

Tuning Exadata

**But
Why?**

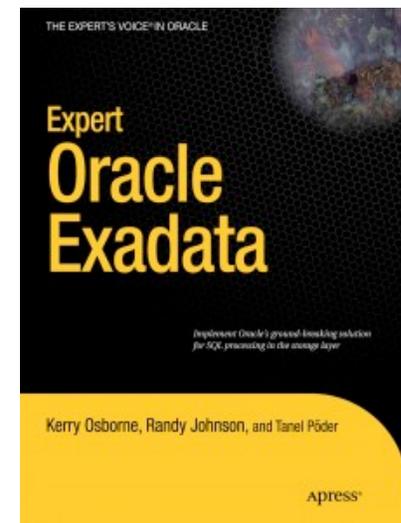


whoami –

Work for Enkitec (www.enkitec.com)
Working with Exadata since early 2010
Many Exadata customers and POCs
Many Exadata Presentations (some to Oracle)
Working on Exadata Book



(kerryosborne.oracle-guy.com)



whoami – (my prejudices)

I am an Exadata Fan Boy
– so please take everything I say with a grain of salt



Agenda

Exadata Basics

Offloading

- Optimizations

- Requirements

- How to tell if it's working *****this is very important*****

Demo

New Way of Thinking

Questions

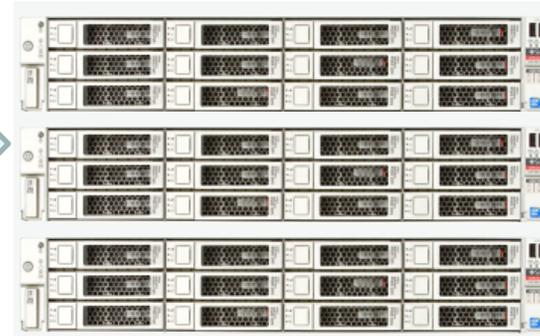
Exadata Basics

Exadata Database Servers



11gR2 / ASM

Exadata Storage Servers



cellsrv

Infiniband

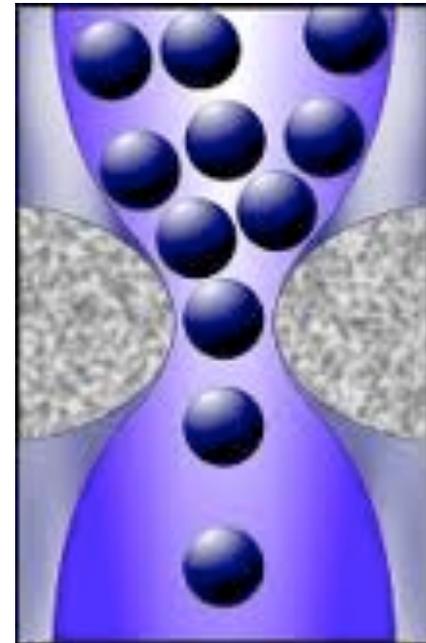
iDB / RDS

The Big Ah Ha!

The Bottleneck on Many (Most) Large Databases is between the Disk and the DB Server(s)!

How to Speed Up?

**Make the Pipe Bigger/Faster
Reduce the Volume**



* The fast way to do anything is not to do it ~ Cary Millsap

Offloading – The “Secret Sauce”

Offloading vs. Smart Scan (what’s the difference)

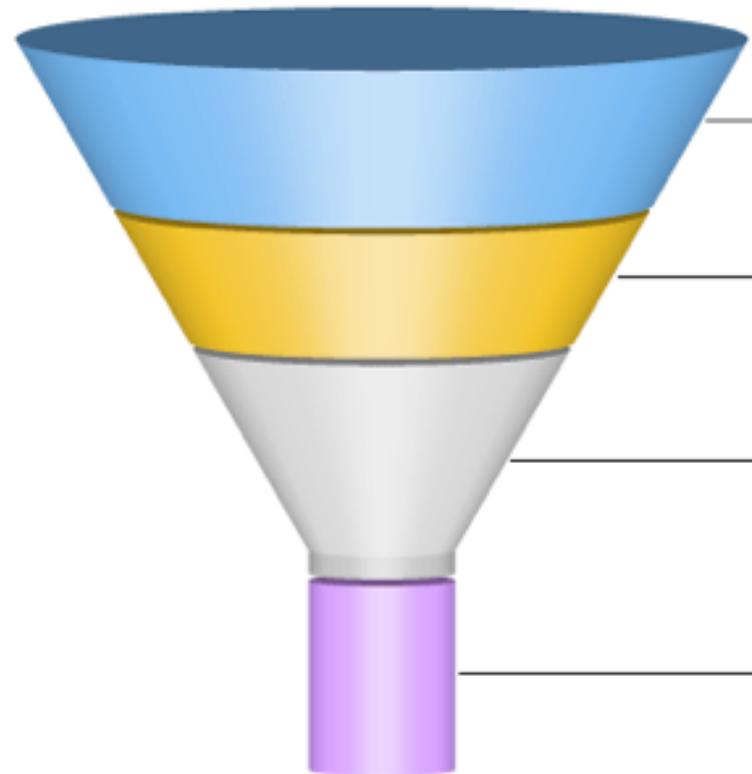
Offloading – generic term meaning doing work at the storage layer instead of at the database layer

Smart Scan – query optimizations covered by “cell smart table/index scan” wait events



Smart Scan Optimizations

Column Projection
Predicate Filtering
Storage Indexes
Simple Joins
Function Offloading
Virtual Column Evaluation
HCC Decompression
Decryption



How to Tune an Exadata (but not a fish)

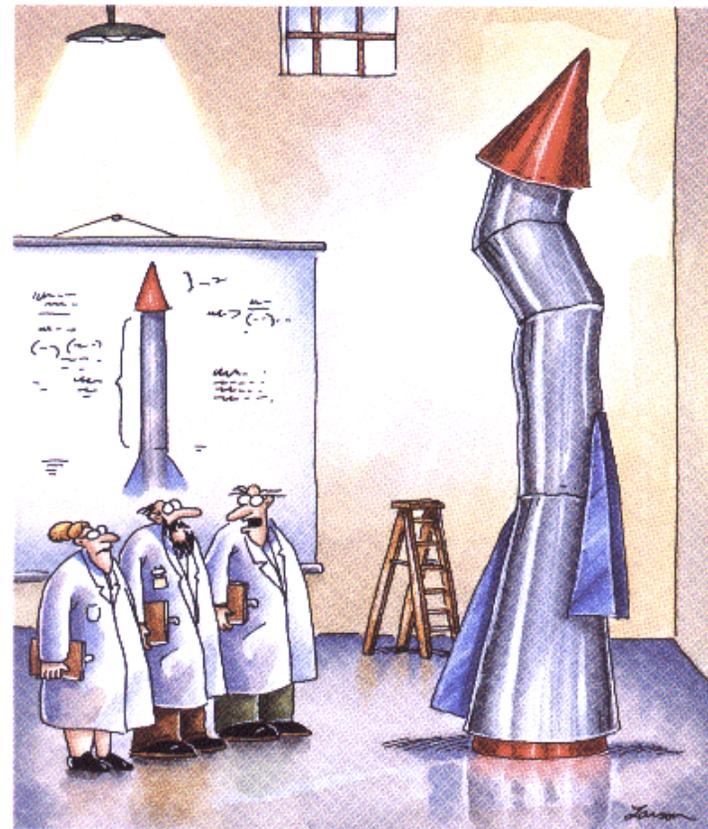
Check to see if you're getting Smart Scans!

If you're not, figure out why and correct the situation!

It's Pretty Simple.

3 things you'll need to know:

- the Optimizations
- the Requirements
- how to Measure



"It's time we face reality, my friends. ... We're not exactly rocket scientists."

Smart Scan Requirements

Full Scan
Direct Path Read
Object Stored On Exadata Storage

Why?

Very Simple Explanation:

Various full scan functions()

- `kcblldrget()` – direct path read function
- `kcfis_read()` – kernel file intelligent storage read (Smart Scan)

*why it's there: checkpointing and non-block data return



Requirement 1: Full Scans

- Table
- Partition
- Materialized View
- Index (FAST FULL SCAN Only)

```
SYS@shareprd1> @op_event_awr.sql
Enter value for event: cell smart%
```

EVENT	OPERATION	COUNT(*)
cell smart index scan	INDEX STORAGE FAST FULL SCAN	124
	INDEX STORAGE SAMPLE FAST FULL SCAN	234
cell smart table scan	MAT_VIEW ACCESS STORAGE FULL	1
	TABLE ACCESS STORAGE FULL	27747

* Query from DBA_HIST_ACTIVE_SESS_HISTORY

Digression - New Exadata Wait Events

cell list of blocks physical read
cell multiblock physical read
cell single block physical read
cell smart file creation
cell smart incremental backup
cell smart index scan
cell smart restore from backup
cell smart table scan

* Note that there are others, these are the most interesting

Requirement 2: Direct Path Reads

Bypass buffer cache – direct to PGA

Decision not part of optimizer's job

Traditionally Used by Parallel Slaves

Non-Parallel Also Possible

- Serial Direct Path Reads (adaptive)
- algorithm not documented (but more aggressive in 11g) *
 - size of segment (table or index or partition)
 - size of buffer cache
 - number blocks already in buffer cache
 - `_small_table_threshold`
 - `_very_large_table_threshold`

* See MOS Note: 50415.1 - WAITEVENT: "direct path read" Reference Note

Requirement 3: Exadata Storage

Kind of Goes Without Saying

- Possible to have non-Exadata storage or mixed
- ASM Diskgroup has an attribute: **cell.smart_scan_capable**
- Must be set to TRUE for Smart Scans to work
- Can't add non-Exadata storage without changing to FALSE

How to Tell if You got a Smart Scan

Millsap It!

- (10046 trace)
- most fool proof

~~Wolfgang It!~~

- unfortunately this doesn't work
- 10053 trace (and the optimizer) has no idea

TP It!

- Tanel's snapper
- v\$sesstat, v\$session_event
- great if it's happening now

KO It!

- My fsx.sql script
- V\$SQL family of views: IO_CELL_OFFLOAD_ELIGIBLE_BYTES
- dbms_sqltune.report_sql_monitor works also
- saved in AWR so works on historical data as well

How to Tell if You got a Smart Scan

```
-- fsx.sql

select sql_id,
       decode(IO_CELL_OFFLOAD_ELIGIBLE_BYTES,0,'No','Yes') Offloaded,
       decode(IO_CELL_OFFLOAD_ELIGIBLE_BYTES,0,0,
              100*(IO_CELL_OFFLOAD_ELIGIBLE_BYTES-
                    IO_INTERCONNECT_BYTES)/
                    IO_CELL_OFFLOAD_ELIGIBLE_BYTES) "IO_SAVED_%"
from v$sql
where sql_text like '&sql_text';
```

* Warning: there are occasions where it's not accurate

The Wrong Tool for the Job?



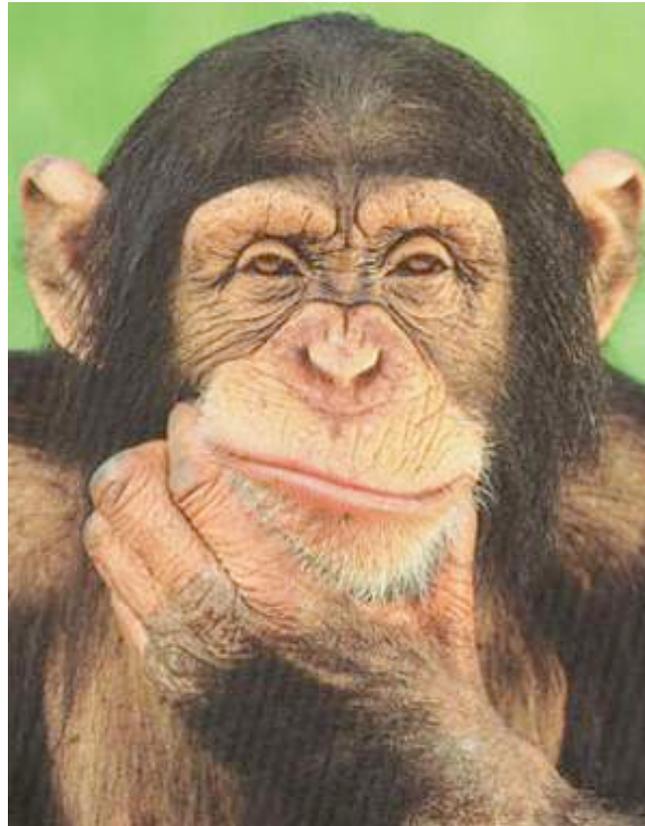
Maybe:

Any of the tools can do the job. Just depends on the circumstance and you're preferences.

Demo Time



A New Way of Thinking



Smart Scans Are Key

Need to be able to tell whether they are being used
Need to be able to tell what optimizations are in play
Need to know what disables them



Other Smart Scan Disablers

IOTs
Clustered Tables
Rowdependencies Enabled
LOBs

No Direct Path Reads *

* Note: You can get direct reads and still not get Smart Scan

Direct Reads?

Gotta Have Em!

How do we get Em?

PX

- not really as easy to control as you might think

Serial Direct Read / Adaptive Direct Read

- Algorithm is complicated
- Works on Indexes as well by the way

Serial Direct Reads

Common Reasons Not to Get Them:

- Big SGA
- Small Table
- Partitioning – decision based on segment size
- Indexes – generally relatively small

Workarounds:

- PX
 - DEGREE=2
 - PARALLEL_DEGREE_POLICY=AUTO ← Scary
- “_serial_direct_read”=always (not OPT_PARAM compatible)

How Do It Know?

The Optimizer Doesn't Know
About Smart Scans

Clues:

The cost calculations are the same
10053 trace file shows nothing
Decision to do direct path read
is not part of optimizer
- and code for Smart Scan is below kcbldr

Results:

When indexes are present, Exadata databases tend to over use them



To Index or Not to Index?

Myth: Exadata doesn't need any indexes

Truth: You'll need indexes for single row access (OLTP)

Note: Moving to Exadata will allow you to get rid of a bunch of indexes that you weren't using in the first place.

Note2: Moving to Exadata may make many indexes that are being used unnecessary.



* Hint: Make them invisible first, and then remove them!

Partitioning?

...Is Still Your Friend!

It's all about eliminating work (time):

Storage Indexes can further reduce this, or in some instances be used instead of Partition Elimination.

But you should not consider them a replacement for Partitioning.

Partitions Also Provide Manageability:

Partitions can have different storage characteristics (HCC)
Don't forget that Serial Direct Read looks at Partition Size

Parallel?

There are many new features in 11gR2:

- Auto DOP
- Parallel Queuing
- In-Memory Parallel

These are not specific to Exadata.

PX will be important in Exadata (uses Direct Path Read)

Every query is parallelized across multiple storage cells

May mean you don't need as high DOP

Auto DOP is probably the wave of the future but still scary

Nulls?

Can't be indexed via B-Tree
Can't do Partition Elimination

But they can be located with Storage Indexes

We may want to re-think the use of nulls

Compression (HCC)

Don't even think about compressing active data...

- Every change migrates the affected row to a new block (OLTP)
- Every change locks the entire Compression Unit
- Partition large objects by time and compress inactive partitions

Mixed Workload Systems

Flash Cache is Key

- Expect 1-2ms Single Block Reads
- If Not Getting Them, Check for FC Problems
- Consider setting CELL_FLASH_CACHE to KEEP
- Remember Indexes Can Be Overused
(optimizer_index_cost_adj)

Last Thoughts

Take Some Time to Test

- Just Because You Can Slam it in Doesn't Mean You Should

Take Some Time to Understand the Exadata Optimizations

- Know What to Expect

Take Some Time to Evaluate Indexes

- Migration is a Golden Opportunity to Get Rid of Some
- Make Sure the Ones You Keep Aren't Overused

Exadata Job Posting

Requirements

- * Previous consulting experience is preferred
- * Full life-cycle implementation experience
- * Candidates should have 10+ years of Oracle Exadata experience
- * Candidates should be willing to travel up to 80-100%
- * Minimum Bachelors degree preferred
- * Citizens/GC/EADs are allowed --- No H1b

Questions / Contact Information



Questions?

Contact Information : Kerry Osborne

kerry.osborne@enkitec.com
kerryosborne.oracle-guy.com
www.enkitec.com