CHIPP: the INAF pilot project for HTC, HPC and HPDA

G. Taffoni, U. Becciani, B. Garilli, G. Maggio, F. Pasian, G. Umana, R. Smareglia, F. Vitello

ABSTRACT

CHIPP ("Computing HTC in INAF – Pilot Project") is an Italian project funded by the Italian Institute for Astrophysics (INAF) and promoted by the ICT office of INAF. The main purpose of the CHIPP project is to coordinate the use of, and access to, already existing high throughput computing and high-performance computing and data processing resources (for small/medium size programs) for the INAF community. Today, Tier-2/Tier-3 systems (1,200 CPU/core) are provided at the INAF institutes at Trieste and Catania, but in the future, the project will evolve including also other computing infrastructures.
During the last two years more than 30 programs have been approved for a total request of 30 Million CPU-h. Most of the programs are HPC, data reduction and analysis, machine learning.
In this poster, we describe in details the CHIPP infrastructures and the results of the first two years of activity.

Computing and Data Infrastructures

The HPC/HTC HOTCAT Cluster at OATs

- **1400 INTEL Haswell E5-4627v3 cores**
- **6GB RAM/Core (8.5TB RAM total)**
- **500 TB parallel storage (BeeGFS)**
- **Storage Nodes INTEL Haswell E5-4627v3 256GB RAM per node**
- **Infiniband ConnectX -3 Pro Dual QSFP+ 54Gbs**

400 Cores out of 1400 are fully dedicated to CHIPP.

The CloudCAT Cluster at OATs is a OpenStack stack with a swift object storage

- **200 INTEL Westmere E5620 @ 2.40GHz cores**
- **8GB RAM/Core**
- **75 TB Object storage**
- **Infiniband 10Gbs**

It is provided to users with Virtual Machine pre-configured with Astronomical Software (e.g. ESO Scisoft) and remote desktop capabilities to allow easy access and usage. It is compliant with EGI-Federated cloud resources.

The HTC MPU Cluster at OACt

- **384 INTEL E5-2620; V2 @ 3.10GHz**
- **5.2GB RAM/Core (1TB RAM total)**
- **70 TB parallel storage (NFS)**
- **10 Gbs ethernet network.**

The Interconnect allows high throughput and low latency (1 microsec).
Storage is distributed on 4 nodes and guarantees 2 GB/sec of throughput.

All the computing infrastructures are equipped with more than 60 software environment for Astronomical **data reduction** and analysis. They offer tools for software development, **profiling and debugging**.

Scope

CHIPP is a pilot project to **develop a distributed computing infrastructure** based on existing computing resources for Italian Astronomers.
CHIPP main goals are:

- Investigate to what extent the Italian astronomical community is interested in using a distributed computing infrastructure;
- Investigate whether the infrastructure meets the Astronomers requirements and collect comment and suggestions to optimize it;
- Investigate the best approach to offer the resources (e.g. periodic call, rolling based, on demand access)
- Implement an efficient user support and investigate which support level is required (system management, support for software development, support for optimization and debugging)

This project is the **prototype and test case for the development of a large distributed computing infrastructure** for astronomers able to offer multiple resource: HPC, HTC, Cloud.

Implementation

- CHIPP coordinates the services offered by the computing centres.
- CHIPP monitors the various phases of the project verifying that the INAF community has benefited from the offered services.
- CHIPP implements and coordinates a distributed supports service that takes care of system support and installation of new libraries and tools.
- CHIPP provides customized development and execution environments for users.
- CHIPP is offering cloud and containers to build science platforms for users and small projects.
- CHIPP is promoting periodic competitive calls for computing resources, one each 6 months for large projects (~80000 core hours) and a rolling open calls for small project (<10000 core hours) of one month.
- Projects supported by CHIPP are chosen on the basis of innovation potential, scientific excellence and relevance criteria.

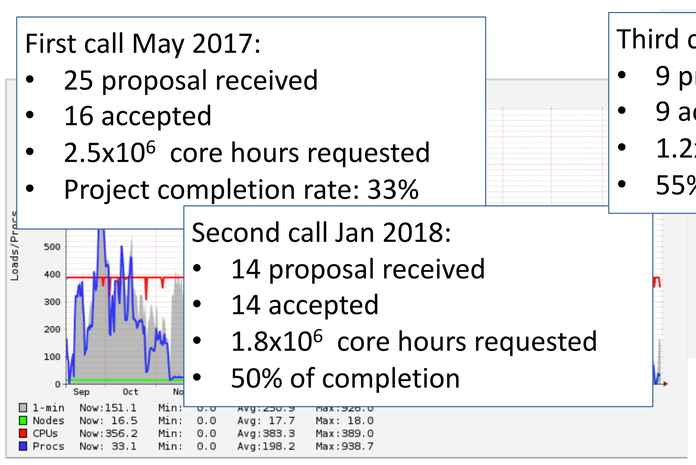
Since may 2017 CHIPP is offering $2.5x10^6$ core hours each ~6 months for INAF

First call May 2017:
- 25 proposal received
- 16 accepted
- $2.5x10^6$ core hours requested
- Project completion rate: 33%

Second call Jan 2018:
- 14 proposal received
- 14 accepted
- $1.8x10^6$ core hours requested
- 50% of completion

Third call Jul 2018:
- 9 proposal received + 3 "rolling"
- 9 accepted
- $1.2x10^6$ core hours requested
- 55% of completion

Fourth call May 2019:
- 13 proposal received
- 13 accepted
- $1.4x10^6$ core hours requested
- On-going



Considerations

- The number of proposal per call decreased with time but it is now steadily around 10 proposal approved each call. This corresponds to greater user awareness of what a computing resource can offer in terms of capabilities and environment.
- The user support service has been extremely successful, however we collected several requests for software debugging and optimization that requires a specialized time consuming support.
- The rolling open calls were also successful: they allow astronomers with an un-planned need of computing resources to easily access to the infrastructure.

https://www.ict.inaf.it/computing/

Conclusions and perspectives

The INAF CHIPP pilot project is offering **computing and storage resources** to astronomers and projects on a call based approach.
CHIPP is **offering support** to astronomers for the use of the resources and to build science platforms based singularity containers, Jupiter notebooks and/or remote desktop.
CHIPP is a successful experiment to **define requirements**, investigate the type of use and necessities of a broad astronomical community.
The experience gained during the project and the requisites collected, are the foundations for the **design and development of a large distributed computing infrastructure for astronomy** in Italy (Tier-1 like)**: the Data-Star project.**
The new infrastructure will benefit of the **European Open Science Cloud** environment and **EuroHPC** project.