Exadata Support Checklist

This serves as a support checklist for Exadata DBAs / DMAs, whether taking delivery of a new machine, learning how to support an existing machine or conducting a regular configuration review.

When taking delivery of a new Exadata machine, learning how to support an existing machine or conducting a regular configuration review, use this checklist as a reference document to ensure that a comprehensive range of Exadata-specific considerations are reviewed.

Best Practices, Support Issues, Critical Issues

- MOS Articles:
  - Exadata Best Practices (Doc ID 757552.1)
  - Exadata Critical Issues (Doc ID 1270094.1)
  - Exadata Database Machine and Exadata Storage Server Supported Versions (Doc ID 888828.1)
  - 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)
  - ExaWatcher utility on Exadata database servers and storage cells (Doc ID 1617454.1)
  - Exadata Smart Scan FAQ (Doc ID 1927934.1)
  - Oracle Sun Database Machine Application Best Practices for Data Warehousing [Doc ID 1094934.1]
  - Tool for Gathering I/O Resource Manager Metrics: metric_iorm.pl [Doc ID 1337265.1]
- Review Best Practices and add changes to the "DBA Change List"
- Review Critical Issues and add changes to the "DBA Change List"
- Review End of Support notifications and plan upgrades

Perform Regular exachk Reviews

- Record score, fail/warnings in customized "DBA" tables (i.e. DB001 schema in TOOLS tablespace).
- Add any changes to the "DBA Change List" and schedule outages
- Raise SR(s) if necessary to confirm remediation / false positives.

Database Configuration

- Standard database setup:
  - SPFILE, memory, RAC, SQLNet
  - Maintain list of non-default parameters and comments in "DBA tables" especially for hidden parameters

    - Do NOT use AMM, use HugePages and set use_large_pages=ONLY
- FlashCache:
    - Are there any objects (tables, indexes, partitions) that we should "pin"?
    - Storage cell can show list of objects which are in the FlashCache at that time need to link back to object_id in the database
    - AWR allows us to determine usage of FlashCache
    - WriteBack FlashCache for OLTP processing offers the ability to WRITE to the FlashCache
- Exadata Smart Flash Log:
    - by default, in ESS 11.2.2.4.0, 512Mb of "flashlog" memory is allocated to help minimize redo log waits
- Exadata Hybrid Columnar Compression (EHCC):
    - Which tables are using EHCC?
    - Which compression method/ratio is being used
    - AWR allows us to determine usage of EHCC
- SmartScans/storage cell offload
    - AWR allows us to determine usage of SmartScans

ASM Configuration

- Where to keep parameter file?
- If SPFILE is kept on ASM diskgroup, when CRS shuts down the diskgroup, it cannot do so "cleanly".
- Should we keep PFILE local?
- Check processes setting
- Disk group redundancy
- If using Data Guard, "NORMAL" redundancy is acceptable, especially on quarter-rack
- Oracle's recommendation is to use "HIGH" redundancy, not viable on quarter-rack machines as it drastically reduces the usable free storage
- Use proper ASM diskgroup attributes (compatible, cell.smart_scan_capable, au_size 4M)
- Review the ASM POWER LIMIT and DISK REPAIR TIME

General O/S Configuration

- Check the exachkcfg autostart service status
- Verify that CSS misscount = 60
- Review users, groups, hostnames, user profile, aliases
- Disk cache policy should be disabled
- Monitor ambient temperature
- Verify RAID controller battery charge and temperature
- Verify hardware and firmware on comp nodes and storage cells are consistent
- Comp nodes and storage cells using WriteBack (not WriteThrough)
- Verify ILOM power up configuration: HOST_AUTO_POWER_ON=disabled, HOST_LAST_POWER_STATE=enabled

Compute Node-Specific Configuration

- Verify no outstanding hardware alerts using "show faulty" with "ipmitool sunoem cli"
- Portmap and nfslock services have to be running if we're using NFS
- Use HugePages
- "Locked memory" should total 75% of physical memory (max)

- Shared memory segment max size = 85% of physical memory
- Sum of processes does not exceed maximum number of semaphors
- Number of semaphores in a semaphore set must be at least as high as the processes parameter in ALL databases
- Size of Shared Memory Segments OS setting for max size = 85% of physical memory
- Verify disk controller configuration on comp nodes
- Verify physical and virtual drive configuration on comp nodes
- Verify that NUMA is NOT enabled on the comp nodes
- Verify that RAC databases use RDS and not UDP to communicate
- Set SQLNET.EXPIRE_TIME = 10 in the RDBMS home

Storage Cell-Specific Configuration

- cellconf check especially for NTP servers, SNMP configuration
- Check for WriteBack/WriteThrough mode (disk cache policy)
- Check for ECC memory errors on the storage cells
- Verify celldisk and flashdisk configuration (no griddisks on Flash!)
- Confirm that total size of all griddisks fully utilizes celldisk capacity
- ipconf parameter file must be consistent with O/S configuration
- Verify that cell services are up and running

InfiniBand Configuration

- Check that IB is the PRIVATE network for cluster communications
- Check for ports disabled due to excessive symbol errors
- Check IB ARP (Address Resolution Protocol) is correctly set up on comp nodes
- Verify IB cable connection quality
- Verify Ethernet cable connection quality
- Verify IB fabric topology
- Check for IB network errors
- Verify IB subnet manager is running on an IB switch
- Verify IB subnet manager is not able to run on anything except an IB switch
- Verify key parameters in the /etc/opensm/opensm.conf file
- Verify IB network throughput (infinicheck) this HAS to be run during a "quiet" time as it evaluates full network throughput

Network Configuration

- Client network should be bonded on bondeth0
- IB network should be bonded on bondib0
- Admin network cannot be bonded and must be eth0 ON ITS OWN SUBNET
- Is 10GiB hardware enabled and used (optical)?
- Do we run a dedicated backup or Data Guard network?
- Verify average ping times from the comp nodes and storage cells to the DNS servers
- May need to set network routes/rule for client network/admin network, especially if OEM is in the DMZ (cannot use the client network)
- Monitor usage

Backup, Recovery, Disaster Recovery and High Availability

- Baremetal restores:
  - storage cells, switches, comp nodes, LVM
- RMAN:
  - backup schedule, logs and reports (success, timing), check for block change tracking
- Configuration file backups:
  - PFILE, ASM PFILE, encryption keys, LVM, tnsnames.ora, listener.ora, sqlnet.ora, DCLI groups, hostnames, oraInventory, network interfaces, network rules and routes, RMAN scripts, crontab, cronjob scripts
- Data Guard:
  - DB params, Data Guard Broker configuration, switchover steps/checklist, status tests
  - Active Data Guard?
  - Make sure that the DGB timeout is longer than other timeouts (clusterware, SQLnet, etc 90 seconds at least)
- ZFS:
  - Use of ZFS storage appliance on InfiniBand to reduce backup/restore time
- FLASHBACK/archive/ORL/SRLs:
  - FLASHBACK DATABASE enabled?
  - Where are the Flashback logs located?
  - Separate out the archive logs from the Fast Recovery Area to avoid conflict with Flashback logs.
  - How long should we be able to Flashback?
  - Archive log location, backup and housekeeping
  - Forced checkpoints, archive log switches, MTTR
  - Log review ORLs, SRLs, logfile groups and sizes, log_buffer parameter
- Recovery Time Objective:
  - How LONG will it take to restore to the appropriate point-in-time?
- Recovery Point Objective:
  - WHAT is the point-in-time?
- Recovery tests:
  - restore of database and configuration files
  - restore of encryption keys
  - switchover tests

Monitoring

- OEM Cloud Control:
- Agents, alerts
- Platinum Support maintain their own OEM agent
- Storage cells:
- alerthistory, list physicaldisk, SNMP setup, customized scripts
- sundiag.sh for failed disks
- Switches:
- ibcheckerrors
- Syslogs:
- for comp nodes, storage cells, switches

Platinum Support

- Patching planning
- List components to be monitored (databases, binaries)
- Access to AMR portal
- Maintain list of "known issues" for which PS should not alert
- PS gateways?
- PS monitoring OEM agents?

Resource Management

- DBRM:
  - Consumer groups with resource limits
  - Maintenance windows?
  - Different resource plans
- IORM:
  - Set up for machines with multiple databases
  - Is Instance Caging in use?
- NetRM:
  - Available after QFSDP Jan 2014 switch firmware update

Database Service Management

- Set up database services and tnsnames.ora files for clients
- Monitor using listener/SCAN listener logs and AWR
- Associate services with consumer groups
- Maintain "DBA table" with connection/service information

Automatic Workload Repository

- Snapshot frequency, retention, retained snapshot periods
- Create AWR "repository" in "DBA tables" for historical purposes
- Maintain periods for comparison (before and after upgrades)
- List of important values to review
- Regular review

Auditing

- DB:
  - DBA audit trail housekeeping, create "summary" trail in "DBA tables"
  - List of audit options/statements enabled
  - Make sure the audit trail AND the FGA audit trail are NOT in the SYSTEM tablespace they need to be in an ASSM tablespace (SYSAUX)
- **ASM:**
  - ASM audit trail housekeeping
- OS:
  - OS DB audit trail (SYSDBA) housekeeping
  - auditd logs (sudo)

Historical Support / Monitoring Data

- Tables containing important admin/support/monitoring data:
- List of "audited" database users and activity
- List of "application" users and access
- Capacity planning storage/safely usable, new data growth, HCC impact, IOPS, memory
- Audit trail summary
- List of non-default and hidden parameters with "comments"
- exachk scores, dates, fail/warnings
- Create external table populated using OPatch
- Serial numbers for each server (external table)
- Product name for each server (external table)
- Associated CSI for each server
- Check for OCM "hardware configuration" set up in MOS
- Maintain list of failed disks (SRs, serial numbers, dates, etc)
- Maintain options, stack software versions, bug fixes, critical issues, upgrade dates, etc

Capacity Planning

- Leverage "DBA tables" for capacity history
- Check storage (total datafile size, total segment size, ASM diskgroup usable space via external table)
- Review HCC configuration and usage (AWR)
- Review FlashCache usage (AWR)
- Review storage cell offloading (AWR)
- Review IOPS, storage, memory, memory structures, disk throughput

Statistics

- Assumes copious amounts of partitioning
- SQL Plan Management:
  - needs some initial "tweaking" for the optimizer to accept the "best" explain plan
  - doesn't work well without bind variables could end up with billions of plans which are checked by the optimizer, causing bad performance
- Incremental global stats gathering:
  - partitioned tables often only update statistics for the local partition
  - global table/index stats are NOT updated unless "incremental global stats" are enabled
  - if NOT enabled, run weekly job to pick the "top 25" more stale tables and explicitly gather their global stats
- Dictionary statistics:
  - run these once a month
- System statistics
  - do not run these without a change in hardware
  - use the EXADATA mode to gather system statistics

General Housekeeping

- Keep trace files for 31 days (alert.log, css.log, syslog, etc)
- gzip the alert.log every day

- gzip the listener/SCAN logs every month
- Purge the audit trail every day (keeping 14-28 days' data)
- Consider using an "audit archive" summary table to maintain data for longer
- Purge the O/S audit trail every day (keeping 31 days)
- Run capacity planning "capture" every day
- Run datafile "shrink" commands weekly
- Run statistic gathering regularly global stats on stale tables (weekly), dictionary stats (monthly)