



Faculty Development Programme
on

“BUILDING SCALABLE DATA
ENGINEERING AND MACHINE
LEARNING APPLICATIONS WITH
HADOOP AND SPARK”

11th to 15th June

ORGANIZING COMMITTEE

Ms.K.Padmaja, Assoc. Prof.

Mr.K.Siva Ramakrishna, Assoc.Prof.

Ms.M.Supriya, Asst.Prof

Ms.T.Krantika Asst.Prof

REGISTRATION LINK:

[HTTPS://FORMS.GLE/DBIJMCUSABCESHS76](https://forms.gle/DBIJMCUSABCESHS76)



IMPORTANT DATES

LAST DATE FOR SUBMISSION OF
APPLICATION

10TH JUNE ,2024

Coordinators

Dr.G.Bindu Madhavi
Associate Professor,
9848443421.

Dr.K.Arpitha
Associate professor,
9959894272.

WHO CAN ATTEND?

FACULTY/RESEARCH SCHOLARS

NO REGISTRATION FEE

REGISTRATION LINK:

[HTTPS://FORMS.GLE/DBIJMCUSABCESHS76](https://forms.gle/DBIJMCUSABCESHS76)

Seats are limited to 50 only. Participants are
selected on first come first serve basis.
Shortlisted candidates will be intimated through
mail. Attendance is mandatory to get the
certificate.

FACULTY DEVELOPMENT PROGRAMME
ON

“BUILDING SCALABLE DATA ENGINEERING AND
MACHINE LEARNING APPLICATIONS WITH
HADOOP AND SPARK”

11TH-15TH JUNE,2024

CHIEF PATRON
SRI. G.R.RAVINDER REDDY
Chairman, GCET

PATRON
DR. S.UDAYA KUMAR
Principal ,GCET

CONVENORS
DR.L.VENKATESWARLU
PROFESSOR & HOD-CSE(AIML)

DR.V.MADHUSUDAN RAO,
PROFESSOR & DEAN ,SCSI

ORGANIZED BY
DEPARTMENT OF COMPUTER SCIENCE AND
ENGINEERING (ARTIFICIAL INTELLIGENCE AND
MACHINE LEARNING)



Striving Towards Perfection..

Geethanjali College of Engineering and Technology
UGC Autonomous Institution,
Accredited with A+ grade by NAAC,
Affiliated to JNTUH,
Cheeryal (V), Keesara (M), Medchal,
Hyderabad, Telangana-501301,
India.
(geethanjaliinstitutions.com)

GEETANJALI COLLEGE OF ENGINEERING AND TECHNOLOGY was established in the year 2005. The campus epitomizes the Geethanjali motto, "Striving towards perfection" in providing the best of infrastructure and ambience. Geethanjali keeps a keen eye on the current trends and innovations happening in the industry and offers learning methods, which are designed to meet the evolving requirements of the industry. The college is recognized as a Scientific and Industrial Research Organization (SIRO) by the Directorate of Scientific and Industrial Research (DSIR) – New Delhi. It is NAAC accredited with an 'A+' grade. The college was conferred "Autonomous" status by UGC with effect from AY 2016-17. It has got ISO 9001:2008 certification. Currently, the college offers Undergraduate programs: B.Tech in CSE, ECE, EEE, ME, CE, CSE-AI&ML, CSE-CS, CSE-DS, and Post-Graduate programs: M.Tech in Computer Science and Engineering and MBA. All eligible programs B.Tech (CSE), B.Tech(ECE), B.Tech(EEE), B.Tech(ME), and B.Tech(CE) of GCET are accredited by NBA, New Delhi.

At Geethanjali, perfection is a passion.

About The Department

B.Tech "Computer Science and Engineering - Artificial Intelligence and Machine Learning" (CSE-AI&ML) was established with an intake of 60 in the academic year 2020-21. Subsequently, increased to 180 in the academic year 2021-22, 240 in the academic year 2023-24.

The Department is chaired by Dr. L. Venkateswarlu, he has a total experience of 26 years in Teaching and Research, an accomplished teacher and researcher demonstrating consistent success. He is also an able administrator and communicates effectively. He has good organizational and negotiation skills. Establishes reachable goals, meticulously plans to ensure participation by all and reaches consensus to attain desirable dreams.

Vision of the Department

To produce globally competent and socially responsible computer science engineers contributing to the advancement of engineering and technology which involves creativity and innovation by providing excellent learning environment with world class facilities.

Mission of the Department

1. To be a centre of excellence in instruction, innovation in research and scholarship, and service to the stake holders, the profession, and the public.
2. To prepare graduates to enter a rapidly changing field as a competent computer science engineer.
3. To prepare graduate capable in all phases of software development, possess a firm understanding of hardware technologies, have the strong mathematical background necessary for scientific computing, and be sufficiently well versed in general theory to allow growth within the discipline as it advances.
4. To prepare graduates to assume leadership roles by possessing good communication skills, the ability to work effectively as team members, and an appreciation for their social and ethical responsibility in a global setting

About FDP

In today's fast-paced digital era, the volume of data generated every minute has exploded dramatically, giving rise to what is known as big data. Characterized by its vast volume, high velocity, and diverse variety, big data presents both challenges and opportunities for professionals. Through advanced analytics, extensive datasets can be transformed via statistical and quantitative analysis into powerful insights that drive efficient decision-making. The Faculty Development Program (FDP) offers a comprehensive understanding of the terminologies and core concepts underlying big data challenges, applications, systems, and the techniques fundamental to big data analysis. Participants will be introduced to prominent frameworks such as Apache Spark, Hadoop, and MapReduce, as well as large-scale data storage technologies like HIVE and HDFS. Additionally, the program covers big data streaming and explores various applications of big data analysis using machine learning (ML), deep learning (DL), and graph processing. Recent research on transformers for text analysis will also be discussed. The FDP is designed for faculty members, research scholars, and practicing engineers/scientists from diverse fields who are keen on learning about the latest advancements and applications in big data technologies and methodologies effectively.

What to Expect

This course is designed for developers and engineers who have programming experience, but prior knowledge of Hadoop and/or Spark is not required. Apache Spark examples and hands-on exercises are presented in Scala and Python. The ability to program in one of those languages is required. Basic familiarity with the Linux command line is assumed. Basic knowledge of SQL is helpful. Programming Environment: Hadoop on Cloudera and Ambari clusters, Spark on local mode and Databricks.

Outcomes:

- Distribute, store, and process data in a Hadoop cluster
- Write, configure, and deploy Spark applications on a cluster
- Use the Spark shell for interactive data analysis
- Process and query structured data using Spark SQL and Hive Query Language
- Understand a wide variety of learning algorithms and Build an end-to-end Machine Learning Model with MLlib in PySpark.

RESOURCE PERSONS

1. Ms. Kratika Sharma,
Corporate Trainer,
Koenig Solutions Pvt. Ltd.

2. Dr. K. Arpitha,
Associate professor,
GCET

3. Mr. Shaik Akbar
Associate professor,
GCET

COURSE DUARATION: 5 DAYS

DAY 1

1. Introduction To Big Data Analytics
2. Introduction To Machine Learning

DAY 2

Module 1
Introduction to Apache Hadoop
• Apache Hadoop Overview
• Data Ingestion and Storage
• Data Processing
• Data Analysis and Exploration
• Other Ecosystem Tools
Apache Hadoop File Storage
• Apache Hadoop Cluster Components
• HDFS Architecture
• Using HDFS
Distributed Processing on an Apache Hadoop Cluster
• YARN Architecture
Introduction to the Hands-On Exercises
Module 2
Apache Spark Basics
• What is Apache Spark?
• Starting the Spark Shell
• Using the Spark Shell
• Getting Started with Datasets and DataFrames
• DataFrame Operations: RDD Overview • RDD Overview • RDD Data Sources • Creating and Saving RDDs • RDD Operations

DAY 3

Module 3
Working with DataFrames and Schemas
• Introduction to DataFrames
• Exercise: Introducing DataFrames
• Exercise: Reading and Writing DataFrames
• Exercise: Working with Columns
Analyzing Data with DataFrame Queries
• Querying DataFrames Using Column Exp.
• Grouping and Aggregation Queries
• Joining DataFrames
Module 4
Introduction to Apache Hive
• About Hive
• Transforming data with Hive QL
Working with Apache Hive
• Exercise: Working with Partitions
• Exercise: Working with Buckets
Module 5
Querying Tables and Views with Apache Spark SQL
• Querying Tables in Spark Using SQL
• Querying Files and Views
• Comparing Spark SQL, Apache Impala and Apache Hive-on-Spark

DAY 4

Module 6
Working with Datasets in Scala
• Datasets and DataFrames
• Creating Datasets
• Loading and Saving Datasets
• Dataset Operations
Module 7
Writing, Configuring, and Running Apache Spark Applications
• Writing a Spark Application
• Building and running an application
• Application Deployment Mode
• The Spark Application Web UI
• Configuring Application Properties
Module 8
Machine Learning with Spark ML
• Common Apache Spark Use Cases
• Iterative Algorithms in Apache Spark: Machine Learning, Graph Processing
• Introduction to MLlib: Various ML algorithms supported by MLlib
• ML model with Spark ML

DAY 5

Hands On spark