Databricks Associate-Developer-Apache-Spark題庫資訊&新版Associate-Developer-Apache-Spark考古題

2025 KaoGuTi最新的Associate-Developer-Apache-Spark PDF版考試題庫和Associate-Developer-Apache-Spark考試問題和答案免費分享：https://drive.google.com/open?id=1jdVKHpzX6Y4ViXfgs3lAY3WyBAmW1Hh7

KaoGuTi是一個為參加Associate-Developer-Apache-Spark認證考試的考生提供Associate-Developer-Apache-Spark認證考試培訓工具的網站。KaoGuTi提供的培訓工具很有針對性，可以幫他們節約大量寶貴的時間和精力。我們的練習題及答案和真實的考試題目很接近。短時間內使用KaoGuTi的模擬測試題你就可以100%通過考試。這樣花少量的時間和金錢換取如此好的結果，是值得的。快將KaoGuTi提供的培訓工具放入你的購物車中吧。

Databricks 經驗認證帶有全球認可及高度重視的價值，對於使用 Apache Spark 的企業而言更是如此。此證照能展現考生對 Apache Spark 的深刻理解及開發 Spark 應用程式能力。對開發者而言，此證照是展示技能及在大數據及數據分析領域能力前行的最佳捷徑。

要為Databricks Associate-Developer-Apache-Spark 認證考試做好準備，考生可以利用各種資源。Databricks 提供了一個免費自學的在線課程，涵蓋了通過考試所需的主題和技能。課程包括實踐練習和測驗以加強學習。此外，還有幾本書籍和在線教程可供選擇，介紹了Apache Spark 及其相關技術。

>> Databricks Associate-Developer-Apache-Spark題庫資訊 <<

熱門的Associate-Developer-Apache-Spark題庫資訊，覆蓋大量的Databricks認證Associate-Developer-Apache-Spark考試知識點

KaoGuTi Databricks的Associate-Developer-Apache-Spark考試培訓資料是所有的互聯網培訓資源裏最頂尖的培訓資料，我們的知名度度是很高的，這都是許多考生利用了KaoGuTi Databricks的Associate-Developer-Apache-Spark考試培訓資料所得到的成果，如果你也使用我們KaoGuTi Databricks的Associate-Developer-Apache-Spark考試培訓資料，我們可以給你100%成功的保障，若是沒有通過，我們將保證退還全部購買費用，為了廣大考生的切身利益，我們KaoGuTi絕對是信的過的。

Databricks Associate-Developer-Apache-Spark考試是一項認證考試，專門為希望展示其Apache Spark專業知識的專業人員。該考試由Databricks提供，這是一個基於雲的平台，可提供數據工程，數據科學和機器學習解決方案。該考試是對候選人對Apache Spark及其生態系統知識的全面測試，包括Spark SQL，Spark Streaming和Spark Mllib。

最新的 Databricks Certification Associate-Developer-Apache-Spark 免費考試真題 (Q177-Q182):

問題 #177
Which of the following is one of the big performance advantages that Spark has over Hadoop?

- A. Spark achieves great performance by storing data and performing computation in memory, whereas large jobs in Hadoop require a large amount of relatively slow disk I/O operations.
- B. Spark achieves great performance by storing data in the DAG format, whereas Hadoop can only use parquet files.

- C. Spark achieves higher resiliency for queries since, different from Hadoop, it can be deployed on Kubernetes.
- D. Spark achieves great performance by storing data in the HDFS format, whereas Hadoop can only use parquet files.
- E. Spark achieves performance gains for developers by extending Hadoop's DataFrames with a user-friendly API.

答案：A

解題說明：
Explanation
Spark achieves great performance by storing data in the DAG format, whereas Hadoop can only use parquet files.
Wrong, there is no "DAG format". DAG stands for "directed acyclic graph". The DAG is a means of representing computational steps in Spark. However, it is true that Hadoop does not use a DAG.
The introduction of the DAG in Spark was a result of the limitation of Hadoop's map reduce framework in which data had to be written to and read from disk continuously.
Graph DAG in Apache Spark - DataFlair
Spark achieves great performance by storing data in the HDFS format, whereas Hadoop can only use parquet files.
No. Spark can certainly store data in HDFS (as well as other formats), but this is not a key performance advantage over Hadoop. Hadoop can use multiple file formats, not only parquet.
Spark achieves higher resiliency for queries since, different from Hadoop, it can be deployed on Kubernetes.
No, resiliency is not asked for in the question. The question is about performance improvements.
Both Hadoop and Spark can be deployed on Kubernetes.
Spark achieves performance gains for developers by extending Hadoop's DataFrames with a user-friendly API.
No. DataFrames are a concept in Spark, but not in Hadoop.

問題 #178
Which of the following code blocks reads in the parquet file stored at location filePath, given that all columns in the parquet file contain only whole numbers and are stored in the most appropriate format for this kind of data?

- A. 1.spark.read.schema(
  2. StructType([
  3. StructField("transactionId", IntegerType(), True),
  4. StructField("predError", IntegerType(), True)]
  5.)).format("parquet").load(filePath)
- B. 1.spark.read.schema([
  2. StructField("transactionId", IntegerType(), True),
  3. StructField("predError", IntegerType(), True)
  4.]).load(filePath, format="parquet")
- C. 1.spark.read.schema(
  2. StructType(
  3. StructField("transactionId", IntegerType(), True),
  4. StructField("predError", IntegerType(), True)
  5.)).load(filePath)
- D. 1.spark.read.schema(
  2. StructType([
  3. StructField("transactionId", StringType(), True),
  4. StructField("predError", IntegerType(), True)]
  5.)).parquet(filePath)
- E. 1.spark.read.schema([
  2. StructField("transactionId", NumberType(), True),
  3. StructField("predError", IntegerType(), True)
  4.]).load(filePath)

答案：A

解題說明：
Explanation
The schema passed into schema should be of type StructType or a string, so all entries in which a list is passed are incorrect.
In addition, since all numbers are whole numbers, the IntegerType() data type is the correct option here.
NumberType() is not a valid data type and StringType() would fail, since the parquet file is stored in the "most appropriate format for this kind of data", meaning that it is most likely an IntegerType, and Spark does not convert data types if a schema is provided.
Also note that StructType accepts only a single argument (a list of StructFields). So, passing multiple arguments is invalid.

Finally, Spark needs to know which format the file is in. However, all of the options listed are valid here, since Spark assumes parquet as a default when no file format is specifically passed.
More info: pyspark.sql.DataFrameReader.schema - PySpark 3.1.2 documentation and StructType - PySpark 3.1.2 documentation

問題 #179
The code block displayed below contains an error. The code block should return a new DataFrame that only contains rows from DataFrame transactionsDf in which the value in column predError is at least 5. Find the error.
Code block:
transactionsDf.where("col(predError) >= 5")

- A. The expression returns the original DataFrame transactionsDf and not a new DataFrame. To avoid this, the code block should be transactionsDf.toNewDataFrame().where("col(predError) >= 5").
- B. The argument to the where method should be "predError >= 5".
- C. Instead of where(), filter() should be used.
- D. The argument to the where method cannot be a string.
- E. Instead of >=, the SQL operator GEQ should be used.

答案：B

解題說明：
Explanation
The argument to the where method cannot be a string.
It can be a string, no problem here.
Instead of where(), filter() should be used.
No, that does not matter. In PySpark, where() and filter() are equivalent.
Instead of >=, the SQL operator GEQ should be used.
Incorrect.
The expression returns the original DataFrame transactionsDf and not a new DataFrame. To avoid this, the code block should be transactionsDf.toNewDataFrame().where("col(predError) >= 5").
No, Spark returns a new DataFrame.
Static notebook | Dynamic notebook: See test 1
(https://flrs.github.io/spark_practice_tests_code/#1/27.html ,
https://bit.ly/sparkpracticeexams_import_instructions)

問題 #180
Which of the following code blocks returns only rows from DataFrame transactionsDf in which values in column productId are unique?

- A. transactionsDf.dropDuplicates(subset="productId")
- B. transactionsDf.unique("productId")
- C. transactionsDf.drop_duplicates(subset="productId")
- D. transactionsDf.distinct("productId")
- E. transactionsDf.dropDuplicates(subset=["productId"])

答案：E

解題說明：
Explanation
Although the question suggests using a method called unique() here, that method does not actually exist in PySpark. In PySpark, it is called distinct(). But then, this method is not the right one to use here, since with distinct() we could filter out unique values in a specific column.
However, we want to return the entire rows here. So the trick is to use dropDuplicates with the subset keyword parameter. In the documentation for dropDuplicates, the examples show that subset should be used with a list. And this is exactly the key to solving this question: The productId column needs to be fed into the subset argument in a list, even though it is just a single column.
More info: pyspark.sql.DataFrame.dropDuplicates - PySpark 3.1.1 documentation Static notebook | Dynamic notebook: See test 1

問題 #181

Which of the following code blocks returns about 150 randomly selected rows from the 1000-row DataFrame transactionsDf, assuming that any row can appear more than once in the returned DataFrame?

- A. transactionsDf.sample(0.85, 8429)
- B. transactionsDf.sample(0.15, False, 3142)
- C. transactionsDf.sample(True, 0.15, 8261)
- D. transactionsDf.sample(0.15)
- E. transactionsDf.resample(0.15, False, 3142)

答案：C

解題說明：
Explanation
Answering this question correctly depends on whether you understand the arguments to the DataFrame.sample() method (link to the documentation below). The arguments are as follows:
DataFrame.sample(withReplacement=None, fraction=None, seed=None).
The first argument withReplacement specified whether a row can be drawn from the DataFrame multiple times. By default, this option is disabled in Spark. But we have to enable it here, since the question asks for a row being able to appear more than once. So, we need to pass True for this argument.
About replacement: "Replacement" is easiest explained with the example of removing random items from a box. When you remove those "with replacement" it means that after you have taken an item out of the box, you put it back inside. So, essentially, if you would randomly take 10 items out of a box with 100 items, there is a chance you take the same item twice or more times. "Without replacement" means that you would not put the item back into the box after removing it. So, every time you remove an item from the box, there is one less item in the box and you can never take the same item twice.
The second argument to the withReplacement method is fraction. This referes to the fraction of items that should be returned. In the question we are asked for 150 out of 1000 items - a fraction of 0.15.
The last argument is a random seed. A random seed makes a randomized processed repeatable. This means that if you would re-run the same sample() operation with the same random seed, you would get the same rows returned from the sample() command. There is no behavior around the random seed specified in the question. The varying random seeds are only there to confuse you!
More info: pyspark.sql.DataFrame.sample - PySpark 3.1.1 documentation
Static notebook | Dynamic notebook: See test 1

問題 #182
.....

Spark ⇐免費下載Associate-Developer-Apache-Spark考試資料

- 免費下載Associate-Developer-Apache-Spark題庫資訊 - 新版Databricks Certified Associate Developer for Apache Spark 3.0 Exam考古題 □ ➡ tw.fast2test.com □□□上的▹ Associate-Developer-Apache-Spark ◃免費下載只需搜尋新版Associate-Developer-Apache-Spark題庫
- Associate-Developer-Apache-Spark考試大綱 □ Associate-Developer-Apache-Spark更新 □ 最新Associate-Developer-Apache-Spark題庫 □ 開啟➽ www.newdumpspdf.com □輸入☀ Associate-Developer-Apache-Spark □☀□並獲取免費下載Associate-Developer-Apache-Spark更新
- Associate-Developer-Apache-Spark題庫資訊：Databricks Certified Associate Developer for Apache Spark 3.0 Exam可靠的認證資源 □ 免費下載□ Associate-Developer-Apache-Spark □只需在✔ tw.fast2test.com □✔□上搜索Associate-Developer-Apache-Spark認證指南
- kemi0713.blogs-service.com, www.stes.tyc.edu.tw, tutor.shmuprojects.co.uk, www.stes.tyc.edu.tw, andicreative.com, www.stes.tyc.edu.tw, campus.academiamentesana.com, www.stes.tyc.edu.tw, mrvsfoodandbeverageblueprint.com, mrsameh-ramadan.com, Disposable vapes

P.S. KaoGuTi在Google Drive上分享了免費的、最新的Associate-Developer-Apache-Spark考試題庫：https://drive.google.com/open?id=1jdVKHpzX6Y4ViXfgs3lAY3WyBAmW1Hh7