STORAGE AND DATA MANAGEMENT

Attempt	1
Marks Obtained	0 / 10
Your score is	0.0%

Completed on	Tuesday , 29 January 2019 , 02:30 PM
Time Taken	00 H 00 M 06 S
Result	Fail

Domains / Topics wise Quiz Performance Report

S.No.	Topic	Total Questions	Correct	Incorrect	Unattempted
1	Storage and Data Management	10	0	1	9

10 Questions	**0** Correct	**1** Incorrect	**9** Unattempted

Show Answers

All

QUESTION 1 INCORRECT STORAGE AND DATA MANAGEMENT

You are working as a SysOps Administrator for a health insurance company. Large amount of documents related to patient's health records are being uploaded to S3 buckets. These documents are uploaded on daily basis from 10 regional offices spread across the globe.

To further analyse this data, you are using AWS Athena along with AWS QuickSight. After initial one month of usage, you found that query performance is degraded. Also costing team has raised concerns about cost inflated above budget. You need to provide a solution to address both these concerns. Which of the following can be used to provide an effective solution? Select 2 correct options.

- [ ] **A.** When you initiate CREATE TABLE statement, use PARTITIONED BY to define keys by which source data is partition by data source identifier & date. ✔
- [x] **B.** After you initiate CREATE TABLE statement, use PARTITIONED BY to define keys by which source data is partition by data source identifier & date. ✖
- [x] **C.** Run CREATE TABLE AS SELECT (CTAS) queries in Athena and specify a data storage format as Parquet or ORC which will convert your existing raw data from other storage formats to columnar format. ✔
- [ ] **D.** Run CREATE TABLE AS SELECT (CTAS) queries in Athena and specify a data storage format as CSV or JSON which will convert your existing raw data from other storage formats to row storage format.

Explanation :

Correct Answer – A, C

Explanation: By partitioning data, you can restrict the amount of data scanned by each query, thus improving performance and reducing cost. Athena leverages Hive for partitioning data. You can partition data by any key. A common practice is to partition the data based on time, often leading to a multi-level partitioning scheme.

Apache Parquet and ORC are columnar storage formats that are optimized for fast retrieval of data and used in AWS analytical applications.

Columnar storage formats have the following characteristics that make them suitable for

using with Athena:

1. Compression by column, with compression algorithm selected for the column data type to save storage space in Amazon S3 and reduce disk space and I/O during query processing.
2. Predicate pushdown in Parquet and ORC enables Athena queries to fetch only the blocks it needs, improving query performance. When an Athena query obtains specific column values from your data, it uses statistics from data block predicates, such as max/min values, to determine whether to read or skip the block.
3. Splitting of data in Parquet and ORC allows Athena to split the reading of data to multiple readers and increase parallelism during its query processing.

Option B is incorrect as PARTITIONED BY needs to be defined during creation of table & not after Table is created.

Option D is incorrect as Row storage format does not improve query performance.

For more information on Partitioning & columnar-storage, refer to following URL's,

https://docs.aws.amazon.com/athena/latest/ug/partitions.html (https://docs.aws.amazon.com/athena/latest/ug/partitions.html)

https://docs.aws.amazon.com/athena/latest/ug/columnar-storage.html (https://docs.aws.amazon.com/athena/latest/ug/columnar-storage.html)

Ask our Experts

QUESTION 2 UNATTEMPTED STORAGE AND DATA MANAGEMENT

You are working for an IT firm. Project Team uploads details of ongoing projects in Amazon S3 buckets. These data consist of various formats like CSV, JSON & Apache Parquet. Top management is looking for interactive dashboards for this data for project review meeting.

This data must be analysed quickly & you do not have any additional budget for setting new servers to analyse this data. Which of the following is best option to meet this requirement?

- A. Amazon Redshift
- B. Amazon EMR
- C. Amazon Athena with Amazon QuickSight. ✔

○ **D.** Amazon Athena with Amazon Glue.

Explanation :

Correct Answer – C

Amazon Athena helps you analyse data stored in Amazon S3. You can use Athena to run ad-hoc queries using ANSI SQL, without the need to aggregate or load the data into Athena. Amazon Athena can process unstructured, semi-structured, and structured data sets. Examples include CSV, JSON, Avro or columnar data formats such as Apache Parquet and Apache ORC. Amazon Athena provides queries for data in S3 without the need to setup or manage any servers. Amazon Athena is ideal for quick, ad-hoc querying and integrates with Amazon QuickSight for easy visualization, it can also handle complex analysis, including large joins, window functions, and arrays.

Option A is incorrect as Amazon Redshift is used when you need to pull together data from many different sources – like inventory systems, financial systems, and retail sales systems – into a common format, and store it for long periods of time, to build sophisticated business reports from historical data. Amazon Redshift provides the fastest query performance for enterprise reporting and business intelligence workloads, particularly those involving extremely complex SQL with multiple joins and sub-queries. For above requirement, it would be additional cost & admin work to setup Amazon Redshift.

Option B is incorrect as For above requirement, Amazon EMR will be costlier as compare to Amazon Athena with Amazon QuickSight.Amazon EMR makes it simple and cost effective to run highly distributed processing frameworks such as Hadoop, Spark, and Presto when compared to on-premises deployments. Amazon EMR is flexible - you can run custom applications and code, and define specific compute, memory, storage, and application parameters to optimize your analytic requirements.

Option D is incorrect as Amazon Glue is ETL (extract, transform & load) service which will load data for analytics. In above requirement Amazon Athena can analyse data from S3, & along with Amazon QuickSight can produce interactive dashboards.

For more information on Amazon Athena & other big data services, refer to following URL, https://aws.amazon.com/athena/faqs/ (https://aws.amazon.com/athena/faqs/)

Ask our Experts

You are working as SysOps Administrator for a media firm. They have implemented hybrid solution where in 3 regional office in Singapore connect to ap-southeast-1 AZ via 10 Gb Direct Connect links.

Employees from these offices upload a large number of video & audio files to S3 buckets which amounts to approximate 2 TB in size. As per latest guidelines to enhance security, all these files need to be encrypted prior to upload to S3 buckets. CTO is looking for a fully managed & cost-effective solution. Which of the following is best option to meet this requirement?

- ○ **A.** Use third party encryption tool on each employee desktop machine which would encrypt data files & then upload all the encrypted files to Amazon S3.
- ○ **B.** Use Hardware Security Module (HSM) appliance on-premise server with sufficient storage to temporarily store, encrypt & then upload all the encrypted files to Amazon S3.
- ○ **C.** Use AWS KMS managed customer master key to enable client-side data encryption, & then to upload encrypted files to S3. ✔
- ○ **D.** Use AWS CloudHSM appliance to manage encryption keys which will encrypt data files & then upload all the encrypted files to Amazon S3.

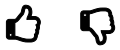
Explanation :

Correct Answer – C
Client-side encryption is the act of encrypting data before sending it to Amazon S3. To enable client-side encryption, you have the following options, use an AWS KMS-managed customer master keys or use a client-side master key.
When using an AWS KMS-managed customer master key to enable client-side data encryption, you provide an AWS KMS customer master key ID (CMK ID).
When uploading an object—Using the CMK ID, the client first sends a request to the AWS Key Management Service (AWS KMS) for a key that it can use to encrypt your object data. AWS KMS returns two versions of a randomly generated data encryption key:

- A plain-text version that the client uses to encrypt the object data
- A cipher blob of the same data encryption key that the client uploads to Amazon S3 as object metadata.

Option A is incorrect as using Third party encryption tool will be costlier option to be deployed on each employee's desktop.
Option B is incorrect as using Hardware Security Module (HSM) will be costly & also since large amount of data is to be uploaded daily, it will need huge temporary storage space.
Option D is incorrect as AWS CloudHSM is not managed solution. AWS will provision CloudHSM device in customer defined VPC environment, but client will have to manage keys residing in the HSM, its lifecycle processes, cryptographic operations and performance management.
For more information on client-side data encryption using AWS KMS, refer to following URL, https://docs.aws.amazon.com/AmazonS3/latest/dev/UsingClientSideEncryption.html (https://docs.aws.amazon.com/AmazonS3/latest/dev/UsingClientSideEncryption.html)

Ask our Experts



QUESTION 4 UNATTEMPTED STORAGE AND DATA MANAGEMENT

You are working as SysOps architect for a financial firm. Firm is coming up with a mobile wallet application which provides client ability to directly transact with multiple bank accounts using Credit cards. For this client credit card & bank details will be stored in encrypted format in AWS cloud infrastructure.

To meet compliance guidelines of Key management process, there are stringent pre-requisite given by Security Team. They are looking for a single tenant key storage with support to asymmetric key integration. Since this is a financial application any downtime will generate a huge financial loss. Which of the following is cost-effective option to meet these prerequisites?

- **A.** Use third party encryption tool on application load balancers which will provide high availability.
- **B.** Use Hardware Security Module (HSM) appliance in multiple locations.
- **C.** Use AWS KMS Server side management.

○ **D.** Use AWS CloudHSM appliance. ✔

Explanation :

Correct Answer – D
AWS CloudHSM provides single tenant key storage & support to asymmetric keys like RSA, ECC.
Also, CloudHSM can be deployed in multiple AZ's to provide redundancy & high availability for key management. AWS will provision CloudHSM device in customer defined VPC environment, but client will have to manage keys residing in the HSM, its lifecycle processes, cryptographic operations and performance management.
Option A is incorrect as Third-party encryption tool will be costlier to implement & deploying redundancy with this will be additional management overhead.
Option B is incorrect as using Hardware Security Module (HSM) will be costlier to implement & deploying redundancy with this will be additional management overhead.
Option C is incorrect as AWS KMS is a multi-tenant key storage & supports only symmetric key encryption.
For more information on using AWS Cloud HSM, refer to following URL,
https://aws.amazon.com/blogs/security/aws-cloudhsm-use-cases-part-one-of-the-aws-cloudhsm-series (https://aws.amazon.com/blogs/security/aws-cloudhsm-use-cases-part-one-of-the-aws-cloudhsm-series)

Ask our Experts

QUESTION 5 UNATTEMPTED STORAGE AND DATA MANAGEMENT

You are working as SysOps Administrator for a financial firm. As per guidelines from legal team you need to save all customer transactions for seven years for compliance & audit purpose. You created a vault for storing archives in S3 Glacier.

You also need to ensure that no changes or deletion is made to these archives for a period of seven years but need to ensure that files can be access multiple times for read purpose. Which of the following policy can be enforced to meet this requirement?

- A. Vault Access Policy
- B. S3 Bucket policy
- C. Glacier Control Policy
- D. Vault Lock Policy ✔

Explanation :

Correct Answer – D
A vault lock policy can be locked to prevent future changes, providing strong enforcement for your compliance controls. You can use the vault lock policy to deploy regulatory and compliance controls, which typically require tight controls on data access.
Option A is incorrect as vault access policy is used to implement access controls that are not compliance related, temporary, and subject to frequent modification.
Option B is incorrect as S3 bucket policy is used to grant permission to your Amazon S3 resources.
Option C is incorrect as there is nothing as Glacier control policy.
For more information on Vault Access Policy & Vault lock Policy, refer to following URL, https://docs.aws.amazon.com/amazonglacier/latest/dev/access-control-resource-based.html (https://docs.aws.amazon.com/amazonglacier/latest/dev/access-control-resource-based.html)

Ask our Experts

QUESTION 6 UNATTEMPTED STORAGE AND DATA MANAGEMENT

You are working as a consultant for a start-up firm. Each year external auditors shared an Audit report on firm's financial statement. As per guidelines from legal team, firm need to save this Audit reports for a period of five years with no amendments of this report by any user.

These reports are archived in vaults of Amazon S3 Glacier. To meet guidelines, from legal team, you need to advice cloud operations team to setup vault lock. Which of following are steps to be followed to lock vault with Glacier API? Choose 2 valid options.

- [x] **A.** Initiate Vault Lock with a vault lock policy. ✔
- [ ] **B.** Initiate Vault Lock with a vault access policy.
- [x] **C.** Complete Vault Lock within 24 hours after Initiate Vault Lock. ✔
- [ ] **D.** Complete Vault Lock within 4 hours after Initiate Vault Lock.
- [ ] **E.** Complete Vault Lock after 24 hours after Initiate Vault Lock.

Explanation :

Correct Answer – A, C

To lock your vault with the Glacier API, you first call Initiate Vault Lock (POST lock-policy) with a vault lock policy that specifies the controls you want to deploy. Initiate Vault Lock (POST lock-policy) attaches the policy to your vault, transitions the vault lock to the in-progress state, and returns a unique lock ID. After the vault lock enters the in-progress state, you have 24 hours to complete the lock by calling Complete Vault Lock (POST lockId) with the lock ID returned from Initiate Vault Lock (POST lock-policy). After the vault is locked it cannot be unlocked.

If you don't complete the vault lock process within 24 hours after entering the in-progress state, your vault automatically exits the in-progress state, and the vault lock policy is removed. You can call Initiate Vault Lock (POST lock-policy) again to install a new vault lock policy and transition into the in-progress state.

Option B is incorrect, as you will need to initiate vault lock with vault lock policy & not vault access policy. A vault lock policy can be locked to prevent future changes, providing strong enforcement for your compliance controls. A vault access policy is used to implement access controls that are not compliance related, temporary, and subject to frequent modification.

Option D is incorrect as Complete Vault Lock (POST lockId) needs to be completed within 24 hours of Initiate Vault Lock. You can initiate Complete Vault Lock within 4 hours, but you have 24 hours to validate your vault lock policy before the lock ID expires.

Option E is incorrect as Complete Vault Lock (POST lockId) needs to be completed within 24 hours & not after 24 hours of Initiate Vault Lock. If you don't complete the vault lock process within 24 hours after entering the in-progress state, your vault automatically exits the in-progress state, and the vault lock policy is removed.

For more information on steps to initiate vault lock, refer to following URL,
https://docs.aws.amazon.com/amazonglacier/latest/dev/vault-lock-how-to-api.html
(https://docs.aws.amazon.com/amazonglacier/latest/dev/vault-lock-how-to-api.html)

Ask our Experts

QUESTION 7 UNATTEMPTED STORAGE AND DATA MANAGEMENT

"s3.amazonaws.com/mybucket/photos/2014/08/puppy.jpg?
x-user=johndoe" is GET URL you have received for accessing your S3 bucket via REST API.What is significance of x-user=johndoe in this URL?

- **A.** Amazon S3 will notify user johndoe once this image is downloaded while capturing these parameters in access log records.
- **B.** Amazon S3 ignores query-string parameters that begin with “x- “, & none of these parameters are included in access log records.
- **C.** Amazon S3 ignores query-string parameters that begin with “x- “, & these parameters are included in access log records. ✔
- **D.** Amazon S3 will allow access based upon user id in query-string parameters that begin with “x- “, & once authorised, this detail is captured in access log records.

Explanation :

Correct Answer – C
You can include custom information to be stored in the access log record for a request by adding a custom query-string parameter to the URL for the request. Amazon S3 ignores query-string parameters that begin with "x-" but includes those parameters in the access log record for the request, as part of the Request-URI field of the log record.
Option A is incorrect as S3 ignore query-string parameter starting with “x- “& does not notify user based upon query-string parameters.
Option B is incorrect as Johndoe will be included as a custom information in access log record.

Option D is incorrect as S3 ignore query-string parameter starting with "x- ".
For more information on log format, refer to following URL;
https://docs.aws.amazon.com/AmazonS3/latest/dev/LogFormat.html
(https://docs.aws.amazon.com/AmazonS3/latest/dev/LogFormat.html)

Ask our Experts

QUESTION 8 UNATTEMPTED STORAGE AND DATA MANAGEMENT

You are working as SysOps Administrator for a global construction company. Team members upload project plans in S3 bucket. For this, you have setup multiple S3 buckets in us-west-1, ap-south-1 & sa-east-1 regions.

Additionally, you need to configure Access logging to provide detailed records for requests made to these buckets. You have setup a new S3 bucket name as "logs-gcc" in us-west-1 region to collect access logs. For us-west-1 region you can select S3 bucket "logs-gcc" as target bucket but not for other 2 regions i.e. ap-south-1 & sa-east-1. What could be possible reason?

- **A.** S3 Server Access logging feature is currently available in US-West-1 region only.
- **B.** Source & target buckets should be in same region. ✔
- **C.** While creating bucket "logs-gcc", specify a prefix for other 2 regions.
- **D.** Amazon S3 Log Delivery group in ap-south-1 & sa-east-1 regions do not have required permission to access bucket "logs-gcc".

Explanation :

Correct Answer – B
When logging is enabled, logs are saved to a bucket in the same AWS Region as the source bucket.
To enable access logging, you must do the following:

• Turn on the log delivery by adding logging configuration on the bucket for which you want Amazon S3 to deliver access logs. We refer to this bucket as the source bucket.
• Grant the Amazon S3 Log Delivery group write permission on the bucket where you want the access logs saved. We refer to this bucket as the target bucket.
To turn on log delivery, you provide the following logging configuration information:
•The name of the target bucket where you want Amazon S3 to save the access logs as objects. You can have logs delivered to any bucket that you own that is in the same Region as the source bucket, including the source bucket itself.
Option A is incorrect as S3 Server Access Logging feature is available in all regions. Its disabled by default & need to enable for S3 buckets.
Option C is incorrect. Its best practice to specify a prefix for all log object keys so that the object names begin with a common string and the log objects are easier to identify. In this case specifying a prefix for different regions is not a valid option as bucket in US-West-1 region will not be collecting access logs from other regions. Source & Target buckets should be in same region.
Option D is incorrect as for access logs collection bucket should be in same region. After creating buckets in each region, you will need to grant the Amazon S3 Log Delivery group write permission on the new buckets where you want the access logs saved.
For more information on source & target buckets, refer to following URL,
https://docs.aws.amazon.com/AmazonS3/latest/dev/ServerLogs.html
(https://docs.aws.amazon.com/AmazonS3/latest/dev/ServerLogs.html)

Ask our Experts

QUESTION 9 UNATTEMPTED STORAGE AND DATA MANAGEMENT

You are working as consultant for a construction firm.The Architecture team save plans for new construction works in an on-premise storage disk. Due to expansion & new project works in last quarter, on-premise storage is getting exhausted.

Firm has very limited budget for storage device & not keen to invest further in storage. Upon further analysis you found that the Architecture team usually works only on recent plans. CTO has enquired you to find a best possible storage option. Which of the following will meet the requirement?

- A. Cached Volumes Gateway ✔
- B. File Gateway
- C. Stored Volumes Gateway
- D. Tape Gateways

Explanation :

Correct Answer – A
You store your data in Amazon Simple Storage Service (Amazon S3) and retain a copy of frequently accessed data subsets locally. Cached volumes offer a substantial cost savings on primary storage and minimize the need to scale your storage on-premises. You also retain low-latency access to your frequently accessed data.
Option B is incorrect as a file gateway supports a file interface into Amazon Simple Storage Service (Amazon S3) and combines a service and a virtual software appliance. In above scenario, this would not meet requirement.
Option C is incorrect as client is looking for saving additional storage space. With Stored Volume Gateway, entire dataset is locally stored on premise gateway while asynchronously back up point-in-time snapshots of this data to Amazon S3.
Option D is incorrect as tape gateway provides a virtual tape infrastructure that scales seamlessly with your business needs and eliminates the operational burden of provisioning, scaling, and maintaining a physical tape infrastructure. In above scenario, this would not meet requirement.
For more information on Storage Gateway, refer to following URL,
https://docs.aws.amazon.com/storagegateway/latest/userguide/StorageGatewayConcepts.html
(https://docs.aws.amazon.com/storagegateway/latest/userguide/StorageGatewayConcepts.html)

Ask our Experts

QUESTION 10 UNATTEMPTED STORAGE AND DATA MANAGEMENT

You are working with an IT firm. They are using File Gateway to store all project documents to S3 buckets using NFS protocol. All these files are encrypted using Amazon KMS.

As a part of consolidation work, you need to update files to buckets created by third party vendor. For this, File Gateway need to access S3 buckets created by third party vendors. Third party vendor has agreed to provide full access to S3 buckets which they own. What additional settings need to be done? Select 2 correct options.

- [ ] **A.** Make sure that the IAM role that your file share uses to access the S3 bucket includes permissions for operations such as s3:GetObjectAcl and s3:PutObjectAcl & a trust policy which allows your account to assume IAM role. ✔
- [ ] **B.** Make sure that the IAM role that your file share uses to access the S3 bucket includes permissions for operations such as s3:GetObject and s3:PutObject & a trust policy which allows your account to assume IAM role.
- [ ] **C.** Disable Give bucket owner full control in the Object metadata settings in the Configure file share setting dialog box.
- [ ] **D.** Enable Give bucket owner full control in the Object metadata settings in the Configure file share setting dialog box. ✔

Explanation :

Correct Answer – A, D

Cross-account access is when an AWS account and users for that account are granted access to resources that belong to another AWS account. With file gateways, you can use a file share in one AWS account to access objects in an Amazon S3 bucket that belongs to a different AWS account.

To use a file share owned by one AWS account to access an S3 bucket in a different AWS account

1. Make sure that the S3 bucket owner has granted your AWS account access to the S3 bucket that you need to access and the objects in that bucket. For information about how to grant this access, see Example 2: Bucket Owner Granting Cross-Account Bucket Permissions in the Amazon Simple Storage Service Developer Guide. For a list of the required permissions, see Granting Access to an Amazon S3 Bucket.
2. Make sure that the IAM role that your file share uses to access the S3 bucket includes permissions for operations such as s3:GetObjectAcl and s3:PutObjectAcl. In addition, make sure that the IAM role includes a trust policy that allows your account to assume that IAM role. For an example of such a trust policy, see Granting Access to an Amazon S3 Bucket.

3. If your file share uses an existing role to access the S3 bucket, you should include permissions for s3:GetObjectAcl and s3:PutObjectAcl operations. The role also needs a trust policy that allows your account to assume this role. For an example of such a trust policy, see Granting Access to an Amazon S3 Bucket.
4. Choose Give bucket owner full control in the Object metadata settings in the Configure file share setting dialog box.

Option B is incorrect File share needs to include permissions for operations such as s3:GetObjectAcl and s3:PutObjectAcl apart from s3:GetObject and s3:PutObject.
Option C is incorrect as Bucket owner will require full control in Object Metadata settings.
For more information on using a File Share for Cross-Account Access, refer to following URL,
https://docs.aws.amazon.com/storagegateway/latest/userguide/managing-gateway-file.html#cross-account-access
(https://docs.aws.amazon.com/storagegateway/latest/userguide/managing-gateway-file.html#cross-account-access)

Ask our Experts

Finish Review (https://www.whizlabs.com/learn/course/aws-csyopaa-practice-tests/quiz/14895)

- Big Data Certification (https://www.whizlabs.com/big-data-certifications/)

Mobile App

- Android Coming Soon
- iOS Coming Soon

Follow us

(https://www.facebook.com/whizlabs.software/)

(https://in.linkedin.com/company/whizlabs-software)

(https://twitter.com/whizlabs?lang=en)

(https://plus.google.com/+WhizlabsSoftware)