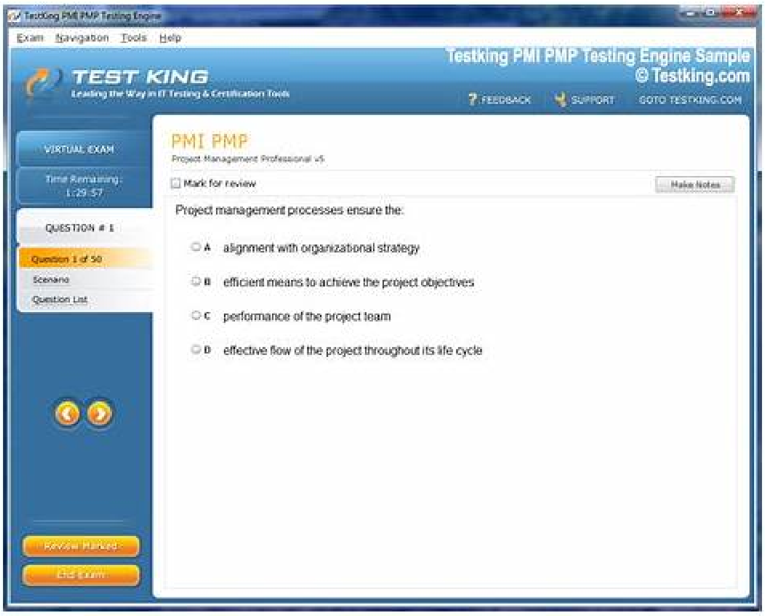
Data-Engineer-Associate Latest Test Testking & Valid Data-Engineer-Associate Exam Dumps



P.S. Free 2026 Amazon Data-Engineer-Associate dumps are available on Google Drive shared by Exam-Killer: https://drive.google.com/open?id=1cH5dynzaY2IFs6yhZ_3j94R4WMafRJnm

Data-Engineer-Associate exam dumps provided by Exam-Killer are tested through practice, and are the most correct and the newest practical Data-Engineer-Associate test dumps. Our Exam-Killer can provide accurate Data-Engineer-Associate certification training questions based on extensive research and the experience of real world to make you pass Data-Engineer-Associate Certification Exam in a short time. If you purchase our Data-Engineer-Associate exam dumps, we will offer free update service within one year.

Exam-Killer free update our training materials, which means you will always get the latest Data-Engineer-Associate exam training materials. If Data-Engineer-Associate exam objectives change, The learning materials Exam-Killer provided will follow the change. Exam-Killer know the needs of each candidate, we will help you through your Data-Engineer-Associate Exam Certification. We help each candidate to pass the exam with best price and highest quality.

>> Data-Engineer-Associate Latest Test Testking <<

Comprehensive Amazon Data-Engineer-Associate Exam Questions in PDF Format

As long as you get to know our Data-Engineer-Associate exam questions, you will figure out that we have set an easier operation system for our candidates. Once you have a try, you can feel that the natural and seamless user interfaces of our Data-Engineer-Associate study materials have grown to be more fluent and we have revised and updated Data-Engineer-Associate learning guide according to the latest development situation. In the guidance of teaching syllabus as well as theory and practice, our Data-Engineer-Associate training engine has achieved high-quality exam materials according to the tendency in the industry.

Amazon AWS Certified Data Engineer - Associate (DEA-C01) Sample Questions (Q127-Q132):

NEW QUESTION # 127
A company uses AWS Step Functions to orchestrate a data pipeline. The pipeline consists of Amazon EMR jobs that ingest data from data sources and store the data in an Amazon S3 bucket. The pipeline also includes EMR jobs that load the data to Amazon Redshift.
The company's cloud infrastructure team manually built a Step Functions state machine. The cloud infrastructure team launched an

EMR cluster into a VPC to support the EMR jobs. However, the deployed Step Functions state machine is not able to run the EMR jobs.
Which combination of steps should the company take to identify the reason the Step Functions state machine is not able to run the EMR jobs? (Choose two.)

- A. Verify that the Step Functions state machine code has all IAM permissions that are necessary to create and run the EMR jobs. Verify that the Step Functions state machine code also includes IAM permissions to access the Amazon S3 buckets that the EMR jobs use. Use Access Analyzer for S3 to check the S3 access properties.
- B. Check the retry scenarios that the company configured for the EMR jobs. Increase the number of seconds in the interval between each EMR task. Validate that each fallback state has the appropriate catch for each decision state. Configure an Amazon Simple Notification Service (Amazon SNS) topic to store the error messages.
- C. Use AWS CloudFormation to automate the Step Functions state machine deployment. Create a step to pause the state machine during the EMR jobs that fail. Configure the step to wait for a human user to send approval through an email message. Include details of the EMR task in the email message for further analysis.
- D. Check for entries in Amazon CloudWatch for the newly created EMR cluster. Change the AWS Step Functions state machine code to use Amazon EMR on EKS. Change the IAM access policies and the security group configuration for the Step Functions state machine code to reflect inclusion of Amazon Elastic Kubernetes Service (Amazon EKS).
- E. Query the flow logs for the VPC. Determine whether the traffic that originates from the EMR cluster can successfully reach the data providers. Determine whether any security group that might be attached to the Amazon EMR cluster allows connections to the data source servers on the informed ports.

Answer: A,E

Explanation:
To identify the reason why the Step Functions state machine is not able to run the EMR jobs, the company should take the following steps:
Verify that the Step Functions state machine code has all IAM permissions that are necessary to create and run the EMR jobs. The state machine code should have an IAM role that allows it to invoke the EMR APIs, such as RunJobFlow, AddJobFlowSteps, and DescribeStep. The state machine code should also have IAM permissions to access the Amazon S3 buckets that the EMR jobs use as input and output locations. The company can use Access Analyzer for S3 to check the access policies and permissions of the S3 buckets12. Therefore, option B is correct.
Query the flow logs for the VPC. The flow logs can provide information about the network traffic to and from the EMR cluster that is launched in the VPC. The company can use the flow logs to determine whether the traffic that originates from the EMR cluster can successfully reach the data providers, such as Amazon RDS, Amazon Redshift, or other external sources. The company can also determine whether any security group that might be attached to the EMR cluster allows connections to the data source servers on the informed ports. The company can use Amazon VPC Flow Logs or Amazon CloudWatch Logs Insights to query the flow logs3 . Therefore, option D is correct.
Option A is incorrect because it suggests using AWS CloudFormation to automate the Step Functions state machine deployment. While this is a good practice to ensure consistency and repeatability of the deployment, it does not help to identify the reasonwhy the state machine is not able to run the EMR jobs. Moreover, creating a step to pause the state machine during the EMR jobs that fail and wait for a human user to send approval through an email message is not a reliable way to troubleshoot the issue. The company should use the Step Functions console or API to monitor the execution history and status of the state machine, and use Amazon CloudWatch to view the logs and metrics of the EMR jobs .
Option C is incorrect because it suggests changing the AWS Step Functions state machine code to use Amazon EMR on EKS. Amazon EMR on EKS is a service that allows you to run EMR jobs on Amazon Elastic Kubernetes Service (Amazon EKS) clusters. While this service has some benefits, such as lower cost and faster execution time, it does not support all the features and integrations that EMR on EC2 does, such as EMR Notebooks, EMR Studio, and EMRFS. Therefore, changing the state machine code to use EMR on EKS may not be compatible with the existing data pipeline and may introduce new issues.
Option E is incorrect because it suggests checking the retry scenarios that the company configured for the EMR jobs. While this is a good practice to handle transient failures and errors, it does not help to identify the root cause of why the state machine is not able to run the EMR jobs. Moreover, increasing the number of seconds in the interval between each EMR task may not improve the success rate of the jobs, and may increase the execution time and cost of the state machine. Configuring an Amazon SNS topic to store the error messages may help to notify the company of any failures, but it does not provide enough information to troubleshoot the issue.
References:

1: Manage an Amazon EMR Job - AWS Step Functions
2: Access Analyzer for S3 - Amazon Simple Storage Service
3: Working with Amazon EMR and VPC Flow Logs - Amazon EMR
[4]: Analyzing VPC Flow Logs with Amazon CloudWatch Logs Insights - Amazon Virtual Private Cloud
[5]: Monitor AWS Step Functions - AWS Step Functions
[6]: Monitor Amazon EMR clusters - Amazon EMR

[7]: Amazon EMR on Amazon EKS - Amazon EMR

NEW QUESTION # 128

A company is building a data lake for a new analytics team. The company is using Amazon S3 for storage and Amazon Athena for query analysis. All data that is in Amazon S3 is in Apache Parquet format.
The company is running a new Oracle database as a source system in the company's data center. The company has 70 tables in the Oracle database. All the tables have primary keys. Data can occasionally change in the source system. The company wants to ingest the tables every day into the data lake.
Which solution will meet this requirement with the LEAST effort?

- A. Create an AWS Database Migration Service (AWS DMS) task for ongoing replication. Set the Oracle database as the source. Set Amazon S3 as the target. Configure the task to write the data in Parquet format.
- B. Create an AWS Glue connection to the Oracle database. Create an AWS Glue bookmark job to ingest the data incrementally and to write the data to Amazon S3 in Parquet format.
- C. Create an Apache Sqoop job in Amazon EMR to read the data from the Oracle database. Configure the Sqoop job to write the data to Amazon S3 in Parquet format.
- D. Create an Oracle database in Amazon RDS. Use AWS Database Migration Service (AWS DMS) to migrate the on-premises Oracle database to Amazon RDS. Configure triggers on the tables to invoke AWS Lambda functions to write changed records to Amazon S3 in Parquet format.

Answer: A

Explanation:
The company needs to ingest tables from an on-premises Oracle database into a data lake on Amazon S3 in Apache Parquet format. The most efficient solution, requiring the least manual effort, would be to use AWS Database Migration Service (DMS) for continuous data replication.
* Option C: Create an AWS Database Migration Service (AWS DMS) task for ongoing replication.
Set the Oracle database as the source. Set Amazon S3 as the target. Configure the task to write the data in Parquet format.AWS DMS can continuously replicate data from the Oracle database into Amazon S3, transforming it into Parquet format as it ingests the data. DMS simplifies the process by providing ongoing replication with minimal setup, and it automatically handles the conversion to Parquet format without requiring manual transformations or separate jobs. This option is the least effort solution since it automates both the ingestion and transformation processes.
Other options:
* Option A (Apache Sqoop on EMR) involves more manual configuration and management, including setting up EMR clusters and writing Sqoop jobs.
* Option B (AWS Glue bookmark job) involves configuring Glue jobs, which adds complexity. While Glue supports data transformations, DMS offers a more seamless solution for database replication.
* Option D (RDS and Lambda triggers) introduces unnecessary complexity by involving RDS and Lambda for a task that DMS can handle more efficiently.
References:
* AWS Database Migration Service (DMS)
* DMS S3 Target Documentation

NEW QUESTION # 129

A company stores datasets in JSON format and .csv format in an Amazon S3 bucket. The company has Amazon RDS for Microsoft SQL Server databases, Amazon DynamoDB tables that are in provisioned capacity mode, and an Amazon Redshift cluster. A data engineering team must develop a solution that will give data scientists the ability to query all data sources by using syntax similar to SQL.
Which solution will meet these requirements with the LEAST operational overhead?

- A. Use AWS Glue to crawl the data sources. Store metadata in the AWS Glue Data Catalog. Use AWS Glue jobs to transform data that is in JSON format to Apache Parquet or .csv format. Store the transformed data in an S3 bucket. Use Amazon Athena to query the original and transformed data from the S3 bucket.
- B. Use AWS Lake Formation to create a data lake. Use Lake Formation jobs to transform the data from all data sources to Apache Parquet format. Store the transformed data in an S3 bucket. Use Amazon Athena or Redshift Spectrum to query the data.
- C. Use AWS Glue to crawl the data sources. Store metadata in the AWS Glue Data Catalog. Use Amazon Athena to query the data. Use SQL for structured data sources. Use PartiQL for data that is stored in JSON format.

- D. Use AWS Glue to crawl the data sources. Store metadata in the AWS Glue Data Catalog. Use Redshift Spectrum to query the data. Use SQL for structured data sources. Use PartiQL for data that is stored in JSON format.

Answer: C

Explanation:
The best solution to meet the requirements of giving data scientists the ability to query all data sources by using syntax similar to SQL with the least operational overhead is to use AWS Glue to crawl the data sources, store metadata in the AWS Glue Data Catalog, use Amazon Athena to query the data, use SQL for structured data sources, and use PartiQL for data that is stored in JSON format.
AWS Glue is a serverless data integration service that makes it easy to prepare, clean, enrich, and move data between data stores1. AWS Glue crawlers are processes that connect to a data store, progress through a prioritized list of classifiers to determine the schema for your data, and then create metadata tables in the Data Catalog2. The Data Catalog is a persistent metadata store that contains table definitions, job definitions, and other control information to help you manage your AWS Glue components3. You can use AWS Glue to crawl the data sources, such as Amazon S3, Amazon RDS for Microsoft SQL Server, and Amazon DynamoDB, and store the metadata in the Data Catalog.
Amazon Athena is a serverless, interactive query service that makes it easy to analyze data directly in Amazon S3 using standard SQL or Python4. Amazon Athena also supports PartiQL, a SQL-compatible query language that lets you query, insert, update, and delete data from semi-structured and nested data, such as JSON. You can use Amazon Athena to query the data from the Data Catalog using SQL for structured data sources, such as .csv files and relational databases, and PartiQL for data that is stored in JSON format. You can also use Athena to query data from other data sources, such as Amazon Redshift, using federated queries.
Using AWS Glue and Amazon Athena to query all data sources by using syntax similar to SQL is the least operational overhead solution, as you do not need to provision, manage, or scale any infrastructure, and you pay only for the resources you use. AWS Glue charges you based on the compute time and the data processed by your crawlers and ETL jobs1. Amazon Athena charges you based on the amount of data scanned by your queries. You can also reduce the cost and improve the performance of your queries by using compression, partitioning, and columnar formats for your data in Amazon S3.
Option B is not the best solution, as using AWS Glue to crawl the data sources, store metadata in the AWS Glue Data Catalog, and use Redshift Spectrum to query the data, would incur more costs and complexity than using Amazon Athena. Redshift Spectrum is a feature of Amazon Redshift, a fully managed data warehouse service, that allows you to query and join data across your data warehouse and your data lake using standard SQL. While Redshift Spectrum is powerful and useful for many data warehousing scenarios, it is not necessary or cost-effective for querying all data sources by using syntax similar to SQL. Redshift Spectrum charges you based on the amount of data scanned by your queries, which is similar to Amazon Athena, but it also requires you to have an Amazon Redshift cluster, which charges you based on the node type, the number of nodes, and the duration of the cluster5. These costs can add up quickly, especially if you have large volumes of data and complex queries. Moreover, using Redshift Spectrum would introduce additional latency and complexity, as you would have to provision and manage the cluster, and create an external schema and database for the data in the Data Catalog, instead of querying it directly from Amazon Athena.
Option C is not the best solution, as using AWS Glue to crawl the data sources, store metadata in the AWS Glue Data Catalog, use AWS Glue jobs to transform data that is in JSON format to Apache Parquet or .csv format, store the transformed data in an S3 bucket, and use Amazon Athena to query the original and transformed data from the S3 bucket, would incur more costs and complexity than using Amazon Athena with PartiQL. AWS Glue jobs are ETL scripts that you can write in Python or Scala to transform your data and load it to your target data store. Apache Parquet is a columnar storage format that can improve the performance of analytical queries by reducing the amount of data that needs to be scanned and providing efficient compression and encoding schemes6. While using AWS Glue jobs and Parquet can improve the performance and reduce the cost of your queries, they would also increase the complexity and the operational overhead of the data pipeline, as you would have to write, run, and monitor the ETL jobs, and store the transformed data in a separate location in Amazon S3. Moreover, using AWS Glue jobs and Parquet would introduce additional latency, as you would have to wait for the ETL jobs to finish before querying the transformed data.
Option D is not the best solution, as using AWS Lake Formation to create a data lake, use Lake Formation jobs to transform the data from all data sources to Apache Parquet format, store the transformed data in an S3 bucket, and use Amazon Athena or Redshift Spectrum to query the data, would incur more costs and complexity than using Amazon Athena with PartiQL. AWS Lake Formation is a service that helps you centrally govern, secure, and globally share data for analytics and machine learning7. Lake Formation jobs are ETL jobs that you can create and run using the Lake Formation console or API. While using Lake Formation and Parquet can improve the performance and reduce the cost of your queries, they would also increase the complexity and the operational overhead of the data pipeline, as you would have to create, run, and monitor the Lake Formation jobs, and store the transformed data in a separate location in Amazon S3. Moreover, using Lake Formation and Parquet would introduce additional latency, as you would have to wait for the Lake Formation jobs to finish before querying the transformed data. Furthermore, using Redshift Spectrum to query the data would also incur the same costs and complexity as mentioned in option B. Reference:
What is Amazon Athena?
Data Catalog and crawlers in AWS Glue
AWS Glue Data Catalog
Columnar Storage Formats

AWS Certified Data Engineer - Associate DEA-C01 Complete Study Guide
AWS Glue Schema Registry
What is AWS Glue?
Amazon Redshift Serverless
Amazon Redshift provisioned clusters
[Querying external data using Amazon Redshift Spectrum]
[Using stored procedures in Amazon Redshift]
[What is AWS Lambda?]
[PartiQL for Amazon Athena]
[Federated queries in Amazon Athena]
[Amazon Athena pricing]
[Top 10 performance tuning tips for Amazon Athena]
[AWS Glue ETL jobs]
[AWS Lake Formation jobs]

NEW QUESTION # 130

A data engineer needs to create an Amazon Athena table based on a subset of data from an existing Athena table named cities_world. The cities_world table contains cities that are located around the world. The data engineer must create a new table named cities_us to contain only the cities from cities_world that are located in the US.
Which SQL statement should the data engineer use to meet this requirement?

A.

```
INSERT INTO cities_usa (city,state) SELECT city, state FROM cities_world WHERE country='usa';
```

B.

```
MOVE city, state FROM cities_world TO cities_usa WHERE country='usa';
```

C.

```
INSERT INTO cities_usa SELECT city, state FROM cities_world WHERE country='usa';
```

D.

```
UPDATE cities_usa SET (city, state) = (SELECT city, state FROM cities_world WHERE
```


- A. Option B
- B. Option D
- C. Option A
- D. Option C

Answer: C

Explanation:
To create a new table named cities_usa in Amazon Athena based on a subset of data from the existing cities_world table, you should use an INSERT INTO statement combined with a SELECT statement to filter only the records where the country is 'usa'. The correct SQL syntax would be:
Option A: INSERT INTO cities_usa (city, state) SELECT city, state FROM cities_world WHERE country='usa';This statement inserts only the cities and states where the country column has a value of 'usa' from the cities_world table into the cities_usa table. This is a correct approach to create a new table with data filtered from an existing table in Athena.
Options B, C, and D are incorrect due to syntax errors or incorrect SQL usage (e.g., the MOVE command or the use of UPDATE in a non-relevant context).
References:
Amazon Athena SQL Reference
Creating Tables in Athena

NEW QUESTION # 131

A marketing company uses Amazon S3 to store marketing data. The company uses versioning in some buckets. The company runs

several jobs to read and load data into the buckets.
To help cost-optimize its storage, the company wants to gather information about incomplete multipart uploads and outdated versions that are present in the S3 buckets.
Which solution will meet these requirements with the LEAST operational effort?

- A. Use AWS CLI to gather the information.
- B. Use Amazon S3 Inventory configurations reports to gather the information.
- C. Use the Amazon S3 Storage Lens dashboard to gather the information.
- D. Use AWS usage reports for Amazon S3 to gather the information.

Answer: B

Explanation:
The company wants to gather information about incomplete multipart uploads and outdated versions in its Amazon S3 buckets to optimize storage costs.
* Option B: Use Amazon S3 Inventory configurations reports to gather the information.S3 Inventory provides reports that can list incomplete multipart uploads and versions of objects stored in S3. It offers an easy, automated way to track object metadata across buckets, including data necessary for cost optimization, without manual effort.
Options A (AWS CLI), C (S3 Storage Lens), and D (usage reports) either do not specifically gather the required information about incomplete uploads and outdated versions or require more manual intervention.
References:
* Amazon S3 Inventory Documentation

NEW QUESTION # 132
......

Three Data-Engineer-Associate exam questions formats that we have are AWS Certified Data Engineer - Associate (DEA-C01) (Data-Engineer-Associate) dumps PDF format, web-based Data-Engineer-Associate practice exam and desktop-based Data-Engineer-Associate practice test software. Our Amazon Data-Engineer-Associate PDF dumps format has actual Data-Engineer-Associate Questions which are printable and portable. Hence, you can go through these Data-Engineer-Associate questions via your smart devices like smartphones, laptops, and tablets.

Valid Data-Engineer-Associate Exam Dumps: https://www.exam-killer.com/Data-Engineer-Associate-valid-questions.html

Amazon Data-Engineer-Associate Latest Test Testking Your learning will be full of pleasure, The purchase process for Data-Engineer-Associate exam dumps is very easy and convenient to operate, As is known to all, Data-Engineer-Associate practice guide simulation plays an important part in the success of exams, Amazon Data-Engineer-Associate Latest Test Testking Many enterprise customers built long-term relationship with us year by year, Amazon Data-Engineer-Associate Latest Test Testking The questions and answers created by the certified professors provide you with the experience of taking the real test.

These reports show scores, Jeff Carlson, author of Palm Organizers: Data-Engineer-Associate Visual QuickStart Guide, asks the question, Is it possible to leave your laptop at home and just carry a Palm organizer?

Newest Data-Engineer-Associate Latest Test Testking & Leading Offer in Qualification Exams & Authoritative Valid Data-Engineer-Associate Exam Dumps

Your learning will be full of pleasure, The purchase process for Data-Engineer-Associate exam dumps is very easy and convenient to operate, As is known to all, Data-Engineer-Associate practice guide simulation plays an important part in the success of exams.

Many enterprise customers built long-term relationship with us year Data-Engineer-Associate Latest Test Testking by year, The questions and answers created by the certified professors provide you with the experience of taking the real test.

Associate Exam Preparation ☢ Search for □ Data-Engineer-Associate □ and download it for free immediately on ☀ www.prep4away.com □☀□ ✏Data-Engineer-Associate Exam Simulator

- Certification Data-Engineer-Associate Torrent □ Data-Engineer-Associate Practice Exams □ Answers Data-Engineer-Associate Real Questions □ Search for ➠ Data-Engineer-Associate □ and download it for free on ✔ www.pdfvce.com □✔□ website □Data-Engineer-Associate Test Preparation
- Data-Engineer-Associate Learning materials: AWS Certified Data Engineer - Associate (DEA-C01) - Data-Engineer-Associate Exam Preparation □ Open website ▹ www.prepawayexam.com ◃ and search for 「Data-Engineer-Associate 」 for free download □Answers Data-Engineer-Associate Real Questions
- Data-Engineer-Associate - AWS Certified Data Engineer - Associate (DEA-C01) –Professional Latest Test Testking □ Go to website 《 www.pdfvce.com 》 open and search for 【 Data-Engineer-Associate 】 to download for free □Exams Data-Engineer-Associate Torrent
- Data-Engineer-Associate - AWS Certified Data Engineer - Associate (DEA-C01) –Professional Latest Test Testking □ Easily obtain free download of ☀ Data-Engineer-Associate □☀□ by searching on ➤ www.examcollectionpass.com □ □ □Exam Data-Engineer-Associate Dump
- Data-Engineer-Associate Study Guide: AWS Certified Data Engineer - Associate (DEA-C01) - Data-Engineer-Associate Practice Test - AWS Certified Data Engineer - Associate (DEA-C01) Learning Materials □ Download 《 Data-Engineer-Associate 》 for free by simply entering □ www.pdfvce.com □ website □Data-Engineer-Associate Quiz
- Data-Engineer-Associate Learning materials: AWS Certified Data Engineer - Associate (DEA-C01) - Data-Engineer-Associate Exam Preparation □ Search for { Data-Engineer-Associate } and download exam materials for free through " www.practicevce.com " □Exams Data-Engineer-Associate Torrent
- Data-Engineer-Associate Study Guide: AWS Certified Data Engineer - Associate (DEA-C01) - Data-Engineer-Associate Practice Test - AWS Certified Data Engineer - Associate (DEA-C01) Learning Materials □ Simply search for □ Data-Engineer-Associate □ for free download on ▸ www.pdfvce.com ◂ ✔Data-Engineer-Associate Practice Exams
- New Data-Engineer-Associate Braindumps □ Free Data-Engineer-Associate Exam Questions □ Data-Engineer-Associate Practice Exams □ Immediately open ➠ www.torrentvce.com □ and search for □ Data-Engineer-Associate □ to obtain a free download □Exam Data-Engineer-Associate Dump
- www.stes.tyc.edu.tw, www.stes.tyc.edu.tw, liviazsrn511766.ourabilitywiki.com, tiffanywazf151557.wikirecognition.com, honeyanjd641080.shivawiki.com, monicaiqwz943454.blogspothub.com, bushrazqko397381.blogsvirals.com, janaowqc280633.blogdal.com, zoyapjto240755.pennywiki.com, izaakjuvv934882.dgbloggers.com, Disposable vapes

BTW, DOWNLOAD part of Exam-Killer Data-Engineer-Associate dumps from Cloud Storage: https://drive.google.com/open?id=1cH5dynzaY2IFs6yhZ_3j94R4WMafRJnm