

## IAM

Is IAM a global or regional service?

1

- Global (x)
- Regional

From the IAM FAQ (<https://aws.amazon.com/iam/faqs>):

"Q: Can users be defined regionally?

Not initially. Users are global entities, like an AWS account is today. No region is required to be specified when defining user permissions. Users are able to use AWS services in any geographic region."

What data is an IAM Request Context?

2

- Calling principal (x)
- Environment data (IP address, user agent, etc.) (x)
- Resource data (e.g., DynamoDB table name) (x)

From the IAM documentation ([http://docs.aws.amazon.com/IAM/latest/UserGuide/reference\\_policies\\_evaluation-logic.html](http://docs.aws.amazon.com/IAM/latest/UserGuide/reference_policies_evaluation-logic.html)):

"The Request Context

When AWS authorizes a request, information about the request is assembled from several sources:

- Principal (the requester), which is determined based on the secret access key. This might represent the root user, an IAM user, a federated user (via STS), or an assumed role, and includes the aggregate permissions that are associated with that principal.
- Environment data, such as the IP address, user agent, SSL enabled, the time of day, etc. This information is determined from the request.
- Resource data, which pertains to information that is part of the resource being requested. This can include information such as a DynamoDB table name, a tag on an Amazon EC2 instance, etc.

This information is gathered into a *request context*, which is a collection of information that's derived from the request. During evaluation, AWS uses values from the request context to determine whether to allow or deny the request. For example, does the action in the request context match an action in the Action element? If not, the request is denied. Similarly, does the resource in the request context match one of the resources in the Resource element? If not, the request is denied.

This is also how the keys work that you can use in the Condition element. For example, for the following policy fragment, AWS uses the date and time from the current request context for the `aws:CurrentTime` key and then performs the `DateGreaterThan` and `DateLessThan` comparisons.

For example:

```
"Condition": {
  "DateGreaterThan": {
    "aws:CurrentTime": "2013-08-16T12:00:00Z"
  },
  "DateLessThan": {
    "aws:CurrentTime": "2013-08-16T15:00:00Z"
  }
}
```

Policy variables like `${aws:username}` also work like this. In the following policy fragment, AWS gets the user name from the request context and uses it in the policy at the place where the `${aws:username}` occurs.

```
"Resource": [ "arn:aws:s3:::mybucket/${aws:username}/*" ]"
```

Can an IAM Role be assigned to an IAM Group?

3

- 
- No (x)
  - Yes

From the IAM FAQ (<https://aws.amazon.com/iam/faqs/>):

**"Q: Can an IAM role be added to an IAM group?**

Not at this time."

Can't find any documentation on this, but they are really different things. You can assign policies (both inline and managed) to groups, and you can also assign policies to roles, but there's no notion of assigning a group a role to give it policies.

Roles are used to grant temporary credentials to either (a) IAM users that assume the role or (b) EC2 instances.

---

Is it possible to set usage quotas on specific IAM users?

4

- 
- Yes
  - No (x)

From the IAM FAQ (<https://aws.amazon.com/iam/faqs/>):

**"Q: Can I set usage quotas on IAM users?**

No. All limits are on the AWS account as a whole. For example, if your AWS Account has a limit of 20 Amazon EC2 instances, IAM users with EC2 permissions can start instances up to the limit. You cannot limit what an individual user can do."

---

When evaluating IAM policies, are requests allowed or denied by default?

5

- 
- Allowed
  - Denied (x)

From the IAM documentation ([http://docs.aws.amazon.com/IAM/latest/UserGuide/reference\\_policies\\_evaluation-logic.html](http://docs.aws.amazon.com/IAM/latest/UserGuide/reference_policies_evaluation-logic.html)):

"When a request is made, the AWS service decides whether a given request should be allowed or denied. The evaluation logic follows these rules:

- By default, all requests are denied. (In general, requests made using the account credentials for resources in the account are always allowed.)
- An explicit allow overrides this default.
- An explicit deny overrides any allows."

---

When evaluating IAM policies, does an explicit Allow override an explicit Deny?

6

- 
- Yes
  - No (x)

From the IAM documentation ([http://docs.aws.amazon.com/IAM/latest/UserGuide/reference\\_policies\\_evaluation-logic.html](http://docs.aws.amazon.com/IAM/latest/UserGuide/reference_policies_evaluation-logic.html)):

"When a request is made, the AWS service decides whether a given request should be allowed or denied. The evaluation logic follows these rules:

- By default, all requests are denied. (In general, requests made using the account credentials for resources in the account are always allowed.)
- An explicit allow overrides this default.
- **An explicit deny overrides any allows."**

---

Can you switch to an IAM role using the account owner's credentials?

7

- Yes
- No (x)

From the AssumeRole API documentation ([http://docs.aws.amazon.com/STS/latest/APIReference/API\\_AssumeRole.html](http://docs.aws.amazon.com/STS/latest/APIReference/API_AssumeRole.html)):

"You cannot call AssumeRole by using AWS account credentials; access will be denied. You must use IAM user credentials or temporary security credentials to call AssumeRole."

Is it possible to create an IAM role that only allows a user to create EC2 instances in a specific VPC?

8

- Yes (x)
- No

Yes, here's an example of a policy that does it:

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": "ec2:RunInstances",
      "Resource": "arn:aws:ec2:us-east-1:417233410536:subnet/*",
      "Condition": {
        "StringEquals": {
          "ec2:Vpc": "arn:aws:ec2:us-east-1:417233410536:vpc/vpc-9c20e4f8"
        }
      }
    },
    {
      "Effect": "Allow",
      "Action": "ec2:RunInstances",
      "Resource": [
        "arn:aws:ec2:us-east-1::image/*",
        "arn:aws:ec2:us-east-1:417233410536:instance/*",
        "arn:aws:ec2:us-east-1:417233410536:volume/*",
        "arn:aws:ec2:us-east-1:417233410536:network-interface/*",
        "arn:aws:ec2:us-east-1:417233410536:key-pair/*",
        "arn:aws:ec2:us-east-1:417233410536:security-group/*",
        "arn:aws:ec2:us-east-1:417233410536:instance/*"
      ]
    }
  ]
}
```

In fact, it's possible to constrain in lots of different ways using a variety of EC2 "condition keys", see here: <http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/ec2-supported-iam-actions-resources.html>.

Are IAM policies evaluated for API requests made with the account owner's credentials?

9

- Yes
- No (x)

From the IAM documentation ([http://docs.aws.amazon.com/IAM/latest/UserGuide/reference\\_policies\\_evaluation-logic.html](http://docs.aws.amazon.com/IAM/latest/UserGuide/reference_policies_evaluation-logic.html)):

"Requests that are made using the credentials of the AWS account owner (the root credentials) for resources in that account are allowed. However, if the request is made using the credentials of an IAM user, or if the request is signed using temporary credentials that are granted by AWS STS, AWS uses the permissions defined in one or more IAM policies to determine whether the user's request is authorized."

Is it possible to use SAML to grant federated users access to AWS APIs (vs SSO access)?

10

- Yes (x)
- No

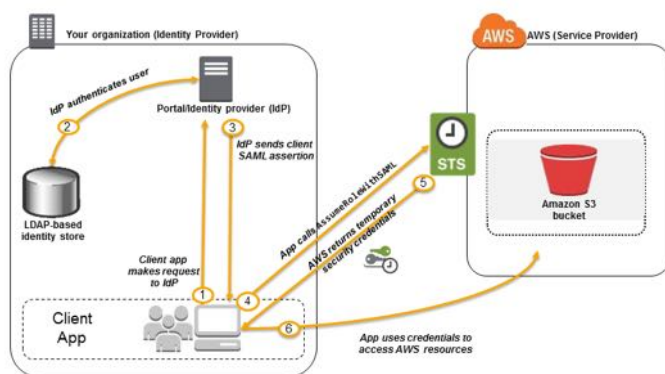
[Note: Remember that once you have valid temporary credentials you can create an SSO console URL using the AWS federation endpoint; it doesn't matter which API you used to get them.]

As per the IAM documentation ([http://docs.aws.amazon.com/IAM/latest/UserGuide/id\\_roles\\_providers\\_saml.html](http://docs.aws.amazon.com/IAM/latest/UserGuide/id_roles_providers_saml.html)):

### "Using SAML-Based Federation for Access to AWS

Imagine that in your organization, you want to provide a way for users to copy data from their computers to a backup folder. You build an application that users can run on their computers. On the back end, the application reads and writes objects in an S3 bucket. Users don't have direct access to AWS. Instead, the application communicates with an identity provider (IdP) to authenticate the user. The IdP gets the user information from your organization's identity store (such as an LDAP directory) and then generates a SAML assertion that includes authentication and authorization information about that user. The application then uses that assertion to make a call to the AssumeRoleWithSAML API to get temporary security credentials. The app can then use those credentials to access a folder in the S3 bucket that's specific to the user.

The following diagram illustrates the flow.



1. A user in your organization uses a client app to request authentication from your organization's IdP.
2. The IdP authenticates the user.
3. The IdP constructs and sends a SAML assertion to the client app.
4. The client app calls the AWS STS AssumeRoleWithSAML API, passing the ARN of the SAML provider, the ARN of the role to assume, and the SAML assertion that was provided by the IdP in the previous step.
5. The API response to the client app includes temporary security credentials.
6. The client app uses the temporary security credentials to call Amazon S3 APIs."

Does a new IAM user have any API permissions by default?

11

- Yes
- No (x)

From the IAM FAQ (<https://aws.amazon.com/iam/faqs/>):

### "Q: How do users call AWS services?

Users can make requests to AWS services using security credentials. A user's ability to call AWS services is governed by explicit permissions. **By default, they have no ability to call service APIs on behalf of the account."**

Which of the following are true about IAM groups?

12

- 
- Groups make it easier to manage user permissions. (x)
  - Groups can be assigned managed IAM policies. (x)
  - Groups can be assigned inline IAM policies. (x)
  - Groups have access keys.
  - Groups can be members of other groups.

From the IAM FAQ (<https://aws.amazon.com/iam/faqs/>):

**"Q: What is a group?"**

A group is a collection of IAM users. Group membership is managed as a simple list:

- Users can be added to or removed from a group.
- A user can belong to multiple groups.
- Groups cannot belong to other groups.
- Groups can be granted permissions using access control policies. This makes it easier to manage permissions for a collection of users, rather than having to manage permissions for each individual user.
- Groups do not have security credentials, and cannot access web services directly; they exist solely to make it easier to manage user permissions. For details, see *Working with Groups and Users*."

---

To which entities can I assign an IAM policy?

13

- 
- Role (x)
  - Group (x)
  - User (x)
  - Certain AWS resource (x)

All of these entities can be assigned a policy. Further, you can also assign a user policies indirectly via group memberships.

---

Can you change the IAM role on a running instance?

14

- 
- Yes
  - No (x)

From the IAM FAQ (<https://aws.amazon.com/iam/faqs/>):

**"Q: Can I associate an IAM role with an already running EC2 instance?"**

No. You can associate only one IAM role with an EC2 instance."

---

Can you associate more than one IAM role with an EC2 instance?

15

- 
- Yes
  - No (x)

From the IAM FAQ (<https://aws.amazon.com/iam/faqs/>):

**"Q: Can I associate more than one IAM role with an EC2 instance?"**

No. You can only associate one IAM role with an EC2 instance at this time."

---

What happens when you delete an IAM role that is presently attached to an EC2 instance?

16

- 
- All access to the EC2 instance is denied immediately (x)
  - Nothing, the instance is unaffected
  - You can't delete a role attached to an instance

From the IAM FAQ (<https://aws.amazon.com/iam/faqs/>):

**"Q: What happens if I delete an IAM role that is associated with a running EC2 instance?"**

Any application running on that instance that's using the role will be denied access immediately."

---

Is it possible to enforce MFA authentication for API calls?

17

- 
- Yes (x)
  - No

From the IAM FAQ (<https://aws.amazon.com/iam/faqs/>):

**"Q. How does AWS MFA work?"**

AWS MFA uses an authentication device that continually generates random, six-digit, single-use authentication codes...You can enforce MFA authentication by adding MFA-related restrictions in IAM policies. To access APIs and resources protected in this way, developers can use temporary security credentials and pass optional MFA parameters in their AWS Security Token Service (STS) API requests (the service that grants temporary security credentials. MFA-validated temporary security credentials can be used to call MFA-protected APIs and resources."

---

Can I force federated users (with GetFederationToken credentials) to use MFAs in API calls?

18

- 
- Yes
  - No (x)

From the IAM FAQ (<https://aws.amazon.com/iam/faqs/>):

**"Q. Does MFA-protected API access work for federated users?"**

Customers will not be able to use MFA-protected API access to control access for federated users. The GetFederatedSession API does not accept MFA parameters. Since federated users can't authenticate with AWS MFA devices, they will be unable to access resources designated using MFA-protected API access."

When IAM evaluates an API request it calculates a "request context", which includes a Calling Principal. What is a Calling Principal? 19

- 
- A root user (x)
  - An IAM user (x)
  - A federated user (via STS) (x)
  - An assumed role (x)
  - An IAM group
  - An EC2 instance id
  - A managed role

From the IAM documentation ([http://docs.aws.amazon.com/IAM/latest/UserGuide/reference\\_policies\\_evaluation-logic.html](http://docs.aws.amazon.com/IAM/latest/UserGuide/reference_policies_evaluation-logic.html)):

#### "The Request Context

When AWS authorizes a request, information about the request is assembled from several sources:

- **Principal (the requester), which is determined based on the secret access key. This might represent the root user, an IAM user, a federated user (via STS), or an assumed role, and includes the aggregate permissions that are associated with that principal.**
- Environment data, such as the IP address, user agent, SSL enabled, the time of day, etc. This information is determined from the request.
- Resource data, which pertains to information that is part of the resource being requested. This can include information such as a DynamoDB table name, a tag on an Amazon EC2 instance, etc.

This information is gathered into a *request context*, which is a collection of information that's derived from the request.

What is an IAM External ID and how is it used?

- An identifier that an AWS managed service provides when assuming a role in its customers' accounts (x)
- An identifier of an external system referenced by AWS resources
- Any identifier for an AWS resource that was created externally to AWS

From the IAM documentation ([http://docs.aws.amazon.com/IAM/latest/UserGuide/id\\_roles\\_create\\_for-user\\_externalid.html](http://docs.aws.amazon.com/IAM/latest/UserGuide/id_roles_create_for-user_externalid.html)):

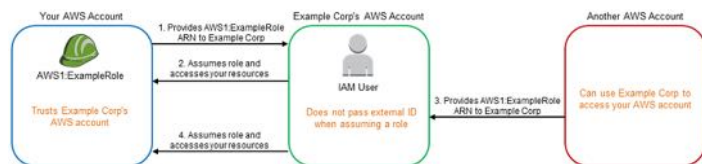
"At times, you need to give a third party access to your AWS resources (delegate access). One important aspect of this scenario is the *External ID*, an optional piece of information that you can use in an IAM role trust policy to designate who can assume the role.

...

In abstract terms, the external ID allows the user that is assuming the role to assert the circumstances in which they are operating. It also provides a way for the account owner to permit the role to be assumed only under specific circumstances. The primary function of the external ID is to address and prevent the "confused deputy" problem.

...

The following diagram illustrates the confused deputy problem.



This diagram assumes the following:

- **AWS1** is your AWS account.
- **AWS1:ExampleRole** is a role in your account. This role's trust policy trusts Example Corp by specifying Example Corp's AWS account as the one that can assume the role.

Here's what happens:

1. When you start using Example Corp's service, you provide the ARN of **AWS1:ExampleRole** to Example Corp.
2. Example Corp uses that role ARN to obtain temporary security credentials to access resources in your AWS account. In this way, you are trusting Example Corp as a "deputy" that can act on your behalf.
3. Another AWS customer also starts using Example Corp's service, and this customer also provides the ARN of **AWS1:ExampleRole** for Example Corp to use. Presumably the other customer learned or guessed the **AWS1:ExampleRole**, which isn't a secret.
4. When the other customer asks Example Corp to access AWS resources in (what it claims to be) its account, Example Corp uses **AWS1:ExampleRole** to access resources in your account

This is how the other customer could gain unauthorized access to your resources. Because this other customer was able to trick Example Corp into unwittingly acting on your resources, Example Corp is now a "confused deputy."

Which are valid IAM principals?

21

- User (x)
- Account (x)
- Service (x)
- Role (x)
- Group
- Bucket

From the IAM policy elements documentation

([http://docs.aws.amazon.com/IAM/latest/UserGuide/reference\\_policies\\_elements.html#Principal](http://docs.aws.amazon.com/IAM/latest/UserGuide/reference_policies_elements.html#Principal)):

"Use the Principal element to specify the user (IAM user, federated user, or assumed-role user), AWS account, AWS service, or other principal entity that is allowed or denied access to a resource.

...

When you create a trust policy for an IAM role that will be assumed by an AWS service, you typically specify the principal using a friendly name for that service..."

What is an EC2 Instance Profile?

22



- 
- A summary of the instance's configuration
  - Used to mount EBS drives to instances at startup
  - Used to associate IAM roles to instance at startup (x)
  - A template for a custom AMI

From the IAM User Guide ([http://docs.aws.amazon.com/IAM/latest/UserGuide/id\\_roles\\_use\\_switch-role-ec2\\_instance-profiles.html](http://docs.aws.amazon.com/IAM/latest/UserGuide/id_roles_use_switch-role-ec2_instance-profiles.html)):

"An instance profile is a container for an IAM role that you can use to pass role information to an EC2 instance when the instance starts."

---

Is it possible to require MFA for API calls?

23

- 
- Yes (x)
  - No

From the IAM documentation ([http://docs.aws.amazon.com/IAM/latest/UserGuide/id\\_credentials\\_mfa\\_configure-api-require.html#MFAProtectedAPI-user-mfa](http://docs.aws.amazon.com/IAM/latest/UserGuide/id_credentials_mfa_configure-api-require.html#MFAProtectedAPI-user-mfa)):

"Scenario: MFA Protection for Access to APIs in the Current Account

In this scenario, you want to make sure that a user in your AWS account can access sensitive API actions only when the user is authenticated using an AWS MFA device.

Imagine that you have account A that contains a group of developers who need to work with EC2 instances. Ordinary developers can work with the instances, but they are not granted permissions for the `ec2:StopInstances` or `ec2:TerminateInstances` actions. You want to limit those "destructive" privileged actions to just a few trusted users, so you add MFA protection to the policy that allows these sensitive Amazon EC2 actions.

In this scenario, one of those trusted users is user Carol. User Alice is an administrator in account A.

1. Alice makes sure that Carol is configured with an AWS MFA device and that Carol knows the ID of the device—the serial number if it's a hardware MFA device, or the device's ARN if it's a virtual MFA device.
2. Alice creates a group named `EC2-Admins` and adds user Carol to the group.
3. Alice attaches the following policy to the `EC2-Admins` group. This policy grants users permission to call the Amazon EC2 `StopInstances` and `TerminateInstances` actions only if the user has authenticated using MFA.  

```
{ "Version": "2012-10-17", "Statement": [{ "Effect": "Allow", "Action": [ "ec2:StopInstances", "ec2:TerminateInstances" ], "Resource": [ "*" ], "Condition": { "Bool": { "aws:MultiFactorAuthPresent": "true" } } } ] }
```
4. If user Carol needs to stop or terminate an Amazon EC2 instance, she (or an application that she is running) calls `GetSessionToken` passing the ID of the MFA device and the current TOTP that Carol gets from her device.
5. User Carol (or an application that Carol is using) uses the temporary credentials provided by `GetSessionToken` to call the Amazon EC2 `StopInstances` or `TerminateInstances` action."

---

Is it possible to require MFA for access to specific AWS resources?

24

- Yes (x)
- No

From the IAM documentation ([http://docs.aws.amazon.com/IAM/latest/UserGuide/id\\_credentials\\_mfa\\_configure-api-require.html#MFAProtectedAPI-user-mfa](http://docs.aws.amazon.com/IAM/latest/UserGuide/id_credentials_mfa_configure-api-require.html#MFAProtectedAPI-user-mfa)):

#### "Scenario: MFA Protection for Resources That Have Resource-based Policies

In this scenario, you are the owner of an S3 bucket, an SQS queue, or an SNS topic and you want to make sure that any user from any AWS account who accesses the resource is authenticated by an AWS MFA device.

This scenario illustrates a way to provide cross-account MFA protection without requiring users to assume a role first. If the user is authenticated by MFA and is able to get temporary security credentials from `GetSessionToken`, and if the user's account is trusted by the resource's policy, the user can access the resource.

Imagine that you are in account A and you create an S3 bucket. You want to grant access to this bucket to users who are in several different AWS accounts, but only if those users are authenticated with MFA.

In this scenario, user Alice is an administrator in account A. User Charlie is an IAM user in account C.

1. In account A, Alice creates a bucket named Account-A-bucket.
2. Alice adds the bucket policy to the bucket. The policy allows any user in account A, account B, or account C to perform the S3 `PutObject` and `DeleteObject` actions in the bucket. The policy includes an MFA condition.  

```
{ "Version": "2012-10-17", "Statement": [{ "Effect": "Allow", "Principal": { "AWS": [ "ACCOUNT-A-ID", "ACCOUNT-B-ID", "ACCOUNT-C-ID" ] }, "Action": [ "s3:PutObject", "s3>DeleteObject" ], "Resource": [ "arn:aws:s3:::ACCOUNT-A-BUCKET-NAME/*" ], "Condition": { "Bool": { "aws:MultiFactorAuthPresent": "true" } } } ] }
```

#### Note

Amazon S3 offers an MFA Delete feature for *root* account access (only). You can enable Amazon S3 MFA Delete when you set the versioning state of the bucket. Amazon S3 MFA Delete cannot be applied to an IAM user, and is managed independently from MFA-protected API access. An IAM user with permission to delete a bucket cannot delete a bucket with Amazon S3 MFA Delete enabled. For more information on Amazon S3 MFA Delete, see [MFA Delete](#).

3. In account C, an administrator makes sure that user Charlie is configured with an AWS MFA device and that he knows the ID of the device—the serial number if it's a hardware MFA device, or the device's ARN if it's a virtual MFA device.
4. In account C, Charlie (or an application that he is running) calls `GetSessionToken`. The call includes the ID or ARN of the MFA device and the current TOTP that Charlie gets from his device.
5. Charlie (or an application that he is using) uses the temporary credentials returned by `GetSessionToken` to call the Amazon S3 `PutObject` action to upload a file to Account-A-bucket.

...

#### Note

The temporary credentials that `AssumeRole` returns won't work in this case because although the user can provide MFA information to assume a role, the temporary credentials returned by `AssumeRole` don't include the MFA information that is required in order to meet the MFA condition in the policy."

If you provide an MFA token when calling `AssumeRole` to get temporary credentials, will those temporary credentials satisfy a `was:MultiFactorAuthPresent` condition?

25

- Yes
- No (x)

From the IAM documentation ([http://docs.aws.amazon.com/IAM/latest/UserGuide/id\\_credentials\\_mfa\\_configure-api-require.html#MFAProtectedAPI-user-mfa](http://docs.aws.amazon.com/IAM/latest/UserGuide/id_credentials_mfa_configure-api-require.html#MFAProtectedAPI-user-mfa)):

"The temporary credentials that `AssumeRole` returns won't work in this case because although the user can provide MFA information to assume a role, the temporary credentials returned by `AssumeRole` don't include the MFA information that is required in order to meet the MFA condition in the policy."

What services support resource-based IAM policies?

26

- EC2 instances
- S3 buckets (x)
- SNS topics (x)
- SQS queues (x)
- Glacier vaults (x)
- OpsWorks stacks (x)
- Elastic Beanstalk environments
- EBS volumes
- VPC resources

From the IAM User Guide ([http://docs.aws.amazon.com/IAM/latest/UserGuide/id\\_roles\\_compare-resource-policies.html](http://docs.aws.amazon.com/IAM/latest/UserGuide/id_roles_compare-resource-policies.html)):

“””

Resource-based policies are supported by the following AWS services and resources:

- **Amazon S3 buckets** – The policy is attached to the bucket, but the policy controls access to both the bucket and the objects in it.
- **Amazon Simple Notification Service (Amazon SNS) topics**
- **Amazon Simple Queue Service (Amazon SQS) queues**
- **Amazon Glacier vaults**
- **AWS OpsWorks stacks**

“””

What is the difference between IAM federation and IAM delegation?

27

- Delegation involves granting access in one account to users in another AWS account; federation refers to granting rights in an AWS account to users external to AWS (x)
- They are synonyms
- Delegation involves granting restricted access to users within an AWS account; federation involves importing users from external sources into AWS

Too much to quote, see the IAM User Guide ([http://docs.aws.amazon.com/IAM/latest/UserGuide/id\\_credentials\\_temp.html](http://docs.aws.amazon.com/IAM/latest/UserGuide/id_credentials_temp.html)) for full details.

What is an IAM Identity Provider?

28

- An entity that manages users outside of AWS and can grant those users permissions to use AWS resources (x)
- An entity that creates IAM users
- An entity that audits IAM users against a source of truth

From the IAM User Guide ([http://docs.aws.amazon.com/IAM/latest/UserGuide/id\\_roles\\_providers.html](http://docs.aws.amazon.com/IAM/latest/UserGuide/id_roles_providers.html)):

**"If you already manage user identities outside of AWS, you can use IAM *identity providers* instead of creating IAM users in your AWS account.** With an identity provider (IdP), you can manage your user identities outside of AWS and give these external user identities permissions to use AWS resources in your account. This is useful if your organization already has its own identity system, such as a corporate user directory. It is also useful if you are creating a mobile app or web application that requires access to AWS resources.

When you use an IdP, you don't have to create custom sign-in code or manage your own user identities; the IdP provides that for you. Your external users sign in through a well-known identity provider, such as Login with Amazon, Facebook, Google, and many others. You can give those external identities permissions to use AWS resources in your account. Identity providers help keep your AWS account secure because you don't have to distribute or embed long-term security credentials, such as IAM access keys, in your application."

What's required to create an SSO URL for federated users?

29

- Temporary credentials obtained from STS (x)
- Calling the AWS federation endpoint (x)
- Constructing a url from the token obtained from the federation endpoint (x)
- Validating the endpoint URL using STS
- Registering the endpoint URL with the console for temporary access
- Using SAML's built-in SSO capability (x)

From the IAM User Guide ([http://docs.aws.amazon.com/IAM/latest/UserGuide/id\\_roles\\_providers\\_enable-console-custom-url.html](http://docs.aws.amazon.com/IAM/latest/UserGuide/id_roles_providers_enable-console-custom-url.html)):

~~~~~

You can write and run code to create a URL that lets users who sign in to your organization's network securely access the AWS Management Console. The URL includes a sign-in token that you get from AWS and that authenticates the user to AWS.

#### Note

If your organization uses an identity provider (IdP) that is compatible with SAML, such as Microsoft's Active Directory Federation Services or open-source Shibboleth, you can set up access to the AWS Management Console without writing code. For details, see [Enabling SAML 2.0 Federated Users to Access the AWS Management Console](#).

To enable your organization's users to access the AWS Management Console, you can create a custom "identity broker" that performs the following steps:

1. Verify that the user is authenticated by your local identity system.
2. Call the AWS Security Token Service (AWS STS) AssumeRole (recommended) or GetFederationToken APIs to obtain temporary security credentials for the user. The credentials are associated with permissions that control what the user can do.
3. Call an AWS federation endpoint and supply the temporary security credentials to get a sign-in token.
4. Construct a URL for the console that includes the token.
5. Give the URL to the user or invoke the URL on the user's behalf.

The URL that the federation endpoint provides is valid for 15 minutes after it is created. The temporary security credentials associated with the URL are valid for the duration you specified when you created them, starting from the time they were created.

#### Important

Keep in mind that the URL grants access to your AWS resources through the AWS Management Console to the extent that you have enabled permissions in the associated temporary security credentials. For this reason, you should treat the URL as a secret. We recommend returning the URL through a secure redirect, for example, by using a 302 HTTP response status code over an SSL connection. For more information about the 302 HTTP response status code, go to RFC 2616, section 10.3.3.

~~~~~

Also

Is it possible to restrict IAM users from creating or modifying tags on an EC2 instance?

30

- Yes (x)
- No

This blog (<http://blogs.aws.amazon.com/security/post/Tx29HCT3ABL7LP3/Resource-level-Permissions-for-EC2-Controlling-Management-Access-on-Specific-Ins>) goes into the details:

"If you choose to use tags as a basis for setting permissions on instances, you will want to restrict which users have permissions to apply and remove tags. For EC2, you will want to restrict which users have permissions to use the `ec2:CreateTags` and `ec2:DeleteTags` actions, so that only these users will be able to change your instance inventory. *Note: We will be enabling tag-specific permissions for `ec2:CreateTag` and `ec2:DeleteTags` in the future, which will enable you to set permissions on a per-tag basis.*"

In the context of IAM identity federation, what is an Identity Broker?

31

- A custom-developed application that authenticates against local identity stores to provide authorized access to AWS resources (x)
- An IAM-provided component that provides seamless integration to external identity providers
- An external component that runs on-premises and integrates with IAM using SAML

From the IAM User Guide ():

To enable your organization's users to access the AWS Management Console, you can create a custom "identity broker" that performs the following steps:

1. Verify that the user is authenticated by your local identity system.
2. Call the AWS Security Token Service (AWS STS) AssumeRole (recommended) or GetFederationToken APIs to obtain temporary security credentials for the user. The credentials are associated with permissions that control what the user can do.
3. Call an AWS federation endpoint and supply the temporary security credentials to get a sign-in token.
4. Construct a URL for the console that includes the token.
5. Give the URL to the user or invoke the URL on the user's behalf.

This blog (<https://aws.amazon.com/blogs/aws/aws-identity-and-access-management-now-with-identity-federation/>) goes into a lot more detail of how to do this with GetFederationToken, including some nice pictures and examples.

## RDS

Is it possible to have a multi-AZ RDS Read Replica?

32

- Yes
- No (x)

From the RDS FAQ (<https://aws.amazon.com/rds/faqs/>):

**"Q: Can I make my Amazon RDS Read Replicas themselves Multi-AZ?"**

Amazon RDS for MySQL and PostgreSQL do not presently support this."

What are the allowable RDS storage types?

33

- Magnetic (Standard) (x)
- General Purpose (SSD) (x)
- Provisioned IOPS (x)

From the RDS documentation ([http://docs.aws.amazon.com/AmazonRDS/latest/UserGuide/CHAP\\_Storage.html](http://docs.aws.amazon.com/AmazonRDS/latest/UserGuide/CHAP_Storage.html)):

**"Amazon RDS provides three storage types: magnetic, General Purpose (SSD), and Provisioned IOPS (input/output operations per second).** They differ in performance characteristics and price, allowing you to tailor your storage performance and cost to the needs of your database. You can create MySQL, PostgreSQL, and Oracle RDS DB instances with up to 6TB of storage and SQL Server RDS DB instances with up to 4TB of storage when using the Provisioned IOPS and General Purpose (SSD) storage types. Existing MySQL, PostgreSQL, and Oracle RDS database instances can be scaled to these new database storage limits without any downtime. For a complete discussion of the different volume types, see the topic Amazon EBS Volume Types."

Is it possible to increase the storage of an existing RDS SQL Server instance?

34

- 
- Yes
  - No (x)

From the RDS FAQ (<http://aws.amazon.com/rds/faqs/#129>):

**"Q: How do I scale the compute resources and/or storage capacity associated with my Amazon RDS Database Instance?**

[...]

Please note that for SQL Server, because of the extensibility limitations of striped storage attached to a Windows Server environment, Amazon RDS does not currently support increasing storage. While we plan to support this functionality in the future, we recommend you to provision storage based on anticipated future storage growth. In the interim, if you need to increase the storage of a SQL Server DB Instance, you will need to export the data, create a new DB Instance with increased storage, and import the data into it. Please refer to the data import guide for SQL Server for more information."

---

Does increasing an RDS instance's storage cause downtime?

35

- Yes
- No (x)

From the RDS User Guide ([http://docs.aws.amazon.com/AmazonRDS/latest/UserGuide/USER\\_PIOPS.StorageTypes.html](http://docs.aws.amazon.com/AmazonRDS/latest/UserGuide/USER_PIOPS.StorageTypes.html)):

"Data storage in Amazon RDS is specified by selecting a storage type and providing a storage size (GB) when you create or modify a DB instance. You can change the type of storage your instance uses by modifying the DB instance, but changing the type of storage in some cases might result in a short outage for the instance. Changing from Magnetic to either General Purpose (SSD) or Provisioned IOPS (SSD) results in an outage. Also, changing from General Purpose (SSD) or Provisioned IOPS (SSD) to Magnetic results in an outage. The outage time is typically 60–120 seconds. For more information about Amazon RDS storage types, see Amazon RDS Storage Types.

**Increasing the allocated storage does not result in an outage.** Note that you cannot reduce the amount of storage once it has been allocated. The only way to reduce the amount of storage allocated to a DB instance is to dump the data out of the DB instance, create a new DB instance with less storage space, and then load the data into the new DB instance."

---

Does modifying an RDS instance's storage type cause downtime?

36

- Yes in all cases
- Yes, but only when changing to or from Magnetic storage (x)
- No

From the RDS User Guide ([http://docs.aws.amazon.com/AmazonRDS/latest/UserGuide/USER\\_PIOPS.StorageTypes.html](http://docs.aws.amazon.com/AmazonRDS/latest/UserGuide/USER_PIOPS.StorageTypes.html)):

"Data storage in Amazon RDS is specified by selecting a storage type and providing a storage size (GB) when you create or modify a DB instance. You can change the type of storage your instance uses by modifying the DB instance, but changing the type of storage in some cases might result in a short outage for the instance. **Changing from Magnetic to either General Purpose (SSD) or Provisioned IOPS (SSD) results in an outage. Also, changing from General Purpose (SSD) or Provisioned IOPS (SSD) to Magnetic results in an outage.** The outage time is typically 60–120 seconds. For more information about Amazon RDS storage types, see Amazon RDS Storage Types."

---

Is RDS Magnetic (Standard) storage single-tenant or is it shared by other customers?

37

- 
- Single-tenant
  - Shared (x)

From the RDS documentation ([http://docs.aws.amazon.com/AmazonRDS/latest/UserGuide/CHAP\\_Storage.html](http://docs.aws.amazon.com/AmazonRDS/latest/UserGuide/CHAP_Storage.html)):

"Amazon RDS provides three storage types: magnetic, General Purpose (SSD), and Provisioned IOPS (input/output operations per second)..."

- Magnetic (Standard) – Magnetic storage, also called standard storage, offers cost-effective storage that is ideal for applications with light or burst I/O requirements. These volumes deliver approximately 100 IOPS on average, with burst capability of up to hundreds of IOPS, and they can range in size from 5 GB to 3 TB, depending on the DB instance engine that you chose. **Magnetic storage is not reserved for a single DB instance, so performance can vary greatly depending on the demands placed on shared resources by other customers.**"

---

Does an RDS instance remain available when its storage capacity or hardware is upgraded?

38

- Yes for both storage capacity and hardware upgrades
- Yes, but only for hardware upgrades
- Yes, but only for storage capacity upgrades (x)
- No

From the RDS documentation ([http://docs.aws.amazon.com/AmazonRDS/latest/UserGuide/CHAP\\_Storage.html](http://docs.aws.amazon.com/AmazonRDS/latest/UserGuide/CHAP_Storage.html)):

**"Q: Will my DB Instance remain available during scaling?"**

The storage capacity allocated to your DB Instance can be increased while maintaining DB Instance availability. However, when you decide to scale the compute resources available to your DB Instance up or down, your database will be temporarily unavailable while the DB Instance class is modified. This period of unavailability typically lasts only a few minutes, and will occur during the maintenance window for your DB Instance, unless you specify that the modification should be applied immediately."

---

How does RDS point-in-time recovery work?

39

- Daily snapshots with applied transaction logs (x)
- Hourly snapshots (recovery limited to last snapshot)
- Weekly snapshots with applied transaction logs

From the RDS FAQ (<http://aws.amazon.com/rds/faqs/#129>):

**"Q: What is the difference between automated backups and DB Snapshots?"**

Amazon RDS provides two different methods for backing up and restoring your DB Instance(s) automated backups and database snapshots (DB Snapshots).

The automated backup feature of Amazon RDS enables point-in-time recovery of your DB Instance. When automated backups are turned on for your DB Instance, Amazon RDS automatically performs a full daily snapshot of your data (during your preferred backup window) and captures transaction logs (as updates to your DB Instance are made). When you initiate a point-in-time recovery, transaction logs are applied to the most appropriate daily backup in order to restore your DB Instance to the specific time you requested. Amazon RDS retains backups of a DB Instance for a limited, user-specified period of time called the retention period, which by default is one day but can be set to up to thirty five days. You can initiate a point-in-time restore and specify any second during your retention period, up to the Latest Restorable Time."

---

When restoring an RDS snapshot or performing an RDS point-in-time recovery, does the new DB instance have a new endpoint?

40

- 
- No, the existing database is paused and the new database is brought up in place with the same endpoint
  - Yes, the new database has a new endpoint (x)

From the RDS FAQ (<http://aws.amazon.com/rds/faqs/#129>):

**"Q: What is the difference between automated backups and DB Snapshots?"**

...

Please note: When you perform a restore operation to a point in time or from a DB Snapshot, a new DB Instance is created with a new endpoint (the old DB Instance can be deleted with the AWS Management Console or DeleteDBInstance API, if so desired). This is done to enable you to create multiple DB Instances from a specific DB Snapshot or point in time."

---

Is database performance affected during when backing up an RDS Multi-AZ deployment?

41

- Yes for both the master and the standby (x)
- Yes, but only for the standby
- Yes, but only for the master
- No

From the RDS documentation

(<http://docs.aws.amazon.com/AmazonRDS/latest/UserGuide/Overview.BackingUpAndRestoringAmazonRDSInstances.html>):

"During the backup window, storage I/O may be suspended while your data is being backed up and **you may experience elevated latency**. This I/O suspension typically lasts for the duration of the snapshot. This period of I/O suspension is shorter for Multi-AZ DB deployments, since the backup is taken from the standby, but latency can occur during the backup process."

also from the RDS FAQ ():

**"Q: How do DB Snapshots and automated backups work with my Multi-AZ deployment?"**

You interact with automated backup and DB Snapshot functionality in the same way whether you are running a standard deployment in a Single-AZ or Multi-AZ deployment. If you are running a Multi-AZ deployment, automated backups and DB Snapshots are simply taken from the standby to avoid I/O suspension on the primary. **Please note that you may experience increased I/O latency (typically lasting a few minutes) during backups for both Single-AZ and Multi-AZ deployments.**

Initiating a restore operation (point-in-time restore or restore from DB Snapshot) also works the same with Multi-AZ deployments as standard, Single-AZ deployments. New DB Instance deployments can be created with either the RestoreDBInstanceFromSnapshot or RestoreDBInstanceToPointInTime APIs. These new DB Instance deployments can be either standard or Multi-AZ, regardless of whether the source backup was initiated on a standard or Multi-AZ deployment."

---

In an RDS multi-AZ deployments, are backups and snapshots taken from the master or the standby?

42

- Master
- Standby (x)

From the RDS documentation

(<http://docs.aws.amazon.com/AmazonRDS/latest/UserGuide/Overview.BackingUpAndRestoringAmazonRDSInstances.html>):

"During the backup window, storage I/O may be suspended while your data is being backed up and you may experience elevated latency. This I/O suspension typically lasts for the duration of the snapshot. **This period of I/O suspension is shorter for Multi-AZ DB deployments, since the backup is taken from the standby**, but latency can occur during the backup process."

---

Are snapshots retained after an RDS instance is deleted?

43



- 
- Yes
  - Yes, but only manually created snapshots (x)
  - Yes, but only automated snapshots
  - No

From the RDS FAQ (<http://aws.amazon.com/rds/faqs/#129>):

**"Q: What happens to my backups and DB Snapshots if I delete my DB Instance?**

...

**Automated backups are deleted when the DB Instance is deleted. Only manually created DB Snapshots are retained after the DB Instance is deleted."**

---

When RDS performs a maintenance event (upgrade, etc) on a multi-AZ database instance, does the primary node change?

44

- Yes (x)
- No

From the RDS documentation

([http://docs.aws.amazon.com/AmazonRDS/latest/UserGuide/USER\\_UpgradeDBInstance.Maintenance.html](http://docs.aws.amazon.com/AmazonRDS/latest/UserGuide/USER_UpgradeDBInstance.Maintenance.html)):

"Running your DB instance as a Multi-AZ deployment can further reduce the impact of a maintenance event, because Amazon RDS will conduct maintenance by following these steps:

1. Perform maintenance on the standby.
2. Promote the standby to primary.
3. Perform maintenance on the old primary, which becomes the new standby.

Note

When you modify the database engine for your DB instance in a Multi-AZ deployment, then Amazon RDS upgrades both the primary and secondary DB instances at the same time. In this case, the database engine for the entire Multi-AZ deployment is shut down during the upgrade.

For more information on Multi-AZ deployments, see High Availability (Multi-AZ)."

---

How do you manually initiate a failover for an RDS instance?

45

- It's an option while rebooting (x)
- Failover is handled automatically and can't be manually initiated
- Invoke the InstanceFailover API call

From the RDS FAQ (<https://aws.amazon.com/rds/faqs/>):

**"Q: Can I initiate a "forced failover" for my Multi-AZ DB Instance deployment?**

Amazon RDS will automatically failover without user intervention under a variety of failure conditions. In addition, **Amazon RDS provides an option to initiate a failover when rebooting your instance**. You can access this feature via the AWS Management Console or when using the RebootDBInstance API call."

---

Is it possible to copy an RDS snapshot across regions?

46

- 
- Yes (x)
  - No

From the RDS User Guide ([http://docs.aws.amazon.com/AmazonRDS/latest/UserGuide/USER\\_CopySnapshot.html](http://docs.aws.amazon.com/AmazonRDS/latest/UserGuide/USER_CopySnapshot.html)):

"Amazon RDS supports two types of DB snapshot copies. First, you can copy an automated DB snapshot to create a manual DB snapshot in the same AWS region. By creating a manual DB snapshot, the DB snapshot is retained; automated DB snapshots are deleted after their retention period expires.

Second, **you can copy either an automated or manual DB snapshot from one region to another region.** By copying the DB snapshot to another region, you create a manual DB snapshot that is retained in that region. You perform the DB snapshot copy in the target region, and use an Amazon RDS ARN to specify the location of the DB snapshot in the source region. "

---

Is it possible to manually copy an RDS automated snapshot and why would you do it?

47

- Yes, to ensure that the snapshot is retained when the instance is deleted (x)
- Yes, to reduce restore time
- Yes, to increase the availability of the snapshot in the event of an AZ failure
- No

From the RDS User Guide ([http://docs.aws.amazon.com/AmazonRDS/latest/UserGuide/USER\\_CopySnapshot.html](http://docs.aws.amazon.com/AmazonRDS/latest/UserGuide/USER_CopySnapshot.html)):

"Amazon RDS supports two types of DB snapshot copies. First, **you can copy an automated DB snapshot to create a manual DB snapshot in the same AWS region. By creating a manual DB snapshot, the DB snapshot is retained;** automated DB snapshots are deleted after their retention period expires.

Second, you can copy either an automated or manual DB snapshot from one region to another region. By copying the DB snapshot to another region, you create a manual DB snapshot that is retained in that region. You perform the DB snapshot copy in the target region, and use an Amazon RDS ARN to specify the location of the DB snapshot in the source region. "

---

What are the allowable RDS database engines?

48

- Aurora (x)
- MySQL (x)
- MariaDB (x)
- PostgreSQL (x)
- Oracle (x)
- SQLServer (x)
- HSQLDB
- Firebird

Just look at the AWS console...

---

Is it possible to encrypt an RDS instance?

49

- Yes (x)
- No

From the RDS User Guide (<http://docs.aws.amazon.com/AmazonRDS/latest/UserGuide/Overview.Encryption.html>):

"You can encrypt your Amazon RDS instances and snapshots at rest by enabling the encryption option for your Amazon RDS DB instance. Data that is encrypted at rest includes the underlying storage for a DB instance, its automated backups, Read Replicas, and snapshots.

Amazon RDS encrypted instances use the industry standard AES-256 encryption algorithm to encrypt your data on the server that hosts your Amazon RDS instance. Once your data is encrypted, Amazon RDS handles authentication of access and decryption of your data transparently with a minimal impact on performance. You don't need to modify your database client applications to use encryption."

- 1-2 minutes (x)
- 3-5 minutes
- 5-10 minutes
- up to 20 minutes

From the RDS FAQ (<https://aws.amazon.com/rds/faqs/>):

**"Q: What happens during Multi-AZ failover and how long does it take?"**

Failover is automatically handled by Amazon RDS so that you can resume database operations as quickly as possible without administrative intervention. When failing over, Amazon RDS simply flips the canonical name record (CNAME) for your DB Instance to point at the standby, which is in turn promoted to become the new primary. We encourage you to follow best practices and implement database connection retry at the application layer.

**Failovers, as defined by the interval between the detection of the failure on the primary and the resumption of transactions on the standby, typically complete within one to two minutes.** Failover time can also be affected by whether large uncommitted transactions must be recovered; the use of adequately large instance types is recommended with Multi-AZ for best results. AWS also recommends the use of Provisioned IOPS with Multi-AZ instances, for fast, predictable, and consistent throughput performance."

- Yes for all instance types
- Yes, but only for MySQL and MariaDB (x)
- No

From the RDS User Guide

([http://docs.aws.amazon.com/AmazonRDS/latest/UserGuide/USER\\_ReadRepl.html#USER\\_ReadRepl.XRgn](http://docs.aws.amazon.com/AmazonRDS/latest/UserGuide/USER_ReadRepl.html#USER_ReadRepl.XRgn)):

"You can create a MySQL or MariaDB Read Replica in a different region than the source DB instance to improve your disaster recovery capabilities, scale read operations into a region closer to end users, or make it easier to migrate from a data center in one region to a data center in another region."

- 1 day (x)
- 7 days
- 14 days
- 30 days

From the RDS User Guide ([http://docs.aws.amazon.com/AmazonRDS/latest/UserGuide/USER\\_WorkingWithAutomatedBackups.html](http://docs.aws.amazon.com/AmazonRDS/latest/UserGuide/USER_WorkingWithAutomatedBackups.html)):

"Amazon RDS can automatically back up all of your DB instances. You can set the backup retention period when you create a DB instance. **If you don't set the backup retention period, Amazon RDS uses a default period retention period of one day.** "

- 10-20 seconds
- 30-60 seconds
- 60-120 seconds (x)
- 120-240 seconds

From the RDS User Guide (<http://docs.aws.amazon.com/AmazonRDS/latest/UserGuide/Concepts.MultiAZ.html>):

"In the event of a planned or unplanned outage of your DB instance, Amazon RDS automatically switches to a standby replica in another Availability Zone if you have enabled Multi-AZ. **The time it takes for the failover to complete depends on the database activity and other conditions at the time the primary DB instance became unavailable. Failover times are typically 60-120 seconds.** However, large transactions or a lengthy recovery process can increase failover time. When the failover is complete, it can take additional time for the RDS console UI to reflect the new Availability Zone."

What conditions trigger an automated failover of a multi-AZ RDS instance?

54

- An Availability Zone outage (x)
- Failure of the primary DB instance (x)
- Change of the DB instance's server type (x)
- Patching the DB instance's operating system (x)
- Manual failover initiated using "Reboot with Failover" (x)
- Client connection timeouts

From the RDS User Guide (<http://docs.aws.amazon.com/AmazonRDS/latest/UserGuide/Concepts.MultiAZ.html>):

Amazon RDS handles failovers automatically so you can resume database operations as quickly as possible without administrative intervention. The primary DB instance switches over automatically to the standby replica if any of the following conditions occur:

- An Availability Zone outage
- The primary DB instance fails
- The DB instance's [storage] type is changed
- The operating system of the DB instance is undergoing software patching
- A manual failover of the DB instance was initiated using Reboot with failover

Is it possible to reduce the storage of an RDS instance?

55

- Yes
- No (x)

From the RDS User Guide ([http://docs.aws.amazon.com/AmazonRDS/latest/UserGuide/USER\\_PIOPS.StorageTypes.html](http://docs.aws.amazon.com/AmazonRDS/latest/UserGuide/USER_PIOPS.StorageTypes.html)):

"Data storage in Amazon RDS is specified by selecting a storage type and providing a storage size (GB) when you create or modify a DB instance. You can change the type of storage your instance uses by modifying the DB instance, but changing the type of storage in some cases might result in a short outage for the instance. Changing from Magnetic to either General Purpose (SSD) or Provisioned IOPS (SSD) results in an outage. Also, changing from General Purpose (SSD) or Provisioned IOPS (SSD) to Magnetic results in an outage. The outage time is typically 60–120 seconds. For more information about Amazon RDS storage types, see Amazon RDS Storage Types.

Increasing the allocated storage does not result in an outage. **Note that you cannot reduce the amount of storage once it has been allocated. The only way to reduce the amount of storage allocated to a DB instance is to dump the data out of the DB instance, create a new DB instance with less storage space, and then load the data into the new DB instance.**"

Does modifying an RDS instance's instance class cause an outage?

56

- 
- Yes (x)
  - No

From the RDS User Guide (<http://docs.aws.amazon.com/AmazonRDS/latest/UserGuide/Overview.DBInstance.Modifying.html>):

If "Apply Immediately" is specified: "Change is applied immediately and an immediate outage will occur."

If "Apply Immediately" isn't specified: "Change is applied during the next maintenance window. Changing this setting causes an outage to occur."

---

Can you use the standby in an RDS Multi-AZ deployment to serve read traffic?

57

- 
- Yes
  - No (x)

From the RDS User Guide (<http://docs.aws.amazon.com/AmazonRDS/latest/UserGuide/Concepts.MultiAZ.html>):

"The high-availability feature is not a scaling solution for read-only scenarios; **you cannot use a standby replica to serve read traffic.** To service read-only traffic, you should use a Read Replica."

## Route 53

---

Does Route 53 support wildcard entries?

58

- 
- Yes (x)
  - No

From the Route 53 FAQ (<https://aws.amazon.com/route53/faqs/>):

### "Q. Does Amazon Route 53 support wildcard entries? If so, what record types support them?"

Yes. To make it even easier for you to configure DNS settings for your domain, Amazon Route 53 supports wildcard entries for all record types. A wildcard entry is a record in a DNS zone that will match requests for any domain name based on the configuration you set. For example, a wildcard DNS record such as *\*.example.com* will match queries for *www.example.com* and *subdomain.example.com*."

---

Can Route 53 point a zone apex at other AWS services (ELBs, S3 buckets, and CloudFront distributions)?

59

- 
- Yes (x)
  - No

From the Route 53 FAQ (<https://aws.amazon.com/route53/faqs/>):

**"Q. Can I point my zone apex (example.com versus www.example.com) at my Elastic Load Balancer?"**

Yes. Amazon Route 53 offers a special type of record called an 'Alias' record that lets you map your zone apex (example.com) DNS name to your ELB DNS name (i.e. elb1234.elb.amazonaws.com). IP addresses associated with Amazon Elastic Load Balancers can change at any time due to scaling up, scaling down, or software updates. Route 53 responds to each request for an Alias record with one or more IP addresses for the load balancer. Queries to Alias records that are mapped to ELB load balancers are free. These queries are listed as "Intra-AWS-DNS-Queries" on the Amazon Route 53 usage report."

and

**"Q. Can I point my zone apex (example.com versus www.example.com) at my website hosted on Amazon S3?"**

Yes. Amazon Route 53 offers a special type of record called an 'Alias' record that lets you map your zone apex (example.com) DNS name to your Amazon S3 website bucket (i.e. example.com.s3-website-us-west-2.amazonaws.com). IP addresses associated with Amazon S3 website endpoints can change at any time due to scaling up, scaling down, or software updates. Route 53 responds to each request for an Alias record with one IP address for the bucket. Route 53 doesn't charge for queries to Alias records that are mapped to an S3 bucket that is configured as a website. These queries are listed as "Intra-AWS-DNS-Queries" on the Amazon Route 53 usage report."

and

**Q. Can I point my zone apex (example.com versus www.example.com) at my Amazon CloudFront distribution?"**

Yes. Amazon Route 53 offers a special type of record called an 'Alias' record that lets you map your zone apex (example.com) DNS name to your Amazon CloudFront distribution (for example, d123.cloudfront.net). IP addresses associated with Amazon CloudFront endpoints vary based on your end user's location (in order to direct the end user to the nearest CloudFront edge location) and can change at any time due to scaling up, scaling down, or software updates. Route 53 responds to each request for an Alias record with the IP address(es) for the distribution. Route 53 doesn't charge for queries to Alias records that are mapped to a CloudFront distribution. These queries are listed as "Intra-AWS-DNS-Queries" on the Amazon Route 53 usage report.

---

How quickly do DNS changes propagate globally?

60

- Within 60 seconds (x)
- Within 120 seconds
- Within 300 seconds

From the Route 53 FAQ (<https://aws.amazon.com/route53/faqs/>):

**"Q. How quickly will changes I make to my DNS settings on Amazon Route 53 propagate globally?"**

Amazon Route 53 is designed to propagate updates you make to your DNS records to its world-wide network of authoritative DNS servers **within 60 seconds** under normal conditions. A change is successfully propagated world-wide when the API call returns an *INSYNC* status listing.

Note that caching DNS resolvers are outside the control of the Amazon Route 53 service and will cache your resource record sets according to their time to live (TTL). The *INSYNC* or *PENDING* status of a change refers only to the state of Route 53's authoritative DNS servers."

---

What is a Route 53 hosted zone?

61

- 
- A collection of resource record sets that share a common domain name suffix (x)
  - A highly available cross-AZ DNS instance
  - A logical segregation of record sets with a DNS domain

From the Route 53 FAQ (<https://aws.amazon.com/route53/faqs/>):

**"Q. What is the difference between a Domain and a Hosted Zone?"**

A domain is a general DNS concept. Domain names are easily recognizable names for numerically addressed Internet resources. For example, *amazon.com* is a domain. A hosted zone is an Amazon Route 53 concept. A hosted zone is analogous to a traditional DNS zone file; it represents a collection of records that can be managed together, belonging to a single parent domain name. All resource record sets within a hosted zone must have the hosted zone's domain name as a suffix. For example, the *amazon.com* hosted zone may contain records named *www.amazon.com*, and *www.aws.amazon.com*, but not a record named *www.amazon.ca*. You can use the Route 53 Management Console or API to create, inspect, modify, and delete hosted zones. You can also use the Management Console or API to register new domain names and transfer in existing domain names into Route 53's management."

---

Using Route 53, is it possible to associate multiple IP addresses to a single record?

62

- Yes (x)
- No

From the Route 53 FAQ (<https://aws.amazon.com/route53/faqs/>):

**"Q. Can I associate multiple IP addresses with a single record?"**

Yes. Associating multiple IP addresses with a single record is often used for balancing the load of geographically-distributed web servers. Amazon Route 53 allows you to list multiple IP addresses for an A record and responds to DNS requests with the list of all configured IP addresses."

---

Can Route 53 provide DNS resolution of records within a VPC without exposing those records to the internet?

63

- Yes (x)
- No

From the Route 53 FAQ (<https://aws.amazon.com/route53/faqs/>):

**"Q. What is Private DNS?"**

Private DNS is a Route 53 feature that lets you have authoritative DNS within your VPCs without exposing your DNS records (including the name of the resource and its IP address(es) to the Internet."

---

What is Route 53 Latency Based Routing (LBR)?

64

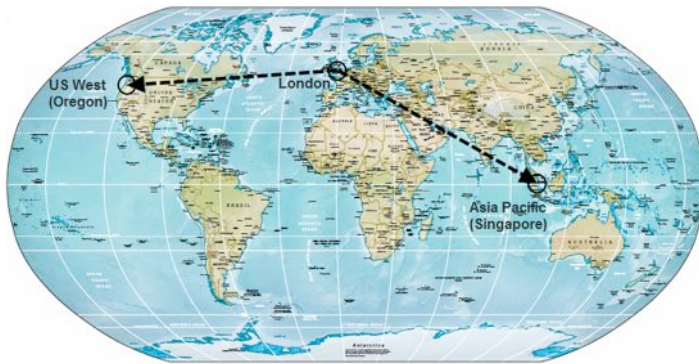
- Allows you to define a group of resource record sets, each in a different country or continent; Route 53 will choose the record set based on the location of the requesting client's IP (x)
- Scales Route 53 to ensure consistent response times when the clients experience increased latency using the Route 53 APIs
- Intelligently updates record sets based on performance of AWS services referenced as aliases

From the Route 53 Developer Guide (<http://docs.aws.amazon.com/Route53/latest/DeveloperGuide/routing-policy.html#routing-policy-latency>):

If your application is hosted on Amazon EC2 instances in multiple Amazon EC2 regions, you can reduce latency for your users by serving their requests from the Amazon EC2 region for which network latency is lowest. Amazon Route 53 latency-based routing lets you use DNS to route user requests to the Amazon EC2 region that will give your users the fastest response.

To use latency-based routing, you create a latency resource record set for the Amazon EC2 resource in each region that hosts your application. When Amazon Route 53 receives a query for the corresponding domain, it selects the latency resource record set for the Amazon EC2 region that gives the user the lowest latency. Amazon Route 53 then responds with the value associated with that resource record set.

For example, suppose you have ELB load balancers in the US West (Oregon) region and in the Asia Pacific (Singapore) region, and that you've created a latency resource record set in Amazon Route 53 for each load balancer. A user in London enters the name of your domain in a browser, and DNS routes the request to an Amazon Route 53 name server. Amazon Route 53 refers to its data on latency between London and the Singapore region and between London and the Oregon region. If latency is lower between London and the Oregon region, Amazon Route 53 responds to the user's request with the IP address of your load balancer in the Amazon EC2 data center in Oregon. If latency is lower between London and the Singapore region, Amazon Route 53 responds with the IP address of your load balancer in the Amazon EC2 data center in Singapore.



Map courtesy of the University of Texas Libraries, The University of Texas at Austin.

What is Route 53 Geolocation Routing?



- 
- Allows you to define resource record sets in multiple AWS regions; Route 53 will send queries to the region that is physically closest to the client (x)
  - Ensures that Route 53 is equally available in all countries
  - Updates interconnects between Route 53 installations to give the best answer to client queries regardless of physical location

From the Route 53 Developer Guide (<http://docs.aws.amazon.com/Route53/latest/DeveloperGuide/routing-policy.html#routing-policy-latency>):

“Geolocation routing lets you choose the resources that serve your traffic based on the geographic location of your users, meaning the location from which DNS queries originate. For example, you might want all queries from Africa to be routed to a web server with an IP address of 192.0.2.111.

When you use geolocation routing, you can localize your content and present some or all of your website in the language of your users. You can also use geolocation routing to restrict distribution of content to only the locations in which you have distribution rights. Another possible use is for balancing load across endpoints in a predictable, easy-to-manage way, so that each user location is consistently routed to the same endpoint.

You can specify geographic locations by continent, by country, or by state in the United States. If you create separate resource record sets for overlapping geographic regions—for example, one resource record set for a continent and one for a country on the same continent—priority goes to the smallest geographic region. This allows you to route some queries for a continent to one resource and to route queries for selected countries on that continent to a different resource. (For a list of the countries on each continent, see [Location](#).)

Geolocation works by mapping IP addresses to locations. However, some IP addresses aren't mapped to geographic locations, so even if you create geolocation resource record sets that cover all seven continents, Amazon Route 53 will receive some DNS queries from locations that it can't identify. You can create a default resource record set that handles both queries from IP addresses that aren't mapped to any location and queries that come from locations for which you haven't created geolocation resource record sets. If you don't create a default resource record set, Amazon Route 53 returns a "no answer" response for queries from those locations."

- A feature whereby Route 53 will reroute traffic away from resource record sets (RRSs) in a group of RRSs of the same type (e.g., latency RRSs or weighted RRSs) that fail health checks (x)
- Lets you control how Route 53 re-routes your traffic in the case of an AZ failure
- Lets you configure high-availability for a hosted zone

From the Route 53 Developer Guide (<http://docs.aws.amazon.com/Route53/latest/DeveloperGuide/dns-failover-simple-configs.html>):

The simplest configuration for which checking the health of your resources is useful is when you have two or more resources that are performing the same function. For example, you might have multiple Amazon EC2 servers running HTTP server software responding to requests for the example.com website. In Amazon Route 53, you create a group of resource record sets that have the same name and type, such as weighted resource record sets or latency resource record sets of type A. You create one resource record set for each resource, and you configure Amazon Route 53 to check the health of the corresponding resource. In this configuration, Amazon Route 53 chooses which resource record set will respond to a DNS query for example.com and bases the choice in part on the health of your resources.

As long as all of the resources are healthy, Amazon Route 53 responds to queries using all of your example.com weighted resource record sets. When a resource becomes unhealthy, Amazon Route 53 responds to queries using only the healthy resource record sets for example.com.



How does it take for Route 53 to execute a DNS failover?

- Under one minute
- Under two minutes (x)
- Under five minutes
- Under ten minutes

From this re:Invent presentation (<https://www.youtube.com/watch?v=f9y-T7mQVxs>):

The top bar represents the time to respond a failover “manually” by personally reacting to a CloudWatch alarm and reconfiguring Route 53 and other components in the best case.

The second bar represents how long it takes for Route 53 itself to execute a DNS failover using the native feature.



Using Route 53, is it possible to create a CNAME record at a zone apex?

68

- Yes
- No (x)

From the Route 53 Developer Guide ():

**"You cannot create a CNAME record at the top node of a DNS namespace, also known as the *zone apex*.** For example, if you register the DNS name example.com, the zone apex is example.com."

Is it possible to use Route 53 to implement split-horizon DNS?

69

- Yes (x)
- No

From the Route 53 Developer Guide (<http://docs.aws.amazon.com/Route53/latest/DeveloperGuide/hosted-zones-private.html>):

**"You can use Amazon Route 53 to configure split-view DNS, also known as split-horizon DNS.** If you want to maintain internal and external versions of the same website or application (for example, for testing changes before you make them public), you can configure public and private hosted zones to return different internal and external IP addresses for the same domain name. Just create a public hosted zone and a private hosted zone that have the same domain name, and create the same subdomains in both hosted zones."

## CloudFront

Does CloudFront accept self-signed certificates in distributions?

70

- 
- Yes
  - No (x)

From the CloudFront documentation (<http://docs.aws.amazon.com/AmazonCloudFront/latest/DeveloperGuide/SecureConnections.html>):

"If you configure CloudFront to use HTTPS when communicating with your origin...CloudFront verifies that your certificate was issued by an established third-party certificate authority...You cannot use a self-signed certificate."

---

Can you upload content through CloudFront?

71

- 
- Yes (x)
  - No

From this press release: <https://aws.amazon.com/about-aws/whats-new/2013/10/15/amazon-cloudfront-now-supports-put-post-and-other-http-methods>

"We are excited to announce that Amazon CloudFront, AWS's easy-to-use and cost-effective content delivery service, has expanded its capabilities to now allow you to upload content to your web servers. This release adds support for five additional HTTP methods, including POST, PUT, DELETE, OPTIONS and PATCH. This means you can improve the performance of dynamic websites that have web forms, comment and login boxes, "add to cart" buttons or other features that upload data from end users. It also means you can now use a single domain name to deliver your whole website through CloudFront thereby accelerating both the download and upload parts of your website.

You can use existing Amazon CloudFront distributions for upload requests by simply enabling support in the AWS management console. When end users upload content, CloudFront will send the upload request back to the origin web server (such as an Amazon S3 bucket, an Amazon EC2 instance, an Elastic Load Balancer, or your own origin server) over an optimized route that uses persistent connections, TCP/IP and network path optimizations."

---

Does using an Origin Access Identity to force users to access S3 content only via CloudFront URLs prevent IAM users from accessing that content via the S3 API?

72

- 
- Yes
  - No (x)

From the CloudFront documentation (<http://docs.aws.amazon.com/AmazonCloudFront/latest/DeveloperGuide/private-content-restricting-access-to-s3.html>):

"[when] CloudFront updates bucket permissions to grant the specified origin access identity the permission to read objects in your bucket [...it...] does not remove existing permissions. If users currently have permission to access the objects in your bucket using Amazon S3 URLs, they will still have that permission after CloudFront updates your bucket permissions. "

---

Is it possible to reserve CloudFront capacity at a discount?

73

- 
- Yes (x)
  - No

From the CloudFront pricing page (<http://aws.amazon.com/cloudfront/pricing/>):

"Reserved Capacity gives you the option to commit to a minimum monthly usage level for 12 months or longer and in turn receive a significant discount. Reserved Capacity agreements begin at a minimum of 10 TB of data transfer per month from a single region. Customers who commit to higher usage receive additional discounts."

---

Can CloudFront restrict access from specific geographic locations?

74

- 
- Yes (x)
  - No

From the CloudFront Developer Guide (<http://docs.aws.amazon.com/AmazonCloudFront/latest/DeveloperGuide/georestrictions.html>):

**"You can use *geo restriction*, also known as *geoblocking*, to prevent users in specific geographic locations from accessing content that you're distributing through a CloudFront web distribution.** To use geo restriction, you have two options:

- Use the CloudFront geo restriction feature. Use this option to restrict access to all of the files that are associated with a distribution and to restrict access at the country level.
- Use a third-party geolocation service. Use this option to restrict access to a subset of the files that are associated with a distribution or to restrict access at a finer granularity than the country level."

---

Can CloudFront serve dynamic content?

75

- 
- Yes (x)
  - No

This blog (<https://aws.amazon.com/blogs/aws/amazon-cloudfront-support-for-dynamic-content/>) goes into the details:

"As you know, content on the web is identified by a URL, or Uniform Resource Locator such as [https://media.amazonwebservices.com/blog/console\\_cw\\_est\\_charge\\_service\\_2.png](https://media.amazonwebservices.com/blog/console_cw_est_charge_service_2.png). A URL like this always identifies a unique piece of content.

A URL can also contain a query string. This takes the form of a question mark ("?",) and additional information that the server can use to personalize the request. Suppose that we had a server at [www.example.com](http://www.example.com), and that can return information about a particular user by invoking a PHP script that accepts a user name as an argument, with URLs like <http://www.example.com/userinfo.php?jeff> or <http://www.example.com/userinfo.php?tina>.

Up until now, CloudFront did not use the query string as part of the key that it uses to identify the data that it stores in its edge locations.

**We're changing that today, and you can now use CloudFront to speed access to your dynamic data at our current low rates, making your applications faster and more responsive, regardless of where your users are located.**

With this change (and the others that I'll tell you about in a minute), Amazon CloudFront will become an even better component of your global applications. We've put together a long list of optimizations that will each increase the performance of your application on their own, but will work even better when you use them in conjunction with other AWS services such as Route 53, Amazon S3, and Amazon EC2."

---

Which of the following considerations would suggest that you use CloudFront signed URLs?

76

- 
- You want to use an RTMP distribution; signed cookies aren't supported for RTMP distributions (x)
  - You want to restrict access to individual files, for example, an installation download for your application (x)
  - Your users are using a client (for example, a custom HTTP client) that doesn't support cookies (x)
  - You want to provide access to multiple restricted files, for example, all of the files for a video in HLS format or all of the files in the subscribers' area of a website
  - You don't want to change your current URLs

From the CloudFront Developer Guide (<http://docs.aws.amazon.com/AmazonCloudFront/latest/DeveloperGuide/private-content-choosing-signed-urls-cookies.html>):

"CloudFront signed URLs and signed cookies provide the same basic functionality: they allow you to control who can access your content. If you want to serve private content through CloudFront and you're trying to decide whether to use signed URLs or signed cookies, consider the following.

**Use signed URLs in the following cases:**

- **You want to use an RTMP distribution. Signed cookies aren't supported for RTMP distributions.**
- **You want to restrict access to individual files, for example, an installation download for your application.**
- **Your users are using a client (for example, a custom HTTP client) that doesn't support cookies.**

Use signed cookies in the following cases:

- You want to provide access to multiple restricted files, for example, all of the files for a video in HLS format or all of the files in the subscribers' area of a website.
- You don't want to change your current URLs."

- You want to restrict access to individual files, for example, an installation download for your application
- Your origin server requires cookies for application logic
- You want to provide access to multiple restricted files, for example, all of the files for a video in HLS format or all of the files in the subscribers' area of a website (x)
- You don't want to change your current URLs (x)

From the CloudFront Developer Guide (<http://docs.aws.amazon.com/AmazonCloudFront/latest/DeveloperGuide/private-content-choosing-signed-urls-cookies.html>):

"CloudFront signed URLs and signed cookies provide the same basic functionality: they allow you to control who can access your content. If you want to serve private content through CloudFront and you're trying to decide whether to use signed URLs or signed cookies, consider the following.

Use signed URLs in the following cases:

- You want to use an RTMP distribution. Signed cookies aren't supported for RTMP distributions.
- You want to restrict access to individual files, for example, an installation download for your application.
- Your users are using a client (for example, a custom HTTP client) that doesn't support cookies.

**Use signed cookies in the following cases:**

- **You want to provide access to multiple restricted files, for example, all of the files for a video in HLS format or all of the files in the subscribers' area of a website.**
- **You don't want to change your current URLs."**

What are the options if you want CloudFront to deliver HTTPS content using your own domain name?

- Not possible
- Server Name Indication (SNI) Custom SSL (x)
- Dedicated IP Custom SSL (x)
- Route 53 CloudFront Configuration

From this announcement ():

«1778»

Custom SSL Options for Amazon CloudFront

**Custom SSL certificate support lets you deliver content over HTTPS using your own domain name** and your own SSL certificate. This gives visitors to your website the security benefits of CloudFront over an SSL connection that uses your own domain name in addition to lower latency and higher reliability.

#### **SNI Custom SSL**

Server Name Indication (SNI) Custom SSL relies on the SNI extension of the Transport Layer Security protocol, which allows multiple domains to serve SSL traffic over the same IP address. Amazon CloudFront delivers your content from each edge location and offers the same security as the Dedicated IP Custom SSL feature (see below).

#### **Dedicated IP Custom SSL**

**If you need to deliver content to browsers that don't support SNI, you can use the Dedicated IP Custom SSL feature.** For this feature CloudFront allocates dedicated IP addresses to serve your SSL content at each CloudFront edge location.

«1777»

Here's some detail on SNI from Wikipedia ([https://en.wikipedia.org/wiki/Server\\_Name\\_Indication](https://en.wikipedia.org/wiki/Server_Name_Indication)):

«1777»

When making a TLS connection the client requests a digital certificate from the web server; once the server sends the certificate, the client examines it and compares the name it was trying to connect to with the name(s) included in the certificate. If a match occurs the connection proceeds as normal. If a match is not found the user may be warned of the discrepancy and the connection may abort as the mismatch may indicate an attempted man-in-the-middle attack. However, some applications allow the user to bypass the warning to proceed with the connection, with the user taking on the responsibility of trusting the certificate and, by extension, the connection.

It is possible for one certificate to cover multiple hostnames. The X.509 v3 specification introduced the *subjectAltName* field which allows one certificate to specify more than one domain and the usage of wildcards in both the common name and *subjectAltName* fields. However it may be impractical—or even impossible, due to lack of a full list of all names in advance—to obtain a single certificate that covers all names a server will be responsible for. A server that is responsible for multiple hostnames is likely to need to present a different certificate for each name (or small group of names). Since 2005, CAcert has run experiments on different methods of using TLS on virtual servers.<sup>[3]</sup> Most of the experiments are unsatisfactory and impractical. For example, it is possible to use *subjectAltName* to contain multiple domains controlled by one person<sup>[4]</sup> in a single certificate. Such "unified communications certificates" must be reissued every time the list of domains changes.

Name-based virtual hosting allows multiple DNS hostnames to be hosted by a single server (usually a web server) on the same IP address. To achieve this the server uses a hostname presented by the client as part of the protocol (for HTTP the name is presented in the host header). However, when using HTTPS the TLS handshake happens before the server sees any HTTP headers. Therefore, it is not possible for the server to use the information in the HTTP host header to decide which certificate to present and as such only names covered by the same certificate can be served from the same IP address.

In practice, this means that an HTTPS server can only serve one domain (or small group of domains) per IP address for secured browsing. Assigning a separate IP address for each site increases the cost of hosting, since requests for IP addresses must be justified to the regional internet registry and IPv4 addresses are now in short supply. The result is that many websites are effectively constrained from using secure communications over IPv4. IPv6 address space is not in short supply so websites served using IPv6 are unaffected by this issue.

**SNI addresses this issue by having the client send the name of the virtual domain as part of the TLS negotiation.<sup>[2]</sup> This enables the server to select the correct virtual domain early and present the browser with the certificate containing the correct name. Therefore, with clients and servers that implement SNI, a server with a single IP address can serve a group of domain names for which it is impractical to get a common certificate.**

«1777»

What is a CloudFront Cache Behavior?

- 
- A set of configuration options for a particular URL path in a distribution (x)
  - A set of distribution-wide configuration properties that can be dynamically selected at runtime based on time-of-day, geography, and other contextual information from the request
  - The configuration for a HTTP verb / URI / media type triplet

From the CloudFront Developer Guide (<http://docs.aws.amazon.com/AmazonCloudFront/latest/DeveloperGuide/distribution-web-values-specify.html#DownloadDistValuesCacheBehavior>):

"A cache behavior lets you configure a variety of CloudFront functionality for a given URL path pattern for files on your website. For example, one cache behavior might apply to all .jpg files in the images directory on a web server that you're using as an origin server for CloudFront."

---

What can be configured in a CloudFront Cache Behavior?

80

- 
- The origin to forward requests to (for multi-origin distributions) (x)
  - Whether or not to forward query strings to the origin (x)
  - Whether or not to require a signed URL (x)
  - Whether or not to require HTTPS (x)
  - The minimum cache retention irrespective of any value in the Cache-Control headers from the origin (x)
  - Compression algorithm
  - Logging options

From the CloudFront Developer Guide (<http://docs.aws.amazon.com/AmazonCloudFront/latest/DeveloperGuide/distribution-web-values-specify.html#DownloadDistValuesCacheBehavior>):

"""

The functionality you can configure for each cache behavior includes:

- The path pattern.
- If you have configured multiple origins for your CloudFront distribution, which origin you want CloudFront to forward your requests to.
- Whether to forward query strings to your origin.
- Whether accessing the specified files requires signed URLs.
- Whether to require end users to use HTTPS to access those files.
- The minimum amount of time that those files stay in the CloudFront cache regardless of the value of any Cache-Control headers that your origin adds to the files.

"""

---

What can be configured in a CloudFront Distribution?

81



- 
- Edge locations to serve data (x)
  - Alternate Domain Names (CNAMEs) (x)
  - SSL Certificate (x)
  - Clients Supported (x)
  - Minimum SSL Protocol Version (x)
  - Default Root Object (x)
  - Logging (x)
  - Whether or not to forward query strings to the origin
  - Whether or not to require a signed URL
  - Whether or not to require HTTPS

From the CloudFront Developer Guide (<http://docs.aws.amazon.com/AmazonCloudFront/latest/DeveloperGuide/distribution-web-values-specify.html#DownloadDistValuesCacheBehavior>):

❖❖❖❖

### **Distribution Details**

- Price Class
- AWS WAF Web ACL
- Alternate Domain Names (CNAMEs)
- SSL Certificate
- Clients Supported
- Minimum SSL Protocol Version
- Default Root Object
- Logging
- Bucket for Logs
- Log Prefix
- Cookie Logging
- Comment
- Distribution State

❖❖❖❖

What can be configured in a CloudFront Origin?

- Origin Domain Name and Path (x)
- Origin Access Identity (OAI) (x)
- S3 bucket access restrictions (x)
- HTTP/HTTPS ports (x)
- Minimum SSL Protocol Version
- Default Root Object
- Logging
- Whether or not to forward query strings to the origin
- Whether or not to require a signed URL
- Whether or not to require HTTPS

From the CloudFront Developer Guide (<http://docs.aws.amazon.com/AmazonCloudFront/latest/DeveloperGuide/distribution-web-values-specify.html#DownloadDistValuesCacheBehavior>):

44/7777

### Origin Settings

- Origin Domain Name
- Origin Path
- Origin ID
- Restrict Bucket Access (Amazon S3 Only)
- Origin Access Identity (Amazon S3 Only)
- Comment for New Identity (Amazon S3 Only)
- Your Identities (Amazon S3 Only)
- Grant Read Permissions on Bucket (Amazon S3 Only)
- HTTP Port (Amazon EC2 and Other Custom Origins Only)
- HTTPS Port (Amazon EC2 and Other Custom Origins Only)
- Origin Protocol Policy (Amazon EC2, Amazon S3 Buckets Configured as Website Endpoints, and Other Custom Origins Only)

44/7777

## Glacier

How long does a Glacier retrieval job take to complete?

83

- Under 1 hour
- 1-2 hours
- 3-5 hours (x)
- Up to 10 hours

From the Glacier FAQ (<https://aws.amazon.com/glacier/faqs/>):

"Q: How can I retrieve data from the service?

You can download data directly from the service using the service's REST API. When you make a request to retrieve data from Glacier, you initiate a retrieval job. Once the retrieval job completes, your data will be available to download for 24 hours. Retrieval jobs typically complete within 3-5 hours."

What is the largest archive that can be uploaded to a Glacier archive in a single request?

84

- 
- 512 KB
  - 4 GB (x)
  - 4 TB
  - 1 PB

From the Glacier FAQ (<https://aws.amazon.com/glacier/faqs/>):

"Q: What is an archive?

An archive is a durably stored block of information. You store your data in Amazon Glacier as archives. You may upload a single file as an archive, but your costs will be lower if you aggregate your data. TAR and ZIP are common formats that customers use to aggregate multiple files into a single file before uploading to Amazon Glacier. The total volume of data and number of archives you can store are unlimited. Individual Amazon Glacier archives can range in size from 1 byte to 40 terabytes. **The largest archive that can be uploaded in a single Upload request is 4 gigabytes.** For items larger than 100 megabytes, customers should consider using the Multipart upload capability. Archives stored in Amazon Glacier are immutable, i.e. archives can be uploaded and deleted but cannot be edited or overwritten."

---

How can you modify a Glacier archive once its been uploaded?

85

- Uploads replace any previous version
- Uploads cannot be edited or overwritten (x)
- Make a request to Amazon to unlock the archive
- Change permissions to allow editing

From the Glacier FAQ (<https://aws.amazon.com/glacier/faqs/>):

"Q: What is an archive?

An archive is a durably stored block of information. You store your data in Amazon Glacier as archives. You may upload a single file as an archive, but your costs will be lower if you aggregate your data. TAR and ZIP are common formats that customers use to aggregate multiple files into a single file before uploading to Amazon Glacier. The total volume of data and number of archives you can store are unlimited. Individual Amazon Glacier archives can range in size from 1 byte to 40 terabytes. **The largest archive that can be uploaded in a single Upload request is 4 gigabytes.** For items larger than 100 megabytes, customers should consider using the Multipart upload capability. **Archives stored in Amazon Glacier are immutable, i.e. archives can be uploaded and deleted but cannot be edited or overwritten.**"

---

What is the smallest possible granularity of the byte range specified for a Glacier range retrieval?

86

- 1 byte
- 1 MB (x)
- 1 GB
- 100 GB

From the Glacier FAQ (<https://aws.amazon.com/glacier/faqs/>):

"Q: Can I retrieve part of an archive?

When initiating a retrieval job using range retrievals, **you provide a byte range that can start at zero (which would be the beginning of your archive), or at any 1MB interval thereafter (e.g. 1MB, 2MB, 3MB, etc).** The end of the range can either be the end of your archive or any 1MB interval greater than the beginning of your range."

also from the Glacier Developer Guide (<http://docs.aws.amazon.com/amazonglacier/latest/dev/downloading-an-archive.html#downloading-an-archive-range>):

"When initiating a retrieval job using range retrievals, you must provide a range that is megabyte aligned. This means that the byte range can start at zero (which would be the beginning of your archive), or at any 1 MB interval thereafter (1 MB, 2 MB, 3 MB, etc.). The end of the range can either be the end of your archive or any 1 MB interval greater than the beginning of your range. Furthermore, if you want to get checksum values when you download the data (after the retrieval job completes), the range you request in the job initiation must also be tree-hash aligned. Checksums are a way you can ensure that your data was not corrupted during transmission. For more information about megabyte alignment and tree-hash alignment, see Receiving Checksums When Downloading Data."

- Yes
- No (x)

From the Glacier Developer Guide (<http://docs.aws.amazon.com/amazonglacier/latest/dev/working-with-archives.html>):

"Except for the optional archive description, Amazon Glacier does not support any additional metadata for the archives. When you upload an archive Amazon Glacier assigns an ID, an opaque sequence of characters, from which you cannot infer any meaning about the archive. You might maintain metadata about the archives on the client-side. The metadata can include archive name and some other meaningful information about the archive."

- Yes (x)
- No

From the Glacier Developer Guide (<http://docs.aws.amazon.com/amazonglacier/latest/dev/uploading-archive-mpu.html>):

"As described in Uploading an Archive in Amazon Glacier, we encourage Amazon Glacier customers to use Multipart Upload to upload archives greater than 100 MB."

The link goes into a lot more detail.

- First initiate a retrieval job, then download the bytes once the job completes (x)
- Submit an email request to Amazon
- Download the bytes immediately using the SDK

From the Glacier Developer Guide (<http://docs.aws.amazon.com/amazonglacier/latest/dev/downloading-an-archive.html>):

"Retrieving an archive from Amazon Glacier is a two-step process:

1. Initiate an archive retrieval job.

When you initiate a job, Amazon Glacier returns a job ID in the response and executes the job asynchronously. After the job completes, you can download the job output.

[...]

2. After the job completes, download the bytes.

[...]"

- Vault (x)
- Archive (x)
- Job (x)
- Notification Configuration (x)
- Owner
- Schema
- Audit Log

From the Glacier Developer Guide (<http://docs.aws.amazon.com/amazonglacier/latest/dev/amazon-glacier-data-model.html>):

"The Amazon Glacier data model core concepts include vaults and archives. Amazon Glacier is a REST-based web service. In terms of REST, vaults and archives are the resources. In addition, the Amazon Glacier data model includes job and notification-configuration resources. These resources complement the core resources.

#### **Vault**

In Amazon Glacier, a vault is a container for storing archives. When you create a vault, you specify a name and select an AWS region where you want to create the vault. Each vault resource has a unique address. [...]

#### **Archive**

An archive can be any data such as a photo, video, or document and is a base unit of storage in Amazon Glacier. Each archive has a unique ID and an optional description. Archive IDs are unique within a vault. Note that you can only specify the optional description during the upload of an archive. Amazon Glacier assigns the archive an ID, which is unique in the AWS region in which it is stored. Each archive has a unique address. [...]

#### **Job**

Retrieving an archive and vault inventory (list of archives) are asynchronous operations in Amazon Glacier in which you first initiate a job, and then download the job output after Amazon Glacier completes the job. With Amazon Glacier, your data retrieval requests are queued and most jobs take about four hours to complete. [...]

#### **Notification Configuration**

Because jobs take time to complete, Amazon Glacier supports a notification mechanism to notify you when a job is complete. You can configure a vault to send notification to an Amazon Simple Notification Service (Amazon SNS) topic when jobs complete. You can specify one SNS topic per vault in the notification configuration [...]."

## **CloudFormation**

Is it possible to automatically create a snapshot of an RDS database created by a CloudFormation stack when the stack is deleted? 91

- Yes (x)
- No

From the CloudFormation documentation:

"DeletionPolicy Options

#### Delete

AWS CloudFormation deletes the resource and all its content if applicable during stack deletion. You can add this deletion policy to any resource type. By default, if you don't specify a DeletionPolicy, AWS CloudFormation deletes your resources.

Note: For Amazon S3 buckets, you must delete all objects in the bucket for deletion to succeed.

#### Retain

AWS CloudFormation keeps the resource without deleting the resource or its contents when its stack is deleted. You can add this deletion policy to any resource type. Note that when AWS CloudFormation completes the stack deletion, the stack will be in Delete\_Complete state; however, resources that are retained continue to exist and continue to incur applicable charges until you delete those resources.

#### Snapshot

**For resources that support snapshots (AWS::EC2::Volume, AWS::RDS::DBInstance, and AWS::Redshift::Cluster), AWS CloudFormation creates a snapshot for the resource before deleting it.** Note that when AWS CloudFormation completes the stack deletion, the stack will be in the Delete\_Complete state; however, the snapshots that are created with this policy continue to exist and continue to incur applicable charges until you delete those snapshots."

What is the default CloudFormation DeletionPolicy?

- 
- Delete (x)
  - Retain
  - Snapshot

From the CloudFormation Documentation (<http://docs.aws.amazon.com/AWSCloudFormation/latest/UserGuide/aws-attribute-deletionpolicy.html>):

"By default, if you don't specify a DeletionPolicy, AWS CloudFormation deletes your resources."

---

When creating an ELB attached to a VPC in a CloudFormation template, what property do you use to control the AZs in which it should be created? 93

- 
- AvailabilityZones
  - Subnets (x)

From the AWS::ElasticLoadBalancing::LoadBalancer CloudFormation resource documentation (<http://docs.aws.amazon.com/AWSCloudFormation/latest/UserGuide/aws-properties-ec2-elb.html#cfn-ec2-elb-availabilityzones>):

"[Description of the] Availability Zones [property]:

The Availability Zones in which to create the load balancer...You can specify the AvailabilityZones or Subnets property, but not both...for load balancers that are in a VPC, specify the Subnets property"

---

What elements of a CloudFormation template are required? 94

- 
- Description
  - Metadata
  - Parameters
  - Mappings
  - Conditions
  - Resources (x)
  - Outputs

From the CloudFormation User Guide (<http://docs.aws.amazon.com/AWSCloudFormation/latest/UserGuide/template-anatomy.html>):

"A template is a JSON-formatted text file that describes your AWS infrastructure. Templates include several major sections. The Resources section is the only section that is required. The first character in the template must be an open brace [...] and the last character must be a closed brace [...]"

Also from the CloudFormation FAQ (<https://aws.amazon.com/cloudformation/faqs/>):

**"Q: What are the elements of an AWS CloudFormation template?"**

AWS CloudFormation templates are JSON formatted text files that are comprised of five types of elements:

1. An optional list of template parameters (input values supplied at stack creation time)
2. An optional list of output values (e.g. the complete URL to a web application)
3. An optional list of data tables used to lookup static configuration values (e.g., AMI names)
4. The list of AWS resources and their configuration values
5. A template file format version number

With parameters, you can customize aspects of your template at run time, when the stack is built. For example, the Amazon RDS database size, Amazon EC2 instance types, database and web server port numbers can be passed to AWS CloudFormation when a stack is created. Each parameter can have a default value and description and may be marked as "NoEcho" in order to hide the actual value you enter on the screen and in the AWS CloudFormation event logs. When you create an AWS CloudFormation stack, the AWS Management Console will automatically synthesize and present a pop-up dialog form for you to edit parameter values.

Output values are a very convenient way to present a stack's key resources (such as the address of an Elastic Load Balancing load balancer or Amazon RDS database) to the user via the AWS Management Console, or the command line tools. You can use simple functions to concatenate string literals and value of attributes associated with the actual AWS resources."

- To specify conditions that must be satisfied before the resource should be created
- To specify an IAM policy for a resource
- To make CloudFormation wait on a signal before completing the creation of the associated resource (x)

From the CloudFormation documentation ():

**"You associate the CreationPolicy attribute with a resource to prevent its status from reaching create complete until AWS CloudFormation receives a specified number of success signals or the timeout period is exceeded.** To signal a resource, you can use the cfn-signal helper script or SignalResource API. AWS CloudFormation publishes valid signals to the stack events so that you track the number of signals sent.

The creation policy is invoked only when AWS CloudFormation creates the associated resource. Currently, the only AWS CloudFormation resources that support creation policies are `AWS::AutoScaling::AutoScalingGroup`, `AWS::EC2::Instance`, and `AWS::CloudFormation::WaitCondition`.

**The CreationPolicy attribute is helpful when you want to wait on resource configuration actions before stack creation proceeds.** For example, if you install and configure software applications on an Amazon EC2 instance, you might want those applications up and running before proceeding. In such cases, you can add a CreationPolicy attribute to the instance and then send a success signal to the instance after the applications are installed and configured. "

- To specify if resources should be replaced or updated in place in a stack update.
- To update a resource's IAM policy
- Controls how CloudFormation responds to updates in an AutoScalingGroup resource (x)

From the CloudFormation User Guide ():

**"You can use the UpdatePolicy attribute to specify how AWS CloudFormation handles updates to the AWS::AutoScaling::AutoScalingGroup resource.**

The update policy is invoked under the following conditions:

- The AutoScalingRollingUpdate policy is applied when you make a change to the Auto Scaling placement group, launch configuration, or subnet group membership of the Auto Scaling group.
- The AutoScalingScheduledAction policy is applied when you update a stack that includes an Auto Scaling group with an associated scheduled action."

- Use the Fn::GetAZs intrinsic function (x)
- Specify a list of AZs in a Mapping element
- Use an cfn-init built-in function

The "cfn-init" choice is just a detractor and is non-sensical.

From the CloudFormation User Guide (<http://docs.aws.amazon.com/AWSCloudFormation/latest/UserGuide/intrinsic-function-reference-getavailabilityzones.html>):

"Fn::GetAZs

The intrinsic function Fn::GetAZs returns an array that lists Availability Zones for a specified region. Because customers have access to different Availability Zones, the intrinsic function Fn::GetAZs enables template authors to write templates that adapt to the calling user's access. That way you don't have to hard-code a full list of Availability Zones for a specified region."

What is the best way to get the region in which a CloudFormation template is running from within the template itself during execution?

---

- Use the Fn::GetRegion intrinsic function
- Specify a list of regions in a Mapping element
- Use the AWS::Region pseudo parameter (x)
- Hard-code the region in the file

There is no Fn::GetRegion function.

From the CloudFormation User Guide (<http://docs.aws.amazon.com/AWSCloudFormation/latest/UserGuide/pseudo-parameter-reference.html>):

"AWS::Region - Returns a string representing the AWS Region in which the encompassing resource is being created, such as us-west-2."

How can you execute a rolling update with CloudFormation?

99



- Create a condition and invoke cfn-signal when each instance is fully configured and added to the ELB
- Use an AutoScalingRollingUpdate UpdatePolicy (x)
- Create a custom resource

The basic idea is that you add an update policy to the ASG resource to tell it to update the ASG (i.e., replace instances) in batches rather than all at once. CloudFormation will guarantee that a certain number of instances remain in service during the upgrade.

From the CloudFormation User Guide (<http://docs.aws.amazon.com/AWSCloudFormation/latest/UserGuide/aws-attribute-updatepolicy.html>):

"You can use the AutoScalingRollingUpdate policy to specify how AWS CloudFormation handles rolling updates for a particular resource."

Example:

```
{
  "ASG": {
    "Type": "AWS::AutoScaling::AutoScalingGroup",
    "Properties": {
      "AvailabilityZones": [
        "us-east-1a",
        "us-east-1b"
      ],
      "DesiredCapacity": "1",
      "LaunchConfigurationName": {
        "Ref": "LaunchConfig"
      },
      "MaxSize": "4",
      "MinSize": "1"
    },
    "UpdatePolicy": {
      "AutoScalingScheduledAction": {
        "IgnoreUnmodifiedGroupSizeProperties": "true"
      },
      "AutoScalingRollingUpdate": {
        "MinInstancesInService": "1",
        "MaxBatchSize": "2",
        "WaitOnResourceSignals": "true",
        "PauseTime": "PT10M"
      }
    }
  },
  "ScheduledAction": {
    "Type": "AWS::AutoScaling::ScheduledAction",
    "Properties": {
      "AutoScalingGroupName": {
        "Ref": "ASG"
      },
      "DesiredCapacity": "2",
      "StartTime": "2017-06-02T20: 00: 00Z"
    }
  }
}
```

- The stack is rolled back (x)
- Creation is halted, but existing resources are left in place
- Notifications are sent

From the CloudFormation FAQ (<https://aws.amazon.com/cloudformation/faqs/>):

**"Q: What happens when one of the resources in a stack cannot be created successfully?"**

By default, the "automatic rollback on error" feature is enabled. This will cause all AWS resources that AWS CloudFormation created successfully for a stack up to the point where an error occurred to be deleted. This is useful when, for example, you accidentally exceed your default limit of Elastic IP addresses, or you don't have access to an EC2 AMI you're trying to run. This feature enables you to rely on the fact that stacks are either fully created, or not at all, which simplifies system administration and layered solutions built on top of AWS CloudFormation."

- Using wait conditions (x)
- Using creation policies and cfn-signal (x)
- Using a custom SNS notification topic
- Not possible

CreationPolicies are the preferred mechanism. From the CloudFormation User Guide ():

"For Amazon EC2 and Auto Scaling resources, we recommend that you use a CreationPolicy attribute instead of wait conditions. Add a CreationPolicy attribute to those resources and use the cfn-signal helper script to signal when an instance has been successfully created."

but either mechanism is valid:

"Using the AWS::CloudFormation::WaitCondition resource and CreationPolicy attribute, you can do the following:

- Coordinate stack resource creation with other configuration actions that are external to the stack creation
- Track the status of a configuration process

**For example, you can start the creation of another resource after an application configuration is partially complete**, or you can send signals during an installation and configuration process to track its progress."

- 
- Yes, for all resources
  - Yes, but only for some resources (x)
  - No

From the CloudFormation FAQ (<https://aws.amazon.com/cloudformation/faqs/>):

**“Q: Why can’t I name all my resources?”**

**Although AWS CloudFormation allows you to name some resources (such as Amazon S3 buckets), CloudFormation doesn’t allow this for all resources.** Naming resources restricts the reusability of templates and results in naming conflicts when an update causes a resource to be replaced. To minimize these issues, CloudFormation will support resource naming on a case by case basis.”

also relevant:

**Q: How does AWS CloudFormation choose actual resource names?**

You can assign logical names to AWS resources in a template. When a stack is created, AWS CloudFormation binds the logical name to the name of the corresponding actual AWS resource. Actual resource names are a combination of the stack and logical resource name. This allows multiple stacks to be created from a template without fear of name collisions between AWS resources.”

---

What are the possible outcomes when CloudFormation updates a resource?

103

- No interruption (x)
- Some interruption (x)
- Full replacement (x)
- Unknown
- Pending approval

From the CloudFormation User Guide (<http://docs.aws.amazon.com/AWSCloudFormation/latest/UserGuide/using-cfn-updating-stacks.html>):

“You modify stack resources by submitting an updated template or by submitting updated input parameters. When you submit an update, AWS CloudFormation updates resources based on differences between what you submit and the stack’s current template. Resources that have not changed run without disruption during the update process. Resources that are updated could be interrupted or replaced, depending on the resources and properties that are being updated. AWS CloudFormation uses one of the following techniques to update resources:

**Update with No Interruption**

AWS CloudFormation updates the resource without disrupting operation of that resource and without changing the resource’s physical name. For example, if you update any properties on an `AWS::CloudWatch::Alarm` resource, AWS CloudFormation updates the alarm’s configuration and, during the update, the alarm’s operation continues without disruption.

**Updates with Some Interruption**

AWS CloudFormation updates the resource with some interruption but the physical name is retained. For example, if you update certain properties on an `AWS::EC2::Instance` resource, the instance might have some interruption while AWS CloudFormation and Amazon EC2 reconfigure the instance.

**Replacement**

AWS CloudFormation recreates the resource during an update, which also generates a new physical ID. AWS CloudFormation creates the replacement resource first, changes references from other dependent resources to point to the replacement resource, and then deletes the old resource. For example, if you update the `Engine` property of an `AWS::RDS::DBInstance` resource, AWS CloudFormation creates a new resource and replaces the current `DBInstance` resource with the new one.

To learn more about updating a particular resource, see the documentation that is associated with that resource. For example, the Amazon EC2 documentation provides details about what changes interrupt an instance. See also the AWS Resource Types Reference, where the effects of updating a resource are listed for each property.”

---

Does CloudFormation change a resource’s name when it updates that resource?

104

- Yes, always
- Yes, but only if the instance needs to be replaced (x)
- Never

From the CloudFormation User Guide (<http://docs.aws.amazon.com/AWSCloudFormation/latest/UserGuide/using-cfn-updating-stacks.html>):

"You modify stack resources by submitting an updated template or by submitting updated input parameters. When you submit an update, AWS CloudFormation updates resources based on differences between what you submit and the stack's current template. Resources that have not changed run without disruption during the update process. Resources that are updated could be interrupted or replaced, depending on the resources and properties that are being updated. AWS CloudFormation uses one of the following techniques to update resources:

#### Update with No Interruption

AWS CloudFormation updates the resource without disrupting operation of that resource and **without changing the resource's physical name**. For example, if you update any properties on an `AWS::CloudWatch::Alarm` resource, AWS CloudFormation updates the alarm's configuration and, during the update, the alarm's operation continues without disruption.

#### Updates with Some Interruption

AWS CloudFormation updates the resource with some interruption but **the physical name is retained**. For example, if you update certain properties on an `AWS::EC2::Instance` resource, the instance might have some interruption while AWS CloudFormation and Amazon EC2 reconfigure the instance.

#### Replacement

AWS CloudFormation recreates the resource during an update, which also **generates a new physical ID**. AWS CloudFormation creates the replacement resource first, changes references from other dependent resources to point to the replacement resource, and then deletes the old resource. For example, if you update the `Engine` property of an `AWS::RDS::DBInstance` resource, AWS CloudFormation creates a new resource and replaces the current `DBInstance` resource with the new one.

To learn more about updating a particular resource, see the documentation that is associated with that resource. For example, the Amazon EC2 documentation provides details about what changes interrupt an instance. See also the AWS Resource Types Reference, where the effects of updating a resource are listed for each property."

## EMR

What are valid scaling options for an EMR cluster?

105

- Add instances to a core instance group (x)
- Remove instances from a core instance group
- Add a core instance group
- Add instances to a master instance group
- Remove instances from a master instance group
- Add a master instance group
- Add a task instance group (x)
- Remove instances from a task instance group (x)
- Add instances to a task instance group (x)

From the EMR Developer Guide (<http://docs.aws.amazon.com/ElasticMapReduce/latest/DeveloperGuide/emr-manage-resize.html>):

**"You cannot shrink the size of the core instance group in a running cluster by reducing the instance count.** However, it is possible to terminate an instance in the core instance group using the AWS CLI or the API. This should be done with caution. Terminating an instance in the core instance group risks data loss, and the instance is not automatically replaced.

Task nodes also run your Hadoop jobs. **After a cluster is running, you can increase or decrease the number of task nodes, and you can add additional task instance groups using the console, CLI, or API."**

I also tested these options out via the Console (the others were grayed out or otherwise not possible)

How do you install non-AWS supported tools and packages on EMR nodes when a cluster is first created?

106

- You can't, it's prohibited
- Using EMR "bootstrap actions" (x)
- Using cloud-init
- Using custom AMLs

This is possible through "bootstrap actions" as described in this blog (<http://blogs.aws.amazon.com/bigdata/post/TxO6EHTHQALSIB/-Getting-Started-with-Amazon-EMR-span-class-matches-Bootstrap-span-Actions>):

"Bootstrap Actions are scripts that run on every machine in the cluster as they are brought online, but before the core Hadoop services like HDFS (name node or data node) and the Hive Metastore are configured and started. For example, Cascading, Apache Spark, and Presto can be deployed to a cluster without any need to communicate with HDFS or Zookeeper."

What are the benefits of the "consistent view" feature of EMRFS?

107

- Eliminates consistency errors as S3 data is passed between steps (x)
- Automatically flushes data to SSD drives on data nodes
- Ensures that there are no inconsistencies in S3 metadata

From the EMR Management Guide (<http://docs.aws.amazon.com/ElasticMapReduce/latest/ManagementGuide/emr-plan-consistent-view.html>):

"EMRFS consistent view monitors Amazon S3 list consistency for objects written by or synced with EMRFS, delete consistency for objects deleted by EMRFS, and read-after-write consistency for new objects written by EMRFS.

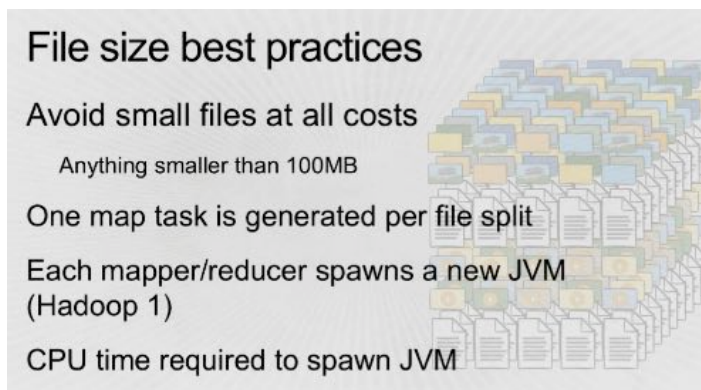
Amazon S3 is designed for eventual consistency. For instance, buckets in the US East (N. Virginia) provide eventual consistency on read-after-write and read-after-overwrite requests. Amazon S3 buckets in the US West (Oregon), US West (N. California), EU (Ireland), EU (Frankfurt), Asia Pacific (Singapore), Asia Pacific (Tokyo), Asia Pacific (Sydney), and South America (Sao Paulo) regions provide read-after-write consistency for put requests of new objects and eventual consistency for overwrite put and delete requests. Therefore, if you are listing objects in an Amazon S3 bucket quickly after putting new objects, Amazon S3 does not provide a guarantee to return a consistent listing and it may be incomplete. This is more common in quick sequential MapReduce jobs which use Amazon S3 as a data store."

Why is it important to avoid small files (under 100MB) when using EMR?

108

- Time is wasted spawning mapper JVMs for each file (x)
- Bandwidth is expensive
- Transfer times can lengthen job completion
- I/O limits can throttle processing

From a 2014 re:Invent presentation (<https://www.youtube.com/watch?v=y5nep2C1eL0&list=PLhr1KZpdzucYj6y9nm5jVtt5fQ1-XC1r&index=12>):



**File size best practices**

- Avoid small files at all costs
- Anything smaller than 100MB
- One map task is generated per file split
- Each mapper/reducer spawns a new JVM (Hadoop 1)
- CPU time required to spawn JVM

What is a good mix of on-demand and spot for a long-running EMR cluster?

109

- Master: on-demand; Core: on-demand, Task: spot (x)
- Master: on-demand; Core: on-demand, Task: on-demand
- Master: on-demand; Core: spot, Task: spot
- Master: spot; Core: spot, Task: spot

From the EMR documentation

(<http://docs.aws.amazon.com/ElasticMapReduce/latest/DeveloperGuide/ChoosingWhattoLaunchasSpotInstances.html>):

The following table shows launch configurations for using Spot Instances in various applications.

Project	Master Instance Group	Core Instance Group	Task Instance Group(s)
Long-running clusters	on-demand	on-demand	spot
Cost-driven workloads	spot	spot	spot
Data-critical workloads	on-demand	on-demand	spot
Application testing	spot	spot	spot

What is a good mix of on-demand and spot for an EMR cluster servicing a cost-driven workload?

110

- Master: on-demand; Core: on-demand, Task: spot
- Master: on-demand; Core: on-demand, Task: on-demand
- Master: on-demand; Core: spot, Task: spot
- Master: spot; Core: spot, Task: spot (x)

From the EMR documentation

(<http://docs.aws.amazon.com/ElasticMapReduce/latest/DeveloperGuide/ChoosingWhattoLaunchasSpotInstances.html>):

The following table shows launch configurations for using Spot Instances in various applications.

Project	Master Instance Group	Core Instance Group	Task Instance Group(s)
Long-running clusters	on-demand	on-demand	spot
Cost-driven workloads	spot	spot	spot
Data-critical workloads	on-demand	on-demand	spot
Application testing	spot	spot	spot

What is a good mix of on-demand and spot for an EMR cluster used for application testing?

111

- Master: on-demand; Core: on-demand, Task: spot
- Master: on-demand; Core: on-demand, Task: on-demand
- Master: on-demand; Core: spot, Task: spot
- Master: spot; Core: spot, Task: spot (x)

From the EMR documentation

(<http://docs.aws.amazon.com/ElasticMapReduce/latest/DeveloperGuide/ChoosingWhattoLaunchasSpotInstances.html>):

The following table shows launch configurations for using Spot Instances in various applications.

Project	Master Instance Group	Core Instance Group	Task Instance Group(s)
Long-running clusters	on-demand	on-demand	spot
Cost-driven workloads	spot	spot	spot
Data-critical workloads	on-demand	on-demand	spot
Application testing	spot	spot	spot

What is a good mix of on-demand and spot for EMR clusters servicing data critical workloads?

112

- Master: on-demand; Core: on-demand, Task: spot (x)
- Master: on-demand; Core: on-demand, Task: on-demand
- Master: on-demand; Core: spot, Task: spot
- Master: spot; Core: spot, Task: spot

From the EMR documentation

(<http://docs.aws.amazon.com/ElasticMapReduce/latest/DeveloperGuide/ChoosingWhattoLaunchasSpotInstances.html>):

The following table shows launch configurations for using Spot Instances in various applications.

Project	Master Instance Group	Core Instance Group	Task Instance Group(s)
Long-running clusters	on-demand	on-demand	spot
Cost-driven workloads	spot	spot	spot
Data-critical workloads	on-demand	on-demand	spot
Application testing	spot	spot	spot

## EBS

What is the availability of EBS General Purpose (SSD) EBS volumes?

113

- 99.9%
- 99.99%
- 99.999% (x)
- 99.9999%

From the EBS FAQ (<https://aws.amazon.com/ebs/faqs/>):

### "Q: What is the EBS General Purpose (SSD) volume type?

The EBS General Purpose (SSD) volumes are backed by the same technology found in EBS Provisioned IOPS (SSD) volumes. The EBS General Purpose (SSD) volume type is designed for 99.999% availability, and a broad range of use-cases such as boot volumes, small and medium size databases, and development and test environments. General Purpose (SSD) volumes deliver a ratio of 3 IOPS per GB, offer single digit millisecond latencies, and also have the ability to burst up to 3000 IOPS for short periods."

What are the baseline IOPS delivered by a General Purpose (SSD) EBS volume?

114

- 1 IOPS/GB
- 3 IOPS/GB (x)
- 5 IOPS/GB
- 10 IOPS/GB

From the EBS FAQ (<https://aws.amazon.com/ebs/faqs/>):

### "Q: What is the EBS General Purpose (SSD) volume type?

The EBS General Purpose (SSD) volumes are backed by the same technology found in EBS Provisioned IOPS (SSD) volumes. The EBS General Purpose (SSD) volume type is designed for 99.999% availability, and a broad range of use-cases such as boot volumes, small and medium size databases, and development and test environments. General Purpose (SSD) volumes deliver a ratio of 3 IOPS per GB, offer single digit millisecond latencies, and also have the ability to burst up to 3000 IOPS for short periods."

What is the burst capacity of a General Purpose (SSD) EBS volume?

115

- 
- 1000 IOPS for short periods
  - 3000 IOPS for short periods (x)
  - 10000 IOPS for short periods
  - 20000 IOPS for short periods

From the EBS FAQ (<https://aws.amazon.com/ebs/faqs/>):

**"Q: What is the EBS General Purpose (SSD) volume type?"**

The EBS General Purpose (SSD) volumes are backed by the same technology found in EBS Provisioned IOPS (SSD) volumes. The EBS General Purpose (SSD) volume type is designed for 99.999% availability, and a broad range of use-cases such as boot volumes, small and medium size databases, and development and test environments. General Purpose (SSD) volumes deliver a ratio of 3 IOPS per GB, offer single digit millisecond latencies, and also have the ability to burst up to 3000 IOPS for short periods."

---

Can I access an EBS volume snapshot via the S3 APIs?

116

- Yes
- No (x)

From the EBS FAQ (<https://aws.amazon.com/ebs/faqs/>):

**"Q: Will I be able to access my snapshots using the regular Amazon S3 APIs?"**

No, snapshots are only available through the Amazon EC2 APIs."

---

Must you detach an EBS volume before taking a snapshot?

117

- Yes
- No
- No, but it's recommended for consistency (x)

From the EBS FAQ (<https://aws.amazon.com/ebs/faqs/>):

**"Q: Do volumes need to be un-mounted in order to take a snapshot? Does the snapshot need to complete before the volume can be used again?"**

No, snapshots can be done in real time while the volume is attached and in use. However, snapshots only capture data that has been written to your Amazon EBS volume, which might exclude any data that has been locally cached by your application or OS. **In order to ensure consistent snapshots on volumes attached to an instance, we recommend cleanly detaching the volume**, issuing the snapshot command, and then reattaching the volume. For Amazon EBS volumes that serve as root devices, we recommend shutting down the machine to take a clean snapshot."

---

What happens to an EBS volume that is used as an EC2 instance root partition when that instance is terminated?

118

- Always deleted
- Possibly deleted based on instance configuration (x)
- Possibly deleted based on volume configuration
- Never deleted

From the EBS FAQ (<https://aws.amazon.com/ebs/faqs/>):

**"Q : What happens to my data when a system terminates?"**

The data stored on a local instance store will persist only as long as that instance is alive. However, data that is stored on an Amazon EBS volume will persist independently of the life of the instance. Therefore, we recommend that you use the local instance store for temporary data and, for data requiring a higher level of durability, we recommend using Amazon EBS volumes or backing up the data to Amazon S3. **If you are using an Amazon EBS volume as a root partition, you will need to set the Delete On Terminate flag to "N" if you want your Amazon EBS volume to persist outside the life of the instance.**"



- Yes (x)
- No

From the EC2 documentation (<http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/ebs-modifying-snapshot-permissions.html>):

"You can share your unencrypted snapshots with your co-workers or others in the AWS community by modifying the permissions of the snapshot. Users that you have authorized can quickly use your unencrypted shared snapshots as the basis for creating their own EBS volumes. If you choose, you can also make your unencrypted snapshots available publicly to all AWS users."

- Yes (x)
- No

From the EC2 documentation (<http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/ebs-copy-snapshot.html>):

"With Amazon EBS, you can create point-in-time snapshots of volumes which we store for you in Amazon Simple Storage Service (Amazon S3). After you've created a snapshot and it has finished copying to Amazon S3 (when the snapshot status is "completed"), **you can copy it from one AWS region to another**, or within the same region."

- Yes
- No (x)

From the EBS FAQ (<https://aws.amazon.com/ebs/faqs/>):

**"Q: Does EBS encryption support boot volumes?"**

No, at this time EBS encryption is only supported for data volumes."

- Yes
- No (x)

There's nothing on this specific topic in the AWS documentation, but it's not possible. A web search will pull up lots of related articles and alternative approaches.

- None, Amazon guarantees a high level of availability regardless of any customer-influenced factors
- Amount of modified data since the last snapshot (x)
- Volume size (x)
- Utilization (what % of the time is it attached to an instance)

From the "AWS Storage Options" white paper ([http://media.amazonwebservices.com/AWS\\_Storage\\_Options.pdf](http://media.amazonwebservices.com/AWS_Storage_Options.pdf)):

"The durability of your Amazon EBS volume depends on both the **size of your volume** and the **amount of data that has changed since your last snapshot**. Amazon EBS snapshots are incremental, point-in-time backups, containing only the data blocks changed since the last snapshot. Amazon EBS volumes that operate with 20 GB or less of modified data since their most recent snapshot can expect an annual failure rate (AFR) between 0.1% and 0.5%. Amazon EBS volumes with more than 20 GB of unmodified data since the last snapshot should expect higher failure rates that are roughly proportional to the increase in modified data."

- Full provisioned capacity (x)
- Only what's used

From the "AWS Storage Options" white paper ([http://media.amazonwebservices.com/AWS\\_Storage\\_Options.pdf](http://media.amazonwebservices.com/AWS_Storage_Options.pdf)):

**"It's important to remember that for Amazon EBS volumes, you are charged for provisioned (allocated) storage, whether or not you actually use it. For Amazon EBS snapshots, you are charged only for storage actually used (consumed)."**

- Yes
- No (x)

From the "AWS Storage Options" white paper ([http://media.amazonwebservices.com/AWS\\_Storage\\_Options.pdf](http://media.amazonwebservices.com/AWS_Storage_Options.pdf)):

" While individual Amazon EBS volumes cannot be resized, if you find that you need additional storage, you have two ways to expand the amount of Amazon EBS space available for your Amazon EC2 instance. The simplest approach is to create and attach a new Amazon EBS volume and begin using it Amazon Web Services – AWS Storage Options October 2013 Page 11 of 34 together with your existing ones. However, if you need to expand the size of a single Amazon EBS volume, you can effectively resize a volume using a snapshot:

1. Detach the original Amazon EBS volume.
2. Create a snapshot of the original Amazon EBS volume's data into Amazon S3.
3. Create a new Amazon EBS volume from the snapshot, but specify a larger size than the original volume.
4. Attach the new, larger volume to your Amazon EC2 instance in place of the original. (In many cases, an OS-level utility must also be used to expand the file system.)
5. Delete the original Amazon EBS volume."

- System boot volumes (x)
- Virtual desktops (x)
- Small to medium sized databases (x)
- Development and test environments (x)
- Critical business applications that require sustained IOPS
- Scenarios where the lowest storage cost is important

From the EC2 User Guide (<http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/EBSVolumeTypes.html>):

Characteristic	General Purpose (SSD)	Provisioned IOPS (SSD)	Magnetic
Use cases	<ul style="list-style-type: none"> <li>• System boot volumes</li> <li>• Virtual desktops</li> <li>• Small to medium sized databases</li> <li>• Development and test environments</li> </ul>	<ul style="list-style-type: none"> <li>• Critical business applications that require sustained IOPS performance, or more than 10,000 IOPS or 160 MiB/s of throughput per volume</li> <li>• Large database workloads, such as: <ul style="list-style-type: none"> <li>◦ MongoDB</li> <li>◦ Microsoft SQL Server</li> <li>◦ MySQL</li> <li>◦ PostgreSQL</li> <li>◦ Oracle</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>• Cold workloads where data is infrequently accessed</li> <li>• Scenarios where the lowest storage cost is important</li> </ul>
Volume size	1 GiB - 16 TiB	4 GiB - 16 TiB	1 GiB - 1 TiB
Maximum throughput	160 MiB/s	320 MiB/s	40-90 MiB/s
IOPS performance	Baseline performance of 3 IOPS/GiB (up to 10,000 IOPS) with the ability to burst to 3,000 IOPS for volumes under 1,000 GiB. See <a href="#">I/O Credits and Burst Performance</a>	Consistently performs at provisioned level, up to 20,000 IOPS maximum	Averages 100 IOPS, with the ability to burst to hundreds of IOPS
API and CLI volume name	gp2	io1	standard

What is the maximum throughput of a General Purpose SSD EBS volume?

127

- 160 MiB/s (x)
- 40-90 MiB/s
- 320 MiB/s
- 640 MiB/s

From the EC2 User Guide (<http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/EBSVolumeTypes.html>):

Characteristic	General Purpose (SSD)	Provisioned IOPS (SSD)	Magnetic
Use cases	<ul style="list-style-type: none"> <li>• System boot volumes</li> <li>• Virtual desktops</li> <li>• Small to medium sized databases</li> <li>• Development and test environments</li> </ul>	<ul style="list-style-type: none"> <li>• Critical business applications that require sustained IOPS performance, or more than 10,000 IOPS or 160 MiB/s of throughput per volume</li> <li>• Large database workloads, such as: <ul style="list-style-type: none"> <li>◦ MongoDB</li> <li>◦ Microsoft SQL Server</li> <li>◦ MySQL</li> <li>◦ PostgreSQL</li> <li>◦ Oracle</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>• Cold workloads where data is infrequently accessed</li> <li>• Scenarios where the lowest storage cost is important</li> </ul>
Volume size	1 GiB - 16 TiB	4 GiB - 16 TiB	1 GiB - 1 TiB
Maximum throughput	160 MiB/s	320 MiB/s	40-90 MiB/s
IOPS performance	Baseline performance of 3 IOPS/GiB (up to 10,000 IOPS) with the ability to burst to 3,000 IOPS for volumes under 1,000 GiB. See <a href="#">I/O Credits and Burst Performance</a>	Consistently performs at provisioned level, up to 20,000 IOPS maximum	Averages 100 IOPS, with the ability to burst to hundreds of IOPS
API and CLI volume name	gp2	io1	standard

What is the IOPS performance of a Magnetic EBS volume?

128

- Averages 100 IOPS with the ability to burst to hundreds of IOPS (x)
- Averages 100 IOPS with the ability to burst to thousands of IOPS
- Averages 1000 IOPS with the ability to burst to thousands of IOPS

From the EC2 User Guide (<http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/EBSVolumeTypes.html>):

Characteristic	General Purpose (SSD)	Provisioned IOPS (SSD)	Magnetic
Use cases	<ul style="list-style-type: none"> <li>• System boot volumes</li> <li>• Virtual desktops</li> <li>• Small to medium sized databases</li> <li>• Development and test environments</li> </ul>	<ul style="list-style-type: none"> <li>• Critical business applications that require sustained IOPS performance, or more than 10,000 IOPS or 160 MiB/s of throughput per volume</li> <li>• Large database workloads, such as: <ul style="list-style-type: none"> <li>◦ MongoDB</li> <li>◦ Microsoft SQL Server</li> <li>◦ MySQL</li> <li>◦ PostgreSQL</li> <li>◦ Oracle</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>• Cold workloads where data is infrequently accessed</li> <li>• Scenarios where the lowest storage cost is important</li> </ul>
Volume size	1 GiB - 16 TiB	4 GiB - 16 TiB	1 GiB - 1 TiB
Maximum throughput	160 MiB/s	320 MiB/s	40-90 MiB/s
IOPS performance	Baseline performance of 3 IOPS/GiB (up to 10,000 IOPS) with the ability to burst to 3,000 IOPS for volumes under 1,000 GiB. See <a href="#">I/O Credits and Burst Performance</a>	Consistently performs at provisioned level, up to 20,000 IOPS maximum	Averages 100 IOPS, with the ability to burst to hundreds of IOPS
API and CLI volume name	gp2	io1	standard

What is the IOPS performance of a General Purpose SSD volume?

- Baseline performance of 3 IOPS/GB (up to 10K IOPS) with the ability to burst to 3K IOPS for volumes under 1TB (x)
- Baseline performance of 10 IOPS/GB (up to 10K IOPS) with the ability to burst to 100K IOPS for volumes under 1TB
- Consistent baseline performance of 10K IOPS

From the EC2 User Guide (<http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/EBSVolumeTypes.html>):

Characteristic	General Purpose (SSD)	Provisioned IOPS (SSD)	Magnetic
Use cases	<ul style="list-style-type: none"> <li>• System boot volumes</li> <li>• Virtual desktops</li> <li>• Small to medium sized databases</li> <li>• Development and test environments</li> </ul>	<ul style="list-style-type: none"> <li>• Critical business applications that require sustained IOPS performance, or more than 10,000 IOPS or 160 MiB/s of throughput per volume</li> <li>• Large database workloads, such as: <ul style="list-style-type: none"> <li>◦ MongoDB</li> <li>◦ Microsoft SQL Server</li> <li>◦ MySQL</li> <li>◦ PostgreSQL</li> <li>◦ Oracle</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>• Cold workloads where data is infrequently accessed</li> <li>• Scenarios where the lowest storage cost is important</li> </ul>
Volume size	1 GiB - 16 TiB	4 GiB - 16 TiB	1 GiB - 1 TiB
Maximum throughput	160 MiB/s	320 MiB/s	40-90 MiB/s
IOPS performance	Baseline performance of 3 IOPS/GiB (up to 10,000 IOPS) with the ability to burst to 3,000 IOPS for volumes under 1,000 GiB. See <a href="#">I/O Credits and Burst Performance</a>	Consistently performs at provisioned level, up to 20,000 IOPS maximum	Averages 100 IOPS, with the ability to burst to hundreds of IOPS
API and CLI volume name	gp2	io1	standard

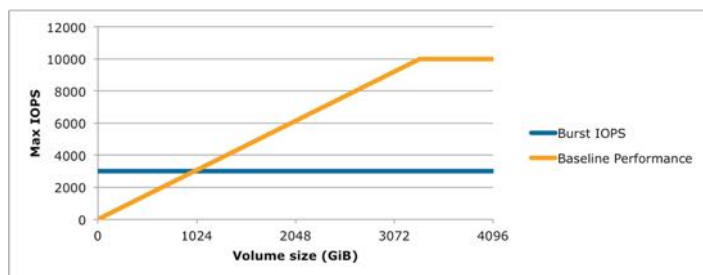
Also from the EC2 User Guide ():

“““

#### I/O Credits and Burst Performance

General Purpose (SSD) volume performance is governed by volume size, which dictates the baseline performance level of the volume and how quickly it accumulates I/O credits; larger volumes have higher baseline performance levels and accumulate I/O credits faster. I/O credits represent the available bandwidth that your General Purpose (SSD) volume can use to burst large amounts of I/O when more than the baseline performance is needed. The more credits your volume has for I/O, the more time it can burst beyond its baseline performance level and the better it performs when more performance is needed.

Each volume receives an initial I/O credit balance of 5,400,000 I/O credits, which is enough to sustain the maximum burst performance of 3,000 IOPS for 30 minutes. This initial credit balance is designed to provide a fast initial boot cycle for boot volumes and to provide a good bootstrapping experience for other applications. Volumes earn I/O credits every second at a baseline performance rate of 3 IOPS per GiB of volume size. For example, a 100 GiB General Purpose (SSD) volume has a baseline performance of 300 IOPS.



“””

What is the IOPS performance of a Provisioned IOPS volume?

- Consistently performs at provisioned level up to 20K IOPS (x)
- Performs within 98% of provisioned level up to 20K IOPS
- Performs within 95% of provisioned level up to 20K IOPS

From the EC2 User Guide (<http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/EBSVolumeTypes.html>):

Characteristic	General Purpose (SSD)	Provisioned IOPS (SSD)	Magnetic
Use cases	<ul style="list-style-type: none"> <li>• System boot volumes</li> <li>• Virtual desktops</li> <li>• Small to medium sized databases</li> <li>• Development and test environments</li> </ul>	<ul style="list-style-type: none"> <li>• Critical business applications that require sustained IOPS performance, or more than 10,000 IOPS or 160 MiB/s of throughput per volume</li> <li>• Large database workloads, such as: <ul style="list-style-type: none"> <li>◦ MongoDB</li> <li>◦ Microsoft SQL Server</li> <li>◦ MySQL</li> <li>◦ PostgreSQL</li> <li>◦ Oracle</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>• Cold workloads where data is infrequently accessed</li> <li>• Scenarios where the lowest storage cost is important</li> </ul>
Volume size	1 GiB - 16 TiB	4 GiB - 16 TiB	1 GiB - 1 TiB
Maximum throughput	160 MiB/s	320 MiB/s	40-90 MiB/s
IOPS performance	Baseline performance of 3 IOPS/GiB (up to 10,000 IOPS) with the ability to burst to 3,000 IOPS for volumes under 1,000 GiB. See <a href="#">I/O Credits and Burst Performance</a>	Consistently performs at provisioned level, up to 20,000 IOPS maximum	Averages 100 IOPS, with the ability to burst to hundreds of IOPS
API and CLI volume name	gp2	io1	standard

What is initial I/O credit balance for a General Purpose SSD EBS volume?

131

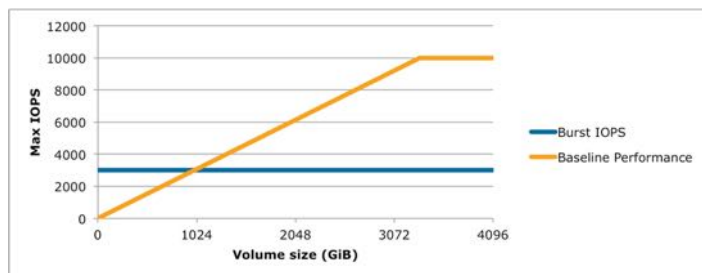
- 5,400,000 (sustained 3K IOPS for 30 minutes) (x)
- 2,700,000 (sustained 3K IOPS for 15 minutes)
- 900,000 (sustained 3K IOPS for 5 minutes)

From the EC2 User Guide (<http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/EBSVolumeTypes.html#IOcredit>):

“”

#### I/O Credits and Burst Performance

**Each volume receives an initial I/O credit balance of 5,400,000 I/O credits, which is enough to sustain the maximum burst performance of 3,000 IOPS for 30 minutes.** This initial credit balance is designed to provide a fast initial boot cycle for boot volumes and to provide a good bootstrapping experience for other applications. Volumes earn I/O credits every second at a baseline performance rate of 3 IOPS per GiB of volume size. For example, a 100 GiB General Purpose (SSD) volume has a baseline performance of 300 IOPS.



“”

What is the maximum throughput of a Provisioned IOPS EBS volume?

132

- 160 MiB/s
- 40-90 MiB/s
- 320 MiB/s (x)
- 640 MiB/s

From the EC2 User Guide (<http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/EBSVolumeTypes.html>):

Characteristic	General Purpose (SSD)	Provisioned IOPS (SSD)	Magnetic
Use cases	<ul style="list-style-type: none"> <li>• System boot volumes</li> <li>• Virtual desktops</li> <li>• Small to medium sized databases</li> <li>• Development and test environments</li> </ul>	<ul style="list-style-type: none"> <li>• Critical business applications that require sustained IOPS performance, or more than 10,000 IOPS or 160 MiB/s of throughput per volume</li> <li>• Large database workloads, such as:               <ul style="list-style-type: none"> <li>◦ MongoDB</li> <li>◦ Microsoft SQL Server</li> <li>◦ MySQL</li> <li>◦ PostgreSQL</li> <li>◦ Oracle</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>• Cold workloads where data is infrequently accessed</li> <li>• Scenarios where the lowest storage cost is important</li> </ul>
Volume size	1 GiB - 16 TiB	4 GiB - 16 TiB	1 GiB - 1 TiB
Maximum throughput	160 MiB/s	320 MiB/s	40-90 MiB/s
IOPS performance	Baseline performance of 3 IOPS/GiB (up to 10,000 IOPS) with the ability to burst to 3,000 IOPS for volumes under 1,000 GiB. See <a href="#">I/O Credits and Burst Performance</a>	Consistently performs at provisioned level, up to 20,000 IOPS maximum	Averages 100 IOPS, with the ability to burst to hundreds of IOPS
API and CLI volume name	gp2	io1	standard

What is the maximum throughput of a Magnetic EBS volume?

133

- 160 MiB/s
- 40-90 MiB/s (x)
- 320 MiB/s
- 640 MiB/s

From the EC2 User Guide (<http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/EBSVolumeTypes.html>):

Characteristic	General Purpose (SSD)	Provisioned IOPS (SSD)	Magnetic
Use cases	<ul style="list-style-type: none"> <li>• System boot volumes</li> <li>• Virtual desktops</li> <li>• Small to medium sized databases</li> <li>• Development and test environments</li> </ul>	<ul style="list-style-type: none"> <li>• Critical business applications that require sustained IOPS performance, or more than 10,000 IOPS or 160 MiB/s of throughput per volume</li> <li>• Large database workloads, such as:               <ul style="list-style-type: none"> <li>◦ MongoDB</li> <li>◦ Microsoft SQL Server</li> <li>◦ MySQL</li> <li>◦ PostgreSQL</li> <li>◦ Oracle</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>• Cold workloads where data is infrequently accessed</li> <li>• Scenarios where the lowest storage cost is important</li> </ul>
Volume size	1 GiB - 16 TiB	4 GiB - 16 TiB	1 GiB - 1 TiB
Maximum throughput	160 MiB/s	320 MiB/s	40-90 MiB/s
IOPS performance	Baseline performance of 3 IOPS/GiB (up to 10,000 IOPS) with the ability to burst to 3,000 IOPS for volumes under 1,000 GiB. See <a href="#">I/O Credits and Burst Performance</a>	Consistently performs at provisioned level, up to 20,000 IOPS maximum	Averages 100 IOPS, with the ability to burst to hundreds of IOPS
API and CLI volume name	gp2	io1	standard

What are some good use cases for Provisioned IOPS EBS volumes?

134



- Critical business applications that require sustained IOPS performance (x)
- Applications that require more than 10k IOPS (x)
- Applications that require more than 160 MiB/s throughput (x)
- Large databases (x)
- Scenarios where lowest storage costs is important

From the EC2 User Guide (<http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/EBSVolumeTypes.html>):

Characteristic	General Purpose (SSD)	Provisioned IOPS (SSD)	Magnetic
Use cases	<ul style="list-style-type: none"> <li>• System boot volumes</li> <li>• Virtual desktops</li> <li>• Small to medium sized databases</li> <li>• Development and test environments</li> </ul>	<ul style="list-style-type: none"> <li>• Critical business applications that require sustained IOPS performance, or more than 10,000 IOPS or 160 MiB/s of throughput per volume</li> <li>• Large database workloads, such as: <ul style="list-style-type: none"> <li>◦ MongoDB</li> <li>◦ Microsoft SQL Server</li> <li>◦ MySQL</li> <li>◦ PostgreSQL</li> <li>◦ Oracle</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>• Cold workloads where data is infrequently accessed</li> <li>• Scenarios where the lowest storage cost is important</li> </ul>
Volume size	1 GiB - 16 TiB	4 GiB - 16 TiB	1 GiB - 1 TiB
Maximum throughput	160 MiB/s	320 MiB/s	40-90 MiB/s
IOPS performance	Baseline performance of 3 IOPS/GiB (up to 10,000 IOPS) with the ability to burst to 3,000 IOPS for volumes under 1,000 GiB. See <a href="#">I/O Credits and Burst Performance</a>	Consistently performs at provisioned level, up to 20,000 IOPS maximum	Averages 100 IOPS, with the ability to burst to hundreds of IOPS
API and CLI volume name	gp2	io1	standard

What are some good use cases for Magnetic SSD EBS volumes?

135

- Cold workloads with infrequently accessed data (x)
- Scenarios where lowest storage costs is important (x)
- Applications that require higher levels of durability
- Applications that run traditional ERP applications

From the EC2 User Guide (<http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/EBSVolumeTypes.html>):

Characteristic	General Purpose (SSD)	Provisioned IOPS (SSD)	Magnetic
Use cases	<ul style="list-style-type: none"> <li>• System boot volumes</li> <li>• Virtual desktops</li> <li>• Small to medium sized databases</li> <li>• Development and test environments</li> </ul>	<ul style="list-style-type: none"> <li>• Critical business applications that require sustained IOPS performance, or more than 10,000 IOPS or 160 MiB/s of throughput per volume</li> <li>• Large database workloads, such as: <ul style="list-style-type: none"> <li>◦ MongoDB</li> <li>◦ Microsoft SQL Server</li> <li>◦ MySQL</li> <li>◦ PostgreSQL</li> <li>◦ Oracle</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>• Cold workloads where data is infrequently accessed</li> <li>• Scenarios where the lowest storage cost is important</li> </ul>
Volume size	1 GiB - 16 TiB	4 GiB - 16 TiB	1 GiB - 1 TiB
Maximum throughput	160 MiB/s	320 MiB/s	40-90 MiB/s
IOPS performance	Baseline performance of 3 IOPS/GiB (up to 10,000 IOPS) with the ability to burst to 3,000 IOPS for volumes under 1,000 GiB. See <a href="#">I/O Credits and Burst Performance</a>	Consistently performs at provisioned level, up to 20,000 IOPS maximum	Averages 100 IOPS, with the ability to burst to hundreds of IOPS
API and CLI volume name	gp2	io1	standard

What is the precise definition of an EBS IOPS?

136



- 
- One I/O per second that is 256KiB or smaller (x)
  - One I/O per second that is 512KiB or smaller
  - One I/O per second that is 1024KiB or smaller
  - One I/O per second that is 4096KiB or smaller

From the EC2 User Guide (<http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/ebs-io-characteristics.html>):

"IOPS are input/output operations per second. **Amazon EBS measures each I/O operation per second (that is 256 KiB or smaller) as one IOPS.** I/O operations that are larger than 256 KiB are counted in 256 KiB capacity units. For example, a single 1,024 KiB I/O operation would count as 4 IOPS; however, 1,024 I/O operations at 1 KiB each would count as 1,024 IOPS."

What is an "I/O Credit Balance" in reference to General Purpose SSD EBS volumes?

137

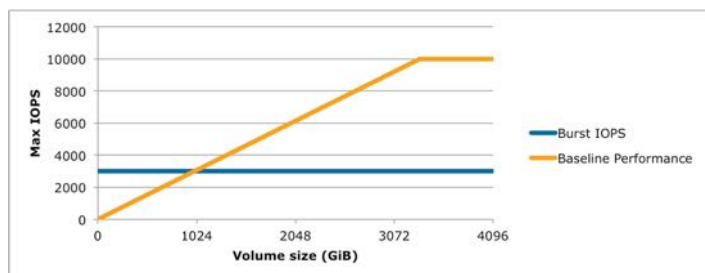
- A pool of IOPS that a volume can utilize to burst beyond baseline performance (x)
- Unused IOPS that reduce the cost of lightly used volumes
- The currently available IOPS on a volume

From the EC2 User Guide (<http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/EBSVolumeTypes.html#IOcredit>):

427791

General Purpose (SSD) volume performance is governed by volume size, which dictates the baseline performance level of the volume and how quickly it accumulates I/O credits; larger volumes have higher baseline performance levels and accumulate I/O credits faster. I/O credits represent the available bandwidth that your General Purpose (SSD) volume can use to burst large amounts of I/O when more than the baseline performance is needed. The more credits your volume has for I/O, the more time it can burst beyond its baseline performance level and the better it performs when more performance is needed.

Each volume receives an initial I/O credit balance of 5,400,000 I/O credits, which is enough to sustain the maximum burst performance of 3,000 IOPS for 30 minutes. This initial credit balance is designed to provide a fast initial boot cycle for boot volumes and to provide a good bootstrapping experience for other applications. Volumes earn I/O credits every second at a baseline performance rate of 3 IOPS per GiB of volume size. For example, a 100 GiB General Purpose (SSD) volume has a baseline performance of 300 IOPS.



When your volume requires more than the baseline performance I/O level, it simply uses I/O credits in the credit balance to burst to the required performance level, up to a maximum of 3,000 IOPS. Volumes larger than 1,000 GiB have a baseline performance that is equal or greater than the maximum burst performance, and their I/O credit balance never depletes. When your volume uses fewer I/O credits than it earns in a second, unused I/O credits are added to the I/O credit balance. The maximum I/O credit balance for a volume is equal to the initial credit balance (5,400,000 I/O credits).

The table below lists several volume sizes and the associated baseline performance of the volume (which is also the rate at which it accumulates I/O credits), the burst duration at the 3,000 IOPS maximum (when starting with a full credit balance), and the time in seconds that the volume would take to refill an empty credit balance.

Volume size (GiB)	Baseline performance (IOPS)	Maximum burst duration @ 3,000 IOPS (seconds)	Seconds to fill empty credit balance
1	3	1,802	1,800,000
100	300	2,000	18,000
214 (Min size for max throughput)	642	2,290	15,790
250	750	2,400	7,200
500	1,500	3,600	3,600
750	2,250	7,200	2,400
1,000	3,000	N/A*	N/A*
3,334 (Min size for max IOPS)	10,000	N/A*	N/A*
16,384 (16 TiB, Max volume size)	10,000	N/A*	N/A*

\* Bursting and I/O credits are only relevant to volumes under 1,000 GiB, where burst performance exceeds baseline performance.

The burst duration of a volume is dependent on the size of the volume, the burst IOPS required, and the credit balance when the burst begins. This is shown in the equation below:

$$\text{Burst duration} = \frac{(\text{Credit balance})}{(\text{Burst IOPS}) - 3(\text{Volume size in GiB})}$$

427791

- 
- Yes
  - No (x)

From the EC2 User Guide (<http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/EBSEncryption.html>):

"This feature is supported with all EBS volume types (General Purpose (SSD), Provisioned IOPS (SSD), and Magnetic), and **you can expect the same IOPS performance on encrypted volumes as you would with unencrypted volumes, with a minimal effect on latency**. You can access encrypted volumes the same way that you access existing volumes; encryption and decryption are handled transparently and they require no additional action from you, your EC2 instance, or your application."

---

What EBS volume types support encryption?

139

- Magnetic (x)
- General Purpose SSD (x)
- Provisioned IOPS (x)

From the EC2 User Guide (<http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/EBSEncryption.html>):

**"This feature is supported with all EBS volume types (General Purpose (SSD), Provisioned IOPS (SSD), and Magnetic)**, and you can expect the same IOPS performance on encrypted volumes as you would with unencrypted volumes, with a minimal effect on latency. You can access encrypted volumes the same way that you access existing volumes; encryption and decryption are handled transparently and they require no additional action from you, your EC2 instance, or your application."

---

Are the snapshots of an encrypted EBS snapshot also encrypted?

140

- Yes (x)
- No

From the EC2 User Guide (<http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/EBSEncryption.html>):

**"Snapshots that are taken from encrypted volumes are automatically encrypted with the same volume encryption key used to encrypt the volume.** Volumes that are created from encrypted snapshots are also automatically encrypted with the same volume encryption key used to create the snapshot. There is no way to directly create an unencrypted volume from an encrypted snapshot; however, you can create an encrypted snapshot from an unencrypted snapshot by creating an encrypted copy of the unencrypted snapshot. For more information, see Copying an Amazon EBS Snapshot."

---

Is it possible to enable encryption for a currently unencrypted EBS volume?

141

- Yes
- No (x)

From the EC2 User Guide (<http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/EBSEncryption.html>):

"There is also no way to encrypt an existing volume. However, you can migrate existing data between encrypted volumes and unencrypted volumes."

More detail about migrating data is here

([http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/EBSEncryption.html#EBSEncryption\\_migrating\\_data](http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/EBSEncryption.html#EBSEncryption_migrating_data)):

"If you have existing data that you would like to store on an encrypted volume, you need to migrate the data from your unencrypted volume to a new encrypted volume.

Likewise, if you have data that currently resides on an encrypted volume that you would like to share with others, you need to migrate the data you want to share from your encrypted volume to a new unencrypted volume."

It goes on to describe how to do this using rsync.

What can you do to achieve a higher I/O throughput than is capable with any provisionable EBS volume?

142

- Join multiple General Purpose SSD or Provisioned IOPS volumes together in a RAID 0 configuration (x)
- Manually shard data across the partitions
- Not possible

From the EC2 documentation (<http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/EBSPerformance.html>):

"Some instance types can drive more I/O throughput than what you can provision for a single Amazon EBS volume. You can join multiple General Purpose (SSD) or Provisioned IOPS (SSD) volumes together in a RAID 0 configuration to use the available bandwidth for these instances. You can also provide redundancy for your volumes with a RAID 1 (mirrored) configuration. "

What should you do if you are getting increased latency when using a newly provisioned EBS volume?

143

- Pre-warm/initialize the volume (x)
- Specify the "pre-format" option when creating the volume
- Write dummy data to each new block before the first read

From the EC2 User Guide (<http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/EBSPerformance.html>):

"There is a significant increase in latency when you first access each block of data on a new EBS volume that was restored from a snapshot (General Purpose (SSD), Provisioned IOPS (SSD), or Magnetic). **You can avoid this performance hit by accessing each block in advance; this process is called *initialization* (formerly known as pre-warming).** "

What is an "EBS Optimized" EC2 instance?

144

- An instance with dedicated network I/O and throughput to EBS volumes (x)
- An instance that can more quickly attach and detach from EBS volumes
- An instance that utilizes EBS volumes more efficiently, reducing cost

From the EC2 User Guide ():

**"An Amazon EBS-optimized instance uses an optimized configuration stack and provides additional, dedicated capacity for Amazon EBS I/O.** This optimization provides the best performance for your EBS volumes by minimizing contention between Amazon EBS I/O and other traffic from your instance.

**EBS-optimized instances deliver dedicated throughput to Amazon EBS,** with options between 500 Mbps and 4,000 Mbps, depending on the instance type you use. When attached to an EBS-optimized instance, General Purpose (SSD) volumes are designed to deliver within 10 percent of their baseline and burst performance 99.9 percent of the time in a given year, and Provisioned IOPS (SSD) volumes are designed to deliver within 10 percent of their provisioned performance 99.9 percent of the time in a given year."

Are EBS volumes highly available?

145

- Yes (x)
- No

From the EC2 User Guide (<http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/AmazonEBS.html>):

"When you create an EBS volume in an Availability Zone, **it is automatically replicated within that zone to prevent data loss due to failure of any single hardware component.** After you create a volume, you can attach it to any EC2 instance in the same Availability Zone. "

[Note: this means that it doesn't make sense to use Raid for redundancy, as noted in the re:Invent presentation <https://www.youtube.com/watch?v=xtuRD2AZt3M>]

- Yes
- No (x)

From the EC2 User Guide (<http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/AmazonEBS.html>):

"An EBS volume can be attached to only one instance at a time **within the same Availability Zone**. However, multiple volumes can be attached to a single instance. If you attach multiple volumes to a device that you have named, you can stripe data across the volumes for increased I/O and throughput performance."

## CloudWatch

How long are CloudWatch metrics stored?

147

- 1 week
- 2 weeks (x)
- 3 weeks
- 4 weeks

From the CloudWatch FAQ (<https://aws.amazon.com/cloudwatch/faqs>):

### "Q: How long are my metrics stored?"

Metrics data are available for 2 weeks. If you want to store metrics data beyond that duration, you can retrieve it using our GetMetricStatistics API as well as a number of applications and tools offered by AWS partners."

What is the CloudWatch monitoring default sampling frequency (without detailed monitoring enabled)?

148

- 1 minute
- 2 minutes
- 5 minutes (x)
- 10 minutes

From the CloudWatch documentation (<http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/using-cloudwatch.html>):

"You can monitor your Amazon EC2 instances using Amazon CloudWatch, which collects and processes raw data from Amazon EC2 into readable, near real-time metrics. These statistics are recorded for a period of two weeks, so that you can access historical information and gain a better perspective on how your web application or service is performing. By default, Amazon EC2 metric data is automatically sent to CloudWatch in 5-minute periods. You can, however, enable detailed monitoring on an Amazon EC2 instance, which sends data to CloudWatch in 1-minute periods."

What EC2 metrics are available in CloudWatch?

149

- 
- CPU credits (x)
  - CPU utilization (x)
  - Disk I/O (x)
  - Network I/O (x)
  - Status checks (x)
  - Memory utilization
  - Disk utilization

From the EC2 documentation (<http://docs.aws.amazon.com/AmazonCloudWatch/latest/DeveloperGuide/ec2-metricscollected.html>)

... see docs for full details

Also, Amazon provides a script to get memory and disk metrics here: <http://aws.amazon.com/code/8720044071969977>

---

How long are CloudWatch Logs stored?

150

- 1 week
- 1 month
- 1 year
- 10 years (x)

From the CloudWatch FAQ (<https://aws.amazon.com/cloudwatch/faqs/>):

**"Q: What is Amazon CloudWatch Logs?"**

Amazon CloudWatch Logs lets you monitor and troubleshoot your systems and applications using your existing system, application and custom log files.

With CloudWatch Logs, you can monitor your logs, in near real time, for specific phrases, values or patterns. For example, you could set an alarm on the number of errors that occur in your system logs or view graphs of latency of web requests from your application logs. You can then view the original log data to see the source of the problem. Log data can be stored and accessed for up to ten years in highly durable, low-cost storage so you don't have to worry about filling up hard drives."

---

How does CloudWatch Logs work?

151

- A CloudWatch Logs agent that you install on your EC2 instance collects log information and sends it back to CloudWatch (x)
- You invoke CloudWatch APIs to activate a CloudWatch Logs agent on your instances that collects log information and sends it back to CloudWatch
- CloudWatch polls specific log files on your EC2 instances using a user-specified IAM role

This link ([http://docs.aws.amazon.com/AmazonCloudWatch/latest/DeveloperGuide/CWL\\_GettingStarted.html](http://docs.aws.amazon.com/AmazonCloudWatch/latest/DeveloperGuide/CWL_GettingStarted.html)) goes into the full detail.

## EC2

Within a region are you charged for data transfer between EC2 instances in different AZs?

152

- 
- Yes (x)
  - No

From the EC2 FAQ (<https://aws.amazon.com/ec2/faqs/>):

**Q: If I have two instances in different availability zones, how will I be charged for regional data transfer?**

Each instance is charged for its data in and data out. Therefore, if data is transferred between these two instances, it is charged out for the first instance and in for the second instance.

---

Is it possible to configure the reverse DNS record for an EIP?

153

- 
- Yes (x)
  - No

From the EC2 FAQ (<https://aws.amazon.com/ec2/faqs/>):

**"Q: Can I configure the reverse DNS record for my Elastic IP address?**

Yes, you can configure the reverse DNS record of your Elastic IP address by filling out this form. Note that a corresponding forward DNS record pointing to that Elastic IP address must exist before we can create the reverse DNS record."

---

What are some of the features of EC2 enhanced networking?

154

- 
- High packet-per-second performance (x)
  - Lower latency (x)
  - Better scalability (x)
  - Reduced cost
  - Higher reliability
  - Better monitoring

From the EC2 FAQ (<https://aws.amazon.com/ec2/faqs/>):

**"Q: Why should I use Enhanced Networking?**

If your applications benefit from high packet-per-second performance and/or low latency networking, Enhanced Networking will provide significantly improved performance, consistence of performance and scalability."

---

Can an EC2 placement group span availability zones?

155

- 
- Yes
  - No (x)

From the EC2 documentation (<http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/placement-groups.html>):

"A placement group can't span multiple Availability Zones."

---

Do EC2 placement group names have to be unique across all AWS accounts?

156

- 
- Yes
  - No (x)

From the EC2 documentation (<http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/placement-groups.html>):

"The name you specify for a placement group a name must be unique within your AWS account."

---

Can you merge EC2 placement groups?

157

- 
- Yes
  - No (x)

From the EC2 documentation (<http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/placement-groups.html>):

"You can't merge placement groups. Instead, you must terminate the instances in one placement group, and then relaunch those instances into the other placement group."

---

Can you can move an existing instance into an EC2 placement group?

158

- 
- Yes
  - No (x)

From the EC2 documentation (<http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/placement-groups.html>):

"You can't move an existing instance into a placement group. You can create an AMI from your existing instance, and then launch a new instance from the AMI into a placement group."

---

Which instance types are memory optimized?

159

- 
- C3
  - R3 (x)
  - T2
  - M3

From the EC2 documentation (<https://aws.amazon.com/ec2/instance-types/>):

C3 — older generation compute optimized (C4 are latest)

R3 — memory optimized

T2 — burstable performance (considered general purpose)

M3 — older general purpose (M4 is latest)

---

What is the baseline performance the T2 server class (micro, small, medium)?

160



- (10%, 20%, 40%) of a single core (x)
- (20%, 40%, 60%) of a single core
- (10%, 40%, 80%) of a single core
- (10%, 15%, 30%) of a single core

From this blog (<https://aws.amazon.com/blogs/aws/low-cost-burstable-ec2-instances/>):

"Here are the specs (prices are for On-Demand Instances in US East (Northern Virginia) Region):

Name	vCPUs	Baseline Performance	RAM (GiB)	CPU Credits / Hour	Price / Hour (Linux)	Price / Month (Linux)
t2.micro	1	10%	1.0	6	\$0.013	\$9.50
t2.small	1	20%	2.0	12	\$0.026	\$19.00
t2.medium	2	40%	4.0	24	\$0.052	\$38.00

The column labeled "Baseline Performance" indicates the percentage of single core performance of the underlying physical CPU allocated to the instance. For example, a t2.small instance has access to 20% of a single core of an Intel Xeon processor running at 2.5 GHz (up to 3.3 GHz in Turbo mode). A t2.medium has access to 40% of the performance of a single core, which you (or your operating system, to be a bit more precise) can use on one or both cores as dictated by demand."

At what rate does the T2 server class accumulate CPU credits (micro, small, medium)?

161

- (3, 6, 12) credits / hour
- (6, 12, 24) credits / hour (x)
- (3, 9, 18) credits / hour
- (12, 24, 48) credits / hour

From this blog (<https://aws.amazon.com/blogs/aws/low-cost-burstable-ec2-instances/>):

>>>

#### CPU Credits

As listed in the table above, each **T2** instance receives CPU Credits at a rate that is determined by the size of the instance. A CPU Credit provides the performance of a full CPU core for one minute.

For example, a **t2.micro** instance receives credits continuously at a rate of 6 CPU Credits per hour. This capability provides baseline performance equivalent to 10% of a CPU core. If at any moment the instance does not need the credits it receives, it stores them in its CPU Credit balance for up to 24 hours. If your instance doesn't use its baseline CPU for 10 hours (let's say), the **t2.micro** instance will have accumulated enough credits to run for almost an hour with full core performance (10 hours \* 6 CPU Credits / hour = 60 CPU Credits).

....

Name	vCPUs	Baseline Performance	RAM (GiB)	CPU Credits / Hour	Price / Hour (Linux)	Price / Month (Linux)
t2.micro	1	10%	1.0	6	\$0.013	\$9.50
t2.small	1	20%	2.0	12	\$0.026	\$19.00
t2.medium	2	40%	4.0	24	\$0.052	\$38.00

<<<

What is the maximum number of accumulated CPU credits for the T2 (burstable) server class?

162

- 
- 6 hours of credits
  - 12 hours of credits
  - 24 hours of credits (x)
  - 48 hours of credits

From this blog (<https://aws.amazon.com/blogs/aws/low-cost-burstable-ec2-instances/>):

"Credits will continue to accumulate if they aren't used, until they reach the level which represents an entire day's worth of baseline accumulation:

- **t2.micro** – 144 – (6 CPU Credits / hour \* 24 hours)
- **t2.small** – 288 (12 CPU Credits / hour \* 24 hours)
- **t2.medium** – 576 (24 CPU Credits / hour \* 24 hours)

No further credits accumulate once an instance reaches this level. In general, suitable workloads for T2 instances will generally maintain a positive credit balance. If you find that you are consistently maxing out on credits, you might consider switching to a smaller instance size to reduce your costs."

---

What is a EC2 burstable instance type "cpu credit"?

163

- The performance of a full core for 10 seconds
- The performance of a full core for 30 seconds
- The performance of a full core for 60 seconds (x)
- The performance of a full core for 120 seconds

From this blog (<https://aws.amazon.com/blogs/aws/low-cost-burstable-ec2-instances/>):

"A CPU Credit provides the performance of a full CPU core for one minute.

For example, a **t2.micro** instance receives credits continuously at a rate of 6 CPU Credits per hour. This capability provides baseline performance equivalent to 10% of a CPU core. If at any moment the instance does not need the credits it receives, it stores them in its CPU Credit balance for up to 24 hours. If your instance doesn't use its baseline CPU for 10 hours (let's say), the **t2.micro** instance will have accumulated enough credits to run for almost an hour with full core performance (10 hours \* 6 CPU Credits / hour = 60 CPU Credits)."

---

Does an ENI require a security group?

164

- Yes (x)
- No

Implied by the EC2 documentation (<http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/using-eni.html>):

>>>

An elastic network interface (ENI) is a virtual network interface that you can attach to an instance in a VPC. An ENI can include the following attributes:

- a primary private IP address
- one or more secondary private IP addresses
- one Elastic IP address per private IP address
- one public IP address, which can be auto-assigned to the elastic network interface for eth0 when you launch an instance, but only when you create an elastic network interface for eth0 instead of using an existing network interface
- **one or more security groups**
- a MAC address
- a source/destination check flag
- a description

<<<

It's also required by the AWS console.

Can you detach an EC2 instance's primary ENI (the one it was assigned on creation)?

165

- Yes
- No (x)

From the EC2 documentation (<http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/using-eni.html>):

"Each instance in a VPC has a default elastic network interface (the primary network interface) that is assigned a private IP address from the IP address range of your VPC. **You cannot detach a primary network interface from an instance.** You can create and attach additional elastic network interfaces. The maximum number of elastic network interfaces that you can use varies by instance type. For more information, see Private IP Addresses Per ENI Per Instance Type."

Is it possible to use an existing ENI as the primary network interface when creating a new EC2 instance?

166

- Yes (x)
- No

From the EC2 documentation (<http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/using-eni.html#best-practices-for-configuring-network-interfaces>):

"When launching an instance from the CLI or API, you can specify the elastic network interfaces to attach to the instance for both the primary (eth0) and additional elastic network interfaces."

Does attaching multiple ENIs to an EC2 instance improve its network bandwidth?

167

- Yes
- No (x)

From the EC2 documentation (<http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/using-eni.html#best-practices-for-configuring-network-interfaces>):

"Attaching another elastic network interface to an instance is not a method to increase or double the network bandwidth to or from the dual-homed instance."

Is it possible to attach a new ENI to a running EC2 instance?

168

- Yes (x)
- No

From the EC2 documentation (<http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/using-eni.html#best-practices-for-configuring-network-interfaces>):

>>>

- You can attach an elastic network interface to an instance when it's running (hot attach), when it's stopped (warm attach), or when the instance is being launched (cold attach).
- A warm or hot attach of an additional elastic network interface may require you to manually bring up the second interface, configure the private IP address, and modify the route table accordingly. Instances running Amazon Linux or Microsoft Windows Server automatically recognize the warm or hot attach and configure themselves.

<<<

Is it possible to associate an EIP to an ENI?

169

- 
- Yes (x)
  - No

From the EC2 documentation ([http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/using-eni.html#associate\\_eip](http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/using-eni.html#associate_eip)):

"If you have an Elastic IP address, you can associate it with one of the private IP addresses for the elastic network interface. You can associate one Elastic IP address with each private IP address."

---

What is the best instance type for development environments, build servers, code repositories, low-traffic web applications, early product experiments, small databases? 170

- 
- T2 (x)
  - M3 / M4
  - C3 / C4
  - R3
  - G2
  - D2
  - I2

From the EC2 documentation (<http://aws.amazon.com/ec2/instance-types>)

---

What instance type offers the lowest price per disk throughput performance? 171

- 
- T2
  - M3 / M4
  - C3 / C4
  - R3
  - G2
  - D2 (x)
  - I2

From the EC2 documentation (<http://aws.amazon.com/ec2/instance-types>)

---

What instance type provides very fast SSD-backed instance storage optimized for very high random I/O performance, and provide high IOPS at a low cost? 172

- 
- T2
  - M3 / M4
  - C3 / C4
  - R3
  - G2
  - D2
  - I2 (x)

From the EC2 documentation (<http://aws.amazon.com/ec2/instance-types>)

---

What instance type has the lowest cost per GiB of RAM? 173

- 
- T2
  - M3 / M4
  - C3 / C4
  - R3 (x)
  - G2
  - D2
  - I2

From the EC2 documentation (<http://aws.amazon.com/ec2/instance-types>)

---

What instance type has the highest performing processors and the lowest price/compute performance?

174

- 
- T2
  - M3 / M4
  - C3 / C4 (x)
  - R3
  - G2
  - D2
  - I2

From the EC2 documentation (<http://aws.amazon.com/ec2/instance-types>)

---

What instance type has a balance of compute, memory, and network resources?

175

- 
- T2
  - M3 / M4
  - C3 / C4 (x)
  - R3
  - G2
  - D2
  - I2

From the EC2 documentation (<http://aws.amazon.com/ec2/instance-types>)

---

What is the best instance type for small and mid-size databases, data processing tasks that require additional memory, caching fleets, and for running backend servers for SAP, Microsoft SharePoint, cluster computing, and other enterprise applications.?

176

- 
- T2
  - M3 / M4 (x)
  - C3 / C4
  - R3
  - G2
  - D2
  - I2

From the EC2 documentation (<http://aws.amazon.com/ec2/instance-types>)

---

What is the best instance type for high performance front-end fleets, web-servers, batch processing, distributed analytics, high performance science and engineering applications, ad serving, MMO gaming, video-encoding, and distributed analytics?

177

- 
- T2
  - M3 / M4
  - C3 / C4 (x)
  - R3
  - G2
  - D2
  - I2

From the EC2 documentation (<http://aws.amazon.com/ec2/instance-types>)

---

What is the best instance type for high performance databases, distributed memory caches, in-memory analytics, genome assembly and analysis, larger deployments of SAP, Microsoft SharePoint, and other enterprise applications?

178

- 
- T2
  - M3 / M4
  - C3 / C4
  - R3 (x)
  - G2
  - D2
  - I2

From the EC2 documentation (<http://aws.amazon.com/ec2/instance-types>)

---

What is the best instance type for 3D application streaming, machine learning, video encoding, and other server-side graphics or GPU compute workloads?

179

- 
- T2
  - M3 / M4
  - C3 / C4
  - R3
  - G2 (x)
  - D2
  - I2

From the EC2 documentation (<http://aws.amazon.com/ec2/instance-types>)

---

What is the best instance type for massively Parallel Processing (MPP) data warehousing, MapReduce and Hadoop distributed computing, distributed file systems, network file systems, log or data-processing applications?

180

- 
- T2
  - M3 / M4
  - C3 / C4
  - R3
  - G2
  - D2 (x)
  - I2

From the EC2 documentation (<http://aws.amazon.com/ec2/instance-types>)

---

What is the best instance type for NoSQL databases like Cassandra and MongoDB, scale out transactional databases, data warehousing, Hadoop, and cluster file systems?

181

- 
- T2
  - M3 / M4
  - C3 / C4
  - R3
  - G2
  - D2
  - I2 (x)

From the EC2 documentation (<http://aws.amazon.com/ec2/instance-types>)

---

What is the maximum number of RIs per availability zone per month?

182

- 
- 5
  - 10
  - 20 (x)
  - 40
  - 100

From the EC2 FAQ (<https://aws.amazon.com/ec2/faqs/>):

**"Q: How many RIs can I purchase?"**

You can purchase up to 20 RIs per Availability Zone each month with the EC2 APIs. If you need additional RIs, complete the form [found here](#). Information about previous generation RI types can be found [here](#)."

---

Can a reserved instance be changed from one instance family (e.g., c1.xlarge) to another (e.g., m1.large)?

183

- 
- Yes
  - No (x)

From the EC2 FAQ (<https://aws.amazon.com/ec2/faqs/>):

**"Q: Can I reassign my RI from one instance family (e.g., c1.xlarge) to another (e.g., m1.large)?"**

No. An RI is associated with a specific instance family for the duration of the RI term; however, you can change from one instance type (e.g., c3.large) to another (e.g., c3.xlarge) in the same family, if it is a Linux/UNIX RI. "

---

Can a reserved instance be moved across regions?

184

- 
- Yes
  - No (x)

From the EC2 FAQ (<https://aws.amazon.com/ec2/faqs/>):

**"Q: Can I move an RI from one Region to another?"**

No. An RI is associated with a specific Region that is fixed for the duration of the RI term."

---

Is it possible to create a one-time EC2 Spot Bid?

185

- 
- Yes (x)
  - No

From the EC2 documentation ():

"*Spot instance request* (or *Spot bid*)—Provides the maximum price (bid price) that you are willing to pay per hour for a Spot instance. When your bid price exceeds the Spot price, Amazon EC2 fulfills your request. **Note that a Spot instance request is either *one-time* or *persistent*.** Amazon EC2 automatically resubmits a persistent Spot request after the Spot instance associated with the request is terminated. Your Spot instance request can optionally specify a duration for the Spot instances."

---

How much warning is given to an instance before EC2 terminates a Spot instance?

186

- 
- 30 seconds
  - 1 minute
  - 2 minutes (x)
  - 5 minutes
  - 10 minutes

From the EC2 documentation (<http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/using-spot-instances.html>):

"*Spot instance interruption*—Amazon EC2 terminates your Spot instance when the Spot price exceeds your bid price or there are no longer any unused EC2 instances. Amazon EC2 marks the Spot instance for termination and provides a Spot instance termination notice, which gives the instance a two-minute warning before it terminates."

---

Can an EC2 Security Group be configured to deny traffic?

187

- 
- Yes
  - No (x)

From the EC2 Security Group documentation ([http://docs.aws.amazon.com/AmazonVPC/latest/UserGuide/VPC\\_SecurityGroups.html](http://docs.aws.amazon.com/AmazonVPC/latest/UserGuide/VPC_SecurityGroups.html)):

"Security Group Basics

The following are the basic characteristics of security groups for your VPC:

- You can create up to 100 security groups per VPC. You can add up to 50 rules to each security group. If you need to apply more than 50 rules to an instance, you can associate up to 5 security groups with each network interface. For more information about network interfaces, see Elastic Network Interfaces (ENI).
- **You can specify allow rules, but not deny rules.**
- You can specify separate rules for inbound and outbound traffic.
- By default, no inbound traffic is allowed until you add inbound rules to the security group.
- By default, an outbound rule allows all outbound traffic. You can remove the rule and add outbound rules that allow specific outbound traffic only.
- Security groups are stateful — responses to allowed inbound traffic are allowed to flow outbound regardless of outbound rules, and vice versa.
- Instances associated with a security group can't talk to each other unless you add rules allowing it (exception: the default security group has these rules by default).
- Security groups are associated with network interfaces. After you launch an instance, you can change the security groups associated with the instance, which changes the security groups associated with the primary network interface (eth0). You can also change the security groups associated with any other network interface."

---

What are the initial rules for a VPC's *default* security group?

188



- Allow all inbound and outbound traffic
- Allow all inbound traffic
- Allow all outbound traffic (x)
- Allow all inbound traffic from other instances in the security group (x)
- Deny all outbound traffic to other instances in the security group

From the EC2 Security Group documentation ([http://docs.aws.amazon.com/AmazonVPC/latest/UserGuide/VPC\\_SecurityGroups.html](http://docs.aws.amazon.com/AmazonVPC/latest/UserGuide/VPC_SecurityGroups.html)):

« 177 »

### Default Security Groups

Your VPC automatically comes with a default security group. Each EC2 instance that you launch in your VPC is automatically associated with the default security group if you don't specify a different security group when you launch the instance.

The following table describes the default rules for a default security group.

Inbound			
Source	Protocol	Port Range	Comments
The security group ID (sg-xxxxxxx)	All	All	Allow inbound traffic from instances assigned to the same security group
Outbound			
Destination	Protocol	Port Range	Comments
0.0.0.0/0	All	All	Allow all outbound traffic

You can change the rules for the default security group.

You can't delete a default security group.

« 177 »

Is it possible to delete a VPCs default security group?

189

- Yes
- No (x)

From the EC2 Security Group documentation ([http://docs.aws.amazon.com/AmazonVPC/latest/UserGuide/VPC\\_SecurityGroups.html](http://docs.aws.amazon.com/AmazonVPC/latest/UserGuide/VPC_SecurityGroups.html)):

« 177 »

### Default Security Groups

Your VPC automatically comes with a default security group. Each EC2 instance that you launch in your VPC is automatically associated with the default security group if you don't specify a different security group when you launch the instance.

The following table describes the default rules for a default security group.

Inbound			
Source	Protocol	Port Range	Comments
The security group ID (sg-xxxxxxx)	All	All	Allow inbound traffic from instances assigned to the same security group
Outbound			
Destination	Protocol	Port Range	Comments
0.0.0.0/0	All	All	Allow all outbound traffic

You can change the rules for the default security group.

**You can't delete a default security group.**

« 177 »

Is it possible to modify a VPCs default security group?

190

- Yes (x)
- No

From the EC2 Security Group documentation ([http://docs.aws.amazon.com/AmazonVPC/latest/UserGuide/VPC\\_SecurityGroups.html](http://docs.aws.amazon.com/AmazonVPC/latest/UserGuide/VPC_SecurityGroups.html)):

447771

### Default Security Groups

Your VPC automatically comes with a default security group. Each EC2 instance that you launch in your VPC is automatically associated with the default security group if you don't specify a different security group when you launch the instance.

The following table describes the default rules for a default security group.

Inbound			
Source	Protocol	Port Range	Comments
The security group ID (sg-xxxxxx)	All	All	Allow inbound traffic from instances assigned to the same security group
Outbound			
Destination	Protocol	Port Range	Comments
0.0.0.0/0	All	All	Allow all outbound traffic

**You can change the rules for the default security group.**

You can't delete a default security group.

447771

Can EC2 instances within the same *non-default* security group talk to each other by default?

191

- No (x)
- Yes

From the EC2 Security Group documentation ([http://docs.aws.amazon.com/AmazonVPC/latest/UserGuide/VPC\\_SecurityGroups.html](http://docs.aws.amazon.com/AmazonVPC/latest/UserGuide/VPC_SecurityGroups.html)):

### "Security Group Basics

The following are the basic characteristics of security groups for your VPC:

- You can create up to 100 security groups per VPC. You can add up to 50 rules to each security group. If you need to apply more than 50 rules to an instance, you can associate up to 5 security groups with each network interface. For more information about network interfaces, see Elastic Network Interfaces (ENI).
- You can specify allow rules, but not deny rules.
- You can specify separate rules for inbound and outbound traffic.
- By default, no inbound traffic is allowed until you add inbound rules to the security group.
- By default, an outbound rule allows all outbound traffic. You can remove the rule and add outbound rules that allow specific outbound traffic only.
- Security groups are stateful — responses to allowed inbound traffic are allowed to flow outbound regardless of outbound rules, and vice versa.
- **Instances associated with a security group can't talk to each other unless you add rules allowing it (exception: the default security group has these rules by default).**
- Security groups are associated with network interfaces. After you launch an instance, you can change the security groups associated with the instance, which changes the security groups associated with the primary network interface (eth0). You can also change the security groups associated with any other network interface."

Can an EC2 Security Group specify different rules for inbound and outbound traffic?

192

- 
- Yes (x)
  - No

From the EC2 Security Group documentation ([http://docs.aws.amazon.com/AmazonVPC/latest/UserGuide/VPC\\_SecurityGroups.html](http://docs.aws.amazon.com/AmazonVPC/latest/UserGuide/VPC_SecurityGroups.html)):

#### "Security Group Basics

The following are the basic characteristics of security groups for your VPC:

- You can create up to 100 security groups per VPC. You can add up to 50 rules to each security group. If you need to apply more than 50 rules to an instance, you can associate up to 5 security groups with each network interface. For more information about network interfaces, see Elastic Network Interfaces (ENI).
- You can specify allow rules, but not deny rules.
- **You can specify separate rules for inbound and outbound traffic.**
- By default, no inbound traffic is allowed until you add inbound rules to the security group.
- By default, an outbound rule allows all outbound traffic. You can remove the rule and add outbound rules that allow specific outbound traffic only.
- Security groups are stateful — responses to allowed inbound traffic are allowed to flow outbound regardless of outbound rules, and vice versa.
- Instances associated with a security group can't talk to each other unless you add rules allowing it (exception: the default security group has these rules by default).
- Security groups are associated with network interfaces. After you launch an instance, you can change the security groups associated with the instance, which changes the security groups associated with the primary network interface (eth0). You can also change the security groups associated with any other network interface."

---

What does "EBS-Backed" mean in reference to an EC2 instance?

193

- 
- The root device is stored on an EBS volume (x)
  - An additional EBS drive is added to the instance at startup
  - Automated backups are retained in EBS

From the EC2 User Guide (<http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/ComponentsAMIs.html>):

"All AMIs are categorized as either *backed by Amazon EBS* or *backed by instance store*. The former means that the root device for an instance launched from the AMI is an Amazon EBS volume created from an Amazon EBS snapshot. The latter means that the root device for an instance launched from the AMI is an instance store volume created from a template stored in Amazon S3."

---

Is it possible to move ENIs between subnets?

194

- 
- Yes (x)
  - No

From the EC2 User Guide (<http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/using-eni.html>):

"You can attach an elastic network interface in one subnet to an instance in another subnet in the same VPC; however, both the elastic network interface and the instance must reside in the same Availability Zone."

---

Is it possible to move ENIs between Availability Zones?

195

- 
- Yes
  - No (x)

From the EC2 User Guide (<http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/using-eni.html>):

"You can attach an elastic network interface in one subnet to an instance in another subnet in the same VPC; however, both the elastic network interface and the instance must reside in the same Availability Zone."

- Yes, but only if the subnets are in the same Availability Zone (x)
- Yes in all cases
- Yes, but only if the subnets have don't have overlapping CIDRs
- No

I just made up the bit about overlapping CIDRs as a distractor.

From the EC2 User Guide (<http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/using-eni.html>):

"You can attach an elastic network interface in one subnet to an instance in another subnet in the same VPC; however, both the elastic network interface and the instance must reside in the same Availability Zone."

- C3/C4 (x)
- D2 (x)
- I2 (x)
- M4 (x)
- R3 (x)
- T2
- G2

From the EC2 User Guide ([http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/enhanced-networking.html#enhanced\\_networking\\_instance\\_types](http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/enhanced-networking.html#enhanced_networking_instance_types)).

- Compute optimized: C3, C4, CC2 (x)
- General purpose: M3, M4, T2 (x)
- GPU: CG1, G2 (x)
- Memory optimized: CR1, R3 (x)
- Storage optimized: D2, H1, HS1, I2 (x)

From the EC2 User Guide ([http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/network\\_mtu.html](http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/network_mtu.html)).

- Low-latency 10GB networking between instances (x)
- Simplified management
- Consolidated CloudWatch metrics
- Ability to terminate all instances as a group

From the EC2 User Guide ():

"A *placement group* is a logical grouping of instances within a single Availability Zone. Using placement groups enables applications to participate in a low-latency, 10 Gbps network. Placement groups are recommended for applications that benefit from low network latency, high network throughput, or both. To provide the lowest latency, and the highest packet-per-second network performance for your placement group, choose an instance type that supports enhanced networking. For more information, see Enhanced Networking."

Do reserved instance capacity reservations apply to instances launched within an EC2 Placement Group?

---

- Yes
- No (x)

From the EC2 User Guide (<http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/placement-groups.html>):

"Reserved instances are applied to On-Demand instances in an Availability Zone. The capacity reservation does not cover instances within a placement group."

Can an EC2 placement group span peered VPCs?

---

201

- Yes (x)
- No

From the EC2 User Guide (<http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/placement-groups.html>):

"A placement group can span peered VPCs; however, you will not get full-bisection bandwidth between instances in peered VPCs. "

What are some recommendations when launching new EC2 placement groups?

---

202

- Launch all the instances you need because you may get an insufficient capacity error if you try to add more instances later (x)
- Use the smallest instance type available so as to maximize your chances of getting the capacity you need
- Use the same instance type for all instances (x)
- Choose instance types that support Enhanced Networking (x)
- Minimize the number of instance groups in any single VPC

From the EC2 User Guide (<http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/placement-groups.html>):

"A *placement group* is a logical grouping of instances within a single Availability Zone. Using placement groups enables applications to participate in a low-latency, 10 Gbps network. Placement groups are recommended for applications that benefit from low network latency, high network throughput, or both. **To provide the lowest latency, and the highest packet-per-second network performance for your placement group, choose an instance type that supports enhanced networking.** For more information, see Enhanced Networking.

First, you create a placement group and then you launch multiple instances into the placement group. **We recommend that you launch the number of instances that you need in the placement group in a single launch request and that you use the same instance type for all instances in the placement group.** If you try to add more instances to the placement group later, or if you try to launch more than one instance type in the placement group, you increase your chances of getting an insufficient capacity error.

If you stop an instance in a placement group and then start it again, it still runs in the placement group. However, the start fails if there isn't enough capacity for the instance.

If you receive a capacity error when launching an instance in a placement group, stop and restart the instances in the placement group, and then try the launch again."

Does EC2 enhanced networking (SR-IOV) require HVM virtualization?

---

203

- Yes (x)
- No

From the EC2 User Guide (<http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/enhanced-networking.html>):

"Enhanced networking is supported only for HVM instances. Enabling enhanced networking with a PV instance can make it unreachable. Setting this attribute without the proper module or module version can also make your instance unreachable."

What are some valid methods to scale up a NAT?

---

204

- 
- Scale up by increasing the instance size (x)
  - Choose an instance with Enhanced Networking (x)
  - Scale out by adding additional NATs in other subnets and splitting traffic across those subnets (x)
  - Autoscale based on the NATs CloudWatch metrics

From Ryan's lectures...

I made up the bit about CloudWatch.

---

What are the properties of an EC2 Reserved Instance?

205

- Instance type (x)
- Availability zone (x)
- Platform description / OS (x)
- Term (x)
- Payment option (x)
- Expected utilization

From the Reserved Instances marketing page (<https://aws.amazon.com/ec2/purchasing-options/reserved-instances/>):

"The Reserved Instance hourly rate is applied to your Amazon EC2 instance usage when the attributes of your instance usage match the attributes of your Reserved Instances.

Amazon EC2 Reserved Instance Attributes

- **Instance Type:** Instance types comprise varying combinations of CPU, memory, storage, and networking capacity. For example, m3.xlarge.
- **Availability Zone:** Amazon EC2 provides you the ability to purchase Reserved Instances within AWS Availability Zones. For example, us-east-1a
- **Platform Description:** Reserved Instances can be purchased for Amazon EC2 running Linux/UNIX, SUSE Linux, Red Hat Enterprise Linux, Microsoft Windows Server, and Microsoft SQL Server platforms.
- **Tenancy:** Each instance that you launch has a tenancy attribute. Generally, instances run with a default tenancy (running on multi-tenant hardware) unless you've explicitly specified to run your instance with a tenancy of dedicated (single tenant hardware).

Amazon EC2 Reserved Instance Payment Attributes

- **Term:** AWS offers Reserved Instances for 1 or 3 year terms. Reserved Instance Marketplace Sellers also offer Reserved Instances often with shorter terms.
- **Payment Option:** You can choose between 3 payment options: All, Partial and No Upfront. If you choose the Partial or No Upfront payment option, the remaining balance will be due in monthly increments over the term."

---

What are the Reserved Instance payment options?

206

- Full Upfront (x)
- Partial (x)
- No Upfront (x)
- Quarterly
- Yearly

From the EC2 Pricing Page (<http://aws.amazon.com/ec2/pricing/#reserved-instances>):

"You can choose between three payment options when you purchase a Reserved Instance. With the **All Upfront** option, you pay for the entire Reserved Instance with one upfront payment. This option provides you with the largest discount compared to On-Demand Instance pricing. With the **Partial Upfront** option, you make a low upfront payment and are then charged a discounted hourly rate for the instance for the duration of the Reserved Instance term. The **No Upfront** option does not require any upfront payment and provides a discounted hourly rate for the duration of the term."

---

What are the possible terms for a Reserved Instance reservation?

207

- 1 year (x)
- 2 years
- 3 years (x)
- 4 years
- 5 years

From the Reserved Instance Getting Started page (<https://aws.amazon.com/ec2/purchasing-options/reserved-instances/getting-started/>):

### "Determine the term length

What percentage of this group do I expect will be running 1 year from now? 3 years from now? Determine the number of instances you want to run and the term length (1 or 3 years)."

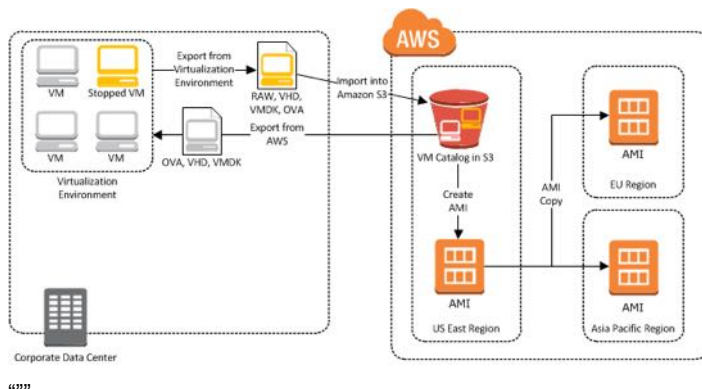
Is it possible to import images from other virtualization environments (Citrix Xen, Microsoft Hyper-V, and VMware vSphere) into EC2? 208

- Yes (x)
- No

From the EC2 documentation (<http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/UsingImportImage.html>):

You can import a virtual machine (VM) from your virtualization environment such as Citrix Xen, Microsoft Hyper-V, or VMware vSphere, and import it as an AMI in Amazon EC. [...]

The following diagram shows the process of exporting a VM from your on-premises virtualization environment to AWS.



Is it possible to import instances from other virtualization environments (Citrix Xen, Microsoft Hyper-V, and VMware vSphere) into EC2? 209

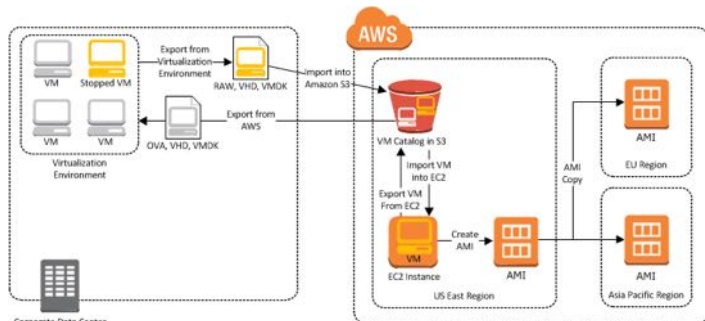
- Yes (x)
- No

From the EC2 documentation (<http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/UsingVirtualMachinesinAmazonEC2.html>):

44778

There are two ways you can launch an instance in Amazon EC2. You can launch an instance from an Amazon Machine Image (AMI), or, you can launch an instance from a virtual machine (VM) that you imported from a virtualization environment such as Citrix Xen, Microsoft Hyper-V, or VMware vSphere. [...]

The following diagram shows the process of exporting a VM from your on-premises virtualization environment to AWS.



44779

What can be modified on an existing Reserved Instance?

210

- Availability Zone (x)
- Switching between EC2-VPC and EC2-Classic (x)
- Changing the instance type within the same instance family (x)
- Subnet association
- Utilization options

From the Reserved Instance documentation (<http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/ri-modifying.html>):

"When your computing needs change, you can modify your Reserved Instances and continue to benefit from your capacity reservation. Modification does not change the remaining term of your Reserved Instances; their end dates remain the same. There is no fee, and you do not receive any new bills or invoices. Modification is separate from purchasing and does not affect how you use, purchase, or sell Reserved Instances. You can modify your whole reservation, or just a subset, in one or more of the following ways:

- Switch Availability Zones within the same region
- Change between EC2-VPC and EC2-Classic
- Change the instance type within the same instance family"

Is it possible to create an AMI from an EBS snapshot?

211



- 
- Yes (x)
  - No

From the EC2 documentation (<http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/creating-an-ami-ebs.html>):

❧ 179 ❧

#### To create an AMI from a snapshot using the console

1. Open the Amazon EC2 console at <https://console.aws.amazon.com/ec2/>.
2. In the navigation pane, under Elastic Block Store, choose Snapshots.
3. Choose the snapshot, and then choose Create Image from the Actions list.
4. In the Create Image from EBS Snapshot dialog box, complete the fields to create your AMI, then choose Create. If you're re-creating a parent instance, then choose the same options as the parent instance.

❧ 179 ❧

---

How do you get an EC2 instance's user data?

212

- `ec2_metadata` command (x)
- Via an API call
- It's pre-installed on the instance during provisioning
- `curl http://169.254.169.254/latest/meta-data` (x)

From the EC2 User Guide (<http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/ec2-instance-metadata.html#instancedata-data-retrieval>):

"To view all categories of instance metadata from within a running instance, use the following URI:  
`http://169.254.169.254/latest/meta-data/`"

I don't see any mention of `ec2_metadata` in the documentation, but I know it works because I use it all the time.

---

How do you get an EC2 instance's user data?

213

- It's part of the instance's metadata (x)
- Via an API call
- It's embedded on the instance during provisioning

From the EC2 User Guide (<http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/ec2-instance-metadata.html#instancedata-data-categories>):

"*Instance metadata* is data about your instance that you can use to configure or manage the running instance. [...] You can also access the *user data* that you supplied when launching your instance. For example, you can specify parameters for configuring your instance, or attach a simple script. You can also use this data to build more generic AMIs that can be modified by configuration files supplied at launch time. For example, if you run web servers for various small businesses, they can all use the same AMI and retrieve their content from the Amazon S3 bucket you specify in the user data at launch. To add a new customer at any time, simply create a bucket for the customer, add their content, and launch your AMI. If you launch more than one instance at the same time, the user data is available to all instances in that reservation."

## S3

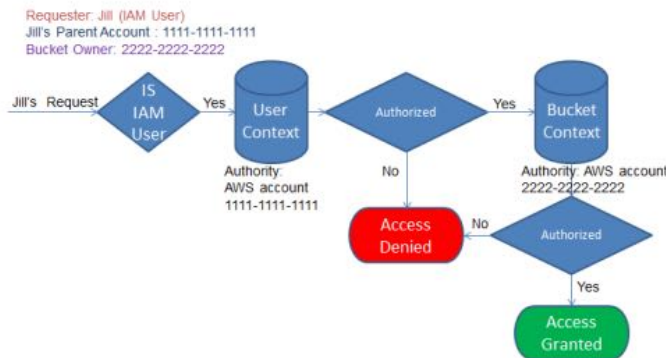
When an IAM user performs an operation on an S3 bucket that is **not** owned by its parent account, which permission contexts are evaluated? 214

- User context (x)
- Bucket context (x)
- Object context

From the S3 documentation (<http://docs.aws.amazon.com/AmazonS3/latest/dev/access-control-auth-workflow-bucket-operation.html>):

"Example 4: Bucket Operation Requested by an IAM User Whose Parent AWS Account Is Not the Bucket Owner

In this example, the request is sent by Jill, an IAM user whose parent AWS account is 1111-1111-1111, but the bucket is owned by another AWS account, 2222-2222-2222.



Jill will need permissions from both the parent AWS account and the bucket owner. Amazon S3 evaluates the context as follows:

1. Because the request is from an IAM user, Amazon S3 evaluates the user context by reviewing the policies authored by the account to verify that Jill has the necessary permissions. If Jill has permission, then Amazon S3 moves on to evaluate the bucket context; if not, it denies the request.
2. In the bucket context, Amazon S3 verifies that bucket owner 2222-2222-2222 has granted Jill (or her parent AWS account) permission to perform the requested operation. If she has that permission, Amazon S3 grants the request and performs the operation; otherwise, Amazon S3 denies the request."

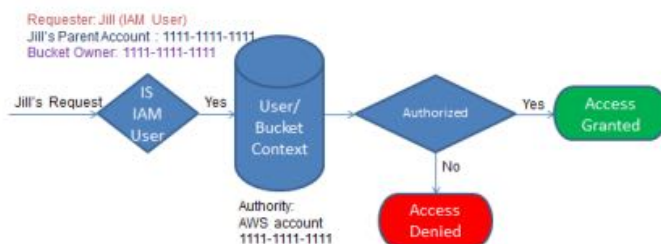
When an IAM user performs an operation on an S3 bucket owned by its parent account, which permission contexts are evaluated? 215

- User context (x)
- Bucket context
- Object context

From the S3 documentation (<http://docs.aws.amazon.com/AmazonS3/latest/dev/access-control-auth-workflow-bucket-operation.html>):

"Example 3: Bucket Operation Requested by an IAM User Whose Parent AWS Account Is Also the Bucket Owner

In the example, the request is sent by Jill, an IAM user in AWS account 1111-1111-1111, which also owns the bucket.



Amazon S3 performs the following context evaluation:

1. Because the request is from an IAM user, in the user context, Amazon S3 evaluates all policies that belong to the parent AWS account to determine if Jill has permission to perform the operation.  
 In this example, parent AWS account 1111-1111-1111, to which the user belongs, is also the bucket owner. As a result, in addition to the user policy, Amazon S3 also evaluates the bucket policy and bucket ACL in the same context, because they belong to the same account.
2. Because Amazon S3 evaluated the bucket policy and bucket ACL as part of the user context, it does not evaluate the bucket context."

When an account owner (not an IAM user) performs an operation on a bucket that it owns, which permission contexts are evaluated?

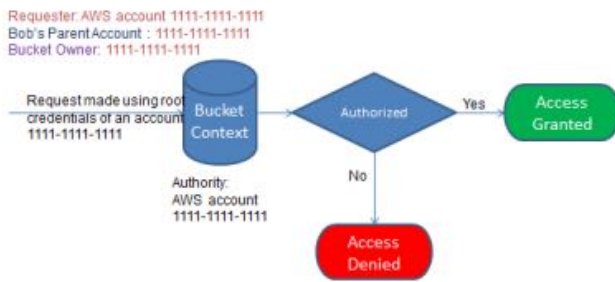
216

- User context
- Bucket context (x)
- Object context

From the S3 documentation (<http://docs.aws.amazon.com/AmazonS3/latest/dev/access-control-auth-workflow-bucket-operation.html>):

“Example 1: Bucket Operation Requested by Bucket Owner

In this example, the bucket owner sends a request for a bucket operation using the root credentials of the AWS account.



Amazon S3 performs the context evaluation as follows:

1. Because the request is made by using root credentials of an AWS account, the user context is not evaluated .
2. In the bucket context, Amazon S3 reviews the bucket policy to determine if the requester has permission to perform the operation. Amazon S3 authorizes the request."

If an account owner A grants another account owner access to an S3 resource, can account owner B then grant another account owner C access to the same resource? 217

- Yes
- No (x)

From the S3 documentation (<http://docs.aws.amazon.com/AmazonS3/latest/dev/access-policy-alternatives-guidelines.html>):

"If an AWS account owns a resource, it can grant those permissions to another AWS account. That account can then delegate those permissions, or a subset of them, to users in the account. This is referred to as permission delegation. But an account that receives permissions from another account cannot delegate permission cross-account to another AWS account."

If an account owner A grants another account owner access to an S3 resource, can account owner B then delegate that access to its own IAM users? 218

- Yes (x)
- No

From the S3 documentation (<http://docs.aws.amazon.com/AmazonS3/latest/dev/access-policy-alternatives-guidelines.html>):

"If an AWS account owns a resource, it can grant those permissions to another AWS account. That account can then delegate those permissions, or a subset of them, to users in the account. This is referred to as permission delegation. But an account that receives permissions from another account cannot delegate permission cross-account to another AWS account."

When a user in one accounts adds an object to an S3 bucket owned by a user another account, by default the bucket owner has permissions to... 219

- Read the object
- Delete the object (x)
- Grant access to the object
- Deny access to the object (x)

From the S3 documentation (<http://docs.aws.amazon.com/AmazonS3/latest/dev/access-policy-alternatives-guidelines.html>):

"A bucket owner cannot grant permissions on objects it does not own. For example, a bucket policy granting object permissions applies only to objects owned by the bucket owner. However, the bucket owner, who pays the bills, can write a bucket policy to deny access to any objects in the bucket, regardless of who owns it. The bucket owner can also delete any objects in the bucket."

When an account owner performs an operation on an S3 bucket owned by another account, which permission contexts are evaluated?

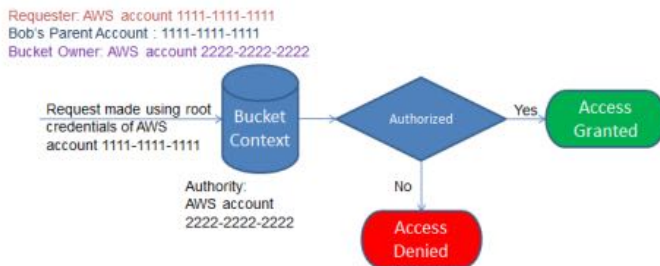
220

- User context
- Bucket context (x)
- Object context

From the S3 documentation (<http://docs.aws.amazon.com/AmazonS3/latest/dev/access-control-auth-workflow-bucket-operation.html>):

"Example 2: Bucket Operation Requested by an AWS Account That Is Not the Bucket Owner

In this example, a request is made using root credentials of AWS account 1111-1111-1111 for a bucket operation owned by AWS account 2222-2222-2222. No IAM users are involved in this request.



In this case, Amazon S3 evaluates the context as follows:

1. Because the request is made using root credentials of an AWS account, the user context is not evaluated.
2. In the bucket context, Amazon S3 examines the bucket policy. If the bucket owner (AWS account 2222-2222-2222) has not authorized AWS account 1111-1111-1111 to perform the requested operation, Amazon S3 denies the request. Otherwise, Amazon S3 grants the request and performs the operation."

When an account owner or IAM user performs an operation on an S3 object, which permission contexts are evaluated?

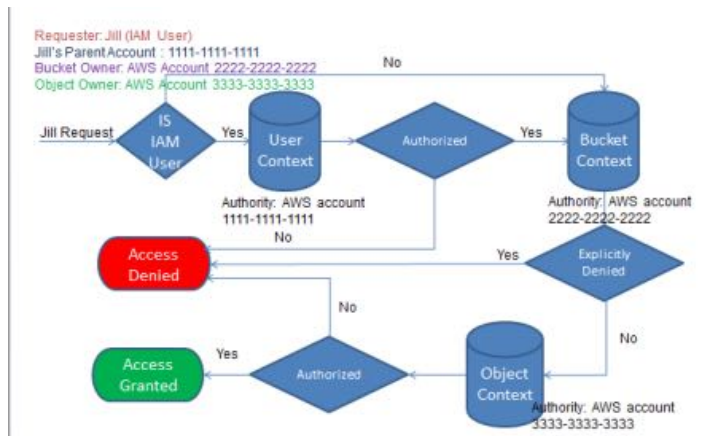
221

- User context (x)
- Bucket context (x)
- Object context (x)

From the S3 documentation (<http://docs.aws.amazon.com/AmazonS3/latest/dev/access-control-auth-workflow-object-operation.html>):

#### "Example 1: Object Operation Request

In this example, IAM user Jill, whose parent AWS account is 1111-1111-1111, sends an object operation request (for example, Get object) for an object owned by AWS account 3333-3333-3333 in a bucket owned by AWS account 2222-2222-2222.



Jill will need permission from the parent AWS account, the bucket owner, and the object owner. Amazon S3 evaluates the context as follows:

1. Because the request is from an IAM user, Amazon S3 evaluates the user context to verify that the parent AWS account 1111-1111-1111 has given Jill permission to perform the requested operation. If she has that permission, Amazon S3 evaluates the bucket context. Otherwise, Amazon S3 denies the request.
2. In the bucket context, the bucket owner, AWS account 2222-2222-2222, is the context authority. Amazon S3 evaluates the bucket policy to determine if the bucket owner has explicitly denied Jill access to the object.
3. In the object context, the context authority is AWS account 3333-3333-3333, the object owner. Amazon S3 evaluates the object ACL to determine if Jill has permission to access the object. If she does, Amazon S3 authorizes the request."

What is the default access for S3 resources (e.g., bucket, objects, lifecycle configuration, website configuration, etc)?

222

- Public
- Private (x)

From the S3 documentation (<http://docs.aws.amazon.com/AmazonS3/latest/dev/s3-access-control.html>):

"By default, all Amazon S3 resources—buckets, objects, and related subresources (for example, lifecycle configuration and website configuration)—are private: only the resource owner, an AWS account that created it, can access the resource. The resource owner can optionally grant access permissions to others by writing an access policy."

Note that this also covers the case where an account owner creates a bucket that another account can add objects to. Unless that other account grants permissions, the owner of the bucket can't delete those objects.

When a user in one account attempts to access an object in an S3 bucket owned by another account, what happens if that bucket does **not** have a bucket policy defined?

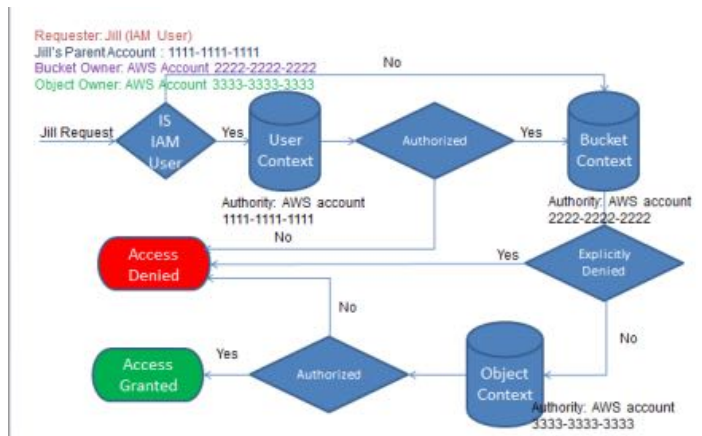
223

- The request is allowed by default
- The request is denied by default
- The request is allowed if the object ACL allows it (x)

From the S3 documentation (<http://docs.aws.amazon.com/AmazonS3/latest/dev/access-control-auth-workflow-object-operation.html>):

#### "Example 1: Object Operation Request

In this example, IAM user Jill, whose parent AWS account is 1111-1111-1111, sends an object operation request (for example, Get object) for an object owned by AWS account 3333-3333-3333 in a bucket owned by AWS account 2222-2222-2222.



Jill will need permission from the parent AWS account, the bucket owner, and the object owner. Amazon S3 evaluates the context as follows:

1. Because the request is from an IAM user, Amazon S3 evaluates the user context to verify that the parent AWS account 1111-1111-1111 has given Jill permission to perform the requested operation. If she has that permission, Amazon S3 evaluates the bucket context. Otherwise, Amazon S3 denies the request.
2. In the bucket context, the bucket owner, AWS account 2222-2222-2222, is the context authority. **Amazon S3 evaluates the bucket policy to determine if the bucket owner has explicitly denied Jill access to the object.**
3. In the object context, the context authority is AWS account 3333-3333-3333, the object owner. Amazon S3 evaluates the object ACL to determine if Jill has permission to access the object. If she does, Amazon S3 authorizes the request."

Also, earlier in that link:

"[...] If the AWS account that owns the object in the request is not same as the bucket owner, in the bucket context Amazon S3 checks the policies if the bucket owner has explicitly denied access to the object. **If there is an explicit deny set on the object, Amazon S3 does not authorize the request.**"

Can a bucket policy control access to a specific S3 object?

224

- Yes (x)
- No

From the S3 documentation (<http://docs.aws.amazon.com/AmazonS3/latest/dev/access-policy-alternatives-guidelines.html>):

"In addition to an object ACL, there are other ways an object owner can manage object permissions. For example:

- If the AWS account that owns the object also owns the bucket, then it can write a bucket policy to manage the object permissions.
- If the AWS account that owns the object wants to grant permission to a user in its account, it can use a user policy."

The referenced link goes into this in a lot more detail.

Can a user modify a policy on a bucket in another account?

225

- 
- Yes (x)
  - No

From the S3 documentation (<http://docs.aws.amazon.com/AmazonS3/latest/dev/access-policy-alternatives-guidelines.html>):

- 

"An AWS account that owns a bucket can grant another AWS account permission to manage access policy. It allows that account to change anything in the policy. To better manage permissions, you may choose not to give such a broad permission, and instead grant only the READ-ACP and WRITE-ACP permissions on a subset of objects. This limits the account to manage permissions only on specific objects by updating individual object ACLs."

also

"Permission Delegation

If an AWS account owns a resource, it can grant those permissions to another AWS account. That account can then delegate those permissions, or a subset of them, to users in the account. This is referred to as permission delegation. But an account that receives permissions from another account cannot delegate permission cross-account to another AWS account."

The referenced link goes into this in a lot more detail.

Are S3 bucket policies applied to API calls made with the account owner's credentials?

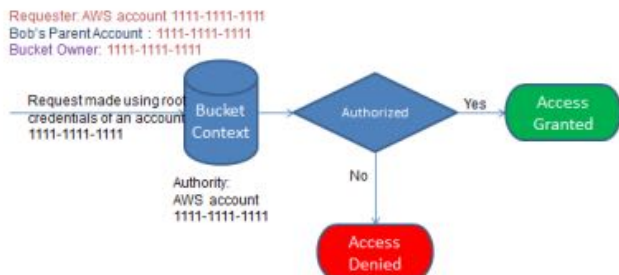
226

- Yes (x)
- No

From the S3 documentation (<http://docs.aws.amazon.com/AmazonS3/latest/dev/access-control-auth-workflow-bucket-operation.html>):

#### "Example 1: Bucket Operation Requested by Bucket Owner

In this example, the bucket owner sends a request for a bucket operation using the root credentials of the AWS account.



Amazon S3 performs the context evaluation as follows:

1. Because the request is made by using root credentials of an AWS account, the user context is not evaluated .
2. In the bucket context, Amazon S3 reviews the bucket policy to determine if the requester has permission to perform the operation. Amazon S3 authorizes the request."

[DLS] I tested this by adding a bucket policy to remove all delete permissions to an S3 object for all principals:

```

{
  "Statement": [
    {
      "Effect": "Deny",
      "Principal": "*",
      "Action": "s3:DeleteObject",
      "Resource": "arn:aws:s3:::cco-foo/*"
    }
  ]
}
  
```

While this allowed me to create objects in the "cco-ffo" bucket as the account owner, I couldn't delete the objects I created. I had to remove the bucket policy."

Are S3 object encrypted at an object or bucket level?

227

- Object (x)
- Bucket

As per the S3 documentation (<http://docs.aws.amazon.com/AmazonS3/latest/dev/UsingServerSideEncryption.html>):

"Server-side encryption is about protecting data at rest. Server-side encryption with Amazon S3-managed encryption keys (SSE-S3) employs strong multi-factor encryption. **Amazon S3 encrypts each object with a unique key.** As an additional safeguard, it encrypts the key itself with a master key that it regularly rotates. Amazon S3 server-side encryption uses one of the strongest block ciphers available, 256-bit Advanced Encryption Standard (AES-256), to encrypt your data."

Can you force all object in an S3 bucket to use server-side encryption?

228



- 
- Yes (x)
  - No

As per the S3 documentation (<http://docs.aws.amazon.com/AmazonS3/latest/dev/UsingServerSideEncryption.html>):

"Amazon S3 supports bucket policies that you can use if you require server-side encryption for all objects that are stored in your bucket. For example, the following bucket policy denies upload object (s3:PutObject) permission to everyone if the request does not include the x-amz-server-side-encryption header requesting server-side encryption.

```
{
  "Version": "2012-10-17",
  "Id": "PutObjPolicy",
  "Statement": [{
    "Sid": "DenyUnEncryptedObjectUploads",
    "Effect": "Deny",
    "Principal": "*",
    "Action": "s3:PutObject",
    "Resource": "arn:aws:s3:::YourBucket/*",
    "Condition": {
      "StringNotEquals": {
        "s3:x-amz-server-side-encryption": "AES256"
      }
    }
  }]
}
```

---

Does S3 server-side encryption encrypt object metadata?

229

- 
- Yes
  - No (x)

As per the S3 documentation (<http://docs.aws.amazon.com/AmazonS3/latest/dev/UsingServerSideEncryption.html>):

"Server-side encryption encrypts only the object data. Any object metadata is not encrypted."

---

What happens to an S3 object when its expiration policy is triggered if versioning **isn't** enabled?

230

- 
- The object is deleted if versioning isn't enabled(x)
  - The object is moved into another user-specified bucket
  - A notification is posted to an SNS topic

From the S3 documentation ():

"[The Expiration lifecycle action] specifies a period in an object's lifetime when Amazon S3 should take the appropriate expiration action. The action Amazon S3 takes depends on whether the bucket is versioning-enabled.

- If versioning has never been enabled on the bucket, Amazon S3 deletes the only copy of the object permanently.
- Otherwise, if your bucket is versioning-enabled (or versioning is suspended), the action applies only to the current version of the object. A versioning-enabled bucket can have many versions of the same object, one current version, and zero or more noncurrent versions.

Instead of deleting the current version, Amazon S3 makes it a noncurrent version by adding a delete marker as the new current version.

- The object is deleted
- The current version is physically deleted
- The current version is logically marked as deleted (x)
- Non-current versions are still available (x)

From the S3 documentation ():

"[The Expiration lifecycle action] specifies a period in an object's lifetime when Amazon S3 should take the appropriate expiration action. The action Amazon S3 takes depends on whether the bucket is versioning-enabled.

- If versioning has never been enabled on the bucket, Amazon S3 deletes the only copy of the object permanently.
- Otherwise, if your bucket is versioning-enabled (or versioning is suspended), the action applies only to the current version of the object. A versioning-enabled bucket can have many versions of the same object, one current version, and zero or more noncurrent versions.

Instead of deleting the current version, Amazon S3 makes it a noncurrent version by adding a delete marker as the new current version.

- Yes (x)
- No

From the S3 documentation (<http://docs.aws.amazon.com/AmazonS3/latest/dev/DNSConsiderations.html>):

"One of the design requirements of Amazon S3 is extremely high availability. One of the ways we meet this requirement is by updating the IP addresses associated with the Amazon S3 endpoint in DNS as needed. These changes are automatically reflected in short-lived clients, but not in some long-lived clients. Long-lived clients will need to take special action to re-resolve the Amazon S3 endpoint periodically to benefit from these changes."

- By the string value of the key name (x)
- By the hash of the key name
- By the hash of the key name + metadata

From the S3 documentation (<http://docs.aws.amazon.com/AmazonS3/latest/dev/request-rate-perf-considerations.html>):

"Amazon S3 maintains an index of object key names in each AWS region. Object keys are stored lexicographically across multiple partitions in the index. That is, Amazon S3 stores key names in alphabetical order. The key name dictates which partition the key is stored in. Using a sequential prefix, such as timestamp or an alphabetical sequence, increases the likelihood that Amazon S3 will target a specific partition for a large number of your keys, overwhelming the I/O capacity of the partition. If you introduce some randomness in your key name prefixes, the key names, and therefore the I/O load, will be distributed across more than one partition."

- 
- A large rate of PUTs (>100/s) with alphabetically sequential key names (x)
  - Prefixing object keys with a hash of itself
  - Reversing the key name prior to object insertion
  - Using CloudFront for dynamic objects

See <http://docs.aws.amazon.com/AmazonS3/latest/dev/request-rate-perf-considerations.html>.

The link goes into details, but the big culprit here is sequential keys. Prefixing keys with a hash is actually a best practice to avoid this, as is reversing the key name. Using CloudFront reduces the load on S3 and attacks the issue directly. So the first option is the only valid answer.

---

What is the ordering of the response from the S3 GET Bucket (List Objects) operation?

235

- Unordered / random
- Alphabetical by key name (x)

From the S3 documentation (<http://docs.aws.amazon.com/AmazonS3/latest/dev/request-rate-perf-considerations.html>):

"Amazon S3 provides a GET Bucket (List Objects) operation, which returns an alphabetical list of key names."

---

Is it possible to permanently delete a specific version of an S3 versioned object?

236

- Yes (x)
- No

Yes, by including the version ID in the DELETE request.

From the S3 documentation (<http://docs.aws.amazon.com/AmazonS3/latest/dev/DeletingObjects.html#DeletingObjectsfromaVersion-EnabledBucket>):

"If your bucket is version-enabled, then multiple versions of the same object can exist in the bucket. When working with version-enabled buckets, the delete API enables the following options:

- Specify a non-versioned delete request—That is, you specify only the object's key, and not the version ID. In this case, Amazon S3 creates a delete marker and returns its version ID in the response. This makes your object disappear from the bucket. For information about object versioning and the delete marker concept, see Object Versioning.
- Specify a versioned delete request—That is, you specify both the key and also a version ID. In this case the following two outcomes are possible:
  - If the version ID maps to a specific object version, then Amazon S3 deletes the specific version of the object.
  - If the version ID maps to the delete marker of that object, Amazon S3 deletes the delete marker. This makes the object reappear in your bucket."

---

What is returned when a GET is issued on an S3 object that has a delete marker?

237

- The latest non-deleted version
- A 404 (Object Not Found) error (x)
- A response header "x-amz-delete-marker: true" (x)

From the S3 documentation (<http://docs.aws.amazon.com/AmazonS3/latest/dev/DeleteMarker.html>):

>>>

#### Working with Delete Markers

A delete marker is a placeholder (marker) for a versioned object that was named in a simple DELETE request. Because the object was in a versioning-enabled bucket, the object was not deleted. The delete marker, however, makes Amazon S3 behave as if it had been deleted.

A delete marker has a key name (or key) and version ID like any other object. However, a delete marker differs from other objects in the following ways:

- It does not have data associated with it.
- It is not associated with an access control list (ACL) value.
- It does not retrieve anything from a GET request because it has no data; you get a 404 error.
- The only operation you can use on a delete marker is DELETE, and only the bucket owner can issue such a request.

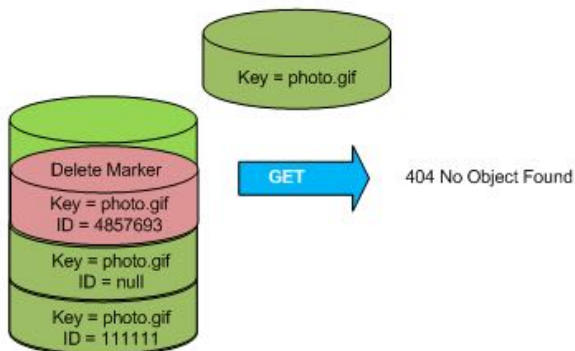
Delete markers accrue a nominal charge for storage in Amazon S3. The storage size of a delete marker is equal to the size of the key name of the delete marker. A key name is a sequence of Unicode characters. The UTF-8 encoding adds from 1 to 4 bytes of storage to your bucket for each character in the name. For more information about key names, see Object Keys. For information about deleting a delete marker, see Removing Delete Markers.

Only Amazon S3 can create a delete marker, and it does so whenever you send a DELETE Object request on an object in a versioning-enabled or suspended bucket. The object named in the DELETE request is not actually deleted. Instead, the delete marker becomes the current version of the object. (The object's key name (or key) becomes the key of the delete marker.) If you try to get an object and its current version is a delete marker, Amazon S3 responds with:

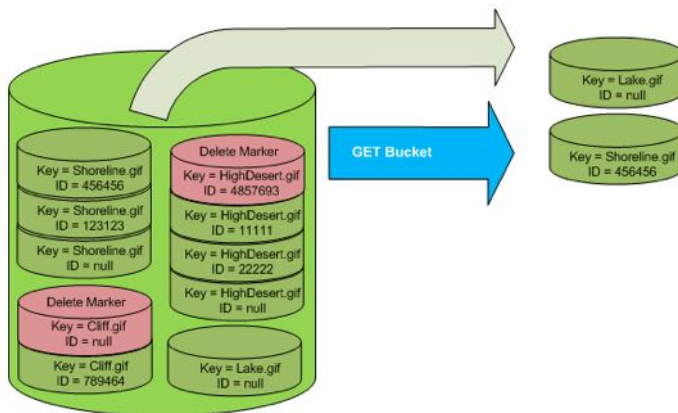
- A 404 (Object not found) error
- A response header, x-amz-delete-marker: true

The response header tells you that the object accessed was a delete marker. This response header never returns false; if the value is false, Amazon S3 does not include this response header in the response.

The following figure shows how a simple GET on an object, whose current version is a delete marker, returns a 404 No Object Found error.



The only way to list delete markers (and other versions of an object) is by using the versions subresource in a GET Bucket request. A simple GET does not retrieve delete marker objects. The following figure shows that a GET Bucket request does not return objects whose current version is a delete marker.



<<<

When cross-region replication is activated for an S3 bucket, are existing objects replicated?

- 
- Yes
  - No (x)

From the S3 documentation (<http://docs.aws.amazon.com/AmazonS3/latest/dev/crr-what-is-isnot-replicated.html>):

"[When replication is activated] Amazon S3 does not replicate the following:

- **Amazon S3 does not retroactively replicate objects that existed before you added replication configuration.**

---

When cross-region replication is activated for an S3 bucket, are objects encrypted with S3-managed server-side encryption replicated?

239

- 
- Yes (x)
  - No

From the S3 documentation (<http://docs.aws.amazon.com/AmazonS3/latest/dev/crr-what-is-isnot-replicated.html>):

"Amazon S3 replicates the following:

- Objects created with server-side encryption using the Amazon S3-managed encryption key. The replicated copy of the object is also encrypted using server-side encryption using the Amazon S3-managed encryption key."

However, objects encrypted with customer-provided or KMS keys are *not* replicated:

"[When replication is activated] Amazon S3 does not replicate the following:

- Objects created with server-side encryption using either customer-provided (SSE-C) or AWS KMS-managed encryption (SSE-KMS) keys are not replicated. For more information about server-side encryption, see Protecting Data Using Server-Side Encryption.

Amazon S3 does not keep the encryption keys you provide after the object is created in the source bucket so it cannot decrypt the object for replication, and therefore it does not replicate the object."

---

Does S3 allow CORS access to S3 objects?

240

- 
- Yes (x)
  - No

From the S3 documentation (<http://docs.aws.amazon.com/AmazonS3/latest/dev/cors.html>):

"Cross-origin resource sharing (CORS) defines a way for client web applications that are loaded in one domain to interact with resources in a different domain. With CORS support in Amazon S3, you can build rich client-side web applications with Amazon S3 and selectively allow cross-origin access to your Amazon S3 resources."

---

What operations are protected by S3 MFA Delete?

241

- 
- All operations
  - Permanently deleting an object version (x)
  - Permanently deletion a non-versioned object
  - Changing an object's versioning state (x)

The third option is a detractor; MFA can only be enabled for versioned objects. It's intended to be another form of protection against accidental deletion.

From the S3 documentation (<http://docs.aws.amazon.com/AmazonS3/latest/dev/Versioning.html#MultiFactorAuthenticationDelete>):

"You can optionally add another layer of security by configuring a bucket to enable MFA (Multi-Factor Authentication) Delete, which requires additional authentication for either of the following operations.

- Change the versioning state of your bucket
- Permanently delete an object version"

---

Is it possible to create a one-time pre-signed S3 download link?

242

- Yes
- No (x)

There's no capability to do this anywhere in the S3 documentation. The only option here is to create a pre-signed URL with a very short expiration.

---

What is the maximum expiration for a pre-signed S3 URL?

243

- 15 minutes
- 24 hours
- 1 month
- multiple years (x)

The S3 documentation on pre-signed URLs (<http://docs.aws.amazon.com/AmazonS3/latest/dev/RESTAuthentication.html#RESTAuthenticationQueryStringAuth>) describes the query string parameters used in a pre-signed URL, and the expires parameter is specified in seconds since the epoch. There is no mention of a limit, but this email thread:

<http://stackoverflow.com/questions/6633492/amazons3-getpresignedurlrequest-max-expires-date>

Indicates that the practical expiry limit is 2038.

---

Are updates to S3 objects fully consistent?

244

- Yes
- No (x)

From the S3 FAQ (<https://aws.amazon.com/s3/faqs/>):

**"Q: What data consistency model does Amazon S3 employ?**

Amazon S3 buckets in all Regions provide read-after-write consistency for PUTS of new objects and eventual consistency for overwrite PUTS and DELETES."

---

Are S3 object deletions fully consistent?

245

- 
- Yes
  - No (x)

From the S3 FAQ (<https://aws.amazon.com/s3/faqs/>):

**"Q: What data consistency model does Amazon S3 employ?"**

Amazon S3 buckets in all Regions provide read-after-write consistency for PUTS of new objects and eventual consistency for overwrite PUTS and DELETES."

---

Are newly inserted S3 objects immediately visible to readers ?

246

- 
- Yes (x)
  - No

From the S3 FAQ (<https://aws.amazon.com/s3/faqs/>):

**"Q: What data consistency model does Amazon S3 employ?"**

Amazon S3 buckets in all Regions provide read-after-write consistency for PUTS of new objects and eventual consistency for overwrite PUTS and DELETES."

---

When an IAM user creates an S3 resource (bucket, object, etc) in another account, what entity owns the resource?

247

- 
- The account in which the resource is created
  - The IAM user
  - The IAM user's parent account (x)

From the S3 documentation (<http://docs.aws.amazon.com/AmazonS3/latest/dev/access-control-overview.html>):

"By default, all Amazon S3 resources are private. Only a resource owner can access the resource. The resource owner refers to the AWS account that creates the resource. For example:

- The AWS account that you use to create buckets and objects owns those resources.
- If you create an AWS Identity and Access Management (IAM) user in your AWS account, your AWS account is the parent owner. **If the IAM user uploads an object, the parent account, to which the user belongs, owns the object.**
- A bucket owner can grant cross-account permissions to another AWS account (or users in another account) to upload objects. In this case, the AWS account that uploads objects owns those objects. The bucket owner does not have permissions on the objects that other accounts own, with the following exceptions:
  - The bucket owner pays the bills. The bucket owner can deny access to any objects, or delete any objects in the bucket, regardless of who owns them.
  - The bucket owner can archive any objects or restore archived objects regardless of who owns them. Archival refers to the storage class used to store the objects. For more information, see Object Lifecycle Management."

---

Is it possible to use S3 cross-region replication to copy objects into another account's bucket?

248

- 
- Yes (x)
  - No

From this blog: [https://aws.amazon.com/blogs/aws/new-cross-region-replication-for-amazon-s3/...](https://aws.amazon.com/blogs/aws/new-cross-region-replication-for-amazon-s3/)

"You will also need to set up an IAM role so that S3 can list and retrieve objects from the source bucket and to initiate replication operations on the destination bucket. Because you have the opportunity to control the policy document, you can easily implement advanced scenarios such as replication between buckets owned by separate AWS accounts."

- Yes (x)
- No

Yes, this is just a feature of HTTP (<http://www.w3.org/Protocols/rfc2616/rfc2616-sec14.html#sec14.35>) and S3 supports it. From the S3 API docs (<http://docs.aws.amazon.com/AmazonS3/latest/API/RESTObjectGET.html>):

«1778»

Range [a parameter to the GET operation]

Downloads the specified range bytes of an object. For more information about the HTTP Range header, go to <http://www.w3.org/Protocols/rfc2616/rfc2616-sec14.html#sec14.35>.

«1778»

If a single part of a multipart S3 upload fails, can that single part be retried?

250

- Yes (x)
- No

From the S3 documentation (<http://docs.aws.amazon.com/AmazonS3/latest/dev/mpuoverview.html>):

"When uploading a part, in addition to the upload ID, you must specify a part number. You can choose any part number between 1 and 10,000. A part number uniquely identifies a part and its position in the object you are uploading. **If you upload a new part using the same part number as a previously uploaded part, the previously uploaded part is overwritten.**"

What are some ways to maximize S3 performance?

251

- Randomize keys to distribute data across partitions (x)
- Use lots of ranged GETs over multiple threads (x)
- Use multipart PUTs to maximize bandwidth (x)
- Use CloudFront (x)
- Store objects in a region close to your users (x)
- Distribute objects across several buckets
- Split objects into smaller chunks

From this 2013 re:Invent slide deck (<http://www.slideshare.net/AmazonWebServices/maximizing-amazon-s3-performance-stg304-aws-reinvent-2013>):

## Wrap up: Maximizing Amazon S3 Performance



Is it possible to disable S3 versioning on a bucket once it's been turned on?

252



- 
- Yes
  - No (x)

From the S3 documentation (<http://docs.aws.amazon.com/AmazonS3/latest/UG/enable-bucket-versioning.html>):

"After you enable versioning on a bucket, it can be in only the enabled or suspended state; you cannot disable versioning on a bucket. "

---

What some mechanisms to secure S3 data?

253

- With bucket policies (x)
- Using MFA Delete (x)
- Storing objects in a separate account and using cross-account access (x)

...from Ryan's course

---

What is the maximum size of an S3 object?

254

- 256 KB
- 1 MB
- 1 GB
- 5 TB (x)

From the S3 FAQ (<https://aws.amazon.com/s3/faqs/>):

**"Q: How much data can I store?**

The total volume of data and number of objects you can store are unlimited. **Individual Amazon S3 objects can range in size from 1 byte to 5 terabytes.** The largest object that can be uploaded in a single PUT is 5 gigabytes. For objects larger than 100 megabytes, customers should consider using the Multipart Upload capability."

---

What is the minimum size of an S3 object?

255

- 1 byte (x)
- 512 bytes
- 1 KB
- 256 KB

From the S3 FAQ (<https://aws.amazon.com/s3/faqs/>):

**"Q: How much data can I store?**

The total volume of data and number of objects you can store are unlimited. **Individual Amazon S3 objects can range in size from 1 byte to 5 terabytes.** The largest object that can be uploaded in a single PUT is 5 gigabytes. For objects larger than 100 megabytes, customers should consider using the Multipart Upload capability."

---

If you expect a very large and sudden spike in S3 request rates, what should you do?

256

- 
- Nothing, S3 will automatically scale to ensure the best performance
  - Submit a support case to let Amazon know so they can prepare for the workload (x)
  - Execute the gets incrementally beforehand to warm the cache

From the S3 Developer Guide (<http://docs.aws.amazon.com/AmazonS3/latest/dev/request-rate-perf-considerations.html>):

"Amazon S3 scales to support very high request rates. If your workload in an Amazon S3 bucket routinely exceeds 100 PUT/LIST/DELETE requests per second or more than 300 GET requests per second, follow the guidelines in this topic to ensure the best performance and scalability. If your request rate grows steadily, Amazon S3 automatically partitions your buckets as needed to support higher request rates. **However, if you expect a rapid increase in the request rate for a bucket to more than 300 PUT/LIST/DELETE requests per second or more than 800 GET requests per second, we recommend that you open a support case to prepare for the workload and avoid any temporary limits on your request rate.** To open a support case, go to Contact Us."

---

Which are valid options to implement S3 encryption?

257

- Server-side encryption with S3-managed keys (x)
- Server-side encryption with KMS-managed keys (x)
- Server-side encryption with CloudHSM-managed keys
- Server-side encryption with Customer-managed keys (x)
- Client-side encryption with KMS-managed keys (x)
- Client-side encryption with a Customer-managed master key (x)
- Client-side encryption with CloudHSM-managed keys (x)

Server-side is described here (<http://docs.aws.amazon.com/AmazonS3/latest/dev/serv-side-encryption.html>) and client side here (<http://docs.aws.amazon.com/AmazonS3/latest/dev/UsingClientSideEncryption.html>).

CloudHSM isn't discussed in the server-side document, but it's definitely an option.

## Storage Gateway

---

Can a Storage Gateway point-in-time snapshot be used to create an EBS volume?

258

- Yes (x)
- No

From the Storage Gateway FAQ (

### "Q. How does the AWS Storage Gateway work?"

...You can also take point-in-time snapshots of your Gateway-Cached volume data in Amazon S3 in the form of Amazon EBS snapshots, enabling you to make space-efficient versioned copies of your volumes for data protection and various data reuse needs."

---

What is the maximum amount of data that can be stored in a Gateway-Stored volume?

259

- 1 TB
- 16 TB (x)
- 32 TB
- 128 TB

From the Storage Gateway FAQ (<https://aws.amazon.com/storagegateway/faqs/>):

### "Q. What is the maximum size of a volume?"

Each Gateway-Stored volume can store up to 16 TB of data. Data written to the volume is stored on your on-premises hardware and asynchronously backed up to AWS for point-in-time snapshots."

- 1 TB
- 16 TB
- 32 TB (x)
- 128 TB

From the Storage Gateway FAQ (<https://aws.amazon.com/storagegateway/faqs/>):

**"Q. What is the maximum size of a volume?"**

Each Gateway-Cached volume can store up to 32 TB of data. Data written to the volume is cached on your on-premises hardware and asynchronously uploaded to AWS for durable storage."

- 192 TB
- 512 TB
- 1 PB (x)
- 5 PB

From the Storage Gateway FAQ (<https://aws.amazon.com/storagegateway/faqs/>):

**"Q. How much volume data can I manage per gateway?"**

...

Each Gateway-Cached gateway can support up to 32 volumes for a maximum of 1 PB of data (32 volumes, each 32 TB in size)."

- 192 TB (x)
- 512 TB
- 1 PB
- 5 PB

From the Storage Gateway FAQ (<https://aws.amazon.com/storagegateway/faqs/>):

**"Q. How much volume data can I manage per gateway?"**

...

Each Gateway-Stored gateway can support up to 12 volumes for a maximum of 192 TB of data (12 volumes, each 16 TB in size)."

- Yes
- No (x)

From the Storage Gateway documentation (<http://docs.aws.amazon.com/storagegateway/latest/userguide/storage-gateway-cached-concepts.html>):

"All gateway-cached volume data and snapshot data is stored in Amazon S3 encrypted at rest using server-side encryption (SSE). However, you cannot access this data with the Amazon S3 API or other tools such as the Amazon S3 console."

- 
- VTLs are stored in Glacier
  - VTLs are stored in S3 (x)
  - VTSs are stored in Glacier (x)
  - VTSs are stored in S3

From the Storage Gateway FAQ (<https://aws.amazon.com/storagegateway/faqs/>):

**"Q. What is the AWS Storage Gateway?"**

...

*Gateway-Virtual Tape Library (VTL):* With Gateway-VTL you can have a limitless collection of virtual tapes. Each virtual tape can be stored in a Virtual Tape Library backed by Amazon S3 or a Virtual Tape Shelf backed by Amazon Glacier. The Virtual Tape Library exposes an industry standard iSCSI interface which provides your backup application with on-line access to the virtual tapes. When you no longer require immediate or frequent access to data contained on a virtual tape, you can use your backup application to move it from its Virtual Tape Library to your Virtual Tape Shelf in order to further reduce your storage costs."

---

How long does it take to load a virtual tape from a Virtual Tape Shelf into a Virtual Tape Library?

265

- 1 hour
- 12 hours
- 24 hours (x)
- 48 hours

From the Storage Gateway User Guide (<http://docs.aws.amazon.com/storagegateway/latest/userguide/managing-vts-vtl.html>):

**"Q. How does the AWS Storage Gateway work?"**

To access virtual tape data in the VTL, you must first retrieve the virtual tape you want from a VTS to your gateway-VTL. **It takes up to 24 hours for the tape to be available in your gateway-VTL."**

---

What is the maximum capacity of a Storage Gateway Virtual Tape Library?

266

- 10 TB
- 25 TB
- 50 TB
- 150 TB (x)
- 300 TB

From the Storage Gateway FAQ (<https://aws.amazon.com/storagegateway/faqs/>):

**"Q. How much data can I store in a Virtual Tape Library?"**

Each Virtual Tape Library (VTL) can store up to 1,500 virtual tapes with a maximum aggregate capacity of 150 TB."

---

How long does it take to retrieve a tape from a Virtual Tape Shelf into a Virtual Tape Library?

267

- 
- Instantaneous
  - 1 hour
  - 12 hours
  - 24 hours (x)

From the Storage Gateway documentation (<http://docs.aws.amazon.com/storagegateway/latest/userguide/storage-gateway-vtl-concepts.html>):

"Retrieving tapes – Tapes archived to the VTS cannot be read directly. To read an archived tape, you must first retrieve it to your gateway-VTL either by using the AWS Storage Gateway console or by using the AWS Storage Gateway API. A retrieved tape will be available in your VTL in about 24 hours."

---

When your backup software ejects a tape from a Virtual Tape Library, what happens to it?

268

- Deleted forever
- Compressed but still available in S3
- Moved to a Virtual Tape Shelf in Glacier (x)

From the Storage Gateway documentation (<http://docs.aws.amazon.com/storagegateway/latest/userguide/storage-gateway-vtl-concepts.html>):

"Archiving tapes – When your backup software ejects a tape, your gateway moves the tape to the VTS for long-term storage. The VTS is located in the AWS region in which you activated the gateway. Tapes in the VTS are stored in Amazon Glacier, an extremely low-cost storage service for data archiving and backup. For more information, go to Amazon Glacier."

---

Can you schedule a Storage Gateway snapshot?

269

- Yes (x)
- No

From the AWS Storage Guide (<http://docs.aws.amazon.com/storagegateway/latest/userguide/WorkingWithSnapshots.html>):

"AWS Storage Gateway provides the ability to back up point-in-time snapshots of your data to Amazon S3 for durable recovery. You can use the snapshot backups later on-premises or in Amazon Elastic Compute Cloud (Amazon EC2), **and you can take snapshots on a one-time or scheduled basis.**"

## Import / Export

---

What import targets are supported by Import / Export?

270

- 
- S3 (x)
  - EBS (x)
  - Glacier (x)
  - Dynamo
  - Redshift
  - RDS

From the Amazon Import/Export documentation (<http://docs.aws.amazon.com/AWSImportExport/latest/DG/whatisdisk.html>):

"Disk is a good choice if you have a small amount of data to import into Amazon S3, Amazon Glacier, or Amazon Elastic Block Store (Amazon EBS), and currently it's the only supported data export solution for Amazon S3. Your data load typically begins the next business day after your storage device arrives at AWS. After the data export or import completes, we return your storage device. For large data sets, Disk can be significantly faster than Internet transfer and more cost-effective than upgrading your connectivity.

Disk supports the following:

- Import to Amazon S3
- Export from Amazon S3
- Import to Amazon EBS
- Import to Amazon Glacier

Disk does not currently support export from Amazon EBS or Amazon Glacier."

---

What export sources are supported by Import / Export?

271

- 
- S3 (x)
  - EBS
  - Glacier
  - Dynamo
  - Redshift
  - RDS

From the Amazon Import/Export documentation (<http://docs.aws.amazon.com/AWSImportExport/latest/DG/whatisdisk.html>):

"Disk is a good choice if you have a small amount of data to import into Amazon S3, Amazon Glacier, or Amazon Elastic Block Store (Amazon EBS), and currently it's the only supported data export solution for Amazon S3. Your data load typically begins the next business day after your storage device arrives at AWS. After the data export or import completes, we return your storage device. For large data sets, Disk can be significantly faster than Internet transfer and more cost-effective than upgrading your connectivity.

Disk supports the following:

- Import to Amazon S3
- Export from Amazon S3
- Import to Amazon EBS
- Import to Amazon Glacier

Disk does not currently support export from Amazon EBS or Amazon Glacier."

---

What are the encryption options for Import / Export?

272

- 
- Optional for imports, mandatory for exports (x)
  - Optional for exports, mandatory for imports
  - Optional for both import and export
  - Mandatory for both import and export

From the Import / Export Developer Guide (<http://docs.aws.amazon.com/AWSImportExport/latest/DG/encrypting-your-data.html>):

"For export from Amazon S3, **AWS always requires data encryption**, either by using a PIN-code device with hardware-based encryption or by using TrueCrypt software encryption. For added security, you can use both methods. You will need to provide the PIN or password, or both, in your manifest file.

**For import jobs, we strongly recommend encrypting your data.** For import to Amazon S3, you can use a PIN-code device with hardware-based encryption or TrueCrypt software to encrypt your data before sending it to AWS Import/Export. You will need to include your PIN code or TrueCrypt password in your import manifest and the TrueCrypt password in your export manifest.

For import to Amazon EBS or Amazon Glacier, you can use a PIN-code device with hardware-based encryption or any software encryption method you choose, or both. AWS uses your PIN to access a PIN-code device, but does not decrypt software-encrypted data for import to Amazon EBS or Amazon Glacier. You will need to include your PIN in your import manifest."

## ELB

---

Is it possible to assign a security group to an ELB?

273

- Yes (x)
- No

From the EC2 FAQ (<https://aws.amazon.com/ec2/faqs/#elastic-load-balancing>):

**"Q: Can I configure a security group for the front-end of the Elastic Load Balancer?"**

If you are using Amazon Virtual Private Cloud, you can configure security groups for the front-end of your Elastic Load Balancer."

---

Is it possible to terminate SSL on an ELB?

274

- Yes (x)
- No

From the EC2 documentation (<http://docs.aws.amazon.com/ElasticLoadBalancing/latest/DeveloperGuide/elb-listener-config.html#https-ssl-listeners>):

"If you use HTTPS or SSL for your front-end listener, you must install an X.509 certificate (SSL server certificate) on your load balancer. The load balancer uses the certificate to terminate the connection and then decrypt requests from clients before sending them to the back-end instances."

---

What are the valid statuses for an instance in an ELB?

275

- 
- InService (x)
  - OutOfService (x)
  - Unknown (x)
  - Shutdown
  - Starting

From the ELB API documentation ([http://docs.aws.amazon.com/ElasticLoadBalancing/latest/APIReference/API\\_InstanceState.html](http://docs.aws.amazon.com/ElasticLoadBalancing/latest/APIReference/API_InstanceState.html)):

**"State**

The current state of the instance.

Valid values: InService | OutOfService | Unknown

Type: String

Required: No"

---

Is it possible to check the status of a particular instance in an ELB?

276

- 
- Yes (x)
  - No

From the ELB API ([http://docs.aws.amazon.com/ElasticLoadBalancing/latest/APIReference/API\\_InstanceState.html](http://docs.aws.amazon.com/ElasticLoadBalancing/latest/APIReference/API_InstanceState.html)):

...check the link for details

---

Is it possible to assign an EIP to an ELB?

277

- 
- Yes
  - No (x)

Can't seem to find anything conclusive on this, but you can't. :)

---

What TCP ports will ELB load balance?

278

- 
- 25 (x)
  - 53
  - 80 (x)
  - 443 (x)
  - 8080
  - 1024-65535 (x)

From the EC2 FAQ (<https://aws.amazon.com/ec2/faqs/#elastic-load-balancing>):

**"Q: What TCP ports can I load balance?**

You can perform load balancing for the following TCP ports: 25, 80, 443, and 1024-65535."

---

Does ELB support IPv6?

279



- 
- Yes (x)
  - No

From the EC2 FAQ (<https://aws.amazon.com/ec2/faqs/#elastic-load-balancing>):

**"Q: Does Elastic Load Balancing support IPv6 traffic?"**

Yes. Each Elastic Load Balancer has an associated IPv4, IPv6, and dualstack (both IPv4 and IPv6) DNS name. IPv6 is not supported in VPC at this time."

---

Can I get a history of ELB API calls?

280

- 
- Yes (x)
  - No

From the EC2 FAQ (<https://aws.amazon.com/ec2/faqs/#elastic-load-balancing>):

**"Q: Can I get a history of Elastic Load Balancing API calls made on my account for security analysis and operational troubleshooting purposes?"**

Yes. To receive a history of Elastic Load Balancing API calls made on your account, simply turn on CloudTrail in the AWS Management Console."

---

Is it possible to store more than one SSL certificate on an ELB?

281

- 
- Yes
  - No (x)

From Ryan's course...

---

What is the default ELB timeout for client and back-end connections?

282

- 
- 15 seconds
  - 60 seconds (x)
  - 120 seconds
  - 300 seconds

From the ELB Developer Guide ():

"For each request that a client makes through a load balancer, the load balancer maintains two connections. One connection is with the client and the other connection is to the back-end instance. For each connection, the load balancer manages an idle timeout that is triggered when no data is sent over the connection for a specified time period. After the idle timeout period has elapsed, if no data has been sent or received, the load balancer closes the connection.

**By default, Elastic Load Balancing sets the idle timeout to 60 seconds for both connections.** If an HTTP request doesn't complete within the idle timeout period, the load balancer closes the connection, even if data is still being transferred. You can change the idle timeout setting for the connections to ensure that lengthy operations, such as file uploads, have time to complete."

---

What is the benefit of ELB cross-zone load balancing?

283

- 
- Ensures that traffic is spread across all back-end instances regardless of availability zone (x)
  - Ensures that traffic is spread evenly across availability zones regardless of the number of instances in those zones
  - Ensures that the ELB itself is operating in multiple availability zones for high-availability

From the ELB Developer Guide (<http://docs.aws.amazon.com/ElasticLoadBalancing/latest/DeveloperGuide/enable-disable-crosszone-lb.html>):

"By default, your load balancer distributes incoming requests evenly across its enabled Availability Zones. **To ensure that your load balancer distributes incoming requests evenly across all back-end instances, regardless of the Availability Zone that they are in, enable cross-zone load balancing.** Cross-zone load balancing reduces the need to maintain equivalent numbers of back-end instances in each Availability Zone, and improves your application's ability to handle the loss of one or more back-end instances. However, we still recommend that you maintain approximately equivalent numbers of instances in each Availability Zone for higher fault tolerance."

---

When using ELB cross-zone load balancing, do you pay inter-zone transfer between the load balancer and back-end instances? 284

- Yes
- No (x)

From the ELB Developer Guide (<http://docs.aws.amazon.com/ElasticLoadBalancing/latest/DeveloperGuide/enable-disable-crosszone-lb.html>):

"When you enable cross-zone load balancing, you are not charged for data transfer between the load balancer nodes and back-end instances."

---

Can you use a self-signed certificate when terminating SSL on an ELB? 285

- Yes (x)
- No

From the ELB Developer Guide (<http://docs.aws.amazon.com/ElasticLoadBalancing/latest/DeveloperGuide/ssl-server-cert.html#create-cert>):

"To create an SSL server certificate, you must generate an RSA private key and create a Certificate Signing Request (CSR). Next, either have your certificate signed by a Certificate Authority (CA), **or generate a self-signed certificate** so that you can test your SSL implementation while waiting for the CA to sign your certificate. [...]"

---

What ELB protocols support the Proxy Protocol? 286

- TCP (x)
- HTTP
- SSL (e.g., for HTTPS terminated on the back end) (x)

From the ELB Developer Guide (<http://docs.aws.amazon.com/ElasticLoadBalancing/latest/DeveloperGuide/enable-proxy-protocol.html>):

"Proxy Protocol is an Internet protocol used to carry connection information from the source requesting the connection to the destination for which the connection was requested. Elastic Load Balancing uses Proxy Protocol version 1, which uses a human-readable header format.

By default, when you use Transmission Control Protocol (TCP) or Secure Sockets Layer (SSL) for both front-end and back-end connections, your load balancer forwards requests to the back-end instances without modifying the request headers. If you enable Proxy Protocol, a human-readable header is added to the request header with connection information such as the source IP address, destination IP address, and port numbers. The header is then sent to the back-end instance as part of the request.

**You can enable Proxy Protocol on ports that use either the SSL and TCP protocols. You can use Proxy Protocol to capture the source IP of your client when you are using a non-HTTP protocol, or when you are using HTTPS and not terminating the SSL connection on your load balancer."**

This blog post (<https://aws.amazon.com/blogs/aws/elastic-load-balancing-adds-support-for-proxy-protocol/>) goes into additional detail.

## ElastiCache

How is an ElasticCache Redis Replication Group structured?

287

- One ElastiCache cluster for the master with additional clusters as replicas, each cluster with a single node (x)
- One ElastiCache cluster for the master with additional clusters as replicas, each cluster with multiple nodes
- One ElastiCache node for the master with additional nodes as replicas

It's a bit weird, but basically there's a master node with replicas collectively called a **replication group**. If a replication group is **multi-AZ enabled**, the replicas will live in different AZs **and** the cluster will automatically fail over to a read replica if the master node fails. For some reason auto-failover is only enabled for multi-AZ replication groups.

The other twist is that the master and replica "nodes" are actually **clusters**. Right now a Redis cluster can only have a single node, so one cluster == one node and they are synonyms in this context. Apparently Amazon is preparing for the day when the master and replica clusters can each contain multiple nodes.

Here's a lot more detail from the ElastiCache FAQ (<https://aws.amazon.com/elasticache/faqs/>):

### "Q: What are Amazon ElastiCache for Redis nodes, clusters, and replications groups?"

An ElastiCache for Redis node is the smallest building block of an Amazon ElastiCache for Redis deployment. Each ElastiCache for Redis node supports the Redis protocol and has its own DNS name and port. Multiple types of ElastiCache for Redis nodes are supported, each with varying amount of CPU capability, and associated memory. An ElastiCache for Redis node may take on a primary or a read replica role. A primary node can be replicated to multiple read replica nodes. An ElastiCache for Redis cluster is a collection of one or more ElastiCache for Redis nodes of the same role; the primary node will be in the primary cluster and the read replica node will be in a read replica cluster. At this time a cluster can only have one node. In the future, we will increase this limit. A cluster manages a logical key space, where each node is responsible for a part of the key space. Most of your management operations will be performed at the cluster level. An ElastiCache for Redis replication group encapsulates the primary and read replica clusters for a Redis installation. A replication group will have only one primary cluster and zero or many read replica clusters. All nodes within a replication group (and consequently cluster) will be of the same node type, and have the same parameter and security group settings."

Can ElastiCache nodes incur downtime during software maintenance?

288

- 
- Yes (x)
  - No

From the ElastiCache FAQ ():

**"Q: What is a maintenance window? Will my Cache Nodes be available during software maintenance?"**

You can think of the Amazon ElastiCache maintenance window as an opportunity to control when software patching occurs, in the event either are requested or required. If a "maintenance" event is scheduled for a given week, it will be initiated and completed at some point during the 60 minute maintenance window you identify.

**Your Cache Nodes could incur some downtime during your maintenance window if software patching is scheduled.** Please refer to Cache Engine Version Management for more details. Patching can be user requested - for example cache software upgrade, or determined as required (if we identify any security vulnerabilities in the system or caching software). Software patching occurs infrequently (typically once every few months) and should seldom require more than a fraction of your maintenance window. If you do not specify a preferred weekly maintenance window when creating your Cache Cluster, a 60 minute default value is assigned. If you wish to modify when maintenance is performed on your behalf, you can do so by modifying your DB Instance in the AWS Management Console or by using the ModifyCacheCluster API. Each of your Cache Clusters can have different preferred maintenance windows, if you so choose."

---

Is it possible to access an ElastiCache node from the public internet?

289

- 
- Yes
  - No (x)

From the ElastiCache FAQ ():

**"Q: Can programs running on servers in my own data center access Amazon ElastiCache?"**

No. Currently, all clients to an ElastiCache Cluster must be within the Amazon EC2 network, and authorized via security groups as described here."

---

Is a Cache Security Group required for ElastiCache nodes deployed within a VPC?

290

- 
- Yes
  - No (x)

From the ElastiCache FAQ (<https://aws.amazon.com/elasticache/faqs/>):

**"Q: How do I control network access to my Cache Cluster?"**

Amazon ElastiCache allows you to control access to your Cache Cluster and therefore the Cache Nodes using Cache Security Groups in non-VPC deployments. A Cache Security Group acts like a firewall controlling network access to your Cache Node. By default, network access is turned off to your Cache Nodes. If you want your applications to access your Cache Node, you can set your Cache Security Group to allow access from EC2 Instances with specific EC2 Security Group membership or IP ranges. This process is called ingress. Once ingress is configured for a Cache Security Group, the same rules apply to all Cache Nodes associated with that Cache Security Group. Cache Security Groups can be configured with the "Cache Security Groups" section of the Amazon ElastiCache Console or using the Amazon ElastiCache APIs.

**In VPC deployments, access to your cache nodes is controlled using the VPC Security Group and the Cache Subnet Group. The VPC Security Group is the VPC equivalent of the Cache Security Group."**

also

**"Q: Can I use Cache Security Groups to configure the cache clusters that are part of my VPC?"**

No, Cache Security Groups are not used when operating in a VPC. Instead they are used in the non VPC settings. When creating a cache cluster in a VPC you will need to use VPC Security Groups."

---

What is ElasticCache "Auto Discovery"?

291

- 
- A feature that enables automatic discovery of cache nodes by memcached clients (x)
  - A feature that allows an ElasticCache cluster to automatically discover clients that meet specific criteria
  - A feature that allows an ElasticCache cluster to automatically reconfigure itself when configuration sets change

From the ElasticCache FAQ (x):

**"Q: What is Auto Discovery and what can I do with it?"**

Auto Discovery is a feature that saves developers time and effort, while reducing complexity of their applications. **Auto Discovery enables automatic discovery of cache nodes by clients when they are added to or removed from an Amazon ElastiCache cluster.** Until now to handle cluster membership changes, developers must update the list of cache node endpoints manually. Depending on how the client application is architected, typically a client initialization, by shutting down the application and restarting it, is needed resulting in downtime. Through Auto Discovery we are eliminating this complexity. With Auto Discovery, in addition to being backwards protocol-compliant with the Memcached protocol, Amazon ElastiCache provides clients with information on cache cluster membership. A client capable of processing the additional information reconfigures itself, without any initialization, to use the most current nodes of an Amazon ElastiCache cluster.

**Q: How does Auto Discovery work?**

An Amazon ElastiCache cluster can be created with nodes that are addressable via named endpoints. With Auto Discovery the Amazon ElastiCache cluster is also given a unique Configuration Endpoint which is a DNS Record that is valid for the lifetime of the cluster. This DNS Record contains the DNS Names of the nodes that belong to the cluster. Amazon ElastiCache will ensure that the Configuration Endpoint always points to at least one such "target" node. A query to the target node then returns endpoints for all the nodes of the cluster in question. After this, you can connect to the cluster nodes just as before and use the Memcached protocol commands such as get, set, incr and decr. For more details, see here. To use Auto Discovery, you will need an Auto Discovery capable client. Auto Discovery clients for Java and PHP are available for download from the Amazon ElastiCache console. Upon initialization, the client will automatically determine the current members of the Amazon ElastiCache cluster using the Configuration Endpoint. When you make changes to your cache cluster by adding or removing nodes or if a node is replaced upon failure, the Auto Discovery client automatically determines the changes and you do not need to initialize your clients manually.

---

When will ElastiCache automatically upgrade a Memcached cluster?

292

- whenever a new version is released
- never
- to address security vulnerabilities (x)

From the ElastiCache FAQ (<https://aws.amazon.com/elasticache/faqs/>):

**"Q: Can I control if and when the engine version powering Amazon ElastiCache Cluster is upgraded to new supported versions?"**

Amazon ElastiCache allows you to control if and when the Memcached protocol-compliant software powering your Cache Cluster is upgraded to new versions supported by Amazon ElastiCache. This provides you with the flexibility to maintain compatibility with specific Memcached versions, test new versions with your application before deploying in production, and perform version upgrades on your own terms and timelines. Version upgrades involve some compatibility risk, thus they will not occur automatically and must be initiated by you. This approach to cache software patching puts you in the driver's seat of version upgrades, but still offloads the work of patch application to Amazon ElastiCache. You can learn more about version management by reading the FAQs that follow. Alternatively, you can refer to the Amazon ElastiCache User Guide. **While Cache Engine Version Management functionality is intended to give you as much control as possible over how patching occurs, we may patch your Cache Cluster on your behalf if we determine there is any security vulnerability in the system or cache software."**

---

What are valid scaling mechanisms for a Memcached ElastiCache cluster?

293

- 
- Add nodes to the cluster (x)
  - Extend the replication group
  - Create a new cluster with a larger instance type (x)
  - Memcached ElastiCache clusters can't be scaled

Replication groups are specific to Redis.

From the ElastiCache User Guide (<http://docs.aws.amazon.com/AmazonElastiCache/latest/UserGuide/Scaling.Memcached.html>):

"Memcached clusters are designed for easy scaling. A Memcached cluster can have from 1 to 20 nodes. To scale your Memcached cluster, merely add or remove nodes."

Also from the User Guide (<http://docs.aws.amazon.com/AmazonElastiCache/latest/UserGuide/Scaling.html#Scaling.UpDown>):

"ElastiCache does not support dynamically changing a cluster's node instance type. Therefore, when you scale up or down you must create a new cluster."

---

What are valid scaling mechanisms for a Redis ElastiCache cluster?

294

- Add nodes to the cluster
- Extend the replication group (x)
- Create a new cluster with a larger instance type (x)
- Memcached ElastiCache clusters can't be scaled

From the ElastiCache User Guide (<http://docs.aws.amazon.com/AmazonElastiCache/latest/UserGuide/Scaling.Redis.html>):

"Scaling your Redis Cluster

**Redis clusters are single-node. Therefore, you cannot scale out/in by partitioning your data across more or fewer nodes. If you need to scale your Redis implementation you can increase or decrease the size of the node instance type in your cluster.** For example, if your cluster is currently running on a cache.m3.large node which has 6.05 GB of memory, you could scale up by moving your cluster to a cache.m3.xlarge node, which has 13.3 GB of memory, or scale down to a cache.m3.medium which has 2.78 GB of memory. If you decide to scale down, you must be sure that the new node type has sufficient memory for your data and overhead. For more information on Redis memory requirements, see [Ensuring You Have Sufficient Memory to Create a Redis Snapshot](#).

For information on changing a cluster's node type, see [Scaling Up/Down: Changing a Cluster's Node Instance Type](#). **Even though you cannot scale your Redis cluster out or in because it has only one node, you can create a Redis replication group (Creating a Replication Group) and balance your reads over a larger or smaller number of clusters by adding or removing read replicas.** For additional information, see [Adding a Read Replica to a Replication Group](#) and [Deleting a Read Replica](#).

---

Which ElastiCache types support backup and restore?

295

- 
- Redis (x)
  - Memcached
  - Both Redis and Memcached

From the User Guide (<http://docs.aws.amazon.com/AmazonElastiCache/latest/UserGuide/Snapshots.html>):

**"Amazon ElastiCache clusters running Redis can back up their data.** The backup can be used to restore a cluster or seed a new cluster. A *backup* is a snapshot of a cluster at a specific moment in time. The backup consists of the cluster's metadata, along with all of the data in the cluster. All backups are written to Amazon Simple Storage Service (Amazon S3), which provides durable storage. At any time, you can restore your data by creating a new Redis cluster and populating it with data from a backup. "

From the ElastiCache FAQ (<https://aws.amazon.com/elasticache/faqs/>):

**Q: Does ElastiCache for Memcached support Backup and Restore?**

No, snapshots are available only for ElastiCache for Redis."

---

Are ElastiCache backups retained when a cluster is deleted?

296

- Yes
- No
- Yes, but only manual backups (x)
- Yes, but only automated backups

From the ElastiCache FAQ (<https://aws.amazon.com/elasticache/faqs/>):

**"Q: What happens to my snapshots if I delete my ElastiCache for Redis cluster?**

When you delete an ElastiCache for Redis cluster, your manual snapshots are retained. You will also have an option to create a final snapshot before the cluster is deleted. Automatic cache snapshots are not retained."

---

Can you use an ElastiCache backup made in one account to warm a cluster in a different account?

297

- Yes
- No (x)

From the ElastiCache FAQ (<https://aws.amazon.com/elasticache/faqs/>):

**"Q: I have multiple AWS accounts using ElastiCache for Redis. Can I use ElastiCache snapshots from one account to warm start an ElastiCache for Redis cluster in a different one?**

Not at this point."

---

Is there a performance impact when taking an ElastiCache snapshot?

298

- Yes (x)
- No

From the ElastiCache FAQ (<https://aws.amazon.com/elasticache/faqs/>):

**"Q: What is the performance impact of taking a snapshot?**

While taking a snapshot, you may encounter increased latencies for a brief period at the node. Snapshots use Redis's built-in BGSAVE and are subject to its strengths and limitations. In particular, the Redis process forks and the parent continues to serve requests while the child saves the data on disk and then exits. The forking increases the memory usage for the duration of the snapshot generation. When this memory usage exceeds that of the available memory of the cache node, swapping can get triggered, further slowing down the node. For this reason, we recommend generating snapshots on one of the read replicas (instead of the primary). Also, we suggest setting the reserved-memory parameter to minimize swap usage. See here for more details."

- Yes
- No
- Yes, but only for Redis clusters (x)
- Yes, but only for Memcached clusters

From the ElastiCache User Guide (<http://docs.aws.amazon.com/AmazonElastiCache/latest/UserGuide/RedisAOF.html>):

"By default, the data in a Redis node on ElastiCache resides only in memory, and is not persistent. If a node is rebooted, or if the underlying physical server experiences a hardware failure, the data in the cache is lost.

If you require data durability, you can enable the Redis append-only file feature (AOF). When this feature is enabled, the node writes all of the commands that change cache data to an append-only file. When a node is rebooted and the cache engine starts, the AOF is "replayed"; the result is a warm Redis cache with all of the data intact.

AOF is disabled by default."

- Yes
- No (x)

From the ElastiCache FAQ (<https://aws.amazon.com/elasticache/faqs/>):

**"Q: Is Redis password functionality supported in Amazon ElastiCache for Redis?"**

No, Amazon ElastiCache for Redis does not support Redis passwords. This is because of the inherent limitations of passwords stored in a configuration file. Instead of relying on Redis passwords, ElastiCache for Redis clusters are associated with an EC2 security group, and only clients within this security group have access to the Redis server."

- 2
- 5 (x)
- 10
- 25

From the ElastiCache FAQ (<https://aws.amazon.com/elasticache/faqs/>):

**"Q: How many read replicas can I create for a given primary cache node?"**

At this time, Amazon ElastiCache allows you to create up to five (5) read replicas for a given primary cache node."



- 
- Synchronous
  - Asynchronous (x)

From the ElastiCache FAQ (<https://aws.amazon.com/elasticache/faqs/>):

**"Q: Will my read replica be kept up-to-date with its primary cache node?"**

Updates to a primary cache node will automatically be replicated to any associated read replicas. **However, with Redis's asynchronous replication technology, a read replica can fall behind its primary cache node for a variety of reasons.** Typical reasons include:

- Write I/O volume to the primary cache node exceeds the rate at which changes can be applied to the read replica
- Network partitions or latency between the primary cache node and a read replica

Read replicas are subject to the strengths and weaknesses of Redis replication. If you are using read replicas, you should be aware of the potential for lag between a read replica and its primary cache node, or "inconsistency". Click here for guidance on how to find out the "inconsistency" of your read replica."

---

Is it possible to create an ElastiCache read replica in a different region than the master cluster?

303

- 
- Yes
  - No (x)

From the ElastiCache FAQ (<https://aws.amazon.com/elasticache/faqs/>):

**"Q: Can I create a read replica in another region as my primary?"**

No. Your read replica may only be provisioned in the same or different Availability Zone of the same Region as your cache node primary."

---

Is there a charge for the ElastiCache multi-AZ feature?

304

- 
- Yes
  - No (x)

From the ElastiCache FAQ (<https://aws.amazon.com/elasticache/faqs/>):

**"Q: How much does it cost to use Multi-AZ?"**

Multi-AZ is free of charge. You only pay for the ElastiCache nodes that you use."

---

Is it possible to take an ElastiCache backup from a read replica?

305

- 
- Yes (x)
  - No

From the ElastiCache FAQ (<https://aws.amazon.com/elasticache/faqs/>):

**"Q: Can I create a snapshot from an ElastiCache for Redis read replica?"**

Yes. Creating a snapshot from a read replica is the best way to backup your data while minimizing performance impact."

---

Is it possible to change an ElastiCache cluster's instance type "in-place"?

306

- 
- Yes, only a restart is required
  - No, you have to recreate the cluster (x)

From the ElastiCache User Guide (<http://docs.aws.amazon.com/AmazonElastiCache/latest/UserGuide/Scaling.html#Scaling.UpDown>):

**"ElastiCache does not support dynamically changing a cluster's node instance type. Therefore, when you scale up or down you must create a new cluster.** If you are using the Redis engine, you can seed the cluster from a backup. New Memcached clusters always start out empty."

---

Is it possible to use an existing snapshot to seed a new Memcached ElasticCache cluster?

307

- 
- Yes
  - No (x)

From the ElastiCache User Guide (<http://docs.aws.amazon.com/AmazonElastiCache/latest/UserGuide/Scaling.html#Scaling.UpDown>):

"ElastiCache does not support dynamically changing a cluster's node instance type. Therefore, when you scale up or down you must create a new cluster. If you are using the Redis engine, you can seed the cluster from a backup. **New Memcached clusters always start out empty.**"

also:

"If your cluster is running the Memcached engine, the new cluster will start out empty, unless your application populates it."

Also, from the ElastiCache FAQ (<https://aws.amazon.com/elasticache/faqs/>):

**Q: Does ElastiCache for Memcached support Backup and Restore?**

No, **snapshots are available only for ElastiCache for Redis.**"

---

Does a customer-initiated reboot of a primary ElastiCache Redis cluster trigger automatic failover to another cluster in a replication group? 308

- 
- Yes
  - No (x)

From the ElastiCache User Guide (<http://docs.aws.amazon.com/AmazonElastiCache/latest/UserGuide/AutoFailover.html>):

"A customer-initiated reboot of a primary cluster does not trigger automatic failover. Other reboots and failures do trigger automatic failover."

---

Is ElastiCache Auto Discovery available for Redis clusters?

309

- 
- Yes
  - No (x)

From the ElastiCache User Guide (<http://docs.aws.amazon.com/AmazonElastiCache/latest/UserGuide/AutoDiscovery.html>):

"Auto Discovery is only available for cache clusters running the Memcached engine. Redis cache clusters are single node clusters, thus there is no need to identify and track all the nodes in a Redis cluster."

---

How long does failover take in a Multi-AZ Redis ElastiCache cluster?

310

- 
- Under a minute
  - A few minutes (x)
  - Several minutes
  - Up to fifteen minutes

From the ElastiCache User Guide (<http://docs.aws.amazon.com/AmazonElastiCache/latest/UserGuide/AutoFailover.html>):

"An ElastiCache replication group consists of a primary cluster and up to five read replicas. During certain types of planned maintenance, or in the unlikely event of a primary cluster or Availability Zone failure, if your replication group is Multi-AZ enabled, ElastiCache will automatically detect the primary cluster's failure, select a read replica cluster and promote it to primary cluster so that you can resume writing to the new primary cluster as soon as promotion is complete. ElastiCache also propagates the DNS of the promoted replica so that if your application is writing to the primary endpoint, no endpoint change will be required in your application. However, because you read from individual endpoints, you will need to change the read endpoint of the replica promoted to primary cluster to the new replica's endpoint.

**The promotion process generally takes just a few minutes, which is much faster than recreating and provisioning a new primary cluster if you do not enable Multi-AZ."**

## VPC

---

Does VPC support multicast or broadcast?

311

- Yes
- No (x)

---

Is it possible to use route tables to restrict access between VPC subnets?

312

- Yes
- No (x)

Each route table has an un-deletable rule that routes anything to the VPCs CIDR block to a target of "local", i.e., the VPCs router. The router knows about all the VPCs attached subnets and how to route amongst them, and there's no way to change that.

From the VPC documentation ([http://docs.aws.amazon.com/AmazonVPC/latest/UserGuide/VPC\\_Route\\_Tables.html](http://docs.aws.amazon.com/AmazonVPC/latest/UserGuide/VPC_Route_Tables.html)):

"Initially, the main route table (and every route table in a VPC) contains only a single route: a local route that enables communication within the VPC ... You can't modify the local route in a route table. Whenever you launch an instance in the VPC, the local route automatically covers that instance; you don't need to add the new instance to a route table."

---

Can you assign a private IP address to an EC2 instance?

313

- Yes (x)
- No

From the VPC FAQ (<https://aws.amazon.com/vpc/faqs/>):

### **Q. How do I assign private IP addresses to Amazon EC2 instances within a VPC?**

When you launch an Amazon EC2 instance within a VPC, you may optionally specify the primary private IP address for the instance. If you do not specify the primary private IP address, AWS automatically addresses it from the IP address range you assign to that subnet. You can assign secondary private IP addresses when you launch an instance, when you create an elastic network interface, or anytime after the instance has been launched or the interface has been created.

---

Can a VPC Customer Gateway be a software appliance?

314

- 
- Yes (x)
  - No

From the VPC documentation  
(<http://docs.aws.amazon.com/AmazonVPC/latest/NetworkAdminGuide/Introduction.html#CustomerGateway>):

"A *customer gateway* is the anchor on your side of that [VPN] connection. It can be a physical or software appliance. "

---

Can you change the size (CIDR block) of an existing VPC?

315

- 
- No (x)
  - Yes

From the VPC FAQ (<https://aws.amazon.com/vpc/faqs/>):

**"Q. Can I change a VPC's size?"**

No. Currently, to change the size of a VPC you must terminate your existing VPC and create a new one. "

Also from the VPC documentation ("[http://docs.aws.amazon.com/AmazonVPC/latest/UserGuide/VPC\\_Subnets.html](http://docs.aws.amazon.com/AmazonVPC/latest/UserGuide/VPC_Subnets.html)"):

"VPC Sizing

You can assign a single CIDR block to a VPC. The allowed block size is between a /28 net mask and /16 netmask. In other words, the VPC can contain from 16 to 65,536 IP addresses. You can't change the size of a VPC after you create it. If your VPC is too small to meet your needs, create a new, larger VPC, and then migrate your instances to the new VPC. To do this, create AMIs from your running instances, and then launch replacement instances in your new, larger VPC. You can then terminate your old instances, and delete your smaller VPC. For more information, see [Deleting Your VPC](#)."

---

What is the smallest possible size for a VPC?

316

- 
- /16 (65,536 addresses)
  - /20 (4,096 addresses)
  - /24 (256 addresses)
  - /28 (16 addresses) (x)

From the VPC documentation ([http://docs.aws.amazon.com/AmazonVPC/latest/UserGuide/VPC\\_Subnets.html#VPC\\_Sizing](http://docs.aws.amazon.com/AmazonVPC/latest/UserGuide/VPC_Subnets.html#VPC_Sizing)):

"VPC Sizing

You can assign a single CIDR block to a VPC. The allowed block size is between a /28 netmask and /16 netmask. In other words, the VPC can contain from 16 to 65,536 IP addresses. You can't change the size of a VPC after you create it. If your VPC is too small to meet your needs, create a new, larger VPC, and then migrate your instances to the new VPC. To do this, create AMIs from your running instances, and then launch replacement instances in your new, larger VPC. You can then terminate your old instances, and delete your smaller VPC. For more information, see [Deleting Your VPC](#)."

---

What is the largest possible size for a VPC?

317

- /14 (262,144 addresses)
- /16 (65,536 addresses) (x)
- /24 (256 addresses)
- /28 (16 addresses)

From the VPC documentation ([http://docs.aws.amazon.com/AmazonVPC/latest/UserGuide/VPC\\_Subnets.html#VPC\\_Sizing](http://docs.aws.amazon.com/AmazonVPC/latest/UserGuide/VPC_Subnets.html#VPC_Sizing)):

#### "VPC Sizing

You can assign a single CIDR block to a VPC. The allowed block size is between a /28 netmask and /16 netmask. In other words, the VPC can contain from 16 to 65,536 IP addresses. You can't change the size of a VPC after you create it. If your VPC is too small to meet your needs, create a new, larger VPC, and then migrate your instances to the new VPC. To do this, create AMIs from your running instances, and then launch replacement instances in your (new, larger VPC. You can then terminate your old instances, and delete your smaller VPC. For more information, see [Deleting Your VPC](#)."

Are VPC peering connections transitive?

318

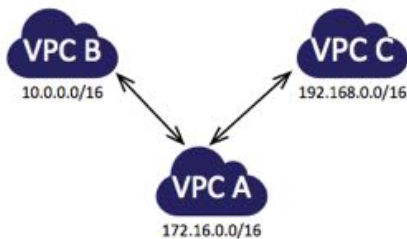
- Yes
- No (x)

From the VPC documentation (<http://docs.aws.amazon.com/AmazonVPC/latest/UserGuide/vpc-peering.html>):

>>>

A VPC peering connection is a one to one relationship between two VPCs. You can create multiple VPC peering connections for each VPC that you own, **but transitive peering relationships are not supported**: you will not have any peering relationship with VPCs that your VPC is not directly peered with.

The following diagram is an example of one VPC peered to two different VPCs. There are two VPC peering connections: VPC A is peered with both VPC B and VPC C. VPC B and VPC C are not peered, and you cannot use VPC A as a transit point for peering between VPC B and VPC C. If you want to enable routing of traffic between VPC B and VPC C, you must create a unique VPC peering connection between them.



<<<

How many VPC subnet addresses does AWS reserve for internal use?

319

- 2
- 3
- 4
- 5 (x)

From the VPC FAQ (<https://aws.amazon.com/vpc/faqs/>):

#### "Q. Can I use all the IP addresses that I assign to a subnet?

No. Amazon reserves the first four (4) IP addresses and the last one (1) IP address of every subnet for IP networking purposes. "

Can a single CGW be used with multiple VGWs?

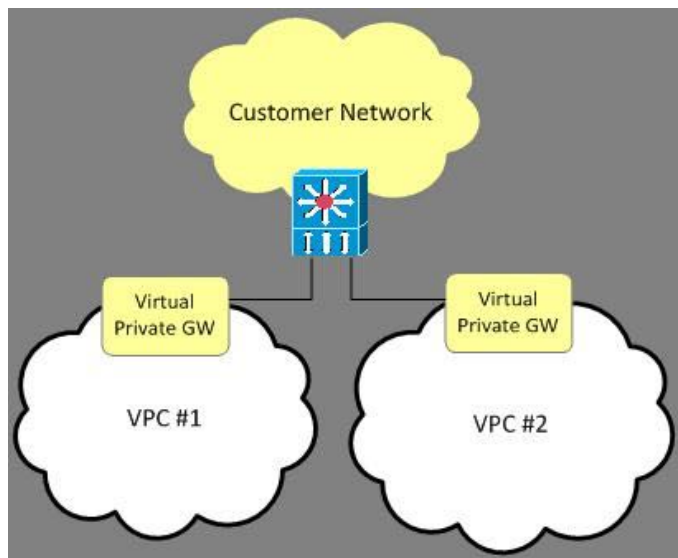
320

- Yes (x)
- No

This article: <http://aws.amazon.com/articles/5458758371599914> has a good explanation:

>>>

Amazon Virtual Private Cloud (Amazon VPC) provides customers with tremendous flexibility in how corporate networks can be connected to one or more VPCs. Customers can connect multiple customer gateways (CGW) to a single VPC virtual private gateway (VGW) or can connect a single router to multiple VPC VGWs. This document describes two approaches when connecting a single customer router to multiple VPCs, as shown in the diagram to the right.



<<<

Can a single VGW be used with multiple CGWs?

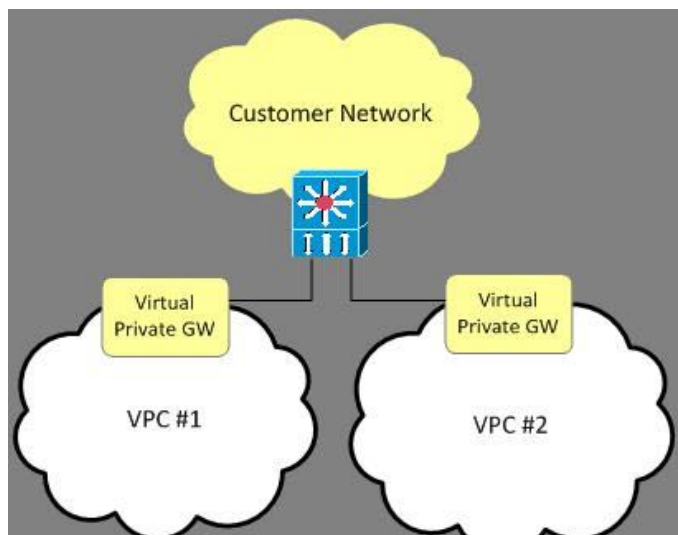
321

- Yes (x)
- No

This article: <http://aws.amazon.com/articles/5458758371599914> has a good explanation:

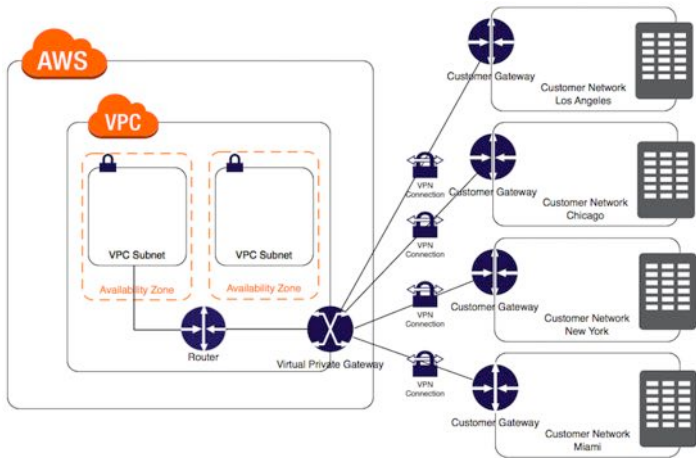
>>>

Amazon Virtual Private Cloud (Amazon VPC) provides customers with tremendous flexibility in how corporate networks can be connected to one or more VPCs. Customers can connect multiple customer gateways (CGW) to a single VPC virtual private gateway (VGW) or can connect a single router to multiple VPC VGWs. This document describes two approaches when connecting a single customer router to multiple VPCs, as shown in the diagram to the right.



<<<

Also from the VPC documentation ([http://docs.aws.amazon.com/AmazonVPC/latest/UserGuide/VPC\\_VPN.html#VPN](http://docs.aws.amazon.com/AmazonVPC/latest/UserGuide/VPC_VPN.html#VPN)):



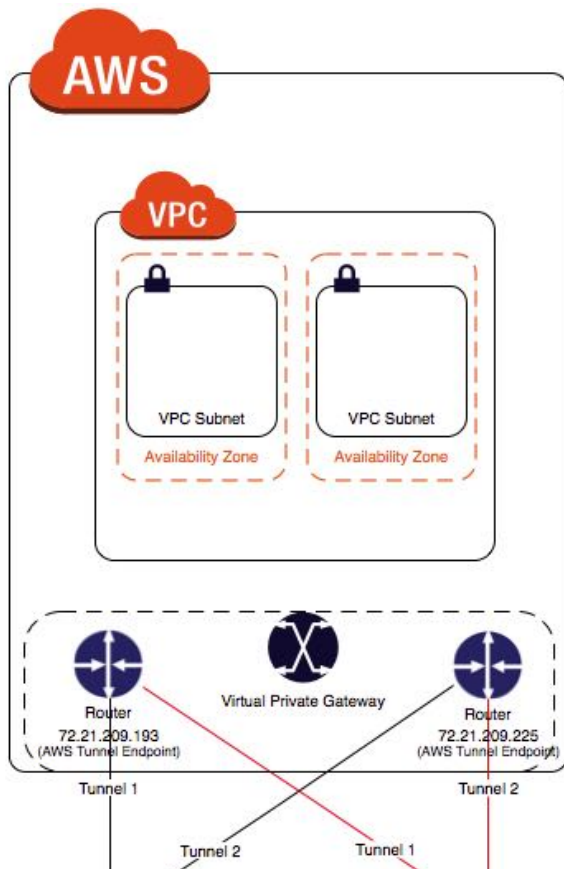
also

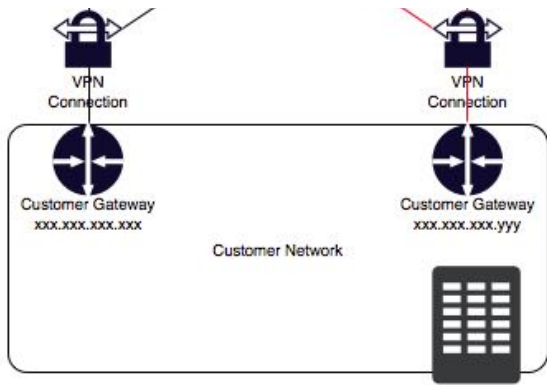
447781

### Using Redundant VPN Connections to Provide Failover

As described earlier, a VPN connection has two tunnels to help ensure connectivity in case one of the VPN connections becomes unavailable. To protect against a loss of connectivity in case your customer gateway becomes unavailable, you can set up a second VPN connection to your VPC and virtual private gateway by using a second customer gateway. By using redundant VPN connections and customer gateways, you can perform maintenance on one of your customer gateways while traffic continues to flow over the second customer gateway's VPN connection. To establish redundant VPN connections and customer gateways on your network, you need to set up a second VPN connection. The customer gateway IP address for the second VPN connection must be publicly accessible and can't be the same public IP address that you are using for the first VPN connection.

The following diagram shows the two tunnels of the VPN connection and two customer gateways.





What is VPC "route propagation"?

322

- A mechanism to automatically update route tables to route traffic to a VPG so that it can reach networks on the customer side of a VPN connection or Direct Connect link (x)
- A mechanism to automatically update route tables when establishing VPN peering links so that peered subnets can communicate
- A mechanism to propagate routes across VPN links using BGP

From the VPC User Guide ([http://docs.aws.amazon.com/AmazonVPC/latest/UserGuide/VPC\\_VPN.html#vpn-configure-routing](http://docs.aws.amazon.com/AmazonVPC/latest/UserGuide/VPC_VPN.html#vpn-configure-routing)):

"Enable Route Propagation in Your Route Table

To enable instances in your VPC to reach your customer gateway, you

must configure your route table to include the routes used by your VPN connection and point them to your virtual private gateway. You can enable route propagation for your route table to automatically propagate those routes to the table for you.

For static routing, the static IP prefixes that you specify for your VPN configuration are propagated to the route table after you've created the VPN connection. For dynamic routing, the BGP-advertised routes from your customer gateway are propagated to the route table when the status of the VPN connection is UP"

Do NACLs filter traffic between instances in the same subnet?

323

- Yes
- No (x)

From the VPC FAQ (<https://aws.amazon.com/vpc/faqs/>):

**"Q. What are the differences between security groups in a VPC and network ACLs in a VPC?"**

Security groups in a VPC specify which traffic is allowed to or from an Amazon EC2 instance. Network ACLs operate at the subnet level and evaluate traffic entering and exiting a subnet. Network ACLs can be used to set both Allow and Deny rules. **Network ACLs do not filter traffic between instances in the same subnet.** In addition, network ACLs perform stateless filtering while security groups perform stateful filtering. "

Can VPCs be peered across regions?

324

- Yes
- No (x)

From the VPC documentation (<http://docs.aws.amazon.com/AmazonVPC/latest/UserGuide/vpc-peering.html#vpc-peering-limitations>):

"You cannot create a VPC peering connection between VPCs in different regions."

Can VPCs be peered across AWS accounts?

325



- 
- Yes (x)
  - No

From the VPC documentation ():

"A VPC peering connection is a networking connection between two VPCs that enables you to route traffic between them using private IP addresses. Instances in either VPC can communicate with each other as if they are within the same network. You can create a VPC peering connection between your own VPCs, **or with a VPC in another AWS account** within a single region."

---

What are the default permissions for a VPC's default NACL?

326

- Allow all inbound
- Allow all outbound
- Allow all inbound and outbound (x)
- Deny all

From the VPC documentation ([http://docs.aws.amazon.com/AmazonVPC/latest/UserGuide/VPC\\_ACLS.html#default-network-acl](http://docs.aws.amazon.com/AmazonVPC/latest/UserGuide/VPC_ACLS.html#default-network-acl)):

#### "Default Network ACL

To help you understand what ACL rules look like, here's what the default network ACL looks like in its initial state. **It is configured to allow all traffic to flow in and out of each subnet.** Each network ACL includes a rule whose rule number is an asterisk. This rule ensures that if a packet doesn't match any of the other rules, it's denied. You can't modify or remove this rule."

also:

"Your VPC automatically comes with a modifiable default network ACL; by default, it allows all inbound and outbound traffic."

---

Can you remove the default "Deny All" rule at the bottom of a NACL's rule list?

327

- Yes
- No (x)

From the VPC documentation ([http://docs.aws.amazon.com/AmazonVPC/latest/UserGuide/VPC\\_ACLS.html#default-network-acl](http://docs.aws.amazon.com/AmazonVPC/latest/UserGuide/VPC_ACLS.html#default-network-acl)):

#### "Default Network ACL

To help you understand what ACL rules look like, here's what the default network ACL looks like in its initial state. It is configured to allow all traffic to flow in and out of each subnet. **Each network ACL includes a rule whose rule number is an asterisk. This rule ensures that if a packet doesn't match any of the other rules, it's denied. You can't modify or remove this rule.**"

---

Which of the following are true about NACL/subnet associations?

328

- Each subnet must be associated with a NACL (x)
- You must explicitly define and associate a custom NACL to each subnet
- VPCs come with a default NACL to which all subnets are initially associated (x)
- A NACL must be associated to a subnet
- A NACL can be associated to multiple subnets (x)
- A subnet can be associated to multiple NACLs

From the VPC documentation ([http://docs.aws.amazon.com/AmazonVPC/latest/UserGuide/VPC\\_ACLS.html#NetworkACL](http://docs.aws.amazon.com/AmazonVPC/latest/UserGuide/VPC_ACLS.html#NetworkACL)):

#### "Associating a Subnet with a Network ACL

To apply the rules of a network ACL to a particular subnet, you must associate the subnet with the network ACL. **You can associate a network ACL with multiple subnets;** however, **a subnet can be associated with only one network ACL.** Any subnet not associated with a particular ACL is associated with the **default network ACL** by default."

- Controls whether or not a VPC includes a DNS server for resolution of public host names
- Controls whether or not newly created hosts received DNS hostnames (x)
- Determines if the hostnames of newly created instances are visible outside of AWS

From the VPC documentation (<http://docs.aws.amazon.com/AmazonVPC/latest/UserGuide/vpc-dns.html>):

**"enableDnsHostnames** - Indicates whether the instances launched in the VPC get DNS hostnames. If this attribute is true, instances in the VPC get DNS hostnames; otherwise, they do not. If you want your instances to get DNS hostnames, you must also set the enableDnsSupport attribute to true."

- Controls whether or not a VPC includes a DNS server for resolution of public host names (x)
- Controls whether or not newly created hosts received DNS hostnames
- Determines if the hostnames of newly created instances are visible outside of AWS

From the VPC documentation (<http://docs.aws.amazon.com/AmazonVPC/latest/UserGuide/vpc-dns.html>):

**"enableDnsSupport** - Indicates whether the DNS resolution is supported for the VPC. If this attribute is false, the Amazon provided DNS service in the VPC that resolves public DNS hostnames to IP addresses is not enabled. If this attribute is true, queries to the Amazon provided DNS server at the 169.254.169.253 IP address, or the reserved IP address at the base of the VPC network range "plus two" will succeed. "

- VGW
- CGW (x)

From the VPC User Guide ():

"A *customer gateway* is a physical device or software application on your side of the VPN connection. When you create a VPN connection, the VPN tunnel comes up when traffic is generated from your side of the VPN connection. **The virtual private gateway is not the initiator; your customer gateway must initiate the tunnels.** If your VPN connection experiences a period of idle time (usually 10 seconds, depending on your configuration), the tunnel may go down. To prevent this, you can use a network monitoring tool to generate keepalive pings; for example, by using IP SLA."

- Yes (x)
- No

From the VPC User Guide ([http://docs.aws.amazon.com/AmazonVPC/latest/UserGuide/VPC\\_VPN.html#vpn-configure-routing](http://docs.aws.amazon.com/AmazonVPC/latest/UserGuide/VPC_VPN.html#vpn-configure-routing)):

"If you have not enabled route propagation for your route table, you must manually update the routes in your route table to reflect the updated static IP prefixes in your VPN connection. For more information, see [Enable Route Propagation in Your Route Table](#)."

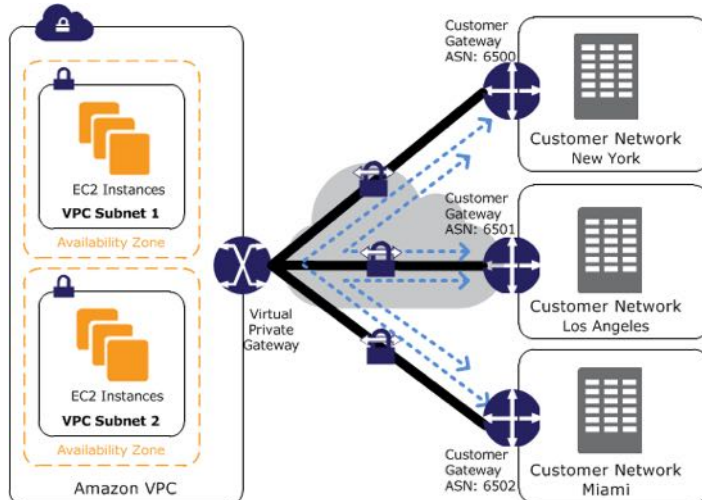
- Allows customers to route between their on-premises networks using a VPG (x)
- Provides redundant fault-tolerance for multiple VPN connections
- Enables routing between peered VPCs

From the VPC documentation ([http://docs.aws.amazon.com/AmazonVPC/latest/UserGuide/VPN\\_CloudHub.html](http://docs.aws.amazon.com/AmazonVPC/latest/UserGuide/VPN_CloudHub.html)):

#### "Providing Secure Communication Between Sites Using VPN CloudHub

If you have multiple VPN connections, you can provide secure communication between sites using the AWS VPN CloudHub. The VPN CloudHub operates on a simple hub-and-spoke model that you can use with or without a VPC. This design is suitable for customers with multiple branch offices and existing Internet connections who'd like to implement a convenient, potentially low-cost hub-and-spoke model for primary or backup connectivity between these remote offices.

The following diagram shows the VPN CloudHub architecture, with blue dashed lines indicating network traffic between remote sites being routed over their VPN connections.



To use the AWS VPN CloudHub, you must create a virtual private gateway with multiple customer gateways. You can use the same Border Gateway Protocol (BGP) Autonomous System Number (ASN) for each, or if you prefer, you can use a unique ASN for each. Customer gateways advertise the appropriate routes (BGP prefixes) over their VPN connections. These routing advertisements are received and re-advertised to each BGP peer, enabling each site to send data to and receive data from the other sites. The routes for each spoke must have unique ASNs and the sites must not have overlapping IP ranges. Each site can also send and receive data from the VPC as if they were using a standard VPN connection.

Sites that use AWS Direct Connect connections to the virtual private gateway can also be part of the AWS VPN CloudHub. For example, your corporate headquarters in New York can have an AWS Direct Connect connection to the VPC and your branch offices can use VPN connections to the VPC. The branch offices in Los Angeles and Miami can send and receive data with each other and with your corporate headquarters, all using the AWS VPN CloudHub."

What networking properties associated with an ENI are retained when an ENI is moved between instances?

334

- Private IP address (x)
- Public IP address (x)
- EIP (x)
- Security group(s) (x)
- MAC address (x)
- Source/destination flag (x)

From the VPC User Guide ([http://docs.aws.amazon.com/AmazonVPC/latest/UserGuide/VPC\\_ElasticNetworkInterfaces.html](http://docs.aws.amazon.com/AmazonVPC/latest/UserGuide/VPC_ElasticNetworkInterfaces.html)):

"An elastic network interface (ENI) is a virtual network interface that can include the following attributes:

- a primary private IP address
- one or more secondary private IP addresses
- one Elastic IP address per private IP address
- one public IP address, which can be auto-assigned to the network interface for eth0 when you launch an instance, but only when you create a network interface for eth0 instead of using an existing network interface
- one or more security groups
- a MAC address
- a source/destination check flag
- a description

You can create an ENI, attach it to an instance, detach it from an instance, and attach it to another instance. An ENI's attributes follow the ENI as it is attached or detached from an instance and reattached to another instance. When you move an ENI from one instance to another, network traffic is redirected to the new instance."

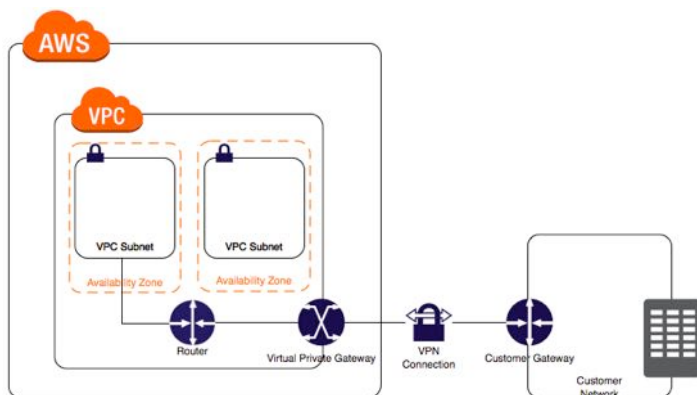
Is a VPG tied to a single AZ within a VPC or is it available to all AZs?

335

- Single AZ
- All AZs (x)

I can't find anything that describes this specifically, but it's implied throughout the VPC documentation. Just like a IGW, VPGs are attached to the VPC globally vs any particular AZ within the in the VPC.

For example, the VPC User Guide ([http://docs.aws.amazon.com/AmazonVPC/latest/UserGuide/VPC\\_VPN.html](http://docs.aws.amazon.com/AmazonVPC/latest/UserGuide/VPC_VPN.html)) has an image that implies this:



When VPCs are peered, is it possible to reference security groups across the connection (e.g., reference a security group in one VPC as a source for ingress traffic in a security group in the other VPC)?

336

- 
- Yes
  - No (x)

From the VPC User Guide (<http://docs.aws.amazon.com/AmazonVPC/latest/UserGuide/vpc-peering.html>):

**"You cannot reference a security group from the peer VPC as a source or destination for ingress or egress rules in your security group.** Instead, reference CIDR blocks of the peer VPC as the source or destination of your security group's ingress or egress rules."

## SQS

---

Does SQS support batch operations?

337

- 
- Yes (x)
  - No

From the SQS documentation (<http://docs.aws.amazon.com/AWSSimpleQueueService/latest/SQSDeveloperGuide/sqs-batch-api.html>):

"In the 2009-02-01 API version of Amazon SQS, only one action—ReceiveMessage—supports batch processing, i.e., processing more than one message with a single call. With the 2011-10-01 API version, Amazon SQS adds batch functionality for sending messages, deleting messages, and changing message visibility timeout values. To send up to ten messages at once, use the SendMessageBatch action. To delete up to ten messages with one API call, use the DeleteMessageBatch action. To change the visibility timeout value for up to ten messages, use the ChangeMessageVisibilityBatch action.

To use the new batch actions, you must use either the Query API or a Software Development Kit (SDK) that supports the new batch actions. Check your specific SDK's documentation to see whether it supports the new Amazon SQS batch actions. The Amazon SQS console does not currently support the batch API actions."

---

Does SQS provide FIFO access to queues?

338

- 
- Yes
  - No (x)

From the SQS FAQ (<https://aws.amazon.com/sqs/faqs/>):

**"Q: Does Amazon SQS provide first-in-first-out (FIFO) access to messages?**

No, Amazon SQS does not guarantee FIFO access to messages in Amazon SQS queues, mainly because of the distributed nature of the Amazon SQS. If you require specific message ordering, you should design your application to handle it."

---

Is it possible to subscribe multiple SQS queues to the same SNS topic?

339

- Yes (x)
- No

This is a common use case.

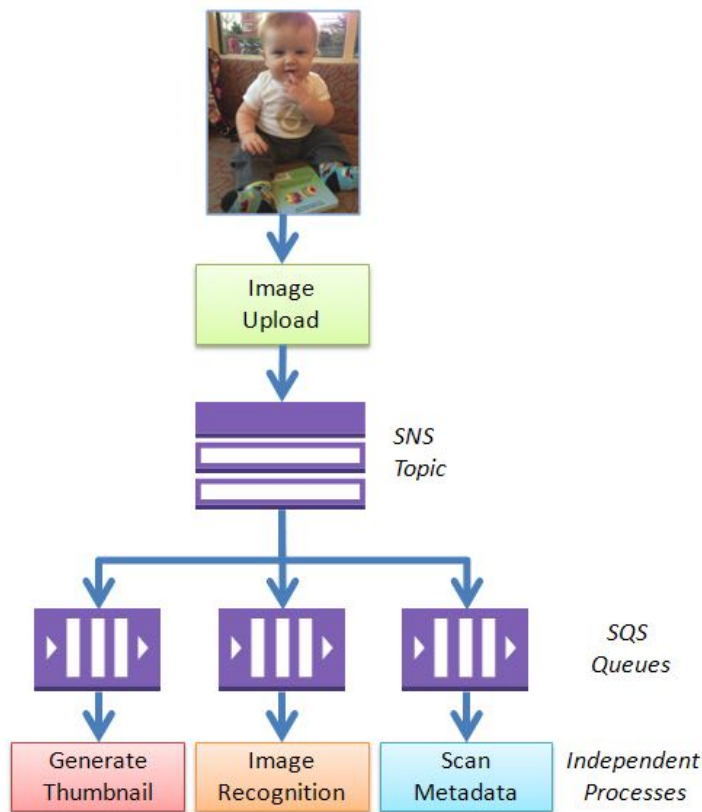
From the SQS documentation (<http://docs.aws.amazon.com/AWSSimpleQueueService/latest/SQSDeveloperGuide/sqssubscribe.html>):

"You can now subscribe an Amazon SQS queue to an Amazon SNS topic using the AWS Management Console for Amazon SQS, which simplifies the process. For example, you can choose from the list of available topics for the selected queue. Amazon SQS then manages the subscription of the queue to the topic and the addition of the necessary permissions. When a message is published to the topic, Amazon SNS sends an Amazon SQS message to the subscribed queue. For more information about Amazon SNS, see [Get Started with Amazon SNS](#). For more information about Amazon SQS, see [Get Started with Amazon SQS](#)"

...and this specifically about the "fanout pattern" on an AWS blog (<https://aws.amazon.com/blogs/aws/queues-and-notifications-now-best-friends/>):

>>>

One common design pattern is called "fanout." In this pattern, a message published to an SNS topic is distributed to a number of SQS queues in parallel. By using this pattern, you can build applications that take advantage parallel, asynchronous processing. For example, you could publish a message to a topic every time a new image is uploaded. Independent processes, each reading from a separate SQS queue, could generate thumbnails, perform image recognition, and store metadata about the image:



<<<

Does the maximum size of an SQS message include its metadata (message attributes)?

- 
- Yes (x)
  - No

From the SQS documentation  
(<http://docs.aws.amazon.com/AWSSimpleQueueService/latest/SQSDeveloperGuide/SQSMessageAttributes.html>):

"Each message attribute consists of the following items:

- **Name** – The message attribute name can contain the following characters: A-Z, a-z, 0-9, underscore(\_), hyphen(-), and period (.). The name must not start or end with a period, and it should not have successive periods. The name is case sensitive and must be unique among all attribute names for the message. The name can be up to 256 characters long. The name cannot start with "AWS." or "Amazon." (or any variations in casing) because these prefixes are reserved for use by Amazon Web Services.
- **Type** – The supported message attribute data types are String, Number, and Binary. You can also provide custom information on the type. The data type has the same restrictions on the content as the message body. The data type is case sensitive, and it can be up to 256 bytes long. For more information, see the Message Attribute Data Types and Validation section.
- **Value** – The user-specified message attribute value. For string data types, the value attribute has the same restrictions on the content as the message body. For more information, see `SendMessage`.

Name, type, and value must not be empty or null. In addition, the message body should not be empty or null. **All parts of the message attribute, including name, type, and value, are included in the message size restriction, which is currently 256 KB (262,144 bytes).**"

---

Is it possible to purge an SQS queue without deleting the queue itself?

341

- 
- Yes (x)
  - No

From the SQS FAQ (<https://aws.amazon.com/sqs/faqs/>):

**"Q: Can I delete all the messages in a queue without deleting the queue itself?"**

Yes, you can delete all the messages in an SQS queue using the `PurgeQueue` action. When you purge a queue, all the messages previously sent to the queue will be deleted. Since your queue and its attributes will remain, there is no need to reconfigure the queue to continue using it. If you only need to delete specific messages, you can use the `DeleteMessage` or `DeleteMessageBatch` actions."

---

Is it possible to receive a message from an SQS queue more than once?

342

- 
- Yes (x)
  - No

From the SQS FAQ (<https://aws.amazon.com/sqs/faqs/>):

**"Q: How many times will I receive each message?"**

Amazon SQS is engineered to provide "at least once" delivery of all messages in its queues. Although most of the time each message will be delivered to your application exactly once, you should design your system so that processing a message more than once does not create any errors or inconsistencies. "

---

Is it possible that clients might see a message on an SQS queue even after the message has been deleted?

343

- 
- Yes (x)
  - No

From the SQS FAQ (<https://aws.amazon.com/sqs/faqs/>):

**"Q: Can a deleted message be received again?"**

Yes, under rare circumstances you might receive a previously deleted message again. This can occur in the rare situation in which a DeleteMessage operation doesn't delete all copies of a message because one of the servers in the distributed Amazon SQS system isn't available at the time of the deletion. That message copy can then be delivered again. You should design your application so that no errors or inconsistencies occur if you receive a deleted message again. "

---

What is the maximum amount of time that a message can remain in an SQS queue?

344

- 
- 2 days
  - 1 week
  - 2 weeks (x)
  - 4 weeks

From the SQS FAQ (<https://aws.amazon.com/sqs/faqs/>):

**"Q: How long can I keep my messages in Amazon SQS queues?"**

The SQS message retention period is configurable and can be set anywhere from 1 minute to 2 weeks. The default is 4 days and once the message retention limit is reached your messages will be automatically deleted. The option for longer message retention provides greater flexibility to allow for longer intervals between message production and consumption. "

---

What is the maximum SQS long polling timeout value?

345

- 
- 20 seconds (x)
  - 5 seconds
  - 10 seconds
  - 60 seconds

From the SQS documentation (<http://docs.aws.amazon.com/AWSSimpleQueueService/latest/SQSDeveloperGuide/sqs-long-polling.html>):

"For most use cases when using long polling, you should set the timeout value to the maximum of 20 seconds. If the 20 second maximum does not work for your application, you can choose a shorter long poll timeout, down to as low as 1 second."

---

What is the maximum delay period for an SQS delay queue?

346

- 
- 1 minute
  - 5 minutes
  - 15 minutes (x)
  - 30 minutes

From the SQS documentation (<http://docs.aws.amazon.com/AWSSimpleQueueService/latest/SQSDeveloperGuide/sqs-delay-queues.html>):

"Delay queues allow you to postpone the delivery of new messages in a queue for a specific number of seconds. If you create a delay queue, any message that you send to that queue will be invisible to consumers for the duration of the delay period. You can use CreateQueue to create a delay queue by setting the DelaySeconds attribute to any value between 0 and 900 (15 minutes). You can also turn an existing queue into a delay queue by using SetQueueAttribute to set the queue's DelaySeconds attribute."



- ReceiveMessage requests don't return until a message appears on the queue, reducing the number of required API calls (x)
- All of the queue servers are queried to service the request, so the ReceiveMessage response always includes all available messages (x)
- Allows for polls longer than 120 seconds

From the SQS Developer Guide (<http://docs.aws.amazon.com/AWSSimpleQueueService/latest/SQSDeveloperGuide/sqs-long-polling.html>):

"One benefit of long polling with Amazon SQS is the reduction of the number of empty responses, when there are no messages available to return, in reply to a ReceiveMessage request sent to an Amazon SQS queue. **Long polling allows the Amazon SQS service to wait until a message is available in the queue before sending a response.** So unless the connection times out, the response to the ReceiveMessage request will contain at least one of the available messages (if any) and up to the maximum number requested in the ReceiveMessage call.

**Another benefit is helping to eliminate false empty responses, where messages are available in the queue but are not included in the response.** This happens when Amazon SQS uses short (standard) polling, the default behavior, where only a subset of the servers (based on a weighted random distribution) are queried to see if any messages are available to include in the response. On the other hand, when long polling is enabled, Amazon SQS queries all of the servers.

Reducing the number of empty responses and false empty responses also helps reduce your cost of using Amazon SQS."

Note that the feature is triggered via the WaitTimeSeconds parameter:

**"Short polling occurs when the WaitTimeSeconds parameter of a ReceiveMessage call is set to 0.** This happens in one of two ways – either the ReceiveMessage call sets WaitTimeSeconds to 0, or the ReceiveMessage call doesn't set WaitTimeSeconds and the queue attribute ReceiveMessageWaitTimeSeconds is 0."

- One of a pre-defined set of properties that clients can query to filter messages on a queue
- Metadata about a message that clients can use to decide how to best process the message (x)
- Message properties such as size, type, time on queue, and others

From the SQS Developer Guide (<http://docs.aws.amazon.com/AWSSimpleQueueService/latest/SQSDeveloperGuide/SQSMessageAttributes.html>):

"Amazon Simple Queue Service (Amazon SQS) provides support for *message attributes*. Message attributes allow you to provide structured metadata items (such as timestamps, geospatial data, signatures, and identifiers) about the message. Message attributes are optional and separate from, but sent along with, the message body. This information can be used by the receiver of the message to help decide how to handle the message without having to first process the message body. Each message can have up to 10 attributes."

## DynamoDB

- Yes
- No (x)

From the CreateTable API documentation ([http://docs.aws.amazon.com/amazondynamodb/latest/APIReference/API\\_CreateTable.html](http://docs.aws.amazon.com/amazondynamodb/latest/APIReference/API_CreateTable.html)):

"In an AWS account, table names must be unique within each region. That is, you can have two tables with same name if you create the tables in different regions."

- 
- None
  - Up to 30 minutes of unused read and write capacity
  - Up to 15 minutes of unused read and write capacity
  - Up to 5 minutes of unused read and write capacity (x)

From the Amazon S3 documentation

(<http://docs.aws.amazon.com/amazondynamodb/latest/developerguide/GuidelinesForTables.html#GuidelinesForTables.Bursting>):

"DynamoDB provides some flexibility in the per-partition throughput provisioning: When you are not fully utilizing a partition's throughput, DynamoDB reserves a portion of your unused capacity for later "bursts" of throughput usage. DynamoDB currently reserves up to 5 minutes (300 seconds) of unused read and write capacity. During an occasional burst of read or write activity, this reserved throughput can be consumed very quickly — even faster than the per-second provisioned throughput capacity that you've defined for your table. However, do not design your application so that it depends on burst capacity being available at all times: DynamoDB can and does use burst capacity for background maintenance and other tasks without prior notice."

---

Does DynamoDB background maintenance consume burst capacity?

351

- Yes (x)
- No

From the Amazon S3 documentation

(<http://docs.aws.amazon.com/amazondynamodb/latest/developerguide/GuidelinesForTables.html#GuidelinesForTables.Bursting>):

"DynamoDB provides some flexibility in the per-partition throughput provisioning: When you are not fully utilizing a partition's throughput, DynamoDB reserves a portion of your unused capacity for later "bursts" of throughput usage. DynamoDB currently reserves up to 5 minutes (300 seconds) of unused read and write capacity. During an occasional burst of read or write activity, this reserved throughput can be consumed very quickly — even faster than the per-second provisioned throughput capacity that you've defined for your table. **However, do not design your application so that it depends on burst capacity being available at all times: DynamoDB can and does use burst capacity for background maintenance and other tasks without prior notice.**"

---

When uploading large data sets into a DynamoDB table with a hash and range key, which is more performant?

352

- Writing all objects with the same hash key one after the other (x)
- Distributing writes across the hash keys

From the DynamoDB documentation

(<http://docs.aws.amazon.com/amazondynamodb/latest/developerguide/GuidelinesForTables.html#GuidelinesForTables.DataUpload>):

"There are times when you load data from other data sources into DynamoDB. Typically, DynamoDB partitions your table data on multiple servers. When uploading data to a table, you get better performance if you upload data to all the allocated servers simultaneously. For example, suppose you want to upload user messages to a DynamoDB table. You might design a table that uses a hash and range type primary key in which UserID is the hash attribute and the MessageID is the range attribute. When uploading data from your source, you might tend to read all message items for a specific user and upload these items to DynamoDB as shown in the sequence in the following table.

UserID	MessageID
--------	-----------

U1	1
U1	2
U1	...
U1	... up to 100
U2	1
U2	2
U2	...
U2	... up to 200

The problem in this case is that you are not distributing your write requests to DynamoDB across your hash key values. You are taking one hash key at a time and uploading all its items before going to the next hash key items. Behind the scenes, DynamoDB is partitioning the data in your tables across multiple servers. To fully utilize all of the throughput capacity that has been provisioned for your tables, you need to distribute your workload across your hash keys. In this case, by directing an uneven amount of upload work toward items all with the same hash key, you may not be able to fully utilize all of the resources DynamoDB has provisioned for your table. You can distribute your upload work by uploading one item from each hash key first. Then you repeat the pattern for the next set of range keys for all the items until you upload all the data as shown in the example upload sequence in the following table:

UserID	MessageID
--------	-----------

U1	1
U2	1
U3	1
...	...
U1	2
U2	2
U3	2
...	...

Every upload in this sequence uses a different hash key, keeping more DynamoDB servers busy simultaneously and improving your throughput performance."

Which of the following are good, performant hash keys for a DynamoDB table?

353

- User ID, where the application has many users. (x)
- Status code, where there are only a few possible status codes.
- Item creation date, rounded to the nearest time period (e.g. day, hour, minute)
- Device ID, where each device accesses data at relatively similar intervals (x)
- Device ID, where even if there are a lot of devices being tracked, one is by far more popular than all the others.

## "Design For Uniform Data Access Across Items In Your Tables

The optimal usage of a table's provisioned throughput depends on these factors:

- The primary key selection.
- The workload patterns on individual items.

The primary key uniquely identifies each item in a table. The primary key can be defined as one attribute (hash type) or two attributes (hash and range type).

When it stores data, DynamoDB divides a table's items into multiple partitions, and distributes the data primarily based upon the hash key element. The provisioned throughput associated with a table is also divided evenly among the partitions, with no sharing of provisioned throughput across partitions.

$$\text{Total Provisioned Throughput} / \text{Partitions} = \text{Throughput Per Partition}$$

Consequently, to achieve the full amount of request throughput you have provisioned for a table, keep your workload spread evenly across the hash key values. Distributing requests across hash key values distributes the requests across partitions.

For example, if a table has a very small number of heavily accessed hash key elements, possibly even a single very heavily used hash key element, request traffic is concentrated on a small number of partitions – potentially only one partition. If the workload is heavily unbalanced, meaning that it is disproportionately focused on one or a few partitions, the requests will not achieve the overall provisioned throughput level. To get the most out of DynamoDB throughput, create tables where the hash key element has a large number of distinct values, and values are requested fairly uniformly, as randomly as possible.

This does not mean that you must access all of the hash keys to achieve your throughput level; nor does it mean that the percentage of accessed hash keys needs to be high. However, do be aware that when your workload accesses more distinct hash keys, those requests will be spread out across the partitioned space in a manner that better utilizes your allocated throughput level. In general, you will utilize your throughput more efficiently as the ratio of hash keys accessed to total hash keys in a table grows.

### Choosing a Hash Key

The following table compares some common hash key schemas for provisioned throughput efficiency:

- **User ID, where the application has many users. (Good)**
- **Status code, where there are only a few possible status codes. (Bad)**
- **Item creation date, rounded to the nearest time period (e.g. day, hour, minute) (Bad)**
- **Device ID, where each device accesses data at relatively similar intervals (Good)**
- **Device ID, where even if there are a lot of devices being tracked, one is by far more popular than all the others. (Bad)**

If a single table has only a very small number of hash key values, consider distributing your write operations across more distinct hash values. In other words, structure the primary key elements to avoid one "hot" (heavily requested) hash key value that slows overall performance.

For example, consider a table with a hash and range type primary key. The hash key represents the item's creation date, rounded to the nearest day. The range key is an item identifier. On a given day, say 2014-07-09, all of the new items will be written to that same hash key value.

If the table will fit entirely into a single partition (taking into consideration growth of your data over time), and if your application's read and write throughput requirements do not exceed the read and write capabilities of a single partition, then your application should not encounter any unexpected throttling as a result of partitioning.

However, if you anticipate scaling beyond a single partition, then you should architect your application so that it can use more of the table's full provisioned throughput.

### Randomizing Across Multiple Hash Key Values

One way to increase the write throughput of this application would be to randomize the writes across multiple hash key values. Choose a random number from a fixed set (for example, 1 to 200) and concatenate it as a suffix to the date. This will yield hash key values such as 2014-07-09.1, 2014-07-09.2 and so on through 2014-07-09.200. Because you are randomizing the hash key, the writes to the table on each day are spread evenly across all of the hash key values; this will yield better parallelism and higher overall throughput.

To read all of the items for a given day, you would need to obtain all of the items for each suffix. For example, you would first issue a Query request for the hash key 2014-07-09.1, then another Query for 2014-07-09.2, and so on through 2014-07-09.200. Finally, your application would need to merge the results from all of the Query requests.

### Using a Calculated Value

A randomizing strategy can greatly improve write throughput; however, it is difficult to read a specific item because you don't know which suffix value was used when writing the item. To make it easier to read individual items, you can use a different strategy: Instead of using a random number to distribute the items among partitions, use a number that you are able to calculate based upon something that's intrinsic to the item.

Continuing with our example, suppose that each item has an OrderId. Before your application writes the item to the table, it can calculate a hash key suffix based upon the order ID. The calculation should result in a number between 1 and 200 that is fairly evenly distributed given any set of names (or user IDs.)

A simple calculation would suffice, such as the product of the ASCII values for the characters in the order ID, modulo 200 + 1. The hash

A simple calculation would suffice, such as the product of the ASCII values for the characters in the order ID, modulo 200 + 1. The hash key value would then be the date concatenated with the calculation result as a suffix. With this strategy, the writes are spread evenly across the hash keys, and thus across the partitions. You can easily perform a GetItem operation on a particular item, because you can calculate the hash key you need when you want to retrieve a specific OrderId value.

To read all of the items for a given day, you would still need to Query each of the 2014-07-09.N keys (where N is 1 to 200), and your application would need to merge all of the results. However, you will avoid having a single "hot" hash key taking all of the workload."

What is the limit on the size of a DynamoDB table?

354

- 10 MB
- 1 GB
- 5 GB
- No Limit (x)

See <http://docs.aws.amazon.com/amazondynamodb/latest/developerguide/Limits.html>

What is a DynamoDB strongly consistent read capacity unit?

355

- 1 read/second of items up to 1 KB per second
- 2 reads/second of items up to 1 KB per second
- 3 reads/second of items up to 1 KB per second
- 4 reads/second of items up to 1 KB per second (x)

From the DynamoDB Developer Guide

(<http://docs.aws.amazon.com/amazondynamodb/latest/developerguide/ProvisionedThroughputIntro.html>):

**"A unit of *read capacity* represents one strongly consistent read per second (or two eventually consistent reads per second) for items as large as 4 KB. A unit of *write capacity* represents one write per second for items as large as 1 KB."**

What is a DynamoDB write capacity unit?

356

- 1 write/second of items up to 1 KB per second (x)
- 2 writes/second of items up to 1 KB per second
- 3 writes/second of items up to 1 KB per second
- 4 writes/second of items up to 1 KB per second

From the DynamoDB Developer Guide

(<http://docs.aws.amazon.com/amazondynamodb/latest/developerguide/ProvisionedThroughputIntro.html>):

"A unit of *read capacity* represents one strongly consistent read per second (or two eventually consistent reads per second) for items as large as 4 KB. **A unit of *write capacity* represents one write per second for items as large as 1 KB.**"

What is the maximum size of a DynamoDB item?

357

- 
- 4 KB
  - 40 KB
  - 400 KB (x)
  - 1000 KB

See <http://docs.aws.amazon.com/amazondynamodb/latest/developerguide/Limits.html>

"Cannot exceed 400 KB which includes both attribute name binary length (UTF-8 length) and attribute value lengths (again binary length). The attribute name counts towards the size limit. For example, consider an item with two attributes: one attribute named "shirt-color" with value "R" and another attribute named "shirt-size" with value "M". The total size of that item is 23 bytes.

For attribute values that are of type binary, the application must encode the data in base64 format before sending it to DynamoDB. Upon receipt of the data, DynamoDB decodes it into an unsigned byte array and uses that as the length of the attribute.

These limits apply to items stored in tables, and also to items in secondary indexes.

For each local secondary index on a table, there is a 400 KB limit on the total size of the following:

- The size of an item's data in the table.
- The size of the local secondary index entry corresponding to that item, including its key values and projected attributes."

---

What is the maximum size of a DynamoDB hash key attribute value?

358

- 
- 256 bytes
  - 512 bytes
  - 1024 bytes
  - 2048 bytes (x)

See <http://docs.aws.amazon.com/amazondynamodb/latest/developerguide/Limits.html>

---

What is the maximum size of a DynamoDB range key attribute value?

359

- 
- 256 bytes
  - 512 bytes
  - 1024 bytes (x)
  - 2048 bytes

See <http://docs.aws.amazon.com/amazondynamodb/latest/developerguide/Limits.html>

---

What is the maximum size of the result set returned by a DynamoDB Query API call?

360

- 
- 256 KB
  - 512 KB
  - 1 MB (x)
  - 5 MB

See <http://docs.aws.amazon.com/amazondynamodb/latest/developerguide/Limits.html>

"The result set is limited to 1 MB per API call. You can use the LastEvaluatedKey from the query response to retrieve more results."

---

What is the maximum size of the scanned data set returned by a DynamoDB Scan API call?

361

- 
- 256 KB
  - 512 KB
  - 1 MB (x)
  - 5 MB

See <http://docs.aws.amazon.com/amazondynamodb/latest/developerguide/Limits.html>

"The maximum size of the scanned data set is 1 MB per API call. You can use the LastEvaluatedKey from the scan response to retrieve more results."

---

Is a change in the provisioned capacity of a DynamoDB propagated to its cross-region replica?

362

- Yes
- No (x)

From the DynamoDB FAQ (<https://aws.amazon.com/dynamodb/faqs>):

**"Q: If I change provisioned capacity on my master table, does the provisioned capacity on my replica table also update?"**

After the replication has been created, any changes to the provisioned capacity on the master table will not result in an update in throughput capacity on the replica table."

The replica table is an independent table that's kept up to date by user-deployed AWS-provided software components (in particular the "DynamoDB Connector") that asynchronously replicates writes to the replica. See this link for all the details: <http://docs.aws.amazon.com/amazondynamodb/latest/developerguide/Streams.CrossRegionRepl.html>

---

When create a replica of a DynamoDB table, does the replica have the master's indexes?

363

- Yes
- No (x)

From the DynamoDB FAQ (<https://aws.amazon.com/dynamodb/faqs>):

**"Q: Will my replica tables have the same indexes as the master table?"**

If you choose to create the replica table from the replication application, the secondary indexes on the master table will NOT be automatically created on the replica table. The replication application will not propagate changes made on secondary indices on the master table to replica tables. You will have to add/update/delete indexes on each of the replica tables through the AWS Management Console as you would with regular DynamoDB tables."

---

What's the best way to copy a DynamoDB table across regions?

364

- Use one of the many open-source libraries to bulk copy data across regions
- Use Data Pipeline (x)
- Use the CrossRegionCopy API
- Create a snapshot and copy it across regions using S3

From the Data Pipeline documentation (<http://docs.aws.amazon.com/datapipeline/latest/DeveloperGuide/dp-template-crossregionddbcopy.html>):

"The DynamoDB Cross Regional Table Copy AWS Data Pipeline template configures periodic movement of data between DynamoDB tables across regions or to a different table within the same region. This feature is useful in the following scenarios:

- Disaster recovery in the case of data loss or region failure
- Moving DynamoDB data across regions to support applications in those regions
- Performing full or incremental DynamoDB data backups"

- No impact; DynamoDB doesn't charge for LSI updates
- Additional provisioned throughput is consumed for reads and writes that involve the LSI (x)
- Depends on LSI configuration

From the DynamoDB Developer Guide

(<http://docs.aws.amazon.com/amazondynamodb/latest/developerguide/LSI.html#LSI.ThroughputConsiderations>):

"When an item in a table is added, updated, or deleted, updating the local secondary indexes will consume provisioned write capacity units for the table. The total provisioned throughput cost for a write is the sum of write capacity units consumed by writing to the table and those consumed by updating the local secondary indexes."

Also from another page in the Developer Guide

(<http://docs.aws.amazon.com/amazondynamodb/latest/developerguide/SecondaryIndexes.html>):

"Queries or scans on a local secondary index consume read capacity units from the table. When you write to a table, its local secondary indexes are also updated; these updates consume write capacity units from the table."

- None (x)
- Additional provisioned throughput is consumed for reads and writes that involve the GSI
- Depends on GSI configuration

From the DynamoDB Developer Guide

(<http://docs.aws.amazon.com/amazondynamodb/latest/developerguide/GSI.html#GSI.ThroughputConsiderations>):

"When you create a global secondary index, you must specify read and write capacity units for the expected workload on that index. The provisioned throughput settings of a global secondary index are separate from those of its parent table. A Query operation on a global secondary index consumes read capacity units from the index, not the table. When you put, update or delete items in a table, the global secondary indexes on that table are also updated; these index updates consume write capacity units from the index, not from the table."

Also from another page in the Developer Guide

(<http://docs.aws.amazon.com/amazondynamodb/latest/developerguide/SecondaryIndexes.html>):

"Every global secondary index has its own provisioned throughput settings for read and write activity. Queries or scans on a global secondary index consume capacity units from the index, not from the table. The same holds true for global secondary index updates due to table writes."

- Cumulative size (x)
- Number of items retrieved

From the DynamoDB Developer Guide

(<http://docs.aws.amazon.com/amazondynamodb/latest/developerguide/ProvisionedThroughputIntro.html>):

"You can use the Query and Scan operations in DynamoDB to retrieve multiple consecutive items from a table or an index in a single request. With these operations, DynamoDB uses the cumulative size of the processed items to calculate provisioned throughput. For example, if a Query operation retrieves 100 items that are 1 KB each, the read capacity calculation is *not*  $(100 \times 4 \text{ KB}) = 100$  read capacity units, as if those items were retrieved individually using GetItem or BatchGetItem. Instead, the total would be only 25 read capacity units  $((100 \times 1024 \text{ bytes}) = 100 \text{ KB, which is then divided by } 4 \text{ KB})$ ."



- The operation is “rounded up” and an additional full capacity unit is consumed (x)
- The operation is “rounded down” and only one capacity unit is consumed
- A fractional capacity unit is consumed

From the DynamoDB Developer Guide

(<http://docs.aws.amazon.com/amazondynamodb/latest/developerguide/ProvisionedThroughputIntro.html>):

For read capacity units:

"If your items are smaller than 4 KB in size, each read capacity unit will give you one strongly consistent read per second, or two eventually consistent reads per second. You cannot group multiple items in a single read operation, even if the items together are 4 KB or smaller. For example, if your items are 3 KB and you want to read 80 items per second from your table, then you need to provision  $80 \text{ (reads per second)} \times 1 \text{ (3 KB / 4 KB = 0.75, rounded up to the next whole number)} = 80$  read capacity units for strong consistency. For eventual consistency, you need to provision only 40 read capacity units.

**If your items are larger than 4 KB, you will need to round up the item size to the next 4 KB boundary.** For example, if your items are 6 KB and you want to do 100 strongly consistent reads per second, you need to provision  $100 \text{ (reads per second)} \times 2 \text{ (6 KB / 4 KB = 1.5, rounded up to the next whole number)} = 200$  read capacity units."

For write capacity units:

"If your items are smaller than 1 KB in size, then each write capacity unit will give you 1 write per second. You cannot group multiple items in a single write operation, even if the items together are 1 KB or smaller. For example, if your items are 512 bytes and you want to write 100 items per second to your table, then you would need to provision 100 write capacity units.

**If your items are larger than 1 KB in size, you will need to round the item size up to the next 1 KB boundary.** For example, if your items are 1.5 KB and you want to do 10 writes per second, then you would need to provision  $10 \text{ (writes per second)} \times 2 \text{ (1.5 KB rounded up to the next whole number)} = 20$  write capacity units."

- Based on the hash of its primary key (x)
- By using the primary key value itself as a hash key
- Based on the hash of the entire item

From the DynamoDB Developer Guide

(<http://docs.aws.amazon.com/amazondynamodb/latest/developerguide/GuidelinesForTables.html>):

"When it stores data, DynamoDB divides a table's items into multiple partitions, and distributes the data primarily based upon the hash attribute value."

Also, see this thread (<http://stackoverflow.com/questions/25142257/are-dynamodb-uuid-hashkeys-better-than-sequentially-generated-ones>) and answered by David Yanacek, a developer on the DynamoDB team:

“100444” and “100445” are not any more likely to be in the same partition than a completely different number, like “12345” for example. Think of a DynamoDB table as a big hash table, where the hash key of the table is the key into the hash table. **The underlying hash table is organized by the hash of the key, not by the key itself.** You'll find that numbers and strings (UUIDs) both distribute fine in DynamoDB in terms of their distribution across partitions.

UUIDs are useful in DynamoDB because sequential numbers are difficult to generate in a scalable way for primary keys. Random numbers work well for primary keys, but sequential values are hard to generate without gaps and in a way that scales to the level of throughput that you can provision in a DynamoDB table. When you insert new items into a DynamoDB table, you can use conditional writes to ensure an item doesn't already exist with that primary key value.

How is the provisioned throughput of a DynamoDB table allocated across partitions?

370

- Evenly distributed with no sharing between partitions (x)
- Dynamically allocated based on average partition load
- Dynamically allocated based on average partition size
- User configurable

From the DynamoDB Developer Guide

(<http://docs.aws.amazon.com/amazondynamodb/latest/developerguide/GuidelinesForTables.html>):

**“The provisioned throughput associated with a table is also divided evenly among the partitions, with no sharing of provisioned throughput across partitions.**

Total Provisioned Throughput / Partitions = Throughput Per Partition

Consequently, to achieve the full amount of request throughput you have provisioned for a table, keep your workload spread evenly across the hash attribute values. Distributing requests across hash attribute values distributes the requests across partitions.

For example, if a table has a very small number of heavily accessed hash attribute values, possibly even a single very heavily used hash attribute value, request traffic is concentrated on a small number of partitions – potentially only one partition. If the workload is heavily unbalanced, meaning that it is disproportionately focused on one or a few partitions, the requests will not achieve the overall provisioned throughput level. To get the most out of DynamoDB throughput, create tables where the hash attribute has a large number of distinct values, and values are requested fairly uniformly, as randomly as possible.

This does not mean that you must access all of the hash attribute values to achieve your throughput level; nor does it mean that the percentage of accessed hash attribute values needs to be high. However, be aware that when your workload accesses more distinct hash attribute values, those requests will be spread out across the partitioned space in a manner that better utilizes your allocated throughput level. In general, you will utilize your throughput more efficiently as the ratio of hash attribute values accessed to the total number of hash attribute values in a table grows.”

Which of the following are examples of “good” DynamoDB hash attributes that optimize for provisioned throughput efficiency?

371

- User ID, where the application has many users (x)
- Status code, where there are only a few possible status codes
- Item creation date, rounded to the nearest time period (e.g. day, hour, minute)
- Device ID, where each device accesses data at relatively similar intervals (x)
- Device ID, where even if there are a lot of devices being tracked, one is by far more popular than all the others

[Note: To me, whether or not the "Item creation date" acceptable depends on context. If there are a consistent and relatively small number of items created per minute, and data is accessed evenly across time, it might be fine. If more recent data is accessed more frequently there would be an issue since the read distribution over the partitions would be weighted towards the more recent items. In that case it would be better to use a hash/range index with the date as the range.]

From the DynamoDB documentation (<http://docs.aws.amazon.com/amazondynamodb/latest/developerguide/GuidelinesForTables.html>):

"The following table compares some common hash attribute schemas for provisioned throughput efficiency:

Hash attribute value	Uniformity
User ID, where the application has many users.	Good
Status code, where there are only a few possible status codes.	Bad
Item creation date, rounded to the nearest time period (e.g. day, hour, minute)	Bad
Device ID, where each device accesses data at relatively similar intervals	Good
Device ID, where even if there are a lot of devices being tracked, one is by far more popular than all the others.	Bad

If a single table has only a very small number of hash attribute values, consider distributing your write operations across more distinct hash attribute values. In other words, structure the primary key elements to avoid one "hot" (heavily requested) hash attribute value that slows overall performance.

For example, consider a table with a composite primary key. The hash attribute represents the item's creation date, rounded to the nearest day. The range attribute is an item identifier. On a given day, say 2014-07-09, all of the new items will be written to that same hash attribute value.

If the table will fit entirely into a single partition (taking into consideration growth of your data over time), and if your application's read and write throughput requirements do not exceed the read and write capabilities of a single partition, then your application should not encounter any unexpected throttling as a result of partitioning.

However, if you anticipate scaling beyond a single partition, then you should architect your application so that it can use more of the table's full provisioned throughput."

Which is the better approach to uploading large data sets to a DynamoDB table?

372

- Sort by the hash key first
- Distribute the writes evenly across the hash keys (x)

From the DynamoDB Developer Guide

(<http://docs.aws.amazon.com/amazondynamodb/latest/developerguide/GuidelinesForTables.html>):

"There are times when you load data from other data sources into DynamoDB. Typically, DynamoDB partitions your table data on multiple servers. **When uploading data to a table, you get better performance if you upload data to all the allocated servers simultaneously.** [...]

**To fully utilize all of the throughput capacity that has been provisioned for your tables, you need to distribute your workload across your hash attribute values."**

Does DynamoDB support conditional writes?

373

- Yes (x)
- No

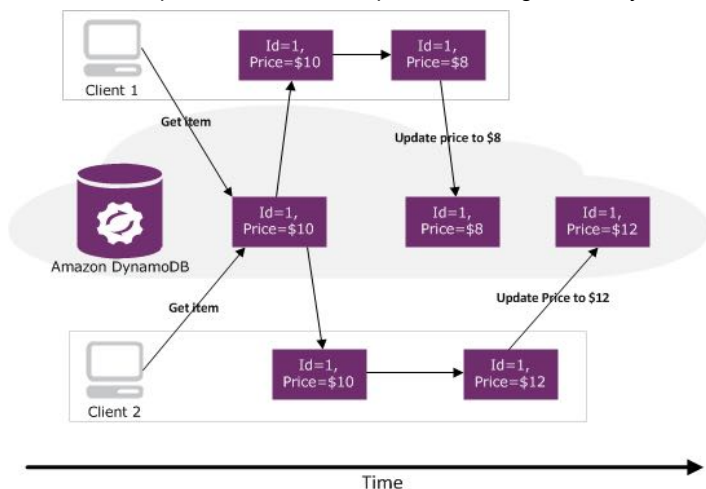
From the DynamoDB Developer Guide

(<http://docs.aws.amazon.com/amazondynamodb/latest/developerguide/WorkingWithItems.html#WorkingWithItems.ConditionalUpdate>):

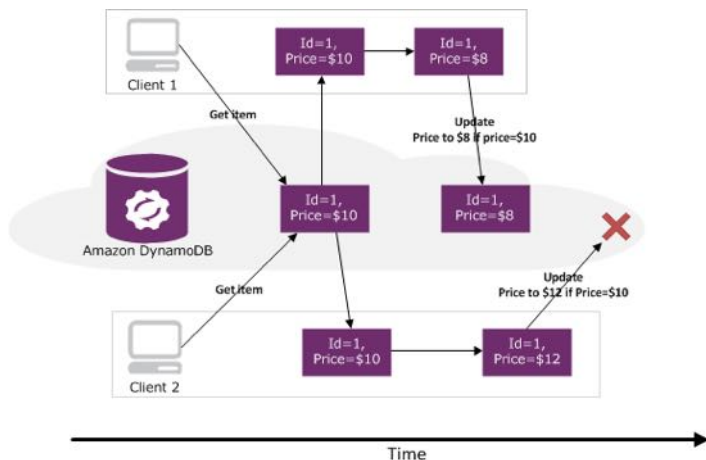
LSI111

### Conditional Writes

In a multi-user environment, multiple clients can access the same item and attempt to modify its attribute values at the same time. However, each client might not realize that other clients might have modified the item already. This is shown in the following illustration in which Client 1 and Client 2 have retrieved a copy of an item (Id=1). Client 1 changes the price from \$10 to \$8. Later, Client 2 changes the same item price to \$12, and the previous change made by Client 1 is lost.



To help clients coordinate writes to data items, **DynamoDB supports conditional writes for PutItem, DeleteItem, and UpdateItem operations.** With a conditional write, an operation succeeds only if the item attributes meet one or more expected conditions; otherwise it returns an error. For example, the following illustration shows both Client 1 and Client 2 retrieving a copy of an item (Id=1). Client 1 first attempts to update the item price to \$8, with the expectation that the existing item price on the server will be \$10. This operation succeeds because the expectation is met. Client 2 then attempts to update the price to \$12, with the expectation that the existing item price on the server will be \$10. This expectation cannot be met, because the price is now \$8; therefore, Client 2's request fails.



Note that conditional writes are *idempotent*. This means that you can send the same conditional write request multiple times, but it will have no further effect on the item after the first time DynamoDB performs the specified update. For example, suppose you issue a request to update the price of a book item by 10%, with the expectation that the price is currently \$20. However, before you get a response, a network error occurs and you don't know whether your request was successful or not. Because a conditional update is an idempotent operation, you can send the same request again, and DynamoDB will update the price only if the current price is still \$20.

To request a conditional PutItem, DeleteItem, or UpdateItem, you specify the condition(s) in the ConditionExpression parameter. ConditionExpression is a string containing attribute names, conditional operators and built-in functions. The entire expression must evaluate to true; otherwise the operation will fail.

LSI111

- 
- Yes
  - No (x)

From the DynamoDB Developer Guide  
(<http://docs.aws.amazon.com/amazondynamodb/latest/developerguide/SecondaryIndexes.html>):

"Local secondary indexes are created at the same time that you create a table. **You cannot add a local secondary index to an existing table, nor can you delete any local secondary indexes that currently exist.**"

---

Is it possible to add or delete a global secondary index (GSI) to/from from a DynamoDB table?

375

- 
- Yes (x)
  - No

From the DynamoDB Developer Guide  
(<http://docs.aws.amazon.com/amazondynamodb/latest/developerguide/SecondaryIndexes.html>):

"Global secondary indexes can be created at the same time that you create a table. **You can also add a new global secondary index to an existing table, or delete an existing global secondary index.**"

---

Can a query against a DynamoDB local secondary index span table partitions?

376

- 
- Yes
  - No (x)

From the DynamoDB Developer Guide  
(<http://docs.aws.amazon.com/amazondynamodb/latest/developerguide/SecondaryIndexes.html>):

**"A local secondary index lets you query over a single partition,** as specified by the hash attribute value in the query."

---

Can a query against a DynamoDB global secondary index span table partitions?

377

- 
- Yes (x)
  - No

From the DynamoDB Developer Guide  
(<http://docs.aws.amazon.com/amazondynamodb/latest/developerguide/SecondaryIndexes.html>):

"A global secondary index lets you query over the entire table, **across all partitions.**"

---

Is a DynamoDB filter expression applied before or after the table query or scan?

378

- 
- Before
  - After (x)

From the DynamoDB Developer Guide  
(<http://docs.aws.amazon.com/amazondynamodb/latest/developerguide/QueryAndScan.html#FilteringResults>):

"With a Query or a Scan operation, you can provide an optional filter expression to refine the results returned to you. **A filter expression lets you apply conditions to the data after it is queried or scanned, but before it is returned to you.** Only the items that meet your conditions are returned."

---

What are some DynamoDB best practices that can improve performance?

379

- 
- Concatenate query attributes into a single LSI (e.g., if you need to query on status and date, create a single range key with status + date (x)
  - Split tables by access frequency (by projecting those specific attributes into a GSI) to reduce query IOPS (x)
  - Ensure that keys are evenly distributed across partitions (x)
  - Shard writes of extremely hot tables by spreading the items across a fixed number of shards and appending a random shard identifier (e.g., an integer from 1 to 10) to an item's hash key for each write; aggregate reads across multiple shards (x)
  - Move less frequently access items into a separate table with lower provisioned I/O (x)
  - Cache read-heavy items (x)
  - Manually specify the ideal number of partitions based on use-case specific performance analysis

Lots of detail in this presentation (<https://www.youtube.com/watch?v=KmHGronoif4>)

## General

---

Does Amazon publish the IP ranges of its servers?

380

- Yes (x)
- No

From the Amazon docs (<http://docs.aws.amazon.com/general/latest/gr/aws-ip-ranges.html>):

"Amazon Web Services (AWS) publishes its current IP address ranges in JSON format. To view the current ranges, download the .json file. To maintain history, save successive versions of the .json file on your system. To determine whether there have been changes since the last time that you saved the file, check the publication time in the current file and compare it to the publication time in the last file that you saved."

---

Which AWS services support automated backups?

381

- RDS (x)
- ElastiCache (Redis only) (x)
- Redshift (x)
- Dynamo
- EC2

This was taken from a slide in Ryan's professional solution architect course (in Lecture 17: Domain 1 - Wrap Up).

---

What is the best AWS storage service for highly structured data that requires sophisticated querying and joining capabilities?

382

- 
- RDS (x)
  - DynamoDB
  - Redshift
  - ElastiCache

From the AWS Storage Options white paper ([http://media.amazonwebservices.com/AWS\\_Storage\\_Options.pdf](http://media.amazonwebservices.com/AWS_Storage_Options.pdf)):

"Amazon RDS is ideal for existing applications that rely on MySQL, Oracle, or SQL Server traditional relational database engines. Since Amazon RDS offers full compatibility and direct access to native database engines, most code, libraries, and tools designed for these databases should work unmodified with Amazon RDS. Amazon RDS is also optimal for new applications with structured data that requires more sophisticated querying and joining capabilities than that provided by Amazon's NoSQL database offering, Amazon DynamoDB."

---

What are reasonable places to store log files?

383

- CloudWatch Logs (x)
- A centralized logging system (AlertLogic, Splunk, SumoLogic, etc) (x)
- S3 Standard Storage (x)
- S3 Reduced Redundancy Storage
- EBS boot device
- Instance storage

Pretty obvious, just including here to keep it fresh during review.

From Amazon's perspective, CloudWatch is preferred and S3 is next preferred depending on the scenario.

## CloudHSM

---

Is a CloudHSM multi-tenant?

384

- Yes
- No (x)

From the HSM FAQ (<https://aws.amazon.com/cloudhsm/faqs/>):

**"Q: Do I share the HSM instance with other AWS customers?"**

No. As part of the service you receive dedicated single tenant access to the HSM appliance."

---

Is it possible to use CloudHSM without a VPC?

385

- Yes
- No (x)

From the CloudHSM FAQ (<https://aws.amazon.com/cloudhsm/faqs/>):

**"Q: I don't currently have a VPC. Can I still use AWS CloudHSM?"**

No. To protect and isolate your CloudHSM from other Amazon customers, CloudHSM must be provisioned inside a VPC. Creating a VPC is easy. Please see the VPC Getting Started Guide for more information."

---

Does an application need to reside in a CloudHSM's VPC to use it?

386

- Yes
- No (x)

From the CloudHSM FAQ (<https://aws.amazon.com/cloudhsm/faqs/>):

**"Q: Does my application need to reside in the same VPC as the CloudHSM instance?"**

No, but the server or instance on which your application and the HSM client is running must have network (IP) reachability to the HSM. You can establish network connectivity from your application to the HSM in many ways, including operating your application in the same VPC, with VPC peering, with a VPN connection, or with Direct Connect. Please see the VPC Peering Guide and VPC User Guide for more details."

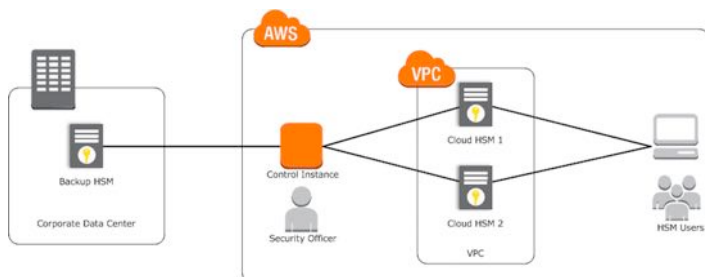
Is CloudHSM highly available?

387

- Yes
- No (x)

From the CloudHSM User Guide (<http://docs.aws.amazon.com/cloudhsm/latest/userguide/configuring-ha.html>):

The recommended configuration for using AWS CloudHSM is to use two HSMs configured in a high-availability (HA) configuration. The failure of a single HSM appliance in a non-HA configuration can result in the permanent loss of keys and data. A minimum of two HSMs are suggested for HA purposes, with each HSM in a different Availability Zone. With this configuration, if one of your HSMs is unavailable, your keys are still available.



HA allows multiple HSMs to be grouped together to form one virtual device, or logical unit, as seen from the client, similar to clustering or RAID technologies. In an HA configuration, service is maintained even if one or more HSMs are unavailable. For example, if three HSMs are combined into an HA group, service is maintained even if two HSMs are offline.

Is it possible to integrate HSM with RDS?

388

- Yes, but only for Oracle instances (x)
- Yes for all instance types
- No

This blog (<https://blogs.aws.amazon.com/security/post/Tx3LXG1HGCVN7BY/AWS-CloudHSM-Is-Now-Integrated-with-Amazon-RDS-for-Oracle-and-Provides-Enhanced>) goes into detail:

**"AWS CloudHSM is now integrated with Amazon RDS for Oracle.** With this new capability, you can let AWS operate your Oracle databases while maintaining control of the master encryption keys. The AWS CloudHSM service helps you meet compliance requirements for data security by making dedicated, single-tenant Hardware Security Module (HSM) appliances available within the AWS cloud. This feature allows you to maintain control of the master encryption keys in CloudHSM instances when encrypting RDS databases with Oracle Transparent Data Encryption (TDE)."

Can I use CloudHSM to store keys or encrypt data used by other AWS services?

389



- 
- Yes (x)
  - No

From the CloudHSM FAQ (<https://aws.amazon.com/cloudhsm/faqs/>):

[Note that you have to write custom applications to do the encryption yourself or use third-party technology...it's not native]

**"Q: Can I use CloudHSM to store keys or encrypt data used by other AWS services?"**

Yes. You can write custom applications and integrate them with CloudHSM, or you can leverage one of the third party encryption solutions available from AWS Technology Partners. Examples include EBS volume encryption and S3 object encryption and key management. Please see the CloudHSM User Guide for a list of supported applications and links to technical application notes that describe third party solutions that work with CloudHSM."

## Costing

---

Does AWS charge for data transfers coming in from the public internet for any of its services?

390

- 
- Yes
  - No (x)

From the "How AWS Pricing Works" white paper ([https://media.amazonwebservices.com/AWS\\_Pricing\\_Overview.pdf](https://media.amazonwebservices.com/AWS_Pricing_Overview.pdf)):

>>>

Free Inbound Data Transfer

There is no charge for inbound data transfer across all Amazon Web Services in all regions. There are no outbound data transfer charges between Amazon Web Services within the same region.

<<<

---

Am I charged for data transfer between EC2 instances in different regions?

391

- 
- Yes (x)
  - No

From the EC2 FAQ (<https://aws.amazon.com/ec2/faqs/>):

**"Q. If I have two instances in different regions, how will I be charged for data transfer?"**

Each instance is charged for its data in and data out at Internet Data Transfer rates. Therefore, if data is transferred between these two instances, it is charged at Internet Data Transfer Out for the first instance and at Internet Data Transfer In for the second instance."

Also see the EC2 pricing page (<https://aws.amazon.com/ec2/pricing/>)

---

Am I charged for data transfer between EC2 instances in the same region?

392

- 
- Yes for transfers between availability zones (x)
  - Yes in all cases
  - Yes for transfers within an availability zone
  - No for transfers between availability zones
  - No in all cases
  - No for transfers within an availability zone

From the EC2 FAQ (<https://aws.amazon.com/ec2/faqs/>):

**"Q: If I have two instances in different availability zones, how will I be charged for regional data transfer?"**

Each instance is charged for its data in and data out. Therefore, if data is transferred between these two instances, it is charged out for the first instance and in for the second instance."

Also see the EC2 pricing page (<https://aws.amazon.com/ec2/pricing/>)

---

Am I charged for data transfer between EC2 instances and other AWS services in the same region?

393

- Yes
- No (x)

From the EC2 FAQ (<https://aws.amazon.com/ec2/faqs/>):

**"Q: How will I be charged and billed for my use of Amazon EC2?"**

You pay only for what you use and there is no minimum fee. Pricing is per instance-hour consumed for each instance type. Partial instance-hours consumed are billed as full hours. **There is no Data Transfer charge between two Amazon Web Services within the same region (i.e. between Amazon EC2 US West and another AWS service in the US West).** Data transferred between AWS services in different regions will be charged as Internet Data Transfer on both sides of the transfer. Usage for other Amazon Web Services is billed separately from Amazon EC2."

Also see the EC2 pricing page (<https://aws.amazon.com/ec2/pricing/>)

---

Is there a cost to transfer data into an EC2 instance from the public internet?

394

- Yes
- No (x)

See the EC2 pricing page (<https://aws.amazon.com/ec2/pricing/>)

---

How much time will you be charged for if you run an EC2 instance for only thirty minutes?

395

- 30 minutes
- one full hour (x)
- nothing

From the EC2 FAQ (<https://aws.amazon.com/ec2/faqs/>):

**"Q: How will I be charged and billed for my use of Amazon EC2?"**

You pay only for what you use and there is no minimum fee. Pricing is per instance-hour consumed for each instance type. **Partial instance-hours consumed are billed as full hours.**

---

What is the approximate cost savings with S3 RRS?

396

- 
- 10%
  - 20% (x)
  - 50%
  - 75%

I can't find any definitive information on this, but I did some quick checks with the Simple Monthly Calculator, and for both 1TB and 100TB of data, the savings is almost exactly 20%.

I also found a blog (<http://harish11g.blogspot.co.nz/2013/05/Amazon-Web-Services-AWS-Cost-Saving-Tips-Amazon-S3-Reduced-Redundancy-storage-rrs-vs-std.html>) that did a more thorough comparison:

**"Savings: From the above case you can conclude that by using RRS option you can save around 20% in your storage costs compared to Amazon S3 standard option."**

## Billing

---

When using Consolidated Billing, can the payer account access AWS resources in the other linked accounts in the Account Family?

397

- 
- Yes
  - No (x)

From the Consolidated Billing documentation (<http://docs.aws.amazon.com/awsaccountbilling/latest/aboutv2/consolidated-billing.html>):

"The payer account is billed for all charges of the linked accounts. However, each linked account is completely independent in every other way (signing up for services, accessing resources, using AWS Premium Support, etc.). The payer account owner cannot access data belonging to the linked account owners (e.g., their files in Amazon S3). Each account owner uses their own IAM user name and password, with account permissions assigned independently of any other account in the Consolidated Billing family."

## Elastic Beanstalk

---

What languages are supported by Elastic Beanstalk?

398

- 
- Java (x)
  - .Net (x)
  - Perl
  - Ruby (x)
  - Python (x)
  - PHP (x)
  - Node.js (x)

From the Elastic Beanstalk FAQ (<https://aws.amazon.com/elasticbeanstalk/faqs/>):

**"Q: What languages and development stacks does AWS Elastic Beanstalk support?"**

AWS Elastic Beanstalk supports the following languages and development stacks:

- Apache Tomcat for Java applications
- Apache HTTP Server for PHP applications
- Apache HTTP Server for Python applications
- Nginx or Apache HTTP Server for Node.js applications
- Passenger for Ruby applications
- Microsoft IIS 7.5 for .NET applications"

---

What application/web servers are supported by Elastic Beanstalk?

399

- 
- Apache Tomcat for Java applications (x)
  - Apache HTTP Server for PHP applications (x)
  - Apache HTTP Server for Python applications (x)
  - Apache HTTP Server for Perl applications
  - Nginx or Apache HTTP Server for Node.js applications (x)
  - Passenger for Ruby applications (x)
  - Microsoft IIS 7.5 for .Net applications (x)
  - Jetty for Java applications
  - Varnish for Java applications

From the Elastic Beanstalk FAQ (<https://aws.amazon.com/elasticbeanstalk/faqs/>):

**"Q: What languages and development stacks does AWS Elastic Beanstalk support?"**

AWS Elastic Beanstalk supports the following languages and development stacks:

- Apache Tomcat for Java applications
- Apache HTTP Server for PHP applications
- Apache HTTP Server for Python applications
- Nginx or Apache HTTP Server for Node.js applications
- Passenger for Ruby applications
- Microsoft IIS 7.5 for .NET applications"

---

Can Elastic Beanstalk automatically provision RDS instances?

400

- Yes (x)
- No

From the Elastic Beanstalk FAQ (<https://aws.amazon.com/elasticbeanstalk/faqs/>):

**"Q: How do I setup a database for use with AWS Elastic Beanstalk?"**

Elastic Beanstalk can automatically provision an Amazon RDS DB Instance. The connectivity information to the DB Instance is exposed to your application by environment variables."

---

What operating systems are supported by Elastic Beanstalk?

401

- Amazon Linux (x)
- Ubuntu Linux
- RedHat Linux
- Windows Server 2008 R2 (x)
- Windows Server 2012 R2

From the Elastic Beanstalk FAQ (<https://aws.amazon.com/elasticbeanstalk/faqs/>):

**"Q: What operating systems does AWS Elastic Beanstalk use?"**

AWS Elastic Beanstalk runs on the Amazon Linux AMI and the Windows Server 2008 R2 AMI. Both AMIs are supported and maintained by Amazon Web Services and are designed to provide a stable, secure, and high performance execution environment for Amazon EC2 cloud computing."

---

What are the available Elastic Beanstalk environment tiers?

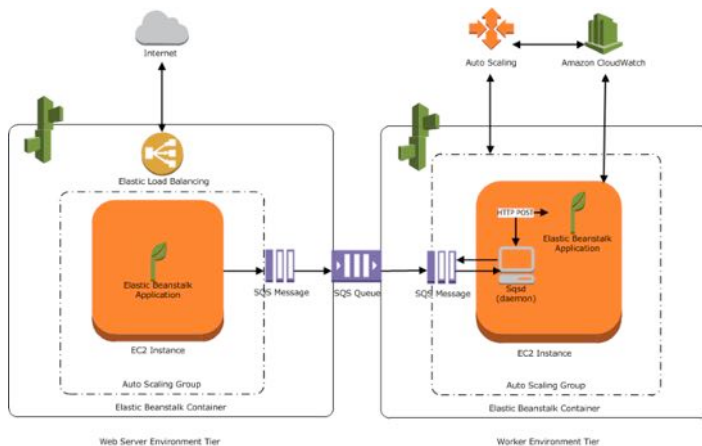
402

- Web Server (x)
- Worker (x)
- Daemon
- Container

From the Elastic Beanstalk Developer Guide

(<https://docs.aws.amazon.com/elasticbeanstalk/latest/dg/concepts.concepts.architecture.html>):

"When you launch an Elastic Beanstalk environment, you choose an environment tier, platform, and environment type. The environment tier that you choose determines whether Elastic Beanstalk provisions resources to support a web application that handles HTTP(S) requests or a web application that handles background-processing tasks. An environment tier whose web application processes web requests is known as a **web server tier**. An environment tier whose web application runs background jobs is known as a **worker tier**. This topic describes the components, resources, and architecture for each type of environment tier."



Is it possible for one Elastic Beanstalk environment to support multiple environment tiers?

403

- Yes
- No (x)

From the Elastic Beanstalk Developer Guide

(<https://docs.aws.amazon.com/elasticbeanstalk/latest/dg/concepts.concepts.architecture.html>):

"One environment cannot support two different environment tiers because each requires its own set of resources; a worker environment tier and a web server environment tier each require an Auto Scaling group, but Elastic Beanstalk supports only one Auto Scaling group per environment."

Is it possible to specify an existing security group in an Elastic Beanstalk configuration?

404

- Yes (x)
- No

From the Elastic Beanstalk Developer Guide (<https://docs.aws.amazon.com/elasticbeanstalk/latest/dg/using-features.managing.ec2.html#using-features.managing.ec2.securitygroups>):

"You can control access to your Elastic Beanstalk application using an *Amazon EC2 security group*. A security group defines firewall rules for your instances. These rules specify which ingress (i.e., incoming) network traffic should be delivered to your instance. All other ingress traffic will be discarded. You can modify rules for a group at any time. The new rules are automatically enforced for all running instances and instances launched in the future."

Is it possible to specify a custom AMI in an Elastic Beanstalk configuration?

405

- 
- Yes (x)
  - No

From the Elastic Beanstalk Developer Guide (<https://docs.aws.amazon.com/elasticbeanstalk/latest/dg/using-features.managing.ec2.html#using-features.managing.ec2.securitygroups>):

"The Amazon Machine Image (AMI) is the Amazon Linux or Windows Server machine image that AWS Elastic Beanstalk uses to launch EC2 instances in your environment. Elastic Beanstalk provides machine images that contain the tools and resources required to run your application.

Elastic Beanstalk selects a default AMI for your environment based on the region, platform, and instance type that you choose. If you have created a custom AMI, replace the default AMI ID with yours"

---

Is it possible to specify a custom AMI in an Elastic Beanstalk configuration?

406

- 
- Yes (x)
  - No

From the Elastic Beanstalk Developer Guide (<https://docs.aws.amazon.com/elasticbeanstalk/latest/dg/using-features.managing.ec2.html#using-features.managing.ec2.securitygroups>):

"The Amazon Machine Image (AMI) is the Amazon Linux or Windows Server machine image that AWS Elastic Beanstalk uses to launch EC2 instances in your environment. Elastic Beanstalk provides machine images that contain the tools and resources required to run your application.

Elastic Beanstalk selects a default AMI for your environment based on the region, platform, and instance type that you choose. If you have created a custom AMI, replace the default AMI ID with yours"

---

When an Elastic Beanstalk environment's EBS settings are modified (volume type or size), do the changes take effect immediately?

407

- 
- Yes, the instances are modified in place
  - Yes, the existing environment is torn down and a new one created in its place (x)
  - No, you have to manually recreate the environment

Verified this via the Console...

---

Which of the following are valid Elastic Beanstalk deployment options?

408

- 
- Rolling in-place deployment in batches (x)
  - CNAME swap (x)
  - Manual

From the Elastic Beanstalk Developer Guide ():

"Deploying a new version of your application to an environment is typically a fairly quick process. **The new source bundle is deployed to an instance and extracted, and then the web container or application server picks up the new version and restarts if necessary.**

If your application takes a long time to start up, however, there can be significant down time when a new version is deployed. Elastic Beanstalk has two features to avoid downtime during application deployment. One is **CNAME Swap**, where you deploy your new version to a second environment and change the DNS settings to start directing traffic to the updated environment once the deployment is complete.

The second feature is a **rolling deployment** (not to be confused with a rolling configuration update). With rolling deployments, Elastic Beanstalk **splits the environments into batches and deploys to one batch at a time**, leaving the rest of the instances in the environment running the old application version. In the middle of a rolling deployment, some instances can be serving requests with the old version, while others from an already completed batch could be serving other requests with the new version at the same time."

- Yes (x)
- No

From the Elastic Beanstalk Developer Guide (<http://docs.aws.amazon.com/elasticbeanstalk/latest/dg/AWSHowTo-vpc.html>):

"Amazon Virtual Private Cloud (Amazon VPC) enables you to define a virtual network in your own isolated section within the Amazon Web Services (AWS) cloud, known as a *virtual private cloud (VPC)*. Using VPC, you can deploy a new class of web applications on Elastic Beanstalk, including internal web applications (such as your recruiting application), web applications that connect to an on-premise database (using a VPN connection), as well as private web service back-ends. **Elastic Beanstalk launches your AWS resources, such as instances, into your VPC.** Your VPC closely resembles a traditional network, with the benefits of using AWS's scalable infrastructure. You have complete control over your VPC; you can select the IP address range, create subnets, and configure routes and network gateways. To protect the resources in each subnet, you can use multiple layers of security, including security groups and network access control lists. "

The link goes into a lot more detail.

- Upload a certificate to IAM (x)
- Use a custom DNS domain mapped to the environment's DNS name (x)
- Submit a request to Amazon
- Update the environment's configuration (x)

From the Elastic Beanstalk Developer Guide (<http://docs.aws.amazon.com/elasticbeanstalk/latest/dg/configuring-https.html>):

"To configure HTTPS for your Elastic Beanstalk application, you must perform the following tasks:

#### Tasks

- Step 1: Create a Custom Domain Name
- Step 2: Create an X509 Certificate
- Step 3: Upload the Certificate to IAM
- Step 4: Update Your Elastic Beanstalk Environment to Use HTTPS"

## Autoscaling

- Yes
- No (x)

From the Auto Scaling documentation (<http://docs.aws.amazon.com/AutoScaling/latest/DeveloperGuide/LaunchConfiguration.html>):

"A *launch configuration* is a template that an Auto Scaling group uses to launch EC2 instances. When you create a launch configuration, you specify information for the instances such as the ID of the Amazon Machine Image (AMI), the instance type, a key pair, one or more security groups, and a block device mapping.

When you create an Auto Scaling group, you must specify a launch configuration. You can specify your launch configuration with multiple Auto Scaling groups. However, you can only specify one launch configuration for an Auto Scaling group at a time, and you can't modify a launch configuration after you've created it. Therefore, if you want to change the launch configuration for your Auto Scaling group, you must create a new launch configuration and then update your Auto Scaling group with the new launch configuration. When you change the launch configuration for your Auto Scaling group, any new instances are launched using the new configuration parameters, but existing instances are not affected."

- 
- Yes
  - No (x)

From the Auto Scaling documentation (<http://docs.aws.amazon.com/AutoScaling/latest/DeveloperGuide/LaunchConfiguration.html>):

"A *launch configuration* is a template that an Auto Scaling group uses to launch EC2 instances. When you create a launch configuration, you specify information for the instances such as the ID of the Amazon Machine Image (AMI), the instance type, a key pair, one or more security groups, and a block device mapping.

When you create an Auto Scaling group, you must specify a launch configuration. You can specify your launch configuration with multiple Auto Scaling groups. However, you can only specify one launch configuration for an Auto Scaling group at a time, and you can't modify a launch configuration after you've created it. Therefore, if you want to change the launch configuration for your Auto Scaling group, you must create a new launch configuration and then update your Auto Scaling group with the new launch configuration. When you change the launch configuration for your Auto Scaling group, any new instances are launched using the new configuration parameters, but existing instances are not affected."

---

What happens when an Auto Scaling group's launch configuration is changed?

413

- Existing instances are terminated
- Existing instances are not affected (x)
- Existing instances are restarted with the new parameters
- New instances are launched with the new parameters (x)

From the Auto Scaling documentation (<http://docs.aws.amazon.com/AutoScaling/latest/DeveloperGuide/LaunchConfiguration.html>):

"A *launch configuration* is a template that an Auto Scaling group uses to launch EC2 instances. When you create a launch configuration, you specify information for the instances such as the ID of the Amazon Machine Image (AMI), the instance type, a key pair, one or more security groups, and a block device mapping.

When you create an Auto Scaling group, you must specify a launch configuration. You can specify your launch configuration with multiple Auto Scaling groups. However, you can only specify one launch configuration for an Auto Scaling group at a time, and you can't modify a launch configuration after you've created it. Therefore, if you want to change the launch configuration for your Auto Scaling group, you must create a new launch configuration and then update your Auto Scaling group with the new launch configuration. When you change the launch configuration for your Auto Scaling group, any new instances are launched using the new configuration parameters, but existing instances are not affected."

---

If not specified, what is the default desired capacity of an Auto Scaling group?

414

- Zero
- One
- The group's minimum number of instances (x)

From the Auto Scaling documentation (<http://docs.aws.amazon.com/AutoScaling/latest/DeveloperGuide/AutoScalingGroup.html>):

When you create a Auto Scaling group, you must specify a name, launch configuration, minimum number of instances, and maximum number of instances. You can optionally specify a desired capacity, which is the number of instances that the group must have at all times. **If you don't specify a desired capacity, the default desired capacity is the minimum number of instances that you specified.** For information about creating an Auto Scaling group, see Creating Auto Scaling Groups.

---

What properties must be specified when creating an Auto Scaling group?

415



- 
- Name (x)
  - Launch configuration (x)
  - Minimum number of instances (x)
  - Maximum number of instances (x)
  - Desired capacity
  - Termination conditions
  - Notification ARN

From the Auto Scaling documentation (<http://docs.aws.amazon.com/AutoScaling/latest/DeveloperGuide/AutoScalingGroup.html>):

When you create a Auto Scaling group, **you must specify a name, launch configuration, minimum number of instances, and maximum number of instances. You can optionally specify a desired capacity**, which is the number of instances that the group must have at all times. If you don't specify a desired capacity, the default desired capacity is the minimum number of instances that you specified. For information about creating an Auto Scaling group, see Creating Auto Scaling Groups.

---

What is the meaning of the Auto Scaling group's "desired capacity" property?

416

- 
- number of instances the group must have at all times (x)
  - a target capacity but not enforced
  - the average of the min and max number of instances

From the Auto Scaling documentation (<http://docs.aws.amazon.com/AutoScaling/latest/DeveloperGuide/AutoScalingGroup.html>):

When you create a Auto Scaling group, **you must specify a name, launch configuration, minimum number of instances, and maximum number of instances. You can optionally specify a desired capacity**, which is the number of instances that the group must have at all times. If you don't specify a desired capacity, the default desired capacity is the minimum number of instances that you specified. For information about creating an Auto Scaling group, see Creating Auto Scaling Groups.

---

Which of the following are valid Auto Scaling termination policies?

417

- OldestInstance (x)
- NewestInstance (x)
- OldestLaunchConfiguration (x)
- NewestLaunchConfiguration (x)
- ClosestToNextInstanceHour (x)
- Default (x)
- LowestCPUUtilization
- HighestCPUUtilization

From the Auto Scaling documentation

(<http://docs.aws.amazon.com/AutoScaling/latest/DeveloperGuide/AutoScalingBehavior.InstanceTermination.html>):

"Auto Scaling currently supports the following custom termination policies:

- **OldestInstance.** Auto Scaling terminates the oldest instance in the group. This option is useful when you're upgrading the instances in the Auto Scaling group to a new EC2 instance type, and want to eventually replace instances with older instances with newer ones.
- **NewestInstance.** Auto Scaling terminates the newest instance in the group. This policy is useful when you're testing a new launch configuration but don't want to keep it in production.
- **OldestLaunchConfiguration.** Auto Scaling terminates instances that have the oldest launch configuration. This policy is useful when you're updating a group and phasing out the instances from a previous configuration.
- **ClosestToNextInstanceHour.** Auto Scaling terminates instances that are closest to the next billing hour. This policy helps you maximize the use of your instances and manage costs.
- **Default.** Auto Scaling uses its default termination policy. This policy is useful when you have more than one scaling policy associated with the group."

Is it possible for an Auto Scaling group to have more than one termination policy?

418

- Yes, they are evaluated in order (x)
- Yes, they are evaluated in reverse order
- No

From the Auto Scaling documentation

(<http://docs.aws.amazon.com/AutoScaling/latest/DeveloperGuide/AutoScalingBehavior.InstanceTermination.html>):

''''''

You can use these policies individually, or combine them into a list of policies that Auto Scaling uses when terminating instances. For example, use the following command to update an Auto Scaling group to use the OldestLaunchConfiguration policy first, and then to use the ClosestToNextInstanceHour policy:

```
aws autoscaling update-auto-scaling-group --auto-scaling-group-name my-asg --termination-policies
"OldestLaunchConfiguration,ClosestToNextInstanceHour"
```

If you use the Default termination policy, make it the last one in the list of termination policies. For example [amend the above command with]

```
--termination-policies "OldestLaunchConfiguration,Default"
```

''''''

Given an Auto Scaling group that spans two Availability Zones with two instances in AZ-1 and three instances in AZ-2, to which instance(s) is the termination policy applied?

419

- The best policy match is selected from all instances across all of the Auto Scaling group's availability zones
- The best policy match is selected only from those instance is AZ-2 (x)
- The best policy match is selected only from those instance is AZ-1 (x)

From the Auto Scaling documentation

(<http://docs.aws.amazon.com/AutoScaling/latest/DeveloperGuide/AutoScalingBehavior.InstanceTermination.html>):

"When you customize the termination policy, Auto Scaling first assesses the Availability Zones for any imbalance. **If an Availability Zone has more instances than the other Availability Zones that are used by the group, then Auto Scaling applies your specified termination policy on the instances from the imbalanced Availability Zone.** If the Availability Zones used by the group are balanced, then Auto Scaling applies the termination policy that you specified."

What factors are considered in an Auto Scaling group's default termination policy?

420

- Try to keep the number of instances the same in all AZs (x)
- Age of launch configuration (x)
- Proximity to the next billing hour (x)
- CPU utilization
- Instance size

From the Auto Scaling documentation

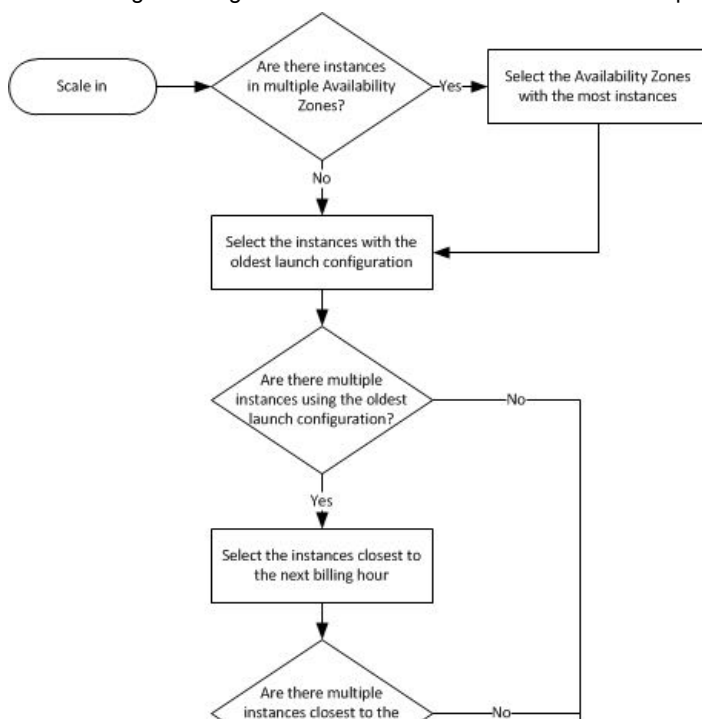
(<http://docs.aws.amazon.com/AutoScaling/latest/DeveloperGuide/AutoScalingBehavior.InstanceTermination.html>):

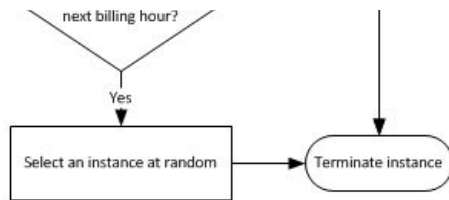
447777

The default termination policy is designed to help ensure that your network architecture spans Availability Zones evenly. When using the default termination policy, Auto Scaling selects an instance to terminate as follows:

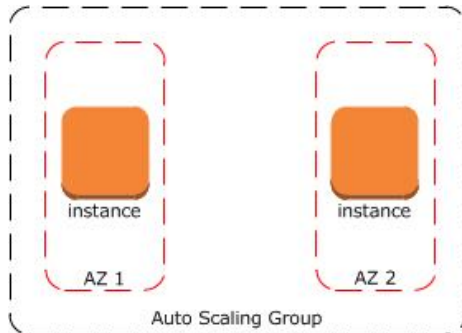
1. **Auto Scaling determines whether there are instances in multiple Availability Zones. If so, it selects the Availability Zone with the most instances.** If there is more than one Availability Zone with this number of instances, Auto Scaling selects the Availability Zone with the instances that use the oldest launch configuration.
2. **Auto Scaling determines which instances in the selected Availability Zone use the oldest launch configuration.** If there is one such instance, it terminates it.
3. If there are multiple instances that use the oldest launch configuration, **Auto Scaling determines which instances are closest to the next billing hour.** (This helps you maximize the use of your EC2 instances while minimizing the number of hours you are billed for Amazon EC2 usage.) If there is one such instance, Auto Scaling terminates it.
4. If there is more than one instance closest to the next billing hour, Auto Scaling selects one of these instances at random.

The following flow diagram illustrates how the default termination policy works.

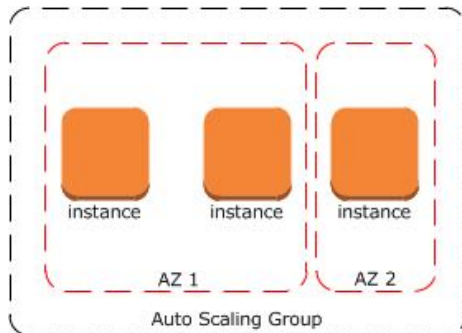




Consider an Auto Scaling group that has two Availability Zones, a desired capacity of two instances, and scaling policies that increase and decrease the number of instances by 1 when certain thresholds are met. The two instances in this group are distributed as follows.



When the threshold for the scale out policy is met, the policy takes effect and Auto Scaling launches a new instance. The Auto Scaling group now has three instances, distributed as follows.



When the threshold for the scale in policy is met, the policy takes effect and Auto Scaling terminates one of the instances. If the group does not have a specific termination policy assigned to it, Auto Scaling uses the default termination policy. Auto Scaling selects the Availability Zone with two instances, and terminates the instance launched from the oldest launch configuration. If the instances were launched from the same launch configuration, then Auto Scaling selects the instance that is closest to the next billing hour and terminates it.

"""

What is an Auto Scaling “health check grace period”?

421

- a delay before health checks are performed on new instances (x)
- the amount of time after a health check fails when another check is issued
- a timeout for health checks

From the Autoscale documentation (<http://docs.aws.amazon.com/AutoScaling/latest/DeveloperGuide/healthcheck.html>):

"Frequently, new instances need to warm up briefly before they can pass the Auto Scaling health check. Auto Scaling waits until the health check grace period that you specified ends before determining the health status of a newly launched instance. Note that the EC2 status checks and ELB health checks can complete before the health check grace period expires, but Auto Scaling does not act on them until the health check grace period expires. To provide ample warm-up time for your instances, set the health check grace period of the Auto Scaling group to cover the expected startup period of your application. If you add a lifecycle hook to perform actions as your instances launch, the health check grace period does not start until you complete the lifecycle hook and the instance enters the InService state."

Is it possible to use a custom health check in an Auto Scaling group?

422

- 
- Yes (x)
  - No

From the Auto Scaling Developer Guide (<http://docs.aws.amazon.com/AutoScaling/latest/DeveloperGuide/healthcheck.html#as-configure-healthcheck>):

"If you have custom health checks, you can send the information from your health checks to Auto Scaling so that Auto Scaling can use this information. For example, if you determine that an instance is not functioning as expected, you can set the health status of the instance to Unhealthy, and then Auto Scaling schedules the instance for replacement."

---

What are the possible Auto Scaling health checks?

423

- EC2 status checks (x)
- ELB health check (x)
- Health explicitly set manually or via API (x)
- Health triggered via EC2 CloudWatch metrics

From the Auto Scaling documentation (<http://docs.aws.amazon.com/AutoScaling/latest/DeveloperGuide/healthcheck.html>):

423

Auto Scaling determines the health status of an instance using one or more of the following:

- EC2 status checks. For more information, see Status Checks for Your Instances in the *Amazon EC2 User Guide for Linux Instances*.
- ELB health checks. For more information, see Configure Health Checks in the *Elastic Load Balancing Developer Guide*.
- Custom health checks. For more information, see Set Instance Health Status Based on Custom Health Checks.

423

---

What is the default Auto Scaling health check?

424

- EC2 status checks (x)
- ELB health check
- Health explicitly set manually or via API
- Health triggered via EC2 CloudWatch metrics

From the Auto Scaling documentation (<http://docs.aws.amazon.com/AutoScaling/latest/DeveloperGuide/healthcheck.html>):

"By default, Auto Scaling health checks use the results of the EC2 status checks to determine the health status of an instance. Auto Scaling marks an instance as unhealthy if its instance status is any value other than running or its system status is impaired."

---

What are the valid Auto Scaling instance states?

425

- Pending (x)
- InService (x)
- Terminating (x)
- Terminated (x)
- Standby (x)
- Unknown
- Pending ELB

From the Auto Scaling documentation (<http://docs.aws.amazon.com/AutoScaling/latest/DeveloperGuide/AutoScalingGroupLifecycle.html>):

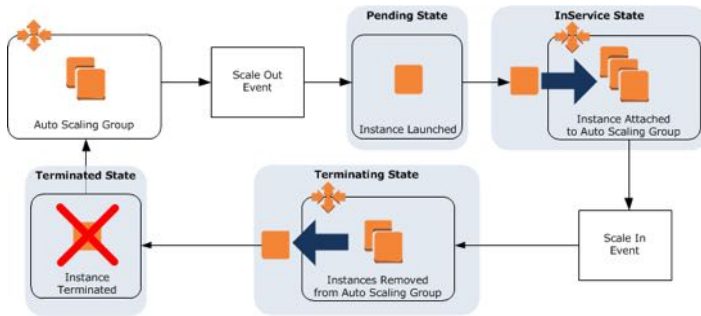
"Auto Scaling Instance States

Instances in an Auto Scaling group can be in one of four main states:

- Pending

- Pending
- InService
- Terminating
- Terminated

The following diagram shows how an instance moves from one state to another.



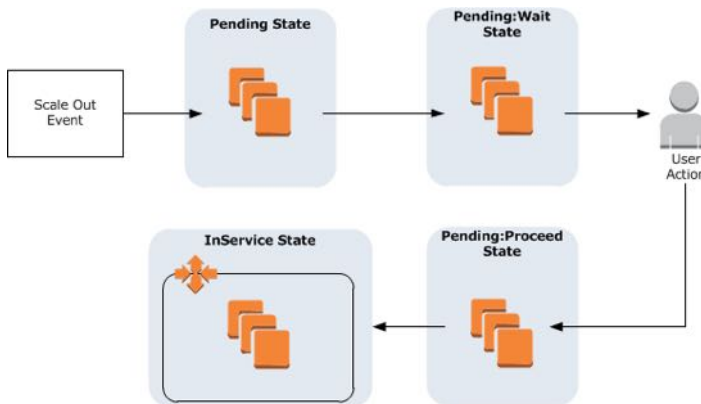
You can take specific actions when an instance is in one of these states:

...

### Auto Scaling Pending State

When an Auto Scaling group reaches a scale out threshold, it launches one or more instances (as determined by your scaling policy). These instances are configured based on the launch configuration for the Auto Scaling group. While an instance is launched and configured, it is in a Pending state.

Depending on how you want to manage your Auto Scaling group, the Pending state can be divided into two additional states: Pending:Wait and Pending:Proceed. You can use these states to perform additional actions before the instances are added to the Auto Scaling group.



...

### Auto Scaling InService State

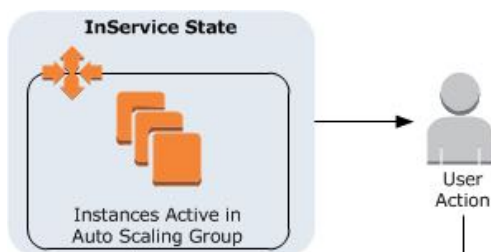
Instances that are functioning within your application as part of an Auto Scaling group are in the InService state. Instances remain in this state until:

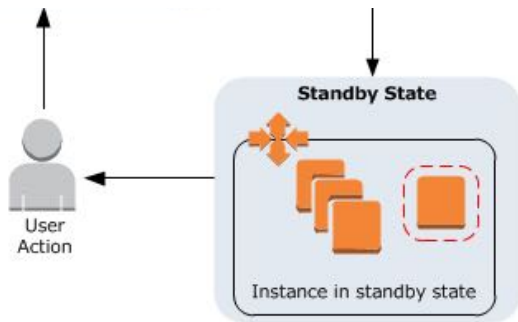
- An Auto Scaling scale in event occurs, reducing the size of the Auto Scaling group
- You put the instance into a Standby state.
- You manually detach the instance from the Auto Scaling group
- The instance fails a required number of health checks or you manually set the status of the instance to Unhealthy.

In addition, any running instances that you attach to the Auto Scaling group are also in the InService state.

You have the option of putting any InService instance into a

Standby state. Instances in this state continue to be managed by the Auto Scaling group. However, they are not an active part of your application until you put them back into service.





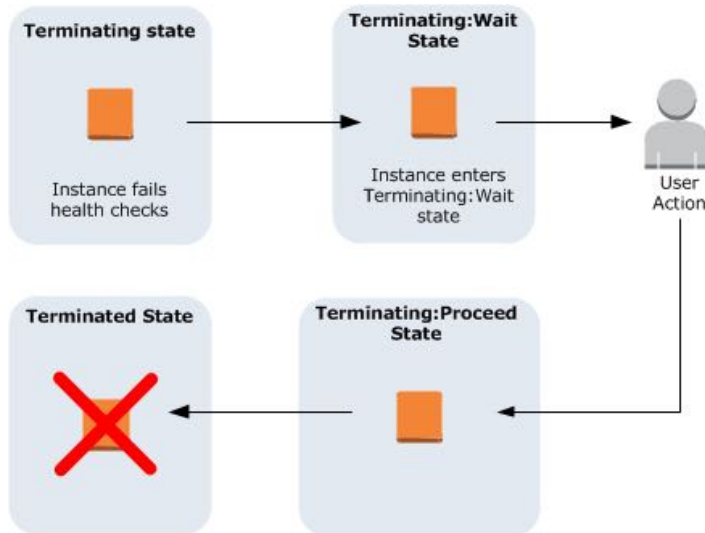
Examples of when you might put instances into the Standby state include:

- To update or modify the instance
- To troubleshoot an instance that isn't performing as expected

### Auto Scaling Terminating State

Instances that fail a required number of health checks are removed from an Auto Scaling group and terminated. The instances first enter the Terminating state, then Terminated.

Depending on how you want to manage your Auto Scaling group, the Terminating state can be divided into two additional states: Terminating:Wait and Terminating:Proceed. You can use these states to perform additional actions before the instances are terminated.



How long can an instance in an Autoscaling Group remain in either the Pending:Wait or Terminating:Wait states before being terminated?

426

- 15 minutes
- 30 minutes
- 60 minutes (x)
- 120 minutes

From the Auto Scaling documentation (<http://docs.aws.amazon.com/AutoScaling/latest/DeveloperGuide/introducing-lifecycle-hooks.html>):

"By default, the instance remains in the Pending:Wait or Terminating:Wait state for one hour. If you take no action during that time, Auto Scaling terminates the instance. You can extend the length of time the instance remains in a waiting state by recording a heartbeat."

After an Auto Scaling group invokes a lifecycle hook, will it respond to further scaling actions while it's waiting for the instance to be put back into service?

427

- Yes
- No (x)

From the Auto Scaling documentation (<http://docs.aws.amazon.com/AutoScaling/latest/DeveloperGuide/lifecycle-hook-considerations.html#lifecycle-hook-cooldowns>):

"Each time Auto Scaling launches or terminates an instance, a cooldown takes effect. This cooldown helps ensure that the Auto Scaling group does not launch or terminate more instances than needed.

When you put a lifecycle hook on an Auto Scaling group, any scaling actions are suspended until the instance is in service. After the instance is in service, the cooldown period starts.

For example, consider an Auto Scaling group that has a lifecycle hook that allows for custom actions as new instances launch. The application experiences an increase in demand, and Auto Scaling launches a new instance to address the need for additional capacity. Because there is a lifecycle hook, the instance is put into the

Pending:Wait

state, which means the instance is not available to handle traffic yet. Scaling actions are suspended for the Auto Scaling group. When the instance is put into service, the cooldown period starts and, when it expires, additional scaling actions can resume."

What of the possible outcomes of an Auto Scaling lifecycle hook?

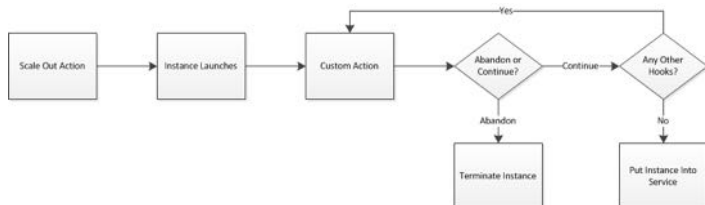
428

- ABANDON (x)
- CONTINUE (x)
- TERMINATE
- RETRY

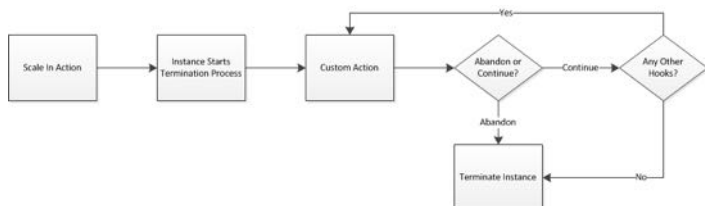
From the Auto Scaling documentation ():

"At the conclusion of a lifecycle hook, an instance can have one of two results: ABANDON or CONTINUE.

If the instance is launching, an ABANDON result means that whatever additional actions you wanted to take on the instance were unsuccessful. Instead of putting the instance into service, Auto Scaling terminates the instance and, if necessary, launches a new one. A CONTINUE result means that your actions were successful, and Auto Scaling can put the instance into service.



If the instance is terminating, an ABANDON result means stop any remaining actions, such as other lifecycle hooks, and move straight to terminating the instance. CONTINUE result means continue with the termination process, but allow any other lifecycle hooks applied to the instance take effect as well.



Note: For terminating instances, both an ABANDON result and a CONTINUE result cause the instance to terminate. The main difference is whether any other actions are allowed to occur on the instance.

What happens when a spot instance being managed by an Auto Scaling group is terminated due to a change in spot price \*and\* that instance has a lifecycle hook set to fire on instance termination?

429



- 
- The instance is terminated immediately (x)
  - The instance is not terminated until `CompleteLifecycleAction` is called

From the Auto Scaling documentation (<http://docs.aws.amazon.com/AutoScaling/latest/DeveloperGuide/lifecycle-hook-considerations.html>):

"You can use lifecycle hooks with Spot Instances. However, a lifecycle hook does not prevent an instance from terminating due to a change in the Spot Price, which can happen at any time. In addition, when a Spot Instance terminates, you must still complete the lifecycle action (using the **complete-lifecycle-action** command or the **CompleteLifecycleAction** action)."

---

Is it possible to suspend Auto Scaling processes to, e.g., troubleshoot instances?

430

- Yes (x)
- No

From the Auto Scaling Developer Guide ([http://docs.aws.amazon.com/AutoScaling/latest/DeveloperGuide/US\\_SuspendResume.html](http://docs.aws.amazon.com/AutoScaling/latest/DeveloperGuide/US_SuspendResume.html)):

"Auto Scaling enables you to suspend and then resume one or more of the Auto Scaling processes in your Auto Scaling group. This can be very useful when you want to investigate a configuration problem or other issue with your web application and then make changes to your application, without triggering the Auto Scaling process."

The document goes into a lot more detail.

What is the best approach to Auto Scaling?

431

- Scale up fast, scale down slow (x)
- Scale up slow, scale down fast
- Scale up slow, scale down slow
- Scale up fast, scale down fast

For scaling up, the general idea is that you want to scale up on a low threshold (say 70% vs 90%) because you want to ensure sufficient capacity while the instances come online.

For scaling down, the idea is that since you pay by the hour, it doesn't matter if the newly created instances are up for five minutes or fifty minutes. Aggressively scaling down isn't likely to save you money.

For example:

Scale up: +2 instances if CPU utilization > 70% for 2 minutes

Scale down: -2 instances if CPU utilization < 40% for 20 minutes

Cooldown: 300 seconds

The policies scale up or down by 2 to keep AZs balanced, the number should match the number of AZs.

This came from a re:Invent presentation (<https://www.youtube.com/watch?v=Z1CEZzxKs>):



## OpsWorks

What are some of the scaling options for OpsWorks instances?

- 
- Manual scaling (24/7) (x)
  - Time-based (x)
  - Load-based (x)
  - CloudWatch-triggered

From the AWS OpsWorks User Guide (<http://docs.aws.amazon.com/opsworks/latest/userguide/workinginstances-autoscaling.html>):

"Automatic scaling is based on two instance types, which adjust a layer's online instances based on different criteria:

- **Time-based instances**

They allow a stack to handle loads that follow a predictable pattern by including instances that run only at certain times or on certain days. For example, you could start some instances after 6PM to perform nightly backup tasks or stop some instances on weekends when traffic is lower.

- **Load-based instances**

They allow a stack to handle variable loads by starting additional instances when traffic is high and stopping instances when traffic is low, based on any of several load metrics. For example, you can have AWS OpsWorks start instances when the average CPU utilization exceeds 80% and stop instances when the average CPU load falls below 60%.

Both time-based and load-based instances are supported for Windows and Linux stacks.

Unlike **24/7 instances**, which you must start and stop manually, you do not start or stop time-based or load-based instances yourself. Instead, you configure the instances and AWS OpsWorks starts or stops them based on their configuration. For example, you configure time-based instances to start and stop on a specified schedule. AWS OpsWorks then starts and stops the instances according to that configuration."

---

Is it possible for an OpsWorks instance to belong to more than one layer?

433

- 
- Yes (x)
  - No

From the OpsWorks User Guide (<http://docs.aws.amazon.com/opsworks/latest/userguide/workinginstances-add.html>):

"If an instance belongs to multiple layers, AWS OpsWorks runs the recipes for each of the instance's layers when a lifecycle event occurs, or when you run a stack or deployment command. However, an instance's layers must be compatible with one another. For example, the HAProxy layer is not compatible with the Static Web Server layer because they both bind to port 80. An instance can be a member of only one of those layers. For more information, see Appendix A: AWS OpsWorks Layer Reference."

---

Is it possible to register an externally-created instance (e.g., created via EC2 or on-premise) to an OpsWorks stack?

434

- 
- Yes
  - Yes, but only for Linux instances (x)
  - Yes, but only for Windows instances
  - No

From the OpsWorks User Guide (<http://docs.aws.amazon.com/opsworks/latest/userguide/registered-instances.html>):

"Instances describes how to use AWS OpsWorks to create and manage groups of Amazon Elastic Compute Cloud (Amazon EC2) instances. You can also incorporate Linux computing resources into a stack that was created outside of AWS OpsWorks:

- Amazon EC2 instances that you created directly by using the Amazon EC2 console, CLI, or API.
- *On-premises* instances running on your own hardware, including instances running in virtual machines."

---

Using OpsWorks, where do you define the "Automatically Assign IP Addresses" properties (public and/or EIP)?

435

- Stack
- Layer (x)
- Instance
- App

This is in a page called "Editing an OpsWorks Layer's Configuration" from the OpsWorks User Guide (<http://docs.aws.amazon.com/opsworks/latest/userguide/workinglayers-basics-edit.html#workinglayers-basics-edit-network>):

#### "Automatically Assign IP Addresses

**You can control whether AWS OpsWorks automatically assigns public or Elastic IP addresses to the layer's instances.** Here's what happens when you enable this option:

- For instance store-backed instances, AWS OpsWorks automatically assigns an address each time the instance is started.
- For Amazon EBS-backed instances, AWS OpsWorks automatically assigns an address when the instance is started for the first time.
- If an instance belongs to more than one layer, AWS OpsWorks automatically assigns an address if you have enabled automatic assignment for at least one of the layers,

#### Note

If you enable automatic assignment of public IP addresses, it applies only to new instances. AWS OpsWorks cannot update the public IP address for existing instances.

If your stack is running in a VPC, you have separate settings for public and Elastic IP addresses. The following table explains how these interact:

		Public IP addresses	
		Yes	No
Elastic IP addresses	Yes	Instances receive an Elastic IP address when they are started for the first time, or a public IP address if an Elastic IP cannot be assigned.	Instances receive an Elastic IP address when they are started for the first time.
	No	Instances receive a public IP address each time they are started.	Instances receive only a private IP address, which is not accessible from outside the VPC.

#### Note

Instances must have a way to communicate with the AWS OpsWorks service, Linux package repositories, and cookbook repositories. If you specify no public or Elastic IP address, your VPC must include a component such as a NAT that allows the layer's instances to communicate with external sites. For more information, see [Running a Stack in a VPC](#).

If your stack is not running in a VPC, Elastic IP addresses is your only setting:

- Yes: Instances receive an Elastic IP address when they are started for the first time, or a public IP address if an Elastic IP address cannot be assigned.
- No: Instances receive a public IP address each time they are started."

#### What is an OpsWorks Lifecycle Event?

436

- A layer-specific trigger that you can use to run event-specific recipes (x)
- A CloudWatch monitoring event
- A CloudTrail log stream

From the OpsWorks User Guide (<http://docs.aws.amazon.com/opsworks/latest/userguide/workingcookbook-events.html>):

"Each layer has a set of five lifecycle events, each of which has an associated set of recipes that are specific to the layer. When an event occurs on a layer's instance, AWS OpsWorks automatically runs the appropriate set of recipes."

#### Which of the following are valid OpsWorks lifecycle events?

437

- 
- Setup (x)
  - Configure (x)
  - Deploy (x)
  - Undeploy (x)
  - Shutdown (x)
  - Unstable
  - Error
  - Ready

These are discussed in detail here: <http://docs.aws.amazon.com/opsworks/latest/userguide/workingcookbook-events.html>, but here's a basic summary:

**Setup** – Occurs on a new instance after it successfully boots; also calls Deploy

**Configure** – Occurs on all of the stack's instances when an instance enters or leaves the online state

**Deploy** – Occurs when you deploy an app

**Undeploy** – Occurs when you delete an app

**Shutdown** – Occurs when you stop an instance; sent about 45 seconds before the backend actually terminates the instance to allow for a clean shutdown

---

Which OpsWorks lifecycle event is invoked on all of a stack's instance whenever an instance leaves or enters the online state? 438

- 
- Setup
  - Configure (x)
  - Deploy
  - Undeploy
  - Shutdown

These are discussed in detail here: <http://docs.aws.amazon.com/opsworks/latest/userguide/workingcookbook-events.html>, but here's a basic summary:

**Setup** – Occurs on a new instance after it successfully boots; also calls Deploy

**Configure** – Occurs on all of the stack's instances when an instance enters or leaves the online state

**Deploy** – Occurs when you deploy an app

**Undeploy** – Occurs when you delete an app

**Shutdown** – Occurs when you stop an instance; sent about 45 seconds before the backend actually terminates the instance to allow for a clean shutdown

---

Does rebooting an instance trigger an OpsWorks lifecycle event? 439

- 
- Yes
  - No (x)

From the OpsWorks User Guide (<http://docs.aws.amazon.com/opsworks/latest/userguide/workingcookbook-events.html>):

"Rebooting an instance does not trigger any lifecycle events."

---

When is the Setup OpsWorks lifecycle event triggered? 440

- 
- After an instance finishes booting (x)
  - After the Setup stack command is manually executed (x)
  - After an instance is restarted
  - After an instance has downloaded all of its recipes
  - Whenever the Deploy command is run on any instance anywhere in the stack

From the OpsWorks User Guide (<http://docs.aws.amazon.com/opsworks/latest/userguide/workingcookbook-events.html>):

"This event occurs after a started instance has finished booting. You can also manually trigger the Setup event by using the Setup stack command."

---

When is the Deploy OpsWorks lifecycle event triggered?

441

- 
- Whenever the Deploy app command is run (x)
  - After the Setup lifecycle event (x)
  - After an instance is restarted
  - After an instance has downloaded all of its recipes

From the OpsWorks User Guide (<http://docs.aws.amazon.com/opsworks/latest/userguide/workingcookbook-events.html>):

"This event occurs when you run a Deploy command, typically to deploy an application to a set of application server instances. The instances run recipes that deploy the application and any related files from its repository to the layer's instances. For example, for a Rails Application Server instances, the Deploy recipes check out a specified Ruby application and tell Phusion Passenger to reload it. You can also run Deploy on other instances so they can, for example, update their configuration to accommodate the newly deployed app.

Note: Setup includes Deploy; it runs the Deploy recipes after setup is complete."

---

When is the Undeploy OpsWorks lifecycle event triggered?

442

- 
- Whenever the Undeploy app command is run (x)
  - When an app is deleted (x)
  - Whenever a custom recipe is redeployed to an instance
  - Just prior to a Deploy event

From the OpsWorks User Guide (<http://docs.aws.amazon.com/opsworks/latest/userguide/workingcookbook-events.html>):

"This event occurs when you delete an app or run an Undeploy command to remove an app from a set of application server instances. The specified instances run recipes to remove all application versions and perform any required cleanup."

---

When is the Shutdown OpsWorks lifecycle event triggered?

443

- 
- Just after an instance shut down is requested, but before the instance is actually terminated (x)
  - When an instance is deleted
  - When an app is stopped

The second option is false and a bit of a trick: you can't delete a running instance, you have to stop it first...at which point the Shutdown event would have already fired.

From the OpsWorks User Guide (<http://docs.aws.amazon.com/opsworks/latest/userguide/workingcookbook-events.html>):

"This event occurs after you direct AWS OpsWorks to shut an instance down but before the associated Amazon EC2 instance is actually terminated. AWS OpsWorks runs recipes to perform cleanup tasks such as shutting down services.

If you have attached an Elastic Load Balancing load balancer to the layer and enabled support for connection draining, AWS OpsWorks waits until connection draining is complete before triggering the Shutdown event.

After triggering a Shutdown event, AWS OpsWorks allows Shutdown recipes a specified amount of time to perform their tasks, and then stops or terminates the Amazon EC2 instance. The default Shutdown timeout value is 120 seconds. If your Shutdown recipes might require more time, you can edit the layer configuration to change the timeout value. "

---

When is the Configure OpsWorks lifecycle event triggered?

444

- After a Setup event (x)
- You associate or disassociate an EIP with an instance (x)
- You attach or detach an ELB to a layer (x)
- After the Configure stack command is manually executed (x)
- After an instance is rebooted
- Before any recipes are executed

So I verified the "After a Setup event" case myself. If you read below, I think that's because running the Setup event first takes the instance offline and then puts it back online, thus triggering a Configure event as described

From the OpsWorks User Guide (<http://docs.aws.amazon.com/opsworks/latest/userguide/workingcookbook-events.html>):

"This event occurs on all of the stack's instances when one of the following occurs:

- An instance enters or leaves the online state.
- You associate an Elastic IP address with an instance or disassociate one from an instance.
- You attach an Elastic Load Balancing load balancer to a layer, or detach one from a layer.

For example, suppose that your stack has instances A, B, and C, and you start a new instance, D. After D has finished running its setup recipes, AWS OpsWorks triggers the Configure event on A, B, C, and D.

If you subsequently stop A, AWS OpsWorks triggers the Configure event on B, C, and D. AWS OpsWorks responds to the Configure event by running each layer's Configure recipes, which update the instances' configuration to reflect the current set of online instances.

The Configure event is therefore a good time to regenerate configuration files. For example, the HAProxy Configure recipes reconfigure the load balancer to accommodate any changes in the set of online application server instances."

---

What is OpsWorks "auto healing"?

445

- A stack will automatically correct common configuration errors
- A layer will automatically add or remove instances in response to configurable events
- An instance will be automatically restarted if its agents stops communicating (x)
- App will automatically redeploy if health checks fail

From the OpsWorks User Guide (<http://docs.aws.amazon.com/opsworks/latest/userguide/welcome.html>):

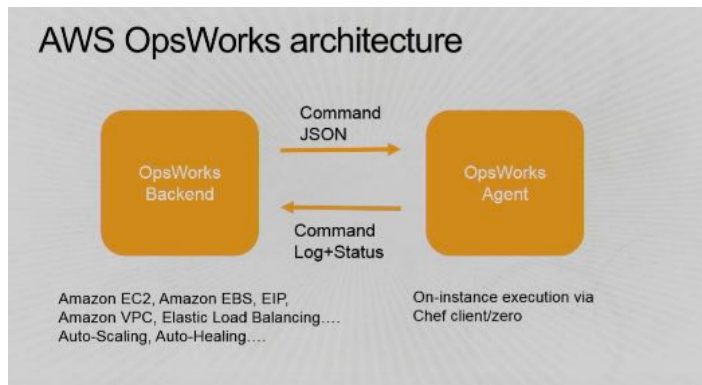
"AWS OpsWorks supports instance autohealing. If an agent stops communicating with the service, AWS OpsWorks automatically stops and restarts the instance."

What functions are performed by the OpsWorks agent?

446

- On-instance execution of recipes via Chef client/zero (x)
- Monitoring and event processing
- Communication with the OpsWorks back-end (x)
- Integration with the instance's auto scale group

From this re:Invent presentation (<https://www.youtube.com/watch?v=cuf7Rqlgeq4>):

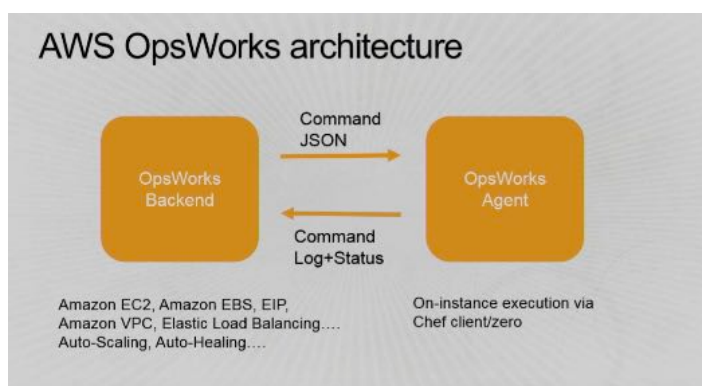


What functions are performed by the OpsWorks backend?

447

- Integration with other AWS services (EC2, EBS, etc) (x)
- Application lifecycle management
- Configuration management
- Real-time metrics collection

From this re:Invent presentation (<https://www.youtube.com/watch?v=cuf7Rqlgeq4>):





- Before
- After (x)

From the OpsWorks User Guide (<http://docs.aws.amazon.com/opsworks/latest/userguide/workingcookbook-assigningcustom.html>):

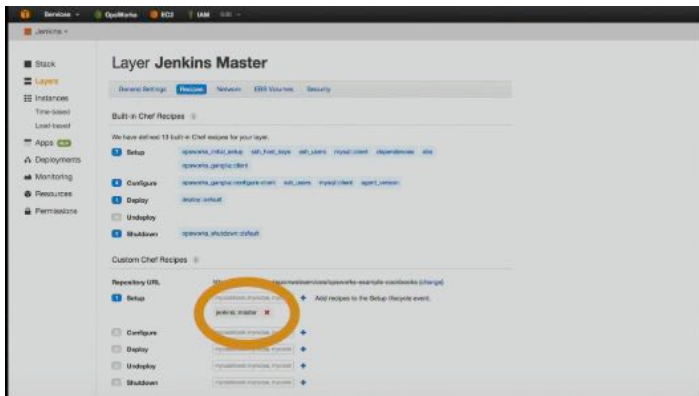
"If you have installed custom cookbooks, you can have AWS OpsWorks run some or all of the recipes automatically by assigning each recipe to a layer's lifecycle event. **After an event occurs, AWS OpsWorks runs the specified custom recipes after the layer's built-in recipes.**"

What would you need to do to add a Jenkins server to an existing OpsWorks stack?

449

- Create a new custom cookbook repository that specifies a new or existing cookbook to configure a Jenkins server (x)
- Create a new custom layer that specifies the Jenkins cookbook for the Setup lifecycle event (x)
- Create a new custom layer that specifies the Jenkins cookbook for the Configure lifecycle event
- Manually invoke a Setup command on the Jenkins instances
- Manually invoke a Setup command on the Jenkins layer

From this presentation (<https://www.youtube.com/watch?v=cuf7Rqlgeq4>):



Does OpsWorks have a command to upgrade the operating system on an instance?

450

- Yes, for all operating systems
- Yes, but only for Amazon Linux or RHEL (x)
- Yes, but only for Windows Server 2012 R2
- No

From the OpsWorks User Guide (<http://docs.aws.amazon.com/opsworks/latest/userguide/workingstacks-commands.html>):

447791

## Run AWS OpsWorks Stack Commands

[...]

### Upgrade Operating System

(Linux only) Upgrades the instances' Amazon Linux or RHEL operating systems to the latest version. For more information, see AWS OpsWorks Operating Systems.

Important

After running Upgrade Operating System, we recommend that you also run Setup. This ensures that services are correctly restarted.

447791

- Instances (x)
- Layers

This is defined below, but basically a Layer is collection of load-balanced instances and a template for those instances. The Apps are deployed to specific instances in those Layers, but not to the Layer itself, which is more akin to a collection of Chef recipes.

For example, a Ruby on Rails Layer would contain the recipes needed to properly configure new Rails servers. You would deploy Rails Apps onto one or more of those Layers using a Deployment, which would then trigger the Layer's Setup and Deploy recipes on those servers.

See the OpsWorks User Guide (<http://docs.aws.amazon.com/opsworks/latest/userguide/workingapps-deploying.html>):

"The primary purpose of [D]eployment is to deploy application code and related files to application server instances. The deployment operation is handled by each instance's Deploy recipes, which are determined by the instance's layer. The details depend on the particular layer, as follows."

and

"When you start an instance, after the Setup recipes complete, AWS OpsWorks automatically runs the instance's Deploy recipes."

and

"The deploy command triggers a Deploy event, which runs the deploy recipes on the selected instances. The deploy recipes for the associated application server download the code and related files from the repository and install them on the instance, so you typically select all of the associated application server instances."

- Instance
- Layer (x)

From the OpsWorks User Guide (<http://docs.aws.amazon.com/opsworks/latest/userguide/workinginstances-autohealing.html>):

"Every instance has an AWS OpsWorks agent that communicates regularly with the service. AWS OpsWorks uses that communication to monitor instance health. If an agent does not communicate with the service for more than approximately five minutes, AWS OpsWorks considers the instance to have failed.

[...]

An instance can be a member of multiple layers. If any of those layers has auto healing disabled, AWS OpsWorks does not heal the instance if it fails.

[...]

If a layer has auto healing enabled—the default setting—AWS OpsWorks automatically replaces the layer's failed instances..."

- Yes, by selecting specific instances when executing the event at the stack level (x)
- Yes, by passing the layer ID into the API or CLI
- Not possible

The documentation doesn't explicitly state that you can't choose to run a command at the layer level, but it's not possible. Here are the instructions on how to do it for a selection of stack instances from the OpsWorks User Guide (<http://docs.aws.amazon.com/opsworks/latest/userguide/workingcookbook-manual.html>):

Although recipes are typically run automatically in response to lifecycle events, **you can manually run recipes at any time on any or all stack instances**. This feature is typically used for tasks that don't naturally map to a lifecycle event, such as backing up instances. To run a custom recipe manually, it must be in one of your custom cookbooks but does not have to be assigned to a lifecycle event. When you run a recipe manually, AWS OpsWorks installs the same deploy attributes that it does for a Deploy event.

#### To manually run recipes on stack instances

1. On the Stack page, click Run command. For Command, select Execute Recipes.

#### Run Command

2. Enter the recipes to be run in the Recipes to execute box by using the standard cookbookname :: recipename format. Use commas to separate multiple recipes; they will run in the order that you list them.
3. Optionally, use the Custom Chef JSON box to add a custom JSON object that defines custom attributes that will be merged into the stack configuration and deployment attributes that are installed on the instances. For more information about using custom JSON objects, see Using Custom JSON and Overriding Attributes.
4. Under Instances, select the instances on which AWS OpsWorks should run the recipes.

When a lifecycle event occurs, the AWS OpsWorks agent receives a command to run the associated recipes. You can manually run these commands on a particular instance by using the appropriate stack command or by using the agent CLI's `run_command` command. For more information on how to use the agent CLI, see Appendix B: AWS OpsWorks Agent CLI.

## Direct Connect

What speeds are available through standard Direct Connect configurations?

454

- 
- 1 GB (x)
  - 10 GB (x)
  - 100 GB

From the Direct Connect User Guide (<http://docs.aws.amazon.com/directconnect/latest/UserGuide/Welcome.html>):

**"AWS Direct Connect links your internal network to an AWS Direct Connect location over a standard 1 gigabit or 10 gigabit Ethernet fiber-optic cable.** One end of the cable is connected to your router, the other to an AWS Direct Connect router. With this connection in place, you can create virtual interfaces directly to the AWS cloud (for example, to Amazon Elastic Compute Cloud (Amazon EC2) and Amazon Simple Storage Service (Amazon S3)) and to Amazon Virtual Private Cloud (Amazon VPC), bypassing Internet service providers in your network path. "

---

Is it possible to use Direct Connect if your network is not already colocated with an existing Direct Connect facility?

455

- Yes (x)
- No

From the Direct Connect User Guide (<http://docs.aws.amazon.com/directconnect/latest/UserGuide/Welcome.html>):

"To use AWS Direct Connect, your network must meet one of the following conditions:

- Your network is colocated with an existing AWS Direct Connect location. For more information on available AWS Direct Connect locations, go to <http://aws.amazon.com/directconnect/>.
- **You are working with an AWS Direct Connect partner** who is a member of the AWS Partner Network (APN). For a list of AWS Direct Connect partners who can help you connect, go to <http://aws.amazon.com/directconnect>.
- **You are working with an independent service provider to connect to AWS Direct Connect."**

---

Does Direct Connect require that the customer network support BGP?

456

- Yes (x)
- No

From the Direct Connect User Guide (<http://docs.aws.amazon.com/directconnect/latest/UserGuide/Welcome.html>):

"[...] your network must meet the following conditions:

- Connections to AWS Direct Connect require single mode fiber, 1000BASE-LX (1310nm) for 1 gigabit Ethernet, or 10GBASE-LR (1310nm) for 10 gigabit Ethernet. You must support 802.1Q VLANs across these connections.
- **Your network must support Border Gateway Protocol (BGP) and BGP MD5 authentication.** Optionally, you may configure Bidirectional Forwarding Detection (BFD)"

---

Is a VPG required to connect a Direct Connect connection to a VPC?

457

- Yes (x)
- No

From the Direct Connect User Guide (<http://docs.aws.amazon.com/directconnect/latest/UserGuide/Welcome.html>):

"To connect to Amazon Virtual Private Cloud (Amazon VPC), you must first do the following:

- Provide a private Autonomous System Number (ASN). Amazon allocates a private IP address in the 169.x.x.x range to you.
- **Create a virtual private gateway and attach it to your VPC.** For more information about creating a virtual private gateway, see Adding a Hardware Virtual Private Gateway to Your VPC in the *Amazon VPC User Guide*"

---

What are some of the benefits of Direct Connect?

458

- 
- Reduce costs when using large volumes of traffic (x)
  - Increased reliability (x)
  - Increased bandwidth (x)
  - Consistent network performance (QoS, low jitter) (x)
  - Enhanced security (x)
  - Simpler network topologies

You start getting cost benefits when transferring GB/s of data—you won't get a lot of savings with lower throughput

You get enhanced security since traffic doesn't flow over the public internet.

Collected from a variety of sources, including Ryan's course:

## Direct Connect Benefits

- Reduce costs when using large volumes of traffic
- Increase reliability
- Increase bandwidth

This re:Invent lecture: <https://www.youtube.com/watch?v=iJeiWU9Ud7w>

and the Direct Connect FAQ (<https://aws.amazon.com/directconnect/faqs/>):

### **"Q. What are the benefits of using AWS Direct Connect and private network connections?"**

In many circumstances, private network connections can reduce costs, increase bandwidth, and provide a more consistent network experience than Internet-based connections."

Is it possible to get sub-1GB Direct Connect connections?

459

- Yes (x)
- No

From the Direct Connect FAQ (<https://aws.amazon.com/directconnect/faqs/>):

### **"Q. What connection speeds are supported by AWS Direct Connect?"**

1Gbps and 10Gbps ports are available. Speeds of 50Mbps, 100Mbps, 200Mbps, 300Mbps, 400Mbps, and 500Mbps can be ordered from any APN partners supporting AWS Direct Connect. Read more about APN Partners supporting AWS Direct Connect."

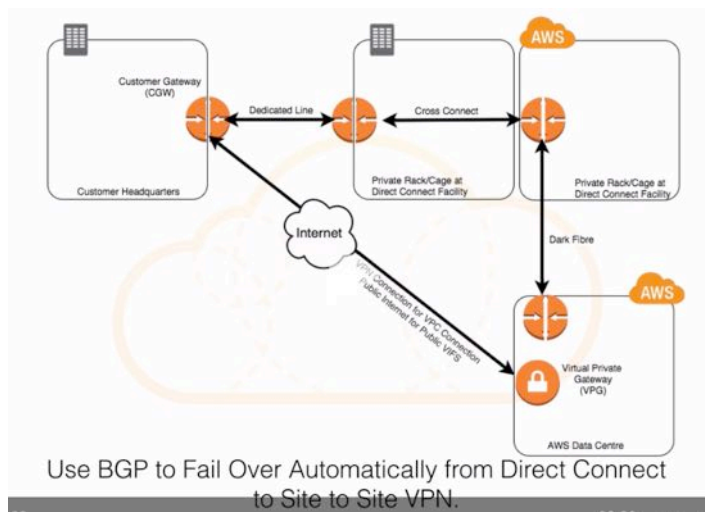
Is it possible to automatically failover from a site-to-site VPN to a Direct Connect connection?

460

- Yes (x)
- No

BGP can do this automatically.

Good diagram from Ryan's course:



What AWS regions can your on-premises networks connect to using Direct Connect?

461

- All regions
- All US regions for US-based connections(x)
- The region nearest to the Direct Connect facility. (x)
- A region selected during configuration

From the Direct Connect FAQ (<https://aws.amazon.com/directconnect/faqs/>):

**"Q. What AWS region(s) can I connect to via this connection?**

Each AWS Direct Connect location enables connectivity to the geographically nearest AWS region. You can access all AWS services available in that region.

Direct Connect locations in the US can also access the public endpoints of the other AWS regions using a public virtual interface."

Are connections to Direct Connect redundant?

462

- Yes
- No (x)

From the Direct Connect FAQ (<https://aws.amazon.com/directconnect/faqs/>):

**"Q. Are connections to AWS Direct Connect redundant?**

Each connection consists of a single dedicated connection between ports on your router and an Amazon router. We recommend establishing a second connection if redundancy is required. When you request multiple ports at the same AWS Direct Connect location, they will be provisioned on redundant Amazon routers."

Is it possible to connect to the public internet using a Direct Connect connection?

463

- 
- Yes
  - No (x)

No, but you can get to AWS public endpoints using a public virtual interface (VIF).

From the Direct Connect FAQ (<https://aws.amazon.com/directconnect/faqs/>):

**"Q: Can I connect to the Internet via this connection?"**

No."

and

**"Q. Can I use the same private network connection with Amazon Virtual Private Cloud (VPC) and other AWS services simultaneously?"**

Yes. Each AWS Direct Connect connection can be configured with one or more virtual interfaces. **Virtual interfaces may be configured to access AWS services such as Amazon EC2 and Amazon S3 using public IP space**, or resources in a VPC using private IP space."

---

How do you establish connectivity between a VPC and your on-premises network over a Direct Connect link?

464

- Using a private VIF (x)
- Using a public VIF
- Using a hosted VIF

From the Direct Connect User Guide (<http://docs.aws.amazon.com/directconnect/latest/UserGuide/WorkingWithVirtualInterfaces.html>):

"You must create a virtual interface to begin using your AWS Direct Connect connection. You can create a public virtual interface to connect to public resources, **or a private virtual interface to connect to your VPC**. You can configure multiple virtual interfaces on a single AWS Direct Connect connection, and **you'll need one private virtual interface for each VPC to connect to**. Each virtual interface needs a VLAN ID, interface IP address, ASN, and BGP key.

To use your AWS Direct Connect connection with another AWS account, you can create a hosted virtual interface for that account. These hosted virtual interfaces work the same as standard virtual interfaces and can connect to public resources or a VPC."

---

How do you establish connectivity between your on-premises network and AWS public endpoints over a Direct Connect link?

465

- Using a private VIF
- Using a public VIF (x)
- Using a hosted VIF

From the Direct Connect User Guide (<http://docs.aws.amazon.com/directconnect/latest/UserGuide/WorkingWithVirtualInterfaces.html>):

"You must create a virtual interface to begin using your AWS Direct Connect connection. **You can create a public virtual interface to connect to public resources**, or a private virtual interface to connect to your VPC. You can configure multiple virtual interfaces on a single AWS Direct Connect connection, and you'll need one private virtual interface for each VPC to connect to. Each virtual interface needs a VLAN ID, interface IP address, ASN, and BGP key.

To use your AWS Direct Connect connection with another AWS account, you can create a hosted virtual interface for that account. These hosted virtual interfaces work the same as standard virtual interfaces and can connect to public resources or a VPC."

---

How do I share a Direct Connect interlink between AWS accounts?

466

- 
- Using a private VIF
  - Using a public VIF
  - Using a hosted VIF (x)

From the Direct Connect User Guide (<http://docs.aws.amazon.com/directconnect/latest/UserGuide/WorkingWithVirtualInterfaces.html>):

"You must create a virtual interface to begin using your AWS Direct Connect connection. You can create a public virtual interface to connect to public resources, or a private virtual interface to connect to your VPC. You can configure multiple virtual interfaces on a single AWS Direct Connect connection, and you'll need one private virtual interface for each VPC to connect to. Each virtual interface needs a VLAN ID, interface IP address, ASN, and BGP key.

**To use your AWS Direct Connect connection with another AWS account, you can create a hosted virtual interface for that account. These hosted virtual interfaces work the same as standard virtual interfaces and can connect to public resources or a VPC."**

---

Does Direct Connect support layer 2 connections?

467

- Yes
- No (x)

From the Direct Connect FAQ (<https://aws.amazon.com/directconnect/faqs/>):

**"Q. Can I establish a Layer 2 connection between VPC and my network?**

No, Layer 2 connections are not supported."

## Hybrid Architecture

---

Which is the most hybrid connectivity option that best optimizes for both consistent performance and excellent security?

468

- VPN over Direct Connect (x)
- Direct Connect alone
- IPSec VPN (CGW/VGW)

See this re:Invent presentation: <https://www.youtube.com/watch?v=PPGXBoeLlcM>.

The presentation goes into detail, but basically AWS can't guarantee any security past its connection to the colocation facility. So you get great consistent performance using DirectConnect, but unverifiable security, and completely verifiable security with VPN but unpredictable performance. Using VPN over Direct Connect gives you the best of both worlds.

## SNS

---

Can SNS push notification to mobile devices ("Mobile Push")?

469

- Yes (x)
- No

From the SNS Developer Guide (<http://docs.aws.amazon.com/sns/latest/dg/SNSMobilePush.html>):

**"With Amazon SNS, you have the ability to send push notification messages directly to apps on mobile devices. Push notification messages sent to a mobile endpoint can appear in the mobile app as message alerts, badge updates, or even sound alerts."**



- A configuration where WAFs are installed on EC2 instances as an embedded reverse proxy
- A configuration where a fleet of autoscaled and load balanced WAFs are deployed in a public subnet routes inspected and filtered traffic to an internal load balancer in a private subnet (x)
- Multiple layers of WAFs successively layered in a public subnet

From this white paper ([https://d0.awsstatic.com/whitepapers/DDoS\\_White\\_Paper\\_June2015.pdf](https://d0.awsstatic.com/whitepapers/DDoS_White_Paper_June2015.pdf)):

In order to inspect all HTTP requests, WAFs sit in-line with your application traffic. Unfortunately, this creates a scenario where WAFs can become a point of failure or bottleneck. To mitigate this problem, you'll need the ability to run multiple WAFs on demand during traffic spikes. This type of scaling for WAF is done via a “WAF sandwich.”

In the “WAF sandwich,” the EC2 instance running your WAF software is included in an Auto Scaling group and placed in between two ELB load balancers. Recall from the Elastic Load Balancing section, you created two load balancers: a basic load balancer in the default VPC, and an internal load balancer. The basic load balancer in your default VPC will be the frontend, public facing load balancer that will distribute all incoming traffic to the WAF EC2 instance. By running the WAF EC2 instance in an Auto Scaling group behind ELB, the instance can scale and add additional WAF EC2 instances should the traffic spike to elevated levels.

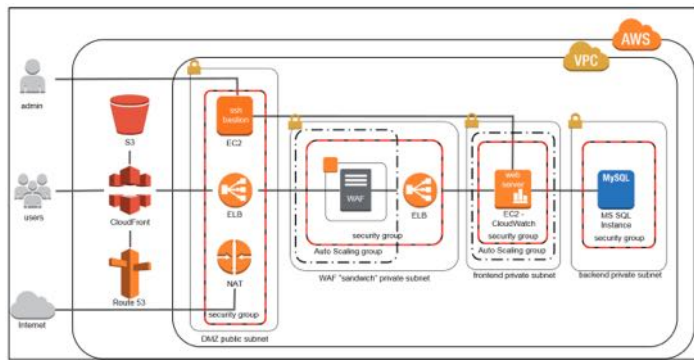


Figure 4: DDoS Resilient Reference Architecture

STS

Which of the following API calls return temporary security credentials?

- AssumeRole (x)
- AssumeRoleWithSAML (x)
- AssumeRoleWithWebIdentity (x)
- GetFederationToken (x)
- GetSessionToken (x)
- GetTemporaryCredentials
- Authenticate

From the IAM User Guide ([http://docs.aws.amazon.com/IAM/latest/UserGuide/id\\_credentials\\_temp\\_request.html](http://docs.aws.amazon.com/IAM/latest/UserGuide/id_credentials_temp_request.html)):

447777

The following table compares features of the actions (APIs) in AWS STS that return temporary security credentials.

AWS STS API	Who can call	Credential lifetime (min/max/default)	MFA support*	Passed policy support*	Restrictions on resulting temporary credentials
<a href="#">AssumeRole</a>	IAM user or user with existing temporary security credentials	15m/1hr/1hr	Yes	Yes	Cannot call <a href="#">GetFederationToken</a> Or <a href="#">GetSessionToken</a> .
<a href="#">AssumeRoleWithSAML</a>	Any user; caller must pass a SAML authentication response that indicates authentication from a known identity provider	15m/1hr/1hr	No	Yes	Cannot call <a href="#">GetFederationToken</a> Or <a href="#">GetSessionToken</a> .
<a href="#">AssumeRoleWithWebIdentity</a>	Any user; caller must pass a web identity token that indicates authentication from a known identity provider	15m/1hr/1hr	No	Yes	Cannot call <a href="#">GetFederationToken</a> Or <a href="#">GetSessionToken</a> .
<a href="#">GetFederationToken</a>	IAM user or root account	IAM user: 15m/36hr/12hr Root account: 15m/1hr/1hr	No	Yes	Cannot call IAM APIs directly. SSO to console is allowed.* Cannot call AWS STS APIs.
<a href="#">GetSessionToken</a>	IAM user or root account	IAM user: 15m/36hr/12hr Root account: 15m/1hr/1hr	Yes	No	Cannot call IAM APIs unless MFA information is included with the request. Cannot call AWS STS APIs except <a href="#">AssumeRole</a> . Single sign-on (SSO) to console is not allowed, but any user with a password (root or IAM user) can sign into the console.*

447777

Which of the following STS APIs support MFA?

- AssumeRole (x)
- AssumeRoleWithSAML
- AssumeRoleWithWebIdentity
- GetFederationToken
- GetSessionToken (x)

From the IAM User Guide ([http://docs.aws.amazon.com/IAM/latest/UserGuide/id\\_credentials\\_temp\\_request.html](http://docs.aws.amazon.com/IAM/latest/UserGuide/id_credentials_temp_request.html)):

“””

The following table compares features of the actions (APIs) in AWS STS that return temporary security credentials.

AWS STS API	Who can call	Credential lifetime (min/max/default)	MFA support*	Passed policy support*	Restrictions on resulting temporary credentials
<a href="#">AssumeRole</a>	IAM user or user with existing temporary security credentials	15m/1hr/1hr	Yes	Yes	Cannot call <code>GetFederationToken</code> or <code>GetSessionToken</code> .
<a href="#">AssumeRoleWithSAML</a>	Any user; caller must pass a SAML authentication response that indicates authentication from a known identity provider	15m/1hr/1hr	No	Yes	Cannot call <code>GetFederationToken</code> or <code>GetSessionToken</code> .
<a href="#">AssumeRoleWithWebIdentity</a>	Any user; caller must pass a web identity token that indicates authentication from a known identity provider	15m/1hr/1hr	No	Yes	Cannot call <code>GetFederationToken</code> or <code>GetSessionToken</code> .
<a href="#">GetFederationToken</a>	IAM user or root account	IAM user: 15m/36hr/12hr  Root account: 15m/1hr/1hr	No	Yes	Cannot call IAM APIs directly.  SSO to console is allowed.*  Cannot call AWS STS APIs.
<a href="#">GetSessionToken</a>	IAM user or root account	IAM user: 15m/36hr/12hr  Root account: 15m/1hr/1hr	Yes	No	Cannot call IAM APIs unless MFA information is included with the request.  Cannot call AWS STS APIs except <code>AssumeRole</code> .  Single sign-on (SSO) to console is not allowed, but any user with a password (root or IAM user) can sign into the console.*

**MFA support.** You can include information about a multi-factor authentication (MFA) device when you call the `AssumeRole` and `GetSessionToken` APIs. This ensures that the temporary security credentials that result from the API call can be used only by users who are authenticated with an MFA device.

“””

Also see [http://docs.aws.amazon.com/IAM/latest/UserGuide/id\\_roles\\_providers\\_enable-console-saml.html](http://docs.aws.amazon.com/IAM/latest/UserGuide/id_roles_providers_enable-console-saml.html):

"You can use a role to configure your SAML 2.0-compliant IdP and AWS to permit your federated users to access the AWS Management Console. "

Which of the following STS APIs can be called by users that do *not* have AWS root or IAM credentials?

- AssumeRole
- AssumeRoleWithSAML (x)
- AssumeRoleWithWebIdentity (x)
- GetFederationToken
- GetSessionToken

From the IAM User Guide ([http://docs.aws.amazon.com/IAM/latest/UserGuide/id\\_credentials\\_temp\\_request.html](http://docs.aws.amazon.com/IAM/latest/UserGuide/id_credentials_temp_request.html)):

For the SAML and WebIdentity calls the caller "must pass a SAML authentication response [or web identity token] that indicates authentication from a known identity provider."

4777

The following table compares features of the actions (APIs) in AWS STS that return temporary security credentials.

AWS STS API	Who can call	Credential lifetime (min/max/default)	MFA support*	Passed policy support*	Restrictions on resulting temporary credentials
<a href="#">AssumeRole</a>	IAM user or user with existing temporary security credentials	15m/1hr/1hr	Yes	Yes	Cannot call <code>GetFederationToken</code> or <code>GetSessionToken</code> .
<a href="#">AssumeRoleWithSAML</a>	Any user; caller must pass a SAML authentication response that indicates authentication from a known identity provider	15m/1hr/1hr	No	Yes	Cannot call <code>GetFederationToken</code> or <code>GetSessionToken</code> .
<a href="#">AssumeRoleWithWebIdentity</a>	Any user; caller must pass a web identity token that indicates authentication from a known identity provider	15m/1hr/1hr	No	Yes	Cannot call <code>GetFederationToken</code> or <code>GetSessionToken</code> .
<a href="#">GetFederationToken</a>	IAM user or root account	IAM user: 15m/36hr/12hr  Root account: 15m/1hr/1hr	No	Yes	Cannot call IAM APIs directly.  SSO to console is allowed.*  Cannot call AWS STS APIs.
<a href="#">GetSessionToken</a>	IAM user or root account	IAM user: 15m/36hr/12hr  Root account: 15m/1hr/1hr	Yes	No	Cannot call IAM APIs unless MFA information is included with the request.  Cannot call AWS STS APIs except <code>AssumeRole</code> .  Single sign-on (SSO) to console is not allowed, but any user with a password (root or IAM user) can sign into the console.*

4777

How do you assume a role in an account in which you don't otherwise have access?

- 
- Ask the role's owner to modify the role's trust policy to trust your account or an identity provider with which you've registered (x)
  - Invoke one of the AssumeRole APIs with the role's ARN and use the temporary credentials to access AWS resources (x)
  - Create an AWS account (setting up federation if needed) and invoke the AssumeRole API with your credentials and the role's ARN

From the IAM User Guide ([http://docs.aws.amazon.com/IAM/latest/UserGuide/id\\_roles\\_use\\_switch-role-api.html](http://docs.aws.amazon.com/IAM/latest/UserGuide/id_roles_use_switch-role-api.html)):

"To assume a role, an application calls the AWS STS AssumeRole API and passes the ARN of the role to use. The AssumeRole API returns a set of temporary security credentials that you can use in subsequent AWS API calls to access resources in the account that owns the role. The temporary credentials have whatever permissions are defined in the role's access policy."

Also from the IAM User Guide ([http://docs.aws.amazon.com/IAM/latest/UserGuide/id\\_roles\\_create\\_for-user.html](http://docs.aws.amazon.com/IAM/latest/UserGuide/id_roles_create_for-user.html)):

"You can use IAM roles to delegate access to your AWS resources. With IAM roles, you can establish trust relationships between your trusting account and other AWS trusted accounts. The trusting account owns the resource to be accessed and the trusted account contains the users who need access to the resource. After you create the trust relationship, an IAM user or an application from the trusted account can use the AWS Security Token Service (AWS STS) AssumeRole API action to obtain temporary security credentials that enable access to AWS resources in your account. The accounts can both be controlled by you, or the account with the users can be controlled by a third party."

## CloudTrail

Is it possible to use CloudWatch to monitor CloudTrail events?

475

- 
- Yes (x)
  - No

From the CloudTrail User Guide (<http://docs.aws.amazon.com/awscloudtrail/latest/userguide/monitor-cloudtrail-log-files-with-cloudwatch-logs.html>):

**"One of the ways that you can work with CloudTrail logs is to monitor them in real time by sending them to CloudWatch Logs.** For a trail that is enabled in all regions in your account, CloudTrail sends log files from all those regions to a CloudWatch Logs log group. You define CloudWatch Logs metric filters that will evaluate your CloudTrail log events for matches in terms, phrases, or values. You assign CloudWatch metrics to the metric filters. You also create CloudWatch alarms that are triggered according to thresholds and time periods that you specify. You can configure an alarm to send a notification when the alarm is triggered so that you can take immediate action. You can also configure CloudWatch to automatically perform an action in response to an alarm. CloudTrail events are protected by SSL encryption as they are delivered from CloudTrail to the CloudWatch Logs log group."

Can CloudTrail log events from multiple accounts into a single S3 bucket?

476

- 
- Yes (x)
  - No

From the CloudTrail User Guide (<http://docs.aws.amazon.com/awscloudtrail/latest/userguide/cloudtrail-receive-logs-from-multiple-accounts.html>):

"You can have CloudTrail deliver log files from multiple AWS accounts into a single Amazon S3 bucket. For example, you have four AWS accounts with account IDs 111111111111, 222222222222, 333333333333, and 444444444444, and you want to configure CloudTrail to deliver log files from all four of these accounts to a bucket belonging to account 111111111111."

## Kinesis

Are Kinesis streams ordered?

477

- 
- Yes, by a customer-supplied key
  - Yes, by a Kinesis-supplied sequence number (x)
  - No, they are randomly ordered

From the Kinesis Developer Guide (<http://docs.aws.amazon.com/kinesis/latest/dev/key-concepts.html>):

"An *Amazon Kinesis stream* is an ordered sequence of data records. Each record in the stream has a sequence number that is assigned by Amazon Kinesis."

---

What is the maximum size of a Kinesis record's data blob?

478

- 
- 1 KB
  - 1 MB (x)
  - 1 GB
  - 1 TB

From the Kinesis Developer Guide (<http://docs.aws.amazon.com/kinesis/latest/dev/key-concepts.html>):

"A *data record* is the unit of data stored in an Amazon Kinesis stream. Data records are composed of a sequence number, partition key, and data blob, which is an immutable sequence of bytes. Amazon Kinesis does not inspect, interpret, or change the data in the blob in any way. **A data blob can be up to 1 MB.**"

---

What is the default retention period for Kinesis records?

479

- 
- 1 hour
  - 12 hours
  - 24 hours (x)
  - 7 days

From the Kinesis Developer Guide (<http://docs.aws.amazon.com/kinesis/latest/dev/key-concepts.html>):

"The length of time data records are accessible after they are added to the stream. **A stream's retention period is set to a default of 24 hours after creation.** You can increase the retention period up to 168 hours (7 days) using the `IncreaseRetentionPeriod` operation, and decrease the retention period down to a minimum of 24 hours using the `DecreaseRetentionPeriod` operation."

---

What is the maximum configurable retention period for Kinesis records?

480

- 
- 1 hour
  - 12 hours
  - 24 hours
  - 7 days (x)

From the Kinesis Developer Guide (<http://docs.aws.amazon.com/kinesis/latest/dev/key-concepts.html>):

"The length of time data records are accessible after they are added to the stream. A stream's retention period is set to a default of 24 hours after creation. **You can increase the retention period up to 168 hours (7 days)** using the `IncreaseRetentionPeriod` operation, and decrease the retention period down to a minimum of 24 hours using the `DecreaseRetentionPeriod` operation."

---

How does Kinesis determine the shard for a newly inserted record?

481

- 
- Using the MD5 hash of a user-specified partition key (x)
  - By choosing a shard at random
  - New records are inserted in round-robin fashion

From the Kinesis Developer Guide (<http://docs.aws.amazon.com/kinesis/latest/dev/key-concepts.html>):

"A *partition key* is used to group data by shard within a stream. Amazon Kinesis segregates the data records belonging to a stream into multiple shards, **using the partition key associated with each data record** to determine which shard a given data record belongs to. Partition keys are Unicode strings with a maximum length limit of 256 bytes. **An MD5 hash function is used to map partition keys to 128-bit integer values and to map associated data records to shards.** A partition key is specified by the applications putting the data into a stream."

---

What is the maximum length of a Kinesis partition key?

482

- 16 bytes
- 256 bytes (x)
- 1 kilobyte
- 16 kilobytes

From the Kinesis Developer Guide (<http://docs.aws.amazon.com/kinesis/latest/dev/key-concepts.html>):

"A *partition key* is used to group data by shard within a stream. Amazon Kinesis segregates the data records belonging to a stream into multiple shards, using the partition key associated with each data record to determine which shard a given data record belongs to. **Partition keys are Unicode strings with a maximum length limit of 256 bytes.** An MD5 hash function is used to map partition keys to 128-bit integer values and to map associated data records to shards. A partition key is specified by the applications putting the data into a stream."

---

What is the provisioned read capacity of a Kinesis shard?

483

- 5 TPS up to 2MB/s (x)
- 10 TPS up to 10MB/s
- 100 TPS up to 100MB/s
- 1000 TPS up to 1000MB/s

From the Kinesis Developer Guide (<http://docs.aws.amazon.com/kinesis/latest/dev/key-concepts.html>):

"A *shard* is a uniquely identified group of data records in an Amazon Kinesis stream. A stream is composed of multiple shards, each of which provides a fixed unit of capacity. **Each shard can support up to 5 transactions per second for reads, up to a maximum total data read rate of 2 MB per second** and up to 1,000 records per second for writes, up to a maximum total data write rate of 1 MB per second (including partition keys). The data capacity of your stream is a function of the number of shards that you specify for the stream. The total capacity of the stream is the sum of the capacities of its shards."

---

What is the provisioned write capacity of a Kinesis shard?

484

- 1K TPS up to 1MB/s including partition keys (x)
- 10K TPS up to 10MB/s including partition keys
- 100K TPS up to 100MB/s including partition keys
- 1000K TPS up to 1000MB/s including partition keys

From the Kinesis Developer Guide (<http://docs.aws.amazon.com/kinesis/latest/dev/key-concepts.html>):

"A *shard* is a uniquely identified group of data records in an Amazon Kinesis stream. A stream is composed of multiple shards, each of which provides a fixed unit of capacity. Each shard can support up to 5 transactions per second for reads, up to a maximum total data read rate of 2 MB per second and **up to 1,000 records per second for writes, up to a maximum total data write rate of 1 MB per second (including partition keys).** The data capacity of your stream is a function of the number of shards that you specify for the stream. The total capacity of the stream is the sum of the capacities of its shards."

## Directory Service

Does SimpleAD support MFA?

485

- Yes
- No (x)

Ryan mentions this in his lectures, but I can't find any documentation on it per se.

The AD Connector **does** support MFA via integration with a RADIUS server.

Does AD Connector support MFA?

486

- Yes (x)
- No

From the Directory Service Administrative Guide ([http://docs.aws.amazon.com/directoryservice/latest/ad-connector/connect\\_mfa.html](http://docs.aws.amazon.com/directoryservice/latest/ad-connector/connect_mfa.html)):

"You can enable multi-factor authentication for your AD Connector directory by performing the following procedure..."

## Redshift

Where does Redshift back up its data?

487

- S3 (x)
- Glacier
- Secondary cluster

From the Redshift FAQ (<https://aws.amazon.com/redshift/faqs/>):

**"Q: How does Amazon Redshift back up my data?"**

Amazon Redshift replicates all your data within your data warehouse cluster when it is loaded and also **continuously backs up your data to S3**. Amazon Redshift always attempts to maintain at least three copies of your data (the original and replica on the compute nodes and a backup in Amazon S3). Redshift can also asynchronously replicate your snapshots to S3 in another region for disaster recovery."

How long does Redshift retain backups by default?

488

- 1 day (x)
- 10 days
- 30 days
- 90 days

From the Redshift FAQ (<https://aws.amazon.com/redshift/faqs/>):

**"Q: How long does Amazon Redshift retain backups? Is it configurable?"**

By default, **Amazon Redshift retains backups for 1 day**. You can configure this to be as long as 35 days."

Does Redshift support Multi-AZ Deployments?

489



- 
- Yes
  - No (x)

From the Redshift FAQ (<https://aws.amazon.com/redshift/faqs/>):

**"Q: Does Amazon Redshift support Multi-AZ Deployments?"**

Currently, Amazon Redshift only supports Single-AZ deployments. You can run data warehouse clusters in multiple AZ's by loading data into two Amazon Redshift data warehouse clusters in separate AZs from the same set of Amazon S3 input files. In addition, you can also restore a data warehouse cluster to a different AZ from your data warehouse cluster snapshots."

---

Can Redshift copy snapshots across regions?

490

- 
- Yes, but to only a single target region (x)
  - Yes, to multiple target regions
  - No

From the Redshift Management Guide (<http://docs.aws.amazon.com/redshift/latest/mgmt/working-with-snapshots.html>):

**"Copying Snapshots to Another Region**

**You can configure Amazon Redshift to automatically copy snapshots (automated or manual) for a cluster to another region.** When a snapshot is created in the cluster's primary region, it will be copied to a secondary region; these are known respectively as the *source region* and *destination region*. By storing a copy of your snapshots in another region, you have the ability to restore your cluster from recent data if anything affects the primary region. **You can configure your cluster to copy snapshots to only one destination region at a time."**

---

Does Redshift support database encryption?

491

- 
- Yes (x)
  - No

From the Redshift Management Guide (<http://docs.aws.amazon.com/redshift/latest/mgmt/working-with-db-encryption.html>):

**"In Amazon Redshift, you can enable database encryption for your clusters to help protect data at rest.** When you enable encryption for a cluster, the data blocks and system metadata are encrypted for the cluster and its snapshots.

Encryption is an optional, immutable setting of a cluster. If you want encryption, you must enable it during the cluster launch process. If you want to go from an encrypted cluster to an unencrypted cluster or the other way around, you must unload your data from the existing cluster and reload it in a new cluster with the chosen encryption setting."

---

Is it possible to disable encryption on an existing Redshift cluster?

492

- 
- Yes
  - No (x)

From the Redshift Management Guide (<http://docs.aws.amazon.com/redshift/latest/mgmt/working-with-db-encryption.html>):

**"In Amazon Redshift, you can enable database encryption for your clusters to help protect data at rest.** When you enable encryption for a cluster, the data blocks and system metadata are encrypted for the cluster and its snapshots.

Encryption is an optional, **immutable setting** of a cluster. If you want encryption, you must enable it during the cluster launch process. If you want to go from an encrypted cluster to an unencrypted cluster or the other way around, you must unload your data from the existing cluster and reload it in a new cluster with the chosen encryption setting."

---

For which workloads you choose Redshift over RDS?

493

- 
- Analytics and reporting (x)
  - Workloads with very large data sets (x)
  - Workloads where analytics can't interfere with OLTP (x)

From the Redshift FAQ (<https://aws.amazon.com/redshift/faqs/>):

**"Q: When would I use Amazon Redshift vs. Amazon RDS?"**

Both Amazon Redshift and Amazon RDS enable you to run traditional relational databases such as MySQL, Oracle and SQL Server in the cloud while offloading database administration. Customers use Amazon RDS databases both for online-transaction processing (OLTP) and for reporting and analysis. Amazon Redshift harnesses the scale and resources of multiple nodes and uses a variety of optimizations to provide order of magnitude improvements over traditional databases for analytic and reporting workloads against very large data sets. Amazon Redshift provides an excellent scale-out option as your data and query complexity grows or if you want to prevent your reporting and analytic processing from interfering with the performance of your OLTP workload."

---

Which AWS services can be a source for the Redshift COPY command?

494

- 
- S3 (x)
  - EMR (x)
  - Remote SSH host via standard output (x)
  - DynamoDB (x)
  - Data Pipeline (x)
  - RDS
  - ElastiCache

From the Redshift FAQ (<https://aws.amazon.com/redshift/faqs/>):

**"You can load data into Amazon Redshift from a range of data sources including Amazon S3, Amazon DynamoDB, Amazon EMR, AWS Data Pipeline and or any SSH-enabled host on Amazon EC2 or on-premises.** Amazon Redshift attempts to load your data in parallel into each compute node to maximize the rate at which you can ingest data into your data warehouse cluster. For more details on loading data into Amazon Redshift please view our Getting Started Guide."