

01. Programming Language :

a. Python

- i. Basic Syntax
- ii. Variables
- iii. Data Types
- iv. Operators
- v. List
- vi. Tuples
- vii. Sets
- viii. Dictionaries
- ix. Conditional Statements (If..Else)
- x. Loops
- xi. Try...Except
- xii. Reading Files (CSV,JSON, TEXT, Excel)
- xiii. Writing Files
- xiv. Functions
- xv. Working with Dates

b. Scala

c. Java

**The practice of hackerrank or leetcode with easy problems
(10-15)**

Time for learning - 2 Weeks

02. Data Structures & Algorithms (Basic):

- a. Time Complexity and Space Complexity (Big O notation)
- b. Arrays
- c. Linked List
- d. Stack
- e. Queue
- f. Tree
- g. Graph

- h. Searching
 - i. Linear Search
 - ii. Binary Search
 - iii. Interpolation Search
- i. Sorting
 - i. Selection Sort
 - ii. Insertion Sort
 - iii. Merge Sort
 - iv. Quick Sort
 - v. Heap Sort

Practice of geeksforgeeks with easy problems (10-12)
Time for learning - 1-2 Months (Depending on previous experience)

- 03. Database Fundamentals :
 - a. DDL (CREATE, DROP, ALTER, TRUNCATE, RENAME)
 - b. DCL (GRANT and REVOKE)
 - c. DML (INSERT, UPDATE, DELETE)
 - d. TCL (COMMIT, ROLLBACK)
 - e. Aggregation (MAX, MIN, FIRST, AVG, COUNT, SUM)
 - f. Integrity Constraints (Primary Key, Foreign Key)
 - g. Data Schema
 - h. ACID Properties
 - i. Views
 - j. Stored Procedures
 - k. ER and Relational Diagrams
 - l. Indexing
 - m. Hashing
 - n. Normalization forms
- 04. SQL Scripting :
 - a. Transactional Databases : MySQL, PostgreSQL

- b. Joins (Left, Inner, Outer, Full, Right)
- c. Sub Queries
- d. UNION Statement
- e. Date Function
- f. Nested Queries
- g. Group By
- h. Having
- i. CASE Statements
- j. Window Functions

Practice of hackerrank or leetcode with easy problems (10-15)

Time for learning - 3-4 Weeks (section 3 and 4)

- 05. BigData Fundamentals :
 - a. BigData Basics and Characteristics?
 - b. 5 V's of BigData
 - c. Vertical vs Horizontal Scaling
 - d. Scaling Up and Scaling Out
 - e. ETL Pipelines
 - f. File formats
 - i. CSV
 - ii. JSON
 - iii. AVRO
 - iv. Parquet
 - v. ORC
 - g. Type of Data
 - i. Structured
 - ii. Unstructured
 - iii. Semi-structured

Time for learning - 1 Week (Only Theory)

- 06. Cluster Computing
 - a. Hadoop Ecosystem
 - i. HDFS
 - ii. Map-Reduce
 - iii. Yarn
 - b. Apache Hive
 - i. How to load data in different file formats
 - ii. Internal Tables
 - iii. External Tables
 - iv. Querying table data stored in HDFS
 - v. Partitioning
 - vi. Bucketing
 - vii. Map-Side Join
 - viii. Sorted-Merge Join
 - ix. UDF in Hive
 - x. SerDe in Hive
- 07. Apache Spark
 - a. Spark Core
 - b. Spark SQL
 - c. Spark Streaming
 - d. Difference Between Hadoop and Spark

Time for learning - 3-4 Weeks (Hands-on and theory)

- 08. Data Processing
 - a. Batch Processing
 - b. Real-Time Processing
 - c. Hybrid Processing

Time for learning - 1-2 Weeks (Understand basic concept)

- 09. Data Warehousing Fundamentals:

- a. OLAP vs OLTP
- b. Dimension Tables
- c. Data Cube
- d. Extract Transform Load (ETL)
- e. E-R Modeling VS Dimensional Modeling
- f. Fact Tables
- g. Star Schema
- h. Snowflake Schema
- i. Warehouse Designing Questions

Time for learning - 1-2 Weeks (Theory)

10. Data Exploration Libraries:

- a. Pandas
 - i. Reading and writing CSV & JSON
 - ii. DataFrames and Series
 - iii. Head, tail
 - iv. Info()
 - v. Dropping columns
 - vi. Sorting
 - vii. Apply
 - viii. Filter
 - ix. Loc and iloc
 - x. Shape, Index, Columns
 - xi. Lambda
 - xii. Basic Arithmetic Functions
 - xiii. Join and Merge
- b. NumPy
 - i. Creating Arrays
 - ii. Indexing and Slicing
 - iii. Copy vs View
 - iv. Shape
 - v. Reshape

- vi. Split
- vii. Join
- viii. Sort, Search, Filter, Split
- c. Matplotlib
 - i. Pyplot
 - ii. Plotting
 - iii. Lines
 - iv. Legends
 - v. Labels
 - vi. Grid
 - vii. Scatter
 - viii. Bars
 - ix. Histogram
 - x. Pie Charts
 - xi. Seaborn

Time for learning - 1-2 Weeks (Theory and HandsOn)

- 11. Data Orchestration (AirFlow) :
 - a. Intro to Airflow
 - b. Implementing Airflow DAGs
 - c. Maintaining and monitoring Airflow workflows
 - d. Building production pipelines in Airflow

Time for learning - 1-2 Weeks (Theory and HandsOn)

- 12. NoSQL:
 - a. Difference between NoSQL vs SQL
 - b. Features of NoSQL
 - c. Types of NoSQL database
 - d. CAP Theorem
 - e. Eventual Consistency
 - f. Tools -
 - i. HBase

- ii. Cassandra
- iii. AWS DynamoDB
- iv. MongoDB

Time for learning - 2-3 Weeks (Theory and HandsOn)
Learn MongoDB or Cassandra

13. Message Queue or Streaming Services :
- a. Apache Kafka
 - b. Apache Beam
 - c. AWS Kinesis

Time for learning - 2-3 Weeks (Theory and HandsOn)
Pick one and learn

14. Dashboarding Tools :
- a. Tableau
 - b. QuickSight
 - c. Data Studio
 - d. Looker

Time for learning - 2 Weeks (Theory and HandsOn)
Build some dashboards (will tell you about projects in future videos)

15. Cloud Services (AWS) :
- a. OnDemand Machines
 - i. AWS EC2
 - b. Access Management
 - i. AWS IAM
 - c. Object Storage
 - i. AWS S3
 - d. Transactional Database Services
 - i. AWS RDS

1. MySQL
2. Arora
3. PostgreSQL

- e. Adhoc Query
 - i. AWS Athena
- f. Data Warehouse
 - i. AWS Redshift
- g. NoSQL Database Services
 - i. AWS DynamoDB
- h. Serverless
 - i. AWS Lambda
- i. ETL Services
 - i. AWS Glue
- j. For Storing and Accessing Credentials
 - i. AWS Secret Manager
- k. Log Services
 - i. AWS Cloudwatch
 - ii. AWS Config
- l. Distributed Data Computation
 - i. AWS EMR
- m. Messaging Queue
 - i. AWS SNS
 - ii. AWS SQS
- n. Real Time Data Processing
 - i. AWS Kinesis
 - ii. AWS Firehose
 - iii. AWS Analytics
- o. Networking (Advance Leve)
 - i. VPC
 - ii. Subnets
 - iii. NACL
 - iv. Security Groups

- v. VPC Peering
- vi. VPN
- p. Security
 - i. KMS
 - ii. WAF

Time for learning - 2-3 Months (Theory and HandsOn)

Learning fundamentals, doing hands-on practice with projects