Centre de Calcul de l'Institut National de Physique Nucléaire et de Physique des Particules

Digitising the Universe with the Large Synoptic Survey Telescope

fabio hernandez

CCIN2P3 cnrs

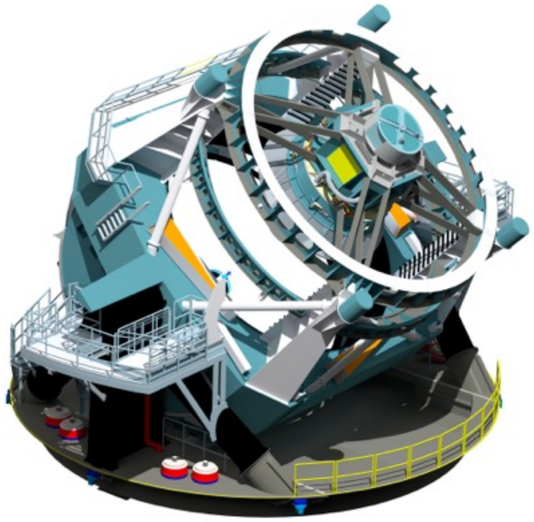
JCAD'2018, Lyon, October 25th, 2018

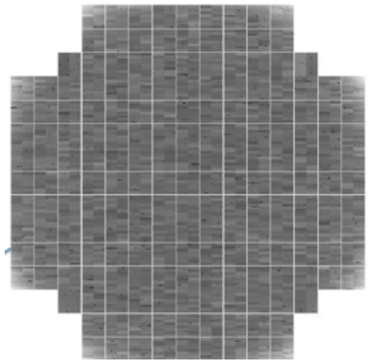
CONTENTS

- LSST overview
- IN2P3 contributions to LSST
- Science platform
- Summary

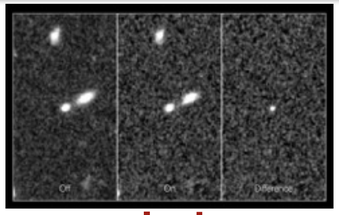
LSST OVERVIEW

LARGE SYNOPTIC SURVEY TELESCOPE

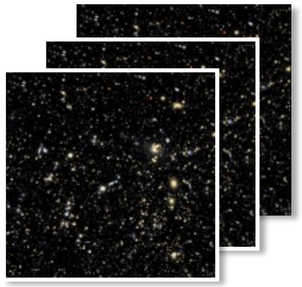




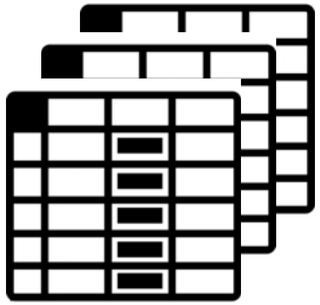
raw images



alerts



calibrated exposures



astronomical catalog

scientific discoveries

LSST aims to deliver a catalog of 20 billion galaxies and 17 billion stars with their associated physical properties

LSST OVERVIEW (CONT.)

- Principle of operations

  *90% of the observing time of the telescope devoted to a **deep-wide-fast survey***

  one complete visit of the southern hemisphere sky every 3-4 nights**, from 2022 for **10 years

  43% of the celestial sphere will be covered by this survey

  each patch of the sky to be visited about 1000 times

- Science themes

  *determining the nature of **dark energy** and **dark matter***

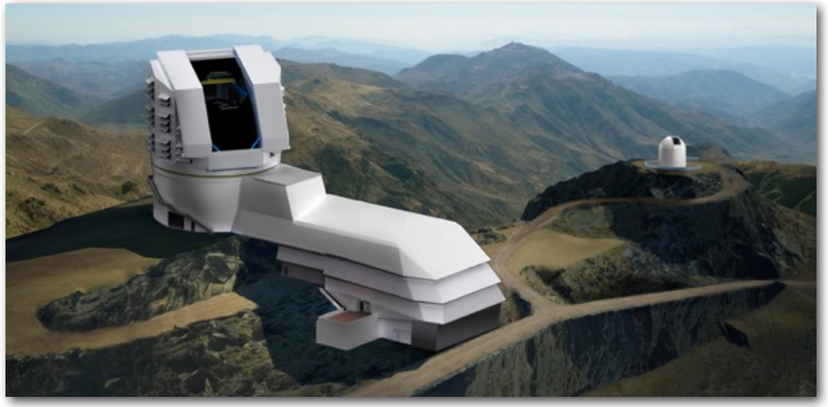
  *taking an inventory of the **solar system***

  *exploring the **transient** optical sky*

  *mapping the structure and evolution of the **Milky Way***

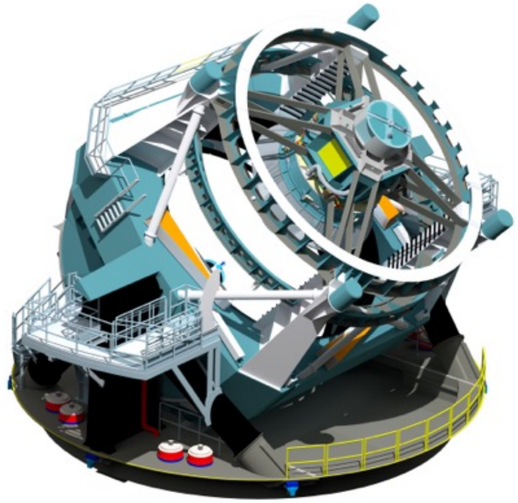
LSST OVERVIEW

OBSERVATORY



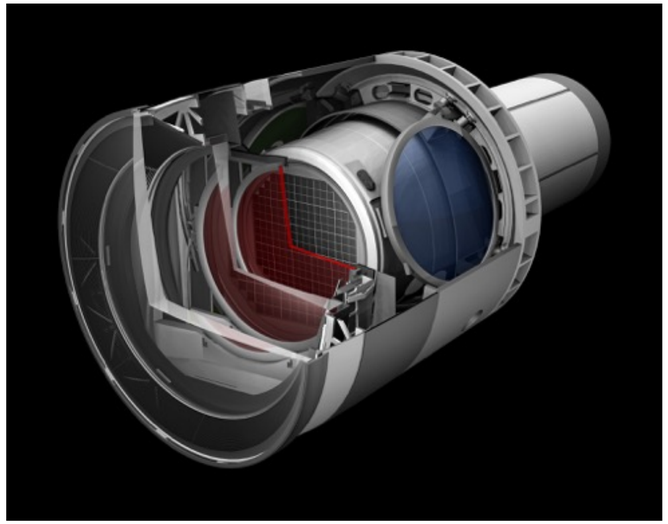
south hemisphere | 2647m a.s.l. | stable air | clear sky | dark nights | good infrastructure

TELESCOPE



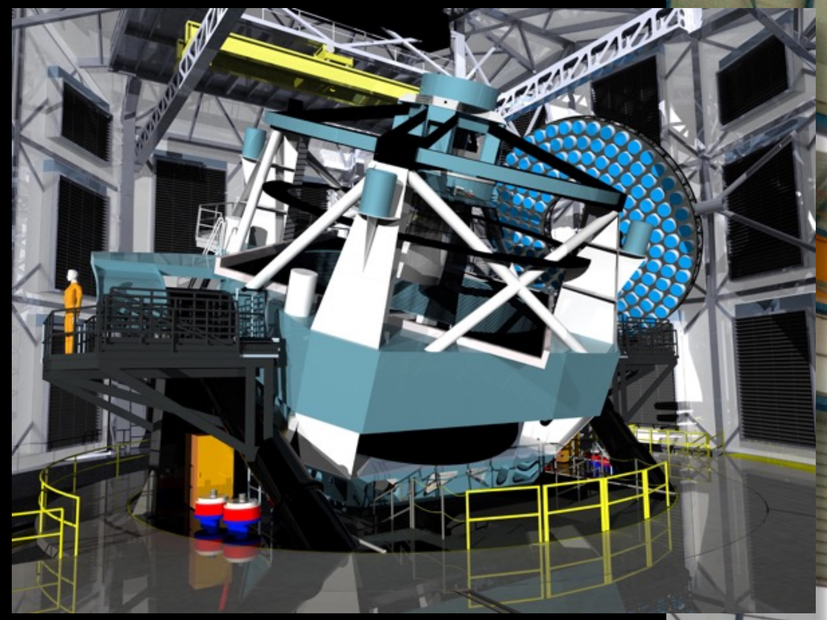
main mirror ⌀ 8.4 m (effective aperture 6.5 m) | large aperture: f/1.234 | wide field of view | compact | 350 ton | to be repositioned about 3M times over 10 years of operations

CAMERA



3.2 G pixels | ⌀ 1.65 m | 3.7 m long | 3 ton | 3 lenses | 3.5° field of view | 9.6 deg^2 | 6 filters ugrizy | focal plane and electronics in cryostat at 173K

Sources: LSST, Gianluca Lombardi (2018-08)



ROTATION: 10 deg/s
3.5° SLEW-SETTLE: 5 s

Sources: LSST, Asturfeito (2018-03)

Data Acquisition

- Raw data

  7.2 GB per image

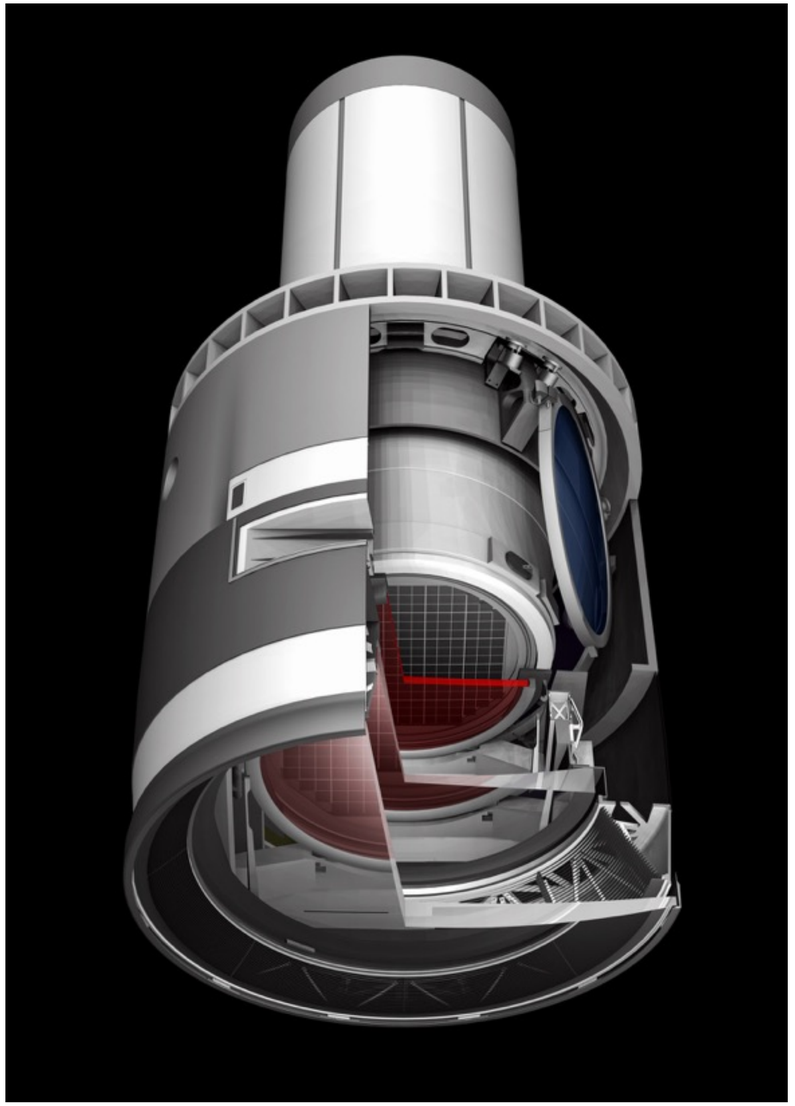
  2000 science images + 450 calibration images per night, 300 nights per year

  20 TB per night, 4.5 PB per year

- Aggregated data over 10 years of operations, including derived data

  images: ~6M exposures, 515 PB

  catalog database: 80 PB



Source: LSST

LSST OVERVIEW: DELIVERABLES

- Deliverable

  *the science-enabling, **ultimate deliverable** of the project will be the **fully reduced data***

  the scientific exploitation of the processed data will be performed by the scientific community

- Open data

  *complete **cumulative data set** (images and catalogs), open to the scientific community of the participating countries, once per year, with no proprietary period*

  ***alerts** of detected variable sources (transients) made available for world-wide distribution **within 60 seconds of observation**, published via standard protocols*

- Open source software: `github.com/lsst`

LSST OVERVIEW: FUNDING AND BUDGET

1

2014-2022 — Construction phase budget: US$ 671M

About 20% of the construction budget devoted to the DATA MANAGEMENT subsystem

2

2019-2034 — Operations phase budget: US$ 41M/year

International collaboration: 25 countries, 39 research institutions

IN2P3

IN2P3

A DISTRIBUTED LABORATORY

in2p3.fr

2500 researchers, engineers and technicians

700 post-docs and PhD students

25 laboratories and research platforms in France, 16 international laboratories

IN2P3 COMPUTING CENTER

- CC-IN2P3

  84 people, 80 FTE, 80% permanent positions

  ~15 M€ overall annual budget

  scientific data center, high throughput computing

  well connected to national and international networks

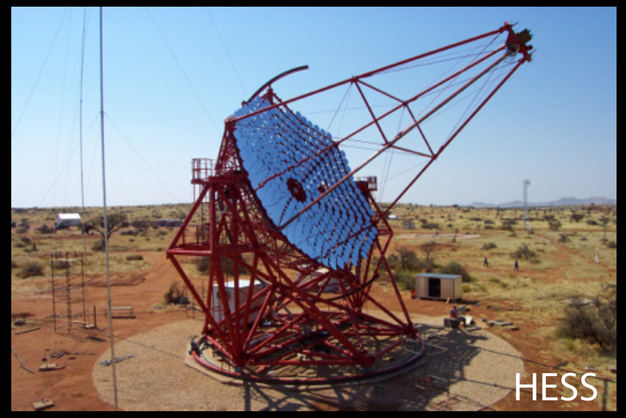
- Shared computing facility supporting the institute's research program

  ~70 projects in high energy physics, nuclear physics and astroparticle physics

- Operations: 24x7

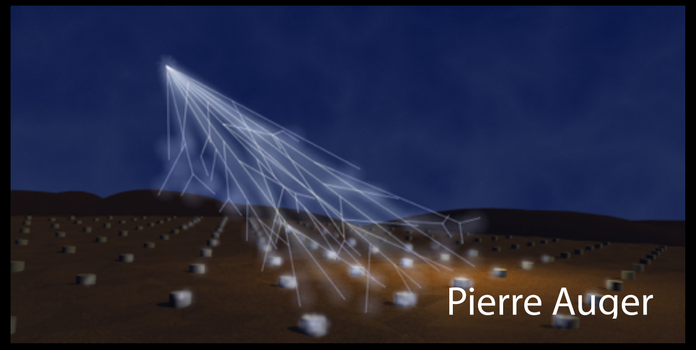
  unattended during nights and weekends

  engineer on duty during off-hours



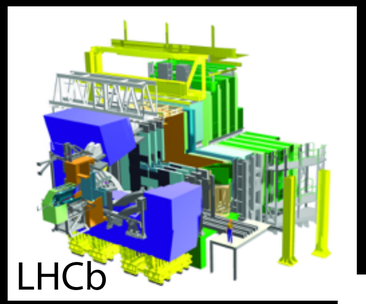
Planck

Fermi

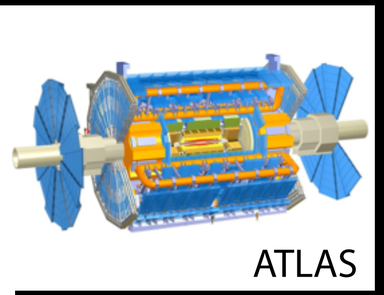


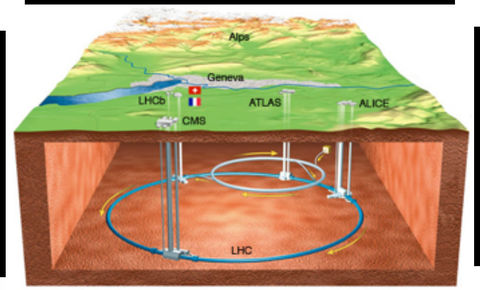
AMS

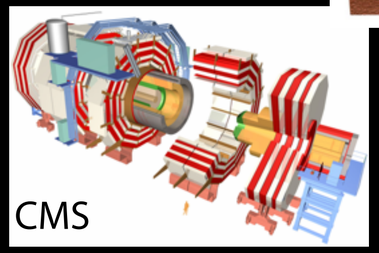
SuperNova Legacy Survey

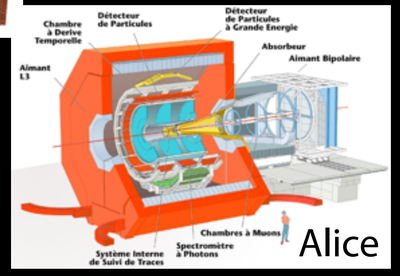


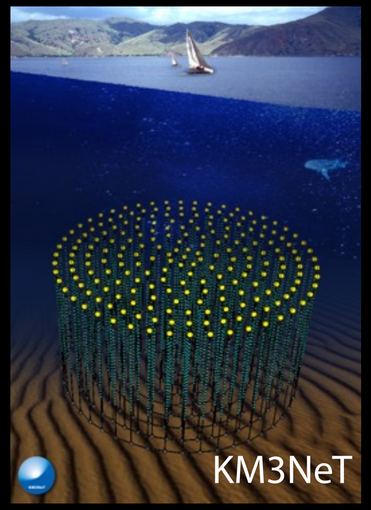
LHC @ CERN











LSST AT IN2P3

- IN2P3 contributes to the construction of the LSST camera

  CCD electronics, filter carousel, filter autochanger and manual loader (design, construction, command and control software)

- IN2P3 is preparing its contribution to **offline data processing** during both the commissioning and operations phases

LSST DATA PROCESSING

LSST DATA MANAGEMENT SUBSYSTEM

- Archival

  *to **record**, **transport** and permanently **store raw data** issued by camera*

- Processing

  *to **detect transients** and **emit alerts** within 60 seconds after observation*

  *once per year, to **produce a data release**: a **self-consistent, immutable dataset**, composed of **processed data since the beginning of the survey***

  *to **develop the software** necessary for processing the data: image processing algorithms (calibration, point spread function, co-addition of images, characterization of objects, processing pipelines, …), catalogue database, middleware (workload management, orchestration, …), data transfer, etc.*

- Publication

  *to **deliver the reduced data** (images + catalogs)*

  to facilitate custom data reduction and individual data analysis

Large Synoptic Survey Telescope

LSST Operations: Sites & Data Flows

HQ Site

Science Operations
Observatory Management
Education & Public Outreach

Base Site

Base Center
Long-term storage (copy 1)

Data Access Center
Data Access & User Services

French Site

Satellite Processing Center
Data Release Production
Long-term Storage (copy 3)

Archive Site

Archive Center
Alert Production
Data Release Production
Calibration Products Production
EPO Infrastructure
Long-term Storage (copy 2)

Data Access Center
Data Access and User Services

Summit Site

Telescope & Camera
Data Acquisition
Crosstalk Correction

Source: LSST

LSST DATA MANAGEMENT CONTRIBUTORS

Princeton University

UNIVERSITY of WASHINGTON

National Optical
Astronomy Observatory

SLAC

SLAC National Accelerator
Laboratory
Stanford University

Infrared Processing and
Analysis Center
California Institute of
Technology

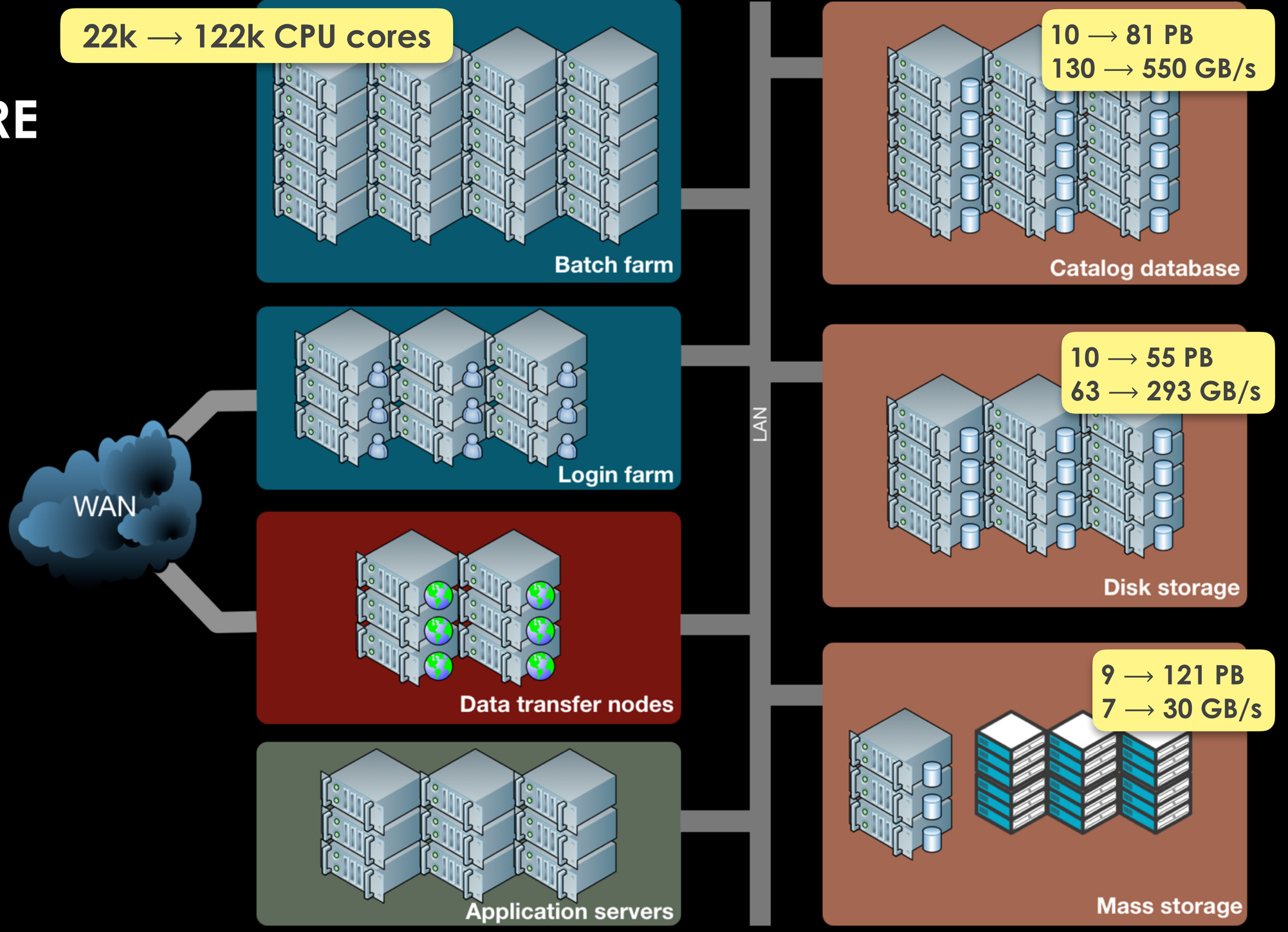
DATA RELEASE PROCESSING CENTRES

National Center for
Supercomputing Applications
University of Illinois at Urbana-
Champaign

CNRS / IN2P3 computing center

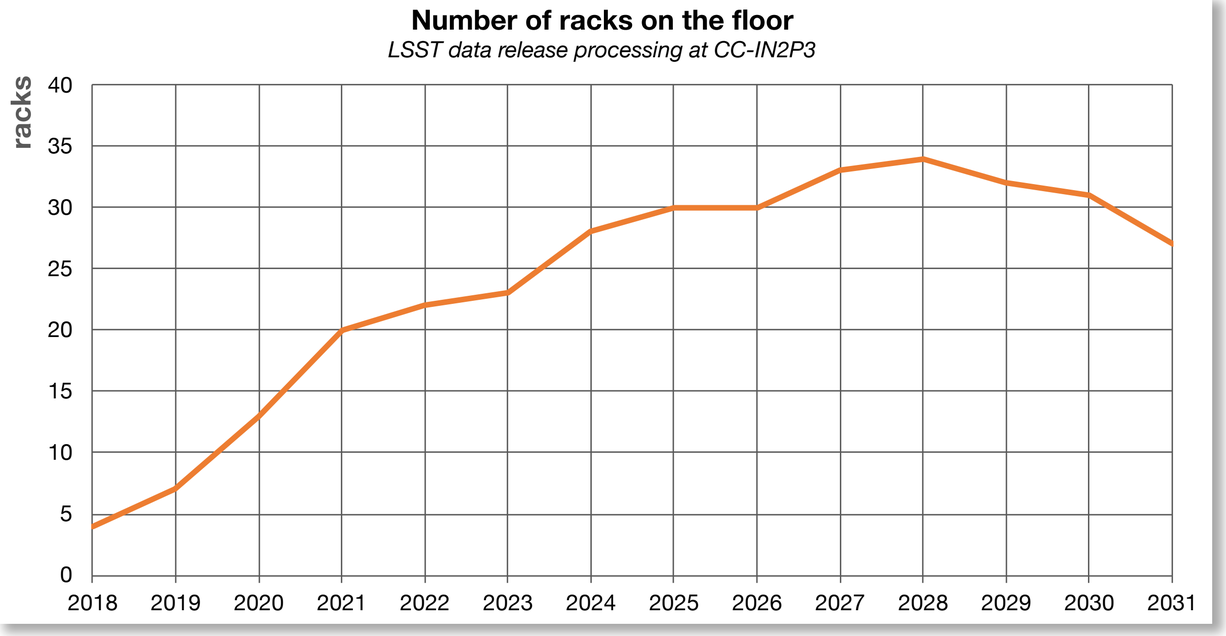
Envisioned architecture

(preliminary figures)



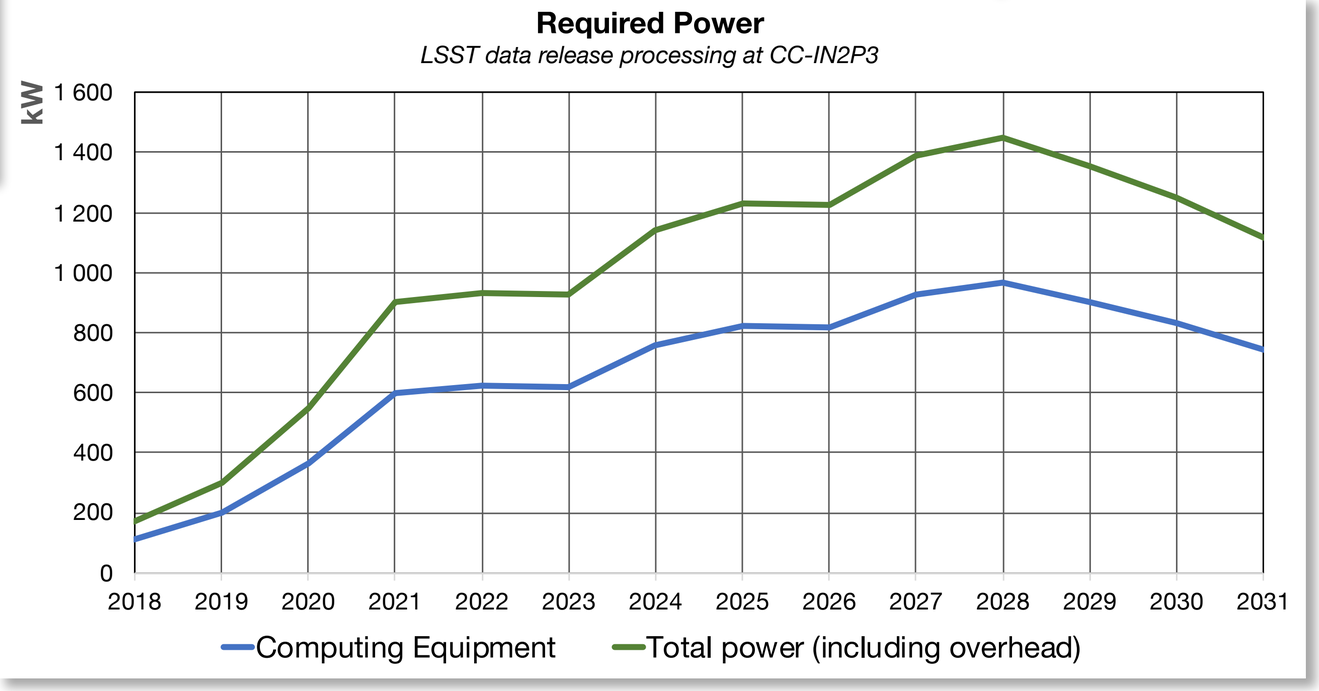
2022 ⟶ 2032

MACHINE ROOM INFRASTRUCTURE



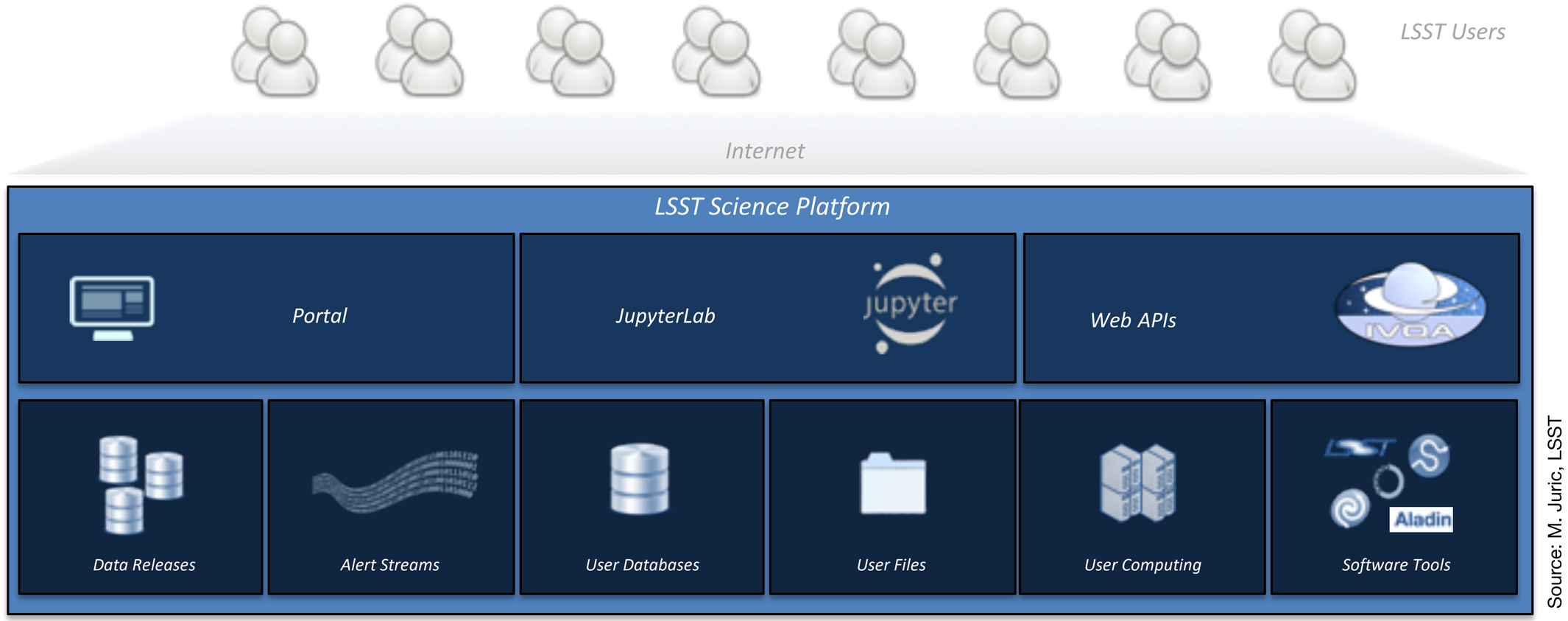
Racks
peak 34 racks

Power
peak 1.4 MW



Science Platform

LSST Science Platform



Source: M. Juric, LSST

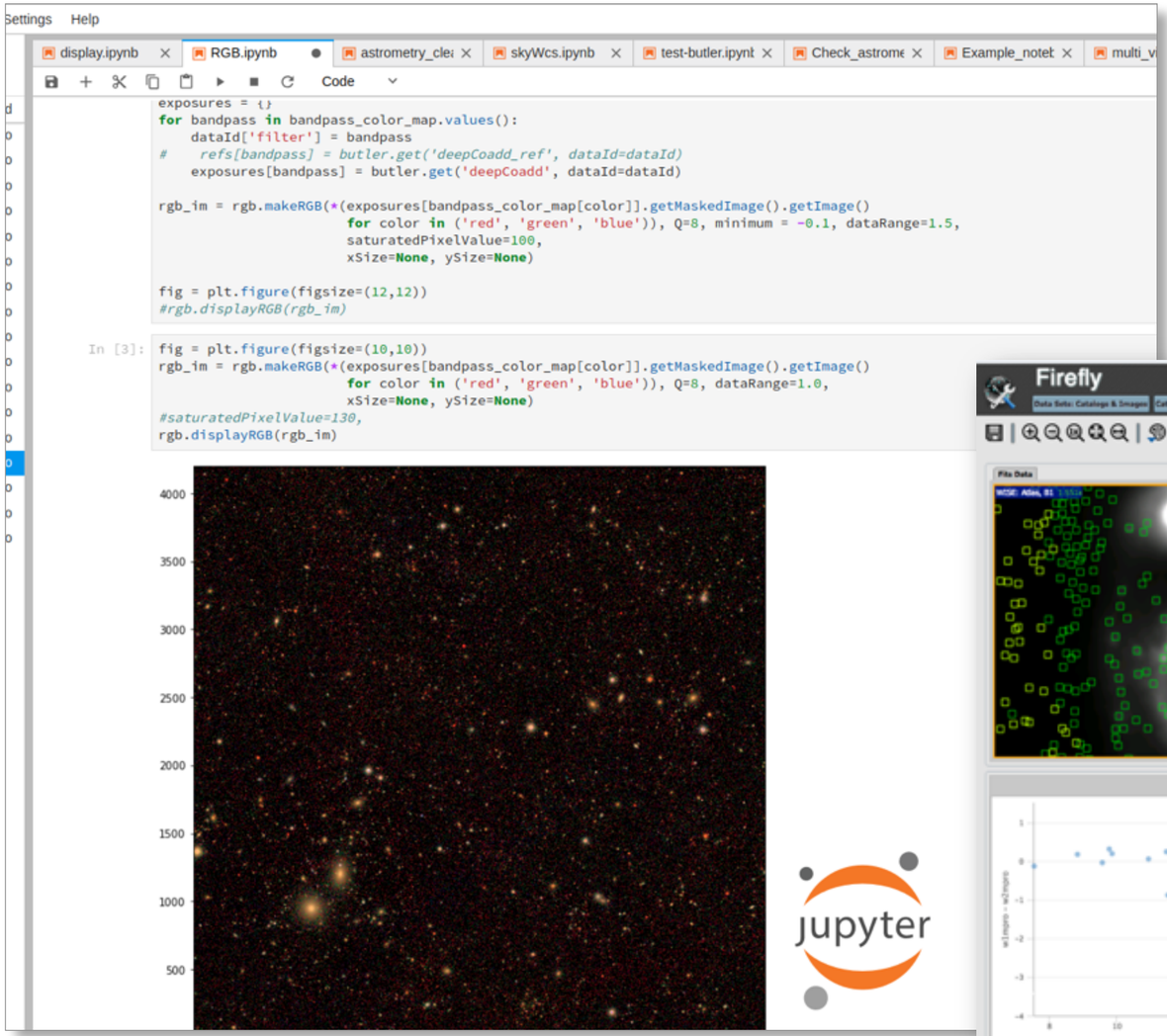
Set of integrated web applications and services, through which the scientific community will access, **visualize**, subset and perform **next-to-the-data analysis** of the data

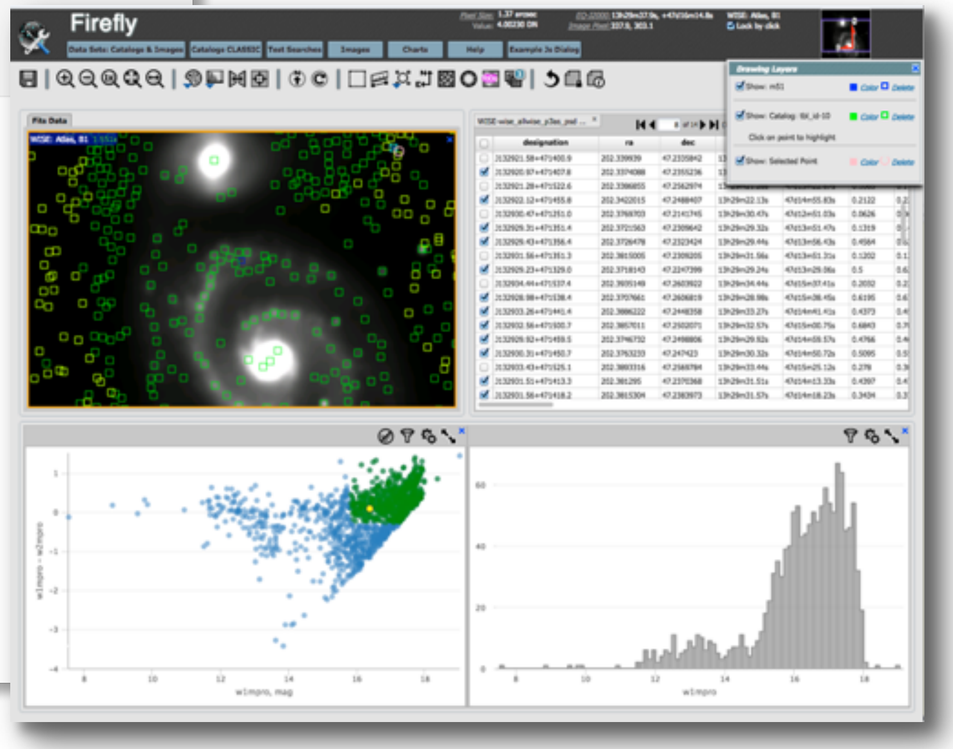
LSST SCIENCE PLATFORM PROTOTYPE

doc.lsst.eu

Interactive computing based on Python notebooks and web-based visualisation tools

Prototype components deployed at CC-IN2P3 and being routinely used by scientists





Sources: D. Boutigny, X. Wu et al.

SOFTWARE DISTRIBUTION

- LSST science pipelines automatically delivered

  *both **stable** and **weekly** releases appear as if they were locally installed under* `/cvmfs/sw.lsst.eu`

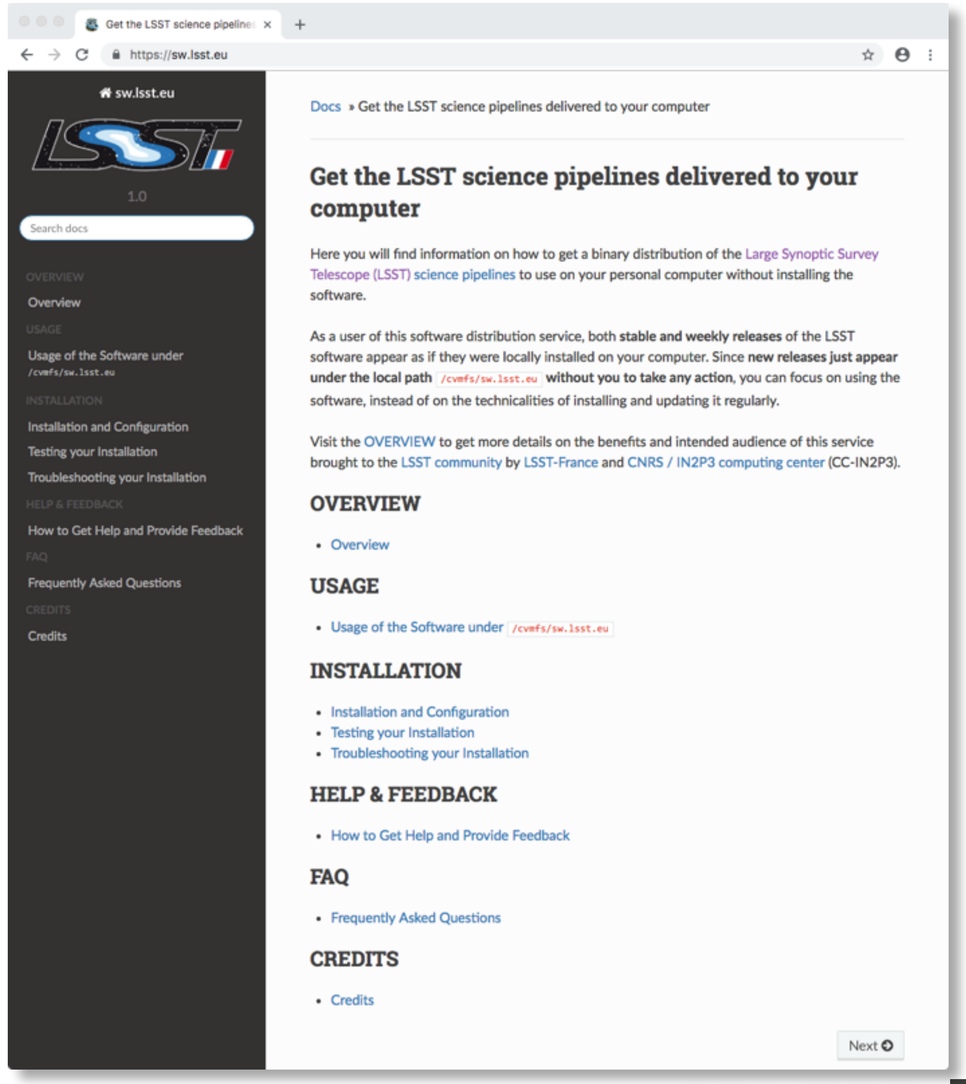
  *mechanism used for delivering the software to computers in both the CC-IN2P3 **login** and **batch farms** as well as to the **scientists' personal computers***

  lower the barriers for end users to use the LSST software

  useful for reproducibility

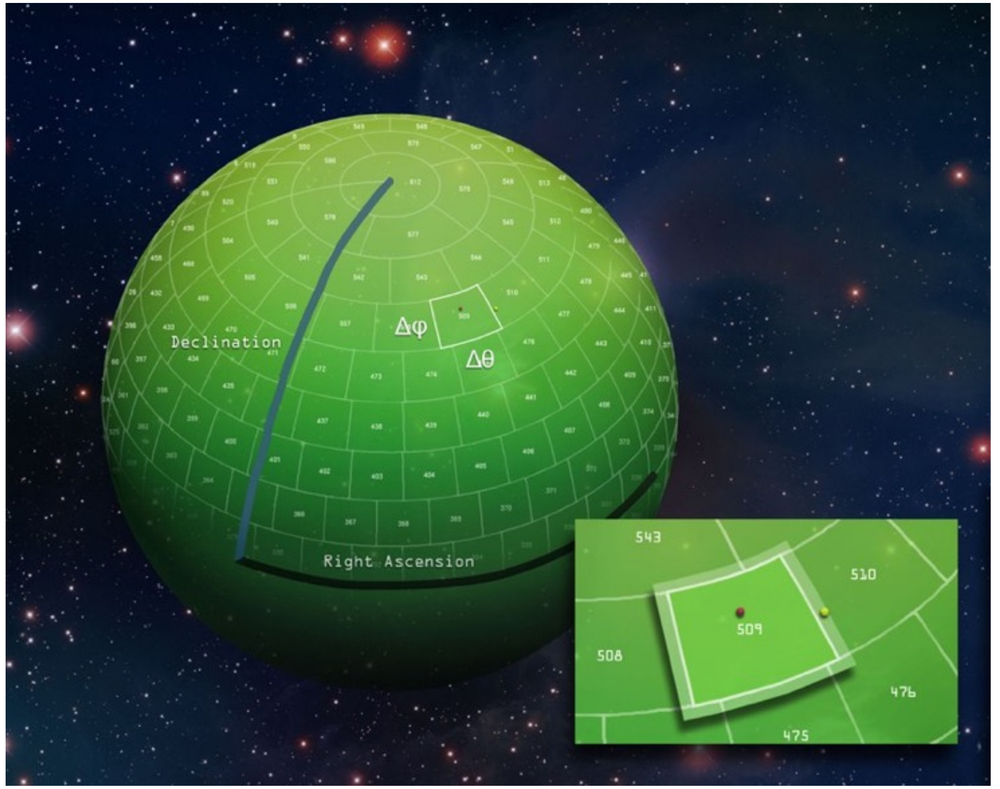
  alternative mechanisms: Docker images, sources

sw.lsst.eu



CATALOG DATABASE

- Qserv: custom, distributed relational database
  - *spatial partitioning by sky coordinates, with overlaps*
  - *map-reduce model*
  - *very high number of rows:* ***~37 trillions***
- CC-IN2P3 hosts and operates one of the development clusters
  - *development effort lead by SLAC, with contribution by IN2P3 LPC Clermont*
- Currently using hardware lent by Dell in the framework of an institutional partnership



SLAC

LPC Particules pLasmas Univers applicationS
Laboratoire de Physique de Clermont

DELL

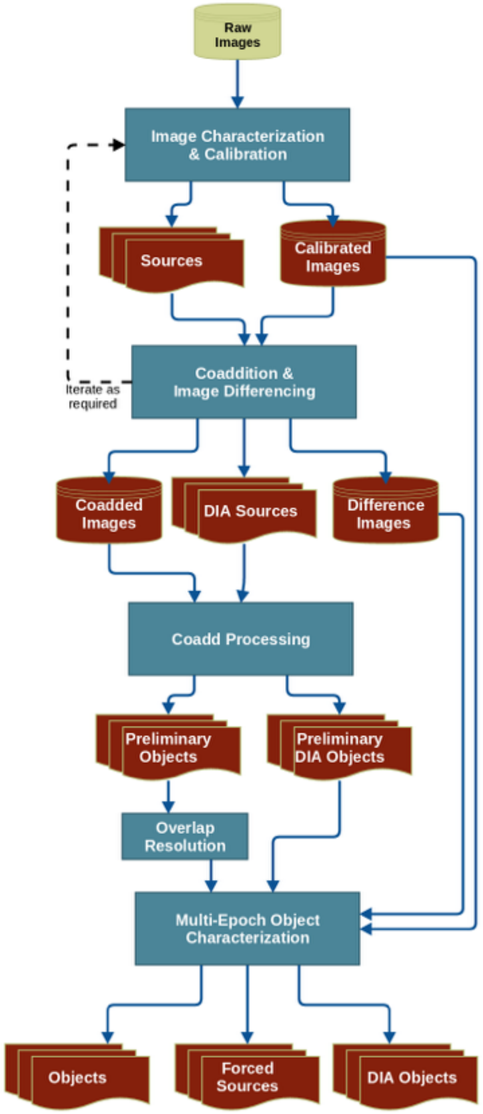
DATA RELEASE PROCESSING

- Bulk data processing for building the annual data release

  *every year, the **entire dataset since the beginning of the survey** is reprocessed to produce an immutable set of calibrated images and catalogs and to update the references for nightly alert production*

- Currently exercising the LSST software for processing simulated data

  Dark Energy Science Collaboration



Connectivity & Data exchange

- Allocated bandwidth between CC-IN2P3 and NCSA: 20 Gbps

  bottleneck link is currently 10 Gbps: expected upgrade to full 20 Gbps before end 2018

- We need to demonstrate capacity to import 20 TB of raw data per night from NCSA (RTT: 110 ms)

  in addition to capacity to exchange data products with NCSA

- Currently exercising regular data exchanges with NERSC (RTT: 150 ms)

CONNECTIVITY & DATA EXCHANGE (CONT.)

Data flow: **NERSC** (GPFS) → **CC-IN2P3** (GPFS) [3 servers, 4 clients]

Aggregated application-level network throughput: **1.5 GB/s (12 Gbps)**

secure HTTP ⇒ integrity, confidentiality

pull model, disk-to-disk transfer, wide area network, 150ms RTT

Connectivity provided by

RENATER
CONNECTEUR DE SAVOIRS

SUMMARY

SUMMARY

- LSST is a world-class, high-profile project in optical astronomy

  high expectations from the scientific community and from the funding agencies about what LSST will bring over the next decade

- IN2P3 is preparing to contribute to the annual production of the LSST data releases

  integral copy of the data (both images and catalogs) to be available at CC-IN2P3 for French scientists members of the project

- Significant investment being made to realise the science discovery potential

  quantified roadmap established, R&D activities ongoing and now entering the deployment phase

We are made of stellar ash. Our origin and evolution have been tied to distant cosmic events. The exploration of the Cosmos is a voyage of self-discovery.

CARL SAGAN, COSMOS

Questions & Comments