COLLABORATE15
TECHNOLOGY AND APPLICATIONS FORUM FOR THE ORACLE COMMUNITY

DBA 3.0 - How to Become a Real-World Exadata DBA

Session ID#: 775

What Your DBA Team Needs to Know to Support Exadata.

Prepared by:
Mark Smith
Senior Staff Consultant
Database Specialists
@hickydoo

REMINDER

Check in on the COLLABORATE mobile app

Who Am I?

- 14 years of DBA experience:
  - From E-Business Suite to Exadata with OLTP and EDW.
  - Exadata DBA / DMA since early 2011.

- Senior Staff Consultant at Database Specialists:
  - Consultancy providing expert Oracle DBA services.
  - Based in San Francisco with DBAs across the United States.
    - Clients from ecommerce, retail, marketing, health care, financial, news media, telecoms, film and television, automotive, aviation, etc.
    - Web site: www.dbspecialists.com

Why Am I Here?

- Exadata is a significant investment:
  - Big promises made, big checks written.
- **Specialized** expertise required to support Exadata **properly**.
- Provisions often not made for support teams:
  - Training, organizational structure, new operating procedures.
- Managing Exadata like a “normal” Oracle database risks membership of the “3x Club”:
  - Performance improvement of **2-3x** instead of **10-15x**.
  - Support headaches, unhappy managers.

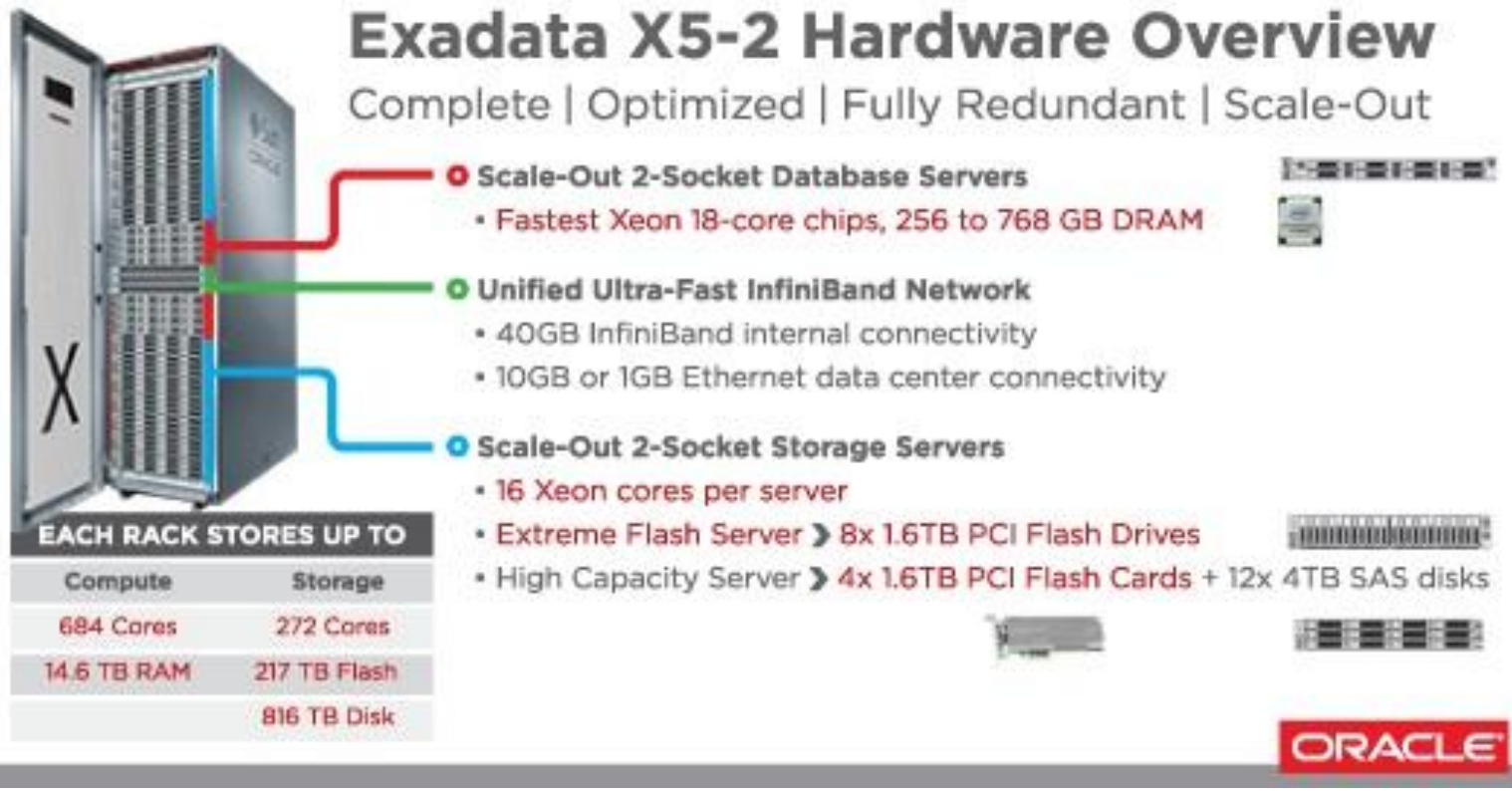
My Exadata Experience

- Exadata V1: HP / Oracle (2010)
  - Poor end-to-end support.
  - Unreliable hardware, no Flash Cache.
- Exadata V2: Sun / Oracle (2011)
  - Significant performance / reliability improvements.
  - Storage cells had firmware / Flash card issues.
- Exadata X3: Oracle (2013)
  - More of everything, very robust, had become “mainstream”.
  - Networking hardening / Platinum Support were problems.

Why is Exadata Different?

- Exadata
  - Integrated hardware / software.
  - SmartScans retrieve only data required for query.
  - EHCC saves storage, reduces I/O, increases performance.
  - Smart Flash Cache intelligently manages large amounts of Flash.
  - InfiniBand networking at 40GiB/s to other engineered systems.
- Non-Exadata
  - Different database / OS / network / storage vendors
  - All data pulled into memory for processing.
  - Row-based compression saves less storage, requires decompression.
  - Smart Flash Cache available with manual operation / configuration.
  - InfiniBand available with manual configuration.

The Exadata X5-2 Machine



Components

- Compute Nodes:
  - Database instance, ASM instance, clusterware.
- Storage Cells:
  - Physical disk for ASM storage and cell filesystem.
  - Flash cards for Exadata Smart Flash Cache.
- IB Switches
  - Manages intra-machine InfiniBand network.
- ILOMs:
  - GUI to manage physical components.

Management Tools

- Standard Oracle management tools:
  - SQLPLUS, ASMCMD, SRVCTL, CRSCTL
- CELLCLI for storage cells.
- DCLI to issue Linux commands to multiple servers.
- IPMITOOL for the ILOMs
- InfiniBand commands for the switches.
  - ibqueryerrors, ibstatus, ibchecklink

DBA? DBMA? DMA? Other?

- DBA 1.0:
  - Single-instance, production databases.
- DBA 2.0:
  - Clustered production databases with standby databases.
  - RAC, ASM, Data Guard.
- DBA 3.0:
  - Exadata and other engineered systems.
  - SmartScan, EHCC, Flash Cache.
  - O/S, storage, network administration skills now required.
  - New title: Database Machine Administrator (DBMA / DMA).

Installing Exadata

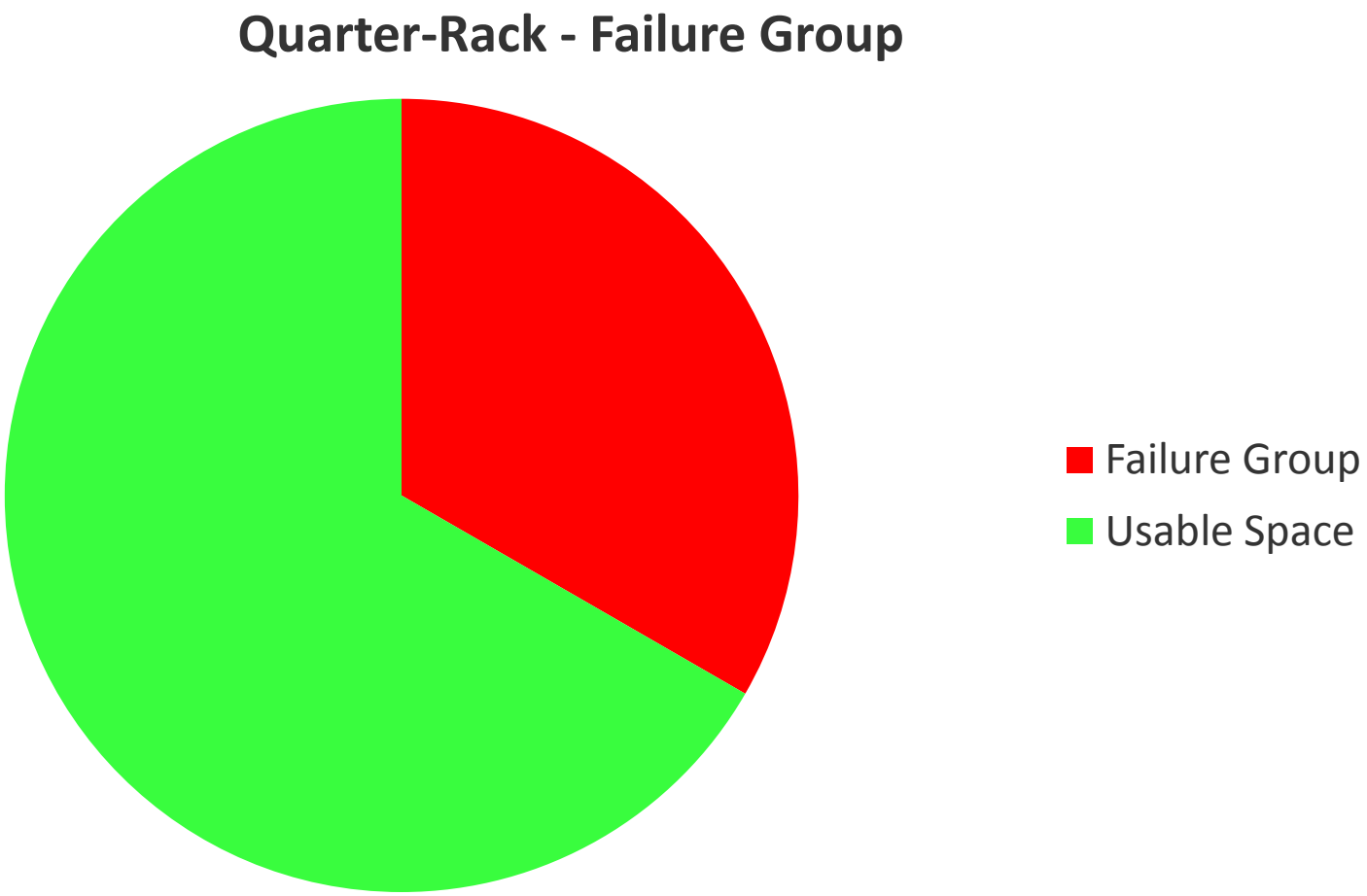
- Oracle Field Services on-site for power-up.
- Oracle ACS on-site for installation
  - Customer completes Oracle Exadata Deployment Assistant
    - **Raise non-default requests AT THIS POINT!**
    - Generates “test” and “deploy” scripts.
    - “Test” scripts must pass before on-site is scheduled.
  - “Deploy” scripts run while ACS on-site.
    - Configures networking, storage, database software
    - **Remember to check for access from other servers, VLANs.**

Redundancy and ASM

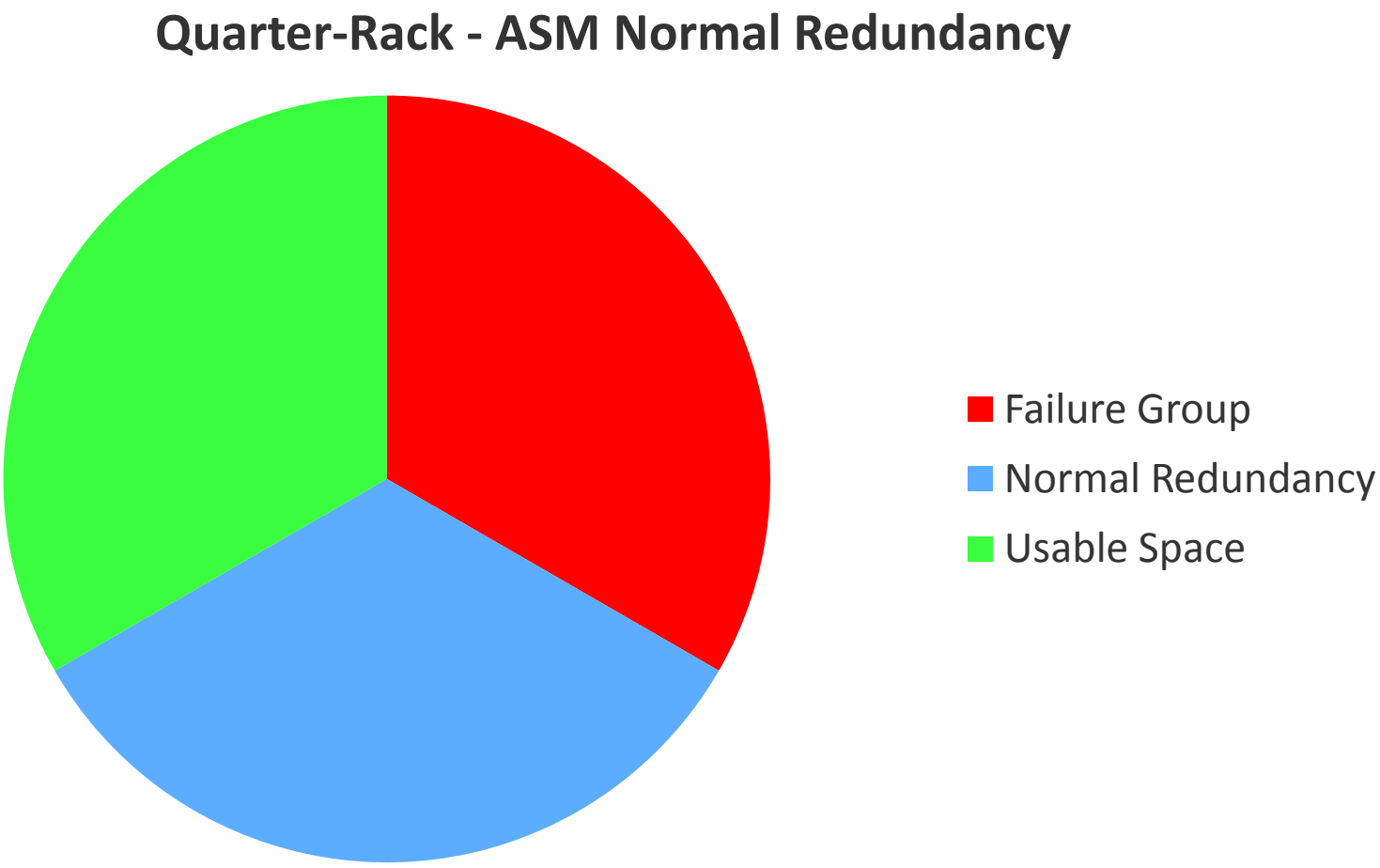
- ASM diskgroup split default is 80:20 (DATA / RECO):
  - 20% of Exadata storage is expen$ive!
  - DBFS diskgroup created from “spare” storage:
    - **Separate Flashback (DBFS) and archive logs (RECO)**

- Measure USABLE_FILE_MB, **not** FREE_MB:
  - ASM redundancy gives 2 or 3 copies of data (normal / high).
    - Tolerates complete loss of 1 or 2 storage cells.
  - USABLE_FILE_MB calculated based on Exadata’s redundancy requirements.
  - With NORMAL redundancy in 1/4-rack, **33% of storage usable.**

Usable Storage – Quarter-Rack



Usable Storage – Quarter-Rack



Disk Failures

- When a cell disk fails:
  - ASM drops disk if offline for > DISK_REPAIR_TIME:
  - ASM mirrors the data on failed disk to other disks:
  - **Keep as many spare disks as possible on-site.**
  - Monitor disk errors with CELLCLI and exachk:
    - Persuade Oracle Support of impending doom.
- Exadata MUST be able to satisfy its redundancy:
  - With NORMAL ASM redundancy, second disk failure before original resync operation completes = **bad day … and night?**

Backups

- Database:
  - RMAN backups should be cell-optimized.
  - **Backups using InfiniBand network are much faster!**
  - ZFS Storage Appliance / Zero Data Loss Recovery Appliance.

- Compute Node:
  - Backup your scripts!
  - Filesystem backups
  - **LVM backups via dbnodeupdate.sh**
  - OCR, voting disk via crsctl
  - Baremetal restore via USB drive contains current ESS / OS

Backups

- Storage cells
  - Backup your scripts (CELLCLI, etc)!
  - Baremetal restore via USB drive contains current ESS / OS.
    - Check ASM alert.logs for signs of failures in other cells.
- IB switches:
  - Extract XML file from ILOM.

Patching and Upgrading

- MUCH more complicated than for just the database:
  - O/S, ESS, firmware, clusterware, Grid Infrastructure, RDBMS.
  - “Bundle Patches” contains software and management tools.
- Maintaining support:
  - Database: some versions were desupported on Exadata “early”.
  - Platinum Support: MUST run **latest-1 Critical Patch Update**.
- Bug fixes:
  - ShellShock, Ghost, etc.
  - Exadata Critical issues.

Why Do I Need Data Guard?

- Large, important databases on Exadata:
  - Even with ZFS / ZLDR appliances, **restores take hours.**
- **Restore to non-Exadata environment can be difficult**:
  - Need to uncompress EHCC data.
  - No storage cells / SmartScans.
- Use standby databases to minimize patch / upgrade outages.
  - (Almost) no need for rolling upgrades - just switchover.
- Consider Active Data Guard

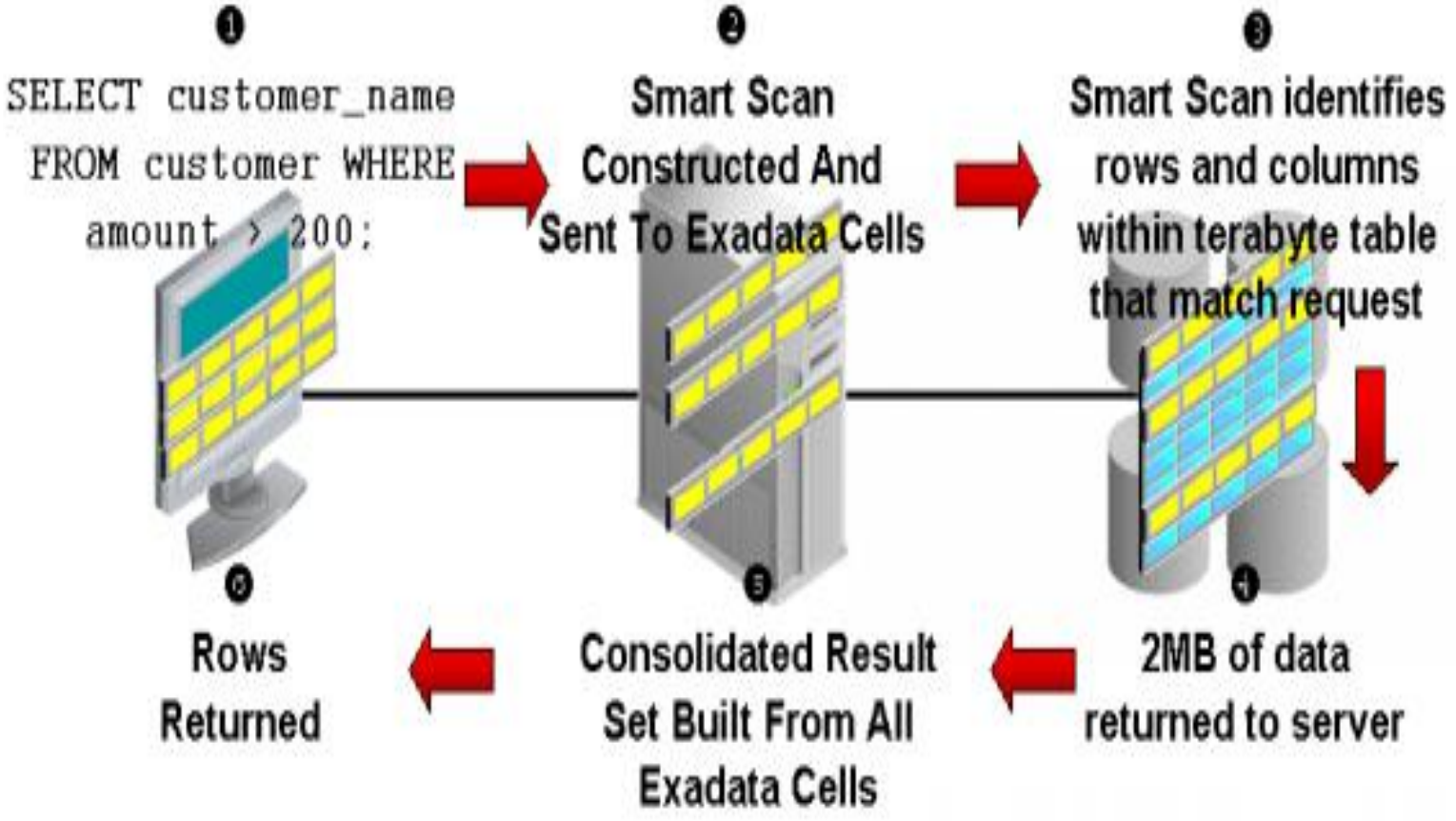
Object, Dictionary and System Statistics

- Global statistics are not gathered automatically on partitioned tables:
  - **Incremental global statistics gathering** needs to be enabled.
  - Alternatively, check for “top” tables with “most stale” global statistics weekly.
- Data dictionary statistics should be updated weekly / monthly:
- It is **VERY IMPORTANT** that system statistics are up-to-date:
  - How will the database know it’s on Exadata without them?
  - Before August 2012, there was no “EXADATA” option.
  - *EXEC DBMS_STATS.GATHER_SYSTEM_STATS(‘EXADATA)’;*

SmartScans (Storage Cell Offloading)

- Most important feature of Exadata.
- Eligible queries are offloaded to the storage cell for processing:
  - Storage cells have their own CPU / RAM / parallelism.
- Storage indexes filter out the data **not required** for the query.
- Returns **significantly** less data back to the compute nodes.
- Frees up instance resources.

SmartScan Example



SmartScan Eligiblity

- The query must use **full-object scans** (table, indexes).
- The query must bypass the SGA and use direct reads into the PGA.
- The query must - ALMOST always – use parallelism.
  - Use _SERIAL_DIRECT_READ to force serial direct reads.
- Query eligibility does not **guarantee** SmartScans.
- Check _SMALL_TABLE_THRESHOLD to see if your buffer cache is too large.

SmartScans and Indexes

- Most likely reason for SmartScan ineligibility is **index usage.**
  - Test whether a query performs better with a FULL hint.
- Important to review **non-unique** indexes for usage.
  - Enable MONITORING and review usage after full batch cycle.
  - If not used, make index INVISIBLE.
  - Save DDL and drop the index.
  - Dropping unused indexes = performance improvement for DML operations.
- V$OBJECT_USAGE view may only show **USER_INDEXES**, not ALL_INDEXES.

Am I Getting SmartScans?

- Explain plan will contain STORAGE FULL operations:
  - *TABLE ACCESS STORAGE FULL employees.*
- … **AND** …
- SmartScan waits will use CELL SMART event names:
  - *cell **smart** table scan* - for full table scans.
  - *cell **smart** index scan* - for full index scans.
- Find metrics for SmartScan usage:
  - GV$ views, AWR reports

Am I Getting a LOT of SmartScans?

- Best single formula is probably “SmartScan Efficiency Ratio”

- A) "*cell physical IO bytes eligible for predicate offload*"
  - data eligible for SmartScans - e.g. 5Tb.
- B) "*cell physical IO interconnect bytes returned by smart scan*"
  - data returned to comp nodes after SmartScans - e.g. 50Mb.

- Total I/O avoided by using SmartScans:
  - A - B = 4.95Tb
- SmartScan ratio:
  - 100 - ((100 / A) * B) = 99%

SmartScans and Parallelism

- Compute nodes have relatively low CPU_COUNT:
  - "Traditional" DW environments can have high CPU_COUNT:
    - Make extensive use of parallelism for analytical queries, etc.

- Check DEGREE setting for tables and indexes:
  - **DO NOT** use DEFAULT.
  - Ensure that large objects have at least DEGREE = 2.
  - Better to use **parallel hints** at the query-level than object-level parallelism.

- **Object inherits an DDL operation's degree of parallelism**:
  - *ALTER INDEX ... REBUILD PARALLEL 20;*

Resource Management

- Implement Database Resource Manager (DBRM) and separate users into Consumer Groups:
  - Restricts resource consumption of users / groups.
    - Prevents “power users” running multiple concurrent queries with excessive parallelism.
  - Limit each consumer group by active sessions and parallel slaves:
    - Max active sessions (4) * max degree of parallelism (4)

    = 16 parallel slaves per instance.

- Use I/O Resource Manager for inter-database management:
  - Multiple databases on same machine.

Exadata Hybrid Columnar Compression (EHCC)

- EHCC is Exadata-specific and can realize both significant savings AND performance benefits:
  - EHCC can be applied to tables and partitions.
- Compression Units (CU) stores groups of rows in a table in columnar format.
  - Values of each column stored and compressed together.
  - More data compressed at once = larger CU = better compression.
- Use EHCC on table / partitions unlikely to change:
  - DML operations will reduce the compression ratio.

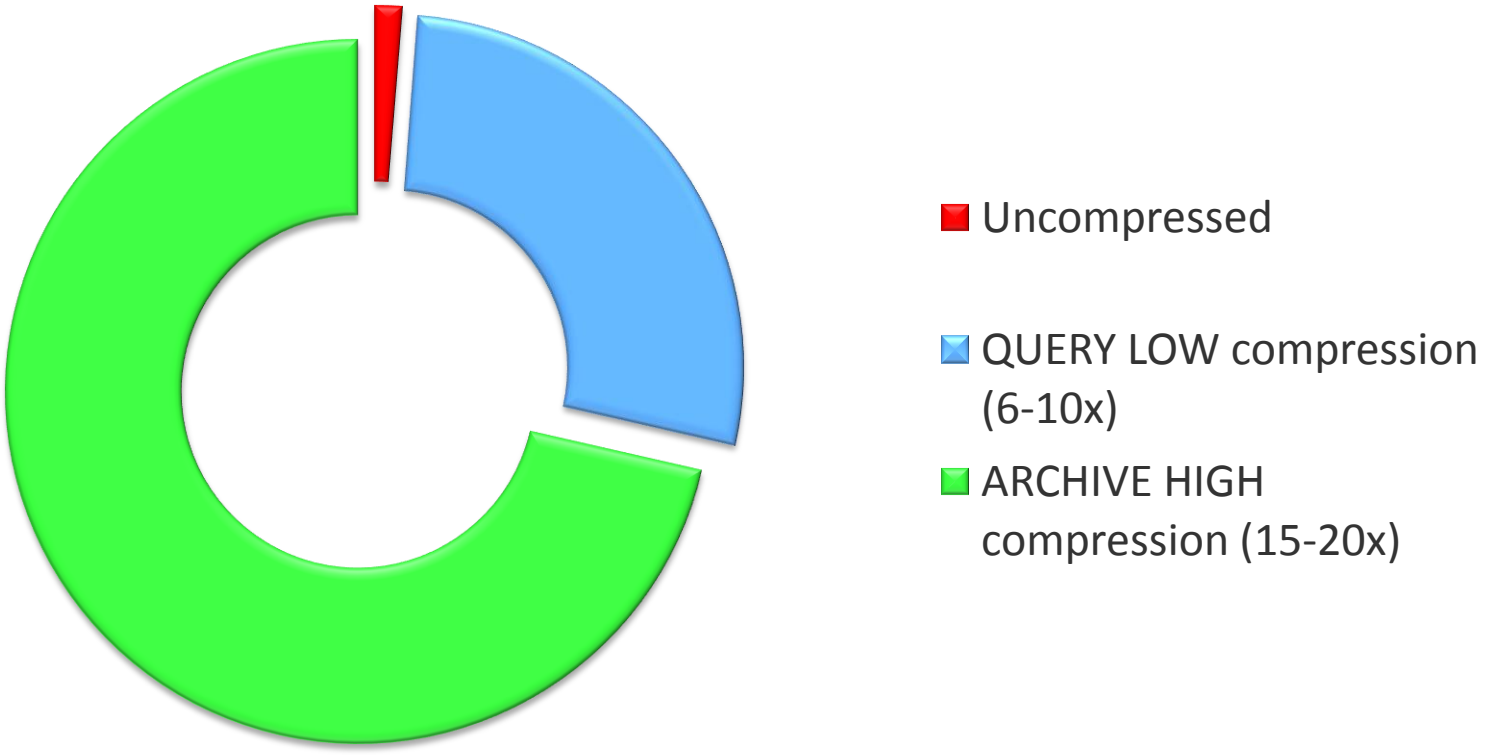
EHCC Compression Types

- 2 compression types (with 2 options each):
  - Warehouse compression – for query performance:
    - QUERY LOW and QUERY HIGH.
    - 6-10x compression ratio.
  - Archival compression – for maximum compression:
    - ARCHIVE LOG and ARCHIVE HIGH.
    - 15x compression ratio (and up).
- Use the Compression Advisory to predict compression ratios before deployment
  - MOS 762974.1.

EHCC Use Case

- Use case:
  - Very large partitioned table (6Tb) with minimal DML.
  - Table has 7 years' of data partitioned by day.
  - Most queries pull back – at most - last 31 days of data.
  - Partitions compressed based on age:
    - Newest 31 days – **not compressed.**
    - Day 32 to Year 2 – **QUERY LOW compression.**
    - Years 2 to 7 – **ARCHIVE HIGH compression.**
  - Deployed EHCC using MOVE on tables / indexes partitions.
  - Space saving around 4Tb with performance **benefit.**
  - **Bitmap indexes** caused problems:
    - Hit bug where index multiplied in size by degree of parallelism!

EHCC Use Case



Am I Using EHCC?

- '*%cell CU%*' statistics show EHCC usage.

- 2 statistics determine rows / compression unit:
  - *EHCC Total Rows for Decompression* – e.g. 5,000,000,000
  - *EHCC CUs Decompressed* – e.g. 5,000,000
  - **Average rows per EHCC compression unit = 1,000.**

- Determine compression used for table / partition / row:
  - DBMS_COMPRESSION procedures.
  - COMPRESS_FOR column in DBA_TABLES / DBA_INDEXES.

Exadata Smart Flash Cache

- Each storage cell has four PCI Flash cards:
  - Combined storage makes up the “Exadata Smart Flash Cache”.
  - Intelligently populated with “hot” data.
    - Consider pinning hot objects (partitions) if Flash Cache ratio is low.
- Acts as “second” database buffer cache.
- Can be used for writes with **Writable Smart FlashCache**:
  - NOT enabled by default – useful for OLTP processing.
  - Accelerates log writes via Smart Flash Log.

Am I Using Flash Cache?

- 2 statistics determine Flash Cache ratio:
  - *physical read total IO requests* – e.g. 3,000,000
  - *cell flash cache read hits* – e.g. 2,850,000
  - **Flash Cache ratio (100 / 3,000,000) * 2,850,000 = 95%**

- **Target state: fit as much of the "hot" data in FlashCache:**
  - Increasing size of Flash Cache in new Exadata models.
  - Flash Cache Compression (cost option).
  - Exponential performance benefit when combined with SmartScans and EHCC.

HugePages and Memory Management

- Set USE_LARGE_PAGES = ONLY
- **DON'T** use Automatic Memory Management (even if you're not using HugePages).
- Use Automatic Shared Memory Management (the 10g version).
- Set minimum component sizes to avoid excessive component resizes / contention.
- Don't make the buffer cache too large!

Database Services

- Separates users / workload into groups:
  - DB_BATCH, DB_ADMIN, DB_REPORTING, etc.
- Use with DBRM to prioritize resource consumption.
- Identify / isolate problem activity.
- Simplify Data Guard switchover:
  - Use both primary / standby SCAN addresses for LDAP / TNS.
    - Main services start on the primary database only.
    - Reporting services start on the physical standby database only

New Operating Procedures

- Specialized SOPs required for DMA:
  - Exadata: Reboot Comp Node / Storage Cell / IB Switch.
  - Exadata: Replace Failed Storage Cell Disk.
  - Exadata: Check for Errors on an IB Switch.
  - Exadata: Component Backups.
  - Exadata: Install ExaWatcher.
  - Exadata: Using ILOMs.
  - Exadata: Performance Ratios and IOPS.
  - Exadata: Capacity Monitoring on Exadata.
  - Exadata: Object, Dictionary and System Statistics.

“Must-Have” Support Notes

- DMA must be aware of new problems / best practices:
  - Exadata Database Machine and Exadata Storage Server Supported Versions (Doc ID 888828.1)
  - Exadata Critical Issues (Doc ID 1270094.1)
  - Information Center: Oracle Exadata Database Machine (Doc ID 1306791.2)
  - Oracle Sun Database Machine X2-2/X2-8, X3-2/X3-8 and X4-2 Security Best Practices (Doc ID 1071314.1)
  - Oracle Exadata Database Machine exachk or HealthCheck (Doc ID 1070954.1)
  - Responses to common Exadata security scan findings (Doc ID 1405320.1)

Monitoring Exadata

- OEM Cloud Control includes Exadata-specific plug-ins.
- Configure email alerts from the storage cells.
- Oracle Configuration Manager:
  - Use OCM or OEM Harvester to send system information back to Oracle Support.
  - Simplifies creation of Support Requests.
  - **OCM-enabled Exadata systems treated with Priority Handling**.

Diagnostics and Performance

- exachk:
  - Scores out of 100 based on best practice compliance.
  - Should be run on a regular basis.
  - **Checks entire machine for problems.**
  - Suggests remedial actions and performance improvements.
  - Can report false positives:
    - UNRECOVERABLE datafiles.
- ExaWatcher:
  - Exadata version of OSW (Black Box)
  - Oracle Support **WILL** ask you for the logs.

Capacity Monitoring

- V1 and V2 machines had finite storage.
- **Very important** to determine available storage and rate of data growth.
- Capture historical information in “DBA” schema:
  - ASM diskgroup, DBA_SEGMENTS.
  - Determine IOPS during peak workload.
  - Monitor performance ratios:
    - SmartScan, EHCC rows per CU, FlashCache.
- Consider AWR Warehouse Repository.

Exadata As a Foundation (The Hub)

- Thanks to InfiniBand connectivity, Exadata integrates with other engineered systems:
  - ZFS / Zero Data Loss Recovery Appliance for fast backups/restores.
  - Exalogic Appliance for application servers.
  - Exalytics Appliance for …?
  - Big Data Appliance:
    - Hadoop environment-in-a-box.
    - Big Data SQL allows Oracle RDBMS database to query data on Hadoop.
    - End users don't need to code their own extraction layer.
    - Advanced Oracle features (security, Data Guard, etc) available.

Exadata As a Foundation (Features)

- Oracle In-Memory Database:
  - Enable extreme OLAP performance via in-memory columnar representation of data.
  - “No" application changes (need to get rid of analytical indexes).
  - No need for OLAP repository / data marts:
    - Reduce data movement.
    - Real-time data can be used for analytical processing.
    - Save on licensing, code maintenance, support costs.
- Oracle 12c Multitenant
  - Use Exadata as a core component of “private cloud”.

DBA 4.0?

- DBA / DMA role will continue to expand:
  - "It's got Oracle / data in its name!"
  - Increasing emphasis on **DMA** role rather than DBA role.

- DBA 4.0:
  - Big Data systems integration.
  - On-demand service provisioning (12c, Multitenant).
  - High-performance features such as In-Memory Database.
  - Expansion into infrastructure architecture.
  - Traditional organization silos can't support expanding scope:
    - Engineered Systems Support?
    - Enterprise Data Management?

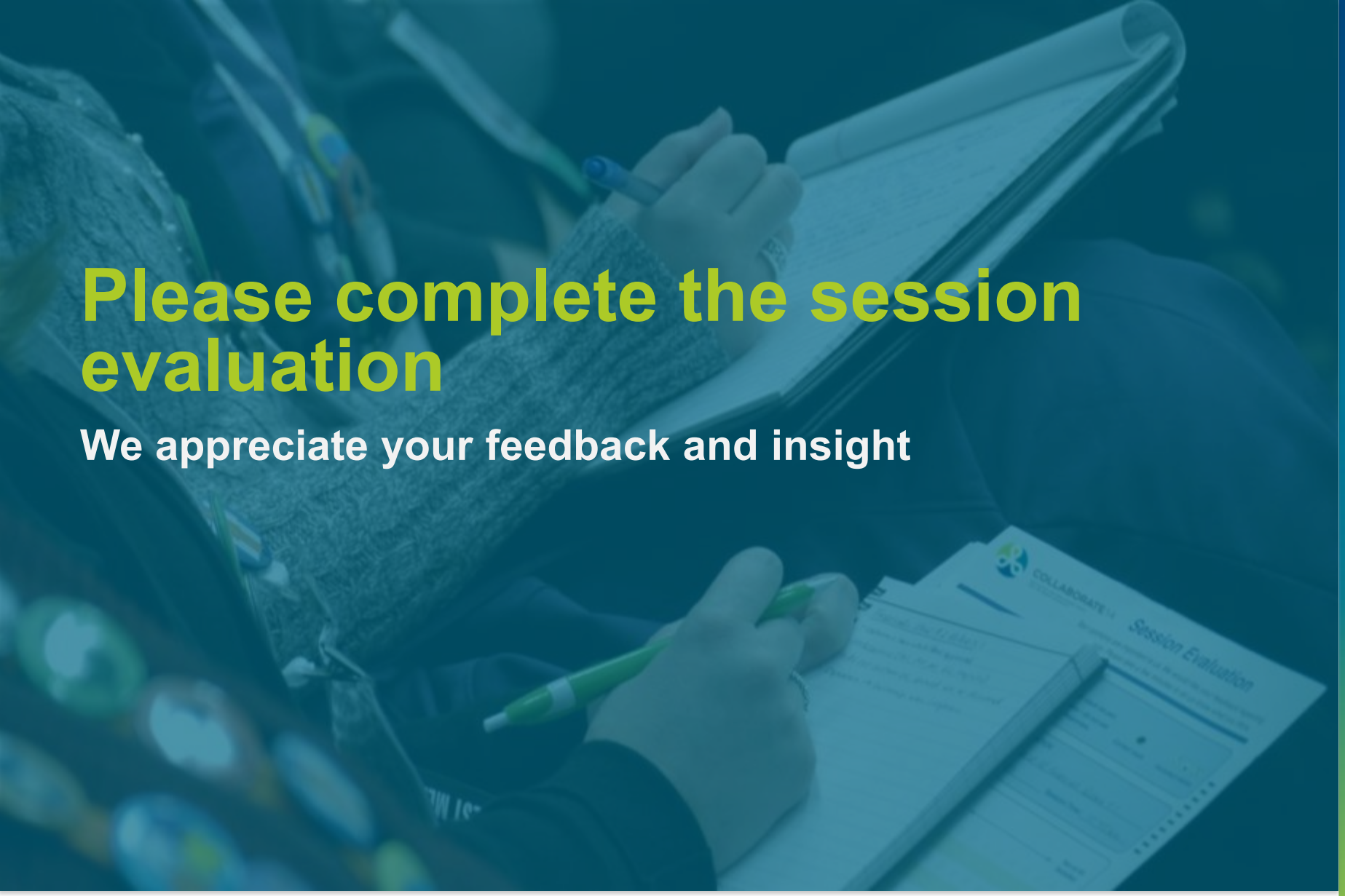
Questions? Comments?

Expert Oracle Consulting and
Remote Database Administration

www.dbspecialists.com
1-888-648-0500

msmith@dbspecialists.com

COLLABORATE15
TECHNOLOGY AND APPLICATIONS FORUM
FOR THE ORACLE COMMUNITY



Please complete the session evaluation

We appreciate your feedback and insight

You may complete the session evaluation either on paper or online via the mobile app

COLLABORATE15
TECHNOLOGY AND APPLICATIONS FORUM FOR THE ORACLE COMMUNITY

#C15LV