Stock Analysis Application

Monday 11 February 2008 16:00 (25 minutes)

The proposed application will automatically manage the analysis of a large mass of financial data. For each financial instrument there is a zip file: its content is one text file per trading day containing high frequency time-series information for that instrument. Overall there are 4 TB of unzipped data: compression reduces it to roughly 100 GB. One analysis run consists in launching one job for each stock; for each instrument about 150 time-series are constructed and analysed; about 700 instruments will be analysed in each run. Many runs are expected, as both the analysis and the time intervals of interest will change during open-ended research on the properties of the data. For a reasonably exhaustive analysis on all the data, about 200 GB of zipped output files are expected.

3. Impact

The application is organised in two tiers: the first one handles the grid infrastructure, while the second one is exclusively concerned with the analysis of the data. The analysis is run in the Worker Node; it expects to have locally available a set of data files for processing, and it will produce a predefined set of local output files. The grid infrastructure code in turn consists of two parts: one to launch and monitor the analysis, and one to prepare the local environment in the WNs for the analysis. The launching and monitoring part is installed in a UI host; it accepts: a file containing the list of data to process, the analysis code to execute, and the grid output directory in a predefined secure SE. The code that prepares the WN local environment: fetches data files from the secure SE, pre-processes them, launches the analysis, clears any local temporary files, and saves them back in the SE.

URL for further information:

https://euindia.ictp.it/stock-analysis-application

4. Conclusions / Future plans

Currently the application facilitates processes that could also be achieved by grid-scripting. This is only a starting point towards a fully fledged distributed grid-application architecture WSx-compliant, integrated in the Information System and ready for QoS as an application-level grid service for financial research. The "second tier" of the application described in (3) can be viewed as a general purpose tool that is useful to any researcher wishing to perform similarly intensive analysis.

Provide a set of generic keywords that define your contribution (e.g. Data Management, Workflows, High Energy Physics)

WSx-architecture QoS Finance general data-intensive analysis

1. Short overview

The primary objective is of analysing a massive financial databases on an instrument-by-instrument basis (one instrument's data analysed at each node) but may have many other application domains. It may be a valuable tool for the grid community at large: transfers and unzips large quantities of data from secure storage to each node, performs identical computationally intensive statistical analysis of the data at each node and then zips and securely stores the voluminous results of this analysis.

Primary authors: Mr CORSO, Ezio (ICTP / EU-Indiagrid team); Dr MICHALAREAS, Giorgios (Department of

Electronics and Computer Science, University of Southampton, UK); Prof. SKOURAS, Spyros (Athens University of Economics and Business)

Co-author: Dr COZZINI, Stefano (ICTP / EU-Indiagrid team)

Presenter: Mr CORSO, Ezio (ICTP / EU-Indiagrid team)

Session Classification: Finance & Multimedia

Track Classification: Application Porting and Deployment