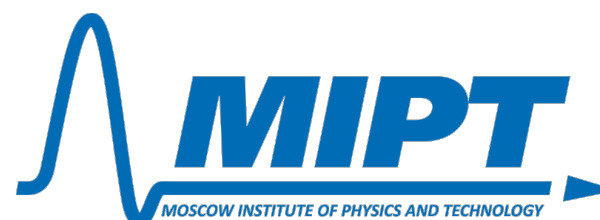




Yandex  
Data Factory



# WLCG R&D status in LHCb

Andrey Ustyuzhanin, Mikhail Hushchyn, Alexander Baranov, Kenenbek Arzymatov

WLCG ML R&D meeting, 1 June 2016

# Data collection

Collection of the historical data is in process. The data is taken from the following sources:

- › LHCb Dirac Web Portal for the aggregate historical data (<http://lhcb-portal-dirac.cern.ch>)
- › <https://ggus.eu> for the grid work anomalies and failures
- › We have asked Stefan Roiser for the more detailed data



# LHCb Grid Simulation

- › SimGrid,
- › Only Tier0, Tier1s with 5 nodes in each one;  
No Tier2s,
- › Three simple job schedulers with ‘pull’ paradigm,
- › Only 2 types of jobs,
- › Dataset replicas locations,
- › Anomalies: random node/tier failures,  
random connection break,
- › Simulation on Synthetic Data

The simulator is pretty far from the LHCb Computing Grid but it is developed.

[https://github.com/skygrid/grid\\_simulation](https://github.com/skygrid/grid_simulation)



# Future Steps of Simulation

Historical data collection:

- › Finish the data collection form LHCb Dirac Web Portal and from other sources

Grid Structure:

- › Add Tier2s
- › Add more nodes to each of the Tiers

Data Storages:

- › Add disks and tapes
- › Add different file types and their replicas in accordance with LHCb

Workload:

- › Add different job types in accordance with LHCb
- › Reproduce Dirac workload scheme, including the job scheduler

Anomalies:

- › Add more anomalies and failures from [ggus.eu](http://ggus.eu)

Thanks for your attention!