

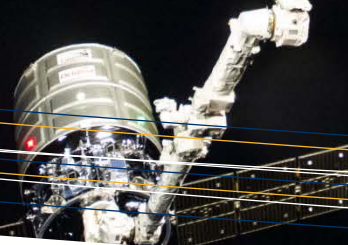
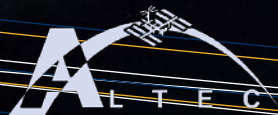


Gaia DPCT at ALTEC

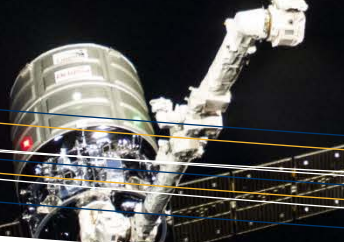
All rights reserved © 2014 - Altec



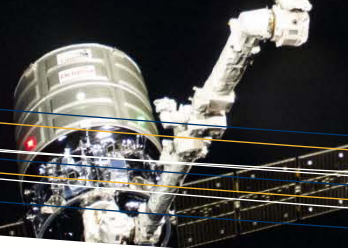
www.altecspace.it



- ALTEC
- Gaia DPCT
- Gaia MDB at DPCT
- Daily Pipelines
- DRC Pipeline
- Data Management
 - Ingestion
 - Retraction & Consolidation
 - Backup
- Data Stores
- Database
 - LOCALDB
 - GSRDB
 - REPDB
- Data Access
 - Services
 - Tools
- DPCT HW Infrastructure
- Q&A

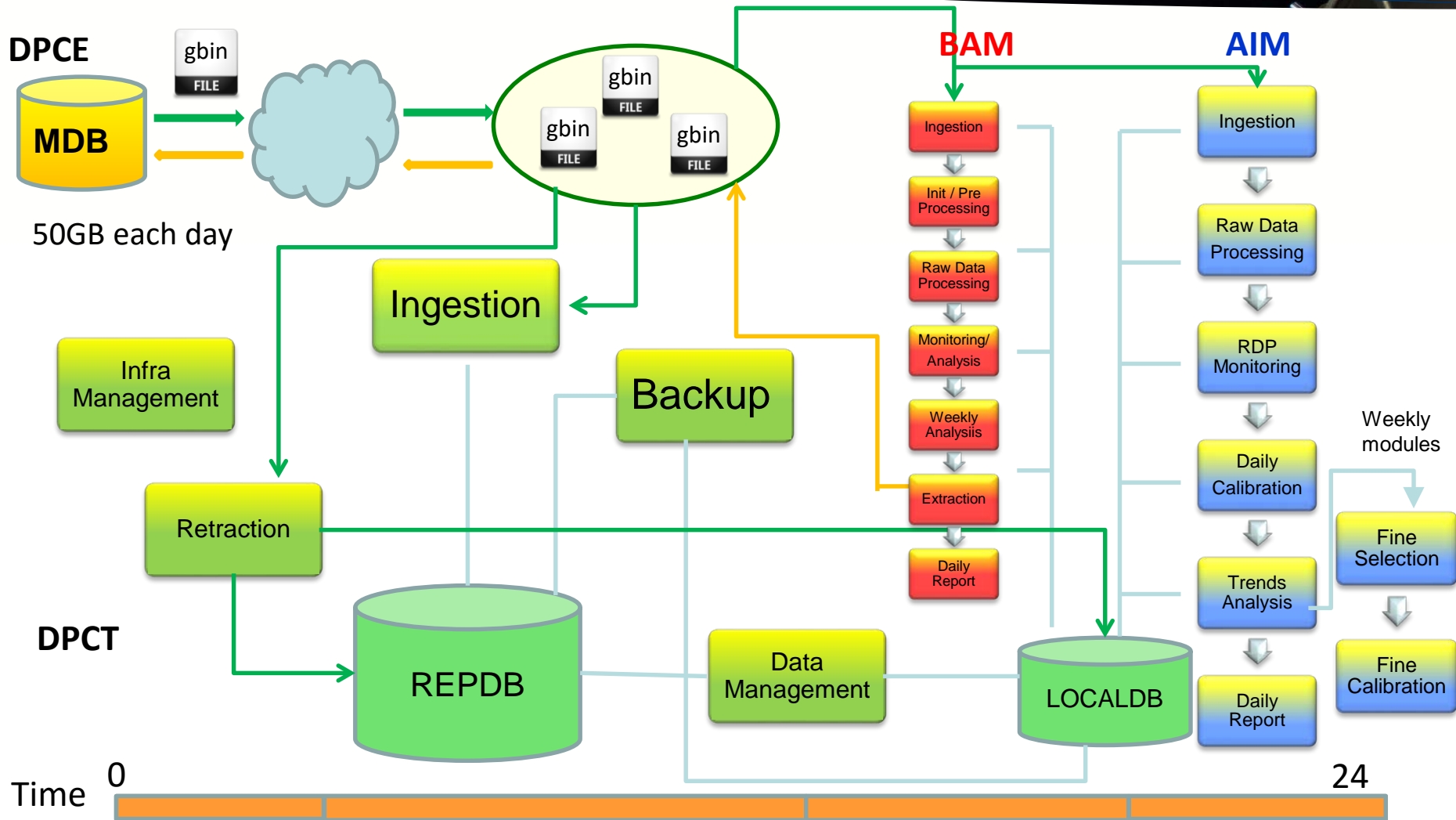


- Public-private company, located in Torino, participated by Thales Alenia Space and ASI (Italian Space Agency)
- ALTEC (Aerospace Logistics Technology Company) operates in:
 - Operating, engineering and logistics support to ISS
 - Astronaut training
 - Support to experiments executed in ISS
 - Operating management of ISS ground segment italian component
 - Development and management of ESA IXV ed ExoMars (ROCC) ground segment
 - Development and management of scientific data processing (Gaia DPCT)
 - Participation in research programs of Distretto Aerospaziale Piemontese (SMAT, STEPS) and european project (FP7: Cross Drive, UIW; H2020: WEKIT)
 - Internal research activities in IT (Big Data, Data processing pipelines)
 - Contribution in management and realization of test facility for optical P/L (partnership with INAF-OATo (OPSYS))
 - Space culture promotion

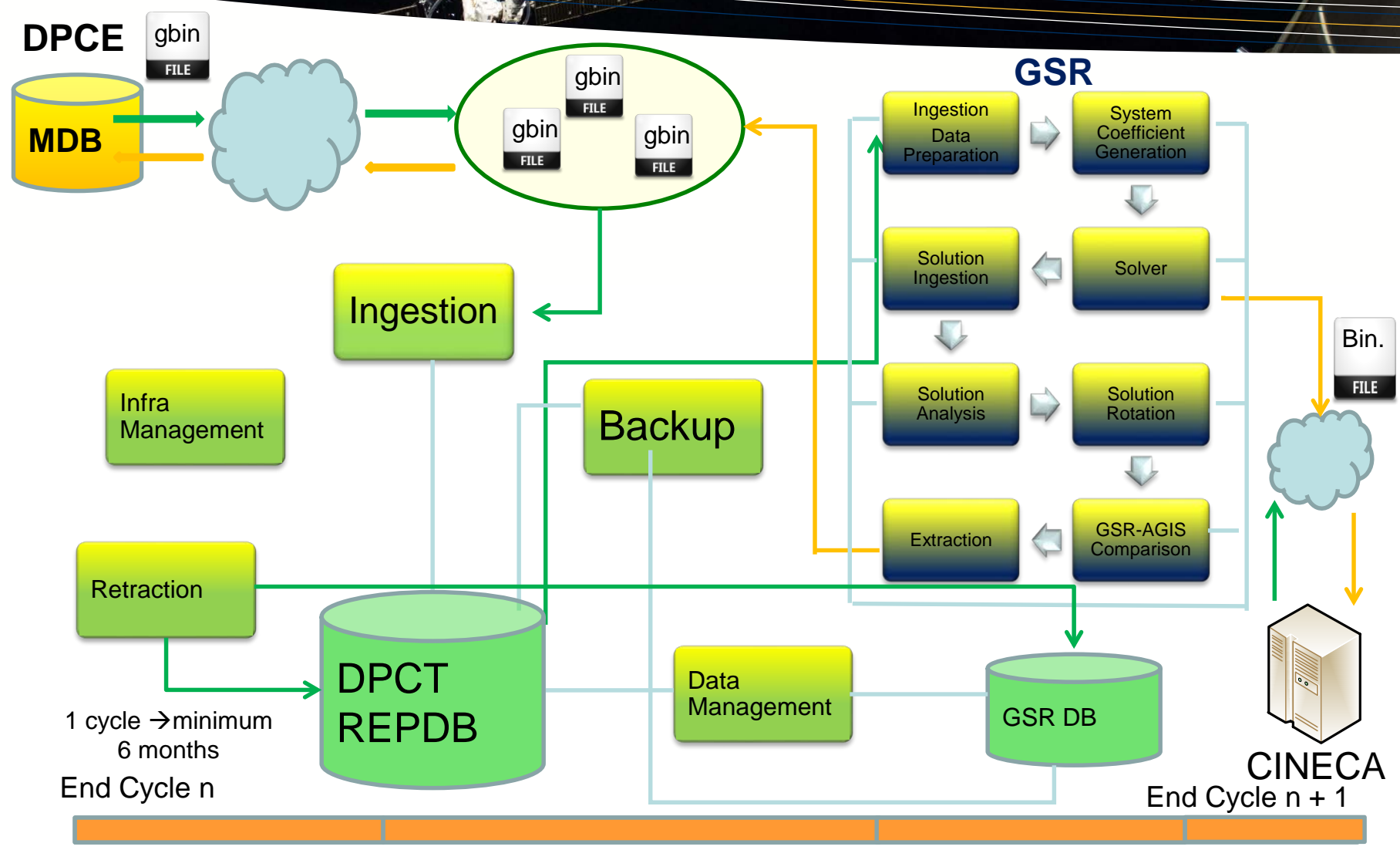


- DPCT is one of six DPCs in the Gaia Data Processing and Analysis Consortium (DPAC), built and hosted in Turin.
- DPCT is a set of products, people and processes to achieve the Astrometric Verification Unit (AVU) in the Gaia mission (Mission Requirements) .
- DPCT is composed by ASI/ALTEC and INAF/OATo
- In addition to mission requirements the DPCT has objectives to medium and long term:
 - Provide infrastructure, software, processes, reprocessing capabilities, data and expertise to support the Italian scientific community for maximum use of all the results of the mission.
 - The DPCT will maintain a copy of the majority of the raw data collected during the mission.

- Data are distributed by DPCE to DPCT in gbin files
 - Data Model is common among DPCs and it can change during mission
- The DPCT contributes to populate the MDB with the output of the AVU systems.
- At the end of the mission, the DPCT could host the whole MDB if the Italian scientific community will decide to have it.
- The DPCT receives data from MDB both on daily and cyclic basis.
- The database REPDB (Repository DataBase) is the data store of DPCT dedicated to manage data coming from MDB.

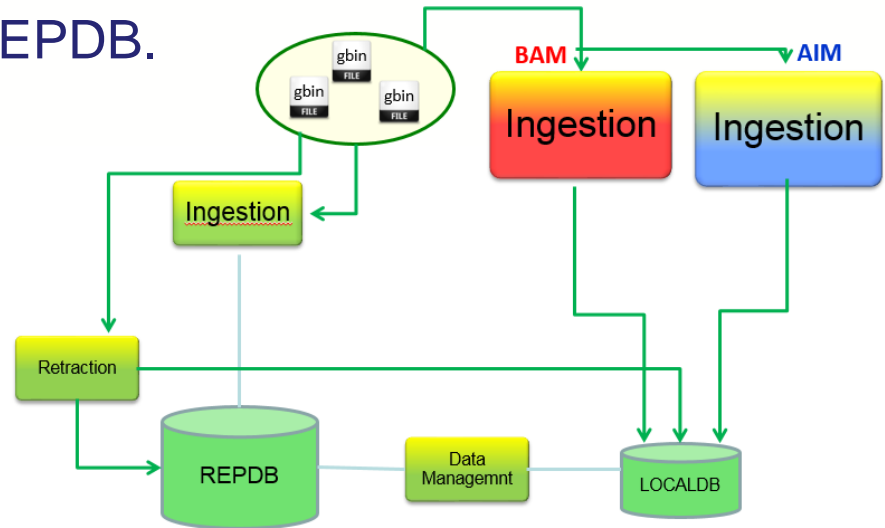


All rights reserved © 2014 - Altec



All rights reserved © 2014 - Altec

- Automatic data ingestion a reception time.
- Daily received files are ingested into DPCT repository (REPDB) and into the processing data store at same time.
- Bulk ingestion for cyclic data.
- Millions of observation ingested into REPDB.
- Success ingestions are events triggering the execution of other modules.



All rights reserved © 2014 - Altec

- The data retraction process consists in the deletion of no more valid data
 - Data can be deleted if there is a wrong processing (at DPCE or at DPCT)
 - If input data are deleted a chains delete all data produce using them
- Regular checks are executed to exclude duplicated and/or missing data
- Pipelines output is archived moving data from the processing DB to the repository DB.
- Data moved in repository (and their input) are remove from processing DB
- Users of REPDB can access to all archived data.

- Consolidated backup strategy on two levels: first level on storage disk (snapshot technology) while second level on the tape.
- First level backup are instantly while tape backups takes days for the biggest database.
- Backups on tape are executed continuously following a strategy based on full and incremental backups.
- Copy of backups will moved in an extenals site to be used in case of disaster.
- Database restore of about 20 TB executed to demonstrate restore capability.



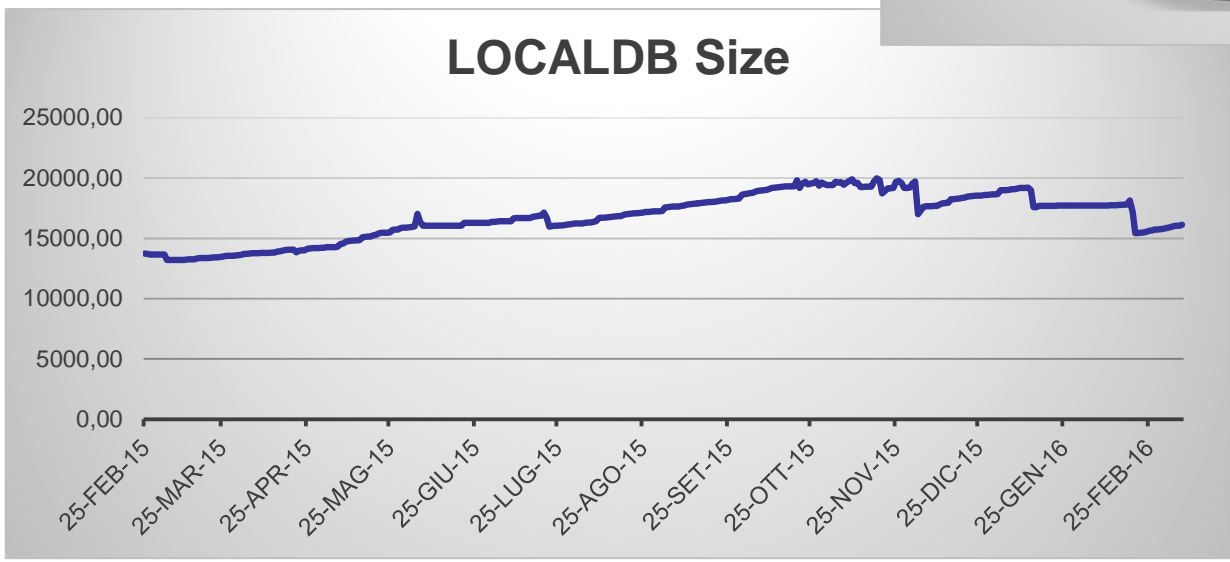
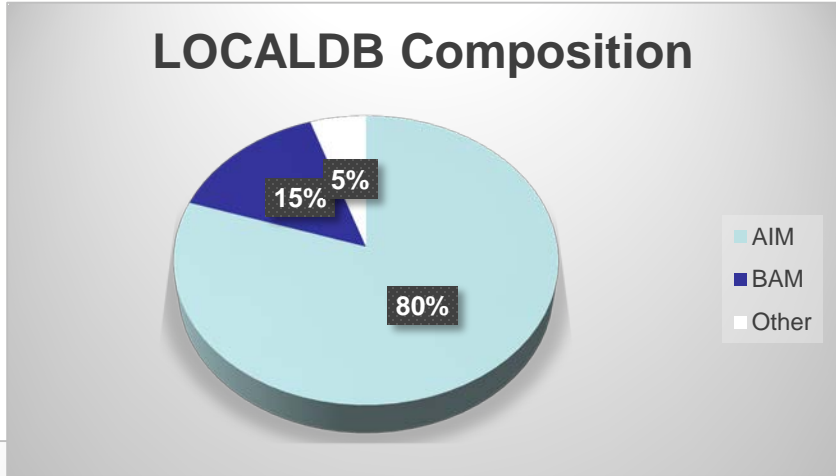
- Data stores: databases and filesystems.
- Database implemented using **ORACLE** technologies allowing to have always all data online.
- There are 3 Oracle 11g RAC (of 3 nodes)
 - REPDB: repository DB containing all received data and all processing output (daily and DRC)
 - LOCALDB: daily processing database
 - GSRDB: DRC processing database
- Data consolidation procedures are executed at database level, needed to control the storage space used by the DBMS and to organize data ready to be queried.

All rights reserved © 2014 - Altec



➤ 08/03/2016

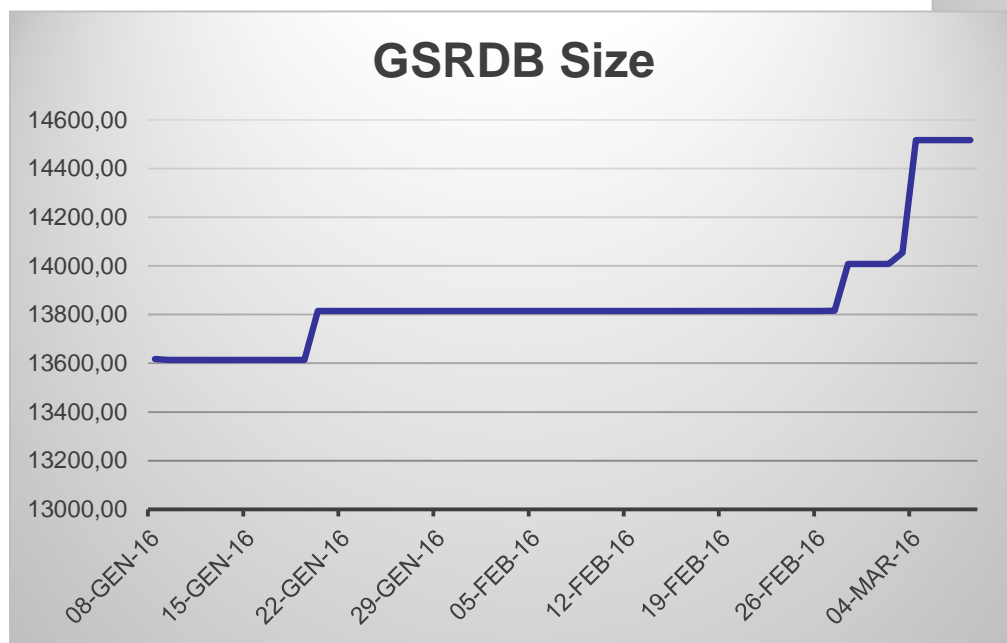
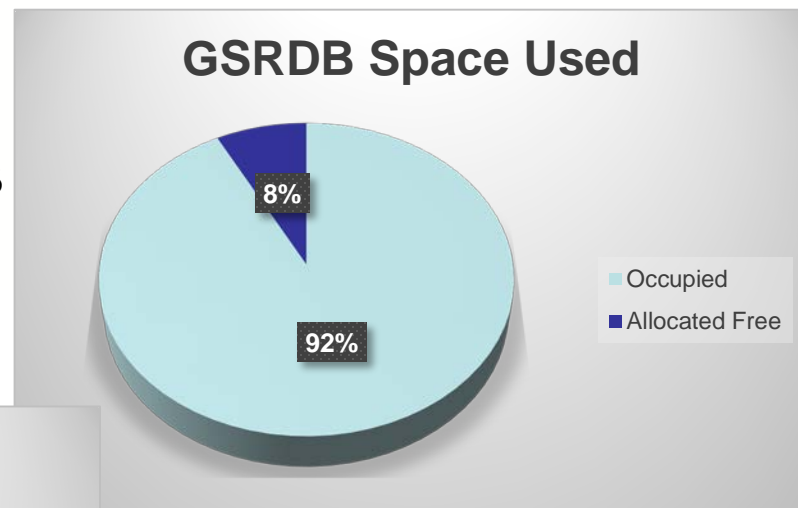
- Used Space: **13929,66** GB
- Allocated Free Space: **2193,418** GB
- Total: **16123,08** GB



All rights reserved © 2014 - Altec

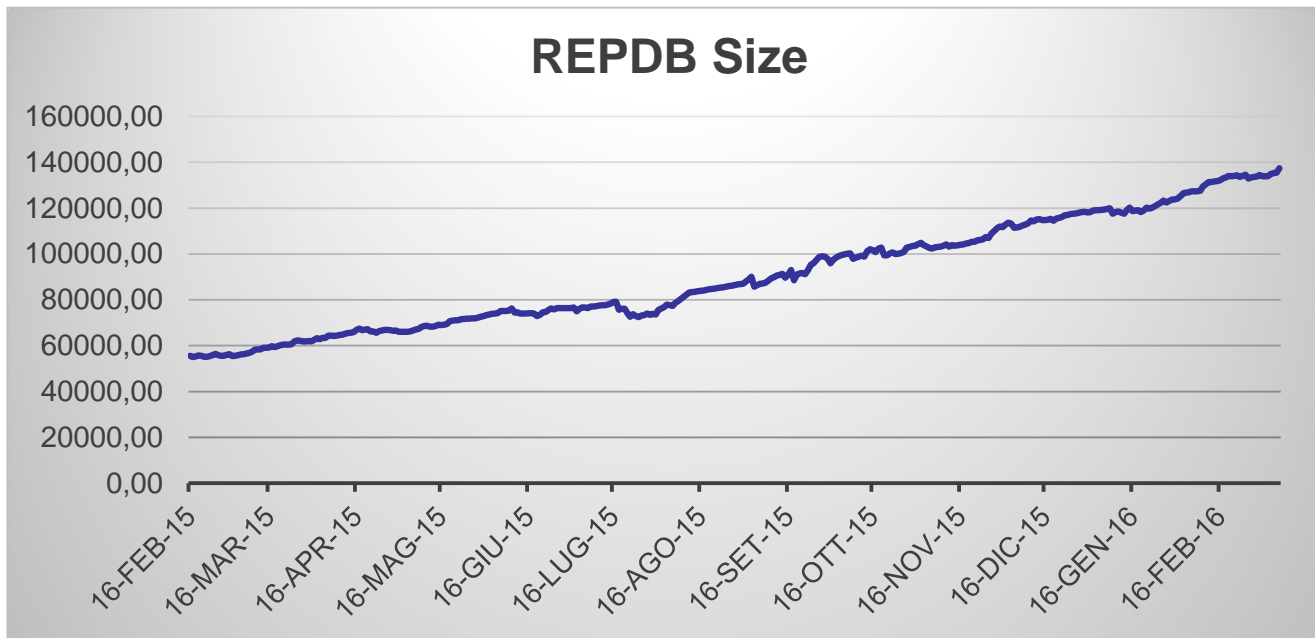
➤ 08/03/2016

- Used space: **13394,52** GB
- Allocated Free space: **1122,30** GB
- Total: **14516,83** GB

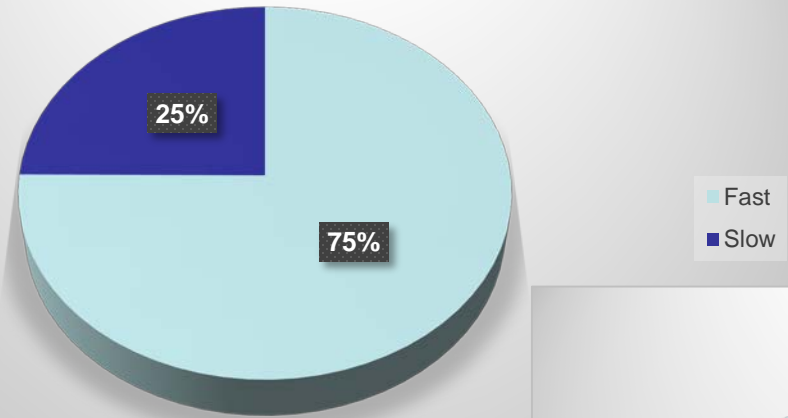


➤ 08/03/2016

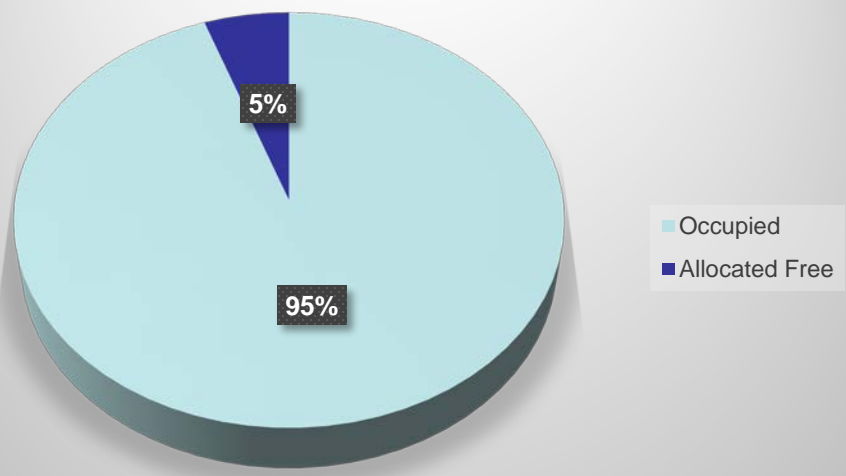
- Used space: **129886,83** Gb
- Allocated Free Space: **7494,45** GB
- Total: **137381,28** GB



Disk Class Occupancy

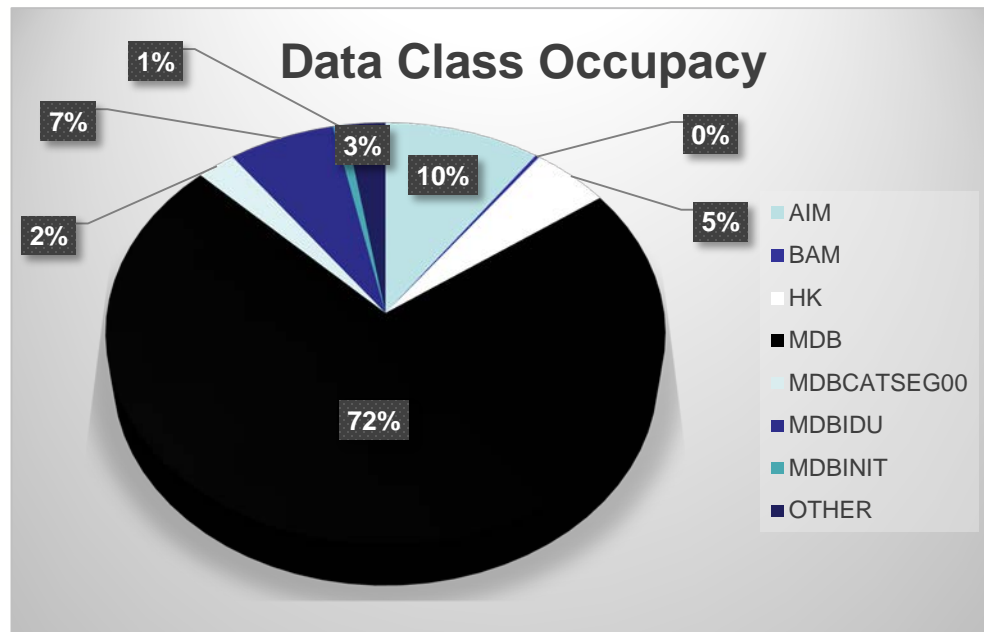


Space Used



All rights reserved © 2014 - Altec

- AIM: 13557,96 GB
- BAM: 348,56 GB
- HK: 6839,75 GB
- MDB: 99404,83 GB
- MDBCATSEG00: 3382,60 GB
- MDBIDU: 9300,36 GB
- MDBINIT: 1021,09 GB
- Other: 3526,14 GB



- DPCT architecture is defined and implemented to provide the data access service to Italian scientists.
- Scientists have read-only access to data managed at DPCT.
- DPCT is providing AVU development teams data packets containing data from different tables/interfaces which can be used to run parts of AVU/AIM and AVU/BAM pipelines.
- DPCT is serving data requests from OATo scientists after that requests are tracked on JIRA
- All people having access to DPCT data access tool have signed the Gaia non-disclosure agreement.
- DPCT has a client area, exposing several services, which is accessible from remote through a secure access.

➤ Tools available in DPCT client area:

- Dedicated interfaces to monitor the processing and see results.
- MDB tools released from CU1 and configured on top of DPCT data stores.
- IDL + Data Mining configured to read data from DPCT repository database.
- Scientists' development environment configured to read data from DPCT repository database.
- Custom workflow environment named DAAS to packet data according to pipeline modules.
- Databases «On demand» are foreseen to support data analysis.
- General Tools such as Topcat, etc..

HIGH PERFORMANCE AND RELIABLE PLATFORM

INTERNET LINK: 1Gbps (300 Mbps guaranteed) via GARR network.

CURRENT STORAGE CAPACITY: 700 TB overall useable disk space distributed between two HP P7400 storage units.

FINAL STORAGE CAPACITY: more than 1 PetaByte at the end of Gaia mission on multiple storage.

PROCESSING CAPACITY: 13 servers HP DL580 G7 with a total of 500 CPU cores and 4TB RAM.

DB SERVERS: 4 HP DL580 G7 dedicated to the database cluster based on Oracle RAC technology.

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.

INFRASTRUCTURE MONITORING AND MANAGEMENT: services based on VMWare virtual environment configured with two HP DL 580 G7 servers clustered and managed by vCenter Server.

BACKUP SYSTEM: first level on storage disks and second level on tape library (HP ESL G3/LTO-5).

HPC INTERCONNECTION: access to HPC super computer at CINECA for dedicated processing.



