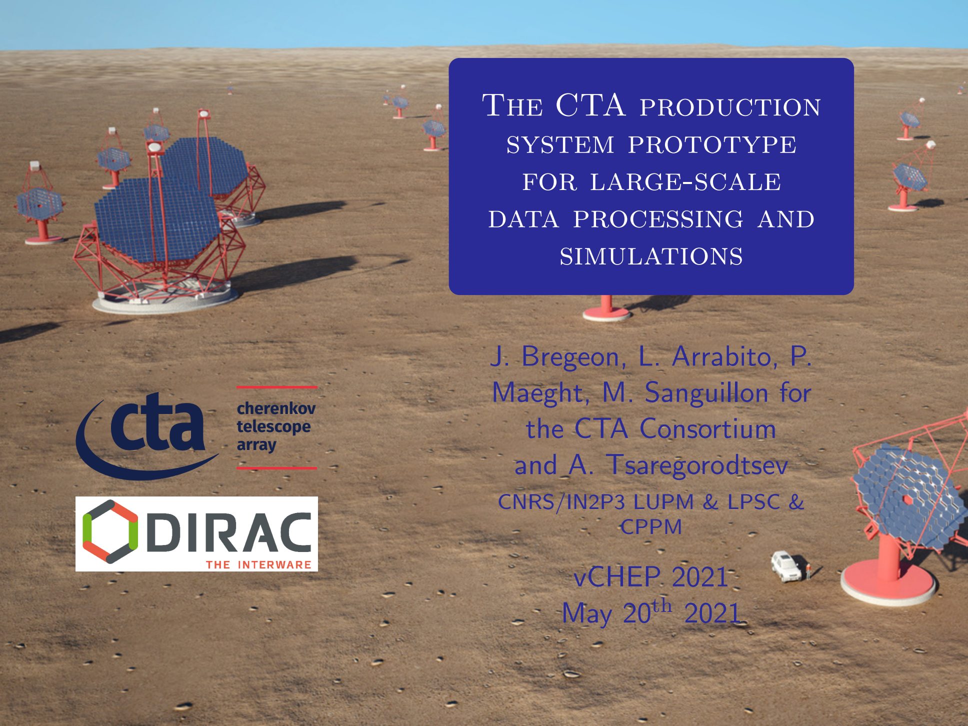


The CTA production system prototype for large-scale data processing and simulations

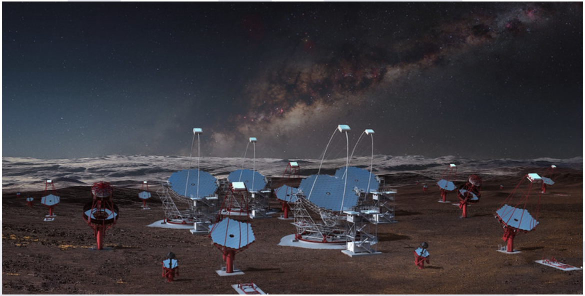
J. Bregeon, L. Arrabito, P. Maeght, M. Sanguillon for the CTA Consortium and A. Tsaregorodtsev

CNRS/IN2P3 LUPM & LPSC & CPPM

vCHEP 2021
May 20th 2021

The CTA Observatory

- The next generation instrument in VHE gamma-ray astronomy (1500 participants in 31 countries)
  - Cosmic–ray origins, High Energy astrophysical phenomena, fundamental physics and cosmology

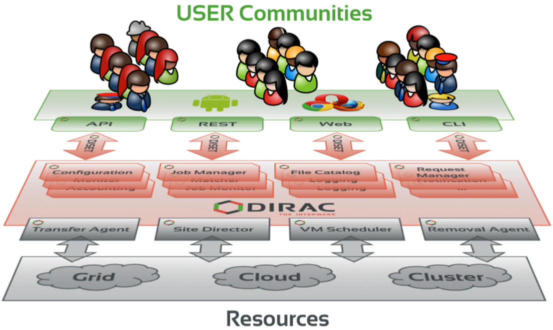


- Two arrays of Cherenkov telescopes
  - North site (La Palma, Spain): O(10) LSTs + MSTs
  - South site (Paranal, Chile): O(50) LSTs + MSTs + SSTs
- Construction starting now, operations for ∼30 years

http://www.cta-observatory.org

DIRAC INTERWARE

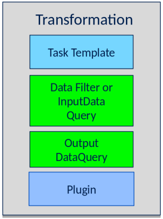
- **Distributed Infrastructure with Remote Agent Control**
  - all–in–one solution to access distributed resources (computing and storage) → *http://diracgrid.org*
  - started by LHCb more than 15 years ago
  - used today by many experiments: LHCb, CLIC, Belle II, CTA...
  - open source software maintained and developed by an international consortium → *http://github.com/dIRACGrid*
  - EGI instance with hundreds of users from many "small" VOs



DIRAC for CTA

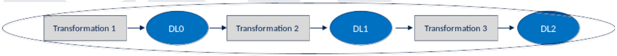
- DIRAC instance dedicated to CTA distributed at 3 sites (CC-IN2P3, PIC, DESY)
- DIRAC v7r1 deployed on 5 core servers
    + extended with `CTADIRAC` plugin (*code on CTAO GitLab*)
    - 1 running WMS services (32 cores, 32 GB RAM)
    - 1 running WMS agents and executors (32 cores, 32 GB RAM)
    - 1 running TS and RMS (16 cores, 8GB RAM)
    - 1 running DMS + 1 DIRAC SE (16 cores, 8GB RAM, 2 TB of disk for the SE)
    - 1 running duplicated DMS, TS, RMS services (8 cores, 32 GB RAM)
    + a VM for the web server
- Services
    - MySql databases at CC-IN2P3 (File Catalog, Transformation DB) and PIC (Accounting, Jobs DB and more...)
    - ELK instance (ElasticSearch++) at CC-IN2P3 for the DIRAC Monitoring system (new for us)
    - CVMFS for software distribution (CC-IN2P3 and DESY)
    - FTS for data movement (CERN instance)

Data driven production system



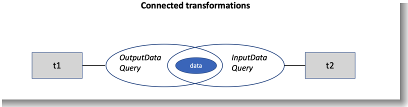
Transformation System

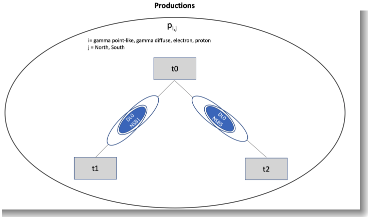
- ▶ transformation: task template, input and output data
- ▶ 2 transformations "connected" if output data of T1 intersect input data of T2
- ▶ workflow is data driven!



Production System — introduced in version v7r0

- ▶ production as a set of "linked" transformation"
- ▶ design to handle the workflow at high level
- ▶ meant to be used by "operators" in *production*





Production system in production

Prepare transformations

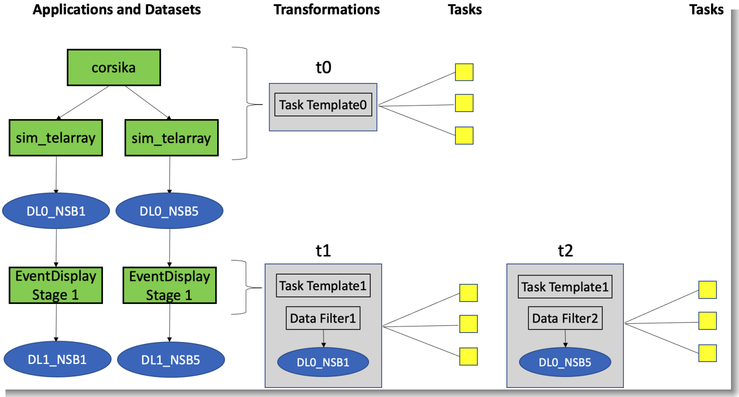
- setup workload into DIRAC Job API
- setup Job to be used through transformation

Build Productions

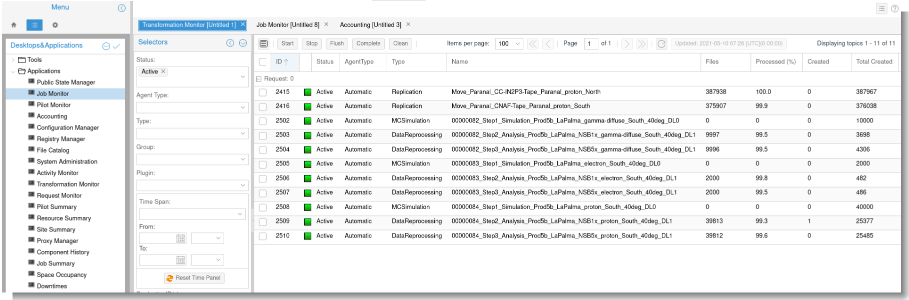
- define input and output datasets, as meta-queries
- assemble Productions from Transformation definition and datasets

Run Productions

- assign simulation jobs to production
- activate production and monitor workflow processing



"Operator" Interface



Request: 0

ID		Status	AgentType	Type	Name	Files	Processed (%)	Created	Total Created
2415	🟩	Active	Automatic	Replication	Move_Paranal_CC-IN2P3-Tape_Paranal_proton_North	387938	100.0	0	387967
2416	🟩	Active	Automatic	Replication	Move_Paranal_CNAF-Tape_Paranal_proton_South	375907	99.9	0	376038
2502	🟩	Active	Automatic	MCSimulation	00000082_Step1_Simulation_Prod5b_LaPalma_gamma-diffuse_South_40deg_DL0	0	0	0	10000
2503	🟩	Active	Automatic	DataReprocessing	00000082_Step2_Analysis_Prod5b_LaPalma_NSB1x_gamma-diffuse_South_40deg_DL1	9997	99.5	0	3698
2504	🟩	Active	Automatic	DataReprocessing	00000082_Step3_Analysis_Prod5b_LaPalma_NSB5x_gamma-diffuse_South_40deg_DL1	9996	99.5	0	4306
2505	🟩	Active	Automatic	MCSimulation	00000083_Step1_Simulation_Prod5b_LaPalma_electron_South_40deg_DL0	0	0	0	2000
2506	🟩	Active	Automatic	DataReprocessing	00000083_Step2_Analysis_Prod5b_LaPalma_NSB1x_electron_South_40deg_DL1	2000	99.8	0	482
2507	🟩	Active	Automatic	DataReprocessing	00000083_Step3_Analysis_Prod5b_LaPalma_NSB5x_electron_South_40deg_DL1	2000	99.5	0	486
2508	🟩	Active	Automatic	MCSimulation	00000084_Step1_Simulation_Prod5b_LaPalma_proton_South_40deg_DL0	0	0	0	40000
2509	🟩	Active	Automatic	DataReprocessing	00000084_Step2_Analysis_Prod5b_LaPalma_NSB1x_proton_South_40deg_DL1	39813	99.3	1	25377
2510	🟩	Active	Automatic	DataReprocessing	00000084_Step3_Analysis_Prod5b_LaPalma_NSB5x_proton_South_40deg_DL1	39812	99.6	0	25485

Run Productions

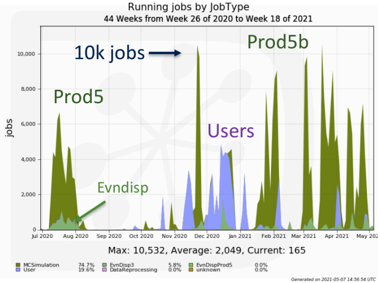
- monitoring mostly through web interface
- failed processing jobs are automatically retried (FailOverRequest module)
- stop/resume productions in case of issues (scripts)
- monitor datasets content (scripts)

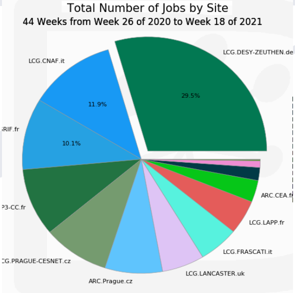
Tools examples

- in vanilla DIRAC
  - `dirac-prod-get-all`
  - `dirac-prod-start / stop`
- in CTADIRAC extension
  - `cta-prod-extend-sim`
  - `cta-prod-monitor`
  - `cta-prod-create-dataset`
  - `cta-prod-show-dataset`

Latest large scale production

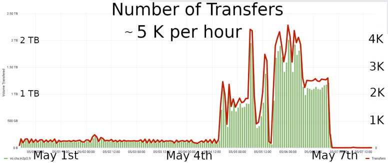
- PROD5(b) goals for CTA
  - improved telescope and camera descriptions
  - fine tuning of telescope positions
- A few numbers
  - ran "almost" non stop for 9 months
  - 200 MHS06.hours (20% by users) in 2.8 millions of jobs
  - 15 different sites across Europe
  - 1.7 PB on disk in 4.7 millions of files

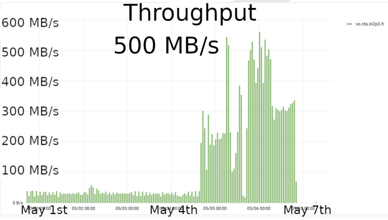




Data Management

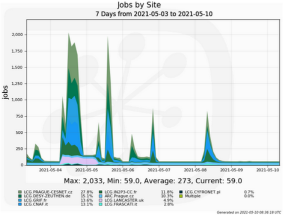
- CTA now handles 6 PB on disk and 2 PB on tape
  - use DIRAC for data management operations
  - data removal
  - moving data from disk to tapes
- DIRAC framework
  - Dirac File Catalog with file and replica meta data
  - Request Management System (asynchronous tasks)
  - Transformation System (combine with Production System!)
- Example: Move (1 PB) of `PROD3` to tape
  - recently setup CERN FTS to run under the hood
  - one Transformation per dataset — meta–data driven!

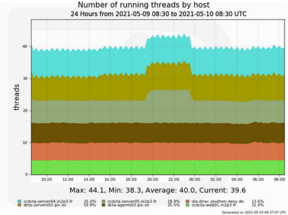


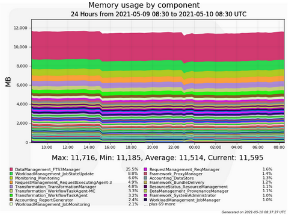


DIRAC Monitoring System

- ELK stack: Elasticsearch, Logstash and Kibana
  - CTA relies on the CC-IN2P3 instance
- need proper configuration of the DIRAC instance (and some work to support latest ELK version)
- gives access to a lot of information in a very efficient way
  - a few examples below
  - now need to dig more and extract specific information for our needs
  - in particular job parameters like memory usage

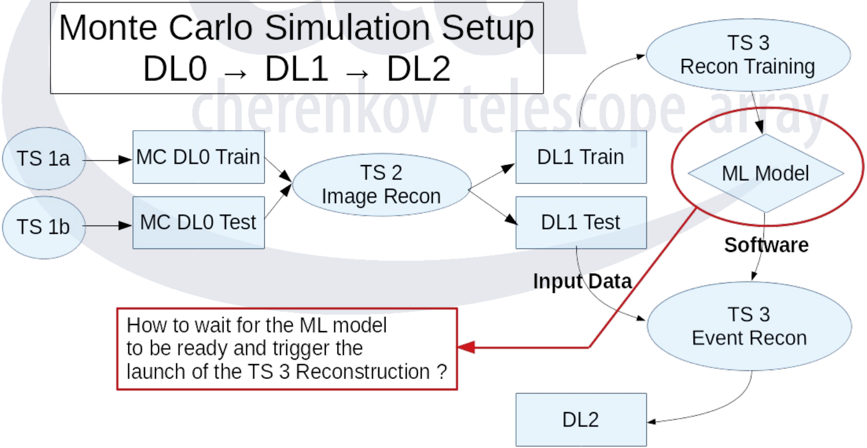






More complex workflows?

- How do we fully automatize this kind of workflow?
  - issue: the look-up table/ML model/BDT are “part” of the software, how do we know that these are ready and that the next step can be run ?
  - → different solutions envisaged, need to try!



Monte Carlo Simulation Setup
DL0 → DL1 → DL2

TS 1a

TS 1b

MC DL0 Train

MC DL0 Test

TS 2
Image Recon

DL1 Train

DL1 Test

TS 3
Recon Training

ML Model

Software

Input Data

TS 3
Event Recon

DL2

How to wait for the ML model to be ready and trigger the launch of the TS 3 Reconstruction ?

Conclusion

- DIRAC is a modular open source tool for distributed computing, driven by an open minded consortium that offers a place for help and discussions.
- CTADIRAC is the DIRAC instance developed for the construction of CTA, it's been up and running for more than 8 years: hundreds of millions of jobs, Petabytes of data.
- CTADIRAC also largely benefits from HEP computing ecosystem: VOMS, CVMFS, FTS...
- CTADIRAC is mostly ready to become the CTA computing resources and workflow management system: in–kind contributions to the CTA Observatory being discussed now, hopefully finalized by end of the year.
- Learn more about CTA performance and see the full list of computing centers that provide resources and support (Thanks to them!) at: *https://www.cta-observatory.org/science/cta-performance/*.