A person wearing a light blue denim shirt over an orange t-shirt is harvesting corn in a field. They are wearing yellow gloves and holding a wooden basket filled with green corn cobs. The background is a blurred green field under a bright sky.

How Oracle Exadata and Database In-Memory Power Precision Marketing at General Mills

Joshua Moe - Solution Architect
Mark Donohue - DBA

October 28, 2015



Agenda

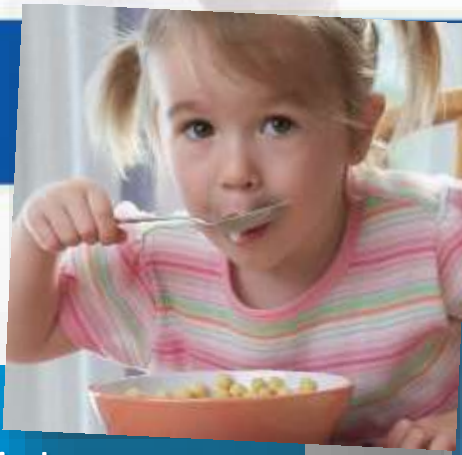


- General Mills History
- Business Case
- Capabilities delivered
- Oracle Exadata Experience
- Oracle Database In-Memory Experience
- Lessons Learned

Our history



Who we are Today



- One of the world's largest food companies
- Products marketed in more than 100 countries on six continents
- 42,000 employees
- \$17.6 billion in fiscal 2015 net sales*





We serve the
WORLD by making
& FOOD
people LOVE



GENERAL MILLS

Business Case



- In-Source existing Marketing Data Warehouse
- Create a Consumer-Centric Data Warehouse
- Capture consumer driven activity
 - Marketing Campaigns
 - Reporting
- Better Understand Consumer behavior
 - Personalized Campaigns
 - Content Personalization

Original Solution Pain Points



- Poor Vendor Response
 - Changes were slow and expensive to implement
 - Limited control over our data
- Performance insufficient to make business-driving marketing decisions:
 - While on Exadata, schema design was inefficient
 - Sizing/configuration sub-optimal
 - Server was oversubscribed with many other customer databases
- Business users demanded access to full dataset
- Difficult to add more data sets
- Desire to support more ad-hoc queries

New Capabilities Desired



- Improved overall database performance
- Effective ad-hoc querying
- Full control of our data
- Add more data sources
- Near real-time data ingestion, instead of batch-only
- Develop more effective and efficient marketing campaigns
- Identify consumer preferences
- Statistics-based processing to identify trends for General Mills brands

Original Configuration




Bettycrocker.com
Pillsbury.com
Boxtops4education.com
etc



Web Activity
Email Newsletter
Campaign Activity



Internal
Transactional DB's



Limited Vendor Data

Hosted Exadata X2-2
Oracle 11gR2



Limited Data Retention due
to data volume constraints



Marketing Campaigns



Reporting & Analytics

Final Configuration




Bettycrocker.com
Pillsbury.com
Boxtops4education.com
etc



Web Activity

Email Newsletter
Campaign Activity



Internal
Transactional DB's

3rd Party
& Vendor Data



Additional data
& Scoring



Internal Exadata X4-2 ¼ Rack
Oracle 12c
150 GB IN-MEMORY per node



Marketing Campaigns



Reporting & Analytics



3rd Party Tools



Experiments

Capabilities Delivered



- Dedicated development team
- Fast turn-around for changes and adding of new data
- Users querying ad-hoc due to big performance boost
- More complete data model
- Greater data retention
- Near real-time loading
- Foundation of next-generation Marketing capabilities
- More projects planned that were never possible before

Technical Solution



- Migrated to Oracle Database 12c on Exadata X4-2 ¼ Rack (in-house)
 - 2 RAC nodes with 24 core, 512 GB RAM per node
 - 3 storage cells with 144 TB total
- Version: Oracle 12.1.0.2 on Oracle Linux 5
- Oracle Database In-Memory
- Advanced (OLTP) Compression
- Near real-time data loads (vs overnight batch)
- Triple the volume of Data
 - More data (5 TB and growing), but better compression
 - Several fact tables over 10 billion
- **50X** improvement on reports and ad-hoc queries
 - Reports that took **days now return in less than 1 hour**
- Production as of March 2015

Our Exadata Experience



- First Exadata for General Mills
- Lives up to the hype for Data Warehouse/Mart
- Drastically shortened development cycle
- Minimal performance tuning needed
- Minimal indexing required
- Storage indexes/caching really work
- Excellent Oracle SCP (Strategic Customer Program) engagement – very helpful

Exadata Tips



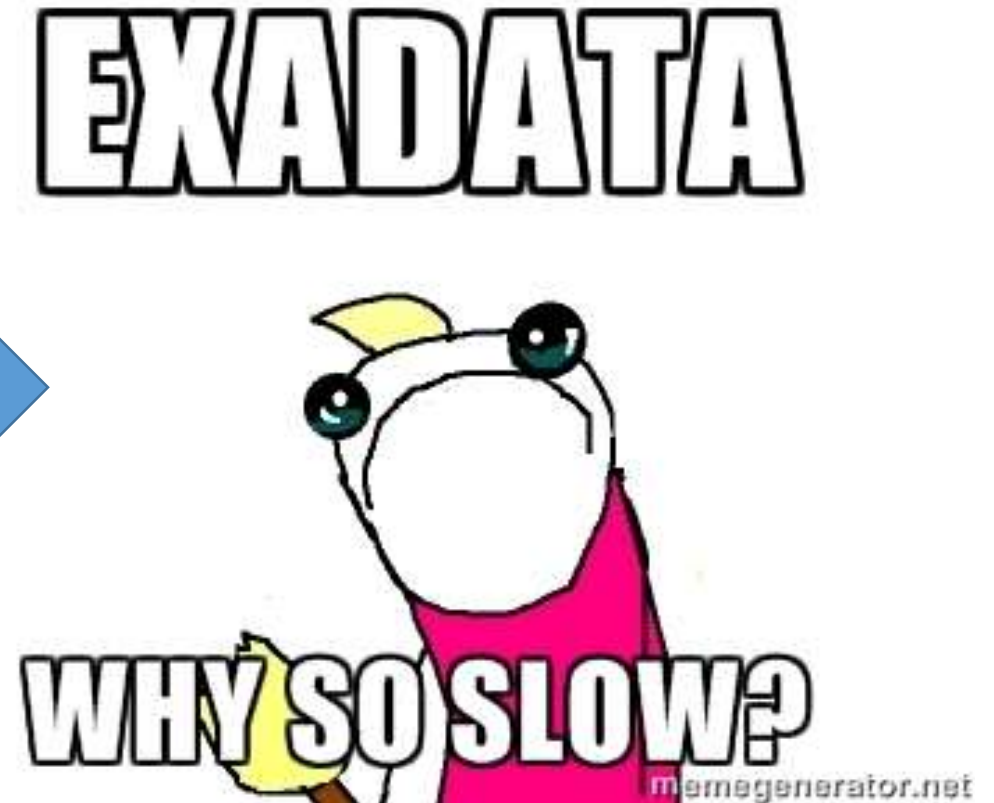
- Design for success
 - Partitioning – HASH or RANGE-HASH, not too fine-grained
 - Compression – use HCC whenever possible
 - Parallelism – use Auto DOP, Resource Manager if required
- Use good general data mart design practices
- Partition-swap loading, where possible
- HCC for bulk-load only (insert /*+append*/)
- OLTP compression for everything else

Parallelism Challenges



- Tuning Auto DOP for mixed workload (app, batch, ad-hoc)
- When queries don't fetch all rows
 - Sessions holds on to parallel worker processes
 - Chokes out parallel workload from other sessions
 - Primarily ad-hoc queries
- End user tools:
 - Oracle SQL Developer
 - PL/SQL Developer
 - SQL Workbench

Parallelism Challenges

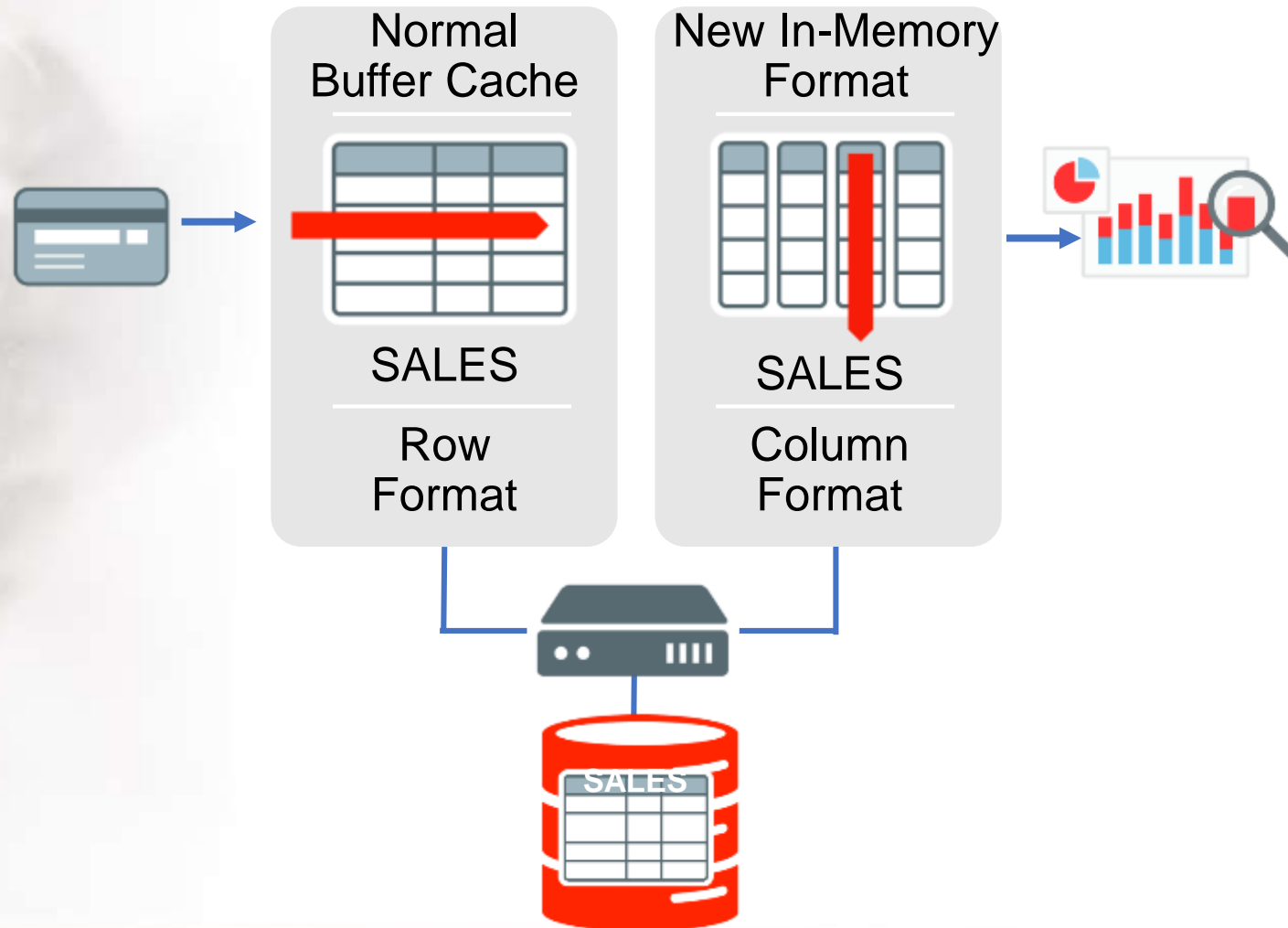


Parallelism Tips



- AUTO DOP needed for In-Memory
 - Recommended for Exadata in general
- PARALLEL_DEGREE_LIMIT alleviated some queuing
- Resource Manager gave better control
- User Training! Close session, or run “select 1 from dual”
- Higher degree = more PGA
 - 2 GB per process limit, by default
 - Parallelism helps “unlock” more PGA per query

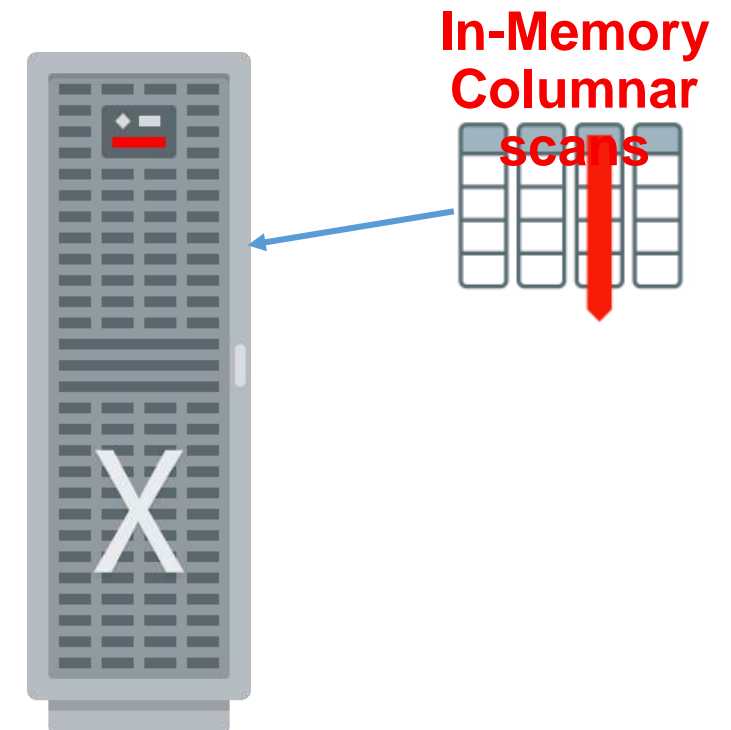
Database In-Memory



Database In-Memory



- In-Memory column store on DB nodes only
- Each node has it's own In-Memory column store
- Table data distributed across nodes
 - Requires Parallel Execution
- Similar benefits to Exadata smart scan
 - Access only the columns need for query
 - Storage indexes prune out unnecessary data
 - Scans & filters data in compressed format



In-Memory Challenges



- Partitioned table with 1 partition on only 1 RAC node
 - ALTER TABLE abc.xyz MODIFY PARTITION PART_1 INMEMORY **DISTRIBUTE BY ROWID RANGE;**
- Still a new product
 - Continued enhancements and bug fixes
 - Partnership with Oracle very beneficial for feedback on improvements
- If join/sort/aggregate spills to TEMP, performance suffers
 - Allocate enough PGA and parallel processes

In-Memory Tips



- HASH partitioning for even distribution across RAC nodes
- Dedicate enough memory to PGA
 - PGA_AGGREGATE_TARGET
- Understand your workload
 - Ad-hoc, batch, applications, etc...
 - Use your in-memory area most effectively
- SQL Monitor reports very helpful for troubleshooting
- In-Memory advisor
 - New advisor from Oracle
 - Recommends which tables benefit most

Slowly Changing Dimensions



- We used SCD “Type 2” most heavily
 - Start & end dates for records
 - History and current stored in same table
- Advantages
 - Precise historical data easily accessible
 - Current state available through simple view (WHERE current_fg='Y')
- Disadvantages
 - Complex merge/update loads
 - More complicated when transactions arrive out of chronological order
 - HCC not feasible (OLTP compression instead)
 - Performance on both loads and queries can suffer
- Recommendation
 - Use for small dimension tables
 - When data volume too high, look for alternate ways of storing history

Other ETL Learnings



- Understand your ETL tools
 - Had poor performance with our chosen ETL tool
 - Reverted to using PL/SQL, external tables, and database links
 - Better bulk-load performance
 - Partition swapping
 - More precise control over loading processes (like SCD)
- Messaging sources where order not guaranteed
 - Duplicates possible when using guaranteed delivery
 - Can receive data out-of-order
 - Re-ordering records complicates loads

Other Development Experiences



- 12c JSON parsing functionality very useful
 - Loading large JSON files via external tables
- 12c online bulk load stats gathering
 - Saves time and simplifies load jobs
- HCC works great, but not for updates
- Parallel DML best only for large loads
- Make sure sequences aren't NOCACHE
- Do a POC

In-Memory vs Exadata



- Depends on the query and workload
 - With complex aggregate queries over all data, performance similar
 - Queries with selective filter predicates or joins, In-Memory faster
 - In-Memory scans very fast, but storage cells scan extremely fast too
 - Both use similar “tricks”: smart scans, bloom filters, vector group by
- In-Memory excels with higher concurrent workload
 - We expect to have higher concurrency in the future
- Understand your bottlenecks
- Both significantly faster than regular Oracle database

Future Considerations



- Leverage In-Memory for some new workloads
 - Potential solution for planned higher concurrency
 - More standardized reporting
- Extend use of newer Oracle DW features
 - Attribute Clustering
 - Zone Maps
 - Approximate Count Distinct where applicable
- Materialized View out-of-place refresh promising
 - Does the refresh outside and then swapping
 - Minimizes downtime for huge mviews

Conclusion



- Exadata and In-Memory a solid platform for success
- Completed a large, complex project in under 1 year
- Far exceeded performance expectations for users
- Excellent support from Oracle (SCP, Platinum Support)
- Confidently delivering new digital marketing capabilities

Thank You!



Questions?

