Clusters IBISCO
Stato e prospettive

Bernardino Spisso

Riunione di fine anno INFN-GR1 Napoli 22/12/2021

UNIONE EUROPEA
Fondo Sociale Europeo
Fondo Europeo di Sviluppo Regionale

Ministero dell'Istruzione, dell'Università e della Ricerca

PON RICERCA E INNOVAZIONE 2014 - 2020

Progetto I.Bi.S.Co.

Infrastruttura per BIg data e Scientific COmputing

CODICE: PIR01_00011

INFRASTRUTTURA: IPCEI-HPC-BDA

CUP: *I66C18000110006*

Progetto I.Bi.S.CO

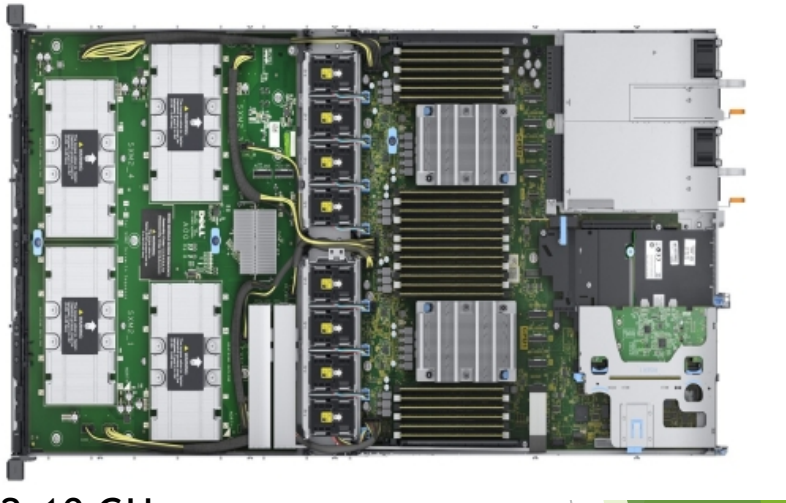
- PON per il potenziamento dei precedenti progetti SCOPE e RECAS
- Risorse INFN per lo storage GRID per circa 10 PB
- Risorse INFN per lo calcolo GRID per circa 6000 core
- **Cluster CPU/GPU condiviso CNR/UNINA/INFN**
- **Cluster CPU/GPU INFN**
- Nuove macchine per i servizi a supporto (Cloud, UI, dhcp, dns, ecc…)
- Nuovo core switch e rack aggiuntivi per ospitare le nuove risorse.

Il progetto I.Bi.S.Co ha previsto due cluster entrambi dotati di GPU ma molto diversi per dimensione e utenza:

- Un cluster condiviso composto da 36 nodi (32 nodi di calcolo + 4 nodi storage) in cui ogni nodo di calcolo possiede 4 GPU e 48-56 core
- Un cluster esclusivo INFN composto da 6 nodi di calcolo + un'unita di storage in cui ogni nodo di calcolo possiede 2 GPU e 128 core

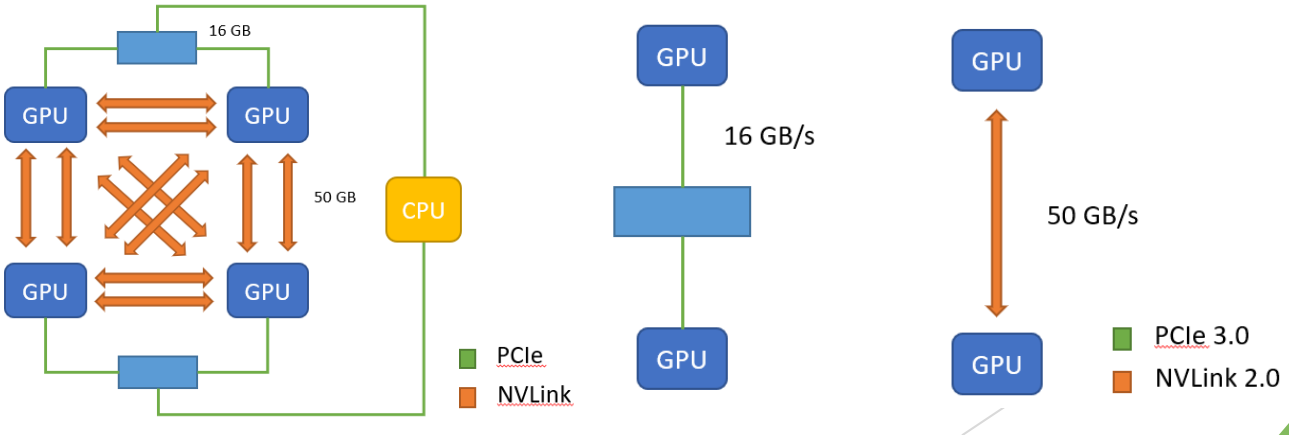
Entrambe i cluster utilizzano la tecnologia Infinibad per l'interconnessione tra i nodi.

Cluster CNR/UNINA/INFN



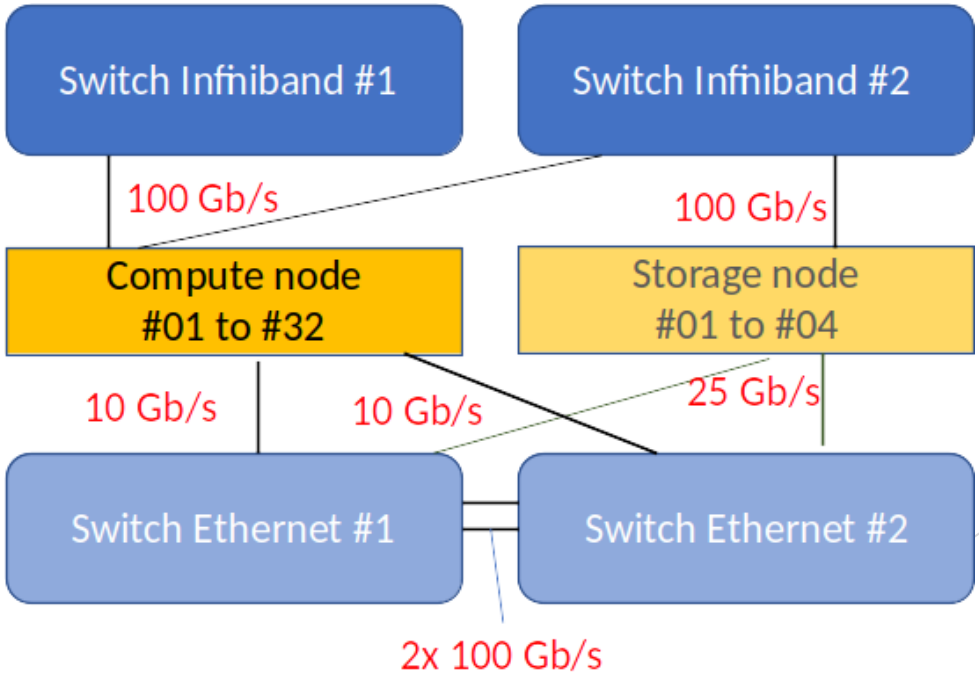
- 32 Nodi di calcolo Dell C4140
- 4 GPU NVIDIA V100 16 GB con bus NVLink.
- 2 CPU Intel Intel Xeon Gold 48-56 core @2.40-2.10 GHz
- Da 750 GB fino 1500 GB di main memory
- 2 dischi a stato solido SATA da 480 GB
- 2x10 Gb/sec Ethernet NIC, 2x100 Gb/sec InfiniBand NIC

Rispetto a PCIe 3.0, NVLink 2.0 utilizza una connessione punto-punto che offre vantaggi in termini di prestazioni di comunicazione.

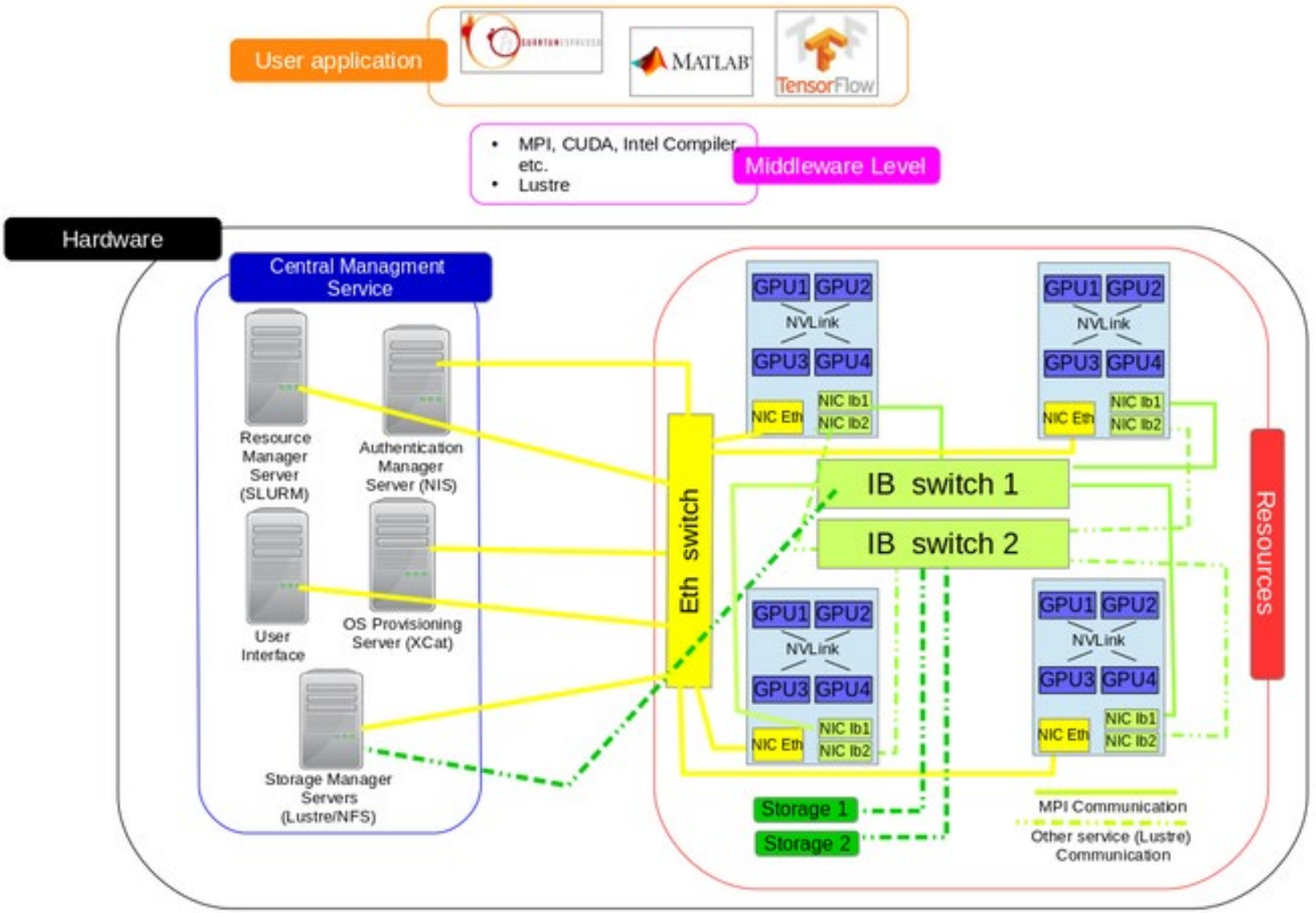


Cluster CNR/UNINA/INFN

- 4 server DELL R740 dedicati allo storage primario dei dati
- 20 HDD SAS da 16 TB e 8 SSD SATA da 1.9 TB.
- Porte InfiniBand a 100Gb/s verso tutti i nodi di calcolo
- Uno spazio separato di 3 PB è disponibile tramite Ethernet e viene utilizzabile come repository a medio termine



Cluster CNR/UNINA/INFN



Cluster CNR/UNINA/INFN

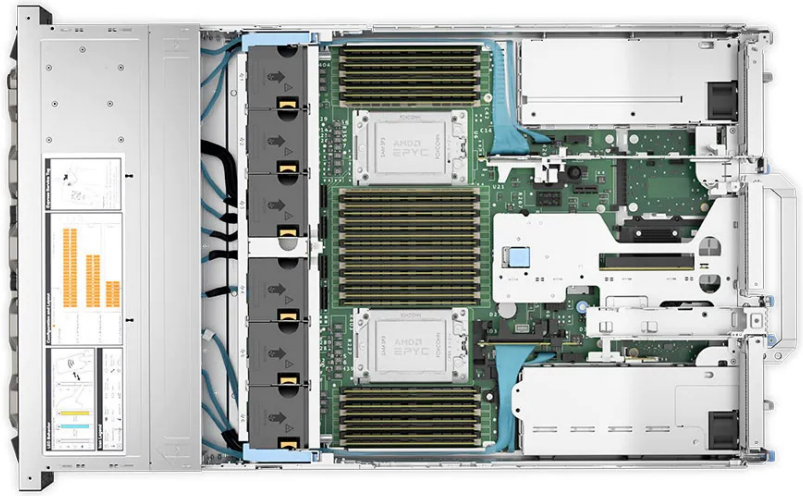
- **NIS server** per la gestione centralizzata degli account utenti
- **Lustre ed NFS server** per la condivisione di spazio sia utente che, ad esempio, per raccogliere il software applicativo.
- **User Interface** come unico punto di accesso alle risorse
- **SLURM server** per la gestione dell'accesso alle risorse e lo scheduling dei job.

Il rilascio delle credenziali per l'accesso alla UI, e quindi alle risorse del cluster, è regolato da delle policy attualmente in fase di rifinitura da parte di un comitato formato da due rappresentati per ogni istituzione. In generale il questo cluster è pensato a applicazioni maggiormente orientate al calcolo GPU o almeno CPU/GPU

È attualmente in allestimento una Wiki:

https://ibiscohpc-wiki.scope.unina.it/start

Cluster INFN

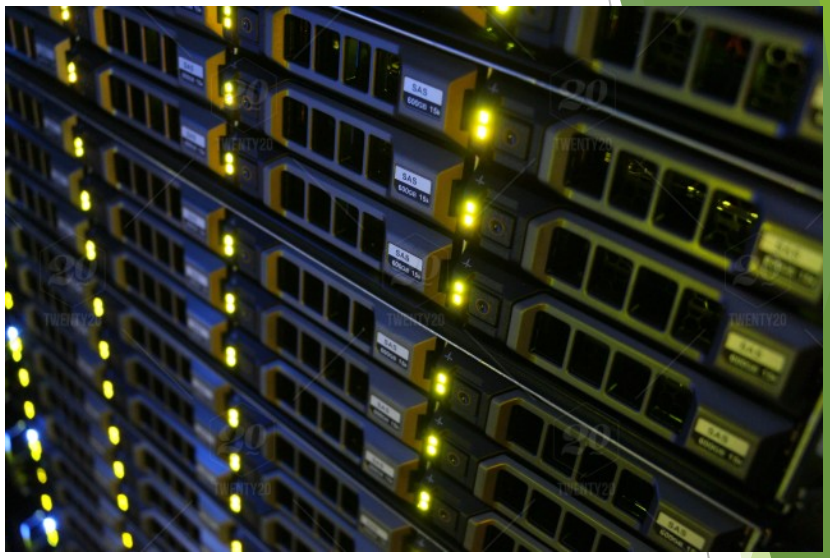


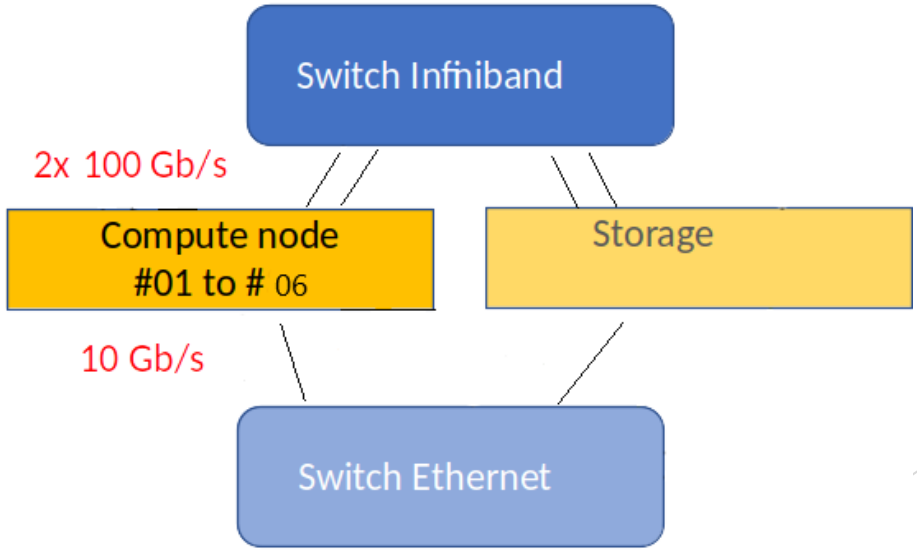
- 6 nodi di calco PowerEdge R7525 many core dotati di GPU
- Sistema di Storage Dell powervault MD3600f per il breve e medio termine circa 300 TB raw
- Interconnessione Infinibad

- 2 CPU AMD EPYC 7742 128 (64x2) core @2.250 GHz
- 2 GPU NVIDIA V100 16 GB PCIe 3.0
- Main memory: 1200 Gb DDR4
- 2 dischi a stato solido SATA da 446.63 GB
- 2 dischi a stato solido SATA da 3576.38 GB
- 2x10 Gb/sec Ethernet NIC, 2x100 Gb/sec InfiniBand NIC

Cluster INFN

La parte storage sarà composta da un unico sistema md3600f connesso ai nodi tramite 2 link infinibad. Tale sistema fornisce importati funzioni, come la ridondanza e le operazioni di lettura e scrittura parallele, in modo trasparente. Nel caso precedente, per ottenere le stesse funzionalità, si è dovuto ricorrere ad uno strato software aggiuntivo.





Stato attuale

Attualmente su tutti i nodi è installato il SO CentOS 7, le librerie CUDA e driver Infinibad. Sono in corso dei test su alcune delle macchine:

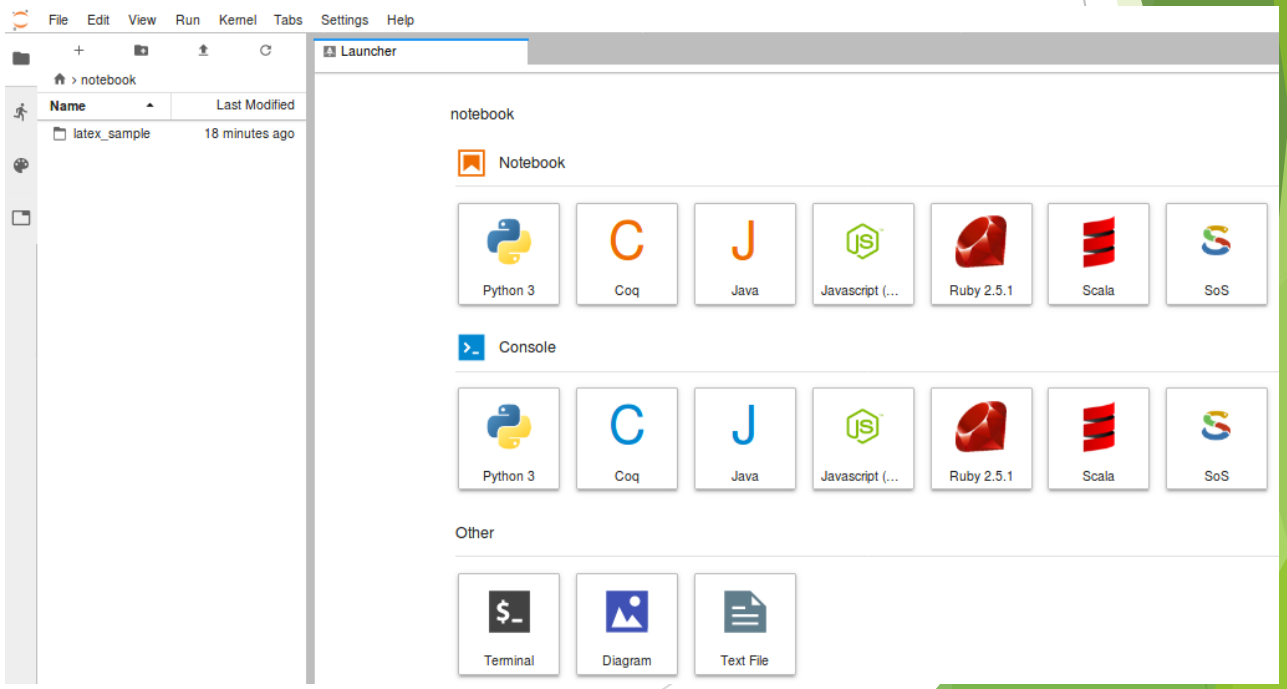
- Ibsico-GPU01 Test sul calcolo GPU (Machine Learning tramite Tensor Flow)
- Ibsico-GPU03 Test sul calcolo massivo su CPU (Geant)
- Ibsico-GPU04 Test su work-flow misti GPU e GPU

To do

- Configurazione dello storage: Lustre e/o NFS per la condivisione di spazio utente e repository.
- User Interface.
- Scheduler: a differenza del cluster, condiviso per omogeneità con il resto del tier2, verrà utilizzato HTCondor.
- Configurazione per operazioni MPI

UI: Riga di comando vs JupyterLab

```
aaronkilik@tecmint ~ $ sudo rsync Templates/* /var/www/html/files/ &
[1] 5782
aaronkilik@tecmint ~ $ jobs
[1]+  Running                 sudo rsync Templates/* /var/www/html/files/ &
aaronkilik@tecmint ~ $
aaronkilik@tecmint ~ $ disown -h %1
aaronkilik@tecmint ~ $
aaronkilik@tecmint ~ $ jobs
[1]+  Running                 sudo rsync Templates/* /var/www/html/files/ &
aaronkilik@tecmint ~ $
```



Politica d’accesso alle risorse

- Solo dedicato alla ricerca o anche per la didattica?
- Ad uso esclusivo o aperto anche alle altre sezioni?
- Sistema di priorità?
- Autenticazione tramite AAI o solo locale?

...

Anche in questo caso bisogna gestire Il rilascio delle credenziali e stabilire delle policy.