The LHCb Grid Simulation.
Proof of concept.

Mikhail Hushchyn1,2, Andrey Ustyuzhanin1,2,3, Kenenbek Arzymatov2,3, Stefan Roiser4, Alexander Baranov1
1) Yandex School of Data Analysis
2) Higher School of Economics
3) Moscow Institute of Physics and Technology
4) CERN
E-mail: mikhail91@yandex-team.ru

22nd International Conference on Computing in High Energy and Nuclear Physics,
San Francisco, October 10-14, 2016

Introduction
The LHCb Grid provides access to the data and computational resources to analyze them for
researchers with different geographical locations. The grid has hierarchical topology with multiple
sites distributed over the world. The sites differ from each other by number of CPUs, amount of disk
storage and connection bandwidth. These parameters are essential for the grid work. Moreover, job
scheduler and data distribution strategy have a great impact on the grid performance. However, it is
hard to choose an appropriate algorithms and strategies as they need a lot of time to be tested on
the real grid.

In this study, we describe the LHCb grid simulator. We compare different algorithms for the job
scheduler and different data distribution strategies.

Workflow

The jobs and data parameters can be changed to test how different models will behave
themself. For the models tesing a list of jobs was generated. Simulations of the jobs running
were done for each pair of the jobs scheduling models (Simple or DAM) and data
management models (1-7 replicas per file).

Comparison with the LHCb Dirac Web Portal

The plots above show that the DAM model is qualitatively close to the one used in the LHCb.
The Simple model results are much more different.

Playing with models
The jobs and data parameters can be changed to test how different models will behave
themselves. For the models testing a list of jobs was generated. Simulations of the jobs running
were done for each pair of the jobs scheduling models (Simple or DAM) and data
management models (1-7 replicas per file).

Conclusion
The simulation allows to evaluate the LHCb Grid behavior for the different workload and
data management models. This helps to select the appropriate models for the optimal grid
performance.

References