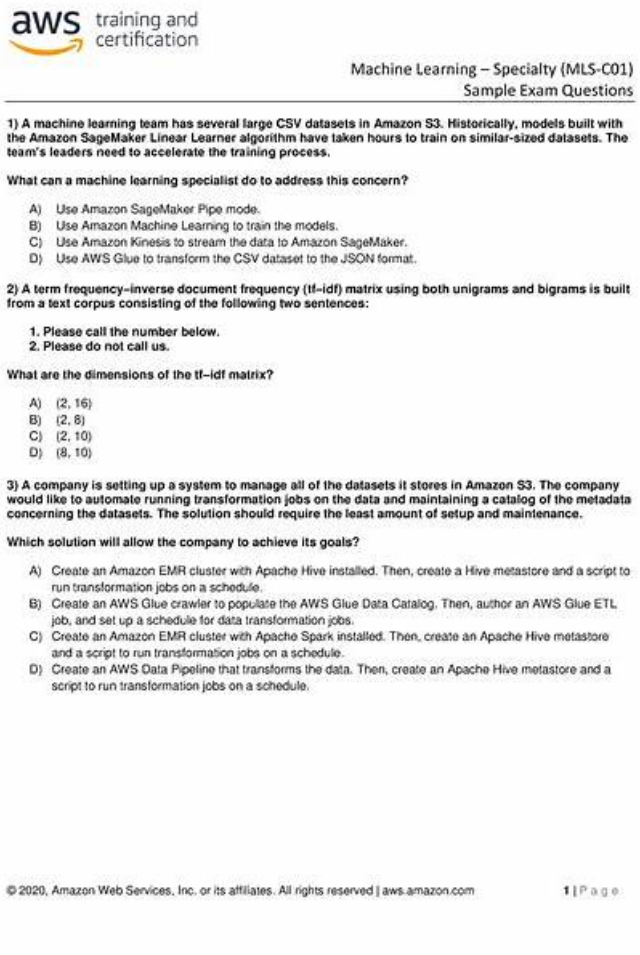
AWS-Certified-Machine-Learning-Specialty Real Exam Questions in Three Formats



aws training and certification

Machine Learning – Specialty (MLS-C01)
Sample Exam Questions

1) A machine learning team has several large CSV datasets in Amazon S3. Historically, models built with the Amazon SageMaker Linear Learner algorithm have taken hours to train on similar-sized datasets. The team's leaders need to accelerate the training process.

What can a machine learning specialist do to address this concern?

A) Use Amazon SageMaker Pipe mode.
B) Use Amazon Machine Learning to train the models.
C) Use Amazon Kinesis to stream the data to Amazon SageMaker.
D) Use AWS Glue to transform the CSV dataset to the JSON format.

2) A term frequency–inverse document frequency (tf–idf) matrix using both unigrams and bigrams is built from a text corpus consisting of the following two sentences:

1. Please call the number below.
2. Please do not call us.

What are the dimensions of the tf–idf matrix?

A) (2, 16)
B) (2, 8)
C) (2, 10)
D) (8, 10)

3) A company is setting up a system to manage all of the datasets it stores in Amazon S3. The company would like to automate running transformation jobs on the data and maintaining a catalog of the metadata concerning the datasets. The solution should require the least amount of setup and maintenance.

Which solution will allow the company to achieve its goals?

A) Create an Amazon EMR cluster with Apache Hive installed. Then, create a Hive metastore and a script to run transformation jobs on a schedule.
B) Create an AWS Glue crawler to populate the AWS Glue Data Catalog. Then, author an AWS Glue ETL job, and set up a schedule for data transformation jobs.
C) Create an Amazon EMR cluster with Apache Spark installed. Then, create an Apache Hive metastore and a script to run transformation jobs on a schedule.
D) Create an AWS Data Pipeline that transforms the data. Then, create an Apache Hive metastore and a script to run transformation jobs on a schedule.

2026 Latest Test4Cram AWS-Certified-Machine-Learning-Specialty PDF Dumps and AWS-Certified-Machine-Learning-Specialty Exam Engine Free Share: https://drive.google.com/open?id=1-1R4R6YSLIQgZJOt_DOcVFDJOCChSrfR

Your personal experience will defeat all advertisements that we post before. When you enter our website, you can download the free demo of AWS-Certified-Machine-Learning-Specialty exam software. We believe you will like our dumps that have helped more candidates Pass AWS-Certified-Machine-Learning-Specialty Exam after you have tried it. Using our exam dump, you can easily become IT elite with AWS-Certified-Machine-Learning-Specialty exam certification.

Understanding functional and technical aspects of AWS Certified Machine Learning Specialty Exam Modeling

The following will be dicussed here:

- Frame business problems as machine learning problems
- Train machine learning models
- Perform hyperparameter optimization
- Evaluate machine learning models
- Select the appropriate model(s) for a given machine learning problem

>> AWS-Certified-Machine-Learning-Specialty Exam Topic <<

AWS-Certified-Machine-Learning-Specialty Valid Braindumps - AWS-Certified-Machine-Learning-Specialty Exam Fees

The Amazon AWS-Certified-Machine-Learning-Specialty PDF is the collection of real, valid, and updated Amazon AWS-Certified-Machine-Learning-Specialty practice questions. The AWS-Certified-Machine-Learning-Specialty PDF dumps file works with all smart devices. You can use the AWS Certified Machine Learning - Specialty PDF questions on your tablet, smartphone, or laptop and start AWS-Certified-Machine-Learning-Specialty Exam Preparation anytime and anywhere. The AWS-Certified-Machine-Learning-Specialty dumps PDF provides you with everything that you must need in Amazon AWS-Certified-Machine-Learning-Specialty exam preparation and enable you to crack the final Amazon AWS-Certified-Machine-Learning-Specialty exam quickly.

The AWS Certified Machine Learning - Specialty certification exam covers a wide range of topics, including data preparation, model training, model deployment, and machine learning algorithms. AWS-Certified-Machine-Learning-Specialty exam is designed to test the candidate's knowledge of AWS services and their ability to apply machine learning techniques to real-world problems. AWS-Certified-Machine-Learning-Specialty Exam consists of multiple-choice and multiple-response questions, and the candidate has 170 minutes to complete it.

Amazon AWS Certified Machine Learning - Specialty Sample Questions (Q151-Q156):

NEW QUESTION # 151
A company needs to quickly make sense of a large amount of data and gain insight from it. The data is in different formats, the schemas change frequently, and new data sources are added regularly. The company wants to use AWS services to explore multiple data sources, suggest schemas, and enrich and transform the data. The solution should require the least possible coding effort for the data flows and the least possible infrastructure management.
Which combination of AWS services will meet these requirements?

- A. AWS Glue for data discovery, enrichment, and transformation
  Amazon Athena for querying and analyzing the results in Amazon S3 using standard SQL Amazon QuickSight for reporting and getting insights
- B. Amazon Kinesis Data Analytics for data ingestion
  Amazon EMR for data discovery, enrichment, and transformation
  Amazon Redshift for querying and analyzing the results in Amazon S3
- C. Amazon EMR for data discovery, enrichment, and transformation
  Amazon Athena for querying and analyzing the results in Amazon S3 using standard SQL Amazon QuickSight for reporting and getting insights
- D. AWS Data Pipeline for data transfer AWS Step Functions for orchestrating AWS Lambda jobs for data discovery, enrichment, and transformation Amazon Athena for querying and analyzing the results in Amazon S3 using standard SQL Amazon QuickSight for reporting and getting insights

Answer: A

Explanation:
Explanation
The best combination of AWS services to meet the requirements of data discovery, enrichment, transformation, querying, analysis, and reporting with the least coding and infrastructure management is AWS Glue, Amazon Athena, and Amazon QuickSight. These services are:
AWS Glue for data discovery, enrichment, and transformation. AWS Glue is a serverless data integration service that automatically crawls, catalogs, and prepares data from various sources and formats. It also provides a visual interface called AWS Glue DataBrew that allows users to apply over
250 transformations to clean, normalize, and enrich data without writing code1 Amazon Athena for querying and analyzing the results in Amazon S3 using standard SQL. Amazon Athena is a serverless interactive query service that allows users to analyze data in Amazon S3 using standard SQL. It supports a variety of data formats, such as CSV, JSON, ORC, Parquet, and Avro. It also integrates with AWS Glue Data Catalog to provide a unified view of the data sources and schemas2 Amazon QuickSight for reporting and getting insights. Amazon QuickSight is a serverless business intelligence service that allows users to create and share interactive dashboards and reports. It also provides ML-powered features, such as anomaly detection, forecasting, and natural language queries, to help users discover hidden insights from their data3 The other options are not suitable because they either require more coding effort, more infrastructure management, or do not support the desired use cases. For example:
Option A uses Amazon EMR for data discovery, enrichment, and transformation. Amazon EMR is a managed cluster platform that runs Apache Spark, Apache Hive, and other open-source frameworks for big data processing. It requires users to write code in

languages such as Python, Scala, or SQL to perform data integration tasks. It also requires users to provision, configure, and scale the clusters according to their needs4 Option B uses Amazon Kinesis Data Analytics for data ingestion. Amazon Kinesis Data Analytics is a service that allows users to process streaming data in real time using SQL or Apache Flink. It is not suitable for data discovery, enrichment, and transformation, which are typically batch-oriented tasks. It also requires users to write code to define the data processing logic and the output destination5 Option D uses AWS Data Pipeline for data transfer and AWS Step Functions for orchestrating AWS Lambda jobs for data discovery, enrichment, and transformation. AWS Data Pipeline is a service that helps users move data between AWS services and on-premises data sources. AWS Step Functions is a service that helps users coordinate multiple AWS services into workflows. AWS Lambda is a service that lets users run code without provisioning or managing servers. These services require users to write code to define the data sources, destinations, transformations, and workflows. They also require users to manage the scalability, performance, and reliability of the data pipelines.
References:
1: AWS Glue - Data Integration Service - Amazon Web Services
2: Amazon Athena - Interactive SQL Query Service - AWS
3: Amazon QuickSight - Business Intelligence Service - AWS
4: Amazon EMR - Amazon Web Services
5: Amazon Kinesis Data Analytics - Amazon Web Services
6: AWS Data Pipeline - Amazon Web Services
7: AWS Step Functions - Amazon Web Services
8: AWS Lambda - Amazon Web Services

NEW QUESTION # 152
A Data Scientist needs to migrate an existing on-premises ETL process to the cloud The current process runs at regular time intervals and uses PySpark to combine and format multiple large data sources into a single consolidated output for downstream processing The Data Scientist has been given the following requirements for the cloud solution
* Combine multiple data sources
* Reuse existing PySpark logic
* Run the solution on the existing schedule
* Minimize the number of servers that will need to be managed
Which architecture should the Data Scientist use to build this solution?

- A. Use Amazon Kinesis Data Analytics to stream the input data and perform realtime SQL queries against the stream to carry out the required transformations within the stream Deliver the output results to a "processed" location in Amazon S3 that is accessible for downstream use
- B. Write the raw data to Amazon S3 Create an AWS Glue ETL job to perform the ETL processing against the input data Write the ETL job in PySpark to leverage the existing logic Create a new AWS Glue trigger to trigger the ETL job based on the existing schedule Configure the output target of the ETL job to write to a "processed" location in Amazon S3 that is accessible for downstream use.
- C. Write the raw data to Amazon S3 Schedule an AWS Lambda function to submit a Spark step to a persistent Amazon EMR cluster based on the existing schedule Use the existing PySpark logic to run the ETL job on the EMR cluster Output the results to a "processed" location m Amazon S3 that is accessible tor downstream use
- D. Write the raw data to Amazon S3 Schedule an AWS Lambda function to run on the existing schedule and process the input data from Amazon S3 Write the Lambda logic in Python and implement the existing PySpartc logic to perform the ETL process Have the Lambda function output the results to a "processed" location in Amazon S3 that is accessible for downstream use

Answer: C

NEW QUESTION # 153
An interactive online dictionary wants to add a widget that displays words used in similar contexts. A Machine Learning Specialist is asked to provide word features for the downstream nearest neighbor model powering the widget.
What should the Specialist do to meet these requirements?

- A. Create word embedding vectors that store edit distance with every other word.
- B. Download word embeddings pre-trained on a large corpus.
- C. Produce a set of synonyms for every word using Amazon Mechanical Turk.
- D. Create one-hot word encoding vectors.

Answer: B

Explanation:
As it is a interactive online dictionary, we need pre-trained word embedding thus the answer is D.
In addition, there is no mention that the online dictonary is unique and does not have a pre- trained word embedding.

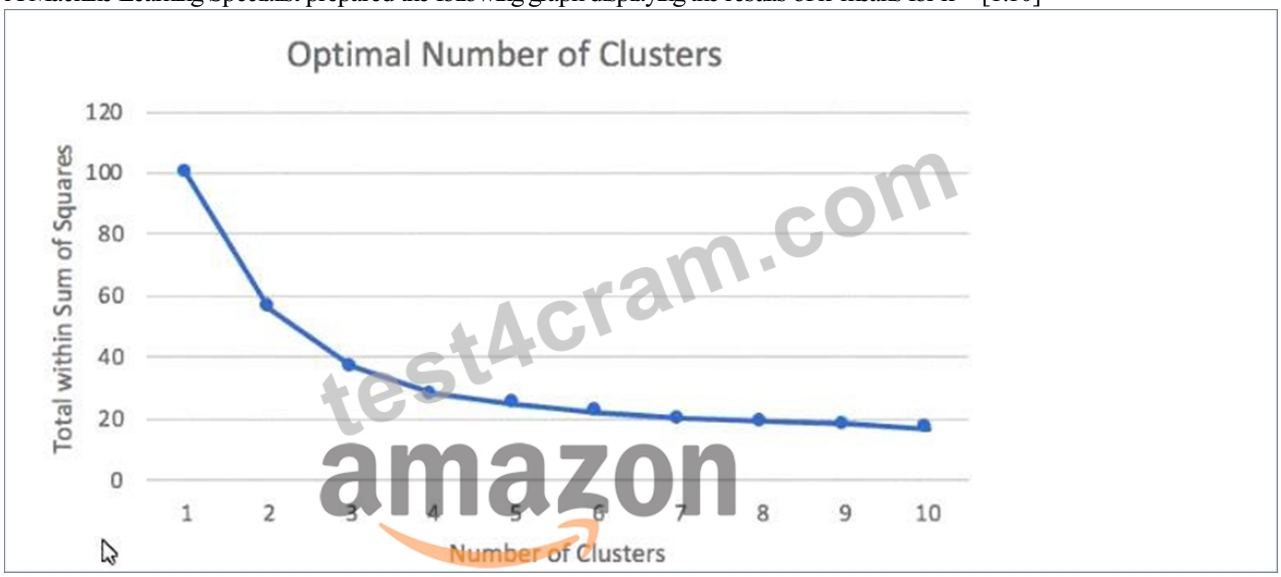
NEW QUESTION # 154
A data scientist is building a forecasting model for a retail company by using the most recent 5 years of sales records that are stored in a data warehouse. The dataset contains sales records for each of the company's stores across five commercial regions The data scientist creates a working dataset with StorelD. Region. Date, and Sales Amount as columns. The data scientist wants to analyze yearly average sales for each region. The scientist also wants to compare how each region performed compared to average sales across all commercial regions.
Which visualization will help the data scientist better understand the data trend?

- A. Create an aggregated dataset by using the Pandas GroupBy function to get average sales for each year for each region Create a bar plot, faceted by year, of average sales for each region Add a horizontal line in each facet to represent average sales.
- B. Create an aggregated dataset by using the Pandas GroupBy function to get average sales for each year for each region Create a bar plot of average sales for each region. Add an extra bar in each facet to represent average sales.
- C. Create an aggregated dataset by using the Pandas GroupBy function to get average sales for each year for each store. Create a bar plot, faceted by year, of average sales for each store. Add an extra bar in each facet to represent average sales.
- D. Create an aggregated dataset by using the Pandas GroupBy function to get average sales for each year for each store. Create a bar plot, colored by region and faceted by year, of average sales for each store.
  Add a horizontal line in each facet to represent average sales.

Answer: A

Explanation:
The best visualization for this task is to create a bar plot, faceted by year, of average sales for each region and add a horizontal line in each facet to represent average sales. This way, the data scientist can easily compare the yearly average sales for each region with the overall average sales and see the trends over time. The bar plot also allows the data scientist to see the relative performance of each region within each year and across years. The other options are less effective because they either do not show the yearly trends, do not show the overall average sales, or do not group the data by region.
References:
* pandas.DataFrame.groupby - pandas 2.1.4 documentation
* pandas.DataFrame.plot.bar - pandas 2.1.4 documentation
* Matplotlib - Bar Plot - Online Tutorials Library

NEW QUESTION # 155
A Machine Learning Specialist prepared the following graph displaying the results of k-means for k = [1:10]



Considering the graph, what is a reasonable selection for the optimal choice of k?

- A. 0

- B. 1
- C. 2
- D. 3

Answer: D

Explanation:
The elbow method is a technique that we use to determine the number of centroids (k) to use in a k-means clustering algorithm. In this method, we plot the within-cluster sum of squares (WCSS) against the number of clusters (k) and look for the point where the curve bends sharply. This point is called the elbow point and it indicates that adding more clusters does not improve the model significantly. The graph in the question shows that the elbow point is at k = 4, which means that 4 is a reasonable choice for the optimal number of clusters.
References:
* Elbow Method for optimal value of k in KMeans: A tutorial on how to use the elbow method with Amazon SageMaker.
* K-Means Clustering: A video that explains the concept and benefits of k-means clustering.

NEW QUESTION # 156
.....

Specialty Valid Braindumps □ Open ☀ www.pdfdumps.com □☀□ enter ➽ AWS-Certified-Machine-Learning-Specialty □ and obtain a free download □Test AWS-Certified-Machine-Learning-Specialty Cram

- myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, www.stes.tyc.edu.tw, www.stes.tyc.edu.tw, www.stes.tyc.edu.tw, digilearn.co.zw, pct.edu.pk, www.skudci.com, peterbonadieacademy.org, www.stes.tyc.edu.tw, ncon.edu.sa, Disposable vapes

2026 Latest Test4Cram AWS-Certified-Machine-Learning-Specialty PDF Dumps and AWS-Certified-Machine-Learning-Specialty Exam Engine Free Share: https://drive.google.com/open?id=1-1R4R6YSLIQgZJOt_DOcVFDJOCChSrfR