Johannes Elmsheuser
Ludwig-Maximilians-Universität München, Germany
25 Sep 2006/LHCC comprehensive review
OUTLINE

1 ATLAS Distributed Analysis

2 Task: Access data in distributed environment

3 Introduction to the tool

4 How to prepare a job to run on the Grid

5 Example of real activity

6 Conclusions
1 ATLAS Distributed Analysis

2 Task: Access data in distributed environment

3 Introduction to the tool

4 How to prepare a job to run on the Grid

5 Example of real activity

6 Conclusions
ATLAS Distributed Analysis

- ATLAS data analysis foreseen on Tier2 sites of three GRIDs: LCG, Nordugrid, OSG
- Data from Production System is currently consolidated by DDM-operations team on 3 sites: CERN, Lyon, BNL
- LCG tool for distributed analysis: GANGA
- OSG production/analysis system tool: Panda

- ATLAS Analysis model implemented by PAT team foresees Athena analysis of AODs/ESDs and interactive use of Athena-aware-ROOT tuples
OUTLINE

1. ATLAS Distributed Analysis

2. Task: Access data in distributed environment

3. Introduction to the tool

4. How to prepare a job to run on the Grid

5. Example of real activity

6. Conclusions
Task description

- Plot different kinematic distributions of:
  - Signal: $H \rightarrow WW \rightarrow \mu\nu\mu\nu$ (PYTHIA, $m_H = 170$ GeV)
  - Main background: $Z/\gamma^* \rightarrow \mu\mu$ (PYTHIA, $60$ GeV < $m_{\mu\mu}$)

- Start from local working area with ATHENA code reading AODs and producing ROOT-tuples with essential information

- Datasets are located on the LCG-GRID
  - Wrap up working area, send to dataset location, produce output ROOT-tuples, retrieve output dataset ROOT-tuples
  - Most of these tasks are managed by: GANGA

- Finally use small ROOT-tuples to produces kinematic distributions
1 ATLAS Distributed Analysis

2 Task: Access data in distributed environment

3 Introduction to the tool

4 How to prepare a job to run on the Grid

5 Example of real activity

6 Conclusions
Ganga is a user tool for job definition and management

- Allows simple switching between testing on a local batch system and large-scale data processing on distributed resources (Grid)

Developed in the context of ATLAS and LHCb:
  - For ATLAS, have built-in support for applications based on Athena framework, for JobTransforms, and for DQ2 data-management system

- Component architecture readily allows extension

- Python framework

Development team: F. Brochu (Cambridge), U. Egede (Imperial), J. Elmsheuser (München), K. Harrison (Cambridge), H. C. Lee (ASCC), D. Liko (CERN), A. Maier (CERN), J. T. Moscicki (CERN), A. Muraru (Bucharest), V. Romanovsky (IHEP), A. Soroko (Oxford), C. L. Tan (Birmingham) and contributions past and present from many others
Ganga II

- Ganga is based on a simple, but flexible, job abstraction
- A job is constructed from a set of building blocks, not all required for every job
Ganga III

- Ganga simplifies running of ATLAS (and LHCb) applications on a variety of Grid and non-Grid backends
Outline

1. ATLAS Distributed Analysis
2. Task: Access data in distributed environment
3. Introduction to the tool
4. How to prepare a job to run on the Grid
5. Example of real activity
6. Conclusions
Integration with the Data Management

- Data is organized in datasets, list of files

- AMI Metadata:
  - query different production steps and generator options

- Dataset browser:

- Dashboard (still under construction):
  - DDM monitoring:
  - Job monitoring: http://arda-dashboard.cern.ch/atlas
**Job preparation**

- Declaring input data:
  - Ganga has DDM/DQ2 integration
  - Select dataset name and specify as input dataset parameter to Ganga
  - Either submit to specific site of dataset location or let Ganga do it automatically during submission

- Splitting parameter:
  - Define to number of parallel sub-jobs as input parameter to Ganga

- Declaring output data:
  - Specify names of output files
  - Stored on remote storage element and optional registered into DDM/DQ2

- Source code:
  - Working area is wrapped up together with jobOptions and re-compiled on remote GRID host
1 ATLAS Distributed Analysis

2 Task: Access data in distributed environment

3 Introduction to the tool

4 How to prepare a job to run on the Grid

5 Example of real activity

6 Conclusions
Job definition (I)

- Job definition from command line on local desktop:
  athena AnalysisSkeleton_jobOptions.py

- Job definition from command line to the GRID:
  ganga athena \
   --inDS csc11.005320.PythiaH170wwll.recon.AOD.v11004107 \ 
   --outputdata AnalysisSkeleton.aan.root \ 
   --split 3 \ 
   --ce ce106.cern.ch:2119/jobmanager-lcglsf-grid_2nh_atlas \ 
   AnalysisSkeleton_jobOptions.py
Job definition within GANJA IPython shell

```python
j = Job()
j.name='5145'
j.application=Athena()
j.application.prepare()
j.application.option_file='$HOME/athena/testarea/11.0.5/InstallArea,jobOptions/
j.splitter=AthenaSplitterJob()
j.splitter.numsubjobs = 10
j.inputdata=DQ2Dataset()
j.inputdata.type='DQ2_LOCAL'
j.inputdata.dataset='csc11.005145.PythiaZmumu.recon.AOD.v11004103'
j.outputdata=DQ2OutputDataset()
j.outputdata.outputdata=['AnalysisSkeleton.aan.root']
j.backend=LCG()
j.backend.CE='ce-fzk.gridka.de:2119/jobmanager-pbspro-atlasS'
j.submit()
```
Datasets

- csc11.005320.PythiaH170wwll.recon.AOD.v11004107
  - $H \rightarrow WW \rightarrow \mu\nu\mu\nu$, $m_H = 170$ GeV
  - PYTHIA 6.3, $\sigma = X$ pb
  - 19 files, 3 jobs

- csc11.005145.PythiaZmumu.recon.AOD.v11004103
  - $Z/\gamma^* \rightarrow \mu\mu$, $60$ GeV $< m_{\mu\mu}$
  - PYTHIA 6.3, $\sigma = 1497$ pb
  - 106 files, 10 jobs
Monitoring with the IPython shell

In [3]:
In [8]: (execfile('/afs/cern.ch/user/e/elsmeus/split/5320_cern.py'))

In [4]:
In [34]: jobs
Out[4]: Summary of jobs:

<table>
<thead>
<tr>
<th>id</th>
<th>status</th>
<th>name</th>
<th>application</th>
<th>backend</th>
<th>backend.actualCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>completed</td>
<td>athena_log_dg2_test1, GRIDER, DQS.Local, 7293</td>
<td>Athena</td>
<td>LSF</td>
<td>ce-fzk.egridal.de:2119/jobmanager-pbspro-atlas</td>
</tr>
<tr>
<td>2</td>
<td>new</td>
<td>athena_log_dg2_test2, GRIDER, DQS.Local, 5320</td>
<td>Athena</td>
<td>LSF</td>
<td>ce-fzk.egridal.de:2119/jobmanager-pbspro-atlas</td>
</tr>
<tr>
<td>3</td>
<td>completed</td>
<td>5185_short</td>
<td>Athena</td>
<td>LSF</td>
<td>ce-fzk.egridal.de:2119/jobmanager-pbspro-atlas</td>
</tr>
<tr>
<td>4</td>
<td>new</td>
<td>5320</td>
<td>Athena</td>
<td>LSF</td>
<td>ce-fzk.egridal.de:2119/jobmanager-pbspro-atlas</td>
</tr>
<tr>
<td>5</td>
<td>completed</td>
<td>5185_long</td>
<td>Athena</td>
<td>LSF</td>
<td>ce-fzk.egridal.de:2119/jobmanager-pbspro-atlas</td>
</tr>
<tr>
<td>6</td>
<td>new</td>
<td>5320</td>
<td>Athena</td>
<td>LSF</td>
<td>ce-fzk.egridal.de:2119/jobmanager-pbspro-atlas</td>
</tr>
<tr>
<td>7</td>
<td>completed</td>
<td>killed</td>
<td>Athena</td>
<td>LSF</td>
<td>ce-fzk.egridal.de:2119/jobmanager-pbspro-atlas</td>
</tr>
<tr>
<td>8</td>
<td>new</td>
<td>killed</td>
<td>Athena</td>
<td>LSF</td>
<td>ce-fzk.egridal.de:2119/jobmanager-pbspro-atlas</td>
</tr>
<tr>
<td>9</td>
<td>completed</td>
<td>submitted</td>
<td>Athena</td>
<td>LSF</td>
<td>ce-fzk.egridal.de:2119/jobmanager-pbspro-atlas</td>
</tr>
<tr>
<td>10</td>
<td>new</td>
<td>submitted</td>
<td>Athena</td>
<td>LSF</td>
<td>ce-fzk.egridal.de:2119/jobmanager-pbspro-atlas</td>
</tr>
<tr>
<td>11</td>
<td>completed</td>
<td>submitted</td>
<td>Athena</td>
<td>LSF</td>
<td>ce-fzk.egridal.de:2119/jobmanager-pbspro-atlas</td>
</tr>
<tr>
<td>12</td>
<td>new</td>
<td>submitted</td>
<td>Athena</td>
<td>LSF</td>
<td>ce-fzk.egridal.de:2119/jobmanager-pbspro-atlas</td>
</tr>
<tr>
<td>13</td>
<td>completed</td>
<td>submitted</td>
<td>Athena</td>
<td>LSF</td>
<td>ce-fzk.egridal.de:2119/jobmanager-pbspro-atlas</td>
</tr>
<tr>
<td>14</td>
<td>new</td>
<td>submitted</td>
<td>Athena</td>
<td>LSF</td>
<td>ce-fzk.egridal.de:2119/jobmanager-pbspro-atlas</td>
</tr>
<tr>
<td>15</td>
<td>completed</td>
<td>submitted</td>
<td>Athena</td>
<td>LSF</td>
<td>ce-fzk.egridal.de:2119/jobmanager-pbspro-atlas</td>
</tr>
<tr>
<td>16</td>
<td>new</td>
<td>submitted</td>
<td>Athena</td>
<td>LSF</td>
<td>ce-fzk.egridal.de:2119/jobmanager-pbspro-atlas</td>
</tr>
<tr>
<td>17</td>
<td>completed</td>
<td>submitted</td>
<td>Athena</td>
<td>LSF</td>
<td>ce-fzk.egridal.de:2119/jobmanager-pbspro-atlas</td>
</tr>
<tr>
<td>18</td>
<td>new</td>
<td>submitted</td>
<td>Athena</td>
<td>LSF</td>
<td>ce-fzk.egridal.de:2119/jobmanager-pbspro-atlas</td>
</tr>
<tr>
<td>19</td>
<td>completed</td>
<td>submitted</td>
<td>Athena</td>
<td>LSF</td>
<td>ce-fzk.egridal.de:2119/jobmanager-pbspro-atlas</td>
</tr>
<tr>
<td>20</td>
<td>new</td>
<td>submitted</td>
<td>Athena</td>
<td>LSF</td>
<td>ce-fzk.egridal.de:2119/jobmanager-pbspro-atlas</td>
</tr>
<tr>
<td>21</td>
<td>completed</td>
<td>submitted</td>
<td>Athena</td>
<td>LSF</td>
<td>ce-fzk.egridal.de:2119/jobmanager-pbspro-atlas</td>
</tr>
<tr>
<td>22</td>
<td>new</td>
<td>submitted</td>
<td>Athena</td>
<td>LSF</td>
<td>ce-fzk.egridal.de:2119/jobmanager-pbspro-atlas</td>
</tr>
<tr>
<td>23</td>
<td>completed</td>
<td>killed</td>
<td>Athena</td>
<td>LSF</td>
<td>ce-fzk.egridal.de:2119/jobmanager-pbspro-atlas</td>
</tr>
<tr>
<td>24</td>
<td>new</td>
<td>killed</td>
<td>Athena</td>
<td>LSF</td>
<td>ce-fzk.egridal.de:2119/jobmanager-pbspro-atlas</td>
</tr>
<tr>
<td>25</td>
<td>completed</td>
<td>killed</td>
<td>Athena</td>
<td>LSF</td>
<td>ce-fzk.egridal.de:2119/jobmanager-pbspro-atlas</td>
</tr>
<tr>
<td>26</td>
<td>new</td>
<td>killed</td>
<td>Athena</td>
<td>LSF</td>
<td>ce-fzk.egridal.de:2119/jobmanager-pbspro-atlas</td>
</tr>
<tr>
<td>27</td>
<td>completed</td>
<td>killed</td>
<td>Athena</td>
<td>LSF</td>
<td>ce-fzk.egridal.de:2119/jobmanager-pbspro-atlas</td>
</tr>
<tr>
<td>28</td>
<td>new</td>
<td>killed</td>
<td>Athena</td>
<td>LSF</td>
<td>ce-fzk.egridal.de:2119/jobmanager-pbspro-atlas</td>
</tr>
<tr>
<td>29</td>
<td>completed</td>
<td>killed</td>
<td>Athena</td>
<td>LSF</td>
<td>ce-fzk.egridal.de:2119/jobmanager-pbspro-atlas</td>
</tr>
<tr>
<td>30</td>
<td>new</td>
<td>killed</td>
<td>Athena</td>
<td>LSF</td>
<td>ce-fzk.egridal.de:2119/jobmanager-pbspro-atlas</td>
</tr>
<tr>
<td>31</td>
<td>completed</td>
<td>killed</td>
<td>Athena</td>
<td>LSF</td>
<td>ce-fzk.egridal.de:2119/jobmanager-pbspro-atlas</td>
</tr>
<tr>
<td>32</td>
<td>new</td>
<td>killed</td>
<td>Athena</td>
<td>LSF</td>
<td>ce-fzk.egridal.de:2119/jobmanager-pbspro-atlas</td>
</tr>
<tr>
<td>33</td>
<td>completed</td>
<td>killed</td>
<td>Athena</td>
<td>LSF</td>
<td>ce-fzk.egridal.de:2119/jobmanager-pbspro-atlas</td>
</tr>
<tr>
<td>34</td>
<td>new</td>
<td>killed</td>
<td>Athena</td>
<td>LSF</td>
<td>ce-fzk.egridal.de:2119/jobmanager-pbspro-atlas</td>
</tr>
<tr>
<td>35</td>
<td>completed</td>
<td>killed</td>
<td>Athena</td>
<td>LSF</td>
<td>ce-fzk.egridal.de:2119/jobmanager-pbspro-atlas</td>
</tr>
</tbody>
</table>

# Monitoring with the GUI

![Monitoring GUI Screenshot](image_url)

<table>
<thead>
<tr>
<th>Job ID</th>
<th>Status</th>
<th>Name</th>
<th>Application</th>
<th>Exec Filename</th>
<th>Backend</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>completed</td>
<td>Executable CloseSSH</td>
<td>CloseSSH</td>
<td></td>
<td>LCG</td>
</tr>
<tr>
<td>2</td>
<td>new</td>
<td>Executable CloseSSH</td>
<td>CloseSSH</td>
<td></td>
<td>LCG</td>
</tr>
<tr>
<td>3</td>
<td>completed</td>
<td>Executable CloseSSH</td>
<td>CloseSSH</td>
<td></td>
<td>LCG</td>
</tr>
<tr>
<td>4</td>
<td>new</td>
<td>Executable CloseSSH</td>
<td>CloseSSH</td>
<td></td>
<td>LCG</td>
</tr>
<tr>
<td>5</td>
<td>completed</td>
<td>Executable CloseSSH</td>
<td>CloseSSH</td>
<td></td>
<td>LCG</td>
</tr>
<tr>
<td>6</td>
<td>new</td>
<td>Executable CloseSSH</td>
<td>CloseSSH</td>
<td></td>
<td>LCG</td>
</tr>
<tr>
<td>7</td>
<td>completed</td>
<td>Executable CloseSSH</td>
<td>CloseSSH</td>
<td></td>
<td>LCG</td>
</tr>
<tr>
<td>8</td>
<td>new</td>
<td>Executable CloseSSH</td>
<td>CloseSSH</td>
<td></td>
<td>LCG</td>
</tr>
<tr>
<td>9</td>
<td>new</td>
<td>Executable CloseSSH</td>
<td>CloseSSH</td>
<td></td>
<td>LCG</td>
</tr>
<tr>
<td>10</td>
<td>completing</td>
<td>Executable CloseSSH</td>
<td>CloseSSH</td>
<td></td>
<td>LCG</td>
</tr>
<tr>
<td>11</td>
<td>completed</td>
<td>Executable CloseSSH</td>
<td>CloseSSH</td>
<td></td>
<td>LCG</td>
</tr>
<tr>
<td>12</td>
<td>completed</td>
<td>Executable CloseSSH</td>
<td>CloseSSH</td>
<td></td>
<td>LCG</td>
</tr>
<tr>
<td>13</td>
<td>new</td>
<td>Executable CloseSSH</td>
<td>CloseSSH</td>
<td></td>
<td>LCG</td>
</tr>
<tr>
<td>14</td>
<td>completed</td>
<td>Executable CloseSSH</td>
<td>CloseSSH</td>
<td></td>
<td>LCG</td>
</tr>
<tr>
<td>15</td>
<td>completed</td>
<td>Executable CloseSSH</td>
<td>CloseSSH</td>
<td></td>
<td>LCG</td>
</tr>
<tr>
<td>16</td>
<td>completed</td>
<td>Executable CloseSSH</td>
<td>CloseSSH</td>
<td></td>
<td>LCG</td>
</tr>
<tr>
<td>17</td>
<td>completed</td>
<td>Executable CloseSSH</td>
<td>CloseSSH</td>
<td></td>
<td>LCG</td>
</tr>
<tr>
<td>18</td>
<td>new</td>
<td>Executable CloseSSH</td>
<td>CloseSSH</td>
<td></td>
<td>LCG</td>
</tr>
<tr>
<td>19</td>
<td>completed</td>
<td>Executable CloseSSH</td>
<td>CloseSSH</td>
<td></td>
<td>LCG</td>
</tr>
<tr>
<td>20</td>
<td>completed</td>
<td>Executable CloseSSH</td>
<td>CloseSSH</td>
<td></td>
<td>LCG</td>
</tr>
<tr>
<td>21</td>
<td>submitted</td>
<td>Executable CloseSSH</td>
<td>CloseSSH</td>
<td></td>
<td>LCG</td>
</tr>
<tr>
<td>22</td>
<td>submitted</td>
<td>Executable CloseSSH</td>
<td>CloseSSH</td>
<td></td>
<td>LCG</td>
</tr>
<tr>
<td>23</td>
<td>completed</td>
<td>Executable CloseSSH</td>
<td>CloseSSH</td>
<td></td>
<td>LCG</td>
</tr>
<tr>
<td>24</td>
<td>completed</td>
<td>Executable CloseSSH</td>
<td>CloseSSH</td>
<td></td>
<td>LCG</td>
</tr>
<tr>
<td>25</td>
<td>completed</td>
<td>Executable CloseSSH</td>
<td>CloseSSH</td>
<td></td>
<td>LCG</td>
</tr>
<tr>
<td>26</td>
<td>completed</td>
<td>Executable CloseSSH</td>
<td>CloseSSH</td>
<td></td>
<td>LCG</td>
</tr>
<tr>
<td>27</td>
<td>completed</td>
<td>Executable CloseSSH</td>
<td>CloseSSH</td>
<td></td>
<td>LCG</td>
</tr>
<tr>
<td>28</td>
<td>completed</td>
<td>Executable CloseSSH</td>
<td>CloseSSH</td>
<td></td>
<td>LCG</td>
</tr>
<tr>
<td>29</td>
<td>new</td>
<td>Executable CloseSSH</td>
<td>CloseSSH</td>
<td></td>
<td>LCG</td>
</tr>
<tr>
<td>30</td>
<td>completed</td>
<td>Executable CloseSSH</td>
<td>CloseSSH</td>
<td></td>
<td>LCG</td>
</tr>
<tr>
<td>31</td>
<td>completed</td>
<td>Executable CloseSSH</td>
<td>CloseSSH</td>
<td></td>
<td>LCG</td>
</tr>
<tr>
<td>32</td>
<td>running</td>
<td>Executable CloseSSH</td>
<td>CloseSSH</td>
<td></td>
<td>LCG</td>
</tr>
<tr>
<td>33</td>
<td>submitted</td>
<td>Executable CloseSSH</td>
<td>CloseSSH</td>
<td></td>
<td>LCG</td>
</tr>
</tbody>
</table>

**Job Details**

- **Job ID**: 36
- **Job Status**: "submitted"
- **Name**: "D020_cern"
- **Input Directory**: `/afs/cern.ch/user/e/elmshue/gangadir/workspace/Local/34/input`
- **Output Directory**: `/afs/cern.ch/user/e/elmshue/gangadir/workspace/Local/34/output`
- **Input Data**: D020Dataset
  - Tag Dataset: ""
  - Type: "D02_LOCAL"
  - Names: ""
  - Match Data: False
  - Dataset: "d811.005320_PythiaH170wwLRes.AOD.v110004107"
- **Merging**: None
- **Input and Output**: ""
- **Application**: Athena
  - 
  **Options**: ""
- **User Setup File**: File
  - Name: 
  - Subdir: 
- **Output Data**: D020OutputDataset
  - Output: ""
Results

Number of muons

$p_\perp$ of muon 1 [GeV]

$p_\perp$ of muon 2 [GeV]

$m_{\mu\mu}$ [GeV]

$\Delta \phi_{\mu\mu}$

$E_T$ final [GeV]

Number of Jets

$m_T(\mu\mu, E_T)$ [GeV]

$Z/\gamma^* \rightarrow \mu\mu$

$H \rightarrow WW \rightarrow \mu\nu\nu\nu$
1. ATLAS Distributed Analysis
2. Task: Access data in distributed environment
3. Introduction to the tool
4. How to prepare a job to run on the Grid
5. Example of real activity

6. Conclusions
Conclusions

- Demonstrated Distributed Analysis of ATLAS/Athena jobs with GANGA on LCG

- **GANGA:**
  - very flexible setup and easy extendable through plugin design
  - Job submission, Monitoring and Output retrieval almost fully automatic

- **Future Improvements:**
  - User view: bulk submission, better Athena job error recovery and book-keeping
  - Developer view: robust data access on all possible sites