# Ganga Tutorial

*Jakub T. Moscicki (KUBA)*
*CERN*

www.eu-egee.org

Information Society and Media

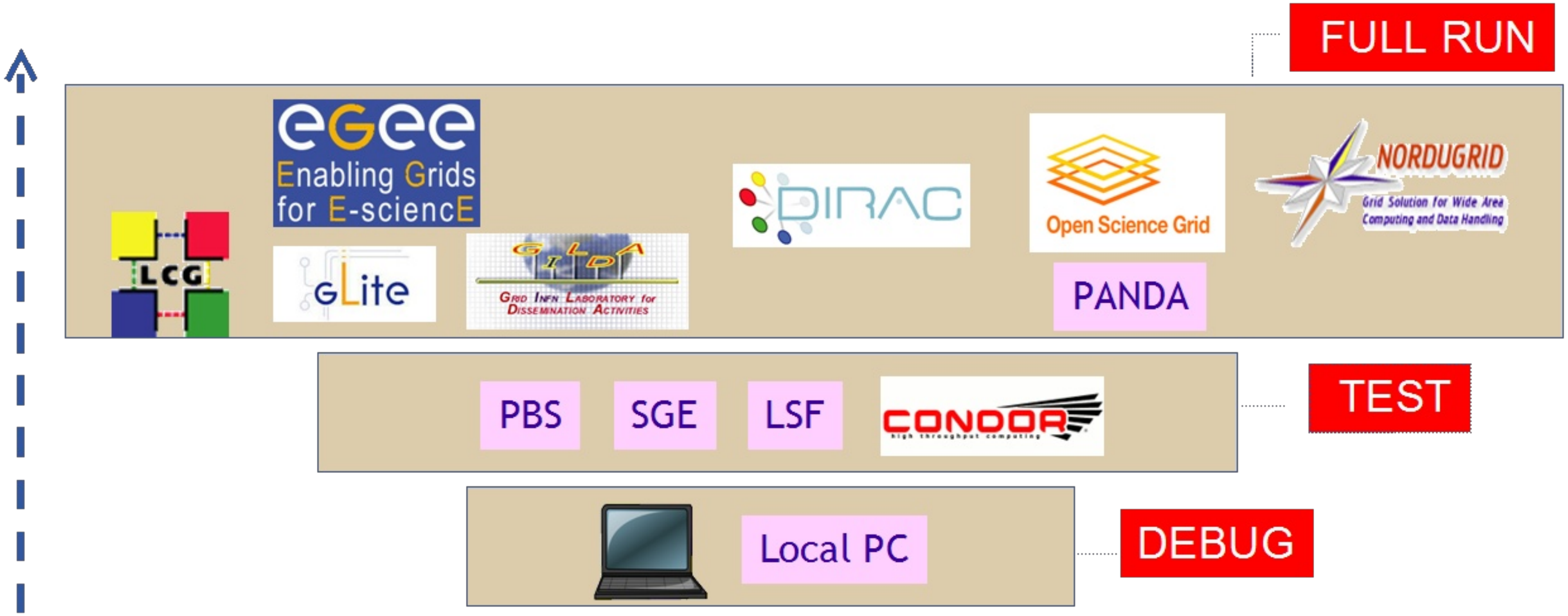
# Agenda

- **Part I: Ganga introduction**
- **Part II: Ganga hands-on**
- **Part III: More about Ganga**

# Part I: Ganga Overview

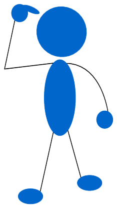
# What Users Want

- **Users want:**
  - Development on the laptop; full analysis on “The Grid™”.
  - To get results quickly, utilizing all of the resources available, wherever they are.
  - A familiar and consistent user interface to all of the resources.
- **Users don’t want:**
  - To know the details of the grids or the resources.
  - To learn yet another tool in order to access some resources
  - To have to reconfigure their application to run on different resources.

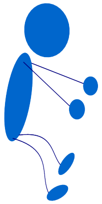

FULL RUN
egee Enabling Grids for E-sciencE
LCG
gLite
GILDA Grid INFN Laboratory for Dissemination Activities
DIRAC
Open Science Grid
PANDA
NORDUGRID Grid Solution for Wide Area Computing and Data Handling
PBS
SGE
LSF
CONDOR high throughput computing
TEST
Local PC
DEBUG

**“configure once, run anywhere”**

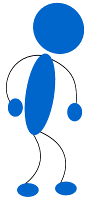
# Frequently Asked Questions



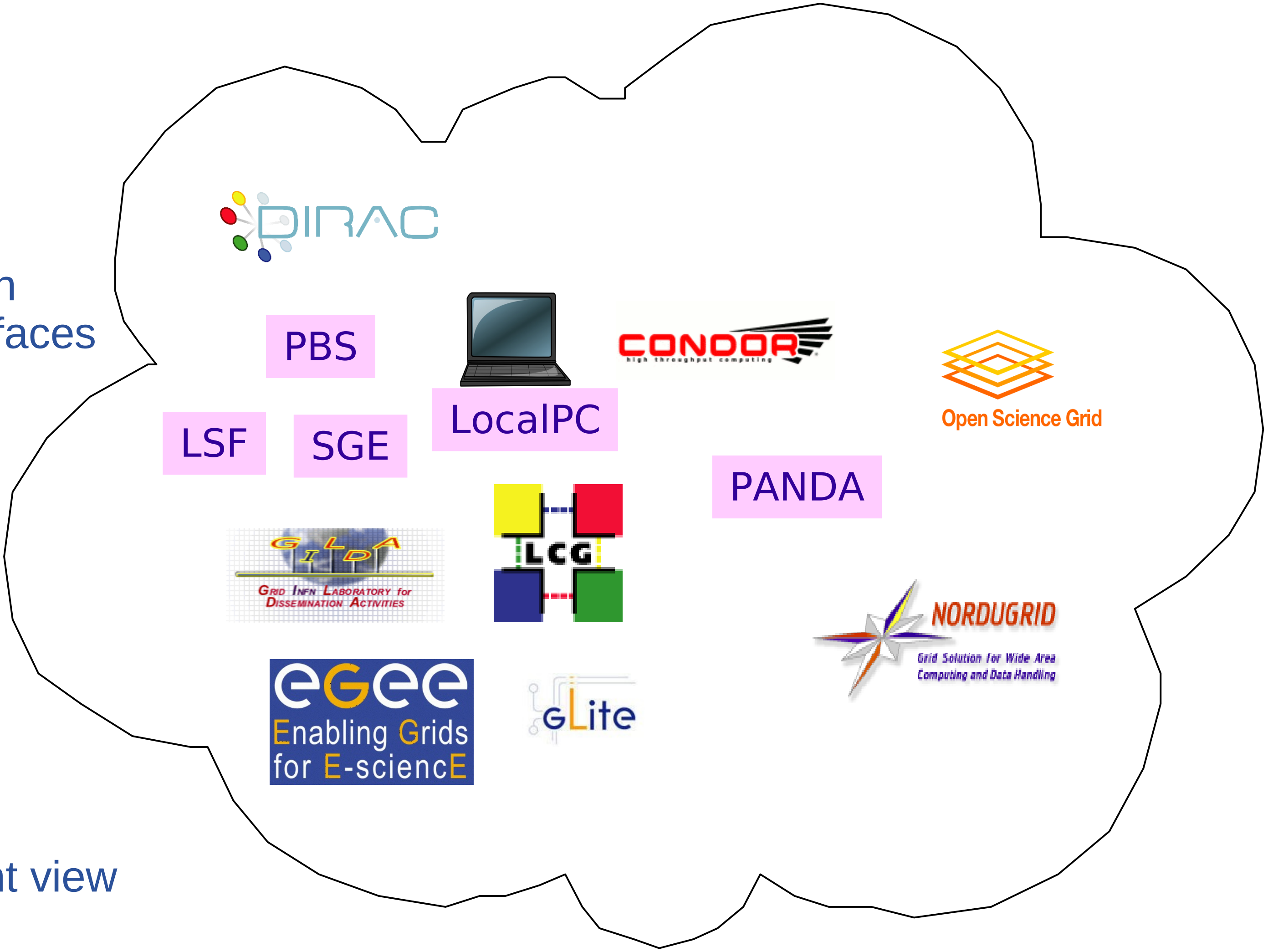
I must learn many interfaces



How to configure my applications?



Do I get a consistent view on all my jobs?



PBS

LocalPC

LSF

SGE

PANDA

DIRAC
CONDOR high throughput computing
Open Science Grid
GILDA Grid INFN Laboratory for Dissemination Activities
LCG
NORDUGRID Grid Solution for Wide Area Computing and Data Handling
egee Enabling Grids for E-sciencE
gLite

# Ganga Overview

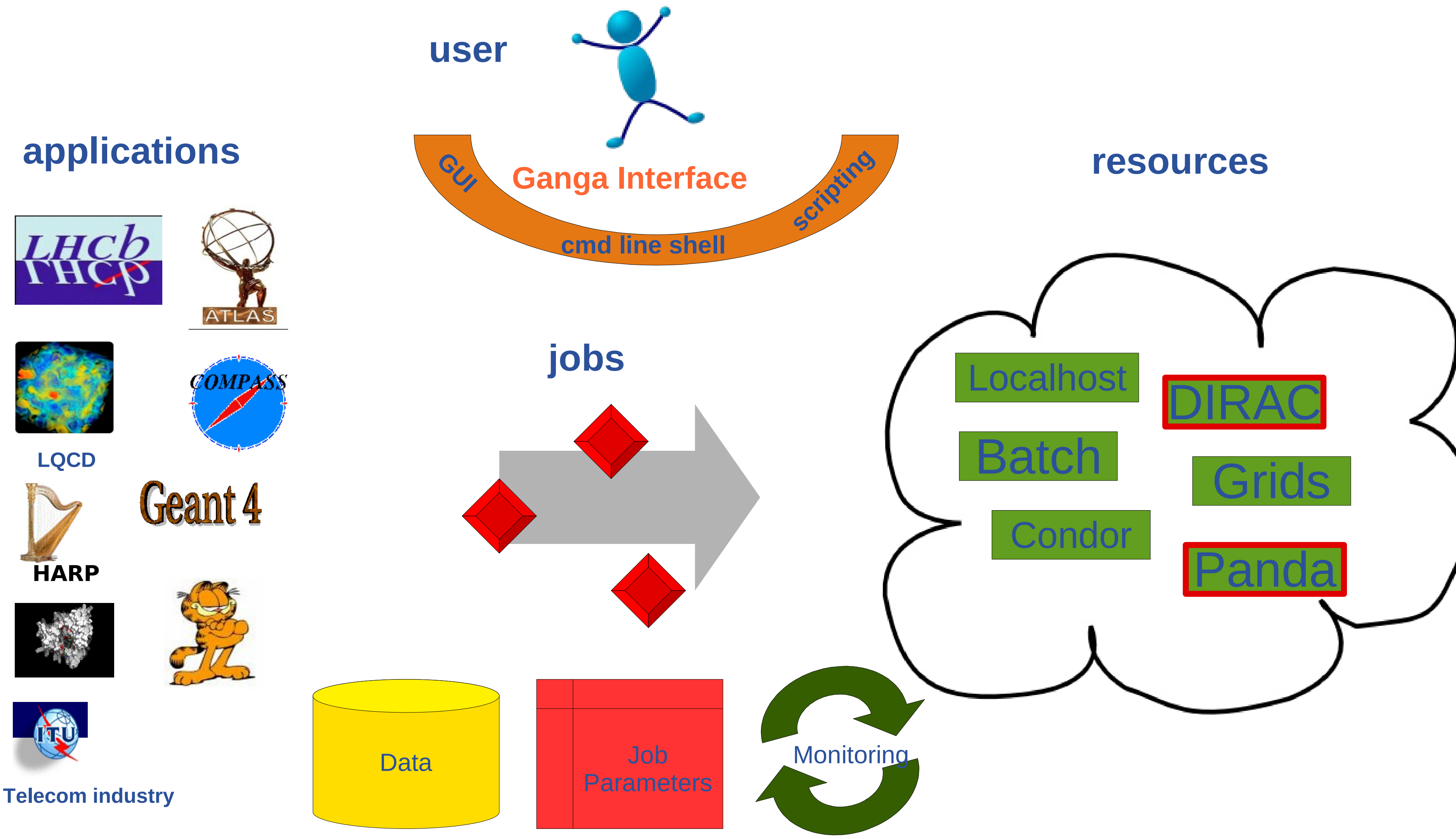
- **Ganga is a user-friendly job management tool.**
  - Jobs can run locally or on a number of batch systems and grids.
  - Easily monitor the status of jobs running everywhere.
  - To change where the jobs run, change one option and resubmit.

- **Ganga is the main distributed analysis tool for LHCb and ATLAS.**
  - Experiment-specific plugins are included.

- **Ganga is an open source community-driven project:**
  - Core development is joint between LHCb and ATLAS
  - Modular architecture makes it extensible by anyone
  - Mature and stable, with an organized development process

# Ganga Overview

*"configure once, run anywhere"*



user

Ganga Interface

GUI

cmd line shell

scripting

applications

LQCD

HARP

Telecom industry

jobs

resources

Localhost

DIRAC

Batch

Grids

Condor

Panda

Data

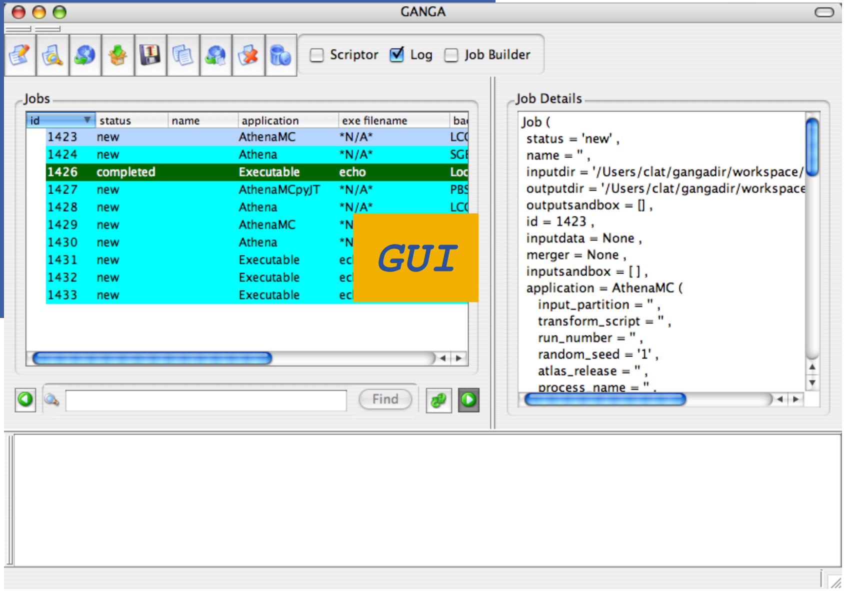
Job Parameters

Monitoring

# User interfaces

```
*** 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
               backend.actualCE
#    1   completed                             Executable
compute.hpc.unimelb.edu.au:2119/jobmanage
```

**CLIP**



| 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 | |
| 1430 | new | | Athena | *N | |
| 1431 | new | | Executable | ec | |
| 1432 | new | | Executable | ec | |
| 1433 | new | | Executable | ec | |

```
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 = '' ,
```

**GUI**

```
#!/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”)
```

**GPI & Scripting**

# Ganga

- **Ganga: Job Management Tool**
  - a utility which you download to your computer
    - or it is already installed in your institute in a shared area
      - *for example: /nfs/sw/ganga/install/5.1.0*
  - Ganga is an **add-on** to installed software
    - Grid UI must be installed separately
    - Batch system should be set-up separately
    - NorduGrid ARC middleware is shipped with Ganga

Ganga
Application Software
LSF Client
LCG UI

# Hello World Examples

#1. Hello World Locally

```
j = Job()
j.backend=Local()
j.submit()
```

#2. Hello World on Nordu Grid

```
j = Job()
j.backend=NG()
j.submit()
```

#3. Hello World on EGEE

```
j = Job()
j.backend=LCG()
j.submit()
```

You may use all features of a first-class programming language to write complex scripts

loops, ifs, variables files, math, network modules,...

# Real Application: the ATLAS data analysis application

Scripting mode

```
$ ganga athena \
--inDS myInputDataset.txt\
--outputdata myOutput.root \
--split 3 \
--maxevt 100 \
--lsf \
jobOptions.py
```

quick

CLIP mode

```
j = Job()
j.application=Athena()
j.application.prepare()
j.application.option_file='jobOptions.py‘

j.inputdata=DQ2Dataset()
j.inputdata.type='DQ2_LOCAL'
j.inputdata.dataset=“myInputDataset.txt”

j.outputdata=DQ2OutputDataset()
j.outputdata.outputdata=[‘myOutput.root']

j.splitter = AthenaSplitterJob(numsubjobs=3)
j.merger = AthenaOutputMerger()

j.backend = LSF()
j.submit()

j2 = j.copy()
j2.backend=LCG( CE=’ce102.cern.ch:2119/jobmanager-lcglsf-grid_2nh_atlas’ )
j2.submit()
```

application

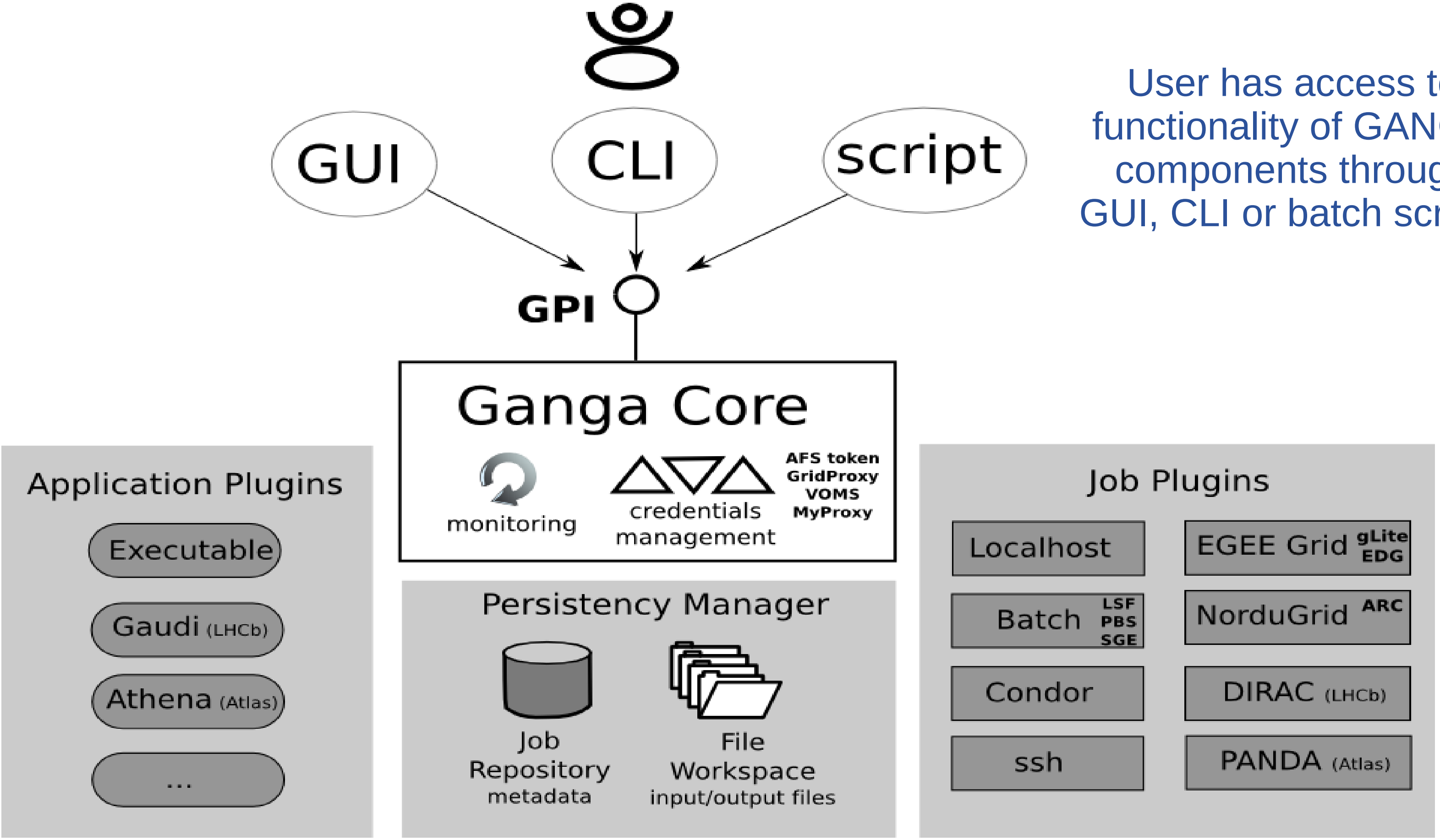
inputdata

outputdata

Splitter & Merger

flexible

# Behind the scenes ...



User has access to functionality of GANGA components through GUI, CLI or batch scripts

application configuration

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

**Data Management**

input data selection, output location specification

GANGA

# GANGA User Communities

**More than 2300 individual users**

LHCb

**HARP**

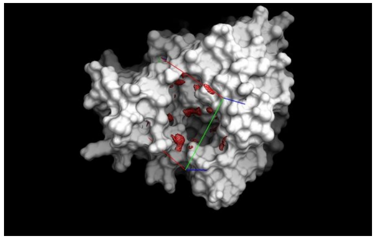
**Garfield**

**Geant 4**

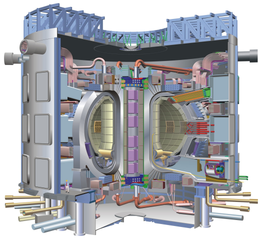
COMPASS

ITU

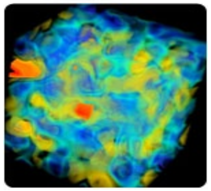
med austron



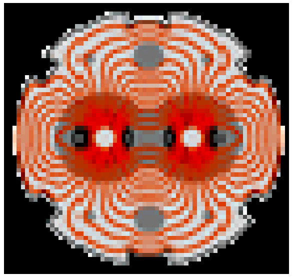
Academia Sinica
Genomics Research Center



Fusion

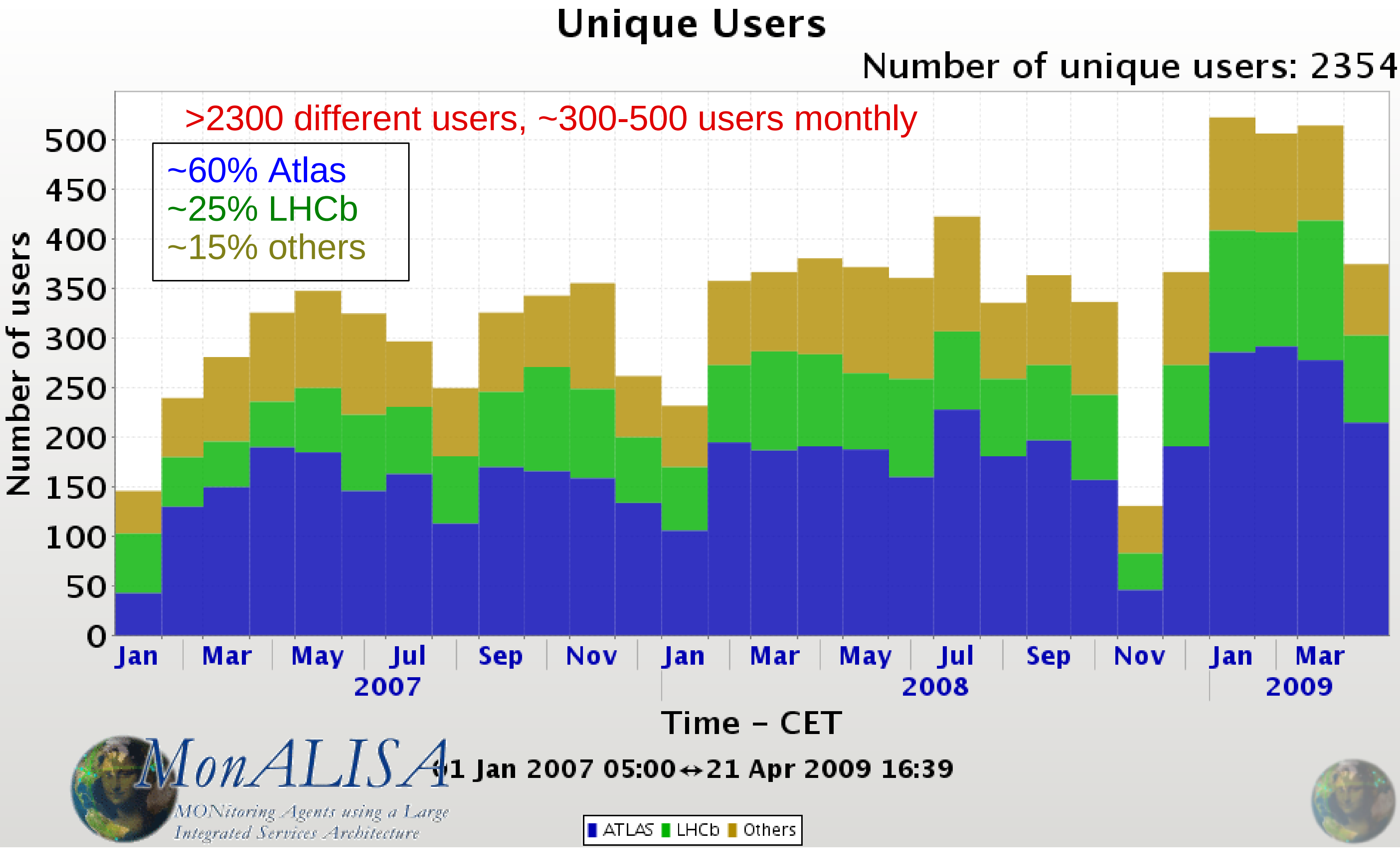


**LQCD**



Sixtrack/LHC

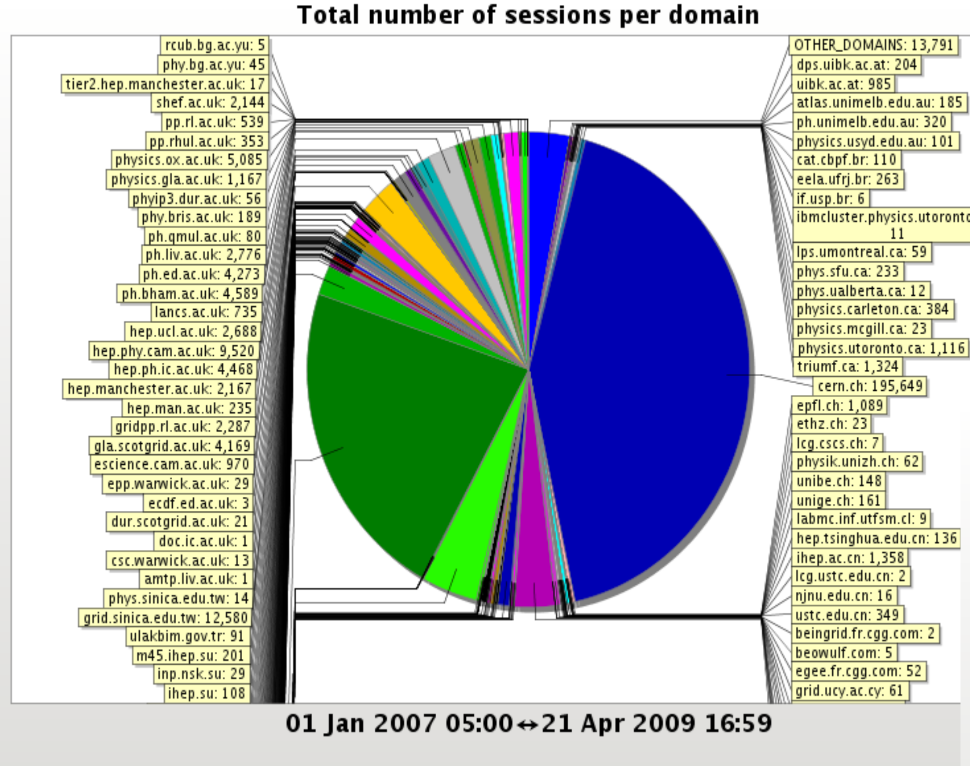
# More than 1400 unique users



>2300 different users, ~300-500 users monthly

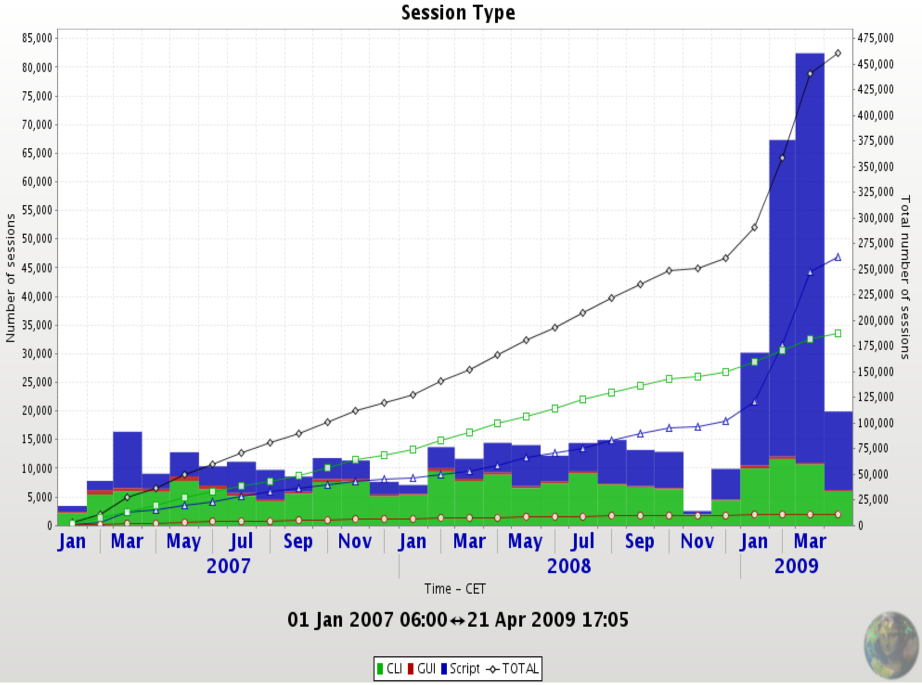
~60% Atlas
~25% LHCb
~15% others

# Ganga usage



Installed at ~200 sites

CLIP and scripts most popular



# Why not portal?

- **Many (not all) scientific applications are:**
  - developed/modified/extended on local machine
  - use local resources (files, environment)
  - are scripted (e.g. higher-level logic is built around them)
- **What user mobility (desktop vs laptop)?**
  - Ganga may use remote repository (and workspace)
  - So user may see his jobs submitted from a different machine and interact with them (resubmit, kill)...
- **Ganga approach:**
  - lean and neat utility with interface in **Python**
  - + GUI if you like
  - + scripting if you like more

# What Ganga can do for you

- **help configure applications**
  - if you teach it about your application
- **organize work**
  - job history: keep track of what user did
  - save job outputs in a consistent way
  - reuse configuration of previously submitted jobs
- **you do not have to learn many interfaces**
  - one set of (simple) commands which are always the same
    - localy and on the Grid
- **Ganga will not hide anything from you**
  - you still have access to automatically generated jdls, job id etc.
  - ganga will tell you what it does to submit jobs
- **Works with no change on Grid production systems, GILDA, ...**

# More info.

- Ganga Home: http://cern.ch/ganga
- Official Ganga User's Guide: http://cern.ch/ganga/user/html/GangaIntroduction/
- Looking for help: project-ganga-developers@cern.ch

# The Ganga development team

- Ganga is supported by HEP

-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), D.Vanderster (CERN)

# Part I: Practical Ganga

www.eu-egee.org

Information Society and Media

# Download, Install, First launch

## Download & Install

```
wget http://cern.ch/ganga/download/ganga-install

python ganga-install \
       --prefix=/usr/local/ganga/prefix \
       --extern=GangaGUI,GangaPlotter \
       5.0.0
```

download installer

installation prefix

Installation of external modules

Ganga version

## First Launch

```
export PATH=$HOME/opt/ganga/install/5.0.0/bin:$PATH
```

start Ganga with inline configurations

```
$ ganga
```

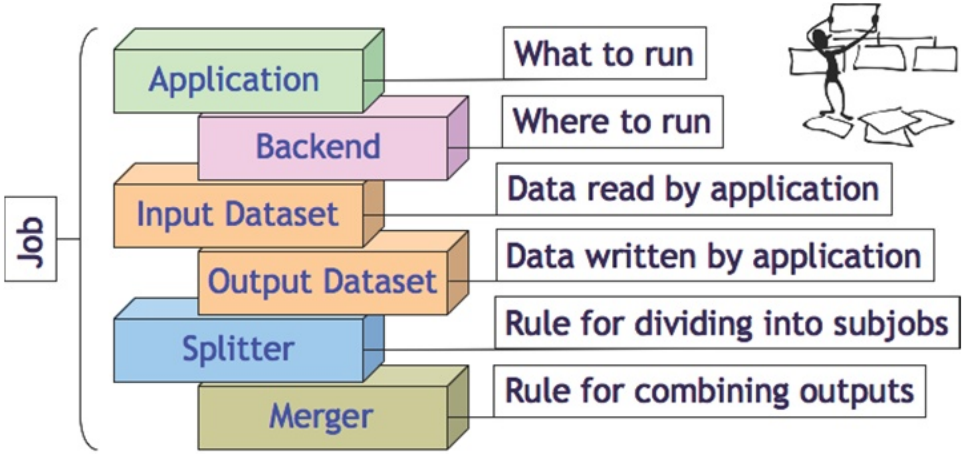
```
*** Welcome to Ganga ***
Version: Ganga-5-0-0
Documentation and support: http://cern.ch/ganga
Type help() or help('index') for online help.

In [1]:
Do you really want to exit ([y]/n)?
```

Ganga CLIP

<ctrl>-D to exit Ganga CLIP

# Submitting a Job with Ganga

## What is a Ganga Job?



**Run the default job locally:**

```
Job().submit()
```

**Default job on the EGEE grid:**

```
Job(backend=LCG()).submit()
```

**Listing of the existing jobs:**

```
jobs
```

**Get help (e.g. on a job):**

```
help(jobs)
```

**Display the nth job:**

```
jobs(n)
```

**Copy and resubmit the nth job:**

```
jobs(n).copy().submit()
```

**Copy and submit to another grid:**

```
j=jobs(n).copy()
j.backend=NG()
j.submit()
```

**Kill and remove the n**$^{th}$ **job:**

```
job(n).kill()
job(n).remove()
```

# Configuration

## Syntax of .gangarc file

```
[Configuration]
TextShell = IPython
... ...
[LCG]
GLITE_ENABLE = True
... ...
```

Python ConfigParser standard

## How to set configurations

```
Hard-coded default configurations
```

release config

site config

```
export GANGA_CONFIG_PATH=/usr/local/ganga/prefix/install/etc/Gilda.ini:GangaTutorial/Tutorial.ini
ganga --config-path=/usr/local/ganga/prefix/install/etc/Gilda.ini:GangaTutorial/Tutorial.ini
```

```
~/.gangarc
ganga -o
```

user config

## Override sequence

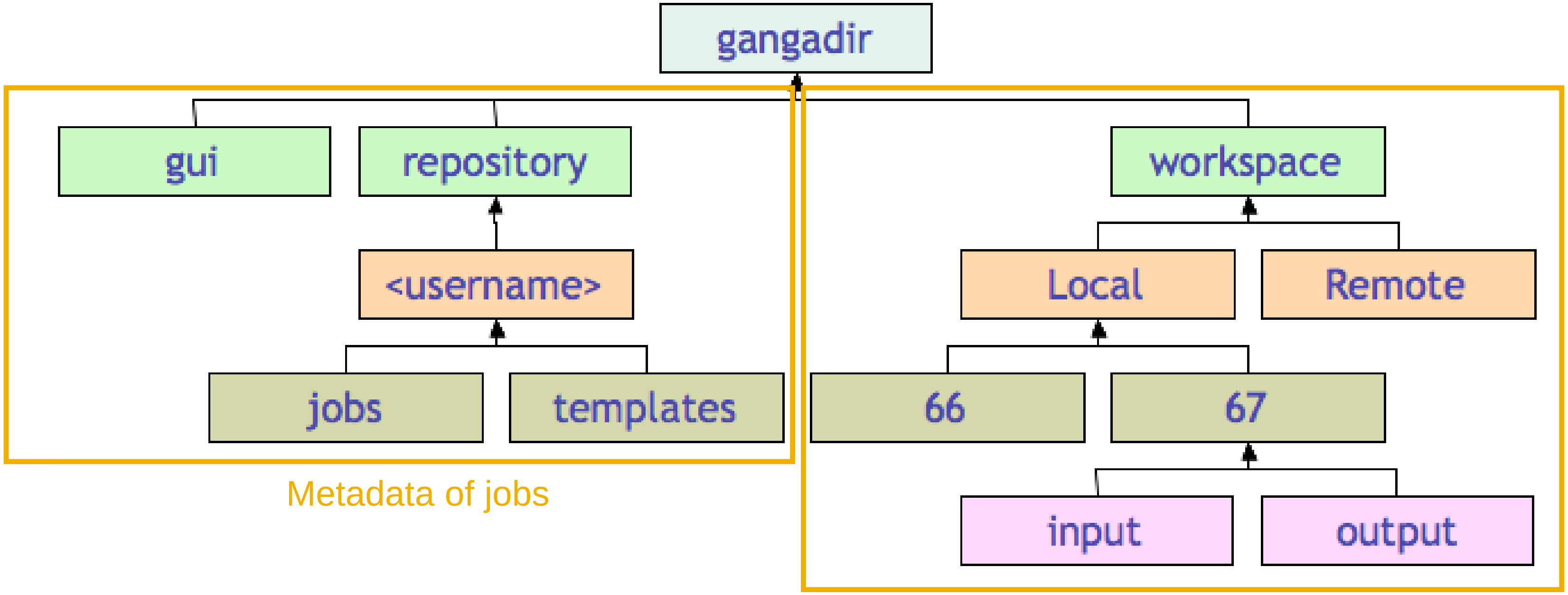
```
user config > site config > release config
```

# `gangadir`

- **gangadir folder is created at the first launch within $HOME directory**

- **To locate it in different directory:**
  - `[Configuration] gangadir = /alternative/gangadir`
  -

- `Job Repository may also be stored remotely in a database`



Metadata of jobs

Data of jobs

# Some handy functions

- <tab> completion
- <page up/down> for cmd history
- system command integration

```
In[1]: j = jobs[1]

In[2]: cat $j.outputdir/stdout
Hello World
```

- Job template

```
In[1]: t = JobTemplate(name=’lcg_simple’)

In[2]: t.backend = LCG(middleware=’EDG’)

In[3]: templates
Out[3]: Statistics: 1  templates
--------------
#   id      status       name   subjobs
application          backend
    backend.actualCE
#    3    template  lcg_simple
Executable              LCG

In[4]: j = Job(templates[3])
In[5]: j.submit()
```

- `In[1]: plugins()`
  - `plugins(‘backends’)`
- `In[2]: help()`
- etc.

# Part II: Ganga hands-on

# Step 0: launch Ganga CLIP

https://twiki.cern.ch/twiki/bin/view/ArdaGrid/EGEETutorialPackage

- Skip the installation step
- Start your Ganga CLIP session:

```
shell> ganga
```

# Step 1: Your first Ganga job - an arbitrary shell script

```
In [1]: !pico myscript.sh
```

```
#!/bin/sh
echo “Hello ${1} !”
echo $HOSTNAME
cat /proc/cpuinfo | grep 'model name’
cat /proc/meminfo | grep 'MemTotal‘
echo “Run on `date`”
```

```
In [2]: !chmod +x myscript.sh
In [2]: j = Job()
In [3]: j.application = Executable()
In [4]: j.application.exe  = File(‘myscript.sh’)
In [5]: j.application.args = [‘Helsinki’]
In [6]: j.backend = Interactive()
In [7]: j.submit()

In [8]: jobs
```

# Step 2: Your first Ganga job - an arbitrary shell script

```
In [9]: j = j.copy()
In [10]: j.backend = Local()
In [11]: j.submit()

In [12]: jobs

In [13]: j.peek()
In [14]:cat $j.outputdir/stdout
```

```
./myscript.sh Budapest
```

# Step 3: your first Ganga job on the Grid

```
In [15]:j = j.copy()
In [16]:j.backend = LCG()
In [17]:j.application.args = [‘Somewhere in the world...’]
In [18]:j.submit()

In [19]:j

In [20]:cat $j.backend.loginfo(verbosity=1)

In [21]:jobs
```

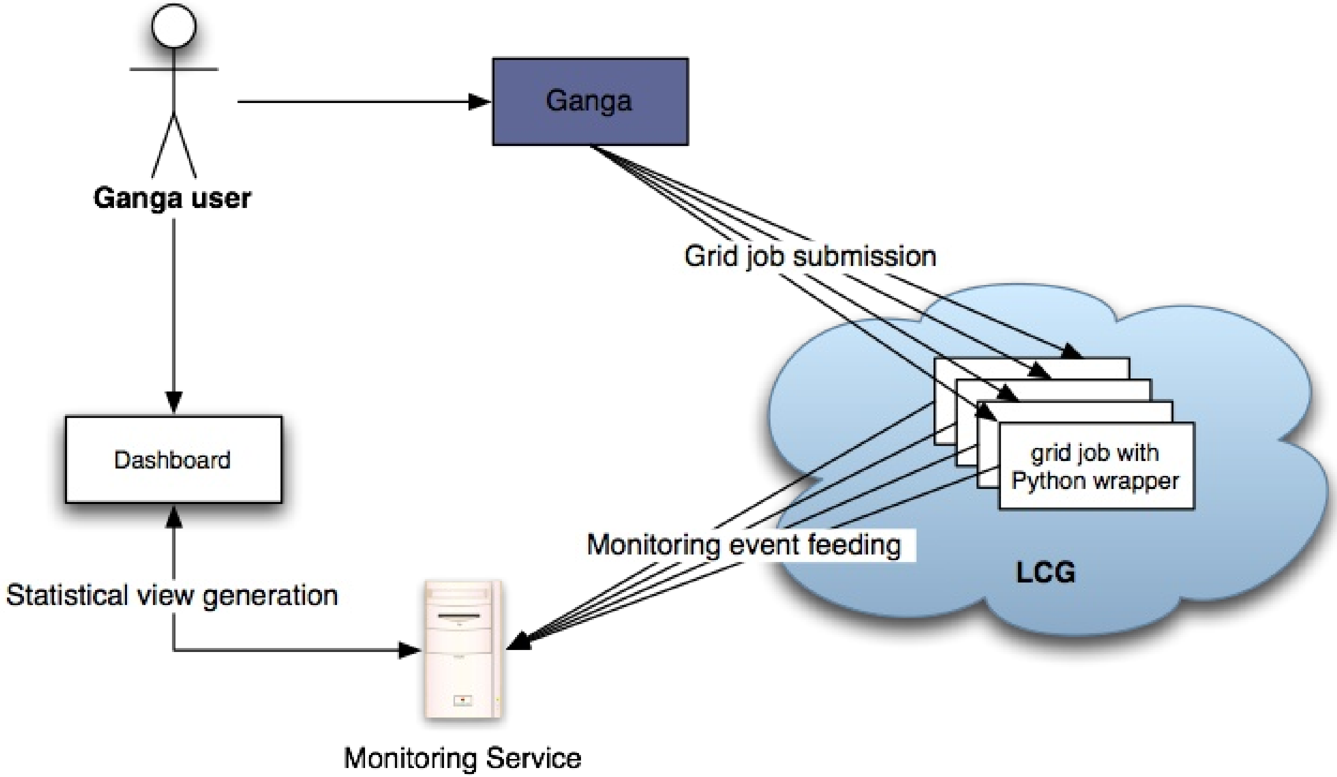
# Exercise: Prime factorization

- **Please follow the instructions from:**

https://twiki.cern.ch/twiki/bin/view/ArdaGrid/EGEETutorialPackage#Exercise_Prime_number_factorizat

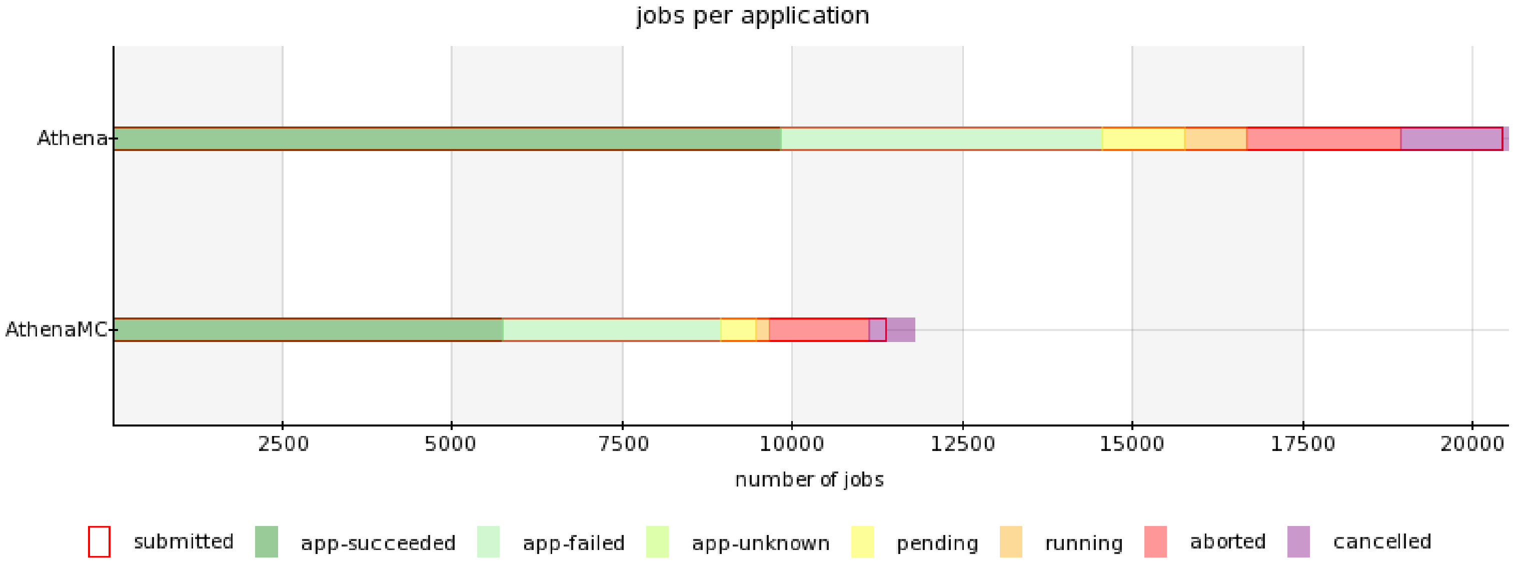
# Part III: Advanced use-cases

# Monitoring & Accounting



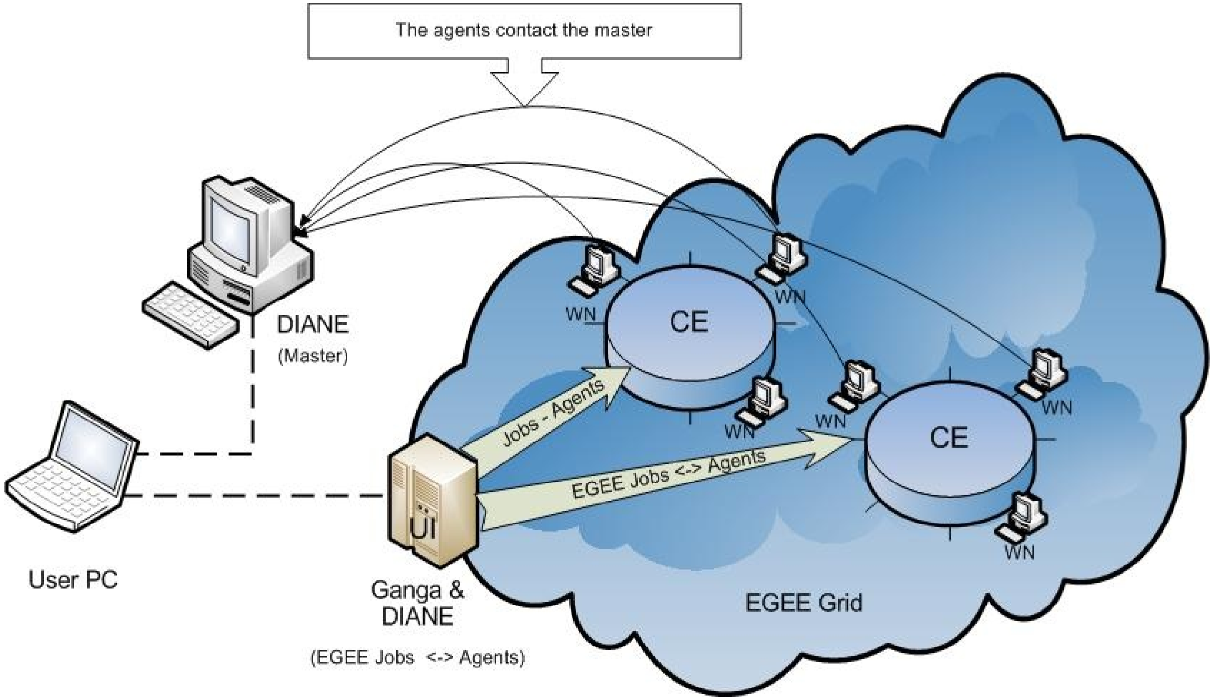
# Dashboard monitoring

- 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



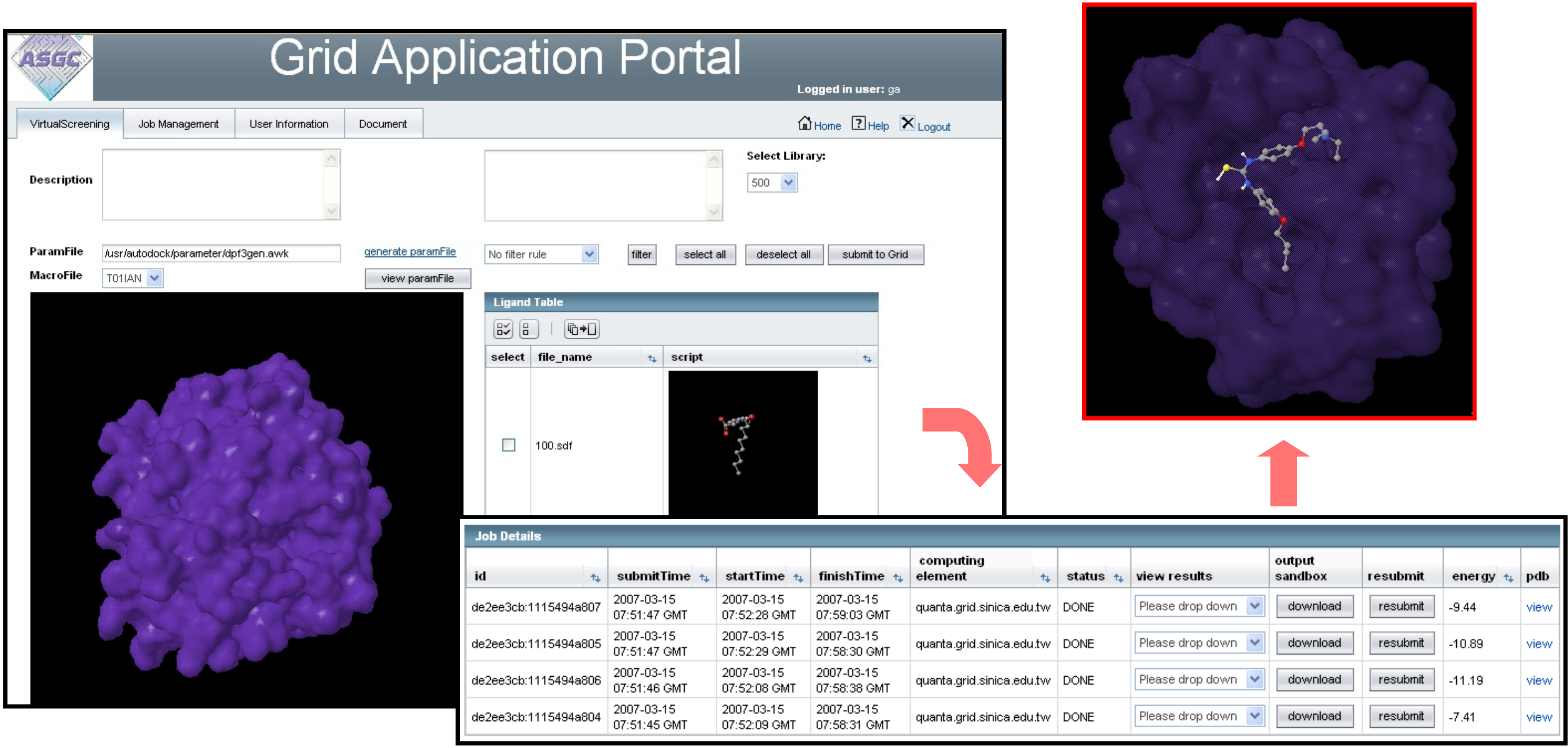
# Integration with frameworks

200K short jobs in 12 hours (500CPUh)



- Reduced Grid overhead
- Efficient workload balancing
- Fast error-recovery
- Automatic execution
- Interactivity

# Web portal for biologists



Grid Application Portal

Logged in user: ga

Description

Select Library: 500

ParamFile: /usr/autodock/parameter/dpf3gen.awk

MacroFile: T01IAN

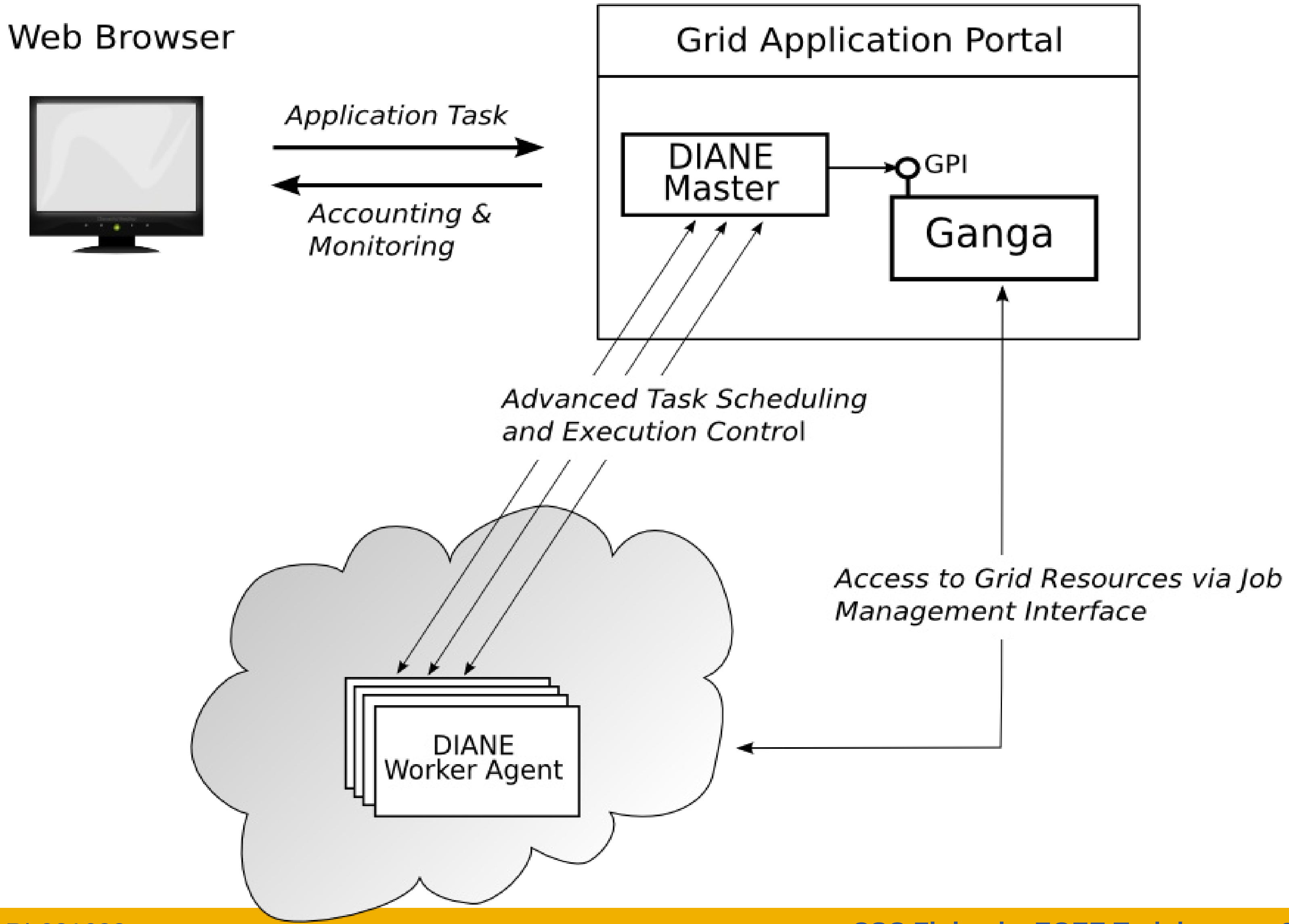
Ligand Table

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

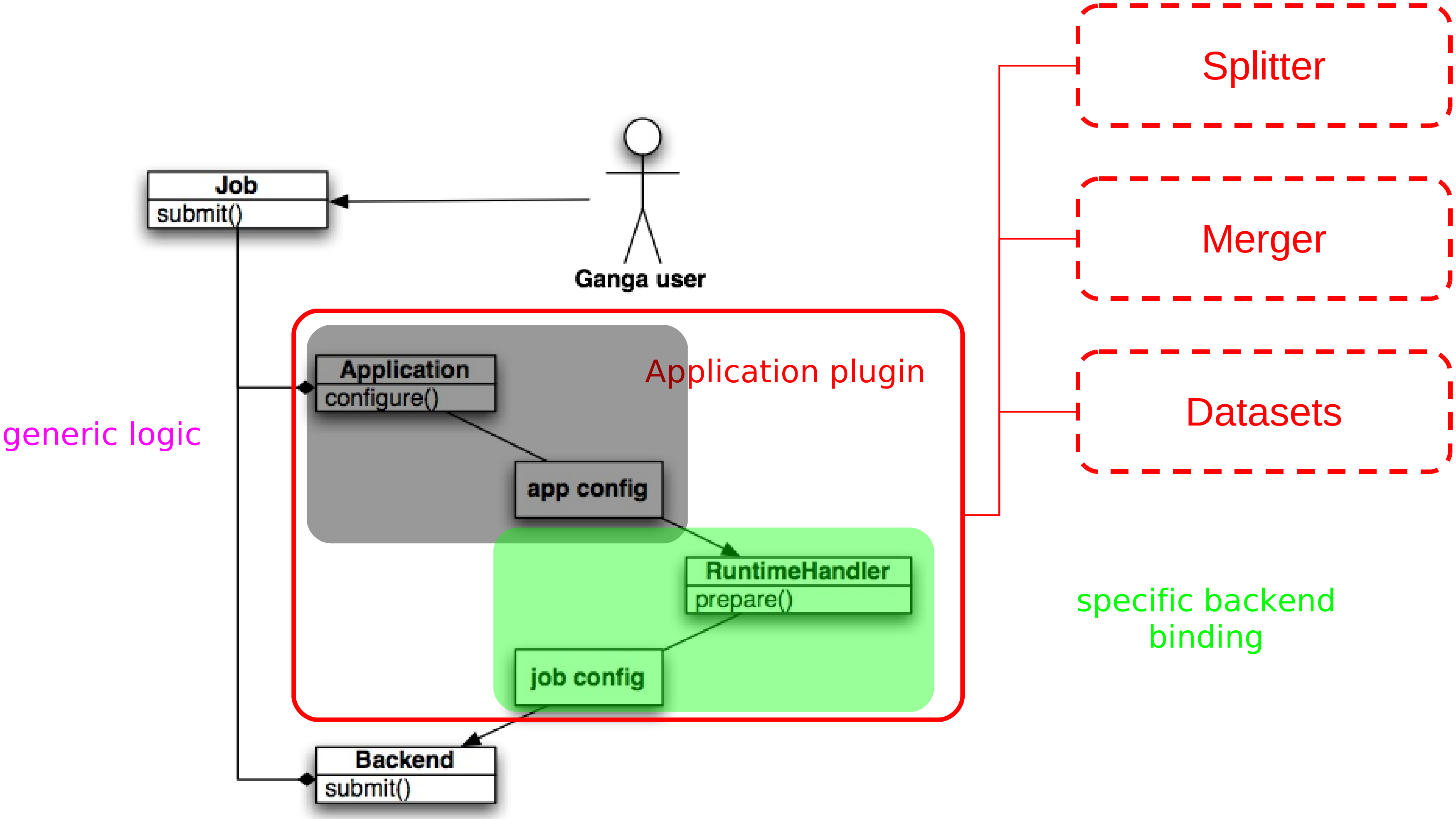
Job Details

| 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


Web Browser
Application Task
Accounting & Monitoring
Grid Application Portal
DIANE Master
GPI
Ganga
Advanced Task Scheduling and Execution Control
Access to Grid Resources via Job Management Interface
DIANE Worker Agent

# Extending Ganga



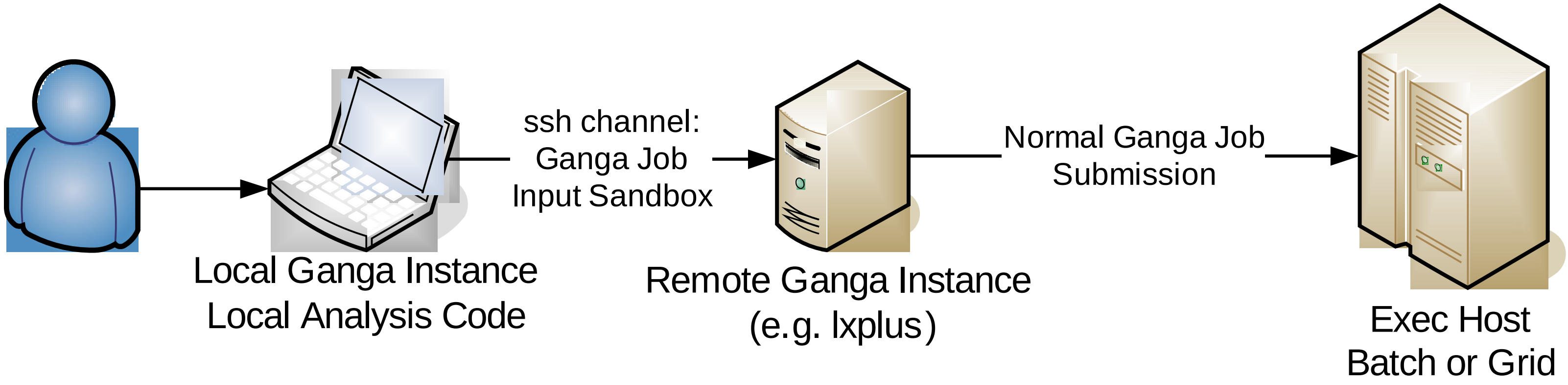
Detailed guide available at:

https://twiki.cern.ch/twiki/bin/view/ArdaGrid/HowToPlugInNewApplicationType

A good example: python/GangaTutorial/Lib/*.py

# Remote backend

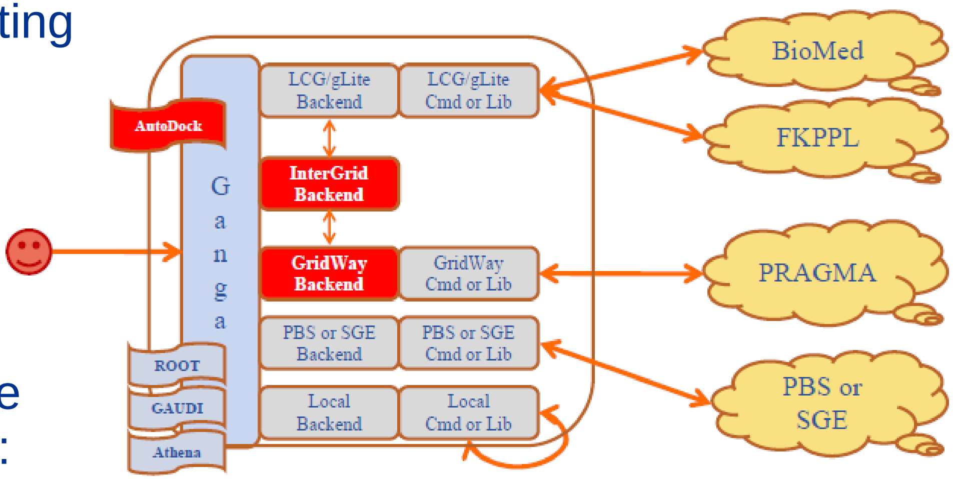
- Remote backend



- ***Remote*** **allows users to submit jobs from the anywhere, even if you don’t have the batch/grid client tools installed:**
  - Local ganga packages the input sandbox, connects to a remote ganga instance via ssh
  - The remote instance actually submits and monitors via the real backend (e.g. LCG/Dirac/Panda)
  - Useful for example to work around quota issues etc…

# GridWay and InterGrid backends

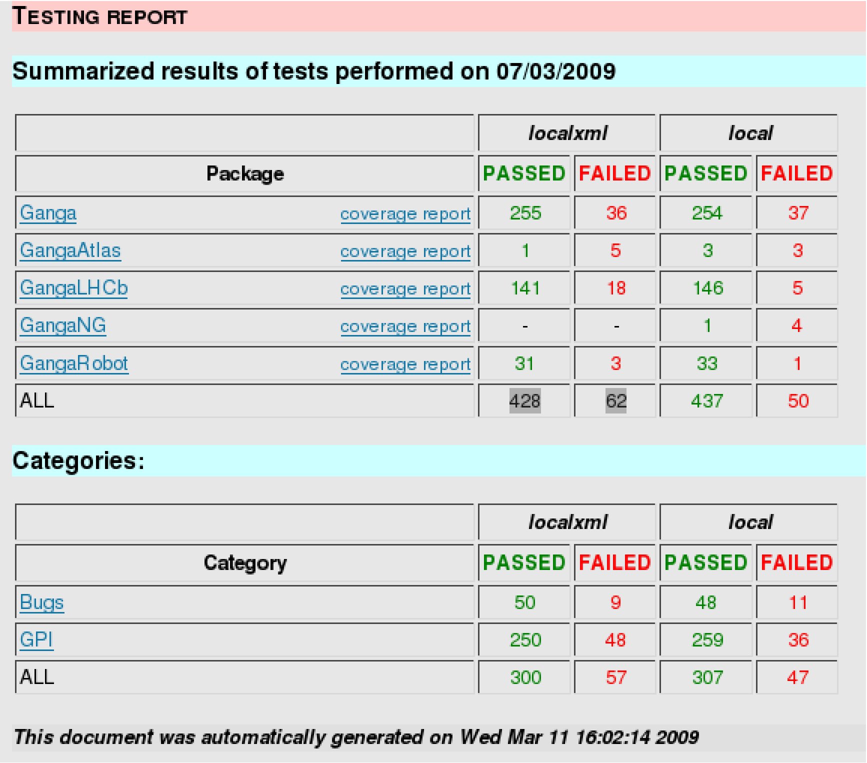
- **There are novel contributions coming from outside the core Ganga team:**
  - KISTI (Korea Institute of Science and Technology Information) and the WISDOM project
    - Continued effort to use grids for Avian flu, Malaria drug searches
  - Some of their resources could already be reached by the existing LCG/gLite backend.
  - But, other resources are Globus sites, managed by GridWay.
    - No GridWay backend, so they developed one.
  - Further, they wanted to hide the backend details from the users:
    - InterGrid backend: selects between LCG and GridWay backends using load information.



  - See Soonwook Hwang et al. “An Approach to Grid Interoperability using Ganga”. EGEE User Forum, Catania, March 2-6 2009

# Testing Framework & Releases

- **One of the strengths of Ganga is the stable release procedure and extensive testing**
- **Release Procedure:**
  - Ganga developers rotate through 6 week terms as release manager
  - Release manager’s job is quite easy, most of the process is automated.
  - When enough tags have been collected, a pre-release is created
- **Testing Framework:**
  - Each pre-release is validated with nearly 500 test cases
  - Tests of new features and to prevent regressions
  - Ideally, each bugfix gets a test case.



TESTING REPORT

**Summarized results of tests performed on 07/03/2009**

| | *localxml* | | *local* | |
|---|---|---|---|---|
| **Package** | **PASSED** | **FAILED** | **PASSED** | **FAILED** |
| Ganga coverage report | 255 | 36 | 254 | 37 |
| GangaAtlas coverage report | 1 | 5 | 3 | 3 |
| GangaLHCb coverage report | 141 | 18 | 146 | 5 |
| GangaNG coverage report | - | - | 1 | 4 |
| GangaRobot coverage report | 31 | 3 | 33 | 1 |
| ALL | 428 | 62 | 437 | 50 |

**Categories:**

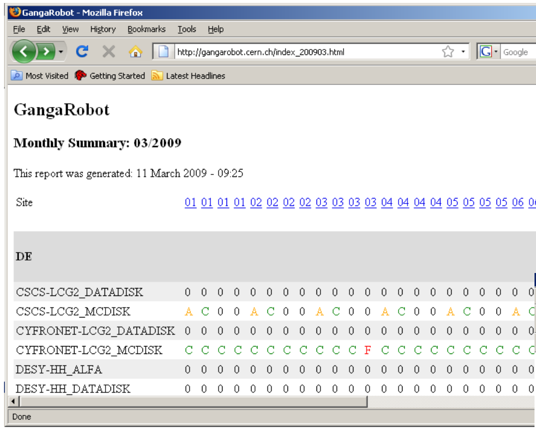
| | *localxml* | | *local* | |
|---|---|---|---|---|
| **Category** | **PASSED** | **FAILED** | **PASSED** | **FAILED** |
| Bugs | 50 | 9 | 48 | 11 |
| GPI | 250 | 48 | 259 | 36 |
| ALL | 300 | 57 | 307 | 47 |

*This document was automatically generated on Wed Mar 11 16:02:14 2009*

Failures usually due to timeouts

# GangaRobot and HammerCloud

- **Ganga is not only useful for end users:**
  - The Python API allows development of Grid applications
- **Two examples come from the ATLAS Distributed Analysis Testing tools.**
- **GangaRobot is used to run many short functional tests daily to continuously validate the DA workflows.**
  - Results are fed into Ganga so that broken sites can be avoided.
- **HammerCloud is used to run large stress tests:**
  - Measuring the behaviour of the storage, networks, databases, etc… under load.
- **See talk "Functional and Large-Scale Testing of the ATLAS Distributed Analysis Facilities with Ganga" 26-Mar-2009 @ 15:20**



GangaRobot - Mozilla Firefox
File Edit View History Bookmarks Tools Help
Most Visited Getting Started Latest Headlines
Done

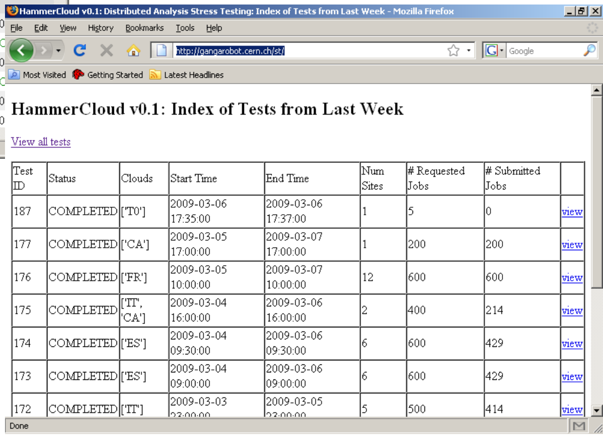
http://gangarobot.cern.ch/index_200903.html

GangaRobot

**Monthly Summary: 03/2009**

This report was generated: 11 March 2009 - 09:25

| Site | 01 | 01 | 01 | 01 | 02 | 02 | 02 | 02 | 03 | 03 | 03 | 03 | 04 | 04 | 04 | 04 | 05 | 05 | 05 | 05 | 06 | 06 | 06 | 06 | 0 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **DE** | | | | | | | | | | | | | | | | | | | | | | | | | |
| CSCS-LCG2_DATADISK | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | | | |
| CSCS-LCG2_MCDISK | A | C | 0 | 0 | A | C | 0 | 0 | A | C | 0 | 0 | A | C | 0 | 0 | A | C | 0 | 0 | A | C | | | |
| CYFRONET-LCG2_DATADISK | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | | | |
| CYFRONET-LCG2_MCDISK | C | C | C | C | C | C | C | C | C | C | C | F | C | C | C | C | C | C | C | C | C | C | | | |
| DESY-HH_ALFA | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | | | |
| DESY-HH_DATADISK | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | | | |



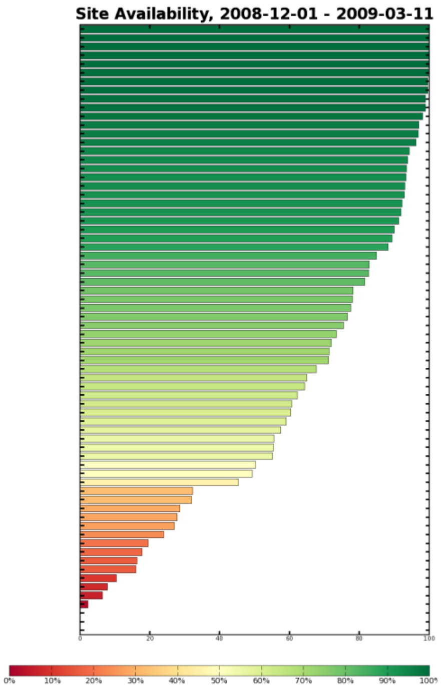
HammerCloud v0.1: Distributed Analysis Stress Testing: Index of Tests from Last Week - Mozilla Firefox
File Edit View History Bookmarks Tools Help
Most Visited Getting Started Latest Headlines
Done

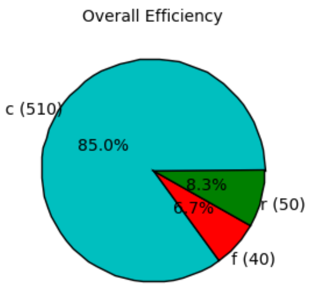
http://gangarobot.cern.ch/st/

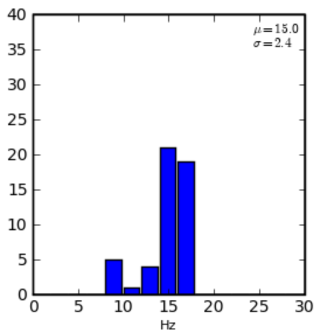
HammerCloud v0.1: Index of Tests from Last Week

View all tests

| Test ID | Status | Clouds | Start Time | End Time | Num Sites | # Requested Jobs | # Submitted Jobs | |
|---|---|---|---|---|---|---|---|---|
| 187 | COMPLETED | ['T0'] | 2009-03-06 17:35:00 | 2009-03-06 17:37:00 | 1 | 5 | 0 | view |
| 177 | COMPLETED | ['CA'] | 2009-03-05 17:00:00 | 2009-03-07 17:00:00 | 1 | 200 | 200 | view |
| 176 | COMPLETED | ['FR'] | 2009-03-05 10:00:00 | 2009-03-07 10:00:00 | 12 | 600 | 600 | view |
| 175 | COMPLETED | ['IT', 'CA'] | 2009-03-04 16:00:00 | 2009-03-06 16:00:00 | 2 | 400 | 214 | view |
| 174 | COMPLETED | ['ES'] | 2009-03-04 09:30:00 | 2009-03-06 09:30:00 | 6 | 600 | 429 | view |
| 173 | COMPLETED | ['ES'] | 2009-03-04 09:00:00 | 2009-03-06 09:00:00 | 6 | 600 | 429 | view |
| 172 | COMPLETED | ['IT'] | 2009-03-03 23:00:00 | 2009-03-05 23:00:00 | 5 | 500 | 414 | view |







http://cern.ch/ganga