Shri Vaishnav Vidyapeeth Vishwavidyalaya, Indore

Shri Vaishnav Institute of Computer Applications

Name of Prgram: BCA (Big Data Analytics) in association with IBM

COURSE CODE	CATEGORY	COURSE NAME	L	T	P	CREDITS	TEACHING & EVALUATION SCHEME				
							THEORY			PRACTICAL	
							END SEM University Exam	Two Term Exam	Teachers Assessment*	END SEM University Exam	Teachers Assessment*
BCCAIBM401	Compulsory	Database Management System	2	0	2	3	60	20	20	30	20

Legends: L - Lecture; T - Tutorial/Teacher Guided Student Activity; P – Practical; C - Credit; Q/A – Quiz/Assignment/Attendance, MST - Mid Sem Test.

***Teacher Assessment** shall be based on following components: Quiz/Assignment/Project/Participation in class (Given that no component shall be exceed 10 Marks)

Course Educational Objectives (CEOs):

- To provide students with basic concepts in information system and the benefits of these systems
- To understand the role, requirement and operations that an analyst needed to analyze, design, and implement the systems
- To provide the knowledge of business data modeling for the designing of efficient information systems
- To explain the various issues related with Data Security.

Course Outcomes (COs): After completing this course the student will be able to:

- Understand the database systems concepts
- Design any Desktop application using an entity relationship diagrams (ERD) to express requirements and demonstrates skills to model data requirements and create data models.
- Apply the knowledge of the subject to any particular database implementation using SQL.
- To learn and understand various Database Architectures and Applications.
- To differentiate between data, information, and knowledge

UNIT-I

Introduction: Introduction to Databases, Purpose of Database System, Database system Vs file system, Database System concepts and architecture, Advantage of DBMS approach, various views of data, data independence, schema and subschema and instances, basic concepts of data models, Database languages, Database administrator and users, data dictionary.

Shri Vaishnav Vidyapeeth Vishwavidyalaya, Indore

Shri Vaishnav Institute of Computer Applications

Name of Prgram: BCA (Big Data Analytics) in association with IBM

COURSE CODE	CATEGORY	COURSE NAME	L	T	P	CREDITS	TEACHING & EVALUATION SCHEME				
							THEORY			PRACTICAL	
							END SEM University Exam	Two Term Exam	Teachers Assessment*	END SEM University Exam	Teachers Assessment*
BCCAIBM401	Compulsory	Database Management System	2	0	2	3	60	20	20	30	20

UNIT-II

ER model: basic concepts, notation for ER diagram, design issues, Entity types, mapping constraint, Concepts of keys: super, candidate, primary, alternate, foreign, weak and strong entity sets, inheritance.

Domains and Relations: domains, relations, kind of relations, relational database.

UNIT-III

Relational Algebra: The structure, relational operations.

Relational Calculus: idea of relational calculus, tuple and domain calculus, Domain relational Calculus, calculus vs algebra.

Normalization: Introduction, non-loss decomposition, FD diagram, Normal forms upto fifth normal form, Codd's rules.

UNIT-IV

SQL: Introduction, basic structure of SQL, Characteristics and Advantage of SQL set operations, aggregate functions, null values, SQL data types and literals. SQL operators, Types of SQL commands. Simple and nested queries, views, join relations. Comparison between tables and views, updatable views, indexes, clustering. Triggers in SQL.

UNIT-V

Transaction, concurrency and Recovery: basic concepts, ACID properties, Transaction states, Implementation of atomicity and durability, concurrency control, deadlock, storage structure types, stable storage implementation, data access, recovery and atomicity- log based recovery, deferred Database modification.

Shri Vaishnav Vidyapeeth Vishwavidyalaya, Indore

Shri Vaishnav Institute of Computer Applications

Name of Prgram: BCA (Big Data Analytics) in association with IBM

COURSE CODE	CATEGORY	COURSE NAME	L	T	P	CREDITS	TEACHING & EVALUATION SCHEME				
							THEORY			PRACTICAL	
							END SEM University Exam	Two Term Exam	Teachers Assessment*	END SEM University Exam	Teachers Assessment*
BCCAIBM401	Compulsory	Database Management System	2	0	2	3	60	20	20	30	20

Text Books:

1. A Silberschatz, H.F Korth, Sudersan, "Database System Concepts", 6th Edition, MGH Publication 2013.
2. C.J. Date, "An introduction to Database Systems", 6th Edition, Pearson 2003.
3. Elmasri&Navathe, "Fundamentals of Database systems",7th Edition, Pearson 2015.
4. B.C. Desai, "An introduction to Database systems", BPB.
5. RaghuRamakrishnan, "Database Management Systems",3rd Edition, TMH 2014.

List of Experiments:

1. To study Basic SQL commands (create database, create table, use, drop, insert) and execute the following queries using these commands:
   - Create a database named Employee".
   - Usethedatabase„Employee"andcreateatable„Emp"withattributes„ename",ecity","salary", "enumber", "eaddress", "depttname".
   - Create another table „Company" with attributes, cname", ccity", "empnumber" in the database „Employee".
2. To study the viewing commands (select, update) and execute the following queries using these commands:
   - Find the names of all employees who live in Delhi.
   - Increase the salary of all employees by Rs.5,000.
   - Find the company names where the number of employees is greater than10,000.
   - Change the Company City to Gurgaon where the Company name is TCS".
3. To study the commands to modify the structure of table (alter, delete) and execute the following queries using these commands.

Shri Vaishnav Vidyapeeth Vishwavidyalaya, Indore

Shri Vaishnav Institute of Computer Applications

Name of Prgram: BCA (Big Data Analytics) in association with IBM

COURSE CODE	CATEGORY	COURSE NAME	L	T	P	CREDITS	TEACHING & EVALUATION SCHEME				
							THEORY			PRACTICAL	
							END SEM University Exam	Two Term Exam	Teachers Assessment*	END SEM University Exam	Teachers Assessment*
BCCAIBM401	Compulsory	Database Management System	2	0	2	3	60	20	20	30	20

- Add an attribute named Designation to the table "Emp".
- Modify the table Emp, Change the data type of salary" attribute to float.
- Drop the attribute, deptname from the table "emp".
- Delete the entries from the table Company" where the number of employees are less than 500.

4. To use (and, or, in , not in, between , not between , like , not like) in compound conditions and execute the following queries using them:
   - Find the names of all employees who live in Gurgaon and whose salary is between Rs.20,000 and Rs.30,000.
   - Find the names of all employees whose names begin with either letter "A or B".
   - Find the company names where the company city is Delhi and the number of employees is not between 5000 and10,000.
   - Find the names of all companies that do not end with letter A.

5. Using aggregate functions execute the following queries:
   - Find the sum and average of salaries of all employees in computer science department.
   - Find the number of all employees who live in Delhi.
   - Find the maximum and the minimum salary in the HR department.

6. To execute the following queries using study the grouping commands (group by, orderby)
   - List all employee names in descending order.
   - Find number of employees in each department where number of employees is greater than5.List all the department names where average salary of a department isRs.10,000.

7. To write SQL queries to
   - Alter table, Emp and make enumber as the primary key.

- Alter table, Companyand add the foreign key constraint.

COURSE CODE	CATEGORY	COURSE NAME	L	T	P	CREDITS	TEACHING & EVALUATION SCHEME				
							THEORY			PRACTICAL	
							END SEM University Exam	Two Term Exam	Teachers Assessment*	END SEM University Exam	Teachers Assessment*
BCCAIBM401	Compulsory	Database Management System	2	0	2	3	60	20	20	30	20

- Add a check constraint in the table Emp such that salary has the value between 0 and Rs.1,00,000.
- Alter table Companyand add unique constraint to column cname.
- Add a default constraint to column ccity of table company with the value Delhi.
- Rename the name of database to ”Employee1“.
- Rename the name of table Emp to Emp1.
- Change the name of the attribute ename to empname.

10. To execute following queries using appropriate SQL statements to
    - Retrieve the complete record of an employee and its company from both the table using joins.
    - List all the employees working in the company TCS“.
11. To study the various set operations and execute the following queries using these commands:
    - List the enumber of all employees who live in Delhi and whose company is in Gurgaon or if both conditions are true.
    - List the enumber of all employees who live in Delhi but whose company is not in Gurgaon.
12. To study the commands for views and execute the following queries using these commands:
    - Create a view having ename and ecity.
    - In the above view change the ecity to Delhiwhere ename is “John“.
    - Create a view having attributes from both the tables.
    - Update the above view and increase the salary of all employees of IT department byRs.1000.

Shri Vaishnav Vidyapeeth Vishwavidyalaya, Indore

Shri Vaishnav Institute of Computer Applications

Name of Prgram: BCA (Big Data Analytics) in association with IBM

.

COURSE CODE	CATEGORY	COURSE NAME	L	T	P	CREDITS	TEACHING & EVALUATION SCHEME				
							THEORY			PRACTICAL	
							END SEM University Exam	Two Term Exam	Teachers Assessment*	END SEM University Exam	Teachers Assessment*
BCCAIBM402	Compulsory	Big Data Technologies(Hadoop)	3	0	2	4	60	20	20	30	20

Legends: L – Lecture; T – Tutorial/Teacher Guided Student Activity; P – Practical; Q/A – Quiz/Assignment/Attendance; MST – Mid Semester Test.

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

Course Education Objectives (CEOs): The objective of this course is to teach students about:

1. To familiarize the students with the need and scope of the subject.
2. Provide an exposure giving a strong foundation to the data analytics practices.
3. Big data computing technologies, Watson studio
4. Understand the challenges posed by distributed applications and how ZooKeeper is designed to Handle.
5. create a basis for the use of advanced investigative and computational methods to convert information to useful knowledge.
6. Develop an understanding of how business analytics is actually performed
7. Develop an understanding of the complete open-source Hadoop ecosystem and its near term future direction
8. Covers foundational techniques and tools required for data science and big data analytics like Hadoop, NoSQLMapReduce, BIGSQL, Watson studio.

Course Outcomes (COs): After the completion of the course the student will be able to:

1. understanding importance and scope of the subject
2. understand the technological foundations of the Big Data
3. differentiate the structured and unstructured data and different databases to store structured and unstructured data like Mongo DB, HBase etc.

4. store and manipulate the different types of data
5. have a knowledge of HADOOP andHadoop ecosystem and its uses in Big Data
6. understand and apply the Big Data Analytics
7. Understand the functions and features of HDP.

COURSE CODE	CATEGORY	COURSE NAME	L	T	P	CREDITS	TEACHING & EVALUATION SCHEME				
							THEORY			PRACTICAL	
							END SEM University Exam	Two Term Exam	Teachers Assessment*	END SEM University Exam	Teachers Assessment*
BCCAIBM402	Compulsory	Big Data Technologies(Hadoop)	3	0	2	4	60	20	20	30	20

8. To model and Design Big Data for analytics
9. Understand the MapReduce model v1 and review java code.
10. understand the professional and ethical responsibility
11. to produce the good decision makers who can use empirical approaches, wide range of data analytic techniques to problem solving.

PRE- REQUISITES:

This course requires the familiarity with linear algebra, calculus, matrix operations, probability theory, statistics, programming, Database Management System, Data Mining and Warehousing

Syllabus

Unit I

Big Data: Introduction and basics, Evolution of Data Management, Definition, Importance, Big Data Types, Structured and unstructured Data, Sources of bigstructured data and unstructured data, Architecture of Big Data Management System, Stages of Big Data Management,

Big Data Technology Foundations: Technology Components, virtualization, distributed computing, Cloud and Big Data, Integration of data types into a big data environment.

Unit II

Big Data and operational Databases: relational, non relational, key-value pair, document, column oriented, graph, spatial databases, MapReduce, Hadoop, Hadoop Foundation and Ecosystem, Appliances and Big Data

Warehouse, Big data Implementation, Big Data Applications. roles in building a complete big data, solution to common business problems.

Hadoop and HDFS: need for a big data strategy in terms of parallel reading of large data files and internode network speed in a cluster, Hadoop Distributed File System (HDFS), function of the NameNode and DataNodes in a Hadoop cluster, files are stored and blocks ("splits") are replicated. Hive, Sqoop.

COURSE CODE	CATEGORY	COURSE NAME	L	T	P	CREDITS	TEACHING & EVALUATION SCHEME				
							THEORY			PRACTICAL	
							END SEM University Exam	Two Term Exam	Teachers Assessment*	END SEM University Exam	Teachers Assessment*
BCCAIBM402	Compulsory	Big Data Technologies(Hadoop)	3	0	2	4	60	20	20	30	20

Unit III

Big Data Analytics: Introduction, Basic and Advanced Analytics, Drivers, Pillars of Analytics: descriptive, predictive and prescriptive. Core Components of analytical data architecture, Performance issues, Parallel vs. distributed processing, Shared nothing data architecture and Massive parallel processing, Elastic scalability, Data loading patterns.

Data Analytics lifecycle: Discovery, Data Preparation, Model Planning, Model Building, Communicating results and findings.

Unit IV

Apache Ambari: The purpose of Apache Ambari in the HDP stack, architecture of Ambari and Ambari' relation to other services and components of a Hadoop cluster, components of Ambari and their functions, initiating start and stop services from Ambari Web Console.

Introduction of Hortonworks Data Platform – HDP: functions and features of HDP, the IBM value-add components, basics of IBM Watson Studio, description of the purpose of the value-add components

Unit V

Shri Vaishnav Vidyapeeth Vishwavidyalaya, Indore

Shri Vaishnav Institute of Computer Applications

Name of Prgram: BCA (Big Data Analytics) in association with IBM

MapReduce and YARN: MapReduce model v1, the limitations of Hadoop 1 and MapReduce, handling the Mapper class with Java, Reducer class and the program driver needed to access MapReduce, the YARN model, Hadoop 2/YARN versus Hadoop 1

ZooKeeper, Slider, and Knox: The challenges with distributed applications and how ZooKeeper is designed to handle them, ZooKeeper: role within the Apache Hadoop infrastructure, the realm of Big Data management, real-world applications, services to manage distributed systems, ZooKeeper CLI and ZooKeeper services.

COURSE CODE	CATEGORY	COURSE NAME	L	T	P	CREDITS	TEACHING & EVALUATION SCHEME				
							THEORY			PRACTICAL	
							END SEM University Exam	Two Term Exam	Teachers Assessment*	END SEM University Exam	Teachers Assessment*
BCCAIBM402	Compulsory	Big Data Technologies(Hadoop)	3	0	2	4	60	20	20	30	20

TEXT BOOKS:

1. "Introduction to InfosphereBigInsights", IBM Career Education
2. "Changing Business with Data Insight", IBM Career Education
3. "Big Insights Analytics for Business Analysts", IBM Career Education
4. "Hadoop: The Definitive Guide Paperback – 2015" by Tom White, Shroff Publishers & Distributers Private Limited - Mumbai; Fourth edition (2015).
5. "Big Data and Hadoop" by V. K. Jain, Khanna Publishers; 1 edition (1 June 2015)
6. "Big Data for Dummies" by Judith Hurwitz, Alan Nugent, Dr. Fern Halper and Marcia Kaufman Wiley, ISBN: 978-1- 118-50422-2, 2013
7. "Data Analytics, Models, Algorithms for Intelligent Data Analysis by Thomas A., Springer Vieweg , ISBN 978-3-8348-2589-6, 2013
8. "HadoopFor Dummies", Dirk deRoos, For Dummies, 2014

REFERENCE BOOKS:

1. "Big Data: A Revolution That Will Transform How We Live, Work, and Think" by Kenneth Cukier, Viktor Mayer-Schönberger; Mariner Books; Edition (2014)
2. **"**Big Data: Using Smart Big Data, Analytics and Metrics to Make Better" by Bernard Marr; Wiley; Edition 1st (2015)

Chairperson	Chairperson	Controller of Examinations	Joint Registrar
Board of Studies	Faculty of Studies	Shri Vaishnav Vidyapeeth	Shri Vaishnav Vidyapeeth
Shri Vaishnav Vidyapeeth	Shri Vaishnav Vidyapeeth	Vishwavidyalaya, Indore	Vishwavidyalaya, Indore
Vishwavidyalaya, Indore	Vishwavidyalaya, Indore		

3. "Mining of Massive Datasets, Chapter 2", by Ullman, Rajaraman,
4. "MapReduce and Parallel DBMS" by Stonebrakeret al.,: Friends or Foes?, Communications of the ACM, January 2010.
5. "MapReduce: A Flexible Data Processing Tool" by Dean and Ghemawat, Communications of the ACM, January 2010.

COURSE CODE	CATEGORY	COURSE NAME	L	T	P	CREDITS	TEACHING & EVALUATION SCHEME				
							THEORY			PRACTICAL	
							END SEM University Exam	Two Term Exam	Teachers Assessment*	END SEM University Exam	Teachers Assessment*
BCCAIBM402	Compulsory	Big Data Technologies(Hadoop)	3	0	2	4	60	20	20	30	20

List of Practical:

1. Installation and configuring Hadoop, configuring HDFS,
2. Installation of Zookeeper, Pig, Sqoop and Hbase
3. Running jobs on Hadoop
4. Working on HDFS
6. Hadoop streaming
7. Creating Mapper function using python.
8. Creating Reducer function using python
9. Python iterator and generators
10. Twitter data sentimental analysis using Flume and Hive
11. Business insights of User usage records of data cards
12. Wiki page ranking with Hadoop
13. Health care Data Management using Apache Hadoop ecosystem