

# Merging OpenStack-based private clouds: the case of CloudVeneto.it

*Paolo Andreotto<sup>1</sup>, Fabrizio Chiarello<sup>1</sup>, Fulvia Costa<sup>1</sup>, Alberto Crescente<sup>1</sup>, Sergio Fantine<sup>2</sup>, Federica Fanzago<sup>1</sup>, Ervin Konomi<sup>1</sup>, Paolo Emilio Mazzon<sup>3</sup>, Matteo Menguzzato<sup>4</sup>, Matteo Segatta<sup>1</sup>, Gianpietro Sella<sup>5</sup>, Massimo Sgaravatto<sup>1</sup>, Sergio Traldi<sup>1</sup>, Marco Verlatto<sup>1</sup>, and Lisa Zangrando<sup>1</sup>*

<sup>1</sup>INFN - Sezione di Padova, Via Marzolo 8, 35131 Padova, Italy

<sup>2</sup>INFN – Laboratori Nazionali di Legnaro, Viale dell'Università 2, 35020 Legnaro (Padova) Italy

<sup>3</sup>Università di Padova, Dipartimento di Ingegneria dell'Informazione (DEI), Via Gradenigo 6/b, 35131 Padova, Italy

<sup>4</sup>Università di Padova, Dipartimento di Fisica e Astronomia "Galileo Galilei", Via Marzolo 8, 35131 Padova, Italy

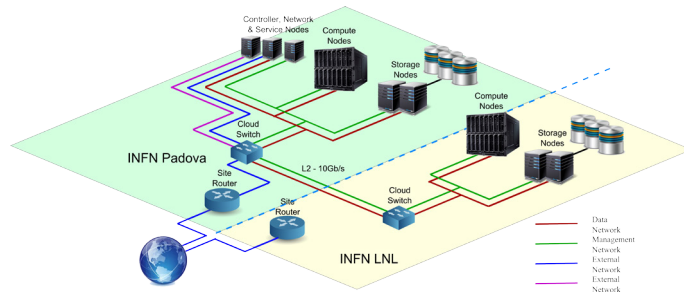
<sup>5</sup>Università di Padova, Dipartimento di Scienze Chimiche (DISC), Via Marzolo 1, 35131 Padova, Italy

**Abstract.** The Cloud Area Padovana, deployed in 2014, is a scientific IaaS cloud, spread between two different sites: the INFN Padova Unit and the INFN Legnaro National Labs. It provides about 1100 logical cores and 50 TB of storage. The entire computing facility, owned by INFN, satisfies the computational and storage demands of more than 100 users belonging to about 30 research projects, mainly related to HEP and nuclear physics. The Padova data centre also has hosted and operated since 2015 an independent IaaS cloud managing network, storage and computing resources owned by 10 departments of the University of Padova, supporting a broader range of scientific and engineering disciplines. This infrastructure provides about 480 logical cores and 90 TB of storage and supports more than 40 research projects. These two clouds share only a limited set of ICT services and tools (mainly for configuration, monitoring and accounting), whereas their daily operations and maintenance are carried out separately by INFN and University personnel. At the end of 2017 we planned to merge the two infrastructures in order to optimise the use of resources (both human and ICT) and to avoid useless duplication of services. We discuss here how we plan to implement this integration, resulting in a single cloud infrastructure named CloudVeneto.it.

## 1 Introduction

At the end of 2013, INFN Padova division and Legnaro National Laboratories (LNL) launched a project for the provision of a cloud service. This led to the implementation of an OpenStack-based cloud infrastructure which was called Cloud Area Padovana [1].

In 2015 ten departments of the University of Padova implemented another private cloud service [2], to provide their users with a pool of resources that could be easily and efficiently used. The implementation of this cloud service owned by the University of Padova was done in tight collaboration with the INFN team responsible for the operations of the Cloud Area



**Figure 1.** The architecture of the Cloud Area Padovana.

Padovana. These two clouds, however, were initially two independent IaaS infrastructures, sharing only a limited set of services and tools (mainly for configuration, monitoring and accounting).

To limit the manpower for the operations of these cloud services, to allow a better usage and sharing of the resources, and to avoid the useless duplication of the services, it was then decided to integrate the two clouds into a unique infrastructure called CloudVeneto.it [3].

The rest of this article is organized as follows. In Sect. 2 and Sect. 3 we describe the architectures respectively of the Cloud Area Padovana and of the Padova University cloud. In Sect. 4 we discuss how this integration, which while writing this paper is being finalized, was envisioned and implemented. In Sect. 5 we illustrate the current status and the usage of the CloudVeneto.it infrastructure. In Sect. 6 we discuss some recent developments. Sect. 7 concludes the article.

## 2 Design and architecture of the Cloud Area Padovana

The architecture of the INFN Cloud Area Padovana, described in detail in [1], is sketched in figure 1. This is an OpenStack [4] cloud infrastructure spread between two sites (INFN Padova Division and INFN LNL) which are about 20 km apart and connected by a 10 Gbps dedicated network link. Cloud and storage services are deployed in Padova, while the hypervisors, where virtual instances are created, are spread among the two sites.

Deployed OpenStack services are Keystone (for authentication and authorization), Nova (for instance management), Neutron (networking), Glance (images), Cinder (block storage), Heat (orchestration), Horizon (dashboard), EC2-api (to provide an EC2 [5] compatible interface) and Ceilometer (accounting).

Cinder backend is implemented with an iSCSI storage system exposed through the NFS Cinder driver (previously the Gluster [6] driver was used) and by a Ceph [7] cluster. The same storage systems are used also for the Glance service to store images. The ephemeral storage of virtual instances is instead implemented by the disk devices on the compute nodes.

As for networking, OpenStack Neutron relying on Open vSwitch and GRE [8] (Generic Router Encapsulation) is used. By using two provider routers (one with the external gateway on the public network and one with the external gateway towards the internal LAN) the instances can alternatively be given a floating (public) IP and be accessed from outside, or they can be reachable from the Padova and LNL LANs without the need to use floating IPs. All cloud instances, also the ones hosted on LNL hypervisors, connect to the internet through the INFN Padova site router.

The design of this private cloud was done with the goal of implementing a highly available infrastructure. The OpenStack services were deployed in two controller-network nodes in an active-active configuration. A three-instance Percona XtraDB cluster [9] is used to host the relational databases needed by the cloud and ancillary services. A HAProxy [10] - Keepalived [11] cluster, composed of three instances, is used to distribute the load among the available instances and to manage the fault tolerance of the services. HAProxy is also used to expose OpenStack services through an SSL compliant interface. Also the AMQP messaging service was deployed in high availability mode, relying on a three-instance RabbitMQ cluster.

### 3 The University of Padova cloud

The architecture of the University of Padova cloud is very similar to the INFN's Cloud Area Padovana. The main difference is related to the storage: a DELL EqualLogic system is used for most of the services: this storage system is used for the block storage service (through the proper Cinder EqualLogic driver), while for the images and for the ephemeral storage of the cloud virtual machines a space created on this storage box and exposed through an NFS server (configured in high availability) is used. Also in the University cloud the services were deployed in high availability mode but, because of lack of hardware, the HAProxy-Keepalived cluster and the RabbitMQ services were not deployed in dedicated machines but on hosts used also for other purposes.

While in the INFN cloud no partitioning of cloud resources was implemented (any virtual machine can be instantiated on any compute node), in the University cloud it was decided to reserve some high-memory hypervisors for instances which require more RAM. This was done using host aggregates and tagging some specific flavors so that instances created through them are targeted to the relevant compute nodes.

The hardware used for the Padova University cloud was physically deployed in the same computing room used to host the services of the Padova's instance of the Cloud Area Padovana.

### 4 The integration of the two clouds

The two cloud services described above shared some services: the monitoring infrastructure (which relies on Ganglia [12], Nagios [13] and Cacti [14]) and the Foreman [15] - Puppet [16] server used for provisioning and configuration. They were, however, two independent infrastructures.

To reduce the manpower needed to manage these cloud services, to favour the sharing of knowledge among the people and to optimize the resource usage, it was decided to merge the two cloud infrastructures into a single IaaS facility which was called CloudVeneto.it. Some constraints were however imposed by the management:

- INFN users had to use only computing and storage resources owned by INFN (and, vice versa, University's resources had to be made available only to users affiliated with the University);
- the facility had to be exposed to users through both a INFN and a University of Padova endpoint.

To implement this integration, it was decided to reconfigure the Cloud Area Padovana into CloudVeneto.it, i.e. in an infrastructure able to meet the requirements of both user communities, and then migrating University users and resources into these integrated services.

It was decided to consider the INFN cloud as reference implementation for this integration since it was running a newer (with respect to the University cloud) version of OpenStack and because the University cloud was smaller, in terms of compute nodes, users and instantiated virtual resources.

The first step in the implementation of CloudVeneto.it was the reconfiguration of the Cloud Area Padovana to expose the cloud services (through the dashboard and via command line tools) with both INFN and University of Padova's endpoints. This was done by properly configuring the HAProxy services and, for the dashboard, by creating specific virtual hosts. Taking as an example the Nova API service, the relevant section in the HAProxy configuration file is the following:

```
...
...

frontend novaapi_pub_infn
    bind <VIP-INFN>:8774 ssl no-sslsv3 crt /etc/cert/infn-cert.pem \
        /etc/cert/uniqueCA.crt
    mode http
    option httpclose
    option forwardfor
    reqadd X-Forwarded-Proto:\ https
    default_backend novaapi

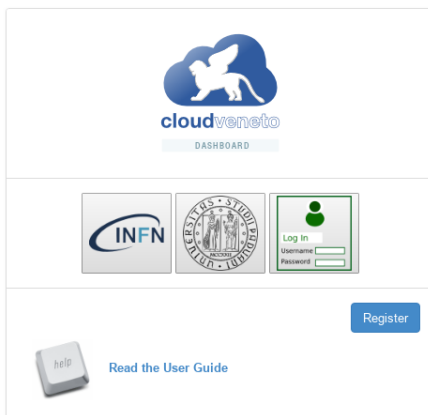
frontend novaapi_pub_unipd
    bind <VIP-UNIPD>:8774 ssl no-sslsv3 crt /etc/cert/unipd-cert.pem \
        /etc/cert/uniqueCA.crt
    mode http
    option httpclose
    option forwardfor
    reqadd X-Forwarded-Proto:\ https
    default_backend novaapi

backend novaapi
    mode http
    balance source
    option httpchk
    server <FQDN-1st-Controller> <IP-1st-Controller>:8774 \
        check inter 2000 rise 2 fall 3
    server <FQDN-2nd-Controller> <IP-2nd-Controller>:8774 \
        check inter 2000 rise 2 fall 3

...
...
```

So, the INFN endpoint and the University one both map to the same set of controller nodes where the service is actually run.

Regarding user authentication, the Cloud Area Padovana was already integrated (relying on the OS-Federation mechanisms) with the INFN Authentication and Authorization (AAI) infrastructure, to allow INFN users to authenticate to the cloud following the same procedure



**Figure 2.** The CloudVeneto.it dashboard.

used to access the other INFN services. To implement the merging of the two clouds the University of Padova’s Single Sign On system was also integrated in the cloud dashboard. This allows users to access the CloudVeneto.it dashboard authenticating using their own identity provider, as shown in figure 2. Authentication using username-password is also supported.

Networking also had to be reconfigured. Each CloudVeneto.it project is given a class-C network, which is then connected to one of the following routers:

- The first router has an external gateway on the INFN public network. This is used for INFN projects which need a public IP.
- The second router has an external gateway on the University of Padova’s public network. This router is used for projects affiliated with the University. The relevant instances can be given a University floating public IP, if needed.
- The third router has the external gateway towards the internal LAN. This project is used for the instances that don’t need a public floating IP, and need instead to be reachable from the Padova and LNL INFN LANs using their private IP.

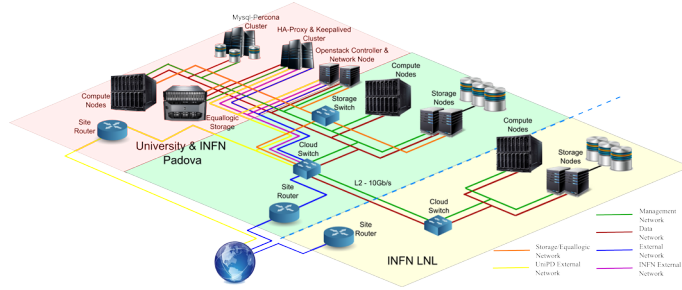
There is actually a fourth router, used for projects which are affiliated neither with INFN nor University (e.g. projects used for Public Administration related users).

Three floating IP sets were therefore created: one for INFN instances which need a public IP, one for University virtual machines, and one to cover the other scenarios.

Regarding the block storage, three Cinder backends were configured in CloudVeneto.it:

- a Ceph backend, available to INFN users;
- an iSCSI backend, implemented through the NFS driver, for INFN users;
- an EqualLogic backend, for University users.

The authorization on these backends was implemented simply relying on the Cinder quota mechanisms. So, for example, INFN projects are given a quota on the Ceph and/or iSCSI backends, but not on the EqualLogic one. This proved to work, even if unfortunately the



**Figure 3.** The architecture of CloudVeneto.it.

**Table 1.** CloudVeneto.it computing resources.

Owner	# compute nodes	# cores (in HT)	# GPUs	RAM (GB)	HepSpec06
INFN Padova	18	808	0	2656	8789
INFN Legnaro	13	416	0	1472	4390
University	13	512	2	2016	5438
Total	44	1736	2	6144	18617

Cinder quota for the different backends is exposed to users only through the command line tools (i.e. not in the dashboard).

Once the Cloud Area Padovana was reconfigured as CloudVeneto.it, the migration of users and resources from the University cloud started. Projects and users were imported on CloudVeneto.it through ad-hoc scripts which also managed to create the relevant project networks, the setting of the quota, the user’s ssh-keys migration, etc. The migration of virtual resources (instances, images, volumes) of the University projects was also automated through a custom script. Regarding images, they were simply downloaded from the University cloud and uploaded into CloudVeneto.it. The migration of instances was done by three steps: creating snapshots, migrating them to the CloudVeneto.it cloud, and finally starting new instances using these snapshots. The migration of the volumes was done creating new volumes on the CloudVeneto.it system, and then copying the data (using dd) from the source cloud. Compute nodes from the University of Padova cloud were then reinstalled and reconfigured in CloudVeneto.it, relying on Puppet modules which in a few minutes implement the needed changes.

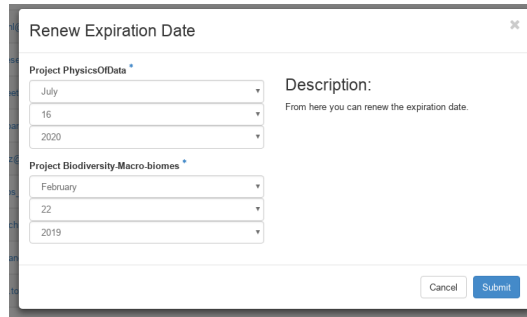
To allow only University users to instantiate virtual machines on the compute nodes owned by the University (and vice versa), the AggregateMultiTenancyIsolation OpenStack scheduler filter is used. Each project maps to an HostAggregate which defines the set of compute nodes that can be used for the instances of that project.

In figure 3 the architecture of the integrated CloudVeneto.it infrastructure is shown.

## 5 Status of CloudVeneto.it

Table 1 summarizes the available computing resources in CloudVeneto.it while, as far as storage concerns, about 230 TB of disk space are available to users.

At the time of writing this paper, about 180 users, belonging to almost 70 projects, are registered in the cloud. They are using the cloud infrastructure for several activities belonging to different scientific domains: multiparticle simulation of high intensity injectors, simulation



**Figure 4.** Renewal of user affiliations.

of complex networks, data assimilation, simulation of accelerator systems, finite element simulations, etc. Reference [17] describes some use cases for the infrastructure.

Since the execution of batch jobs is a use case common to many projects, an elastic on-demand batch cluster service is provided to CloudVeneto.it users. Users can instantiate on the cloud a HTCondor [18] batch system which, using the elastiq [19] system, is automatically scaled up or down as needed. This allows an optimal usage of cloud resources, as they are automatically released when they are not needed anymore.

## 6 New developments

The module managing users and project registration, described in [20], was recently refactored in particular to simplify the registration process and to support user renewal. Users belonging to different projects have now different expiration dates. The affiliation of a user to a certain project can then be renewed using the OpenStack dashboard, by the manager of that project, or by the cloud administrator (as shown in the example of figure 4). The module was also enhanced so that every action concerning user management (new accounts, renewals, etc) is properly logged, recorded in a database, and visible to the cloud administrator in the OpenStack dashboard.

There are also new developments in accounting. The Gnocchi [21] service was deployed in the cloud, to replace the MongoDB storage. The home-made development CAOS [22] service was modified to interact with Gnocchi and to provide additional accounting and monitoring information.

## 7 Conclusions

We discussed in this paper the CloudVeneto.it infrastructure, which is the result of INFN's Cloud Area Padovana and of Padova University cloud merging. It now provides about 1800 cores and 230 TB of storage to a wide user community of different research and scientific disciplines.

We foresee a considerable growth of user communities, as well as of the compute and storage infrastructure. The deployment of new services is also planned. Regarding the storage, the provisioning of an object storage service is being implemented. This will be done using the Ceph storage cluster, which will expose both S3 and Swift interfaces. Provisioning of some other high level services on top of the CloudVeneto.it IaaS is also foreseen. In particular the goal is to enable users to instantiate general purpose or specialized (e.g. for Big

Data workloads) container clusters (managed by Kubernetes orchestrator) with auto-scaling capability.

## References

- [1] P. Andreetto et al, **The Cloud Area Padovana: from pilot to production**, J. Phys.: Conf. Ser. 898 052007
- [2] P. E. Mazzon, M. Menguzzato, G. Sella, M. Sgaravatto, **Progetto CloudVeneto.it - Status Report**, Technical Report, May 2016
- [3] Website for the CloudVeneto.it infrastructure, <http://cloudveneto.it>
- [4] Home page for the OpenStack project, <http://www.openstack.org>
- [5] Amazon EC2, <https://aws.amazon.com/ec2/>
- [6] The Gluster web site, <http://www.gluster.org/>
- [7] Home page for the Ceph project, <http://ceph.com/>
- [8] Hanks S et al., **Generic routing encapsulation (GRE)** (2000). <http://tools.ietf.org/html/rfc2784.html>
- [9] Percona XtraDB Cluster, <https://www.percona.com/software/mysql-database/percona-xtradb-cluster>
- [10] Tarreau W, **HAProxy-The Reliable, High-Performance TCP/HTTP Load Balancer**, <http://haproxy.1wt.eu>
- [11] Cassen A, **Keepalived: Health checking for LVS and high availability**, (2002), <http://www.linuxvirtualserver.org>
- [12] Massie M, Chun B, Culler D, **The Ganglia Distributed Monitoring System: Design, Implementation, and Experience**, Journal of Parallel Computing, vol. 30, no. 7, July 2004.
- [13] Barth W, **Nagios. System and Network Monitoring**, No Starch Press, u.s (2006)
- [14] Cacti home page, <http://www.cacti.net/>
- [15] Foreman home page, <http://theforeman.org/>
- [16] Puppet Labs home page, <https://puppetlabs.com/>
- [17] <http://www.cloudveneto.it/testimonials.php>
- [18] Thain D, Tannenbaum T, and Livny M, **Distributed Computing in Practice: The Condor Experience**, Concurrency and Computation: Practice and Experience, Vol. 17, No. 2-4, pages 323-356, February-April, 2005.
- [19] The elastiQ software, <https://github.com/dberzano/elastiq>
- [20] C. Aiftimiei et al, **Implementation and use of a highly available and innovative IaaS solution: the Cloud Area Padovana**, J. Phys.: Conf. Ser., Volume 664, Issue 2, article id. 022016 (2015).
- [21] Gnocchi – Metric as a Service, <https://gnocchi.xyz/>
- [22] F. Chiarello, S. Traldi, P. Andreetto, **CAOS: a tool for OpenStack accounting management**, to appear in Proc. of CHEP 2018 conference