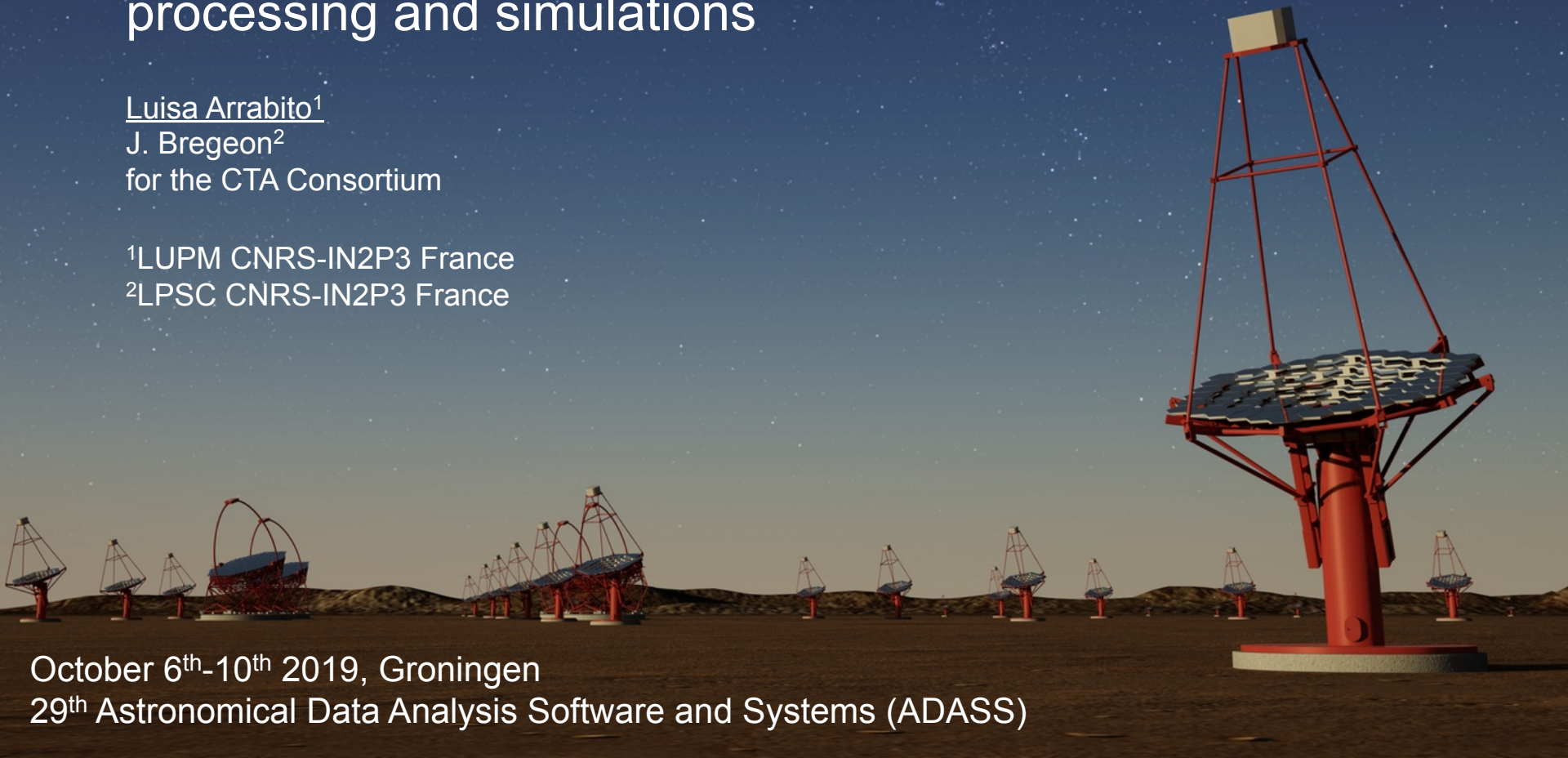
cherenkov telescope array

A DIRAC-based prototype for the Cherenkov Telescope Array data management, processing and simulations

Luisa Arrabito[1]
J. Bregeon[2]
for the CTA Consortium

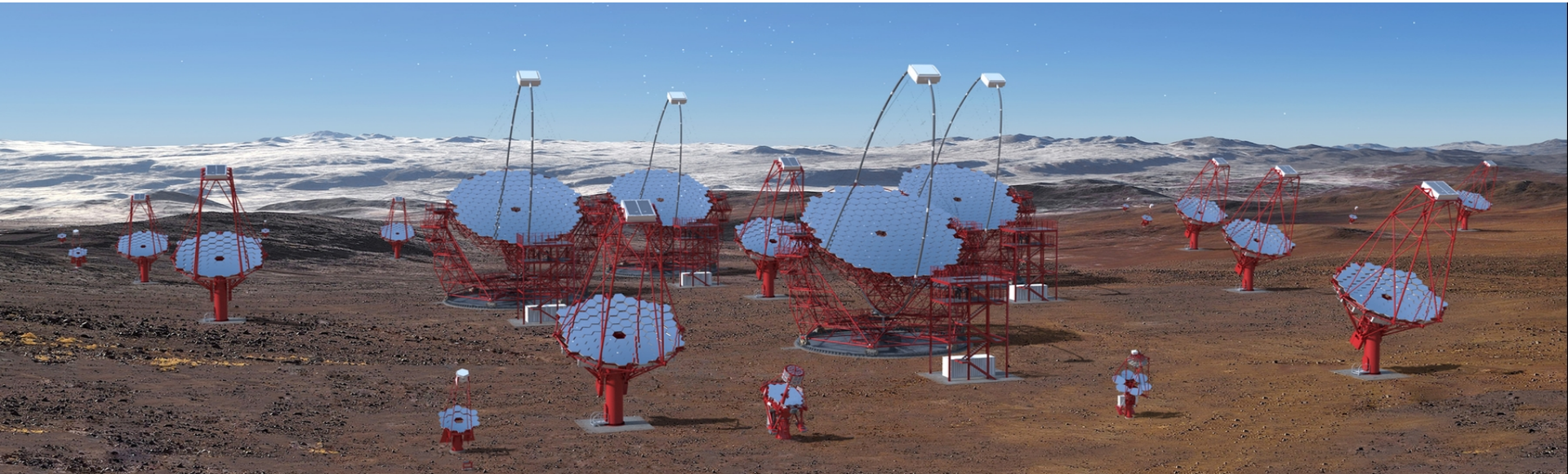
[1]LUPM CNRS-IN2P3 France
[2]LPSC CNRS-IN2P3 France



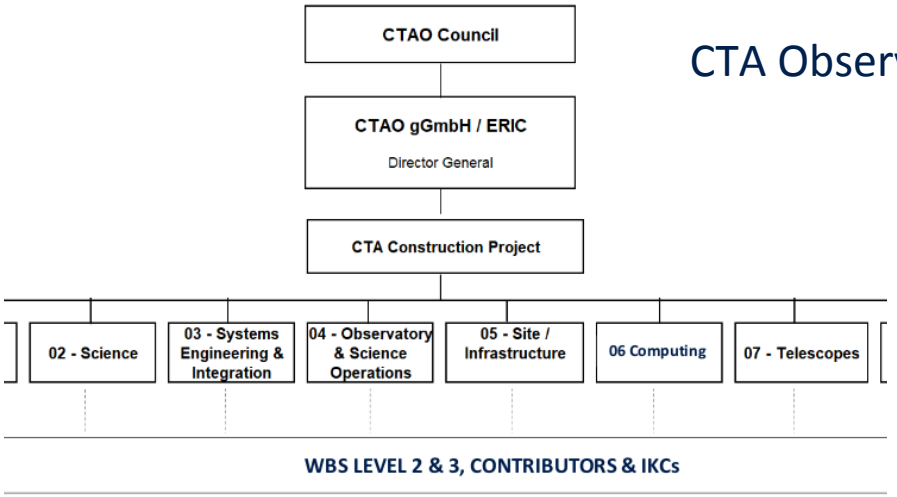
October 6th-10th 2019, Groningen
29th Astronomical Data Analysis Software and Systems (ADASS)

CTA (Cherenkov Telescope Array)

- Next generation IACT, VHE gamma-ray 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-sized, 15 medium-sized telescopes
  - Southern site (Paranal, Chile): 4 large- sized, 25 medium-sized, 70 small-sized telescopes
- Project schedule
  - Construction and deployment: 2017-2025
  - Science operations: from 2022, for ~30 years

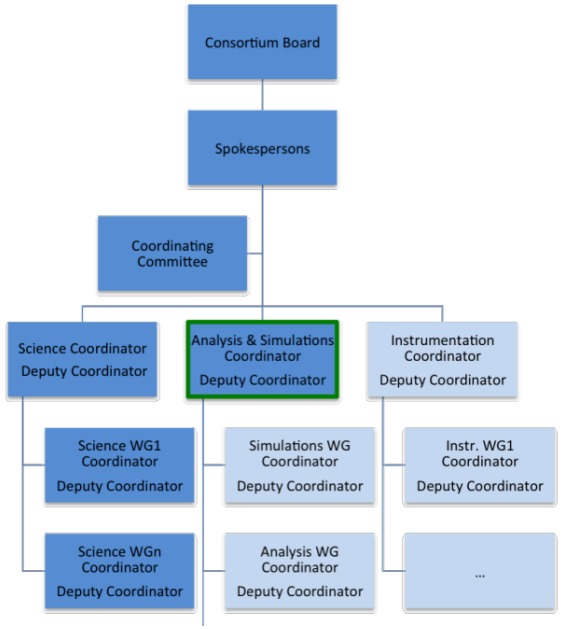


CTA Observatory and Consortium



CTA Observatory

CTA Consortium

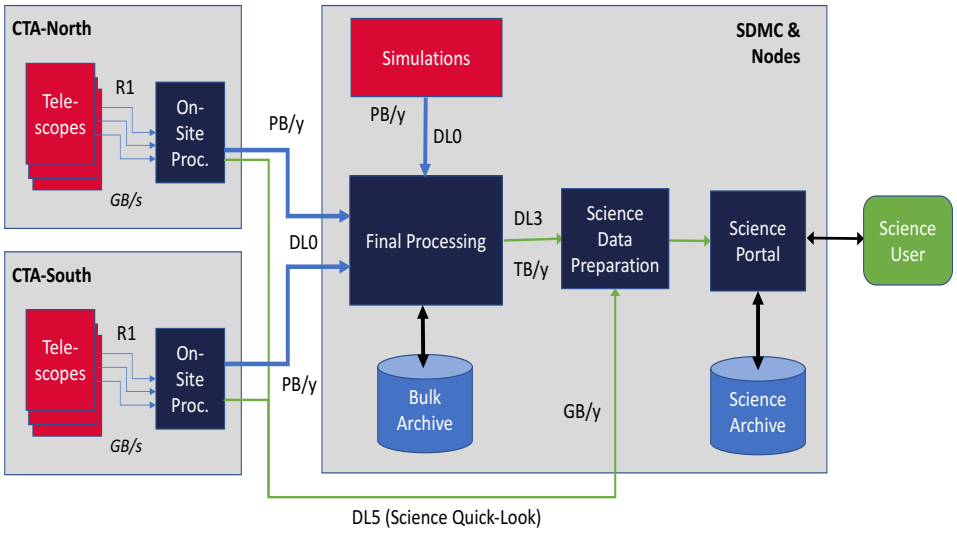


- CTA Observatory (CTAO) is the legal entity responsible for the construction, operation, maintenance and upgrade of the observatory
- CTA Consortium (CTAC) member institutes make in-kind contributions to CTA construction
- CTA Software/Computing
  - CTAO Computing Department
    - Architecture, Design, Specification
    - Coordination of implementation via IKC, contracts with companies, in-house implementation
  - CTAC
    - From prototype development to IKC

CTA Computing Challenges

- CTA has data management challenges with large-scale data processing and simulation needs
  - + 27 PB/year
  - 1200 – 9000 cores/year on average after 15 years of operations

Data Flow



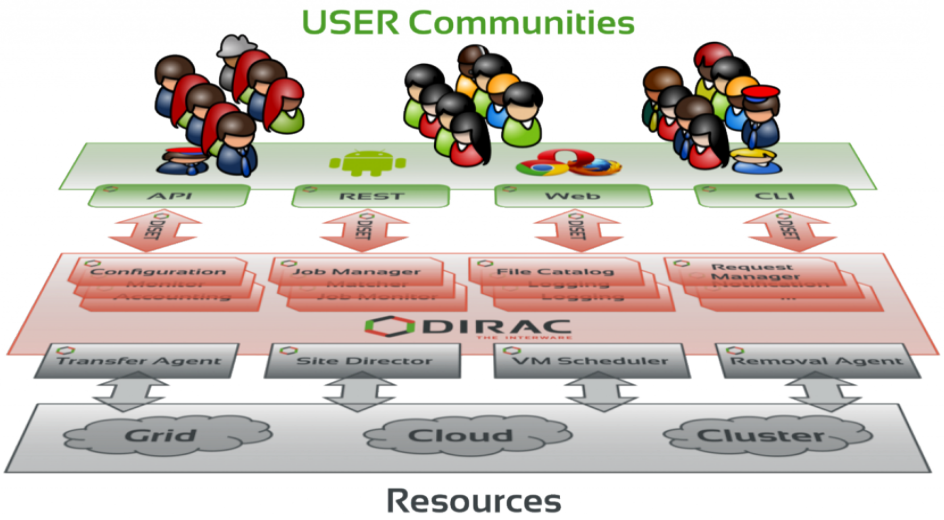
- On-site computing
  - Near real-time processing
  - Next-day data processing for quick-look and science alerts
  - On-site buffer and data transfer
- Off-site computing
  - Simulations and final processing
  - Bulk Archive ~ PB/year
  - Science data preparation
  - Science Archive ~ GB/year
  - Open access through Science Portal

CTA and DIRAC

- CTA aims for efficient fully automatic processing
  - Raw data to be processed within 1 month
  - 1 full re-processing per year
  - Regular simulations to calculate the Instrument Response Functions
- Distributed computing model for Off-site computing
  - Baseline with 4 first-class Data Centers
- DIRAC is a software framework to manage data and workload on a distributed infrastructure
  - Developed at CERN by LHCb collaboration to use WLCG resources
  - Then generalized to be used by several other experiments/projects (high energy physics, astronomy, life science, etc.)
  - Proposed for CTA Final Processing and Simulations

DIRAC Introduction

- An open source software framework for distributed computing
  - https://github.com/DIRACGrid/DIRAC
- A layer between users and resources of different kinds
- Experiment agnostic, extensible, flexible
- Current users communities
  - LHCb, ILC, Belle II, T2K, CTA, Pierre Auger Observatory, Eiscat 3D, BioMed, We-nmr etc.
- Other experiments/observatories are interested (SKA, Virgo, ...)
- EGI Core Service (DIRAC4EGI): https://dirac.egi.eu/DIRAC
- Proposed as common tool within the ESCAPE H2020 project

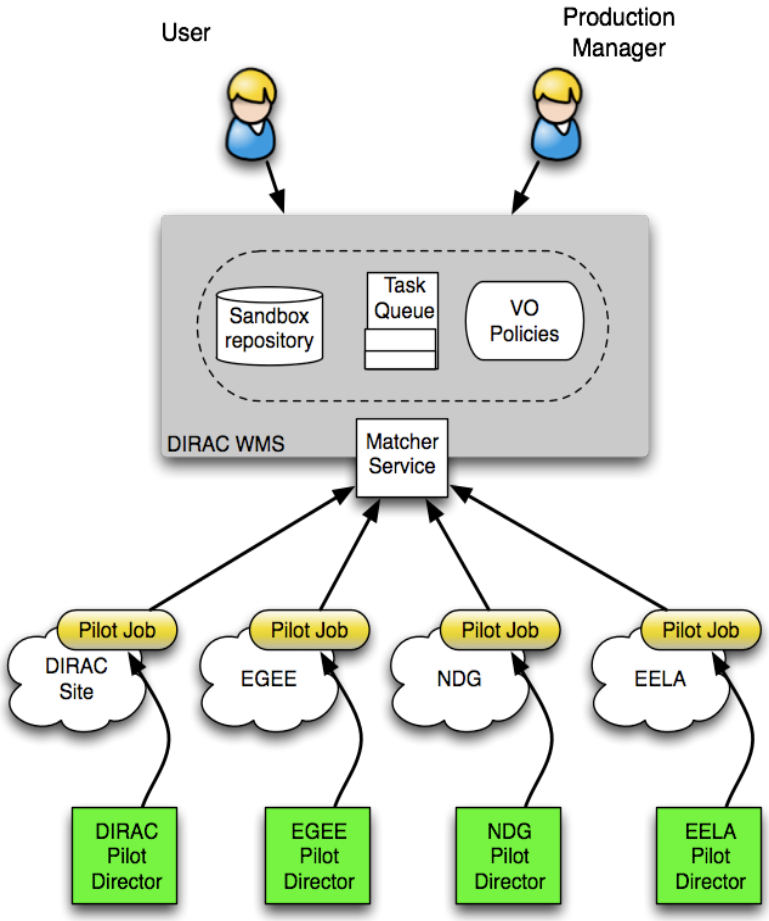


DIRAC Software Architecture

- Based on Service Oriented Architecture and composed of several Systems interacting together through Client/Service communications
- Each System composed of:
  - Services, Agents, DBs, Client interface (CLI, API, REST, web)
- Main DIRAC Systems:
  - Workload Management System (WMS)
  - Data Management System (DMS)
  - Transformation System (TS)
  - Request Management (RMS)
  - Resource Status System (RSS)
  - Accounting
  - ...

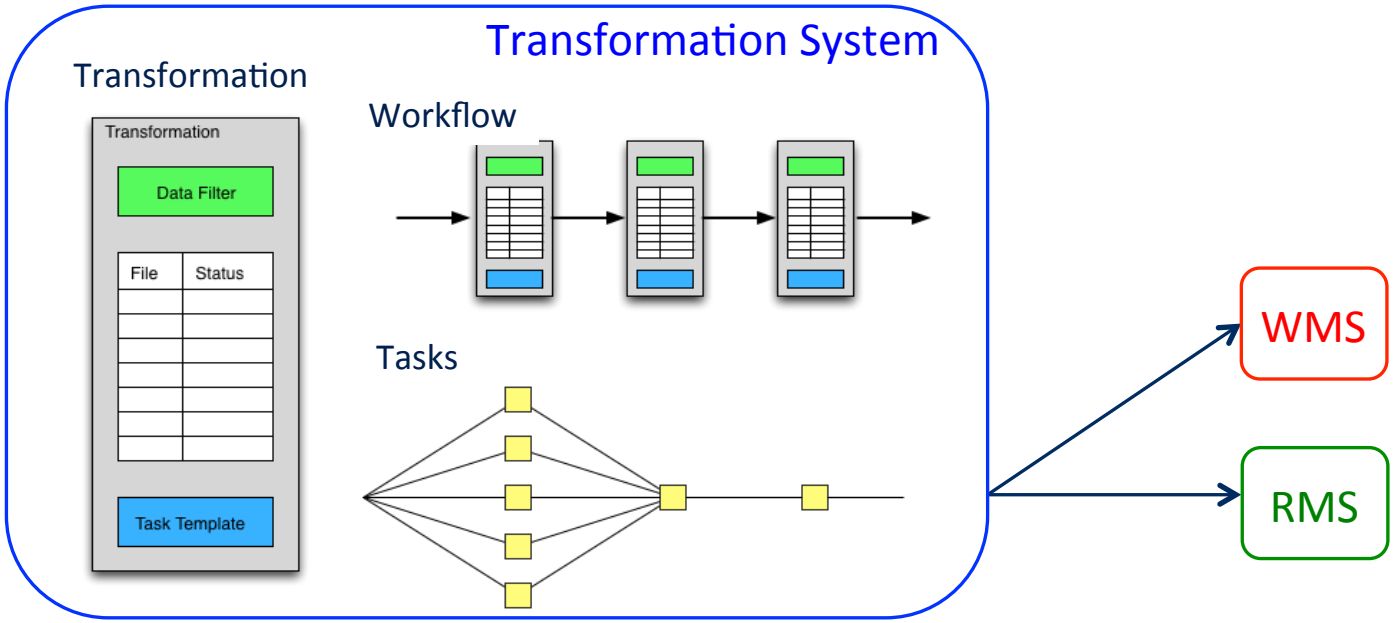
DIRAC Workload Management System

- Much more than an interface for ‘job submission’
- **Implements the pilot mechanism**
  - Pull scheduling paradigm and standard configurable environment for jobs
- Allows combination of heterogeneous resources in a transparent way
  - Grids, clouds, standalone clusters, etc.
- Management of users’ priorities
- Allows very detailed job monitoring
- Users submit jobs directly to the WMS or through the Transformation System (used for large productions, next slide)



DIRAC Transformation System for Data-Driven Workflows

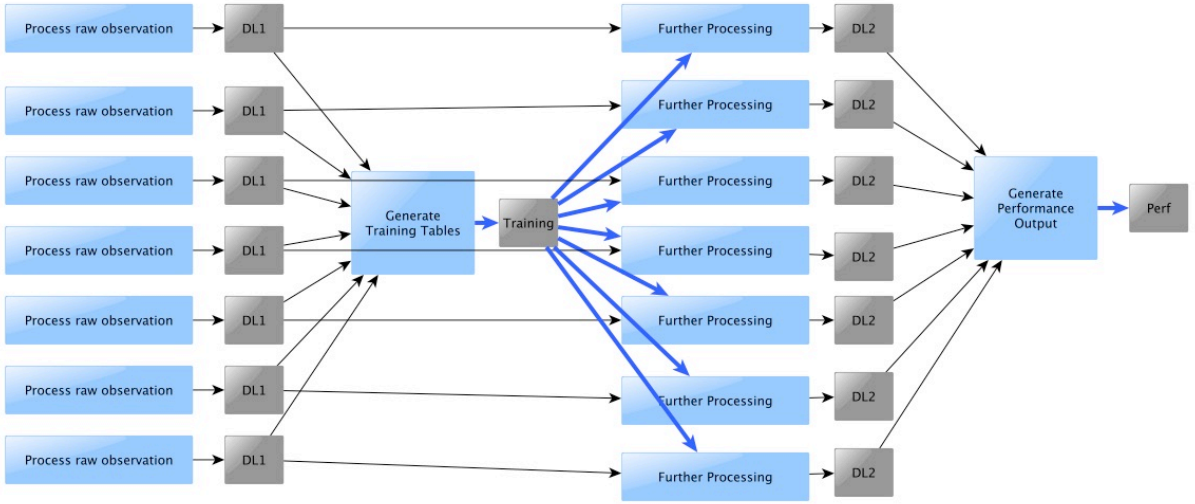
- Data-driven workflows as chains of data transformations
- Transformation: input data filter + recipe to create tasks
  - Tasks are created as soon as data with required properties are registered into the file catalog
- Tasks:
  - Jobs submitted to the WMS
  - Data operation requests submitted to the RMS (for bulk data operations)



DIRAC Production System for data driven workflows

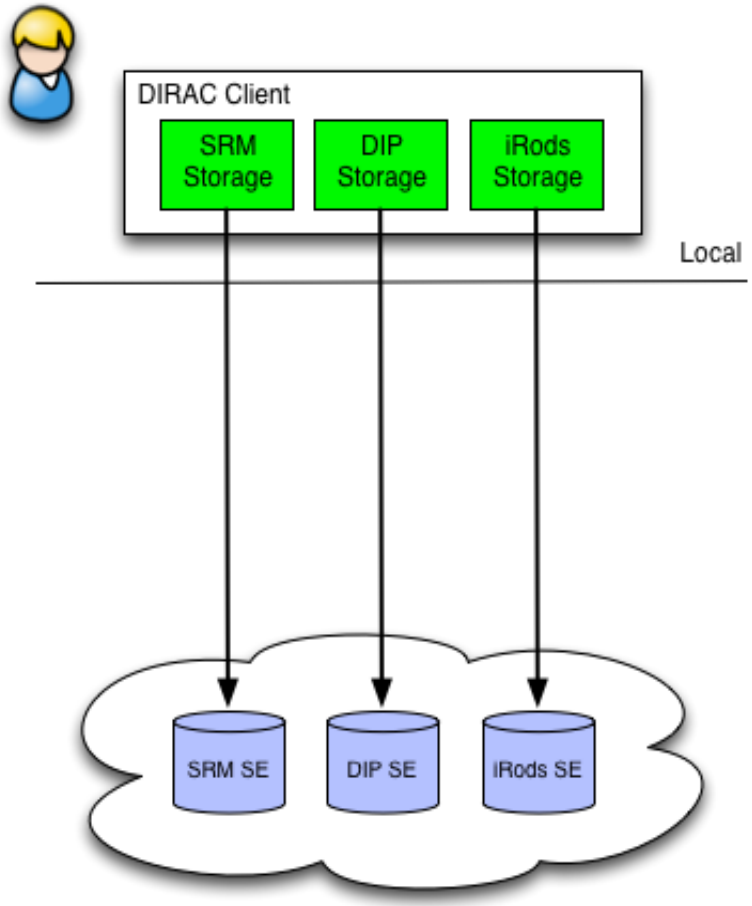
- Transformations can be combined together to form workflows of arbitrary complexity
- This is achieved by the DIRAC Production System (CTA contribution to DIRAC core)
  - Developed on the top of the Transformation System
  - Allows full automatization of multi-step workflow execution

- CTA workflow example
  - Split/merge, machine learning model production, etc.



DIRAC Data Management (DMS)

- Storage element abstraction with a client implementation for each access protocol
  - DIPS – DIRAC data transfer protocol
  - FTP, HTTP, WebDAV
  - HEP centers specific protocols (SRM, XROOTD, RFIO, DCAP, etc.)
  - Cloud specific data access protocols (S3, Swift, CDMI)
- DIRAC is dealing with large volumes of scientific data
  - 10's of Petabytes
  - 10^7-10^8 of files and directories
- Massive data operations supported
  - Asynchronous execution
  - Automatic failure recovery
  - Data integrity checking
  - Automated data driven workflows



DIRAC File Catalog (DFC)

- DFC as Replica Catalog
  - A service to keep track of all the physical file replicas in all storage elements
  - Defines a single logical name space for all the managed data
  - Organizes files hierarchically like in common file systems
- DFC as Metadata Catalog
  - Support for user-defined metadata as key-value pairs (*e.g.* simulation conditions, provenance data, etc.). Example from CTA:
    - primary = {gamma, proton, electron, ...}
    - zenith = {20, 40, 60, ...}
    - air_shower_sim_prog = corsika
    - ...
  - Allow for efficient searches. Example from CTA:

    find air_shower_sim_prog = corsika air_shower_sim_prog_version = 7.0 primary=gamma zenith=40 site=Paranal
  - Support of Datasets
    - Alias to a given query
    - Useful for frequent queries, *e.g.:*

      cta-dump-dataset Prod4_Paranal_gamma_20deg_North_dl0
- Scalability
  - e.g. 30 million of replicas in CTA DFC works fine
  - Confirmed by dedicated performance measurements

CTA-DIRAC prototype

- CTA-DIRAC servers
  - 5 servers at CC-IN2P3, PIC, DESY-ZN
    - Running all DIRAC Systems just described and more
  - MySQL servers hosting all DIRAC DBs
  - 1 web-server for DIRAC portal
- CTA-DIRAC software extension
  - Mainly extension of the DIRAC Job API to easily configure CTA jobs
  - Utility to setup CTA software environment for all jobs (supporting multiple software locations)
  - Utility to register files with custom meta-data and directory structure in the DFC
- CernVM-FS (CVMFS) to easily manage sw installation and access by distributed jobs
  - 1 stratum-0 at CC-IN2P3 and 2 stratum-1 at CC and DESY-ZN

Current computing model for MC simulations (CTA preparatory phase)

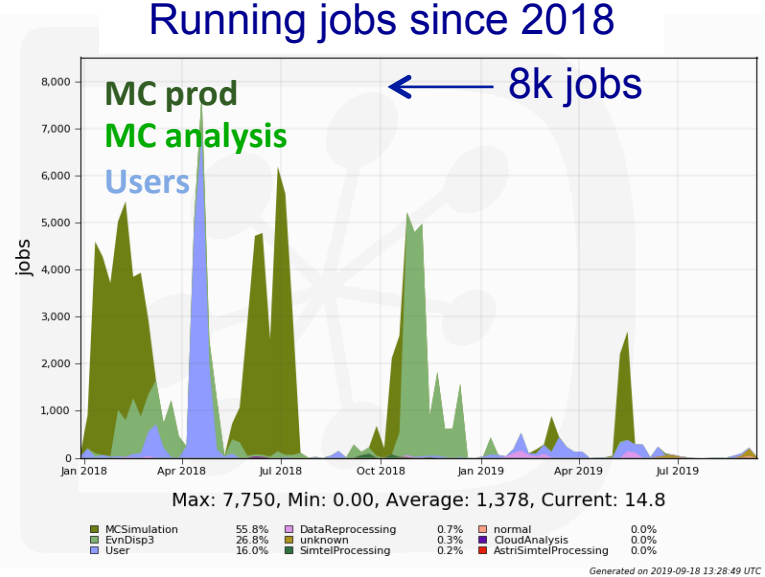
- Massive MC simulations during the CTA preparatory phase
  - CTA site selection, telescope layout, Instrument Response Functions
- Use CTA-DIRAC prototype to access EGI grid resources
  - About 15 sites for computing
  - 6 Storage Elements: ~ 6 PB in total
- Computing model used in this phase
  - MC production jobs run at all sites
    - Output data stored at 6 SEs
  - MC analysis jobs run at selected sites with good connectivity to SEs
  - Users jobs also running in parallel
- Future computing model will be distributed but not necessarily grid-based
  - Will have to ensure fast data-processing and no data loss

Grid sites supporting CTA Virtual Organization

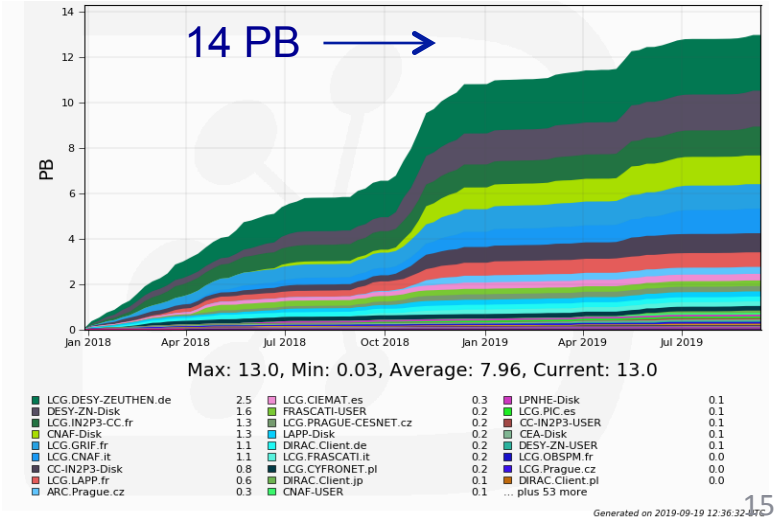
CTA-DIRAC exploitation for MC simulations (CTA preparatory phase)

- CTA-DIRAC exploitation (since 2013)
  - > 15 million executed jobs
  - 30 million of replicas in the DFC
  - All productions launched via the Transformation System
  - New resources integrated
    - Scalability tests with Clouds done in 2018
- Resource usage
  - 100-200 million HS06 CPU hours/year
  - 5-10 PB transferred data/year
  - 4.6 PB currently on disk/tape distributed over 6 sites

Running jobs since 2018



Transferred data by destination since 2018



Conclusions

- We have developed a DIRAC-based prototype to handle the massive MC simulations of CTA during its preparatory phase
  - Millions of jobs and CPU hours every year over 15 sites
  - Handling 10's of Petabytes, millions of files and directories
  - Automated workflows management
    - CTA contribution to DIRAC core (Production System)
- CTA-DIRAC proposed for the future data-processing and simulations of CTA
- Future work to adapt CTA-DIRAC for the operation phase (resource description, implementation of policies, interfaces etc.)
- Successful experience with DIRAC
  - It can certainly be useful for other communities