



Finanziato
dall'Unione europea
NextGenerationEU



Ministero
dell'Università
e della Ricerca



Italiadomani
PIANO NAZIONALE
DI RIPRESA E RESILIENZA



INAF
ISTITUTO NAZIONALE
DI ASTROFISICA

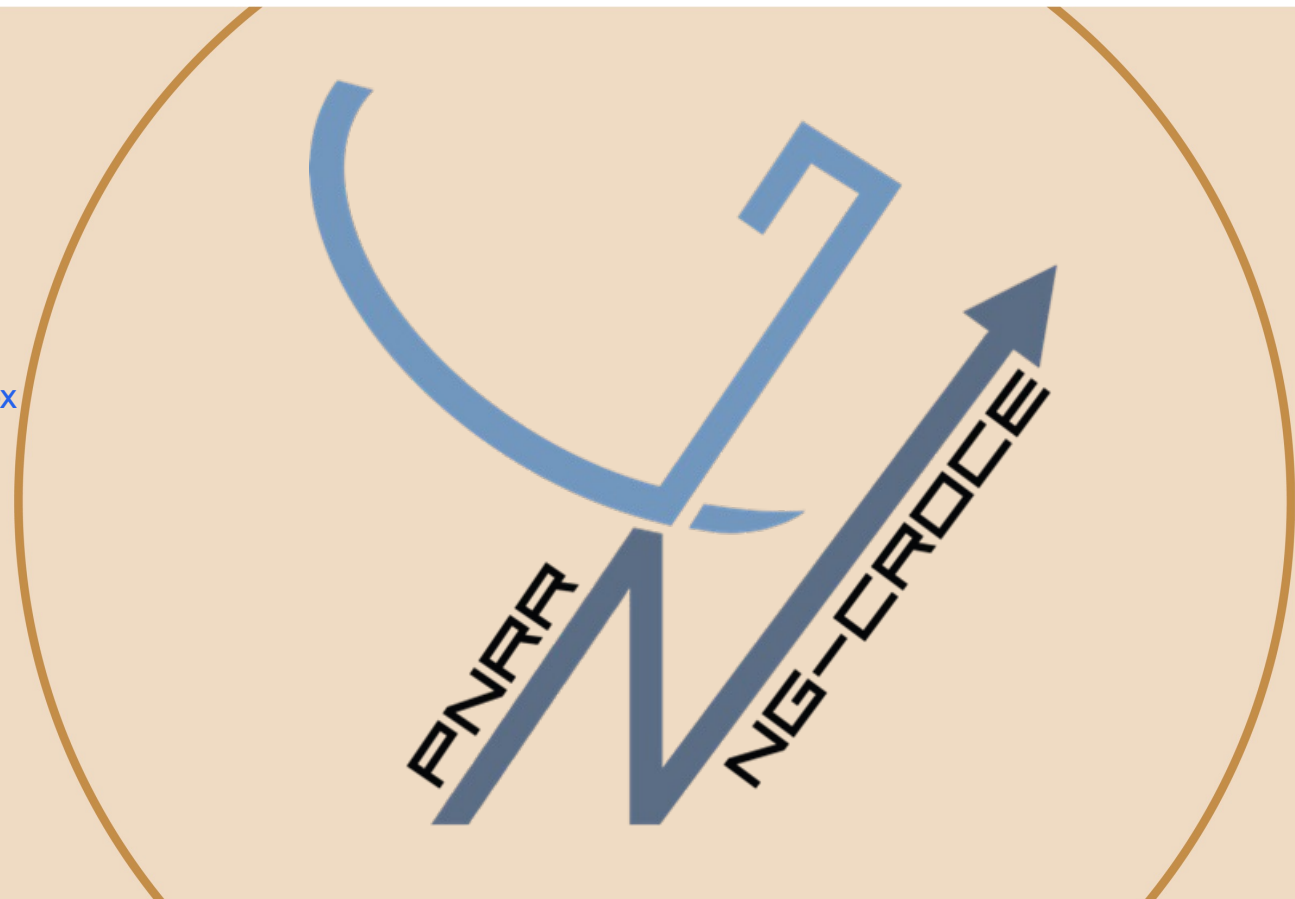
Cluster Provisioning & Scheduling

Architetture Scalabili con Slurm, Ceph e Proxmox

Final Meeting

Martedì 30 Giugno – Giovedì 02 Luglio

OAC - Osservatorio Astrofisico di Arcetri



Giacomo Lorenzo



Finanziato
dall'Unione europea
NextGenerationEU



Ministero
dell'Università
e della Ricerca



Italiadomani
PIANO NAZIONALE
DI RIPRESA E RESILIENZA



INAF
ISTITUTO NAZIONALE
DI ASTRONOMIA

I PILASTRI DELL'INFRASTRUTTURA



Architettura

Login nodes, Compute nodes e il ruolo di Proxmox Virtualization.



Provisioning

Deployment automatizzato delle risorse e gestione dello storage con Ceph.



Scheduling

Ottimizzazione dei job attraverso lo scheduling e High Availability.



Finanziato
dall'Unione europea
NextGenerationEU



Ministero
dell'Università
e della Ricerca



Italiadomani
PIANO NAZIONALE
DI RIPRESA E RESILIENZA



INAF
ISTITUTO NAZIONALE
DI ASTROFISICA

NODI DI LOGIN VS NODI DI CALCOLO

Login Nodes (Frontend)

- Punto di accesso unico per gli utenti tramite SSH.
- Sottomissione job, compilazione codice e editing.
- Non esegue calcoli pesanti (prevenzione crash).

Compute Nodes (Backend)

- Dedicati esclusivamente all'esecuzione dei job.
- Nessun accesso diretto esterno.
- Ottimizzati per performance CPU/GPU e RAM.



Interconnessione

La separazione garantisce la stabilità sistema. Proxmox gestisce questi nodi istanze isolate con risorse garantite (H Passthrough se necessario).



Finanziato
dall'Unione europea
NextGenerationEU



Ministero
dell'Università
e della Ricerca



Italiadomani
PIANO NAZIONALE
DI RIPRESA E RESILIENZA



INAF
ISTITUTO NAZIONALE
DI ASTRONOMIA

PROXMOX VE E HIGH AVAILABILITY

Proxmox funge da **Hypervisor** per l'intero cluster, permettendo una gestione flessibile dei nodi virtuali.

Quorum & HA

In caso di guasto hardware di un nodo fisico, le VM (Login o Slurm Controller) vengono migrate automaticamente su nodi sani grazie allo stack HA integrato.

- **KVM/LXC:** Mix di macchine virtuali e container leggeri
- **Web GUI:** Monitoraggio centralizzato dei nodi.
- **Live Migration:** Spostamento dei carichi, uguale a meno downtime.
- **Fencing:** Protezione contro il "split-brain"



Finanziato
dall'Unione europea
NextGenerationEU



Ministero
dell'Università
e della Ricerca



Italiadomani
PIANO NAZIONALE
DI RIPRESA E RESILIENZA



INAF
ISTITUTO NAZIONALE
DI ASTRONOMIA

CEPH: IL CUORE DEI DATI

Storage Unificato

Ceph fornisce Block Storage per le VM di Proxmox e Shared File System (CephFS) per le home directory degli utenti nei nodi di calcolo.

Resilienza Nativa

Dati replicati su più dischi/nodi. Autorigenerazione in caso della perdita di uno o più dischi.

```
# Proxmox + Ceph integration  
pveceph init -network 10.0.0.0/24
```



Finanziato
dall'Unione europea
NextGenerationEU



Ministero
dell'Università
e della Ricerca



Italiadomani
PIANO NAZIONALE
DI RIPRESA E RESILIENZA



INAF
ISTITUTO NAZIONALE
DI ASTRONOMIA

SCHEDULING: SLURM WORKFLOW

Slurm è lo standard de-facto per orchestrare l'esecuzione dei job.

```
# Sottomissione job
sbatch -partition=gpu -nodes=2 script.

# Controllo stato
squeue -u $USER

# Info nodi
sinfo
```

Funzioni Chiave

Allocazione: Riserva CPU/RAM/GPU.

Priorità: Gestisce code (Partitions) in base all'utente.

Fairshare: Bilancia l'uso delle risorse nel tempo.



Finanziato
dall'Unione europea
NextGenerationEU



Ministero
dell'Università
e della Ricerca



Italiadomani
PIANO NAZIONALE
DI RIPRESA E RESILIENZA



INAF
ISTITUTO NAZIONALE
DI ASTRONOMIA

OTTIMIZZAZIONE DELLE PRESTAZIONI

Senza Scheduler

45% Efficienza

Con Slurm

85% Efficienza

Slurm + Backfill

96% Saturazione

Il **Backfill Scheduling** permette a Slurm di inserire job piccoli nei "buchi" temporali lasciati dai job grandi, migliorando il ROI dell'hardware.



Finanziato
dall'Unione europea
NextGenerationEU



Ministero
dell'Università
e della Ricerca



Italiadomani
PIANO NAZIONALE
DI RIPRESA E RESILIENZA



INAF
ISTITUTO NAZIONALE
DI ASTRONOMIA

MONITORING & ALERTING



Grafana / Prometheus

Metriche real-time su carico
CPU, traffico Ceph e code
Slurm.



Alerting

Notifiche istantanee in caso
di nodi "DOWN" o
performance degradate.



Accounting

SlurmDBD traccia l'
storico per reportis
billing.



Finanziato
dall'Unione europea
NextGenerationEU



Ministero
dell'Università
e della Ricerca



Italiadomani
PIANO NAZIONALE
DI RIPRESA E RESILIENZA



INAF
ISTITUTO NAZIONALE
DI ASTRONOMIA

PERCHÉ QUESTA COMBINAZIONE?

Scalabilità Totale

L'unione di Proxmox (Hypervisor), Ceph (Storage) e Slurm (Orchestratore) crea un ecosistema robusto, open-source e facile da scalare da 3 a 300 nodi.

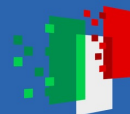
- **Costi ridotti:** Nessuna licenza proprietaria
- **HA integrata:** Servizi critici sempre onli
- **Flessibilità:** Gestione VM e Bare Metal
- **Standard:** Tecnologie leader nel settore



Finanziato
dall'Unione europea
NextGenerationEU



Ministero
dell'Università
e della Ricerca



Italiadomani
PIANO NAZIONALE
DI RIPRESA E RESILIENZA



INAF
ISTITUTO NAZIONALE
DI ASTRONOMIA



Domande?

Grazie per l'attenzione!