MCQZ

ATo identify the _____ required we need to perform data profiling

 ans⊘Degree of Transformation ---------------

Complexity

Cost

Time

Execution can be completed successfully or it may be stopped due to some error. If some error occurs, execution will be terminated abnormally and all transactions will be _____

 ans⊘Committed to the database---------------

Rolled back

Companies collect and record their own operational data, but at the same time they also use reference data obtained from _____ sources such as codes, prices etc.

Operational

None of these

Internal

 ans⊘External---------------

Ad-hoc access means to run such queries which are known already.

True

 ans⊘False---------------

_____ in agriculture extension is that pest population beyond which the benefit of spraying outweighs its cost.

Profit Threshold Level

 ans⊘Economic Threshold LevelA ---------------

Medicine Threshold Level

None of these

People that design and build the data warehouse must be capable of working across the organization at all levels

 ans⊘True---------------

False

The _____ is only a small part in realizing the true business value buried within the mountain of data collected and stored within organizations business systems and operational databases.

Independence on technology

 ans⊘Dependence on technology---------------

None of these

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

 ans⊘Silver Bullet---------------

Golden Bullet
Suitable Hardware
Compatible Product
Multidimensional databases typically use proprietary _____ format to store pre-summarized cube structures.
► File
► Application
► Aggregate
► Database
A dense index, if fits into memory, costs only _____ disk I/O access to locate a record by given key.
► One
► Two
► lg (n)
► n
All data is _____ of something real.
I An Abstraction
II A Representation
Which of the following option is true?
ans ◊► I Only---------------

► II Only
► Both I & II
► None of I & II
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.
ans ◊► Pipeline Parallelism---------------

► Overlapped Parallelism
► Massive Parallelism
► Distributed Parallelism
Non uniform distribution, when the data is distributed across the processors, is called _____.
ans ◊► Skew in Partition---------------

► Pipeline Distribution
► Distributed Distribution
► Uncontrolled Distribution
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
ans ◊► Sequentially---------------

► *In Parallel*
► *Distributed*
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.
 ans ◊► *Exploratory---------------*

► *Non-Exploratory*
► *Computer Science*
Data mining evolve as a mechanism to cater the limitations of _____ systems to dealmassive data sets with high dimensionality, new data types, multiple heterogeneous data resources etc.
► *OLTP*
► *OLAP*
► *DSS*
► *DWH*
_____ is the technique in which existing heterogeneous segments are reshuffled, relocated into homogeneous segments.
 ans ◊► *Clustering---------------*

► *Aggregation*
► *Segmentation*
► *Partitioning*
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*
 ans ◊► *Both Pearson correlation and Euclidean distance---------------*

► *None of these*
For a DWH project, the key requirement are _____ and product experience.
► *Tools*
 ans ◊► *Indust---------------*
ry
► *Software*
► *None of these*
Pipeline parallelism focuses on increasing throughput of task execution, NOT on _____ sub-task execution time.
► *Increasing*
 ans ◊► *Decreasing---------------*

► *Maintaining*
► *None of these*
Focusing on data warehouse delivery only often end up _____.
► *Rebuilding*
► *Success*

► *Good Stable Product*
► *None of these*
Pakistan is one of the five major _____ countries in the world.
 ans ◊► *Cotton-growing ---------------*

► *Rice-growing*
► *Weapon Producing*
_____ is a process which involves gathering of information about column through execution of certain queries with intention to identify erroneous records.
 ans ◊► *Data profiling---------------*

► *Data Anomaly Detection*
► *Record Duplicate Detection*
► *None of these*
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*
 ans ◊► *Any Direction---------------*

► *Two Direction*
► *None of these*
DSS queries do not involve a primary key
► *True*
► *False*
_____ contributes to an under-utilization of valuable and expensive historical data, and inevitably results in a limited capability to provide decision support and analysis.
 ans ◊► *The lack of data integration and standardization---------------*

► *Missing Data*
► *Data Stored in Heterogeneous Sources*
DTS allows us to connect through any data source or destination that is supported by _____
 ans ◊► *OLE DB---------------*

► *OLAP*
► *OLTP*
► *Data Warehouse*
Data Transformation Services (DTS) provide a set of _____ that lets you extract, transform, and consolidate data from disparate sources into single or multipledestinations supported by DTS connectivity.
 ans ◊► *Tools---------------*

► *Documentations*
► *Guidelines*
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 _____.
ans◊► Execution of package---------------

► Creation of package
► Connection of package
To judge effectiveness we perform data profiling twice.
► One before Extraction and the other after Extraction
ans◊► One before Transformation and the other after Transformation---------------

► One before Loading and the other after Loading
The need to synchronize data upon update is called
► Data Manipulation
► Data Replication
ans◊► Data Coherency---------------

► Data Imitation
Taken jointly, the extract programs or naturally evolving systems formed a spider web, also known as
► Distributed Systems Architecture
ans◊► Legacy Systems Architecture---------------
► Online Systems Architecture
► Intranet Systems Architecture
Node of a B-Tree is stored in memory block and traversing a B-Tree involves _____ page faults.
► O (n)
► O (n2)
► O (n lg n)
► O (lg n)
Which statement is true for De-Normalization?
► Redundant data is a performance liability at query time, but is a performance benefit at update time.
► Redundant data is a performance benefit at both query time and update time.
► Redundant data is a performance liability at both query time and update time.
ans◊► Redundant data is a performance benefit at query time, but is a performance liability at update time. ---------------

It is observed that every year the amount of data recorded in an organization is
Doubles
Triples
Quartiles
Remains same as previous year
Pre-computed _____ can solve performance problems
Aggregates
Facts
Dimensions

The degree of similarity between two records, often measured by a numerical value between _____, usually depends on application characteristics.
0 and 1
0 and 10
0 and 100
0 and 99
The purpose of the House of Quality technique is to reduce _____ types of risk.
 ans ✆Two---------------

Three
Four
All
NUMA stands for _____
 ans ✆Non-uniform Memory Access---------------

Non-updateable Memory Architecture
New Universal Memory Architecture
There are many variants of the traditional nested-loop join. If the index is built as part of the query plan and subsequently dropped, it is called
Naive nested-loop join
Index nested-loop join
Temporary index nested-loop join ans ✆---------------

None of these
The Kimball s iterative data warehouse development approach drew on decades of experience to develop the _____.
 ans ✆Business Dimensional Lifecycle ---------------

Data Warehouse Dimension
Business Definition Lifecycle
OLAP Dimension
For a smooth DWH implementation we must be a technologist.
True
False
During the application specification activity, we also must give consideration to the organization of the applications.
True ans ✆---------------

False
The most recent attack is the _____ attack on the cotton crop during 2003- 04, resulting in a loss of nearly 0.5 million bales.
Boll Worm ans ✆---------------

Purple Worm
Blue Worm
Cotton Worm

The users of data warehouse are knowledge workers in other words they are_____
in the organization.
ans ◊Decision make ---------------
r
Manager
Database Administrator
DWH Analyst
_____ breaks a table into multiple tables based upon common column values.
ans ◊Horizontal splitting ---------------

Vertical splitting
As apposed to the out come of classification , estimation deal with _____
valued outcome.
Discrete
Isolated
ans ◊Continuous ---------------

Distinct
The goal of _____is to look at as few block as possible to find the matching records.
Indexing
Partitioning
Joining
none of these
nested loop join ans ◊---------------

none of these
The technique that is used to perform these feats in data mining modeling, and this act
of model building is something that people have been doing for long time, certainly
before the _____ of computers or data mining technology.
Access Advent ans ◊---------------

Ascent Avowal
A data warehouse may include
ans ◊► Legacy systems ---------------

► Only internal data sources
► Privacy restrictions
► Small data mart
De-Normalization normally speeds up
ans ◊► Data Retrieval ---------------

► Data Modification
► Development Cycle
► Data Replication
In horizontal splitting, we split a relation into multiple tables on the basis of
ans ◊► Common Column Values ---------------

► *Common Row Values*
► *Different Index Values*
► *Value resulted by ad-hoc query*
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 C and 'b' I/Os for each data block then the total cost of accessing table-B is _____ logical I/Os approximately.
 ans ◊► (a + b)M ---------------

► *(a - b)M*
► *(a + b + M)*
► *(a * b * M)*
For a given data set, to get a global view in un-supervised learning we use
 ans ◊► One-way Clustering ---------------

► *Bi-clustering*
► *Pearson correlation*
► *Euclidean distance*
In DWH project, it is assured that _____ environment is similar to the production environment
► *Designing*
 ans ◊► Development ---------------

► *Analysis*
► *Implementation*
For good decision making, data should be integrated across the organization to cross the LoB (Line of Business). This is to give the total view of organization from:
► *Owner's Perspective*
 ans ◊► Customer's Perspective ---------------

► *Decision Maker's Perspective*
► *Employee's Perspective*
Which is the least appropriate join operation for Pipeline parallelism?
Hash Join
Inner Join
Outer Join
Sort-Merge Join
Data mining derives its name from the similarities between searching for valuable business information in a large database, for example, finding linked products in gigabytes of store scanner data, and mining a mountain for a _____ of valuable ore.
Furrow
Streak
Trough
 ans ◊Vein---------------

With data mining, the best way to accomplish this is by setting aside some of your data in a _____ to isolate it from the mining process; once the mining is complete, the results can be tested against the isolated data to confirm the model's validity.
Cell
Disk
Folder
Vault
We must try to find the one access tool that will handle all the needs of their users.
True
False
Investing years in architecture and forgetting the primary purpose of solving business problems, results in inefficient application. This is the example of _____ mistake.
Extreme Technology Design
 ans⧸Extreme Architecture Design---------------

The automated, prospective analyses offered by data mining move beyond the analysis of past events provided by respective tools typical of _____.
OLTP
OLAP
Decision Support systems
None of these
There are many variants of the traditional nested-loop join, if there is an index is exploited, then it is called……
Naïve nested loop join index
Nested loop join temporary index
 ans⧸Index nested-loop joins---------------

SUBJECTIVE

Snippet of my DWH paper today:
Some of the MCQs came from past papers. Attaching for you guys.
1. Horizontal splitting of tables
2. Nested loop join and its variants
3. Hash based join and its working
4. Define DWH, data mining & knowledge discovery in database
5. Kimball's life cycle model
6. Why should companies entertain students to visit their company's place?
7. Static & dynamic attributes and examples from agriDWH
8. Write SQL to fetch total number of female students registered in BS telcom (5 marks)
9. Is it possible to have erroneous data in DWH? How this will impact business processes? (5 marks)
10. How is hardware utilization different in DWH? (2 marks)

11. In case of non uniform distribution, what will be the impact on performance? (5 marks)
12. Why pilot project methodology is highly recommended in DWH?

Tufail Soft Technology
tufailsoft.weebly.com
tufailsoft@gmail.com
All project's Assignments and coding available on reasonable price