ISTITUTO NAZIONALE DI FISICA NUCLEARE

Sezione di Bologna

INFN/code-xx/xxx
26/02/2016

CCR-xx/2016/P

INTEGRATION OF HETEROGENEOUS CLOUDS IN THE INFN INFRASTRUCTURE

Giuseppe Andronico[1], Stefano Bagnasco[2], Giacinto Donvito[3], Claudio Grandi[4], Dario Menasce[5], Massimo Sgaravatto[6], Daniele Spiga[7], Federico Zani[8]

[1)]INFN-Sezione di Catania, [2)]INFN-Sezione di Torino, [3)]INFN-Sezione di Bari, [4)]INFN-Sezione di Bologna, [5)]INFN-Sezione di Milano Bicocca, [6)]INFN-Sezione di Padova, [7)]INFN-Sezione di Perugia, [8)]INFN-Sezione di Roma Tor Vergata

Abstract

INFN has a project to build a Cloud infrastructure, the INFN Corporate Cloud (INFN-CC), to help rationalising the management of its computing resources. The INFN Corporate Cloud will provide a core set of uniformly managed, highly reliable resources deployed on a small number of sites. While the INFN-CC would in principle be able to address all use cases, we foresee that a significant amount of resources will be made available to INFN by other means. In this document we try to provide the criteria to discriminate if the exploitation of an external Cloud infrastructure by INFN is worth the investment. This may represent a set of guidelines for who is designing such an infrastructure and also for the management who needs to decide whether to fund a project or not.

Published by **SIS–Pubblicazioni**
Laboratori Nazionali di Frascati

1 EXPLOITING CLOUD TECHNOLOGIES ON THE INFN INFRASTRUCTURE

Cloud technologies offer new opportunities to rationalize and improve effectiveness of the INFN computing infrastructure, both for what concerns basic services and scientific computing. The INFN *Commissione Calcolo e Reti* (CCR) has been evaluating and prototyping Cloud technologies since a few years and the recent approval of the INDIGO-DataCloud EC project, coordinated by INFN, represents an important achievement and a unique opportunity for this activity. A document[1)] is in preparation where we will present a possible architecture for a Cloud-based INFN computing infrastructure and a realistic roadmap for implementing it.

2 USE CASES

In the above-mentioned document[1)] we define the use cases for which the use of Cloud technologies may be useful. They are summarized here in order to better contextualize the process of integrating heterogeneous Clouds.

2.1 Central IT services

This use case addresses the provisioning of central services such as: Authentication and Authorization services, DBMS, Web applications, etc.

2.2 Local IT services

This use case addresses the physically delocalized hosting of services addressing the needs of individual INFN units and managed by personnel of the units themselves.
The range of services that can exploit a distributed Cloud infrastructure is wide: mail services (mail relays, mailing lists, IMAP servers), web servers and web applications, collaborative tools (wiki, issue tracking, software repositories), test and development, training and education, etc.

2.3 Scientific computing

This use case is the same currently addressed by Grid computing: provisioning of massive computing power in batch-mode with dynamic resource allocation, provisioning of storage capacity and data distribution over the WAN, automation of complex workflows.

2.4 Support to analysis

This use case addresses the needs of end users and of small collaborations, including the so-called “last-mile” of analysis. It includes: provisioning of interactive computing power, provisioning of local batch computing power (e.g. Batch System-as-a-Service), ad-hoc solutions for specific needs; personal and group storage.

3 THE INFN CORPORATE CLOUD

Even though the overall architecture will be compatible with a heterogeneous environment,

the existence of a homogeneous infrastructure will offer additional functionalities. For this reason we foresee the creation of a multi site INFN Corporate Cloud (INFN-CC). The INFN-CC architecture is described in a separate document[2)].
INFN-CC tightly couples a few homogeneous OpenStack installations that share a number of services, while being independent, but still coordinated, on other aspects.
Users will access INFN-CC via a web dashboard or via APIs as if accessing a single OpenStack installation, but will be able to deploy resources in multiple sites.
The focus of INFN-CC is on resource replication, distribution and high availability, both for network services and for user applications.
INFN-CC represents a single, though distributed, administrative domain.
The architecture of INFN-CC is particularly fit for a wide range of use cases where a strict relation exists among resources that are distributed over different sites.
Most of these use cases are related to the delivery of IT services for the INFN community, be they of local interest for users belonging to a single INFN structure or of general interest for the whole community.

4 HETEROGENEOUS CLOUDS

The INFN Corporate Cloud will provide a core set of uniformly managed, highly reliable resources deployed on a small number of sites. While the INFN-CC would in principle be able to address all use cases, we foresee that a significant amount of resources will be made available to INFN by other means. Examples are infrastructures (co-) funded through external projects (e.g. ReCaS), local agreements with Universities or public administrations (e.g. the Clouds in Padua and Turin), managed via Cloud middlewares other than OpenStack or simply that is not convenient to integrate in the INFN-CC.
We are not addressing here strategic infrastructures for scientific computing that, due to their specificity, are managed independently and for which INFN has other means of control (e.g. Tier-1 and Tier-2 sites). Furthermore, we are not addressing the exploitation of opportunistic resources but rather of infrastructures where INFN has some level of ownership.
These infrastructures still retain a high value for INFN not only because they may incorporate additional resources at a convenient cost but also because they represent a sort of biodiversity on which it is possible to maintain and build distributed know-how, study alternative solutions and maintain connection with a wider range of communities.
Nevertheless not any cloud infrastructure is worth the effort of being built and exploited by INFN and in general the number of such infrastructures should not be excessive. In this section we try to provide the criteria to discriminate if the exploitation of an external Cloud infrastructure by INFN is worth the investment. This may represent a set of guidelines for who is designing such an infrastructure and also for the management who needs to decide whether to fund a project or not.
We would like to stress however that the INFN Cloud community inside the *Commissione*

Calcolo e Reti has the expertise to help in the decision process for any specific technical case.

5 EVALUATION OF SUSTAINABILITY AND CONVENIENCE

In the following we discriminate between the IT infrastructure (computers, storage, network) and the basic infrastructure (walls, cooling, electric power and manpower support).
The Director(s) of the involved INFN structures must guarantee that coverage of the basic infrastructural costs will not be explicitly requested to INFN. This means that either the infrastructure is already available (e.g. the site is already an INFN official computing centre of adequate size) or the costs are covered through ordinary funding of the structure or by the project itself. This infrastructure must be suitable for hosting the IT infrastructure.
The effort requested to INFN, not only for the creation but also for the maintenance of the IT infrastructure, should be made explicit in a Memorandum of Understanding together with the expected benefits. The proposed MoU, together with any other relevant agreement of INFN with the involved partners that may have impact on the evaluation of the proposal, should be made available to the INFN management in order to determine the economic and scientific benefit of the project.

6 TECHNICAL REQUIREMENTS

The resources provided on the Cloud must be identifiable as in use by INFN and accounted accordingly.
The Cloud must comply with the security rules defined by the INFN Security Group and by the Harmony group.
It is recommended that the Cloud supports a Federated Identity Authentication and Authorization system.
The access to the wide area network should follow the GARR rules: access to certain networks should be granted only to users that possess the appropriate rights. In particular if the Cloud is hosted on a site connected to LHCONE or LHCOPN, the access to these networks must only be granted to the INFN Virtual Organizations that are entitled to use them.
The following are specific requirements that must be satisfied to use the Cloud (according to the MoU) for the specific use cases described in the previous sections of the document.

6.1 Scientific computing

If it is foreseen that the Cloud resources can be pledged to VOs, then the allocation and the access to these resources must be possible via the mechanisms and with the performance agreed with the VOs themselves. More explicitly:

- The site must offer at least one of the interfaces considered by each supported VO for resource allocation and access (e.g. EC2, Grid CE interface, SRM,...).
- The access from the local computing resources to the local storage of the VOs must

comply with the VO requirements for protocols and performance; in general the access to the local area network must be possible with the required performance.
- The access to the wide area network must be possible with the required performance, in particular access to the VO storage from the WAN must be guaranteed with the required bandwidth.
- The support to the VOs must follow the usual communication channels and the site must provide local or remote contact persons.

6.2 Support to analysis

We expect these Clouds to be able to provide support for analysis also to non-local users. The exploitation model is that the resources are not allocated to individual users or group of users, but to the administrators of other INFN structures or VO application managers that in turn allocate the resources to their users and support them.
The Cloud site must have the technical competences to address the requests by the designated contact people or other INFN administrators. This may concern for instance the deployment of specific services like virtual private networks, Batch-System-as-a-Service, configuration of the network to allow access to remote storage, etc… The managers of the Cloud must be available to discuss with them to find suitable solutions. No interaction between the administrators of the Cloud and the remote end-users is foreseen: any interaction of the users is done through their designated contact people or local administrators.

6.3 Central and local IT services

We expect that the only infrastructure that will support central and local IT services for other sites (other than the two already described) is the INFN-CC.
Nevertheless we expect that the local IT services of the site that hosts the Cloud may be hosted on the Cloud itself. In this case the maximum CCR contribution for the maintenance of the Cloud will be the amount provided if the services were managed in a traditional way.

7 EVALUATION OF COSTS

As an estimate of costs involved in the adoption of Cloud technologies, we take as a reference the INFN-Torino deployment, originally meant to virtualize a Tier-2 and then expanded to host a number of different applications. Since the Torino infrastructure was designed a few years ago with several constraints, we assume an updated design to take into account some experience gained and a recently published OpenNebula reference architectural document[3)].
The described deployment will have 100 hypervisors, but will easily scale up to a few hundreds. As an order-of-magnitude estimate, 100 hypervisors could provide the equivalent of 2000 job slots and associated services.
In the following, only extra costs for the IaaS management are included; we do not include any resource (worker nodes, persistent disk space, networking) that is provided to users and would be needed regardless of the adoption of Cloud technologies. All prices quoted below include VAT.

The minimal design includes one server for the OpenNebula front-end and its DB, a backend storage with two redundant servers with 10Gb/s connectivity and one extra network connection (alongside the public and private networks) across machines to serve as service and storage network.
A more realistic deployment would include:

- Two redundant servers for the front-end and the DB.
- A resilient backend storage with two redundant servers, to host both the "image data store" (image repository) and the "system data store" (holding running instances). To reduce costs, the system data store is directly exported only to a fraction of the hypervisors (a few units, say 5), thus allowing live migrations of "service" VMs only.
- One extra physical server to run infrastructure monitoring services and configuration management tools. Since such services are commonly used in most non-virtualized data centres, this is not to be considered an extra cost.

Besides the normal network connectivity needed by service and computing nodes, the reference architecture recommends one extra 1 Gbps port for the storage/service network per each hypervisor (say 100 including the service hypervisors). We decide to adopt this configuration for the service hypervisors only, collapsing the service network onto the private one for the others.
Taking as a benchmark the CONSIP [CONSIP] standard 1U servers, the two front-end nodes will cost about 3.2 k€ each. Adding the 10 Gbps NICs and the transceivers, the storage servers will cost about 3.8 k€ each.
For the storage backend, as a very rough estimate we again use standard equipment from the CONSIP catalogue: building a fully double-chain storage system by adding one JBOD expansion box to each of the storage servers, providing 14.4 TB each with 10kRPM SAS disks, costs about 25.5 k€ including the SAS cards. A cluster file system such as GlusterFS or GPFS will be needed to manage the dual storage; we assume the free open-source GlusterFS. In this way we have a very resilient high-performance storage backend, which can serve a deployment with a few tens of "migratable" virtual machines.
Extra connectivity includes two 10 Gbps ports for the storage servers (at an estimated cost of about 0.5 k€ each including transceivers) and one extra 1 Gbps port for each service hypervisor for the service and storage network, which would cost about 50 €/port (assuming the site already has a modular core switch for the public and private networks).

Front end servers	6.4 k€
Backend storage, incl. servers	33.1 k€
Extra network ports	1.25 k€
Total	40.75 k€

This is an order-of-magnitude estimate of the extra costs that are needed to build a very

resilient and high-performance infrastructure, with some scalability for the future, using discrete readily available components. In several real scenarios such costs can be reduced; for example, the backend storage may be obtained as a fraction of a larger storage that will be used also to provision persistent storage to tenants (e.g. for the home directories of virtual farms, which is not an extra cost since such space needs to be provided even without a cloud infrastructure), or a lower grade of redundancy may be deemed sufficient for the storage infrastructure. Also, the size of the storage is overestimated for many applications. Another often-adopted solution is to use two virtual machines for the front-end, which actually does not need much computational power (2-4 cores according to the recommended reference architecture), or to dispense with high availability for the front-end. On the other side, the cost may get higher should one want to use a Ceph object storage cluster for the image data store, independent of the shared system data store, as suggested in the reference architecture for larger deployments. We did not find the extra complexity (and costs) justified in the actual INFN-Torino deployment, which is fully based on GlusterFS.
So we can take 40 k€ as a pessimistic estimate of the extra costs needed to set up a highly resilient and performing IaaS system serving up to some thousands of job slots or equivalent; such equipment may have an expected life of 5 years.

8 ACKNOWLEDGEMENTS

The authors of this document are a subset of the authors of [1)] from where the introductory text has been taken.

9 REFERENCES

(1) Temporary link: https://docs.google.com/document/d/10Tuq19N_ndlmmNSj8k26Yh6WNRdr_etVQiV4HO_a_bM/edit?pli=1#
(2) Temporary link: http://wiki.infn.it/cn/ccr/cloud/infn_cc/documento_architetturale
(3) https://support.opennebula.pro/hc/en-us/articles/204210319