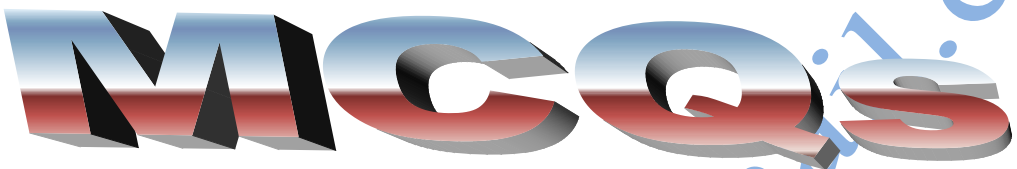
Prepared by: Dr. Tariq Hanif
Midterm-13
For more help @:
Email:qirathanif@gmail.com
Website:drqirathanif.jimdo.com
Cell# 03037300008



CS614 Data Warehousing

01. _____ is one class of decision support environment.

OLAP 30page

OLTP
Data Cleansing
ETL
2. The confusion created by data redundancy makes it difficult for companies to
Create online processing capabilities.
Work in batch processing load.
Use a distributed database.
Integrate data from different sources.
3. Effects of de-normalization on database performance are

Unpredictable

Predictable
Conventional
Unsurprising
5. DOLAP model facilitates _____ computing paradigm.

Mobile

Permanent
Rigid
Strict
7.Extract, Transform, Load (ETL) process consist of steps which are _____.

Independent and Interrelated

Independent or Interrelated
Dependent and Interrelated
Dependent or Interrelated
9. _____ is an application of intelligence and experience.
Skill
Power

Wisdom

Knowledge
11. Collapsing tables can be done on the _____ relationship(s)
Only One-to-One
Only Many-to-Many
Only One-to-Many

Both One-to-One and Many-to-Many

12. Transactional fact tables do not have records for events that do not occur. These are called

Not Recording Facts

Fact-less Facts
Null Facts
Empty Facts
13. Semantically "Dirty Data" class of anomalies includes which of the following:
I) Lexical Errors
II) Integrity Constraints Violation
III) Business Rule Contradiction
IV) Irregularities
V) Duplication
(I) and (II) only
(I), (II), and (III)

(II), (III), and (V) only

(I), (IV), and (V) only
14. 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

Two Direction
Partitions
15. One major goal of horizontal splitting is

Splitting rows for exploiting parallelism

Splitting columns for exploiting parallelism
Splitting schema for exploiting parallelism
Splitting relationships for exploiting parallelism
16. MOLAP usually builds "cubes" in proprietary file format of a multi-dimensional database (MDD) or a user defined data structure, therefore _____ is not supported.

ANSI

Microsoft
Oracle
SAP
17. A company has implemented data warehouse for analytical purpose. Quantity sold is stored as a fact. This quantity sold is

Additive Fact

Non-Additive Fact
Associative Fact
Non-Associative Fact
18. Typically a data mart is much smaller to data warehouse and it is pretty easy to take its _____ as compare to data warehouse.

Backup

Cube
Load
Schema
19. "Change Data Capture" is one of the challenging technical issues in _____

Data Extraction

Data Loading
Data Transformation
Data Cleansing
20. Within the data warehousing domain, data _____ is applied especially when several databases are merged.
Extraction
Loading

Cleansing

Join

1. Taken jointly, the extract programs or naturally evolving systems formed a spider web, also known as
Distributed Systems Architecture

Legacy Systems Architecture

Online Systems Architecture
Intranet Systems Architecture
2. Suppose the amount of data recorded in an organization is doubled every year. This increase is
Linear
Quadratic
Logarithmic

Exponential

5. ER is a _____ design technique that seeks to remove the redundancy in data.

Logical

Physical
Data Dependent
Transaction Dependent
6. _____ is the lowest level of detail or the atomic level of data stored in the warehouse.
Cube

Grain

Virtual Cube
Aggregate
7. It is called a _____ violation, if we have null values for attributes where NOT NULL constraint exists.
Load
Transform

Constraint

Extraction
9. In _____ system, the contents change with time.

OLTP

DSS
ATM
OLAP
10. It is observed that every year the amount of data recorded in an organization

Doubles

Triples
Quartiles
Remains same as previous year
11. Normalization is the process of efficiently organizing data in a database by _____ a relational table into smaller tables by projection.

Composing
Joining / Merging
Combining

Decomposing

12. 3NF removes even more data redundancy than 2NF but it is at the cost of

Simplicity and Performance

Complexity
Number of tables
Relations

16. When tables are populated for the first time, it is a full data refresh. This may be called as:

1. Block Insert
2. Block Slamming
3. Bulk Insert
4. Bulk Slamming

Which of the following option is true?
Option 1 & 3

Option 1 & 2

Option 1 & 4
Option 1, 2 & 3
17. The TQM philosophy of management is _____. All members of a total quality management organization strive to systematically manage the improvement of the organization through the ongoing participation of all employees in problem solving efforts across functional and hierarchical boundaries.

Customer-Oriented

Employee-Oriented
Employer-Oriented
Organization-Oriented
18. Identify the correct option. One Petabyte (PB) equals to ___
2^{52} or 10^{13} bytes

2^{50} or 10^{15} bytes

2^{50} or 10^{10} bytes
2^{48} or 10^{12} bytes
19. Pre-computed _____ can solve performance problems

Aggregates

Facts

Dimensions
Primary Keys
20. Single value attributes during recording of a transaction are _____

Dimensions

Facts
Aggregates
Constraints

CS614 Data Warehousing

1. Development of data warehouse is hard because data sources are

Unstructured & Heterogeneous

Structured & Heterogeneous
Unstructured & Homogeneous
Structured and Homogeneous
3. Select the statement which is true for Insurance Data Warehouse

It has Long Operational Business Cycle

It has Long Development & Implementation Cycle
It has Short Operational Business Cycle
It has Short Development & Implementation Cycle
4. Redundancy causes anomalies which are called
Selection Anomalies

Update Anomalies

SQL Anomalies
Data Warehouse Anomalies
6. 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.

Redundant data is a performance benefit at query time, but is a performance liability at update time.

7. Pre-join technique is used to avoid

Run time join

Compile time join
Load time join
8. OLAP is used for analytical process. For analytical processing we need

Multi-level aggregates

Record level access
Data level access
Row level access
9. The cube clause which is a part of SQL: 1999 is

GROUP BY CUBE (V_1, V_2 .... V_n)

SELECT BY CUBE (V_1, V_2 .... V_n)
JOIN BY CUBE (V_1, V_2 .... V_n)
None of these
10. ER is a logical design technique that seeks to remove the _____ in data.

Redundancy

Normalization
Anomalies
11. Non recording facts have a disadvantage that it has

Lack of Information

Redundant Information
Repeated Information
Normalized Information
12. Once the data has been transformed and ready to be loaded in to data warehouse, we adopt one of two prevalent _____ strategies.

Loading

Transformation
Quality
Indexing
13. Syntactically Dirty Data class of anomalies includes which of the following:

1. Lexical Errors
2. Integrity Constraints Violation
3. Business Rule Contradiction
4. Irregularities
5. Duplication

Option 1 and 4

Option 2 and 3
Option 2, 3, and 5
Option 1, 4, and 5
14. Records referring to the same entity are represented in different formats in the different data sets or are represented erroneously. Thus, duplicate records will appear in

the merged database. The issue is to identify and eliminate these duplicates. The problem is known as the _____ .

Merge/Purge Problem

Cleansing Problem
Transformation Problem
Data Quality Problem

15. since this form is useful for longitudinal comparisons illustrating trends of continuous improvement. Many traditional data quality metrics, such as free-of-error, completeness, and consistency take this form. This statement is about which of the following:

Simple Ratio

Min Operation
Max Operation
Weighted Average

16. To handle dimensions that require the aggregation of multiple data quality indicators, which of the following operation can be applied

Minimum or Maximum

Complex Ratio
Aggregate Average

17. 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.

None of these
Operational
Internal

External

19. _____ is about taking/collecting data from different heterogeneous sources.

Data Warehouse

Data Mart
Data Mining

20. In ROLAP access to information is provided via relational database using _____ standard SQL.

ANSI

Microsoft
Oracle
SAP

1.A typical example of the crisis in credibility in the naturally evolving architecture is the decision of CEO based on politics and personalities on receiving two different reports for the same query. We say CEO is

Very Subjective and Non-Scientific

Very Objective and Non-Scientific
Very Subjective and Scientific
3. Financial data warehouses have some severe drawbacks that are not found elsewhere. For example it is almost impossible to reconcile down to the rupee. This is because of many reasons. Select the statement which shows the possible reason(s).

The accounting periods may be different in different operational systems or the classifications of regions may change

The accounting periods may be different in Data Warehouse application
Data warehouse uses dynamic classifications of regions
During aggregation data warehouse neglect amount in rupees

6. One major goal of horizontal splitting is

Splitting rows for exploiting parallelism

Splitting columns for exploiting parallelism
Splitting schema for exploiting parallelism
8. ER Model can be simplified in -------- ways
One

Two

Three
Four
10. A company has implemented data warehouse for analytical purpose. Quantity sold is stored as a fact. This quantity sold is

Additive Fact

Non-Additive Fact
11. Fact-less fact table is a fact table without numeric fact columns. It is used to capture relationship between _____

Dimensions

Attributes
Tables
Facts
12. Full and Incremental extraction techniques are types of _____

Logical Extraction

Physical Extraction
Both Logical and Physical Extraction
None of these

13. Rearranging the grouping of source data, delivering it to the destination database, and ensuring the quality of data are crucial to the process of loading the data warehouse. Data _____ is vitally important to the overall health of a warehouse project.

1. Cleansing
2. Cleaning
3. Scrubbing

Which of the following options is true?

Option 1 only

Option 2 only
Option 1 & 2 only
Option 1, 2 & 3

16. As consumers, human beings judge the quality of things during their life-time.

I	Consciously
II	Subconsciously
III	Unconsciously

Which of the following statement is true?
I Only
II Only
III Only

I & II Only

17. All data is _____ of something real.

I	An Abstraction
II	A Representation

Which of the following option is true?

I Only

II Only
Both I & II
None of I & II

18. _____queries deal with number of variables spanning across number of tables (i.e. join operations) and looking at lots of historical data.

OLTP
DBMS

DSS

None of these

19.Collapsing tables can be done on the _____ relationships

Many-to-Many

Both One-to-One and Many-to-Many

None of these
One-to-One

20. In data warehouse, a query results in retrieval of hundreds of records from very large table. The ratio of number of records retrieved to total number of record present is high and selectivity is
Low

High

Average
Can not be calculated

4. OLAP is a (n) _____ of application.

Classification

Amalgamation
Unification
Blending

7. Extract, Transform, Load (ETL) process consist of steps which are _____.

Independent and Interrelated

Independent or Interrelated
Dependent and Interrelated
Dependent or Interrelated

9. _____ is an application of intelligence and experience.
Skill
Power

Wisdom

Knowledge

11. Collapsing tables can be done on the _____ relationship(s)
Only One-to-One
Only Many-to-Many
Only One-to-Many

Both One-to-One and Many-to-Many

12. Transactional fact tables do not have records for events that do not occur. These are called

Not Recording Facts

Fact-less Facts

Null Facts
Empty Facts
13. Semantically "Dirty Data" class of anomalies includes which of the following:

I) Lexical Errors
II) Integrity Constraints Violation
III) Business Rule Contradiction
IV) Irregularities
V) Duplication
(I) and (II) only
(I), (II), and (III)

(II), (III), and (V) only

(I), (IV), and (V) only
14. 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

Two Direction
Partitions
15. One major goal of horizontal splitting is

Splitting rows for exploiting parallelism

Splitting columns for exploiting parallelism
Splitting schema for exploiting parallelism
Splitting relationships for exploiting parallelism
16. MOLAP usually builds "cubes" in proprietary file format of a multi-dimensional database (MDD) or a user defined data structure, therefore _____ is not supported.

ANSI

Microsoft
Oracle
SAP

17. A company has implemented data warehouse for analytical purpose. Quantity sold is stored as a fact. This quantity sold is

Additive Fact

Non-Additive Fact
Associative Fact
Non-Associative Fact
18. Typically a data mart is much smaller to data warehouse and it is pretty easy to take its _____ as compare to data warehouse.

Backup

Cube
Load
Schema
19. "Change Data Capture" is one of the challenging technical issues in _____

Data Extraction

Data Loading
Data Transformation
Data Cleansing

1. Taken jointly, the extract programs or naturally evolving systems formed a spider web, also known as
Distributed Systems Architecture

Legacy Systems Architecture

Online Systems Architecture
Intranet Systems Architecture
3. The most common use of range partitioning in data warehouse is on

Date

Most redundant column
Fact
Dimensions
5. ER is a _____ design technique that seeks to remove the redundancy in data.

Logical

Physical
Data Dependent
Transaction Dependent
8. In the Information Age, the _____ learning organization is at a distinct disadvantage. This term means "impaired functioning."
Functional

Dysfunctional

Purposeful
Serviceable
OLAP
14.The goal of star schema design is to simplify _____
Logical data model

Physical data model

Conceptual data model
Semantic data model
16. When tables are populated for the first time, it is a full data refresh. This may be called as:

1. Block Insert
2. Block Slamming
3. Bulk Insert
4. Bulk Slamming

Which of the following option is true?
Option 1 & 3

Option 1 & 2

Option 1 & 4
Option 1, 2 & 3
CS614 Data Warehousing
3. De-Normalization normally speeds up

Data Retrieval

Data Modification
Development Cycle
Data Replication
4. In horizontal splitting, we split a relation into multiple tables on the basis of

Common Column Values

Common Row Values
Different Index Values
Value resulted by ad-hoc query
7. One of the OLAP characteristics is Multi-dimensional, which is _____ for OLAP.

Essential

Optional
Discretionary
Not Obligatory
8. Non recording facts have a disadvantage that it has

Lack of Information

Redundant Information
Repeated Information
Normalized Information
9. During ETL process of an organization, suppose you have data which can be transformed using any of the transformation method. Which of the following strategy will be your choice for least complexity?

One-to-One Scalar Transformation

One-to-Many Element Transformation
Many-to-Many Element Transformation
Many-to-One Element Transformation
11. _____ is an application of information and data.
Skill

Knowledge

Intelligence
Power
13. "The environment is smart enough to develop or compute higher level aggregates using lower level or more detailed aggregates". Which of the following approach is described by the above statement?

Aggregate awareness

Cube partitioning
Indexing
MOLAP cube aggregation
15. Syntactically "Dirty Data" class of anomalies includes _____
I) Lexical Errors
II) Integrity Constraints Violation
III) Business Rule Contradiction
IV) Irregularities
V) Duplication

(I) and (IV) only

(II) and (III) only
(II), (III), and (IV) only
(I), (IV), and (V) only
16. Experience showed that for a single pass of magnetic tape that scanned 100% of the records, only _____ of the records, sometimes were actually required.

5%

30%
50%
80%
19. In full extraction, data is extracted completely from the source system. Therefore there is no need to keep track of changes to the _____

Data Source

DWH
Data Mart
Data Destination

21. In MOLAP, there may be many reasons for the increase in cube size. List down any two. 02

Answer: There may be many reasons for the increase in cube size, such as increase in the number of dimensions, or increase in the cardinality of the dimensions, or increase in the amount of detail data or a combination of some or all these aspects. **(Pg.87)**

22. The problems associated with the extracted data can correspond to non-primary keys. List down any four problems associated with the non-primary key.02

Answer: 1. Different encoding in different sources. 2. Multiple ways to represent the same information.

3. Sources might contain invalid data. 4. Two fields with different data but same name. **(Pg. 163)**

23. In MOLAP, the number of possible aggregates is very large but some of the aggregates will have null values. Why? Justify with an example. 03

Answer: Although the number of possible aggregates is very large, but NOT all the aggregates may have values, there can be and will be quite a few aggregates which will have null values. For example, many of the items sold in winter are not sold in summer and not even kept in the store (and vice-a-versa). Consequently, there are no corresponding sales, and if the cube is generated that includes all the items, there will be many null aggregates, resulting in a very sparse cube. This will result in requirement of large amount of memory, most of which would be wasted.

24. What is "ranking" in data source selection? Explain with an example.03

Answer: Ranking is all about selecting the "right" source system. Rank establishment has to

be based on which source system is known to have the cleanest data for a particular attribute. Obviously you take the data element from the source system with the highest rank where the element exists**. For example**, consider the case of the gender data coming from two different source systems A and B. It may be the case that the highest quality data is from source system A, where the boxes for the gender were checked by the customers themselves. But what if someone did not check the gender box? Then you go on to the next cleanest source system i.e. B, where the gender was guessed based on the name.

25. Consider a fact table with name "Sales". The grain of table might be stated as "Sales volume by Day by Product by Store". Identify few facts (at least three) that can be used to populate such table. 05

26. Clustering is considered to be one of the most important automatic data cleansing techniques. Name any three automatic data cleansing techniques other

than clustering. Also mention if any drawback is associated with clustering technique in this automatic data cleansing process. 05

Answer: 1) Statistical 2) Pattern Based 3) Clustering

Some of the data cleansing techniques are listed. Let's discuss each of them in detail.

Statistical: Identifying outlier fields and records using the values of mean, standard deviation, range, etc., based on Chebyshev's theorem, considering the confidence intervals for each field. Outlier values for particular fields are identified based on automatically computed statistics. For each field the average and the standard deviation are utilized and based on Chebyshev's theorem those records that have values in a given field outside a number of standard deviations from the mean are identified. The number of standard deviations to be considered is customizable. Confidence intervals are taken into consideration for each field.

Pattern-based: Identify outlier fields and records that do not conform to existing patterns in the data. Combined techniques (partitioning, classification, and clustering) are used to identify patterns that apply to most records. A pattern is defined by a group of records that have similar characteristics ("behavior") for p% of the fields in the data set, where p is a user-defined value (usually above 90).

Clustering: Identify outlier records using clustering based on Euclidian (or other) distance. Existing clustering algorithms provide little support for identifying outliers. However, in some cases clustering the entire record space can reveal outliers that are not identified at the field level inspection. The main drawback of this method is computational time. The clustering algorithms have high computational complexity. For large record spaces and large number of records, the run time of the clustering algorithms is prohibitive. **(Pg. 164)**

22. In Data Extraction, "change data capture" is considered as the most challenging activity. Why?02

Change Data Capture is therefore, typically the most challenging technical issue in data extraction.

Two CDC sources • Modern systems • Legacy systems

Without Change Data Capture, database extraction is a cumbersome process in which you move the entire contents of tables into flat files, and then load the files into the data Warehouse. This ad hoc approach is expensive in a number of ways.

25. Identify the given statements as correct or incorrect and justify your answer in either case.

1. **"Transformation is the process in which we extract the data from single/multiple data sources".**

2. **"Offline Extraction is a type of Logical Data Extraction".**05

   Answer:

   2."Offline Extraction is a type of Physical Data Extraction"

21. Differentiate between MOLAP and ROLAP in terms of implementation. 02

Answer: MOLAP physically builds "cubes" for direct access - usually in the proprietary file format of a multi-dimensional database (MDD) or a user defined data structure. Therefore ANSI SQL is not supported.

ROLAP or a Relational OLAP provides access to information via a relational database using ANSI standard SQL. **(Pg. 78)**

22. Differentiate between One-to-One Scalar transformation and One-to-Many transformation on data warehouse.02

Answer: Simple one-to-one scalar transformations - 0/1 → M/F

One-to-many element transformations - 4 x 20 address field → House/Flat, Road/Street, Area/Sector, City.

23. What are the good features of DOLAP that distinguish it from other techniques?03

24. One of the steps of domain value validation is: "The occurrences of each domain value within each coded attribute of the database". In your point of view, how is this step performed. Give a real life example to explain your answer.03

Ans: Value validation is the process of ensuring that each value that is sent to the data warehouse is accurate. You may had that experience in which you look at the contents of one of your major flat files or database tables and intuitively pick that the data is incorrect. No way could that employee be born in 2004! You know your company doesn't hire infants. You may also discover another incorrect record. How someone could be born in 1978 but hired in 1977?

25. How dimensional modeling Differ from ER Modeling?05

ER	DMs
Constituted to optimize OLTP performance	Constituted to optimize DSS query performance
Models the micro relationships among data elements	Models the macro relationships among data elements with an overall deterministic strategy.
A wild variability of the structure of ER models	All dimensions serve as equal entry points to the fact table.
Very vulnerable to changes in the user's querying habits, because suchs schemas are asymmetrical	Changes in user querying habits can be catered by automatic SQL generators

26. splitting of Single Fields Transformation is used to store individual components of names and addresses in separate fields in data warehouse. In your point of view, what are the main reasons of doing this? 05

21. What is the difference between MOLAP and DOLAP in terms of their implementation? 02

Answer: MOLAP: OLAP implemented with a multi-dimensional data structure.

DOLAP: OLAP implemented for desktop decision support environments.

MOLAP physically builds "cubes" for direct access - usually in the proprietary file format of a multi-dimensional database (MDD) or a user defined data structure.

DOLAP allows download of "cube" structures to a desktop platform without the need for shared relational or cube server. **(Pg. 78)**

23. **What is O (1) time. How MOLAP uses O (1) for increasing the efficiency? 03**

Ans: The performance in a MOLAP cube comes from the O(1) look-up time for the array data
structure. Recall that to access an array, only the indexes are required i.e. there is no scanning of the array (like a file data structure), there is no hashing it a constant access time operations, similar to a random access memory (or RAM). The only time the time complexity goes beyond O(1) is when the cube size is so large that it cannot fit in the main memory, in such a case a page or a block fault will occur.

24. Suppose window size of BSN (Basic Sorted Neighborhood) method is increased. In your point of view, how is the complexity of the BSN method affected? 03Ans: **Complexity Analysis of BSN Method**

- Time Complexity: O(n log n)
- O (n) for Key Creation
- O (n log n) for Sorting
- O (w n) for matching, where w ≤ 2 ≤ n
- Constants vary a lot
- At least three passes required on the dataset.
- For large sets disk I/O is detrimental.
- Complexity or rule and window size detrimental.

25. If wrong data is used at government level stored in data warehouse, how will decision making at government level produce undesirable results?05

Ans: Decisions taken at government level using wrong data resulting in undesirable results. **Administration**: The government analyses data collected by population census to decide
which regions of the country require further investments in health, education, clean drinking water, electricity etc. because of current and expected future trends. If the rate of birth in one region has increased over the last couple of years, the existing health facilities and doctors employed might not be sufficient to handle the number of current and expected patients. Thus, additional dispensaries or employment of doctors will be needed. Inaccuracies in analyzed data can lead to false conclusions and misdirected

26. Consider the following data, showing items sold and the discount during Monday and Thursday. Identify the additive and non additive data from the following given table? Explain with reasons.

Day	No Of Items sold
Monday	10
Thursday	15
TOTAL	25

Day	% Discount
Monday	9
Thursday	7
TOTAL	24%

Additive facts are easy to work with

- Summing the fact value gives meaningful results
- Additive facts:
- Quantity sold
- Total Rs. sales
- **Non-additive facts:**
- Averages (average sales price, unit price)
- Percentages (% discount)

- Ratios (gross margin)
- Count of distinct products sold

22. The problems associated with the extracted data can correspond to non-primary keys. List down any four problems associated with the non-primary key. 02

Ans: Non primary key problems…

1. Different encoding in different sources. 2. Multiple ways to represent the same information.

3. Sources might contain invalid data. 4. Two fields with different data but same name.

Primary key problems 1. Same PK but different data. 2. Same entity with different keys.

3. PK in one system but not in other. 4. Same PK but in different formats.

25. Consider a fact table with name "Sales". The grain of table might be stated as "Sales volume by Day by Product by Store". Identify few facts (at least three) that can be used to populate such table. 05

22. In Data Extraction, "change data capture" is considered as the most challenging activity. Why?02

23. Identify the given statement as correct or incorrect and justify your answer in either case.

"One of the basic purposes of an OLTP system is to represent the historical picture of an organization". 03

Answer: OLTP systems to track history, purged after 90 to 180 days.

Actually don't want to keep historical data for OLTP system. **(Pg. 122)**

24. In your point of view what may be the possible reasons to use enrichment during data transformation? Explain with an example.03

Answer: The task is the rearrangement and simplification of individual fields to make them more useful for the data warehouse environment. You may use one or more fields from the same input record to create a better view of the data for the data warehouse. This principle is extended when one or more fields originate from multiple records, resulting in a single field for the data warehouse. **(pg.136)**

26. Consider the following facts table having name "PrdocutSales":

ProductID	RegionID	Period	Quantity
01	N	Monthy	25
02	N	Monthly	50 fact
02	S	Weekly	30

Identify the dimensions, facts and primary key from the above table. 05

22. List down any four ways of "handling missing data" during "data cleansing process".02

Answer: 1. Dropping records.

2. "Manually" filling missing values.
3. Using a global constant as filler.
4. Using the attribute mean (or median) as filler.

5. Using the most probable value as filler. **(Pg. 162)**s

25. Identify the given statements as correct or incorrect and justify your answer in either case. "Incremental Data Extraction is a type of Physical Data Extraction".

1. **"In Incremental Data Extraction, there's no need to keep track of changes to the data source since the last successful extraction". 05marks**

Answer:

1. Incremental Data Extraction is a type of **Extracting Changed Data. (Pg. 149)**
2. **Incremental data extraction** i.e. what has changed, say during last 24 hrs if considering nightly extraction.

"Full Extraction "there's no need to keep track of changes to the data source since the last successful extraction.