Distributed Archive System for the Cherenkov Telescope Array

Eva Sciacca, S. Gallozzi, A. Antonelli, A. Costa
INAF, Astrophysical Observatory of Catania
INAF, Astronomical Observatory of Rome
eva.sciacca@oact.inaf.it
e-Research Summer Hackfest
Outline

• CTA: What, Who & Why
• Data Requirements and Data Lifecycle
• Data Model
• CTA Archive Overview
• Proposed DEMO solution & Status
CTA: What (1)

Aims:

- Understanding the origin of cosmic rays and their role in the Universe
- Understanding the nature and variety of particle acceleration around black holes
- Searching for the ultimate nature of matter and physics beyond the Standard Model

The CTA project is an initiative to build the world’s largest and most sensitive ground-based high-energy gamma-ray observatory.
The Cherenkov Telescope Array (CTA) will explore our Universe in depth in Very High Energy (VHE, $E > 10$ GeV to above $100$ TeV) gamma-rays.

- ~20 telescopes for the North-site (Canarie)
- ~100 telescopes for the South-site (Chile)

An expected RAW data rate of **0.4-5.3 GB/s** for both sites.

CTA will contain ~120 telescopes of three different sizes (Small / Medium / Large):

- SST($\varnothing \sim 4$m) $E \approx 10$ TeV, FoV~10deg
- MST($\varnothing \sim 12$m) $E \sim [0.1;1]$ TeV, FoV~6-8deg
- LST($\varnothing \sim 24$m) $E < 0.1$ TeV, FoV~4-5deg
CTA: Who & Why

CTA is an ESFRI Project

European Strategy Forum on Research Infrastructures

31 countries, ~1200 participants, ~180 institutes, ~400 FTE

A deeper sensitivity (~ factor 10) on a wider band ([20;30]GeV to 100 TeV) more than 1000 new gamma ray sources are foreseen to be discovered.
Data Requirements

Without data compression and assuming 165 operational nights/yr:

ASTRI/Prot. → ~0.8 TB/night → ~0.3 PB/year

Mini-Array → ~3 TB/night → ~6.1 TB/night A.R. → ~1.0 PB/year A.R.

CTA → ~8.5 GB/s → ~40 TB/night → ~4 PB/year → ~20 PB/year

A.R. (A.R. = After Reduction → input+processed data including calibs, intermediate reduction and MC simulation data) this is the OPTIMISTIC SCENARIO

The pessimistic one can take ~>100PB/year !

The CTA Archive system must store, manage, preserve and provide easy access (IVOA Access) to a such huge amount of data for a long time.
## CTA Data Levels

<table>
<thead>
<tr>
<th>Data Level</th>
<th>Short Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 0 (DL0)</td>
<td>DAQ-RAW</td>
<td>Data from the Data Acquisition hardware/software.</td>
</tr>
<tr>
<td>Level 1 (DL1)</td>
<td>CALIBRATED</td>
<td>Physical quantities measured in each separate camera: photons, arrival times, etc., and per telescope parameters derived from those quantities.</td>
</tr>
<tr>
<td>Level 2 (DL2)</td>
<td>RECONSTRUCTED</td>
<td>Reconstructed shower parameters (per event, no longer per telescope) such as energy, direction, particle ID, and related signal discrimination parameters.</td>
</tr>
<tr>
<td>Level 3 (DL3)</td>
<td>REDUCED</td>
<td>Sets of selected (e.g. gammaray- candidate) events, along with associated instrumental response characterizations and any technical data needed for science analysis.</td>
</tr>
<tr>
<td>Level 4 (DL4)</td>
<td>SCIENCE</td>
<td>High Level binned data products like spectra, sky maps, or light curves.</td>
</tr>
<tr>
<td>Level 5 (DL5)</td>
<td>OBSERVATORY</td>
<td>Legacy observatory data, such as CTA survey sky maps or the CTA source catalog.</td>
</tr>
</tbody>
</table>
Data Life Cycle
The CTA Archive organization is based on OAIS (Open Archive Information System) standards.

- Databases (Data & Metadata)
- Hardware (storage devices, support Infras, ...)
- Software (ingest, download, browse&management)
- Data Access (database, DBMS, queries, ...)

Off-site archive

Catania, 06/07/2016
Archive Framework

- **INGEST** unit involves a collection of software and/or middleware able to receive bulk data of different types coming from the array and to prepare them for storage, performing basic operation like data indexing, dependencies and compression.

- **STORAGE** guarantees the efficient retrieval of ingested data, and providing simple archive hierarchy management and maintenance. Storage also supervises the status of the media used in the archive, providing a guarantee of error control and data security.

- **ADMINISTRATION** unit deals with all the operations related to the CTA archive system and its management. It will assure archive performance and standards/requirements fulfillment by means of dedicated monitoring functionality and recover of failures.

- **ACCESS** unit consists of a collection of software and on-line services that provide efficient access to the data to the other CTA components (e.g. the data processing pipelines). Furthermore, it will make CTA users able to access CTA data accordingly to their specific data access privileges.

- **DATA TRANSFER** unit will guarantee the transfer of data and data products between the on-site and the off-site zone of the archive system.
Archive System Architecture

- For **ASTRI prototype** → **Single Mirrored Archive System** for the off-site Archive is enough solution for the management of the whole data-flow chain. The AS has to be mirrored.

- For **ASTRI-Miniarray project** → **Distributed Archive System** is recommended in order to test all relevant technologies for the whole CTA Archive system. A demo data-model has been identified and will be produced within the **E.U. INDIGO-DataCloud collaboration**.

- For **CTA Observatory** → **Distributed Archive System** for the off-site Archive is the only solution for the management of such a large amount of data and can be deployed by the federation of different storage entities.
Archive Requirements

- **Bandwidth**: an efficient, fast, high throughput and low-latency connection.
- **Computing**: an efficient and managed queuing system to run and manage lots of jobs.
- **Storage**: easily horizontal-scalable storage resources, for long term to be silent-corruption free.
- **Easy Access**: to a very large amount of data and meta-data for different users and use-cases.

Linear scalability request for such a great amount of data (O(PB)/year) and the necessity to distribute geographically the storage are the most challenging request for a “centralized archive” approach.
Distributed Archive Advantages

• lower cost respect to a single huge data centre,
• easy manageability & maintenance by single-site human resources
• distributed database of meta-data within the architecture
• easily scalability by adding new nodes on the system
• disaster recovery free by different sites redundancy
Currently **exporting DATA-File relational table to CouchBase DB** in order to export searchable keywords in a dedicated **OneData embedded solution**!
## Sample Data Model

<table>
<thead>
<tr>
<th>Field</th>
<th>Type</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>PFNP</td>
<td>varchar(255)</td>
<td>/archive/ASTRI/PHYSICAL/41/2016_03_08/DL0/000</td>
</tr>
<tr>
<td>PFN</td>
<td>varchar(255)</td>
<td>astri_000_41_001_00001_R_000005_000_1002.lv0.gz</td>
</tr>
<tr>
<td>ORIGIN_ID</td>
<td>varchar(2)</td>
<td>41</td>
</tr>
<tr>
<td>PROGRAM_ID</td>
<td>varchar(3)</td>
<td>001</td>
</tr>
<tr>
<td>FILENAME</td>
<td>varchar(255)</td>
<td>astri_000_41_001_00001_R_000005_000_1002.lv0</td>
</tr>
<tr>
<td>FILEVERS</td>
<td>int(16)</td>
<td>1</td>
</tr>
<tr>
<td>PROP_ID</td>
<td>varchar(16)</td>
<td></td>
</tr>
<tr>
<td>USER_ID</td>
<td>varchar(16)</td>
<td></td>
</tr>
<tr>
<td>SHW_CONF</td>
<td>varchar(255)</td>
<td></td>
</tr>
<tr>
<td>DET_CONF</td>
<td>varchar(255)</td>
<td></td>
</tr>
<tr>
<td>PARTICLE</td>
<td>varchar(255)</td>
<td></td>
</tr>
</tbody>
</table>

RAW data keywords
Scheduler Keywords
Monte Carlo keywords
FITS Data keywords
Searchable keywords
Implementation Plan - 1

1st realization (prototype phase):

1) create a set of geographically distributed resources (INAF & INFN sites) populated by few TBs of simulated and real datasets from CTA prototypes (i.e. ASTRI and miniArray)

2) federate together these resources using Indigo tools (e.g. OneData provider)

3) develop a database schemaless description of data and a suitable interface to query and retrieve data from a OneData Client → export CTA Data Model to CouchBase DB embedded in OneData

4) test INDIGO solutions for A&A and I/O with federated & distributed storage

5) implementing other tools (related to visualization and management of providers)
Implementation Plan - 2

2nd realization (pre-production phase):

1) add a GRID infrastructure to a working archive prototype →
   A. migrate the INAF-site repositories to GRID enabled sites (testing at INFN-LNF) creating dedicated Storage Element to CTA V.O.
   B. enable GRID computations to CTA VO and redirect CTA storage in the dedicated local resources
   C. again federate together using the prototype archive setup

2) expand computation and storage in order to be compliant for the preproduction of CTA

3) enable access from CTA Tools and CTA Science Gateway and Proposal Handling Platform.
Prototype Development

For the purpose of this event we have prepared a test environment on the local machine:

- Set up VirtualBox VM with CentOS 7
- Installed Docker 1.11
- Installed Docker-Compose 1.7
- Create a sample deployment using Docker Compose
  - Use GitHub scenarios at https://github.com/onedata/getting-started.git.
  - Run Scenario 2.0: pre-configured Oneprovider with pre-configured Onezone.
- Ingest a CTA sample dataset.
- Configure a sample Data Model on the CouchBase DB.
- Query the sample dataset (based on searchable keywords).
Questions?