

Experience deploying an analysis facility for the Rubin Observatory's Legacy Survey of Space and Time (LSST) data

Gabriele Mainetti^{1,*}, Fabio Hernandez¹, Fabrice Jammes², and Quentin Le Boulc'h¹

¹CNRS, CC-IN2P3, 21 avenue Pierre de Coubertin, CS70202, 69627 Villeurbanne CEDEX, France

²Université Clermont Auvergne, CNRS/IN2P3, Laboratoire de Physique de Clermont, F-63000 Clermont-Ferrand, France

Abstract. The Vera C. Rubin Observatory is preparing for the execution of the most ambitious astronomical survey ever attempted, the Legacy Survey of Space and Time (LSST). Currently in its final phase of construction in the Andes mountains in Chile and due to start operations in 2025 for 10 years, its 8.4-meter telescope will nightly scan the southern sky and collect images of the entire visible sky every 4 nights using a 3.2 Gigapixel camera, the largest imaging device ever built for astronomy. Automated detection and classification of celestial objects will be performed by sophisticated algorithms on high-resolution images to progressively produce an astronomical catalog eventually composed of 20 billion galaxies and 17 billion stars and their associated physical properties.

In this paper, we briefly present the infrastructure deployed at the French Rubin data facility (operated by IN2P3 computing center, CC-IN2P3) to deploy the Rubin Science Platform, a set of web-based services to provide effective and convenient access to LSST data for scientific analysis. We describe the main services of the platform, the components that provide those services and our deployment model. We also present the Kubernetes-based infrastructure we are experimenting with for hosting the LSST astronomical catalog, a petabyte-scale relational database developed for the specific needs of the project.

1 Introduction

This article outlines the tools being deployed by the CNRS/IN2P3 Computing Center¹ (CC-IN2P3) to provide researchers a convenient way to explore and analyse the data that will be produced as part of the Legacy Survey of Space and Time (LSST) by the Vera C. Rubin Observatory.

The Vera C. Rubin Observatory is in the final stages of its construction on the Chilean Andes. Equipped with an 8.4-meter telescope and the largest digital camera ever built (3.2 gigapixel), it will execute the most ambitious survey ever attempted: the Legacy Survey of Space and Time (LSST) collecting imagery covering the entire visible sky every four nights.

The four main science goals of LSST are to investigate dark matter and dark energy, map the Milky Way, take an inventory of the solar system and study the transient sky [1]. To achieve those goals, the Observatory will produce, by the end of the survey after 10 years of

*e-mail: gabriele.mainetti@cc.in2p3.fr

¹<https://cc.in2p3.fr>

operation, an astronomical catalog containing 37 billion objects (20 billion galaxies, 17 billion stars) as well as a set of 5.5 million images. Science-ready images and the astronomical catalog will be regularly delivered by the Observatory to science collaborations.

This paper is structured as follows. The section 2 introduces the role of CC-IN2P3 as a data facility for the Rubin Observatory. Section 3 describes the analysis facility, the technology behind it and the resources deployed by the CC-IN2P3 to run it. The following sections cover the analysis facility main components and, in particular, section 4 describes the database component and section 5 the data analysis platform.

2 CC-IN2P3 as a Rubin Observatory Data Facility

Image data recorded by the Rubin Observatory will be processed by 3 centers, one in the USA (US Data Facility - USDF²) and the other two in Europe (UK Data Facility - UKDF³ and France Data Facility - FrDF), see Figure 1.

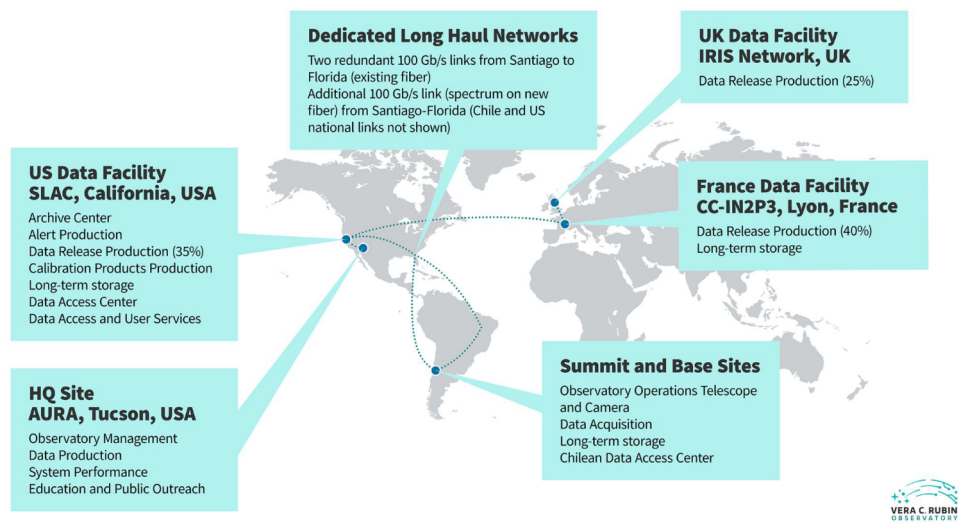


Figure 1: Images flow from the Summit Site where the telescope is located in Chile to the three Rubin Data Facilities which collectively provide the computational capacity for processing the images taken by the Observatory for the duration of the survey.

FrDF is located in Lyon, France at the CC-IN2P3 and will provide the computing and storage resources to annually process 40% of the raw images recorded since the beginning of the survey, with 35% provided by the USDF and 25% by the UKDF. In addition, CC-IN2P3 is preparing to provide long-term storage for the entire raw images data set as well as for selected data products. By the end of the survey, the size of the Rubin data sets could reach several hundred petabytes, with 15 PB for the final version of the astronomical catalog alone.

²USDF is located at SLAC National Accelerator Laboratory, <https://www.slac.stanford.edu>

³Supported by the IRIS Infrastructure <https://www.lsst.ac.uk> and <https://www.iris.ac.uk>

3 The analysis facility

The Rubin Observatory's data analysis facility deployed at CC-IN2P3 is designed with three main objectives: to provide access to Rubin data, to integrate the working environment with the rest of the site's environment and to reach high levels of scalability and resilience.

3.1 Data Access

Our main objective is to provide researchers with an intuitive platform providing a convenient and effective way to access and to analyse the extensive survey data (both images and catalog) generated by the Rubin Observatory. This objective is achieved with two major components of the platform:

1. **Qserv**: Qserv [2, 3] is an open source Massively Parallel Processing database which hosts the Rubin astronomical catalog (see Section 4 below for details), and
2. **Rubin Science Platform (RSP)**: the RSP [4] is an online interactive analysis hub, allowing researchers to access and analyse the data via a user-friendly web-based interface, see Section 5 for more details.

3.2 Integration with CC-IN2P3 Environment

We want the analysis facility to be seamlessly integrated into the existing CC-IN2P3 working environment to make easy for the researchers the transition between the platform and their familiar environment. This integration involves many aspects, including authentication and transparent access to all the storage areas individual users can usually access from a terminal session or from the batch farm, including their \$HOME directory and other project-specific disk-based storage areas. Site-wide single sign on authentication at CC-IN2P3 is based on Keycloak [5].

3.3 Scalability and Resilience

The platform is designed to be both scalable and resilient to ensure that it can adapt to evolving demands linked to the large volumes of astronomical data. The use of Kubernetes (K8S) [6] as the orchestration tool for the platform's containers allows us to effectively manage and scale the components as needed.

For this purpose, CC-IN2P3 provides two K8S clusters dedicated to the Qserv and the Rubin Science Platform: one used as test-bench is based on OpenStack [7] virtual machines and another one used for production on top of bare-metal⁴, see Table 1 for a summary.

CC-IN2P3 provides also 4 data transfer nodes to share with the community catalog data ready for ingestion into Qserv via Caddy [8] HTTP servers; these servers are also used internally to populate the Qserv databases deployed at the site (see Section 4.2).

4 Qserv

The astronomical catalog produced by LSST will include the physical properties of 20 billion galaxies and 17 billion stars, resulting in a total of 15 petabytes of data by the end of the survey. The Qserv database management system [2, 3] has been specifically developed by Rubin project members at SLAC, with contributions from IN2P3, to handle this large volume

⁴The bare metal cluster is composed of DELL PowerEdge R440 and DELL PowerEdge R540 servers

Table 1: Details of the components of the Kubernetes clusters dedicated to the Rubin analysis platform at CC-IN2P3.

Cluster	Production	Test-bench
Infrastructure	Bare-metal	OpenStack
Total Nodes	25	8
Qserv Nodes	17	3
RSP Nodes	5	1
Local Storage	50 TB	1 TB

of data and to respond to the specific data extraction requests scientists typically perform against astronomical catalogs.

Qserv is a shared-nothing Massively Parallel Processing Relational (SQL) Database: this means the processing is split among servers and a leader node handles communications with each of the individual nodes as a map-reduce process. Nodes don’t share resources with other nodes. A representation of this architecture is shown in Figure 2.

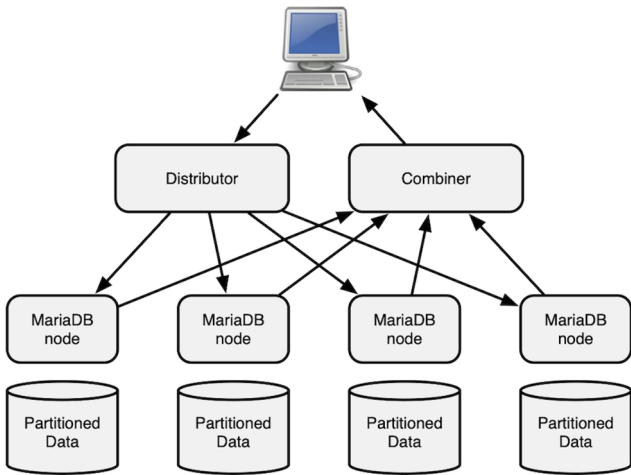


Figure 2: Shared-nothing database architecture, from [9]

Developed as an open source software project, it includes data partitioning over the celestial sphere’s region of equal area, data replication to ensure resilience and high-availability, shared scanning to reduce the total I/O load, and sciSQL [10] User Defined Functions (UDFs) to simplify spherical geometry, statistics and photometry queries.

4.1 Deployment: qserv-operator

Deployment of Qserv is based on Kubernetes operator SDK [11], a framework to manage and automate the deployment of complex applications. This approach makes the installation of Qserv really easy because only two commands are necessary⁵:

⁵A demo of the deployment is available at <https://is.gd/FK62Wa>.

```
$ kubectl apply -f manifest/operator.yaml
$ kubectl apply -k manifest/<instance>
```

where `<instance>` refers to the customization needed for a specific infrastructure. The `qserv-operator` [12] code is available in GitHub.

4.2 Data ingestion: `qserv-ingest`

Qserv has a powerful distributed data ingestion algorithm integrated: `qserv-ingest`. The ingestion workflow’s schema is shown in Figure 3; `qserv-ingest` uses Argo Workflow [13] to orchestrate the ingestion tasks which concurrently execute on all nodes in the cluster. The workflow’s input data to be loaded into the catalog’s tables must be previously partitioned (using a partitioner included in Qserv) and stored in CSV format.



Figure 3: Schematic view of the data ingestion process into Qserv database.

Each transaction can handle a significant number of files for ingestion. Asynchronous REST requests are issued to drive the ingestion of each input file, enabling recovery in case of errors. The workflow also performs validation steps and executes benchmarks on the ingested database.

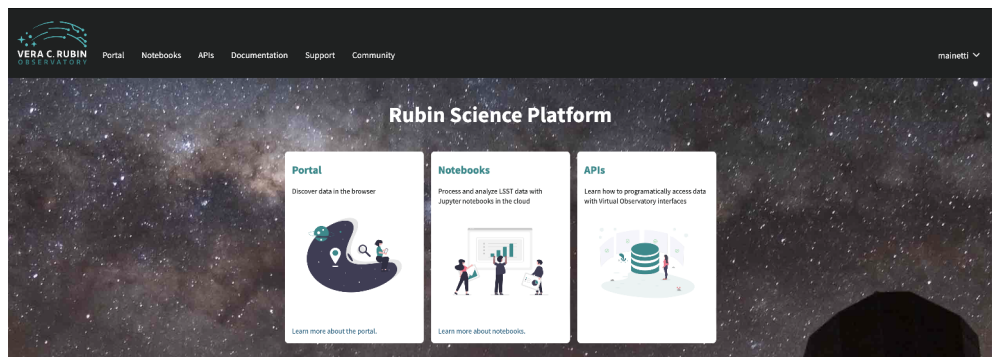
During a test campaign in 2022, `qserv-ingest` was able to ingest 22 million files for an aggregated 40TB of data in 5 hours. At CC-IN2P3 we use `qserv-ingest` as the workflow to ingest data in Qserv.

4.3 Status of Qserv at CC-IN2P3

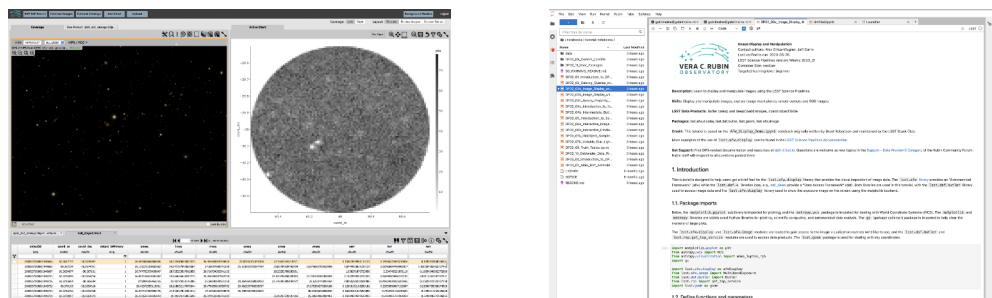
The production Qserv instance at CC-IN2P3 is currently composed of 17 nodes, 15 `qserv-worker` and 2 `qserv-master` nodes. There are 5 astronomical catalogs ingested for an aggregated size of 90 TB, as shown in Table 2.

Table 2: Status of Qserv-managed astronomical catalogs at CC-IN2P3.

Catalog	Size (TB)	Number of rows (Billions)
idf-dp0.2-catalog	36.6	139
frdf-dp0.2-catalog	35.9	121
dp01_dc2_catalogs	1.1	1.7
skysim5000_v1.1.1	13.6	20.5
cosmoDC2_v1.1.4	3.7	5.5



(a) Web portal access



(b) Tables and images viewer

(c) Jupyter notebook platform

Figure 4: A view of the RSP

The DP0.1 and DP0.2 catalogs above are catalogs generated by the LSST data processing campaigns [14, 15] using simulated sky images and they are a preview of what we expect once the Rubin Observatory will start to produce real data.

5 The Rubin Science Platform

The **Rubin Science Platform** (RSP) is a coherent set of web-based applications and services designed by the Rubin project’s Data Management team to offer scientists a unified, near-to-the-data interactive analysis tool. It uses Firefly [16] for data visualisation and plotting and provides access to both Qserv-managed catalogs as well as external catalogs. For advanced, programmatic analysis using Python it integrates a Jupyter [17] notebook platform configured to include the LSST Science Pipelines [18]. Finally, it acts also as gateway to Qserv catalogs for Virtual Observatory⁶ Table Access Protocol [19] compatible tools.

A view of the RSP is shown in Figure 4.

An experimental instance of the Rubin Science Platform deployed at CC-IN2P3 is online and accessible to registered project members at the address <http://data-dev.lsst.eu>.

⁶<https://www.ivoa.net>

5.1 Deployment

Designed as a collection of applications and services, it has been developed by the Rubin Observatory Science Quality and Reliability Engineering⁷ (SQuaRE) team.

The RSP components run on top of Kubernetes clusters and each component is configured via Helm [20] charts. Individual Helm charts are managed via the Phalanx [21] git repository [22].

The use of ArgoCD [23] allows to synchronize these application deployment manifests into the Kubernetes cluster of each environment.

Authentication is managed via the **gafaelfawr** ([24], [25], [26]) identity management application, which supports LDAP [27] and OpenId Connect [28] providers (e.g. Github). CC-IN2P3's RSP instance uses Keycloak [5] as OpenID Connect (OIDC) provider and LDAP as provider of the user's group identifiers.

In addition to gafaelfawr, there are 4 core applications providing key functionality for other applications:

- **argocd** for deployment orchestration,
- **cert-manager** for certificates management,
- **ingress-nginx** for traffic routing,
- **vault-secrets-operator** for secret management

Seventeen components are currently activated on the CC-IN2P3 RSP instance, some of them requiring adaptations to the specifics of the CC-IN2P3 environment.

6 Summary

The Rubin Observatory will record a large number of images of the Southern sky over several years. We outlined the role of CC-IN2P3 as one of the three main data facilities that will process these images to regularly produce an updated version of an astronomical catalog and we presented how we deployed an analysis facility on top of Kubernetes, highlighting its two key components: the scalable, shared and resilient database designed for serving the Rubin astronomical catalog and the Rubin Science Platform to perform interactive analysis of the data to be regularly released by the Observatory.

References

- [1] Z. Ivezić et al., *Astrophys. J.* **873**, 111 (2019)
- [2] D.L. Wang et al., *Qserv: A Distributed Shared-Nothing Database for the LSST Catalog*, in *State of the Practice Reports* (Association for Computing Machinery, 2011), SC '11, ISBN 9781450311397, <https://doi.org/fnk9nm>
- [3] *Qserv: petascale distributed database*, <https://github.com/lsst/qserv>
- [4] M. Jurić, D. Ciardi, G.P. Dubois-Felsmann and L.P. Guy, *LSE-319 - LSST Science Platform Vision Document* (2019), Vera C. Rubin Observatory Systems Engineering Technical Note, <https://lse-319.lsst.io>
- [5] *KeyCloak*, <https://www.keycloak.org>
- [6] *Kubernetes*, <https://kubernetes.io>
- [7] *OpenStack*, <https://www.openstack.org>
- [8] *Caddy, an open source web server*, <https://caddyserver.com>

⁷<https://github.com/lsst-sqre>

- [9] F. Mueller et al., *DMTN-243 - Qserv: A Distributed Petascale Database for the LSST Catalogs* (2022), Vera C. Rubin Observatory Data Management Technical Note, <https://dmtn-243.lsst.io/>
- [10] *sciSQL 0.3: Science tools for mysql*, <https://smonkewitz.github.io/scisql/>
- [11] *Operator Framework for kubernetes*, <https://operatorframework.io>
- [12] *qserv-operator: a qserv operator for Kubernetes*, <https://github.com/lsst/qserv-operator>
- [13] *Argo Workflow a workflow engine for orchestrating parallel jobs on kubernetes*, <https://argoproj.github.io/argo-workflows>
- [14] W. O'Mullane, *RTN-001 - Data Preview 0: Definition and planning*. (2021), Vera C. Rubin Observatory Technical Note, <https://rtn-001.lsst.io/>
- [15] Q. Le Boulc'h, F. Hernandez, G. Mainetti, *The Rubin Observatory's Legacy Survey of Space and Time DP0.2 processing campaign at CC-IN2P3*, in *Proc. of CHEP 2023* (to appear)
- [16] *Firefly a browser-based interactive particle visualization app*, <http://firefly-viz.com>
- [17] *Jupyter*, <https://jupyter.org>
- [18] J. Bosch et al., *An Overview of the LSST Image Processing Pipelines*, in *Astronomical Data Analysis Software and Systems XXVII*, edited by P.J. Teuben, M.W. Pound, B.A. Thomas, E.M. Warner (2019), Vol. 523 of *Astronomical Society of the Pacific Conference Series*, p. 521, 1812.03248
- [19] P. Dowler, G. Rixon, D. Tody, M. Demleitner, *Table Access Protocol Version 1.1*, IVOA Recommendation 27 September 2019 (2019), <https://10.5479/ADS/bib/2019ivoa.spec.0927D>
- [20] *Helm the package manager for kubernetes*, <https://helm.sh>
- [21] *Phalanx a gitops repository for rubin observatory's kubernetes environments*, <https://phalanx.lsst.io>
- [22] *Phalanx*, <https://github.com/lsst-sqre/phalanx>
- [23] *ArgoCD a declarative, gitops continuous delivery tool for kubernetes*, <https://argo-cd.readthedocs.io>
- [24] *Gafaelfawr rsp identity management*, <https://github.com/lsst-sqre/gafaelfawr>
- [25] R. Allbery, *DMTN-224 - RSP identity management implementation strategy* (2023), Vera C. Rubin Observatory Data Management Technical Note, <https://dmtn-224.lsst.io/>
- [26] R. Allbery, *DMTN-234 - RSP identity management design* (2023), Vera C. Rubin Observatory Data Management Technical Note, <https://dmtn-234.lsst.io/>
- [27] *LDAP*, <https://ldap.com>
- [28] *OpenId*, <https://openid.net>