Course Instructor: Dr Chiranjib Sur

Course Webpage: https://chiranjibsuruf.github.io/courses/da331.html

Email: chiranjib@iitg.ac.in

Objectives:

Prerequisites:

Course Code: DA331

Course Name: Big Data Analytics: Tools and Techniques

Credits: 2-0-2-6

Syllabus: Fundamentals of Big Data: Understanding big data, datasets, data analysis, data analytics, big data characteristics, types of data, case studies; Big data adoption and planning considerations: data procurement, big data analytics lifecycle, case study examples; Big data storage concepts: cluster computing, file system, distributed file systems, Relational & non-relational databases, scaling up & scaling out storage; No-SQL: Data types, Creating, Updating & Deleting documents, Querying, An example No-SQL database; Distributed computing framework: Introduction, file system, MapReduce programming model, examples of distributed computing environment framework; Stream data processing: tools such as Apache Spark, Apache Storm; Analytics with distributed computing framework: supervised learning examples, unsupervised learning examples.

Textbooks:

- Thomas Erl, Wajid Khattak and Paul Buhler, Bigdata fundamentals, concepts, drivers & techniques, 1st Edition, Pearson, 2016.

- Pramod J. Sadalage, Martin Fowler, Addison-Wesley, NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence, 1st Edition, Pearson, 2012.

- Sandy Ryza, Uri Laserson, Sean Owen and Josh wills, Advanced Analytics with Spark – patterns for learning from data at scale, 1st Edition, O'reilly, 2017.

References:

- Sean T. Allen, Matthew Jankowski, and Peter Pathirana, Storm Applied - Strategies for real-time event processing, 1st Edition, Manning Publications, 2015.

- Shannon Bradshaw, Eoin Brazil, Kristina Chodorow, MongoDB: The Definitive Guide, 3rd Edition, O'reilly, 2019.

- Bill Chambers, Matei Zaharia, Spark: The Definitive Guide, 1st Edition, O'Reilly, 2018.

- Tom White, Hadoop: The Definitive Guide, 4th Edition, Shroff/O'Reilly, 2015.

- Balamurugan Balusamy, Nandhini Abirami R, Seifedine Kadry, Amir H., Big Data: Concepts, Technology, and Architecture, 1st Edition, Pearson, 2015.