

Farm di Analisi Interattiva e Infrastruttura di Calcolo Scientifico a Pisa

E. Mazzoni

21 Novembre 2017



INFN-Pisa in numeri

- › 100 m² di sala
- › Circa 10000 core di calcolo sia HTC che HPC
- › 2PB di disco in un unico file-system GPFS
- › Più di 20 Virtual Organization Grid supportate
- › Più di 200 utenti geograficamente distribuiti che accedono il sito
- › Tier2 per CMS e Belle2





Di che cosa si compone

Accesso

- › gridui.pi.infn.it
- › localui.pi.infn.it

Autenticazione

- › Kerberos
- › AAI

Storage

- › AFS
- › GPFS
- › CVMFS

Risorse computazionali

- › LSF

Grid

- › CE, SRM, ecc...



Accesso

gridui.pi.infn.it {1,2,3}

- › Nascono come vere e proprie User Interface grid ma funzionano anche come macchine di login pubblico e sottomissione locale al sistema di batch
- › Sia per utenti locali che remoti
- › Home GPFS, Scientific Linux 6

localui.pi.infn.it {1,2,3}

- › Nascono come macchine di login pubblico per accesso alle risorse via sottomissione locale (diretta al sistema di batch)
- › Sia per utenti locali che remoti
- › Home GPFS, Scientific Linux 6

setupui.pi.infn.it

- › Non è una macchina di login vera e propria
- › Serve per crearsi la propria home in GPFS
- › Necessario farci login solo la prima volta che si intende accedere al sito



Autenticazione

Kerberos

- Disponibile per utenti "pisani" su tutte le User Interface
- Sicurezza: la password non lascia la propria macchina (se configurata)
- Permette l'uso di AFS con token in batch
 - `fai-krb-setup`
 - Code: `fai` e `local` in automatico

AAI

- Funziona per tutto il personale INFN
- Utenti afferenti a CSN4, CMS e Belle2 automaticamente, tutti gli altri a richiesta
- Meccanismo standard username e password



Storage

AFS

- › Namespace globale /afs/pi.infn.it....
- › Server-client, disponibile per tutti i sistemi operativi
- › Controllo accesso ai file con meccanismi propri
- › Back-up settimanale con storico di un mese
- › 40TB

GPFS

- › Namespace locale /gpfs/ddn
- › File system posix standard con ACL estese
- › Pensato per elevate prestazioni
- › Nessun back-up solo ridondanza hardware dei sistemi
- › 2PB

CVMFS

- › Namespace globale /cvmfs/cern.ch
- › Usato per distribuire il software di esperimenti LHC e non solo
- › Accesso in sola lettura

Tutti accessibili da tutte le nostre risorse con il solito path



Storage: GPFS

Unica infrastruttura per tutte le necessità di archiviazione: aree home, aree di gruppo, SRM

- Uso dei fileset per la gestione di permessi e quote di gruppo
- Quota per utente solo per le home
- SRM → StoRM usa una fetta del filesystem
- Accesso POSIX: quindi anche la parte SRM è accessibile in lettura senza bisogno di tool GRID-specifici

home → /gpfs/ddn/users/[username] 200GB/utente ~20TB in totale
dhquota aggiornato ogni ora

gruppo → /gpfs/ddn/belle2 ancora non esiste si potrebbe dedicargli
parte dello storage 2018

SRM → /gpfs/ddn/srm/belle RW via StoRM (grid)
RO in locale per utenti belle2
attualmente 3TB crescerà nel 2018



Risorse computazionali

Hardware

- **HTC** (seriale/grid): 2GB/core
- **HPC** (parallelo): >2GB/core + IB
 - Diablo: 4x(24 core, 52GB) QDR CFD
 - Mefisto: 4x(36 core, 192GB) QDR CFD ing. Strutturale
 - Ghibli: 128x(8 core, 16GB) DDR CFD
 - Tramontana: 152x(8 core, 24GB) QDR CFD
 - Zefiro: 32x(64 core, 512GB) QDR CSN4
- **FAI**: Farm Analisi Interattiva
- **theonuc**: sistemi ad elevata RAM/core dedicato ai teorici nucleari
- **GGPU**: vari sistemi con coprocessori CUDA di varie generazioni

Utilizzo

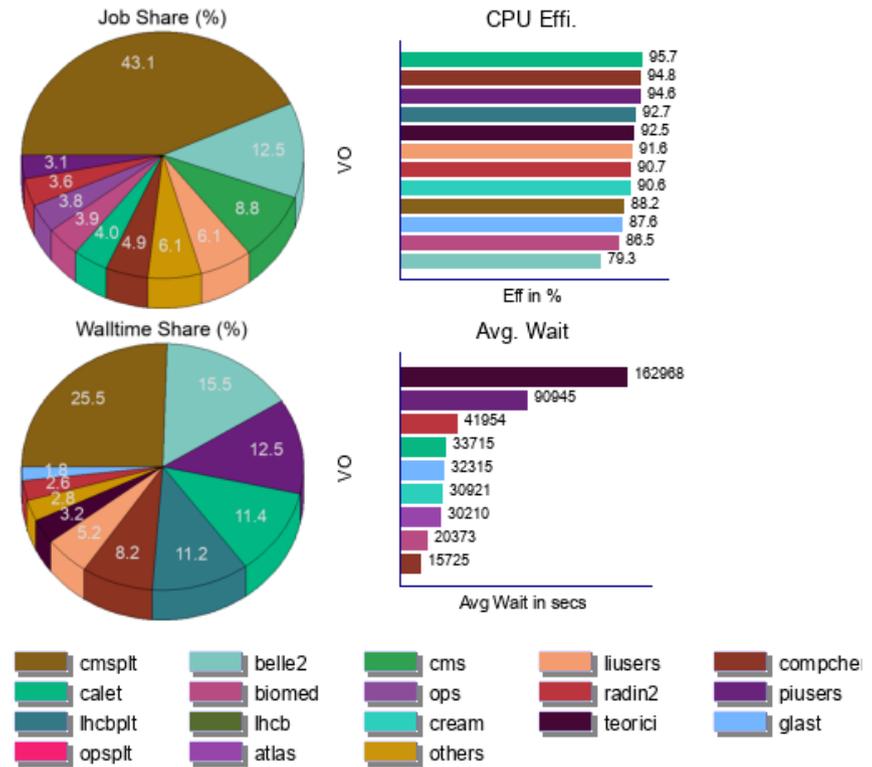
- Accesso alle risorse mediato attraverso LSF
- Cluster INFN-PISA LSF7: risorse HTC, FAI, theonuc, gridui
- Cluster INFN-HPC LSF9: risorse HPC, localui



Utilizzo del sito

[Current Status](#)
[Last 3 Hours](#)
[Last 6 Hours](#)
[Last 12 Hours](#)
[Last Day](#)
[Last Week](#)
[Last Month](#)
[Last 3 Months](#)
[Last 6 Months](#)
[Last Year](#)
[Full Period](#)

VO/Group	Total Jobs	Succ Jobs	Succ Rate(%)	Walltime (sec)	CPU Time (sec)	CPU Eff(%)	Walltime Share(%)	Avg Wait (sec)
cmsplt	2117594	2013308	95.08	17923769379	15802276676	88.16	25.55	8008
belle2	611166	593235	97.07	10856560101	8607765842	79.29	15.48	4459
piusers	150532	134645	89.45	8774457671	8300363221	94.60	12.51	90945
calet	197725	178313	90.18	7981170849	7635121643	95.66	11.38	33715
lhcbplt	117250	109289	93.21	7884437112	7305716130	92.66	11.24	3926
compchem	242827	191877	79.02	5761724761	5460520803	94.77	8.21	15725
liusers	297183	97861	32.93	3626775047	3323134860	91.63	5.17	4487
teorici	42246	35551	84.15	2249222306	2081398898	92.54	3.21	162968
radin2	178505	177406	99.38	1790531561	1623824765	90.69	2.55	41954
glast	17148	10459	60.99	1269392487	1112001541	87.60	1.81	32315
biomed	192668	183967	95.48	1097900089	950167014	86.54	1.57	20373
cms	432762	428814	99.09	572815794	318792398	55.65	0.82	721
cream	54693	53288	97.43	232850966	211061972	90.64	0.33	30921
ape	351	38	10.83	66687788	66113637	99.14	0.10	54090



Monitoring completo LSF: <https://farmsmon.pi.infn.it/lsfmon>



Farm Analisi Interattiva: FAI

Perchè

- Nasce ~4 anni fa per caso e necessità
- Caso: acquisto contemporaneo di varie macchine da parte di vari gruppi per fare tutti la stessa cosa
- Necessità: avere una piattaforma per fare attività interattiva, in pratica ognuno voleva la propria macchina per fare analisi

Come

- Gruppi pagano, IT gestisce
- Non sono User Interface ma Worker Node quindi ci si accede via LSF
- Una coda dedicata ma di solo tipo interattivo
- Stessa struttura software di una UI: posso eseguire e compilare
- UI+LSF+nodo FAI provvedono al trasporto X11 fino al tuo terminale

Cosa

- Garantisce una shell interattiva su una macchina
`bsub -ls -q fai /bin/bash`

Risorsa nazionale?

- Lo divenne immediatamente dopo l'abilitazione di AAI su le User Interface



FAI + Belle2

Evoluzione

- Acquisite nuove risorse HW: 32 core + HT grazie al finanziamento di 8k€ da Dot.1 Pisa (CSN1 non finanzia questo tipo di infratsruture)
- Migliorata la fruibilità: aggiunto alle User Interface un meccanismo "smart" per il trasporto della parte grafica
- Più flessibile: "cloud-di-scoglio" si usa Docker+LSF per dare un ambiente non più solo una shell su un WN in questo momento disponibili ScientificLinux6 e CentOS7
- Possibilità di personalizzazione: creato repo locale agganciato ad AAI per gestire le immagini Docker (portus.pi.infn.it) è possibile sviluppare l'immagine sul proprio portatile e quindi renderla disponibile all'intera infrastruttura

Proviamo ad usarla



FAI uso

- Primo login per creare la home

```
ssh username@setupui.pi.infn.it
```

con le vostre credenziali AAI

- Max 5 min e l'account è pronto per essere usato, si accede ad una UI

```
ssh username@gridui.pi.infn.it
```

ricordarsi di chiedere il forward di X11 se necessario

- Si accede alla FAI scegliendo l'ambiente di cui si ha bisogno

```
bsub -Is -n 1 -q fai sl6
```

sl6 → per Scientific Linux 6 (default)

cs7 → per CentOS 7

- Finito il download dell'immagine Docker (tutta o parte secondo la necessità) si otterrà una shell bash e si è pronti a lavorare

Maggiori dettagli su Confluence:

<https://confluence.desy.de/pages/viewpage.action?spaceKey=BI&title=Italian+Analysis+Facility>



X2go

- Il problema principale nell'uso di risorse interattive centralizzate è il trasporto della parte grafica
- In linea di principio tutto è già disponibile, ssh sa fare il forward di X11 in maniera trasparente all'utente ma di fatto la latenza della rete lo rende inutilizzabile
- È necessario cambiare modo di trasportare X11 sulla rete per questo sulle User Interface è stato installato un sistema di "Remote Desktop" chiamata [X2go](#)
- Per sfruttarlo è necessario installare l'opportuno client sulla propria macchina configurarlo ed utilizzarlo al posto del tradizionale ssh

Parametri configurazione

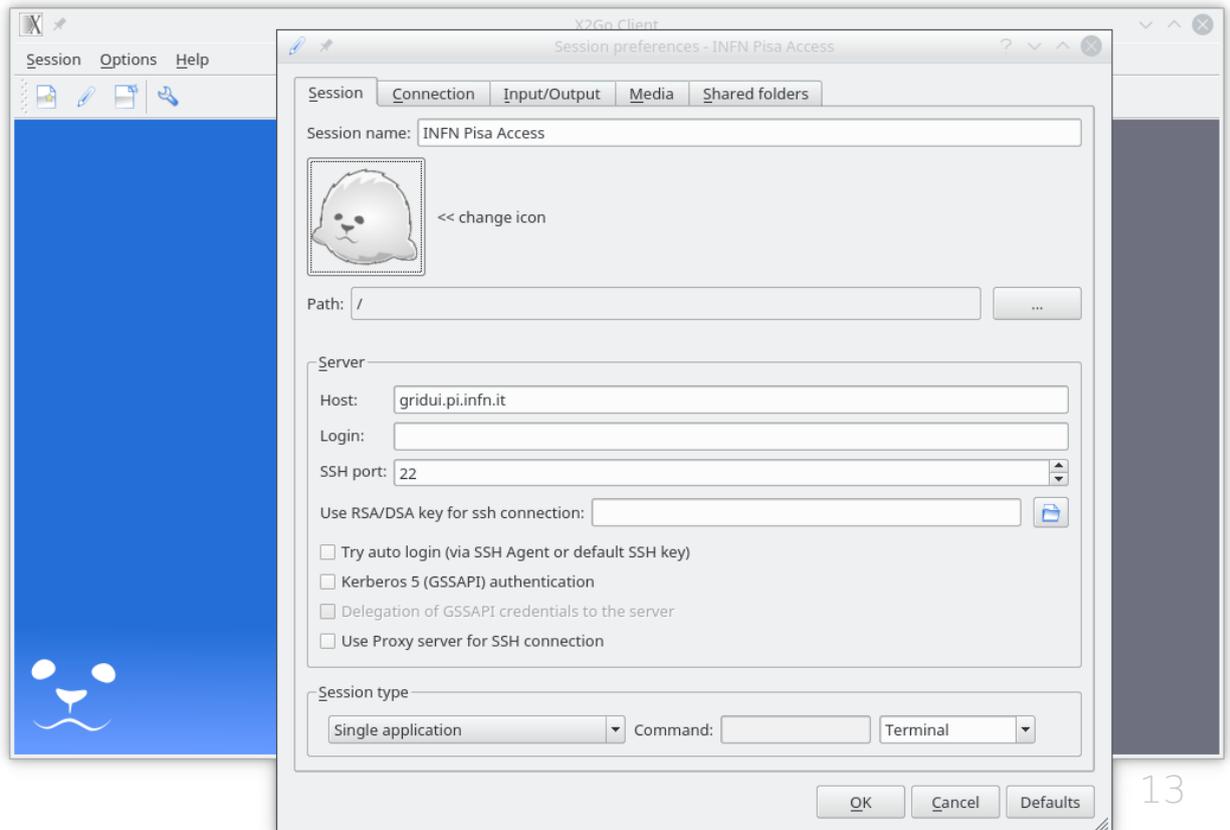
Session name: quello che si vuole

Host: gridui.pi.infn.it

Login: username

Session type: single application

Command: terminal





Un po' di mani in pasta

- › setupui.pi.infn.it
- › gridui.pi.infn.it
- › Un po' di kerberos .k5login
- › Bsub
- › Setting software di belle e prova istogramma
- › Ripetere con x2go
- › Altri tool es. mathematica

Maggiori dettagli su Confluence:

<https://confluence.desy.de/pages/viewpage.action?spaceKey=BI&title=Italian+Analysis+Facility>



Futuro

- Nel 2018 installazione di CPU e Storage finanziati per Belle2
- Centro poli funzionale (HTC, HPC, GGPU, Interattivo) ed aperto a tutta la comunità INFN grazie ad AAI. Gli ingredienti di Calcolo Scientifico ci sono tutti, sta alla "fantasia" dell'utenza che banchetto apparecchiare
- Espansione dell'infrastruttura di Calcolo Scientifico di INFN-Pisa con l'utilizzo di una secondo CED in cui migrare la parte HTC del centro. Si tratta di una collaborazione fra INFN e Università nell'ambito del nuovo centro di calcolo che UniPi ha realizzato all'interno della Ex-sede INFN a San Piero a Grado. Due sedi fra loro vicine e ben collegate grazie alla rete in fibra dell'Ateneo.
- Infrastruttura di storage di tipo Enterprise paragonabile a quella del CNAF in grado di assorbire "facilmente" anche forti espansioni
- Può esserci spazio per incrementare l'attività di Belle2 per compensare quanto non sarà disponibile dato l'allagamento se la cosa è di interesse per la comunità