How to Prepare For Databricks Developer Certification for Apache Spark

Exam details:

There were 40 MCQs for which you will get 90minutes to solve. And one must score 50% atleast to pass the exam. Exam is not tough but, tricky. 90% of questions were from Spark core, Spark streaming, Spark SQL. There were only two questions from Machine learning and GraphX, one from each. Out of 40, around five questions were theoritical remaining about coding. Questions can be from all the 3 languages ( Java, Scala, Python) but, are not language specific. So being able to read and understand the code in all the three languages is sufficient to clear the exam.

Topics to be covered:

1. All transformations and actions on Rdds.

<http://homepage.cs.latrobe.edu.au/zhe/ZhenHeSparkRDDAPIExamples.html>

1. Word count example in all the languages. There was a question for me on word count in java.
2. Lazy execution model of DAG.
3. Different persistence levels for RDD.
4. Stateful operations (Window & UpdateStateByKey) in spark streaming.
5. Spark SQL- Creating DataFrames and querying it with and without SQL.
6. Machine learning – K-means, Regression and other algos, knowing the API is enough.
7. Tuning and Debugging guide from spark docs and Learning Spark.

* Best practices for less time of computation and to avoid Memory bottlenecks.
* Accumulators and Broadcast variables.
* Sending Functions to spark and Understanding the Closures.

<http://spark.apache.org/docs/1.6.0/programming-guide.html#passing-functions-to-spark>

1. Deployment guide.

Resources:

* ‘Learning Spark’ book.
* Spark documentation. I followed 1.6.1 version.

<http://spark.apache.org/docs/1.6.1/>

* Advanced Exercise from spark-summit.

<https://spark-summit.org/2014/training/>

* For hands on ,practice all the example available in spark documentation and Learning spark.