ScienceBox 2.0

Evolving the demonstrator package for CERN Storage and Analysis services

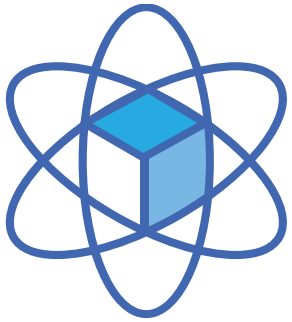
Samuel Alfageme
Enrico Bocchi
Jimil Desai

May 9, 2023

What is ScienceBox?

- **Demonstrator package for (some of) CERN's Storage and Analysis services**

EOS



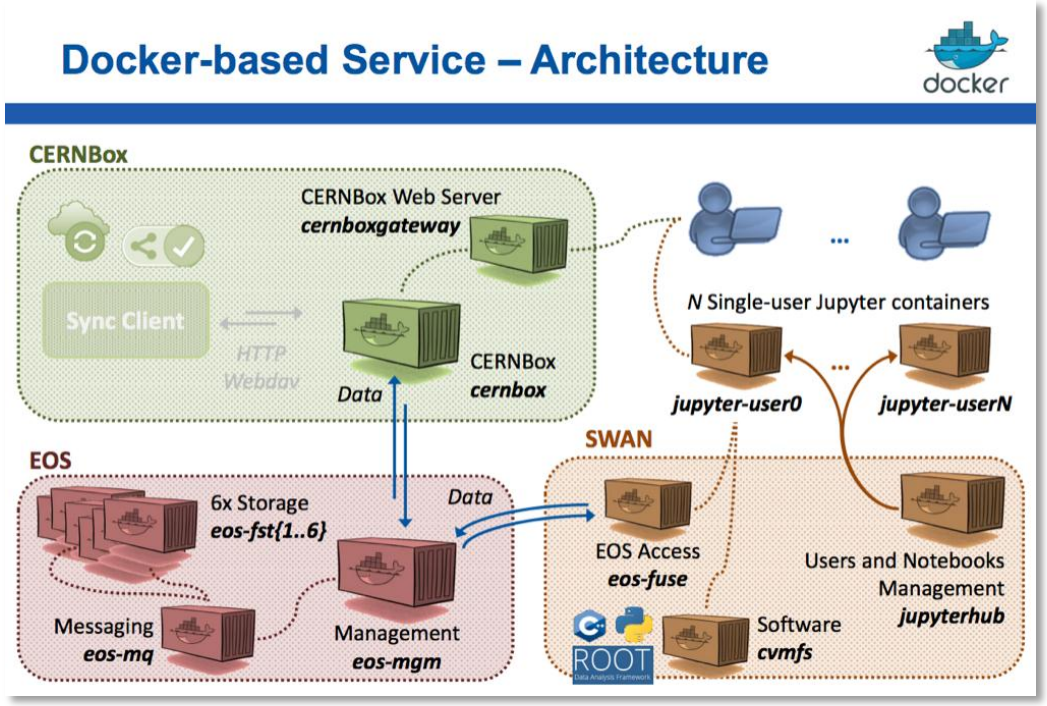
CERNBox

SWAN

CVMFS

- Enable non-CERN users to evaluate and deploy storage, sync&share, and analysis services on external premises
- Offer a turn-key self-configuring package to hide the configuration complexity

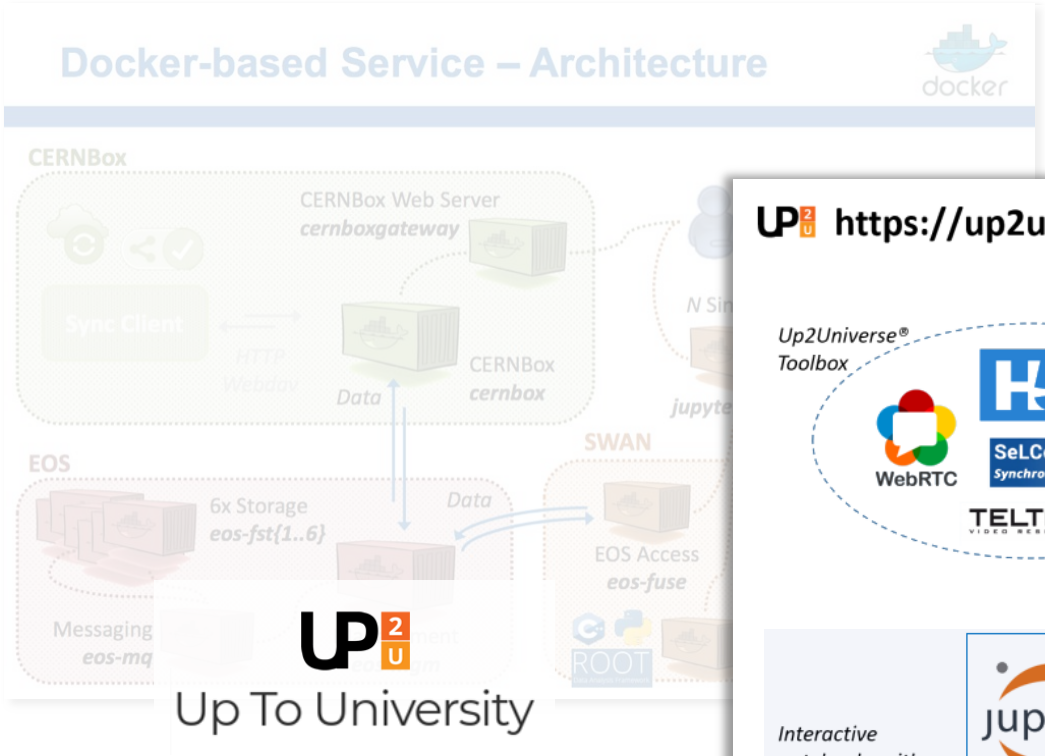
ScienceBox Heritage



***Early ScienceBox**, c.2017*

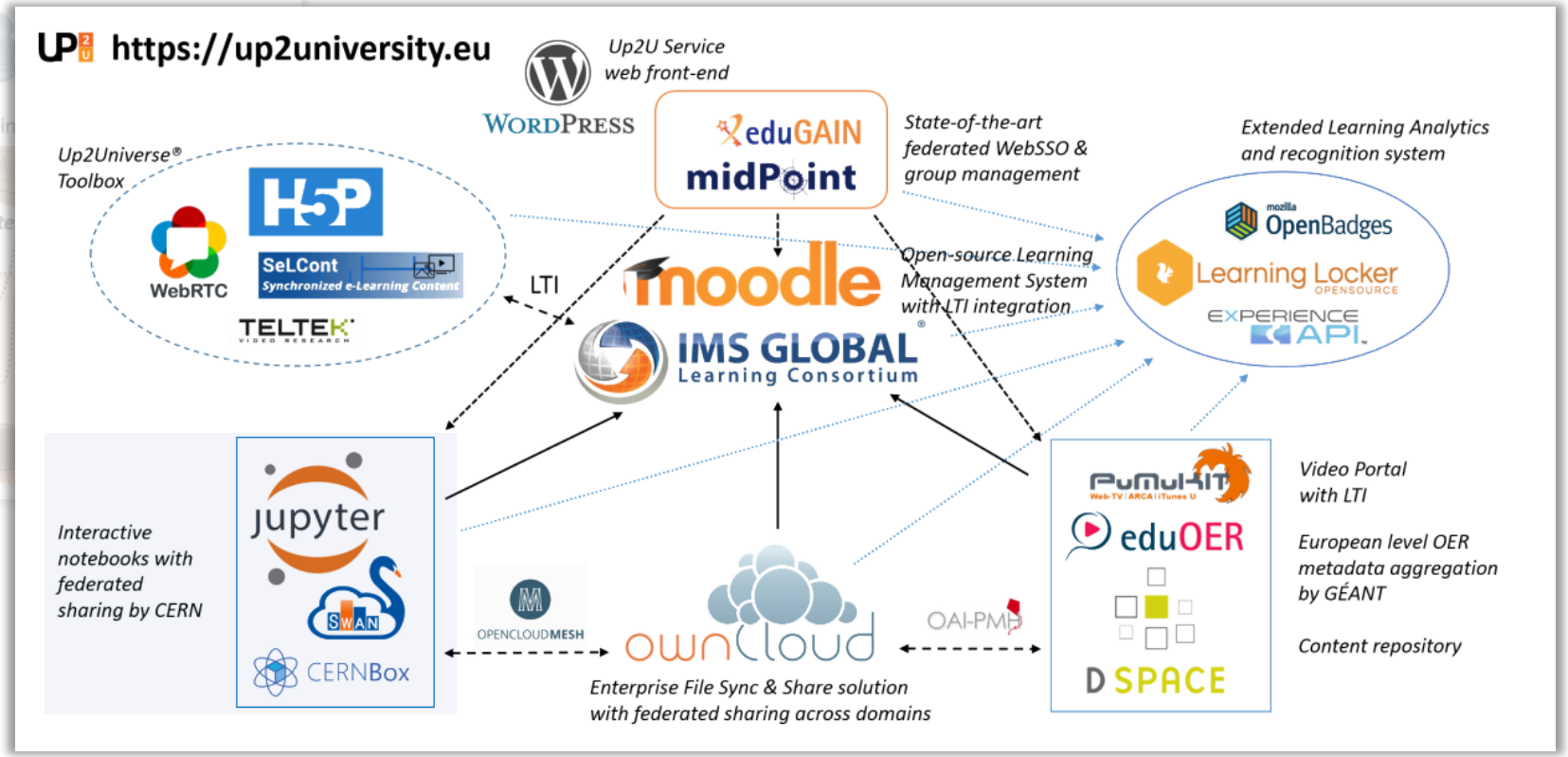
- `docker-compose` on one host
- Static YAMLs and bash wrappers

ScienceBox Heritage



***Early ScienceBox**, c.2017*

- `docker-compose` on one host
- Static YAMLs and bash wrappers



ScienceBox for Up To University (Up2U)

- Kubernetes-managed multi-host clusters
- Integration with Up2U IDP, Storage Federation
- Still plenty of bash and workarounds

ScienceBox Heritage

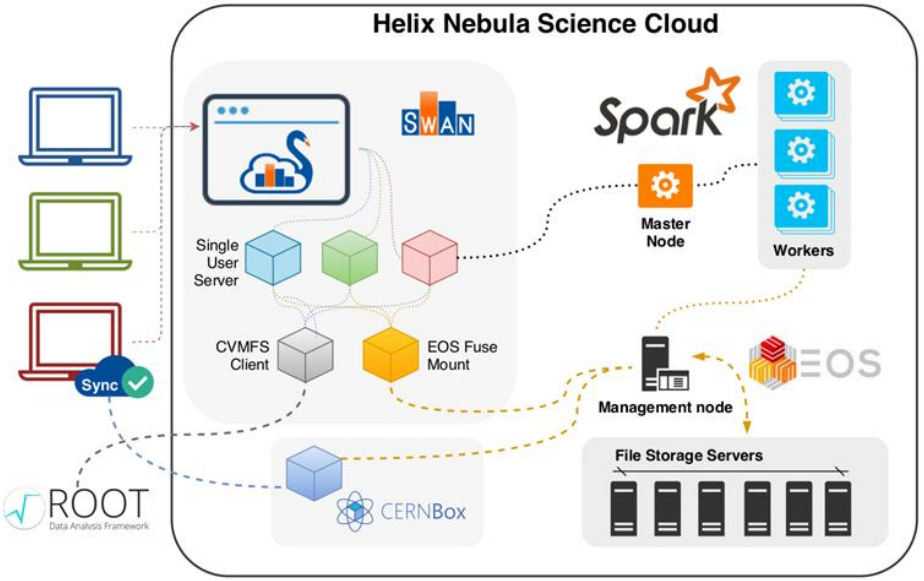
***ScienceBox for HEP**, c.2018*

- K8s on Helix Nebula Science Cloud
- 2000+ CPUs, 10+ TB memory
- HEP Physics analysis by TOTEM

2018 IEEE/ACM International Conference on Utility and Cloud Computing Companion (UCC Companion)

Big Data Tools and Cloud Services for High Energy Physics Analysis in TOTEM Experiment

Valentina Avati†, Milosz Blaszkiewicz†, Enrico Bocchi*, Luca Canali*, Diogo Castro*, Javier Cervantes*, Leszek Grzanka†, Enrico Guiraud*, Jan Kaspar*, Prasanth Kothuri*, Massimo Lamanna*, Maciej Malawski†, Aleksandra Mnich†, Jakub Moscicki*, Shravan Murali*, Danilo Piparo*, Enric Tejedor*
†AGH Universtity of Science and Technology, Krakow, Poland, Email: {grzanka,malawski}@agh.edu.pl
* CERN, CH-1211 Geneva 23, Switzerland, Email: {first.last}@cern.ch



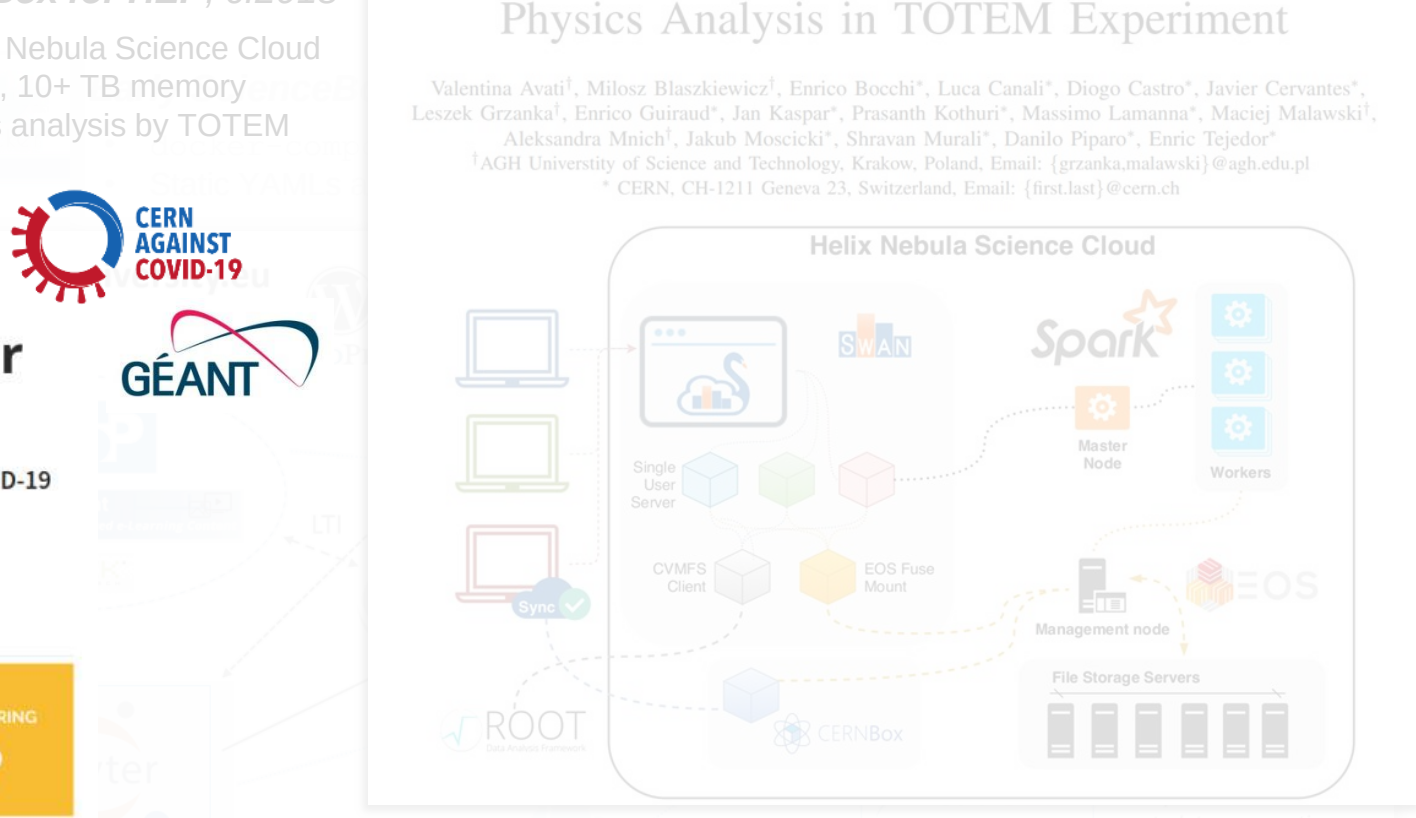
ScienceBox for Up To University (Up2U)

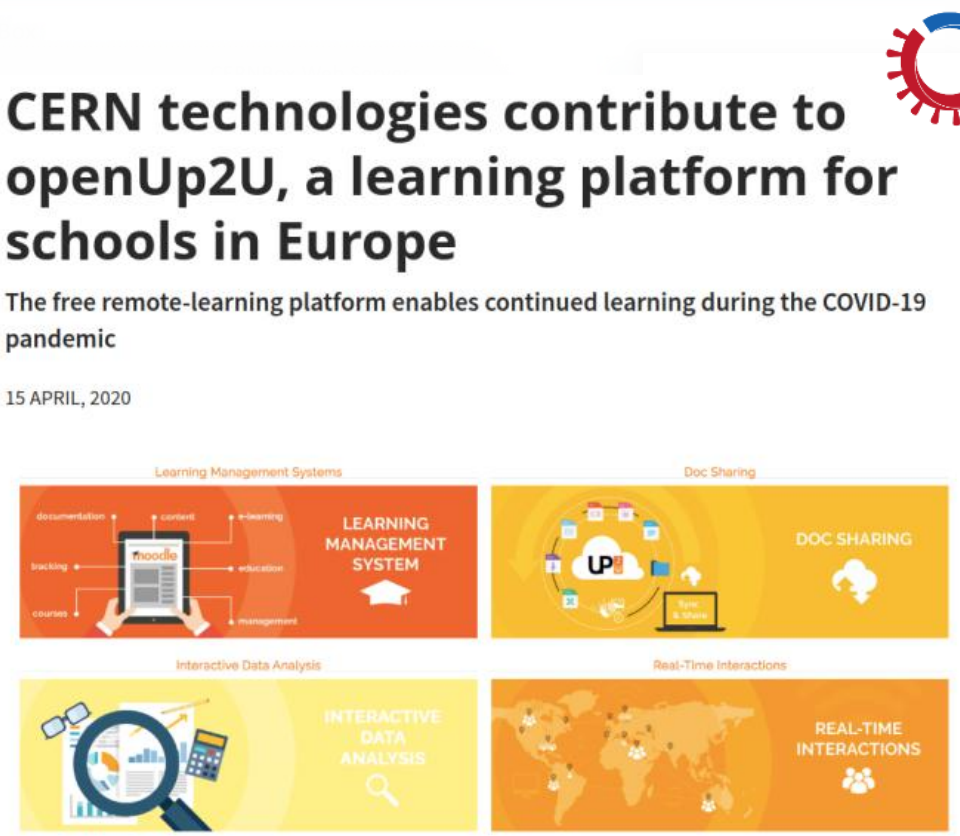
- Kubernetes-managed multi-host clusters
- Integration with Up2U IDP, Storage Federation
- Still plenty of bash and workarounds

ScienceBox Heritage

***ScienceBox for HEP**, c.2018*

- K8s on Helix Nebula Science Cloud
- 2000+ CPUs, 10+ TB memory
- HEP Physics analysis by TOTEM





CERN AGAINST COVID-19

GÉANT

***ScienceBox as Free Learning Platform**, 2020*

- Free-access remote-learning platforms for EU students
- Hosted on AWS, funded by GÉANT

Why ScienceBox 2.0?

1. Use modern, widely-adopted container technologies
2. Improve maintainability and modularity
3. Ease contributions to the package

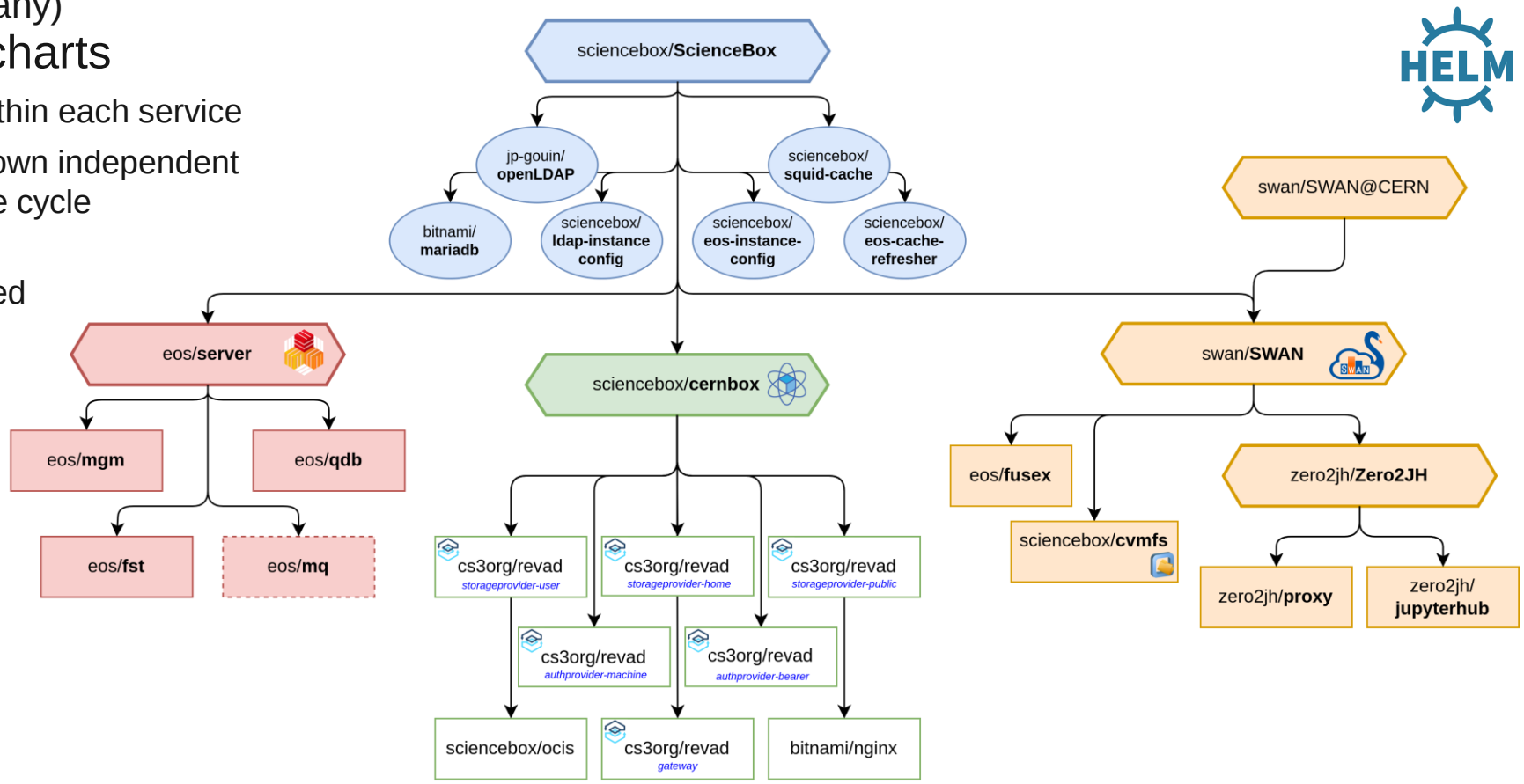
Why ScienceBox 2.0?

1. Use modern, widely-adopted container technologies
2. Improve maintainability and modularity
3. Ease contributions to the package

- ScienceBox re-architected as a collection of **Helm Charts**
  - Re-use (where available) existing charts from EOS, CERNBox, SWAN, CVMFS
  - Add the glue for stand-alone deployments (LDAP, Identity Provider, many configuration bits, …)
  - Allow for deployment of single components and addition of new ones
    - Disposable service for testing, debugging/development
    - Integration with external services // Onboarding of new services through Helm charts

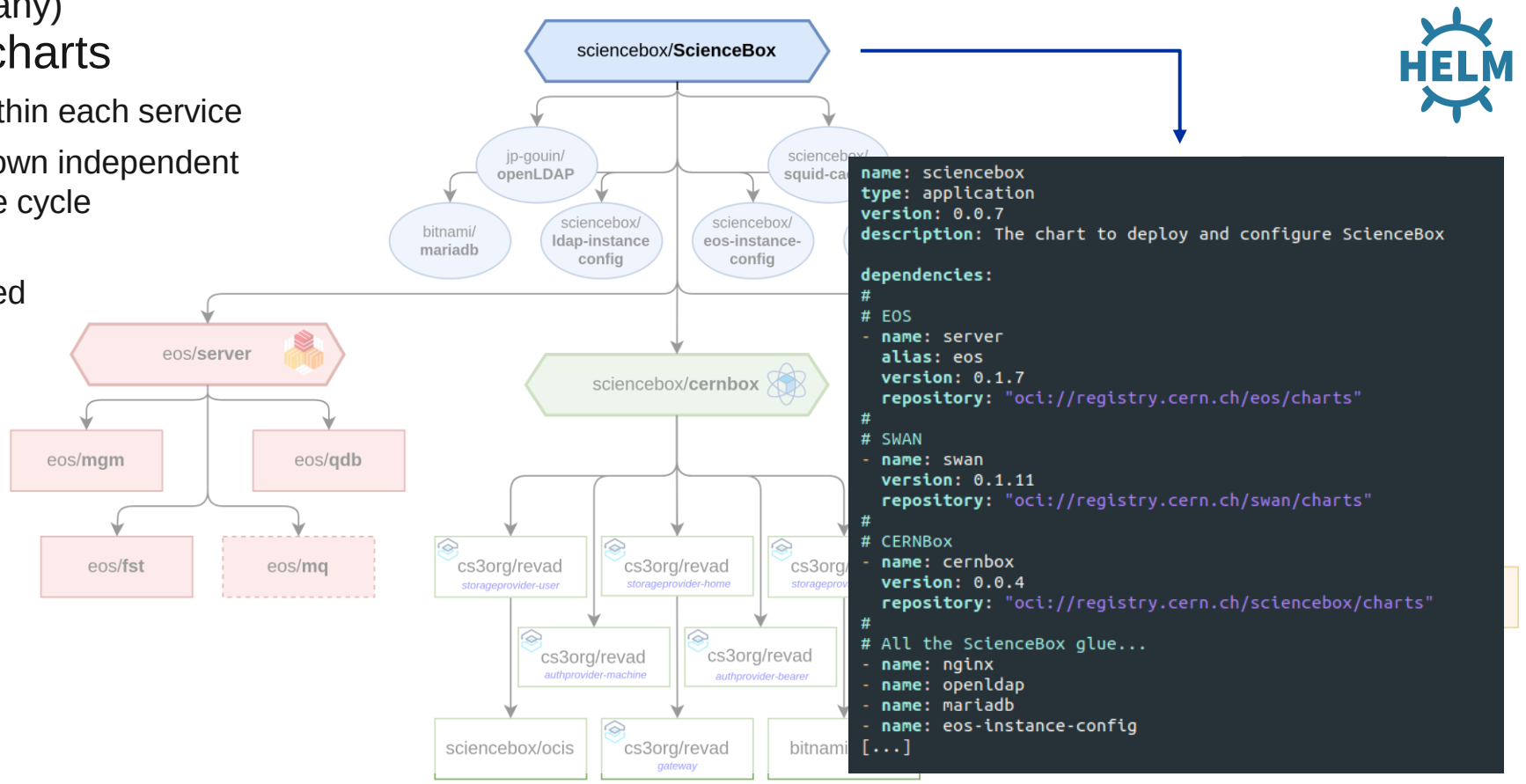
Renewed Architecture

- ScienceBox is root to (many) single and umbrella charts
  - Confine internal complexity within each service
  - Each service (and chart) has own independent development and release cycle
  - ScienceBox updates dependencies on the need



Renewed Architecture

- ScienceBox is root to (many) single and umbrella charts
  - Confine internal complexity within each service
  - Each service (and chart) has own independent development and release cycle
  - ScienceBox updates dependencies on the need



mboxed: Demonstrator in minikube

- **How to get started?**

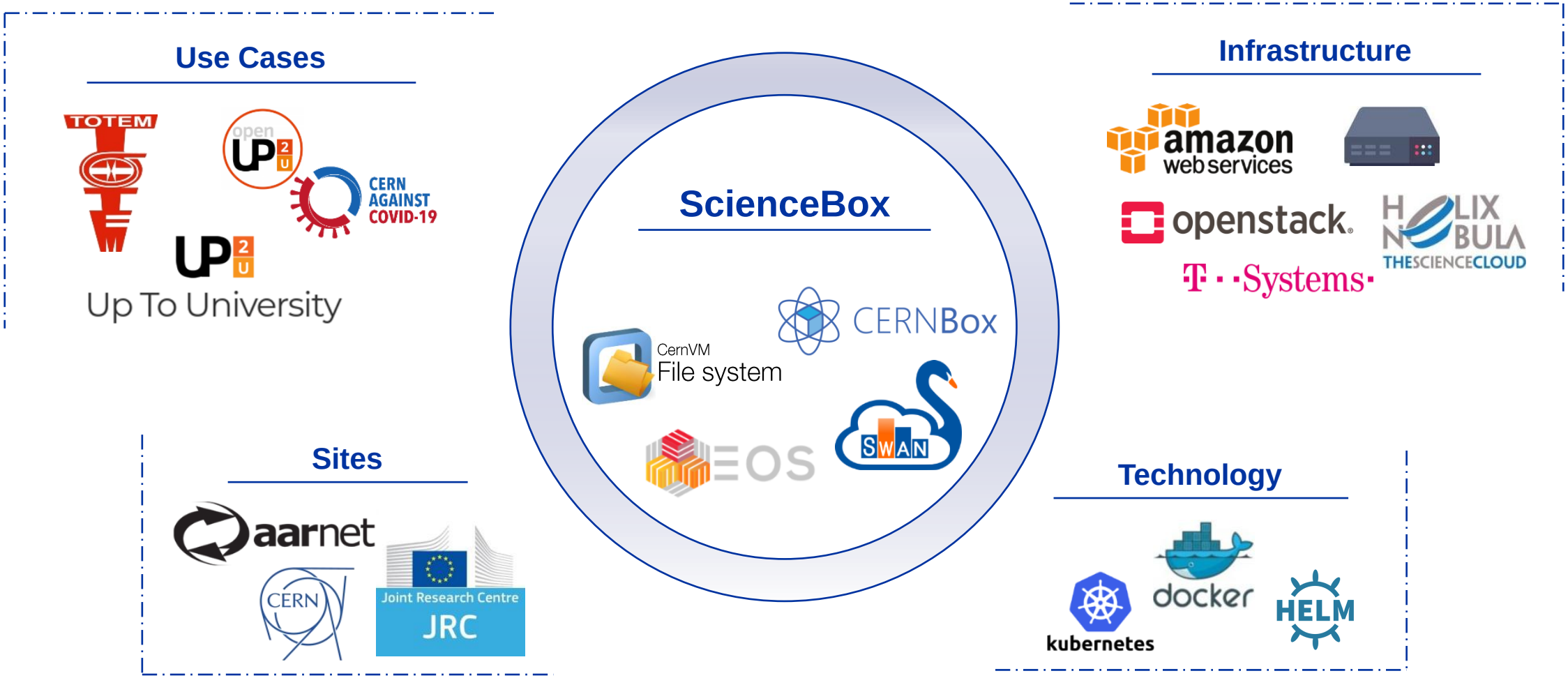
```
[root@mboxed-demo-centos ~]# git clone https://github.com/sciencebox/mboxed.git && cd mboxed
[...]
[root@mboxed-demo-centos ~]# ./SetupInstall.sh
[...]
[root@mboxed-demo-centos ~]# ./ScienceBox.sh
[...]
Release "sciencebox" does not exist. Installing it now.
NAME: sciencebox
LAST DEPLOYED: Thu Apr 27 13:56:55 2023
NAMESPACE: default
STATUS: deployed
REVISION: 1
TEST SUITE: None

ScienceBox is being installed!
  Now it is a good time to grab a coffee...

  The deployment should be ready in a few minutes.
  You can check the containers status by typing `kubectl get pods`.

  Once all the containers are running (or completed), you can reach ScienceBox from your browser at
  https://mboxed-demo-centos.cern.ch/sciencebox
```

mboxed: Demonstrator in minikube

Time for a demo

youtu.be/knzfCZPrZUE

Wrapping Up


Use Cases
TOTEM
open UP 2 U
CERN AGAINST COVID-19
UP 2 U
Up To University
ScienceBox
CernVM File system
CERNBox
EOS
SWAN
Infrastructure
amazon web services
openstack
HELIX NEBULA THESCIENCECLOUD
T-Systems
Sites
aarnet
CERN
Joint Research Centre JRC
Technology
kubernetes
docker
HELM

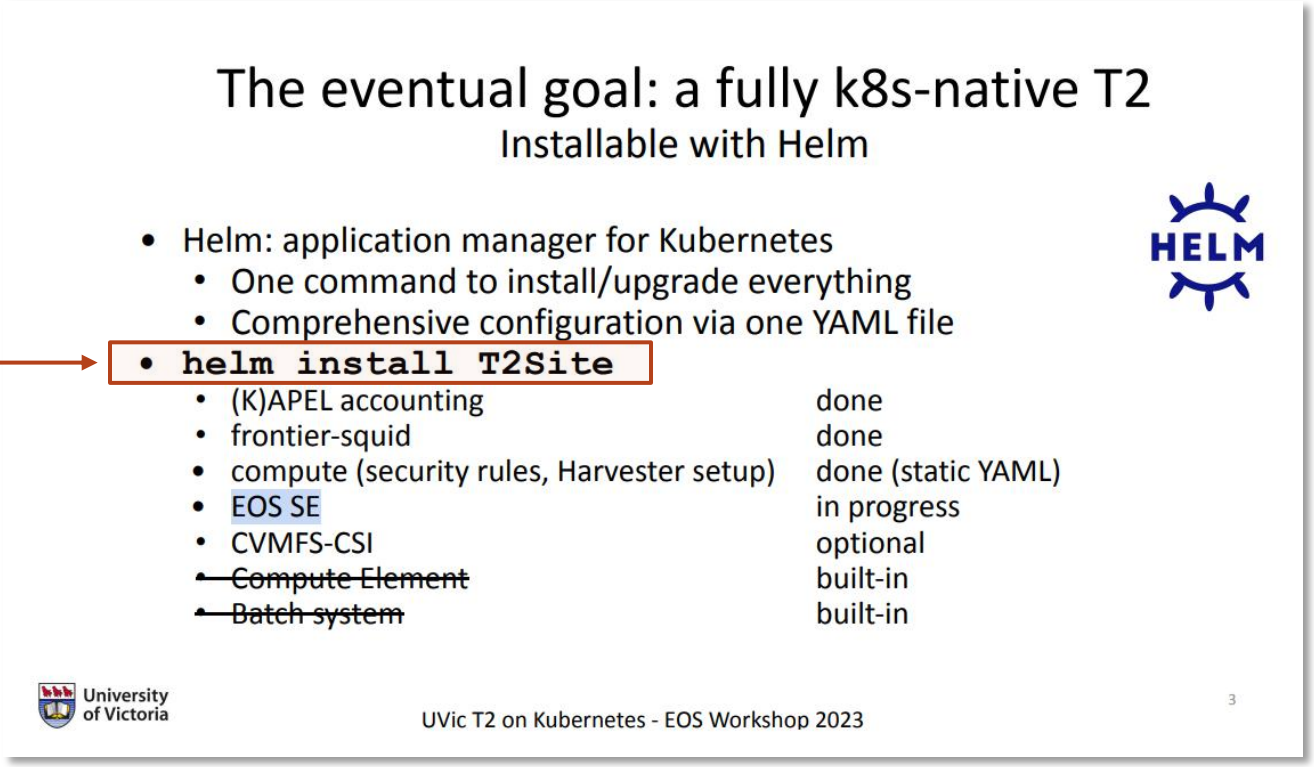
Coming Up Next: ScienceBox for OnBoard

onboard.app.cern.ch

Coming Up Next: ScienceBox for Tier 2 sites

- **ScienceBox for WLCG T2s**
  - Deploy a subset of components for compute workloads

Credits to Ryan Taylor, University of Victoria, Canada

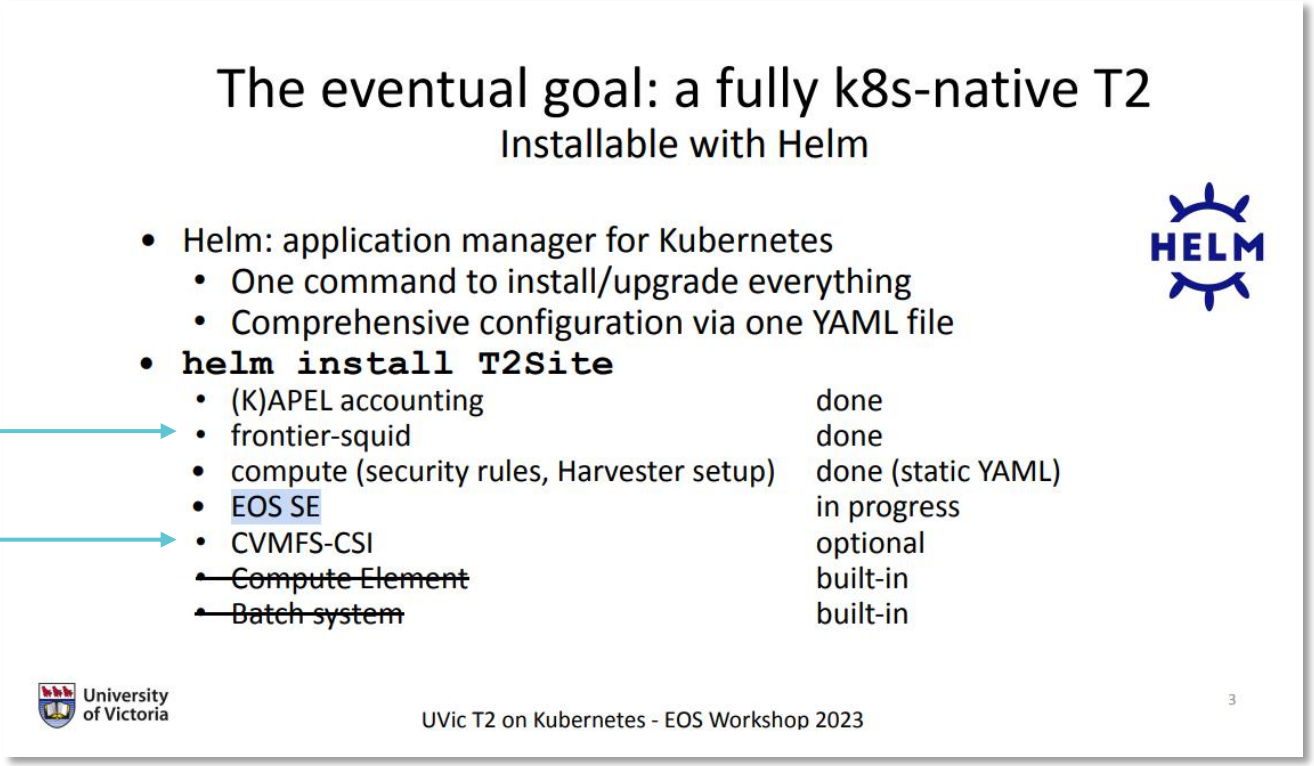


Coming Up Next: ScienceBox for Tier 2 sites

- **ScienceBox for WLCG T2s**
  - Deploy a subset of components for compute workloads

1. Ubiquitous `/cvmfs`

Credits to Ryan Taylor, University of Victoria, Canada

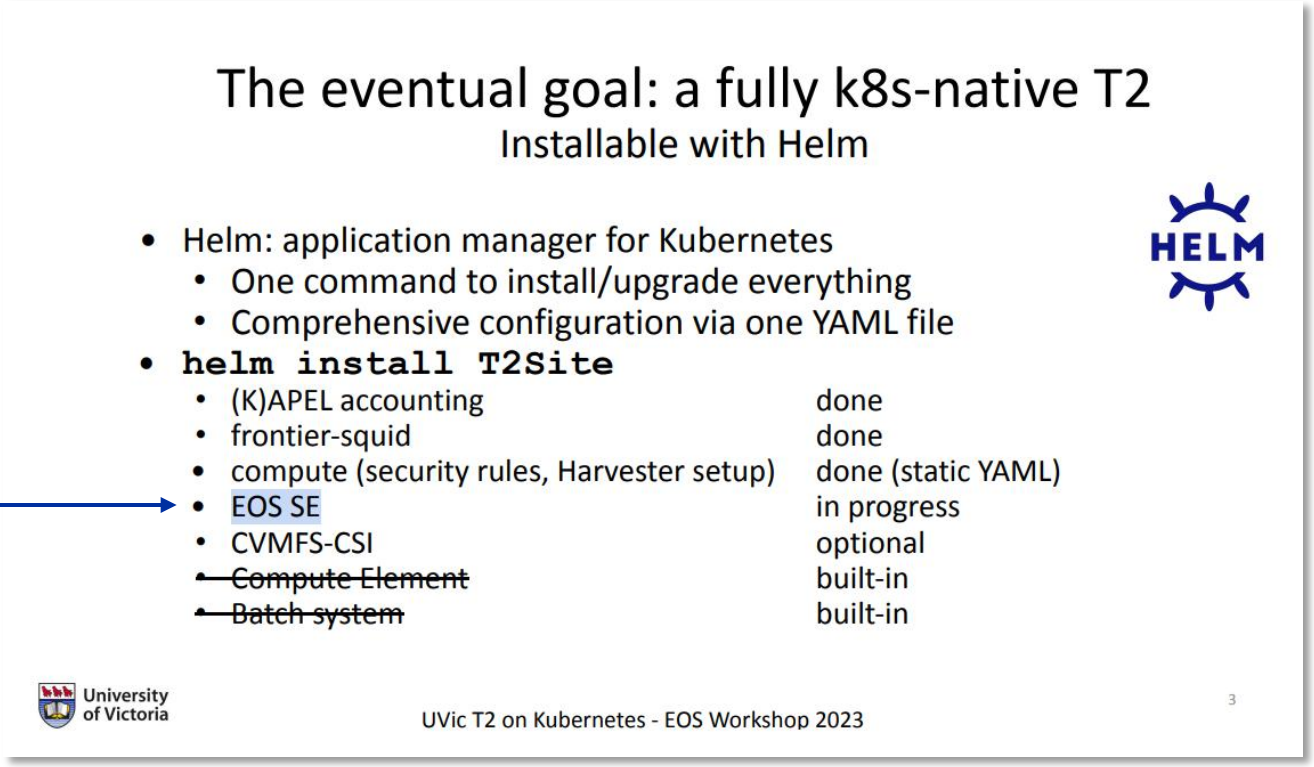


Coming Up Next: ScienceBox for Tier 2 sites

- **ScienceBox for WLCG T2s**
  - Deploy a subset of components for compute workloads

1. Ubiquitous `/cvmfs`
2. Virtualized EOS on top of existing storage (e.g., Ceph)

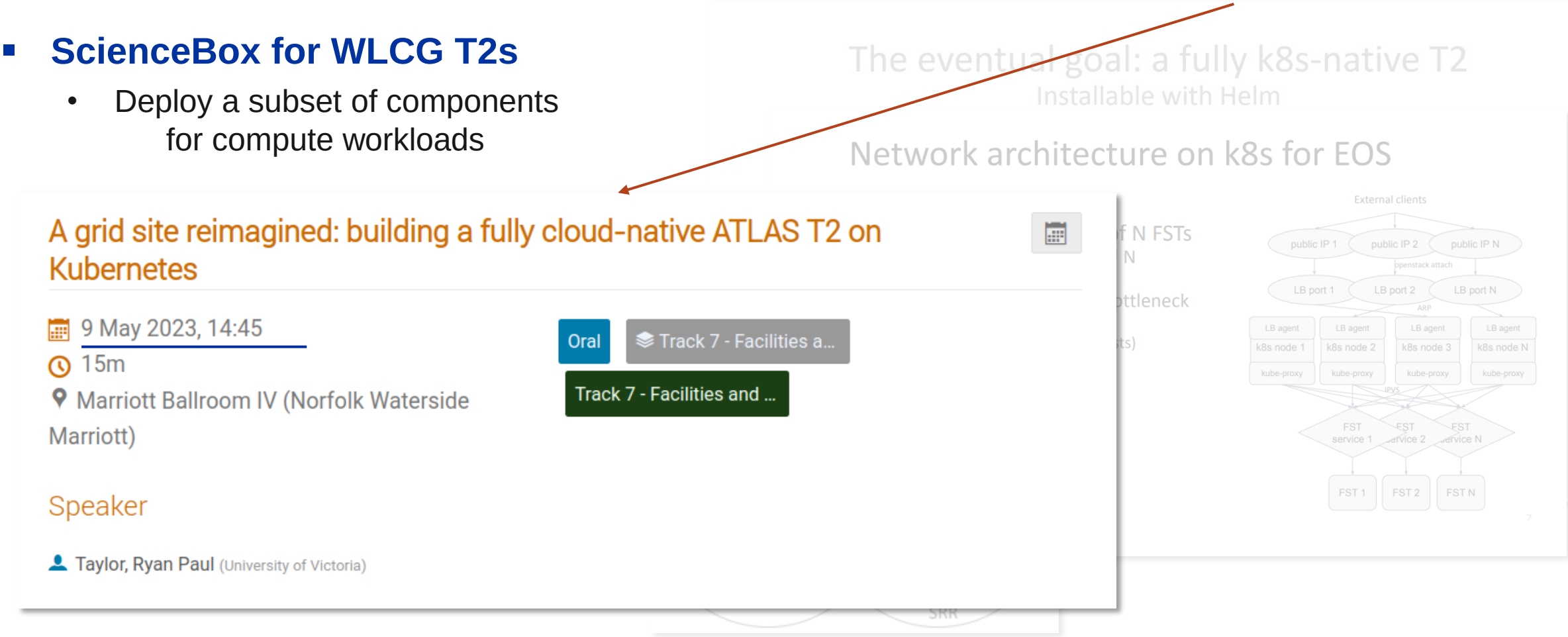
Credits to Ryan Taylor, University of Victoria, Canada



Coming Up Next: ScienceBox for Tier 2 sites

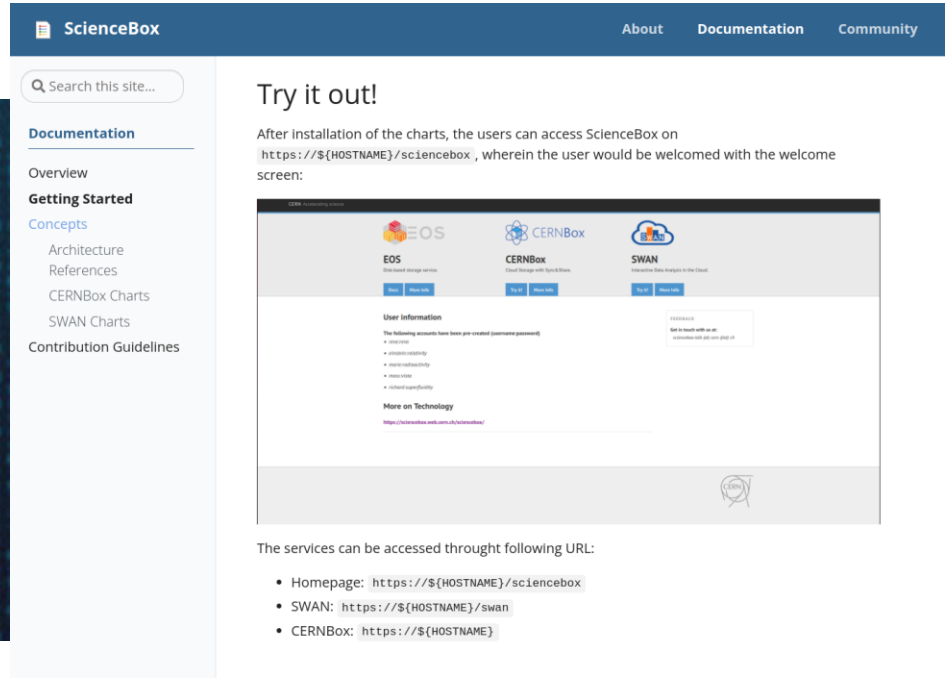
Credits to Ryan Taylor, University of Victoria, Canada

- **ScienceBox for WLCG T2s**
  - Deploy a subset of components for compute workloads



Where to find ScienceBox

- **ScienceBox**
  - Project homepage: sciencebox.web.cern.ch/
  - ArtifactHUB: artifacthub.io/packages/helm/sciencebox/sciencebox
  - New and improved documentation



Where to find ScienceBox

- **ScienceBox**
  - Project homepage: sciencebox.web.cern.ch/
  - ArtifactHUB: artifacthub.io/packages/helm/sciencebox/sciencebox
  - New and improved documentation

- **Code repositories**
  - ScienceBox Organization on GitHub: https://github.com/sciencebox/
  - One-click minikube-based deployment: https://github.com/sciencebox/mboxed

Contributions and Testing Welcome!

- **More on ScienceBox services**
  - {eos, cernbox, swan, cvmfs}.web.cern.ch

Thank you!

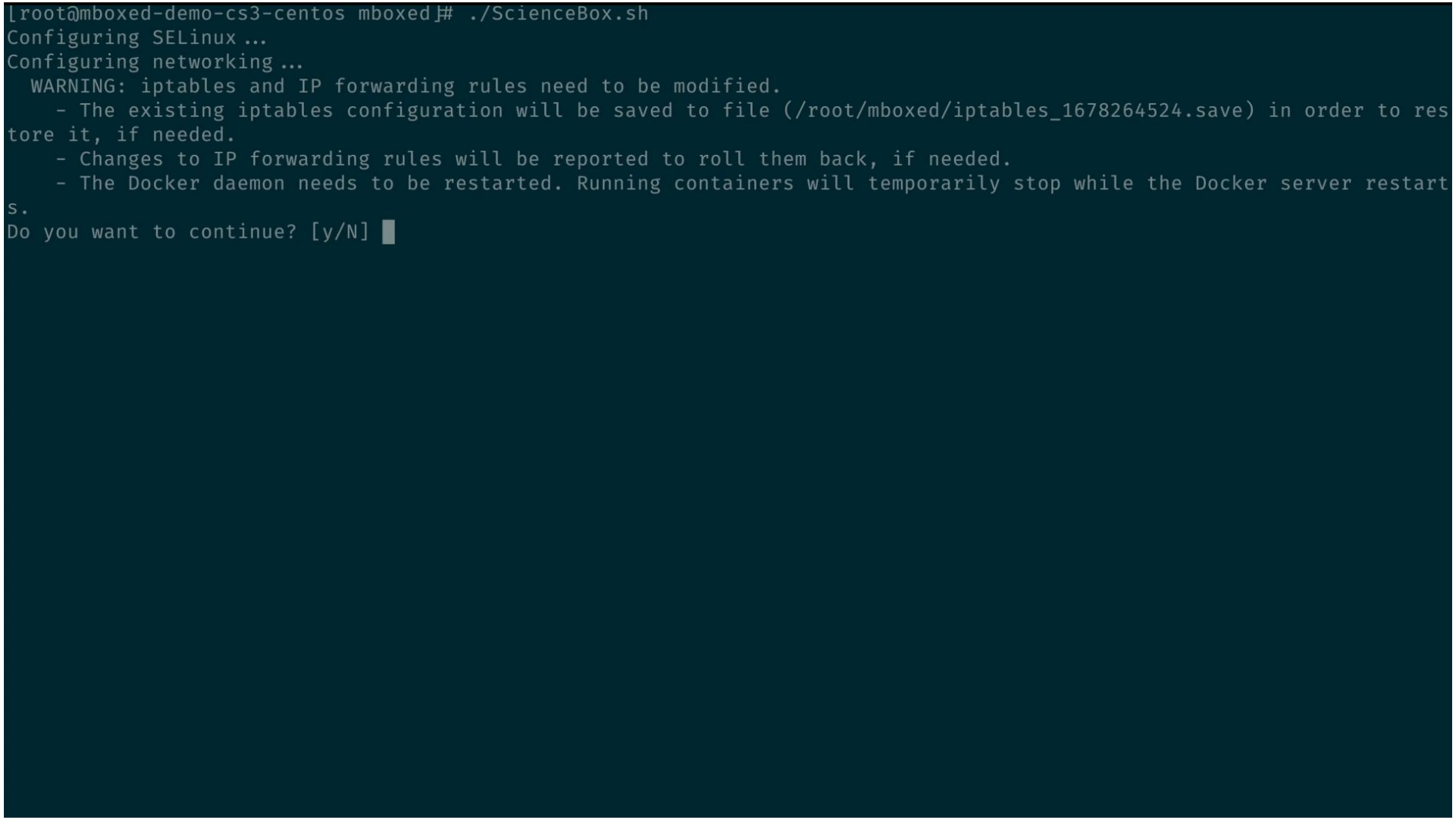
ScienceBox 2.0

Evolving the demonstrator package for CERN Storage and Analysis services

Enrico Bocchi
enrico.bocchi@cern.ch

Video Demo Screenshots

mboxed: Installing ScienceBox



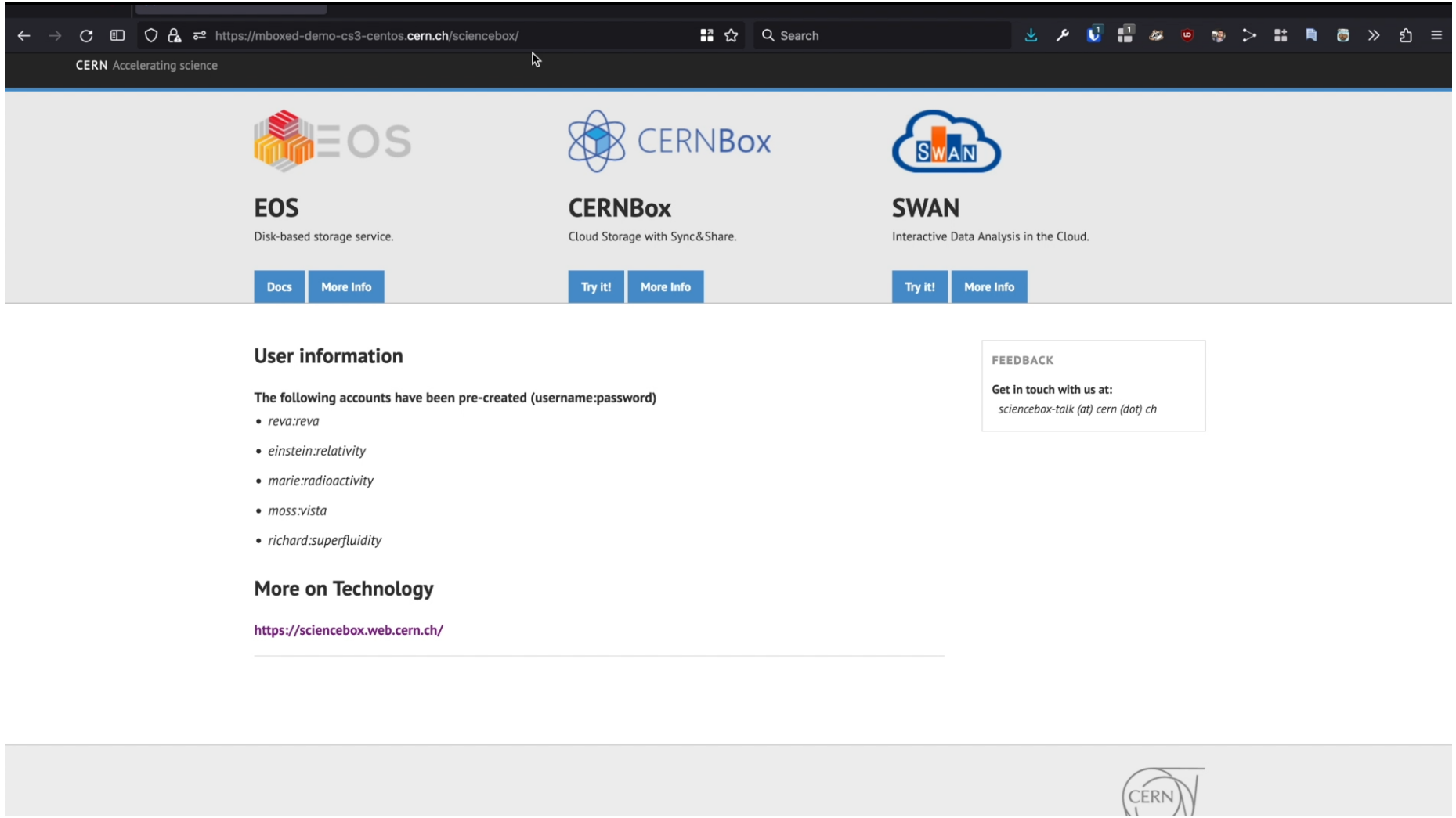
mboxed: Installing ScienceBox

```
    ▪ sudo mv /root/.kube /root/.minikube $HOME
    ▪ sudo chown -R $USER $HOME/.kube $HOME/.minikube

💡  This can also be done automatically by setting the env var CHANGE_MINIKUBE_NONE_USER=true
🔎  Verifying Kubernetes components ...
    ▪ Using image gcr.io/k8s-minikube/storage-provisioner:v5
🌟  Enabled addons: default-storageclass, storage-provisioner
🏄  Done! kubectl is now configured to use "minikube" cluster and "default" namespace by default
Enabling ingress addon ...
    ▪ Using image k8s.gcr.io/ingress-nginx/controller:v1.1.0
    ▪ Using image k8s.gcr.io/ingress-nginx/kube-webhook-certgen:v1.1.1
    ▪ Using image k8s.gcr.io/ingress-nginx/kube-webhook-certgen:v1.1.1
🔎  Verifying ingress addon ...
🌟  The 'ingress' addon is enabled
Waiting for ingress controller to be available ...
Release "sciencebox" does not exist. Installing it now.
NAME: sciencebox
LAST DEPLOYED: Wed Mar  8 09:36:31 2023
NAMESPACE: default
STATUS: deployed
REVISION: 1
TEST SUITE: None


ScienceBox is being installed!
  Now it is a good time to grab a coffee ...

  The deployment should be ready in a few minutes.
  You can check the containers status by typing `kubectl get pods`.

  Once all the containers are running (or completed), you can reach ScienceBox from your browser at
  https://mboxed-demo-cs3-centos.cern.ch/sciencebox
```

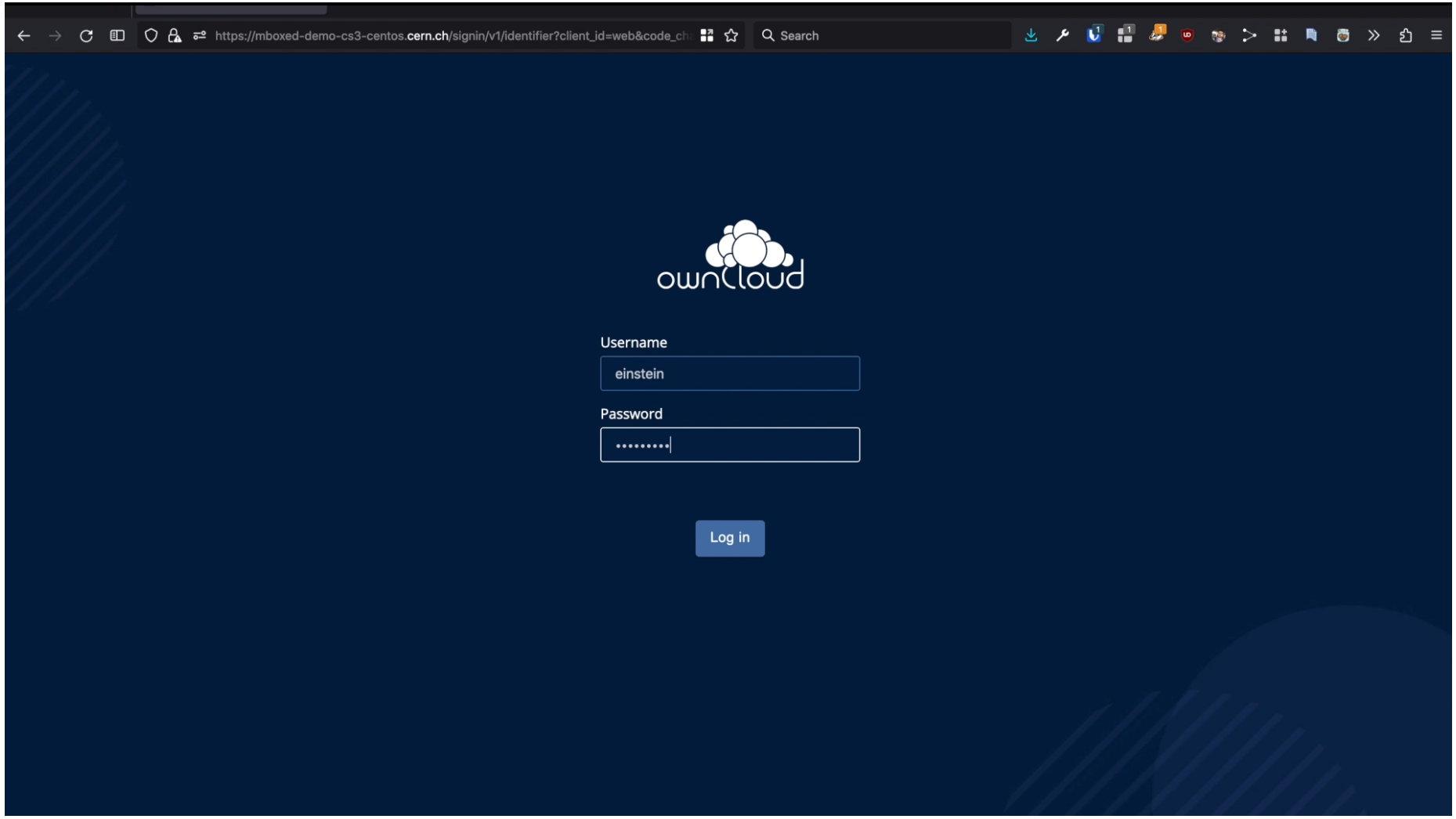
mboxed: Installing ScienceBox

```
[root@mboxed-demo-cs3-centos mboxed]# kubectl get pods
NAME                                          READY   STATUS      RESTARTS   AGE
hub-68d4ddd7d5-f44wm                          1/1     Running     0          9m51s
proxy-7cffc688b8-5gz8z                        1/1     Running     0          9m52s
sciencebox-authproviderbearer-7844ccc5cd-575fx      1/1     Running     0          9m52s
sciencebox-authprovidermachine-74b44dd875-f7qvn     1/1     Running     0          9m52s
sciencebox-cernbox-web-748dc49659-vthkw       1/1     Running     0          9m52s
sciencebox-cvmfs-kkrds                        2/2     Running     0          9m52s
sciencebox-eos-instance-config-rfgtz          0/1     Completed   1          9m52s
sciencebox-eos-user-cache-refresher-lv44c     0/1     Completed   1          9m52s
sciencebox-fst-0                              1/1     Running     0          9m52s
sciencebox-fst-1                              1/1     Running     0          7m40s
sciencebox-fst-2                              1/1     Running     0          7m27s
sciencebox-fst-3                              1/1     Running     0          7m17s
sciencebox-fusex-ddsq7                        1/1     Running     0          9m52s
sciencebox-gateway-8499b95795-md8xj           1/1     Running     0          9m52s
sciencebox-ldap-0                             1/1     Running     1          9m52s
sciencebox-ldap-1                             1/1     Running     1          9m13s
sciencebox-ldap-2                             1/1     Running     1          8m43s
sciencebox-ldap-instance-config-g789d         0/1     Completed   0          9m52s
sciencebox-mariadb-0                          1/1     Running     0          9m52s
sciencebox-mgm-0                              4/4     Running     0          9m52s
sciencebox-nginx-welcome-page-74c8d77495-bxmlq      1/1     Running     0          9m52s
sciencebox-ocis-86687645d8-7cpsr              1/1     Running     0          9m52s
sciencebox-qdb-0                              1/1     Running     0          9m52s
sciencebox-qdb-1                              1/1     Running     0          9m29s
sciencebox-qdb-2                              1/1     Running     0          9m19s
sciencebox-storageproviderhome-8c667d568-h8888      1/1     Running     0          9m52s
sciencebox-storageproviderpublic-6d69947468-cwlfv   1/1     Running     0          9m52s
sciencebox-storageprovideruser-57d4cd9976-57n4t     1/1     Running     0          9m52s
[root@mboxed-demo-cs3-centos mboxed]#
```

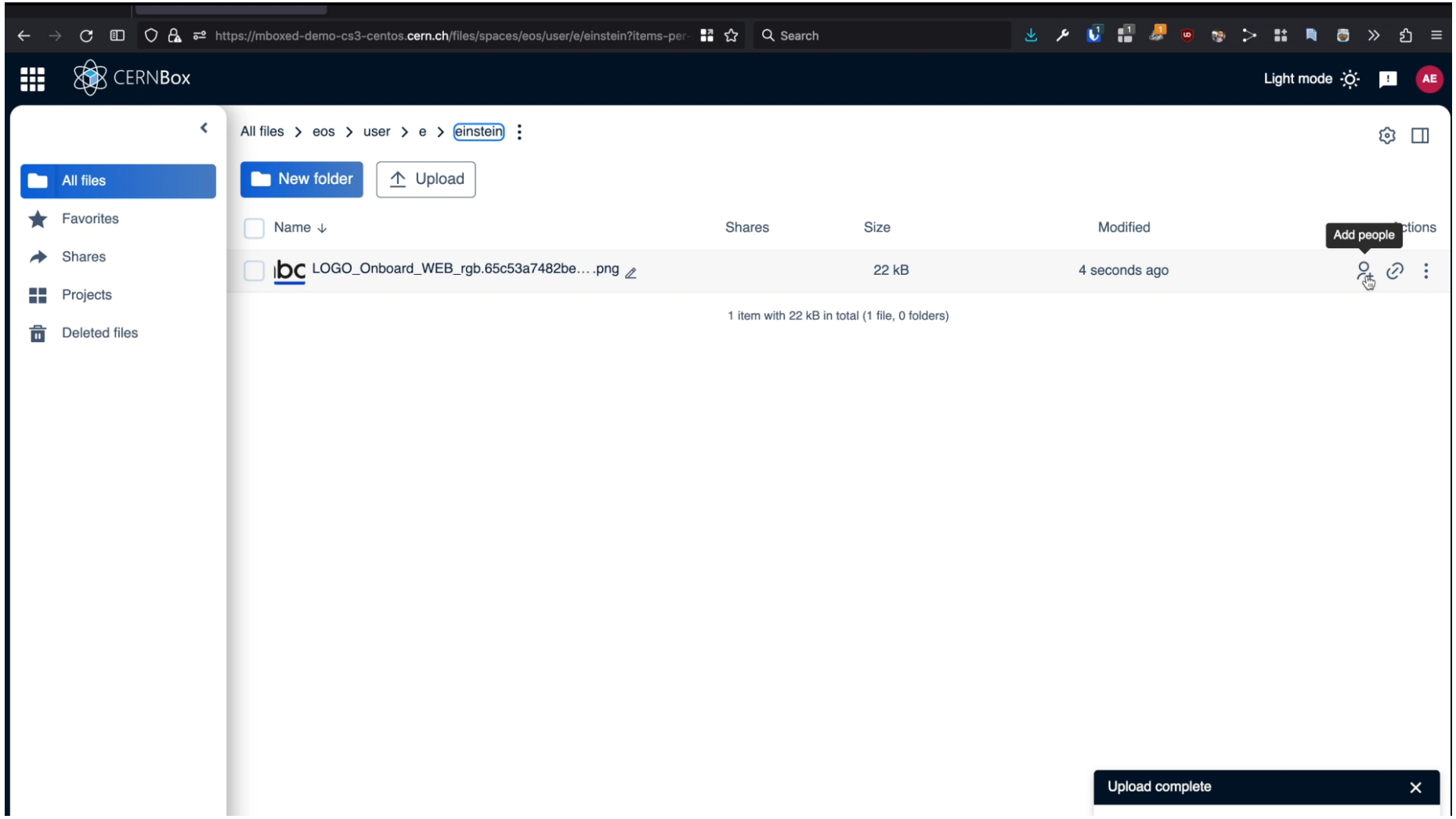
mboxed: Landing Page



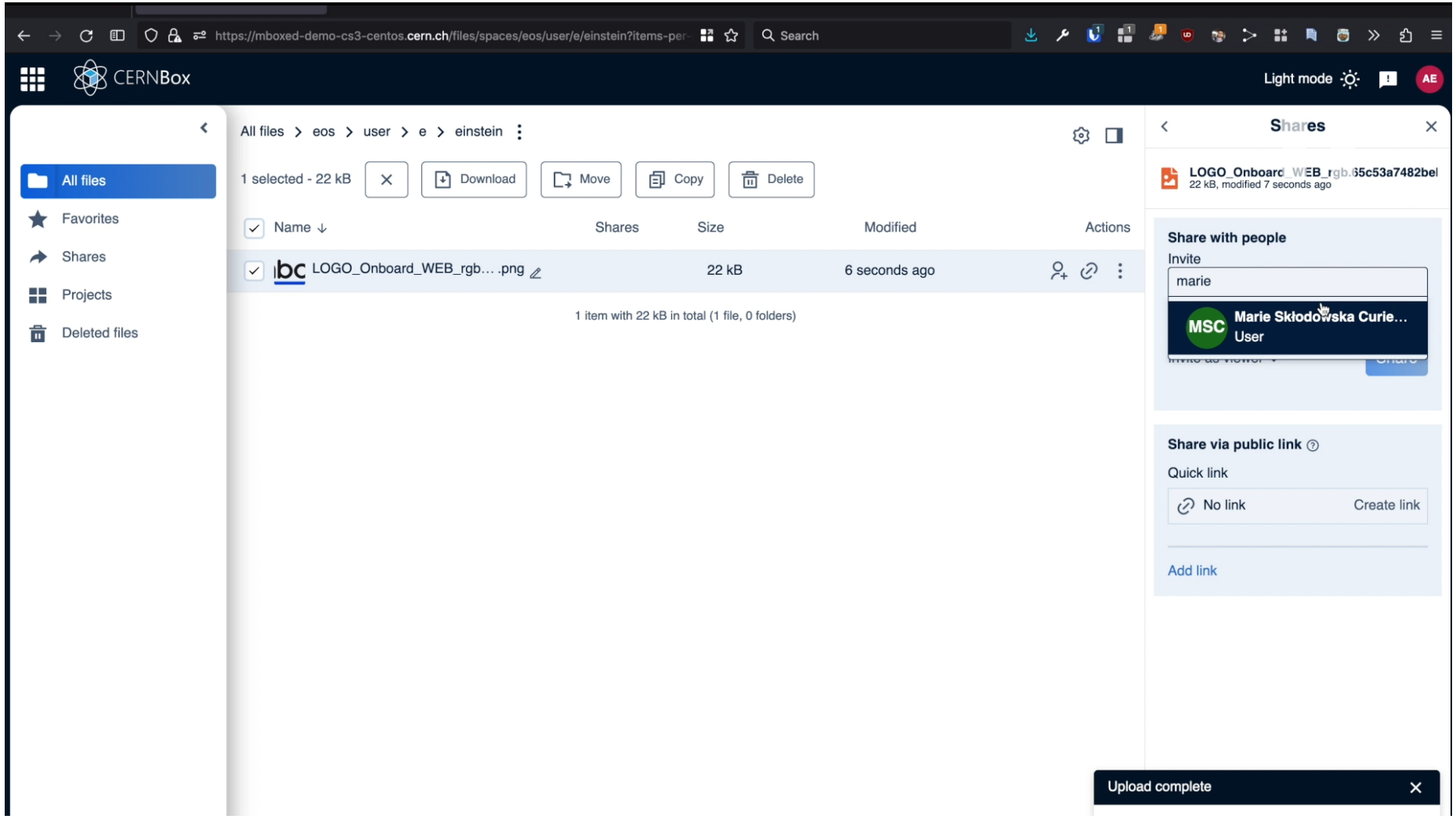
mboxed: Logging in



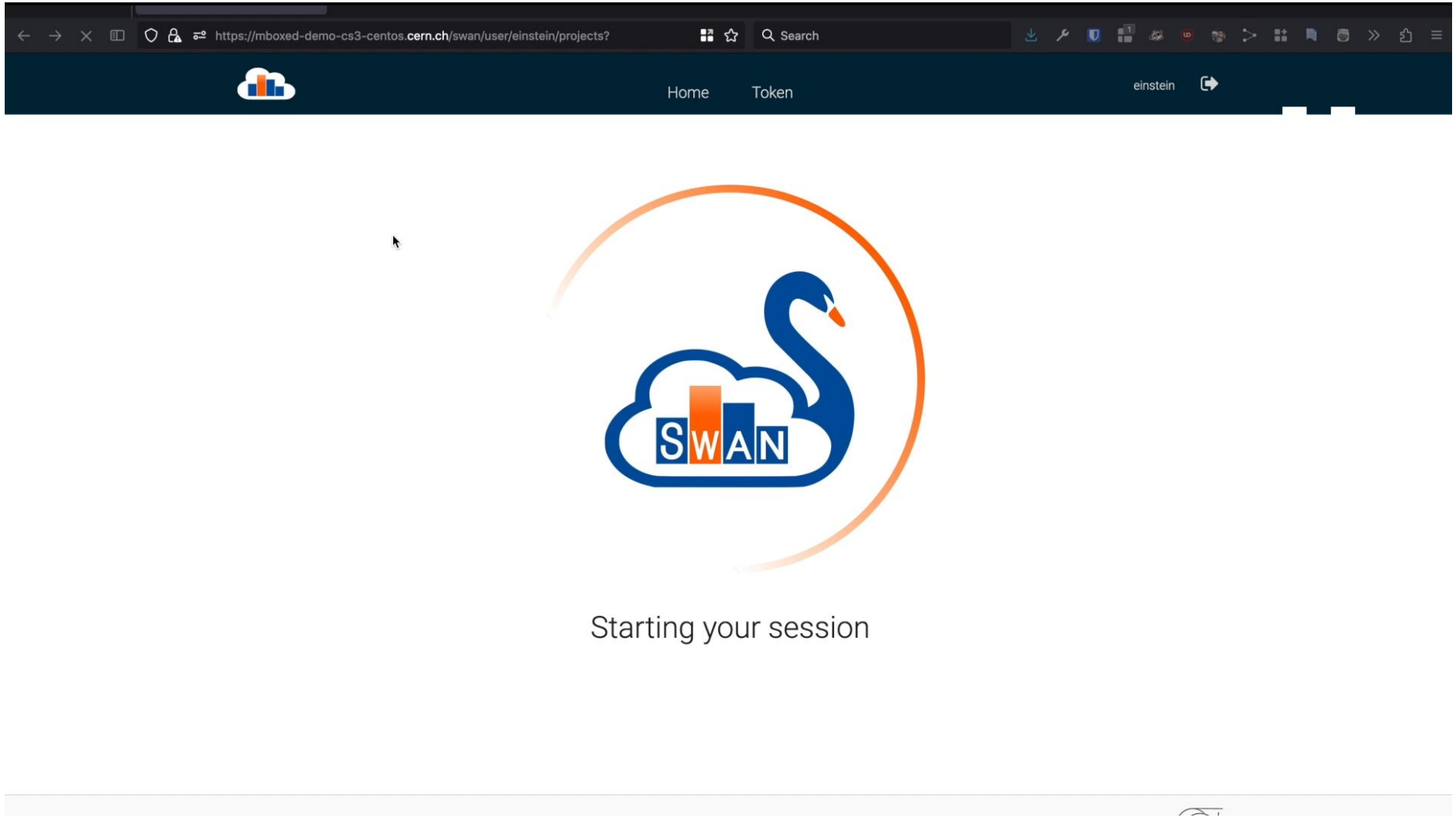
mboxed: Upload and Share



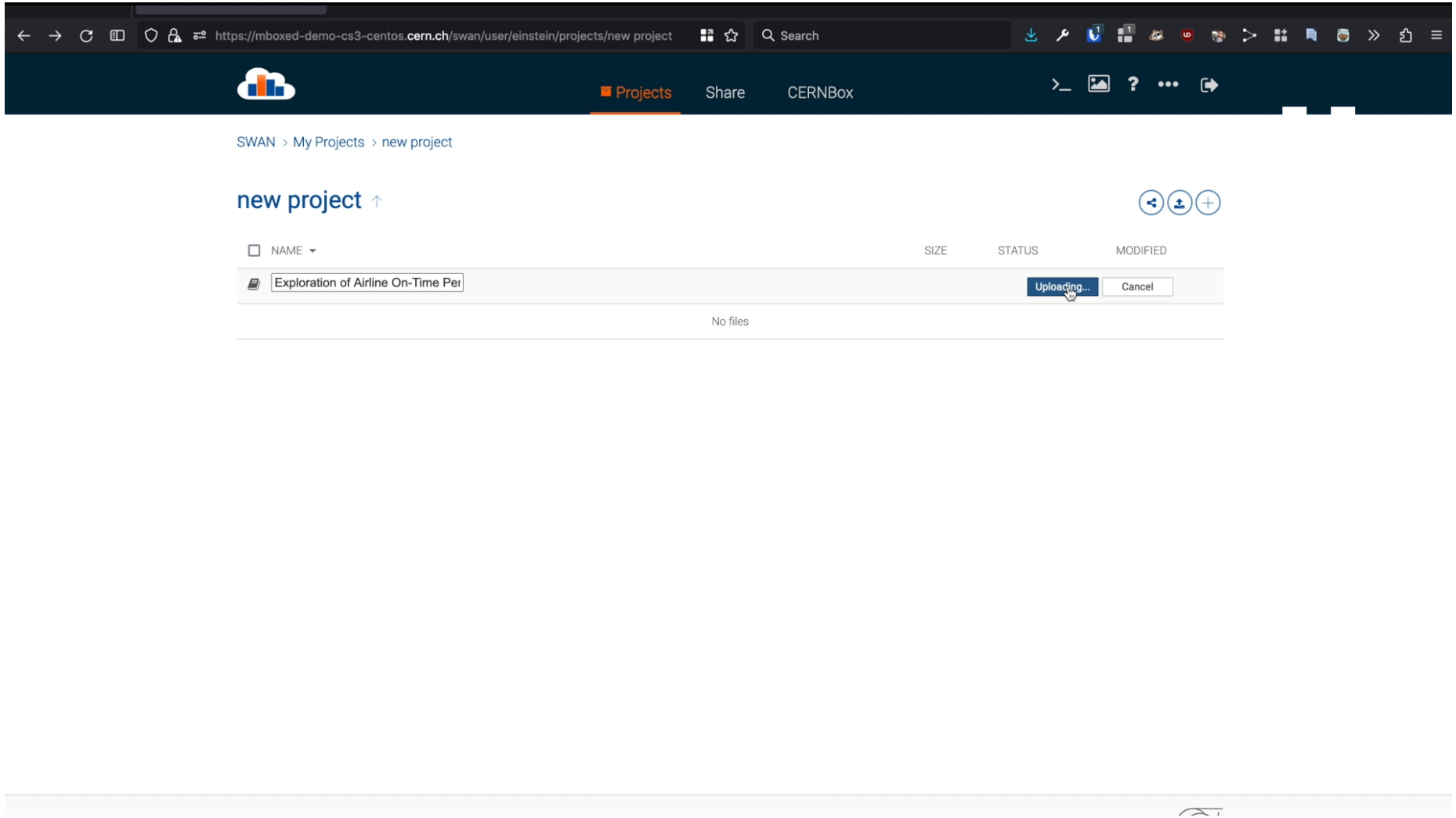
mboxed: Upload and Share



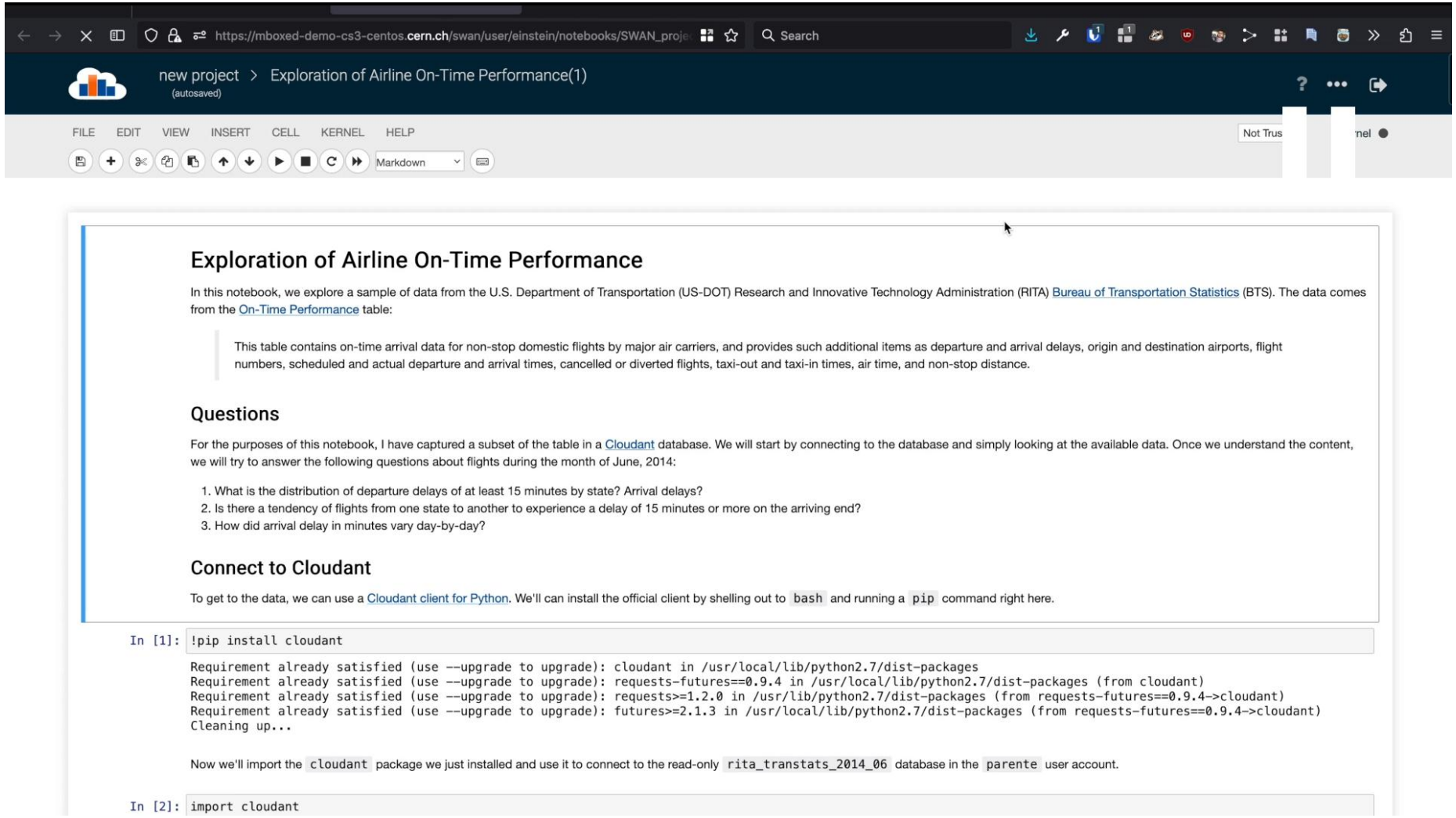
mboxed: Swan


https://mboxed-demo-cs3-centos.cern.ch/swan/user/einstein/projects?
Search
Home
Token
einstein
SWAN
Starting your session

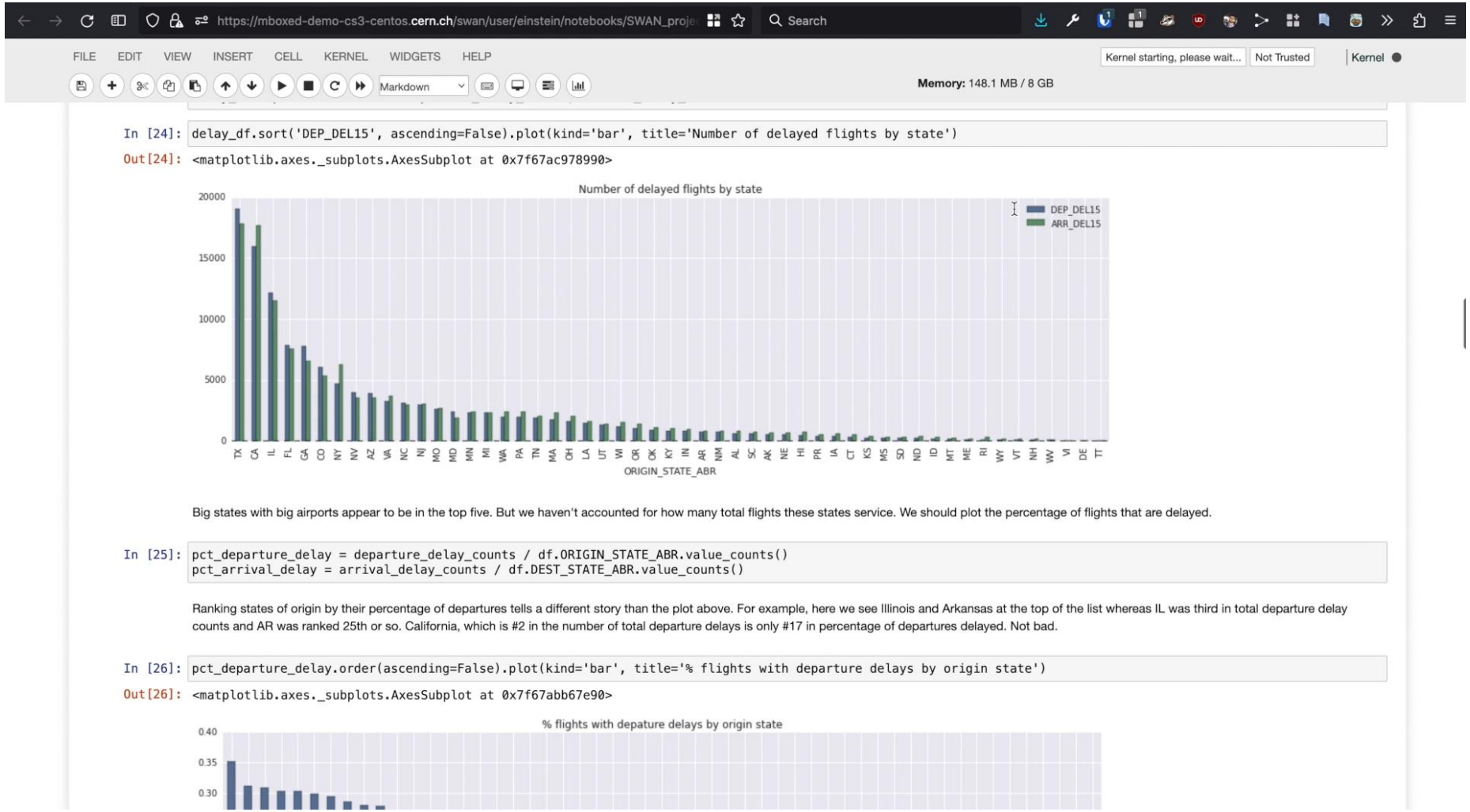
mboxed: Swan


https://mboxed-demo-cs3-centos.cern.ch/swan/user/einstein/projects/new project
Search
Projects
Share
CERNBox
SWAN › My Projects › new project
new project
NAME
SIZE
STATUS
MODIFIED
Exploration of Airline On-Time Per
Uploading...
Cancel
No files

mboxed: Swan Python Notebook



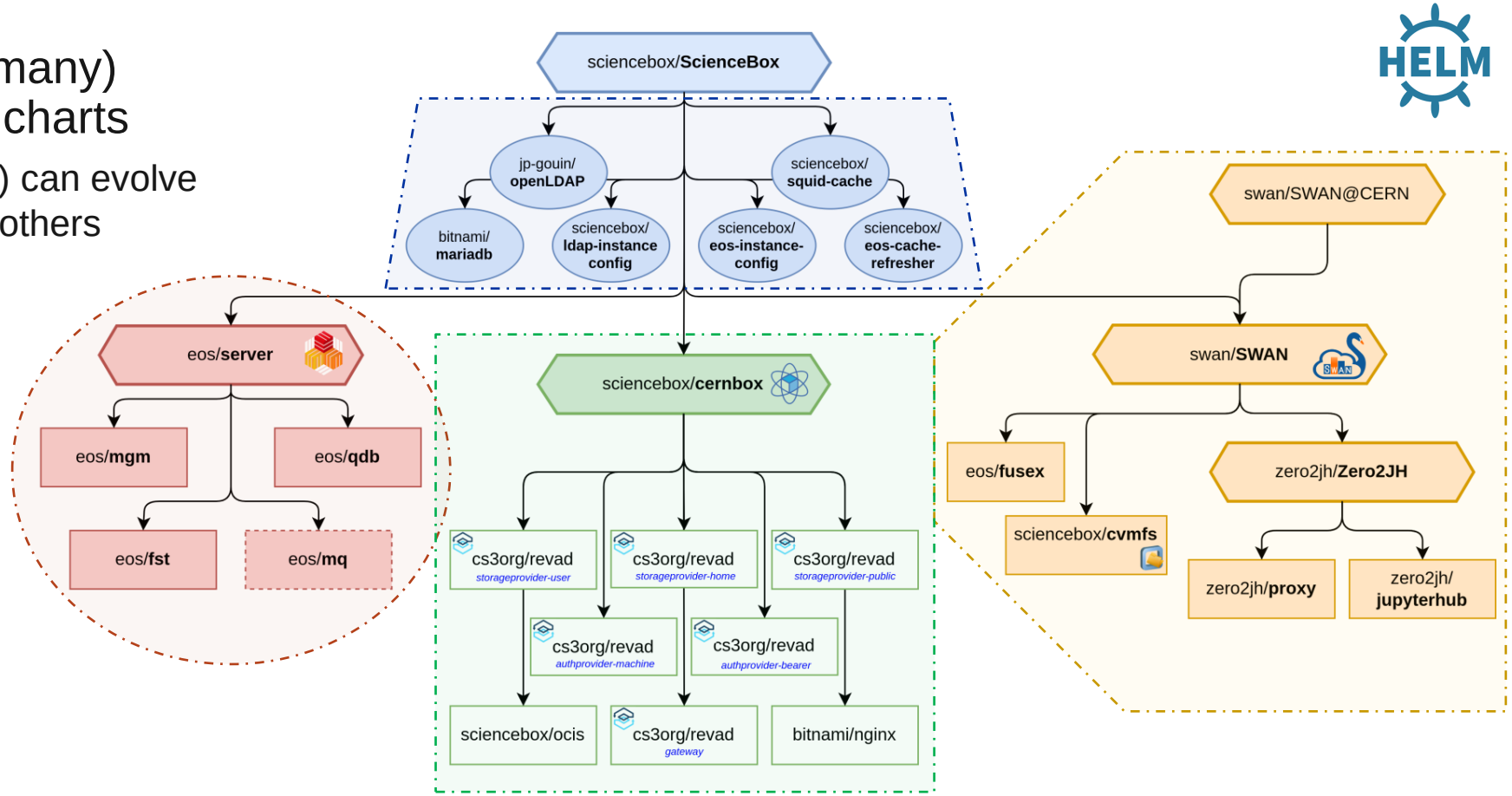
mboxed: Swan Python Notebook



Backup

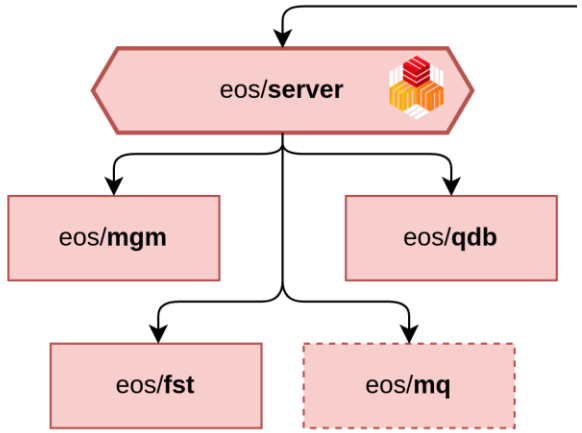
Renewed Architecture

- ScienceBox is root to (many) single and umbrella charts
  - Each service (and chart) can evolve independently on the others



EOS Deployment

```
[root@mboxed-demo-centos ~]# helm install eos oci://registry.cern.ch/eos/charts/server
NAME: eos
LAST DEPLOYED: Thu Apr 27 15:20:43 2023
NAMESPACE: default
STATUS: deployed
REVISION: 1
TEST SUITE: None


[root@mboxed-demo-centos ~]# kubectl get pods
NAME        READY   STATUS    RESTARTS   AGE
eos-fst-0   1/1     Running   0          6m30s
eos-fst-1   1/1     Running   0          4m48s
eos-fst-2   1/1     Running   0          4m38s
eos-fst-3   1/1     Running   0          4m28s
eos-mgm-0   2/2     Running   0          6m30s
eos-qdb-0   1/1     Running   0          6m30s
eos-qdb-1   1/1     Running   0          6m21s
eos-qdb-2   1/1     Running   0          6m10s
```



Coming Up Next: ScienceBox for Tier 2 sites

Credits to Ryan Taylor, University of Victoria, Canada

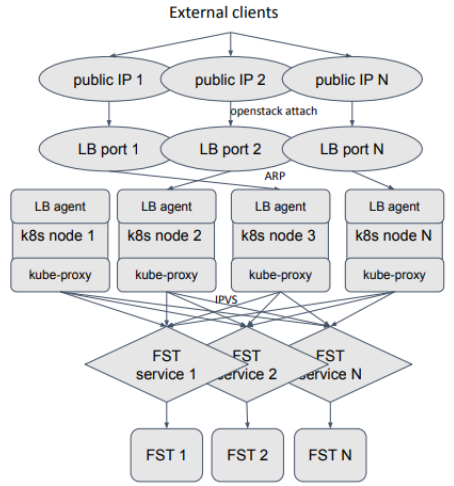
- **ScienceBox for WLCG T2s**
  - Deploy a subset of components for compute workloads
  - Expanding charts to onboard T2 requirements for storage access
    - Multihoming, Third Party Copy
    - Authentication, Grid Certs, …
    - Load-balancing, aggregated throughput

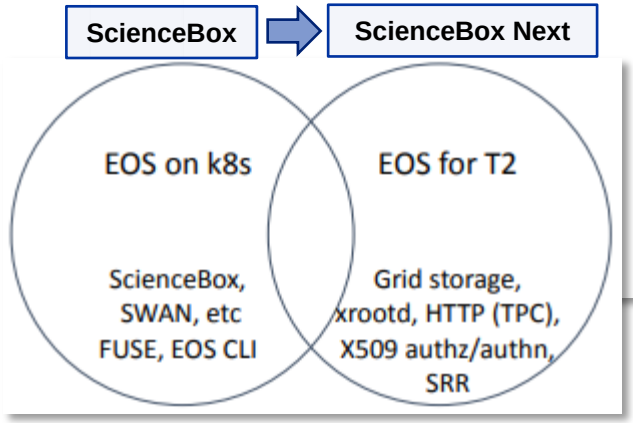
The eventual goal: a fully k8s-native T2

Installable with Helm

Network architecture on k8s for EOS

- One LB service for each of N FSTs
  - Total bandwidth = 1 NIC * N
  - L3 routing: 1 IP per FST
  - Ingress controller not a bottleneck
  - Solves multi-homing
    - With hostAliases (/etc/hosts)





Where to find ScienceBox

- **ScienceBox**
  - Project homepage: sciencebox.web.cern.ch/
  - ArtifactHUB: artifacthub.io/packages/helm/sciencebox/sciencebox
  - Mailing list: sciencebox-talk@cern.ch

- **Code repositories**
  - ScienceBox Organization on GitHub: https://github.com/sciencebox/
  - One-click minikube-based deployment: https://github.com/sciencebox/mboxed

- **More on ScienceBox services**
  - {eos, cernbox, swan, cvmfs}.web.cern.ch

Contributions and Testing Welcome!