

ENEA GRID

Corso di introduzione all'uso ed all'amministrazione

Autore: Alessandro Secco
alessandro.secco@nice-italy.com

Lezione 3

- Riepilogo lezione 2
- AFS: terminologia e repliche
- LSF Admin: concetti di base
- Aggiunta di un host
- Aggiunta di una risorsa boolean
- Configurazione di una coda

Riepilogo lezione 2

- Nella lezione precedente si è parlato di argomenti inerenti LSF dal lato utente.
- Come trasferire dati
 - Client AFS
- LSF utente: concetti di base
 - Comandi di interrogazione del cluster
 - Richiesta di risorse
- LSF: lancio di un job
 - Compilazione di un job in AFS
 - Lancio e monitoraggio di un job

AFS: terminologia e repliche

AFS: terminologia

- Come visto nella prima lezione, AFS è un **file system distribuito** nato per girare in ambiente multi-centro, come può essere l'Enea.
- **CELLA**: insieme di client e server in cui gira il software AFS, con amministratori ed utenti propri.
- **VOLUME**: contenitore di file in relazione tra loro, contenuto in una partizione di un file server.
- **MOUNT POINT**: directory del file system in cui viene reso disponibile un volume.
- **REPLICA READ-ONLY**: volume Read-Only, copia di un volume Read-Write
- **CACHE**: copia parziale dei file di AFS in locale sul client
- **UTENTE AFS/Kerberos 5**: un account definito in AFS/Kerberos 5. Potrebbe anche non esistere nel sistema operativo.
- **TOKEN**: rappresenta le credenziali, salvate nella memoria del kernel, che serviranno per accedere ai servizi forniti da AFS.
- **AMMINISTRATORE**: un utente AFS in grado di creare volumi, mount point, repliche, ecc.

AFS: attori principali

- **FILE SERVER:** monta le partizioni fisiche contenenti i volumi read-write e read-only
- **DB SERVER:** in origine era una serie di demoni che gestivano il database di tutti gli utenti, permettevano l'autenticazione dei client, ... Da quando AFS lavora con Kerberos 5, il DB server garantisce l'integrazione tra i due ambienti.
- **CLIENT:** una macchina che si autentica sul DB server e mostra i volumi forniti dal file server all'utente finale o all'amministratore.

AFS: miniguia alle repliche

- Le celle AFS disponibili, vengono montate dentro /afs. Per esempio, la cella enea si trova in /afs/enea.it
- A partire da /afs/enea.it, si trovano tutti i mount point dei volumi Enea.
 - **fs ls /afs/enea.it/software** mostra il nome del volume presente in quella directory
- Ogni mount point può essere:
 - **Read-only**: i file del volume non possono essere scritti direttamente
 - **Read-write**: i file possono essere scritti, se si possiedono i permessi.

AFS: miniguia alle repliche

- Per capire se si sta lavorando su un volume read-only, bisogna interrogare AFS.
 - Si supponga di lavorare in
 - `/afs/enea.it/software/lsf/conf`
 - Si lanci un comando di ispezione del volume:
 - `fs examine /afs/enea.it/software/lsf/conf`
 - Se il nome termina con `.readonly`, allora si tratta di un volume read-only.
- Il non poter scrivere in una directory, non è garanzia di trovarsi su un volume read-only !

AFS: miniguia alle repliche

- Come modificare un volume readonly ?
 - Si applica la procedura vista precedentemente per capire il nome del volume
 - Ci si sposta in una directory “speciale” di AFS contenente i volumi read-write, caratterizzata da avere un punto davanti al nome della cella:
 - `cd /afs/.enea.it/software/lsf/conf`
 - Dopo aver modificato i file nel volume read-write, si prende il token di amministratore di AFS e si aggiornano le repliche read-only.
 - `klog <amministratore>` (chiede una password)
 - `vos rel <nomevolume>`
- Per chi non fosse amministratore, esiste l'utility WARC, documentata dal seguente link:
 - <http://www.afs.enea.it/project/eneagrid/afs/#Utility>

LSF Admin: concetti di base

LSF Admin

- Esistono diversi tipi di amministratore LSF:
 - **Amministratore di LSF:** è l'utente che può far ripartire il sistema a caldo, modificare i file di configurazione, intervenire sui job di tutti gli utenti. In Enea è l'utente "lsf".
 - **Amministratore secondario:** tecnicamente ha tutti i poteri dell'amministratore di LSF, ma potrebbe avere restrizioni di file-system sulla modifica dei file di configurazione.
 - **Amministratore di coda:** non può configurare LSF, ma all'interno di una coda può intervenire sui job di tutti gli altri utenti

Percorso dei file

- In Enea LSF è installato in:
 - **/afs/enea.it/software/lsf**
 - **conf**: file di configurazione del cluster e risorse
 - **conf/lSBATCH/<nomecluster>/configdir**: configurazione delle code, dei limiti su host e utente e politiche riguardanti il lancio dei job.
 - **docs**: documentazione di LSF in PDF e HTML
 - **7.0/<piattaforma>**: contiene le directory dei file binari. Per ogni piattaforma è stato fatto un link a “fs sysname”.
 - **bin**: comandi utente e admin
 - **etc**: demoni di LSF
 - **lib**: librerie (API) di LSF
 - Queste 3 directory sono linkate in **/afs/enea.it/software/lsf** sostituendo @sys in luogo della piattaforma, per fornire un ambiente uniforme di accesso ai file binari.
 - **work**: contiene le informazioni di accounting e i log degli eventi usati per la fault tolerance
 - **log_<nomecluster>**: contiene i log di sistema del cluster

Terminologia aggiuntiva

- **PROCESSO**: indica un programma in esecuzione. Per semplificare, si considera:
 - Job seriale = 1 processo
 - Job multithreading = n thread su 1 host. Il thread è visto come un processo.
 - Job parallelo = n processi su m host.
 - Per i job batch un processo è considerato tale solo se usa intensivamente la CPU.
- **SLOT**: è uno spazio in grado di contenere un singolo processo
- **MASTER**: il server che si occupa di gestire il cluster LSF.
- **MASTER di backup**: se il master si ferma, uno o più host possono prenderne il posto.
- **SLAVE**: i server controllati dal master

Conoscere master e cluster

- **Isid** è un comando che informa sul nome del master e sul cluster dell'host da cui si è lanciato il comando.
- Il master è indicato nella variabile d'ambiente LSF_MASTER_LIST come prima chiave. La variabile è letta da Isf alla partenza dei demoni.
- Se il master non è attivo, la seconda chiave di LSF_MASTER_LIST diventa il nuovo master, e così via.

Demoni di LSF

- LSF usa dei processi che girano in background (demoni).
- I demoni, che si vedranno in dettaglio in altre lezioni, hanno alcune caratteristiche di cui bisogna tener conto:
 - La loro comunicazione è volutamente blanda, per aumentare la scalabilità di LSF. Questo può causare lentezza nella risposta del sistema (bsub impiega circa 10-15 secondi a schedulare un job).
 - L'amministratore può regolare le tempistiche di tutti i demoni, con molta, molta cautela.
- Alcuni demoni: lim, res, sbatchd, pim, mbatchd, mbschd, ...
- I demoni vanno fatti partire da **root** con lo script `lsf_daemons` che si trova in **7.0/<piattaforma>/etc**

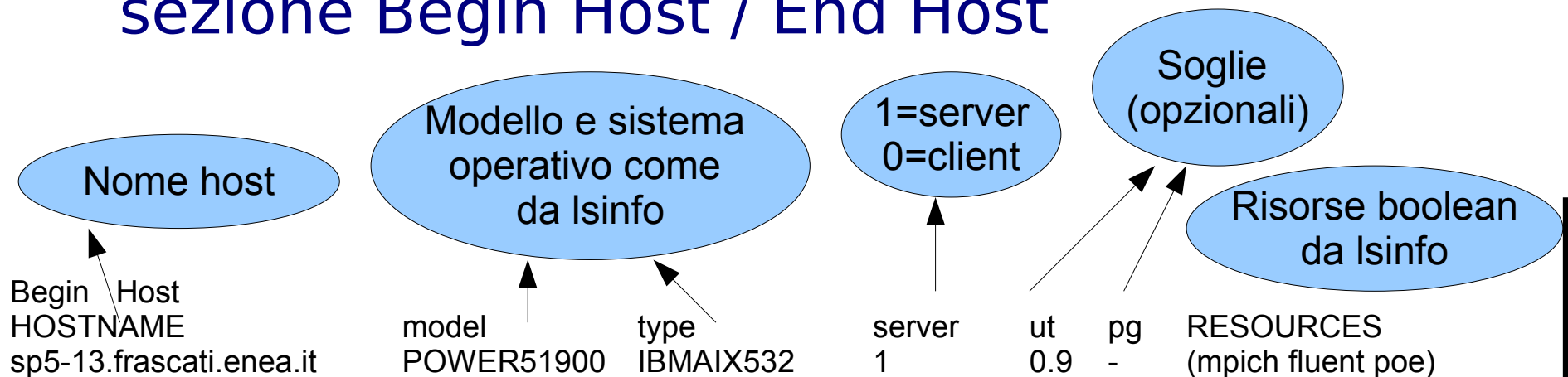
Aggiunta di un host

Aggiunta di un host

- Per motivi di sicurezza e stabilità, la politica usata da Enea per l'aggiunta di host, è quella di usare una configurazione statica basata su file.
- File in **/afs/enea.it/software/lsf/conf**:
 - **lsf.cluster.<nomecluster>**
 - **hosts**
 - **NOTA**: bisogna modificare il volume read-write !!
- File locali (accesso da root):
 - **/etc/lsf.sudoers**
 - **lsf_daemons**
 - **/bin/afsws/bin/pagsh**

lsf.cluster.<nomecluster>

- Questo file contiene l'elenco degli amministratori, l'elenco dei client e server di LSF, le licenze in uso dai demoni, la disposizione delle risorse nel cluster.
- Per aggiungere un server si modifica la sezione Begin Host / End Host



hosts

- Il file hosts è simile al file /etc/hosts, e viene usato da LSF per sovrascrivere le informazioni da /etc/hosts e DNS in caso di macchine con più schede di rete o configurazioni particolari.
- Se, per esempio, una macchina comunica su 2 IP, potrebbe essere necessario aggiungere una configurazione come segue:
 - `xxx.yyy.zzz.www nomehost nomealias`
 - `aaa.bbb.ccc.ddd nomehost`

File locali

- Infine si aggiungono i file locali per la partenza dei demoni e l'integrazione con AFS:
 - Integrazione con AFS:
 - **In -s /afs/enea.it/software/lsf/conf/lsf.sudoers /etc**
 - Deve esistere **/usr/afsws/bin/pagsh** (cercare in distribuzione di AFS e copiare o creare link)
 - Partenza dei demoni di LSF (solo server):
 - **/afs/enea.it/software/etc/lsf_daemons start**
 - Mettere **lsf_daemons** tra le ultime operazioni al boot del sistema.

Riconfigurazione di LSF

- Una volta modificati i file di configurazione che si trovano in `/afs/.enea.it/software/lsf/conf` si deve:
 - Aggiornare le copie read-only di AFS
 - Riconfigurare lsf in 2 passi:
 - **lsadmin reconfig**: avvisa il master lim di rileggersi i file di configurazione e di ripartire.
 - **badmin reconfig**: comunica al master batch daemon di riconfigurarsi.
 - Vedere che l'host sia stato aggiunto:
 - **lsload <nomehost>**

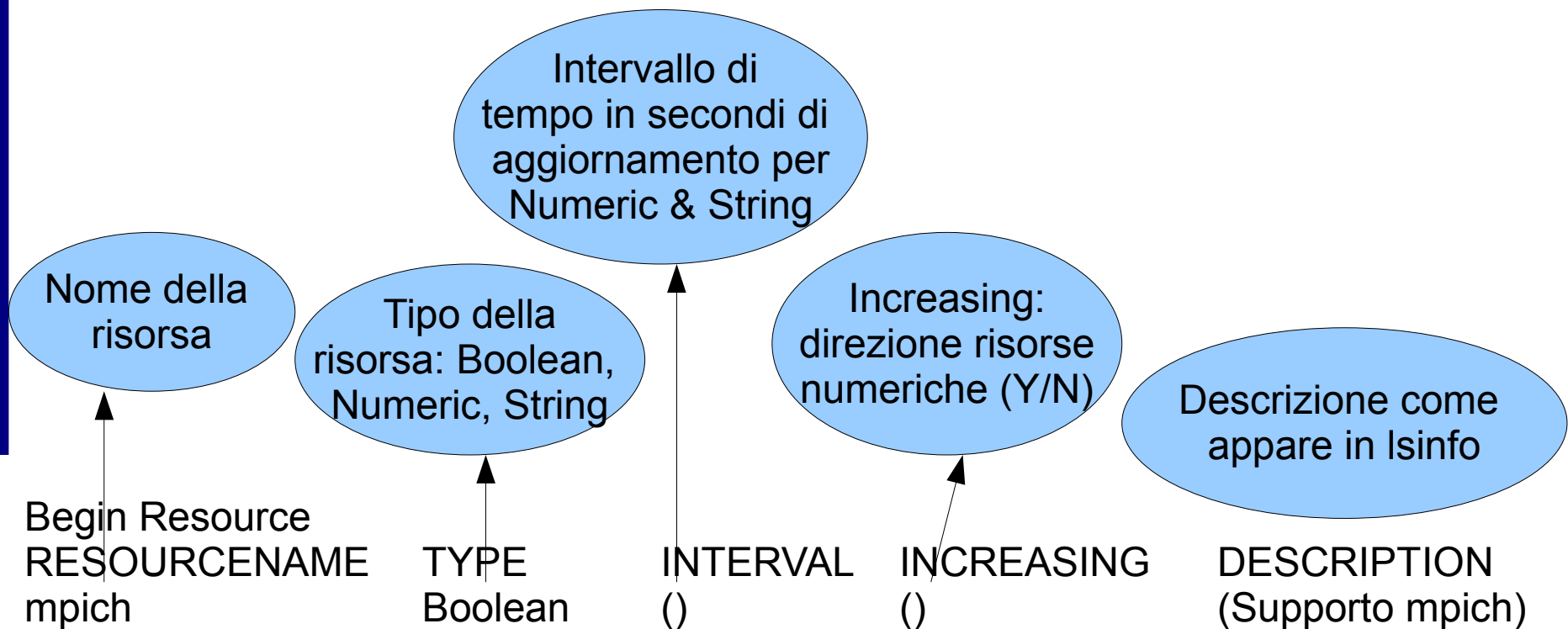
Aggiunta di una risorsa boolean

Il file Isf.shared

- Isf.shared è il file contenente tutte le risorse di LSF e si trova nella directory conf.
- Sezione “Begin Cluster”
 - Contiene i cluster di Isf connessi tra loro
- Sezione “Begin HostType”
 - Contiene tutte le keyword inerenti ai sistemi operativi
- Sezione “Begin HostModel”
 - Contiene i nomi dei modelli, il loro cpufactor, e la keyword dell'architettura
- Sezione “Begin Resource”
 - Contiene la definizione di tutte le risorse stringhe, numeriche, boolean

Risorsa boolean

- Basta aggiungere una riga alla sezione Begin Resource



Risorsa boolean

- Successivamente va modificato `lsf.cluster.<nomecluster>`, aggiungendo la risorsa sull'host desiderato alla colonna `RESOURCES`, come visto in precedenza.
- Infine:
 - Aggiornare i volumi readonly
 - `lsadmin reconfig`
 - `badmin reconfig`

Risorse numeriche (cenni)

- Non è così semplice aggiungere risorse dinamiche numeriche (Come calcolo il loro valore ? Come le distinguo da quelle statiche ?)
- LSF fornisce uno strato di software, che si vedrà in lezioni successive) per lanciare script adatti a questo scopo.
- Per gli impazienti:
 - `man lim` (cercare ELIM)
 - `man lsf.cluster` (cercare ResourceMap)
 - `man lsf.shared`

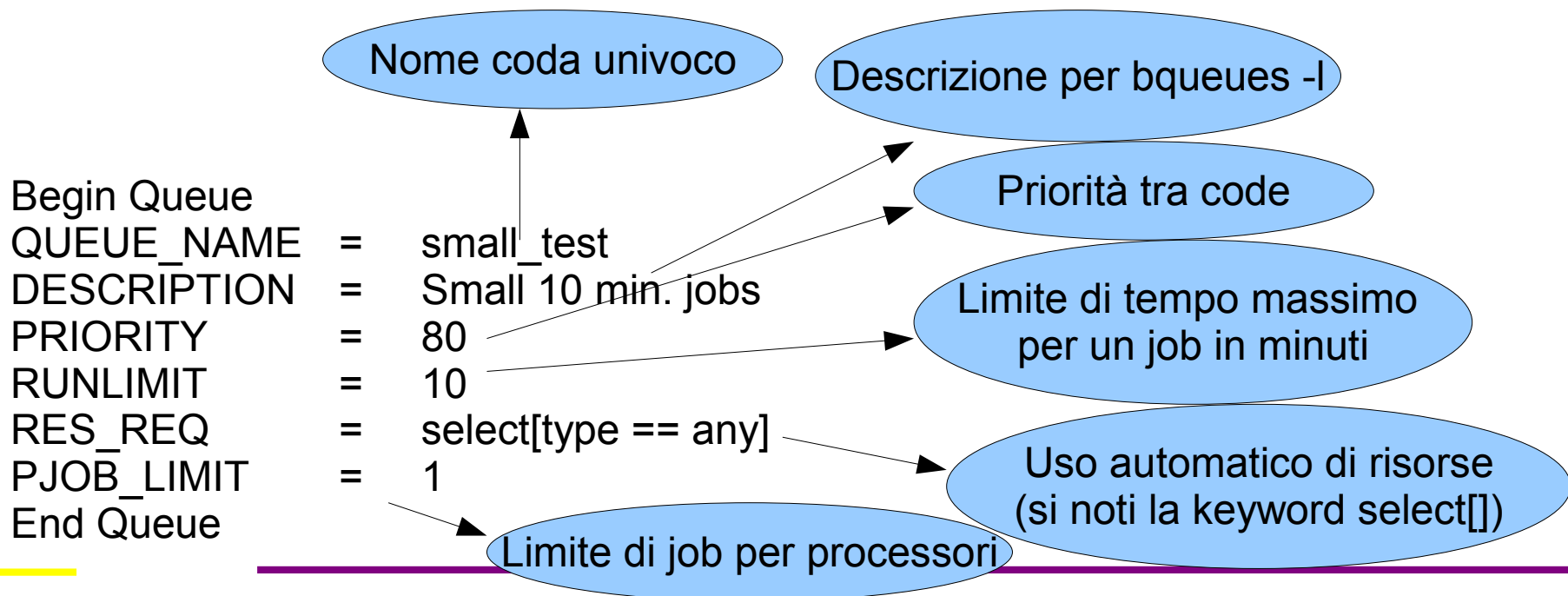
Configurazione di una coda

Il file `lsb.queues`

- `lsb.queues` si trova in `conf/lsbatch/<nomecluster>/configdir`
- È un file formato da tante sezioni “Begin Queue” ognuna delle quali definisce una coda di LSF.
- Esistono decine di parametri di configurazione delle code che si possono trovare in **`man lsb.queues`**.

Aggiunta di una coda

- Si supponga di voler aggiungere una coda per job brevi: bisogna aggiungere a `lsb.queues` una sezione vuota delimitata da “Begin Queue/End Queue” ed inserire alcuni parametri.



Aggiunta di una coda

- Dopo la modifica di `lsb.queues` (nel volume `read-write`), vanno replicate le copie `readonly` del volume `software.lsf.conf`.
- Quando modifico un file della directory `conf/lsbatch/<nomecluster>/configdir` devo lanciare solamente
 - `badmin reconfig`
- Infine, per verificare che tutto sia corretto:
 - `bqueues -l small_test`
 - `bsub -q small_test sleep 10`

Isb.queues: altri parametri utili

- **CPULIMIT**: limite imposto sul tempo di CPU (solo per job seriali)
- **STACKLIMIT**: limite stack (su alcune architetture è obbligatorio)
- **HOSTS**: limita la coda ad alcuni hosts
- **USERS**: limita la coda ad alcuni utenti
- **PRE_EXEC/POST_EXEC**: esegue script prima e dopo il job sull'host di esecuzione
- **JOB_STARTER**: usa uno script per lanciare il job

Esercizio

- Da lin4p.frascati.enea.it, collegarsi a **172.1.50.11**
 - user: lavadmin
 - password: gemuxx
- Si troverà l'installazione di una versione free di Isf utile per fare dei test.
- **COMPITO:** aggiungere una *risorsa* ed una *coda* e documentare il funzionamento delle modifiche.
- **NOTA BENE:** si tratta di un cluster virtuale in emulazione: è meglio non collegarsi in tanti.