Distributed Databases

1) A(n) _____ is a database stored on multiple computers in multiple locations that are NOT connected by a data communications link.
A) distributed database
B) decentralized database
C) unlinked database
D) data repository

2) Which of the following are business conditions that encourage the use of distributed databases?
A) Business unit autonomy
B) Data sharing needs
C) Data communication reliability
D) All of the above

3) Which of the following environments uses the same DBMS at each node with a central or master DBMS coordinating database access across nodes?
A) Centralized; maximum
B) Centralized; minimum
C) Homogeneous; nonautonomous
D) Federated; nonautonomous

4) Which of the following environments uses a different DBMS at each node and supports local databases for unique data requests?
A) Gateways; local
B) Centralized; minimum
C) Homogeneous; cross-linked
D) Heterogeneous; federated

5) Which of the following supports a simple path to other databases, without the benefits of one logical database?
A) Linked; differential
B) Gateways
C) Autonomous; cross-linked
D) Federated; nonautonomous

6) Which of the following characterizes homogeneous environments?
A) Contains simple pathways
B) Same DBMS used at all locations
C) Some users require only local access
D) Cross-linked systems easily communicate

7) _____ is a design goal for a distributed database, which says a user does not need to know the location of data to use the data.
A) Location transparency

B) Location autonomy
C) Location dependency
D) Distributed locatability

8) Which of the following is true of distributed databases?
A) Less reliable
B) Better local control
C) Slower response time
D) None of the above
9) Which of the following is true of data replication?
A) Additional storage requirements
B) Faster response
C) Node decoupling
D) All of the above

10) All of the following are disadvantages of data replication EXCEPT:
A) storage requirements.
B) reduced network traffic at prime time.
C) complexity and cost of updating.
D) none of the above.

11) Which of the following is an advantage to data replication?
A) Reliability
B) Fast response
C) Node decoupling
D) All of the above

12) Snapshot replication is most appropriate for:
A) a data warehouse application.
B) an application where the processing of data follows a workflow across the business units.
C) a transaction processing system.
D) none of the above.

13) One way to generate store and forward messages for completed transactions to be broadcast across a network is through the use of:
A) stored procedures.
B) triggers.
C) functions.
D) SQL statements.

14) With a pull strategy of replication, the _____ node determines when a database is updated.
A) user
B) source
C) target
D) all of the above

15) Replication should be used when:
A) data currency and timing don't matter.
B) there is a highly heterogeneous network.
C) there are no or few triggers.
D) all of the above.

16) Which of the following is true about horizontal partitioning?
A) Data can be stored to optimize local access.
B) Data are less secure.
C) There is consistent access speed.
D) None of the above.

17) All of the following are advantages of vertical partitioning EXCEPT:
A) efficiency.
B) security.
C) ease of querying.
D) easier to set up than horizontal partitioning.

18) A centralized strategy has _____ expandability.
A) excellent
B) good
C) poor
D) medium

19) A synchronized replication strategy has a(n) _____ reliability.
A) excellent
B) good
C) fair
D) medium

20) An integrated partition strategy is _____ to manage.
A) excellent
B) difficult
C) easy
D) medium

21) Which of the following are factors in deciding on database distribution strategies?
A) Organizational forces
B) Frequency of data access
C) Reliability needs
D) All of the above

22) A distributed database must:
A) keep track of data locations.
B) present a single logical database that is physically distributed.
C) provide consistency among data copies.
D) all of the above.

23) In a distributed database, a transaction that requires reference to data at one or more nonlocal sites is called a _____ transaction.
A) link
B) tight
C) global
D) loose

24) With _____, users can act as if all the data were located at a single node.
A) location transparency
B) local autonomy
C) client-based control
D) all of the above

25) A design goal for distributed databases to allow programmers to treat a data item replicated at several sites as though it were at one site is called:
A) location transparency.
B) replication accessibility.
C) replication transparency.
D) data accessibility.

26) With _____, all of the actions of a transaction are either committed or not committed.
A) location transparency
B) replication transparency
C) failure transparency
D) commit transparency

27) An optimization strategy that allows sites that can update to proceed and other sites to catch up is called:
A) read-only commit.
B) lazy commit.
C) sequenced commit.
D) linear set up.

28) _____ ensures that a transaction is successfully completed or else it is aborted.
A) Failure protocol
B) TCP/IP protocol
C) Commit protocol
D) None of the above

29) A design goal for distributed databases that states that although a distributed database runs many transactions, it appears that a given transaction is the only one in the system is called:
A) linear run.
B) concurrency transparency.
C) transaction CPU transparency.
D) location transparency.

30) The step in which a distributed database decides the order in which to execute the distributed query is called:
A) decomposition.
B) localization.
C) step optimization.
D) none of the above.

31) A joining operation in which only the joining attribute from one site is transmitted to the other site is called a(n):
A) attribute linkup.
B) key data element transfer.
C) key data element join.
D) semijoin.

32) The sequence of instructions required to process a transaction is called the:
A) unit of work.
B) logical program unit.
C) rollback.
D) commit.
33) A _____ allows a single SQL statement to refer to tables in more than one remote DBMS.
A) distributed UOW
B) join SQL statement
C) distributed request
D) data transfer command

State True/False

34) A distributed database is *not* just a collection of files. **T**

35) Distributed databases make data sharing more difficult because of the data communication problems. **F**

36) Distributed databases do not easily satisfy both transaction and analytical processing systems. **F**

37) In a homogeneous environment, the same DBMS is used at each location. **T**

38) In a heterogeneous environment, not all users use the same DBMS. **T**

39) Users of distributed database systems always need to know the location of the data. **F**

40) Local autonomy means that data are not accessible from remote sites. **F**

41) With an asynchronous distributed database, all data across the network is continuously kept up to date. **F**

42) A distributed database may require more costly and complex software than a centralized database. **T**

43) Replicated databases usually result in tightly coupled nodes. **F**

44) Replication is often used for noncollaborative data. **T**

45) Replication should NOT be used if timely updates are important. **T**

46) Applications such as decision support or data warehousing often do not require current data and are supported by periodic snapshots. **T**

47) Shared ownership of data is most appropriate as business activities move across time zones.**F**

48) One way to generate messages for near-real-time replication is through the use of database triggers. **T**

49) Applications that can tolerate out-of-date data are not the best candidates for data replication. **F**

50) Horizontal partitioning offers increased efficiency because data are close to use areas. **T**

---

1) Which of the following is a reason for using an object-oriented design?

A) The analysis model is not formal enough to be implemented in a programming language.
B) The actual system must be adapted to the environment in which the system will actually be implemented.
C) The analysis results can be validated using object-oriented design.
D) All of the above.

2) All of the following are benefits of object-oriented modeling EXCEPT:
A) the ability to tackle more challenging problem domains.
B) increased consistency among analysis, design, and programming activities.
C) decreased communication among the users, analysts, designers, and programmers.
D) reusability of analysis, design, and programming results.
AACSB: Use of Information Technology

3) The Unified Modeling Language:
A) is a notation useful for graphically depicting an object-oriented analysis or design model.
B) allows one to capture design decisions of a system.
C) promotes communication among key personnel involved in development.
D) all of the above.

4) A(n) _____ is a concept, abstraction, or thing that has a state, behavior, and identity.
A) relationship
B) attribute
C) key
D) object

5) A(n) _____ encompasses an object's properties and the values of those properties.
A) state
B) attribute
C) method
D) behavior

6) Which of the following refers to a set of objects that share common structures and behaviors?
A) Supertype
B) Object class
C) Model
D) Entity group

7) A diagram that shows the static structure of an object-oriented model is called a(n):
A) structure diagram.
B) class diagram.
C) entity diagram.
D) none of the above.

8) A graph of instances that are compatible within a class diagram is called a(n):
A) object group.

B) method chart.
C) object diagram.
D) ERD.

9) Which of the following is a function or service provided by all instances of a class?
A) Operation
B) Implementation
C) Class group
D) Query

10) _____ is the technique of hiding the internal implementation details of an object from its external view.
A) Grouping
B) Encapsulation
C) Classification
D) Mining

11) Which of the following operations does NOT alter the state of an object?
A) Update
B) Query
C) Constructor
D) None of the above

12) An operation that creates a new instance of a class is called a(n):
A) query operation.
B) update operation.
C) scope operation.
D) constructor operation.

13) Which type of operation has side effects?
A) Update
B) Append
C) Query
D) Scope

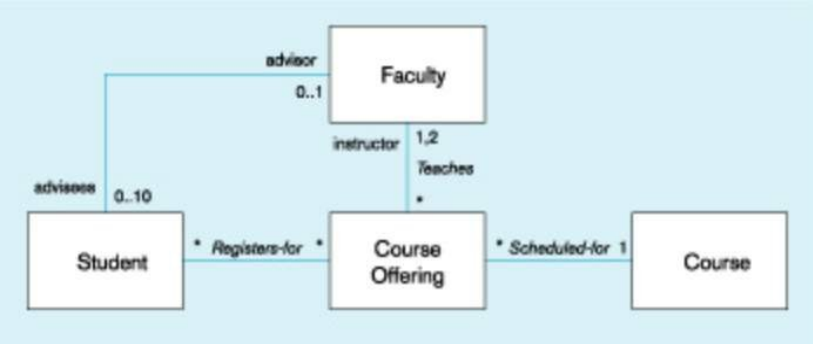
14) An operation that applies to a class rather than an object instance is a(n):
A) constructor operation.
B) query operation.
C) class-scope operation.
D) update operation.

15) A(n) _____ is shown as a solid line between the participating classes.
A) connector
B) update
C) entity
D) association

16) The end of an association where it connects to a class is called a(n):
A) connector.
B) terminator.
C) initiator.
D) association role.

17) _____ indicates how many objects participate in a given relationship.
A) Bound
B) Multiplicity
C) Role
D) Relationship

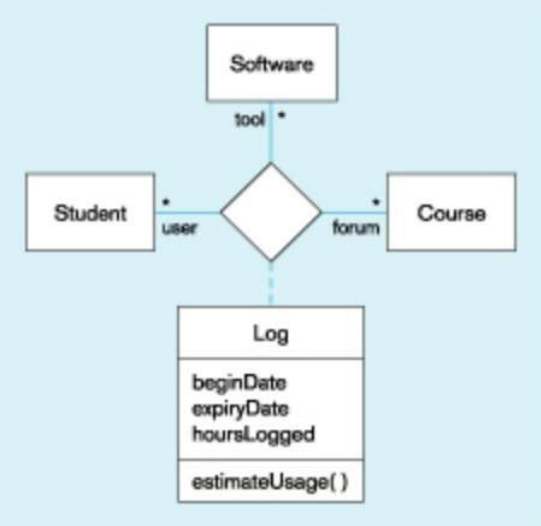
18) In the figure below, which of the following is true?



A) A faculty may advise only one student.
B) A faculty may advise up to a maximum of 10 students.
C) A course offering may be offered for multiple courses.
D) A student can have more than one advisor.

19) An association _____ is an association that has attributes or operations of its own.
A) link
B) group
C) class
D) body

20) In the figure below, which of the following is true?



A) Students use various software tools for different courses.
B) Students can only register for one course.
C) Students use only one type of software.
D) Each course uses specific software.

21) A class that has direct instances is called a(n) _____ class.
A) abstract
B) multiplicity
C) large
D) concrete

22) According to the UML Notation Guide, overlapping means:
A) a descendant may be descended from more than one of the subclasses.
B) a descendant may not be descended from more than one of the subclasses.
C) all subclasses have been specified, and no others are expected.
D) the list of objects is stable, but relationships will change.

23) According to the UML Notation Guide, complete means:
A) a descendant may come from more than one supertype.
B) a descendant may be not descended from more than one of the subclasses.
C) all subclasses have been specified, and no others are expected.
D) the list of objects is *not* stable.

24) A(n) _____ is an attribute of a class that specifies a value common to an entire class.
A) completion attribute
B) class-scope attribute
C) overall descriptor
D) summary descriptor

25) A(n) _____ defines the form or protocol of an operation, but not its implementation.
A) meta-operation
B) inquisitor

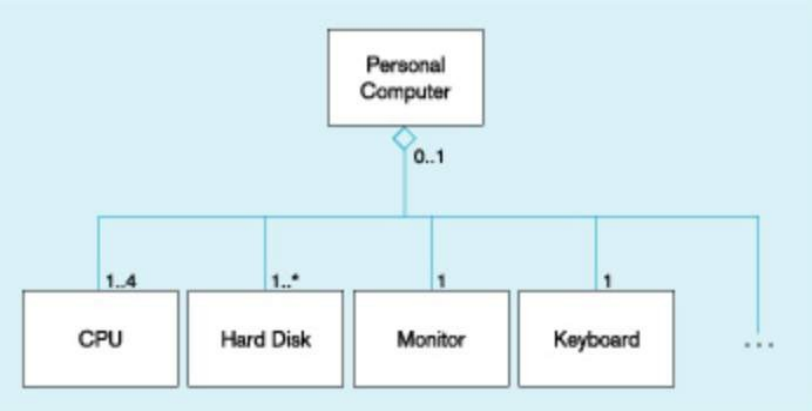
C) abstract operation
D) operation descriptor

26) A _____ is the implementation of an operation.
A) meta-operation
B) method
C) query
D) constructor

27) _____ means that the same operation can apply to two or more classes in different ways.
A) **Polymorphism**
B) Inheritance
C) Combined operations
D) Checkpoint

28) The process of replacing a method inherited from a superclass by a more specific implementation of the method in a subclass is called:
A) inheritance.
B) encapsulation.
C) polymorphism.
D) overriding.

29) When an object is an instance of more than one class, it is called:
A) multiplicity.
B) polymorphism.
C) multiple classification.
D) multiple associations.
30) In the figure below, what relationship is shown?



A) Overlaps
B) Disjoints
C) Rejoinders
D) Aggregation

31) A part object which belongs to only one whole object and which lives and dies with the

whole object is called a(n):
A) dependent.
B) weak entity.
C) composition.
D) declarative.

32) An object encapsulates both data and behavior. **T**

33) A state represents how an object acts. **F**

34) An object class is a set of objects that share a common structure and behavior. **T**

35) A class diagram shows the dynamic structure of an object-oriented model. **F**

36) An operation is a function or service that is provided by all of the classes. **T**

37) Encapsulation is the technique of hiding the internal implementation details of an object from its external view. **T**

38) A constructor operation does not alter an instance of a class. **F**
39) A class-scope operation applies to a class rather than an object instance. **T**

40) An association role is the end of an association where it connects to a class. **T**

---

1) Which of the following is true of poor data and/or database administration?
A) Data timing problems
B) Multiple entity definitions
C) Unknown meanings of stored data
D) All of the above

2) _____ is a technical function responsible for database design, security, and disaster recovery.
A) Data administration
B) Database administration
C) Tech support
D) Operations

3) An information repository supplies information to:
A) users.
B) automated CASE tools.
C) database management systems.
D) all of the above.

4) Which of the following functions do cost/benefit models?
A) Database planning
B) Database analysis

C) Database design
D) Operations
5) Which of the following functions develop integrity controls?
A) Database planning
B) Database analysis
C) Database implementation
D) Database design

6) Which of the following functions model business rules?
A) Database planning
B) Database analysis
C) Database design
D) Operations

7) A data warehouse administrator has which of the following responsibilities?
A) Build and administer an environment supportive of decision-support applications
B) Build a stable architecture for the data warehouse
C) Develop service level agreements with suppliers and consumers of data for the data warehouse
D) All of the above

8) An organization should have one data warehouse administrator for every:
A) 100 users.
B) 100 gigabytes of data in the enterprise data warehouse.
C) department.
D) processor.

9) An open-source DBMS is:
A) a free source-code RBMS that provides the functionality of an SQL-compliant DBMS.
B) a beta release of a commercial RDBMS.
C) an object-oriented database management system.
D) source code for a commercial RDBMS.

10) Which of the following threats involves outside parties using information to embarrass a company?
A) Accidental loss
B) Theft and fraud
C) Loss of confidentiality
D) Loss of data integrity

11) Guidelines for server security should include all of the following EXCEPT:
A) securing the network between client and server.
B) guidelines for password lengths.
C) password naming conventions.
D) guidelines for frequency of password changes.

12) Which of the following is a type of network security?
A) Authentication of the client workstation
B) Password naming conventions
C) Guidelines for frequency of password changes
D) All of the above

13) Security measures for dynamic Web pages are different from static HTML pages because:
A) dynamic Web pages are built "on the fly."
B) static Web pages contain more sensitive data.
C) the connection requires full access to the database for dynamic pages.
D) all of the above.

14) The W3C standard for Web privacy is called:
A) the Web privacy act.
B) Platform for Privacy Preferences.
C) Freedom of Web Information Act.
D) the Magna Carta.

15) While views promote security by restricting user access to data, they are not adequate security measures because:
A) an unauthorized person may gain access to a view through experimentation.
B) all users can read any view.
C) a view's data does not change.
D) none of the above.

16) A trigger can be used as a security measure in which of the following ways?
A) To prohibit inappropriate actions
B) To cause special handling procedures to be executed
C) To cause a row to be written to a log file
D) All of the above

17) Controls designed to restrict access and activities are called:
A) schemas.
B) business rules.
C) encryption controls.
D) authorization rules.

18) Which of the following is a principal type of authorization table?
A) Subject
B) Transaction
C) View
D) Index

19) A device to measure or detect fingerprints or signatures is called a(n) _____ device.
A) biometric
B) view

C) ink
D) laser

20) A credit-card sized plastic card with an embedded microprocessor chip with the ability to store, process and output electronic data in a secure manner is called a(n):
A) smart chip.
B) smart card.
C) e-credit card.
D) secure card.

21) An audit trail of database changes is kept by a:
A) change control device.
B) subschema.
C) before image.
D) journalizing facility.

22) A DBMS periodically suspends all processing and synchronizes its files and journals through the use of a:
A) checkpoint facility.
B) backup facility.
C) recovery manager.
D) database change log.

23) A discrete unit of work that must be processed completely or not at all within a computer system is called a:
A) change control device.
B) transaction.
C) before image.
D) journalizing facility.

24) A _____ is a DBMS module that restores the database to a correct condition when a failure occurs.
A) backup facility
B) recovery manager
C) restart facility
D) transaction logger

25) _____ is used to undo unwanted database changes.
A) Rollback
B) Rollforward
C) Restart
D) Encryption

26) Forward recovery is faster than restore/rerun because:

A) transactions do not have to be repeated.
B) security can be avoided.
C) images are mirrored.
D) systems are more complete.
27) The preferred method of fixing an aborted transaction is:
A) repairing the schema.
B) switching.
C) duplication of data.
D) backing out the transaction.

28) When incorrect data have been introduced, the database is best recovered by:
A) backward recovery, if the error has been discovered soon enough.
B) human intervention for only a few errors.
C) restarting from the most recent checkpoint and processing subsequent transactions.
D) all of the above.

29) The actions that must be taken to ensure data integrity is maintained during multiple simultaneous transactions are called _____ actions.
A) logging
B) concurrency control
C) transaction authorization
D) multiple management

30) The extent of the database resource that is included with each lock is called the level of:
A) impact.
B) management.
C) lock granularity.
D) none of the above.

31) A(n) _____ prevents another transaction from reading and therefore updating a record until it is unlocked.
A) record controller
B) exclusive lock
C) authorization rule
D) shared lock

32) A(n) _____ is a procedure for acquiring the necessary locks for a transaction where all necessary locks are acquired before any are released.
A) record controller
B) exclusive lock
C) authorization rule
D) two-phase lock

33) An optimistic approach to concurrency control is called:
A) versioning.

B) denormalization.
C) deadlock resolution.
D) none of the above.
34) A repository of information about a database that documents data elements of a database is called a:
A) schema.
B) subschema.
C) view.
D) data dictionary.

35) A(n) _____ stores metadata about an organization's data and data processing resources.
A) DBA
B) information repository
C) organizational system catalog
D) data dictionary

36) Which of the following is not a component of a repository system architecture?
A) An informational model
B) The repository engine
C) A data transformation process
D) The repository database

37) Which of the following functions are supported by a repository engine?
A) Object management
B) Relationship management
C) Version management
D) All of the above
38) Which of the following is NOT an area of concern when trying to maintain a well-tuned database?
A) Memory usage
B) CPU usage
C) Input/output contention
D) User interface design

39) A(n) _____ is submitted by a DBA to test the current performance of a database or predict the response time for queries.
A) elusive query
B) performance test
C) heartbeat query
D) none of the above

40) The data administrator takes responsibility for the overall management of data resources.
Answer: TRUE
Diff: 1 Page Ref: 464
Topic: Introduction
AACSB: Use of Information Technology

41) Databases are generally the property of a single department within an organization. **F**
42) The role of database administration is typically a more hands-on, physical involvement with the management of databases. **T**
43) Specifications for transactions do not need to be reviewed quickly. **F**

44) The role of a data warehouse administrator emphasizes integration and coordination of metadata and data across many data sources. **T**

45) Open-source software always comes with complete documentation. **F**
46) The goal of database security is the protection of data from accidental or intentional threats to its integrity and access. **T**

47) Loss of data integrity does not impact the quality of data in a database. **F**

48) A view can be built to present only the data to which a user requires access. **T**

49) A domain is a way to create a structure that acts like a table. **F**

50) A trigger can be used for security purposes to prohibit inappropriate actions, such as changing a salary value outside of a business day. **T**

51) Authorization rules are controls incorporated in the data management system that restrict access to data and also restrict the actions that people may take when they access the data. **T**

Part2:
1) What new trends and technologies have driven the need for changes in the data

Trends and Technologies that are changing the role of data administration and database administration roles
Cloud Demand: There are demand of the database for application that could be hosted on private or public cloud. the requirement of the cloud is collaboration with all team supporting the application stack.
Database as a Service: Now a days administrator needs to provide database as a service as part of development life-cycle.

2) What are some of the core roles of the traditional data administrator?
- **Software installation and Maintenance.**
- **Data Extraction, Transformation, and Loading.**

3) What factors should one consider when choosing an open source DBMS?
- **Usability, Visualisation & Reporting, Security, Functionality, Support, Development, Integration, Scalability and Cost and Suitability.**

4) What threats to data security must be addressed in a comprehensive data security plan?

Answer:

5) What are integrity controls? Describe the different types.

To manage the integrity of data, which is a fundamental component of information security

---

TRUE/FALSE

1. Autoglobal array elements are referred to with an index number. **F**

2. You must use the `global` keyword to reference a global variable within the scope of a function. **T**

3. When using the variable name as an index for the `$GLOBALS` array, you omit the leading dollar sign from the name. **T**

4. The method attribute defines how the form data is submitted to the server. **T**

5. The value of the method attribute will be either post or get. **T**

6. When the form data is submitted using the get method, the form data is appended to the URL specified in the form's method attribute. **F**

7. When a form's submit button is pressed, each field on the form is sent to the Web server as a name/value pair. **T**

8. When the post method is used to submit a form, the name portion of the name/value pair becomes the value assigned to the array element. **F**

9. When using the get method to submit form data, the form data is separated from the URL by a colon. (:). **F**

10. The get method restricts the number of characters that can be appended to a URL to 255. **T**

11. PHP includes a feature called magic quotes, which automatically adds a backslash (\) to any single quote, double quote, or NULL character in submitted form data. **T**

12. By default, `magic_quotes_gpc` is the only magic quote directive enabled in the php.ini configuration file. **T**

13. The alternate method to escape strings is the `stripslashes()` function. **T**

14. The `blank()` function can be used to determine if a value was entered in a form. **F**

15. The best way to ensure valid form data is to build the Web form with controls that only allow the user to enter acceptable responses **T**

16. The `round()` function can be used to ensure that numbers have the appropriate number of digits after the decimal point, if any. **T**

17. Regular expression functions are some of the best tools for verifying string data meets the strict formatting required for email addresses. **T**

18. A good programming practice is to stop processing a form when an error is found and display the error to the user. **F**

19. A sticky form is redisplayed with the values that the user entered the last time the form was submitted. **T**

20. With complex escaping from XHTML, you close one PHP block, insert some XHTML elements, and then open another PHP block to continue the script. **F**

21. In PHP, an e-mail message is sent using the `email()` function. **F**

22. The `subject` argument of the mail function must be plain text without XHTML tags or character entities. **T**

23. `From`, `Cc`, `Bcc` and `Date` headers are examples of additional headers. **T**

24. The `mail()` function returns a value of `TRUE` if the message was sent successfully and FALSE if it was not. **T**

25. In an all-in-one form, when the user clicks the submit button, the script submits the form to a separate processing script. **F**

26. The `issent()` function can be used to determine if the `$Submit` variable has been set (if the Submit button has been pressed). **F**

27. The `unset()` function is used to uninitialize a variable after it has been set. **T**

28. The post method is ideal for embedding options in a hyperlink. **F**

29. When using text hyperlinks to navigate within a Web page template, the values that specify which dynamic content to show must be appended to the filename in the "href" attribute of the anchor tag. **T**

30. You cannot use buttons to target content to a dynamic section of a Web page template. **F**

MULTIPLE CHOICE

1. Autoglobals or superglobals are predefined ____.

a.	variables	c.	**arrays**
b.	constants	d.	functions

2.Autoglobals are associative arrays whose elements are referred to with a(n) ____ key instead of an index number.

a.	numeric	c.	symbolic
b.	**alphanumeric**	d.	alphabetic

3. The `$_ENV` autoglobal contains ____ variables that programs use to interact with the system.

a.	system	c.	**program**
b.	environmental	d.	script

4.The ____ function displays the elements of the `$_ENV` array and their values.

a.	`env_info()`	c.	**`phpinfo()`**
b.	`systeminfo()`	d.	`php_info()`

5. You can use ____ autoglobal array to refer to the global version of a variable from inside a function.

a.	`$GLOBALS`	c.	`$_GLOBALS`
b.	`$_GLOBAL`	d.	**`$GLOBAL`**

6.The value of the ____ attribute identifies the program on the Web server that will process the form data when the form is submitted.

a.	**action**	c.	name
b.	method	d.	id

7. When you click a form's submit button, each field on the form is sent to the Web server as a ____ pair.

a.	true/false	c.	variable/value
b.	**name/value**	d.	yes/no

8. When you use the ____ method to submit form data to the processing script, the form data is appended to the URL specified by the action attribute.

a.	**get**	c.	request
b.	post	d.	submit

9.The PHP feature that automatically adds a backslash to any single quote, double quote, or `NULL` character contained in form data that a user submits to a PHP script is ____.

a.	base quotes	c.	**magic quotes**
b.	special quotes	d.	special quotes

10. The ____ function is used to reverse the changes made by magic quotes or the addslashes() function.

a.	`removeslashes()`	c.	`deleteslashes()`
b.	**`stripslashes()`**	d.	`addslashes()`

Chapter 9 Data Warehousing

1) The analysis of summarized data to support decision making is called:
A) operational processing.
B) informational processing.
C) artificial intelligence.
D) data scrubbing.

2) The characteristic that indicates that a data warehouse is organized around key high-level entities of the enterprise is:

A) subject-oriented.
B) integrated.
C) time-variant.
D) nonvolatile.

3) When we consider data in the data warehouse to be time-variant, we mean:
A) that the time of storage varies.
B) data in the warehouse contain a time dimension so that they may be used to study trends and changes.
C) that there is a time delay between when data are posted and when we report on the data.
D) none of the above.

4) Which of the following advances in information systems contributed to the emergence of data warehousing?
A) Improvements in database technology, particularly the relational data model
B) Advances in computer hardware, especially affordable mass storage and parallel computer architectures
C) Advances in middleware products that enabled enterprise database connectivity across heterogeneous platforms
D) All of the above

5) Which of the following factors drive the need for data warehousing?
A) Businesses need an integrated view of company information.
B) Informational data must be kept together with operational data.
C) Data warehouses generally have better security.
D) None of the above.

6) Which of the following organizational trends does not encourage the need for data warehousing?
A) Multiple, nonsynchronized systems
B) Focus on customer relationship management
C) Downsizing
D) Focus on supplier relationship management

7) Informational systems are designed for all of the following EXCEPT:
A) running a business in real time.
B) supporting decision making.
C) complex queries.
D) data mining.
8) Operational and informational systems are generally separated because of which of the following factors?
A) A data warehouse centralizes data that are scattered throughout disparate operational systems and makes them readily available for decision support applications.
B) A properly designed data warehouse adds value to data by improving their quality and consistency.
C) A separate data warehouse eliminates contention for resources that results when informational

applications are confounded with operational processing.
D) All of the above.
9) A data mart is a(n):
A) enterprisewide data warehouse.
B) smaller system built upon file processing technology.
C) data warehouse that is limited in scope.
D) generic on-line shopping site.

10) One characteristic of independent data marts is complexity for end users when they need to access data in separate data marts. This complexity is caused by not only having to access data from separate databases, but also from:
A) the possibility of a new generation of inconsistent data systems, the data marts themselves.
B) lack of user training.
C) denormalized data.
D) incongruent data formats.

11) All of the following are limitations of the independent data mart EXCEPT:
A) separate extraction, transformation, and loading processes are developed for each data mart.
B) data marts may not be consistent with one another.
C) there is no capability to drill down into greater detail in other data marts.
D) it is often more expedient to build a data mart than a data warehouse.

12) A dependent data mart:
A) is filled with data extracted directly from the operational system.
B) is filled exclusively from the enterprise data warehouse with reconciled data.
C) is dependent upon an operational system.
D) participates in a relationship with an entity.
13) An operational data store (ODS) is a(n):
A) place to store all unreconciled data.
B) representation of the operational data.
C) integrated, subject-oriented, updateable, current-valued, detailed database designed to serve the decision support needs of operational users.
D) small-scale data mart.

14) A logical data mart is a(n):
A) data mart consisting of only logical data.
B) data mart created by a relational view of a slightly denormalized data warehouse.
C) integrated, subject-oriented, detailed database designed to serve operational users.
D) centralized, integrated data warehouse.

15) All of the following are unique characteristics of a logical data mart EXCEPT:
A) logical data marts are not physically separate databases, but rather a relational view of a data warehouse.
B) the data mart is always up-to-date since data in a view is created when the view is referenced.
C) the process of creating a logical data mart is lengthy.
D) data are moved into the data warehouse rather than a separate staging area.

16) The real-time data warehouse is characterized by which of the following?
A) It accepts near-real time feeds of transaction data.
B) Data are immediately transformed and loaded into the warehouse.
C) It provides near-real-time access for the transaction processing systems to an enterprise data warehouse.
D) All of the above.
17) _____ technologies are allowing more opportunities for real-time data warehouses.
A) Web
B) MOLAP
C) RFID
D) GPS

18) All of the following are some beneficial applications for real-time data warehousing EXCEPT:
A) just-in-time transportation.
B) e-commerce. For example, an abandoned shopping cart can trigger an email promotional message.
C) fraud detection in credit card transactions.
D) data entry.

19) Data that are detailed, current, and intended to be the single, authoritative source of all decision support applications are called _____ data.
A) reconciled
B) subject
C) derived
D) detailed

20) A database action that results from a transaction is called a(n):
A) transition.
B) event.
C) log entry.
D) journal happening.
21) Data that are never physically altered once they are added to the store are called _____ data.
A) transient
B) override
C) periodic
D) complete

22) Which of the following is an objective of derived data?
A) ease of use for decision support systems
B) faster response time for user queries
C) support data mining applications
D) All of the above

23) A star schema contains both fact and _____ tables.
A) narrative
B) cross functional
C) dimension
D) starter
24) Every key used to join the fact table with a dimension table should be a _____ key.
A) primary
B) surrogate
C) foreign
D) secondary
25) The level of detail in a fact table determined by the intersection of all the components of the primary key, including all foreign keys and any other primary key elements, is called the:
A) span.
B) grain.
C) selection.
D) aggregation.

26) Conformed dimensions allow users to do the following:
A) share nonkey dimension data.
B) query across fact tables with consistency.
C) work on facts and business subjects for which all users have the same meaning.
D) all of the above.

27) Factless fact tables may apply when:
A) we are tracking events.
B) we are tracking sales.
C) we are taking inventory of the set of possible occurrences.
D) both A and C.

28) An expanded version of a star schema in which all of the tables are fully normalized is called a(n):
A) snowflake schema.
B) operational schema.
C) DSS schema.
D) complete schema.
29) All of the following are ways to handle changing dimensions EXCEPT:
A) overwrite the current value with the new value.
B) for each dimension attribute that changes, create a current value field and as many old value fields as we wish.
C) create a new dimension table row each time the dimension object changes.

D) create a snowflake schema.

30) _____ is an ill-defined term applied to databases where size strains the ability of commonly used relational DBMSs to manage the data.
A) Mean data
B) Small data
C) Star data
D) Big data

31) _____ is/are a new technology which trade(s) off storage space savings for computing time.
A) Dimensional modeling
B) Column databases
C) Fact tables
D) Snowflake schemas

32) A class of database technology used to store textual and other unstructured data is called:
A) mySQL.
B) NoSQL.
C) KnowSQL.
D) PHP.
33) The use of a set of graphical tools that provides users with multidimensional views of their data is called:
A) on-line geometrical processing (OGP).
B) drill-down analysis.
C) on-line analytical processing (OLAP).
D) on-line datacube processing (ODP).

34) OLAP tools that use the database as a traditional relational database are called:
A) ROLAP tools.
B) MOLAP tools.
C) slice and dice.
D) none of the above.

35) Rotating the view of a multidimensional database for a particular data point is called data:
A) cubing.
B) drill-down.
C) dicing.
D) pivoting.

36) Which of the following is true of data visualization?
A) It is easier to observe trends and patterns in data.
B) Correlations and clusters in data can be easily identified.
C) It is often used in conjunction with data mining.
D) All of the above.

37) Going from a summary view to progressively lower levels of detail is called data:
A) cubing.
B) drill-down.
C) dicing.
D) pivoting.

38) Which of the following data-mining techniques identifies clusters of observations with similar characteristics?
A) Case reasoning
B) Rule discovery
C) Clustering and signal processing
D) Neural nets

39) Which of the following data-mining techniques searches for patterns and correlations in large data sets?
A) Case reasoning
B) Rule discovery
C) Signal processing
D) Neural nets

40) Which of the following data-mining applications identifies customers for promotional activity?
A) Population profiling
B) Target marketing
C) Usage analysis
D) Product affinity
41) Advances in computer hardware, particularly the emergence of affordable mass storage and parallel computer architectures, was one of the key advances that led to the emergence of data warehousing. **T**

42) The development of the relational data model did not contribute to the emergence of data warehousing. **F**

43) The need for data warehousing in an organization is driven by its need for an integrated view of high-quality data. **T**

44) When multiple systems in an organization are synchronized, the need for data warehousing increases. **F**

45) Informational systems are designed to support decision making based on historical point-in-time and prediction data. **T**

46) A separate data warehouse causes more contention for resources in an organization. **F**

47) An independent data mart is filled with data extracted from the operational environment without the benefit of a data warehouse. **T**

48) A data mart is a data warehouse that contains data that can be used across the entire organization. **F**

49) Organizations adopt data mart architectures because it is easier to have separate, small data warehouses than to get all organizational parties to agree to one view of the organization in a central data warehouse. **T**

50) Independent data marts do not generally lead to redundant data and efforts. **F**

Short Answer
1) Discuss the history of data warehousing.
A huge improvement in database technology and advancments in computer hardware including mass storage. There also an emergence of end-user computing with powerful interfaces and tools.
There is also advances in middleware which enables database connectivity for everyone.

2) Explain the difference between operational and informational systems as well as the primary factors that contribute to the need for separation.

Operational system – a system that is used to run a business in real time, based on current data; also called a system of record
Informational system – a system designed to support decision making based on historical point-in-time and prediction data for complex queries or data-mining applications

The primary purpose of operational systems is to run a business on a current basis while informational systems support managerial decisions.

3) Explain the four basic steps to build an independent data mart.
Answer:

Subject-oriented: e.g. customers, patients, students, products
Integrated: consistent naming conventions, formats, encoding structures; from multiple data sources
Time-variant: can study trends and changes
Non-updatable: read-only, periodically refreshed

4) Discuss the limitations of the independent data mart architecture.
Answer:
Mini-warehouses, limited in scope
Separate ETL for each independent data mart
Data access complexity due to multiple data marts

Part1:
1) Data governance can be defined as:
A) a means to slow down the speed of data.
B) high-level organizational groups and processes that oversee data stewardship.

C) a government task force for defining data quality.
D) none of the above.

2) High-quality data are data that are:
A) accurate.
B) consistent.
C) available in a timely fashion.
D) all of the above.

3) Data quality ROI stands for:
A) return on investment.
B) risk of incarceration.
C) rough outline inclusion.
D) none of the above.

4) Data quality is important for all of the following reasons EXCEPT:_
A) it minimizes project delay.
B) it aids in making timely business decisions.
C) it provides a stream of profit.
D) it helps to expand the customer base.
5) Quality data can be defined as being:
A) unique.
B) inaccurate.
C) historical.
D) precise.

6) Conformance means that:
A) data have been transformed.
B) data are stored, exchanged or presented in a format that is specified by its metadata.
C) data are stored in a way to expedite retrieval.
D) none of the above.

7) One characteristic of quality data which pertains to the expectation for the time between when data are expected and when they are available for use is:
A) currency.
B) consistency.
C) referential integrity.
D) timeliness.

8) External data sources present problems for data quality because:
A) data are not always available.
B) there is a lack of control over data quality.
C) there are poor data capture controls.
D) data are unformatted.

9) Data quality problems can cascade when:
A) data are not deleted properly.
B) data are copied from legacy systems.
C) there is redundant data storage and inconsistent metadata.
D) there are data entry problems.

10) The best place to improve data entry across all applications is:
A) in the users.
B) in the level of organizational commitment.
C) in the database definitions.
D) in the data entry operators.

11) Which of the following are key steps in a data quality program?
A) Conduct a data quality audit.
B) Apply TQM principles and practices.
C) Estimate return on investment.
D) All of the above.

12) One simple task of a data quality audit is to:
A) interview all users.
B) statistically profile all files.
C) load all data into a data warehouse.
D) establish quality metrics.
13) One way to improve the data capture process is to:
A) allow all data to be entered manually.
B) provide little or no training to data entry operators.
C) check entered data immediately for quality against data in the database.
D) not use any automatic data entry routines.

14) TQM stands for:
A) Thomas Quinn Mann, a famous data quality innovator.
B) Total Quality Manipulation.
C) Transforming Quality Management.
D) Total Quality Management.

15) The methods to ensure the quality of data across various subject areas are called:
A) Variable Data Management.
B) Master Data Management.
C) Joint Data Management.
D) Managed Data Management.

16) All of the following are popular architectures for Master Data Management EXCEPT:
A) Identity Registry.
B) Integration Hub.

C) Persistent Object.
D) Normalization.

17) In the _____ approach, one consolidated record is maintained, and all applications draw on that one actual "golden" record.
A) persistent
B) identity registry
C) federated
D) integration hub

18) All of the following are ways to consolidate data EXCEPT:
A) application integration.
B) data rollup and integration.
C) business process integration.
D) user interaction integration.

19) Data federation is a technique which:
A) creates an integrated database from several separate databases.
B) creates a distributed database.
C) provides a virtual view of integrated data without actually creating one centralized database.
D) provides a real-time update of shared data.

20) _____ duplicates data across databases.
A) Data propagation
B) Data duplication
C) Redundant replication
D) A replication server

21) Event-drive propagation:
A) provides a means to duplicate data for events.
B) pushes data to duplicate sites as an event occurs.
C) pulls duplicate data from redundant sites.
D) none of the above.
22) The major advantage of data propagation is:
A) real-time cascading of data changes throughout the organization.
B) duplication of non-redundant data.
C) the ability to have trickle-feeds.
D) none of the above.

23) A characteristic of reconciled data that means the data reflect an enterprise-wide view is:
A) detailed.
B) historical.
C) normalized.

D) comprehensive.

24) Informational and operational data differ in all of the following ways EXCEPT:
A) level of detail.
B) normalization level.
C) scope of data.
D) data quality.

25) A method of capturing only the changes that have occurred in the source data since the last capture is called _____ extract.
A) static
B) incremental
C) partial
D) update-driven
26) A technique using artificial intelligence to upgrade the quality of raw data is called:
A) dumping.
B) data reconciliation.
C) completion backwards updates.
D) data scrubbing.

27) All of the following are tasks of data cleansing EXCEPT:
A) decoding data to make them understandable for data warehousing applications.
B) adding time stamps to distinguish values for the same attribute over time.
C) generating primary keys for each row of a table.
D) creating foreign keys.
28) An approach to filling a data warehouse that employs bulk rewriting of the target data periodically is called:
A) dump mode.
B) overwrite mode.
C) refresh mode.
D) update mode.

29) Which type of index is commonly used in data warehousing environments?
A) Join index
B) Bit-mapped index
C) Secondary index
D) Both A and B
30) Loading data into a data warehouse involves:
A) appending new rows to the tables in the warehouse.
B) updating existing rows with new data.
C) purging data that have become obsolete or were incorrectly loaded.
D) all of the above.

31) Data may be loaded from the staging area into the warehouse by following:
A) SQL Commands (Insert/Update).
B) special load utilities.

C) custom-written routines.
D) all of the above.

32) The process of combining data from various sources into a single table or view is called:
A) extracting.
B) updating.
C) selecting.
D) joining.

33) The process of transforming data from a detailed to a summary level is called:
A) extracting.
B) updating.
C) joining.
D) aggregating.

34) Which of the following is a basic method for single field transformation?
A) Table lookup
B) Cross-linking entities
C) Cross-linking attributes
D) Field-to-field communication

State True/False

35) Quality data are not essential for well-run organizations. **T**

36) ETL is short for Extract, Transform, Load. **T**

37) There are six major steps to ETL. **F**

38) A data steward is a person assigned the responsibility of ensuring the organizational applications properly support the organization's enterprise goals for data quality. **F**

39) A data governance committee is always made up of high-ranking government officials. **F**

40) Dirty data saves work for information systems projects. **F**

41) Data quality is essential for SOX and Basel II compliance. **T**

42) Quality data does not have to be unique. **F**

43) Completeness means that all data that are needed are present. **T**

44) Retention refers to the amount of data that is not purged periodically from tables. **F**

45) Generally, records in a customer file never become obsolete. **F**

46) Data which arrive via XML and B2B channels is always guaranteed to be accurate. **F**

47) The uncontrolled proliferation of spreadsheets, databases and repositories leads to data quality problems. **T**

48) A data quality audit helps an organization understand the extent and nature of data quality problems. **T**

49) A data stewardship program does not help to involve the organization in data quality. **F**

50) Total quality management (TQM) focuses on defect correction rather than defect prevention. **F**

Part2

1) Discuss data governance and what needs to be included in a data governance program.
   Since in today's world, the internet is a huge thing, everything is online. Due to that, organization of data is becoming more vulnerable. In order to protect, regulate and manage organization's data, it needs data governance. It minimizes risks by providing proper security. And establish internal rules for data use. Not only that, but it also improves internal and external communication, making it easier to multitask.

2) Why is data quality important?
   in today's world, data is important, and a requirement of quality is increasing. One wants to be aware and updated every second of the day. Hence, our data is the most valuable thing to be studied by analysts and other agencies. Is it important because if the quality of the data is good, then analysts can get more information from it. also, good quality data provides better information that is not going to be false and spread rumors. In order for data to have a higher quality, it has to be more up to date, relevant and as accurate as possible to be considered.

3) What are some of the characteristics of quality data?
   Accuracy: For quality data, your data must contain accurate information and it should be precise.

   Completeness: Quality data must have complete information of signal, if data will be distorted then it will be of no use .

   Reliability: As we all know that "little knowledge is harmful", it is same with our data, it should be complete as well as reliable.

   Relevance: Relevant data or data on which we can trust with our closed eye is must in today's time. Wrong kind of information or data spreading can lead to serious accidents.

Timeliness : It means our data must be up to date. Out dated data can mislead someone. For ex- There are application which tell us latest news. Suppose it is showing same news everyday, then you will not get updated with daily news. This makes you unaware of what 's happening around. So, to become aware, your app data must be updated every day, so data must be updated .

4) What are some of the reasons for deteriorated data quality?
it can be either a human mistake (spreadind data without validation) , machine mistake(unknown bugs, data corruption, viruses), or not updating data.

5) What are the key steps in a data quality program?
Get business buy-in
Perform data quality audit
Establish data stewardship program
Improve data capture processes
Apply modern data management principles and technology
Apply total quality management (TQM) practices

6) Discuss Master Data Management.
Master Data Management disciplines, technologies, and methods to ensure the currency, meaning, and quality of reference data within and across various subject areas

Three main architectures:

Identity registry: master data remains in source systems; registry provides applications with location
Integration hub: data changes broadcast through central service to subscribing databases
Persistent: central “golden record” maintained; all applications have access. Requires applications to push data. Prone to data duplication.

7) What are some of the approaches to data integration?
8) Discuss the ETL process.

9) What are the characteristics of data after ETL?