Book Name	data warehousing fundamentals by Paulraj Ponniah
Chapter No. 6	Requirements as a driving force for data warehouse

It consists of two parts

- REVIEW QUESTIONS
- **EXERCISES**

REVIEW QUESTIONS
1. "In a data warehouse, business requirements of the users form the single and most powerful driving force." Do you agree? If you do, state four reasons why. If not, is there any other such driving force?
Aswer: Yes, Because, Every task that is performed in every phase in the development of the data warehouse is determined by the requirements. Every decision made during the design phase—whether it may be the data design, the design of the architecture, the configuration of the infrastructure, or the scheme of the information delivery methods—is totally influenced by the requirements.
2. How do accurate information diagrams turn into sound data models for your data marts? Explain briefly.
Answer: information package diagrams are part of the requirements definition document

truly reflect the actual business requirements. Otherwise, your data model will not signify
what the users really want to see in the data warehouse

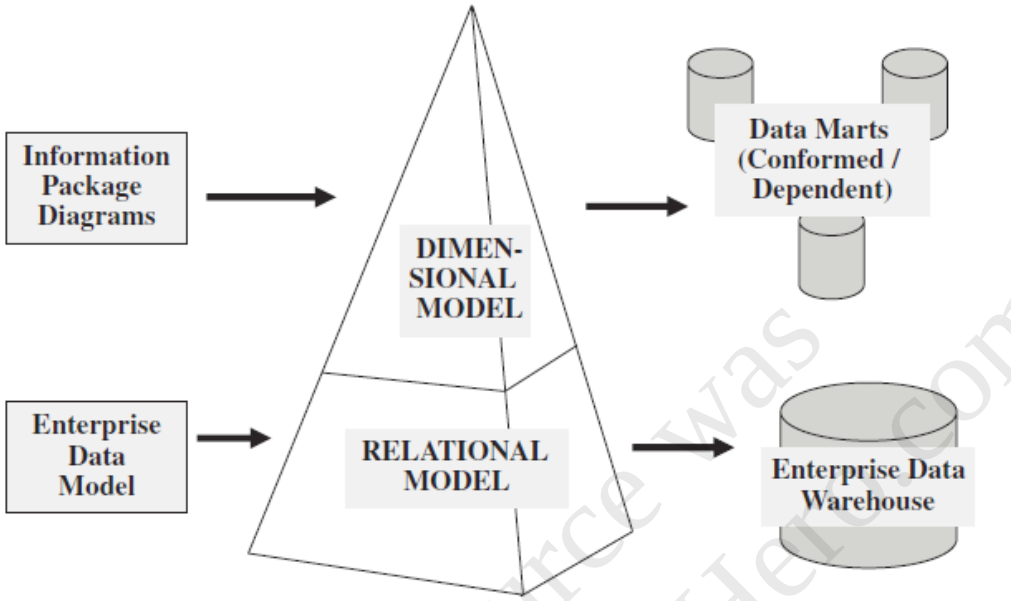

Figure 6-2 Requirements driving the data model.

3. Name five architectural components that are strongly impacted by business requirements.
Explain the impact of business requirements on any one of those five components.

- Source Data
  - Operational source systems
  - Computing platforms, operating systems, databases, files
  - Departmental data such as files, documents, and spreadsheets
  - External data sources
- Data Staging
  - Data mapping between data sources and staging area data structures
  - Data transformations
  - Data cleansing
  - Data integration
- Data Storage
  - Size of extracted and integrated data
  - DBMS features

- Growth potential
- Centralized for enterprise data warehouse
- Data marts—conformed, dependent, independent, federated

- Information Delivery
  - Types and number of users
  - Types of queries and reports
  - Classes of analysis
  - Dashboards/scorecards
  - Data mining operations
  - Front-end DSS applications
- Metadata
  - Operational metadata
  - ETL (data extraction, transformation, and loading) metadata
  - End-user metadata
  - Metadata storage
- Management and Control
  - Data loading
  - External sources
  - Alert systems
  - End-user information delivery

4. What is the impact of requirements on the selection of vendor tools and products?
Do requirements directly determine the choice of tools?

Requirements do not directly impact the selection of tools. Do not select the tools based on requirements and then adjust the architecture to suit the tools. This is like putting the cart before the horse.
Design the data warehouse architecture and then look for the proper tools to support the architecture.
A specific tool, ideally suited for the functions in one data warehouse, may be a complete misfit in another data warehouse. That is because the architectures are different. What do we mean by the statement that the architectures are different? Although the architectural components are generally the same in both data warehouses,
the scope, size, content, and the make-up of each component are not the same.

5. List any four aspects of information delivery that are directly impacted by business requirements. For two of those aspects, describe the impact.

- Queries and reports

- Types of analysis
- Information distribution
- Real time information delivery
- Decision support applications
- Growth and expansioN

Queries and Reports

Find out who will be using predefined queries and preformatted reports. Get the specifications. Also, get the specifications for the production and distribution frequency for the reports. How many users will be running the predefined queries? How often? The second type of queries is not a set of predefined ones. In this case, the users formulate their own queries and they themselves run the queries. Also in this class is the set of reports in which the users supply the report parameters and print fairly sophisticated reports themselves. Get as many details of this type of queries and this type of report sets as you can.

Power users may run complex queries, most of the time as part of an interactive analysis session. Apart from analysis, do your power users need the ability to run single complex queries?

Information Distribution

Where are your users? Are they in one location? Are they in one local site connected by a local area network (LAN)? Are they spread out on a wide area network (WAN)? These factors determine how information must be distributed to your users. Clearly indicate these details in the requirements definition.

In many companies, users get access to information through the corporate intranet.Webbased technologies are used. If this is the case in your company, Web-based technologies must be incorporated into the information delivery component. Let your requirements definition be explicit about these factors.

6. How do business requirements affect the choice of DBMS? Describe any three of the ways in which the selection of DBMS is affected.

many of the user requirements affect the selection of the proper DBMS. The relational DBMS products on the market are usually bundled with a set of tools for processing queries, writing reports, interfacing with other products, and so on. Your choice of the DBMS may be conditioned by its tool kit component. And the business requirements are likely to determine the type of tool kit component needed.

Ways of selection:

the following elements

of business requirements affect the choice of the DBMS:

Level of User Experience. If the users are totally inexperienced with database systems, the DBMS must have features to monitor and control runaway queries. On the other hand, if many of your users are power users, then they will be formulating their own queries. In this case, the DBMS must support an easy SQL-type language interface.

Types of Queries. The DBMS must have a powerful optimizer if most of the queries are complex and produce large result sets. Alternatively, if there is an even mix of simple and complex queries, there must be some sort of query management in the database software to balance the query execution.

Need for Openness. The degree of openness depends on the back-end and front-end architectural components and those, in turn, depend on the business requirements.

Data Loads. The data volumes and load frequencies determine the strengths in the areas of data loading, recovery, and restart.

7. What are MDDBs? What types of business requirements determine the use of MDDBs in a data warehouse?

(MDDBs): multidimensional database systems

Therefore, one important criterion for the database management system is that the system must be open. It must be compatible with the chosen back-end and front-end tools. More recently many composite data warehouse appliance products are available.

8. How do requirements affect the choice of the metadata framework? Explain very briefly.

You already know that metadata in a data warehouse is not merely data dictionary entries.
When a user queries the data warehouse, metadata acts as the information resource to connect the query parameters with the database components.

Metadata Management. If your metadata component does not have to be elaborate, then a DBMS with an active data dictionary may be sufficient. Let your requirements definition reflect the type and extent of the metadata framework.

9. What types of user requirements dictate the granularity or the levels of detail in a data warehouse?

how your users plan to use the data warehouse for analysis
kind of analysis you need to provide drill-down and roll-up
facilities for analysis. During data model

10. How do you estimate the storage size? What factors determine the size?

Data Staging Area. Calculate storage estimates for the data staging area of the overall corporate data warehouse from the sizes of the source system data structures for each business subject. Figure the data transformations and mapping into your calculation. For the data marts, initially estimate the staging area storage based on the business dimensions and metrics for the first data mart.
Overall Corporate Data Warehouse. Estimate the storage size based on the data structures for each business subject. You know that data in the data warehouse is stored by business subjects. For each business subject, list the various attributes, estimate their field lengths, and arrive at the calculation for the storage needed for that subject.
Data Marts—Conformed, Independent, Dependent, or Federated. While defining requirements, you create information diagrams. A set of these diagrams constitutes a data mart. Each information diagram contains business dimensions and their attributes. The information diagram also holds the metrics or business measurements that are meant for analysis. Use the details of the business dimensions and business measures found in the information diagrams to estimate the storage size for the data marts. Begin with your first data mart.
Multidimensional Databases. These databases support OLAP or multidimensional analysis. How much online analytical processing (OLAP) is necessary for your users?

The corporate data warehouse or the individual conformed or dependent data mart supplies the data for the multidimensional databases. Work out the details of OLAP planned for your users and then use those details to estimate storage for these multidimensional databases.

Howbig will your datawarehouse be? How much storage will be needed for all the data repositories? What is the total storage size? Answers to these questions will impact the type and size of the storage medium.

EXERCISES

1. Match the columns:	
1. information package diagrams 2. need for drill-down 3. data transformations 4. data sources 5. data aging 6. sophisticated analysis 7. simple and complex queries 8. data volume 9. specialized DSS 10. corporate data warehouse	A. determine data extraction B. provide OLAP C. provide data feed D. influences load management E. query management in DBMS F. low levels of data G. larger staging area H. influence data design I. possible pollution source J. data staging design
Answer:	
1. 2. 3. 4.A 5.I 6. 7. 8. 9. 10.C(to Dss Systems)	

2.It is a known fact that data quality in the source systems is poor in your company. You are assigned to be the data quality assurance specialist on the project team. Describe what details you will include in the requirements definition document to address the data quality problem.

Answer

I need to identify the following:

- Data Pollution Sources

- System conversions and migrations
- Heterogeneous systems integration
- Inadequate database design of source systems
- Data aging
- Incomplete information from customers
- Input errors
- Internationalization/localization of systems
- Lack of data management policies/procedures

- Types of Data Quality Problems
  - Dummy values in source system fields
  - Absence of data in source system fields
  - Multipurpose fields
  - Cryptic data
  - Contradicting data
  - Improper use of name and address lines
  - Violation of business rules
  - Reused primary keys
  - Nonunique identifiers

3. As the analyst responsible for data loads and data refreshes, describe all the details you will look for and document during the requirements definition phase.

Answer:

- Define the initial load.
- Determine how often each major group of data must be kept up-to-date in the data warehouse.
- How much of the updates will be nightly updates?
- Does your environment warrant more than one update cycle in a day?
- How are the changes going to be captured in the source systems?
- Define how the daily, weekly, and monthly updates will be initiated and carried out.
- If your plan includes real time data warehousing, specify the method for real time updates.

4. You are the manager for the data warehouse project at a retail chain with stores all across the country and users in every store. How will you ensure that all the details necessary to decide on the DBMS are gathered during the requirements phase? Write a memo to the senior analyst directly responsible to coordinate the requirements definition phase.
Answer: take care for data warehouse going to reside in one central location, or is it going to be distributed? The answer to this question will establish whether the selected DBMS must support distributed databases. Take care to: Data Warehouse Growth. Your business requirements definition must contain information on the estimated growth in the number of users, and in the number and complexity of queries. The growth estimates will have a direct relation to how the selected DBMS supports scalability
5. You are the query tools specialist on the project team for a manufacturing company with the primary users based in the main office. These power users need sophisticated tools for analysis. How will you determine what types of information delivery methods are needed? What kinds of details are to be gathered in the requirements definition phase?
Answer ○ Find out who will be using predefined queries and preformatted reports. Get the specifications. ○ How often? ○ type of queries is not a set of predefined ones. In this case, the users formulate their own queries and they themselves run the queries. ○ Also in this class is the set of reports in which the users supply the report parameters and print fairly sophisticated reports themselves. ○ Get as many details of this type of queries and this type of report sets as you can. ○ Power users may run complex queries, most of the time as part of an interactive analysis session. Apart from analysis, ○ do your power users need the ability to run single complex queries?