TECHGENX

Data Science course

Please Note: All below course content will be covered in practical scenarios and regular assignments will be shared. All sessions will be recorded and shared with student for future reference (free of cost). Along with below course.

Objective – Data Analytics Training Course New Delhi, Data Science Training Course New Delhi

The objective of Data Analytics and Data Science course is to master study of data science to become a successful data scientist. The course aims to equip the data scientist to successfully carry out data analysis to include tools for carrying out massive data management , statistical modelling and provide algorithm for data mining such as clustering and associate rule mining to name a few. The course primarily covers the complete range of Statistical Software Course & R-Programming Course and machine language learning techniques as defined in the Data Science study.

Scope of the program

After undertaking the course, one aims to achieve the proficiency in the following:

- Understand the basic role played by the Data scientist in analyzing the Data Analysis Life cycle.
- Analyze Big data by the use of Statistical Software and R statistically.
- Learn Predictive Analytics, Machine Learning & Data mining Techniques
- Insight in to various Machine Learning Techniques and their implementation using R.
- Handling tools and techniques involved in sampling, filtering and data transformation

Who Should Enroll

The course is a blend of two major open source tools available viz.Statistical Software and R language. The course is ideal for you if you are:

- A Professional working on Database management and streaming of Big Data.
- IT or Management student who are passionate about problem solving methodologies.
- Professionals who are expert in their domain and strive to learn technology for business and technology integration.

Section : 1 Analytics with Statistical Software Programming

1. Introduction to Analytics and Basic Statistics
2. Hypothesis Testing
3. T-Test
4. ANOVA
5. Chi-Square
6. Linear Regression
7. Logistic Regression
8. Cluster Analysis

9. Time Series

Section :2 Python Programming

1. Introduction to Python
2. Data Structure
3. Getting Data from Differences Sources (Import/Export)
4. Functions
5. Data Transformation
6. Data Formatting
7. Restructuring Data
8. Looping Structure
9. Graphics

Section 3 : Analytics with Python Programming

1. Basic Statistics
2. Hypothesis Testing
3. T Test, Z Test, F Test
4. Correlation
5. ANOVA, MANOVA, ANCOVA
6. Linear regression
7. Logistic regression
8. Cluster Analysis
9. Time Series
10. Decision Tree
11. Random Forest
12. Text Analytics
13. Sentiment Analysis
14. Social Media Analytics

Section 4: Advance Excel

1. Functions
2. Data Analysis
3. Pivot Tables
4. Interactive charts
5. What-if-analysis, GOAL SEEK
6. Linking Multiple Sheets
7. Sub Total Reports, Auto Filter
8. Password Protecting Worksheets
9. Dashboard
10. Record a Macro/VBA
11. Editing a VBA Code

Big Data Analytics with Hadoop

As the need for big data analytics gains prominence world over; there is a subsequent growth in demand for Hadoop skill to process big data.

Program Objective

This course aims to take you through all the Big Data and Hadoop analytics concepts through step by step, well structured modules. It is our objective at Analytic square that by the end of this program you should be able to –

- Have a basic understanding of Hadoop Distributed File System as well as MapReduce framework
- Create a Hadoop cluster
- Work with Sqoop and Flume on Data Loading Techniques
- Learn to program in YARN, MapReduce and even write them
- Do data analytics
- Work on your own Big Data Analytics project implementing Hadoop

Who Should Join

If you are interested in big data and want to become a proficient Hadoop Developer this course is just right for you. You can benefit from this course if you are a –

- Software professionals
- ETL developers
- Project Managers
- Analytics Professionals
- Testing experts
- Students with knowledge of Core Java

Section 1: Introduction

- Lecture 1: **What is Big Data**
- Lecture 2: **What is Hadoop**
- Lecture 3: **Distributed System and Hadoop**
- Lecture 4: **RDBMS and Hadoop**

Section 2: Starting Hadoop

- Lecture 5: **Single node Hadoop Cluster**
- Lecture 6: **Configuring Hadoop**
- Lecture 7: **Hadoop Architecture**
- Lecture 8: **Hadoop Components**
- Lecture 9: **Name and Data Nodes**

- Lecture 10: **Command Line Interface**
- Lecture 11: **Running Hadoop**
- Lecture 12: **Web-based cluster UI-Name Node UI, Map Reduce UI**
- Lecture 13: **Hands-On Exercise: Using HDFS commands**
- **Section Quiz**

Section 3: UNDERSTANDING MAPREDUCE

- Lecture 14: **How Map Reduce Works**
- Lecture 15: **Data flow in Map Reduce**
- Lecture 16: **Map operation**
- Lecture 17: **Reduce operation**
- Lecture 18: **Map Reduce Program In JAVA using Eclipse**
- Lecture 19: **Counting words with Hadoop—Running your first program**
- Lecture 20: **Writing Map Reduce Drivers, Mappers and Reducers in Java**
- Lecture 21: **Real-world "Map Reduce" problems**
- Lecture 22: **Hands-On Exercise: Writing a Map Reduce Program and Running a Map Reduce Job**
- Lecture 23: **Java Word Count Code Walkthrough**
- **Section Quiz**

Section 4: Hadoop Ecosystem

- Lecture 24: **Hive**
- Lecture 25: **Sqoop**
- Lecture 26: **Pig**
- Lecture 27: **Hbase**
- **Section Quiz**

Section 5: Hive

- Lecture 28: **Installation of Hive**
- Lecture 29: **Introduction to Apache Hive**
- Lecture 30: **Getting data into Hive**
- Lecture 31: **Hive's Architecture**
- Lecture 32: **Hive-HQL**
- Lecture 33: **Query Execution**
- **Section Quiz**

Section 6: Sqoop

- Lecture 34: **Installing and Configure Sqoop**
- Lecture 35: **Import RDBMS data to Hive using Sqoop**
- Lecture 36: **Export from to Hive to RDBMS using Sqoop**

Section 7: Pig

- Lecture 37: **Introduction and Installation of Pig**
- Lecture 38: **Pig Architecture**
- Lecture 39: **Pig Latin – Reading and writing data using Pig**

Section 8: HBase

- Lecture 40: **Installation**
- Lecture 41: **Architecture of Hbase**
- Lecture 42: **Managing large data sets with HBase**
- **Final Quiz**