

SRM VALLIAMMAI ENGINEERING COLLEGE

(An Autonomous Institution)
SRM Nagar, Kattankulathur – 603 203

DEPARTMENT OF ARTIFICIAL INTELLIGENCE AND DATA SCIENCE

QUESTION BANK



VIII SEMESTER
1922805 – DATA QUALITY AND TRANSFORMATION
Regulation – 2019
Academic Year 2025 – 2026 EVEN

Prepared by

Mr.R. Danu, Asst.Prof/ AI-DS



SRM VALLIAMMAI ENGINEERING COLLEGE

SRM Nagar, Kattankulathur – 603203.



DEPARTMENT OF ARTIFICIAL INTELLIGENCE AND DATA SCIENCE QUESTION BANK

SUBJECT: 1922805 – DATA QUALITY AND TRANSFORMATION

YEAR / SEM: IV / VIII

UNIT I - INTRODUCTION

Introduction to the concept of Data Quality-Data Quality and Types of Data-Data Quality and Types of Information Systems-Research Issues and Application Domains in Data Quality: Research Issues in Data Quality-Application Domains in Data Quality-Research Areas Related to Data Quality

PART - A (2 - MARKS)

Q. No	QUESTIONS	Competence	BT Level
1.	Define DQ.	Remembering	BTL-1
2.	List the examples of the importance of data quality in organizational processes.	Remembering	BTL-1
3.	Explain Corporate house-holding.	Understanding	BTL-2
4.	Differentiate private and public initiatives.	Understanding	BTL-2
5.	Define integrity.	Remembering	BTL-1
6.	How data quality has been addressed from research perspective?	Remembering	BTL-1
7.	What is completeness error?	Remembering	BTL-1
8.	Define semi structured data.	Remembering	BTL-1
9.	Differentiate elementary and aggregate data.	Understanding	BTL-2
10.	Assess component data items.	Understanding	BTL-2
11.	What is federated data?	Remembering	BTL-1
12.	What is web data?	Remembering	BTL-1
13.	What are the 3 categories of data based on change frequency?	Remembering	BTL-1
14.	Define stable data.	Remembering	BTL-1
15.	Brief DW.	Understanding	BTL-2
16.	Define CIS.	Remembering	BTL-1
17.	What are the five main types of information systems?	Remembering	BTL-1
18.	Explain the two hierarchy of anatomy.	Understanding	BTL-2
19.	Explain peer-to-peer information system.	Understanding	BTL-2
20.	Sketch on Main issues in data quality.	Understanding	BTL-2
21.	Define organizational data dependency in the context of data quality.	Remembering	BTL-1
22.	What is the role of metadata in ensuring data quality?	Understanding	BTL-2
23.	Differentiate transactional data and analytical data.	Remembering	BTL-1
24.	What is meant by structural data issues?	Remembering	BTL-1

PART - B (13 - MARKS)

1.	Why data quality is relevant? Explain the importance of data quality in organizational process.	(13)	Evaluating	BTL-5
2.	Discuss the concept of Data Quality.	(13)	Applying	BTL-3

3.	(i) Briefly describe Data Quality. (ii) Explain the different types of data.	(7) (6)	Analyzing	BTL-4
4.	Define data quality and types of information systems in detail.	(13)	Applying	BTL-3
5.	Explain the Main Research Issues and Application Domains in Data Quality.	(13)	Applying	BTL-3
6.	Explain the Research Areas Related to Data Quality in the areas of computer science.	(13)	Applying	BTL-3
7.	Describe the application domains in data quality.	(13)	Analyzing	BTL-4
8.	Discuss the five main types of information systems with a diagram.	(13)	Analyzing	BTL-4
9.	Explain the consequences of poor quality of data often experienced in everyday life.	(13)	Applying	BTL-3
10.	Summarize some of the major initiatives in both the private and public domains.	(13)	Analyzing	BTL-4
11.	Address the data quality from a research perspective in different areas.	(13)	Analyzing	BTL-4
12.	Explain the relation movies data quality problem.	(13)	Applying	BTL-3
13.	Explain how different types of information systems influence data quality requirements.	(7) (6)	Evaluating	BTL-5
14.	Analyze the importance of research methodologies in identifying data quality challenges.	(5) (8)	Evaluating	BTL-5
15.	Explain about the three different criteria proposed for distributed databases: distribution, heterogeneity, and autonomy.	(13)	Applying	BTL-3
16.	Explain the main issues in data quality with a pictorial representation.	(13)	Analyzing	BTL-4
17.	Analyze the three distinct application domains of DQ.	(13)	Analyzing	BTL-4

PART - C (15 - MARK)

1.	Explain in detail about the Main Research Issues and Application Domains in Data Quality.	(15)	Evaluating	BTL-5
2.	Investigate the Research Areas Related to Data Quality.	(15)	Creating	BTL-6
3.	Explain data quality and types of information system with a neat diagram.	(15)	Analyzing	BTL-4
4.	Explain the different types of data based on different perspective.	(15)	Analyzing	BTL-4
5.	Evaluate the major application domains of data quality and discuss how quality issues differ across each domain.	(15)	Creating	BTL-6

UNIT II - DATA QUALITY DIMENSIONS

Accuracy-Completeness: Completeness of Relational Data and Web Data- Time-related Dimensions: Currency, Timeliness and Volatility-Consistency: Integrity Constraints-Data Edits. Other Data Quality Dimensions - Approaches to the Definition of Data Quality Dimensions - Schema Quality Dimensions.

PART - A (2 - MARKS)

Q. No	QUESTIONS	Competence	BT Level
1.	List some of the data quality dimensions.	Remembering	BTL-1
2.	Define accuracy.	Remembering	BTL-1
3.	List and explain two kinds of accuracy.	Remembering	BTL-1
4.	What is semantic accuracy?	Remembering	BTL-1

5.	What are the two main issues to be addressed for solving the object identification problem?		Remembering	BTL-1
6.	List the three types of accuracy for relational model.		Remembering	BTL-1
7.	What is duplication in accuracy?		Remembering	BTL-1
8.	Explain the metrics used for the calculation of accuracy.		Understanding	BTL-2
9.	What is strong accuracy error?		Remembering	BTL-1
10.	List and explain three types of completeness.		Remembering	BTL-1
11.	How completeness in the relational model can be characterized?		Understanding	BTL-2
12.	Represent the Completeness of different elements in the relational model by a diagram.		Understanding	BTL-2
13.	Show the graphical representation of completability.		Remembering	BTL-1
14.	What is volatility?		Remembering	BTL-1
15.	Define Currency.		Remembering	BTL-1
16.	What are the different types of dependency?		Remembering	BTL-1
17.	Define edit-imputation problem.		Remembering	BTL-1
18.	Explain geographical and geospatial domain.		Understanding	BTL-2
19.	What type of documentation needed to maximize interpretability?		Remembering	BTL-1
20.	List the three main approaches adopted for proposing comprehensive sets of the dimension definitions.		Remembering	BTL-1
21.	Define semantic completeness with an example.		Understanding	BTL-2
22.	What is temporal consistency?		Understanding	BTL-2
23.	List two key challenges in measuring volatility of data.		Remembering	BTL-1
24.	Define interpretability in the context of data quality dimensions.		Remembering	BTL-1
PART - B (13 - MARKS)				
1.	Describe briefly the data quality dimension: Accuracy.	(13)	Analyzing	BTL-4
2.	Discuss the two types of accuracy in detail with example.	(13)	Analyzing	BTL-4
3.	Explain the metrics used for the calculation of accuracy.	(13)	Applying	BTL-3
4.	Describe in detail about different types of accuracy used for calculation of DQ.	(13)	Analyzing	BTL-4
5.	Describe briefly the data quality dimension: Completeness.	(13)	Analyzing	BTL-4
6.	Specify the three types of completeness in detail.	(13)	Analyzing	BTL-4
7.	Exemplify completeness of relational data.	(13)	Applying	BTL-3
8.	Illustrate with an example to explain the presence and absence of NULL values in the completeness of the relational model.	(13)	Applying	BTL-3
9.	Explain the relationship between accuracy, consistency, and completeness with real-time data examples.	(13)	Analyzing	BTL-4
10.	Discuss the schema-related dimensions that affect the overall data quality of a database.	(6) (7)	Evaluating	BTL-5
11.	(i) Demonstrate the relational model without null values with open world assumption.	(6)	Applying	BTL-3
	(ii) Illustrate the relational model with null values with closed world assumption.	(7)		
12.	Illustrate the completeness of web data.	(13)	Applying	BTL-3
13.	Exemplify the completeness of tuples, attributes and relations in Student relation.	(13)	Applying	BTL-3

14.	Discuss in detail about Time-Related Dimensions: Currency, Timeliness, and Volatility.	(13)	Analyzing	BTL-4
15.	Explain in detail about the DQ dimension: Consistency.	(13)	Analyzing	BTL-4
16.	Discuss the domains, the specific data quality dimensions are proposed for.	(13)	Analyzing	BTL-4
17.	Illustrate the following DQ dimensions. (i) Accessibility (ii) Quality of Information Sources	(6) (7)	Applying	BTL-3

PART - C (15 - MARK)

1.	Explain (i) Completeness of relational data (ii) Completeness of web data	(8) (7)	Applying	BTL-3
2.	Discuss the Time-Related Dimensions: Currency, Timeliness, and Volatility.	(15)	Analyzing	BTL-4
3.	Examine the DQ dimension: Consistency.	(15)	Analyzing	BTL-4
4.	Analyze the empirical approach to defining data quality dimensions and evaluate its impact on data assessment.	(15)	Applying	BTL-3
5.	Investigate the following approaches (i) Theoretical Approach (ii) Empirical Approach	(10) (5)	Creating	BTL-6

UNIT III - MODELS FOR DATA QUALITY

Introduction-Extensions of Structured Data Models: Conceptual Models-Logical Models for Data Description-The Polygen Model for Data Manipulation-Data Provenance. Extensions to Semi-structured Data Models-Management Information System Models: IP-MAP model-Extensions of IP-MAP-Data Models.

PART-A (2 - MARKS)

1.	Distinguish conceptual and logical schema.		Understanding	BTL-2
2.	Give an example of quality dimension represented in the Entity Relationship Model.		Understanding	BTL-2
3.	What is polygen domain?		Remembering	BTL-1
4.	Define restrict operator.		Remembering	BTL-1
5.	What is Data provenance?		Remembering	BTL-1
6.	What are the two types of provenance defined in the literature?		Remembering	BTL-1
7.	Define reverse query approach.		Remembering	BTL-1
8.	List the three possible types of execution or propagation schemes in forward propagation approach.		Remembering	BTL-1
9.	Define D ² Q model.		Remembering	BTL-1
10.	List the main features of the D ² Q model.		Remembering	BTL-1
11.	Define quality association function.		Remembering	BTL-1
12.	Define IP-MAP.		Remembering	BTL-1
13.	Define any two IP-MAP construct blocks.		Remembering	BTL-1
14.	What are event process chain diagrams?		Remembering	BTL-1
15.	List the advantages of IP-UML formalism.		Remembering	BTL-1
16.	What are the different types of model comes under data quality profile?		Remembering	BTL-1
17.	Define quality data class.		Remembering	BTL-1

18.	Explain the IP-MAP elements.		Understanding	BTL-2
19.	What are the limitations of IP-MAP?		Remembering	BTL-1
20.	How an input flow to a process can be transformed in the domain of a specific organization?		Understanding	BTL-2
21.	What is model-driven data assessment?		Remembering	BTL-1
22.	Define structural provenance.		Understanding	BTL-2
23.	Mention two limitations of polygen-based data manipulation.		Remembering	BTL-1
24.	Define quality constraint rules in conceptual models.		Understanding	BTL-2

PART - B (13 - MARKS)

1.	Illustrate the conceptual model in detail.	(13)	Analyzing	BTL-4
2.	Explain how the complete data quality schema is made for the conceptual model with a diagrammatic representation?	(13)	Analyzing	BTL-4
3.	Explain the models (i) Logical Models for Data Description (ii) The Polygen Model for Data Manipulation	(6) (7)	Applying	BTL-3
4.	Elaborate Data Provenance in detail.	(13)	Applying	BTL-3
5.	Compare conceptual, logical, and semi-structured models in terms of their support for data quality enforcement.	(13)	Evaluating	BTL-5
6.	Explain the extensions introduced in management information system models for enhanced data evaluation.	(13)	Analyzing	BTL-4
7.	Discuss Data and data quality schema (D ² Q) with a neat diagram.	(13)	Analyzing	BTL-4
8.	Discuss IP-MAP model.	(13)	Analyzing	BTL-4
9.	Elaborate Extensions of IP-MAP model with a neat sketch.	(13)	Applying	BTL-3
10.	Illustrate the Organizations, processes, and information flows in a Cooperative Information System with a neat diagram.	(13)	Applying	BTL-3
11.	Explain the Data Model of the Information Flows of an Organization	(13)	Applying	BTL-3
12.	Discuss about the Quality Profile Model.	(13)	Analyzing	BTL-4
13.	Discuss the stereotypes which are introduced in order to model the overall set of dimension-related requirements.	(13)	Analyzing	BTL-4
14.	Elaborate IP-UML.	(13)	Applying	BTL-3
15.	Demonstrate information production map.	(13)	Applying	BTL-3
16.	Summarize the main features of the D ² Q model.	(13)	Applying	BTL-3
17.	Explain the two client relation and mapping relation with an example.	(13)	Applying	BTL-3

PART - C(15 - MARKS)

1.	Investigate Extensions of Structured Data Models in detail.	(15)	Creating	BTL-6
2.	Discuss Data Provenance with an example.	(15)	Analyzing	BTL-4
3.	Demonstrate about the Extensions of Semi structured Data Models.	(15)	Applying	BTL-3
4.	Discuss in detail about the management information system model.	(15)	Analyzing	BTL-4
5.	Examine how IP-MAP and its extensions support organization-wide quality tracking with suitable illustrations.	(15)	Analyzing	BTL-4

UNIT IV- DATA QUALITY MEASUREMENT AND IMPROVEMENT

Basics on Data Quality Methodologies: Inputs and Outputs-Classification of Methodologies-Comparison among Data-driven and Process-driven strategies. Assessment Methodologies-Comparative Analysis of general purpose methodologies-CDQM Methodology.

PART - A (2 - MARKS)			
1.	List the Knowledge involved in the DQ measurement and improvement process.	Remembering	BTL-1
2.	What are the relevant elements involved in a DQ methodology?	Remembering	BTL-1
3.	What are the three different cost categories in DQ methodology?	Remembering	BTL-1
4.	Draw a neat sketch on the Inputs and outputs of a DQ measurement and improvement methodology.	Remembering	BTL-1
5.	List the Data quality methodologies classified according to several criteria.	Remembering	BTL-1
6.	Compare Data-driven vs. process-driven.	Understanding	BTL-2
7.	Compare General-purpose vs. special-purpose.	Understanding	BTL-2
8.	Define new data acquisition.	Remembering	BTL-1
9.	Outline the use of data edits/integrity constraints.	Remembering	BTL-1
10.	What are the two main phases characterize process-driven strategies?	Remembering	BTL-1
11.	List down the three main activities followed in assessment methodologies.	Remembering	BTL-1
12.	Summarize the phases of the assessment methodology.	Understanding	BTL-2
13.	Define objective/quantitative assessment.	Remembering	BTL-1
14.	Define subjective/qualitative assessment.	Remembering	BTL-1
15.	List few of the general purpose methodologies.	Remembering	BTL-1
16.	Summarize some of the Common phases for the assessment process.	Understanding	BTL-2
17.	Summarize some of the Common phases for the improvement process.	Understanding	BTL-2
18.	Brief about TDQM methodology.	Remembering	BTL-1
19.	Define quality verification phase and quality improvement phase.	Remembering	BTL-1
20.	Write short notes on TQDM.	Remembering	BTL-1
21.	What is quality cost estimation in DQ methodologies?	Remembering	BTL-1
22.	Define corrective data improvement.	Remembering	BTL-1
23.	List two challenges in implementing process-driven strategies.	Remembering	BTL-1
24.	Define quality verification metrics.	Understanding	BTL-2

PART - B (13 - MARKS)				
1.	Discuss in detail about the inputs and outputs on Data Quality Methodologies.	(13)	Analyzing	BTL-4
2.	Illustrate the types of knowledge involved in the DQ measurement and improvement process.	(13)	Applying	BTL-3
3.	Describe the classification of Data quality methodologies according to numerous criteria.	(13)	Applying	BTL-3
4.	Determine the comparison among Data-driven and Process-driven Strategies.	(13)	Applying	BTL-3
5.	Distinguish the three major strategies, using three distinct data quality activities on Data Quality Methodologies.	(13)	Analyzing	BTL-4
6.	Explain the role of assessment methodologies in identifying data defects and prioritizing improvement.	(13)	Applying	BTL-3
7.	Discuss the usual process followed in assessment methodologies.	(13)	Analyzing	BTL-4
8.	Discuss in detail about the Basics on Data Quality Methodologies.	(13)	Analyzing	BTL-4
9.	Demonstrate the goal of assessment methodologies.	(13)	Applying	BTL-3
10.	Describe the main phases of the assessment methodology with a neat sketch.	(13)	Applying	BTL-3

11.	(i) Describe about objective/quantitative assessment with an example. (ii) Describe about subjective/qualitative assessment with an example	(7) (6)	Applying	BTL-3
12.	Compare and Analyze General-purpose Methodologies.	(13)	Analyzing	BTL-4
13.	Discuss the common phases of the (i) Assessment process (ii) Improvement process	(7) (6)	Analyzing	BTL-4
14.	Demonstrate the Basic Common Phases Among Methodologies.	(13)	Applying	BTL-3
15.	Illustrate the TDQM Methodology.	(13)	Applying	BTL-3
16.	Discuss about the Istat Methodology with a neat diagram.	(13)	Analyzing	BTL-4
17.	Discuss the comparison framework used for evaluating various DQ methodologies.	(13)	Analyzing	BTL-4

PART-C (15-MARKS)

1.	Investigate about the CDQM methodology in detail.	(15)	Creating	BTL-6
2.	Illustrate the New technological architecture for Government-to-Business interactions with a neat sketch.	(15)	Applying	BTL-3
3.	Critically analyze how CDQM integrates organizational processes, data standards, and improvement activities.	(15)	Evaluating	BTL-5
4.	Discuss in detail about all the phases in the assessment methodologies.	(15)	Analyzing	BTL-4
5.	Illustrate in detail about the basics on data quality methodologies.	(15)	Applying	BTL-3

UNIT V - DATA TRANSFORMATION

Introduction-Benefits and Challenges-Functions of data transformations: Extraction and parsing-Translation and mapping-Filtering, aggregation, and summarization-Enrichment and imputation-Indexing and ordering - Anonymization and encryption - Modeling, typecasting, formatting, and renaming.

PART - A (2 - MARKS)

1.	List the steps involved in data transformation.	Remembering	BTL-1
2.	Define data transformation.	Remembering	BTL-1
3.	Explain smoothing.	Understanding	BTL-2
4.	Define aggregation.	Remembering	BTL-1
5.	Define discretization.	Remembering	BTL-1
6.	Write short notes on generalization.	Remembering	BTL-1
7.	What are the techniques used for normalization?	Remembering	BTL-1
8.	Define Z-Score Normalization.	Remembering	BTL-1
9.	List some of the advantages of Data Transformation in Data Mining.	Remembering	BTL-1
10.	List some of the disadvantages of Data Transformation in Data Mining.	Remembering	BTL-1
11.	What is schema-level transformation?	Understanding	BTL-2
12.	Define tokenization in data extraction.	Understanding	BTL-2
13.	State any two challenges in data anonymization.	Remembering	BTL-1
14.	What is multi-step parsing?	Understanding	BTL-2
15.	How to convert data formats and structures?	Understanding	BTL-2
16.	Which function Protects sensitive data?	Understanding	BTL-2
17.	List the ETL tools used for data transformation.	Remembering	BTL-1

18.	What are the six main stages involved in data extraction?		Remembering	BTL-1
19.	After extraction, the data may need to undergo transformation. List the tasks included.		Understanding	BTL-2
20.	Define ETL.		Remembering	BTL-1
21.	Define ELT.		Remembering	BTL-1
22.	Define web scraping and parsing.		Remembering	BTL-1
23.	List out few Tools for Effective Data Extraction.		Remembering	BTL-1
24.	What are the five best practices for data extraction?		Remembering	BTL-1
PART – B (13 MARKS)				
1	Discuss in detail about data transformation.	(13)	Analyzing	BTL-4
2	Illustrate all the steps involved in data transformation.	(13)	Applying	BTL-3
3	List out the advantages and disadvantages of Data Transformation in Data Mining.	(13)	Applying	BTL-3
4	Demonstrate the benefits and challenges of data transformation.	(13)	Applying	BTL-3
5	Describe all the functions of data transformation in detail.	(13)	Analyzing	BTL-4
6	Describe about the extraction and parsing function of data transformation.	(13)	Analyzing	BTL-4
7	Exemplify data integration.	(13)	Applying	BTL-3
8	Discuss about ETL and ELT integration.	(13)	Analyzing	BTL-4
9	Explain how enrichment and imputation techniques improve the value and usability of incomplete datasets.	(13)	Applying	BTL-3
10	Explain about data Enrichment and imputation with an example.	(13)	Applying	BTL-3
11	Discuss about indexing and ordering of data in detail.	(13)	Analyzing	BTL-4
12	Compare data Anonymization and encryption in detail.	(13)	Analyzing	BTL-4
13	Elaborate Modeling, typecasting, formatting, and renaming of data.	(13)	Applying	BTL-3
14	Elaborate data transformation steps.	(13)	Applying	BTL-3
15	Describe briefly the role of extraction and parsing in data transformation.	(13)	Analyzing	BTL-4
16	Data Anonymization vs Encryption: What's The Difference?	(13)	Analyzing	BTL-4
17	Discuss the significance of translation and mapping in large-scale data migration systems.	(13)	Applying	BTL-3
PART- C (15 MARKS)				
1	Write detailed notes on data transformation.	(15)	Remembering	BTL-1
2	Assess the functions of data transformation in detail.	(15)	Evaluating	BTL-5
3	Investigate data extraction and parsing with example.	(15)	Creating	BTL-6
4	Describe about Anonymization and encryption of data.	(15)	Analyzing	BTL-4
5	Investigate the role of data aggregation, filtering, and summarization in modern analytics pipelines, with examples.	(15)	Applying	BTL-3