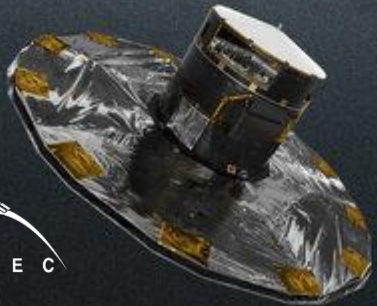
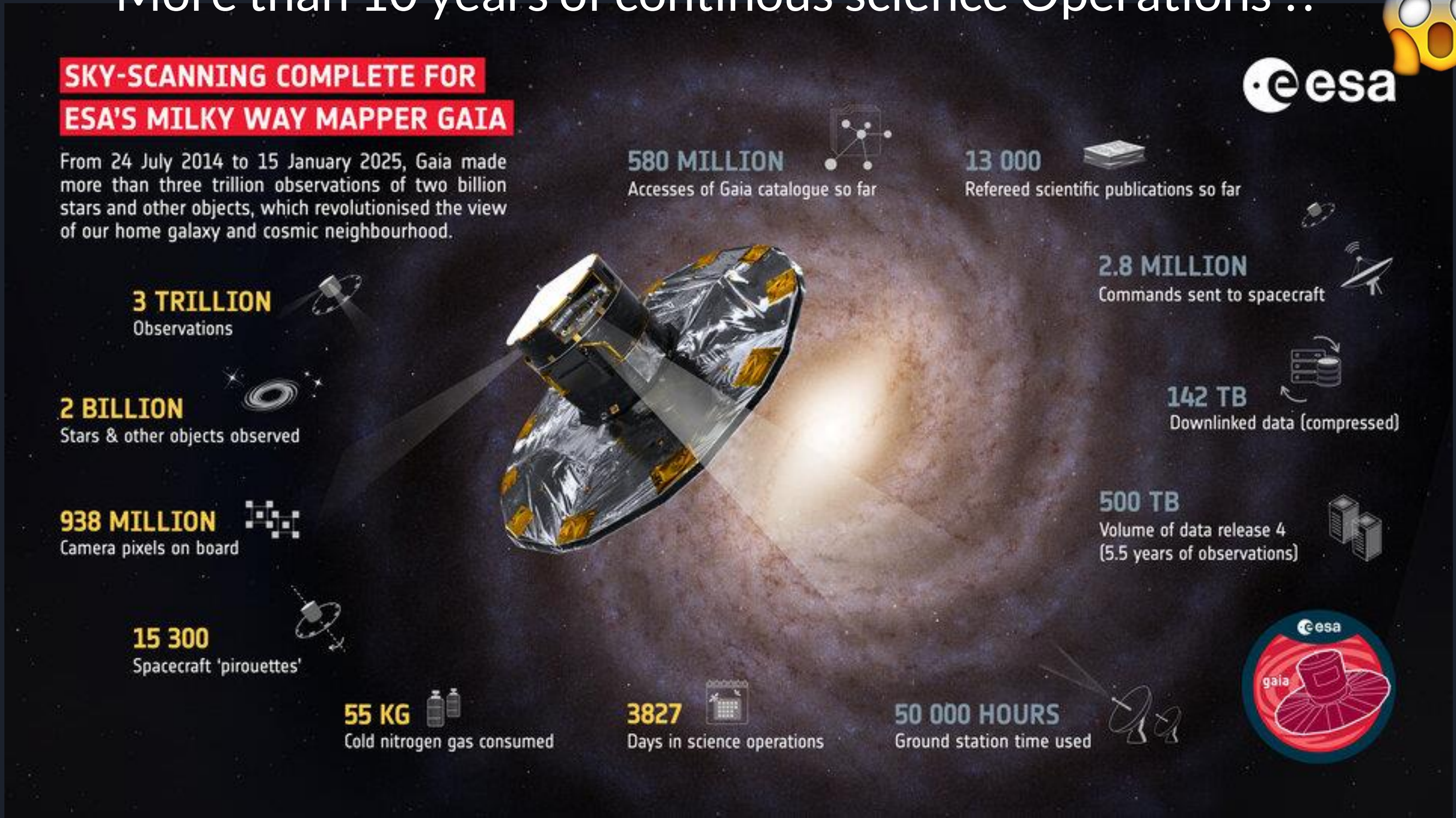
The Gaia mission: towards a whole-sky Legacy Data management and archive infrastructure at OPS4 INAF-ASI facility

Deborah Busonero
on behalf of the INAF and DPCT team



More than 10 years of continous science Operations !!



Gaia science data acquisition numbers and beyond

Gaia science observation phase	
Average distance of Gaia space telescope from Earth (in km)	1,510,000
Number of days in science operations (25 July 2014 - 15 January 2025)	3827
Operational data collected	
Volume of science data collected (in GB)	141,038
Number of object transits through the focal plane	267,336,261,480
Number of astrometric CCD measurements	2,635,171,720,297
Number of photometric CCD measurements	530,821,357,368
Number of object transits through the RVS instrument	17,471,272,143
Number of spectroscopic CCD measurements	52,062,041,736
Total number of CCD measurements (astrometric + photometric + spectroscopic)	3,218,055,119,401
Statistics	
Average number of objects scanned per second	809
Average number of CCD measurements per second	9,732
Nitrogen gas consumed (in kg)	55
Camera pixels onboard	937,782,000
Commands sent to the spacecraft	2,800,000
Ground station time used (hours)	50,000
Spacecraft pirouettes (full scan circles)	15,300

Next Gaia data deliveries to the scientific community worldwide:

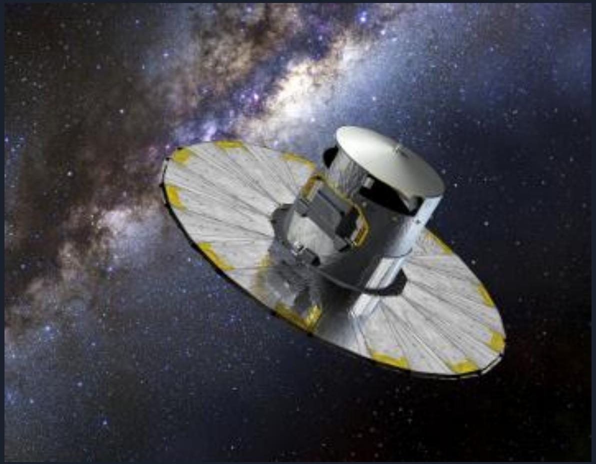
- **GDR4: 2nd quarter of 2026 (5.5 years of satellite science data)**
- **GDR5: not before the end of 2030 (10.5 years of satellite data)**

A few more info on Gaia and its Data

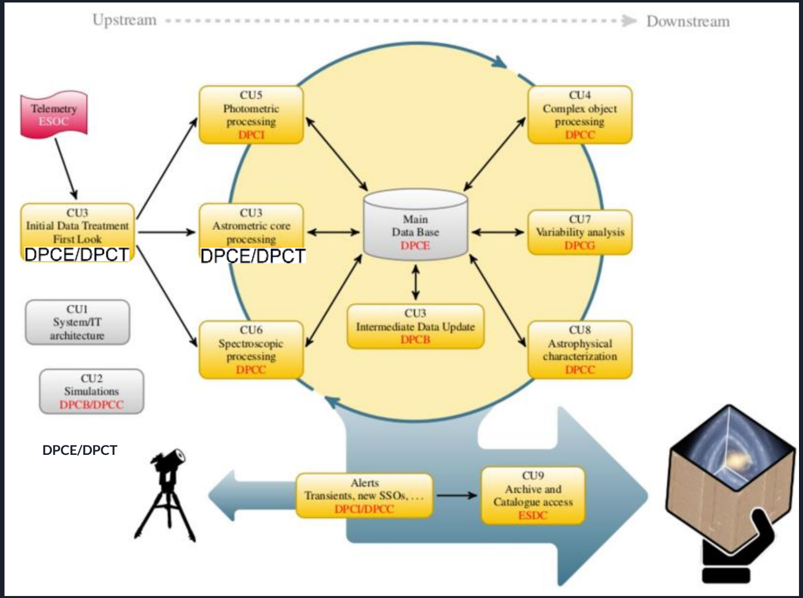
- Up to now, observed and measured 2+ billion sources
- Each source has been observed on average ~200 times
- Each source transit is comprised by
  - telemetry data
  - actual raw data (6.5k ~ 35k windows, 18x12 px, every 4.41s for every one of the 62 AF CCDs)
  - data and metadata generated by multiple processing pipelines across europe (9 Cus – 6 DPC)
  - and much more…

- Estimated size of the full dataset @the end of data reduction in 2030) ~ tenth of PB

- Estimated size of the full dataset @DPC in Torino (including outputs) ~ 3PB



Raw data are NOT images



DPCT: one of the 6 data processing centers (DPC) of the Gaia SGS, hosted in ALTEC in Turin. Under dedicated ASI industrial and scientific contracts, its construction and operation is the result of the work of an integrated INAF-OATo / ALTEC team.

DPCT provides the infrastructure support (in terms of HW, DB and software framework) to run the CU3 sw systems that are part of the AVU (Astrometric Verification Unit): AIM, BAM and GSR from raw data to catalogue data.

Entering the **post-operations phase and Gaia' legacy era**

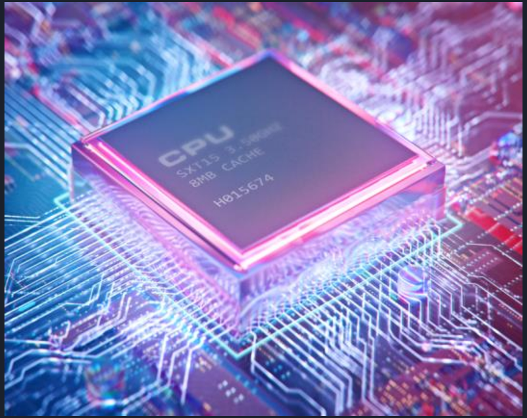
From data acquisition, reduction, analysis and scientific quality validation for the given Project/mission, to **Big Data** data management (including long term preservation) and (**deep: HTC and HPC**) exploitation (including re-processing/calibration pipelines).

Team INAF team, with the support of ALTEC and **additional resources** procured via more recent initiatives funded by PNRR, is working at our Data Processing facility (DPCT@ALTEC) to face the challenges of the mission legacy:

- **Data Management and Computational issues**
- **New paradigms in data science**
- **Infrastructural aspects**

>> All implied by the deep exploitation of the first Big Data system of the faint astronomical sky built only from space-borne data.

Processing Requirements @ DPCT



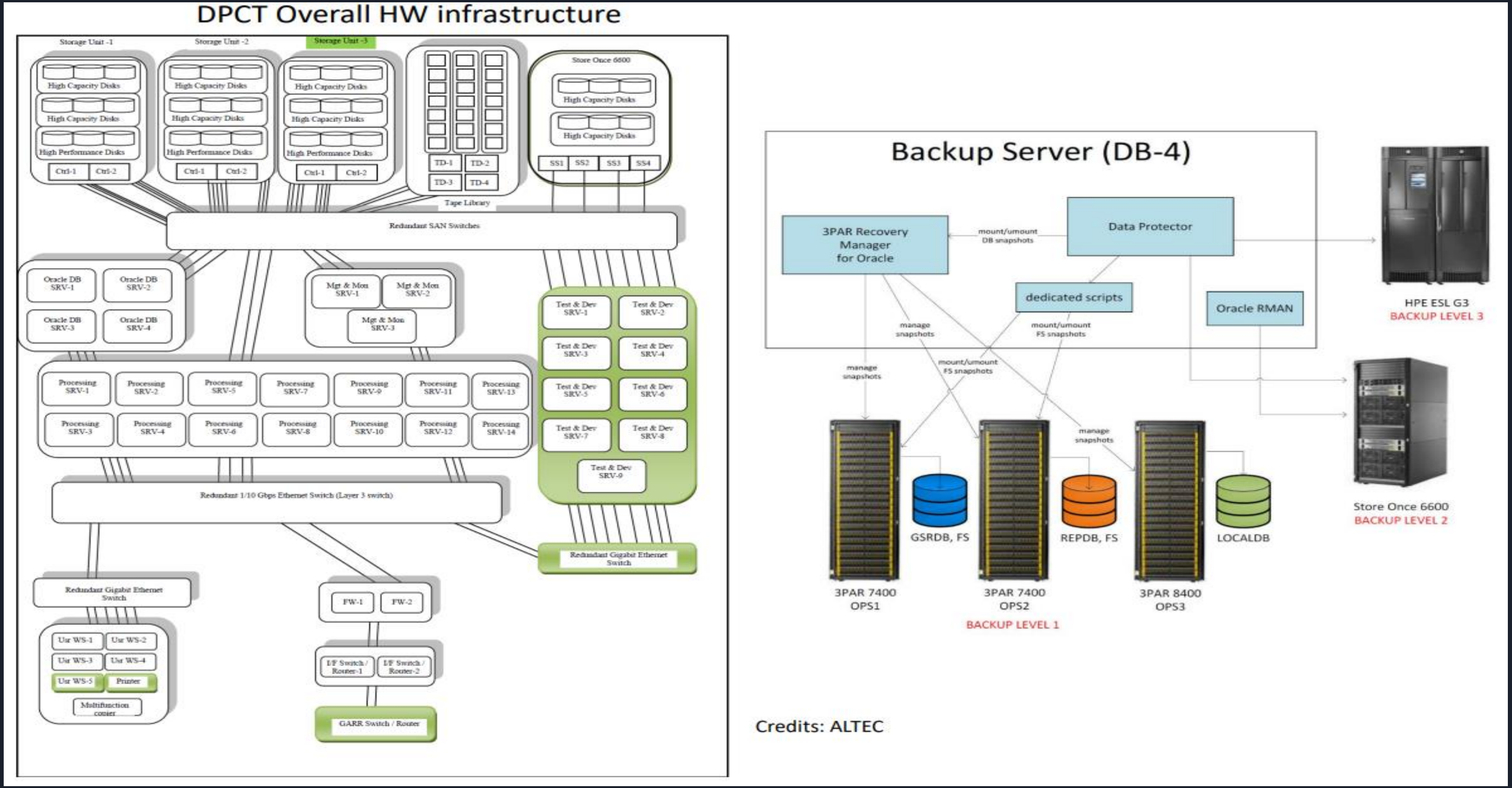
1. Handle a massive dataset: $10^6 \sim 10^7$ transits / day
2. Pipeline integrated DBMS capable of supporting OLTP (online transactional processing) and OLAP (online analytical processing) operations
3. Max processing time for Daily pipeline < 24h
4. Heterogeneous workloads
    - Computation
    - Monitoring
    - Visualization
5. Store all mission data and all produced outputs of both daily and cyclic processing

- Operations and Test and development Platforms
- Procurement paradigm: performed incrementally according to mission needs

INTERNET LINK : from 1 to 10 Gbps via GARR
HOT STORAGE SYSTEM CAPACITY: **2.5+ PB overall raw disk space** distributed between two HP P7400 storage units and **one P8400**.
COMPUTING : **14 servers** HP DL580 G7/G9 with a total of about **600 CPU cores and 4.5TB RAM**.
DEV & TEST: 7 servers HP
DB SERVERS: **3 servers** HP DL580 G7 (**32 cores**, **256MB RAM each**) based on **Oracle RAC technology** (**DBMS Oracle**).
NETWORK CONNECTION: LAN network up to 10 Gbps. SAN network redundant at 8 Gbps.
SECURITY SERVICE: redundant firewall based on pfSense, enabling secure remote access via VPN.
INFRA MONITORING AND MANAGEMENT: services based on VMWare virtual environment configured with two HP DL 580 G7 servers clustered and managed by vCenter Server.
BACKUP SERVERS: HP DL580 G7 dedicated to DB and filesystem backups from data volume snaphots.
3 LEVELS BACKUP : L1 on primary storage array, L2 on disks (StoreOnce 6600) and L3 on tape libraries (HP ESL G3).
HPC INTERCONNECTION: access to HPC super computer at CINECA for dedicated processing.

Direct link to HPC systems at CINECA via dedicated MOU

HW infrastructure 1/2



HW infrastructure 2/2

The DPCT hardware infrastructure is based on a distributed environment, including a computational grid, a database grid and a storage area network.

The DBMS choice is Oracle that provides advanced availability and scalability features. Oracle allows multiple computers to run the Oracle DBMS software simultaneously while accessing a single database, thus providing a clustered database.

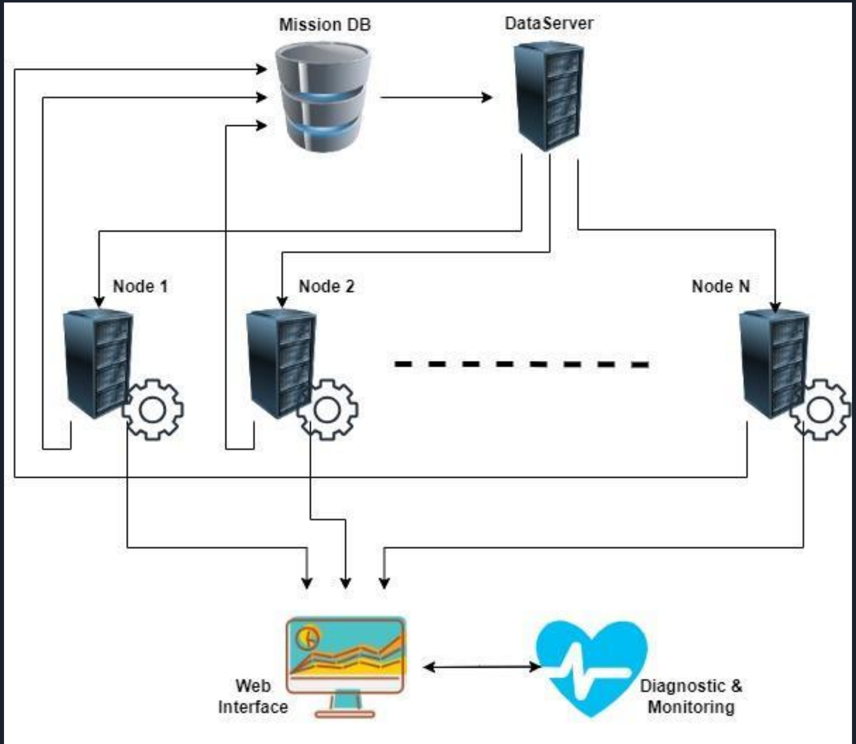
The database grid will use the following Oracle products:

- Oracle Server
- Oracle RAC
- Oracle Partitioning
- Oracle ASM to manage storage used by database

Simplified Functional Schema

For the execution of each production run:

- Data is retrieved from Mission DB and sent to the DataServer (DS)
- The DS performs checks, selects data, creates data packages to be sent to the processing nodes
- Each node runs the scientific SW.
  - Statistical data collection is performed at this level
- Outputs produced by the nodes is used to populate the Web Interface, and to perform Monitoring and Diagnostic Operations, while being stored back to the Mission DB



Production & Test Environments

To be able to develop your project you also need an appropriate dev and testing environment, together with a solid SW life-cycle

PROD ENV: 14 servers with a total of about 600 CPU cores and 4.5TB RAM.

DEV & TEST: half as much (is it enough?)

- Ensure consistency with sw baseline and configuration
- Unit Tests → Use Cases → System Test → Integration Test → Prod
- SRN → STR → Release
- ALL documentation adhere to ECSS standards (European cooperation for space standardization)
- Test automation via Jenkins + Ant Tasks
- 3 SW systems competing for HW resources

Considerations

- Is it reasonable / necessary to have a full scale test env?
- Plan for a procedure to keep Prod and Test envs configuration aligned

Infrastructure Software

- DPCT_SoftwareSystem
  - manages workflow
  - node deployment
  - db interaction
- DPCT_SoftwareConfiguration
  - manages the conf of daily and DRC workflows
  - separate conf for all project modules
  - plot descriptors
- DPCT_ConnectorLibrary
  - Integration between the Infra SWs and the scientific SWs

The integration between Infra and Scientific code managed by the ConnectorLibrary required:

- A Shared effort between INAF and ALTEC on specific parts of the source code
- Required tight cooperation and coordination between research institution and industry for the developent and maintenance

Scientific Code - Overview

- 5 pipelines:
  - AIM - Astrometric Instrument Modeling - daily/cyclic
  - BAM - Basic Angle Monitoring - daily/cyclic
  - GSR - Global Sphere Reconstruction - cyclic
- Parallelization is managed by the infrastructure
  - but many parts have been parallelized to allow testing of large batches of data
- Source code is under version control using SVN (chosen by ESA and imposed to the Consortium)
- Implements established design patterns such as: Chain of Responsibility / Singletons / Abstract Factory
- The data packages send to the processing nodes are not stored "as is": they are runtime items, built from an underlying db and following a specific DM
- the whole AIM and BAM pipelines are performed locally on the DPCT infrastructure
- Computational requirements for GSR exceeds the capability of the DPCT
  - the SOLVER is executed on LEONARDO (see Cesare talk this conference)

Workflow - CD/CI - Issue Tracking

- Centralized management utility VMware vCenter: deployment of VMs across the servers
- Ingestion, Archive, and daily Workflow, are automated using custom GAIA SW system developed in Java
- A monitoring web interface updated in real time (in house sw)
- All main project's branches on SVN (both for infrastructure and scientific SW) are monitored by a Jenkins instance that manages
    - svn checkouts
    - compilation
    - tests with apache ant tasks
- Daily pipelines programmatically generate a report each day
- Issue tracking is managed with JIRA

GAIA TOOLS

During Operations

In Operations direct use of **parallel queries** (via Oracle RAC technology) on DB to meet stringent **DPAC requirements** on processing time of Daily and Cyclic pipelines. Up to **~1000 simultaneous workflows** estimated: however, limited time for code optimization and firmware limitations (HPE infrastructure vs Oracle DBMS) allowed us to achieve ~ **250 at peak performances**.

Daily Pipelines use data in the time domain (objects transits on focal plane) and **DB occupancy depends on cadence and tables** are naturally organized/accessed by time:

- **AstroObservations Table** = the actual Data Lake made of the 16 bit ADU pixels registered on focal plane digital detectors and sent to ground: 0.3 PBy.
- **IDU – AstroElemetaries Table** = Intermediate Data 0.4 PBy containing the actual measurements

The computational *complexity* is driven by the ***number observations*** taken on average **per day** by Gaia.

Relativistic reference systems: tetrads

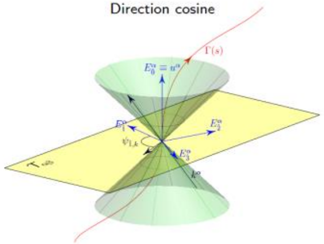
Tetrads can be used when the modeling of a single direction wrt a reference frame is needed (instead, e.g. of an arc).

$$\cos\psi_{k,\hat{a}} = \frac{\mathcal{T}_{\alpha\beta}\, k^\alpha E^\beta_{\hat{a}}}{\sqrt{\mathcal{T}_{\alpha\beta}\, k^\alpha k^\beta}}$$

↓

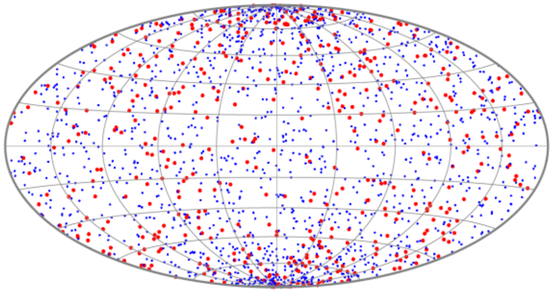
$$\cos\psi_{\mathbf{r}',\hat{\mathbf{X}}} = \frac{\mathbf{r}'\cdot\hat{\mathbf{X}}}{r'}$$

$(\mathbf{r}' \to \mathbf{v})$



Primaries and Secondaries

The reference frame materialized by the primaries is used by other pipeline processes to include the other stars into the Gaia sphere.



For the massive HPC aspects to reconstruct the celestial sphere with GSR, and the need for the facilities @Cineca

Cyclic Pipeline (General Relativistic Sphere Reconstruction, GSR) utilize data per source (**n~2 billions sources found by Gaia**) and **for each** of the **n sources all of the measurements** (the **IDU – AstroElemetaries Table**) are processed in a whole big astrometric reconstruction of the whole celestial sphere.
GSR is a unique INAF full-GenRel model to 0.1 μas, where both observables (photons) and receiving Observers are relativistic (only two existing worldwide)
[As sources are discovered they are naturally placed on the celestial sphere utilisizing a **HEALPix** sphere partioning schema we implemented in Orcacle via their Spatial product.]

Therefore the **n** Gaia sources must be accessed spatially as well, and intermediate relation tables must be constructed to efficiently build a 2D indexing (time and position). Relation tables that can be huge are partitioned using Oracle specific methods.

Thus, computationl complexity > **n** times that of the Daily pipelines

Processing requirements on the Gaia Science Ground Segment led to the use of massive DB's, based on Oracle DBMS technology.

Legacy (strategic) functions: legacy requirements

Gaia's Legacy

SCIENCE

- Establish & maintain deep/dense Reference Frame at different wavelength
- Development of new mission concepts, their design and profiling >> business line
- **Support mission operations (Euclid, CSST, LISA,….) >> business line (ex. 1, next slide)**
- Support of operations of large ground-based facilities
- Support and operate deep exploitation and interoperability with 'external' data
- **Support the development of innovative reduction and analysis methods by operating partial or full (from raw data) re-calibration/processing functions**

Supporting thousands of simultaneous complex queries/ transactions

SPACE SCIENCE & TECHNOLOGY

- Combining data from the science telemetry channel and that of the housekeeping data >> **Potential business line**

Transitioning towards a new DMS 1/3

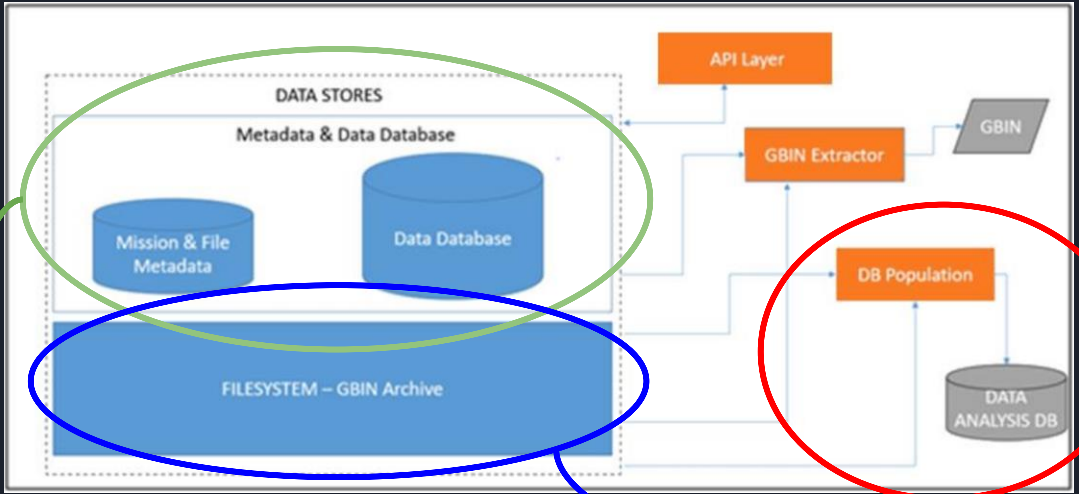
Why

1. Mission extended far beyond the expected timeframe
2. Massive databases/tablespaces (~TB) with detrimental effect on DMBS performance and maintenance
3. Daily / Cyclic operations require different DBMS paradigms
4. Heterogeneous requirements for the data model and its indexing/partitioning:
    - Hybrid DM: Relational with serialized objects (BLOB)

How

1. New Data Management System : Hybrid FS (ZFS) and DB solution
    - data lake with all mission data as GBINS
    - 2 different databases for medata and mission metadata
    - ODA dedicated to the scientific exploitation of the data
2. New data model:
    - Vertical cut of the original tables
    - removed / exploded every BLOB

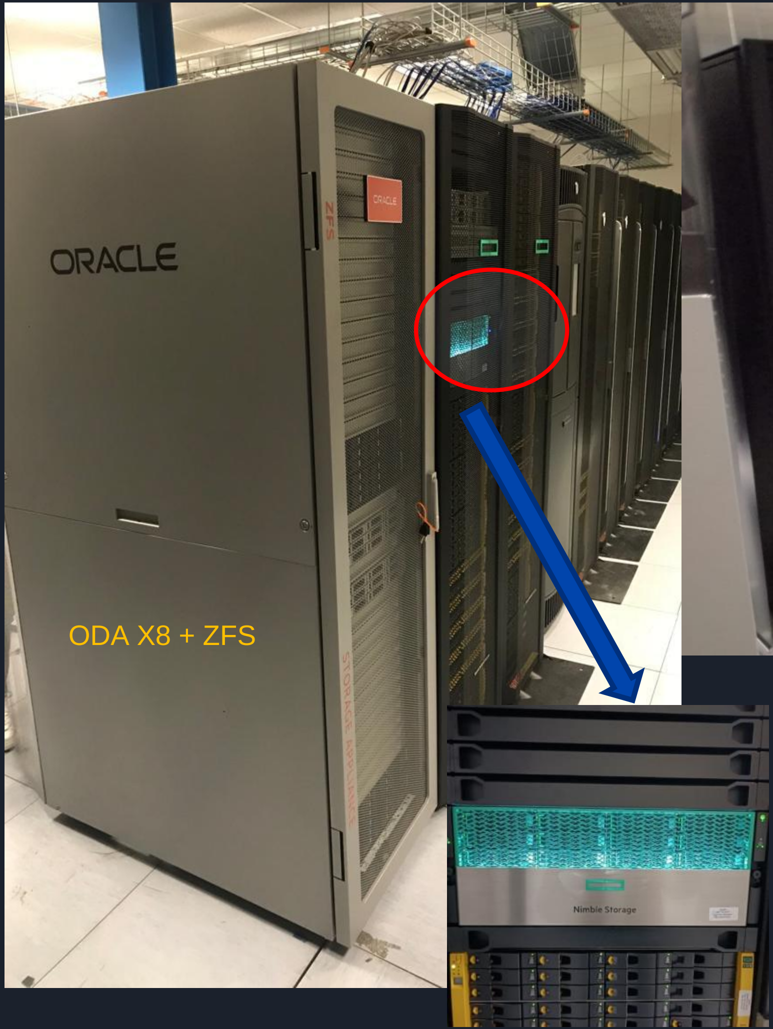
Transitioning towards a new DMS 2/3 - HW & Logical Design



Oracle Database Appliance X8-2-HA, TAA

HPE Nimble HF40

Oracle ZFS Storage Appliance Racked System ZS9-2



INAF OPS4 today: Facility IN PLACE!

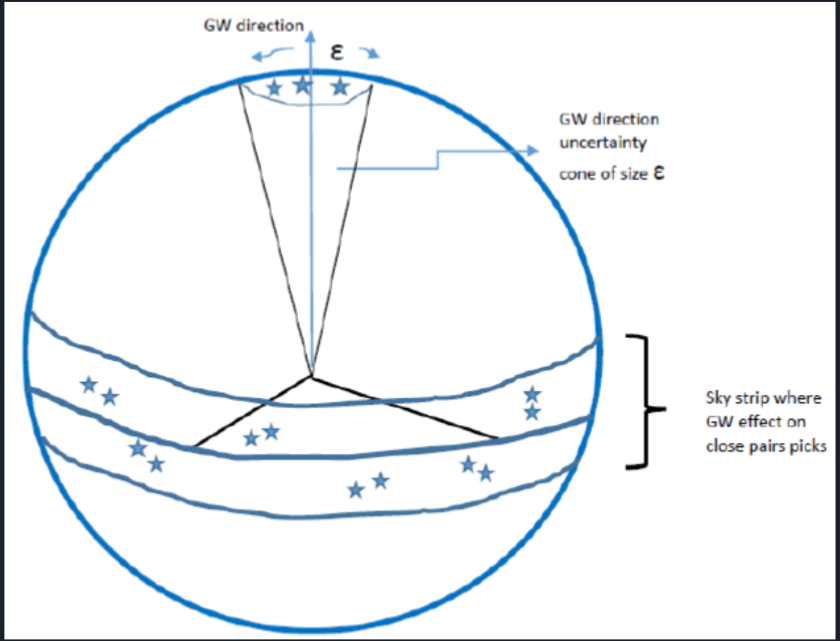
(See Licata's contribution)

Gaia challenge use case

Interoperability Data Lake for Gaia Use Case (IGUC) Requirements:

For any sky direction of interest:

1) **Identify Sources (<u>SourceId</u>, S)** on the celestial sphere in cone toward direction of presumed incoming GW and in the corresponding meridian band (see figure);
2) **Pairing**, in given regions, of all of suitable sources (i.e., with angular distances >0.2 '', appropriate magnitudes and color ranges, photometric stability, etc…) (as shown in figure);
3) **Extract all of the elementary, or epoch, observations (<u>AstroElementaries,</u>Obs) for each pair:**
   a. ***Astrometrically calibrate angular separations***
   b. ***Build separation time series for analysis***



Analyze according to multi-direction fundamental equations (Crosta, MGL et al. 2024)

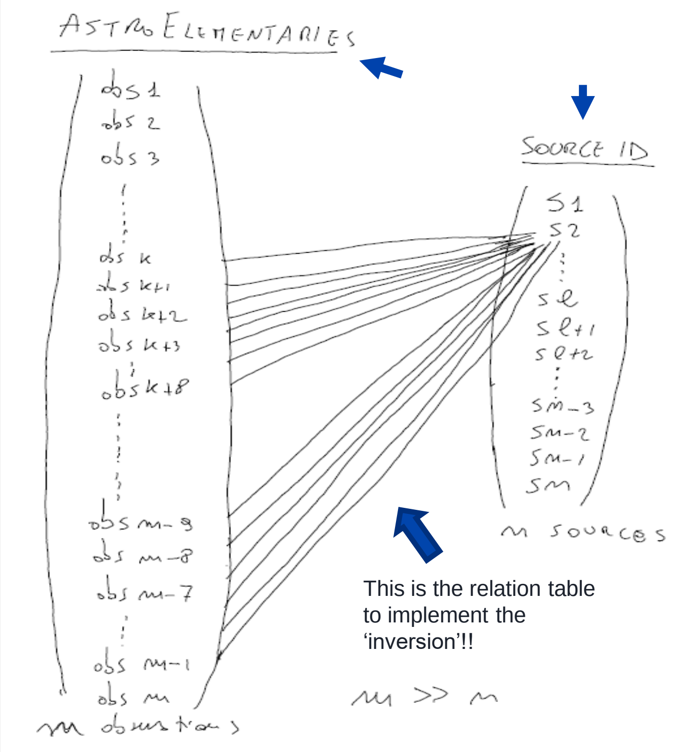
See Licata's talk for more informations

IGUC: Sizing the effort- dimensions and computational complexity

Gaia mission: *natural*, i.e. observation-driven, data archiving schema based on that – no more suitable for the Legacy

I. SourceId table is sky referenced, indexed, via HEALPix : estimated 10^6 **suitable pairs** to G < 17 for each GW direction;

II. current challenge: Extract from IDUAstroElementary Table all of the observations of each pair formed in steps 1 and 2. This means that the relation in Figure must be inverted. Given the dimensions of the IDUAstroElementary table this operation requires proper partitioning of the CrossMatch table (shown in Figure as connections from obs-i to s-j)



Current results: 0.15 sec for the retrieval of all obs of a given source (pair) for 5.5 years of Gaia observations (DR4), i.e., 800 observations per source as sky average (actual numbers depend on ecliptic latitude). Therefore, 0.15 x 10^6 sec for each trial GW direction, or, 42 h of extraction time!! The issue: it is not linearly scaleble

Unrealistic processing time to analyse GW sky at a resolution of, say, 1 square degree. HTC breakthrough required, and\or resort to mitigation strategies (reduce no. of pairs)!!

Transitioning towards a new DMS 3/3

Original GAIA DM and Data format

GAIA DM it's an hybrid between relational and document based:

- Every object is represented as a table but some fields are objects as well
- Impossible to query on some of its data/metadata
- Access to the data requires specific sw tools

GAIA Data Format

- Gbin data format is a zipped serialized java object
- very limited interoperability
- very lightweight (structure not included)

New DM and Data Format

OPS4 DM is fully relational

- all fields are query-able with standard SQL
- data can be accessed directly on the DB (vertical cut) or retrieved fully from the archive

Data Format:

- Experimenting with HDF5
- greatly improved interoperability
- increased storage requirements (by a factor of 2)

See Licata's talk for more informations


THANK you Gaia!