



Contribution ID: 32

Type: Poster

Automatic detection of error sources of failed grid jobs with data mining algorithms

Tuesday 23 September 2008 16:13 (1 minute)

Describe the activity, tool or service using or enhancing the EGEE infrastructure or results. A high-level description is needed here (Neither a detailed specialist report nor a list of references is required).

Grid error handling is a crucial task, which is very difficult due to the complexity of the Grid. It's hard to trace back errors to their real source, even though the error reporting system is well established and delivers meaningful error codes. This is a good starting point for another layer of detecting problematic Grid components. The QAOES (Quick Analysis Of Error Sources) tool applies Data Mining algorithms to detect correlations between the characteristics of Grid jobs (site, user, ...).

Report on the impact of the activity, tool or service. This should include a description of how grid technology enabled or enhanced the result, or how you have enabled or enhanced the infrastructure for other users.

There are two different approaches proposed by QAOES: firstly, a prototype to specifically distinguish between user's and site's faults by evaluating the correlations of Grid job parameters (e.g. user, site, file, submission time, ...); secondly, various Data Mining algorithms are applied using the Oracle Data Miner tool. The discovered problematic Grid components are presented as a report and in a web interface.

Describe the added value of the grid for your activity, or the value your tool or service adds for other grid users. This should include the scale of the activity and of the potential user community, and the relevance for other scientific or business applications.

By detecting the correlations between failed Grid jobs additional information is produced, which helps finding out the real source of the errors and contributes to the reliability of the grid by decreasing the time to solve problems. For one of the LHC experiments, CMS, there are currently about 30000 analysis jobs running every day, which are submitted by about 70 users to 60 different sites. Both, the Grid users as well as the Site Administrators benefit from the additional information delivered by QAOES.

Primary author: Ms MAIER, Gerhild (Universitaet Linz)

Co-authors: Dr GAIDIOZ, Benjamin (CERN); Prof. KRANZLMUELLER, Dieter (Universitaet Linz); Mrs ANDREEVA, Julia (CERN); Dr LAMANNA, Massimo (CERN)

Presenter: Ms MAIER, Gerhild (Universitaet Linz)

Session Classification: Demos and Posters

Track Classification: Poster