Implication of Restful web services in Bigdata Analytics

Gopal Krishna	Ashok kumar P S	Manjunath Swamy	Noor Basha	Punith Palagi
Assoc Prof, Dept. of CSE	Prof., Dept of ISE	Prof., Dept of CSE	Asst Prof, Dept. of CSE	Asst Prof, Dept. of CSE
AIT	DBIT	AMCE	VIT	NHCE
Chikmagaluru, India	Bengaluru, India	Bengaluru, India	Bengaluru, India	Bengaluru, India
nithin.gopal@yahoo.com	**ashokydit@gmail.com**	**venkateshptcedbit@gmail.com**	**mdnoor202@gmail.com**	**puneetpalagi@gmail.com**

***Abstract*- We are in Big Data technology, through which we can do generic evaluation of specific elements within the big volume of data like, peta data. Where there are many challenges in bigdata processing like, scalability issues, heterogeneity, handling of exception, authentication, etc. with respect to from the acquisition of data to final out come of data. To over come above significant problem, the scientific requests are familiar throughout in the big applications with variety of domains and scientific tools are addresses on perspective to single province itself.**

Apache Hadoop provide one such solution for above problem, but by analyzing the solution we identify some issues. The better solution for the above issue we purposed a restful web services. The main importance of the Restfulll Web services is scalability and heterogeneity with a variety of platform. In addition by improving MapReduce technique in Reorder phase leads to get a desirable result. In short in this paper we design and develop a framework for Big Data analytics using RestFul Web Services

Keywords— restful Webservices, Bigdata Analytics, URL, Hadoop, Map reduce

I. INTRODUCTION

In this digital world, big data plays a very significant role in data driving decisions, innovation and productivity. The Bigdata & Analytics Program are changes accountability of information to contribute creative solutions to organisation by bringing new challenges of everyday.

The term *Bigdata* specifies the management of huge information and examination with technologies. Three standard traditions to identify the Bigdata technology are i.e.

1. Volume (quantity of data)
2. Velocity (speedup of data creation & transmission)
3. Variety (forms of ordered and unstructured information)

Volume: if the volume of information is bigger than tera bytes and peta bytes is nothing but a volumens of data.

Velocity: for the processing of bigdata we need tremendous amount of speed,

Variety: In Bigdata, data may contain assortment of origins, generally it might be, ordered semi ordered and unordered.

An ordered data might be easily sorted and inserted into warehouse but unordered data is randomised in nature and very hard to evaluate.

Semi ordered data do not have fixed fields.

II. BACKGROUND: BIG DATA ANALYTIC WITH HADOOP

Hadoop-MapReduce is a frame work, it executes the map and reduce work to distributed environment in server node, i.e. the significance of MapReduce algorithm is to process big voluminous data set over set of servers. The input to MapReduce algorithm is a file which is stored in DFS, where the cluster is split up into equal sized planned groups to facilitate and make simpler huge amount of data set into process large cluster of hardware in a trustworthy approach.

Mapreduce is a two phase data computation in Hadoop,

1. map phase
2. reduce Phase, i.e. voluminous data set is to be transformed into ordered key & value pair [3].

Figure 1 illustrates the data flow of MapReduce computation.

MapReduce process consist of following phase

- Splitting
- Mapping
- Reordering
  - Sorting
  - merging
- Reducing

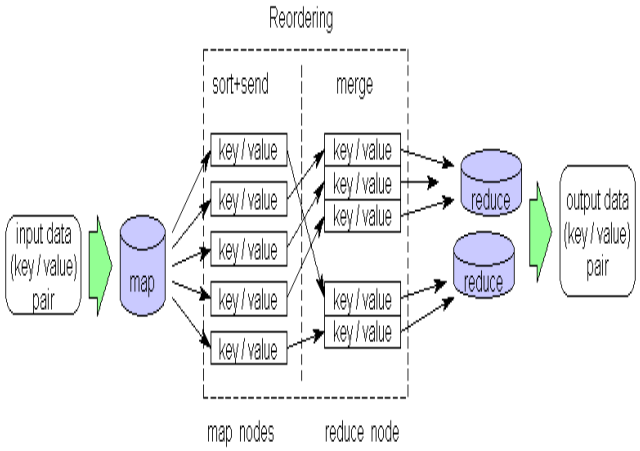

Fig 1 -- Data Flow of Mapreduce computation

Mapper will take the benefit of buffering then write it into the disk, where each mapper holds a circular memory buffer with a default size of 100MB, which can be tweak. When the memory buffer is filled to specific level, a dedicated thread is triggered, which spill the content from buffer to disk.

When the map tasker fit the data inside the reducer track memory, then the output files from different nodes in sorted map accomplishes the memory merge.

At the merge event data is applied to the reducer stage, which execute some supplementary data processing finally the data is going to be held in HDFS.

mapreduce algorithm work on key & value brace i.e.

- Input -- <key, value>
- Output -- <key, value>

Map-Reduce core functionality:

Code written in Java with the Hadoop Streaming API

Data flows of Mapreduce computation has 02 means of information, i.e. map-phase and reduce-phase, where,

map phase

- ireader – input data is divide into appropriate portions of breaks, then assign into map-function.
- mnode – master-node split the complex input data into small quires and assign into worker-nodes.
- map function – maps the data file into intermediate key & value pair.
- Worker-node - it executes small queries and sends back to master node.

reduce phase

- Reduce function – it consider intermediary value, reduces the solution and send back to the framework
- master node – consider the sub queries and join them in a specified method.
- owriter – writes file output

III. Big Data Analytics Using RESTful Web Services

A. RESTful Web services

webservice is a set of standard and protocol applied for exchange the data among various requests, i.e interoperability in applications like, Communication between two different programming languages like, java and python or java and c# dot net or application usage with different operating system like, windows-based and Linux-based applications.

RESTful architecture is used in the webservices application is known as RESTful webservices, basically REST is a web based standard architecture, for communication of data and RESTful webservices uses http Protocol.

In RESTful webservices architecture, REST server merely allows for access the entities, where as REST client access and present the entities. URI's identifies each and every entity, i.e. various resources, like Image, Text, json and xml, in which json is globally accepted resource set used for webservices individually.

http method	RESTful webservice method	Operation performed
POST	CREATE	Create a resources
GET	READ	Get resources
PUT	UPDATE	Update a resource
DELETE	DELETE	Delete a resource

Table2: HTTP methods in RESTful websevices

Apache hadoop allows for high performance native protocol for recovering HDFS,

i.e. Hadoop programs are executing within the Hadoop cluster, where all users can associate to HDFS from the outside world.

For the concept of proof, we have implemented an application Big Data analytics based on industry standard RESTful approach shown in figure 2.

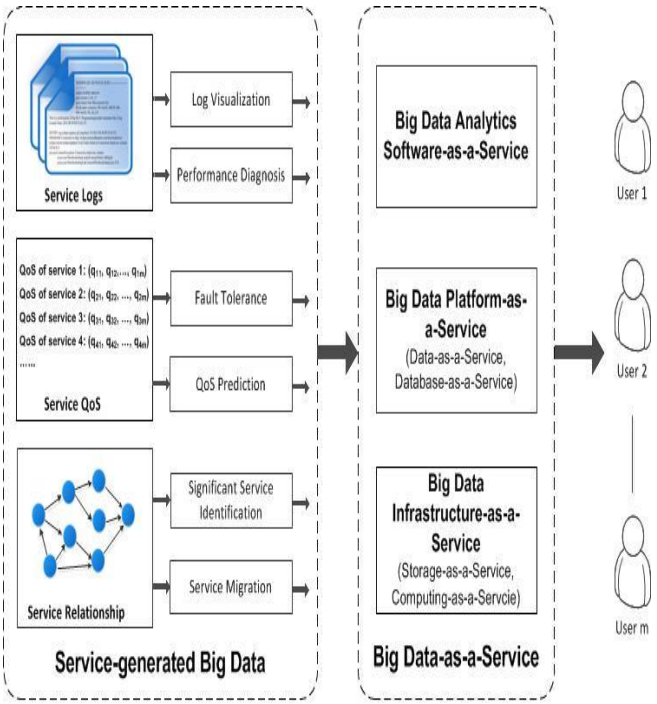

Fig 2—Architecture of services generated in Big Data

(note: reference [11])

Hadoop MapReduce framework uses reordering phase, which needs to merge and then sort all the generated keys at one place, which cause problem if the generated keys are large. Therefore, merging and sorting process should not conduct at single place.

TABLE1: CUSTOMERFORM.JSON

```
customerForm.json
{
"firstName": "Ashok",
"lastName": "Kumar",
"age":40,
"address":{
"streetAddress":"4th main",
"city":"Bengaluru",
"State":"Karnataka",
"Postal code":"560078" },
"phoneNumber":
[ { "type":"Home",
"number":"9845323230"} ]

}
```

In purpose "Big Data Analytics using RESTful Services" method we are going to provide such architecture, which is helpful to overcome recent problems.

The RESTful web services uses JSON data to transfer across nodes, a sample code of json specified in Table 1.

Here DataLine object is also mention to write data from client to datanode. The object contains Index property, which helps to maintain recovery management

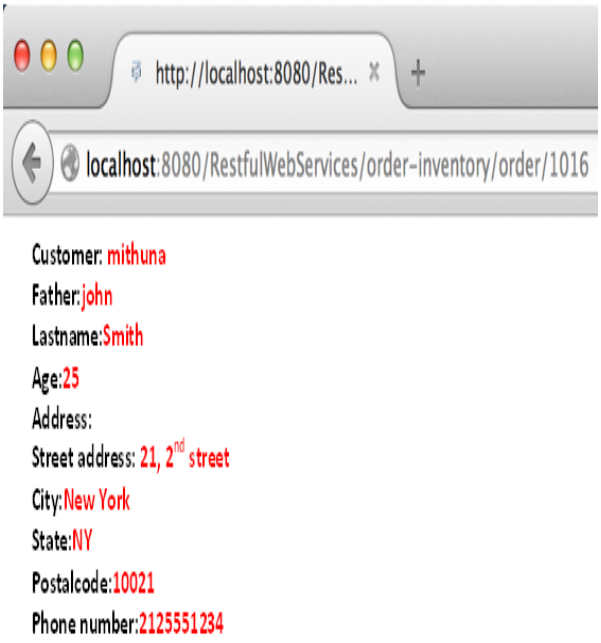

Fig3: sample shot of customer form

Schemes. Initial experiment shows that to mapping task takes 20 sec to maps 100 Kb file in single node with .NET environment. However, the reducer task takes generally 200 ms. Overall performance can be improved by adopting faster approach of exacting data from file to main memory. Currently major time consume on getting data from txt file to main memory.

IV. CONCLUSION

The data Analytics arena changing the responsibility of information research concept, it contributes the creative solutions to organization by bringing new challenges of everyday. For, the analysis of these collected data requires scalable, flexible, and well performed tools are required to provide analysis of data and perception in an appropriate manner. The outcome of this research paper is in the form of a design and development of naive framework for Big Data analytics based on Restful web services. Finally the ability of this framework design is to provide better scalability with in the component interaction and independent deployment of component for big data analytical tool.

REFERENCES

[1] http://gartner.com/it-glossary/big-data/

[2] J. Dean and S. Ghemawat. MapReduce: Simplified Data Processing on Large Clusters. In Communications of the ACM, 51 (1): 107-113, 2008

[3] Figure from slide deck on MapReduce from Google academic cluster, http://tinyurl.com/4zl6f5. Available under Creative Commons Attribution 2.5 License.

[4] http://www-01.ibm.com/software/data/bigdata

[5] http://www.nytimes.com/2012/03/29/technology/new-us-research-will-aim-at-flood-ofdigital-data.html

[6] Hadoop Map-Reduce flow, http://hadoop.apache.org/docs/r1.2.1/mapred_tutorial.html

[7] C. Eaton, D. Deroos, T. Deutsch, G. Lapis and P.C.Zikopoulos, Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data, Mc Graw-Hill Companies, 978-0-07-179053-6, 2012.

[8] S. Madden,"From Databases to Big Data", IEEE Internet Computing, v.16, pp.4-6, June 2012.

[9] S. Singh and N. Singh, "Big Data Analytics", 2012 International Conference on Communication, Information & Computing Technology Mumbai India, IEEE, October 2011.

[10] J. Dean and S. Ghemawat. MapReduce: Simplified Data Processing on Large Clusters. In Communications of the ACM, 51 (1): 107-113, 2008

[11] Zibin Zheng, Jieming Zhu, and Michael R. Lyu, Service-generated Big Data and Big Data-as-a-Service: An Overview, 2nd IEEE International Congress on Big Data, Santa Clara Marriott, CA, USA, June 27-July 2, 2013

[12] R. Cattell, (2010) "Scalable SQL and NoSQL Data Stores," ACM SIG-MODRecord, vol. 39.

[13] Scalable SQL and NoSQL Data Stores Rick Cattell Originally published in 2010, last revised December 2011

[14] Oracle White Paper— In-database Map-Reduce 2009

[15] KORNELIUSZ, "MongoDB vs MySQL: Comparison Between RDBMS and Document Oriented Database" MAY 2, 2014

[16] https://www.tutorialspoint.com/restful webservices

[17] Shilpa Patil, Ashok kumar P S, Prasadgouda Patil, "Data Analysis & Classification Methodology for Knowledge Discovery in Big Data", IEEE International Conference on Inventive Systems and Control, PP 1-4, ICISC-2017

[18] Gopal Krishna, Ashok kumar P S, Thirthe Gowda M T, "Significance of Restful web services in Bigdata Analytics", IEEE International Conference on Inventive Systems and Control, PP 11-14, ICISC-2017