



JAYPEE INSTITUTE OF INFORMATION TECHNOLOGY

(Deemed University)

A-10, Sector 62, Noida, Gautam Buddh Nagar – 201 309 (U.P.)

AICTE Approved | UGC Approved | NAAC Accredited | NIRF Ranked

An Institution Dedicated to Excellence in Higher Education

Introduces Online Certificate Courses

Big Data Processing With Hadoop and Spark



Preamble

This course will introduce the characteristics of Big Data and its application in Big Data Analytics. The features, benefits, limitations and applications of Hadoop and Spark will be discussed. You will explore the components of Hadoop ecosystem, map reduce and learn how to use Hive, HBase and Pig to process Big Data. In this course, you will also learn how to leverage Apache Spark to deliver insights into Big Data. The course provides an overview of Resilient Distributed Datasets, Spark SQL and running Scala in Spark.

Duration of the Course: 3 months

Mode of Operation: Online

Number of Lectures of 1 hour duration: 20 L

Number of Practical Sessions of 1 – 2 hours duration: 24 hours session (12*2) + 2*2 hours evaluation

To be covered in each Lecture

| | |
|------|---|
| L-1 | : Big Data Characteristics, Scaling, Big Data Tools and Ecosystem, Big Data Use Cases |
| L-2 | : Apache Hadoop And The Hadoop Ecosystem |
| L-3 | : Map Reduce |
| L-4 | : Hive 1 |
| L-5 | : Hive 2 |
| L-6 | : Nosql DB HBase 1 |
| L-7 | : Nosql DB HBase 2 |
| L-8 | : Integrating Hive and HBase |
| L-9 | : Pig 1 |
| L-10 | : Pig 2 |
| L-11 | : Spark, Components, Hadoop Ecosystem Vs Spark |

| | |
|-----------------|---|
| L-12-14 | : Running Scala in Spark shell, Functional Programming, Programming With Scala |
| L-15,16 | : Resilient Distributed Datasets (RDD) In Spark, RDD Operations |
| L-17 -20 | : Spark SQL Introduction, Dataframes, Spark SQL Architecture, Data Formats, Dataframe Using SQL Query, RDD Vs Dataframes VS Datasets, Spark Mllib |

To be covered in each Practical Session

| | |
|----------------|------------------------|
| P-1 | : HDFS |
| P-2 | : Mapreduce |
| P-3 | : Hive |
| P-4 | : Hive |
| P-5 | : HBase |
| P-6 | : Pig |
| P-7 | : Scala in Spark Shell |
| P-8,9 | : Scala Programming |
| P-10-12 | : Spark SQL |

Pre-requisite, if any: Basic SQL

Schedule of the Classes: (Saturday/ Sunday – 2 hours on each day)

Nature of the Course: Broad based

Name of the Faculty Coordinators: Dr Shikha Mehta, Dr Parmeet Kaur

Name(s) of the Faculty to be involved in conduction of the Course: Dr Shikha Mehta, Dr Parmeet Kaur

Target Participants: Job seekers/ Current Senior students who wish to upgrade their professional skills

Minimum Qualifications for participants: Preferably Undergraduate in any specialization also pursuing Undergraduate degree will also be considered

IPR Rights of the Content: JIIT

Mode of evaluation of the participants after every 10 Lecture Sessions: Online Test

Mode of evaluation of the participants after 4 Practical Sessions: Online Test

For course related query please mail to:

Dr. Shikha Mehta: shikha.mehta@jiit.ac.in Dr.ParmeetKaur: parmeet.kaur@jiit.ac.in

For course registration, please click the link:

<https://forms.gle/AkVidsjMp5ArwNj7>