Data Warehouse Applications
in financial institution

Kalle Tomingas
kalle.tomingas at gmail.com
Doctoral Student at TTU
BI Consultant at Mindworks Industries

Outline

- WHY
- HOW
- DW in financial institution
- DW development process and ETL
- BI applications
Why Data Warehouse?

- Data Warehousing is a technique to properly assemble and manage data from various sources to answer business questions not previously known or possible.
- Data Warehousing is a process, not a product or project.

Process of transformations:
- data -> information -> knowledge -> decisions -> acts

Evolution of techniques:
- Reporting -> DM -> DW -> BI

Ideology considerations:
- OLTP vs. OLAP
- Kimball vs. Inmon
- structural and semantic integration of data
- physical vs. virtual integration of data

BI and Data Warehousing delivery process:
- sources -> ETL -> DW -> DM -> applications, models and reports -> consumer

DW/BI Application Areas in Financial Institution

- Customer Relationship Management (CRM)
- Sales Automation
- Sales reporting
- Financial reporting
- Regulators reporting
- Credit risk management
- Operational risks management
- Strategic and tactical planning
- Predictive modeling
How to do a Data Warehouse?
A Conceptual Architecture for DW/BI:

- IBM’s “Corporate Information Factory”

![Diagram of IBM's Corporate Information Factory]

How to do a Data Warehouse?
A Conceptual Architecture for DW/BI:

- Teradata’s “Active Data Warehousing”

![Diagram of Teradata's Active Data Warehousing]
How to do a Data Warehouse?

- Department paradigm

- Enterprise paradigm

Optimized cost structure
Store once, use many
Single view of enterprise
How to do a Data Warehouse? (Business)

- Business initiative, motivation, needs and requirements
- Management support
- Owner, sponsor and budget
- One clear focus
- Project, that becomes a process
- Timeline and milestones
- Delivery of results
- Resources and know how
- Big picture
- Problem solving, data fixing, support and maintenance

How to do a Data Warehouse? (Technical)

- Hardware architectures
  - SMP, NUMA, MPP
- Relational Database Systems for BI
  - Traditional RDBMS systems: IBM, Oracle, MSSql, Sybase, ...
  - Special appliances: Teradata, Netezza, Exadata, Datalegro, Neoview, (Tandem, Tolerant), ...
  - New players: GreenPlum, Vertica, Cognito, Calpoint, Infobright, ExaSol, ParAccel, Asterdata, GridSQL, ...
  - Open source: PostgreSQL, MySql, CouchDB, (Ingres), ...
- ETL (Extract, Transform, Load) tools
  - Informatica, ODI, Ab Initio, Genio, MsSSIS(DTS), BO, SAS, ...
- Data Modeling and design tools
  - Modeling tools: ERwin, ER/Studio, PowerDesigner, Rational Rose, ...
  - Industry data models: FTI’s GFDM, Teradata’s (FS) LDM, SAS, IBM, Oracle...
- Reporting and query tools
  - Business Objects, Cognos, Informatica, Hyperion, Microstrategy, SAS, ...
- Data Mining
  - SAS, SPSS, Unica, ...
How to do a Data Warehouse? (Choices)

- What Gartner is guessing:
  - DB systems
  - BI tools
  - ETL tools
  - DM tools

- TPC (www.tpc.org)
  - Tpc-D
  - Tpc-R
  - Tpc-H

- Other
  - Analytics and consultancy companies
  - Internet
  - Blogs
  - Vendors

Data Warehousing techniques:

- How to model and design it?
- How to get data in?
  - Insert statements
  - odbc/oledb/jdbc
  - Load utilities
- How to transform data?
  - ETL vs ELT
- How to get data out?
  - query tools
  - reporting tools
  - interfaces and api-s to other systems
  - applications
- How to document and manage all this?
Enterprise Data Warehouse
(Technical figures and metrics)

- Teradata DW solution
  - 5-node distributed MPP RDBMS
  - 15 TB db space / 10+ TB data
  - 3 environments (development, test, production)

- DTS/ODI loading tools
  - 500+ data loading packages
  - 6 to 12 hours daily loading and calculation cycle
  - 3 countries

- Users and applications
  - 1000+ named users
  - 300+ active users
  - 100+ services and applications

Enterprise Data Warehouse
(Organizational figures)

- Business drive
  - Owner of EDW initiative: Financial Division
  - Focus on 2-3 different areas: Risk management, CRM, Reporting

- Staff
  - Some managers
  - 1 and half dba’s
  - 30+ analyst and developers
  - 10+ data stewards
  - 5+ change managers and testers
  - 5+ service administrators

- Geography
  - 3 Baltic countries in focus (+waiting list)
  - Co-operation with mother company
ETL Development Process

- Source interface development/agreement (excel, word, erwin, sql)
- DW data model design/customization (erwin, sql)
- Mappings for data transformation (excel)
- ETL development
  - Source data analysis (sql)
  - Target structures development (sql)
  - Transformation design (sql, dts, odi)
  - Documentation (word, excel, sharepoint)
- Access Layer and Application development
- Testing
- Deployment
- Data validation

ETL Development: Source data analysis

- Input:
  - diagrams
  - definitions
  - logical mapping
  - query tools
  - “tons” of sql

- Output:
  - changes in source interface
  - changes in source data
  - changes in target structures
  - changes in logical mappings
ETL Development: Impact Analysis

- Complicated loading package:
  - 96 transformation tasks
  - 110 dependencies
  - 128 tables
  - 152 source relations
  - 115 target relations

- How to decompose to smaller independent packages?
- How to estimate impact (internal, external)?
- How to analyze it?
- How to be sure?

ETL Development: Impact Analysis

- Structure and dependencies extraction from package metadata
- Grouping and coloring different subject areas
ETL Development: Impact Analysis

- Adding functional dependencies
- Elimination of manually set dependencies
- Grouping and Regrouping
- Visual dependency graph analysis

ETL Development: Impact Analysis

- Final decomposition based on graph analysis, subject areas, manual dependency elimination and functional dependencies analysis
DW/BI Applications

- BI Environment
  - Reporting: Business Objects (BO), Excel
  - Data Mining: SPSS/Clementine
  - Web based CRM environment
  - ETL tools for data loading and calculations: ODI, DTS

- CRM applications
  - Analytical (segmentation, offers, limits, income, assets, liabilities and overdue calculations)
  - Operational (sales and contact management, ad management)
  - Daily calculations and synchronization

- Sales and Financial reporting
  - Group reporting
  - Business measurement
  - Activity Based Costing (ABC)
  - Fund Transfer Pricing (FTP)
  - Profitability
  - Economic Value Added (EVA)

- Regulators reporting
  - Central Bank, card organizations, foreign trade reporting

- Risk management
  - Online customer’s rating and scoring
  - Risk modeling and calculations (PD,LGD,RWA)
  - Risk management and reporting
  - Fraud detection

---

DW/BI Applications: CRM

- Analytical CRM
  - segmentation calculations
  - offers calculations
  - limits and liabilities calculations
  - overdue calculations

- Operational
  - sales and contact management in teller’s desktop
  - campaign management
  - ad management in web

- Daily calculations and synchronization
  - closed loop marketing
DW/BI Applications: Risk Management

- Risk Management: Credit risk, Market risk, Operation risk, Capital calculation
- BASEL II requirements
- Online rating and scoring
- Defaults and losses modeling, prediction and calculation
  - Probability of Default (PD)
  - Loss Given Default (LGD)
  - Exposure at Default (EAD)
  - Expected Loss (EL = PD x LGD x EAD)
- Capital Calculations
  - Risk Weighted Assets (RWA) calculations
  - ...

DW/BI Applications: Activity Based Costing (ABC)

- Input data:
  - Activity hierarchy’s and weights
  - Resources (business units, )
  - Cost objects (Segments, Products, Channels)
  - GL costs
  - Aggregated sales figures (quantities)
- Monthly calculations
  - Weight adjustments
  - Mappings
  - Graph calculations
  - Adjustments and estimations
  - Cost distribution calculations
- Output
  - Unit costs
DW/BI Applications: Financial calculations and reporting

- ABC units and general cost distribution to agreement level
- Revenue distribution to agreement level
- Fund Transfer Pricing (FTP) calculations
- Treasury net profit and maturity mismatch distribution to agreement level
- Cost and revenue distribution to department level
- Other...