HOWTO: Implementing Oracle ASM Successfully

Alex Gorbachev
23 September, 2010
San Francisco, CA
Alex Gorbachev

- CTO, The Pythian Group
- Blogger
- OakTable Network member
- Oracle ACE Director
- BattleAgainstAnyGuess.com
- President, Oracle RAC SIG
Why Companies Trust Pythian

• Recognized Leader:
  • Global industry-leader in remote database administration services and consulting for Oracle, Oracle Applications, MySQL and SQL Server
  • Work with over 150 multinational companies such as Forbes.com, Fox Interactive media, and MDS Inc. to help manage their complex IT deployments

• Expertise:
  • One of the world’s largest concentrations of dedicated, full-time DBA expertise.

• Global Reach & Scalability:
  • 24/7/365 global remote support for DBA and consulting, systems administration, special projects or emergency response
Pre-ASM Oracle Storage
Pre-ASM Oracle Storage

- Too complex
- 3rd party vendors
  - Additional licensing cost
- Many groups involved
  - Storage engineers
  - SA’s
  - DBA’s
- Loosing layout visibility
Pre-ASM Oracle Storage

- Too complex
- 3rd party vendors
  - Additional licensing cost
- Many groups involved
  - Storage engineers
  - SA’s
  - DBA’s
- Loosing layout visibility

- Too simple
  - Doesn’t scale
  - Not reliable
  - Not flexible
Why ASM?

- Simplify
  - Fewer vendors involved
  - Fewer layers
  - Move many operations under DBA umbrella
    - oh well, Storage Administrator role
- Cost efficient
  - No need to license volume manager
  - Cheap storage boxes are good for ASM
  - Manageable
ASM Basic Principles

- Organizing Oracle files on physical disks
  - “Purpose-built file system”

- Mirroring
  - reliability & failure tolerance

- Striping
  - performance & scalability

- Rebalancing
  - manageability
ASM Striping

- Extents and Allocation Units
- Coarse striping & Fine striping
- Random striping -> equal distribution
- You cannot disable ASM striping
ASM Striping

- Extents and Allocation Units
- Coarse striping & Fine striping
- Random striping -> equal distribution
- You cannot disable ASM striping
ASM Striping

- Extents and Allocation Units
- Coarse striping & Fine striping
- Random striping -> equal distribution
- You cannot disable ASM striping
ASM Striping

- Extents and Allocation Units
- Coarse striping & Fine striping
- Random striping -> equal distribution
- You cannot disable ASM striping
ASM Striping

- Extents and Allocation Units
- Coarse striping & Fine striping
- Random striping -> equal distribution
- You cannot disable ASM striping
ASM Striping

- Extents and Allocation Units
- Coarse striping & Fine striping
- Random striping -> equal distribution
- You cannot disable ASM striping
ASM Striping

- Extents and Allocation Units
- Coarse striping & Fine striping
- Random striping -> equal distribution
- You cannot disable ASM striping
ASM Striping

- Extents and Allocation Units
- Coarse striping & Fine striping
- Random striping -> equal distribution
- You cannot disable ASM striping

[Diagram of file1 and file2 striped across three disks]
What’s the Best Hardware Striping for ASM?
What’s the Best Hardware Striping for ASM?

Hardware Striping

RAID
What’s the Best Hardware Striping for ASM?

- Allocate disk spindles or partitions to ASM
  - 1 physical disk = 1 LUN = 1 ASM disk
- Can mirror disks in RAID-1
  - 2 x RAID-1 physical disks = 1 LUN
- If you really have to use RAID...
  - Watch for striping conflicts
    - stripe width 640K & AU size 1MB = bad
  - Align AU with striping
    - 1 AU = stripe size
    - 1 AU = stripe width
  - Avoid RAID-5
What’s the Best Hardware Striping for ASM?

• Allocate disk spindles or partitions to ASM
  • 1 physical disk = 1 LUN = 1 ASM disk

• Can mirror disks in RAID-1
  • 2 x RAID-1 physical disks = 1 LUN

• If you really have to use RAID...
  • Watch for striping conflicts
    • stripe width 640K & AU size 1MB = bad

• Align AU with striping
  • 1 AU = stripe size
  • 1 AU = stripe width

• Avoid RAID-5
ASM Rebalancing

- `asm_power_limit` - rebalancing speed
  - Async rebalancing in 11.2.0.2
- No auto-magic re-layout based on performance
- You can force rebalancing for a DG
- Intelligent Data Placement - Hot / Cold
ASM Rebalancing

- `asm_power_limit` - rebalancing speed
  - Async rebalancing in 11.2.0.2
- No auto-magic re-layout based on performance
- You can force rebalancing for a DG
- Intelligent Data Placement - Hot / Cold
ASM Rebalancing

- `asm_power_limit` - rebalancing speed
  - Async rebalancing in 11.2.0.2
- No auto-magic re-layout based on performance
- You can force rebalancing for a DG
- Intelligent Data Placement - Hot / Cold
ASM Rebalancing

- `asm_power_limit` - rebalancing speed
  - Async rebalancing in 11.2.0.2
- No auto-magic re-layout based on performance
- You can force rebalancing for a DG
- Intelligent Data Placement - Hot / Cold
ASM Rebalancing

- `asm_power_limit` - rebalancing speed
  - Async rebalancing in 11.2.0.2
- No auto-magic re-layout based on performance
- You can force rebalancing for a DG
- Intelligent Data Placement - Hot / Cold
ASM Rebalancing

- `asm_power_limit` - rebalancing speed
  - Async rebalancing in 11.2.0.2
- No auto-magic re-layout based on performance
- You can force rebalancing for a DG
- Intelligent Data Placement - Hot / Cold
What’s the Best LUN Size for ASM?

• All disks in the same DG should be sized equally
• All disks in the same DG should perform the same
• LUN size is the increment of adding space
• Don’t make too big - harder to manage
  • Bug for 2+ TB disks
  • Size within 500GB-1TB is usually OK
• Don’t make disks too small & too many disks
• Don’t make DG more than few TB in 10g
ASM Mirroring

- Primary extent & mirror extent
- Write is done to all extent copies
- Read is done always from primary extent by default
  - Silent corruptions?
  - Use ASMCMD “remap” command
- Intelligent Data Placement for secondary extents
ASM Mirroring

- Primary extent & mirror extent
- Write is done to all extent copies
- Read is done always from primary extent by default
  - Silent corruptions?
  - Use ASMCMD “remap” command
- Intelligent Data Placement for secondary extents
ASM Mirroring

- Primary extent & mirror extent
- Write is done to all extent copies
- Read is done always from primary extent by default
  - Silent corruptions?
  - Use ASMCMD "remap" command
- Intelligent Data Placement for secondary extents
ASM Mirroring

- Primary extent & mirror extent
- Write is done to all extent copies
- Read is done always from primary extent by default
  - Silent corruptions?
  - Use ASMCMD “remap” command
- Intelligent Data Placement for secondary extents
ASM Mirroring

- Primary extent & mirror extent
- Write is done to all extent copies
- Read is done always from primary extent by default
  - Silent corruptions?
  - Use ASMCMD “remap” command
- Intelligent Data Placement for secondary extents
ASM Mirroring

- Primary extent & mirror extent
- Write is done to all extent copies
- Read is done always from primary extent by default
  - Silent corruptions?
  - Use ASMCMD "remap" command
- Intelligent Data Placement for secondary extents
ASM Mirroring

- Primary extent & mirror extent
- Write is done to all extent copies
- Read is done always from primary extent by default
  - Silent corruptions?
  - Use ASMCMD “remap” command
- Intelligent Data Placement for secondary extents
ASM Mirroring

- Primary extent & mirror extent
- Write is done to all extent copies
- Read is done always from primary extent by default
  - Silent corruptions?
  - Use ASMCMD “remap” command
- Intelligent Data Placement for secondary extents
ASM Mirroring

- Primary extent & mirror extent
- Write is done to all extent copies
- Read is done always from primary extent by default
  - Silent corruptions?
- Use ASMCMD “remap” command
- Intelligent Data Placement for secondary extents
ASM Failure Groups

- Group of disks can fail at once
- Mirror extents between failure groups
- Beware of space provisioning
- Beware of disk partnering (double disk failures)
ASM Failure Groups

- Group of disks can fail at once
- Mirror extents between *failure groups*
- Beware of space provisioning
- Beware of disk partnering (double disk failures)
ASM Failure Groups

- Group of disks can fail at once
- Mirror extents between *failure groups*
- Beware of space provisioning
- Beware of disk partnering (double disk failures)
ASM 10g issues

• ASM upgrade requires full RAC outage
• Fixed 1MB allocation unit
• Overhead with multi-TB databases
• Recovery from failure is inefficient
  • ASM Mirroring is not VLDB ready
• Extended clusters read IO is inefficient
• Limited FS functionality and tools
ASM 11gR1 - variable extent size

• 10g
  • 1 extent = 1 AU

• 11g
  • 1-20,000 extents
    1 extent = 1 AU
  • 20,001 - 40,000
    1 extent = 8 AU’s
  • 40,001 - …
    1 extent = 64 AU’s

• 1 TB = 1+ mil. extents

• 1 TB = 53,572 extents
• 64MB AU:
  • 1 TB = 16,384 extents
• 100 TB = 66,788 extents
ASM 11gR1 - recovery from failures

- Fast Mirror Resync
  - OFFLINE disks
  - disk_repair_time attribute
  - ASM extent change tracking
  - Suitable for transient failures and maintenance

- Fast Rebalancing in restricted mount
  - Normal rebalancing => many lock/unlock extent map
  - Restrict mode rebalancing => no locks
  - Restrict mounted DG => service outage
ASM 11gR1 - Preferred Mirror Read

• For extended clusters
  • Storage mirrored across 2 or 3 datacenters

• 10g
  • Read is first done on primary extent

• 11g
  • Preferred read failure groups (to local disks)
  • asm_preferred_read_failure_groups in init.ora
ASM Instance

- Special kind of Oracle instance
  - Clustered in RAC
- Manages metadata
  - ASM disks and diskgroups
  - ASM extent map
- Performs rebalancing
- Must be started before database instance
ASM Software Stack
ASM Software Stack

Multipathing

Operating System
ASM Software Stack

Grid Infrastructure

Multipathing

Operating System
ASM Software Stack

ASM

Grid Infrastructure

Multipathing

Operating System
ASM Software Stack

Database

ASM

Grid Infrastructure

Multipathing

Operating System
ASM Software Stack

- Database
- Grid Infrastructure
- Multipathing
- Operating System
- ASM
- ASM Kernel Modules (ACFS)
ASMLib?
ASMLib?
ASMLib?

- Role of ASMLib
  - supposedly simplified discovery
    - covered by UDEV / dev mapper / MP
  - reduced number of open descriptors

- Originally designed as API for storage vendors
  - ASMLib on Linux - reference implementation

- ASMLib is useless in vast majority of deployments
ASM Layers

Diagram from Oracle whitepaper
ASM Overhead?

- ASM is **not** in the IO code path for Oracle Database
- ASM only operates extent maps and disk headers
  - Database -> thousands IO’s per second vs ASM -> few IO’s per minute
- Additional ASM instance - memory and processes but not much
- Database instance keeps ASM extent maps in SGA
- Mirroring generates additional write IO
- ADVM - another layer
- ACFS - yet another layer
ASM Licensing

- Prior to 11.2.0.2 - ASM is “free”

- All ACFS features are part of “Oracle Cluster File System - Cloud Edition” as of 11.2.0.2
  - Restricted use license for Oracle database files included with all database editions (SE/SE1/EE)
  - I.e. all ASM features for the database are still free
    - External-tables, DataPump exp/imp dumps, flat files for SQL Loader ?????
  - Pricing is still not published
How Many Diskgroups Do You Need?

• It depends…

• Use SAME (Stripe And Mirror Everything)
• Start with 2 - DATA + FRA for all databases
• Make conscious decision to split into more diskgroups
  • need to be very convincing
  • redo logs are often a candidate
  • tiered storage is an option
  • per database split (cold failover, database moves)
DATA and FRA on the Shared Physical Disks?

• YES if following SAME principles (maximize perf/util)
  • DATA on the outer tracks of disk and FRA on the inner tracks
  • Exadata is a typical example

• NO if following resiliency & stability principles
  • Performance degradation
    • Exadata has IORM that prioritizes backup IO lower
  • Disk failure makes both database and backup corrupted
ASM Implementation Checklist

- Plan space and **performance** capacity (IOPS & MB/s)
  - AWR or Statspack
  - Write IO vs read IO
  - Redo, tempfiles, datafiles, etc
- Create provisioning standards (DG naming, disk sizing)
- Performance of each LUN and total (ORION is great)
- Use OMF -- "*We have standards*" doesn’t cut it!
- Don’t screw up multipathing
- Very very careful with thin provisioning
- Master RMAN
- Use latest possible version
ASM Implementation Checklist

- Plan space and performance (IOPS & MB/s)
- AWR or Statspack
- Write IO vs read IO
- Redo, tempfiles, datafiles, etc
- Create provisioning standards (DG naming, disk sizing)
- Performance of each LUN and total (ORION is great)
- Use OMF -- “We have standards” doesn’t cut it
- Don’t screw up multipathing
- Very very careful with thin provisioning
- Master RMAN
- Use latest possible version
Operational Support Checklist

- Watch space usage
- Watch performance of each LUN
- Watch failed disks
- Monitor ASM instance alert.log
- Use REMAP regularly to check secondary extents
- Consider a separate role for ASM management
- Break the silos - bring ASM management to DBA team
- Avoid rebalancing in critical hours
- Every DBA and storage admin should practice failure handling
What’s next?

• Book draw
  • Leave me your cards or email+phone

Q & A

Email me - gorbachev@pythian.com
Read my blog - http://www.pythian.com
Follow me on Twitter - @AlexGorbachev
Join Pythian fan club on Facebook & LinkedIn