Loading and accessing DC2 data on Qserv

D. Boutigny, S. Elles *et al.*

LAPP/CNRS/IN2P3, Annecy

LSST France - LPC - Juin 2019

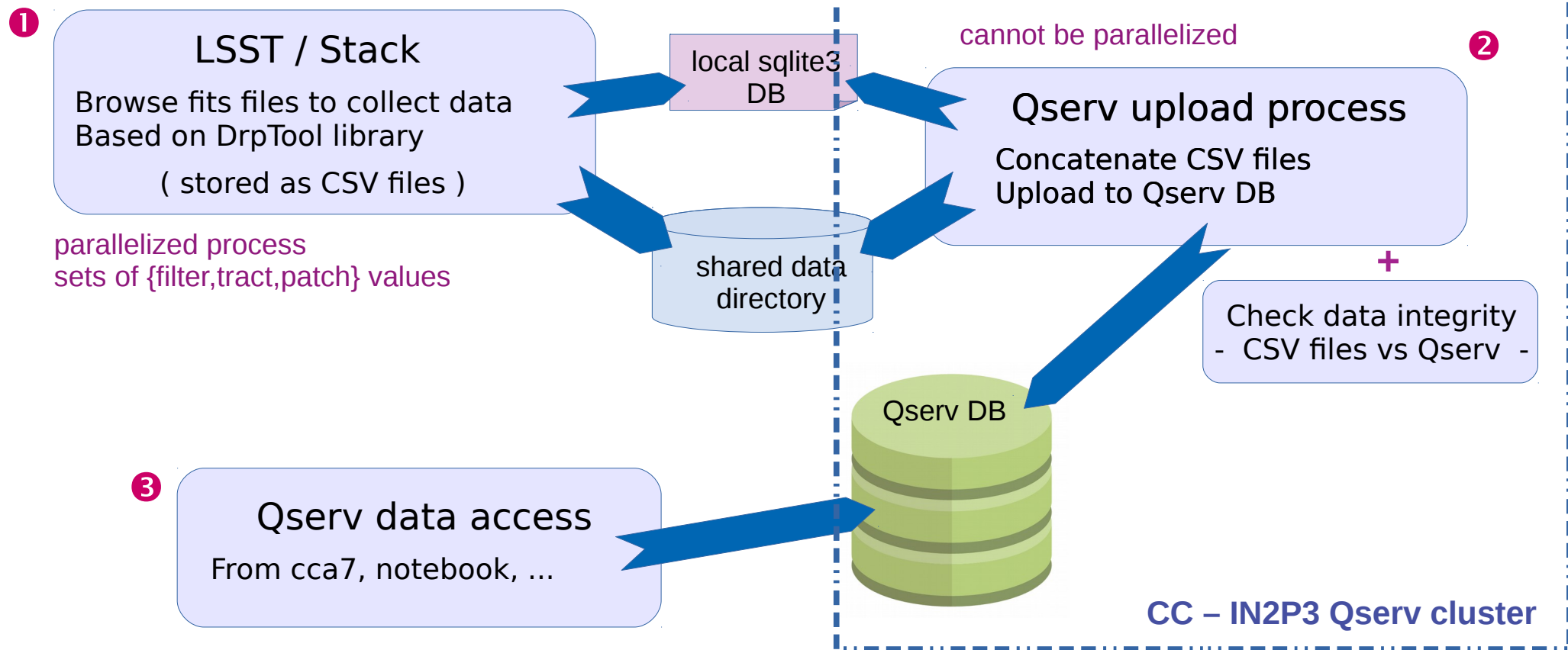---

Outline ⇨ Qserv Database for the LSTT Stack software

from LSST data catalogs to the Qserv database

From LSST datasets/catalogs to the Qserv database
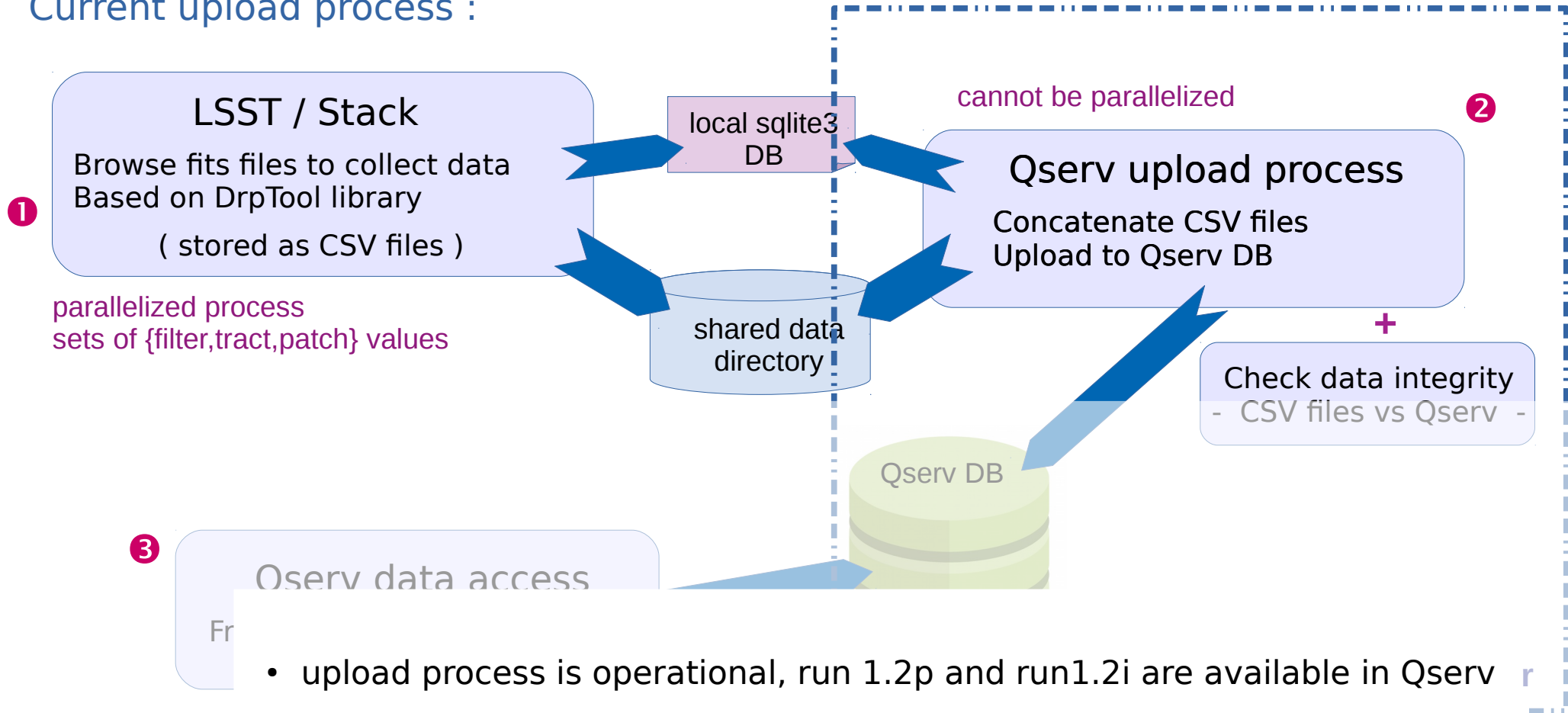
Qserv is developed at SLAC + IPAC
Design optimized for astronomical queries (parallel distributed SQL database)

Current upload process :

1 **LSST / Stack**
Browse fits files to collect data
Based on DrpTool library

(stored as CSV files)

parallelized process
sets of {filter,tract,patch} values

local sqlite3
DB

shared data
directory

cannot be parallelized

2 **Qserv upload process**
Concatenate CSV files
Upload to Qserv DB

+

Check data integrity
- CSV files vs Qserv -

Qserv DB

3 **Qserv data access**
From cca7, notebook, ...

CC – IN2P3 Qserv cluster

From LSST datasets/catalogs to the Qserv database

Current upload process :

LSST / Stack

Browse fits files to collect data
Based on DrpTool library

(stored as CSV files)

❶

parallelized process
sets of {filter,tract,patch} values

local sqlite3
DB

shared data
directory

cannot be parallelized

❷

Qserv upload process

Concatenate CSV files
Upload to Qserv DB

+

Check data integrity
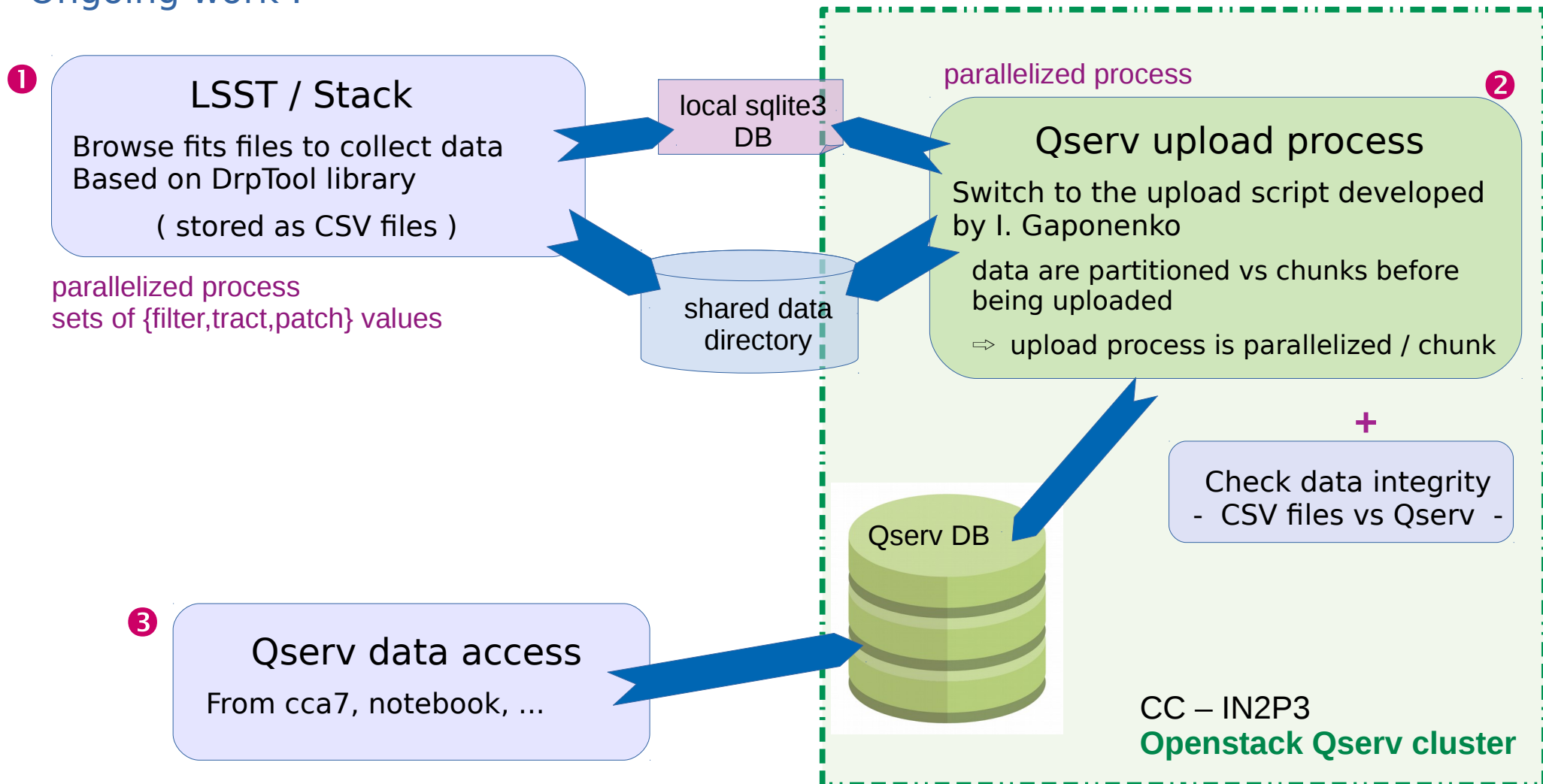- CSV files vs Qserv -

Qserv DB

❸

Qserv data access

Fr

- upload process is operational, run 1.2p and run1.2i are available in Qserv   r

- Qserv DB is accessible from cca7, notebook, … as shown by D. Boutigny in his talk

Status :

but the upload process is too slow because step **❷** is not parallelized
⇒ run1.2i :   step **❶** ~0.7 day + step **❷** ~1.5 days

From LSST datasets/catalogs to the Qserv database

Ongoing work :

❶ **LSST / Stack**

Browse fits files to collect data
Based on DrpTool library

(stored as CSV files)

parallelized process
sets of {filter,tract,patch} values

local sqlite3 DB

shared data directory

parallelized process

❷ **Qserv upload process**

Switch to the upload script developed by I. Gaponenko

data are partitioned vs chunks before being uploaded

⇒ upload process is parallelized / chunk

+

Check data integrity
- CSV files vs Qserv -

Qserv DB

❸ **Qserv data access**

From cca7, notebook, …

CC – IN2P3
Openstack Qserv cluster

Status : Openstack Qserv cluster is running
– necessary to avoid unwanted operations on the official Qserv DB -
new upload script is installed

⇒ work in progress

4

Data ingestion – next steps

- Finalize the transition to I. Gaponenko script
  and update the official CC-IN2P3 Qserv cluster upload process

- Adapt the ingest process to DPDD formatted data

- Measure the performance of the upload process

- ...