Available online at http://www.journalcra.com

International Journal of Current Research
Vol. 7, Issue, 06, pp. 17301-17304, June 2015

INTERNATIONAL JOURNAL OF CURRENT RESEARCH

ISSN: 0975-833X

RESEARCH ARTICLE

ANALYTICAL SUMMARY OF COST EFFECTIVENESS USING HADOOP MAPREDUCE ON MARKET

***Surekha Lanka and Uzma Hashmi**

Faculty of Computing and Information Technology, King Abdul Aziz University, Jeddah, K.S.A

ARTICLE INFO

Article History:
Received 26th March, 2015
Received in revised form
12th April, 2015
Accepted 29th May, 2015
Published online 30th June, 2015

Key words:
Databases,
Key,
Value,
Map Reduce,
Framework,
Hadoop

ABSTRACT

In my paper I will evaluate features for processing Map reduce of Data science which ensures business alignment, growing skills, share knowledge and manage best practices and also cost effectiveness of the Hadoop. Data science is about to build data products not just answering the questions and products empower others to use the data. May help communicate your results, may empower others to do their own analysis actually data science is able to take data and able to understand it, to process it. To extract value from it, to visualize it, to communicate it that's going to be a huge imprint skill

Citation: Surekha Lanka and Uzma Hashmi, 2015. "Analytical summary of cost effectiveness using hadoop mapreduce on market", International Journal of Current Research, 7, (6), 17301-17304.

INTRODUCTION

Huge Big data reside of large data sets are not performed by the normal computing techniques. There are three challenges of big data, Volume comes in the size of the data, velocity for the latency of data processing relative to the growing demand for the interactivity, variety that diversity of sources, formats, quality, structures. Data science is about data products. Data - driven apps like spellchecker, machine translator. Interactive visualizations like Google flu application, global burden of disease. Online databases like enterprise data warehouse and Sloan digital sky survey. Data science is about to build data products not just answering the questions and products empower others to use the data. May help communicate your results, and may empower others to do their own analysis. By the figure X axis we take date sources and Y axis for bytes. Here map rides different fields of study by indicating 3 v's.

Hadoop Map Reduce

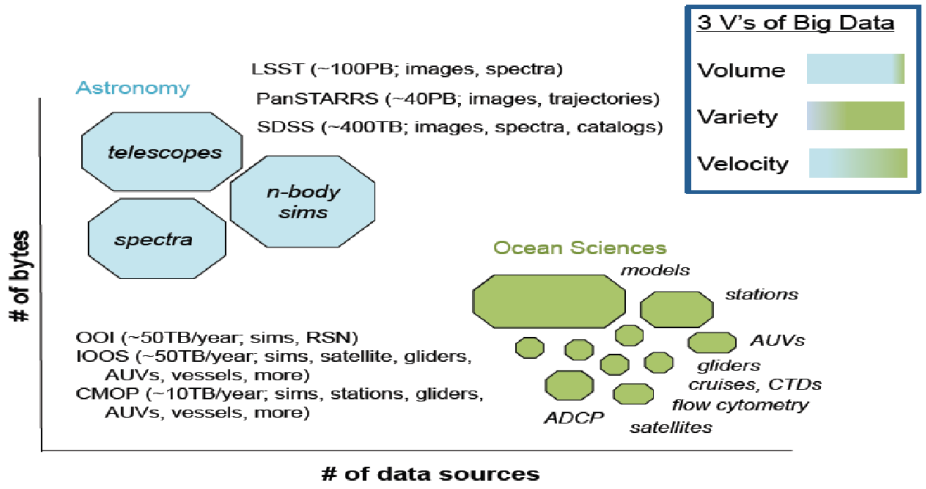
This is an open software work which was implemented in Java for saving and performing on big data in the distributes on large sets the two main components of Hadoop are Hadoop Distributed file system and Map reduce. So in this paper, we are talking about map reduce. Hadoop map reduces task is to split the input data into various chunks, those are handled by maps tasks in a special way. The map reduces framework works on <key,value> pairs. Framework views the input of the <key, value > and produces a set of <key,value> as an output of the task. The Programmer specifies two functions

```
map (in_key, in_value) -> list (out_key, intermediate_value).
reduce (out_key, list(intermediate_value)) -> list(out_value)
```

```
map(String input_key, String input_value):
// input_key: document name
// input_value: document contents
for each word w in input_value:
EmitIntermediate (w, 1);
reduce(String output_key, Iterator intermediate_values):
// output_key: word
// output_values: ????
int result = 0;
for each v in intermediate_values:
result += v;
Emit(intermediate_key, result);
```

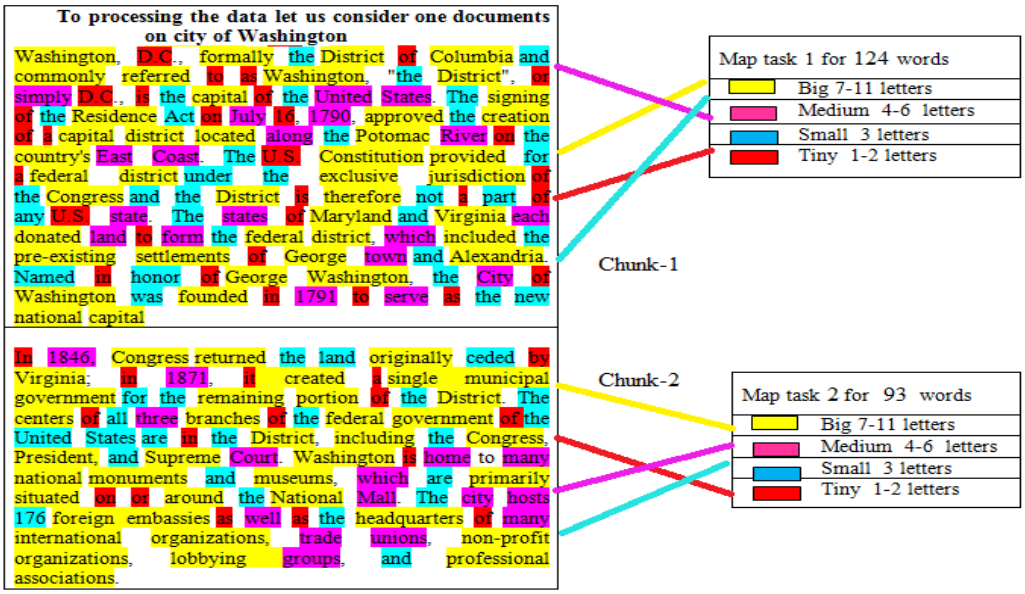
Maps are the individual tasks that transfigure the input records into intermediate records. These records do not need to be of the same type as the input records. A given input pair may map to zero or many output pairs.

Corresponding author: Surekha Lanka,
Faculty of Computing and Information Technology, King Abdul Aziz University, Jeddah, K.S.A.



To processing the data let us consider one documents on city of Washington

Washington, D.C., formally the District of Columbia and commonly referred to as Washington, "the District", or simply D.C., is the capital of the United States. The signing of the Residence Act on July 16, 1790, approved the creation of a capital district located along the Potomac River on the country's East Coast. The U.S. Constitution provided for a federal district under the exclusive jurisdiction of the Congress and the District is therefore not a part of any U.S. state. The states of Maryland and Virginia each donated land to form the federal district, which included the pre-existing settlements of George town and Alexandria. Named in honor of George Washington, the City of Washington was founded in 1791 to serve as the new national capital. In 1846, Congress returned the land originally ceded by Virginia; in 1871, it created a single municipal government for the remaining portion of the District. The centers of all three branches of the federal government of the United States are in the District, including the Congress, President, and Supreme Court. Washington is home to many national monuments and museums, which are primarily situated on or around the National Mall. The city hosts 176 foreign embassies as well as the headquarters of many international organizations, trade unions, non-profit organizations, lobbying groups, and professional associations.



Top sub industries with Hadoop framework [2]

RANK	INDUSTRY
1	Business intelligence, analytics & performance management
2	Advertising, sales and marketing
3	Advertising network or exchange
4	Monitoring and security
5	Social
6	Education and training
6	Customer relationship management
6	Data and document management
6	Database management
6	Music
10	Video
10	Gaming

(http://www.alliedmarketresearch.com/hadoop-market) by the figure we consider a document on Washington city. This document implemented given to map reduce (key, value) and finding out how many Big, medium, small and tiny words are there in the document. Map reduce made the document into chunks 1 and 2 to find the words as given by (key value). For ex: (yellow, 37). The Map Reduce is further integrated by the companies are required to make a new skill base where the capital investments rapidly outweigh the investment infrastructure. The existing data warehouse subsequent to the business intelligence has to be leveraged that requires the Big Data with Hadoop to integrate to ensure leverage in the skills and existing tools.

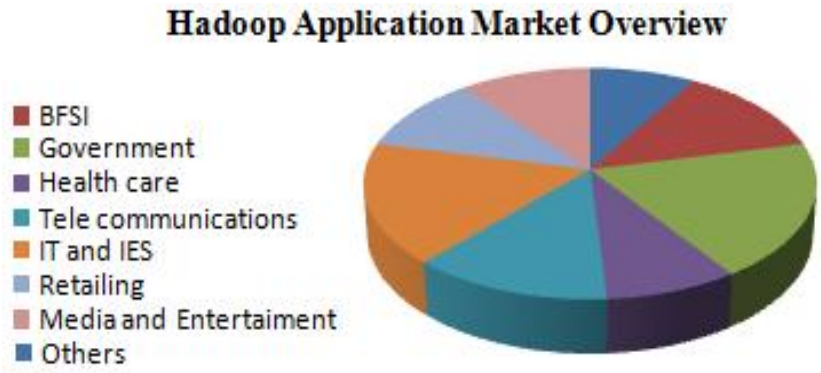
Global Hadoop Market

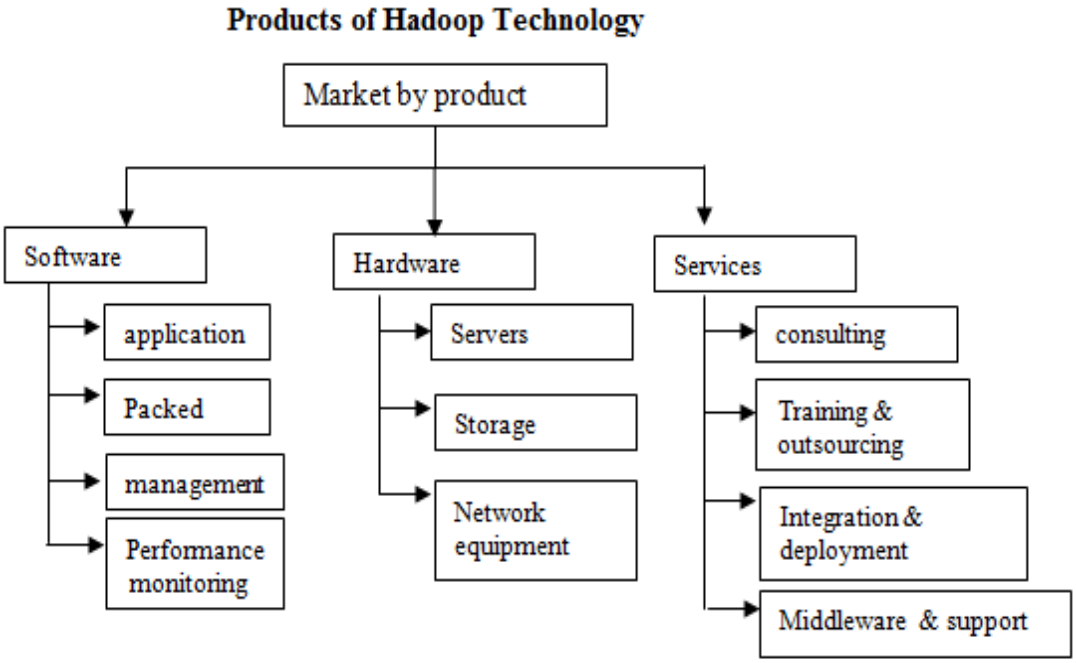
Hadoop market is expanding widely at an outstanding rate. When compared to the regular data analysis tool like RDBMS, Hadoop technology provides Actual cost effective and quick results. Many companies in programming languages of the tech stack data, main frameworks and databases now on Hadoop technology. This open source framework can be accessed for the Distributed processing and analytics of immense data set with clusters of server (http://www.alliedmarketresearch.com/hadoop-market).

Now the Hadoop platform is used by companies in a range of markets like advertising, education & training, data and document management and social, music, gaming, videos, sales and marketing tech and advertising networks and exchanges (http://www.alliedmarketresearch.com/hadoop-market).

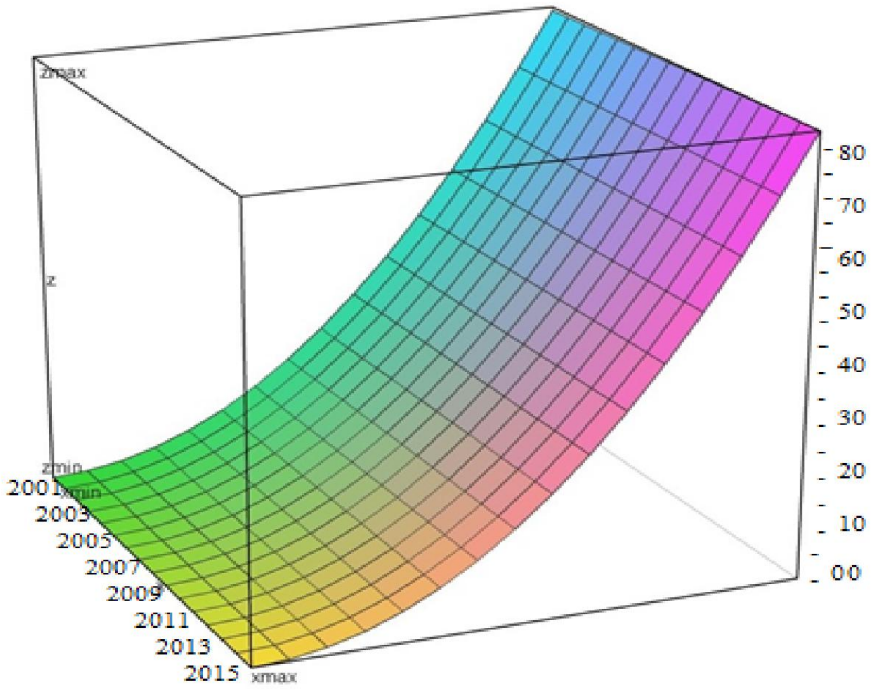
Hadoop software, hardware and services market analysis

This software market again categorized into several software applications, package software, management software and performance monitoring software. Now presents days software application is the largest section in the software market of Hadoop. North America excepted to lead the revenues, especially in this market because technology adoption is more. Indian big data industries are estimated to increase fivefold from the current level by 2015 (http://www.alliedmarketresearch.com/hadoop-market). Coming to the Hardware market is sectioned into servers, storage and network equipment. here storage market is the largest section the hardware market.the Hadoop services market is sub categorized into consulting services, like Hadoop training and outsourcing. Consulting services market making largest revenue compared to all Hadoop services.





Current Hadoop market scenario



The Cost Effectiveness of Hadoop Big Data

Rapid economic growth of Hadoop market, actually in big data analytics over based tool of data analysis provides cost effective and fast data processing. Hadoop Market observed it and make it has better future prospects in all the sectors, Hadoop provides a much better solution over RDBMS (Relational Database Management System) and other traditional tools (http://www.alliedmarketresearch.com/hadoop-market#). The cost effective scalability of Hadoop is the key aspect of the Big Data. The example can be taken as the cluster of the peta byte Hadoop that requires 125 to 250 nodes that has the cost of $ 1 million. Similarly, the annual cost of the supported Hadoop distribution will be around $4000 per node that is only a small enterprise fraction of data warehouse $10 to $100 million. The Big Data on the Hadoop may seem a great deal, however the innovative enterprise already has Hadoop today. The point of concern is how to leverage and what pace it will evolve as a critical mission and the focus of IT (Assunção *et al*., 2014). The graph shows us the Hadoop market scenario from 2001 – 2015.

Job Opportunities with Hadoop

Coming to the market of big data, it seems to be promising and the rising trend will keep progressing with time. So the job trend or Market is not a short lived phenomenon as Big Data. Hadoop provides and the clear carrier with advantages like accelerated career growth and increases pay package for skills on Hadoop. And its technologies are here to stay. Hadoop is a potential to pop up job rules, whether you are a fresher or an experienced professional.

Conclusion

Map Reduce jobs are easy to run a programmer can implement with his known language like Java, python, c++ and map reduce recovery the failures and it is fast responding to the queries. The main advantages are key-value stores have implemented much nicer support for fast and scalable random reads and writes. So Hadoop is the having high demand in the mark so The digital world of today has stepped into the era of Big Data where business opportunities are rising on a daily basis due to the innovative technologies introduced for data management that enables the organizations to examine their data. The unstructured and semi structured data generated in tremendous quantity at the speed of the network, which are information sources that identifies the need of the customers to the organization.

REFERENCES

Assunção, M. D., Calheiros, R. N., Bianchi, S., Netto, M. A. and Buyya, R. 2014. Big Data computing and clouds: Trends and future directions. *Journal of Parallel and Distributed Computing*.

Brown, B., Chui, M. and Manyika, J. 2011. Are you ready for the era of 'big data'. *McKinsey Quarterly*, 4, 24-35.

http://hadoop.apache.org/docs/r1.2.1/mapred_tutorial.html#Source+Code

http://www.alliedmarketresearch.com/hadoop-market

http://www.alliedmarketresearch.com/hadoop-market#

http://www.cloudera.com/content/cloudera/en/documentation/hadoop-tutorial/CDH5/Hadoop-Tutorial.html

http://www.edureka.co/blog/5-reasons-to-learn-hadoop/

https://www.cbinsights.com/blog/big-data-hadoop-vc-companies/

****