Future Trends and Technologies: Big Data Analytics pro Web Processing Service

Deepti Joshi, Harish Karnatak, Arijit Roy

Indian Institute of Remote Sensing, Indian Space Research Organization, Department of Space, Govt. of India, Dehradun, deepti@iirs.gov.in

Abstract

Recent evolution is focused on interaction with large amount of datasets and challenges to design analysis technique. To meet these challenges, Big Data Analytics concept emerged which require advance technologies to process large volume of data efficiently by using distributed and parallel file systems and in terms of GIS it is an incipient technology which allows analysis of unstructured large volume datasets to solve complex spatial decision problems. Big Data Analytics enunciated through powerful and distributed framework offered by Apache Hadoop MapReduce. Hadoop and its ecosystem of Big Data technologies offer distributed storage, large scalable data processing platform which facilitate enterprises to build emerging types of data driven applications that are at the heart of their digital strategy.

In former study, Indian Biodiversity data repository generates in Geo-RDBMS environment using PostgreSQL and POSTGIS during Biodiversity characterization study. The raster data is published as WMS and WFS standard which allows online query and visualization. The developed system is taking input from users as Shape file upload, WKT as AOI to perform raster based operation using python, GDAL/OGR, JavaScript and execute raster data analysis over the web using PHP and PostGIS.

Indian Biodiversity contains large volume of geo-spatial data which can be processed through Big Data Analytics very efficiently. By using spatial data (satellite based primary information, secondary geospatially derived or modelled information and geospatially referenced field sample plots) from Biodiversity Information System (http://bis.iirs.gov.in), Climatic data, Plant species dataset, Biogeography and other datasets will also give valuable input towards climate change study. This study will characterize how to process enormous amount of data into web environment which solicit WPS (Web processing service). WPS specification endues interface rules that standardize inputs and outputs for geospatial processing services on web. Geospatial Data Processing can take advantage of the OGC WPS as web interface to allow for the dynamic deployment of user processes. However, intensive computation is required for processing of large amount of geospatial problem and it can be accomplished through Big Data analytics. High performance computing cluster will be used to configure Apache Hadoop framework for large scale processing to perform intensive computation for spatial data modeling and data required by WPS will available at the server. Considering effectively, the climate change in Indian Biodiversity requires further research to overcome this dilapidation.

Keyword: Big Data, Biodiversity, Services Oriented Architecture, OGC web services, GDAL/OGR, and Geoprocessing, Hadoop

1. Introduction

The term Big Data which is used to define the utility of data, organized and amorphous in anenormous amount and it is massive in nature. Big Data concept has come as internet usage is growing. Variation, capacity, pace, mutability and involution, these are the factor which describes Big Data very preciously. In order to avail deep sightedness of big data, describes related technologies and that provides web environment. The relationship between the factors to establish web based method for querying with Big Data is studied in detail.

2. Related Technologies

2.1 *WPS*

Earlier, we are simply concentrating on non-spatial data sets to run in web environment but by the time being vast revolution has come into internet domain and individuals are focusing on spatial datasets also. Web Processing Service (WPS) term emerge because of spatial data which provides interface for geo-processing in web domain given by Open Geospatial Consortium (OGC). The WPS interface where client connected to the server to display, narration and executing predestined geographic method using HTTP (Hypertext transfer Protocol) and XML (Extensible Markup Language), to interchange the data sets.

WPS provides flexibility, distributed computing, reusability and scalability when data is interchanging. It defines complex data, Textual Data viz. numerical values, Bounding Box Data viz. geographic coordinate for rectangular area.
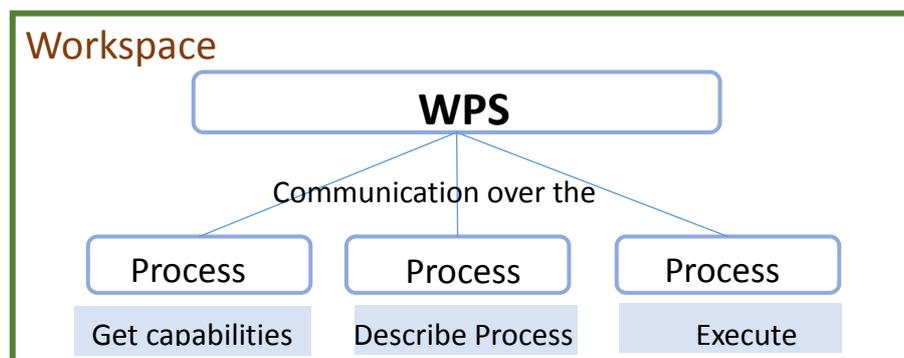

2.1 Connection between WPS and Big Data

Latterly few years huge amount of data exchange over the web. In Web environment, such a large amount of data to process, has become a challenge. Big data is primarily generated from social networking services, real-time database and it provides organized and semi-organized datasets. Big Data concept has emerged because of more internet usage.

The OGC Web Processing Service (WPS) Interface Standard permit the dynamic distribution of processes. To process such large amount of data in the internet domain WPS involve and access to processing on web has been implemented. The Group on Earth Observations (GEO) used WPS to develop the Global Earth Observing system of System (GEOSS).

2.3   GIS

Geographic Information System (GIS) designed to evaluate, seize collection, control, and accomplishall spatial dataassociated with locations on Earth's surface. It is a system which integrates information technology and geography to solve real time spatial problems.Integration of GIS with Multicriteria Decision analysis techniques allows solving semi structure problems related to natural resources.

Traditionally, GIS has retrieved geospatial content stored in a relational database management system optimized typically for datasets of less than a gigabyte.GIS analyst has the capability to use or misuse the control that GIS brings to the world's problems by functioning on Mapping & Visualization, Spatial Database and Spatial analysires.
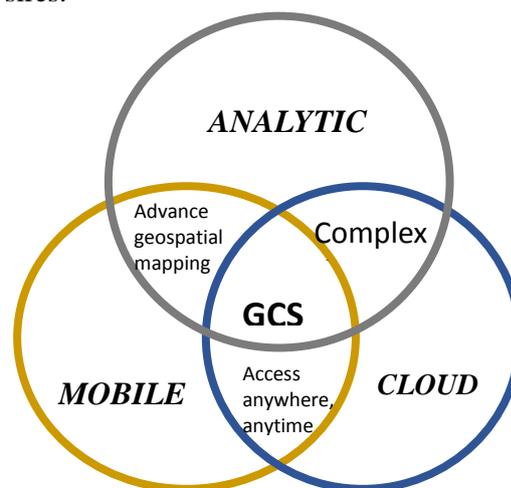
2.4   GIS with Big Data

The concept of Big Data for GIS is an emerging technology which allows analysis of unstructured large volume datasets to solve complex spatial decision problems. Spatio-temporal sensor, image, simulation, statistics data are comes under Big Data which is also called geo-spatial data. Big data analytics require advance technologies to process large volume of datathat would take too much time and cost to load into a relational database for analysis efficiently by using distributed and parallel file systems.As Big Data technology is able to process enormous amounts of data in the shortest time, it has comes as a new frontline for geospatial analysis.

2.5   Hadoop with Big Data

GIS Implements for Hadoop is an open source toolkit proposed for Big Spatial Data Analytics.Hadoop provides the framework forBig Data analytics concept to work non-spatial and spatial data.Two successful projects namely Apache Hadoop (2001), Google MapReduce (2004, 2007) developed for cloud computing including Google File System (GFS) are typical example of Big Data analytics where complex computations are performed using distributed file system.Hadoop is composed of Hadoop Distributed File System (HDFS) and Map Reduce runtime system. In technical perspective, we can establish Hadoop environment and check the applicability of Hadoop in geospatial processing.

Hadoop is a principal part of operations of computers for many organizations, just as Google, Yahoo, Facebook, Twitter and LinkedIn.OGC access services: Web Map Service (WMS),

Web Coverage Service (WCS), Web Feature Service (WFS) and Sensor Observation Service (SOS) Interface Standards are well suitable for service interfaces in a geospatial data processing.

3. Related Work

The OGC Web Processing Service (WPS) Interface Standard was applied by Terradue to develop a Cloud Service that allows development on Hadoop framework. Terradue chosen the 52° North WPS open source component as the interface to a Cloud service, which can called as WPS-Hadoop. The University of Pretoria has implemented "Processing as a Service" in the cloud with WPS where PAAS combines with WPS. Feng Chia University has provided WPS access to service oriented architecture (SOA) based debris flow monitoring system.

George Percivall et al. discussed application of WPS in GEOSS. Climate changeanalysis on Biodiversity data can be done by using WPS. The GEOSS AIP "eHabitat" generate the same scenario. It uses web service which allows end-users to evaluate the possibility to looking-for continually habitats.To access and process the spatial data, OGC WMS and WPS were used as open standard. Feng Chia University which is GIS center,developed a WPS for dispatch of emergency vehicles.

A characteristic geoprocessing process takes an input dataset, executes a process on that dataset and returns the outcome of the operation as an output. For online geospatial data processing, Web Processing Service specification was released in 2005 by Open Geospatial Consortium, that incorporate complex spatial process through a standardized service interface based on the Hypertext Transfer Protocol (HTTP) (Foerster et al., 2006).

Claudia Vitolo et al. (2014) elaborated presently available implementations related to web-based technologies for processing large and heterogeneous datasets and discuss their relevance within the context of environmental data processing, simulation and prediction. It follows schematic structure and describes typical web service architectures and complexity of communication between client and server. They discussed the range of available tools and technologies for web based environment modelling dealing with Big Environmental Data.

4. Result and Discussion

In this paper, we analyse the work done on Big Data and offers web based processing based on Hadoop open source. It also provides overall concept on spatial data by describing GIS with enormous amount of data. It emphasizing on future research work related to Geospatial data strategy, Remote Sensing, Disaster Management, Agriculture, Environment studies (Climate change study, Biodiversity data study) and Earth observations.

The OGC WPS is the next foremost part to meeting the appropriator needs. However, my paper is showing some preliminary research with the Big data problem including we based

processing. The reach the aim for geospatial web based processing, a lot of work need to be done. This is our upcoming interest work.

5. References

1. Claudia Vitolo, Yehia Elkhatib, Dominik Reusser, Christopher J.A. Macleod, Wouter Buytaert. (2014), Web technologies for environmental Big Data, Environmental Modelling & Software 63: 185-198
2. Definitions of Big Data, Opentracker Analytics (2015), http://www.opentracker.net/article/definitions-big-data
3. Empowering GIS with Big Data, http://www.gislounge.com/
4. George Percivall (2013), Big Processing of Geospatial Data,http://www.opengeospatial.org/blog/1866
5. George Percivalla, Lionel Menardb, Lan-Kun Chungc, Stefano Nativi d and Jay Pearlman, GEO-PROCESSING IN CYBERINFRASTRUCTURE:MAKING THE WEB AN EASY TO USE GEOSPATIAL COMPUTATIONAL PLATFORM
6. IEEE Transactions on Big Data(2015) , http://www.computer.org/web/tbd
7. Kun Zheng and Yanli Fu (2013), Research on Vector Spatial Data Storage Schema Based on HadoopPlatform, International Journal of Database Theory and Application Vol.6, pp.85-94
8. M. Chen et al. (2014), Big Data: Related Technologies, Challenges and Future Prospects,SpringerBriefs in Computer Science, DOI 10.1007/978-3-319-06245-7__2
9. Theodor Foerster, Andre Brühl, Bastian Schäffer (2011), RESTful Web Processing Service
10. WPS Concepts, Geographic Information Processingfor the Web,http://geoprocessing.info/wpsdoc/Concepts
11. S. Akhter a,b, K. Aida a,b and Y. Chemin c (2010), GRASS GIS ON HIGH PERFORMANCE COMPUTING WITH MPI, OPENMP ANDNINF-G PROGRAMMING FRAMEWORK, International Archives of the Photogrammetry, Remote Sensing and Spatial Information Science, Volume XXXVIII, Part 8