Clustering MetaLib at Brazilian Government Library Consortium

CAPES – Coordination for the Improvement of the Higher Education Personnel

Ronan Moraes – Technical Manager
Gustavo Portella – IT Senior Analyst
Summary

• Introducing CAPES
• Portal de Periódicos CAPES
  – Brief Description
  – The Portal in Numbers
• The New Portal
  – Project Overview
  – Hardware Architecture
  – Software Outline
• Clustering Solution
  – MetaLib Cluster
  – Benefits
• Conclusion
Introducing CAPES

• CAPES is a governmental foundation responsible for the Brazilian strictu sensu post graduation system (Master and PhD levels).

• CAPES main tasks are:
  – Evaluation of the post graduation system;
  – Access and communication of the scientific production;
  – Investment on high level human resources in Brazil and worldwide;
  – Promotion of international scientific cooperation
  – Formation of teachers for basic education.

• In this context, Portal de Periódicos is an important research instrument for the benefit of the entire Brazilian education system.
• Brief Description
  – Portal de Periódicos is a CAPES initiative and was created on November 2000 as an instrument for on line access to updated scientific content.
  – It is a tool for democratizing scientific information in order to reduce regional differences in the Brazilian research and post graduation system.
  – It contributes to increase national scientific production and to enlarge the Brazilian insertion in the international academic community.
  – The acquisition of new titles by CAPES is a decision based on the demand requested by academic institutions and collegiate committees.
The Portal in Numbers

Considered the Brazilian biggest library consortia, it includes:

- more than 15,000 full text journals;
- 126 abstract databases;
- and six patent databases.

One of the world’s most reached library consortia, as it is accessed by 308 institutions placed on the entire Brazilian territory.

The Portal also offers access to books, technical standards, audio and visual content and online training.
The New Portal

• Project Overview
  – In 2006, CAPES started the infrastructure upgrade project with the support of RNP – the National Research Network.
  – The New Portal should implement the specific requirements:
    • Allow the management of local resources and contracts with the editors;
    • Generate reliable statistics information about resource access by institutions, categories and other criteria;
    • Make the process of content publication in the Portal to be handled only by the administrative users of CAPES (without the help of the IT team);
    • Offer customized information and services to the users, according to its interests;
    • Use RNP backbone infra-structure;
    • **Optimize the use of the available resources.**
The New Portal

- Hardware Architecture

<table>
<thead>
<tr>
<th>Item</th>
<th>Description</th>
<th>Model</th>
<th>Brand</th>
<th>#</th>
<th>Cores</th>
<th>Clock</th>
<th>RAM</th>
<th>Disk</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Server SUN X4600 M2</td>
<td>8356</td>
<td>AMD</td>
<td>8</td>
<td>4</td>
<td>2.3Ghz</td>
<td>64GB</td>
<td>4 SAS 146GB</td>
</tr>
<tr>
<td>2</td>
<td>Server SUN X4600 M2</td>
<td>8356</td>
<td>AMD</td>
<td>8</td>
<td>4</td>
<td>2.3Ghz</td>
<td>64GB</td>
<td>4 SAS 146GB</td>
</tr>
<tr>
<td>3</td>
<td>Server SUN X4600 M2</td>
<td>8220</td>
<td>AMD</td>
<td>8</td>
<td>2</td>
<td>2.8Ghz</td>
<td>32GB</td>
<td>4 SAS 146GB</td>
</tr>
<tr>
<td>4</td>
<td>Server SUN X4600 M2</td>
<td>8220</td>
<td>AMD</td>
<td>8</td>
<td>2</td>
<td>2.8Ghz</td>
<td>32GB</td>
<td>4 SAS 146GB</td>
</tr>
<tr>
<td>5</td>
<td>Server SUN X4600 M2</td>
<td>8220</td>
<td>AMD</td>
<td>8</td>
<td>2</td>
<td>2.8Ghz</td>
<td>32GB</td>
<td>4 SAS 146GB</td>
</tr>
<tr>
<td>6</td>
<td>Server SUN X4150 M2</td>
<td>X5460</td>
<td>Intel</td>
<td>2</td>
<td>4</td>
<td>3.16Ghz</td>
<td>16GB</td>
<td>4 SAS 146GB</td>
</tr>
<tr>
<td>7</td>
<td>Server SUN X4150 M2</td>
<td>X5460</td>
<td>Intel</td>
<td>2</td>
<td>4</td>
<td>3.16Ghz</td>
<td>16GB</td>
<td>4 SAS 146GB</td>
</tr>
<tr>
<td>8</td>
<td>Server SUN X4150 M2</td>
<td>X5460</td>
<td>Intel</td>
<td>2</td>
<td>4</td>
<td>3.16Ghz</td>
<td>16GB</td>
<td>4 SAS 146GB</td>
</tr>
<tr>
<td>9</td>
<td>Server SUN X4150 M2</td>
<td>X5460</td>
<td>Intel</td>
<td>2</td>
<td>4</td>
<td>3.16Ghz</td>
<td>16GB</td>
<td>4 SAS 146GB</td>
</tr>
<tr>
<td>10</td>
<td>Server SUN X4150 M2</td>
<td>X5460</td>
<td>Intel</td>
<td>2</td>
<td>4</td>
<td>3.16Ghz</td>
<td>16GB</td>
<td>4 SAS 146GB</td>
</tr>
<tr>
<td>11</td>
<td>Server SUN X4150 M2</td>
<td>X5460</td>
<td>Intel</td>
<td>2</td>
<td>4</td>
<td>3.16Ghz</td>
<td>16GB</td>
<td>4 SAS 146GB</td>
</tr>
<tr>
<td>12</td>
<td>Server Dell PowerEdge 2850</td>
<td>Xeon</td>
<td>Intel</td>
<td>2</td>
<td>2</td>
<td>2.8Ghz</td>
<td>6GB</td>
<td>3 SCSI 36GB</td>
</tr>
<tr>
<td>13</td>
<td>Server HP DL 380 G2</td>
<td>Xeon</td>
<td>Intel</td>
<td>2</td>
<td>2</td>
<td>2.4Ghz</td>
<td>2GB</td>
<td>2 SCSI 36GB</td>
</tr>
<tr>
<td>14</td>
<td>Storage Netapp 2020</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Foundry, FastIron Edge X424</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>Foundry, FastIron Edge X424</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>Backup system</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* There is also a KVM module for remote maintenance.
### The New Portal

- **Software Outline**
  - **Red Hat 5**
    - Operating System
  - **ExLibris:**
    - SFX
    - MetaLib *(implementing Heartbeat cluster solution)*
    - Verde
  - **Joomla**
    - Content management of the Portal
    - MySQL database
  - **EzProxy**
    - Access control in application layer
  - **OpenLDAP**
    - User catalog and authorization control
  - **Shibboleth**
    - Federative integration with institutions
  - **Heartbeat**
    - Cluster stack solution

---

**Important:** every installation and data path is mapped to the storage through NFS. Like other ExLibris tools, MetaLib has one (and only) installation mapped to the storage.
The New Portal

- Software Outline
• MetaLib Cluster
  – Cluster solution with Heartbeat considering:
    • IP address virtualization: each node has its own IP address, and there is an extra IP address that is used for virtualization. MetaLib installation “knows” only the virtual IP address;
    • File System virtualization: the MetaLib path is in the storage and is mapped to a remote NFS path on each node of the cluster;
    • MetaLib application: standard Metalib application service and z39.50 gateway protocol.
  – Configuration Details:
    • Server priority definition: active and passive nodes;
    • Automatic service migration based on availability analysis;
    • IPMI interface and protocol configuration;
    • Forced migration for hardware maintenance;
    • Scalability can be achieved by adding more nodes to the cluster;
    • Shell script configuration.
• MetaLib Cluster
  – Shell script configuration:
    • MetaLib shell script configuration based on ExLibris’ script;
    • Start and Stop service options were maintained similar to the original script;
    • Status service options programmed for detection of correct service behavior or failure.
  – Verifications for the Status option:
    • HTTP on port 80: connectivity through GET request;
    • TCP on port 4331: MetaLib main application check through ordinary TCP socket communication;
    • TCP on port 7331: gateway evaluation also with socket connectivity test.
Conclusion

- CAPES’ Portal de Periódicos is a project that demands high availability of ExLibris tools MetaLib and SFX.

- The cluster solution implemented have the benefits:
  - Automatic service migration to the redundant node;
  - Possibility to use more than 2 nodes: 1 active and N passives;
  - Use forced migration to the passive node for scheduled maintenance (hardware or software) on the active node;

- For the future:
  - Implement a more refined solution that also includes MetaLib’s Oracle service
  - Extend the solution to SFX services;
Thank you!

Ronan Moraes – ronan.moraes@capes.gov.br
Gustavo Portella – gustavo.portella@capes.gov.br