Business Intelligence, 3e **(Sharda/Delen/Turban)**
Chapter 2 Data Warehousing

1) In the Isle of Capri case, the only capability added by the new software was increased processing speed of processing reports.
Answer: FALSE
Diff: 2 Page Ref: 38

2) The "islands of data" problem in the 1980s describes the phenomenon of unconnected data being stored in numerous locations within an organization.
Answer: TRUE
Diff: 2 Page Ref: 41

3) Subject oriented databases for data warehousing are organized by detailed subjects such as disk drives, computers, and networks.
Answer: FALSE
Diff: 2 Page Ref: 42

4) Data warehouses are subsets of data marts.
Answer: FALSE
Diff: 1 Page Ref: 43

5) One way an operational data store differs from a data warehouse is the recency of their data.
Answer: TRUE
Diff: 2 Page Ref: 43-44

6) Organizations seldom devote a lot of effort to creating metadata because it is not important for the effective use of data warehouses.
Answer: FALSE
Diff: 2 Page Ref: 46

7) Without middleware, different BI programs cannot easily connect to the data warehouse.
Answer: TRUE
Diff: 2 Page Ref: 48-49

8) Two-tier data warehouse/BI infrastructures offer organizations more flexibility but cost more than three-tier ones.
Answer: FALSE
Diff: 2 Page Ref: 50

9) Moving the data into a data warehouse is usually the easiest part of its creation.
Answer: FALSE
Diff: 2 Page Ref: 52

10) The hub-and-spoke data warehouse model uses a centralized warehouse feeding dependent data marts.
Answer: TRUE
Diff: 2 Page Ref: 52

11) Because of performance and data quality issues, most experts agree that the federated architecture should supplement data warehouses, not replace them.
Answer: TRUE
Diff: 2 Page Ref: 54

12) Bill Inmon advocates the data mart bus architecture whereas Ralph Kimball promotes the hub-and-spoke architecture, a data mart bus architecture with conformed dimensions.
Answer: FALSE
Diff: 2 Page Ref: 55

13) The ETL process in data warehousing usually takes up a small portion of the time in a data-centric project.
Answer: FALSE
Diff: 3 Page Ref: 59

14) In the Starwood Hotels case, up-to-date data and faster reporting helped hotel managers better manage their occupancy rates.
Answer: TRUE
Diff: 1 Page Ref: 66

15) Large companies, especially those with revenue upwards of $500 million consistently reap substantial cost savings through the use of hosted data warehouses.
Answer: FALSE
Diff: 2 Page Ref: 68

16) OLTP systems are designed to handle ad hoc analysis and complex queries that deal with many data items.
Answer: FALSE
Diff: 2 Page Ref: 70

17) The data warehousing maturity model consists of six stages: prenatal, infant, child, teenager, adult, and sage.
Answer: TRUE
Diff: 2 Page Ref: 73

18) A well-designed data warehouse means that user requirements do not have to change as business needs change.
Answer: FALSE
Diff: 2 Page Ref: 77

19) Data warehouse administrators (DWAs) do not need strong business insight since they only handle the technical aspect of the infrastructure.
Answer: FALSE
Diff: 2 Page Ref: 82

20) Because the recession has raised interest in low-cost open source software, it is now set to replace traditional enterprise software.
Answer: FALSE
Diff: 2 Page Ref: 83

21) The "single version of the truth" embodied in a data warehouse such as Capri Casinos' means all of the following EXCEPT
A) decision makers get to see the same results to queries.
B) decision makers have the same data available to support their decisions.
C) decision makers get to use more dependable data for their decisions.
D) decision makers have unfettered access to all data in the warehouse.
Answer: D
Diff: 3 Page Ref: 40

22) Operational or transaction databases are product oriented, handling transactions that update the database. In contrast, data warehouses are
A) subject-oriented and nonvolatile.
B) product-oriented and nonvolatile.
C) product-oriented and volatile.
D) subject-oriented and volatile.
Answer: A
Diff: 3 Page Ref: 40

23) Which kind of data warehouse is created separately from the enterprise data warehouse by a department and not reliant on it for updates?
A) sectional data mart
B) public data mart
C) independent data mart
D) volatile data mart
Answer: C
Diff: 2 Page Ref: 43

24) All of the following statements about metadata are true EXCEPT
A) metadata gives context to reported data.
B) there may be ethical issues involved in the creation of metadata.
C) metadata helps to describe the meaning and structure of data.
D) for most organizations, data warehouse metadata are an unnecessary expense.
Answer: D
Diff: 2 Page Ref: 45-46

25) A Web client that connects to a Web server, which is in turn connected to a BI application server, is reflective of a
A) one tier architecture.
B) two tier architecture.
C) three tier architecture.
D) four tier architecture.
Answer: C
Diff: 2 Page Ref: 49-50

26) Which of the following BEST enables a data warehouse to handle complex queries and scale up to handle many more requests?
A) use of the web by users as a front-end
B) parallel processing
C) Microsoft Windows
D) a larger IT staff
Answer: B
Diff: 3 Page Ref: 51

27) Which data warehouse architecture uses metadata from existing data warehouses to create a hybrid logical data warehouse comprised of data from the other warehouses?
A) independent data marts architecture
B) centralized data warehouse architecture
C) hub-and-spoke data warehouse architecture
D) federated architecture
Answer: D
Diff: 3 Page Ref: 53

28) Which data warehouse architecture uses a normalized relational warehouse that feeds multiple data marts?
A) independent data marts architecture
B) centralized data warehouse architecture
C) hub-and-spoke data warehouse architecture
D) federated architecture
Answer: C
Diff: 3 Page Ref: 53

29) Which approach to data warehouse integration focuses more on sharing process functionality than data across systems?
A) extraction, transformation, and load
B) enterprise application integration
C) enterprise information integration
D) enterprise function integration
Answer: B
Diff: 3 Page Ref: 58-59

30) In which stage of extraction, transformation, and load (ETL) into a data warehouse are data aggregated?
A) transformation
B) extraction
C) load
D) cleanse
Answer: A
Diff: 3 Page Ref: 59

31) In which stage of extraction, transformation, and load (ETL) into a data warehouse are anomalies detected and corrected?
A) transformation
B) extraction
C) load
D) cleanse
Answer: D
Diff: 3 Page Ref: 59

32) Data warehouses provide direct and indirect benefits to using organizations. Which of the following is an indirect benefit of data warehouses?
A) better and more timely information
B) extensive new analyses performed by users
C) simplified access to data
D) improved customer service
Answer: D
Diff: 3 Page Ref: 61

33) All of the following are benefits of hosted data warehouses EXCEPT
A) smaller upfront investment.
B) better quality hardware.
C) greater control of data.
D) frees up in-house systems.
Answer: C
Diff: 2 Page Ref: 68

34) When representing data in a data warehouse, using several dimension tables that are each connected only to a fact table means you are using which warehouse structure?
A) star schema
B) snowflake schema
C) relational schema
D) dimensional schema
Answer: A
Diff: 3 Page Ref: 68-69

35) When querying a dimensional database, a user went from summarized data to its underlying details. The function that served this purpose is
A) dice.
B) slice.
C) roll-up.
D) drill down.
Answer: D
Diff: 3 Page Ref: 70-71

36) Which of the following online analytical processing (OLAP) technologies does NOT require the precomputation and storage of information?
A) MOLAP
B) ROLAP
C) HOLAP
D) SQL
Answer: B
Diff: 2 Page Ref: 71-72

37) Active data warehousing can be used to support the highest level of decision making sophistication and power. The major feature that enables this in relation to handling the data is
A) country of (data) origin.
B) nature of the data.
C) speed of data transfer.
D) source of the data.
Answer: C
Diff: 2 Page Ref: 77

38) Which of the following statements is more descriptive of active data warehouses in contrast with traditional data warehouses?
A) strategic decisions whose impacts are hard to measure
B) detailed data available for strategic use only
C) large numbers of users, including operational staffs
D) restrictive reporting with daily and weekly data currency
Answer: C
Diff: 3 Page Ref: 81

39) How does the use of cloud computing affect the scalability of a data warehouse?
A) Cloud computing vendors bring as much hardware as needed to users' offices.
B) Hardware resources are dynamically allocated as use increases.
C) Cloud vendors are mostly based overseas where the cost of labor is low.
D) Cloud computing has little effect on a data warehouse's scalability.
Answer: B
Diff: 3 Page Ref: 83

40) All of the following are true about in-database processing technology EXCEPT
A) it pushes the algorithms to where the data is.
B) it makes the response to queries much faster than conventional databases.
C) it is often used for apps like credit card fraud detection and investment risk management.
D) it is the same as in-memory storage technology.
Answer: D
Diff: 3 Page Ref: 85

41) With _____ data flows, managers can view the current state of their businesses and quickly identify problems.
Answer: real-time
Diff: 2 Page Ref: 40

42) In _____ oriented data warehousing, operational databases are tuned to handle transactions that update the database.
Answer: product
Diff: 2 Page Ref: 42

43) The three main types of data warehouses are data marts, operational _____, and enterprise data warehouses.
Answer: data stores
Diff: 2 Page Ref: 43

44) _____ describe the structure and meaning of the data, contributing to their effective use.
Answer: Metadata
Diff: 1 Page Ref: 45

45) Most data warehouses are built using _____ database management systems to control and manage the data.
Answer: relational
Diff: 2 Page Ref: 51

46) A(n) _____ architecture is used to build a scalable and maintainable infrastructure that includes a centralized data warehouse and several dependent data marts.
Answer: hub-and-spoke
Diff: 2 Page Ref: 52

47) The _____ data warehouse architecture involves integrating disparate systems and analytical resources from multiple sources to meet changing needs or business conditions.
Answer: federated
Diff: 2 Page Ref: 54

48) Data _____ comprises data access, data federation, and change capture.
Answer: integration
Diff: 3 Page Ref: 57

49) _____ is a mechanism that integrates application functionality and shares functionality (rather than data) across systems, thereby enabling flexibility and reuse.
Answer: Enterprise application integration (EAI)
Diff: 3 Page Ref: 58

50) _____ is a mechanism for pulling data from source systems to satisfy a request for information. It 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.
Answer: Enterprise information integration (EII)
Diff: 3 Page Ref: 59

51) Performing extensive _____ to move data to the data warehouse may be a sign of poorly managed data and a fundamental lack of a coherent data management strategy.
Answer: extraction, transformation, and load (ETL)
Diff: 3 Page Ref: 61

52) The _____ Model, also known as the EDW approach, emphasizes top-down development, employing established database development methodologies and tools, such as entity-relationship diagrams (ERD), and an adjustment of the spiral development approach.
Answer: Inmon
Diff: 2 Page Ref: 65

53) The _____ Model, also known as the data mart approach, is a "plan big, build small" approach. A data mart is a subject-oriented or department-oriented data warehouse. It is a scaled-down version of a data warehouse that focuses on the requests of a specific department, such as marketing or sales.
Answer: Kimball
Diff: 2 Page Ref: 65

54) _____ modeling is a retrieval-based system that supports high-volume query access.
Answer: Dimensional
Diff: 2 Page Ref: 68

55) Online _____ is arguably the most commonly used data analysis technique in data warehouses.
Answer: analytical processing
Diff: 1 Page Ref: 69

56) Online _____ is a term used for a transaction system that is primarily responsible for capturing and storing data related to day-to-day business functions such as ERP, CRM, SCM, and point of sale.
Answer: transaction processing
Diff: 2 Page Ref: 69

57) In the Michigan State Agencies case, the approach used was a(n) _____ one, instead of developing separate BI/DW platforms for each business area or state agency.
Answer: enterprise
Diff: 2 Page Ref: 76

58) The role responsible for successful administration and management of a data warehouse is the _____, who should be familiar with high-performance software, hardware, and networking technologies, and also possesses solid business insight.
Answer: data warehouse administrator (DWA)
Diff: 2 Page Ref: 82

59) _____, or "The Extended ASP Model," is a creative way of deploying information system applications where the provider licenses its applications to customers for use as a service on demand (usually over the Internet)
Answer: SaaS (software as a service)
Diff: 2 Page Ref: 83

60) _____ (also called in-database analytics) refers to the integration of the algorithmic extent of data analytics into data warehouse.
Answer: In-database processing
Diff: 2 Page Ref: 85

61) What is the definition of a data warehouse (DW) in simple terms?
Answer: In simple terms, a data warehouse (DW) is a pool of data produced to support decision making; it is also a repository of current and historical data of potential interest to managers throughout the organization.
Diff: 2 Page Ref: 40

62) A common way of introducing data warehousing is to refer to its fundamental characteristics. Describe three characteristics of data warehousing.
Answer:
· **Subject oriented.** Data are organized by detailed subject, such as sales, products, or customers, containing only information relevant for decision support.
· **Integrated.** Integration is closely related to subject orientation. Data warehouses must place data from different sources into a consistent format. To do so, they must deal with naming conflicts and discrepancies among units of measure. A data warehouse is presumed to be totally integrated.
· **Time variant (time series).** A warehouse maintains historical data. The data do not necessarily provide current status (except in real-time systems). They detect trends, deviations, and long-term relationships for forecasting and comparisons, leading to decision making. Every data warehouse has a temporal quality. Time is the one important dimension that all data warehouses must support. Data for analysis from multiple sources contains multiple time points (e.g., daily, weekly, monthly views).
· **Nonvolatile.** After data are entered into a data warehouse, users cannot change or update the data. Obsolete data are discarded, and changes are recorded as new data.
· **Web based.** Data warehouses are typically designed to provide an efficient computing environment for Web-based applications.
· **Relational/multidimensional.** A data warehouse uses either a relational structure or a multidimensional structure. A recent survey on multidimensional structures can be found in Romero and Abelló (2009).
· **Client/server.** A data warehouse uses the client/server architecture to provide easy access for end users.
· **Real time.** Newer data warehouses provide real-time, or active, data-access and analysis capabilities (see Basu, 2003; and Bonde and Kuckuk, 2004).
· **Include metadata.** A data warehouse contains metadata (data about data) about how the data are organized and how to effectively use them.
Diff: 3 Page Ref: 42-43

63) What is the definition of a data mart?
Answer: A data mart is a subset of a data warehouse, typically consisting of a single subject area (e.g., marketing, operations). Whereas a data warehouse combines databases across an entire enterprise, a data mart is usually smaller and focuses on a particular subject or department.
Diff: 2 Page Ref: 43

64) Mehra (2005) indicated that few organizations really understand metadata, and fewer understand how to design and implement a metadata strategy. How would you describe metadata?
Answer: Metadata are data about data. Metadata describe the structure of and some meaning about data, thereby contributing to their effective or ineffective use.
Diff: 2 Page Ref: 45-46

65) According to Kassam (2002), business metadata comprise information that increases our understanding of traditional (i.e., structured) data. What is the primary purpose of metadata?
Answer: The primary purpose of metadata should be to provide context to the reported data; that is, it provides enriching information that leads to the creation of knowledge.
Diff: 2 Page Ref: 46

66) In the MultiCare case, how was data warehousing able to reduce septicemia mortality rates in MultiCare hospitals?
Answer:

- The Adaptive Data Warehouse™ organized and simplified data from multiple data sources across the continuum of care. It became the single source of truth required to see care improvement opportunities and to measure change, integrated teams consisting of clinicians, technologists, analysts, and quality personnel were essential for accelerating MultiCare's efforts to reduce septicemia mortality.
- Together the collaborative effort addressed three key bodies of work–standard of care definition, early identification, and efficient delivery of defined-care standard.

Diff: 3 Page Ref: 47-48

67) Briefly describe four 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 may also come from an OLTP or ERP system.
- **Data extraction and transformation.** Data are extracted and properly transformed using custom-written or commercial ETL software.
- **Data loading.** Data are loaded into a staging area, where they are transformed and cleansed. The data are then ready to load into the data warehouse and/or data marts.
- **Comprehensive database.** Essentially, this is the EDW to support all decision analysis by providing relevant summarized and detailed information originating from many different sources.
- **Metadata.** Metadata include software programs about data and rules for organizing data summaries that are easy to index and search, especially with Web tools.
- **Middleware tools**. Middleware tools enable access to the data warehouse. There are many front-end applications that business users can use to interact with data stored in the data repositories, including data mining, OLAP, reporting tools, and data visualization tools.

Diff: 2 Page Ref: 48-49

68) There are several basic information system architectures that can be used for data warehousing. What are they?
Answer: Generally speaking, these architectures are commonly called client/server or n-tier architectures, of which two-tier and three-tier architectures are the most common, but sometimes there is simply one tier.
Diff: 2 Page Ref: 49-50

69) More data, coming in faster and requiring immediate conversion into decisions, means that organizations are confronting the need for real-time data warehousing (RDW). How would you define real-time data warehousing?
Answer: Real-time data warehousing, also known as active data warehousing (ADW), is the process of loading and providing data via the data warehouse as they become available.
Diff: 2 Page Ref: 77

70) Mention briefly some of the recently popularized concepts and technologies that will play a significant role in defining the future of data warehousing.
Answer:

- Sourcing (mechanisms for acquisition of data from diverse and dispersed sources):
  - Web, social media, and Big Data
  - Open source software
  - SaaS (software as a service)
  - Cloud computing
- Infrastructure (architectural–hardware and software–enhancements):
  - Columnar (a new way to store and access data in the database)
  - Real-time data warehousing
  - Data warehouse appliances (all-in-one solutions to DW)
  - Data management technologies and practices
  - In-database processing technology (putting the algorithms where the data is)
  - In-memory storage technology (moving the data in the memory for faster processing)
  - New database management systems
  - Advanced analytics

Diff: 3 Page Ref: 83-86