

FUTURE GEN TECHNOLOGIES

DATA SCIENCE SYLLABUS

Introduction to Data Science

- Need for Data Scientists
- Foundation of Data Science
- What is Business Intelligence
- What is Data Analysis, Data Mining, and Machine Learning
- Analytics vs Data Science
- Value Chain
- Types of Analytics
- Lifecycle Probability
- Analytics Project Lifecycle

Data

- Basis of Data Categorization
- Types of Data
- Data Collection Types
- Forms of Data and Sources
- Data Quality, Changes and Data Quality Issues, Quality Story
- What is Data Architecture
- Components of Data Architecture
- OLTP vs OLAP
- How is Data Stored?

Big Data

- What is Big Data?
- 5 Vs of Big Data

- **Big Data Architecture, Technologies, Challenge and Big Data Requirements**
- **Big Data Distributed Computing and Complexity**
- **Hadoop**
- **Map Reduce Framework**
- **Hadoop Ecosystem**

Data Science Deep Dive

- **What is Data Science?**
- **Why are Data Scientists in demand?**
- **What is a Data Product**
- **The growing need for Data Science**
- **Large-Scale Analysis Cost vs Storage**
- **Data Science Skills**
- **Data Science Use Cases and Data Science Project Life Cycle & Stages**
- **Map-Reduce Framework**
- **Hadoop Ecosystem**
- **Data Acquisition**
- **Where to source data**
- **Techniques**
- **Evaluating input data**
- **Data formats, Quantity and Data Quality**
- **Resolution Techniques**
- **Data Transformation**
- **File Format Conversions**
- **Anonymization**

Intro to R Programming

- **Introduction to R**
- **Business Analytics**
- **Analytics concepts**

- **The importance of R in analytics**
- **R Language community and eco-system**
- **Usage of R in industry**
- **Installing R and other packages**
- **Perform basic R operations using command line**
- **Usage of IDE R Studio and various GUI**

R Programming Concepts

- **The datatypes in R and its uses**
- **Built-in functions in R**
- **Subsetting methods**
- **Summarize data using functions**
- **Use of functions like head(), tail(), for inspecting data**
- **Use-cases for problem solving using R**

Data Manipulation in R

- **Various phases of Data Cleaning**
- **Functions used in Inspection**
- **Data Cleaning Techniques**
- **Uses of functions involved**
- **Use-cases for Data Cleaning using R**

Data Import Techniques in R

- **Import data from spreadsheets and text files into R**
- **Importing data from statistical formats**
- **Packages installation for database import**
- **Connecting to RDBMS from R using ODBC and basic SQL queries in R**
- **Web Scraping**
- **Other concepts on Data Import Techniques**

Exploratory Data Analysis (EDA) using R

- What is EDA?
- Why do we need EDA?
- Goals of EDA
- Types of EDA
- Implementing of EDA
- Boxplots, cor() in R
- EDA functions
- Multiple packages in R for data analysis
- Some fancy plots
- Use-cases for EDA using R

Data Visualization in R

- Storytelling with Data
- Principle tenets
- Elements of Data Visualization
- Infographics vs Data Visualization
- Data Visualization & Graphical functions in R
- Plotting Graphs
- Customizing Graphical Parameters to improvise the plots
- Various GUIs
- Spatial Analysis
- Other Visualization concepts

Big Data and Hadoop Introduction

- What is Big Data and Hadoop?
- Challenges of Big Data
- Traditional approach Vs Hadoop
- Hadoop Architecture
- Distributed Model

- **Block structure File System**
- **Technologies supporting Big Data**
- **Replication**
- **Fault Tolerance**
- **Why Hadoop?**
- **Hadoop Eco-System**
- **Use cases of Hadoop**
- **Fundamental Design Principles of Hadoop**
- **Comparison of Hadoop Vs RDBMS**

Understand Hadoop Cluster Architecture

- **Hadoop Cluster and Architecture**
- **5 Daemons**
- **Hands-On Exercise**
- **Typical Workflow**
- **Hands-On Exercise**

Map Reduce Concepts

- **Map Reduce Concepts**
- **What is Map Reduce?**
- **Why Map Reduce?**
- **Map Reduce in real world and Map Reduce Flow**
- **What is Mapper, Reducer, and Shuffling?**
- **Word Count Problem**

Spark

Apache Spark

- **Introduction to Apache Spark**

- **Why Spark**
- **Batch Vs. Real-Time Big Data Analytics**
- **Batch Analytics – Hadoop Ecosystem Overview**
- **Real-Time Analytics Options**
- **Streaming Data – Storm**
- **In Memory Data – Spark, What is Spark?**
- **Spark benefits to Professionals**
- **Limitations of MR in Hadoop**
- **Components of Spark**
- **Spark Execution Architecture**
- **Benefits of Apache Spark**
- **Hadoop vs Spark**

Introduction to Scala

- **Features of Scala**
- **Basic Data Types of Scala**
- **Val vs Var**
- **Type Inference**
- **REPL**
- **Objects & Classes in Scala**
- **Functions as Objects in Scala**
- **Anonymous Functions in Scala**
- **Higher Order Functions**
- **Lists in Scala**
- **Maps**
- **Pattern Matching**
- **Traits in Scala**
- **Collections in Scala**

Spark Core Architecture

- **Spark & Distributed Systems**
- **Spark for Scalable Systems**
- **Spark Execution Context**
- **What is RDD**
- **RDD Deep Dive and Dependencies**
- **RDD Lineage**
- **Spark Application In Depth and Spark Deployment**
- **Parallelism in Spark**
- **Caching in Spark**

Spark Internals

- **Spark Transformations, Actions, Cluster and SQL Introduction**
- **Spark Data Frames**
- **Spark SQL with CSV, JSON, and Database**

Spark Streaming

- **Features of Spark Streaming**
- **Micro Batch**
- **Dstreams**
- **Transformations on Dstreams**
- **Spark Streaming Use Case**

Statistics + Machine Learning

Statistics

What is Statistics?

- **Descriptive Statistics**
- **Central Tendency Measures**

- **The Story of Average**
- **Dispersion Measures**
- **Data Distributions**
- **Central Limit Theorem**
- **What is Sampling**
- **Why Sampling**
- **Sampling Methods**
- **Inferential Statistics**
- **What is Hypothesis testing**
- **Confidence Level**
- **Degrees of freedom**
- **what is pValue**
- **Chi-Square test**
- **What is ANOVA**
- **Correlation vs Regression**
- **Uses of Correlation and Regression**

Machine Learning

Machine Learning Introduction

- **ML Fundamentals**
- **ML Common Use Cases**
- **Understanding Supervised and Unsupervised Learning Techniques**
- **Clustering**
- **Similarity Metrics**
- **Distance Measure Types: Euclidean, Cosine Measures**
- **Creating predictive models**
- **Understanding K-Means Clustering**
- **Understanding TF-IDF, Cosine Similarity and their application to Vector Space Model**
- **Case study**

- **Implementing Association rule mining**
- **Case study**
- **Understanding Process flow of Supervised Learning Techniques**
- **Decision Tree Classifier**
- **How to build Decision trees**
- **Case study**
- **Random Forest Classifier**
- **What is Random Forests**
- **Features of Random Forest**
- **Out of Box Error Estimate and Variable Importance**
- **Case study**
- **Naive Bayes Classifier**
- **Case study**
- **Project Discussion**
- **Problem Statement and Analysis**
- **Various approaches to solving a Data Science Problem**
- **Pros and Cons of different approaches and algorithms**
- **Linear Regression**
- **Case study**
- **Logistic Regression**
- **Case study**
- **Text Mining**
- **Case study**
- **Sentimental Analysis**
- **Case study**



Getting Started with Python

- **Python Overview**
- **About Interpreted Languages**

- **Advantages/Disadvantages of Python pydoc**
- **Starting Python**
- **Interpreter PATH**
- **Using the Interpreter**
- **Running a Python Script**
- **Python Scripts on UNIX/Windows, Editors and IDEs**
- **Using Variables**
- **Keywords**
- **Built-in Functions**
- **Strings Different Literals**
- **Math Operators and Expressions**
- **Writing to the Screen**
- **String Formatting**
- **Command Line Parameters and Flow Control**

Sequences and File Operations

- **Lists**
- **Tuples**
- **Indexing and Slicing**
- **Iterating through a Sequence**
- **Functions for all Sequences**
- **Using Enumerate()**
- **Operators and Keywords for Sequences**
- **The xrange() function**
- **List Comprehensions**
- **Generator Expressions**
- **Dictionaries and Sets**

Deep Dive – Functions Sorting Errors and Exception Handling

- **Functions**

- **Function Parameters**
- **Global Variables**
- **Variable Scope and Returning Values.**
- **Sorting**
- **Alternate Keys**
- **Lambda Functions**
- **Sorting Collections of Collections, Dictionaries and Lists in Place**
- **Errors and Exception Handling**
- **Handling Multiple Exceptions**
- **The Standard Exception Hierarchy**
- **Using Modules**
- **The Import Statement**
- **Module Search Path**
- **Package Installation Ways**

Regular Expressionist's Packages and Object – Oriented Programming in Python

- **The Sys Module**
- **Interpreter Information**
- **STDIO**
- **Launching External Programs**
- **path directories and Filenames**
- **Walking Directory Trees**
- **Math Function**
- **Random Numbers**
- **Dates and Times**
- **Zipped Archives**
- **Introduction to Python Classes**
- **Defining Classes**
- **Initializers**
- **Instance Methods**

- **Properties**
- **Class Methods and Data Static Methods**
- **Private Methods and Inheritance**
- **Module Aliases and Regular Expressions**

Machine Learning Using Python

- **Introduction to Machine Learning**
- **Areas of Implementation of Machine Learning**
- **Why Python**
- **Major Classes of Learning Algorithms**
- **Supervised vs Unsupervised Learning**
- **Learning NumPy**
- **Learning Scipy**
- **Basic plotting using Matplotlib**
- **Machine Learning application**

Supervised and Unsupervised learning

- **Classification Problem**
- **Classifying with k-Nearest Neighbours (kNN)**

Algorithm

- **General Approach to kNN**
- **Building the Classifier from Scratch**
- **Testing the Classifier**
- **Measuring the Performance of the Classifier**
- **Clustering Problem**
- **What is K-Means Clustering**
- **Clustering with k-Means in Python and an**

Application Example

- **Introduction to Pandas**
- **Creating Data Frames**
- **Grouping/Sorting**
- **Plotting Data**
- **Creating Functions**
- **Converting Different Formats**
- **Combining Data from Various Formats**
- **Slicing/Dicing Operations**

Introduction to Hadoop

- **Introduction to Scikit-Learn**
- **Inbuilt Algorithms for Use**
- **What is Hadoop and why it is popular**
- **Distributed Computation and Functional Programming**
- **Understanding Map Reduce Framework Sample Map Reduce Job Run**

Tableau

- **Future gen training & course syllabus can make the student to challenge on the success of international interviews.**

ABOUT US

- **Future Gen Technologies Training center is an engineering training institute providing trainings on job oriented professional courses for . Our training programs cover Civil Engineering, Mechanical Engineering, Electrical Engineering & IT Engineering.**
- **The Aim of the Future Gen is to develop quality engineers by providing the real time practical knowledge.**
- **Future Gen Training center has a team of highly experienced & qualified teaching staff. Many of them are experienced with real time projects. Our courses training mainly focus on instructors past experience and practical examples. This will help the trainees to have a better understanding of the concept.**
- **Future gen technologies also comprise a placement cell and enjoy placement tie ups with many companies. Student gets the placement benefits after the completion of the course. Many of our students have been successfully placed with some reputed companies in India & other countries.**

WHY FUTURE GEN ?

- **Future Gen provides the real time training on 100% job oriented courses on international projects & make every student to work on individual project in the training period.**
- **Knowledge of Students who trained under Future Gen will be as good as experienced professionals.**
- **Future gen provides digital training by In-depth demonstration on every aspect of course topics.**
- **Future Gen provides the complete support after completion of your course**
- **We provide excellent quality of infrastructure to our students**
- **Future gen provides the special training classes on personality development and interview skills after completion of course with 100% placement assistance.**
- **Highly qualified faculty with industry experience**
 - **Training will be provided based on 20% theory and 80% practical concept**
- **Get fee Refund if not satisfied**

WWW.FUTUREGENTECHNOLOGIES.COM

Thank & Regards,

FUTURE GEN TECHNOLOGIES,

AMEERPET,KPHB HYDERABAD

FUTURE GEN TECHNOLOGIES