



Welcome to Module 13: Well-Architected Pillar 5 - Cost Optimization. This module covers the Cost Optimization pillar of the Well-Architected Framework. While this module discusses many best practices, it also talks about how to choose components based on cost, how to procure Amazon Elastic Compute Cloud-or Amazon EC2-instances for the lowest cost, and how to analyze or audit your resources for inefficient costs or budget overruns.

What's in This Module



- Part 1: Principles of the Cost-Optimization Pillar
- Part 2: Optimizing the Cost of Your Infrastructure
- Part 3: Dedicated Instances and Dedicated Hosts
- Part 4 Trusted Advisor
- Part 5: Optimizing Costs with Caching
- Part 6: AWS Cost Calculation Tools
- **Exercise 4:** Improve this Architecture
- Appendix A: Cost Optimization Questions

© 2018, Amazon Web Services, Inc. or its Affiliates. All rights reserved

In Part 1, you will learn the principles of the Cost Optimization pillar.

In Part 2, you will learn how to optimize the cost of your infrastructure.

In Part 3, you will learn about dedicated instances and dedicated hosts.

In Part 4, you will learn about AWS Trusted Advisor and its ability to help you manage costs.

In Part 5, you will learn about optimizing costs with caching.

In Part 6, you will explore several cost calculation tools.

Finally, you will end this module by working on an architecture improvement exercise.

Module Objectives



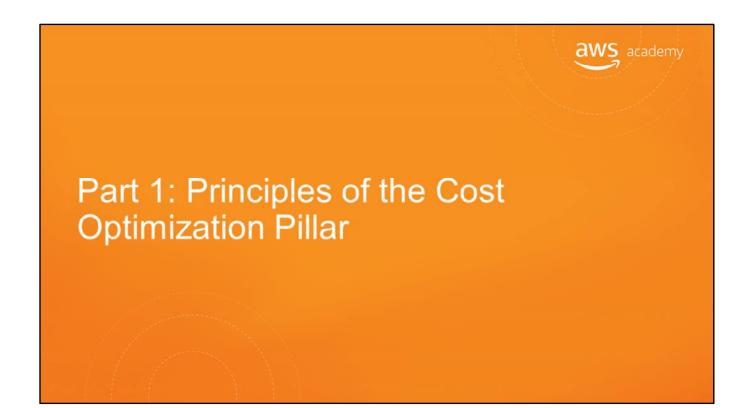
Upon completion of this module, you will be able to:

- Understand the principles of the cost optimization pillar.
- Discover how to optimize the costs of your infrastructure.
- Follow best practices to eliminate unneeded costs or suboptimal resources.

© 2018, Amazon Web Services, Inc. or its Affiliates. All rights reserved

Upon completion of this module, you will be able to:

- Understand the principles of the cost optimization pillar.
- Discover how to optimize the costs of your infrastructure.
- And follow best practices to eliminate unneeded costs or suboptimal resources.



In Part 1, we will review the principles of the Cost Optimization pillar.

Cost Optimization Principles



The ability to avoid or eliminate unneeded cost or suboptimal resources.

- Cost-effective resources
- Matched supply and demand
- Expenditure awareness
- Optimizing over time



© 2018, Amazon Web Services, Inc. or its Affiliates. All rights reserved.



In the cloud, you can follow several principles that help you achieve cost optimization. Cost optimization is the ability to avoid or eliminate unneeded costs or suboptimal resources. You want to make sure that the resources you are using are right-sized for the project so that they are cost-effective, and that resources are matched to supply and demand. You should also be aware of what you are spending, and also make sure that resources are optimized over time.

If you follow best practices for cost optimization, you should have a good cost comparison with on-premises. However, it's always possible to reduce your costs in the cloud as your applications and environments are migrated and mature over time. Cost optimization should not end until the cost of identifying money-saving opportunities is more than the amount of money you are actually going to save. Until that point is reached, you should continually monitor your expenditures and look for new ways to save on cost. For example, you should evaluate when new features are available.

Select the link to learn more about cost optimization in the AWS Well-Architected Framework whitepaper. https://d0.awsstatic.com/whitepapers/architecture/AWS Well-Architected Framework.pdf

Design Principles



- Adopt a consumption model.
- Measure overall efficiency.
- Stop spending money on data center operations.
- Analyze and attribute expenditures.
- Use managed services to reduce cost of ownership.



© 2018, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

In the cloud, you can follow a number of principles that help you achieve cost optimization.

- Adopt a consumption model: You can pay only for the computing resources that you consume, and increase or decrease usage depending on business requirements, not by using elaborate forecasting. For example, development and test environments are typically only used for 8 hours a day during the work week. You can stop these resources when they are not in use for a potential cost savings of 75 percent, or 40 hours versus 168 hours.
- Measure overall efficiency: You can measure the business output of the system and the costs that are associated with delivering it. Use this measure to understand the gains you make from increasing output and reducing costs.
- Stop spending money on data center operations: We do the heavy lifting of racking, stacking, and powering servers, so you can focus on your customers and business projects instead of IT infrastructure.
- Analyze and attribute expenditures: The cloud makes it more straightforward to accurately identify the usage and cost of systems, which then allows transparent attribution of IT costs to individual business owners. This helps measure return on investment—or ROI—and gives system owners an opportunity to optimize their resources and reduce costs.
- Use managed services to reduce cost of ownership: In the cloud, managed services remove the operational burden of maintaining servers for tasks like sending email or managing databases. And because managed services operate at cloud scale, they can offer a lower cost per transaction or service.



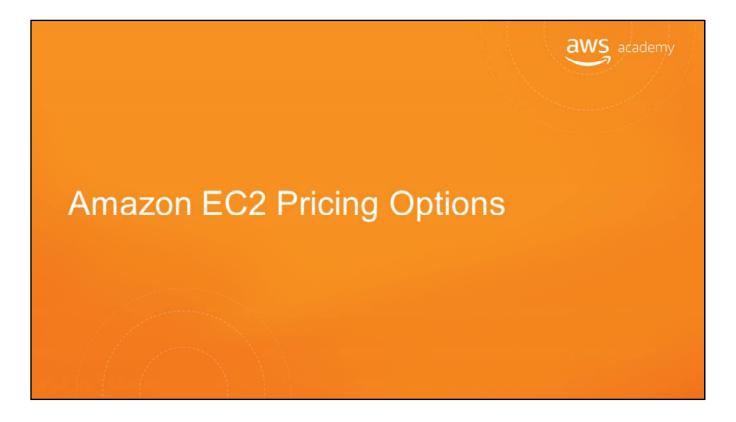
The key services for cost optimization are listed on this chart. The key AWS feature that supports cost optimization is cost allocation tags, which help you to understand the costs of a system. These tags are used to tag all resources from Amazon EC2 instances to Amazon Elastic Block Store—or Amazon EBS—volumes with cost center tags. These tags enable you to know which cost center is incurring the most AWS charges. This information can help you consolidate or trim down expenses.

The following services and features are important in the four areas of cost optimization:

- Cost-effective resources: You can use Reserved Instances and prepaid capacity to reduce your costs. You can use Cost Explorer to see patterns in how much you spend on AWS resources over time, identify areas that need further inquiry, and see trends that you can use to understand your costs.
- Matched supply and demand: AWS Automatic Scaling allows you to add or remove resources to match demand without overspending.
- Expenditure awareness: Amazon CloudWatch alarms and Amazon Simple Notification Service—or Amazon SNS—notifications will warn you if you go over, or are forecasted to go over, your budgeted amount.
- Optimizing over time: The AWS Blog and What's New section on the AWS website are resources for learning about newly launched features and services. AWS Trusted Advisor inspects your AWS environment and finds opportunities to save money by eliminating unused or idle resources, or committing to Reserved Instance capacity.



In Part 2, we will review how to optimize the cost of your infrastructure.



Let's review pricing options for Amazon EC2.

Amazon EC2 Purchasing Options On-Demand Instances Spot Instances Per-second Billing (Amazon Linux and Ubuntu only) Per-hour Billing (All other OSs) Per-hour Billing (All other OSs) Per-hour Billing Learn more.

As part of the AWS Free Tier, new AWS customers can get started with Amazon EC2 and some other services each month for up to one year after sign-up. What's available for free through the AWS Free Tier varies from service to service.

Amazon EC2 usage of Amazon Linux and Ubuntu-based instances that are launched in On-Demand, Reserved and Spot form will be billed on 1-second increments, with a minimum of 60 seconds.

On-Demand Instances let you pay for compute capacity with no long-term commitments or upfront payments.

Spot Instances allow you to purchase compute capacity with no upfront commitment, and through rates that are usually lower than the On-Demand rate. If the Spot Instance price moves higher than your maximum price, your instance will be shut down by Amazon EC2.

Dedicated Instances are Amazon EC2 instances that run in a virtual private cloud, or VPC, on hardware that's dedicated to a single customer. Your Dedicated Instances are physically isolated at the host hardware level from instances that belong to other AWS accounts. Dedicated Instance pricing has two components. The first component is an hourly per instance usage fee, and the second component is a dedicated per region fee. Note: You pay this once per hour regardless of how many Dedicated Instances you run.

Dedicated Hosts are physical Amazon EC2 servers with instance capacity that is fully dedicated for your use. Dedicated Hosts can help you reduce costs by allowing you to use your existing server-bound software licenses. Dedicated Hosts can be purchased On-Demand or as a Reserved Instance. Regardless of the quantity or the size of instances that you choose to launch on a

particular Dedicated Host, you pay hourly for each active Dedicated Host, and you are not billed for instance usage.

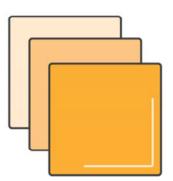
To learn more about AWS pricing, select the link to view the whitepaper on How AWS Pricing Works.

https://d0.awsstatic.com/whitepapers/aws pricing overview.pdf

Amazon EC2 On-Demand Instances aws academy



- Pay for compute capacity per second (Linux) or by the hour (all other OSs).
- No long-term commitments.
- No upfront payments.
- Increase or decrease your compute capacity depending on the demands of your application.



Amazon EC2 On-Demand Instances allow you to pay for compute capacity per second for Linux, or by the hour for all other operating systems, or OSs. There are no long term commitments and no upfront payments. You can increase or decrease your compute capacity depending on the demands of your application.

When you are first starting out, On-Demand Instances are a good option.

Amazon EC2 On-Demand Instances aws academy



Recommended for:

- Users who want the low cost and flexibility of Amazon EC2 without an upfront payment or long-term commitment.
- Applications with short term, spiky, or unpredictable workloads that cannot be interrupted.
- Applications being developed or tested on Amazon EC2 for the first time.

© 2018, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Amazon EC2 On-Demand Instances are recommended for:

- Users who want the low cost and flexibility of Amazon EC2 without an upfront payment or long-term commitment.
- Applications with short term, spiky, or unpredictable workloads that cannot be interrupted.
- Applications that are being developed or tested on Amazon EC2 for the first time.

Amazon EC2 Spot Instances



- Bid for unused AWS capacity.
- Prices controlled by AWS based on supply and demand.
- Termination Notice provided 2 minutes prior to termination, stored in metadata (now sent as a CloudWatch event as well as in instance metadata).



© 2018, Amazon Web Services, Inc. or its Affiliates. All rights reserved

With Amazon EC2 Spot Instances, you get to name your own prices for the same range of Amazon EC2 instance types that are available in On-Demand pricing by bidding for unused AWS capacity. Prices are controlled by AWS based on supply and demand.

At the end of the time billing period, you pay the market rate. If the market rate changes and goes above your maximum bid, you lose your compute resource, and our instance is terminated.

If your instance is marked for termination, the Termination Notice is stored in the instance's metadata 2 minutes before its termination time. The notice is accessible at http://169.254.169.254/latest/meta-data/spot/termination-time, and includes the time when the shutdown signal will be sent to the instance's operating system.

Do not use a spot instance for any type of workload that cannot be interrupted. Remember to use Amazon EBS for the backend of any Spot Instance so your data is not lost.

Amazon EC2 Spot Instances



- Hibernate, stop, or terminate Spot instances when they are interrupted.
- Best approach to temporary requests for large numbers of servers.
- Spot Fleet: a collection (or fleet) of Spot instances that is launched based on criteria that you specify.

2018, Amazon Web Services, Inc. or its Affiliates. All rights reserved.



You can specify whether Amazon EC2 should hibernate, stop, or terminate Spot Instances when they are interrupted. The default action is to terminate Spot Instances when they are interrupted. You can choose the interruption behavior that meets your needs.

The best approach is to include a temporary request when you need a fleet of servers. This is referred to as Spot Fleet, which is a collection—or fleet—of Spot Instances that is launched based on criteria that you specify.

Be sure to architect your infrastructure for interruption:

- Decouple components.
- Separate interactive processes and backend processes.
- Use frameworks, such as Amazon EMR.
- Design for interruption.
- Use Amazon Simple Queue Service, or Amazon SQS.
- Use Amazon Simple Workflow Service, or Amazon SWF.
- Place data in durable stores, such as Amazon S3 or Amazon DynamoDB.
- Save progress regularly.

Relevant applications on Spot Instances should poll for the termination notice at 5-second intervals, which gives the application almost the entire 2 minutes to complete any needed processing before the instance is terminated.

Select the link to view a blog post about the Spot instance Termination Notice. Notice: http://aws.amazon.com/blogs/aws/new-ec2-spot-instance-termination-notices/

Spot Use Cases



Use Case	Types of Applications	
Batch processing	Generic background processing (scale out computing)	
Web/data crawling	Analyze data	
Financial	Hedge fund analytics, energy trading, etc.	
Amazon EMR	Hadoop (large data processing)	
Grid computing	Scientific trials or simulations in chemistry, physics, and biology	
Transcoding	Transform videos into specific formats	
Gaming	Back-end servers for Facebook games	
Testing	Scale to large server pool to test software, websites, etc.	

© 2018, Amazon Web Services, Inc. or its Affiliates. All rights reserved

Although Spot Instances are terminated when the market price exceeds your bid price, Spot Instances are good for many different kinds of workloads.

Some example use cases for Spot Instances include:

- · Batch processing
- Web and data crawling to analyze data
- Financial workloads like hedge fund analytics, energy trading, and so on.
- Using Amazon EMR with Hadoop for large-volume data processing
- Grid computing for scientific trials; simulations in chemistry, physics and biology
- Transcoding to transform videos into specific formats
- Gaming, such as being used as backend servers for Facebook games
- Testing that enables scaling to a large server pool to test software, websites, and so on.

Amazon EC2 Reserved Instance Types academy

No Upfront

- Access a Reserved Instance without an upfront payment.
- 1-year or 3-year term (Standard and Convertible).

Up to 75% discount compared to
On-Demand Instance pricing.

Partial Upfront

- Part of the Reserved Instance must be paid at the start of the term.
- 1-year or 3-year term (Standard and Convertible).

All Upfront

- Full payment made at the start of the term.
- No other costs incurred for the remainder of the term, regardless of usage.
- 1-year or 3-year term (Standard and Convertible).

© 2018, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

There are three payment options with **Reserved Instances**:

- The **No Upfront** option does not require any upfront payments, and it provides a discounted rate for the duration of the term. 1-year or 3-year terms-both standard and convertible-are available.
- With **Partial Upfront**, you make a low upfront payment, and then you are charged a discounted rate for the instance for the duration of the Reserved Instance term. 1-year or 3-year terms-both standard and convertible-are available.
- With **All Upfront**, you pay for the entire Reserved Instance with one upfront payment. 1-year or 3-year terms-both standard and convertible-are available.

When you purchase Reserved Instances, note that you are required to specify the instance type, operating system, and Availability Zone. In addition, you are billed for the term, even when the instance is not in use. With Linux, you are billed per second, and for all other operating systems, you are billed for every hour. For situations that require only the limited use of instances, we recommend using On Demand Instances or Spot Instances instead.

Amazon EC2 Convertible Reserved Instance academy

- Change the instance family and other parameters associated with a Reserved Instance at any time.
- No limits.
- Use to take advantage of EC2 price reductions over time.
- No cost.
- Requirements:
 - Convertible Reserved Instance must be active.
 - Must be not pending a previous exchange request.

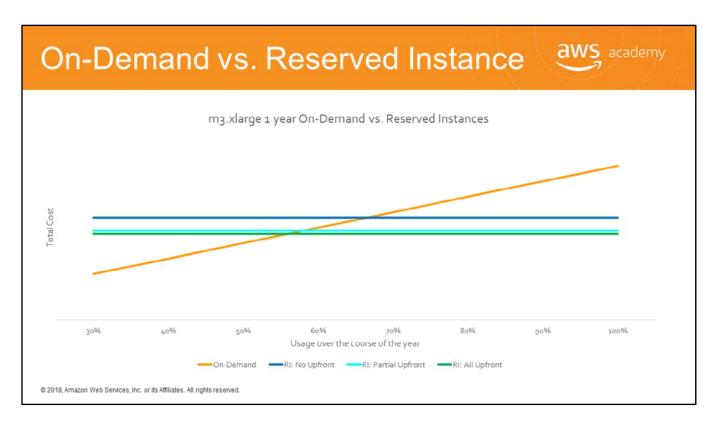
© 2018, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Amazon EC2 Convertible Reserved Instances are different from standard Amazon EC2 instances because they can be exchanged during the term for another Convertible Reserved Instance with new attributes. These new attributes include instance family, instance type, platform, scope, or tenancy. You can also modify some attributes of a Convertible Reserved Instance.

There are no limits to how many times you perform an exchange, as long as the target Convertible Reserved Instance is of an equal or higher value than the Convertible Reserved Instances that you are exchanging.

Exchanging Convertible Reserved Instances is free, but you might need to pay a true-up cost if the value is lower than the value of the Reserved Instances you're exchanging to.

For example, you can convert C3 Reserved Instances to C4 Reserved Instances to take advantage of a newer instance type, or convert C4 Reserved Instances to M4 Reserved Instances if your application needs more memory. You can also use Convertible Reserved Instances to take advantage of Amazon EC2 price reductions over time.



What are the break-even points of these various Reserved Instance purchasing options, compared to purchasing On-Demand Instances? Let's look at the following scenario: You decide that you need an m3.xlarge instance that runs Linux, and is located in the US-East Region. You wonder what level of usage—or the number of hours in a month that you will run that instance—you will need. Does it make sense to purchase a Reserved Instance, or should you purchase an On-Demand Instance?

In this scenario, if you are running the instance at least 60 percent of the time, you will begin saving money by purchasing an All Upfront Reserved Instance instead of an On-Demand Instance. If you are not in a position to purchase an All Upfront Reserved Instance, consider the Partial Upfront Reserved Instance or No Upfront Reserved Instance options where the break-even points are at approximately 65 percent usage and 72 percent usage, respectively.

In other words, if you run the instance at least 65 percent of the time, and you cannot make an All Upfront Instance purchase, you will save money over the year by purchasing a Partial Upfront Reserved Instance.

If you would rather make no upfront payments at all, but can commit to running this instance at least 72 percent of the time for at least a year, the No Upfront Reserved Instance payment option allows you to pay a low monthly fee to take advantage of savings over using an On-Demand Instance.

Reserved Instance Marketplace



- Flexibility
 - Sell your unused Amazon EC2 Reserved Instances.
 - Buy Amazon EC2 Reserved Instances from other AWS customers.
 - Change your Reserved Instances as your needs change.
- Diverse term and pricing options
 - Shorter terms.
 - Opportunity to save on upfront pricing.
- Identical capacity reservations

© 2018, Amazon Web Services, Inc. or its Affiliates. All rights reserved

Along with the implementation of the Reserved Instance pricing model, we also released the Reserved Instance Marketplace. So, if you have Reserved Instances that you no longer need, you can sell them in the Reserved Instance Marketplace. This applies to US-only bank accounts.

It has flexibility. You can sell your unused Amazon EC2 Reserved Instances to other businesses and organizations. You can also buy Amazon EC2 Reserved Instances from other AWS customers who are listed as third-party sellers. Third-party sellers on the Reserved Instance Marketplace provide a wide selection of term lengths and pricing options to choose from. Throughout the course of your term, your needs might change. You might need more or fewer instances than anticipated, or you might need to move an instance to a new AWS Region or change other options for your instance, such as instance type.

There are diverse term and pricing options. Buying Reserved Instances from the Reserved Instance Marketplace provides a wider selection of prices and terms than buying Reserved Instances directly from us. When you purchase a Reserved Instance from the Reserved Instance Marketplace, you will be taking over the third-party seller's original 1- or 3-year term. As such, you will only be required to fulfill the remainder of the term from the point at which you purchased the Reserved Instance from a third-party seller. In addition, the upfront cost is determined by the third-party seller, which provides you with the opportunity to spend less on your upfront costs compared to the price of a Reserved Instance that you purchase directly from us. However, usage or

recurring fees, such as monthly fees, will remain the same as the fees that were set when the Reserved Instances were originally purchased from us.

With identical capacity reservations, Reserved Instances from the Reserved Instance Marketplace, offer the same capacity reservations as Reserved Instances that are purchased directly from us.

You must have a US bank account to sell Reserved Instances on the Reserved Instance Marketplace. Support for non-US bank accounts will be coming soon.

Scheduled Instances



Reserved instances for predefined blocks of time on a recurring basis for a one-year term, with prices that are generally 5–10 percent lower than the equivalent On-Demand rates.

- Run reserved instances on a daily, weekly, or monthly basis.
- After you purchase your reserved instances, they are available to launch during the time windows that you've specified.
- Can be launched via the AWS Management Console, AWS CLI, AWS Tools for Windows PowerShell, and the RunScheduledInstances function.

© 2018, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Scheduled Instances are Reserved Instances that are provisioned for predefined blocks of time on a recurring basis for a 1-year term, with prices that are generally 5-10 percent lower than the equivalent On-Demand Instance rates. These instances can be run on a daily, weekly, or monthly basis.

For example, Reserved Instances could be used as Scheduled Instances for a certain batch job or a big software job that needs to happen at the end of every quarter. In this case, you can schedule Reserved Instances to take care of the job more quickly by using Amazon EC2 instances. With scheduling, you can take advantage of a specific block of time when you can run a Reserved Instance without having the server for up for 12 months out of the year to run a small batch job that happens only four times a year.

Scheduled Reserve Instance pricing is available if you purchase Reserved Instances on a recurring schedule, which allows you to pay for compute capacity per second for Linux, or by the hour for all other operating systems. You can obtain a capacity reservation ahead of time for the time periods that you will need it.

After you purchase your Reserved Instances, they are available to launch during the time windows that you specified. They can be launched via the AWS Management Console; the AWS Command Line Interface, or AWS CLI; AWS Tools for Windows PowerShell; and the RunScheduledInstances function.

Scheduled Instances

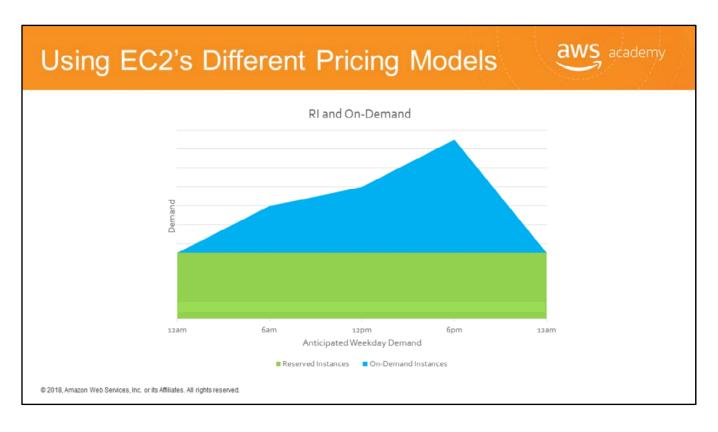


Use cases:

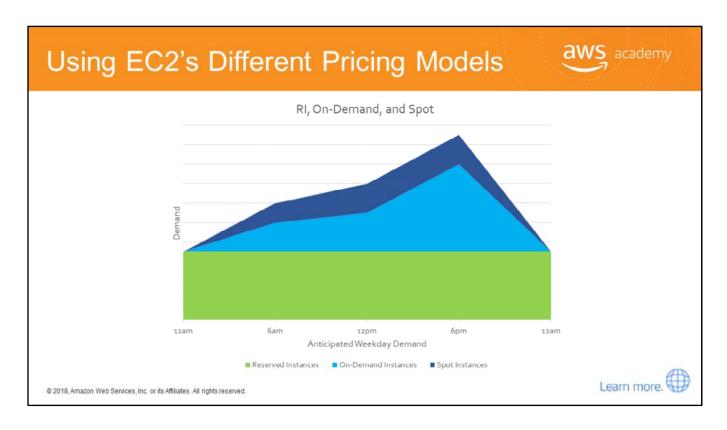
- A bank performs value at risk calculations every weekday afternoon.
- A phone company does a multi-day bill calculation run at the start of each month.
- A trucking company optimizes routes and shipments on Monday, Wednesday, and Friday mornings.
- An animation studio performs a detailed, compute-intensive 3D rendering every night.

© 2018, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

In addition to the batch job that happens quarterly, use cases for Scheduled Instances include risk calculations that are performed by a bank every weekday afternoon or multi-day bill calculations that are performed by a phone company at the start of each month. Other use cases include a trucking company that performs shipment optimization on Monday, Wednesday and Friday mornings; and an animation studio that performs a detailed, compute-intensive 3D rendering every night.



This chart illustrates a very basic way to use Reserved Instances and On-Demand Instances together to accommodate fluctuations in demand over time. In this example, multiple Reserved Instances were purchased and are running. However, as the day goes on and demand increases, the need for more instances increases as well. This customer supplements their capacity with On-Demand Instances, which can be shut down when they are not needed later in the evening.



In this second chart, this customer takes a more complex approach that attempts to use all three price models. In this case, some of the supplemental instances come first as On-Demand Instances, but further needs are addressed with Spot Instances. This allows the customer to save money by using Spot Instances over On-Demand Instances. However, this practice exposes the customer to unexpected instance termination because they would lose those instances if they are outbid. This situation could lead to lost data or insufficient capacity for their customers, which means that a model like this should only be implemented in circumstances when the sudden termination of instances is acceptable and is handled appropriately.

One AWS customer that has used all three models together is Pinterest. Select the link to learn more.

http://www.allthingsdistributed.com/2012/08/tco-and-return-on-agility.html

Blended Approach

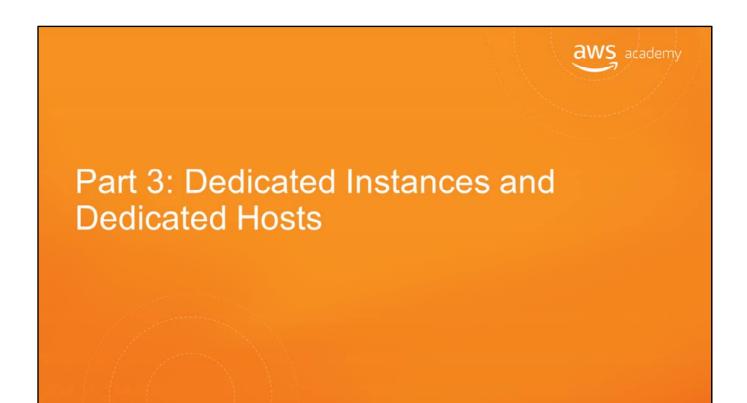


- Choose instance type that matches requirements.
 - Start with memory requirements and architecture type (32-bit or 64-bit).
 - Then, choose the closest number of virtual cores required.
- Scale across Availability Zones.
 - Smaller increments give more granularity for deploying to multiple AZs.
- Start with On-demand instances and then assess usage for Reserved Instances.

© 2018, Amazon Web Services, Inc. or its Affiliates. All rights reserved

A blended approach means choosing the instance type that matches the requirements. What are the requirements for memory and architecture type—that is, either 32-bit or 64-bit architectures? Choose the instance with the closest number of virtual cores that you require. Scaling across Availability Zones in smaller increments gives you more granularity for deploying to multiple Availability Zones, which is also known as a multi-AZ deployment. The more Availability Zones you have, the more flexible and highly available your application will be.

You can also start with On Demand instances, and then assess your usage to determine how many instances could be converted to Reserved Instances so that you can save money. If you know the baseline usage, using Reserved Instances can save you money over time because they offer discounted, or effective, rates. At seasonal peaks, you can use On-Demand Instances to handle the increased load.



In Part 3, we will discuss Dedicated Instances and Dedicated Hosts.

Dedicated Instances and Dedicated Hosts aws academy



Characteristic	Dedicated Instances	Dedicated Hosts
Dedicated physical servers	1	1
Per-instance billing (subject to a \$2 per Region fee)	✓	Х
Per-host billing	X	✓
Visibility of sockets, cores, host ID	×	✓
Affinity between a host and instance	X	✓
Targeted instance placement	X	✓
Automatic instance placement	✓	✓
Add capacity using an allocation request	Х	√

Let's take a moment to compare Dedicated Instances and Dedicated Hosts. One thing to notice about Dedicated Instances is that they are subject to per-instance billing, but Dedicated Hosts are not. However, if you want to add capacity by using an allocation request, you can do that for a Dedicated Host, but you cannot do this for the Dedicated Instance.

Dedicated Instances



- Single customer tenancy on compute hardware.
- Physically isolated at the host hardware level:
 - From other instances that are not dedicated instances.
 - From instances that belong to other AWS accounts.
- Useful for workloads with unique compliance requirements.
- May not be needed unless physical isolation is required.

© 2018, Amazon Web Services, Inc. or its Affiliates. All rights reserved

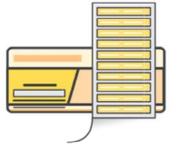
We provide various options for provisioning your Amazon EC2 instances such as On-Demand Instances, Reserved Instances, Spot Instances, and Dedicated Instances. Of these options, the Dedicated Instance option is very useful for customers with unique compliance requirements, like the Health Insurance and Portability and Accountability Act, or HIPAA.

Dedicated Instances are Amazon EC2 instances that run in a virtual private cloud, or VPC, on hardware that is dedicated to a single customer for additional isolation. These instances are physically isolated at the host hardware level from instances that are not Dedicated Instances and from instances that belong to other AWS accounts. Multiple Dedicated Instances for a single customer can share the same physical device.

Dedicated Hosts



- A dedicated host is a physical EC2 server with instance capacity fully dedicated for your use.
- Dedicated Hosts can help you:
 - Use your existing server-bound licenses
 - Meet compliance and regulatory requirements



© 2018 Amazon Web Services Inc. or its Affiliates. All rights reserved



A Dedicated Host is a physical Amazon EC2 server with instance capacity that is fully dedicated for your use. Dedicated Hosts can help you reduce costs by allowing you to use your existing server-bound software licenses, including Windows Server, SQL Server, and SUSE Linux Enterprise Server—all subject to your license terms. Dedicated Hosts can also help you meet compliance requirements, and they can be purchased On-Demand, or on an hourly basis. Reservations can provide up to a 70 percent discount compared to the On-Demand price.

Benefits of Dedicated Host include:

- Save Money on Licensing Costs: With Dedicated Hosts, you can save money by using your own per-socket or per-core software licenses in Amazon EC2.
- Help Meet Compliance and Regulatory Requirements: With Dedicated Hosts, you can place your instances in a VPC on a specific, physical server. This scenario enables you to deploy instances by using configurations that help address requirements for corporate compliance and regulations.

Select the link to learn more about Dedicated Hosts. https://aws.amazon.com/ec2/dedicated-hosts/



Part 4: AWS Trusted Advisor

In Part 4, we'll review AWS Trusted Advisor, which is another tool to help you make sure your instance are right-sized, and that you are not spending money that you don't need to.

Using AWS Trusted Advisor



AWS Trusted Advisor provides best practices (checks) in:

- Cost optimization
- Security
- Fault tolerance
- Performance improvement
- Service checks

AWS Trusted Advisor checks are available with Business and Enterprise Support plans

© 2018, Amazon Web Services, Inc. or its Affiliates. All rights reserved

AWS Trusted Advisor comes with every AWS account, and it can be accessed from the console. AWS Trusted Advisor helps you provision your resources by following best practices to improve system performance and reliability, increase security, and look for opportunities to save money.

AWS Trusted Advisor is like your customized cloud expert. It provides four of the most popular performance and security recommendations to all AWS customers, including:

- Cost optimization
- Security
- Fault tolerance
- Performance improvement; and
- Service checks

AWS Trusted Advisor checks and additional guidance is available with Business and Enterprise Support plans.

Using AWS Trusted Advisor



Best practices available to all customers:

- Service Limits
- Security Groups Specific Ports Unrestricted
- IAM Use
- MFA on Root Account
- Amazon S3 Bucket Permissions
- Amazon EBS Public Snapshots
- Amazon RDS Public Snapshots



© 2018, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

The best practices that are available to all customers include:

- Service limits
- Security group rules that allow unrestricted access to specific ports
- IAM use
- Multi-factor authentication—or MFA—on the root account
- Amazon S3 bucket permissions
- Amazon EBS public snapshots; and
- Amazon Relational Database Service—or Amazon RDS—public snapshots

Select the link to learn more about AWS Trusted Advisor best practices. https://aws.amazon.com/premiumsupport/trustedadvisor/best-practices/

AWS Trusted Advisor Features



AWS Trusted Advisor provides a suite of features for you to customize recommendations and to proactively monitor your AWS resources.

Notifications



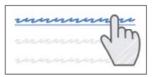
Access Management



AWS Support API



Action Links



Recent Changes



Exclude Items



5-Min Refresh



© 2018, Amazon Web Services, Inc. or its Affiliates. All rights reserved.



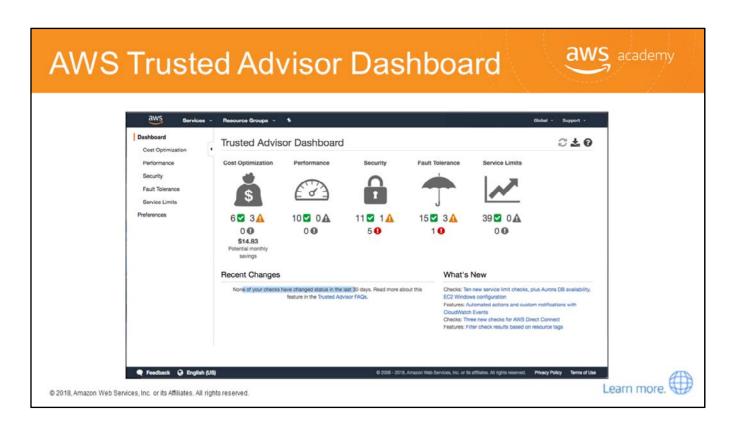
Let's review some features and functionality of Trusted Advisor.

Within the console, you have:

- AWS Trusted Advisor Notifications, which helps you stay up-to-date with your AWS resource deployment. You will be notified by weekly email when you opt in for this service, and it is free.
- You can use AWS Identity and Access Management, or IAM, to control access to specific checks or check categories.
- You can retrieve and refresh Trusted Advisor results programmatically by using the AWS Support API.
- Action Links are hyperlinks on items within a Trusted Advisor report. Using Action Links takes you directly to the console, where you can take action on the Trusted Advisor recommendations.
- With the Recent Changes feature, you can track recent changes of a check status on the console dashboard. The most recent changes appear at the top of the list to bring them to your attention.
- The Exclude Items feature allows you to customize the Trusted Advisor report. You can exclude items from the check result if they are not relevant.

You can refresh individual checks or refresh all the checks at once by choosing the Refresh All button in the summary dashboard. A check is eligible for refresh 5 minutes after it was last refreshed.

Select the link to learn more about AWS Trusted Advisor. https://aws.amazon.com/premiumsupport/trustedadvisor/



Displayed an example of the Trusted Advisor Dashboard. It shows a summary of the four best practice categories, as well as recent changes. The cost optimization check shows recommendations for saving money. Trusted Advisor also offers checks for performance, security, and fault tolerance.

Select the link to access the Trusted Advisor Dashboard. https://console.aws.amazon.com/trustedadvisor/

Case Study: Hungama





Using AWS Trusted Advisor helped us save 33% on our monthly bill, and we'll continue to use it to optimize our infrastructure and costs on AWS.

> Amit Vora CTO, Hungama Digital Media



Hungama is a leading aggregator, developer, publisher and distributor of Bollywood and South-Asian entertainment content.

© 2018, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

- Hungama has used AWS for server and storage management since 2008.
- They deliver content to consumers in 47 countries across mobile, Internet, and Internet protocol television (IPTV) services.
- The company uses Amazon S3 to host more than 60 TB of content and Amazon EC2 and Amazon RDS for server and storage management.
- As the company grew rapidly, more departments used AWS for development, causing an increase in monthly costs.

Let's look at a case study of how Hungama uses Trusted Advisor to optimize usage and cut costs to save 33 percent on their monthly bill.

Hungama:

- Has used AWS for server and storage management since 2008.
- Delivers content to consumers in 47 countries across mobile, internet, and internet protocol television—or IPTV–services.
- Uses Amazon S3 to host more than 60 TB of content, and uses Amazon EC2 and Amazon RDS for server and storage management.
- As the company grew rapidly, more departments used AWS services for development, which caused an increase in monthly costs.

Hungama Uses Trusted Advisor



Hungama reduced monthly costs by 33 percent by using AWS Trusted Advisor.

How did they do that?

By using AWS Trusted Advisor's Cost Optimizing checks. Here are a few things they did:



The Low Utilization Amazon EC2 Instances check revealed over-provisioned instance sizes, and instances spun up for special projects were not terminated after completion.

In response, the audit team used this information to right-size their instances. They also categorized production and development servers, and automated the process of shutting down development servers during non-business hours.

© 2018, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

How did Hungama use Trusted Advisor to reduce monthly costs by 33 percent? They did it by using Trusted Advisor's cost optimizing checks.

The Low Utilization Amazon EC2 Instances check on Trusted Advisor checks the Amazon EC2 instances that run at any time during the last 14 days. It alerts you if the daily central processing unit—or CPU—usage was 10 percent or less, and if network I/O was 5 MB or less on 4 or more days.

This check revealed that Hungama had overprovisioned instance sizes, and instances that were spun up for special projects were not terminated after completion. The team used this information to right-size their instances. They also categorized production and development servers, and they automated the process of shutting down development servers during non-business hours.

The estimated monthly savings are calculated by using the current usage rate for On-

Demand Instances and the estimated number of days that the instance might be underused. Actual savings will vary if you use Reserved Instances or Spot Instances, or if the instance does not run for a full day. To get daily usage data, you can download the report for this check.

Hungama's Reserved Instances



Hungama reduced monthly costs by 33% by using AWS Trusted Advisor.

How did they do that?

By using AWS Trusted Advisor's Cost Optimizing checks. Here are a few things they did:



The Reserved Instance Optimization check identified additional opportunities for optimization of the RI instances they had purchased.

In response, Hungama changed how they reserved their instances and based reservations on the specific usage patterns of their different instance categories (dev/prod/test/etc).

© 2018, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

The Reserved Instance Optimization check from Trusted Advisor checks your Amazon EC2 computing consumption history, and it calculates an optimal number of Partial Upfront Reserved Instances. Recommendations are based on the hour-by-hour usage of the previous calendar month, and is aggregated across all consolidated billing accounts.

Hungama's Reserved Instance changes occurred before the current pricing model of No Upfront, Partial Upfront, and All Upfront was implemented. However, this check is still extremely useful to customers who use Reserved Instances. In response to the findings of Trusted Advisor, Hungama changed how they reserved their instances. They based their reservations on the specific usage patterns of their different instance categories, such as development, production, test, and so on.

Hungama Snapshots



Hungama reduced monthly costs by 33% by using AWS Trusted Advisor.

How did they do that?

By using AWS Trusted Advisor's Cost Optimizing checks. Here are a few things they did:



The Underutilized Amazon EBS volumes check identified a number of unused or underutilized EBS volumes that were often leftover from previous test projects.

In response, the audit team created snapshots of many of the underutilized EBS volumes, which they stored on Amazon S3, and then deleted the volumes. This resulted in a reduction of over 90% on the number of snapshots generated weekly.

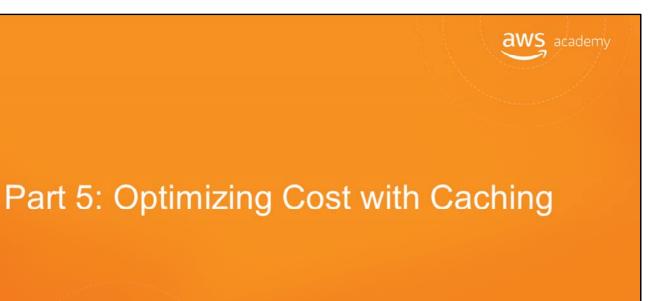
© 2018, Amazon Web Services, Inc. or its Affiliates. All rights reserved.



The Underutilized Amazon EBS volumes check on Trusted Advisor checks Amazon EBS volume configurations, and issues a warning when volumes appear to be underused. If a volume remains unattached or has very low write activity for a period of time—excluding boot volumes—the volume is probably not being used.

In response to this check, the Hungama audit team created snapshots of many of the underused Amazon EBS volumes, which they stored on Amazon S3. They then deleted the volumes. This action resulted in a reduction of over 90 percent on the number of snapshots that were generated weekly.

To learn more about how Hungama uses AWS, select the link. https://aws.amazon.com/solutions/case-studies/hungama/



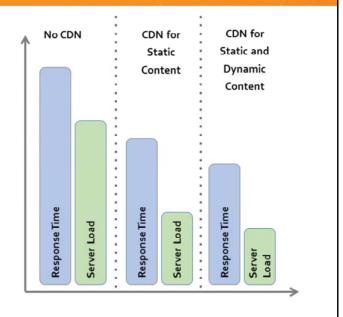
In Part 5, we will look at optimizing cost with caching.

Serving Content with Caching



The more you can offload, the less infrastructure you need to maintain, scale, and pay for.

- Offload popular traffic to Amazon CloudFront and Amazon S3.
- Introduce caching.



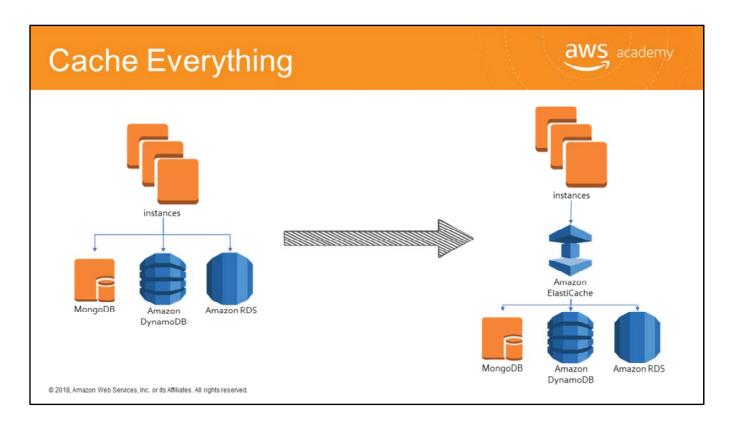
© 2018, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

The term content delivery network—or CDN—refers to Amazon CloudFront. CloudFront is a content delivery web service. It gives developers and businesses a way to distribute content to end users with low latency, high data transfer speeds, and no minimum usage commitment. Depending on your application types, it might make better sense to distribute your static or dynamic content through CloudFront, instead of scaling out your infrastructure.

Let's look at the response time versus server load if you don't use a content delivery network. Without a CDN, the response time goes up, and the load on the origin is much higher.

When only static content is offloaded by using a content delivery network like Amazon CloudFront, the response time and server load is reduced by half.

If both static and dynamic content are delivered through Amazon CloudFront, you have larger savings in terms of server response time and an even larger amount of savings on the server load. This scenario means that you can have smaller and fewer instances at the origin if that work is offloaded to Amazon CloudFront.



Databases can be cached to Amazon ElastiCache. Database types include NoSQL instances like MongoDB or Amazon DynamoDB, and also include relational database servers.



Part 6: AWS Cost Calculation Tools

In Part 6, we will cover a couple of cost calculation tools that you can use from the console.

AWS Total Cost of Ownership (TCO) aws academy

- Estimate cost savings when using AWS
- Use a detailed set of reports that can be used in executive presentations
- Modify assumptions that best meet your business needs



 Describe your infrastructure in four steps, or enter detailed configurations

© 2018, Amazon Web Services, Inc. or its Affiliates. All rights reserved.



2. Get an instant summary report



3. Download a full report including detailed cost breakdowns



We can help you reduce Total Cost of Ownership—or TCO—by reducing the need to invest in large capital expenditures. We provide a pay-as-you-go model that empowers you to invest in the capacity you need, and use it only when the business requires it.

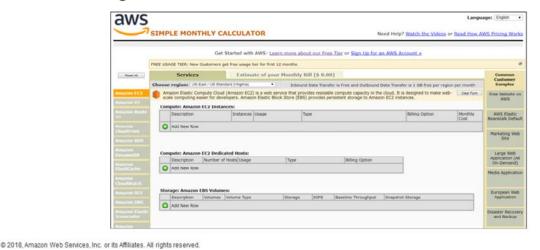
The AWS TCO calculators allow you to estimate the cost savings when using AWS and provide a detailed set of reports that can be used in executive presentations. The calculators also give you the option to modify assumptions that best meet your business needs.

Select the link to launch the AWS Total Cost of Ownership Calculator. https://awstcocalculator.com/

AWS Simple Monthly Calculator



Estimate the cost of running your application or solution in the AWS Cloud based on usage.



Learn more.

The AWS Simple Monthly Calculator is a tool that can help you estimate the cost of running your application or solution in the AWS Cloud based on usage. This calculator will help you determine how adding a new server or an additional service will impact your overall bill.

Select the link to access the AWS Simple Monthly Calculator. https://calculator.s3.amazonaws.com/index.html



Exercise: Improve This Architecture

In the initial "Improve This Architecture" exercise, your team implemented all options to improve security, reliability, and performance without considering the cost implications.

In this exercise, you will consider making improvements that include a justification or analysis of the cost trade-offs.

Cost Optimization



- Your team performed a security, reliability, and performance assessment and put in every option they thought would increase performance. Your Cost Explorer indicates that your forecast is higher than your budget.
- Review the diagram and ask questions to identify where improvements can be made to improve the Cost Optimization of the system.
- Recommend ways to redesign the solution in accordance with the Well-Architected Cost Optimization pillar.

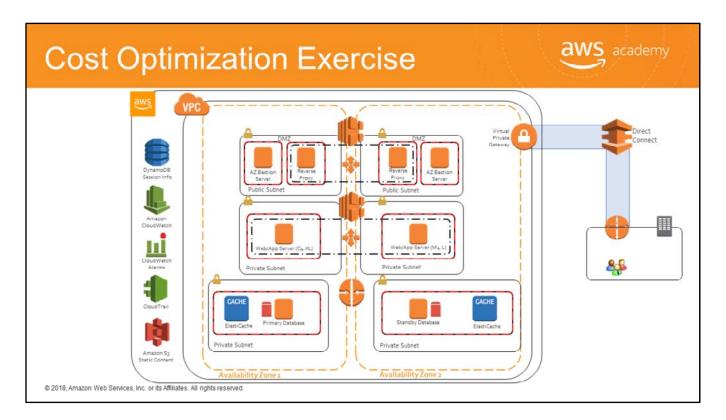
© 2018, Amazon Web Services, Inc. or its Affiliates. All rights reserved

In the prior architecture exercise, you:

- Performed a security, reliability, and performance assessment.
- You put in every option that you thought would increase performance.

For this exercise:

- Your Cost Explorer indicates that your forecast is higher than your budget.
- You will review the diagram, and ask questions to identify where you can improve the cost optimization of the system.
- You will also recommend ways to redesign the solution in accordance with the Cost-Optimization pillar of the Well-Architected Framework.



Consider the following issues:

- Are there unbalanced instance sizes?
- Why should you use a C4 Amazon EC2 instance for a web server?
- Why run large web or application servers behind small reverse proxies?
- Why have do-it-yourself databases, over Amazon RDS?
- Why use Amazon ElastiCache? What purpose does it serve?
- Should you use Memcached or Redis?
- How many nodes are in the ElastiCache Cluster?
- What is the bandwidth of the AWS Direct Connect connection?
- What is the throughput of the Amazon DynamoDB instance?
- Does Amazon CloudWatch use detailed monitoring or default interval?
- This architecture does not use Amazon Glacier. Could the customer realize cost savings by moving some data off Amazon S3 and on to Amazon Glacier?
- What is the cost of using Amazon CloudTrail versus the cost of no auditing?
- Why does this architecture use two load balancers? Could Amazon SQS be used?
- Why does this architecture use two bastion hosts?
- Could any of these workloads be done with AWS Lambda?

Additional Resources



- Video
 - Cost Optimization on AWS
- Documentation
 - AWS Cloud Economics Center
- Tools
 - MWS Total Cost of Ownership (TCO) Calculator
 - AWS Detailed Billing Reports
 - AWS Simple Monthly Calculator
 - AWS Cost Explorer

© 2018, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Here are some additional resources for optimizing costs for AWS products and services.

Select each link to learn more.

Review



- Revealed the principles of the cost optimization pillar.
- Discovered how to optimize the costs of your infrastructure.
- Understand best practices to eliminate unneeded costs or suboptimal resources.

To finish this module:

Complete: Knowledge Assessment

© 2018, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

In review, you have:

- Revealed the principles of the cost optimization pillar.
- Discovered how to optimize the costs of your infrastructure.
- And understand best practices to eliminate unneeded costs or suboptimal resources.

To finish this module, please complete the corresponding knowledge assessment.