ScienceBox

E. Bocchi
On behalf of the SWAN team

https://cern.ch/swan

Oct 11th, 2019
SWAN Users' Workshop

Why ScienceBox?

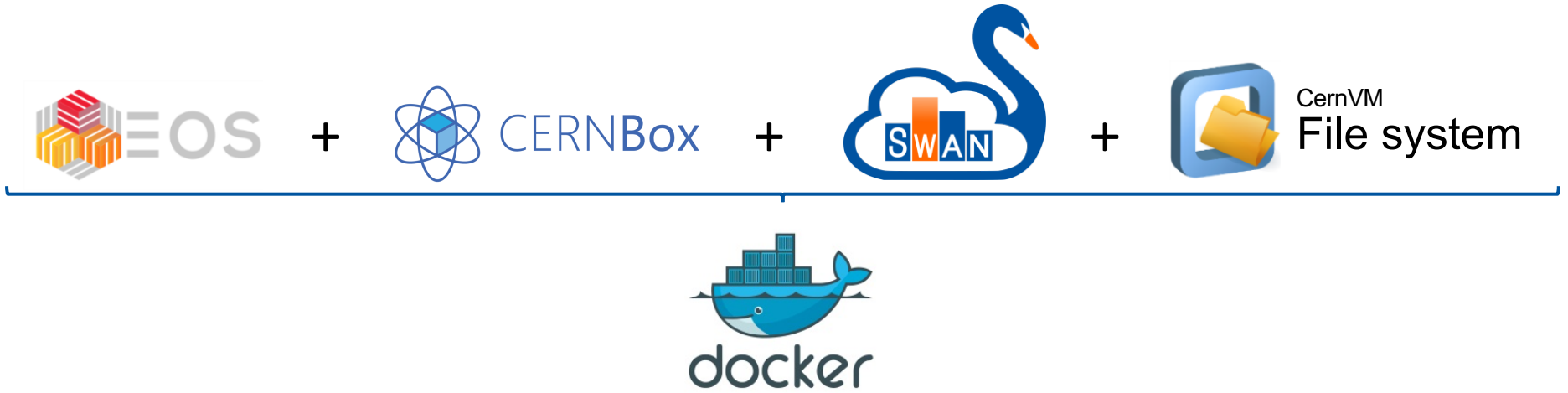
> Growing interest in CERN cloud software from external collaborators
- High Energy Physics sites
- National Research and Education Networks
- European projects

> Facilitate distribution and deployment outside CERN
- Simplified installation leveraging on container technologies
- Allow collaboration and contributions from expanding community

> Disposable deployment for development at CERN
- Testing software updates
- New functionalities (e.g., GPU support)

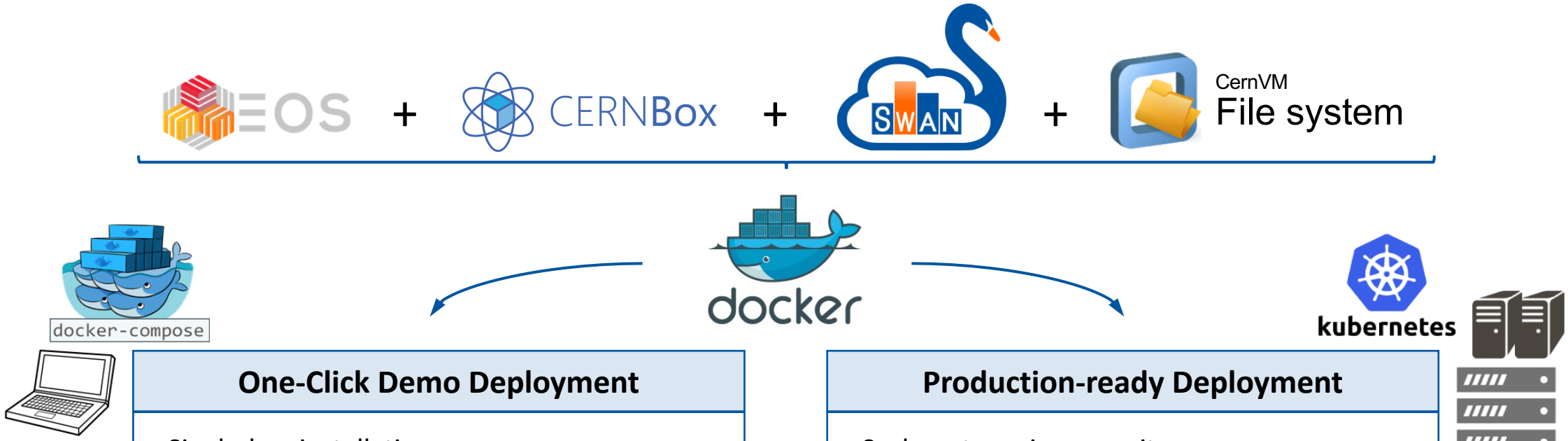
ScienceBox

> Self-contained Docker-based software package



ScienceBox

> Self-contained Docker-based software package



One-Click Demo Deployment

- Single-box installation
- Download and run in 5 minutes

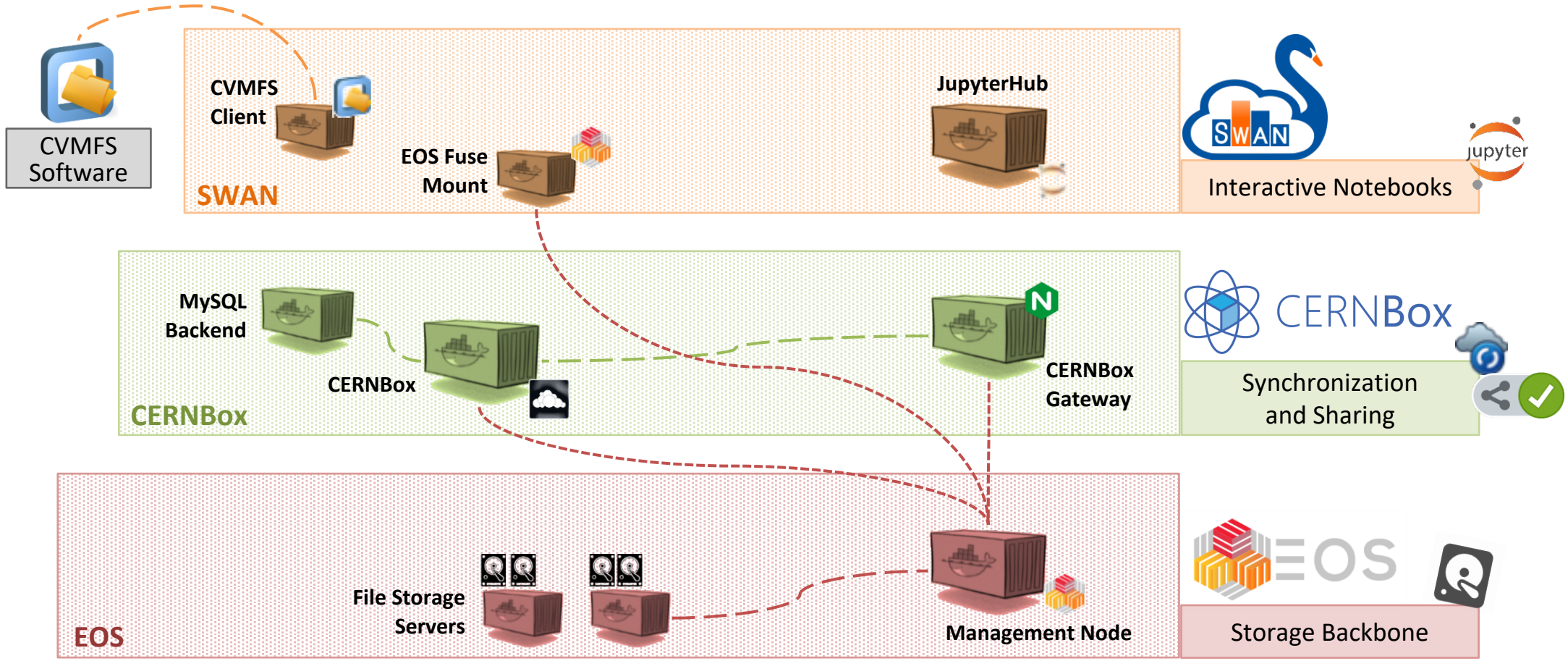
https://github.com/cernbox/uboxed

Production-ready Deployment

- Scale out service capacity
- Tolerant to node failures

https://github.com/cernbox/kuboxed

ScienceBox Architecture

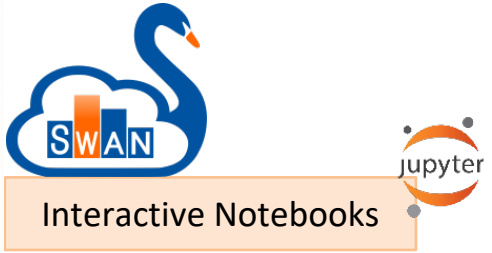


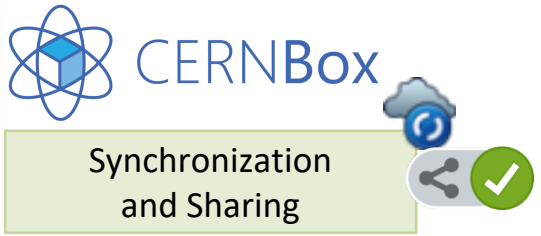
ScienceBox Architecture

> Minimal deployment with 12 containers
  - Less than 20GB required on disk

> Scale out computing and storage
  - Add more resources to the cluster
  - Spawn additional containers to use the new capacity

> Replicate containers for high-availability
  - Web frontends, proxies, DBs for CERNBox and SWAN
  - Management node for EOS





Use Cases

Up to University

> Allow students in high-schools to adopt tools used in science
  - SWAN – Full data analysis ecosystem in a web browser
  - CERNBox – Cloud storage for easy sharing and access form any device

Up To University

> ScienceBox in production for Up2U users for 1.5 years
  - Deployed at Poznan Supercomputing and Networking Center, Poland
  - Kubernetes on VMs, Ceph volumes for persistent storage

> Pilot service at CERN – http://up2u.cern.ch
  - CERNBox and SWAN on Kubernetes VMs
  - EOS on VMs and bare metal disks

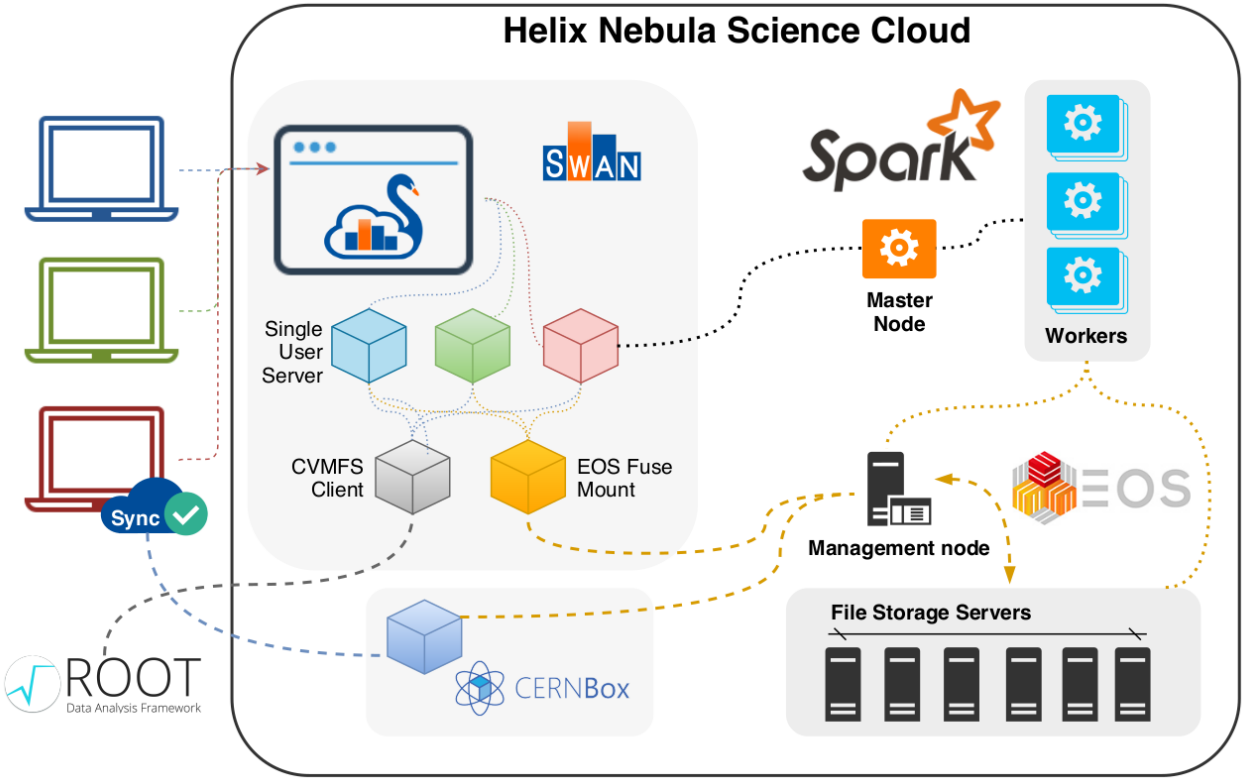
openstack

Totem Analysis on Helix Nebula Cloud

> Deployment on commercial cloud
  - 2000+ CPUs
  - 10+ TB memory
  - Virtually unlimited block storage

> ScienceBox with Apache Spark for massive computations

> Full TOTEM Analysis
  - Dataset: 4.7 TB, 1153 files
  - Data imported via xrootd
  - Results synchronized back via CERNBox client



Big Data Tools and Cloud Services for High Energy Physics Analysis in TOTEM Experiment - V. Avati et al.
https://ieeexplore.ieee.org/document/8605741

More External Sites and Collaborators

> External SWAN deployments inspired by ScienceBox

- Australia's Academic and Research Network (AARNET)

  AARNet's CloudStor SWAN deployment facilitated by CERN's ScienceBox
  Michael D'Silva @10:40

- SURFSARA, The Netherlands
- Joint Institute for Nuclear Research (JINR), Russia
- Academia Sinica Grid Computing Centre (ASGC), Taiwan

Future Outlook

CS3MESH

> EU-funded project
- 6M EUR, 12 partners, 2020-2022

> Goal: Global collaborative environment for research
- Share documents, files, projects, data, …
- Connected Application Hubs
- Data/metadata-aware workflows ➔ FAIR
  - Find, Access, Interoperate, Reuse

> Federation of existing CS3 sites
- 30+ sites (e.g. CERNBox, DesyBox, Universities, …)
- 300K+ users
- cs3community.org

CS3

Cloud Storage Services
for Synchronization and Sharing

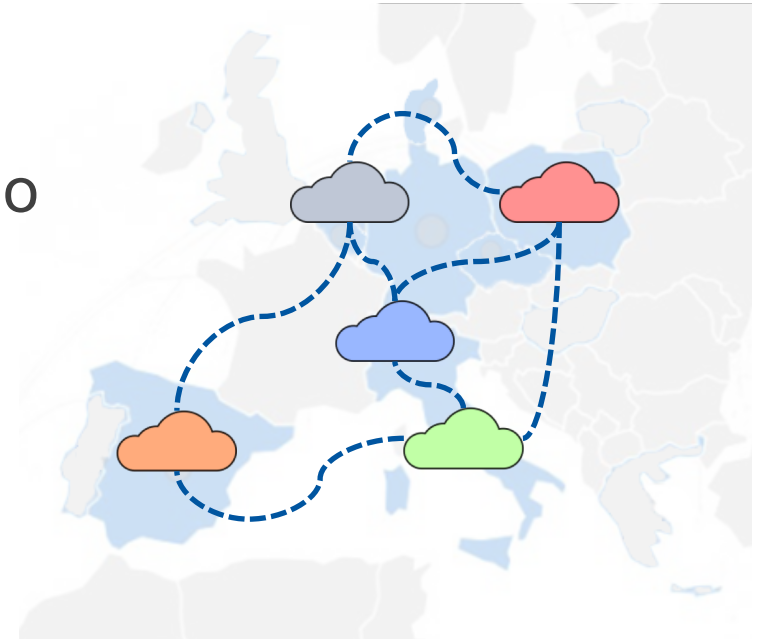
CERN

ScienceBox for CS3MESH

> ScienceBox is the reference platform in CS3MESH for distribution and deployment of cloud software

> SWAN will become part of the federated service portfolio
- Bigger community using CERN software
- Contribute back and improve SWAN for CERN users

> Benefits for SWAN users at CERN
- Share SWAN projects beyond the CERN borders
- Work easily with your experiment collaborators outside CERN



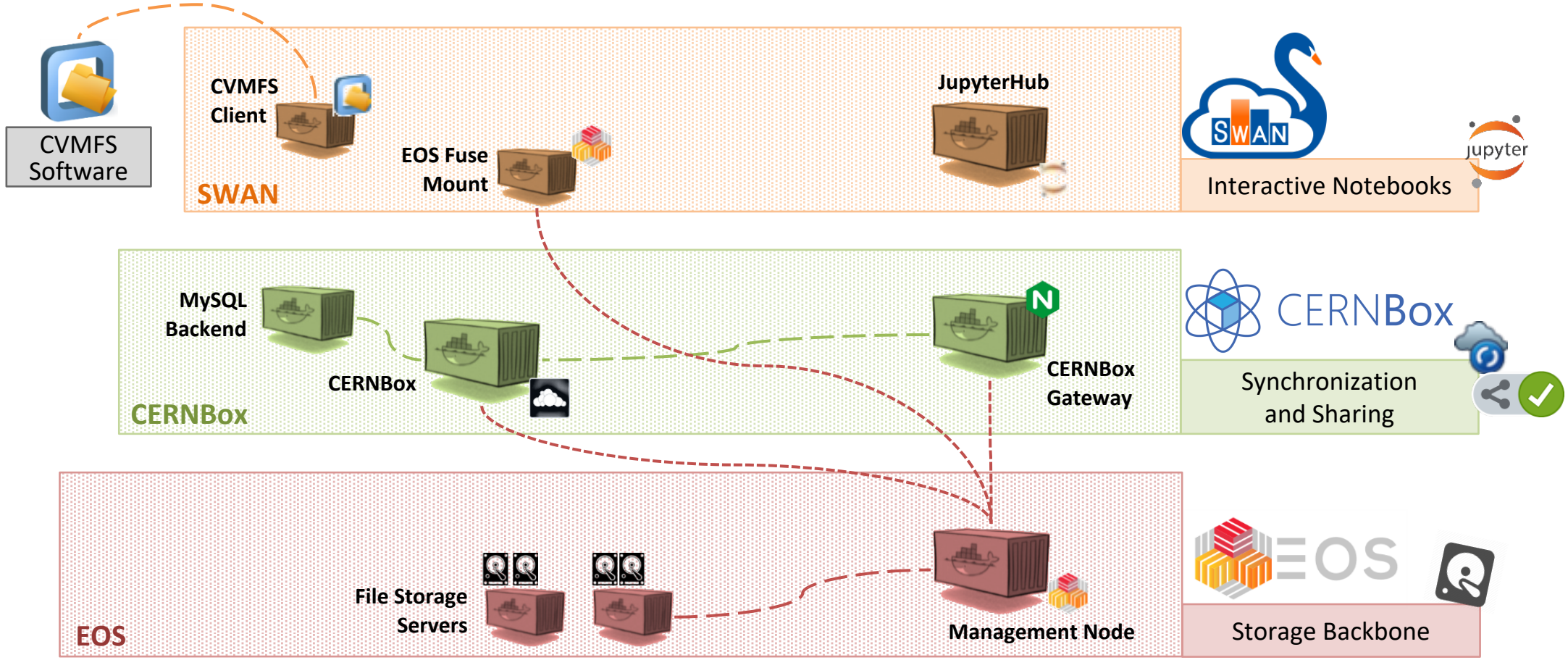
ScienceBox

Thank you

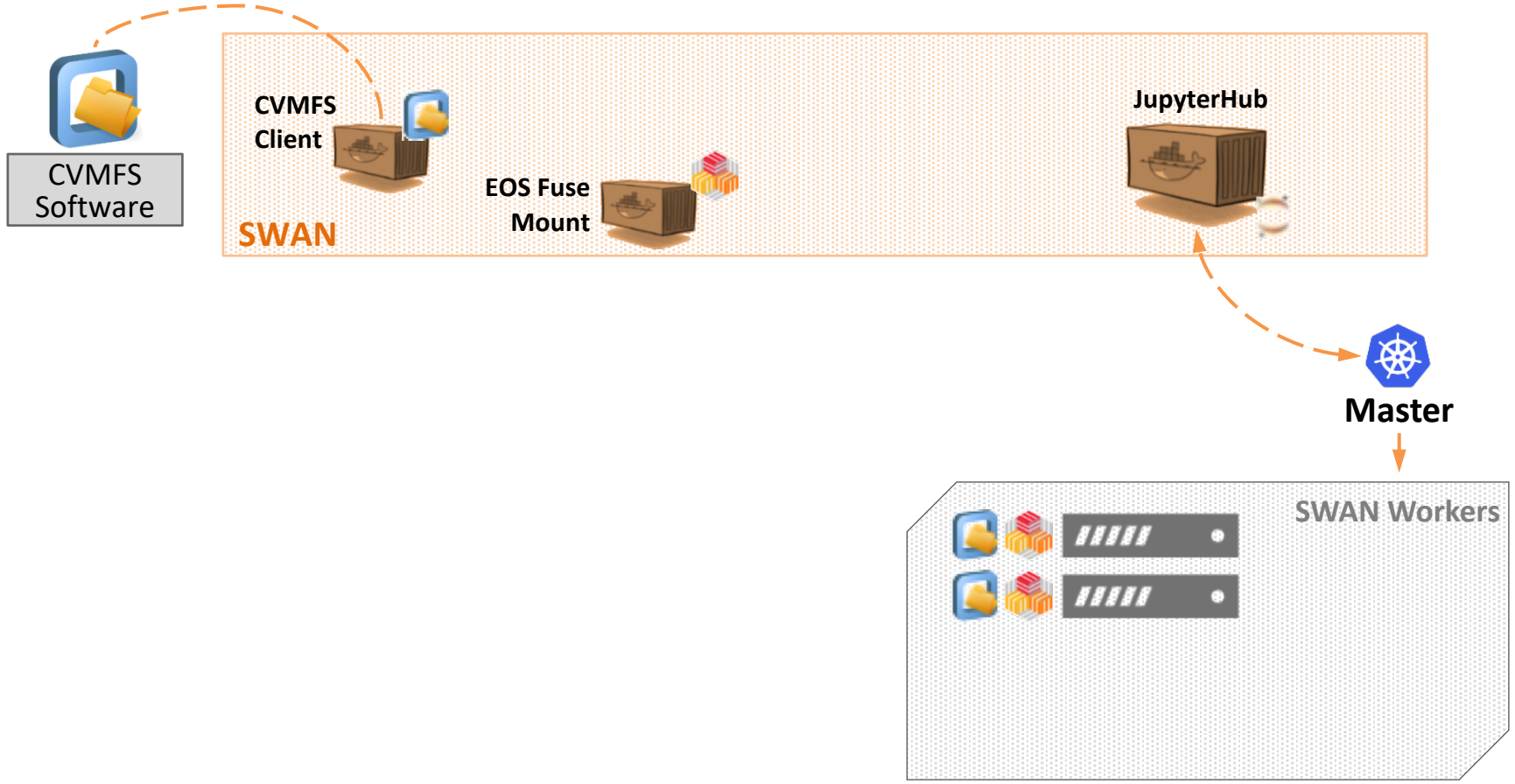
Enrico Bocchi
enrico.bocchi@cern.ch

Backup Slides

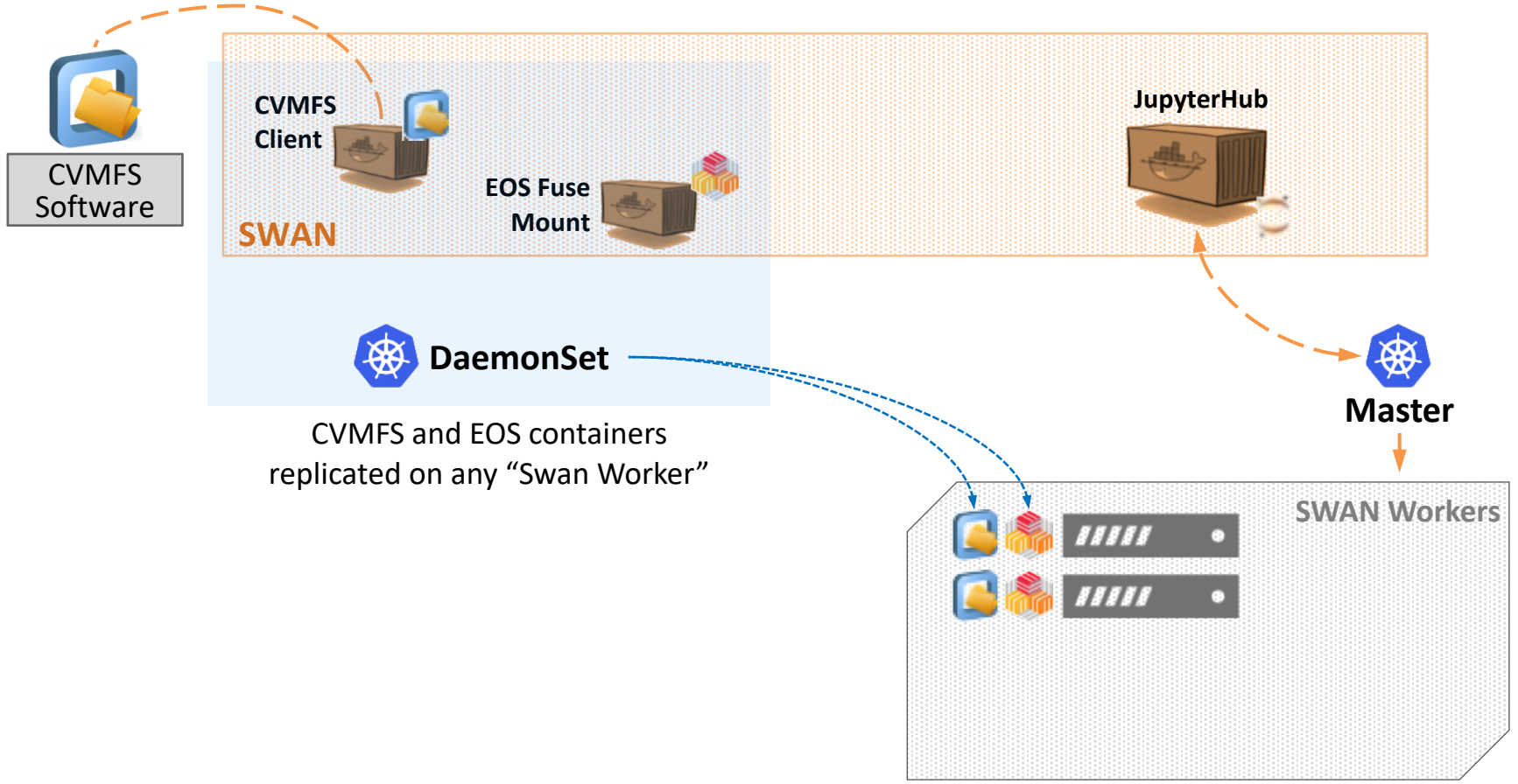
Architecture



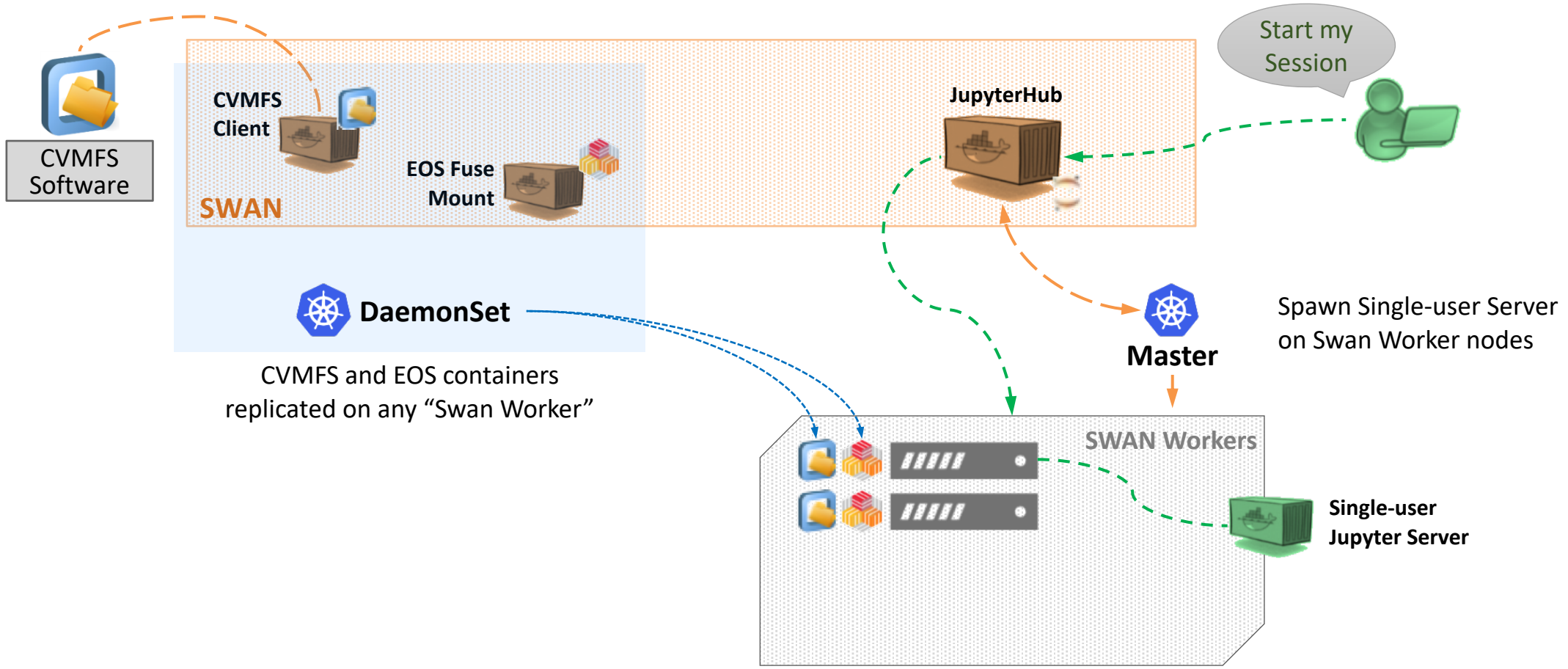
Architecture



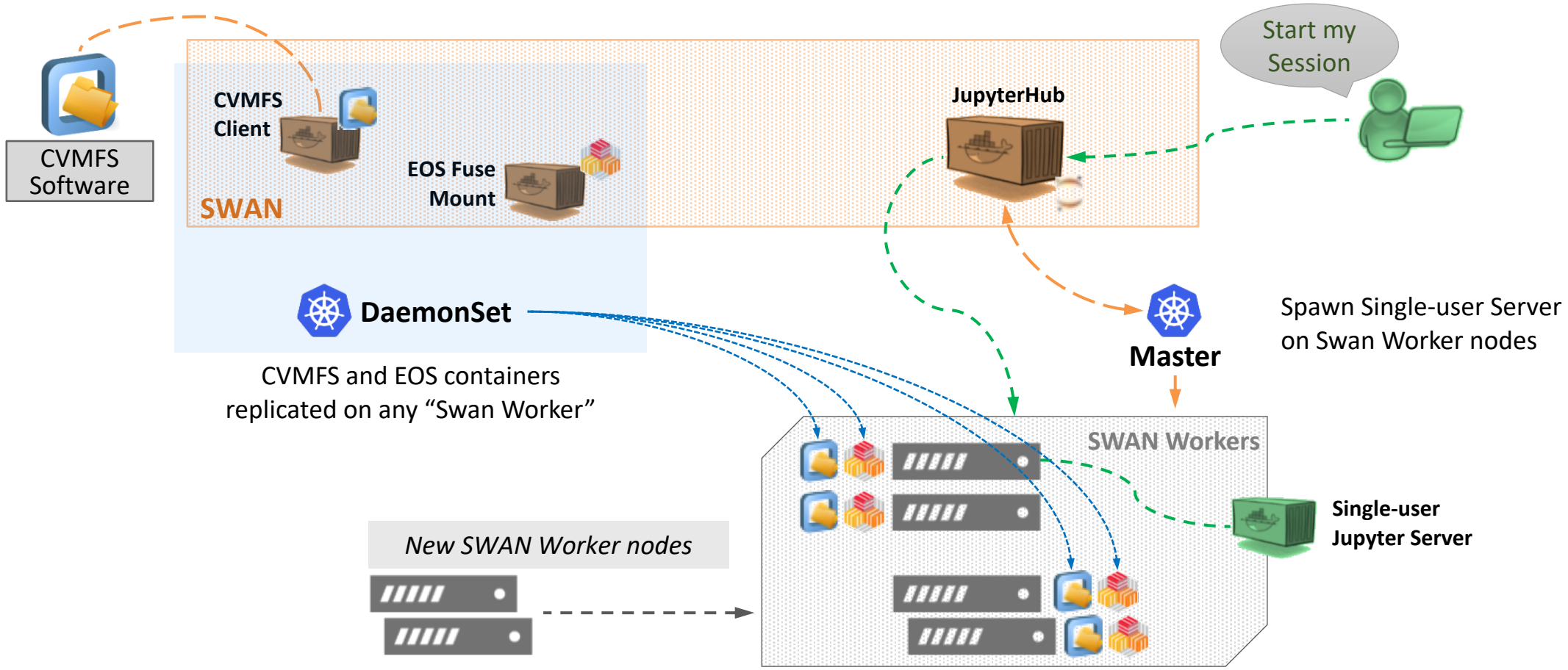
Architecture



Architecture



Architecture



SWAN production to Kubernetes

SWAN production to Kubernetes

> Modernize SWAN infrastructure
- Replicated, highly-available containers
- Add capacity in minutes (e.g., for tutorials and trainings)
- Leverage on Cloud Containers service by IT-CM

> Improve user experience
- Roll out updates with no impact on user's session and service capacity
- Dedicated cluster-wide CVMFS cache
- Private EOS mount for each user (under evaluation)

SWAN production to Kubernetes

> Single Docker image for SWAN production and ScienceBox

- SWAN production is an instance of ScienceBox
- Configuration and customizations applied at run time
- Allow contributions from ScienceBox community to improve SWAN at CERN