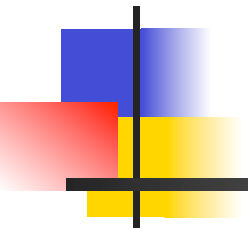
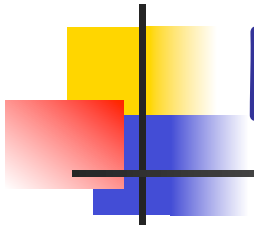
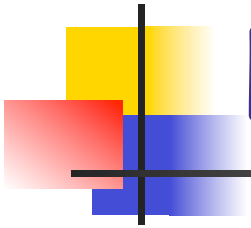


# Fileset, quote, storage pool

Alessandro Brunengo INFN-Genova



# Fileset



# Fileset

- GPFS supporta il concetto di fileset, che e' sostanzialmente un sottoalbero del file system che dal punto di vista amministrativo si comporta come un file system indipendente
  - e' possibile definire una quota per fileset
  - e' possibile definire user/group quota per fileset
  - si possono definire policy di collocazione e movmentazione di file per fileset
  - si possono creare snapshot di singoli fileset
- Il fileset e' identificato da una stringa di caratteri che deve essere univoca all'interno del file system

# Fileset e i-node space

- Il fileset puo' essere
  - indipendente: lo spazio di i-node e' dedicato al fileset
    - questo ottimizza funzioni di scan dei file di un fileset, ad esempio nella applicazione di policy
  - dipendente: lo spazio di i-node e' quello del file set *root* o di un altro fileset indipendente
- Alla creazione del file system viene automaticamente creato il fileset *root*
  - non puo' essere cancellato
  - la radice del fileset *root* coincide con la root del file system
  - contiene file di sistema, come i file di quota

# Junction del fileset

- Il fileset creato contiene una root directory vuota, e non e' visibile
- la accessibilita' del fileset viene realizzata creando un junction point all'interno del *root* fileset o di un fileset visibile
  - la junction ha l'aspetto di una normale directory (comprese le permission) ma non si possono eseguire le operazioni di rmdir e unlink su di essa

# Creazione del fileset

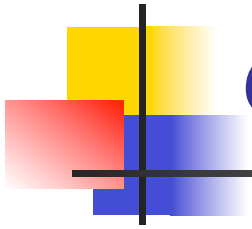
- mmcrfileset <dev> <fs-name> [--inode-space <spec>] [-p <afm-attribute>…]
  - --inode-space new: fileset indipendente
  - --inode-space <existing-fileset>: crea un fileset dipendente che condivide l'i-node space con il fileset specificato

# Link/unlink il fileset

- mmlinkfileset <dev> <fs-name> -J <junction path>
  - rende visibile il fileset posizionando la sua root sotto <junction path>, che viene creata col comando
- mmunlinkfileset <dev> {<fs-name>|-J <junction-path>} [-f]
  - rende il fileset invisibile
    - i file vengono conservati (ed i blocchi restano allocati!)
    - -f per forzare l'operazione, che fallisce se esistono file open entro il fileset

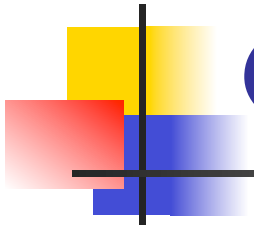
# Altre operazioni sul fileset

- mmlsfileset <dev> …
  - visualizza le caratteristiche di uno o di tutti i fileset
  - vedere la man page
- mmchfileset <dev> <fs-name> …
  - modifica i parametri del fileset (nuova junction point, dimensione di i-node space per fileset indipendenti, attributi AFM)
- mmdelfileset <dev> <fs-name>

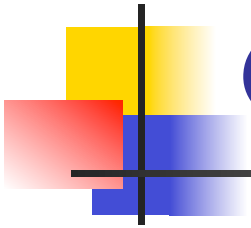


# df sul fileset

- Per default, df eseguito su una dir di un fileset mostra lo spazio available/ used dell'intero file system
- Se un file system e' configurato con il parametro **--filesetdf**, e sul fileset e' definita una quota, df mostra lo spazio available/used del solo fileset, come definito dalla quota



# Quota



# Quota management

- GPFS supporta il controllo di quota per l'allocazione di dati e/o i-nodes
- Il controllo di quota puo' essere attivato per file system, relativamente a user, group e fileset
  - a partire dalla release 3.5 puo' essere attivato il controllo di user o group quota all'interno del singolo independent fileset
- I dati sulla quota vengono mantenuti in memoria e nei file user.quota, group.quota e fileset.quota nella root del file system
- La quota e' imposta attraverso la definizione di soft limit, hard limit, e grace timeout
- Attenzione ai file system con replica di dati/metadati
  - la replica dei dati raddoppia lo spazio disco contreggiato
  - la replica dei metadati raddoppia gli i-nodes conteggiati

# Abilitazione della quota

- Per abilitare la gestione della quota su un file system alla sua creazione, si deve eseguire **mmcrfs** con il parametro -Q yes
  - questo attiva automaticamente la quota ad ogni mount
- Per abilitare e attivare il controllo di quota su un file system precedentemente creato, si deve modificare il parametro a file system smontato su tutti i nodi:

```
# mmumount <device-filesystem> -a
# mmchfs <device-filesystem> -Q yes
# mmmount <device-filesystem> -a
```

una volta attivato il controllo di quota si deve compilare la statistica sull'attuale utilizzo di file e data blocks:

```
# mmcheckquota <device-filesystem>
```

# Disabilitazione del controllo di quota

- Per disabiltare il quota management, si deve effettuare l'operazione inversa:

```
# mmumount <device-filesystem> -a
# mmchfs <device-filesystem> -Q no
# mmmount <device-filesystem> -a
```

I file della quota rimarranno nel file system, ma non saranno piu' utilizzati

# Attivazione/disattivazione del controllo di quota

- E' possibile disattivare temporaneamente il controllo di quota su un file system, anche limitatamente a user, group o fileset quota:

  **# mmquotaoff [-u | -g | -j] <filesystem>**

- Per riattivare il controllo di quota precedentemente disattivato:

  **# mmquotaon [-u | -g | -j] <filesystem>**

- Si puo' visualizzare lo stato del quota management sul file system con il comando

  **# mmlsfs <filesystem> -Q**

```
flag  value              description
------------------       --------------------
-Q    group;fileset      Quotas enforced
```

# Default quota

- E' possibile definire valori di quota di default ed abilitare il filesystem ad assegnare tali valori a nuovi user, group o fileset
- I comandi per attivare o disattivare l'assegnazione della quota per default sono **mmdefquotaon** ed **mmdefquotaof**
  - il file system deve avere il controllo di quota abilitato (**-Q yes**)
  - l'assegnazione di limiti di default puo' essere fatta indipendentemente per file system e per user, group o fileset
- per visualizzare lo stato della attivazione di quota di default:

```
# mmdefquotaon -g fs1

# mmlsfs  fs1 -Q

flag value          description
---- -------------- ------------------------
-Q  group           Quotas enforced
     group          Default quotas enabled
```

# Assegnazione della quota

- La quota si assegna tramite il comando **mmedquota**
  - il comando permette di definire i parametri della quota tramite una sessione di editing
  - e' possibile assegnare la quota uguagliandola ad altri user, group o fileset (**mmedquota -p**)
    - questo permette l'utilizzo del comando da script
- **mmdefedquota** deve essere utilizzato per definire i valori della **quota di default** per user, group o fileset (**mmedquota -d**)

# Visualizzazione della quota

- Il comando che visualizza lo stato di un singolo utente/ gruppo/fileset e' **mmlsquota**
  - il comando mostra blocchi ed inodes utilizzati, utilizzabili (hard e soft limit), ed ***in_doubt***
    - i valori ***in_doubt*** sono la quantita' di quota assegnata come disponibile ai client del cluster ma non ancora conteggiata
    - puo' essere quota usata o non usata
    - in occasione di una failure di un nodo, i valori ***in_doubt*** potrebbero rimanere non aggiornati
  - la somma di **quota utilizzata** ed ***in_doubt*** non puo' superare l'**hard limit**
  - utilizzare **mmlsquota -d** per visualizzare i valori di default quota
- Per avere un report sull'utilizzo della quota in un file system, utilizzare **mmrepquota**

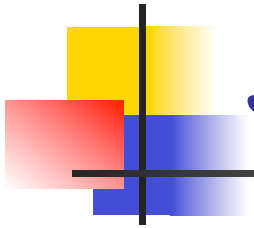
# Controllo della quota

- mmcheckquota esegue una rivalutazione della quota utilizzata effettivamente sul file system
  - **mmlsquota** fornisce i valori attualmente memorizzati sul file system manager, che potrebbero essere **piu' o meno out-of-date**
  - in particolare, mmcheckquota esegue un controllo sulla quota ***in_doubt***, rimuovendo quella parte di quota ***in_doubt*** che potrebbe essere rimasta erroneamente conteggiata a causa di **node failure**
- **mmcheckquota** deve essere utilizzato qualora la percentuale della quota ***in_doubt*** dovesse divenire importante
  - questo e' un segno di un possibile conteggio errato di quota in_doubt

# Restore dei quota file

- I file contenenti la configurazione delle quote possono corrompersi
  - in questo caso compaiono errori MMFS_QUOTA nei log file
- E' possibile recuperare tali file da backup
  - ad esempio, per recuperare la confiugrazione della user quota:
    - restorare il file user.quota in userquota.bck
    - eseguire il comando mmchechquota -u userquota.bck fs1
    - ricalcolare l'utilizzo attuale: mmcheckquota fs1
- In assenza di backup:

1. **# mmumount fs1 -a**
2. **# mmchfs -Q no**
3. **# mmmount fs1**
4. **# rm /gpfs/fs1/*.quota**
5. **# mmumount fs1**
6. **# mmchfs -Q yes**
7. **# mmount fs1 -a**
8. **eseguire mmedquota per ridefinire le quote**
9. **# mmcheckquota fs1**



# Storage pool

# Storage Pools

- Lo storage pool e' un raggruppanemto (named) dei dischi che costituiscono il file system
- Il contenuto di ogni singolo file e' assegnato ad uno storage pool specifico, in base a policy rules
  - placement policies (dove mettere il file alla creazione)
  - migration policies (spostare file da un pool ad un altro)
  - deletion policies (rimuovere file da uno storage pool)
- Gli storage pool servono a:
  - Implementare tiered storage (diversa qualita' di dischi per diversi dati)
  - Avere prestazioni omogenee tra i dischi usati per un file
  - Realizzare storage dedicato (per user, per project o per directory subtree)
  - Limitare la perdita di dati in caso di failure
  - Limitare l'impatto di un RAID rebuild sulle prestazioni

# Internal storage pool (I)

- Gli internal storage pool sono pool che contengono dischi conosciuti e gestiti da GPFS (gli NSD)
  - Lo strage pool e' definito da una stringa univoca per il file system
  - L'appartenenza ad uno storage pool e' un attributo del disco definito al momento di inserire un disco nel file system (mmcrfs o mmadddisk) o spostato tramite migration policy
    - inserendo nel file system un disco in un pool inesistente, il pool viene automaticamente creato
    - lo storage pool di default e' "system", e deve essere sempre definito in un file system

# Internal storage pool (II)

- System storage pool
  - lo storage pool system e' l'unico che possa contenere metadati; puo' contenere anche dati; deve esistere ed e' il pool di default
- User storage pool
  - uno o piu' pool in cui e' possibileinserire dischi contenti esclusivamente dati (disk usage: dataOnly)
- Combinanto l'attributo "disk usage" dei dischi e gli storage pool e' possibile realizzare la separazione di dati e metadati in diversi storage pool:
  - inserire nel pool system solo dischi per uso metadataOnly o descOnly
  - inserire dischi per uso dataOnly in pool diversi da system

# Storage pool e block size

- E' possibile definire valori diversi per la block size degli user storage pool e per il system storage pool
  - user storage pool block size: quella del file system (-B)
  - system storage pool block size: definita dal parametro --metadata-block-size

# Placement policy

- Quando un file viene creato i dati vengono inseriti in un pool definito da una placement policy
- Per default si utilizza il pool system
- Definizione di una placement policy:
  - scrivere in un file la policy di placement
  - attivare la policy di placement tramite il comando

    # mmchpolicy *Device Policy-file-name*

# Placement policy: esempio

- Creare il file placement.txt per collocare i file del solo fileset *fset1* file system */dev/fs1* nel pool *data*:

  RULE 'rule1' SET POOL 'data' FOR FILESET ('fset1')
  RULE 'default' SET POOL 'system'

- Attivare la policy:

  # mmchpolicy /dev/fs1 placement.txt

- Visualizzare la placement policy attiva:

  # mmlspolicy /dev/fs1 –L

- Restorare la policy di default:

  # mmchpolicy /dev/fs1 DEFAULT

# External storage pool

- GPFS supporta la definizione di pool esterni (low cost near-line o off-line storage, come tape library), non gestiti da GPFS
- Questi vengono definiti tramite policy (vedi sessione ILM/Policy) in cui e possibile specificare l'interfaccia tra GPFS ed il gestore dello storage esterno
- GPFS gestisce i metadati dei file collocati negli external pool ed accede all'occorrenza a tali dati migrandoli dal pool esterno tramite l'interfaccia
  - in questo modo GPFS supporta tutti gli attributi estesi dei file anche se collocati su storage esterno
- Tramite policy e' possibile migrare dati da/verso lo storage esterno in occasione di eventi come ad esempio occupazione percentuale del file system inline

# Filesets and Storage pools

- Storage pools allow the creation of disk groups within a file system (hardware partitioning)
- Filesets is a sub-tree of the file system namespace (Namespace partitioning).
  - Behave like separate file systems
  - can be used as administrative boundaries to set quotas
  - In 3.5 Independent Filesets
    - Using separate i-node space
- Policy to be used to connect SP and Filesets
- Default policy writes everything to system SP