International Journal of Engineering Science Invention (IJESI)
ISSN (Online): 2319 – 6734, ISSN (Print): 2319 – 6726
www.ijesi.org ||Volume 7 Issue 2 Ver. IV || February 2018 || PP.64-68

Coexistence, Trends and Comparative Study between Traditional Relational Database Management System and Big Data

Parag C. Shukla[1], Dr. Bankim L. Radadiya[2]
[1]Assistant Professor & HEAD, Department of MCA, Atmiya Institute of Technology & Science, Rajkot, India,
[2] Director-IT, Navsari Agricultural University, Navsari, Gujarat, India,
Corresponding Author: Parag C. Shukla

ABSTRACT :*Traditional RDBMS rise from 20th century and nowadays we find the buzz word Big Data. As per the google trends, in 2011 the word big data has cross the popularity line of RDBMS worldwide. Even many international companies started their investment and work in big data. It doesn't mean that RDBMS is going to die, still many companies are bit comfortable with RDBMS. Not only transaction dispensation, but for altogether much attentive index concerned with query on that type of data. Big data and Hadoop is an alternative kind of environment for different types of analysis and diverse kind of data and two of them are co-exists. As per the trends big data is buzzword nowadays. People are choosing big data over RDBMS if they want to store structured as well as unstructured data and if they are preferring open-source as well as with faster speed. There are lot of difference between RDBMS and big data like variety, architecture, throughput, Scalability, Latency response time, cost, data processing etc. In Traditional RDBMS environment, all the enterprise's data is stored in central server whereas in big data environment data resides in distributed file system. Distributed File System scale by scaling in or out horizontally as compared to typical database server that scales vertically. In this research paper we studied big data and relational database management system and finds the trends, co-existence and comparison between both. It is not about rip and replace. It will not be possible to get rid of RDBMS or massively parallel processing, but instead use the right tool for right job.*
KEYWORDS:*Traditional RDBMS, Comparison, Difference between RDBMS and Big data, SQL Vs NoSQL, Hadoop Vs RDBMS, Coexistence between Data Warehouse and Big Data*

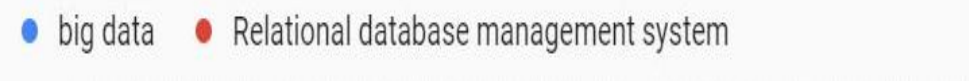
Date of Submission: 09-02-2018 Date of acceptance: 24-02-2018

I. INTRODUCTION

Traditional RDBMS rise from 20th century and nowadays we find the buzz word Big Data. As per the google trends, in 2011 the word big data has cross the popularity line of RDBMS worldwide. Even many international companies started their investment and work in big data. It doesn't mean that RDBMS is going to die, still many companies are bit comfortable with RDBMS. Not only transaction processing, but for all much attentive index concerned with query on that type of data. Big data and Hadoop is an alternative kind of environment for diverse kinds of scrutiny and different kind of facts and two of them are co-exists. As per the trends big data is buzzword nowadays. People are choosing big data over RDBMS if they want to store structured as well as unstructured data and if they are preferring open-source as well as with faster speed. There are lot of difference between RDBMS and big data like variety, architecture, throughput, Scalability, Latency response time, cost, data processing etc. In Traditional RDBMS environment, all the enterprise's data is housed in central server whereas in big data environment data resides in distributed file system. Distributed File System scale by scaling in or out horizontally as compared to typical database server that scales vertically. In Traditional BI, data is generally analysed in an offline mode whereas in big data it is analysed in both real time as well as in offline mode.

II. TRENDS BETWEEN BIG DATA AND RDBMS

As per google trends the popularity line of Big data has cross the popularity of RDBMS in 2011 world-wide.



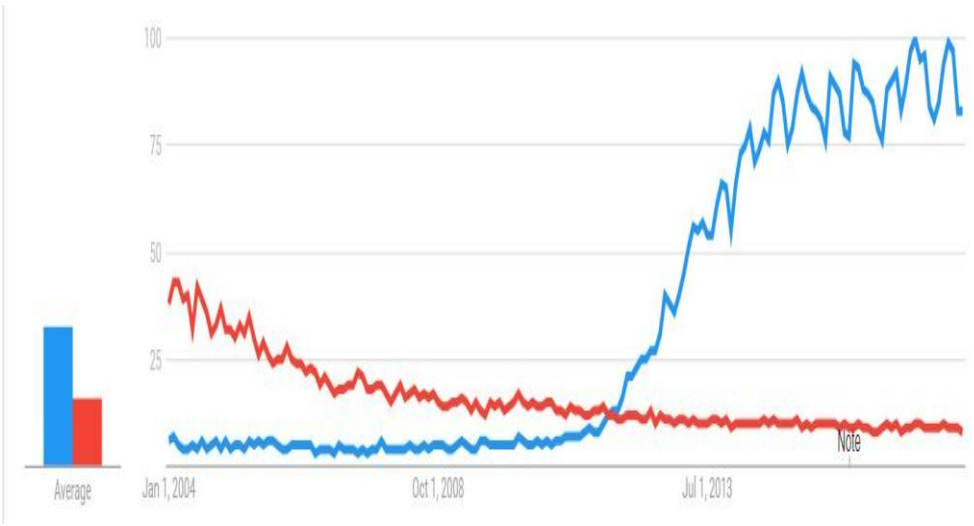


Figure-1 World-wide trends between RDBMS and Big Data [Source:google trends)

In India, as per google trends the popularity line of Big data has cross the popularity of RDBMS in Sep 2012 world-wide.

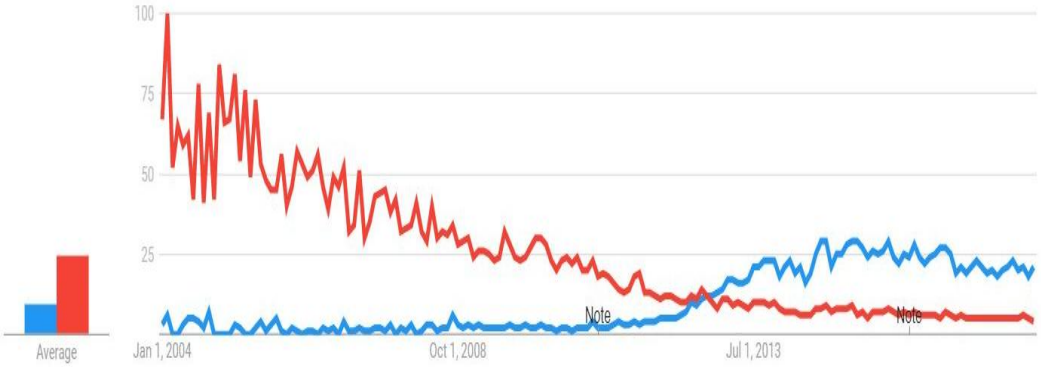


Figure-2 Trends between RDBMS and Big Data in India [Source:google trends)

III. COMPARISION BETWEEN RDBMS AND BIG DATA

In Traditional RDBMS environment, all the enterprise's data is stored in central server whereas in big data environment data resides in distributed file system. Distributed File System measure by scaling in or out horizontally as compared to typical database server that scales vertically. In Traditional RDBMS, data is generally analysed in an offline mode whereas in big data it is analysed in both real time as well as in offline mode. Traditional RDBMS is about structured data and it is here that data is taken to processing functions (move data to code) whereas big data is about variety: structured, semi-structured, and unstructured data and here the processing functions are taken to the data.

Typical Data Warehouse Environment

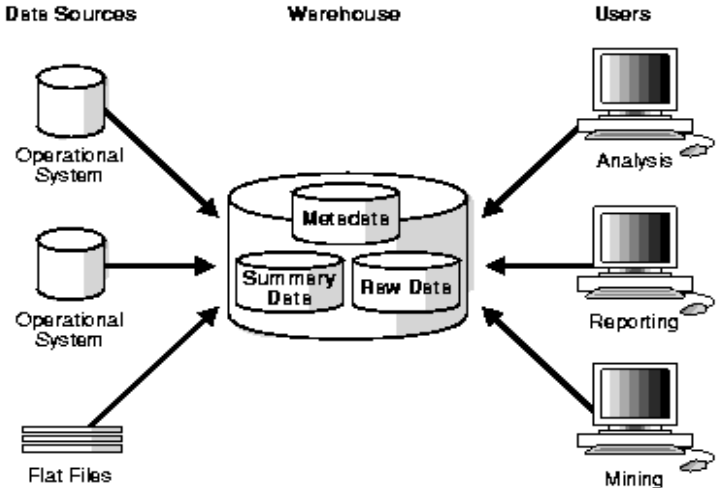


Figure-3 DATA WAREHOUSE ENVIRONMENT

Functioning or transactional or everyday business facts is collected from Enterprise Resource Planning (ERP) systems, Customer Relationship Management (CRM), legacy systems and numerous third-party applications.

The facts from this foundation may vary in format [data could have been housed in any RDBMS such as Oracle, MSSQL, DB2, MySQL or spreadsheet or txt).Data may come from information sources located in same geography or different geographies.This data is then combined, prepared up, altered and uniform through the process of Extract, Transform and Loading (ETL).The transformed data is then loaded into the enterprise data warehouse or data marts.A host of market leading business intelligence and analytics tools are then used to enable decision making from the use of ad-hoc queries, SQL, enterprise dashboards, data mining etc.

Typical Hadoop Environment

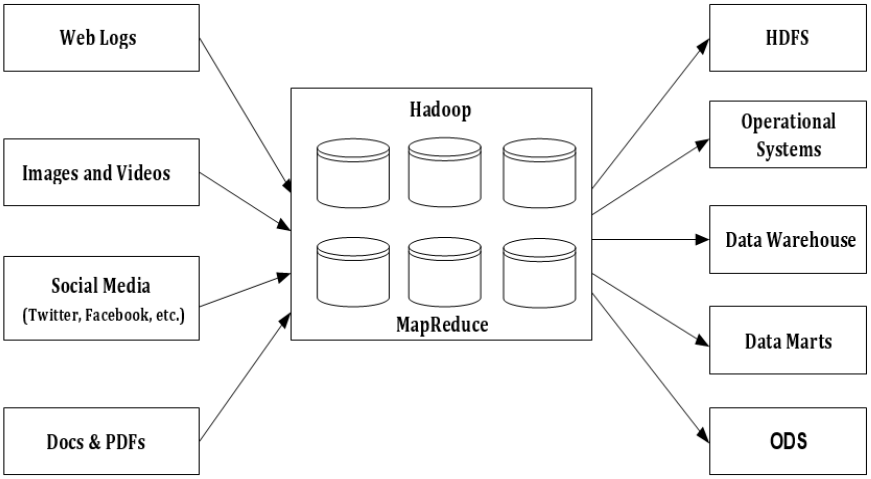


Figure-4 HADOOP ENVIRONMENT

As is obvious from the Figure, the data foundations are dissimilar from web logs to images, audios and videos to social media data to the various docs, pdfs etc.Here the data in focus is not just the data within the firm's firewall but also data exist in outside the business's firewall.This information is placed in HDFS (Hadoop Distributed File System)**.** If essential be, this can be repopulated back to operative systems or fed to the initiative data warehouse or data marts or Operational Data Source (ODS) to be picked for additional dispensation and analyse.

HADOOP	RDBMS
Stores data in native format	Structured format
Scale horizontally and vertically	Scale up vertically
Key-Value Pairs	Relational table
Functional Programming	Declarative Queries
Off-line batch processing	On-line transaction processing
CAP Theorem	ACID Properties
Dynamic Schema	Pre-Defined Schema
Largelypreferred for large datasets	Not preferred for large datasets

Bestfitforhierarchicalstorageasit follows thekey-valuepairofstoringdatasimilartoJSON(JavaScript ObjectNotation)	Not best fit for hierarchical data
Example: MongoDB, HBase, Cassandra, Redis, Neo4j, CouchDB, Couchbase etc	Examples:Oracle,DB2,MySQL,MSSQL, PostgreSQL,etc.
Model-less Approach	Relational Model
Reliesheavilyoncommunitysupport	Excellentsupportfromvendors
Open-Source	Open-Source as well as Closed
Doesnothavegoodsupportforcomplexquerying	Supports complex querying

Table-1 HADOOP Vs RDBMS

IV. COEXISTENCE OF BIG DATA AND DATA WAREHOUSE

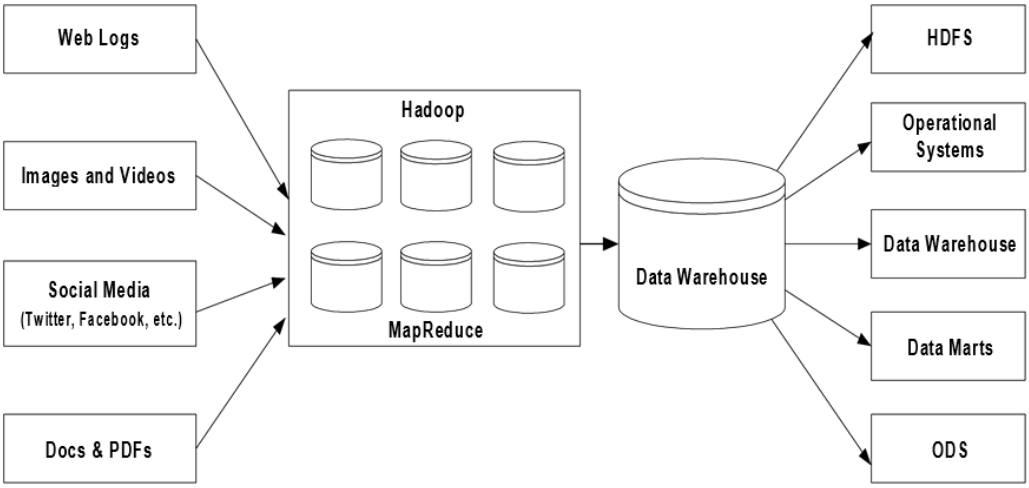


Figure-5 COEXISTENCE BETWEEN HADOOP AND RDBMS

A coexistence policy that syndicates the greatest of legacy data warehouse and analytics environment with the novel control of big data results is the best of both the worlds.It is not about rip and replace. It will not be conceivable to get rid of RDBMS or massively parallel processing, but instead use the right tool for right job.As we are conscious that many companies are bit contented working with obligatory data warehouse for standard data warehouse and analytics reporting.For example, the quarterly sales report, customer dashboard etc.The data warehouse can endure with its standard capacity drawing data from legacy operational systems, loading the historical facts to establishment of traditional data warehouse reporting and analytics requirements.
However, one will not be able to disregard the control that Hadoop brings to the table with diverse types of scrutiny on dissimilar types of data.The same functioning systems, which till now was involved in powering the data warehouse, can also populate the big data environment when they're needed for computation-rich dispensation or for raw data examination.It will be tight complementary act to steer the workload to the right platform based on what that platform was designed to do.Just as one cannot ignore the powerful analytics capabilities of Hadoop, one will not be able to ignore the revolutionary development in RDBMS such as in memory processing etc.Both Data warehouse and Hadoop co-exist in today's environment.

V. CONCLUSION

We came to conclusion that trends of big data is more popular compare to relational database management system. Both are best at their feet. Both are providing platform to store the data and to analyze the information to the user. There are several differences between Big Data and RDBMS. In Traditional RDBMS environment, all the enterprise's data is stored in central server whereas in big data environment data resides in distributed file system. Distributed File System measure by scaling in or out horizontally as compared to typical database server that scales vertically. In Traditional RDBMS, data is generally analysed in an offline mode whereas in big data it is analysed in both real time as well as in offline mode. Traditional RDBMS is about structured data and it is here that data is taken to processing functions (move data to code) whereas big data is about variety: structured, semi-structured, and unstructured data and here the processing functions are taken to the data. But still we can say that both are co-existence. It is not about rip and replace. It will not be possible to get rid of RDBMS or massively parallel processing, but instead use the right tool for right job.

Acknowledgements

We wish to thank google trends for providing data of current trends, apache foundation, oracle corporation & sincere thanks to our mentor for inspiring to do research work.

REFERENCES

[1]. Big Data and Analytics – Wiley Publication, Seema Acharya, Subhashini Chellapan
[2]. https://docs.oracle.com/cd/B10501_01/server.920/a96520/concept.htm.
[3]. Wei Fan, Albert Bifet," Mining Big Data: Current Status, and Forecast to the Future", SIGKDD Explorations, Volume 14, Issue 2.
[4]. Srinivasan,"SOA and WOA Article, Traditional vs. Big Data Analytics, "Why big data analytics is important to enterprises"
[5]. Elena Geanina ULARU, Florina Camelia PUICAN, Anca APOSTU, ManoleVelicanu," Perspectives on Big Data and Big Data Analytics", Database Systems Journal vol. III, no. 4/2012
[6]. Big Data Survey Research Brief", Tech. Rep.SAS, 2013
[7]. DhrubaBorthakur, The Hadoop Distributed File System: Architecture and Design.
[8]. NESSI-Big Data White Paper," Big Data –a new world of opportunities" December 2012
[9]. Singh S. 2008. Agricultural Mechanization Policy. Proceedings of Tractor & Farm Machinery Manufacturers' Meet, Nov. 16-17, 2007, CIAE, Bhopal.
[10]. Kulkarni SD. 2005. Food Safety and Security Issues in India: Challenges and Approach. Lead Paper for presentation in National Seminar on Post Production Systems and Strategies to the Issues and Challenges of Food Safety and Security during Sept. 22-23, 2005 at TNAU, Coimbatore – 641003
[11]. Shukla P, Radadiya B, Atkotiya K. An Emerging Trend of Big Data for High Volume and Varieties of Data to Search of Agricultural Data. Orient.J. Comp. Sci. and Technol;8(2)
[12]. White Paper by DataStax Corporation - 2012.
[13]. https://en.wikipedia.org/wiki/Comparison_of_structured_sto rage_software
[14]. https://www.qubole.com/blog/hadoop-vs-traditional/
[15]. https://www.w3trainingschool.com/difference-big-data-hadoop-traditional-rdbms
[16]. The Apache Software Foundation Announces Apache Cassandra Release 0.6 : The Apache Software Foundation Blog.
[17]. http://www.eenadupratibha.net/pratibha/Engineering-Colleges/Engineering-Jobs/content-four-bigdata-unit2.html
[18]. https://www.qubole.com/blog/hadoop-and-data-warehouse/
[19]. https://mapr.com/blog/exploring-relationship-between-hadoop-and-data-warehouse-part-2/
[20]. https://tdwi.org/articles/2015/01/27/hadoop-replace-data-warehouse.aspx

International Journal of Engineering Science Invention (IJESI) is UGC approved Journal with Sl. No. 3822, Journal no. 43302.

Parag C. Shukla, Dr. Bankim L. Radadiya"Coexistence, Trends and Comparative Study between Traditional Relational Database Management System and Big Data" International Journal of Engineering Science Invention (IJESI), vol. 07, no. 02, 2018, pp. 64–68.