

Question No: 1 (Marks: 1) – Please choose one

A data warehouse may include

- ▶ **Legacy systems** (Page 135)
- ▶ Only internal data sources
- ▶ Privacy restrictions
- ▶ Small data mart

Question No: 2 (Marks: 1) – Please choose one

De-Normalization normally speeds up

- ▶ **Data Retrieval** (Page 51)
- ▶ Data Modification
- ▶ Development Cycle
- ▶ Data Replication

Question No: 3 (Marks: 1) – Please choose one

In horizontal splitting, we split a relation into multiple tables on the basis of

- ▶ **Common Column Values** (Page 54)
- ▶ Common Row Values
- ▶ Different Index Values
- ▶ Value resulted by ad-hoc query

Question No: 4 (Marks: 1) – Please choose one

Multidimensional databases typically use proprietary _____ format to store pre-summarized cube structures.

- ▶ **File** (Page 79)
- ▶ Application
- ▶ Aggregate
- ▶ Database

Question No: 5 (Marks: 1) – Please choose one

A dense index, if fits into memory, costs only _____ disk I/O access to locate a record by given key.

- ▶ **One** (Page 223)
- ▶ Two
- ▶ $\lg(n)$
- ▶ n

ایماندار کو غصہ دیر سے آتا ہے اور جلدی دور ہو جاتا ہے

Question No: 6 (Marks: 1) – Please choose one

All data is _____ of something real.

I An Abstraction

II A Representation

Which of the following option is true?

▶ **I Only** (Page 180)

▶ II Only

▶ Both I & II

▶ None of I & II

Question No: 7 (Marks: 1) – Please choose one

The key idea behind _____ is to take a big task and break it into subtasks that can be processed concurrently on a stream of data inputs in multiple, overlapping stages of execution.

▶ **Pipeline Parallelism** (Page 214)

▶ Overlapped Parallelism

▶ Massive Parallelism

▶ Distributed Parallelism

Question No: 8 (Marks: 1) – Please choose one

Non uniform distribution, when the data is distributed across the processors, is called _____.

▶ **Skew in Partition** (Page 218)

▶ Pipeline Distribution

▶ Distributed Distribution

▶ Uncontrolled Distribution

Question No: 9 (Marks: 1) – Please choose one

The goal of ideal parallel execution is to completely parallelize those parts of a computation that are not constrained by data dependencies. The smaller the portion of the program that must be executed _____, the greater the scalability of the computation.

▶ None of these

▶ **Sequentially** (Page 204)

▶ In Parallel

▶ Distributed

Question No: 10 (Marks: 1) – Please choose one

If 'M' rows from table-A match the conditions in the query then table-B is accessed 'M' times. Suppose table-B has an index on the join column. If 'a' I/Os are required to read the data block for each scan and 'b' I/Os for each data block then the total cost of accessing table-B is _____ logical I/Os approximately.

▶ $(a + b)M$

▶ $(a - b)M$

▶ $(a + b + M)$

▶ $(a * b * M)$

Question No: 11 (Marks: 1) – Please choose one

Data mining is a/an _____ approach, where browsing through data using data mining techniques may reveal something that might be of interest to the user as information that was unknown previously.

- ▶ **Exploratory** (Page 249)
- ▶ Non-Exploratory
- ▶ Computer Science

Question No: 12 (Marks: 1) – Please choose one

Data mining evolved as a mechanism to cater the limitations of _____ systems to deal massive data sets with high dimensionality, new data types, multiple heterogeneous data resources etc.

- ▶ **OLTP** (Page 254)
- ▶ OLAP
- ▶ DSS
- ▶ DWH

Question No: 13 (Marks: 1) – Please choose one

_____ is the technique in which existing heterogeneous segments are reshuffled, relocated into homogeneous segments.

- ▶ **Clustering** (Page 264)
- ▶ Aggregation
- ▶ Segmentation
- ▶ Partitioning

Question No: 14 (Marks: 1) – Please choose one

To measure or quantify the similarity or dissimilarity, different techniques are available. Which of the following option represent the name of available techniques?

- ▶ Pearson correlation is the only technique
- ▶ Euclidean distance is the only technique
- ▶ **Both Pearson correlation and Euclidean distance** (Page 270)
- ▶ None of these

Question No: 15 (Marks: 1) – Please choose one

For a given data set, to get a global view in un-supervised learning we use

- ▶ **One-way Clustering** (Page 271)
- ▶ Bi-clustering
- ▶ Pearson correlation
- ▶ Euclidean distance

Question No: 16 (Marks: 1) – Please choose one

In DWH project, it is assured that _____ environment is similar to the production environment

- ▶ Designing
- ▶ **Development** (Page 314)
- ▶ Analysis
- ▶ Implementation

Question No: 17 (Marks: 1) – Please choose one

For a DWH project, the key requirement are _____ and product experience.

- ▶ Tools
- ▶ **Industry** (Page 320)
- ▶ Software
- ▶ None of these

Question No: 18 (Marks: 1) – Please choose one

Pipeline parallelism focuses on increasing throughput of task execution, NOT on _____ sub-task execution time.

- ▶ Increasing
- ▶ **Decreasing** (Page 215)
- ▶ Maintaining
- ▶ None of these

Question No: 19 (Marks: 1) – Please choose one

Many data warehouse project teams waste enormous amounts of time searching in vain for a _____.

- ▶ **Silver Bullet** (Page 315)
- ▶ Golden Bullet
- ▶ Suitable Hardware
- ▶ Compatible Product

Question No: 20 (Marks: 1) – Please choose one

Focusing on data warehouse delivery only often end up _____.

- ▶ **Rebuilding** (Page 315)
- ▶ Success
- ▶ Good Stable Product
- ▶ None of these

Question No: 21 (Marks: 1) – Please choose one

Pakistan is one of the five major _____ countries in the world.

- ▶ **Cotton-growing** (Page 330)
- ▶ Rice-growing
- ▶ Weapon Producing

Question No: 22 (Marks: 1) – Please choose one

_____ is a process which involves gathering of information about column through execution of certain queries with intention to identify erroneous records.

- ▶ **Data profiling** (Page 439)
- ▶ Data Anomaly Detection
- ▶ Record Duplicate Detection
- ▶ None of these

Question No: 23 (Marks: 1) – Please choose one

Relational databases allow you to navigate the data in _____ that is appropriate using the primary, foreign key structure within the data model.

- ▶ Only One Direction
- ▶ **Any Direction (Page 19)**
- ▶ Two Direction
- ▶ None of these

Question No: 24 (Marks: 1) – Please choose one

DSS queries do not involve a primary key

- ▶ **True (Page 21)**
- ▶ False

Question No: 25 (Marks: 1) – Please choose one

_____ contributes to an under-utilization of valuable and expensive historical data, and inevitably results in a limited capability to provide decision support and analysis.

- ▶ **The lack of data integration and standardization (Page 330)**
- ▶ Missing Data
- ▶ Data Stored in Heterogeneous Sources

Question No: 26 (Marks: 1) – Please choose one

DTS allows us to connect through any data source or destination that is supported by _____

- ▶ **OLE DB (Page 373)**
- ▶ OLAP
- ▶ OLTP
- ▶ Data Warehouse

Question No: 27 (Marks: 1) – Please choose one

Data Transformation Services (DTS) provide a set of _____ that lets you extract, transform, and consolidate data from disparate sources into single or multiple destinations supported by DTS connectivity.

- ▶ **Tools (Page 373) rep**
- ▶ Documentations
- ▶ Guidelines

Question No: 28 (Marks: 1) – Please choose one

Execution can be completed successfully or it may be stopped due to some error. In case of successful completion of execution all the transactions will be _____

- ▶ **Committed to the database (Page 419)**
- ▶ Rolled back

بری صحبت سے تنہائی بہتر ہے اور تنہائی سے نیک صحبت بہتر ہے

Question No: 29 (Marks: 1) – Please choose one

If some error occurs, execution will be terminated abnormally and all transactions will be rolled back. In this case when we will access the database we will find it in the state that was before the _____.

▶ **Execution of package** (Page 419)

- ▶ Creation of package
- ▶ Connection of package

Question No: 30 (Marks: 1) – Please choose one

To judge effectiveness we perform data profiling twice.

- ▶ One before Extraction and the other after Extraction
- ▶ **One before Transformation and the other after Transformation** (Page 441)
- ▶ One before Loading and the other after Loading