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

France Data Facility Overview

fabio hernandez, quentin le boulc'h, gabriele mainetti

Data Facilities and Multisite Processing, Lyon, February 11th, 2025

IN2P3

National Institute of Nuclear and Particle Physics

2700+ scientists, engineers and technicians

800 doctoral students and post-doctoral fellows

10 national platforms and research infrastructures

21 national and international laboratories

www.in2p3.fr


Satellites spatiaux :
EUCLID
PLANCK
FERMI
AMS
DUNE
SPIRAL 2
AGATA
SUPERNEMO
Caen
Modane
Genève
Pise
Toulon
LHC
T2K
Tsukuba
Chicago
La Palma
Kaiping
CTA
KM3NeT
VIRGO
BELLE 2
JUNO
Désert d'Atacama
La Serena
AUGER
Gamsberg
Malargue
RUBIN OBS.
(LSST)
HESS
*Environ 80 autres expériences ou projets
Futures expériences - engagement en discussion

CC-IN2P3

- 80+ people
- CPU: 650 compute nodes, 57 k threads
- Storage: 80+ PB disk, 175+ PB tape

EQUIPMENT PURCHASE

- ## Annual funding cycle

  dedicated line of funding for Rubin equipment from CNRS (IN2P3 mother organisation)

  multi-year planning

- ## Purchase schedule

  annual equipment purchase for Rubin typically happens in the second semester every year

  purchase procedure subject to regulations about public equipment purchase in France

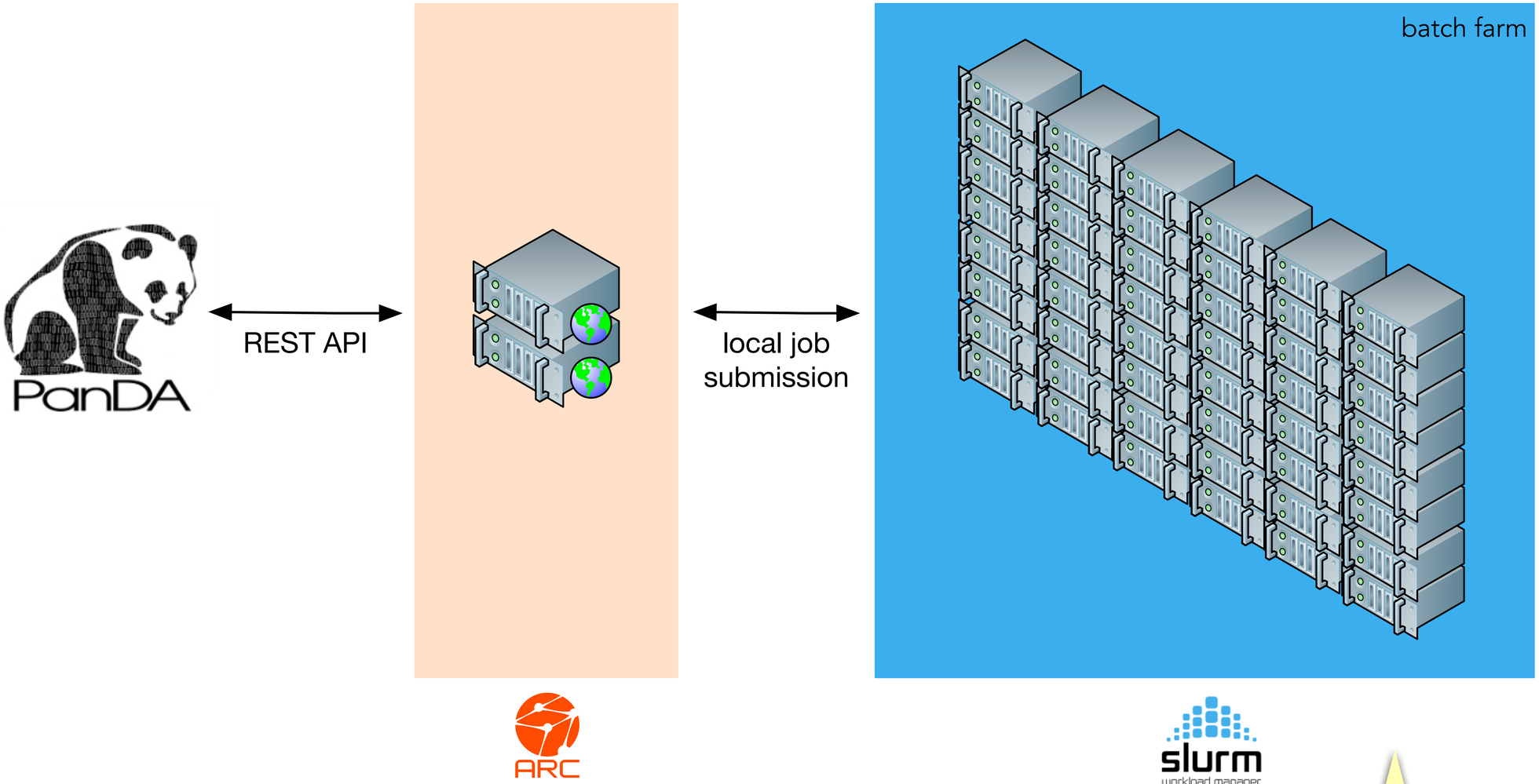
  equipment enters production by the end of the year

- ## Equipment production lifetime

  CPU and disk storage: so far 5 years, now switching to 7 years

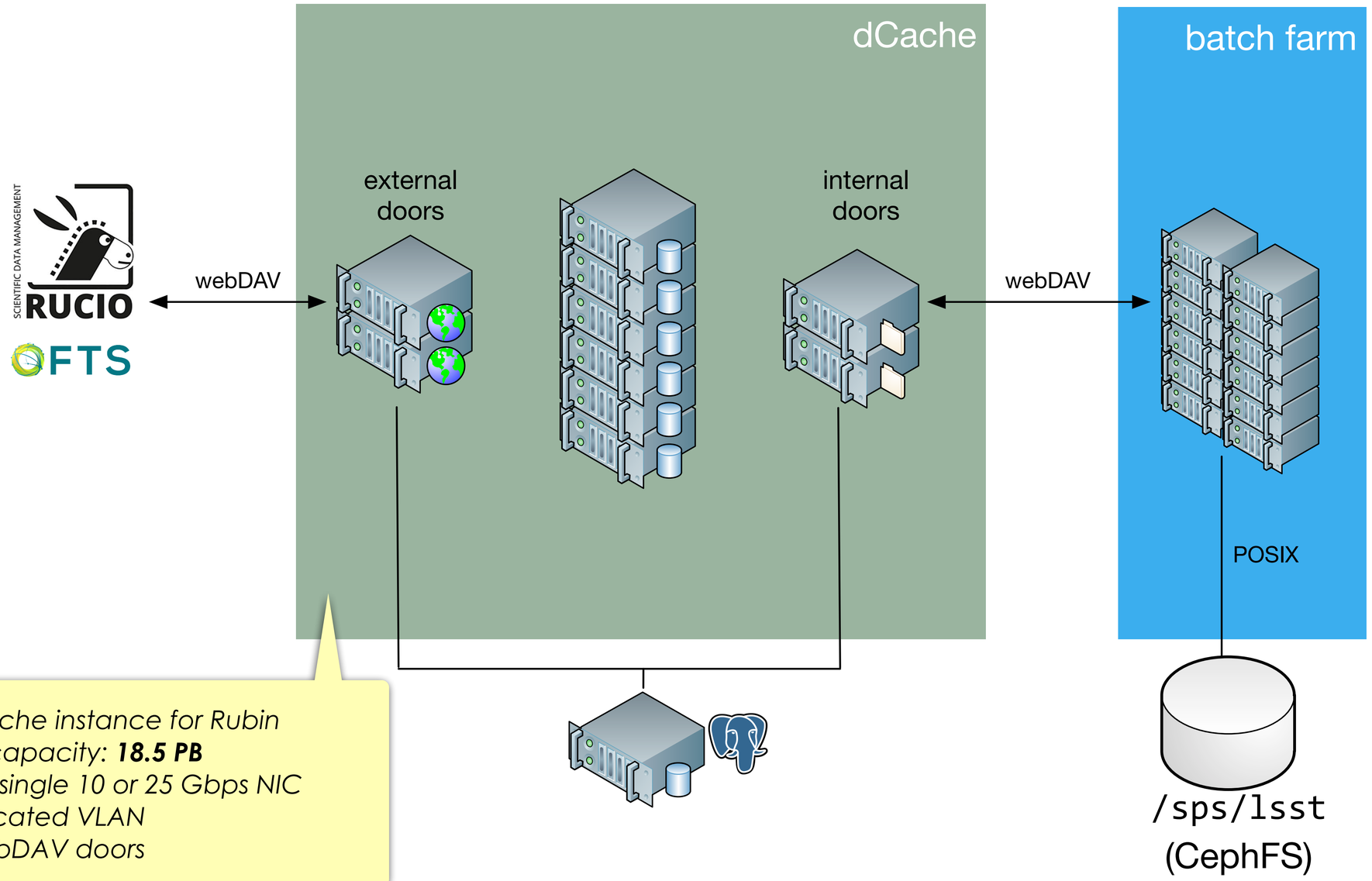
  tape storage: typically 7 years, subject to technological evolution of tape recording and tape drives

COMPUTE



current allocation for Rubin composed of
***5000 CPU threads**, **9+ GB per thread**, hyperthreading enabled*
dedicated Slurm partition

STORAGE



dedicated dCache instance for Rubin
installed capacity: ***18.5 PB***
62 disk pools with single 10 or 25 Gbps NIC
dedicated VLAN
4 webDAV doors

BUTLERS

- Individual scientists can have their own butlers

  datastore on CephFS (under `/sps/lsst`) and registry database on a dedicated Postgres instance

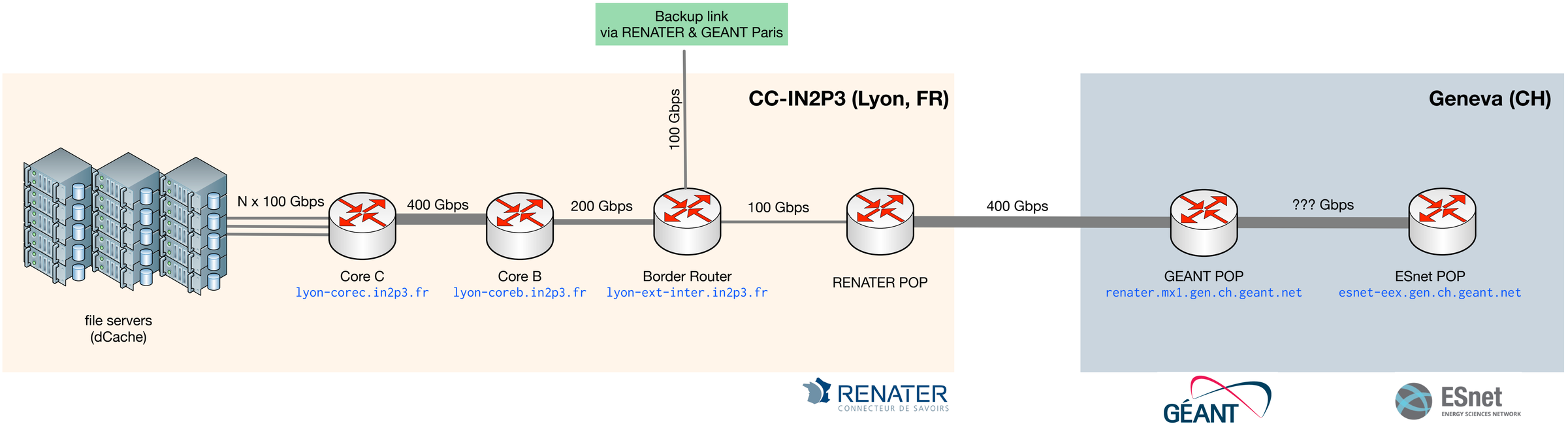
  each individual has their own Postgres database

- Data release processing

  datastore on dCache (webDAV)

  registry database on a dedicated Postgres instance (full flash)

International Connectivity



RENATER connects CC-IN2P3 to ESnet via GEANT point of presence in Geneva (primary) or Paris (secondary)

International Connectivity (cont.)

- Two Perfsonar instances: one for bandwidth measurements and another for latency
- Bandwith

  https://ccperfsonar3.in2p3.fr

  dedicated to collect throughput measurements against other Rubin sites

  single 100 Gbps NIC, connected to the same subnetwork of dCache servers

  currently only probing LANCS: other sites to be added as soon as we get the details of their instances

- Latency

  https://ccperfsonar2.in2p3.fr

- Details: PREOPS-1144

PLATFORM FOR SCIENCE ANALYSIS

- Experimental instance of Qserv connected to a local instance of the Rubin Science Platform
  - *single Kubernetes cluster on bare metal*
  - *5 hosts for RSP: https://data-dev.lsst.eu*
  - *17 hosts for Qserv (2 masters + 15 workers)*
- RSP configured to use CC-IN2P3 single sign on
- Exploratory test of access the UK IDAC's Qserv via TAP from local RSP
  - *no noticeable performance impact induced by network latency*
  - *details in DMTN-298 by Mainetti et al.*
- Final shape of analysis platform at IN2P3 still unclear
  - *operating a local instance of Qserv is excluded for budgetary reasons and because it doesn't look essential for the analysis foreseen in France*
  - *ongoing work to identify needs specific for the science cases relevant for French scientists*

SOFTWARE DISTRIBUTION

- Stratum 0 of CernVM-FS repository sw.lsst.eu operated by FrDF
  as well as stratum one

  both CERN and Open Science Grid provide a replica stratum one (details)
- Contents
  weekly and stable releases, built from sources

  in the form of both a Conda environment and relocatable Apptainer container images with the Conda environment embedded

  AlmaLinux (`x86_64`, `aarch64`), CentOS (`x86_64`) and macOS (`x86_64`, `arm64`)

  ⚠ *CentOS and macOS `x86_64` to be deprecated hopefully very soon*
- Deployed at all the Rubin data facilities
  and beyond (e.g., NERSC), individual personal computers

ONGOING WORK

- Working closely with CM team executing DC2 and RC2 single-site processing campaigns

  job submission via BPS + PanDA at USDF

  some products (mainly plots) sent back to USDF for analysis using data replication services

- Experimenting with Slurm compute nodes equipped with CPUs with aarch64 architecture

  4 compute nodes running RHEL v9, 256 CPU threads per worker

  no issues noticed so far executing typical workloads (DC2, RC2)

  next step is to measure performance per watt and total cost of ownership, to assess the interest of this architecture for Rubin workloads

Ongoing work (cont.)

- Exploring mechanisms to quantify resource usage of select pipetasks

  goal: to help pipetask developers identify opportunities for improvement

  focus on RAM usage and CPU

  see talk tomorrow by Antoine and Quentin

- Preparing the configuration for archiving raw data on tape

  taking into account the constraints imposed by how Rubin intends to manage data using Rucio + FTS, Butler, as well as dCache and HPSS

  slot devoted to this topic on Wednesday: to collect inputs from experts and explore paths forward

Questions & Comments