The Framework of Remote Sensing Image Map Service on Hadoop

Junfeng KANG

Jiangxi University of Science and Technology,
Ganzhou, China, 342100
Email: jfkang@sina.com

Zhenhong DU

Department of Earth Science,
Zhejiang Provincial Laboratory of GIS,
Zhejiang University.Hangzhou, China. 310028

Xiaosheng LIU

Jiangxi University of Science and Technology,
Ganzhou, China, 342100

Abstract—**Through completely discussing the data storage strategy on Hadoop, and analyzing the characteristics of remote sensing image map service , we designed a framework of remote sensing image map service on hadoop.**

Keywords-Hadoop; remote sensing image; map service

I. Introduction

With the development of Earth observation technology, the amount of Remote sensing image data is growing very fast. Traditional storage and management method couldn't adapt to the need, so how provide effective storage and management method and high-performance mapping service for the growing mass remote sensing image data, has become a pressing problem.

Common commercial image map services platform includes Google company's Google Earth[1], Microsoft virtual earth, and National Administration of Surveying, Mapping and Geoinformation of China (NASG)'s TIANDITU[2].

These public remote sensing image platforms mostly focus on providing map service for public users, but taking into account the need for mass data storage, real-time management and security, obviously professional users cannot directly publish their data on these platforms. Based on these commercial remote sensing image platform's advantages, we proposed an enterprise-class remote sensing image map service framework on Hadoop.

II. Remote sensing image data storage strategies on Hadoop

Common remote sensing images services usually include map services, sharing service, and high-performance computing services. For remote sensing images map service, remote sensing images commonly used image pyramid model for storage management, the visit efficiency for HDFS was determined by the splitted image data's size. For remote sensing images sharing service, mostly used for data download service, only need to public the data visit service. For high-performance computing services, the strategy was more complicated, one should split the data into the distributed filesystem to suit for MapReduce[3] program model, because of HDFS's "The price on moving computing is cheaper than moving data".

We concluded the remote sensing image strategies on Hadoop:

(1) HDFS's default storage method:

If file was larger than HDFS's default size, HDFS would split the data into Block. This process can treat as a transparent process, for example, if a file named A store in HDFS, the block name in HDFS should be A000, A001, A002 etc, but user still use name A to get the file, because the namenode of HDFS saves the relation between A and A000.

(2) Remote sensing image map service:

As described above, Our research adopted remote sensing image pyramid model to provide map service, which using Hiblert[4] code to manage the splitted files, then these files would use "HDFS's default storage method" to store in HDFS.

(3) pre-split by default block size:

Mostly remote sensing image's size is larger than HDFS default block's size, so we chose to pre-split the file into default size, and recorded these files' relation, this would reduce the Hadoop servers' burden. This method can improve the MapReduce programs' performance, as showed in Cray's experiment[5], the MapReduce program's performance boost linear close to the block number.

(4) Small file compression method:

Usually the size of remote sensing image file which pre-split into pyramid model was smaller than block size, too many small files would lower the Hadoop's performance. Hadoop provided a method that can use compression file as MapReduce program's input, to improve the program 's performance.

(5) Hadoop's default file manage method:

For download service, remote sensing images only need to directly store in HDFS, but for MapReduce program, the method could not fully make use of Hadoop servers' parallel ability.

(6) other method:

Like "Remote sensing image map service", some MapReduce Program need use professional algorithm to pre-process data for better performance.

III. Remote sensing image map service on Hadoop

Usually remote sensing image map service would allow user viewing the map through browser, figure 1 showed a sketch map, that published to standard map service(such as WFS, WMS, WCS) by map server(such as ArcGIS Server, GeoServer etc)。
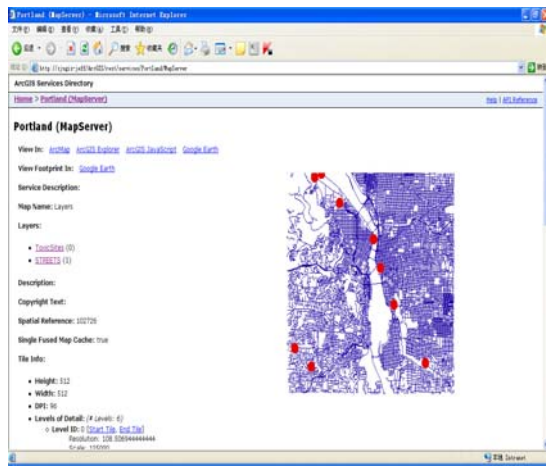
Figure 1. The Interface Example of Web Map Service

A Static map service and dynamic map service

According to data access type, map service included static map service and dynamic map service, static map service referred to previously building the pyramid model through splitting original remote sensing image, but dynamic map service split the remote sensing image when needed. Figure 2 showed the traditional map service's basic structure,

Static map service can provide better experience for users, because of pre-building pyramid model; this could be a huge burden when the amount of remote sensing image grows faster and faster. So Prejudge the possible behavior would be a better solution. For example, for sharing service, remote sensing image should choose "HDFS's default storage method", but the sketch image could create dynamically.
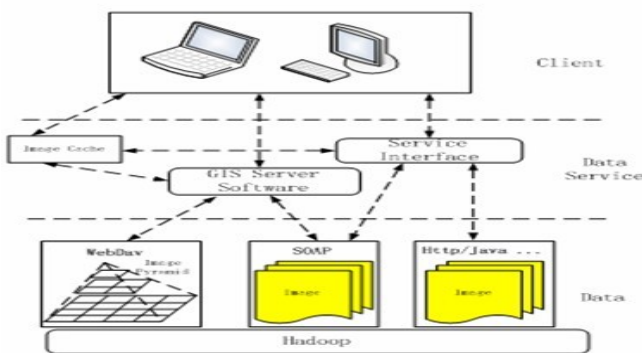

Figure 2. Tranditional Framework of Web Map Service

B Remote sensing image map service on Hadoop

When file stored in HDFS's DataNode as block, the relationship between original file and blocks was also saved in NameNode of HDFS. So visiting file on HDFS need three steps: first getting the relationship on NameNode, then parallel reading the block from each DataNode, finally merging these blocks into original file.

Pre-building pyramid model was be a efficient method for map service, but too many small pre-split image files would lower the performance on HDFS. So we proposed a new solution by merging these small files and saving the index.

As showed in figure 3, our research designed a remote

sensing framework on Hadoop, included three layers, data layer, data service layer, client. data service layer would provide static and dynamic remote sensing image service, data layer utilized Hadoop's interfaces to support different data service. We summarized three methods for the remote sensing image service:
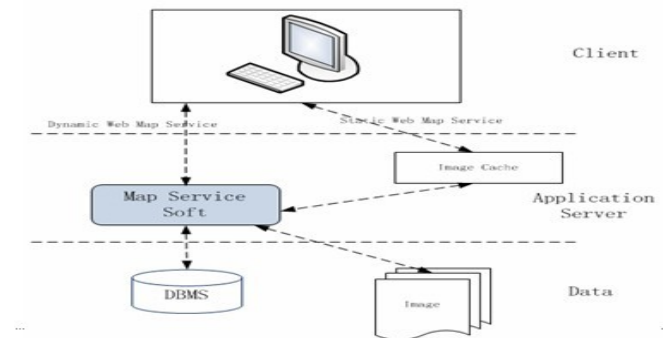

Figure 3. The Framework of Web Map Service on Hadoop

(1)WebDav interface:

WebDav could mount HDFS as normal Linux file system[6], so user could visit the file stored in HDFS like normal file. But too many pre-building pyramid files would be big burden for HDFS.

(2) combination of SOAP protocol and GIS Server:

Hadoop provided SOAP interface, and most GIS Server support SOAP protocol, so combining SOAP protocol and GIS Server would be a better method, but user need to do some work to modify GIS Server.

(3) Http protocol or Java interface:

Though utilizing Hadoop's Http protocol or Java interface, user could visit the file in HDFS, then published as web service.

IV. summarize and future work

Based on discussing how to manage remote sensing image on Hadoop, the paper designed a remote sensing image service framework to providing static and dynamic web map service. In the future, we will focus on how to utilize Hadoop's MapReduce program model to provide high performance computing service, such as designed a dynamic MapReduce splitting image algorithm.

References

[1] Wikipedia. Google earth. http://en.wikipedia.org/wiki/ Google_earth[EB/OL]. 2011-3-11.
[2] Baidu Encyclopedia. TIANDITU. http://baike.baidu.com/view/ 4559007.htm[EB/OL], 2010.
[3] Dean, J. and S. Ghemawat, MapReduce: Simplified data processing on large clusters[J]. USENIX Association Proceedings of the Sixth Symposium on Operating Systems Design and Implementation (OSDE '04), 2004: p. 137-149.
[4] Xueqing DENG. Research of Service Architecture and Algorithms for Grid Spatial Data[D].The PLA Information Engineering University, 2003.
[5] Cary A, Sun Z, Hristidis V et al. Experiences on Processing Spatial Data with MapReduce[M]: Springer Berlin / Heidelberg, 2009: 302-319.
[6] Wbite T.Hadoop: The Definitive Guide, Second Edition [M]:O'Reilly, Yahoo Press,2010.