

HOW ONE COMPANY LEVERAGED UNIVA GRID ENGINE TO ENABLE NEW SCIENCE

CUSTOMER'S GOAL

The customer developed an improved method to replace an existing ETL process that was used to manage unstructured data for the research group in a drug discovery organization. The application's design goal was to enable "new science" as the amount of data to be computed was at a scale not practically supported without Hadoop. At the point of deployment the team had to devise a model that would support the CIO's mandate of a single system. No stand-alone application silos were permitted.

THE SOLUTION

The new application, developed using Cloudera's Distribution including Hadoop, was integrated into the existing cluster running Univa® Grid Engine™ enabling the team to meet the CIO's single system mandate.

IMPLEMENTATION HIGHLIGHTS

The integration allows Hadoop jobs to run in the cluster that is managed by Univa Grid Engine non-visible to the scientists and researchers. To ensure persistence of HDFS additional drives were added to nodes where the MapReduce jobs will run.

BENEFITS

The integration of the Hadoop environment into the existing cluster significantly reduced the capital expenses (capex) with the elimination of an expensive stand-alone Hadoop cluster. The incremental costs were "just a bunch of disks" (JBOD) for the nodes supporting the Hadoop jobs. By leveraging existing cluster expertise and a single pool of computational resources to run mixed workloads, operational expenses were reduced.

WHAT THE CUSTOMER GAINED

- **Ability to conduct new science at a scale previously not feasible**
- **Significant cost avoidance**
- **Increased utilization of existing cluster**

Univa Corporation provides the evolution of Grid Engine, the most widely deployed and distributed resource management software platform used by enterprises and research organizations across the globe.

