

Advanced Data Management Technologies

Unit 3 — Building a Data Warehouse

J. Gamper

Free University of Bozen-Bolzano
Faculty of Computer Science
IDSE

Acknowledgements: I am indebted to Michael Böhlen and Stefano Rizzi for providing me their slides, upon which these lecture notes are based.

Outline

- 1 Methodological Framework to Build a DW
- 2 DW Project Management

Outline

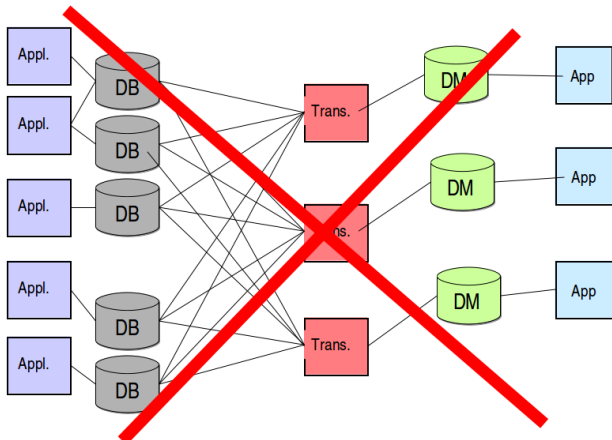
1 Methodological Framework to Build a DW

2 DW Project Management

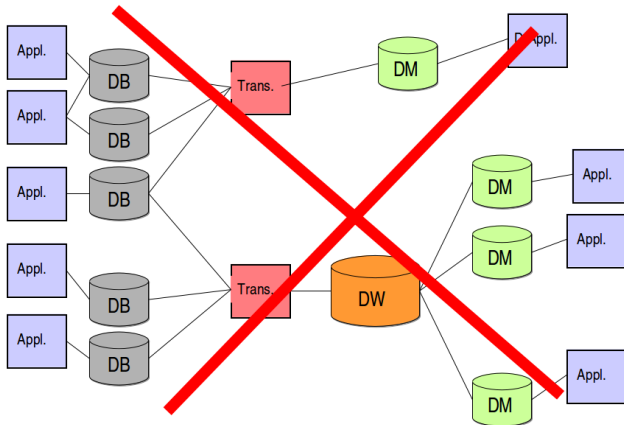
Methodological Framework

- Building a DW is a **very complex** task
- It requires an **accurate planning** aimed at devising satisfactory answers to organizational and architectural questions
- A large number of organizations **lack experience and skills** that are required to meet the challenges involved in DW projects
- Reports of DW project failures state that a major cause lies in the **absence of a global view** of the design process,
 - i.e., absence of a **design methodology**

Many Ways not to Do/1



Many Ways not to Do/2



Top-Down Approach

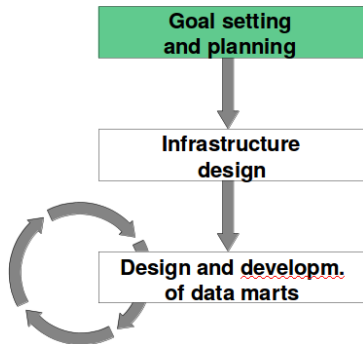
- **Top-down approach:** Analyze **global business needs**, plan how to develop a data warehouse, design it, and implement it **as a whole**
- Looks promising as it is **based on a global picture** of the goal to achieve, and in principle it ensures consistent, well integrated DW
- High cost estimates with **long-term implementations discourage** company managers
- Analyzing/integrating all relevant sources at the same time is a **very difficult task**, even because it is not very likely that they are all available and stable at the same time.
- It is **extremely difficult to forecast** the specific needs of every department, which can result in the analysis process coming to a standstill.
- Since **no working system is delivered in the short term**, users cannot check for this project to be useful, so they lose trust and interest in it.

Bottom-Up Approach

- **Bottom-up approach:** DW is **incrementally built** by iteratively creating several data marts
 - Each data mart is based on a set of facts that are linked to a specific department and that can be interesting for a user group
- Leads to concrete **results in a short time**
- Does **not require huge investments**
- Enables designers to investigate **one area at a time**
- Gives managers a **quick feedback** about the actual benefits of the system being built
- Keeps the interest for the project constantly high
- May determine a **partial vision** of the business domain

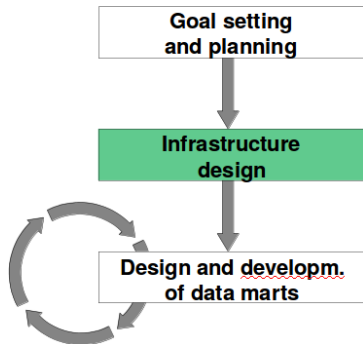
The Life-cycle – Goal Setting and Planning

- Set system goals, borders, and size
- Select an approach for design and implementation
- Estimate costs and benefits
- Analyze risks and expectations
- Examine the skills of the working team



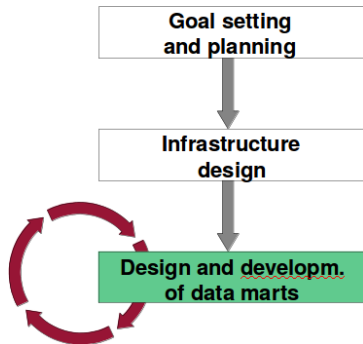
The Life-cycle – Infrastructure Design

- Analyze and compare the possible architectural solutions
- Assess the available technologies and tools
- Create a preliminary plan of the whole system

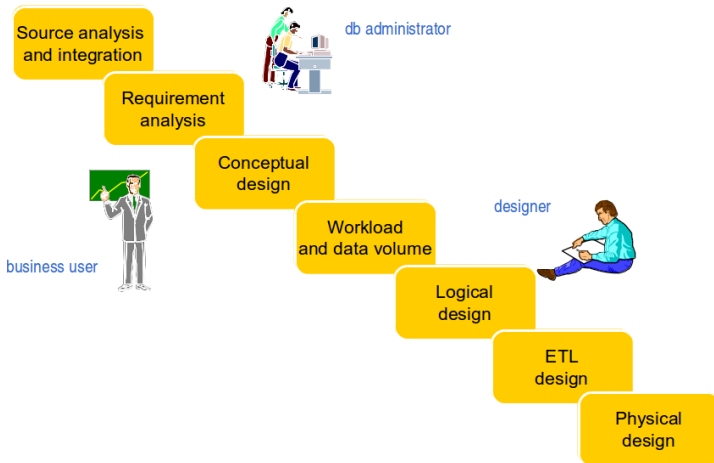


The Life-cycle – Design and Development of DMs

- Every iteration causes a new DM and new applications to be created and progressively added to the DW system.



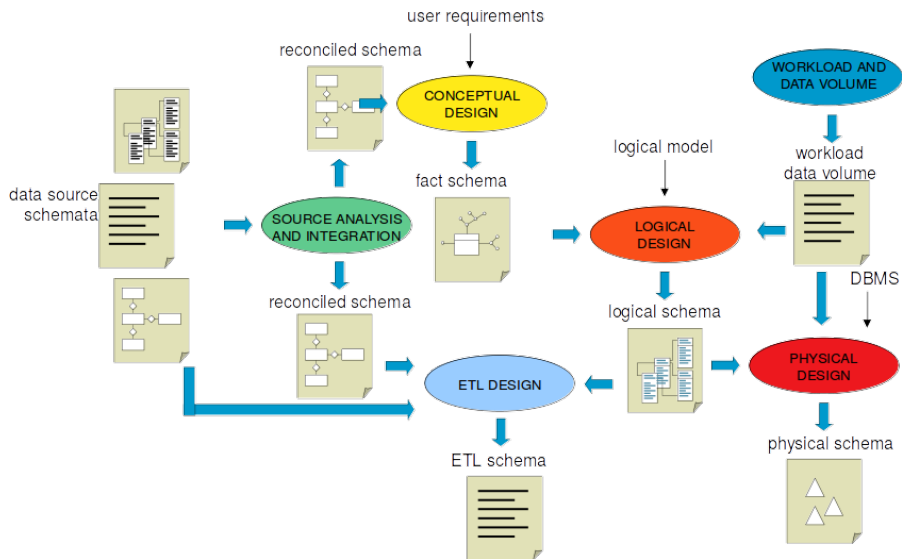
Data Mart Design Phases



Different Approaches for DM Design

- Supply-driven (data-driven) approach
- Demand-driven (requirement-driven) approach
- Mixed approach

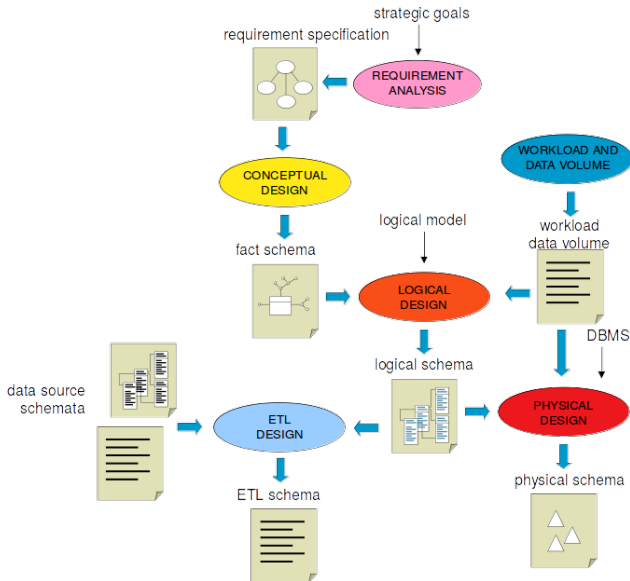
Supply-driven (Data-driven) DM Design/1



Supply-driven (Data-driven) DM Design/2

- Approach
 - Begin with an analysis of the **data sources**
 - User requirements show designers which groups of data should be selected
- Pros
 - Initial conceptual schema for DMs can be **automatically derived** from the reconciled layer
 - ETL design is **extremely streamlined** because every single information piece in a DM is directly associated with source attributes
 - Resulting DMs are **stable in time** since they are rooted in source schemata
 - In general, project goals can be reached in a **short time**
- Cons
 - User requirements play a **minor role** when specifying the contents
 - Designers have a **limited support** during the specification of facts, dimensions, and measures

Demand-driven (Requirement-driven) DM Design/1



Demand-driven (Requirement-driven) DM Design/2

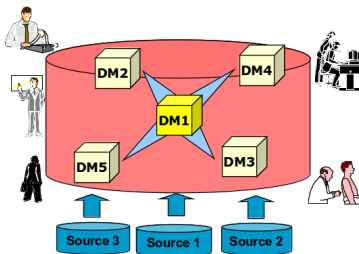
- Approach
 - Begin with the definition of information **requirements of DM users**
 - The problem of how to map those requirements into existing data sources is addressed at a later stage
- Pros
 - User requirements play a **leading role**
- Cons
 - Designers need **strong leadership** and mediation qualities to properly integrate different viewpoints
 - Required data might **not be available** in data sources
 - In general more **time intensive** since users do not have a clear understanding of the business goals

Mixed Approach to DM Design

- Requirement and data source analysis are done at the **same time**
 - user requirements
 - reconciled layer
- User requirements help to **reduce the complexity** of the reconciled layer
- Mixed approach is typically the best solution

The First Data Mart

- Is the one playing the most strategic role for the enterprise
- Should be a backbone for the whole DW
- Should lean on available and consistent data sources



Outline

1 Methodological Framework to Build a DW

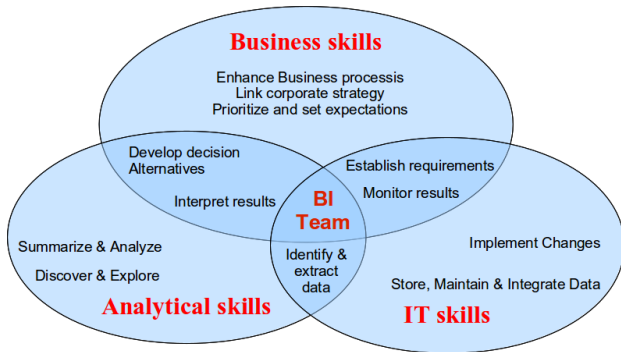
2 DW Project Management

DW Project Management

- DW projects are **large and different** from ordinary SW projects
- 12-36 months and \$1+ million per project
- Data marts are smaller and safer (bottom up approach)
- Reasons for failure
 - Lack of proper design methodologies
 - High HW+SW cost
 - Deployment problems (lack of training)
 - Organizational changes are difficult (new processes, data ownership, . . .)
 - Ethical issues (security, privacy, . . .)
- Creation of a **Business Intelligence Competence Center (BICC)** is crucial for success

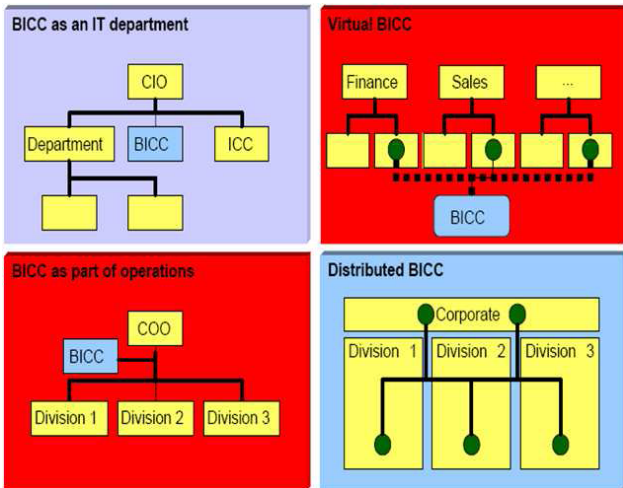
Business Intelligence Competence Center (BICC)/1

- Combines competences from different but crucial sectors
- Leads and is responsible for the DW project



Business Intelligence Competence Center (BICC)/2

- BICC requires a **change** in the organization (difficult!)
- No best place, but has **strategic importance**



Summary

- Building a DW is a **complex task**
 - There is a **lack of experience** and of a **methodological framework**;
 - **Top-down** versus **Bottom-up** design
 - **Supply-driven** versus **demand-driven** DM design
 - **First DM** plays a crucial role
- **Life-cycle of DW**: goal setting and planning, infrastructure design, iterative design and development of DMs
- Creation of a **Business Intelligence Competence Center** is crucial for success.