Bigfoot – 1

- Big data in the cloud
- Automated and self-tuned deployments of data analytics on virtualized systems
- MapReduce, HDFS, Spark and OpenStack
- Main contributions:
  - OpenStack: analytics as a service
  - Hadoop: scheduling and work sharing
  - DinoDB: MPP framework for RAW data exploration using a SQL interface
Bigfoot – 2

• OpenStack: analytics as a service
  – Launch a job with one click
  – **Automatic configuration** of HDFS, Spark and Hadoop
  – No manual provisioning of virtual machines
  – Fully integrated, available in **OpenStack Juno**

• Hadoop: scheduling and large scale data summarization
  – HFSP scheduler optimizes performance of concurrent jobs
  – Optimized **OLAP** operators for **PIG**, included in mainstream PIG releases

• DinoDB: Explore RAW data in HDFS with SQL queries