

Practice Sheet -01

Data Science & Artificial Intelligence

Warehousing

- Q1** What is the primary goal of data normalization in the context of databases?
- (A) To increase data redundancy.
 - (B) To minimize data security.
 - (C) To reduce data redundancy and improve data integrity.
 - (D) To improve data compression.
- Q2** In the process of data normalization, what does the First Normal Form (1NF) require?
- (A) Elimination of repeating groups.
 - (B) No duplicate rows.
 - (C) Each column must have a unique name.
 - (D) No NULL values in the table.
- Q3** Which normal form ensures that there are no partial dependencies in a database table?
- (A) First Normal Form (1NF)
 - (B) Second Normal Form (2NF)
 - (C) Third Normal Form (3NF)
 - (D) Boyce-Codd Normal Form (BCNF)
- Q4** In a database, which of the following is an example of a partial dependency?
- (A) An employee's name depends on their employee ID.
 - (B) A customer's address depends on their customer ID.
 - (C) A department's manager depends on the department's location.
 - (D) A product's price depends on the product's name.
- Q5** Which normal form is considered stricter than the Boyce-Codd Normal Form (BCNF)?
- (A) First Normal Form (1NF)
 - (B) Second Normal Form (2NF)
 - (C) Third Normal Form (3NF)
 - (D) Fourth Normal Form (4NF)
- Q6** Which of the following is a commonly used method for eliminating data redundancy through data normalization?
- (A) Aggregation
 - (B) Indexing
 - (C) De-normalization
 - (D) Splitting tables and using foreign keys
- Q7** Which normal form allows no transitive dependencies?
- (A) First Normal Form (1NF)
 - (B) Second Normal Form (2NF)
 - (C) Third Normal Form (3NF)
 - (D) Boyce-Codd Normal Form (BCNF)
- Q8** In the context of data normalization, what does a candidate key represent?
- (A) A super key that uniquely identifies a table.
 - (B) A unique identifier for a table.
 - (C) A key used to connect two tables in a relational database.
 - (D) A key that identifies multiple attributes.
- Q9** Which of the following is an example of a super key?
- (A) An attribute that is not unique in the table.
 - (B) A combination of attributes that can uniquely identify a row.
 - (C) An attribute that is a foreign key.
 - (D) An attribute that contains NULL values.
- Q10** What is the primary benefit of normalizing a database?
- (A) To increase data redundancy.
 - (B) To improve data compression.
 - (C) To improve data integrity and reduce data redundancy.
 - (D) To make data retrieval faster.



- Q11** You have a dataset of temperatures in degrees Fahrenheit. You want to discretize it into three equal-width bins: Cold, Mild, and Hot. If the temperature range is 32°F to 100°F, what is the width of each bin?
- (A) 24°F (B) 34°F
(C) 33.33°F (D) 25.5°F
- Q12** In data discretization, what is the purpose of converting continuous data into discrete intervals or categories?
- (A) To simplify data analysis
(B) To increase data volume
(C) To make data visualization more challenging
(D) To preserve the original data structure
- Q13** You're discretizing a dataset of ages into three equal-frequency bins. If you have 60 records, how many records would be in each bin?
- (A) 15 (B) 20
(C) 25 (D) 30
- Q14** Which discretization method divides the data into bins such that each bin contains an approximately equal number of data points?
- (A) Equal-width discretization
(B) Equal-frequency discretization
(C) Unsupervised discretization
(D) Supervised discretization
- Q15** You have a dataset with the following values: 5, 8, 15, 22, 30, 35, 48, 52. If you want to discretize these values into two bins using equal-width discretization, what is the midpoint value that separates the two bins?
- (A) 25.5 (B) 30.5
(C) 40.5 (D) 26.5
- Q16** Which of the following is a disadvantage of data discretization?
- (A) Improved data interpretability
(B) Loss of information
(C) Simplified data analysis
(D) Faster data processing

Q17

You're discretizing a dataset into three bins using equal-width discretization. If the minimum value is 10 and the maximum value is 70, what is the width of each bin?

- (A) 20 (B) 30
(C) 40 (D) 60

- Q18** In supervised discretization, what is used to guide the bin boundaries?
- (A) Class labels
(B) Data range
(C) Mean values
(D) Standard deviation
- Q19** You have a dataset of student grades ranging from 0 to 100. If you want to discretize these grades into five equal-width bins, what is the width of each bin?
- (A) 20 (B) 25
(C) 15 (D) 10
- Q20** What is the main objective of data discretization in data mining?
- (A) To increase data complexity
(B) To ensure data is stored efficiently
(C) To transform numerical data into text data
(D) To simplify data analysis and pattern recognition
- Q21** What is the purpose of data sampling in the context of data transformation?
- (A) To reduce the size of the dataset for faster processing.
(B) To remove outliers from the dataset.
(C) To perform clustering on the data.
(D) To normalize the data.
- Q22** Which of the following is a common sampling method used in data transformation to ensure every element in the population has an equal chance of being selected?
- (A) Stratified sampling
(B) Systematic sampling
(C) Convenience sampling
(D) Snowball sampling



- Q23** In data transformation, what is oversampling primarily used for?
- (A) Reducing the size of the dataset.
(B) Balancing imbalanced datasets.
(C) Improving data visualization.
(D) Removing outliers.
- Q24** Which of the following sampling methods involves selecting a random sample of fixed-size k elements from the population at regular intervals?
- (A) Stratified sampling
(B) Simple random sampling
(C) Cluster sampling
(D) Systematic sampling
- Q25** What is the primary purpose of stratified sampling in data transformation?
- (A) To select data randomly from the entire population.
(B) To divide the population into homogenous strata and sample from each stratum.
(C) To select the largest data values in the dataset.
(D) To sort the data in ascending order.
- Q26** You have a dataset with 200 records. If you use simple random sampling to select a sample of 30 records, what is the probability of any specific record being selected?
- (A) 0.15 (B) 0.20
(C) 0.05 (D) 0.30
- Q27** In systematic sampling, if you have a population of 500 elements and you want to select a sample of 50 elements, what would be the sampling interval (k)?
- (A) 10 (B) 25
(C) 5 (D) 100
- Q28** You are conducting stratified sampling on a dataset with three strata. Stratum A has 200 records, Stratum B has 150 records, and Stratum C has 100 records. If you want to select a total of 40 records for your sample, how many records should you sample from Stratum B?
- (A) 8 (B) 12
(C) 10 (D) 20
- Q29** You're using cluster sampling for a study, and there are 10 clusters with an average of 25 elements per cluster. If you want a sample size of 50 elements, how many clusters should you select?
- (A) 2 (B) 4
(C) 5 (D) 10
- Q30** You have a dataset with 400 records, and you decide to use stratified sampling. Stratum A has 100 records, Stratum B has 200 records, and Stratum C has 100 records. If you want to select a sample of 50 records, how many records should you sample from Stratum A?
- (A) 20 (B) 25
(C) 10 (D) 15
- Q31** What is the purpose of data compression?
- (A) To increase the size of data for better quality.
(B) To reduce data redundancy and minimize the size for efficient storage and transmission.
(C) To convert data into a different format.
(D) To encrypt data for security.
- Q32** Which compression type preserves all the original data and is reversible?
- (A) Lossy compression
(B) Lossless compression
(C) Run-Length Encoding
(D) Huffman coding
- Q33** Which of the following is a common lossless compression algorithm used for text data?
- (A) JPEG (B) ZIP
(C) MP3 (D) MPEG
- Q34** Which of the following lossy compression formats is commonly used for audio files?
- (A) BMP (B) PNG
(C) WAV (D) MP3



- Q35** What does "bitrate" refer to in the context of data compression?
(A) The rate at which data is compressed.
(B) The number of bits used to represent one unit of data.
(C) The quality of the compressed data.
(D) The randomness of the data.
- Q36** What is the purpose of delta encoding in data transformation?
(A) To encrypt data for security.
(B) To increase data redundancy for error correction.
(C) To reduce data redundancy by storing only the differences between sequential data.
(D) To convert data into a different format.
- Q37** Which algorithm is commonly used for lossless compression of images?
(A) JPEG (B) GIF
(C) PNG (D) MP3
- Q38** In data compression, what is "entropy"?
(A) The level of data corruption during compression.
(B) The randomness or unpredictability of data.
(C) The speed at which data is compressed.
(D) The quality of the compressed data.
- Q39** Which compression method is commonly used in video streaming for reducing the amount of data sent over the internet?
(A) JPEG
(B) Run-Length Encoding
(C) H.264 (MPEG-4 AVC)
(D) Lempel-Ziv-Welch (LZW)
- Q40** What is the main purpose of Huffman coding in data compression?
(A) To increase the size of data for better quality.
(B) To convert data into a different format.
(C) To encrypt data for security.
(D) To reduce data redundancy and minimize the size for efficient storage and transmission.
- Q41** What is the primary purpose of data warehousing?
(A) To store transactional data
(B) To process real-time data
(C) To support decision-making
(D) To conduct data analysis
- Q42** What is the ETL process in data warehousing?
(A) Extract, Transform, Load
(B) Extract, Transfer, Load
(C) Extract, Transport, Load
(D) Extract, Transmit, Load
- Q43** Which of the following is a dimension table in a data warehouse?
(A) Product sales by date
(B) Customer names and addresses
(C) Daily website traffic
(D) Revenue by region
- Q44** What is a fact table in data warehousing?
(A) A table that contains descriptive data
(B) A table that stores transactional data
(C) A table that stores metadata
(D) A table that stores historical data
- Q45** In data warehousing, what does the term "star schema" refer to?
(A) A schema design with multiple interconnected tables
(B) A schema design with a central fact table connected to dimension tables
(C) A schema design that uses a hierarchy of dimensions
(D) A schema design that represents unstructured data
- Q46** Which data modeling technique is often used for data warehousing to represent hierarchical structures?
(A) Entity-Relationship Diagrams (ERD)
(B) Dimensional Modeling
(C) Normalization
(D) Data Mining
- Q47**



What is the purpose of surrogate keys in data warehousing?

- (A) To provide additional security for the data.
- (B) To encrypt sensitive information.
- (C) To provide unique identifiers for dimension records.
- (D) To store historical data.

Q48 Which type of schema is designed to optimize queries and reporting in a data warehouse?

- (A) Snowflake schema
- (B) Star schema
- (C) Normalized schema
- (D) Denormalized schema

Q49 What is the purpose of slowly changing dimensions (SCDs) in data warehousing?

- (A) To track changes in data over time.
- (B) To improve data loading performance.
- (C) To reduce storage costs.
- (D) To encrypt sensitive data.

Q50 Which SQL operation is used to combine rows from different tables in a data warehouse?

- (A) JOIN
- (B) SELECT
- (C) INSERT
- (D) UPDATE

Q51 In data warehousing, what does "ETL" stand for?

- (A) Extract, Transfer, Load
- (B) Extract, Transform, Load
- (C) Extract, Transmit, Load
- (D) Extract, Table, Load

Q52 Which of the following is an example of a business intelligence tool often used with data warehouses?

- (A) Microsoft Word
- (B) Tableau
- (C) Photoshop
- (D) Notepad

Q53 What is the primary function of a data warehouse's data marts?

- (A) To store raw, unprocessed data.
- (B) To support data extraction.
- (C) To provide specialized subsets of data.
- (D) To perform real-time processing.

Q54 Which of the following is a common data modeling technique in data warehousing?

- (A) Entity-Relationship Diagrams (ERD)
- (B) Data Lake
- (C) NoSQL
- (D) OLAP

Q55 What is the purpose of OLAP (Online Analytical Processing) in data warehousing?

- (A) To store raw transactional data.
- (B) To support real-time data processing.
- (C) To facilitate complex data analysis and reporting.
- (D) To back up data

Q56 In data warehousing, what is the role of a star schema's fact table?

- (A) To store transactional data.
- (B) To store metadata.
- (C) To hold dimension data.
- (D) To create hierarchical structures.

Q57 What is the purpose of a data warehouse's metadata repository?

- (A) To store user passwords.
- (B) To manage encryption keys.
- (C) To store information about the data stored in the warehouse.
- (D) To store transactional data.

Q58 Which of the following is a common data modeling technique for a data warehouse that reduces data redundancy?

- (A) Normalization
- (B) Denormalization
- (C) Star schema
- (D) Snowflake schema

Q59 What is a "data warehouse bus" in the context of data warehousing?

- (A) A physical bus used to transport data.
- (B) A set of related data marts.
- (C) A data integration tool.
- (D) A security feature.

Q60



[Android App](#)

| [iOS App](#)

| [PW Website](#)

What is the purpose of a snowflake schema in data warehousing?

- (A) To optimize query performance.
- (B) To simplify data modeling.
- (C) To reduce data redundancy.
- (D) To create hierarchical structures.

Q61 What is a "star join" in data warehousing?

- (A) A type of dimension table.
- (B) A query optimization technique.
- (C) A data loading process.
- (D) A type of data encryption.

Q62 In data warehousing, what is the role of a bridge table in a snowflake schema?

- (A) To store sensitive data.
- (B) To store historical data.
- (C) To create relationships between dimension tables.
- (D) To optimize query performance.

Q63 Which of the following is not a common data modeling technique in data warehousing?

- (A) Entity-Relationship Diagrams (ERD)
- (B) Dimensional Modeling
- (C) NoSQL
- (D) Relational Modeling

Q64 What is a "starflake schema" in data warehousing?

- (A) A combination of star and snowflake schemas.
- (B) A schema design for unstructured data.
- (C) A data warehousing benchmark.
- (D) A data security protocol

Q65 What is the primary goal of data normalization in data warehousing?

- (A) To improve query performance.
- (B) To create hierarchical structures.

(C) To reduce data redundancy.

(D) To support real-time processing.

Q66 In data warehousing, what is the primary purpose of a data transformation process?

- (A) To encrypt data.
- (B) To load data into the warehouse.
- (C) To process and clean data for analysis.
- (D) To store data.

Q67 Which data modeling technique is often used for unstructured data in data warehousing?

- (A) Star schema
- (B) Dimensional Modeling
- (C) NoSQL
- (D) Normalization

Q68 What is the role of a data warehouse architect in the data warehousing process?

- (A) To design data visualizations.
- (B) To create ETL jobs.
- (C) To plan and design the data warehouse infrastructure.
- (D) To write SQL queries.

Q69 Which of the following is an advantage of denormalization in data warehousing?

- (A) Reduced query performance
- (B) Improved data integrity
- (C) Simplified query complexity
- (D) Increased data redundancy

Q70 What is the primary goal of data profiling in data warehousing?

- (A) To monitor data warehouse performance.
- (B) To track changes in data over time.
- (C) To improve query performance.
- (D) To understand the content and quality of data.



Answer Key

Q1 (C)
Q2 (A)
Q3 (C)
Q4 (C)
Q5 (D)
Q6 (D)
Q7 (C)
Q8 (B)
Q9 (B)
Q10 (C)
Q11 (A)
Q12 (A)
Q13 (B)
Q14 (B)
Q15 (B)
Q16 (B)
Q17 (B)
Q18 (A)
Q19 (B)
Q20 (D)
Q21 (A)
Q22 (B)
Q23 (B)
Q24 (D)
Q25 (B)
Q26 (B)
Q27 (A)
Q28 (B)
Q29 (A)
Q30 (A)

Q31 (B)
Q32 (B)
Q33 (B)
Q34 (D)
Q35 (B)
Q36 (C)
Q37 (C)
Q38 (B)
Q39 (C)
Q40 (D)
Q41 (C)
Q42 (A)
Q43 (B)
Q44 (B)
Q45 (B)
Q46 (B)
Q47 (C)
Q48 (B)
Q49 (A)
Q50 (A)
Q51 (B)
Q52 (B)
Q53 (C)
Q54 (A)
Q55 (C)
Q56 (A)
Q57 (C)
Q58 (A)
Q59 (B)
Q60 (D)

[Android App](#)[iOS App](#)[PW Website](#)

Q61 (B)
Q62 (C)
Q63 (C)
Q64 (A)
Q65 (C)

Q66 (C)
Q67 (C)
Q68 (C)
Q69 (C)
Q70 (D)



[Android App](#) | [iOS App](#) | [PW Website](#)

Hints & Solutions

Q1 Text Solution:

Data normalization is primarily aimed at organizing data in a database to minimize redundancy and enhance data integrity.

Q2 Text Solution:

1NF requires that each column in a table should contain only atomic (indivisible) values, eliminating any repeating groups of data.

Q3 Text Solution:

3NF ensures that there are no partial dependencies, meaning that non-prime attributes are functionally dependent on the primary key.

Q4 Text Solution:

A partial dependency occurs when non-prime attributes (in this case, the manager) depend on part of the candidate key (location).

Q5 Text Solution:

4NF is stricter than BCNF and deals with multi-valued dependencies in addition to functional dependencies.

Q6 Text Solution:

Data normalization often involves splitting tables and using foreign keys to eliminate redundancy and ensure data integrity.

Q7 Text Solution:

3NF ensures that there are no transitive dependencies, meaning that non-prime attributes do not depend on other non-prime attributes.

Q8 Text Solution:

A candidate key is a key that can uniquely identify each row in a table.

Q9 Text Solution:

A super key is a set of attributes that can uniquely identify a row in a table.

Q10 Text Solution:

Normalizing a database helps improve data integrity and reduce data redundancy, making it more efficient and maintaining consistent data.

Q11 Text Solution:

Explanation: To create three equal-width bins, you divide the range by the number of bins:
 $(100^{\circ}\text{F} - 32^{\circ}\text{F}) / 3 = 68^{\circ}\text{F} / 3 = 24^{\circ}\text{F}$.

Q12 Text Solution:

Data discretization is used to simplify the analysis of continuous data by grouping it into discrete intervals or categories.

Q13 Text Solution:

To create equal-frequency bins, you divide the total number of records by the number of bins:
 $60 \text{ records} / 3 \text{ bins} = 20 \text{ records per bin}$.

Q14 Text Solution:

Equal-frequency discretization divides the data into bins where each bin contains a similar number of data points.

Q15 Text Solution:

To create two equal-width bins, you find the midpoint between the minimum and maximum values: $(5 + 52) / 2 = 30.5$.

Q16 Text Solution:

One of the disadvantages of data discretization is the loss of information due to the conversion of continuous data into discrete categories.

Q17 Text Solution:

To create three equal-width bins, you divide the range by the number of bins: $(70 - 10) / 3 = 60 / 3 = 30$.

Q18 Text Solution:

In supervised discretization, class labels are used to guide the bin boundaries, ensuring that the discretization is done with respect to class distinctions.



Q19 Text Solution:

To create five equal-width bins, you divide the range by the number of bins: $(100 - 0) / 5 = 100 / 5 = 20$.

Q20 Text Solution:

The main objective of data discretization in data mining is to simplify data analysis and pattern recognition by converting continuous data into discrete intervals or categories.

Q21 Text Solution:

Data sampling is used to select a subset of data from a larger dataset, often for the purpose of reducing its size while retaining essential characteristics. This can lead to faster processing and analysis of data.

Q22 Text Solution:

Systematic sampling involves selecting every n th element from the dataset in a systematic manner, ensuring that each element has an equal chance of being included in the sample.

Q23 Text Solution:

Oversampling is a technique used to address imbalanced datasets by generating synthetic examples of the minority class to balance the class distribution.

Q24 Text Solution:

Systematic sampling selects every n th element from the population at regular intervals, leading to a random sample of fixed size k .

Q25 Text Solution:

Stratified sampling divides the population into subgroups (strata) based on specific characteristics and then samples from each stratum to ensure representation of each group. You can use these examples as a basis for creating additional questions on data transformation and sampling methods, changing the content, and format as needed.

Q26 Text Solution:

In simple random sampling, each record has an equal probability of being selected. Therefore,

the probability of a specific record being selected is $30/200 = 0.20$.

Q27 Text Solution:

To calculate the sampling interval (k), you divide the population size by the sample size: $500 / 50 = 10$.

Q28 Text Solution:

To allocate the sample size proportionally, you would sample from each stratum based on its size relative to the total population. Stratum B constitutes $150/450$ ($1/3$) of the population, so you would sample $1/3$ of the total sample size: $(1/3) \times 40 = 12$.

Q29 Text Solution:

To achieve a sample size of 50 elements, you need to select clusters. Given that each cluster has 25 elements on average, you would need to select 2 clusters ($2 \times 25 = 50$).

Q30 Text Solution:

To allocate the sample size proportionally, you would sample from each stratum based on its size relative to the total population. Stratum A constitutes $100/400$ ($1/4$) of the population, so you would sample $1/4$ of the total sample size: $(1/4) \times 50 = 12.5$.

Q31 Text Solution:

Data compression is used to reduce the size of data, making it more efficient for storage and transmission by eliminating redundant information.

Q32 Text Solution:

Lossless compression techniques maintain all the original data, and the compression process is fully reversible.

Q33 Text Solution:

ZIP is a common lossless compression algorithm used for text data, as well as other types of data like documents, spreadsheets, and more.

Q34 Text Solution:

MP3 is a widely used lossy compression format for audio files, allowing significant reduction in



file size while maintaining acceptable audio quality.

Q35 Text Solution:

Bitrate refers to the number of bits used to represent a specific amount of data, typically measured in bits per second (bps).

Q36 Text Solution:

Delta encoding is a method of data transformation that stores the differences between successive values instead of storing each value separately, reducing redundancy in sequential data.

Q37 Text Solution:

PNG (Portable Network Graphics) is a common lossless compression format used for images, preserving the original image quality without loss.

Q38 Text Solution:

Entropy is a measure of the randomness or unpredictability of data, which is used in various compression algorithms to analyze and reduce data redundancy.

Q39 Text Solution:

H.264, also known as MPEG-4 AVC, is a commonly used compression standard for video streaming, reducing the amount of data transmitted over the internet while maintaining video quality.

Q40 Text Solution:

Huffman coding is a commonly used technique in data compression to assign variable-length codes to different symbols in a way that minimizes the overall data size. This reduces redundancy and improves storage and transmission efficiency.

Q41 Text Solution:

The primary purpose of data warehousing is to support decision-making by providing a centralized repository of structured data for analysis and reporting.

Q42 Text Solution:

The ETL process stands for Extract, Transform, Load. It involves extracting data from source systems, transforming it to meet business needs, and loading it into the data warehouse.

Q43 Text Solution:

Customer names and addresses are typically stored in a dimension table in a data warehouse.

Q44 Text Solution:

A fact table in data warehousing stores transactional data, such as sales revenue, quantity sold, or any measurable event.

Q45 Text Solution:

A star schema in data warehousing is a schema design where a central fact table is connected to dimension tables, resembling a star.

Q46 Text Solution:

Dimensional modeling is often used in data warehousing to represent hierarchical structures.

Q47 Text Solution:

Surrogate keys are used in data warehousing to provide unique identifiers for dimension records, helping to maintain data integrity.

Q48 Text Solution:

The star schema is designed to optimize queries and reporting in a data warehouse by simplifying data retrieval through a central fact table connected to dimension tables.

Q49 Text Solution:

Slowly changing dimensions (SCDs) are used to track changes in data over time, allowing historical analysis.

Q50 Text Solution:

The SQL operation used to combine rows from different tables in a data warehouse is JOIN.

Q51 Text Solution:

In data warehousing, ETL stands for Extract, Transform, Load.

Q52 Text Solution:

Tableau is a popular business intelligence tool often used with data warehouses for data visualization and analysis.

Q53 Text Solution:

The primary function of a data warehouse's data marts is to provide specialized subsets of data to specific user groups or departments.

Q54 Text Solution:

Entity-Relationship Diagrams (ERD) is a common data modeling technique, but it is often used in conjunction with other techniques in data warehousing.

Q55 Text Solution:

The purpose of OLAP in data warehousing is to facilitate complex data analysis and reporting, enabling users to perform multidimensional analysis.

Q56 Text Solution:

The fact table in a star schema is used to store transactional data, such as sales, revenue, or quantity sold.

Q57 Text Solution:

The purpose of a data warehouse's metadata repository is to store information about the data stored in the warehouse, including data lineage, data definitions, and data transformation details.

Q58 Text Solution:

Normalization is a data modeling technique that reduces data redundancy but is often not used in its purest form in data warehousing.

Q59 Text Solution:

In data warehousing, a "data warehouse bus" typically refers to a set of related data marts designed to address specific business areas or departments.

Q60 Text Solution:

The purpose of a snowflake schema in data warehousing is to create hierarchical structures by normalizing dimension tables.

Q61 Text Solution:

In data warehousing, a "star join" refers to a query optimization technique that leverages the structure of a star schema to improve query performance.

Q62 Text Solution:

A bridge table in a snowflake schema is used to create relationships between dimension tables, allowing for many-to-many relationships.

Q63 Text Solution:

NoSQL is not typically a data modeling technique used in traditional data warehousing, as it is a different database paradigm.

Q64 Text Solution:

A "starflake schema" is a combination of both star and snowflake schema elements in data warehousing.

Q65 Text Solution:

The primary goal of data normalization in data warehousing is to reduce data redundancy, leading to more efficient storage and maintenance of data.

Q66 Text Solution:

The primary purpose of a data transformation process in data warehousing is to process and clean data for analysis, making it suitable for reporting and querying.

Q67 Text Solution:

NoSQL databases are often used for unstructured data in data warehousing, as they can handle diverse data types and structures.

Q68 Text Solution:

The role of a data warehouse architect is to plan and design the data warehouse infrastructure, including schema design and data integration strategies.

Q69 Text Solution:

Denormalization in data warehousing can simplify query complexity and improve query performance by reducing the need for complex joins.

Q70 Text Solution:

The primary goal of data profiling in data warehousing is to understand the content and quality of data, including data distribution,

completeness, and potential data quality issues.

[Android App](#)[iOS App](#)[PW Website](#)