Studio per lo sviluppo del prototipo del centro dati scientifico ASPIS

ALLEGATO 6

The Space Science Data Center

The Space Science Data Center (SSDC) of the Italian Space Agency has its roots in the early '90s with the Data Center of the BeppoSax X-ray astronomy satellite, then evolved into the ASI Science Data Center (ASDC), established in 2000 as a multi-mission facility linked to high-energy astrophysics data managing and exploitation.
The two distinctive scopes of SSDC are the archiving, analysis, distribution of space mission data, and the development of scientific tools to extract high-level scientific information from the data of interest. These goals have been traditionally accomplished by following an "operative approach" involving both scientists and industrial partners. The center is constituted by personnel from the Italian Space Agency, the National Institute for Astrophysics (INAF), and the National Institute for Nuclear Physics (INFN), with an Information and Communication Technology (ICT) support provided by industrial partners (that currently are Telespazio and Serco).
This approach allowed for the development of software tools oriented to the real needs of the scientific community of reference. Some of these tools have become milestones for the community, such as the Multi Mission Interactive Archive (MMIA), the SED (Spectral Energy Distribution) Tool and MATISSE. It is located at the Italian Space Agency Headquarters in Rome (Italy).

1. The SSDC Infrastructure

Personnel and competencies

The scientific competencies at SSDC are primarily provided by INAF, INFN and additional institutions participating through formal agreements with ASI. The current scientific staff is approximately composed of 40 persons.

The ICT support is provided by a dedicated industrial contract and is supplied with the following methods:

- On-call through support requests created by the scientific personnel
- By scheduled activities previously agreed with the ASI management or scientific staff
- By autonomous task derived from specific responsibilities assigned to ICT staff (i.e. system maintenance and monitoring)

Currently, the ICT support staff is composed of 7 persons.

Considering the competencies and the roles defined by the industrial contract, the ICT support can be involved in the following task:

- System design for new applications/tools, including analysis of requirements, high-level architecture definition, data model definition
- System maintenance of hardware and software components
- User support related to the SSDC infrastructure, including the configuration of network policies, workstation configuration, access to internal and external services
- Operational tasks targeted to specific missions/projects
- Design and development of software components, including interfaces to commonly used scientific data formats, instrument simulators, web-oriented tools for data visualization and analysis, data processing pipelines, science products archives, astronomical catalogs and databases

- Application of FAIR principles and Virtual Observatory techniques to scientific products, including the design and the development of data distribution services
- Refactoring of existing software components and/or their integration into the SSDC infrastructure

The ICT support team has solid expertise on the following list of tools and technologies:

- MySQL / MariaDB
- PostgreSQL
- VMWare/Vsphere
- Apache Tomcat
- vsftp/proftp
- Iptables
- Nagios
- DHCP Server
- Mail Server
- Nfs Server
- Samba Server
- CVS
- GIT

The main coding/scripting languages adopted by the ICT support team are:

- C/C++
- Java/JSP
- Javascript
- PHP
- Perl
- Fortran
- Browse
- CSS
- HTML
- SQL
- Linux shell scripting

Hardware resources available for ASPIS

Computational resources for ASPIS will be part of a shared virtualization framework.
This virtualization framework is based on a cluster of four nodes running the VMware ESXi hypervisor, the version in use is 6.5.
The available hosting infrastructure is located in a dedicated room with environmental thermal control and UPS devices for emergency power supply. This cluster of physical servers in rack configuration is managed by SSDC personnel and regulated by an industrial contract defining service level agreements and clear responsibilities.
The cluster configuration is composed of three "Dell Poweredge R630" nodes and one "HPE Proliant DL 380 G9" node, offering a total of 7 CPU and 220 logical cores, with a RAM amount of 1792 GB total.
These system resources are shared among running virtual machines and the assigned amount of resources for each machine can be dynamically configured, depending on the current requirements arising from the related processing activity.
The amount of storage available for ASPIS will be part of the storage system described below.

The storage system is based on a set of Oracle ZS3-2 storage servers mounting a total of 400 TB shared storage space adopting the ZFS file system for high availability. The available storage space is obviously shared among the several active projects hosted into the SSDC infrastructure, tailoring the virtual storage allocated to each virtual machine set required by a given project. Each virtual machine can access its reserved storage area (defined as "share") which size can be dynamically configurable, storage sharing can also be enabled between different machines, whenever required.
The storage system is accessed through two redundant controllers to guarantee business continuity. It is also possible to configure the system to accurately determine the grade of reliability and performance for each application.
In the actual configuration, two types of redundancy are provisioned:

- *Mirrored*: high performance and high redundancy, obtained by doubling the assigned physical storage size with respect to the required logical storage size
- *Double Parity*: medium performance and high redundancy, obtained by considering spare disk devices for fault tolerance

The allocated storage is divided into logical groups defined as "Projects". Each Project group has its own set of shares.
For each share, the policy specifying the available permissions is defined and the storage size assigned can be dynamically increased or decreased.
Many export protocols are provisioned: NFS, SMB, FTP, SFTP. For each chosen protocol, a security policy can be defined by assigning read/write/access permissions on a hostname/IP criterion.
The dedicated network connection to the WAN is based on a redundant network access point maintained by the GARR NOC in collaboration with the ASI IT department. The current bandwidth for Inbound/Outbound traffic is 1.0 Gbps and is currently being upgraded to 10 Gbps.
As a tentative allocation, we foresee to dedicate 32 cores, 256 GB RAM and 50 TB of storage for ASPIS. However, this is meant only as an estimate to illustrate the expected computational resources available to properly sizing the proposal and it is not meant as a mandatory requirement to be fulfilled. Detailed resources will be discussed and agreed with the proponent team.

Software resources and internal interfaces

Due to the nature of the hosting infrastructure, the hardware virtualization technology determines the high flexibility available to any hosted software product. The single requirement is the need to use an operating system compliant with the VMWare ESXI virtualization technology version 6.5, meaning that a very broad range of solutions can be adopted (practically any operating system family/vendor). Further documentation can be found in the VMWare Compatibility Guide and at the following URL to search for configuration compatibility (considering version 6.5 of ESXI technology): https://www.vmware.com/resources/compatibility/search.php
Analogously, no specific interface requirements exist about the running environment, including required software libraries, operating system dependencies, build environments. Once the virtual machine is allocated to the project, following the resource specification provided it can be autonomously configured and managed by the project responsible.
It is also possible to migrate an externally defined virtual machine into the SSDC virtualization cluster, assuming the compatibility of its operating system, as stated above.