

Grid technologies for large-scale projects



N. S. Astakhov, A. S. Baginyan, S. D. Belov, A. G. Dolbilov, A. O. Golunov,
I. N. Gorbunov, N. I. Gromova, I. A. Kashunin, V. V. Korenkov,
V. V. Mitsyn, V. V. Palichik, S. V. Shmatov, T. A. Strizh, E. A. Tikhonenko,
V. V. Trofimov, N. N. Voitishin, V. E. Zhiltsov

Laboratory of Information Technologies

JINR

Alushta, Crimea

2016

Grid technologies - a way to success



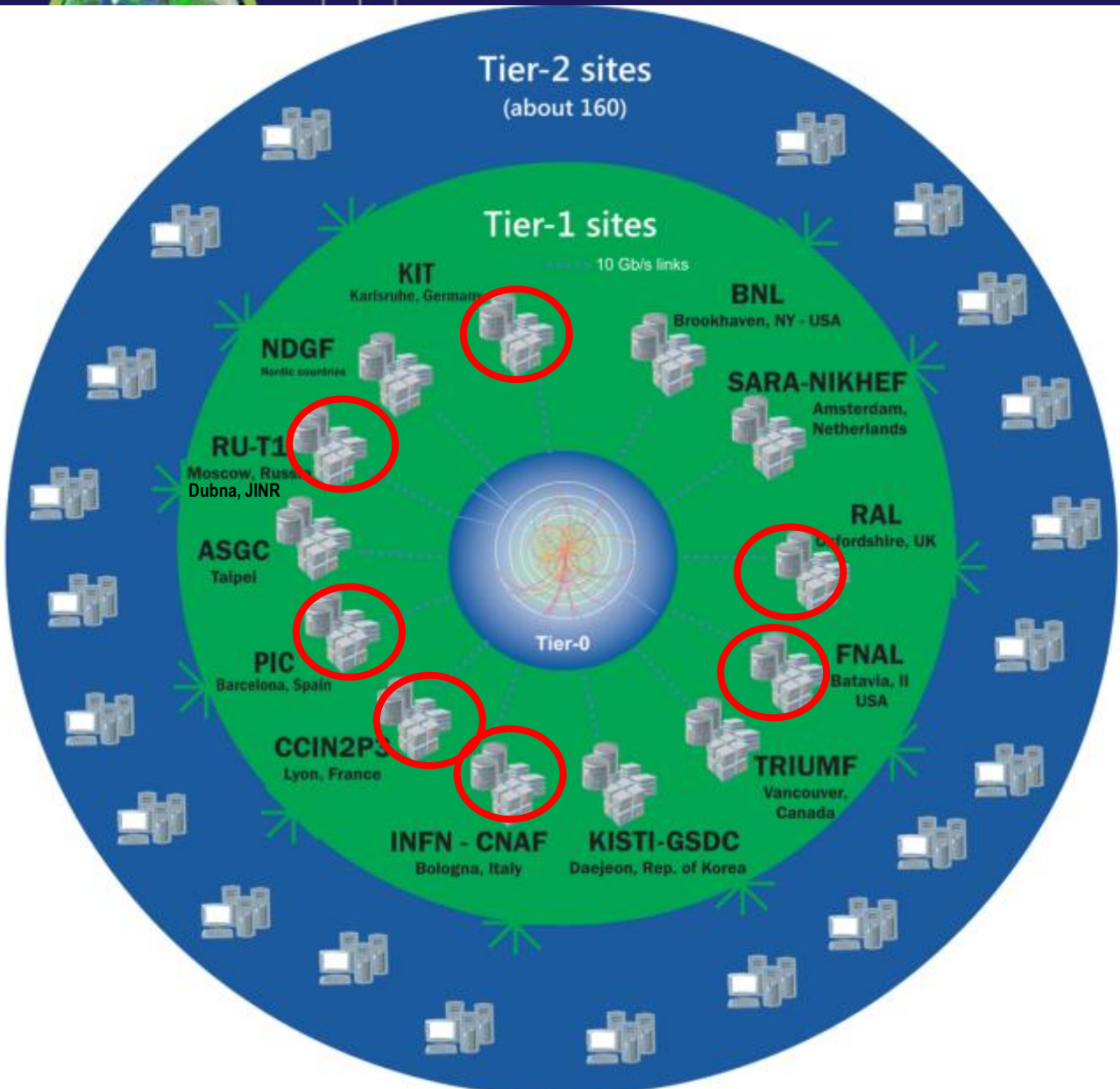
On a festivity dedicated to receiving the Nobel Prize for discovery of Higgs boson, CERN Director professor Rolf Dieter Heuer directly called the grid-technologies one of three pillars of success (alongside with the LHC accelerator and physical installations).



Without implementation of the grid-infrastructure on LHC it would be impossible to process and store enormous data coming from the collider and therefore to make discoveries.

Nowadays, every large-scale project will fail without using a distributed infrastructure for data processing.

LHC Computing Model



Tier-0 (CERN):

- Data recording
- Initial data reconstruction
- Data distribution

Tier-1 (11 → 14 centres):

- Permanent storage
- Re-processing
- Analysis
- Simulation

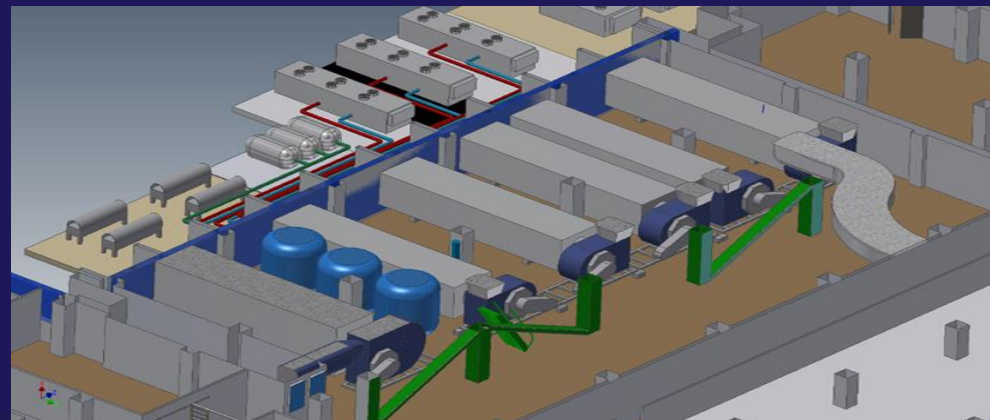
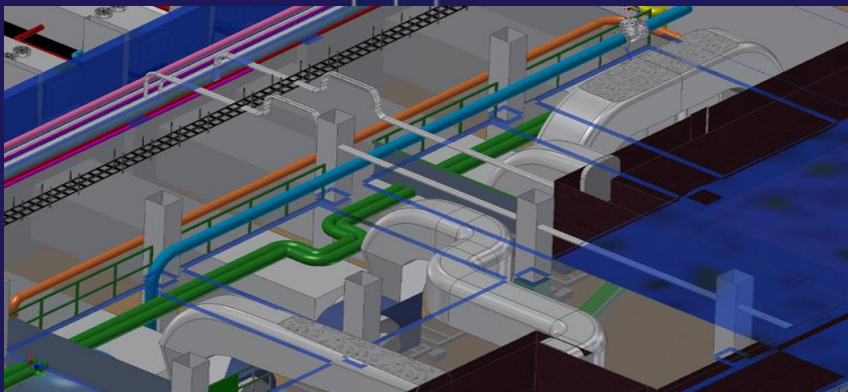
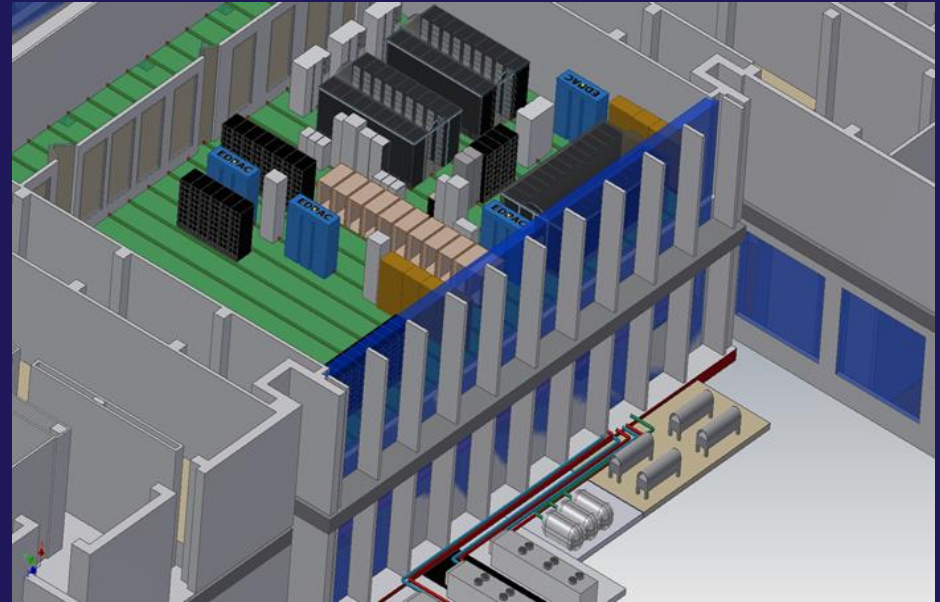
Tier-2 (>200 centres):

- Simulation
- End-user analysis

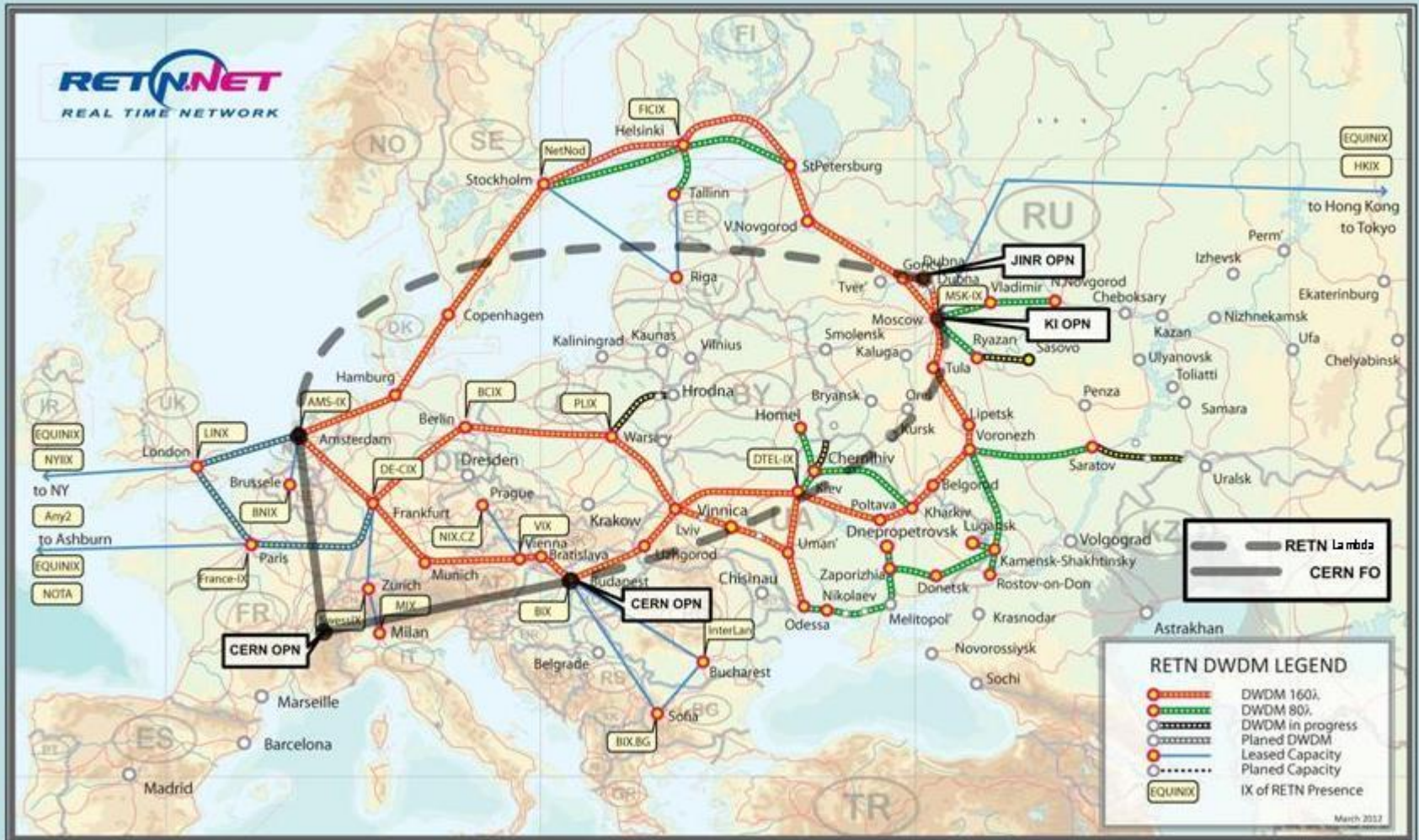


Creation of CMS Tier1 in JINR

- Engineering infrastructure (a system of uninterrupted power supply, climate - control);
- High-speed reliable network infrastructure with a dedicated reserved data link to CERN (LHCOPN);
- Computing system and storage system on the basis of disk arrays and tape libraries of high capacity;
- 100% reliability and availability.



JINR Tier-1 Connectivity Scheme



Components of the Tier1



Cooling system



Computing elements



Tape Robot

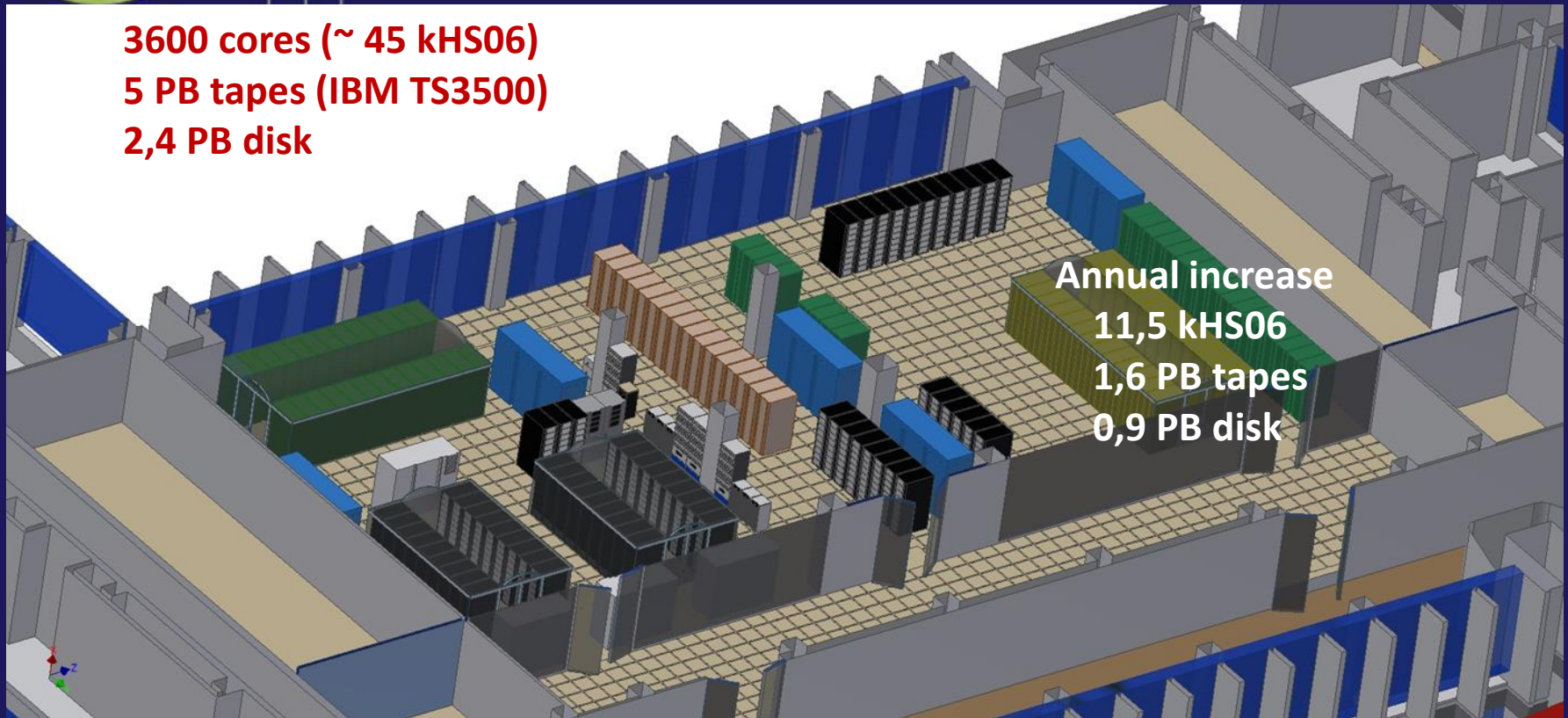


Uninterrupted power supply



Current configuration and plans

3600 cores (~ 45 kHS06)
5 PB tapes (IBM TS3500)
2,4 PB disk



Annual increase
11,5 kHS06
1,6 PB tapes
0,9 PB disk

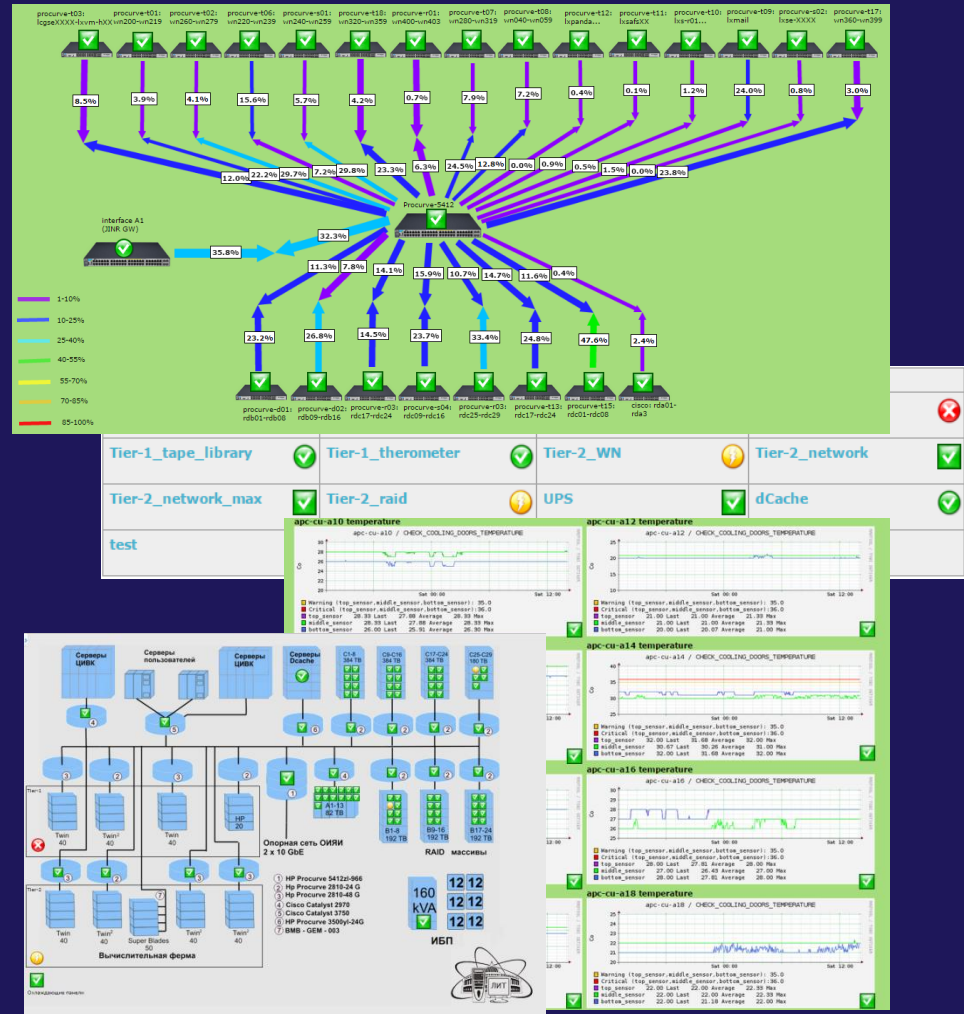


JINR Tier-1 monitoring system

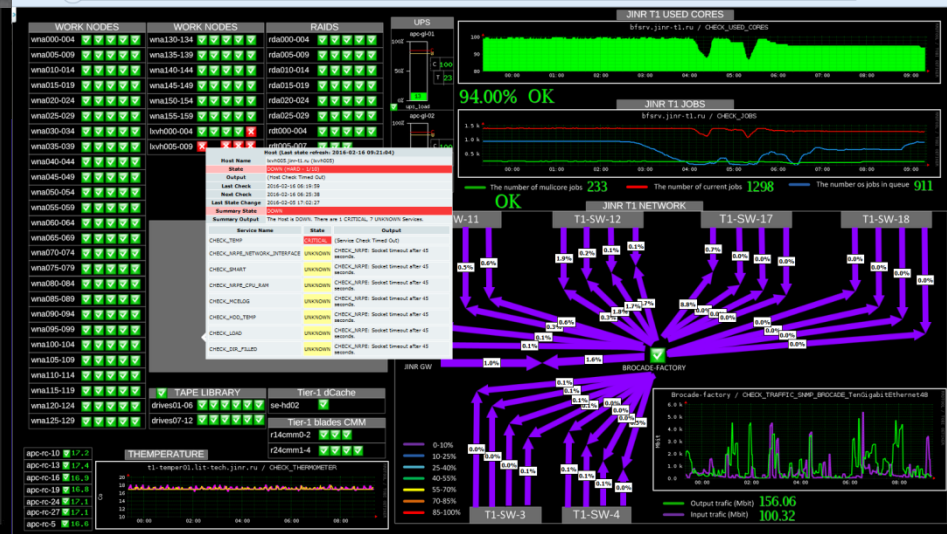
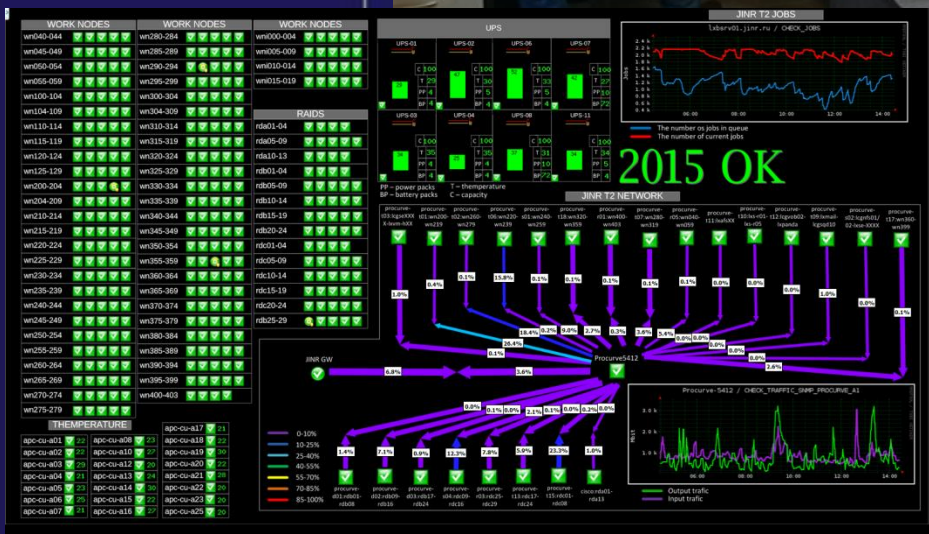
JINR Tier-1 monitoring system provides real-time information about:

- * work nodes;
- * disk servers;
- * network equipment;
- * uninterruptible power supply elements;
- * cooling system.

It also can be used for creating network maps and network equipment load maps, for drawing state tables and different plots.



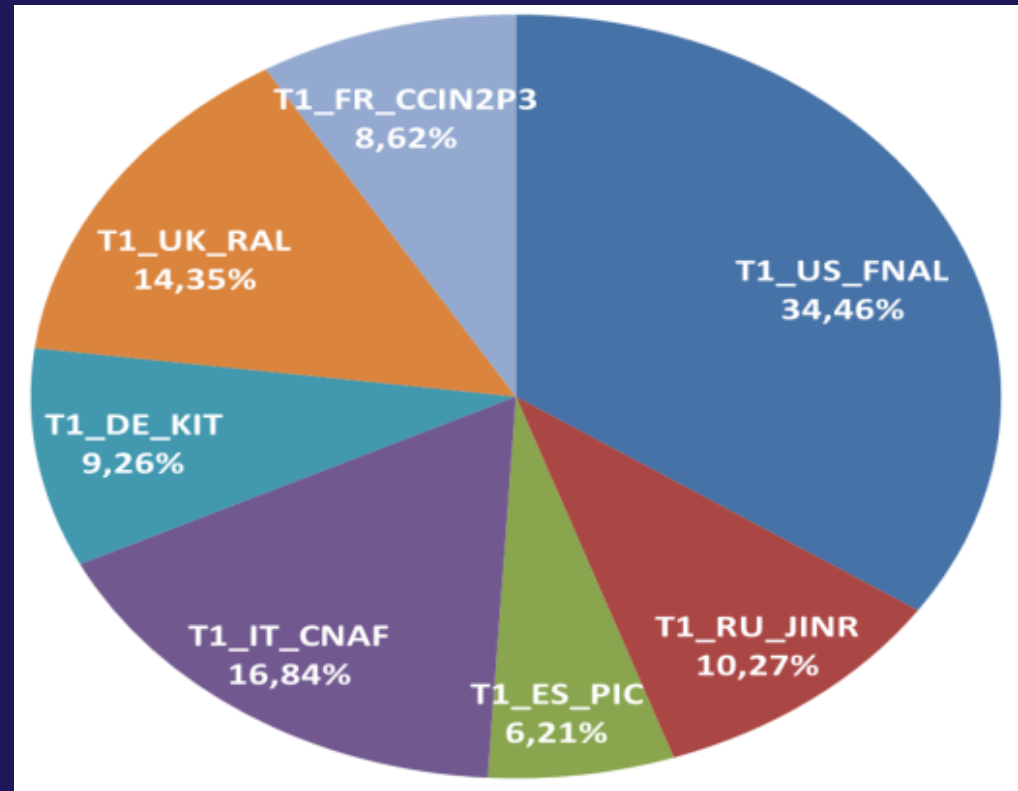
JINR Tier's Dashboard



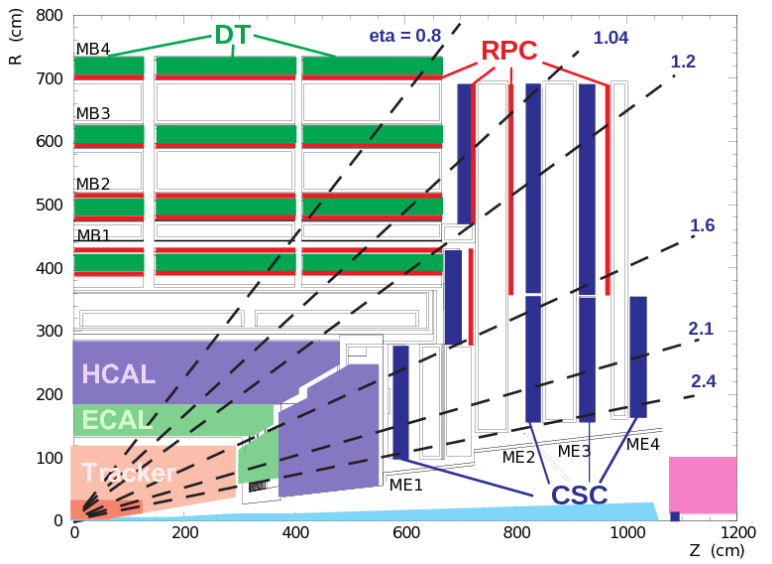
Usage of Tier1 centers by the CMS experiment (last month)



JINR Tier-1 CMS
617 413 jobs



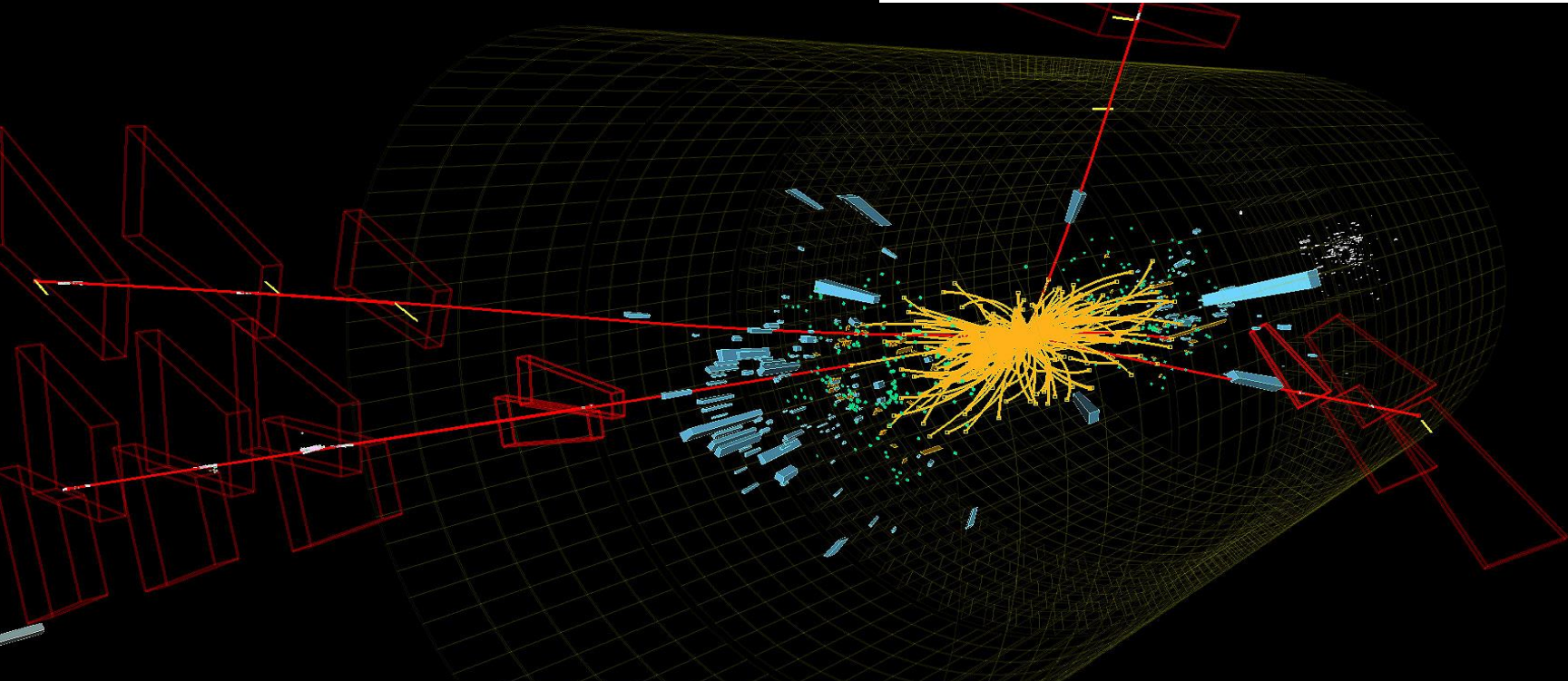
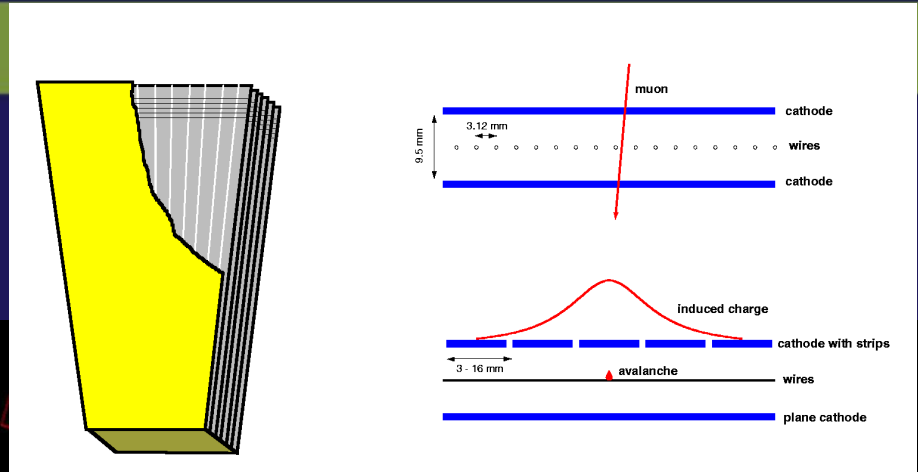
JINR LIT CMS group Cathode Strip Chambers (CSC)



CMS Experiment at the LHC, CERN

Data recorded: 2011-Jun-05 09:01:21.346043 GMT(04:01:21 CDT)

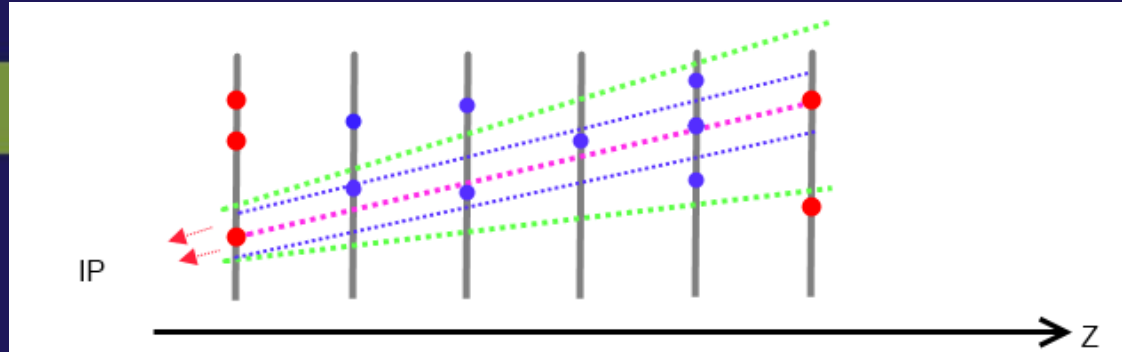
Run / Event: 166512 / 337493970



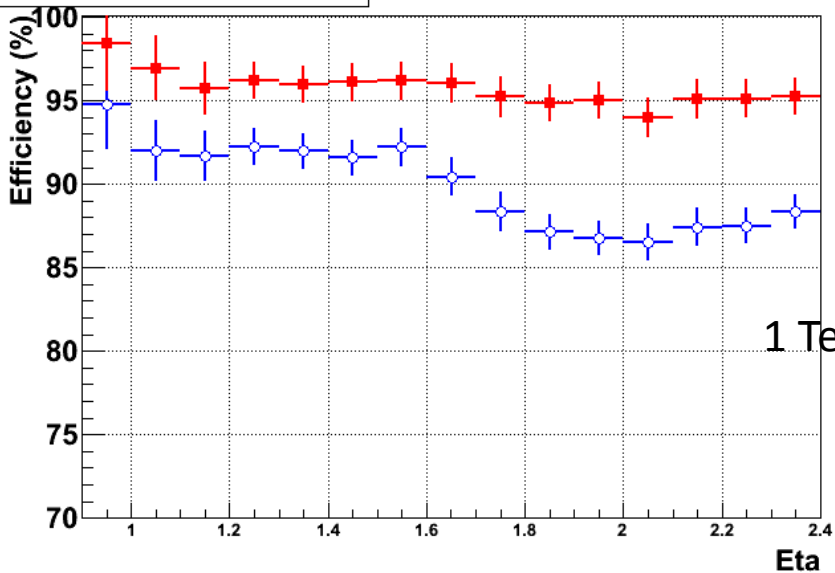
Development of a new CSC segment building algorithm



The new CSC segment building algorithm takes into account the interaction point

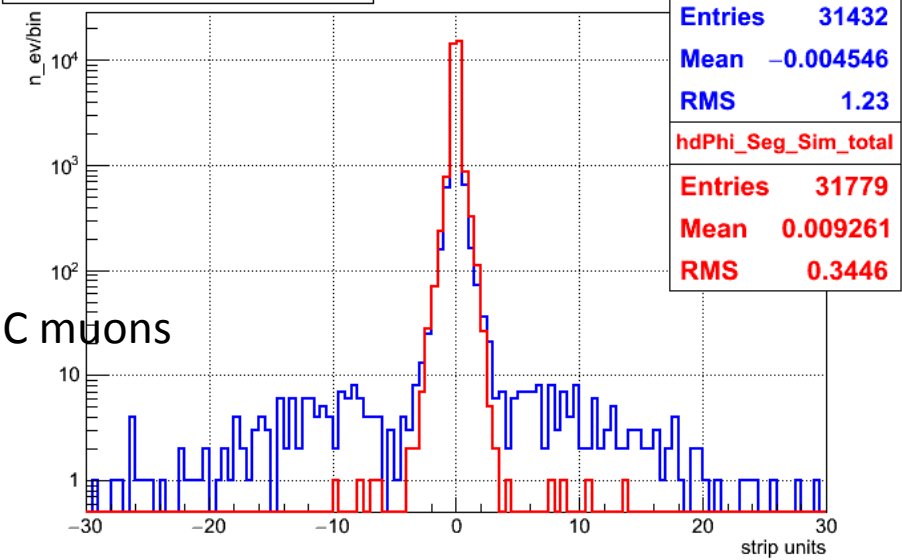


hRH_EfficiencyvsEta



1 TeV MC muons

dPhi_Seg_Sim_total

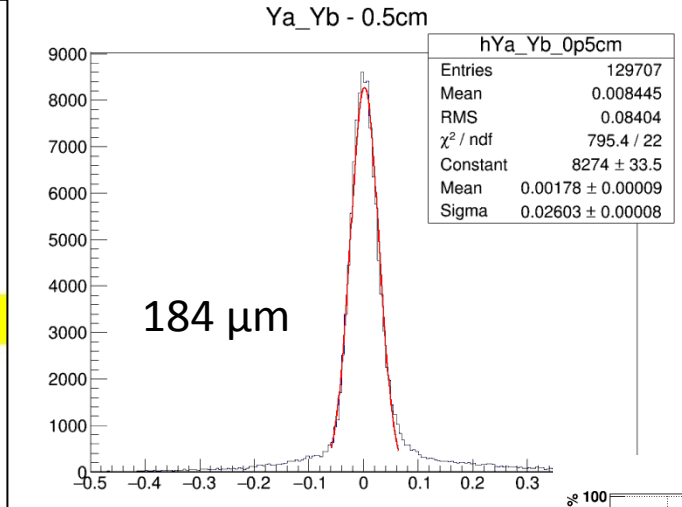
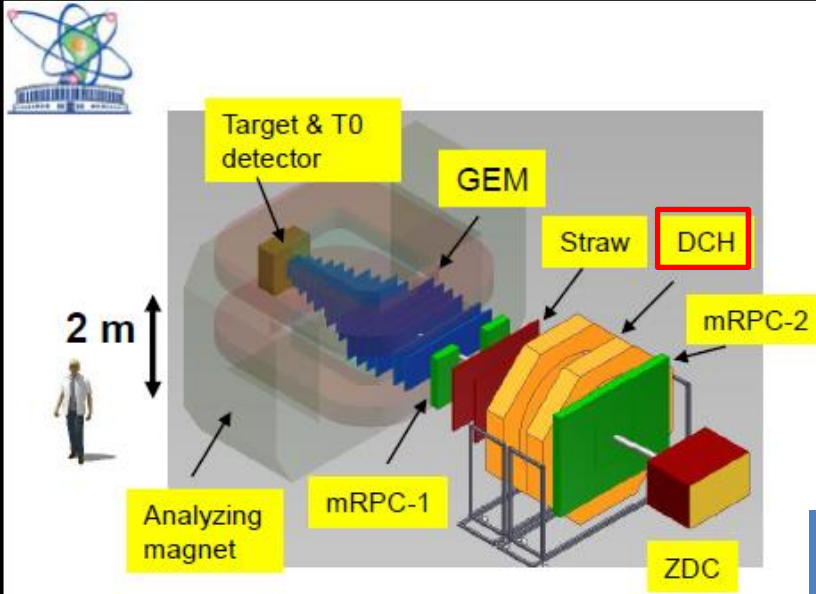


Reconstruction efficiency vs pseudorapidity

Difference in angles of the reconstructed and simulated segments (strip_unit ~ 3mrad)

~5000 CRAB jobs
~5TB skimmed datasets

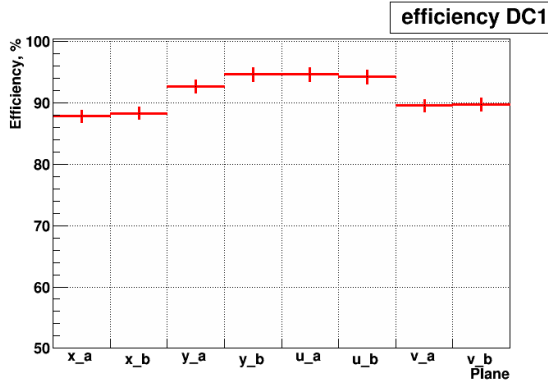
Baryonic Matter at Nuclotron



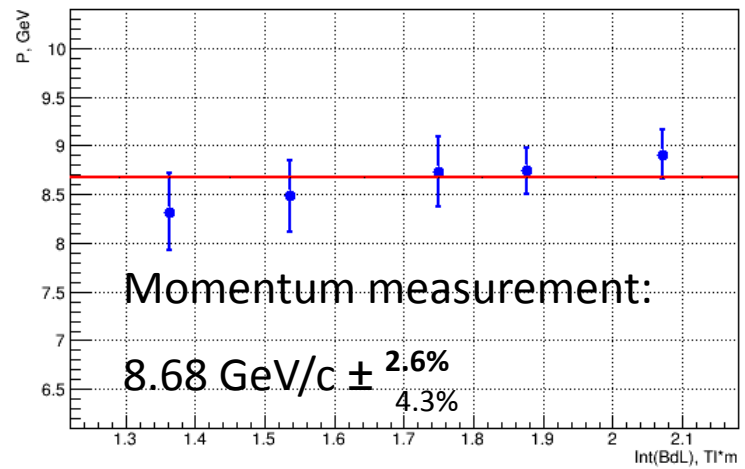
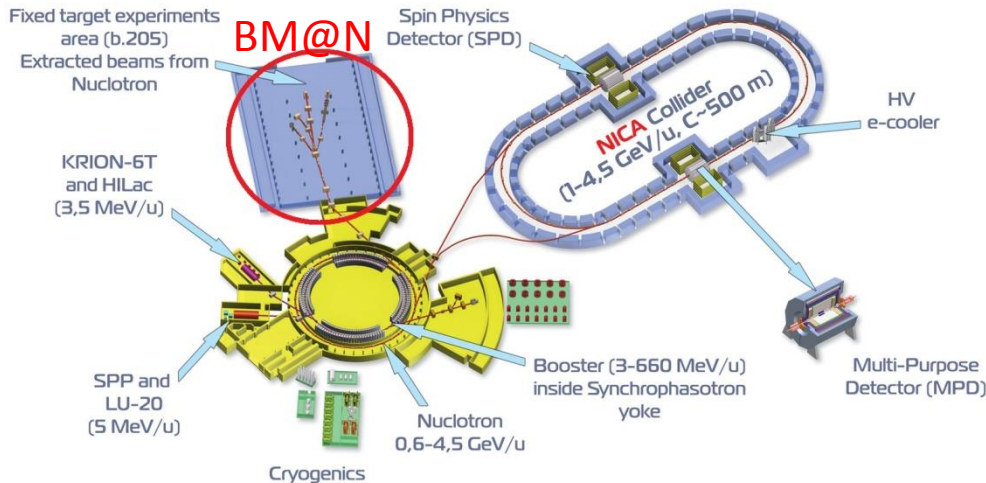
DCH Detector Resolution

DCH Detector Efficiency

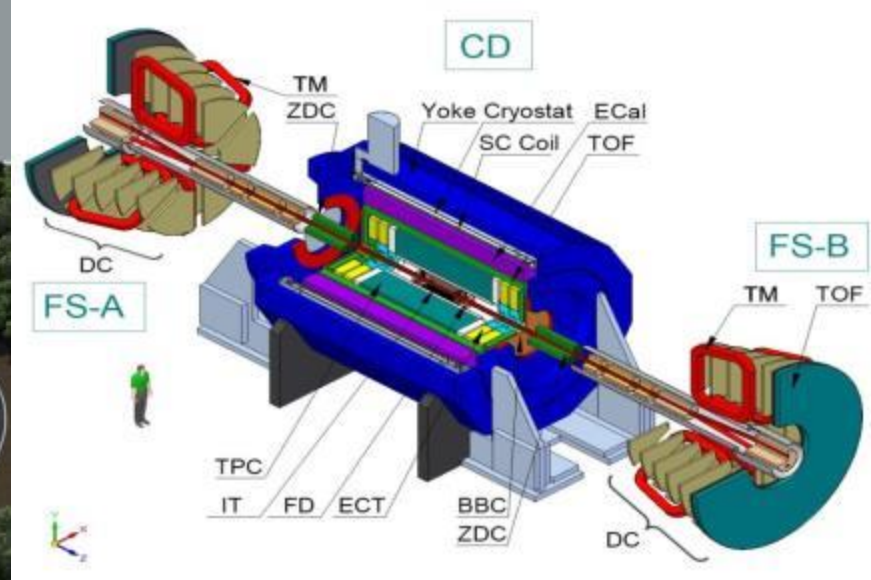
- Fixed target experiment
- Proton and gold beams
- 1-6 GeV per nucleon



Superconducting accelerator complex NICA (Nuclotron based Ion Collider fAcility)



NICA Project



NICA project data amount estimation:

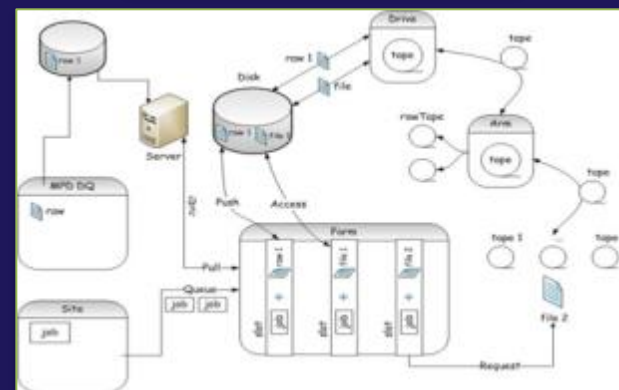
- High frequency of registered events (up to 6 KHz);
- in Au-Au collisions at the expected energies more than 1000 charged particles per event will be produced;
- overall number of events – 19 billions;
- overall amount of data produced per year – 30 PB and 8.4 PB more after analysis.

The model of a distributed computer infrastructure



The model for detailed investigation and estimation:

- ✓Tape robot,
- ✓Disk array,
- ✓CPU Cluster.





Conclusions

- * The concept of grid perfectly fits the LHC project, making possible tremendous amounts of data transfers and job processing.
- * The JINR Tier-1 site along with the Russia Tier-2 sites gives the possibility for physicists from JINR, member states to fully participate in processing and analysis of the LHC experimental data.
- * An enormous experience in building and maintaining big data processing and storing center was obtained that can be very useful for the development of large scale projects at JINR and other member states.

Thank you for your attention!

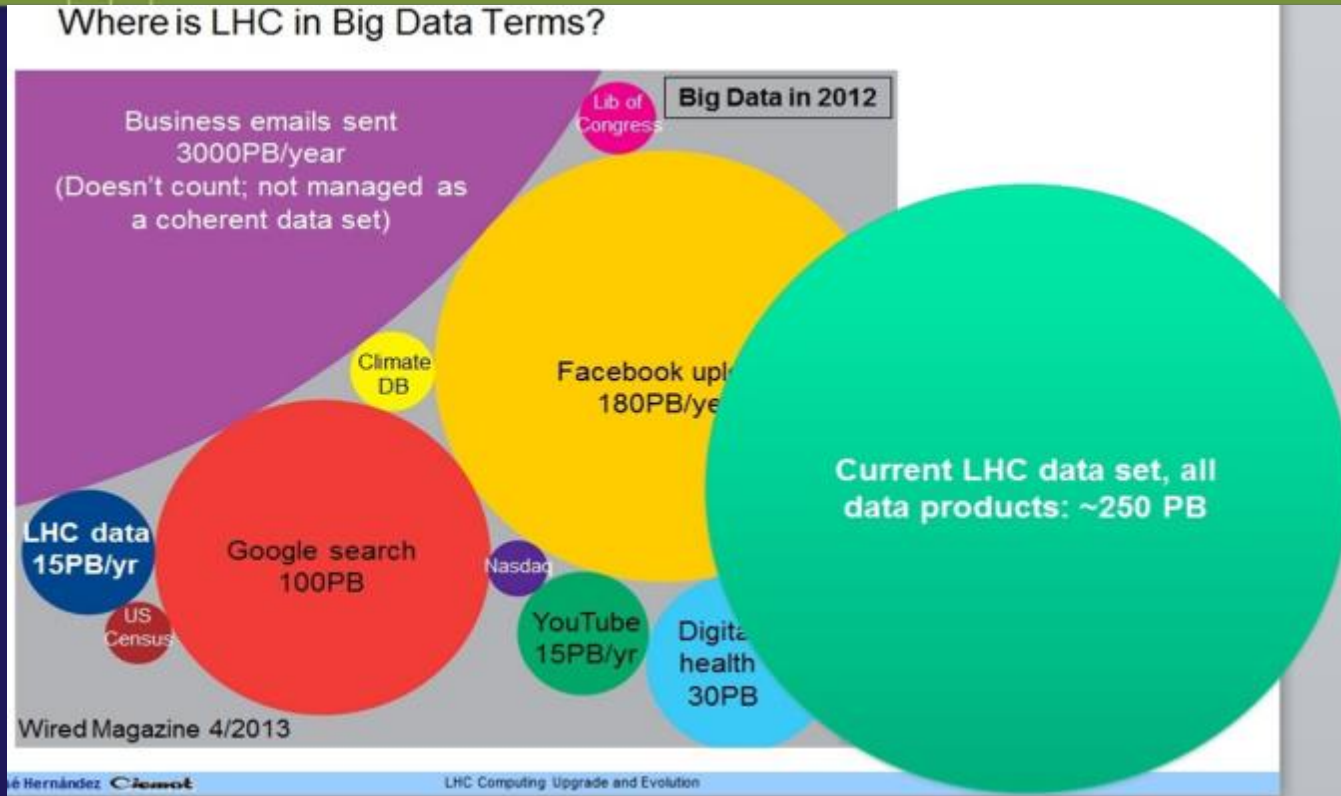




Back-up slides



Stepping into Big Data era



The comparison plot of the worldwide processed data shows that the amount of data that comes from the LHC fits the term of Big Data.

The expected amount of data received from LHC should become 2.5 times bigger in Run2, that will require the increase in resources and the optimization of their usage.