Jaypee Institute of Information Technology

Declared Deemed to be University under Sec 3 of UGC Act 1956

Date & Time: 30/11/2023 & 01:27:40

Lecture Wise Breakup

Course Code.	21B12CS318	**Semester**	2023EVENSEM
Course Name	Big Data Ingestion		
Credits	3	**Contact Hours**	3-0-0
Faculty (Names)		**Faculty Coordinator**	Shikha Mehta,Bharat Gupta
		Teacher(s) Alphabetically	Bharat Gupta,Shikha Mehta

Course Outcomes		**Cognitive Levels**
C332-2.1	Explain the fundamental concepts of Big Data and Data Analytics.	Understand (Level 2)
C332-2.2	Understand the various formats of Big Data and their sources.	Understand (Level 2)
C332-2.3	Infer the need and challenges of Big Data Ingestion.	Understand (Level 2)
C332-2.4	Various types of storage for Big Data such as Hadoop,Distributed File Systems, NoSQL and NewSQL.	Apply (Level 3)
C332-2.5	Apply BDI tools as Sqoop and Flume to ingest data into a Big Data system.	Apply (Level 3)

Module No.	**Title of the Module**	**Topics in the Module**	**No. of Lectures for the module**
1	Introduction to BigData, Architectureand Patterns	Review of Big Data landscape, Big Data: Why and Where,Characteristics of Big Data (V's of Big Data) andDimensions of Scalability, Data Ingestion, Data Collection,Data processing, Data Storage Layer, Data Querying andData Visualization Layer, Concepts of Data Ingestion, DataStorage, Data Quality, Data Operations.	8
2	Big Data Sourcesand Formats	Structured vs. Semi-structured vs. Unstructured, Batch vs.Streams, Understanding Data Lakes, Exploring the Relational Data Model of CSV Files, Exploring the Semi-	6

Module No.	Title of the Module	Topics in the Module	No. of Lectures for the module
		structured Data Model of JSON data, Exploring the RC andORC File Formats, Exploring Streaming Sensor Data,Exploring Streaming Twitter Data.	
3	Big Data Sourcesand Formats	Need, Parameters, Challenges, Key Functions, Big DataIngestion Tools: Common Features, Objectives, Benefits,Examples.	3
4	Big Data StorageTechnologies	Big Data Technologies: Hadoop, NoSQL and NewSQL,Using Hadoop to Store Data (HDFS, HBASE), FromDBMS to BDMS, Redis: An Enhanced Key-Value Store,Semi-structured Data – AsterixDB, Solr: Managing Text,Relational Data – Vertica.	8
5	Using Sqoop forBig Data Ingestion	Sqoop Import, Import Data from MySql to HDFS, OtherVariations of Sqoop Import Command, Sqoop ExportCommand, Sqoop Jobs.	8
6	Using Flume forBig Data Ingestion	What is Flume, and where it is used, Difference betweenFlume and Sqoop, How Flume Works, What is FlumeAgent, What are the Components of Flume Agent, How7 Data Flows between Various Components of the Flume.	7
Total Lectures			40

Components	Maximum Marks
T1	20
T2	20
End Term	35
TA (Attendance: 10, Assignment/Quiz/Mini-Project: 15)	25
Total	100

Recommended Reading Material

Author(s), Title, Edition, Publisher, Year of Publication etc. (Text Books, Reference Books, Journals, Reports, Websites etc. in the IEEE format)

Text Book(s):

Dey, N., Hassanien, A. E., Bhatt, C., Ashour, A., & Satapathy, S. C. (Eds.). (2018). Internet of Things and Big Data Analytics Toward Next-Generation Intelligence (pp. 3-549). Berlin: Springer.