Analysis of USGS Earthquake Data

NANDITA SATHE[1,*]

[1] School of Informatics and Computing, Bloomington, IN 47408, U.S.A.
[*] Corresponding authors: nsathe@iu.edu

project-001, March 25, 2017

Geo-spatial data fits into definition of Big Data as it has all three 'V's viz. high-velocity, high-volume and high-variety. Big Data Analytics tools now allow us to analyze the huge volumes of geo-spatial data. Data of earthquakes that take place globally is a major part of crucial geo-spatial data. This application analyzes data related to earthquake which can be utilised in further research. US Geological Survey's (USGS) Earthquake Hazards Program monitor and report earthquakes, assess earthquake impacts and hazards, and research the causes and effects of earthquakes [1].

Keywords: I524, geospatial, MongoDB, D3.js, Apache Spark, Python, USGS, Ansible

https://github.com/nsathe/sp17-i524/blob/master/project/S17-IO-3017/report/report.pdf

1. INTRODUCTION

The USGS estimates that several million earthquakes occur in the world each year, although many go undetected because they occur in remote areas or have very small magnitudes [2]. Thus earthquakes pose significant risk globally to the mankind. USGS collects volumes of geospatial data pertaining to earthquakes and makes it available for analysis. This project intends to analyze this data. The application will be deployed on cloud. Deployment will be automated using Ansible.

2. TECHNOLOGIES USED

Technologies used for development and deployment of this project are listed below.

1. Cloudmesh - For connecting to different cloud environments.
2. Ansible -For deploying software and associated packages.
3. Python - Writing script for data analysis and data processing
4. Apache Spark - For data processing
5. Mongo-DB - For storing Geo-spatial data
6. D3.js - As a visualization tool

3. EXECUTION PLAN

This is how I intend to execute the project on week-by-week basis. Although my intention is to follow the plan deligently, it is possible that because of technical and other un-foreseen challenges deadlines may be pushed ahead.

1. **6 Mar 2017 - 12 Mar 2017** Create virtual machines on Chameleon cloud using Cloudmesh and submit the project proposal.
2. **13 Mar 2017 - 19 Mar 2017** Deploy Mongo DB to Chameleon cloud using Cloudmesh and develop initial Ansible playbook to install the required software packages.
3. **20 Mar 2017 - 26 Mar 2017** Write script in Python for downloading USGS data at run-time. Write Python and Spark scripts for data analysis and processing.
4. **27 Mar 2017 - 02 Apr 2017** Implement visualization using D3.js. Update Ansible playbook to install D3js package.
5. **03 Apr 2017 - 09 Apr 2017** Test on different cloud systems. Define quantitative benchmarks. Tentatively benchmarks will be for data insertion time and data processing time.
6. **10 Apr 2017 - 16 Apr 2017** Create deployable software package in Python.
7. **17 Apr 2017 - 23 Apr 2017** Update and finalize the Project Report

4. BENCHMARK

As mentioned in Section 3, benchmarks are tentatively for data insertion time and data processing time on different clouds.

5. ACKNOWLEDGEMENTS

This project is undertaken as part of the I524: Big Data and Open Source Software Projects course at Indiana University. The author would like to thank Prof. Gregor von Laszewski and his associates from the School of Informatics and Computing for providing all the technical support and assistance.

6. LICENSING

TBD

7. CONCLUSION

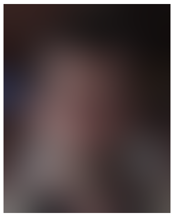
TBD

REFERENCES

[1] USGS, "Earthquake hazards program," Web Page. [Online]. Available: https://earthquake.usgs.gov/

[2] ——, "About us - program overview," Web Page. [Online]. Available: https://earthquake.usgs.gov/aboutus/

AUTHOR BIOGRAPHIES



Nandita Sathe is PMP certified project manager by profession. She will obtain MS in Data Sciences from Indiana University in May 2018. Her interests are in data analytics and machine learning.

8. WORK BREAKDOWN

The work on this project was distributed as follows between the authors:

Nandita Sathe. She completed all the work related to development of this application including research, testing and writing the project report.