Union of Light Ion Centres in Europe

Shared services and tools based on the Ganga job definition and management framework.

EGI User Forum 2011 – Vilnius, Lithuania.

U. Egede (Imperial College London), **M. Kenyon** (CERN), J. Moscicki (CERN), I.A. Dzhunov (University of Sofia), L. Kokoszkiewicz (CERN), M. Cinquilli (CERN), L. Sargsyan (CERN/Yerevan Physics Institute), E. Karavakis (CERN), J. Andreeva (CERN)

The ULICE project is co-funded by the European Commission under FP7
Grant Agreement Number 228436.

Overview

- The EGI Introductory Package.
  - Tools that were initially developed by and for specific user communities.
- Tool adoption case-studies.
  - Grid job submission & management.
  - Error-reporting infrastructure for users.

The ULICE project is co-funded by the European Commission under FP7
Grant Agreement Number 228436.

The EGI Introductory Package

- "Simple, complete solution for running and monitoring computing tasks on the grid."
  - Audience: small to medium-sized user communities.
- Comprises the following components.
  - Ganga: user interface for job submission and management.
  - DIANE: automatic control and scheduling of Ganga jobs.
  - Mini-dashboard: monitoring of Ganga and/or Diane tasks & error reporting.

The ULICE project is co-funded by the European Commission under FP7
Grant Agreement Number 228436.

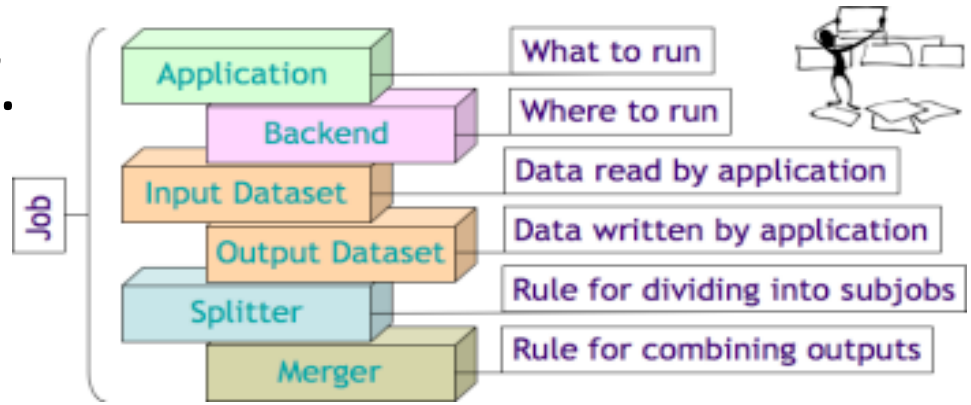
Ganga overview

- Ganga project was started by ATLAS & LHCb in 2001.
  - Entirely written in Python from the outset.
  - Extensible framework; simple and well documented procedure.
- Underwent a major redesign in 2005.
  - Core codebase is stable, though new features are in active development.
  - Assign ownership of sub-packages.
- Release procedure.
  - Rotating release manager.
  - Rigorous testing prior to each release.
- Registered in the EGI application database.

The ULICE project is co-funded by the European Commission under FP7
Grant Agreement Number 228436.

Ganga overview

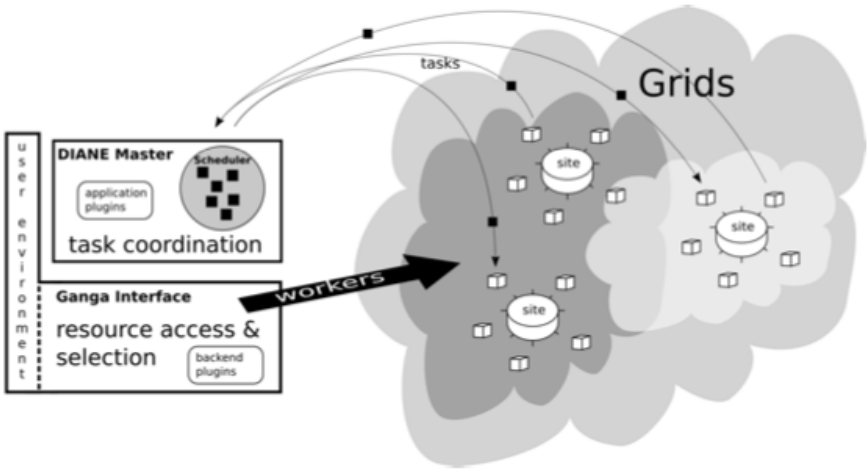
- Easy to use frontend for the configuration, execution and management of computational tasks.
  - Interactive Python prompt, script submission, GUI.
- Submission to a range of platforms in a consistent manner (localhost, PBS, LSF, SGE, Condor, gLite, ARC, Globus).
- Key concept: “The Job”.



The ULICE project is co-funded by the European Commission under FP7
Grant Agreement Number 228436.

DIANE – Distributed Analysis Environment

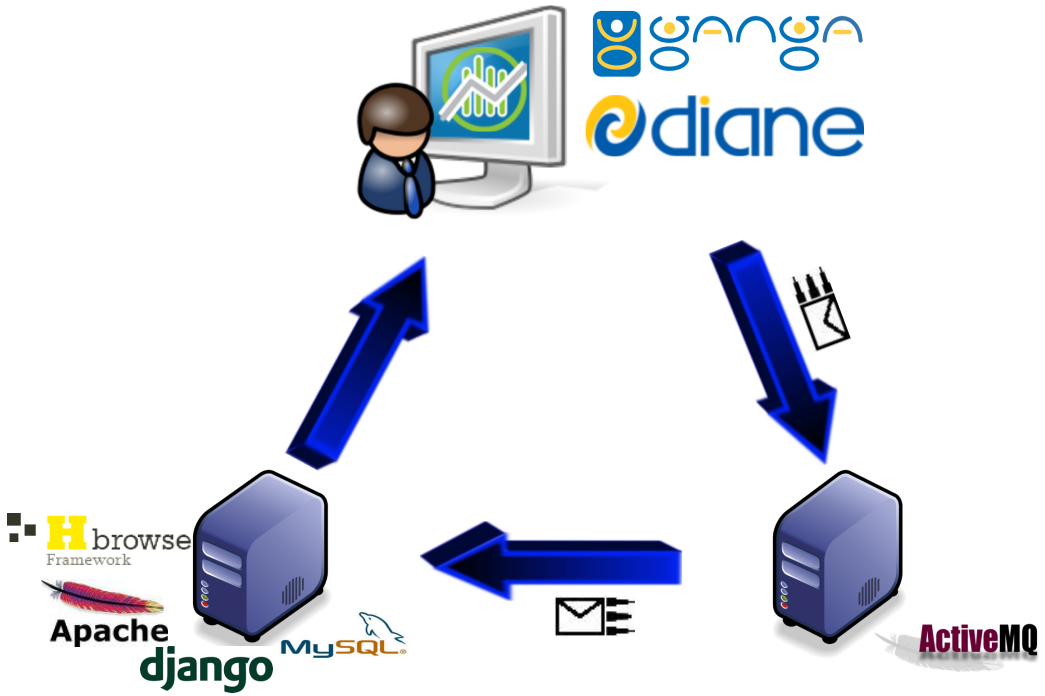
- DIANE development started within CERN IT in 2000.
  - Registered in the EGI application database.
- Job execution control framework.
  - Automatic load balancing, scheduling and failure recovery.
- Utilises a pilot-job mechanism (also known as “late binding”).
  - Pilots controlled by Ganga.
  - Workload fed directly to pilots.



The ULICE project is co-funded by the European Commission under FP7
Grant Agreement Number 228436.

Mini-dashboard: task monitoring

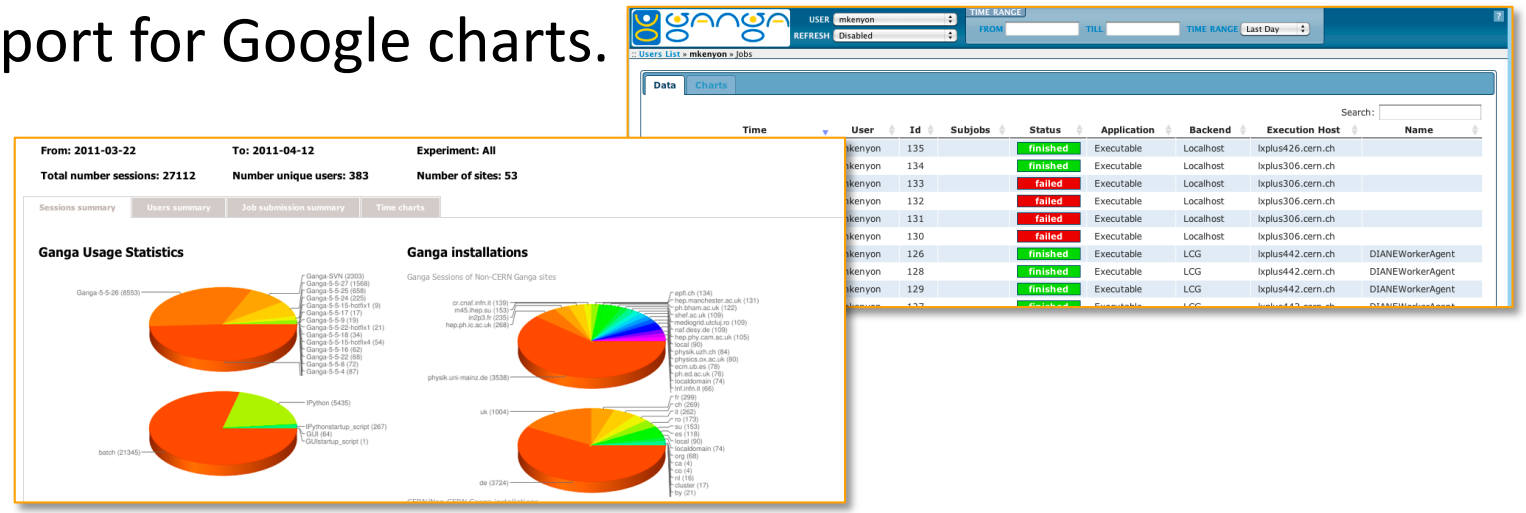
- A service to monitor the state of submitted Ganga and/or DIANE tasks.



The ULICE project is co-funded by the European Commission under FP7
Grant Agreement Number 228436.

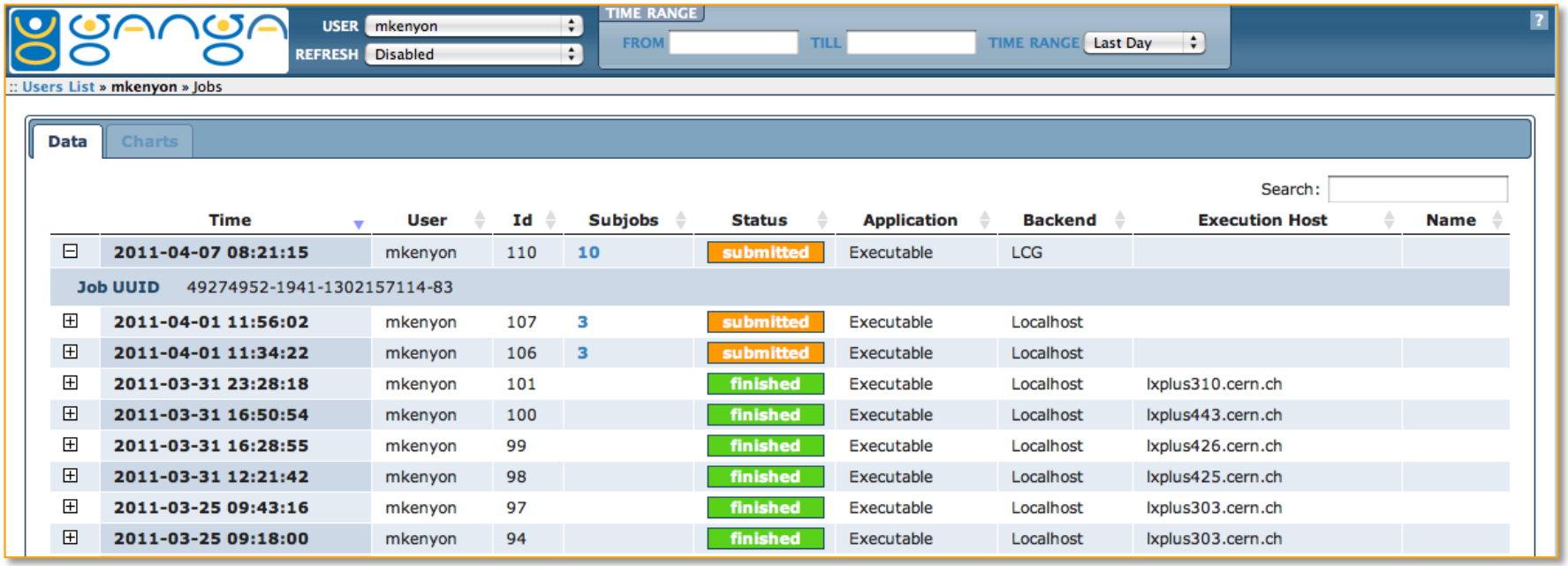
Mini-dashboard: task monitoring

- Based around the hBrowse framework.
  - HTML/Javascript client that consumes JSON data.
  - Highly configurable.
  - Plugin architecture (dynamic tables, user selection, field filters).
  - Support for Google charts.



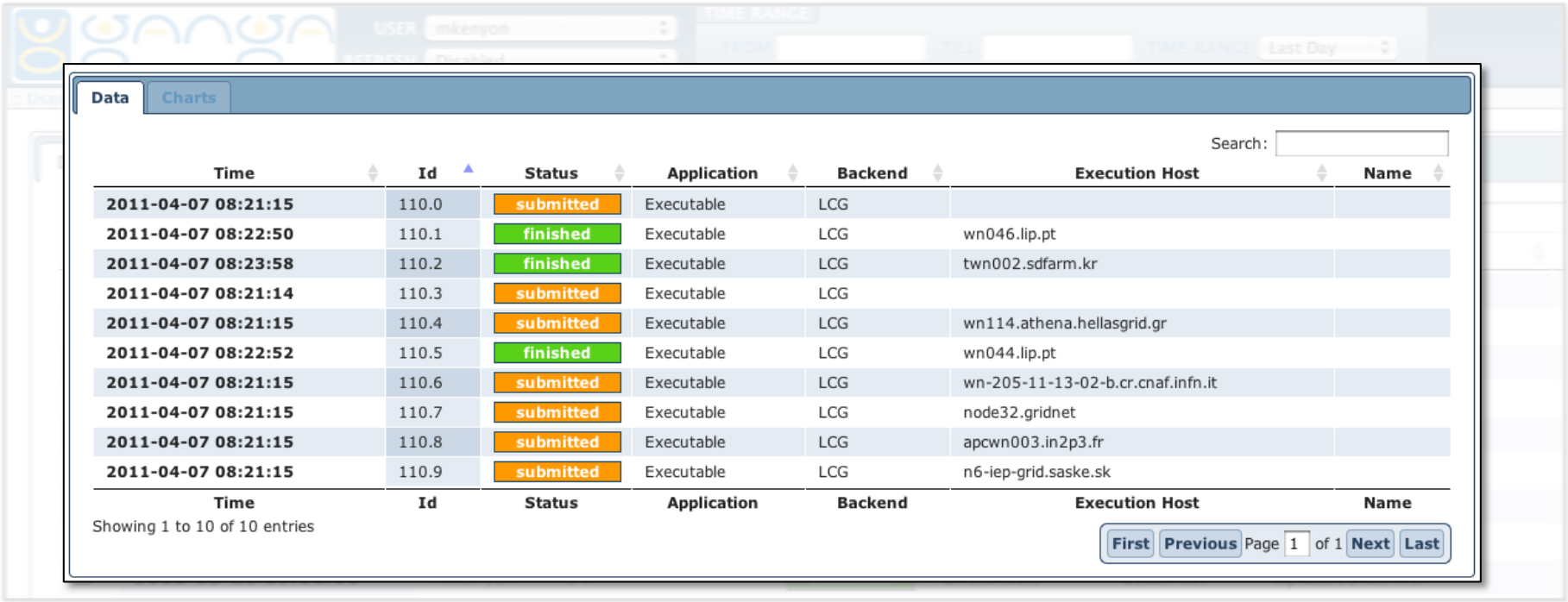
The ULICE project is co-funded by the European Commission under FP7
Grant Agreement Number 228436.

Mini-dashboard: task monitoring



The ULICE project is co-funded by the European Commission under FP7
Grant Agreement Number 228436.

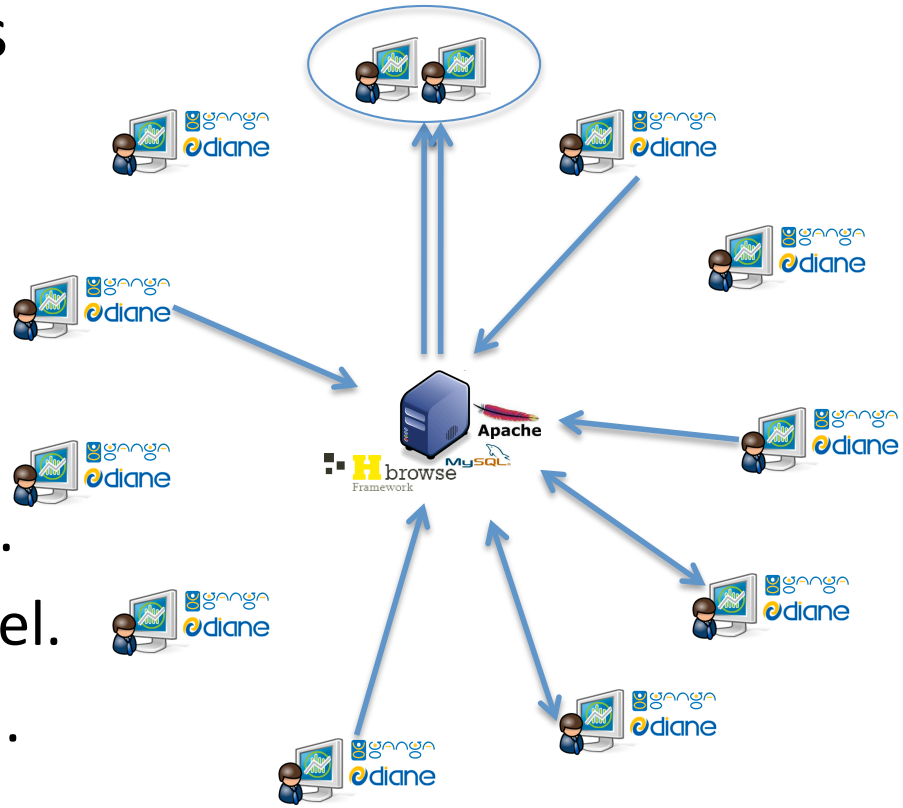
Mini-dashboard: task monitoring



The ULICE project is co-funded by the European Commission under FP7
Grant Agreement Number 228436.

Mini-dashboard: error reporting

- Support teams need the full picture to help users.
- Grid jobs are complex beasts
  - Range of software versions.
  - Configuration settings.
  - Environment variables.
- Error reporting tool.
  - Repository for diagnostic data.
  - Traditional “expert-user” model.
  - Or, community support model.



The ULICE project is co-funded by the European Commission under FP7
Grant Agreement Number 228436.

Mini-dashboard: error-reporting tool

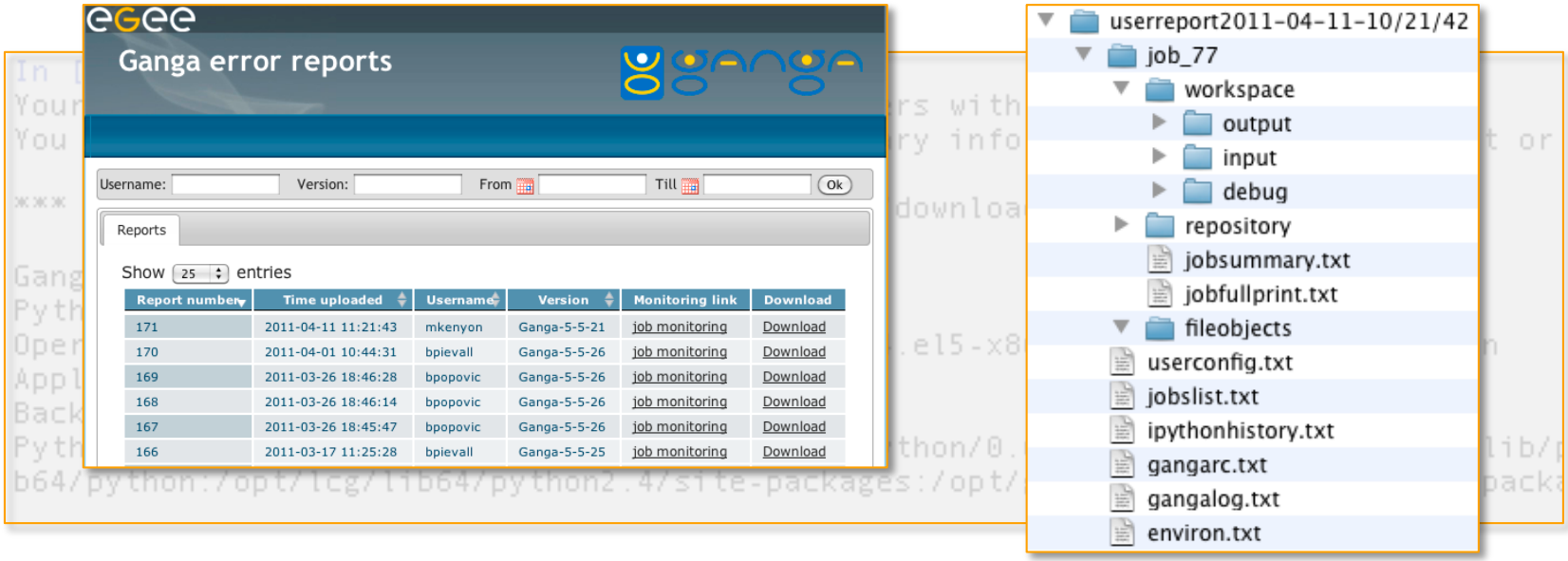
- Ganga function to POST a job error report to a central repository.

```
In [2]:report(jobs(77))
Your error report was uploaded to ganga developers with the following URL.
You may include this URL and the following summary information in your bug report or

*** http://gangamon.cern.ch/django/errorreports/download/171 ***

Ganga Version : Ganga-5-5-21
Python Version : 2.4.3
Operation System Version : Linux-2.6.18-194.17.4.el5-x86_64-with-redhat-5.5-Boron
Application Name : Executable
Backend Name : Localhost
Python Path : /cluster/local/Ganga5/external/ipython/0.6.13_ganga_patch1/noarch/lib/p
b64/python:/opt/lcg/lib64/python2.4/site-packages:/opt/glite/lib/python2.3/site-packa
```


The ULICE project is co-funded by the European Commission under FP7
Grant Agreement Number 228436.

Mini-dashboard: error-reporting tool

- Ganga function to POST a job error report to a central repository.
- Report available to community support team.
- Command history, detailed logs, environment...



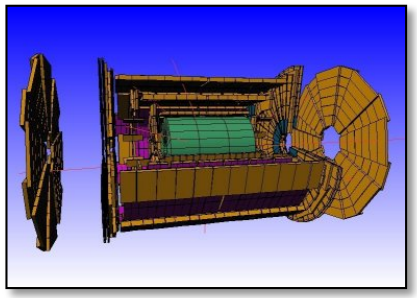
The ULICE project is co-funded by the European Commission under FP7
Grant Agreement Number 228436.

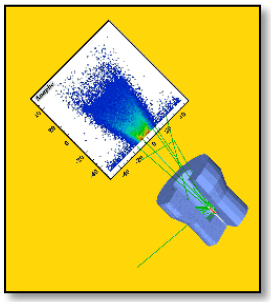
Case study 1
Geant4 code validation.

The ULICE project is co-funded by the European Commission under FP7
Grant Agreement Number 228436.

Community case study: Geant4

- Geant4 is a toolkit for simulating the trajectories of particles through matter.
- Finds application within HEP, nuclear and accelerator physics, medical and space science.
- Vast, object-oriented suite, with over 600,000 lines of complex source code.
  - Developers are distributed around the world.





The ULICE project is co-funded by the European Commission under FP7
Grant Agreement Number 228436.

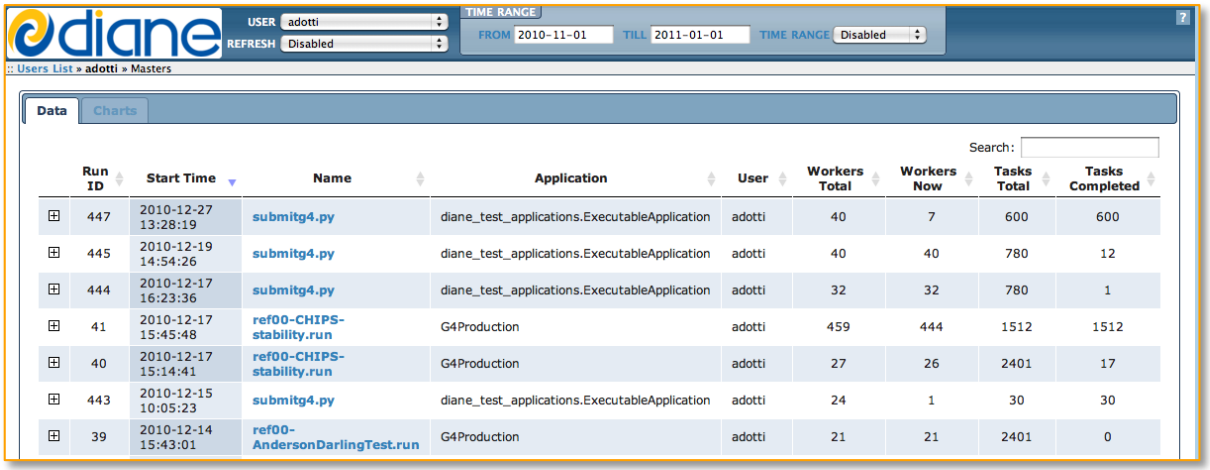
Geant4: Code validation

- Statistical regression tests compare simulation results between the previous and pending releases.
  - Wide range of (physics) input parameters are used.
  - 1000 independent tasks generated, each one simulating 5000 events.
  - Total CPU time required per release = a few CPU years.
- Intensive validation period prior to release (every 6 months).
  - Quick succession of candidate releases.
  - Tests partially or wholly repeated multiple times.
- The Ganga/DIANE framework has been used to test releases since June 2007.
- Validation metrics; performance of code (time/event) and stability (application crashes).

The ULICE project is co-funded by the European Commission under FP7
Grant Agreement Number 228436.

Geant4: Code validation

- Last validation Dec. 2010; Geant4 v9.4 (report)
  - Ganga/DIANE used to run code validation on 18 grid sites.
  - 80 million events produced in 2-3 days (c.f. 1 week for previous releases).
  - Exposed rare Geant4 software crashes (1 crash per 10,000 generated events).



	Run ID	Start Time	Name	Application	User	Workers Total	Workers Now	Tasks Total	Tasks Completed
⊞	447	2010-12-27 13:28:19	submitg4.py	diane_test_applications.ExecutableApplication	adotti	40	7	600	600
⊞	445	2010-12-19 14:54:26	submitg4.py	diane_test_applications.ExecutableApplication	adotti	40	40	780	12
⊞	444	2010-12-17 16:23:36	submitg4.py	diane_test_applications.ExecutableApplication	adotti	32	32	780	1
⊞	41	2010-12-17 15:45:48	ref00-CHIPS-stability.run	G4Production	adotti	459	444	1512	1512
⊞	40	2010-12-17 15:14:41	ref00-CHIPS-stability.run	G4Production	adotti	27	26	2401	17
⊞	443	2010-12-15 10:05:23	submitg4.py	diane_test_applications.ExecutableApplication	adotti	24	1	30	30
⊞	39	2010-12-14 15:43:01	ref00-AndersonDarlingTest.run	G4Production	adotti	21	21	2401	0

The ULICE project is co-funded by the European Commission under FP7
Grant Agreement Number 228436.

Geant4: Value added by EGI IP

- Geant4 validation team observations.
  - New DIANE/Ganga infrastructure [...] allows for robust and reliable logging and monitoring of Ganga jobs.
  - DIANE's error detection allows the automatic exclusion of mis-configured nodes ("the main problem in previous validation periods").
  - Grid experience substantially improved since last campaign; increase of 10x in number of jobs executed for the same allocated resources.
- "GRID usage has improved substantially... mainly due to the improved software and to the increased ability to monitor and debug jobs."

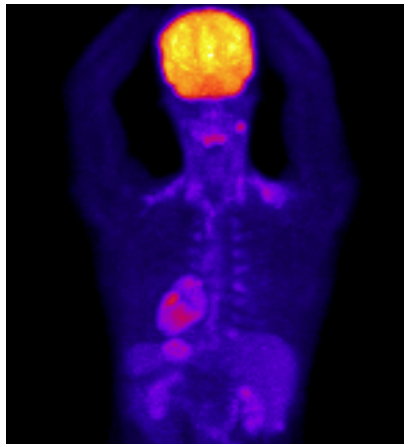
The ULICE project is co-funded by the European Commission under FP7
Grant Agreement Number 228436.

Case study 2
OpenGate Virtual Laboratory Environment

The ULICE project is co-funded by the European Commission under FP7
Grant Agreement Number 228436.

Community case study: OpenGate

- GATE software; nuclear medicine simulations for imaging.
  - Based on Geant4 toolkit.
  - Simulate particle tracks through matter.
    - Initial set of properties (type, location, direction).
    - Large number of particles (simulations).
    - Typical simulations take hours to days.

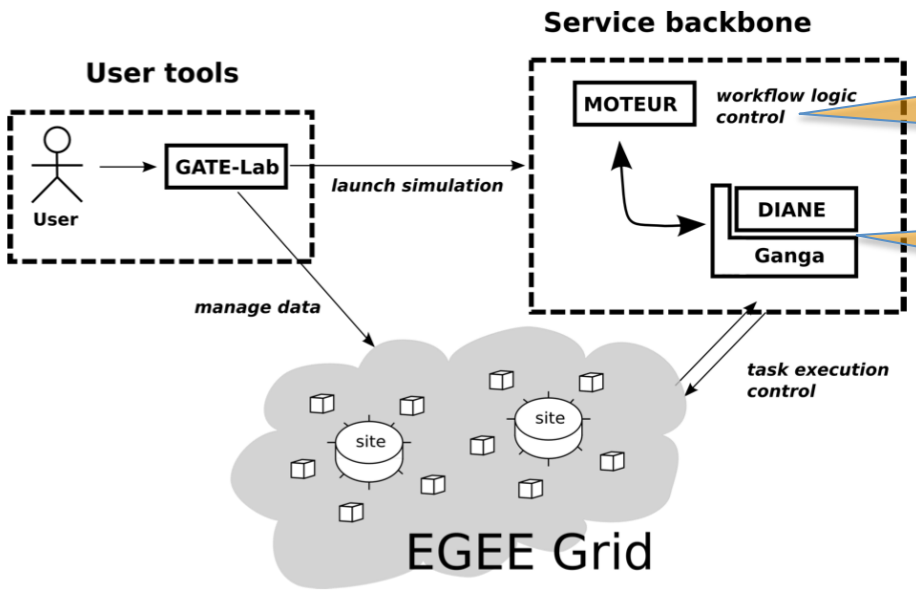


- Tasks well suited to being parallelized.
  - Split simulation into sub-simulations (static partitioning).
  - But; all MC sub-tasks must complete to get the final result.

The ULICE project is co-funded by the European Commission under FP7
Grant Agreement Number 228436.

Community case study: OpenGate

- OpenGate collaboration developed a dynamic splitting method for GATE.
  - Camarasu-Pop et al. J Grid Computing (2010) 8:241–259



Monitors no. of simulated events; generates new tasks if required; passes them to DIANE.

Task processing engine; pilot jobs on grid nodes; STDOUT & ERR returned periodically.

The ULICE project is co-funded by the European Commission under FP7
Grant Agreement Number 228436.

Community case study: OpenGate

- N tasks are dispatched by DIANE to the Grid.
  - Each task will continue to run until the desired number of particles (summed across all tasks) is reached.
- On average, a grid node will simulate a fraction (1/N) of the total number of particles.
- DIANE performs a periodic check of total number of particles simulated.
- Run is terminated when all particles have been simulated.

The ULICE project is co-funded by the European Commission under FP7
Grant Agreement Number 228436.

OpenGate: Value added by EGI IP

- Makespan reduced; for simulating $20x10^6$ events.
  - 8.5 hours on dual-core PC (2008).
  - ~24 hours on the grid with 100 classical jobs. 78% success rate.
  - 1.75 hours on the grid with 100 DIANE worker agents. 100% of simulation complete.
- In period July 2009 – Aug 2010;
  - 360 DIANE RunMaster instances were activated in the backbone, handling 58,000 worker agent jobs.
- Generic solution; any application managed by MOTEUR can now interface with DIANE.
- Solution has entered production environment for non-clinical radiation therapy researchers at Creatis.

The ULICE project is co-funded by the European Commission under FP7
Grant Agreement Number 228436.

Case study 3
CMS Error Reporting Tool.

The ULICE project is co-funded by the European Commission under FP7
Grant Agreement Number 228436.

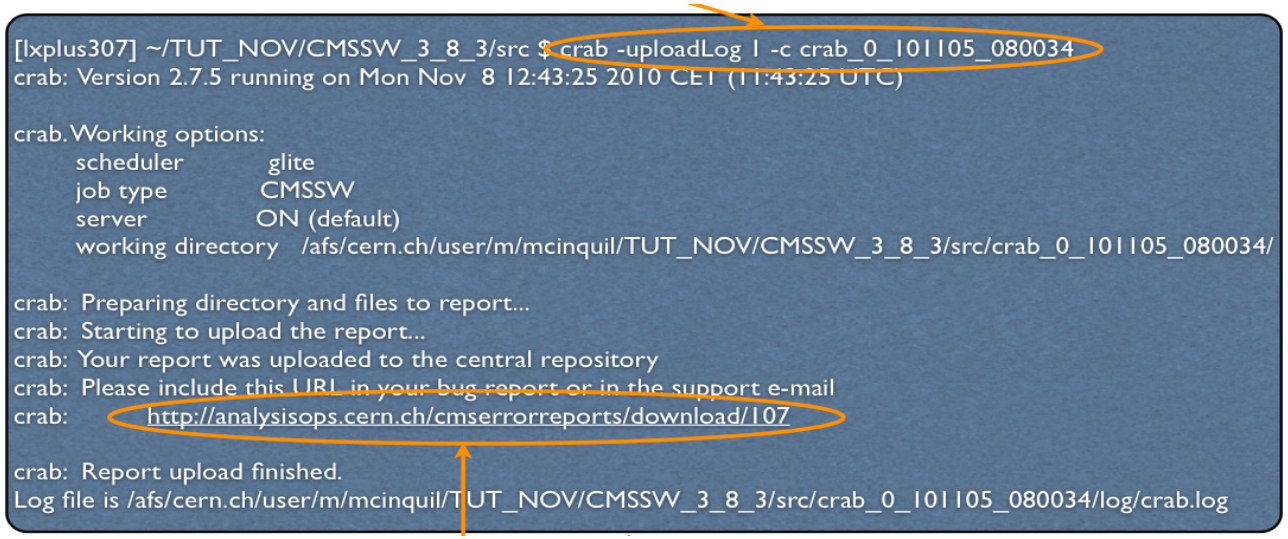
CMS error reporting tool

- Users perform analysis via CMS Remote Analysis Builder (CRAB).
- Jobs running on heterogeneous services.
  - CMS services/middleware/batch systems.
  - Differing use cases.
- Dedicated analysis operation debugging team.
  - Proactively find and debug problems.
  - Handle support requests from CMS user community mailing list.
  - 3,500 users from 40 countries across many time zones.

The ULICE project is co-funded by the European Commission under FP7
Grant Agreement Number 228436.

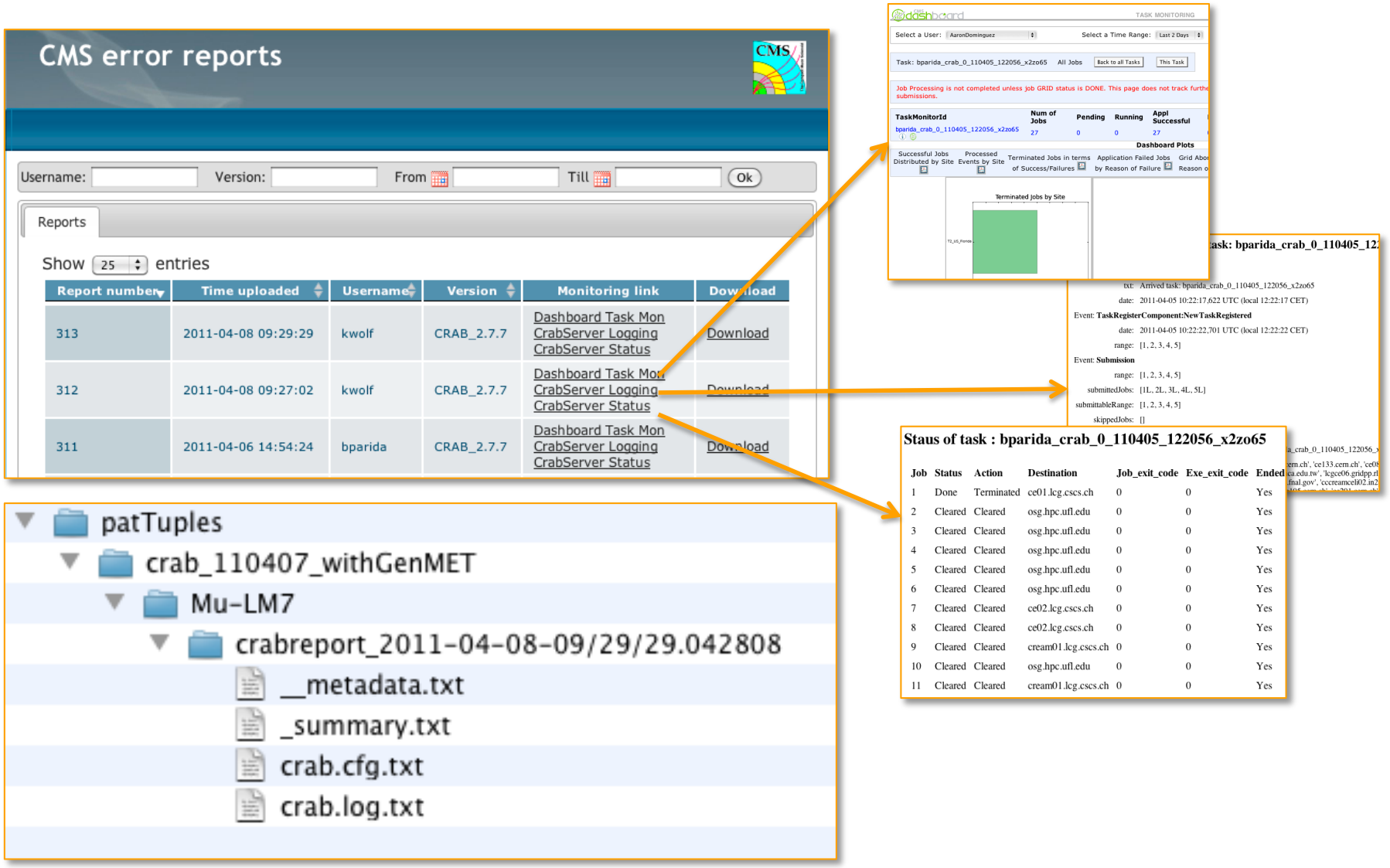
CMS error reporting tool

- In order to optimise the support procedure, CMS adopted the error-reporting tool.
- A CRAB plugin was developed that uploads debugging information to a central repository.



The ULICE project is co-funded by the European Commission under FP7
Grant Agreement Number 228436.

CMS error reporting tool



The ULICE project is co-funded by the European Commission under FP7
Grant Agreement Number 228436.

CMS error-reporting: early feedback

- Positive; the system works well, and is considered very useful.
- Streamlines a user-support request.
  - Reduces email traffic to support lists.
- Has allowed CMS to centralise and formalise their user-support mechanism.

The ULICE project is co-funded by the European Commission under FP7
Grant Agreement Number 228436.

Summary

- EGI Introductory Package has been adopted and adapted by a wide range of user communities.
- Based around mature, stable and well documented tools.
- Startup overhead is surprisingly low (see demo later today).
- DIANE allows researchers to use resources more efficiently than direct job submission alone.
- Mini-dashboard provides customisable tools for monitoring jobs and optimising user-support.

The ULICE project is co-funded by the European Commission under FP7
Grant Agreement Number 228436.

Further information

- EGI introductory package
  - https://twiki.cern.ch/twiki/bin/view/ArdaGrid/EGIIntroductoryPackage
- Ganga and DIANE are registered in the EGI application database, and are part of the EGEE Respect tool suite.
- Demonstration: Lambda room 4pm today.

The ULICE project is co-funded by the European Commission under FP7
Grant Agreement Number 228436.