

DULCIAN


Oracle9i Data Warehouse Review

Robert F. Edwards
Dulcian, Inc.

DULCIAN

Agenda


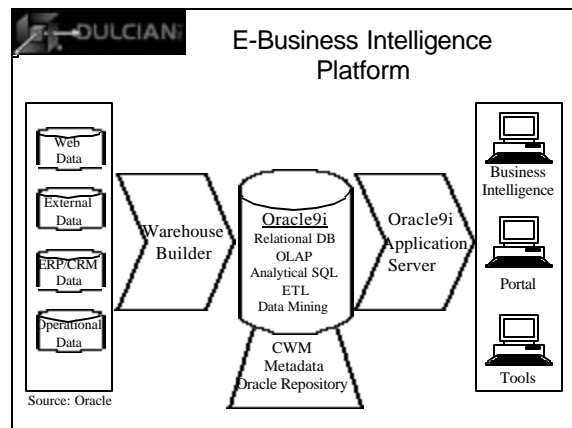
- ◆ Oracle9i Server
- ◆ OLAP Server
- ◆ Analytical SQL
- ◆ Data Mining
- ◆ ETL Infrastructure
- ◆ 9i Warehouse Builder



DULCIAN

Oracle 9i Server Overview

- ◆ 9i Server: “True Data Warehouse Platform”
- ◆ Core of E-Business Intelligence Platform
- ◆ 9i Exceeds DW Core Requirements
- ◆ Oracle 9i Server Components
 - > Relational Database
 - > OLAP Services – *New in 9i*
 - > Analytical SQL – *New functions added*
 - > Data Mining – *New in 9i*
 - > ETL Infrastructure – *New in 9i*
- ◆ 9i Warehouse Builder

DULCIAN


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.

DULCIAN


Performance Enhancements

- ◆ Bitmap Join Index – Spans multiple tables
- ◆ Materialized Views – Query Rewrite, Fast Refresh
- ◆ Full Outer Joins – ANSI syntax (left/right/full)
- ◆ With Clause – Named sub-queries
- ◆ Adaptive Direct I/O Operations – Dynamic I/O
- ◆ Automatic Memory Tuning – Runtime Memory



DULCIAN Scalability Enhancements

- ◆ List Partitioning
 - List of discrete values allowed, e.g.,
 - 'Canada', 'Mexico', 'United States'
 - More precise control over data loaded into partitions
- ◆ Parallel Queries Improved
 - Finer grain of parallel query distribution
 - Enhanced dynamic load balancing



DULCIAN Manageability Enhancements

- ◆ DB Resource Manager – new controls on RCGs
- ◆ Enhanced Statistics – DBMS_STATS
- ◆ Execution Plans in Cache – view while running
- ◆ Summary Advisor – recommendations based on schema and workload history
- ◆ Query Rewrite – can force queries to use materialized views, or abort.

DULCIAN OLAP Overview


- ◆ Oracle9i OLAP Server
 - All data and OLAP functions now reside on the Oracle 9i server.
 - Analytic SQL Functions support most queries
 - Java APIs support building of complex queries
 - Administration by 9i Enterprise Manager
 - Priced Option

DULCIAN Application Support

- ◆ Four basic Business Intelligence applications supported:
 - Reports – static, parameterized reports
 - Ad-hoc query and Reporting – on-line queries
 - Multidimensional Analysis – complex queries
 - Planning – budgets, financial analysis, forecasts
- ◆ New trends in Business Intelligence
 - Collaborative analysis and planning
 - Convergence of operational and BI applications

DULCIAN Architecture

- ◆ OLAP Fully Integrated with 9i database
- ◆ Two data structures types -
 - Relational database is basic data store
 - Highly Scalable - very large data sets supported
 - Lower Overhead than Multi-Dimensional DB
 - Materialized Views used to store aggregates
 - Analytical Workspace (cube)
 - Used for specialized analytical calculations



DULCIAN Administration


- ◆ Administration by Enterprise Manager
 - Instance, Session, Configuration management
 - Performance Monitoring
 - Job Scheduling
- ◆ OEM OLAP Tool
 - Data Modeling – full multidimensional model
 - Build Materialized Views
- ◆ Support
 - High Availability / Security / Backup / Recovery

DULCIAN Scalability and Performance

- ◆ 9i OLAP scales well for very large data sets.
- ◆ Performance vs. MD OLAP server
 - Under 250 GB – MD is faster, but both have sub-second response times – not noticeable
 - Over 250 GB – 9i OLAP faster than MD as size of data set increases
 - Execution plan optimized for complex queries and large data sets
- ◆ Materialized Views
 - Efficiently manage aggregated data

DULCIAN Java OLAP API

- ◆ Java OLAP API provides Object Oriented environment for building complex analytical queries
- ◆ Provides the OLAP interface to Oracle 9i
- ◆ Java API Components -
 - Metadata Provider
 - SQL Generator
 - Query Processor
 - Multidimensional Engine
- ◆ Analytic Workspace – special external data type
 - Used for certain predictive analytical calculations
 - Assist Express-to-9i OLAP data migration
 - Persistent data store




DULCIAN Development Environment

- ◆ JDeveloper 9i
- ◆ Business Intelligence (BI) Beans
 - Analytically aware building blocks -
 - Tables – provides row oriented view of data
 - Cross Tab – provides multidimensional view of data
 - Chart – provides graph view of data
 - Query Builder – select data in multidimensional context
 - Calculation Builder – define “custom measures”
- ◆ Beans Repository
 - Develop and store analytical objects on 9i database


DULCIAN 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 only
- ◆ New Analysis and Planning application for 9i - available in 2002
- ◆ Migration tools available




DULCIAN Analytical Functions – 8i

- ◆ Introduced in Oracle 8i:
 - Cube and Rollup (aggregation)
 - Lead / Lag
 - Ranking
 - Moving Window Averages (Moving Average, Cumulative Sum)
 - Period over Period Comparisons
 - Ratio to Report
 - Statistical Functions (Linear Regression, Correlations)




DULCIAN Analytical Functions – 9i

- ◆ Introduced with Oracle9i:
 - Inverse Percentiles
 - Hypothetical Rank and Distributions
 - Histograms
 - First / Last Aggregates
 - Grouping Sets (aggregation)
 - Concatenated Grouping Sets (agg.)
 - Composite Column in Group By (agg.)



DULCIAN Oracle Data Mining

- ◆ Oracle Darwin – acquired from Thinking Machines
 - Stand alone tool to DM engine embedded in database
 - Flat file to database and network accessible
 - Unix GUI to Windows, Web browser and Wizards
- ◆ Machine Learning Techniques
- ◆ Model-based Scoring
- ◆ Java API for Data Mining
- ◆ Priced option on 9i EE



DULCIAN Learning Techniques

- ◆ Supervised (Classification & Prediction)
 - Tool searches data for patterns and relationships between specified fields
 - Results built as models used to score new data
- ◆ Unsupervised (Association Rules)
 - Tool searches for associations and clusters without an objective
 - Commonly used on historical data

DULCIAN Models and Scoring

- ◆ Models are built using three equal sets of data:
 - Training – build initial model
 - Testing – first model refined with test sample
 - Evaluation – predictions made using refined model
- ◆ Scoring – Process of predicting outcomes, using models
- ◆ New data scored using models
 - Batch – applied to a table of records
 - Prediction stored in another table
 - On-demand – applied to single record
 - Prediction returned to calling application

DULCIAN Java API for Data Mining

- ◆ Java API is used for development of all data mining functions
 - Data preparation
 - Model building
 - Model-based scoring
- ◆ API is adaptable to different user groups
 - Complex settings exposed for advanced users
 - Automated for business users

DULCIAN ETL Features


- ◆ Change Data Capture – detects changes in source data
 - Synchronous – Real-time extraction
 - Asynchronous – Deferred extraction
- ◆ External Tables – access flat files as tables
- ◆ Merge (Upsert) – single step update and insert
- ◆ Multi-table Inserts – single step inserts to multiple tables

DULCIAN ETL Features (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 – restriction lifted
 - Separate metadata and data copy operations
 - Now accommodates different source/target block sizes
- ◆ Resumable Statements – repairable error harmless
- ◆ Parallel DML Operations on partitions

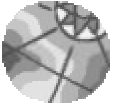
DULCIAN Warehouse Builder

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



DULCIAN 9i Warehouse Builder

- ◆ OWB 3i is a Beta release
- ◆ Production release is 9i Warehouse Builder
- ◆ New Mapping Architecture
- ◆ New Mapping and Transformation Paradigm
- ◆ Major Enhancements



DULCIAN 9i Mapping Architecture

- ◆ Previous architecture
 - > 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
- ◆ Operators
 - > Operator: output is a subset of the input rows
 - AGGREGATES and FILTERS
 - > Transform: output is equal to the input rows
 - Standard SQL functions, e.g. TO_CHAR

DULCIAN 9i Mapping and Transform Features

- ◆ Join Operator – join multiple sources
- ◆ Split Operator – output to multiple targets
- ◆ Multi-stage Transformation – pipeline SQL expr.
- ◆ 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

DULCIAN 9i 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
- ◆ Performance enhancements

DULCIAN Contact Information

Robert F. Edwards
 Dulcian, Inc.
 732-744-1116
 redwards@dulcian.com

www.dulcian.com