Kirthi Banothu
Junior, Computer Engineering

Mentored by the
XSEDE Group
Project Manager: Martin Biernat
Wayne Louis Hoyenga

Project Objective

Implement open source virtual machine version of Hadoop running on Illinois Campus Cluster.

Illinois Campus Cluster: Campus wide resource that meets the needs for faculty's research computing cycles.

Apache Hadoop: Open source software that enables distributed processing of large data sets across clusters of commodity servers.

Approach

Our goal is to deploy Hadoop onto the campus cluster such that the users can leverage big-data analytics.
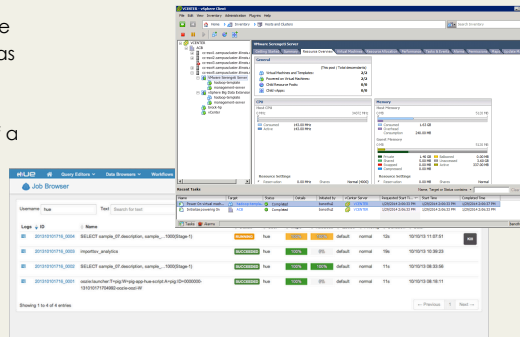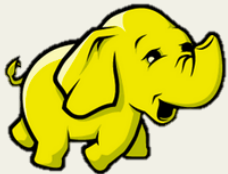
Our procedure is to deploy the infrastructure on top of Serengeti. Serengeti will allow for us to help automate deployment and the management of Hadoop clusters for the users. In order to virtualize the environment, we will be using vSphere.

Status

With assistance from Wayne and Martin, the vSphere foundation, as well as Serengeti has been deployed onto a virtual machine for utilization.
Following this will be the implementation of a GUI for process management.

VMware Serengeti Server running on virtual machine.

Questions

• What are some potential applications for which a user could leverage the capabilities of big-data analytics?

• What features and applications of Hadoop are of most important to you?

• How could an interface be created to help provide a user with the best experience?