Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

The GAIA use case in Spoke 3 and Innovation Grants

Enrico Licata on behalf of the INAF team and
Leonardo / Cherrydata team

INAF ISTITUTO NAZIONALE DI ASTROFISICA

esa gaia

LEONARDO

ALTEC

CHERRY DATA

Agenzia Spaziale Italiana

Archives and Data Management Systems in the Big Data Era
February 26 – 28, 2025
CNR Bologna- Via Piero Gobetti 101, Bologna (BO)

The Team

ALTEC

Lorenzo Bramante
Ruben De March
Daniele Gontero
Rosario Messineo
Luigi Squillante
Leonardo Tolomei

INAF ISTITUTO NAZIONALE DI ASTROFISICA

Deborah Busonero
Giacomo Coran
Massimo Costantini
Mariateresa Crosta
Lorenzo Filippello
Sara Gelsumini
Cristina Knapic
Mario G. Lattanzi
Enrico Licata

CHERRY DATA

Chiara Francalanci
Paolo Giacomazzi

LEONARDO

Carolina Berucci

Filippo Balla
Gennaro Chiorazzo
Fabrizio Lupi
Daniel Procopio
Sonia Regis

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

Spoke 3 - WP4 IDL and IGUC

Interoperability Data Lake for the Gaia Use Case

The purpose of Spoke 3 is the development of innovative applications and software capable of fully exploiting cutting-edge HPC technologies and big data storage solutions, to achieve excellence in the areas of astronomy, high-energy astrophysics, astroparticle physics, and cosmology.

WP4 builds upon best practices and already implemented frameworks for managing data and software with FAIR and Open Science principles, to develop innovative frameworks capable of addressing the Big Data Challenge

The IGUC innovation grant originates within the National Center for HPC, Big Data and Quantum Computing and is developed through a joint collaboration between INAF and Leonardo S.p.A. to study various technological solutions, such as DMS and the use of alternative DBMS, to manage, store, and access big data for the GAIA use case.

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

Objectives

Technological testing of various DBMS and Data Management Systems, starting from the GAIA use case, with the objective of estimating and comparing the performance of the different systems, in a way that is as hardware and scale invariant as possible

OPS4@DPCT

- Oracle DBMS
- ZFS Filesystem
- HDF5 File format
- Oracle ODAx8

Spoke 3 - WP4 IDL

- Postgres DBMS,
- Rucio DMS,
- HDF5 File format,
- INFN Data Lake machines

IGUC - Leonardo/Cherrydata

- AyraDB,
- ext4 Filesystem,
- HDF5 File format,
- INAF infrastructure (bare metal and virtualized)

GAIA Test Case: Dataset 1 - Cone search + Meridian

Given a direction defined by (α,δ) and a radius ε, we have that a generic source of coordinates (α′,δ′) is inside the cone search if:

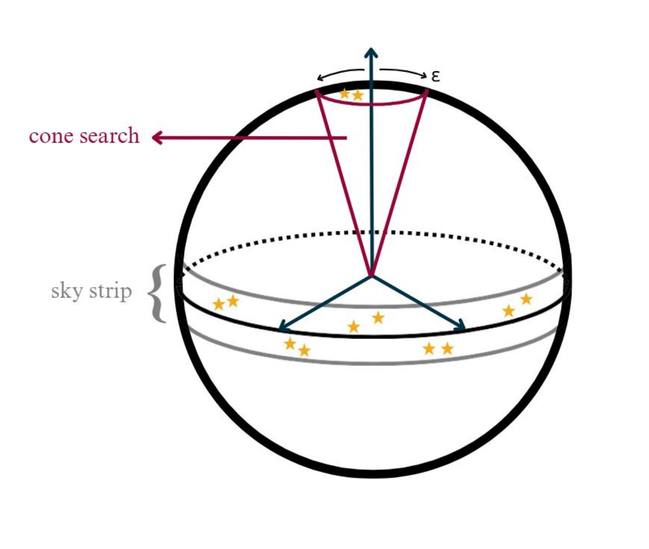
$$cos(\theta) \geq cos(\epsilon)$$

where

$$cos(\theta) = [cos(\delta)\, cos(\delta')\, cos(\alpha' - \alpha) + sin(\delta)\, sin(\delta')]$$

While it falls within the plane of semi-width ε and perpendicular to the direction (α,δ) if:

$$- sin(\epsilon) \leq cos(\theta) \leq sin(\epsilon)$$



Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

GAIA Test Case: Dataset 1 - Cone search + Meridian

Select all sources and related transits for the specified regions of space and the specified timeframe

Search details:

- Cone search direction:
  - α = 0 [rad],
  - δ = PI/4 [rad] = 45 (deg)
- Cone radius & semi-width of meridian band:
  - ε = 0,002182 [rad] = 1/8 (deg)
- TransitID range:
  - Start = 64151930880000000
    Revolutions 4640.62,
    UTC 2016-12-31T23:56:32.680453840
  - End = 65866106880131071
    Revolutions 4764.62,
    UTC 2017-01-31T23:56:31.676574736

Search results:

- Identified ~ $4.5*10^6$ Sources

In order to perform a first test on real data we chose to limit the dataset size to ~ 1TB. To achieve this, we had to limit the timeframe for the transits to 1 month of data (over 10 years of mission)

We now have a dataset of around 1.3TB of Gbins

- ~100 GB of CompleteSource
- ~100 GB of Match
- ~1.1 TB of AstroElementary

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

GAIA Test Case: Dataset 2 - 20k Cone searches

Select all sources and related transits for the specified regions of space over the entire mission

Search details:

- Cone search direction:
  - α_k, δ_k = 1 of 20.000 directions equally spaced along an homogeneous spiral from celestial north pole
- Cone radius
  - $\varepsilon = 4{,}71 \cdot 10^{-4}$ [rad] ~ 96,7 arcsec
- TransitID range:
  - all available mission data

Expected search results:

- Identified ~ $2.0 \cdot 10^{6}$ Sources

This dataset is still not available, since the volume of the returned data will be well beyond the current scope of the project.

The number of gbins selected by this dataset and their total volume is overestimated due to some of the issues inherent in the gbin data format and their organization

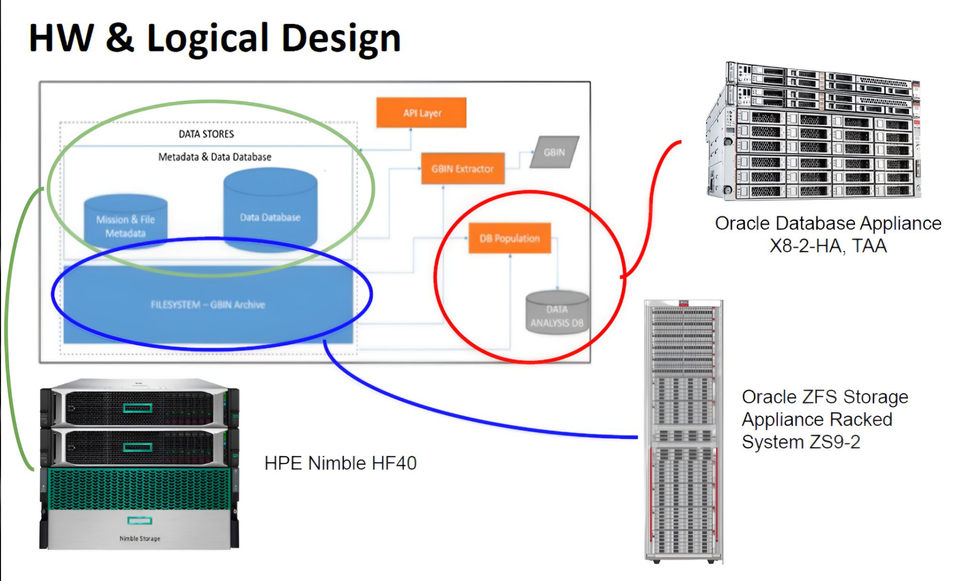
More on this in the following slides

OPS4 @ DPCT - Oracle + ZFS

The legacy project transitions from operational real-time data processing (OLTP) to a hybrid data management approach (OLAP + FS)

New DMS based on the interaction between Data Lake (GBIN format) managed by ZFS and two different DBs hosting metadata and mission metadata implementing a FAIR paradigm.

The submission of query/Data request is managed via an API layer, and outputs are delivered as files (GBINS / FITS / HDF5) or are used to populate a specific Data Analysis DB hosted on the ODA



For more details refer to the presentation: "Gaia Legacy" by Deborah Busonero

Finanziato dall'Unione europea NextGenerationEU | Ministero dell'Università e della Ricerca | Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA | ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

OPS4 @ DPCT - Oracle + ZFS

Oracle Spatial is a key feature required to efficiently identify and extract datasets like those of the GAIA use case :

- Cone Search: Distance from SDO_Geometry 2001 (point)
- Meridian Search: Distance from SDO_GEOMETRY 2002 (line) defined by 3 points:
  - 1st @ α as defined by search details
  - 2nd and 3rd @ α+180 degrees, separated by minimum TOLERANCE (this avoided computing the difference between to spherical caps)

Sample Query

```
select * from datadb_c04.completesource
where SDO_WITHIN_DISTANCE(                                   -- meridian
    COORDS,                                        -- sdo_geometry column
    SDO_GEOMETRY(
     2002,             -- line (Oracle code to identify a 2D line)
     20000202,                                     -- SRID (i.e., ICRS)
     null,
     SDO_ELEM_INFO_ARRAY(1,2,1),                   -- Line string whose
                 vertices are connected by straight line segments
     SDO_ORDINATE_ARRAY(0,acos(-1)/4-acos(-1)/2,
acos(-1), acos(-1)/2 - acos(-1)/4, 2*acos(-1)
-0.000000000000002, acos(-1)/4-acos(-1)/2)
                                    -- 3 vertices to define meridian
          -- [alpha_0, delta_0-90] [alpha_0+180, 90-delta_0]
[alpha_0+360-2*TOLERANCE, delta_0-90]), 'distance =
0.00218') = 'TRUE';
```

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

GBIN file format

ALL GAIA SKY

- 2013 Gbins for CompleteSource
- 2170 Gbins for Match
- 71.000 Gbins for AstroElementary

Multiple AE Gbins (up to 7) insist on the same transitID range: this leads to an ambiguity on the identification of the correct gbin given a specific transitid

Impossible to select only the data required: no available off-the-shelf software to shrink the gbins and extract only the data required by the DR

GBIN features

- Compressed serialized java objects
- Requires a specific java sw with the correct DM to be able to access the data.
- Lightweight and suitable for operations
- Not suitable for scientific exploitation and dissemination

A tool based on Apache NiFi to extract the required fields from the gbins is currently under development @DPCT

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

Spoke 3 WP4 IDL

- DBMS: postgreSQL
  - Datamodel an indexing based on the metadata (ATTRIBUTES) of the provided HDF5 Data-Lake
- DMS: Rucio instance hosting the Data-Lake
- Cut and Merge custom SW:
  - Able to retrieve specific parts of the original HDF5 file and provide only the required GROUPS / DATASETS / ATTRIBUTES to the user
- Web interface to allow the submission of queries from the users

Jportal Retrieval System

- Query on objectName, sourceId, or other metadata
- Results pair metadata retrieved from PostgreSQL to the corresponding files from Rucio Data Lake
- Data downloadable thanks to MinIO

Tap Endpoints following IVOA standards

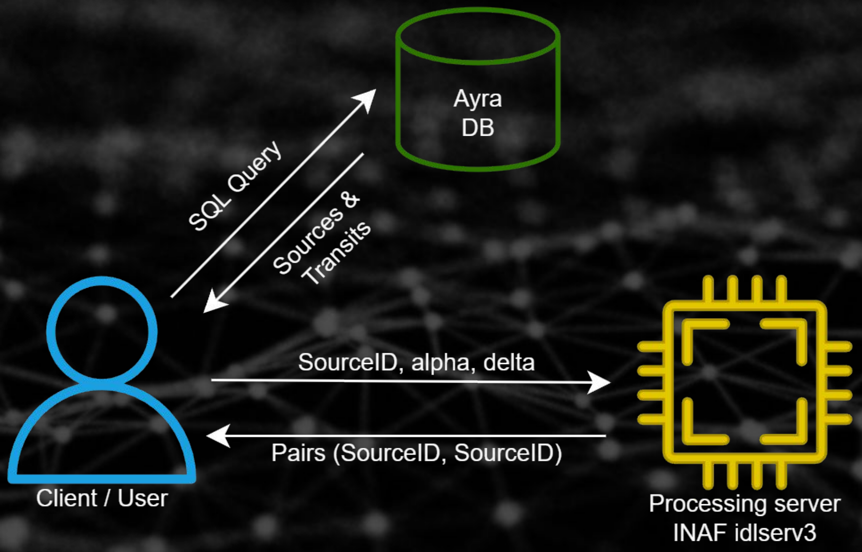
- enables the execution of queries from other clients such as TOPCAT

For more details refer to the presentation: “Data Lakes Spoke 3 and IDL” by Coran & Costantini

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

IGUC - Leonardo & Cherrydata

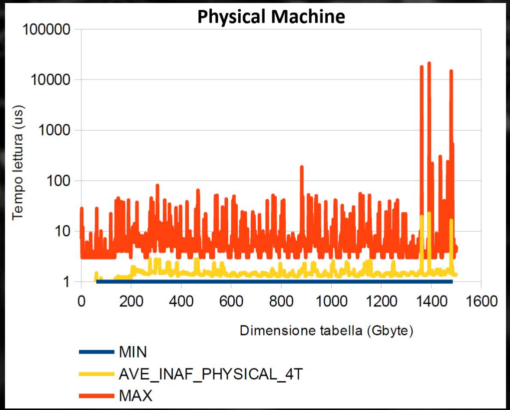
- DBMS: AyraDB
  - database developed by Cherrydata,
  - Key-Value core
  - SQL operations
- Custom SW for data extraction from the HDF5 files
- Data-Lake: HDF5 files
- Deployment of a test AyraDB cluster on INAF infrastructure@OATs

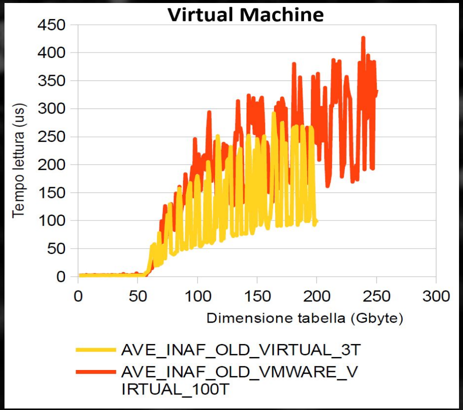


IGUC - Leonardo & Cherrydata

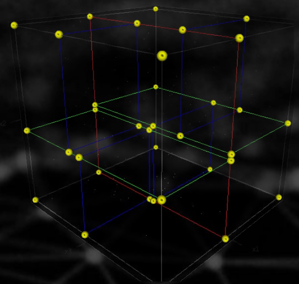
A series of performance test has been executed to estimate di I/O access time with 2 HW configurations

- Virtual Machine: gaiaserv1.ia2.inaf.it, VMWARE configuration, 100 TB drive
- Physical Machine: calcolo02.ia2.inaf.it, 4TB drive





Pairs of neighbouring sources are computed using a 3 dimensional KdTree algorithm:



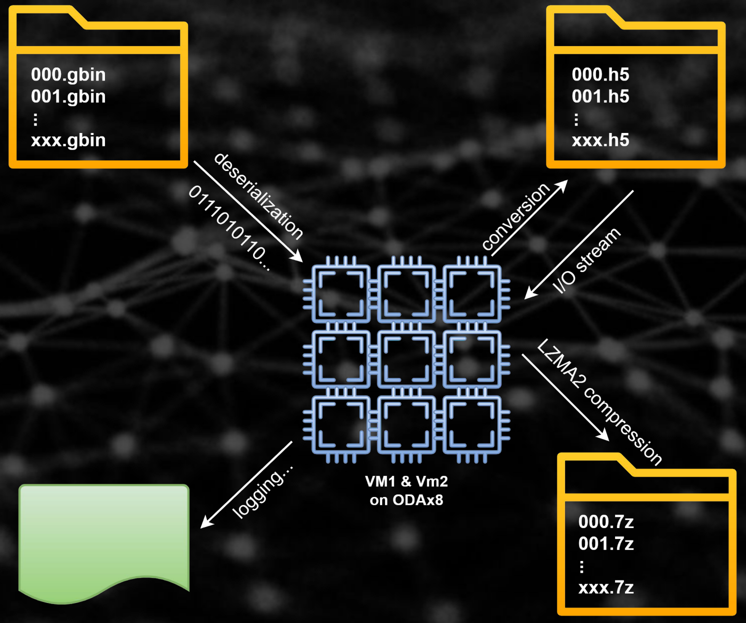
- The angular space is projected onto a sphere in a Cartesian space
- Now it is possible to index the sources using a kdTree of order 3
- The result is a search with a complexity of Nlog(N), where N is the number of sources

Finanziato dall'Unione europea NextGenerationEU | Ministero dell'Università e della Ricerca | Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA | ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

HDF5 Converter

The software heavily leverages reflection and recursion to explore each input objects, regardless of its type, structure or complexity, and creates a corresponding HDF5 file mapping: Java objects, arrays and primitives are mapped respectively to HDF5 Groups, Datasets, and Attributes.

- Gbin files are read from a folder and all its subfolders
- Using MDBDM (Gaia Datamodel) and HDFQL (library) an .h5 version of each gbin is created and stored in a temp folder
- Each h5 is then read through a bufferedStream, compressed using LZMA2 compression algorithm and stored into an output folder
- The hierarchical structure of the original gbin is conserved into the final HDF5



Finanziato dall'Unione europea NextGenerationEU | Ministero dell'Università e della Ricerca | Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA | ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

HDF5 Converter

This apparently trivial and embarrassingly parallel operation resulted being a challenge in terms of memory, I/O, and computation time. This lead to the selection of the ODAx8 as the execution environment

2 Available VMs on ODAx8:

- 48 CPUs each
- 180 GB RAM
- 12TB of NFS for input, temp and output

Each gbin is around 1GB of compressed data, inflating to ~10x when deserialized in RAM → impossible to read 90+ at once

This required the creation of a custom gbin reader, allowing the selection of the number of objects to deserialize (batch read)

HDF5 (x36.3 gbin size) → 7z (x2.26 gbin size)

Each HDFQL operation is analogue to the execution of an SQL statement. This required a careful consideration of the execution statements.

The refactoring of this part of the code, reduced the number of I/O operations of a factor of 10^5

Tests performed on the FS of the ODAx8 showed that a block size of 4MB was optimal for sequential read operations

A custom version of the deserializer (from GT) is in development to leverage this information

Estimated time to convert 1.3TB ~ 7gg

HDF5 Converter

We decided to create entries also for null Objects/Groups or empty Arrays/Datasets

- preserve the structure of the original DM in the exported HDFs,
- leads to a measurable increase in the final volume of the data estimated ~20%

Due to the policies on the distribution of GAIA intermediate data outside the DPAC consortium, we decided to randomize all data non strictly required for the execution of the test

- preserve the volume of the dataset to have a convincing test case
- this leads to a measurable decrease in the compression rate (data have more entropy) ~17%

Sample Configuration

```
######################
## GENERAL SETTINGS ##
######################

# input folder: all gbins found inside the specified folder
# and subfolders will be exported into HDF5 format.
# 1 HDF5 file will be created for each available gbin.
# keeping the same filename
inputFolder = E:\\CherryData\\GBins\\test_random

# folder that will contain the exported HDF5 files
# WARNING: if this folder already contains HDF files with conflicting fil
# the old HDF file will be overwritten
# WARNING: the folder must already exist
# WARNING: the .h5 files will be deleted after the creation of the compre
tempFolder = E:\\CherryData\\Temp

# folder that will contain the exported and compressed HDF5 files
# WARNING: if this folder already contains HDF files with conflicting fil
# the old HDF file will be overwritten
# WARNING: the folder must already exist
outputFolder = E:\\CherryData\\HDF5

# folder that will contain the log file keeping track of the conversion p
# if the conversion progress is interrupted can be restarted from where i
# reading this logfile.
# WARNING: delete this file if you want to restart the conversion process
# WARNING: the folder must already exist
logFolder = E:\\CherryData

# This parameter is used to filter the "get" methods returned by JAVA Ref
# it should be set to a substring of the desired fully qualified class na
# to avoid the invocation of native java methods inherited from the Objec
# default value: gaia
packageFilter = gaia

# Enables/disables the creation of empy groups in place of null objects
# WARNING: enabling this option will increase significantly the storage
# space required by the generated HDF5, but will conserve the original st
# defaults value = false
exportNullObjects = true

# Enables/disables the temp file deletion. Mainly for debug purpose
# default value = true
removeTempFiles = true

# Sets the GbinReader version
# Available values: 3, 4, 5
# default value = 4
gBinReaderVersion = 4

# enables / disables SHUFFLE and ZLIB options for datasets compression
# WARNING: depending on the type of data this option might increase the size of the output file
# default value: false
enableCompression = false
```

```

###############################
## READ/WRITE OPTIMIZATION ##
###############################

# Maximum number of objects to export to HDF5
# 0 means ALL available objects inside input gbin
# default value = 0
maxObjects = 0

# Maximum number of objects to be deserialized at a given time
# this property is used to limit the amount of RAM required at runtime
# default value = 10000
chunkSize = 10000

# Maximun number of objects to be stored in ram, before writing them to disk
# WARNING: higher values reduce IO activity, but heavily impact RAM
# default value = 500
flushSize = 500

# Sets the maximum number of objects to be stored insied a single h5 file to
# limit its volume on disk
# WARNING: if set to 0, removes the limit
# WARNING: hdfMaxObjects must be greater then flushSize
# default value = 10000 ~ 650MB
hdfMaxObjects = 1000


#readBlockSize = 8192
#writeBlockSize = 4096


##########################
## DATA RANDOMIZATION ##
##########################

# enables / disables data randomization.
# Each field will be replaced with a random value of the same type
# default value = false
randomizeData = true

# List of fields that will not be randomized.
# Required to specify the exact field name
# WARNING: this property is enable only when randomizeData is set to TRUE
useOriginalValuesFor = transitid sourceid alpha alphaStarError delta deltaError
```

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

Next Steps

- Start the ingestion of Dataset 1 (Cone+Meridian) into DataAnalysisDB on the ODAx8@DPCT
- Perform extraction (using Apache NiFi) and start conversion of Dataset 2
- Complete the conversion of Dataset 1 to HDF5
  - This might still require a few tweaks on the HDF5Converter (maybe we can talk about it during the workshop)
- Transmit the converted Dataset to our partners Leonardo/Cherrydata and OATs
  - Perform the ingestion into the different systems
  - Start performance testing on the first massive dataset

Bibliography

1. Gaia use case, Sara Gelsumini, Deborah Busonero - Spoke 3 II Technical Workshop, Bologna Dec 17 -19, 2024
2. IDL DM Study and Archiving Ingestion Tests - Giacomo Coran – INAF OATs - Spoke 3 II Technical Workshop, Bologna Dec 17 -19, 2024
3. Status of the IGUC Project - Deborah Busonero (INAF); Paolo Giacomazzi (CherryData) - Spoke 3 II Technical Workshop, Bologna Dec 17 -19, 2024
4. GAIA mission legacy @DPCT: introducing a new data management system - E. Licata et al - Proceedings, 34th Annual Astronomical Data Analysis Software and Systems Conference


THANK YOU!
PURE OV
CAUTION
RISK OF FALLING