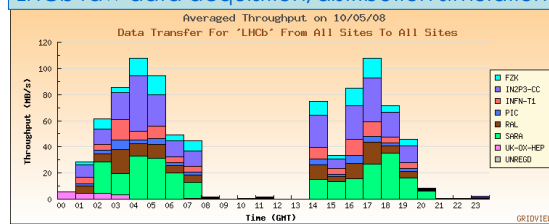
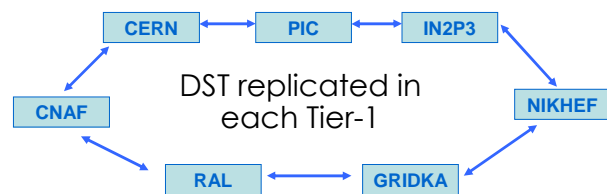


- In order to exploit the distributed resources over the Grid, the LHCb computing model is based on hierarchical system of computing centres called Tier.
- When The LHCb experiment will come online the raw data will be acquired at a rate of 2 kHz, with an average event size of 35 kByte, hence with a throughput of 70 MB/s stored on tape at CERN (Tier-0).
- One copy of raw data will be distributed to 6 Tier-1 sites where will be stored on tape as a backup copy and processed. The output, rDST, will be locally used to perform preliminary analysis process

LHCb raw data acquisition/distribution simulation



- The rDST data samples are then selected by means of selection algorithms (data stripping) tuned to reduce the size of the samples to be easily analyzed by the final users.
- The output of data stripping is in the DST format. DST data samples are distributed from the Tier-1 sites where they are produced to the other Tier-1 centres, such to have a copy of data available at any site for users analysis.
- LHCb's Tier-2 computing centres (14 centres), differently to the other LHC experiments, are planned to be used mainly for the production of Monte Carlo simulated events.
- The LHCb Italian Tier2 computing centre is located at CNAF



The distributed analysis model foresees the usage of the LHCb Tier-1 centres for data analysis. In order to analyze the data the LHCb collaboration will exploit the Grid distributed computing resources spread over the world. Two main software tools will be used : DIRAC and GANGA.

- DIRAC is a software mainly used to easily access distributed computing resources
- GANGA is the Grid user interface to allow user an easy way to submit analysis jobs in a data driven mode, through DIRAC that takes care to execute the job over the Grid



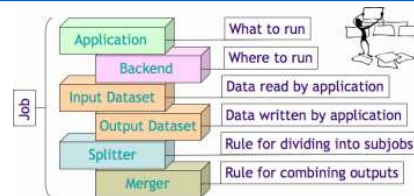
The DIRAC (Distributed Infrastructure with Remote Agent Control) project began as a solution for the LHCb experiment at CERN to carry out massive Monte Carlo simulation, data processing and distributed user analysis on various computing resources.



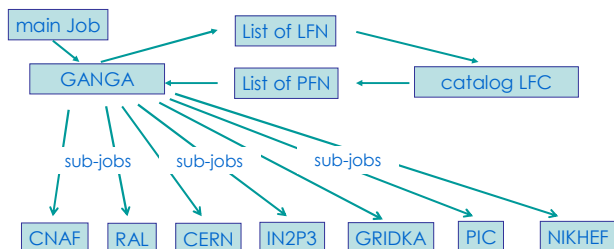
Within the LHCb collaboration the analysis are realized by means of GANGA (Gaudi/Athena and Grid Alliance) specifically designed to provide the user with a very easy-to-use mean to submit jobs to the Grid. GANGA deals with configuring the LHCb applications, submitting the jobs and keeping track of the results.

An analysis job in GANGA consists of a set of building blocks:

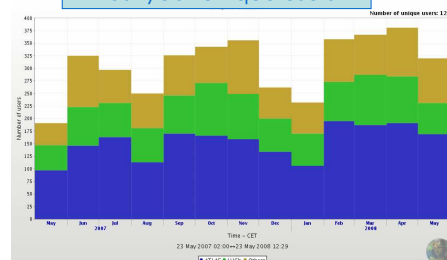
- **application** → the analysis software to be executed
- **backend** → the processing system where the application will run (local, batch system, remote computing node of the Grid through DIRAC). In the LHCb computing model it is foreseen to mainly use DIRAC as backend order to have a centralized accounting and job prioritization system.
- **input dataset** → list of file names to be processed.
 - The user selects data sets as a Logical File Name. To a given LFN may correspond several replicas (at least one) located at Tier-1 sites. In this way the submission to a site is data driven.
- **splitter** → the possibility to split the job in sub-jobs and submits them independently
- **merger** → the possibility to merge the output of the sub-jobs into a single output.



Job splitting and data drive submission



Last year Unique users



GANGA is largely used within LHCb collaboration. The number of total unique users from May 2007 to May 2008 running analysis jobs through GANGA were in average about 260.