Contribution ID: **59**   Type: **Parallel**

Scalable proxy cache for Grid Data Access

Thursday, May 24, 2012 1:55 PM (25 minutes)

This contribution describes a prototype grid proxy cache system developed at Nikhef, motivated by a desire to construct the first
building block of a future https-based Content Delivery Network for multiple-VO grid infrastructures. Two goals drove the project:
firstly to provide a "native view" of the grid for desktop-type users, and secondly to improve performance for physics-analysis type use cases, where multiple passes are made over the same set of data (residing on the grid). We further constrained the design by
requiring that the system should be made of standard components wherever possible.

The prototype that emerged from this exercise is a horizontally-scalable, cooperating system of web server / cache nodes, fronted by a customized webDAV server. The webDAV server is custom only in the sense that it supports HTTP redirects (providing horizontal scaling) and that the authentication module has, as back end, a proxy delegation chain that can be used by the cache nodes to retrieve files from the grid.

The prototype was deployed at Nikhef and tested at a scale of several terabytes of data and approximately one hundred fast cores of
computing. Both small and large files were tested, in a number of scenarios, and with various numbers of cache nodes, in order to
understand the scaling properties of the system. For properly-dimensioned cache-node hardware, the system showed speedup of
several integer factors for the analysis-type use cases. These results and others are presented and discussed in this contribution.

Primary author: Mr CIRSTEA, Cristian (Technische Universtiteit Eindhoven)

Co-authors: GROEP, David (NIKHEF); KEIJSER, Jan Just (NIKHEF); TEMPLON, Jeff (NIKHEF (NL)); KOEROO, Oscar Arthur (Sticht. Fund. Onderzoek der Materie (FOM)); STARINK, Ronald (Unknown)

Presenter: TEMPLON, Jeff (NIKHEF (NL))

Session Classification: Distributed Processing and Analysis on Grids and Clouds

Track Classification: Distributed Processing and Analysis on Grids and Clouds (track 3)