

**Iniciado em** quinta, 21 set 2023, 22:55  
**Estado** Finalizada  
**Concluída em** quinta, 21 set 2023, 22:57  
**Tempo empregado** 1 minuto 38 segundos  
**Avaliar** 10,00 de um máximo de 10,00(100%)


Questão 1

Correto

Atingiu 1,67 de 1,67

Você é membro de uma equipe de aprendizado de máquina em um grande varejista on-line. Sua equipe é responsável pela análise da concorrência no varejo. Você tem uma fonte de streaming de dados de produtos concorrentes que precisa ingerir em seu data lake. Você precisa usar os dados do produto do concorrente na streaming para relacionar aos dados do produto correspondente em seu catálogo de produtos. Usando essa correspondência, seus cientistas de dados podem produzir painéis de análise competitiva em uma ferramenta de BI.

Qual das opções a seguir oferece a melhor ingestão de dados e a solução de comparação de produtos mais eficiente?

- ☐ a. Kinesis Data Firehose -> Elasticsearch -> Kibana
- ☒ b. Kinesis Data Firehose -> S3 -> Lake Formation -> QuickSight  A alternativa está CORRETA. O Kinesis Data Firehose pode transmitir diretamente para o S3. O Lake Formation tem uma transformação FindMatches que permite identificar registros correspondentes em seu conjunto de dados, mesmo quando os registros não têm um identificador exclusivo comum. Nenhum campo corresponde exatamente. Nesse cenário, você fará a correspondência de produtos em seu catálogo de produtos com suas origens de produtos concorrentes, mesmo que as entradas de produtos sejam estruturadas de maneira diferente. O QuickSight permite que seus cientistas de dados produzam seus painéis de análise.
- ☐ c. Kinesis Data Streams -> S3 -> Lake Formation -> QuickSight
- ☐ d. Kinesis Data Streams -> Elasticsearch -> Kibana

Sua resposta está correta.

A resposta correta é: Kinesis Data Firehose -> S3 -> Lake Formation -> QuickSight



## Questão 2

Correto

Atingiu 1,67 de 1,67

Você trabalha como especialista em aprendizado de máquina para uma empresa de bicicletas elétricas. As bicicletas elétricas que sua empresa produz possuem sensores de IoT que transmitem informações de uso e manutenção para o data lake de sua empresa. Você está coletando os dados de IoT da bicicleta e armazenando-os em um bucket do S3 que pode ser usado para seus modelos de aprendizado de máquina. Você faz parte da equipe que tem a atribuição de usar os dados de IoT para prever quando as bicicletas elétricas de seus clientes precisam de manutenção.

Os dados de IoT produzidos pelas bicicletas elétricas não são estruturados e, às vezes, dependendo do fabricante da peça de IoT, os dados têm uma estrutura de esquema diferente. Você precisa limpar e classificar os dados de IoT antes de usá-los em seu modelo de aprendizado de máquina. Como você pode construir um script ETL para realizar a limpeza e classificação necessárias sabendo que você tem dados de mensagem com diferentes estruturas de esquema?

- ☐ a. Use o AWS Glue para criar uma série de transformações que usam o Apache Spark para passar os dados de uma transformação para outra. Cada transformação executa uma tarefa diferente de limpeza e/ou transformação.
- ☒ b. Use o AWS Glue para criar uma série de transformações que passam os dados de uma transformação para outra. Cada transformação executa uma tarefa diferente de limpeza e/ou transformação. ✓

Sua resposta está correta.

A resposta correta é: Use o AWS Glue para criar uma série de transformações que passam os dados de uma transformação para outra. Cada transformação executa uma tarefa diferente de limpeza e/ou transformação.

## Questão 3

Correto

Atingiu 1,67 de 1,67

Você trabalha como especialista em aprendizado de máquina para uma empresa de serviços financeiros. Sua equipe de aprendizado de máquina foi incumbida de criar um modelo de análise quantitativa para seus gerentes de portfólio de fundos mútuos no departamento quantitativo da empresa. Você tem várias fontes de dados do provedor de dados financeiros que precisa usar em seu modelo. Você está procurando a plataforma de fonte de dados ideal para ingerir dados em seu ambiente de notebook jupyter de aprendizado de máquina. Quais opções **NÃO** são uma plataforma de fonte de dados que você pode usar? (Selecione DOIS)

- ☐ a. Redshift
- ☒ b. RDS ✓ A alternativa está correta. O RDS não é uma fonte de dados viável para a ingestão de dados em seu ambiente de notebook jupyter de aprendizado de máquina.
- ☐ c. EMR
- ☐ d. Athena
- ☒ e. DynamoDB ✓ A alternativa está correta. O DynamoDB não é uma fonte de dados viável para ingerir dados em seu ambiente de notebook jupyter de aprendizado de máquina.

Sua resposta está correta.

As respostas corretas são: DynamoDB, RDS



#### Questão 4

Correto

Atingiu 1,66 de 1,66

Qual das seguintes afirmações é verdadeira sobre Data Lake e Data Lakehouse?

- ☐ a. Um Data Lake é um sistema de armazenamento que organiza dados de forma estruturada e hierárquica, enquanto um Data Lakehouse é um repositório de dados em que os dados são armazenados em seu formato bruto, sem organização.
- ☒ b. Um Data Lake é um repositório de dados em que os dados são armazenados em seu formato bruto, sem organização, enquanto um Data Lakehouse é uma evolução do Data Lake que adiciona camadas de gerenciamento e organização para facilitar análises mais eficazes. ✓
- ☐ c. Um Data Lake e um Data Lakehouse são conceitos obsoletos que não são mais usados em ambientes de dados modernos.
- ☐ d. Um Data Lake e um Data Lakehouse são termos intercambiáveis usados para descrever o mesmo conceito de armazenamento de dados em seu formato bruto, sem organização.

Sua resposta está correta.

A resposta correta é: Um Data Lake é um repositório de dados em que os dados são armazenados em seu formato bruto, sem organização, enquanto um Data Lakehouse é uma evolução do Data Lake que adiciona camadas de gerenciamento e organização para facilitar análises mais eficazes.

#### Questão 5

Correto

Atingiu 1,67 de 1,67

Você é um especialista em aprendizado de máquina que trabalha para um site de compras de varejo on-line. Sua equipe de aprendizado de máquina é responsável por criar um ambiente de aprendizado de máquina para prever vendas online e otimização do pipeline de produtos. Sua equipe também precisa otimizar a solução de ingestão de dados em seu data lake, que é a principal fonte de seus modelos de aprendizado de máquina. Sua solução de ingestão também facilitará a análise (análise em tempo real e interativa de dados históricos), análise de fluxo de cliques, bem como recomendações de produtos. Qual opção atende melhor aos requisitos de sua equipe?



- ☐ a. Use o Athena como o catálogo de dados de seus arquivos de data lake, use o Kinesis Data Streams e o Kinesis Data Analytics para gerar insights de dados quase em tempo real, use o Kinesis Data Firehose para análises de clickstream e use o Glue para criar recomendações personalizadas de produtos.
- ☐ b. Use o Athena como o catálogo de dados de seus arquivos de data lake, use o Kinesis Data Streams e o Kinesis Data Analytics para insights de dados históricos, use o Glue para análises de clickstream e crie recomendações personalizadas de produtos.
- ☒ c. Use Glue como o catálogo de dados de seus arquivos de data lake, use Kinesis Data Streams e Amazon Kinesis Data Analytics para insights de dados em tempo real, use Kinesis Data Firehose para entregar seus dados ElasticSearch para análise de clickstream e use EMR para gerar recomendações personalizadas de produtos. ✓ A alternativa está correta. Glue é a escolha certa para o seu catálogo de dados, usando o catálogo de dados Glue. O Kinesis Data Streams combinado com o Kinesis Data Analytics satisfaz seu requisito de análise quase em tempo real. O Kinesis Data Firehose para ElasticSearch atende ao seu requisito de clickstream, e o EMR usa trabalhos de ignição para atender ao seu requisito de recomendação em escala.
- ☐ d. Use o Glue como o catálogo de dados de seus arquivos de data lake, use o Kinesis Data Streams e o Kinesis Data Analytics para insights de dados históricos, use o Kinesis Data Firehose para entregar seus dados ao ElasticSearch para análises de clickstream e aproveite os painéis do Kibana para criar recomendações personalizadas de produtos.

Sua resposta está correta.


A resposta correta é: Use Glue como o catálogo de dados de seus arquivos de data lake, use Kinesis Data Streams e Amazon Kinesis Data Analytics para insights de dados em tempo real, use Kinesis Data Firehose para entregar seus dados ElasticSearch para análise de clickstream e use EMR para gerar recomendações personalizadas de produtos.

## Questão 6

Correto

Atingiu 1,66 de 1,66

Você trabalha no departamento de análise de dados de uma empresa de software de compartilhamento de viagens. Você precisa usar o algoritmo de aprendizado de máquina K-means para separar os dados de viagem otimizados da sua empresa em clusters com base nas coordenadas de viagem. Como você usaria o AWS Glue da melhor maneira para criar o esquema de dados necessário para classificar os dados da viagem?

- ☒ a. Use rastreadores Glue para rastrear seus dados de compartilhamento de viagens.  A alternativa está correta. A melhor maneira de criar o esquema para seus dados é usar um rastreador Glue que utiliza um classificador ou vários classificadores. (Consulte a documentação do rastreador AWS Glue <https://docs.aws.amazon.com/glue/latest/dg/add-crawler.html>).
- ☐ b. Use o Glue para transformar e nivelar seus dados para que você possa classificar os dados da viagem com base nas coordenadas.
- ☐ c. Use o Glue para gerar automaticamente o código para classificar os dados da viagem com base nas coordenadas.
- ☐ d. Use Glue FindMatches para localizar e remover registros duplicados em seus dados.

Sua resposta está correta.

Para obter um exemplo, consulte a postagem do blog AWS Machine Learning intitulada Serverless unsupervised machine learning with AWS Glue and Amazon Athena: <https://aws.amazon.com/blogs/machine-learning/serverless-unsupervised-machine-learning-with-aws-glue-and-amazon-athena/>.

A resposta correta é: Use rastreadores Glue para rastrear seus dados de compartilhamento de viagens.

◀ ADM (Augmented Data Management, gerenciamento de dados aumentado) Glue Databrew para tratamento de dados automatizado

Seguir para...

Aula 05 - Implementando um pipeline de aprendizado de máquina com Amazon SageMaker ▶

