



Shri Vaishnav Vidyapeeth Vishwavidyalaya

Shri Vaishnav Institute of Information Technology

B.Tech. (CSE- Cloud and Mobile Computing-IBM)

Choice Based Credit System (CBCS) 2020-21

SEMESTER VI

SUBJECT CODE	Category	SUBJECT NAME	TEACHING & EVALUATION SCHEME								
			THEORY			PRACTICAL		Th	T	P	CRED ITS
			END SEM University Exam	Two Term Exam	Teachers Assessment*	END SEM University Exam	Teachers Assessment*				
BTIBM604	DCC	Big Data Engineering - Spark & Scala	60	20	20	30	20	3	0	2	4

Legends: L - Lecture; T - Tutorial/Teacher Guided Student Activity; P – Practical; C - Credit;

***Teacher Assessment** shall be based on following components: Quiz/Assignment/ Project/Participation in Class, given that no component shall exceed more than 10 marks.

Course Educational Objectives (CEOs):

Students will acquire knowledge on:

1. Introduce students to Big Data Analysis
2. Introduce to Understanding Data Science and Notebooks.
3. Understating machine learning concept and Introduce JAQL, pig and HIVE
4. Managing the Big SQL, Data stream, Hortonworks Data Platform (HDP), Watson Studio

Course Outcomes (COs):

At the end of the mobility period, students will be able to:

1. Read data from persistent storage and load it into Apache Spark,
2. Manipulate data with Spark and Scala,
3. Express algorithms for data analysis in a functional style,
4. Recognize how to avoid shuffles and recomputation in Spark,
5. Analyze data with Watson Studio

Syllabus:

UNIT-I

10HRS

Introduction to Big Data and Analytics

Introduction to big data, Big data overview, IBM 4'v, Describing the functions and features of HDP, Developing and understanding of the complete open-source Hadoop ecosystem, Understanding the purpose of Apache Ambari in the HDP stack, Exploring some of the directory structure on the Linux system, Describing the nature of the Hadoop Distributed File System (HDFS), Explaining the function of the NameNode and DataNodes in an Hadoop cluster, Explaining how files are stored and blocks ("splits") are replicated, Filing access and basic commands with HDFS, Describing the MapReduce model v1, Listing the limitations of Hadoop 1 and MapReduce 1, Mapper class and Reducer class, Describing the YARN model, Comparing Hadoop 2/YARN with Hadoop 1, Run MapReduce and YARN jobs, Understanding the overall architecture of Ambari, Listing the functions of the main components of Ambari, Explaining how to start and stop services from Ambari Web Console, Managing Hadoop clusters with

Chairperson

Board of Studies
Shri Vaishnav Vidyapeeth
Vishwavidyalaya, Indore

Chairperson

Faculty of Studies
Shri Vaishnav Vidyapeeth
Vishwavidyalaya, Indore

Controller of Examination

Shri Vaishnav Vidyapeeth
Vishwavidyalaya, Indore

Joint Registrar

Shri Vaishnav Vidyapeeth
Vishwavidyalaya, Indore



Shri Vaishnav Vidyapeeth Vishwavidyalaya

Shri Vaishnav Institute of Information Technology

B.Tech. (CSE- Cloud and Mobile Computing-IBM)

Choice Based Credit System (CBCS) 2020-21

SEMESTER VI

SUBJECT CODE	Category	SUBJECT NAME	TEACHING & EVALUATION SCHEME								
			THEORY			PRACTICAL		Th	T	P	CRED ITS
			END SEM University Exam	Two Term Exam	Teachers Assessment*	END SEM University Exam	Teachers Assessment*				
BTIBM604	DCC	Big Data Engineering - Spark & Scala	60	20	20	30	20	3	0	2	4

Legends: L - Lecture; T - Tutorial/Teacher Guided Student Activity; P – Practical; C - Credit;

***Teacher Assessment** shall be based on following components: Quiz/Assignment/ Project/Participation in Class, given that no component shall exceed more than 10 marks.

Apache Ambari to Start the Apache Ambari web console, Sqoop Introduction, Sqoop commands processing, Zookeeper introduction

UNIT–II Apache Spark and Scala

9HRS

Working with Spark RDD with Scala, Listing and describing the architecture and components of the Spark unified stack, Understanding the principles of Spark programming, Listing and describing the Spark libraries, Launching and using Spark's Scala and Python shells, Understanding the nature and purpose of Apache Spark in the Hadoop ecosystem, Listing and describing the architecture and components of the Spark unified stack, Describing the role of a Resilient Distributed Dataset (RDD), Understanding the principles of Spark programming, Listing and describing the Spark libraries, Launching and using Spark's Scala and Python shells

UNIT–III Understanding Data Science and Notebooks

8HRS

Data Scientists overview, Recognizing the iterative nature of a data science project, Outlining the benefits of using Data Science Notebooks, Describing the mechanisms and tools used with Data Science Notebooks, Comparing and contrasting the major Notebooks used by Data Scientists, Data and notebooks in Jupyter, How notebooks help data scientists, Essential packages: NumPy, SciPy, Pandas, Scikit-learn, NLTK, BeautifulSoup, Data visualizations: matplotlib, PixieDust, Using Jupyter “Magic” commands, Start Jupyter - it will open in a web browser, Importing the lab, Exploring the component panels

UNIT–IV BigSQL

7HRS

Overview of Big SQL, Understanding how Big SQL fits in the Hadoop architecture, Start and stop Big SQL using Ambari and command line, Connecting to Big SQL using command line, Connecting to Big SQL using IBM Data Server Manager, Starting Hadoop components, Executing basic Big SQL statements, Describing and creating Big SQL schemas and tables, Describing and listing the Big SQL data types, Working with various Big SQL DDLs, Loading data into Big SQL tables using best practices, Creating and dropping simple Big SQL table, Creating sample tables, Moving data into HDFS, Loading data into Big SQL tables, Creating and working with views, Creating external tables, Describing Big SQL supported file formats,

Chairperson

Board of Studies
Shri Vaishnav Vidyapeeth
Vishwavidyalaya, Indore

Chairperson

Faculty of Studies
Shri Vaishnav Vidyapeeth
Vishwavidyalaya, Indore

Controller of Examination

Shri Vaishnav Vidyapeeth
Vishwavidyalaya, Indore

Joint Registrar

Shri Vaishnav Vidyapeeth
Vishwavidyalaya, Indore



Shri Vaishnav Vidyapeeth Vishwavidyalaya

Shri Vaishnav Institute of Information Technology

B.Tech. (CSE- Cloud and Mobile Computing-IBM)

Choice Based Credit System (CBCS) 2020-21

SEMESTER VI

SUBJECT CODE	Category	SUBJECT NAME	TEACHING & EVALUATION SCHEME								
			THEORY			PRACTICAL		Th	T	P	CRED ITS
			END SEM University Exam	Two Term Exam	Teachers Assessment*	END SEM University Exam	Teachers Assessment*				
BTIBM604	DCC	Big Data Engineering - Spark & Scala	60	20	20	30	20	3	0	2	4

Legends: L - Lecture; T - Tutorial/Teacher Guided Student Activity; P – Practical; C - Credit;

***Teacher Assessment** shall be based on following components: Quiz/Assignment/ Project/Participation in Class, given that no component shall exceed more than 10 marks.

Query Big SQL tables using various DMLs, Connecting to Big SQL, Query data with Big SQL, Working with the ARRAY type, Working with Big SQL functions, Storing data in an alternate file format (Parquet), Configuring the Big SQL Server, Configuring the Big SQL Scheduler, Backup and restore Big SQL, Configuring authentication for Big SQL, Managing security with Apache Ranger, Enabling SSL encryption, Configuring authorization of Big SQL objects, Configuring impersonation in Big SQL, Listing the supported data sources, Configuring Fluid Query with Big SQL

UNIT–V Watson Studio

7HRS

Setting up a project, Working with collaborators, Managing data assets, Sign up for a Watson Studio account, Creating a new project, Managing a project, Adding collaborators, Loading data Managing the object storage, Overview of Jupyter notebooks, Creating notebooks, Coding and running notebooks, Sharing and publishing notebooks, Creating a notebook, Using notebooks, Working with external data.

TEXT BOOKS:

1. DT Editorial Services, “Big Data Black Book”, 2016
2. Subhashini Chellappan & Seema Acharya, “Big Data Analytics”, 2019
3. Radha Shankarmani & M. Vijaylakshmi, ” Big Data Analytics”, 2016
4. Vibrant Publishers “Hadoop Big Data”, 2017

REFERENCES:

1. IBM Study Material
2. IBM Ebook

Chairperson

Board of Studies
Shri Vaishnav Vidyapeeth
Vishwavidyalaya, Indore

Chairperson

Faculty of Studies
Shri Vaishnav Vidyapeeth
Vishwavidyalaya, Indore

Controller of Examination

Shri Vaishnav Vidyapeeth
Vishwavidyalaya, Indore

Joint Registrar

Shri Vaishnav Vidyapeeth
Vishwavidyalaya, Indore



Shri Vaishnav Vidyapeeth Vishwavidyalaya

Shri Vaishnav Institute of Information Technology

B.Tech. (CSE- Cloud and Mobile Computing-IBM)

Choice Based Credit System (CBCS) 2020-21

SEMESTER VI

SUBJECT CODE	Category	SUBJECT NAME	TEACHING & EVALUATION SCHEME								
			THEORY			PRACTICAL		Th	T	P	CRED ITS
			END SEM University Exam	Two Term Exam	Teachers Assessment*	END SEM University Exam	Teachers Assessment*				
BTIBM604	DCC	Big Data Engineering - Spark & Scala	60	20	20	30	20	3	0	2	4

Legends: L - Lecture; T - Tutorial/Teacher Guided Student Activity; P – Practical; C - Credit;

***Teacher Assessment** shall be based on following components: Quiz/Assignment/ Project/Participation in Class, given that no component shall exceed more than 10 marks.

LIST OF EXPERIMENTS:

1. Steps to run vmware or cloudera
2. Processing of all hadoop commands on terminal
3. Working with hdfs storage and retrieval of data
4. Mapper class implementation using java
5. Reducer class implementation using java
6. Yarn implementation
7. Sqoop command processing
8. Spark implementation
9. Scala implementation
10. Ambari console and start and stop process
11. Jupyter notebook implementation
12. Working with ddl, dml, dql, tcl, dcl commands in big sql
13. Working with big sql views, array, functions like query
14. Creating Watson studio project
15. Working with Watson datasets, collaborators etc.

Chairperson

Board of Studies
Shri Vaishnav Vidyapeeth
Vishwavidyalaya, Indore

Chairperson

Faculty of Studies
Shri Vaishnav Vidyapeeth
Vishwavidyalaya, Indore

Controller of Examination

Shri Vaishnav Vidyapeeth
Vishwavidyalaya, Indore

Joint Registrar

Shri Vaishnav Vidyapeeth
Vishwavidyalaya, Indore