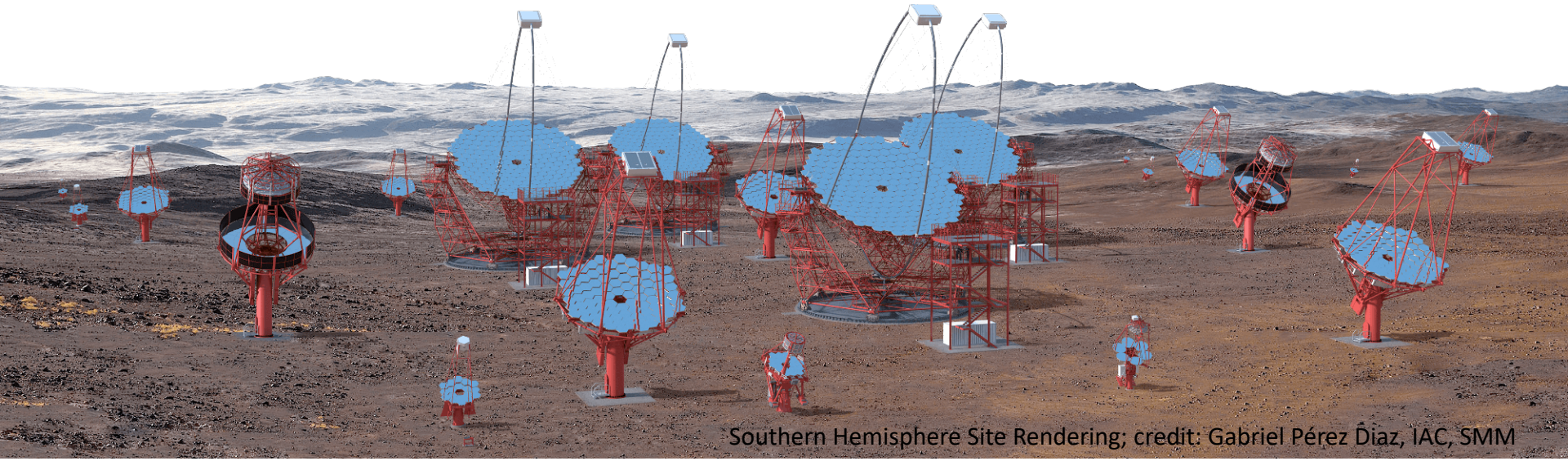
Pipeline execution and tracing with CTADIRAC



Southern Hemisphere Site Rendering; credit: Gabriel Pérez Diaz, IAC, SMM

Michèle Sanguillon[(1)], Luisa Arrabito[(1)], Johan Bregeon [(2)], Karl Kosack [(3)], Patrick Maeght[(1)]

[(1)] LUPM, Montpellier, France [(2)] LPSC, Grenoble, France [(3)] CEA, Paris, France

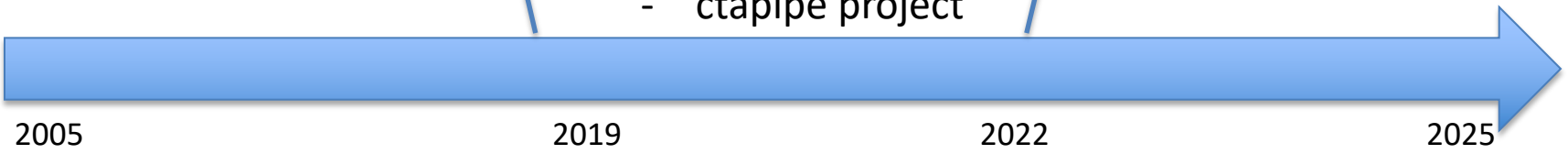
CONTEXT

- Next Generation Gamma-Ray Astronomy (succeeding H.E.S.S., MAGIC and VERITAS)
- Two arrays of 99 (South) and 19 (North) Cherenkov telescopes (4, 12 et 24 m in diametre)
- Observatory open to the Astronomy community
- Timeline:

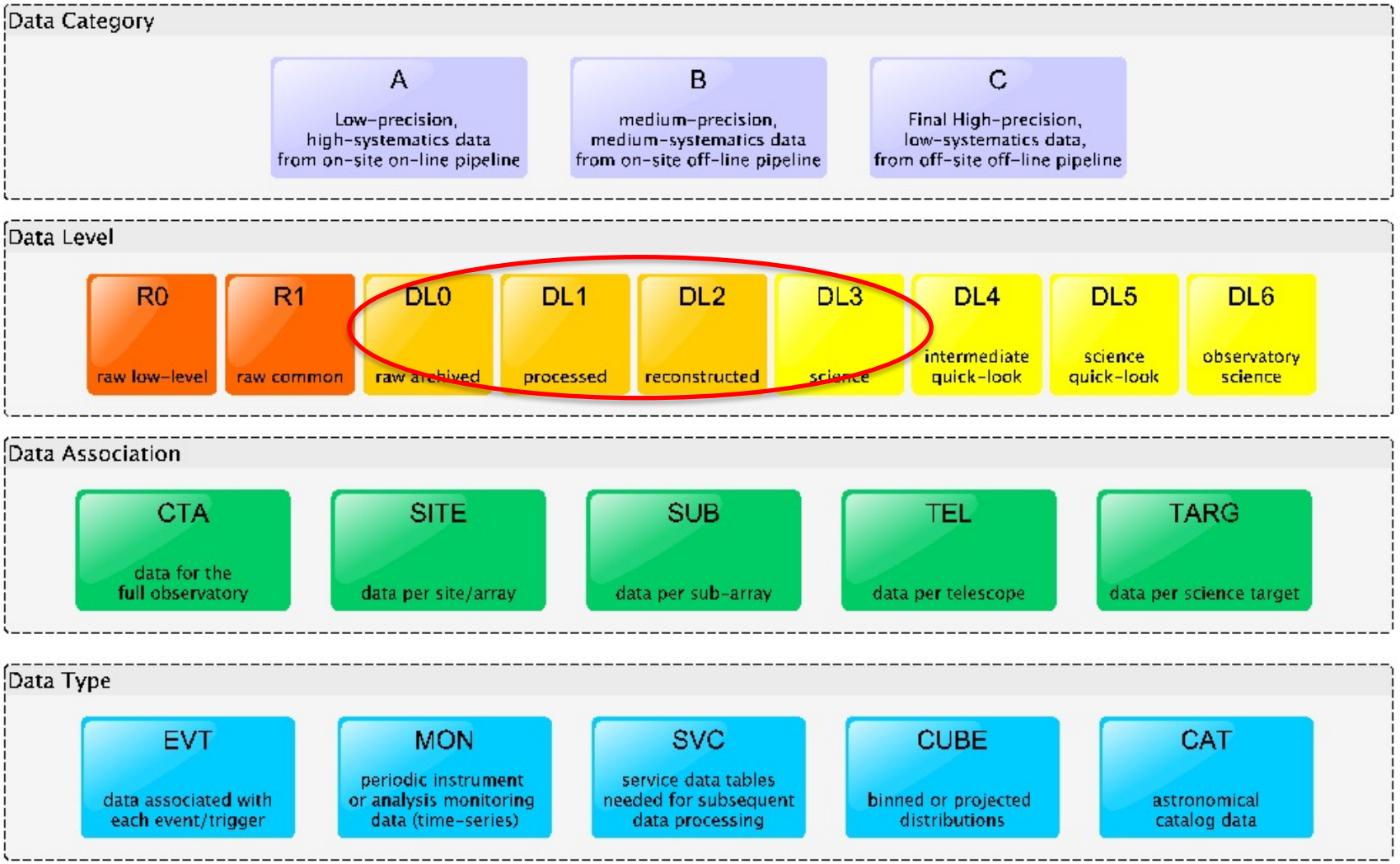
Pre-Construction phase:
- Simulations

Pre & Production phase:
- Data model definition
- Software development
  - Prototype OPUS
  - ctapipe project

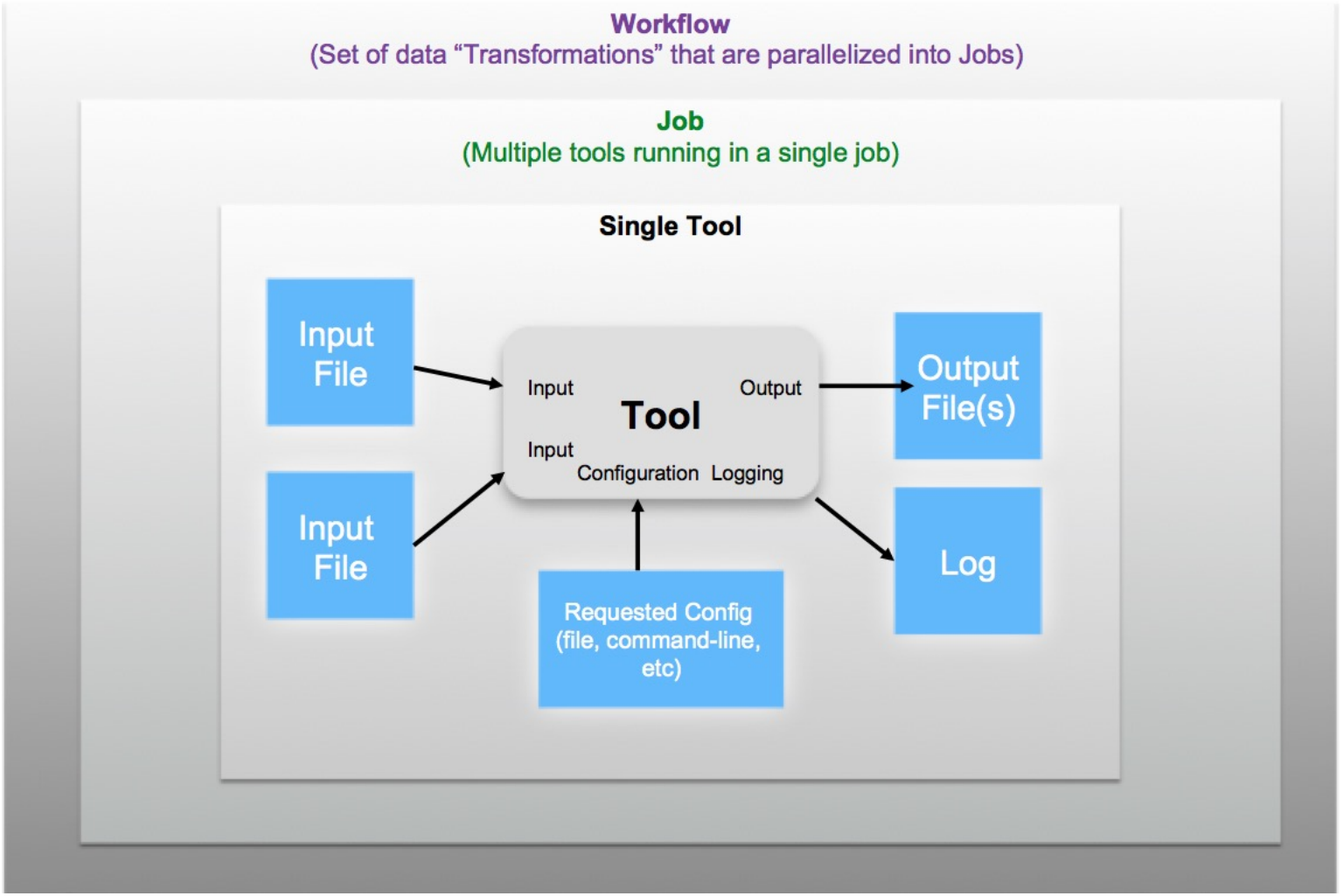
Operation phase:
- Acquisitions
- High level data available



Data Model


Data Category
A
Low-precision, high-systematics data from on-site on-line pipeline
B
medium-precision, medium-systematics data from on-site off-line pipeline
C
Final High-precision, low-systematics data, from off-site off-line pipeline
Data Level
R0
raw low-level
R1
raw common
DL0
raw archived
DL1
processed
DL2
reconstructed
DL3
science
DL4
intermediate quick-look
DL5
science quick-look
DL6
observatory science
Data Association
CTA
data for the full observatory
SITE
data per site/array
SUB
data per sub-array
TEL
data per telescope
TARG
data per science target
Data Type
EVT
data associated with each event/trigger
MON
periodic instrument or analysis monitoring data (time-series)
SVC
service data tables needed for subsequent data processing
CUBE
binned or projected distributions
CAT
astronomical catalog data

Reconstruction pipeline



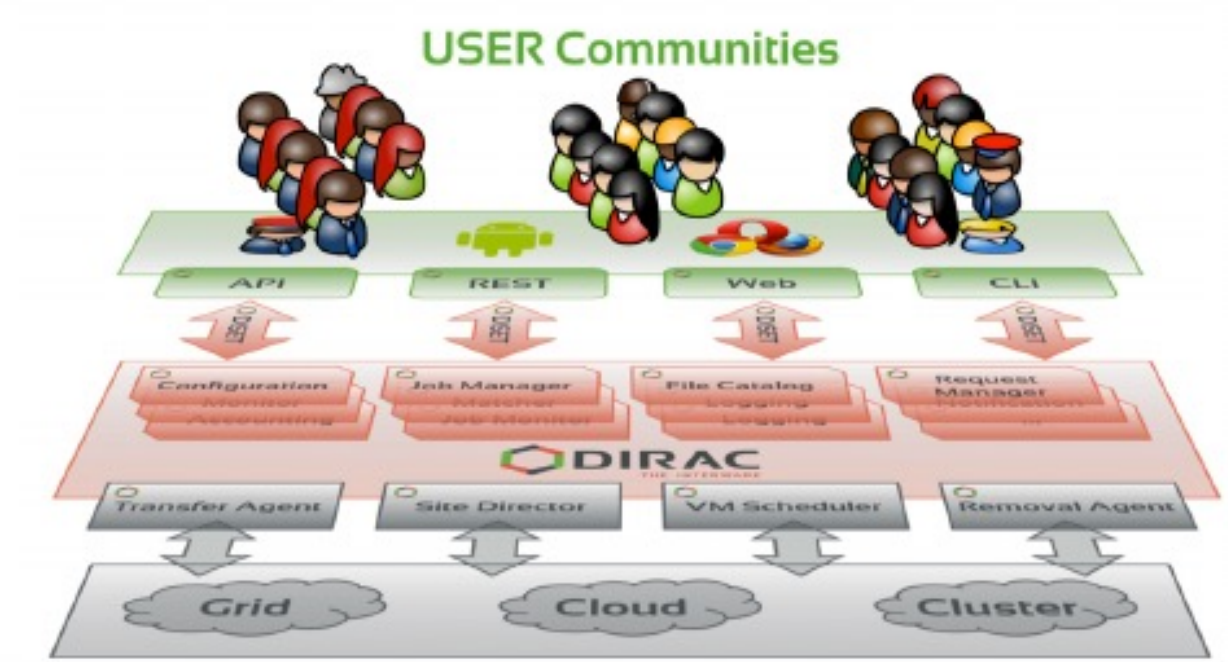
ctapipe

- Ctapipe (Karl Kosack et al.) :
  - « Framework for prototyping the low-level data processing algorithms for the Cherenkov Telescope Array »
  - Documentation: https://cta-observatory.github.io/ctapipe/
  - Framework containing different tools (Python)
  - Provenance Tracking is integrated in the ctapipe framework => Provenance module (ctapipe.core.provenance)
  - Capture of Provenance in each tool of ctapipe
  - Storage of Provenance in a file
  - Source: **https://github.com/cta-observatory/ctapipe**

ctapipe


```
[{'activity_name': 'ctapipe-display-muons',
  'activity_uuid': '93fc2206-59de-4852-868d-ec045ec2
5d1d',
  'config': {'MuonDisplayerTool': {'events': 'proton
_20deg_180deg_run22___cta-prod3-demo-2147m-LaPalma-b
aseline.simtel.gz'}},
  'duration_min': 0.01649999999998819,
  'input': [{'role': 'dl0.sub.evt',
    'url': '/Users/bourgeat/Documents/CTA/Provenance
/ctasoft/ctapipe/tests/proton_20deg_180deg_run22___c
ta-prod3-demo-2147m-LaPalma-baseline.simtel.gz'}],
  'output': [{'role': 'dl1.tel.evt.muon',
    'url': '/Users/bourgeat/Documents/CTA/Provenance
/ctasoft/ctapipe/tests/muons.hdf5'}],
  'start': {'time_utc': '2019-05-07T06:28:44.605'},
  'status': 'completed',
  'stop': {'time_utc': '2019-05-07T06:28:45.595'},
  'system': {'arguments': ['/Users/bourgeat/anaconda
3/lib/python3.6/site-packages/ipykernel_launcher.py'
,
    '-f',
    '/Users/bourgeat/Library/Jupyter/runtime/kernel-
d793d0e4-abd3-4989-a71c-19b68194ea9e.json'],
   'ctapipe_resources_version': '0.2.15',
   'ctapipe_svc_path': None,
   'ctapipe_version': '0.6.1',
```

DIRAC et CTA-DIRAC



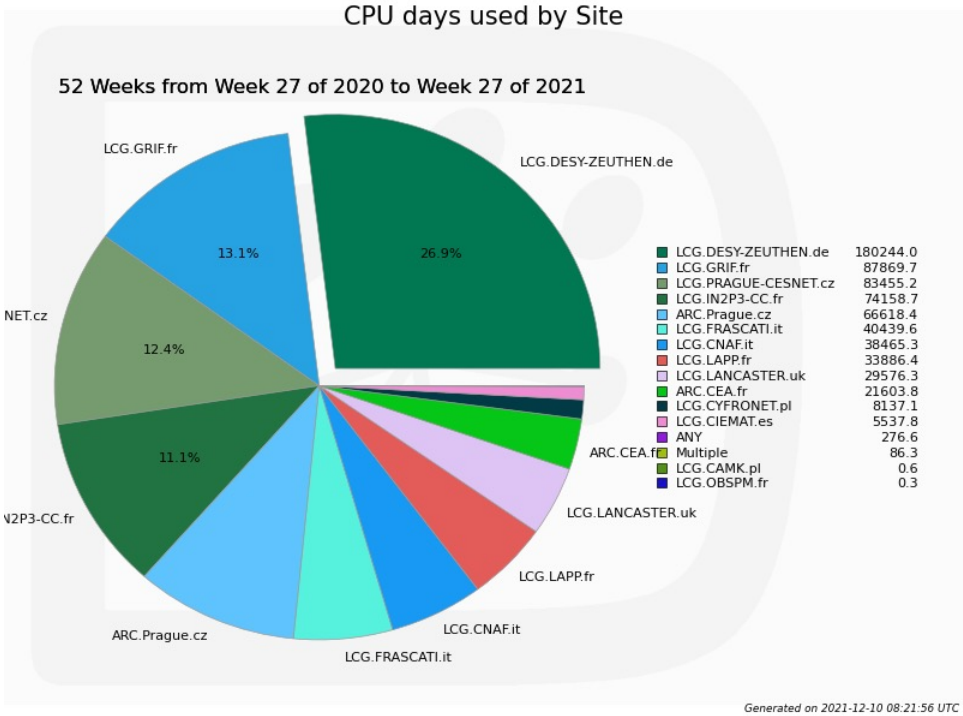
CTADIRAC: Prototype based on DIRAC - L. Arrabito - LUPM used to manage distributed computing on heterogeneous resources (grid, cloud, clusters) for the CTA experiment

CTA-DIRAC

Grid resources used for Prod5(b) (mid 2020 to mid 2021)

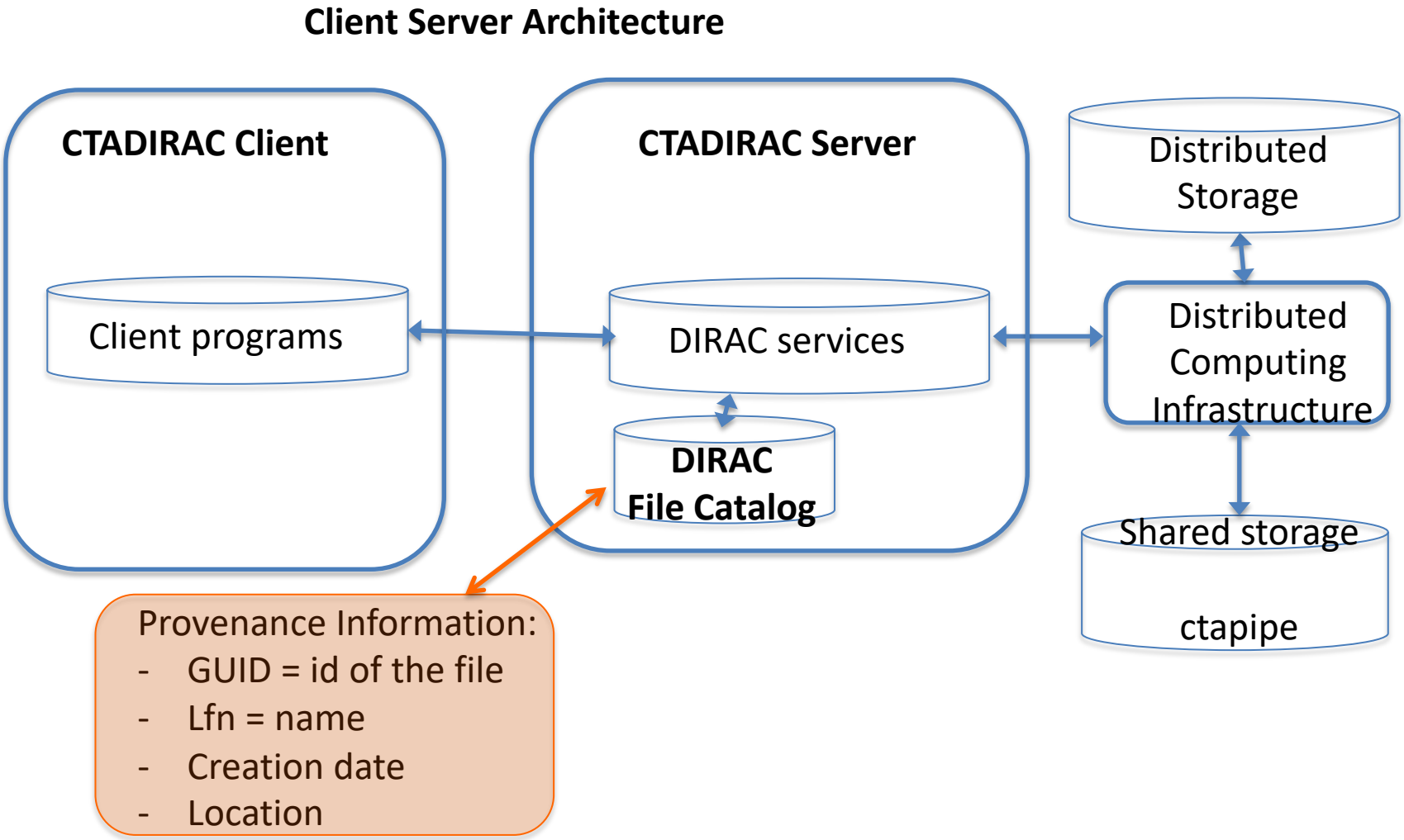
Total CPU: 164 million CPU hours HS06

Total disk: ~1.7 PB



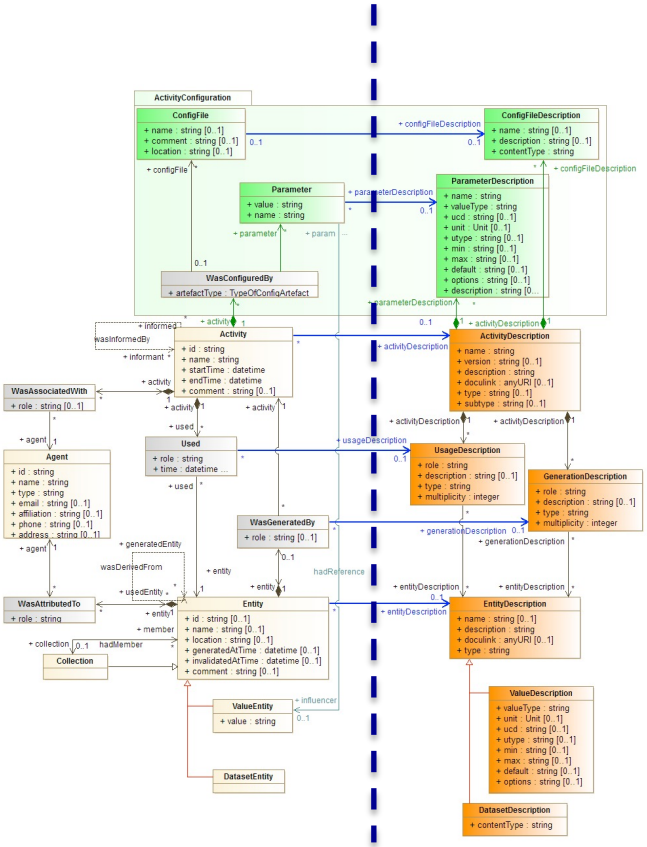
Site	Used (TB)
DESY-ZEUTHEN	158
IN2P3-CC	112
GRIF (LPNHE+CEA+LLR)	486 (154+4+328)
IN2P3-LAPP	70
INFN-T1	676
UKI-NORTHGRID-LANCS-HEP	163
Total	1665

CTA-DIRAC


Client Server Architecture
CTADIRAC Client
Client programs
CTADIRAC Server
DIRAC services
DIRAC
File Catalog
Distributed
Storage
Distributed
Computing
Infrastructure
Shared storage
ctapipe
Provenance Information:
- GUID = id of the file
- Lfn = name
- Creation date
- Location

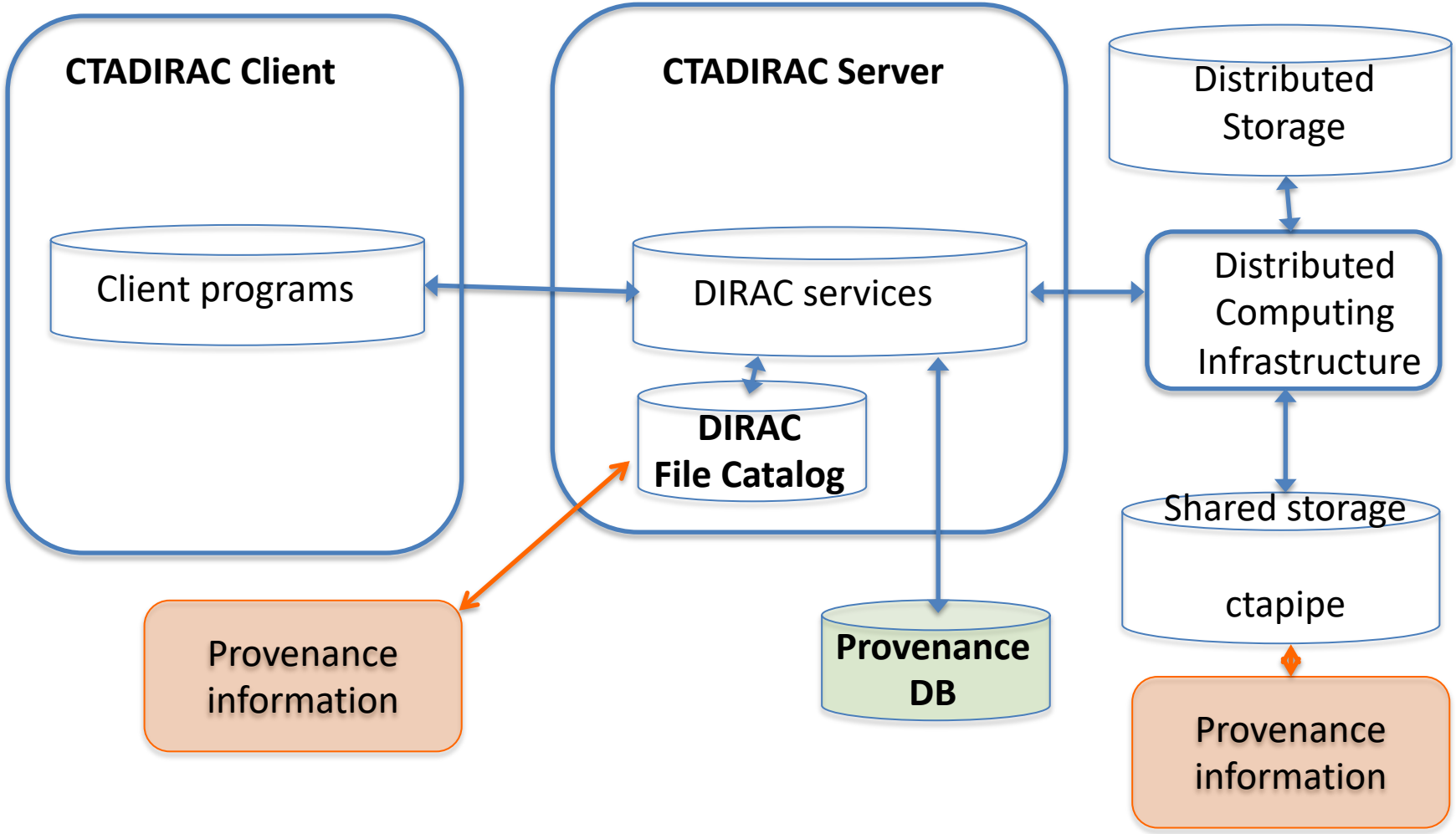
IVOA Provenance DM

Information relative to each execution of tool



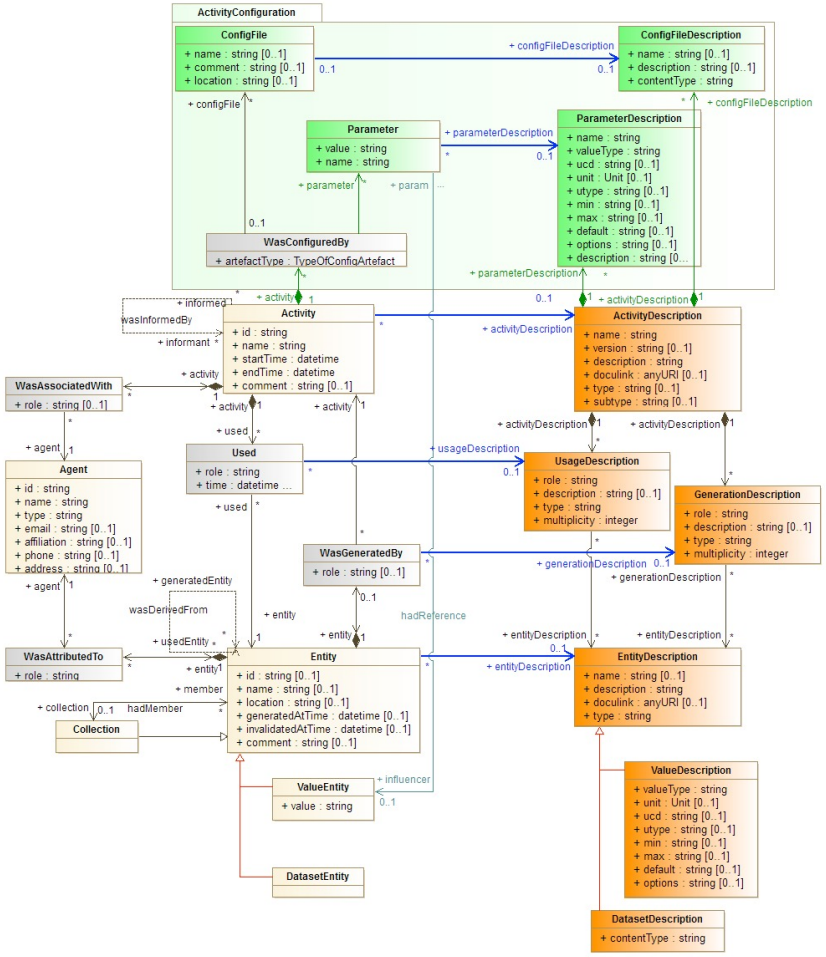
Descriptions relative to each type of tool

Goal


CTADIRAC Client
Client programs
CTADIRAC Server
DIRAC services
DIRAC
File Catalog
Distributed
Storage
Distributed
Computing
Infrastructure
Shared storage
ctapipe
Provenance
information
Provenance
DB
Provenance
information

IMPLEMENTATION

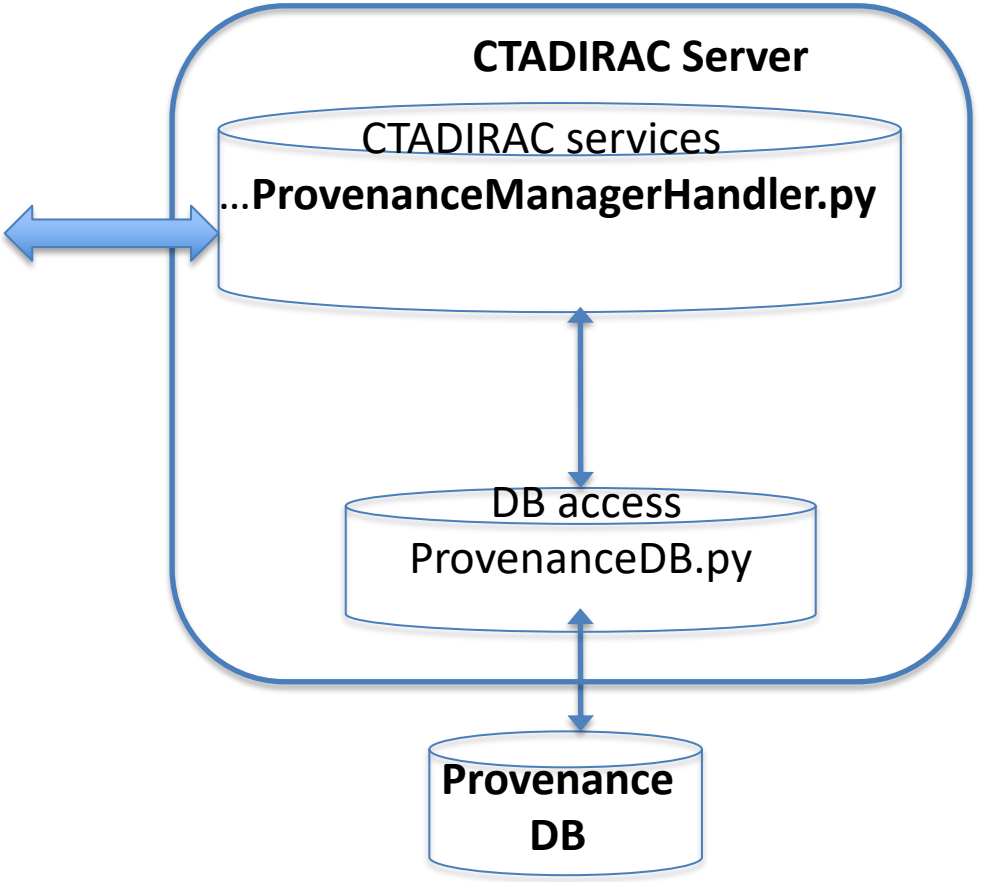
Provenance DB



- Postgresql DB
- Sqlalchemy module
- Each table has an auto incremented internal key (not in the model)
- Model restrictions:
  - ActivityDescription : unicity of (name, version)
  - UsageDescription & GenerationDescription: unicity of (ActivityDescription, EntiityDescription, role) => multiplicity = 1
  - Link Parameter to ValueEntity not yet implemented

CTADIRAC Server

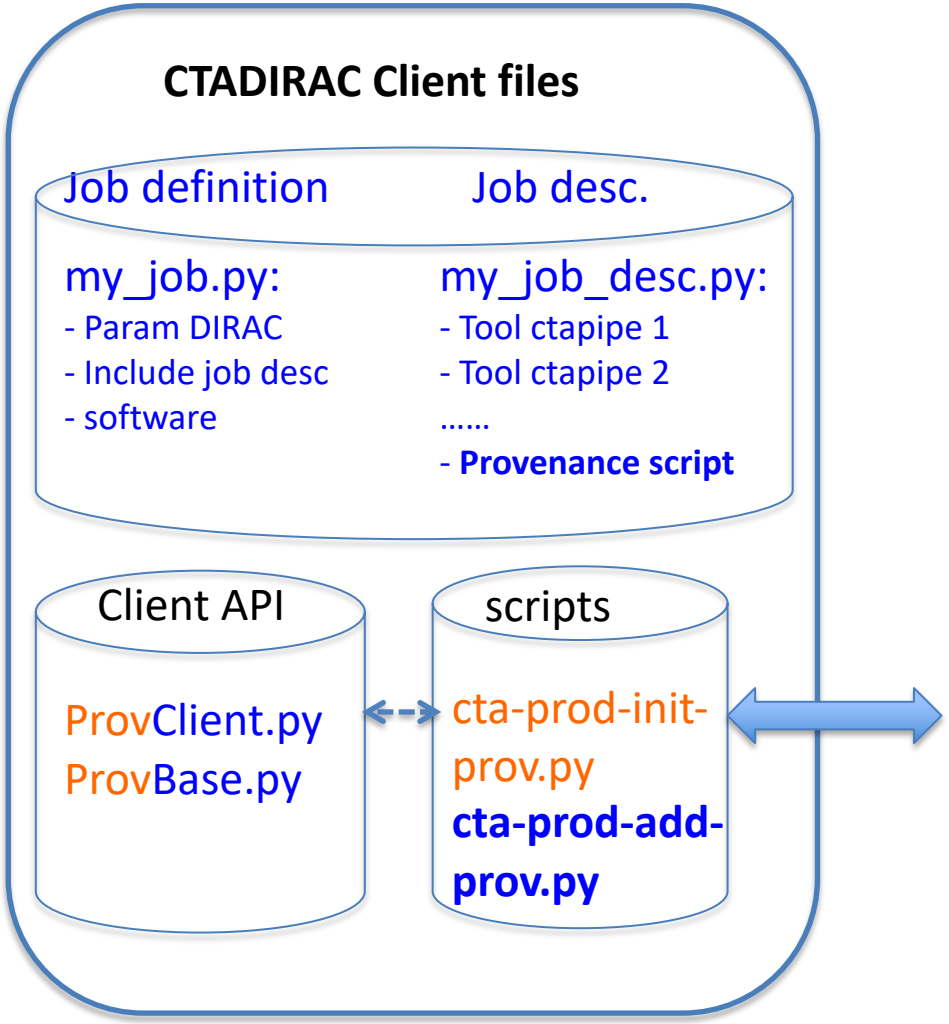
- Server:
  - Provenance Manager
  - Provenance DB access



CTADIRAC Client

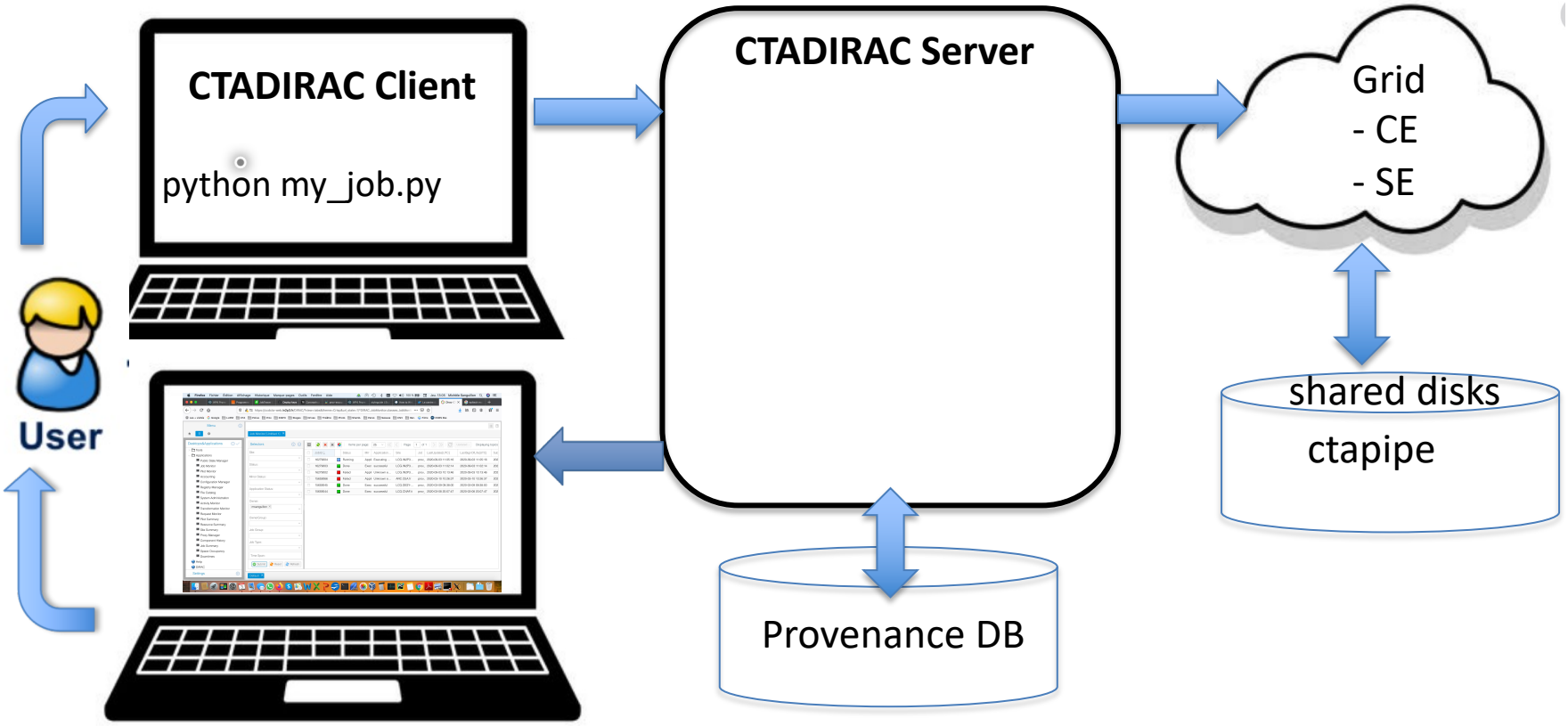
- API
  - ProvBase (Objects)
  - ProvClient (Functionalities)
- Scripts
  - Script to add the descriptions and the agent (CTAO)
  - Script to add the provenance execution information
- Job definition
  - DIRAC job definition
  - Job description (steps)

Used to add description information
Used to add execution information



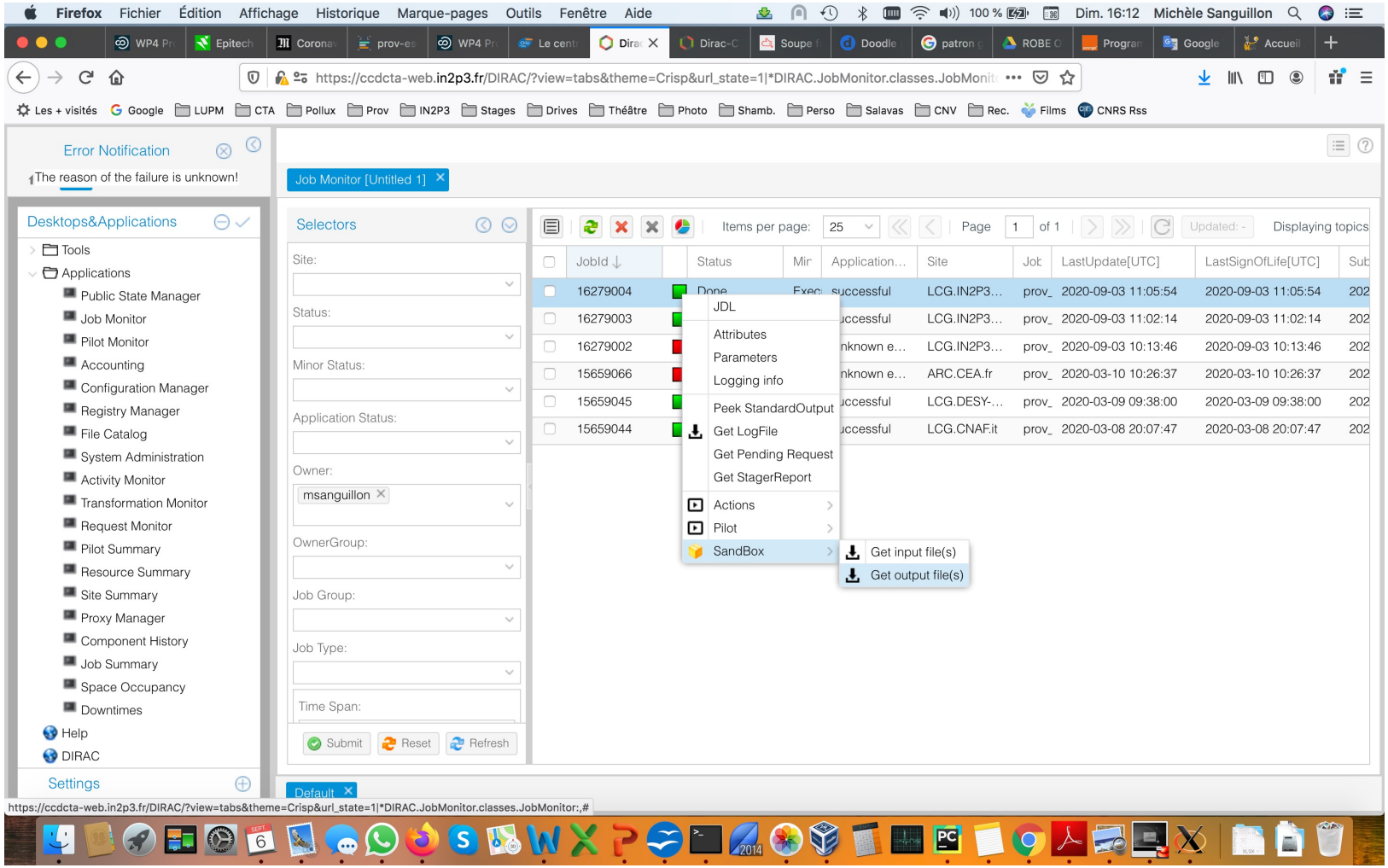
USAGE

Job execution usage

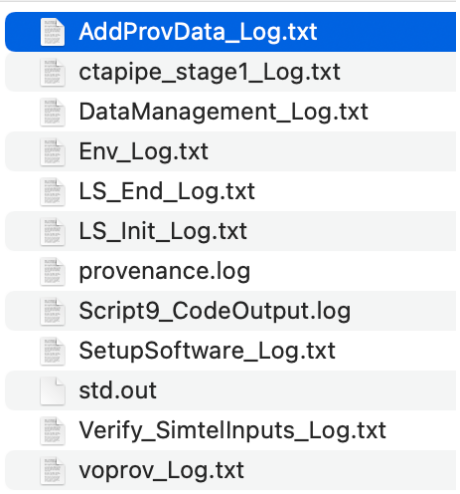


Job monitoring

https://ccdcta-cert.in2p3.fr/



Sandbox content




```
{"comment": "", "name": "ctapipe-display-muons", "activityDescription_key": 1, "startTime":
"2020-09-03T11:05:32.873", "endTime": "2020-09-03T11:05:38.361", "id":
"a4342513-329d-4d3d-806b-20d08b6164ad"}
{"agent_key": 1, "activity_key": 3, "role": null}
{"agent_key": 1, "entity_key": 1, "role": null}
{"activity_key": 3, "role": "dl0.sub.evt", "usageDescription_key": 1, "entity_key": 1, "time":
null}
{"agent_key": 1, "entity_key": 2, "role": null}
{"activity_key": 3, "role": "dl1.tel.evt.muon", "generationDescription_key": 1, "entity_key":
2}
{"generatedAtTime": null, "comment": null, "name": "status", "classType": "value",
"entityDescription_key": 3, "location": null, "invalidatedAtTime": null, "id":
"a4342513-329d-4d3d-806b-20d08b6164ad_status"}
{"activity_key": 3, "role": "status", "generationDescription_key": 2, "entity_key": 4}
Done
```

DB content

```
ctaprovtest=> select * from "activities";
 internal_key |                  id                  |         name          |        startTime        |         endTime         | comment | activityDescription_key
--------------+--------------------------------------+-----------------------+-------------------------+-------------------------+---------+-------------------------
            1 | 5dc245d2-aa16-4026-9da9-8c763b148a2c | ctapipe-display-muons | 2020-09-03T10:13:17.790 | 2020-09-03T10:13:24.010 |         |                       1
            2 | 4ed870fb-a3a0-48e9-8444-45dd63460996 | ctapipe-display-muons | 2020-09-03T11:01:49.802 | 2020-09-03T11:01:55.353 |         |                       1
            3 | a4342513-329d-4d3d-806b-20d08b6164ad | ctapipe-display-muons | 2020-09-03T11:05:32.873 | 2020-09-03T11:05:38.361 |         |                       1
            4 | 993734a7-52b7-496a-a9ed-87ddbf79f697 | ctapipe-display-muons | 2020-09-03T13:05:47.367 | 2020-09-03T13:05:53.020 |         |                       1
(4 rows)

ctaprovtest=> select * from "activityDescriptions";
 internal_key |         name          | version |     description      |      type      | subtype | doculink
--------------+-----------------------+---------+----------------------+----------------+---------+----------
            1 | ctapipe-display-muons | 0.6.2   | Muon reconstruction | Reconstruction |         |
(1 row)

ctaprovtest=> select * from "entities";
 internal_key |                   id                        |                                                         name
|    location     |   generatedAtTime    | invalidatedAtTime | comment | classType | entityDescription_key
--------------+---------------------------------------------+------------------------------------------------------------------------------------------------------------------
+-----------------+----------------------+-------------------+---------+-----------+-----------------------
            1 | 1C1A6113-481A-C0E9-ADF3-26A9E3F19AC6        | /vo.cta.in2p3.fr/user/s/sanguillon/proton_20deg_180deg_run22___cta-prod3-demo-2147m-LaPalma-baseline.simtel.gz
| {CC-IN2P3-USER} | 2019-05-02 12:00:52 |                   |         | dataset   |                     1
            2 | 9CD2D6F0-E512-7ED3-0825-271B083201C1        | /vo.cta.in2p3.fr/user/s/sanguillon/LaPalma/PROTON/provtest/0000/Data/-01xxx/muons_22.hdf5
| {DESY-ZN-Disk}  | 2020-09-03 11:02:03 |                   |         | dataset   |                     2
            3 | 4ed870fb-a3a0-48e9-8444-45dd63460996_status | status
|                 |                      |                   |         | value     |                     3
            4 | a4342513-329d-4d3d-806b-20d08b6164ad_status | status
|                 |                      |                   |         | value     |                     3
            5 | 993734a7-52b7-496a-a9ed-87ddbf79f697_status | status
|                 |                      |                   |         | value     |                     3
(5 rows)

ctaprovtest=> select * from "used";
 internal_key |    role     | time | activity_key | entity_key | usageDescription_key
--------------+-------------+------+--------------+------------+----------------------
            1 | dl0.sub.evt |      |            2 |          1 |                    1
            2 | dl0.sub.evt |      |            3 |          1 |                    1
            3 | dl0.sub.evt |      |            4 |          1 |                    1
(3 rows)
```

I thank you for your attention