Course practice exam

Latest Submission Grade 90%

1. Question 1

HTAP (Hybrid Transactional/Analytical processing) enables business to run advanced analytics in near-real-time on data stored and processed in which of the following?

1 / 1 point

- ELT
- ETL
- OLTP
- OLAP

Correct

Online Transactional processing commonly referred to as OLTP systems work with operational data. HTAP enables business to run advanced analytics in near-real-time on data stored and processed by OLTP systems.

2. Question 2

Which of the following are common use cases for using Azure Synapse Link for Azure Cosmos DB?

Choose all that apply.

1 / 1 point

- [x] Real-time personalization

Correct

In retail, many web-based retailers will perform real-time basket analysis to make product recommendations to customers who are about to purchase products. This increased revenues for these organizations as the provided targeted suggestions at the point of sales.

- [x] IOT predictive maintenance

Correct

Industrial IOT innovations have drastically reduced downtimes of machinery and increased overall efficiency across all fields of industry. One of such innovations is predictive maintenance analytics for machinery at the edge of the cloud.

- [ ] Integration with APIs such as Gremlin API, Cassandra API, and Table API
- [x] Supply chain analytics, forecasting and reporting

Correct

Azure Synapse Link for Cosmos DB allows these organizations to store data from their sales systems, ingest real-time telemetry data from in vehicle systems and integrate date from their ERP systems into a common operational store in Azure Cosmos DB and then leverage the data from Synapse analytics to enable both predictive analytics scenarios such as stock out monitoring and supply chain bottleneck management.

3. Question 3

With Synapse Link, you can directly connect to your Azure Cosmos DB containers from Azure Synapse Analytics. Azure Synapse Analytics currently supports Synapse Link with which of the following?

Choose all that apply

1 / 1 point

- [ ] Dedicated SQL Pool
- [x] Serverless SQL pool

Correct

Azure Synapse Analytics supports Synapse Link with serverless SQL pools.

- [x] Synapse Apache Spark

Correct

Azure Synapse Analytics supports Synapse Link with Synapse Apache Spark.

4. Question 4

True or false?

In Azure Cosmos DB you can enable an analytical store on an existing container.

1 / 1 point

- (•) False
- ( ) True

Correct

Analytical store can only be enabled for new containers. To use analytical store for existing containers, migrate data from your existing containers to new containers using Azure Cosmos DB migration tools.

5. Question 5

Within Azure Synapse Link for Azure Cosmos DB, which of the following is a Column-oriented store optimized for analytical queries?

0 / 1 point

- ( ) Transactional store
- ( ) Cosmos DB store
- (•) Analytical store

Incorrect

Try going back and reviewing Design hybrid transactional and analytical processing using Azure Synapse Analytics.

6. Question 6

Currently once the Azure Synapse Link feature is enabled on an account you cannot disable it. Enabling Synapse Link on an account. When will Synapse Link begin to incur billing charges?

1 / 1 point

- As soon as containers are created
- As soon as Synapse Link is initially enabled
- As soon as the Analytical store is enabled, and containers are created
- As soon as the Analytical store is enabled

Correct

Enabling Synapse Link on the account has no billing implications until containers are created with the analytical store enabled.

7. Question 7

In Azure Cosmos DB enabling an analytical store is only available at the time of creating a container. How can you effectively disable an analytical store?

Choose all that apply.

1 / 1 point

- [x] Delete the container

Correct

Enabling analytical store is only available at the time of creating a container and cannot be completely disabled without deleting the container.

- [ ] From within the API disable the analytical store feature
- [x] Set the default TTL to 0 (Null)

Correct

Setting the default analytical store TTL value to 0 or null effectively disables the analytical store by no longer synchronize new items to it from the transactional store and deleting items already synchronized from the analytical store.

8. Question 8

Azure Cosmos DB automatically has system properties such as _ts, _self, _attachments, _rid, and _etag associated with every document. These system document properties are seldom useful for analytical store query purposes and can be removed using which of the following PySpark code snippet examples

1 / 1 point

system_document_properties = {'_attachments','_etag','_rid','_self','_ts'}

customer_columns = list(set(dfCustomer.columns) - system_document_properties)

customer_columns = list(set(dfCustomer.columns) = system_document_properties)

system_document_properties = {'_attachments','_etag','_rid','_self','_ts'}

system_document_properties - {'_attachments','_etag','_rid','_self','_ts'}

customer_columns = list(set(dfCustomer.columns) = system_document_properties)

Correct

This is the correct syntax to remove the system properties.

9. Question 9

The following snippet of code has been added to the first cell of a notebook.

1: dfCustomer = spark.read\

2: .format("cosmos.olap")\

3: .option("spark.synapse.linkedService", "AdventureWorksSQL")\

4: .option("spark.cosmos.container", "Customer")\

5: .load()

Assuming all the requirements to run this code are already in place, what is the purpose of the code on line number 2? **.format("cosmos.olap")**

1 / 1 point

indicates that we want to read from an Azure Cosmos DB analytical store

Format the contents of the DataFrame as aggregated values

Read from an Azure SQL Database.

Correct

cosmos.olap indicates that we are want to read from an Azure Cosmos DB analytical store.

10. Question 10

By default, the OPENJSON table-valued function returns three columns. These are?

1 / 1 point

Value

Correct

Value contains the value of the property.

- [x] Key

Correct

Key contains the name of the specified property or the index of the element in the specified array.

- [ ] Name
- [ ] Schema Name
- [x] Type

Correct

Contains the JSON type of the value. This is represented as an int value (from 0 to 5). This column is only returned when you use the default schema.

Course practice exam

Latest Submission Grade 85%

1. Question 1

How many drivers does a Cluster have?

1 / 1 point

- [x] Only one
- [ ] Configurable between one and eight
- [ ] Two, running in parallel.

Correct

A Cluster has one and only one driver.

2. Question 2

In Azure Databricks, what type of process are the driver and the executors?

1 / 1 point

- [ ] Python processes
- [ ] C++ processes
- [x] Java processes

Correct

The driver and the executors are Java processes.

3. Question 3

How do you list files in DBFS within a notebook?

1 / 1 point

- [ ] %fs dir /my-file-path
- [ ] ls /my-file-path
- [x] %fs ls /my-file-path

Correct

Correct. You added the file system magic to the cell before executing the ls command.

4. Question 4

We can read a CVS file when using a notebook and a spark session. Which of the following can be used to view the first couple of thousand characters of a file?

1 / 1 point

%fs dir /mnt/training/wikipedia/pageviews/

%fs ls /mnt/training/wikipedia/pageviews/

%fs head /mnt/training/wikipedia/pageviews/pageviews_by_second.tsv

Correct

We can use %fs head ... to view the first couple thousand characters of a file.

5. Question 5

How do you create a DataFrame object?

1 / 1 point

Introduce a variable name and equate it to something like myDataFrameDF =

Use the DF.create() syntax

Use the createDataFrame() function

Correct

This approach is the correct way to create DataFrame objects.

6. Question 6

Which of the following statements describes a wide transformation?

1 / 1 point

A wide transformation can be applied per partition/worker with no need to share or shuffle data to other workers.

A wide transformation requires sharing data across workers. It does so by shuffling data.

A wide transformation applies data transformation over a large number of columns.

Correct

Wide transformation shares data across workers by shuffling data between executors.

7. Question 7

Which feature of Spark of optimization is used in shuffling operations during wide transformations?

0 / 1 point

Tungsten Record Format

Catalyst Optimizer

Lazy Execution

Incorrect

Try going back and reviewing Data processing in Azure Databricks.

8. Question 8

Which of the listed methods for renaming a DataFrame's column are correct?

Select two.

1 / 1 point

- [ ] df.alias("timestamp", "dateCaptured")
- [x] df.toDF("dateCaptured")

Correct

This is a valid renaming method.

- [x] df.select(col("timestamp").alias("dateCaptured"))

Correct

This is a valid renaming method.

9. Question 9

You need to change a column name from “dob” to “DateOfBirth” on a spark DataFrame. Which of the following syntax is valid?

1 / 1 point

- (•) .withColumnRenamed("dob","DateOfBirth")
- ( ) .ColumnRename("dob","DateOfBirth")
- ( ) .RenameColumn("dob","DateOfBirth")

Correct

This is correct and will rename the column “dob” to “DateOfBirth”

10. Question 10

A Microsoft-managed Azure Databricks workspace virtual network (VNet) exists within the customer subscription. Information exchanged between this VNet and the Microsoft-managed Azure Databricks Control Plane VNet is sent over a secure TLS connection using which Ports?

Select two options.

1 / 1 point

- [ ] Port 53
- [x] Port 5557

Correct

The VNet and the Microsoft-managed Azure Databricks Control Plane VNet uses port 5557.

- [ ] Port 6667
- [ ] Port 443
- [x] Port 22

Correct

The VNet and the Microsoft-managed Azure Databricks Control Plane VNet uses port 22.

11. Question 11

You are starting to use Azure Databricks and you want to do specific network customizations, such as deploying Azure Databricks data plane resources in your own VNet. Which of the following will you configure?

1 / 1 point

- (•) VNet Injection
- ( ) VNet Peering
- ( ) You cannot create a custom configuration with VNets

Correct

If you're looking to do specific network customizations, you could deploy Azure Databricks data plane resources in your own VNet. In this scenario, instead of using the managed VNet, which restricts you from making changes, you "bring your own" VNet where you have full control.

12. Question 12

In which modes does Azure Databricks provide data encryption?

0 / 1 point

- ( ) In-transit only
- ( ) At-rest and in-transit
- ( ) At-rest only

Incorrect

You didn’t select an answer.

13. Question 13

What does Azure Data Lake Storage (ADLS) Passthrough enable?

1 / 1 point

- (•) Commands running on a configured cluster can read and write data in ADLS without configuring service principal credentials.

Automatically mounting ADLS accounts to the workspace that are added to the managed resource group.

User security groups that are added to ADLS are automatically created in the workspace as Databricks groups.

Correct

Correct. Also authentication to ADLS from Azure Databricks clusters is automatic, using the same Azure AD identity one uses to log into Azure Databricks.

14. Question 14

What is the Databricks Delta command to display metadata?

1 / 1 point

DESCRIBE DETAIL tableName

SHOW SCHEMA tablename

MSCK DETAIL tablename

Correct

You display metadata by using DESCRIBE DETAIL tableName.

15. Question 15

Which of the following can be used to successfully perform an UPSERT in a Delta dataset?

1 / 1 point

Use MERGE INTO my-table USING data-to-upsert

Use UPSERT INTO my-table /MERGE

Use UPSERT INTO my-table

Correct

That's the correct syntax to perform UPSERT in a Databricks.

16. Question 16

What is a lambda architecture and what does it try to solve?

1 / 1 point

An architecture that splits incoming data into two paths - a batch path and a streaming path. This architecture helps address the need to provide real-time processing in addition to slower batch computations.

An architecture that employs the latest Scala runtimes in one or more Databricks clusters to provide the most efficient data processing platform available today

An architecture that defines a data processing pipeline whereby microservices act as compute resources for efficient large-scale data processing

Correct

The lambda architecture is a big data processing architecture that combines both batch- and real-time processing methods.

17. Question 17

What happens if the command option("checkpointLocation", pointer-to-checkpoint directory) is not specified?

1 / 1 point

When the streaming job stops, all state around the streaming job is lost, and upon restart, the job must start from scratch

The streaming job will function as expected since the checkpointLocation option does not exist

It will not be possible to create more than one streaming query that uses the same streaming source since they will conflict

Correct

Setting the checkpointLocation is required for many sinks used in Structured Streaming. For those sinks where this setting is optional, keep in mind that when you do not set this value, you risk losing your place in the stream.

18. Question 18

What's the purpose of Activities in Azure Data Factory?

0 / 1 point

To link data stores or computer resources together for the movement of data between resources

To represent a data store or a compute resource that can host execution of an activity

To represent a processing step in a pipeline

Incorrect

Try going back and reviewing Analyze streaming data and create production workloads.

19. Question 19

What sort of pipeline is required in Azure DevOps for creating artifacts used in releases?

1 / 1 point

A Build pipeline

- A Release pipeline
- An Artifact pipeline

Correct

The output of a Build pipeline is one or more artifacts that can be used within release pipelines for automated deployments.

20. Question 20

In Azure Databricks you can deploy more than one Workspace. Best practice is to use the Hub and Spoke Model. Which of the following steps should be carried out to create a best practice Hub and Spoke Model in Azure Databricks?

1 / 1 point

- [x] Join the Workspace spokes with the central networking hub using VNet Peering

Correct

Best practice for Hub and Spoke is to join the Workspace spokes with the central networking hub using VNet Peering.

- [ ] Put all the common networking resources in a central hub VNet but excluding the custom DNS server.
- [x] Deploy each Workspace in its own VNet

Correct

Best practice for Hub and Spoke is to deploy each Workspace in its own VNet.

- [x] Put all the common networking resources in a central hub Vet, including the custom DNS server

Correct

Best practice for Hub and Spoke is to put all the common networking resources in a central hub Vet, including the custom DNS server.

- [ ] Join the Workspace spokes with the central networking hub using VNet Association
- [ ] Deploy each Workspace in the same VNet

Course practice exam

Latest Submission Grade 94.67%

1. Question 1

Azure Data Lake Storage Gen2 supports which of the following authorization mechanisms?

Select all that apply.

1 / 1 point

- [x] Role-based access control (Azure RBAC)

Correct

Feedback: Azure Data Lake Storage Gen2 supports Role-based access control.

- [x] Shared access signature (SAS) authorization

Correct

Feedback: Azure Data Lake Storage Gen2 supports Shared access signature (SAS).

- [x] Access control lists (ACL)

Correct

Feedback: Azure Data Lake Storage Gen2 supports Access control lists (ACL).

- [x] Shared Key authorization

Correct

Feedback: Azure Data Lake Storage Gen2 supports Shared Key authorization.

- [ ] Microsoft SQL Passthrough Authentication

2. Question 2

Azure Data Lake Storage Gen2 plays an important role in a wide range of big data architectures. There are four stages for processing big data solutions that are common to all architectures. From the following options, select the four stages.

Select four that apply.

1 / 1 point

- [x] Model and serve

Correct

Model and serve is one of the four stages for processing big data solutions that are common to all architectures.

- [x] Store

Correct

Store is one of the four stages for processing big data solutions that are common to all architectures.

- [ ] Transform Data
- [x] Prep and train

Correct

Prep and train is one of the four stages for processing big data solutions that are common to all architectures.

- [x] Ingestion

Correct

Ingestion is one of the four stages for processing big data solutions that are common to all architectures.

3. Question 3

Your company has a Data Lake Storage Gen2 account. If you want to upload a single file by using a tool that you don't have to install or configure, which tool should you use?

1 / 1 point

- (•) The Azure portal
- ( ) Azure Data Factory
- ( ) Azure Storage Explorer

Correct

The Azure portal requires no installation or configuration. To upload a file, you only have to sign in and a select an Upload button.

4. Question 4

In Azure Storage clients can use a shared key, or shared secret for authentication and to access resources. A Shared key is supported by which of the following storage models?

Select four that apply.

0.8 / 1 point

- [x] Blobs

Correct

Clients can use a shared key and it supports Blobs.

- [x] Queues

Correct

Clients can use a shared key and it supports queues.

- [x] Files

Correct

Clients can use a shared key and it supports files.

- [ ] Tables
- [ ] Disks

You didn't select all the correct answers

5. Question 5

You are required to grant access to a third-party app that will be uploading pictures to one of your Blob Stores. From a security perspective which of the following is recommended to delegate access and specify constraints such as permissions?

1 / 1 point

- ( ) Azure Active Directory
- (•) Shared Access Signature (SAS)
- ( ) Storage Account Key

Correct

As a best practice, for untrusted clients, use a shared access signature (SAS). A SAS is a string that contains a security token that can be attached to a URI. Use a SAS to delegate access to storage objects and specify constraints, such as the permissions and the time range of access.

6. Question 6

Security alerts are triggered when anomalies in activity occur. Azure provides various services including monitoring, recommendations, and remediation advice.

Which of the following Notification features are available with Azure for security alerts?

Select two that apply.

0.6666666666666666 / 1 point

- [x] Email Alerts

Correct

Security alerts are triggered when anomalies in activity occur. These security alerts are integrated with Azure Security Center, and are also sent via email to subscription administrators, with details of suspicious activity and recommendations on how to investigate and remediate threats.

- [x] Azure Security Center

Correct

Security alerts are triggered when anomalies in activity occur. These security alerts are integrated with Azure Security Center, and are also sent via email to subscription administrators, with details of suspicious activity and recommendations on how to investigate and remediate threats.

- [x] SMS messages

This should not be selected

Try going back and reviewing Secure your Azure Storage account

7. Question 7

Event processing pipelines provide an end-to-end solution for ingesting, transforming, and analyzing data streams. Which of the following components is responsible for the ingestion and transformation of streaming event data?

1 / 1 point

- ( ) An event producer
- (•) An event processor
- ( ) An event consumer

Correct

An event processor is responsible for the ingestion and transformation of streaming event data.

8. Question 8

Azure Event Hubs is a cloud-based, event-processing service that can receive and process millions of events per second. An entity that reads data from the Event Hubs is called what?

Select two that apply.

1 / 1 point

- [ ] Reader
- [x] Subscriber

Correct

Feedback: An entity that reads data from Event Hubs can be called a Subscriber.

- [ ] Publisher
- [x] Consumer

Correct

An entity that reads data from Event Hubs can be called a consumer.

9. Question 9

Azure Stream Analytics includes native support for five kinds of temporal windowing functions. Select the correct types of Windowing functions from the following list.

Select five that apply.

1 / 1 point

- [x] Hopping

Correct

Hopping is a valid temporal windowing function in Azure Stream Analytics.

- [x] Snapshot

Correct

Snapshot is a valid temporal windowing function in Azure Stream Analytics.

- [ ] EndTime

- [x] Sliding

Correct

Sliding is a valid temporal windowing function in Azure Stream Analytics.

- [ ] StartTime

- [x] Session

Correct

Session is a valid temporal windowing function in Azure Stream Analytics.

- [x] Tumbling

Correct

Tumbling is a valid temporal windowing function in Azure Stream Analytics.

10. Question 10

Which of the following functions generate events for points in time when the content of the window actually changed?

1 / 1 point

- ( ) Snapshot
- ( ) Hopping
- ( ) Tumbling
- ( ) Session
- (•) Sliding

Correct

Sliding windows generate events for points in time when the content of the window actually changed. To limit the number of windows it needs to consider, Azure Stream Analytics outputs events for only those points in time when an event entered or exited the window.