# **Market Requirements Document**

## Feature Name: Objectivity/Superconnector

Version: 3	Completed By:	Leon Guzenda	Date Submitted: 12/10/07
Version: 2	Completed By:	Leon Guzenda	Date Submitted: 12/5/06
Version: 1	Completed By:	Leon Guzenda	Date Submitted: 11/13/06

## **Description of the Problem**

Objectivity/DB is an ODBMS engine, rather than an application. It does very little when the user "opens the box". It is attractive to object oriented developers, but it may take many days to build a prototype application. We need something to jumpstart developers in our prime markets. We also need more than a single product line.

#### **Description of the Requested Feature**

- 1. A customizable data fusion and visual analytics framework for: rapidly ingesting textual (or certain types of multimedia information) into an Objectivity/DB federation; performing various types of information analysis and correlation; and providing advanced visualizations, datamining and ad hoc queries. The ingest may be from files, streams or external DBMSs. There will be a standard XML ingest component, but the user must be able to rapidly integrate other sources.
- 2. A simple API for porting analysis, correlation and visualization components to the framework.
- 3. A simple workflow management component for controlling ingest, analysis and reporting tasks.
- 4. Visual components for: object modeling; physical resource configuration; information security; interfacing to external security mechanisms; datamining and relationship visualization.

The major application components may be open source or partner applications or tools.

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

This is a new feature that leverages existing strengths:

- Eclipse compatibility,
- Powerful object modeling
- \* A scalable high ingest database engine
- ✤ XML import
- Distributed Parallel Query Engine
- Ultrafast connection navigation.

It will also showcase major Release 10 features, including:

- Federated Queries (PQE)
- Unified Language Binding
- Enhanced Queries
- Encrypted pages/objects
- Grid enablement
- ♦ Generic iterator

# **Conceptual Overview**

The diagram below shows how the various components interrelate.



The Workflow and Security Administration component is used to:

- Adapt the object model to handle new kinds of data, if necessary.
- Configure the logical to physical layout of the federation (e.g. directing XML objects to one group of databases and video objects to another).
- Tie in external security mechanisms.
- Schedule the pipeline of ingest, analysis, correlation, reporting and visualization tasks.

A typical flow would be:

- Ingest a group of XML files
- Use a keyword analyzer to identify words in the text.
- Use a keyword Correlator to build associations between keyword occurrences.

- Use a semantic component to apply ontology rules and build semantic associations between objects, e.g. grouping together anything to do with maritime vessels.
- Use a geospatial gazetteer to identify place names and other geopolitical connections.
- Correlate all objects into geopolitical groups.
- Identify timelines and chronological clues in the text, e.g. "do something" after "event X", or "on a Public Holiday".
- Run a predefined report looking for all information relating to a particular country, type of ship and timeframe.
- Manually run a visualization program to help analyze the new data and its connections to old data. Answer questions such as "Are there any links between these two objects?" and display the paths.

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

As stated above, developers currently have to learn Objectivity/DB and write new applications before being able to do anything useful with Objectivity/DB. There are some sample applications, but they are very limited in scope.

## What languages must support this capability?

The framework will be compatible with all existing APIs. However, its components may be written in C++, Java or Python. There must be Java equivalents of any C++ module and vice versa.

## Which platforms must be supported?

• All Tier 1 platforms.

## Do any competitors already have this feature?

- There are numerous third party equivalents for DB2, Oracle, MySQL and SQL Server, often using ODBC as the interface. However, none of these products is likely to match the ingest and navigational performance of Objectivity/DB.
- There are no known equivalents for db4objects, ObjectStore or Versant. The closest match is in the Complex Event Processing, or Streamed Event Processing domain.

## Customers who require this feature

• Intelligence community customers and integrators (NGC has expressed an interest).

- Complex financial and other IT applications.
- Manufacturing applications.
- Very Large Database [VLDB] applications, e.g. in Big Science.

#### Revenue at risk, or which could be won

This product is a new line of business, as it can be sold as a freestanding application or as a modifiable framework. A full Business Plan will be produced in support of this MRD.

#### When is this required?

- Phase 1, based on Release 9.3 capabilities June 2007
- Phase 2, fully functional With Release 10.

#### **Additional Notes**

1. We will also need:

- Marketing collateral, including promotional material and a special area on our web site.
- Technical Publications.
- New QA material to prove that the framework works and is interoperable with other platform and language combinations.

2. Licensing costs are to be determined. One possibility is to make the framework Open Source, but to charge for runtime (End User) licenses. The framework may itself be an adaptation of one from the open source community, such as IBM's UIMA. Another option is to license or buy the source code of the framework that one of our Canadian customers, the Communications Security Establishment, has developed.

3. Note that the standard components may all be currently available (or adapted) open source tools or applications. Appendix A lists some potential candidates. It is far from exhaustive, but it gives an idea of what is currently available.

### APPENDIX A – CANDIDATE TECHNOLOGIES (All Open Source unless stated otherwise)

#### A1. Workflow and Scheduling

#### \* **<u>Quartz Enterprise Job Scheduler</u>**

- Quartz is a full-featured, open source job scheduling system that can be integrated with, or used along side virtually any J2EE or J2SE application from the smallest stand-alone application to the largest e-commerce system.
- It is primarily time driven, but it has listener events.
- Enhydra Shark Project Fully featured workflow solutions, including the Together Workflow GUI and server.

Eile Connection Refreshing Misc Help				
Rennsitory management Package management Process instantiation management	ment Process instantiation management			
Process monitor User management Application mapping Cache management Worklist manageme	ent   l			
Select process				
Opened packages				
Application Repository				
Process definition-Check Credit, Version-1				
Process definition-customer Service - Request foi				
Instantiated process-Sales Order Processing-				
Participant Repository stock				
Process definition-Customer Service - notify custo				
Process definition-Produce widgets, Version-1				
Shipment	_			
	- 11			
pian part i	,∼			
🕥 Start 🕺 Suspend 🍙 Resume 🔤 Terminate 💥 Abort 🛛 🖆 Show history 📃 Description				
🔇 Delete 🚳 Reevaluate assignments 🚳 Check deadlines 🍦 Check limits				
🜀 Go back 💽 Go forth 🗞 Get previous options 🧏 Get next options 🍂 Go anywhere 📰 Fast process deletion				

AJAX Timeline – a DHTML-based AJAXy widget for visualizing time-based events. It is like Google Maps for time-based information.



Intalio – An Open Source, standards compliant Business Process Modeling System that is integrated with Eclipse.

## A2. Security

- Making our Open Source framework secure Through its Science and Technology Directorate, the Department of Homeland Security has given \$1.24 million in funding to Stanford University, Coverity and Symantec to hunt for security bugs in opensource software and to improve Coverity's commercial tool for source code analysis
- \* Security Assertion Markup Language [SAML]
- Gabriel is a security framework for Java. By using access control lists and permissions, Gabriel enables components to check access to actions. On top of that Gabriel protects methods like EJB does but without the overhead.
- Diamelle Open Source Identity and Access Management.

#### A3. Text Analytics

- IBM Unstructured Information Management Architecture [UIMA] Supports Semantic Search and could in itself be the framework that we use to construct Superconnector.
- SecondString fuzzy string matching algorithms.
- Attensity a potential partner. They sell a complete text analytics suite.
- Other potential partners include Autonomy, Clearforest, Convera, Inxight (already an informal partner), Marklogic and Megaputer.

#### A4. Ontology

- Kaon Open Source ontology management infrastructure. Includes RDF support.
- Protege-2000 Protégé-2000 an open-source, Java tool that provides an extensible architecture for the creation of customized knowledge-based applications. Protégé-2000 now provides beta level support for editing Semantic Web ontologies in OWL.
- **Ontology Works** is a commercial partner and HPKS should be a plugable component.

#### A5. Geospatial

- cpGeo A research project that indicate the kinds of technology involved.
- Geographic Resources Analysis Support System, commonly referred to as GRASS GIS, is a Geographic Information System (GIS) used for data management, image processing, graphics production, spatial modeling, and visualization of many types of data.
- deegree is a Java Framework offering the main building blocks for Spatial Data Infrastructures. Its entire architecture is developed using standards of the Open Geospatial Consortium (<u>http://www.opengeospatial.org/</u>) and <u>ISO/TC 211</u> (ISO Technical Committee 211 -- Geographic Information/Geomatics).
- **Geographic Names Information System** [GNIS] USA and Antarctic gazetteer.
- GEOnet Names Server Foreign place names gazetteer.
- Design For An Integrated Gazetteer Database A Mitre report about a project to define a schema for an integrated gazetteer.

#### A6. Image Analysis

- Section 2018 Secti
- Vigra a novel computer vision architecture and a comprehensive suite of algorithms.
- TINA (TINA Is No Acronym) is an open source environment developed to accelerate the process of image analysis research. TINA provides functionality to assist in all areas of image analysis including handling of images, image features and geometrical data; statistical and numerical analysis of data; GUI development as well as transmission and containment of data. TINA also provides a range of high-level analysis techniques for both machine vision (3D object location, 2D object recognition, temporal-stereo depth estimation, etc) and medical image analysis (MR tissue segmentation, blood flow analysis, etc).

#### A7. Grid Support

\* **Open Grid Services Architecture** [OGSA]

#### A8. Genomics

The Institute for Genomic Research [TIGR] – good source of a variety of genomic tools (microarray, sequencing/finishing, alignment, gene finding/annotation etc.)

#### **A9.** Business Intelligence

Pentaho – "The Pentaho Business Intelligence Project provides enterprise-class reporting, analysis, dashboard, data mining and workflow capabilities that help organizations operate more efficiently and effectively. The software offers flexible deployment options that enable use as embeddable components, customized BI application solutions, and as a complete out-of-the-box, integrated BI platform. Ranked #1 in Open Source Business Intelligence."