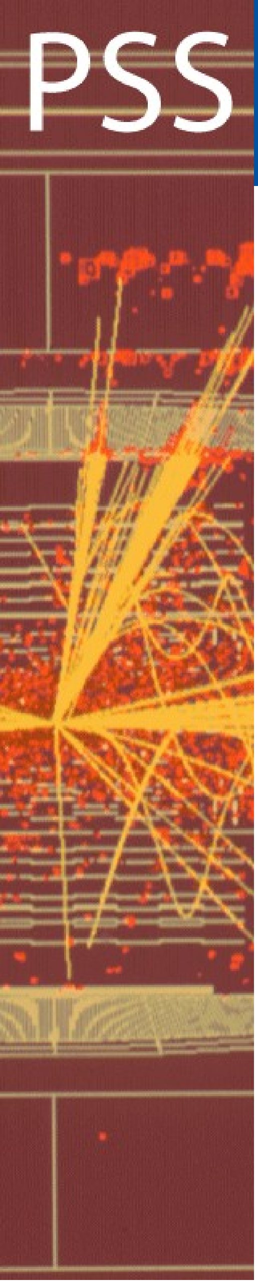


Physics Services Support

CERN IT Department

# The Ganga job management system

*Jakub T. Mościcki, CERN IT*

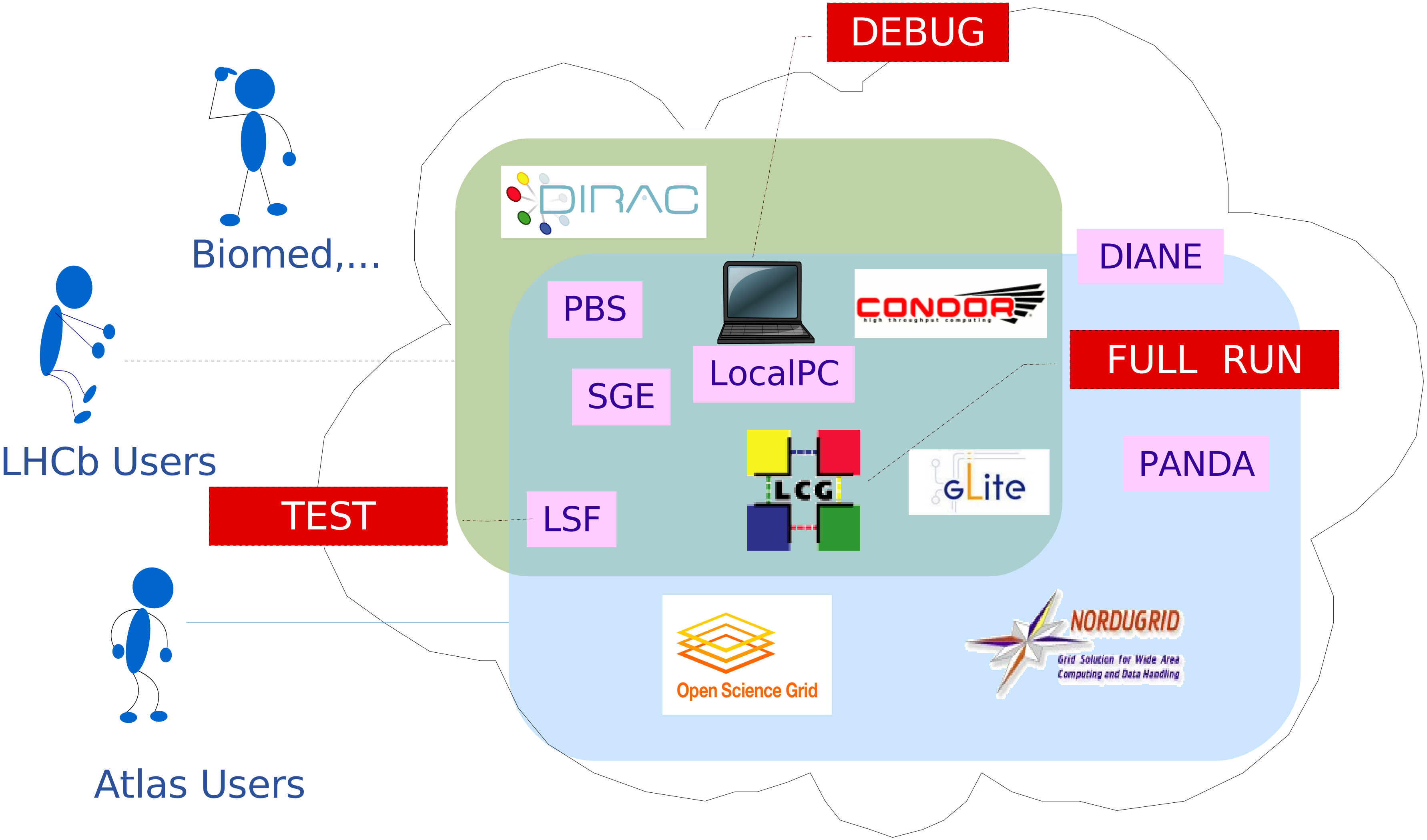
**http://cern.ch/ganga**

# The Vision

- **Help the “Physicist” to do his/her analysis work**
  - Help configuring and submitting analysis jobs (Job Wizard)
  - Help users to keep track of what they have done
  - Hide completely all technicalities
  - Provide a palette of possible choices and specialized plug-ins:
    - pre-defined application configurations
    - batch/grid systems, etc.
    - GANGA “knows” what the user is running, so it can really help him/her.
  - Single “desktop” for a variety of tasks
  - Friendly user interface is essential
    - do not impose a single working style, adapt to the users!

# Motivation

- **In practice users deal with multiple computing backends**



# Who does Ganga

- Ganga is an ATLAS/LHCb joint project

-Support for development work

-Core team:

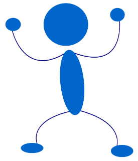
- F.Brochu (Cambridge), U.Egede (Imperial), J. Elmsheuser (Munich), K.Harrison (Cambridge), H.C.Lee (ASGC Taipei), D.Liko (CERN), A.Maier (CERN), J.T.Moscicki (CERN), A.Muraru (Bucharest), W.Reece (Imperial), A.Soroko (Oxford), CL.Tan (Birmingham)

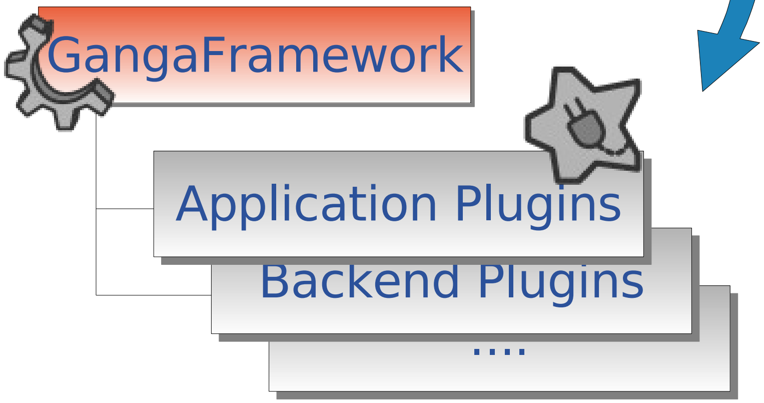
Ganga is GPL, has been commercially used
so recently we had to put the license

# Lightweight tool

- **Ganga: a lightweight tool in python**
  - a utility which you download to your computer
    - or it is already installed in your institute in a shared area
      - *for example: /afs/cern.ch/sw/ganga/install/4.3.0*
    - it is an **add-on** to installed software
  - comes with a set of plugins for **Atlas** and **LHCb**
    - **open** - other applications and backend may be easily added
      - *even by users*

Ganga
Application Software
LSF Client
LCG UI





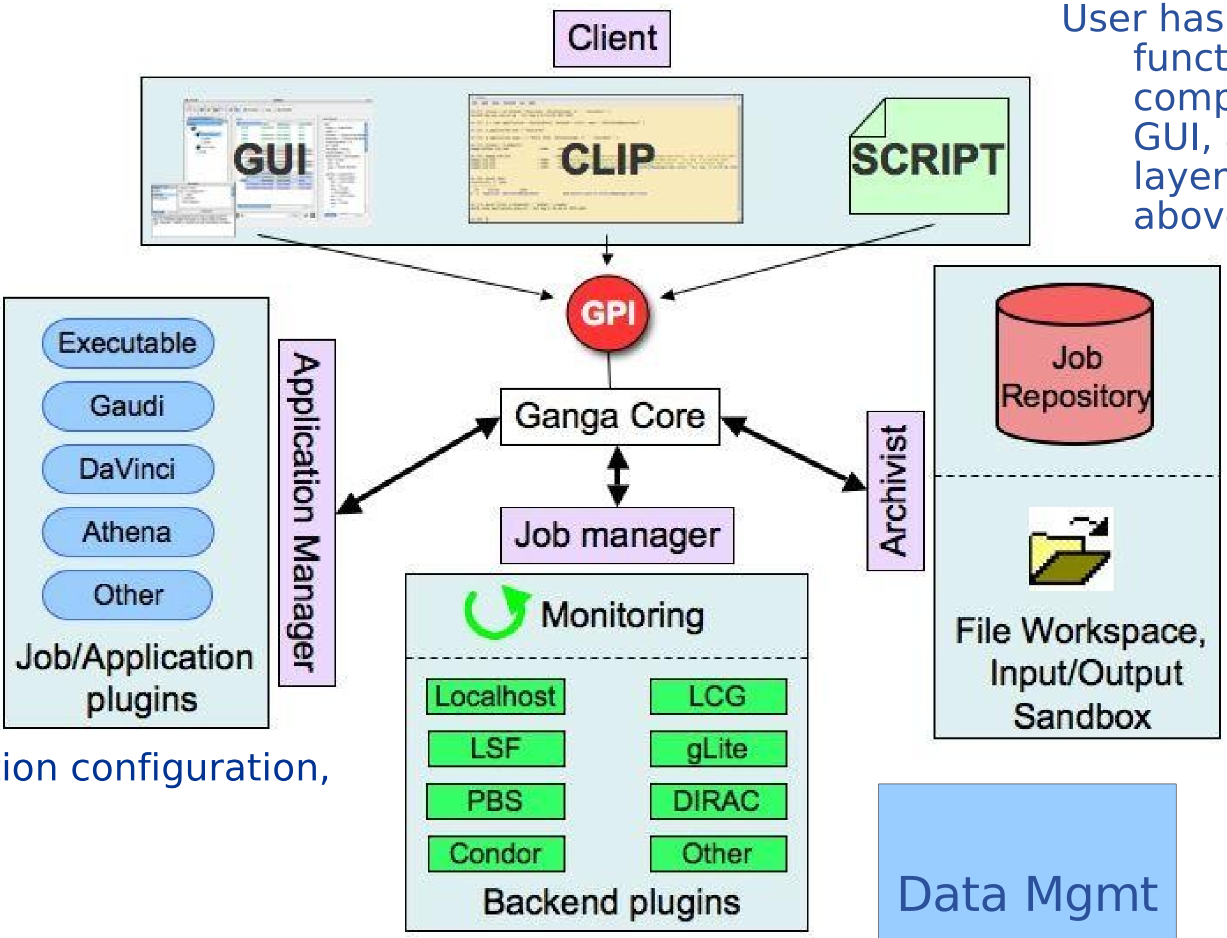
GangaFramework

Application Plugins

Backend Plugins

....

# Architecture



User has access to functionality of GANGA components both through GUI, and through CLI, layered one over the other above the software bus

storage and recovery of job information in a local or remote DB storage of sandbox files

application configuration, ...

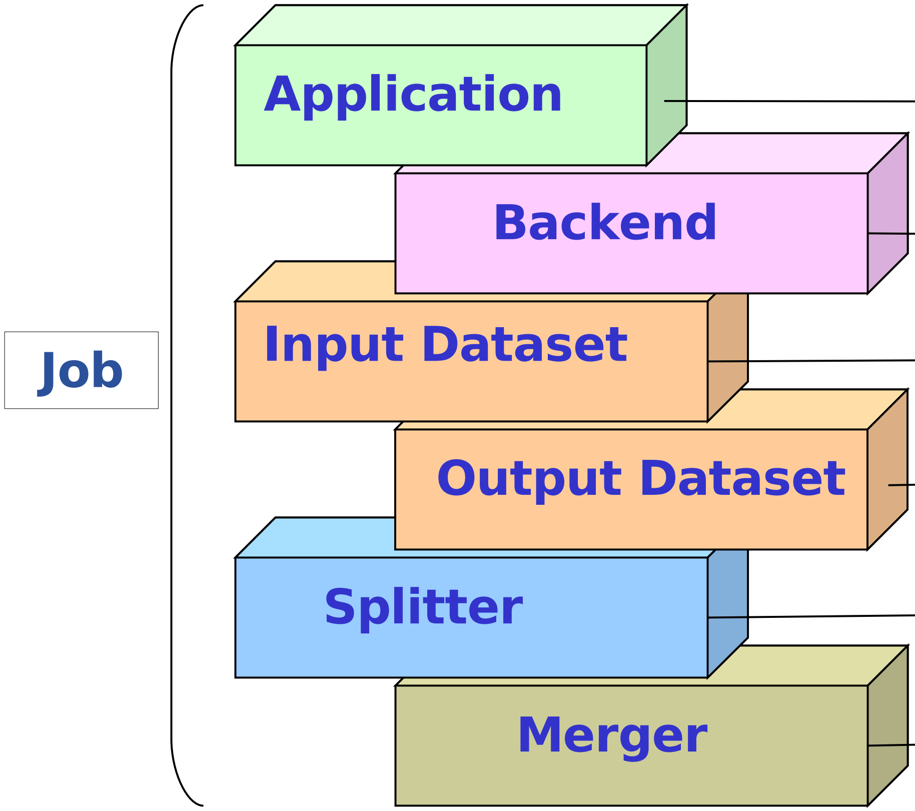
job submission: selection between Grid and local system

Data Mgmt

input data selection, output location specification

# Job Abstraction

- **simple but powerful job abstraction**



What to run

Where to run

Data read by application

Data written by application

Rule for dividing into subjobs

Rule for combining outputs

# Ganga Interface

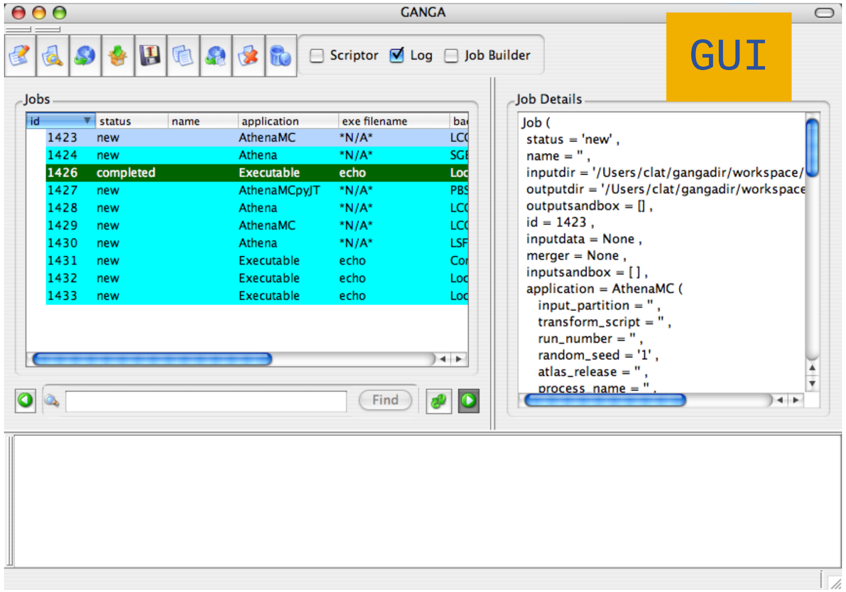
- **Python-based interface**
- **3 different views**

CLIP

```
*** Welcome to Ganga ***
Version: Ganga-4-2-8
Documentation and support: http://cern.ch/ganga
Type help() or help('index') for online help.

In [1]: jobs
Out[1]: Statistics: 1  jobs
--------------
#   id      status         name    subjobs       application
#    1   completed                                Executable
In [2]: 
```

GUI



| id | status | name | application | exe filename | ba |
|---|---|---|---|---|---|
| 1423 | new | | AthenaMC | *N/A* | LC |
| 1424 | new | | Athena | *N/A* | SG |
| 1426 | completed | | Executable | echo | Lo |
| 1427 | new | | AthenaMCpyJT | *N/A* | PB |
| 1428 | new | | Athena | *N/A* | LC |
| 1429 | new | | AthenaMC | *N/A* | LC |
| 1430 | new | | Athena | *N/A* | LS |
| 1431 | new | | Executable | echo | Co |
| 1432 | new | | Executable | echo | Lo |
| 1433 | new | | Executable | echo | Lo |

```
Job (
 status = 'new' ,
 name = '' ,
 inputdir = '/Users/clat/gangadir/workspace/
 outputdir = '/Users/clat/gangadir/workspace
 outputsandbox = [] ,
 id = 1423 ,
 inputdata = None ,
 merger = None ,
 inputsandbox = [ ] ,
 application = AthenaMC (
   input_partition = '' ,
   transform_script = '' ,
   run_number = '' ,
   random_seed = '1' ,
   atlas_release = '' ,
   process_name = '' ,
```

Scripting (GPI)

```
#!/usr/bin/env ganga
#-*-python-*-
import time
j = Job()
j.backend = LCG()
j.submit()
while not j.status in [‘completed’,’failed’]:
    print(‘job still running’)
    time.sleep(30)
```

```
> ./myjob.exec

> ganga ./myjob.exec

In [1]:execfile(“myjob.exec”)
```

# Job Submission

- **shell command**

```
> ganga athena --inDS csc11.005300.PythiaH130zz4l.recon.AOD.v11004103 \
               --outputdata AnalysisSkeleton.aan.root \
               --lsf --maxevt 100 \
               AnalysisSkeleton_jobOptions.py
```

quick

- **interactive command line in python**

```
In[1]: j = Job()
In[2]: j.application=Athena()
In[3]: j.application.prepare()
In[4]: j.application.option_file='AnalysisSkeleton_jobOptions.py'
In[5]: j.inputdata=DQ2Dataset()
In[6]: j.inputdata.type='DQ2_LOCAL'
In[7]: j.inputdata.dataset="csc11.005300.PythiaH130zz4l.recon.AOD.v11004103"
In[8]: j.outputdata=DQ2OutputDataset()
In[9]: j.outputdata.outputdata=['AnalysisSkeleton.aan.root']
In[10]:j.outputdata.location = '$HOME/athena/output'
In[11]:j.backend=LSF()
In[12]:j.submit()

In[13]: j2 = j.copy()
In[14]: j2.backend=LCG()
In[15]: j2.submit()
```

**mix**

flexible

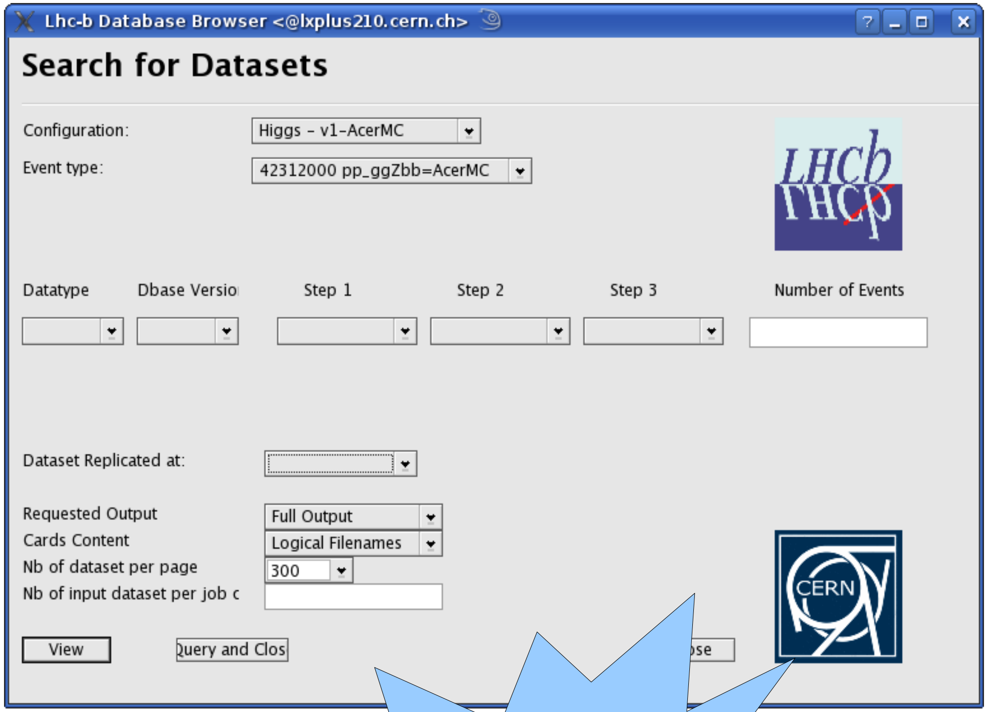
# Application integration

- **Additional application-specific services**

```
In[1]: d = browseBK()

Out[1]: LHCbDataset (
 files = [ ... ]
 )

In[2]: j.inputdata = d
```



Lhc-b Database Browser <@lxplus210.cern.ch>

**Search for Datasets**

Configuration: Higgs – v1-AcerMC

Event type: 42312000 pp_ggZbb=AcerMC

Datatype | Dbase Versio | Step 1 | Step 2 | Step 3 | Number of Events

Dataset Replicated at:

Requested Output: Full Output

Cards Content: Logical Filenames

Nb of dataset per page: 300

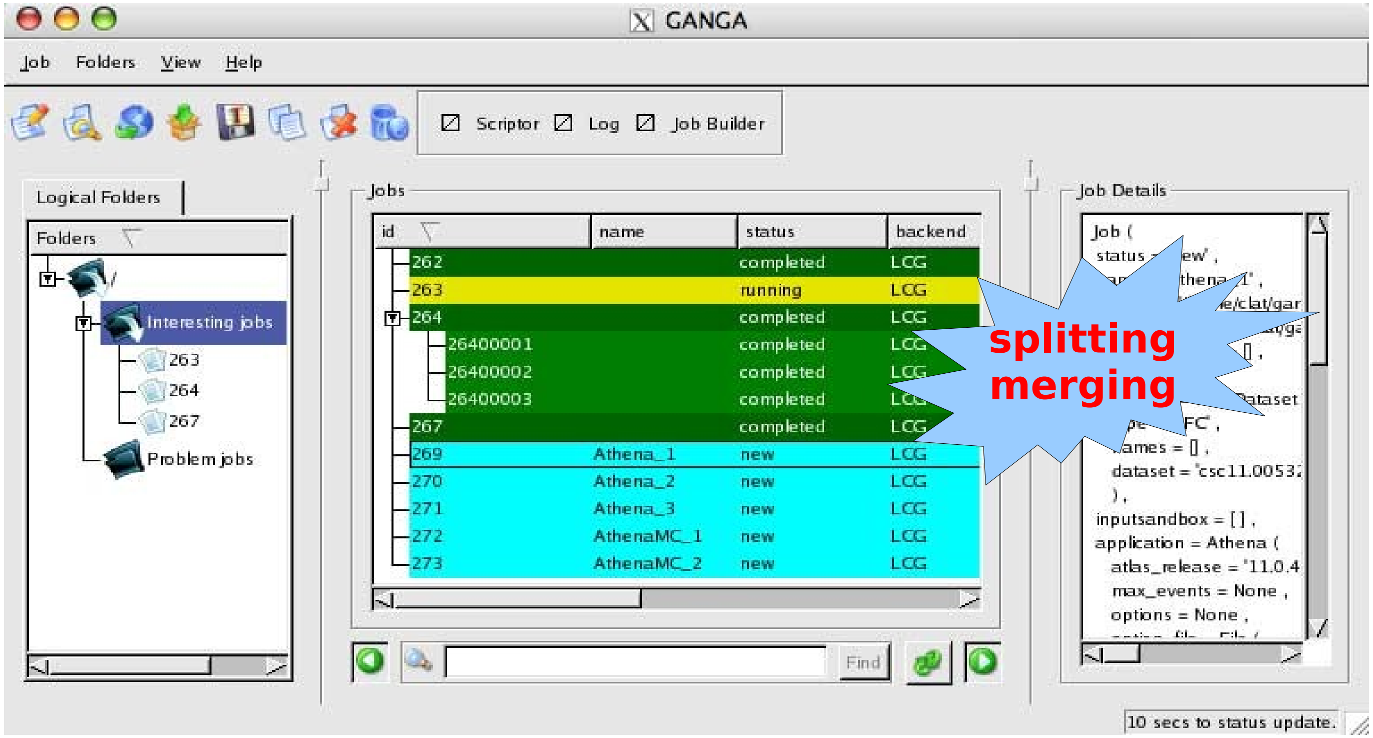
Nb of input dataset per job c

**mix CLIP/GUI**

**query data from the same tool**

# Job History

- **access previously created jobs**
  - peek into output and job configuration
  - reuse parts of job configuration
  - save commonly used configurations as job templates



| id | name | status | backend |
|---|---|---|---|
| 262 | | completed | LCG |
| 263 | | running | LCG |
| 264 | | completed | LCG |
| 26400001 | | completed | LCG |
| 26400002 | | completed | LCG |
| 26400003 | | completed | LCG |
| 267 | | completed | LCG |
| 269 | Athena_1 | new | LCG |
| 270 | Athena_2 | new | LCG |
| 271 | Athena_3 | new | LCG |
| 272 | AthenaMC_1 | new | LCG |
| 273 | AthenaMC_2 | new | LCG |

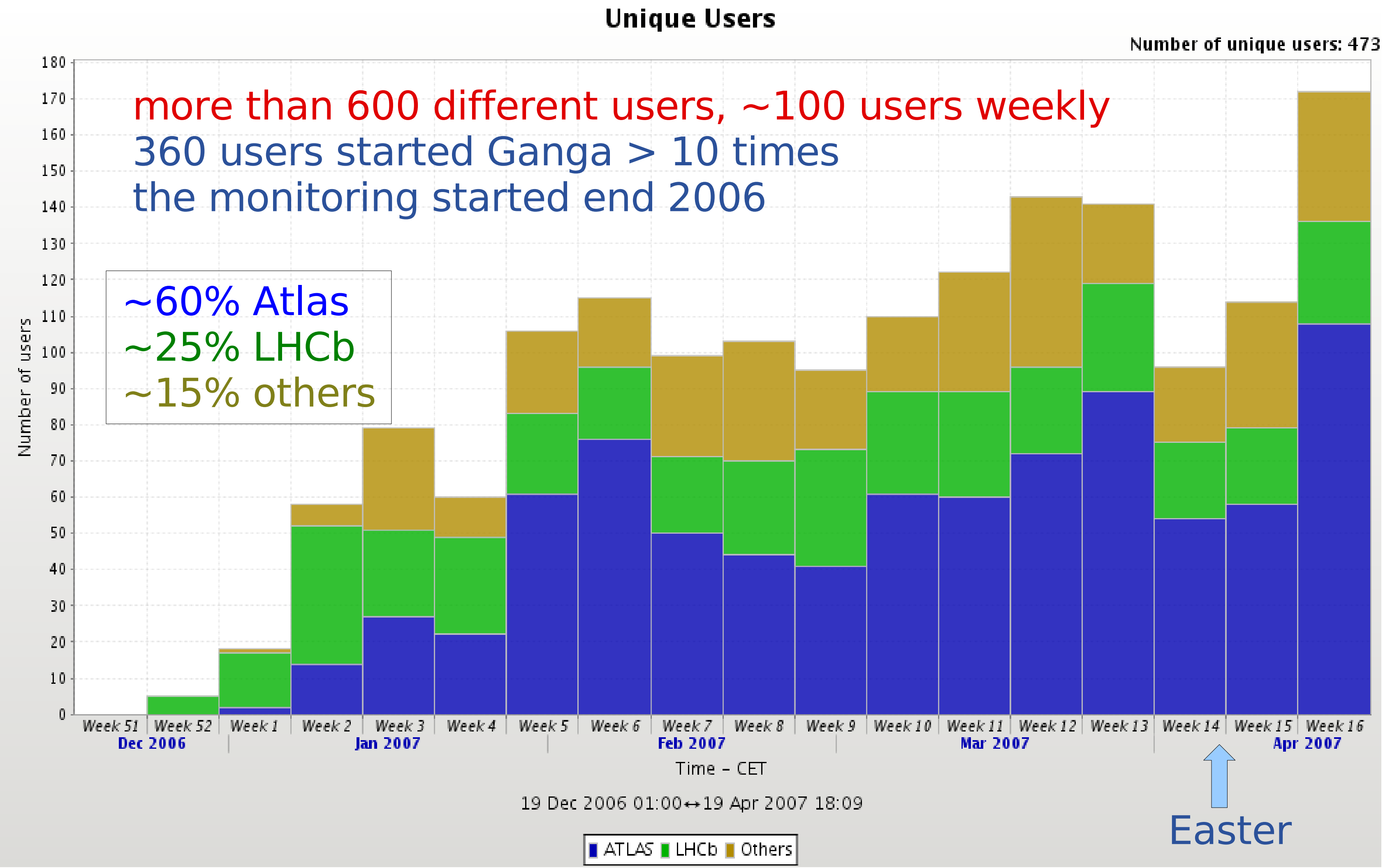
```
Job (
 status =   ew' ,
        thena_1' ,
              e/clat/gar
                    /ga
                    [] ,

                 Dataset
   pe     FC' ,
    ames = [] ,
   dataset = 'csc11.00532
   ) ,
 inputsandbox = [] ,
 application = Athena (
   atlas_release = '11.0.4
   max_events = None ,
   options = None ,
```

10 secs to status update.

many more features to explore

# Usage of Ganga



more than 600 different users, ~100 users weekly

360 users started Ganga > 10 times

the monitoring started end 2006

~60% Atlas

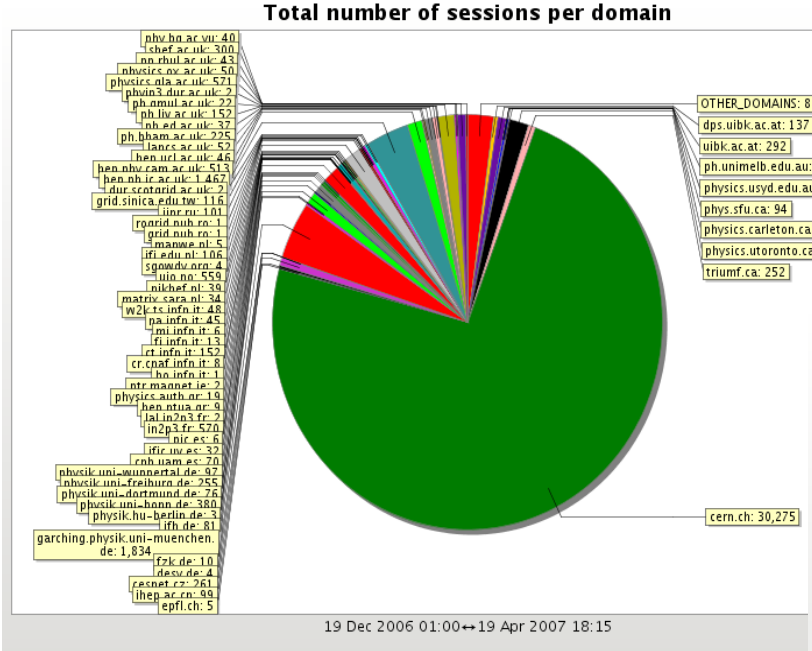
~25% LHCb

~15% others

Easter

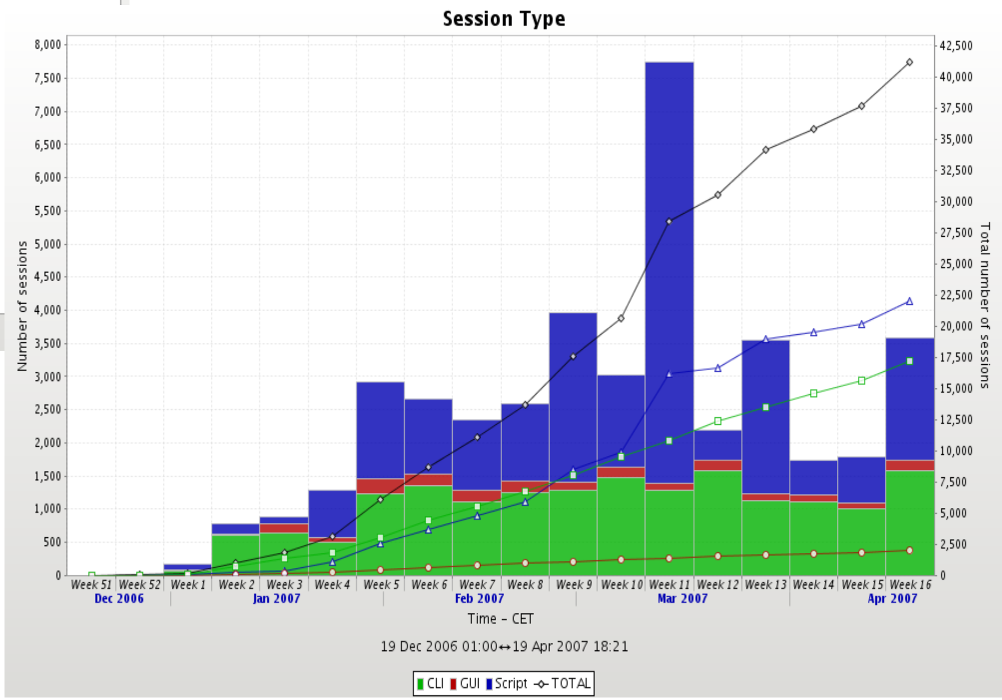
http://gangamon.cern.ch:8888

# Usage of Ganga



~75% at CERN or lxplus
over 50 local sites

CLIP and scripts most popular

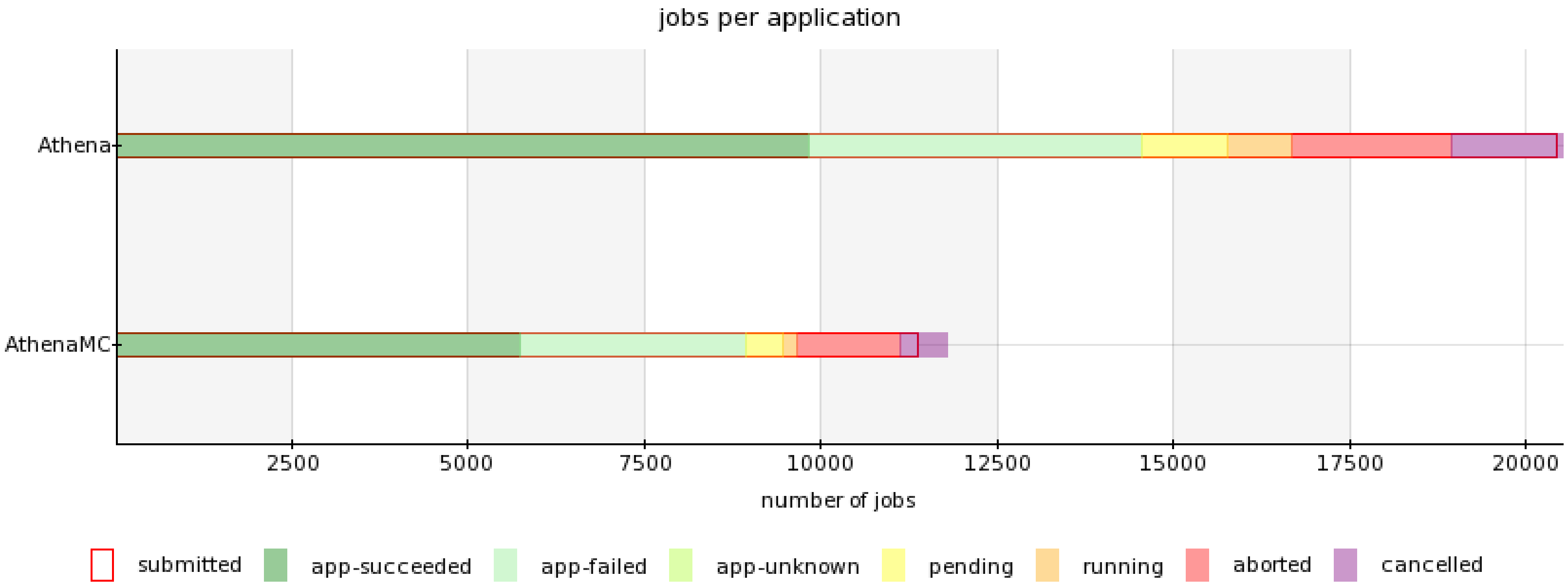


# Ganga in Atlas

- **Data Analysis**
  - many systems: LCG/gLite, OSG/PANDA, NorduGrid, Condor
- **Usage Example**
  - BPhysics Group activities in 2007
    - production of validation samples (AthenaMC application)
      - *50K events produced so far*
        - single muons and simple benchmark B decays
    - analysis of CSC data
      - *10K events analyzed so far, soon expected 450K signal events and 700K background events*
    - 5 regular users now, expected to grow to 15-20 in the future
    - using mainly LCG now

# Ganga in Atlas

- **analysis and simulations jobs run in ATLAS**
  - one month (mid-March to mid-April 2007)
    - 20K analysis jobs, 10K simulation jobs
  - only LCG jobs, others not shown
  - data collected by a ARDA Dashboard sensor integrated with Ganga



# Ganga in LHCb

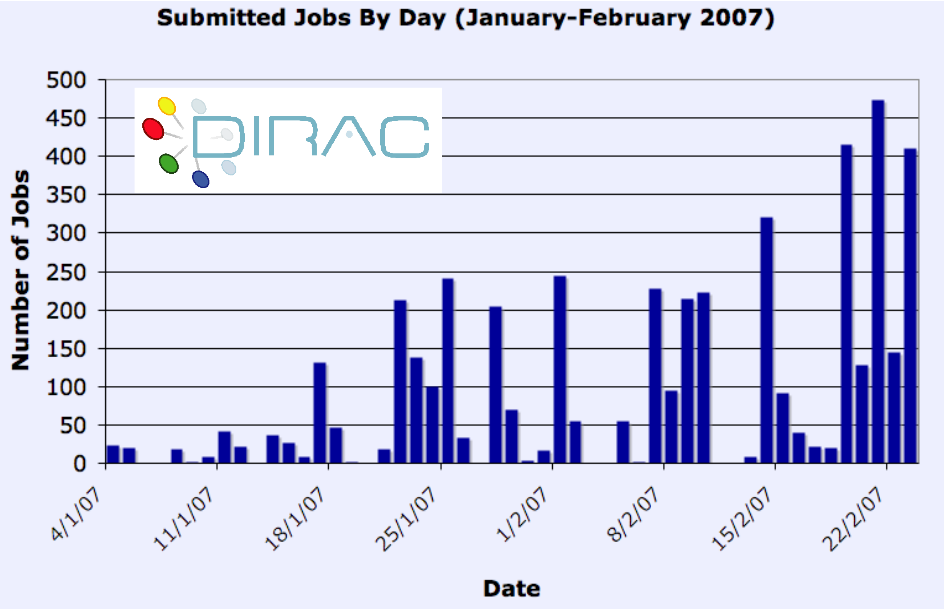
- **Data Analysis**
  - **DIRAC** is the LHCb WMS system (and LCG gateway)
    - Ganga and DIRAC provide the “full” solution for LHCb
- **User examples**
  - UK groups: 2006 activities

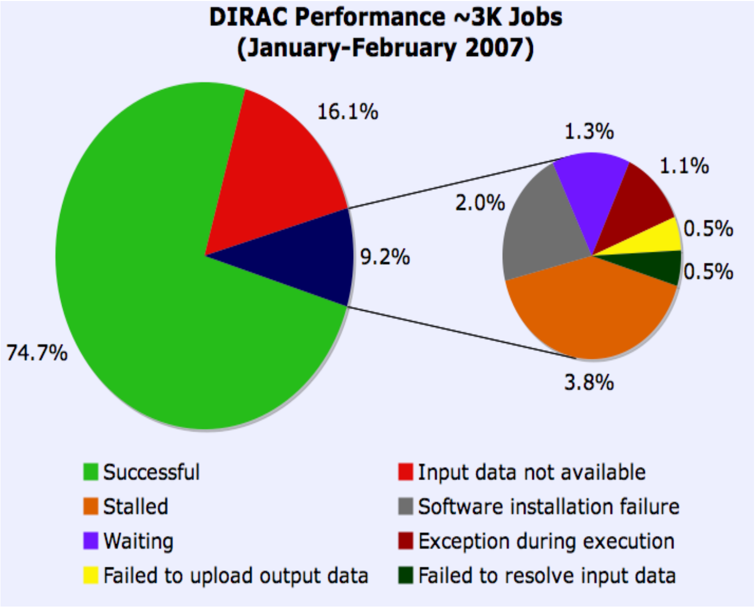
    “Studies of signal and background simulation samples for understanding CP-violation in various B decays"

    - Evaluation of background for $B^+ \to D^0(K^0\pi^+\pi^-)K^+$
      - *Analysis run on* 400k $B^+ \to D^0(K^0\pi^+\pi^-)K^{*0}$
    - Flavour tagging with protons
      - *Analysis run on* 100k $B_s \to J/\psi\ \phi$ *tagHLT*
    - 75 million events analyzed with DaVinci application
      - *using DIRAC (Grid) and Condor (local resources)*
  - Toy Monte-Carlo studies
      - *RooFit application on LCG (Grid) and SGE (local resources)*
      - *Ganga used from within pyRoot and with CINT*

# Ganga in LHCb

- **Daily testing of the distributed analysis framework based on Ganga**
  - set of Ganga scripts submitting and monitoring LHCb jobs
  - the LHCb data analysis framework provides a more robust and fault-tolerant environment on the Grid





Source: Stuart K. Paterson, LHCb Software Week, 1st March 2007

# Ganga Communities

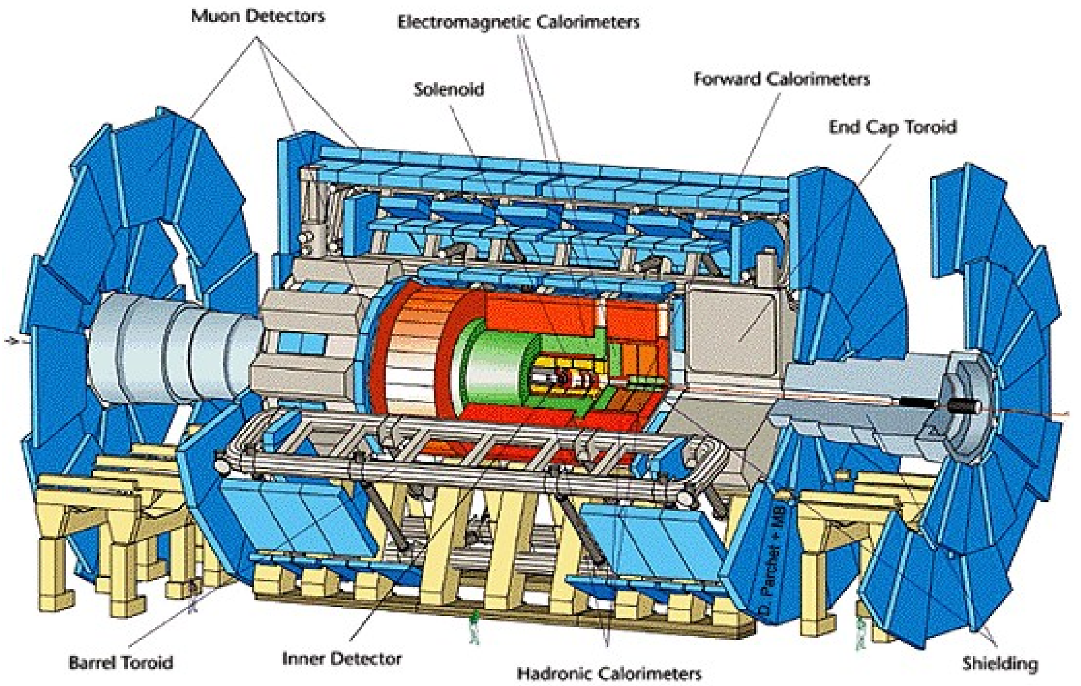
commercial application
ATLAS
LHCb
Cambridge
Ontology
Geant 4
COMPASS
HARP
Garfield
W I E N 2k
ITU
International
Telecommunication
Union
Academia Sinica
Genomics Research Center
med
austron

# Garfield on Grid

Example/DEMO

Thanks to Rob!

# Atlas MDTs

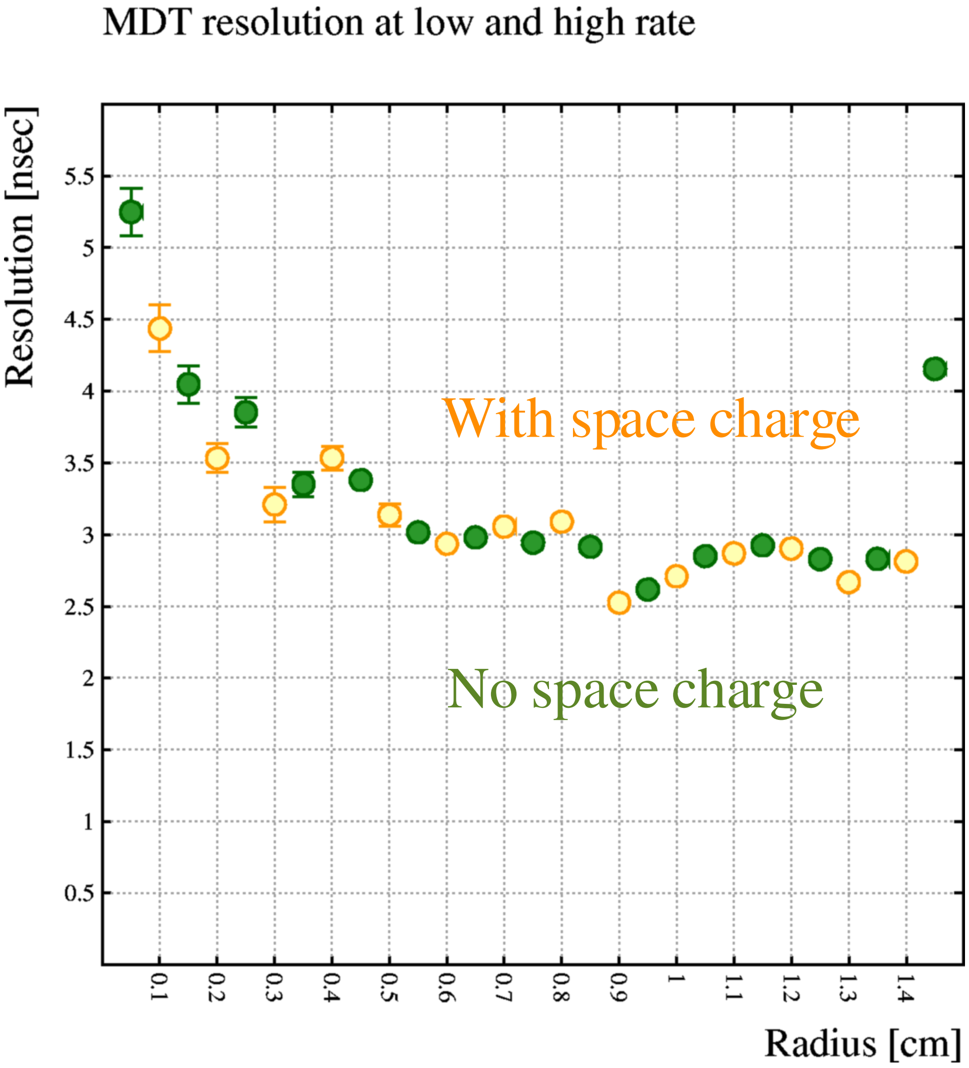

Muon Detectors
Electromagnetic Calorimeters
Solenoid
Forward Calorimeters
End Cap Toroid
D. Parchet + MB
Barrel Toroid
Inner Detector
Hadronic Calorimeters
Shielding

# Purpose of the calculation

- The Atlas muon chambers consist of
  - round tubes with an inner radius of 1.46 cm,
  - a 50 μm diameter W-Re wire,
  - an Ar-based gas under a pressure of 3 bar.
- These tubes can measure the passage of a muon with a spatial resolution better than 100 μm.
- At high rate however, space charge builds up and the resolution deteriorates.
- The purpose of this example is computing the resolution, with and without space charge and, for Ar $N_2$ $CH_4$ (not the usual Atlas MDT gas).

# Resolution curve

- Repeat for radii ranging from 0 to 1.45 cm, both with and without space charge.
- Would have expected a bigger difference.



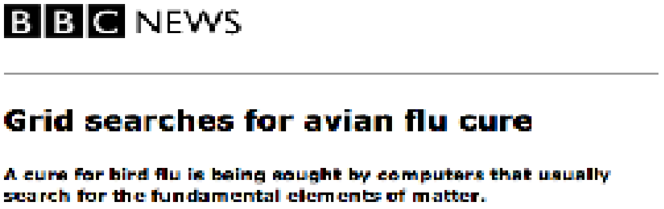
# Technically ...

- This MC part of this calculation is easy to split:
  - each radius is independent,
  - for each radius, statistics can be collected in a large number of independent short jobs,
  - in each job, the random number generators are initialised according to the time of the day at which the job starts to run,
  - each job returns an histogram as only output.

- The histograms are collected by a program which does fits and shows the results.

# Bioinformatics

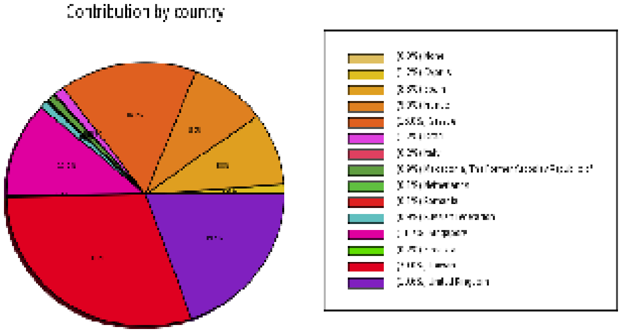
- **Ganga is used as Grid abstraction layer**
- **DIANE – Grid Scheduling add-on**
  - Biomedical application: **Avian Flu Drug Search 2006**



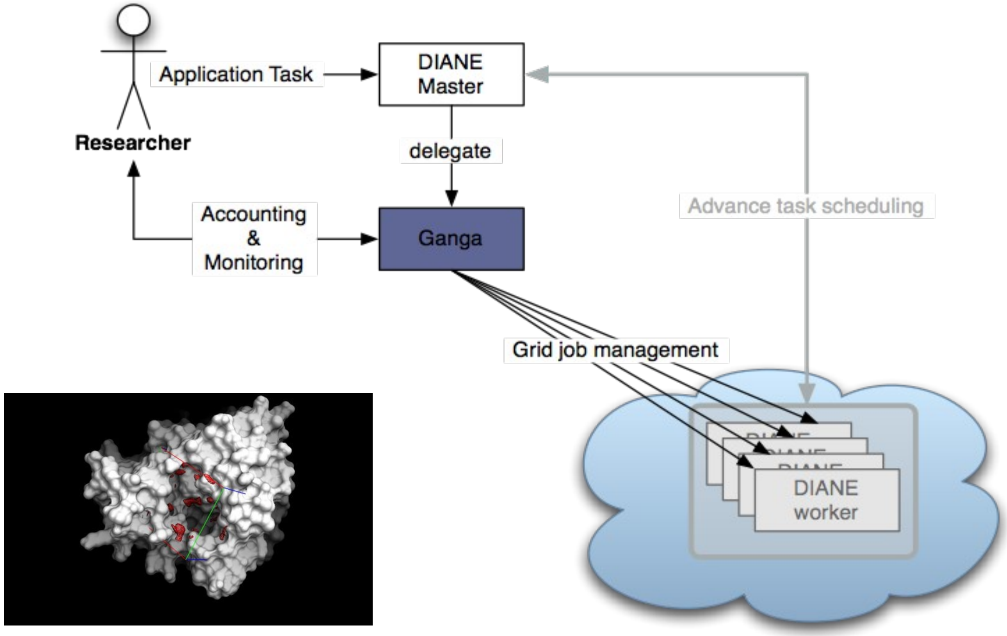
BBC NEWS

**Grid searches for avian flu cure**

**A cure for bird flu is being sought by computers that usually search for the fundamental elements of matter.**

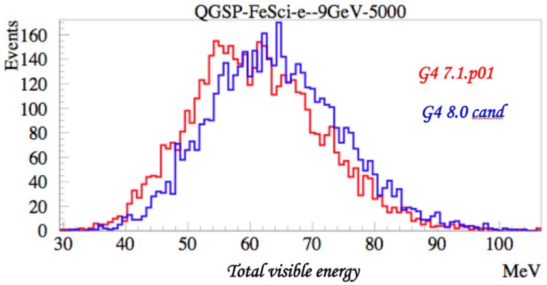


**Statistics of 2580 Ganga Grid jobs**

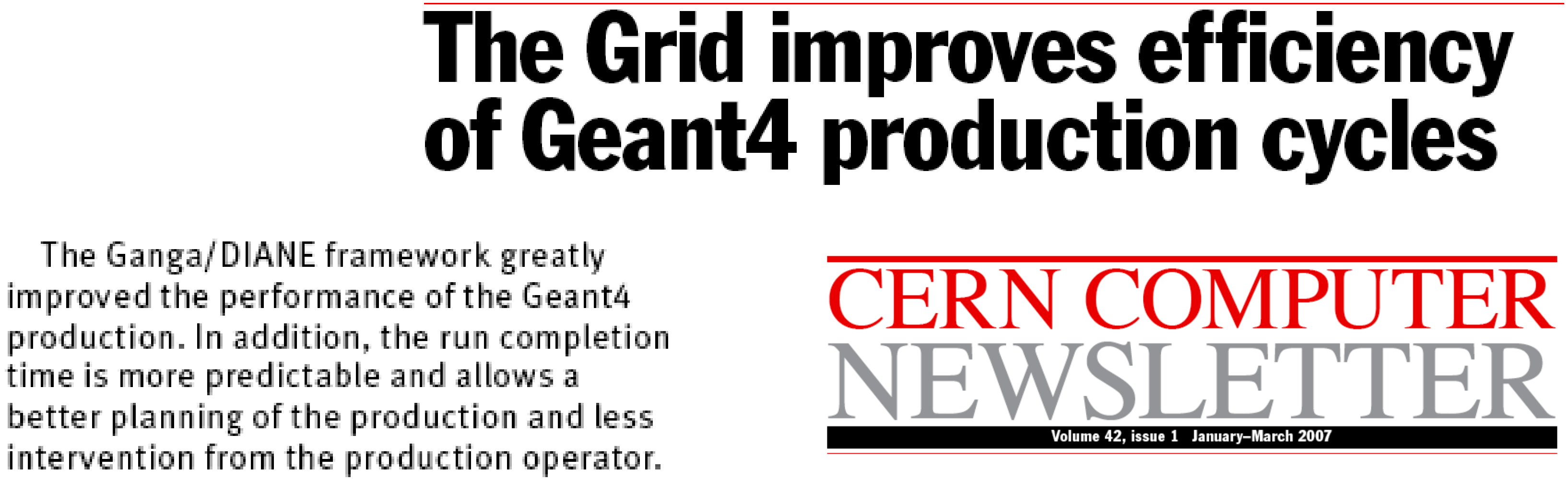


# Geant4 Regression Testing

- **Regression tests run for Geant 4 major releases - twice per year**
  - Search for differences in calorimeter observables
  - Need a few years of CPU, concentrated in short period



- **Geant 4 team now runs these tests on the Grid, using Ganga in conjunction with the DIANE job-distribution framework**



**The Grid improves efficiency of Geant4 production cycles**

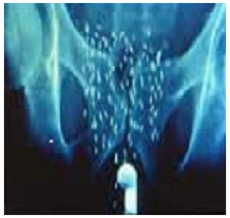
The Ganga/DIANE framework greatly improved the performance of the Geant4 production. In addition, the run completion time is more predictable and allows a better planning of the production and less intervention from the production operator.

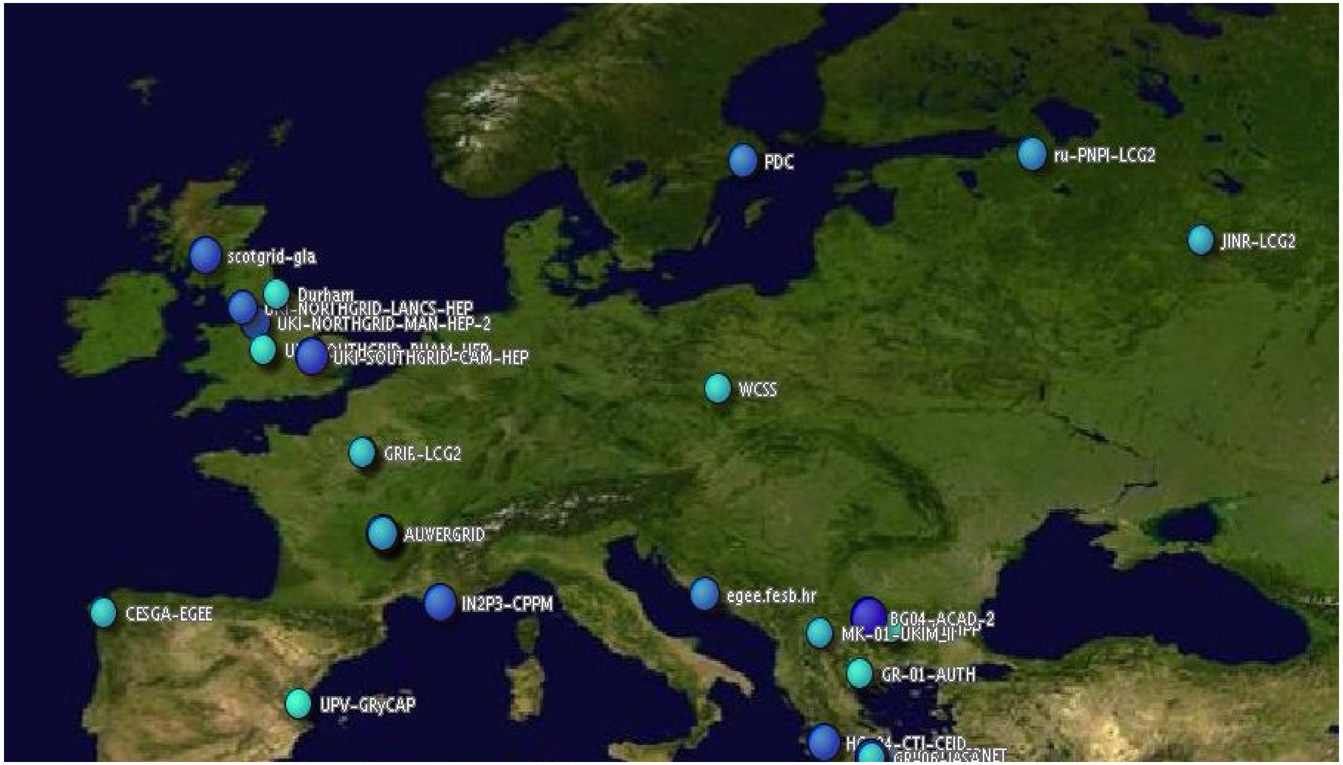
CERN COMPUTER NEWSLETTER

Volume 42, issue 1 January–March 2007

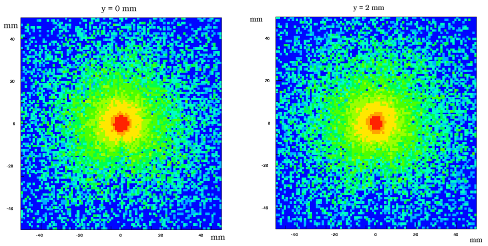
# Geant4 LowE

- **Geant 4 medical applications**
  - radiotheraphy
  - Ganga/DIANE on the Grid for treatment planning



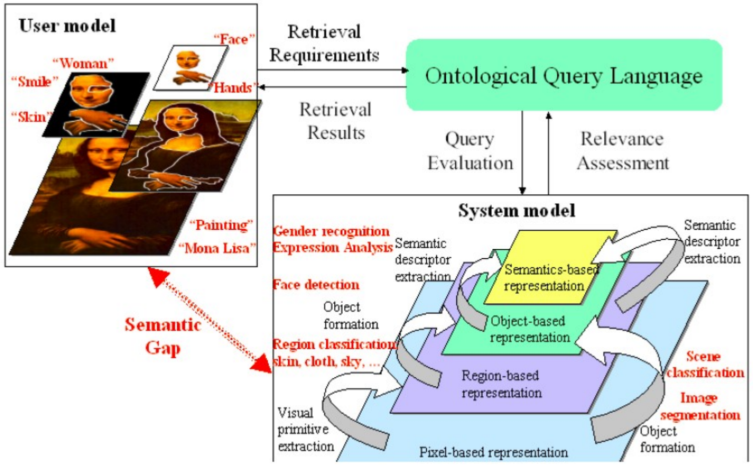


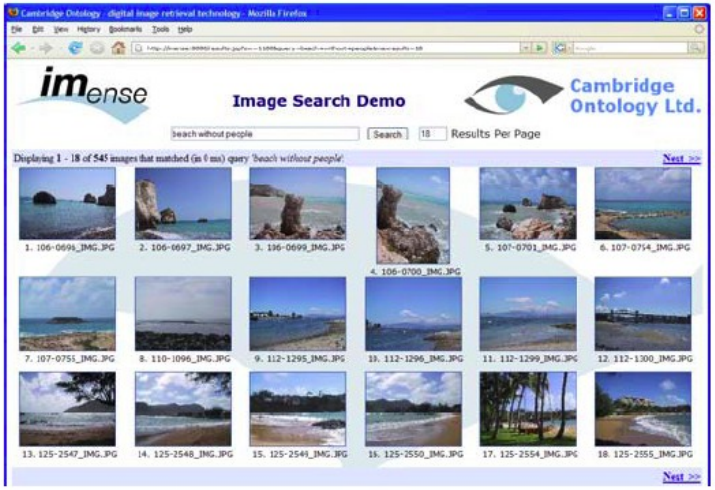
energy deposit histograms updated as the Grid job runs



# Cambridge Ontology

- **Cambridge Ontology is developing solutions for image retrieval based on analysis of image content**
  - combines state-of-the-art computer vision with machine learning, natural-language processing, and information-retrieval methods





beach without people

18

Displaying 1 - 18 of 545 images that matched (in 0 ms) query *'beach without people'*

- Large amounts of CPU time needed for image processing
- Mini-PIPSS project set up to enable processing to be done on the Grid
- Ganga is used for the job submission
  - Specialised Ganga components for Cambridge Ontology applications

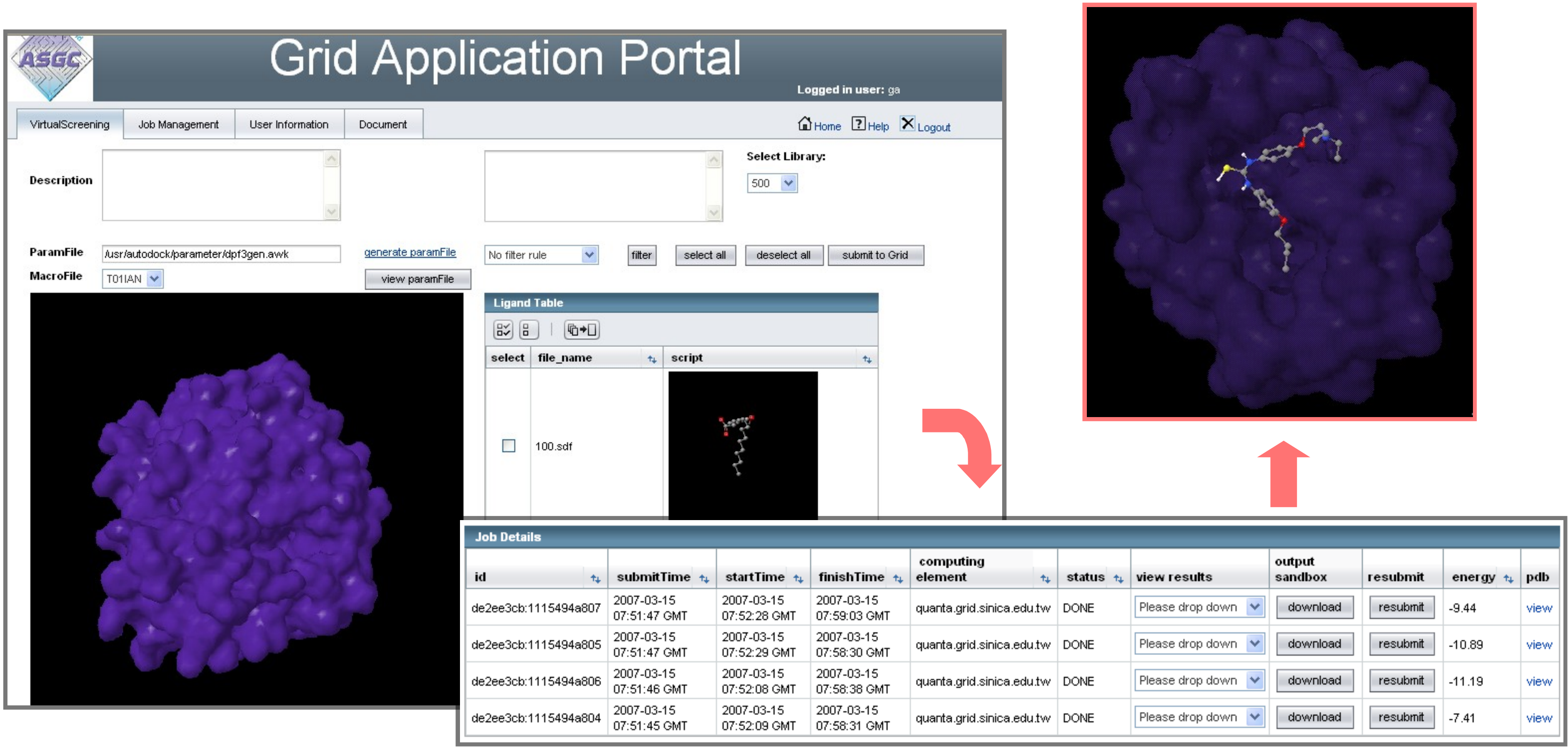
# Summary

- **Ganga is easy to use**
  - job submission and application configuration
  - rich functionality and batteries included
    - 13 backend plugins, 12 application plugins
  - choice of 3 interfaces: CLIP, GUI, script
- **Ganga is becoming <u>the</u> DA tool in Atlas/LHCb**
  - the main Grid interface for data analysis in Atlas and LHCb
  - numerous groups using it for their activities
- **Ganga is extensible and reusable**
- **Ganga is used**
  - 100 users per week, 600 users since January 2007
  - many user communities are picking it up
- **Excellent collaboration with EP/SFT projects**

# http://cern.ch/ganga

- documentation
- tutorials
- savannah
- presentations
- release information
- download

# Grid Portal for Biologists



Grid Application Portal

Logged in user: ga

VirtualScreening | Job Management | User Information | Document

Home | Help | Logout

Description

Select Library: 500

ParamFile /usr/autodock/parameter/dpf3gen.awk

generate paramFile

MacroFile T01IAN

view paramFile

No filter rule

filter | select all | deselect all | submit to Grid

Ligand Table

| select | file_name | script |
|---|---|---|
| □ | 100.sdf | |

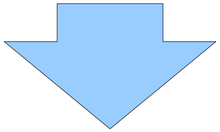
Job Details

| id | submitTime | startTime | finishTime | computing element | status | view results | output sandbox | resubmit | energy | pdb |
|---|---|---|---|---|---|---|---|---|---|---|
| de2ee3cb:1115494a807 | 2007-03-15 07:51:47 GMT | 2007-03-15 07:52:28 GMT | 2007-03-15 07:59:03 GMT | quanta.grid.sinica.edu.tw | DONE | Please drop down | download | resubmit | -9.44 | view |
| de2ee3cb:1115494a805 | 2007-03-15 07:51:47 GMT | 2007-03-15 07:52:29 GMT | 2007-03-15 07:58:30 GMT | quanta.grid.sinica.edu.tw | DONE | Please drop down | download | resubmit | -10.89 | view |
| de2ee3cb:1115494a806 | 2007-03-15 07:51:46 GMT | 2007-03-15 07:52:08 GMT | 2007-03-15 07:58:38 GMT | quanta.grid.sinica.edu.tw | DONE | Please drop down | download | resubmit | -11.19 | view |
| de2ee3cb:1115494a804 | 2007-03-15 07:51:45 GMT | 2007-03-15 07:52:09 GMT | 2007-03-15 07:58:31 GMT | quanta.grid.sinica.edu.tw | DONE | Please drop down | download | resubmit | -7.41 | view |

- **Interface created by biologists** (Model-View-Controller design pattern)
  - The **Model** makes use of **Ganga** as a submission tool and **DIANE** to better handle docking jobs on the Grid
  - The **Controller** organizes a set of actions to perform the virtual screening pipeline; The **View** represents biological aspects

# User requirements

- **User requirements**
  - interact with all backend systems in a very similar way
    - submit, kill, monitor jobs
  - configure the applications easily and transparently across desired backends
  - organize work
    - job history: keep track of what user did
    - save job outputs in a consistent way
    - reuse configuration of previously submitted jobs



**Ganga**

# Introduction

## Mission statement

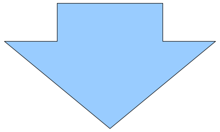
- Running jobs on the Grid
  - Make it easy and integrated with application environment
  - Allow quick transition between local PC, cluster, Grid...
  - Organize work, keep history of jobs,...

- **Users**
  - Atlas, LHCb
  - EGEE applications
- **Developers/supporters**
  - ATLAS, LHCb, ARDA
    multiple funding agencies

High-Energy Physics:
data analysis
private data productions
Monte Carlo

# Wish list

- **Wish list**
  - easy to learn and use
  - powerful: not limiting the features of the backends
  - close to the application (which is typically compiled locally)
  - not imposing single working style: scripts, command line, GUI,...



**Ganga**