

Vikram University, Ujjain – Certificate Course in Big Data Analytics w.e.f.
2021-2022

Vikram University, Ujjain

Board of studies in Computer science (Faculty of Engineering Science)

SYLLABUS of

Certificate Course in Big Data Analytics

Exclusively for University Teaching Department (ICS, VUU)

**Three Month Certificate Course in Big Data Analytics Programme of
UTD (ICS, VUU)**

(Effective from Academic Session 2021-22)

Certificate Course in Big Data Analytics

Duration of the Course (in weeks) : **90 days**

Minimum Eligibility Criteria and pre-requisites, if any : 10+2 pass with knowledge of basics of Computers

COURSE STRUCTURE

S.No.	Course code	Title of Paper	Theory External Marks	Min. Pass Marks	Internal Mark	Min. Pass Marks	Max Marks
1	CBDA-101	Introduction to Data Analytics	75	27	25	9	100
2	CBDA-102	Big Data Analytics	75	27	25	9	100
3	CBDA-103	Internship/ Industrial Training/ Project Work	150 <u>300</u>	54	50 <u>100</u>	28	200 <u>400</u>

Minimum pass marks in each Paper (Theory) - 36%

Minimum pass marks in each Paper (Internal Assessment) - 36%

Minimum pass marks in Practical (Theory) - 36%

Detailed Syllabus

CBDA-101: Introduction to Data Analytics

UNIT 1

Descriptive Statistics: Introduction to the Course. Descriptive Statistics ,Probability Distribution. Inferential Statistics through Hypothesis test, Regression

UNIT 2

Machine Learning: Differentiate Algorithmic and model based framework. Regression :Ordinary Least Square, K- Nearest Neighbours Regression and classification.

UNIT 3

Supervised Learning with Regression and Classification techniques Bias-Variance Dichotomy Model Validation Approaches Logistic Regression Linear Discriminant Analysis Quadratic Discriminant Analysis Regression and Classification Trees Support Vector Machines

UNIT 4

Unsupervised Learning and Challenges for Big Data Analytics Clustering Associative Rule Mining Challenges for big data analytics

UNIT 5

Prescriptive analytics Creating data for analytics through designed experiments Creating data for analytics through Active Learning Creating data for analytics through Reinforcement learning

Text:

1. R. Panneerselvam, “Research Methodologies,” PHI.
2. C.R. Kothari: Research methodology, Methods and Techniques, New Age Publication.
3. S.N.Sivanandam ,S.N.Deepa, “Introduction to Neural Networks using MATLAB 6.0“, TATA MCGraw- Hill publications William Stallings,"Cryptography and Network security",Third Edition, Pearson Ed

CBDA-102: Big Data Analytics

UNIT I: Introduction to Big Data and Hadoop

Types of Digital Data, Introduction to Big Data, Big Data Analytics, History of Hadoop, Apache Hadoop, Analysing Data with Unix tools, Analysing Data with Hadoop, Hadoop Streaming, Hadoop Echo System, IBM Big Data Strategy, Introduction to Infosphere BigInsights and Big Sheets.

UNIT II: HDFS (Hadoop Distributed File System)

The Design of HDFS, HDFS Concepts, Command Line Interface, Hadoop file system interfaces, Data flow, Data Ingest with Flume and Scoop and Hadoop archives, Hadoop I/O: Compression, Serialization, Avro and File-Based Data structures.

UNIT III: Map Reduce

Anatomy of a Map Reduce Job Run, Failures, Job Scheduling, Shuffle and Sort, Task Execution, Map Reduce Types and Formats, Map Reduce Features.

Unit IV: Hadoop Eco

Pig: Introduction to PIG, Execution Modes of Pig, Comparison of Pig with Databases, Grunt, Pig Latin, User Defined Functions, Data Processing operators.

Hive: Hive Shell, Hive Services, Hive Metastore, Comparison with Traditional Databases, HiveQL, Tables, Querying Data and User Defined Functions.

HBase: HBasics, Concepts, Clients, Example, Hbase Versus RDBMS.

Big SQL: Introduction

UNIT V: Data Analytics with R

Machine Learning: Introduction, Supervised Learning, Unsupervised Learning, Collaborative Filtering. Big Data Analytics with BigR.

Text Books:

- Tom White “Hadoop: The Definitive Guide” Third Edit on, O’reily Media, 2012.
- Seema Acharya, Subhasini Chellappan, "Big Data Analytics" Wiley 2015.

References:

- Michael Berthold, David J. Hand, "Intelligent Data Analysis”, Springer, 2007.
- Jay Liebowitz, “Big Data and Business Analytics” Auerbach Publications, CRC press (2013)
- Tom Plunkett, Mark Hornick, “Using R to Unlock the Value of Big Data: Big Data Analytics with Oracle R Enterprise and Oracle R Connector for Hadoop”, McGraw-Hill/Osborne Media (2013), Oracle press.
- Anand Rajaraman and Jeffrey David Ulman, “Mining of Massive Datasets”, Cambridge University Press, 2012.
- Bill Franks, “Taming the Big Data Tidal Wave: Finding Opportunities in Huge Data Streams with Advanced Analytics”, John Wiley & sons, 2012.
- Glen J. Myat, “Making Sense of Data”, John Wiley & Sons, 2007.