Market Requirements Document

Feature Name: Objectivity/DB 3.0 Knime Integration

Version: 0.1 Date Submitted: 6/8/2015 Completed By: Brian Clark

Description of the Problem

Project 'Purple' is bringing Objectivity products into the "Big Data" and "Fast Data" space leveraging a number of existing open source technologies. KNIME is the leading open platform for data-driven innovation helping organizations to stay ahead of change.

This MRD will set out the requirements for the integration 'purple' with Knime for workflow creation.

Background

"KNIME, pronounced [naim], is a modern data analytics platform that allows you to perform sophisticated statistics and data mining on your data to analyze trends and predict potential results. Its visual workbench combines data access, data transformation, initial investigation, powerful predictive analytics and visualization. KNIME also provides the ability to develop reports based on your information or automate the application of new insight back into production systems. KNIME Analytics Platform is open source and available under GPL License. It can be extended with KNIME Commercial Software to include professional support, productivity and collaboration functionality, providing the best of both worlds".

Description of the Requested Feature

There are 2 requirements:

1. Knime provides built-in 'Nodes' to connect to some well known databases. We will need to add our own Nodes to connect to Objectivity/DB and extract data.
2. Knime will need to integrate with Apache Spark (through the Intel ATK). It is expected Intel/Knime will perform this integration.

Part of an existing feature or does it require another feature, if so, which one?

- New work.

How is this problem being solved now, and why isn't that acceptable?

- New.

What languages must support this capability?

- Initially Java.

Which platforms must be supported?

- Integration with Knime should run initially on Linux.

Do any competitors already have this feature?

- .

Customers who require this feature

- Some existing customers e.g. CGG, new customers in the ‘big data’ space.

Revenue at risk, or which could be won

- Could lead to more early adopters.

When is this required?

- Increments through ‘Purple’ MVP October.

Additional Notes

1. **Implementation notes:**
   a. CGG demonstration
   b. Need to prioritize what features are needed for each increment.

2. **Related Material**
   We will also need:
   Field Training.
   Quality Assurance.

3. **Software requirements**
   a. Objectivity/HDFS integration [http://objyshare.obj.com:8080/display/PM/MRD_Objectivity_HDFS_Integration_V0.1]
   b. Objectivity/RDD [http://objyshare.obj.com:8080/display/PM/Spark+Connector+MRD]
   c. Objectivity/REST Server [http://objyshare.obj.com:8080/display/PM/MRD_Objectivity_REST_Server_V0.5]
   d. Objectivity/Administration Console [http://objyshare.obj.com:8080/display/PM/MRD_Objectivity_Administration_Console_V0.1]
   e. Apache Spark [https://spark.apache.org]
   f. Apache Oozie [http://oozie.apache.org]
   g. YARN [http://hadoop.apache.org/docs/stable/hadoop-yarn/hadoop-yarn-site/YARN.html]
   h. Cloudera Quickstart [http://www.cloudera.com/content/cloudera/en/downloads/quickstart_vms.html]
   i. Hortonworks Sandbox [http://hortonworks.com/products/hortonworks-sandbox/] [http://docs.hortonworks.com/HDPDocuments/HDP2/HDP-2.1.5/bk_dataintegration/content/ch_using-oozie.html]
   j. MapR Sandbox [http://doc.mapr.com/display/MapR3/MapR+Sandbox+for+Hadoop#MapRSandboxforHadoop-InstallingonVMwarePlayer/VMwareFusion]

4. **Hardware Requirements**
   a. Hadoop cluster