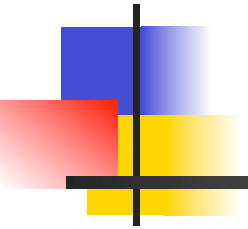


Miscellanea

Alessandro Brunengo INFN-Genova

Pending questions (I)

- Da GPFS FAQ (http://www-01.ibm.com/support/knowledgecenter/SSFKCN/com.ibm.cluster.gpfs.doc/gpfs_faqs/gpfsclustersfaq.html)
  - cNFS lock failover and failback **do not work properly on RHEL 6.*x*** due to Linux kernel issues. After replacing **/usr/sbin/sm-notify** with one of version 1.2.5 or higher, lock failover and failback will work only if there is no lock competition among clients. If there is lock competition, lock failover and failback will not work. Please reference Bugzilla 959006 - infinite loop of resends until the blocked lock is satisfied after server reboot at https://bugzilla.redhat.com/index.cgi. and https://patchwork.kernel.org/patch/2469651/

Pending questions (II)

- mmrestripefs sposta blocchi di un file da uno storage pool ad un altro solo quando il file e' hill-placed
  - hill-placed file: un file cha ha (uno o piu') blocchi su uno storage pool diverso da quello definito nei suoi attributi
- Quando si cambia il disk usage di un disco da dataAndMetadata a metadatOnly, i file contenuti sul disco non diventano hill-placed
  - nei file attribute lo storage pool rimane quello a cui appartiene il disco
- Questo e' il motivo per cui mmrestripe nella esercitazione 2 fallisce: tenta di ribilanciare i blocchi dei file all'interno dello storage pool, ma non puo' perche' lo storage pool non permette di allocare blocchi per i dati
- L'unico modo e' applicare una policy di migrazione
  - **RULE "free-system' MIGRATE FROM POOL 'system' TO POOL 'pool1'**

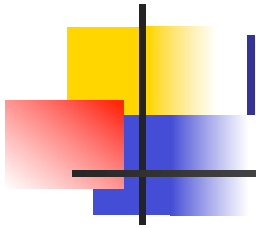
Modifica dell'indirizzo IP di un nodo

- Prima di operare la modifica dell'IP, il nodo deve essere rimosso dal cluster
- Se il nodo ha ruoli di server, questi devono essere rimossi
  - primary o secondary configuration server: mmchcluster per spostare il ruolo su altro nodo del cluster
  - quorum node: mmchnode per rimuovere il ruolo
  - NSD server
    - mmumount del file system ch ha NSD serviti da tale server
    - mmchnsd per rimuovere il nodo dall'elenco degli NSD server
  - CNFS server: utilizzare mmchnode –cnfs-interface=DELETE per rimuovere il nodo dal cluster CNFS
- Il nodo puo' ora essere rimosso dal cluster
  - mmshutdown sul nodo
  - mmdelnode per rimuovere il nodo dal cluster
- Si opera la modifica della configurazione IP del nodo
- Se la modifica dell'IP address comporta modifiche nelle subnets, applicare le modifiche tramite mmchconfig
- Si reinserisce il nodo nel cluster, tramite mmaddnode
- Si ripristinano i ruoli di server

Modifica dell'indirizzo IP di tutti i nodi

Quando tutti i nodi del cluster devono cambiare IP (reindirizzamento di sito) si devono eseguire le seguenti operazioni:

- cluster shutdown
- aggiungere a tutti i nodi il nuovo indirizzo e nome mantenendo i vecchi attivi (eventualmente utilizzando record in /etc/hosts)
- eseguire **mmchnode --daemon-interface** e **--admin-interface** per configurare i nuovi nomi/IP
- se necessario, operare modifiche alla configurazione delle subnets
- cluster startup
- rimozione dei vecchi indirizzi IP



IPv6

- GPFS (dalla 3.5) supporta IPv6
- Se alla creazione del cluster un node name risolve in un indirizzo IPv6, cluster e' automaticamente abilitato a comunicare via IPv6
- Per abilitare IPv6 su un cluster gia' esistente:
  - cluster shutdown + mmchconfig enableIPv6=yes

oppure:

  - mmchconfig enableIPv6=prepare
  - restart (uno ad uno) dei nodi del cluster; quando fatto:
  - mmchconfig enableIPv6=commit

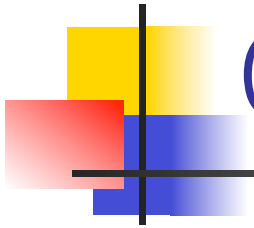
dopo e' possibile aggiungere nodi IPv6

Export file system definition tra cluster

- Puo' capitare di dover migrare la configurazione di un file system da un cluster ad un altro
  - il cluster deve essere distrutto e ricostruito da zero
  - si vuole spostare un file system da un cluster ad un altro senza spostare i dati
- Esiste una procedura:
  - verificare che tutti i dischi del file system siano ready e up
  - smontare il file system su tutti i nodi (clean umount)
  - eseguire: **mmexportfs [*fs-name* | all] –o *exportFile***
    - ora il file system e gli NSD relativi non saranno piu' visibili dal vecchio cluster
  - sul nuovo cluster: **mmimportfs [*fs-name* | all] –i *exportFile* –S *stanzaFile***
    - la configurazione di NSD e file system vengono importate nel nuovo cluster, ed il file system puo' essere montato
    - in *stanzaFile* vanno inserite le nuove definizioni di NSD server per il nuovo cluster

GPFS e firewall: daemon communication

- mmfsd: in ascolto sulla porta TCP 1191
  - configurabile tramite **mmchconfig tscTcpPort=*PortNum***
- mmsdrserv (daemon che gira su config server per servire i file di configurazione quando mmfsd e' down): in ascolto sulla porta TCP 1191
  - configurabile tramite **mmchconfig mmsdrservPort=*PortNum***
- source port per esecuzione di alcuni comandi (client side di mmfsd): porta effimera
  - configurabile tramite **mmchconfig tscCmdPortRange=*lowPort-highPort***
- administrative communication: porte usate dagli applicativi utilizzati specificati dagli switch -r e -R in mmcrcluster (ssh/scp: TCP port 22)



Cenni di monitoring

- mmpmon
- dstat

mmpmon (I)

- Utility per raccogliere dati relativi all'I/O di GPFS
- Puo' raccogliere:
  - dati di I/O eseguiti dal nodo per file system
  - dati di I/O eseguiti dal nodo globalmente
  - istogramma dei dati di I/O selezionati per size range e, all'interno di ogni range, per intervallo di latenza
- Puo' essere utilizzato per raccogliere la statistica anche su altri nodi del cluster
  - solo nodi appartenenti allo stesso cluster
- Fornisce contatori: va integrata con script di analisi dell'output

mmpmon (II)

mmpmon [-i CommandFile] [-d IntegerDelayValue] [-p [-r IntegerRepeatValue] [-s] [-t IntegerTimeoutValue]

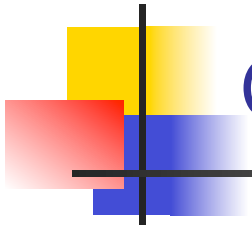
- -i: file che contiene i comandi per mmpmon
- -d: millisecondi tra due successive collection di dati
- -r: quante ripetizioni
- -t: timeout delle richieste a mmfsd
- -p: output adatto ad analisi da script
- -s: sopprime prompt

Utilizzabile in modalita' interattiva.

mmpmon (II)

- Utility piuttosto complessa da usare! Input cmd:

Request	Description
fs_io_s	"Display I/O statistics per mounted file system"
io_s	"Display I/O statistics for the entire node" on page 100
nlist add *name*[*name*...]	"Add node names to a list of nodes for mmpmon processing" on page 102
nlist del	"Delete a node list" on page 103
nlist new *name*[*name*...]	"Create a new node list" on page 104
nlist s	"Show the contents of the current node list" on page 104
nlist sub *name*[*name*...]	"Delete node names from a list of nodes for mmpmon processing" on page 105
once *request*	Indicates that the request is to be performed only once.
reset	"Reset statistics to zero" on page 101
rhist nr	"Changing the request histogram facility request size and latency ranges" on page 111
rhist off	"Disabling the request histogram facility" on page 112. This is the default.
rhist on	"Enabling the request histogram facility" on page 113
rhist p	"Displaying the request histogram facility pattern" on page 114
rhist reset	"Resetting the request histogram facility data to zero" on page 116
rhist s	"Displaying the request histogram facility statistics values" on page 117
source *filename*	"Using request *source* and prefix directive *once*" on page 123
ver	"Displaying mmpmon version" on page 119

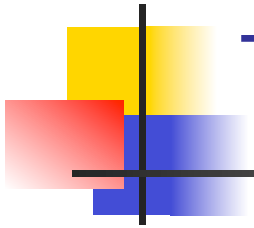


dstat

- dstat viene distribuita con un plugin per GPFS
- il plugin raccoglie solo poche statistiche di natura globale:
  - dstat --gpfs: GPFS I/O (read e write)
  - dstat --gpfs-ops: GPFS file operaton (open, close, read, write, readdir, inode ops)
- con la release GPFS 3.5 viene distribuito un plugin di gpfs-ops per dstat che fornisce molte piu' informazioni
  - le informazioni del plugin vengono prese da mmpmon!

dstat: gpfs-ops plugin

- Installare dstat (versione 0.6 o 0.7)
- I plugin sono in /usr/lpp/mmfs/samples/util/:
  - dstat_gpfsops.py.dstat.0.6 per dstat 0.6
  - dstat_gpfsops.py.dstat.0.7 per dstat 0.7
- Copiare il plugin opportuno in /usr/share/dstat/dstat_gpfs_ops.py
- Utilizzare variabili di ambiente per definire i parametri da visualizzare (vedi le prime 50 righe del plugin per la documentazione)



Todo:

- Aggiungi mmfsdiag
- Aggiungi utility in samples