



INAF

Scenario per lo sviluppo e il supporto alle attività cloud

Fabio Pasian, Claudio Vuerli
INAF – O.A.Trieste



Istituto Nazionale di Astrofisica

- 16 Strutture di ricerca (Osservatori Astronomici e Istituti ex-CNR) spesso ospitano personale Universitario
- una ventina di sedi; due sedi osservative a La Palma (isole Canarie) e Mount Graham (Arizona, USA)
- circa 1100 dipendenti, di cui circa 700 sono ricercatori e tecnologi;
- oltre al personale di ruolo, circa 500 ricercatori sono associati a INAF e operano nelle Strutture dell'Ente con forme diverse di collaborazione: di questi, oltre 300 sono giovani con contratti a tempo determinato (TD, assegni e borse di ricerca).



Progetti

- La comunità partecipa attivamente e spesso occupa ruoli di coordinamento all'interno di grandi progetti:
 - ESFRI (CTA, E-ELT, SKA)
 - missioni spaziali del programma “Cosmic Vision 2025” dell’Agenzia Spaziale Europea (Euclid, Solar Orbiter, JUICE)
 - molti altri progetti non meno impegnativi sia scientificamente che per risorse impiegate (es. LST)
- Per la realizzazione dei loro obiettivi scientifici, tali progetti affrontano notevoli sfide tecnologiche sia nel campo della costruzione strumentale, che in quello delle infrastrutture digitali.
- I progetti sono intrinsecamente data-intensive e pongono sfide computazionali di difficile soluzione.



Alcune considerazioni

- Qualche numero:
 - SKA avrà un data flow di 950PB/giorno di dati raw, 80GB/sec processati
 - Euclid elaborerà > 15 PB di dati in modo distribuito su > 9 Science Data Centres nazionali (cataloghi: 10^{10} oggetti, 1500 par)
- Nel caso di strumenti ottici/IR, vengono conservati TUTTI i dati, raw e calibrati → E-ELT
- L'astronomia è una scienza osservativa
- C'è una lunga tradizione di standard (FITS), storage («SaveTheBits»), archiviazione (dati e metadati), dati pubblici, accesso e interoperabilità (Osservatorio Virtuale)
- Anche il calcolo dei singoli astronomi è data-intensive (elaborazione/analisi dei propri dati)
- Simulazioni (la missione Planck è stata simulata > 200 volte); modellistica; calcolo teorico; fluidodinamica



Infrastrutture informatiche

- INAF ha infrastrutture piccole/medie di storage e di calcolo ad alte prestazioni
 - cluster paralleli,
 - alcuni sistemi basati su GPU e ibridi CPU-GPU.
- Alcuni di questi, per lo più finanziati dal MIUR tramite PON, sono interconnessi alle infrastrutture Grid italiana (IGI) ed europea (EGI).
- Risorse di calcolo utilizzate
 - per scopi di ricerca interni all'INAF,
 - per offrire servizi nell'ambito delle attività operative (missioni spaziali, telescopi a terra)
 - in parte dai dipartimenti/sezioni di astronomia delle università ospitati presso le Strutture dell'Ente,
 - come cofinanziamento in progetti europei e nazionali compartecipati dall'INAF.
- L'INAF non eroga servizi di calcolo ad alte prestazioni verso l'esterno se non nell'ambito di progetti di ricerca con partner esterni, ma usa le infrastrutture per i suoi scopi istituzionali e di ricerca.



Attività in informatica applicata

- Studio dei sistemi HPC e Grid/Cloud finalizzato alle tematiche di ricerca in astronomia e astrofisica:
 - elaborazione di grandi quantità di dati osservativi,
 - produzione di modelli, fluidodinamica,
 - gestione e l'archiviazione dei dati,
 - data mining,
 - visualizzazione,
 - information retrieval.
- Queste attività hanno portato anche alla progettazione e realizzazione di sistemi software per il calcolo, librerie e ambienti di sviluppo ed esecuzione di applicazioni ad elevato parallelismo, da realizzarsi su sistemi ad alte prestazioni e/o tramite calcolo distribuito.



e-infrastrutture multi-disciplinari

- INAF è attiva nel campo delle infrastrutture distribuite di calcolo e di archiviazione/gestione dati da oltre dieci anni
- partecipazione a progetti finanziati da EU/FP
 - osservazioni remote e presenza virtuale → REMOT, EVALSO
 - interoperabilità tra gli archivi di dati astronomici distribuiti a livello mondiale → EuroVO, EuroPlanet, EGSO, HELIO, Cassis
 - realizzazione di infrastrutture di calcolo distribuito e relative applicazioni all'astrofisica → EGEE, EGI-Inspire, SciBUS, ER-flow
- definizione del modello computazionale per CTA (ESFRI) e organizzazione del relativo Virtual Team finalizzato alla creazione dei gateway scientifici e meccanismo di single-sign-on
- chair delle organizzazioni mondiali IVOA International Virtual Observatory Alliance (2007-10) e IPDA International Planetary Data Alliance (2007-09)
- coordinamento della comunità Astrofisica in EGEE-III ed EGI



Altre considerazioni

- Nonostante siano grandi produttori di dati, gli astronomi non sono in media skilled nell'uso di computer; c'è una cultura informatica adeguata solamente all'interno di piccoli gruppi
- Grid è stata un'esperienza sostanzialmente deludente:
 - per le risorse da impegnare (sysadmin)
 - per la ripida curva di apprendimento
 - per l'assenza di supporto domain-specific al porting
 - per un fit molto parziale con le esigenze della comunità
- Cloud appare invece più interessante:
 - per il concetto di accesso trasparente
 - soprattutto per i piccoli gruppi con grandi requisiti di calcolo
 - per i progetti, se include l'intera piramide (incluso HPC)



Infrastruttura digitale nazionale 1

- E' urgente la realizzazione e messa in opera di servizi multidisciplinari e geograficamente distribuiti, per poter rispondere nel modo più efficace alle sfide scientifiche e tecnologiche posti dai progetti di nuova generazione.
- E' necessario un forte coordinamento tra istituzioni → va sfruttato il nuovo ruolo che GARR si trova a svolgere all'interno della comunità scientifica.
- Oltre all'ovvia necessità di condivisione di risorse che portino a un'ottimizzazione delle capacità di calcolo e di gestione distribuita dei dati, ci sono svariati aspetti tecnologici che solo un coordinamento inter-Ente può realizzare in modo efficace; esempio Single-Sign-On.



Infrastruttura digitale nazionale 2

- Applicazioni di calcolo ed elaborazione dati sono ampiamente diversificate, in termini computazionali e data flow.
- Da semplici cluster locali di fascia bassa a grosse infrastrutture costituite da risorse di calcolo del tutto eterogenee (anche con architetture diverse).
- La soluzione tecnologica che è meglio rispondente a queste necessità è un ambiente integrato basato su una rete digitale ad alta efficienza che permetta all'utente di accedere in modo trasparente a risorse federate ed eterogenee (HPC-HTC-dati) tramite una infrastruttura digitale all'uopo costruita e mantenuta a livello operativo ("Cloud per la Ricerca").
- Indispensabile modificabilità ed espansione nel tempo.



Infrastruttura digitale nazionale 3

- E' indispensabile
 - procedere a uno stretto coordinamento tra i Consorzi di supercalcolo (CINECA 2.0), il Consortium GARR e gli Enti di Ricerca (almeno INFN, ENEA, INAF, CNR, INGV, OGS), più altri partner.
 - realizzare la necessaria sinergia che permetta all'infrastruttura digitale multidisciplinare e geograficamente distribuita di essere realizzata e messa in opera tempestivamente



Infrastruttura digitale nazionale 4

INAF ribadisce:

- la volontà di utilizzare l'infrastruttura digitale esistente, propria e di altri Istituti, in modo coordinato allo scopo di ottimizzare l'utilizzo delle risorse;
- la volontà di potenziare centri anche non-INAF che garantiscano un'operatività continua, con risorse da condividere all'interno della futura infrastruttura digitale;
- la volontà di condividere risorse INAF esistenti, e da acquisire in futuro, all'interno di un sistema coordinato di Cloud per la Ricerca;
- la disponibilità a mantenere operative le risorse contribute all'infrastruttura digitale condivisa;



Infrastruttura digitale nazionale 5

INAF ribadisce inoltre:

- la disponibilità a supportare con proprie risorse umane
 - lo sviluppo tecnologico della Cloud per la Ricerca secondo le necessità stabilite con gli altri partner;
 - il training del proprio personale di ricerca e tecnico finalizzato all'utilizzo e alla gestione, rispettivamente, dell'infrastruttura digitale condivisa;
 - il mantenimento e lo sviluppo di competenze specialistiche disponibili nei propri centri e la pianificazione concertata, in base alle esigenze nazionali, degli investimenti di potenziamento formativo (borse di dottorato, borse di studio, ...) per la crescita delle stesse.
- la volontà di pianificare in modo coordinato, in base alle esigenze comuni a tutti i partner, degli investimenti di potenziamento infrastrutturale e formativo;
- la volontà di partecipare, congiuntamente agli altri partner, alle iniziative nazionali ed europee nel campo, incluse le domande di finanziamento sui canali opportuni.