The Cherenkov Telescope Array production system for data-processing and Monte Carlo simulation

Luisa Arrabito[1]
K. Bernloehr[2], J. Bregeon[1], P. Cumani[3], T. Hassan[3], A. Haupt[4], G. Maier[4], T. Michael[5], A. Moralejo[3], N. Neyroud[6]
for the CTA Consortium
F. Stagni[7], A. Tsaregorodtsev[8] for the DIRAC Consortium

[1]LUPM CNRS-IN2P3 France
[2]MPIK Germany
[3]IFAE Spain
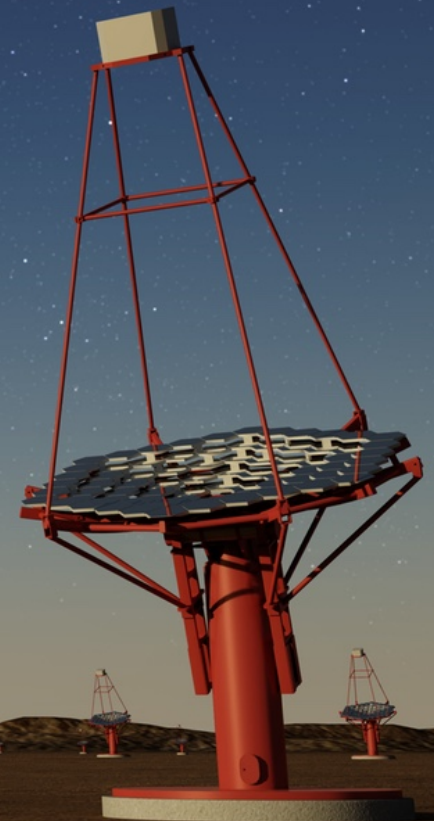[4]DESY Germany
[5]CEA Saclay France
[6]LAPP CNRS-IN2P3 France
[7]CERN
[8]CPPM CNRS-IN2P3 France

Outlook

- CTA overview
- Production setup for MC simulations
- Tests on cloud resources integration
- Data driven workflow management
- Conclusions

CTA (Cherenkov Telescope Array)

- Next generation IACT, VHE gamma-rays Observatory
- Worldwide collaboration, 1500 members
- Scientific goals
  - Cosmic ray origins, High Energy astrophysical phenomena, fundamental physics and cosmology

- Two Cherenkov telescope arrays
  - Northern Site (La Palma, Spain): 4 large size, 15 mid-size telescopes
  - Southern site (Paranal, Chile): 4 large size, 25 mid-size, 70 small size telescopes
- Project schedule
  - Construction and deployment: 2019-2025
  - Science operations: from 2022, for ~30 years
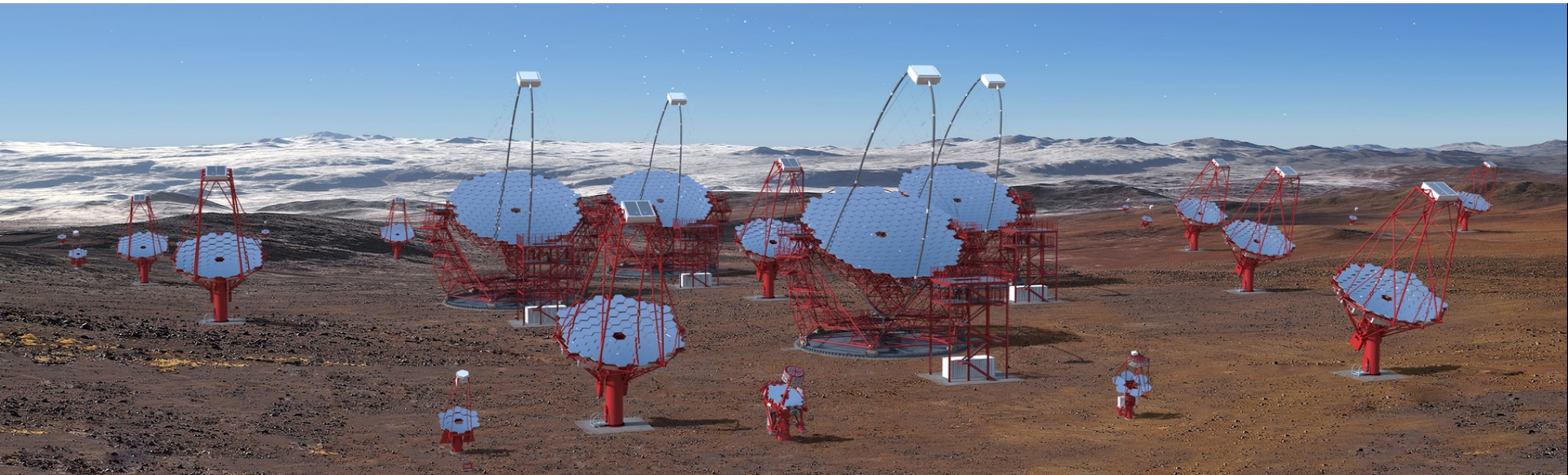
CTA computing model at a glance
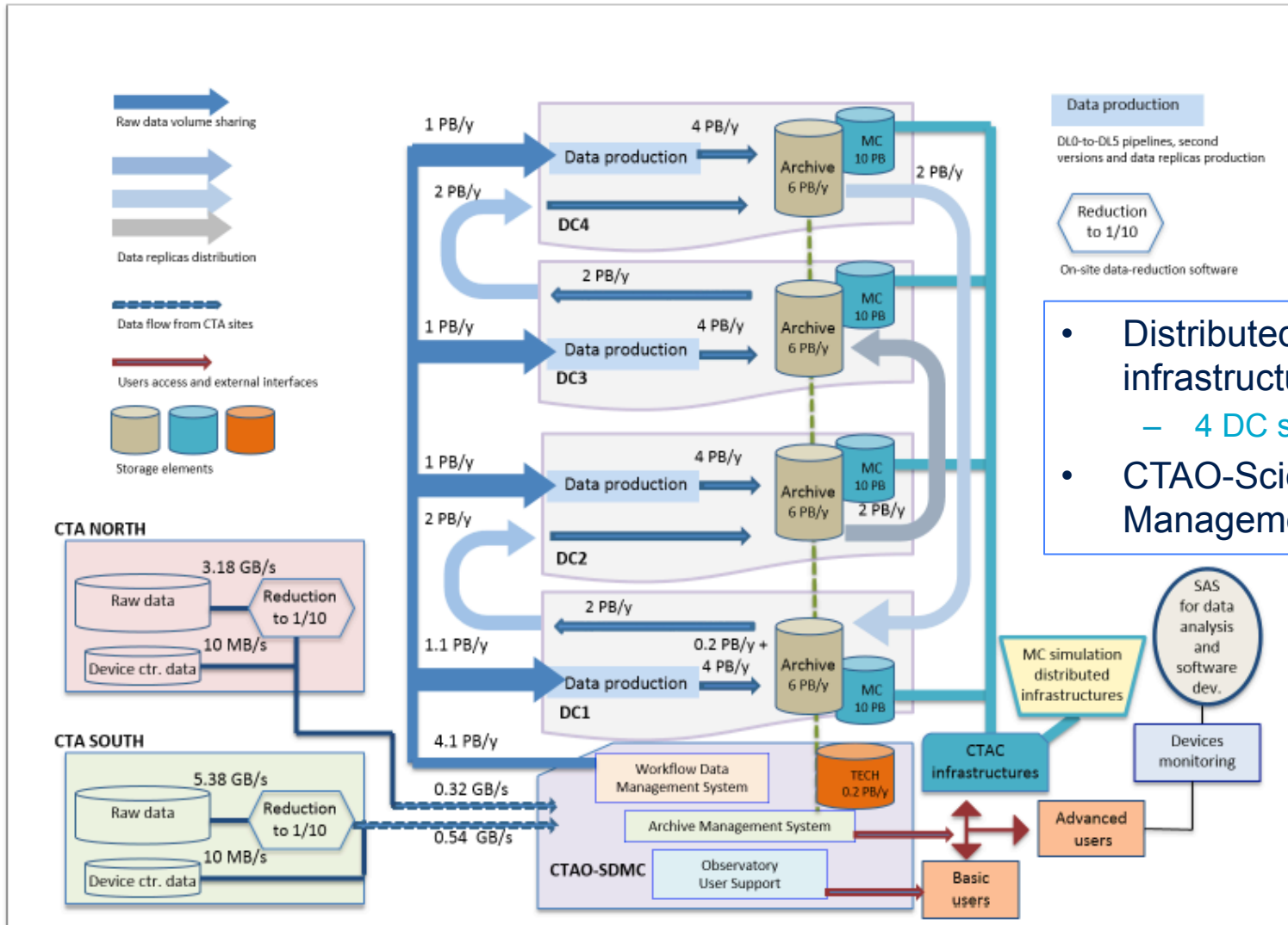
CTA data volume

- Raw-data rate
  - CTA South: 5.4 GB/s
  - CTA North: 3.2 GB/s
  - 15% of observation time per year
- Raw-data volume
  - 40 PB/year
  - 4 PB/year after reduction
- Total data-volume
  - 27 PB/year (including calibrations, reduced data and all replicas)

Cumulative generated data (PB)

Legend: ■ Cumulative Volume South site   ■ Cumulative Volume North site
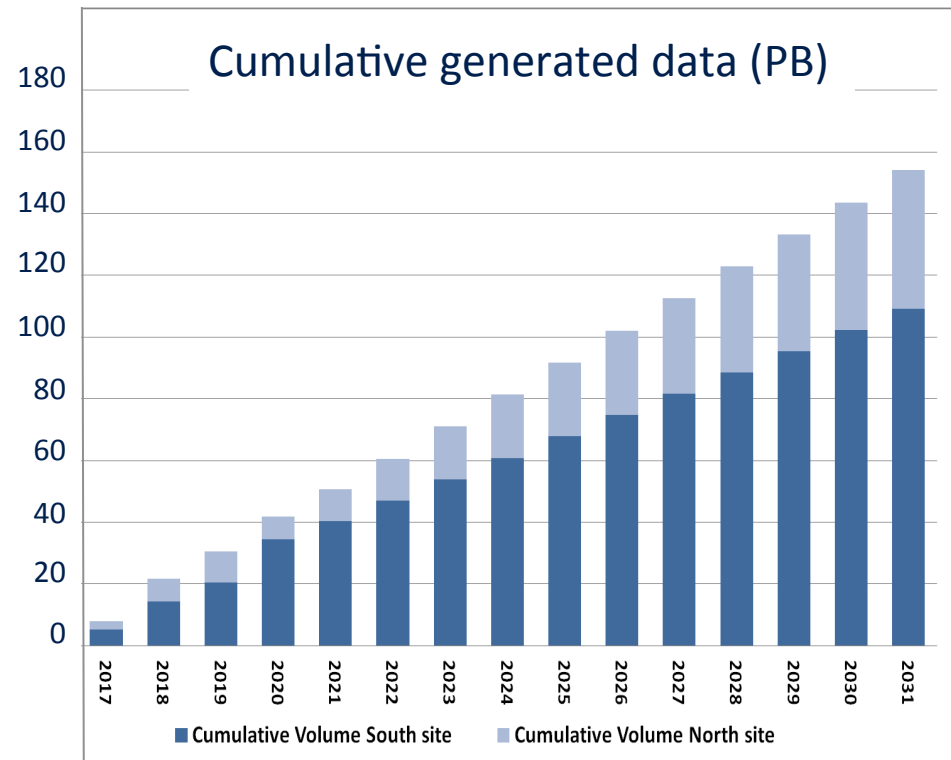
Current computing model for MC simulations

Grid sites supporting CTA Virtual Organization

- Use EGI grid resources (CTA Virtual Organization)
  - ~ 20 sites in Europe
  - 6 sites provide in total 4 PB
- MC production jobs run at all sites
  - Output data are stored at 6 SEs (1 distributed replica)
- MC analysis jobs run at sites with good connectivity to SEs
- Users jobs also running in parallel

Production setup

- Using DIRAC for Workload and Data Management
  - Workload Management System
    - Optimization of the computing resources usage by an efficient scheduling of the computing tasks
    - Resource integration (CREAM, ARC, Cloud, HPC, standalone clusters)
  - Transformation System
    - Running multi-steps data-processing workflows in a automatized way
  - Data Management System
    - All data operations (transfers, replication, removal, etc.-
    - File Catalog: replica and metadata
- CVMFS used to easily manage sw installation and access by distributed jobs
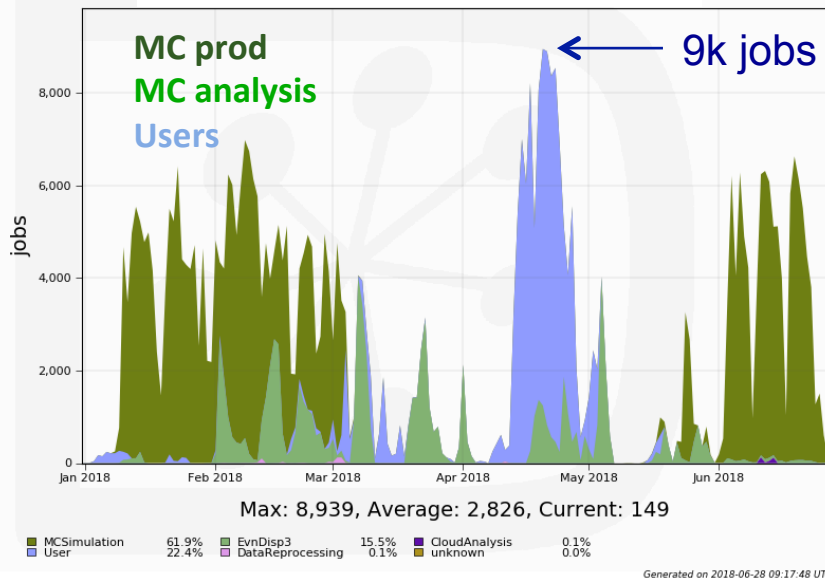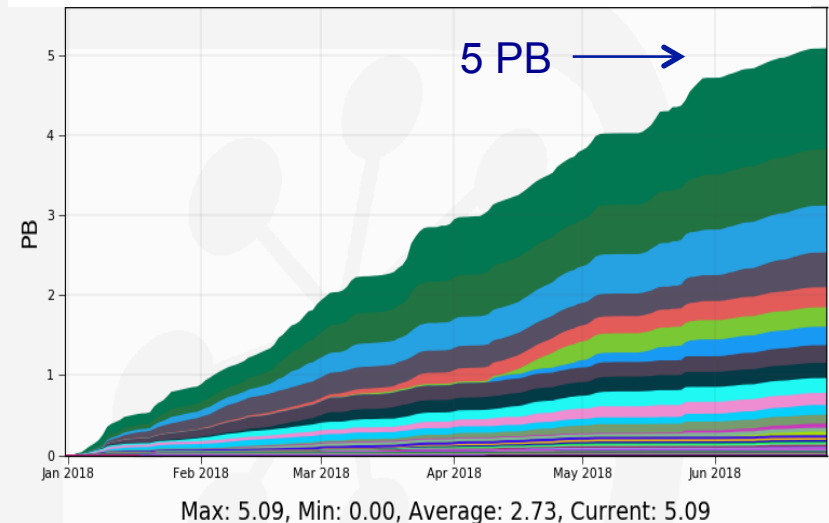
Monte Carlo campaigns during past years

- MC Campaigns to assess CTA design since 2013
  - CTA site selection, telescope layout, Instrument Response Functions
- Resources used in 2018
  - 130 M HS06 CPU hours
  - 5 PB transferred data (3.4 PB currently on disk/tape)
  - 28 M files registered in the catalog

Running jobs in 2018

MC prod
MC analysis
Users

9k jobs

Max: 8,939, Average: 2,826, Current: 149

MCSimulation	61.9%	EvnDisp3	15.5%	CloudAnalysis	0.1%
User	22.4%	DataReprocessing	0.1%	unknown	0.0%

Transferred data by destination in 2018

5 PB

Max: 5.09, Min: 0.00, Average: 2.73, Current: 5.09

Generated on 2018-06-28 09:17:48 UTC
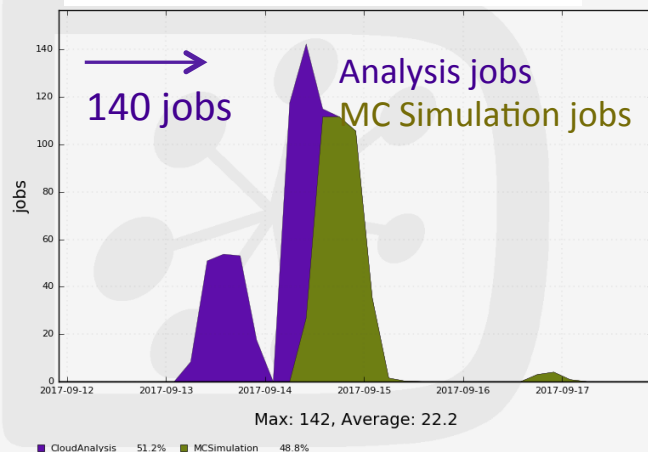
Tests on cloud resources integration

- Since last year, testing the integration of cloud resources in the CTA production system
- Using VMDIRAC module for transparent integration
  - Clouds are just seen as additional sites
  - Jobs behave as standard DIRAC grid jobs
- Cloud resources used
  - Private: 2 sites of the France Grilles Federated Cloud (CC-IN2P3, LUPM)
  - Commercial: 3 companies in the context of the HNSciCloud project
- Functional tests in 2017
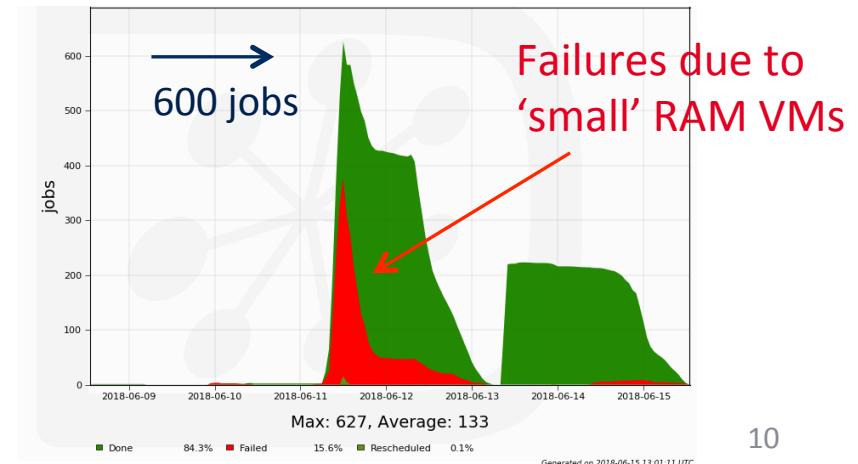- First scalability tests in 2018
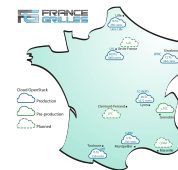
Tests on cloud resources integration

- Running standard CTA jobs
  - MC simulation
  - Analysis jobs accessing Input Data at remote grid SEs
  - 1 job/VCPU -> 1 VM with 4 VCPUs takes 4 jobs
- Functional tests at LUPM and CC-IN2P3 (OpenStack) ✓ **Successfull**
  - Up to 50 VMs available at each site
- First scalability tests on 1 commercial cloud (OpenStack) ✓ **Successfull**
  - Up to 250 VMs with flavour: 4 VCPUs, 8 GB RAM
    - Failures due to 'small' RAM VMs
  - Up to 60 VMs but with larger RAM (4 VCPUs, 32 GB RAM)
  - VM size not completely customizable
- No problem accessing remote Input Data
  - CPU efficiency > 90%

Jobs on the LUPM Cloud

140 jobs

Analysis jobs
MC Simulation jobs

Max: 142, Average: 22.2

CloudAnalysis 51.2%   MCSimulation 48.8%

Generated on 2017-09-23 05:58:47 UTC

Analysis jobs on a commercial cloud

600 jobs

Failures due to 'small' RAM VMs

Max: 627, Average: 133

Done 84.3%   Failed 15.6%   Rescheduled 0.1%

Generated on 2018-06-15 13:01:11 UTC
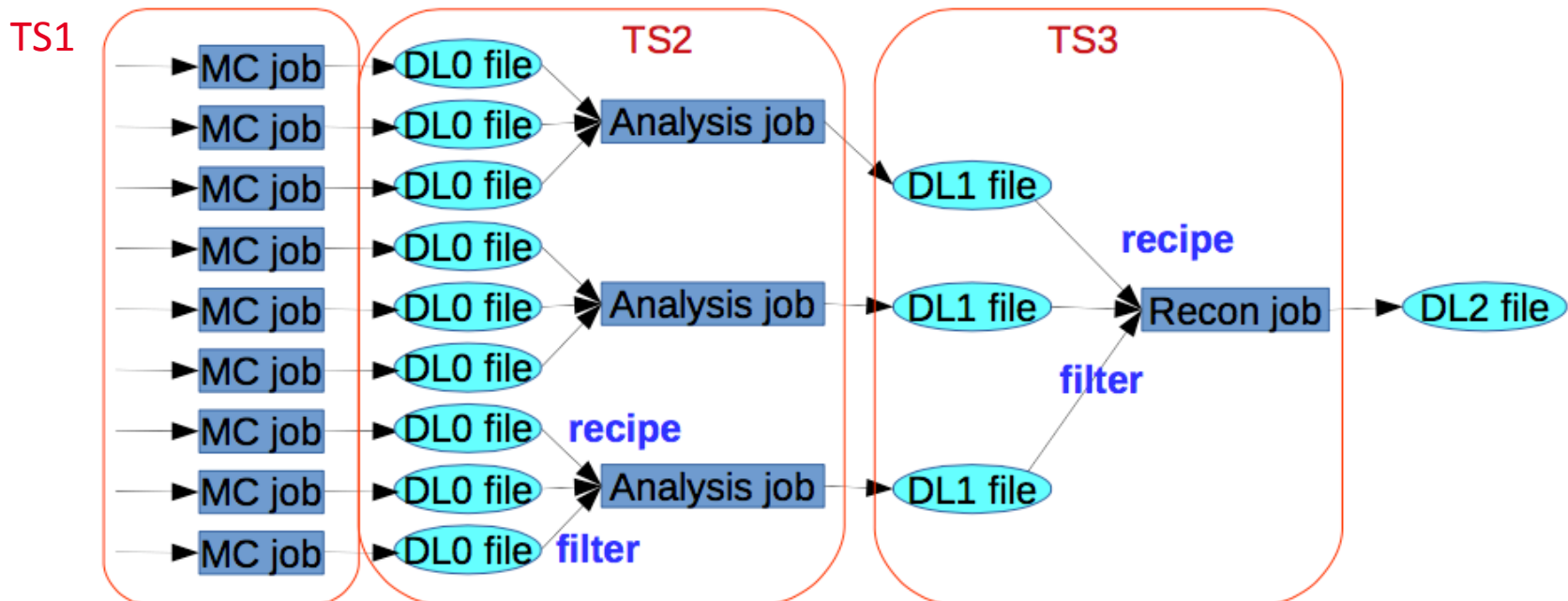
Data driven Workflow Management

- CTA workflow management heavily rely on the DIRAC Transformation System
    - Automated Tasks, workhorse for MC production and analysis
    - A **Transformation** is an input *data filter + a recipe* to create jobs
    - Fully data-driven: jobs are created as soon as data with required properties are registered into the file catalog

Data driven Workflow Management

- Transformations manually created and monitored one by one
  - Need further automatization
  - Development started of a new light high-level system (called Production System)
- Production System architecture
  - As any other DIRAC System (DB, Service, Agents, Client)

Data driven Workflow Management
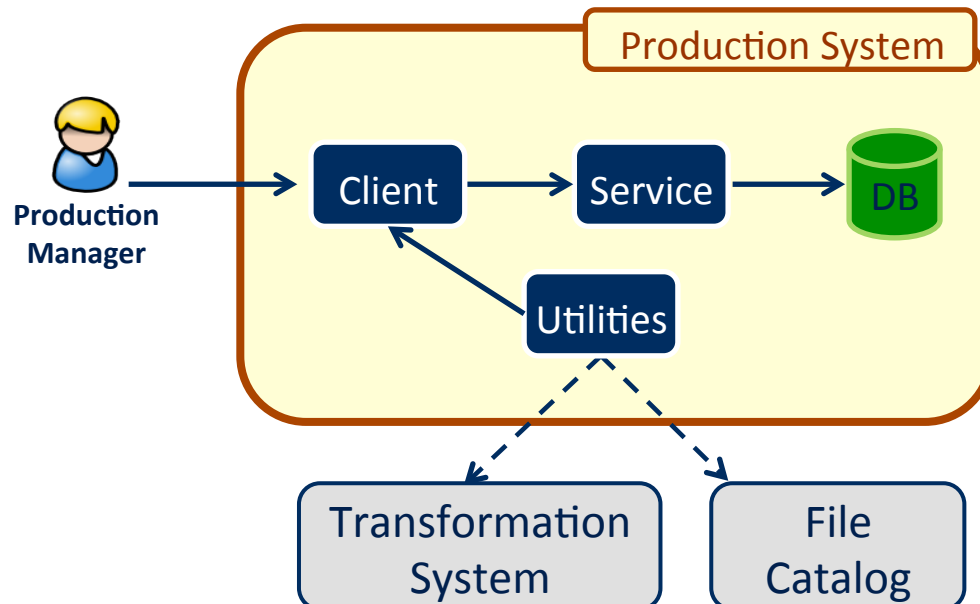
- Production System
  - Production: set of 'linked transformations'
    - *E.g.* 2 transformations t1, t2 are 'linked' if:
      - InputQuery2 logically intersects OutputQuery1
    - Enhancement of the transformation definition to characterise the inputs/outputs of a transformation through meta-queries
  - Automatic transformation instantiation based on the production definition
  - Support multiple workflows schemes (sequential, split, merge)
  - Prototype implementation done early 2018
  - Aim to be a general system used by other communities
    - On-going discussion about design choices in an open forum

Conclusions

- A production setup, based on DIRAC, used since 7 years for Monte Carlo simulations and analysis

- Cloud resource integration successfully tested

- Most likely HPC and GPUs resources will be integrated in future (as use-cases appear)

- Prototype implementation of a high-level workflow management system

- We will have to ensure the integration of the production setup with the other CTA systems (*e.g.* Archive System)