



**GridPP**  
UK Computing for Particle Physics

---

# ***Ganga: A Grid User Interface***

***GridPP Collaboration Meeting  
1<sup>st</sup> - 3<sup>rd</sup> September, 2008  
Mark Slater, Birmingham University***

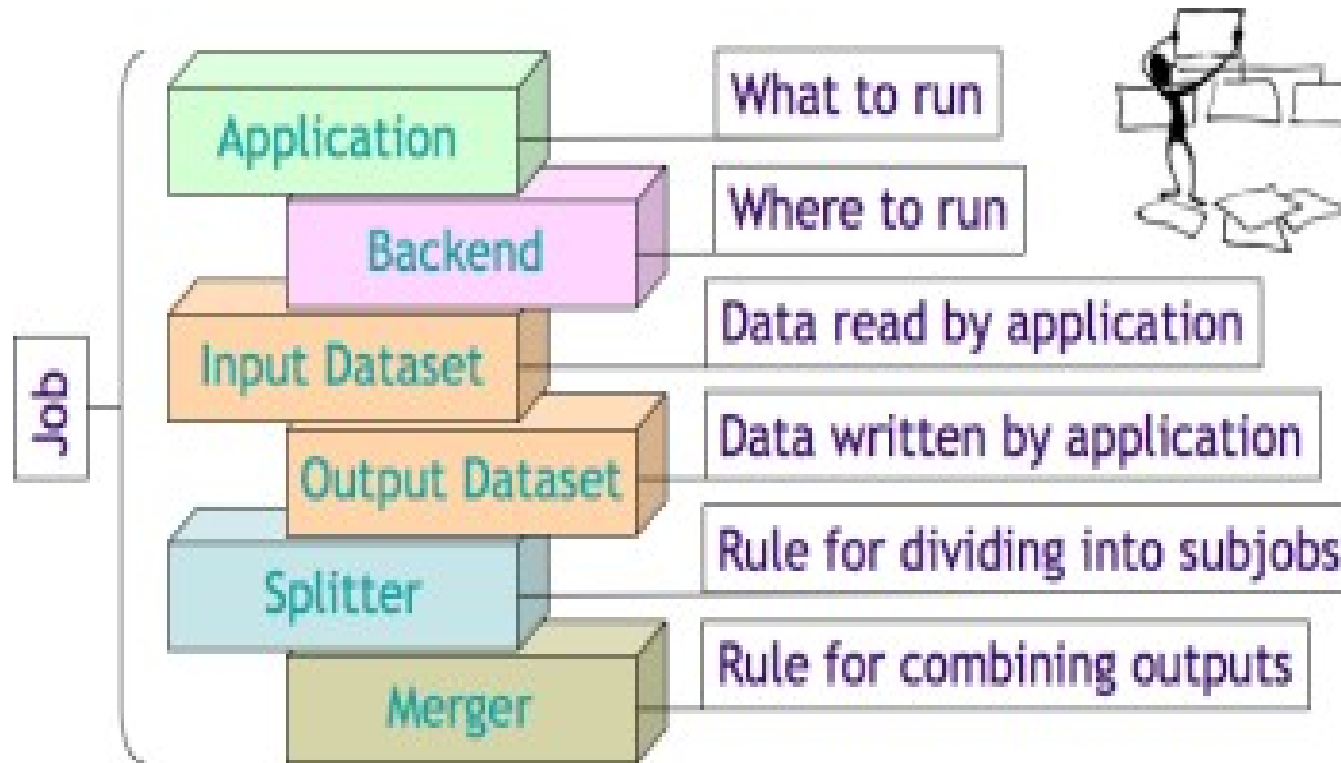
# Ganga Overview



GridPP  
UK Computing for Particle Physics

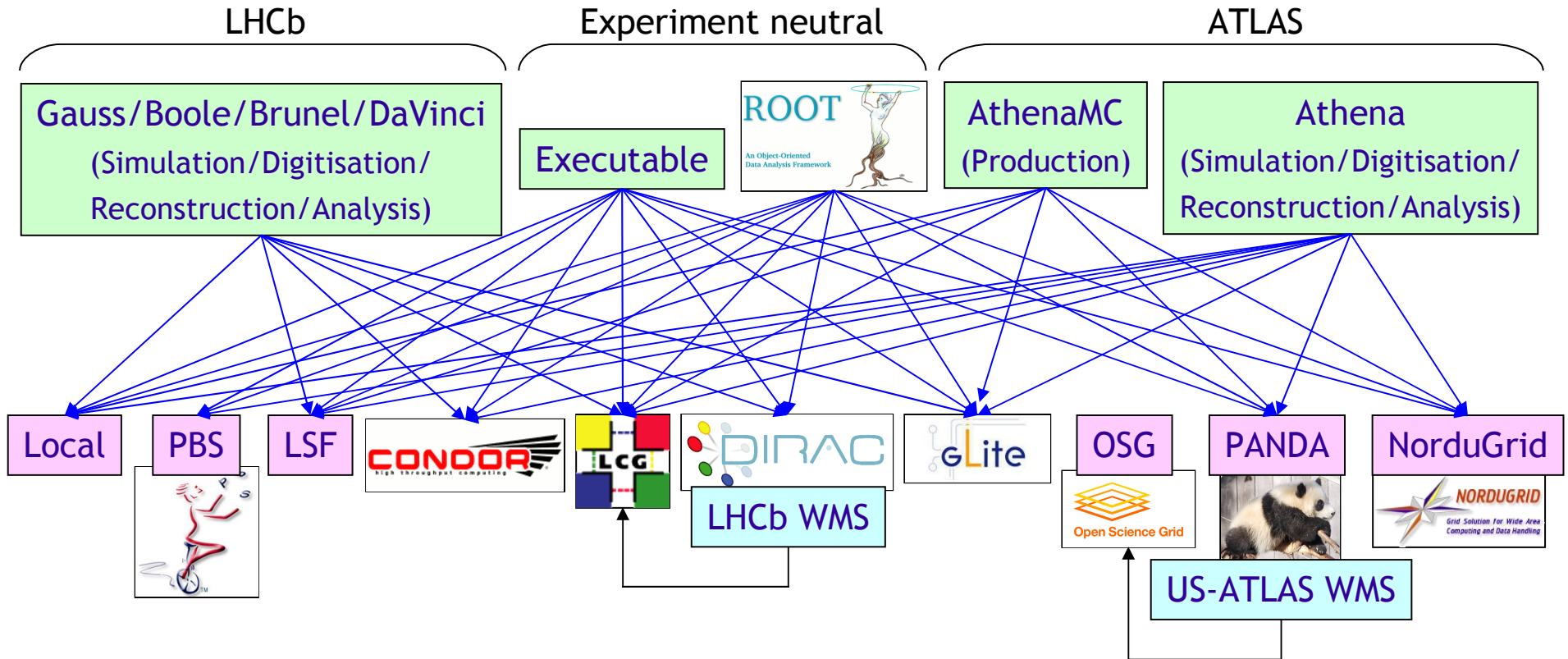


- Ganga is a user interface that **simplifies the submission of jobs** to both the Grid and other batch systems
- It is built on the idea that a job can be broken down into **independent modules** that can be any one of several types of **plugin**:

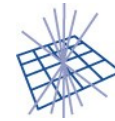


# Ganga Overview

- To give an idea of the power of this approach, here are listed the current combinations of just the possible **applications and backends**
- This doesn't include the possible **input/output data types and job splitting/merging** that is also possible!



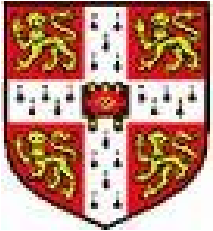
# Ganga Developers



**GridPP**  
UK Computing for Particle Physics



- Ganga is a truly international project, but with a very significant contribution from the UK:**



**University of Birmingham**  
**Karl Harrison, Mark Slater, Chun Lik Tan**



**University of Cambridge**  
**Frederic Brochu**



**CERN**  
**Hung-Chun Lee, Andrew Maier, Jakub Moscicki, Daniel Colin Van Der Ster**



**University of Edinburgh**  
**Greig Cowan**



LUDWIG-  
MAXIMILIANS-  
UNIVERSITÄT  
MÜNCHEN

**Imperial College London**  
**Ulrik Egede, Will Reece, (Mike Williams)**

**Imperial College  
London**

**Ludwig Maximilian University of Munich**  
**Johannes Elmsheuser**

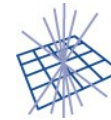


**University of Oslo**  
**Katarina Pajchel, Bjoern Hallvard Samset**



**University of Oxford**  
**Aliaksandr Saroka**

# Using Ganga



GridPP  
UK Computing for Particle Physics



- Jobs can be submitted to Ganga either through the **command line, IPython interface or GUI**
- However it is submitted though, **a job is broken down into the parts described previously**
- Therefore any job, **regardless of complexity, can be defined clearly and simply in a few lines of code**

## 1) Create the job

```
j = Job()
```

## 2) Setup App

```
j.application=Athena()  
j.application.exclude_from_user_area=["*.o","*.root*","*.exe"]  
j.application.prepare(athena_compile=False)  
j.application.option_file=['AnalysisSkeleton_topOptions.py']
```

## 3) Define input data

```
j.inputdata=DQ2Dataset()  
j.inputdata.dataset="fdr08_run2.0052283.physics_Muon.merge.AOD  
.o3_f8_m10"
```

## 4) Define output data

```
j.outputdata=DQ2OutputDataset()  
j.outputdata.outputdata=['AnalysisSkeleton.aan.root']
```

## 5) split into subjobs

```
j.splitter=DQ2JobSplitter()  
j.splitter.numfiles=3
```

## 6) Merge output

```
j.merger=AthenaOutputMerger()
```

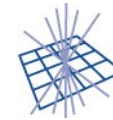
## 7) Where to run

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

## 8) Submit!

```
j.submit()
```

# Using Ganga



GridPP  
UK Computing for Particle Physics



- After submission to the specified backend, jobs are **tracked and updated** and job status can be checked in both **IPython and the GUI**:

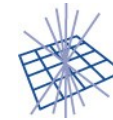
The screenshot shows the Ganga GUI interface. On the left, a table lists various jobs with columns for ID, status, name, application, exe filename, and backend. On the right, a 'Job Details' panel shows the configuration for a specific job.

id	status	name	application	exe filename	backend
6	completed		Executable	close.sh	LOG
7	new		Executable	close.sh	LOG
8	new		Executable	close.sh	LOG
9	completed		Executable	close.sh	LOG
10	completed		Executable	close.sh	LOG
12	completed		Athena	"TVA"	LSF
13	new		Executable	echo	Local
14	completed		Athena	"TVA"	LSF
15	completed		Athena	"TVA"	LOG
16	completed		Athena	"TVA"	LSF
17	completed	athena_log_dg2_test1_GRIDKA_DG2_LOCAL_7233	Athena	"TVA"	LOG
18	new	athena_log_dg2_test2_LYON_DG2_LOCAL_5300	Athena	"TVA"	LOG
19	completed	athena_log_dg2_test2_LYON_DG2_LOCAL_5300	Athena	"TVA"	LOG
20	completed		Athena	"TVA"	LSF
23	submitted	5320	Athena	"TVA"	LOG
24	submitted	5320	Athena	"TVA"	LOG
25	completed	5145_short	Athena	"TVA"	LOG
27	completed	5300	Athena	"TVA"	LOG
28	new	5300	Athena	"TVA"	Local
29	completed	5300	Athena	"TVA"	LOG
30	new	5145	Athena	"TVA"	LOG
31	completed	5145_short	Athena	"TVA"	LOG
32	running	5300	Athena	"TVA"	LOG
34	submitted	5320_seri	Athena	"TVA"	LOG

```
Job {
  status = 'submitted',
  name = '5320_seri',
  inputdir = '/afs/cern.ch/user/e/elmehaus/gangad/worksapce/Local/5320/Inp4',
  outputdir = '/afs/cern.ch/user/e/elmehaus/gangad/worksapce/Local/5320/Outp',
  outputandbox = [],
  id = 34,
  inputdata = DG2Dataset(
    tagdataset = '',
    type = 'DG2_LOCAL',
    names = [],
    match_ce = False,
    dataset = 'tsc11.005320.PythiaH170wvll.recon.ACD.v1.1004107'
  ),
  merger = None,
  inputandbox = [],
  application = Athena,
  atlas_release = '11.0.5',
  max_events = None,
  options = None,
  user_setupfile = File(
    name = '',
    subdir = ''
  ),
  option_file = File(
    name = '/afs/cern.ch/user/e/elmehaus/athena/testarea/11.0.5/Physics/Anz',
    subdir = ''
  ),
  user_area = File(
    name = '/afs/cern.ch/user/e/elmehaus/gangad/worksapce/Local/UserAn',
    subdir = ''
  ),
  outputdata = DG2OutputDataset(
    output = [],
    datasetname = ''
  )
}
```

- Importantly, all details of **submission and data retrieval** for the Grid (or other backends) are **handled by Ganga**

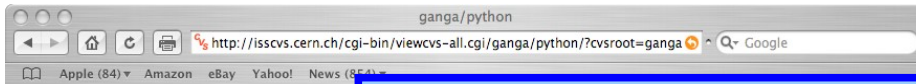
# Release Procedure



GridPP  
UK Computing for Particle Physics



- Ganga is being actively maintained and has a well-developed release procedure



(1) Code committed to CVS by package maintainer(s)

ganga/python

Current directory: [ganga] / ganga / python

File

- Ganga/
- GangaAtlas/
- GangaCamtology/
- GangaCronus/
- GangaExtraMon/
- GangaGUI/
- GangaLHCb/
- GangaNG/
- GangaPanda/
- GangaPlotter/
- GangaTest/
- GangaTutorial/

Testing report

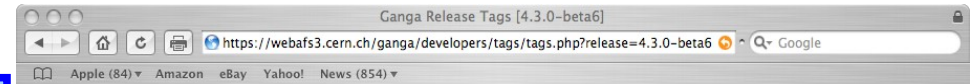
Summarized results for tests performed on 15/03/2007

Package	local	
	PASSED	FAILED
Ganga	93	12
GangaAtlas	1	1
GangaLHCb	3	7
ALL	97	20

Categories:

Category	local	
	PASSED	FAILED
Bugs	34	9
GPT	60	9

(3) Suite of over 100 tests is run, with results available on web



(2) Packages are tagged for release, with changes documented

```
lxlplus2041 ~ > ganga-release 4.2.10
CONFIGURATION:
  platf      = slc3_gcc323
  installdir = /afs/cern.ch/sw/ganga/install/slc3_gcc323
  top        = /afs/cern.ch/sw/ganga
  tag        = Ganga-4-2-10
  tagsdir_tcp = /afs/cern.ch/sw/ganga/www/developers/tags
  version    = 4.2.10
  configdir  = /afs/cern.ch/user/m/noscicki/LHCB/ganga/release/tools/./config
  tooledir   = /afs/cern.ch/user/m/noscicki/LHCB/ganga/release/tools
  releasenotesdir = /afs/cern.ch/user/m/noscicki/LHCB/ganga/release/tools/./
  reldir     = /afs/cern.ch/sw/ganga/install/slc3_gcc323/4.2.10
  cveroot    = :kserver:isssvcs.cern.ch:/local/rep/ganga
  tmpdir     = /afs/cern.ch/sw/ganga/install/tmp/slc3_gcc323/4.2.10
  tagsdir    = /afs/cern.ch/sw/ganga/www/developers/tags/data/4.2.10

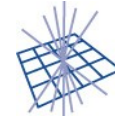
COMMANDS:
  0 - Show current configuration
  1 - COLLECT: collect tags from the tagsdir into PACKAGES file and create the new release notes file
  2 - PRE: create the PRE-RELEASE in the tmpdir
  3 - TEST: run the test suite in the pre-release tmpdir
  4 - TAG: make a GLOBAL TAG using the a pre-release in the tmpdir
  5 - RELEASE: make a release based on a previously created GLOBAL TAG and create the distribution kits
  6 - PREPARE NEXT development cycle: prepare tags dir for the next releases
  7 - Quit

Enter go
```

(4) Release made, using purpose-written tool

- There is generally a new release every two weeks allowing the developers to react very quickly to bugs found by the community

# Release Procedure



**GridPP**  
UK Computing for Particle Physics



- *With the release of Ganga version 5, the release procedure has received a significant overhaul*
- *The most important change has been the introduction of a 'round-robin' assignment of release manager responsibilities*
- *These responsibilities include:*
  - *Creation of pre-releases and releases*
  - *Running of the automated testing framework*
  - *Maintenance of the Savannah bug database*
- *This allows everyone in the development team to obtain a good understanding of the whole Ganga project*
- *In addition to this, there have been other updates:*
  - *Improved release scripting*
  - *Migrated to a fresh CVS repository*

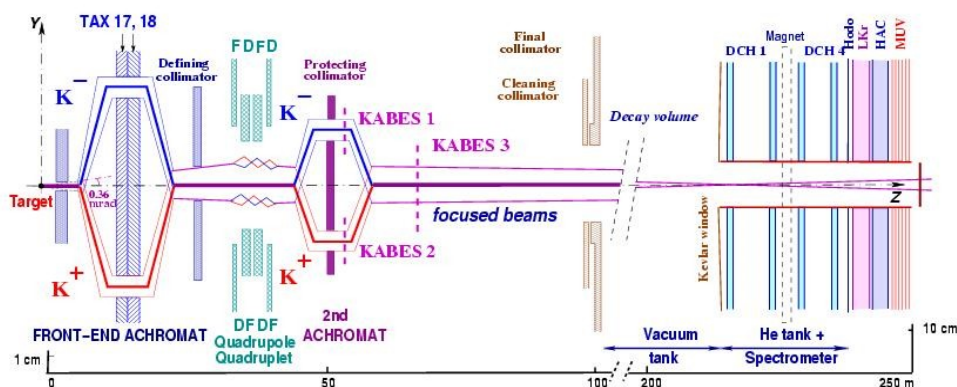
# Projects involving Ganga



GridPP  
UK Computing for Particle Physics

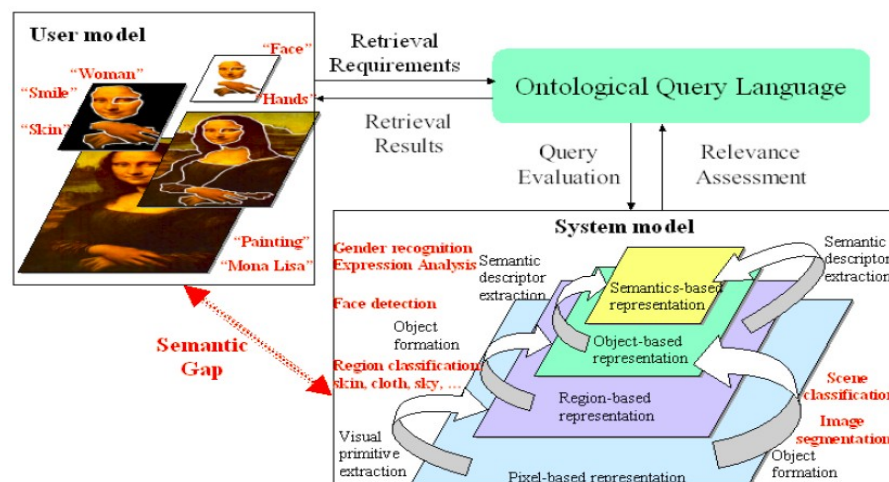


- The main users are the **Atlas** and **LHCb** experiments where **Ganga** is one of the **principle ways** of submitting jobs to the Grid
- However, other users outside **Atlas** and **LHCb** are also using **Ganga** for grid submission and job management, e.g. **Cambridge Ontology** and **NA48**

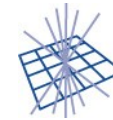


- **NA48** is a **fixed target experiment** at **CERN** looking at **CP violation**
- They are currently looking at using **Ganga** to manage their **Monte Carlo production** at on both **local clusters** and the **Grid**

- **Imense Ltd.** is developing solutions for **image retrieval based on image content**
- **Ganga** is being used to manage the jobs that **process the images** on the grid and **control web spiders** that find more images
- Try it at <http://imense.org>



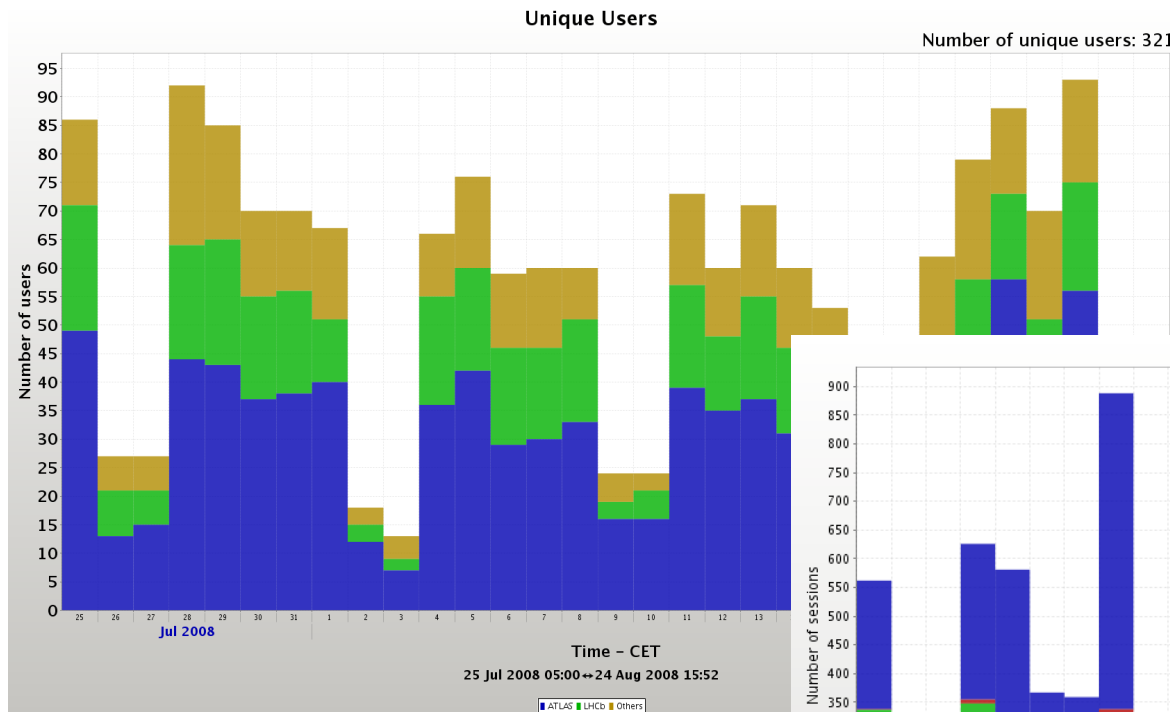
# Usage Statistics



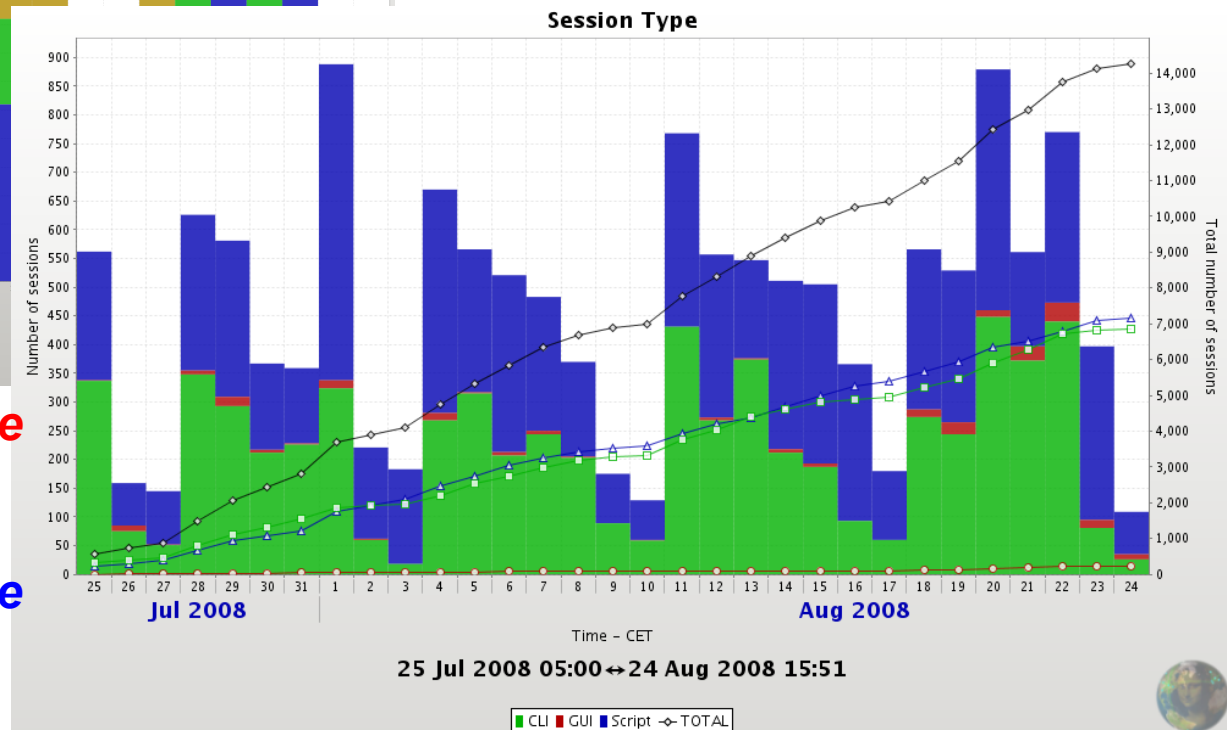
GridPP  
UK Computing for Particle Physics



- There is a healthy and active user base that has been growing steadily for the last 12 months



- There have been 1394 unique users in the last year
- This is up from 842 during the previous year

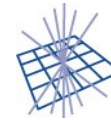


# Core Updates



- **Ganga Version 5 was released in the last few months**
- **This updated many things 'under-the-hood' as well as in the user interface:**
  - ➔ **The job repository system was optimised**
  - ➔ **Type checking added**
  - ➔ **Improved configuration system**
  - ➔ **Slicing on jobs was introduced**
  - ➔ **Summary printing on subjobs**
- **In addition to these changes, the release procedure was improved**
- **These improvements were made to stabilise the core framework and to ease the development cycle**
- **They also greatly improve the user interface, allowing quicker and easier job management**

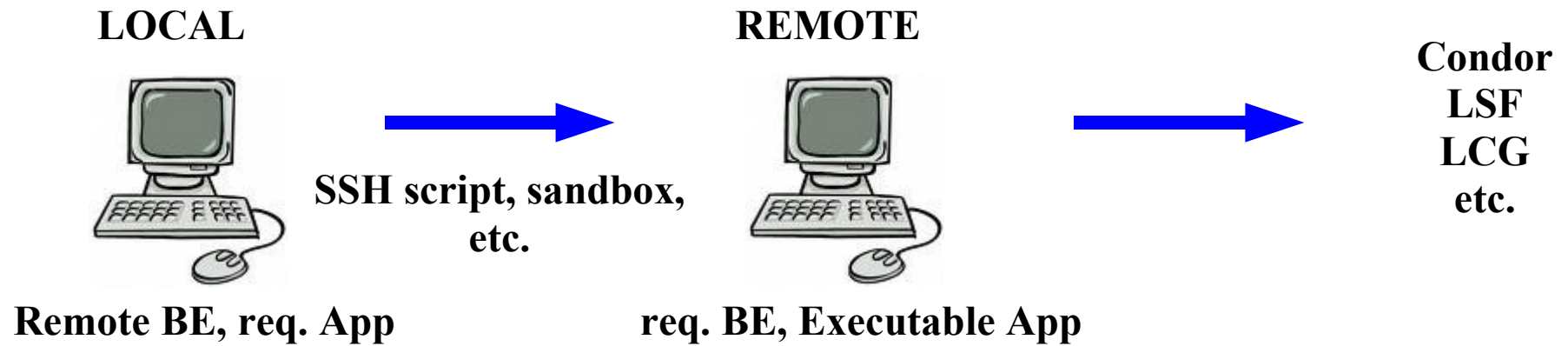
# Remote Backend



**GridPP**  
UK Computing for Particle Physics



- One of the **major core updates** with **Ganga Version 5** was the introduction of an **additional backend**
- This allows the user to submit jobs **to a remote machine from a local machine** using any backend or application available
- This allows, for example, jobs to be submitted:
  - ➔ To **LxPlus (CERN)** **from a home institute**
  - ➔ To the **Grid** from a computer **without a UI installed** via one that has
- It also allows all these separate jobs running at different places to be tracked in one place



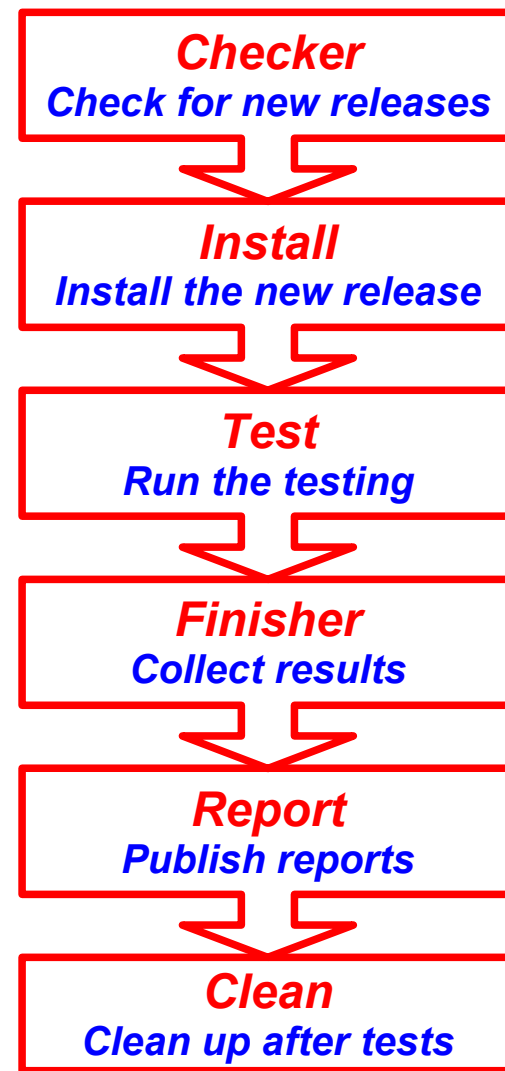
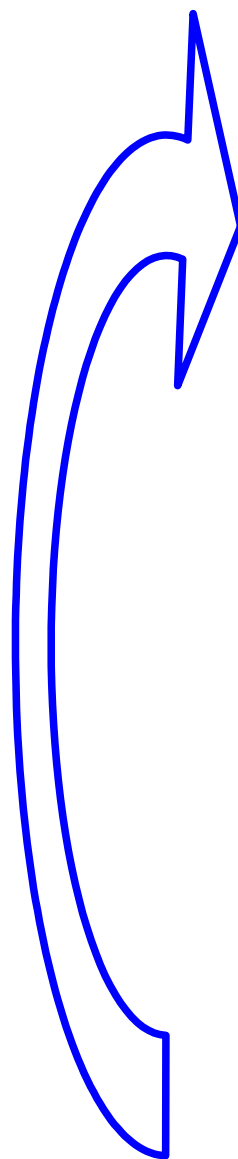
# Testing Outside CERN



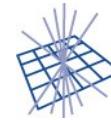
**GridPP**  
UK Computing for Particle Physics



- At present, testing of new releases is **only done at CERN** during the release process
- Alex Shires (a GridPP funded summer student) has developed a **remote testing framework** to run at sites other CERN
- This will allow testing of **non-CERN centric issues and installation**
- It takes advantage of the **Ganga Robot** that allows a complex set of actions to be run in sequence within the context of a Ganga session
- It integrates this with the current testing framework allowing the process to be **completely automated**
- Work is **nearly complete** and will very shortly be **integrated into the full release process**



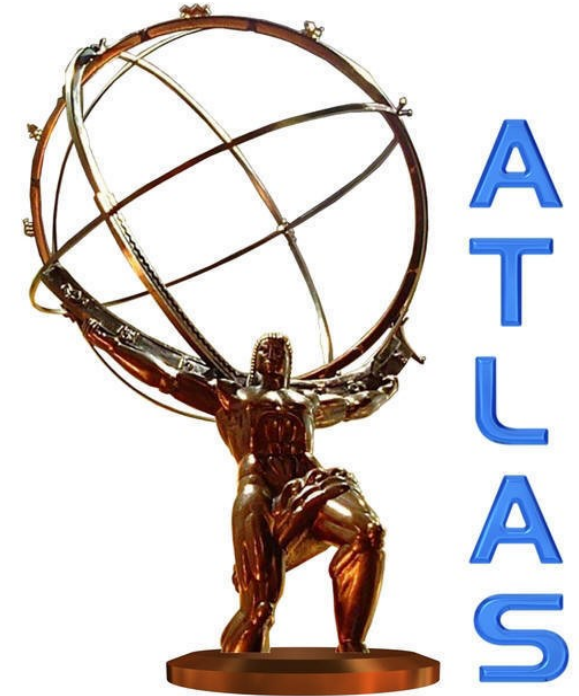
# Atlas Updates



**GridPP**  
UK Computing for Particle Physics



- **Recent updates to Atlas include:**
  - ➔ **Support for multiple DQ2 datasets for a job**
  - ➔ **Container Datasets**
  - ➔ **Better support for local datasets**
  - ➔ **AthenaRootAccess**
  - ➔ **Improved support for Space Tokens**
  - ➔ **Improved support for the Panda system (OSG)**
- **It is hoped future updates will include:**
  - ➔ **Improve CMT environment detection**
  - ➔ **Streamline Local/LCG code**
  - ➔ **Use DQ2Clients on the worker nodes**
  - ➔ **Improve TAG and GangaTnT**
  - ➔ **Improve User monitoring**
- **Ganga was used and reviewed as part of the UK involvement with the Full Dress Rehearsal (FDR) and the feedback was very positive**



# LHCb Updates



**GridPP**  
UK Computing for Particle Physics



- **Recent updates to LHCb include:**
  - ➔ **Full support for submission using jobs configured with options in Python**
  - ➔ **RAW data reading now possible**
  - ➔ **User controlled event loop now supported (GaudiPython)**
  - ➔ **Further integration with development environment for new Gaudi and LHCb releases**
  - ➔ **Support for both 32 and 64 architectures**
  - ➔ **Merging of DST files**
- **It is hoped future updates will include:**
  - ➔ **Upgrade to next version of DIRAC workload management tool**
  - ➔ **Support for new file structure in LHCb releases**
  - ➔ **Integration into quality control of data from the detector (using the Ganga Robot)**



# User Support

- There are several ways for users to get help with using Ganga

**Mailing Lists**

**If none of these solve the problem, bug reports can be posted that should be solved by the following release**

# Typical Support - Atlas

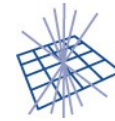


**GridPP**  
UK Computing for Particle Physics



- *As an example of the support available for the two major experiments, here is an overview of the **Atlas support offered** (LHCb is very similar)*
- *Support for Ganga Atlas is currently based around the following:*
  - ➔ ***Tutorial Document (recently updated for 5.0)***
  - ➔ ***Twiki based FAQ***
  - ➔ ***Hypernews for direct help***
- *The **FAQ** covers most of the important issues:*
  - ➔ ***Common errors***
  - ➔ ***Dealing with DQ2Datasets***
  - ➔ ***General work flow***
- *To coordinate both support issues and development, the dedicated Atlas developers meet briefly twice a week*
- *On average, **5-10 questions a week** are posted to the hypernews forum*
- *Questions are answered **within a day** and generally resolved **within 3-4 days***

# Tutorial Sessions



**GridPP**  
UK Computing for Particle Physics

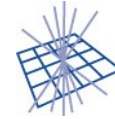


- *Ganga is widely advertised in the **analysis tutorial sessions** held regularly at CERN and elsewhere*
- *These are usually well subscribed with between **40-80 people attending***
- ***UK only analysis sessions** are also being run (e.g. Edinburgh, Birmingham)*
- *During these sessions, **specific Ganga experts** are usually present*



# Conclusions

---



**GridPP**  
UK Computing for Particle Physics



- *Ganga is becoming **more and more widely used** as a job submission and management tool*
- *It is used by **two of the major experiments** at the LHC as well as in other cases*
- *Thanks to a very active release schedule, we are able to **respond very quickly** to any bugs/feature requests posted by users*
- *The recent release of Ganga 5 has added **even more functionality** as well as **stabilising the framework and interface***
- *User support and documentation has been improved and continues to address the vast majority of the issues of the client base*
- *We are now ready to help **analyse the first data coming out of the LHC!***