

Qserv: the shared nothing petascale database

Status report for LSST-France 2020

Fabrice Jammes, LPC Clermont-Ferrand

Comparison with Google Cloud: Qserv wins

- **Qserv and Google Big Query have quite similar performance** but:
 - Big Query is **not affordable** for LSST
 - It is **not as flexible** as Qserv
 - LSST engineers have **no expertise** on it
- **Google Kubernetes Engine is easy to use** but:
 - It is **not as flexible** as on-premise setup (i.e. NCSA and CC-IN2P3)
 - It is **~15% slower** than on-premise setup (remote storage)
 - It is **expensive** (~70000\$ for 3 weeks testing with 0.5% of LSST final data)

Toward production

Qserv is running on **Kubernetes** at NCSA

- memory: ~4 TB
- CPUs: ~400
- Data: 200 TB, mostly from **Gaia**



Qserv is now managed by software instead of system administrator

thanks to the qserv-operator

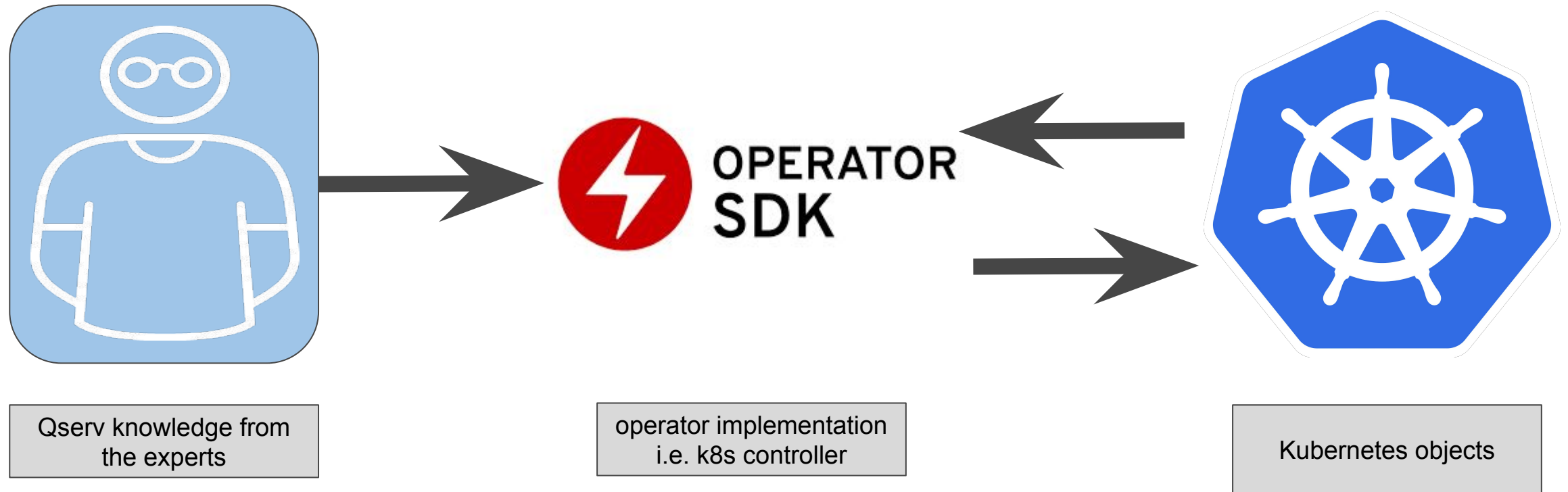


A new large scale test is under preparation at NCSA

Qserv Operator for Kubernetes



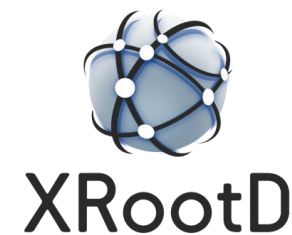
Operator embeds Qserv knowledge from the experts



- A Kubernetes Operator is an abstraction for deploying non-trivial applications on Kubernetes.
- The Operator pattern aims to capture the key aim of a human operator who is managing a service or set of services

Toward high availability

- Qserv data replication system: *Igor Gaponenko*
 - tested at Google and NCSA on thousand of chunks and TB of data
- First prototype for Qserv multi-master: *John Gate*
- Xrootd multi master is up and running: *A. Hanushevsky, F. Jammes*
 - proposal for improvements at Google Summer of Code 2020



Advanced features are coming

- Qserv **scalable dataloader**: *Igor Gaponenko*
 - will be soon tested at CC-IN2P3 by *Sabine Elles*, using DC2 data
(DPDD files using GCR catalogs available for run1.2i - tests ongoing on CC Openstack)



- Qserv **terascale secondary index**: *Nate Pease*
 - ObjectId/ChunkId
 - Stored inside **Redis**, a distributed, in-memory key-value database



Thanks!

Contact:

Fabrice JAMMES

LPC

Clermont-Ferrand

fabrice.jammes@in2p3.fr

