INAF
ISTITUTO NAZIONALE
DI ASTROFISICA
NATIONAL INSTITUTE
FOR ASTROPHYSICS

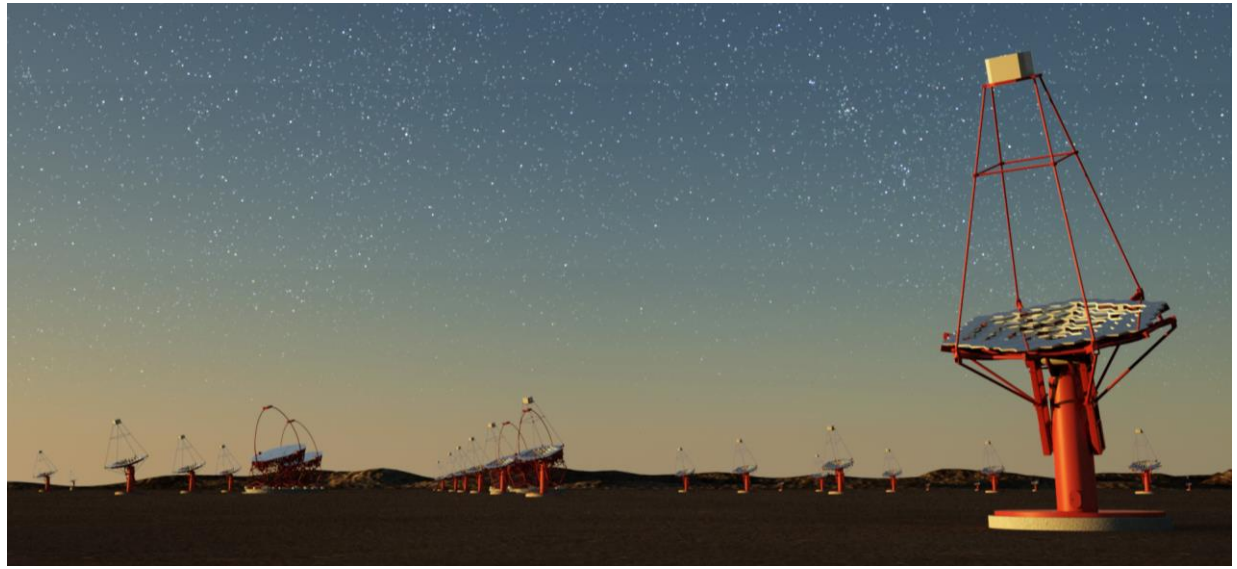
Distributed Archive System for the Cherenkov Telescope Array

Eva Sciacca[1], *Stefano Gallozzi*[2], Angelo Antonelli[2,3], Alessandro Costa[1]

[1]INAF, Astrophysical Observatory of Catania
[2]INAF, Astronomical Observatory of Rome
[3]ASDC, ASI-Science Data Center

Cherenkov Telescope Array

https://cta-observatory.org



WHAT: CTA is the worldwide project for the future of Very High Energy gamma-ray astronomy.
~20 telescopes for the North-site (Canarie) ~100 telescopes for the South-site (Chile)

WHO: the CTA Consortium consists of more than 1,200 scientists and engineers from 32 countries from 5 continents and has become a truly global (ESFRI) project.

OUR AIM: One of the major technological challenge is related to the data-handling and archiving of the huge amount of data (from 20 to 100 PB/year) coming from the observatory facilities.

CTA Data Model

Data Level	Short Name	Description
DL0	DAQ-RAW	Acquired raw data.
DL1	CALIBRATED	Calibrated camera data.
DL2	RECONSTRUCTED	Reconstructed shower parameters (such as energy, direction, particle ID).
DL3	REDUCED	Sets of selected events with associated instrumental response characterizations needed for science analysis.
DL4	SCIENCE	High Level binned data products (such as spectra, sky maps, or light curves).
DL5	OBSERVATORY	Legacy observatory data (such as survey sky maps or source catalog).

Data Requirements

Without data compression and assuming 165 operational nights/yr:

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

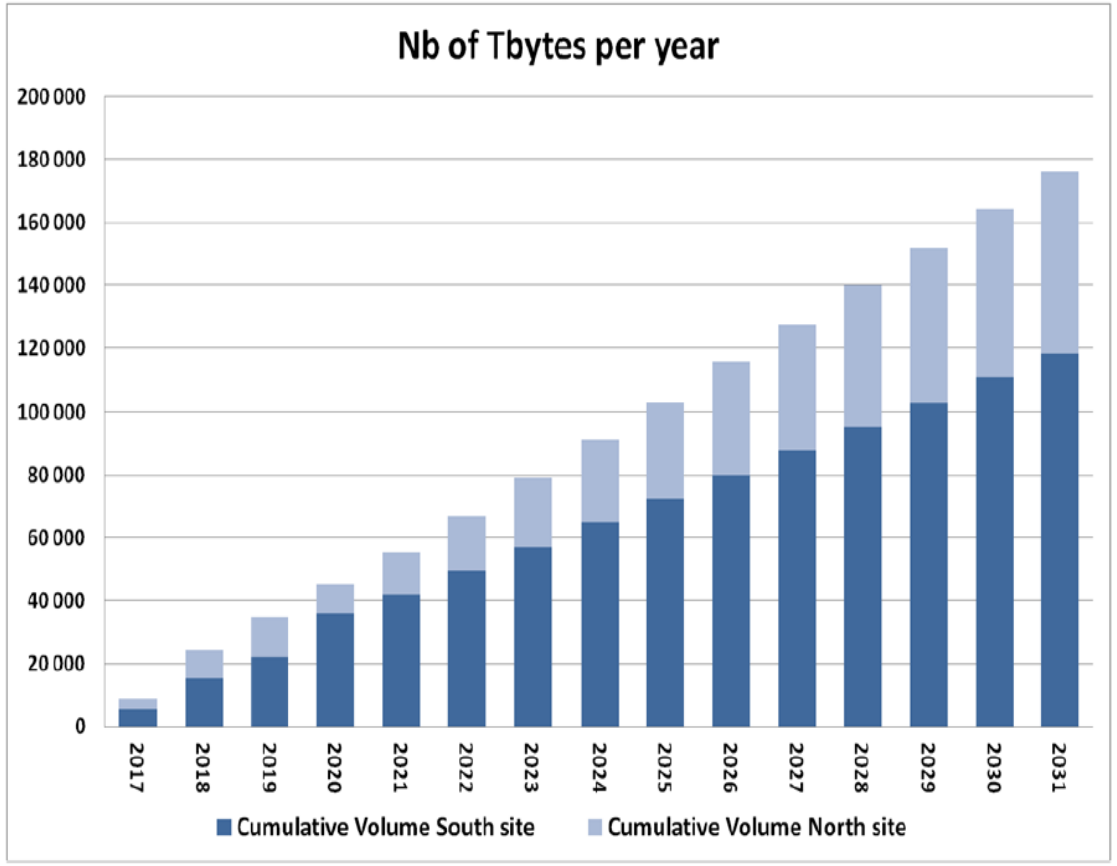
Mini-Array
→ ~3 TB/night
→ ~6.1 TB/night A.R.
→ **~1.0 PB/year** A.R.

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)

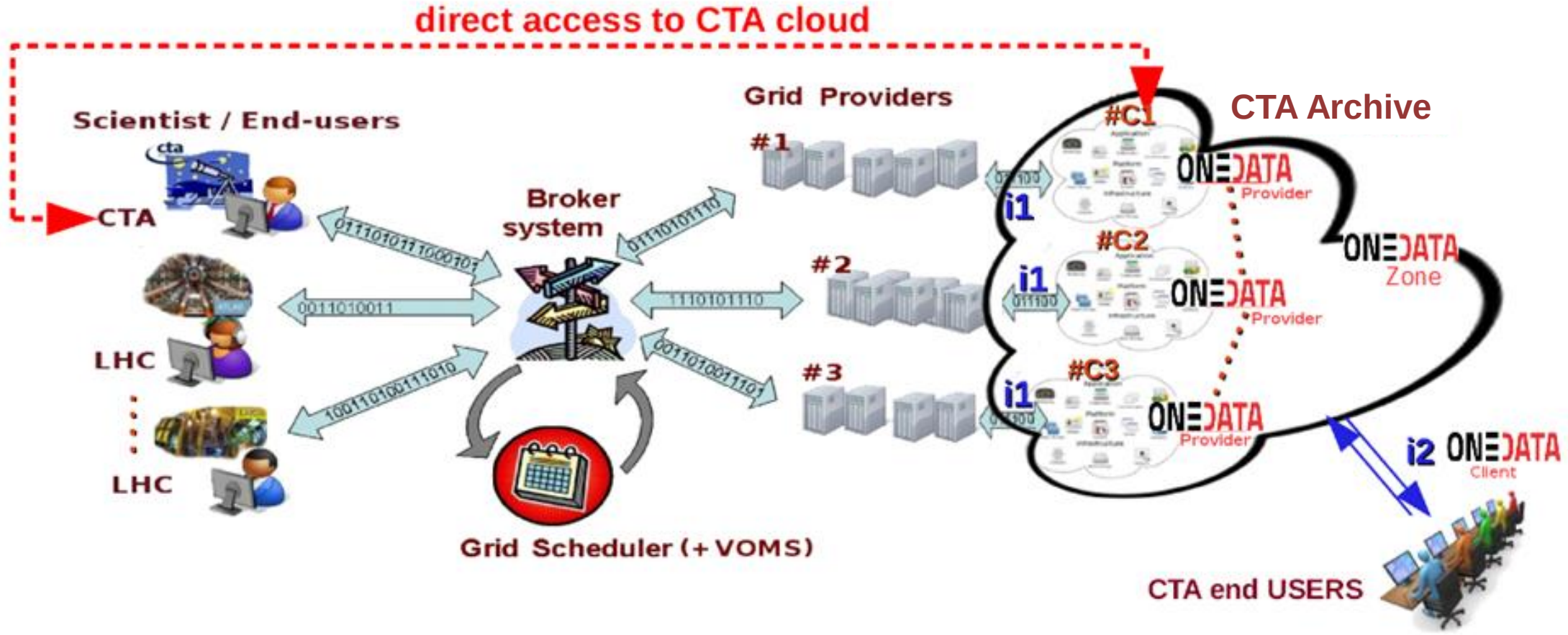
this is the **OPTIMISTIC SCENARIO**

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



The CTA Archive system must store, manage, preserve and provide easy access to such huge amount of data for a long time.

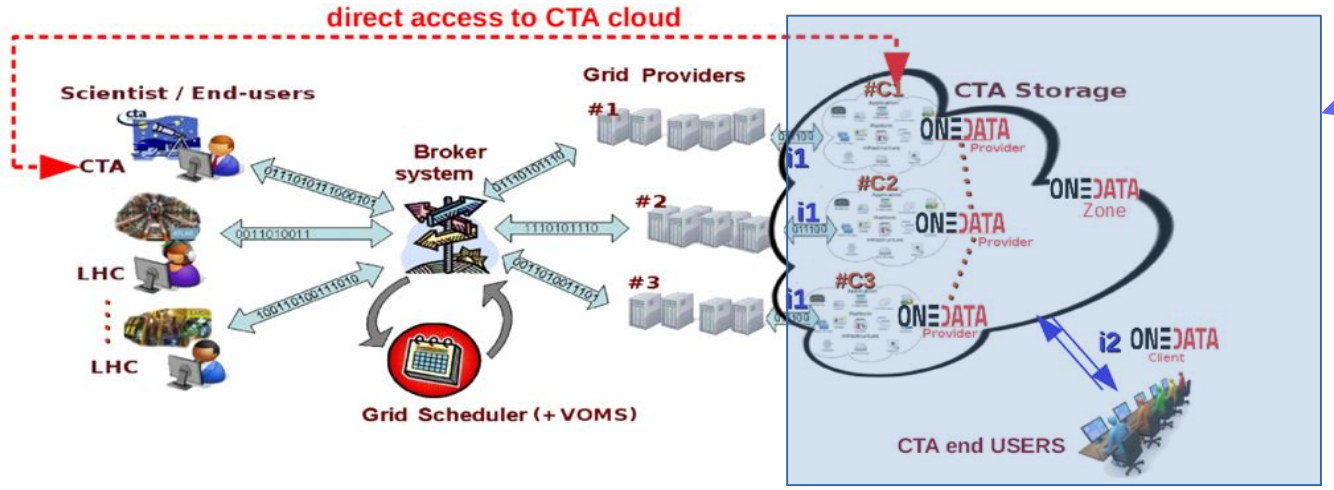
CTA Archive Prototype



CTA Archive Prototype

CTA Collaboration & Community participate to the INDIGO-Data Cloud H2020 Project AS **"Use Case"** for the **INDIGO infrastructure**.

The aim of our commitment was the very fruitful multi-disciplinary collaboration with INDIGO Communities in order to include the BigData challenges coming from the CTA Archive as an INTERNAL INDIGO Use CASE / Case Study
→ to be investigated with a distributed approach ←



INDIGO solution effort
In the Distributed
Federation of Storage

INDIGO - DataCloud

SOFTWARE → STARTING POINT

OneData solutions are ready for *CTA A&A*

Running the Tests

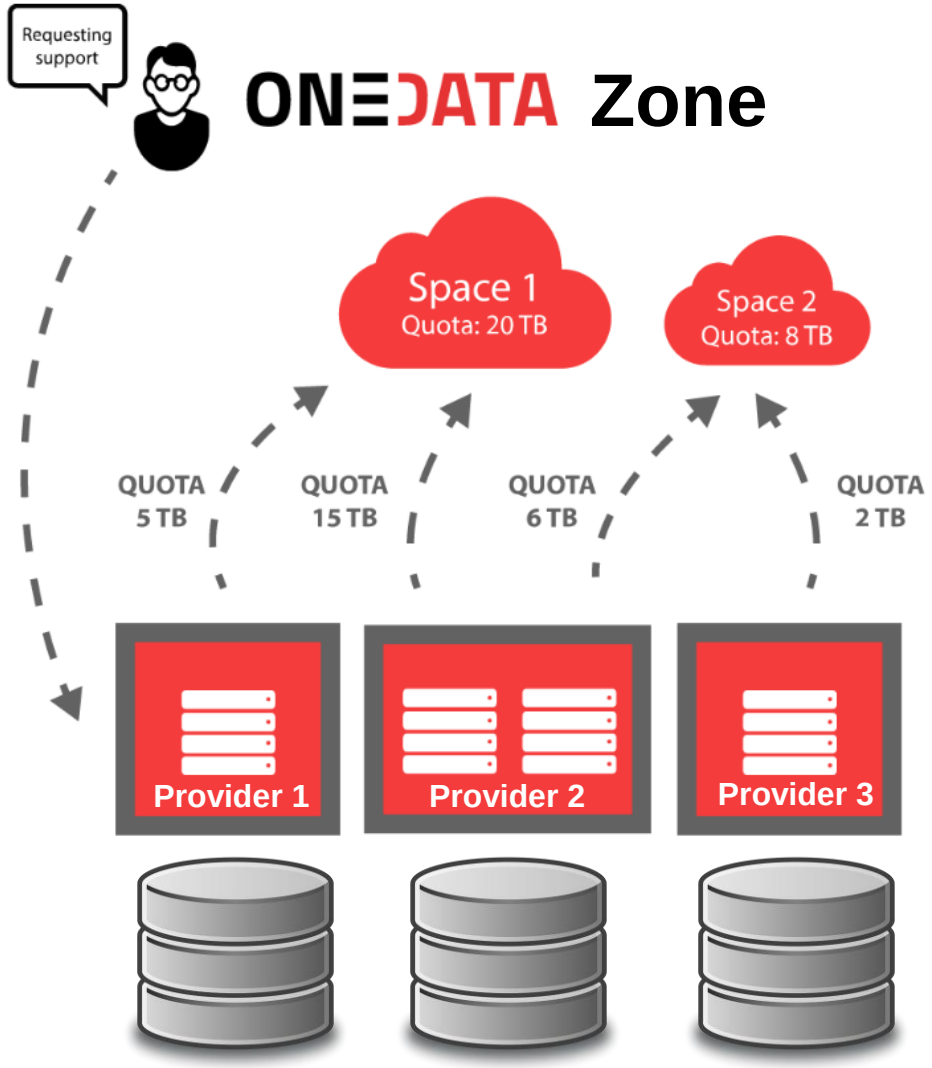
- The test infrastructure has been setup using **VirtualBox Virtual Machines** and **Docker containers**.
- Demo **datasets** coming from the **ASTRI project** are uploaded to the **CTA OneZone** within a **space** supported by the two providers.
- The ingested data are enriched with **Metadata** thanks to the **Cloud Data Management Interface** (CDMI) or, alternatively, the REST API can be used.
- **Metadata queries** are performed using **REST-API** and **indexing functions** (associated to the Space) on pre-defined extended attributes (Metadata).
- The **CouchBase** database (embedded in OneData) can be used alternatively to query and retrieve the metadata using **Elastic Search engines** (e.g. N1QL) or common **MapReduce functions** using the standard CouchBase console and the SDK from the client side. This will enable versatile access to the whole CTA dataset to higher level application frameworks and end-users analysis tools.

OneData Overview

OneData system **virtualizes** storage systems provided by storage resource providers **distributed** globally.

The most important concepts of the platform are:

- **Spaces** - distributed virtual volumes, where users can organize their data
- **Providers** - entities who support spaces with actual storage resources
- **Zones** - federations of providers which enable creation of closed or interconnected communities.

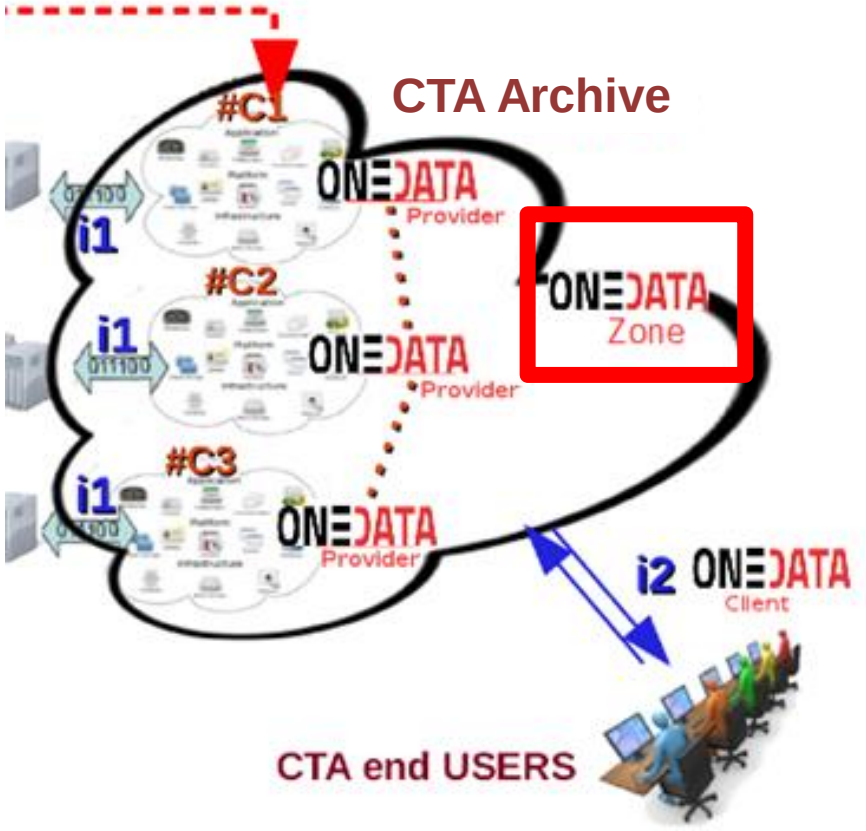


CTA OneZone

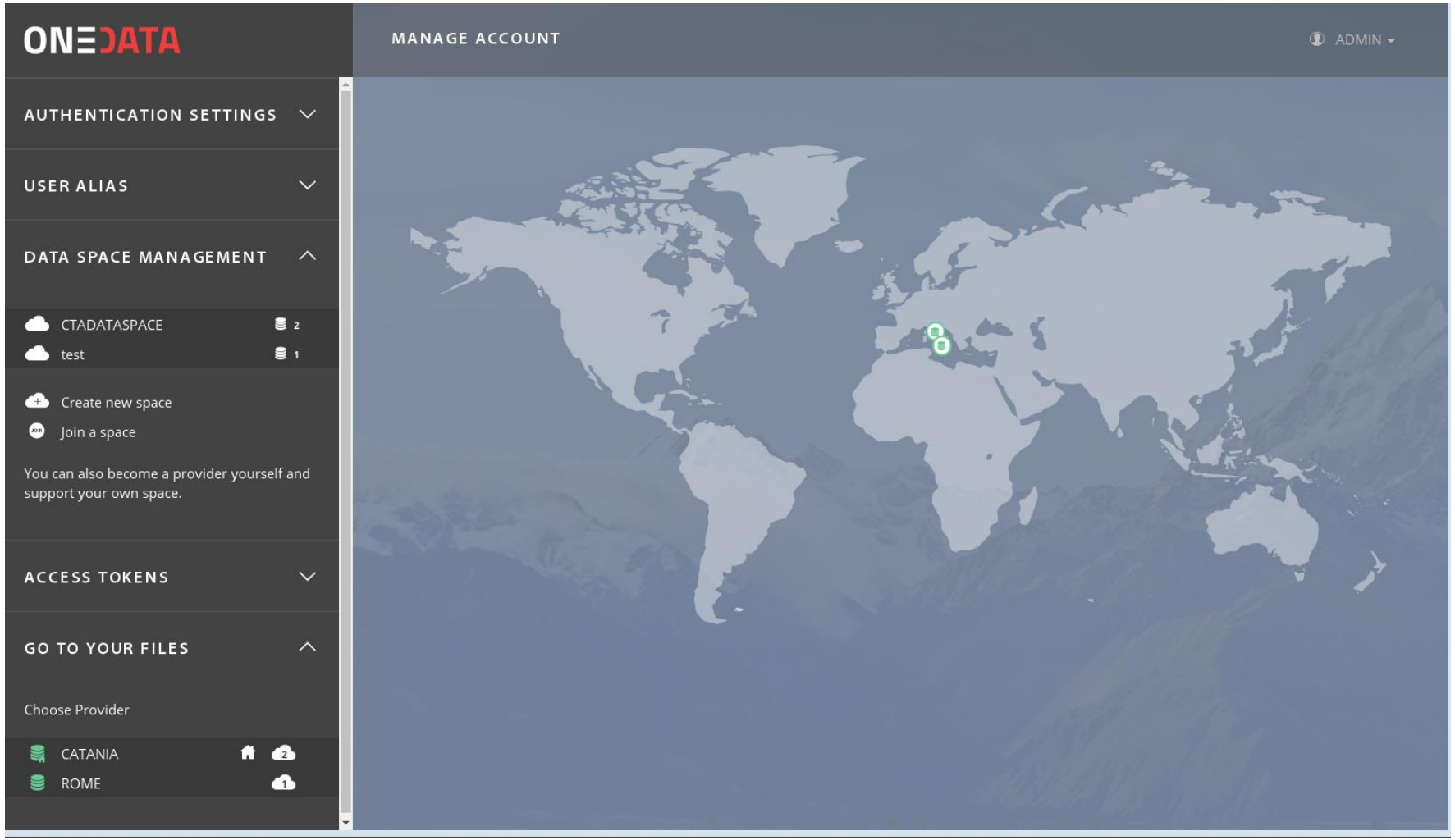
OneZone is the gateway for users to the OneData system. It is responsible for connecting to the **authentication** and **authorization** infrastructure.

It allows users to:

- ✓ create **user spaces**
- ✓ generate space **support tokens**, that can be used to support user spaces with storage from a dedicated storage provider
- ✓ **monitor availability** of storage providers that support user spaces
- ✓ see the geographical distribution of storage providers
- ✓ choose storage provider for spaces

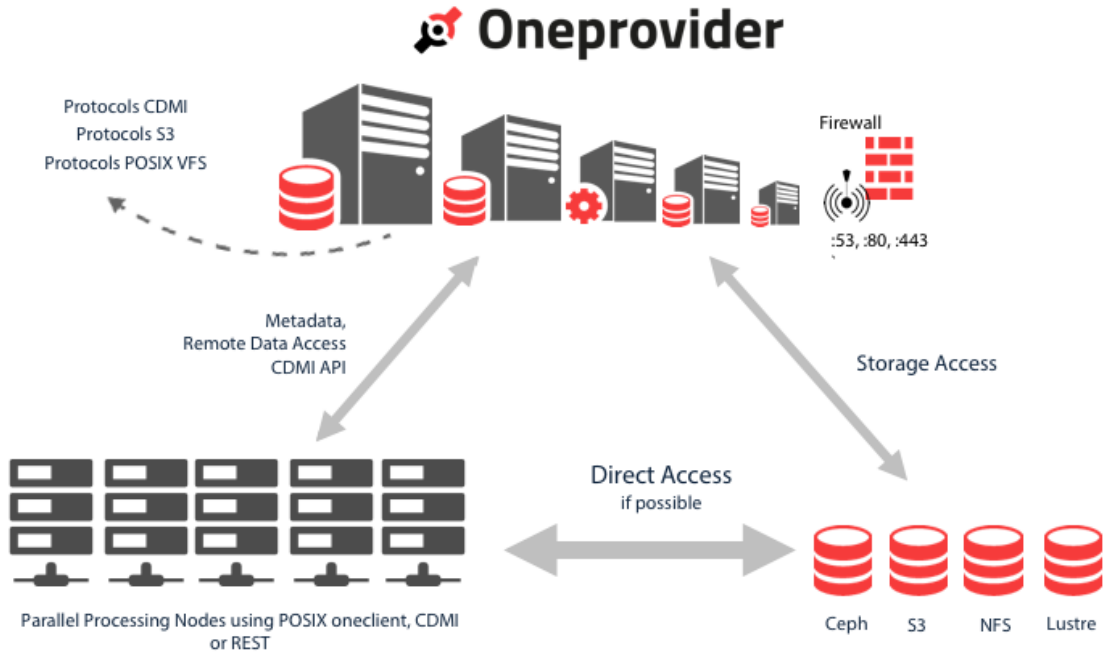


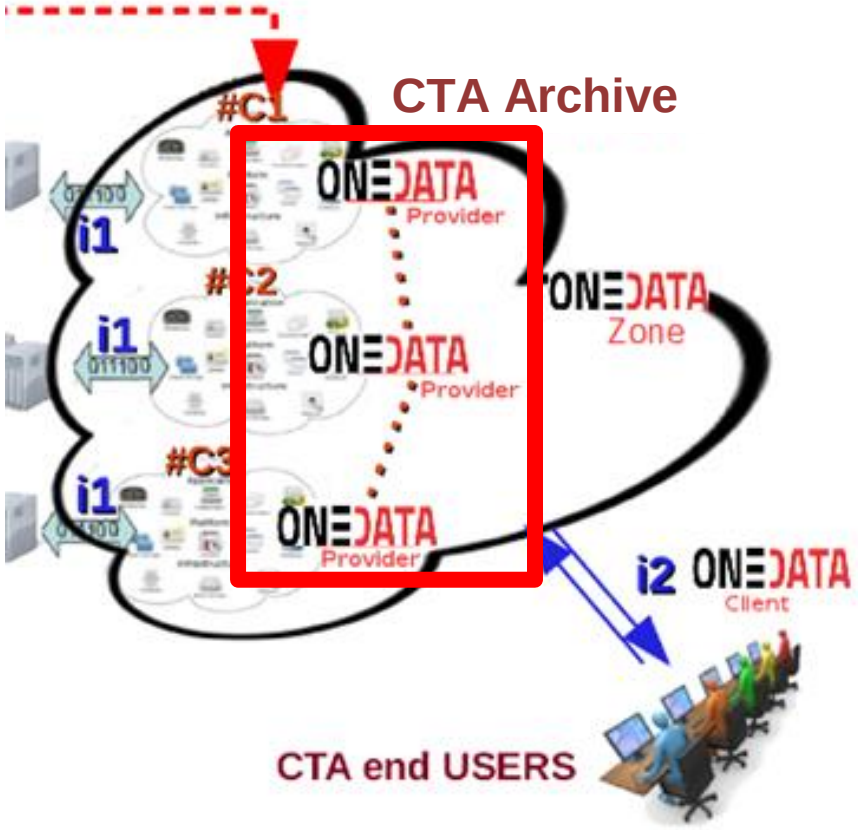
CTA OneZone


ONEDATA
MANAGE ACCOUNT
ADMIN
AUTHENTICATION SETTINGS
USER ALIAS
DATA SPACE MANAGEMENT
CTADATASPACE
2
test
1
Create new space
Join a space
You can also become a provider yourself and support your own space.
ACCESS TOKENS
GO TO YOUR FILES
Choose Provider
CATANIA
2
ROME
1
Spaces
Providers

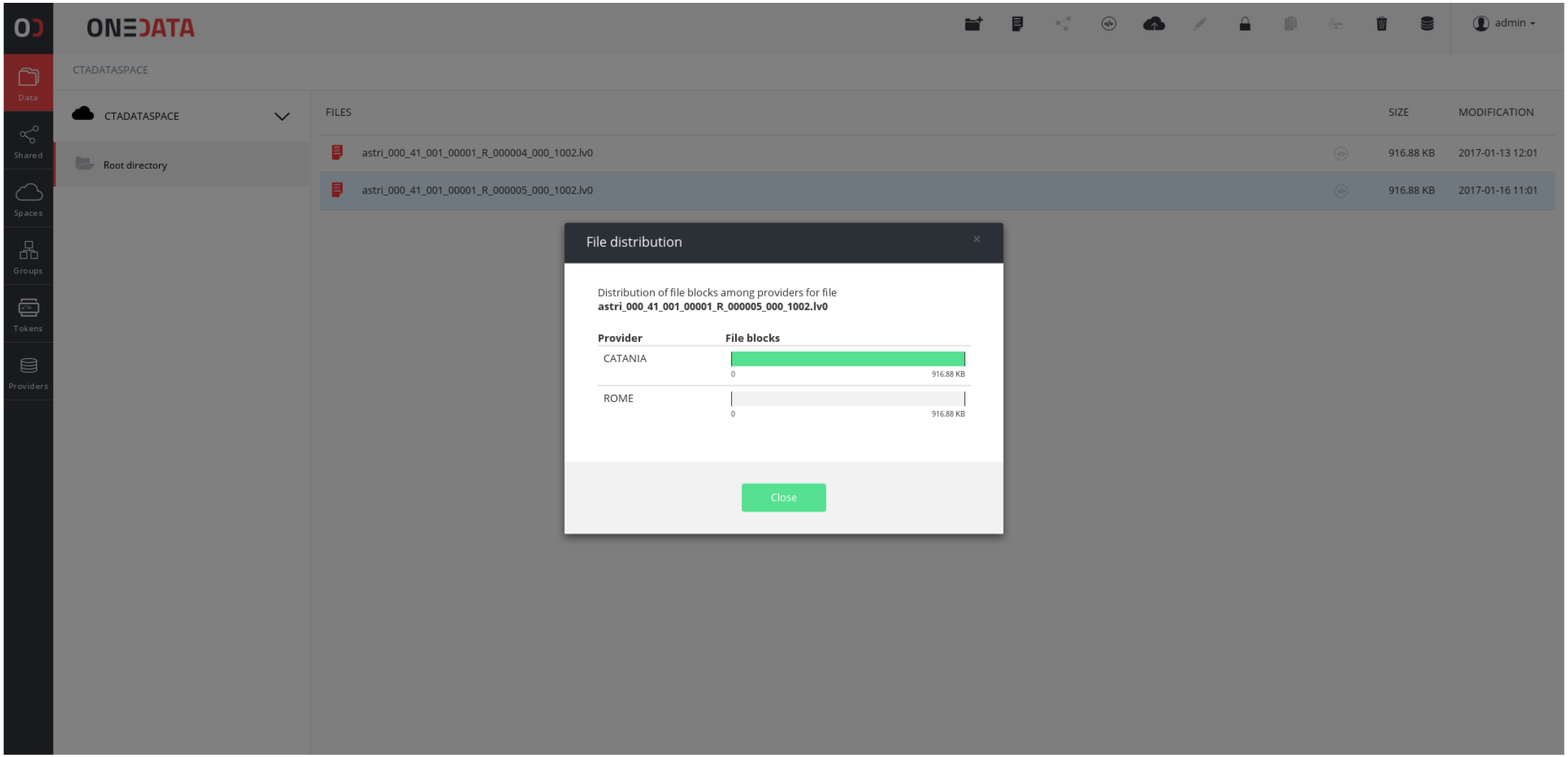
CTA OneProvider(s)

OneProvider exposes storage resources. It is deployed in a data or computing center, on the nodes equipped with high speed **connections** to **storage resources**.





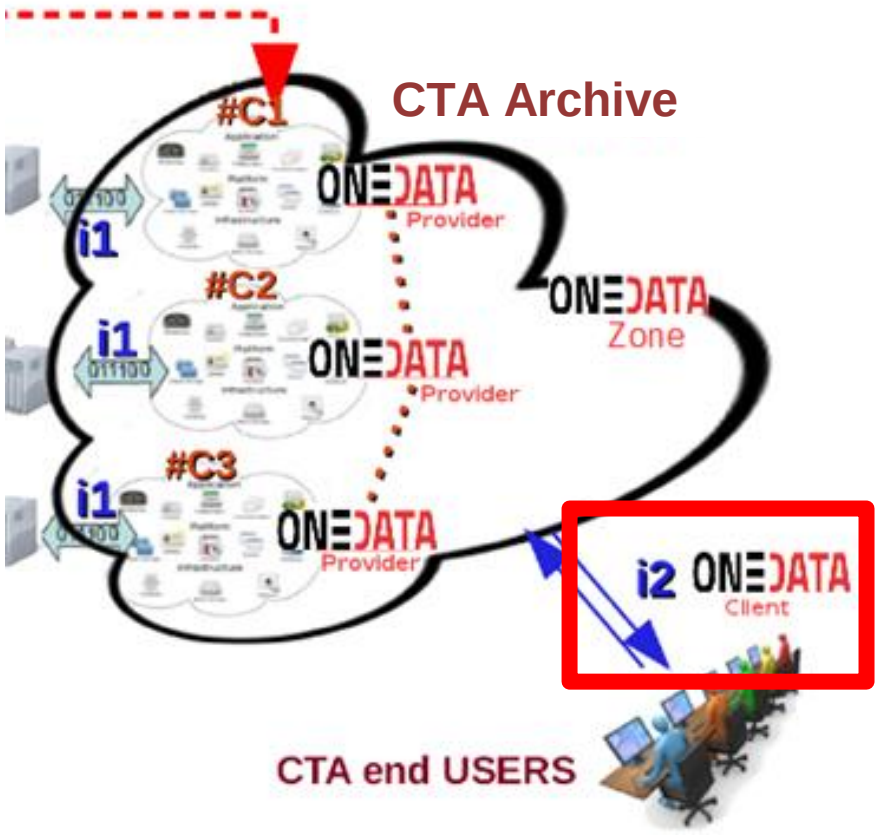
CTA OneProvider(s)



CTA OneClient

OneClient is a command-line based application for **accessing** and **managing** user spaces via **virtual file system**.

User spaces are **mounted** in the local file system tree.



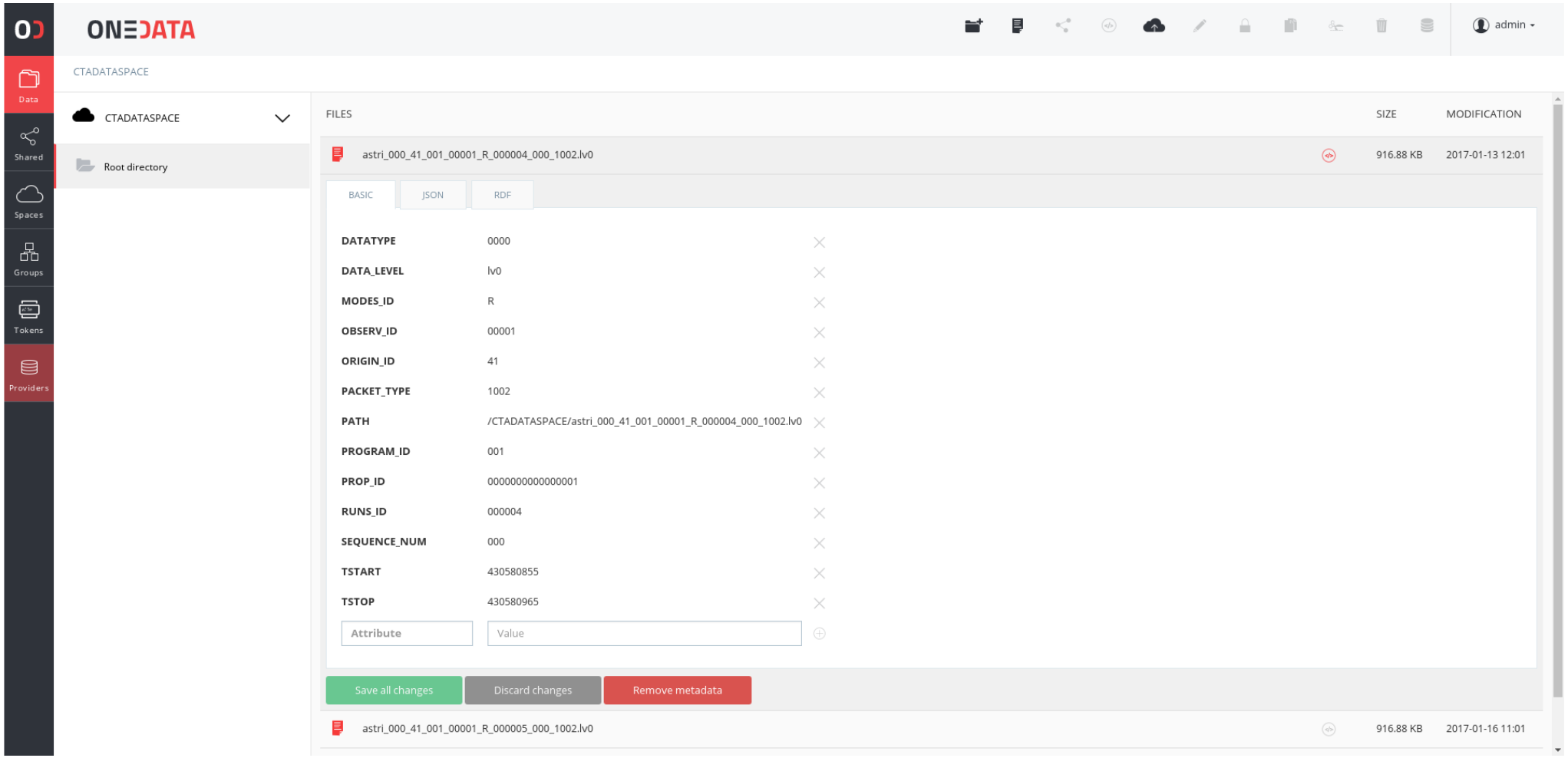
Metadata

Metadata in OneData are organized into 3 levels:

- ✓ **Filesystem attributes** - basic metadata related to file system operations such as file size, creation and modification timestamps, POSIX access rights, etc.,
- ✓ **Extended attributes** - these attributes enable assigning custom key-value pairs.
- ✓ **User metadata** - this level provides most flexibility and OneData itself does not assume any schema related with these metadata. For each resource, user can assign a separate document in one of supported metadata formats (currently JSON and RDF).

The filesystem and extended level attributes are accessible via **REST-API**, **CDMI** and the GUI or directly through queries to the embedded database.

Metadata


ONEDATA
admin
Data
Shared
Spaces
Groups
Tokens
Providers
CTADATASPACE
CTADATASPACE
Root directory
FILES
SIZE
MODIFICATION
astri_000_41_001_00001_R_000004_000_1002.lv0
916.88 KB
2017-01-13 12:01
BASIC
JSON
RDF
DATATYPE 0000
DATA_LEVEL lv0
MODES_ID R
OBSERV_ID 00001
ORIGIN_ID 41
PACKET_TYPE 1002
PATH /CTADATASPACE/astri_000_41_001_00001_R_000004_000_1002.lv0
PROGRAM_ID 001
PROP_ID 0000000000000001
RUNS_ID 000004
SEQUENCE_NUM 000
TSTART 430580855
TSTOP 430580965
Attribute
Value
Save all changes
Discard changes
Remove metadata
astri_000_41_001_00001_R_000005_000_1002.lv0
916.88 KB
2017-01-16 11:01

Metadata

Sample Ingestion

```
curl -k -H $TOKEN_HEADER -H $CDMI_VSN_HEADER -H 'Content-Type:
application/cdmi-object'
-d '{"metadata" : {"PROGRAM_ID" : "001"}}' -X PUT "$ENDPOINTDATA"
```

Sample indexing function

```
function(meta) {
    if(meta['PROGRAM_ID']) {
        return meta['PROGRAM_ID'];
    }
    return null;
}
```

Query using a REST-API call

```
curl -v -k --tlsv1.2 -Ss -H "X-Auth-Token: $TOKEN" \
-X GET "https://$HOST:8443/api/v3/oneprovider/query-
index/$INDEX_ID?key=\"0001\"&stale=false"
```

Distributed Archive Advantages

- The distributed architecture allows to **lower costs** with respect to a single huge data center including **easy manageability** and **maintenance**.
- The solution takes care of **redundancy policy**: involved databases of metadata are distributed together with the storage sites allowing a very high throughput and availability of inter-communications with the best data-model scheduling organization.
- It is **fault tolerant** and **risk-management free**: it has no single point of failure and can easily solve any disaster recovery event thanks to the redundancy of the distributed approach with a robust database management system.

Issues and Future Works

- A more stable version of OneData is needed for the full CTA Archive production.
- Improve Metadata query: possibility to performe more complex queries.
- Test OneData **roles** and data **permissions** (through connection with an Authentication and Authorization Infrastructure) and test of the **replication policies** between providers.
- **Prototype deployment** of the CTA Archive in 3 sites (INAF-Catania, INAF-Rome, ASDC) to enable CTA users to test it.
- Prototype deployment with **Data-Grid functionalities** for CTA specific users (simulation & pipelines)
- A look forward to **Cloud-Services** to be ready for CTA Workload Management System (DIRAC) migration from the DataGrid Environment to the Cloud Paradigm.

References

- CTA web page: http://www.cta-observatory.org/
- ASTRI web page: http://www.brera.inaf.it/astri/
- YouTube demo: https://youtu.be/UhOWnJluIgE
- OneData documentation: https://onedata.org/docs/index.html
- OneData @ docker hub: https://hub.docker.com/u/onedata/