

Juju Charms per il deployment e la gestione di servizi (e come svilupparli)

Claudio Pisa - GARR - Dipartimento di Calcolo e Storage Distribuito

Roma, 8-10 ottobre 2019

Workshop GARR

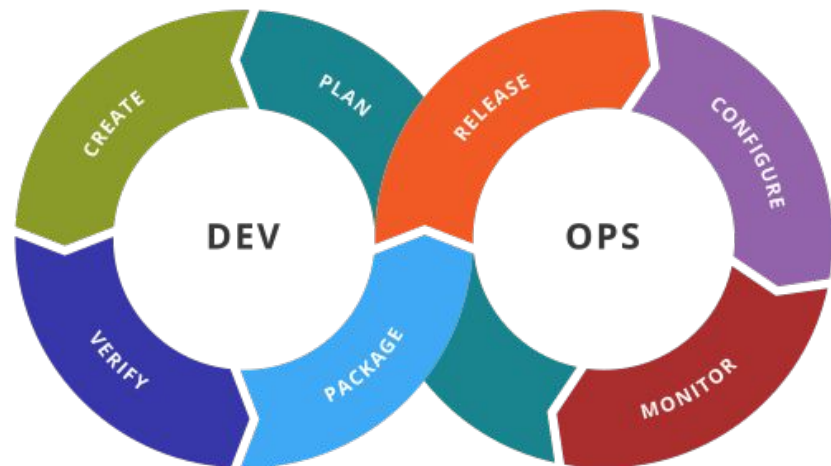
GARR Distributed Computing and Storage Department



- Role:
 - provider of resources (“long tail of science”)
 - resource aggregator (federation)
 - embody a replicable model for storage and computing provisioning
- Goals:
 - harmonize (SSO / federation of resources)
 - build secure and open infrastructures
 - enhance the user experience

GARR CSD - Guidelines

- Automation and Replicability
 - Infrastructure as Code
 - self-deployment
 - Open documentation
 - Continuous integration
 - Monitoring and self-healing
- Security and Privacy
 - GDPR
 - ISO 27001



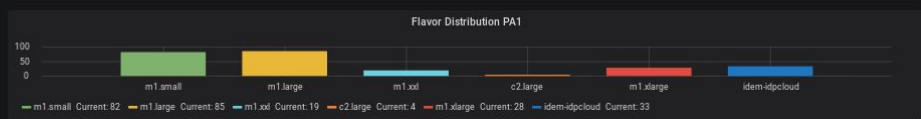
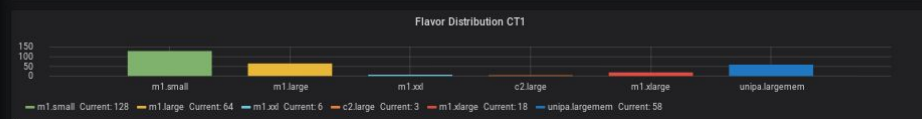
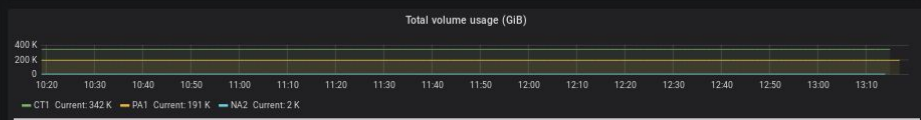
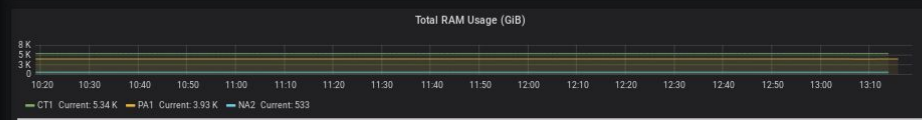
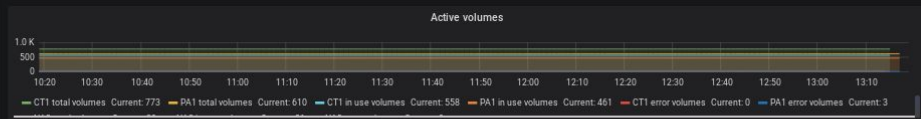
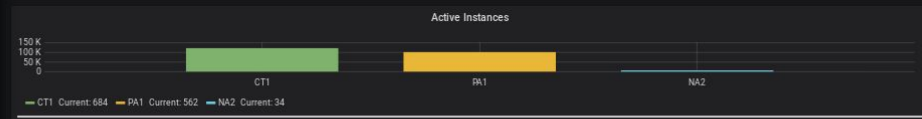
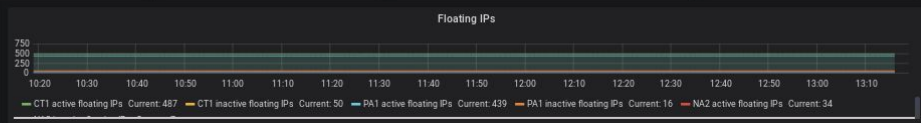
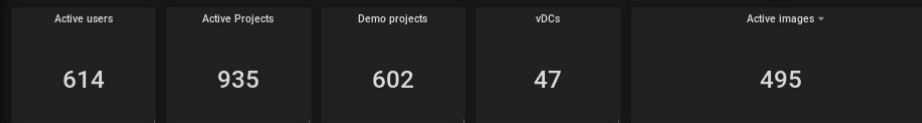
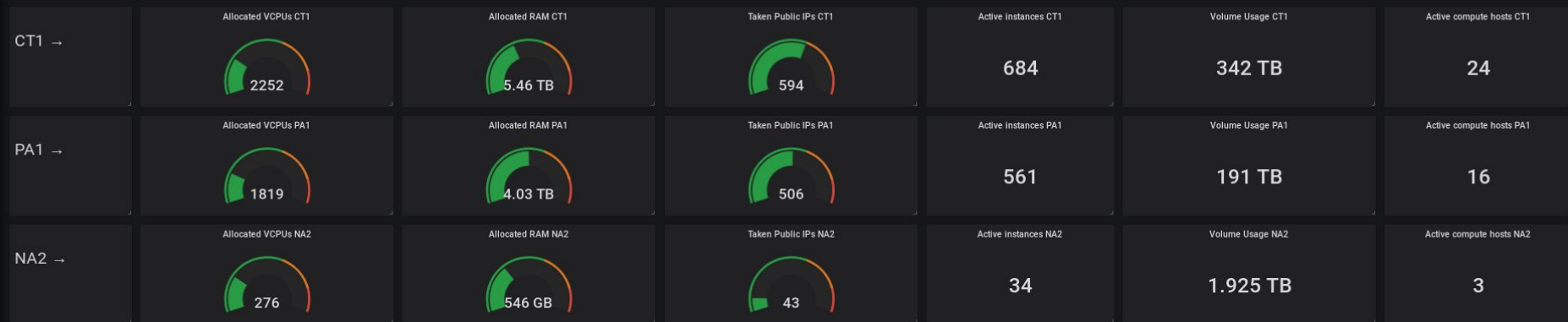
GARR Cloud Services

- Infrastructure as a Service:
 - GARR Cloud (OpenStack)
- Platform as a Service:
 - GARR Container Platform (Kubernetes)
 - Deployment as a Service (Juju)
- Software as a Service:
 - GARR Workplace (OnlyOffice)





- Based on **OpenStack**
 - cloud OS for data centers
- Features:
 - IDEM authentication
 - Ceph based storage
 - replicated and erasure-coded
- Users:
 - Single Users
 - Virtual Data Centers
 - administrative delegation

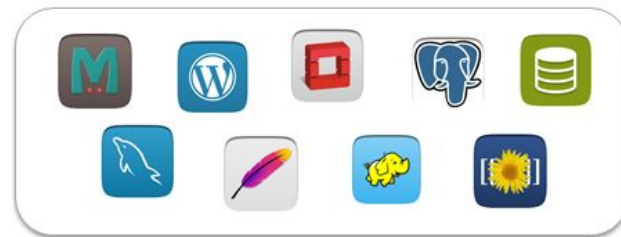




- Environment for automating deployment, scaling, and management of containerized applications
 - based on **Kubernetes**
- Cluster
 - Deployed on bare metal
 - Provides GPU resources
 - Ceph based storage
- Helm support
- Multitenant
 - Unified authentication with OpenStack
 - result of GN4-2
 - contributed upstream

DaaS (Deployment as a Service)

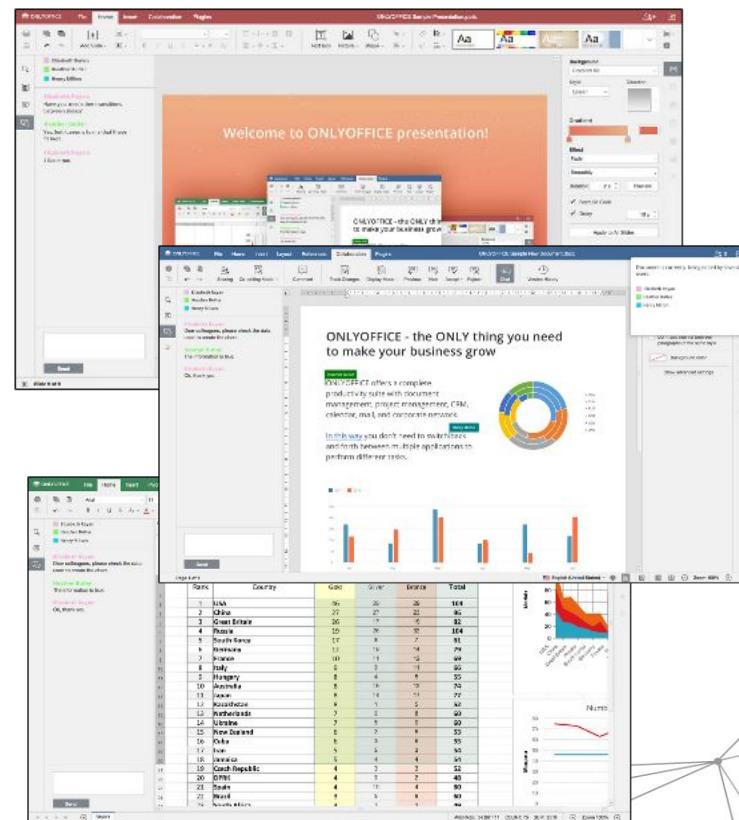
- A graphical user interface to install applications on the GARR Cloud Platform
 - based on Juju
 - application catalogue at <https://jaas.ai/>
- <https://cloud.garr.it/apps/>



<https://daas-pa.cloud.garr.it>
<https://daas-ct.cloud.garr.it>

GARR Workplace

- A web-based office and collaborative suite hosted on the GARR cloud
 - based on OnlyOffice
 - community and commercial licenses
- Features:
 - Document editor
 - highly compatible with popular document formats
 - Collaborative tools
 - e-mail, calendar, wiki, chat, etc.
 - CRM
 - Project management
 - Pluggable cloud storage
 - e.g. GARRBox



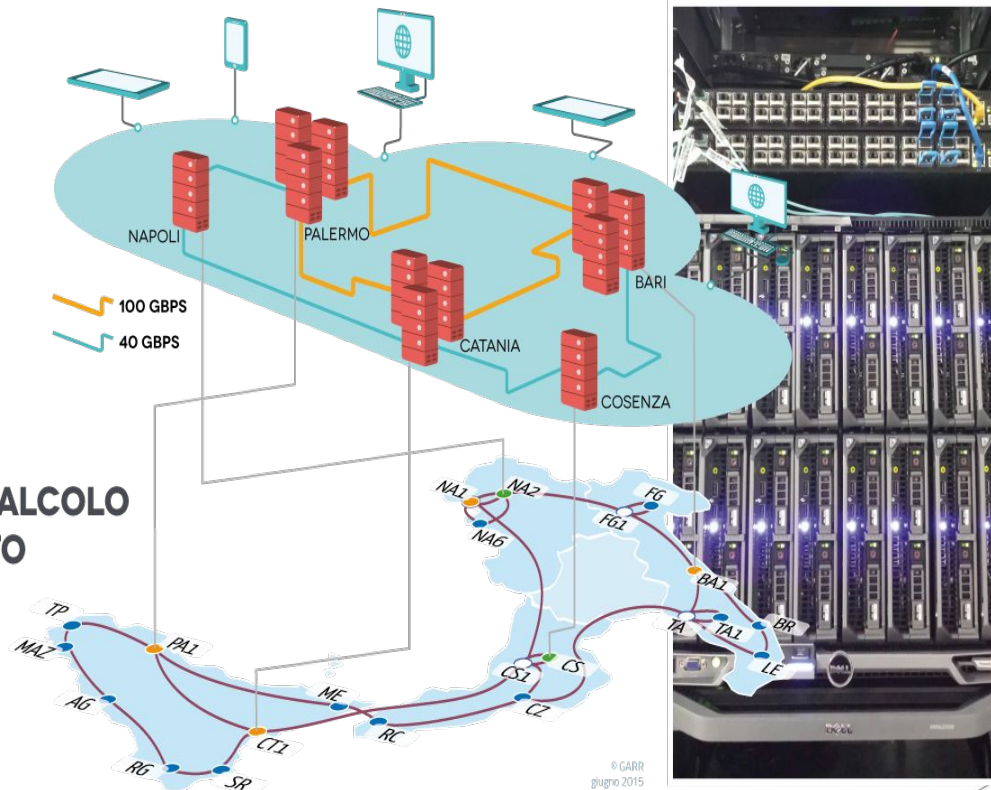
GARR Computing and Storage Infrastructure

INFRASTRUTTURA DI CALCOLO E STORAGE DISTRIBUITO

📍 5 siti distribuiti

📄 8.448 virtual CPU

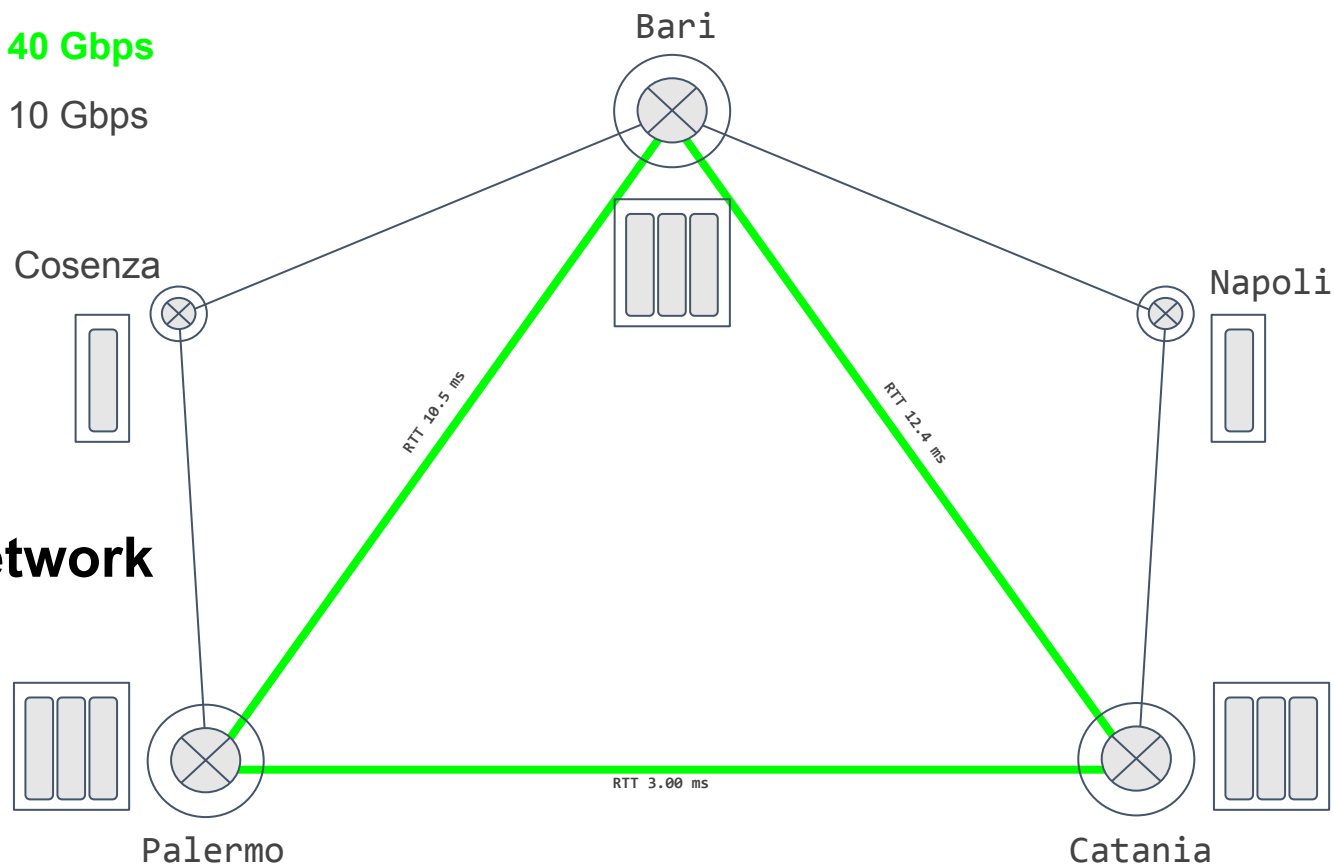
💻 10 PB spazio storage

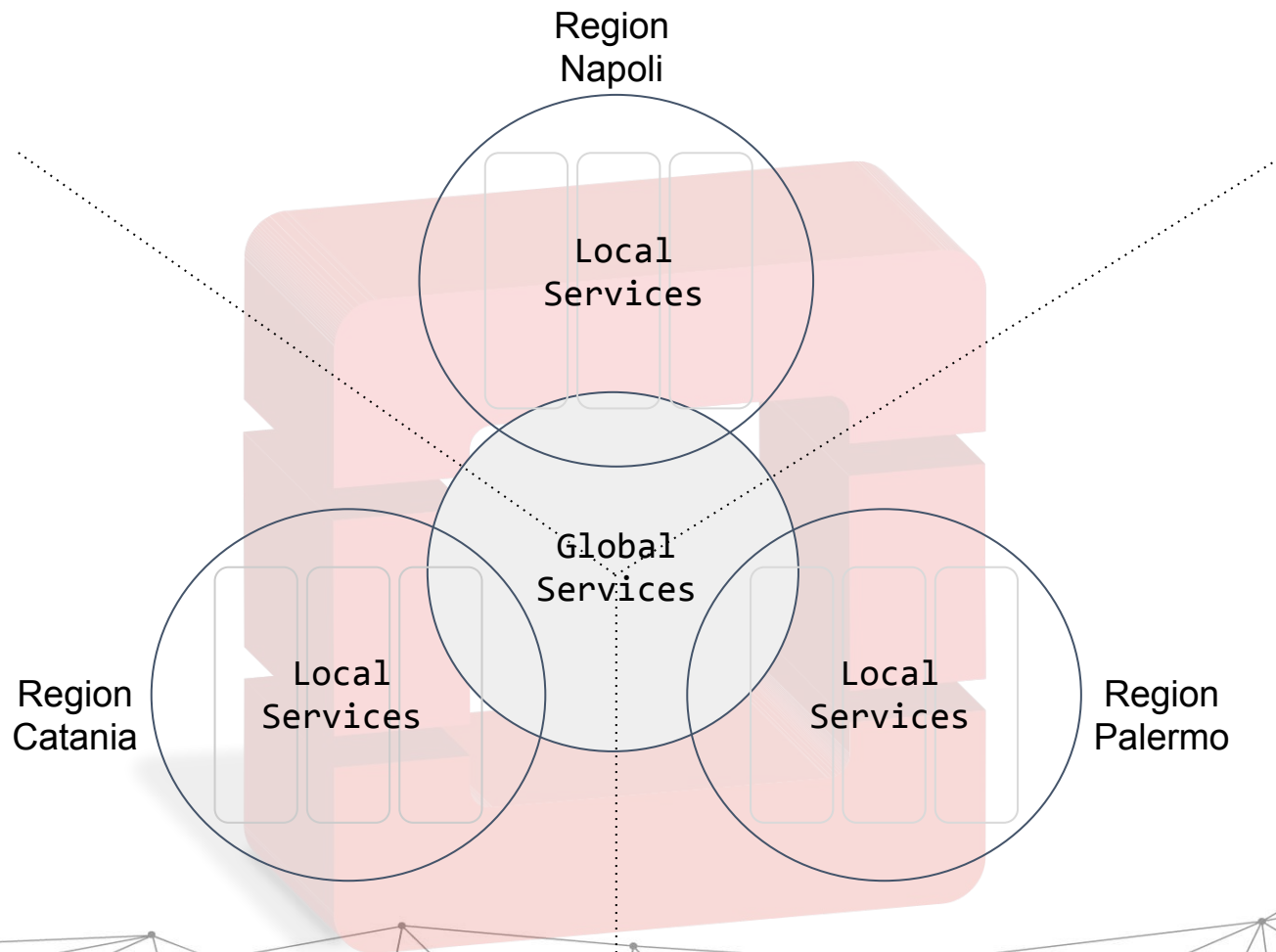


40 Gbps

10 Gbps

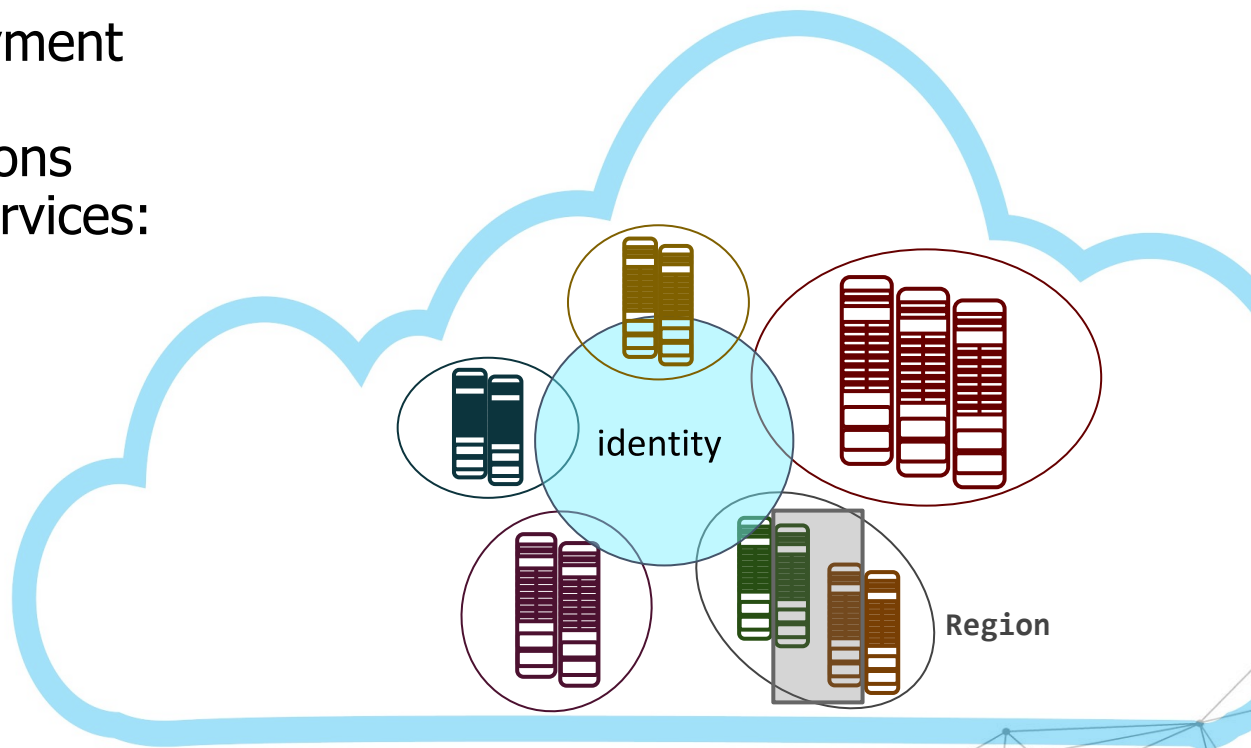
Network





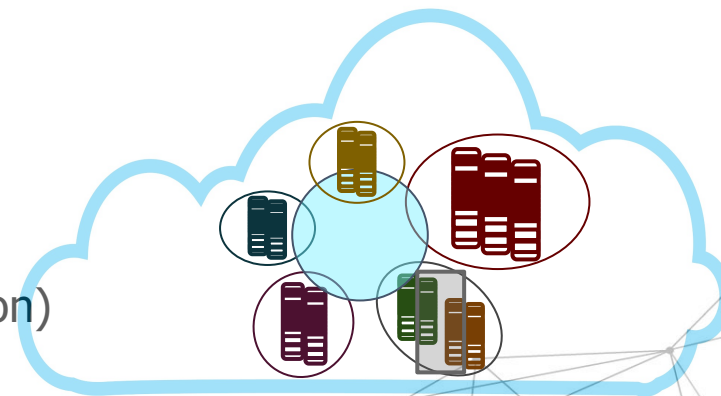
Multi-region/multi-domain Federation Model

- Region: separate deployment of OpenStack
 - Linked to other regions through common services:
 - identity service
 - dashboard
 - image service



GARR Cloud Federation

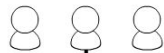
- Simple Federation recipe
 - Reference architecture with use cases
 - git repository and knowledge base
 - <https://git.garr.it/cloud/federation/>
 - <https://cloud.garr.it/doc/federation/>
 - Federated authentication
 - Administrative delegation
 - work in (advanced) progress
- Our recipe is working: join us or just copy us!
 - UniTo, PoliTo, UniPD (federating)
 - 3 INGV regions (external federation)
 - HPC4AI project
 - EAPConnect (East EU Nren, external federation)
 - Hungary
 - ...



Open Infrastructures

- Open Documentation
 - replicable model
 - Infrastructure as (open source) Code
- Open Federation
 - members of the GARR community can federate their computing resources
- Open Source / Free Software contributions:
 - Horizon
 - k8s-keystone-auth
 - Juju charms: ceph, keystone saml/sso, default gw, moodle...

Internet



- CLI clients(nova, cinder, neutron and so on)
- Cloud management tools
- GUI tools

HTTP(S)

Horizon
OpenStack Dashboard

heat-api heat-api-cfn

Queue

heat-engine

OpenStack Orchestration

Database LDAP

keystone-all

Optional

OpenStack Identity Service

ceilometer-collector

ceilometer-agent-notification

ceilometer database

ceilometer-agent-compute

Queue

ceilometer-agent-central

ceilometer-api

ceilometer-agent-evaluator

Log or HTTP callback

ceilometer-agent-notifier

OpenStack Telemetry

swift-proxy-server

trove-api

nova-api

nova-scheduler nova-console

cinder-api

glance-api glance store

neutron-server

sahara-all

swift-object-server

Queue

Trove Database

Nova database

Queue

Glance database

Neutron L2 agent *

Queue

swift-account-server

trove-taskmanager

nova-conductor

nova-cert

Cinder database

glance-registry

neutron-L3-agent *

neutron-dhcp-agent

Account database

trove-conductor

nova-consoleauth

nova-compute

cinder-volume

Volume provider

Neutron database

sahara database

Object database

Container database

Guest agent

Instance

cinder-scheduler

OpenStack Image service

Neutron 3rd party plugin

Optional, depends on plugin *

Openstack Object Storage

OpenStack Database Service

OpenStack Compute

OpenStack Block Storage

OpenStack Image service

OpenStack Networking

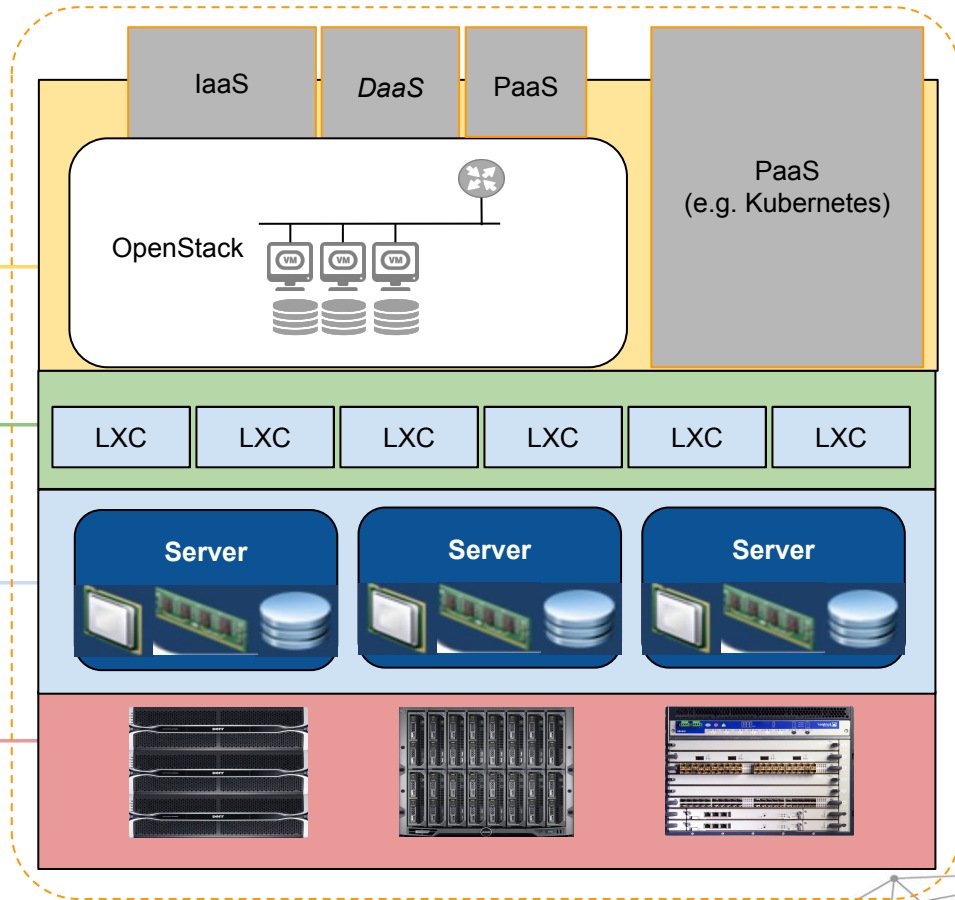
OpenStack Data Processing

1. Application Services

2. Infrastructure *Virtualization*

3. Operating System

4. Physical resources

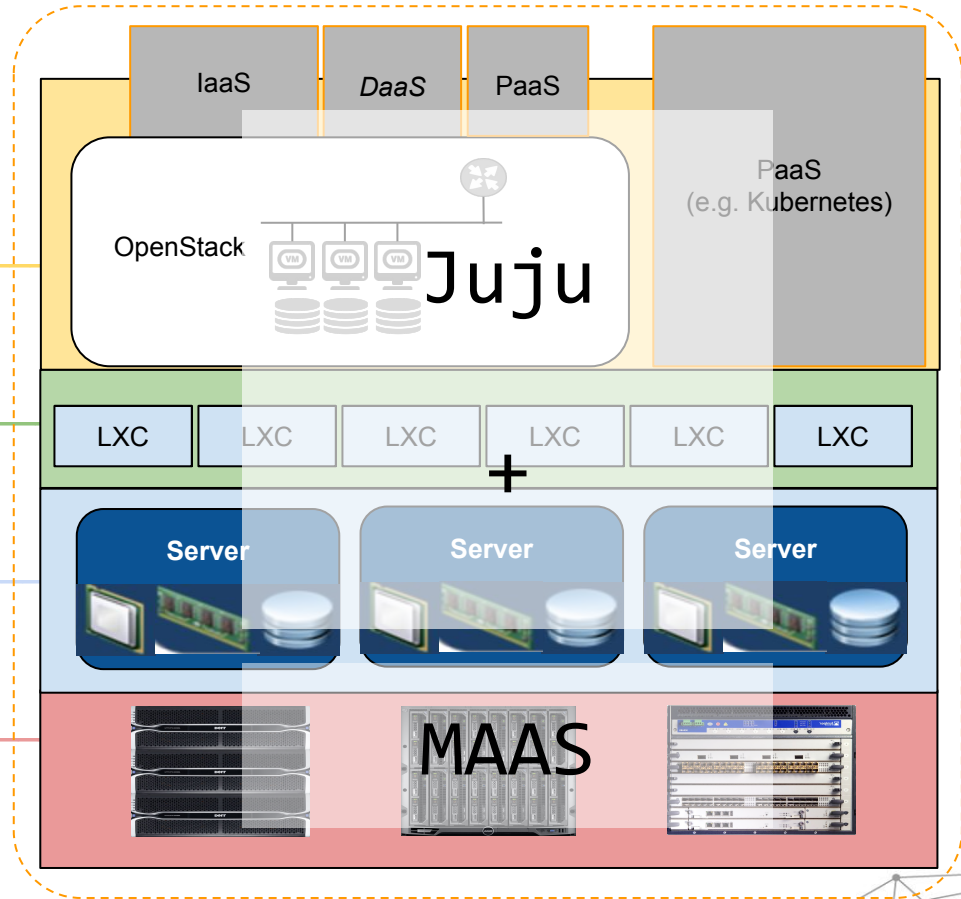


1. Application Services

2. Infrastructure *Virtualization*

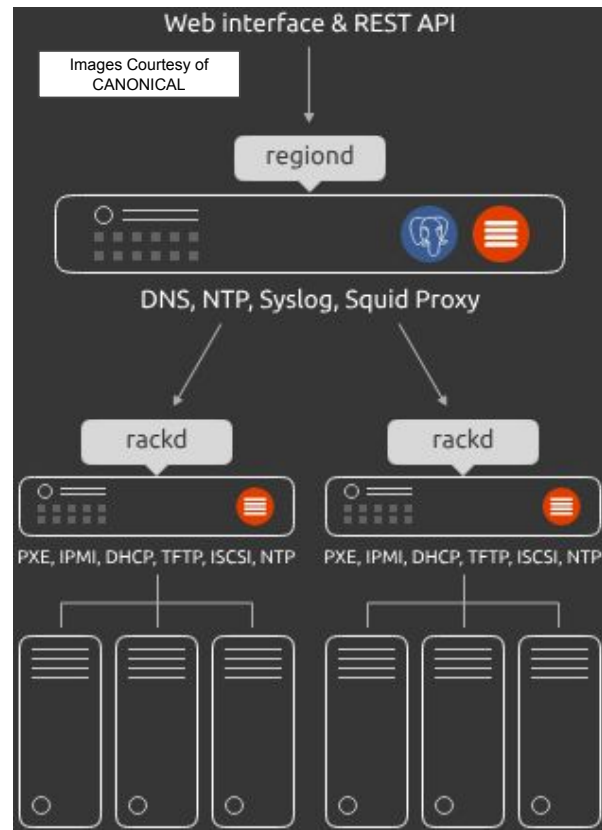
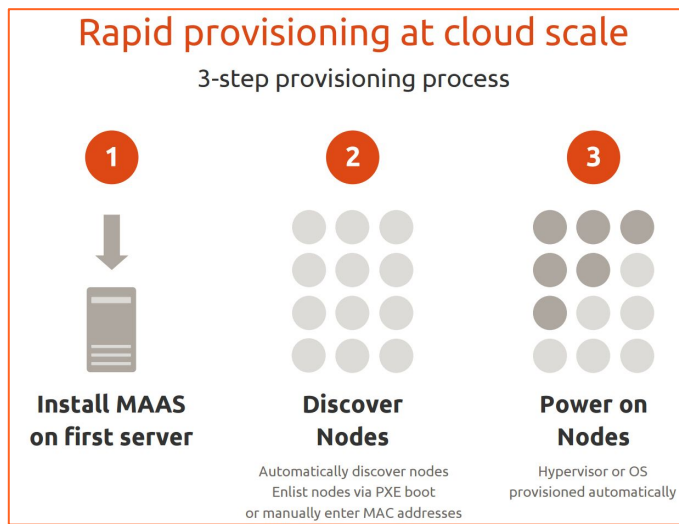
3. Operating System

4. Physical resources



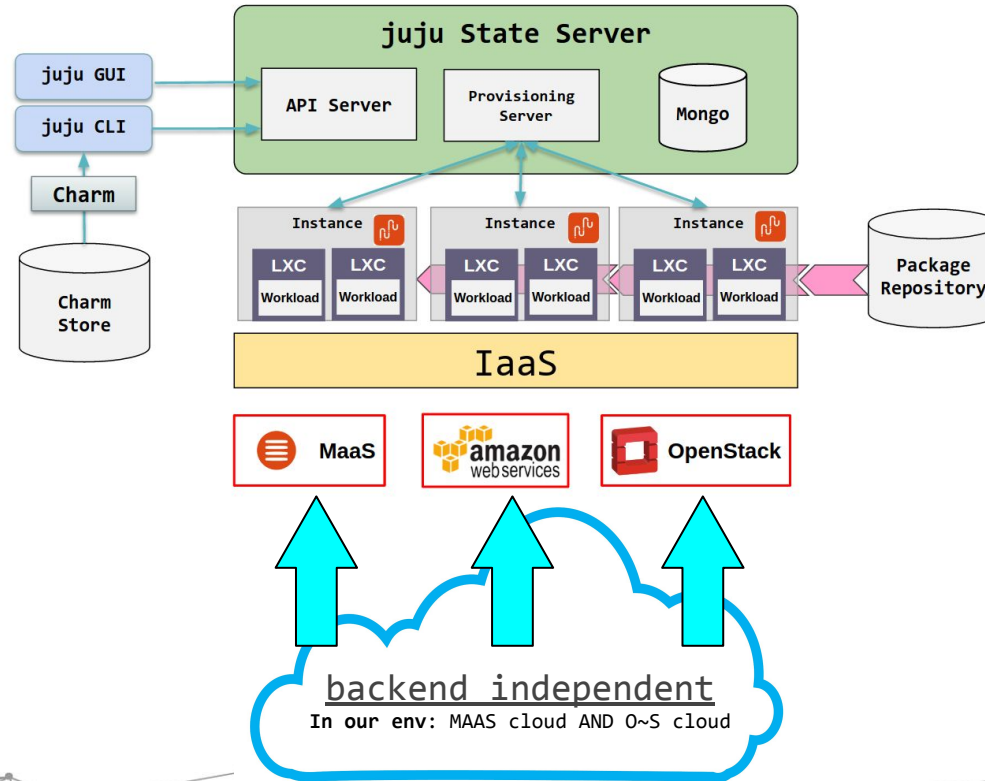
Metal As A Service (MAAS)

- Discovers and deploys physical servers
- Allocates resources to match requirements
- Undeploys servers when they are no longer needed
- Allows cross datacenter provisioning



- Deployment, configuration and management cloud-native tool
- Free and open source
- Exposes a high-level declarative language
- Supports different backends:
 - OpenStack
 - Kubernetes
 - Amazon AWS
 - Microsoft Azure
 - Google CGE
 - VMWare
 - LXD
 - MAAS

Juju Architecture



Juju Charms

- Charm
 - deployment "brick" / microservice
 - a collection of scripts called *hooks* + metadata
 - can expose configurable parameters
 - several charms available online in the charm store
- Hook
 - executable file with a pre-defined name which defines how to handle an event
 - install, start, config-changed, ...
 - can be written in any scripting/programming language

Juju Charms

- Charm relations
 - establish a logical connection between services
 - in practice: configuration parameters are exchanged between charms
- Bundle
 - a YAML file which describes charms, their configurations and their relations

Charm simple example

- Writing a charm can be simple
- Example: wordpress
 - but could be your web application
- Example charm structure:
 - `metadata.yaml`
 - `hooks/install`
 - `hooks/start`
 - `hooks/database-relation-changed`
 - `hooks/website-relation-joined`



```
1  name: wordpress
2  summary: The wordpress CMS
3  maintainer: GARR CSD <cloud-support@garr.it>
4  description: The Wordpress Content Management System
5  tags: [cms]
6  subordinate: false
7  series: [bionic]
8  provides:
9    website:
10      interface: http
11  requires:
12    database:
13      interface: mysql
```



```
1  #!/bin/bash
2
3  status-set maintenance "Installing wordpress"
4  apt-get install -y wordpress
5
6  status-set maintenance "Configuring wordpress"
7  echo "
8      Alias /blog /usr/share/wordpress
9      <Directory /usr/share/wordpress>
10         Options FollowSymLinks
11         AllowOverride Limit Options FileInfo
12         DirectoryIndex index.php
13         Order allow,deny
14         Allow from all
15     </Directory>
16     <Directory /usr/share/wordpress/wp-content>
17         Options FollowSymLinks
18         Order allow,deny
19         Allow from all
20     </Directory>" > /etc/apache2/sites-available/wordpress.conf
21  a2ensite wordpress
22  status-set maintenance "Wordpress installed"
```



```
1 #!/bin/bash
2 status-set maintenance "Starting wordpress"
3 service apache2 restart
4 status-set blocked "Waiting for database relation"
```



```
1  #!/bin/bash
2  database=`relation-get database`
3
4  [ -z "$database" ] && exit 0
5
6  # retrieve the remaining values
7  # which have been set by the mysql charm
8  user=`relation-get user`
9  password=`relation-get password`
10 host=`relation-get private-address`
11
12 # write configuration file
13 echo "<?php
14 define('DB_NAME', '${database}');
15 define('DB_USER', '${user}');
16 define('DB_PASSWORD', '${password}');
17 define('DB_HOST', '${host}');
18 define('WP_CONTENT_DIR', '/usr/share/wordpress/wp-content');
19 ?>" >/etc/wordpress/config-default.php
20
21 open-port 80
22 status-set active "Ready"
23
```



```
1  #!/bin/bash
2
3  relation-set "hostname=$(unit-get private-address)"
4  relation-set "port=80"
5
```

Deploying a cloud application with Juju

```
# deploy a local web application and a database from the charm store
```

```
juju deploy ./wordpress
```

```
juju deploy mysql
```

```
# and connect them
```

```
juju add-relation wordpress mysql
```



Cloud applications and high availability

- Deploying applications on the cloud introduces a paradigm shift:
 - any component of the cloud infrastructure can fail at any moment
 - as opposed to expensive, redundant hardware
 - high availability has to be moved to higher layers of the stack

“no single component can guarantee 100% uptime (and even the most expensive hardware eventually fails)” - Netflix technology blog

Chaos Engineering - Chaos Monkey

- Chaos Monkey randomly terminates virtual machine instances and containers that run inside of your production environment
- The name comes from the idea of unleashing a wild monkey with a weapon in your data center (or cloud region) to randomly shoot down instances and chew through cables — all the while we continue serving customers without interruption
- Exposing engineers to failures more frequently incentivizes them to build resilient services





Edit

Web IDE

Replace

Delete

```
1 series: bionic
2 applications:
3   wordpress:
4     charm: "/home/ubuntu/wordpress"
5     num_units: 3
6     to: [ "0", "1", "2" ]
7   percona-cluster:
8     charm: cs:percona-cluster-279
9     num_units: 3
10    to: [ "3", "4", "5" ]
11   haproxy:
12     charm: cs:haproxy
13     num_units: 3
14     to: [ "3", "4", "5" ]
15     expose: true
16     options:
17       peering_mode: active-active
18 machines:
19   "0": {}
20   "1": {}
21   "2": {}
22   "3": {}
23   "4": {}
24   "5": {}
25 relations:
26 - [ wordpress:database, percona-cluster:db ]
27 - [ wordpress:website, haproxy:reverseproxy ]
```

3 applications

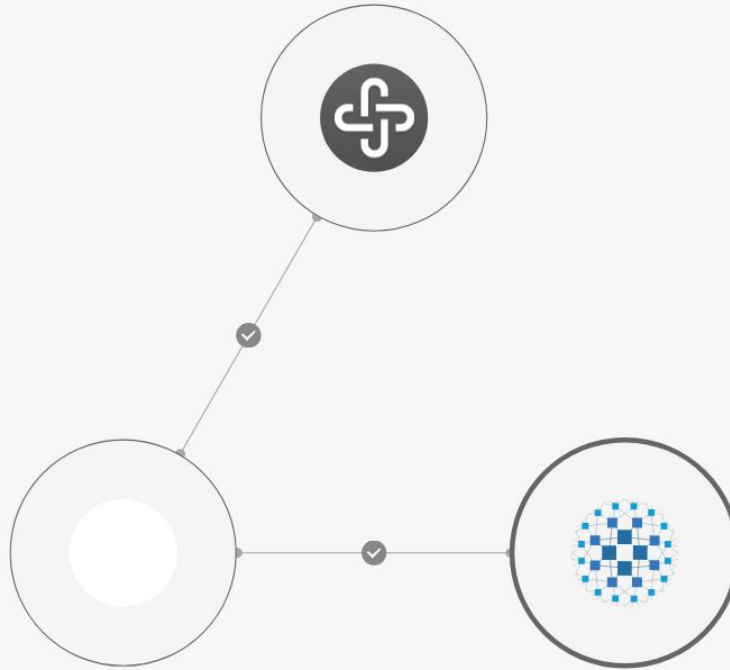
3 machines

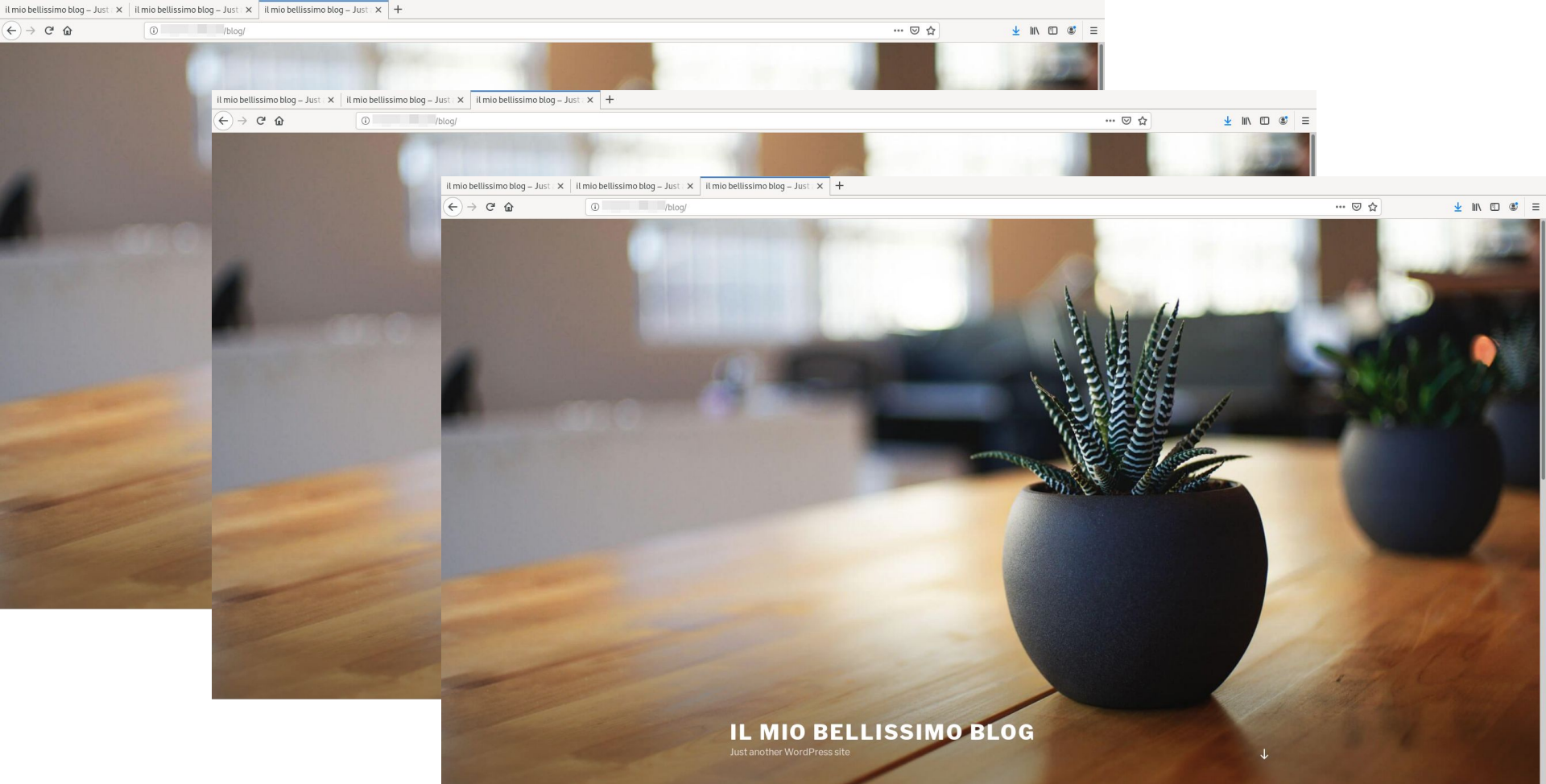
status

3 wordpress

3 haproxy

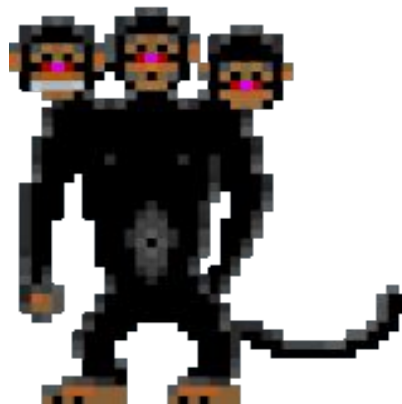
3 percona-cluster





Wrap up

- we wrote a simple charm for our web application
- we deployed our application on a cloud infrastructure
- we used Juju to compose our charm with other charms to obtain a scalable, highly available application
 - monkey-proof



Takeaways

- We can use Juju in several ways:
 - to deploy and manage cloud infrastructures
 - to deploy and manage scalable, highly-available applications
 - to make our own applications scalable and highly-available
- Juju is cloud agnostic
 - the same tool for private, community and public clouds
 - the same tool for physical hardware up to pods
- Note that Juju can use other configuration management systems (e.g. Ansible) for its hooks
- Juju limitations:
 - needs an agent to be installed (i.e. it is not OS independent)
 - does not manage networking appliances
 - needs scripting/programming skills to implement complex charms



<https://www.hackthecloud.it/>



Grazie

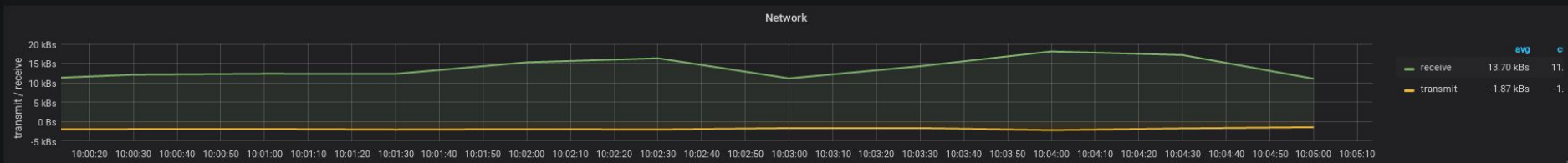
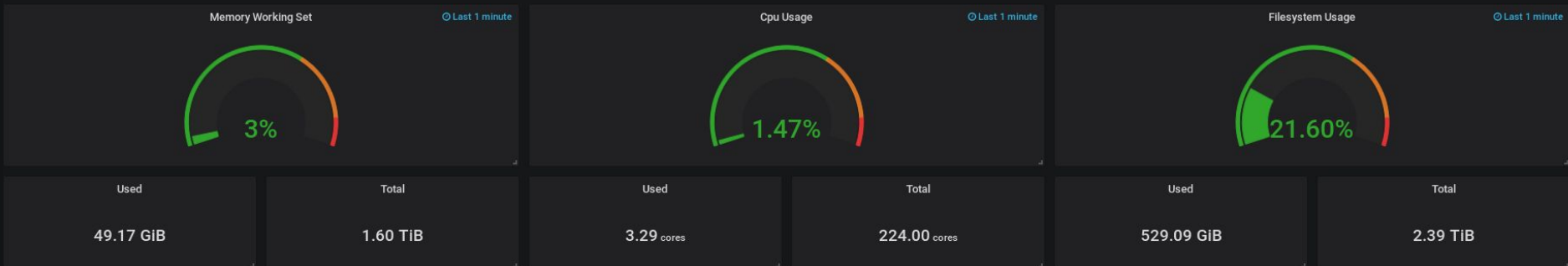
<https://cloud.garr.it>

<mailto:claudio.pisa@garr.it>

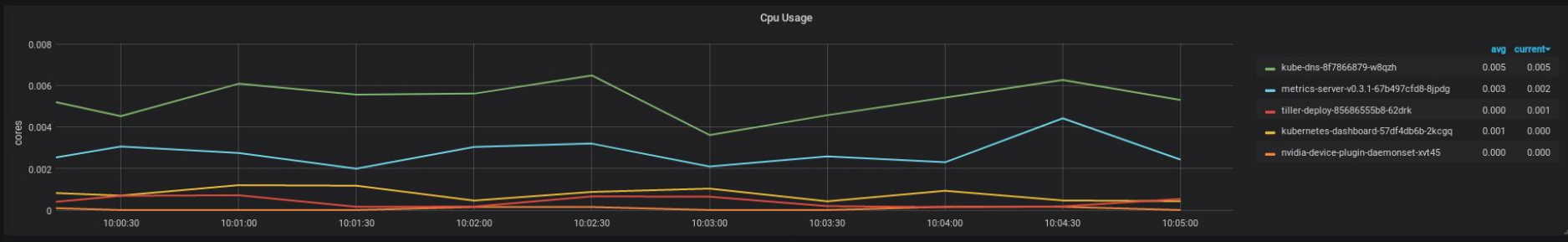
<mailto:cloud-support@garr.it>



all pods



each pod



GARR Cloud: the crew

Alex Barchiesi

Alberto Colla

Fulvio Galeazzi

Claudio Pisa



43

GARR Cloud: the crew

Alex Barchiesi

Alberto Colla

Fulvio Galeazzi

Claudio Pisa

New!

Matteo Di Fazio

New!

Marco Lorini

New!

Delia Passalacqua



44

Juju vs the rest of the world

vs ansible, terraform, puppet

Juju Layers

docker layer

ansible layer

GARR Computing and Storage



600 Users

1100 VMs

3500 vCPUs 9 TB_RAM

1 PB_storage 900 IPv4

Container platform
(GPU inside)

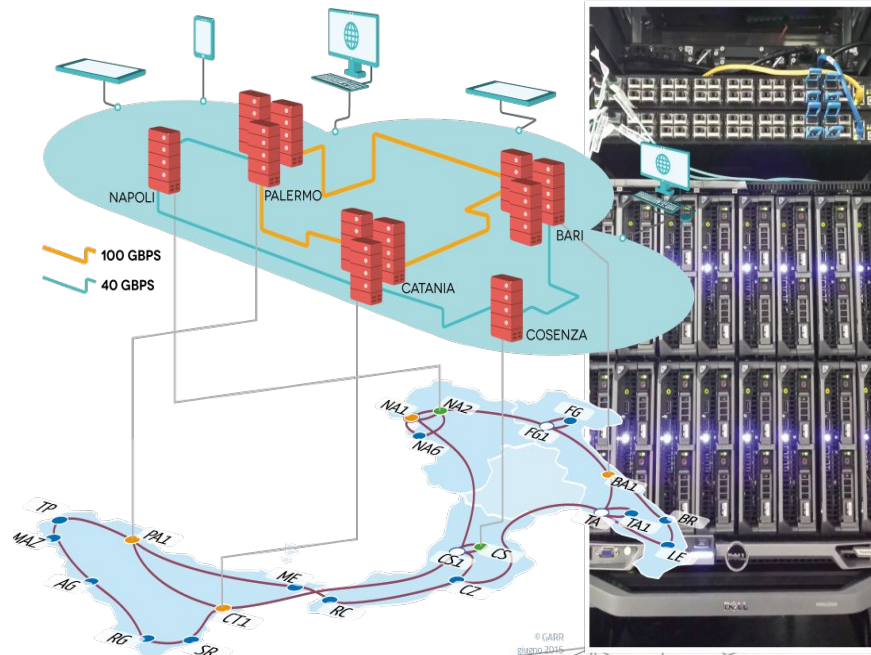


The engine


MAAS


JUJU


ceph



DaaS (deployment a.a s.)

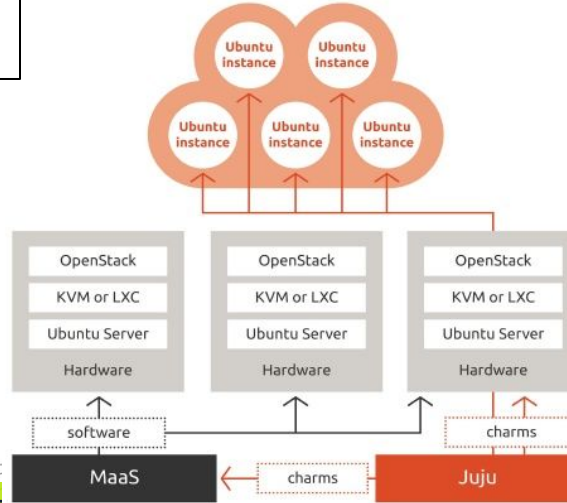
more powerful than a PaaS, easier than a IaaS



Frontend UI/CLI

Juju on OpenStack Cloud

backend



=



<https://daas-pa.cloud.garr.it>
<https://daas-ct.cloud.garr.it>

Delegation of authority (via domains, policy and metadata filters)

REQUIREMENT: You stay in control of your resources

manages:

- identity only
- region inclusion

cloud Admin

LOCAL REGION (Domain)

physical HW

Virtual Datacenter

Region Admin

inside a physical datacenter manages:

- quota
- HW (hypervisors)
- Network
- VdC assignment to Projects
- Physical resources reservation

Region

organization

organization

(project scope)

Project Admin

Project Admin

Project Admin

inside a project manages:

- membership roles
- Internal Networks

Project Admin

member

member

member

member

member

inside his project manages:

- VM
- Volumes

member

member

(few) details about deployment
requirements

networking prescriptions

1. private network

- 1.1. ipmi (usually untagged)
- 1.2. **pxe/boot/management (best practice untagged - simplifies the setup of pxe)**
- 1.3. Storage ceph “priv” (to be used by OSDs only)
- 1.4. Storage ceph “pub” (used by MON and Ceph clients, it normally is a private network)
- 1.5. OpenStack data+mgmt

2. public network

- 2.1. Public ip (for infrastructure and Cloud service frontend)
Needed minimum 40 IP (suggested to have a 3 members HA of each service: /25 subnet)
- 2.2. OpenStack floating ip - (this is basically the number of VM that you foresee to have publicly exposed)
To be evaluated according to computing resources/use case

Best practice:

- link aggregate (bond) all interfaces and set a virtual interface (vlan) for each of the previously mentioned networks, except PXE.
- keep IPMI network separated.
- NB configure ILO/IDRAC with IPMI over LAN

Note: The networking must be configured on the switches/routers while MAAS takes into account the server configuration (ref <https://docs.maas.io/2.5/en/installconfig-networking>)

firewall: ports needed

egress: 80, 443, 5000, 35357, 8774, 8776, 8778, 9292, 9696, 6080

ingress (only needed on public network from subnets which need access to OpenStack and NB from GARR-keystone network): 80, 443, 8774, 8776, 8778, 9292, 9696, 6080, 5000

N.B. open port 5000 to test “local” keystone functionality, after including the region in the federation it can be closed again.

Automation and deployment: MAAS + juju

MAAS (responsible of physical machine deployment - ref: <https://maas.io/>)

Note MAAS is not a demanding service in terms of CPU, RAM, Disk, Bandwidth.

1 physical machine (2 if in HA):

- LXD host (MAAS + Juju + OpenStack clients)
- Hypervisor host (mainly for VM hosting Juju controller)

container LXC:

- Region controller (needs to reach ipmi network (ipmi network routed towards MAAS)
- Rack controller (may be co-located with Region)
- NAT (Note this is a Gateway for outgoing requests of services with only private networks, e.g. installation/upgrade of DB)

Best practice:

- configure HA on the hosted services
- Link aggregation of network interfaces (bond) + virtual interfaces (vlan)
- You can also install Region and Rack controller on the same LXC
- You can install juju and OpenStack client on the same LXC container (may be also deployed on separate containers)
- NAT configuration: (iptables -t nat -A POSTROUTING -o <NIC_NAME> -j MASQUERADE)

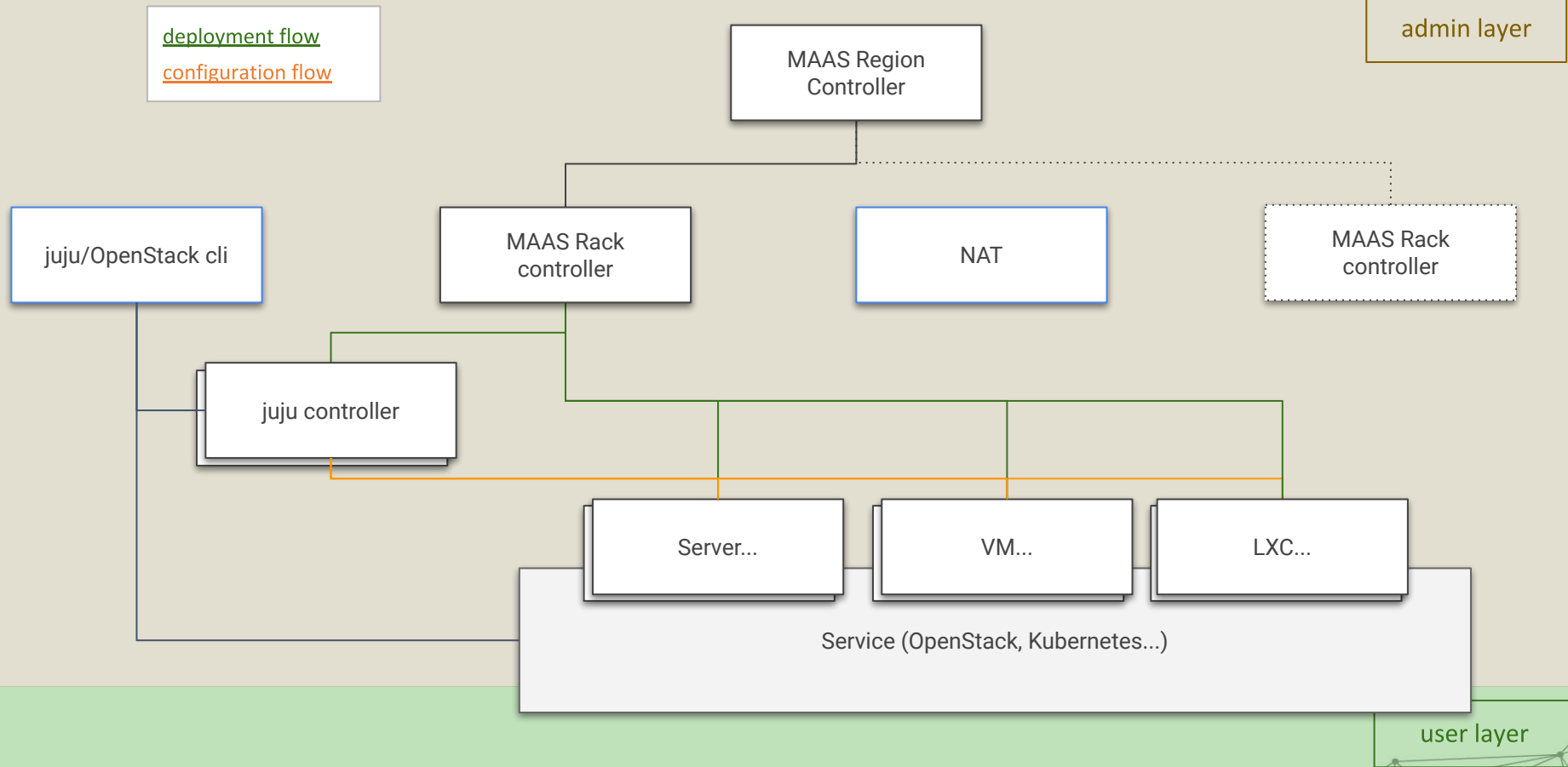
juju controller (ref: <https://docs.jujucharms.com/2.5/en/>)

- will be *bootstrapped* under MAAS supervision as a VM (Best-practice: several VMs on separate servers for HA)

Note juju controller is a key service to deploy and manage any service (i.e. OpenStack)

[deployment flow](#)
[configuration flow](#)

admin layer





next steps

governance

accounting (blockchain?)

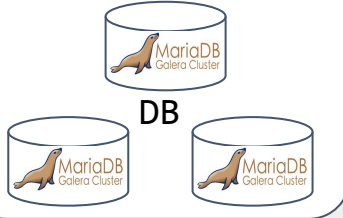
federation of federations (keystone to keystone trust mesh)

Global Services

keystone

keystone

keystone



glance

glance

glance

Swift proxy

Swift proxy

Swift proxy



Object storage

Servizi locali

Servizi locali

Global Services

DNS

HA proxy

keystone



DB



glance

Swift proxy



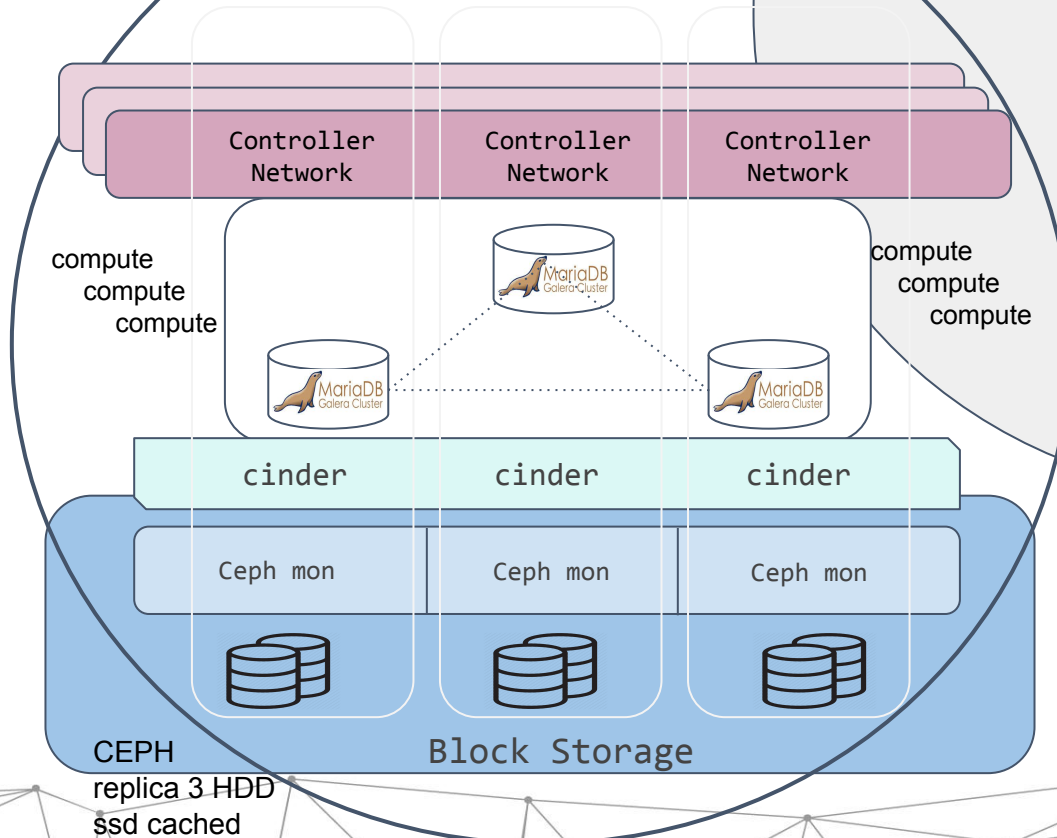
Object storage

local services

local services

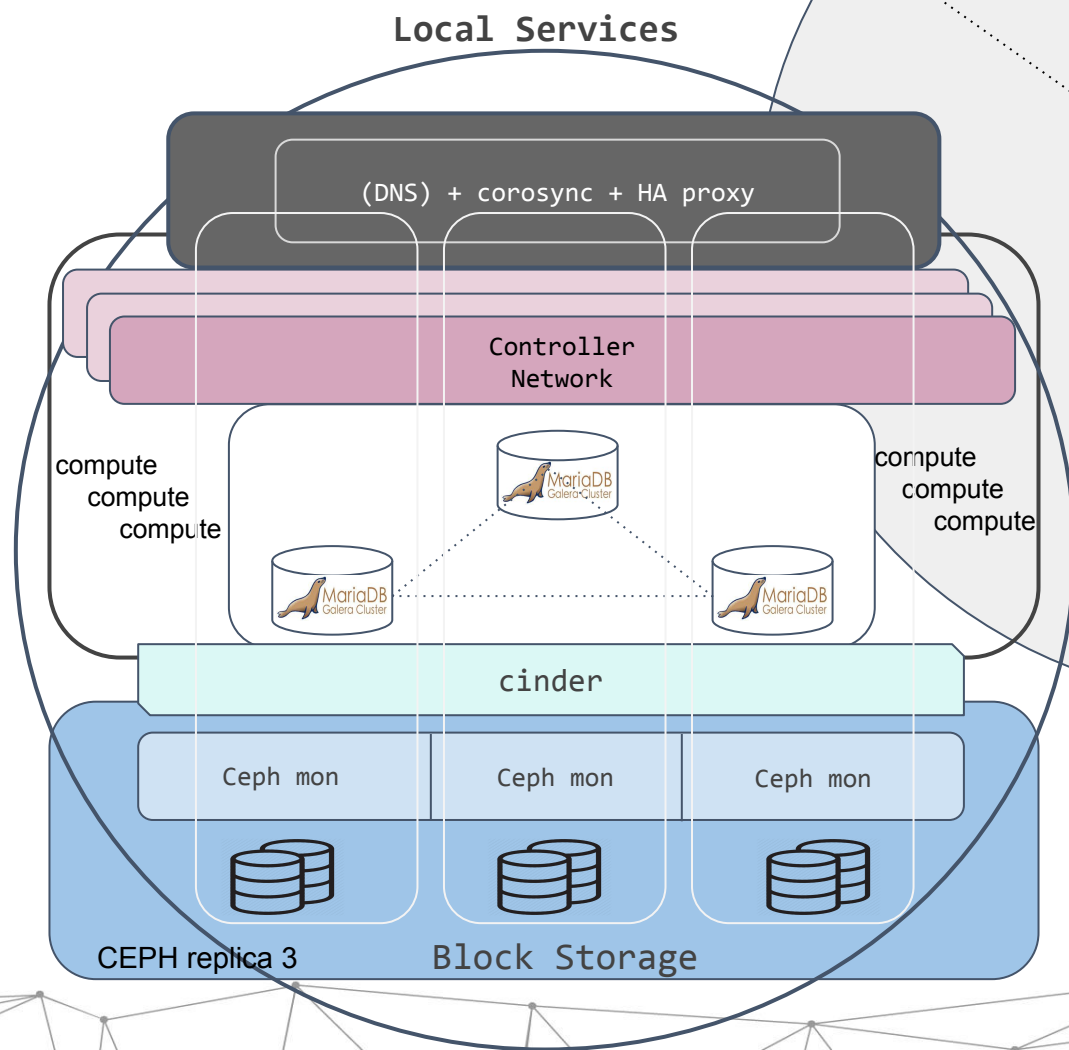
local Services

Global
Services

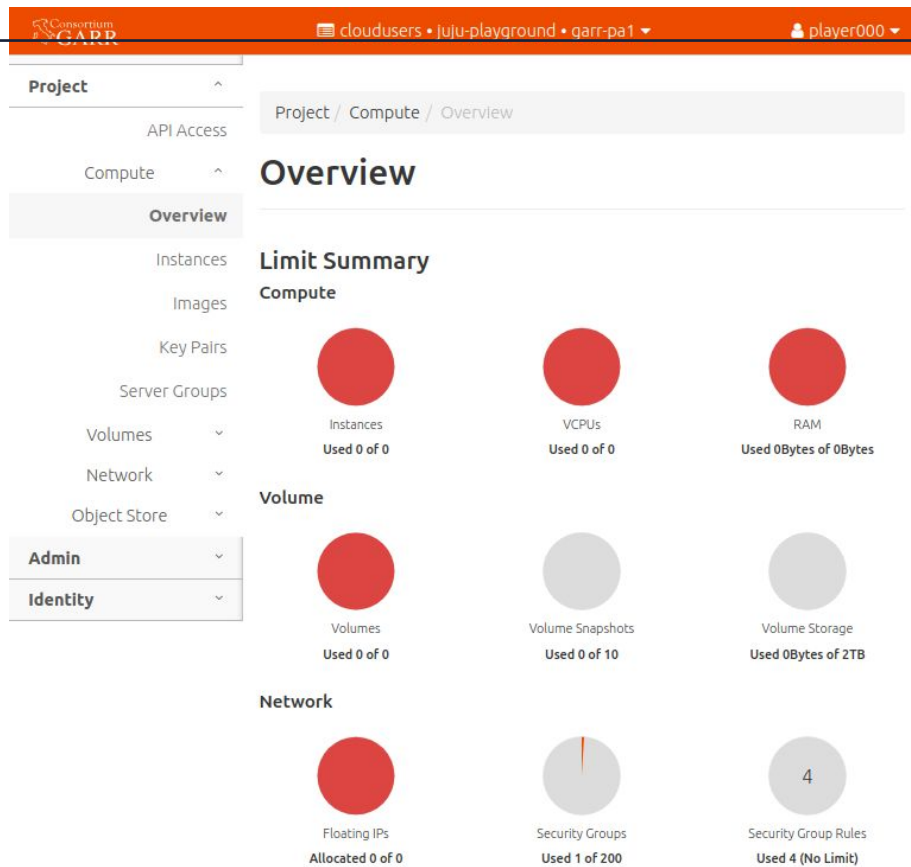
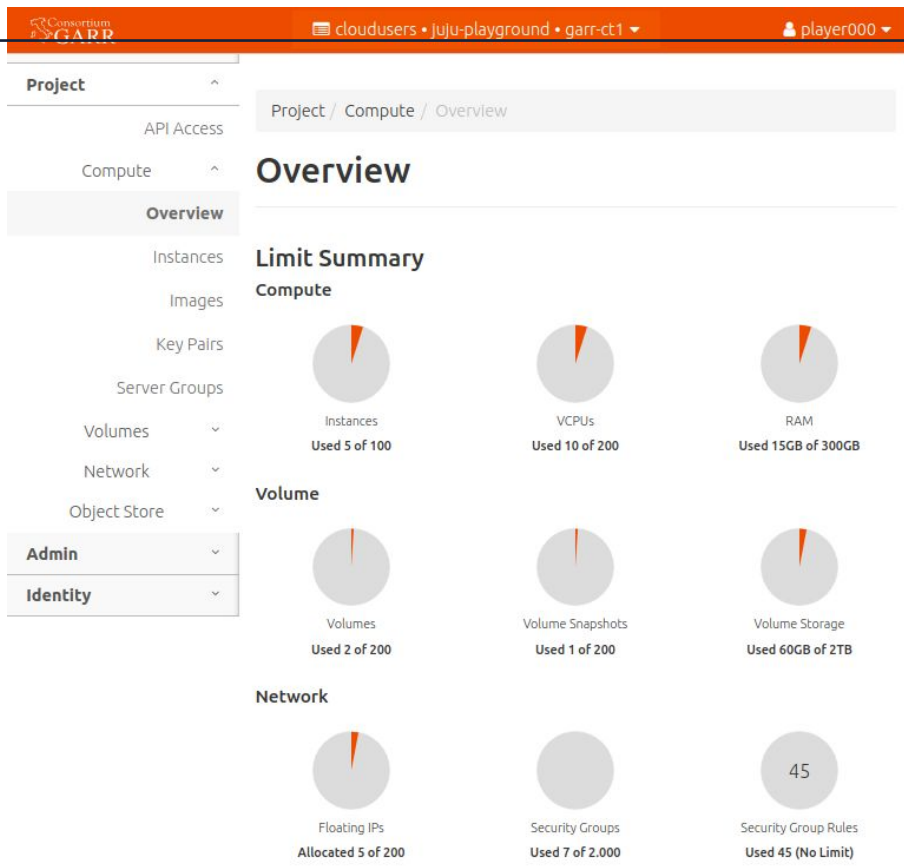


Local Services

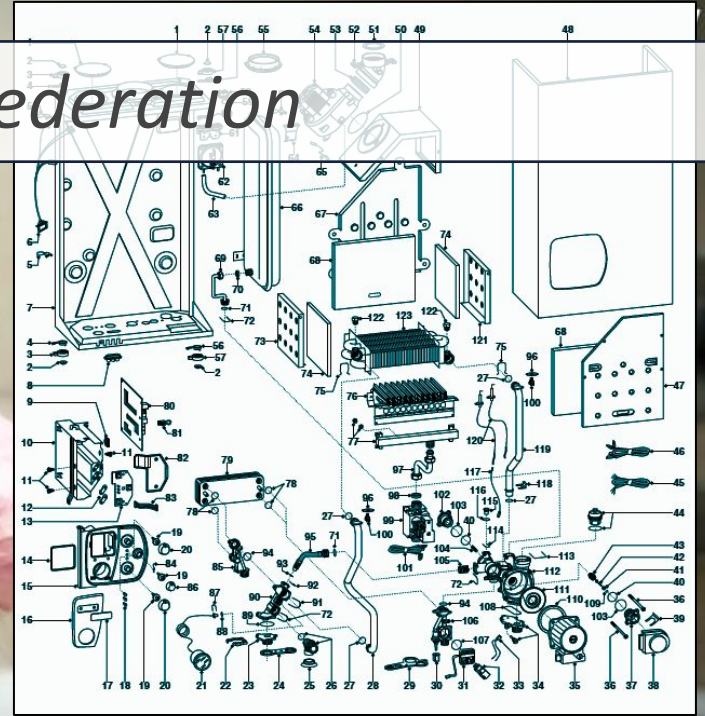
Global
Services



Regions, projects and quotas



our concept of *Federation*



the simplest (for federated region admin) the better
the less requirements the more inclusive

Ceph

4 Layers recipe:

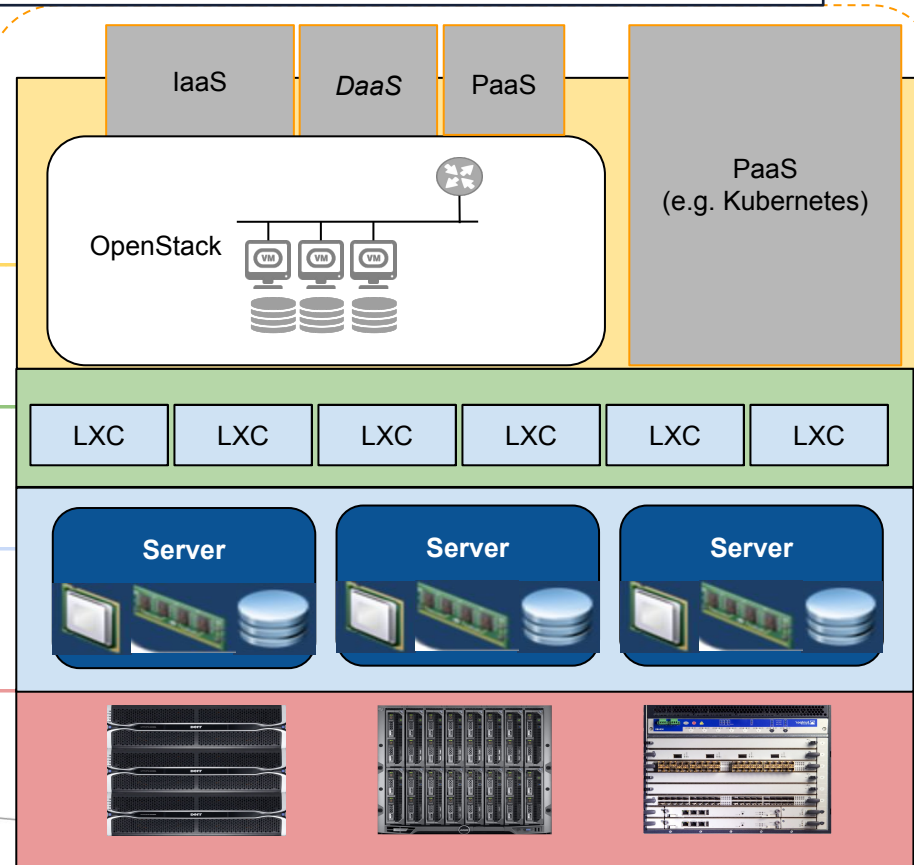


1. Application Services

2. Infrastructure *Virtualization*

3. Operating System

4. Physical resources



4 Layers recipe:

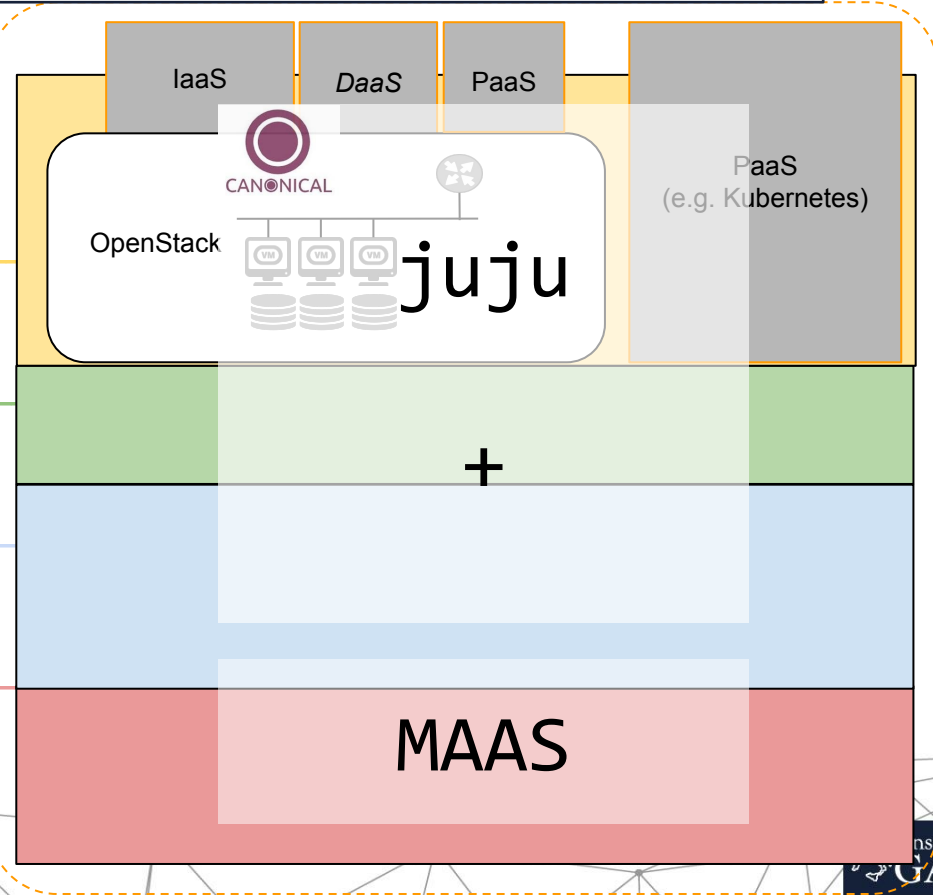


1. Application Services

2. Infrastructure *Virtualization*

3. Operating System

4. Physical resources



-Operating system: **Ubuntu Xenial**

-4 NIC - link aggregation **40Gbps**

-vlan + **Bridge** linux

-**blade** is the GW for LXC

-**iptables**:

-fwd, nat + security LXC

-blade interchangeable from p.v. LXC

MAAS + juju*

LinuxContainer (LXC with LXD management)

-opensource

-modern kernel support

-Efficient resource usage (compared to VM)

-Near Bare Metal runtime performance

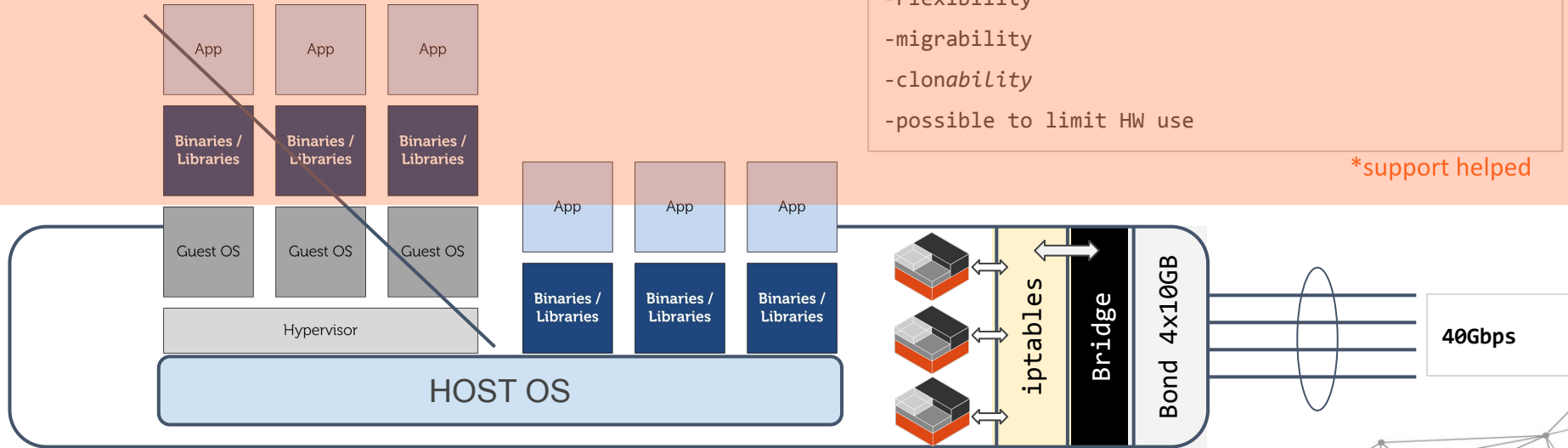
-Flexibility

-migrability

-clonability

-possible to limit HW use

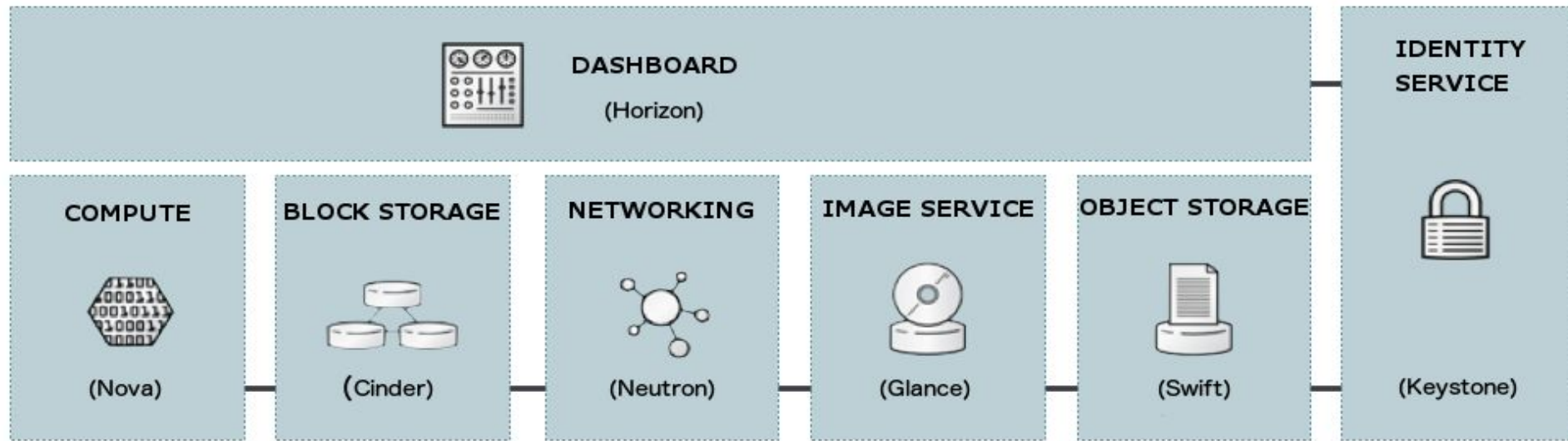
***support helped**



blade

OpenStack Components (umbrella project for)

- **Horizon** (Dashboard)
- **Keystone** (Identity Management)
- **Nova** (Compute, where VMs are run)
- **Glance** (Image Service, where templates are)
- **Cinder** (Block Storage, persistent storage for VMs)
- **Swift** (Object Storage, snapshots and not frequently updated data)
- **Neutron** (Networking and SDN)

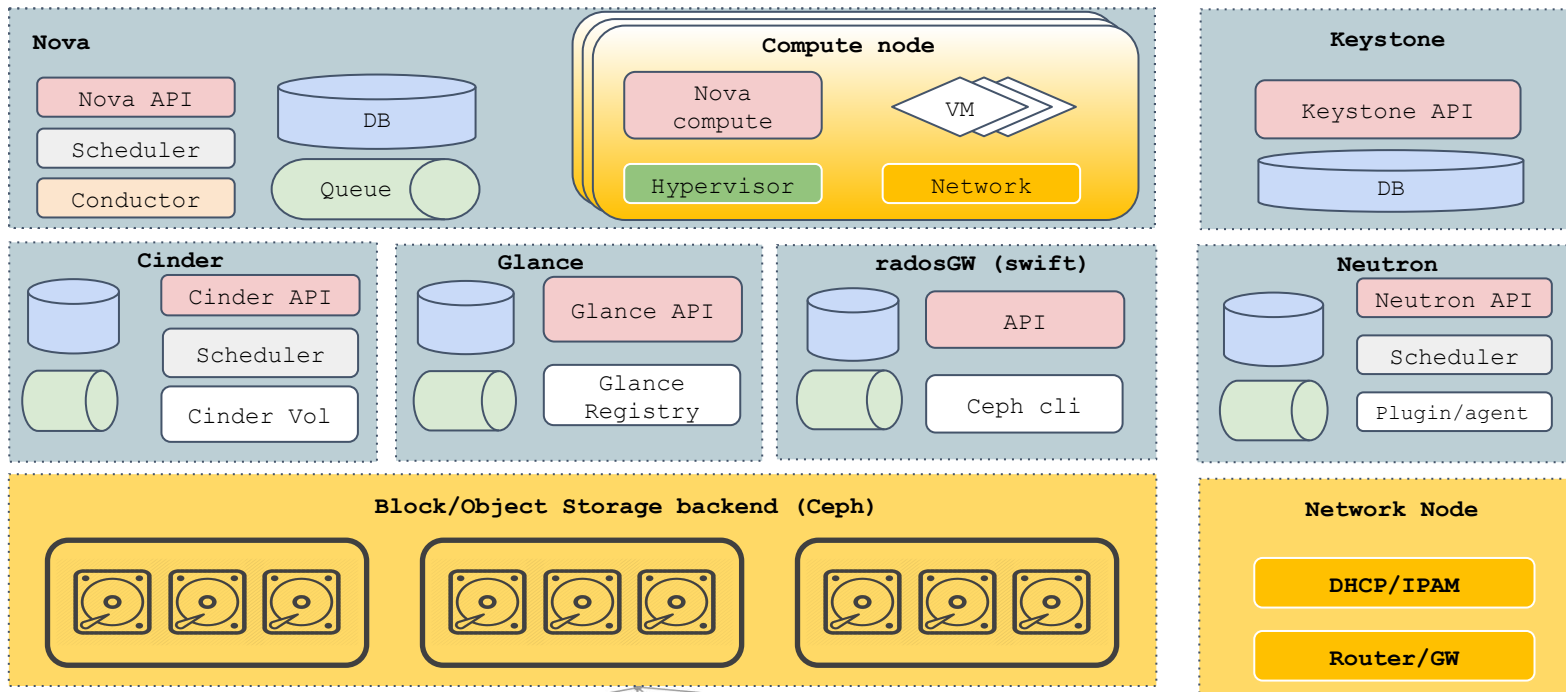


OpenStack through VM provisioning

- Most common and complex process
- Involves interaction of most components



Dashboard:
Horizon or CLI



what is available till now: quite a lot...

3 + 2 regions up and running

4 deployments openstack for a total of about **20.000 vcpu** (and counting till 100k - 120k)

more than **500** users (demo account for 6 months)

Virtual Data Centers delivery in minutes (~500)

Physical Data Center administrative **delegation** (you administer what you own and offload to other regions)

DaaS a GARR hack for advanced PaaS or simplified IaaS (via juju with OpenStack cloud backend)

Kubernetes container platform 4 NVIDIA GPUs on bare metal

Federated access (SAML-idem and OIDC-google)

but the important thing is...

simple **Federation *recipe*** (git and knowledge base available, references before)

Reference **architecture** with use cases

Deployment of OpenStack bare metal (+LXD) region up & running in a **couple of hours**

7 federated (3 federations using the model) regions (ongoing):

- HPC4AI project,
- Politecnico To,
- Uni Padova,
- 3 INGV region (external federation, see confederation)
- EAPConnect (East EU Nren, external federation)
- Hungary
- RECAS INFN Bari (maintenance)

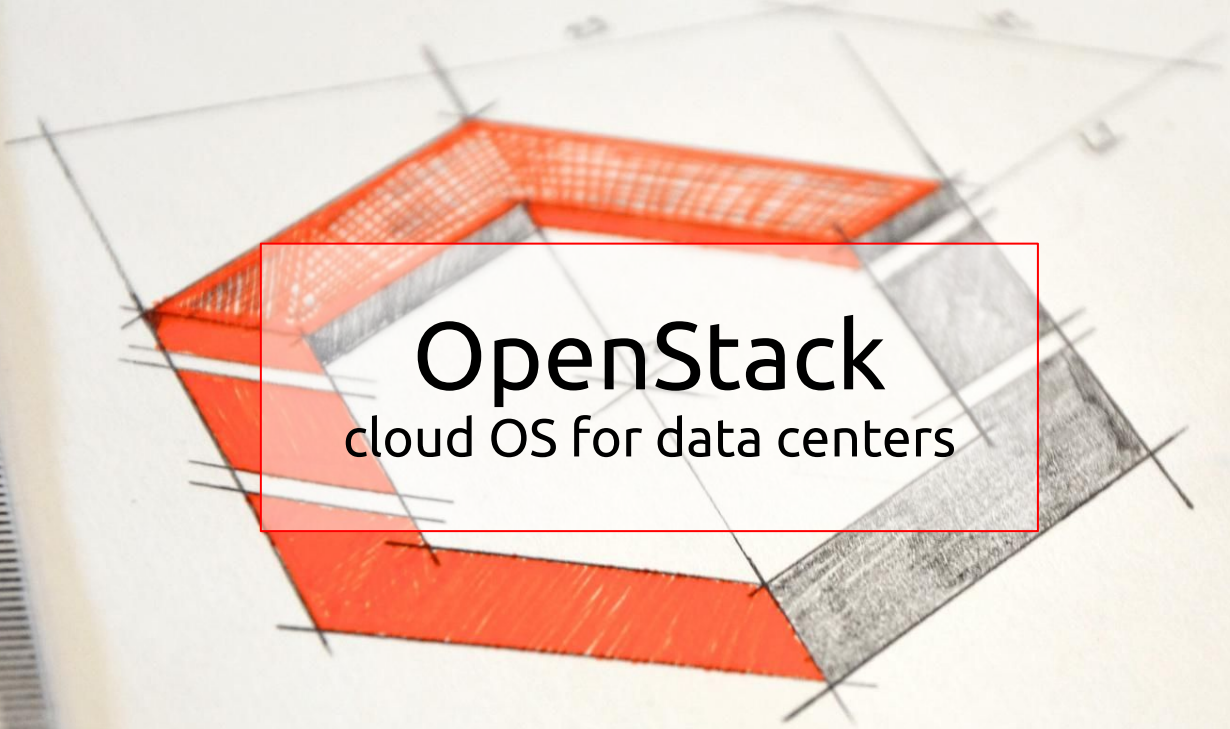


federation recipe:

<https://cloud.garr.it>

<https://git.garr.it/cloud/federation>

“To produce the ubiquitous Open Source cloud computing platform that will meet the needs of public and private cloud providers regardless of size, by being simple to implement and massively scalable.”

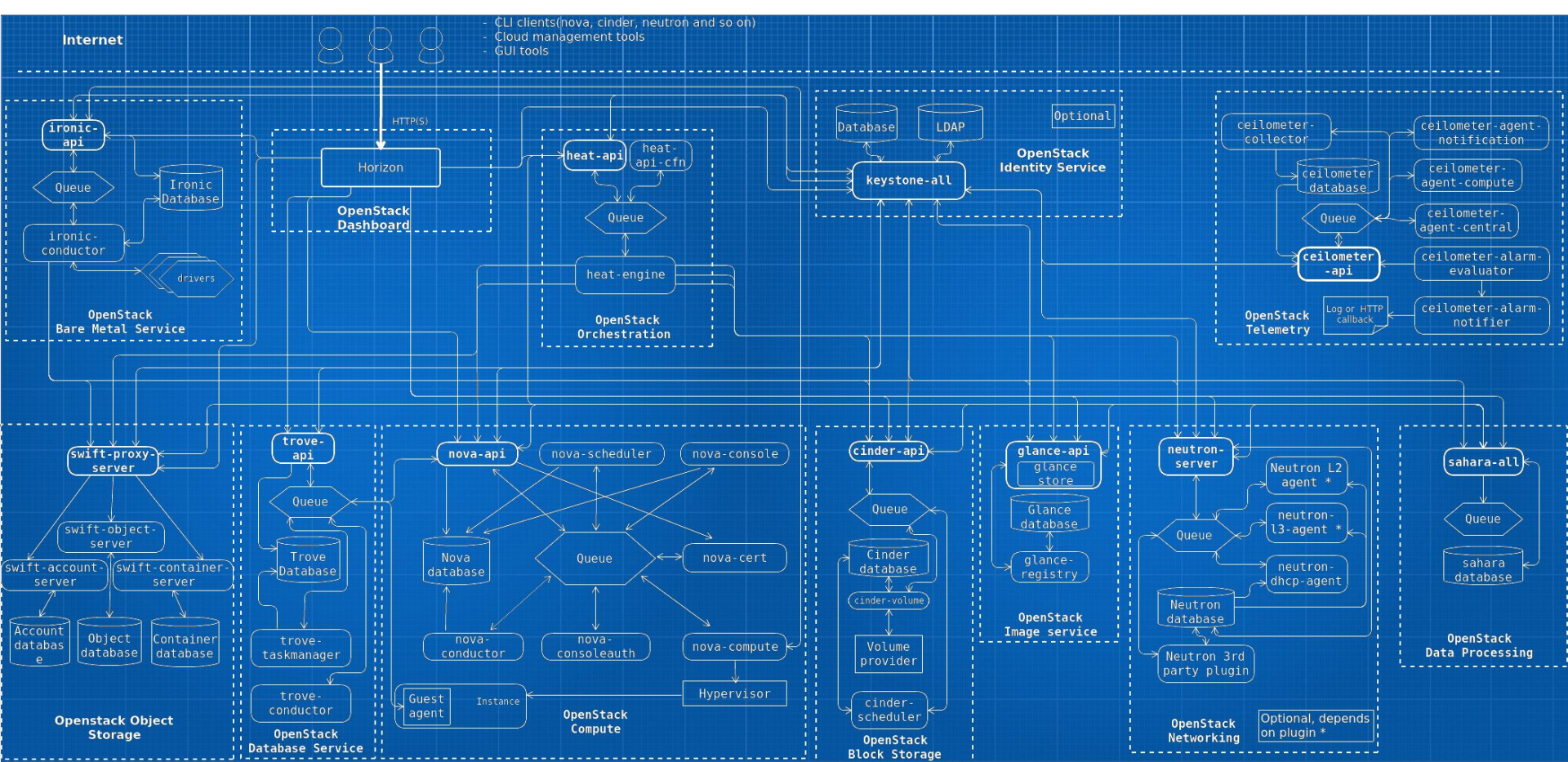


OpenStack

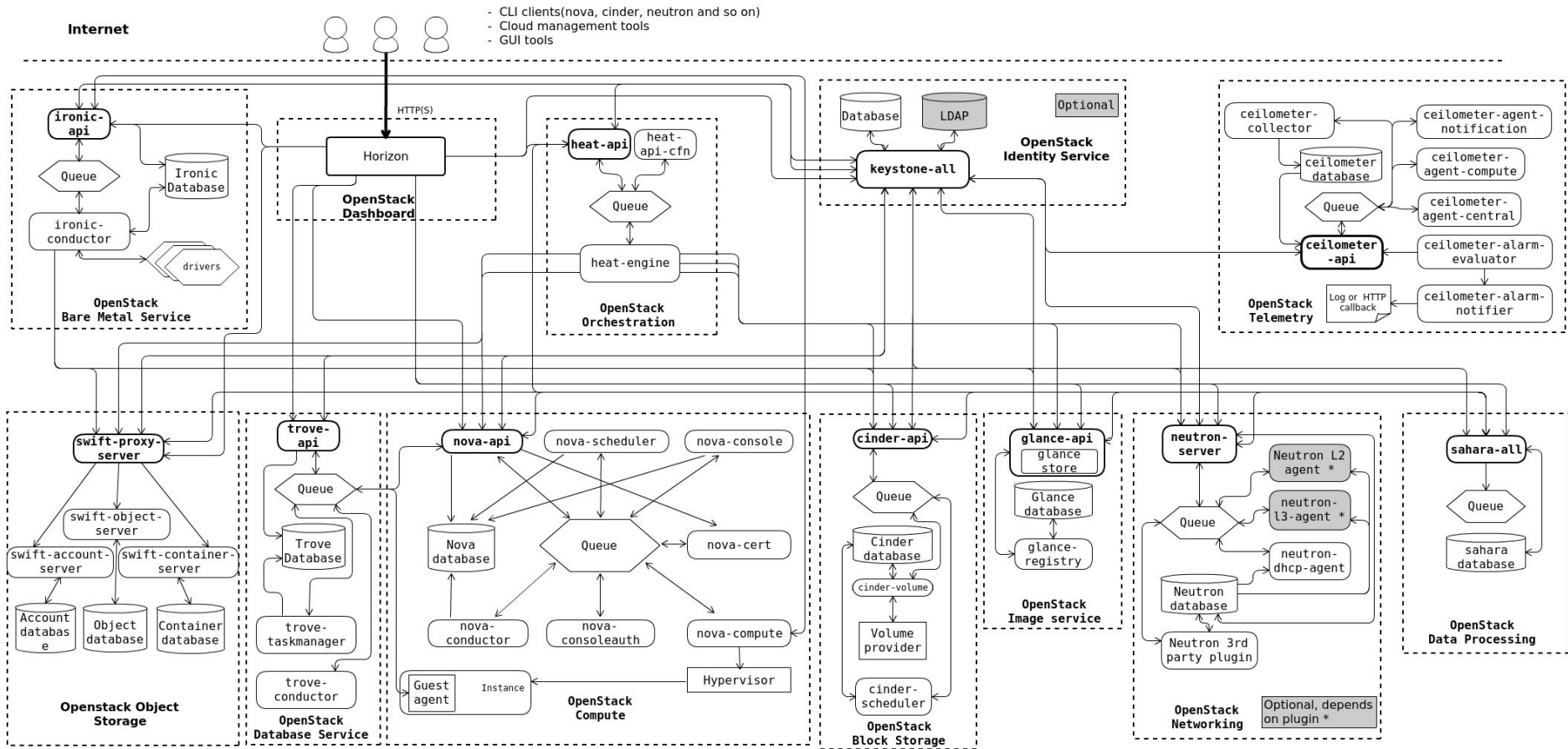
cloud OS for data centers



- Juju allows configuring, managing, maintaining, deploying and scaling cloud services quickly and efficiently on multiple providers:
 - private or public clouds
 - bare metal, leveraging MAAS to control the hardware
- Uses script collections called Charms which specify how to deploy a service
- Juju can manage and scale models consisting of many charms, creating complex architectures like OpenStack.
- Juju can be controlled via a web GUI, the command line, or API.



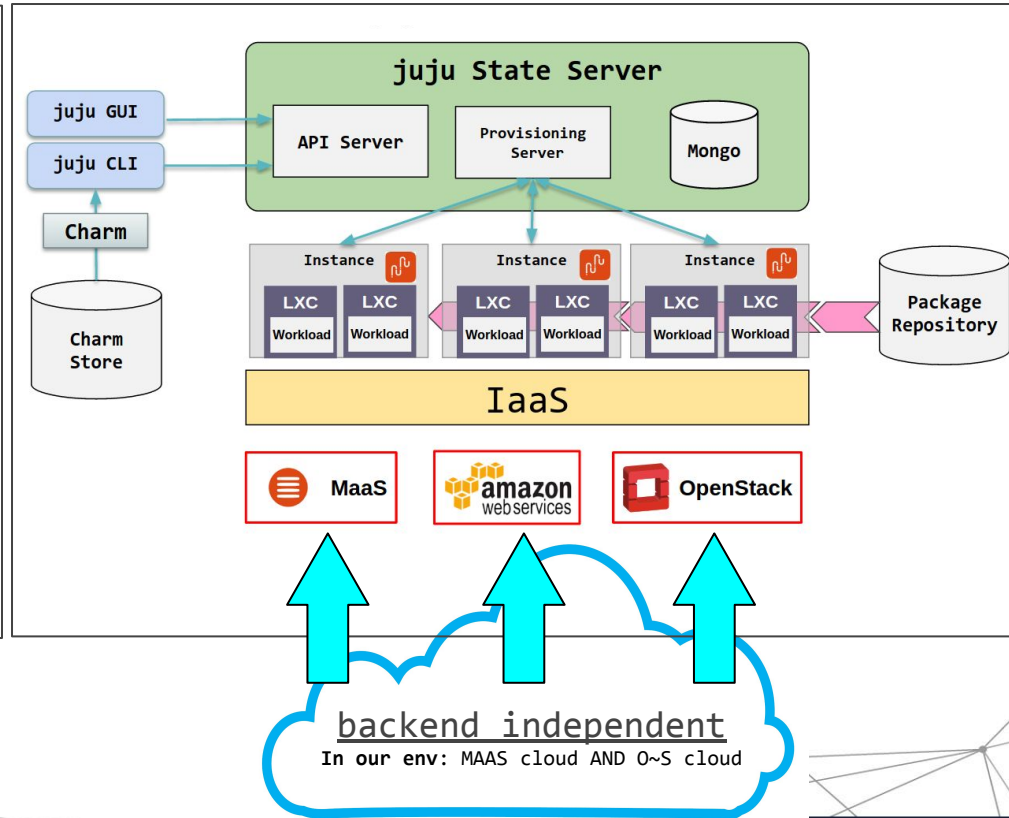
OpenStack Architecture



OpenStack Architecture

juju workflow

- engine follows a **reactive pattern**, triggered by events that cause hook handlers to run
- Multiple **handlers** may match for a given hook and **run** in a **non-determined order**
- The **state engine** is **evaluated** every event occurred
- The engine **runs until convergence** to a stable state



GARR Computing and Storage

A word cloud on a dark blue background. The central and largest text is 'Infrastructure-as-Code' in a light yellow-green color. Surrounding it are various other terms in different shades of purple and white, including 'Security', 'Documentation', 'replicable', 'Automation', 'scalable', 'ISO27001', 'self-healing', 'almost-no-manpower', 'delegation-of-authority', 'resource-sharing', 'Monitoring', 'open-source', and 'self-deploying'.

self-healing
ISO27001
scalable
Automation
Security
almost-no-manpower
replicable
Documentation
Infrastructure-as-Code
delegation-of-authority
resource-sharing
Monitoring
open-source
self-deploying

Deploying an HA cloud application with Juju

```
# deploy a local web application and a database from the charm store
```

```
juju deploy ./wordpress
```

```
juju deploy mysql
```

```
# and connect them
```

```
juju add-relation wordpress mysql
```

Deploying an HA cloud application with Juju

```
# implement high-availability

juju remove-application mysql

juju deploy percona-cluster

juju add-relation wordpress percona-cluster

juju deploy haproxy

juju add-relation wordpress haproxy

juju config haproxy peering_mode=active-active

juju expose haproxy
```

```
# scale the application

juju add-unit -n 2 percona-cluster

juju add-unit -n 2 wordpress

juju add-unit -n 2 haproxy
```

Metal As A Service (MAAS)

- Discovers and deploys physical servers
- Allocates resources to match requirements
- Undeploys servers when they are no longer needed
- Allows cross datacenter provisioning

