Oracle9i Data Warehouse Review

Robert F. Edwards
Dulcian, Inc.

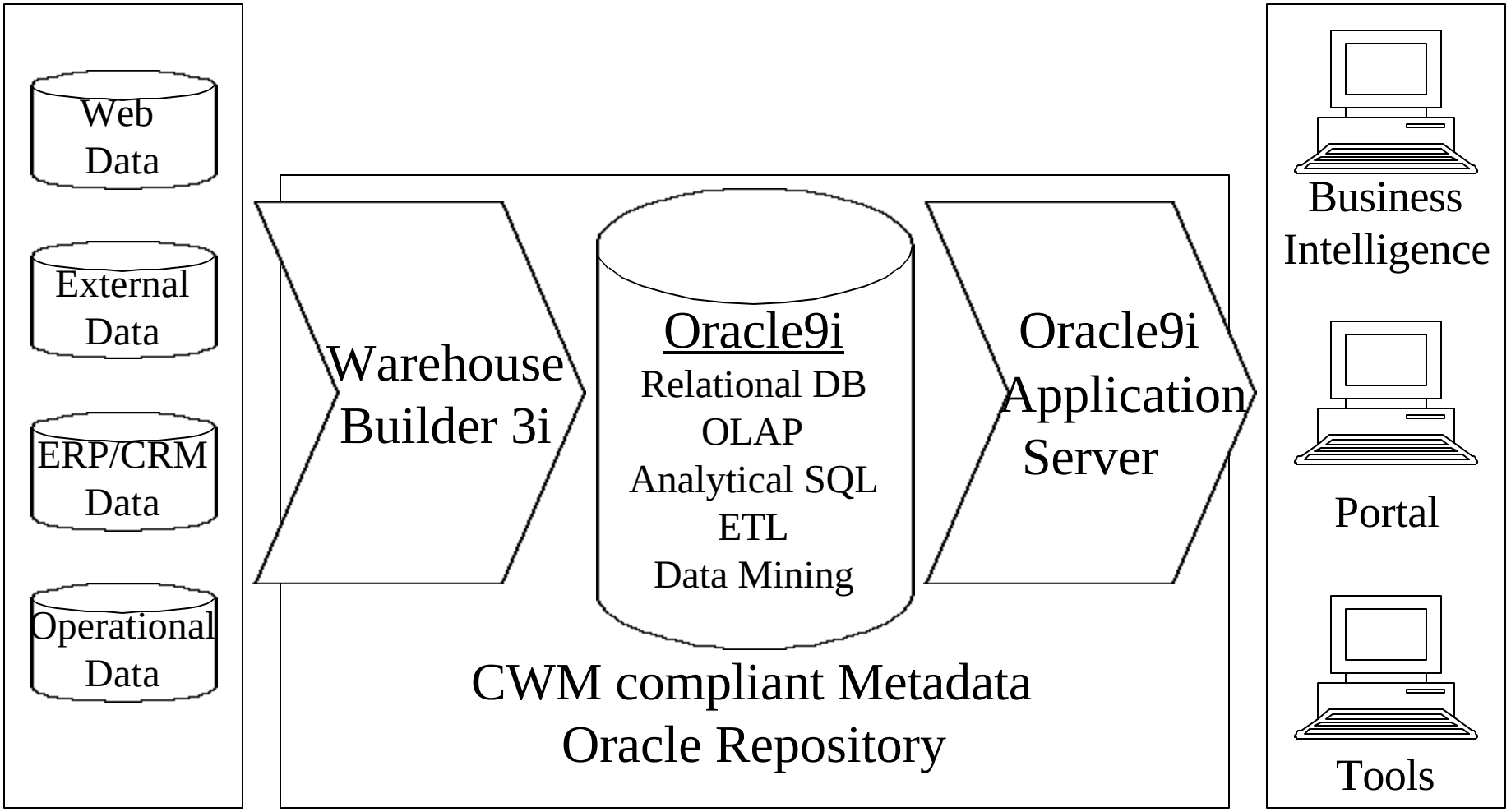
Agenda

- Oracle9i Server
- OLAP Server
- Analytical SQL
- Data Mining
- ETL
- Warehouse Builder 3i

Oracle 9i Server Overview

- 9i Server = Data Warehouse Platform
- DW Core Requirements
- Oracle 9i Server Components
  - Relational Database
  - OLAP Services – *Added in 9i*
  - Analytical SQL – *New functions added in 9i*
  - ETL Infrastructure – *Added in 9i*
  - Data Mining – *Added in 9i*
- Core of E-Business Intelligence Platform

E-Business Intelligence Platform



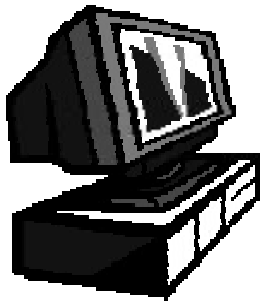
Source: Oracle

DW Core Requirements

- Performance – Most important requirement
  - Query Response Time - Critical
- Scalability – Growth: Data and Users
  - Data Growth – very large data volumes
  - Additional Users – many users supported
- Manageability – Simple to maintain
  - Growth does not require additional resources

Performance Enhancements

- Bitmap Join Index – Spans multiple tables
- Enhanced Materialized Views - Query Rewrite
- Full Outer Joins
- With Clause – Complex Queries & Sub-queries
- Adaptive Direct I/O Operations – Dynamic I/O
- Automatic Memory Tuning – Runtime Memory



Scalability Enhancements

- List Partitioning
  - List of discrete values
  - Precise control over data loaded into partitions
- Parallel Queries
  - Finer grain of parallel query distribution
  - Enhanced dynamic load balancing
  - Inter-node parallel query improved
    - Clustered systems
    - MPP platforms

Manageability Enhancements

- Appropriate Resource Allocation to Queries
- Maximization of Throughput
- DBAs and Users can view status of active jobs
- DBA Specified Abort and Re-Queuing
- DB Resource Manager (introduced in Oracle8i)
  - Resource Consumer Groups (RCG)
  - Limit Active Sessions per RCG
  - Query Governing per RCG
  - Criteria-based Change of RCG

Oracle9i Server Overview

- 9i Integrates four data engines:
  - Relational Database – *Always here!*
  - OLAP Server – *previously just Oracle Express*
  - ETL Infrastructure – *Custom built or third party tool*
  - Data Mining – *previously just Oracle Darwin*
- Analytical SQL Functions
- Java APIs – build complex transforms and queries
- Enhancements Open to Software Vendors
  - Implemented in SQL
- Tight integration with Warehouse Builder

OLAP Overview

- On-Line Analytical Processing
  - Scalable, High Performance Calculation Engine
  - End User Business Intelligence Analysis Tool
  - High-level aggregates and customized data needs

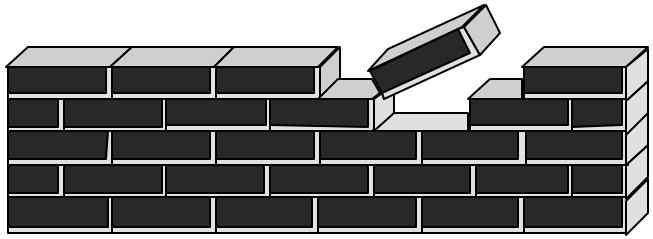
- OLAP on Oracle9i Server
  - All data resides on the relational database
  - Most queries satisfied with Analytic SQL functions
  - Java API supplements SQL functions
  - Administration by 9i Enterprise Manager

OLAP Server

- OLAP Services
  - Architecture and Support
  - Java OLAP API
  - Java Development Environment
  - Express-to-9i Migration Plan
  - Analytical Functions
  - OLAP SQL Functions

Architecture and Support

- Fully Integrated with 9i database
- Relational Database is OLAP Data Store
  - Relational - not a proprietary array structure
  - Scalable - very large volumes (Terabytes) supported
  - Lower Overhead than Multi-Dimensional DB
- Administration with Oracle Enterprise Manager
- Data Modeling with 9i Dictionary and Meta Data
- High Availability and Security support by 9i



Java OLAP API

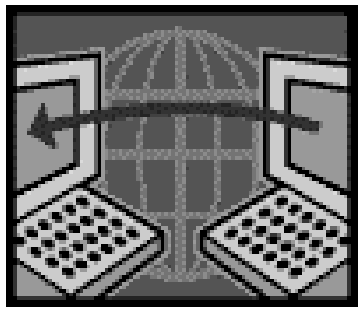
- Java OLAP API provides Object Oriented environment for building complex analytical queries
  - Multi-dimensional object model
  - Full range of functions to support analytical Apps
  - Supports asymmetrical queries (nested rankings)
  - Support for calculated dimension members and measures (facts)
  - Incremental query construction
  - Multi-dimensional cursors
- Supplements analytical SQL functions

Development Environment

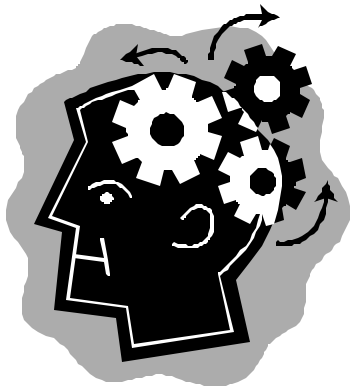
- Business Intelligence (BI) Beans
  - Analytically aware building blocks -
    - Tables, cross tabs, graphs, query & calc builders
- JDeveloper 9i environment
- Runtime Repository for Analytical Objects
  - Supports collaborative analysis
- Data Dictionary and Oracle Repository meta data
  - Fully describe multidimensional analytical model
- Meta Data administered by OEM and OWB

OLAP Migration Plan

- Express and DB converge: OLAP Services
- Support continues for Express 6.x
  - Maintenance & compatibility with 9i DB
  - Financial and Sales Analyzer continue support
- New features added to 9i OLAP Services
- New Analysis and Planning app for 9i
- Migration tools will be available



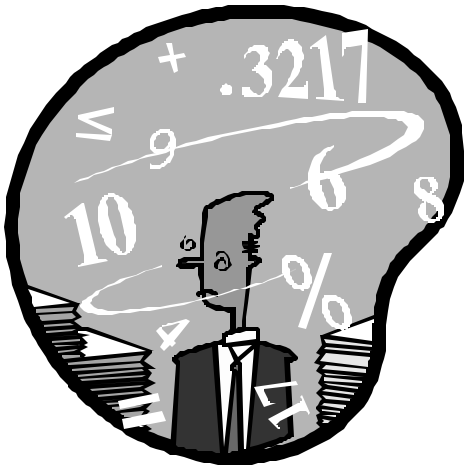
Analytical SQL



- Introduced in Oracle8/8i
- Additional functions in Oracle9i
- Support for OLAP
  - Aggregations, including Grouping functions
- Support for Data Mining
  - Sampling and Statistics
- Most functions added to ANSI SQL

Analytical Functions – 8i

- Introduced with Oracle8i:
  - Ranking
  - Moving Window Aggregates
  - Period over Period Comparisons
  - Ratio to Report
  - Statistical Functions –
    - Linear Regression
    - Correlations



Analytical Functions – 9i

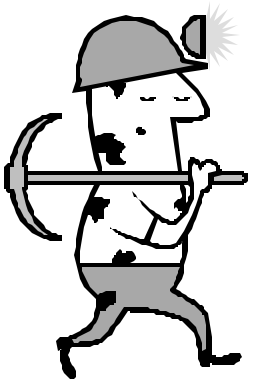
- Introduced with Oracle9i:
  - Inverse Percentiles
  - Hypothetical Rank and Distributions
  - Histograms
  - First / Last Aggregates
  - Grouping Sets
  - Composite Group By (column sets)

OLAP SQL Functions

- Aggregation Functions
  - 8.0 – Cube and Rollup (Group By)
  - 9i – Grouping Sets
    - Grouping Sets – Sets of aggregation
    - Concatenated Grouping Sets – Set Combinations (like Cartesian product)
    - Composite Columns in Group By – Column sets are treated like a single unit

Data Mining

- Data Mining
  - Discover hidden patterns and relationships
  - Machine-learning Mining Algorithms
  - Scoring Models
  - Java API Interface
  - Available in 9i EE



Data Mining

- Oracle Data Mining
- Machine-learning Algorithms
- Model-based Scoring
- Java Functions

Oracle Data Mining

- Oracle buys Darwin and Thinking Machines
- Initially Unix-based and stand alone, flat file I/P
- Data Mining Suite (Darwin 3.7)
  - Accesses database, network able, wizards introduced
- Integrated with CRM 11i – Marketing application
- Architecture Changes -
  - Shift from stand alone to database platform
  - Access to database and network
  - Unix Interface to Windows GUI
  - Access via Web browser

Data Mining Algorithms

- Machine-learning Data Analysis techniques
- Classification and Prediction
  - Naïve Bayes algorithm – supervised learning
  - Binary and Multiple-class outcomes
- Association Rules
  - Detect co-occurring events hidden in data
  - Unsupervised learning
- Results of analysis are built as models used to score new data

Model Scoring

- Scoring – Process of Predicting Outcomes
- New data scored using Naïve Bayes models
  - Batch – applied to a table of records
    - Prediction stored in another table
  - On-demand – applied to single record
    - Prediction returned to calling application

Java Functions

- Java API is used for development of all data mining functions
  - Data preparation
  - Model Building
  - Model-based Scoring
- Java Data Mining (JDM)
  - Emerging industry data mining standard

ETL Overview

- Extraction, Transformation & Loading
  - Extracts source data from database tables and flat files of operational applications
  - Transforms (converts) source data to a consistent format usable by the data warehouse and BI apps
  - Loads transformed data to the warehouse tables
  - Builds aggregates and customized data requirements for data marts and OLAP servers
- Formerly had to be built in-house using PL/SQL or other languages, or use a third party tool
- Data Migration consumed extensive project resources
- Warehouse Builder is Oracle's ETL development tool

Oracle 9i ETL

- Scalable ETL infrastructure
- Major new features / enhancements
  - Change Data Capture
  - External File Access
  - Upserts and Multi-table Inserts
  - Table Functions
  - Transportable Tablespaces
- Leveraged by Warehouse Builder 3i design and deployment functions

ETL Enhancements

- Change Data Capture – detects changes in source data
  - Replication and LogMiner
- External Tables – access flat files with SQL like DB tables, without staging
- Upserts – single step update and insert
- Multi-table Inserts – single step inserts to multiple tables

ETL Enhancements (2)

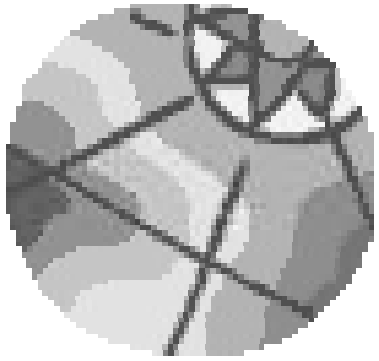
- Table Functions – complex transformations
  - Complex functionality not available with SQL alone
  - Implemented in PL/SQL, Java, C, C++
  - Pipelined, fully parallelized, and scalable
- Transportable Tablespaces – accommodates multiple block sizes, eases data movement from OLTP to warehouse
  - Block size can be defined for each tablespace
  - OLTP – 4-8 KB blocks
  - Data Warehouse – 16-32 KB blocks

Warehouse Builder Overview

- Warehouse Builder is Oracle's DW design and ETL development tool
- Design, deploy and manage warehouses, data marts, and business intelligence apps
- Tightly integrated with Oracle 9i Server
- Maintains Oracle Repository (Meta Data)
- Leverages Oracle9i's ETL infrastructure

Warehouse Builder 3i

- OWB 3i is a major release
- New Mapping and Transformation paradigm
- New Mapping Architecture
- Major Enhancements



3i Mapping and Transform Features

- Join Operator – join multiple sources
- Split Operator – output to multiple targets
- Multi-stage Transformation – pipeline SQL expr.
- Multiple Targets – multiple tables in single pass
- Key Lookup – lookup and replace values
- Set Operator – support for Union, Minus, Intersect
- Inline Expressions – SQL expression in Select
- Incremental Code Generation – generate and view code up to any point within the data flow

3i Mapping Architecture

- Previous architecture (OWB 2.x)
  - Two level maps – High-level and Detailed
  - Source and target dependency for transform maps
- New architecture
  - Connectivity through parameters
  - Transforms are independent of sources and targets
- Operator: output is a subset of input rows
  - AGGREGATES and FILTERS
- Transform: output is equal to the input rows
  - Standard SQL functions, e.g. TO_CHAR

3i Enhanced Features

- Enhanced flat file Integrator
- Meta Data Reconciliation –source and target DBs
- Upgrade or drop warehouse schema
- Complete inline SQL expression builder
- Multiple-user support on OWB repository
- International Name and Address Cleansing
- Reverse Engineering PL/SQL into repository
- Total Data Warehousing Management
- Significant performance enhancements

Summary

- 9i Server: A True Data Warehouse Platform
  - Integrated DB, OLAP, ETL and Data Mining engines
- Core of E-Business Intelligence Platform
  - Warehouse Builder / Oracle9i / 9i Application Server
- All data stored and processed on the relational DB
  - ETL support for DW and Data Mining
  - Analytical SQL and Java API support OLAP and DM
  - Internet/Web enabled, with JDeveloper environment
- Warehouse Builder 3i
  - New Mapping and Transform paradigm
  - Major Enhancements

Contact Information

Robert F. Edwards

Dulcian, Inc.

732-744-1116

redwards@dulcian.com

www.dulcian.com