*International Journal of Engineering Applied Sciences and Technology, 2016*
**Vol. 2, Issue 2, ISSN No. 2455-2143, Pages 27-29**
Published Online December-January 2016 in IJEAST (http://www.ijeast.com)

# TRENDS OF WORKLOAD PREDICTION OF BIG DATA ON CLOUD COMPUTING ENVIRONMENT

Supreet Kaur Sahi
Asst Prof, SGTBIMIT

*Abstract*— **In this paper author has presented different trends of big data on cloud environment. Workload prediction of big data on cloud depends upon type of applications. Big data consist of different applications like database management, data analysis, data visualization, data regression etc. In this paper tools corresponding to these applications are also presented.**

*Keywords*— **Big data, cloud environment, IaaS, PaaS**

## I. INTRODUCTION

Standardizations, self-service and flexibility are some of the features of cloud computing which makes it powerful technology. Cloud computing has eliminated need to maintain expensive hardware, software and dedicated storage space [1][2]. Massive growth in scaling of data or big data arises from cloud computing environment. This scaling leads to fluctuations of workload. Big data addressing is time consuming and challenging task as it demands high computational infrastructure for data analysis and processing. The trend of big data and cloud computing is presented in this study. In this articles research issues are studied with focus on availability, data integrity, scalability and governance.

Big data is unstructured data such as output records, weather data, financial transactions that requires analyst for generation of systematic analysis. Big data is a collection of large number of data sets that cannot be managed by available database management tools. Searching, sharing, capturing, transferring, analyzing are challenges of the big data. The complexity of conversion of data to systematic data varies from organization capabilities of management of unstructured data [3][4].

In cloud computing, IaaS (Infrastructure as a Service) allows deployment of nodes while PaaS (Platform as a Service) helps in scaling capacity on demand and reduce costs. The Big Data processing for enterprises of all sizes are permitted by the cloud, but there are still complexities in gathering of the business data from the big unstructured business data available [4][5][6].

Whether dealing with social media, analyzing log files, following click streams, managing biological sequence or analyzing transactions to avoid threats, big data is used in one way or other [7]. Cloud computing features of elasticity and on demand provisioning make big data analytics accessible to more teams.

## II. CATEGORIES OF TOOLS AVAILABLE

Data can be collected or transferred in a cloud data sink like Amazon S3. Government agencies and other organization use the maze of data for analyzing security related or simply pattern of consumers. According to survey by IBM on average, 2.5 billion gigabytes of data is created on daily basis consist of 200 million tweets and 30 billion contents Facebook content shared each month [8]. There are different tools available on cloud for understanding big data. These tools are categorized based on following parameters.

1. Database management
2. Data cleaning
3. Data analysis
4. Data mining
5. Data visualization
6. Data integration
7. Data languages
8. Data collection

With big data the major problem that arises is of storage. Cloud services can act as a storage provider. Workload prediction of big data on cloud depends on amount of storage required. Before mining of data starts one must have to understand and initiate cleanup process of data to create well-structured data sets [9]. Some of the tools that help in storage of big data are as follow [10]:

1. Hadoop: It is an open source software framework available for distributed storage of large datasets on cloud clusters. Using this, scaling of data can be done without worrying about failure of data. It provides massive storage capabilities, high processing power and ability to handle large amount of concurrent transaction.

2. Cloudera: It is an enterprise solution that helps businesses to manage and store internal data of

organizations.

3. MongoDB: It is excellent tool for managing data changes frequently or unstructured or semi structured data. Common uses of this tools are mobile apps data storage, product manuals, real-time catalogues, application delivering etc.

4. Talend: It combines real-time data applications with embedded quality. It is open source system.

5. Open Refine: It is an open source tool dedicated for cleanup process of messy data. With help of this tool one can explore unstructured data quickly and easily.

6. DataCleaner: It transforms semi-structured data sets into readable data sets. It also provides data warehouse and data management services.

7. RapidMiner: This is data mining tool used for predictive analysis. It can be used to integrate API through algorithms.

8. IBM SPSS Modeler: It provides complete suite of data mining that includes text analysis, entity analytics, decision management and optimization. It can run on virtually any type of database.

9. Oracle data mining: It allows modeling for discovering customer behavior and targeting best customer. It enables data analyst to work with database in drag and drop manner.

10. Teradata: It provides end-to-end solutions in data warehousing, marketing applications and analytics. It helps in implementation, training and consulting.

11. FramedData: It has very good features about analyzing the data and making reports about those customers, which are not happy with the company products.

12. Kaggle: This is the world largest science community. If anyone is struck in any data-mining problem in big data it can provides solutions.

13. Qubole: It simplifies, scales and speeds big data analysis workload against data on Google, Azure, AWS clouds.

14. BigML: It offers a powerful machine-learning interface for importing data and getting predictions.

15. Statwing: This tool provides visual of complex analysis.

16. Tableau: This visualization tool focus on business intelligence. This tool helps in data visualization by creating bar charts, maps, plots without much needed coding.

17. Silk: It helps in building interactive charts and maps. It is simple data visualization and analytical tool available.

18. CartoDB: It is data visualization tool that helps in visualizing locations without much needed coding.

19. Chartio: It allows combination of data sources and queries in a browser. Powerful dashboards can be created with few clicks. It helps in fetching data from anywhere without knowledge of SQL.

20. Plotl.ly: This system helps in creation of 2d and 3d charts. The free version of this tool helps in creation of unlimited public charts.

21. Datawrapper: It is an open source tool that created embedded charts for data analysis. It invites contributions for improvements from users.

22. Blockspring: This data integration tool can harness data to tools like excel and Google sheets. It allows easy connection to any third party software. It can very well connected to AWS, import.io, Google Spreadsheet etc.

23. Pentaho: It helps data integration with simple drag and drop user interface. They offer embedded and business analytics services.

24. R Software: It is a language used for statistical and graphics based computing.

25. Python: It provides wide range of library for different tasks.

26. Import.io: It is a tool for data extraction. With the help of simple user interface, webpages can be transformed into easy to use spreadsheet that can

analyze, visualize and use data- driven decisions.

## III. CONCLUSION

The workload of big data on cloud computing will depend on applications one is working upon. Tools for different applications like data visualization, data mining, data integration etc. are presented in this article. This article can be used as base for understanding research trends in cloud computing and big–data.

## IV. REFERENCE

[1] [1] B. Hayes. “Cloud Computing,” Communications of the ACM, 2008.

[2] [2] Thomas Mendel, Vice President, EMEA “MARKET OVERVIEW: CLOUD INFRASTRUCTURE SERVICES 2012 Maturing Vendor Offerings in a Busy Market”: HfS Research.

[3] [3] Big Data Offers Big Opportunities for Retail, Financial, Web Companies. <http://www.eweek.com/enterprise-apps/big-data-offers-big-opportunities- for-retail- financial-web-companies/>.

[4] [4] O’Driscoll A, Sleator RD. Synthetic DNA: the next generation of big data __storage. Bioengineered 2013:4.

[5] [5] Xiaofeng M, Xiang C. Big data Management: Concepts, Techniques and Challenges [J]. Journal of Computer Research and Development, 2013, 50(1): 146-169.

[6] [6] Mayer-Schönberger V, Cukier K. Big data: A revolution that will transform how we live, work, and think [M]. Houghton Mifflin Harcourt, 2013.

[7] [7] “Big data in cloud” retrieved from http://www.rightscale.com/solutions/cloud- computing-uses/big-data, Jan 10, 2016

[8] [8] Maamar Ferkoun, “Cloud Computing and Big data: an Ideal combination” reterived from https://www.ibm.com/blogs/cloud-computing/2014/02/cloud-computing-and-big- data-an-ideal-combination/, Feb 2014

[9] [9] “Big Data Cloud Database & Computing” retrieved from https://www.qubole.com/resources/article/big--data--cloud--database--computing/, Aug 2016

[10] [10] “Guide to Big Data Analytics: Platforms, Software, Companies Tools, Solutions and Hadoop”, http://cloudnewsdaily.com/big-data-analytics/, Aug 2016