DIRAC for users, MC productions and CTACG ressources

J. Bregeon, L. Arrabito, A. Haupt
IN2P3–LUPM, DESY

La Palma, ASGW parallel session, November 7th 2017
- DIRAC for users
  - what is DIRAC
  - wiki pages: main, install client, simulations status, available software (cvmfs), PROD3(b) status, link to tutorials
  - user support (issues)
  - highlight: COMDIRAC for data management

- Monte Carlo simulation and analysis
  - Prod3(b) continues: La Palma reference, Paranal NSB, Paranal 60 deg
  - technical tests: zstd compression test, corsika version, build architecture
  - atmosphere: transmission (clouds) and density profiles
  - divergent pointing, degraded telescopes

- CTACG resources, now and for 2018
  - what we have consumed so far
  - requests for 2018
What is DIRAC again?

- **DIRAC**: Distributed Infrastructure with Remote Agent Control is a software framework for distributed computing.
- **CTADIRAC** is a plugin built on top of DIRAC to meet CTA needs in the pre-construction phase:
  - i.e. CTA users portal to access large distributed computing and storage resources.
- CTA DIRAC servers and dbs are hosted at CC-IN2P3, PIC and DESY-Zeuthen (including service/agents redundancy).
- **CTA DIRAC team**: L. Arrabito (LUPM), J. Bregeon (LUPM) and A. Haupt (DESY).
  - plus help from the large DIRAC community and from our main computing centres.

→ **CTADIRAC** is also proposed as a prototype for CTA Computing Resource Management System (see Luisa’s talk at the DPPS session).
DIRAC for CTA

- Refer to the CTADIRAC project on the Redmine forge

- step by step CTADIRAC client installation and "hello world"
- tutorials to run applications like corsika, simtel_array, EventDisplay and Mars
  - no tutorial yet for ctapipe but we should do that soon
- tutorials for simple data management
- available software distributed via CVMFS

- PROD3(b) status
- simulation and analysis productions summary
Open an Issue, or send an email...

"we" like to know what "you" are doing, in particular if you will need large amount of CPU and/or disk space
COMDIRAC for Data Management

▶ COMDIRAC is a DIRAC plugin meant to provide Unix like commands for most DIRAC component
▶ COMDIRAC is now shipped with the CTADIRAC client
  ▶ particularly useful as a Unix like access to the file catalog, as shown below

# Move into file catalog directory
[pc]$ dpwd
/vo.cta.in2p3.fr/MC/PROD3/LaPalma/gamma/evndisp/1394
[pc]$ dls -IH Data/069xxx
/vo.cta.in2p3.fr/MC/PROD3/LaPalma/gamma/evndisp/1394/Data/069xxx:
  −rwxrwxrx— 1 ar rabito cta_prod 190.5MiB 2017−09−16 00:26:47
gamma_20deg_180deg_tid00069200___cta−prod3−demo−2147m−LaPalma−baseline_evndisp−f05−NN−DL2.root

# Download a file
[pc]$ dcd Data/069xxx
[pc]$ dget gamma_20deg_180deg_tid00069200___cta−prod3−demo−2147m−LaPalma−baseline_evndisp−f05−NN−DL2.root
[pc]$ ls gam*
gamma_20deg_180deg_tid00069200___cta−prod3−demo−2147m−LaPalma−baseline_evndisp−f05−NN−DL2.root
COMDIRAC is a DIRAC plugin meant to provide Unix like commands for most DIRAC component

COMDIRAC is now shipped by defaults with the CTADIRAC client

particularly useful as a Unix like access to the file catalog, as shown below

```
# Upload a file on a storage element
[pc]$ dcd /vo.cta.in2p3.fr/
[pc]$ dput arandomfile.py user/b-bregeon/myfile.py
[pc]$ dls user/b-bregeon/myfile.py
/vo.cta.in2p3.fr/user/b-bregeon/myfile.py:
    rwxrwxr-x 0 bregeon cta_prod 60553 2017-10-31 15:52:01 myfile.py
```

```
# find files by meta data
[pc]$ dfind /vo.cta.in2p3.fr/MC/PROD3 configuration_id=2
/vo.cta.in2p3.fr/MC/PROD3/Paranal/gamma/simtel/1575/Data/000xxx/
     simtel.zst
/vo.cta.in2p3.fr/MC/PROD3/Paranal/gamma/simtel/1575/Data/000xxx/
     gamma_20deg_180deg_run18___cta–prod3–demo_desert—2150m–Paranal–baseline.
     simtel.zst
```
## Simulation and analysis productions summary

### Running and planned

<table>
<thead>
<tr>
<th>#</th>
<th>Status</th>
<th>% Done</th>
<th>Priority</th>
<th>Subject</th>
<th>Category</th>
<th>Assigned To</th>
<th>Parent task</th>
<th>Due date</th>
</tr>
</thead>
<tbody>
<tr>
<td>17019</td>
<td>In progress</td>
<td></td>
<td>Normal</td>
<td>Run a test production with reference configuration for both sites to assess minimal resources consumption</td>
<td>MC Production</td>
<td>Bregeon Johan</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21541</td>
<td>In progress</td>
<td></td>
<td>Normal</td>
<td>Prod3b High NSB production</td>
<td>MC Production</td>
<td>Bregeon Johan</td>
<td></td>
<td></td>
</tr>
<tr>
<td>22108</td>
<td>Assigned</td>
<td></td>
<td>Low</td>
<td>Run prod3b production at 60 deg zenith angle for Paranal</td>
<td>MC Production</td>
<td>Bregeon Johan</td>
<td></td>
<td></td>
</tr>
<tr>
<td>22246</td>
<td>In progress</td>
<td></td>
<td>Normal</td>
<td>Test corsika_simtel package with zstd compression</td>
<td>MC Production</td>
<td>Bregeon Johan</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Done

<table>
<thead>
<tr>
<th>#</th>
<th>Status</th>
<th>Priority</th>
<th>Subject</th>
<th>Assigned To</th>
<th>Category</th>
<th>Due date</th>
</tr>
</thead>
<tbody>
<tr>
<td>21547</td>
<td>Closed</td>
<td>Normal</td>
<td>Prod3b High NSB production - analyse with Eventdisplay</td>
<td>Bregeon Johan</td>
<td>MC Analysis</td>
<td></td>
</tr>
<tr>
<td>17833</td>
<td>Closed</td>
<td>Normal</td>
<td>Cross-check of evndisp command line options</td>
<td>Bregeon Johan</td>
<td>MC Analysis</td>
<td></td>
</tr>
<tr>
<td>17713</td>
<td>Closed</td>
<td>Normal</td>
<td>Pro3b - La Palma - Eventdisplay Analysis - SC-MSTs</td>
<td>Arrabito Luisa</td>
<td>MC Analysis</td>
<td></td>
</tr>
<tr>
<td>17208</td>
<td>Closed</td>
<td>Normal</td>
<td>Small sub-layouts, prod3b 20 deg diffuse Paranal (MARS analysis)</td>
<td>Arrabito Luisa</td>
<td>MC Analysis</td>
<td></td>
</tr>
<tr>
<td>16994</td>
<td>Closed</td>
<td>High</td>
<td>Pro3b - La Palma - Eventdisplay Analysis - 40 deg zenith angle</td>
<td>Arrabito Luisa</td>
<td>MC Analysis</td>
<td></td>
</tr>
<tr>
<td>16740</td>
<td>Closed</td>
<td>High</td>
<td>Pro3b - La Palma - MARS Analysis</td>
<td>Arrabito Luisa</td>
<td>MC Analysis</td>
<td></td>
</tr>
<tr>
<td>16669</td>
<td>Closed</td>
<td>Normal</td>
<td>Pro3b - La Palma - Eventdisplay Analysis</td>
<td>Arrabito Luisa</td>
<td>MC Analysis</td>
<td></td>
</tr>
</tbody>
</table>
La Palma Baseline reference

- (Issue #17019)
- software: corsika 2017-04-19 and evendisp prod3b_d20170602
- ConfigurationId=0
- 20 deg South pointing is Done

Paranal Baseline High NSB

- (Issue #21541) + (Issue #21547)
- corsika 2017-09-01 and evendisp prod3b_d20170922
- ConfigurationId=1
- output files have meta data to tag the NSB level: ×1, ×5, ×30
- 20 deg North and South pointing are Done.

Corsika zstd compression test

- (Issue #22246)
- software: corsika 2017-09-28-zstd ready, need an evendisp version (prod3b_d20170922 with zstd compression)
- ConfigurationId=2
- Test ok at CC-IN2P3 and DESY, asked sites to install zstd
- CPU (faster compression) and disk space (∼ 10%) gains

Paranal Prod3b 60deg Zenith angle

- (Issue #22108)
- corsika package: original Prod3(b) or similar one removing ASTRI (bad MC model)
- analysis package we’ll be needed too
- optimization steps for number of events and energy range for the different particle types will be needed
Idea: simplified setup to reproduce PROD3(b) results with more flexibility

- baseline layout 4 LSTs + 15 MSTs, NectarCam only
- run efficiently a full simulation and analysis chain: Corsika+simtel_array, EventDisplay Conversion, Calibration, Reconstruction and Analysis

La Palma Prod3(b) baseline South pointing simulation is now done, see (Issue #17019)

→ that should be the reference for all further studies varying one parameter at a time (atmosphere transparency, NSB, mirror degradation...)

- note: gamma South pointing, 20 deg Zenith production holds in 10 DL2 like evndisp files (just 2 GB overall), from which (any)one can derive the IRFs

- Gernot confirmed that results are identical to original PROD3(b)
Atmosphere related simulations: 2 requests made by the CCF group (6 months ago)

- Link to the "Atmospheric Simulation forum"
  1. many protons to start the study of the Cherenkov Transparency Coefficient
     - no progress made on my side, but people have run small productions on their own
  2. different atmosphere transmission and density profiles to study the effect of the atmosphere
     - slow progression, now need to get atmosphere transmission tables and/or test cloud option in simtel array

- Some discussions took place at the CCF meeting a few weeks ago, work needs to restart more seriously
Other specific studies

- IRFs
  - more zenith angle points for IRFs interpolations
  - study divergent pointing
  - performances with degraded telescope models

- Analysis technique
  - machine learning specific simulations (from ASWG Forum)
    - different spectrum, increased gamma diffuse statistics, "single" telescope simulations

- Architecture/version
  - moving from Corsika 6.9 to 7.5
  - moving from SL6 to CentOS7 platform (OS, libraries, gcc version)
### Disk Space Status

<table>
<thead>
<tr>
<th>Site</th>
<th>Available</th>
<th>Used</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>CYFRONET</td>
<td>0.0</td>
<td>614.5</td>
<td>628.5</td>
</tr>
<tr>
<td>DESY</td>
<td>20.5</td>
<td>1036.0</td>
<td>1056.5</td>
</tr>
<tr>
<td>CNAF</td>
<td>2.3</td>
<td>493.7</td>
<td>496.0</td>
</tr>
<tr>
<td>CC-IN2P3</td>
<td>86.4</td>
<td>346.3</td>
<td>432.7</td>
</tr>
<tr>
<td>GRIF (LPNHE+CEA)</td>
<td>9.6</td>
<td>212.2</td>
<td>232.7</td>
</tr>
<tr>
<td>LAPP</td>
<td>0.0</td>
<td>118.1</td>
<td>118.1</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>109.2</strong></td>
<td><strong>2820.8</strong></td>
<td><strong>2930.0</strong></td>
</tr>
</tbody>
</table>

- numbers from recent simulations
  - Paranal 20 deg NSB : 50.4 TB North + South pointing
  - La Palma baseline 20 deg South pointing : 23 TB
  - enough disk space for increased statistics on these simulations and/or run Atmosphere related simulations
- **PROD3(b)** Paranal 60 deg zenith angle (≈ 200 TB, free disk)
Computing needs for 2018

- Wiki page: Estimation of Computing resources for 2018
- 2017 request: 180 MHS06, we’ve used 100 MHS06 so far
- PROD4: updated telescope models and camera simulation
  - need to run Paranal and La Palma at zenith 20, 40 and 60 deg, North and South pointing
  - based on PROD3(b) Paranal + La Palma: 100 MHS06 + 1.2 PB
  - PROD4 will require 130 MHS06 and 1.6 PB (assuming no SCT)
- consider an overhead for specific studies and user activity
- consider deleting PROD2, and moving PROD3 to tape

→ 2018 request: 200 MHS06, +600 TB disk, 2 PB of tape
+ ask for access to a GPU cluster for machine learning
**Conclusions**

- **CTADIRAC**
  - used for large scale MC productions in the past 5 years
  - user portal to CTA distributed resources (10k CPUs, 3 PB)
  - proposed as a CRMS prototype, more developments expected in 2018

- **Productions**
  - 2017: PROD3(b) simulation and analysis, beginning of ctapipe, hopefully atmosphere related simulations
  - 2018: PROD4 simulations and analysis, expect more request for specific studies, expect ctapipe ramp up

- **CTACG resources**
  - 2017: 180 MH0S6 requested, 100 MHS06 consumed, only 100 TB of disk free now, but 200 TB can be deleted
  - 2018: request 200 MHS06, 600 TB of disk and 2 PB of tape