DATA SCIENCE QUESTION BANK

Unit 1

1	Describe data science technology stack/ describe he rapid information factory ecosystem.
2	Explain schem-on-read and schema-on-write .
3	Define data lake,data vault and data warehouse.
4	Differentiate data lake, data vault and data warehouse.
5	Explain any five data processing tools in data science technology stack.
6	Short note on : • Spark • Mesos • Akka • Cassandra • Kafka • Elastic search • R • Scala • Python • MQTT
1	Definition of data science framework.
2	Describe CRISP-DM
3	Describe HORUS in detail with help of an example.
4	Explain in detail the top layers of a layered framework.
1	Describe the functional requirement in the business layer of the data science framework.
2	Explain sun model with help of example .
3	Write a note on types of dimensions in sun model.
4	List out the non functional requirements in the business layer of the data science framework.
5	What are the common pitfalls with requirement.
6	Explain engineering a practical business layer in detail.
1	Explain utility layer in detail.
2	Describe the various data processing utilities in the layered framework utility layer.
3	Write a code to convert text(csv) to horus.
4	Explain retrieve utility of data processing utility in detail with example.
5	Explain assess utility of data processing utility in detail.
6	Explain data vault and transform utilities data processing utility in detail.
7	Explain the various data science utilities in details.
8	Describe the maintenance utilities in the layered framework utility layer.
9	Describe the processing utilities in the layered framework utility layer.
10	Write a note on engineering a practical utility layer.

Unit 2

1	Explain the thing to be recorded in the operational management layer.
2	Explain the concept of scheduling in detail.
2	Explain in brief the indicators used in the audit layer.
3	Describe the Yoke solution used in the control sublayer.
4	Explain the five fundamental steps that form the core of the data science process.
5	Write a short note on the six supersteps for processing the data.
1	What is a data swamp? Explain the steps to avoid a data swamp.
2	Explain analytical model usage in data lake in detail.
3	Explain how the data science ecosystem can be connected to different data sources in Python.
4	Explain in detail the various shipping information terms used in any logistic company.
5	Explain any three shipping term as per Incoterm 2010.

Unit 3

1	What are the different ways to handle errors in the Assess super step.
2	Explain the data quality dimension of Assess super step.
3	Write a short note on the missing value treatment.
1	Write a note on Process super step.
2	Explain T-P-O-L-E design principles in detail.
3	Explain the various Time data entities in proces ss super step.
4	Explain the various Person data entities in process super step.
5	Explain the various Object data entities in process super step.
6	Explain the various Location data entities in process super step.
7	Explain the various Event data entities in process super step.
8	Explain the concept of 5 Why's used in the analysis phase of data science.
9	Write a note on how to complete the 5 Why's .
10	Explain the basic steps on which data science works.

Unit 4

1	Explain the various steps of data exploration and preparation in Transform super step.
2	Explain the various Feature Extraction Techniques in Transform super step with example.
3	What is hypothesis testing? Explain two most popular hypothesis testing test.
4	Explain the concept of Precision –Recall , Sensitivity –Specificity in detail.
5	Write a note on cross validation test.
6	Explain the concept of linear regression in detail.
7	Explain the concept of logistic regression in detail.

8	Explain the various Clustering techniques in Transform superstep.
9	Explain hierarchical clustering technique in transform super step.
10	Explain partial clustering technique
11	Write a note on PCA (Principal Component Analysis)
12	Write a note on Decision Treee
13	Write a note on Support Vector Mahine, network, clusters and grid.
14	Write a note on Association,classification and clustering patterns in data mining
15	Explain connectivity based clustering
16	Explin distribution based clustering
17	Write a note on Bayesian classification
18	Write a note on machine learning
19	Write a note on reinforcement learning
20	Write a note on bootstrap aggregation (bagging)
21	Write a note on Random Forest and computer vision
22	Explain Natural Language Processing (NLP)
23	Write a note on Neural Network

Unit 5

1	Explain Organize super step in detail.
2	Write a note on Association rule mining.
3	Explain Report super step in detail.
4	List and explain the various graphic visualization option available .
5	List and explain the various complex graphs .