Implementation of Cloud Computing Environment for Processing of Airborne Laser Scanning Data

Jaesun Park[1], Muwook Pyeon[1], Yangdam Eo[1], Hanku Lee[2], and Hyeokju Lee[2]

[1] Department of Advanced Technology Fusion, Konkuk University, Seoul, Korea
[2] Department of Internet & Multimedia Engineering, Konkuk University, Seoul, Korea
{xteen88, neptune, eoandrew, hlee, hjlee09}@konkuk.ac.kr

Abstract. Cloud computing technologies have been interested in the field of IT and studies on distributed storing huge data into lots of different servers and managing them have been largely conducted. In addition, it is essential to process such huge data in the field of space information and related studies on this issue have also been processed. In this study, a creation of DSM is attempted using regular grid generation and mean value interpolation methods, which can be used to process airborne LiDAR data, in a cloud computing environment as a part of such studies. Then, its processing times are measured according to the number of processing nodes. As a result, it is verified that the processing speed is reduced according to increases in the number of nodes.

Keywords: Airborne LiDAR, Big Data, Cloud Computing, DSM

1 Introduction

Recently studies on cloud computing in the field of IT have been actively performed. The cloud computing is a way of computing for providing IT resources as a form of service based on Internet [1]. It has been recognized as an important Internet service platform in global Internet service providers, such as Google, Yahoo, and so on, in which they have developed a large scale cluster based distributed computing platform technology based on low cost commercial nodes and applied it [6]. In addition, the amount of space data in the field of space information has been dramatically increased due to the significant development of GIS(Geographic Information System), LiDAR(Laser imaging Detection and Ranging), and etc. Thus, in this study, a creation of DSM(Digital Surface Model) is attempted using regular grid generation and mean value interpolation methods, which can be used to process airborne LiDAR data, in a cloud computing environment and its applicability is verified through measuring the processing time according to the number of processing nodes.

2 Processing Method of Airborne LiDAR Data

A LiDAR system measures distances using laser from a sensor that can obtain the precise position and attitude data of an aircraft to natural geographical features, such as mountains, valleys, and etc., and artificial geographical features, such as buildings, bridges, an etc., and provides terrain modeling data based on the measured numerical data [2,3]. In this study, a cloud computing processing method was used as a regular grid generation and interpolation processing process for the work after applying the filtering of the previously mentioned primitive data in a data processing flowchart. Because the LiDAR data is obtained with uneven point locations and the data after applying a filtering process represents a more uneven distribution state, it exhibits an uneven data density [5]. Thus, it is necessary to treat the LiDAR data as a form of even grid data according to application [4]. The regular grid generated in this study can be understood as a reference of configuring points, which include 3D coordinates, to 2D cells with a specific interval. The first thing to be considered for generating the regular grid is the size of the grid. The regular grid was generated based on the values of x and y in a 3D coordinate value (x, y, z) belonged in a point. The number of regular grids (G_n) is the multiplication of the difference between the maximum values of x (x_{max}) and y (y_{max}) and the minimum values of x (x_{min}) and y (y_{min}) and the regular grid distances ($x_{Cellsize}, y_{Cellsize}$). The second thing to be considered is the density of points, which are employed in the grid. As the general density of the airborne LiDAR data is 3 points/㎡, the size of the regular grid is to be generated by more than such density. In this study, the average method was used to calculate altitude values.

3 Implementation of a Cloud Computing Environment

3.1 Implementation of Cloud Computing Environment

The cloud computing environment configured in this study was consisted of three sections, such as an LAS data collection section from the airborne LiDAR, an HDFS section that distributedly stores the collected LAS data, and a data processing section that processes the distributedly stored LAS data using the MapReduce. Also, a library of libLAS was used to apply the primitive data obtained in the LiDAR system [7]. The data processing performed in this is as follows.

- The LAS data was stored on HDFS by transforming the data as a 3D coordinate format (x,y,z) using the libLAS in distributed manner
- The data stored in the HDFS was transformed into grids with an interval of 1m through mapping the data using the MapReduce and that was transmitted to a Reduce process by generating a middle value like ({x_0, y_0)}, list{z}).
- The z value of the middle value transmitted from the Map was interpolated and the output was stored as a form of ({x_0, y_0}, z_1) to HDFS for generating DSM.

3.2 Results & Setup of Experimental

The cloud server used in the experiments was a single enterprise scale cluster that consists of 7 computational nodes (slave nodes) and 1 head node (master node). The only way to access the cluster is through the master node. All nodes are running on Linux OS (CentOS 5.5). In addition, other Hadoop environments were determined to

follow the basic configuration. The proposed system was tested with real LiDAR(LAS format) data. The experimental data area was 10.44 square kilometers and 3.5 points were applied to one square meter.
In this experiment, the number of nodes was configured as a variable that was varied by 1, 2, 4, and 8 and the processing times for each number of nodes were measured. The processing time with 8 nodes was reduced about 43% and 10% compared with that of one node and four nodes respectively.

4 Conclusion

In this study a cloud computing environment was introduced to rapidly process a large scale LiDAR data and DAM was generated by using an average interpolation method through generating regular grids. As a result, the processing times were reduced according to increases in the number of distributed processing nodes. Also, it was verified that it can be applied to process a large scale data. In future studies, more various experimental data sets are to be applied and the design and application of a cloud computing environment for various and complex filtering processes will be implemented instead of using an interpolation method in its application.

Acknowledgement. This research was supported by a grant from High-tech Urban Development Program - Development of fundamental technology for next generation hyper-live map Project(11CHUD-C061156-01) funded by Ministry of Land, Transport and Maritime Affairs.

References

1. Rajkumar Buyya, Chee Shin Yeo, Srikumar Venugopal, James Broberg, Ivona Brandic : Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility. Future Generation Computer Systems, vol. 25, pp. 599--616 (2009)
2. Meng, X.; Wang, L., Silván-Cárdenas, J.L., Currit, N.: A multi-directional ground filtering algorithm for airborne LIDAR. ISPRS Journal of Photogrammetry & Remote Sensing, vol. 64, pp.117--124 (2009)
3. Shan, J., Sampath, A.: Urban DEM generation from raw LiDAR data: a labeling algorithm and its performance. ISPRS Journal of Photogrammetry & Remote Sensing, vol. 71, pp.217--226 (2005)
4. Peter Axelsson,: Processing of laser scanner data—algorithms and applications. ISPRS Journal of Photogrammetry & Remote Sensing, vol. 54, pp.138–147 (1999)
5. Keqi Zhang, Jianhua Yan, Shu-Ching Chen: Automatic Construction of Building Footprints From Airborne LIDAR Data. IEEE Transactions on Geoscience and Remote Sensing, vol. 44(9), pp.2523--2533 (2006)
6. Stephen E. Arnold : The Google Legacy, Infonortics (2005)
7. libLAS, http://liblas.org