Data Warehousing

Jens Teubner, TU Dortmund
`jens.teubner@cs.tu-dortmund.de`

Winter 2015/16

Part III

Planning and Starting a Data Warehouse Project

Planning a Data Warehouse Project

"Single Source of Truth" makes data warehouse a key component.

$\rightarrow$ Proper planning and management are vital.

DW projects have unique difficulties

- **Data** lies at the heart of DW, not application
    - $\rightarrow$ Usual strategies may not apply
- "Customers:" executives, not necessarily with IT background

Ready for a data warehouse project?

1. Strong management sponsor
2. Compelling business motivation
3. Feasibility

Strong Management Sponsor

DW projects tend to be **expensive** and **long running**

→ Seek a **strong** sponsor (times might get tougher)

→ Ideally: **multiple** sponsors, in case someone backs off

Multiple sponsors might have multiple expectations.

Don't make your project IT-only!

→ Seek sponsor in **business organization**, not only in IT.

→ Sponsor should have a **vision**, understand **business value** of DW

Compelling Business Motivation

Must convince **business organization** of the project

- Demonstrate **business value**
- Show **alignment with strategy** of the company

Possible motivators:

- Competition and external changes
- Internal crisis or difficulties
- Company acquisition ($\rightarrow$ need data integration)
- Always seek a **concrete business use case**

Feasibility

Primary concern: **data**

- **Data profiling:** explore available data
  - $\rightarrow$ Analyze data volume, value distributions, (foreign) key constraints, quality/consistency
  - $\rightarrow$ Which data is available in which system?
  - $\rightarrow$ Tools might help with this
- Enough data available (and accessible) to address business motivation?
- Be wary of **poor data quality**

Possibly also a concern:

- Technology
- People/staffing

Once you have green light for your project,

- **define the scope** of the data warehouse project,
    - Which business processes are part of the project?
    - Expected costs? Expected benefits?
    - Responsibilities? Time line? How do you measure success?
- **recruit project members**,
    - Need a good mixture of IT and business people
- **kick off**.

Database Design Process / Requirements Analysis

Remember the database design process ($\rightsquigarrow$ "Information Systems"):
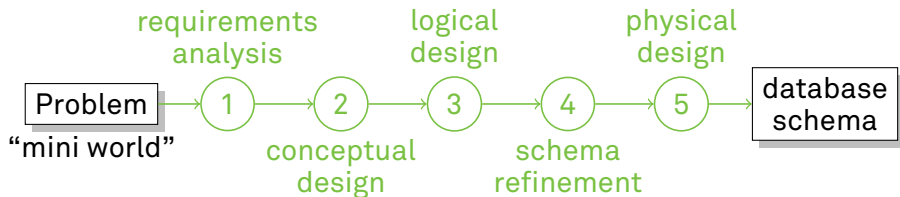
- Meet with customers
- Create ER diagram
- Turn ER diagram into (relational) database schema
- Refine schema
- Logical $\rightarrow$ physical schema; meet performance needs

Data Warehouse Design Process

We'll use a similar strategy to design a data warehouse:

- **Interview** "customers" (requirements analysis)
    - $\rightarrow$ Understand their problem to phrase it in your own terms.
- Create a **high-level model** that can be communicated with your "customers"
- From the high-level model, gradually work toward **logical and physical data models**.

Observe how this resembles the strategy

requirements analysis $\rightarrow$ ER model $\rightarrow$ relational model .

Requirements Analysis

"normal" database	**data warehouse**

- Focus on **subjects** (concepts)

- Focus on **processes**

- verbs → **relationships**

- "by" words; **dimensions**

processes: organization's core business processes

- *E.g.*, billing, shipping, taking orders, receiving orders, handling service calls, etc.

"by" words: natural groupings of the business process

- *E.g.*, "We need to look at claim payments **by** *policy holder*, *agent*, and *coverage*."

Existing/asked-for **reports** can be a good guide, too.

Sales Report				
	Q1/2013	Q2/2013	Q3/2013	Q4/2013
California	1770	1815	1815	1850
Los Angeles	*910*	*930*	*925*	*940*
San Francisco	*860*	*885*	*890*	*910*
Texas	1655	1710	1705	1695
Austin	*510*	*495*	*535*	*505*
Dallas	*595*	*610*	*615*	*605*
Houston	*550*	*605*	*555*	*585*
Total	**3425**	**3525**	**3520**	**3545**

Enterprise Data Warehouse Bus Matrix

From interviews, create an **enterprise data warehouse bus matrix**:

|	Date	Policy Holder	Coverage	Covered Item	Agent	Policy	Claim	Claimant	Payee
Underwriting Transactions	✔	✔	✔	✔	✔	✔			
Policy Premium Billing	✔	✔	✔	✔	✔	✔			
Agents' Commissions	✔	✔	✔	✔	✔	✔			
Claims Transactions	✔	✔	✔	✔	✔	✔	✔	✔	✔

- **Rows:** Business processes   (⤳ facts)
- **Columns:** "by" words, dimensions, grouping   (⤳ dimensions)

Opportunity Matrix

Often, it is also helpful to create an **opportunity matrix**:

	Underwriting & Actuarial	Marketing & Sales	Customer Service	Finance
Underwriting Transactions	✔	✔	✔	
Policy Premium Billing	✔	✔	✔	✔
Agents' Commissions		✔		✔
Claims Transactions	✔	✔	✔	✔

$\rightarrow$ Which organization/workgroup is involved in which process?
$\rightarrow$ Useful for justification

requirements
analysis

define scope   DW design
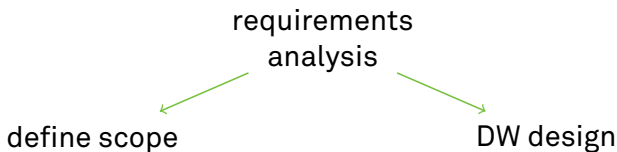
Project scope:

- Initially focus on **single business process**
  - $\rightarrow$ Choose a business process that is **manageable**, yet yields **business value**
  - $\rightarrow$ Requirements analysis is a good basis to decide on process
- Define **goals** and **milestones** to reach, estimate **cost**.

Prioritization Grid
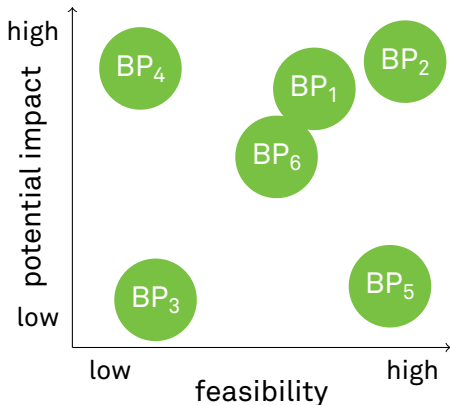
- business analysts → potential business impact
- IT department → feasibility

Summary

1. Ready for a data warehouse project?
    - → sponsor, business motivation, feasibility

2. Requirements analysis
    - → Focus on processes, not subjects
    - → Enterprise Data Warehouse Bus Matrix
    - → Opportunity Matrix
    - → Prioritization Grid

3. Define project scope
    - → Focus (on single business process)
    - → Pick high-impact, high-feasibility processes first