Optimizing ASM Deployments for Resiliency, Data Protection, Utilization, and Performance

Alvin Richards alvinr@netapp.com

Technical Architect

Storage Mgmt. & Application Integration

Agenda

- ASM Overview
- Considerations and Recommendations
  - Data Resilience
  - Performance
  - Storage Utilization
  - Data Protection
- Summary

Agenda

- ASM Overview
- Considerations and Recommendations
  - Data Resilience
  - Performance
  - Storage Utilization
  - Data Protection
- Summary

Oracle® Automatic Storage Management

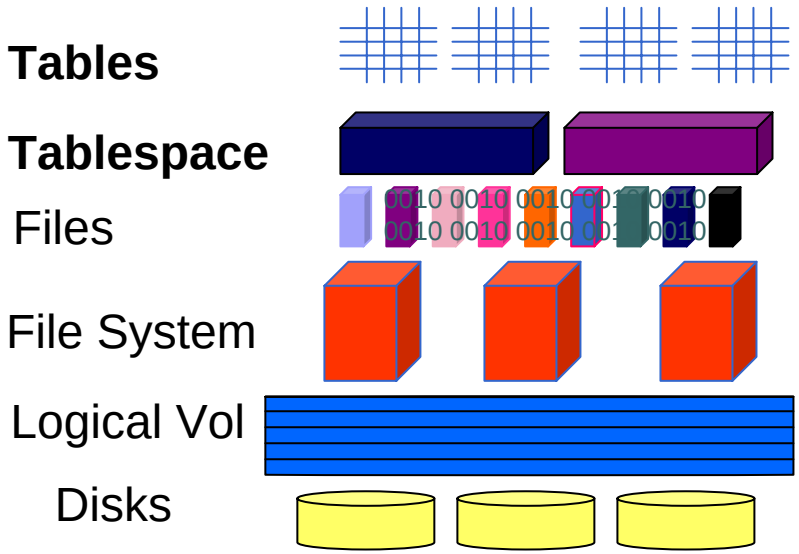
A portable **storage manager** designed to manage Oracle Database 10*g* database files – offered at no additional cost

- Volume Manager
- Clustered File System (for RAC)

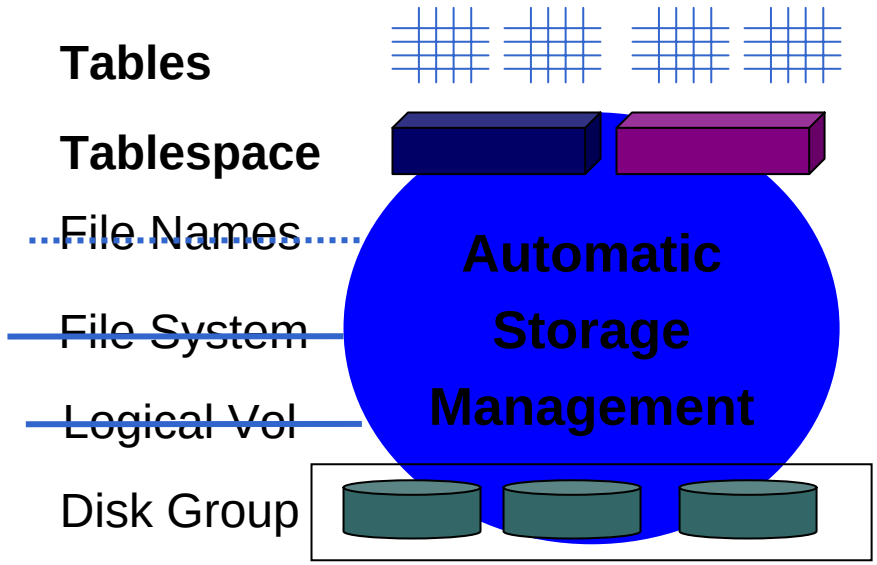
Reduce Cost and Complexity Without Compromising Performance or Availability

ASM: The Operational Stack

Before ASM



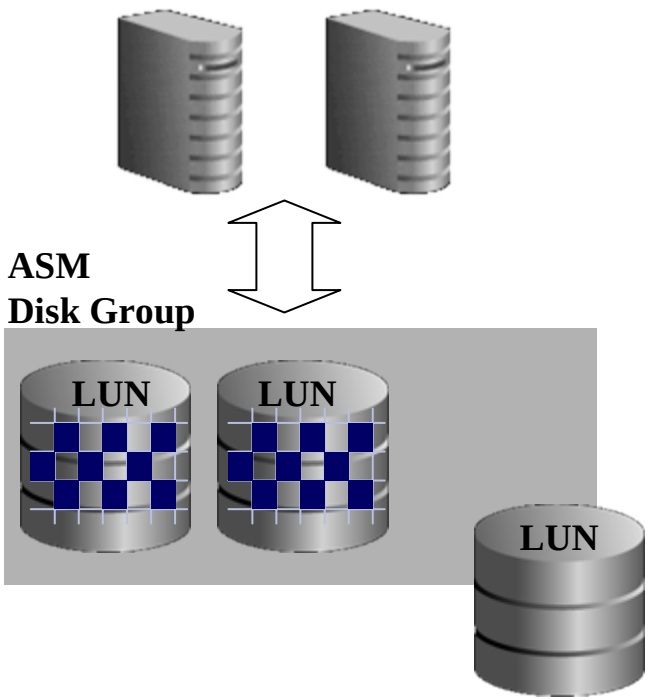
ASM



Networked Storage

(SAN, NAS, DAS)

ASM Key Features



1. Volume Management
2. Database File System with performance of RAW I/O
3. Supports clustering (RAC) and single instance
4. Automatic data distribution
5. On-line add/drop/resize disk with minimum data relocation
6. Automatic file management
7. Flexible mirror protection

Automation

Taking Control of Storage

- **Traditionally DBAs did not control storage**
  - Storage or Host Admin's responsibility
  - DBA went to these guys when something went wrong (or slowly)
- **ASM Changes your DBA world**
  - ASM enable a DBA to manage storage directly
  - DBA now has to consider
    - How do I deal with disk failures? Data resilience
    - How do I back up storage? Data protection
    - How do I make the best use of the raw storage? Storage Utilization
    - How do I make it run fast enough? Performance

Agenda

- ASM Overview
- Considerations and Recommendations
  - Data Resilience
  - Performance
  - Storage Utilization
  - Data Protection
- Summary

Data Resilience Considerations

- Disk failures
- Block corruptions
- Lost writes

Protection against Disk Failure

- **Must retain data despite disk failure(s)**
- **Alternatives**
  - ASM Redundancy (mirroring)
  - Hardware RAID-1 (mirroring)
  - Hardware RAID-5 (single disk parity protection)
  - Hardware RAID-6 (double disk parity protection)
- **Choice depends on**
  - Business requirements
  - Budget (cost, availability, performance & utilization)

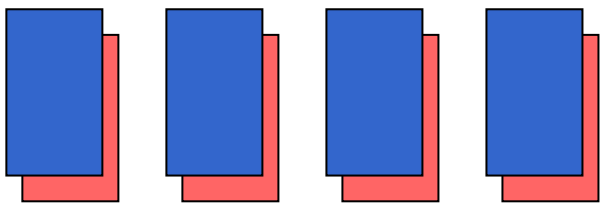
ASM can leverage both software and hardware protection

Mirroring versus Parity Protection

Mirroring (RAID-1)

PRO
- Best redundancy
- Best performance
- Low recovery overhead

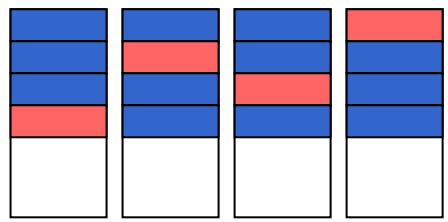
CON
- Requires double capacity
- Cannot survive failure of mirror pair



Parity Protection (RAID-5)

PRO
- Requires less capacity

CON
- Less redundancy
- Less performance
- High recovery overhead



RAID Comparison - Space and Performance Implications

RAID Level	Failure Protection	# of Disks *(Cost Hit)*	Performance Impact
RAID-1 *(Mirroring)*	Multiple disks, *except* mirror pairs	2N *(N or 100%)*	2x writes; none to add or reconstruct
RAID-5 *("RAID")*	1 failed disk or uncorrectable disk (hard) errors	N+1; N≈7 *(1/N or 14%)*	XOR calculation on writes, data reconstructed and when disks added
RAID-6 (RAID-DP™)	2 failed disks or 1 failed disk *and* hard disk errors	N+2; N≈14 *(2/N or 14%)*	XOR calculation on writes and when data reconstructed; *none when adding disks*

RAID-6 provides RAID-1 protection & efficiency with RAID-5 costs

Block Corruption - Causes and Prevention

- **Caused by HBAs, Switches, Network, etc.**
- **Active checking (occurs on write)**
  - Detection and prevention of
    - Writes that logically or physically corrupt blocks
    - Writes of partial or incomplete blocks
    - Writes to incorrect locations or by other applications
  - Oracle® HARD allows vendors to re-compute block checksums
  - Increase latency for writes, as checksum is recomputed (typically < 5%)
  - Database configuration (init.ora, `alter system`)
    - `DB_BLOCK_CHECKSUM = OFF | TYPICAL | FULL`
    - `DB_BLOCK_CHECKING = OFF | LOW | MEDIUM | FULL`
- **Passive checking (user initiated)**
  - Check for block corruption on backup (e.g. RMAN)
  - `dbv` can be run against database files (online, backups etc.)
  - `alter table <table> validate structure cascade`

Lost Writes
- “Disks lose writes? No way!”

- **Disk driver gets successful write completion, but subsequent read of block returns old contents**
  - Disk block X contains data D1
  - Disk driver writes data D2, and write completes successfully
  - Subsequent read returns data D1 instead of expected data D2
- **Block checksums don’t help**
  - Data and checksum get lost together
  - On read, old data and checksum still match each other
- **Zoned checksums cannot detect lost writes in some cases**
  - Data and checksums stored on different blocks and updated using separate I/Os
  - If both I/Os are lost, checksum and data will match and corruption cannot be detected

ASM cannot detect lost writes, requires storage based solution

Data Resilience - NetApp Recommendations

- **Use Hardware RAID-DP™ (RAID-6)**
  - ASM Mirroring (RAID-1) is not as reliable RAID-DP
    - RAID-1 cannot survive the double disk failure of the mirror pair
  - Mirroring requires double or triple the disk capacity depending on the redundancy chosen
  - RAID-DP requires only two additional disks per RAID group

- **Protect against "lost writes"**
  - NetApp storage servers automatically detect lost writes
  - Transparent to the application, lost write is reconstructed on read automatically
  - Highly efficient, does not require read-after-write verification, sampling or other expensive processing

Data Resilience
- NetApp Recommendations

- **Use block corruption detection solution**
  - Deploy NetApp SnapValidator™ to eliminate corrupt block writes
    - Implements Oracle® HARD checks before a write is acknowledged, prevents corrupt blocks from being written
    - Supports NFS, iSCSI, FCP
  - Or automate block level validation of backups
    - e.g. use NetApp SnapManager® for Oracle

Agenda

- ASM Overview
- Considerations and Recommendations
  - Data Resilience
  - Performance
  - Storage Utilization
  - Data Protection
- Summary

Performance Considerations

- ASM striping versus storage striping
- Utilizing spindle I/Os
- Host based I/O balancing
- Prioritization of I/O
- Storage network protocols and throughput

Physical Distribution of I/O Load - Utilizing all available Disk Spindles

- **Disks spindles have a physical I/O limit**
- **Higher I/Os can be sustained by spreading the data across multiple spindles**
- **Solutions**
  - ASM Striping
  - LVM Striping
  - Storage Striping
- **Choice**
  - LVM and ASM are both host based striping, pick only one
  - ASM Striping and Storage Striping are complementary
    - ASM balances I/O on the host
    - Storage Striping balances I/O to physical disks

ASM Striping Versus Storage Striping

ASM striping

- **Pros**
  - Distributes across LUNs
- **Cons**
  - Not well balanced across ALL physicals disks
  - Requires more LUNs on the host to effectively balance

Storage striping

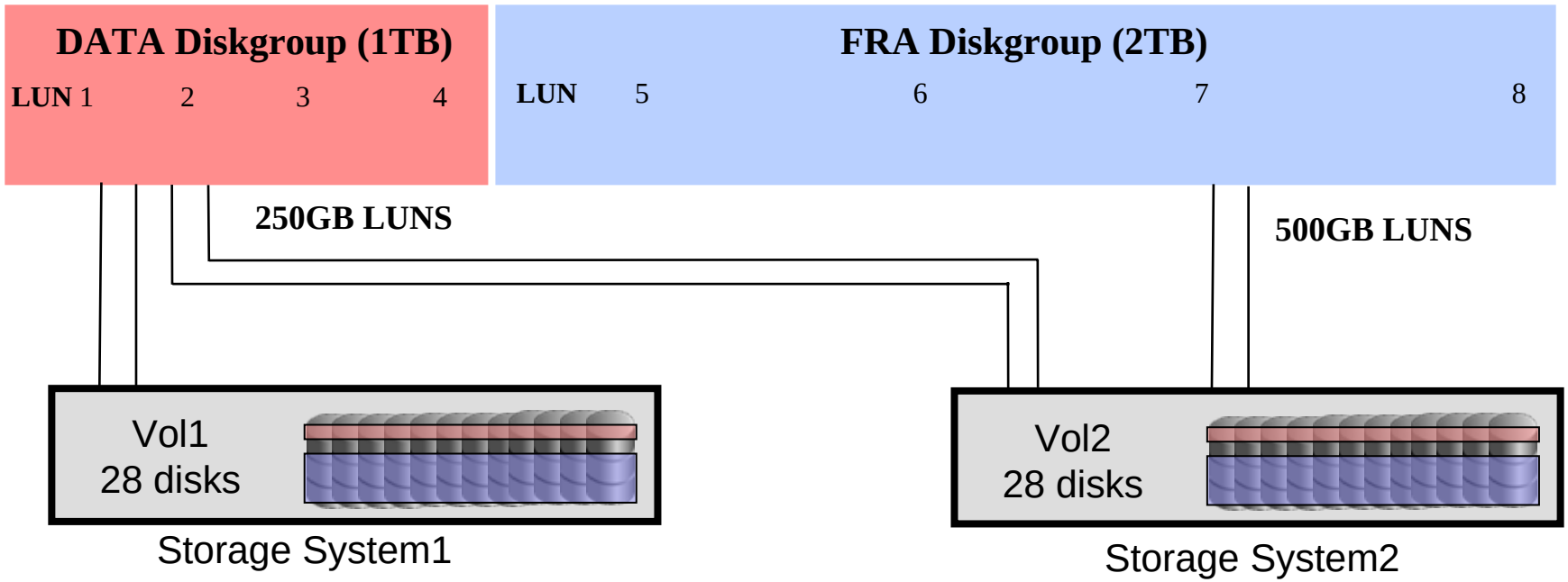
- **Pros**
  - Distributes across ALL physical disks
  - Fewer LUNs to manage to maximize spindles
- **Cons**
  - Large incremental growth (e.g. whole disk)

Managing striping at multiple levels is hard, and as a DBA you must be very aware of how the storage system is laid out.

Host Based Distribution of I/O Load - Using ASM to Balance the I/O Load

- **ASM can help spread the I/O load**
  - Balances I/O across Disks in a Disk Group
  - Eliminates “hot spots” by dynamic relocation of data from Disk to Disk
- **ASM rebalances when**
  - Disk is added or dropped from a Disk Group
  - `alter diskgroup <diskgroup> rebalance`
- **ASM Disk is both a unit of space and unit of I/O**
  - ASM does not know how the LUN is mapped to Disks e.g.
    - LUNs 1-3 are backed by 5 disks, LUN 4 is backed by another 5 disks
    - Balancing the load across these ASM Disks does not balance load on the storage system

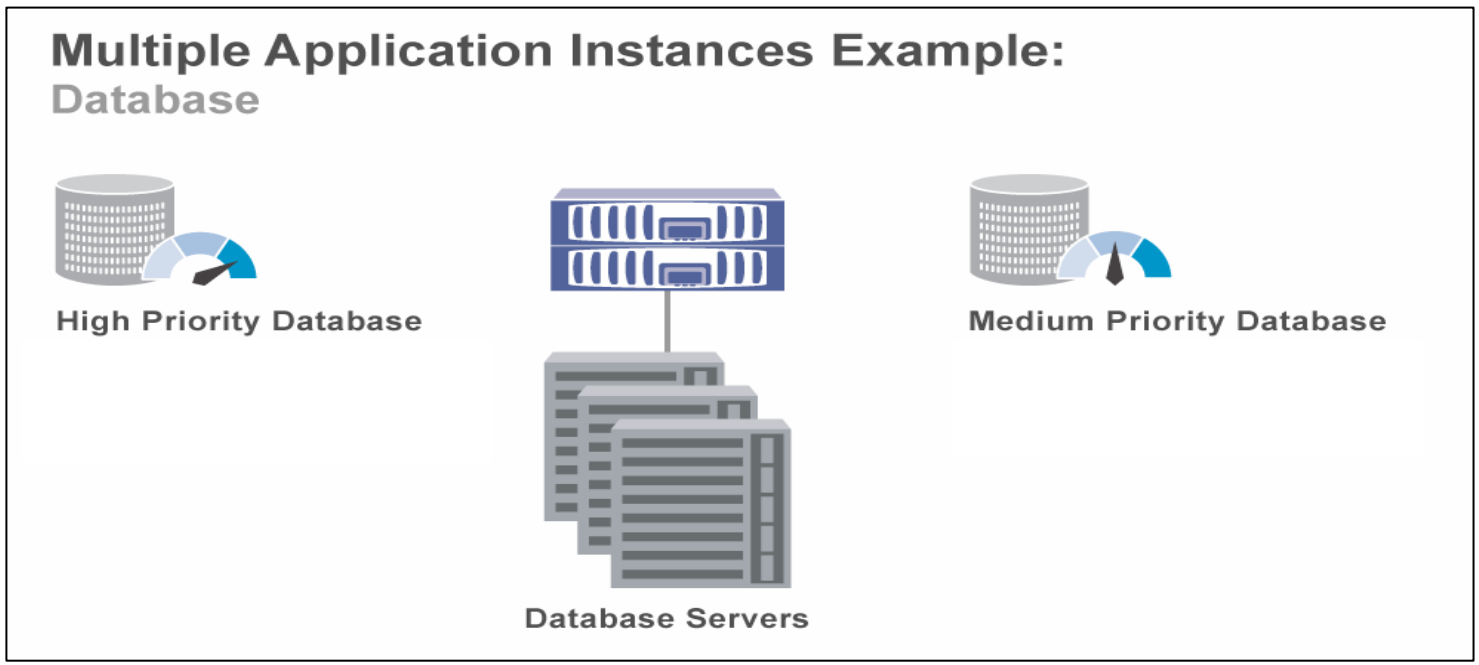
Using ASM to Balance the I/O Load



Managing load balancing at multiple levels is hard, balancing on the Host does not mean well balanced storage and vice versa

Other I/O Considerations - Not all I/O is created equal

- Storage is being consolidated, i.e. many applications can use the same networked storage
- Need to ensure that application's I/Os are prioritized appropriately
- ASM does not prioritize I/Os



Storage Networking Protocol - Performance Comparison

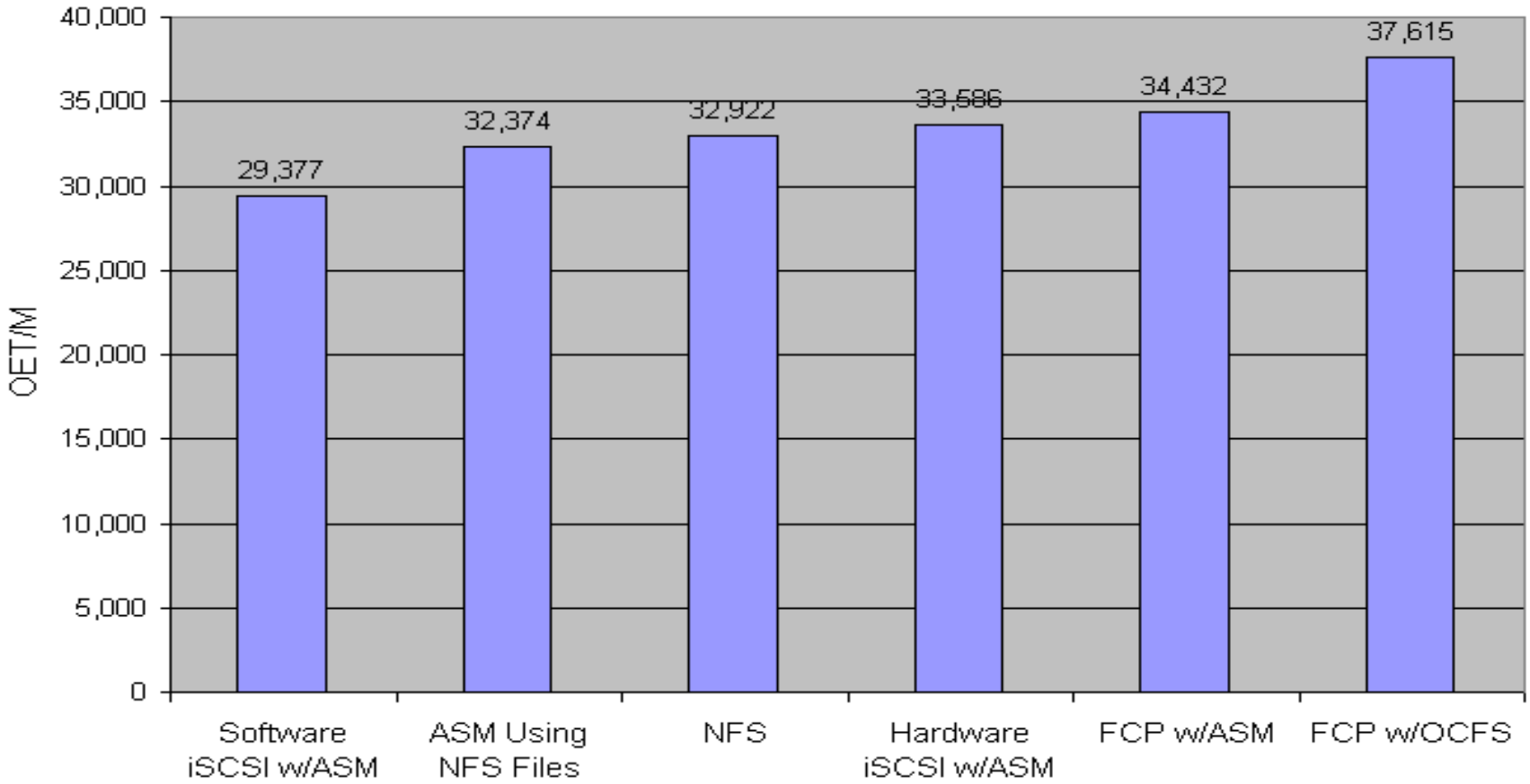
- **Oracle® 10*g*R1 RAC, 3 nodes with**
  - Oracle ASM Using Software iSCSI LUNs
  - Oracle ASM Using NFS Files
  - Oracle ASM Using Hardware iSCSI LUNs
  - Oracle ASM Using FCP LUNs
  - NFS With Directio
  - OCFS with FCP LUNs
- **3 x Dell PowerEdge 2650 Servers E/W:**
  - Dual 3.2 Ghz CPU with 4 GB RAM
  - GbE for cluster interconnect (Jumbo Frames)
  - Qlogic QLA 4010 iSCSI initiator for hardware iSCSi testing
  - Qlogic QLA 2342 FCP initiator for FCP testing
  - RHEL 3.0 Update 4
- **Ensured host are CPU bound and <u>not</u> I/O bound**

Workload Characteristics

- **OLTP type load - wholesale supplier order processing**
- **Designed to emulate the real life activities**
  - Each order can consist of multiple database transactions.
  - I/O accessed through primary and secondary keys.
  - The number of items per business transaction is 10.
  - Each item ordered requires
    - 1 row selection with data retrieval
    - 1 row selection with retrieval and update
    - 1 row insertion

Throughput Comparisons

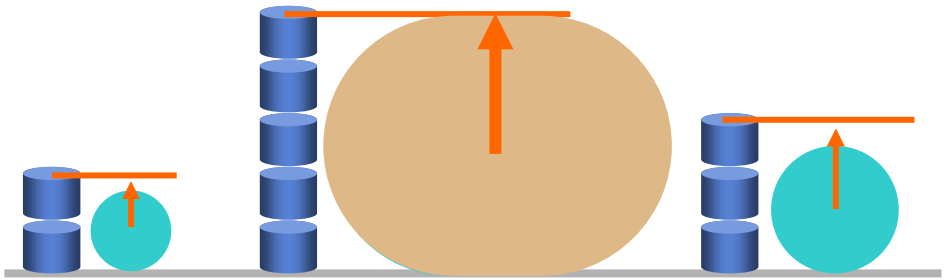
Required throughput will help you determine the protocol and cost

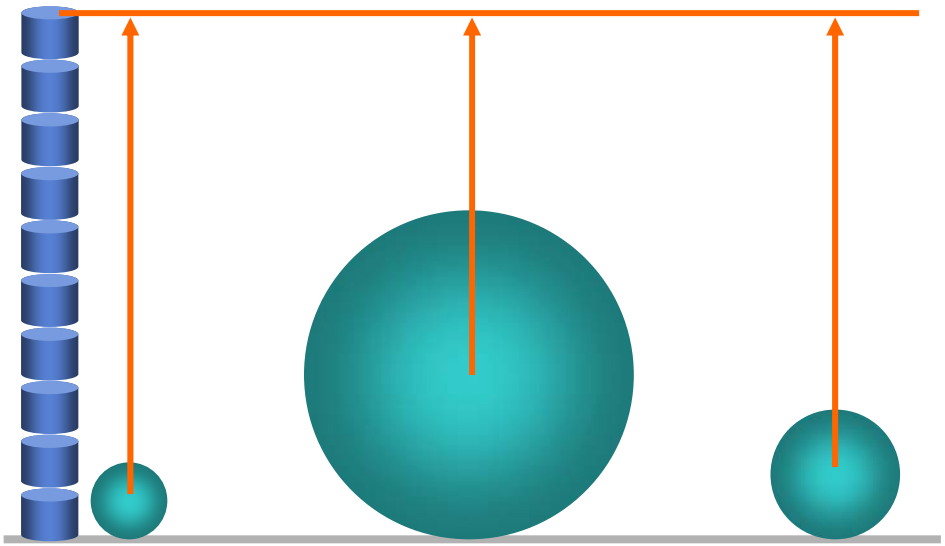


Performance - NetApp Recommendations

- **Stripe your data**
  - Combine ASM and FlexVol™ volumes for storage striping
  - Data ONTAP® will load balance all LUNs across a very large storage pool, so no conflict with ASM Balancing
  - Ensure that each LUN is backed by the same number and type of physical disks when using ASM Balancing
- **Configure disks for I/O requirements not just space**
  - Use FlexVol to maximize the I/O available from all spindles
- **Ensure that Application I/O is prioritized**
  - Use FlexShare™ to ensure control of service between Volumes on shared storage server
- **Choose your storage protocol**
  - Balance between cost/complexity and performance
  - NetApp gives flexibility
    - All storage protocols are supported in a single box
    - Can expose ASM Disks over different storage protocols without data migration
  - Knowing your throughput needs will help you choose

Performance NetApp Recommendations - Using all physical spindles with FlexVol™





Regular Volumes:

- Volume performance limited by number of disks spindles it has
- “Hot” volumes can’t be helped by disks on other volumes
- RAID reconstruction required if disks are added to existing volume

FlexVol:

- All spindles are available to all volumes
- No RAID reconstruction when adding more disks

Agenda

- ASM Overview
- Considerations and Recommendations
  - Data Resilience
  - Performance
  - Storage Utilization
  - Data Protection
- Summary

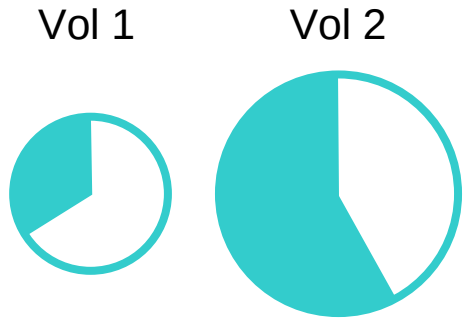
Storage Utilization Considerations

- **Free space management**
- **Storage of database clones**
- **Data tiering**

Volume Free Space Management - Pools Of Wasted Storage

How space is wasted:

- **Free space fragmented across volumes**
- **Free space not available to other volumes**
- **Over allocation wastes available disk space**



Need to consider:

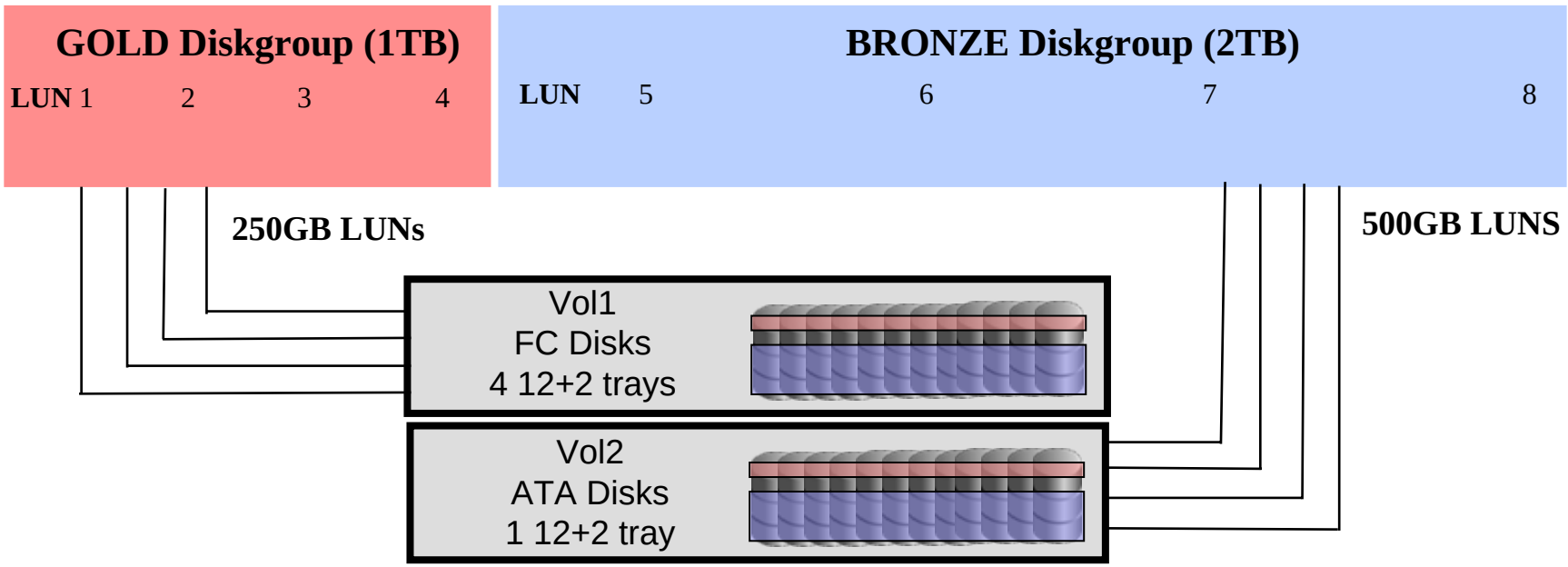
- **Careful storage planning with Storage Admin to maximize utilization for each Volume**
- **ASM Disks can be re-sized**

Challenges Cloning an ASM Database

- **Traditionally each clone is a physical copy**
  - Cloned 500GB Database requires 1TB of storage (original + clone)
- **Host based data movement is inefficient and takes time**
  - Each block needs to be read and written
  - The larger the database, the longer this takes
- **Cannot simply copy the ASM Disks**
  - ASM stamps the Disk with the Disk Group name
  - Two Disks with the same name and same Disk Group not allowed
  - May be added to alternate ASM Instance
- **Oracle® recommended practice is to use RMAN to create Database clone**
  - Takes time because it's a physical copy
  - Impacts the host and network because it's a physical copy
  - Doubles the space because it's a physical copy

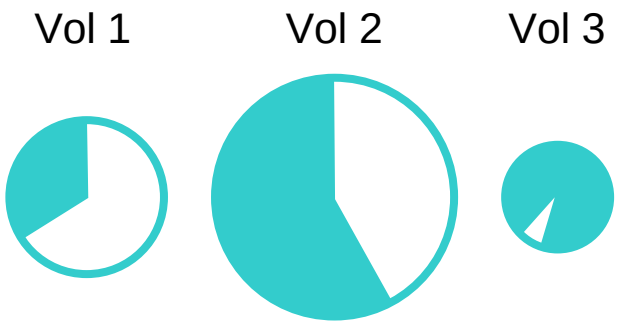
Data Tiers using ASM

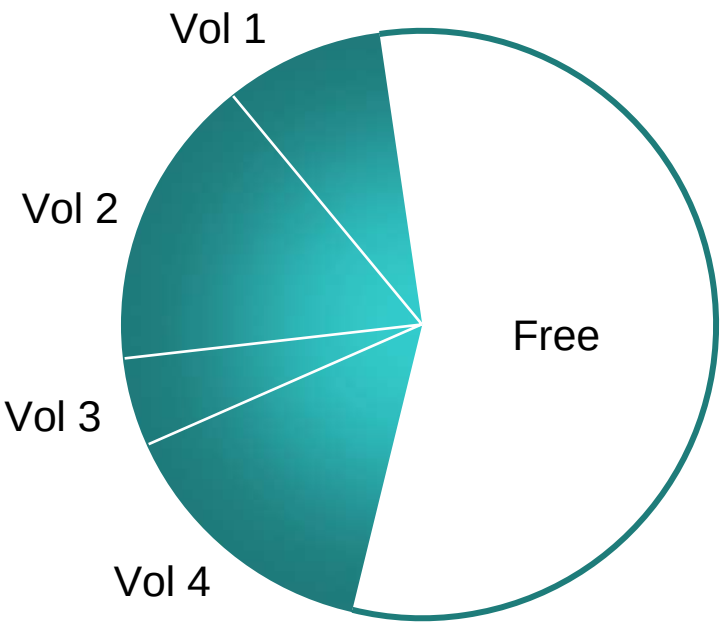
- **Create Disk Group with Disks of similar capability**
  - Disk size, speed, IOPs etc
- **Disk Group represents your Service Level Objective, e.g.**
  - FC 15K disks are “gold”
  - ATA 7.5K disks are “bronze”
- **Choose the appropriate Disk Group based on needs**
  - Log & Datafiles are on “gold”
  - Archive Logs are on “bronze”
- **Useful for ILM / compliance**



Single Storage System

Utilization Recommendations - Coalescing free space with FlexVol™





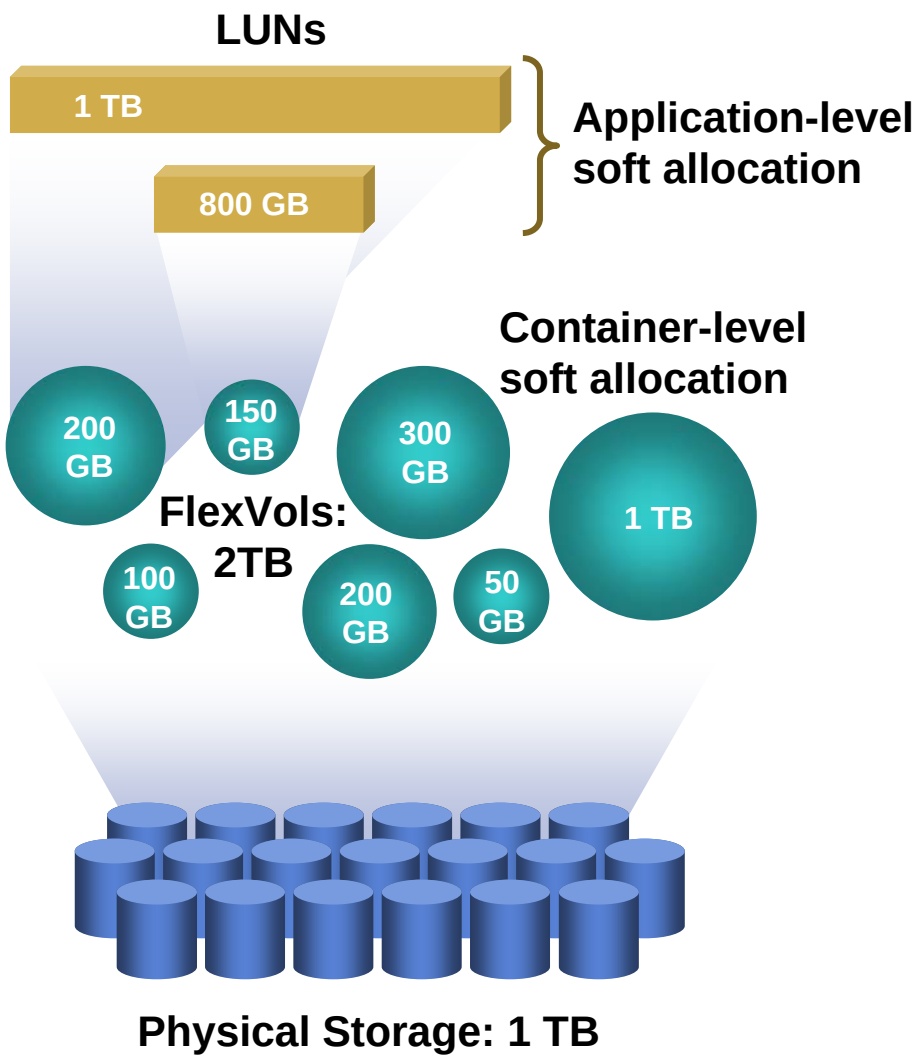
Regular Volumes:

- Free space fragmented across volumes
- Free space not available to other volumes

FlexVol:

- Disks & RAID groups owned by aggregate
- FlexVol volumes carved out of the aggregate
- No pre-allocation of free space
- Free space available for use by other or new volumes
- Can dynamically grow and shrink

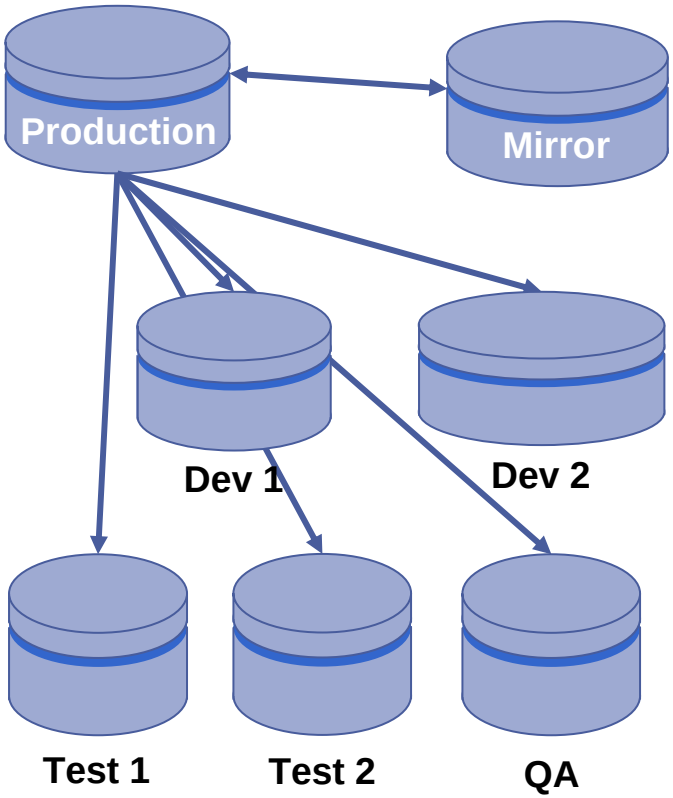
Utilization Recommendations - Thin provisioning with FlexVol™

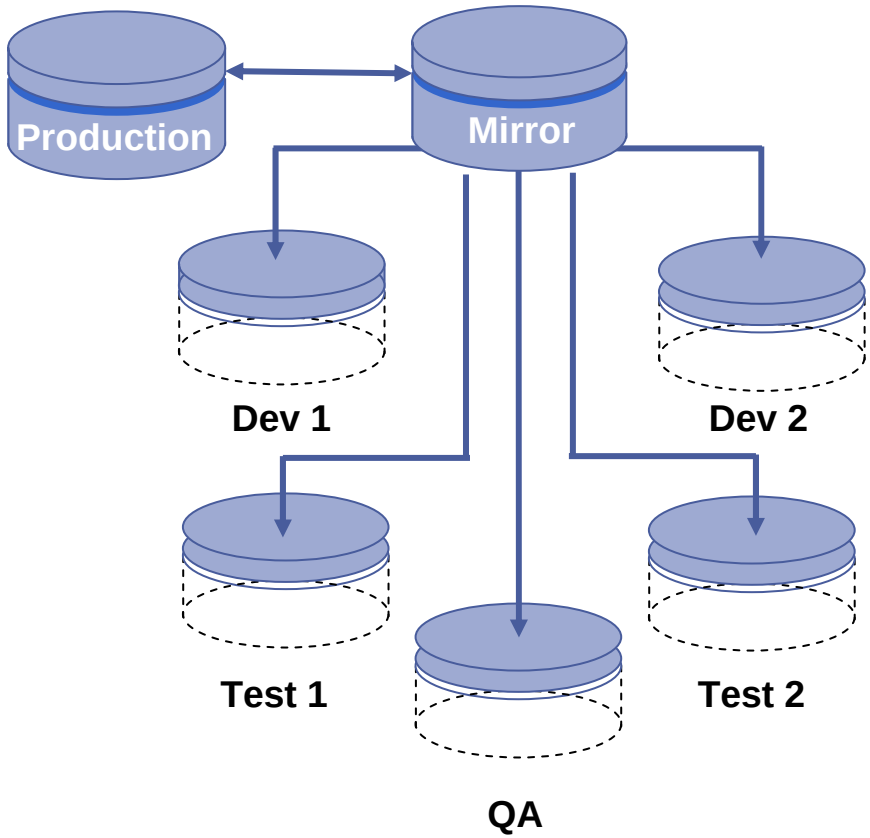


FlexVol:

- **Separates space visible to users from physical disks**
- **Increases control of space allocation**
  - Flexible provisioning
  - Better utilization
  - Higher granularity
  - Application over-allocation containment

Utilization Recommendations - Fast and space efficient ASM Clones





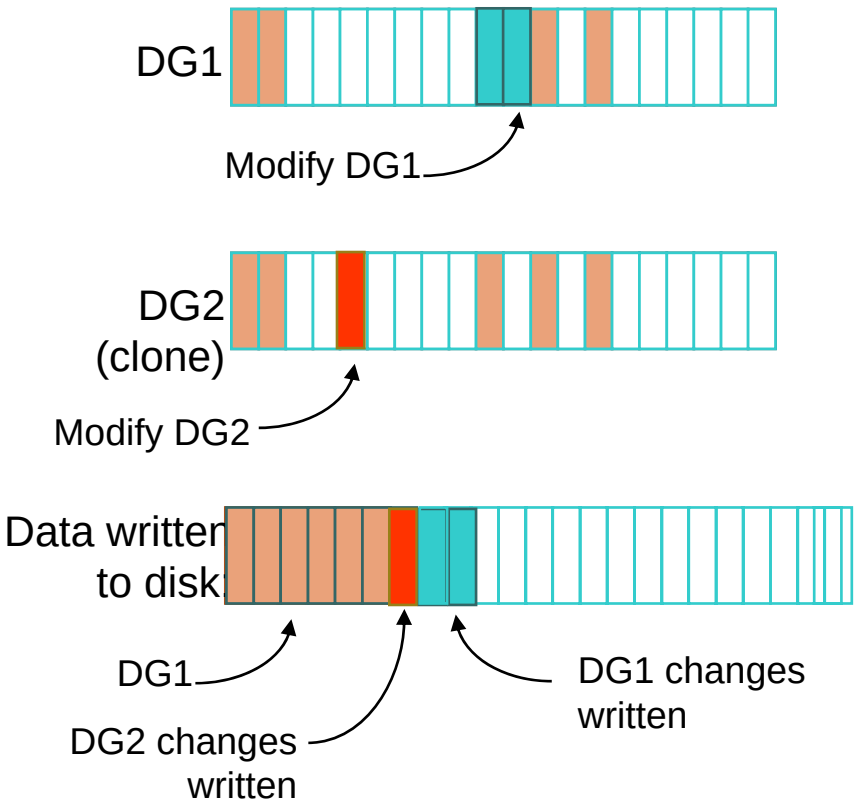
- **Physical copies consume space and time**
- **FlexClone™ copies are near-instantaneous and only store changed blocks**

Utilization Recommendations - Fast and space efficient ASM Clones

- **Create logical copy of ASM Disks using FlexClone™**
  - No physical movement of data required
    - No impact on Host CPU
    - No impact on Network
  - When the Clone is created, it takes up no additional space
    - Only store the changed blocks
    - 500GB Database clone takes only a few additional KB
- **Automate the cloning with SnapManager® for Oracle®**
  - Allows cloned ASM disks to be added back to same ASM Instance
  - Simplified management using a single ASM Instance
  - Automates all the steps to create an ASM clones

Simple, automated management, fast and very storage efficient clones of ASM backed Databases

Utilization Recommendations - How FlexClone™ makes cloning efficient



- **Start with Disk Group DG1**
- **Create a clone DG2**
- **Modify the cloned Disk Group DG2**
- **Modify the original Disk Group DG1**

Result:

- **Independent copies, efficiently stored**
- **Cost and time efficient Dev/Test, Patch testing etc.**

Agenda

- ASM Overview
- Considerations and Recommendations
  - Data Resilience
  - Performance
  - Storage Utilization
  - Data Protection
- Summary

Data Protection Considerations

- **Backup & restore windows**
- **Host versus storage based backups**
- **Granularity of restore**

Backup Considerations

- **Smaller maintenance windows**
  - Business moving towards 24x7x365 operation
  - Smaller windows for backup
- **SLAs mandate short recovery times**
  - Smaller windows to restore Database on failure
  - Require more frequent backups
- **Online versus offline**
  - Offline backups simpler but Database not available
  - Online backups have performance impact
    - 15-20% host CPU utilization impact
    - Physical I/O increase (whole log block writes)
  - As database size grows, will impact backup time

Restore Considerations

Need to control Mean Time To Recovery

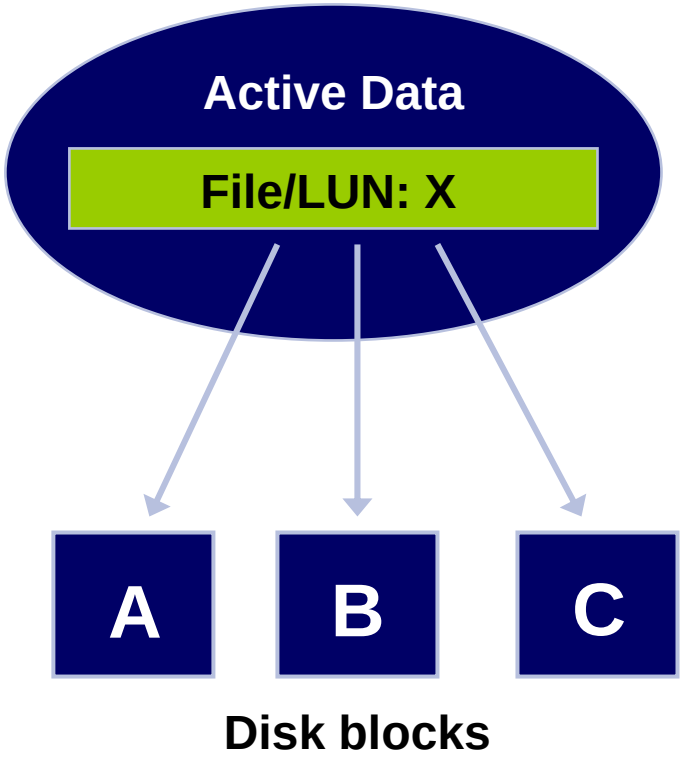
- **MTTR = restore time + recovery time**
- **Restore time depends on**
  - Amount of data to transfer
  - Time dependent on I/O rate of backup store (tape, disk)
  - Host CPU and network contention
- **Recovery time depends on**
  - Amount of logs that need to be applied
  - Directly related to transaction rate and time between backups

Host Versus Storage Based Solutions

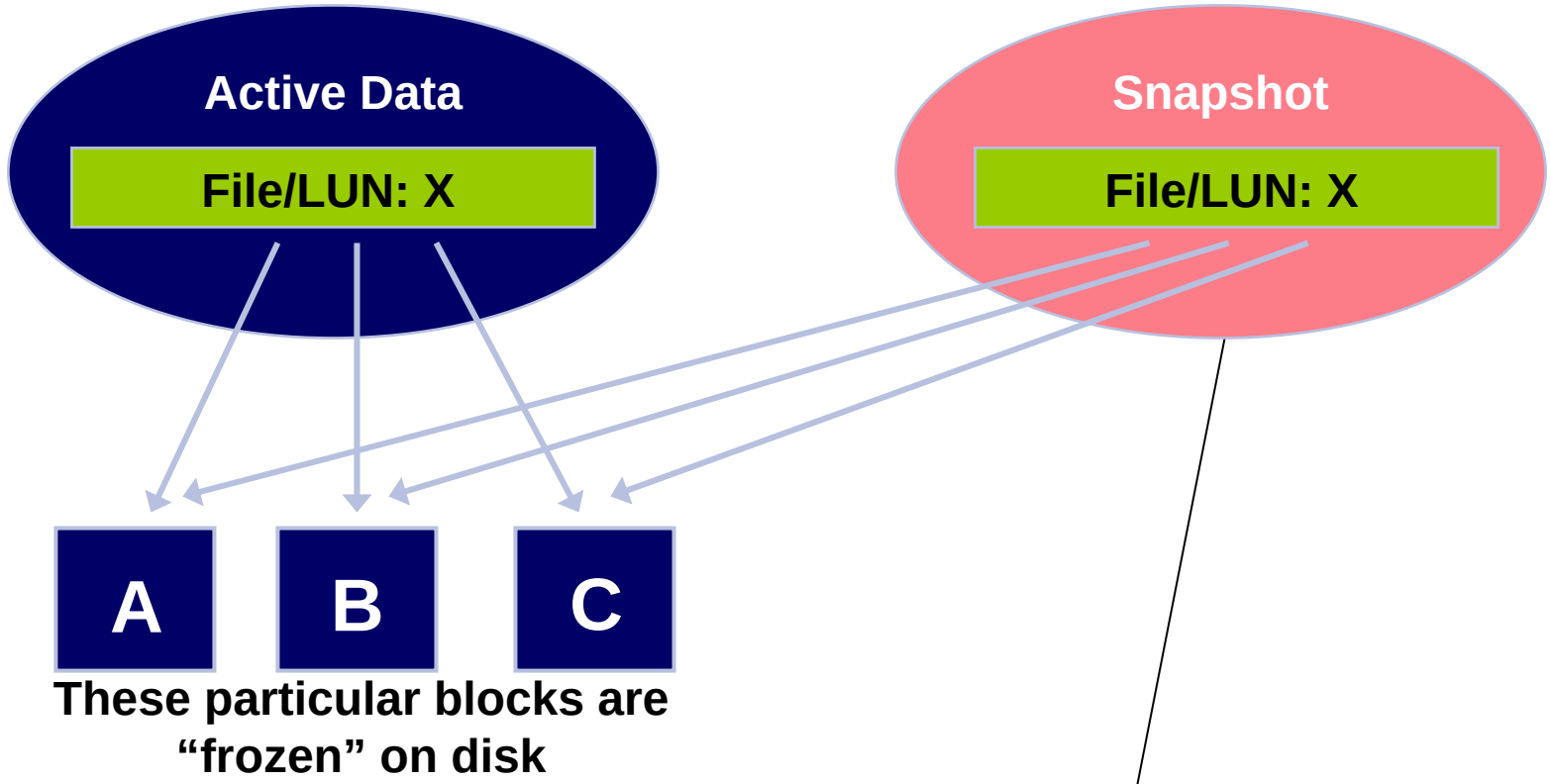
- **Host based backups (RMAN, NetBackup etc.)**
  - Read all required blocks from storage
  - Write all block to secondary store (e.g. tape)
  - Tool may have“incremental” feature to reduce the physical I/O
  - Independent of Database layout
- **Storage based backups**
  - Can utilize Snapshot™ like functionality
  - Consistent quick backup & restore time
    - Not Database size dependent
    - No physical I/O required
  - Very Space efficient
    - Stores just changed blocks
  - No host CPU impact
  - Can be sensitive to Database layout

Storage based solutions enable faster and more frequent backups, deliver shorter recovery times and enable you to meet the SLAs.

How storage Snapshot™ Copies work - Data created

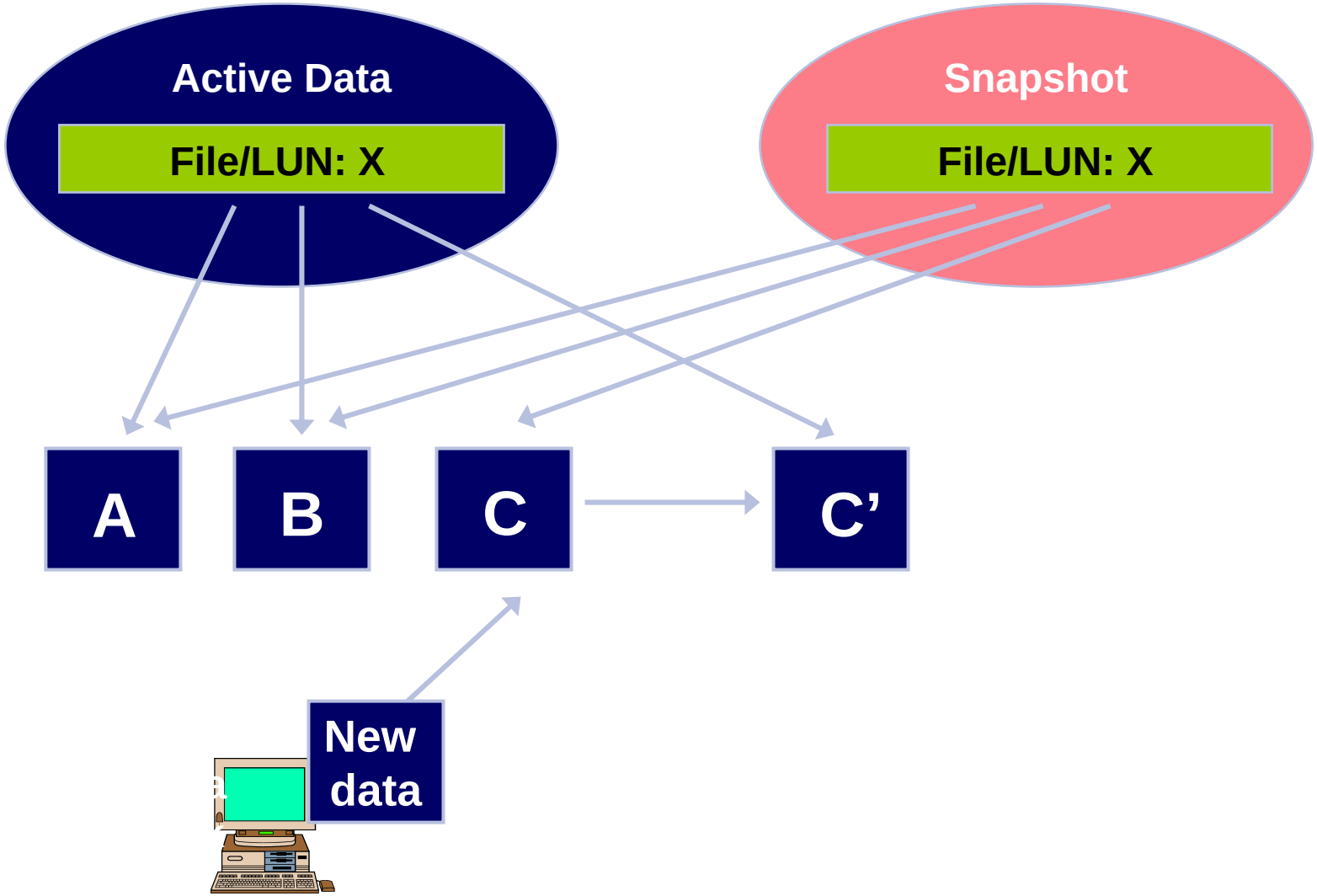


How storage Snapshot™ Copies work - Snapshot copy is created



- Consistent point-in-time copy
- Ready to use
- Consumes no space!

How storage Snapshot™ copies work - Original data changed



Backup & Recovery - ASM Challenges

- **Challenges with ASM:**
  - ASM is not storage Snapshot™ aware
  - Difficult to identify dataset for Snapshot copy as database storage layout is hidden behind ASM disk groups
- **Requirements for storage technology**
  - Consistent copy of the ASM Disks within an ASM Disk Group
    - Need an exact point in time Snapshot copy across volumes or storage servers
  - Understand the ASM file system
    - Need to ensure that data relocation has not occurred during the backup
  - Copy ASM diskgroup for backup verification or DB clones
    - Requires renaming of ASM Disk Group
  - Ability to restore a file “in-place” into a Disk Group
    - Requires understanding of ASM extent map
    - Requires sophisticated technology for partial LUN or file restoration

Maximizing Recovery efficiency

Recovery efficiency based on:

- **Granularity that storage can recover**
  - Typically LUN or File level
  - Advanced storage allows for partial LUN restore
- **What needs to be restored**
  - Whole database
  - Partial database (tablespaces or datafiles)
- **What is using the Disk Group**
  - If a single Database, whole Disk Group can be restored
  - If multiple Databases, then individual files within the Disk Group needs to be restored

Data Protection - NetApp Recommendations

- **Utilize storage based backup and restore**
  - Perform more frequent and faster backups
    - Eliminate host CPU impact of “hot backup”
    - Time & Space efficient
    - Reduction in Mean Time To Recovery (MTTR)
  - Create consistent Snapshot™ copies across Volumes and Storage Systems of ASM Disk Groups
  - Restore files in place using Partial LUN Restore
  - Automated by SnapManager® for Oracle®
- **Separate Disk Groups by Database**
  - Allows more efficient storage based restores
- **Separate Data by Disk Groups**
  - Put static and dynamic data into their own Disk Groups
  - Separate Logs and Data into their own Disk Groups

Agenda

- ASM Overview
- Considerations and Recommendations
  - Data Resilience
  - Performance
  - Storage Utilization
  - Data Protection
- Summary

It's real simple: NetApp adds tremendous value to Oracle® ASM deployments

	Oracle ASM	Oracle ASM + NetApp
Data Resilience		
Single Disk Failure	Yes	Yes
Double Disk failure	No	Yes
Passive Block corruption detection	Yes	Yes
Active Block corruption detection	Yes	Yes
Lost disk writes	No	Yes
Performance		
Strip across ASM Disks	Yes	No
Balance across ASM Disks	Yes	No
Strip across Physical Disks	No	Yes
Balance across Physical Disks	No	Yes
I/O prioritization	No	Yes
Storage Utilization		
Free space management across physical disks	No	Yes
Thin provisioning of ASM Disks	No	Yes
Space efficient Cloning	No	Yes
Data Protection		
Storage Snapshot™ copy based Backups	No	Yes
Storage Snapshot copy based Restores	No	Yes

More Information

- Come and see us at the Booth #1614 for demos of
  - SnapManager® for Oracle®
  - FlexVol™, FlexClone™ and FlexShare™
- NetApp Oracle partner page
  - http://www.netapp.com/partners/oracle/
- Best Practice Guidelines for Oracle
  - http://www.netapp.com/library/tr/3369.pdf
- SnapManager for Oracle
  - http://www.netapp.com/library/tr/3426.pdf
- Performance Report: Oracle Database 10*g* RAC:
  - http://www.netapp.com/library/tr/3423.pdf
- RAID-DP™: NetApp Implementation of RAID Double Parity for Data Protection
  - http://www.netapp.com/library/tr/3298.pdf