



Contribution ID: 223

Type: oral presentation

A study on dynamic data placement for the ATLAS Distributed Data Management system

Monday 13 April 2015 16:45 (15 minutes)

This contribution presents a study on the applicability and usefulness of dynamic data placement methods for data-intensive systems, such as ATLAS distributed data management (DDM). In this system the jobs are sent to the data, therefore having a good distribution of data is significant. Ways of forecasting workload patterns are examined which then are used to redistribute data to achieve a better overall utilisation of computing resources and to reduce waiting time for jobs before they can run on the grid. This method is based on a tracer infrastructure that is able to monitor and store historical data accesses and which is used to create popularity reports. These reports provide detailed summaries about data accesses in the past, including information about the accessed files, the involved users and the sites. From this past data it is possible to then make near-term forecasts for data popularity in the future. This study evaluates simple prediction methods as well as more complex methods like neural networks. Based on the outcome of the predictions a redistribution algorithm deletes unused replicas and adds new replicas for potentially popular datasets. Finally, a grid simulator is used to examine the effects of the redistribution. The simulator replays workload on different data distributions while measuring the job waiting time and site usage. The study examines how the average waiting time is affected by the amount of data that is moved, how it differs for the various forecasting methods and how that compares to the optimal data distribution.

Primary author: BEERMANN, Thomas (Bergische Universitaet Wuppertal (DE))

Co-authors: STEWART, Graeme (University of Glasgow (GB)); MATTIG, Peter (Bergische Universitaet Wuppertal (DE))

Presenter: BEERMANN, Thomas (Bergische Universitaet Wuppertal (DE))

Session Classification: Track 5 Session

Track Classification: Track5: Computing activities and Computing models