

# Data Warehousing

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

Summer 2019

## Part III

# Planning and Starting a Data Warehouse Project

# Planning a Data Warehouse Project

“Single Source of Truth” makes data warehouse a key component.

→ Proper planning and management are vital.



DW projects have unique difficulties

- **Data** lies at the heart of DW, not application
  - 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

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 (→ need data integration)
- Always seek a **concrete business use case**

Primary concern: **data**

- **Data profiling:** explore available data
  - Analyze data volume, value distributions, (foreign) key constraints, quality/consistency
  - Which data is available in which system?
  - 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

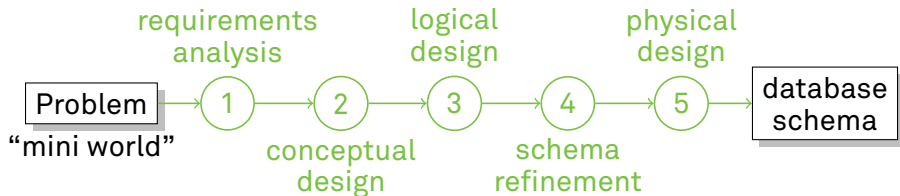
# Starting Your Data Warehouse Project

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 ( $\leadsto$  “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)
  - 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 → ER model → relational model .

## “normal” database

- Focus on **subjects**  
(concepts)
- verbs → **relationships**

## data warehouse

- Focus on **processes**
- “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.*”

# Requirements Analysis

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

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

# 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	✓	✓	✓	✓	✓	✓	✓	✓	✓

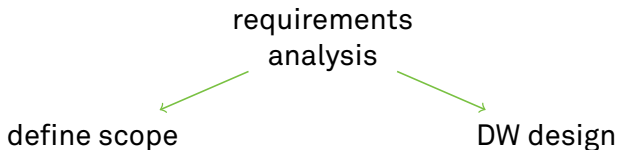
- **Rows:** Business processes (↪ facts)
- **Columns:** “by” words, dimensions, grouping (↪ dimensions)

# Opportunity Matrix

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

	<i>Underwriting &amp; Actuarial</i>	<i>Marketing &amp; Sales</i>	<i>Customer Service</i>	<i>Finance</i>
Underwriting Transactions	✓	✓	✓	
Policy Premium Billing	✓	✓	✓	✓
Agents' Commissions		✓		✓
Claims Transactions	✓	✓	✓	✓

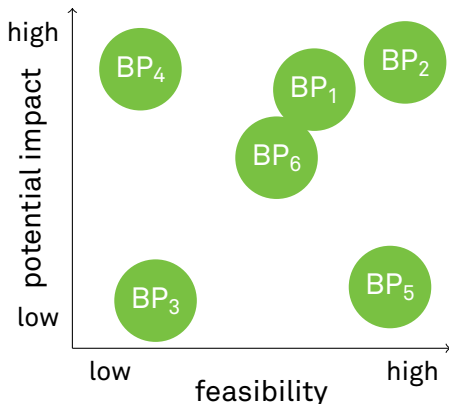
- Which organization/workgroup is involved in which process?
- Useful for justification



## Project scope:

- Initially focus on **single business process**
  - Choose a business process that is **manageable**, yet yields **business value**
  - 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

- 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