Exam Questions AWS-Certified-Big-Data-Specialty

Amazon AWS Certified Big Data - Speciality

https://www.2passeasy.com/dumps/AWS-Certified-Big-Data-Specialty/

NEW QUESTION 1
The operations team and the development team want a single place to view both operating system
and application logs.
How should you implement this using AWS services? Choose two answers

A. Using AWS CloudFormation, create a CloudWatch Logs LogGroup and send the operating system and application logs of interest using the CloudWatch Logs Agent
B. Using AWS CloudFormation and configuration management, set up remote logging to send events via UDP packets to CloudTrail
C. Using configuration management, set up remote logging to send events to Amazon Kinesis and insert these into Amazon CloudSearch or Amazon Redshift, depending on available analytic tools
D. Using AWS CloudFormation, create a CloudWatch Logs LogGrou
E. Because the CloudWatch log agent automatically sends all operating system logs, you only have to configure the application logs for sending off-machine
F. Using AWS CloudFormation, merge the application logs with the operating system logs, and use IAM Roles to allow both teams to have access to view console output from Amazon EC2

Answer: AC

NEW QUESTION 2
You are working with customer who has 10 TB of archival data that they want to migrate to Amazon Glacier. The customer has a 1Mbps connection to the Internet.
Which service or feature provide the fastest method of getting the data into Amazon Glacier?

A. Amazon Glacier multipart upload
B. AWS Storage Gateway
C. VM Import/Export
D. AWS Import/Export

Answer: D

NEW QUESTION 3
You want to securely distribute credentials for your Amazon RDS instance to your fleet of web server
instances. The credentials are stored in a file that is controlled by a configuration management system.
How do you securely deploy the credentials in an automated manner across the fleet of web server instances, which can number in the hundreds, while retaining the ability to roll back if needed?

A. Store your credential files in an Amazon S3 bucke
B. Use Amazon S3 server-side encryption on the credential file
C. Have a scheduled job that pulls down the credential files into the instances every 10 minutes
D. Store the credential files in your version-controlled repository with the rest of your cod
E. Have a post-commit action in version control that kicks off a job in your continuous integration system which securely copies the new credentials files to all web server instances
F. Insert credential files into user data and use an instance lifecycle policy to periodically refresh the files from the user data
G. Keep credential files as a binary blob in an Amazon RDS MySQL DB instance, and have a script on each Amazon EC2 instance that pulls the files down from the RDS instance
H. Store the credential files in your version-controlled repository with the rest of your cod
I. Use a parallel file copy program to send the credential files from your local machine to the Amazon EC2 instances

Answer: D

NEW QUESTION 4
A us-based company is expanding their web presence into Europe. The company wants to extend their AWS infrastructure from Northern Virginia (us-east-1) into the Dublin (eu-west-1) region. Which of the following options would enable an equivalent experience for users on both continents?

A. Use a public-facing load balancer per region to load-balancer web traffic, and enable HTTP health checks
B. Use a public-facing load balancer per region to load balancer web traffic, and enable sticky sessions
C. Use Amazon Route S3, and apply a geolocation routing policy to distribution traffic across both regions
D. Use Amazon Route S3, and apply a weighted routing policy to distribute traffic across both regions

Answer: C

NEW QUESTION 5
You need to configure an Amazon S3 bucket to serve static assets for your public-facing web application. Which methods ensure that all objects uploaded to the bucket are set to public read? Choose 2 answers

A. Set permissions on the object to public read during upload
B. Configure the bucket ACL to sell all objects to public read
C. Configure the bucket policy to set all objects to public read
D. Use AWS identity and access Management roles to set the bucket to public read
E. Amazon S3 objects default to public read, so no action is needed

Answer: BC

NEW QUESTION 6
You have started a new job and are reviewing your company's infrastructure on AWS You notice one
web application where they have an Elastic Load Balancer (&B) in front of web instances in an Auto Scaling Group When you check the metrics for the ELB in CloudWatch you see four healthy instances in Availability Zone (AZ) A and zero in AZ B There are zero unhealthy instances.
What do you need to fix to balance the instances across AZs?

A. Set the ELB to only be attached to another AZ
B. Make sure Auto Scaling is configured to launch in both AZs
C. Make sure your AMI is available in both AZs
D. Make sure the maximum size of the Auto Scaling Group is greater than 4

Answer: B

NEW QUESTION 7
In AWS, which security aspects are the customer's responsibility? Choose 4 answers

A. Life-Cycle management of IAM credentials
B. Security Group and ACL settings
C. Controlling physical access to compute resources
D. Path management on the EC2 instance's operating system
E. Encryption of EBS volumes
F. Decommissioning storage devices

Answer: ABDE

NEW QUESTION 8
A photo-sharing service stores pictures in Amazon Simple Storage Service (S3) and allows application sign-in using an opened connect-compatible identity provider. Which AWS Security Token Service approach to temporary access should you use for the Amazon S3 operations?

A. Cross-Account Access
B. AWS identity and Access Management roles
C. SAML-based Identity Federation
D. Web identity Federation

Answer: C

NEW QUESTION 9
The project you are working on currently uses a single AWS CloudFormation template to deploy its
AWS infrastructure, which supports a multi-tier web application. You have been tasked with organizing the AWS CloudFormation resources so that they can be maintained in the future, and so that different departments such as Networking and Security can review the architecture before it goes to Production.
How should you do this in a way that accommodates each department, using their existing workflows?

A. Organize the AWS CloudFormation template so that related resources are next to each other in the template, such as VPC subnets and routing rules for Networking and Security groups and IAM information for Security
B. Separate the AWS CloudFormation template into a nested structure that has individual templates for the resources that are to be governed by different departments, and use the outputs from the networking and security stacks for the application template that you control
C. Organize the AWS CloudFormation template so that related resources are next to each other in the template for each department's use, leverage your existing continuous integration tool to constantly deploy changes from all parties to the Production environment, and then run tests for validation
D. Use a custom application and the AWS SDK to replicate the resources defined in the current AWS CloudFormation template, and use the existing code review system to allow other departments toapprove changes before altering the application for future deployments

Answer: B

NEW QUESTION 10
You have launched an Amazon Elastic Compute Cloud (EC2) instance into a public subnet with a primary private IP address assigned, an internet gateway is attached to the VPC, and the public route table is configured to send all internet-based internet. Why is the internet unreachable from this instance?

A. The Internet gateway security group must allow all outbound traffic
B. The instance does not have a public IP address
C. The instance "Source/Destination check" property must be enabled
D. The instance security group must allow all inbound traffic

Answer: B

NEW QUESTION 10
A company needs to deploy virtual desktops to its customers in a virtual private cloud, leveraging
existing security controls. Which set of AWS services and features will meet the company's requirements?

A. Virtual private network connection, AWS Directory services, and ClassicLink
B. Virtual private network connection, AWS Directory services, and Amazon WorkSpaces
C. AWS Directory service, Amazon WorkSpaces, and AWS Identity and Access Management
D. Amazon Elastic Compute Cloud, and AWS identity and access management

Answer: B

NEW QUESTION 11
When attached to an Amazon VPC which two components provide connectivity with external
networks? Choose 2 answers

A. Elastic IPS (EIP)
B. NAT Gateway (NAT)
C. Internet Gateway {IGW)
D. Virtual Private Gateway (VGW)

Answer: CD

NEW QUESTION 15
As part of your continuous deployment process, your application undergoes an I/O load performance
test before it is deployed to production using new AMIs. The application uses one Amazon Elastic Block Store (EBS) PIOPS volume per instance and requires consistent I/O performance.
Which of the following must be carried out to ensure that I/O load performance tests yield the correct results in a repeatable manner?

A. Ensure that the I/O block sizes for the test are randomly selected
B. Ensure that the Amazon EBS volumes have been pre-warmed by reading all the blocks before the test
C. Ensure that snapshots of the Amazon EBS volumes are created as a backup
D. Ensure that the Amazon EBS volume is encrypted
E. Ensure that the Amazon EBS volume has been pre-warmed by creating a snapshot of the volume before the test

Answer: B

NEW QUESTION 16
You have an Auto Scaling group associated with an Elastic Load Balancer (ELB). You have noticed that
instances launched via the Auto Scaling group are being marked unhealthy due to an ELB health check, but these unhealthy instances are not being terminated.
What do you need to do to ensure trial instances marked unhealthy by the ELB will be terminated
and replaced?

A. Change the thresholds set on the Auto Scaling group health check
B. Add an Elastic Load Balancing health check to your Auto Scaling group
C. Increase the value for the Health check interval set on the Elastic Load Balancer
D. Change the health check set on the Elastic Load Balancer to use TCP rather than HTTP checks

Answer: B

NEW QUESTION 18
An existing application stores sensitive information on a non-boot Amazon EBS data volume attached
to an Amazon Elastic Compute Cloud instance. Which of the following approaches would protect the sensitive data on an Amazon EBS volume?

A. Snapshot the current Amazon EBS volum
B. Restore the snapshot to a new, encrypted Amazon EBS volume Mount the Amazon EBS volume
C. Create and mount a new, encrypted Amazon EBS volum
D. Move the data to the new volum
E. Delete the old Amazon EBS volume
F. Unmount the EBS volum
G. Toggle the encryption attribute to Tru
H. Re-mount the Amazon EBs volume
I. Upload your customer keys to AWS CloudHS
J. Associate the Amazon EBS volume with AWS CloudHS
K. Re-mount the Amazon EBS volume

Answer: A

NEW QUESTION 19
When using the following AWS services, which should be implemented in multiple Availability Zones
for high availability solutions? Choose 2 answers

A. Amazon Simple Storage Service
B. Amazon Elastic Load Balancing
C. Amazon Elastic Compute Cloud
D. Amazon Simple Notification Service
E. Amazon DynamoDB

Answer: BC

NEW QUESTION 20
A company needs to monitor the read and write IOPs metrics for their AWS MySQL RDS instances
and send real-time alerts to their operations team. Which AWS services can accomplish this? Choose 2 answers

A. Amazon Simple Email Service
B. Amazon CloudWatch
C. Amazon Simple Queue Service
D. Amazon Route 53
E. Amazon Simple Notification Service

Answer: BE

NEW QUESTION 21
Your application uses CloudFormation to orchestrate your application's resources. During your testing phase before application went live, your Amazon RDS instance type was changed and caused the instance to be re-created, resulting in the loss of test data.
How should you prevent this from occurring in the future?

A. Within the AWS CloudFormation parameter with which users can select the Amazon RDS instance type, set AllowedValues to only contain the current instance

type
B. Use an AWS CloudFormation stack policy to deny updates to the instanc
C. Only allow UpdateStack permission to IAM principles that are denied SetStackPolicy
D. In the AWS CloudFormation template, set the AWS::RDS::DBInstance's DBInstanceClass property to be read-only
E. Subscribe to the AWS CloudFormation notification "BeforeResourceUpdate" and call CancelStackUpdate if the resource identified is the Amazon RDS instance
F. In the AWS ClousFormation template, set the DeletionPolicy of the AWS::RDS::DBInstance's DeletionPolicy property to "Retain"

Answer: E

NEW QUESTION 26
How can the domain's zone apex, for example,"myzoneapexdomain.com", be pointed towards an
Elastic Load Balancer?

A. By using an Amazon Route 53 Alias record
B. By using an A record
C. By using an AAAA record
D. By using an Amazon Route 53 CNAME record

Answer: A

NEW QUESTION 29
Which of the following are characteristics of Amazon VPC subnets? Choose 2 answers

A. Each subnet maps to a single Availability Zone
B. A CIDR block mask of /25 is the smallest range supported
C. Instances in a private subnet can communicate with the internet only if they have an Elastic IP.
D. By default, all subnets can route between each other, whether they are private or public
E. Each subnet spans at least 2 Availability zones to provide a high-availability environment

Answer: AD

NEW QUESTION 30
A company is preparing to give AWS Management Console access to developers. Company policy
mandates identity federation and role based access control. Roles are currently assigned using groups in the corporate Active Directory. What combination of the following will give developers access to the AWS console? Choose 2 answers

A. AWS Directory Service AD connector
B. AWS Directory Service Simple AD
C. AWS identity and Access Management groups
D. AWS identity and Access Management roles
E. AWS identity and Access Management users

Answer: AD

NEW QUESTION 31
A customer wants to track access to their Amazon Simple Storage Service (S3) buckets and also use
this information for their internal security and access audits. Which of the following will meet the Customer requirement?

A. Enable AWS CloudTrail to audit all Amazon S3 bucket access.
B. Enable server access logging for all required Amazon S3 bucket
C. Enable the Requester Pays option to track access via AWS Billing
D. Enable Amazon S3 event notifications for Put and Post.

Answer: B

NEW QUESTION 32
A user is running one instance for only 3 hours every day. The user wants to save some cost with the
instance. Which of the below mentioned Reserved Instance categories is advised in this case?

A. The user should not use RI; instead only go with the on-demand pricing
B. The user should use the AWS high utilized RI
C. The user should use the AWS medium utilized RI
D. The user should use the AWS low utilized RI

Answer: A

NEW QUESTION 34
Your company operates a website for promoters to sell tickets for entertainment events. You are
using a load balancer in front of an Auto Scaling group of web server. Promotion of popular events can cause surges of websites visitors. During scaling-out at theses times, newly launched instances are unable to complete configuration quickly enough, leading to user disappointment.
What option should you choose to improve scaling yet minimize costs? Choose 2 answers

A. Create an AMI with the application pre-configure
B. Create a new Auto Scaling launch configuration using this new AMI, and configure the Auto Scaling group to launch with this AMI
C. Use Auto Scaling pre-warming to launch instances before they are require
D. Configure pre- warming to use the CPU trend CloudWatch metric for the group
E. Publish a custom CloudWatch metric from your application on the number of tickets sold, and create an Auto Scaling policy based on this

F. Using the history of past scaling events for similar event sales to predict future scaling requirement
G. Use the Auto Scaling scheduled scaling feature to vary the size of the fleet
H. Configure an Amazon S3 bucket for website hostin
I. Upload into the bucket an HTML holding page with its 'x-amz-website-redirect-location' metadata property set to the load balancer endpoin
J. Configure Elastic Load Balancing to redirect to the holding page when the load on web servers is above a certain level

Answer: DE

NEW QUESTION 38
A user has setup an RDS DB with Oracle. The user wants to get notifications when someone modifies the security group of that DB. How can the user configure that?

A. It is not possible to get the notifications on a change in the security group
B. Configure SNS to monitor security group changes
C. Configure event notification on the DB security group
D. Configure the CloudWatch alarm on the DB for a change in the security group

Answer: C

NEW QUESTION 41
You have a web application that is currently running on a collection of micro instance types in a single AZ behind a single load balancer. You have an Auto Scaling group configured to scale from 2 to 64 instances. When reviewing your CloudWatch metrics, you see that sometimes your Scaling group is running 64 micro instances. The web application is reading and writing to a DyanamoDB-configured backend and configured with 800 Write Capacity units and 800 Read Capacity units. Your customers are complaining that they are experiencing load times when viewing you website. You have investigated the DynamoDB CloudWatch metrics; you are under the provisioned read and Write Capacity units and there is no throttling.
How do you scale your service to improve the load times and ensure the principles of high availability?

A. Change your Auto Scaling group configuration to include multiple AZs
B. Change you Auto Scaling group configuration to include multiple AZs, and increase the number of Read Capacity units in your DynamoDB table by a factor of three, because you will need to be calling DynamoDB from three AZs
C. Add a second load balancer to your Auto Scaling group so that you can support more inbound connections per second
D. Change your Auto Scaling group configuration to use larger instances and include multiple AZs instead of one

Answer: D

NEW QUESTION 44
Which of the following notification endpoints or clients are supported by Amazon Simple Notification Service? Choose 2 answers

A. Email
B. CloudFront distribution
C. File Transfer Protocol
D. Short Message Service
E. Simple Network Management Protocol

Answer: BC

NEW QUESTION 47
A user is trying to setup a recurring Auto Scaling process. The user has setup one process to scale up every day at 8 am and scale down at 7 PM. The user is trying to setup another recurring process which scales up on the 1st of every month at 8 AM and scales down the same day at 7 PM. What will Auto Scaling do in this scenario?

A. Auto Scaling will execute both processes but will add just one instance on the 1st
B. Auto Scaling will add two instances on the 1st of the month
C. Auto Scaling will schedule both the processes but execute only one process randomly
D. Auto Scaling will throw an error since there is a conflict in the schedule of two separate Auto Scaling Processes

Answer: D

NEW QUESTION 50
You are designing a web application that stores static assets in an Amazon Simple Storage Service (S3) bucket. You expect this bucket to immediately receive over 150 PUT requests per second. What should you do to ensure optimal performance?

A. Use multi-part upload.
B. Add a random prefix to the key names.
C. Amazon S3 will automatically manage performance at this scale.
D. Use a predictable naming scheme, such as sequential numbers or date time sequences, in the key names

Answer: B

NEW QUESTION 54
You have an ASP.NET web application running in Amazon Elastic BeanStalk. Your next version of the application requires a third-party Windows installer package to be installed on the instance on first boot and before the application launches.
Which options are possible? Choose 2 answer

A. In the application's Global.asax file, run msiexec.exe to install the package using Process.Start() in the Application_Start event handler
B. In the source bundle's .ebextensions folder, create a file with a .config extensio
C. In the file, under the "packages" section and "msi" package manager, include the package's URL

D. Launch a new Amazon EC2 instance from the AMI used by the environmen
E. Log into the instance, install the package and run syspre
F. Create a new AM
G. Configure the environment to use the new AMI
H. In the environment's configuration, edit the instances configuration and add the package's URL to the "Packages" section
I. In the source bundle's .ebextensions folder, create a "Packages" folde
J. Place the package in the folder

Answer: BC

NEW QUESTION 59
You are deploying an application to track GPS coordinates of delivery in the United States.
Coordinates are transmitted from each delivery truck once every three seconds. You need to design an architecture that will enable realtime processing of these coordinates from multiple consumers. Which service should you use to implement data ingestion?

A. Amazon Kinesis
B. AWS Data Pipeline
C. Amazon AppStream
D. Amazon Simple Queue Service

Answer: A

NEW QUESTION 62
Which of the following are true regarding AWS Cloud Trail? Choose 3 answers

A. Cloudtrail is enabled globally
B. Cloudtrail is enabled by default
C. Cloudtrail is enabled on a per-region basis
D. Cloudtrail is enabled on a per-service basis
E. Logs can be delivered to a single Amazon S3 bucket for aggregation
F. Logs can only be processes and delivered to the region in which they are generated

Answer: ACE

NEW QUESTION 63
A user has created an ELB with Auto Scaling. Which of the below mentioned offerings from ELB helps
the user to stop sending new requests traffic from the load balancer to the EC2 instance when the instance is being deregistered while continuing in-flight requests?

A. ELB sticky session
B. ELB deregistration check
C. ELB connection draining
D. ELB auto registration Off

Answer: C

NEW QUESTION 65
When will you incur costs with an Elastic IP address (EIP)?

A. When an EIP is allocated.
B. When it is allocated and associated with a running instance.
C. When it is allocated and associated with a stopped instance.
D. Costs are incurred regardless of whether the EIP is associated with a running instance.

Answer: C

NEW QUESTION 69
A user has launched an EC2 instance from an instance store backed AMI. The infrastructure team
wants to create an AMI from the running instance. Which of the below mentioned steps will not be performed while creating the AMI?

A. Define the AMI launch permissions
B. Upload the bundled volume
C. Register the AMI
D. Bundle the volume

Answer: A

NEW QUESTION 71
A company needs to monitor the read and write IOPs metrics for their AWS MySQL RDS instance and
send real-time alerts to their operations team. Which AWS services can accomplish this? Choose 2 answers

A. Amazon Simple Email Service
B. Amazon CloudWatch
C. Amazon Simple Queue Service
D. Amazon Route 53
E. Amazon Simple Notification Service

Answer: BE

NEW QUESTION 75
The Trusted Advisor service provides insight regarding which four categories of an AWS account?

A. Security, fault tolerance, high availability, and connectivity
B. Security, access control, high availability, and performance
C. Performance, cost optimization, security, and fault tolerance
D. Performance, cost optimization, access control, and connectivity

Answer: C

NEW QUESTION 77
You have been asked to handle a large data migration from multiple Amazon RDS MySQL instances to
a DynamoDB table. You have been given a short amount of time to complete the data migration. What will allow you to complete this complex data processing workflow?

A. Create an Amazon Kinesis data stream, pipe in all of the Amazon RDS data, and direct data toward DynamoDB table
B. Write a script in you language of choice, install the script on an Amazon EC2 instance, and then use Auto Scaling groups to ensure that the latency of the mitigation pipelines never exceeds four seconds in any 15-minutes period.
C. Write a bash script to run on your Amazon RDS instance that will export data into DynamoDB
D. Create a data pipeline to export Amazon RDS data and import the data into DynamoDB

Answer: D

NEW QUESTION 78
You have a video Trans coding application running on Amazon EC2. Each instance pools a queue to find out which video should be Trans coded, and then runs a Trans coding process.
If this process is interrupted, the video will be Trans coded by another instance based on the queuing system. You have a large backlog of videos which need to be Trans coded and would like to reduce this backlog by adding more instances. You will need these instances only until the backlog is reduced. Which type of Amazon EC2 instance should you use to reduce the backlog in the most cost- effective way?

A. Dedicated instances
B. Spot instances
C. On-demand instances
D. Reserved instances

Answer: B

NEW QUESTION 81
What is one key difference between an Amazon EBS-backed and an instance-store backed instance?

A. Amazon EBS-backed instances can be stopped and restarted
B. Instance-store backed instances can be stopped and restarted
C. Auto scaling requires using Amazon EBS-backed instances
D. Virtual Private Cloud requires EBS backed instances

Answer: A

NEW QUESTION 82
A data engineer is about to perform a major upgrade to the DDL contained within an Amazon Redshift cluster to support a new data warehouse application. The upgrade scripts will include user permission updates, view and table structure changes as well as additional loading and data manipulation tasks. The data engineer must be able to restore the database to its existing state in the event of issues.
Which action should be taken prior to performing this upgrade task?

A. Run an UNLOAD command for all data in the warehouse and save it to S3
B. Create a manual snapshot of the Amazon Redshift cluster
C. Make a copy of the automated snapshot on the Amazon Redshift cluster
D. Call the wait For Snap Shot Available command from either the AWS CLI or an AWS SDK

Answer: B

NEW QUESTION 86
A customer is collecting clickstream data using Amazon kinesis and is grouping the events by IP address into 5-minute chunks stored in Amazon S3.
Many analysts in the company use Hive on Amazon EMR to analyze this data. Their queries always reference a single IP address. Data must be optimized for querying based on UP address using Hive running on Amazon EMR. What is the most efficient method to query the data with Hive?

A. Store an index of the files by IP address in the Amazon DynamoDB metadata store for EMRFS
B. Store the Amazon S3 objects with the following naming scheme: bucketname/source=ip_address/year=yy/month=mm/day=dd/hour=hh/filename
C. Store the data in an HBase table with the IP address as the row key
D. Store the events for an IP address as a single file in Amazon S3 and add metadata with key:Hive_Partitioned_IPAddress

Answer: B

NEW QUESTION 88

An online gaming company uses DynamoDB to store user activity logs and is experiencing throttled writes on the company's DynamoDB tables. The company is NOT consuming close to the provisioned capacity. The table contains a large number of items and is partitioned on user and sorted by date. The table is 200GB and is currently provisioned at 10K WCU and 20K RCU.
Which two additional pieces of information are required to determine the cause of the throttling? (Select Two.)

A. The structured of any GSIs that have been defined on the table
B. CloudWatch data showing consumed and provisioned write capacity when writes are being throttled
C. Application-level metric showing the average item size and peak update rates for each attribute
D. The structure of any LSIs that have been defined on the table
E. The maximum historical WCU and RCU for the table

Answer: BD

NEW QUESTION 89
A large grocery distributor receives daily depletion reports from the field in the form of gzip archives of CSV files uploading to Amazon S3. The files range from 500MB to 5GB. These files are processes daily by an EMR job.
Recently it has been observed that the file sizes vary, and the EMR jobs take too long. The distributor needs to tune and optimize the data processing workflow with this limited information to improved the performance of the EMR job.
Which recommendation should an administrator provide?

A. Reduce the HDFS block size to increase the number of task processors
B. Use bzip2 or Snappy rather than gzip for the archives
C. Decompress the gzip archives and store the data as CSV files
D. Use Avro rather than gzip for the archives

Answer: B

NEW QUESTION 90
A company operates an international business served from a single AWS region. The company wants to expand into a new country. The regulator for that country requires the Data Architect to maintain a log of financial transactions in the country within 24 hours of production transaction. The production application is latency insensitive. The new country contains another AWS region.
What is the most cost-effective way to meet this requirement?

A. Use CloudFormation to replicate the production application to the new region
B. Use Amazon CloudFront to serve application content locally in the country; Amazon CloudFront logs will satisfy the requirement
C. Continue to serve customers from the existing region while using Amazon Kinesis to stream transaction data to the regulator
D. Use Amazon S3 cross-region replication to copy and persist production transaction logs to a budget the new country's region

Answer: D

NEW QUESTION 93
An organization is designing an application architecture. The application will have over 100 TB of data and will support transactions that arrive at rates from hundreds per second to tens of thousands per second, depending on the day of the week and time of day. All transaction data must be durably and reliably stored. Certain read operations must be performed with strong consistency.
Which solutions meets these requirements?

A. Use Amazon DynamoDB as the data store and use strong consistent reads when necessary
B. Use an Amazon Relational Database Service (RDS) instance sized to meet the maximum transaction rate and with the High Availability option enabled.
C. Deploy a NoSQL data store on top of an Amazon Elastic MapReduce (EMR) cluster, and select the HDFS High Durability option.
D. Use Amazon Redshift with synchronous replication to Amazon Simple Storage Service (S3) and row- level locking for strong consistency.

Answer: A

NEW QUESTION 95
A company's social media manager requests more staff on the weekends to handle an increase in customer contacts from a particular region. The company needs a report to visualize the trends on weekends over the past 6 months using QuickSight.
How should the data be represented?

A. A line graph plotting customer contacts v
B. time, with a line for each region
C. A pie chart per region plotting customer contacts per day of week
D. A map of the regions with a heatmap overlay to show the volume of customer contacts
E. A bar graph plotting region vs volume of social media contacts

Answer: A

NEW QUESTION 99
An organization uses Amazon Elastic MapReduce (EMR) to process a series of extract-transform-load (ETL) steps that run in sequence. The output of each step must be fully processed in subsequent steps but will not be retained.
Which of the following techniques will meet this requirement most efficiently?

A. Use the EMR File System (EMRFS) to store the outputs from each step as objects in Amazon Simple Storage Service (S3).
B. Use the s3n URI to story the data to be processes as objects in Amazon S3.
C. Define the ETL steps as separate AWS Data Pipeline activities.
D. Load the data to be processed into HDFS and then write the final output to Amazon S3.

Answer: A

NEW QUESTION 103
A customer has an Amazon S3 bucket. Objects are uploaded simultaneously by a cluster of servers from multiple streams of data. The customer maintains a catalog of objects uploaded in Amazon S3 using an Amazon DynamoDB table. This catalog has the following fields StreamName, TimeStamp, and ServerName, TimeStamp, and ServerName, from which ObjectName can be obtained.
The customer needs to define the catalog to support querying for a given stream or server within a defined time range.
Which DynamoDB table scheme is most efficient to support these queries?

A. Define a Primary Key with ServerName as Partition Key and TimeStamp as Sort Ke
B. Don NOT define a Secondary Index or Global Secondary Index.
C. Define a Primary Key with StreamName as Partition Key and TimeStamp followed by ServerName as Sort Ke
D. Define a Global Secondary Index with ServerName as Partition Key and TimeStamp followed by StreamName.
E. Define a Primary Key with ServerName as Partition Ke
F. Define a Local Secondary Index with StreamName as Partition Ke
G. Define a Global Secondary Index with TimeStamp as Partition Key.
H. Define a Primary Key with ServerName as Partition Ke
I. Define a Local Secondary Index with TimeStamp as Partition Ke
J. Define a Global Secondary Index with StreamName as Partition key and TimeStamp as Sort Key.

Answer: A

NEW QUESTION 106
A company is using Amazon Machine Learning as part of a medical software application. The application will predict the most likely blood type for a patient based on a variety of other clinical tests that are available when blood type knowledge is unavailable.
What is the appropriate model choice and target attribute combination for the problem?
A. Multi-class classification model with a categorical target attribute
B. Regression model with a numeric target attribute
C. Binary Classification with a categorical target attribute
D. K-Nearest Neighbors model with a multi-class target attribute

Answer: C

NEW QUESTION 107
A data engineer in a manufacturing company is designing a data processing platform that receives a
large volume of unstructured data. The data engineer must populate a well- structured star schema in Amazon Redshift.
What is the most efficient architecture strategy for this purpose?

A. Transform the unstructured data using Amazon EMR and generate CSV dat
B. COPY data into the analysis schema within Redshift.
C. Load the unstructured data into Redshift, and use string paring functions to extract structured data for inserting into the analysis schema.
D. When the data is saved to Amazon S3. Use S3 Event Notifications and AWS Lambda to transform the file conten
E. Insert the data into the analysis schema on Redshift.
F. Normalize the data using an AWS Marketplace ETL tool persist the result to Amazon S3 and use AWS Lambda to INSERT the data into Redshift.

Answer: B

NEW QUESTION 111
An organization uses a custom map reduce application to build monthly reports based on many small
data files in an Amazon S3 bucket. The data is submitted from various business units on a frequent
but unpredictable schedule. As the dataset continues to grow, it becomes increasingly difficult to process all of the data in one day. The organization has scaled up its Amazon EMR cluster, but other optimizations could improve performance.
The organization needs to improve performance minimal changes to existing processes and applications.
What action should the organization take?
A. Use Amazon S3 Event Notifications and AWS Lambda to create a quick search file index in DynamoDB.
B. Add Spark to the Amazon EMR cluster and utilize Resilient Distributed Datasets in-memory.
C. Use Amazon S3 Event Notifications and AWS Lambda to index each file into an Amazon Elasticsearch Service cluster.
D. Schedule a daily AWS Data Pipeline process that aggregates content into larger files using S3DistCp.
E. Have business units submit data via Amazon Kinesis Firehose to aggregate data hourly into Amazon S3.

Answer: A

NEW QUESTION 113
A company needs a churn prevention model to predict which customers will NOT review their yearly
subscription to the company's service. The company plans to provide these customers with a promotional offer. A binary classification model that uses Amazon Machine Learning is required. On which basis should this binary classification model be built?

A. User profiles (age, gender, income, occupation)
B. Last user session
C. Each user time series events in the past 3 months
D. Quarterly results

Answer: C

NEW QUESTION 114
A company that provides economics data dashboards needs to be able to develop software to display
rich, interactive, data-driven graphics that run in web browsers and leverages the full stack of web standards (HTML, SVG and CSS).
Which technology provides the most appropriate for this requirement?

A. D3.js
B. Python/Jupyter
C. R Studio
D. Hue

Answer: C

NEW QUESTION 116
A customer needs to determine the optimal distribution strategy for the ORDERS fact table in its
Redshift schema. The ORDERS table has foreign key relationships with multiple dimension tables in this schema.
How should the company determine the most appropriate distribution key for the ORDRES table?

A. Identity the largest and most frequently joined dimension table and ensure that it and the ORDERS table both have EVEN distribution
B. Identify the target dimension table and designate the key of this dimension table as the distribution key of the ORDERS table
C. Identity the smallest dimension table and designate the key of this dimension table as the distribution key of ORDERS table
D. Identify the largest and most frequently joined dimension table and designate the key of this dimension table as the distribution key for the orders table

Answer: D

NEW QUESTION 119
A data engineer is running a DWH on a 25-node Redshift cluster of a SaaS service. The data engineer
needs to build a dashboard that will be used by customers. Five big customers represent 80% of usage, and there is a long tail of dozens of smaller customers.
The data engineer has selected the dashboarding tool.
How should the data engineer make sure that the larger customer workloads do NOT interfere with the smaller customer workloads?

A. Apply query filters based on customer-id that can NOT be changed by the user and apply distribution keys on customer id
B. Place the largest customers into a single user group with a dedicated query queue and place the rest of the customer into a different query queue
C. Push aggregations into an RDS for Aurora instanc
D. Connect the dashboard application to Aurora rather than Redshift for faster queries
E. Route the largest customers to a dedicated Redshift cluster, Raise the concurrency of the multi- tenant Redshift cluster to accommodate the remaining customers

Answer: D

NEW QUESTION 120
A solutions architect works for a company that has a data lake based on a central Amazon S3 bucket.
The data contains sensitive information. The architect must be able to specify exactly which files each user can access. Users access the platform through SAML federation Single Sign On platform.
The architect needs to build a solution that allows fine grained access control, traceability of access to the objects, and usage of the standard tools (AWS Console, AWS CLI) to access the data.
Which solution should the architect build?

A. Use Amazon S3 Server-Side Encryption with AWS KMS-Managed Keys for strong data.Use AWS KMS to allow access to specific elements of the platfor
B. Use AWS CloudTrail for auditing
C. Use Amazon S3 Server-Side Encryption with Amazon S3 Managed Key
D. Set Amazon S3ACI to allow access to specific elements of the platfor
E. Use Amazon S3 to access logs for auditing
F. Use Amazon S3 Client-Side Encryption with Client-Side Master Ke
G. Set Amazon S3 ACI to allow access to specific elements of the platfor
H. Use Amazon S3 access logs for auditing
I. Use Amazon S3 Client-Side Encryption with AWS KMS-Managed keys for storing data.Use AMS KWS to allow access to specific elements of the platfor
J. Use AWS CloudTrail for auditing

Answer: B

NEW QUESTION 125
An administrator is deploying Spark on Amazon EMR for two distinct use cases: machine learning
algorithms and ad hoc querying. All data will be stored in Amazon S3. Two separate clusters for each use case will be deployed. The data volumes on Amazon S3 are less than 10 GB.
How should the administrator align instance types with the cluster's purpose?

A. Machine Learning on C instance types and ad-hoc queries on R instance types
B. Machine Learning on R instance types and ad-hoc queries on G2 instance types
C. Machine Learning on T instance types and ad-hoc queries on M instance types
D. Machine Learning on D instance types and ad-hoc queries on I instance types

Answer: D

NEW QUESTION 128
A company hosts a portfolio of e-commerce websites across the Oregon, N.Virginia, Ireland and
Sydney AWS regions. Each site keeps log files that captures user behavior. The company has built an application that generates batches of product recommendations with collaborative filtering in Oregon. Oregon was selected because the flagship site is hosted there and provides the largest collection of data to train machine learning models against. The other regions do NOT have enough historic data to train accurate machine learning models.
Which set of data processing steps improves recommendations for each region?

A. Use the e-commerce application in Oregon to write replica log files in each other region
B. Use Amazon S3 bucket replication to consolidate log entries and builds a single model in Oregon
C. Use Kinesis as a butler for web logs and replicate logs to the Kinesis streams of a neighboring region

D. Use the CloudWatch Logs agent to consolidate logs into a single CloudWatch logs group

Answer: C

NEW QUESTION 133
A media advertising company handles a large number of real-time messages sourced from over 200
websites. The company's data engineer needs to collect and process records in real time for analysis using Spark Streaming on Amazon Elastic MapReduce (EMR). The data engineer needs to fulfill a corporate mandate to keep ALL raw messages as they are received as a top priority.
Which Amazon Kinesis configuration meets these requirements?

A. Publish messages to Amazon Kinesis Firehose backed by Amazon Simple Storage Service (S3). Pull messages off Firehose with Spark Streaming in parallel to persistence to Amazon S3
B. Publish messages to Amazon Kinesis Stream
C. Pull messages off Stream with Spark Streaming in parallel to AWS messages from Streams to Firehose backed by Amazon Simple Storage Service (S3)
D. Publish messages to Amazon Kinesis Firehose backed by Amazon Simple Storage (S3). Use AWS Lambda messages from Firehose to Streams for processing with Spark Streaming
E. Publish messages to Amazon Kinesis Streams, pull messages off with Spark Streaming and write data new data to Amazon Simple Storage Service (S3) before and after processing

Answer: D

NEW QUESTION 138
An administrator needs to manage a large catalog of items from various external sellers. The administration needs to determine if the items should be identified as minimally dangerous, dangerous or highly dangerous based on their textual description. The administrator already has some items with the danger attribute, but receives hundreds of new item descriptions every day without such classification.
The administrator has a system that captures dangerous goods reports from customer support team or from user feedback. What is a cost –effective architecture to solve this issue?

A. Build a set of regular expression rules that are based on the existing example
B. And run them on the DynamoDB streams as every new item description is added to the system.
C. Build a kinesis Streams process that captures and marks the relevant items in the dangerous goods reports using a Lambda function once more than two reports have been filed.
D. Build a machine learning model to properly classify dangerous goods and run it on the DynamoDB streams as every new item description is added to the system.
E. Build a machine learning model with binary classification for dangerous goods and run it on the DynamoDB streams as every new item description is added to the system.

Answer: C

NEW QUESTION 139
A new algorithm has been written in Python to identify SPAM e-mails. The algorithm analyzes the free text contained within a sample set of 1 million e-mails stored on Amazon S3. The algorithm must be scaled across a production of 5 PB, which also resides in Amazon S3 storage
Which AWS service strategy is best for this use case?

A. Copy the data into Amazon ElasticCache to perform text analysis on the in-memory data and export the results of the model into Amazon machine learning
B. Use Amazon EMR to parallelize the text analysis tasks across the cluster using a streaming program step
C. Use Amazon Elasticsearch service to store the text and then use the Python Elastic search client to run analysis against the text index
D. Initiate a python job from AWS Data pipeline to run directly against the Amazon S3 text files

Answer: C

Explanation:
Reference: https://aws.amazon.com/blogs/database/indexing-metadata-in-amazon-elasticsearch- service-using-aws-lambda-and-python/

NEW QUESTION 141
A travel website needs to present a graphical quantitative summary of its daily bookings to website
visitors for marketing purposes. The website has millions of visitors per day, but wants to control costs by implementing the least-expensive solution for this visualization. What is the most cost-
effective solution?

A. Generate a static graph with a transient EMR cluster dail
B. And store it in Amazon S3
C. Generate a graph using MicroStrategy backed by a transient EMR cluster
D. Implement a Jupyter front-end provided by a continuously running EMR cluster leveraging spot instances for task nodes
E. Implement a Zeppelin application that runs on a long-running EMR cluster

Answer: C

NEW QUESTION 142
A system needs to collect on-premises application spools files into a persistent storage layer in AWS.
Each spool file is 2 KB. The application generates 1 M files per hour. Each source file is automatically deleted from the local server after one hour. What is the most cost-efficient option to meet these requirements?

A. Write file contents to an Amazon DynamoDB table
B. Copy files to Amazon S3 standard storage
C. Write file content to Amazon ElastiCache
D. Copy files to Amazon S3 infrequent Access storage

Answer: A

NEW QUESTION 145
There are thousands of text files on Amazon S3. The total size of the files is 1 PB. The files contain
retail order information for the past 2 years. A data engineer needs to run multiple interactive queries to manipulate the data. The data Engineer has AWS access to spin up an Amazon EMR cluster. The data Engineer needs to use an application on the cluster to process this data and return the results in interactive time frame. Which application on the cluster should be the data engineer use?

A. Oozie
B. Apache Pig with Tachyon
C. Apache Hive
D. Presto

Answer: D

NEW QUESTION 148
An administrator is processing events in near real-time using Kinesis streams and Lambda. Lambda
intermittently fails to process batches from one of the shards due to a 5 –minute time limit. What is a possible solution for this problem?

A. Add more Lambda functions to improve concurrent batch processing
B. Reduce the batch size that lambda is reading from the stream
C. Ignore and skip events that are older than 5 minutes and put them to Dead Letter Queue (DLQ)
D. Configure Lambda to read from fewer shards in parallel

Answer: BD

NEW QUESTION 149
An enterprise customer is migrating to Redshift and is considering using dense storage nodes in its
Redshift cluster. The customer wants to migrate 50 TB of data. The customer's query patterns involve performing many joins with thousands of rows. The customer needs to know how many nodes are needed in its target Redshift cluster. The customer has a limited budget and needs to avoid performing tests unless absolutely needed. Which approach should this customer use?

A. Start with many small nodes
B. Start with fewer large nodes
C. Have two separate clusters with a mix of small and large nodes
D. Insist on performing multiple tests to determine the optimal configuration

Answer: D

NEW QUESTION 150
A web-hosting company is building a web analytics tools to capture clickstream data from all of the
websites hosted within its platform and to provide near-real-time business intelligence. This entire system is built on AWS services. The web-hosting company is interested in using Amazon kinesis to collect this data and perform sliding window analytics. What is the most reliable and fault-tolerant technique to get each website to send data to Amazon Kinesis with every click?

A. After receiving a request each web server sends it to Amazon kinesis using the Amazon kinesis PutRecord APL Use the SessionID as a parturition key and set up a loop to retry until a successresponse is received
B. After receiving a request each web server sends it to Amazon kinesis using the Amazon Kinesis Producer Library addRecord method
C. Each web server bluffers the request until the count reaches 500 and sends them to Amazon kinesis using the Amazon kinesis PutRecord API call
D. After receiving a request each web server sends it to Amazon Kinesis using the Amazon kinesis PutRecord AP
E. Use the exponential back off algorithm for retries until a successful response is received

Answer: A

NEW QUESTION 153
An organization needs a data store to handle the following data types and access patterns
• Faceting
• Search
• Flexible schema (JSON) and fixed schema
• Noise word elimination
Which data store should the organization choose?

A. Amazon Relational Database Service (RDS)
B. Amazon Redshift
C. Amazon DynamoDB
D. Amazon Elasticsearch Service

Answer: C

NEW QUESTION 156
.....

THANKS FOR TRYING THE DEMO OF OUR PRODUCT

Visit Our Site to Purchase the Full Set of Actual AWS-Certified-Big-Data-Specialty Exam Questions With Answers.

We Also Provide Practice Exam Software That Simulates Real Exam Environment And Has Many Self-Assessment Features. Order the AWS-Certified-Big-Data-Specialty Product From:

https://www.2passeasy.com/dumps/AWS-Certified-Big-Data-Specialty/

Money Back Guarantee

AWS-Certified-Big-Data-Specialty Practice Exam Features:

* AWS-Certified-Big-Data-Specialty Questions and Answers Updated Frequently

* AWS-Certified-Big-Data-Specialty Practice Questions Verified by Expert Senior Certified Staff

* AWS-Certified-Big-Data-Specialty Most Realistic Questions that Guarantee you a Pass on Your FirstTry

* AWS-Certified-Big-Data-Specialty Practice Test Questions in Multiple Choice Formats and Updatesfor 1 Year