INFN-Grid progress report

grid.infn.it

Antonia Ghiselli
INFN-CNAF
Bologna

Commissione Scientifica Nazionale 1
22 giugno 2004
Pisa
content

- INFN-Grid project: introduction
- INFN-Grid Production Infrastructure
  - Sites and resources
- INFN Production Grid organization and support
- INFN-Grid services
  - Present release
  - Next developments
- Grid use model (Atlas DC use case)
- Conclusions
**INFIN-Grid – goals (started at 2000)**

1. **To promote computational grid technologies research & development:**
   - **Middleware**
     1. Through european and international projects
        1. DataGrid, DataTAG, GLUE
     2. Internal R&D activities

2. **To implement the INFN grid infrastructure**
   1. National layout: 22 sites

3. **To set up the national Grid Infrastructure for the national research community**
   1. **FIRB:** *Grid.it (open to INAF, INGV, Bioinfo (PD,GE,MI))*

4. **To participate to the implementation of the global Grid infrastructure for the LHC community**
   1. **LCG:** Tier1 and n*Tier2

5. **To set up the eInfrastructure for the European Research Area**
   1. **EU FP6:** *EGEE*: robust grid services for an european infrastructure based on national Grids
Organizzazione

Giunta esecutiva dell’INFN

esperimenti → Executive Board → Referee group

Technical Board

Progetti Europei e nazionali

Infrastruttura di produzione

Iniziative r&d

Project manager mirco.mazzucato
Technical coordinators antonia.ghiselli,lamberto.luminari
Representative of INFN T1 federico.ruggieri

INFN Computing coordinators:
   ATLAS: laura.perini
   CMS: paolo.capiluppi
   LHC-B: domenico.galli
   ALICE: massimo.masera
   LCG: alberto.masoni
   APE: rapuano
   BABAR: morandin
   VIRGO: fulvio.ricci
   CDF: stefano.belforte

INFN contacts for other Sciences giorgio.maggi,leonardo.merola

Production Italian Grid ROC manager and EGEE ROC Coordinator cristina vistoli

Production Grid planning luciano gaido
INFN-Grid – collaborations and results

- EU – Datagrid (closed March 2004) : middleware development
  - **WMS** = job submission to the Grid,
    - CE and SE selection on the basis of job requirements specification, CPU load, CE-SE network conditions…..
    - Support for interactive jobs
    - Job checkpointing
    - Support for parallel jobs
  - **Virtual Organization authentication and authorization service**: **VOMS** (VO Membership Service, EDG/EDT)

- EU – DataTAG (closed March 2004) : inter-grid Interoperability; EU-US collaboration within the GLUE framework
  - Grid Resources Information modeling: **GLUE schema** for Computing and Storage Element
  - Authorization/authentication service : **VOMS-VOX** integration (EDT-Fnal/CMS coll.)
  - First WorldGrid demo by nov.2002 within IST2002 and SC2002 events
  - **GridICE**, Grid monitoring system based on GLUE schemas extension

- Italian Grid.it (on going): Grid management and support infrastructure
  - First tools in production
  - R&D on Resource Utilization Policies and Data Management
Infn-grid infrastructure

- Grid as “coordinated resource sharing” on a large scale for a multi-institutional and dynamic virtual organisation.

- sites 22
  - 3 T2 : MI, TO, LNL
  - 8 T2-T3 : PI, CT, TO, MI, PD, cnaf, Roma1
  - 3 sites grid.it (firb) : cnaf, PD, NA + siti INAF (TS, CT)
  - 14 T3: Bari, Bologna, Cagliari, Ferrara, Genova, Lecce, Legnaro, Napoli, Parma, Pavia, Roma1, Roma2, Roma3, e Trieste
  - T1 : cnaf

- Risorse di esperimento:
  - CMS: LNL, T1, ....
  - Atlas: LNF, NA, MI, T1, Roma1....
  - Babar : Ferrara, Padova
  - Virgo : Roma1, NA, Pisa, T1
Italian – Grid now (Site/resource map)

INFN
- CMS T2 T2/3
- Atlas T2 T2/3
- Alice T2 T2/3
- LHCb T2 T2/3
- Babar
- VIRGO

T2 (50-100 nodes)
T3 (10-15 nodes)
T1 Cnaf (~300)

grid.it resources
- INFN (15-25 nodes)
- INAF (5-10 nodes)
- INGV (NEC computers),
- BIO (tbd)
- general purpose resources (8-15 nodes)

Tot. ~ 1000 nodes
INFN - Grid infrastructure organization

Experiments (VOs) → Coordination Committee (now EB) → Grid Technical coordination

- VO representatives, Grid technical coord., Operations resp. grid experts
  - Deployment Planning
  - resource Policy application
  - .......

- Grid Services support
  - VO admin
  - New VO admin & support

- Site-man Resource admin

- Release distribution, documentation and porting

Grid Resource Coordination

- Resource availability
- Resource allocation
- Policy for shared resources

Service planning coordination

EGEE/CIC

EGEE/ROC

Support for New VO-users

From INFN-Grid, LCG, EGEE or National Middleware development projects
Release, Operation & Support organization

**Support:** 8SA1 + 3cnaf / T1 + 2unf. = **13p**
- CMS: N.Smirnov(PD), M.Corvo(PD) D.Bonaccorsi(cnaf)
- Atlas: G.Negri (MI/cnaf), S.Resconi(MI)
- Alice: S.Bagnasco(TO), G.Lore(cnaf)

- Servizi generali
  - Cnaf: D.Cesini, A.Paolini,
  - PD: M.Verlato
  - TO: A.Amoroso
  - NA: G.Tortone
  - LNL: S.Fantinel

- **CMT:** 1cnaf, 6SA1 man power= **7p**
  - Cnaf: A.Italiano, A.Cavalli, P.Veronesi, M.Donatelli
  - PD: F.Fanzago
  - MI: L.Vaccarossa (Atlas)
  - BA, G.Donvito

- **Release:** 4firb + 1sa1 man power= **5p**
  - PD: E.Ferro, A.Caltroni
  - CNAF: A.Ferraro, R.Zappi, L.Carota(sa1)

- **Site resource manager 20p**
  - Cnaf: A.Chierici
  - TO: S.Lusso
  - MI: L.Vaccarossa (Atlas)
  - LNL: M.Biasotto
  - RM1: A.Desalvo(Atlas)
  - NA: Taurino
  - CT: G.Andronico
  - PD: Ferrari
  - PI: M.Davini
  - FE: D.Andreotti
  - PV: C.DeVecchi
  - PG: M.Mariotti
  - LNF, ML.Ferrer
  - RM2, C.Dellefratte
  - CA: D.Mura
  - BA: G.Donvito
  - LE: E. Fasanelli
  - PR: R.Alfieri
  - BO : F.Semeria
  - TS : M.Macorini
Present Grid services for users
Release infn-grid 2.x based on LCG 2.x

General Services

- Security based on Globus GSI (Grid Security Infrastructure)
  - Standard Protocols: X.509 certificates, PKI, GSS-API
  - Login once (credential delegation)

- VO oriented Authentication/Authorization tools (VOMS with attribute extension of X.509 certificate)

- Grid Monitoring System

- VO oriented User Support systems

Blue: developed by INFN
User interacts with CA, VO and Resource Providers

1. Certificates are issued by a set of well-defined Certification Authorities (CA's).

2. Grant authorization at the VO level.
   1. Each VO has its own VOMS server.
   2. Contains (group / role / capabilities) triples for each member of the VO.

3. RP’s evaluate authorization granted by VO to a user and map into local credentials to access resources.

CAs: Policies and procedures → mutual thrust

- CERN
- CESNET
- CNRS
- GermanGrid
- Grid-Ireland
- INFN
- NIKHEF
- NorduGrid
- LIP
- Russian DataGrid
- DATAGRID-ES
- GridPP
- US–DOE Root CA
- US–DOE Sub CA
- CrossGrid

CA’s

VO-Manager
(administer user membership, roles and Capabilities)

Grid agreement

Resource Provider
(map into Local credential)
Grid.IT Production Grid: Operations Portal

- User documentation
- Site managers documentation
- Software repository
- Monitoring
- Trouble tickets system
- Knowledge base

Welcome to the INFN Production Grid for Scientific Applications!

INFN-GRID is a research project which features solutions and innovations in methodologies and technologies for the implementation and widespread use of large-scale platforms and grids. We participate to several National and International research projects on Grid Computing.

We’re coordinating our objectives with the strategies of the European Community to build the Next Generation Grid.

Our efforts are evaluated in terms of our grid capability to solve very critical, real problems in the medium-long term. The best standards in ICT are assumed as the technological starting point (e.g. OOP, Web services, Globus), over which new technologies are studied and built.

Read the latest news from October 31, 2003

http://grid-it.cnaf.infn.it
Get your personal certificate

Access to the grid
Get your certificate

Step 2: Get your personal certificate

To access the GRID you need a Personal Certificate (released by a Certification Authority) to be installed in a User Interface where you got an account.

1. Install the **Certification Authority Certificate** on your browser 🌐
2. Identify yourself to the **Registration Authority** in your department and ask him for an ID 🌐
3. Ask for your **Personal Certificate** using the ID given to you by the RA 🌐
4. Install your **Personal Certificate** on your browser (the same browser of step 1). You have to wait for a couple of days to receive a mail with a web link to the page containing your certificate.
5. Export your **Personal Certificate** from your browser 🌐
6. Copy your **Personal Certificate** in your home directory of a User Interface where you got an account

All these steps are described in detail in the following document:

- [INFN-GRID personal certificates howto](mailto:INFN-GRID personal certificates howto) - [PDF] - [TXT]

Go to: **Step 3: Register to a VO**
How to register to a VO

Access to the grid

Register to a VO

Step 3: Register to a VO

Using your personal certificate, you can be authenticated by the grid, but not authorized. If not authorized, you are not allowed, for instance, to submit jobs. To be authorized you must belong to a Virtual Organisation which is a kind of user group usually working on the same project and using the same application software on the grid. Your request will be evaluated by a VO manager.

Please note that to proceed with your registration your personal certificate has to be installed in your browser: it will be used to authenticate your identity.

- Click here to subscribe infngrid, theophys and virgo Virtual Organisation
- Click here to subscribe gridit, bio, ingv and inaf Virtual Organisation

Supported VOs

<table>
<thead>
<tr>
<th>VOs</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>bio</td>
<td>Biology group</td>
</tr>
<tr>
<td>ingv</td>
<td>INGV Bologna</td>
</tr>
<tr>
<td>inaf</td>
<td>INAF</td>
</tr>
<tr>
<td>gridit</td>
<td>General Grid.it Project VO</td>
</tr>
<tr>
<td>alice</td>
<td>LHC Alice experiment</td>
</tr>
<tr>
<td>atlas</td>
<td>LHC Atlas experiment</td>
</tr>
<tr>
<td>crns</td>
<td>LHC CMS experiment</td>
</tr>
</tbody>
</table>
Grid groups within the support system

Trouble Ticketing System

[Helpdesk Website](http://helpdesk.oneorzero.com)
GridICE
GridICE

- A monitoring system for the Grid Infrastructure
  - Selects grid entities (resources and services) per VOs and sites
- Automatic discovery based on Grid Information Service (Globus/MDS2.x) and Glue schema extensions.
- Layered architecture:
  - Measurement service (local monitoring interfaced to the GIS)
  - Publisher service
  - Data collector service with auto-discovery feature
  - Data Analyzer + Detection and notification service (on going)
  - Presentation service via web interface
- Modularity, flexibility and interoperability
- Ongoing activities:
  - Integration of network resource monitoring
  - Notification mechanism
- Utilizzato in infn-grid, LCG/EGEE, esperimenti (CMS, Atlas..)
Grid ICE components

1: LDAP Query
2: available CE/SE
3: LDAP Query
4: CEIDs, WNs,

Steps 3,4 repeated for every CE/SE
<table>
<thead>
<tr>
<th>Site</th>
<th>Q#</th>
<th>Slot#</th>
<th>SlotFree</th>
<th>SlotLoad</th>
<th>Power</th>
<th>WN#</th>
<th>CPU#</th>
<th>CPULoad</th>
<th>% Available</th>
<th>Total</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>ba.infn.it</td>
<td>3</td>
<td>186</td>
<td>0</td>
<td>100%</td>
<td>63</td>
<td>18</td>
<td>100%</td>
<td></td>
<td>209K</td>
<td>1.3 Tb</td>
<td>23%</td>
</tr>
<tr>
<td>bo.infn.it</td>
<td>3</td>
<td>30</td>
<td>0</td>
<td>100%</td>
<td>10</td>
<td>31</td>
<td>100%</td>
<td></td>
<td>10K</td>
<td>67.4 Gb</td>
<td>6%</td>
</tr>
<tr>
<td>bo.ingv.it</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>0</td>
<td>5</td>
<td>0%</td>
<td></td>
<td>-</td>
<td>23.8 Gb</td>
<td>31%</td>
</tr>
<tr>
<td>ca.infn.it</td>
<td>3</td>
<td>45</td>
<td>0</td>
<td>100%</td>
<td>15</td>
<td>3</td>
<td>100%</td>
<td></td>
<td>-</td>
<td>139.3 Gb</td>
<td>2%</td>
</tr>
<tr>
<td>cern.ch</td>
<td>4</td>
<td>42</td>
<td>0</td>
<td>100%</td>
<td>36</td>
<td>1353</td>
<td>100%</td>
<td></td>
<td>-</td>
<td>67.5 Gb</td>
<td>2%</td>
</tr>
<tr>
<td>cnaf.infn.it</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>cr.cnaf.infn.it</td>
<td>10</td>
<td>4704</td>
<td>1524</td>
<td>67%</td>
<td>450</td>
<td>138</td>
<td>22%</td>
<td></td>
<td>943K</td>
<td>57.8 Gb</td>
<td>65%</td>
</tr>
<tr>
<td>ct.infn.it</td>
<td>3</td>
<td>342</td>
<td>276</td>
<td>19%</td>
<td>11</td>
<td>25</td>
<td>3%</td>
<td></td>
<td>146K</td>
<td>1.8 Tb</td>
<td>9%</td>
</tr>
<tr>
<td>fe.infn.it</td>
<td>3</td>
<td>42</td>
<td>0</td>
<td>100%</td>
<td>14</td>
<td>40</td>
<td>100%</td>
<td></td>
<td>26K</td>
<td>23.8 Gb</td>
<td>7%</td>
</tr>
<tr>
<td>ifae.es</td>
<td>7</td>
<td>536</td>
<td>0</td>
<td>100%</td>
<td>117</td>
<td>6</td>
<td>100%</td>
<td></td>
<td>428K</td>
<td>3.9 Tb</td>
<td>23%</td>
</tr>
<tr>
<td>lnf.infn.it</td>
<td>3</td>
<td>18</td>
<td>18</td>
<td>0%</td>
<td>0</td>
<td>0</td>
<td>0%</td>
<td></td>
<td>16K</td>
<td>722.3 Gb</td>
<td>1%</td>
</tr>
<tr>
<td>lnl.infn.it</td>
<td>1</td>
<td>142</td>
<td>113</td>
<td>20%</td>
<td>1</td>
<td>0</td>
<td>0%</td>
<td></td>
<td>466K</td>
<td>159.0 Gb</td>
<td>68%</td>
</tr>
<tr>
<td>mi.infn.it</td>
<td>3</td>
<td>150</td>
<td>0</td>
<td>100%</td>
<td>50</td>
<td>1</td>
<td>100%</td>
<td></td>
<td>106K</td>
<td>1.2 Tb</td>
<td>20%</td>
</tr>
<tr>
<td>na.infn.it</td>
<td>6</td>
<td>108</td>
<td>0</td>
<td>100%</td>
<td>34</td>
<td>91</td>
<td>100%</td>
<td></td>
<td>95K</td>
<td>816.5 Gb</td>
<td>6%</td>
</tr>
<tr>
<td>oat.ts.astro.it</td>
<td>3</td>
<td>6</td>
<td>0</td>
<td>100%</td>
<td>2</td>
<td>40</td>
<td>100%</td>
<td></td>
<td>-</td>
<td>31.0 Gb</td>
<td>10%</td>
</tr>
<tr>
<td>pd.infn.it</td>
<td>3</td>
<td>192</td>
<td>186</td>
<td>3%</td>
<td>2</td>
<td>0</td>
<td>1%</td>
<td></td>
<td>249K</td>
<td>278.2 Gb</td>
<td>57%</td>
</tr>
<tr>
<td>pg.infn.it</td>
<td>3</td>
<td>12</td>
<td>12</td>
<td>0%</td>
<td>0</td>
<td>0</td>
<td>0%</td>
<td></td>
<td>3K</td>
<td>214.9 Gb</td>
<td>1%</td>
</tr>
<tr>
<td>pl.infn.it</td>
<td>3</td>
<td>27</td>
<td>0</td>
<td>100%</td>
<td>9</td>
<td>33</td>
<td>100%</td>
<td></td>
<td>11K</td>
<td>9.7 Gb</td>
<td>4%</td>
</tr>
<tr>
<td>pv.infn.it</td>
<td>3</td>
<td>18</td>
<td>18</td>
<td>0%</td>
<td>0</td>
<td>0</td>
<td>0%</td>
<td></td>
<td>7K</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>roma1.infn.it</td>
<td>6</td>
<td>150</td>
<td>87</td>
<td>2%</td>
<td>21</td>
<td>23</td>
<td>19%</td>
<td></td>
<td>108K</td>
<td>36.0 Gb</td>
<td>16%</td>
</tr>
<tr>
<td>roma2.infn.it</td>
<td>3</td>
<td>24</td>
<td>0</td>
<td>100%</td>
<td>8</td>
<td>29</td>
<td>100%</td>
<td></td>
<td>19K</td>
<td>29.7 Gb</td>
<td>5%</td>
</tr>
<tr>
<td>to.infn.it</td>
<td>4</td>
<td>96</td>
<td>0</td>
<td>100%</td>
<td>24</td>
<td>5</td>
<td>100%</td>
<td></td>
<td>57K</td>
<td>416.2 Gb</td>
<td>79%</td>
</tr>
</tbody>
</table>

TOTAL: 79 6870 2234 68% 867 1841 59% 2M 628 1083 50% 11.2 Tb 34.0 Tb 25%
### VO totals

<table>
<thead>
<tr>
<th>VO name</th>
<th>Jobs Running</th>
<th>Jobs Queued</th>
<th>Jobs Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>inaf</td>
<td>99</td>
<td>39</td>
<td>138</td>
</tr>
<tr>
<td>ingv</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>atlas</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>dteam</td>
<td>4</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>cms</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
</tbody>
</table>

### Site Name: **ba.infn.it**

<table>
<thead>
<tr>
<th>VO name</th>
<th>Jobs Running</th>
<th>Jobs Queued</th>
<th>Jobs Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>inaf</td>
<td>76</td>
<td>0</td>
<td>76</td>
</tr>
<tr>
<td>ingv</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

### Site Name: **mi.infn.it**

<table>
<thead>
<tr>
<th>VO name</th>
<th>Jobs Running</th>
<th>Jobs Queued</th>
<th>Jobs Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>atlas</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>dteam</td>
<td>4</td>
<td>0</td>
<td>4</td>
</tr>
</tbody>
</table>

### Site Name: **na.infn.it**

<table>
<thead>
<tr>
<th>VO name</th>
<th>Jobs Running</th>
<th>Jobs Queued</th>
<th>Jobs Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>inaf</td>
<td>22</td>
<td>39</td>
<td>61</td>
</tr>
</tbody>
</table>

### Site Name: **pd.infn.it**

<table>
<thead>
<tr>
<th>VO name</th>
<th>Jobs Running</th>
<th>Jobs Queued</th>
<th>Jobs Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>cms</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>inaf</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>
## JOB details: site pd.infn.it - VO cms

<table>
<thead>
<tr>
<th>Local ID: 15745</th>
<th>Global ID: <a href="https://grid014.ct.infn.it:9000/jilcxLufGkaFY8n4U44cod">https://grid014.ct.infn.it:9000/jilcxLufGkaFY8n4U44cod</a></th>
</tr>
</thead>
<tbody>
<tr>
<td>Job Name</td>
<td>Job Status</td>
</tr>
<tr>
<td>15745</td>
<td>R</td>
</tr>
<tr>
<td>Execution Target</td>
<td>CPU Time</td>
</tr>
<tr>
<td>lcms25</td>
<td>115</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Local ID: 15744</th>
<th>Global ID: <a href="https://grid014.ct.infn.it:9000/_m0u7kDh9infX40hr36o1g">https://grid014.ct.infn.it:9000/_m0u7kDh9infX40hr36o1g</a></th>
</tr>
</thead>
<tbody>
<tr>
<td>Job Name</td>
<td>Job Status</td>
</tr>
<tr>
<td>15744</td>
<td>R</td>
</tr>
<tr>
<td>Execution Target</td>
<td>CPU Time</td>
</tr>
<tr>
<td>lcms26</td>
<td>1235930</td>
</tr>
</tbody>
</table>

Generated: Mon, 21 Jun 2004 12:34:47 +0200
Data presentation (3)
Grid services for scientific applications

- Grid job submission
  - **RJS (Remote Job Submission) to specific computers (Globus):**
    - The user submits a job on a WAN computing system providing the address
  - **RJS to Grid domains (INFN)**
    - Via a Grid scheduler without knowing the destination computers

- Grid Information System (UK) (**GLUE schema**)

- Data Management (CERN)
  - **Data replica management**
    - **Remote Data location** via File catalogue and metadata catalogue (RLS Replica Location Service)
    - **Data transfer and access,**
      - **GridFTP:** Provides high-performance, reliable data transfer

- Storage Resource Management (UK) for
  - Storage allocation
  - File pinning.....

- Grid User Interface : **UI, Genius (INFN)**

- Job Logging & bookkeeping (INFN)
Grid services (LCG-2 release)
‘real’ Grid Job Submission is allowed via the Workload Management Service (INFN within EU-DataGrid)

- The user interacts with Grid via a **Workload Management System** (not directly with GRAM)
- The Goal of WMS is the **distributed scheduling and resource management in a Grid environment**.
- What does it allow Grid users to do?
  - To submit their jobs via a “**Job description language**”
  - To execute them
  - To get information about their status
  - To retrieve their output

- **The WMS tries to optimize the usage of resources**
  (funzione che sara’ ottimizzata in EGEE)
Job Preparation

- Information to be specified
  - Job characteristics
  - Requirements and Preferences of the computing system
  - Software dependencies
  - Job Data requirements

- It is specified using a Job Description Language (JDL)
  - It allows to define a set of attributes, the WMS takes into account when making its scheduling decision
Based upon Condor’s *CLASSified ADvertisement language (ClassAd)*, that is a fully extensible language and that allows to define easily requirements and preferences to be used by the matchmaking algorithm

- **Executable** (mandatory)
  - The command name

- **Arguments** (optional)
  - Job command line arguments

- **StdInput, StdOutput, StdErr** (optional)
  - Standard input/output/error of the job

- **Environment** (optional)
  - List of environment settings

- **InputSandbox** (optional)
  - List of files on the UI local disk needed by the job for running
  - The listed files will automatically staged to the remote resource

- **OutputSandbox** (optional)
  - List of files, generated by the job, which have to be retrieved
Job Description Language (JDL): relevant attributes 3/4

**Requirements**
- Job requirements on computing resources
- Specified using attributes of resources published in the Information Service
- If not specified, default value defined in UI configuration file is considered
  - Default: other.Active (the resource has to be able to accept jobs)

**Rank**
- Expresses preference (how to rank resources that have already met the Requirements expression)
- Specified using attributes of resources published in the Information Service
- If not specified, default value defined in the UI configuration file is considered
  - Default: -other.EstimatedTraversalTime (the lowest estimated traversal time)
Job Description Language (JDL): “data” attributes 4/4

◆ **InputData** (optional)
  - Refers to data used as input by the job: these data are published in the Replica Catalog and stored in the SEs)
  - PFNs and/or LFNs

◆ **DataAccessProtocol** (mandatory if InputData has been specified)
  - The protocol or the list of protocols which the application is able to speak with for accessing **InputData** on a given SE

◆ **OutputSE** (optional)
  - The Uniform Resource Identifier of the output SE
  - RB uses it to choose a CE that is compatible with the job and is close to SE
GUI & APIs

Remote Machine Type
- Architecture: INTEL
- OS Type: LINUX
- OS Version:

Connectivity
- Inbound
- Outbound

Remote Site Job Manager
- LRMS Type: LSF
- LRMS Version:

Remote Site Parameters
- Min Number of Free CPUs: 3
- Min Main Memory: 256 Mb
- Max Available CPU Time: 0.0 sec
- Time to Traverse Queue: 0.2 sec

requirements = other.GlueHostArchitecturePlatformType == "INTEL" &
other.GlueHostOperatingSystemName == "LINUX" &
other.GlueHostNetworkAdapterInboundIP ==
true &
other.GlueHostMainMemoryRAMSize >= 256.0 &
other.GlueCEinfoLRMSType == "LSF" &
other.GlueCEStateFreeCPUs >= 3 &
other.GlueCEStateStatus == "Production";
WMS Components

- WMS is currently composed of the following parts:
  1. **User Interface** (UI) : access point for the user to the WMS
  2. **Resource Broker** (RB) : the broker of GRID resources, responsible to find the “best” resources where to submit jobs
  3. **Job Submission Service** (JSS) : provides a reliable submission system
  4. **Information Index** (II) : a specialized Globus GIIS (LDAP server) used by the Resource Broker as a filter to the information service (IS) to select resources
  5. **Logging and Bookkeeping** services (LB) : store Job Info available for users to query
WMS Architecture in EDG v2.x

UI

Network Server

RLS

Match-Maker/Broker

Inform. Service

RB node

Log Adapter

Log Monitor

Job Contr. - CondorG

Match-Maker/Broker

CE characts & status

SE characts & status

RB storage

Logging & Bookkeeping

Log Adapter

Workload Manager

Job Adapter
Other WMS UI Commands

- **dg-job-list-match**
  Lists resources matching a job description
  Performs the matchmaking without submitting the job

- **dg-job-cancel**
  Cancels a given job

- **dg-job-status**
  Displays the status of the job

- **dg-job-get-output**
  Returns the job-output (the OutputSandbox files) to the user

- **dg-job-get-logging-info**
  Displays logging information about submitted jobs (all the events “pushed” by the various components of the WMS)
  Very useful for debug purposes

- **dg-job-id-info**
  A utility for the user to display job info in a formatted style
WMS release 2 functionalities

- User APIs
- GUI
- Support for interactive jobs
- Job checkpointing
- Support for parallel jobs
- Support for automatic output data upload and registration
- VOMS (Management Service) support
- Support for job dependencies (via DAGman)
  - “Lazy” scheduling: job (node) bound to a resource (by RB) just before that job can be submitted (i.e. when it is free of dependencies)
What is a DAG?

- Directed Acyclic Graph
- Each node represents a job
- Each edge represents a (temporal) dependency between two nodes
  - e.g. NodeC starts only after NodeA has finished
WMS: Future activities

- New functionalities

  - Support for job partitioning (available soon)
    - Use of job checkpointing and DAGMan mechanisms
      - Original job partitioned in sub-jobs which can be executed in parallel
      - At the end each sub-job must save a final state, then retrieved by a job aggregator, responsible to collect the results of the sub-jobs and produce the overall output
  
  - Grid Accounting: DGAS (testing in progress)
    - Based upon a computational economy model
      - Users pay in order to execute their jobs on the resources and the owner of the resources earn credits by executing the user jobs
    - help the Workload Management System in balancing the workload by supporting economic resource brokering
  
  - VO-based Resource access Policy support: G-PBOX (development in progress); Grid resource sharing requires:
    - to deploy VO-wide policies.
    - to respect local site policies.
    - to specify policies relating to the behavior of the grid as a whole.
    - RB can take decision on the basis of VO/user policies
VO oriented Policy system: G-PBOX

Policy Examples:

- Users belonging to group /vo/a may only submit 10 jobs a day.
- Users belonging to group /vo/b should have their jobs submitted on the max priority queue.
- “some user” is banned from the CNAF site.

Requirements. The system should:

- Be VO-based and distributed.
- Be highly configurable and able to define and enforce previously unknown types of policies.
- Leave total control on local sites to local admins.
- Be capable of express policies requiring a global view of the grid.
- Be compliant to existing protocols and not require their redesign.

Objective: help the Workload Management System in Grid resource selection.
Data Management (CERN): requirements

In EU-DataGrid and Globus the technique adopted for optimizing data access and providing fault tolerance is data replication.

The Requirements are:

- Planning, scheduling and monitoring execution of data requests and computations

- Management of data replication
  - Register and query for replicas
  - Select the best replica for a data transfer

- Security
  - Protect data on storage systems
  - Support secure data transfers
  - Protect knowledge about existence of data

- Virtual data
  - Desired data may be stored on a storage system ("materialized") or created on demand
Data Management services

- **A Replica Location Service (RLS)** is a distributed registry service that records the locations of data copies and allows discovery of replicas.

- Maintains mappings between *logical* identifiers and *target names*.
  - Physical targets: Map to exact locations of replicated data.
  - Logical targets: Map to another layer of logical names, allowing storage systems to move data without informing the RLS.

- RLS was designed and implemented in a collaboration between the Globus project and the DataGrid project.

- EDG: RM + RLS (problemi di lentezza... altre soluzioni?)

- Necessita’ di RLS locali: problema della consistenza delle repliche, che si sta studiando in INFN-Grid.
Storage Resource Management adopted by LCG

Data are stored on disk pool servers or Mass Storage Systems based on different technologies.

SRM (Storage Resource Manager) is a Grid Service that takes care of local storage interaction (to several secondary and tertiary storage systems) and provides a Grid interface to outside world (hiding the details of the local storage management from the user).

(Original design protocol: LBL, JNL, FNAL, CERN)

- Storage resource management takes into account
  - Transparent access to files (migration to/from disk pool)
  - File pinning
    - Files can be read from disk caches rather than from tape
  - Space reservation
  - File status notification….

- attualmente c’è la versione per castor e dcache
Grid Storage Element Interfaces

First version of SE: disk server with GridFTP and NFS server protocols

New SE version:

- **SRM interface**
  - Management and control
  - SRM (with possible evolution)

- **Posix-like File I/O**
  - File Access
  - Open, read, write
  - Not real posix (like rfio)

---

**File access**

---

**Storage Management**

**INFN-Grid**

**SToRM**

POSIX API File I/O

SRM interface

Storage allocation quota management

rfio
dcap
chirp
Castor MSS
dCache
NeST
Disk /GPFS
General User Interface : GENIUS

- Based on WEB portal architecture
- Support for generic applications
- Basic Requirement: Grid transparent access
- It must be accessed from everywhere and by “everything” (desktop, laptop, PDA, WAP phone).
- It must be redundantly “secure” at all levels: 1) secure for web transactions, 2) secure for user credentials, 3) secure for user authentication, 4) secure at VO level.
- All available grid services must be incorporated in a logic way, just “one mouse click away”.
- Its layout must be easily understandable and user friendly.
- Evolution in EGEE/NA4
INFN-Grid, TB: Attività – 2004
coord. A.Ghiselli, L.Luminari

Progetti
- INFN-Grid
- DataGrid (chiuso)
- DataTAG (chiuso)
- Grid.IT (2005)
- EGEE (2006)
- Coregrid (2006)

Attività’
- WMS,
- GridICE,
- VOMS,
- GLUE schema
- Grid-Policy
- Data Management
  - STOrM
  - Consistenza delle repliche
  - Grid-database
- Advance Reservation
- Network Services
- Genius (Portali)
- Release
- Grid di produzione: ROC-CIC

Coordinatori
- F.Prelz, M.Sgaravatto
- G.Tortone, S.Fantinel
- R.Cecchini, V.Ciaschini
- S.Andreozzi
- V.Ciaschini, GL.Rubini
- R.Zappi, L.Magnoni
- F.Donno
- E.Ambrosi
- T.Ferrari, E.Ronchieri
- T.Ferrari
- R.Barbera
- E.Ferro
- C.Vistoli, L.Gaido
INFN-GRID Technical Board

Attività’, Progetti manpower funded e unfunded

Grid di produzione

Management Operations

usr

Grid di produzione

R&D

WMS

Pbox

DM

AA

InfoMod

Mon

portali

OGSA

EGEE

Firb-Grid

CoreGrid

GridCC

EGEE Appl. & Dissem.

Progetti
Global Grid services view

Grid management

Operation

Support

Grid use planning

Release&doc

Users: WMS-UI, Genius

Applications and WMS-API

Collective services
WMS(Grid scheduler), replica consistency

Resource or Core services
GRAM
Storage allocation, Advance reservation, network service

Grid Resources and Local Services layer
Compute Element, Storage Element, Network, Local Authorization Service, Local policy Service, Local Scheduler

Network Layer

General Services

CA services

VOMS

G-Policy and G-Accounting

Monitoring

Network Layer
Grid use model

❖ ipotesi

❖ Un esperimento è interessato a risorse distribuite in diversi domini di grid (risorse proprie, risorse condivise nazionali, risorse condivise internazionali).

❖ I domini di grid hanno proprie organizzazioni di management, operation e support che devono collaborare (INFN-Grid, GridPP, NorduGrid, Grid3, ....). EGEE è una interconnessione di Grid (non una grid sola) che oltre alla release software cerca di definire come mettere insieme le diverse strutture organizzative. *Cio’ funzionera’ se ogni struttura nazionale mantiene una sua autonomia.*

❖ L’esperimento è interessato a risorse hardware con particolari requisiti e concorda le condizioni di utilizzo (policy).

❖ approccio

❖ Grid virtuale dell’esperimento

❖ Accordo e Applicazione delle Politiche di uso delle risorse per ogni dominio di Grid

❖ Supporto distribuito e coordinato
VO-Virtual Grid on top of Multi-Grids

Coordinated VO-support

Risorse private

INFIN-Grid
same middleware
ex: LCG-2

VO-User (UI)

VO-monitoring
VOMS
RB/II

VO-RLS
RB/II

US-Grid
same core services

VO-User
Grid use model: esempio di Atlas Data Challenge in ambiente omogeneo (INFN-Grid, LCG/EGEE)

◆ Atlas Virtual Grid resources:
  - INFN-Grid: T1, MI, Roma1, LNF, NA, TO, LNL..... to be defined with CC
  - LCG/EGEE: Nikhef, RAL, (T1 di Atlas + .....) to be defined with .......

◆ Atlas Virtual Grid services: to be defined with planning group (and CIC?)
  - VO-ldap service located at Nikhef and managed by A. De Salvo.
  - 3 RB: one at CERN, managed by CERN.IT + Miguel Branco + A De Salvo, 1 at CNAF and 1 in Milano, managed by INFN-Grid ROC
  - GridICE monitoring server located in Milano
  - RLS.... Managed by the CERN OP
  - >5 UI in the ATLAS sites (increase number for analysis phase)

◆ Policy: alta priorità nell’accesso delle risorse per 1 mese (produzione per la simulazione dei dati, circa altrettanto per la ricostruzione)

◆ Storage allocation: circa 1TB locali ai CE

Atlas DC

- Supporto
  - Atlas support team: DeSalvo, Resconi, Negri
  - Site e grid service support di INFN-Grid come parte dell’EGEE support
  - Uso piu’ intensivo del Supporto per l’analisi a causa di un numero piu’ elevato di utenti.

- Monitoring
  - LCG/EGEE GOC
  - GridICE specifico di Atlas
    - Notifica dei down
    - Monitoring dei job specifici di Atlas
    - Stato delle risorse e dei servizi
Conclusions

- Grid services are ready for production Grids and already in use in LCG in EU, INFN-Grid in Italy and in Grid3 in US.

- They are still evolving for more functionalities, robustness and security (EGEE).

- **Standardization is towards:** *Bringing Grid & Web Services Together* (Dr. Daniel Sabbah, Vice President of Strategy & Technology, IBM Software Group, Globus World in January 2004): linee guida per EGEE
  
  - **Web Service Resource Framework** is the new grid model: Modeling Stateful Resources with Web Services
    - WS-Notification
      - Provides a publish-subscribe messaging capability for Web Services
    - WS-Resource framework
      - A family of Web services specifications that clarify how "state" and Web services combine

- Il successo dell’utilizzo della Grid da parte degli esperimenti passa anche dall’interazione con l’organizzazione della grid di produzione ‘nazionale’, per assicurarsi il massimo di sinergia per il funzionamento delle risorse e servizi, e il massimo del supporto per identificare e risolvere velocemente i problemi. Va evitata una eccessiva centralizzazione ed eccessiva nell’utilizzo delle Grid....