

Resource Predictors in HEP Applications

Monday, 13 February 2006 17:00 (20 minutes)

The ATLAS experiment uses a tiered data Grid architecture that enables possibly overlapping subsets, or replicas, of original datasets to be located across the ATLAS collaboration. Many individual elements of these datasets can also be recreated locally from scratch based on a limited number of inputs. We envision a time when a user will want to determine which is more expedient, downloading a replica from a site or recreating it from scratch. To make this determination the user or his agent will need to understand the resources necessary both to recreate the dataset locally and to download any available replicas.

We have previously characterized the behavior of ATLAS applications and developed the means to predict the resources necessary to recreate a dataset. This paper presents our efforts first to establish the relationship between various Internet bandwidth probes and observed file transfer performance, and then to implement a software tool that uses data transfer bandwidth predictions and execution time estimates to instantiate a dataset in the shortest time possible. We have found that file transfer history is a more useful bandwidth predictor than any instantaneous network probe. Using databases of application performance and file transfer history as predictors and using a toy model to distribute files and applications, we have tested our software tool on a number of simple Chimera-style DAG's and have realized time savings which are consistent with our expectations from the toy model.

Primary authors: SCHOPF, Jennifer (Argonne National Laboratory); HUTH, John (Harvard University); HURST, Peter (Harvard University); GRINSTEIN, Sebastian (Harvard University)

Presenter: HUTH, John (Harvard University)

Session Classification: Distributed Data Analysis

Track Classification: Distributed Data Analysis