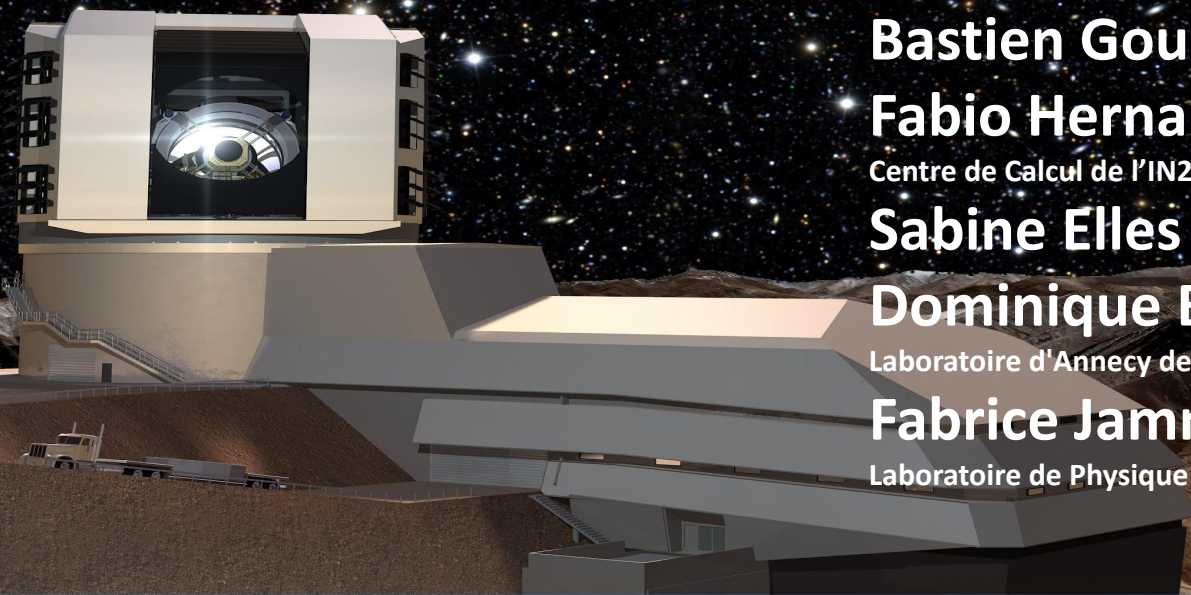


Container Orchestration, Cloud, and Petabytes of Data: The Rubin Observatory Example



Bastien Gounon

Fabio Hernandez

Centre de Calcul de l'IN2P3

Sabine Elles

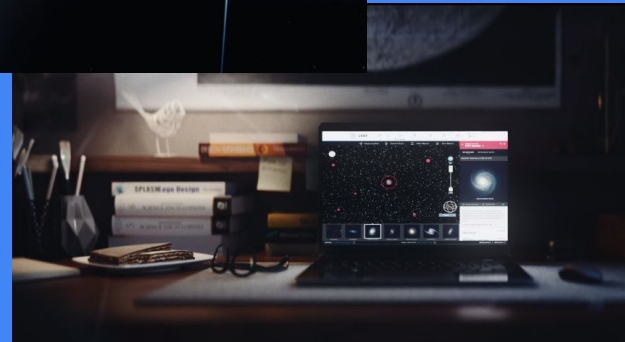
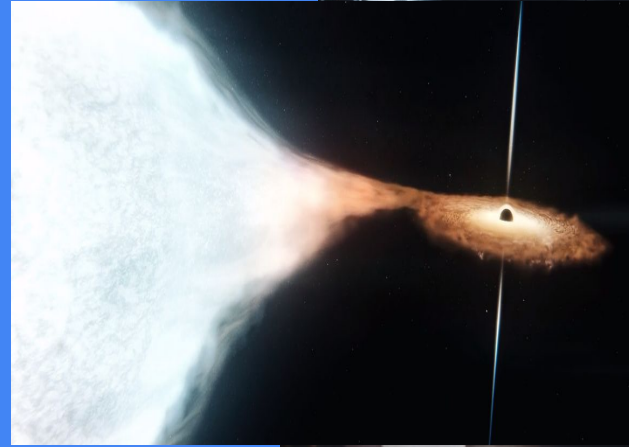
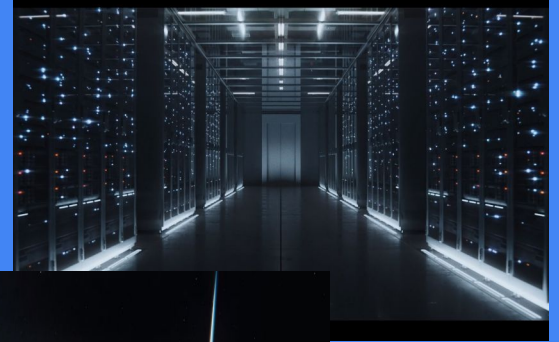
Dominique Boutigny

Laboratoire d'Anney de Physique des Particules

Fabrice Jammes

Laboratoire de Physique de Clermont

- 1 Large Synoptic Survey Telescope
- 2 The largest astronomical catalog
- 3 Cloud-Native: Kubernetes
- 4 Cloud-Native: Gitops & CI
- 5 Cloud-Native: Kubernetes Operators
- 6 Cloud-Native: Storage management
- 7 Cloud-Native: Workflows



A project that makes you dream

A revolutionary telescope

The **largest digital camera** in the world

The **largest celestial catalogs** ever made

Funding

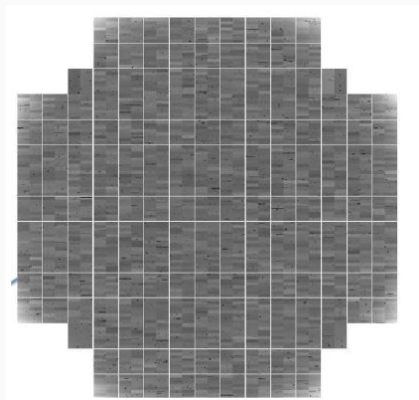
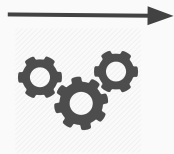
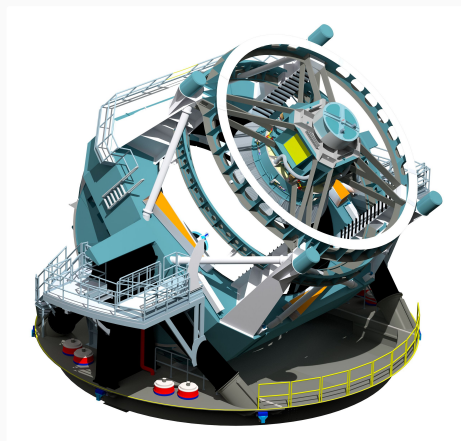
~\$1 billion, 20% dedicated to data management

Key role of CNRS/IN2P3

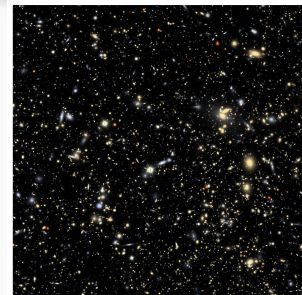
Objective:
Define the nature of dark energy



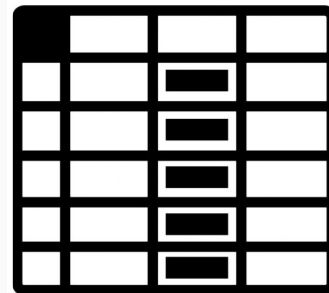
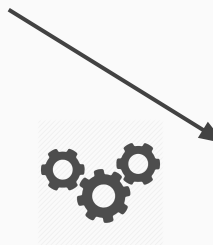
The largest astronomical catalog



Raw data



Processed image



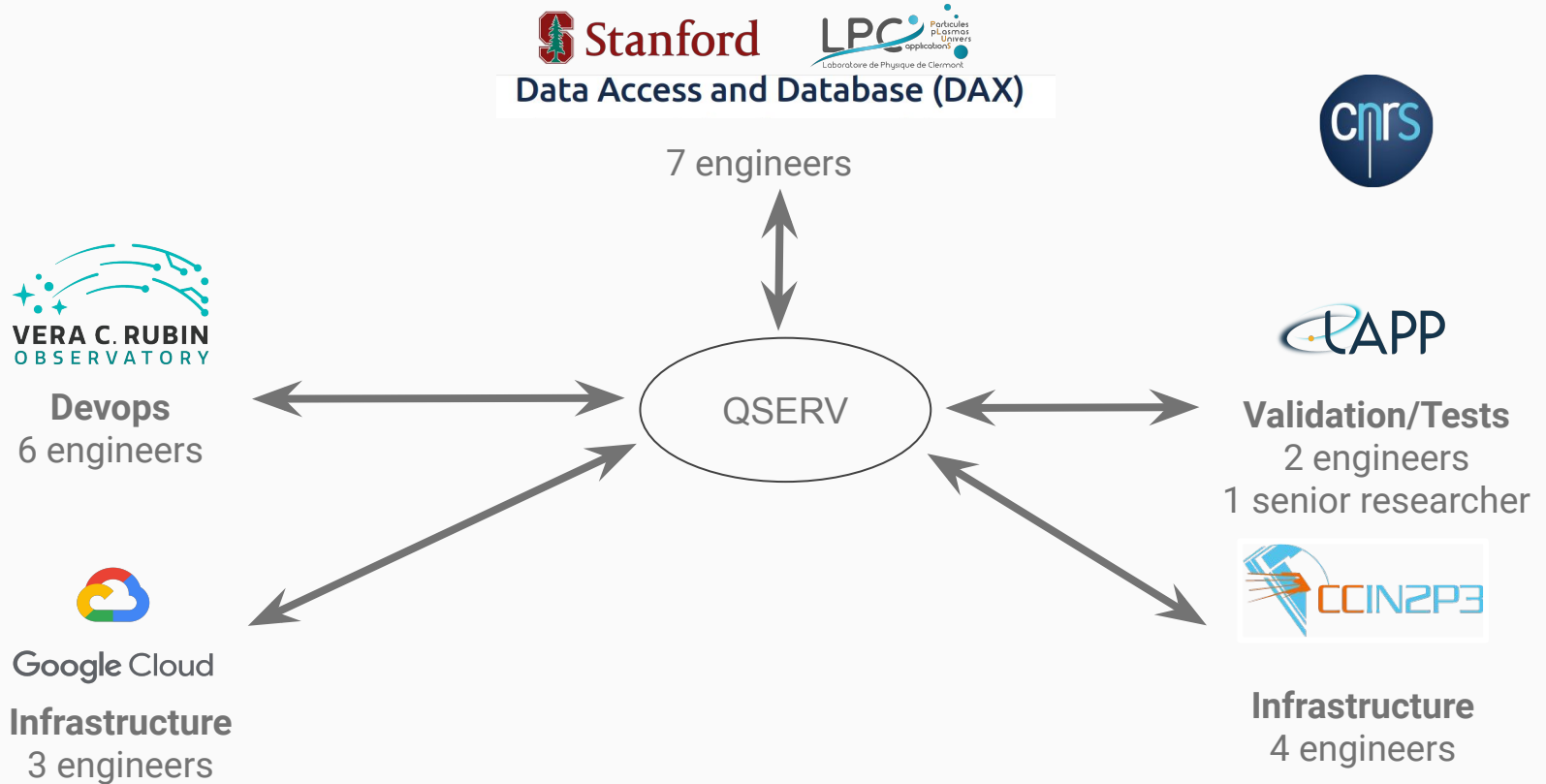
Catalog (stars, galaxies, objects, sources, transients, exposures, etc.)

LSST will produce a catalog of **40 billion galaxies and stars** and their associated physical properties, i.e. **500 PB** of data

Qserv

The Petascale database

International context

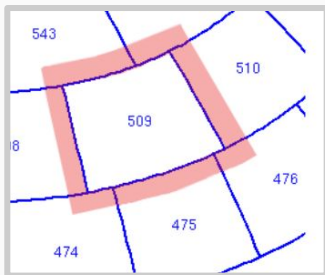
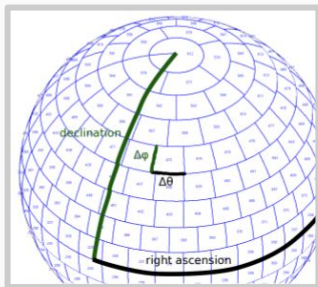


Qserv design



Relational database, 100% open source

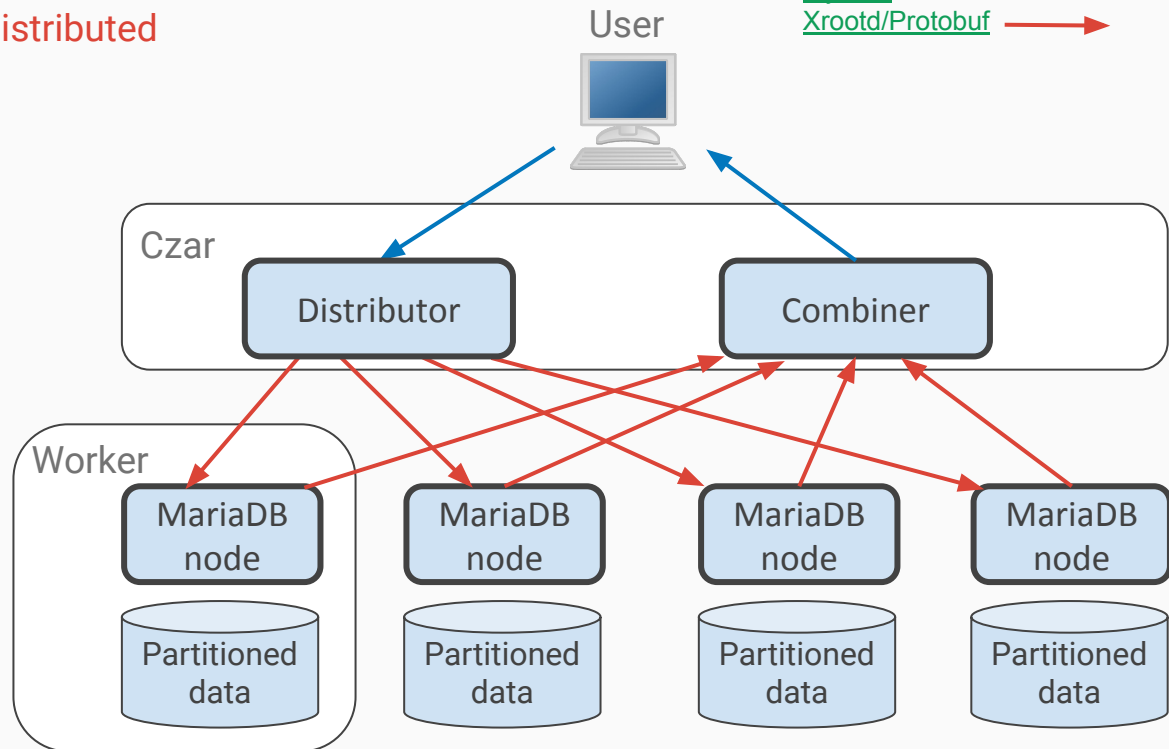
Spatially-sharded with overlaps

Map/reduce-like processing, highly distributed



Legend:

MySQL 
Xrootd/Protobuf 



~1000 workers, 20 chunks/5TB per workers

Highly automated deployment

Targets:

In France

CC-IN2P3 will analyze **50% of the data** stream and provide **access to the entire catalog**

In the US

Google hosts the **Interim Data Facility**

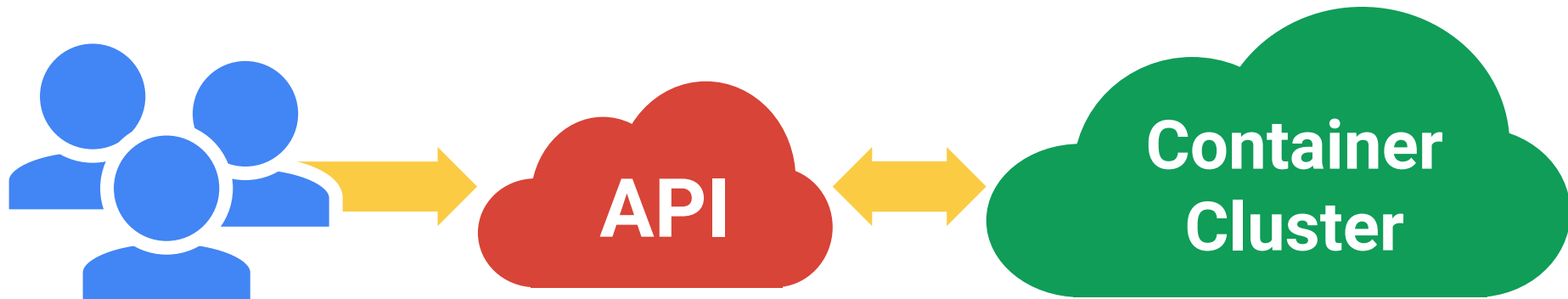
~1000 machines per database instance

Coordination of Rubin Observatory, IN2P3 and Google
Kubernetes accepted by the project and validated for **20%** of the target



Cloud-Native Kubernetes

All you really care about





Portability

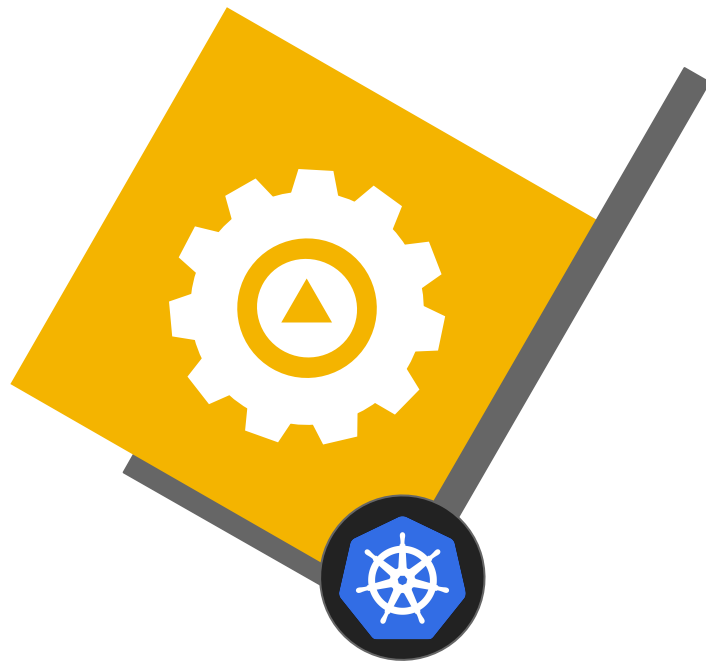
Build your apps on-prem,
lift-and-shift into cloud when you
are ready

Before Kubernetes

~3 months to deploy Qserv inside a new
cluster

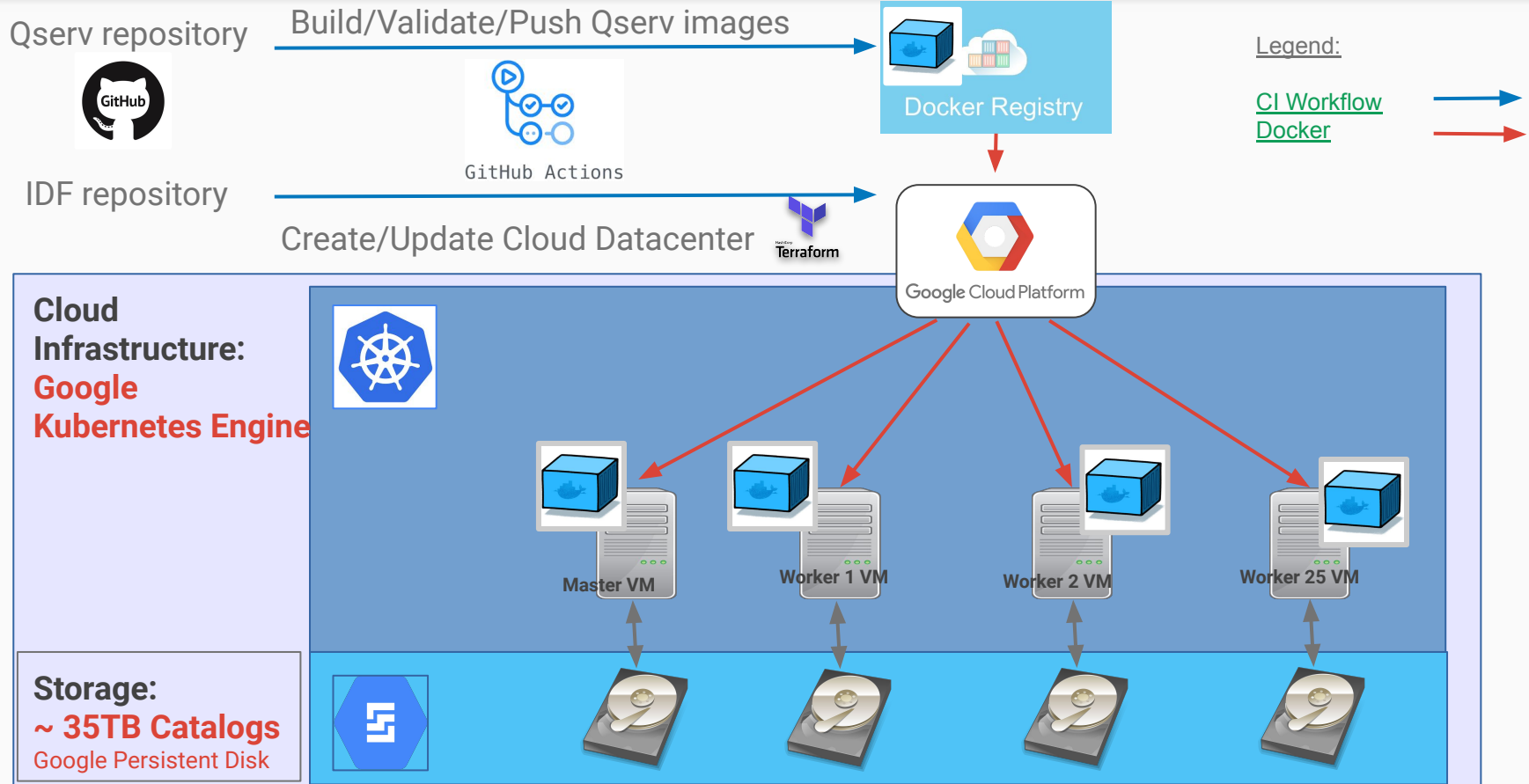
With Kubernetes

5 minutes to 1 day



Cloud-Native Gitops & CI

Automated deployment: Cloud Native



CI in practice: Qserv integration tests

Code Issues Pull requests 1 Actions Security Insights Settings

Workflows New workflow

All workflows

Filter workflow runs

2,475 workflow runs

| Event | Status | Branch | Actor |
|------------------|--|------------------|-------------|
| Tickets/dm 29567 | CodeQL #289: Pull request #28 synchronize by fjammes | tickets/DM-29567 | 6 hours ago |
| Tickets/dm 29567 | CI #787: Pull request #28 synchronize by fjammes | tickets/DM-29567 | 6 hours ago |
| Tickets/dm 29567 | Static code analysis #554: Pull request #28 synchronize by fjammes | tickets/DM-29567 | 6 hours ago |

Build image

Analyze image

Run e2e tests

Push qserv-operator image to public r...

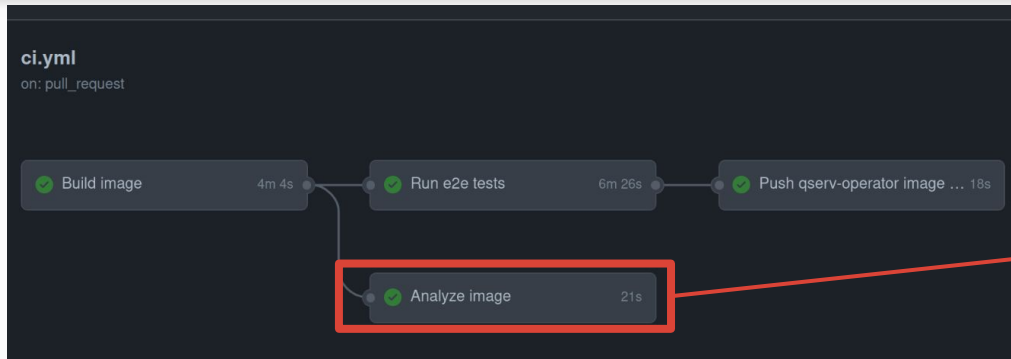
Integration tests with kind/k8s

ci.yml

on: pull_request

```
graph LR; A[Build image 4m 4s] --> B[Run e2e tests 6m 26s]; A --> C[Analyze image 21s]; B --> D[Push qserv-operator image ... 18s]
```

CI in practice: Qserv image scanning



Analyze image

succeeded 6 hours ago in 21s

- > ✓ Set up job
- > ✓ Download image
- > ✓ Load image in local registry
- > ✓ Scan operator image
- > ✓ upload Anchore scan SARIF report
- > ✓ Complete job

anchore

Vulnerability
Scanning &
Policy-Compliance
for Containers

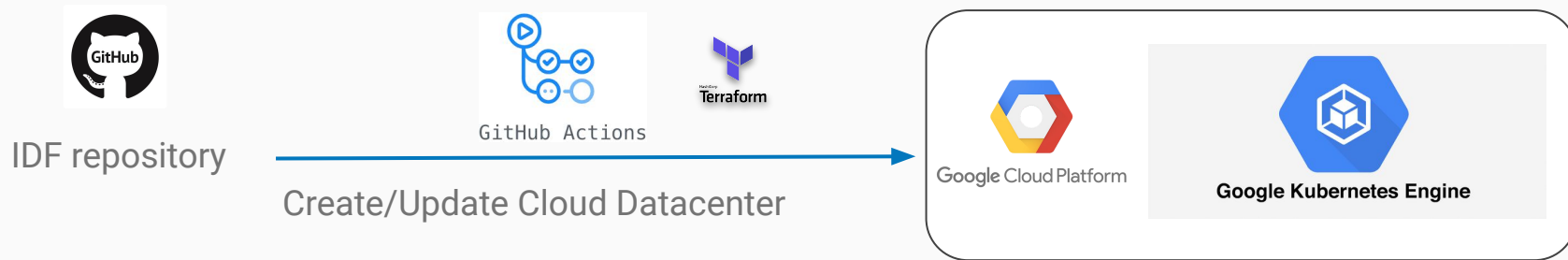
Requests 1 Actions Security Insights Settings

Code scanning

Add more scans

| Latest scan | Pull request | Workflow | Lines scanned | Duration | Result |
|----------------|--------------|----------|----------------|----------|----------|
| 43 minutes ago | #28 | CodeQL | 2.48k / 2.6k ⓘ | 5m 49s | 0 alerts |

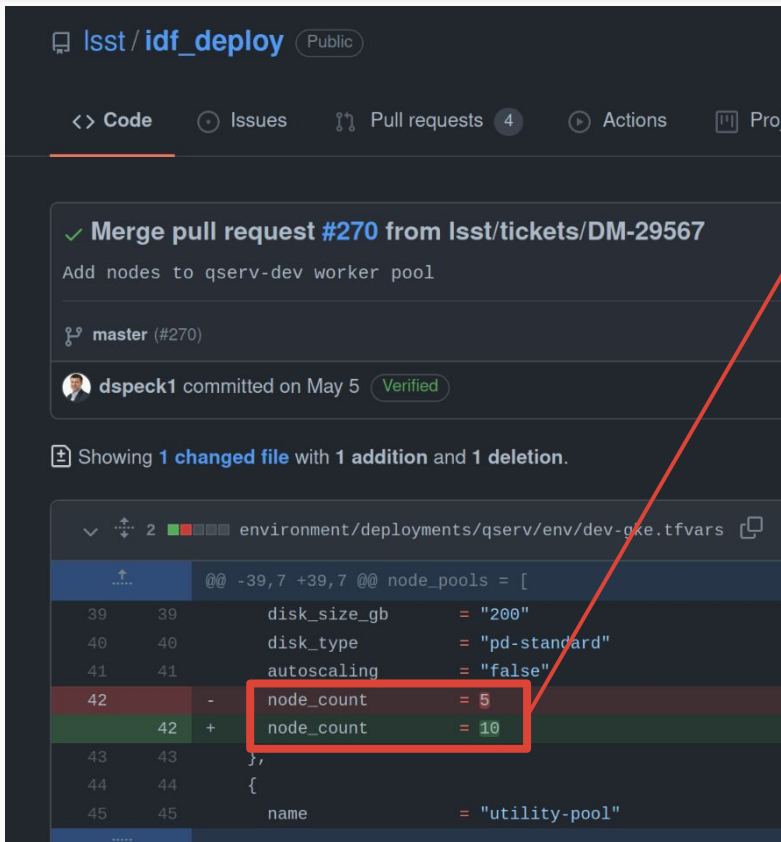
Gitops: CI + IaC



- Delegate access to infrastructure management
- Track who does what on infrastructure
- Recreate infrastructure from scratch
- Ease Kubernetes maintenance/upgrade

Kubernetes is fully managed by Google Cloud / GKE

In practice



Isst / idf_deploy Public

<> Code Issues Pull requests 4 Actions Pro

✓ Merge pull request #270 from Isst/tickets/DM-29567

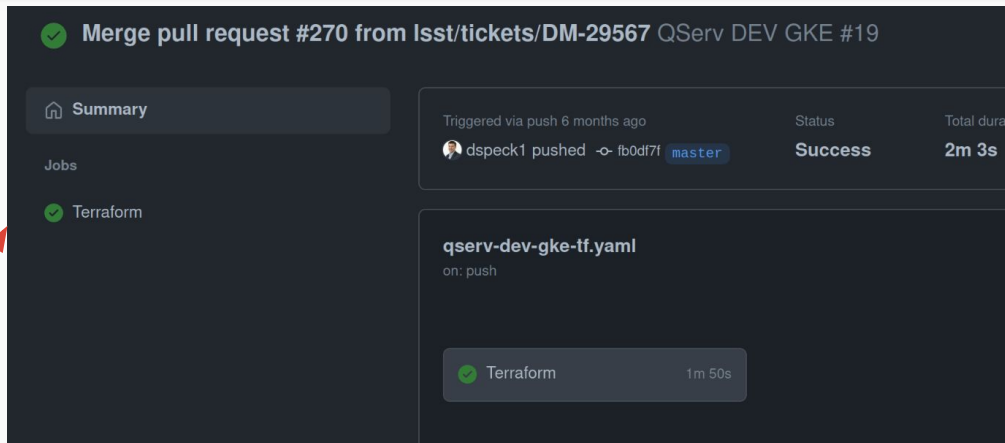
Add nodes to qserv-dev worker pool

master (#270)

dspeck1 committed on May 5 Verified

Showing 1 changed file with 1 addition and 1 deletion.

```
environment/deployments/qserv/env/dev-gke.tfvars
@@ -39,7 +39,7 @@ node_pools = [
39 39     disk_size_gb     = "200"
40 40     disk_type        = "pd-standard"
41 41     autoscaling      = "false"
42 -     node_count       = 5
42 +     node_count       = 10
43 43   },
44 44   {
45 45     name              = "utility-pool"
```



✓ Merge pull request #270 from Isst/tickets/DM-29567 Qserv DEV GKE #19

Summary

Triggered via push 6 months ago Status Total duration

dspeck1 pushed -> fb0d7f1 master Success 2m 3s

Jobs

✓ Terraform

qserv-dev-gke-tf.yaml

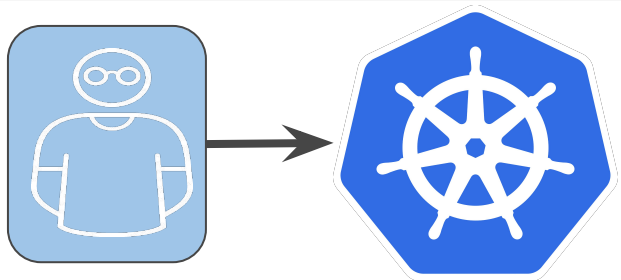
on: push

✓ Terraform 1m 50s

Add five nodes to the GKE cluster
Kubernetes will then allow to easily scale Qserv

Cloud-Native Kubernetes operators

How does an operator works?



Allow to deploy a complex application with only a few lines of yaml

Software Developer
Kubernetes user

K8s API



Kubernetes operator

Native Kubernetes
resources

Custom resource

```
apiVersion: qserv.lsst.org/v1alpha1
kind: Qserv
metadata:
  name: qserv
  namespace: database
spec:
  czar:
    image: qserv/lite-qserv:2021.10.1-rc1
  replicas: 1
  storage: 1Ti
  worker:
    image: qserv/lite-qserv:2021.10.1-rc1
  replicas: 10
  ...
```

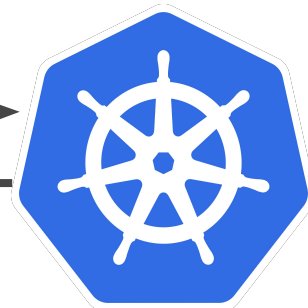
Custom Kubernetes controller

Watch Event

Reconcile

Custom resource definition

here Qserv



Deployments
StatefulSets
Autoscalers
Secrets
Config maps

Qserv is available on operatorHub

<https://operatorhub.io/operator/qserv-operator>



OperatorHub.io

Qserv operator

Create and maintain highly-available Qserv clusters on Kubernetes

Search OperatorHub... Contribute

Home > Qserv operator

Qserv operator

Install

Overview

Qserv Operator manages the full lifecycle of Qserv at scale, in order to ease and fully automate deployment and management of Qserv clusters.

The operator aims to provide the following:

- Basic Install to Qserv components.
- Out-of-box Intra-Cluster discovery support.

Mind you, this is still a work-in-progress implementation.

CHANNEL
alpha

VERSION
2021.9.1-rc1 (Current) ▾

CAPABILITY LEVEL ⓘ

- Basic Install
- Seamless Upgrades
- Full Lifecycle
- Deep Insights
- Auto Pilot

PROVIDER
Vera C. Rubin Observatory

Seamless upgrade is a work in progress

Cloud-Native Storage management

Storage management

GKE: Dynamic storage provisioning

User deploy Qserv instance

Create PVClaims

Google Storage creates automatically PersistentVolume+Google Disks (ssd+hdd)

On-premise:

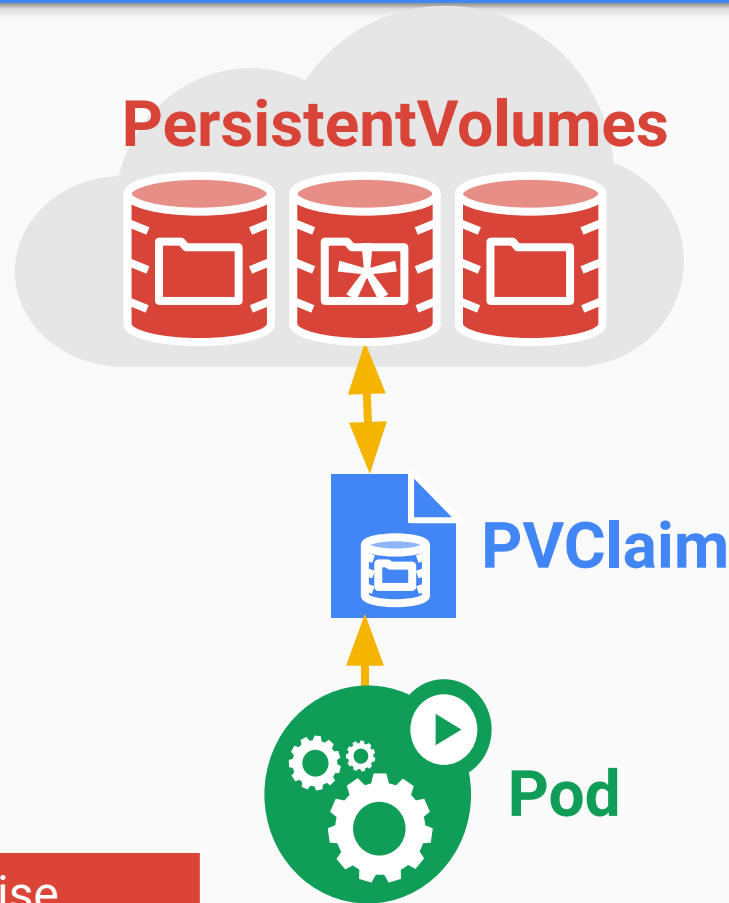
Storage is manually declared to Kubernetes (via PV) and created



Cluster Admin



User



Easier on GKE, but better performance on-premise

Cloud-Native Workflows

A powerful data ingest workflow

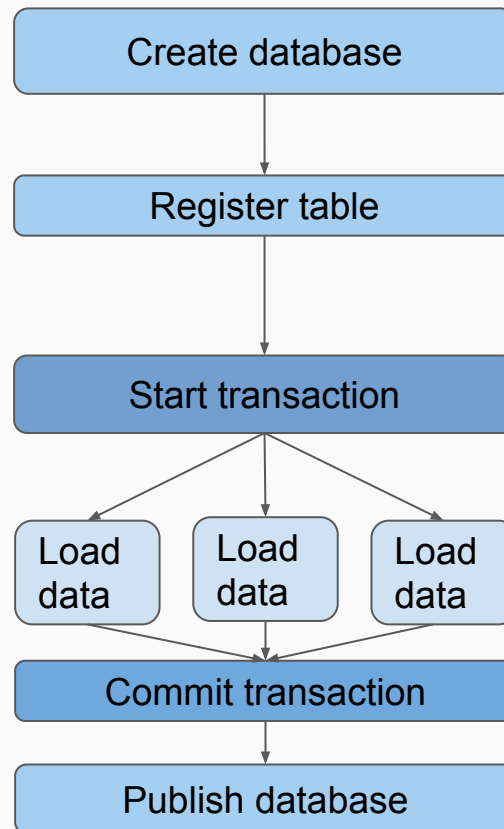
Qserv has a powerful distributed ingest algorithm
Flexible but require orchestrating tasks (DAG)

Argo Workflow project help us a log



Case study 2021: Implementation of a large-scale data loading algorithm

Ingestion of 15000 files and 15TB in 1h30

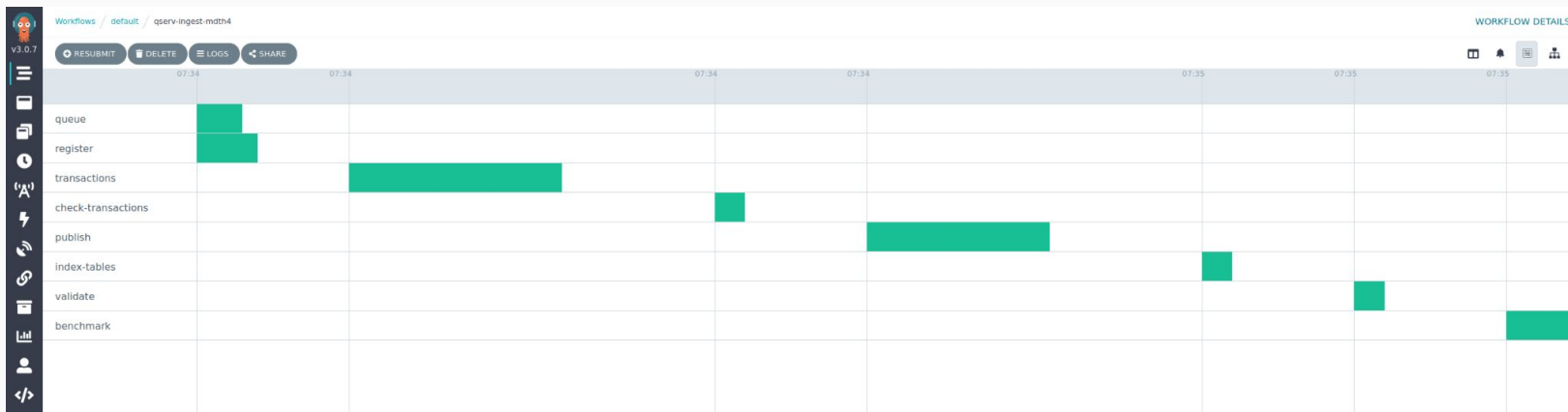


Argo: screenshots



```
fjammes@clrinport18 ~$ argo get @latest | tail -n 15
```

| STEP | TEMPLATE | PODNAME | DURATION |
|----------------------|--------------|-------------------------------|----------|
| ✓ qserv-ingest-mdth4 | main | | |
| ✓ queue | ingest-step | qserv-ingest-mdth4-2075476264 | 3s |
| ✓ register | ingest-step | qserv-ingest-mdth4-964548720 | 4s |
| ✓ transactions | transactions | qserv-ingest-mdth4-1041421248 | 14s |
| ✓ check-transactions | ingest-step | qserv-ingest-mdth4-3195504171 | 2s |
| ✓ publish | ingest-step | qserv-ingest-mdth4-4256901816 | 12s |
| ✓ index-tables | index-tables | | |
| ✓ index-tables | ingest-step | qserv-ingest-mdth4-1866502525 | 2s |
| ✓ validate | ingest-step | qserv-ingest-mdth4-493206715 | 2s |
| ✓ benchmark | benchmark | | |
| ✓ benchmark | ingest-step | qserv-ingest-mdth4-1797710727 | 5s |



Public cloud: pros and cons

Pros

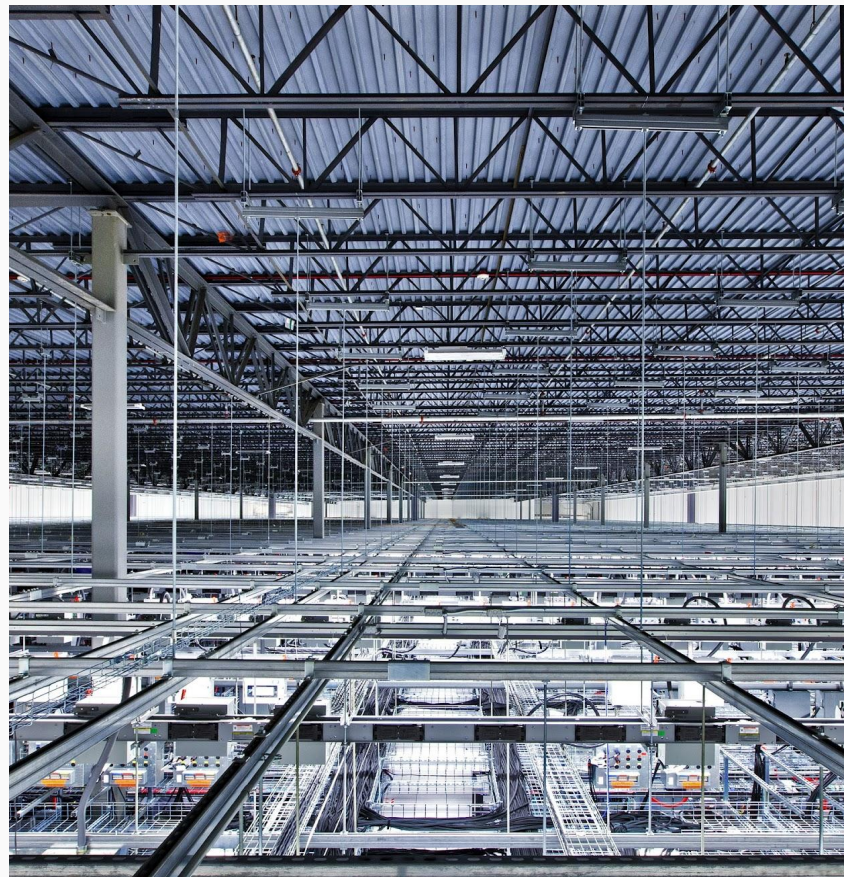
- ★ Flexibility (access, provisioning)
- ★ Excellent support
- ★ Low maintenance
- ★ Cool proprietary features

Cost-effective over time if organizations learn to use and operate it more efficiently

Cons

- ★ Cost difficult to understand
- ★ Vendor lock-in
- ★ Hide Kubernetes internals (black box)
- ★ Run slower than bare-metal (~25%)

The higher the average server load, the less attractive the cloud is financially



- 1 Qserv is going on
- 2 Container orchestration helps a lot
- 3 Commercial cloud is worth considering

Conclusion

Q&A

Bastien Gounon, Fabio Hernandez
Centre de Calcul de l'IN2P3

Sabine Elles, Dominique Boutigny
*Laboratoire d'Annecy de Physique
des Particules*

Fabrice JAMMES
*Laboratoire de Physique de
Clermont*

