

Norfolk, Virginia, USA • May 8-13, 2023

CHEP 2023

Computing in High Energy & Nuclear Physics



ScienceBox 2.0

Evolving the demonstrator package for
CERN Storage and Analysis services

May 9, 2023

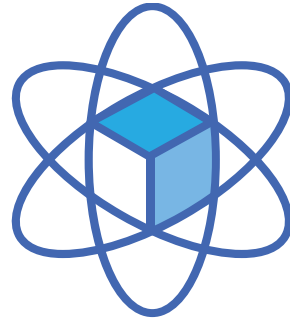
Samuel Alfageme
Enrico Bocchi
Jimil Desai

What is ScienceBox?

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



EOS



CERNBox



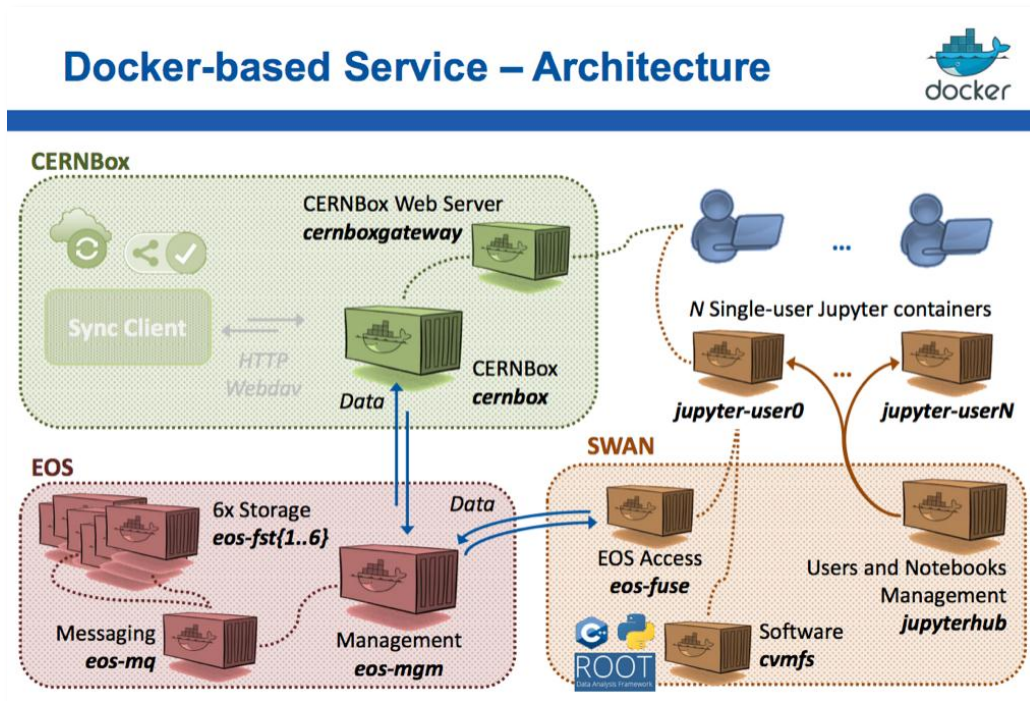
SWAN



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

Docker-based Service – Architecture

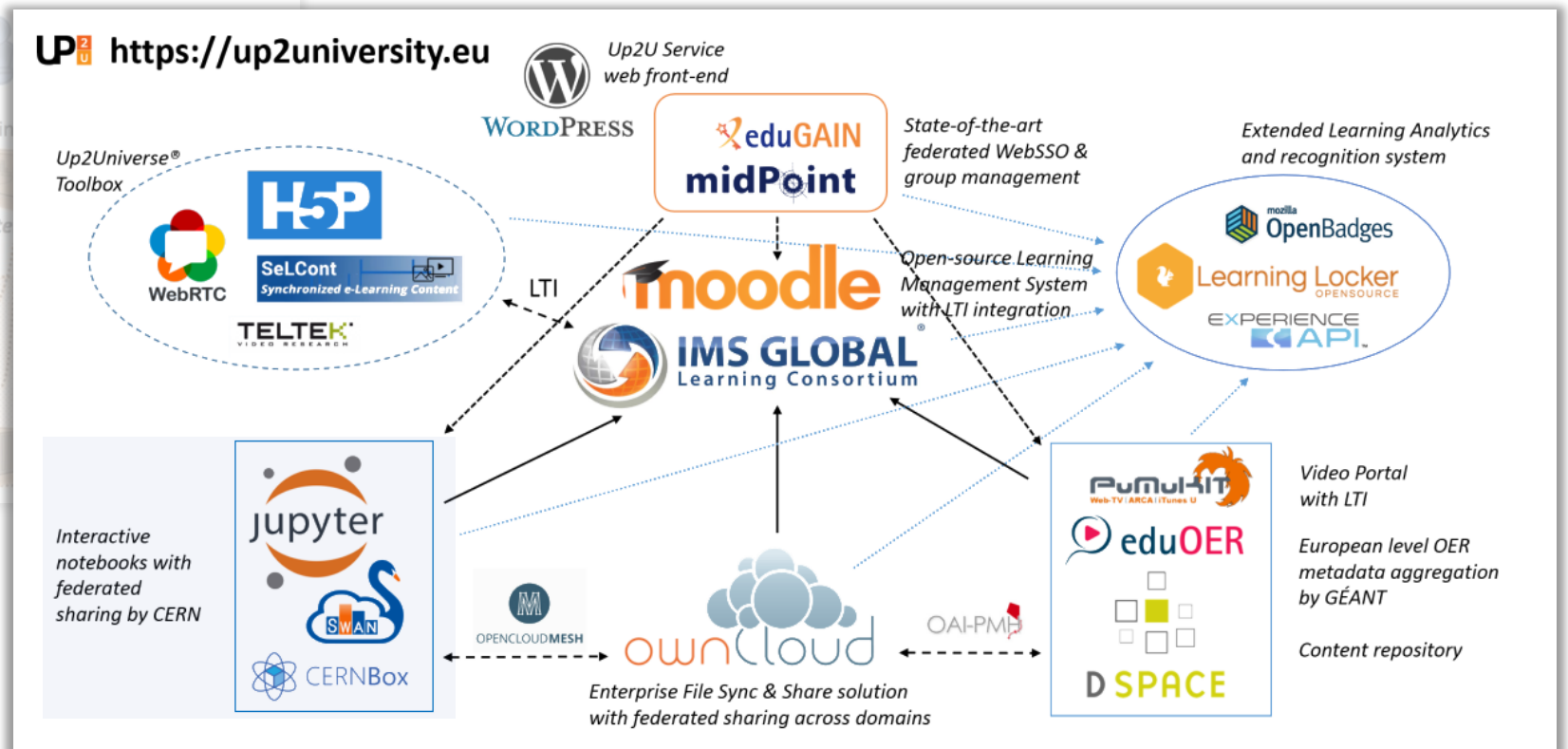
The diagram illustrates the Docker-based Service Architecture for Up To University. It is organized into three main functional areas:

- CERNBox:** This section includes a Sync Client, a CERNBox Web Server (labeled *cernboxgateway*), and the CERNBox component (labeled *cernbox*). The Sync Client interacts with the CERNBox Web Server via HTTP Webdav. Data flows from the CERNBox Web Server to the CERNBox component.
- EOS:** This section includes 6x Storage (labeled *eos-fst{1..6}*) and Messaging (labeled *eos-mq*). Data flows from the CERNBox component to the EOS storage.
- SWAN:** This section includes EOS Access (labeled *eos-fuse*), ROOT, and jupyter. Data flows from the EOS storage to the EOS Access, which then connects to the ROOT and jupyter components.

The diagram also features the Up To University logo and a URL: <https://up2u>.

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

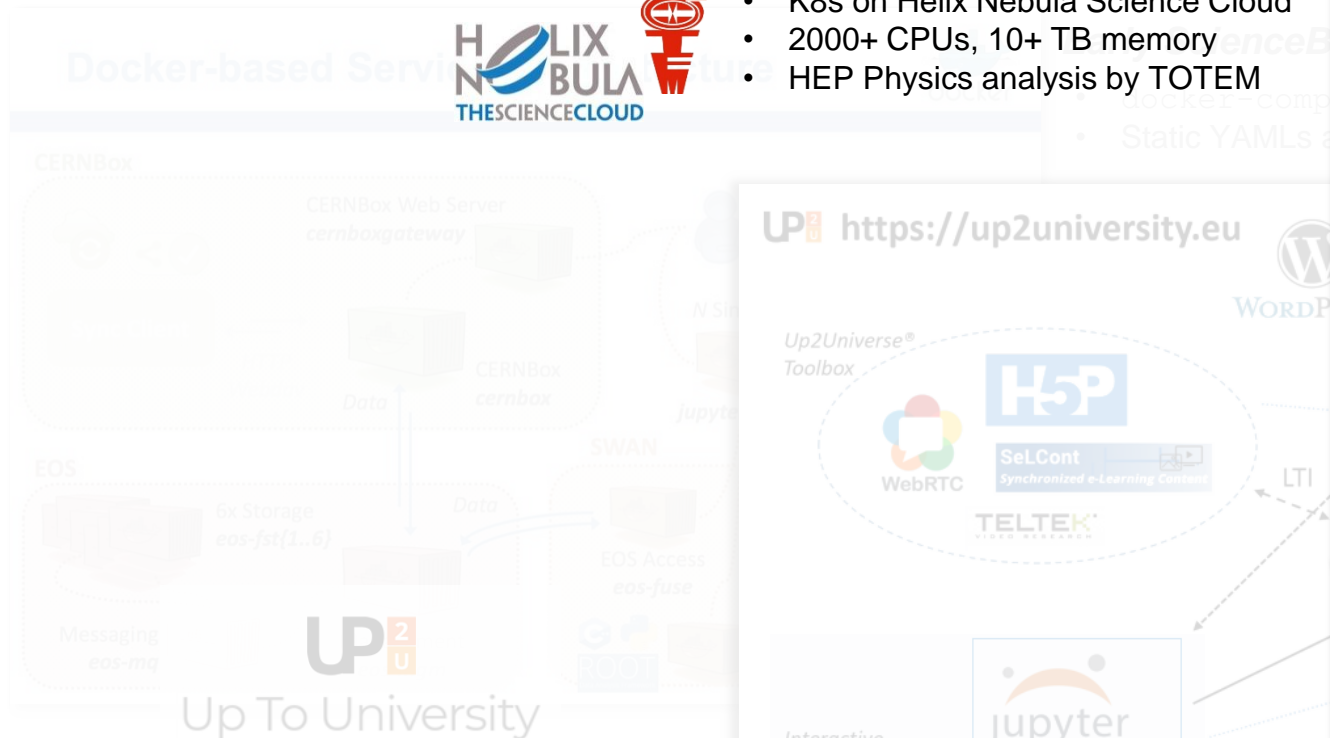
- docker-compose on one host
- Static YAMLS and bash wrappers



ScienceBox Heritage

ScienceBox for HEP, c.2018

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



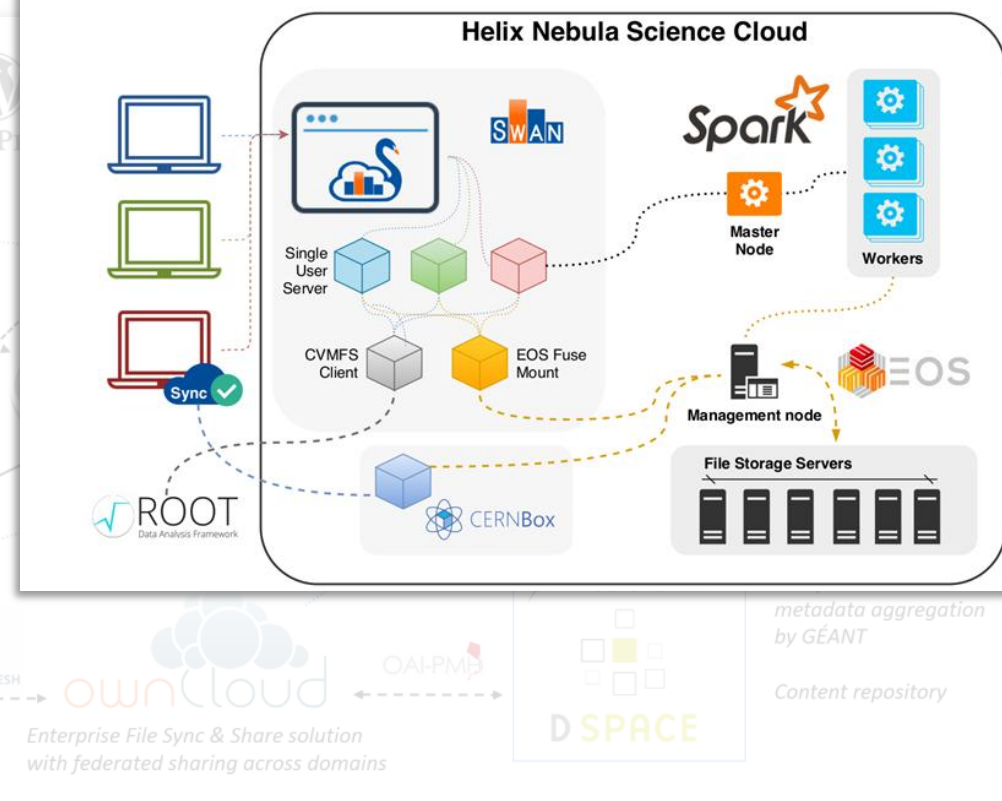
ScienceBox for Up To University (Up2U)

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

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 University of Science and Technology, Krakow, Poland, Email: {grzanka,malawski}@agh.edu.pl
^{*}CERN, CH-1211 Geneva 23, Switzerland, Email: {first.last}@cern.ch



ScienceBox Heritage



ScienceBox for HEP, c.2018

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

CERN technologies contribute to openUp2U, a learning platform for schools in Europe

The free remote-learning platform enables continued learning during the COVID-19 pandemic

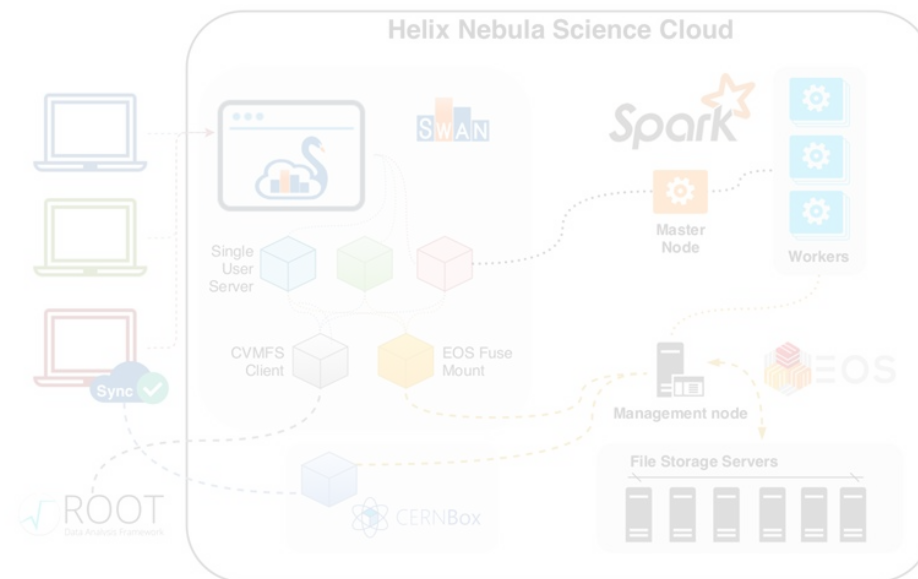
15 APRIL, 2020



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 University of Science and Technology, Krakow, Poland, Email: {grzanka,malawski}@agh.edu.pl
* CERN, CH-1211 Geneva 23, Switzerland, Email: {first.last}@cern.ch



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

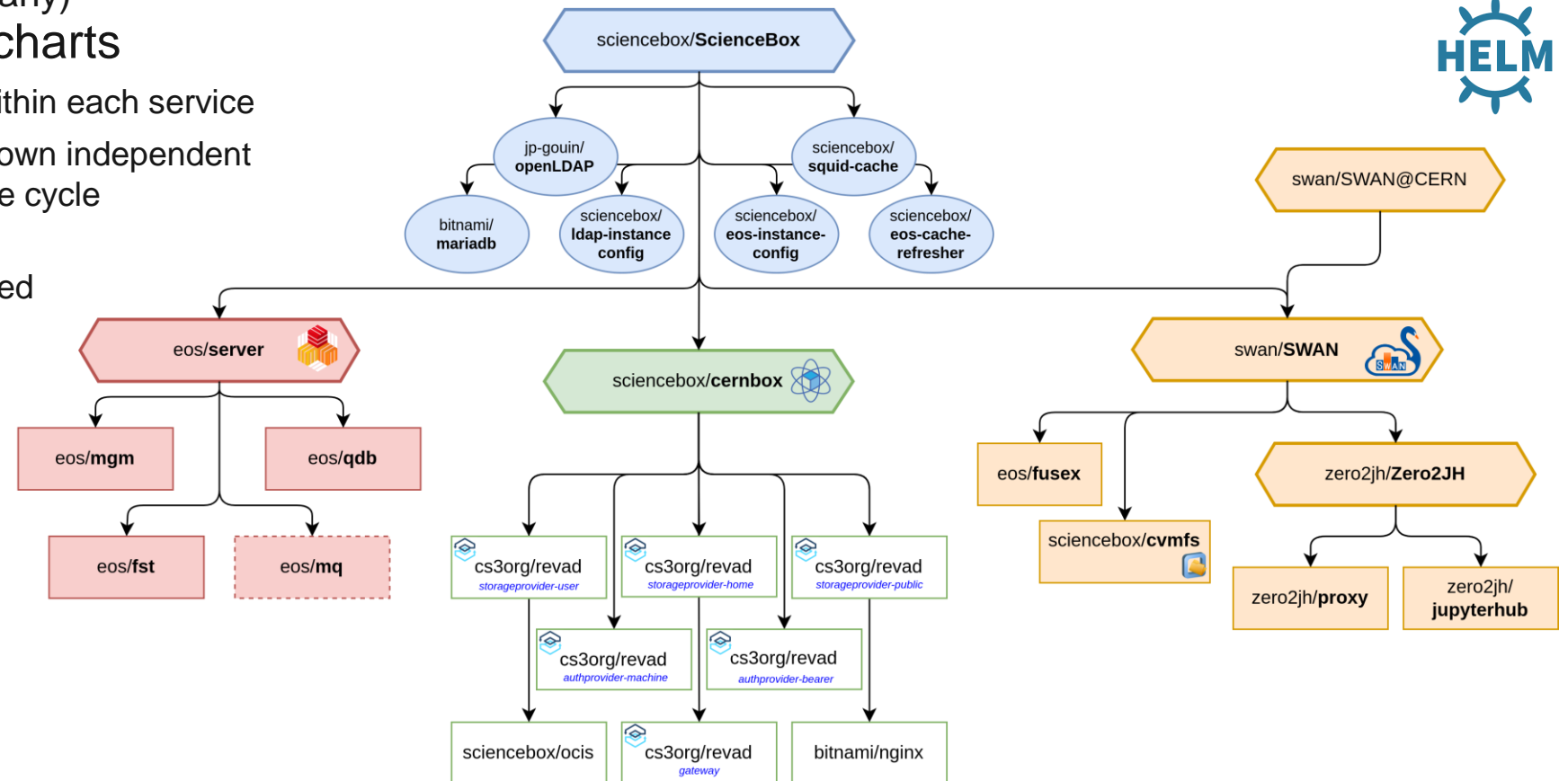
Why ScienceBox 2.0?

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

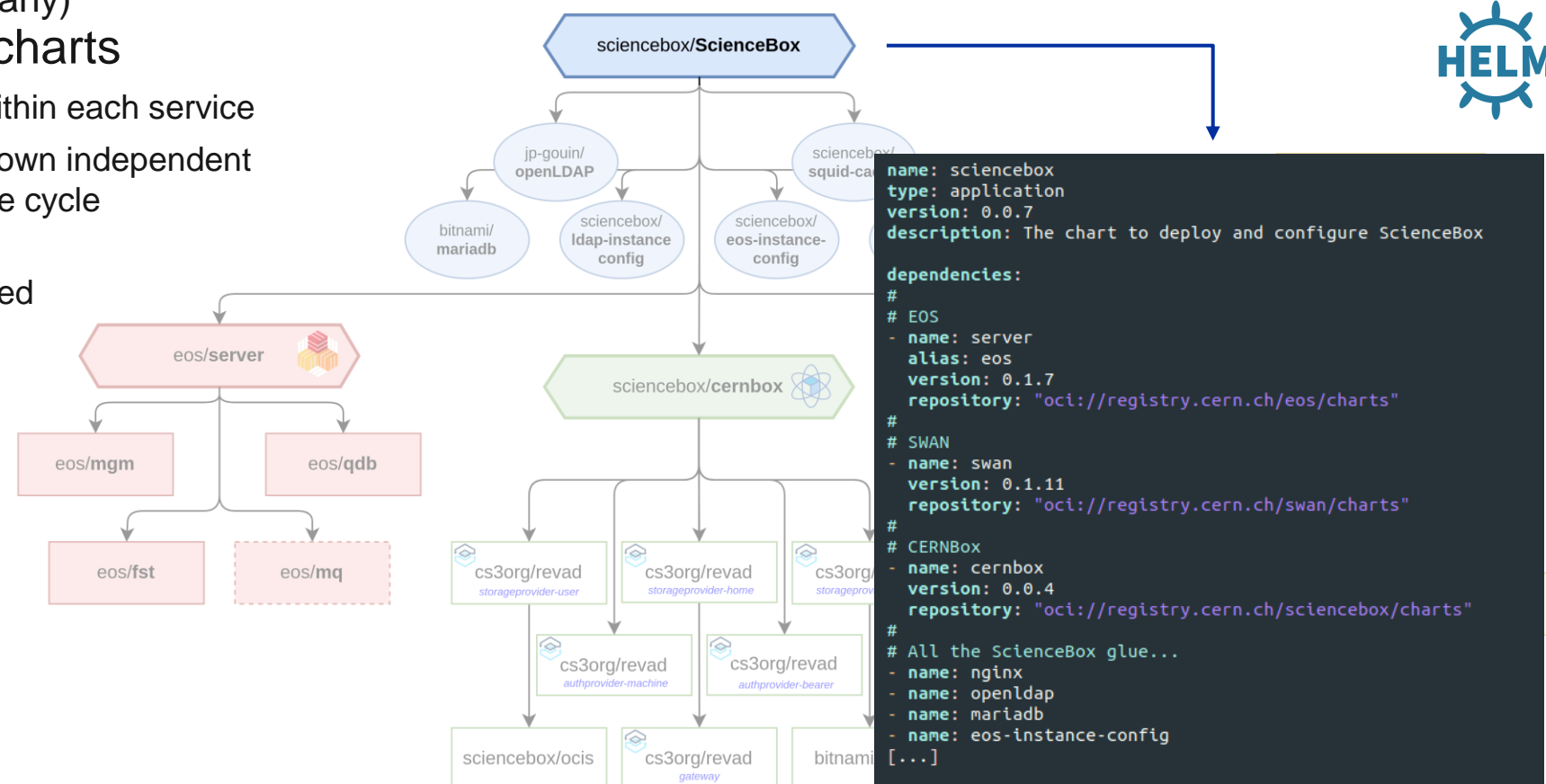
Renewed Architecture

- 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
```

boxed: Demonstrator in minikube

Time for a demo

youtu.be/knzfCZPrZUE

Wrapping Up

Use Cases



Sites



ScienceBox



Infrastructure



Technology



Coming Up Next: ScienceBox for OnBoard

onboard.app.cern.ch

An open-source dashboard for
science best practices,
open data and reproducible results.

Enabling researchers to get access to cutting-edge technologies
in a few clicks.

Learn more »



Guidance & Best Practices



Collaborations &
Multidisciplinarity



Open & Reproducible
Science



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

The eventual goal: a fully k8s-native T2
Installable with Helm



- Helm: application manager for Kubernetes
 - One command to install/upgrade everything
 - Comprehensive configuration via one YAML file
- **helm install T2Site**
 - (K)APEL accounting done
 - frontier-squid done
 - compute (security rules, Harvester setup) done (static YAML)
 - EOS SE in progress
 - CVMFS-CSI optional
 - ~~Compute Element~~ built-in
 - ~~Batch system~~ built-in



UVic T2 on Kubernetes - EOS Workshop 2023

3

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

1. Ubiquitous /cvmfs

The eventual goal: a fully k8s-native T2
Installable with Helm



- Helm: application manager for Kubernetes
 - One command to install/upgrade everything
 - Comprehensive configuration via one YAML file
- **helm install T2Site**
 - (K)APEL accounting done
 - frontier-squid done
 - compute (security rules, Harvester setup) done (static YAML)
 - EOS SE in progress
 - CVMFS-CSI optional
 - ~~Compute Element~~ built-in
 - ~~Batch system~~ built-in



UVic T2 on Kubernetes - EOS Workshop 2023

3

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

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

The eventual goal: a fully k8s-native T2
Installable with Helm



- Helm: application manager for Kubernetes
 - One command to install/upgrade everything
 - Comprehensive configuration via one YAML file
- **helm install T2Site**
 - (K)APEL accounting done
 - frontier-squid done
 - compute (security rules, Harvester setup) done (static YAML)
 - **EOS SE** in progress
 - CVMFS-CSI optional
 - ~~Compute Element~~ built-in
 - ~~Batch system~~ built-in



UVic T2 on Kubernetes - EOS Workshop 2023

3

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

A grid site reimaged: building a fully cloud-native ATLAS T2 on Kubernetes

📅 9 May 2023, 14:45

🕒 15m

📍 Marriott Ballroom IV (Norfolk Waterside Marriott)

Speaker

👤 Taylor, Ryan Paul (University of Victoria)

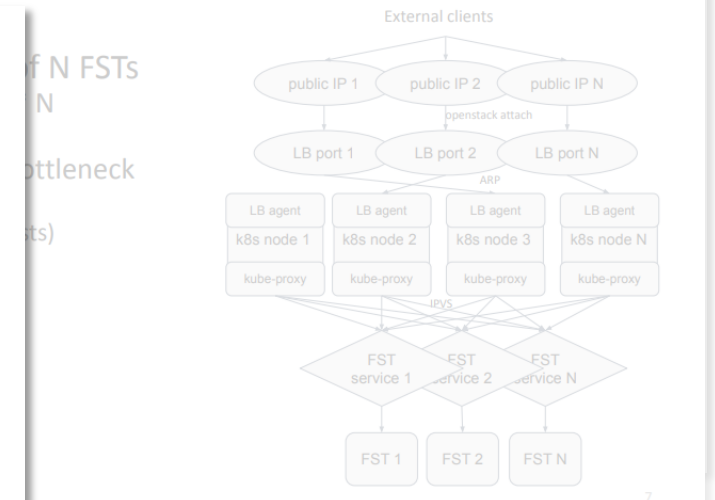
Oral

📁 Track 7 - Facilities a...

Track 7 - Facilities and ...

The eventual goal: a fully k8s-native T2
Installable with Helm

Network architecture on k8s for EOS

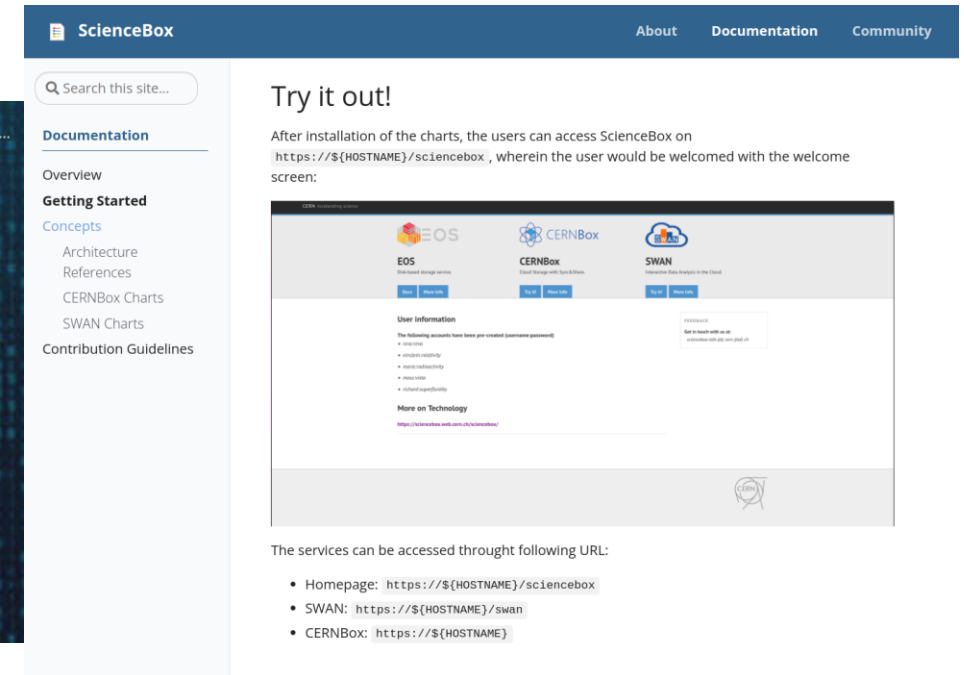
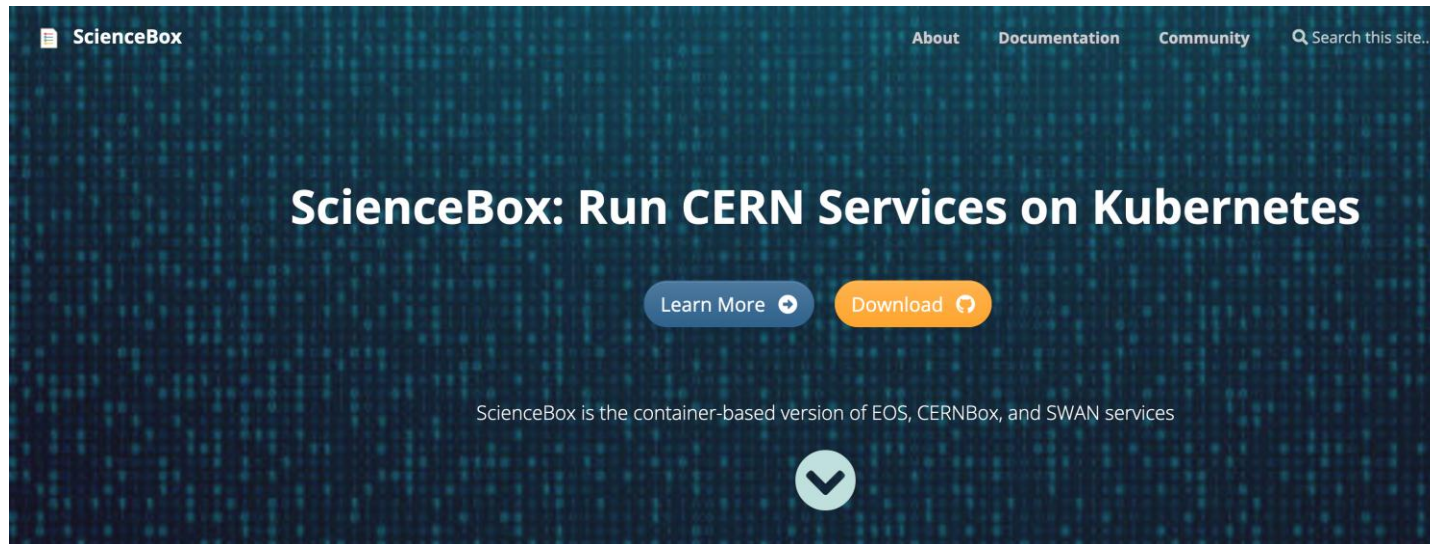


Where to find ScienceBox



■ 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>

■ More on ScienceBox services

- [eos](https://eos.web.cern.ch/), [cernbox](https://cernbox.web.cern.ch/), [swan](https://swan.web.cern.ch/), [cvmfs](https://cvmfs.web.cern.ch/).web.cern.ch

**Contributions and
Testing Welcome!**



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

```
[root@mboxed-demo-cs3-centos mboxed]# ./ScienceBox.sh
Configuring SELinux ...
Configuring networking ...
WARNING: iptables and IP forwarding rules need to be modified.
- The existing iptables configuration will be saved to file (/root/mboxed/iptables_1678264524.save) in order to restore it, if needed.
- Changes to IP forwarding rules will be reported to roll them back, if needed.
- The Docker daemon needs to be restarted. Running containers will temporarily stop while the Docker server restarts.
Do you want to continue? [y/N] █
```

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 ...

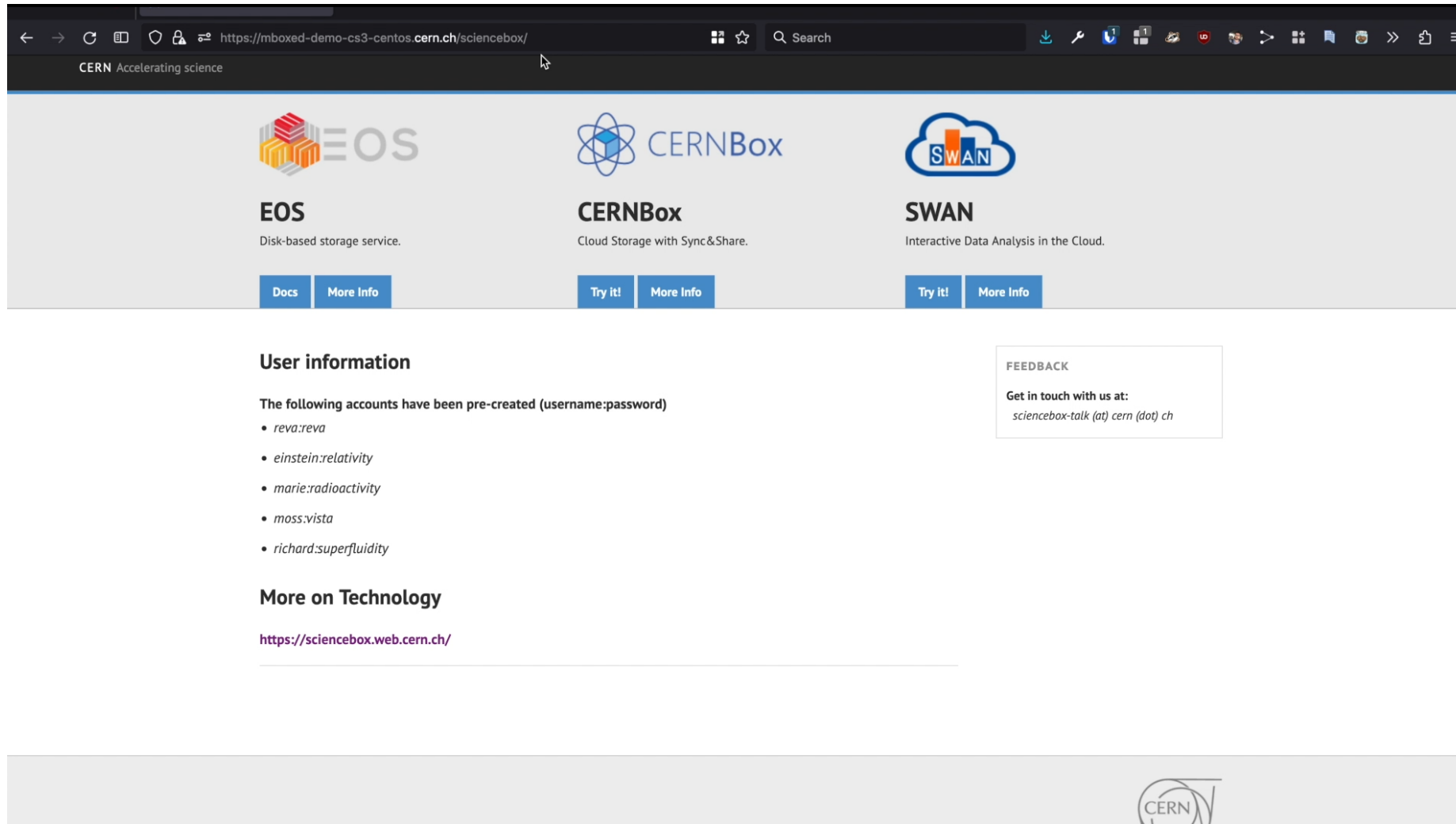
The deployment should be ready in a few minutes.
You can check the containers status by typing `kubectl get pods`.

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

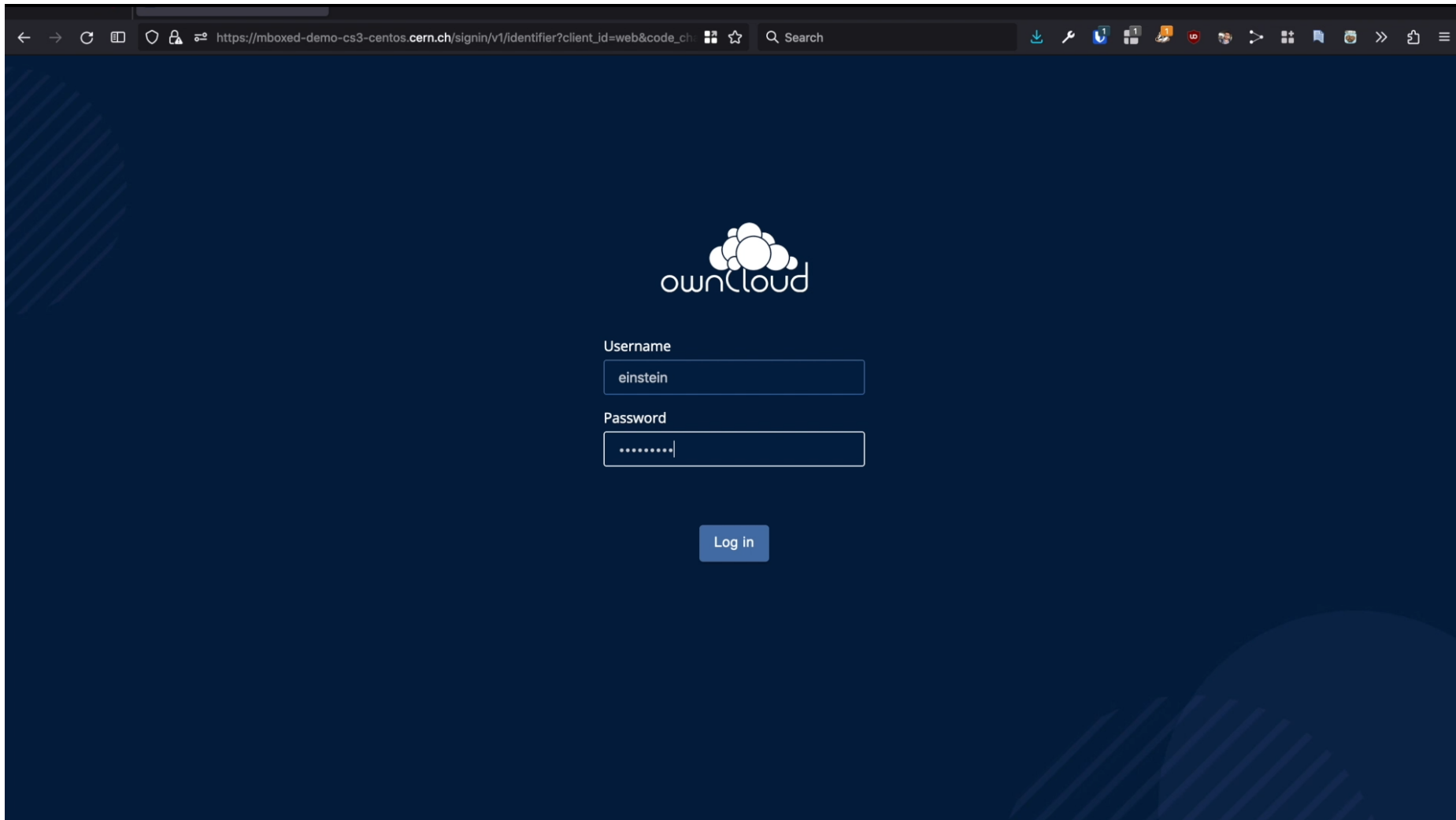
mboxed: Installing ScienceBox

```
[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]#
```

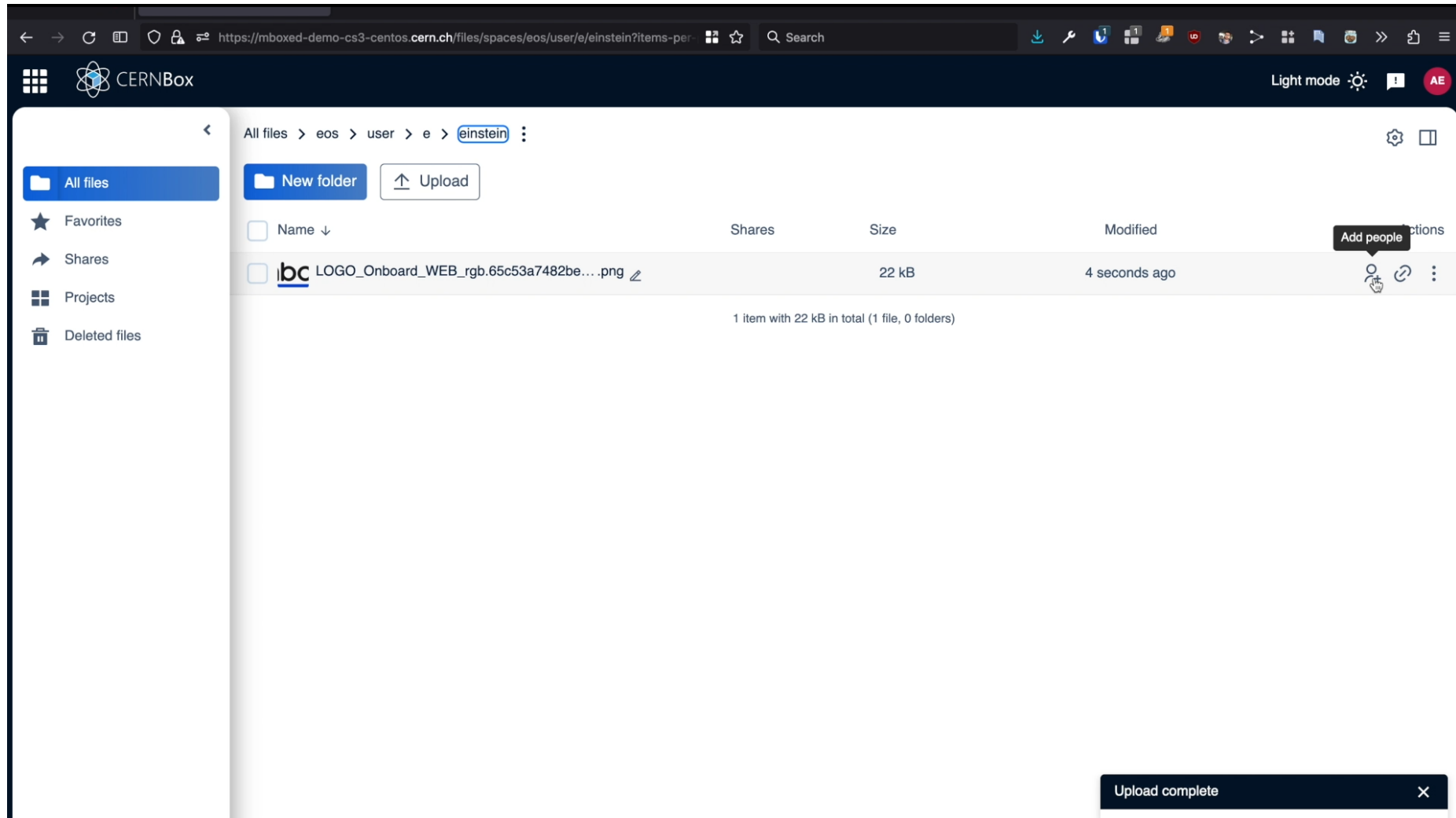
mboxed: Landing Page



mboxed: Logging in



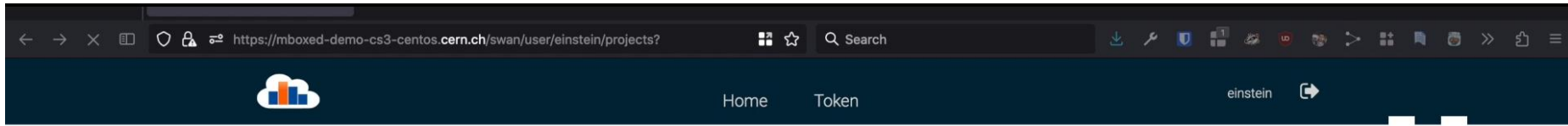
mboxed: Upload and Share



mboxed: Upload and Share

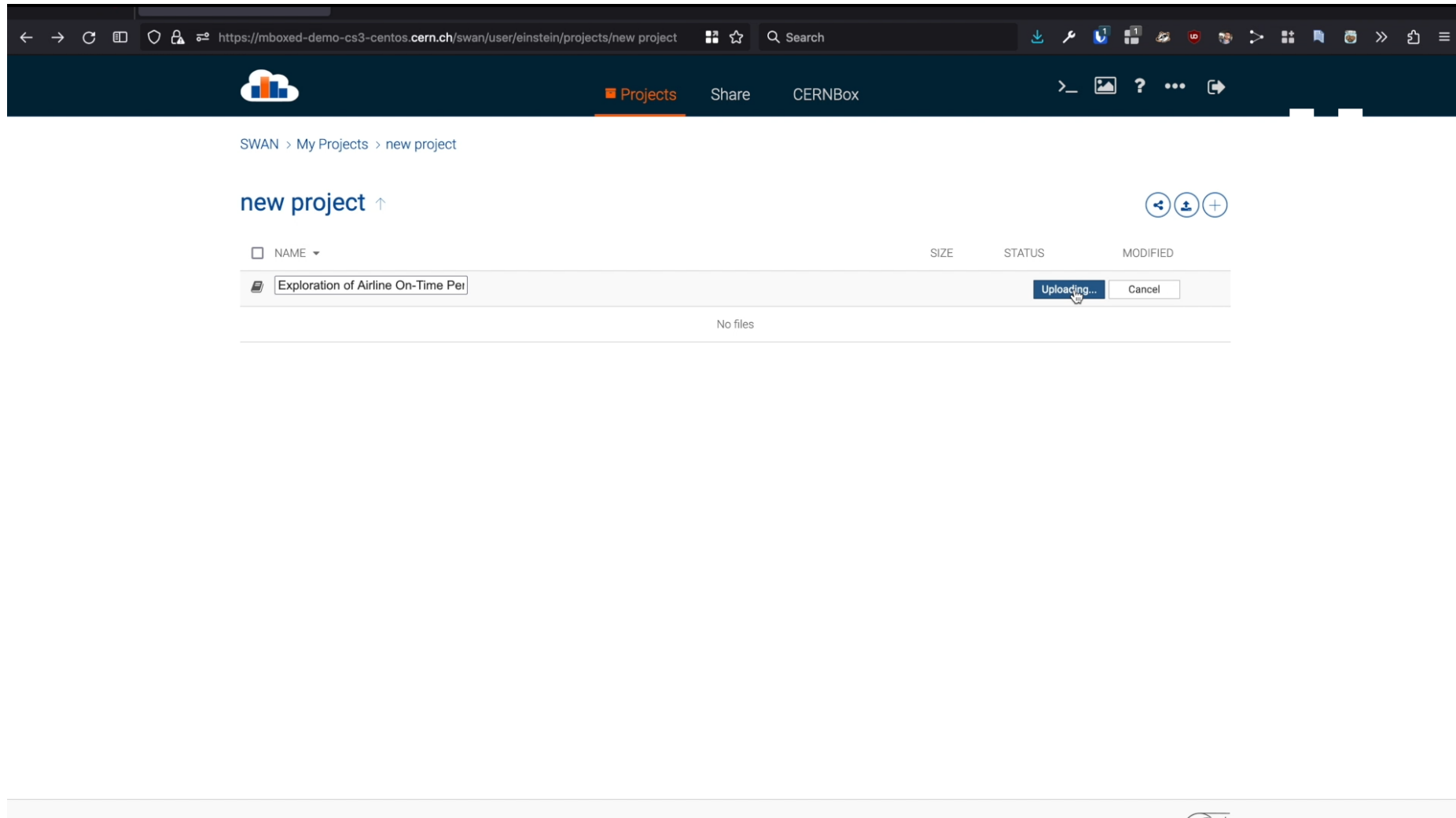
The screenshot displays the mboxed web interface within a browser. The address bar shows the URL: `https://mboxed-demo-cs3-centos.cern.ch/files/spaces/eos/user/e/einstein?items-per-`. The interface features a dark header with the CERNBox logo, a search bar, and a user profile icon labeled 'AE'. On the left, a sidebar contains navigation links: 'All files', 'Favorites', 'Shares', 'Projects', and 'Deleted files'. The main content area shows a file list with columns for 'Name', 'Shares', 'Size', 'Modified', and 'Actions'. A file named 'LOGO_Onboard_WEB_rgb... .png' (22 kB, modified 6 seconds ago) is selected. Above the file list, there are buttons for 'Download', 'Move', 'Copy', and 'Delete'. A 'Shares' panel on the right offers options to 'Share with people' (via an 'Invite' dropdown showing 'marie' and a user card for 'Marie Skłodowska Curie...') and 'Share via public link' (with a 'Quick link' section containing 'No link' and 'Create link' buttons, and an 'Add link' button). A dark notification bar at the bottom right states 'Upload complete'.

mboxed: Swan

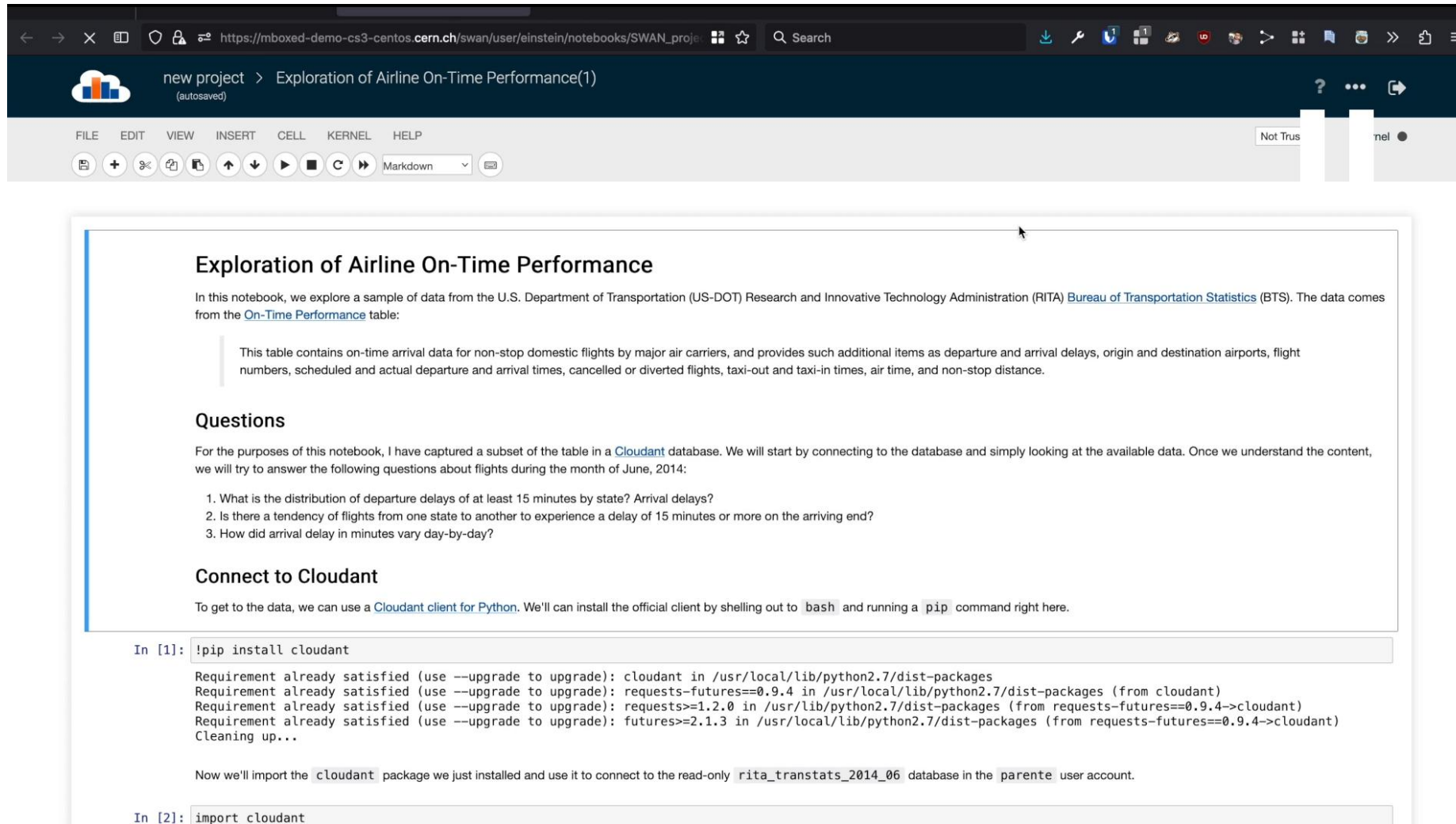


Starting your session

mboxed: Swan



mboxed: Swan Python Notebook



The screenshot shows a web browser window with the URL `https://mboxed-demo-cs3-centos.cern.ch/swan/user/einstein/notebooks/SWAN_proje`. The notebook interface has a dark blue header with the title "new project > Exploration of Airline On-Time Performance(1)" and a sub-label "(autosaved)". Below the header is a menu bar with options: FILE, EDIT, VIEW, INSERT, CELL, KERNEL, and HELP. A toolbar contains icons for file operations and a dropdown menu currently set to "Markdown".

The notebook content is titled "Exploration of Airline On-Time Performance". It begins with an introduction: "In this notebook, we explore a sample of data from the U.S. Department of Transportation (US-DOT) Research and Innovative Technology Administration (RITA) [Bureau of Transportation Statistics](#) (BTS). The data comes from the [On-Time Performance](#) table:

This table contains on-time arrival data for non-stop domestic flights by major air carriers, and provides such additional items as departure and arrival delays, origin and destination airports, flight numbers, scheduled and actual departure and arrival times, cancelled or diverted flights, taxi-out and taxi-in times, air time, and non-stop distance.

Questions

For the purposes of this notebook, I have captured a subset of the table in a [Cloudant](#) database. We will start by connecting to the database and simply looking at the available data. Once we understand the content, we will try to answer the following questions about flights during the month of June, 2014:

1. What is the distribution of departure delays of at least 15 minutes by state? Arrival delays?
2. Is there a tendency of flights from one state to another to experience a delay of 15 minutes or more on the arriving end?
3. How did arrival delay in minutes vary day-by-day?

Connect to Cloudant

To get to the data, we can use a [Cloudant client for Python](#). We'll can install the official client by shelling out to `bash` and running a `pip` command right here.

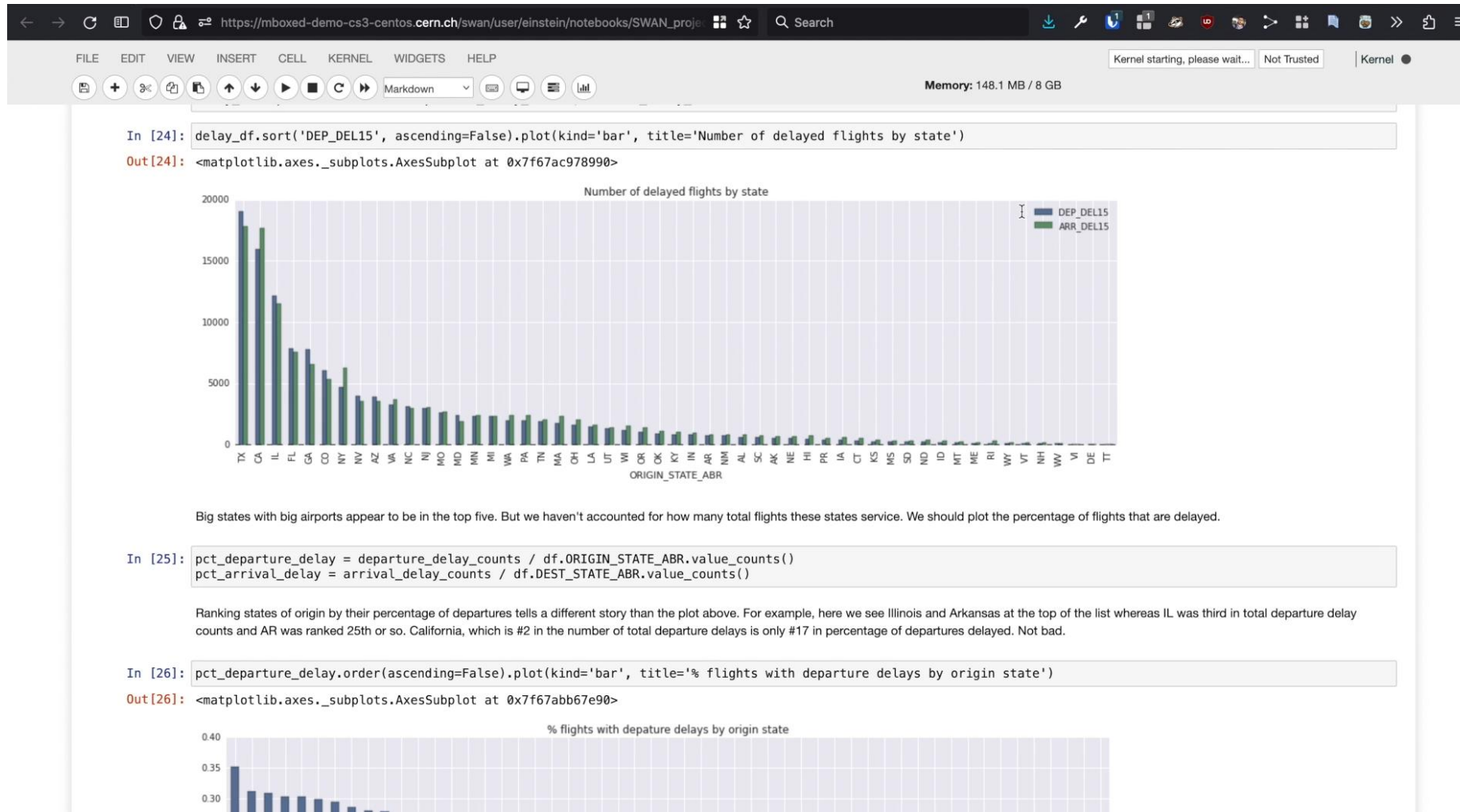
```
In [1]: !pip install cloudant
```

Requirement already satisfied (use --upgrade to upgrade): cloudant in /usr/local/lib/python2.7/dist-packages
Requirement already satisfied (use --upgrade to upgrade): requests-futures==0.9.4 in /usr/local/lib/python2.7/dist-packages (from cloudant)
Requirement already satisfied (use --upgrade to upgrade): requests>=1.2.0 in /usr/lib/python2.7/dist-packages (from requests-futures==0.9.4->cloudant)
Requirement already satisfied (use --upgrade to upgrade): futures>=2.1.3 in /usr/local/lib/python2.7/dist-packages (from requests-futures==0.9.4->cloudant)
Cleaning up...

Now we'll import the `cloudant` package we just installed and use it to connect to the read-only `rita_transtats_2014_06` database in the `parente` user account.

```
In [2]: import cloudant
```

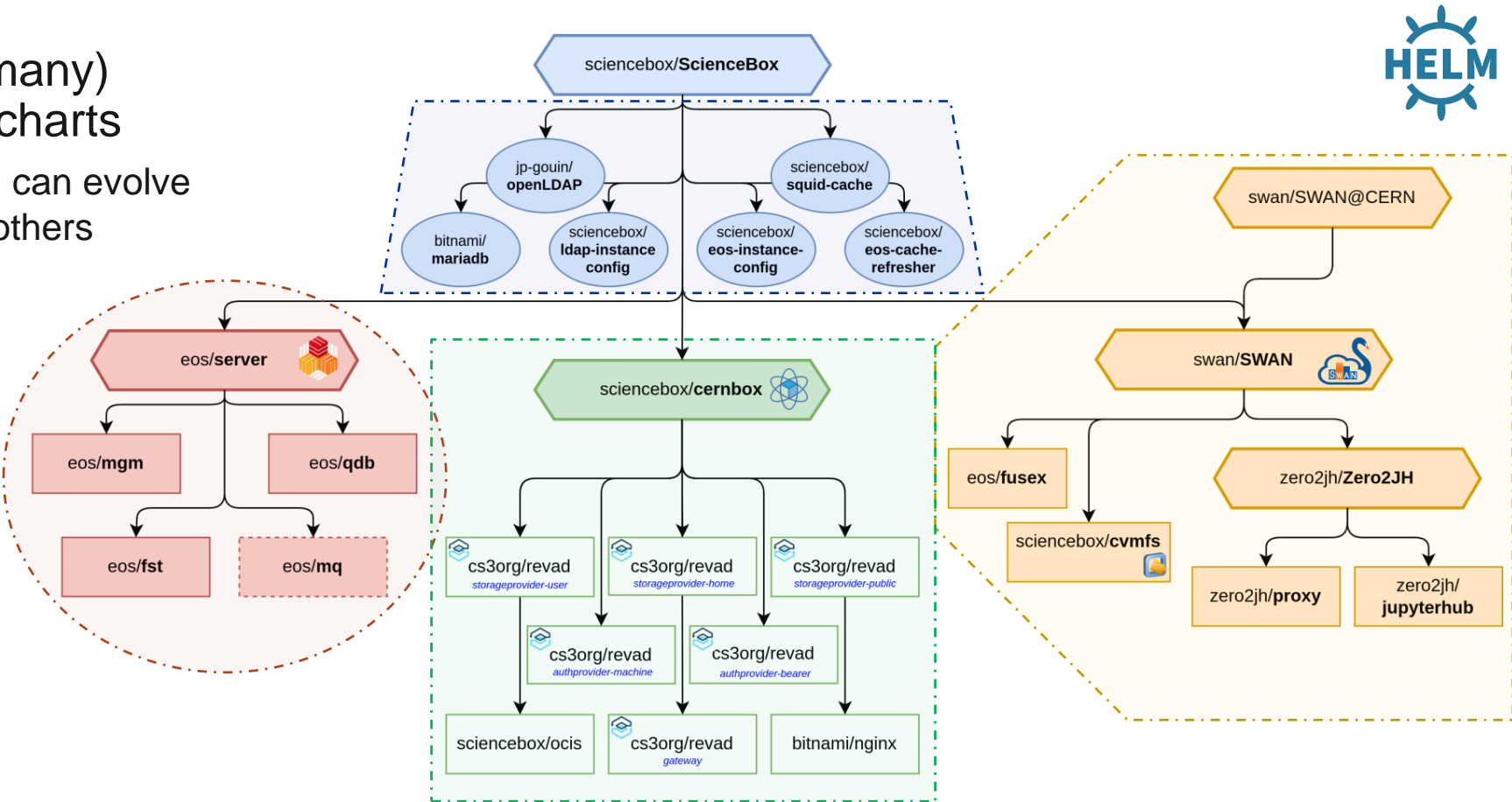

mboxed: Swan Python Notebook



Backup

Renewed Architecture

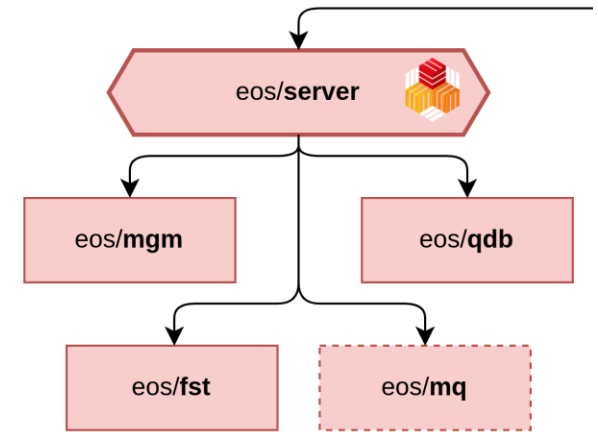
- ScienceBox is root to (many) single and umbrella charts
 - Each service (and chart) can evolve independently on the others



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

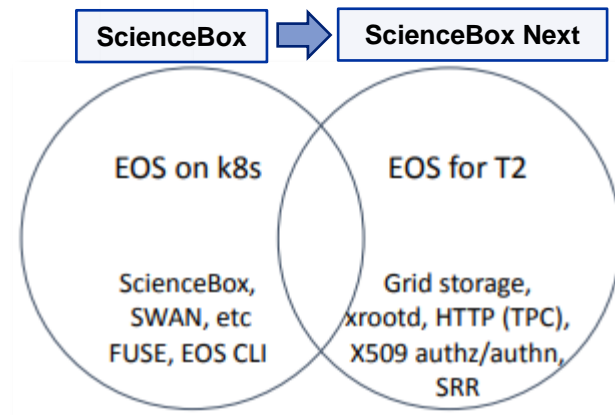
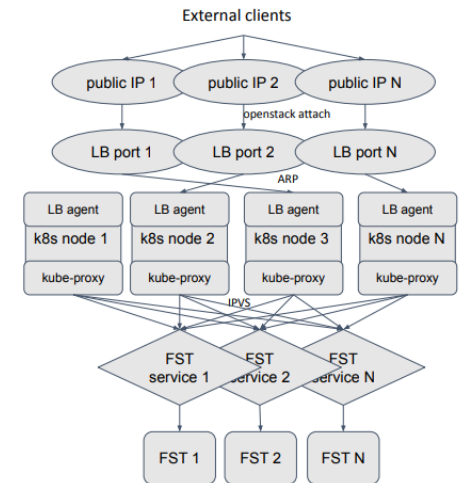
■ ScienceBox for WLCG T2s

- 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](https://eos.web.cern.ch/), [cernbox](https://cernbox.web.cern.ch/), [swan](https://swan.web.cern.ch/), [cvmfs](https://cvmfs.web.cern.ch/).web.cern.ch

**Contributions and
Testing Welcome!**