Market Requirements Document

Feature Name: Objectivity 3.0 Demonstration

Version: 0.1 Date Submitted: 5/28/2015 Version: 0.2 Date Submitted: 7/8/2015 Update requirements Completed By: Brian Clark 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. We need a way to demonstrate our capabilities and differentiators.

This MRD will set out the requirements for a demonstration application.

Background

Traditionally we have built a variety of demonstrations and samples to show off our capabilities. These are usually driven by work done in the field based on customers evaluations and proof of concepts. We have an opportunity here to choose the right application and data sources to show off our capabilities and differentiators from the competition to promote lead generation and sales & marketing activities.

Description of the Requested Feature

There are some general requirements:

- 1. Must run on local VM (Cloudera Quickstart VM, Hortonworks Sandbox, MapR Sandbox);
- 2. Must run in the cloud (Amazon, Databricks);
- 3. Must run on Hadoop clusters;
- 4. Must run on the Intel ATK framework (Cloudera QuickStart VM).

Requirements for input:

- 1. Multiple different data sources;
- 2. Data sources must be representative of real world use cases, including time-series and geo-spatial data, and support real world queries;
- 3. Must show ingest in parallel using Spark Streaming on YARN;
- 4. Must show ingest in parallel from source data on HDFS;
- 5. Must show combing of fast data with big data;

- 6. Must scale up to billions of objects and relationships;
- 7. Must include Spark processing steps to generate meta data, via RDD transformations;
- 8. Should show use of Spark MLlib and GraphX;
- 9. Should show results using REST Server;
- 10. Should monitor progress with our dashboard, (and integration with 3rd party dashboard?)

Requirements for query:

- 1. Our core strength is navigation query, so that should be primary focus. How are two objects of any types connected through any number of connections of any connection types is the killer;
- 2. Ease of specifying complex navigation queries through path pattern matching should be shown;
- 3. Use of DO (Declarative Objectivity) via REST Server;
- 4. Fast lookup and using custom indexing should be shown.

Requirements for output:

- 1. Must be visually appealing to the demonstration target consumer;
- 2. Should show off our dashboard, (and integration with 3rd party dashboard?)

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

• New demonstration dependent on components of 'purple'.

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

• New demonstration application for new product purple being addressed and resulting solution.

What languages must support this capability?

• Initially Java.

Which platforms must be supported?

• Demonstration should run initially on Linux Windows and Mac OS X to be considered later.

Do any competitors already have this feature?

• Major implementations of Hadoop (MapR, Cloudera, and Hortonworks) support running in this type of environment.

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 end of June.

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.objy.com:8080/display/ PM/MRD_Objectivity_HDFS_Integration_V0.1]
- b. Objectivity/RDD [http://objyshare.objy.com:8080/display/PM/ Spark+Connector+MRD]
- c. Objectivity/REST Server [http://objyshare.objy.com:8080/display/PM/ MRD_Objectivity_REST_Server_V0.5]
- d. Objectivity/Administration Console [http://objyshare.objy.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-yarnsite/YARN.html]
- h. Cloudera Quickstart [http://www.cloudera.com/content/cloudera/en/ downloads/quickstart_vms.html]
- i. Hortonworks Sandbox [http://hortonworks.com/products/hortonworkssandbox/] [http://docs.hortonworks.com/HDPDocuments/HDP2/ HDP-2.1.5/bk_dataintegration/content/ch_using-oozie.html]
- j. MapR Sandbox [<u>http://doc.mapr.com/display/MapR3/</u> <u>MapR+Sandbox+for+Hadoop#MapRSandboxforHadoop-</u> <u>InstallingonVMwarePlayer/VMwareFusion</u>]

4. Hardware Requirements

a. Hadoop cluster