



Name: Jhummarwala Abdul Taiyab Abuzar  
Enrollment No: 129990907004  
Branch: Computer Engineering, GTU  
Thesis Title: Open Cloud based distributed Geo-ICT services

## **ABSTRACT**

Geographic Information System (GIS) consists of applications which gather geospatial data from a variety of sources, operate upon them and provide results or maps which are utilized for planning and development purposes across the domains of forestry, rural and urban planning, disaster management, etc. Due to widespread adaption of GPS, terabyte sized geodatasets having billions of geographic features are available. Analysis of such huge datasets which requires large amounts of storage, compute and memory infrastructure is not possible with a single computer using traditional GIS. These are readily available in Cloud Computing and compels its application to GIS. The Geospatial data is collected in raster and vector formats. Shapefiles (.shp) format is the most widely used and prominent vector data format. The aim of this thesis is development of a distributed processing system capable of processing and visualizing large amount of geospatial vector data available in form of Shapefiles. The available shapefile dataset consists of more than 330,000 shapefiles which was collected over a span of a decade, requires 750 GB of storage and has hundreds of billions of geographic features. The distributed geo-processing software framework, named "GeoDigViz", has been developed. It is based on Hadoop for distributed processing and uses HDFS to store terabytes of geospatial data. The model framework employs multi-level indexing and has enabled processing and utilization of hundred of thousand of Shapefiles collectively without specialized consideration of their heterogeneous structure. It is scalable to support terabytes of such data and in addition, is based on free and open source geo-processing software for processing and visualization. The model framework has enabled processing and utilization of hundred of thousand of Shapefiles collectively without specialized consideration of their heterogeneous structure. It provides visual access, filtering and extraction of subsets of data from the collection of shapefiles for performing complex spatio-temporal analysis later using specialized GIS software. The model relieves the geo-scientists from the complexities of working with big-geospatial data and distributed systems and they can focus on derivation of current insights from the underlying geospatial phenomenon.

### **List of Publications:**

1. Abdul, J., Sumit, P., Potdar, M.B., 2017. GeoDigViz: A spatio-temporal model for massive analysis of Shapefiles. IEEE International Conference on Computing, Communication and Automation (ICCCA2017).
2. Abdul, J., Mazin, A., Potdar, M.B., 2016. Geospatial Hadoop (GS-Hadoop): An efficient MapReduce based engine for distributed processing of Shapefiles. 2nd IEEE International Conference on Advances in Computing, Communication, & Automation (ICACCA).
3. Abdul, J., Potdar, M.B. and Chauhan, P., 2014. Parallel and Distributed GIS for Processing Geo-data: An Overview. International Journal of Computer Applications (IJCA). 106(16).