The LSST Data Processing and Data Analysis Centers in France

Dominique Boutigny
Laboratoire d'Annecy de Physique des Particules

LineA Webinar – April 26th, 2018
The LSST project

Modified Paul-Baker optical formula
D = 8.4 m (6.7 effective)
f/d = 1.23
350 t mobile structure

Étendue = surface X field of view
→ LSST: 319 m².deg²
The LSST Camera

Field of view: 9.6 deg$^2$

3.2 Gpixels – 0.2” / pixel
189 CCD (4k x 4k) deep depleted

Highly segment electronics → the full focal plane is read in 2 s

2 x 15 s exposure → 40 s total time / visit
5 s to slew to a new position

Dominique Boutigny
6 filters 320 – 1070 nm
• Main tool to determine photometric redshifts

The filter exchange system design and construction is a French contribution to LSST

France is also contributing to
• CCD procurement and Electronics
• Slow control
• Stand alone Optical Bench
Fast – Wide - Deep

- Survey time : 10 years
- Main survey area : 18 000 deg²
- 2.75 $10^6$ visits in 10 years
- $<825>$ visits / pointing

<table>
<thead>
<tr>
<th></th>
<th>u</th>
<th>g</th>
<th>r</th>
<th>i</th>
<th>z</th>
<th>y</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single visit</td>
<td>23.9</td>
<td>25.0</td>
<td>24.7</td>
<td>24.0</td>
<td>23.3</td>
<td>22.1</td>
</tr>
<tr>
<td>10 years</td>
<td>26.1</td>
<td>27.4</td>
<td>27.5</td>
<td>26.8</td>
<td>26.1</td>
<td>24.9</td>
</tr>
</tbody>
</table>

LSST will visit Deep Drilling Fields several times / night → **Transient events detection**

LSST is particularly optimized for
- Fast detection of faint objects
- Transient event detection
- Precise measurement of Star positions and colors
- Precise measurement of Galaxy shapes and colors

IN2P3 is mainly involved in the Dark Energy Science Collaboration (DESC)
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.

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.

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.
France has signed up an MOA with LSST where
  • France will process 50% of the Level 2 Data Release Processing (DRP)
    • The other 50% is processed at NCSA
    • France will hold a complete copy of the LSST data

NCSA is responsible to coordinate the whole DRP

In exchange of this contribution France gets 45 data rights

⇒ *Note that Data Access and Science analysis is not covered by this agreement*
Dominique Boutigny

**Summit and Base Sites**
- Telescope and Camera Operations
- Data Acquisition
- Long-term storage (copy 1)
- Chilean Data Access Center

**Archive Site**
- Archive Center
  - Alert Production
  - Data Release Production (50%)
  - Long-term Storage (copy 2)
  - Data Access Center
  - Data Access and User Services

**French satellite center**
(CC-IN2P3, Lyon, France)
- Data Release Production (50%)
- French DAC

**SLAC Center**
- Data Products Production Support
- Science Operations and Community Support

**HQ Site**
- Science Operations headquarters
- Operations office
- Education and Public Outreach

**Archive Site**
- Archive Center
  - Data Release Production (50%)
  - Long-term Storage (copy 2)
  - Data Access Center
  - Data Access and User Services
The IN2P3 Computing Center
CC-IN2P3

2 computer rooms

CPU:
- 31,000 cores (330,000 HEPSpec06)

Disk storage:
- DAS: 22 PB
- Shared Filesystem GPFS: 2.2 PB

Mass Storage:
- 4 tape robots (340 PB nominal capacity)
- 50 PB in use

Main computing center in France for HEP, Nuclear Physics, Astroparticles and Cosmology

- ~65 highly skilled computing engineers
- Operating 24/24 – 365/365

- Tier-1 for the 4 LHC experiments within the Worldwide LHC Computing Grid
- Supporting ~70 experiments / projects
Capacity planning for LSST DRP@CCIN2P3

From Fabio Hernandez (LSST project leader at CC-IN2P3)
The choice of the Data Management team has been to re-use existing algorithms whenever possible but to rewrite everything from scratch

- Considerable experience from previous projects especially SDSS and also from current ones: HSC, DES, …

Develop a modular, efficient and versatile image analysis framework

- ~50 M$ funding for construction and commissioning of the LSST stack and associated middleware
- All the code is open source: https://github.com/lsst

Code is in Python and C++

- Standard users are dealing with Python modules

Designed to support any CCD-based instrument: SDSS – HSC – CFHT – DES…

- DM stack is the official software package for HSC image processing

⇒ The LSST DM stack is a complex software system. Building up experience with it was (is) a very important goal for IN2P3
Building up experience

**Split DRP Data Challenge in 2013:**

- Process SDSS Stripe 82 with DM stack
- 50% at CC-IN2P3 – 50% at NCSA and 10% in common for comparison
- Output in MySQL database
  - Had tremendous problems ingesting data but finally ok
- The whole thing was successful
  - Validated the Split DRP (NCSA / CC-IN2P3) approach

**Reprocess precursor datasets**

- CFHT – Galaxy clusters – SN detection and light curve
- HSC SXDS field
- The driving idea here was to
  - concentrate on relatively small fields to avoid scaling problem
  - keep a strong connection with science
- We also closely monitored the I/O in order to anticipate bottlenecks
Re-measure clusters masses using shear signal from background galaxies


Develop a full pipeline to process data from the raw images up to the cluster mass measurement

MACSJ2243.3-0935

Mariana Penna-Lima Vitenti (postdoc at LAPP and moving to Brasília this summer) is developing a state of the art mass fitter
HSC image processed through the DM stack at CC-IN2P3
The Dark Energy Science Collaboration has started its 2nd Data Challenge

Full simulation starting from a extragalactic catalog from a large scale cosmological simulation + Galaxies + Stars + Strong Lensing + Supernovae…

**Main survey:**
- Area: 300 deg$^2$
- Survey depth: 10 years
- Using the latest optimized cadence model + dithering scheme
- Visits: (ugrizy) = (56, 80, 184, 184, 160,160) x 30 fields ~27 000 visits

**Ultra Deep Drilling Field:**
- 1.25 deg$^2$
- 10 years
- 20 000 visits

Use the phosim ray-tracer to simulate realistic images including optics, atmosphere, sky background, sensor effects (tree rings, Brighter-Fatter, etc.)
- Image simulation run at NERSC

Hope to have everything done by this summer
IN2P3 will process the DC2 datasets through DM stack to produce Data Release catalogs + difference imaging.

We consider this DESC Data Challenge as an LSST Data Challenge in preparation of our future role in the DRP:

- > 50 Million CPU hours (HS06)
- ~1.5 PB of output images and catalogs

Use the SRS workflow engine developed at SLAC to automatize production.

Explore:
- Data transfer through transatlantic link
- I/O and memory footprint
- Optimization of the process and infrastructure

We hope to learn a lot from this big enterprise.

© Nicolas Chotard
RGB image from one sky patch (4k x 4k)

Coaddition:
- r : 11 visits
- i : 9 visits
- g : 8 visits
The database: a key component

The challenge is to design an SQL database system able to store trillions of objects while keeping the access time at a reasonable value.

Qserv: developed at SLAC – Design optimized for astronomical queries

Massively parallel - distributed - fault tolerant relational database

Total size: ~83 PB
(11 Data Release)
CC-IN2P3 has a partnership with Dell
Setup Qserv test bench
  • 50 servers (2 test benches x 25 servers)
  • 400 cores
  • 0.5 PB disk storage
Heavily used bu Qserv developers for integration and performance tests

In parallel IN2P3 is developing another Qserv framework to test the Database from the *science perspective*
  • Ingest galaxy cluster data and repeat analysis
  • Also plan to ingest the DESC DC2 catalogs

Alternatives to Qserv are also considered by the Data Management team
⇒ Final decision soon
LSST is planning to build a Data Access Center at NCSA to serve the worldwide science community

Note:
Note: Data rights does not imply Data Access ⇒ LSST will charge the access to the NCSA DAC

In France we want to give a full data access to the IN2P3 scientists

We plan to deploy a Data Access Center and a Science User Interface
  • Will re-use components being developed at IPAC as much as possible

4 possible scopes for our “Science Platform”:
  1. Limited to French community ⇒ Mainly DESC but also transients, stellar,...
  2. Europe ⇒ within a EU project
  3. Worldwide together with NCSA and maybe others?

No decision yet – Many political and financial aspects
Plans

We are planning to investigate all the components of the Science User Platform by deploying prototypes at IN2P3

- Continue to use a “Science Driven Approach”
- Test using actual analyses on real precursor data or on DESC Simulated data

We also plan to check if the current model can be adapted toward a distributed approach

- Large scale data access and heavy computation within a Data Center (a.k.a CC-IN2P3 or NCSA or any other international DAC)
- Work on Jupyter notebooks, Firefly, final analysis… in remote laboratories

No reason to limit the model to LSST data only – Can be extended to multi-wavelength / multi-probes datasets

We are of fully open to future International collaborations

- Europe of course (large UK contribution in LSST)
- Brazil - Our collaboration with Mariana Penna Lima and Sandro Vitenti is a first step