IDS IMPORTANT QUESTIONS	
SNo	**UNIT-1**
1	Define Data science and Bigdata. How Data science and Bigdata are related.
2	Compare Structural and Unstructural Data.
3	Define Data Science.
4	List out uses of Data Science.
5	List the processing steps in data science.
6	What is Exploratory Data Analysis(EDL)?
7	Classify various facets of Data science.
8	Compare Natural Language and Machine Generated data.
1	Explain various data types / facets used in Data science and Big data.
2	Identify various steps for preprocessing the data and explain them. **(OR)** State and explain the steps Data Cleansing, Integrating and Transforming Data. (OR) Discuss in detail about the data preparation step of data science process.
3	Discuss the various processing steps / life cycle of in Data Science. **(OR)** Describe the concept of Data Science process.
4	Classify the term big data ecosystem. **(OR)** Discuss about the following components of big data technologies. i)Distributed File System ii)Distributed Programming Framework iii)Data Integration Framework iv)Machine Learning Framework
5	Discuss about NoSQL Databases. Explain different types of NoSQL Databases for handling big data.
6	Define Data Science. Discuss benefits and uses of data science.
7	Differentiate structured, semi-structured and unstructured data.
8	Discuss about the machine learning framework and tools used for handling big data.
	UNIT-2
1	Define Machine Learning. List out applications of Machine Learning in data science.
2	List some python tools used in Machine Learning.
3	What is feature engineering?
4	Outline the problems in handling big data.
5	Define Machine Learning.
6	Identify various tools of python used in ML.
7	List some of the basic data structures used in mitigating the problems while using big data.
8	Outline various programming tips to handle huge volumes of data.
1	Design a case study on predicting malicious URLs.
2	How to design machine learning model and explain various steps involved in it.
3	State and explain different types of Machine Learning with examples.
4	Identify various data structures helpful to handle large volumes of Data.
5	Discuss about various steps of ML modeling process.
6	Explain about different problems faced while handling big data.
7	Summarize the different applications of ML in Data Science.
8	Identify the various solutions for handling big data.
9	Compare Supervised, Unsupervised and Semi-Supervised Learning.
10	Explain programming tips to handle large data.
11	Discuss about various tools available in python for dealing with Data Science using Machine Learning.
12	List out various types of Machine Learning and explain any two of them using example programs.
	UNIT-3
1	List different components of Hadoop.
2	Define Brewer's Theorem.
3	List different components of spark ecosystem.
4	List the ACID properties of relational databases.
5	What is YARN? Explain its roles.

6	Define CAP Theorem.
7	What is spark?
8	Discuss the BASE properties of NoSQL databases.
1	List and Explain various components of Hadoop Framework. **(OR)** Describe how Map-Reduce can be used to achieve parallelism in Hadoop Framework.
2	Discuss in detail about ACID properties of relational databases.
3	Illustrate the CAP theorem with a neat sketch. **(OR)** Explain how the CAP Theorem affects the design of distributed systems and databases. **(OR)** Illustrate Brewer's theorem with neat sketch.
4	Explain different types of NoSQL databases.
5	Write a case study to classify the type of disease using data from NoSQL databases.
6	Explain the BASE properties of NoSQL databases.
7	Design a case study on Assessing risk while loaning money using Hadoop Framework.
8	List and explain the various components of Spark Framework.
	UNIT-4
1	What is the need for using graph databases?
2	Explain about NLTK.
3	Compare connected data and graph databases.
4	List the applications of Text Mining.
5	Define connected data. How to represent it?
6	Define NLTK. List two advantages of NLTK.
7	List the four basic structures in Neo4j.
8	Explain different techniques in text mining.
1	Discuss about Neo4j architecture.
2	Explain how Cypher is used as a query language for exploring graph databases.**(OR)** Describe the syntax and structure of a basic Cypher query for retrieving data from a graph database.
3	Design a case study focused on the classification of Reddit posts utilizing NLTK and SQLite.
4	Identify various applications of graph databases in real world scenarios.
5	Differentiate graph database and RDBMS. **(OR)** What is Cypher, and how does it differ from SQL in querying graph databases?
6	What are the key advantages of using Neo4j over traditional relational databases for graph-based data?
7	Discuss about different text mining techniques with an example.**(OR)** Discuss the terms stemming, lemmatization and Decission tree classifier.
8	Explain the architecture of Neo4j and how it supports graph database management.
9	List out the applications of graph databases.
	UNIT-5
1	Identify different data visualization options available for visualizing data.
2	Explain about Crossfilter library.
3	State the steps involved in creating an interactive dashboard using dc.js
4	List different dashboard development tools available for visualizing data.
5	What is data visualization?
6	Describe the usage of JavaScript Map Reduce library.
7	What are the steps required for creating an interactive dashboard using dc.js.
8	List different data visualization options available for visualizing data.
1	Discuss in detail how CrossFilter,MapReduce library helps in providing visualization of explored data.
2	Describe the procedure to create an interactive dashboard using dc.js.
3	Classify different dashboard development tools available for visualizing data.
4	State and explain various data visualization options. **(OR)** Explain different ways of communicate the insights from the data analysis to end user.
5	Explain the two main steps in the MapReduce process: "Map" and "Reduce," with examples.
6	Explain why most user interfaces were built using HTML5 rather than sophisticated tools and libraries available in market.
7	Apply the Data Science process for any real world problem.