# **Python**

* An Introduction To Python
* Beginning Python Basics
  + Python Data Structures & Data Types, operators, Type Casting
  + Python Program Flow
  + Conditional Statements
  + Looping and Iteration
  + Break and Continue
  + The Range Function
  + Break & Continue
  + Assert
* Functions & Modules
  + Built-in functions like : map, zip, reduce, filter, any, chr, ord, sorted, globals, locals, all, etc.
  + User Defined Functions
  + Lambda function
  + Recursion Function
* Exceptions Errors: Exception Handling with try
* File Handling
* Classes In Python
  + Concept of class,
  + object and instances Constructor,
  + class attributes and destructors Real time use of class in live projects Inheritance
  + overlapping and overloading operators
  + Adding and retrieving dynamic attributes of classes
  + Programming using Oops support
* Python Mysql Database Access
  + DB Connection
  + Creating DB Table INSERT, READ,UPDATE, DELETE operations
  + COMMIT & ROLLBACK operation
  + Handling Errors
* The Os Module
* Working with XML, DOM and SAX

# **Numpy**

The NumPy ndarray: A Multidimensional Array Object

* Universal Functions: Fast Element-wise Array Functions
* Data Processing Using Arrays
* File Input and Output with Arrays
* Storing Arrays on Disk in Binary Format
* Saving and Loading Text Files
* Linear Algebra
* Random Number Generation
* ndarray Object Internals
* Advanced Array Manipulation
* Broadcasting
* Structured and Record Arrays
* NumPy Matrix Class
* Advanced Array Input and Output
* Memory-mapped Files

# **Pandas**

* Introduction to pandas Data Structures
  + Series, DataFrame
* Essential Functionality
  + Reindexing, Dropping entries from an axis, Indexing, selection, and filtering,
  + Arithmetic and data alignment Function application and mapping,
  + Sorting and ranking, Axis indexes with duplicate values
* Summarizing and Computing Descriptive Statistics
  + Correlation and Covariance, Unique Values, Value Counts, and Membership
* Handling Missing Data
  + Filtering Out Missing Data, Filling in Missing Data
* Hierarchical Indexing
  + Reordering and Sorting Levels, Integer Indexing, Panel Data
* Data Loading, Storage, and File Formats
  + Reading and Writing Data in Text Format
  + Binary Data Formats
  + Interacting with HTML and Web APIs
  + Interacting with Databases
* Data Wrangling: Clean, Transform, Merge, Reshape
  + Combining and Merging Data Sets
  + Reshaping and Pivoting
  + Data Transformation
  + String Manipulation
* Data Aggregation and Group Operations
  + Group-wise Operations and Transformations
  + Pivot Tables and Cross-Tabulation
  + Date and Time Data Types and Tools
  + Converting between string and datetime
* Time Series Basics
  + Date Ranges, Frequencies, and Shifting
  + Time Zone Handling
  + Periods and Period Arithmetic
  + Resampling and Frequency Conversion
  + Time Series Plotting
* Moving Window Functions

# **PySpark Course Content**

The key objectives of this course are as follows;

* Learn the basic concept of Spark which are listed down below :
  + Basic understanding of Hadoop architecture
  + Spark Architecture
  + Spark Execution Concepts
  + to interpret DAG (Directed Acyclic Graph) for Spark Execution
  + RDD (Resilient Distributed Datasets) API
    - RDD Transformations
    - RDD Actions
  + Spark DataFrame and data sets
  + Installation of spark clustered
* Learn and hands on the Spark DataFrameAPI
  + Transformations
    - The .map(...) transformation
    - The .filter(...) transformation
    - The .flatMap(...) transformation
    - The .distinct(...) transformation
    - The .sample(...) transformation
    - The .leftOuterJoin(...) transformation
    - The .repartition(...) transformation
  + Actions
    - The .take(...) method
    - The .collect(...) method
    - The .reduce(...) method
    - The .count(...) method
    - The .saveAsTextFile(...) method
    - The .foreach(...) method
  + Create Schemas and Assign DataTypes
  + Python to RDD communications
  + Catalyst Optimizer refresh
  + Speeding up PySpark with DataFrames
  + Creating DataFrames
  + Generating our own JSON data
  + Creating a DataFrame
  + Creating a temporary table
  + Simple DataFrame queries
  + DataFrame API query
  + Interoperating with RDDs
  + Inferring the schema using reflection
  + Programmatically specifying the schema
  + Querying with the DataFrame API
  + Number of rows
  + Running filter statements
  + Querying with SQL
  + Number of rows
  + Running filter statements using the where Clauses
  + DataFrame scenario – on-time flight performance
  + Preparing the source datasets
  + Joining flight performance and airports
  + Visualizing our flight-performance data
  + Spark Dataset API