

BIG DATA ANALYTICS

COURSE CODE : ITA6008

LTPJC
30044

Dr.M.Balasubramani
Cabin No: PRP 315 D

Contact Hours

Tuesday :12:00 to 01:00 PM

Wednesday:03:00 to 04:00 PM

Objectives

1. To understand the big data platform and its use cases.
2. To impart knowledge in applying skills and tools to manage and analyze the big data.
3. To apply analytics on structured and unstructured data.

Expected Outcomes

- Demonstrate knowledge of the fundamental elements and concepts related to big data.
- Analyze the core architectural concepts to meet the challenges in implementing big data systems.
- Design and develop a Big Data Environment according to the benchmarks.
- Setup a Big Data Environment and implement security techniques.

Expected Outcomes

- Evaluate the use of data through cleansing, warehousing, analytics, and visualization to the ultimate business decision.
- Analyze the data using various statistical methods.
- Develop applications using large scale analytics tools to solve open big data problems.

Module - 01

Introduction to Big Data Analytics

- Big Data Overview, State of practice in analytics
- Role of Data Scientists
- Examples of Big Data Analytics
- Data Analytics Lifecycle

Module - 02

Introduction to Big Data Analytics

- Components of Hadoop
- Analyzing Big data with Hadoop, Design of HDFS,
- Developing a Map reduce Application.

Module – 03

Map Reduce

- Distributed File System(DFS)
- Map Reduce, Algorithms using Map Reduce
- Communication cost Model
- Graph Model for Map Reduce Problem.

Module – 04

Hadoop Environment

- Setting up a Hadoop Cluster, Hadoop Configuration
- Security in Hadoop, Administering Hadoop
- Hadoop Benchmarks
- Hadoop in the cloud.

Module – 05

Big Data Analytics methods using R

- Introduction to R-Attributes, R Graphical user interfaces
- Data import and export, attribute and Data Types.
- Descriptive Statistics, Exploratory Data Analysis.

Module – 06

Statistical methods for evaluation

- Hypothesis Testing
- Difference of Means, Wilcoxon Rank-Sum Test, Type I and Type II errors, power and sample size, ANOVA

Module – 07

Advanced Analytics - technologies and tools

- Analytics for unstructured data
- The Hadoop ecosystem – pig – Hive- HBase- Mahout- NoSQL

Module – 08

Contemporary issues

- Applications and currents trends of Big Data in Industry.

Text Books

- **Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data by EMC Education Services, 2015, publishing.**
- Anand Raja Raman and Jeffrey David Ullman, Mining of Massive Datasets, 2012, Cambridge University Press. [\[Reference\]](#)
- Tom White, Hadoop: The Definitive Guide, 3rd Edition, O'Reilly Media. [\[Reference\]](#)

Marks Split up

S.No	Mark Title	Max.Mark	Weightage (%)
1	Mid Term[21.7.23 to 23.7.23]	60	30
2	Digital Assignment [19.07.2023]	10	10
3	Quiz [14.07.2023 & 04.08.2023]	20	20
4	Final Assessment Test	100	40

Marks Split up & Dates for Jth Component Project Reviews

S.No	Mark Title	Max.Mark	Date
1	Review - I	20	5 th – 7 th July, 2023
2	Review - II	30	18 th – 20 th July, 2023
3	Review - III	50	2 nd – 4 th Aug, 2023

Jth component Google form link:

https://docs.google.com/forms/d/e/1FAIpQLSfcjmpM9l6PiA5N0rqXWVtZy4ZoP4Nc9IH7N5x73tnWDJASQA/viewform?usp=sf_link

- **Deadline – 02.07.2023 by 6:00 pm**



Recommended Technologies

- Open-source databases
- R (programming language)
- Cloud solutions (such as Azure and AWS)
- PHP and Javascript
- C++, Python
- SAS
- Tableau

Jth component

- ❑ Big Data for cybersecurity / Crime detection / Malicious user detection
- ❑ Health Status (Lung / Heart / Diabetes / Breast / Throat / Bone Cancer Prediction)
- ❑ Anomaly detection in cloud servers
- ❑ Agriculture yield Prediction
- ❑ Weather Forecasting / Rainfall Prediction
- ❑ Recommender system (Product / Tourist / Movie and Restaurant)
- ❑ Land / House Price Prediction
- ❑ Gross Domestic Product Prediction
- ❑ Election result Prediction
- ❑ Sales Prediction
- ❑ Sensex and stock Prediction
- ❑ Social Media – Tweets

UCI Machine Learning Repository

<https://archive.ics.uci.edu/ml/datasets.html>

Project Components

- Analysis on existing systems
- Requirement gathering



Review - 1

- Literature Review
- Problem identification
- Development



Review - 2

- Execution and Evaluation
- Submission of Report



Review - 3

Jth component (template)

- <https://drive.google.com/file/d/1M6AQeUi11jji-BAsELQ36BoO3VZzuMjc/view?usp=sharing>

Happy Learning ...

All the very best