DW Construction & Maintenance

- DW Return Of Investment (ROI)
- Development methodology
- DW maintenance
- Improving DW performance
- DW security
DW ROI

- DW = Data + ETL + Meta data + OLAP + DM + Management & Control
- Source data => DW => Analysis Services
- Should we build a DW?
  - Feasible?
  - ROI?
DW Feasibility

- **Organization's internal complexity**
  - **A:** feasible
  - **B:** Not feasible
  - **C:** must
  - **D:** feasible

**D:** banks, telecomm, insurance companies
DW ROI Analysis

- Quantitative Analysis
  - ROI = Return / Cost
  - Payback period

- Example
  - Option I:
    - Invest: $8M; Return: yrs 3-6, $4M/yr
  - Option II:
    - Invest: $5M; Return: yrs 3-6, $2.5M/yr
**Example**

- **Cash discount rate:** 12%
- **ROI calculation:**
  - **Option I:**
    - \( \text{ROI} = \frac{400 \times 1.12^3 + 400 \times 1.12^2 + 400 \times 1.12 + 400}{800 \times 1.12^6} = 121\% \)
  - **Option II:**
    - \( \text{ROI} = \frac{250 \times 1.12^3 + 250 \times 1.12^2 + 250 \times 1.12 + 250}{500 \times 1.12^6} = 136\% \)
Qualitative Analysis
- Improve product quality & reduce production cost
- Coordinate internal relationship & promote team work
- Improve customer relationship & satisfaction
- Improve management capability
- Facilitate short & long term decision making
- Quicker response to market changes
An IDC survey on 62 companies in 1996
45% ROI between 3% - 1,838%
Within 3 years,
  - 90% companies, ROI $\geq$ 40%
  - 50% companies, ROI $\geq$ 160%
  - 25% companies, ROI $\geq$ 600%
Average investment, $22M
Average payback period, 2.3 year
DW Development Methodology

- Water fall
  - Better for transaction databases
- Spiral
  - Preferred by DW systems
- Build prototype, install/operation, maintain
DW Maintenance

- Four key issues
  - Data update period
  - Referential integrity
  - Data environment information
  - Data backup & recovery
Data update period

- Product vendor changed in DB, how often should we update the DW?
  - Depends on DB update frequency & customer’s requirement

Diagram:

- Product 1 Vender A
- Product 1 Vender B
- DB
- DW
Referential integrity in DB

- FK must appear in PK
- Can not delete PK if FK exists
DW Maintenance

- Referential integrity in DW
  - In 2006, *Degree* required is BS
  - In 2010, *Degree* required is MS
  - How to maintain DW referential integrity about *Degree*?
    - Periodic flash shots
    - Record all updates
DW Maintenance

- Periodic flash shots

Flash shot 1  →  Flash shot 2  →  Flash shot 2  →  ….
2008.1  →  2008.6  →  2006.1

- Record all updates

Flash shot + Update records = Complete data
DW Maintenance

- Data environment information
  - Save data and its environment information
  - Same data but in different environment
    - The background information when the data were generated
    - High performance CPU in 1995, 2005
      - Mhz, Ghz
DW Maintenance

- Data backup
  - What data need to be backup?
    - Separate current and old data
    - Backup current data
  - Complete backup or incremental backup
  - Time of backup
    - OLTP usually in the night
    - Consider backup the updated data in the same time
  - Backup media
    - Depends on the size of DW
Improving DW Performance

- Improving I/O performance
- Reduce query range
- Adopt parallel optimization techniques
- Select appropriate initialization parameters
Improving DW Performance

- Improving I/O performance
  - Combine data tables
    - Fewer I/O access
  - Build data block sequence
    - Sort data blocks & save in one page for fewer I/O access
  - Introduce redundancy
    - For attributes used in many tables, duplicate & save them in each table
Improving DW Performance

- Improving I/O performance
  - Aggregate on the primitive data & export directly to reduce I/O operations

- Reduce Query Range
  - Save appropriate data granularity according to analysis types
  - Partition tables to smaller ones
  - Build more indexes to speed up query responses
Improving DW Performance

- Adopt parallel optimization techniques
  - Since operation on DW are mostly “read only”, some processes can be done in parallel
  - ETL in parallel
Improving DW Performance

- Adopt parallel optimization techniques
  - Parallel query processes

```
Query 1          Query 2
    ↓            ↓
  Table 1

Query
  ↓
Operation on tables 1 & 2
  ↓
Operation on table 3
  ↓
Combine
```
DW Security

- DBMS provides certain data protection mechanism; but not on DW
  - E.g. access control

- 4 types of security in DW
  - Individual – own data
  - Group – group’s data
  - Hierarchy – all data below in the hierarchy
  - Conglomeration - all
DW Security

- Security management
- 2 types
  - External
    - Fire wall, pswd
  - Internal
Internal security management

Encryption on data

- Structure encryption
  - E.g. table with 3 attributes, A, B, C
  - Encrypt on A & C only

- Multiple encryption
  - Different subject uses different encryption schemes
  - DES, AES, RSA, ...

The highly summarized data should be kept more secured