

Assignment 3 - Problem 3

Summary on “Research on the Construction of Data Management System of Massive Satellite Images” conducted by H. Wang, X. Tang, S. Shi, F. Ye.

Summary:

This paper focuses on the difficulties related to the massive data gathered by the satellites which causes issues to the storage organization. This study introduced and developed a distributed database based on cluster storage system with respect to the chosen storage system type, the design of database models, and the creation of the management system by examining the big-data characteristics of satellite pictures. One of the major challenges in the field of spatial information research is storing and handling large amounts of satellite image data [1]. The researchers concentrated on addressing the major problems related to the storage of huge satellite images and investigated multi-processor parallel archiving, relation queries on JSON, speedy statistics analysis method of grid imagery coverage, and achieved massive satellite image management with daily data increments of terabytes and annual data increments of petabytes of data [1].

The research is segregated in an orderly manner where the data is analyzed first. Here the data is analyzed based on the size of the data with respect to the correlation between them. The data is also analyzed with respect to the coexistence of the structured as well as unstructured data along with the quick update frequency of the data [1]. The researchers also propose to utilize a cluster storage system so that it provides flexibility with respect to storage management and scalability of the massive dataset to be scaled up or out. It also suggests making use of a distributed database model for remote sensing image databases. A management system was suggested that would be the entry point for the users to carry out analyses which recommended multi-process parallel archiving process and solved the issue concerning the relational query method based on JSON along with image coverage using a quick statistical technique based on grids.

To conclude, this research established a multi-layer architecture for managing large amounts of satellite imagery data, involving levels of infrastructure, data, compute, as well as application levels.

Block diagram of Canada's winter storm data management system:

Considering the concept of data management discussed in the research paper mentioned in the assignment description, I analyzed and created a block diagram architecture of Canada's winter storm data management system. The figure is attached below as **Figure 3.3.1**. The data of the winter storm is segregated into raw image data, basic product, senior product, and thematic product. This was segregated based on the database analyses by [1]. This data is bifurcated into database storage and file storage data. The data is segregated based on the type of data - if the data is huge containing a big size of individual image data then these images are not suitable for database storage and so these images are grouped into file storage [1]. For database storage, the system uses a distributed NoSQL database which has multiple data nodes for storage which in turn communicates with a meta server [2]. This server interacts with the client requests from the client pool. Similar to the database storage, for file storage, the system uses a distributed file system which consists of storage devices to store file data which further communicates with the NFS server [3]. This server also handles the client requests from the client pool.

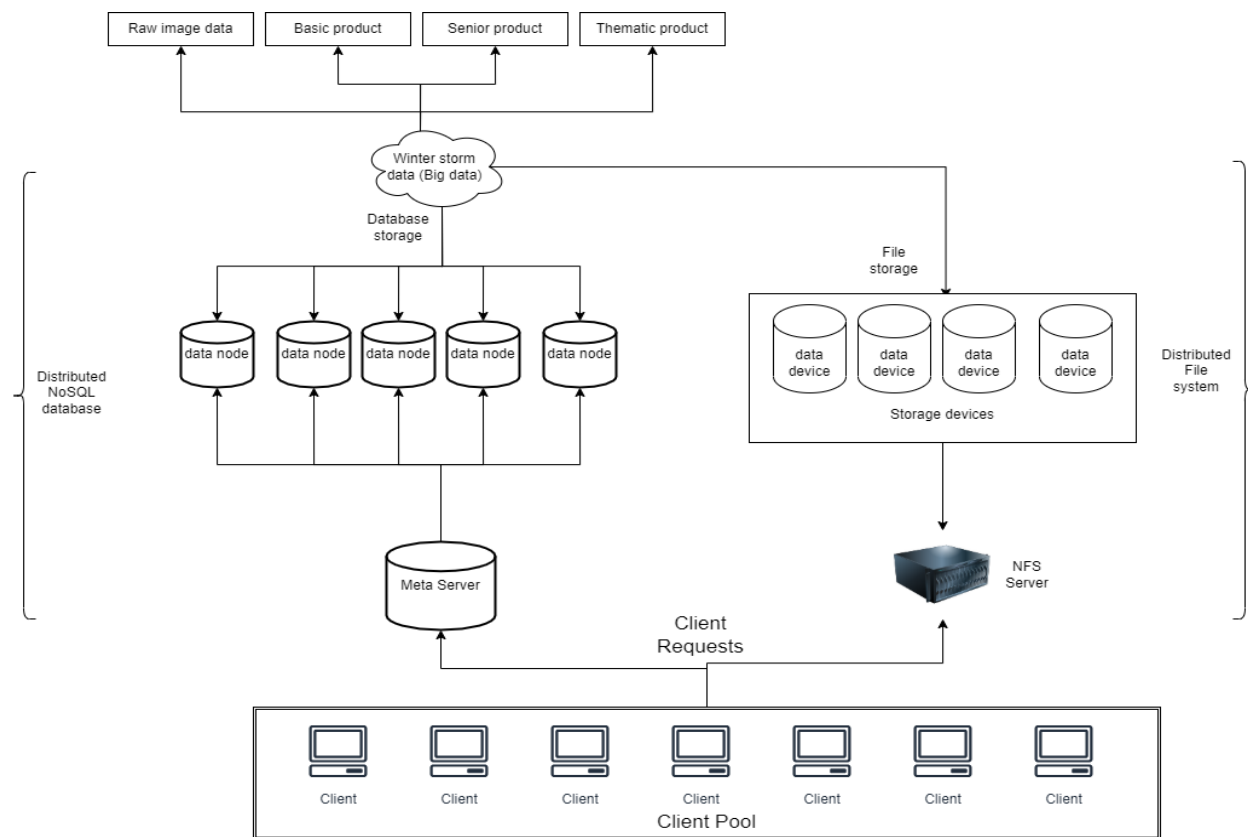


Figure 3.3.1 Canada's winter storm data management system

References:

- [1] H. Wang, X. Tang, S. Shi and F. Ye, "Research on the Construction of Data Management System of Massive Satellite Images," *2018 International Workshop on Big Geospatial Data and Data Science (BGDDS)*, 2018, pp. 1-4, doi: 10.1109/BGDDS.2018.8626845.
- [2] Vatika Sharma, Meenu Dave; "NoSQL and Hadoop Technologies On Oracle Cloud"; *International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)*; Volume 2, Issue 2, March – April 2013.
- [3] "What is DFS (Distributed File System)? - GeeksforGeeks", *GeeksforGeeks*, 2022. [Online]. Available: <https://www.geeksforgeeks.org/what-is-dfsdistributed-file-system/>. [Accessed: 24- Feb- 2022].