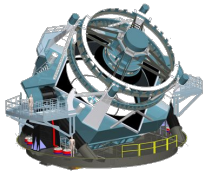
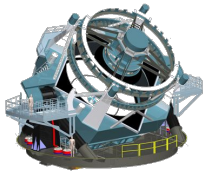


Computing Overview

LSST-France

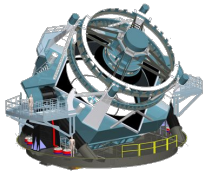
Dominique Boutigny - LAPP

February 3-5, 2020



News - IN2P3

- IN2P3 computing review October 24th, 2019
  - Berrie Giebels - Volker Beckmann(remotely) - Rodolphe Cledassou- Pierre-Étienne Macchi
  - https://indico.in2p3.fr/event/19728/
  - People looked happy
  - But unfortunately we haven't received any other feedback…
- The positive thing is that we obtain more or less the resources that we need to work
  - Essential role of Fabio to coordinate all the DRP project (and much more) at CC-IN2P3
    - But we also need input from IN2P3 as Fabio's official mandate as DRP Coordinator is over



News - Project

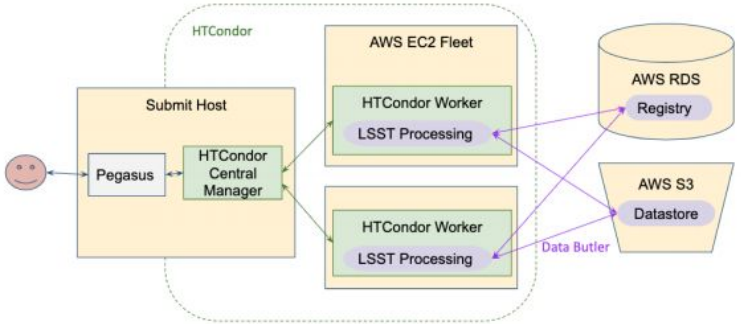
Politics

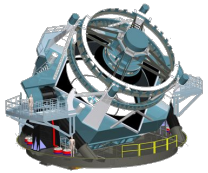
- As DOE will cover the extra operation cost associated to the new LSST Data Policy, the project is considering to move the US DRP from NCSA to a DOE lab: SLAC, FNAL or BNL

Technics:

- *butler gen3* is being released now
  - Will have a huge impact on the way we are processing jobs
  - Consequences also for DESC
    - Slack discussion channel #desc-dm-gen3
  - We need to start investigating *gen3* ASAP
    - Understand how to build workflows with *gen3*
- DM stack will probably produce *parquet* files in the future
  - Spark and DASK are being considered

See : https://dmtn-137.lsst.io/





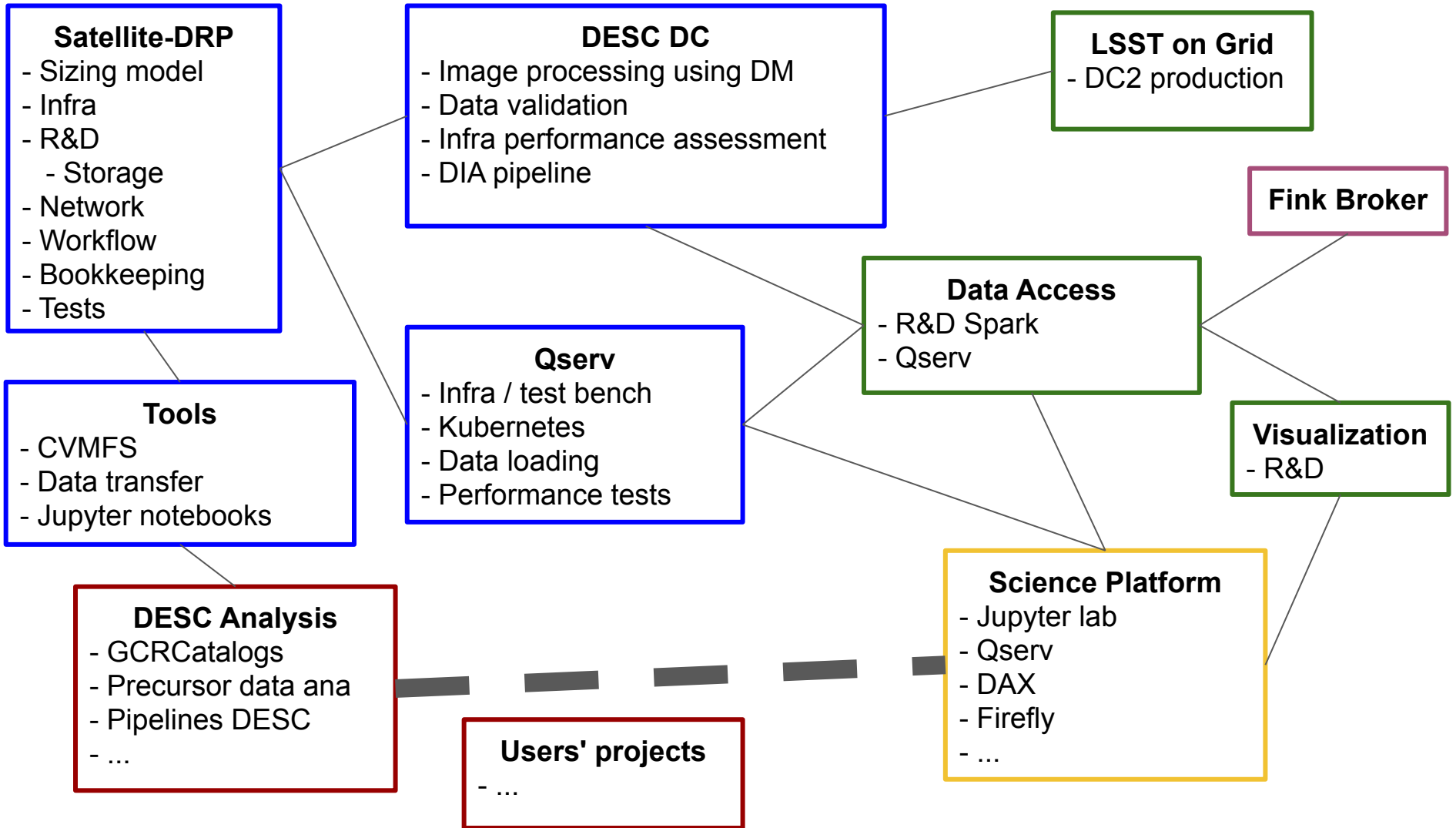
News - LSST-France

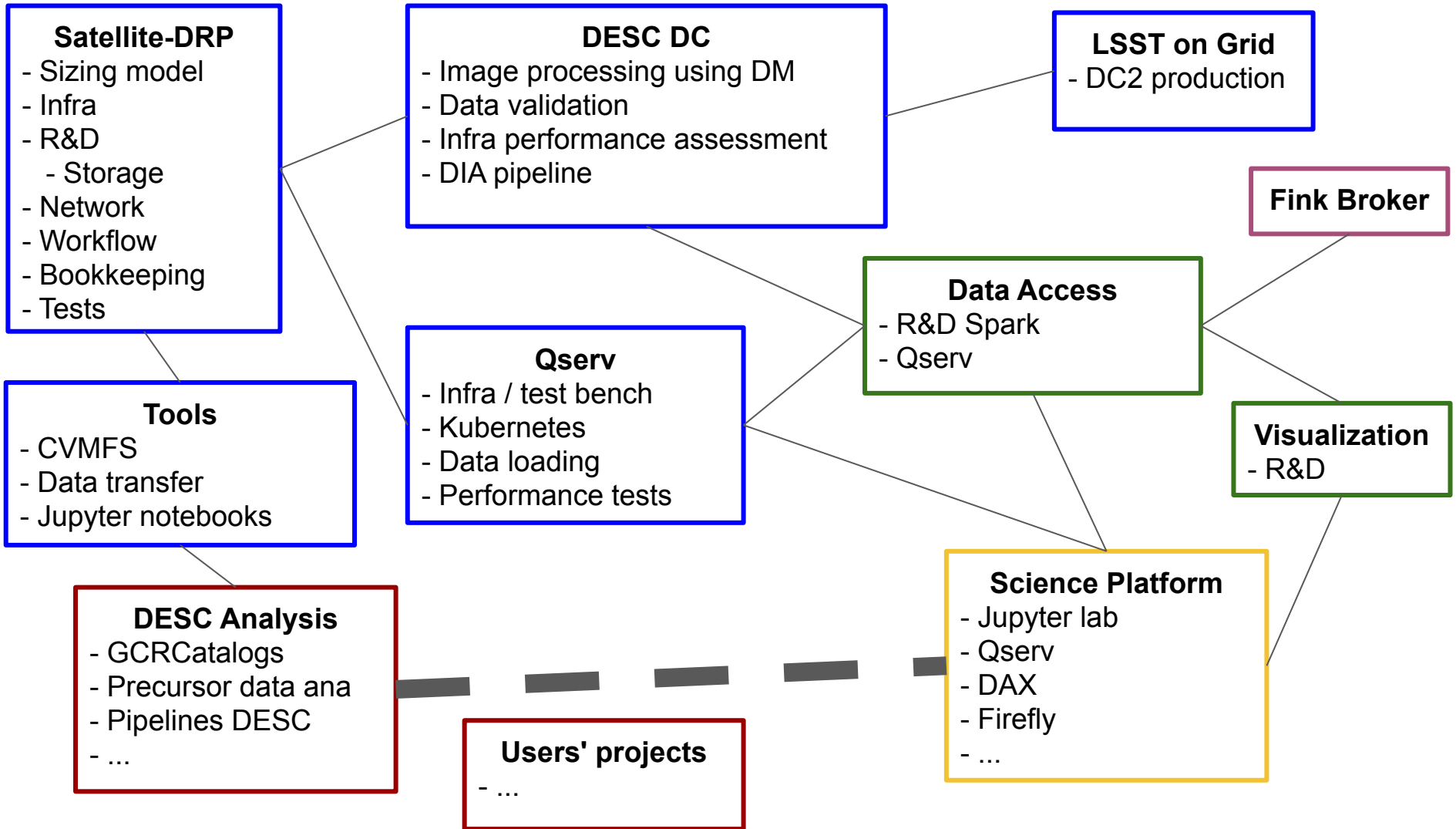
Johann has been appointed co-convener of the **CO**mputing Working Group together with Heather Kelly

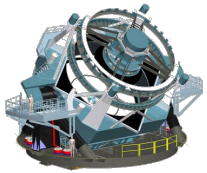
- Congrats !

A 2 year position for a Data Scientist is opened at LAPP in the framework of ESCAPE WP5

- Deploy, test and possibly extend the LSST Science Platform as a use case / example for ESAP (ESFRI Science Analysis Platform)


Satellite-DRP
- Sizing model
- Infra
- R&D
- Storage
- Network
- Workflow
- Bookkeeping
- Tests
DESC DC
- Image processing using DM
- Data validation
- Infra performance assessment
- DIA pipeline
LSST on Grid
- DC2 production
Fink Broker
Data Access
- R&D Spark
- Qserv
Qserv
- Infra / test bench
- Kubernetes
- Data loading
- Performance tests
Tools
- CVMFS
- Data transfer
- Jupyter notebooks
Visualization
- R&D
DESC Analysis
- GCRCatalogs
- Precursor data ana
- Pipelines DESC
- ...
Science Platform
- Jupyter lab
- Qserv
- DAX
- Firefly
- ...
Users' projects
- ...


Satellite-DRP
- Sizing model
- Infra
- R&D
- Storage
- Network
- Workflow
- Bookkeeping
- Tests
DESC DC
- Image processing using DM
- Data validation
- Infra performance assessment
- DIA pipeline
LSST on Grid
- DC2 production
Fink Broker
Data Access
- R&D Spark
- Qserv
Qserv
- Infra / test bench
- Kubernetes
- Data loading
- Performance tests
Visualization
- R&D
Tools
- CVMFS
- Data transfer
- Jupyter notebooks
DESC Analysis
- GCRCatalogs
- Precursor data ana
- Pipelines DESC
- ...
Science Platform
- Jupyter lab
- Qserv
- DAX
- Firefly
- ...
Users' projects
- ...



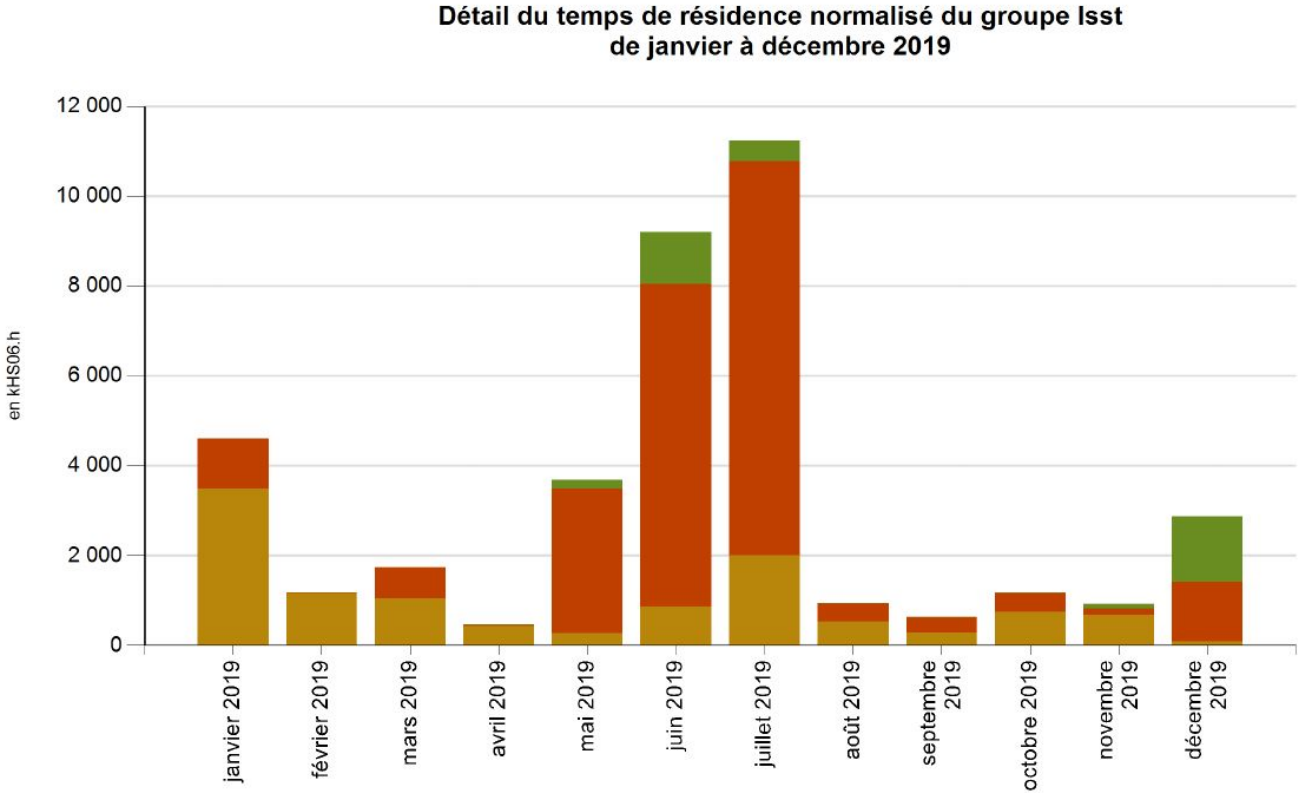
Overall Strategy (as presented as the October review)

- DRP is the top priority
- All the rest is driven by science (not engineering)

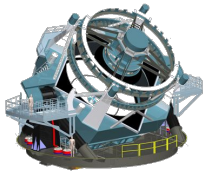
Goals:

- Host a complete LSST dataset in France
- Set up a computing environment to guarantee an efficient data access to the IN2P3 community
- Design this environment in order to be scalable (in order to extend the data access to a larger community if decided)
- Do not reinvent the wheel
- Guarantee a complete compatibility (interoperability ?) between the DESC environment in the US and in France
- Use IN2P3 expertise and flexibility to test and propose alternative technical solutions to LSST and DESC
- Systematically test software components to check their usability in a science context

CC-IN2P3 usage in 2019



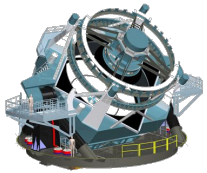
Total: 38.6 MHS06.h / 160 MHS06.h demandé (24%)



CC-IN2P3 request for 2020

	Trimestre 1	Trimestre 2	Trimestre 3	Trimestre 4
CPU	40 M	40 M	40 M	40 M
SPS	+700 TB	+600	+0	+500 TB
dCache LSST VO	+100 TB	+100 TB	+0	+0
Throng /pbs/throng/lsst	+1 TB	+1 TB	+0	+0
Bandes (HPSS)	+200 TB	+100 TB	+100 TB	+0

- 10 GPU (constantly all the time)
- + some services : OpenStack, PostgreSQL, CVMFS, network...



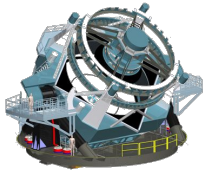
Test on the new Jean Zay system at IDRIS

Jean Zay is the new HPC system at IDRIS: 13.9 Pflops with

- 1528 "standard" 40 core nodes
- 261 nodes with 4 GPUs Nvidia V100 SXM2 32 GB

Got a 15 000 hours allocation to test DESC imSim simulation production + test to write output directly at CC-IN2P3 using high bandwidth network connection

- Took me ~2 weeks to install LSST stack → OK now
- Good interaction with IDRIS users support



LSST / DESC environment at CC-IN2P3

LSST

- All LSST software available through CVMFS
- All reference catalogs
- Several datasets - More can be copied upon request

DESC

- All previous DC2 pre-releases
- Complete GCRCatalogs environment maintained
  - conda software not strictly the same as at NERSC
    - Plan to work on this soon
    - But ok in the meantime
  - Accessible from Jupyter notebooks through ***stackyter.py*** or from regular shells
  - setup everything with source /pbs/throng/lsst/software/desc/setup.sh
    - Nothing more
    - Even give you access to a working ds9
  - Please ask if you need something specific
  - Slack channel #in2p3