



IL DATA PROCESSING CENTER ITALIANO (DPCT)

Michele MARTINO / Filomena SOLITRO / Rosario MESSINEO
ALTEC S.p.A.
Corso Marche 79
Torino

Chi è ALTEC



ALTEC è 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.

ALTEC è una società pubblico-privata partecipata da Thales Alenia Space (51%), dall'Agenzia Spaziale Italiana ASI (29%) e dal consorzio pubblico ICARUS (20%) costituito dagli enti locali e da Finmeccanica.

ALTEC ha sede a Torino e ha proprio personale distaccato presso NASA-JSC e presso i centri EAC e Col-CC.

I servizi di ALTEC vanno dal supporto ingegneristico e logistico, all'addestramento degli astronauti, al supporto agli esperimenti in particolare di biomedicina, al processamento di dati scientifici, allo sviluppo e alla gestione del segmento di terra di programmi spaziali e alla promozione della cultura spaziale.

Cos'è il DPCT



Il DPCT, **Data Processing Center Torino**, è il DPC italiano dedicato alla missione GAIA, uno dei sei DPCs previsti dal Gaia Data Processing and Analysis Consortium (DPAC).

E' il risultato di uno sforzo **congiunto** tra il team scientifico dell'Osservatorio Astronomico di Torino e del team industriale di ALTEC.

La sua realizzazione è finanziata dall'ASI attraverso un contratto scientifico (INAF-OATo) ed un contratto industriale (ALTEC).

IL DPCT è realizzato ed ospitato nel centro ALTEC di Torino.



Perchè il DPCT



Obbiettivi del DPCT nella missione Gaia

- 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,
- Eseguire l'esperimento GAREQ.

Obbiettivo del DPCT a medio-lungo termine:

- Fornire una infrastruttura, il software, i processi, le reprocessing capabilities, i dati e le competenze a supporto della comunità scientifica Italiana per il massimo utilizzo dei risultati della missione, che vada oltre l'accesso e lo sfruttamento scientifico del catalogo prodotto dal consorzio DPAC (di competenza di ASDC) come estrazione del Mission Data Base di cui il DPCT manterrà copia.

Il design del DPCT ha considerato entrambi questi requisiti fin dall'inizio..

DPCT History



Lo sviluppo del DPCT è formalmente iniziato a **luglio del 2008** con un contratto di INAF ad ALTEC dedicato prevalentemente:

- alla definizione del centro, delle interfacce e dell'integrazione nei confronti del DPAC, e la preparazione della documentazione da sottomettere alla Ground Segment CDR dell'ESA (mag/giu.2009);
- alla predisposizione della piattaforma HW e SW di sviluppo, e ad una prima implementazione prototipale delle funzionalità critiche del SW di infrastruttura;
- alla selezione del DBMS da utilizzare;
- all'approfondimento e sperimentazione delle problematiche relative allo sviluppo ed integrazione del SW scientifico.

Il lavoro è proseguito, da **giugno 2010**, con il contratto triennale di ASI attualmente in corso.

DPCT History



Questa seconda fase è stata dedicata:

- all'implementazione progressiva del SW di infrastruttura, del modello dati specifico del DPCT e dell'implementazione delle interfacce con i SW scientifici;
- all'ampliamento della piattaforma HW / SW di sviluppo e test
- all'integrazione progressiva del SW di infrastruttura con le release dei SW scientifici ed all'esecuzione dei test di integrazione
- alla definizione e partecipazione alle campagne di test integrati a livello DPAC, che hanno progressivamente coinvolto tutti i DPCs:
 - 2b2 test con il DPCE (2010-2011),
 - End-to-End Stage 2 per il processing giornaliero dei dati (AIM / BAM) (2011-2012)
 - End-to-End Stage 3 (fase 2) relativo al processing DRC (GSR) (2012)
 - Operational Rehearsal #01 e #02, relativo alla verifica operativa sia dei sistemi che dei teams (2012)
 - Operational Rehearsal #3 (schedulata per fine Aprile 2013)
 - End-to-End Stage 3 (fase 3) relativa all'esecuzione di tutte le pipeline (AIM, BAM, GSR)

DPCT History

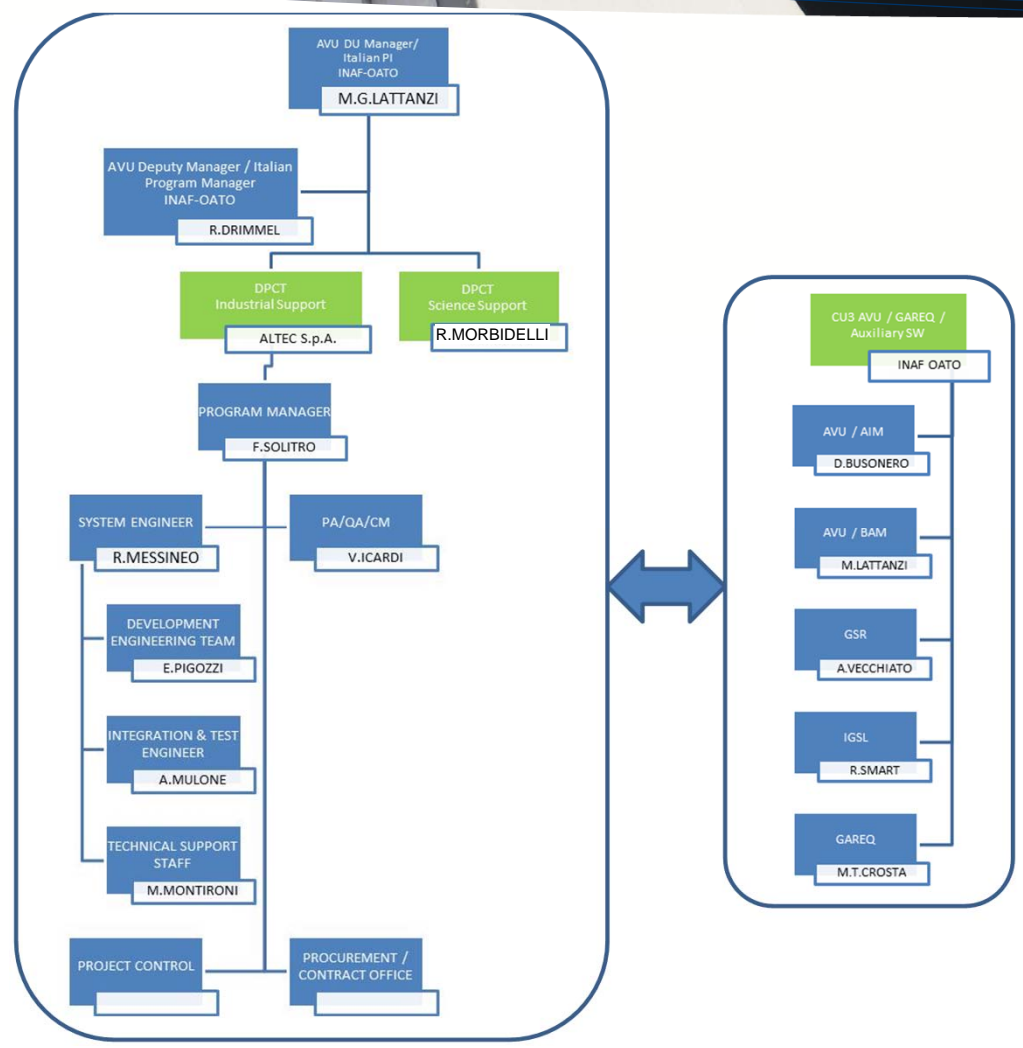


- alla definizione dell'architettura HW / SW della prima versione della piattaforma operativa del DPCT, alla sua acquisizione ed installazione (in corso), con lo scopo di avere il DPCT pronto alla FAR (prevista a giugno-luglio 2013);
- alla preparazione dei piani, delle procedure e del team operativo (sia dell'infrastruttura che scientifico) necessari per l'esecuzione delle operazioni di supporto al commissioning ed alla missione scientifica;
- alla preparazione / partecipazione alle review ESA di programma
 - Ground Segment Implementation Review (2011)
 - Ground Segment Readiness Review (marzo 2013)
 - Final Acceptance Review (prevista giugno-luglio 2013)

Organizzazione team di sviluppo

La figura mostra l'organizzazione dedicata allo sviluppo del DPCT e l'integrazione del team di ALTEC con il team OATo.

Ciò è facilitato anche dalla disponibilità in ALTEC dell'area operativa Gaia DPCT dedicata al progetto a cui il personale OATo ha libero accesso.



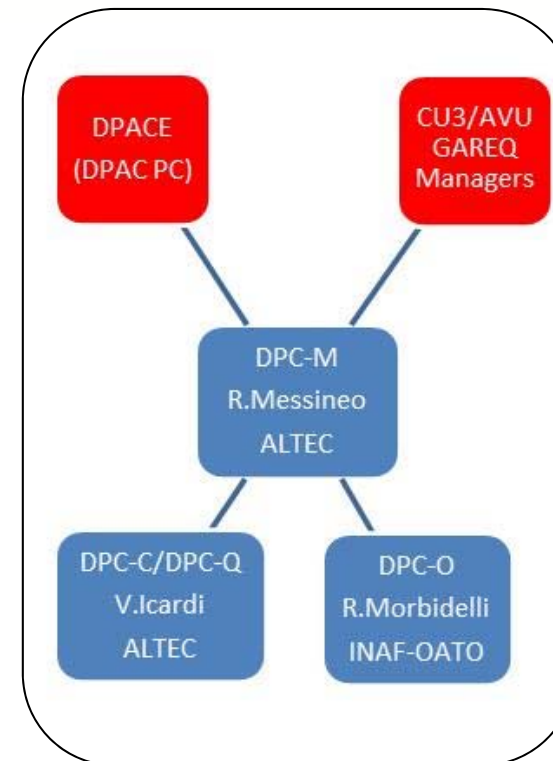
All rights reserved © 2012, - Altec

Organizzazione team di sviluppo



All'interno dell'organizzazione dei DPC in DPAC ALTEC fornisce alcune figure chiave:

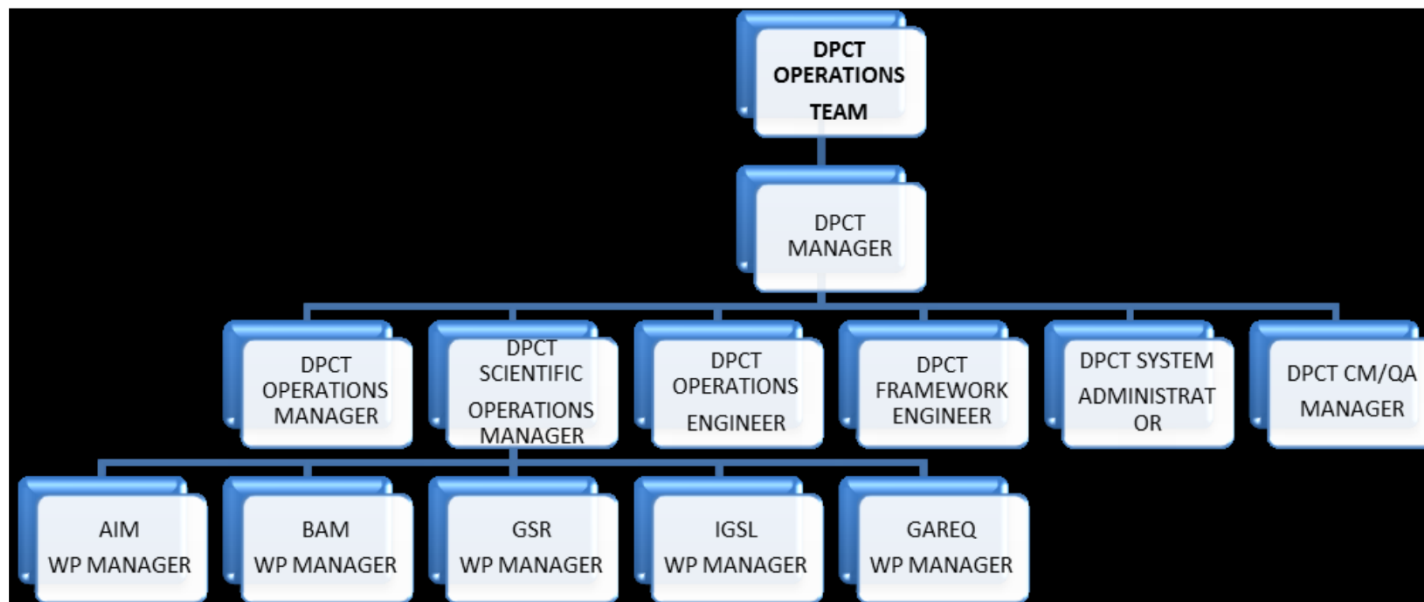
- **Il DCPT Manager:**
ha la responsabilità del DPC e riporta direttamente sia al DPACE ed al DPAC Project Coordinator, sia ai managers della CU3.
- **il DPCT QA/PA:**
ha la responsabilità di tutti gli aspetti di qualità e Produc Assurance del DPC e fornisce supporto in tale area anche ai team scientifici.



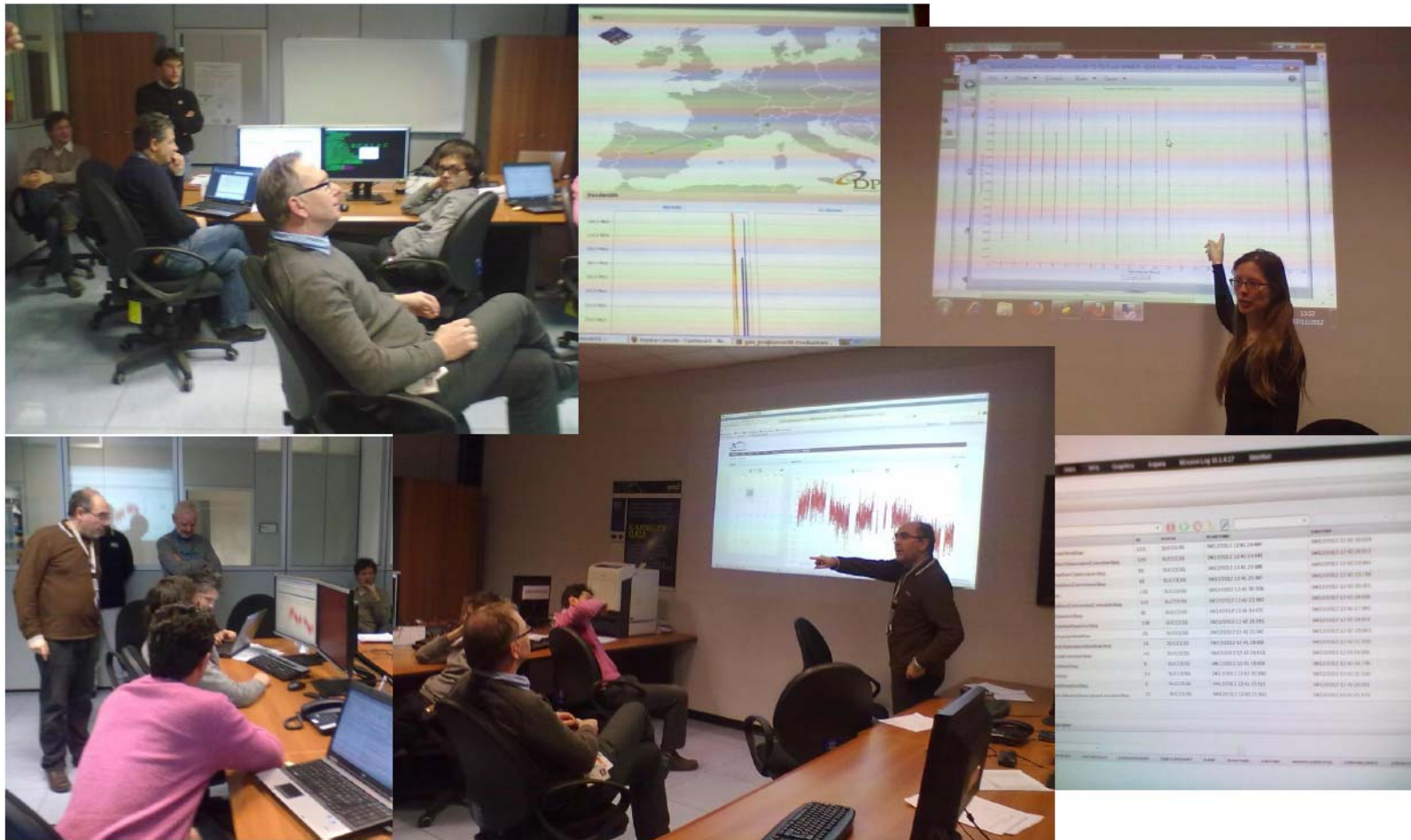
DPCT Team Operativo



- Questa organizzazione evolverà nei prossimi mesi nell'ottica della gestione operativa della missione.
- ALTEC manterrà la responsabilità sull'operatività di tutta l'infrastruttura del centro e sull'interfaccia verso il DPAC (DPCT Manager);
- L'INAF-OATo seguirà gli aspetti scientifici e, attraverso la figura del Scientific Operation Manager, coordinerà i diversi team coinvolti.



DPCT Team durante la OR#02



All rights reserved © 2012, - Altec

DPCT – Problematicità iniziali



- **SW scientifici diversi tra di loro** in termini di:
 - Volume dei dati trattati
 - Quantità di processing richiesto
 - **Complessità e maturità degli algoritmi, che si traduce in differenti schedule implementative, anche con il rischio di revisioni sostanziali del SW in corso d'opera**
 - Modalità operative
- **Quantità di dati da trattare enorme**, dell'ordine delle centinaia di Terabytes, sia strutturati in DB che su file system
- Vincoli sui tempi di processing dei singoli moduli con **alte performance richieste sia in termini di I/O che di esecuzione**
- Missione operativa con tempi lunghi (5 anni di missione + 3 ulteriori di data reduction)
- **Schedule di sviluppo molto compressa**
- Budget limitato, soprattutto per la piattaforma HW e del SW di sistema
- Incertezze sulla data di lancio e sulla dimensione finale del MDB

DPCT – Architettura



- **Prima scelta architetture essenziale è stata quella del DBMS Oracle:**
 - Disponibilità e scalabilità avanzate
 - Gestione di grandi moli di dati anche in ambito scientifico (es. LHC al CERN)
 - Disponibilità di un cluster database : Real Application Cluster (RAC)
 - Possibilità di gestire lo storage in modalità diretta: ASM
 - Disponibilità delle funzionalità di Partitioning per evitare contese nel calcolo parallelo
 - Know-how esistente sia in ALTEC che in INAF-OATo
 - Ottima collaborazione e forte disponibilità in termini di supporto diretto al progetto da parte della società Oracle, anche sul lungo periodo
- Il DBMS non è un elemento comune in DPAC: ogni centro ha fatto la sua scelta, sia in funzione del processing specifico, sia degli accordi esistenti ed il know-how disponibile

DPCT – Architettura

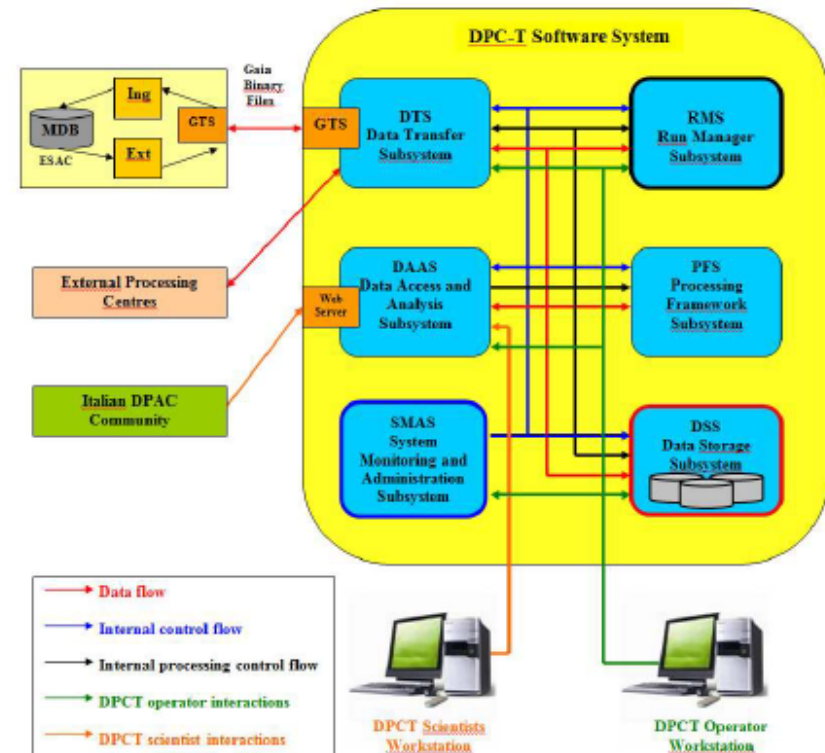


- **Seconda scelta essenziale dell'architettura è il disaccoppiamento e la standardizzazione delle interfacce tra il SW scientifico e il SW di infrastruttura:**
 - Rendere trasparente al team scientifico la complessità dell'infrastruttura
 - Concentrarsi sugli aspetti scientifici e trattamento del dato e non su problematiche tecniche
 - Supportare uno sviluppo in parallelo dei prodotti SW, ciascuno con schedule ed obiettivi differenti
 - Semplificare quanto più possibile l'integrazione incrementale nelle varie fasi del progetto
- **Altri elementi caratteristici:**
 - SW sviluppato in JAVA (standard in DPAC)
 - Massimo riutilizzo di librerie esistenti messe a disposizione dal progetto
 - Utilizzo di tecnologie open source largamente diffuse

Architettura SW



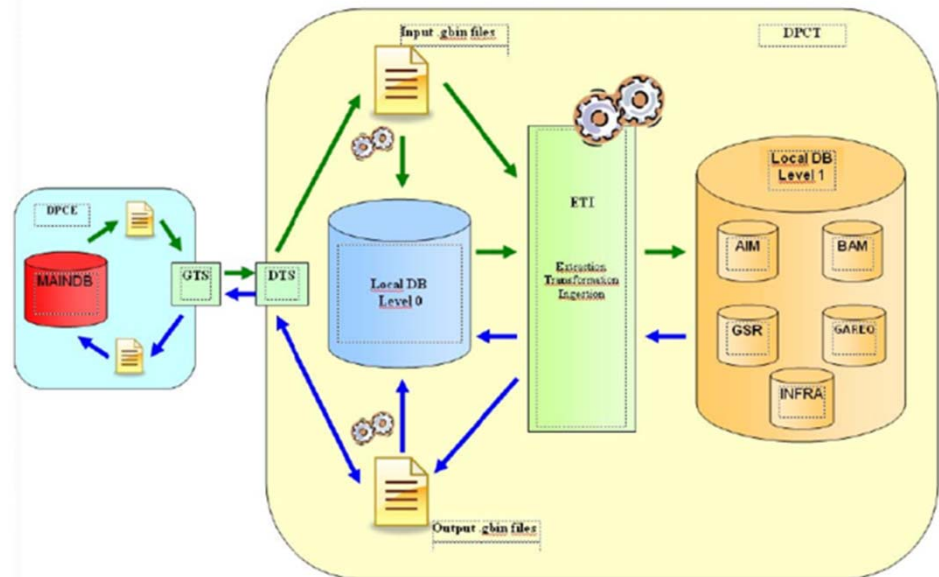
- **DTS**: basato sul COTS Aspera, esteso in termini di funzionalità per la gestione automatizzata delle notifiche sui trasferimenti dati tra i centri permettendo una totale automatizzazione dei processi senza ricorrere a sistemi esterni
- **RMS**: coordina e monitora l'esecuzione dei processi, sulla base del concetto di workflow (sequenza di task). Ogni task viene mappata su uno o più job che vengono poi eseguiti dal modulo PFS.
- **PFS**: esegue e controlla i jobs che contengono gli algoritmi scientifici e l'esecuzione di attività di gestione dell'infrastruttura, distribuendoli sulle risorse di calcolo disponibili (HW processing cluster).



Architettura SW



- **DSS:** Fornisce tutte le funzionalità di accesso al DB. L'architettura prevede due livelli:
 - **livello 0** che rappresenta il repository di tutti i dati ricevuti dal DPCT (speculare del Mission DB di ESAC);
 - **livello 1** (di dimensioni più ridotte) ottimizzato per il supporto ai processing scientifici ed alle informazioni di gestione del sistema
- **SMAS:** integra tutte le funzionalità di gestione e monitoraggio dell'infrastruttura HW/SW del DPCT
- **DAAS:** fornisce ai team operativi gli strumenti di accesso alle informazioni del DPCT, sia in termini di esecuzione dei processi, sia dei prodotti generati, sia per il retrieval e l'analisi dei dati



Architettura HW



L'architettura HW del DPCT prevede:

- **Piattaforma di sviluppo e test**, realizzata in modo incrementale negli scorsi anni, da mantenere in esercizio per tutta la durata del programma a supporto della manutenzione correttiva ed evolutiva del SW, nonché del troubleshooting
- **Piattaforma operativa**, la cui implementazione prevede almeno tre steps incrementali:
 - **Step 1** (in corso), relativo all'infrastruttura necessaria per il supporto al commissioning del satellite e per l'esecuzione della pipeline giornaliera (AIM e BAM). Da completare entro la FAR ad L-4 mesi.
 - **Step 2**, relativo all'infrastruttura necessaria al supporto della pipeline di tipo DRC (esecuzione del GSR) da implementare successivamente al lancio del satellite
 - **Step 3**, relativo al completamento della piattaforma in termini di ampliamento dello spazio di storage e in funzione della post mission data reduction.

Architettura HW



- Nessun vincolo / commonality a livello di DPAC è stato richiesto: ogni DPC ha scelto la sua infrastruttura
- Per il DPCT, le guidelines sono state:
 - Prodotti con performance ed affidabilità elevate
 - Architettura aperta, flessibile ed espandibile
 - Sistemi di storage con alte prestazioni di I/O e alta espandibilità (fino a 1 PB)
 - Modelli ed architetture up-to-date su cui il vendor garantisca supporto sul lungo periodo
 - Massimizzazione del know-how esistente sui prodotti da parte dei team tecnici di ALTEC mantenendo massima commonality con altri progetti in corso
 - Fornitore in grado di offrire alta affidabilità, soprattutto nella fase di post-vendita e di manutenzione, se possibile in grado di coprire tutte le tipologie di elementi dell'architettura, in particolare il sistema di storage
 - ad un prezzo compatibile con i budget disponibili.
- Quanto sopra ha trovato in Hewlett Packard il partner di riferimento per la fornitura dell'infrastruttura HW e dei SW di sistema.

Configurazione HW DPCT Operativo



- **Storage System:** 2xHP 3PAR StoreServ7400 (ciascuno con 4 controller nodes, 72 GB high performance disks and 144 high capacity disks).
- **SAN Switches:** Redundant HP 8/80 (8Gbps I/F)
- **Database server:** 4x HP DL570G7, 4 CPU with 8-cores each and 256 GB RAM
- **Processing server :** 4 x HP DL570G7, 4 CPU with 8-cores each and 256 GB RAM
- **Management server :** 2x HP DL570G7, 4 CPU with 8-cores each and 128 GB RAM (virtual environment)
- **Specific Processing server:** SGI Altix 450: 8 CPU Dual Core, 256 GB RAM
- **Backup System :** HP ESL G3 Tape library, 4xLTO-5 Drives, 100 tapes)
- **Network:**Redundant Cisco 4500X switch (with 10Gbps ports), and a cluster of four Cisco 3750X switches
- **User WS:** 4xHp Z820 user workstations
- **Firewall :**based on Pfsense installed on 2xHP DL380G8
- **Internet Link :** bandwidth provided by GARR (nominally 300 Mbps max, scalable up to 1 Gbps)



Configurazione HW Sviluppo e Test



- HP EVA 4400 Storage Array, dual controller, con 48 TB su dischi ad alta capacità, e 18 TB su dischi ad alte prestazioni
- SAN Switch in comune con la piattaforma operativa
- HP Proliant Server Processing/Database
 - 5 HP DL580 G7 (4px4core, 128GB RAM)
 - 2 HP DL380 G8 (2px8core, 128GB RAM)
 - 1 HP DL380 (management server)
- SGI Altix 450: 8 CPU Dual Core, 256 GB RAM (in comune alla piattaforma operativa)
- Backup system, Network, firewall e connessione Internet in comune con la piattaforma operativa.



Configurazione HW DPCT Operativo



- Installazione del sistema è attualmente in corso in collaborazione con il team HP
- Tutto l'HW durante la vita operativa avrà un SLA per problemi HW del tipo «6 Hr call to repair»
- L'HW verrà configurato per inviare informazioni di performance e log ai sistemi HP che potranno effettuare un monitoraggio ed analisi remota del comportamento del sistema
- Ciò avrà lo scopo di anticipare quanto più possibile l'identificazioni dell'insorgere di potenziali problemi HW.
- A valle del completamento dell'installazione saranno anche riviste le procedure operative da utilizzare da parte del team di ALTEC per la gestione delle varie situazioni di failure ed in particolare la gestione delle «richieste di escalation» nella risoluzione dei problemi di particolare complessità.



Conclusioni



Il DPCT si sta progressivamente avvicinando all'appuntamento del lancio del satellite Gaia secondo i piani previsti.

Questo è il risultato del lavoro:

- delle persone dell'ASI, che hanno coordinato tutti gli aspetti programmatici e contrattuali, permettendo il rispetto delle milestone del programma ed evitando che questi potessero impattare la già compressa schedule di attività;
- del team scientifico dell'OATo e del team di ALTEC che hanno saputo integrarsi, mettendo a fattor comune la cultura e l'approccio scientifico con la cultura e l'approccio industriale, perseguendo insieme lo stesso obiettivo.

Conclusioni



Il Gaia DPC di Torino 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.

Il Centro ha inoltre le potenzialità per:

- Lo sviluppo di sistemi avanzati e tecnologie per il 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.

