Advances in Computer Science and Information Technology (ACSIT)
Print ISSN: 2393-9907; Online ISSN: 2393-9915; Volume 1, Number 3; November, 2014 pp. 103-106

http://www.krishisanskriti.org/acsit.html

Reflection Seismic Data Analysis using Big Data

Manish Gupta[1], Gourav Gupta[2], [2]Mukesh Kumar Sharma and [3]Inder Singh

[1]Material Sciences Research Lab, ITMU, Huda Sector 23A,Gurgaon, Haryana
[2] Department of Computer Science, TIT&S, Birla Colony, Bhiwani, Haryana
[3]Department of Mathematics, J.V.M.G.R.R. College, Charkhi Dadri, Haryana

Abstract

In this paper sincere attempt has been made to examine the huge amount of reflection seismic data with the help of high computational Big Data techniques i.e. (HiveQL and Apache Hadoop with MapReduce).This technique also used to scrutinize the Data for the prediction of different types of mineral rocks, oil and natural gases beneath the earth.

Keywords*: Big Data, Cloudera,HiveQL, Hadoop, MapReduce, Reflection Seismic Data.*

1. INTRODUCTION

Drilling is very expensive, risky and time consuming process to detect the oil/gas well, valuable minerals lying beneath the earth. Recently Seismic imaging technique is use to identify the materials beneath the earth.As the propagation of seismic waves gives many information regarding the medium in which they travel. The speed of the wave depends not only material property of the medium but also the angle of incidence. Reflection data are used to prospect petroleum, natural gases and investigate earth's internal structure.

With the help of reflection and refraction of seismic waves at an interface give the information about the density and thickness of the reflecting rock. Every time the seismic pulse meets a change in rock properties .If the velocity and time are known then we can find out the depth of the event. In seismic Surveys, reflected waves are combined and interpreted on seismograms. These data can also be obtain by artificial technique like land data acquisitions, marine data acquisitions and Transition zone recording instead of occurrence of earthquake.

About 1970's there was a new era of Geodesy due to the rapid development of GPS technology. Weijun and Gang [17] represented a new model based on GPS technology, that can improve Earthquake monitoring and forecasting.

In the middle of 1990's, Earthquake disasters monitoring and forecasting has been an important topic for research point of view by the Data Scientists. In the last of 1990's, IBM define the concept of Big data on the basis of 3V's i.e. volume, velocity and variety. After that Google has extended this concept 3 V's to 5 V's and used firstly this idea in search engine optimization. Many companies like Amazon, Microsoft, Yahoo and Facebook etc. have been working on the above topic since then. They have contributed in a wide range towards its application in various fields e.g. in Geophysical and petroleum industries[1,7], Oil and Gas Companies [2,11],Monitoring of assets and environmental conditions[12], Geophysics Aerodynamics[8],Health care application[9] etc.Big Data also play an significant role in seismic Data processing .

2. BIG DATA DEFINITION

In 1990s after a pioneer work by Meta Group[3],we can say that Big data is the term for a collection of data sets so large and complexthat it becomes difficultto process using on-hand database management tools or traditional data processing applications. Fig1. shows block diagram of 5V's of Big Data. Big Data is a conventional data base system, when our data is big, moves too fast, or doesn't fit the structures of our data base architectures[5,6].

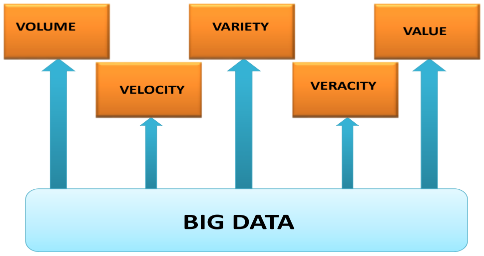

Fig. 1: Block diagram of '5V' of Big Data

3. BIG DATA FOR SEISMIC DATA ANALYSIS

These techniques are primarily used by exploration and production companies in order to locate oil and natural gas deposition . Fig2. shows the diagram of Marine seismic survey set up .

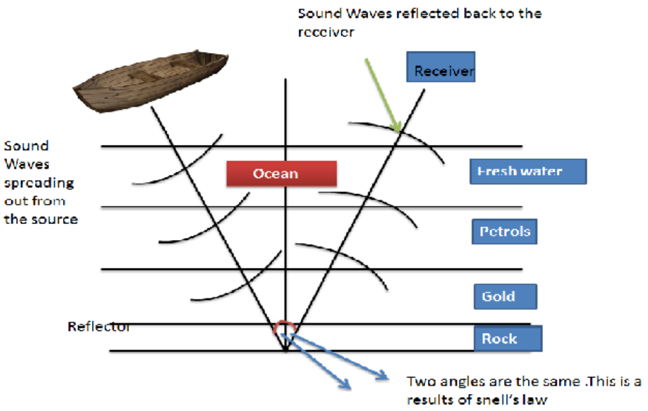

Fig. 2: Diagram of Marine seismic survey set up.

In this process, seismic data is recorded from probable area where Oil and Gas deposits. Seismic Waves are generated through a source by means of explosion .The amplitude and arrival times of the reflected Waves at the interfaces between rock layers are recorded by receiver .By comparing the time it took for the seismic waves to travel from different source and receiver locations, we can locate the depth of the common subsurface point that the waves reflected. For more detail we can see on Seismic Data Processing[13] .

Rizvandi and Zomaya [18] used MapReduce technique to analyze seismic images data and explained to fit PKTM (Prestack Kirchhoff Time Migration) algorithm for running on Google’s MapReduce framework. The relation between PKTM completion time and the number of map/reduce tasks on pseudo-distributed MapReduce mode was analyzed by them.

Recently, Cloudera group developed a Seismic Hadoop project to demonstrate how to store and process Seismic data in a Hadoop cluster. Systematic mechanism is given below

4. FLOW CHART DIAGRAM

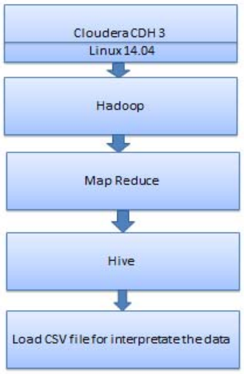

Fig. 3: Flow chart of Cloudera (CDH3).

Fig. 3. Shows the step by step flow chart of simulation work .Cloudera (CDH3) VM[14] have been install on the windows 7 .Cloudera (CDH3) is open source platform and associate with large numbers of technologies Big Data,Hadoop etc. It also contains the linux environment .On the top of the CDH3 Apache Hadoop based framework is set .Apache Hadoop[10] is uses for distributed processing of huge data sets across the clusters of computers using simple Map reducing programing Model. Mapreduce is written in Java programming language .CSV data files have been analysis using HiveQL [15]at the top most part of the Hadoop .

5. SIMULATION WORK

For Reflection Seismic data interpretation we have taken the data recorded by Anderson and Lieberman [16] .These collection of data have store as in CSV file format. These data contain the informations about density (g/cc) of different minerals types, rock types and liquids types under the Sea water. HiveQL[15] along with Apache Hadoop[10] and Map-Reduce have been used to analyze huge amount of reflection Seismic data.All the simulation work performed by Hive and Hadoop on the top of the Cloudera CHD3 VM .Following are the steps for simulation .

1. Installation of Cloudera CH3 on windows.
2. Start Hive services and and ensure that Hive daemons are running in your Hadoop cluster.
3. Use HiveQL to create a table with same column name as given in CSV file.
4. Load the CSV file in Hive table.
5. Execute the HiveQL query to get the desired results.

6. CREATE HIVE TABLE

```
Hive>create table seismichive (mineral_typeString,Rock_typeString,liquid_typeString,Density double)

Row format delimited fields terminate by “;”;
```

Load Data to Hive Table

```
Hive>load data local in path ‘/home/cloudera/seismicdata.csv’ into table seismichive;
```

Calculate the minerals Types,Rock types and Liquid types

```
Hive>select Mineral_type,count (mineral_type)from seismichive Group by mineral_type;
```

Output

Mineral Type	Density(g/cc)
Pyrite	4.93
Magnetite	5.20
Haematite	5.12
Quartz	2.65

selectRock_type,count (Rock_type)from seismichive Group by Rock_type;

Output

Rock Type	Density(g/cc)
Dolomite	2.84
Limestone	2.73
Sandstone	2.65

selectLiquid_type,count (Liquid_type)from seismichive Group by Liquid_type;

Output

Liquid	Density(g/cc)
Water(20% NaCl)	1.14
Water Fresh	1.00
Oil	0.80-0.85

7. CONCLUSION

In this paper, using reflection seismic data with the help Hadoop technique we can differentiate particular minerals, oil, rocks and salts beneath the bottom of the sea. Simulation work along with example has been presented. Undoubtedly, we have progress a lot but we have still ‘miles to go’. We have a vision of successfully practically implementation of the developed model in large scale and for it we required a lot of research and international cooperation .

REFERENCES

[1] A. Baaziz and Luc Quoniam, “How use Big Data Technology to optimize operations in upstream Petroleum Industry,”International Journal of Innovation,Sao Paulo,V.1(1),pp., 30-42,jan/dec 2013.

[2] M.Brule”Tapping the power of Big Data for the oil and gas industry”, IBM Software white paper for petroleum industry, May 2013.

[3] M. Schroech, R.Shockley, J.Smart,D.R. Morals, P.Tufano, “Analytics : The real-world use of big data, ”IBM Global Services,pp., 1-20,October 2012.

[4] E.Pednault, “Big Data platforms,Tools and Research at IBM,”IBM Research, pp.,1-35,2011.

[5] J.S., Ward and A. Barker, “Underdefined by Data: A Survey of Big Data definition,” 11 september 2013.

[6] G .Wang,T.S. Eugeneng, A.Shaikh, “Programming your Network at Run-time for Big Data Application,” ACM,Finland,2012.

[7] A.Hems, A.Soofi, E.Peren, “How innovative oil and gas companies are using Big Data to outmaneuver the competition, ” A Microsoft white paper,pp.,1-12,May 2013.

[8] Jonathan, “Data Science and Aerodynamics,” 1987.

[9] B.Hamilton, “Big Data is the future of Health Care,” Cognizant 20-20 insights,pp.,1- 6,September 2012.

[10] http://hadoop.apache .org/

[11] K.Hollaway, “Optimize Exploration and Production with Data – Driven model,”JohnWiley,North Carolina,USA,2014.

[12] J.Feblowitz, “The Big Deal about Big Data in Upstream oil and gas, ” paper &presentation, IDC Energy Insights, October 2012.

[13] J.Will, “Seismic Data Science reflection Seismology and hadoop,” jan 2012.

[14] http://www.cloudera.com/content/cloudera/en/home.html

[15] http://docs.treasuredata.com/articles/hive

[16] A.M.A. Selim, A.A.Ali and AM Abdel Salam, “The Geological Interpretates in Reflection Seismic Data,”Thesis,2009.

[17] S. Weijunand L Gang, “Research of Gps technology in Seismic Monitoring System,”
the 14th world conference on earthquake engineering,Beijing, China,12-17 oct.,2008.

[18] N.B. Rizvandi and A.Y. Zomaya,”High Performance Computing of Seismic Data on Mapreduce,”School of information technology, The University of Sydeny.

, ,