The Data Warehouse

# True-False Questions

|  |  |
| --- | --- |
|  | According to Inmon, a data warehouse is a subject-oriented, integrated database designed to support DSS functions where the data is volatile and relevant.  Answer: False |
|  | A data warehouse is typically physically separated from transaction processing systems.  Answer: True |
|  | Only a small fraction of the data that is captured, processed and stored in the enterprise is actually ever made available to executives and decision makers.  Answer: True |
|  | The most common component of a data warehouse environment is the operational data store.  Answer: True |
|  | A data mart is a subset of a data warehouse.  Answer: True |
|  | The essence of the data warehouse concept is a recognition that the characteristics and usage patterns of operational systems used to automate business processes and those of a DSS are fundamentally similar and symbiotically linked.  Answer: False |
|  | One difference between data warehouses and operational data stores is the presence of *metadata*.  Answer: True |
|  | Metadata are data about data.  Answer: True |
|  | Metadata are detailed data that have been aggregated and condensed into a more useful form.  Answer: False |
|  | The very essence of the DW environment is that the data contained within the boundaries of the warehouse are integrated. This integration manifests itself through consistency in naming convention and measurement attributes, accuracy, and common aggregation.  Answer: True |
|  | The time horizon for data in the data warehouse is typically significantly longer than operational data sores.  Answer: True |
|  | One of the benefits of integrated data is the establishment of a common unit of measure for all synonymous data elements from dissimilar databases.  Answer: True |
|  | According to Inmon, because data warehouses are highly de-normalized, they are highly redundant.  Answer: False |
|  | One objective of the data warehouse environment is to minimize the impact on operational systems.  Answer: True |
|  | The end user in a data warehouse environment deals directly with the application messaging layer.  Answer: False |
|  | The process management layer in a data warehouse can be thought of as a scheduler.  Answer: True |
|  | The processes needed to prepare the data to be loaded into a data warehouse are performed in the data staging layer.  Answer: True |
|  | A data warehouse topology requires a centralized data warehouse that is accessed by one or more decision support tools.  Answer: False |
|  | One of the most important factors in the development of a data warehouse is a comprehensive architectural framework.  Answer: True Reference: p. 60 |
|  | *Transformation mapping* metadata records how data from operational data stores and external sources are transformed on the way into the warehouse.  Answer: True Reference: p. 57 |

# Multiple Choice Questions

|  |  |
| --- | --- |
| 21. | The most common component in the DW environment is the \_\_\_\_\_\_\_\_\_\_\_\_\_. Its primary day-to-day function is to store the data for a single, specific set of operational applications.  a. data mart  b. data warehouse  c. operational data store  d. data staging tables  **Answer: c** |
| 22. | Which of the following is not true of a data warehouse?  a. Implicit in its definition is that the data warehouse is physically separated from all other operational systems.  b. The data warehouse replaces the need for all other reporting systems within an organization.  c. The data warehouse holds aggregated data and atomic data for management.  d. None of the above.  **Answer: b** |
| 23. | An alternative to the data warehouse concept is a lower-cost, scaled-down version referred to as a:  a. data mart.  b. metadata warehouse.  c. operational data store.  d. None of the above.  **Answer: a** |
| 24. | Which of the following is not true of a data mart?  a. The data mart is often viewed as a way to gain entry into the realm of data warehouses and to make the mistakes on a smaller scale.  b. Vendors of data warehouse applications have found it easier to deal with a small group of isolated users than with the IS department of an entire organization.  c. The data mart is more efficient than a fully-developed data warehouse.  d. None of the above.  **Answer: c** |
| 25. | Which of the following is not a characteristic of a data warehouse?  a. Data integrated  b. Volatile  c. Subject oriented  d. Time variant  **Answer: b**  **Reference: p. 46** |
| 26. | The essence of the data warehouse environment is that the data contained within the boundaries of the warehouse are \_\_\_\_\_\_\_\_\_\_\_\_\_.  a. integrated  b. consistent  c. streamlined  d. accurate  **Answer: a**  **Reference: p. 47** |
| 27. | The concept of time variant data implies which of the following statements?  a. Data are simply assumed to be accurate as of some moment in time and not necessarily *right now*.  b. Data are assumed to be accurate at the moment they were loaded into the data warehouse.  c. Data are assumed to vary over time.  d. Both a and b  **Answer: d**  **Reference: p. 49** |
| 28. | Which of the following activities would not normally be associated with a data warehouse?  a. Loading  b. Updating  c. Accessing  d. None of the above.  **Answer: b**  **Reference: p. 50** |
| 29. | The *\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_* represents the source data for the DW. This layer is comprised, primarily, of operational transaction processing systems and external secondary databases.  a. information access layer  b. operational and external layer  c. data access layer  d. process management layer  **Answer: b**  **Reference: p. 51** |
| 30. | Which layer of the data warehouse architecture does the end user deal directly with?  a. Data access layer  b. Application messaging layer  c. Information access layer  d. None of the above.  **Answer: c** |
| 31. | The *\_\_\_\_\_\_\_\_\_\_\_\_\_\_* serves as a sort of interface or middleman between the operational and information access layers and the data warehouse itself. This layer spans the various databases contained within the DW and facilitates common access by the DW users.  a. data access layer  b. application messaging layer  c. information access layer  d. None of the above.  **Answer: a** |
| 32. | Which of the following would not be a good example of metadata?  a. The directory of where the data is stored.  b. The rules used for summarization and scrubbing.  c. Where the operational data came from.  d. All of the above are examples of metadata.  **Answer: d**  **Reference: p. 55** |
| 33. | Which layer within the data warehouse architecture focuses on scheduling tasks that must be accomplished to build and maintain the data warehouse and data directory information?  a. Data access layer  b. Process management layer  c. Application messaging layer  d. None of the above.  **Answer: b** |
| 34. | Which of the following is a valid data warehouse configuration?  a. Centralized data warehouse  b. Virtual data warehouse  c. Distributed data warehouse  d. All of the above.  **Answer: d** |
| 35. | The *\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_* has to do with transporting information around the enterprise computing network. This layer is also referred to as the "middleware,” but it can typically involve more that just networking protocols and request routing.  a. application messaging layer  b. process management layer  c. data access layer  d. information access layer  **Answer: a** |
| 36. | The *\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_* is where the actual data used for decision support throughout the organization are located.  a. information access layer  b. operational and external layer  c. physical data warehouse layer  d. process management layer  **Answer: c** |
| 37. | The final component of the DWA is the *\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_*. This layer includes all of the processes necessary to select, edit, summarize, combine and load data warehouse and information access data from operational and/or external databases.  a. data staging layer  b. operational and external layer  c. physical data warehouse layer  d. process management layer  **Answer: a** |
| 38. | The process that records how data from operational data stores and external sources are transformed on the way into the warehouse is referred to as:  a. summarization algorithms.  b. transformation mapping.  c. back propagation.  d. extraction history.  **Answer: b** |
| 39. | The *\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_* applied to the detail data are of importance to any decision maker analyzing or interpreting the meaning of the summaries. These metadata can also save time by making it easier to decide which level of summarization is most appropriate for a given analysis context.  a. summarization algorithms  b. transformation mapping  c. back propagation  d. extraction history  **Answer: a** |
| 40. | Whenever historical information is analyzed, meticulous update records must be kept. Often a decision maker will begin the process of constructing a time-based report by reviewing the *\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_* because any changes to the business rules must be ascertained in order to apply the right rules to the right data.  a. summarization algorithms  b. transformation mapping  c. back propagation  d. extraction history  **Answer: d** | |

## Essay Questions

|  |  |
| --- | --- |
| 41. | *What are the seven deadly sins of data warehouse implementation?*   1. “If you build it, they will come.” 2. Omission of a data warehouse architectural framework. 3. Underestimating the importance of documenting all assumptions and potential conflicts. 4. Abuse of methodology and tools. 5. Abuse of the data warehouse life cycle. 6. Ignorance concerning the resolution of data conflicts. 7. Failure to document the mistakes made during the first DW project. |
| 42. | *What are the characteristics of a data warehouse?*  Subject orientation: data are organized based on how the users refer to it.  Integrated: all inconsistencies regarding naming convention and value representations are removed.  Nonvolatile: data are stored in read-only format and do not change over time.  Time Variant: data are not current but normally time-series.  Summarized: operational data are mapped into a decision-usable format.  Large Volume: time-series datasets are normally quite large.  Not Normalized: DW data can, and often are, redundant.  Metadata: data about data are stored.  Data Sources: internal and external unintegrated operational systems. |
|  | *What is a data warehouse?*  A data warehouse is “a collection of integrated, subject-oriented databases designed to support the DSS function, where each unit of data is nonvolatile and relevant to some moment in time. Implicit in this definition is that: (1) the data warehouse is physically separated from all other operational systems and (2) the data warehouse holds aggregated data and transactional data for management separate from those used for online transaction processing. |
|  | *List and explain the different layers in the data warehouse architecture.*  The operational and external database layer represents the source data for the DW. This layer is comprised, primarily, of operational transaction processing systems and external secondary databases. The information access layer of the DWA is the layer that the end-user deals with directly. In particular, it represents the tools that the end user normally uses day-to-day to extract and analyze the data contained within the DW. The data access layer serves as a sort of interface or middleman between the operational and information access layers and the data warehouse itself. This layer spans the various databases contained within the DW and facilitates common access by the DW users. In order to provide for universal data access, it is absolutely necessary to maintain some form of data directory or repository of metadata information. Metadata are data about the data stored within the DW. The process management layer focuses on scheduling the various tasks that must be accomplished to build and maintain the data warehouse and data directory information. The application messaging layer has to do with transporting information around the enterprise computing network. This layer is also referred to as the "middleware," but it can typically involve more than just networking protocols and request routing. The physical data warehouse layer is where the actual data used for decision support throughout the organization is located. The final component of the DWA is the data staging layer. Data staging (sometimes referred to as copy or replication management) includes all of the processes necessary to select, edit, summarize, combine, and load data warehouse and information access data from operational and/or external databases. |
|  | *What are metadata? Why are they so important to a data warehouse?*  Metadata are simply an abstraction from data. They have been defined as data about data. It is high-level data that provides us with a concise description of lower-level data. Metadata are an essential ingredient in the transformation of raw data into knowledge. Metadata are useful in finding correlations within the data. |