Exadata

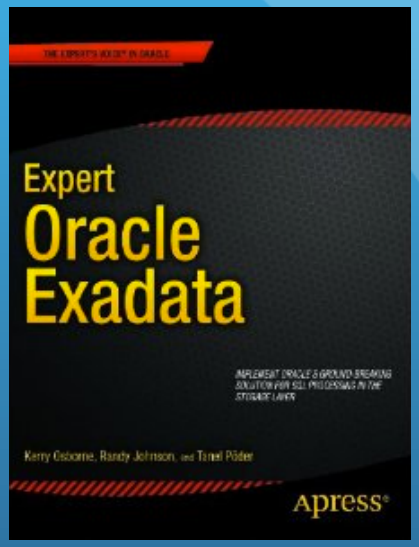
Presented by: Kerry Osborne

February 23, 2012

enkitec

whoami -

Worked with Oracle Since 1982 (V2)
Working with Exadata since early 2010
Work for Enkitec (www.enkitec.com)
(Enkitec owns a Half Rack – V2/X2)
Many Exadata customers and POCs
Many Exadata Presentations (many to Oracle)
Exadata Book



Blog: kerryosborne.oracle-guy.com

OakTable.net

ORACLE ACE Director

enkitec

What's the Point?

Can we get near Exadata performance …
… without buying an Exadata?

- Commodity Hardware
- Published Specs
- Specs are Easily Reproduced (or Exceeded)
- So the Question Comes Up Frequently …

Note: This presentation was originally proposed as a session for OpenWorld 2010 by Kevin Closson.

enkitec

Poll – Can You Get Exadata Like Performance w/o Buying an Exadata?

Yes – I think I can build a better mousetrap (for less money)

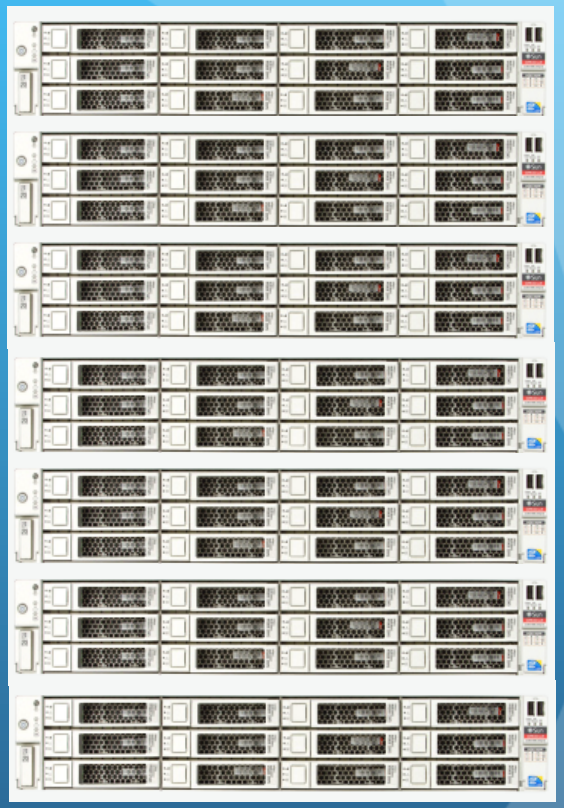
No – It absolutely cannot be done

Maybe – I think I might be able to get pretty close

enkitec

What is Exadata?

Exadata Storage Servers



Exadata Database Servers

Infiniband

11gR2 / ASM

iDB / RDS

cellsrv

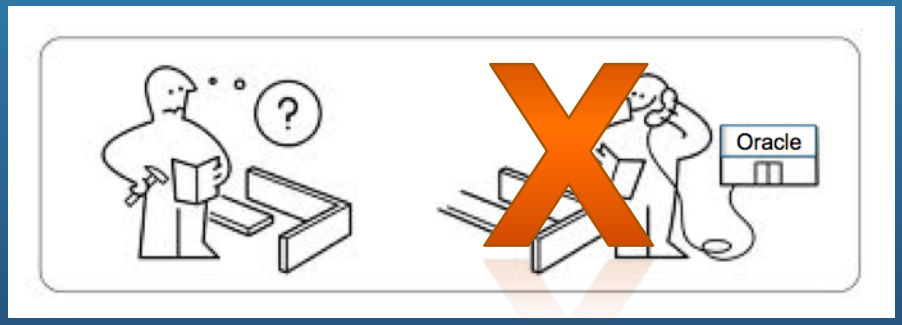
*Half Rack

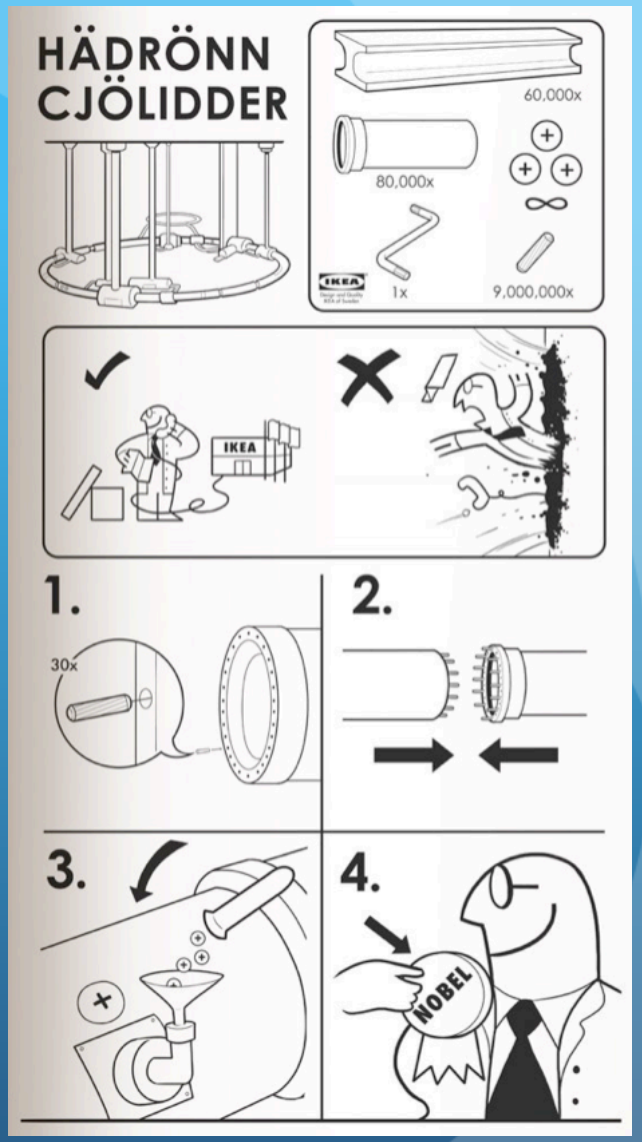
enkitec

What's the Plan?

Important Architectural Features

- Flash Cache
- Pipe (Infiniband)
- Compute Resources (CPU's)
- Total Storage
- Redundancy (RAC?)
- Manageability (Dial Home, ILOM, etc…)
- … remember there are tradeoff's





enkitec

Filling the Buckets*

Compute Capacity (cores) – 132 (48+84)
Storage (TB Usable) ~ 80 (252 raw)
Flash Cache (TB) ~ 2.6
Pipe (Gb/s) – 40
Redundancy ?
Manageability ?
Cost ?

*Half Rack

enkitec

First Iteration

Copy Exadata Specs

- 4 - Sun ~~X4170~~ Z Servers (need **4270**'s due to PCIe slots)
- 16 – 8G F~~C~~ s a~~n~~ Switch
- EMC ~~VNX~~ (need to step up to VMAX)
  - ~~10~~ – 20(ast Cache (need **20** due to RAID 1)
  - 84 - 7 Drives

But We're Already Off in the Weeds!

And we're not accounting for additional CPU's on storage tier.

enkitec

Storyville

Imagine a system that spends 4.5 hours every night doing a batch update of a Billion+ row table – one row at a time.

Which buckets are most important?

- Storage?
  - Capacity
  - Throughput
  - Latency
- Pipe?
- CPU?
- Memory?

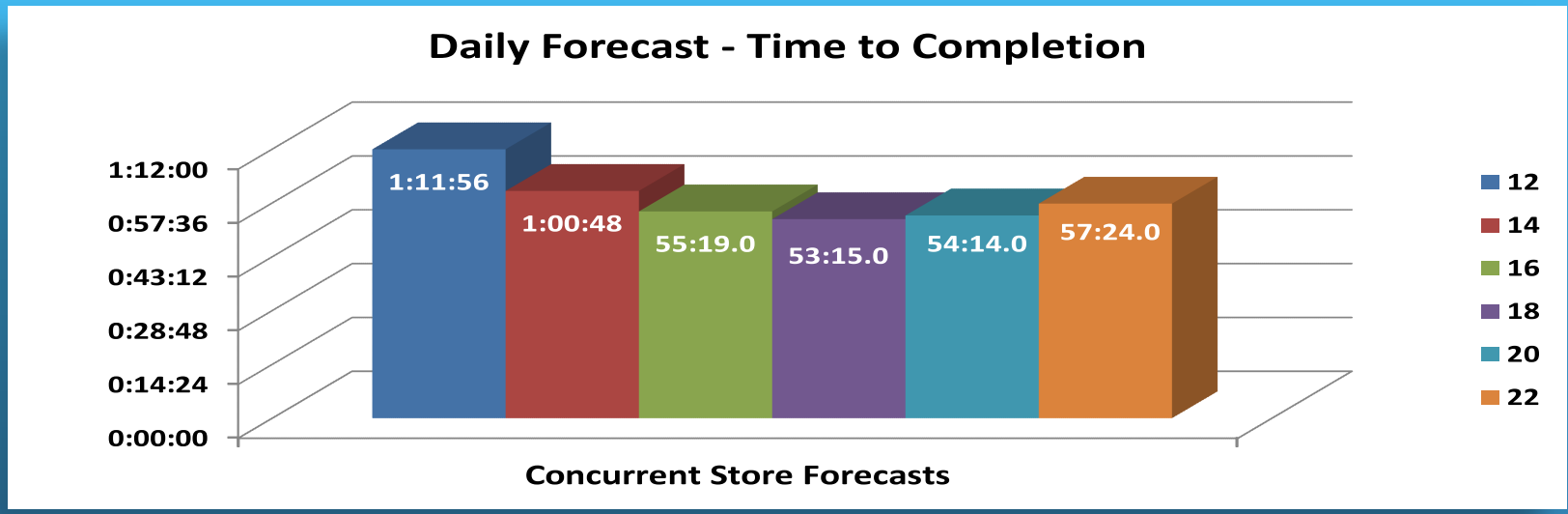
enkitec

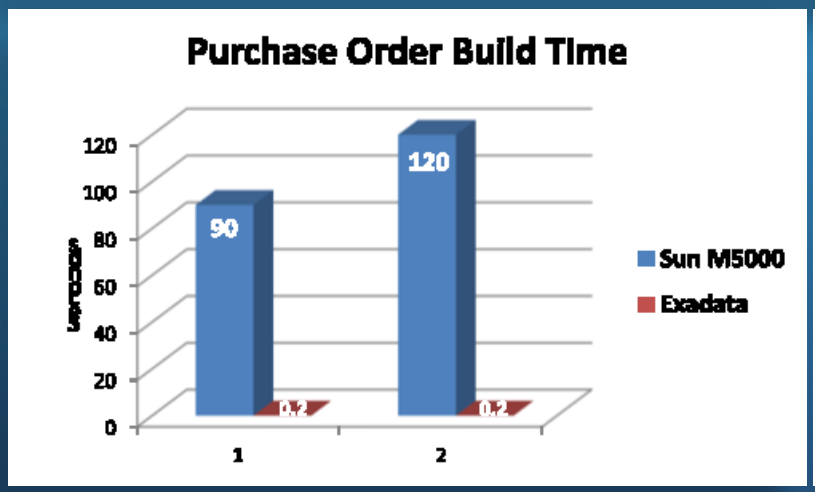
High Performance: Large Scale Retail Comparison

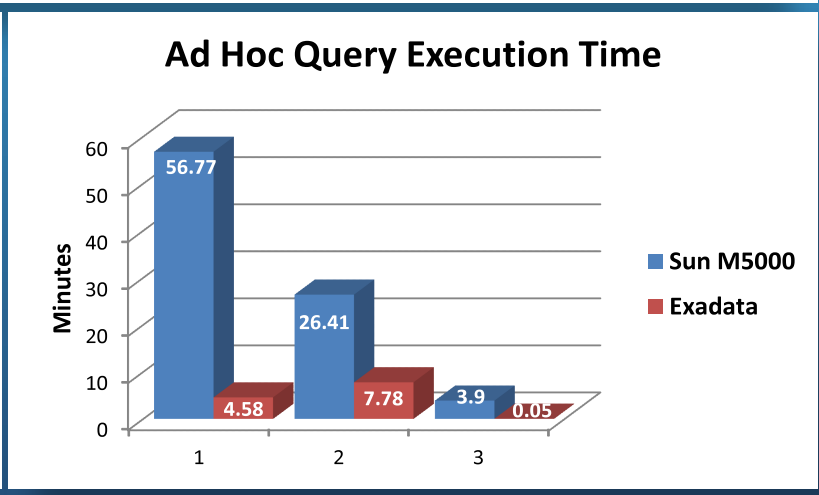
Customer Environment 32 Core (RISC) Max RAM Solid State SAN	Enkitec Exadata Environment Quarter Rack - 16 Core Intel
Test 1 - Nightly Forecast 4 Concurrent Stores Execution Time: 4.5 hours	**Test 1 - Nightly Forecast** 18 Concurrent Stores Execution Time: 53 Minutes
Test 2 - PO Build Plan Execution Time: 120 seconds each	**Test 2 - PO Build Plan** Execution Time: 0.2 seconds
Test 3 - Ad Hoc Queries 56 minutes 27 minutes 4 minutes	**Test 3 - Ad Hoc Queries** 4.5 minutes 8 minutes 3 seconds

enkitec

High Performance:
Large Scale Retail Overview







Customer Decided to Pursue DIY Route -

Second Iteration

SSD &
Big SGA

- 1 – Dell R91 x8 C , 256G RAM, PCIe)
  - 70G buffe c
- 4 – 8G HBA's a itch(es)
- Hitachi SAN (A 00)
  - 2 trays o 0G D (30)
  - 1 tray o sty sp things (38x450G)

No redundancy
Not enough storage
Did get write back cache
Long Running Queries still take a while
But Not Bad!

enkitec

DIY Results:

Batch Job: ~ 50 minutes
56 Min Query: ~ 15 minutes

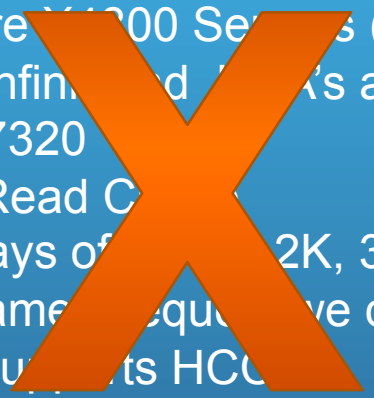
Costs:

- hardware roughly the same as half rack
- Oracle software quite a bit less

Third Iteration

Lots O'
CPU & A
Big Pipe

- 2 - Sun Fire X4200 Servers (plenty of CPU, memory, PCIe)
- 4 – QDR Infiniband HCA's and switch(es)
- Sun ZFS 7320
  - 2TB Read Cache
  - 4 – trays of 24, 2K, 3T Drives
  - per James Lee we can use RDMA
  - also supports HCC



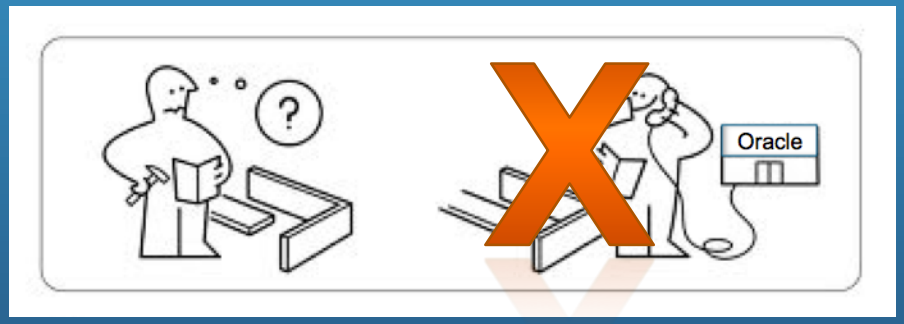
Getting close, but what's it going to cost?

enkitec

Digression: Got Balance?

- **DB Grid must generate I/O requests**
  - Generating I/O requests require CPU
- **Storage must be able to deliver the I/O**
  - **Need enough devices, etc…**
- **Transport mechanism must be adequate**
- **DB Grid must ingest the I/O**
  - Consuming I/O requires CPU

Basic idea is that we must be able to consume what is produced.



enkitec

Hardware Conclusion

- Exadata Architecture Provides a Roadmap
  - Flash Based Storage
  - Big Pipes (Infiniband)
  - Low Latency (RDMA)
  - RAC Provides Ability to Scale Out
- Unlikely that you can build it for anywhere near the cost
- But you can probably build something adequate for specific WL's

enkitec

Hardware is only half the story:

Remember:

- CPU's on Storage Cells Can Be Used For DB Processing
- So We Need More CPU on DB Servers To Compensate
- And the associated DB/RAC licensing costs
- We May Also Need More DB Server Memory
- All Because of the Storage Software

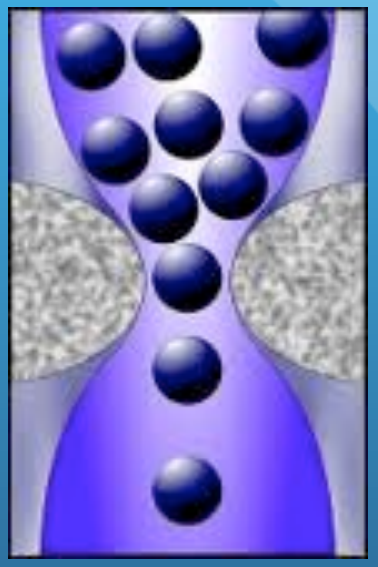
enkitec

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!

enkitec

Offloading – The "Secret Sauce"

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

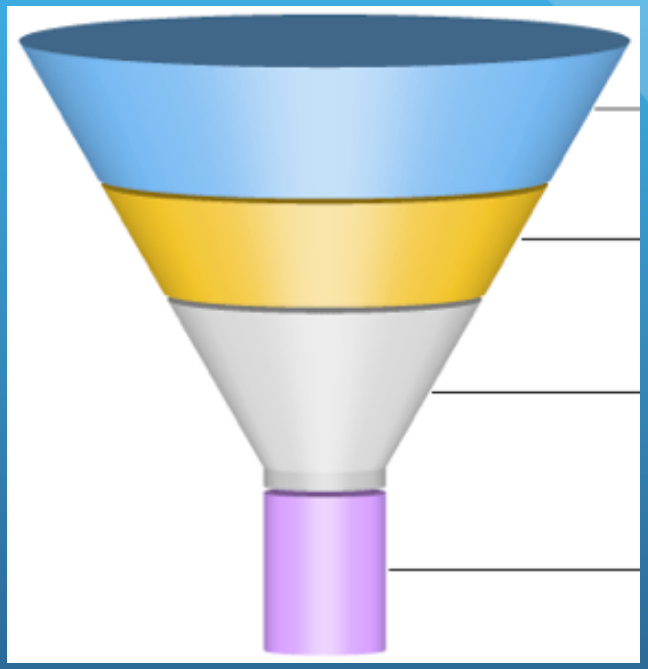
Offloading – generic term meaning doing work at the storage layer instead of at the database layer

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

enkitec

Smart Scan Optimizations

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

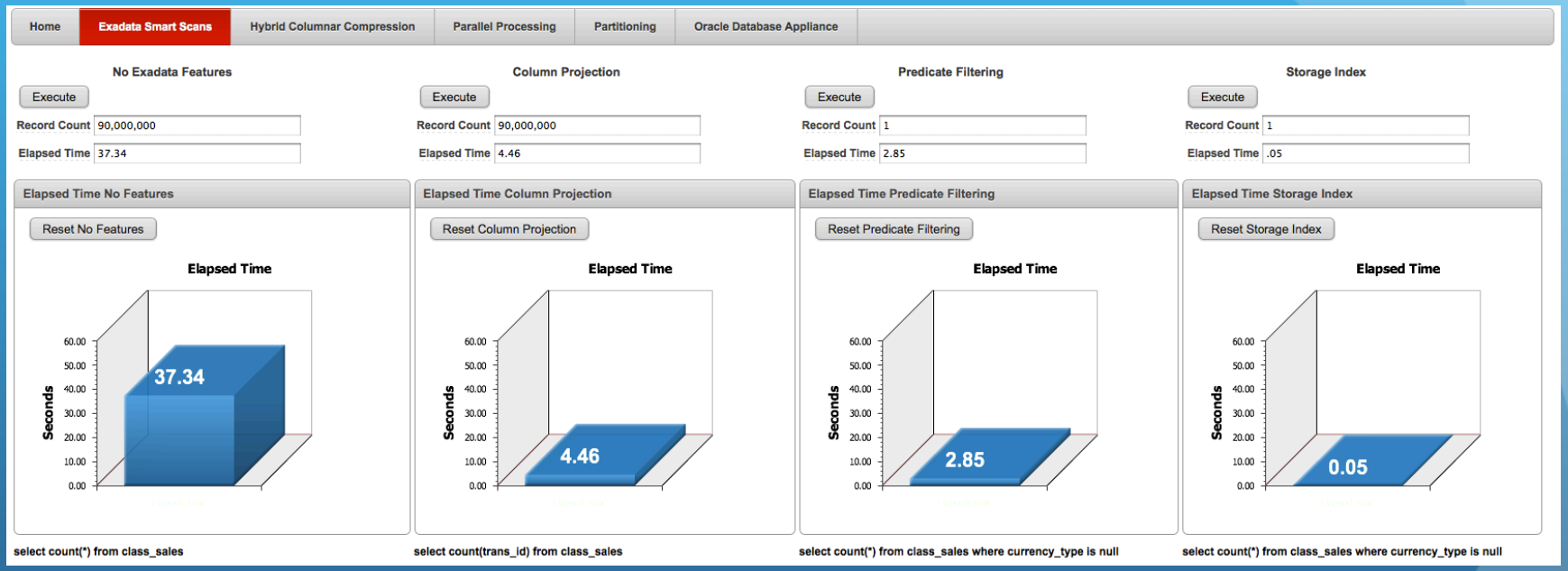


enkitec

Demo Time

enkitec

Exadata Software Performance



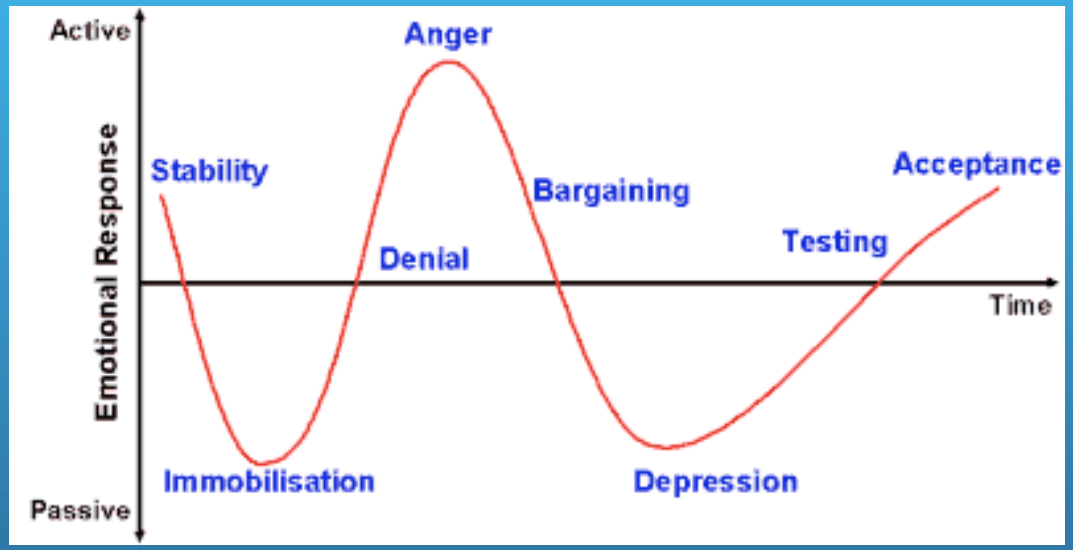
enkitec

High Transaction Volume: Telco Provider

- Customer Runs Dell, 16 Core Machines in Multiple RAC Instances
- Very High Volume of OLTP and Data Warehouse Type Queries on Same Database
- Performance Differences Were Too Excessive to Graph

SQL	Current	Exadata	Times Faster
Process 1: 6-Month Data Volume	52 min	19.5 sec	160 x
Process 2: 3-Month Data Volume	51 min	11.5 sec	269 x
Process 3: 1-Year Data Volume	50 min	37.5 sec	81 x
Process 4: 2-Month Data Volume	48 min	9.4 sec	308 x
Update SCN_CALL_PARTY_LOG	13 min	1.05 sec	744 x
Update SCN_CALL_PARTY_IDENT_LOG	7 min	.23 sec	1871 x
Select SCN_CALL_PARTY_EXTDATA_LOG	6.75 min	.47 sec	868 x

The Kübler-Ross grief cycle



Exposure to Exadata

enkitec

Questions?

Contact Information : Kerry Osborne

kerry.osborne@enkitec.com

kerryosborne.oracle-guy.com

www.enkitec.com

ORACLE Platinum Partner
Specialized
Exadata

A Member Of The Inc. 5000

DALLAS 100
Fastest Growing Companies in Dallas