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.

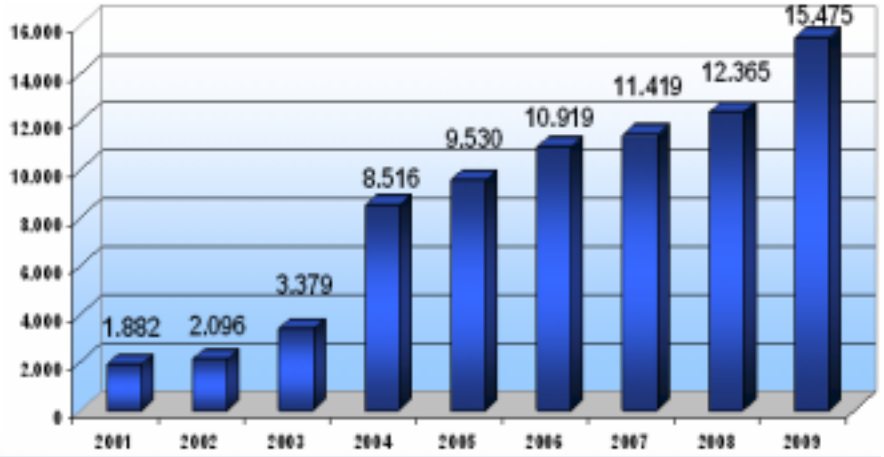
Portal de Periódicos CAPES

- 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.

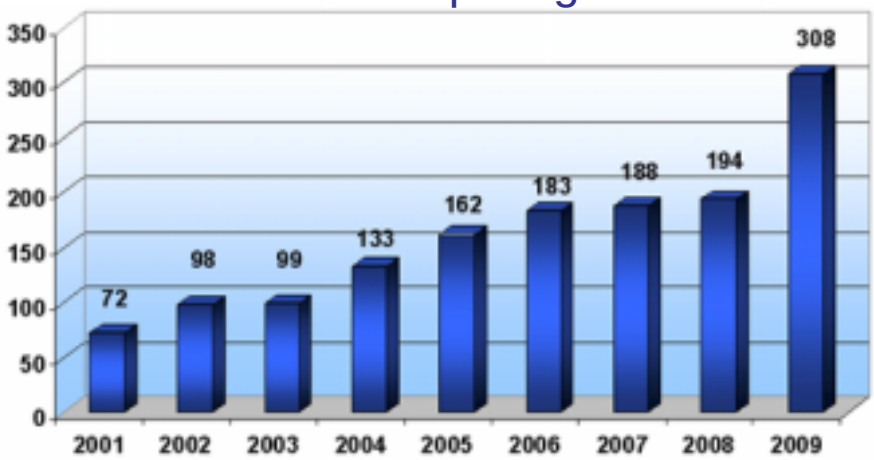
Portal de Periódicos CAPES

- 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.

Number of Available Titles



Number of Participating Institutions



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

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

* There is also a KVM module for remote maintenance.

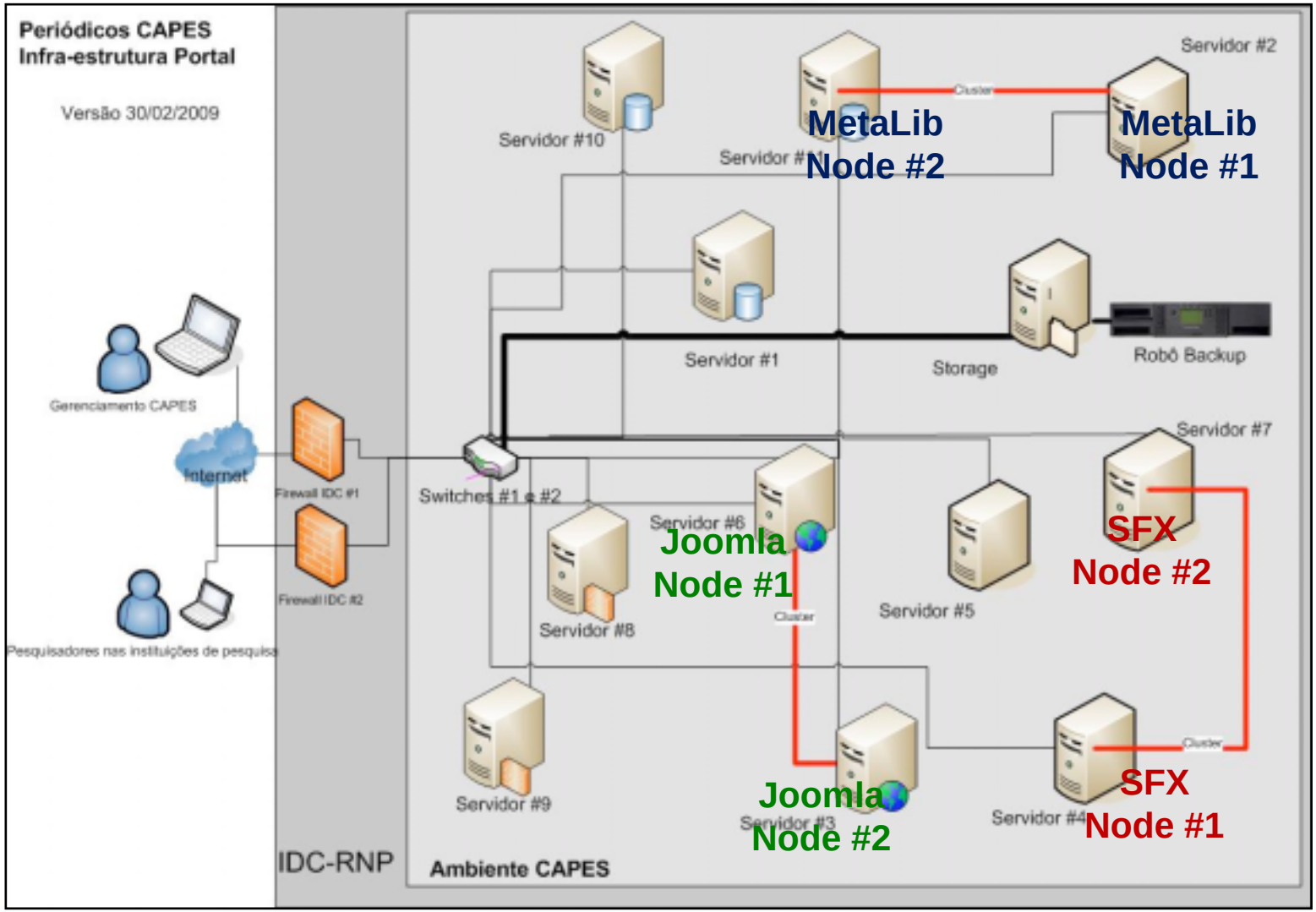
The New Portal

- 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



Clustering Solution

- 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 instalation “knows” only the virtual IP address;
    - File System virtualization: the MetaLib path is int 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.

Clustering Solution

- 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