Euclid and LSST

Éric Aubourg • APC

Key issues in cosmology

We only know 4% of the content of the Universe

— What is the nature of dark matter ?

— What is driving the acceleration of the expansion of the Universe ? An extra component, with negative pressure and unknown nature ("dark energy")? Should we modify the theory of gravitation?

An important goal is currently to study the properties of dark energy

— Recover the history of the expansion of the Universe.

— Study the growth of large scale structures.

Methodology

Map « tracers » of large scale structures.

Use distance ladders (supernovae, BAO…).

Detect far objects to unwind the history of the expansion.

Need to map a large volume : large surface and depth.

Projects for 2020

Two major projects :

— The Euclid Satellite, dedicated to the study of dark energy, though baryon oscillations and weak lensing.

— The LSST ground telescope, multi-purpose, that uses to study dark energy all the key probes: baryon oscillations, lensing, clusters and supernovae.

LSST will produce the biggest volume of data of the two.

Euclid

Satellite launched to L2, ~ 2020

Visible imager for lensing (1 billion sources used / 10)

Infrared imager for photo-z (used for lensing analysis)

Infrared spectrograph for BAO (10 million sources)

6-year mission, mapping 15 000 deg^2 of the sky.

~ 2 Pb of raw data (850 Gb/day)

Euclid

Euclid does not have narrow band visible filters

Need to merge with ground-based data for photo-z

Main computing challenge : merge with LSST, if pixel-level coprocessing is required.

Different ground segment processing stages will be distributed among SDC (Science Data Centers) in Europe. CC-IN2P3 will be one of them.

LSST

Large Synoptic Survey Telescope, 2020 in Chili.

Caméra has 3.2 Gpixel (6.4 Gb), takes one image every 15 seconds : 15-30 Tb of data per night.

Science objectives are wide and diverse.

In ten years of operation, LSST will perform 1000 “visits” on each accessible sky patch (full coverage in three nights).

— deep sky through image stacking (or equivalent techniques)

— transient objects studies (supernovae), moving objects (astéroïdes).

Survey mode only : the telescope delivers reduced data to the community.

LSST

5 000 billions source detections

32 000 billions photometric measurements

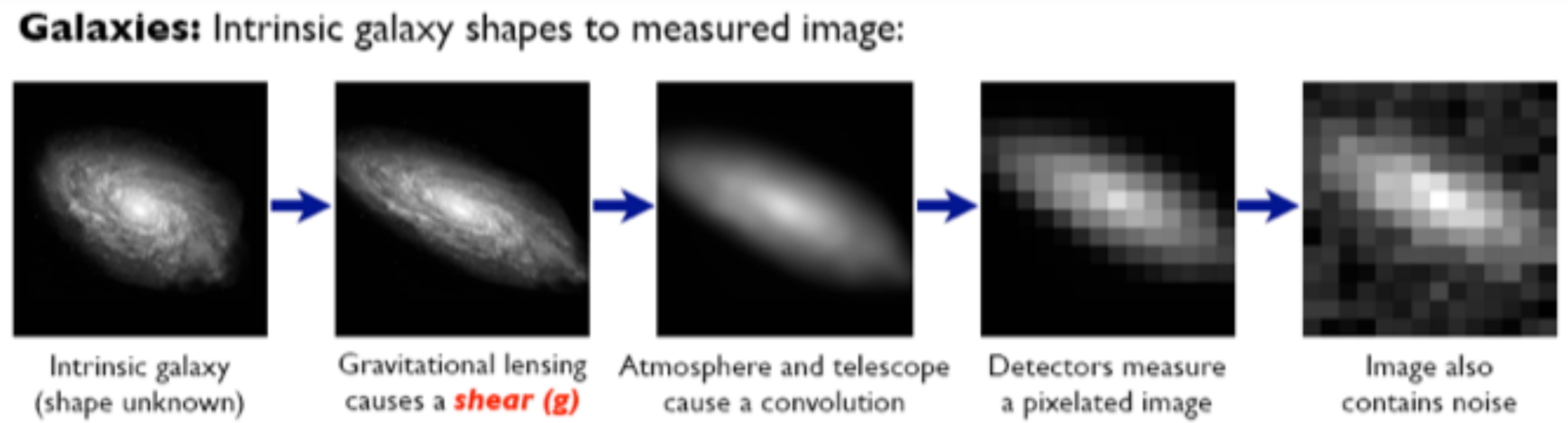
38 billions objects followed (incl. 24 billions galaxies)

Alerts on transient objects are posted less then 60 seconds after the image readout.

LSST : some key science objectives

Weak lensing (or weak shear)

Measurement of the *total* mass in front of a source, through the gravitational shear induced on the source.



LSST : some key science objectives

The analysis of the gravitational shear effect requires the knowledge of the distance of the sources.

Too many objects to acquire a spectrum for each source.

Photometric redshifts redshifts are used: photometric measurement in five bands, and probabilistic reconstruction of the distance.

Can be improved with more bands, through for instance joint measurement in the infrared with the Euclid satellite.

Requires an accurate photometric calibration.

LSST : some key science objectives

Instrument and atmosphere effects average out between the different "visits" of a field

- Co-adding images in not optimal
- Multi fit on a stack of images, with for each image a model of atmosphere and instrumental effects.

LSST : some key science objectives

Supernovae search

- Stellar explosion. The luminosity can reach the one of the host galaxy.
- Can be used as “standard candle”: distance ladder to measure the expansion of the Universe.

Done through the detection of “new” objets in the images

- Image subtraction algorithms

LSST : some key science objectives

Solar system and asteroids

Detection of moving objects, orbit reconstruction.

LSST : the consortium

Nine laboratories from IN2P3 in France

Chile

35 US universities

Google

Data processing will be done 50% at NCSA, 50% at CC-IN2P3 (full mirror of raw and reduced data)

LSST

Catalogues are published each year (a few months of reprocessing each year for the data release)

- Images will be public, but it will be necessary to compute where the images (10-100 Pb) are located: data access centers.

Reconstruction of time series: non sequential access to data is required.

LSST

The data management stack is implementing optimal algorithms in C++ (+ python), on a distributed file system.

Catalogs are stored in a distributed database.

A specific database system (QServ) has been developed, that allows parallel queries to be executed on a farm of MySQL servers (based on Scalia/xrootd)

LSST From the User's Perspective

Level 1

- A stream of ~10 million time-domain events per night, detected and transmitted to event distribution networks within 60 seconds of observation.
- A catalog of orbits for ~6 million bodies in the Solar System.

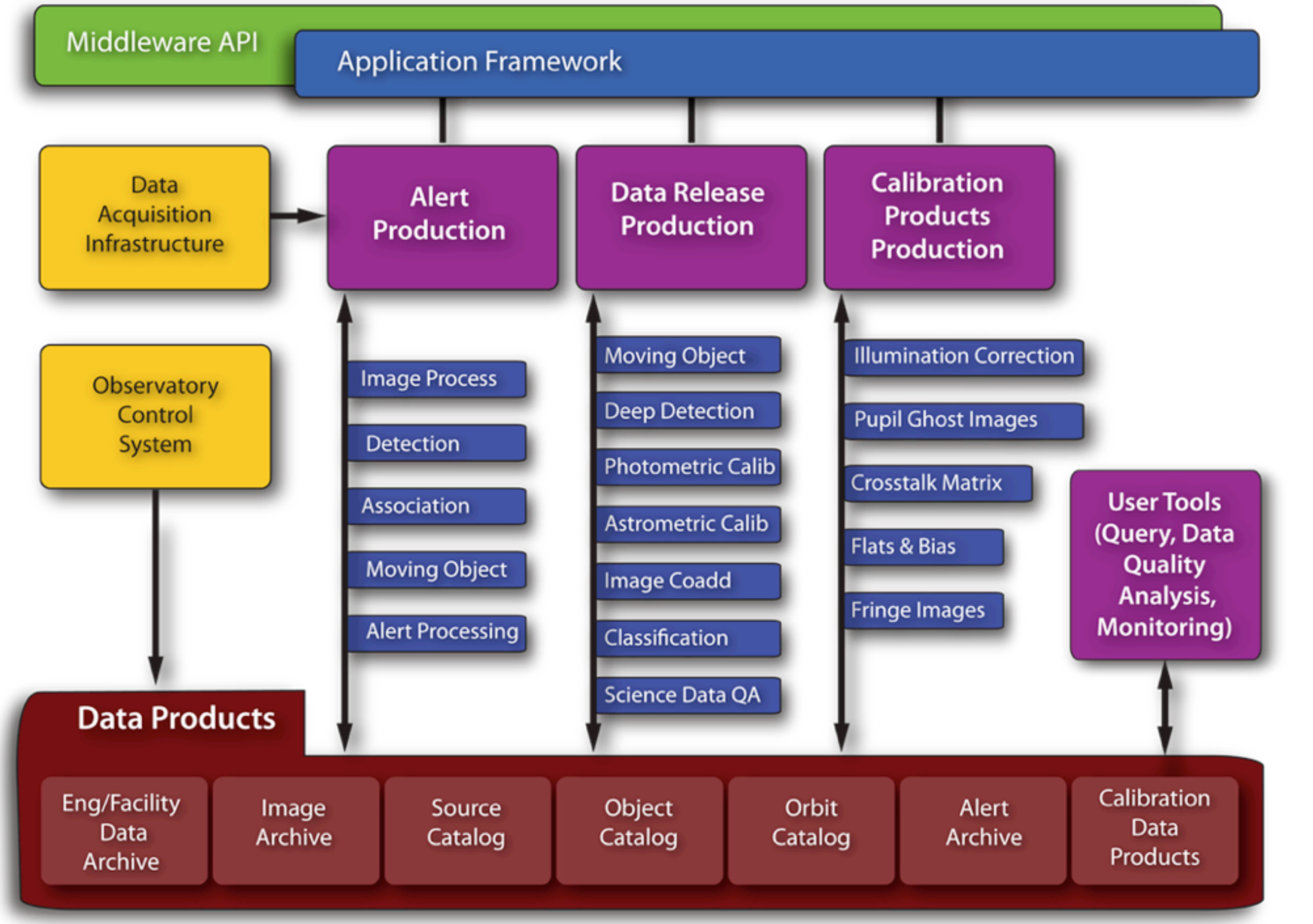
Level 2

- A catalog of ~37 billion objects (20B galaxies, 17B stars), ~7 trillion observations ("sources"), and ~30 trillion measurements ("forced sources"), produced annually, accessible through online databases.
- Deep co-added images.

Level 3

- Services and computing resources at the Data Access Centers to enable user-specified custom processing and analysis.
- Software and APIs enabling development of analysis codes.

Application Layer - framework-based pipelines process raw data to products



Multifit implementation

In the data release pipeline

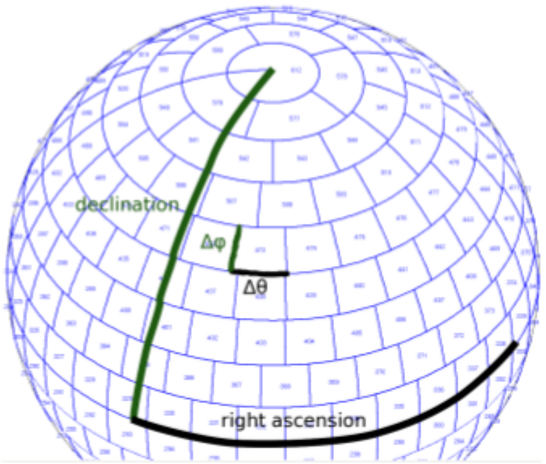
Images are coadded to allow for a deep detection of objects

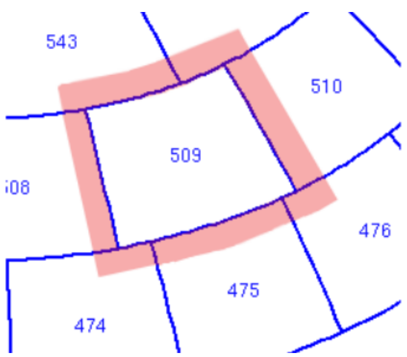
Final photometry is **not** done on the stacks, but back on the individual images

Simultaneous multi-fit of all images in all bands, with parameters depending on the object, the filter, the current image…

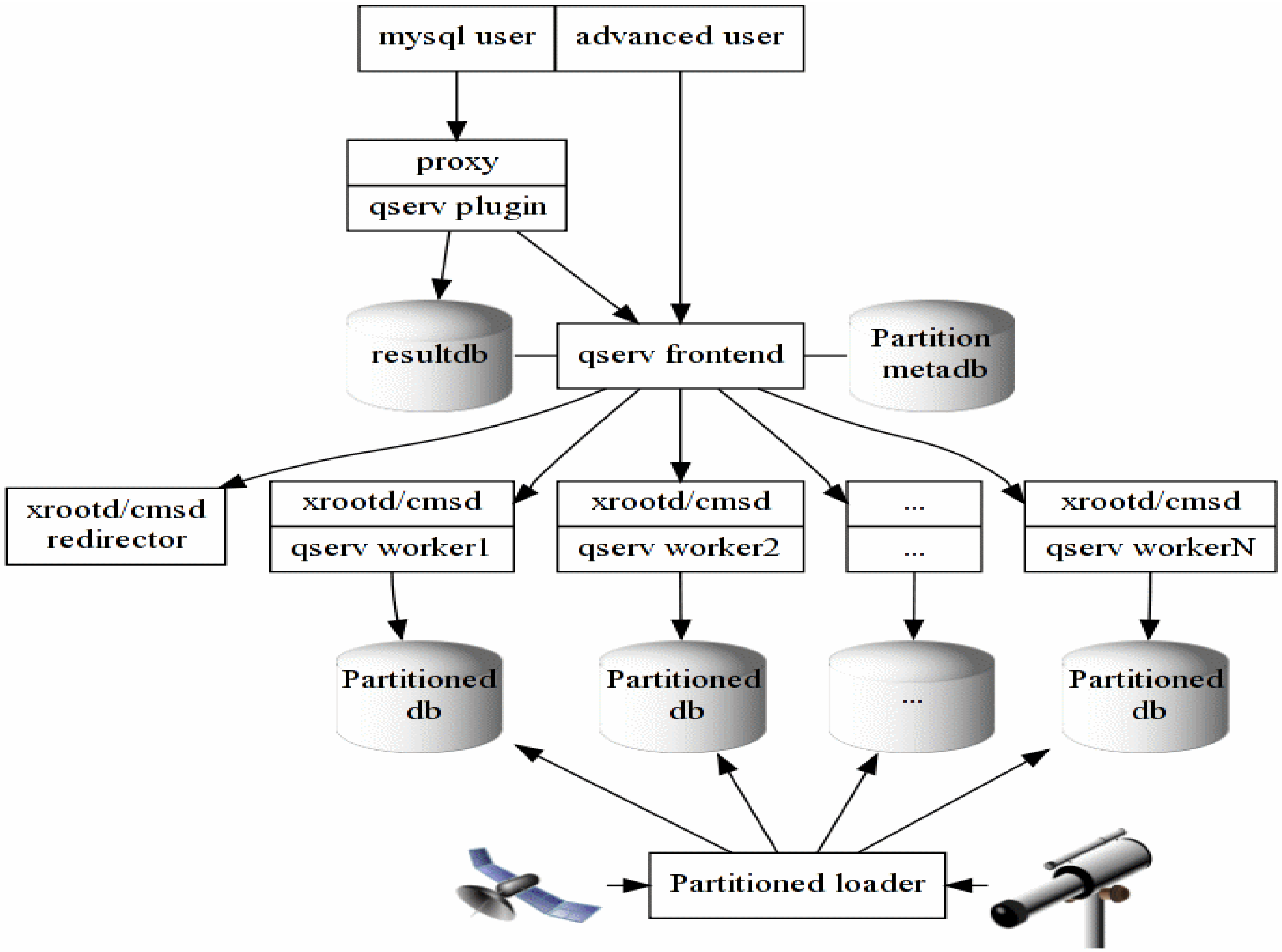
Tests on SDSS Stripe 82 : more than 0.5 mag deeper than doing photometry on the stack like the SDSS pipeline.

LSST-Euclid multifit?


declination
Δφ
Δθ
right ascension


543
510
509
476
475
474


509


mysql user
advanced user
proxy
qserv plugin
resultdb
qserv frontend
Partition metadb
xrootd/cmsd redirector
xrootd/cmsd
qserv worker1
xrootd/cmsd
qserv worker2
...
...
xrootd/cmsd
qserv workerN
Partitioned db
Partitioned db
...
Partitioned db
Partitioned loader

SQL-base ad-hoc solution

+ SQL well-mastered by users, reliable technology

+ Optimized for spherical geometry, search of neighboring objects

+ Code is owned by LSST, and largely RDMS-agnostic

- Some tools have to be rewritten

  partitioned loader, failure recovery, load balancing and high-availability

– Tests

	Scale	Inter-active	Table scans	Large joins	Notes
The PDR test (2011)	150 nodes, 32TB, 2B objects, 55B sources	4-9 sec	3-8 min	10 min – 5 h	Problems with >4 concurrent queries (<20K segment-queries)
JHU (2012)	20 nodes, 100TB, 2B objects, 80B sources	~5 sec	< 7 min		Numerous problems with unstable hardware
IN2P3 (2013)	300 nodes, 10TB, 0.4B objects, 14B sources	1.2-4 sec	10 sec – 10 min	~ 5 min	Showed good scaling and low dispatch overhead, proved concurrency

– Demonstrations

- Concurrency (up to 100K in-flight segment-queries, on ~100 nodes)
- Fault tolerance (catching errors, transparent fail over to a replica)
- Shared scanning (30-query scan: 5m27s, avg speed for a single query: 3m)

LSST : data distribution and partnerships

Cloud actors expressed interest:

Google is a member of the LSST consortium.

Amazon offers free hosting of all public data, and can sell CPU time.