



**SIDDHARTH INSTITUTE OF ENGINEERING & TECHNOLOGY:: PUTTUR
(AUTONOMOUS)**

Siddharth Nagar, Narayanavanam Road – 517583

QUESTION BANK(DESCRIPTIVE)

Subject with Code	BIG DATA ANALYTICS & AI APPLICATIONS-23CS0910	Course & Branch	B.Tech- CSM & CAI
Year & Sem	III B.Tech & II Sem	Regulation	R23

UNIT-1

INTRODUCTION TO BIG DATA AND ANALYTICS ECOSYSTEM

1.	a)	Define Big Data.	[L1] [CO1]	[2M]
	b)	List the characteristics of big data.	[L1] [CO1]	[2M]
	c)	Differentiate between structured and unstructured data.	[L1] [CO1]	[2M]
	d)	What is Big Data Analytics?	[L1] [CO1]	[2M]
	e)	State any two differences between Data Lake and Data Warehouse.	[L1] [CO1]	[2M]
2.	a)	Explain Big Data and its importance in modern organizations.	[L2] [CO1]	[5M]
	b)	Describe the 5 V's of Big Data with suitable examples.	[L2] [CO1]	[5M]
3.	a)	Illustrate the different types of data in Big Data with examples.	[L3] [CO1]	[5M]
	b)	What is Big Data Analytics? Explain its applications in real-world scenarios.	[L2] [CO1]	[5M]
4.		Describe the Hadoop ecosystem its major components with neat sketch.	[L2] [CO1]	[10M]
5.	a)	Discuss about Descriptive and Diagnostic Analytics with examples.	[L2] [CO1]	[5M]
	b)	Compare Predictive and Prescriptive Analytics with suitable use cases.	[L4] [CO1]	[5M]
6.	a)	Describe YARN architecture and its role in resource management.	[L2] [CO1]	[5M]
	b)	Explain the working of MapReduce with Map and Reduce functions.	[L2] [CO1]	[5M]
7.		Illustrate the concept of NoSQL databases and their need in Big Data.	[L3] [CO1]	[10M]
8.	a)	Describe different types of NoSQL databases with examples.	[L2] [CO1]	[5M]
	b)	What is HBase and Explain about how it supports Big Data storage.	[L2] [CO1]	[5M]
9.	a)	Illustrate about Column-oriented Database and Graph based Database.	[L3] [CO1]	[5M]
	b)	Discover the knowledge about major challenges of Big Data faced by organizations.	[L3] [CO1]	[5M]
10.	a)	Explain the role of Zookeeper and Oozie in Hadoop ecosystem.	[L2] [CO1]	[5M]
	b)	Discuss about Key value stores and Document oriented Database.	[L2] [CO1]	[5M]
11.		Compare Data Lake and Data Warehouse in detail.	[L4] [CO1]	[10M]

UNIT –II**Big Data Tools and Frameworks**

1.	a)	Define Apache Spark.	[L1] [CO2]	[2M]
	b)	What is an RDD in Apache Spark?	[L1] [CO2]	[2M]
	c)	List any two features of Apache Spark.	[L1] [CO2]	[2M]
	d)	What is in-memory computing in Spark?	[L1] [CO2]	[2M]
	e)	Define Data Frame in Apache Spark.	[L1] [CO2]	[2M]
2.		Explain Apache Spark architecture in detail, including all major components.	[L2] [CO2]	[10M]
3.	a)	Illustrate the different deployment modes of Spark on Hadoop.	[L2] [CO2]	[5M]
	b)	Explain Resilient Distributed Datasets (RDDs) and their characteristics.	[L3] [CO2]	[5M]
4.	a)	Explain Data Frames and list their features.	[L2] [CO2]	[5M]
	b)	Explain the role of SQL Context in Spark with an example.	[L2] [CO2]	[5M]
5.		Discover the knowledge about RDDs in detail. Explain their creation methods, properties, and advantages.	[L4] [CO2]	[10M]
6.	a)	Describe Spark SQL and its key features.	[L2] [CO2]	[5M]
	b)	Describe the working of Spark Streaming for real-time analytics.	[L2] [CO2]	[5M]
7.	a)	Describe the architecture and working of Apache Hive.	[L2] [CO2]	[5M]
	b)	Discuss about Apache Pig and the need for Pig Latin.	[L2] [CO2]	[5M]
8.		Explain Spark SQL, Data Frames, and Datasets with suitable examples and operations.	[L2] [CO2]	[10M]
9.		Compare Hadoop and Spark in detail with respect to performance, processing model, scalability, fault tolerance, and use cases.	[L4] [CO2]	[10M]
10.	a)	Describe about Kafka architecture and its core components.	[L2] [CO2]	[5M]
	b)	Illustrate about the advantages and use cases of Apache Kafka.	[L3] [CO2]	[5M]
11.	a)	Explain Impala architecture and its advantages over Hive.	[L2] [CO2]	[5M]
	b)	Compare Apache Pig and MapReduce.	[L4] [CO2]	[5M]

UNIT-III**Machine Learning on Big Data**

1	a)	Define MLlib. Mention any two features of MLlib.	[L1] [CO3]	[2M]
	b)	What is the role of data preprocessing in big data machine learning?	[L1] [CO3]	[2M]
	c)	What is dimensionality reduction? Mention one use case.	[L1] [CO3]	[2M]
	d)	Define accuracy and precision in model evaluation.	[L1] [CO3]	[2M]
	e)	What is a machine learning pipeline in Spark MLlib?	[L1] [CO3]	[2M]
2	a)	Explain the workflow of MLlib models with a neat diagram.	[L2] [CO3]	[5M]
	b)	Illustrate about the major algorithms supported by MLlib.	[L3] [CO3]	[5M]
3	a)	Explain Logistic Regression and its types with suitable examples.	[L2] [CO3]	[5M]
	b)	Differentiate between classification and regression with real-world examples.	[L4] [CO3]	[5M]
4	a)	Describe Principal Component Analysis (PCA) for dimensionality reduction.	[L2] [CO3]	[5M]
	b)	Explain cross-validation techniques used for model evaluation.	[L2] [CO3]	[5M]
5		Describe the complete data preprocessing pipeline for big data machine learning with examples.	[L2] [CO3]	[10M]
6		Discover the knowledge about supervised learning techniques for large datasets, focusing on classification and regression models.	[L4] [CO3]	[10M]
7	a)	Discuss about evaluation metrics used for regression models.	[L2] [CO3]	[5M]
	b)	Describe the working of supervised learning with an example.	[L2] [CO3]	[5M]
8		Discuss clustering and dimensionality reduction techniques used in unsupervised learning.	[L2] [CO3]	[10 M]
9	a)	Explain the concept of distributed training in machine learning.	[L2] [CO3]	[5M]
	b)	Compare synchronous and asynchronous training methods.	[L5] [CO3]	[5 M]
10	a)	Explain K-Means clustering and its applications.	[L2] [CO3]	[5 M]
	b)	Analyze about optimization techniques used in large-scale machine learning systems.	[L4] [CO3]	[5 M]

UNIT –IV
AI Applications on Big Data

1.	a)	Define Predictive Maintenance (PdM) using Big Data and AI.	[L1] [CO4]	[2M]
	b)	What is AIoT in smart cities?	[L1] [CO4]	[2M]
	c)	List two performance metrics used in evaluating real-time AI systems.	[L1] [CO4]	[2M]
	d)	What is Hyper-personalization in retail analytics?	[L1] [CO4]	[2M]
	e)	List any two benefits of AI-based fraud detection in banking.	[L1] [CO4]	[2M]
2.		Evaluate the impact of AI in transforming healthcare from reactive to proactive systems.	[L5] [CO4]	[10M]
3.	a)	Explain the working process of Predictive Maintenance using IoT and Big Data.	[L2] [CO4]	[5M]
	b)	Analyze anomaly detection and failure prediction techniques in PdM.	[L4] [CO4]	[5M]
4.	a)	Illustrate about the key challenges in implementing AI-based predictive maintenance.	[L3] [CO4]	[5M]
	b)	Explain how Machine Learning enables real-time fraud detection in banking.	[L2] [CO4]	[5M]
5.		Analyze the role of Big Data platforms in supporting AI applications across industries.	[L4] [CO4]	[10M]
6.	a)	Explain the role of IoT sensors and AI in smart waste management.	[L2] [CO4]	[5M]
	b)	Analyze intelligent transportation systems in smart cities using AI.	[L4] [CO4]	[5M]
7.	a)	Analyze the importance of data governance in AI-driven big data applications.	[L4] [CO4]	[5M]
	b)	Discuss common fraud types detected using AI in banking.	[L2] [CO4]	[5M]
8.		Describe the integration of AI and IoT (AIoT) in smart cities with suitable examples.	[L2] [CO4]	[10M]
9.	a)	Describe anomaly detection and failure prediction techniques in PdM.	[L2] [CO4]	[5M]
	b)	Illustrate about how AI supports genomics and personalized medicine.	[L3] [CO4]	[5M]
10.	a)	Explain predictive demand forecasting in retail using Big Data.	[L2] [CO4]	[5M]
	b)	Analyze scalability and reliability issues in real-time AI systems.	[L4] [CO4]	[5M]
11.		Compare Predictive Maintenance and Preventive Maintenance approaches with advantages and limitations.	[L4] [CO4]	[5M]

UNIT V**Advanced Topics and Case Studies**

1.	a)	What is TensorFlowOnSpark?	[L1] [CO5]	[2M]
	b)	Define Explainable AI (XAI).	[L1] [CO5]	[2M]
	c)	Define Edge Computing.	[L1] [CO5]	[2M]
	d)	State any two ethical issues in Big Data AI.	[L1] [CO6]	[2M]
	e)	What is Feature Importance in XAI?	[L1] [CO5]	[2M]
2.	a)	Explain about how TensorFlow integrates with Apache Spark for large-scale deep learning	[L2] [CO5]	[5M]
	b)	Describe the role of Spark Data Frames in preprocessing Big Data for deep learning.	[L2] [CO5]	[5M]
3.		Compare Cloud Computing and Edge Computing for AI applications in terms of latency, scalability, and security.	[L5] [CO5]	[10M]
4.	a)	Analyze the design principles of Edge Architecture.	[L4] [CO5]	[5M]
	b)	Describe the role of visualization techniques in interpreting deep learning models.	[L2] [CO5]	[5M]
5.	a)	Compare LIME and SHAP techniques.	[L5] [CO5]	[5M]
	b)	Explain bias and discrimination issues in AI models.	[L2] [CO5]	[5M]
6.	a)	Analyze about Feature Importance, Attribution, and Visualization approaches in XAI.	[L4] [CO5]	[5M]
	b)	Explain about how Edge Computing supports low-latency AI applications.	[L2] [CO6]	[5M]
7.		Illustrate the applications of Edge Computing in Healthcare, Manufacturing, and Autonomous Vehicles.	[L3] [CO6]	[10M]
8.		Analyze Ethical Issues in Big Data AI and propose suitable Data Governance strategies.	[L4] [CO6]	[10M]
9.	a)	Describe the role of Data Lifecycle Management in ensuring ethical AI usage.	[L2] [CO6]	[5M]
	b)	Discuss about the key components of Data Governance.	[L2] [CO6]	[5M]
10.	a)	Analyze the regulatory compliance requirements in data governance.	[L4] [CO6]	[5M]
	b)	Describe the role of AI in Smart Cities using Edge Architecture.	[L2] [CO6]	[5M]