An Approach for Processing Underground Spatial-Temporal Data by Cloud Computing

Renfeng A. Xu
Shenzhen Polytechnic 518055 Shenzhen, China
Email: xurf@szpt.edu.cn

Dingju B. Zhu
Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China
School of Electronic Engineering and Computer Science, Peking University- 100871 Beijing, China
Email: 102268832@qq.com

***Abstract*—The underground spatial-temporal data of an urban is so large that it is impossible to process them by PC. A PC cluster or a supercomputer is more suitable for them. In order to fully make use of multiple computing nodes, the parallel programs are needed. Traditional parallel programs are usually developed in parallel computing mode such as using MPI which is more complex in programming than hadoop a cloud computing mode. Underground spatial-temporal data has four dimensions including temporal dimension and spatial three dimensions. The relationship among different dimensions is little, which is a very key feature suitable for hadoop to process. In a word, vast underground spatial-temporal data is difficult to be deal with serial computing, but is feasible by parallel computing or cloud computing, and underground spatial-temporal data cloud computing is easier in programming and more scalable and more tolerant but as fast as underground spatial-temporal data parallel computing.**

***Index Terms*—spatial-temporal data, cloud computing, underground**

I. Introduction

The underground data can be used to search oil or mineral or pipeline or place for high building or path for subway. Underground data is larger than surface data in a city. Such large data has to be processed by high performance computer [1] or cluster [2].

Parallel computing [3] is designed for high performance computer and cluster. There are many parallel program languages such OPENMPI, OPENMP, HPC. OPENMPI is used by most parallel computing applications. There are three disadvantages of traditional parallel program mode: first, it is difficult to code good parallel programs; second, when the data processed by a parallel program is changed, the program has to be rewrite or revised; third, even one process goes down will cause all parallel processes go down.

Hadoop is known as a program mode for cloud computing that can overcome those disadvantages.

Manuscript received September 25, 2012; revised December 24, 2012.

Cloud computing [4] is a kind of new mode for distributed parallel computing. Cloud computing is easier and more scalable and more tolerant than parallel computing.

II. Underground Spatial-temporal Data Cloud Computing Mode

The underground spatial-temporal data can be divided into different files according to the four dimensions of the data (Fig. 1). A file usually contains the data of different depth in a block at a time, so that the file will not be too small. The data in different time and different blocks can be stored into different files, so that the files can be processed in parallel. If a file is larger than 64M Byte, the file will be cut by hadoop automatically intofiles equal to or less than 64M. Different files are distributed onto different nodes, then Hadoop will distribute processes onto the nodes to deal with files locally.

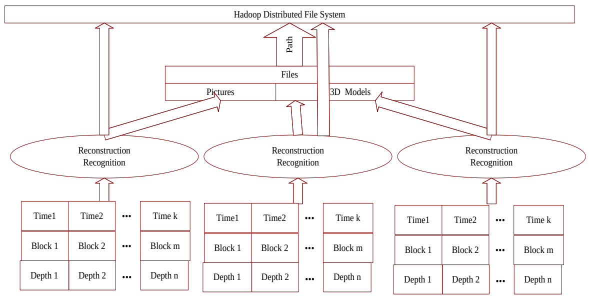

Figure 1. Underground spatial-temporal data file system model

III. Advantage of Underground Spatial-temporal Data Cloud Computing

The underground spatial-temporal data parallel computing mode (Fig. 2) and the underground spatial-temporal data cloud computing (Fig. 3) have different effects when they are used to process underground data.

In the underground spatial-temporal data cloud computing mode, programmer needs not to care about how to distribute data and processes, and the communication among processes are simplified to

doi: 10.12720/joace.1.2.164-165

mapping and reducing. Therefore, underground spatial-temporal data cloud computing system is easier to be programmed than underground spatial-temporal data parallel computing.

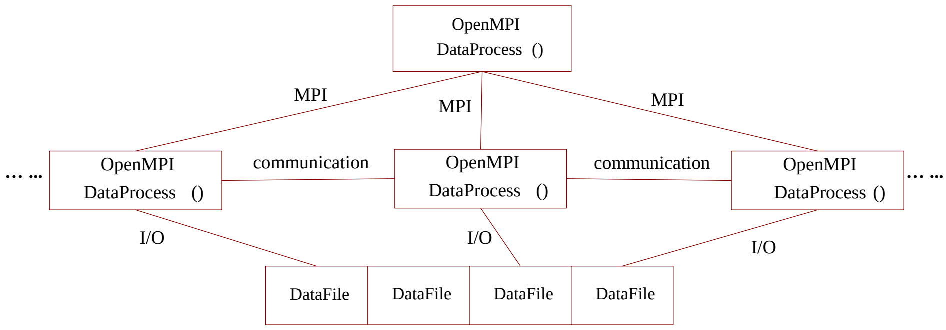

Figure 2. Underground spatial-temporal data parallel computing model

In the underground spatial-temporal data cloud computing mode, when new files are added or old files are deleted or revised, the programs need not to be revised, for hadoop can distribute processes in different way according to the changes of the number and locations of the files. Therefore, underground spatial-temporal data cloud computing system is more scalable than underground spatial-temporal data parallel computing.

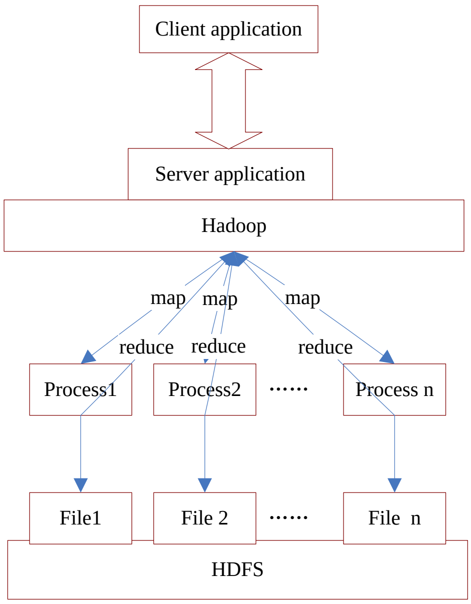

Figure 3. Underground spatial-temporal data cloud computing model

In the underground spatial-temporal data cloud computing mode, when new files are added or old files are deleted or revised, the programs need not to be revised, for hadoop can distribute processes in different way according to the changes of the number and locations of the files.

In the underground spatial-temporal data cloud computing mode, when a node goes down, the data in the node will be resumed in another node from the copy of the data stored in other nodes, and the process in the node will be restarted in the node, where the data has been resumed, to process the data locally.

IV. Conclusion

This paper brings forward an approach to process underground spatial-temporal data by cloud computing. Through discussion, we find that cloud computing is a better approach than other approaches including traditional serial computing and traditional parallel computing for underground spatial-temporal data.

References

[1] A. Chow, D. Hopkins, R. Drost, and R. Ho, "Exploiting capacitance in high-performance computer systems," in *Proc. 4th Annual IEEE International Symposium on VLSI Design, Automation, and Test,* 2008, pp. 55-58.

[2] J. Dean and S. Ghemawat, "MapReduce: Simplified data processing on large clusters," *Communications of the ACM*, vol. 51, no. 1, pp. 10-11, 2008

[3] K. Asanović, "A view of the parallel computing landscape," *Communications of ACM*, vol. 52, no. 10, pp. 56-67, 2009.

[4] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. H. Katz, A. Konwinski, G. Lee, D. A. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, "Above the Clouds: A Berkeley View of Cloud Computing," Tech. Rep. UCB/EECS-2009-28, EECS Department, University of California, Berkeley, 2009.

Renfeng A. Xu was born in May 1961 in Shenyang Liaoning China. She graduated from the Liaoning University in 1983, and gained the bachelor degree with a major in computing. Currently, she is an associate professor in Shenzhen Polytechnic in the field of Database, Data Mining, and Cloud Computing.

She used to work for Liaoning Transportation Research Institute. In 1997, she was employed by the Shenzhen Polytechnic which is one of the best higher occupation institutions in China. Regarding her outstanding working performance, she was given different important positions, such as Vice Director of Computer Application Engineering Department, Director of Software Engineering Department, Vice Dean with school of Electronic and Information Engineering. Now, She is the vice dean with school of Computing Engineering in charge of teaching, adults educating and training. She published 5 textbooks, 6 academic papers.

Ms. Xu is a member of Computer Basic Education Council of Colleges and Universities in China, a member of a council Guangdong Computer Society, Bidding Expert of Shenzhen Government Project Procurement. She has been also honored as Shenzhen Excellent Teacher, Awarded the title of Local-level Leadership Talent of Shenzhen in 2010.