RED HAT FORUM
Europe, Middle East & Africa

redhat.

CSCS OPENSTACK FEDERATION WITH RED HAT SINGLE SIGN-ON, CEPH STORAGE AND EXTERNAL SWIFT

MASSIMO BENINI

SOFTWARE ENGINEER

SWISS NATIONAL SUPERCOMPUTING CENTRE

MARCO PASSERINI

SYSTEM ENGINEER

SWISS NATIONAL SUPERCOMPUTING CENTRE

CSCS

ETH zürich

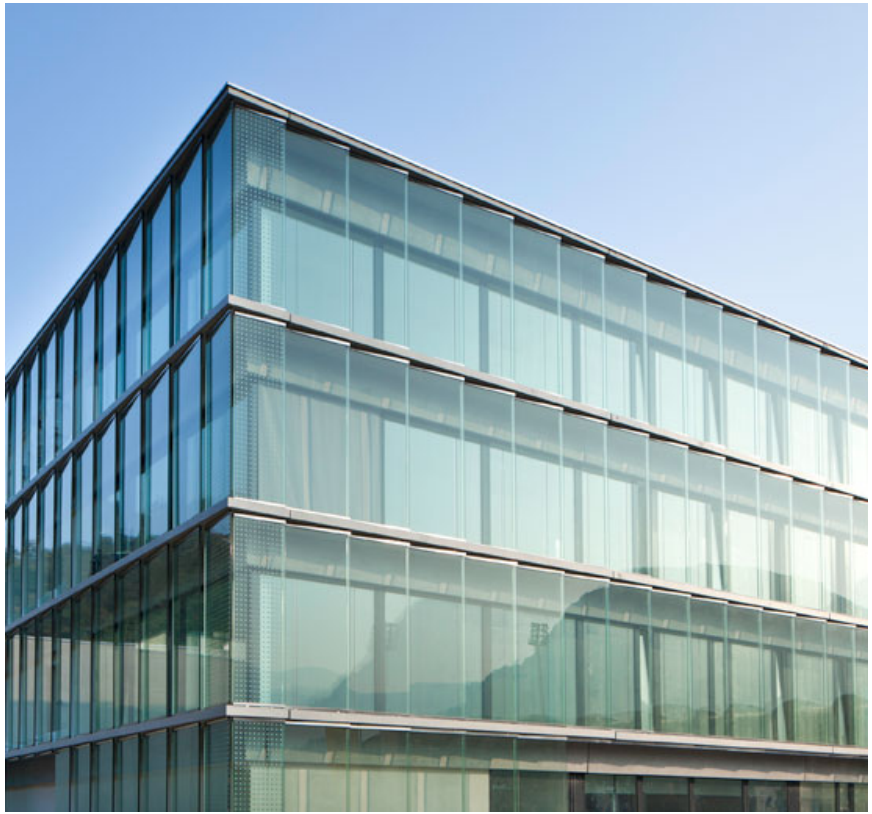
CSCS
Centro Svizzero di Calcolo Scientifico
Swiss National Supercomputing Centre

ETH zürich

CSCS (Swiss National Supercomputing Centre) is an HPC Centre whose mission is to develop and provide the key supercomputing capabilities required to solve important problems for science and/or society.

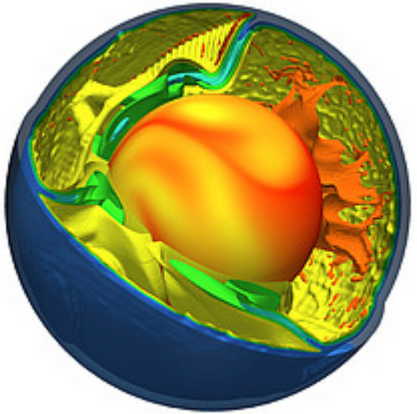
Agenda

- **CSCS Overview**
- Red Hat Engagement
- RH-SSO Federation
- Storage



CSCS in Brief (1)

- CSCS, the Swiss National Supercomputing Centre, develops and provides the key supercomputing capabilities required to solve important problems for science and/or society
- Unit of the Swiss Federal Institute of Technology in Zurich (ETH Zurich), located in Lugano
- CSCS's resources are open to academia, industry and the business sector
- Disciplines such as physics, materials science and cosmology traditionally use high-performance computers like those operated by CSCS



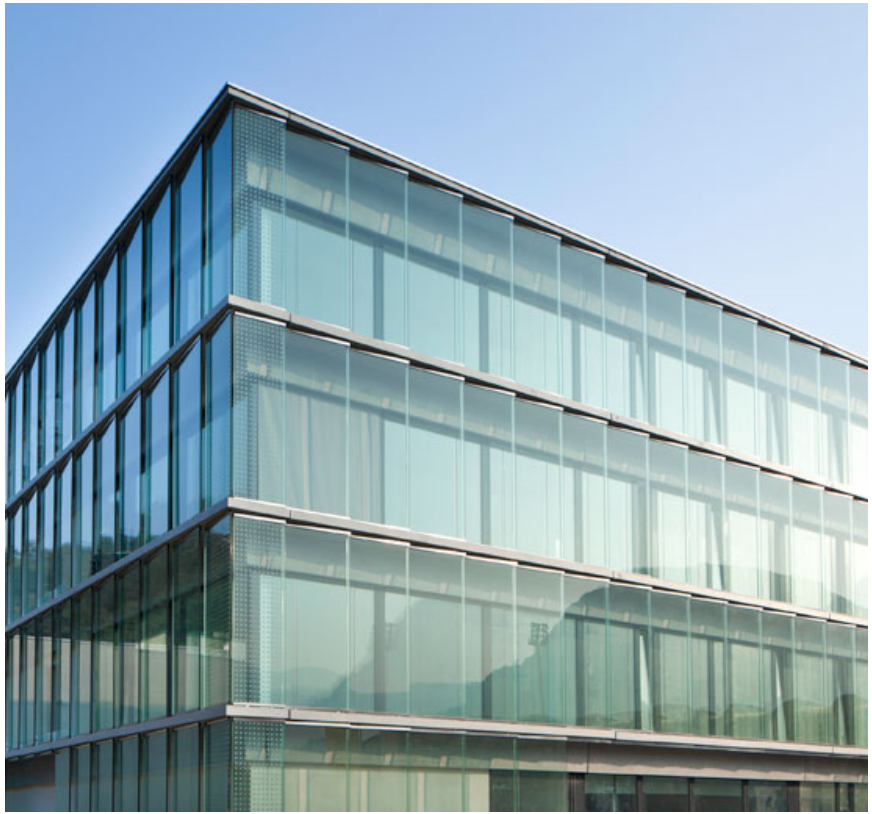
Computer models are extremely important to understand processes in the Earth's interior. They help comprehend plate-tectonic processes and the resulting earthquakes or volcanic activity better. Such simulations are thus essential for hazard and risk assessment. (Photo: Paul Tackley's research group / ETH Zurich)

CSCS in Brief (2)

- 2000 m^2 machine room with no single supporting pillar or any partitioning
- Operates the very latest supercomputers and works with the world's leading computing centers and hardware manufacturers
- Some operational HPC supercomputers:
  - Piz Daint (Cray XC40/XC50)
  - Kesch + Escha (Meteoswiss, Cray CS-Storm)
  - Mönch (NEC Cluster)
  - Phoenix (LHC CERN, Grid Cluster)
  - Monte Leone (High-memory cluster)
  - Gran Tavé (KNL R&D)

Agenda

- CSCS Overview
- **Red Hat Engagement**
- RH-SSO Federation
- Storage

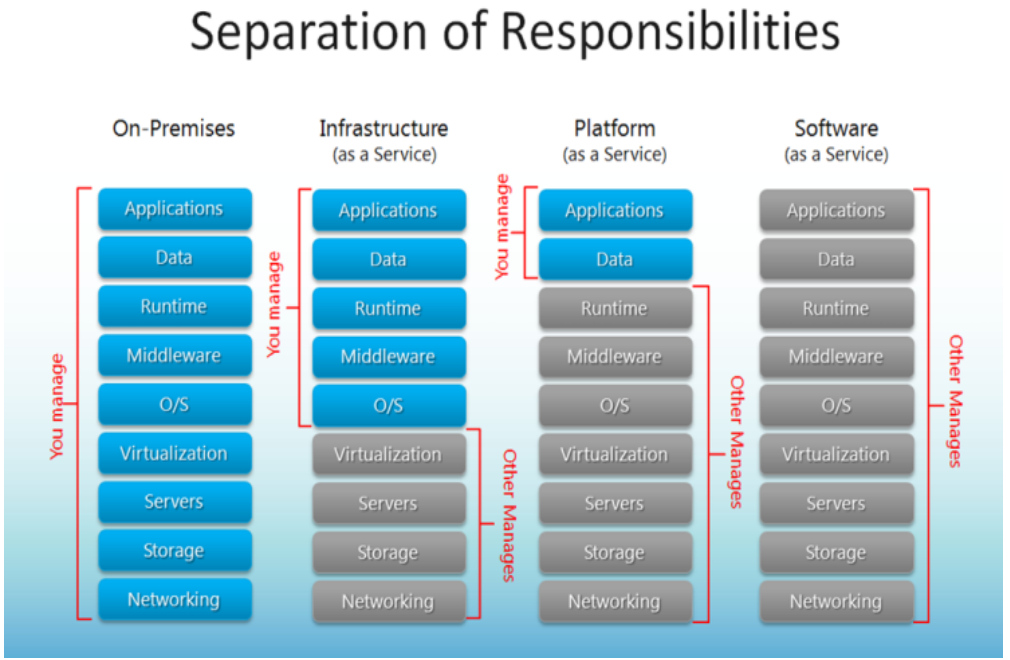


Motivation

- Need to **augment our infrastructure in a Service Oriented manner**, to accommodate new use cases coming from various user communities
  - User communities want to create web portals where they can show and share results
- OpenStack fits nicely with these requirements

Benefits of Infrastructure-as-a-Service

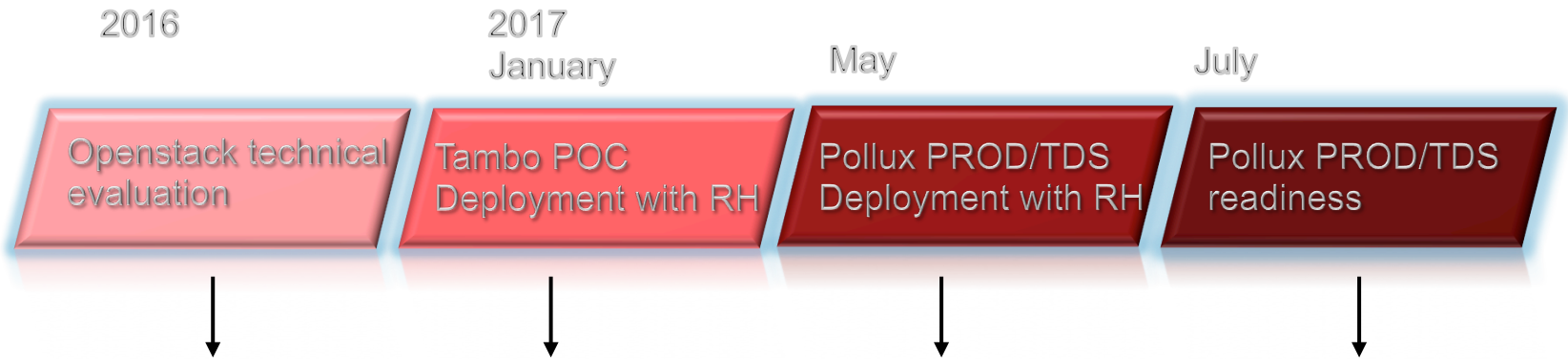
- Variable costs, pay-as-you-go model
- Immediate resource availability
- Dynamic scaling
- APIs and automation
- Interoperability
- RBAC
- Allows IT to shift focus
- Clear distinction of layers and responsibilities



OpenStack Deployment Constraints

- Reuse of existing **LDAP/Kerberos** infrastructure for authentication
  - Avoids creating an isolated OpenStack “island”
- Be prepared to **Federate services** with other external IdPs
  - But keep the CLI functionality working
- Big datasets will be stored in **Object Storage**
- We would like reuse our storage capacity on the **SAN**
  - Leverages economies of scale
  - Reuse the GPFS infrastructure for Swift Object storage

OpenStack Deployment Timeline



Devstack
RDO

Recycled hardware
RHOSP10
Hyperconverged for:
- AAI investigation (PAM,CLI,KRB,Keycloak)

Storage
- Manila
- GPFS

CVR

New hardware
RHOSP11
RHSSO
RH Training L1
DVR

First customers onboard
Finalization for production

Pollux Hardware

- **1x director**
  - Lenovo 3550 M5
  - CPU: 2x Intel E5-2603 v4 6C
  - RAM: 64 GB
  - NIC: 1x Intel X710 (Dual 10 Gb), 1x IPMI, 1x 1 Gb
  - HDD: 2x 120GB SSD
- **3x controllers**
  - Lenovo 3650 M5
  - CPU: 2x Intel E5-2620 v4 8C
  - RAM: 128 GB
  - NIC: 1x Intel X710 (Dual 40 Gb), 1x IPMI, 1x 1 Gb
  - HDD: 2x 120GB SSD
- **5x compute**
  - Lenovo 3650 M5
  - CPU: 2x Intel E5-2660 v4 14C
  - RAM: 512 GB
  - NIC: 1x Intel X710 (Dual 40 Gb), 1x IPMI, 1x 1 Gb
  - HDD: 2x 120GB SSD
- **5x compute nodes (big mem)**
  - HP DL360 G9
  - CPU: 2x Intel E5-2667 v3 8C
  - RAM: 768 GB
  - NIC: 1x HP 10Gb (Dual), 1x HP FDR 40Gb, 4x 1Gb
  - HDD: 2x 120GB SSD
- **3x Ceph storage nodes**
  - Lenovo 3650 M5
  - CPU: 2x Intel E5-2620 v4 8C
  - RAM: 128 GB
  - NIC: 1x Intel X710 (Dual 40 Gb), 1x IPMI, 1x 1 Gb
  - HDD:
    - 120GB SSD local drives RAID1
    - 18x SATA 2TB drives for data
    - 6x SSD 400GB drives for journaling
- **4x Swift nodes** (Spectrum Scale CES)
  - Supermicro SYS-5018R-WR
  - CPU: 1x Intel(R) Xeon(R) CPU E5-2637 v4 @ 3.50GHz, 4C
  - RAM: 128GB
  - NIC: 1x Intel XL710 (Dual 40 Gb), 1x IPMI, 1x 1 Gb
  - External SAN storage: Netapp E5600 (data), IBM FS900 Flash (metadata and Swift DBs)

RHOSP11 Services

We are currently operating the following OpenStack services:

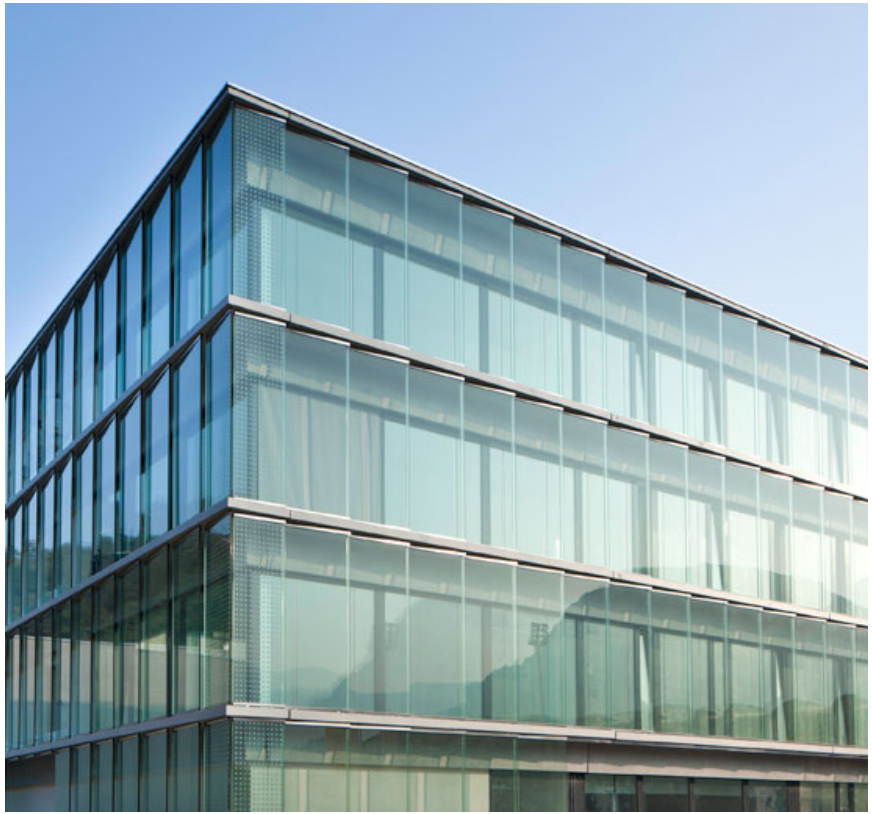
- aodh
- ceilometer
- **cinder**
- **glance**
- gnocchi
- heat
- **keystone**
- mistral
- **neutron**
- **nova**
- panko
- placement
- sahara
- **swift**

Integration with Other Services

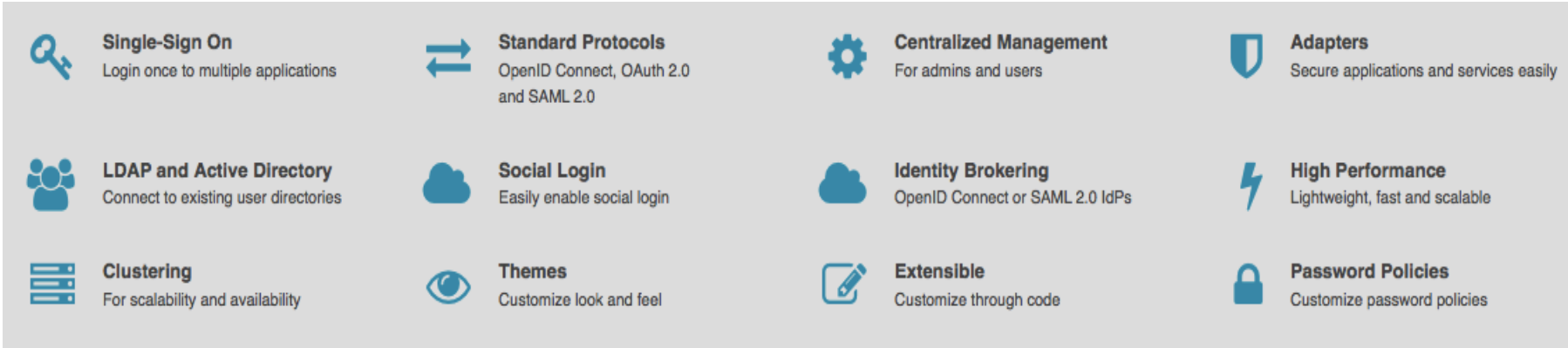
- Nagios
- Collectd
- Graylog
- IBM TSM
- LDAP/KRB
- External Swift (IBM Spectrum Scale CES Object)

Agenda

- CSCS Overview
- Red Hat Engagement
- **RH-SSO Federation**
- Storage

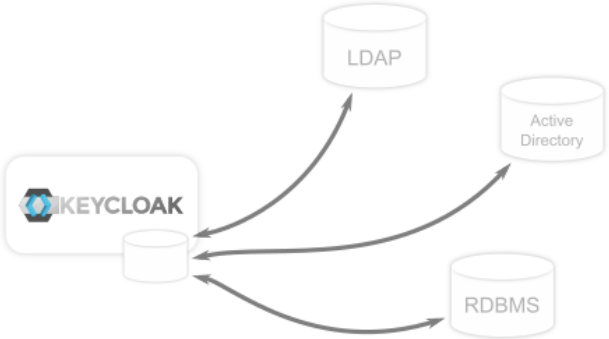


KeyCloak-RHSSO

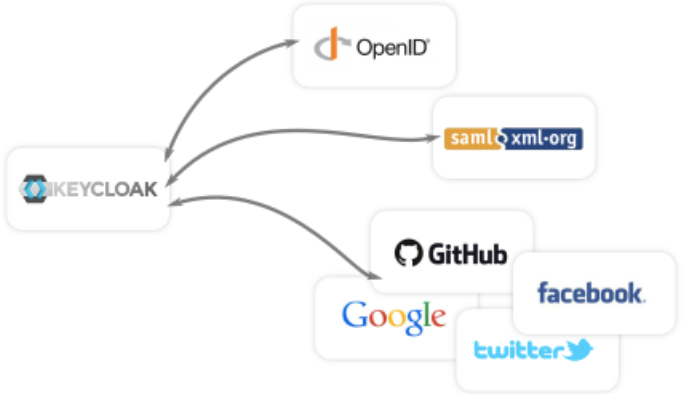


- Identity and Access Management solution aimed at modern applications and services
- Based on standard protocols

KeyCloak-RHSSO



User Federation, Kerberos bridge



Identity Brokering and Social Login

RHSSO

- Choice driven by our requirements:

  1. Need to maintain our users accounting unchanged (LDAP username and Kerberos password) → keystone natively don't allow this configuration.

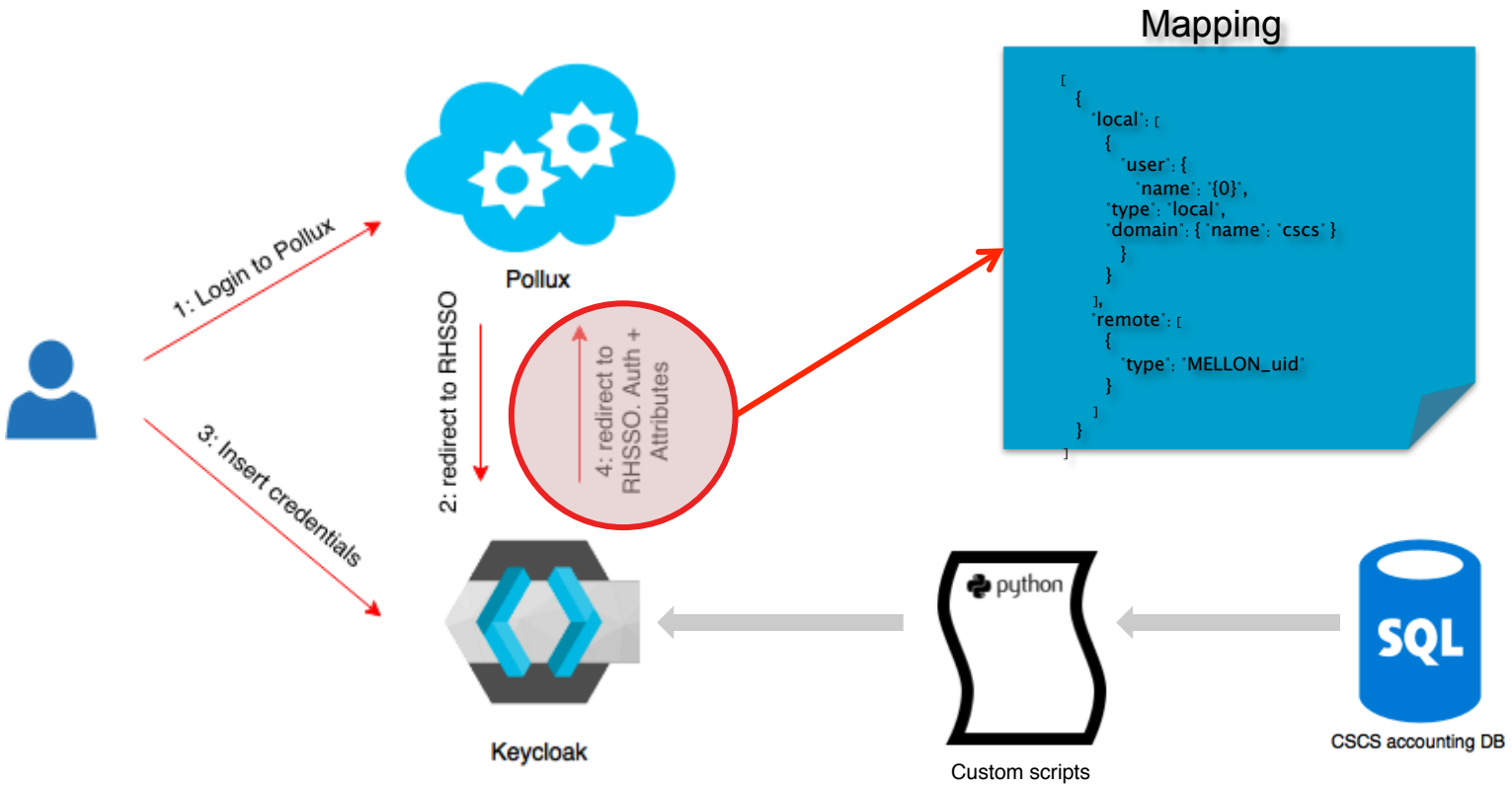
  2. Be prepared to Federate services with other external IdPs but keeping the CLI functionality working → RHSSO, acting as Identity Broker, is perfectly suitable for this. RH assure the CLI functionality in the OSP11 release.

- CLI set environment script: https://github.com/eth-cscs/openstack
  - GPLv3
  - easy automation with scripts
- `mod-auth-mellon` apache module for SAML

CLI code snippets (GPLv3)

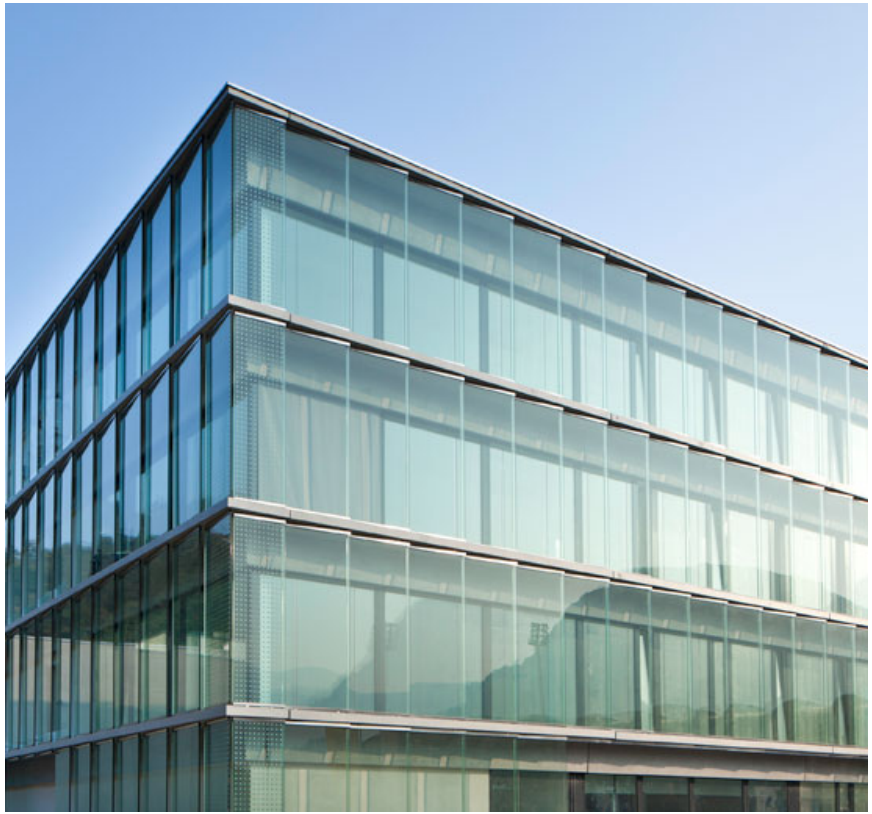
```
...

export OS_IDENTITY_API_VERSION=3
export OS_AUTH_URL=https://pollux.cscs.ch:13000/v3
export OS_IDENTITY_PROVIDER=cscskc
export OS_IDENTITY_PROVIDER_URL=https://kc.cscs.ch/auth/realms/cscs/protocol/saml/
export OS_PROTOCOL=mapped
export OS_INTERFACE=public

...

#Getting the unscoped token:
echo "[openstack --os-auth-type v3samlpassword token issue]"
UNSCOPED_TOKEN="$(openstack --os-auth-type v3samlpassword token issue --format value --column id)”

...

#Getting the scoped token:
echo "[openstack --os-project-id $PROJECT_ID token issue]"
SCOPED_TOKEN="$(openstack --os-project-id $PROJECT_ID token issue --format value --column id)"

...

echo " * Setting custom 'swift' alias"
alias swift='swift --os-auth-token $OS_TOKEN --os-storage-url https://object.cscs.ch:8443/v1/AUTH_$OS_PROJECT_ID'
echo " * Environment ready for openstack CLI with scoped project: $PROJECT_NAME"
```

Mapping


Mapping
[
{
"local": [
{
"user": {
"name": "{0}",
"type": "local",
"domain": { "name": "cscs" }
}
}
],
"remote": [
{
"type": "MELLON_uid"
}
]
}
]
Pollux
1: Login to Pollux
2: redirect to RHSSO
3: Insert credentials
4: redirect to RHSSO. Auth + Attributes
Keycloak
python
Custom scripts
SQL
CSCS accounting DB

Agenda

- CSCS Overview
- Red Hat Engagement
- RH-SSO Federation
- **Storage**



Storage Environment (1)

- Requirements
  - We need enough space for **block and image storage** (**30TB**)
  - For **object storage** our customers want:
    - To scale to millions of files
    - PB of data
    - High bandwidth
  - We have lots of space on **SAN** available for use
  - We want to use our **tape library** for data backups

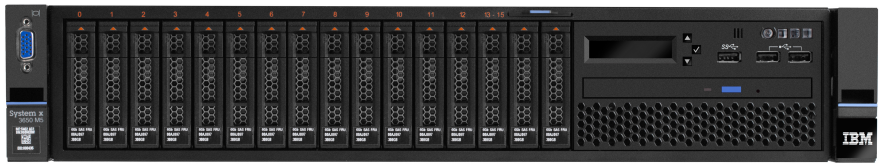
Storage Environment (2)

Implementation:

- **Ceph Jewel** 10.2.7-27.el7cp RHOSP11
  - Cinder block storage
  - Glance image storage
- **IBM Spectrum Scale CES Object** (GPFS)
  - Swift object storage
    - Used also for volume backups

Storage Environment – Ceph Hardware

- PROD:
  - **3x** servers Lenovo 3650 M5
    - CPU: 2x Intel E5-2620 v4 8C
    - RAM: 128 GB
    - NIC:
      - 1x Intel X710 (Dual 40 Gb) bonded for storage network and management
      - 1x 1Gb for provisioning
      - 1x 1Gb IPMI interface
    - HDD:
      - 2x 120GB SSD local drives RAID1
      - **18x SATA 2TB drives for data**
      - **6x SSD 400GB drives for journaling**
- TDS:
  - 3 servers with similar hardware configuration
  - HDD:
    - 3x SATA 2TB drives for data
    - 1x SSD 400GB drives for journaling

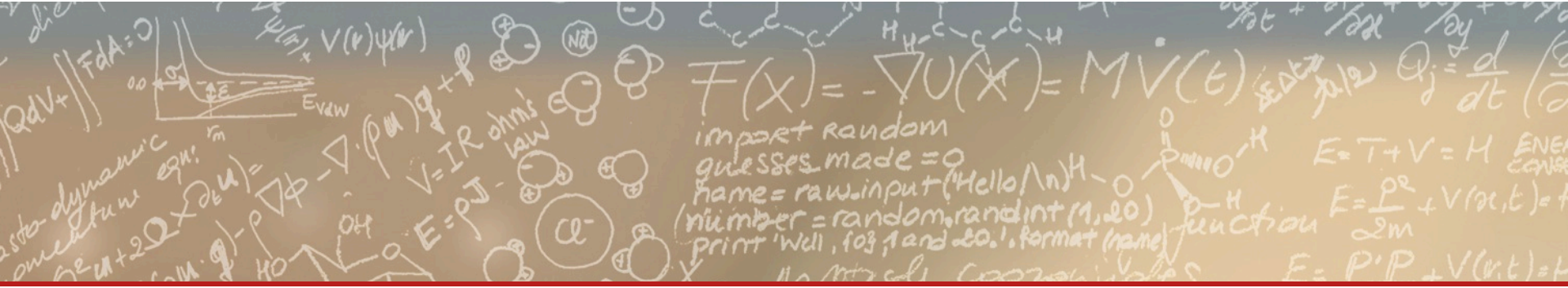


Storage Environment - Configuration

- **3 replicas**
  - PG groups calculated with http://ceph.com/pgcalc/
- Block storage **volume types**
  - Bronze 1.2 GB/s 1000 IOPS
  - Gold 1.2 GB/s 10000 IOPS
  - Platinum 1.2 GB/s 30000 IOPS
- **Backups**
  - Can be triggered by users
  - Backed up daily to Swift, then backed up to TSM
- **Benchmarks**
  - Aggregate bandwidth on Ceph (3 servers)**:**
    - **770 MB/s write**
    - **700 MB/s read**
  - Each storage server could potentially reach 2GB/s

Lessons Learned

- **A good design is key**
- Reliable and **supported hardware** is very important
- Changing the setup after deployment is very challenging
- Network design is complex
  - Needs the VM connectivity requirements and security policies in advance
- **Integration with legacy systems** is difficult (GPFS, monitoring, logging, accounting, AAI..)
  - Must have requirements in advance
- Implementation of **additional services** not trivial
- We have now a much clearer idea on how to install an OpenStack environment



Q&A

Massimo Benini: benini@cscs.ch
Marco Passerini: passerini@cscs.ch

RED HAT FORUM

Europe, Middle East & Africa