Business Intelligence, 2e **(Turban/Sharda/Delen/King)**
Chapter 2 Data Warehousing

1) Before implementing an active data warehouse solution, DirecTV pulled data from the server every night in batch mode, a process that was taking too long and straining the system.
Answer: TRUE
Diff: 2 Page Ref: 30

2) A real-time data warehouse together with a decision support system that leverages integrated data can provide significant financial benefits for an organization.
Answer: TRUE
Diff: 2 Page Ref: 32

3) A data warehouse differs from an operational database in that most data warehouses have a product orientation and are designed to handle transactions that update the database.
Answer: FALSE
Diff: 1 Page Ref: 32

4) A data warehouse maintains historical data that do not necessarily provide current status, except in real-time systems.
Answer: TRUE
Diff: 2 Page Ref: 33

5) Once the data are entered into the data warehouse, users cannot change or update the data.
Answer: TRUE
Diff: 2 Page Ref: 33

6) There are three main types of data warehouses, which are data marts, operational data stores, and enterprise data warehouses.
Answer: TRUE
Diff: 2 Page Ref: 33

7) An independent data mart is a small warehouse designed for a strategic business unit (SBU) or a department whose source is an EDW.
Answer: FALSE
Diff: 2 Page Ref: 33

8) Operational data store is used for the medium- and long-term decisions associated with the enterprise data warehouse (EDW).
Answer: FALSE
Diff: 3 Page Ref: 33

9) The data for an oper mart come from an ODS.
Answer: TRUE
Diff: 2 Page Ref: 34

10) Effectiveness, extensibility, reusability, interoperability, efficiency and performance, evolution, entitlement, flexibility, segregation, user interface, versioning, versatility, and low maintenance cost are some of the key requirements for building a successful metadata-driven enterprise.
Answer: TRUE
Diff: 1 Page Ref: 36

11) There are several levels of metadata management maturity that describe where an organization is in terms of how and how well it uses its metadata.
Answer: TRUE
Diff: 2 Page Ref: 36

12) There are ethical considerations involved in the collection and ownership of the information contained in
metadata, including privacy and intellectual property issues.
Answer: TRUE
Diff: 2 Page Ref: 36

13) There are many metaware tools that business users can use to access data stored in the data repositories, including data mining, reporting tools, and data visualization.
Answer: FALSE
Diff: 1 Page Ref: 37

14) In a three-tier architecture, operational systems contain the data and the software for data acquisition in the first tier, the data warehouse is a second tier, and the third tier includes the DSS/BI/BA engine.
Answer: TRUE
Diff: 2 Page Ref: 38

15) The centralized data warehouse helps to simplify data management and administration and reduce data redundancy.
Answer: FALSE
Diff: 2 Page Ref: 42

16) Because of performance and data quality issues, most experts agree that federated approaches work well to replace data warehouses.
Answer: FALSE
Diff: 1 Page Ref: 42

17) According to conventional wisdom, independent data marts are a poor architectural solution.
Answer: TRUE
Diff: 2 Page Ref: 44

18) ETL tools transport data between sources and targets, document how data elements change as they move between source and target, exchange metadata with other applications, and administer all runtime processes and operations.
Answer: TRUE
Diff: 2 Page Ref: 48

19) A hosted data warehouse has less functionality than an onsite data warehouse, but it does not consume computer resources on client premises for computer upgrades, software licenses, in-house development, and in-house support and maintenance.
Answer: FALSE
Diff: 2 Page Ref: 54

20) A data warehouse needs to support scalability, which pertains to 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.
Answer: TRUE
Diff: 3 Page Ref: 62

21) When DirectTV decided to implement an active data warehouse solution, the goal of the new system was to send fresh data to the call center at least daily. Once the capabilities of the solutions became apparent, that goal:
A) dropped to fresh data of less than 15 minutes to improve responsiveness.
B) dropped to fresh data every 12 hours to improve responsiveness.
C) increased to every 2 days to reduce maintenance costs.
D) increased to every 5 days to reduce maintenance costs.
Answer: A
Diff: 2 Page Ref: 30

22) Data warehouse is a(n) _____, integrated, time-variant, nonvolatile 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
Diff: 2 Page Ref: 32

23) 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
Diff: 2 Page Ref: 33

24) A data warehouse contains _____ about how data are organized and how to use them effectively.
A) a data directory
B) a data index
C) data fields
D) metadata
Answer: D
Diff: 2 Page Ref: 33

25) 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 (an) _____.
A) data mart
B) operational data store
C) dependent data mart
D) independent data mart
Answer: D
Diff: 2 Page Ref: 33

26) Which of the following are created when operational data need to be analyzed multidimensionally?
A) Oper marts
B) Customer information file
C) Dependent data marts
D) Independent data marts
Answer: A
Diff: 2 Page Ref: 34

27) Which of the following is not a data source to a data warehouse?
A) ERP
B) Legacy
C) POS
D) ETL
Answer: D
Diff: 2 Page Ref: 37

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
Diff: 2 Page Ref: 38

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
Diff: 1 Page Ref: 38

30) The _____ have inconsistent data definitions and different dimensions and measures, making it difficult to analyze data across those marts.
A) enterprise data marts
B) operational data marts
C) dependent data marts
D) independent data marts
Answer: D
Diff: 2 Page Ref: 42

31) Users demanding access via PDAs and through speech recognition and synthesis is becoming more commonplace, further complicating _____ issues.
A) data extraction
B) data load
C) data integration
D) OLAP
Answer: C
Diff: 1 Page Ref: 45

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 multidimensional databases?
A) Information integration
B) Data management integration
C) SQL data integration
D) Enterprise information integration (EII)
Answer: D
Diff: 2 Page Ref: 47

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
Diff: 2 Page Ref: 47

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
Diff: 3 Page Ref: 49

35) Which of the following is not a direct benefit of a data warehouse?
A) End users can perform extensive analysis in numerous ways.
B) A consolidated view of the data provides a single version of the truth.
C) Simplified data access
D) Improved customer service and satisfaction
Answer: D
Diff: 2 Page Ref: 49

36) Guidelines that need to be considered when developing a vendor list include all of the following except:
A) financial strength
B) trade shows
C) ERP linkages
D) market share
Answer: B
Diff: 2 Page Ref: 52

37) A star schema contains a central _____ surrounded by several dimension tables.
A) database
B) fact table
C) data tree
D) data table
Answer: B
Diff: 2 Page Ref: 55

38) Which of the following is not one of the failure factors in data warehousing?
A) Cultural issues are ignored.
B) inappropriate architecture
C) unrealistic expectations
D) high levels of data summarization
Answer: D
Diff: 3 Page Ref: 61

39) _____ is a critical aspect of data warehousing that includes reconciling conflicting data definitions and formats organization-wide.
A) Data modification
B) Fact refinement
C) Data purification
D) Data cleansing
Answer: D
Diff: 2 Page Ref: 62

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
Diff: 2 Page Ref: 63

41) Data often are fragmented in distinct operational systems, so managers often make decisions with partial information at best. _____ 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.
Answer: Data warehousing
Diff: 1 Page Ref: 32

42) _____ 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.
Answer: Dependent data mart
Diff: 2 Page Ref: 33

43) _____ is a small data warehouse designed for a strategic business unit (SBU) or a department.
Answer: Independent data mart
Diff: 1 Page Ref: 33

44) _____ provides a fairly recent form of customer information files (CIF). It is a type of database often used as an interim staging area for a data warehouse.
Answer: Operational data store (ODS)
Diff: 2 Page Ref: 33

45) An _____ is a large-scale data warehouse that is utilized across the enterprise for decision support.
Answer: enterprise data warehouse (EDW)
Diff: 1 Page Ref: 34

46) In three-tier architecture for data warehouse, _____ 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.
Answer: operational systems
Diff: 3 Page Ref: 38

47) The _____ 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.
Answer: federated approach
Diff: 2 Page Ref: 42

48) _____ 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.
Answer: Data integration
Diff: 2 Page Ref: 45

49) EII (enterprise information integration) tools use predefined metadata to populate views that make integrated data appear relational to end-users. _____ may be the most important aspect of EII, because it allows data to be tagged either at the time of creation or later.
Answer: Extensible markup language (XML)
Diff: 2 Page Ref: 47

50) One of the benefits of a well-designed data warehouse is that business rules can be stored in a _____ repository and applied to the data warehouse centrally.
Answer: metadata
Diff: 3 Page Ref: 48

51) A data warehouse contains numerous _____ that define such things as how the data will be used, summarization rules, standardization of encoded attributes, and calculation rules.
Answer: business rules
Diff: 2 Page Ref: 48

52) The _____ is a scaled-down version of the data warehouse that centers on the requests of a specific department, such as marketing or sales.
Answer: data mart
Diff: 1 Page Ref: 52

53) The data warehouse design is based upon the concept of _____ modeling, which is a retrieval-based model that supports high-volume query access.
Answer: dimensional
Diff: 2 Page Ref: 55

54) A(n) _____ contains the attributes needed to perform decision analysis, descriptive attributes used for query reporting, and foreign keys to link to dimension tables.
Answer: fact table
Diff: 2 Page Ref: 55

55) A(n) _____ data warehouse has nearly the same, if not more, functionality as an on-site data warehouse, but it does not consume computer resources on client premises.
Answer: hosted
Diff: 2 Page Ref: 55

56) Once the data are properly stored in a data warehouse, that data can be used in various ways to support organizational _____.
Answer: decision making
Diff: 2 Page Ref: 56

57) During data modeling, expertise is required to determine what data are needed, define business rules associated with the data, and decide what _____ and other calculations may be necessary.
Answer: aggregations
Diff: 3 Page Ref: 56

58) The main issues pertaining to _____ 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.
Answer: scalability
Diff: 3 Page Ref: 64

59) _____ is the process of loading and provides data via the data warehouse as they become available.
Answer: Real-time data warehousing (RDW) or active data warehousing (ADW)
Diff: 2 Page Ref: 65

60) _____ is the person responsible for the administration and management of a data warehouse.
Answer: Data warehouse administrator (DWA)
Diff: 2 Page Ref: 70

61) List four fundamental characteristics of a data warehouse.
Answer:
- Subject-oriented
- Integrated
- Time variant (time series)
- Nonvolatile
- Real time
- Web based
- Contains internal and external data
- Contains metadata

Diff: 2 Page Ref: 32

62) Describe the major components of the data warehousing process.
Answer:
- *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.

Diff: 2 Page Ref: 37

63) What are the issues to consider when deciding which architecture for data warehousing to use?
Answer:
- Which database management system (DBMS) to use?
- Will parallel processing and/or partitioning be used?
- Will data migration tools be used to load the data warehouse?
- What tools will be used to support data retrieval and analysis?

Diff: 2 Page Ref: 40

64) Describe or sketch two alternative architectures to the basic architectural design types in data warehousing.
Answer: See Figure 2.5 in the textbook.
Diff: 2 Page Ref: 41

65) List five factors that potentially affect the architecture selection decision.
Answer: Any five of the following:
1. Information interdependence between organizational units
2. Upper management's information needs
3. Urgency of need for a data warehouse
4. Nature of end-user tasks
5. Constraints on resources
6. Strategic view of the data warehouse prior to implementation
7. Compatibility with existing systems
8. Perceived ability of the in-house IT staff
9. Technical Issues
10. Social/political factors

Diff: 2 Page Ref: 43

66) Describe various integration technologies that enable data and metadata integration.
Answer:
- *Enterprise application integration (EAI)*. 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 (SOA)*. Coarse-grained services that are well defined and documented
- *Enterprise information integration (EII)*. An evolving tool space that promises real-time data integration from a variety of sources such as relational databases, Web services, and multidimensional databases
- *Extraction, transformation, and load (ETL)*. Instrumental in the process and use of data warehouses

Diff: 3 Page Ref: 47

67) Describe various issues that affect whether an organization will purchase data transformation tools or build the transformation process itself.
Answer:
- Data transformation tools are expensive.
- Data transformation tools 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.

Diff: 2 Page Ref: 48

68) Describe a star schema.
Answer: 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.
Diff: 3 Page Ref: 55

69) List five benefits of a hosted data warehouse.
Answer: Any five of the following:
- 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

Diff: 2 Page Ref: 55

70) Identify five data warehouse best practices.
Answer: Any five of the following:
- The project must fit with corporate strategy and business objectives.
- There must be complete buy-in to the project.
- Manage user 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 cleansed and are of a quality understood by the organization.
- Do not overlook training requirements.
- Be politically aware.

Diff: 2 Page Ref: 61