***Business Intelligence***

**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

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

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

4) A data warehouse maintains historical data that do not necessarily provide current status, except in real-time systems.

Answer: TRUE

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

Answer: TRUE

6) There are three main types of data warehouses, which are data marts, operational data stores, and enterprise data warehouses.

Answer: TRUE

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

8) Operational data store is used for the medium- and long-term decisions associated with the enterprise data warehouse (EDW).

Answer: FALSE

9) The data for an oper mart come from an ODS.

Answer: TRUE

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

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

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

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

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

Page Ref: 38

15) The centralized data warehouse helps to simplify data management and administration and reduce data redundancy.

Answer: FALSE

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

Page Ref: 42

17) According to conventional wisdom, independent data marts are a poor architectural solution.

Answer: TRUE

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

43) \_\_\_\_\_\_\_\_ is a small data warehouse designed for a strategic business unit (SBU) or a department.

Answer: Independent data mart

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)

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

Answer: enterprise data warehouse (EDW)

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

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

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

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)

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

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

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

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

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

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

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

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

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

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)

Page Ref: 65

60) \_\_\_\_\_\_\_\_ is the person responsible for the administration and management of a data warehouse.

Answer: Data warehouse administrator (DWA)

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

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.

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?

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.

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

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

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.

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.

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

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.

***Business Intelligence***

**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

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

3) Subject oriented databases for data warehousing are organized by detailed subjects such as disk drives, computers, and networks.

Answer: FALSE

4) Data warehouses are subsets of data marts.

Answer: FALSE

5) One way an operational data store differs from a data warehouse is the recency of their data.

Answer: TRUE

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

7) Without middleware, different BI programs cannot easily connect to the data warehouse.

Answer: TRUE

8) Two-tier data warehouse/BI infrastructures offer organizations more flexibility but cost more than three-tier ones.

Answer: FALSE

9) Moving the data into a data warehouse is usually the easiest part of its creation.

Answer: FALSE

10) The hub-and-spoke data warehouse model uses a centralized warehouse feeding dependent data marts.

Answer: TRUE

11) Because of performance and data quality issues, most experts agree that the federated architecture should supplement data warehouses, not replace them.

Answer: TRUE

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

13) The ETL process in data warehousing usually takes up a small portion of the time in a data-centric project.

Answer: FALSE

14) In the Starwood Hotels case, up-to-date data and faster reporting helped hotel managers better manage their occupancy rates.

Answer: TRUE

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

16) OLTP systems are designed to handle ad hoc analysis and complex queries that deal with many data items.

Answer: FALSE

17) The data warehousing maturity model consists of six stages: prenatal, infant, child, teenager, adult, and sage.

Answer: TRUE

18) A well-designed data warehouse means that user requirements do not have to change as business needs change.

Answer: FALSE

19) Data warehouse administrators (DWAs) do not need strong business insight since they only handle the technical aspect of the infrastructure.

Answer: FALSE

20) Because the recession has raised interest in low-cost open source software, it is now set to replace traditional enterprise software.

Answer: FALSE

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Page Ref: 85

41) With \_\_\_\_\_\_\_\_ data flows, managers can view the current state of their businesses and quickly identify problems.

Answer: real-time

Page Ref: 40

42) In \_\_\_\_\_\_\_\_ oriented data warehousing, operational databases are tuned to handle transactions that update the database.

Answer: product

Page Ref: 42

43) The three main types of data warehouses are data marts, operational \_\_\_\_\_\_\_\_, and enterprise data warehouses.

Answer: data stores

Page Ref: 43

44) \_\_\_\_\_\_\_\_ describe the structure and meaning of the data, contributing to their effective use.

Answer: Metadata

Page Ref: 45

45) Most data warehouses are built using \_\_\_\_\_\_\_\_ database management systems to control and manage the data.

Answer: relational

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

47) The \_\_\_\_\_\_\_\_ data warehouse architecture involves integrating disparate systems and analytical resources from multiple sources to meet changing needs or business conditions.

Answer: federated

48) Data\_\_\_\_\_\_\_\_ comprises data access, data federation, and change capture.

Answer: integration

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)

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)

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)

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

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

54) \_\_\_\_\_\_\_\_modeling is a retrieval-based system that supports high-volume query access.

Answer: Dimensional

55) Online \_\_\_\_\_\_\_\_ is arguably the most commonly used data analysis technique in data warehouses.

Answer: analytical processing

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

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

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)

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)

60) \_\_\_\_\_\_\_\_ (also called in-database analytics) refers to the integration of the algorithmic extent of data analytics into data warehouse.

Answer: In-database processing

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.

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.

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.

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.

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.

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

Answer:

∙ The Adaptive Data Warehousetm 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.

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.

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.

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.

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):

o Web, social media, and Big Data

o Open source software

o SaaS (software as a service)

o Cloud computing

∙ Infrastructure (architectural–hardware and software–enhancements):

o Columnar (a new way to store and access data in the database)

o Real-time data warehousing

o Data warehouse appliances (all-in-one solutions to DW)

o Data management technologies and practices

o In-database processing technology (putting the algorithms where the data is)

o In-memory storage technology (moving the data in the memory for faster processing)

o New database management systems

o Advanced analytics