Astronomy ESFRI & Research Infrastructure Cluster
ASTERICS - 653477

Large Synoptic Survey Telescope

- Qserv integration into science pipeline -

Nicolas Chotard, Dominique Boutigny, Sabine Elles

CNRS / IN2P3 / ASTERICS

2nd ASTERICS-OBELICS Workshop

Oct. 16Th, 2017, Barcelona, Spain

Large Synoptic Survey Telescope

The Large Synoptic Survey Telescope is an instrument designed to make **high precision images** of the **whole accessible sky in 4-D (x, y, z, t)**

-

A 10 year time-lapse movie of the southern sky

Time domain science

- Novae – Supernovae – GRBs
- Source characterization
- Instantaneous discovery

Moving sources

- Asteroids and comets
- Proper motions of stars

Mapping the Milky Way

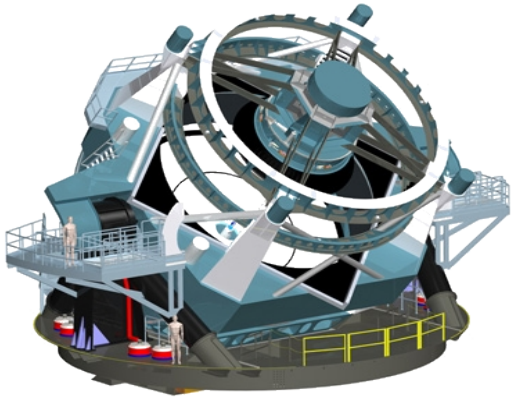
- Tidal stream
- Galactic structure
- Complementary to GAIA

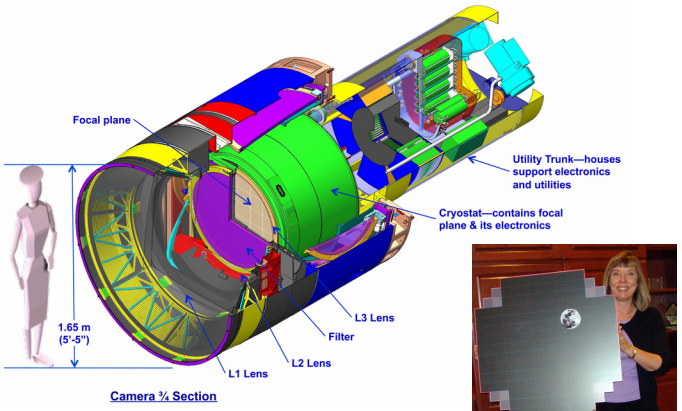
Dark energy and dark matter

- Gravitational lensing (strong and weak)
- Evolution of large scale galactic structures
- Trace the nature of dark energy

Where and how?

Average seeing: 0.67 arcsec





Telescope

- 8.4 m (6.7m effecif)
- Fully corrected (sphericity, coma, astigmatisme)
- A 350 tons mobile structure

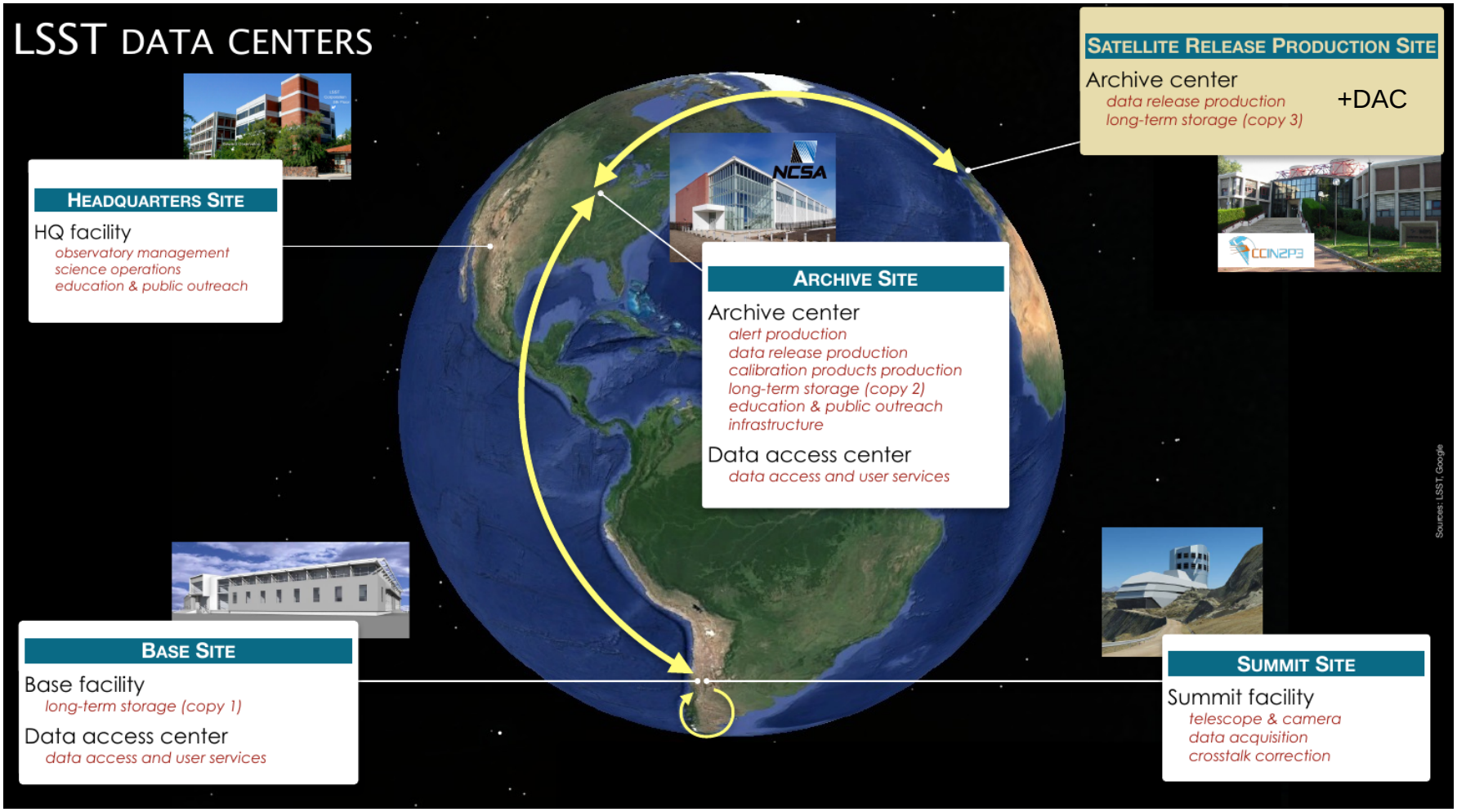
Camera

- 3.2 billion pixels @ 0.2 arcsecond / pixel
- 21 rafts
- 9 CCD / raft

Data centers

- ~15 TB of raw data / night
- Final image collection : 0.5 Exabytes
- Final database size (DR11) = 15 PB

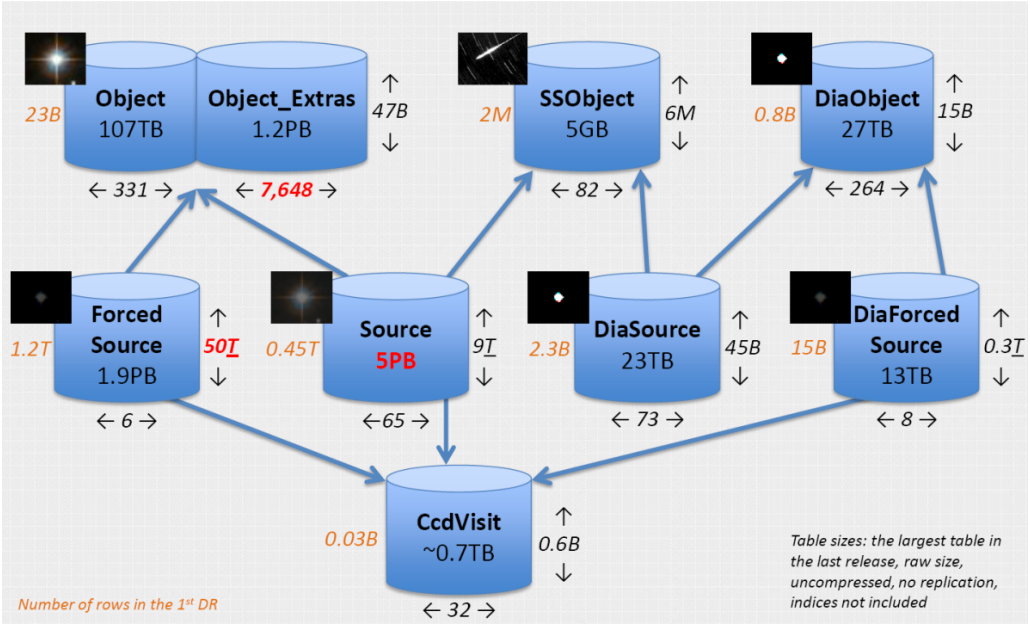
- Summit (Pachón) to Base (La Serena) = 2 x 100 Gbps
- Base(La Serena) to Archive (NCSA) = 2 x 40 Gbps

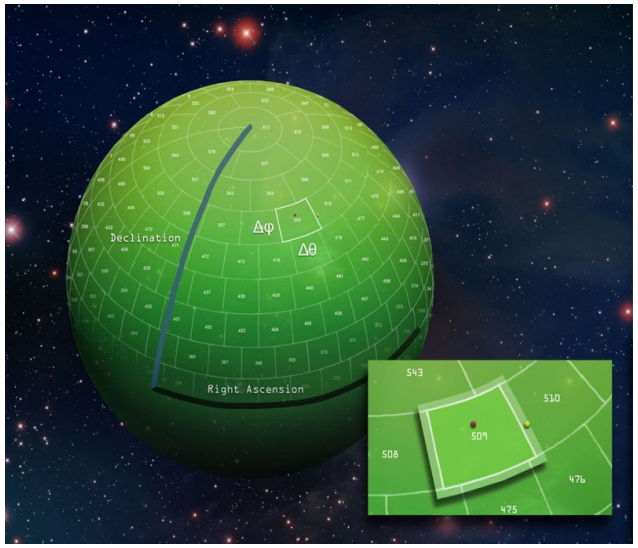


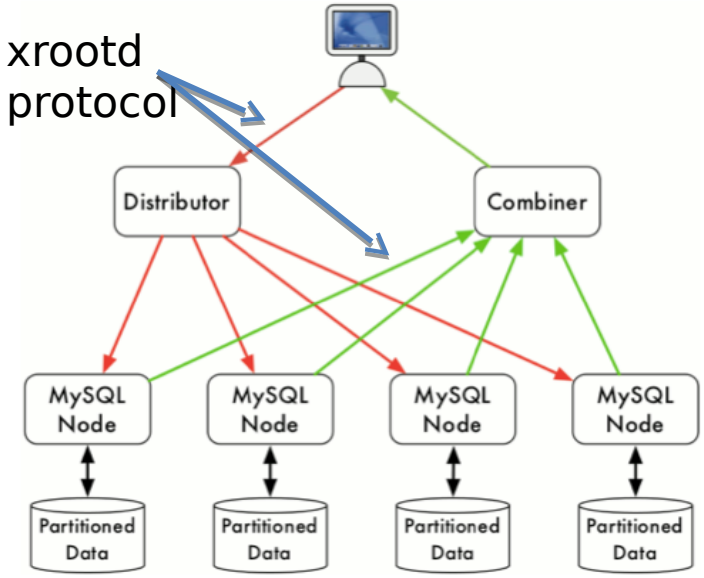
The database: a key component

Challenge: design an open source SQL database system able to store trillions of objects while keeping a reasonable access time

Qserv: developed at SLAC + IPAC
Design optimized for astronomical queries







Massively parallel – distributed – fault tolerant
relational database

Data access

Data reprocessing

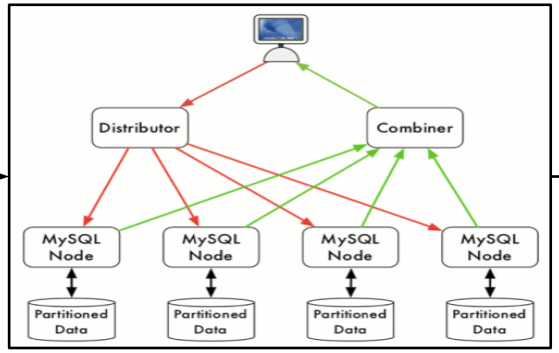
Data access

Data analysis

Automatized data reprocessing
(**LSST stack** + Workflow engine)

Catalogs

Qserv instance



Local files with data

Python interface

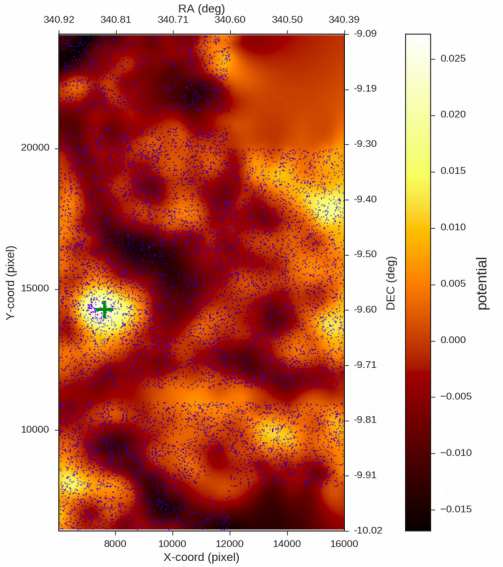
Local queries

Distant queries

Science pipelines

Qserv integration

- Test Qserv on real data processed through the LSST software stack
  - Different queries (magnitudes, position, etc.)
  - Different configuration of the DB (number of stripes and chunks)
  - Different catalogs (sources, coadds)
  - ➢ Test its capabilities and performances
- Qserv integration into science analysis pipelines
  - Automatic inclusion of LSST stack-processed data into a Qserv instance
  - Direct queries in this database from a science pipeline
  - Construction/test of python tools to query the data
- ➔ Test case: Clusters pipeline
  - Galaxy cluster mass estimate
  - LSST stack data used in all steps of the analysis
  - CFHT data already processed
  - 5 filters, several areas of the sky



Current work

- Short term
  - Set-up of a test-case Qserv instance @ CC-IN2P3 (OpenStack)
  - Incremental load of data (prototype already developed)
  - Reprocess and load more data (telescope, filters, sky area, etc)
- Longer term
  - Permanent Qserv instance @ CC-IN2P3 dedicated to analysis
  - Automatic and incremental ingestion of new LSST stack-processed data in the CC-IN2P3 Qserv instance
  - Python tools to query these data
- Implementation of these tools in DESC science pipelines
- Qserv tests and validation on real science cases in LSST/DESC

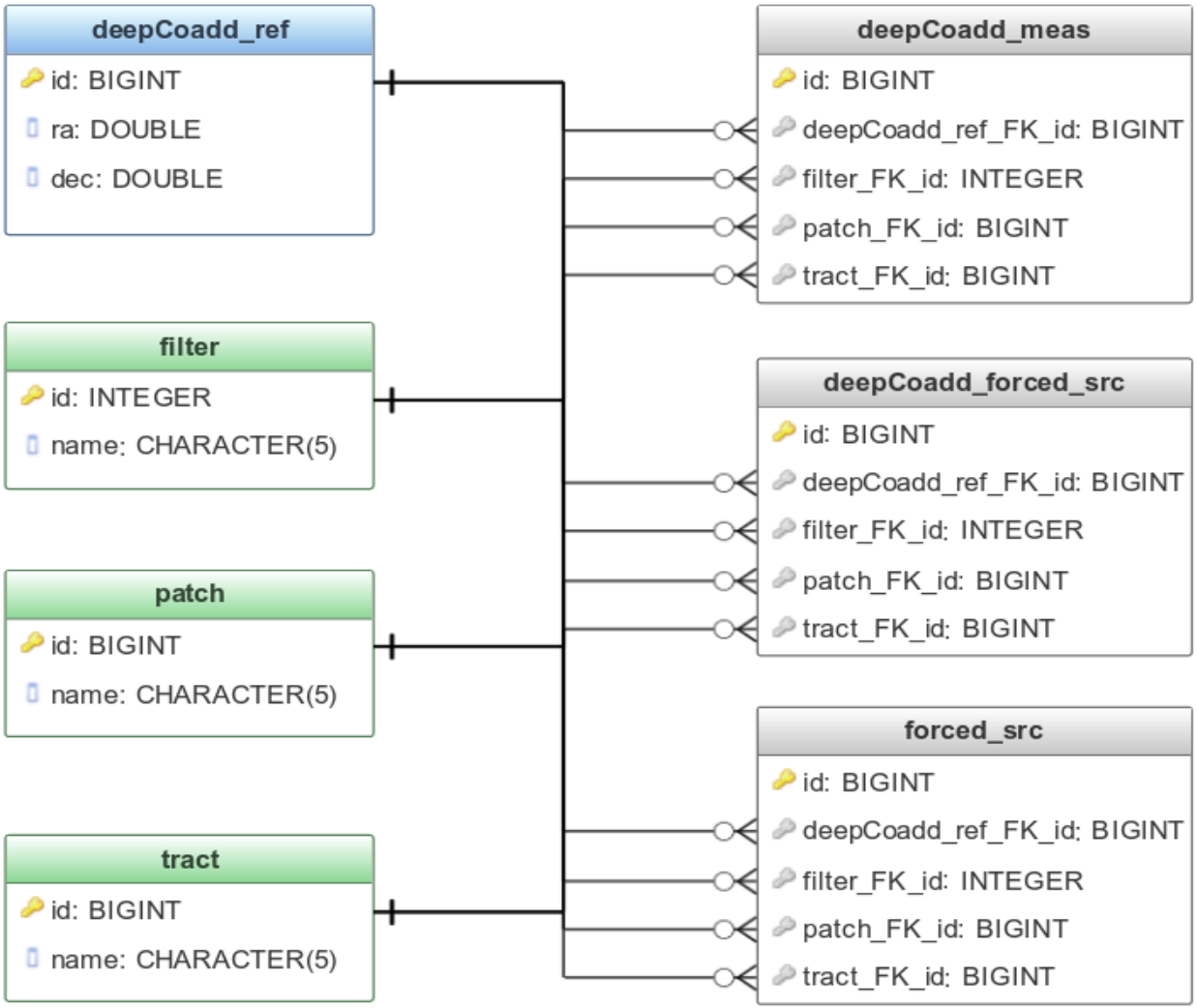
Conclusions

- LSST softwares (stack + Qserv) already available for users
  - All LSST softwares are open source and available on github
  - But still under developments
  - A huge work of validation has to be done
  - A connection between those tools is also needed
- The work undertaken here is crucial to
  - Understand the limitation of these softwares
  - Make sure that we can improve them while it is still time
  - See if we can already use them in different context (telescopes, data-sets, experiments)
  - Use them for science analyses before LSST starts taking data

Aknowledgement

H2020-Astronomy ESFRI and Research Infrastructure Cluster (Grant Agreement number: 653477)

Data base schema



Incremental load of catalogs

- Tool to create/update the Qserv DB based on the data-set produced by the LSST soft-wares
  - DB structure was defined in collaboration with the stack developers/physicists
  - Use as far as possible the DB access tools provided by Qserv developers (collaboration with LPC-Clermont)
- Guidelines
  - Asynchronous process that allows parallelisation
  - Data integrity guaranteed by MySQL
  - Main challenge: dealing properly with the error-recovery
- Status
  - Prototype developed and tested locally with a small data-set
  - Migration to the Qserv platform available at CC IN2P3 machines is ongoing