Backing up Very Large Databases

Charles Silvan
GATE Informatic
Switzerland
Agenda

- Environment
- Flashcopy Principles
- TSM for ACS
- Numbers
- Issues
- Scalability
Customer Environment

- Four Datacenters worldwide
- Each one over two sites, separated by 20-40 km
- Everything LVM mirrored, HACMP
- Factory approach to installation and maintenance
- Each datacenter hosting 10-20 SAP DB2 databases of n TB’s
SAN Layout

Site A

- Edge 1A
- Core 1A
- DS8000
- TSM Server A
- Tape Library

Site B

- Edge 1B
- Core 1B
- DS8000
- TSM Server B
- Tape Library

C. Silvan

Back up Very Large Databases
Flashcopy/Flashrestore Principles

- Set of LUN pairs: source-target, same sizes
- Flashcopy establishment: pointers copied to target LUN’s = Point-in-time snapshot
- Data on target LUN immediately accessible
- After all data is copied to targets: can reverse relationship: Flashrestore
- After first full copy: can use Incremental Flashcopy
Flashcopy Backup to Tape

- Database set to "write suspend" mode
- Flashcopy pairs establishment
- Database resumed
- Target LUN’s are accessed from Backup Mover system
- Importvg, mount, etc
- Database is started on Mover System
- Online DB backup to TSM tape, in LAN-less mode
Flashcopy Benefits

- Flashcopy Backups: minimize impact on production database systems
  - short quiesce time
  - Zero load on production server
  - only reading load on production disks

- Costs:
  - double the disk space requirement
  - setup
Flashrestore Benefits

- Reduces restore time from hours to minutes (basically independent of database size)
- Eliminates data transfer from tape
- Forward recovery can start immediately after reverse Flashcopy relationship is established

Note: in case of tape restore: LAN-less to production system
Flashcopy Landscape

![Diagram showing Flashcopy Landscape with components like TSM1, Prd1, Prd2, TSM2, BM1, BM2, SAN, MAN Links, LTO Tape Library, DS8K#1, DS8K#2, Source, Target, HACMP, LVM Mirror.](image)
TSM for ACS
(= TDP for Flashcopy)

• TSM for Advanced Copy Services provides a solution for Flashcopy backup in SAP environment

• Software stack:
  SAP, DB2, DP for SAP, DP for Flashcopy, TSM Client, TSM Storage agent, CIM Agent, DS8K SMI-S API, CIM client (Pegasus)

• Main limitation: with LVM mirroring, DB must reside on one disk subsystem
Complexity Aspects

- Multiple systems
- Java: installers, products
- Code levels dependencies: DS8000, CIM Agent
- Log files
- Password (DB2, TSM, CIM, DS8000)

All of this make automated installs and maintenance complicated
Production Issues

High percentage of failed backups, for many different reasons:

- Setup: NFS, passwords, rexec, LVM/DB changes, code upgrades, DS8000 LUNs, etc
- Tapes: drives not available, tape preemption
- Performance: MAN links limits, disk subsystems

As DB’s were expected to grow by a factor 5 to 10, customer asked IBM:

*How do you backup a database of 20, 40, 80 TB?*
IBM Montpellier Study

- Real size tests, over several months, by top IBM specialists, and lots of hardware
- Documented in a Redbook: SG24-7289
  « Infrastructure Solutions: Design, Manage, and Optimize a 20 TB SAP NetWeaver Business Intelligence Data Warehouse »
- Also includes proposals for scaling up to 60 TB
Some Conclusions from Montpellier

- Dedicate DS8K ports for tape backups, one per tape stream
- Spread Flashcopy source and targets on all arrays: when no backup, use full DS8K power; when backups, equivalent to separating them
- Backup time: 4.5 hours for 20 TB, using 16 LTO-3 drives = 4.4 TB/hr, 300 GB/hr/stream
Transfer Rates Arithmetics

- Backup or restore 42 TB in 7 hours = 6 TB/hr. Assuming 300 GB/hr per stream (=83 MB/s):
  - need 20 parallel streams
  - aggregate data rate: 1660 MB/s
- For 63 TB DB in 7 hours: 9 TB/hr:
  - 30 parallel streams
  - aggregate: 2500 MB/s
- If over the MAN: 9 - 13 x 2 Gbps links
- Conclusions:
  - cannot afford backups going over the MAN
  - cannot afford taking a second copy
Flashcopy Numbers

- Full Flashcopy of 42 TB DB: assuming 1.8 TB/hr, or 500 MB/sec read + write (Montpellier numbers), requires 23 hours on one DS8000
- For 63 TB: 33 hours
- To go down to 8 hours elapsed: spread the load on 3 to 4 DS8000's
How many DS8000's?

- Capacity-wise: 2 DS8000 (per site) for a 42 TB DB
- Performance-wise: I/O loads:
  - Production: ? read, ? write, over 24 hours
  - Full Flashcopy: 1700 MB/s read, 1700 MB/s write, over 7 hours
  - Backup to tape: 1660 MB/s read, over 7 hours
  - Logs: 23 MB/s write, 23 MB/s read. Peak: 50 MB/s
- Total I/O Capacity of one DS8000: about 1100 MB/sec
- Backup to tape plus Full Flashcopy represent about 3 x DS8000 I/O capability

- So: need at least 4 x DS8000, for performance
Backup Movers

- 20 / 30 tape streams: need more than one Backup Mover
- Go for 4 Backup Movers, one per DS8000
- Each one with 6 / 8 tape streams
- Using 4 DS8000 ports, and 4 tape HBA's (at 4 Gbps)
- Have HACMP Clusters of Backup Movers? If one Backup Mover fails, the whole DB backup has failed
New 5x2 LPAR Database
Tape Drive Allocation

- Need to guarantee that 20 tape drives will be available at backup time.
- If the TSM server used for more than one DB: not possible
- The only way is to use a Library Manager-only TSM server, which allows to dedicate drives
Alternate Backups

- Goal: avoid Copy storagepool by backing up alternatively to two tape libraries, one on each site
- One site disaster: recovery from -1 backup
- Intelligence is in the (SAP) client scripts: backup to other Mgmtclass (=library) than last backup
- Still to one TSM server
- Needs new definitions: primary stgpoools on remote library and new set of Management classes: more complexity
Archive Logs

Archive Logs are even more critical than DB backups: if one log is missing, cannot roll-forward past that point

- Continuous flow of 2 GB objects, about 100 GB/hr
- Want to keep logs from last N days on disk, for restore
- If need to restore from tape, must have parallelism
- Cannot live more than a few hours without archiving (and deleting) the logs
MIGDELAY

- To keep last N days on disk: use MIGDELAY

- Problems:
  - MIGDELAY vs MIGCONTINUE
  - No practical was to get the age of archive files in a disk storage pool
Collocation Issues

- Collocation has been invented to avoid files of one node or filespace from being spread on many tape volumes.
- In this case, we need to migrate files from one filespace to multiple tape volumes (for parallel restores): not possible if same node and filespace.
TSM Passwords

- TSM password must be correct, at all times, on:
  - All production hosts and their take-over hosts
  - All Backup Mover systems
- Cannot use CLUSTER=YES
- Use ASNODE, but still need to authenticate from each host
- Each node still needs to be backed up
- -> in TSM server: 18 x 2 nodenames, 36 password files
Testing

- Can you afford a test environment with 20 hosts, 8 DS8000, 40 - 60 tape drives?
- Test/validation needed for:
  - Setup changes
  - Software level changes
  - DS8000 microcode changes
- Development Labs have the same problem!
- How do you verify that a backup image is OK, without having to do a full restore?
General Issues

Most products were not designed for these sizes:

- Large system effects
- Restart capabilities
- Failure containment

- Error analysis
- Performance monitoring
Scalability

- **Scalability options:**
  - technology: faster disks and tapes. Will it keep up with data growth?
  - horizontal scaling: more DS8000's, more tape drives, more Backup Movers, more LPAR’s. Can go a very long way.

- But only up to a point:
  - complexity
  - skills
  - exposure to one failing element
Conclusion

- Q: How do you back up very large databases?
- A: Very carefully

... or not at all.