PhD student in grid computing

http://www.nikhef.nl/grid/trigger/phdstudent.php

ATLAS, the largest high energy physics experiment in the world, is investigating extensions of their data acquisition and trigger system over a wide area network. These so-called remote online farms will run (quasi) real-time applications on a computing grid infrastructure. This PhD research is specifically intended to address problems that arise from running such applications in an intrinsically unstable environment. The ultimate goal is the deployment of a stable and reliable (quasi) real-time application (the ATLAS higher level trigger) on top of the distributed LHC computing grid infrastructure.

Computer science aspects

- Resource management.

The LHC computing grid infrastructure has a large number of users. All of them have different kinds of shares, priorities and policies. Nevertheless, a real-time application requires a fairly stable resource pool, including the possibility for advanced planning and scheduling. This functionality is non-existing at the moment and needs to be developed.

- Fault tolerance.

Real-time applications are fundamentally different from other applications in their fault tolerance: data must be provided and processed instantly. Performance should largely be independent of failures in any individual component. The research is expected to contain a study of different approaches to deal with failures that are intrinsic to a grid environment, like redundancy and buffering in various places.

- Performance.

Performance is probably the most important aspect in this research. Performance means in this case that the entire chain needs to be operated reliably over longer periods of time without intervention, contributing significantly to the overall results of the online application. If the real-time application fails, data is lost forever. Mean Time Between Failure and Mean Time To Repair will be important parameters.

- Software management.

The heterogeneous grid environment provides an extra challenge to software management. Real-time applications put stringent requirements on version management. In general at any given moment in time, only one specific version of the software may be used for data processing on a limited number of different architectures. As a result we need to investigate virtualization techniques (*e.g.* Xen) to be able to exploit the entire LHC computing grid and not just a limited number of sites that happen to have the correct architecture and operating system installed.

Application aspects

- Integration with the data acquisition system.

The emphasis of this task will be on security and database management. ATLAS will be conservative with respect to connections of their databases and data acquisition system to the outside world and grid sites have a similar attitude. Moreover it needs to be guaranteed that data are treated consistently, *i.e.* long term trust relations are essential.

- User interface.

Eventually this application has to be managed by non-experts and integrated with the rest of the user interface for the online computer systems.