HDP Certified Spark Developer Exam Objectives

# Core Spark

* Write a Spark Core application in Python or Scala
* Initialize a Spark application
* Run a Spark job on YARN
* Create an RDD
* Create an RDD from a file or directory in HDFS
* Persist an RDD in memory or on disk
* Perform Spark transformations on an RDD such as filtering and aggregations
* Perform Spark actions on an RDD
* Create and use broadcast variables and accumulators
* Configure Spark properties
* Ingest data using Spark Session
* Sort results and write out to HDFS or other supported destinations

# Spark SQL

* Create Spark DataFrames from an existing RDD
* Perform operations on a DataFrame
* Write a Spark SQL application
* Use Hive with ORC from Spark SQL
* Write a Spark SQL application that reads and writes data from Hive tables
* Invoke SQL API or SparkSession SQL functionality to select and produce results
* Using join capabilities produce analytic results
* Rename DataFrame/Dataset columns to produce best results
* Spark Streaming
* Use Spark structured streaming to ingest data in real time
* Invoke streaming transformations and aggregations to produce analytic results
* Invoke spark-submit utility on existing Spark application using proper arguments

# Other skills:

* Should learn the HDFS and MapReduce.
* Understand Hadoop 2.x architecture.
* Thoroughly read the chapters 1 to 9 in “Learning Spark” by O’Reilly publication.
* Practice lot of questions of Spark core and Spark SQL in Spark-shell.
* Learn Machine learning theories like K-mean and Regression.
* Learn how to work in RDD in Spark.
* Learn GraphX basics like Vertex and Edge RDD.
* Should learn Lineage Graph.
* Acquire proper knowledge of the variables of Accumulator and Broadcast.