Syllabus:-

1. Introduction to Bigdata and Hadoop
2. Introduction to Big Data & Big Data Challenges
3. Limitations & Solutions of Big Data Architecture
4. Comparison between Traditional system vs. Hadoop Ecosystem.
5. Hadoop & its Features
6. Hadoop Ecosystem
7. Hadoop Storage: HDFS (Hadoop Distributed File System)
8. Hadoop Processing: MapReduce Framework
9. Hadoop Architecture and HDFS
10. Brief Introduction to MapReduce [MR]
11. MR2 Installation on Linux
12. YARN Architecture and its components.
13. Brief Introduction Hadoop Distribution File System [HDFS]
14. Anatomy of MapReduce Program
15. Input Splits, Relation between Input Splits and HDFS Blocks
16. MapReduce: Combiner & Partitioner
17. Overview of Different files format and Compression technique.
18. Introduction to Apache Sqoop
19. Installation of Sqoop on Linux
20. Sqoop Architecture.
21. Sqoop components – Import, Export, Merge, Job, etc.,
22. Introduction to Apache Hive
23. Installation of Hive on Linux
24. Hive Architecture and Components
25. Hive Metastore
26. Limitations of Hive
27. Comparison with Traditional Database
28. Hive Data Types and Data Models
29. Hive Analytical functions
30. Hive Partition
31. Hive Bucketing
32. Hive Tables (Managed Tables and External Tables)
33. Importing Data
34. Querying Data & Managing Outputs
35. Hive Script & Hive UDF
36. Supported file format in Apache Hive
37. POC on Sqoop and Hive
38. Apache HBase
39. HBase Installation on Linux
40. HBase v/s RDBMS
41. HBase Architecture.
42. Components in HBase
43. Apache Spark [Batch Processing]
44. Spark Installation on Linux
45. Spark Architecture
46. MapReduce vs. Spark
47. Spark Components
48. Spark RDD, DataFrame, SparkSQL
49. Handling File Format in Spark
50. POC on Spark for understanding Batch processing.
51. Apache Kafka
52. Kafka Installation on Linux
53. Architecture of Apache Kafka
54. Kafka vs. Flume
55. Kafka components
56. Apache Spark Streaming [streaming processing]
57. Apache Spark Streaming Architecture
58. Basic components
59. Performance Tuning
60. Fault-tolerance Semantics
61. Kafka and Spark Integration
62. Apache OOzie
63. [Overview](https://oozie.apache.org/docs/4.0.1/DG_Overview.html) and Installation on Linux
64. [Workflow Functional Specification](https://oozie.apache.org/docs/4.0.1/WorkflowFunctionalSpec.html)
65. [Coordinator Functional Specification](https://oozie.apache.org/docs/4.0.1/CoordinatorFunctionalSpec.html)
66. [Bundle Functional Specification](https://oozie.apache.org/docs/4.0.1/BundleFunctionalSpec.html)
67. [Command Line Tool](https://oozie.apache.org/docs/4.0.1/DG_CommandLineTool.html)
68. [Workflow Re-runs Explained](https://oozie.apache.org/docs/4.0.1/DG_WorkflowReRun.html)
69. [HCatalog Integration Explained](https://oozie.apache.org/docs/4.0.1/DG_HCatalogIntegration.html)