

TEST BANK



BUSINESS INTELLIGENCE

A MANAGERIAL APPROACH



Efraim Turban | Ramesh Sharda
Jay E. Aronson | David King

Chapter 2

Data Warehousing

True-False Questions

1. A real-time, enterprise-level data warehouse combined with a strategy for its use in decision support can leverage data to provide massive financial benefits for an organization.

Answer: True

Difficulty: Easy

Page Reference: 39

2. A data warehouse differs from an operational database in that most data warehouses have a product orientation and are tuned to handle transactions that update the database.

Answer: False

Difficulty: Easy

Page Reference: 39

3. A data warehouse uses either a relational, network, or multidimensional structure.

Answer: False

Difficulty: Easy

Page Reference: 40

4. Once the data are entered into the data warehouse, users cannot change or update the data.

Answer: True

Difficulty: Medium

Page Reference: 40

5. An independent data mart ensures that the end-user is viewing the same version of the data that are accessed by all other data warehouse users.

Answer: False

Difficulty: Moderate

Page Reference: 40

6. Operational data store is used for the medium- and long-term decisions associated with the enterprise data warehouse.

Answer: False

Difficulty: Moderate

Page Reference: 41

7. The data for an operational data store are created from oper marts.

Answer: False

Difficulty: Moderate

Page Reference: 41

8. Metadata can generally be defined in terms of usage and pattern.

Answer: True

Difficulty: Moderate

Page Reference: 41

9. There are five levels of metadata management maturity that can be used to describe where an organization is in terms of how and how well it utilizes its metadata.

Answer: True

Difficulty: Medium

Page Reference: 42

10. Data warehouses contain vast amounts of internal data but are limited because they do not contain external data.
- Answer:** False **Difficulty:** Moderate **Page Reference:** 43
11. The two most common data warehouse architectures are three-tiered and four-tiered.
- Answer:** False **Difficulty:** Easy **Page Reference:** 44
12. In a three-tier architect for data warehousing, data from the warehouse are processed once and deposited in an additional multidimensional database, organized for easy multidimensional analysis and presentation, or replicated in data marts.
- Answer:** False **Difficulty:** Moderate **Page Reference:** 44
13. Page loading speed is an important consideration in designing Web-based applications.
- Answer:** True **Difficulty:** Easy **Page Reference:** 47
14. A centralized data warehouse is similar to the hub-and-spoke architecture except that there are no independent data marts.
- Answer:** False **Difficulty:** Moderate **Page Reference:** 50
15. The centralized data warehouse helps to simplify data management and administration.
- Answer:** True **Difficulty:** Easy **Page Reference:** 50
16. Because of performance and data quality issues, most experts agree that federated approaches work well to replace data warehouses.
- Answer:** False **Difficulty:** Easy **Page Reference:** 51
17. IT managers are often faced with challenges as the ETL process typically consumes 90 percent of the time in a data-centric project.
- Answer:** False **Difficulty:** Easy **Page Reference:** 54
18. The top-down development approach for data warehouse is also known as the data mart approach.
- Answer:** False **Difficulty:** Easy **Page Reference:** 58
19. Database management systems tend to be no redundant, normalized, and relational, whereas data warehouses are redundant, unnormalized, and multidimensional.
- Answer:** True **Difficulty:** Easy **Page Reference:** 65

20. Data cleansing is a critical, but difficult aspect of data warehousing and it involves reconciling conflicting data definitions and formats organizationwide.

Answer: True

Difficulty: Moderate

Page Reference: 66

Multiple Choice Questions

21. Data warehouse is a _____, integrated, time-variant, non-volatile collection of data in support of management's decision making process.

- a. analysis-oriented
- b. object-oriented
- c. subject-oriented
- d. model-oriented

Answer: c

Difficulty: Moderate

Page Reference: 39

22. Once data are entered into the warehouse, users cannot change or update the data. Obsolete data are discarded, and changes are recorded as new data. This _____ characteristic is one of the characteristics of data warehousing.

- a. changeable
- b. nonvolatile
- c. nonperishable
- d. static

Answer: b

Difficulty: Moderate

Page Reference: 40

23. Which of the following describes how data are organized and how to use them effectively?

- a. data directory
- b. data indexing
- c. data filed
- d. metadata

Answer: d

Difficulty: Moderate

Page Reference: 40

24. The high cost of data warehouses limits their use to large companies. As an alternative, many firms use a lower-cost, scaled-down version of a data warehouse referred to as _____.

- a. data mart
- b. operational data store
- c. dependent data mart
- d. independent data mart

Answer: d

Difficulty: Moderate

Page Reference: 40

25. Which of the following are created when operational data need to be analyzed multidimensionally?

- a. oper marts
- b. data marts
- c. dependent data marts
- d. independent data marts

Answer: a

Difficulty: Moderate

Page Reference: 41

26. Which of the following type of metadata describes the meaning of the data in a specific domain?

- a. technical
- b. business
- c. structural
- d. semantic

Answer: d

Difficulty: Moderate

Page Reference: 41

27. Which of the following is not a level of metadata management maturity?

- a. ad hoc
- b. discovered
- c. managed
- d. investigated

Answer: d

Difficulty: Moderate

Page Reference: 42

28. Which of the following is one of the components of data warehousing process that enables users to access the data warehouse?

- a. middleware tools
- b. users interface
- c. query tools
- d. OLAP

Answer: a

Difficulty: Easy

Page Reference: 44

29. The advantage of three-tier architecture for data warehousing is its separation of the functions of the data warehouse, which eliminates resource constraints and makes it possible to easily create data _____.

- a. banks
- b. cubes
- c. bases
- d. marts

Answer: d

Difficulty: Moderate

Page Reference: 44

30. The _____ have inconsistent data definitions and different dimensions and measures, making it difficult to analyze data across the marts.
- a. data marts
 - b. operational data marts
 - c. dependent data marts
 - d. independent data marts

Answer: d

Difficulty: Moderate

Page Reference: 50

31. As data warehouses grow in size, the issues of _____ also exasperate.
- a. data alignment
 - b. data addition
 - c. data integration
 - d. data combination

Answer: c

Difficulty: Easy

Page Reference: 52

32. Which of the following is an evolving tool space that promises real-time data integration from a variety of sources such as relational databases, Web services, and multi-dimensional databases?
- a. information integration
 - b. data management integration
 - c. SQL data integration
 - d. enterprise information integration

Answer: d

Difficulty: Moderate

Page Reference: 53

33. ETL process consists of extract, transform, and load. Transformation occurs by using _____ or lookup tables or by combining the data with other data.
- a. rules
 - b. policies
 - c. strategies
 - d. procedures

Answer: a

Difficulty: Moderate

Page Reference: 54

34. Karacsony indicates that there is a direct correlation between the extent of _____ data and the amount of ETL processes. When data are managed correctly as an enterprise asset, ETL efforts are significantly reduced.
- a. enormous
 - b. bad
 - c. redundant
 - d. wrong

Answer: c

Difficulty: Moderate

Page Reference: 55

35. Which of the following is a benefit of a data warehouse?

- a. consolidated view of corporate data
- b. better and more timely information
- c. simplification of data access
- d. all of the above

Answer: d

Difficulty: Easy

Page Reference: 56

36. The data warehouse collects not only _____ data but also near real-time data from each vending machine (viewed as a store) that could be transmitted via wireless connection to headquarters.

- a. current
- b. historical
- c. future
- d. all of the above

Answer: b

Difficulty: Easy

Page Reference: 57

37. A star schema contains a central _____ surrounded by several dimension tables.

- a. data bases
- b. fact table
- c. data tree
- d. data table

Answer: b

Difficulty: Moderate

Page Reference: 61

38. Which of the following is the failure factor in data warehousing identified by Turban et al.?

- a. cultural issues are ignored
- b. inappropriate architecture
- c. missing information
- d. all of the above

Answer: d

Difficulty: Easy

Page Reference: 65

39. _____ is a critical aspect of data warehousing that includes reconciling conflicting data definitions and formats organizationwide.

- a. Data modification
- b. Fact refinement
- c. Data purification
- d. Data cleansing

Answer: d

Difficulty: Moderate

Page Reference: 66

40. Which of the following is needed to determine how data are to be retrieved from a data warehouse, and will assist in the physical definition of the warehouse by helping to define which data require indexing?
- a. indexing modeling
 - b. retrieval modeling
 - c. access modeling
 - d. tactic modeling

Answer: c

Difficulty: Moderate

Page Reference: 67

Fill In the Blanks

41. Decision makers require concise, dependable information about current operations, trends, and changes. Data often are fragmented in distinct operational systems, so managers often make decisions with partial information at best. Data warehousing cuts through this obstacle by accessing, integrating, and organizing key operational data in a form that is consistent, reliable, timely, and readily available where needed.

Difficulty: Easy

Page Reference: 39

42. Dependent data mart is a subset that is created directly from the data warehouse. It has the advantages of using a consistent data model and providing quality data.

Difficulty: Moderate

Page Reference: 40

43. Operational data store provides a fairly recent form of customer information files. It is a type of database often used as an interim staging area for a data warehouse.

Difficulty: Moderate

Page Reference: 40

44. Independent data mart is a small data warehouse designed for a strategic business unit or a department.

Difficulty: Easy

Page Reference: 40

45. An enterprise data warehouse is a large-scale data warehouse that is utilized across the enterprise for decision support.

Difficulty: Easy

Page Reference: 41

46. In three-tier architecture for data warehouse, operational systems contain the data and the software for data acquisition in one tier, the data warehouse is another tier, and the third tier includes the decision support and the client.

Difficulty: Hard

Page Reference: 44

47. The **federated approach** is a concession to the natural forces that undermine the best plans for developing a perfect system. It uses all possible means to integrate analytical resources from multiple sources to meet changing needs or business conditions.

Difficulty: Moderate

Page Reference: 50

48. **Data integration** comprises three major processes that, when correctly implemented, permits data to be accessed and made accessible to an array of ETL and analysis tools and data warehousing environment.

Difficulty: Easy

Page Reference: 52

49. EII tools use pre-defined metadata to populate views that make integrated data appear relational to end-users. **XML** may be the most important aspect of EII, because it allows data to be tagged either at creation time or later.

Difficulty: Moderate

Page Reference: 53

50. One of the benefits of a well-designed data warehouse is that business rules can be stored in a **metadata repository** and applied to the data warehouse centrally.

Difficulty: Hard

Page Reference: 55

51. A data warehouse contains numerous **business rules** that define such things as how the data will be used, summarization rules, standardization of encoded attributes, and calculation rules.

Difficulty: Moderate

Page Reference: 55

52. The **data mart** is a scaled down version of the data warehouse that centers on the requests of a specific department such as marketing or sales.

Difficulty: Easy

Page Reference: 60

53. The data warehouse design is based upon the concept of **dimensional modeling**, which is a retrieval-based model that supports high-volume query access.

Difficulty: Moderate

Page Reference: 61

54. **Dimension tables** are tables that address how data will be analyzed.

Difficulty: Easy

Page Reference: 61

55. The star schema design provides extremely fast query response time, simplicity, and ease of maintenance for **read-only** database structures.

Difficulty: Hard

Page Reference: 62

56. It is important to identify the right level of **granularity** that will be needed, during the scoping of a data warehouse project. A low level of that will result in more data being stored in the warehouse.

Difficulty: Easy

Page Reference: 62

57. During data modeling, expertise is required to determine what data are needed, define business rules associated with the data, and decide what **aggregations** may be necessary.

Difficulty: Hard

Page Reference: 67

58. The main issues pertaining to **scalability** are the amount of data in the warehouse, how quickly the warehouse is expected to grow, the number of concurrent users, and the complexity of user queries.

Difficulty: Hard

Page Reference: 68

59. **Real-time data warehousing or active data warehousing** is the process of loading and provides data via the data warehouse as they become available.

Difficulty: Moderate

Page Reference: 69

60. **Data warehouse administrator** is the person responsible for the administration and management of a data warehouse.

Difficulty: Easy

Page Reference: 73

Essay Questions

61. *List the characteristics of a data warehouse.*

- Subject-oriented
- Integrated
- Time-variant
- Nonvolatile
- Summarized
- Not normalized
- Contains internal and external data
- Contains metadata

Difficulty: Moderate

Page Reference: 39

62. *Describe the data warehousing process.*

- Data sources - Data are sourced from multiple independent operational “legacy” systems and possibly from external data providers (such as the U.S. Census).
- Data extraction - Data are extracted using custom-written or commercial software called extraction, transformation, and load (ETL).
- Data loading - Data are then loaded into a staging area where they are transformed and cleansed.
- Comprehensive database - The enterprise data warehouse to support all decision analysis.
- Metadata - Metadata are maintained so that they can be assessed by IT personnel and users.
- Middleware tools - Middleware tools enable access to the data warehouse.

Difficulty: Moderate

Page Reference: 43

63. *What are the issues to consider when deciding which architecture for data warehousing using?*

- Which database management system to use?
- Will parallel processing and/or partitioning be utilized?
- Will data migration tools be used to load the data warehouse?
- What tools will be used to support data retrieval and analysis?

Difficulty: Moderate

Page Reference: 47

64. *Describe alternatives architects to the basic architectural design types in data warehousing.*

See Figure 2.5 (a-e) Alternative Data Warehouse Architectures

Difficulty: Moderate

Page Reference: 47

65. *List the 10 factors identified by Ariyachandra and Watson (2005) that potentially affect the architecture selection decision.*

- Information interdependence between organizational units
- Upper management’s information needs
- Urgency of need for a data warehouse
- Nature of end-user tasks
- Constraints on resources
- Strategic view of the data warehouse prior to implementation
- Compatibility with existing systems
- Perceived ability of the in-house IT staff
- Technical Issues
- Social/political factors

Difficulty: Moderate

Page Reference: 51

66. *Describe various integration technologies enable data and metadata integration today.*

- Enterprise application integration - it provides a vehicle for pushing data from source systems into the data warehouse. It involves integrating application functionality, and is focused on sharing functionality across systems, thereby enabling flexibility and reuse.
- Service oriented architecture - coarse-grained services that are well-defined and documented.
- Enterprise information integration - an evolving tool space that promises real-time data integration from a variety of sources such as relational databases, Web services, and multi-dimensional databases.
- Extraction, transformation, and load - instrumental in the process and use of data warehouses.

Difficulty: Moderate

Page Reference: 53

67. *Describe various issues that affect whether an organization will purchase data transformation tools or build the transformation process itself.*

- Data transformation tools are expensive.
- They may have a long learning curve.
- It is difficult to measure how the IT organization is doing until it has learned to use the tools.

Difficulty: Easy

Page Reference: 55

68. *List the benefits of a hosted data warehouse.*

- Minimal investment in infrastructure
- Frees up capacity on in-house systems
- Frees up cash flow
- Powerful solutions are affordable
- Powerful solutions provide for growth
- Better quality equipment and software
- Faster connections
- Ability to access data from remote locations
- Allows a company focus on core business
- Meets storage needs for large volumes of data

Difficulty: Moderate

Page Reference: 62

69. *Describe a star schema.*

Data warehouse design is based upon the concept of dimensional modeling. The dimensional model is implemented with a star schema. The star schema is the means by which dimensional modeling is implemented. A star schema contains a central fact table. A fact table contains the attributes needed to perform decision analysis, descriptive attributes used for query reporting, and foreign keys to link to dimension tables. The fact tables describe what data can be analyzed; dimension tables describe how data can be analyzed.

Difficulty: Moderate

Page Reference: 62

70. *Identify at least five data warehouse best practices.*

- The project must fit with corporate strategy and business objectives.
- There must be complete buy-in to the project.
- Manage expectations.
- The data warehouse must be built incrementally.
- Build in adaptability.
- The project must be managed by both IT and business professionals.
- Develop a business supplier relationship.
- Only load data that have been cleaned and are of a quality understood by the organization.
- Do not overlook training requirements.
- Be politically aware.

Difficulty: Moderate

Page Reference: 64