ISSN : 0974 - 7435 Volume 10 Issue 23

2014

BioTechnology

An Indian Journal

FULL PAPER

BTAIJ, 10(23), 2014 [14637-14643]

Economic effect of big data compressed storage technology in Rockburst experiment

Yan-ping Bai[1], Yu Zhang*[2,3]

[1]College of Management, Capital Normal University, Beijing, (CHINA)

[2]School of Computer Science, Beijing University of Civil Engineering and Architecture, Beijing 100044, (CHINA)

[3]State Key Laboratory for GeoMechanics and Deep Underground Engineering, China University of Mining & Technology, Beijing, 100083, (CHINA)

E-mail : zhangyu@bit.edu.cn; yibaibaby@126.com

ABSTRACT

State Key Laboratory for GeoMechanics and Deep Underground Engineering (GDLab) has accumulated more than 500 TB data of Rockburst experiment. But so far the amount of analysed data is less than 5% in GDLab. Data storage dilemma is restricting the study mechanism of Rockburst. In this paper, we applied big data technology into analyse of Rockburst, and makes deep analysis about characteristic of Rockburst data. Basing on this, a big data based data storage systems (BDSS) for Rockburst experiment was proposed. BDSS based on Hadoop for Rockburst with online data loading and rapid retrieval of data. In Storage node machine cluster in BDSS, Big Data Compressed Storage Algorithm was proposed. The algorithm can provide average compressed ratio about 3.26%. Experimental analysis shows that the algorithm has excellent performance in Rockburst and solves the Data storage dilemma. Research work of this paper laid some foundation of Rockburst.

KEYWORDS

Economic effect; Big data; Compressed storage technology; Rockburst.

INTRODUCTION

Rockburst is a kind of geological disaster in underground engineering with excavation, which has become one of worldwide underground engineering problems[1]. In 2006, State Key Laboratory for GeoMechanics and Deep Underground Engineering, GDLab for short, successfully reproduced process of Rockburst in indoors. Since then, research on the mechanism of Rockburst rose to a new level. A lot of research work on the mechanism of Rockburst has been done by GDLab, and make a series of valuable research results[2-6]. But there are some dilemmas during the research progress. A most important one of the dilemmas is data storage dilemma. Data storage dilemma is because of a large number of experimental data is produced in Rockburst research. Furthermore, this data is inevitable and determined by the characteristic of Rockburst.

For example, one Rockburst experiment numbered as “yqsii6#”, it generated 33217 txt files in an hour. Hard disk needs 12GB storage space to save all these 33217 files. Along with the further research work, the data obtained also growth with geometric speed. The GDLab has accumulated more than 500 TB of data, but in so far, its analysis ratio is no more than 5 percent.

In the face of mountains of data, only effectively manage these data and analysis these data using modern means, which can establish theoretical basis of Rockburst mechanism and forecast. This is the focus of this paper.

In a variety of types of data[7], quickly acquiring the ability value information is the big data technology. Big data have the characteristics of four aspects:

1): Volume: From the traditional GB level, jumped to the TB level, even PB level;

2): Variety: The Internet of things, cloud computing, mobile Internet, network, mobile phone, tablet computer, PC and various sensors spreading all over each corner of the earth are a data source.

3): Value: Valuable information needs to be extracted from the mass of data.

4): Velocity: Fast processing speed, this is essentially different from the traditional data mining technology .

Big Data" as a new concept [8], now is not directly given the proper noun by Chinese government and given policy support. However, network plan on December 8, 2011, 4 key technologies as information processing technology One of the innovation projects has been put forward by engineering and information department, including the mass data storage, data mining, image and video intelligence analysis and so on, which are important components of big data. While the other 3 key technical innovation projects, respectively is the information technology, the letter information transmission technology and information security technology, these 3 technologies are closely related with the "Big Data".

OUR APPROACH

Rockburst big data

In recent years, GDLab has done a lot of research works, and accumulated a great amount of research data at the same time. For example, one Rockburst experiment numbered as “yqsii6#”, it generated 33217 txt files in an hour. Hard disk needs 12GB storage space to save all these 33217 files. Along with the further research work, the data obtained also growth with geometric speed. How to scientifically managing and using these data become the biggest problem which GDLab facing now.

Only effectively managing and using the modern computer method for analysis and processing and extraction these data, we can better study phenomenon of Rockburst and explore the prediction mechanism of it. So it can avoid the damage of Rockburst, at the same time bring huge economic and social benefits.

Collection of rockburst big data

System collect signal using two types of sensors from Pengxiang and PAC. Parameters of the sensors are shown in figure 1.

Figure 1 : Two types of sensors and their parameters

Massive-data dilemma

Take the same example of “yqsii6#”. As we mentioned above:

1): It generated 33217 txt files in an hour. Hard disk needs 12GB storage space to save all these 33217 files.

2): The size of above four pictures in JPG format is 294 KB and they need 300KB storage space in hard disk to save them.

We can get the conclusions as below:

1): Data size of "yqsii6#" for an hour can be calculated by formula 1.

12GB+4*300KB*33217=51.86GB (1)

2): According to the mainstream desktop hard drive with 2t~3t capacity in the market, it will be occupied within 39hours~48hours only by one experiment such as "yqsii6#". As shown in table 1, it involved 268 samples. We can't image how big the number of data is. But we can see it should be growth with geometric speed.

Here we take three typical Rockburst experiments as example, studying the characteristics of Rockburst experimental data. In the following part of this paper, we will use the same experiments too. Data storage details of three typical Rockburst experiments are described in table 1.

Table 1 : Data storage details of three typical Rockburst experiments

Experiment No.	Number of TXT files	Occupied Disk Space (MB)
O-1#	29438	1536
G-1#	41645	2161
GO-1#	71351	3748

From table 1 we can see that each experiment has how many txt files and how large the disk space it occupied.

BIG DATA COMPRESSED STORAGE ALGORITHM IN ROCKBURST EXPERIMENT

As we mentioned above, also different experiment has different number of txt files, each txt file has accurate 4096 values. Under this principle, each experiment may have hundred millions values as showed in table 2. It's very huge.

Table 2 : Data storage details of three typical Rockburst experiments with values

Experiment No.	Number of TXT files	Number of values	Occupied Disk Space (MB)
O-1#	**29438**	**120578048**	**1536**
G-1#	**41645**	**170577920**	**2161**
GO-1#	**71351**	**292253696**	**3748**

It's impossible to analyse these data in the past. We proposed a big data based algorithm solving this dilemma as below.

System structure

We have put forward a big data based data storage systems for Rockburst experiment, BDSS for short, which based on Hadoop for Rockburst with online data loading and rapid retrieval of data. The structure of BDSS is showed in figure 2.

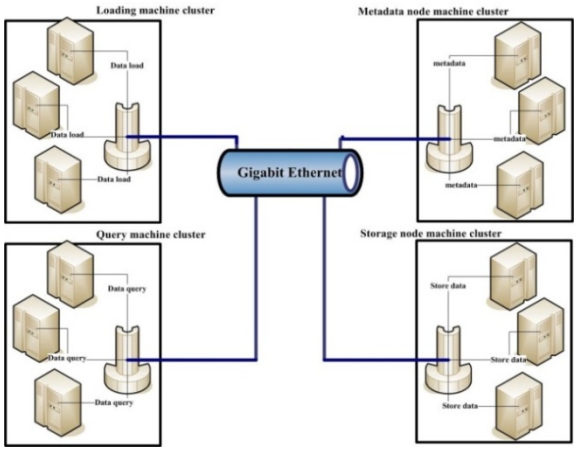


Figure 2 : System structure of BDSS

As illustrated in figure 2, Storage node machine cluster is at the lower right corner. The storage node machine cluster provides persistent storage capacity, long-term preservation of historical data. System storage data sources as small block usually take once or several times from the loading machine cluster as the data block unit.

Under such mechanism, we got some law of experiment data as illustrated in section 3.2.

Big data compressed storage algorithm

We proposed and implemented the big data compressed storage algorithm. Here we illustrated some key principles of the algorithm.

1) Combined all txt files into a new one txt file.

2) New txt file is combined by triples {X, Y1Y2Y3Y4, Data}.

X represents the txt file's sequence number is X.

Y1Y2Y3Y4 is a number with 4 bits. It represents the line number in a txt file.

Data is the specific data in original txt files.

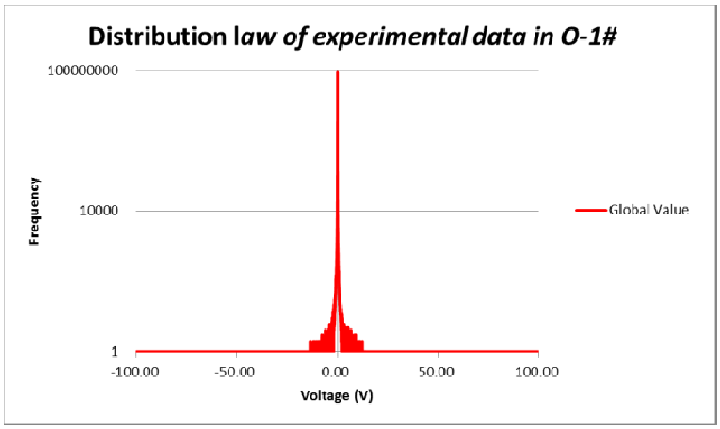
So, {X, Y1Y2Y3Y4, Data} is represents the specific data in line Y1Y2Y3Y4 in Xth txt in one experiment.

3) If the data is zero, we simply discard the value.

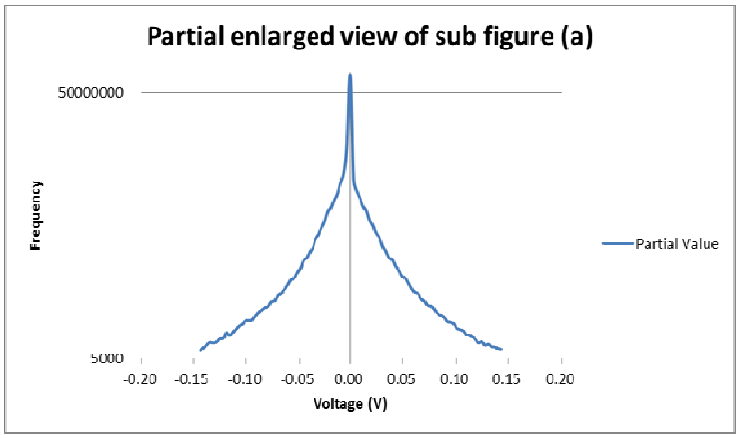
EXPERIMENTS AND ANALYSIS

Law of experimental data

Firstly, we separately got the distribution law of experimental data in three experiments as showed in figure 3, 4 and 5. Each figure has two sub figures. Sub figure (a) is the distribution law of experimental data, and sub figure (b) is the partial enlarged view of sub figure (a).

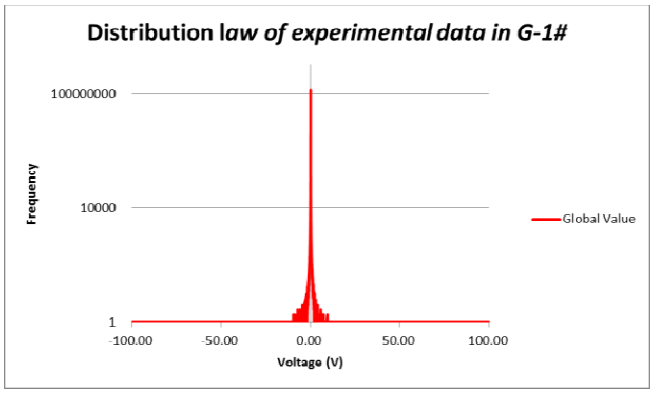


(a)

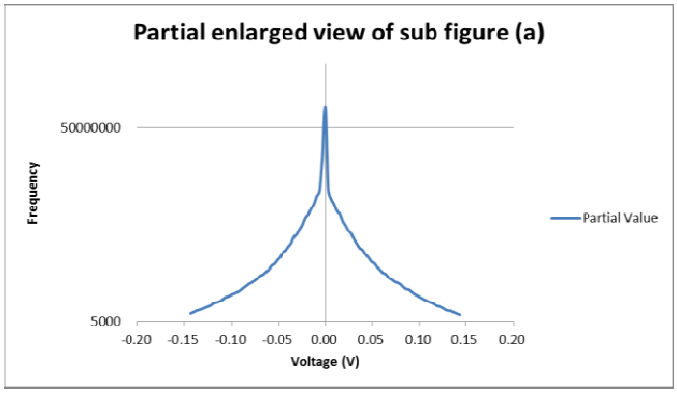


(b)

Figure 3 : Distribution law of experimental data in O-1#

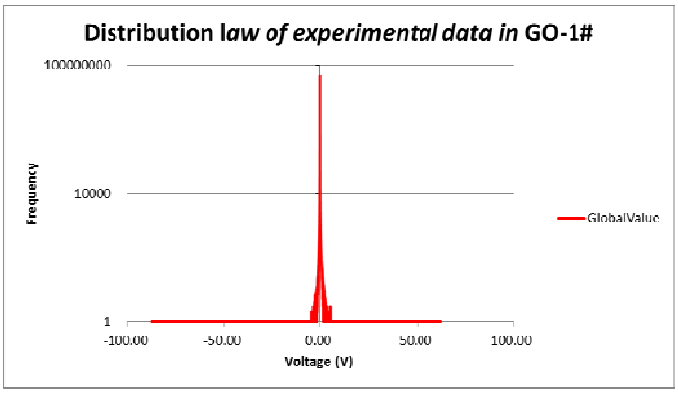


(a)

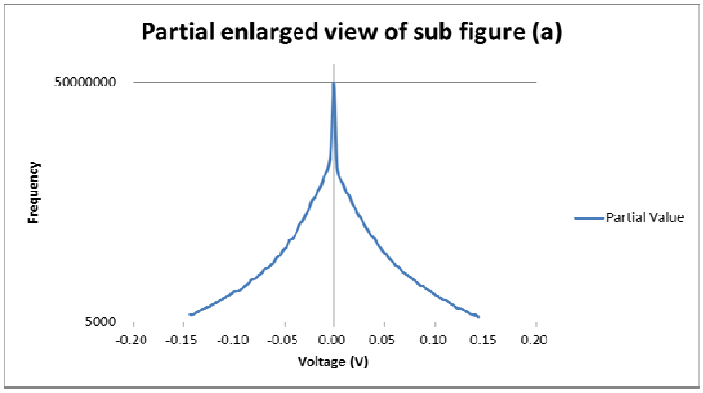


(b)

Figure 4 : Distribution law of experimental data in G-1#



(a)



(b)

Figure 5 : Distribution law of experimental data in GO-1#

From all above we found that the distribution law of experimental data is a regular distribution. Then we made some further research on it, and proposed the big data compressed storage algorithm in section 3.2.

To verify the efficient of the algorithm, we conducted a series of experiments as below.

Compressed ratio of big data compressed storage algorithm

We compared occupied disk space before compressed and after compressed with Big Data Compressed Storage Algorithm. The results are listed in table 3.

Table 3 : Data storage details before and after compressed under the algorithm

Experiment No.	Occupied Disk Space (MB)-before compressed BC	Occupied Disk Space (MB)-after compressed AC	Compressed Ratio =AC/BC
O-1#	1536	51	3.32%
G-1#	2161	56	2.59%
GO-1#	3748	145	3.87%

From table 3 we can get the conclusion that big data compressed storage algorithm has significant efficient, which means the average compressed ratio is 3.26%. In figure 6, we compare occupied disk space between original data and after compressed by our algorithm.

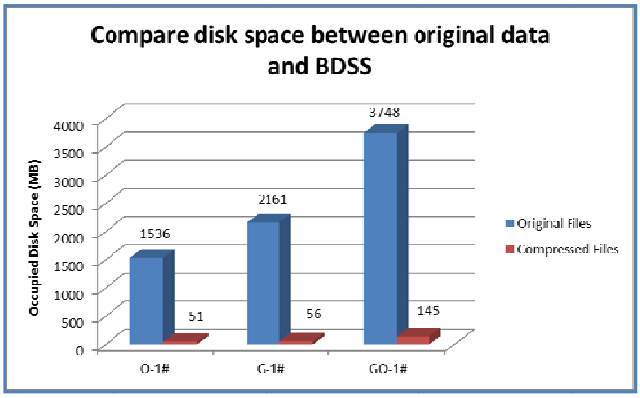


Figure 6 : Compare disk space between original data and BDSS

CONCLUSIONS

From all above, we can draw the conclusions that Big Data Compressed Storage Algorithm is suitable for Rockburst, and the algorithm has excellent performance.

In the future, we will go on analysis the distribution law of experimental data and improve our algorithm to get a better compress ratio than ever.

ACKNOWLEDGMENTS

This work was supported by National Natural Science Foundation of China under Grant No. 51444003; and also supported by 2014 Technology Projects of Beijing Education Committee (No.KM201410028011); and also supported by 2012 Outstanding Talents Projects of Beijing Organization Department (No.2012D005017000005); and also State Key Laboratory for GeoMechanics and Deep Underground Engineering, China University of Mining & Technology (No. SKLGDUEK1215); and also supported by projects of BUCEA Y12-35 and No.00331613033; and finally supported by 2014 Exllent Teacher Training programme granted by Beijing Teachers Higher Education Training Center 067145301400

REFERENCES

[1] F.JIANG, et al., "The Current Statues and Developing Tendency of Rock Blasting Forecast at Home and Abroad," [J]. Industrial Safety and Environment Protection, 29(8), 19-22, (2003).

[2] M. HE, et al., "Experimental study of rockbursts in underground quarrying of Carrara marble," International Journal of Rock Mechanics and Mining Sciences, 52, 1-8, (2012).

[3] M. He, et al., "Rockburst process of limestone and its acoustic emission characteristics under true-triaxial unloading conditions," International Journal of Rock Mechanics and Mining Sciences, 47, 286-298, (2010).
[4] M. He, et al., "Study on Rock Mechanics in Deep Mining Engineering," Chinese Journal of Rock Mechanics and Engineering, 24, 2803-2813, (2005).
[5] M. He, et al., "Experimental Investigation of Bedding Plane Orientation on the Rockburst Behavior of Sandstone," Rock mechanics and rock engineering, 1-16, (2012).
[6] M. He, et al., "Research and Monitoring Mechanisms of deep coal mine disaster," Coal Science & Technology Magazine, (2007).
[7] KE P, LI Y, NI F. An evolvable cellular automata based data encryption algorithm, International Journal of Wireless and Mobile Computing, , 6(1), 66-71, (2013)
[8] WHITE T. Hadoop: The definitive guide, O'Reilly Media, Inc., (2012).