Information Systems Technological Infrastructure in the Polytechnic Institute of Setúbal – Steps to Evolution

Nuno Pina Gonçalves [1], João Filipe dos Santos Daniel [2]

[1]Systems and Informatics Department, Superior School of Technology of Setúbal, Rua Vale de Chaves, Estefanilha 2910-761 Setúbal, Portugal, nuno.pina@estsetubal.ips.pt. [2] Information System Department, Superior School of Business Administration of Setúbal, Rua Vale de Chaves, Estefanilha 2910-761 Setúbal, Portugal, joao.daniel@estsetubal.ips.pt.

Keywords
IT infrastructure, technological evolution, integrated information systems

1. EXECUTIVE SUMMARY

Nowadays organizations and institutions face an important challenge. The technological infrastructure supporting the information systems is usually far from the ideal solution. Information Technologies (IT) projects tend not to have an easy relationship between the technological requirements and the financial resources. The start point of an acquisition is always complex. Sometimes it's necessary to reach a commitment solution between various fundamental factors all of them important and difficult to conciliate. Some of these factors are performance, availability, scalability, fault tolerance and backups.

The Polytechnic Institute of Setúbal (IPS) is a public higher education institution with five schools.

The information system project (SI-IPS) started in the Superior School of Technology of Setúbal (ESTSetúbal) six years ago, integrating and managing all the information related to students, teachers and collaborators. SI-IPS is now in expansion to the other schools. The project also integrates information related to human resources from all the schools. SI-IPS is an adapted version of the SIGARRA - Information System for the Aggregate Management of Resources and Academic registers, developed by Porto University and nowadays in use on all faculties of Porto University.

In the initial acquisition of the technological infrastructure supporting SI-IPS, it was necessary a commitment solution between the factors previously mentioned due to political and financial issues. The system visibility, credibility, mass usage and user dependency, empowered its own technological evolution. In the beginning, SI-IPS was a necessary investment, now it is a competitive advantage.

This paper describes the used methodology to achieve the solution for the actual requirements of SI-IPS technological infrastructure. The evolution of the available services usage and the applications supporting them are analyzed, using the factors performance, availability, scalability, fault tolerance and backups. This paper focuses our decision to follow the initial path, relying in Oracle 10g (Database, Real Application Cluster, Application Server, Forms and Reports) and Novell SUSE Linux Enterprise Server. A hardware and software technical solution is exposed in order to guarantee and improve the factors mention above. Along the time, some indicators related to the factors were analyzed to validate the decisions that were made for the technological infrastructure evolution and are presented in this paper, being proposed some future work.

2. SI-IPS

2.1. IPS

The Polytechnic Institute of Setúbal, a public institution of higher education, was created in 1979, and includes the Presidency Service, a Welfare Social Service and five Schools that offer a range of graduate courses in different areas such as Technology, Education, Business Administration and Health Care.

The IPS includes five high schools, four of which are installed on the campus of Estefanilha (ESE, ESTSetúbal, ESCE and ESS) in Setubal, and the fifth (ESTBarreiro) in Barreiro (34 km north from Setúbal). In total, students of IPS are around 6500, while the number of teachers is 505.

2.2. SI-IPS PROJECT (IPS INFORMATION SYSTEM)

Six years ago, IPS applications were islands of information. It was very difficult to gather integrated information of students and teachers. At the same time, the IPS and their schools needed management indicators that had to obtain them in various applications without integration and even with redundancy and duplication of information processes. It was not possible to the community to gain access to their information online and it would not be possible without a new integrated system based on the Web.

Most students of IPS live far from their schools, and many of them work during the day and study at night. It is a competitive advantage to give them access to information about classes and academic acts. Through various web portals, they do not need to be in school to have access to this information.

The idea to develop in house a new system was placed apart, because IPS did not have the sufficient skilled human resources to manage a project with all these requisites in useful time. All the systems with this kind of requisites usually take several years to become in production in a stable way. It was created a team that started to analyze all the available academic information systems all over the country. Two years later the team defined finally that there was only a really stable, solid and tested system that meet the identified requirements to be installed in IPS – SIGARRA.

SIGARRA - Information System for the Aggregate Management of Resources and Academic registers constitute a central tool to manage Superior educational establishments. The system development was started in Porto University in 1992 and has by now 13 collaborators. Nowadays it is in use on all faculties of Porto University and started its implementation in other public institutions of higher education such has IPS (Gonçalves 2007), (Ribeiro et al. 1999), (Ribeiro et al. 1999) and (Ribeiro et al. 1997).

SIGARRA's system architecture is an integrated Web based information system, a student's management applications (GA) and human resources management application (GRH), its main characteristics are:

- Information integration;
- Web interface;
- Modularity and configurability

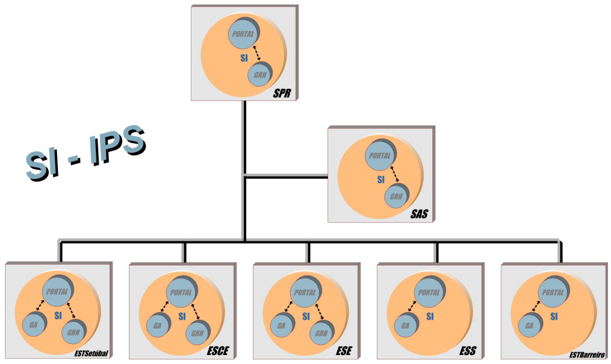

Figure 1: SI-IPS and its organic units

As IPS has five schools and two more organic units (figure 1), all of them used different procedures to the organizational processes. The main idea was to standardize the processes in all the schools. The adopted implementation methodology consisted in the migration of the old systems at each school one by one to minimize the impact in services and increase the team expertise. Beside the lack of standardized procedures another challenge was the fact that the legacy systems didn't have enough documentation, so the migration process needed to be validated several times to assure data consistency. While the migration was in progress, the web portal started being installed and customized.

3. CURRENT INFRASTRUCTURE

The SI-IPS is based on technology Oracle 10g and Novell SUSE Linux Enterprise Server 9 (SLES9).

The Oracle products used are:

- Oracle Real Application Cluster 10gR2 – Oracle Database Server 10gR2 (RAC-DB)
- Oracle Application Server 10gR1 – J2EE and Web Cache (AS)
- Oracle Application Server 10gR1 - Forms and Reports Services (FR)

The SI-IPS consists of the following servers:

- AS Server: DELL PowerEdge 2850 with SLES9 and Oracle AS;
- BDA Server: DELL PowerEdge 2850 with SLES9, Oracle RAC-DB e Oracle FR;
- BDB Server: DELL PowerEdge 2850 with SLES9 and Oracle RAC-DB;
- TESTES Server: DELL PowerEdge 2950 with SLES9 and Veritas Netbackup;

AS Server – Application Server is responsible for providing the various institutional portals (SI), as the application of human resources (GRH). Users access via Web (SI) or via intranet (in the case of GRH) to the SI-IPS through this machine, which handles these requests and makes communication with both of database nodes (BDA and BDB) where it gets the necessary information to the construction of all pages, dynamically generated by the system.

BDA and BDB Servers form the database nodes of the IPS cluster. In each of them is installed an instance of the database Oracle 10gR2 in Real Application Cluster configuration. This allows both instances to manage exactly the same information, offering two repositories identical and redundant to obtain information. All information on the SI-IPS - the construction of dynamic pages that form the portals, the student's academic data - is stored in the database of IPS by these two instances.

Oracle Real Application Clusters present in both databases, can maintain a single database IPS, using two separate instances on servers, which allows achieving greater redundancy and performance in access to data. In the event of a collapse of one of the instances, the other ensures that the system remains in production. One example is when it is needed to put one of the nodes in maintenance (Oracle Critical Updates, SLES9 Online Updates, among other). Another example is when in fact something goes wrong in a production environment and one of the nodes collapses by a problem in hardware or software. In these cases, the RAC gives us the assurance that even if a transaction goes wrong, we have another instance that keeps the system in production. The advantage is obvious for those who make the administration of systems, but is still more important to the IPS community that uses the features of the system.

The Oracle RAC automatically manages the distribution of load between the nodes BDA and BDB. This load increases with the number of applications (http) made by the server AS, and the applications (SQL) used by employees who manage the portals and applications relevant to their organic units. The automatic management in the distribution of the load by the instances of the cluster, minimizes the consumption of resources, and maximizes the performance in the responses to requests.

BDA server has also installed the Oracle Application Server 10gR1 Forms and Reports Services. It is through this service that the application for management of students is made available, operating separately, in a very closed and secure environment, accessible only by academics sectors from the various schools.

To maintain the IPS database, was acquired a SAN (storage area network), CLARIION CX300, which makes use of two processors to manage the data inputs/outputs, with 2.0TB of space in SCSI disks. This solution ensures high availability and security in the data it manages. The data of production systems are kept in groups of RAID1 for greater security in the data, while the other data (backups,…) are kept in groups of RAID5. This equipment is connected through a fiber-optic switch to all servers of the SI-IPS. The switch allows redundant paths which are configured with optical fiber from the servers to storage. The current configuration provides four different paths to/from each server. In addition to increased redundancy, this solution enhances performance.

The backups of SI-IPS are made through automated scripts that make use of Oracle RMAN and Oracle Export. The backup strategy is disk-to-disk-to-tape. First of all backups are made directly to the SAN, focusing performance.

TESTES server has installed Veritas Netbackup, configured for both of Oracle RAC nodes. The Veritas Netbackup handles the management of the SI-IPS backups to tape, making the copy from the SAN to the tapes that are housed in a LIBRARY PowerVault 132T, equipped with 9.6 GB of space on tape.

There is also a server in the IPS Presidency Service (off campus) to which is sent automatically and secure, copies of backups. In Figure 2 is the technological infrastructure of the current SI-IPS.

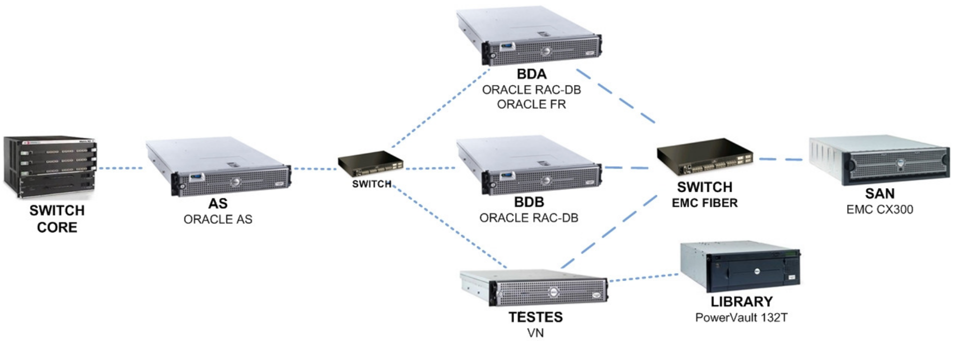

Figure 2 - Current technology infrastructure

4. EVOLUTION FACTORS

SI-IPS is in production in ESTSetúbal and Presidency Service. It remains the entry into production of institutional portals of four other organic units. Thus, a calculation was made on the future with a planned reality to a size three times greater, which is estimated to be the impact caused by putting into production the other portals. This calculation was based on the number of teachers, staff and students of each institution of higher education and organic units in the IPS.

4.1. System's access

The verification of the type and number of hits to institutional portals of SI-IPS was the first approach for analysis and estimate of resources required.

Thus, has seen an analysis of the actual number of accesses to the SI-IPS in different periods of time, for a month during the first quarter of this year, and during the last year now.

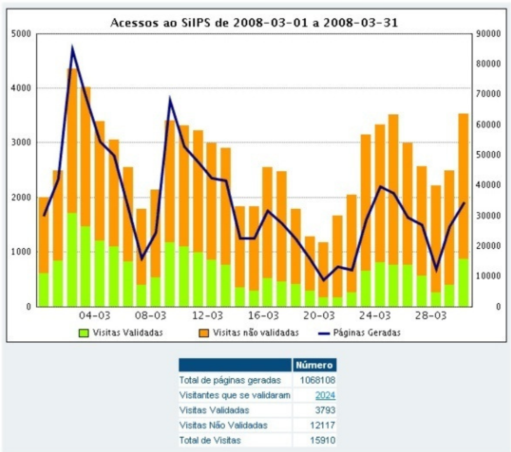

Figure 3 - Views during March 2008

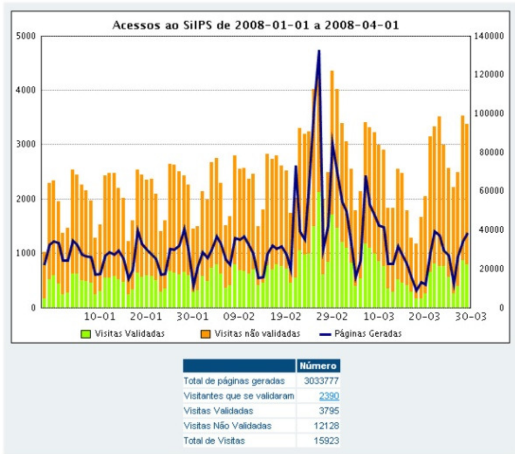

Figure 4 - Views in the first quarter of the year

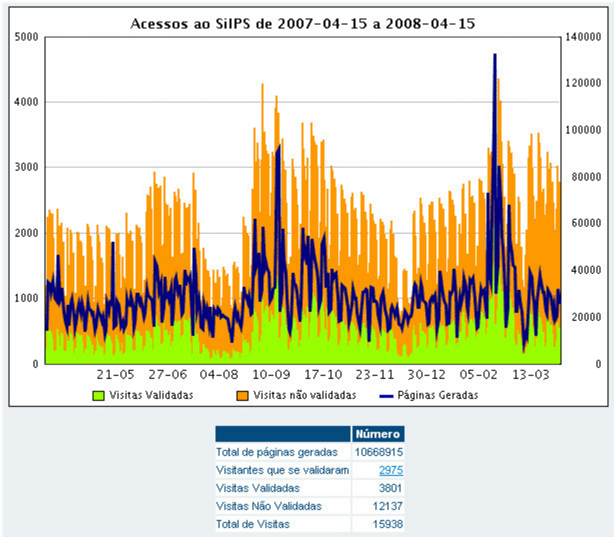

Figure 5 - Views during the last year

Note in Figures 3, 4 and 5 that there are some peaks of access to the system. These are due to the fact that at the beginning of each semester (February / March and September / October), are made the inscriptions and management of schedules. The schedules in ESTSetúbal are generated every six months based on the number of students who attended classes in the previous year, resulting in a provisional map of classes. In this sense, the system has an increase of hits not only for the generation and management of schedules, as well as for their future viewing by the students and teachers.

Through the values generated by the system and displayed in the previous charts, it was possible to project the reality of the future SI-IPS in terms of access and pages generated by the system. It is expected that in future, the numbers are increased by three times. Note that IPS users refer to the faculty, staff and students of IPS. Figure 6 makes the comparison between the number of users IPS and the pages generated by the system current and projected into the future.

	Today	Tomorrow
Number of Users	2.500	8.000
Pages generated / Year	10.668.915	32.006.745
Visitors / Year	15.938	47.814
Pages generated / Month	889.076	2.667.229
Visitors / Month	1.328	3.985

Figure 6 - Statistics of access to institutional portals

4.2. Performance

Performance was considered the most important indicator across the analysis because it is what creates greater concern about the future size of the community who use the services provided by SI-IPS.

Several factors were considered, however two of them were chosen because they are the most important and comprehensive: percentage of CPU Usage and percentage of Memory Usage (Iyengar et al. 1997), (Kimberly et al. 2000) and (Squillante et al. 1999). The server AS, which supports the Oracle Application Server 10gR1 - J2EE and Web Cache, and is responsible for managing the access to institutional portals, filed last year an increase of 2% to 17% of CPU Usage, and an increase of 7 % To 9% of Memory Usage. With the values obtained, we estimate a CPU Usage of 51% and a Memory Usage of 15%, when the SI-IPS entries in production for all organic units of the IPS.

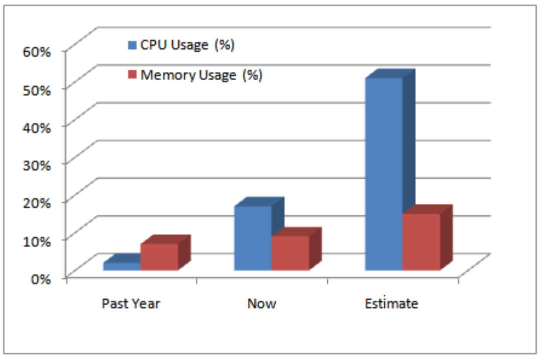

Figure 7 - Analysis of the consumption of resources of the AS server

BDA server, which supports the Oracle Application Server 10gR1 - Forms and Reports Services, is responsible for managing the application for management of students, presented last year an increase of 3% to 14% of CPU Usage, and an increase of 4% to 6% of Memory Usage. With the values obtained, it is envisaged a future CPU Usage of 42%, and a Memory Usage of 12%. This server also supports one of the instances of the database of the cluster, so that the values found should add half of the occupation of resources consumed by Oracle RAC.

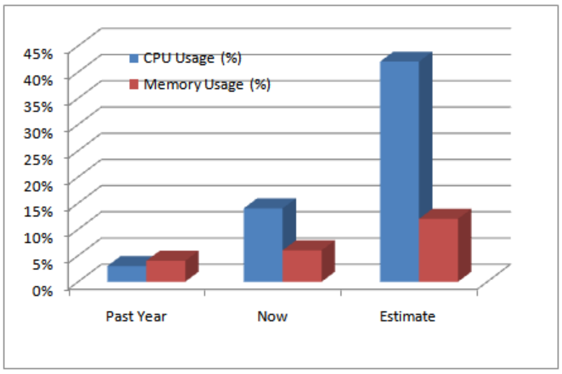

Figure 8 - Analysis of the consumption of resources for BDA server - Oracle Forms and Reports Services

BDA and BDB servers, which support the Oracle Real Application Cluster (RAC) 10gR2, and are responsible for managing the two database instances, presented last year an increase of 12% to 18% of CPU Usage, and a increase of 28% to 32% of Memory usage. With the figures obtained, it is envisaged a future of 54% of CPU Usage, and a Memory Usage of 44%.

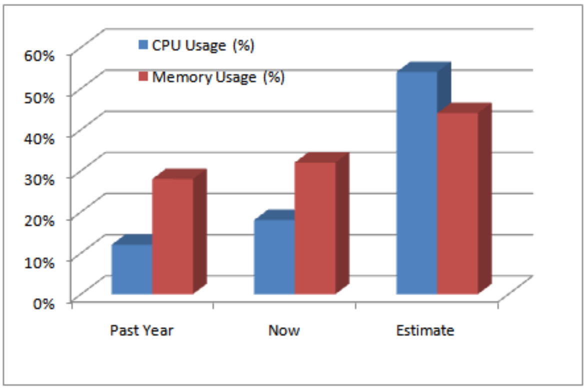

Figure 9 - Analysis of the consumption of resources and BDA BDB servers (RAC)

4.3. Fault tolerance

With the increase in the number of users of the system, the risk of failure increases. The increase in the number of requests made to the system through its various applications means an increase in resources needed for servers, which leads to the possibility of collapse by any of the servers that fail to respond to a certain volume of requests. To avoid such situations, are privileged some scenarios with tolerance to failures.

The solution to the SI-IPS passed by using clusters on most critical systems, and adding new servers to provide services that were in the same machine. The goal was the reduction of resources consumed by machine, providing them with better performance and less likely to come into collapse by over-processing.

One of the points to review was the access to institutional portals, managed only by a single machine which then distributes the requests by the database nodes. Adding a server and configuring both in active-active cluster allows users to have two redundant paths redundant to access SI-IPS. This was one of the claims on the initial purchase, but it was not possible for financial reasons. Thus, it will reduce by 50% the likelihood of losses of access to institutional portals, caused by deficiencies in these servers.

Oracle RAC is the most important in the whole infrastructure of the SI-IPS and is responsible for the management of more critical information. The current environment already contains a cluster with two database instances. But we intend to take a step forward in terms of technology, by adding another database instance the cluster. This step is essential to prevent and minimize future collapses in the cluster, reducing by 33% the likelihood of losses in Oracle RAC, caused by deficiencies in our databases.

The application for management of students is available through the Oracle AS installed on one of the cluster of database, and as such, it is also our desire to move this application to a dedicated server. In terms of access problems, the application will no longer be dependent on the smooth operation of the database node where it is installed.

4.4. Virtualization

Having a system in production that will give us the guarantee of reliability is critical, and this is the priority when is time to make an investment towards the evolution of the technological infrastructure. But with time, it is necessary to have a similar system to work in parallel, where it is possible to make tests, without the possibility of a system in production enter on a collapse by any unsuccessful test.

Examples of useful tasks in a testing environment are:

- Implementation of SLES9 updates;
- Implementation of Oracle updates;
- Administration of various Oracle products;
- Create a test database, identical to the one in production, for the employees of the team who need to do massive and heavy queries, or develop applications with access to the database;
- Test scenarios of severe collapse in the IPS database, with the recovery of all information through the use of backups.

By performing these tasks in a test environment, it will withdraw some consumption of resources in the production environment, and it is possible to test the system intensively, and above all, there will be no risk to the system fails in production due to some unsuccessful test.

The implemented solution involves the use of VMWare to create a virtual scenario that represents all the infrastructure of the SI-IPS, using a single server (TESTES) where you can install and configure various virtual servers corresponding to each of the servers in production.

5. STRATEGY FOR ACQUISITION

The process of acquisition of equipment and applications for a project with the size of the SI-IPS is costly and there are many factors involved. The break with the past and resistance to change are two factors, but there are more such as resistance to spending and ROI - "Return On Investment" obtained with the acquisition, as well as factors associated with organizational culture and context of the institution against possible external financial support - such as Virtual Campus (e-U).

Even with the whole scenario of financial support at European level since the e-U, yet we must think very well where the money goes to. It is therefore necessary to make its division by network equipment, servers, software, facilities, training and e-learning. Thus, it is important to guarantee that the decision-makers are aware of the advantages of the proposed solution.

In the project SI-IPS, there were only directly addressed the purchases of servers and software - Oracle, configurations, facilities and training.

With the financial ceiling that was available, the original solution would always be a compromise solution, a solution that would allow, taking into account the factors of development previously discussed, provide the infrastructure for a compromise between the performance chronological and the reality of implementation. Thus, after the first year, putting into production the applications on ESTSetúbal, both the decision makers and users were loyal and believed that these were really an added value for the institution. In this context it was possible to justify, with the analysis in this article of the factors of development, the need to provide the infrastructure for a few more improvements.

Then the servers to invest were placed in ascending order of importance:

1. BDC Server, Oracle Application Server Cluster 10gR2 - Oracle 10gR2 Database Server, to add another instance of the database to existing cluster;
2. AS2 Server, Oracle Application Server Cluster 10gR2 - J2EE and Web Cache, to form a cluster with the current server;
3. GA Server, Oracle Application Server 10gR2 - Forms and Reports Services, to support the application for management of students;

As such consultations were made for the supply of equipment, based on the available budget, dividing it by servers needed in order to meet the priorities defined previously. In figure 10 is the percentage of investment made by each of the servers purchased.

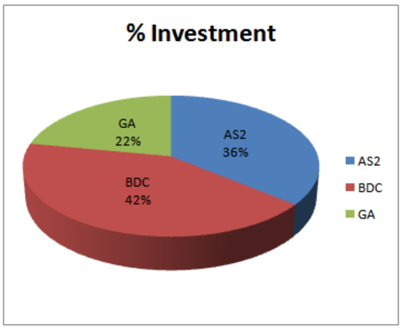

Figure 10 - Percentage Investment per server

6. FUTURE INFRASTRUCTURE

The future infrastructure will also be based on Oracle 10g and Novell SUSE Linux Enterprise Server 9 (SLES9).

The Oracle products to be used are:

- Oracle Real Application Cluster Server 10gR2 - Oracle 10gR2 Database Server (RAC-DB)
- Oracle Application Server Cluster 10gR2 - J2EE and Web Cache (ASC)
- Oracle Application Server 10gR2 - Forms and Reports Services (FR)

The SI-IPS will consist of the following servers:

- AS Server : Dell PowerEdge 2850 with SLES9 and Oracle ASC;
- AS2 Server: Dell PowerEdge 1950 with SLES9 and Oracle ASC;
- BDA Server: Dell PowerEdge 2850 with SLES9 and Oracle RAC-DB;
- BDB Server: Dell PowerEdge 2850 with SLES9 and Oracle RAC-DB;
- BDC Server: Dell PowerEdge 2850 with SLES9 and Oracle RAC-DB;
- GA Server: Dell PowerEdge 1950 with SLES9 and Oracle FR;
- TESTS Server: Dell PowerEdge 2850 with SLES9, Veritas Netbackup and VMWare;

The AS and AS2 servers form a cluster, enabling a balanced distribution of the load caused by access to institutional portals. In this case the cluster examines the state of both instances and chooses one with more resources available. The information provided is identical in both, obtained with higher performance, due to have the double of resources to meet the requests. Thus, users who access the SI-IPS by web (SI) or intranet (in the case of GRH), without realizing are forwarded to one of two machines, each request they make to the system. The Oracle Application Server Cluster will communicate with the Oracle Real Application Clusters. This last one chooses one of the database nodes (BDA, BDB, CDB) to obtain the necessary information to the construction of the pages, dynamically generated by the system.

The BDA, BDB BDC servers will form the nodes of the IPS cluster, where in each node will be installed an instance of the Oracle Database 10gR2 with Real Application Cluster, maintaining a single database IPS. This will allow even greater redundancy and performance in access to the database. In the event of a collapse of an instance, we have two others to ensure that the system remains in production. Thus the Oracle RAC will have one more instance of the database which can automatically manage the load distribution. The load will decrease by machine in the same proportion to which we were added to the cluster (Torsten et al. 2001).

Thus, compared to the past, we will have two paths to access institutional portals (one more than previously), and we will have six paths to access the databases from the portals (we had previously two).

The SAN (Storage Area Network) CLARIION CX300 will be configured for the new node of the cluster database to access the corresponding partitions. This means adding a set with 4 new routes from/to the new instance of the database. Thus, we will pass from eight to twelve routes of fiber optics, from the SAN to an instance of the database IPS. Again, as well as greater redundancy, this solution enhances performance.

TESTES server remain the machine to manage the backups to tapes through the Veritas Netbackup, which will be set for the three nodes of Oracle RAC. The novelty will be the installation and configuration of VMWare to create a virtual scenario that represents the infrastructure of the SI-IPS, with all future servers represented by a virtual instance, in a closed environment, with identical environment. This solution will enable advanced tests on the platform SI-IPS, such as the preparation of the Oracle environment for new updates of software, information retrieval without the need for production database, installation and testing of new software, among others. In figure 11 is the technological infrastructure to be implemented in the future SI-IPS.

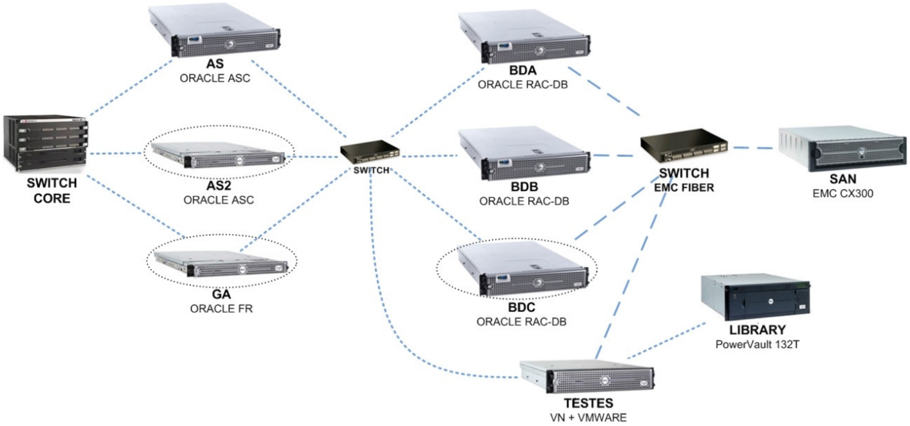

Figure 11 - Future technology infrastructure

7. CONCLUSIONS AND FUTURE WORK

This article discussed the strategy to meet the technological solutions adopted to meet the needs of today's technological infrastructure to support the project SI-IPS. The article examined the evolution in the use of available services and applications supporting it, taking into account the factors of development, performance, availability, scalability, fault tolerance and backups. The technical solution was exposed in terms of hardware and software that guarantees an improvement in terms of factors of development that were discussed. Indicators were analyzed, over time, in order to validate the choices made to evaluate the results obtained with the development of infrastructure technological structure.

It is not always possible at once to purchase hardware and software solutions that are the ideals, however, whenever possible, the best solution is to phase the purchases over time, given that the first purchase will have to respect a compromise between performance, availability, scalability, fault tolerance and the backups, and the financial ceiling available. In the next steps will be necessary to first prove that the application solutions are really effective, loyalty their users and convince decision makers of its added value for the organization. Thus, it will be possible in a phased manner, to provide the improvements to the infrastructure and conditions for achieving an ideal solution. However, it is noteworthy that these steps must be relatively close (maximum periods of one year) to ensure that the hardware and software customers do not become obsolete.

As future work, will be implemented technological solutions proposed for the infrastructure and new indicators will be analyzed to support and validate such solution. At the same time, there is a beginning of a new cycle, based on further improvements to the technological infrastructure and acquisitions.

8. REFERENCES

Gonçalves, N. Fernandes J. (2007). *Portal Implementation in the Superior School of Technology Setúbal – Key issues and Evaluation*. EUNIS 2007 - International Conference on European University Information Systems.

Iyengar, A. MacNair, E. Nguyen, T. (1997). *An analysis of Web server performance*. Global Telecommunications Conference, *GLOBECOM '97*: IEEE, vol.3, 1943-1947

Kimberly, K. David, P. (2000). *Towards a simplified database workload for computer architecture evaluations*. Workload Characterization for Computer System Design. Kluwer Academic Publishers

Ribeiro, L. David, G. (1998). *Sistema de Informação da Faculdade de Engenharia da Universidade do Porto: Resumo e Justificação. Relatório Técnico Faculdade de Engenharia da Universidade do Porto.*

Ribeiro, L. David, G. (1999). *SiFEUP - um Sistema de Informação ao Serviço da Faculdade*. Sistemas e Tecnologias de Informação: Desafios para o Século XXI

Ribeiro, L. David, G. (1999). *Getting Management Support from an University Information System.* EUNIS 99 Information Technology Shaping European Universities. 132-137

Ribeiro, L. David, G. Azevedo, A. Santos, M. (1997). *Developing an Information System at the Engineering Faculty of Porto University*. Proceedings of the EUNIS 97 - European Cooperation in Higher Education Information Systems. 282-287

Squillante, M.S. Yao, D.D. Li Zhang. (1999). *Web traffic modeling and Web server performance analysis. Proceedings of the 38th IEEE Conference on Decision and Control. vol.5,* 4432-4439

Torsten, G. Klemens, B. Hans-Jorg S. (2001). *High-Level Parallelization in a Database Cluster: A Feasibility Study Using Document Services.17th International Conference on Data Engineering (ICDE'01).*