Cloudera

CDP-500

Cloudera Administrator Cloud Certification Exam

For More Information – Visit link below:

https://www.examsempire.com/

Product Version

1. **Up to Date products, reliable and verified.**
2. **Questions and Answers in PDF Format.**

Latest Version: 6.0

Question: 1

You are onboarding a new Cloudera Data Platform (CDP) environment in AWS. Your security team requires all network traffic to be encrypted both in transit and at rest. Which of the following options are BEST suited for achieving this goal with minimal administrative overhead?

A. Enable encryption at rest using AWS KMS managed keys for all S3 buckets and use TLS for all communication between CDP components. Configure all EC2 instances with AWS Systems Manager Agent for patching and vulnerability scanning.
B. Implement IPsec tunnels between all subnets in the VPC and use self-signed certificates for internal communication. Enable server-side encryption (SSE) with S3 managed keys (SSE-S3).
C. Utilize AWS CloudHSM for all key management and implement mutual TLS authentication for all CDP components. Enable S3 bucket encryption with client-side encryption using keys managed in CloudHSM.
D. Rely solely on the default encryption provided by S3 and EC2. Configure network ACLs for basic security.
E. Enable encryption at rest using AWS KMS Customer Managed Keys (CMK) for all S3 buckets and use TLS for all communication between CDP components. Additionally, implement security groups with the principle of least privilege for network access control.

Answer: E

Explanation:
Using KMS CMKs gives you control and auditability over the encryption keys. TLS ensures encryption in transit. Security Groups based on the principle of least privilege provide robust network access control. Options B and C introduce unnecessary complexity. Option A uses AWS managed keys which give less control. Option D is insufficient for security.

Question: 2

A.

```
"properties": {
  "hardwareProfile": {
    "vmSize": "Standard_D4s_v3"
  },
  "zones": ["1"]
}
```

B.

```
"properties": {
  "hardwareProfile": {
    "vmSize": "Standard_D4s_v3"
  }
}
```

C.

```
"properties": {
  "hardwareProfile": {
    "vmSize": "Standard_D4s_v3"
  },
  "zones": ["zone1"]
}
```

D.

```
"properties": {
  "hardwareProfile": {
    "vmSize": "Standard_D4s_v3"
  },
  "availabilitySet": {
    "id": "/subscriptions//resourceGroups//providers/Microsoft.Compute/availabilitySets/"
  }
}
```

E.

```
"properties": {
  "hardwareProfile": {
    "vmSize": "Standard_D4s_v3"
  },
  "zones": ["1", "2", "3"]
}
```

Answer: A

Explanation:
The 'zones' property in the ARM template should be configured with an array of strings representing the availability zones. Using correctly specifies that the VM should be deployed in zone 1. Zone numbers are string types in the ARM template definition. Using availability sets, while valid for HA, doesn't directly specify zone placement as requested by the question. Option E is possible to achieve high availability, but the question asks for a specific zone.

Question: 3

You are configuring network security for a CDP environment in GCP You need to ensure that only specific CIDR blocks can access the Cloudera Manager UI. Which of the following GCP networking constructs is MOST appropriate for implementing this requirement?

A. Cloud DNS
B. VPC Service Controls
C. Firewall Rules
D. Cloud NAT
E. Cloud Armor

Answer: C

Explanation:
GCP Firewall Rules allow you to control network traffic based on source IP addresses (CIDR blocks), destination IP addresses, protocols, and ports. VPC Service Controls are for broader perimeter security. Cloud DNS manages DNS records. Cloud NAT provides NAT services. Cloud Armor protects web applications from attacks.

Question: 4

A new regulatory requirement mandates that you implement data masking on sensitive PII data stored in Hive tables within your CDP deployment on AWS. Which of the following approaches is the MOST efficient and compliant way to achieve this, assuming you want to minimize code changes and maintain data usability for authorized users?

A. Implement a custom IJDF (User Defined Function) in Hive to mask the data on read.
B. Export the Hive tables to S3, use AWS Glue to transform and mask the data, and then re-import the data back into Hive.
C. Utilize Apache Ranger's data masking policies to dynamically mask the data based on user roles and policies.
D. Encrypt the entire Hive tables using Hive's built-in encryption features.
E. Create views on top of the Hive tables, masking the data in the views, and restrict access to the underlying tables.

Answer: C

Explanation:
Apache Rangers data masking policies provide a centralized and policy-driven approach to data masking. This avoids code changes and allows for dynamic masking based on user roles, fulfilling the requirements of the question. Other options involve more manual intervention or less flexible control. Encrypting the entire table does not meet the specific data masking requirement. Using a custom IJDF is less manageable than Ranger policies. Views restrict, rather than mask the data.

Question: 5

You are setting up a hybrid CDP deployment, with some workloads running on-premise and others in Azure. You need to configure secure connectivity between the two environments. Which of the following options provides the MOST secure and reliable connection for data transfer and application communication?

A. Public internet with VPN
B. Azure ExpressRoute
C. Public internet without VPN
D. SSH Tunneling
E. Azure Storage Explorer

Answer: B

Explanation:
Azure ExpressRoute provides a dedicated, private network connection to Azure, offering better security, reliability, and performance compared to public internet connections. VPN is more secure than just the public internet. SSH Tunneling isn't suited for production or large amount of data transfer. Azure Storage Explorer is a utility to access azure storage account data from client machine.

Question: 6

You have a requirement to monitor the CPU utilization of all the worker nodes in your CDP cluster running on AWS. Which of the following monitoring tools, when integrated with Cloudera Manager, provides the MOST comprehensive and scalable solution?

A. CloudWatch
B. Ganglia
C. Nagios
D. Prometheus with Grafana
E. AWS x-Ray

Answer: D

Explanation:
Prometheus, coupled with Grafana for visualization, offers a powerful and scalable solution for monitoring time-series data, including CPU utilization. It's well-suited for dynamic cloud environments and integrates well with containerized deployments (if applicable). CloudWatch is specific to AWS, making it not the best choice. Ganglia is older technology, and it doesn't scale as well. Nagios provides host and service monitoring but lacks the deep metrics analysis capabilities. AWS X-Ray is a distributed tracing system, not for server level monitoring

Question: 7

Which of the following are critical prerequisites before onboarding a Cloudera Data Platform (CDP) environment on a public cloud provider (select all that apply)?

A. Properly configured VPC/VNet with appropriate subnets and routing.
B. A detailed disaster recovery plan.
C. Appropriate IAM roles/permissions for CDP to access cloud resources.
D. Implementation of a continuous integration/continuous delivery (CIICD) pipeline.
E. A well-defined security strategy, including network security, encryption, and access controls.

Answer: A,C,E

Explanation:
Before deploying CDP, it's crucial to have the network infrastructure (VPC/VNet), cloud access permissions (IAM roles), and security strategy in place. While a DR plan and CI/CD pipeline are important, they are not strictly prerequisites for initial onboarding. They are important for operation but not specifically onboarding.

Question: 8

You are configuring a Cloudera Data Engineering (CDE) service on Azure. You need to set the correct environment variables for Spark applications to access data stored in Azure Data Lake Storage Gen2 (ADLS Gen2). Which of the following environment variables (with their correct values) are REQUIRED for Spark to authenticate with ADLS Gen2 using a Service Principal?

A.
```
fs.azure.account.oauth2.client.endpoint=; fs.azure.account.oauth2.client.id=; fs.azure.account.oauth2.client.secret=
```
B.
```
dfs.adls.oauth2.access.token.provider.type=ClientCredential; dfs.adls.oauth2.client.id=; dfs.adls.oauth2.credential=; dfs.adls.oauth2.refresh.url=
```
C.
```
fs.azure.account.key..dfs.core.windows.net=
```
D.
```
spark.hadoop.fs.azure.account.oauth2.client.endpoint=; spark.hadoop.fs.azure.account.oauth2.client.id=;
spark.hadoop.fs.azure.account.oauth2.client.secret=; spark.hadoop.fs.azure.account.oauth2.msi.tenant=
```
E.
```
dfs.adls.oauth2.access.token.provider.type=ClientCredential; dfs.adls.oauth2.client.id=; dfs.adls.oauth2.credential=; dfs.adls.oauth2.refresh.url=;
spark.hadoop.fs.azure.account.oauth2.msi.tenant=
```

Answer: D

Explanation:
Option D uses the correct 'spark.hadoop.fs.azure.account' prefix, which is necessary for Spark to recognize these configuration properties. It also includes the ‘oauth2.msi.tenant’ property for the tenant ID. The property is not required. Option C uses account key which is not optimal. Option B uses deprecated property names. Option A has incorrect property names.

Question: 9

You are auditing the security configuration of a CDP environment running on AWS. You discover that some EC2 instances are using default security groups. What potential security risks does this pose?

A. Increased cost due to higher network traffic.
B. Unnecessary resource consumption.
C. Default security groups typically allow all inbound traffic within the group, potentially exposing instances to unnecessary network access and vulnerabilities.
D. The EC2 instances will be automatically terminated by AWS.
E. Limited network bandwidth.

Answer: C

Explanation:
Default security groups often have overly permissive rules, allowing inbound traffic from any instance within the same security group. This can expose instances to vulnerabilities and unauthorized access if not properly managed. A default security group will not automatically terminate any instance. No cost or resource consumption will affect the network.

Question: 10

You are implementing a data lake on GCP using CDR Data needs to be ingested from various sources with different levels of sensitivity. How can you classify the data based on sensitivity levels and enforce different access control policies using only native GCP services integrated with CDP components?

A. Use Cloud IAM roles to control access to the entire data lake.
B. Employ Cloud Data Loss Prevention (DLP) to identify sensitive data and then use Cloud IAM and BigQuery authorized views to control access.
C. Manually tag the data files with sensitivity labels and write custom scripts to enforce access controls based on these tags.
D. Utilize Cloud Storage bucket ACLs to control access to individual data files.
E. Utilize Cloud IAM roles to control access for some and Cloud Storage bucket ACLs to control access on the rest of data files.

Answer: B

Explanation:
Cloud DLP helps automatically identify sensitive data based on predefined or custom rules. Cloud IAM then lets you control access based on service account, user, groups. BigQuery authorized views allows for more fine grained control on BigQuery datasets. Cloud IAM is helpful but does not provide DLP capabilities. Tagging is manual and not scalable. ACLs are manageable only at storage bucket level.

Thank You for Trying Our Product

Special 16 USD Discount Coupon: NSZUBG3X

Email: support@examsempire.com

Check our Customer Testimonials and ratings available on every product page.

Visit our website.

https://examsempire.com/