

ReCaS Datacenter cluster HPC, guida al suo utilizzo

Febbraio, 2022

Dal 2022 per le attività HPC sono disponibili nuove risorse che sostituiranno integralmente quelle usate fino ad ora. In particolare anche il Resource manager verrà sostituito, infatti HTCondor si occuperà anche della gestione delle risorse HPC. L'idea di base per l'utilizzo di queste nuove risorse sta nell'esigenza di sottomettere job HPC che vengano eseguiti al massimo su un nodo, sfruttando la possibilità di parallelizzare il codice fino ad un massimo di 255 Threads.

Indice:

1. Caratteristiche del cluster HPC
2. Accesso alle risorse, nodi di frontend
3. Compilazione codice
4. FileSystem condiviso
5. Compilatori
6. Batch system
7. Sottomettere un job

1. Caratteristiche del cluster HPC del ReCaS Datacenter

Il cluster e' composto da 6 macchine, naturalmente identiche, che hanno le seguenti caratteristiche:

- 2 socket da 128 core AMD EPYC 7742 64-Core Processor
- 2 TB di RAM
- Una scheda grafica NVIDIA A100
- Una scheda Ethernet a 10Gb/s in fibra ottica

2. Accesso alle risorse.

L'accesso alle risorse del cluster HPC del ReCaS Datacenter e garantito solo ed esclusivamente in modalità batch. Ne consegue che la sottomissione di job batch possa essere effettuata solo dopo avere fatto il login via ssh sulla seguente macchina:

- frontend.recas.ba.infn.it

3. Compilazione codice

Nel caso il codice applicativo sia parallelo e si voglia usare la versione disponibile di openMPI si deve procedere alla compilazione attraverso un job interattivo sottomesso al batch System. In questo modo ci sarà a disposizione una sessione interattiva su uno dei nodi del cluster sul quale si potrà procedere alla compilazione del codice. Sui nodi è presente il seguente codice:

- openmpi-devel
- openmpi

Comando per sottomettere un job interattivo:

- condor_submit <path_submit_file> -interactive
- <https://www.recas-bari.it/images/manuali/ITAManualeHTCondor.pdf>

4. Filesystem condiviso

Sul cluster HPC oltre che sulle due macchine di frontend saranno disponibili i seguenti filesystem condivisi

- /lustre:

IBM GPFS filesystem usato per memorizzare i dati degli utenti che vengono scritti in un'unica copia cio' vuol dire che non e' garantita la consistenza del contenuto della propria aria. Su questo filesystem viene applicata una quota a livello di gruppo locale/esperimento di appartenenza superata la quale l'intero gruppo non potra' piu' usare il filesystem per scrivere dati.

- /lustrehome:

IBM GPFS filesystem usato per fornire spazio disco per le home degli utenti. I file in questo filesystem vengono gestiti in doppia copia di modo da garantire la consistenza del contenuto della propria area.

5. Compilatori

Allo stato attuale il compilatore disponibile di default e' il seguente:

- gcc 4.8.5

E' anche disponibile devtoolset-11 accessibile col seguente comando:

- scl enable devtoolset-11 bash
- In questa bash troverete disponibili tutte le ultime versione degli strumenti per lo sviluppo di codice

Per la GPU invece cuda è disponibile al seguente path e su richiesta possiamo installare la versione necessaria:

- /usr/local/cuda-11
- /usr/local/cuda-11.5

6. Batch system

Le risorse di calcolo presenti nel cluster HPC siano queste le GPU che le CPU vengono gestite da un batch system che nel nostro caso specifico è'

- HTCondor 8.8

una guida completa per il suo utilizzo è disponibile al seguente link:

- <https://www.recas-bari.it/images/manuali/ITAManualeHTCondor.pdf>

7. Sottomettere un job

La sottomissione di un job HPC dovrebbe essere fatta seguendo le linee guida di HTCondor che trovate al seguente link.

- <https://htcondor.readthedocs.io/en/latest/users-manual/parallel-applications.html#mpi-applications-within-htcondor-s-vanilla-universe>

In particolare è necessario specificare come requirements un numero di cpu maggiore di 15 in alternativa richiedere una GPU. In assenza di uno dei due requirement il job verrà gestito da condor come un regolare job HTC. La sottomissione di job HPC è riservato solo agli utenti abilitati.

In particolare aggiungete le seguenti righe al vostro submit file se volete richiedere una GPU

```
request_gpus = 1  
request_cpus = 1
```