Building Service Platforms using OpenStack and CEPH: A University Cloud at Humboldt University

Malte Dreyer[1], Jens Döbler[1], Daniel Rohde[1]

[1]Computer and Media Service, Humboldt-Universität zu Berlin, Unter den Linden 6, 10099 Berlin, Germany, malte.dreyer@cms.hu-berlin.de, jd@cms.hu-berlin.de, d.rohde@cms.hu-berlin.de

Keywords
OpenStack, CEPH, Cloud Platforms, IaaS, PaaS, XaaS

1. ABSTRACT

For providing Infrastructure as a Service to the institutes, OpenStack was selected as a platform at the Computer and Media Service of Humboldt University. CEPH was chosen as the storage backend within this platform. The paper describes the project and first results. An overview to OpenStack is given and the results of initial CEPH performance tests are shown. The technical setup is depicted. Moving from traditional data center management to IaaS involves organizational changes as well as changes to service and business models. The first findings of this process are discussed and an outlook is provided.

2. Introduction

The Computer and Media Service at Humboldt University (HU) Berlin is the central IT service provider and data center for the HU. Within the institutes and other organizational units, de-centralized IT staff is responsible for the local IT; hence some services are provided in a wholesale and retail market model.

With around 120 services, the central resource capacity is utilized to a high degree and new services are requested at even increasing rates, while the IT budget remains static. For example, new tasks in the area of research data arise for data centers, requiring a high level of technical expertise and special domain knowledge. Currently, research data tools and related managing infrastructure are often developed in the context of fixed-term projects. Within these constraints, ensuring sustainability of software and information is difficult to manage. With around 430 professors from diverse scientific disciplines at HU, there is no apparent scalable solution to support these new tasks centrally. Besides other reasons, like organizational development and renewing the current technical infrastructure, it was decided to alleviate the burdens of operating the infrastructure for the institutes by building an Infrastructure as a Service (IaaS) environment. It is planned to extend this scenario towards Platform as a Service (PaaS) models successively to increase the amount of self service offerings.

After an evaluation of open source systems, OpenStack was chosen in January 2014, followed by workshops and staff training to build the first bigger testing environment. For cloud storage interfaces and scalability aspects, CEPH storage was decided to be integrated in the IaaS scenario. To support bootstrapping and knowledge building, a company with experience in OpenStack was closely involved. For dogfooding reasons, some new data center services currently in development will be based on the new IaaS architecture. It is planned to apply the same service patterns to more and more services in the next years, using "Anything as a Service" (XaaS) as a basic paradigm, also for organizational development.

To discuss the proper service models for the operation and extension of this cloud as well as establishing the user community, workshops have been held with HU institutes and two different funding models were identified.

3. The OpenStack Platform

OpenStack defines itself as "... a cloud operating system that controls large pools of compute, storage, and networking resources throughout a datacenter ..." (OpenStack: The Open Source Cloud Operating System) and many functionalities are similar to Amazon's AWS platform, even down to the level of APIs.

It consists of several modules or projects, addressing different aspects of the infrastructure, like virtual machine management, identity management, block storage, object storage, network abstraction, automation, bare metal deployment, user front end or specialized services for the management of databases, hadoop or redis as well as utilization metering or network function virtualization. (OpenStack Project Teams)

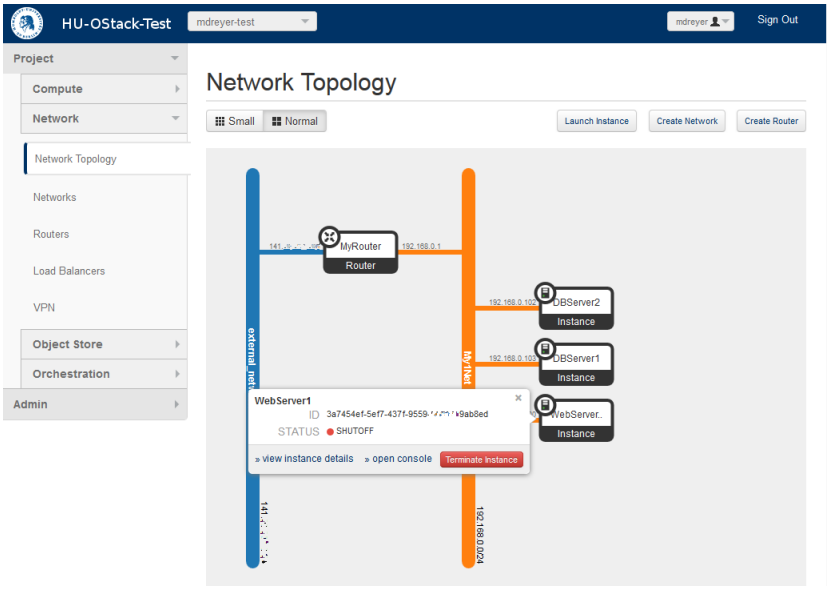

Figure 1. OpenStack dashboard user interface showing a tenant's network and resources

New projects are reviewed by the technical committee and evaluated for maturity which is expressed by the three statuses "Incubation", "Graduation to Integrated" or "First Integrated Release". (Openstack/governance)

Currently it has around two releases each year with the version name starting letters in alphabetic order and the name derived from a city or county near to the corresponding OpenStack design summit.

Since its inception in 2010 by NASA and Rackspace, OpenStack has developed at impressive rates. The latest Juno release from November 2014 was built by around 3,000 developers affiliated to more than 200 organizations. Compared to the Havana release end of 2013 it tripled the amount of developers involved within one year (Bitergia's blog). The project is hosted by the OpenStack Foundation established end of 2012.

4. Hardware and Configuration

For the servers, standard hardware was used and we started with a minimal setup, yet providing sufficient initial performance:

The compute nodes are 4 x 16 Cores, each 128GB RAM, the controller nodes are 2 x 12 Cores, each 64GB RAM, for the CEPH server nodes 10 x 12 Cores, each 64G RAM with 5 TB net (16 TB gross) hard disk, 200GB SSD for CEPH journal and 3 CEPH monitor servers are used.

In order to simplify the installation of operating systems and software and to avoid problems, for instance with network latencies, the servers are currently operated in just one location. In a later

development stage, the setup will be deployed in two major locations at Berlin Mitte and Adlershof with a distance of over 25 kilometers (over 15.5 miles) between them.

The servers are connected redundantly to two Brocade switches (ICX7750-48) over 10G copper connectors (10GBase-T), which provide the necessary bandwidth for network communication. The preferred operating system is Ubuntu Server 14.04 LTS. Each CEPH server has two redundant connections to a hard disk array (Infortrend EonStor DS ESDS-S16F-G2840/R with 16 Hitachi SATA hard disks) over fiber channel. The staging area is much smaller: 2 compute nodes, 2 controller nodes, and 3 CEPH nodes including CEPH monitors.

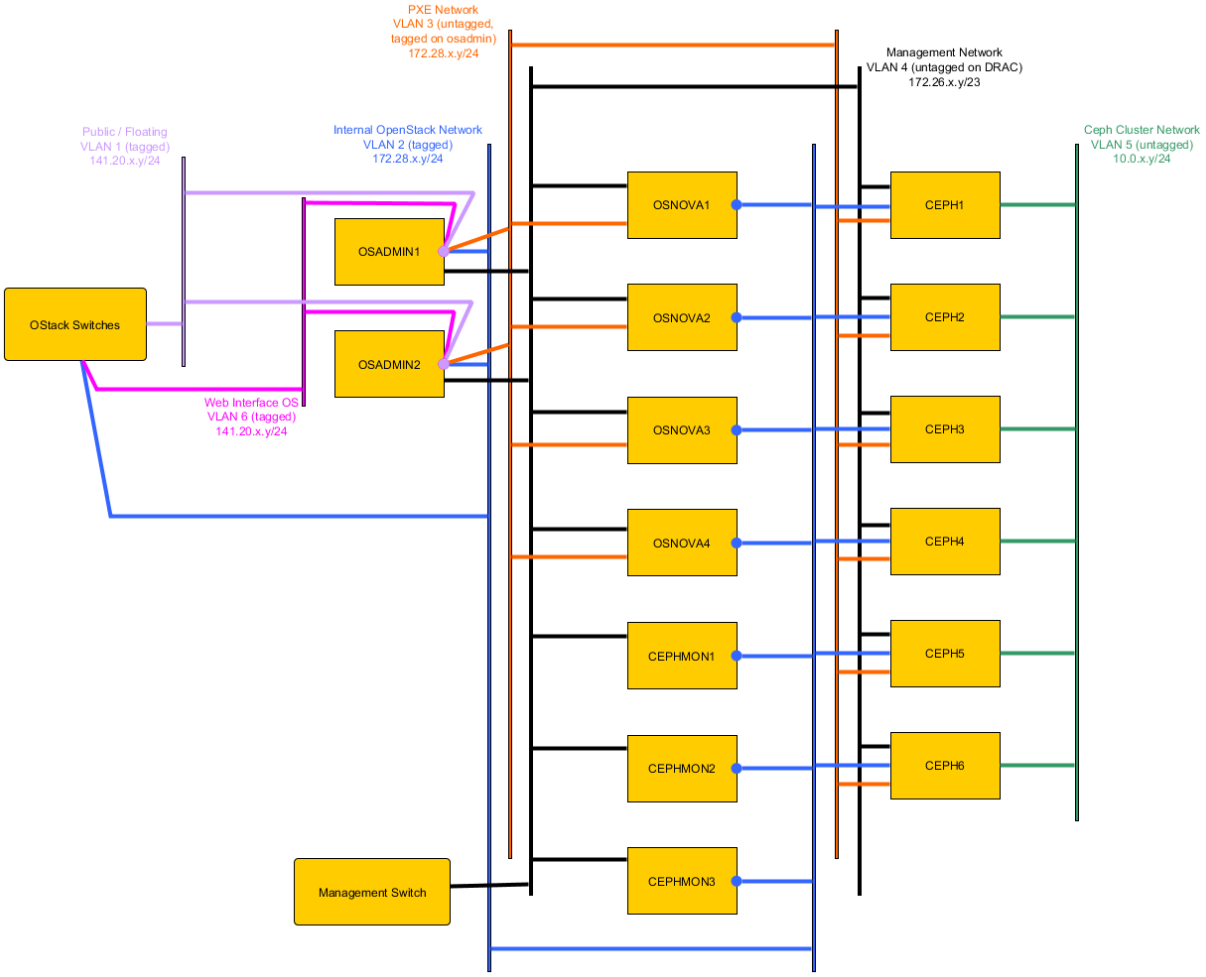

Figure 2. OpenStack physical testbed topology at HU

The OpenStack and CEPH roll-out was done with Ansible configuration management scripts initially provided by the involved company. The OpenStack platform was installed using all modules from the Icehouse release, but not all of them are enabled in the production environment. The compute nodes (Nova) using libvirt (KVM/QEMU) and the network management (Neutron) using VxLAN (SDN) and the Open vSwitch (OvS).

5. CEPH Storage

CEPH is a software defined storage solution that uses commodity hardware to scale horizontally. It is deployed as a volume storage backend to be used for e.g. object storage within OpenStack. Reliance and fault tolerance is achieved by storing multiple copies of objects on disks attached to different servers. Data placement is determined algorithmically instead of using a central map. This distributes

the load of locating objects over all clients, avoiding a central server (or cluster) which could be overloaded by numerous connections. The overall performance can be increased by adding additional object servers.

6. CEPH Performance

Performance tests were performed with VMs in OpenStack. Each VM had access to a 100 GB Cinder volume, which in turn corresponds to a RADOS block device in the CEPH cluster. Default settings are used; the volumes consist of 25600 4 MB objects and the data is striped over the physical drives in the CEPH cluster. The utility fio and direct access to the virtual disk device was used for benchmarks. The impact of multiple I/O streams was determined by runs with 1, 2, 4 and 8 identical VMs and identical I/O profiles.

Write performance was tested with sequential writes using 1 MB block size. In a single stream 295 MB/s are transferred, it profits from the performance of the SSD that store the write journals. The performance has almost reached saturation with 4 concurrent transfers a 514 MB/s (aggregated) and increases slightly to 537 MB/s with 8 transfers. Almost no difference is found for sequential writes with 1 MB and 32 KB block sizes. This can be attributed to the SSD journals and the data striping over all physical disks.

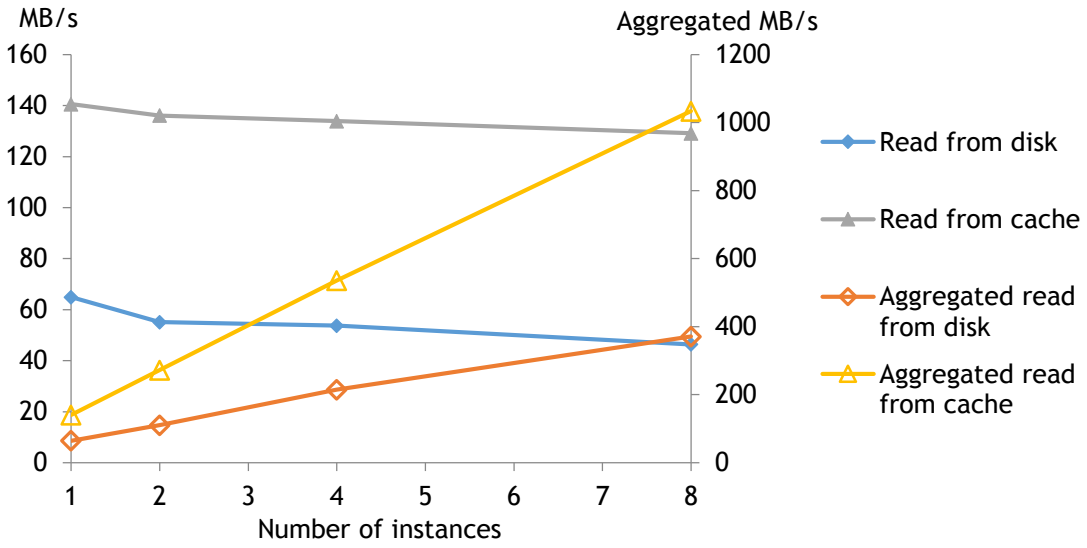

Figure 3. Read performance for sequential reads depending on number of instances.

Measurement of read performance turned out to be more difficult. Initially the results were not reproducible, due to caching effects. The objects are stored on the OSDs as files in an XFS file system and are cached in the RAM of the CEPH server after reading the object. Following reads on the object (which correspond to 4 MB chunks of the block device in the present case) access the RAM cache and no disk read occurs. The solution was to either drop the caches on all CEPH servers or reading the block devices to force the data into the cache before performing a benchmark run. In a single stream of sequential reads (32K blocks) 66 MB/s are transferred upon reading from disk and 144 MB/s with cached data. In both cases multiple streams lead to an almost linear increase. With 8 streams we find aggregated bandwidths of 380 MB/s (disk read) and 1058 MB/s (cached). The scale up suggests that saturation is not reached with 8 streams.

7. Conclusion and Outlook

Moving from traditional data center management to IaaS involves organizational changes, also effecting e.g. administrations techniques, networking and storage technologies as well as service and business models. Therefore the implementation of OpenStack can't be treated just as any other service to be newly introduced, which usually can be installed by applying existing knowledge and deploying services on top of the existing physical infrastructure one by one. As OpenStack means also a shift to software abstraction layers, many prior necessary tasks get obsolete - like for single service provisioning - and new ones arise - like the automating of deployment. For the HU Computer and Media Service, having separate groups for storage, networking, virtualization or databases, the implementation of IaaS can just be handled by incorporating the existing distributed knowledge of each group and developing the necessary knowledge and skills in each group simultaneously. Where managing techniques are changing, also new knowledge and new skills have to be built and incorporated into daily processes. Involving an external company for workshops, training and support helped to deal with the complexity of OpenStack and general acceptance.

After trying to deploy an OpenStack in OpenStack environment for staging purposes, another practical finding is, that a dedicated hardware environment is necessary as a staging area, because e.g. the network dependencies coming from the neutron component. It has to be close to the production setting in terms of hardware components.

After migrating to the Juno version of OpenStack, the cloud will be opened gradually to more and more users. It is expected, that the real usage examples will strongly influence the service attributes.

There are many new technologies to be understood and deployed in the context of OpenStack and IaaS. This is demanding much time and effort to master. Besides the technological challenges, OpenStack enables for very flexible service design. The huge developer community and the increasing number of projects for more services are promising, as long as the community model and community gardening are keeping pace with this growth.

8. REFERENCES

OpenStack: The Open Source Cloud Operating System. (n.d.). Retrieved February 13, 2015, from https://www.openstack.org/software/

OpenStack Project Teams. (n.d.). Retrieved February 13, 2015, from https://wiki.openstack.org/wiki/Project_Teams

Openstack/governance - OpenStack Technical Committee Decisions. (n.d.). Retrieved February 13, 2015, from http://git.openstack.org/cgit/openstack/governance/tree/reference/incubation-integration-requirements.rst

Bitergia's blog. (n.d.). Retrieved February 13, 2015, from http://blog.bitergia.com/category/openstack/

9. AUTHORS' BIOGRAPHIES

Malte Dreyer is the technical director of the Computer and Media Service of Humboldt University since 12/2012. He was the director of the department of research and development at Max Planck Society, Max Planck Digital Library for six years before. In this role he developed research and publication data infrastructure for the Max Planck Society's institutes, as well as many research tools. His interests now are in the field of scalable information management architectures and infrastructures in the intersection of organizational perspectives on ICT from data centers and information management organizations. He has a degree in computer engineering.

Jens Döbler has a PhD in Chemistry and worked for 6 years as a postdoc on theoretical chemistry. For the last 7 years he worked in the IT department of the HU-Berlin, focusing on storage and backup.

Daniel Rohde is the head of the department of System Software and Communications in the Computer and Media Service of Humboldt University. He has a degree in computer science and is working in the IT service center since 12 years with the major focuses on web server and web applications.