Oracle Data Warehousing

Laying Foundations for High Performance Data Warehouses on Oracle

Jb.sastry@ge.com
State of Literature on DW Performance

Still searching…………..

High degree of customization

Divergent technology

Unique Needs

We Must Build Anecdotal Libraries
Classic DW Struggles

- Dimensional Modeling vs. ER
- ETL vs. End-User priorities
- Summaries - Nesting vs. de-normalization
- Indexes- Bitmaps vs. B-Trees
DM Needs Star Transformation

JOINS, JOINS and MORE JOINS
Selecting Partition Keys

Which single Key?

ETL window (Partition Exchange on Load Date)
Or
Biggest Queries (Store once use many times principle)

Smartkey temptations
(Huge value but apply filters on FACT instead of Dims)
Considerations for Pre-Aggregating

The Two dimensional SAS mindset vs. Over hyped OLAP

Should Oracle be a dumb data repository?

Needs high speed JOINs
(more efficient than star transforms and bitmap join indexes)

Needs effortless data Transposing (Scalability of MODEL clause)

If not, can it scale for complex analysis??
Indexing Choices

Bitmaps conducive to Star transformations
(They do not scale on DML)

Bitmaps – the death spiral (Tim Gorman)

GLOBAL vs. LOCAL Index choices
(GLOBAL indexes for the addl partitioning keys)
Design Factors

Logical Considerations

Physical Considerations
Influential Factors

Timeliness
Levels of aggregation
Types of usage – Ad hoc, drilldown vs. canned
Execute on demand for volatile objects vs. scheduled
Ratio of Power users to the normal ones
Nature of Hierarchies of Dimensions (Agility)
Data load frequency and the volume to be processed
Keeping a low ETL Window

CHUNK Sized ETL
Partitioning PLUS
Data Volumes
Re-invent ETL/Staging
CHUNK Sized ETL

Daily Load Volume → BULK ETL → CHUNK ETL
Chunking Loads

1. Minimize impact to online users
2. Optimize data volume per chunk

A 40% gain realized on a 5Mill/Day operation on 50K chunks!!!
Art of Staging

Diagram:
- Legacy
- Staging
- Data Mart Facts/s
- ODS Structures
- Cubes
- Aggregates
SCDs vs. RCDs

ETL on Oracle is not conducive to UPDATEs

RCDs are a huge problem to keep ETL Window down

Wage battles
Upfront at Dim
Modeling time

Creative Process Designs
‘LAZY-UPDATE’ TRICK

UPDATE a Weekly_DEL Flag for updated records

ADD as NEW records

Use a VIEW to filter out ‘deleted’ records

Weekly house keeping

Should this be a product development suggestion to Oracle!?
An Quick Intro to Hybrid Systems

A split FACT approach
Hybridization of DWs – Create Sub-FACTs

Legacy data

Batch Loads

TX FACT

TB FACT - GF

TB FACT - ADJ

TB FACT Union

User Interface

JV Master
Hybridization of DWs – Unified FACT-VIEWS
‘Fashioning’ Data Usage

# of Users Accessing Summaries at each level

Data Volume Accessed At each Level of Aggregation
10g Features of Interest

Ora_rowscns
Row level System Commit Number (SCN) and TIMESTAMP

Robust Oracle Streams for change propagation
This topic requires a major discussion in itself

Rename Tablespaces for TTS
This is a very useful feature for bulk moving data segments across databases

Sorted HASH Cluster Tables
Tables can be stored in hash clusters after pre-sorting on selected columns
10g Features

External Tables for Read/write
External tables now can be read from and written to in parallel

User defined metrics and tracking
User defined metrics on production data to trigger alerts, messages and events

HTML DB
Light weight, operational reports from log tables to SysMan portals

RCG Enhancements
Resource Consumer Groups now monitor idle time and trigger session terminations
10g Features

Oracle OLAP
Waiting to hear the pros and cons of this database embedded Express Engine

SQL – MODEL clause
Meets common spreadsheet-like transposing needs

Job Scheduler
DBMS_JOB interface has been used in evolving this scheduler

Data Pump
This enhancement will probably revolutionize the ETL architectures like never before.
Lessons Learned

Move from ER/DM Puritanism to a practical MIX

Platform independence is an impractical dream

Never hesitate to customize
Be Creative

Manage performance expectations Adaptively

Usage of DWs is changing rapidly

The BEST is yet to come

No packaged solution heaven
Q & A

Jb.sastry@ge.com