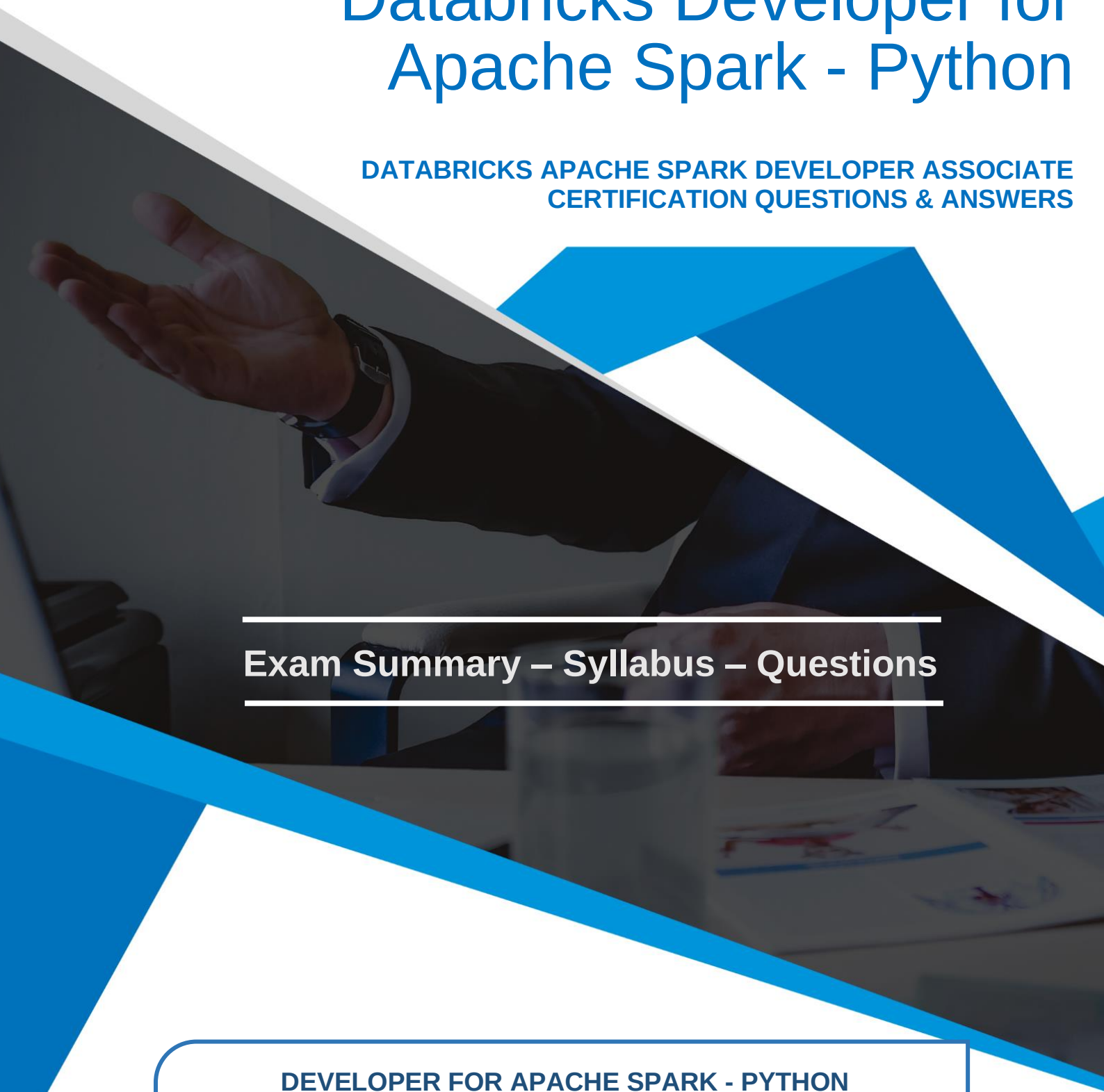
Databricks Developer for Apache Spark - Python

DATABRICKS APACHE SPARK DEVELOPER ASSOCIATE CERTIFICATION QUESTIONS & ANSWERS



Exam Summary – Syllabus – Questions

DEVELOPER FOR APACHE SPARK - PYTHON

Databricks Certified Associate Developer for Apache Spark

60 Questions Exam – 70% Cut Score – Duration of 120 minutes

www.CertFun.com

Table of Contents

Know Your Developer for Apache Spark - Python Certification Well:

The Developer for Apache Spark - Python is best suitable for candidates who want to gain knowledge in the Databricks Specialty. Before you start your Developer for Apache Spark - Python preparation you may struggle to get all the crucial Apache Spark Developer Associate materials like Developer for Apache Spark - Python syllabus, sample questions, study guide.

But don't worry the Developer for Apache Spark - Python PDF is here to help you prepare in a stress free manner.

The PDF is a combination of all **your queries like**-

- What is in the Developer for Apache Spark - Python syllabus?
- How many questions are there in the Developer for Apache Spark - Python exam?
- Which Practice test would help me to pass the Developer for Apache Spark - Python exam at the first attempt?

Passing the Developer for Apache Spark - Python exam makes you Databricks Certified Associate Developer for Apache Spark. Having the Apache Spark Developer Associate certification opens multiple opportunities for you. You can grab a new job, get a higher salary or simply get recognition within your current organization.

Databricks Developer for Apache Spark - Python Apache Spark Developer Associate Certification Details:

Exam Name	Databricks Certified Associate Developer for Apache Spark
Exam Code	Developer for Apache Spark - Python
Exam Price	$200 (USD)
Duration	120 mins
Number of Questions	60
Passing Score	70%
Schedule Exam	**Kryterion Webassesor**

Sample Questions	Databricks Developer for Apache Spark - Python Sample Questions
Practice Exam	Databricks Developer for Apache Spark - Python Certification Practice Exam

Developer for Apache Spark - Python Syllabus:

Topic	Weights
Apache Spark Architecture Concepts	17%
Apache Spark Architecture Applications	11%
Apache Spark DataFrame API Applications	72%

Databricks Developer for Apache Spark - Python Sample Questions:

Question: 1

Which of the following statements is NOT true for broadcast variables?

a) It provides a mutable variable that a Spark cluster can safely update on a per-row basis.
b) It is a way of updating a value inside of a variety of transformations and propagating that value to the driver node in an efficient and fault-tolerant way.
c) You can define your own custom broadcast class by extending org.apache.spark.util.BroadcastV2 in Java or Scala or pyspark.AccumulatorParams in Python.
d) Broadcast variables are shared, immutable variables that are cached on every machine in the cluster instead of serialized with every single task.
e) The canonical use case is to pass around a small large table that does fit in memory on the executors.

Answer: a, b, c

Question: 2

Which of the following are valid execution modes?

a) Kubernetes, Local, Client
b) Client, Cluster, Local
c) Server, Standalone, Client
d) Cluster, Server, Local
e) Standalone, Client, Cluster

Answer: b

Question: 3

Which of the following code blocks adds a column predErrorSqrt to DataFrame transactionsDf that is the square root of column predError?

a) transactionsDf.withColumn("predErrorSqrt", sqrt(col("predError")))
b) transactionsDf.withColumn("predErrorSqrt", sqrt(predError))
c) transactionsDf.select(sqrt(predError))
d) transactionsDf.withColumn("predErrorSqrt", col("predError").sqrt())
e) transactionsDf.select(sqrt("predError"))

Answer: a

Question: 4

The code blown down below intends to join df1 with df2 with inner join but it contains an error. Identify the error.

d1.join(d2, “inner”, d1.col(“id”) === df2.col(“id"))

a) The join type is not in right order. The correct query should be d2.join(d1, d1.col(“id”) === df2.col(“id"), “inner”)
b) There should be two == instead of ===. So the correct query is d1.join(d2, “inner”, d1.col(“id”) == df2.col(“id"))
c) Syntax is not correct d1.join(d2, d1.col(“id”) == df2.col(“id"), “inner”)
d) We cannot do inner join in spark 3.0, but it is in the roadmap.

Answer: c

Question: 5

What command we can use to get the number of partition of a dataframe named df?

a) df.rdd.getPartitionSize()
b) df.getPartitionSize()
c) df.getNumPartitions()
d) df.rdd.getNumPartitions()

Answer: d

Question: 6

Which of the following DataFrame methods is classified as a transformation?

a) DataFrame.count()
b) DataFrame.show()
c) DataFrame.select()
d) DataFrame.foreach()
e) DataFrame.first()

Answer: c

Question: 7

The code block displayed below contains an error. The code block is intended to join DataFrame itemsDf with the larger DataFrame transactionsDf on column itemId. Find the error. Code block: transactionsDf.join(itemsDf, "itemId", how="broadcast")

a) The syntax is wrong, how= should be removed from the code block.
b) The join method should be replaced by the broadcast method.
c) Spark will only perform the broadcast operation if this behavior has been enabled on the Spark cluster.
d) The larger DataFrame transactionsDf is being broadcasted, rather than the smaller DataFrame itemsDf
e) broadcast is not a valid join type.

Answer: e

Question: 8

If spark is running in client mode, which of the following statement about is correct?

a) Spark driver is randomly attributed to a machine in the cluster
b) Spark driver is attributed to the machine that has the most resources
c) Spark driver remains on the client machine that submitted the application
d) The entire spark application is run on a single machine.

Answer: e

Question: 9

Which of the following three DataFrame operations are classified as an action? (Choose 3 answers)

a) PrintSchema()
b) Show()
c) First()
d) limit()
e) foreach()
f) cache

Answer: b, c, e

Question: 10

If we want to create a constant integer 1 as a new column ‘new_column’ in a dataframe df, which code block we should select?

a) df.withColumnRenamed('new_column', lit(1))
b) df.withColumn(new_column, lit(1))
c) df.withColumn(”new_column”, lit(“1”))
d) df.withColumn(“new_column”, 1)
e) df.withColumn(“new_column”, lit(1))

Answer: e

Study Guide to Crack Databricks Apache Spark Developer Associate Developer for Apache Spark - Python Exam:

- Getting details of the Developer for Apache Spark - Python syllabus, is the first step of a study plan. This pdf is going to be of ultimate help. Completion of the syllabus is must to pass the Developer for Apache Spark - Python exam.
- Making a schedule is vital. A structured method of preparation leads to success. A candidate must plan his schedule and follow it rigorously to attain success.
- Joining the Databricks provided training for Developer for Apache Spark - Python exam could be of much help. If there is specific training for the exam, you can discover it from the link above.
- Read from the Developer for Apache Spark - Python sample questions to gain your idea about the actual exam questions. In this PDF useful sample questions are provided to make your exam preparation easy.
- Practicing on Developer for Apache Spark - Python practice tests is must. Continuous practice will make you an expert in all syllabus areas.