European Weather Cloud: A community cloud service tailored for Meteorology

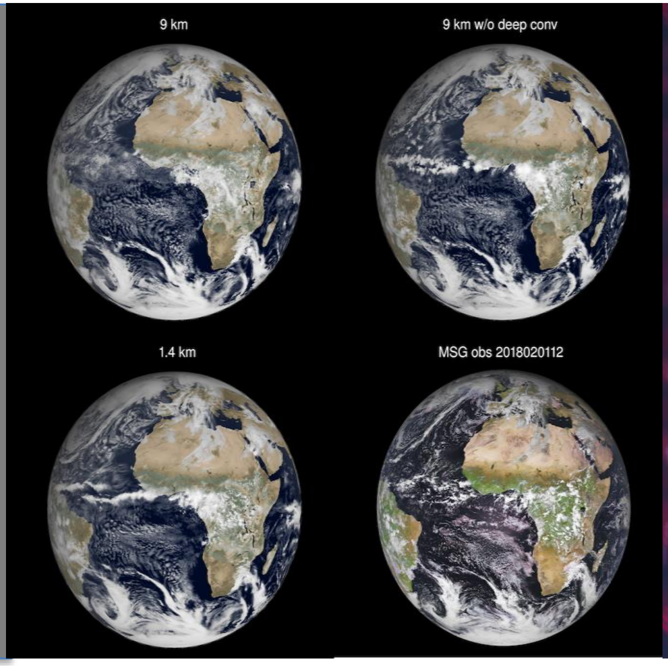
19th Workshop on high performance computing in meteorology
September 20-24, 2021

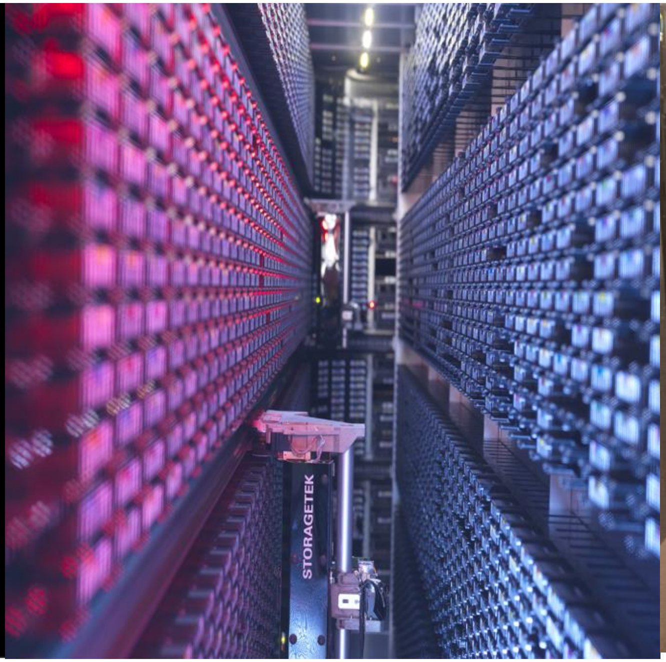
Vasileios Baousis, Umberto Modigliani, Florian Pappenberger, Martin Palkovic, Stephan Siemen, Xavier Abellan, Charalampos Kominos

Vasileios A. Baousis (PhD)

ECMWF EUMETSAT EUROPEAN WEATHER CLOUD
CLOUD COMPUTING-BASED INFRASTRUCTURE, FOCUSED ON THE NEEDS OF THE METEOROLOGICAL COMMUNITY

What is the Vision?





*The European Weather Cloud aims to become the **cloud-based collaboration platform for meteorological application development & operations in Europe** and contributes to the digital transformation of the European Meteorological Infrastructure*

"a **community** cloud"

ECMWF EUMETSAT EUROPEAN WEATHER CLOUD
CLOUD COMPUTING-BASED INFRASTRUCTURE, FOCUSED ON THE NEEDS OF THE METEOROLOGICAL COMMUNITY

European Weather Cloud

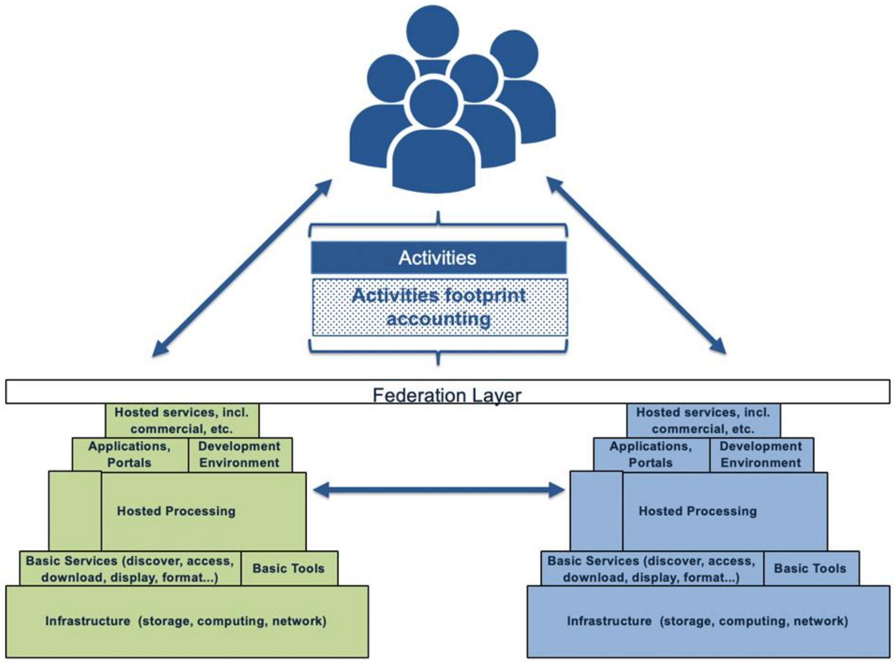
www.europeanweather.cloud

- Three-year pilot project started in January 2019 -ECMWF and EUMETSAT
  - ➢ Basic goal is **to bring the computation resources (Cloud) closer to our Big data** (meteorological archive and satellite data)

- The project includes:
  - Building infrastructure
  - Organising and implementing use cases
  - Addressing challenges: technical, policy, governance

- ECMWF's Pilot infrastructure was built with open

 source software-Ceph and Openstack.

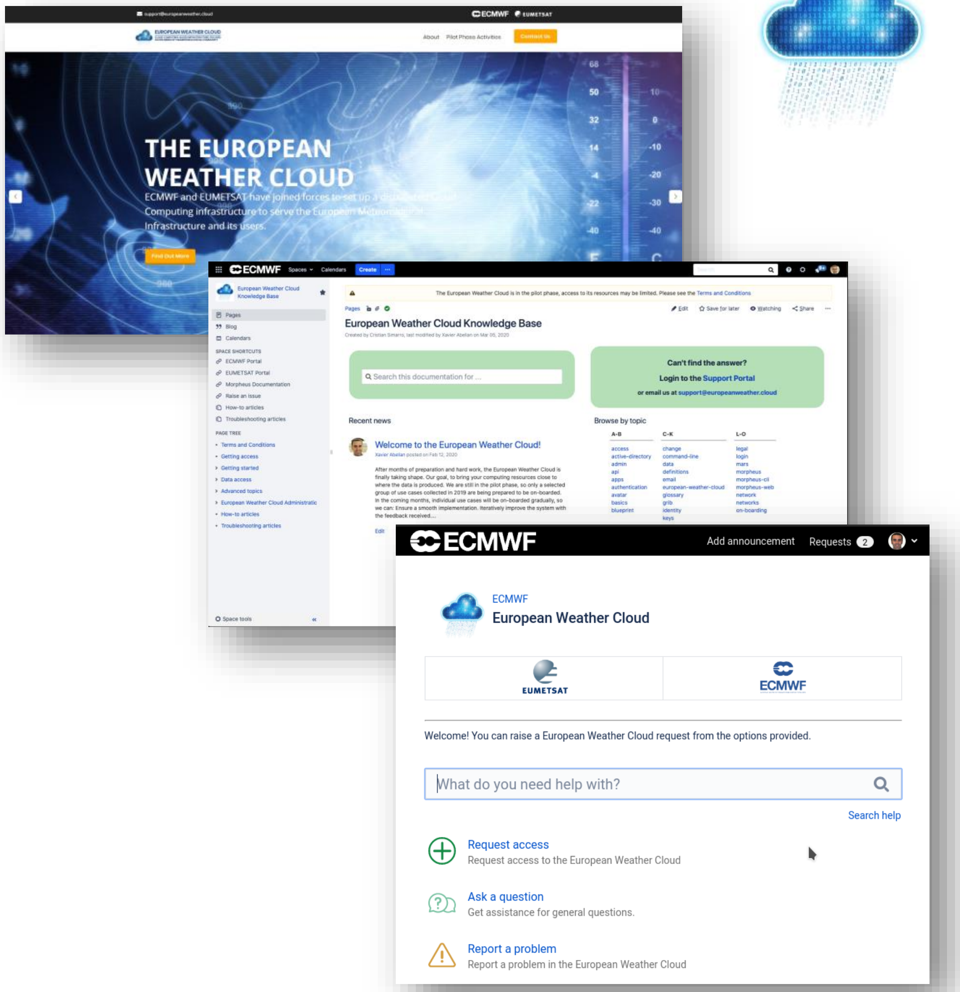


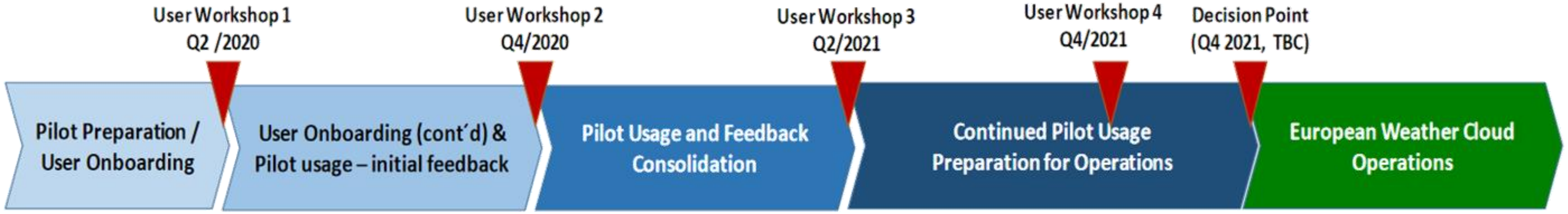
European Weather Cloud is a Community Cloud- EMI (E&E and MS)

ECMWF EUMETSAT EUROPEAN WEATHER CLOUD
CLOUD COMPUTING-BASED INFRASTRUCTURE, FOCUSED ON THE NEEDS OF THE METEOROLOGICAL COMMUNITY

European Weather Cloud timeline

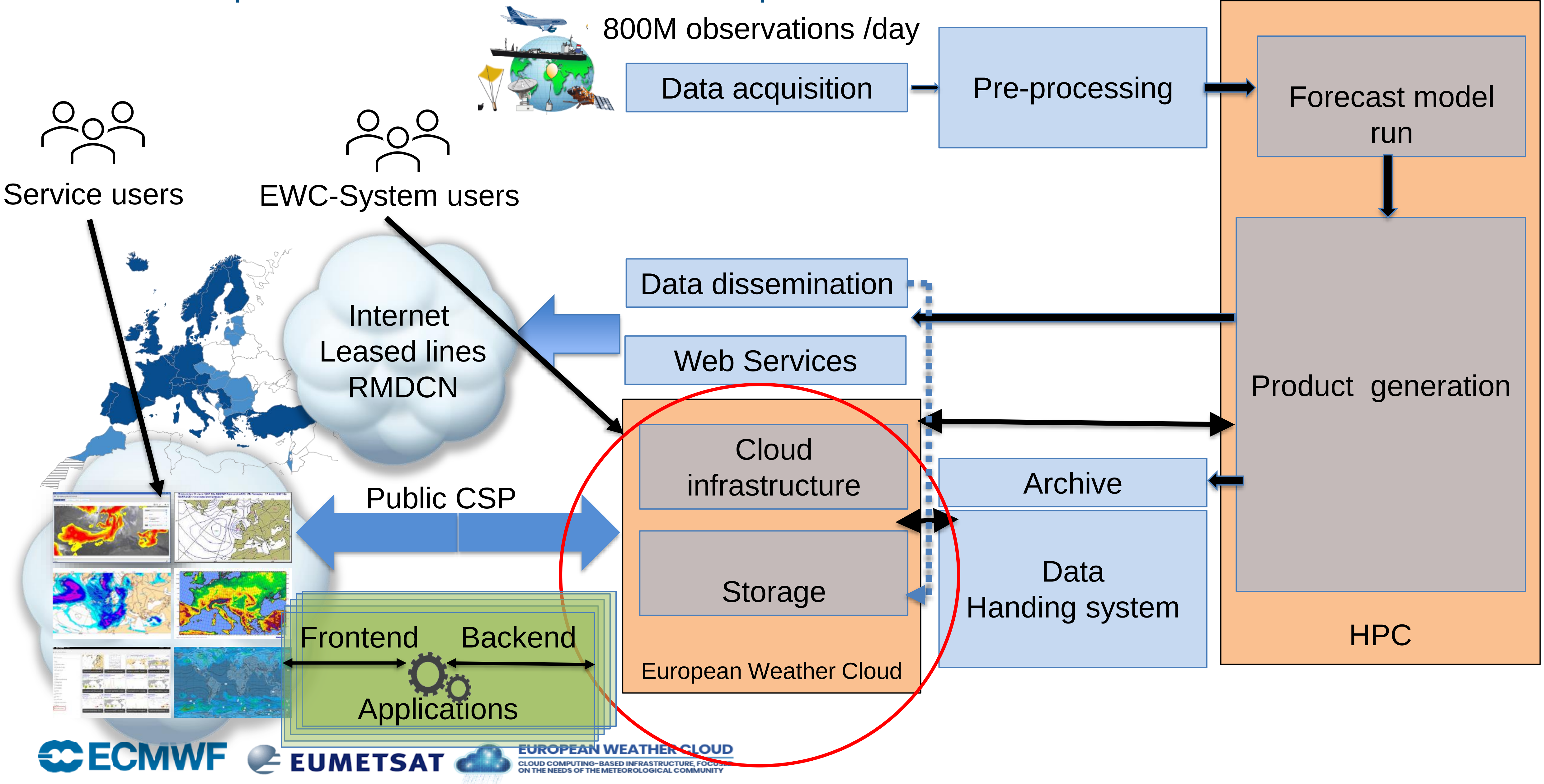
- Q1/2019- Pilot phase started
- Pilot infrastructures
  - Deployed 2019
  - User journeys validation - February 2020.
  - ECMWF’s Cloud and storage cluster’s validation.
- User on-boarding started - February 2020
- Q2/2020-Knowledge Base and Service Portal
- Q2/2020-Joint landing page and support portal in place
- 2022 Operational phase



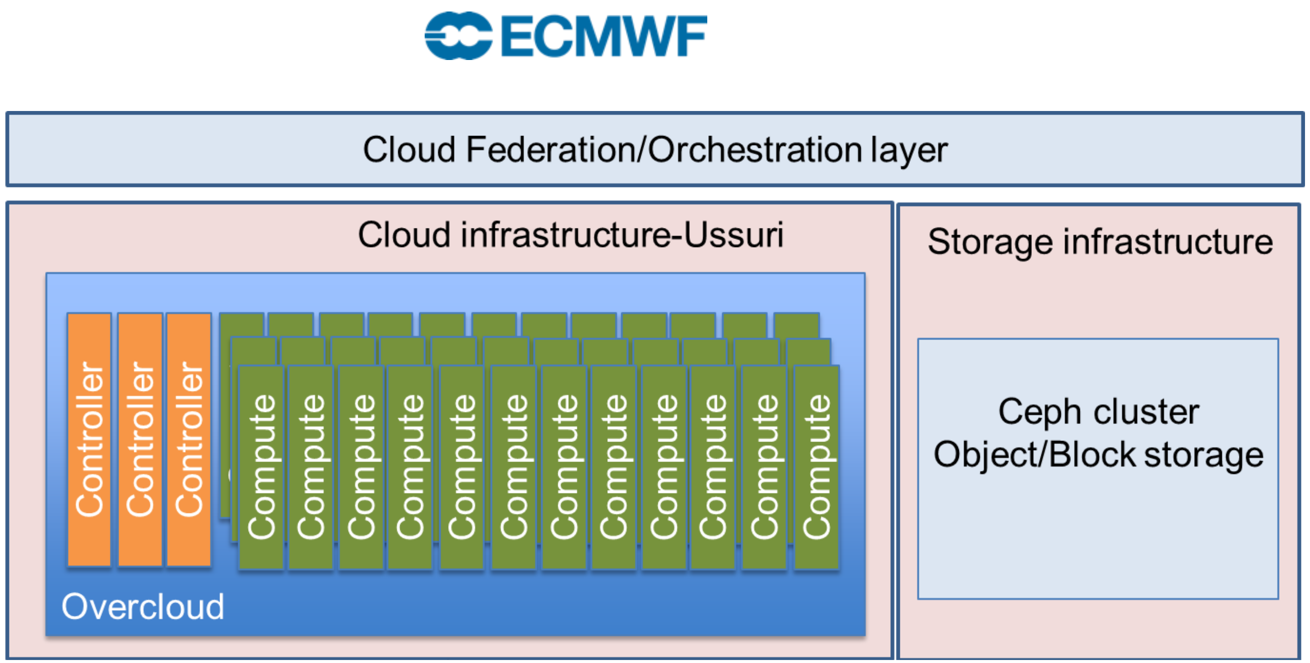


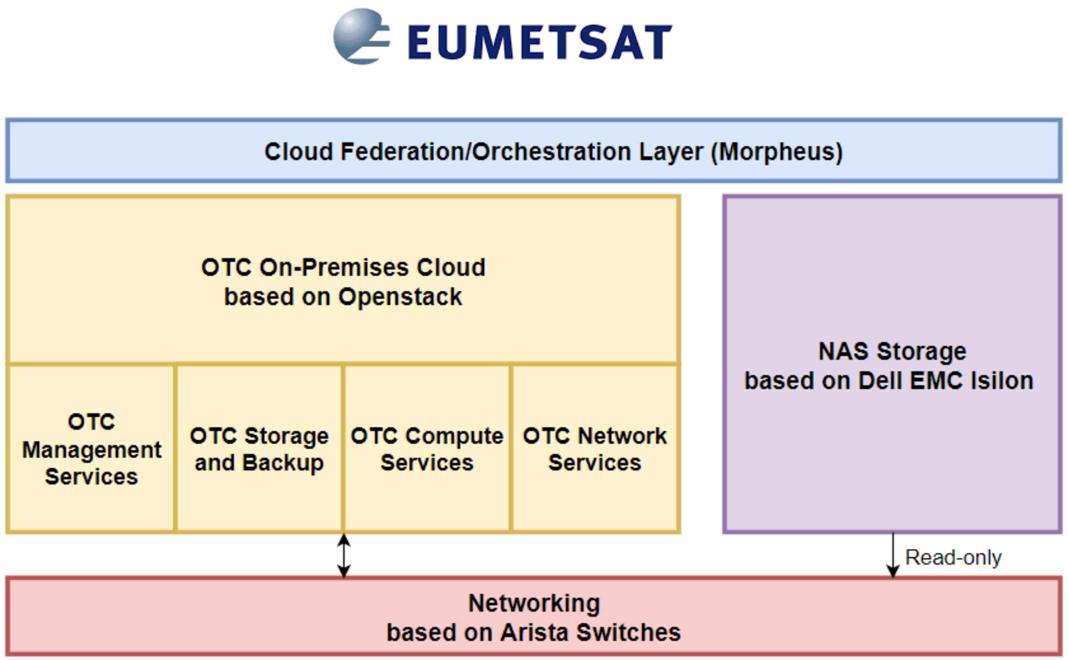
ECMWF EUMETSAT EUROPEAN WEATHER CLOUD
CLOUD COMPUTING-BASED INFRASTRUCTURE, FOCUSED ON THE NEEDS OF THE METEOROLOGICAL COMMUNITY

ECMWF's production workflow and European Weather Cloud

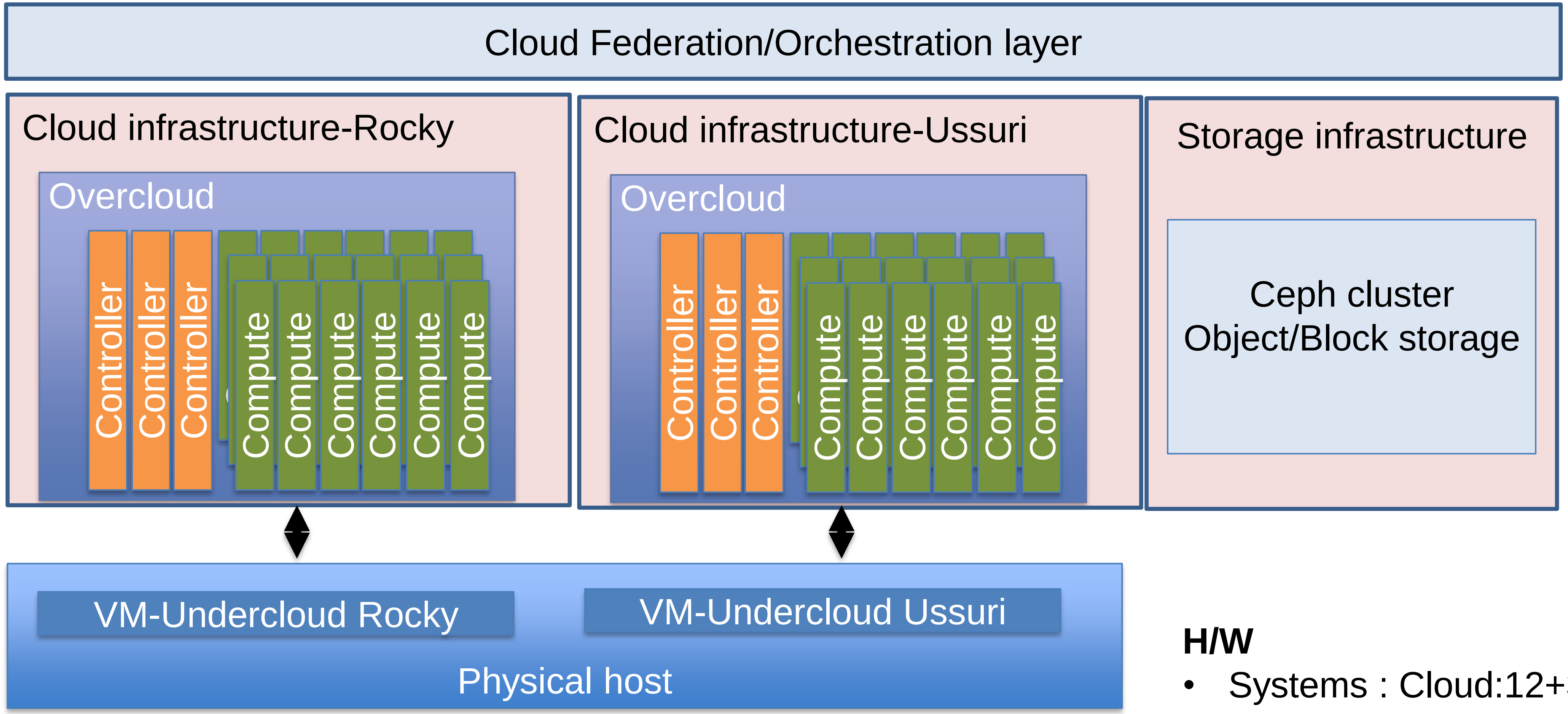


European Weather Cloud Pilot infrastructure





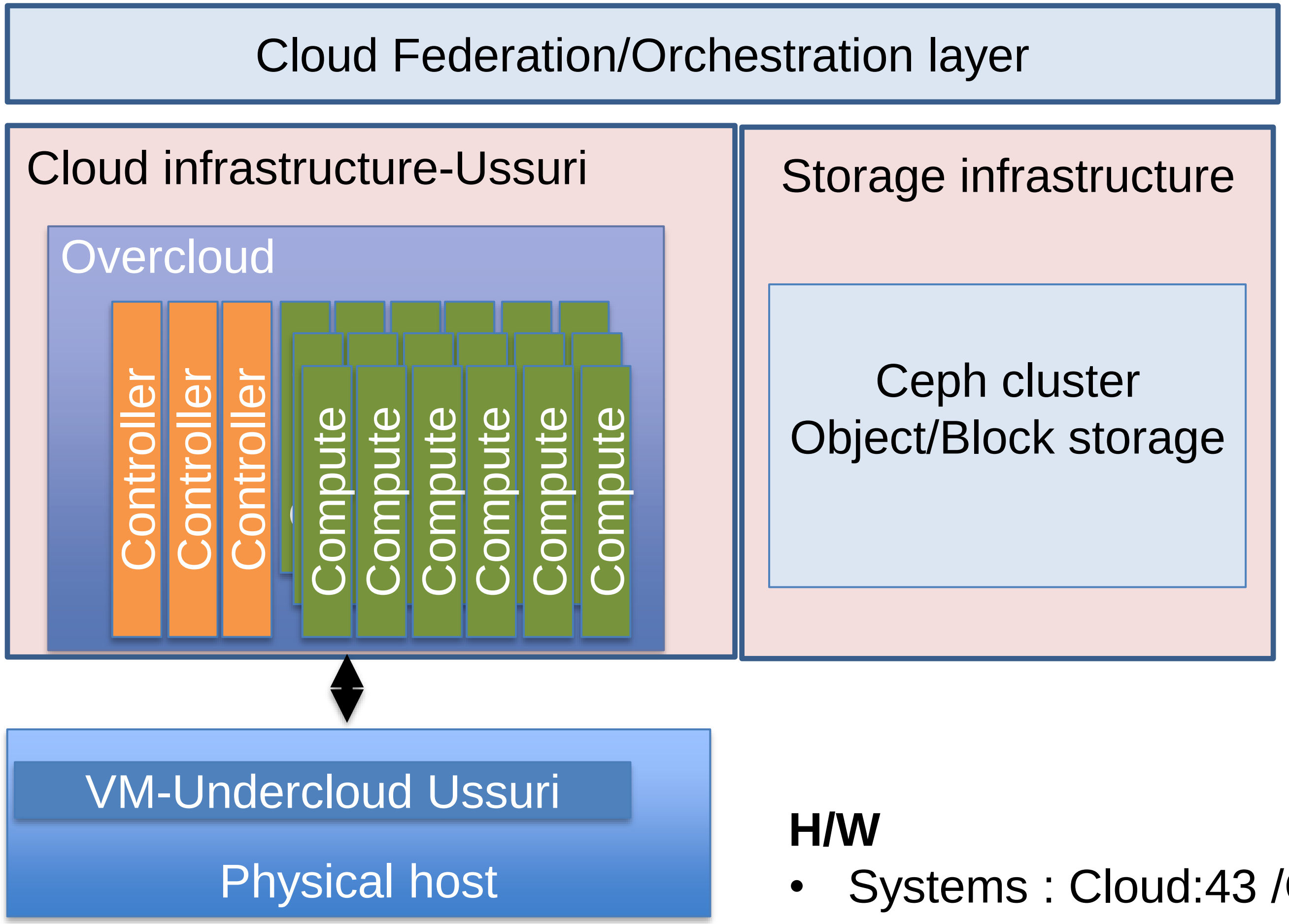
European Weather Cloud pilot infrastructure deployment @ ECMWF



H/W

- Systems : Cloud:12+31=43 /Ceph: 23
- Cores : ~3000
- RAM : ~21TB
- Storage: ~1PB (HDD+SSD)
- GPUs : 2x5 NVIDIA Tesla V100

European Weather Cloud **final** pilot infrastructure @ ECMWF



H/W

- Systems : Cloud:43 /Ceph: 23
- Cores : ~3000
- RAM : ~21TB
- Storage: ~1PB (HDD+SSD)
- GPUs : 2x5 NVIDIA Tesla V100

ECMWF EUMETSAT EUROPEAN WEATHER CLOUD
CLOUD COMPUTING-BASED INFRASTRUCTURE FOCUSED ON THE NEEDS OF THE METEOROLOGICAL COMMUNITY

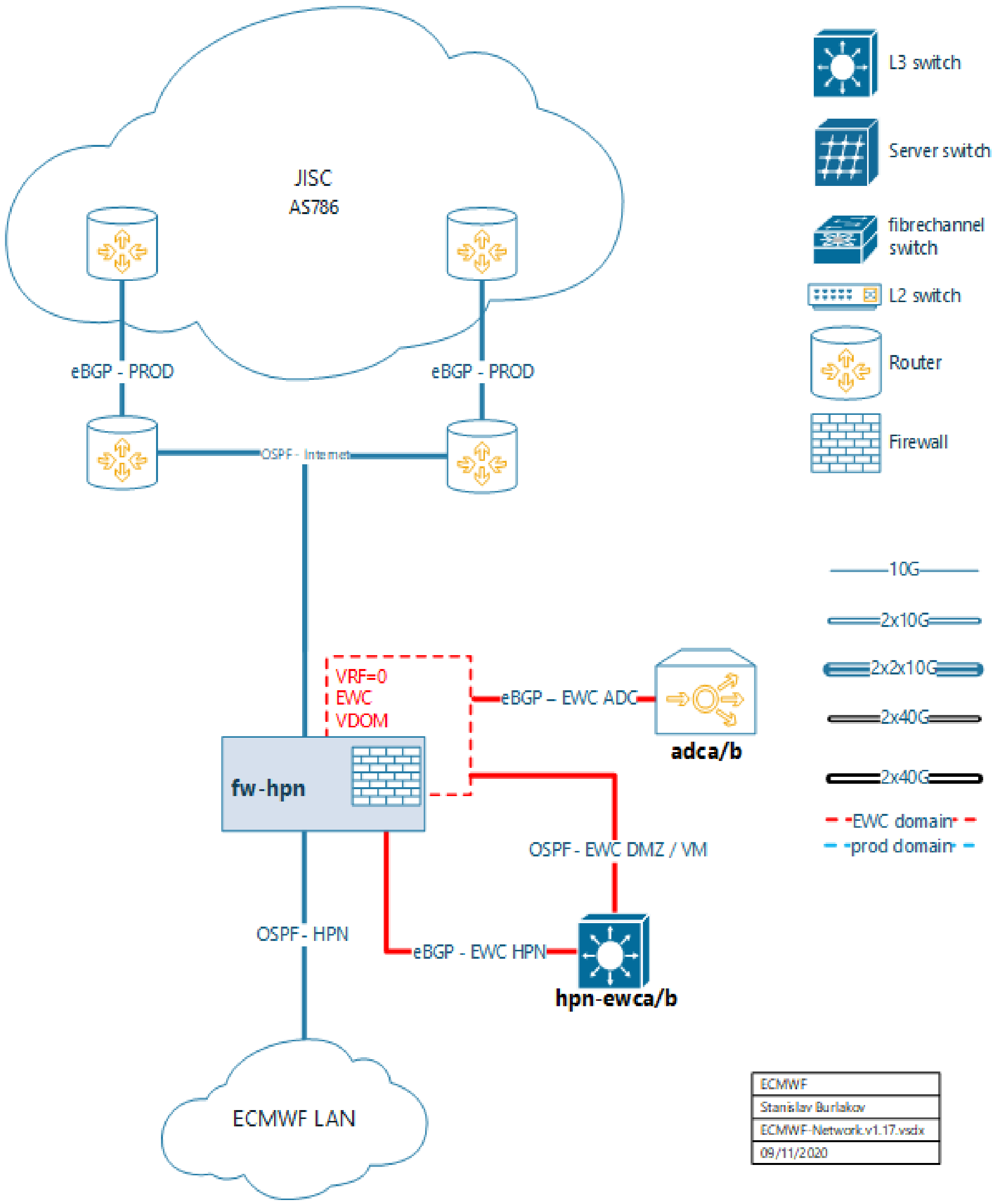
European Weather Cloud pilot infrastructure @ECMWF

Openstack:
- Systems : 43
- Network : 2x25G NICs
- vCPUs : 43 *72 ~3000
- RAM : 21TB

Ceph:
- Systems : 18 with HDDs +5 SSDs
- Networking : 2x25G NIC Public network , 2x25G NIC cluster network
- RAM : 192G
- Storage : 2 RAID1 SSD, 24 RAID0/1.8Gb/➔Raw 0.9936PB (HDD+SSD)

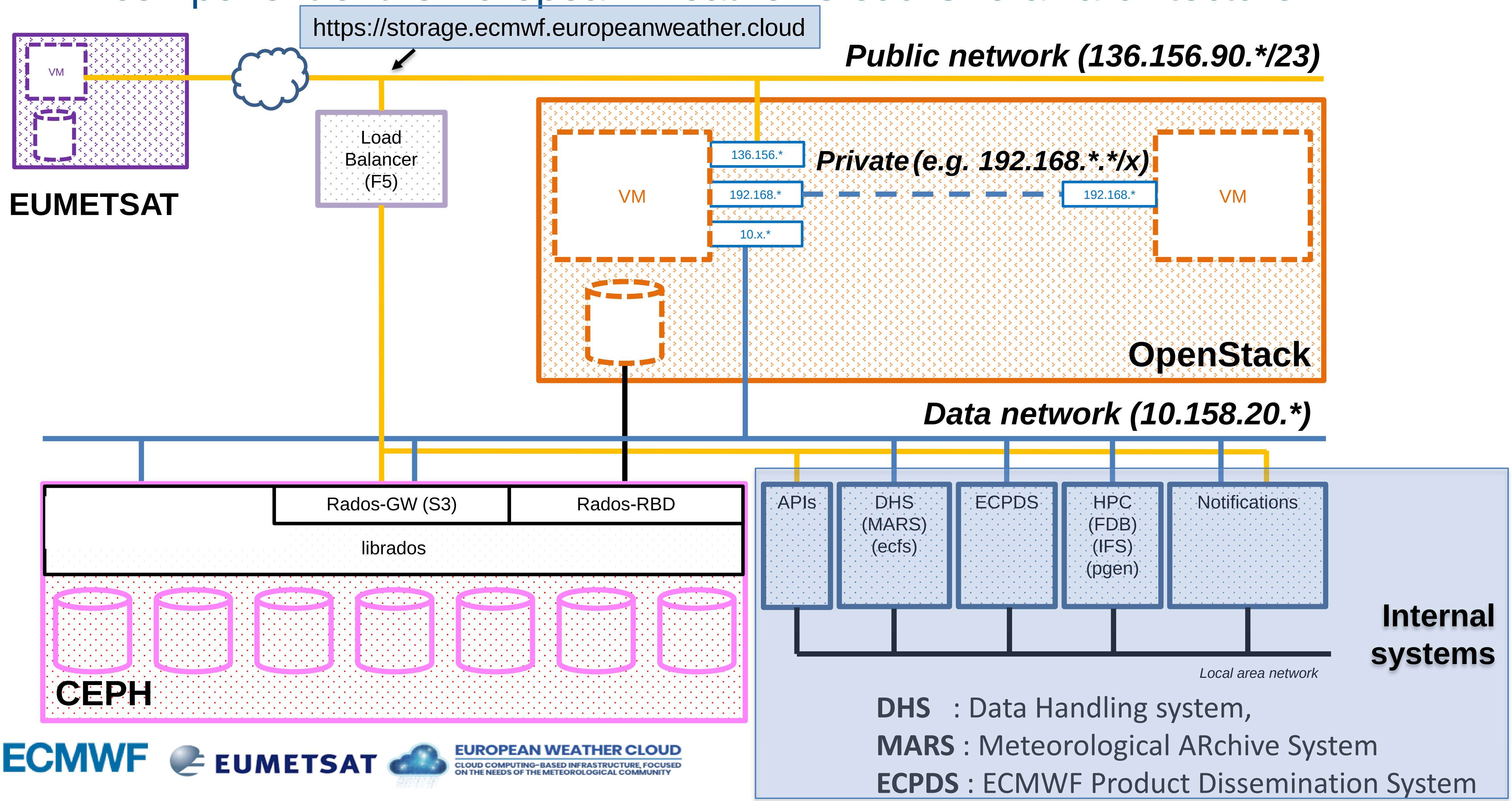
Networking
- L2 Network (spine-leaf with MLAG). Systems are connected to two ToR Arista switches for HA (balance-slb, layer2+3 hashing)
- Internet access through firewall (separate VO) for only specific ports (ssh, http(s), 6443)
- Access to the Openstack GUI and API (only to EUMETSAT) through our Load balancer (separate partition).
- Public bandwidth 20Gbps shared with all ECWMF systems

Cloud orchestrator
- morpheus.ecmwf.int VM running at our DMZ connecting users to the Openstack and Ceph.
- morpheus-dev.ecmwf.int ➔development



ECMWF EUMETSAT EUROPEAN WEATHER CLOUD
CLOUD COMPUTING-BASED INFRASTRUCTURE FOCUSED ON THE NEEDS OF THE METEOROLOGICAL COMMUNITY

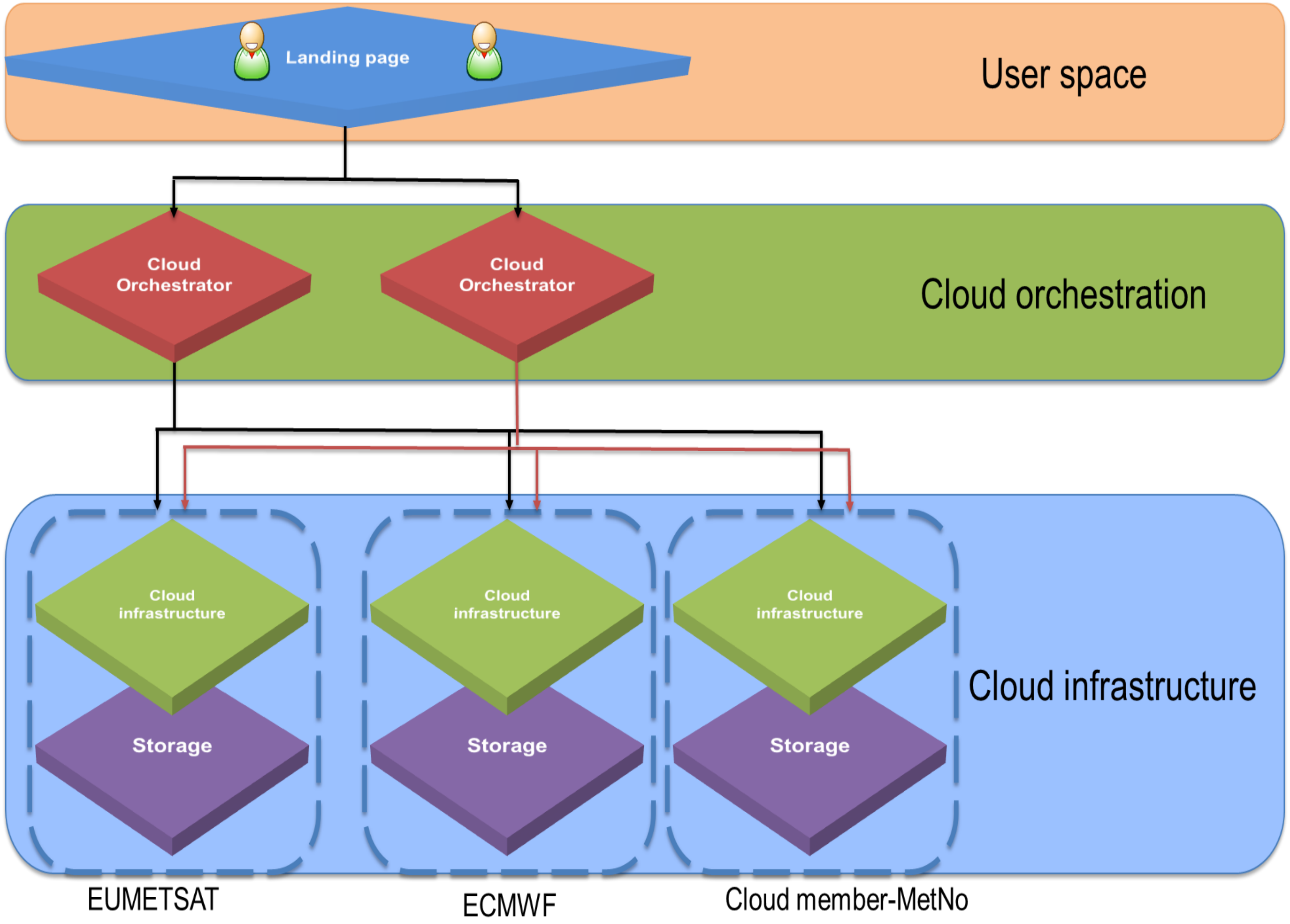
ECMWF component of the European Weather Cloud Overall architecture



ECMWF EUMETSAT EUROPEAN WEATHER CLOUD
CLOUD COMPUTING-BASED INFRASTRUCTURE, FOCUSED ON THE NEEDS OF THE METEOROLOGICAL COMMUNITY

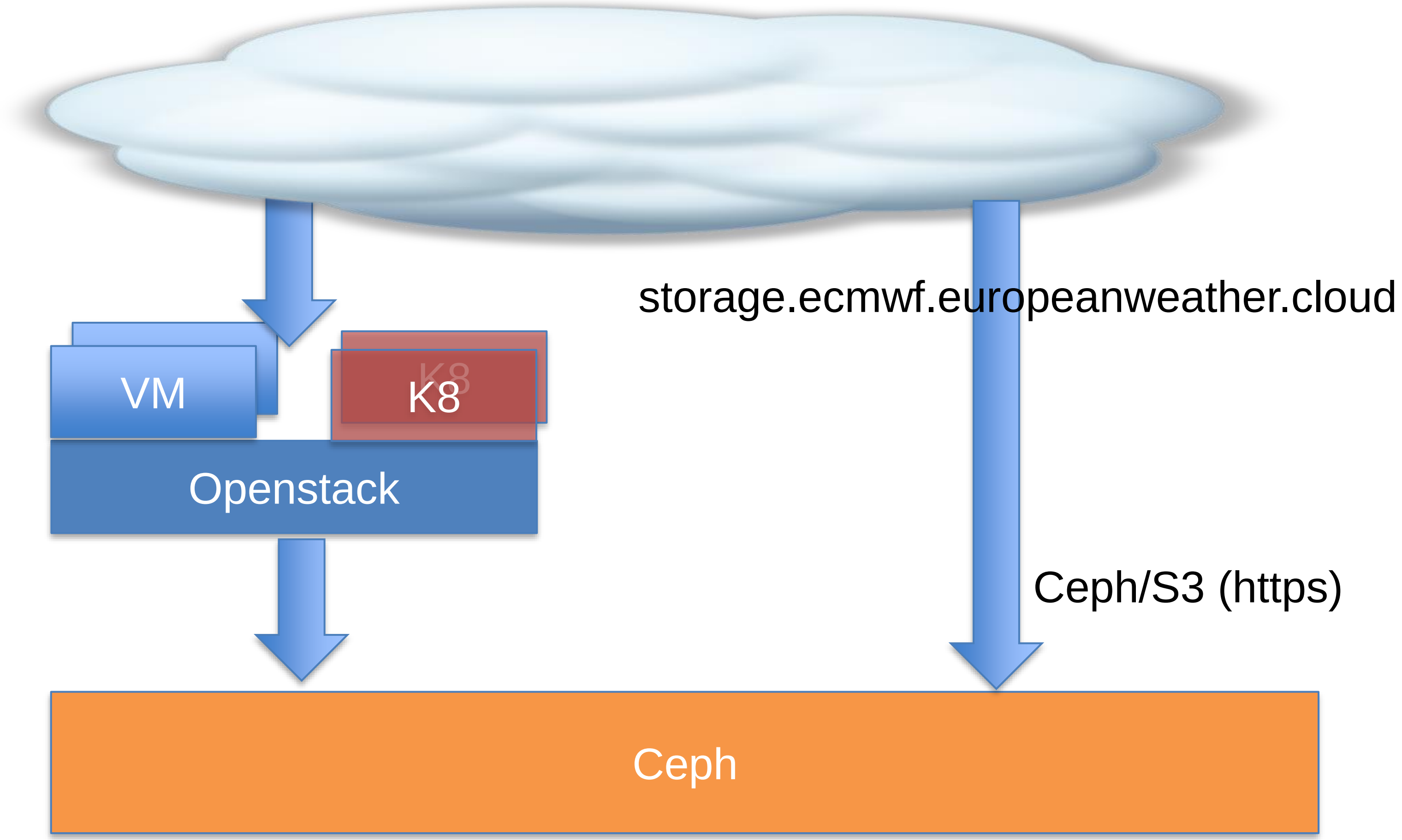
Exploring cloud federation with Member States

- Design of the European Weather Cloud allows for other partners to join
- Members of the federation can offer their own infrastructure which can be accessed from other federated systems
- Properly authorised European Weather Cloud users can then place computing tasks on the infrastructure
    - Per-user explicit consent and technical setup under partner control
- Technical tests and trials underway
- Setting up an organisational / legal framework via ECMWF/EUMETSAT formal processes

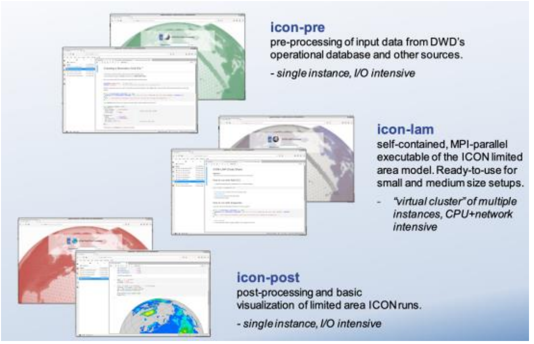


Data offering through EWC

- Storage service through storage.<ecmwf|eumetsat>.europeanweather.cloud
- ECWMF Ceph/S3 storage.ecmwf.europeanweather.cloud (https)

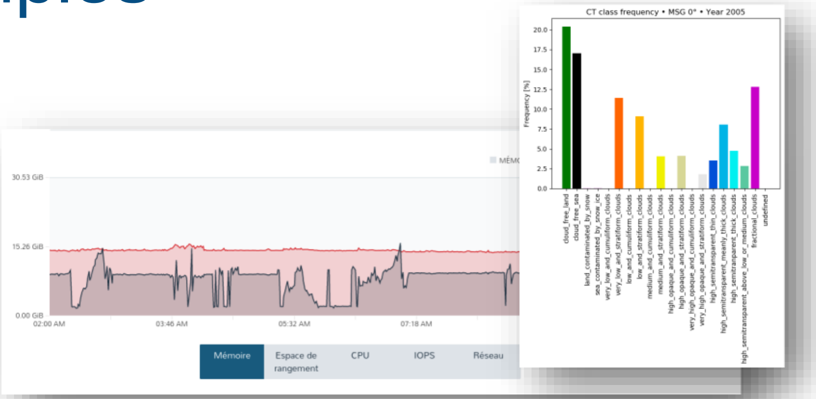


European Weather Cloud – Use case examples

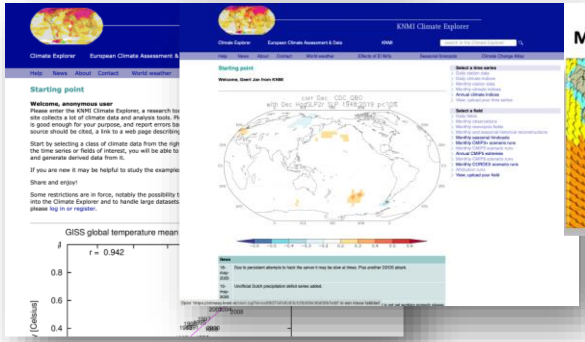


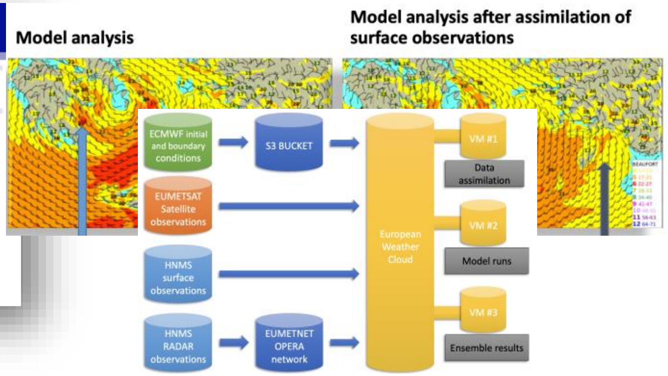
DWD/notebooks to train and develop the ICON model

Oxford University Jupyter notebook environments for ML on weather & climate data sets



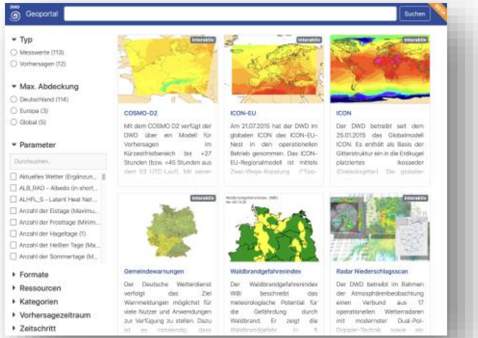
Forecast and climatology of cloud cover for Energy and Spatial sectors Météo-France
Hosted on both ECMWF and EUMETSAT



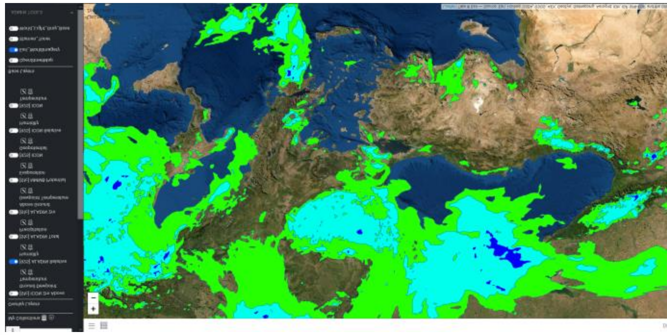


KNMI Climate Explorer running on the European Weather Cloud

HNMS uses ECMWF forecast as boundary condition for model and assimilation trials



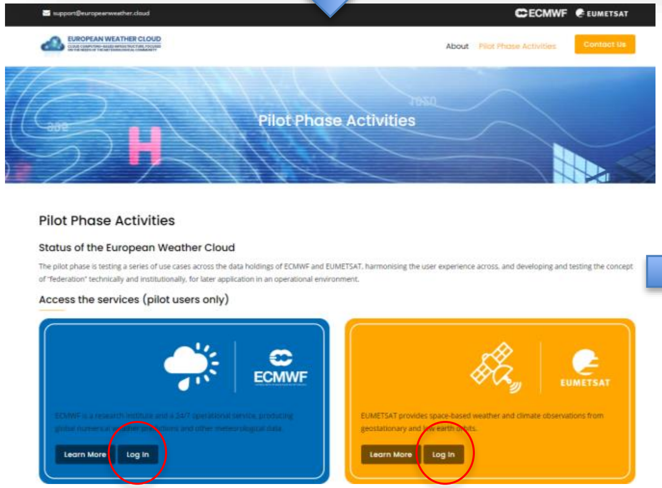
OGC web map services integrating maps in DWD's Geoportal

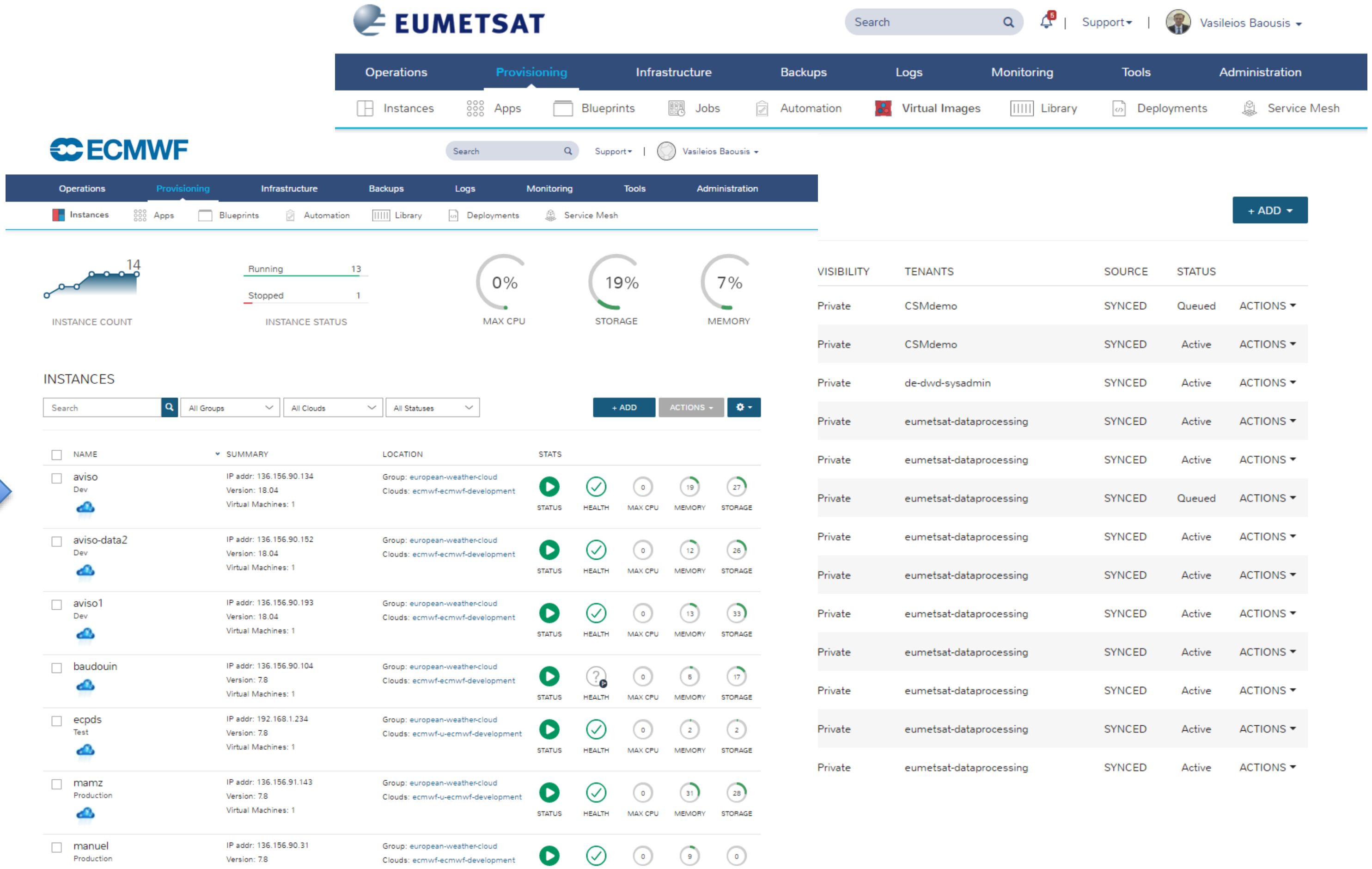


South-East European Multi-Hazard Early Warning Advisory System Common Interface Platform

ECMWF EUMETSAT EUROPEAN WEATHER CLOUD
CLOUD COMPUTING-BASED INFRASTRUCTURE FOCUSED ON THE NEEDS OF THE METEOROLOGICAL COMMUNITY

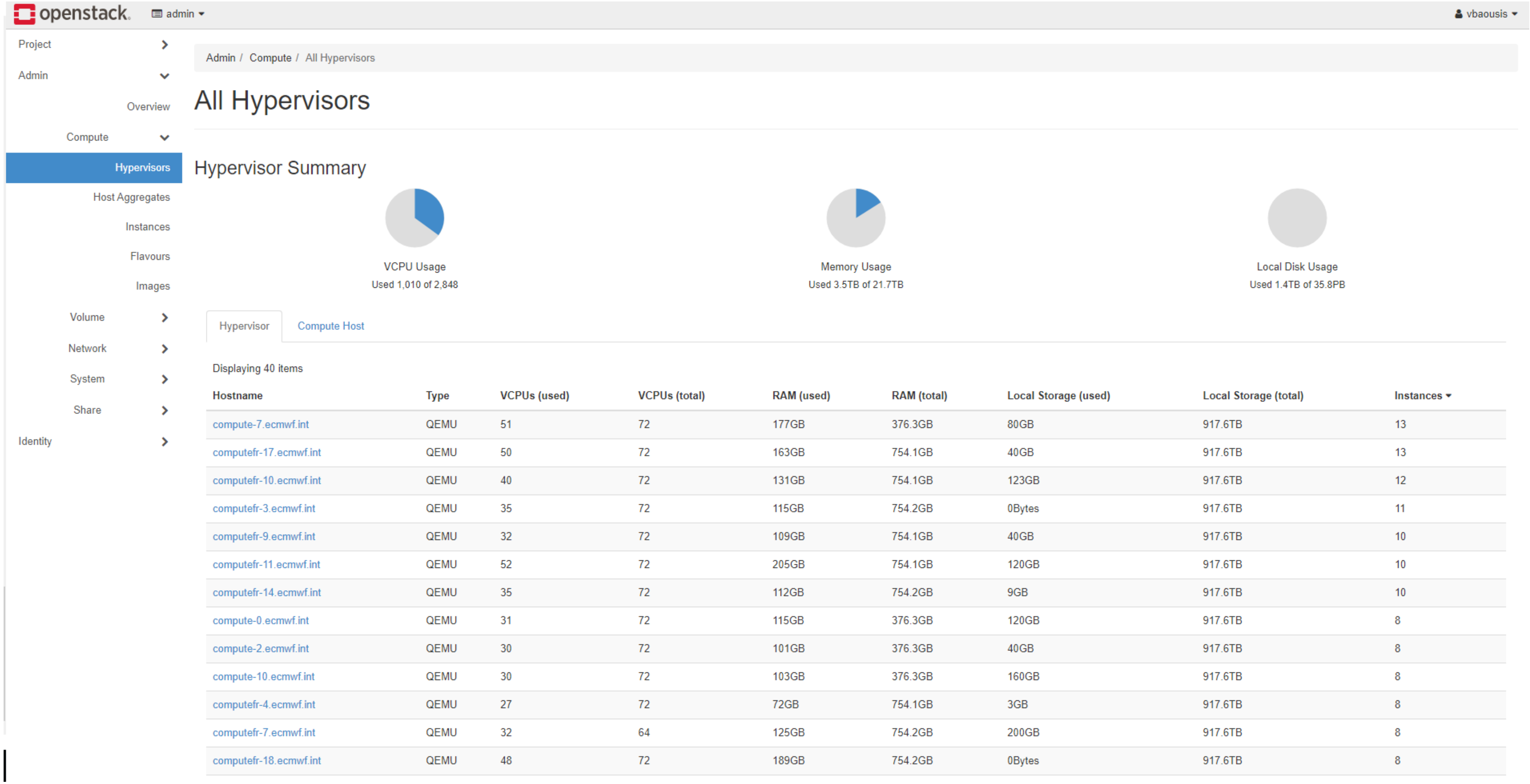
Cloud Management Orchestrator -Morpheus





ECMWF EUMETSAT **EUROPEAN WEATHER CLOUD**
CLOUD COMPUTING-BASED INFRASTRUCTURE, FOCUSED ON THE NEEDS OF THE METEOROLOGICAL COMMUNITY

ECMWF European Weather Cloud-Openstack



Hostname	Type	VCPUs (used)	VCPUs (total)	RAM (used)	RAM (total)	Local Storage (used)	Local Storage (total)	Instances
compute-7.ecmwf.int	QEMU	51	72	177GB	376.3GB	80GB	917.6TB	13
computefr-17.ecmwf.int	QEMU	50	72	163GB	754.1GB	40GB	917.6TB	13
computefr-10.ecmwf.int	QEMU	40	72	131GB	754.1GB	123GB	917.6TB	12
computefr-3.ecmwf.int	QEMU	35	72	115GB	754.2GB	0Bytes	917.6TB	11
computefr-9.ecmwf.int	QEMU	32	72	109GB	754.1GB	40GB	917.6TB	10
computefr-11.ecmwf.int	QEMU	52	72	205GB	754.1GB	120GB	917.6TB	10
computefr-14.ecmwf.int	QEMU	35	72	112GB	754.2GB	9GB	917.6TB	10
compute-0.ecmwf.int	QEMU	31	72	115GB	376.3GB	120GB	917.6TB	8
compute-2.ecmwf.int	QEMU	30	72	101GB	376.3GB	40GB	917.6TB	8
compute-10.ecmwf.int	QEMU	30	72	103GB	376.3GB	160GB	917.6TB	8
computefr-4.ecmwf.int	QEMU	27	72	72GB	754.1GB	3GB	917.6TB	8
computefr-7.ecmwf.int	QEMU	32	64	125GB	754.2GB	200GB	917.6TB	8
computefr-18.ecmwf.int	QEMU	48	72	189GB	754.2GB	0Bytes	917.6TB	8

European Weather Cloud infrastructure - OpenStack

- First cluster deployment September 2019 –Rocky Openstack release
- A development openstack/ceph cluster similarly configured
- Configuration & Experience
  - ~1600VCPUs and 11TB RAM without any significant problem.
  - External Ceph cluster integration worked was straightforward ceph-config.yaml →
  - Two external networks (public and private for fast access to our data archive-MARS)
  - Most of our VMs are attached to both external networks (no FIPS) and/or private tenant network=> Challenging VM routing without dynamic routing on the switches=> Workaround with dhcp hooks and configuring VM images routing
  - Some problems with the NIC bond interface configuration with our switches : LACP configuration?=> Single NIC deployment
  - Problems with LBaaS (https://bugs.launchpad.net/tripleo/+bug/1832866)
    - – Octavia problems with certificates overridden on each deployment

– **Live Updating**
  - We moved for a Single NIC to a Multiple NIC deployment
  - The whole first cluster was redeployed, and the network was re-configured with DVR(Distributed Virtual Routing) configuration.

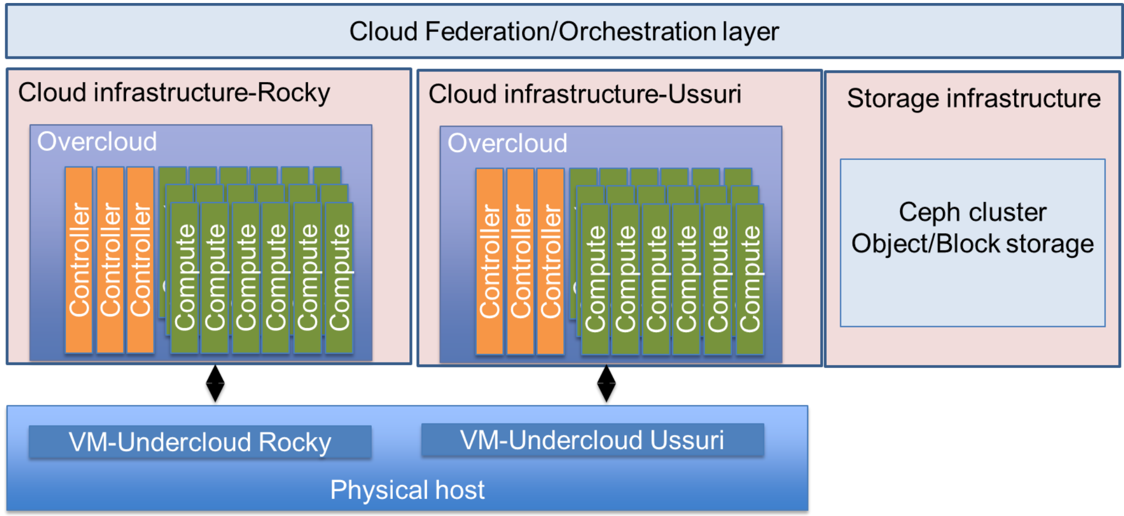
– **Good performance overall.**
  - Verified by a 3rd party- February 2020


```
parameter_defaults:
  DockerCephDaemonImage: ceph/daemon:tag-stable-3.0-jewel-centos-7
  CephClusterFSID: 'FSID'
  CephClientKey: 'ClientKey'
  CephExternalMonHost: 'IPS'
 NovaEnableRbdBackend: true
  CinderEnableRbdBackend: true
  CinderBackupBackend: ceph
  GlanceBackend: rbd
  GnocchiBackend: rbd
  NovaRbdPoolName: vms
  CinderRbdPoolName: volumes
  CinderBackupRbdPoolName: backups
  GlanceRbdPoolName: images
  GnocchiRbdPoolName: metrics
  CephClientUserName: openstack
  CinderEnableIscsiBackend: false
  NeutronNetworkType: vxlan
```

European Weather Cloud OpenStack infrastructure

- More hardware added to OpenStack & Ceph clusters.
- Second cluster first build 30/5/2020 (Ussuri)

- Problems-changes-challenges
  - New build method based on Ansible rather than Mistral
    - Some hiccups (the user used to deploy the stack : stack / heat-admin
  - CentOS8 base OS both for the host systems and service containers.
  - We continued with OVS and not OVN (due to FIP problems).
  - Bugs reported, resolved, help received from the community
    - Octavia https://bugs.launchpad.net/tripleo/+bug/1881420
    - Ceph https://bugs.launchpad.net/tripleo/+bug/1880579
    - OVN DVR and FIP https://bugs.launchpad.net/networking-ovn/+bug/1828891
- Problems resolved June 2020.

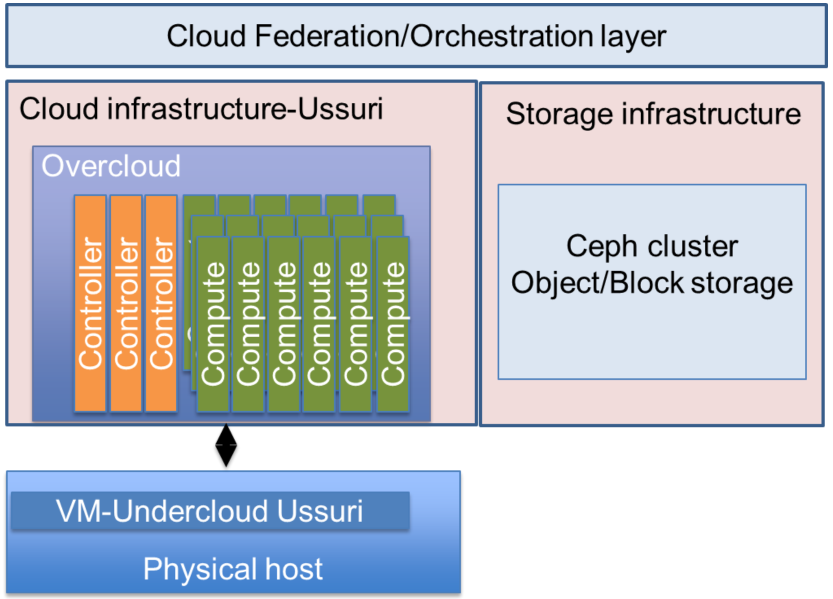


Building an OpenStack cloud explained: yaml file building the entire cloud

Most of the yaml files have a few modifications to match our network/ceph etc environments

```
openstack overcloud  deploy --templates   \
  -e /home/stack/containers-prepare-parameter.yaml \
  -e /home/stack/templates/node-info.yaml \
  -r /home/stack/templates/roles_data.yaml \
  -n /home/stack/templates/network_data.yaml \
  -e /home/stack/templates/environments/network-environment-OVS.yaml \
  -e /home/stack/templates/environments/network-isolation.yaml \
  -e /home/stack/templates/environments/ceph-ansible/ceph-ansible-external.yaml \
  -e /home/stack/templates/ceph-config.yaml \
  -e /home/stack/templates/environments/docker-ha.yaml \
  -e /home/stack/templates/environments/ssl/enable-tls.yaml \
  -e /home/stack/templates/environments/ssl/inject-trust-anchor-hiera.yaml \
  -e /home/stack/templates/environments/ssl/inject-trust-anchor.yaml \
  -e /home/stack/templates/environments/ssl/tls-endpoints-public-dns.yaml \
  -e /home/stack/templates/environments/predictable-placement/custom-domain.yaml  \
  -e /home/stack/templates/cloudname.yaml \
  -e /home/stack/templates/environments/manila-cephfsnative-config.yaml \
  -e /home/stack/templates/environments/ceph-ansible/ceph-mds.yaml \
  -e /home/stack/templates/manila-cephfsnative-config.yaml \
  -e /home/stack/templates/environments/enable-legacy-telemetry.yaml \
  -e /home/stack/templates/environments/neutron-ovs-dvr.yaml \
  -e /home/stack/templates/environments/services/octavia.yaml \
  -e /home/stack/templates/overcloud_dashboard_hardening.yaml \
  -e /home/stack/templates/novafixes.yaml \
  --timeout 1500
```

European Weather Cloud OpenStack infrastructure



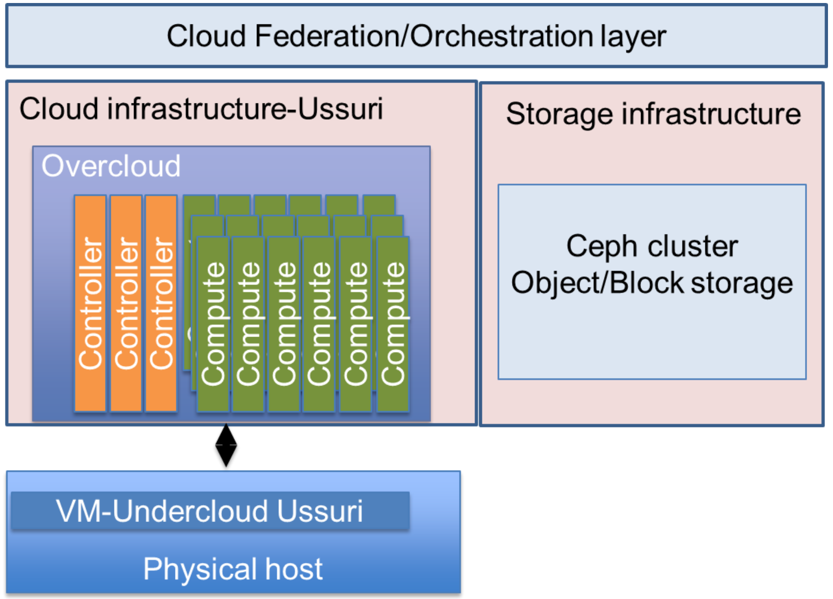
Configuration of Nvidia GPUs.
Problems

- Since we haven't implemented IPv6 to our Ussuri cluster when we installed and configured the GPUs drivers to a node, OVS was complaining
  - During boot time, OVS was trying to bind to IPv6 addresses=> resulting to considerable increased boot time.
  - A workaround was to explicitly remove IPv6 configuration to our all GPU nodes.
  - All installed as normal compute nodes and configured nova.conf with ansible playbooks.
    - Much easier to reconfigure the GPU profiles offered
  - GPU profiles assignment to VMs based on flavors

Hyp/sor	Model	Nova profile	VM/GPU /Model	VMs /host	Frame BufferSize	Max. Display Resolution
gpu01	GRID V100-16C	nvidia-301	1	2	16384	4096×21602
gpu02	GRID V100-8C	nvidia-300	2	4	8192	4096×21602
gpu03	GRID V100-8C	nvidia-300	2	4	8192	4096×21602
gpu04	GRID V100-4C	nvidia-299	4	8	4096	4096×21602
gpu05	GRID V100-4C	nvidia-299	4	8	4096	4096×21602

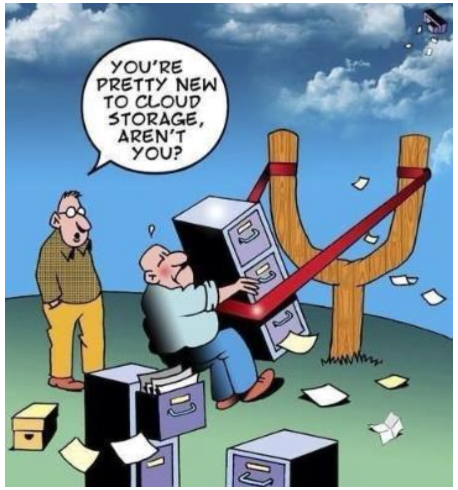
European Weather Cloud infrastructure – Ceph Storage Cluster

- Ceph is built and maintained separately to OpenStack.
- Based on CentOS7/Ceph Nautilus (14.2.22)
- Hardware :
  - Systems: 23XDell PowerEdge R740xd
  - Storage : 2 system SSDs(RAID1)+24 HDD or SSD (1.7TB) with two RAID controllers
  - Networking : 2x25Gbps cluster and 2x25Gbps public NICs
  - RAM : 192GB
  - Deployed with **ceph-deploy** and maintained with puppet
  - **mon** =22, **mgr**:5, **rgw**=3 (load balanced), **osd**=552

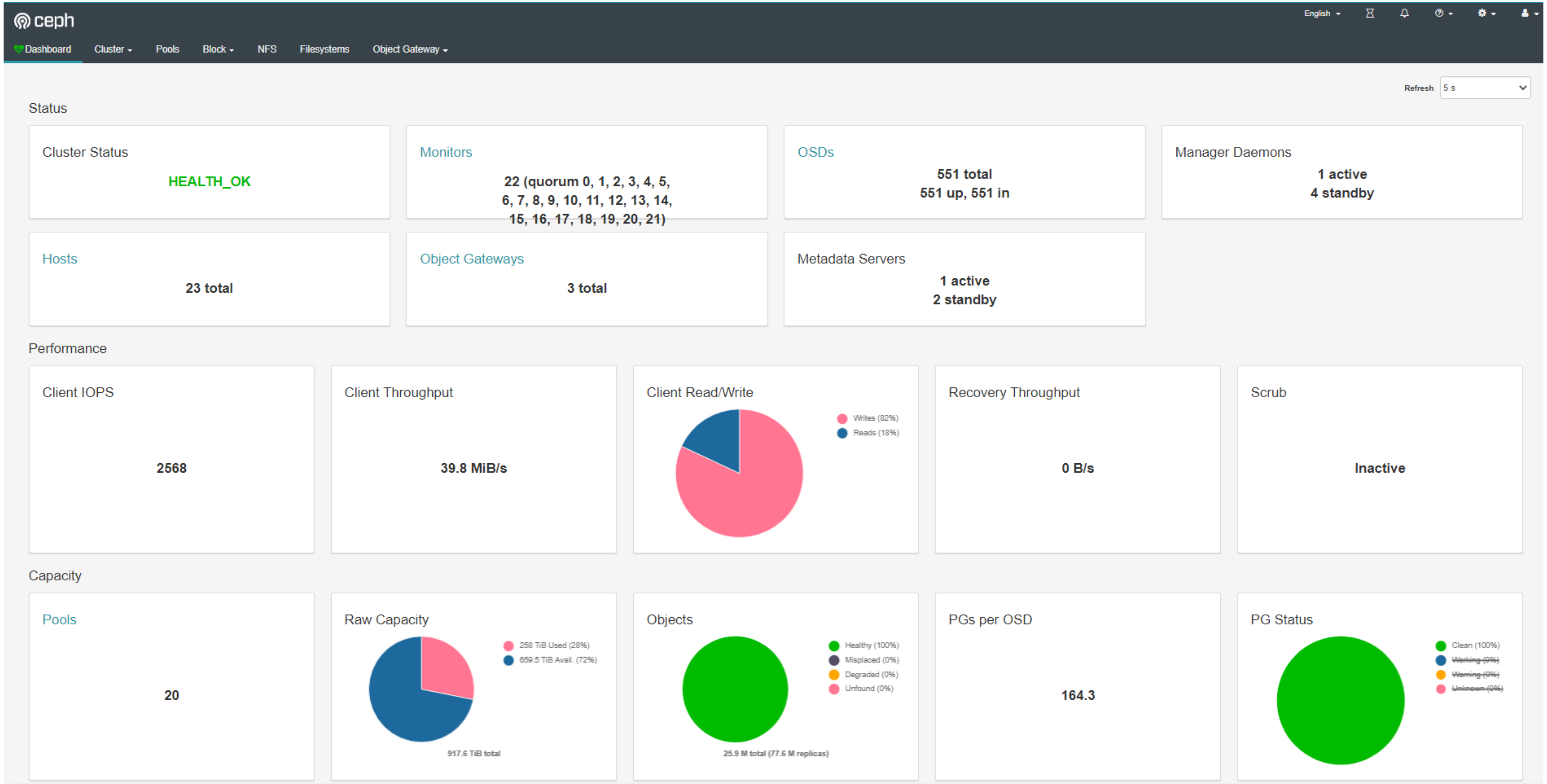


- Numa configuration based on the recommendations using numad service (link)

- First build was ~2 year ago, expanded to its current capacity 1 year ago.
- 3rd party validation with some minor suggestions for improvements
- Both our Openstack clusters use the same Ceph infrastructure (and the same rbd pools)
- Besides some usual HDD failures, Ceph performs well.



ECMWF component of the European Weather Cloud- Ceph storage cluster



ECMWF EUMETSAT EUROPEAN WEATHER CLOUD
CLOUD COMPUTING-BASED INFRASTRUCTURE, FOCUSED ON THE NEEDS OF THE METEOROLOGICAL COMMUNITY

ECMWF component of the European Weather Cloud- Ceph storage cluster

ceph

English | Dashboard | Cluster | Pools | Block | NFS | Filesystems | Object Gateway

Pools

Pools List | Overall Performance

Create | 20

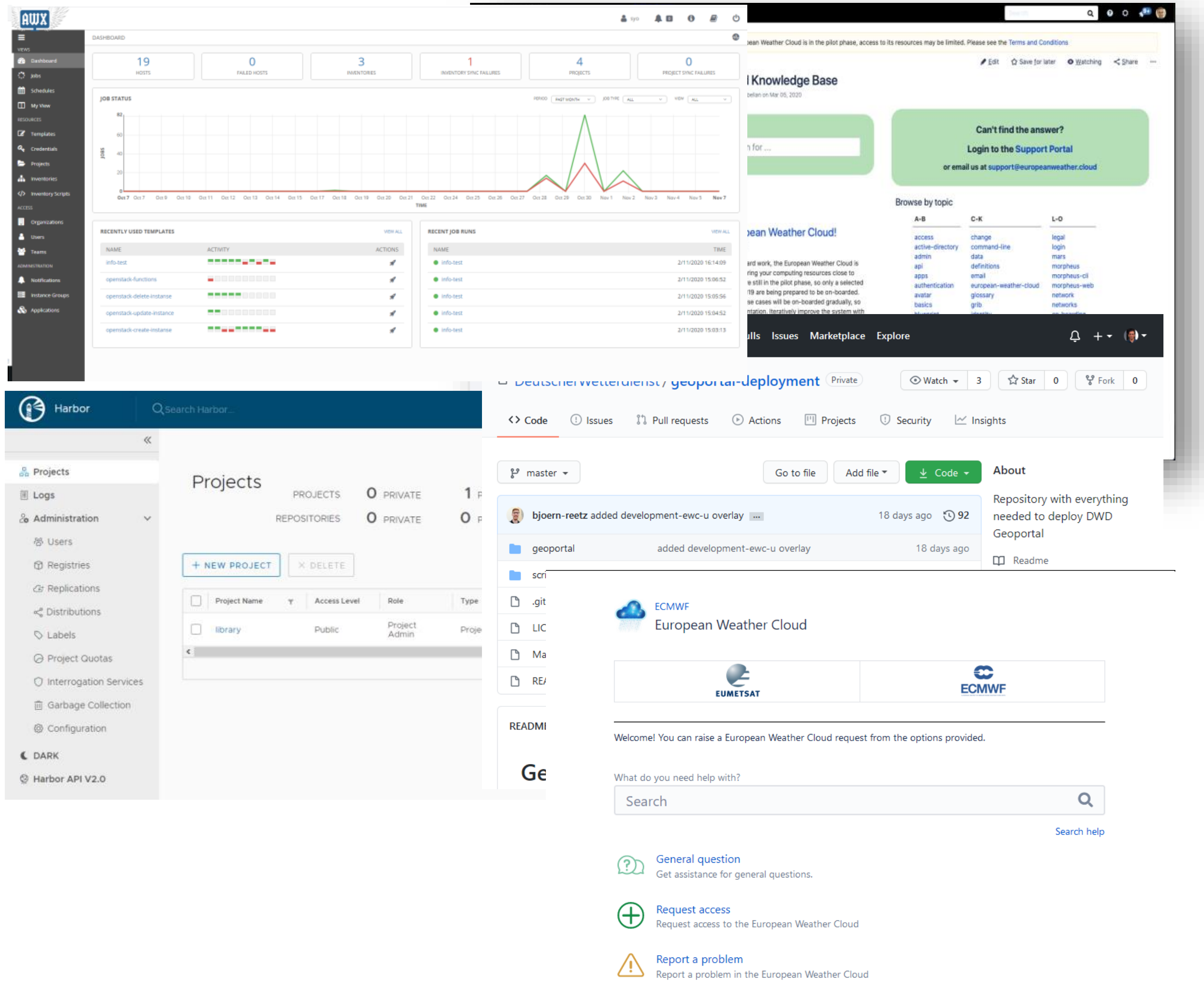
Name	Type	Applications	PG Status	Replica Size	Last Change	Erasure Coded Profile	Crush Ruleset	Usage	Read bytes	Write bytes	Read ops	Write ops
.rgw.root	replicated	rgw	8 active+clean	3	1386		replicated_rule	0%			0 /s	0 /s
backups	replicated	rbd	2048 active+clean	3	1823		replicated_rule	0%			0 /s	0 /s
chkopool	replicated	rbd	64 active+clean	3	35214		replicated_rule	0%			0 /s	0 /s
default.rgw.buckets.data	replicated	rgw	1 active+clean+scrubbing, 2047 active+clean	3	33333		replicated_rule	4.58%			0 /s	0 /s
default.rgw.buckets.index	replicated	rgw	8 active+clean	3	6042		replicated_rule	0%			0 /s	0 /s
default.rgw.buckets.non-ec	replicated	rgw	8 active+clean	3	6328		replicated_rule	0%			0 /s	0 /s
default.rgw.control	replicated	rgw	8 active+clean	3	1388		replicated_rule	0%			0 /s	0 /s
default.rgw.log	replicated	rgw	8 active+clean	3	1391		replicated_rule	0%			6.2 /s	4.1 /s
default.rgw.meta	replicated	rgw	8 active+clean	3	1391		replicated_rule	0%			0 /s	0 /s
images	replicated	rbd	1024 active+clean	3	37417		replicated_rule	0.72%			374.4 /s	0.2 /s
mars	replicated	rbd	512 active+clean	3	5615		replicated_rule	0.01%			0 /s	0 /s
mars-erasure42	erasure	rbd	1024 active+clean	6	10017	mars-erasure42	mars-erasure42	0%			0 /s	0 /s
mars-erasure52	erasure	rbd	1024 active+clean	7	13037	mars-erasure52	mars-erasure52	0%			0 /s	0 /s
metrics	replicated	rbd	1 active+clean+scrubbing, 1023 active+clean	3	1584		replicated_rule	0.02%			8.2 /s	0 /s
test-pool-ssd	replicated	rbd	512 active+clean	2	21715		replicated_ssd	0%			0 /s	0 /s
vms	replicated	rbd	1 active+clean+scrubbing+deep, 2047 active+clean	3	35135		replicated_rule	0.09%			7.2 /s	46.5 /s
volumes	replicated	rbd	1 active+clean+scrubbing+deep, 16383 active+clean	3	37313		replicated_rule	7.29%			NaN undefined /s	273.9 /s

0 selected / 17 total

ECMWF EUMETSAT EUROPEAN WEATHER CLOUD
CLOUD COMPUTING-BASED INFRASTRUCTURE, FOCUSED ON THE NEEDS OF THE METEOROLOGICAL COMMUNITY

European Weather Cloud: From IaaS to PaaS model

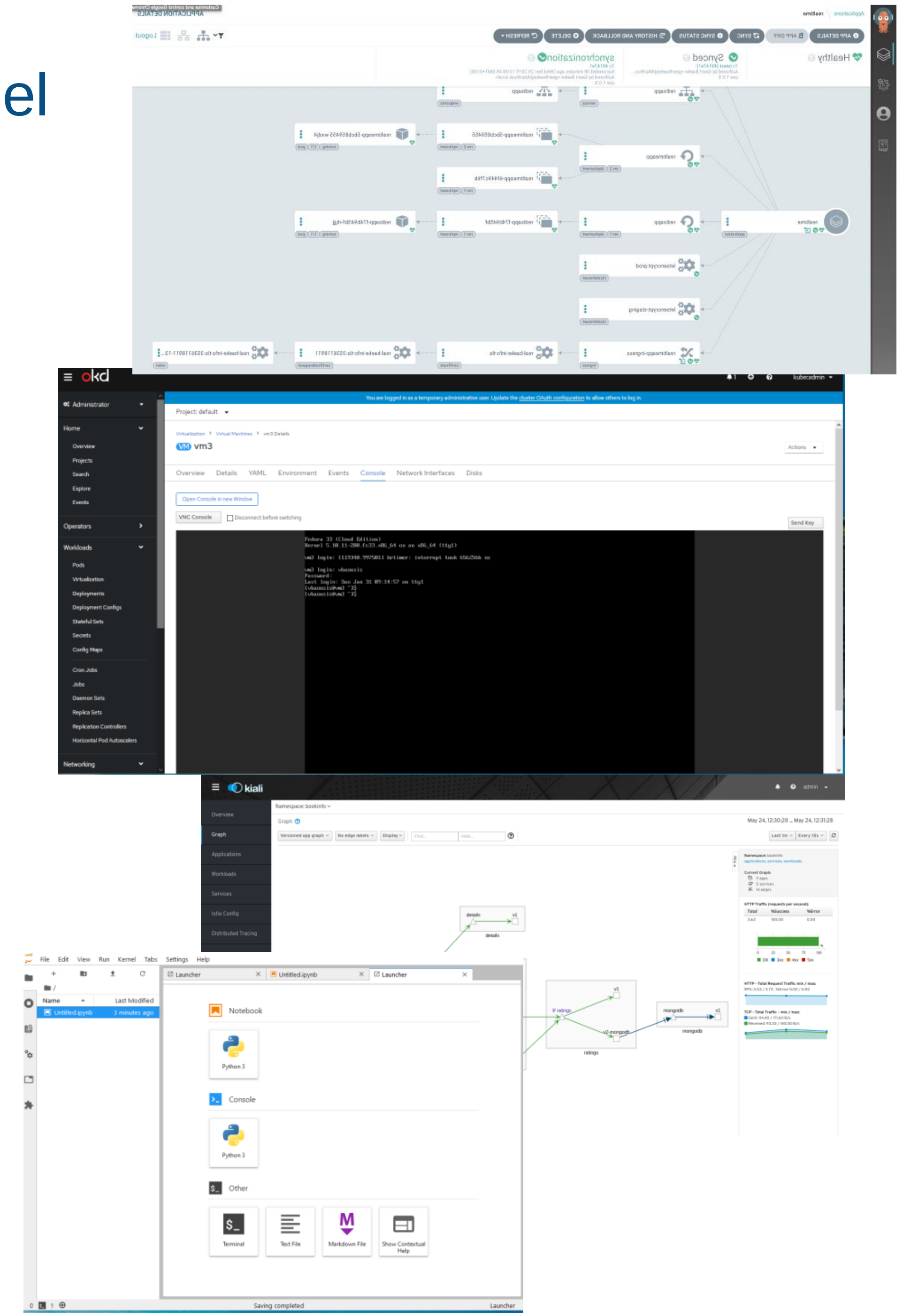
- User base diversity
- Mainly meteorological applications
  - & general-purpose platforms and applications
- Multi-tier applications
- Resource demanding workloads-Support ML/AI
- Fast access to data
- Containerization and orchestration
- Quick/easy creation/disposal of platforms (e.g. Jupyterlab supported with GPUs)
- User and application security, privacy and isolation (certificate creation etc)
- User accounting and reporting
- Data privacy



EUROPEAN WEATHER CLOUD
CLOUD COMPUTING-BASED INFRASTRUCTURE, FOCUSED ON THE NEEDS OF THE METEOROLOGICAL COMMUNITY

European Weather Cloud: From IaaS to PaaS model
Kubernetes operators

- *Operators* take human operational knowledge and encode it into software that is more easily shared with consumers
- A method of packaging, deploying, and managing a Kubernetes application. Repeatable, health checks, easily updated, encapsulate knowledge. Using Ansible, Helm
- Operator SDK
- Examples
  - Opportunistic VMs with Kubevirt within K8 : https://kubevirt.io/
  - GitOps CI/CD tool for Kubernetes https://argoproj.github.io/argo-cd/
  - Serverless with knative
  - JupyterHub
  - SSO with Keycloack
  - Apache Kafka
  - Security and dynamic SSL certificates
  - Higher storage abstraction (with Ceph/Cinder)
  - Machine learning (ML) workflows creation
  - Kubernetes cluster federation



ECMWF EUMETSAT EUROPEAN WEATHER CLOUD
CLOUD COMPUTING-BASED INFRASTRUCTURE, FOCUSED ON THE NEEDS OF THE METEOROLOGICAL COMMUNITY

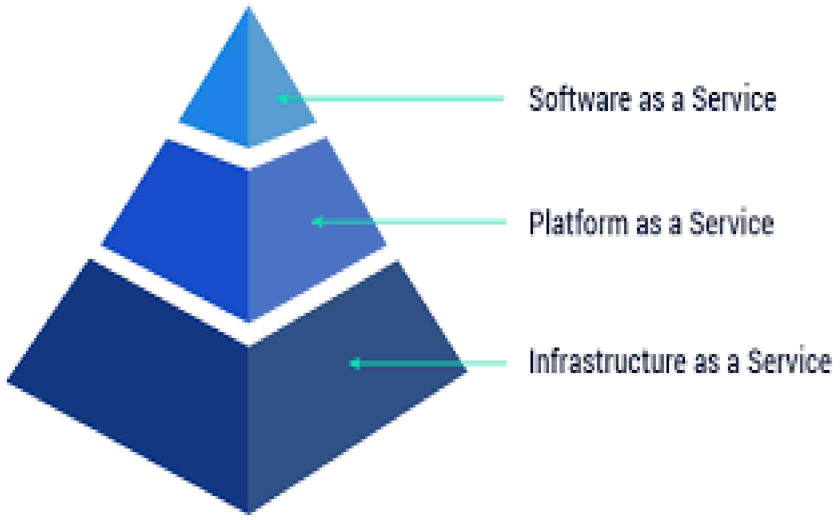
HPC and Cloud convergence (same examples)

- Fenix (BSC, CEA, CINECA, CSCS, JSC)
  - Computing Services: Interactive, Scalable and VM services
  - Data services : Active & (Federated) Archival Data Repositories, Data mover-location-transport services
  - Federation services : Authentication and Authorization Services (AAI), Fenix User and Resource management services (FURMS)
- GENCI
  - French Research Infrastructure for HPC (Access through Cloud)
- EGI Cloud federation
  - Multi-cloud IaaS with SSO,VM image catalogue, discovery, accounting, monitoring, GUI dashboard, cloud compute , container compute, online storage training infrastructure, applications, notebooks
  - Data staging between HPC and Cloud
- Cambridge University CSD3 Supercomputer (CSD3 + Openstack) : Essentially using ironic instead of xCAT for the Slurm cluster (running node converted to image and pushed to the other nodes)
  - Many more steps for the HPC and Cloud convergence.

The next steps

- Accounting & review of allocations
- Transition to Production infrastructure and conclude the governance framework of the European Weather Cloud.

- Provide building blocks for services building
  - Go beyond infrastructure (IaaS → PaaS → SaaS)
  - Build a platform and share within the meteorological community
- Continue trials on federating with other clouds and systems
- Synergies with other projects and initiatives like Destination Earth (DestinE), European Open Science Cloud(EOSC), GAIA-X, participate/contribute to EU funded projects pertaining to the evolution of Cloud and Edge computing.



Openstack infrastructure summit link

ECMWF EUMETSAT EUROPEAN WEATHER CLOUD
CLOUD COMPUTING-BASED INFRASTRUCTURE, FOCUSED ON THE NEEDS OF THE METEOROLOGICAL COMMUNITY

European Weather Cloud: A community cloud service tailored for Meteorology

19th Workshop on high performance computing in meteorology
September 20-24, 2021

Q/A

Vasileios Baousis, Umberto Modigliani, Florian Pappenberger, Martin Palkovic, Stephan Siemen, Xavier Abellan, Charalampos Kominos

Vasileios A. Baousis (PhD)

ECMWF EUMETSAT EUROPEAN WEATHER CLOUD
CLOUD COMPUTING-BASED INFRASTRUCTURE, FOCUSED ON THE NEEDS OF THE METEOROLOGICAL COMMUNITY