



Contribution ID: 401

Type: oral presentation

Interactive Data Analysis on the Grid using Globus 3 and JAS3

Thursday, September 30, 2004 5:30 PM (20 minutes)

The aim of the service is to allow fully distributed analysis of large volumes of data while maintaining true (sub-second) interactivity. All the Grid related components are based on OGSA style Grid services, and to the maximum extent uses existing Globus Toolkit 3.0 (GT3) services. All transactions are authenticated and authorized using GSI (Grid Security Infrastructure) mechanism - part of GT3. JAS3, and experiment independent data analysis tool is used as the interactive analysis client.

The system consists of three main service components:

Dataset Catalog Service:

The Dataset Catalog supports browsing for an interesting dataset, or searching for data using a query language which operates on metadata stored in the catalog. The catalog makes few assumptions about the metadata stored in the catalog, except that the metadata consists of key-value pairs, stored in a hierarchical tree. The Dataset Catalog Service is designed to allow easy interfacing to existing data catalog back-ends.

Dataset Analysis Grid Service:

This service is responsible for resolving the dataset id from the catalog service, and transferring chunks of data to worker nodes for analysis processing. This service also manages the worker nodes, distributes analysis code to the worker nodes and retrieves intermediate results from the worker nodes before sending merged results back to the analysis client.

Worker Execution Services:

This service runs on each worker node and is responsible for processing analysis requests.

In this presentation we will demonstrate the current system, and will describe some of the choices made in architecting the system, in particular the challenges of obtaining interactive response times from GT3.

Primary authors: ANANTHAN, B. (Tech-X corporation); ALEXANDER, D. (Tech-X corporation); TURRI, M. (SLAC); JOHNSON, T. (SLAC); SERBO, V. (SLAC)

Presenter: JOHNSON, T. (SLAC)

Session Classification: Distributed Computing Systems and Experiences

Track Classification: Track 5 - Distributed Computing Systems and Experiences