



Centri di Processamento e Gestione Dati

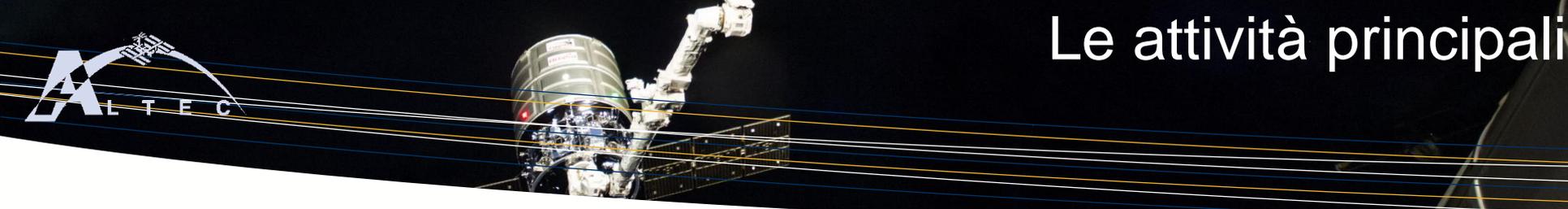
8.Oct.2015

Michele Martino / Rosario Messineo



- E' il centro di eccellenza italiano per la fornitura di servizi ingegneristici e logistici a supporto delle operazioni e dell'utilizzazione della Stazione Spaziale Internazionale e dello sviluppo e della realizzazione delle missioni di esplorazione planetaria.
- E' una società pubblico-privata partecipata da Thales Alenia Space, dall'Agenzia Spaziale Italiana.
- Ha sede a Torino in corso Marche ed ha proprio personale distaccato presso il NASA Johnson Space Center a Houston, presso l'European Astronauts Center a Colonia e al Columbus Control Centre a Monaco.



- 
- The top of the slide features a dark background with a photograph of the International Space Station (ISS) in orbit. On the left side, the ALTEC logo is displayed, consisting of a stylized 'A' followed by the letters 'L T E C' in a spaced-out font. Several thin, horizontal lines in blue and yellow cross the top of the slide.
- Il supporto operativo, ingegneristico e logistico alla ISS, l'addestramento degli astronauti, ed il supporto agli esperimenti in orbita, in particolare di biomedicina
 - La gestione operativa dei componenti italiani del ground segment della ISS (ALTEC Mission Support Center e ALTEC Engineering Support Center)
 - Lo sviluppo e la gestione del segmento di terra dei programmi ESA IXV ed ExoMars (ROCC)
 - Lo sviluppo e gestione operativa di centri processamento dati scientifici (Gaia DPCT)
 - La partecipazione ai programmi di ricerca del Distretto Aerospaziale Piemontese (SMAT, STEPS) ed Europei (FP7: Cross Drive, UIW; H2020: WEKIT)
 - Lo svolgimento di attività di ricerca interna in ambito IT (Big Data, Data processing pipelines) e tecnologico (utilizzo di palloni stratosferici e vincolati)
 - Il contributo alla realizzazione e gestione di test facility per P/L ottici in collaborazione con INAF-OATo (OPSYS)
 - la promozione della cultura spaziale.

**SUPPORTO
ISS**



PMM



LOGISTICS CENTER



FLIGHT



CREW TRAINING



HUMAN SPACE

**CENTRO
CONTROLLO
E GESTIONE
DATI**



ALTEC MSC



ALTEC ESC

**CENTRO
CONTROLLO E
GESTIONE
DATI**



IXV MCC



GAIA DPCT

**ESPLORAZION
E
PLANETARIA**



EXOMARS ROCC



**PROGRAMMI
DI RICERCA**



SMAT F2



STEPS 2



FP7



FP7

**TEST
FACILITIES**



OPSS

**PROMOZIONE
CULTURA
SPAZIALE**



- ALTEC è ospitata all'interno del Centro Multi-Funzionale Spaziale (CMFS).
- Collegamento diretto al nodo del GARR di Torino (fin dal 1998) con un link a 1Gbps (300 Mbps garantiti) e dispone di un collegamento secondario con la rete COLT
- Collegamento alla rete ASINET per la ricezione dei dati real time dell'ISS, incluso i flussi audio e video. Attraverso ASINET ALTEC è collegata alla stazione di terra di Malindi.
- Sala server con circa 30 rack e 70 server per l'hosting delle infrastrutture ICT a supporto delle attività e dei programmi aziendali.
- Integrazione con il super computer italiano gestito dal CINECA per utilizzarne la potenza di calcolo.
- Sale missioni dedicate alle attività operative.



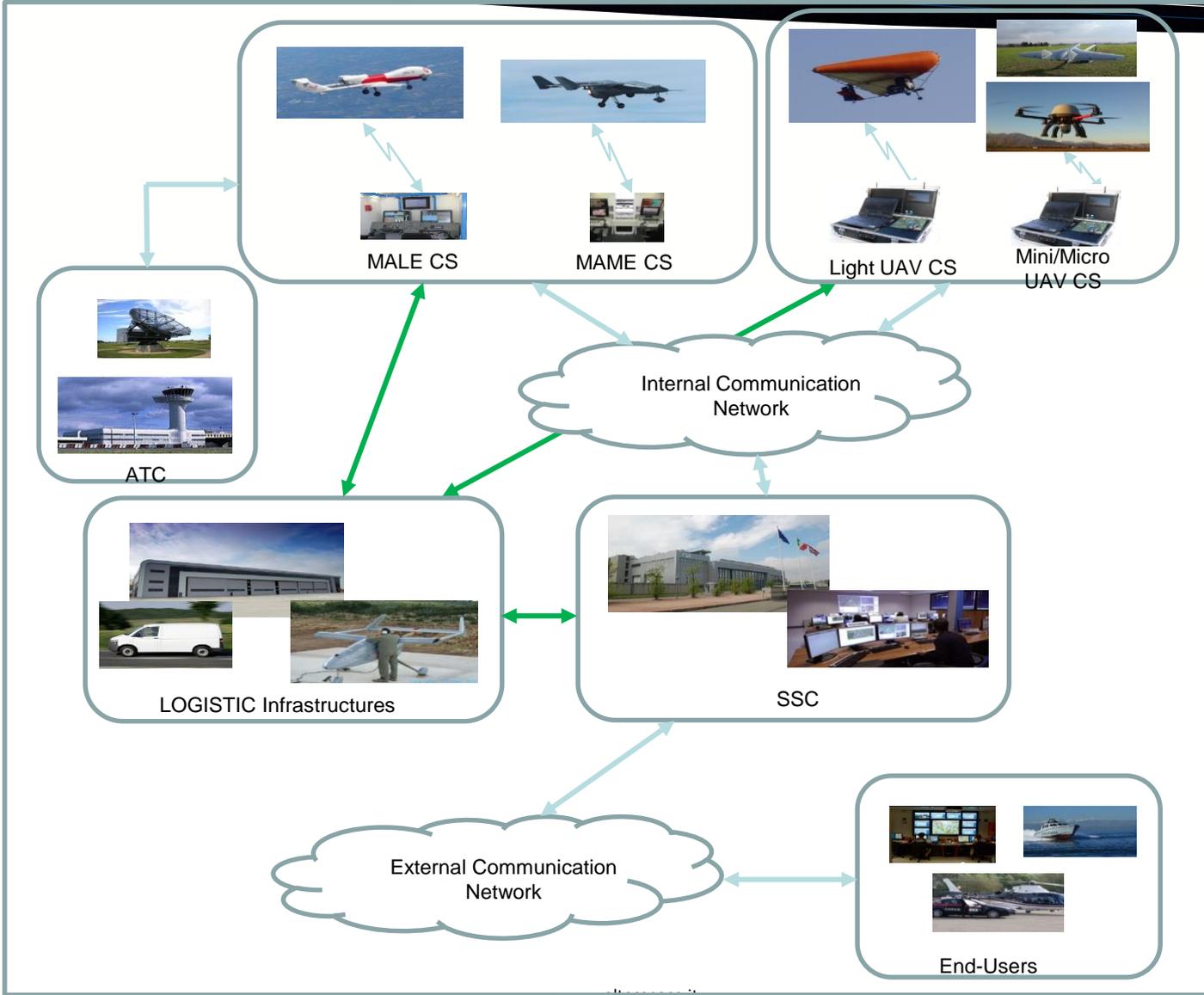


- Lo svolgimento delle attività «core» richiedono competenze sistemistiche e tecnologiche, nonché infrastrutture IT e di comunicazione, costruite da ALTEC nel corso degli anni sia con investimenti aziendali che di programma.
- Questi investimenti sono stati condotti sulla base di alcuni principi:
 - Mantenere la massima commonality a livello architetturale delle soluzioni tecnologiche adottate, pur nella diversa specificità applicativa
 - Garantire il travaso delle esperienze acquisite nelle varie attività / progetti e quanto più possibile riutilizzare soluzioni tecnologiche e software applicativi
 - Omogeneizzare i processi di gestione operativa delle infrastrutture
 - Utilizzare quanto più possibile soluzioni architetture HW aperte
 - Sfruttare le opportunità, ove applicabili, di soluzioni SW open source
 - Sperimentare tecnologie emergenti utilizzandole prima possibile all'interno delle infrastrutture esistenti
 - Mantenere il know-how del personale nel tempo
 - Stabilire rapporti di collaborazione con vendor IT selezionati per sfruttare al meglio le opportunità di evoluzione tecnologica e di accordi commerciali

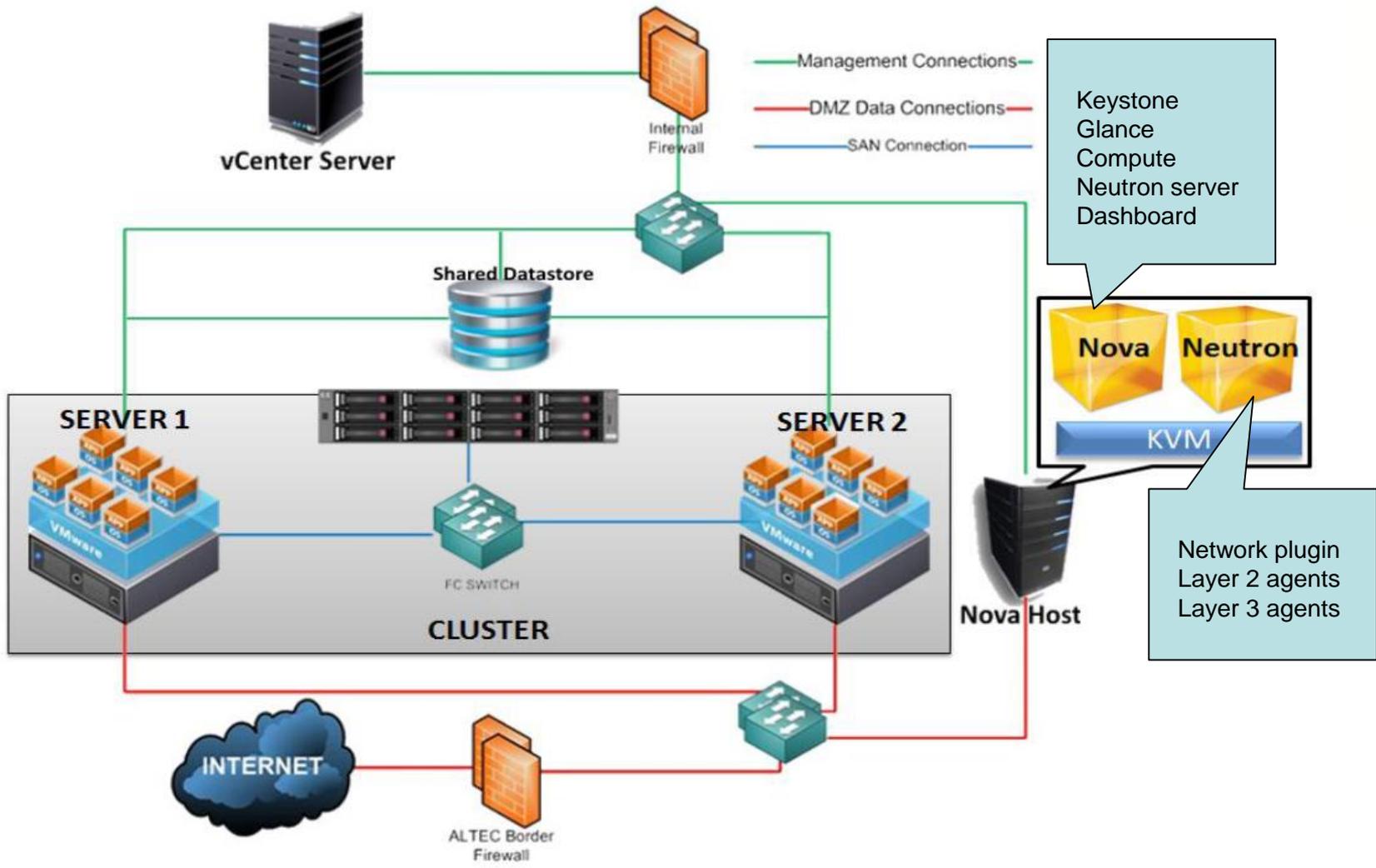
- In questo contesto la virtualizzazione di tutto l'ambiente IT del centro ALTEC (salvo poche eccezioni) con tecnologia VmWare, rappresenta un elemento chiave della nostra infrastruttura:
 - Semplicità di gestione, deployment e manutenzione degli ambienti
 - Massimizzazione dell'utilizzo dell'HW disponibile e riduzione dei costi (il Mission Support Center dedicato alla ISS riutilizzato completamente per la realizzazione del Mission Control Center di IXV)
 - Maggior rapidità nel rispondere alle necessità degli sviluppatori
 - Possibilità di scalare le soluzioni sviluppate
 - Elemento essenziale dei nuovi progetti di sviluppo (ExoMars ROCC, SMAT SSC)
- Ambienti virtualizzati: infrastruttura di produzione aziendale, servizi su DMZ, Multi-purpose Mission Support Center, Columbus Engineering Support Center, servizi generali del DPCT, ExoMars ROCC (development, integration, operational e disaster recovery platforms), progetto SMAT SSC, ambiente di sviluppo dei programmi, ambienti per attività R&D



- Il progetto SMAT (Sistema di Monitoraggio Avanzato del Territorio) del Distretto Aerospaziale Piemontese, è stata l'opportunità per mettere a frutto le competenze acquisite in ambito ground segment di missioni spaziali, ad ambiti diversi quali quelli di gestione di flotte di UAV, e nello stesso tempo valutare e provare l'implementazione di nuove tecnologie per risolvere problemi specifici.
- Nello sviluppo della SSC (Stazione di Supervisione e Coordinamento) è stata ampiamente utilizzata l'infrastruttura virtualizzata con VmWare, ma essa è stata integrata con la soluzione di cloud computing di OpenStack (in collaborazione con ISMB).
- Openstack è stato qui utilizzato per permettere il deployment dinamico ed automatico delle VMs dedicate alla gestione real time di una singola missione di uno o più UAV, gestito e controllato da un sistema centralizzato – Mission Index - alimentato dai processi di planning tattico delle missioni da eseguire.
- Lo scenario operativo di riferimento è quello di una flotta di UAV di tipologia e configurazioni differenti distribuiti su un territorio di tipo nazionale, in siti logistici diversi, in grado di eseguire missioni concorrenti, ma indipendenti l'una dall'altra.
- Due SSC sono in grado di gestire tutte le operazioni e processamento di dati e sono in ridondanza l'uno dell'altra.



All rights reserved © 2014 - Altec



All rights reserved © 2014 - Altec

- A partire da questa esperienza si vuole ora estendere la sperimentazione di Openstack sui seguenti aspetti:
 - Utilizzo a supporto delle attività di sviluppo e test delle nostre unità operative offrendo in modo più efficiente l'accesso alle risorse computazionali
 - Sperimentare l'utilizzo di soluzioni di virtualizzazione alternativi a VmWare
 - Approfondire le caratteristiche del cloud per identificare meglio ulteriori applicazioni nella tipologia di infrastrutture ed architetture che oggi ALTEC gestisce, in particolare:
 - per gli aspetti legati alle tecnologie di Big Data
 - per i potenziali utilizzi in configurazioni a supporto dell'exploitation dei dati contenuti nell'archivio del DPCT di Gaia

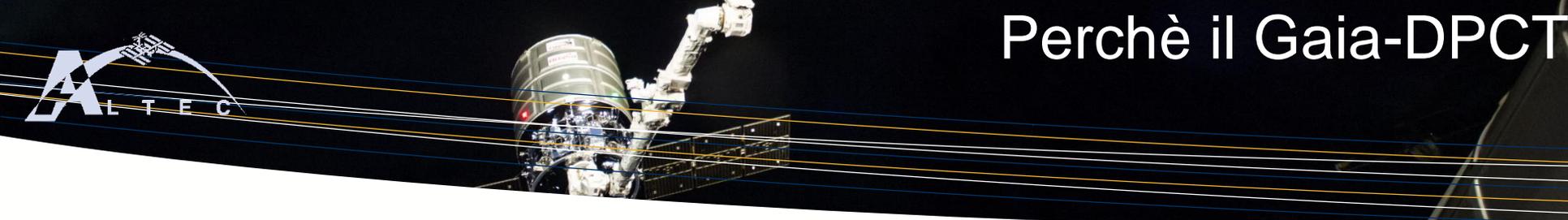


- Altro elemento sviluppato all'interno del progetto SMAT riguarda la gestione automatizzata e distribuita di pipeline di processamento in modalità automatica attraverso l'utilizzo di workflows:
 - Il deployment dei componenti di sistema avviene su macchine virtuali (VmWare)
- Questo sviluppo costituisce un primo prototipo di generalizzazione dei concetti implementati nel framework del DPCT di Gaia.
- Un secondo prototipo è stato applicato all'attività di sviluppo di un sistema per la gestione e processamento dei dati del satellite NASA SMAP (di cui ALTEC è early adopter insieme ad Arpa Piemonte).
 - In questo caso il deployment dei componenti del sistema è stato realizzato tramite la tecnologia docker (containers technology)
- Queste attività sono servite a consolidare requirements e architettura del sistema ASDP (un prodotto interno di ALTEC), attualmente in fase di sviluppo, e che è stato proposto ad ESA per l'utilizzo nel programma ExoMars 2018 nella gestione delle pipeline di processing dei dati scientifici del Rover.



Sotto contratto ASI e in stretta connessione con l'INAF-Osservatorio Astronomico di Torino, ALTEC ha progettato, realizzato e conduce il Data Processing Center Italiano per il Programma GAIA, che fa parte di una rete di 6 Centri Europei.



- 
- Il DPCT rappresenta un contributo importante dell'Italia alla missione Gaia
 - Due sono i requisiti/obiettivi primari dello sviluppo del DPCT

Requisiti specifici di missione

- Realizzare l'Astrometric Verification Unit (AVU) della missione, attraverso l'esecuzione giornaliera delle pipeline dei sistemi scientifici AIM e BAM, e l'esecuzione ciclica (DRC) del sistema GSR;
- Ospitare ed integrare l'Initial Gaia Source List (IGSL) DB,
- Supportare l'esecuzione dell'esperimento GAREQ.

Obiettivi di medio-lungo termine

- Fornire l'infrastruttura, il SW, i processi, le reprocessing capabilities, i dati e le competenze a supporto della comunità scientifica Italiana per il massimo utilizzo di tutti i risultati della missione,
- Questo aspetto va ben oltre l'accesso e lo sfruttamento scientifico del catalogo (prodotto dal consorzio DPAC), permettendo l'accesso e l'estrazione di dati dal Mission Data Base di cui il DPCT manterrà copia.



Il lavoro del DPCT è caratterizzato da una stretta collaborazione tra team scientifico e team industriale e utilizza la infrastruttura del centro di controllo/project office@ALTEC



ALTEC/OATo Staff meeting #61

@DPCT, 29/9/2015

➤ Missione

- ✓ Lancio del satellite Gaia **19.Dic.2013.**
- ✓ Ricezione dei primi dati di HK **18.Dic.2013.**
- ✓ Inizio ricezione dati ed attivazione dei sistemi daily AVU **3.Gen.2014.**
- ✓ Primo ciclo di processamento DPAC è terminato a settembre 2015.
- ✓ La pipeline operativa di GSR entrerà in produzione a fine 2015.

➤ Processamenti eseguiti:

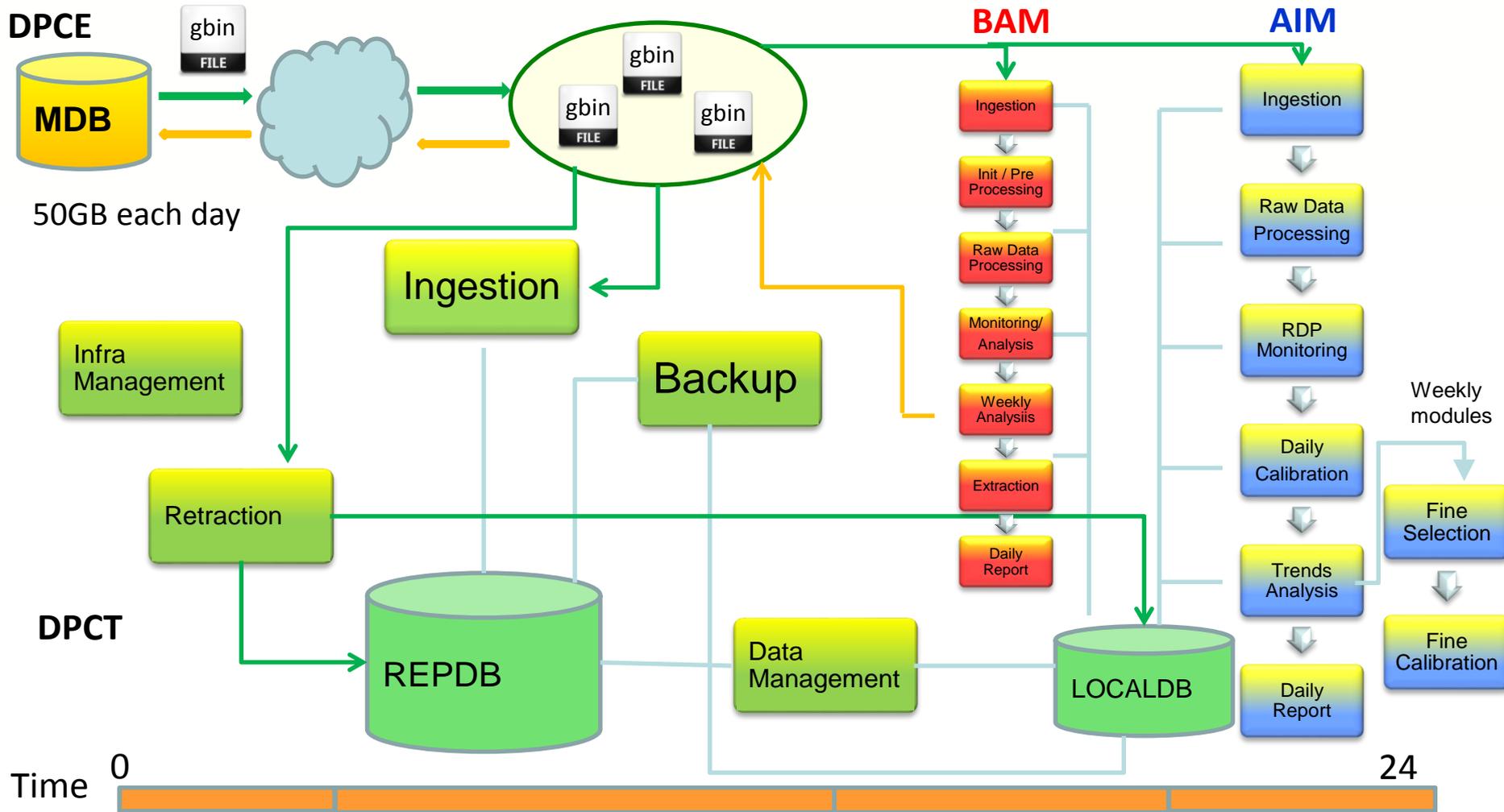
- ✓ **600** run di **BAM**
- ✓ **600** run di **AIM**
- ✓ **75.000** workflow e **7.000.000** job

➤ I dati trattati ad oggi dal DPCT:

- ✓ **50 TB** di dati di input in formato gbin.
- ✓ Database di archivio di **85 TB.**
- ✓ Database di processamento di **20 TB.**
- ✓ Processo di data retraction attivo.

➤ Team operativo:

- ✓ Operatori organizzati in turni coordinati dall'Operations Manager,
- ✓ Verifica giornaliera dei processamenti automatici
- ✓ Costante collegamento con il gruppo scientifico AVU di OATo
- ✓ Briefing settimanali interni ed esterni.
- ✓ **6150** post nel Mission Log
- ✓ Costante coinvolgimento del team di integrazione e del team di ground nelle attività operative
- ✓ Turni di reperibilità al di fuori dell'orario di lavoro (24x7, 365 giorni all'anno)
- ✓ Monitoraggio remoto di HP per la parte HW operativo.



All rights reserved © 2014 - Altec

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.

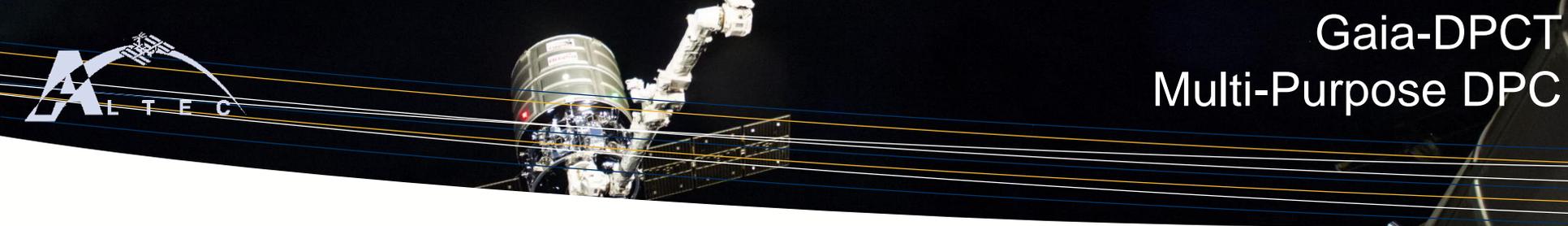
SECURITY SERVICE: redundant firewall based on pfSense, enabling secure remote access via VPN.

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.





➤ Peculiarità dell'architettura del DPCT:

- totale scalabilità
- possibilità di costruirla progressivamente in step successivi in base alle reali necessità.
- HW di tipo «open», cioè gli elementi funzionali di tale architettura sono completamente «vendor independent» (al di là della scelta del fornitore unico HP basato su opportunità tecnico/economiche dei prodotti offerti);.
- Possibilità di estendere e replicare elementi funzionali dell'architettura e soluzioni tecnologiche implementate anche in altri programmi.

➤ Dal punto di vista infrastrutturale e di capacità di hosting, il centro ALTEC è in grado di offrire una opportunità per ulteriori Programmi di Data Processing, mettendo in sinergia alcune soluzioni ed investimenti del Gaia DPCT:

- sala operativa, sistemi di comunicazione, CED, servizi di supporto per le operazioni remote, gestione operativa – con livelli di monitoraggio e manutenzione da parte dei vendors molto stretti, i.e. 6 hrs call to repair 24x7.

- Progettazione ed operazione di una infrastruttura HW «very large» and «reliable», sia in termini di storage (1 PB) che di processamento (oltre 500 cores).
- Utilizzo di tecnologie virtualizzate di gestione dei dati su volumi fisici distribuiti in unità di storage differenti (es. thin provisioning, snapshots, remote copy, disk-to-disk backup, disk-to-tape backup, etc) e di tecnologie di Storage Area Network.
- Realizzazione di un framework SW scalabile e distribuito per il processamento automatizzato ed unattended di SW scientifici.
- Design, sviluppo, implementazione e gestione di «very large and high performance database» basati su prodotti Oracle.
- Design ed implementazione di SW complessi in linguaggio object oriented.
- Design ed implementazione di complessi sistemi di raccolta ed analisi di log (esempio: Elasticsearch, Logstash and Kibana).



ORACLE®

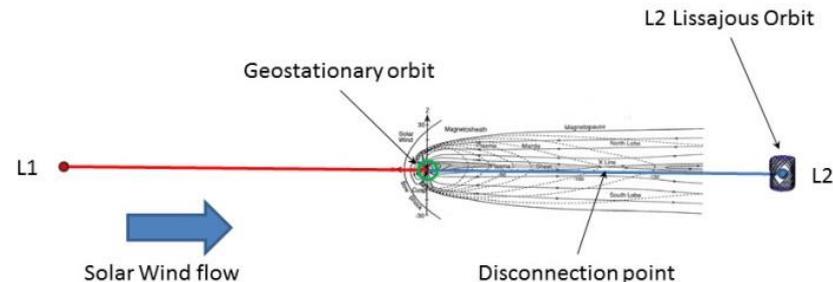
- Tecnologie di virtualizzazione per la gestione dei componenti di servizio dell'infrastruttura.
- High performance network (10 Gbps).
- Sistemi di scambio dati con altri centri su Internet a banda larga (fino a 1 Gbps) e completamente automatizzati.
- Gestione degli aspetti di sicurezza informatica e servizi di accesso remoto all'infrastruttura attraverso VPN e certificati digitali.
- Integrazione end-to-end di servizi di High Performance Computing (HPC) offerti in remoto dal super-computer «Fermi» del Cineca, per eseguire specifiche parti della pipeline scientifica.



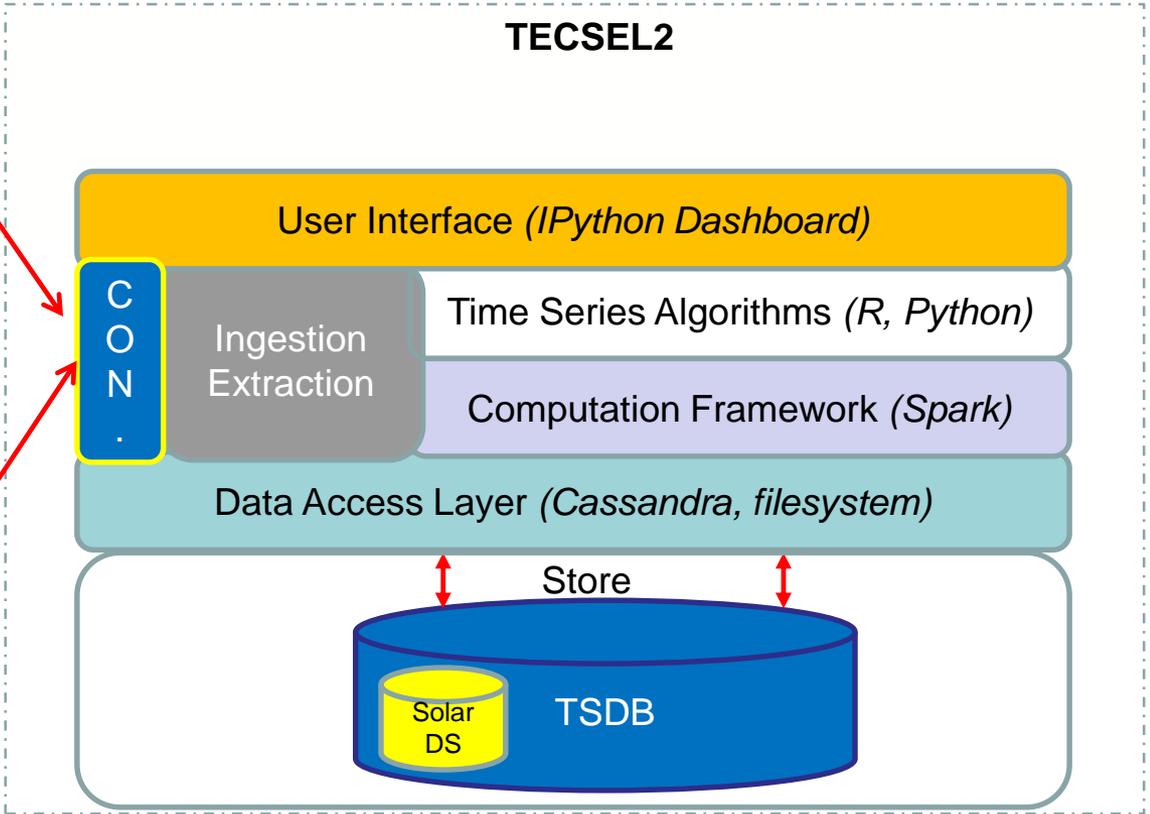
- La tecnologia Oracle RDBMS è utilizzata sia per le pipeline di processamento sia per l'archiviazione dei dati.
 - 2 Oracle RAC (of 3 nodes) database nella versione 11g
 - Nessun problema di prestazioni ed affidabilità riscontrato dopo quasi 2 anni di missione.
 - Configurazione fisica delle strutture dati definita nel dettaglio.
 - Tabelle partizionate su due livelli tramite procedure automatiche eseguite giornalmente
 - Tablespace allocati su tre diversi classi di dischi.
 - LOB allocati su tablespace dedicati.
 - DBA con elevate competenze di gestione di VLDB.
 - La gestione dello spazio disco utilizzato da Oracle è un problema se non gestito.
 - Con l'ottimizzazione della configurazione fisica del db e le attività di consolidamento dati l'utilizzo dei volumi disco dei database del DPCT è passato da un fattore 4 ad un fattore 3, rispetto alla dimensione dei dati di input.
 - I dati gestiti dal DPCT sono sempre online ed accessibile tramite i tool di data access del DPCT.
 - In fase di studio il passaggio alla versione 12c di Oracle.

- I dati di Gaia dovranno essere disponibili ancora per molti anni alla comunità scientifica per l'esecuzione di nuove calibrazioni, a supporto di nuovi esperimenti scientifici e tecnologici (navigazione spaziale, studio del degrado dei componenti elettronici causato dall'esposizione alle radiazioni per lunghi periodi).
- Non è detto che l'approccio ed i metodi adottati per i sistemi di missione siano ottimali in una fase di "exploitation scientifica" dei dati, dove nuovi approcci, metodologie e tecniche saranno richiesti.
- Per questo motivo, ALTEC ha aderito nel 2013 come partner al "**Master on Big Data Engineering**" organizzato dal Politecnico di Torino e finanziato dalla Regione Piemonte (uno dei primi a livello nazionale), acquisendo risorse dedicate orientate alla formazione di nuovi profili tecnici nel settore ("data scientist").
- Il centro del DPCT offre un ambiente integrato, composto di infrastruttura, servizi e personale, a disposizione della comunità INAF per l'utilizzo dell'archivio di missione di Gaia. L'evoluzione della piattaforma per massimizzare l'exploitation dei dati di missione verrà discussa nel prossimo anno.

- TECSEL2 (TEMPoral Characterization of the remote SENSors response to radiation damage in L2) è il primo progetto di ricerca interno interamente sviluppato con tecnologie «Big Data» (NoSQL database, in memory time series processing ecc.)

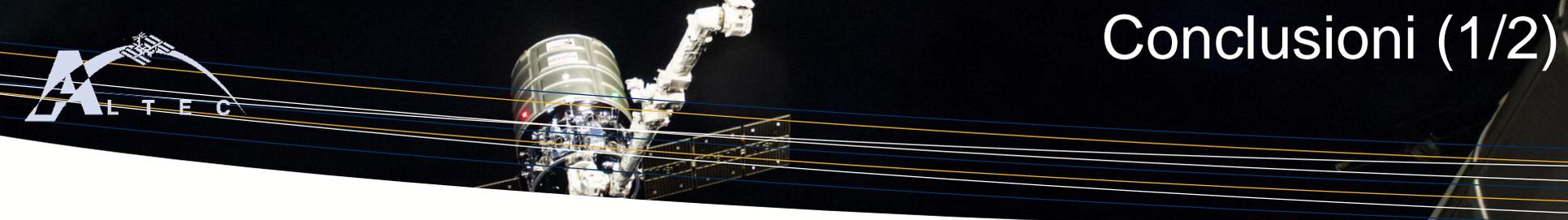


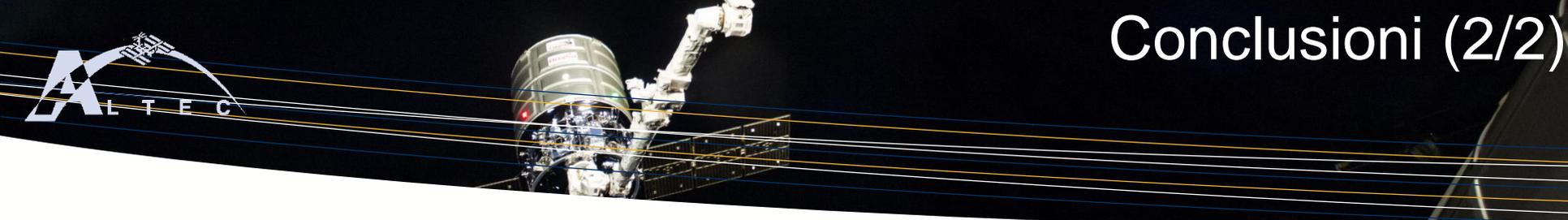
- Obiettivo scientifico: caratterizzazione dei CCDs del piano focale del satellite soggetti a forti danni da radiazioni in un ambiente non protetto dal campo magnetico terrestre, nel punto lagrangiano L2 dove Gaia opera.
- Obiettivo tecnico: avere un ambiente «Big Data» utilizzabile e connesso con l'infrastruttura del DPCT per sperimentare nuove tecniche di data analisi.



Gli algoritmi di TECSEL2 sono stati implementati in R e Python per essere eseguiti sul Spark (distribuzione Cloudera).

- Dal framework di processamento utilizzato dal DPCT e dalle operazioni di missioni di lunga durata condotte in ambito ISS, è nato un progetto interno per lo sviluppo di un sistema di gestione delle pipeline automatiche (ASDP).
- Il sistema fornisce servizi di:
 - gestione e scheduling dei workflow,
 - gestione delle risorse di processamento (cluster, sistemi a code ecc.),
 - gestione delle configurazioni,
 - tracking dei dataset processati,
 - gestione dei metadati dei prodotti delle pipeline,
 - interfaccia verso data store,
 - interfaccia verso sistemi di controllo e monitoraggio esterni (tramite messaggi).
- Il sistema è stato progettato per essere utilizzato in ambito cloud come PaaS.
- Questa soluzione, scalabile, è stata adottata per la processing pipeline dei dati degli esperimenti di bordo di Exomars.

- 
- The top of the slide features the ALTEC logo on the left, which consists of a stylized 'A' with a satellite dish and the letters 'ALTEC' below it. To the right, a satellite is shown in space, with its solar panels and instruments visible against a dark background with some light streaks.
- Per ALTEC la collaborazione con INAF è di particolare importanza in quanto ha permesso e permette di sviluppare competenze tecnologiche all'avanguardia che ci consentono di competere a livello Europeo
 - Lo sviluppo e la gestione operativa del DPCT da parte del team integrato INAF-ALTEC rappresentano un esempio di collaborazione tra enti scientifici ed industria di grande valore e che può costituire una opportunità da considerare anche in altri progetti
 - Il DPCT può essere considerato un “key asset” per ASI, INAF ed ALTEC nei seguenti campi applicativi:
 - Integrazione di database e gestione archivi di grosse dimensioni
 - Supporto alle operazioni scientifiche
 - Architetture e pipeline SW di riduzione dati ad alta complessità
 - High Performance Computing
 - SW per il controllo, monitoraggio e calibrazione di strumenti scientifici.

- 
- Attraverso il DPCT sono state realizzate infrastrutture operative complesse che possono offrire numerose opportunità per:
 - Lo sviluppo di sistemi avanzati e tecnologie di supporto alla ricerca scientifica,
 - Lo sviluppo di iniziative legate alla conservazione dei dati per l'utilizzo scientifico a lungo termine (con capacità di riprocessamento)
 - La fornitura di servizi ed infrastrutture per la comunità scientifica
 - L'obiettivo è massimizzare il ritorno degli investimenti effettuati sia in termini di know-how e sviluppo applicativo, sia in termini di data exploitation.
 - ALTEC sta investendo soprattutto sui giovani. Il team che ha realizzato e sta operando il DPCT è costituito prevalentemente da giovani laureati del Politecnico e dell'Università di Torino.
 - Siamo disponibili a supportare attività di stage o di tesi nell'ambito delle tecnologie IT e di Data Science.
 - Siamo disponibili a valutare la partecipazione a progetti e/o collaborazioni per l'analisi e sviluppo di nuove applicazioni, per lo sfruttamento di dati scientifici anche per applicazioni tecnologiche.