Data-and-Compute Intensive Processing: Middle-tier or Database? Trade-Offs and Case Study

Kuassi Mensah – Marcelo Ochoa
Oracle
The following is intended to outline our general product direction. It is intended for information purposes only, and may not be incorporated into any contract. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions. The development, release, and timing of any features or functionality described for Oracle’s products remains at the sole discretion of Oracle.
Agenda

- New Data-and-Compute Intensive Applications
- Middle-tier Approaches
- Database Approaches
- Case Study: Lucene Domain Index
New Data-and-Compute Intensive Applications
New Data-and-Compute Intensive Applications *

- Search
- Data Intensive High Volume Mission Critical Applications
- Social Web Computing
  - Database-backed Web Sites
- eXtreme Transaction Processing
- Astronomy
- CyberSecurity

*Inspired from
Common Challenges

• Implementing Complex Algorithms
  • Bloom Filter, N-Gram
• Processing Extremely Large Databases with Very Low Latency
• Guarantee Unlimited Scalability
  • Capacity on Demand
• Overcome Massive Network Roundtrips to the Database
• Continuous Availability

Some of these challenges are still subject to research but How do you address these, today?
Designing Data-and-Compute Intensive Applications

- Database Approaches
- Data Intensive Problems
- Data-and-Compute Intensive Problems
- Compute Intensive Problems
- Non Challenging Problems
- Middle-tier Approaches
Middle-tier Approaches
Middle-tier Approaches

Pros
- Can Scale up to 1000s Processing Nodes
- Caching Reduces Massive Network Roundtrips to the Database
- Perfect for Compute Intensive Problems
- Good for Data-and-Compute intensive

Cons
- Network Roundtrip to the Database is EXPENSIVE!!!
- Very Large Databases (TeraBytes) may not be CACHE-ABLE
How to Avoid Massive Data Shipping

• Client Result Cache
• Data Grid
• In Memory Database
Client Query Result Cache

- Database Server Configuration (init.ora)
  - `client_result_cache_size=200M`
  - `client_result_cache_lag=5000`
- Client Configuration (sqlnet.ora)
  - `OCI_QUERY_CACHE_SIZE=200M`
  - `OCI_QUERY_CACHE_MAXROWS=20`
- Hints for Caching the Result Set
  - `select /*+ result_cache */ * from employees`
- Alternatively without code change
  - `alter table emp result_cache (mode force);`
- The Cache is proactively invalidated upon changes to the Result Set on the server side
- Available with OCI, PHP, Ruby, ODP.Net, JDBC-OCI, ODBC
Oracle Coherence as Data Broker

- Oracle Coherence brokers Data Supply with Data Demand
- Scale out Data Grid in middle tier using commodity hardware

Ever Expanding Universe of Users

Web Servers

Application Servers

Java Objects

Data Demand

Data Supply

Data Sources
Oracle In-Memory Database Cache
Optimized for Real-Time Applications

- Delivers *real-time responsiveness* and very high throughput

- Deployed in the application tier, as in-memory database cache for Oracle Database
Top 5 Reasons Customers Choose In-Memory Database Cache

- Predictable and bounded response time
  - Microseconds to sub-milliseconds
  - In-process performance (network round-trips not required)
- High Availability, data persistence and recoverability
- Standard SQL relational model, standard APIs
  - No rewrite of business logic or interface
- Caching Oracle database tables with automatic data synchronization
- Very high transaction rate
  - A direct result of low response time
Database Approaches
Database Approaches

Pros

• Business Logic Runs Directly in the Database and Process Data *In Situ*,
• Eliminates Network Roundtrip
• Code Shipping in Cheap
• Less Moving Parts
• Perfect for Data-Intensive Problems
• Good for Data-and-Compute intensive

Cons

• Few Languages Run Natively in the Database
• Database Nodes Can Scale up to 100s Nodes
Languages that Run Directly in the Database

- PL/SQL
  - APEX
- Java-in-the-Database
  - Java SE, JDBC, SQLJ
Java in the Database

Java SE/EE

Java

JDBC Calls

SQL

Java in the Database

Any Language

Stored Procedure Call

Java

JDBC Calls

SQL

Typically 10+ times Faster!
Java in the Database: What For

- Trigger-based Notification System using RMI
- Secure Credit-Card Processing using JSSE
- Custom Alert applications that monitor business data
- Sending emails with attachment from within the database
- Produce PDF files from Result Set
- Execute external OS commands and external procedures
- Implement Md5 CRC
- Publish Repository Content to Portal
- Portable Logistic Applications
- Implement Parsers for various File Formats (txt, zip, xml, binary)
- Implement Image Transformation and Format Conversion (GIF, PNG, JPEG, etc)
- Implement database-resident Content Management System
- HTTP Call-Out
- JDBC Call-Out
- RMI Call-Out to SAP
- Web Services Call-Out
- Messaging across Tiers
- RESTful Database Web Services*
- Lucene Domain Index*

* http://marceloochoa.blogspot.com/
Database Approach

Case Study: Lucene Domain Index

(companion presentation)
OracleJVM and Java Stored Procedures

In this release, OracleJVM goes JDK 1.5, gets a JXT compiler (goodbye to NCONP), a JDI-like interface, JWS, a closure tool, and enhancements to utilities.

To learn more, read "Java and Web Services Developer's Perspective on Oracle Database: 11g" white paper.

Doc & Blog
- Frequently Asked Questions: Oracle JVM and Java Stored Procedures
- Java Developer's Guide
- JJU online Database programming

OOW 2007 Presentations
- "Getting the Most from the Java VM in the Oracle Database 11g"
- "Java Stored Procedures or Java Middleware: Trade Offs, Techniques and Tips from Database and Middleware teams"

Technical Whitepapers
- "Executing operating system commands from PL/SQL" (PDF)
- "Java and Web Services Developer's Perspective on Oracle Database 11g" (PDF)
- "Java and Web Services Developer’s Perspective on Oracle Database 11g" (PDF)
- "Reducing your Costs and Extending your Database with Java in the Oracle Database" (PDF)
- "Unlock the Power of Java Stored Procedures" (PDF)
- "Java in the Oracle Database 11g: Work Customer Case Studies" (PDF)
- "Developing Java Stored Procedures in Java" (PDF)

Technical Articles
- "Parametrized custom Oracle database Error Messages" July 2007
- "Securing Java Execution: A DBA’s Perspective on Oracle JVM Security Mechanisms" July 2003
- "Simplify with Java Stored Procedures" Jan 2003
- "Database-assisted Web Publishing using Java Stored Procedure"