

## ScienceBox

### Converging to Kubernetes containers in production for on-premise and hybrid clouds for CERNBox, SWAN, and EOS

Enrico Bocchi<sup>1,\*</sup>, Luca Canali<sup>1</sup>, Diogo Castro<sup>1</sup>, Prasanth Kothuri<sup>1</sup>, Hugo Gonzalez Labrador<sup>1</sup>, Maciej Malawski<sup>2</sup>, Jakub T. Mościcki<sup>1</sup>, and Piotr Mrowczynski<sup>1</sup>

<sup>1</sup>CERN, Esplanade des Particules 1, 1211 Geneva 23, Switzerland

<sup>2</sup>AGH University of Science and Technology, Aleja Adama Mickiewicza 30, 30-059 Kraków, Poland

**Abstract.** Docker containers are the de-facto standard to package, distribute, and run applications on cloud-based infrastructures. Commercial providers and private clouds expand their offer with container orchestration engines, making the management of resources and containerized applications tightly integrated. The Storage Group of CERN IT leverages on container technologies to provide ScienceBox: An integrated software bundle with storage and computing services for general purposes and scientific use. ScienceBox features distributed scalable storage, sync&share functionalities, and a web-based data analysis service, and can be deployed on a single machine or scaled-out across multiple servers. ScienceBox has proven to be helpful in different contexts, from High Energy Physics analysis to education for high schools, and has been successfully deployed on different cloud infrastructure and heterogeneous hardware.

## 1 Introduction

Container technologies are rapidly becoming the preferred means by developers and system administrators to package applications, distribute software, and run services [1]. The Storage Group of the IT department at the European Organization for Nuclear Research (CERN) has successfully exploited them as a basis for ScienceBox [2]: A Docker-based version of CERN cloud technologies for storage, sync&share, and data analysis.

ScienceBox enables the deployment of a fully-fledged replica of CERN production services on-premises: EOS (the CERN storage technology for LHC data and users' files), CERNBox (cloud synchronization and sharing for science), and Service for Web based ANalysis (SWAN) (data analysis platform in a browser based on Jupyter notebooks). It can be installed on a single machine, a cluster of hosts, or using cloud-provided container orchestrators and has been successfully deployed on OpenStack-based clouds as well as on commercial platforms such as Amazon Web Services [3] and the Open Telekom Cloud (OTC) [4].

In this paper, we firstly describe ScienceBox services, architecture, and deployment models in Sec. 2. Secondly, we report on how it was successfully used for High Energy Physics (HEP) analysis on a commercial cloud infrastructure and for education purposes in the context of the EU project Up To University (Up2U) in Sec. 3. Lastly, we summarize ScienceBox capabilities and envisage how it can be used and improved in the future in Sec. 4.

---

\*e-mail: [enrico.bocchi@cern.ch](mailto:enrico.bocchi@cern.ch)

## 2 ScienceBox

ScienceBox is a self-contained Docker-based version of CERN core cloud services. It delivers open-source software used at CERN for petabyte-scale production services in the context of HEP analysis data storage and analysis. In what follows, we describe ScienceBox reporting on the functionalities it delivers (Sec. 2.1), its architecture (Sec. 2.2), and the deployment models it supports (Sec. 2.3).

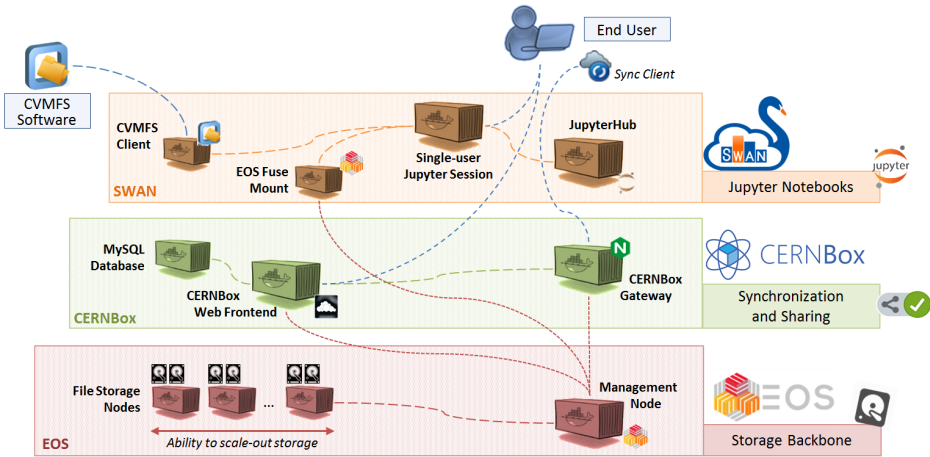
### 2.1 Service Portfolio

ScienceBox delivers an integrated set of services for storage, sync&share, and data analysis and enables their deployment on-premises. Such services include:

- EOS [5], the distributed storage service used at CERN to host all the LHC physics data and users' files. EOS consists of two major components server-side: i) The Management node (MGM), where all the metadata are stored, and ii) the File storage node (FST), where the actual payload of the files is stored on disk. Diverse access methods are supported client-side, including xrdcp, ROOT, WebDAV, and a native FUSE implementations.
- CERNBox [6], the sync and share platform for science that leverages EOS storage and provides cloud-based storage and sharing functionalities. CERNBox is based on the open-source ownCloud software and expands it by providing tight integration with other services available at CERN, including office suites, flowcharts editors, and Jupyter notebooks.
- SWAN [7], the Jupyter Notebook service at CERN that provides an advanced data analysis environment through a web interface. SWAN builds on top of the upstream Jupyter software by integrating storage from EOS, adding a custom user interface, and introducing the concept of *projects*: Collections of Jupyter notebooks, input datasets, images, plots, etc.. The custom interface allows the sharing of SWAN projects leveraging on the CERNBox sharing APIs. SWAN can also be interfaced with powerful computational backends including Apache Spark clusters or local GPUs for mass processing.
- CVMFS [8], the service for software distribution at a global scale adopted by the Worldwide LHC Computing Grid (WLCG) for the delivery of experiment analysis software. CVMFS provides read-only access to clients by mounting a POSIX file system at `/cvmfs` and uses standard HTTP protocol as the transport layer, allowing for the use of off-the-shelf web server and web caching software.

### 2.2 Architecture

Fig. 1 depicts the internal architecture of ScienceBox. The three main blocks represent the main services delivered through containers: EOS at the bottom, CERNBox in the middle, and SWAN on the top. EOS comprises two types of containers, one providing the MGM component and the second running the FST daemon, and provides storage to CERNBox (via native `xrootd` calls and HTTPS protocol) and to SWAN (via a client that provides fuse mount capabilities). CERNBox is delivered through a group of 3 containers, one routing synchronization client traffic to EOS and web traffic to the web frontend, another providing the core functionalities and the web interface, while the third is an upstream container [9] providing a MySQL database endpoint. SWAN makes use of a container that runs JupyterHub for user authentication and management of users' sessions. As SWAN uses EOS storage and software from CVMFS, two additional containers with special privileges run aside to provide the required access to storage and software via fuse mounts. When a user accesses the SWAN service, another container (labeled as "Single-user Jupyter Session" in Fig. 1) running the



**Figure 1.** ScienceBox Architecture. EOS, CERNBox, and SWAN are shown as decoupled blocks with tight interactions to provide end users with a fully-fledged working environment. CVMFS delivers the software required by SWAN. Web frontends of CERNBox and SWAN are directly accessible by end users, while a sync client can be configured to use ScienceBox as the remote server for synchronization.

Jupyter server for the user is spawned. This is done as a security measure (users actions are confined inside the container) and to provide full isolation among users.

End users can access the services provided by ScienceBox either by using their web browser and accessing the web frontends of CERNBox and SWAN or by installing on their devices (e.g., laptops, mobile) the CERNBox synchronization client. EOS can be accessed directly by installing the fuse client on users' machines or by configuring ScienceBox to expose relevant ports for native `xrootd` protocol communication.

### 2.3 Supported Deployment Models

ScienceBox can be installed<sup>1</sup> on a variety of machines, ranging from personal laptops, to clusters of Virtual Machines (VMs), to container orchestration services managed by cloud providers. What makes this possible is the fact that ScienceBox makes use of immutable Docker images that can be used in different contexts. At the time of writing, two main deployment contexts are supported:

- A one-click deployment that allows the installation of ScienceBox in a completely automated fashion via a simple script that ensures all the required software is available (or installs it, otherwise) and runs the required containers. This deployment mode is based on `docker-compose` and allows the execution of all the services on a single machine by instantiating 14 different containers each providing one building block (e.g., EOS MGM, CERNBox web frontend, etc.) of the whole package. The one-click deployment is suitable for demonstration purposes or for providing service to a limited number of users.
- A scalable production deployment that allows deploying ScienceBox over a cluster of machines. In this scenario, ScienceBox leverages the functionalities provided by Kubernetes, a container orchestration engine that manages the scheduling, the life cycle, and the external resources (e.g., storage volumes, network connectivity and addressing, etc.) of containers.

<sup>1</sup>Documentation and installation instruction are available at <https://sciencebox.web.cern.ch>.

ScienceBox services and containers are engineered to provide the ability to seamlessly scale-out storage and computing according to the needs. For instance, it is possible to run multiple copies of the EOS FST container and grant them access to different block storage devices to scale-out storage. Similarly, it is possible to support multiple concurrent SWAN users sessions by adding more machines to the cluster and by replicating the containers providing access to EOS and CVMFS. Kubernetes also allows to run multiple instances of the same container for load-balancing purposes and to achieve high availability. Kubernetes, indeed, allows defining health-checks on each container to assess whether the service running inside is operating as expected or not. In the latter case, it automatically directs the traffic to healthy containers and restarts the failing one.

### 3 Use Cases

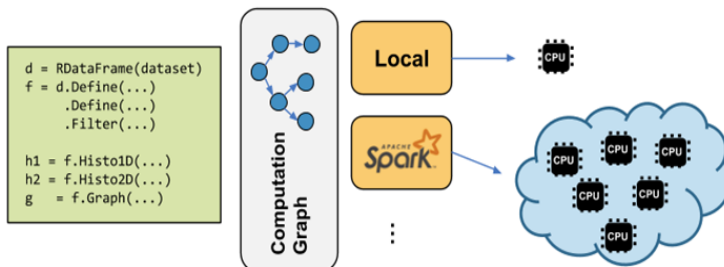
ScienceBox has proven to be successful for diverse use cases and in various deployment scenarios. In what follows, we report on ScienceBox deployments for HEP analysis on commercial cloud providers (Sec. 3.1) and for education and outreach (Sec. 3.2).

#### 3.1 TOTEM Experiment analysis on the Helix Nebula Science Cloud

In 2018, ScienceBox was at the core of an investigation on Big Data tools for processing TOTAL Elastic and diffractive cross section Measurement (TOTEM) [10] experiment data. In this context, ScienceBox was deployed on the Open Telekom Cloud (OTC) through the Helix Nebula Science Cloud (HNSciCloud) project [11], an initiative targeting the procurement of cloud resources from commercial providers and publicly-funded science clouds.

The deployment comprises a full ScienceBox installation in its Kubernetes-based flavor (see Sect. 2.3) complemented by an Apache Spark cluster running on plain VMs. Overall, 2,400 CPUs were available (out of which 2,000 were used by Spark), 10 TB of memory, and 22 TB of block storage. Network connectivity consisted of 10 Gbps links across all the VMs in the OTC datacenter, while uplink to the Internet was granted through several 1 Gbps links and public IP addresses for web frontends of deployed services.

After installation and testing, access was granted to a group of TOTEM scientists who actively used the deployment for 6 months in place of the production service at CERN. They worked on a TOTEM dataset of 4.7 TB that was copied to the EOS instance on the HNSciCloud and re-implemented their analysis workflow using RDataFrame [12], a new declarative



**Figure 2.** RDataFrame interface to computing resources. A pseudo-code snippet shows the high-level interface provided by RDataFrame, which internally determines how to distribute the computing tasks according to the available resources. Users can run their analysis on a small set of CPUs or at scale leveraging the Spark resources and task distribution layer without further modifications to the code.

interface of the ROOT [13] analysis framework. `RDataFrame` provides a high-level declarative interface that allows for implicit parallelization of computing tasks by partitioning and distributing the workload to multiple Spark worker nodes, as depicted in Fig. 2.

SWAN was used as the platform to perform HEP analysis. SWAN allows the exploitation of Spark resources through a specialized connector that interfaces with the Spark scheduler and monitors the progressions of the tasks inside the SWAN interface. With this setup, TOTEM scientists validated the obtained physics results and demonstrated that it is possible to reduce the time required for the final steps of their analysis by a factor 280x when compared to single core execution (i.e., from 8 hours to less than 2 minutes). We invite the interested reader to refer to [14] for further details.

### 3.2 Education and Outreach for Up2U

Up To University (Up2U) [15] is a European project involving companies, research labs, and universities that aims at creating a bridge between high schools and higher education. The objective is to provide tools and technologies commonly used in academia and big science to secondary school students to prepare them for their future careers. The Storage Group of the CERN IT contributed EOS, CERNBox, and SWAN to the project via ScienceBox in order to simplify sharing and re-usage of educational notebooks among students, teachers, and scientists, and to foster the utilization of sync&share and data analysis platforms in schools.

This led the deployment of ScienceBox at Poznan Supercomputing and Networking Center (PSNC) in Poland for the production Up2U service as well as at CERN for a pilot service accessible to a number of selected users. In the former case, ScienceBox was deployed on a fully virtualized infrastructure based on OpenStack on top of which Kubernetes was installed. On the other hand, at CERN, EOS was installed on physical hardware and interfaced with CERNBox and SWAN containers deployed in Kubernetes. In both cases, the deployment proved to be stable over 2 years providing service to more than 200 users overall.

## 4 Conclusions and Future Outlook

In this paper we presented ScienceBox, a Docker-based package that allows deploying CERN storage, sync&share and data analysis services on a single machine for demonstration purposes or at a scale across multiple servers. It proved to be suitable for deployment on bare-metal hosts, on VMs, or on mixed infrastructures and for diverse use cases, ranging from HEP analysis to education and outreach.

We are currently investigating the feasibility of running production services at CERN in containers by leveraging on the experience gained with ScienceBox. We envisage to introduce containers gradually, starting with instances for testing and development. This would allow us to accumulate further operational experience with container technologies and with the management of their life cycle, from the build phase to distribution and deployment, from the update of a live system to performance assessment and emergency response on incidents.

Also, we foresee ScienceBox to evolve into a modular and fully customizable bundle where each service component can be deployed independently from the others through Helm charts. This provides all-round configuration flexibility and the ability to install on-premises the complete stack of ScienceBox services or only a subset of them. In this scenario, the ease of deployment provided by container technologies and the modular architecture of ScienceBox would foster the distribution of open-source scientific software across multiple institutions to increase the interoperability beyond the borders of single clouds.

Lastly, ScienceBox has been chosen as the reference deployment model for CS3MESH4EOSC [16], a European project that aims at creating a federation of data and

higher-level services across the private clouds of research centers, science laboratories, and National Research & Education Networks. The goal of the project is to provide friction-free collaboration between European researchers, students, scientists, and engineers leveraging on established technology enablers: Open Cloud Mesh to interconnect cloud-based sync&share services and EduGAIN to access identity federations. ScienceBox will, therefore, support such technologies to be fully interoperable with existing services and to be easily integrated with on-site applications, storage, and computing systems.

## Acknowledgements

This work was supported in part by the Polish Ministry of Science and Higher Education, grant DIR/WK/2018/13.

## References

- [1] K. Matthias, S. Kane, "Docker: Up & Running - Shipping Reliable Containers in Production", O'Reilly Media, 2018.
- [2] ScienceBox, <https://sciencebox.web.cern.ch>
- [3] Amazon Web Services, <https://aws.amazon.com>
- [4] Open Telekom Cloud, <https://open-telekom-cloud.com/en>
- [5] A. J. Peters, E. A. Sindrilaru, G. Adde, "EOS as the present and future solution for data storage at CERN", Journal of Physics: Conference Series, Vol. 664, 2015.
- [6] L. Mascetti, H. Labrador, M. Lamanna, J. T. Mościcki, A. J. Peters, "CERNBox + EOS: end-user storage for science", Journal of Physics: Conference Series, Vol. 664, 2015.
- [7] D. Piparo, E. Tejedor, P. Mato, L. Mascetti, J. T. Mościcki, M. Lamanna, "SWAN: A service for interactive analysis in the cloud", Future Generation Computer Systems, Vol. 78, 2016.
- [8] J. Blomer, C. Aguado, P. Buncic, H. Artem, "Distributing LHC application software and conditions databases using the CernVM file system", Journal of Physics: Conference Series, Vol. 331, 2011.
- [9] MariaDB – Free and open-source implementation of MySQL under the GNU General Public License, [https://hub.docker.com/\\_/mariadb](https://hub.docker.com/_/mariadb)
- [10] The TOTEM Experiment: Total Cross Section, Elastic Scattering and Diffraction Dissociation at the LHC, <http://totem.web.cern.ch>
- [11] M. Gasthuber, H. Meinhard, R. Jones, "HNSciCloud - Overview and technical Challenges", Journal of Physics: Conference Series, Vol. 898, 2017.
- [12] E. Guiraud, A. Naumann, and D. Piparo, "TDataFrame: functional chains for ROOT data analyses", <https://doi.org/10.5281/zenodo.260230>, 2017.
- [13] I. Antcheva et al., "ROOT - A C++ Framework for Petabyte Data Storage, Statistical Analysis and Visualization", Computer Physics Communications, Vol. 108, 2009.
- [14] V. Avati et al., "Big Data Tools and Cloud Services for High Energy Physics Analysis in TOTEM Experiment", 2018, IEEE/ACM International Conference on Utility and Cloud Computing Companion (UCC Companion), Zurich, pp. 5-6.
- [15] Up to University: Bridging the gap between schools and universities through informal education, <https://up2university.eu/>
- [16] Science Mesh : Interactive and agile/responsive sharing mesh of storage, data, & applications for the European Open Science Cloud (CS3MESH4EOSC), <https://cordis.europa.eu/project/id/863353>