

European Weather Cloud architecture

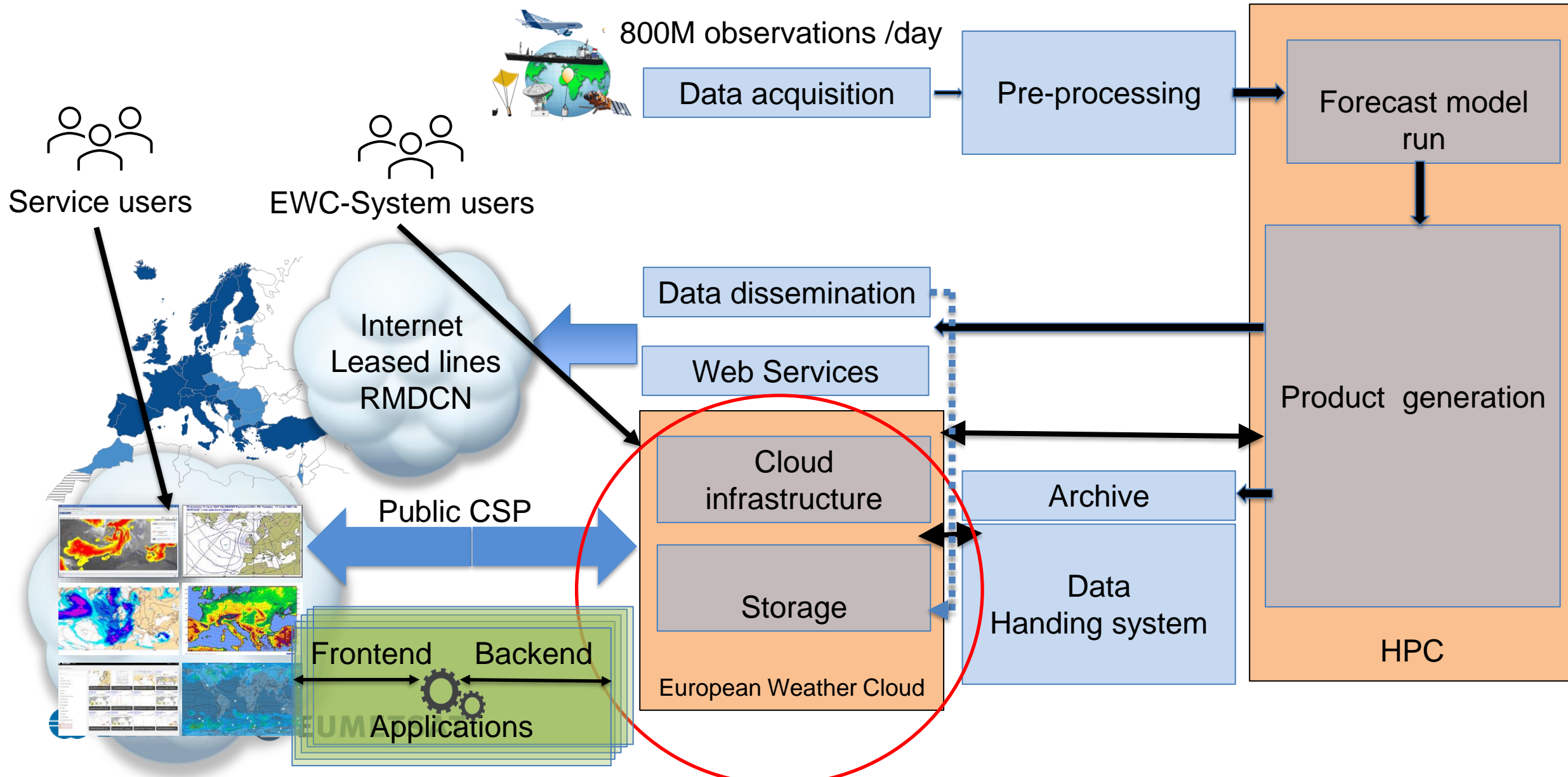
European Weather Cloud User Workshop
27th September 2023

Vasileios A. Baousis (PhD)-ECMWF
Ismael Cuadrado Cordero (PhD)-EUMETSAT

Agenda

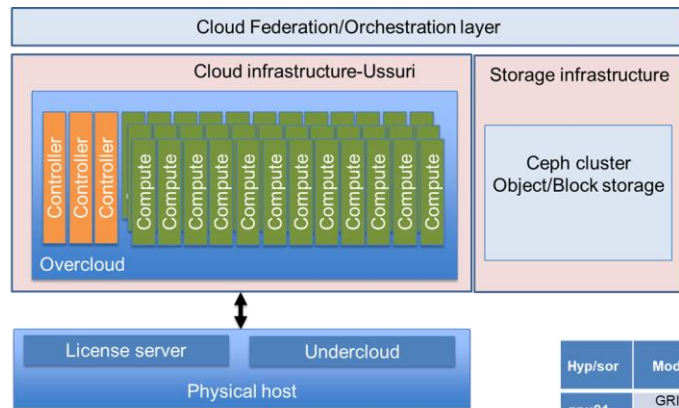
- European Weather Cloud : Symmetric service offering from both organizations
 - ECMWF's production workflow & European Weather Cloud
 - EWC ECMWF & EUMETSAT infrastructure
- EWC Federation
- Security & Shared responsibility

ECMWF's production workflow and European Weather Cloud



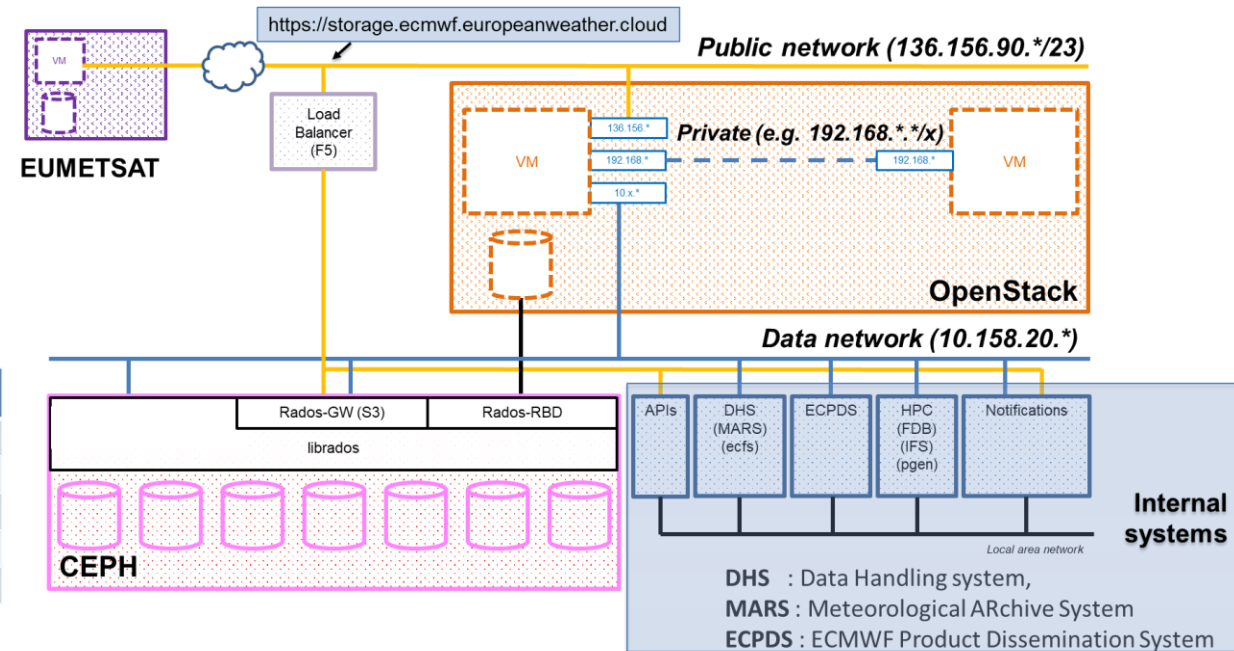
EWC Pilot infrastructure @ ECMWF

To be decommissioned this week



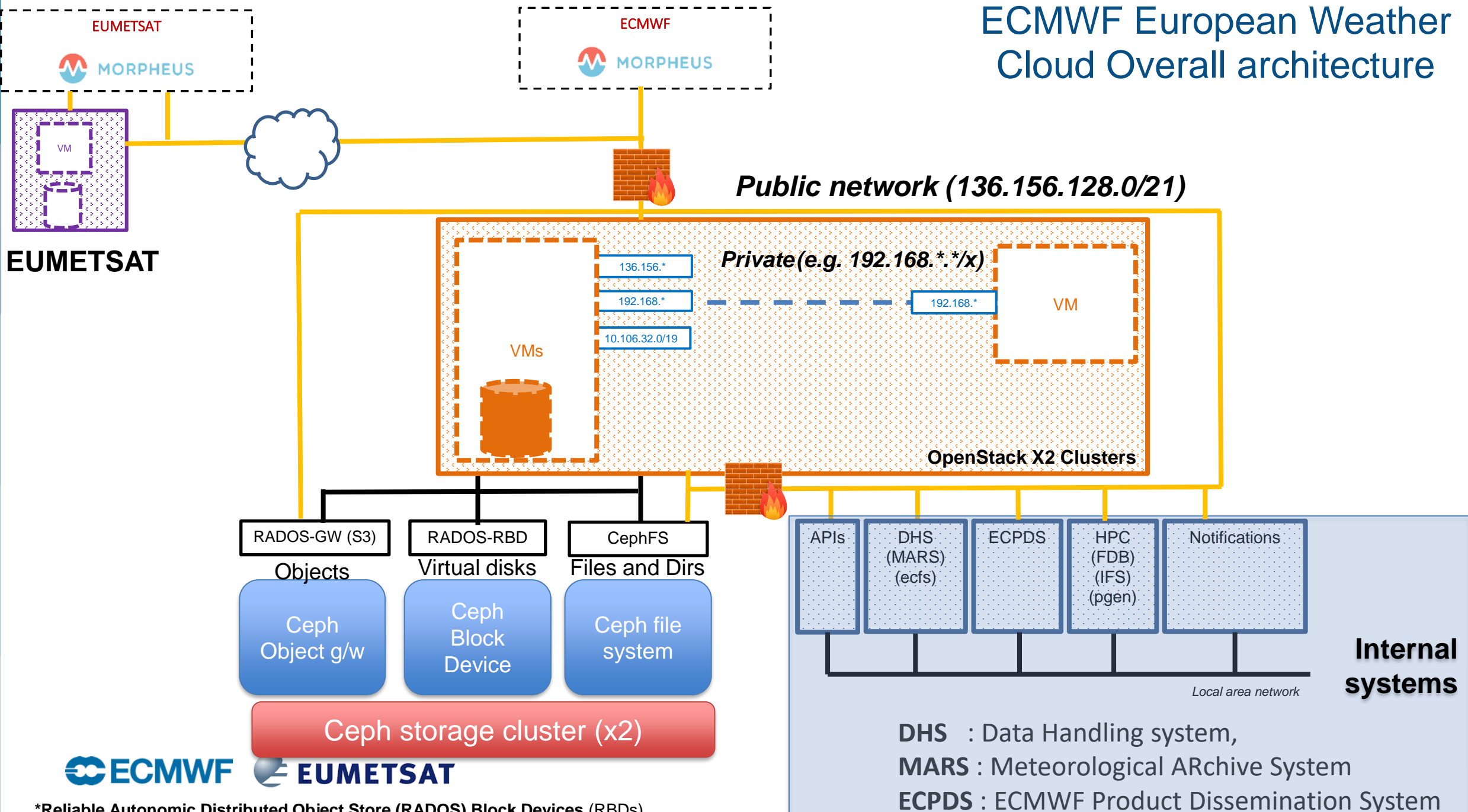
- Systems : Cloud:43 /Ceph: 23
- Cores : ~3000
- RAM : ~21TB
- Storage : ~1PB (HDD+SSD)
- GPUs : 2x5 NVIDIA Tesla V100

Hyp/sor	Model	Nova profile	VM/GPU /Model	VMs /host	Frame BufferSize	Max. Display Resolution
gpu01	GRID V100-16C	nvidia-301	1	2	16384	4096x21602
gpu02	GRID V100-8C	nvidia-300	2	4	8192	4096x21602
gpu03	GRID V100-8C	nvidia-300	2	4	8192	4096x21602
gpu04	GRID V100-4C	nvidia-299	4	8	4096	4096x21602
gpu05	GRID V100-4C	nvidia-299	4	8	4096	4096x21602



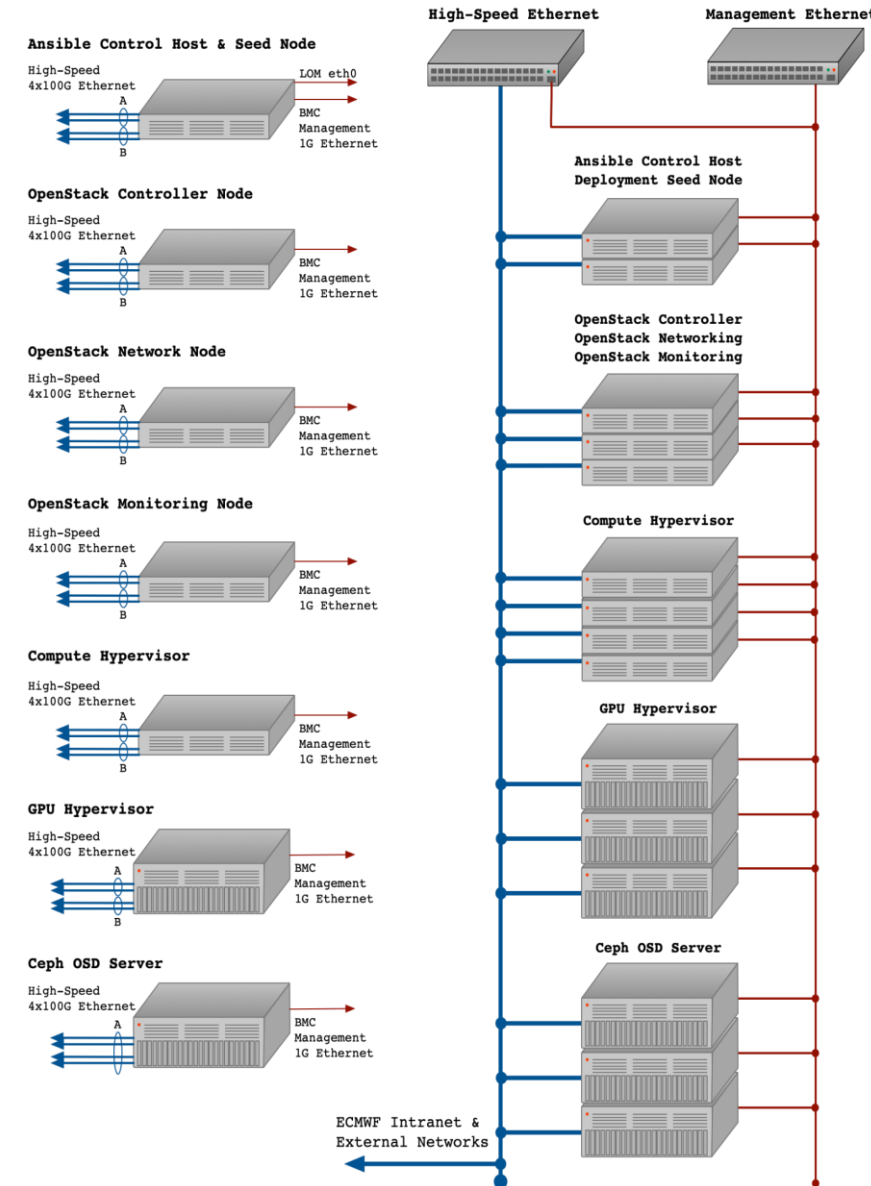
ECMWF built and managed

ECMWF European Weather Cloud Overall architecture



ECMWF/EWC Infrastructure in Bologna

- **Cloud infrastructure based on OpenStack**
 - Yoga release currently deployed
 - Upgrade to the latest stable release on annual basis
 - 2 independent clusters (each on different room)
 - 18 compute and 8 CPU, 3 network and 3 monitors nodes per cluster
 - SRIOV enabled, CPU pinning for some hosts etc
- **Storage infrastructure based on Ceph**
 - Quincy release currently deployed
 - Upgrade to the latest stable release on annual basis
 - 2 independent clusters (each on different room)
 - 9 nodes/cluster -42X18TB HDD/system
- **Networking: High-Speed Ethernet:**
 - Control plane hosts with either 2x2x100G or 4x100G bonds.
 - All switches in the High-Speed Ethernet fabric Juniper QFX5120-32C.
 - IPv6 (coming soon)



ECMWF/EWC Infrastructure in Bologna

1. Ceph nodes (9/cluster)

- **System** : DELL PowerEdge R7525,
- **CPU** : AMD 7713 2.0GHz, CPU Total: 256
- **Network** :2X2 ports Mellanox ConnectX-6 DX Dual Port 100GbE QSFP56],1 X2 NetXtreme BCM5720 Gigabit Ethernet PCIe
- **Memory** : 1TB (16 x 64GB RDIMM, 3200MT/s, Dual Rank)
- **Disks** : 2X 15.36TB Enterprise NVMe , 42X 18TB Hard Drive SAS ISE 12Gbps 7.2K 512e 3.5in Hot-Plug

Total/Cluster: 6.3 PiB raw/ 3 (replication) =2.1PB usable

2. Compute nodes (18 cluster)

- **System** : PowerEdge R7525
- **CPU** : AMD 7713 2.0GHz,64C/128T,256M **CPUTotal: 256**
- **Network** :2X2 ports Mellanox ConnectX-6 DX Dual Port 100GbE QSFP56],1 X2 NetXtreme BCM5720 Gigabit Ethernet PCIe
- **Memory**: 1TB (16 x 64GB RDIMM, 3200MT/s, Dual Rank)
- **Disks** : 2X960GB SSD vSAS Read Intensive 12Gbps 512e 2.5in Hot-Plug ,AG Drive SED, 1DWPD

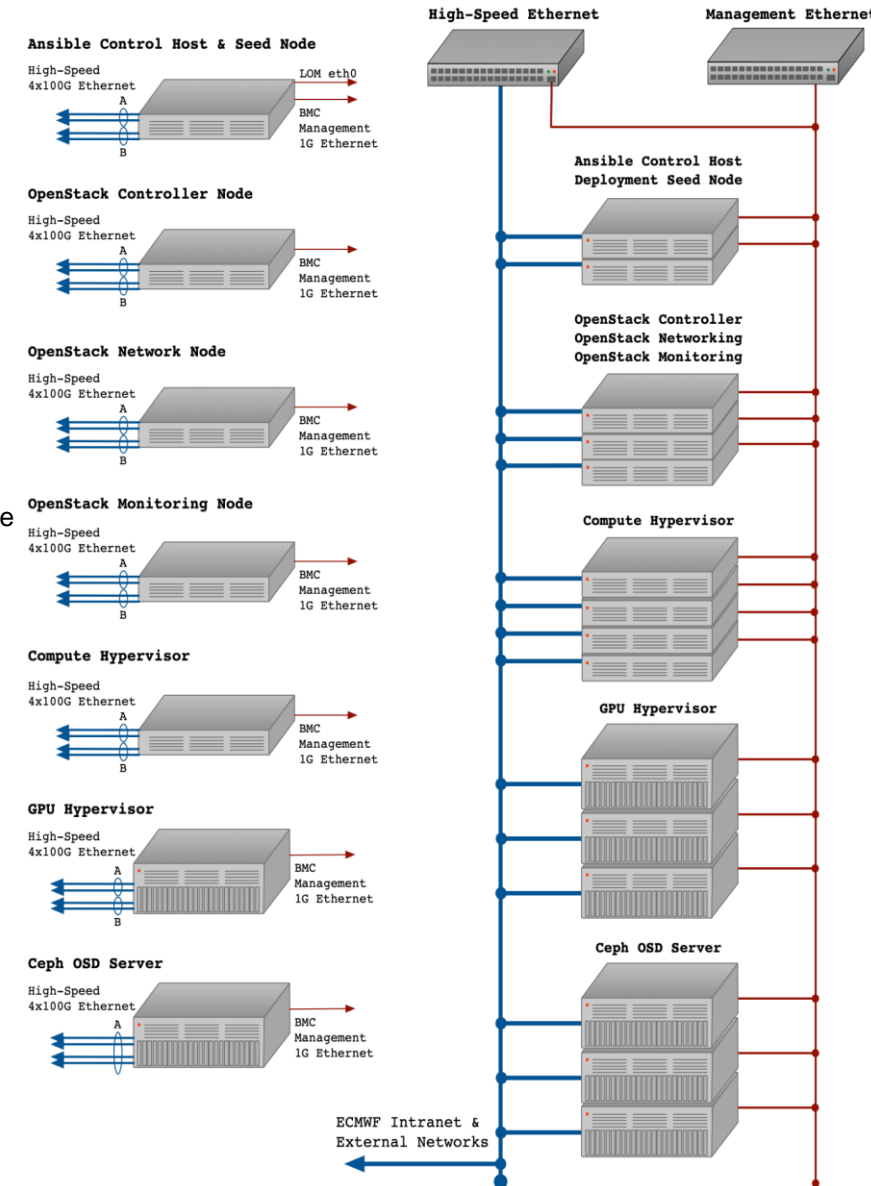
3. GPU Compute nodes (8/cluster)

- **System** :PowerEdge R7525
- **CPU**: AMD 7543 2.8GHz,32C/64T,256M, **CPU Total =128**
- **GPUs** :2 X NVIDIA Ampere A100, PCIe,/node

4.Network/Monitor nodes (3/3/cluster)

- **System** : PowerEdge R7525
- **CPU** : AMD 7713 2.0GHz,64C/128T,256M **CPUTotal: 256**
- **Network** : 2X2 ports Mellanox ConnectX-6 DX Dual Port 100GbE QSFP56],
- **Memory**: 512 GB (16 x 32GB RDIMM, 3200MT/s, Dual Rank)
- **Disks** : 2X960GB SSD vSAS Read Intensive 12Gbps 512e 2.5in Hot-Plug ,AG Drive SED, 1DWPD

Total : 5632 vCPUs/Cluster or 11264 total (with hyperthreading)



EWC Infrastructure in Bologna/GPUs

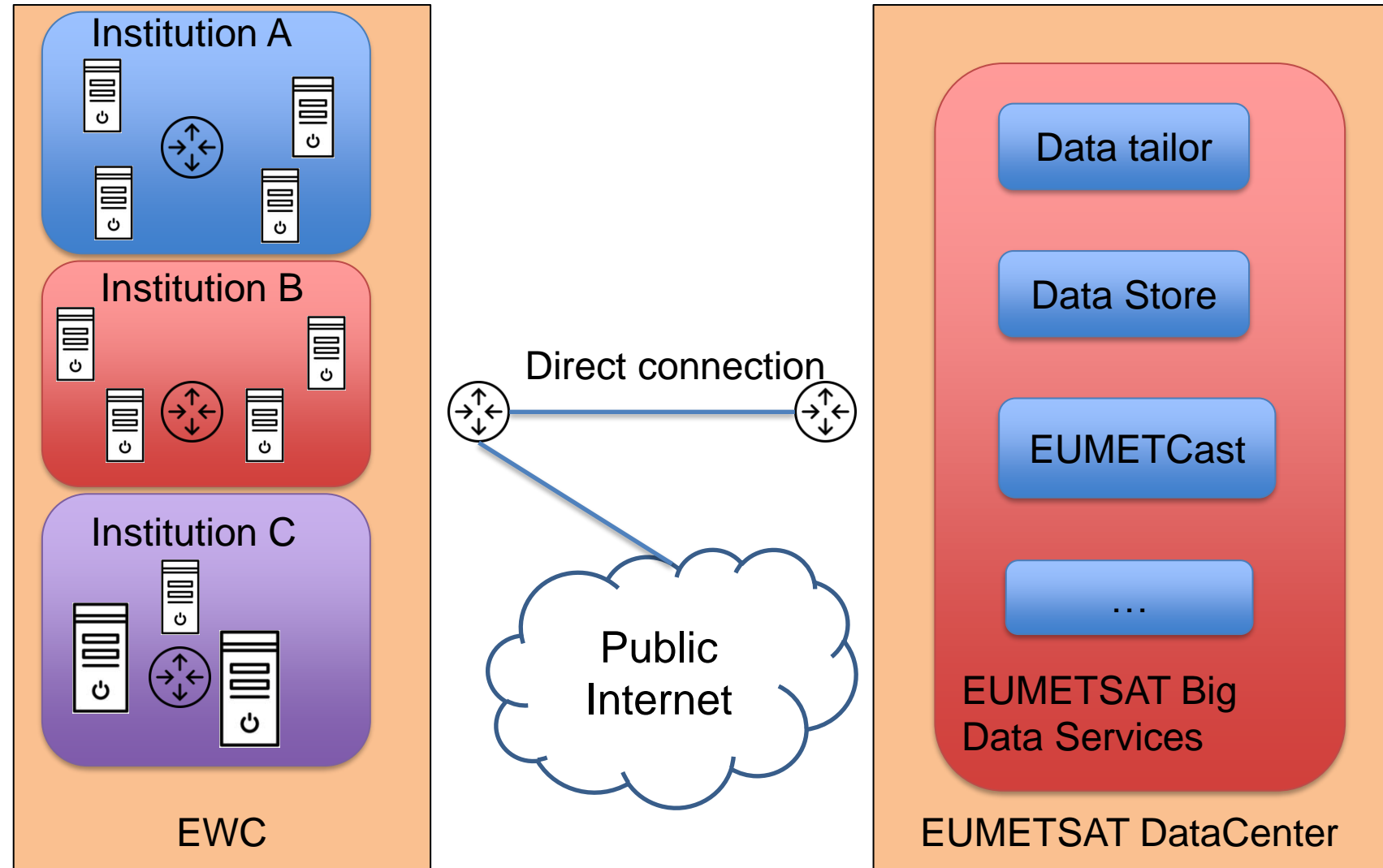
Compute/Network/Monitor nodes (8/cluster)

- **GPUs** :2 X NVIDIA Ampere A100, PCIe,/node
- CPU pinning & NUMA passthrough.
- SRIOV networking available.
- 8GB RAM per VCPU.

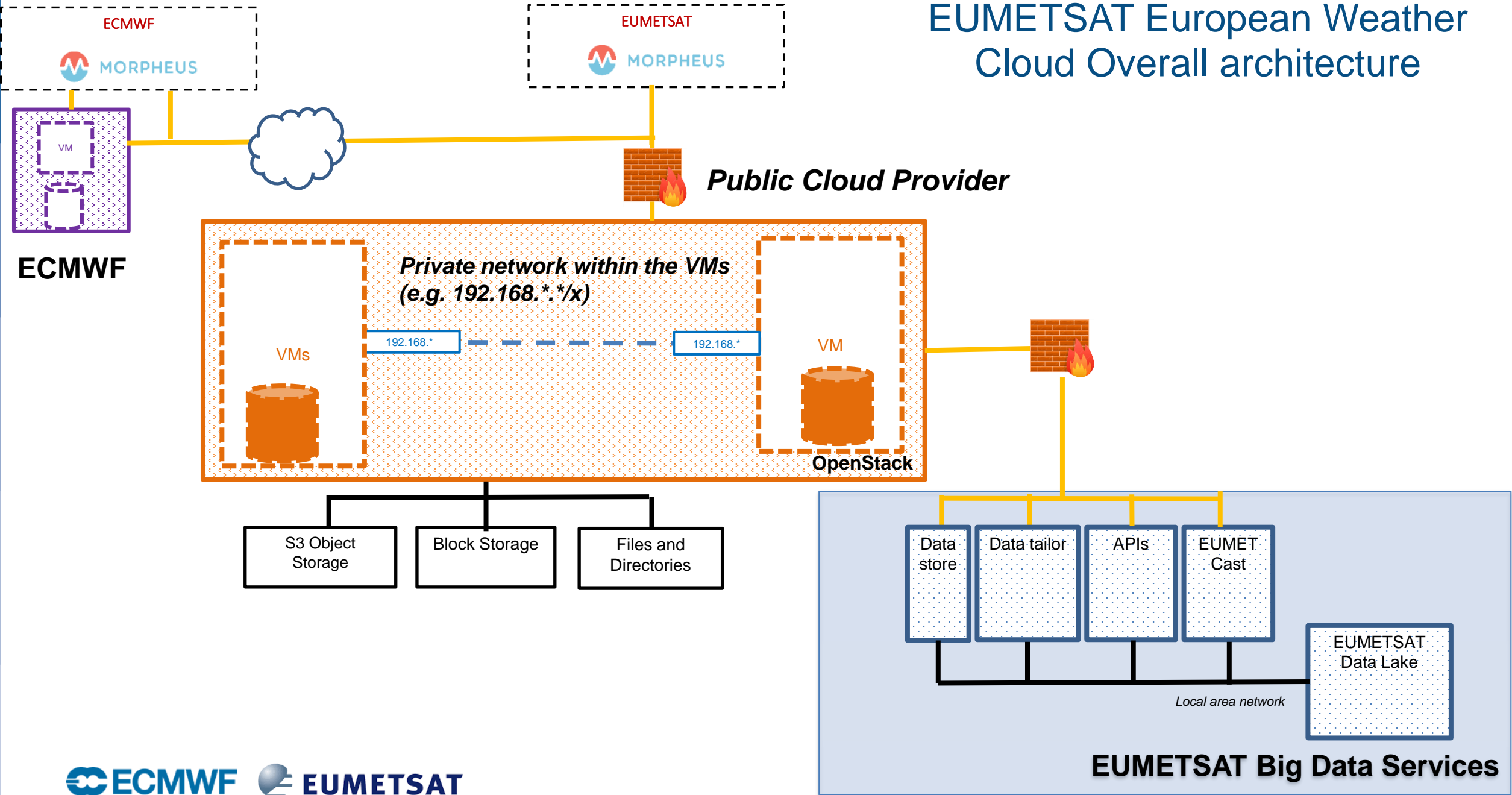
Hypervisor	Mode	GPUs/host	Model	Nova profile	VMs/host	Frame BufferSize	Maximum Display Resolution
computegpu000	MIG	2	GRID A100D-1-10C	nvidia-699	2	framebuffer=10240M,	max_resolution=4096x2400
computegpu000	MIG		GRID A100D-2-20C	nvidia-700	6	framebuffer=20480M	max_resolution=4096x2400
computegpu001	MIG	2	GRID A100D-1-10C	nvidia-699	2	framebuffer=10240M,	max_resolution=4096x2400
computegpu001	MIG		GRID A100D-2-20C	nvidia-700	6	framebuffer=20480M	max_resolution=4096x2400
computegpu002	MIG	2	GRID A100D-1-10C	nvidia-699	2	framebuffer=10240M,	max_resolution=4096x2400
computegpu002	MIG		GRID A100D-2-20C	nvidia-700	6	framebuffer=20480M	max_resolution=4096x2400
computegpu003	MIG	2	GRID A100D-1-10C	nvidia-699	2	framebuffer=10240M,	max_resolution=4096x2400
computegpu003	MIG		GRID A100D-2-20C	nvidia-700	6	framebuffer=20480M	max_resolution=4096x2400
computegpu004	non-MIG	2	GRID A100D-40C	nvidia-697	4	framebuffer=40960M	max_resolution=4096x2400
computegpu005	non-MIG	2	GRID A100D-40C	nvidia-697	4	framebuffer=40960M	max_resolution=4096x2400
computegpu006	non-MIG	2	GRID A100D-40C	nvidia-697	4	framebuffer=40960M	max_resolution=4096x2400
computegpu007	non-MIG	2	GRID A100D-40C	nvidia-697	2	framebuffer=40960M	max_resolution=4096x2400
computegpu007	non-MIG		GRID A100D-80C	nvidia-698	1	framebuffer=81920M	max_resolution=4096x2400

EUMETSAT's data offering and European Weather Cloud

- EUMETSAT utilises public Cloud resources to host EWC users
- Each tenancy receives a certain amount of resources, including access to GPUs
- Computing representatives ensure the distribution of these resources between the users
- Users get faster-than-the-internet access to EUMETSAT's data via EUMETSAT's data services



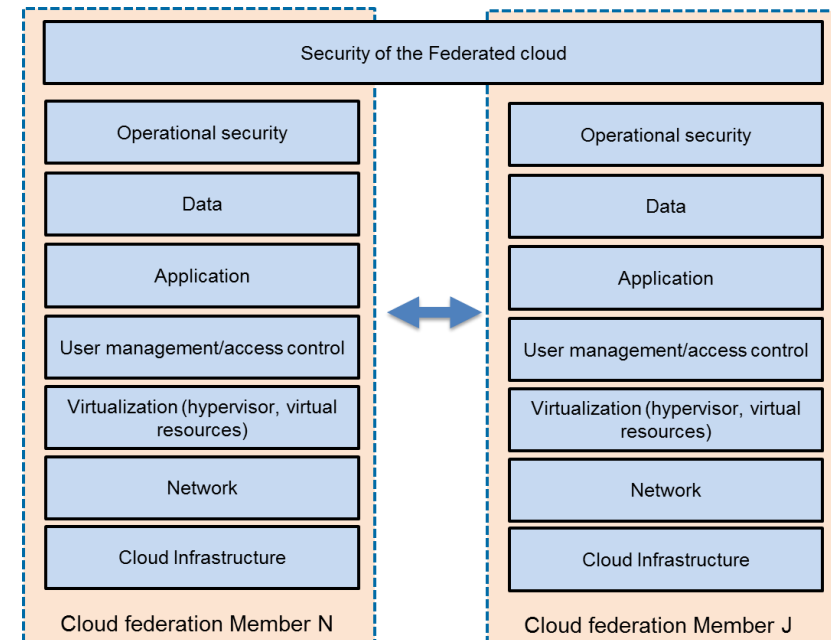
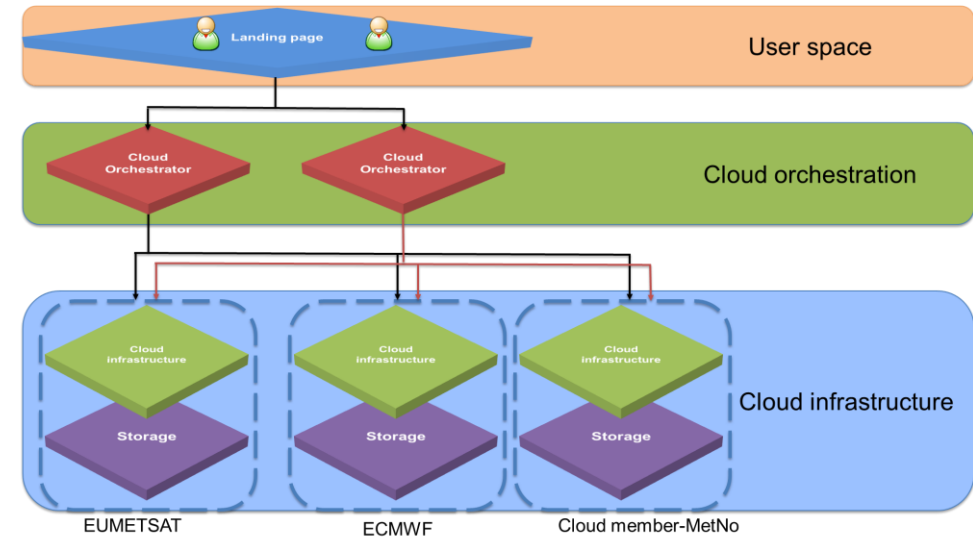
EUMETSAT European Weather Cloud Overall architecture



*Reliable Autonomic Distributed Object Store (RADOS) Block Devices (RBDs)

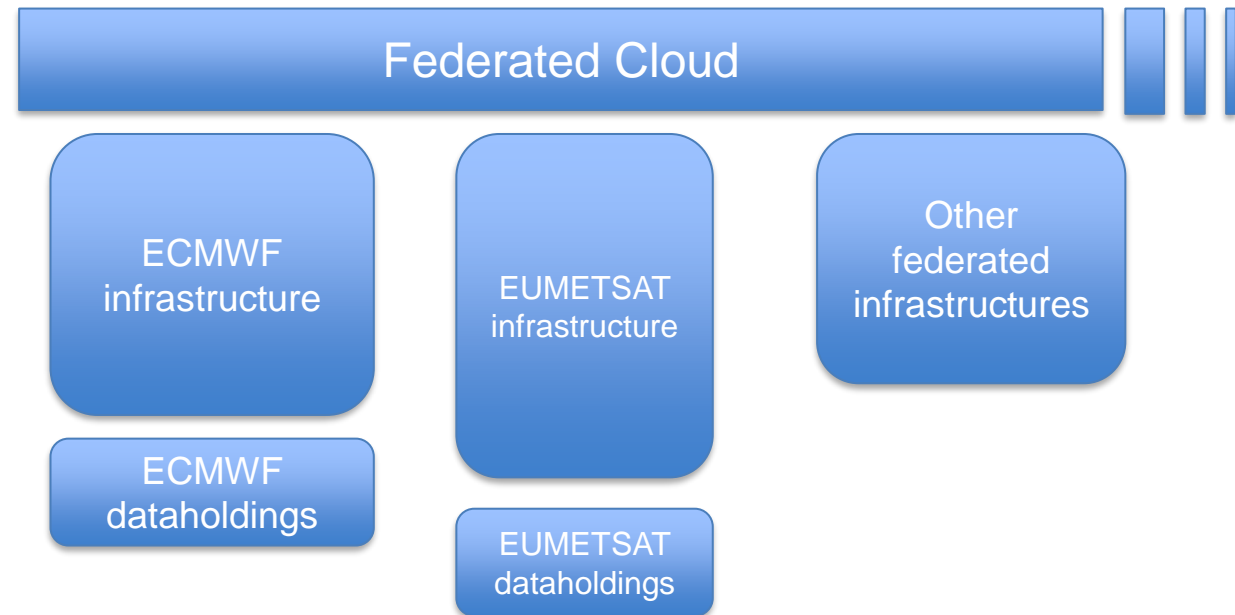
Federation with Member States

- 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
- Security is a shared responsibility between Cloud service providers(CSP) and between CSP and users.



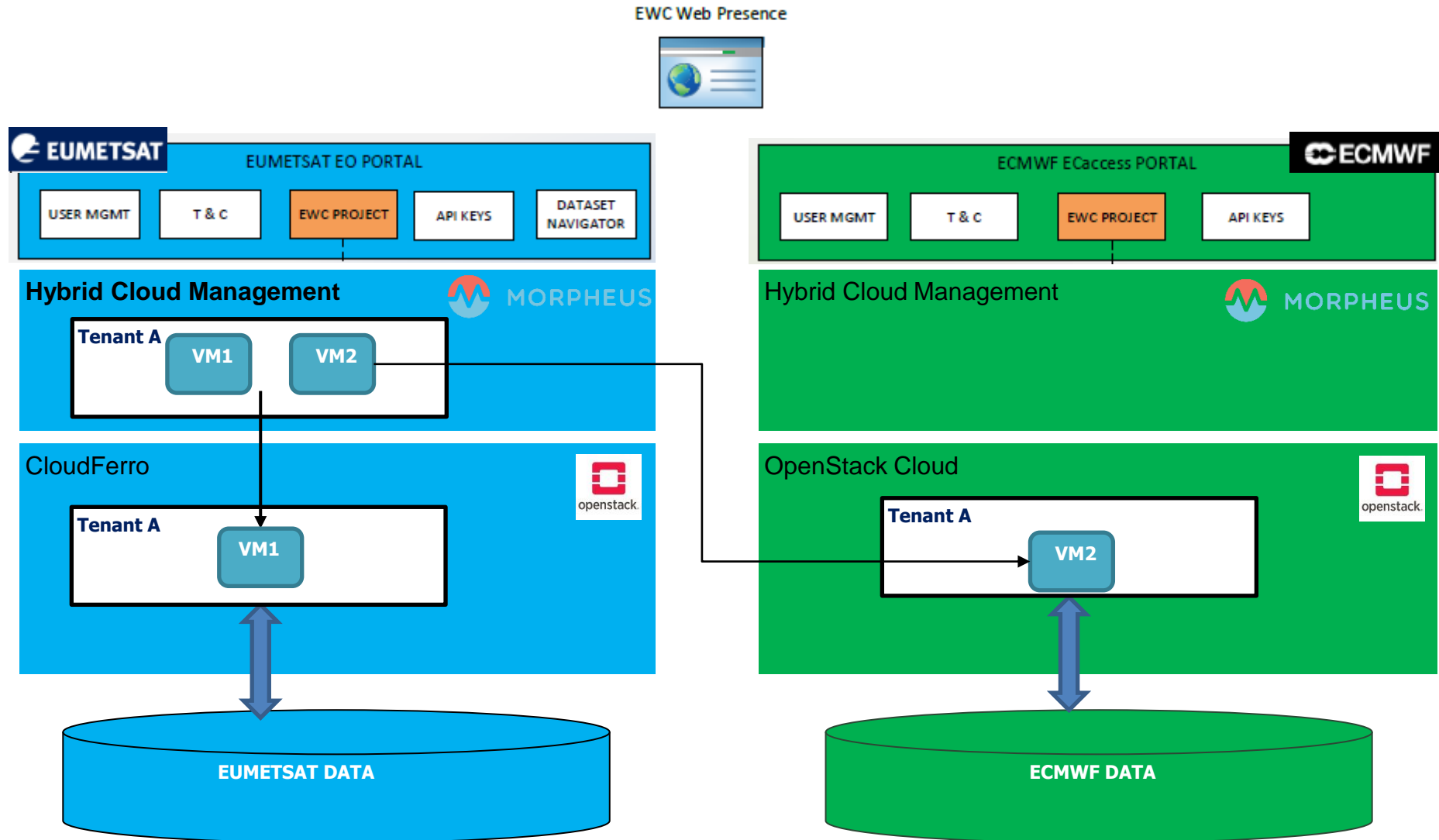
European Weather Cloud as a shared infrastructure

- The user perceives a seamless service, and users from both organizations get privileged access to computation and data access of all federated partners.
- Infrastructural advantages
 - Federated members retain control of their cloud, data and user database
 - Organizations can offer their users a larger computation power
 - Increases cooperation between user bases



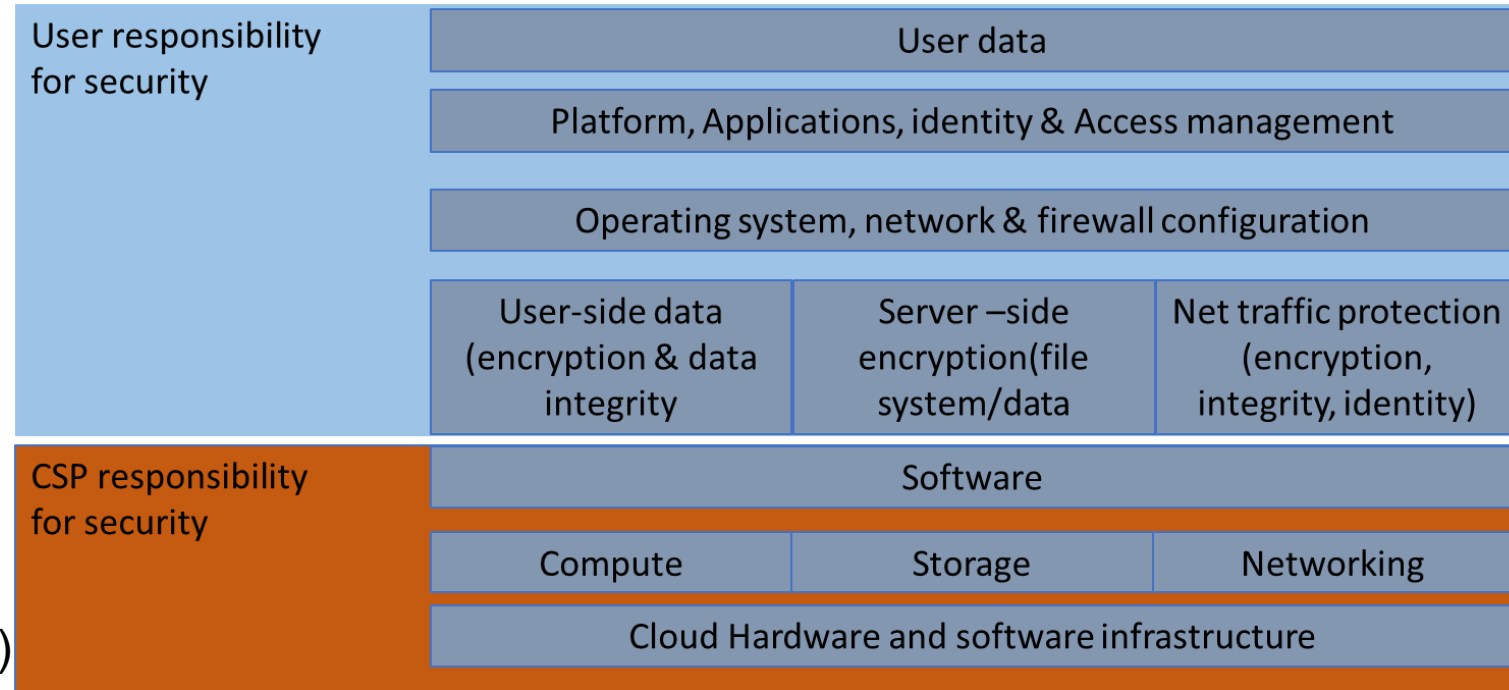
Federation with Member States

- Users can deploy, within the same application, computing resources on the infrastructures of all federated partners
- These computing resources remain always “close to the data” so that the user can select the infrastructure in which to deploy their resources, using data from all partners in the most efficient way



Shared responsibility-Security by design

- Provided images
 - Rocky and Ubuntu prepared by E&E
 - Automatic package update enabled
 - Fail2ban enabled
- Firewall
- OpenStack security groups (rules for ingress and egress traffic)
- Security is a shared responsibility between Cloud Service Providers (E&E) and between CSP and users i.e. E&E and Member/Cooperating State users.



European Weather Cloud architecture

European Weather Cloud User Workshop
27th September 2023

Vasileios A. Baousis (PhD)-ECMWF
Ismael Cuadrado Cordero (PhD)-EUMETSAT