Pyspark Syllabus:

Python Programming Spark:

(a) Python Setup
(b) Python Object and Data Structure Basics
(c) Python Comparison Operators
(d) Python Statements
(e) Methods and Functions

Core Spark:

(a) Writing a Core Spark application using Python
(b) How we can initialize an Spark application
(c) Running Spark jobs on cluster using YARN
(d) How to create an RDD
(e) How to create an RDD using file or using a directory in HDFS
(f) How we can persist an RDD on disk or in memory
(g) How to apply Spark transformations on an RDD using filtering and aggregations
(h) Ways to perform actions on an Spark RDD
(i) Ways to create and use broadcast variables and accumulators
(j) How to configure Spark properties
(k) Ways to ingest data using SparkSession
(l) How we can sort the results and write this out to HDFS(Hadoop)
or other destinations supported

Spark SQL:

(a) How to Create Spark DataFrames from an existing RDD
(b) How we can Perform operations on a DataFrame
(c) How to Write a Spark SQL application
(d) Using Hive with ORC from Spark SQL
(e) Writing a Spark SQL application that directly reads and writes data from Hive tables
(f) Ways to Invoke SQL API or SparkSession SQL functionality to select and produce results
Using join capabilities produce analytic results
(g) Hw to rename a DataFrame/Dataset columns to produce best results

Spark Streaming:

(a) Invoking and using Spark structured streaming to ingest data in real time
(b) Invoking streaming transformations and aggregations to produce analytic results
(c) Invoking spark-submit utility on existing Spark application using proper arguments