**MANAV RACHNA INTERNATIONAL INSTITUTE OF RESEARCH AND STUDIES**

(Deemed to be University under section 3 of the UGC Act 1956)

**NAAC 'A' Grade University**

**BCS-DS-730: BIG DATA ANALYTICS (CC, BA)**

Periods/week Credits Max. Marks : 200

L :3 T: 0 3.0 Continuous Evaluation : 100

Duration of Exam: 3 Hrs End Sem Examination : 100

**Pre-Requisite: Database Management System (BCS-DS-404)**

**Course Type: Program Electives**

**Course Outcomes:** Students will be able to-

BCS-DS-730.1. Explain and identify Big Data Fundamentals and its Business Implications.

BCS-DS-730.2. Apply the architectures and platforms introduced for Big data, in particular Hadoop and MapReduce.

BCS-DS-730.3. Access and process Data on Distributed File System.

BCS-DS-730.4. Develop Big Data Solutions using Hadoop Eco System.

BCS-DS-730.5. Understand Big Data Visualization Techniques.

BCS-DS-730.6. Understand and appreciate Analytics for Big data at Rest and in Motion.

**PART-A**

**Unit-1:Big Data Concepts**

1.1 What is Big Data, Volume, Velocity, Variety,

1.2 Why it’s Important, Risks of Big Data, Need of Big Data,

1.3 Structure of Big Data,Exploring Big Data,

1.4 Filtering Big Data, the Need for Standards,

1.5 Big Data and Analytics, Adoption Architecture, Benefits & Barriers,

1.6 Trends for Big Data Analytics

**Unit-2:Hadoop Fundamentals**

2.1 Hadoop Architecture, Hadoop File System (HDFS),

2.2 HDFS Administration , Map / Reduce concepts,

2.3 Setup of a Hadoop Cluster,

2.4 Managing Job Execution,

2.5 Move data into Hadoop using Flume, Data Loading,

2.6 Overview of workflow engine,

**Unit-3:Query languages for Hadoop Eco System**

3.1 Jaql Basics,

3.2 Jaql data types,

3.3 Input/output with Jaql,

3.4 Working with operators and expressions,

3.5 Use of Pig

3.6 Use of Hive

**PART-B**

**Unit-4:Hadoop Reporting and Analysis**

4.1 Approaches to Big Data reporting and analysis,

4.2 Big Data Access Technologies for Reporting and Analysis,

4.3 Business Intelligence and Hadoop Architecture,

4.4 Direct Batch Reporting on Hadoop,

4.5 Live Exploration of Big Data,

4.6 Indirect Batch Analysis on Hadoop

**Unit-5:Analytics for Big Data at Rest & in Motion-I**

5.1 Data Stream overview,

5.2 Streams Processing Language Basics,

5.3 Streams Processing Language Development,

5.4 SPL Programming Introduction,

5.5 Adapter Operators,

5.6 Relational and Utility Operators

**Unit-6:Analytics for Big Data at Rest & in Motion-II**

6.1 Windowing and Joins,

6.2 Punctuation, aggregation and Sorting,

6.3 Timing and Coordination,

6.4 Lists, Sets, and Maps,

6.5 Nodes and Partitions, Debugging,

6.6 Adapters and Toolkits

**Text Books / Reference Books:**

1. None specified, 2016, Big Data Analytics, I edition, IBM ICE Publication.
2. Seema Acharya, Subhasini Chellappan, 2015, Big Data Analytics, Wiley.
3. John M., 2015, Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data, I edition, EMC Education service.
4. Tom White, Hadoop: The Definitive Guide, Storage and Analysis at Internet Scale, 4th edition, O'Reilly Media.
5. 2016, Big data Black Book: DT Editorial Services, Dream tech Press.
6. Chris Eaton, Dirk, 2012, Understanding Big data, 1st edition, McGraw Hill.

**Software required/Weblinks:**

Ibm.com

www. searchbusinessanalytics.techtarget.com

www.mastersindatascience.org

**Instructions for paper setting:** Seven questions are to be set in total. First question will be conceptual covering entire syllabus and will be compulsory to attempt. Three questions will be set from each Part A and Part B (one from each unit) Student needs to attempt two questions out of three from each part. Each question will be of 20 marks.

**Distribution of Continuous Evaluation:**

|  |  |
| --- | --- |
| Sessional- I | 30% |
| Sessional- II | 30% |
| Assignment/Tutorial | 20% |
| Class Work/ Performance | 10% |
| Attendance | 10% |

**Evaluation Tools:**

Assignment/Tutorials

Sessional tests

Surprise questions during lectures/Class Performance

End Semester Examination

**COURSE ARTICULATION MATRIX :**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **CO Statement**  **(BCS-DS-730)** | PO  1 | PO  2 | PO  3 | PO  4 | PO  5 | PO  6 | PO  7 | PO  8 | PO  9 | PO  10 | PO  11 | PO  12 | PSO  1 | PSO  2 | PSO  3 |
| BCS-DS-730.1 | 3 | 2 | 2 | - | 3 | - | - | - | - | - | - | - | 1 | 2 | - |
| BCS-DS-730.2 | 1 | 2 | - | - | - | 2 | 3 | - | - | - | - | - | 1 | 2 | 3 |
| BCS-DS-730.3 | 2 | 3 | - | - | - | - | - | - | - | - | - | - | - | - | 3 |
| BCS-DS-730.4 | 1 | 2 | 3 | - | - | - | - | - | - | - | - | - | 1 | - | 2 |
| BCS-DS-730.5 | - | 2 | 3 | - | - | - | - | - | - | - | - | - | - | 2 | 3 |
| BCS-DS-730.6 | 1 | 2 | 3 | - | - | - | - | - | - | - | 2 | 3 | 1 | 2 | 3 |

**MANAV RACHNA INTERNATIONAL INSTITUTE OF RESEARCH AND STUDIES**

(Deemed to be University under section 3 of the UGC Act 1956)

**NAAC 'A' Grade University**

**BCS-DS-772: BIG DATA ANALYTICS LAB (CC, BA)**

Periods/week Credits Max. Marks : 100

P: 2 1.0 Continuous Evaluation : 50

Duration of Examination: 2 Hrs End Sem Examination : 50

**Co-Requisite: Big Data Analytics (BCS-DS-730)**

**Course Type: Program Electives**

**Course Outcomes:** Students will be able to-

BCS-DS-772.1: Understand the hadoop file systems.

BCS-DS-772.2: Import the data files, manipulate and transform the data.

BCS-DS-772.3: Use the market relevant tools, used in BDA.

BCS-DS-772.4: Apply various kinds of filters and extract the information.

BCS-DS-772.5: Use models to solve various problems of big data.

BCS-DS-772.6: Handle data in available in various formats.

**List of Experiments**:

1. Installation of IBM/Biginsight on VMWare and/or Cloudera on Virtual Box
2. Hadoop:Interaction with the hadoop file system using “hdfs dfs -put/mkdir/ls/put/get/cp/rm” etc.
3. Hadoop:Learnto work on Hbase Shell. Create the table, load, select etc.
4. Hbase: Apply different kind of filtersand sort operationsin the table created in Hbase shell.
5. Jaql:Create/downloada JSONFile, Read it in the Jaql, apply filter and transform. Write the output to a file.
6. Jaql:Create/downloadtwo JSON files, read both of them and perform join operations.
7. Pig:Load a log file into HDFS and perform work count operation on it using commands: Load, foreach, groupby, tokenize, flatten, generate and count.
8. Pig:Create two CSV files in PIG, load them into variables and perform filter and / or join operations.
9. Hive: Cretae / Downloada Database,Table and load data into it from a publically available dataset site, alternatively create your own file and load it.
10. Hive: Perform select operations using where, group by and orderby clause.
11. Hive: Create atleast 2different views on the already created table and perform select operations.

**Text Books/ Reference Books/ Internet References:**

* + - 1. Seema Acharya and Subhashini Chellapan, Big Data and Analytics by

1. <https://www.cloudera.com/documentation.html>
2. <https://www.ibm.com/support/knowledgecenter/en/SSPT3X_4.2.0/com.ibm.swg.im.infosphere.biginsights.welcome.doc/doc/welcome.html>

**Software required/Weblinks:**

VMWare, IBM BigInsights

VirtualBox, Cloudera

**Note:** At least 5 more exercises to be given by the teacher concerned.

**Distribution of Continuous Evaluation:**

|  |  |
| --- | --- |
| Viva- I | 30% |
| Viva- II | 30% |
| File/Records | 20% |
| Class Work/ Performance | 10% |
| Attendance | 10% |

**Evaluation Tools:**

Experiments in lab

File work/Class Performance

Viva (Question and answers in lab)

End Semester Practical Examination

**COURSE ARTICULATION MATRIX:**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **CO Statement**  **(BCS-DS-772)** | **PO**  **1** | **PO**  **2** | **PO**  **3** | **PO**  **4** | **PO**  **5** | **PO**  **6** | **PO**  **7** | **PO**  **8** | **PO**  **9** | **PO**  **10** | **PO**  **11** | **PO**  **12** | **PSO**  **1** | **PSO**  **2** | **PSO**  **3** |
| BCS-DS-772.1 | 1 | - | - | 2 | 3 | 2 | 2 | - | 1 | 1 | - | 1 | 3 | 1 | - |
| BCS-DS-772.2 | 1 | 3 | 3 | 3 | 3 | 3 | 2 | - | 1 | 1 | - | 1 | 3 | 1 | - |
| BCS-DS-772.3 | 1 | 3 | 3 | 3 | 3 | 3 | 2 | - | - | 2 | - | 1 | 3 | 1 | - |
| BCS-DS-772.4 | 1 | 3 | 3 | 3 | 3 | 3 | 2 | - | - | 2 | - | 1 | 3 | 1 | - |
| BCS-DS-772.5 | 1 | 3 | 3 | 3 | 3 | 3 | 2 | - | 1 | - | - | 1 | 3 | 1 | - |
| BCS-DS-772.6 | 1 | 3 | 3 | 3 | 3 | 3 | 2 | - | 1 | - | - | 1 | 3 | 1 | - |