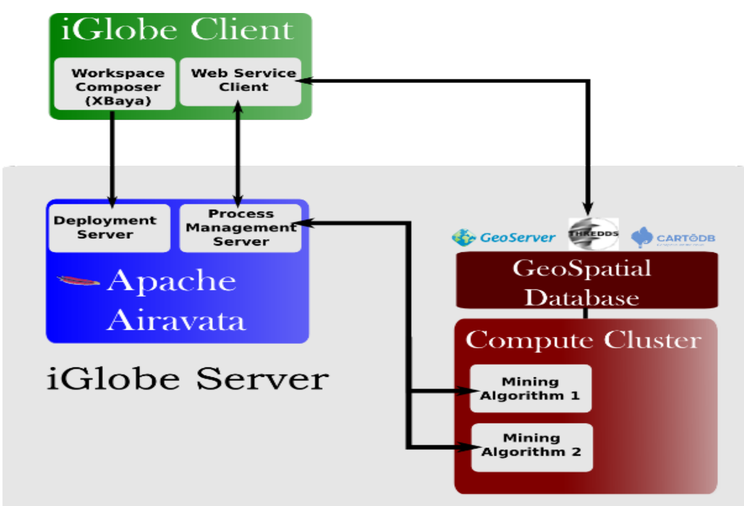
ARISTOTLE CLOUD FEDERATION

Use Case 1: A Cloud-Based Framework for Visualization and Analysis of Big Geospatial Data

V. Chandola (University at Buffalo) with R. Vatsavai (North Carolina State University), P. Hogan (NASA Ames Research Center), B. Bhaduri (Oak Ridge National Laboratory), M. Gee (University of Chicago)

Motivation: Geo-data has historically been one of the most valuable formats of data across a broad spectrum of application domains. Typically, such data comes with spatial and temporal modalities. Examples include Earth Observation (EO) data collected by air and space borne remote sensing instruments and Climate Simulations, produced by climate models. With advances in sensor and computing technologies, the temporal and spatial resolution of such data as well as the diversity in the type of observations has increased manifold over the last decade, resulting in massive troves of geospatial data scattered around the world within the data archives of researchers, government, and the private sector. Integrative visualization and analysis of such data is vital for many scientific and research efforts. However, geospatial data presents unique challenges besides the typical big data challenges. These include the heterogeneity in terms of access protocols, file formats, geographic projections, spatial and temporal resolutions, etc. For example, most remote sensing products are typically served as images in data archives whereas scientific simulation data is available in formats such as the Hierarchical Data Format (HDF). Researchers typically pull the desired data into a local computer, and visualization and analysis of such data is typically done *ad hoc* using dedicated software for each task, often limited by the local resource constraints (computing, networking and storage).

Activity: To streamline the integrated visualization and analysis of geo-data, we will develop a cloud-based solution that will allow researchers to seamlessly integrate heterogeneous geo-data from a variety of sources into a cloud-based analysis engine. Since the analytics are delegated to the cloud, it puts minimal burden on the local computing, storage, and networking resources. The proposed framework will deliver visual output to the local computer for display, comprising an interactive system allowing users to locate, visualize, and subset desired data and then create workflows to execute sophisticated analysis in the cloud. The solution will be developed around iGlobe [1], an interactive analysis and visualization tool for geo-data. The proposed iGlobe client-server product addresses the heterogeneity across disparate data formats and access protocols in addition to providing a unique 3D visualization environment.



First, we will develop the cloud-based server component of iGlobe as a *thin layer* in the cloud between the iGlobe client and the analytic tools, to allow scalable, demand-driven execution of analytics on big geospatial data, as shown in the figure.

Proposed architecture of iGlobe-based solution

#1541215

Second, we will develop several core machine learning and statistical analysis-based tools that are vital for the research community. Our overarching focus is on developing scalable tools that can leverage the unique computing capabilities of the underlying cloud architecture (distributed and parallel). Specific instances of tools include: correlations across space and time, multivariate change analysis, and statistical downscaling.

References

[1] Chandola, V., Vatsavai, R. & Budhendra, B. (2011). iGlobe: an interactive visualization and analysis framework for geospatial data. *COM. Geo.*, 21. Retrieved from: http://dl.acm.org/citation.cfm?id=1999341.

#1541215