

Week	Day	Date	Topic & Skill	Content Coverage
1	Day 1	Monday, 3 March, 2025	Programming Logic and Techniques (PLT)	<p>Flow Charts –</p> <ul style="list-style-type: none"> •Definition and importance of flow charts, Real-world applications of flow charts. •Common flowchart symbols (Start/End, Input/Output, Process, Decision), Creating simple flow charts. •Step-by-step approach to designing flow charts, Practice session: Creating flow charts for basic tasks. <p>Algorithms –</p> <ul style="list-style-type: none"> •Definition, Importance, and Types of algorithms (Sequential, Conditional, Iterative). •Identifying inputs/outputs, Breaking down problems into steps, Example algorithms (Sorting, Searching). •Writing algorithms for simple tasks (e.g., finding factorial, sorting numbers). <p>Pseudo Code –</p> <ul style="list-style-type: none"> •Definition, Purpose, and Basic structure of pseudo code. •Translating flow charts into pseudo code, Writing simple pseudo code for algorithms. •Writing pseudo code for common problems (e.g., Fibonacci series, prime number check).
	Day 2	Tuesday, 4 March, 2025	Introduction to Web Technology, HTML, Basic Tags and CSS	<p>Html</p> <ul style="list-style-type: none"> •<html> •<body> •<head> •<title> •<p> •<div> •, , •<h1> to <h6> •, <i>, <u> •
, <hr> •<form> •<table> •, audio, video •<a> anchor tag <p>Introduction of CSS</p> <ul style="list-style-type: none"> •How to connect CSS with html: inline, internal, external •Basic CSS properties: width, height, background, border, font, foreground, padding, margin, color
	Day 3	Wednesday, 5 March, 2025	Introduction to Web Service	<ul style="list-style-type: none"> •What are web services •Web service architecture •REST vs SOAP •Swagger Docs for REST APIs •Http methods •Demo with Swagger pre-defined REST endpoints
	Day 4	Thursday, 6 March, 2025	Introduction to Cloud Computing and Python basics	<p>Introduction to Cloud Computing</p> <ul style="list-style-type: none"> •What is cloud computing •Why cloud computing •Examples of Cloud Computing (IaaS, SaaS, PaaS) •Types of cloud computing (private, public, hybrid, community) •Cloud Service Provider (AWS, Azure, GCP) <p>Introduction to Python</p> <ul style="list-style-type: none"> •Input output statements in python (print, input) •Print function with keyword arguments (sep, end) •Data types in python (str, int, float, complex) •Expression •Operators •Type casting •Conditional statements •Looping statements •Jumping statements •Some special functions
	Day 5	Friday, 7 March, 2025	Python : List, Tuple, and Set	<ul style="list-style-type: none"> •Introduction to Data structure in python •List in python •Introduction to List with Properties •List methods •Append, Insert, Remove, Copy (shallow and deep copy), Count, Extend, Index, Sort (ascending and descending), Reverse, Clear, Pop •List comprehensions •Indexing •Positive indexing, Negative indexing, Retrieve, update •Slicing •Positive Slicing, Negative Slicing, Retrieve, Update, Delete, insert •Searching and sorting with list •Introduction to tuple •Properties of tuple •List vs tuple •Indexing and slicing on tuple •Tuple specific methods: count, index •Python in-build functions with List and Tuple •sum •max •min •avg •len •sort •Convert string into tuple and list •Tuple and list specific programs •Set : data structure •Introduction to Sets •Introduction to sets: Unordered collection of unique elements. •Creating sets: set() constructor, curly braces {}.

	Day 6	Saturday, 8 March, 2025	Python : Dictionary, User Defined Functions	<ul style="list-style-type: none"> o Dictionary Methods and Operations • Dictionary methods: keys(), values(), items(), etc. • Adding and removing key-value pairs. • Dictionary comprehension. • Nesting dictionaries and practical examples. o Use Cases and Real-world Examples • Solving problems using sets and dictionaries. • Examples from data manipulation, filtering, and analysis. • Best practices for choosing between sets and dictionaries. • Function in Python o Introduction to function o Syntax and examples o Local and global variables o Global keyword <ul style="list-style-type: none"> • Parameters in function <ul style="list-style-type: none"> Positional arguments, Keyword arguments, Positional only arguments, Keyword only arguments, Default arguments, Variable length arguments, Keyword Variable length arguments, Pass function as an argument • Lambda function • Map, reduce and filter • Global and local variables • Return keyword in function <ul style="list-style-type: none"> Return a value, Return multiple values, Return function
2	Day 1	Monday, 10 March, 2025	Basics of DBMS, DDL, DML,	<ul style="list-style-type: none"> • Introduction to MySQL • Datatypes <ul style="list-style-type: none"> o Numeric o Date and time o String • DDL – DML commands <ul style="list-style-type: none"> o Create o Drop o Alter o Truncate • Constraints <ul style="list-style-type: none"> o Unique o Not null o default o Primary key o Foreign key o Check • DML – DML Commands <ul style="list-style-type: none"> o Update o Delete o insert
	Day 2	Tuesday, 11 March, 2025	Operators and Built-in Functions	<ul style="list-style-type: none"> • Operators <ul style="list-style-type: none"> o Arithmetic (+, -, /, %, *) o Bitwise (&,) o Comparison (<, >, =, <=, >=, <>) o Logical <ul style="list-style-type: none"> AND, ALL, ANY, BETWEEN, EXISTS, IN, LIKE, NOT, OR, SOME • DQL <ul style="list-style-type: none"> o Select query o Group by clause o Order by clause • In-build functions <ul style="list-style-type: none"> o String functions <ul style="list-style-type: none"> Char length, Ascii, concat, Instr, Lcase, Ucase, Substr, Lpad, Rpad, Trim o Numeric functions <ul style="list-style-type: none"> Avg, Count, Pow, Min, Max, Round, Sqrt, Floor o Date and time functions <ul style="list-style-type: none"> Current_date, Day_diff, Date, Current_time, Last_day, Sys_date, adddate
	Day 3	Wednesday, 12 March, 2025	Joins and Normalization	<ul style="list-style-type: none"> • Joins <ul style="list-style-type: none"> o Inner join o Left join o Right join o Full outer join o Cross join o Self join • Normalization <ul style="list-style-type: none"> o Anomalies o Keys (candidate keys) o Primary key and foreign key constraints o 1st Normal Form o 2nd Normal Form o 3rd Normal Form o Boyce cod normal form
	Day 4	Thursday, 13 March, 2025	Mock Test-1 + Debriefing	Recap - Practice - Mocktests
	Day 5	Friday, 14 March, 2025	Mock Test-2 + Debriefing	Recap - Practice - Mocktests
	Day 6	Saturday, 15 March, 2025	Assessment	Milestone 1 Assessment

	Day 1	Monday, 17 March, 2025	Complex Joins, Subqueries	<ul style="list-style-type: none">• Subqueries<ul style="list-style-type: none">◦ Single row subqueries◦ Multi row subqueries◦ Correlated subqueries queries• Joins<ul style="list-style-type: none">◦ Joins with subqueries◦ Joins with aggregate functions
	Day 2	Tuesday, 18 March, 2025	Analytical and Advanced Functions	<ul style="list-style-type: none">• RANK• DENSE_RANK• ROW_NUMBER• CUME_DIST• LAG• LEAD
	Day 3	Wednesday, 19 March, 2025	Procedures, Functions and Cursor & Date and Time functions	<ul style="list-style-type: none">• Procedures• Functions• IN• OUT• INOUT• Cursors User defined, Pre-definedJoins with date and time functions
	Day 4	Thursday, 20 March, 2025	TCL, Triggers and Views	<ul style="list-style-type: none">• TCL commands<ul style="list-style-type: none">◦ Commit◦ Rollback◦ Save point• Triggers - using insert, update, delete operations• View<ul style="list-style-type: none">◦ Creating/replacing views◦ Dropping views

	Day 5	Friday, 21 March, 2025	NoSQL and MongoDB	<ul style="list-style-type: none"> • Introduction to NoSQL • Overview of relational databases vs NoSQL databases. • Types of NoSQL databases: Document-oriented, Key-Value, Column-family, Graph databases. • Characteristics of NoSQL databases: Schema-less, scalability, flexibility. • Use cases for choosing NoSQL databases. • Introduction to MongoDB <ul style="list-style-type: none"> ◦ Overview of MongoDB: Document-oriented database. ◦ MongoDB's JSON-like BSON format. ◦ Collections and Documents: Equivalent to tables and rows in relational databases. ◦ Installation of MongoDB locally. • MongoDB CRUD Operations <ul style="list-style-type: none"> ◦ Inserting documents. ◦ Querying documents. ◦ Updating documents. ◦ Deleting documents. • MongoDB Data Modeling <ul style="list-style-type: none"> ◦ Introduction to BSON data types. ◦ Schema design best practices. ◦ Embedded documents and arrays. ◦ Relationships and normalization in MongoDB. • Indexing and Query Optimization <ul style="list-style-type: none"> ◦ Importance of indexing. ◦ Creating and managing indexes. ◦ Query optimization techniques. ◦ Explain plan and performance monitoring. • Introduction to MongoDB Atlas <ul style="list-style-type: none"> ◦ Overview of MongoDB Atlas. ◦ Creating a free MongoDB Atlas account.
	Day 6	Saturday, 22 March, 2025	Linux Basics and Linux	<ul style="list-style-type: none"> • Linux Fundamentals -Overview • What is Linux and Linux Distributions • Components of Linux System • Basic Features of Linux System • Linux File System • Bash and PowerShell Scripting • Daemons • Graphical Servers and Desktop Environment • Applications • What is Powershell • BootLoader • Linux Kernel • Linux Commands related to editor • Linux commands related to version control • Linux General Commands • SED editor
4	Day 1	Monday, 24 March, 2025	Command Line Programming	<ul style="list-style-type: none"> • Ability to obtain data using command line from various resources like CSV, JSON, XML etc.. • Ability to create reusable command line tools • Ability to apply scrubbing operations using command line operators • Ability to manage data using command line operators • Ability to explore data using command line operators • Ability to work with parallel pipelines using command line operators
	Day 2	Tuesday, 25 March, 2025	Git and DevOps	<p>Learning Objective Git and Agile</p> <p>Introduction to Version Control Systems</p> <p>Introduction to Git</p> <p>Git Installation</p> <p>Basic Terminal Commands</p> <p>Git init</p> <p>Git config</p> <p>Tracked and Untracked Files</p> <p>Git log</p> <p>Git commit</p> <p>Git checkout</p> <p>Git reset</p> <p>Git ignore</p> <p>Git add and commit</p> <p>Introduction to Github</p> <p>Building Github repository</p> <p>Pushing local repository to Github</p> <p>Commit and Pull request in Github</p> <p>Branching in Github</p> <p>Merge conflict</p> <p>Git Rebase</p> <p>Merging branches and resolving conflict in Github</p> <p>Learning outcomes- Git</p> <p>Introduction to Devops</p>

	Day 3	Wednesday, 26 March, 2025	Functions and OOPs in Python	<ul style="list-style-type: none"> • Revision of data structure in python (List, Set, Tuple, Dict) • Function in python Objects & Class <ul style="list-style-type: none"> o Parameters in function <ul style="list-style-type: none"> Positional arguments, Keyword arguments, Positional only arguments, Keyword only arguments, Default arguments, Variable length arguments, Keyword Variable length arguments, Pass function as an argument o Lambda function o Map, reduce and filter o Global and local variables o Return keyword in function <ul style="list-style-type: none"> Return a value, Return multiple values, Return function • OOPs in python <ul style="list-style-type: none"> o Introduction to OOPs principles o General OOPs vs Python OOPs o Class and object o Data encapsulation o Abstraction o Data hiding o Methods in class <ul style="list-style-type: none"> Normal methods, Class methods, Static methods, Data members in class, Instance variables, Class variables o Methods with arguments <ul style="list-style-type: none"> Pass object as an argument, Object as an return type, Method overloading
	Day 4	Thursday, 27 March, 2025	constructor, destructor , inheritance	<ul style="list-style-type: none"> • Special methods <ul style="list-style-type: none"> __init__ __str__ __new__ • Constructor and destructor in python <ul style="list-style-type: none"> Constructor with default arguments Constructor without arguments Constructor with arguments Destructor: Define and invocation of destructor • Inheritance in python <ul style="list-style-type: none"> Introduction and properties of inheritance MRO (method resolution order) Types of inheritance Constructor and destructor with inheritance • Method overriding <ul style="list-style-type: none"> Method overriding Super method Super method with special methods (__init__, __str__)
	Day 5	Friday, 28 March, 2025	Python Exception Handling, SQL and NoSQL Connectivity	<ul style="list-style-type: none"> • Exception Handling <ul style="list-style-type: none"> o Introduction to Exceptions <ul style="list-style-type: none"> • Definition of Exceptions • Why Exception Handling is important o Basics of Exception Handling <ul style="list-style-type: none"> • Syntax of try and except blocks • Handling multiple exceptions • Using else and finally blocks o Common Built-in Exceptions <ul style="list-style-type: none"> • SyntaxError • IndentationError • NameError • TypeError • ValueError • ZeroDivisionError • Custom exceptions (brief introduction) o Handling Specific Exceptions <ul style="list-style-type: none"> • Using except with specific exception types • Using as to get exception details o Raising Exceptions <ul style="list-style-type: none"> • Using the raise statement o Exception Hierarchy <ul style="list-style-type: none"> • Understanding the hierarchy of exceptions in Python • Handling exceptions at different levels o Best Practices for Exception Handling <ul style="list-style-type: none"> • Properly documenting code with exceptions • Avoiding broad except clauses • Using finally for cleanup operations o Practical Examples and Exercises <ul style="list-style-type: none"> • Writing code snippets to demonstrate exception handling

5	Day 6	Saturday, 29 March, 2025	Python Unit Testing and NumPy	<ul style="list-style-type: none"> • Pytest <ul style="list-style-type: none"> o Introduction to Testing <ul style="list-style-type: none"> o Briefly explain the importance of testing in software development. o Different types of testing (unit testing, integration testing, etc.). o Benefits of automated testing. o Setting Up Pytest <ul style="list-style-type: none"> o Install Pytest using pip. o Directory structure for test files and the naming convention. o Organize test files and test functions. o Writing Your First Test <ul style="list-style-type: none"> o Basic structure of a Pytest test function. o Demonstrate writing simple test functions for basic assertions. o Common assertion methods provided by Pytest. • Hands-On Exercise • Numpy <ul style="list-style-type: none"> o Introduction to NumPy <ul style="list-style-type: none"> • NumPy as a fundamental library for numerical operations in Python. • Benefits of using NumPy over standard Python lists. o Installing NumPy <ul style="list-style-type: none"> • Demonstrate how to install NumPy using pip. • Verify the installation and version. o NumPy Arrays <ul style="list-style-type: none"> o Introduce NumPy arrays as the core data structure. o Discuss the advantages of NumPy arrays over lists. o Demonstrate creating arrays using np.array() and exploring array attributes (shape, size, dtype). o Other Numpy functions <ul style="list-style-type: none"> • asarray(), ones(), zeros(), empty()
	Day 1	Tuesday, 1 April, 2025	Deep dive NumPy and Intro to Pandas	<p>Numpy</p> <ul style="list-style-type: none"> • Array Operations and Broadcasting <ul style="list-style-type: none"> o Explore basic array operations (addition, subtraction, multiplication, etc.). o Broadcasting for performing operations on arrays of different shapes. o Element-wise operations and their efficiency. o Indexing and Slicing <ul style="list-style-type: none"> • Indexing and slicing in NumPy arrays. • Boolean indexing • Fancy indexing • Copy of ndarray o Array Manipulation <ul style="list-style-type: none"> • Discuss reshaping, flattening, and concatenating arrays. • Introduce stacking and splitting of arrays. • Explore methods for adding and removing elements. o Hands-On Exercises <p>Pandas</p> <ul style="list-style-type: none"> o Introduction to Pandas <ul style="list-style-type: none"> • Importance of data manipulation in data science and analysis. • Pandas a library for handling structured data in Python. • Benefits of using Pandas for data cleaning and analysis. o Installing Pandas <ul style="list-style-type: none"> • Demonstrate how to install Pandas using pip. • Verify the installation and version. o Pandas Series <ul style="list-style-type: none"> • Series are one-dimensional labeled arrays. • Creating Series from lists, arrays, and dictionaries. • Explore Series attributes and methods for data manipulation. o Pandas DataFrame <ul style="list-style-type: none"> • Pandas DataFrames are two-dimensional labeled data structures. • Creating DataFrames from dictionaries, lists, and external files.
	Day 2	Wednesday, 2 April, 2025	Deep dive Pandas	<ul style="list-style-type: none"> o Data Inspection and Exploration <ul style="list-style-type: none"> o Methods for inspecting data, such as head(), tail(), and info(). o Importance of understanding the structure and types of data. o Descriptive statistics using describe(). o Data Cleaning with Pandas <ul style="list-style-type: none"> o Common data cleaning tasks, such as handling missing values. o Methods for handling missing data, including dropna() and fillna(). o Handle duplicates and outliers. o Indexing and Selection in DataFrames <ul style="list-style-type: none"> o loc and iloc for label-based and integer-based indexing. o slicing o boolean indexing o Boolean indexing for filtering data. o Data Manipulation and Transformation <ul style="list-style-type: none"> o Methods for reshaping and transforming data, including groupby() and pivot_table(). o Demonstrate how to apply functions to data using apply() and map(). o Explore merging and concatenating DataFrames. o Hands-On Exercises
	Day 3	Thursday, 3 April, 2025	Visualization using Matplotlib and Seaborn	<ul style="list-style-type: none"> o Introduction to Data Visualization <ul style="list-style-type: none"> o Importance of data visualization in conveying information. o Introduce key concepts such as data types (categorical, numerical) and chart types (scatter plots, bar charts, etc.). o Role of visualization in data analysis and storytelling. o Installing Matplotlib and Seaborn <ul style="list-style-type: none"> o Install Matplotlib and Seaborn using pip. o Importance of using virtual environments for dependency management. o Verify the installation and version. o Matplotlib <ul style="list-style-type: none"> o Introduction to Matplotlib <ul style="list-style-type: none"> • Matplotlib as a versatile plotting library. • Create basic plots (line plots, scatter plots) using pyplot. • The anatomy of a Matplotlib figure (axes, labels, title). o Customizing Matplotlib Plots <ul style="list-style-type: none"> • Explore customization options for plots, including colors, markers, and line styles. • Set titles, labels, and legends. o Working with Different Types of Plots <ul style="list-style-type: none"> • Line Plot • Scatter Plot • Bar Plot • Histogram • Box Plot • Pie Chart • Area Plot o Seaborn <ul style="list-style-type: none"> o Introduction to Seaborn <ul style="list-style-type: none"> • Role of Seaborn in statistical data visualization. • Seaborn's integration with Matplotlib and Pandas.
	Day 4	Friday, 4 April, 2025	Mock Test-1 + Debriefing	Recap - Practice - Mocktests
	Day 5	Saturday, 5 April, 2025	Mock Test-2 + Debriefing	Recap - Practice - Mocktests

Day 1	Monday, 7 April, 2025	Assessment	Milestone 2 Assessment
Day 2	Tuesday, 8 April, 2025	Data Warehouse Fundamentals	<p>Understanding traditional Relational Database Management Systems (RDBMS)</p> <ul style="list-style-type: none"> Challenges encountered with RDBMS Issues related to scaling up in RDBMS Different RDBMS cluster approaches <p>Data Warehousing Architecture</p> <ul style="list-style-type: none"> Introduction to Data Warehousing Data Warehousing Architecture Overview Components of Data Warehousing Architecture Importance of Data Warehousing in modern enterprises <p>Designing Data Warehouse and Introduction to Big Data</p> <ul style="list-style-type: none"> Concepts of Fact Tables and Dimension Tables Designing Fact Tables and Dimensions ETL (Extract, Transform, Load) process in Data Warehousing Best practices in designing ETL pipelines <p>Big Data Introduction and Getting Started with Hadoop</p> <ul style="list-style-type: none"> Overview of Big Data and its challenges Introduction to Hadoop as a Big Data solution Getting started with Hadoop Understanding the Hadoop framework <p>Hadoop Framework and HDFS</p> <ul style="list-style-type: none"> Stepping into the Hadoop ecosystem Hadoop 2.x - YARN (Yet Another Resource Negotiator) Understanding the role of YARN in Hadoop Practical insights into Hadoop applications <p>HDFS (Hadoop Distributed File System)</p> <ul style="list-style-type: none"> Introduction to HDFS Why HDFS is crucial for Big Data processing Working on HDFS - hands-on exercises Real-world applications of HDFS
Day 3	Wednesday, 9 April, 2025	Big Data, Hadoop and HDFS as Data Lake	<p>MapReduce Programming Paradigm</p> <ul style="list-style-type: none"> Introduction to MapReduce Understanding the MapReduce programming paradigm Key concepts and principles behind MapReduce <p>Closer Look at MapReduce</p> <ul style="list-style-type: none"> Deep dive into the components of MapReduce Exploring the Map phase and Reduce phase Hands-on examples of MapReduce programs <p>Practical Approach to MapReduce</p> <ul style="list-style-type: none"> Real-world use cases of MapReduce Best practices and optimization techniques Challenges and solutions in implementing MapReduce <p>Understanding MapReduce with an Example</p> <ul style="list-style-type: none"> Walkthrough of a practical MapReduce example Explanation of Map and Reduce functions Analyzing the output and results <p>Hadoop 1.x vs Hadoop 2.x</p> <ul style="list-style-type: none"> Evolution of Hadoop from version 1.x to 2.x Key enhancements and features in Hadoop 2.x Impacts on MapReduce programming model <p>HDFS as a Data Lake</p> <ul style="list-style-type: none"> Overview of Hadoop Distributed File System (HDFS) Utilizing HDFS as a Data Lake Storing and managing large-scale data in HDFS File Formats and Their Use <p>Cases</p> <ul style="list-style-type: none"> Introduction to various file formats (CSV, Parquet, ORC, Avro, etc.) Use cases for different file formats in Hadoop Choosing the right file format for specific scenarios <p>Hadoop Installation (Step by Step)</p> <ul style="list-style-type: none"> Step-by-step guide to installing Hadoop
Day 4	Thursday, 10 April, 2025	Hive and Scoop	<p>Apache Hive Basics</p> <ul style="list-style-type: none"> Introduction to Apache Hive Key features and advantages of using Hive Hive architecture and components <p>Hive Illustration - External Table in Hive (Partition & Bucketing)</p> <ul style="list-style-type: none"> Understanding External Tables in Hive Implementing partitioning and bucketing for optimization Hands-on illustration with external tables <p>Hive Illustration - Loading Different File Formats (Serialization & Deserialization)</p> <ul style="list-style-type: none"> Overview of file formats in Hive (CSV, Parquet, ORC, Avro) Serialization and deserialization in Hive Loading data in different file formats <p>Hive Illustration - Loading Data into Hive Tables</p> <ul style="list-style-type: none"> Step-by-step process of loading data into Hive tables Handling various data sources and formats Best practices for efficient data loading <p>Hive Illustration - Simple Operations on Hive Table</p> <ul style="list-style-type: none"> Performing basic operations on Hive tables (INSERT, UPDATE, DELETE) Data manipulation and transformation in Hive Practical examples of simple operations <p>Hive Illustration - Query Operations on Hive Table (RegEX)</p> <ul style="list-style-type: none"> Introduction to Regular Expressions (RegEX) in Hive Query operations using RegEX on Hive tables Real-world scenarios and use cases <p>Hive Illustration - Querying Complex Structures</p> <ul style="list-style-type: none"> Handling complex data structures in Hive (Arrays, Maps, Structs) Querying and processing nested data Advanced examples of querying complex structures <p>Hive Illustration - Views</p> <ul style="list-style-type: none"> Creating and managing views in Hive

	Day 5	Friday, 11 April, 2025	Spark and PySpark	<p>Getting Started - Spark Basics</p> <ul style="list-style-type: none"> • Introduction to Apache Spark • Key features and advantages • Spark ecosystem overview <p>Spark vs Hadoop</p> <ul style="list-style-type: none"> • Comparative analysis of Spark and Hadoop • Use cases and scenarios for choosing Spark <p>Spark - Architecture</p> <ul style="list-style-type: none"> • High-level architecture of Apache Spark • Components and their roles in Spark architecture <p>Driver Program</p> <ul style="list-style-type: none"> • Understanding the role of the Driver Program in Spark • Interaction with the Spark cluster <p>Cluster Manager</p> <ul style="list-style-type: none"> • Overview of Cluster Managers in Spark (e.g., Apache Mesos, Apache YARN) • Managing resources in a Spark cluster <p>Executor, Tasks</p> <ul style="list-style-type: none"> • Executor role in Spark • Task execution in Spark clusters <p>Direct Acyclic Graph (DAG)</p> <ul style="list-style-type: none"> • Introduction to DAG in the context of Spark • Understanding the execution flow in Spark <p>Spark Terminologies</p> <ul style="list-style-type: none"> • Explanation of key Spark terminologies • Clarification of terms used in Spark ecosystem <p>Introduction to RDD</p> <ul style="list-style-type: none"> • Overview of Resilient Distributed Datasets (RDD) in Spark • Characteristics and advantages of RDD <p>Pyspark - Hands-on</p> <ul style="list-style-type: none"> • Practical introduction to PySpark
	Day 6	Saturday, 12 April, 2025	RDD and Spark Cluster	<p>Spark Program Lifecycle</p> <ul style="list-style-type: none"> • Overview of the lifecycle of a Spark program • Understanding the stages of Spark program execution <p>RDDs - Building Blocks of Spark</p> <ul style="list-style-type: none"> • Introduction to Resilient Distributed Datasets (RDDs) • Characteristics and benefits of using RDDs in Spark <p>RDD - Transformations</p> <ul style="list-style-type: none"> • Explanation of RDD transformations • Commonly used transformations in Spark <p>RDD Operations - Actions</p> <ul style="list-style-type: none"> • Overview of RDD actions in Spark • Examples of actions triggering the execution of Spark computations <p>RDD Lineage</p> <ul style="list-style-type: none"> • Understanding RDD lineage in Spark • Fault tolerance mechanisms provided by RDD lineage <p>Concepts of Execution Methods - Lazy Execution, etc.</p> <ul style="list-style-type: none"> • Lazy execution in Spark and its benefits • Strategies for optimizing Spark program execution <p>Data Pipeline</p> <ul style="list-style-type: none"> • Designing and implementing a data pipeline in Spark • Connecting various stages in a Spark data pipeline <p>RDD from Python Collections</p> <ul style="list-style-type: none"> • Creating RDDs from Python collections • Hands-on exercises for working with Python collections in Spark <p>RDD from File</p> <ul style="list-style-type: none"> • Loading data into Spark RDD from external files • File formats and considerations for RDD creation <p>Scheduling Workflows using Oozie</p> <ul style="list-style-type: none"> • Introduction to Apache Oozie • Designing and scheduling Spark workflows with Oozie
	Day - 1	Monday, 14 April, 2025	Working with Spark Data frames	<p>Transition to DataFrames</p> <ul style="list-style-type: none"> • Challenges with RDDs and the need for DataFrames. • Introduction to DataFrames and their benefits. • Transformation from RDDs to DataFrames. <p>Working with DataFrames</p> <ul style="list-style-type: none"> • Creating DataFrames: Loading data from various sources (CSV, JSON, etc.). • Basic operations and transformations on DataFrames. • Hands-on exercises for practice. <p>Aggregations in Spark DataFrames</p> <ul style="list-style-type: none"> • Overview of aggregation operations (groupBy, agg, etc.). • Examples of common aggregate functions (sum, avg, max, min, count). • Hands-on exercises for aggregations. <p>DataFrames from CSV</p> <ul style="list-style-type: none"> • Loading data from CSV files into Spark DataFrames. • Handling CSV-specific challenges (e.g., handling headers, handling missing data). • Hands-on exercises for working with CSV data. <p>Advanced DataFrame Operations</p> <ul style="list-style-type: none"> • Joins, unions, and other advanced DataFrame operations. <p>Introduction to Spark SQL</p> <ul style="list-style-type: none"> • Overview of Spark SQL • Integration of SQL queries with Spark • Key features and use cases of Spark SQL <p>Spark SQL as an ETL Tool</p> <ul style="list-style-type: none"> • Understanding ETL processes with Spark SQL <p>Spark SQL Performance Tuning</p> <ul style="list-style-type: none"> • Strategies for optimizing performance in Spark SQL • Best practices for improving query performance <p>Spark Streaming and Real-Time Data Analytics</p> <ul style="list-style-type: none"> • Introduction to Spark Streaming for real-time data processing
	Day - 2	Tuesday, 15 April, 2025	Spark Streaming and Kafka	<ul style="list-style-type: none"> • Introduction to Kafka - The Need for Kafka • Understanding the challenges addressed by Apache Kafka • Importance of real-time, distributed, and fault-tolerant data streaming • Use cases and scenarios where Kafka excels • Kafka Components • Overview of key components in the Kafka ecosystem • Broker, Producer, Consumer, Zookeeper, and other essential elements • Roles and responsibilities of each component in a Kafka cluster • Kafka Workflow • Deep dive into the typical workflow of Kafka • Producing and consuming messages • Kafka topics and partitions • Message serialization and deserialization • Spark Streaming • Introduction to Apache Spark Streaming • Real-time data processing using micro-batch processing • Integration of Spark Streaming with Kafka for scalable and fault-tolerant stream processing • Use cases and benefits of Spark Streaming

7	Day - 3	Wednesday, 16 April, 2025	Introduction to Kafka	<ul style="list-style-type: none"> • Introduction to Kafka - The Need for Kafka • Understanding the challenges addressed by Apache Kafka • Importance of real-time, distributed, and fault-tolerant data streaming • Use cases and scenarios where Kafka excels • Kafka Components • Overview of key components in the Kafka ecosystem • Broker, Producer, Consumer, Zookeeper, and other essential elements • Roles and responsibilities of each component in a Kafka cluster • Kafka Workflow • Deep dive into the typical workflow of Kafka • Producing and consuming messages • Kafka topics and partitions • Message serialization and deserialization • Spark Streaming • Introduction to Apache Spark Streaming • Real-time data processing using micro-batch processing • Integration of Spark Streaming with Kafka for scalable and fault-tolerant stream processing • Use cases and benefits of Spark Streaming
	Day - 4	Thursday, 17 April, 2025	Scala Introduction & Introduction to Airflow	<ul style="list-style-type: none"> • Variables, Methods, Classes, and Objects <ul style="list-style-type: none"> ◦ Introduction to basic concepts in Scala programming ◦ Declaration and usage of variables ◦ Defining methods and functions ◦ Creating classes and objects in Scala • Package and Package Objects <ul style="list-style-type: none"> ◦ Overview of packages in Scala ◦ Organizing code into packages ◦ Utilizing package objects for shared functionality • Higher-Order Functions in Scala <ul style="list-style-type: none"> ◦ Understanding higher-order functions ◦ Passing functions as arguments ◦ Returning functions as results ◦ Practical examples of higher-order functions in Scala • Scala Collections <ul style="list-style-type: none"> ◦ Introduction to Scala collections ◦ Immutable and mutable collections ◦ Common operations on collections ◦ Selection of appropriate collection types for specific use cases • SBT (Scala Build Tool) <ul style="list-style-type: none"> ◦ Overview of SBT as a build tool for Scala projects ◦ Project structure and configuration in SBT ◦ Managing dependencies with SBT ◦ Building, testing, and running Scala applications using SBT <p>Introduction to Airflow</p> <ul style="list-style-type: none"> • Overview of Apache Airflow • Role in orchestrating complex workflows and data pipelines • Key features and benefits of using Apache Airflow • Airflow Architecture & Components <p>Deep dive into the architecture of Apache Airflow</p>
	Day - 5	Friday, 18 April, 2025	Working with Airflow & Advanced Apache Airflow	<ul style="list-style-type: none"> • Airflow UI <ul style="list-style-type: none"> ◦ In-depth exploration of the Airflow web-based user interface ◦ Navigating through DAGs, tasks, and logs ◦ Utilizing the UI for monitoring and managing workflows • Reading Logs <ul style="list-style-type: none"> ◦ Understanding how logs are generated and stored in Airflow ◦ Reading and interpreting task logs for troubleshooting ◦ Leveraging logging features for debugging workflows • Task Actions <ul style="list-style-type: none"> ◦ Exploring actions that can be performed on tasks ◦ Retry, mark success, and other task actions ◦ Managing tasks through the Airflow UI • Deleting a DAG <ul style="list-style-type: none"> ◦ Best practices and considerations for deleting a DAG ◦ Steps to safely remove a DAG from the Airflow instance ◦ Impact of DAG deletion on running workflows • Xcoms in Airflow <ul style="list-style-type: none"> ◦ Introduction to Xcoms (Cross-Communication) in Airflow ◦ Sharing data between tasks within a DAG ◦ Use cases and examples of using Xcoms • Implementing Xcoms <ul style="list-style-type: none"> ◦ Practical implementation of Xcoms in Airflow ◦ Configuring tasks to exchange data using Xcoms ◦ Handling data dependencies between tasks • Variables in Airflow <ul style="list-style-type: none"> ◦ Overview of Airflow Variables ◦ Storing and managing dynamic values ◦ Using variables for parameterizing workflows • Dummy Operator <ul style="list-style-type: none"> ◦ Understanding the Dummy Operator in Airflow
	Day - 6	Saturday, 19 April, 2025	Data Management Overview	<ul style="list-style-type: none"> • Overview of Data Management <ul style="list-style-type: none"> ◦ Definition and significance of data management ◦ Evolution of data management in enterprises ◦ Challenges and opportunities in modern data management • Importance of Data Management at Enterprises <ul style="list-style-type: none"> ◦ Understanding the critical role of data in business operations ◦ Impact of effective data management on decision-making ◦ Compliance and regulatory considerations • Multidimensional Data Representation and Manipulation <ul style="list-style-type: none"> ◦ Exploring multidimensional data models ◦ Manipulating data in multiple dimensions ◦ Use cases and benefits of multidimensional data representation • Design Practices and Methodologies <ul style="list-style-type: none"> ◦ Overview of design practices for efficient data management ◦ Methodologies for designing scalable and maintainable data solutions ◦ Incorporating best practices in data management design • Understanding Data Warehouse, Data Lake, Lake House <ul style="list-style-type: none"> ◦ Differentiating between data warehouse, data lake, and lake house ◦ Use cases and scenarios for each data storage architecture ◦ Integration and interoperability between these data storage solutions • Data Management Architecture <ul style="list-style-type: none"> ◦ Building a comprehensive data management architecture ◦ Components and layers of a robust data architecture ◦ Scalability, flexibility, and performance considerations • Quality Check - How to Build a Quality Check Pipeline <ul style="list-style-type: none"> ◦ Implementing data quality checks in a data pipeline ◦ Setting up validation and verification processes ◦ Ensuring data quality throughout the data lifecycle • Governance - How to Create Governance Layer in Pipeline <ul style="list-style-type: none"> ◦ Establishing data governance principles

8	Day - 1	Monday, 21 April, 2025	Software Engineering Essentials	<ul style="list-style-type: none"> Event-Based Spark Streams <ul style="list-style-type: none"> Understanding event-based stream processing in Apache Spark Real-time data processing with Spark Streaming Use cases and scenarios for event-based streaming in Spark Stream Processing with Flink <ul style="list-style-type: none"> Introduction to Apache Flink for stream processing Data streams and their characteristics Time-based window operators for processing streaming data Stateful operators in Flink Reading and writing from/to external systems in Flink Debugging in Stream Processing <ul style="list-style-type: none"> Strategies for debugging streaming applications Identifying and resolving common issues in real-time data processing Debugging tools and techniques for Spark and Flink Docker and Kubernetes <ul style="list-style-type: none"> Introduction to containerization with Docker Orchestrating containers with Kubernetes Benefits of containerization and orchestration in stream processing Microservices (Only Theory) <ul style="list-style-type: none"> Overview of microservices architecture Decoupling and scaling components in a microservices environment Considerations for building microservices in a streaming data ecosystem FAST API (Document and Ready Demo) <ul style="list-style-type: none"> Introduction to FAST API for building APIs in Python Documenting APIs with Swagger
	Day - 2	Tuesday, 22 April, 2025	Cloud Fundamentals	<p>The Classical Enterprise</p> <ul style="list-style-type: none"> Understanding the traditional enterprise model Characteristics and challenges of classical enterprises <p>Myths of Cloud Computing</p> <ul style="list-style-type: none"> Debunking common misconceptions about cloud computing Clarifying misconceptions related to security, reliability, and cost <p>Service Delivery Models</p> <ul style="list-style-type: none"> Overview of cloud service delivery models (IaaS, PaaS, SaaS) Choosing the right service model for specific business needs <p>Degree of Abstraction</p> <ul style="list-style-type: none"> Understanding the level of abstraction in cloud services Benefits and considerations of abstraction in the cloud <p>Cloud Attributes</p> <ul style="list-style-type: none"> Key attributes defining cloud computing (scalability, flexibility, etc.) Impacts of cloud attributes on business operations <p>Service Offerings</p> <ul style="list-style-type: none"> Exploring various cloud service offerings Infrastructure, platform, and software services in the cloud <p>Managed Services</p> <ul style="list-style-type: none"> Introduction to managed services in the cloud Outsourcing operational responsibilities to cloud providers <p>Cost Economics</p> <ul style="list-style-type: none"> Cost considerations in cloud computing Analyzing the economics of cloud services <p>Elasticity</p> <ul style="list-style-type: none"> Understanding elasticity in the context of cloud computing Scaling resources based on demand and workload <p>Virtualization</p> <ul style="list-style-type: none"> Overview of virtualization technology in the cloud Benefits and use cases of virtualization
	Day - 3	Wednesday, 23 April, 2025	Azure Fundamentals, Services and Infrastructure	<p>Definitions, Stories & Business Concerns</p> <ul style="list-style-type: none"> Classical Enterprise, Why Cloud & Evolution of Cloud Service Models, Abstraction Levels (v1), SPIDERS Cloud Attributes, Managed Services & Deployment Models Pricing & Scaling Models Hands-on Demo: Creating an Azure account, Setting up a basic project environment in Azure Introduction to Virtualization Containers vs VMs, PaaS & Services Taxonomy Price Economics, Data Velocity & Distributed Computing Apps for Cloud & Security Model Azure Networking Components Azure Active Directory Hands-on Demo: Configuring Azure Virtual Network and security groups, Integrating Azure Active Directory for access control
	Day - 4	Thursday, 24 April, 2025	Azure Storage Services	<ul style="list-style-type: none"> Blob Storage File Storage ADLS Introduction Explain Azure Data Lake storage Use Monitoring and Analytics to Gain Operational Insights Hands-on Demo: Uploading files to Azure Blob Storage, Configuring Azure Data Lake Storage for big data analytics
	Day - 5	Friday, 25 April, 2025	Azure Database Services and Data Pipelines	<ul style="list-style-type: none"> Azure SQL Azure CosmosDB ADF - Quick Start Data Streams in Azure Streaming Analytics Hands-on Demo: Creating and configuring Azure SQL databases, Building real-time data pipelines with Azure Data Factory
	Day - 6	Saturday, 26 April, 2025	Advance Azure Data and Analytics Services	<ul style="list-style-type: none"> Azure Data Factory (ADF) Deep-dive Azure Databricks Deep-dive HDInsight Architecting and implementing a data engineering solution with Azure Data Services Hands-on Demo: Performing big data analytics with Azure Databricks, Setting up and running Hadoop clusters with Azure HDInsightData Visualization and Reporting
	Day - 7	Sunday, 27 April, 2025	Project Integration and Presentation	<ul style="list-style-type: none"> Integrating data sources and pipelines into a cohesive project Project Presentation Preparation Final Q&A and Feedback Session Hands-on Demo: Finalizing project components and ensuring data integrity, Practice presentations and peer feedback sessions
9	Day - 1	Monday, 28 April, 2025	Data Visualization and Reporting	<ul style="list-style-type: none"> Intro to Power BI Power BI Dashboard Creation Building Interactive Reports Hands-on Demo: Creating dashboards and reports in Power BI, Connecting Power BI to Azure data sources for real-time insights
	Day - 2	Tuesday, 29 April, 2025	Project Integration and Presentation	<ul style="list-style-type: none"> Integrating data sources and pipelines into a cohesive project Project Presentation Preparation Final Q&A and Feedback Session Hands-on Demo: Finalizing project components and ensuring data integrity, Practice presentations and peer feedback sessions
	Day - 3	Wednesday, 30 April, 2025	Mock Test-1 + Debriefing	Recap - Practice - Mocktests
	Day - 5	Friday, 2 May, 2025	Mock Test-2 + Debriefing	Recap - Practice - Mocktests

	Day - 6	Saturday, 3 May, 2025	Assessment	Milestone 3 Assessment
10	Day - 1	Monday, 5 May, 2025	Assessment	Mock Client Interviews
	Day - 2	Tuesday, 6 May, 2025	Capstone Project	Group Project
	Day - 3	Wednesday, 7 May, 2025	Capstone Project	Group Project
	Day - 4	Thursday, 8 May, 2025	Capstone Project	Group Project
	Day - 5	Friday, 9 May, 2025	Capstone Project	Group Project
	Day - 6	Saturday, 10 May, 2025	Capstone Project	Group Project
11	Day - 1	Monday, 12 May, 2025	Capstone Project	Group Project
	Day - 2	Tuesday, 13 May, 2025	Assessment	Milestone - 4 : Capstone Project Evaluation