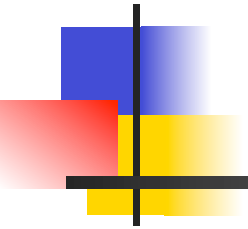
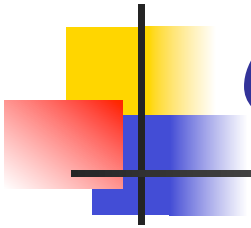


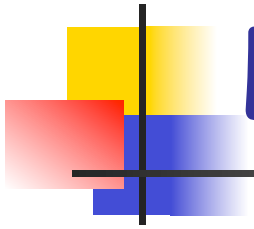
Problem Tracking

Alessandro Brunengo INFN-Genova



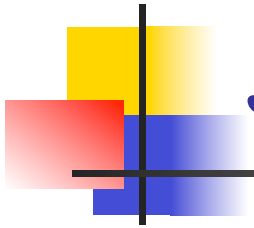
Contact IBM

- Nel documento "Problem Determination Guide" c'e' un intero capitolo su **come si contatta IBM**
- E' evidente l'apprensione di IBM per i tentativi di recovery eseguiti dai suoi clienti
  - GPFS e' complicato e fortemente auto-correttivo
  - Quando l'auto-correzione non va, il problema potrebbe essere molto serio
- Se il problema pare complesso, i dati sono molto importanti, e non recuperabili altrove, teniamo presente l'eventualita' di chiamare IBM.
  - … non che questo ci garantisca dalla perdita di dati…



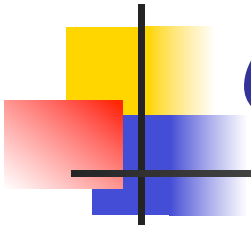
Log file di mmfsd

- /var/adm/ras/mmfs.log.<date>.<node>
  - ultimi logs linked in mmfs.log.latest e mmfs.log.previous
  - prima sorgente di informazioni
- Attenti: siamo in un cluster environment
  - quello che accade ad un nodo puo' **dipendere da altri nodi, o influenzarli**
  - puo' essere necessario analizzare i log di tutti i nodi del cluster
  - provare a modificare ed eseguire /usr/lpp/mmfs/samples/gatherlogs.samples.sh



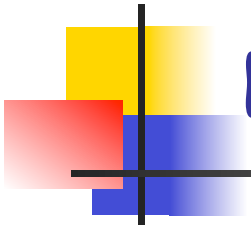
System logs

- mmfsd scrive logs nel log file di sistema /var/log/messages
  - grep mmfs: /var/log/messages
  - necessario per identificare eventuali problemi hardware o di sistema che possono aver generato problemi a GPFS
- Vari errori, tra cui
  - **MMFS_FSSTRUCT**: problemi sulla struttura (su disco) di un file system
  - **MMFS_LONGDISKIO** (warning)
  - **MMFS_GENERIC**: internal GPFS check problem
  - **MMFS_PHOENIX**: messaggio dell'high availability layer
- Solitamente associati ad uno o piu' internal code
  - Riferimenti sui significati dei codici nel cap. 11 di "**GPFS Problem Determination Guide**", parte della documentazione della release di GPFS.



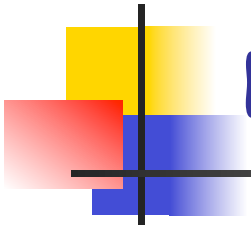
Collect status information

- gpfs.snap
  - esegue il collect di una grande quantita' di informazioni sulla configurazione e lo stato (GPFS e sistema) di uno o piu' nodi
  - utile piu' per fornire i dati al supporto IBM che per fare un vero problem tracking (troppi dati, praticamente tutto)



mmfsadm

- IBM dice che si deve usare solo se indicato da IBM
- Legge dati da GFPS daemon senza usare lock
  - funziona anche con problemi di locking interni
  - puo' far andare in crash mmfsd o il nodo
- L'output varia da release a release
  - non si devono fare script basati sul suo output



mmfsadm usage

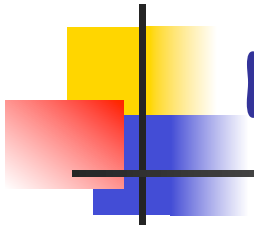
- interattiva, oppure:
  - mmfsadm dump <what>
  - lo scope di mmfsadm e' il nodo di esecuzione
    - spesso importante guardare l'output sul file system manager
  - mmfsadm help per vedere cosa si puo' fare

mmfsadm dump usage

dump what: Dump data structures and statistics, where 'what' can be:

alloc, alloc all, alloc stats, allocmgr, allocmgr all, allocmgr stats, allocmgr hint, brt, buffers, cfgmgr, condvar, config, DACspy, dealloc, dealloc stats, dealloc all, disk, dmapi, eventsExporter, files, filesets, filocks, fs, fsck, fsmgr, ialloc, ialloc all, iallocmgr, iallocmgr all, indblocks, indirect, instance, iocounters, iohist, kthread, llfile, lock, log, lstat, malloc, mb, mmap, mutex, mutex all, nsd, afm, perfmon, nsdrg, pdisk, pdiskdevs, vdisk, vdiskbufs, vtracks pgalloc, pgalloc all, pit, quorumState, quota, reclock, reclockSleepers, reclockStats, res, sanergy, sgmgr, stripe, sxlock, thread, thread all, threadstacks, threadstats, tmstats, tokenmgr, tokens, tmcomm, updatelogger, disk, verbs, version, vfsstats, vnodes, waiters, winsec,

or all



mmdiag

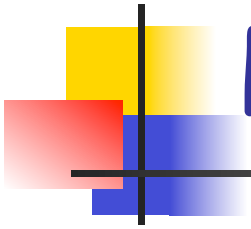
- Accede ad informazioni sullo stato del GPFS daemon
- Si sovrappone in parte a mmfsadm dump
- Si puo' specificare:
  - --waiters, --threads, --memory
  - --network, --config, --trace, --tokenmgr
- Anche informazioni statistiche:
  - --iohist, --stats

GPFS tracing facility

- GPFS integra il supporto al tracing di svariate sue parti
- Per raccogliere informazioni di trace:
  - verificare che la directory per la raccolta dei dati di dump esista
    - mmdiag --config | grep dataStructureDump
    - mkdir -p <dataStructureDump>
  - utilizzare mmtracectl per configurare parametri del tracing se i default non vanno bene (istruiti da IBM, ovviamente)
  - far partire il trace: **mmtracectl –start**
  - indurre l'evento da analizzare
  - fermare il tracing: mmtracectl --stop (non dimenticarsi!)
- mmtracectl --trace-recycle=<val> per attivare il trace allo startup se il problema e' li'

Problemi sui configuration file

- mmrefresh [-f] [-a | -N <node>]
  - ricarica i file di configurazione up-to-date sul nodo <node>
  - -f: forza la rigenerazione dei file, anche se sembrano up-to-date
- mmsdrrestore [-p <fromnode>]
  - riprendi la configurazione integralmente
  - -p se si deve specificare da quale nodo



Problemi sul file system

- Restricted mode mount (last resort)
  - quando la perdita di un disco fa perdere tutto il file system…
  - dopo aver cercato di recuperare i dischi…
  - dopo aver eseguito mmfsck…
  - dopo aver eseguito l'editing a mano dell'NSD descriptor (alla Vladimir, per intenderci)…
  - allora prova **mount -o rs**

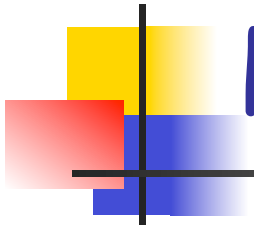
Quali file abbiamo perso?

- mmfileid <device> {-d <diskdesc> | -F <descfile>} [-N <node>] ...
- Identifica i file che hanno parti in aree di un disco
- per <diskdesc> si puo' usare ":{NSDname|BROKEN}[:start[:stop]]"
  - DNSname: identifica il disco da usare
  - BROKEN: scan di tutti i dischi del file system per trovare file su parti danneggiate
- in alternativa si puo' usare un file <descfile> che contiene un descriptor per linea

mmfileid output (esempio)

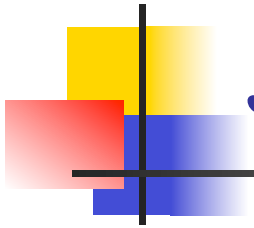
- Address 2201958 is contained in the Block allocation map (inode 1)
- Address 2206688 is contained in the ACL Data file (inode 4, snapId 0)
- Address 2211038 is contained in the Log File (inode 7, snapId 0)
- 14336 1076256 0 /gpfsB/tesDir/testFile.out
- 14344 2922528 1 /gpfsB/x.img

- Linee che iniziano per Address indicano che la zona del disco contiene metadati o file strutturali (malemalemale...)
- Le altre linee indicano <inode> <indirizzo del disco> <snapshot di appartenenza> <nome del file cui appartiene quella zona e che potrebbe essere perduto>



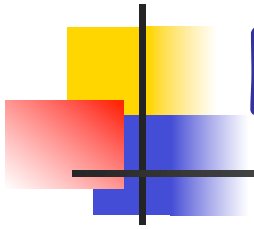
Problemi

- Molti problemi diversi
- Cause diverse
  - spesso problemi di layer sottostanti (SAN, rete, disco) o laterali (OOM Killer)
- C'e' solo una regola valida sempre:
  - Calma.
- Possono aiutare: pensare, accedere alla documentazione, accedere ai forum di utenti (e Vladimir, ovviamente)



Sito wiki

- Mailing list gpfs@lists.infn.it
  - avvisi urgenti
  - richieste di aiuto di prima istanza
- In preparazione sito wiki su cui inserire informazioni utili al riguardo
  - esempi di utlizzo di utility
  - analisi di problemi comuni
- Vladimir vi fara' vedere qualcosa…



Discussione aperta

Aperta la sessione Problem Storming