Distributed Computing for Pierre Auger Observatory

Jiri Chudoba for the Pierre Auger Collaboration
Institute of Physics of the CAS and CESNET

Goals of PAO

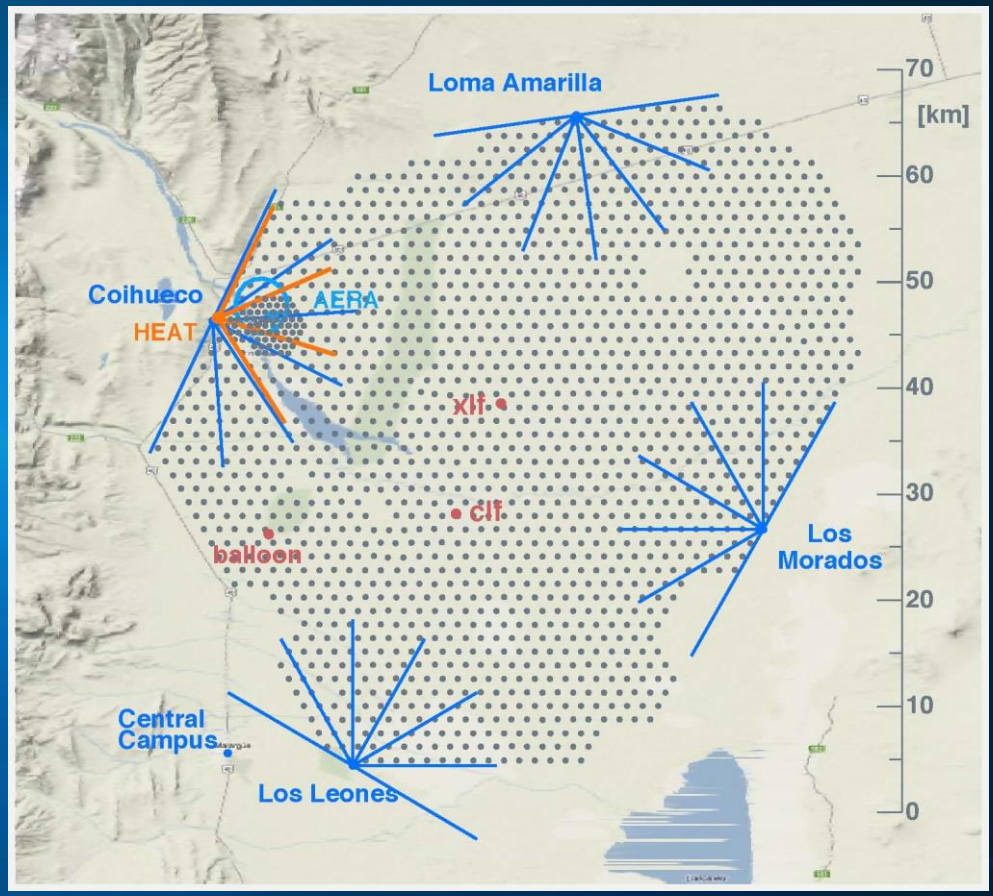
- Properties of Ultra High Energy Cosmic Rays
  - Sources
  - Energy spectrum
  - Composition
  - Existence of GZK cut-off

Tools

- Measurement of cosmic rays showers

Observatory – Hybrid Detector

- Surface detector
  - 3000 km^2
  - 1671 water Cherenkov detectors
  - Spacing 1500 m (E>$10^{18.5}$ eV)
  - infill array 750 m (E>$10^{17.5}$ eV)
- Fluorescence detector
  - 4x6 (FD) + 3 (HEAT) telescopes on 4 sites
- Radio detection techniques
  - AERA - 153 stations (17 km^2)
- Auger Upgrade
  - adding scintillation detectors
  - installation 2016 - 2019



Worldwide Collaboration

More than 500 members from 16 countries

Argentina
Australia
Brasil
Colombia*
Czech Republic
France
Germany
Italy
Mexico
Netherlands
Poland
Portugal
Romania
Slovenia
Spain
USA

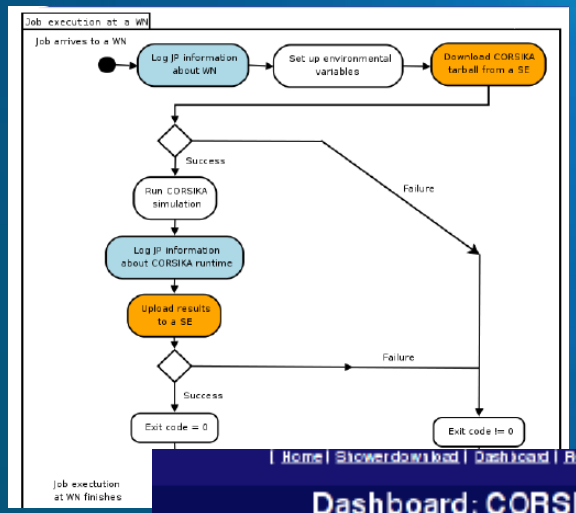
**associated*

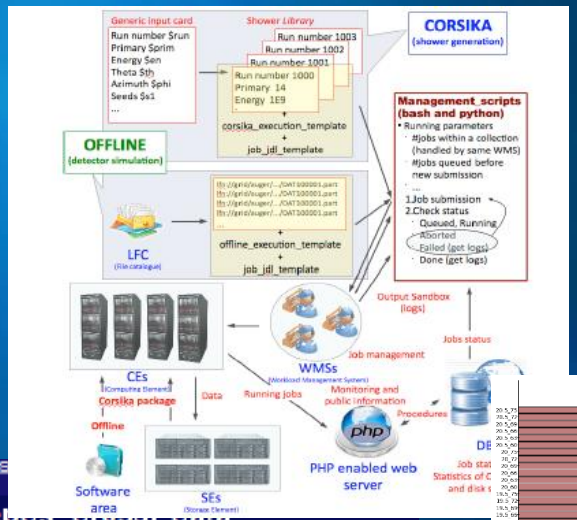
Computing requirements

- Real data processing and storage at CC IN2P3 (plus mirror at FERMILAB)
- Monte Carlo simulations on available resources
  - libraries of showers with different parameters
    - mostly CORSIKA plus Auger Offline
    - models, energy, angle, primary particle, ...
  - several CPU hours per 1 shower
  - resources:
    - local farms
    - grid sites since 2007

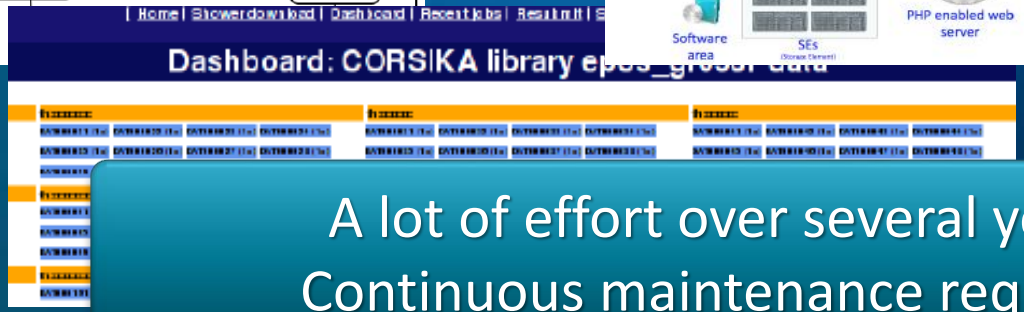
Management of Distributed Resources

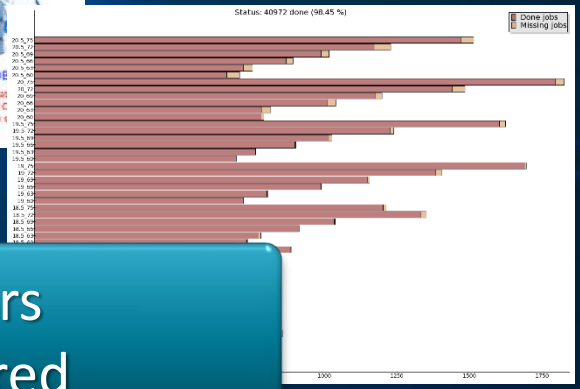
- Evolution from set of scripts around lcg mw to production system with MySQL backend
- Django dashboard, PHP scripts, SimDB





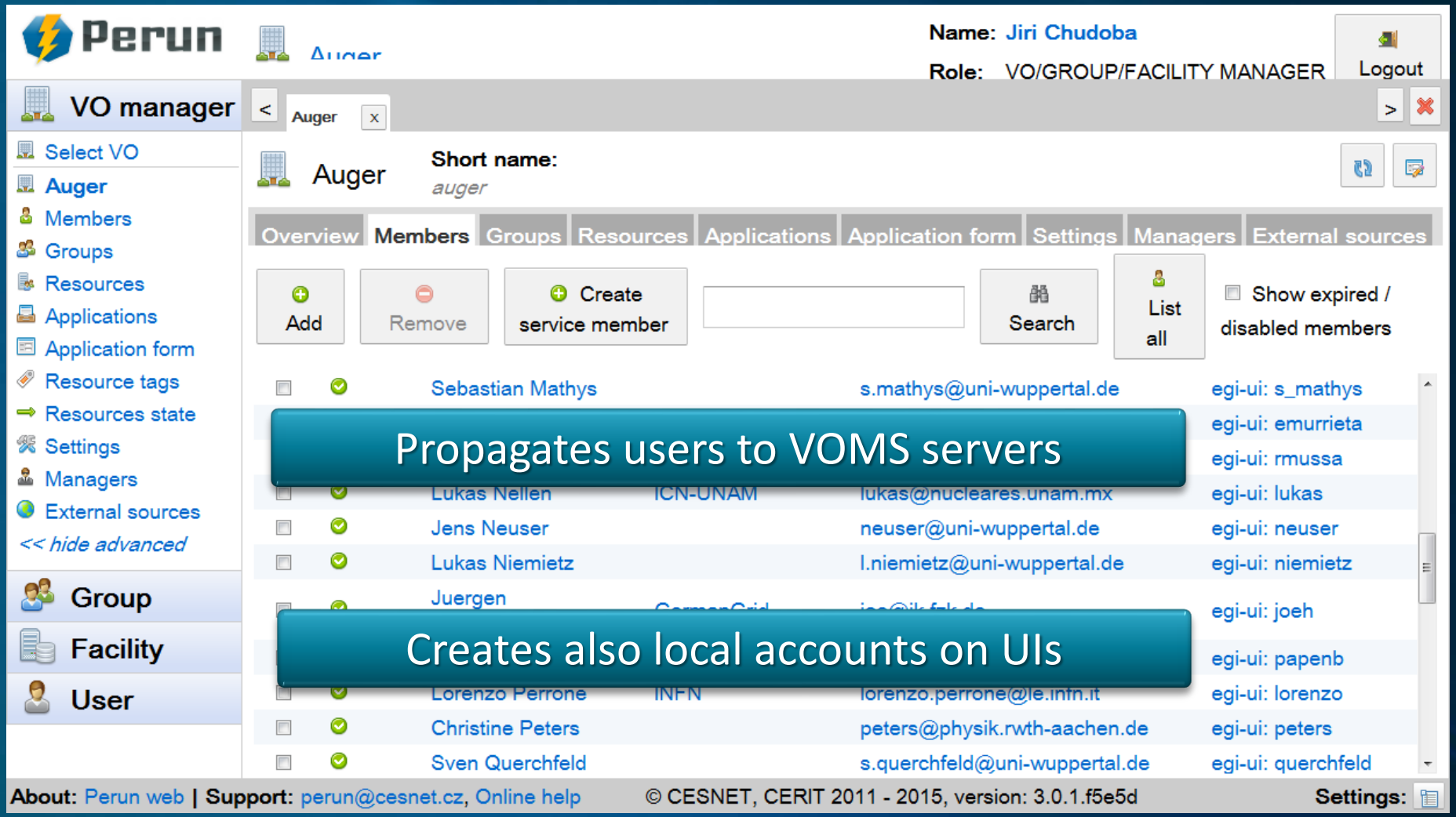
Corsika version:	6990
Fluka version:	2011.2.6
Offline version:	V2r7p7
Energies:	17.00 17.50 18.00 18.50 19.00 19.50 20.00 20.50
Theta:	60 63 66 69 72 75
Atmosphere:	1
Number of shower per energy and theta bin:	50
Total of shower:	41618
Productions:	NuEccQgsjetI_gr144 (35500 showers) NuEccQgsjetIRec_gr144 (6118 showers)
Other information:	Diverse fixed first interaction heights
Corsika status:	Corsika Status
Download Corsika LFN available in Grid:	Corsika LFN
Download CORSIKA list available in SimDB:	
Offline status:	Offline Status
Download Offline LFN:	Offline LFN
Download OFFLINE list available in SimDB:	



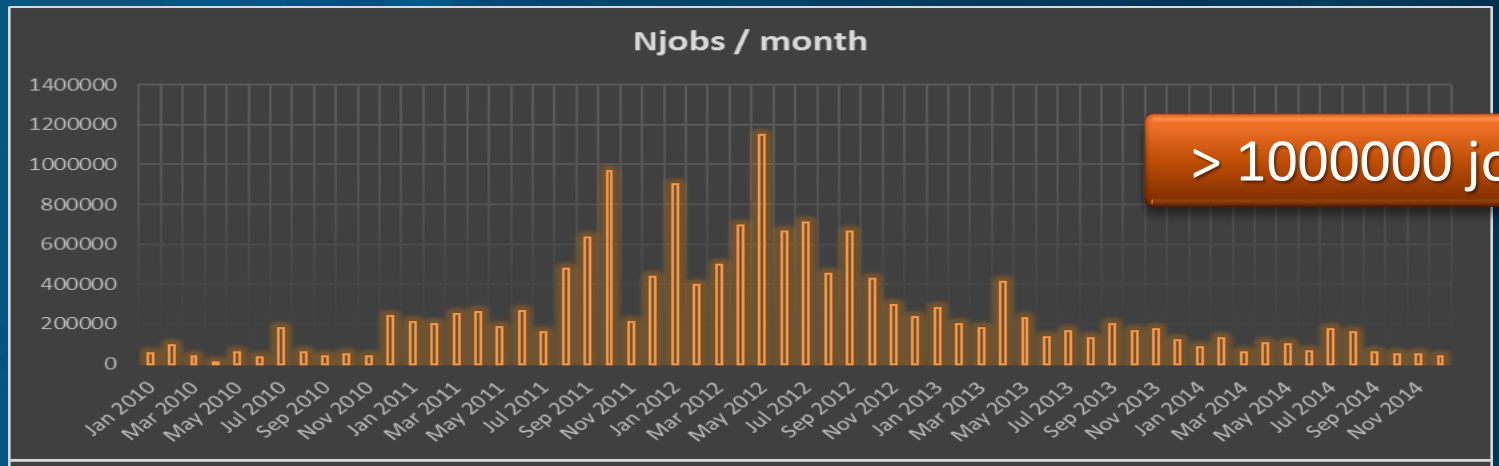


A lot of effort over several years
Continuous maintenance required

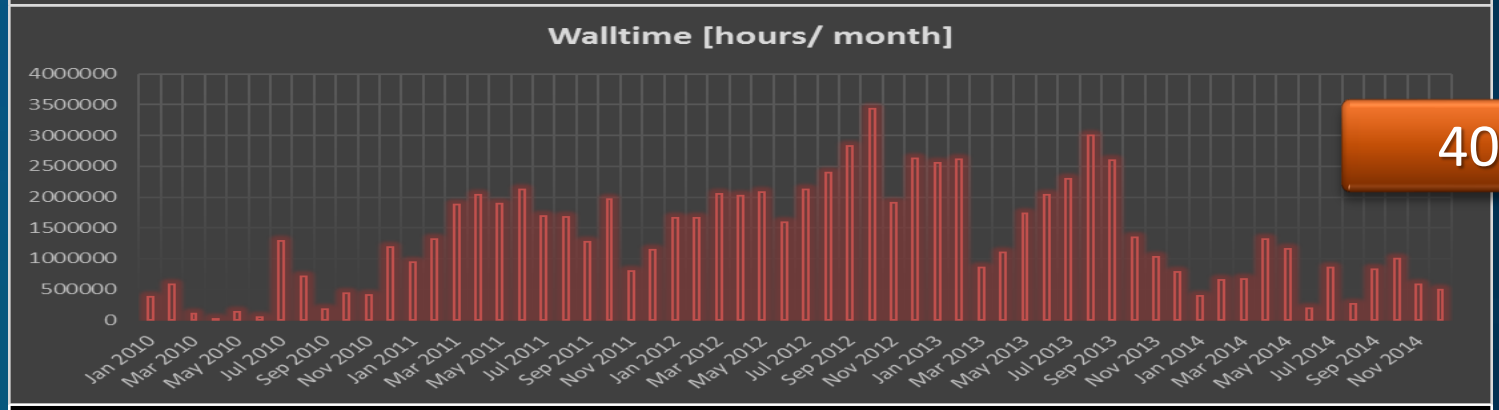
VO Management - PERUN



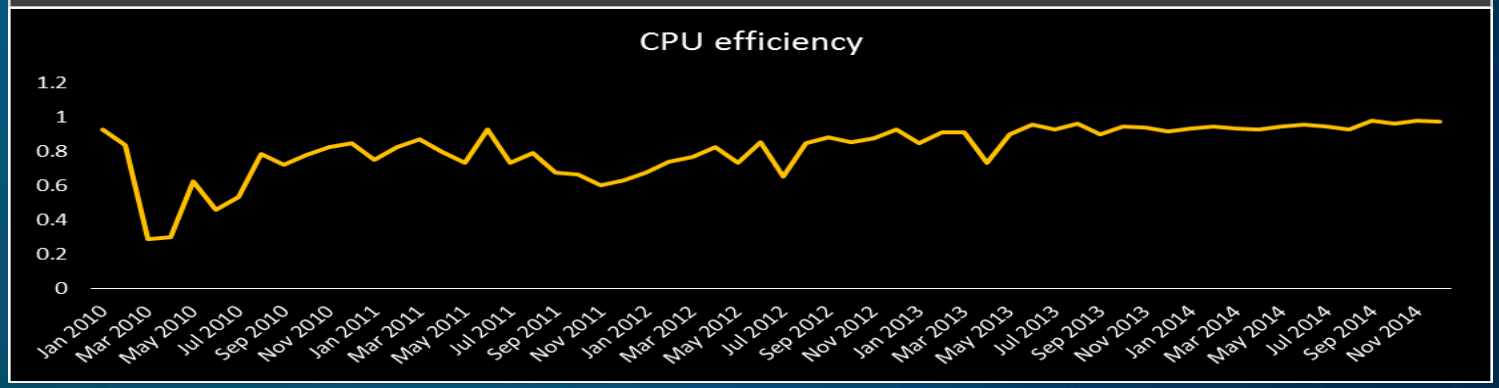
Statistics (EGI, 2010 – 2014)



> 1000000 jobs/month



4000 cores



Decision to use DIRAC

- WMS, LFC not used by LHC experiments
- DIRAC Welcomed Features
  - Pilot jobs
    - better usage of some sites
  - Monitoring and accounting
    - better control over individual users productions
  - “Free” Support
  - DIRAC instance managed by external team
  - Continuous development

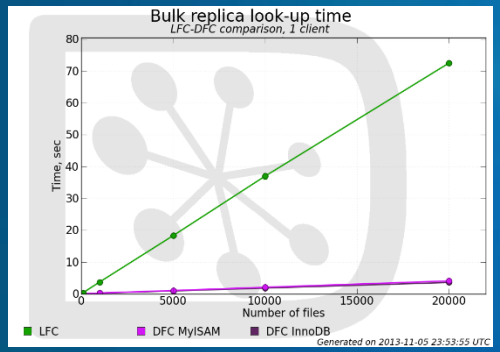
See other talks about DIRAC at CHEP15 for more DIRAC features

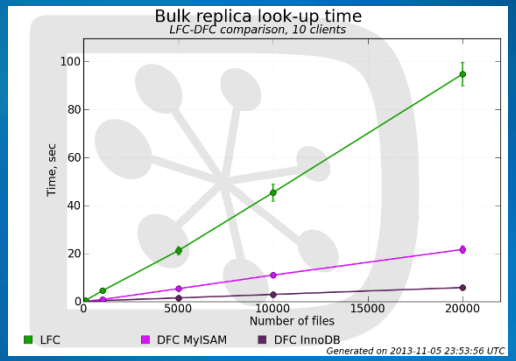
DIRAC for Bulk Production

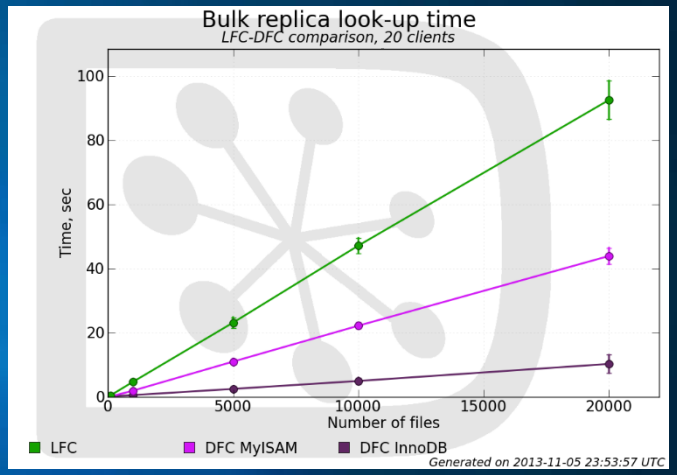
- Parametric jobs
  - Integer parameter defines line in a file with parameters (random number seed, run nr.)
  - 10000 jobs submitted in 150 s
  - Easy for standard CORSIKA production
  - Other workflows more difficult

LFC to DFC migration

- Tested on the full catalog (36 mil. files)
- Performance verified on test instances







Decision

- No bulk migration
- New production to DFC
- keep LFC running for needed period
- Use DFC as Metadata Catalog too

CVMFS

- First tests with OSG
  - but different path for EU and USA sites
- EGI CVMFS Task force lead by Catalin Condurache
- All Auger EGI sites now have /cvmfs/auger.egi.eu
  - months to complete
- Experiments with end users desktops and local clusters
  - attention to correct cache setup

Contributions to DIRAC

- Datasets implementation
  - dataset = set of files matching metaquery
  - static and dynamic datasets
  - cached informational data (#files, size, ...)
  - methods for adding, removing, status query, releasing, freezing, replication, distribution over SEs

Implementation done by M. Adam (NPI CAS, CZ)

Contributions to DIRAC

- Datasets implementation

```
FC:/auger/user/m/madam>dataset add augerTestSet primaryParticle=proton energyPeV>13500
Successfully added dataset /auger/user/m/madam/augerTestSet
FC:/auger/user/m/madam>dataset status /auger/user/m/madam/augerTestSet

augerTestSet:
=============
    Key              Value
=================================================================================
  1 NumberOfFiles    2
  2 MetaQuery        {'energyPeV': {'>': 13500}, 'primaryParticle': 'proton'}
  3 Status           Dynamic
  4 DatasetHash      B06B30385C5A253816FBB7781E208C56
  5 TotalSize        200
  6 UID              2
  7 DirID            5
  8 OwnerGroup       auger_user
  9 Owner            madam
 10 GID              2
 11 Mode             509
 12 ModificationDate 2015-04-07 15:45:35
 13 CreationDate     2015-04-07 15:45:35
 14 DatasetID        60
```

Implementation done by M. Adam (NPI CAS, CZ)

Contributions to DIRAC

- Datasets implementation

```
FC:/auger/user/m/madam>dataset files /auger/user/m/madam/augerTestSet
/auger/user/m/madam/auger-prod/proton/prod001
/auger/user/m/madam/auger-prod/proton/prod002
FC:/auger/user/m/madam>
FC:/auger/user/m/madam>dataset freeze /auger/user/m/madam/augerTestSet
Successfully frozen dataset /auger/user/m/madam/augerTestSet
FC:/auger/user/m/madam>register file /auger/user/m/madam/auger-prod/proton/prod005 test 100 FZU-USER
File successfully added to the catalog
FC:/auger/user/m/madam>meta set /auger/user/m/madam/auger-prod/proton/prod005 energyPeV 15000
/auger/user/m/madam/auger-prod/proton/prod005 energyPeV 15000
FC:/auger/user/m/madam>dataset files /auger/user/m/madam/augerTestSet
/auger/user/m/madam/auger-prod/proton/prod001
/auger/user/m/madam/auger-prod/proton/prod002
FC:/auger/user/m/madam>dataset release /auger/user/m/madam/augerTestSet
Successfully released dataset /auger/user/m/madam/augerTestSet
FC:/auger/user/m/madam>dataset files /auger/user/m/madam/augerTestSet
/auger/user/m/madam/auger-prod/proton/prod001
/auger/user/m/madam/auger-prod/proton/prod002
/auger/user/m/madam/auger-prod/proton/prod005
```

Implementation done by M. Adam (NPI CAS, CZ)

Conclusions

- DIRAC as the next framework for bulk production
- Adding support for individual users
- CVMFS simplifies Auger Offline installation
- Improving Data Management

Acknowledgement

- A. Tsaregorodtsev, DIRAC Technical Coordinator
- M. Adam – datasets implementation
- Members of Auger Distributed Computing Team
  - G. Isar, G. Rubio, A. Sevcenco