



## International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 10, October 2015

# NextGen Big DWH: Big Data Oriented Data Warehouse Architecture for Improved Business Intelligence

Gerard Deepak<sup>1</sup>, Reevan Chris Pinto<sup>2</sup>

M.E. Scholar, Dept. of Computer Science and Engineering, UVCE, Bangalore University, Bangalore, India<sup>1</sup>

B.E. Student, Dept. of Computer Science and Engineering, UVCE, Bangalore University, Bangalore, India<sup>2</sup>

**ABSTRACT:** Data Warehousing is highly essential for achieving Business Intelligence in an Enterprise. A traditional Data Warehouse is built in par with the Inmon's Architecture which follows Extract, Transform and Load (ETL) strategy for data pre-processing and Online Analytical Processing (OLAP) for Analysis. With several recent trends like the Online Social Networks (OSNs), e-commerce and increasing number of internet users, the amount of data has risen exponentially. The Data is highly dynamic where existing Data Warehouse Architectures are unable to keep in par with large amount of data for processing. Though the ETL strategy performs fairly well, it consumes a lot of time for real-time data processing. To enhance the processing capability of large volumes of Data, several Big Data Technologies and frameworks are introduced. In this paper, a Big Data Oriented Data Warehouse Architecture is proposed where the Big Data Technologies are accommodated in the Data Warehouse Architecture in a highly logical manner with an essence of chronological arrangement of the Big Data technologies. A detailed Empirical Evaluation of the proposed architecture is conducted based on a survey involving big data experts in order to validate the proposed Data Warehouse Architecture incorporating Big Data Technologies. Incorporation of Intelligent and Semantic agents is also achieved for customizing and making the Analysis of Enterprise Level data more efficient and in turn paving a way for improved Business Intelligence at the Enterprise Level.

**KEYWORDS:** Big Data, Big Data Oriented Data Warehouse, Business Intelligence, DWH, Data Warehouse, Data Warehouse Architecture, NextGen Data Warehouse Architecture

### I. INTRODUCTION

The present age is the information age where information is in the highest demand. Gone are the days where data was considered as the utmost commodity and whose representations were given the highest importance. Today the most important Commodity for an Enterprise is Knowledge to drive Business. Of course without data there would be no information and without information there would be no knowledge. Creation of Knowledge bases and engineering of intelligent agents for extracting knowledge from data must be given prime importance in this era of Knowledge and Business Intelligence. There is a need for proper analytics strategy to incorporate Business Intelligence into a particular Business for an Enterprise. The Business Intelligence is directly proportional to the Marketing Intelligence which is a mandate to drive a business into a phenomenal new level. The Business Intelligence added to a Business gives an entirely new direction and better is the outcome when compared to driving a business without any predictions and analysis.

Data Warehousing was introduced to achieve Business Intelligence by converting the day-to-day Transactional Data or Operational Data into a highly summarized Analytical Data which in turn needs to be subject to analytical tools for business predictions. Data Warehousing served and still serves as a highly significant technology for achieving significant robustness in business predictions and thus paving a pathway for imbibing Business Intelligence. The business strategy has taken a major turn in the recent few years where e-commerce and m-commerce has made it big in the business front. Traditional Marketing Strategies have also evolved into Digital Marketing Techniques with Digitalization making its debut in the technology front. With the introduction of Marketing Techniques like Viral



# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 10, October 2015

Marketing and Online Marketing there is a huge shift in paradigm in the perspective of shopping, retailing and Business. The entire Business Trends have gone to the “Web” which generates a large amount of data every second. The Data in the World Wide Web increases exponentially owing to the large number of users in the Web.

There is an urgent need for Management of such huge volumes of data that is generated every second over the Web. The data is highly sophisticated, scattered and randomized. Some amount of data may be structured but a large amount of unstructured data is generated. Data that is generated is of a high degree of variation and it ranges from spatial to multimedia, from text to numeric, from audio to video, from web pages to web documents. All these different formats of data need efficient processing techniques for proper analysis. All the unwanted data which can be categorized as noise needs to be filtered. In the previous decade, data was limited in the absence of business trends like e-commerce and online social networks. The OLAP (Online Analytical Processing) strategy was thoroughly sufficient for Data Processing and building a Data Warehouse which catered to users from different perspectives to analyze and use the data. Since there is an exponential rise in the data every second, traditional OLAP is insufficient for data processing and constructing a Data Warehouse.

With introduction of Big Data and Analytics for data analysis, Data Warehouse can be constructed at a quicker and a faster rate when compared to the Traditional ETL methodologies owing to the amount of unstructured data that are available today due to increased number of users. To support this, business trends have moved online. According to Bill Inmon, Big Data can never replace a Data Warehouse as it's a myth which states that Big Data has over ridden a Data Warehouse as Data Warehouse is architecture while Big Data is a Technology. Data Warehouse mainly concentrates on Analytics and Business Intelligence where as Big data focuses on processing large amounts of data in terms of volume at a faster rate. Big Data can be incorporated into a Data Warehouse for enhancing the performance of the Warehouse and it can never be an overriding factor for Data Warehouse.

**Motivation:** Since the data is exponentially increasing every second owing to the number of users as well as the increased business trends over the World Wide Web, a need arises for a faster and much more efficient technology for large scale data processing. Data Warehouse uses ETL for data pre-processing which slows down the entire phenomenon of pre-processing the data in the present day times of high data density. In order to overcome the slow speed of data processing in the conventional ETL transformations, the concept of integration of Big Data Technologies is proposed for a Data Warehouse.

**Contribution:** The architecture for Data Warehouse that integrates Big Data Technologies is proposed and is compared with the existing Data Warehouse architecture. The Big data technology is integrated into the existing Data Warehouse and an Empirical Analysis of the new Big Data Oriented Data Warehouse Architecture is conducted. Finally, a survey is conducted where professionals from the Data Warehousing domain, Business Intelligence Domain, ERP Domain and Big Data Domain were questioned about their opinion about incorporating Big Data Analytics into the Data Warehouse. The Survey results are evaluated in order to prove that the proposed Architecture is valid. A methodology for incorporating Agent Technologies and Semantic Agents for extracting useful patterns from the Warehouse is also included in the architecture proposed.

**Organization:** This paper is organized into 5 main sections. Section 2 provides an overview of related research works. Section 3 presents the Proposed Architecture. The empirical justifications are documented in the section 4. The paper is concluded in section 5.

## II. RELATED WORK

Vijay Kumar et al., [1] have formulated a detailed study and analysis of the basic data warehouse architecture where the two major styles of a Data Warehouse is studied namely the Bill Inmon's style of Data Warehouse and the Ralph Kimball's style. Also an approach towards enterprise oriented Business Intelligence is discussed in the paper. Indira Gandhi et al., [2] have studied in detail the Management and Challenges involved in Big Data and put forth the urgent need for a new Enterprise Level Architecture for Big Data Oriented Data Processing. They have surveyed in detail the various sources of big data and have proposed a valid life cycle model for data.



# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 10, October 2015

Lakshmi et al., [3] have related Big data and Business Intelligence and have conducted a comprehensive study considering real time use cases of Business Intelligence. They have also proposed various methodologies for Big Data processing in a Data Warehouse Platform for an Enterprise to achieve a high degree of Business Intelligence. Ramandeep et al., [4] have investigated the need for Database Management System, Data Warehousing and Data Mining for inferring the appropriate need of these technologies for driving the business needs of an organization flawlessly and intrinsically.

W.H. Inmon, [5] gives a clear cut idea of what a Data Warehouse is and How Data Mining is done over a data Warehouse. He puts forth OLAP and Data Summarization and formally defines a Data Warehouse. Dittrich et al., [6] have proposed a strategy for processing Big Data using Hadoop and MapReduce and have achieved faster processing of a large data cluster and have also studied in detail about parallelism of Data Processing for achieving a faster processing capability.

### III. PROPOSED ARCHITECTURE

According to Bill Inmon, the father of Data Warehousing, a Data Warehouse is irreplaceable especially by the Big Data Technologies. Standalone Big Data Technologies cannot replace a Data Warehouse as Data Warehouse is an architecture where as Big Data is a Technology [7]. Similarly there are new allegations that the Data Lake has already replaced the Data Warehouse but the truth is that the Enterprise Wide Data Warehouse cannot be replaced but a brand new architecture is needed for the Data Warehouse which is an extension and enhancement of the Bill Inmon Architecture. The newly proposed architecture *NextGen Big DWH* that is depicted in Figure 1 clearly incorporates and integrates the Big Data Technologies into the Data Warehouse Architecture for an Enterprise in order to get the best in class performance of Data processing and achieve the best results for catering to the informational and knowledge needs of several classes of users who depend on the Data Warehouse for useful information.

Tamara Dull clearly compares the Data Warehouse and Data Lakes and concludes that both the Data Warehouse and the Data Lake are Data Storage Repositories [8] but a Data Warehouse cannot become a Data Lake and a Data Lake cannot be equated to a Data Warehouse. They are two separate entities which cannot be replaced. Off late both Data Warehouse and Data Lakes are necessary at an Enterprise as they have their own functionalities. Henceforth the conclusions drawn by Bill Inmon and Tamara Dull serve as the basis for the proposition of the new architecture for a Data Warehouse where the Big Data Technologies are integrated at an Enterprise Level. Owing to this fact, this new architecture can be called as "*Big Data Oriented NextGen Data Warehouse Architecture*" depicted in the Figure 1. There is a clear cut demarcation between the existing architecture of the data Warehouse as well as the Technologies of Big Data. Though there exists a differentiation between the existing Warehouse architecture and the Big Data Technologies, a loosely coupled integration marks the characteristic feature of the proposed architecture.

The proposed architecture for the Data Warehouse targets in chronological arrangement of the ten most viable Big Data technologies and including them to the existing Data Warehouse Architecture. The proposed architecture stands out because it engulfs both the ETL technologies as well as Big Data Technologies for Data Processing. It also allows Aggregation of data from Data Lake to the Data Warehouse in order to make the data more available to all the classes of users. The architecture is sub-divided into two major phases, Phase 1 architecture has no changes, ie, it is the existing Inmon Architecture of a Data Warehouse. The Phase 2 architecture is the entire Big Data Technologies that are arranged in a correct chronological order to facilitate big data analytics. The Phases 1 and 2 of the Data Warehouse Architectures are Loosely Coupled and are integrated to provide a single large environment for the Data Warehouse.

# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 10, October 2015

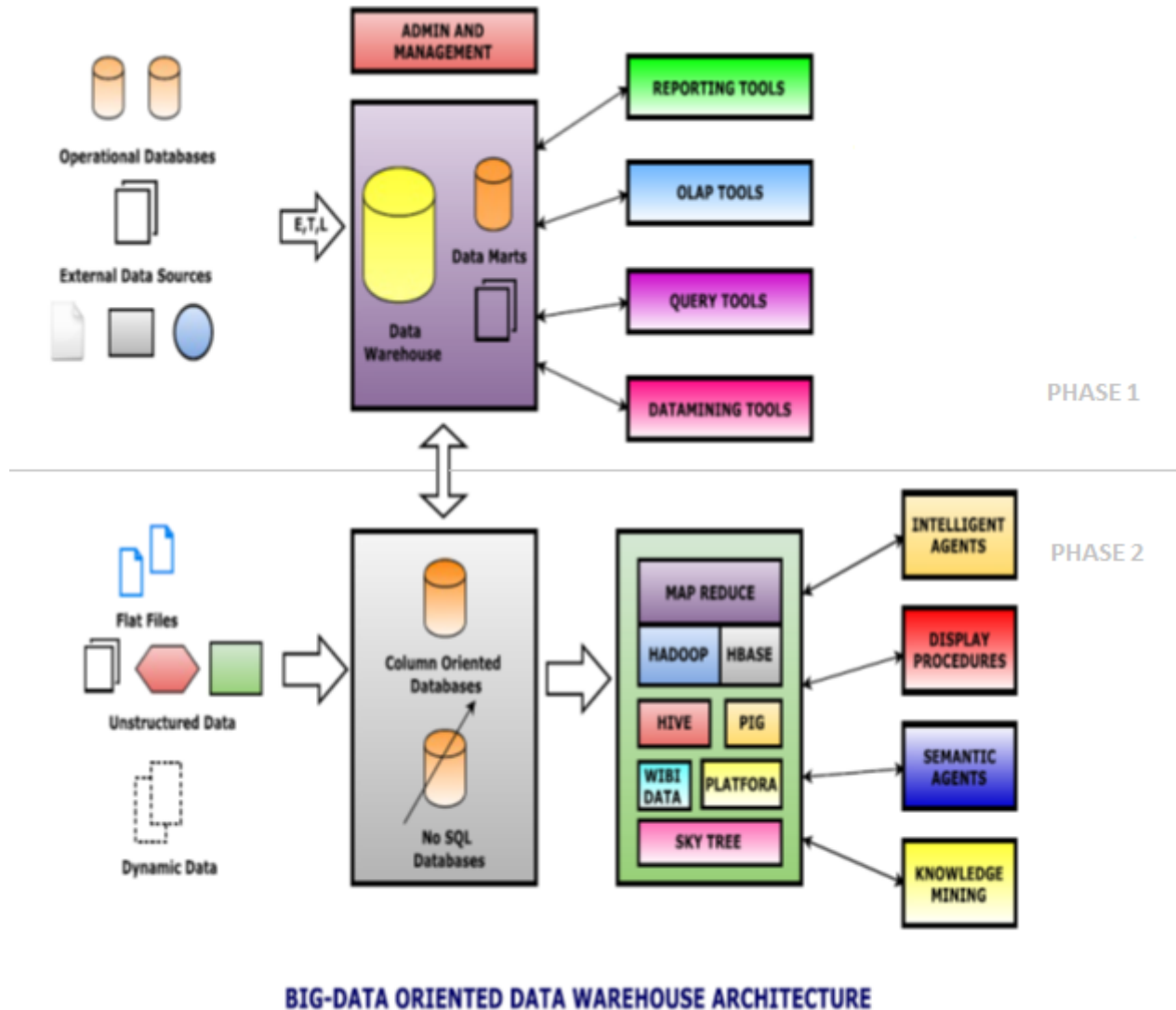


Figure 1: Big-Data Oriented NextGen Data Warehouse Architecture

The Big Data Oriented Data Warehouse architecture's Phase 1 is a replica of Bill Inmon's Data Warehouse Architecture where the External Data Sources and Operational Databases serve as the source for data. The Data from these sources undergoes E, T, and L (Extract, Transform and Load) operations for pre-processing of data which is fed into the Data Warehouse that is centrally located. The Data Warehouse is an integration of a Large Data Warehouse Database with Data Marts which contain highly subject oriented data. The Data Warehouse Environment is further associated with a Repository of Metadata. The Data Warehouse contains Summarized data which can be extracted through Reporting Tools, OLAP Tools, Query and Data Mining Tools. An Admin and Management Component associated with the Data Warehouse are designated for the overall administration of the Data Warehouse.

The Phase 1 architecture which is an adaptation of the Inmon's Conventional Data Warehouse Architecture is loosely coupled with the actual Phase 2 Big Data Technologies arranged in a meaningful and a coherent manner. This architectural component comprises of a large amount of unstructured, dynamic and constantly changing data including the flat file that are transformed into intermediate Big Data Technological Databases such as Column Oriented



# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 10, October 2015

Databases, No SQL Databases or Schema Less Databases. The Data in these Big Data Friendly Databases are in turn processed by a series of technologies.

The data from the Big Data Oriented Databases is further passed into a Pool of Big Data Technologies namely the Map Reduce and Hadoop. Hadoop is the partial implementation of the Map Reduce which effectively processes the raw unstructured large amounts of data in quite a low time. HBASE is a Database mounted over Hadoop which serves as a repository of data. Further the data is processed by technologies like HIVE, PIG, WibiData, Platfora, Skytree based on the situation for the processing of Big Data. The Database Technologies that support Big Data are included in the proposed architecture and constitute Column Oriented Databases and No SQL Databases. In a Column Oriented Database, there occurs classification and organization of data under a similar class into a serialized column. The Column Oriented Database is actually used as it has a phenomenal number of advantages in terms of retrieval efficiency and organization of large amount of data. The Column Oriented Database has Self Indexing Capability where the data values themselves serve as indexes for retrieval which facilitate the phenomenon of data retrieval. Owing to these advantages, Column Oriented Databases are preferred for storing large values of data which is to be processed by Big Data Technologies.

No SQL Databases also referred to as No Schema Databases or even called as Non SQL or Non-Relational Databases are those Databases which are organized in any other manner except that of RDBMS in order to speed up the retrieval process while involving large amounts of data [9]. The specific modeling strategies of a No SQL database could be a key-value type organization or even a document type of organization of data which allows retrieval/update of data. No SQL Databases are integrated in the proposed architecture as their processing capability is much higher than the traditional RDBMS type of databases. Moreover, there is no specificity for organization of data for further data processing. No SQL Databases facilitates easy processing of the large volumes of data by Big Data Technologies. However, No SQL databases are the Broad Class of Non- Relational databases which may include MongoDB, Solr, ElasticSearch, HBASE, Apache Cassandra [9], etc.

The large volumes of data are further processed by the various Big Data Technologies that are coherently collected and organized. Map Reduce is a parallel algorithmic programming framework for distributed processing of large volumes of data which works on a paradigm of Map and Reduce, where the sorting and filtering of data is done in the Map phase and the data is summarized in the Reduce phase [10]. Map Reduce is the most primary among all available big data technologies for facilitating large scale data processing in the most feasible manner. Hadoop being one of the earliest implementations of Map Reduce is made available as an open source that paves a way for the distributed processing of large volumes of data. Hadoop is integrated with the proposed architecture owing to its efficiency, open source nature, easy interoperability, better processing capability and ease of use. The latest version of Hadoop includes Marketing Analytics feature which is a Business Intelligence feature to analyze and predict the marketing trends promoting Marketing Research.

HIVE is by itself an important commodity or an infrastructure for a Data Warehouse [11]. HIVE is built on Hadoop and holds features for data aggregation and summarization. Furthermore, HIVE is a platform engineered for data analytics. HIVE facilitates various, Analytical Procedures, Statistical Tool and Knowledge Engineering Tools to be subjected to the Hadoop Data Pool by creating a pavement for integration. HIVE also offers data compression services, indexing facilities, Metadata storage services to reduce the processing time. Accounting to these advantages and unique features, HIVE is incorporated into the proposed NextGen architecture of the Data Warehouse.

PIG is another technology that serves as an integrating agent to the Hadoop Data Pool with a highly simplified Perl like syntax. It has a high degree of similarity to that of HIVE but differs with the language syntax. Like HIVE, PIG also is an open source with easy availability and interoperability. Thus, PIG is integrated into the proposed NextGen Big DWH architecture. Platfora big data technology is associated with increasing the speed of data analytics, data availability, sharing of data and incorporating a proper channelized view towards the analyzed data. WibiData is included in the proposed architecture as it incorporates web analytics [12] into the Hadoop Data Environment. WibiData incorporates web based integration to Hadoop data, personalization and customized representation of data in the Hadoop Data Environment. WibiData is mainly proposed to be integrated to the Data Warehouse Environment





# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 10, October 2015

mainly to focus on the Web representation of Data in the Hadoop Data Cluster owing to the advancements in the World Wide Web in the present day times. A web based customization is a pre requisite and a compulsive mandate for data in the Hadoop Environment, thus proposing for the inclusion of WibiData into the proposed NextGen architecture. Owing to the necessity of machine learning strategy for processing the large volumes of data, SkyTree technology for big data processing is included in the NextGen Architecture of the Data Warehouse. Skytree is a platform for machine learning associated with data analytics capability which performs with a high degree of efficiency [12].

Intelligent Agents are used for extracting useful data automatically from the Hadoop Data Environment. These agents are trained to work efficiently with the Hadoop Data Cluster to enhance the usability of the Information used for catering to the various needs of information of several classes of users. Semantic Agents are similar to Intelligent Agents but are mainly web based agents that focus on information extraction and optimizations. They focus on associating information to a specific class of users' and give a new meaning to the knowledge extracted to cater to the users' needs. Display Procedures and Knowledge Mining tools may be analytical statistical procedures or rule based methods for extracting useful knowledge from the Hadoop Data Environment.

## IV. EMPIRICAL JUSTIFICATIONS

The empirical justifications for the proposed architecture are done based on a set of research questions that are discussed in detailed in this section. The research questions to be elicited and reasoned are as follows:

**RQ1: There already exist various architectures for a Data Warehouse. Why should a new architecture be proposed?**

It is true several architectures already exists for a Data Warehouse which follows Inmon's reference architecture and they use the traditional E, T, L (Extract, Transform, Load) methodology for processing data. ETL methodology for data processing was perfect for the past decade when the amount of data was limited, but in the most recent times with increase in the number of online users of Social Networks and boom in e-commerce trends, there is a compulsion in the need for enhancing the speed of data processing. Traditional RDBMS, OLAP and ETL technologies are insufficient for processing data. With the debut of Big Data Technologies for Data Analytics, there is a need for a new Architecture for Data Warehouse which has the flavour of Big Data for Rapid Data Processing (RDP).

**RQ2: Why should Big Data Technologies integrate with the Data Warehouse Architecture?**

Big Data Technologies is the latest trend involved in Data Processing. Big data technologies process large volumes of data and most of the Big Data Technologies are available as Open Source. Big data technologies can be subject to data of all kinds whether structured, unstructured or dynamic data, they are all easily processable by Big Data Technologies This is the reason why Big Data Technologies are incorporated with the Data Warehouse Architecture.

**RQ3: Why only a few Big Data Technologies are included in the architecture?**

Only the most prominent Big Data Technologies are included in the architecture as almost all the functionalities needed by the NextGen Big Data Warehouse Architecture is covered in the included Big Data Technologies. There is no need for including other technologies in the Data Warehouse Architecture apart from the mentioned ones as the totality of functionalities needed for the Data Warehouse in under the cover of the mentioned Technologies.

**RQ4: Why is this architecture called Big Data Oriented NextGen Data Warehouse architecture?**

The nomenclature of the architecture is so because it has an essence of Big Data and drives a Data Warehouse at the Enterprise Level. It's called NextGen because it gives birth to a new Data Warehouse Architecture which clearly is an enhancement of the already existing Data Warehouse Architecture with an essence of technologies to process large volumes of data. The Nomenclature of the Architecture is clear, precise, quite straightforward and inferential.

**RQ5: Big Data as a Technology is capable to stand alone. Why Big Data is included in the architecture?**

Big Data Technology is definitely capable of its standalone existence. But the Objective of Big Data Technologies is processing large amounts of data. One of the Objective of Data Warehouse is Data Processing and Analysis. The



# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 10, October 2015

objectives of Big Data Technologies and the Data Warehouse converge at some point. Owing to this reason, both Big Data and Data Warehouse is integrated together despite the heterogeneity involved in them to give rise to a new architecture for Data Warehouse with an essence of Big Data.

## **RQ6: Why should Intelligent Agents be incorporated in the proposed architecture?**

Intelligent Agents are included in order to make data analytics, predictions, querying and visualization of data more appropriate and user oriented. They tend to make the process of Knowledge Extractions much easier and less complex. This lays the foundation for a better and a much efficient data analysis and in turn enhancing the Business Intelligence in the Enterprise.

## **RQ7: What is the main objective of the proposed NextGen Architecture for a Data Warehouse?**

The primary and the most basic purpose of the Data Warehouse is Data Processing. The large volumes of Data are processed and converted to summarize data by both traditional ETL and Big Data Technologies. Once the analytical data is obtained they are subjected to auxiliary objectives namely querying, knowledge mining, data aggregation, data visualization, etc. Other indirect but yet the most primary purpose of the Data Warehouse is to achieve Business Intelligence in an Enterprise,

## V. CONCLUSIONS

In this paper a new architecture for Data Warehousing is proposed which integrates the traditional Data Warehouse Architecture with the Big Data Technologies. The proposed architecture is highly interoperable and can accommodate even the highly dynamic data for analysis. The NextGen Data Warehouse architecture is an amalgamation of ETL technologies as well as the Big Data Technologies for Data Processing. Data Analytics is also seasoned well by inducing semantic and intelligent agents for rendering of information, thus laying a strong foundation for achieving Business Intelligence at the Enterprise Level.

The Big Data technologies are selectively included into the architecture and are evaluated in every aspect for logically ordering them in achieving proper Business Intelligence and Render the best in class performance of the predicted proposed Data Warehouse. The proposed NextGen Business Oriented Architectures clearly accommodates the Big Data viable databases namely Column Databases and No SQL Databases in the Architecture. The proposed architecture is highly interoperable and scalable. This can be reference architecture and customized NextGen Data Warehouse architectures can implement the Data Warehouses. An empirical evaluation based on several research questions is done and justifications are answered for various credibility issues of the proposed Data Warehouse architecture. The given architecture is definitely justified by the research questions and was an opinion level evaluation and proposition of several Data Warehouse experts.

## ACKNOWLEDGEMENT

I whole heartedly thank all the professionals specially the Data Warehouse architects, Big Data Analysts, Business-Oriented Data Warehouse users, Domain Level experts, Technical Architects, Project Associates and Project Managers for taking part in the survey and playing a vital role in proposing this Data Warehouse Architecture.

## REFERENCES

- [1] G Vijay Kumar and M Sreedevi, "Data Warehousing Practices in Business Initiatives." in *Oriental Journal of Computer Science and Technology*, Vol. 1, Issue 1, 2008, pp.55-58.
- [2] K Indira Gandhi and C Shreedar, "Survey on Big Data: Management and Challenges," in *International Journal of Computer Trends and Technology*, Vol. 20, No 1, Feb 2015, pp. 33-36.
- [3] P R S M Lakshmi, SanthiSri and N Veeranjanyulu "A Study of Real-Time Business Intelligence and Big Data," in *Journal for Information Engineering*, Vol. 4, 2015, pp.1-6.



# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

**Vol. 3, Issue 10, October 2015**

- [4] Ramndeeep Kaur, Amanpreet Kaur, Sarabjeet Kaur, Amandeep Kaur and Ranbir Kaur, "An Overview of Database Management Systems, Data Warehousing and Data Mining," in *International Journal of Advanced research in Computer and Communication Engineering*, Vol. 2, Issue 7, 2013 pp.2508-2512.
- [5] W.H. Inmon, "The Data Warehouse and Data Mining," in *Communications of the ACM*, Vol.39 No. 11, 1996, pp.49-50.
- [6] Jens Dittrich, Jorge-Arnulfo and Quaine-Ruiz, "Efficient Big Data Processing, in *Proceedings of the VLDB Endowment* Vol. 5, Issue 12, August 2012, pp.2014-2015.
- [7] Bill Inmon, "Big Data Implementation Vs. Data Warehouse," "<http://www.beyenetwork.com/view/17017>", Last Accessed on 20 Oct 2015.
- [8] Tamara Dull "<http://blogs.sas.com/content/customeranalytics/2015/08/28/isnt-a-data-lake-just-the-data-warehouse-revisited>", Last Accessed on 20 Oct 2015.
- [9] Wikipedia Article "[https://en.wikipedia.org/wiki/ NoSQL](https://en.wikipedia.org/wiki/NoSQL)", Last Accessed 20 Oct, 2015.
- [10] Wikipedia Article "[https://en.wikipedia.org/wiki/ MapReduce](https://en.wikipedia.org/wiki/MapReduce)", Last Accessed 20 Oct, 2015.
- [11] Wikipedia Article "[https://en.wikipedia.org/wiki/ Apache\\_Hive](https://en.wikipedia.org/wiki/Apache_Hive)", Last Accessed 20 Oct, 2015.
- [12] Wikipedia Article "<http://www.techrepublic.com /blog/big-data-analytics/10-emerging-technologies-for-big-data>", Last Accessed 20 Oct, 2015.