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

CHEP2023 Norfolk 8-12 May 2023

F. Hernandez, F. Jammes, Q. Le Boulc’h, G. Mainetti
Outline

1. The Rubin Observatory and LSST
2. The analysis facility
3. The CC-IN2P3 infrastructure
4. Conclusions and Perspectives
The Rubin Observatory and LSST
The Vera C. Rubin Observatory
• Cerro Pachón @ Chili (2647m asl)
• Main mirror 8.4m Ø
• 9.6 deg² Field of View
• 3.2 G pixels camera
• f/1.234 aperture

The Legacy Survey of Space and Time (LSST)
• Composition of the Universe: Dark Energy and Dark Matter
• The Solar System inventory
• The changing sky: variable and transient objects observation
• The Milky Way cartography

- 37 Billion objects (20B Galaxies, 17B Stars) catalog
- Each object observed more than 800 times
- Full visible sky survey in 3 nights
- One pointing each 38s, 10M galaxies per pointing

Images Source: Rubin Observatory
The analysis facility
The analysis facility

- **Objectives:**
  - Provide researchers with a platform to easily access/analyze survey data (images and catalogs)
  - Integrate it with the CC-IN2P3 (e.g., authentication, $HOME and other file systems, …) for a smooth transition between environments
  - Deploy a scalable and resilient platform

- **Technology:**
  - **Kubernetes**
  - Open source development: [https://github.com/lsst-sqre](https://github.com/lsst-sqre), [https://www.lsst.io/](https://www.lsst.io/)

- **Two main components:**
  - **Qserv**: the astronomical catalog database
  - **Rubin Science Platform** (RSP): the interactive analysis platform
Qserv

- Developed by SLAC with contributions from IN2P3
- Shared-nothing Massively Parallel Processing Relational Database
- Spherical partitioning with overlap, sciSQL (UDF)
- Shared scans (concurrent query load)
- Data Replication
- 100 % Open Source

---

Qserv design

Relational database, 100% open source
Spatially-sharded with overlaps
Map/reduce-like processing, highly distributed

---

apiVersion: qserv.lsst.org/v1beta1
kind: Qserv
metadata:
  name: qserv
spec:
  queryService:
    type: NodePort
    nodePort: 38904
    storageClassName: "qserv-local-storage"
    storage: "100Gi"
  worker:
    replicas: 15
    replicationResources:
      limits:
        cpu: 36
    tolerations:
      - key: "dedicated"
        operator: "Equal"
        value: "qserv"
        effect: "NoSchedule"

---

- Based on the Kubernetes operator-sdk framework
- 2 commands to deploy it on cloud or bare metal clusters
  - k apply -f manifest/operator.yaml
  - k apply -k manifest/<instance>

https://qserv.lsst.io

Deployment demo: https://is.gd/FK62Wa
Data Ingestion: Qserv Ingest

Tool used for data ingestion
Developed by F. Jammes (IN2P3)

Implementation of a large-scale data loading algorithm:
2M files and ~40TB ingested in 5h

https://github.com/lsst-dm/qserv-ingest

Gabriele Mainetti
Rubin Science Platform

2.2
Rubin Science Platform

• What is it?
  • Web environment for interactive data analysis

• Why?
  • To provide astronomers with an all-in-one tool allowing easy and quick data access/analysis integrating:
    • catalogs and tables viewer (as TOPCAT, STILTS)
    • image viewer and analyzer (as DS9)
    • advanced analysis with LSST python stack via Jupyter
    • gateway to Qserv catalogs (Interoperability) for Virtual Observatory (VO) tools

!!! It’s not (yet?) ChatGPT, so the request must be translated into ADQL format !!!
Developed mainly by the Rubin SQuaRE Team

Configured via Helm Chart (https://phalanx.lsst.io)

One config per data facility and one config per application

Deployed via ArgoCD

4 Core applications:

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

Authentication managed via in-house **gafaelfawr** IDM application (supporting OpenId, LDAP, github, …)

Many others applications available
Rubin Science Platform

Qserv data access

External catalog access

External images access
Jupyter platform

Chart creation

Table and image visualization
The CC-IN2P3 infrastructure
CC-IN2P3 role and resources

- CC-IN2P3 is the French Data Facility (FrDF, with USDF, UKDF)
- Annual processing of 40% of the cumulated data
- Long term storage of selected subsets of the data releases
- Eventually several hundreds Pb stored on disc/tapes, 15 Pb for the astronomical catalog of the last data release

Bare-metal cluster dedicated to Qserv and RSP:
- 25 worker-nodes
- 5 DELL PowerEdge R440, 20 DELL PowerEdge R540
- 3 worker-nodes as K8S control plane
- 17 worker-nodes dedicated to Qserv (via taint)
- RSP worker-nodes:
  - RAM: 256 GB
  - Local storage: 50 TB

OpenStack cluster as test-bench:
- 8 VM
- 3 VM as K8S control plane
- 4 VM dedicated to Qserv
- 1 VM dedicated to the RSP and other
- 4 data transfer nodes to expose Qserv data via Caddy web server
Conclusions and Perspectives
Conclusions and Perspectives

- Rubin-LSST challenge: the amount of data and the computing resources needed to process it
  - “Overview of the distributed image processing infrastructure to produce the Legacy Survey of Space and Time (LSST)” (Track 1, Thursday, 11:30am, F. Hernandez)
  - “The Rubin Observatory’s Legacy Survey of Space and Time DP0.2 processing campaign at CC-IN2P3” (Track 4, Thursday, 2:15pm, F. Hernandez)

- Scalable, shared and resilient database to deal with the astronomical catalog deployed
- Scalable analysis platform to access and analyze data deployed

- Collect the RSP user feedback to improve it with new functionalities
- Complete the RSP integration with the CC-IN2P3 environment
- Scaling up the amount of data in QServ (~55T at the moment)
Backup
RSP Current Status

- Deployed on the production cluster
- 17 applications activated
- Simulated DP02 and DP01 catalogs available
- CC-IN2P3 environment accessible from the notebook platform
- /sps/lsst not yet accessible
- Images not yet available (SPS needed to expose the butler)
- Development mainly US:
  - Use of $ Google services (e.g. GCS)
  - Efforts have been needed to adapt some RSP applications to the CC-IN2P3 environment, but modification proposals well-accepted and integrated
# Qserv Current Status

<table>
<thead>
<tr>
<th>Catalog</th>
<th>Taille (To)</th>
<th># Lignes (Milliards)</th>
</tr>
</thead>
<tbody>
<tr>
<td>idf-dp0.2-catalog</td>
<td>36.6</td>
<td>139</td>
</tr>
<tr>
<td>dp01_dc2_catalogs</td>
<td>1.1</td>
<td>1.7</td>
</tr>
<tr>
<td>skysim5000_v1.1.1</td>
<td>13.6</td>
<td>20.5</td>
</tr>
<tr>
<td>cosmoDC2_v1.1.4</td>
<td>3.7</td>
<td>5.5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Database</th>
<th>#chunks</th>
<th>Data [GB]</th>
<th>in all replicas</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>in unique chunks</td>
<td>regular</td>
</tr>
<tr>
<td></td>
<td>unique</td>
<td>replicas</td>
<td>data</td>
</tr>
<tr>
<td>cosmoDC2_v1.1.4</td>
<td>1730</td>
<td>1744</td>
<td>3569.4</td>
</tr>
<tr>
<td>dp01_dc2_catalogs</td>
<td>1398</td>
<td>1412</td>
<td>915.3</td>
</tr>
<tr>
<td>dp02_dc2_catalogs</td>
<td>1478</td>
<td>1492</td>
<td>31746.3</td>
</tr>
<tr>
<td>skysim5000_v1.1.1_parquet</td>
<td>18738</td>
<td>18752</td>
<td>13171.2</td>
</tr>
<tr>
<td>Total [TB for data]</td>
<td>23344</td>
<td>23400</td>
<td>49.4</td>
</tr>
</tbody>
</table>
Gafaelfawr

- Provides authentication and identity management services.
- It also manages the tokens
- OpenID Connect compatible
- @CC-IN2P3 : Keycloak+LDAP
Qserv Ingest

Transaction 1
Each transaction monitor ingest of ~200 000 files
Handle recovery on error

Transaction 10

Ingest workflow

Kubernetes

Qserv

Ingest controller
Worker 1
Worker 2
Worker 15

CEPH
~2M input files

Web server
Web server
Web server

Each worker load
~130000 files inside MySQL

Legend:
Async REST queries monitoring ingest of 1 file
HTTP query to download 1 file

Source: F. Jammes
Rubin FoV

15x

Source: P. Antilogus