The Analyst Workbench

Overview

This system is used by analysts who work with various sorts of statistics and observations of and relationships between statistics. It provides them with the ability to store and manage sets of statistics in a central database and to perform various types of processing on them.

The key functions of the system are to provide:

- Storage of statistics
- Calculation of derived statistics, based on raw statistics that have been imported. This can be automatic, based on rules, or on demand from analyst requests.
- Management of analyst observations which are attached to statistics sets or individual pieces of data (creation, query, removal).
- Provide analysts with a user interface allowing statistics and observations to be queried, analysed, viewed and visualised.

The normal usage of the system involves analysts requesting external sets of statistics to be loaded and analysed. The Statistics Calculator is used to produce derived statistics (such as distributions, correlations and characteristic statistics) based on the raw base statistics sets. Analysts then use the user interface to investigate the statistics and record observations that they believe are significant. They can also request further calculations to be performed to provide additional insight to support their investigation.

Functional Structure

A simple sketch of the system's functional structure can be found in Figure 1.
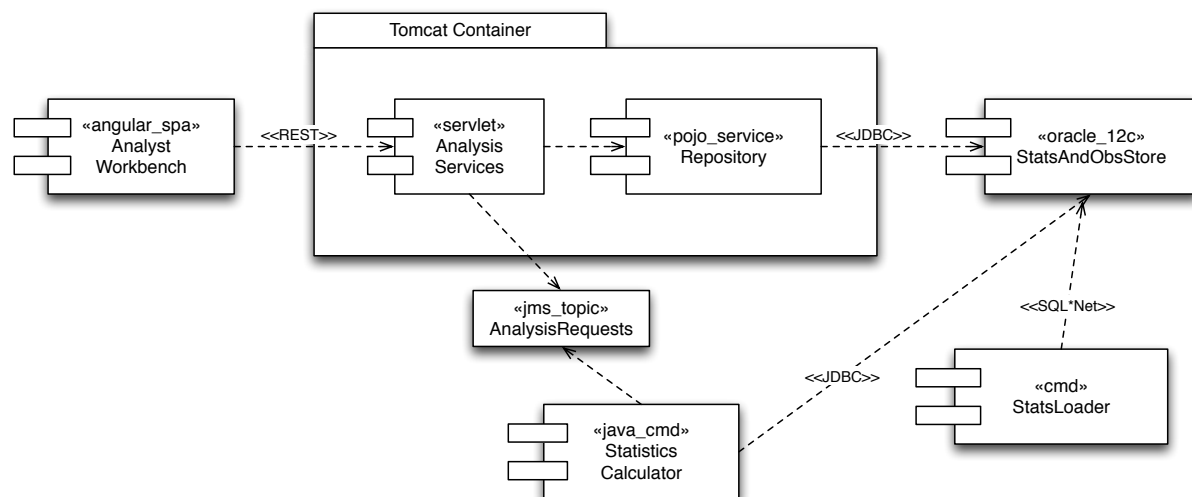

Figure 1 - Functional Structure

The elements of this system and the primary responsibilities of each are as follows:

- Analyst Workbench – a single-page-application user interface, running in a browser, which provides a rich, interactive analysis environment for analysts.

- Analysis Services – the set of network accessible services that provide the user interface with access to data and processing services. This component is also responsible for triggering automatic analysis of new sets of statistics, by recognising the arrival of the statistics and requesting a standard set of calculations to be performed by the Statistics Calculator.
- Repository – a software abstraction providing access to the underlying Oracle database that prevents the services components needing to deal with the specifics of database schema and access.
- Tomcat Container – the runtime container hosting the network services and supporting components.
- StatsAndObsStore – an Oracle relational database that is used to store all of the information within the system.
- StatisticsCalculator – provides a range of calculations on statistics sets within the system such as correlations between variables, value distributions, characteristic statistics over a set and similar. The calculator is triggered by requests arriving on the AnalysisRequests message topic.
- AnalysisRequests - a message topic that stores requests for the Statistics Calculator to perform calculations on sets of statistics held in the StatsAndObsStore.
- StatsLoader – a command line tool used to load external statistics sets from flat files.