ScienceBox 2.0
Evolving the demonstrator package for CERN Storage and Analysis services

May 9, 2023
What is ScienceBox?

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

- 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

ScienceBox for Up To University (Up2U)
- Kubernetes-managed multi-host clusters
- Integration with Up2U IDP, Storage Federation
- Still plenty of bash and workarounds

Big Data Tools and Cloud Services for High Energy Physics Analysis in TOTEM Experiment
Valentina Avati¹, Milosz Biaszkiewicz¹, Enrico Bocchi¹, Luca Canali¹, Diogo Castro¹, Javier Cervantes¹, Leszek Grzanka¹, Enrico Guiraud¹, Jan Kaspar¹, Prasanth Kothuri¹, Massimo Lamanna¹, Maciej Malawski¹, Aleksandra Minichi¹, Jakub Moscicki¹, Shravan Murari¹, Danilo Piparo¹, Enric Tejedor¹
¹AGH University of Science and Technology, Krakow, Poland, Email: [grzanka,malawski]@agh.edu.pl
* CERN, CH-1211 Geneva 23, Switzerland, Email: [first.last]@cern.ch

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
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
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

```yaml
name: sciencebox
type: application
version: 0.0.7
description: The chart to deploy and configure ScienceBox
dependencies:
  # EOS
  - name: server
    alias: eos
    version: 8.1.7
    repository: "oci://registry.cern.ch/eos/charts"
  # SWAN
  - name: swan
    version: 0.1.11
    repository: "oci://registry.cern.ch/swan/charts"
  # CERNBox
  - name: cernbox
    version: 0.6.4
    repository: "oci://registry.cern.ch/sciencebox/charts"
  # All the ScienceBox glue...
  - name: mgmt
    repository: "oci://registry.cern.ch/sciencebox/charts"
  - name: openldap
  - name: mariadb
  - name: eos-instance-config
[...]`
```
mboxed: Demonstrator in minikube

- How to get started?

```bash
[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
- UP2U
- CERN Against COVID-19
- Up To University

Sites
- CERN
- Joint Research Centre (JRC)
- aarnet

Technology
- CERNBox
- EOS
- SWAN
- Kubernetes
- Docker
- Helm

Infrastructure
- Amazon Web Services
- OpenStack
- T-Systems
- Nobula
- Helix

ScienceBox

CHEP 2023
Coming Up Next: ScienceBox for OnBoard

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 »
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

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

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

- 9 May 2023, 14:45
- Marriott Ballroom IV (Norfolk Waterside Marriott)

Speaker

Taylor, Ryan Paul (University of Victoria)
Where to find ScienceBox

- **ScienceBox**
  - Project homepage: [sciencebox.web.cern.ch/](sciencebox.web.cern.ch/)
  - ArtifactHUB: [artifacthub.io/packages/helm/sciencebox/sciencebox](artifacthub.io/packages/helm/sciencebox/sciencebox)
  - New and improved documentation
Where to find ScienceBox

- **ScienceBox**
  - Project homepage: [sciencebox.web.cern.ch/](http://sciencebox.web.cern.ch/)
  - ArtifactHUB: [artifacthub.io/packages/helm/sciencebox/sciencebox](https://artifacthub.io/packages/helm/sciencebox/sciencebox)
  - New and improved documentation

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

- **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

```
[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
## Installing ScienceBox

```
[root@mboxed-demo-cs3-centos mboxed]# kubectl get pods

<table>
<thead>
<tr>
<th>NAME</th>
<th>READY</th>
<th>STATUS</th>
<th>RESTARTS</th>
<th>AGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>hub-6d4dd7d5-f44w</td>
<td>1/1</td>
<td>Running</td>
<td>0</td>
<td>9m51s</td>
</tr>
<tr>
<td>proxy-7cfcc68888-5g2z</td>
<td>1/1</td>
<td>Running</td>
<td>0</td>
<td>9m52s</td>
</tr>
<tr>
<td>sciencebox-authproviderbearer-7844ccc5cd-575fx</td>
<td>1/1</td>
<td>Running</td>
<td>0</td>
<td>9m52s</td>
</tr>
<tr>
<td>sciencebox-authprovidermachine-74b44d0875-f7vq</td>
<td>1/1</td>
<td>Running</td>
<td>0</td>
<td>9m52s</td>
</tr>
<tr>
<td>sciencebox-cernbox-web-748dc49659-vthkw</td>
<td>1/1</td>
<td>Running</td>
<td>0</td>
<td>9m52s</td>
</tr>
<tr>
<td>sciencebox-cvmfs-5krds</td>
<td>2/2</td>
<td>Running</td>
<td>0</td>
<td>9m52s</td>
</tr>
<tr>
<td>sciencebox-eos-instance-config-rftgz</td>
<td>0/1</td>
<td>Completed</td>
<td>1</td>
<td>9m52s</td>
</tr>
<tr>
<td>sciencebox-eos-user-cache-refresherv-lv44c</td>
<td>0/1</td>
<td>Completed</td>
<td>1</td>
<td>9m52s</td>
</tr>
<tr>
<td>sciencebox-fst-0</td>
<td>1/1</td>
<td>Running</td>
<td>0</td>
<td>9m52s</td>
</tr>
<tr>
<td>sciencebox-fst-1</td>
<td>1/1</td>
<td>Running</td>
<td>0</td>
<td>7m40s</td>
</tr>
<tr>
<td>sciencebox-fst-2</td>
<td>1/1</td>
<td>Running</td>
<td>0</td>
<td>7m27s</td>
</tr>
<tr>
<td>sciencebox-fst-3</td>
<td>1/1</td>
<td>Running</td>
<td>0</td>
<td>7m17s</td>
</tr>
<tr>
<td>sciencebox-fusex-ddsq7</td>
<td>1/1</td>
<td>Running</td>
<td>0</td>
<td>9m52s</td>
</tr>
<tr>
<td>sciencebox-gateway-8499b95795-md8xj</td>
<td>1/1</td>
<td>Running</td>
<td>0</td>
<td>9m52s</td>
</tr>
<tr>
<td>sciencebox-ldap-0</td>
<td>1/1</td>
<td>Running</td>
<td>1</td>
<td>9m52s</td>
</tr>
<tr>
<td>sciencebox-ldap-1</td>
<td>1/1</td>
<td>Running</td>
<td>1</td>
<td>9m13s</td>
</tr>
<tr>
<td>sciencebox-ldap-2</td>
<td>1/1</td>
<td>Running</td>
<td>1</td>
<td>8m43s</td>
</tr>
<tr>
<td>sciencebox-ldap-instance-config-g89d9</td>
<td>0/1</td>
<td>Completed</td>
<td>0</td>
<td>9m52s</td>
</tr>
<tr>
<td>sciencebox-mariadb-0</td>
<td>1/1</td>
<td>Running</td>
<td>0</td>
<td>9m52s</td>
</tr>
<tr>
<td>sciencebox-mgm-0</td>
<td>4/4</td>
<td>Running</td>
<td>0</td>
<td>9m52s</td>
</tr>
<tr>
<td>sciencebox-nginx-welcome-page-74c8d77495-bxmlq</td>
<td>1/1</td>
<td>Running</td>
<td>0</td>
<td>9m52s</td>
</tr>
<tr>
<td>sciencebox-ocis-86687645d8-7cprs</td>
<td>1/1</td>
<td>Running</td>
<td>0</td>
<td>9m52s</td>
</tr>
<tr>
<td>sciencebox-qdb-0</td>
<td>1/1</td>
<td>Running</td>
<td>0</td>
<td>9m52s</td>
</tr>
<tr>
<td>sciencebox-qdb-1</td>
<td>1/1</td>
<td>Running</td>
<td>0</td>
<td>9m29s</td>
</tr>
<tr>
<td>sciencebox-qdb-2</td>
<td>1/1</td>
<td>Running</td>
<td>0</td>
<td>9m19s</td>
</tr>
<tr>
<td>sciencebox-storageproviderhome-8c667568-h88888</td>
<td>1/1</td>
<td>Running</td>
<td>0</td>
<td>9m52s</td>
</tr>
<tr>
<td>sciencebox-storageproviderpublic-6d69947688-cwlfv</td>
<td>1/1</td>
<td>Running</td>
<td>0</td>
<td>9m52s</td>
</tr>
<tr>
<td>sciencebox-storageprovideruser-57d4cd9766-57n4t</td>
<td>1/1</td>
<td>Running</td>
<td>0</td>
<td>9m52s</td>
</tr>
</tbody>
</table>
```

[root@mboxed-demo-cs3-centos mboxed]#
mboxed: Landing Page

User information

The following accounts have been pre-created (username:password)
- ronnie_robert
- einstein_relativity
- manihandaradioactivity
- marcella
- richard_superfluid

More on Technology

https://sciencebox.web.cern.ch/
mboxed: Logging in
mboxed: Upload and Share
mboxed: Upload and Share
mboxed: Swan

Starting your session
mboxed: Swan
Exploration of Airline On-Time Performance

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 install the official client by shelling out to `bash` and running a `pip` command right here.

```
In [11]: !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==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_transstats_2014_06` database in the `pieante` user account.

```
In [12]: import cloudant
```
Big states with big airports appear to be in the top five. But we haven’t accounted for how many total flights these states service. We should plot the percentage of flights that are delayed.

```python
In [25]: pct_departure_delay = departure_delay_counts / df.ORIGIN_STATE_ABR.value_counts()
pct_arrival_delay = arrival_delay_counts / df.DEST_STATE_ABR.value_counts()

Ranking states of origin by their percentage of departures tells a different story than the plot above. For example, here we see Illinois and Arkansas at the top of the list whereas IL was third in total departure delay counts and AK was ranked 20th or so. California, which is 42 in the number of total departure delays is only #17 in percentage of departures delayed. Not bad.
```

```python
In [26]: pct_departure_delay.order(ascending=False).plot(kind='bar', title='% flights with departure delays by origin state')
```
Backup
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    AGE
eos-fst-0 1/1 Running 6m30s
eos-fst-1 1/1 Running 4m48s
eos-fst-2 1/1 Running 4m38s
eos-fst-3 1/1 Running 4m28s
eos-mgm-0 2/2 Running 6m30s
eos-qdb-0 1/1 Running 6m30s
eos-qdb-1 1/1 Running 6m21s
eos-qdb-2 1/1 Running 6m10s
### Coming Up Next: ScienceBox for Tier 2 sites

- **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

---

**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 host/aliases (/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!