DP-201 Q&As

Designing an Azure Data Solution

Pass Microsoft DP-201 Exam with 100% Guarantee

Free Download Real Questions & Answers **PDF** and **VCE** file from:

https://www.pass4itsure.com/dp-201.html

100% Passing Guarantee
100% Money Back Assurance

Following Questions and Answers are all new published by Microsoft Official Exam Center

QUESTION 1

You are evaluating data storage solutions to support a new application.

You need to recommend a data storage solution that represents data by using nodes and relationships in graph structures.

Which data storage solution should you recommend?

A. Blob Storage

B. Cosmos DB

C. Data Lake Store

D. HDInsight

Correct Answer: B

For large graphs with lots of entities and relationships, you can perform very complex analyses very quickly. Many graph databases provide a query language that you can use to traverse a network of relationships efficiently.

Relevant Azure service: Cosmos DB

References:

https://docs.microsoft.com/en-us/azure/architecture/guide/technology-choices/data-store-overview

QUESTION 2

You need to design a real-time stream solution that uses Azure Functions to process data uploaded to Azure Blob Storage. The solution must meet the following requirements:

1.

Support up to 1 million blobs.

2.

Scaling must occur automatically.

3.

Costs must be minimized. What should you recommend?

A. Deploy the Azure Function in an App Service plan and use a Blob trigger.

B. Deploy the Azure Function in a Consumption plan and use an Event Grid trigger.

C. Deploy the Azure Function in a Consumption plan and use a Blob trigger.

D. Deploy the Azure Function in an App Service plan and use an Event Grid trigger.

Correct Answer: C

Create a function, with the help of a blob trigger template, which is triggered when files are uploaded to or updated in Azure Blob storage. You use a consumption plan, which is a hosting plan that defines how resources are allocated to your function app. In the default Consumption Plan, resources are added dynamically as required by your functions. In this serverless hosting, you only pay for the time your functions run. When you run in an App Service plan, you must manage the scaling of your function app.

References: https://docs.microsoft.com/en-us/azure/azure-functions/functions-create-storage-blob-triggered-function

QUESTION 3

HOTSPOT

You are designing a data processing solution that will run as a Spark job on an HDInsight cluster. The solution will be used to provide near real-time information about online ordering for a retailer.

The solution must include a page on the company intranet that displays summary information.

The summary information page must meet the following requirements:

1. Display a summary of sales to date grouped by product categories, price range, and review scope.

2. Display sales summary information including total sales, sales as compared to one day ago and sales as compared to one year ago.

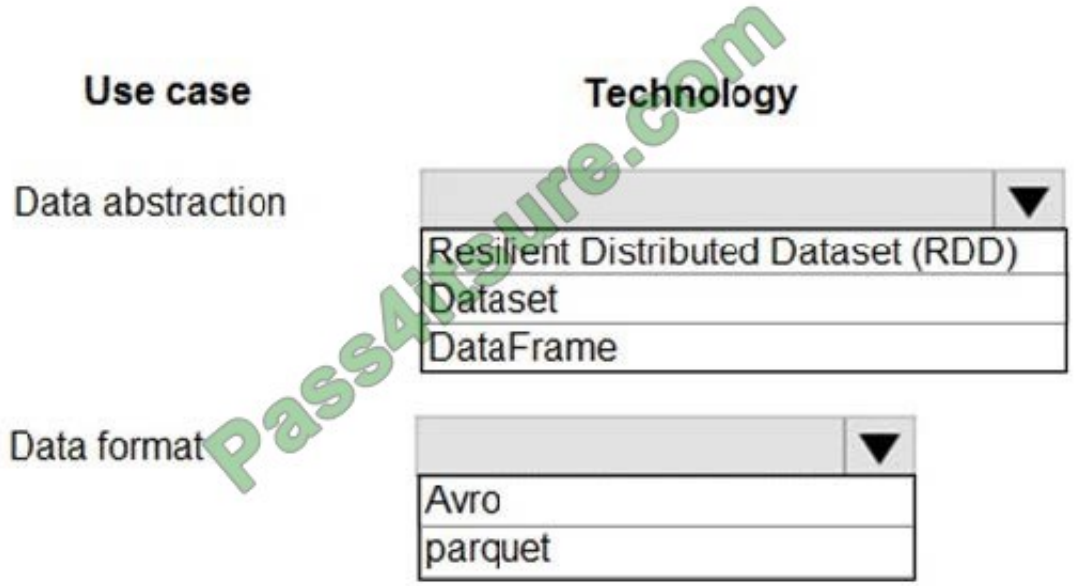
3. Reflect information for new orders as quickly as possible.

You need to recommend a design for the solution.

What should you recommend? To answer, select the appropriate configuration in the answer area.

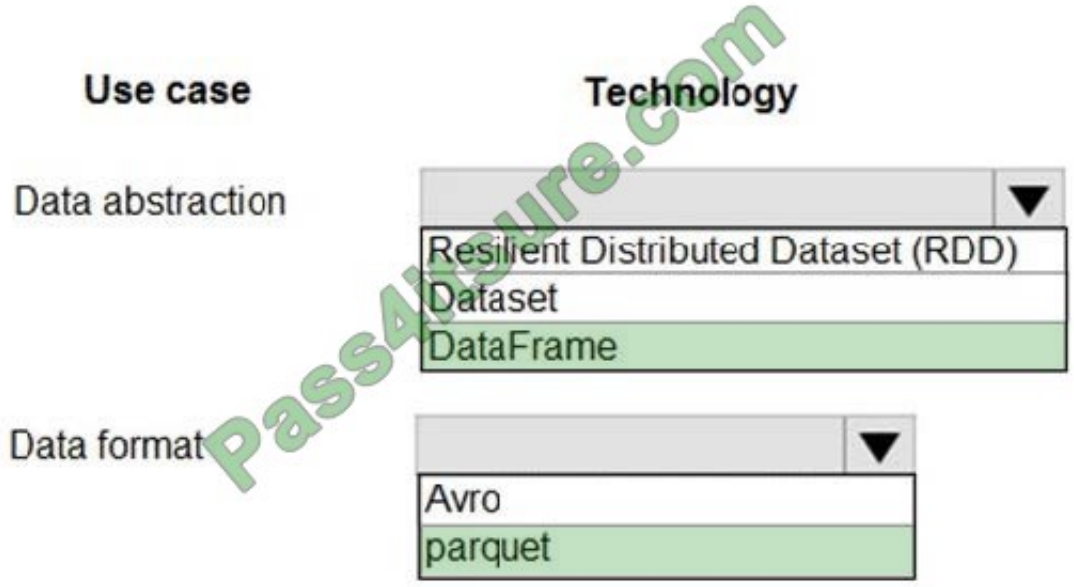
Hot Area:

Answer Area



Correct Answer:

Answer Area



Explanation:

Box 1: DataFrame

DataFrames

Best choice in most situations.

Provides query optimization through Catalyst.

Whole-stage code generation.

Direct memory access.

Low garbage collection (GC) overhead.

Not as developer-friendly as DataSets, as there are no compile-time checks or domain object programming.

Box 2: parquet

The best format for performance is parquet with snappy compression, which is the default in Spark 2.x. Parquet stores data in columnar format, and is highly optimized in Spark.

Incorrect Answers:

DataSets

Good in complex ETL pipelines where the performance impact is acceptable.

Not good in aggregations where the performance impact can be considerable.

RDDs

You do not need to use RDDs, unless you need to build a new custom RDD.

No query optimization through Catalyst.

No whole-stage code generation.

High GC overhead.

References:

https://docs.microsoft.com/en-us/azure/hdinsight/spark/apache-spark-perf

QUESTION 4

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while

others might not have a correct solution.

After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.

You are designing an HDInsight/Hadoop cluster solution that uses Azure Data Lake Gen1 Storage.

The solution requires POSIX permissions and enables diagnostics logging for auditing.

You need to recommend solutions that optimize storage.

Proposed Solution: Ensure that files stored are smaller than 250MB.

Does the solution meet the goal?

A. Yes

B. No

Correct Answer: B

Ensure that files stored are larger, not smaller than 250MB.

You can have a separate compaction job that combines these files into larger ones.

Note: The file POSIX permissions and auditing in Data Lake Storage Gen1 comes with an overhead that becomes apparent when working with numerous small files. As a best practice, you must batch your data into larger files versus writing

thousands or millions of small files to Data Lake Storage Gen1. Avoiding small file sizes can have multiple benefits, such as:

1.

Lowering the authentication checks across multiple files

2.

Reduced open file connections

3.

Faster copying/replication

4.

Fewer files to process when updating Data Lake Storage Gen1 POSIX permissions

References: https://docs.microsoft.com/en-us/azure/data-lake-store/data-lake-store-best-practices

QUESTION 5

Your company is an online retailer that can have more than 100 million orders during a 24-hour period, 95 percent of which are placed between 16:30 and 17:00. All the orders are in US dollars. The current product line contains the following three item categories:

1.

Games with 15,123 items

2.

Books with 35,312 items

3.

Pens with 6,234 items

You are designing an Azure Cosmos DB data solution for a collection named Orders Collection. The following documents is a typical order in Orders Collection.

```
"OrderTime": "16:35",
"id": " d0379ca2-f912-5h7f-k159-340ffa1z18e4"
"item": {
    "id": "08g17u57-1j58-6511-4x65-
    2qb5bf723u5s",
    "Title": "Living the Data Dream",
    "Category": "Books",
    "PurchasePrice": 12.56,
    "Currency": "USD"
}
```

Order Collection is expected to have a balanced read/write-intensive workload. Which partition key provides the most efficient throughput?

A. Item/Category

B. OrderTime

C. Item/Currency

D. Item/id

Correct Answer: A

Choose a partition key that has a wide range of values and access patterns that are evenly spread across logical partitions. This helps spread the data and the activity in your container across the set of logical partitions, so that resources for data storage and throughput can be distributed across the logical partitions.

Choose a partition key that spreads the workload evenly across all partitions and evenly over time. Your choice of partition key should balance the need for efficient partition queries and transactions against the goal of distributing items across multiple partitions to achieve scalability.

Candidates for partition keys might include properties that appear frequently as a filter in your queries. Queries can be efficiently routed by including the partition key in the filter predicate.

References: https://docs.microsoft.com/en-us/azure/cosmos-db/partitioning-overview#choose-partitionkey

To Read the Whole Q&As, please purchase the Complete Version from Our website.

Try our product !

100% Guaranteed Success
100% Money Back Guarantee
365 Days Free Update
Instant Download After Purchase
24x7 Customer Support
Average 99.9% Success Rate
More than 800,000 Satisfied Customers Worldwide
Multi-Platform capabilities - Windows, Mac, Android, iPhone, iPod, iPad, Kindle

We provide exam PDF and VCE of Cisco, Microsoft, IBM, CompTIA, Oracle and other IT Certifications.
You can view Vendor list of All Certification Exams offered:

https://www.pass4itsure.com/allproducts

Need Help

Please provide as much detail as possible so we can best assist you.
To update a previously submitted ticket: