



cherenkov
telescope
array



Storage of Provenance information: example of CTADIRAC

Tests in progress
thanks to the help and the
contribution of Luisa Arrabito



cherenkov
telescope
array

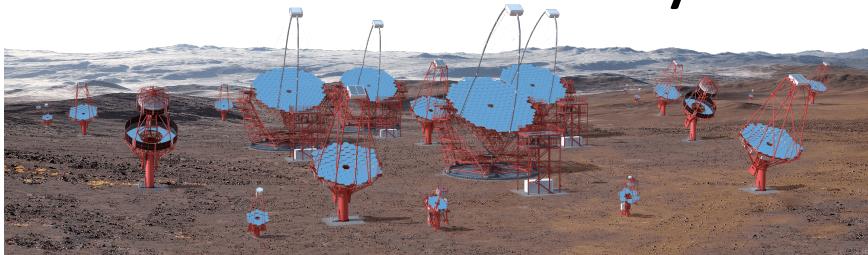


CONTEXT

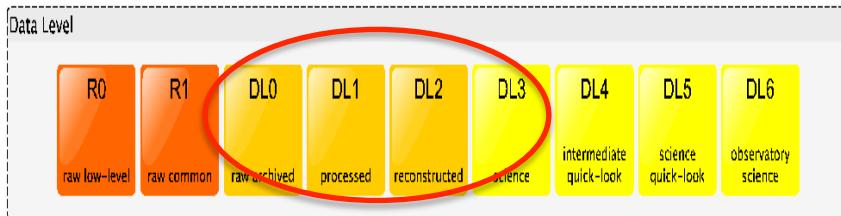
Context



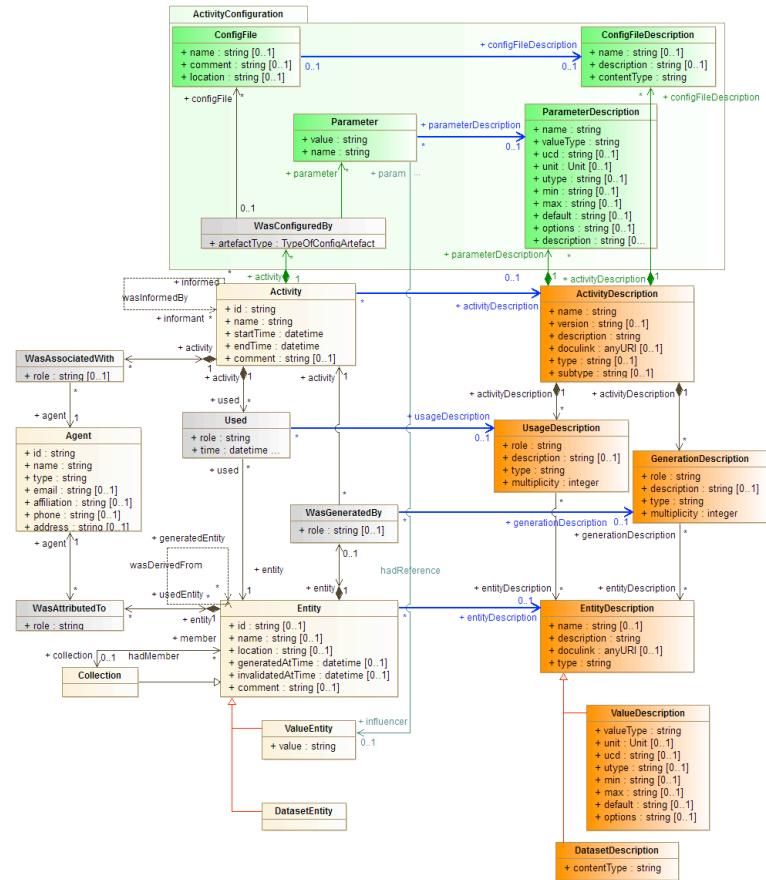
- CTA Observatory

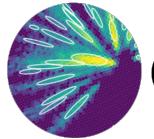


- DL0-DL3 pipeline



- IVOA



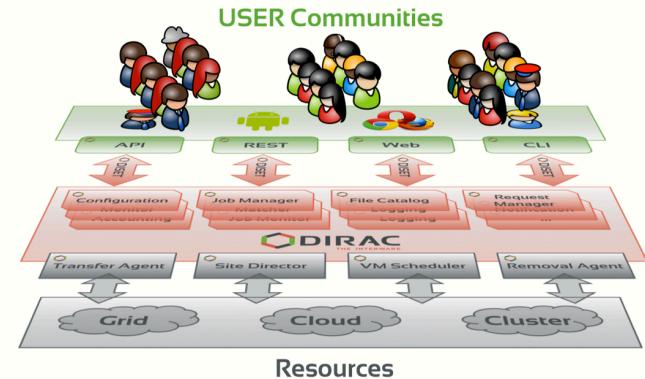


ctapipe

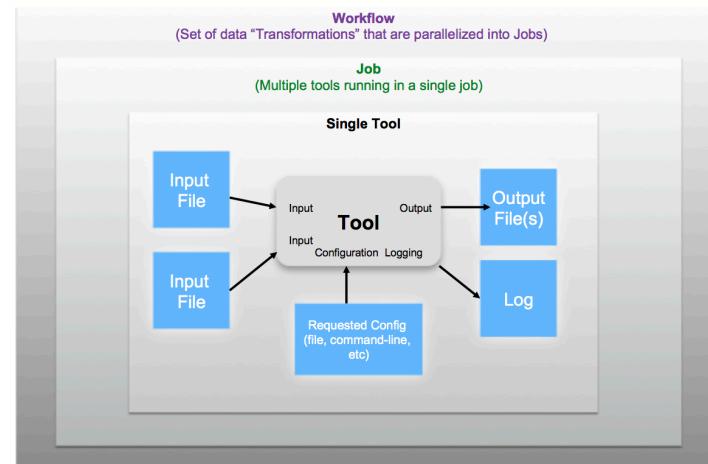
Pipeline context



- **CTADIRAC** (Prototype based on DIRAC - L. Arrabito - LUPM) used to manage distributed computing on heterogeneous resources (grid, cloud, clusters)



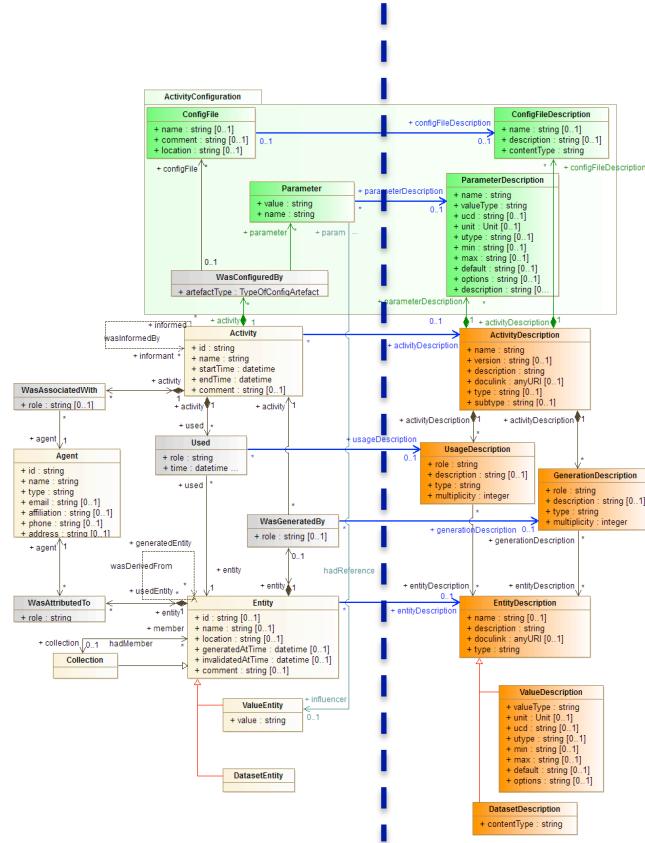
- **ctapipe** library (including the Provenance module which captures Provenance information - K. Kosack - CEA)



IVOA Provenance DM

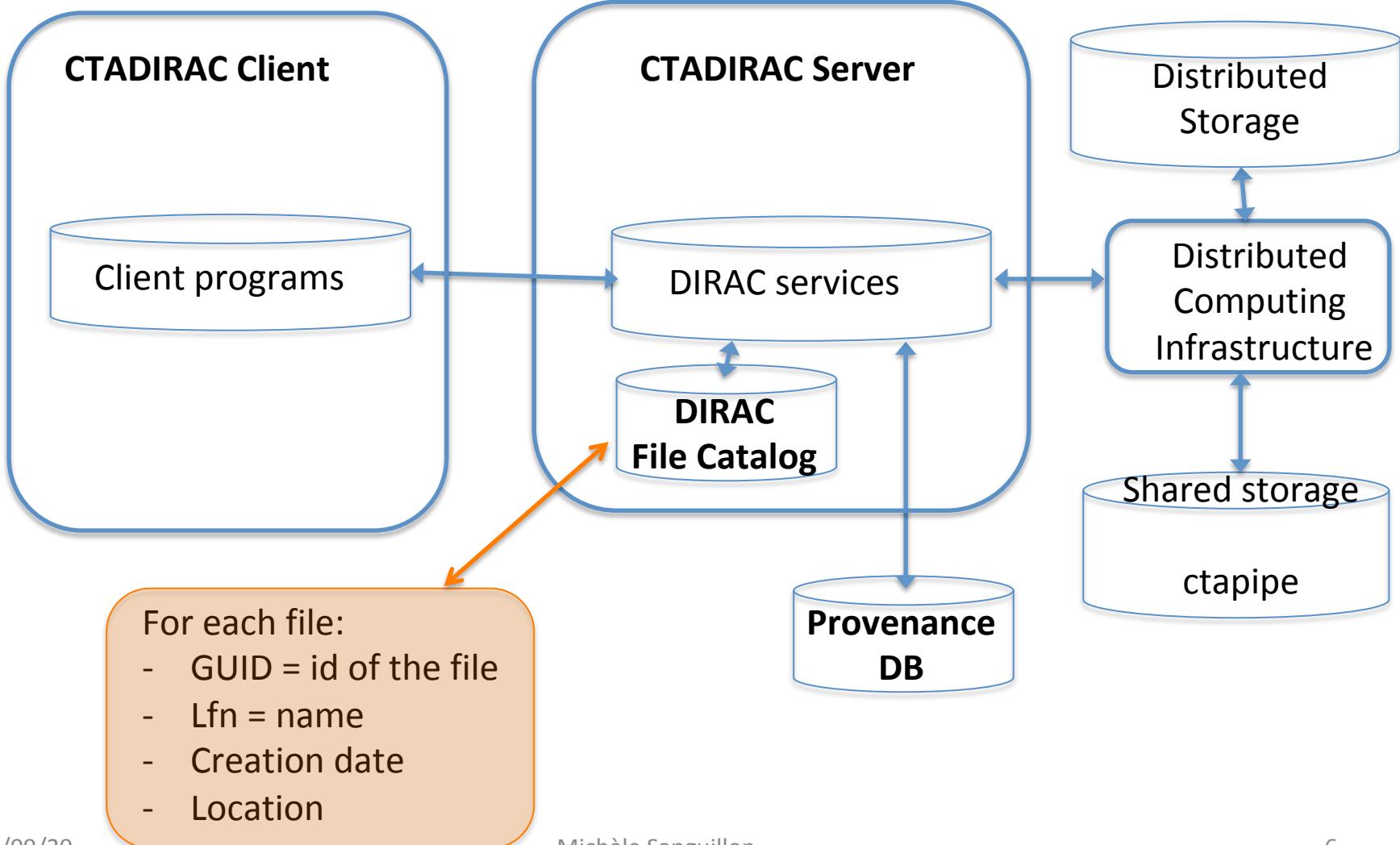


Information relative to each execution of tool



Descriptions relative to each type of tool

CTADIRAC



ctapipe

- Provenance module
 - `ctapipe.core.provenance`
 - Developed so that each tool automatically register provenance information
 - This module has been  designed to return a JSON file with all provenance information

```
[{"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": 0.0164999999998819,
 "input": [{"role": "d10.sub.evt",
 "url": "/Users/bourgeat/Documents/CTA/Provenance
/ctasoft/ctapipe/tests/proton_20deg_180deg_run22_cta-
prod3-demo-2147m-LaPalma-baseline.simtel.gz"}],
 "output": [{"role": "d11.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"},
```



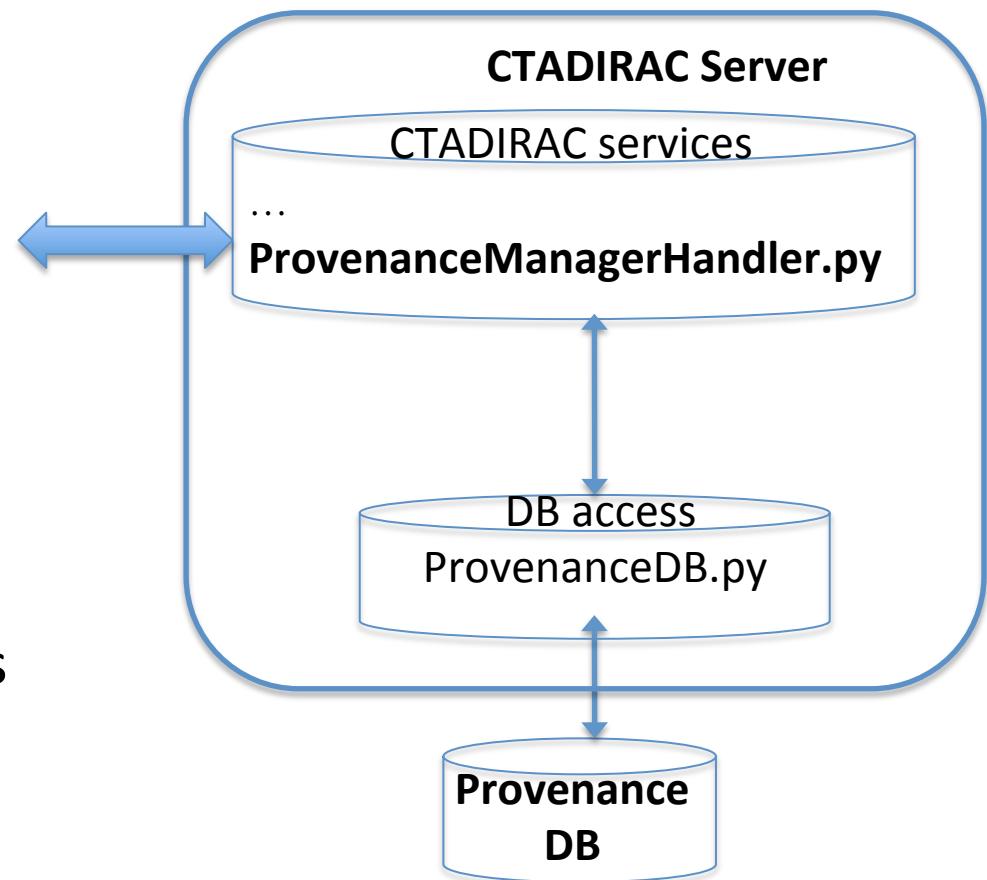
cherenkov
telescope
array



IMPLEMENTATION

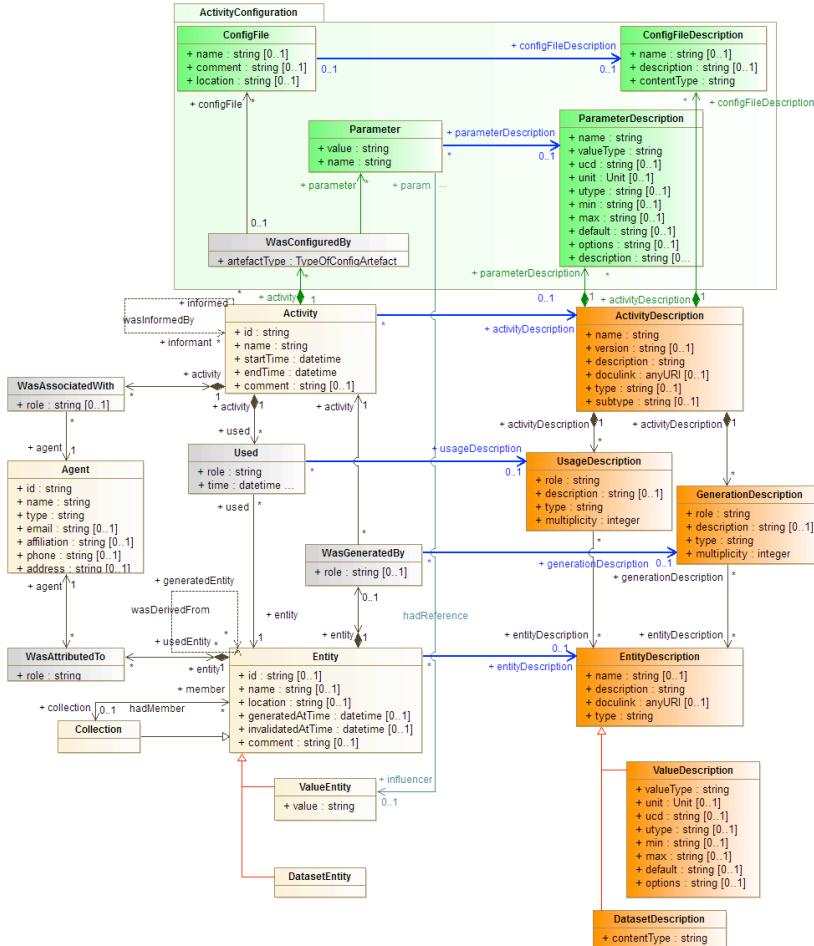
CTADIRAC Server

- Server:
 - Provenance Manager
 - Provenance DB access





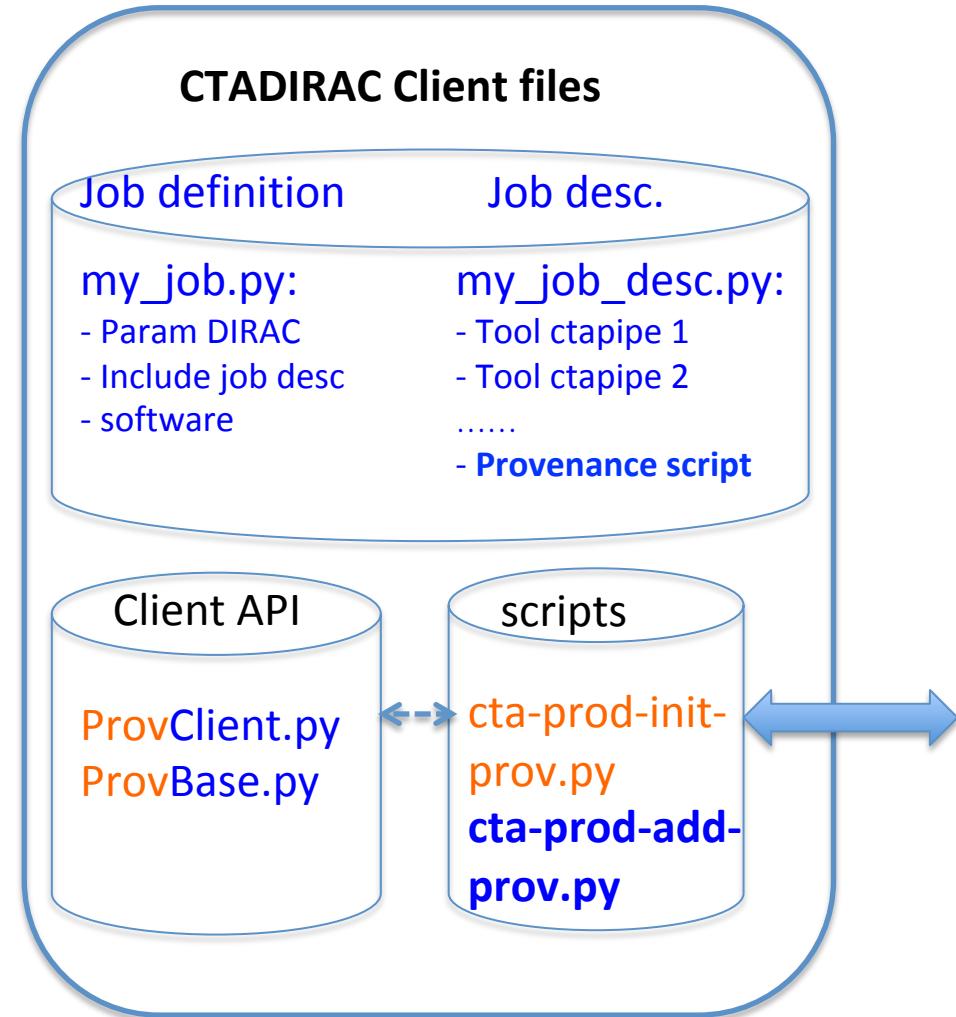
Provenance DB



- 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, EntityDescription, role) => multiplicity = 1
 - Link Parameter to ValueEntity not yet implemented

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

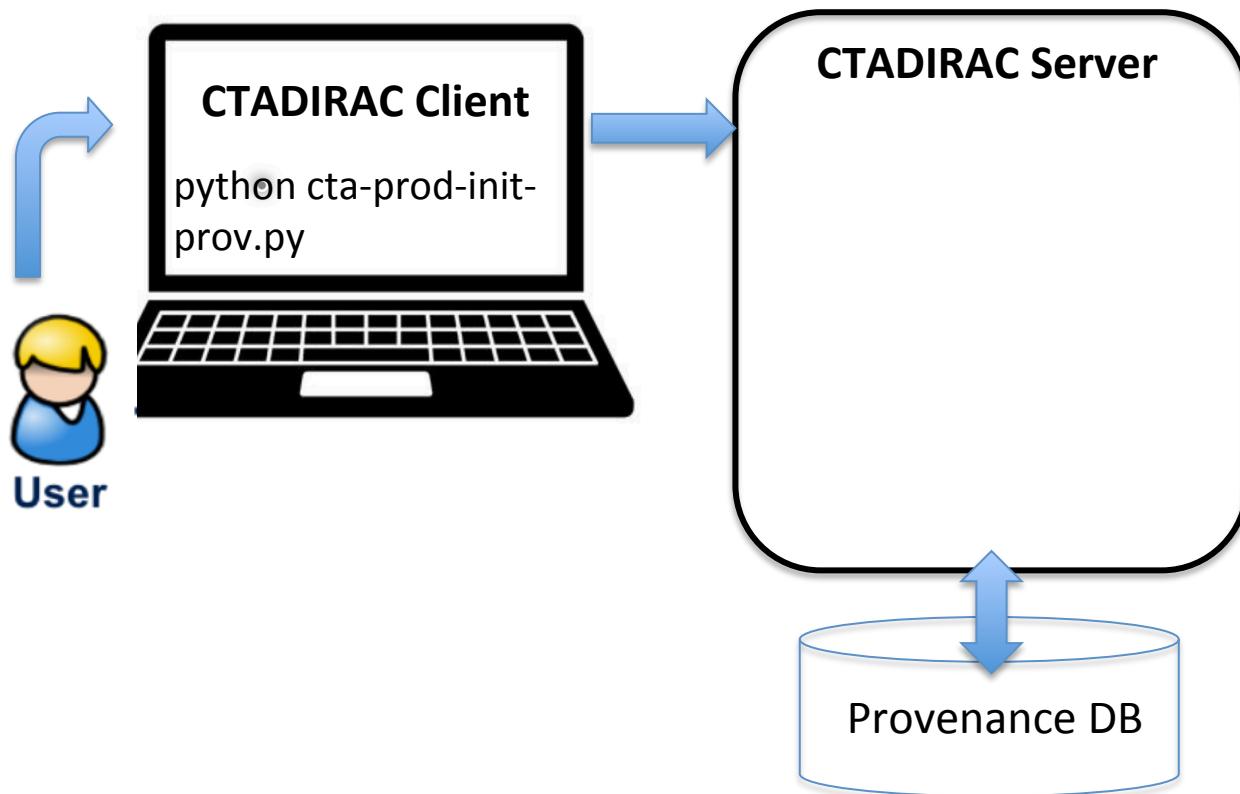


cherenkov
telescope
array

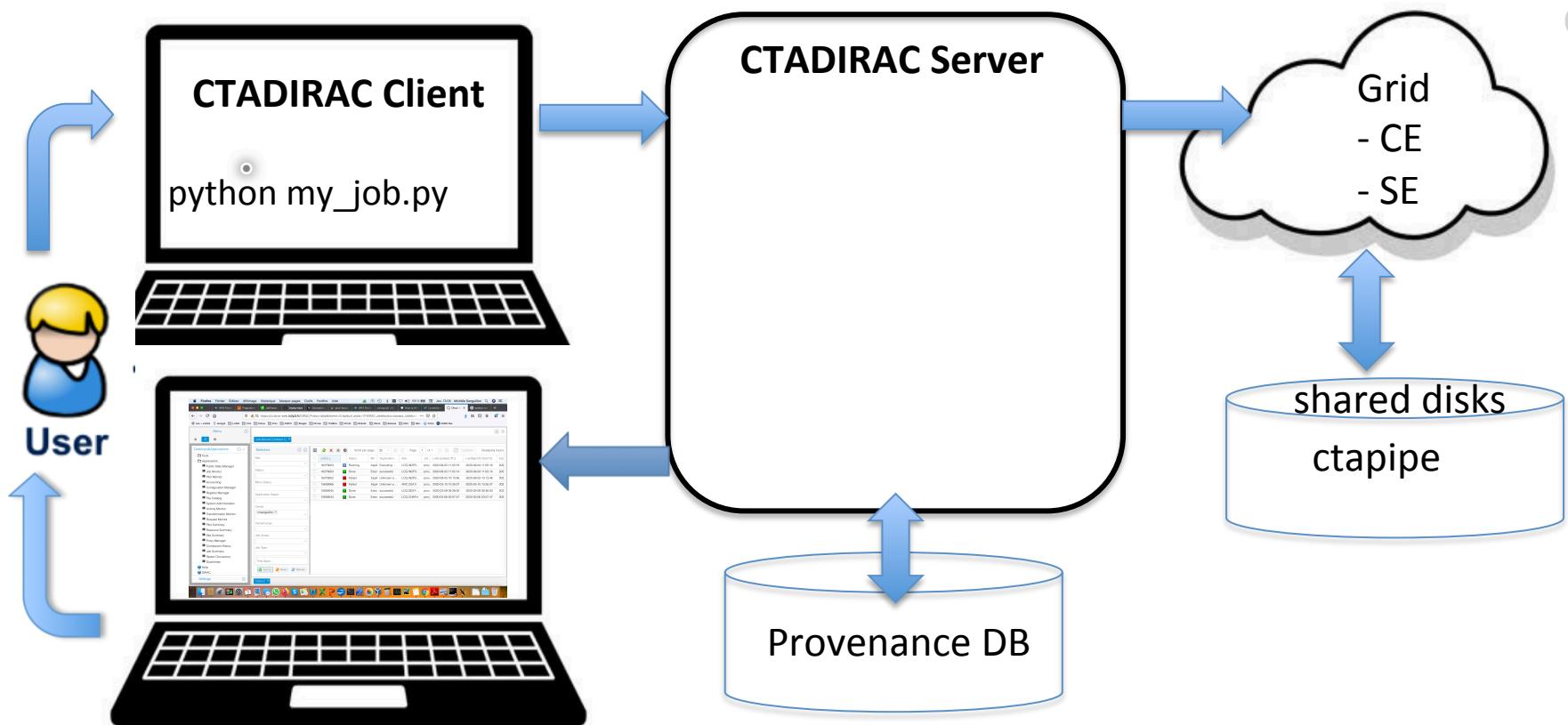


USAGE

Adding descriptions



Job execution usage



Job monitoring

<https://ccdcta-web.in2p3.fr/>

The screenshot shows a Firefox browser window with the following details:

- Address Bar:** https://ccdccta-web.in2p3.fr/DIRAC/?view=tabs&theme=Crisp&url_state=1*DIRAC.JobMonitor.classes.JobMonitor;#
- Toolbar:** Includes standard Firefox icons like Back, Forward, Stop, Home, and Refresh.
- Bookmarks Bar:** Shows various bookmarks including "WP4 Pro", "Epitech", "Corona", "prov-es", "WP4 Pro", "Le cent", "Dirac X", "Dirac-C", "Soupe f", "Doodle", "Google patron", "ROBE", "Program", "Google Accueil", and "Accueil".
- Address Bar:** https://ccdccta-web.in2p3.fr/DIRAC/?view=tabs&theme=Crisp&url_state=1*DIRAC.JobMonitor.classes.JobMonitor;#
- Page Content:**
 - Error Notification:** A message stating "The reason of the failure is unknown!"
 - Job Monitor [Untitled 1]:** A table listing jobs with columns: JobId, Status, Mir, Application..., Site, Job, LastUpdate[UTC], LastSignOfLife[UTC], Sub. The table contains several entries, e.g., JobId 16279004, Status Done, Application LCG.IN2P3..., Site prov_, LastUpdate 2020-09-03 11:05:54.
 - Selectors:** A sidebar with dropdown menus for Site, Status, Minor Status, Application Status, Owner (msanguillon), OwnerGroup, Job Group, Job Type, and Time Span. Buttons for Submit, Reset, and Refresh are at the bottom.
 - Context Menu:** A context menu is open over a row in the table, listing options like Done, Executed, successful, JDL, Attributes, Parameters, Logging info, Peek StandardOutput, GetLogFile, Get Pending Request, Get StagerReport, Actions, Pilot, and SandBox. The "SandBox" option is highlighted.
- Bottom Bar:** Shows the Mac OS Dock with various application icons.

Sandbox content

	AddProvData_Log.txt
	Ctapipe_Log.txt
	DataManagement_Log.txt
	LS_End_Log.txt
	LS_Init_Log.txt
	output_lfns.txt
	provDict.txt
	SetupSoftware_Log.txt
	std.out
	Verify_Ctapip...nputs_Log.txt

```
{
  "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-87dbbf79f697 | 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 | 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 | | | dataset | 1
| {CC-IN2P3-USER} | 2019-05-02 12:00:52 | | | | | |
 2 | 9CD2D6F0-E512-7ED3-0825-271B083201C1 | /vo.cta.in2p3.fr/user/s/sanguillon/LaPalma/PROTON/provtest/0000/Data/-01xxx/muons_22.hdf5 | | | dataset | 2
| {DESY-ZN-Disk} | 2020-09-03 11:02:03 | | | | | |
 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-87dbbf79f697_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)
```



cherenkov
telescope
array



I thank you for your attention