IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Volume-4, Special Issue-5, Jul 2016 E-ISSN: 2347-2693

Proposed Scalable Architecture for Analyzing Big Data in Education System

Nishant Agnihotri[1], Aman Kumar Sharma[2*]

Department of Computer Science, Himachal Pradesh University, Shimla

Available online at: www.ijcseonline.org

Abstract: **Big data have emerged very fast in the last couple of years and it came up with a number of solutions to different problems varies from effective decision making, real time decision making, and effective implementation of E-Governance etc. Information and Communication technology is the major responsible for this boom in the data. It has also given birth to number of challenges like Effective data architectures, Effective analysis algorithm, Frameworks to handle this big data. As the size of data is increasing repeatedly, it gives a challenge of scaling the frameworks which could be able to handle such data. On the basis of aforesaid problems and challenges this paper has proposed a scalable architecture for the analysis of Big Data in higher Education System.**

Keywords: Big Data, ICT, Scalability, Frameworks, Architectures, Education system.

I. INTRODUCTION

Now a day's data are emerging on a doubling rate per year and it is going beyond the control of traditional systems or legacy systems. Data which falls under this category is termed as Big Data. Initial Big data were considered to have three characteristics, i.e. Volume, Velocity and Verity. With the emergence of research two more characteristics of big data Veracity and Value were added to the 3Vs of Big Data.

- **Volume:** it is the size of the data being generated through various sources mostly falls under ICT.
- **Velocity:** it the speed or the rate at which data is received on servers in the form of images, videos and documents [9] [10].
- **Variety:** refers to the type of data being stored on the server's i.e. raw documents, images, videos or audio songs.
- **Veracity:** refers to how true the data are in the term of the integrity of the content.
- **Value:** refers how big data generated the revenue to the different organizations in terms of money. It is the direct benefit from the better decision making in terms of hike in economy of any organization.

Big data have also given benefits to the number of other industries like social networking industries, small scale and large scale organizations, production units. Apart from it, Big data analysis (BDA) has also helped in real time decision making in the fields of election analysis, education reformation, social media analysis, flood detection and many other fields.

Computational architecture is the major challenge of the scientists dealing with the big data computation and analysis. Architectures and frameworks are major demand to solve this purpose of computation. Reduction in the time of time of job handling is also a big challenge which has been taken care by MapReduce as of now. Fig.1.Shows the complete working of the MapReduce job handling procedure. No matter MapReduce provides what kinds of architecture, but on the same time there need to have a good framework or architecture for the successful implementation of the MapReduce work procedure to get effective results of the data.

This paper has focused on proposing a scalable computational framework to handle the Big Data. As scalability has become a major concern of any system with the increase in the computational capacities of the systems or with the hike in the size of the data. This paper is further organized in 4 sections covering the related work done in the field of Big Data from .

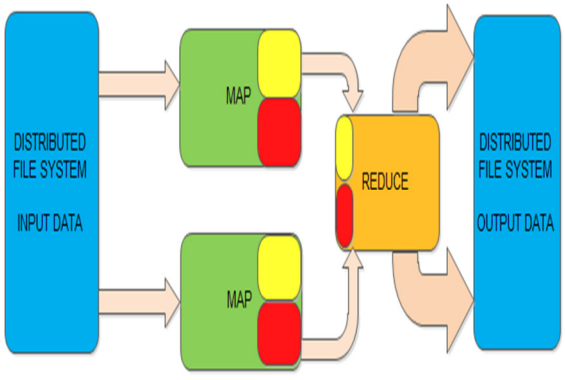

Fig. 1.*Data Flow in Map Reduce*

Scalability to various architectures. Later this paper will cover proposed framework for the scalable computation of big data followed by benefits, conclusion and future work and references.

II. RELATED WORK DONE

A Munar et.al.[1] Have discussed about the financial information, management architecture for global banking. They proposed and architecture over the traditional approach of the distributed computing paradigms. Proposed architecture was capable to consolidate, validate, enrich the process with different big data analytics techniques where the data was gathered from various systems use to encounter in the routine processing of the banking environment.

Samuel Marchal et.al.[2] Have worked on the architecture of the big data which was capable of handling the large scale security. Proposed architecture was based on the collecting and storing honeypot, DNS (Domain Name System) data, HTTP traffic and IP flows in a distributive manner. They discussed several correlation schemes which were relying on their data and the application. This proposed system was capable of fitting the five states of big data into it and it worked on offline environment.

Apart from the proposed architectures in the field of big data scalability is also discussed in the field of cloud computing where architectures are proposed for improving the services of the cloud. NishantAgnihotri et.al.[3] Proposed a scalability improvement system for the cloud computing. They took the scalability as the major factor and introduced an architecture which was capable of improving the PaaS (platform as a service) scalability. Proposed system generated the benefits like scaling down delays, scaling down cost, scaling up connectivity.

Enrique Moguel et.al.[4] Worked on Multilayered Big data architecture for remote sensing in Eolic parks. They described a new system where strategies were made to handle the big data generated from the Eolic parks. This system covers the data ranging from the remote sensing and designed to cope up with the specific contexts of the data.

LakshmishRamaswamy et.al.[5]Has taken the concept of quality-centric Big Data architecture for federated sensor services. They proposed a cloud based big data architecture for handling and supporting sensor services. They also discussed the various benefits from the implementation of the proposed architecture.

Benefits of big data are the direct result of the application fields associated with it and are discussed by many researchers.

Asur et.al.[6] Has discussed how Big data is generated from social media and analysis is performed over social media data for predicting future. They proposed and proved their work on the basis of data generated from social media and proved it with the real time result of movies over the box office. They implemented their work on the bases of movies before its release and after release.

N. Agnihotri et.al. [7] Have discussed about the benefits of big data in the field of e-governance. They discussed how the effective analysis of the Big Data helps the government to effectively implementation of different policies. They discussed the various benefits of implementation of the big data analysis in the public sector.

Chen et.al. [8] Discussed various areas dealing with web analysis and discussed future and further references of the analysis in web analysis. The areas which they considered covered campaign advertising, voter-mobilization, policy discussion and donations.Most of the work in such fields are done by the government itself and remaining contribution is given by the academicians.

III. PROPOSED FRAMEWORK

P. Michalik et.al. [11] gavea conceptual definition of big data architecture in the education system. They designed an algorithm for processing the data in any educational institute and also discussed various sources of the data. Various sources of data in an organization can be

- Server logs
- Documents
- Social networks
- Public portals
- Information systems

Most of the educational institutes have aforesaid data collection methods. Analysis took place over all these parameters Fig. 2. Represents the role of analysis in an educational institute.The decision is the major issues which totally depend on analysis of the educational related data in order to give improvement concepts. As the data is growing at a doubling rate every year it will be a great challenge to provide a platform which will be capable enough to scale up its boundaries according to data being collected from various sources.

This paper has focused on proposing an architecture which will be able to extend its boundary with the emergence of large data in computerized environment. In addition to itan algorithm of the proposed architecture is discussed which will show the flow of data in scalable architectures for an educational system to provide an effective analysis as it has many benefits discussed by D. Stodder[11].

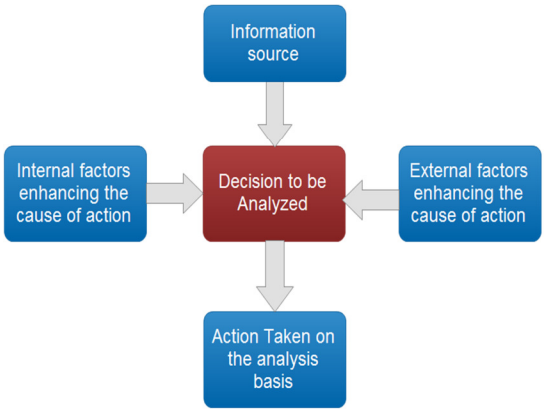

Fig. 2.*Role of Analysis in Education System*

In Fig. 2. Decision analysis is totally based on the analysis reports to be generated from Hadoop Distributed Filesystem with the helpof some analysis algorithms. Fig. 3. Represents how the complete process of decision to be analyzed will took place in the Big Data analysis .

Proposed system shown in Fig.4. Has a capacity of adding more data and scaling the computational limits of the framework. Proposed system has the capacity of adding the buffer in front of Mapper and Reducer. System will be having a provision of adding the buffers as many required depending upon the need of computational data. The addition of the buffer on the mapper end will help to manage the large datasets without enhancing the value of Mappers and reducer. The limited scope of the computational tools will then be able to enhance the boundary of handling more data.

Further the incremental value of buffers will be added with the modifications in the proposed algorithm with this architecture.

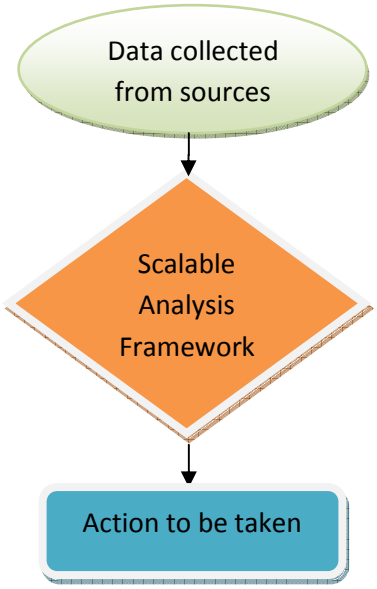

Fig. 3.*Decision Analysis Process*

Algorithm 1: Data Processing in Scalable Architecture

1. **Data:** Fetch data from DFS
2. **Result:** output to DFS
3. **Select Buffer** : Call Algorithm Buffer selection for scalability
4. **Repeat:**
   a. Map similar pairs MP mp1,mp2…..mpn
   b. Select similar pairs of data
   c. Collect it in buffer
   d. Call Algorithm [2] Buffer selection for scalability
   e. Reduce similar pair RP rp1,rp2….rpn
5. Until DFS empty

Algorithm 1**:** *Data Processing in Scalable Architecture*

Algorithm 2: Buffer selection for scalability Improvement

1. **Add**: data to buffers from DFS
2. **Compare** : Size of DFS Sd
   a. If Sd> Input data(ID)&&Sd==ID
   Skip buffer
   b. Else Sd< Input data(ID)
   Size of buffer Sb= DFS-ID
   Add buffer B1……Bn
3. **Forward**: data to Mapper/Reducer
4. Until DFS empty

Algorithm 2: *Buffer selection for scalability*

IV. BENEFITS FROM SCALABLE ARCHITECTURE

As we have discussed about how data is doubling at a pace of two times a year after the last few years. It is basically a challenge to architectures to handle such data with a swinging flow. This proposed architecture will be capable of following

1. To cope up with the changing size of data to provide fast and sustainable environment.
2. Consumer satisfaction is most challenging jobs for any system which will be carried out in a perfect manner by this system.
3. Addition of buffer with the change of the data is easy because of customizable algorithms.
4. Will help educational institutes to take better decision without considering size of data.

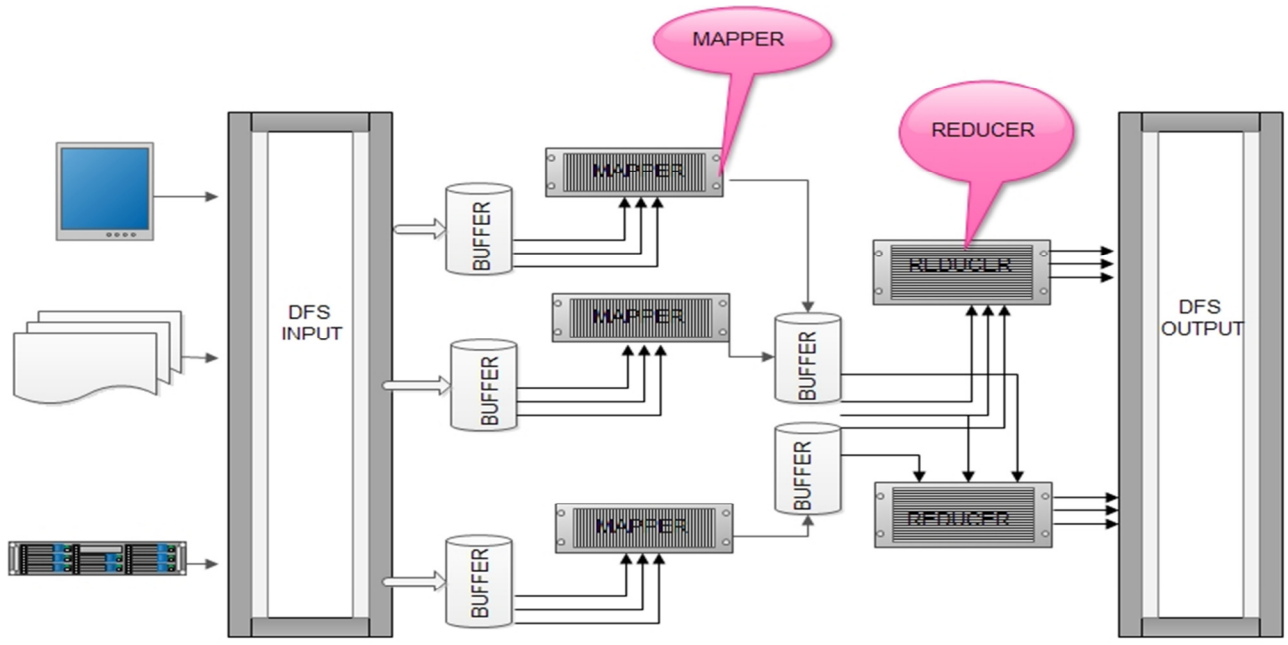

Fig.4. *Proposed Scalable Architecture*

V. CONCLUSION AND FUTURE WORK

This paper has mainly focused on a scalable platform for the changing data size and computational demands. Scalable systems are always admired by the end user and consumers. This system will be capable of handling large data sets by adding the buffer values in comparison to the DFS. Future work of this approach will be toward providing such scalable platforms for real time stream analysis where capturing and managing of real time incoming large data sets is a big challenge.

REFERENCE

[1] A Munar, E. Chmer and I. Sales, "A Big Data Financial Information Management Architecture for Global Banking", *IEEE International Conference on Future Internet of Things and Cloud*, pp: 385-388, 2014.

[2] S.Marshal, X. Jiang, R. State and T. Engel, "A Big Data Architecture for large Scale Security Monitoring", *IEEE International Congress on Big Data*, pp: 56-63, 2014.

[3] N. Agnihotri and A.K.Sharma, "Evaluating PaaS Scalability And Improving Performance Using Scalability Improvement Systems", *International Journal of Research in Engineering and Technology*, pp: 598-602, 2014.

[4] E.Moguel, J.C. Preciado, F.S. Figueroa, Miguel A. Preciado and Juan Hernandez, " Multilayer Big Data Architecture for Remote Sensing in Eolic Parks", *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2015.

[5] L.Ramaswamy, V. Lawson and S.V. Gogineni, "Towards A Quality-Centric Big Data Architecture for Federated Sensor Services", *IEEE International Congress on Big Data*, pp: 86-93, 2013.

[6] S. Asur and B. A. Huberman, "Predicting the future with Social Media", *IEEE/WIC/ACM, International conference on web intelligent agent technology*, pp. 492-499, 2010,

[7] N. Agnihotri and A.K. Sharma, "Big data analysis and its need for effective E-governance", *International Journal of Innovations & Advancement in Computer Science,* Vol. 4, pp. 219-224, 2015.

[8] H. Chen, R. H. L. Chiang and V. C. Storey, "Business Intelligence and Analytics: From Big Data to Big Impact" MIS Quarterly, pp 1165–1188, 2012.

[9] Official website of SAS company http://www.sas.com/big-data/

[10] Official webpage of IBM company: http://www-01.ibm.com/software/data/bigdata/

[11] D. Stodder," Applying analytics with big data for customer intelligence", TDWI checklist report, online : www.tdwi.org.

