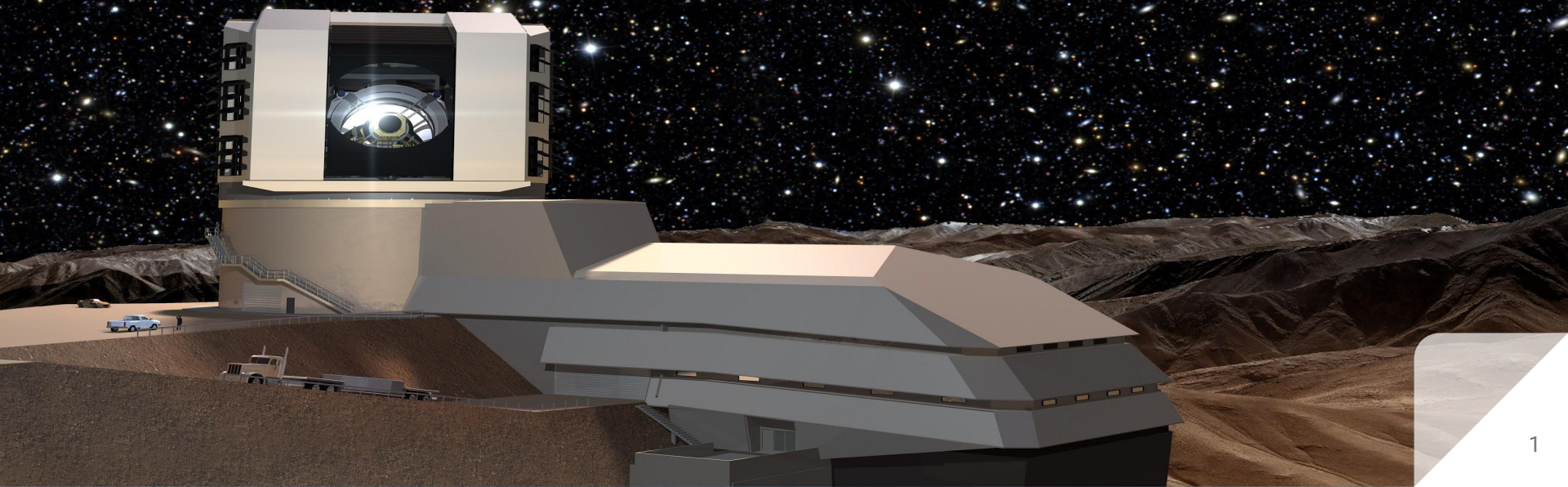
Cosmic Peta-Scale Data Analysis at IN2P3

Fabrice Jammes

Scalable Data Systems Expert
LSST-France Technical Coordinator for Databases



Who we are

Database and Data access team

- ★ 10 engineers at SLAC + 1 LPC-IN2P3 (~10 FTE)
  - *Software development*

Operations teams

- ★ 5 engineers at CC-IN2P3
  - *Large Scale development platform*
  - *System administration, Monitoring*
- ★ 5 engineers at NCSA
  - *Prototype Data Access Center*

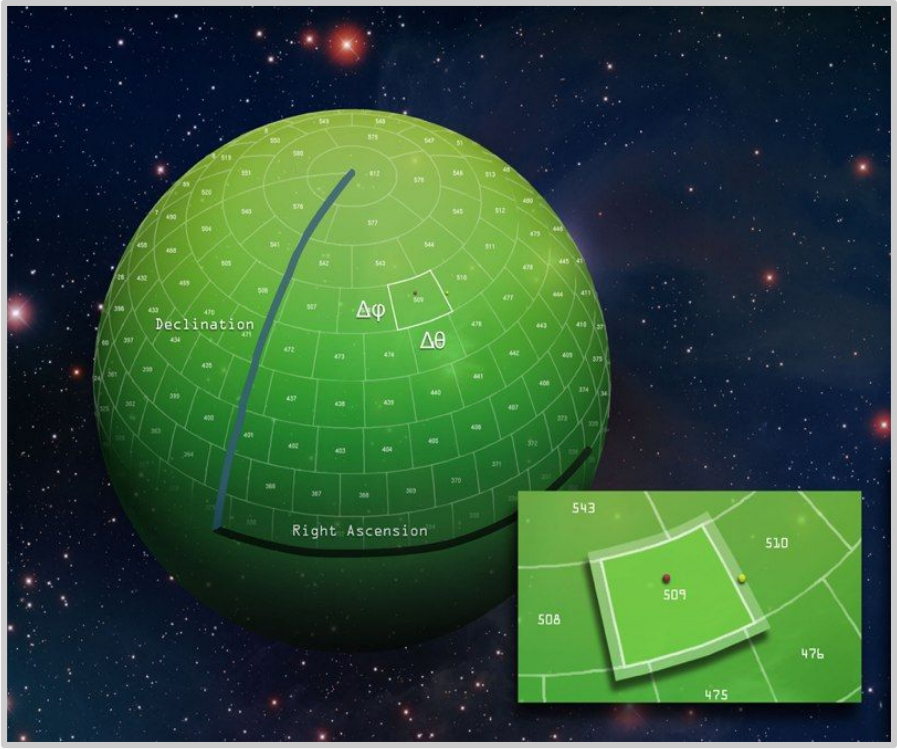
Research and development

- ★ 5 engineers/researchers at LPC-IN2P3/LAPP/LIMOS (2 FTE)
  - *Data-loading*
  - *Cloud-computing, containers, CEPH*
  - *Large Scale Continuous integration*

What we do

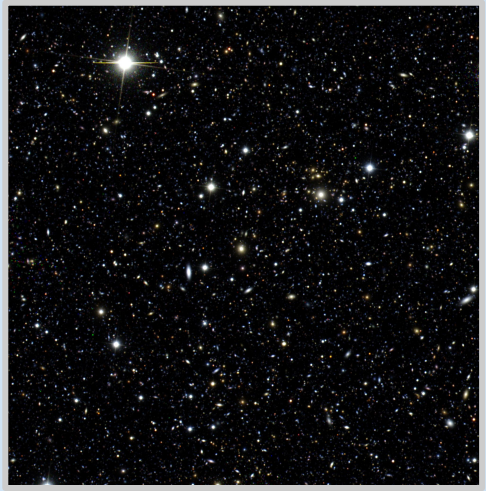
Data Access and **Database**

- ★ Data and metadata
- ★ Images and databases
- ★ Persisting and querying
- ★ For pipelines and users
- ★ Real time Alert Production and annual Data Release Production
- ★ For Archive Center and all Data Access Centers
- ★ For USA, France and international partners
- ★ Persisted and virtual data
- ★ **Estimating, designing, prototyping, building, and productizing**



Data

Images
Persisted: **~38 PB**
Temporary: **~½ EB**



- ★ ~3 million “visits”
- ★ ~47 billion“objects”
- ★ ~9 trillion “detections”

- ★ Largest table: **~5 PB**
- ★ Tallest table: ~50 trillion rows
- ★ Total (all data releases, compressed): **~83 PB**

Ad-hoc user-generated data
Rich provenance

Analytics

Aiming to enable majority of analytics via database
Aiming to enable rapid turnaround on exploratory queries

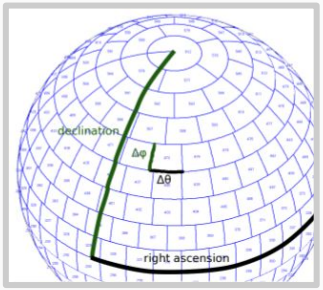
In a region
Get an object or data for small area - <10 sec
Across entire sky
Scan through billions of objects - ~1 hour
Deeper analysis (Object_*) - ~8 hours
Analysis of objects close to other objects
~1 hour, even if full-sky
Analýsis that requires special grouping
~1 hour, even if full sky
Time series analysis
Source, ForcedSource scans - ~12 hours
Cross match & anti-cross match with external catalogs
~1 hour

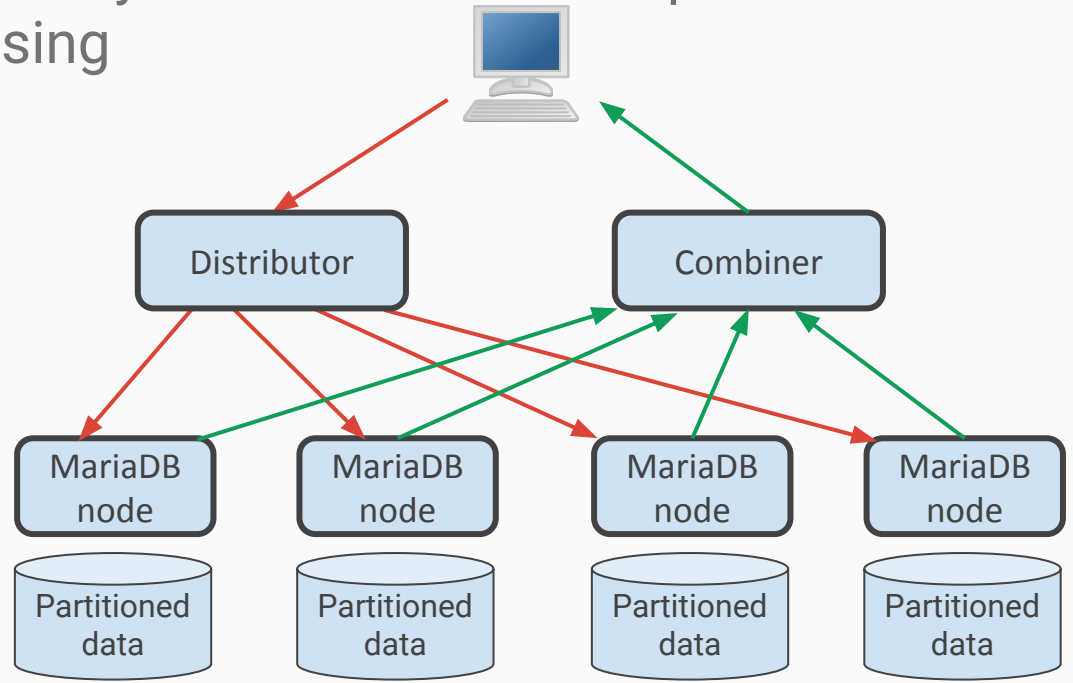
Sizing the system for
~100 interactive +
~50 complex
simultaneous DB queries.

Same for images

Qserv design

- ★ 100% Open source *(based on xrootd, mysql-proxy, mariadb, Google protobuf, flask, ...)*
- ★ Relational database, spatially-sharded with overlaps
- ★ Map/reduce-like processing





Tests and demonstrations

Target for production

~500 nodes clusters in 2 international data-centers

Running now

Development platform (CC-IN2P3)

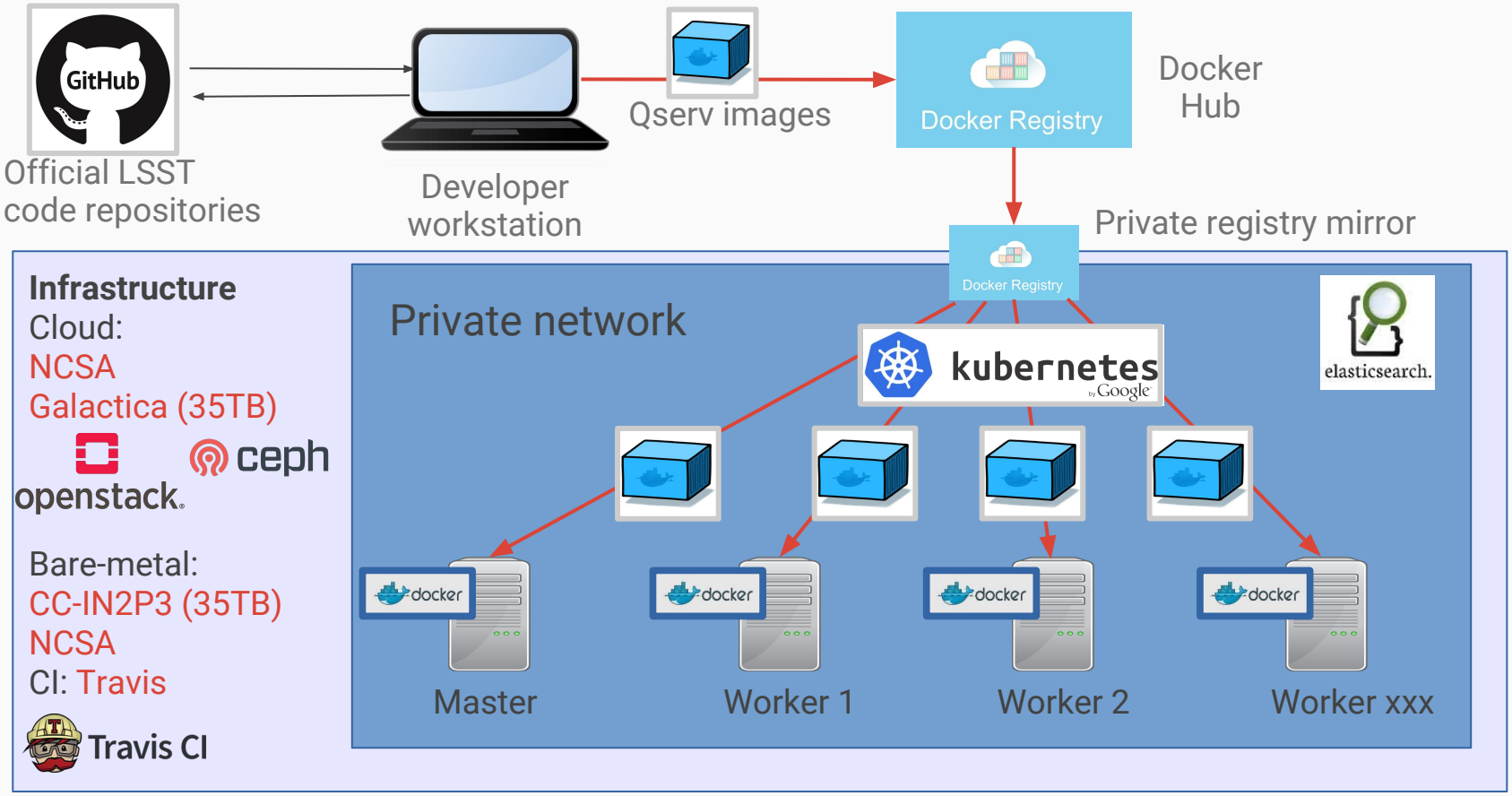
400 cores, 800 GB memory
500 TB storage,

★ **~35 TB data set** on 2*25 nodes
★ *Computing KPI on **~70TB dataset***

Prototype Data Access Center (NCSA)

500 cores, 4 TB memory
700 TB storage,
SDSS Stripe 82 and
Wide-Field Infrared Survey Explorer (WISE) data loaded

Automated Qserv deployment



Summary

- ★ Big Data with Complex Analytics
- ★ Spatially-sharded, map/reduce-like RDBMS
- ★ Open source + custom glue
- ★ Optimized for astronomical data sets at scale
- ★ Have working prototype
- ★ Turning it into a production system
- ★ Want to learn more?
  - http://ls.st/4gh (Database Design doc)
  - http://ls.st/6ym (User Manual)
- ★ Are you an adventurous super early adopter? You can try it now
  - http://ls.st/89y (Qserv Documentation)

Thanks!

Contact:

Fabrice JAMMES
LPC
Clermont-Ferrand

fabrice.jammes@in2p3.fr