



# Calcolo distribuito ad alte alta produttività

---

*Esercitazione guidata n.2*  
*Cluster ad Alte Prestazioni*

*Tutor: Damiano Perri*

1.

Che problema affrontiamo?

# Esercitazione n°2

Realizzazione di un cluster ad alta produttività



*Lo scopo di questo lavoro ha come obiettivo la realizzazione di un cluster ad alte prestazioni, cioè di un insieme di nodi interconnessi tra loro capaci di distribuirsi il carico di lavoro in maniera autonoma in ambienti orientati all'esecuzione di applicazioni particolarmente onerose dal punto di vista dell'utilizzo del processore.*

# Esercitazione n°2

Realizzazione di un cluster ad alta produttività



*Il software alla base del progetto è HTCondor sviluppato all'interno della University of Wisconsin–Madison.*

*HTCondor gestisce i job, le macchine di esecuzione, i dati, utilizza la rete disponibile, gestisce il workflow e inoltre può essere monitorato dagli admin tramite console.*

# 2.

Quando un sistema di questo  
tipo è utile?

# Esercitazione n°2

Realizzazione di un cluster ad alta produttività



*Prendiamo in considerazione un software che debba calcolare l'interazione che lo scudo termico di uno shuttle al rientro nell'atmosfera terrestre possa avere con l'aria. Un calcolo del genere presenta numerose variabili e sarebbe troppo oneroso da sostenere per una singola macchina, di conseguenza l'idea alla base del progetto prevedere di suddividere il lavoro tra più computer.*

# Esercitazione n°2

Realizzazione di un cluster ad alta produttività



*A scadenze di tempo regolari, dette checkpoint, i calcoli svolti della cpu vengono salvati su disco e successivamente integrati insieme per ottenere il risultato finale desiderato.*

*Un calcolo che poteva richiedere anni, può essere così essere portato a compimento in mesi o addirittura settimane.*

# Esercitazione n°2

Realizzazione di un cluster ad alta produttività



*Esempi di applicazioni:*

- *studi relativi alla Climatologia*
- *studi relativi ad equazioni fluidodinamiche in Fisica*
- *studi a livello atomico della materia in Chimica*
- *ricerca sulle interazioni tra le proteine artificiali e l'organismo in Farmacologia*
- *studi contro le malattie degenerative*



3.

Punti a favore e a sfavore

# Esercitazione n°2

Punti a favore



*Il vantaggio che si ottiene realizzando un'architettura di questo tipo è sia economico che prestazionale: invece che avere un'unica macchina ad altissime prestazioni e dal costo elevato, si distribuiscono i calcoli su più unità detti nodi (anche con caratteristiche hardware eterogenee tra loro) e con un costo di messa in opera contenuto*

# Esercitazione n°2

Punti contro



*Non tutte le macchine sono adatte, servono sistemi di dissipazione adeguati per gestire il calore generato*

*I problemi di configurazione possono essere difficili da gestire: i job non partono, partono ma non vengono gestiti dalla cpu, partono ma rimangono idle, ecc*

3.

Cosa è richiesto?

# L'obiettivo da raggiungere

1/2



*Configurare almeno 2 nodi (ma consigliati 3 nodi di cui 2 worker e un master)*

*Installare Condor su tutte le macchine del cluster*

*Creare dei lavori (jobs) da sottomettere a Condor*

*Punto Bonus:*

*Realizzare una configurazione che se rileva un input dell'utente (magari l'utente sta navigando in rete e digita sulla tastiera un carattere) sospende temporaneamente i calcoli e li riprende non appena la macchina torna IDLE*

# L'obiettivo da raggiungere

2/2



*Documentare tutto quello che si è realizzato e produrre una relazione dettagliata sul lavoro svolto da consegnare in vista dell'esame*

*Inserire nella relazione foto e descrivere il problema.*

*Verificare che i jobs assegnati a Condor vengano effettivamente svolti, documentando i risultati ottenuti.*

*In caso di insuccesso durante qualche fase spiegare nella relazione cosa è andato storto e le difficoltà riscontrate*

3.

Installare il software base

# I pacchetti fondamentali



*Installare condor*

*sudo dnf -y install condor*

*A differenza del progetto precedente dove erano richiesti ulteriori programmi, come ad esempio HeartBeat e Corosync, in questo caso non sono necessari ulteriori applicativi.*

*Configurare la rete per i nodi e il firewall*



# I primi passi



*Configurare il file /etc/hosts*

*Ogni nodo dovrà avere come nome oltre a se stesso anche se stesso.condor.*

*es per nodo1 con ip 192.168.1.40:*

*192.168.1.40 nodo1 nodo1.condor*

*192.168.1.50 nodo2 nodo2.condor*

*Ecc..*

# I primi passi



*Creare un nuovo file 00personal\_condor.config a partire da quello presente in etc/condor/config.d  
Qui sono presenti le configurazioni che l'amministratore del nodo stabilisce valide per la macchina. La copia creata avrà nome condor\_config.local*

# I primi passi



*Definire*

*CONDOR\_HOSTS* cioè il nodo central manager

*COLLECTOR\_NAME* central manager,..

*NETWORK\_INTERFACE* nome dell'interfaccia di rete

*HOSTALLOW\_WRITE* = qui si definiscono i nodi con  
facoltà di scrittura, in questo caso è stato messo

*\*.condor* in quanto abbiamo definito i due nodi con la  
parte finale del nome in comune; la subnet o il dominio

# I primi passi



*Definire*

*DAEMON\_LIST* questo è il parametro che indica quali demoni devono essere in esecuzione sul nodo corrente, il nodo master (node1)

*Il nodo 1 avrà COLLECTOR, MASTER, NEGOTIATOR SCHEDD, STARTD.*

*Mentre il nodo slave (node2) avrà MASTER e STARTD*

# I primi passi



*Fatto ciò abilitare condor e inserirlo tra i demoni in autostart come nella esercitazione1*  
*Configurare poi una cartella di lavoro, definire un job e sottometterlo alla rete.*

# Bonus 1/2



*Per raggiungere il massimo in questa relazione è richiesto di definire opportune policy del sistema dove l'esecuzione dei jobs termina non appena viene rilevata attività da tastiera OPPURE quando la cpu viene utilizzata oltre il 60%.*

*Fare cioè in modo di eseguire i jobs solo con la macchina in IDLE e che vengano stoppati e ripresi in modo automatico.*

# Bonus 2/2



*Testare e relazionare cosa succede utilizzando Condor su una macchina con più processori fisici / logici  
Ad esempio impostando da Virtualbox più di 1 core per almeno un nodo.*

*ES:  
Come viene gestita la coda se un nodo ha 2 core invece che uno solo ?*

Grazie !  
**Domande?**

Per eventuali contatti:  
[Damiano.perri@gmail.com](mailto:Damiano.perri@gmail.com)