****

**Format No. QSP/7.1/01.F01 (B) Issue No.05 Rev. No 5 Dated: Jan 1, 2017**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**UNIVERSITY OF PETROLEUM & ENERGY STUDIES**

**College of Engineering Studies**

**Dehradun**

**COURSE PLAN**

Programme : B. Tech – (CSE + Big Data)

Course : Big Data Overview

Course Code : CSBD1001

No. of credits : 3

Semester : I

Session : 2017-18 (July)

Batch : 2017-21

Prepared by : Mr. Ajay Rawat , Ms. Kalpana Rangra

Email : [arawat@ddn.upes.ac.in](mailto:arawat@ddn.upes.ac.in),

krangra@ddn.upes.ac.in

**Approved By**

Faculty HOD/ Prog. Head

UPES Campus Tel : +91-135-2770137

“Energy Acres” Fax : +91 135- 27760904

P.O. Bidholi, , Dehradun

**COURSE PLAN**

1. **PREREQUISITE:**
   1. Basic understanding of computers
2. **PROGRAM OUTCOMES (POs) and PROGRAM SPECIFIC OUTCOMES (PSOs) for Big Data:**

**B1. PROGRAM OUTCOMES (POs)**

PO1: **Engineering knowledge**: Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.

PO2: **Problem analysis**: Identify, formulate, review research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.

PO3: **Design/development of solutions**: Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations.

PO4: **Conduct investigations of complex problems**: Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.

PO 5 **Modern tool usage:** Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modeling to complex engineering activities with an understanding of the limitations.

PO6: **The engineer and society:** Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal and cultural issues and the consequent responsibilities relevant to the professional engineering practice.

PO7: **Environment and sustainability**: Understand the impact of the professional engineering solutions in societal and environmental contexts, demonstrate the knowledge of, and need for sustainable development.

PO8: **Ethics**: Apply ethical principles and commit to professional ethics, responsibilities, and norms of the engineering practice.

PO9: **Individual and Teamwork**: Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.

PO10: **Communication**: Communicate effectively on complex engineering activities with the engineering community and with society, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.

PO11: **Project management and** **finance**: Demonstrate knowledge and understanding of the engineering and management principles and apply these to one’s own work, as a member and leader in a team, to manage projects and in multidisciplinary environments.

PO12: **Life**-**long learning**: Recognize the need for, and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change.

**B2. PROGRAM SPECIFIC OUTCOMES for B.tech CSE-Big Data**

PO13: Perform system and application programming using computer system concepts, concepts of Data Structures, algorithm development, problem solving and optimizing techniques.

PO14: Apply software development and project management methodologies using concepts of front-end and back-end development and emerging technologies and platforms.

PO15: **Big Data**: Design solutions to challenging and ever growing real world data engineering problems and examine it to uncover hidden patterns, correlations, insights and make better data driven decisions.

1. **COURSE OUTCOMES FOR BIG DATA Overview:**

**At the end of this course student should be able to**

CO.1. Illustrate the basic concepts of data, Big Data.

CO.2. Infer how data is growing rapidly and the concept of big data ecosystem.

CO.3. Distinguish between structured data, semi-structured data, and unstructured data and different V’s in the Big data ecosystem.

CO.4. Compare files, tabular and relational databases, NoSQL data stores and distributed systems.

CO.5. Contrast the concept of data lake, Data Lake sources, data lake storage, and distinguish between vertical scaling and horizontal scaling.

**Table: Correlation of POs and PSOs v/s COs**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| PO/CO | PO1 | PO2 | PO3 | PO4 | PO5 | PO6 | PO7 | PO8 | PO9 | PO  10 | PO  11 | PO  12 | PSO  13 | PSO  14 | PSO  15 |
| CO1 | 2 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| CO2 | 2 | 2 |  |  |  |  |  |  |  |  |  |  | 2 | 2 | 2 |
| CO3 |  | 2 |  | 2 |  |  |  |  |  |  |  |  | 3 |  | 3 |
| CO4 | 2 |  | 2 |  |  |  |  |  |  |  | 2 |  |  | 2 | 3 |
| CO5 | 2 | 2 | 2 |  |  |  |  |  |  |  | 2 | 2 | 2 | 2 | 3 |

1: Slight (Low) 2: Moderate (Medium) 3: Substantial (High)

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | Engineering Knowledge | Problem analysis | Design/development of solutions | Conduct investigations of complex problems | Modern tool usage | The engineer and society | Environment and sustainability | Ethics | Individual or team work | Communication | Project management and finance | Life-long Learning | System and application programming | Software Project Management | Applying Big Data |
| Course Code | Course Title | PO1 | PO2 | PO3 | PO 4 | PO 5 | PO6 | PO 7 | PO8 | PO9 | PO 10 | PO 11 | PO12 | PSO13 | PSO14 | PSO15 |
| CSBD 1001 | Big Data Overview | 2 | 1 | 1 |  |  |  |  |  |  |  | 1 |  | 1 | 1 | 2 |

1: Slight (Low) 2: Moderate (Medium) 3: Substantial (High)

1. **PEDAGOGY**

* **Presentation**
* **White Board**

1. **COURSE COMPLETION PLAN**

|  |  |
| --- | --- |
| **Total Class room sessions** | 36 |
| **Total Quizzes** | 00 |
| **Total Test** | 02 |
| **Total Assignment** | 02 |

One Session =60 minutes

1. **EVALUATION & GRADING**

Students will be evaluated based on the following 3 stages.

* 1. Internal Assessment - 30%
  2. Mid-term Examination - 20%
  3. End term Examination - 50%

**F1. INTERNAL ASSESSMENT: WEIGHTAGE – 30%**

Internal Assessment shall be done based on the following:

|  |  |  |
| --- | --- | --- |
| Sl. No. | Description | % of Weightage out of 30% |
| 1 | Class Tests and Quizzes | 40% |
| 2 | Assignments (Problems/Presentations) | 30% |
| 3 | Attendance and performance in the class and presentation | 30% |

**F2*. Internal Assessment Record Sheet (including Mid Term Examination marks)*** *will be displayed online at the end of semester i.e. last week of regular classroom teaching.*

**F3. CLASS TESTS/QUIZZES:** Two Class Tests based on descriptive type theoretical & numerical questions and Two Quizzes based on objective type questions will be held; one class test and one quiz at least ten days before the Mid Term Examination and second class test and a second quiz at least ten days before the End Term Examination. Those who do not appear in Viva-Voce and quiz examinations shall lose their marks.

*The marks obtained by the students will be displayed on LMS a week before the start of Mid Term and End Term Examinations respectively.*

**F4. ASSIGNMENTS:** After completion of each unit or in the mid of the unit, there will be home assignments based on theory and numerical problems. Those who fail to submit the assignments by the due date shall lose their marks.

**F5. GENERAL DISCIPLINE:** Based on student’s regularity, punctuality, sincerity, and participation in the interactions.

*The marks obtained by the students will be displayed on ICOS at the end of the semester.*

**F6. MID TERM EXAMINATION: WEIGHTAGE – 20%**

Mid Term examination shall be Two Hours duration and shall be a combination ofShort and Long theory Questions.

***Date of showing Mid Term Examination Answer Sheets: Within a week after completion of the mid-Sem examination.***

**F7. END TERM EXAMINATION: WEIGHTAGE – 50%**

End Term Examination shall be Three Hours duration and shall be a combination of Short and Long theory/numerical Questions.

**F8. GRADING:**

The overall marks obtained at the end of the semester comprising all the above three mentioned shall be converted to a grade.

1. **DETAILED SESSION PLAN**

Note: The Online Sessions are highlighted in yellow color.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Module /Session** | **Big Ideas/ Topics** | **Course Outcomes Addressed** | **Required Learning Resources (including media)** | **Pedagogy/ Discussion(s)/ Postings** | **Assessment** |
|  | Introduction to course plan |  | Suggested Readings | Lecture |  |
|  | Data, information, units of data storage |  | Refer Blackboard, Suggested Reading | Readings/ brief video/ Presentation |  |
|  | Number system |  | Suggested Readings | Lecture |  |
|  | DBMS: introduction, significance, limitation |  | Suggested Readings | Lecture |  |
|  | Definition of IT, Data ware housing, Data mining |  | Refer Blackboard, Suggested Reading | Readings/ brief video/ Presentation |  |
|  | Big Data: definition, explosion, evolution, challenges |  | Suggested Readings | Lecture |  |
|  | Types of data, sources of Big Data, DBMS vs Big Data |  | Suggested Readings | Lecture |  |
|  | Characteristics of Big Data |  | Refer Blackboard, Suggested Reading | Readings/ brief video/ Presentation |  |
|  | Scalability |  | Suggested Readings | Lecture |  |
|  | Categories of Big Data, elements of Big Data, Roles in Big Data |  | Suggested Readings | Lecture | **Assignment1** |
|  | Big Data platform, Big Data ecosystem |  | Refer Blackboard, Suggested Reading | Readings/ brief video/ Presentation |  |
|  | Big Data pipeline |  | Suggested Readings | Lecture |  |
|  | Big Data analysis |  | Suggested Readings | Lecture |  |
|  | Big Data usecases |  | Refer Blackboard, Suggested Reading | Readings/ brief video/ Presentation |  |
|  | Distributed processing, centralized processing, traditional ways |  | Suggested Readings | Lecture |  |
|  | Block architecture, Block management |  | Suggested Readings | Lecture |  |
|  | Google file system, Distributed storage |  | Refer Blackboard, Suggested Reading | Readings/ brief video/ Presentation |  |
|  | Test 1 |  | Suggested Readings | Lecture |  |
|  | HDFS introduction |  | Suggested Readings | Lecture |  |
|  | Data storage: mechanism, flat file,tabular format, relational databases |  | Refer Blackboard, Suggested Reading | Readings/ brief video/ Presentation |  |
|  | HDFS architecture |  | Suggested Readings | Lecture |  |
|  | File formats: AVRO, sequence file, ORC, parquet |  | Suggested Readings | Lecture |  |
|  | File formats: text, JSON,CSV |  | Refer Blackboard, Suggested Reading | Readings/ brief video/ Presentation |  |
|  | Introduction to NOSQL |  | Suggested Readings | Lecture |  |
|  | NOSQL: types, examples |  | Suggested Readings | Lecture |  |
|  | NOSQL use case |  | Refer Blackboard, Suggested Reading | Readings/ brief video/ Presentation |  |
|  | Map Reduce |  | Suggested Readings | Lecture |  |
|  | Apache Spark |  | Suggested Readings | Lecture | **Assignment2** |
|  | Distributed computing |  | Refer Blackboard, Suggested Reading | Readings/ brief video/ Presentation |  |
|  | Introduction to Data lake |  | Suggested Readings | Lecture |  |
|  | Components of Data lake |  | Suggested Readings | Lecture |  |
|  | Sources of data and data stores in Data lake |  | Refer Blackboard, Suggested Reading | Readings/ brief video/ Presentation |  |
|  | Storing data in database and Data warehousing 1 |  | Suggested Readings | Lecture |  |
|  | Storing data in database and Data warehousing 2 |  | Suggested Readings | Lecture |  |
|  | Polyglot persistence, CAPs theorm |  | Refer Blackboard, Suggested Reading | Readings/ brief video/ Presentation |  |
|  | Test 2 |  | Suggested Readings | Lecture |  |

1. **SUGGESTED READINGS:**
2. Big Data Overview – Volume 1, Pearson and Xebia Press

**H1. REFERENCE BOOKS:**

1. Big Data Now: O'Reilly Media
2. The Big Data Market: O'Reilly Media
3. Big Data: A Revolution That Will Transform How We Live, Work, Think - Book by Kenneth Cukier and Viktor Mayer-Schönberger

**H2. OTHER RESOURCES**

**H3. VIDEO RESOURCES:**

1. <https://www.youtube.com/watch?v=3eQzwaxqjoA>
2. https://archive.bigdatauniversity.com/

**H4. WEB RESOURCES:**

1. <http://dataconomy.com/big-data-blogs/>
2. http://www.ibmbigdatahub.com/blogs

**GUIDELINES**

***Cell Phones and other Electronic Communication Devices*:** Cell phones and other electronic communication devices (such as Blackberries/Laptops) are not permitted in classes during Tests or the Mid/Final Examination. Such devices MUST be turned off in the class room.

***E-Mail and online learning tool:*** Each student in the class should have an email id and a pass word to access the LMS system regularly. Regularly, important information – Date of conducting class tests, guest lectures, via online learning tool. The best way to arrange meetings with us or ask specific questions is by email and prior appointment. All the assignments preferably should be uploaded on online learning tool. Various research papers/reference material will be mail/upload on online learning platform time to time.

***Attendance:*** Students are required to have a **minimum attendance of 75%** in each subject. Students with less than said percentage shall **NOT** be allowed to appear in the end semester examination.

**Course outcome assessment:** To assess the fulfillment of course outcomes two different approaches have been decided. The degree of the fulfillment of course outcomes will be assessed in different ways through direct assessment and indirect assessment. In Direct Assessment, it is measured through quizzes, tests, assignment, Mid-term and/or End-term examinations. It is suggested that each examination is designed in such a way that it can address one or two outcomes (depending upon the course completion). Indirect assessment is done through the student survey which needs to be designed by the faculty (sample format is given below) and it shall be conducted towards the end of course completion. The evaluation of the achievement of the Course Outcomes shall be done by analyzing the inputs received through Direct and Indirect Assessments and then corrective actions suggested for further improvement.

***Passing criterion:*** Student has to secure minimum 40% marks of the “highest marks in the class scored by a student in that subject (in that class/group class)” individually in both the ‘End-Semester examination’ and ‘Total Marks’ in order to pass in that paper.

* Passing Criterion for B. Tech: Minimum 40% of the highest marks in the class

**Sample format for Indirect Assessment of Course outcomes**

|  |
| --- |
| NAME: |
| ENROLLMENT NO: |
| SAP ID: |
| COURSE: |
| PROGRAM: |

Please rate the following aspects of course outcomes of Big Data Overview.

Use the scale 1-4\*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Sl. No. |  | 1 | 2 | 3 | 4 |
| 1 | CO.1. Illustrate the basic concepts of data, Big Data. |  |  |  |  |
| 2 | CO.2. Infer how data is growing rapidly and the concept of big data ecosystem. |  |  |  |  |
| 3 | CO.3. Distinguish between structured data, semi-structured data, and unstructured data and different V’s in the Big data ecosystem. |  |  |  |  |
| 4 | CO.4. Compare files, tabular and relational databases, NoSQL data stores and distributed systems. |  |  |  |  |
| 5 | CO.5. Contrast the concept of data lake, Data Lake sources, data lake storage, and distinguish between vertical scaling and horizontal scaling. |  |  |  |  |

3

Below Average

Good

1

**\***

Very Good

Average

4

2