### CSA4208 Bigdata Analytics
### Credits: 4C

| Learning Outcomes | Suggested Pedagogical Processes |
|---|---|
| Understand the concept and challenge of big data and why existing technology is inadequate to analyze the big data. | Explain the concepts and challenges of Big data. |
| Collect, manage, store, query and analyze various form of big data. | Discuss about various forms of big data. |
| Learn the concepts, architecture of Hadoop. | Explain the concepts, architecture of Hadoop using black board and powerpoint presentation. |
| Acquire the concepts, architecture of NoSQL. | Explain the concepts, architecture of NoSQL using black board and PowerPoint presentation. |
| Learn about the concepts and algorithms of Map Reduce. | Discuss the concepts and various algorithms of Map Reduce. |
| Get familiarity with Mining data streams. | Explain how to mine data streams with different data sets. |
| Understand the concepts of Link Analysis. | Explain the concepts of Link Analysis. |
| Attain the knowledge of Clustering. | Explain the concepts of Clustering. |

| Unit No. | Title of Unit and Contents |
|---|---|
| I | **Introduction to Big Data**<br>1.1 Introduction about distributed file system, Big Data and its importance<br>1.2 Four Vs, Drivers for Big data<br>1.3 Big data analytics<br>1.4 Big data applications<br>1.5 Algorithms using map reduce<br>1.6 Algorithms Matrix-Vector Multiplication by Map Reduce |
| II | **Introduction to Hadoop**<br>2.1 What is Hadoop?<br>2.2 Core Hadoop Components<br>2.3 Hadoop Ecosystem<br>2.4 Physical Architecture<br>2.5 Hadoop limitations |
| III | **NoSQL**<br>3.1 Introduction to NoSQL<br>3.2 NoSQL business drivers<br>3.3 NoSQL data architecture patterns: Key-value stores, Graph stores, Column family (Bigtable) stores, Document stores<br>3.4 Variations of NoSQL architectural patterns<br>3.5 Using NoSQL to manage big data: What is a big data NoSQL solution?, Understanding the types of big data problems, Analyzing big data with a shared-nothing architecture, Choosing distribution models: master-slave versus peer-to-peer, Four ways that NoSQL systems handle big data problems |
| IV | **Map Reduce and the New Software Stack** |

| | | |
|---|---|---|
| | 4.1 | Distributed File Systems: Physical Organization of Compute Nodes, Large Scale File-System Organization |
| | 4.2 | Map Reduce: The Map Tasks, Grouping by Key, The Reduce Tasks, Combiners, Details of Map Reduce Execution, Coping With Node Failures |
| | 4.3 | Algorithms using Map Reduce: Matrix-Vector Multiplication by Map Reduce, Relational-Algebra Operations, Computing Selections by Map Reduce, Computing Projections by Map Reduce, Union, Intersection, and Difference by Map Reduce, Computing Natural Join by Map Reduce, Grouping and Aggregation by Map Reduce, Matrix Multiplication, Matrix Multiplication with One Map Reduce Step |
| V | **Mining Data Streams** | |
| | 5.1 | The Stream Data Model: A Data-Stream- Management System, Examples of Stream Sources, Stream Query, Issues in Stream Processing |
| | 5.2 | Sampling Data in a Stream: Obtaining a Representative Sample, The General Sampling Problem, Varying the Sample Size |
| | 5.3 | Filtering Streams: The Bloom Filter, Analysis |
| VI | **Link Analysis** | |
| | 6.1 | Efficient computation of Page Rank: Page Rank Iteration Using Map Reduce, Use of Combiners to Consolidate the Result Vector |
| | 6.2 | Topic sensitive Page Rank, link Spam |
| VII | **Clustering** | |
| | 7.1 | Page Rank Definition, Structure of the Web, Dead Ends |
| | 7.2 | Using Page Rank in a Search Engine |
| | 7.3 | Hubs and Authorities |
| | 7.4 | CURE Algorithm |
| | 7.5 | Stream-Computing |
| | 7.6 | A Stream-Clustering Algorithm |
| | 7.7 | Initializing and Merging Buckets |
| | 7.8 | Answering Queries |

**Learning Resources**

1. Boris lublinsky, Kevin t. Smith, Alexey Yakubovich, "Professional Hadoop Solutions", ISBN: 9788126551071, Wiley, 2015
2. Chris Eaton, Dirk deroos et al. , "Understanding Big data ", McGraw Hill, 2012
3. Tom White, "HADOOP: The definitive Guide", O Reilly 2012
4. Vignesh Prajapati, "Big Data Analytics with R and Hadoop", Packet Publishing, 2013
5. Jy Liebowitz, "Big Data and Business analytics",CRC Press, 2013
6. Ullman, Rajaraman, Mining of Massive Datasets, 2014