to answer questions that were previously considered beyond your reach. Until now, there was no practical way to harvest this opportunity [3] Just how big a phenomenon big data actually is was eloquently captured in a remark by Google's Eric Schmidt. He pointed out that we are creating as much information every two days as we did from the dawn of civilization up until 2003. On a daily basis, this translates into around 2.5 Exabyte's of data. [4] Figure ?? : Daily Amount of Data Created [5] Upon checking the Facebook for the data manipulated per day Facebook recently gave a look into the massive amount of data that the company has to deal with on a daily basis [6] Figure ?? : Facebook data per day [7] III.

4 Big Data Analytics

Big Data is everywhere, from sensors that monitor and manage traffic loads to the flood of tweets and Facebook "likes." But how do organizations make sense of the mountain of structured and unstructured data that now shape how world work, live and play? More importantly, how do they sort through a maze of information in different data structures, formats and sources that traditionally was presented in mindnumbing tabular formats and one dimensional pie charts? [8] Big data analytics is where advanced analytic techniques operate on big data sets. Hence, big data analytics is really about two things-big data and analytics-plus how the two have teamed up to create one of the most profound trends in business intelligence (BI) today. Let's start by defining advanced analytics, then move on to big data and the combination of the two. [9] New and evolving analytical processing technologies now make possible what was not possible before. Examples include: [10] ? New systems that handle a wide variety of data from sensor data to web and social media data. Improved analytical capabilities (sometimes called advanced analytics) including event, predictive and text analytics. ? Operational business intelligence that improves business agility by enabling automated real-time actions and intraday decision making. ? Faster hardware ranging from faster multi-core processors and large memory spaces, to solidstate drives and virtual data storage for handling hot and cold data. Cloud computing including on-demand software-as-a service (SaaS) analytical solutions in public clouds and data platforms and virtualization in private clouds.

The architecture of the big data will be like this given below:

? The data can be extracted from multiple sources such as unstructured data across net, Internet based data and various data from social networking sites etc. ? The data is stored and cleansed as the same process of the traditional data warehouse system. ? The big data is then used for the analytical reporting purpose. ? The data is taken from within and outside of the organization ? When there is an system that is optimized for these kind of acquiring, cleansing and storing it will be very effective. IV.

Need for Big Data Analytics

Companies are capturing and digitizing more information than ever before. According to IDC, the world produced approximately 1.8 zetta byte's of data in 2011. Fuelling this data explosion are over 5 billion mobile phones, 30 billion pieces of content shared on Facebook per month, 20 billion internet searches per month, and millions of networked sensors connected to mobile phones, energy meters, automobiles, shipping containers, retail packaging and more. As the volume of data continues to grow, businesses struggle to get the right information

at the right time to remain relevant. Big data is a platform for transforming all of this data into action able intelligence for business decision making [11].

Between now and 2020, the sheer volume of digital information is predicted to increase to 35 trillion gigabytes -much of it coming from new sources including blogs, social media, internet search, and sensor networks. [12] The way organizations do business is changing rapidly much business try to predict the frequently changing consumer behaviour. With the huge growth in internet and huge volume of data accumulated through the social networking sites makes it an advantage for large enterprise to drill down to get more detailed analytical reports.

Big data can and will impact every nation, industry, company and individual around the globe, whether it's in terms of understanding our galaxy, optimizing healthcare, selecting an ideal retail location or finding the perfect date.

A study by McKinsey Global Institute estimates that big data can add \$300 billion worth of value to the US healthcare system and can increase retailers' operating margins by as much as 60%.[13] There is no doubt that those who collect, analyze and act on their data success fully will gain a competitive advantage in their market.

V.

5 Big Data Analytics Adoption

In March 2012, The White House announced a national "Big Data Initiative" that consisted of six Federal

The initiative included a National Science Foundation "Expeditions in Computing" grant of \$10 million over 5years to the AMPLab [15] at the University of California, Berkeley.[16] The AMPLab also received funds from DARPA, and over a dozen industrial sponsors and uses big data to attack a wide range of problems from predicting traffic congestion[17] to fighting cancer. [18] The White House Big Data Initiative also included a commitment by the Department of Energy to provide \$25million in funding over 5 years to establish the Scalable Data Management, Analysis and Visualization (SDAV)Institute,[19] led by the Energy Department's Lawrence Berkeley National Laboratory. The SDAV Institute aims to bring together the expertise of six national laboratories and seven universities to develop new tools to help scientists manage and visualize data on the Department's supercomputers.

The U.S. state of Massachusetts announced the Massachusetts Big Data Initiative in May 2012, which provides funding from the state government and private companies to a variety of research institutions.[20] The Massachusetts Institute of Technology hosts the Intel Science and Technology Center for Big Data in the MIT Computer Science and Artificial Intelligence Laboratory, combining government, corporate, and institutional funding and research efforts. [21] The European Commission is funding a 2-yearlong Big Data Public Private Forum through their Seventh Framework Program to engage companies, academics and other stakeholders in discussing Big Data issues. The project aims to define a strategy in terms of research and innovation to guide supporting actions from the European Commission in the successful implementation of the Big Data economy. Outcomes of this project will be used as input for Horizon2020, their next framework program. [22] VI.

6 Advantages of Big Data

The traditional analytics system have been fetching data from the source system such as ERP and CRM system, many analytics query will involve the aggregated form of data like "Sales in the north region during the year" which will be helpful in making future analytical decision. But now there is need for much better analytics with changing customer behaviour there is need for organizations to do predictive analysis, there is a need to go beyond the general OLAP analysis, organizations now need systems that can fully utilize the potential of hardware innovations such as In-Memory computing and In-Memory analytics. With the availability of mass storage devices it is possible to have Big data analytics for any enterprises.

Big Data analytics not only enable organizations to quickly access and track the ongoing stream of Year posts. It also empowers decision makers with the deepdive capability they need to accurately query and analyze the opinions shared by their organization's customers, suppliers, stakeholders, shareholders, and employees [23] Probabilistic logic could become instrumental in using Big Data analytics because it can help organization slever age the probabilities of future events that the software predicts. By adopting a probabilistic mindset, organizations can gain flexibility that is not possible when using a cause and-effect (objective logic) approach [24] Organizations can no longer focus solely on delivering the best product or service. To succeed they must uncover hidden customer, employee, vendor and partner trends and in sights. Organizations need to anticipate behavior and then take proactive action and empower the team with intelligent next steps to exceed customer expectations. Predictive analysis helps organizations to achieve real time in sights that increase the understanding of customer behavior, improve response to customer and deliver tangible business value to customer which ultimately drive profitability in business. [24] VII.

7 Real Time Analytics

With the huge volume of data across all platforms generally it will impact the performance of any system. Hence performance is an important consideration in building the business intelligence systems that use Big data for real time analytics. But With the availability of high speed multi-core processors and dual core processors, 64 bit processors and the availability of main memory for less price in the market along with the emergence of new

generation software that has ability to run on effectively on these hardware platforms enable organizations to think about real time analytics based out of big data. It will be the next big shift in the business intelligence field. When enterprises have In Memory database with 64 bit processors along with good servers with better performance and storage considerations the real time analytics based on the big data can be easily achieved.

VIII.

9 Future Research

Cloud computing is one of the emerging technologies. A comprehensive study on using Big data over cloud data warehouse will be promising. There are various challenges in storing the Big data currently hence a study on better storage option for Big data will be good future research. The usage of big data, text analytics and unstructured data for the data warehousing needs to be deeply studied along with integration with cloud in the future research.

10 IX.

11 Conclusion

Today's business needs real time analytics to succeed in the market there are many business intelligence tools that help organizations to extract data from various source system and provide analytical reports based on that. But with the changing consumer behaviour and unpredictable markets it is tough to decide on a good business strategy. The best business decision can be made only if there is the knowledge based on all available data that any organizations have among that big data plays a vital role. When the business intelligence and data warehousing systems can be well equipped to handle these huge volume of unused bigdata it will provide further deeper insights into business and help the enterprise to make better decision making. When there is enough real time analytical data along with there ports based on big data any enterprises will make better business decision to make their business to run better.



Figure 1: C

¹© 2013 Global Journals Inc. (US) Global Journal of Computer Science and Technology

²© 2013 Global Journals Inc. (US)

³Cdepartments and agencies committing more than \$200 million to big data research projects.[14]

34

2.5B - content items shared
2.7B - 'Likes'
300M - photos uploaded
100+PB - disk space in a single HDFS cluster
105TB - data scanned via Hive (30min)
70,000 - queries executed
500+TB - new data ingested

Figure 2: Figure 3 :CFigure 4 :

183 [Parikh] , Jay Parikh , Facebook . http://news.cnet.com/8301-1023_3-57498531-93/
184 [facebook-processes-more-than-500-tb-of-data-daily/](http://news.cnet.com/8301-1023_3-57498531-93/)

185 [Zaharia] , Matei Zaharia .

186 [Ma] , Justin Ma .

187 [Franklin] , Michael Franklin .

188 [Abbeel] , Pieter Abbeel .

189 [Bigdata] , Bigdata . csail.mit.edu.2013-02 22. [http://bigdata.csail.mit.edu/22.Cordis.europa.](http://bigdata.csail.mit.edu/22.Cordis.europa.eu.2012-09-01http://cordis.europa.eu/search/index.cfm?fuseaction=proj.document&PJ_RCN=13267529)
190 [eu.2012-09-01http://cordis.europa.eu/search/index.cfm?fuseaction=proj.document&](http://cordis.europa.eu/search/index.cfm?fuseaction=proj.document&PJ_RCN=13267529)
191 [PJ_RCN=13267529](http://cordis.europa.eu/search/index.cfm?fuseaction=proj.document&PJ_RCN=13267529)

192 [Bayen (2011)] , Alexandre Bayen . [https://amplab.cs.berkeley.edu/publication/](https://amplab.cs.berkeley.edu/publication/scalingthemobile-millennium-system-in-the-cloud-2/)
193 [scalingthemobile-millennium-system-in-the-cloud-2/](https://amplab.cs.berkeley.edu/publication/scalingthemobile-millennium-system-in-the-cloud-2/) October 2011.

194 [Patterson (2011)] ‘://energy.gov/ articles/secretary-chu-announcesnew-institute-helpscientists-improve-
195 massive-data-setresearch-doe 20’. David Patterson . [http://www.mass.gov/governor/pressoffice/](http://www.mass.gov/governor/pressoffice/pressreleases/2012/2012530-governor-announces-big-datainitiative.html)
196 [pressreleases/2012/2012530-governor-announces-big-datainitiative.html](http://www.mass.gov/governor/pressoffice/pressreleases/2012/2012530-governor-announces-big-datainitiative.html) *The New*
197 *YorkTimes* “Computer Scientists May Have What It Takes to Help Cure Cancer”, 5 December 2011. March
198 29, 2012http. May 30, 2012. Commonwealth of Massachusetts.

199 [Khera] *Data Visualization Market*[http://mars commons.marsdd.com/the-datavisualizers/market](http://marscommons.marsdd.com/the-datavisualizers/market), Neha Khera .

200 [Sicular (2013)] *Gartner’s BigData Definition Consists of Three Parts, Not to Be Confused with*
201 *Three*, Svetlana Sicular . [http://www.forbes.com/sites/gartnergroup/2013/03/27/](http://www.forbes.com/sites/gartnergroup/2013/03/27/gartners-big-data-definition-consists-of-three-parts-notto-be-confused-with-three-vs/)
202 [gartners-big-data-definition-consists-of-three-parts-notto-be-confused-with-three-vs/](http://www.forbes.com/sites/gartnergroup/2013/03/27/gartners-big-data-definition-consists-of-three-parts-notto-be-confused-with-three-vs/)
203 [Mar/2013](http://www.forbes.com/sites/gartnergroup/2013/03/27/gartners-big-data-definition-consists-of-three-parts-notto-be-confused-with-three-vs/). Gartner, Inc.

204 [Perkowski] *McKinsey Global Institute: Big data: The next frontier for innovation, competition and productivity*,
205 Mike Perkowski . 13.

206 [Obama Administration Unveils “Big Data” Initiative: Announces 200MillionInNewRDInvestments]
207 http://www.nsf.gov/news/news_summ.jsp?cntn_id=123607&org=NSF&from=news17 *Obama*
208 *Administration Unveils “Big Data” Initiative: Announces \$200 Million In New R & D Investments*,
209 (Timothy Hunter) (Teodor Moldovan)

210 [Sap and Hana] *Solution brief, SAP, Transform your future with better business insights using predictive analytics*,
211 Sap , Hana .

212 [TechCrunch: Eric Schmidt: Every 2 Days We Create As Much Information As We Did Up To ()]
213 *TechCrunch: Eric Schmidt: Every 2 Days We Create As Much Information As We Did Up To*,
214 <http://techcrunch.com/2010/08/04/schmidt-data> 2003.

215 [Dennis (2013)] *The SoftBenefit of Big Data Analytics*, Joseph Dennis ,
216 Kelly . [Differentlyhttp://blogs.sap.com/innovation/analytics/](http://blogs.sap.com/innovation/analytics/the-softbenefit-of-big-data-analytics-thinking-differently-025327)
217 [the-softbenefit-of-big-data-analytics-thinking-differently-025327](http://blogs.sap.com/innovation/analytics/the-softbenefit-of-big-data-analytics-thinking-differently-025327) January 14, 2013.