

The LHCb Grid Simulation. Proof of concept.



Mikhail Hushchyn^{1,3}, Andrey Ustyuzhanin^{1,2,3}, Kenenbek Arzymatov^{2,3}, Stefan Roiser⁴, Alexander Baranov¹

- 1) Yandex School of Data Analysis
- 2) Higher School of Economics
- 3) Moscow Institute of Physics and Technology
- 4) CERN

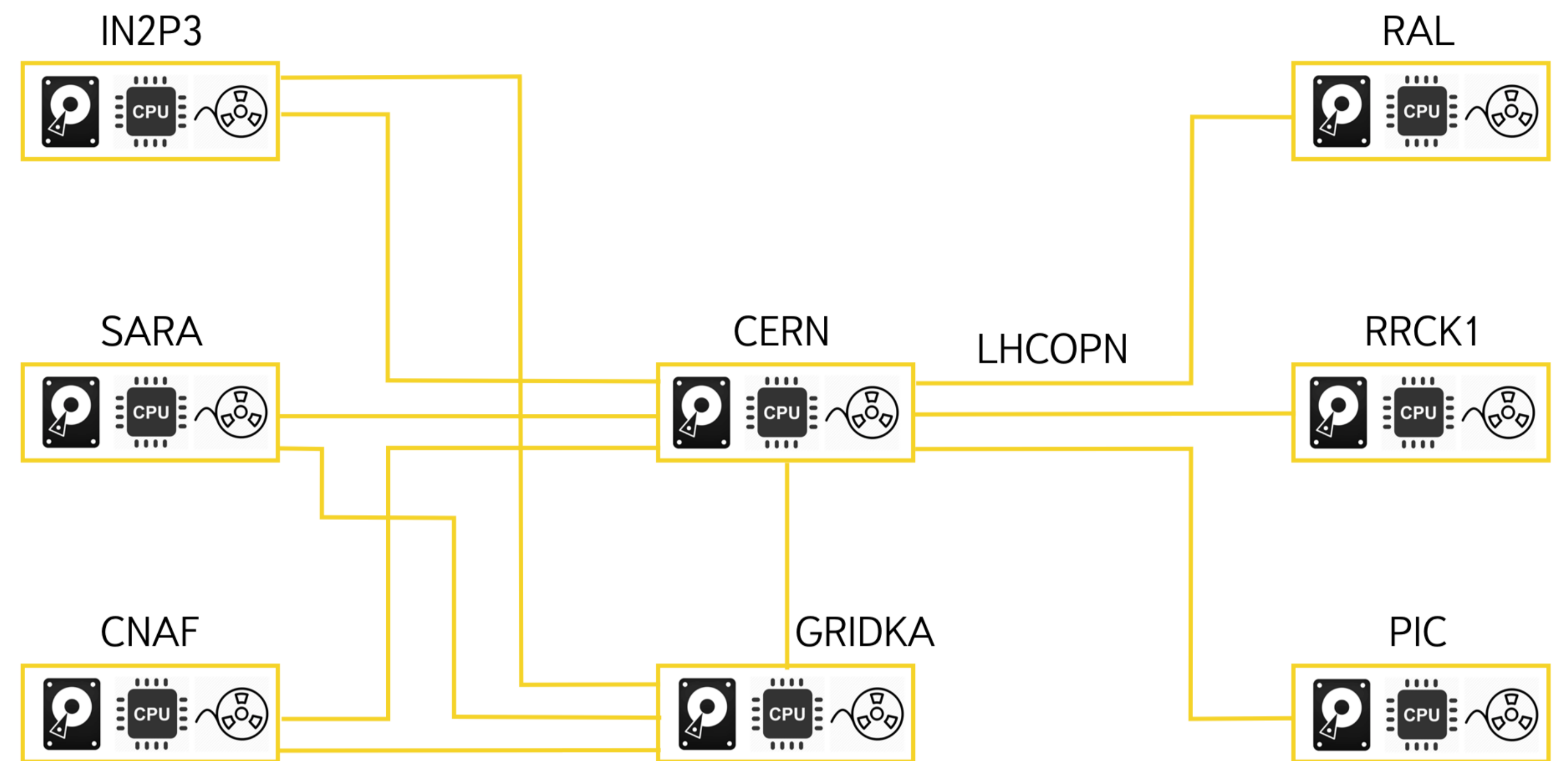
E-mail: mikhail91@yandex-team.ru

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

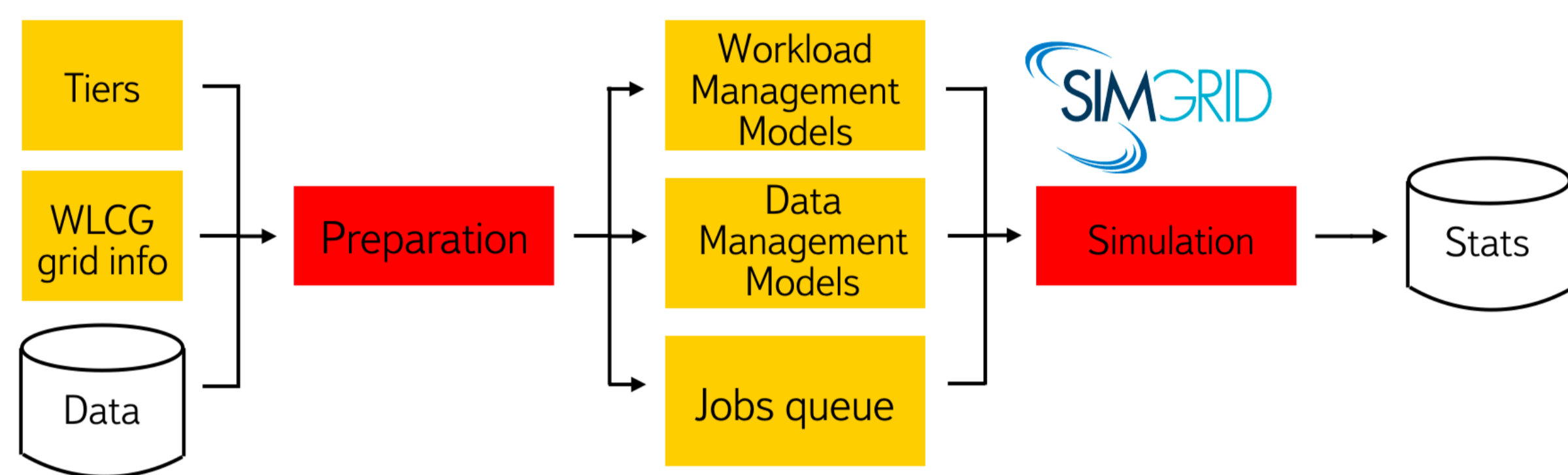
Introduction

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 simulator includes the LHCb Tier0 at CERN and 7 Tier1s with their number of CPUs, disk and tape storages, optical network.

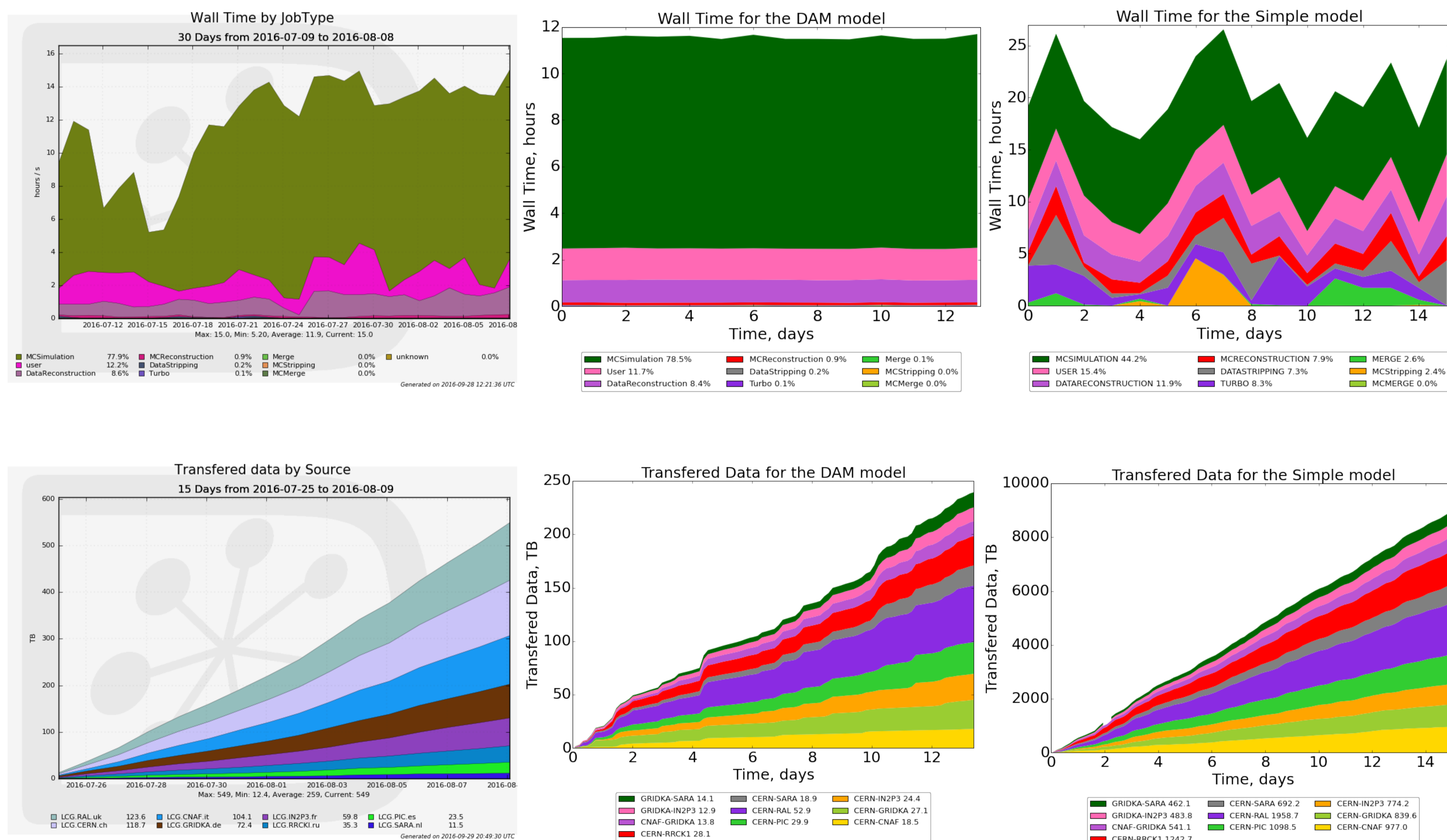
The main job and data types were simulated. Different statistics for them such as CPU time, number of jobs, input/output data size were taken from the LHCb Dirac Web Portal [1] and were used for the jobs generation.

In this study, we have tested two jobs scheduling models: Simple and Data Availability Matching (DAM). The Simple model distributes jobs by their place in the queue. The DAM model takes into account a job input data availability on a requesting tier. For the data management model the current replication strategy was used [2].

The SimGrid [3] simulator was used for the LHCb Grid work simulation.

Comparison with the LHCb Dirac Web Portal

LHCb Dirac Web Portal



Simulation

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

- [1] The LHCb Dirac Web Portal <http://lhcb-portal-dirac.cern.ch/DIRAC/>
- [2] J.P.Baud, Ph.Charpentier, K.Ciba, R.Graciani, E.Lanciotti, Z.Màthèl, D.Remenska, R.Santana The LHCb Data Management System, Journal of Physics: Conference Series 396 (2012) 032023
- [3] SimGrid <http://simgrid.gforge.inria.fr/>

