SuperDataScience

AWS Machine Learning
Practical Exams Reward

1. **You are working for a analytic company and the data that you want to perform analytics is stored in S3 as JSON file. You want increase the speed at which the data is analyzed using Athena. What can you do?**

   A Use AWS Quicksight instead of AWS Athena
   B Convert the file from JSON format to Parquet format
   C Use PCA to reduce the dimensions in the data
   D Convert the file from JSON format to CSV format

2. **John has stored most of his work-related documents in his S3 bucket in CSV format. He wants to run analytics on the data and also visualize them. Which are the AWS service that would help you achieve this?**

   A AWS Glue, AWS Athena and QuickSight
   B AWS Sagemaker, AWS RedShift and QuickSight
   C AWS Glue, AWS Polly and AWS EMR
   D AWS Sagemaker, AWS Athena and AWS EMR

3. **You have created a transforming ETL job written in Pyspark using AWS Glue for the large amount of data stored in S3. But now, you feel it is taking too much of a time to complete the process. What could be a better alternative and also you want to lower the cost of this process?**

   A You can upload the data to redshift and use redshift spectrum to perform the required ETL job with high speed and low cost
   B You can use RDS to store the data and use Athena to perform ETL job to get the desired performance
   C You can create a EMR cluster with Spark, Hive and Flink to perform the ETL job. You can optimize the parameters to get the desired performance
   D None

4. **You have stream of data coming in JSON format and you want to convert it into Parquet format. So that it can be easily analyzed by AWS Athena?**

   A Use AWS Kinesis Stream to stream the data and RDS can convert JSON to Parquet format

   B Use AWS Firehose to stream the data and use Redshift to convert JSON to Parquet format

   C Use AWS Kinesis Block Stream to stream the data and it can convert JSON to Parquet format

   D Use AWS Firehose to stream the data and it can convert JSON to Parquet format

5. **You have access to data coming from DynamoDB and the company you are working, want to perform analytics on that using Athena, in a most efficient way. Which is the most suitable solution?**

   A Use AWS DMS to transform the data to RDS and then use Athena

   B Use AWS Redshift to transform the data to structured format and then use Athena

   C Use AWS Glue to transform the data to Parquet format and then use Athena

   D Use AWS Athena without any modifications to the data

6. **Which of the AWS services can be used as Business Intelligent tool to visualize the data?**

   A Athena

   B Seaborn

   C Matplotlib

   D Quicksight

7. **John is tasked to create a EMR (Elastic Map Reduce) cluster to host the data lake. He has downloaded the required requirements along with Spark on the EMR cluster. How can he run Sagemaker Spark Application on EMR?**

   A By submitting EMR request
   B By submitting spark application jar
   C By submitting SDK call
   D By submitting spark-sdk call

8. **Which of the Apache frame work helps you to pre-process the data and integrate with Sagemaker model?**

   A Apache Hive
   B Apache Flink
   C Apache Spark
   D Apache Rim

9. **You working on a ML project and you want to detect anomaly in the data. Which tool can help you with this?**

   A Quicksight
   B Athena
   C Redshift spectrum
   D Hadoop cluster

10. **Which of the following services provide access to the Hadoop environment?**

    A AWS RDS
    B AWS SDK
    C AWS EMR
    D AWS PK

ANSWER KEY:

1. B	2. A	3. C	4. D	5. C
6. D	7. B	8. C	9. A	10. C