Geographical failover for the EGEE-WLCG Grid collaboration tools

CHEP 2007
Victoria, Canada, 2-7 September

Alessandro Cavalli, Alfredo Pagano
(INFN/CNAF, Bologna, Italy)

Cyril L'Orphelin, Gilles Mathieu, Osman Aidel
(IN2P3/CNRS Computing Centre, Lyon, France)

Rafal Lichwala (PSNC, Poznan, Poland)
Outline

• The Failover System
• Technical solutions
  – the DNS and the new domain
  – Oracle replication
• Use cases
  – CIC Portal
  – SAM Admin
  – GSTAT, GRIDICE
  – GOCDB
  – SAM

• Future plans and improvements
  – Distributed agents
  – Monitoring system
  – CIC Portal ?
Failover: definition and principle

- A backup operation that automatically switches to a standby database, server or network if the primary system fails or is temporarily shut down for servicing.

- Failover is an important fault tolerance function of mission-critical systems that rely on constant accessibility.

- Failover automatically and transparently to the user redirects requests from the failed or down system to the backup system that mimics the operations of the primary system.
Failover: definition and principle

• How much availability must we guarantee?

<table>
<thead>
<tr>
<th>Availability</th>
<th>Downtime/Year</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>90.0%</td>
<td>36 days, 12 hours</td>
</tr>
<tr>
<td>2</td>
<td>99.0%</td>
<td>87 hours, 36 min</td>
</tr>
<tr>
<td>3</td>
<td>99.9%</td>
<td>8 hours, 45.6 min</td>
</tr>
<tr>
<td>4</td>
<td>99.99%</td>
<td>52 min, 33.6 sec</td>
</tr>
<tr>
<td>5</td>
<td>99.999%</td>
<td>5 min, 15.4 sec</td>
</tr>
<tr>
<td>6</td>
<td>99.9999%</td>
<td>31.5 seconds</td>
</tr>
</tbody>
</table>
Downtime causes

- Magic words are:
  - Redundancy
  - Remove Single Points of Failure (SPOF)
Propose, implement and document failover procedures for the collaboration, management and monitoring tools used in EGEE/WLCG Grid.

- The mentioned tools (listed later in this talk) are daily and heavily used by COD teams, regional and sites operators and other user categories, for grid management and control purposes.

- These are the reasons for an availability requirement that is high and which tend to become higher in future.
• Born as EGEE SA1 Operations COD task

• Reminder: who are the “CODs”?
  – Teams provided by EGEE federations, working in pairs (one lead and one backup) on a weekly rotation
  – Role:
    ▪ Watch the problems detected by the grid monitoring tools
    ▪ Problem diagnosis
    ▪ Report these problems (GGUS tickets)
    ▪ Follow and escalate them if needed (well defined procedure)
    ▪ Provide help, propose solutions
    ▪ Build and maintain a central knowledge database (WIKI)
Links to other projects

• **LCG-3D**
  - Some differences
    ▪ Our failover activity deals with operational tools
    ▪ LCG-3D deals with databases and data transfers.
  - Some similarities and shared goals
    ▪ work on databases replication and switches
    ▪ Same concern in disaster recovery solutions
    ▪ Often involves the same (or at least related) teams.
  - discussions and working sessions engaged

• **Other failover related examples**
  - TODO
DNS based failover

• **DNS choice:**
  – Well supported by local staff at our institutes
  – Easy to understand how to exploit its features
  – Very stable and consolidated (born in the ’70s)
  – Widely used as element for failover solution by ISPs and IT companies

• **the DNS approach consist in:**
  – mapping the service name to one or more destinations
  – update this mapping whenever some failure is detected

• **this must be coupled with procedures that:**
  – keep data in sync where it is needed
  – kill unnecessary processes on the system in failure
  – enable needed processes on the replacing system.
DNS downsides

- ISP caching policies
- Caching at OS level
- Caching by the web browsers
- TO BE COMPLETED
Technical details & new gridops.org domain

• TODO
Available operations tools:

- CIC Portal
- GSTAT
- GRIDICE
- GOCDB
- SAM
- SAMAP

• Many tools, many links on www.gridops.org
• TO BE REVISED
geographical failover

User asking for CIC

Local DNS

Root DNS

gridops.org

contacted cic-main

CIC portal (main)

CIC portal (replica)

Operator

Monitoring

Victoria, Canada, 2-7 Sep 2007
geographical failover: automatic switch

Local
Root
gridops.org
nsupdate
User
Operator
Monitoring

monitoring
activate if needed
CIC portal (replica)
CIC portal (main)

NOT YET AVAILABLE

Enabling Grids for E-sciencE
geographical failover: active-active

The diagram illustrates a user asking for information from a GSTAT (main) and a GSTAT2 (replica) server. The user is connected to a local DNS that resolves to gridops.org, which in turn resolves to GSTAT and GSTAT2. The nsupdate process is responsible for updating the DNS records.

The DNS hierarchy starts with the root domain, which is then resolved to the local DNS, and finally to the gridops.org domain. The user interacts with the GSTAT and GSTAT2 servers, which are part of the geographical failover setup. The diagram also includes an operator and monitoring system.
Use case: CIC portal failover

• Replication added early on the list
  – Highly critical tool
  – Planned or unexpected service outages could break the continuity of daily activity

• First switch successfully done
  – Replica instance used in production during one whole week
  – Normal use of the portal during this time
  – No problem reported
The CIC portal is based on three distinct components:
- A web portal module (php and html, css files)
- A database module (Oracle)
- A data processing system (named Lavoisier)

Each component can work with any of the other, master or replica: 8 possible scenarios!
CIC portal failover (cont.)

• **Web module replication**
  - Portable code
  - Environment & configuration on replica (Apache, PHP, libs)
  - Host certificate for the replica

• **Data processing system (Lavoisier) replication**
  - Environment and configuration on replica (Apache ANT, Globus toolkit, java web services)
  - Deployment of a second instance of Lavoisier
  - Settings on replica (e-mail address for local error reporting…)

• **Database replication**
  - Dump of master DB exported to replica
  - Well established procedure, involving 2 persons and an interruption of service
  - We are working on better solutions
Use case: SAMAP – functionality

• SAMAP (SAM Admin's Page) - web-based monitoring tool for submitting SAM Framework test jobs on demand
• based on SAM-client (Site Availability Monitoring Framework in EGEE)
• additional functionality implemented in response to the site administrator's needs
• provided functionality:
  – SAM job submission on demand
  – check status of the running test jobs
  – cancel submitted SAM jobs
  – publish SAM job results
  – show logging info of the running test jobs
  – schedule regular SAM job submission (cron task management)
SAMAP architecture divided into two independent parts: Portal part and UI (grid User Interface) part

- Portal part integrated with the CIC Portal
- UI part installed on dedicated UI machine
- SAMAP installed in two independent instances and linked to proper DNS domain entries
- Synchronization of instances via CVS repository
- Separate WMS servers available for both instances
- Easy switch from main to backup instance by DNS entry modification
- Full transparency for end users
• TODO
Future plans

• TODO