Real-Time Data Warehousing & Fraud Detection with Oracle 11gR2

Dr.-Ing. Holger Friedrich
Agenda

- Introduction
- Scope & Challenges
- Tools & Infrastructure
- Architecture & Implementation
- Prospects
- Conclusions
sumIT AG

- sumIT AG offers consulting and implementation services in Switzerland
- Oracle Certified Partner
- We are experts in
  - Data Warehousing and
  - Business Intelligence solutions using Oracle technology
- Our motto: *Get Value From Data*
- Visit our web-site: www.sumit.ch
• Founded by big International Sports Assoc
• Mission: prevent negative impact of online betting on sport events
• Means: Monitoring & analysing online betting quotes
• Events: World Cup, Olympics, …
Betting Model Sketch

- Actor
- Event
- Time
- Bet-Type
- Bet-Object
- Odds
- Customer
- Bookmaker
Scope

- Load/update of master data
- Load/update of transactional data
- Simple data quality assessment & feedback
- Active data retrieval from remote sources
- Data preparation for analysis
- Flexible visual data representation of analysis
- Support of ad-hoc reporting
- Simple supervision of load processes
Challenges

- Flexible, automatic daily master data update
- Transactional data retrieval every 30 sec
- Data ready for online analysis within ca. 5 min
- Integration of multiple data feeds
- Flexible retrieval & handling of XML-data
- High availability especially during events
- Adaptation to very different workloads
- Quick and cost efficient implementation
Software Tools

- **Oracle RDBMS 11gR2**
  - Oracle Warehouse Builder (ODIEE)
  - Oracle Workflow
  - XML DB
  - Partitioning
  - ASM
- **Oracle Business Intelligence EE 10g**
- **Oracle Data Guard 11gR2**
- **SQL Developer**
Infrastructural & Setup

- **Mid-Tier Wintel Server** running OBIEE
- **DB-Tier Lintel Server** running Oracle 11gR2 under Oracle Linux
- **NetApp Storage**
Oracle Std Architecture

Base data warehouse schema
Atomic-level data, 3nf design
Supports general end-user queries
Data feeds to all dependent systems

Application-specific performance structures
Summary data / materialized views
Dimensional view of data
Supports specific end-users, tools, and applications
Processing Steps

Data Provider → Staging Area → Data Store → Data Mart

- Retrieval & XSL-transform
- Data Mart load
- Retrieval Prep: URL-Generation
- Transform & ODS-load

30-seconds cycle  5-minute cycle
Oracle 11gR2 Dev Support

• Oracle Warehouse Builder 11gR2 is pre-installed
• OWB features used here (selection):
  - Workflow & Scheduler integration
  - XML DB APIs
  - Advanced operators for table functions, queues, ...
  - Dimensional modeling and loading support
  - Operator copying, grouping etc
• SQL-Developer used for monitoring and triggering job executions
• Advanced DB features like automatic DOP and Ref-Partitioning
Retrieval Preparation

- URL-generation in 30 sec cycles
- sets of URLs generated flexibly based on
  - audit infos of prior runs,
  - calendar data, and
  - master data
- URL-storage in Queue Table
- OWB 11gR2 provides modeling & code generation for handling required types and functionality (table functions, user defined types, queue tables)
Demo - Retrieval Prep
Retrieval & Staging

- Execution in 30 second cycles
- Retrieval of pending URLs from queue
- Per URL retrieval of data as XML-Feed
  - storage of data in relational staging table
  - transformation using XSL-File
  - data quality assessment and data selection
  - code transformations and lookups
- XML DB & OWB support
  - online retrieval of XML-feeds from remote sources
  - de-normalisation & transformation using XSL-Files
  - loading XSLed XML into relational staging table
XSL Transformation File
Demo - Retrieval & Stage
Transformation & DS Load

- Traditional Code-Lookups with Outer Joins
- Data merged or inserted in 3-NF data store schema
- Oracle Warehouse Builder 11gR2 raises productivity by providing
  - advanced look-up operators
  - expression and operator copy
  - operator grouping in UI
  - one-window GUI
- Reduced configuration & maintenance effort by automatic degree of parallelism DB feature
Automatic DOP

- SQL statement
- Statement is hard parsed and optimizer determines the execution plan
- If estimated time greater than threshold, optimizer determines ideal DOP
  - Actual DOP = MIN(PARALLEL_DEGREE_LIMIT, ideal DOP)
- If estimated time less than threshold, statement executes serially
- Statement executes in parallel
Parallel Stmnt Queuing

SQL statements

Statement is parsed and Oracle automatically determines DOP

If not enough parallel servers available queue statements

FIFO Queue

When the required number of parallel servers become available the first stmt on the queue is dequeued and executed

If enough parallel servers available execute immediately
Demo - ODS Load
Data Mart Model & Load (I)

- State-of-the-art star schema data mart design
- Modeling & Load through OWB mappings
- OWB provides wizards and special operators
- Load code is generated on demand based on dimension and cube definition
- Load operators handle complex tasks, e.g.
  - Implementation MOLAP or ROLAP (Star/Snow Flake)
  - Slowly Changing Dimensions
  - Orphan Handling (load and delete)
  - Loading multiple hierarchies
  - …
Observations
- Many people still don’t use dims & cubes in the DB
- It’s a complex topic, therefore the handling in OWB is not simple too
- Developers ignore the OWB functionality, because of:
  - lack of trust in the code generation capabilities
  - high confidence in their own coding capabilities
  - lack of understanding of how the OWB UI works
  - lack of expertise regarding dimensional modeling in general

OWB can dramatically increase Data Mart development performance and code quality
OWB 11gR2 bursts with additional DM features
Demo - DM Model & Load
End-user Interface OBIEE

- OBIEE 10g is used to provide dashboards

- KPI definitions provided by expert analysts from the online betting business
OBIEE Screenshots
Process Monitoring

• Classical methods available
  - OWB Run-Time Audit Browser
  - Oracle Workflow Monitor
  - Self developed APEX application

• Alternative with SQL Developer
  - Quick setup
    (import of report definitions)
  - Based on SQL Dev’s user report interface
  - no infrastructure required (e.g. Apache)
  - based on
    • OWB public Run-Time Views & execution API
    • System Views (Jobs, Scheduler etc)
Process Audits & Actions
Prospects

- **Oracle Warehouse Builder**
  - Use of Code Templates and Template Mappings, and Real-Time Maps in order to
    - Allow data retrieval via mid-tier server, while DB server has no internet access
    - Nevertheless keep all design data in the same repository

- **Oracle Business Intelligence EE 11g**
  - use of MOLAP interface with asymmetric drilling etc.

- **Oracle Data Mining**
  - Going beyond experts’ prior knowledge using knowledge discovery algorithms instead of static KPIs
  - flexibly adapt analysis to altered fraud strategies
“Oracle database 11gR2 allows you to efficiently implement all kinds of Data Warehouses, without the need to further invest in other tools”