Business Intelligence (BI) and Data Warehousing (DW)

Data Warehouse System life Cycle

Jackline Ssanyu

February 28, 2020

Top-Down Vs Bottom-Up

There are two methodological approaches to data warehouse design

- Top-down
- Bottom-up

Top-Down Approach

With top-down approach:

- analyze global business needs
- plan how to develop a data warehouse
- design it , possibly using fact constellation schema, and implement it as a whole.

This approach is promising because:

- It will achieve excellent results since it is based on a global picture of the goal to achieve.
- it ensures a consistent, well-integrated data warehouse.

Disadvantages of top-down approach

- high cost estimates with long-term implementations discourage company managers from embarking on these kind of projects.
- analyzing and bringing together all relevant resources is a very difficult task, also 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 involved in a project, which can result in the analysis process coming to a standstill.
- since no prototype is going to be 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

- Data warehouses are incrementally built and several data marts are iteratively created.
- Each data mart is based on a set of facts that are linked to a specific company department and that can be interesting for a user group.
- Bottom-up approach coupled with quick prototyping remarkably reduces the time and cost needed for implementation.
- This way, Company managers will notice how useful the project being carried out is and will still be of interest to them.
- Bottom-up approach is the most widely accepted method.

Disadvantages of Bottom-up

- If the first prototype doesn't provide the best results, company managers and users may not feel the need to continue with the project.
- If not designed properly may hinder additional data marts to be easily added.

Business Dimensional Life Cycle (BDLC)

BDLC stands for the time needed for designing, developing and implementing data warehouse systems. The phases are:

- Project planning - define system goals and properties, assess impact to company practices, estimate costs, benefits and required resources. Develop a preliminary plan.
- Requirements definition- Designers should detect key factors of decision-making processes and turn them into design specifications. Includes data, technology and application requirements.
- Dimensional modeling - user data requirements are used for dimensional modeling which leads to the definition of logical schema which should then be transformed into a physical design implemented into the selected DBMS. Data staging design and development is also done in this phase.

Business Dimensional Life Cycle (BDLC)...

- Architectural design - technology requirements are used for architectural design. It involves definition of ETL and data analysis tools which will be used, product selection and installation.
- Application design - application requirements are used for designing the required report formats, user interfaces having interactive navigation of data and automatic knowledge extraction.
- Deployment phase - leads to system startup and maintenance to provide users with support and training.

Rapid Warehousing Methodology

- Rapid warehousing methodology is an iterative approach to managing data warehousing projects.
- The approach divides large projects into smaller, much less risky subprojects.
- Each subproject takes advantage of the data warehouse environments developed during previous subprojects.
- It expands them to add new features and evolves them.

Phases of Rapid Warehousing Methodology

- Assessment - this ascertains whether a company is ready to undertake the data warehousing project and setting goals, risks and benefits.
- Requirements - gathers data, architectural and application specifications of the system.
- Design - analysis specifications are refined to generate logical, physical data designs and data staging design. Implementation tools are also selected.
- Construction and final test - one data warehouse is implemented and populated with data extracted from data sources. Front-end applications are developed and tested.
- Deployment - the system is delivered and started up after properly training its end users.

Phases of Rapid Warehousing Methodology....

- Maintenance and administration - maintenance can implement additional features, upgrade the data warehouse to meet new needs, and check data for quality.
- Review - Each subproject includes three review processes: implementation check, post-deployment check to ensure that the organization is ready to use the warehouse, and final check for costs and benefits to be assessed.

Questions

- Data staging is one of the activities carried out during the design and implementation of a data warehouse. Explain what this activity is, and what it involves.