



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.lead4pass.com/dp-201.html>

100% Passing Guarantee
100% Money Back Assurance

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

-  **Instant Download** After Purchase
-  **100% Money Back** Guarantee
-  **365 Days** Free Update
-  **800,000+** Satisfied Customers





QUESTION 1

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 2

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 Azure SQL Database that will use elastic pools. You plan to store data about customers in a table. Each record uses a value for CustomerID.

You need to recommend a strategy to partition data based on values in CustomerID.

Proposed Solution: Separate data into customer regions by using horizontal partitioning.

Does the solution meet the goal?

A. Yes

B. No

Correct Answer: B

We should use Horizontal Partitioning through Sharding, not divide through regions.

Note: Horizontal Partitioning - Sharding: Data is partitioned horizontally to distribute rows across a scaled out data tier. With this approach, the schema is identical on all participating databases. This approach is also called "sharding". Sharding can be performed and managed using (1) the elastic database tools libraries or (2) self-sharding. An elastic query is used to query or compile reports across many shards.

References: <https://docs.microsoft.com/en-us/azure/sql-database/sql-database-elastic-query-overview>

QUESTION 3

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 Azure SQL Database that will use elastic pools. You plan to store data about customers in a table. Each record uses a value for CustomerID.

You need to recommend a strategy to partition data based on values in CustomerID.

Proposed Solution: Separate data into customer regions by using vertical partitioning.

Does the solution meet the goal?

A. Yes

B. No

Correct Answer: B

Vertical partitioning is used for cross-database queries. Instead we should use Horizontal Partitioning, which also is called charding.

References:

<https://docs.microsoft.com/en-us/azure/sql-database/sql-database-elastic-query-overview>

QUESTION 4

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

Use case	Technology
Data abstraction	<div style="border: 1px solid black; padding: 2px;"><div style="background-color: #f0f0f0; padding: 2px; display: flex; justify-content: space-between; align-items: center;">▼</div><div style="padding: 2px;"><p>Resilient Distributed Dataset (RDD)</p><p>Dataset</p><p>DataFrame</p></div></div>
Data format	<div style="border: 1px solid black; padding: 2px;"><div style="background-color: #f0f0f0; padding: 2px; display: flex; justify-content: space-between; align-items: center;">▼</div><div style="padding: 2px;"><p>Avro</p><p>parquet</p></div></div>

Correct Answer:

Answer Area

Use case	Technology
Data abstraction	<div style="border: 1px solid black; padding: 2px;"><div style="background-color: #f0f0f0; padding: 2px; display: flex; justify-content: space-between; align-items: center;">▼</div><div style="padding: 2px;"><p>Resilient Distributed Dataset (RDD)</p><p>Dataset</p><p style="background-color: #d9ead3;">DataFrame</p></div></div>
Data format	<div style="border: 1px solid black; padding: 2px;"><div style="background-color: #f0f0f0; padding: 2px; display: flex; justify-content: space-between; align-items: center;">▼</div><div style="padding: 2px;"><p>Avro</p><p style="background-color: #d9ead3;">parquet</p></div></div>

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 5

You need to recommend a solution for storing customer data. What should you recommend?

- A. Azure SQL Data Warehouse
- B. Azure Stream Analytics
- C. Azure Databricks
- D. Azure SQL Database

Correct Answer: C

From the scenario:

Customer data must be analyzed using managed Spark clusters.



All cloud data must be encrypted at rest and in transit. The solution must support: parallel processing of customer data.

References:

<https://www.microsoft.com/developerblog/2019/01/18/running-parallel-apache-spark-notebook-workloads-on-azure-databricks/>

[Latest DP-201 Dumps](#)

[DP-201 VCE Dumps](#)

[DP-201 Study Guide](#)



To Read the [Whole Q&As](#), please purchase the [Complete Version](#) from [Our website](#).

Try our product !

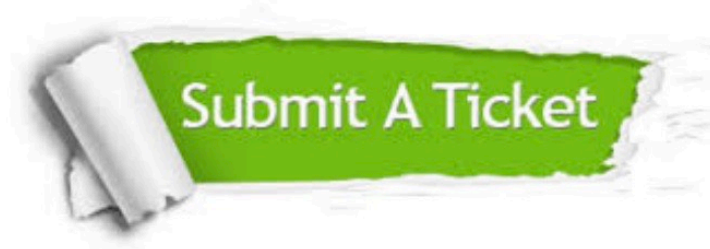
100% Guaranteed Success
100% Money Back Guarantee
365 Days Free Update
Instant Download After Purchase
24x7 Customer Support
Average 99.9% Success Rate
More than 800,000 Satisfied Customers Worldwide
Multi-Platform capabilities - [Windows](#), [Mac](#), [Android](#), [iPhone](#), [iPod](#), [iPad](#), [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.lead4pass.com/allproducts>

Need Help

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



 <p>One Year Free Update Free update is available within One Year after your purchase. After One Year, you will get 50% discounts for updating. And we are proud to boast a 24/7 efficient Customer Support system via Email.</p>	 <p>Money Back Guarantee To ensure that you are spending on quality products, we provide 100% money back guarantee for 30 days from the date of purchase.</p>	 <p>Security & Privacy We respect customer privacy. We use McAfee's security service to provide you with utmost security for your personal information & peace of mind.</p>
---	---	--

Any charges made through this site will appear as Global Simulators Limited.
All trademarks are the property of their respective owners.
Copyright © lead4pass, All Rights Reserved.