

# DHTCS-IT

## Distributed High Throughput Computing and Storage in Italy

Giuliano Taffoni – Fabio Pasian – Claudio Vuerli

**Istituto Nazionale di Astrofisica – O.A.Trieste**

Workshop ICT INAF, Pula (Cagliari), 16-19 Settembre 2014

# Il progetto DHTCS-IT

Finanziamento delle attività di ricerca a valenza internazionale 2013 (MIUR):

« Il progetto ha l'obiettivo di **consolidare e far evolvere** la attività che è stata realizzata nell'ambito di **IGI** in un quadro di portata internazionale. Il progetto intende, inoltre, riconfigurare sul panorama nazionale le attività nel campo del calcolo distribuito e agire da **"collante" tra le varie iniziative esistenti** (o pianificate) previste per specifici ambiti territoriali o per comunità di utilizzatori definiti, come i Progetti **PON**, i Progetti **POR**, i progetti **Smart Cities** e **Cluster Tecnologici** e nel campo internazionale i progetti **ESFRI**. »



# Finanziamento di DHTCS-IT

- Gestito da INFN, con partecipazione CNR, INAF, INGV e GARR (tramite INFN)
- Finanziamento MIUR 2013 come attività di ricerca a valenza internazionale (es. al pari di SKA, E-ELT e SRT per INAF)
- totali: 3.500.000 € → 3.150.000
  - per INAF: 350.000 € → 315.000
  - Responsabile progetto: Dott. Fabio Pasian



# Obiettivi di DHTCS-IT

- potenziare e far evolvere l'infrastruttura IGI
- fornire servizi HTC e **Cloud** ai gruppi di ricerca dei soggetti partecipanti (CNR, INAF, INFN, INGV, GARR)
- supportare in modo adeguato, dal punto di vista infrastrutturale, i progetti ESFRI dei soggetti partecipanti
- coordinamento e integrazione con altri progetti (es. GARR-X Progress)





# Italian Grid Infrastructure

- Infrastruttura Italiana Grid
- Partner dell'European Grid Initiative
- E' la NGI Italiana sino a fine 2014
  - *"legal organizations, supported by governments, and providing a unique representation at the European and international levels for all the communities related to national grid infrastructures: from resources providers to scientific users"*
- Era una associazione temporanea di scopo (INFN, INAF, ENEA, etc)
- L'infrastruttura è oggi gestita dal Dip. Calcolo del GARR.
- Gestisce: > 15000 Cores e storage distribuiti tra 50 centri tra cui alcuni INAF (CT, TS)
- Gestisce i servizi di monitoring e coordinamento.



# Da Grid a Cloud

- EGI → EGI federated cloud
  - *a seamless grid of academic private clouds and virtualised resources, built around open standards and focusing on the requirements of the scientific community.*
- Sviluppata anche con la collaborazione di INFN (e.g. Bari, CNAF, etc)
- IGI → supporto a cloud resources



# High Throughput Computing?

- *“use of many computing resources over long periods of time to accomplish a computational task”*
- *“a computing paradigm that focuses on the efficient execution of a large number of loosely-coupled tasks”*
- In pratica molti Astronomi usano e necessitano di HTC...



# Cloud Computing – for whom

- Total control over deployed applications
- Elastic resource consumption based on real need
- Immediately processed workloads – no more waiting time
- Service performance scaled with elastic resource consumption





# DHTCS-IT: Nuovi Servizi

- DHTCS-IT come evoluzione/estensione dei servizi offerti da IGI
- Evoluzione verso servizi cloud open integrati in Grid o gestiti in modo indipendente/complementare
- Rafforzamento della vocazione multidisciplinare dell'infrastruttura



# DHTCS-IT: Nuova tecnologia

- DHTCS-IT come evoluzione tecnologica di IGI
- Evoluzione del middleware, integrazione delle applicazioni e gestione dei servizi comuni
- Risorse hardware che permettano la costituzione di un'infrastruttura pilota nazionale in grado di offrire risorse di calcolo e di storage condivise



# Come realizzarlo?

- Principalmente basato su Open Source
  - Sistema Operativo Linux
  - Middleware Cloud OpenStack
  - File system(s): GlusterFS/Lustre/GPFS
  - Personal Cloud storage: ownCloud
  - Federazione con orchestratore (es. CLEVER) e standards (es. OCCl, CDMI)
- Abbiamo bisogno di una rete efficiente: GARR-X e sua evoluzione



# DHTCS-IT@INAF



- INAF supporta la creazione di e-infrastruttura pluridisciplinare italiana (position paper Nov 2012)
- Realizzazione di un nodo DHTCS a INAF-OATS, integrato con l'infrastruttura Italiana



# Budget e Infrastruttura

- ~300 Mila Euro
- ~800 Cores (64 core x Motherboard su 4 sockets)
- Xeon Phi coprocessors
- Tesla K20X NVIDIA GPUs
- 500TB di High performance parallel FS (Lustre/GPFS)
- 40Gbps InfiniBand



# Chi potrà usarlo

- Progetti (Euclid, SKA, PRIN, etc.)
- Singoli ricercatori
- Gruppi di ricerca



# Come usarlo

- Piccola risorsa HPC per test e sviluppo
- Risorsa HTC per i ricercatori italiani
  - Risorsa HTC per Astronomi INAF
  - Integrata con servizio Archivi IA2
- Risorsa Cloud: IaaS
- Riusciremo a farli convivere?



# Risorsa Cloud

- Basata su Open standards
- Integrata con Owncloud
- Modello di IaaS
  
- Perché:
  - sw preservation (e.g. Midas)
  - Infrastruttura integrata
  - Partner di EGI fed cloud





# DHTCS-IT@INAF

- Un nodo di una rete pluridisciplinare multi-ente
- Intende essere partner di EGI FedCloud
- Intende essere (a scopo sperimentale) partner di CANFAR (Cloud canadese per l'astronomia)
- Sarà integrato con IA2 (totale **integrazione tra archivi, storage e calcolo**)
- Potrebbe essere il seme di una e-infrastructure INAF, composta da più nodi integrati.

