Designing Highly Available, Cost-efficient, Fault-tolerant Scalable Systems

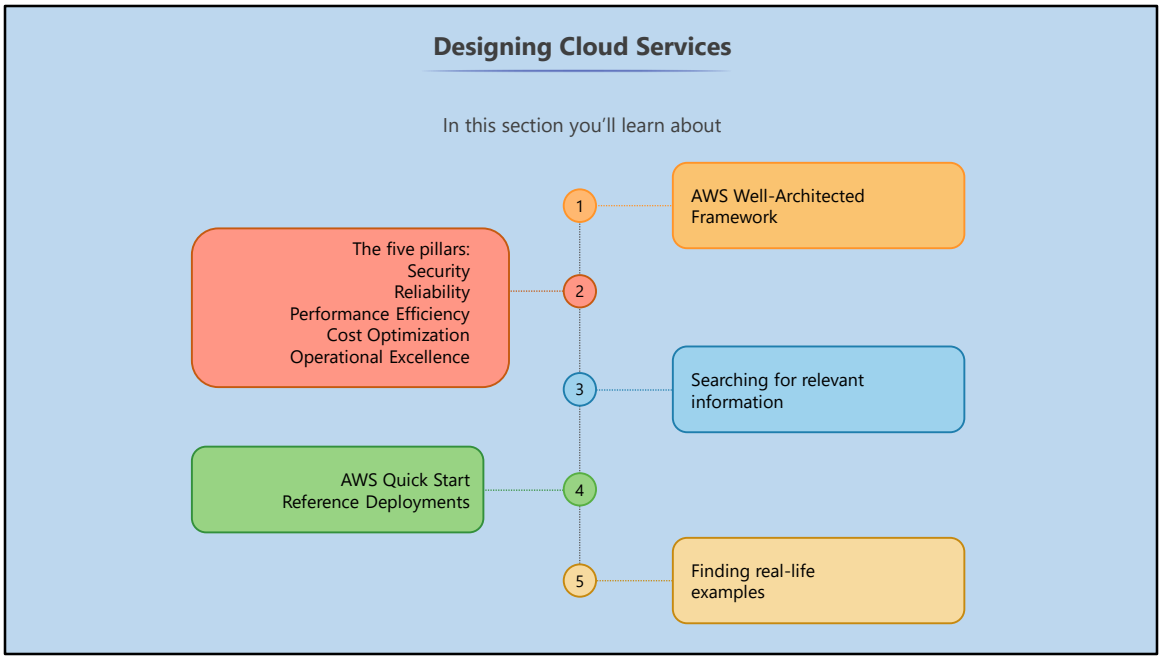
Welcome to the second lesson of the AWS Solutions Architect Associate level course—"How to Design a Cloud Service."

Learning Objectives

- Discuss Cloud versus Traditional Infrastructure
- Define AWS Well-Architected Framework
- Describe Planning and Designing Cloud Infrastructure
- Illustrate AWS Monitoring and Logging Tools
- Explain Hybrid Cloud Infrastructure

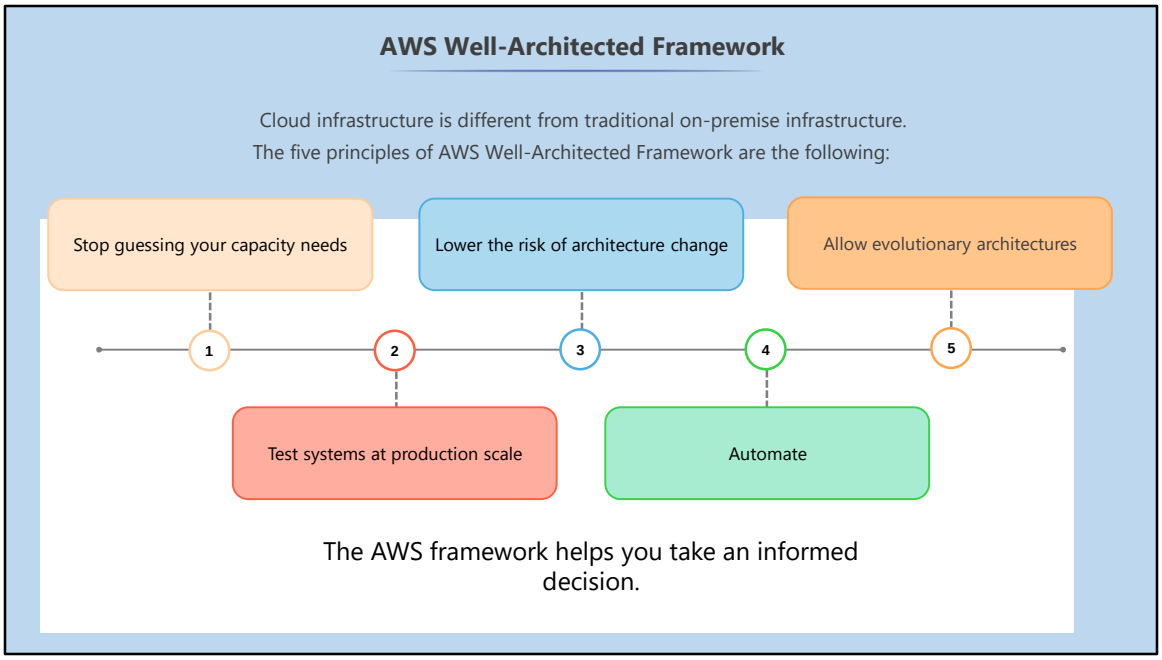
How to Design Cloud Services

This section is an overview of the cloud design principles.



In this section you'll learn about

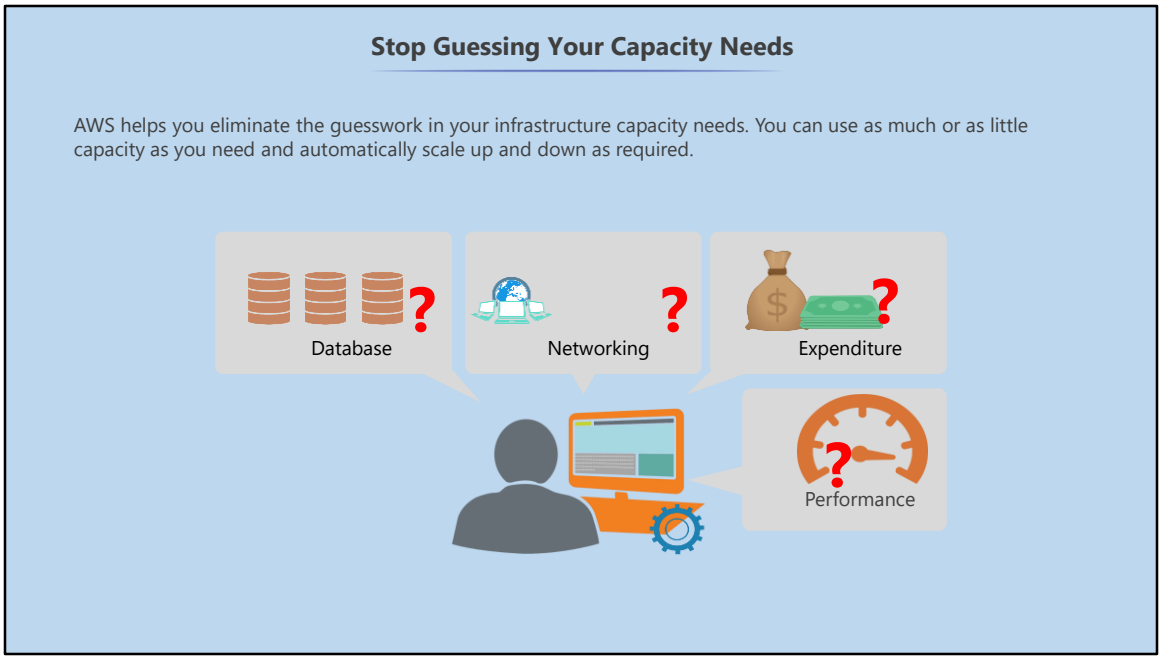
- AWS Well-Architected Framework and why it exists
- Details of the four pillars of the AWS Well-Architected Framework
- Where to find information on the AWS infrastructure
- What are AWS Quick Start Reference Deployments and their uses
- Examples of companies and their AWS usage



AWS Well-Architected Framework helps you understand the pros and cons of the decisions you take when building systems on AWS.

The five principles of AWS Well-Architected Framework are:

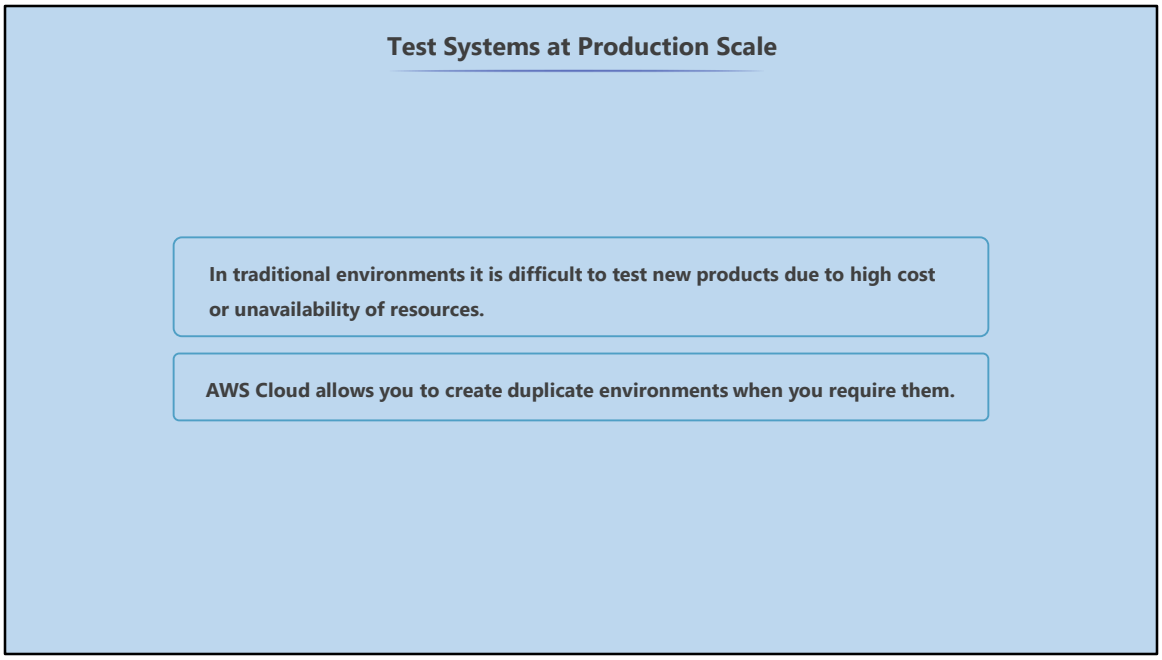
1. Stop guessing your capacity needs
2. Test systems at production scale
3. Lower the risk of architecture change
4. Automate to make architectural experimentation easier
5. Allow for evolutionary architectures



Imagine you are designing a new application for your business.

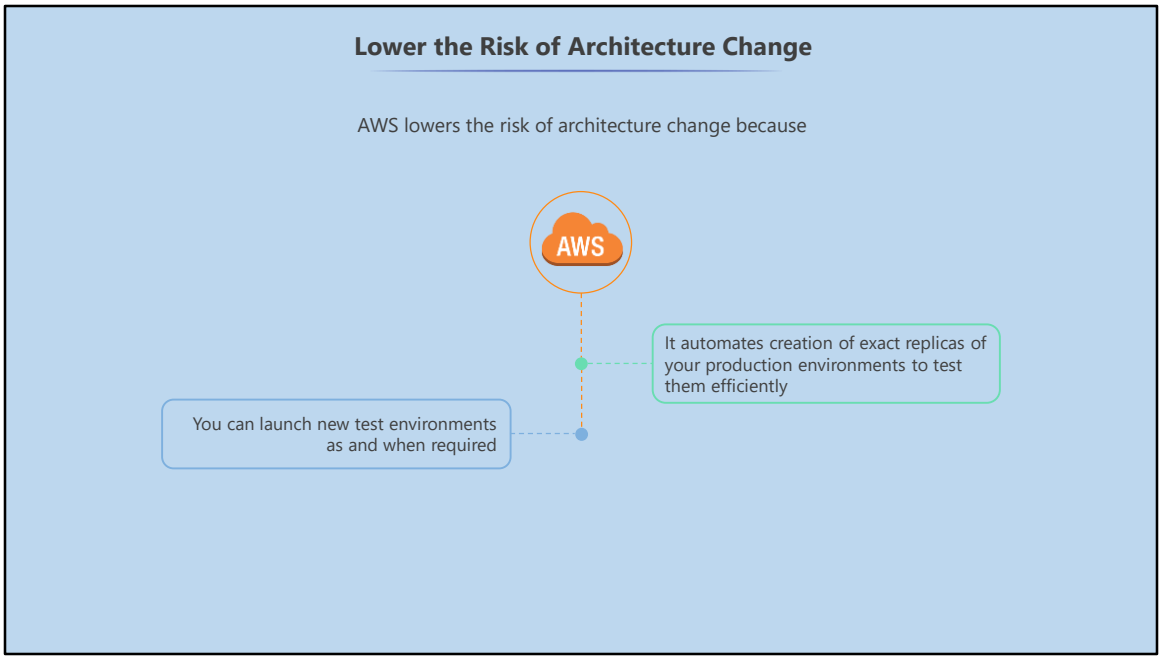
All the infrastructure decisions are made using estimates or guesses. You might overestimate and end up buying too much hardware resulting in expensive, idle resources. Alternatively, you might underestimate and have to deal with the performance implications of limited capacity.

Cloud computing solves these problems. AWS helps you eliminate the guesswork in your infrastructure capacity needs. You can use as much or as little capacity as you need and automatically scale up and down as required.



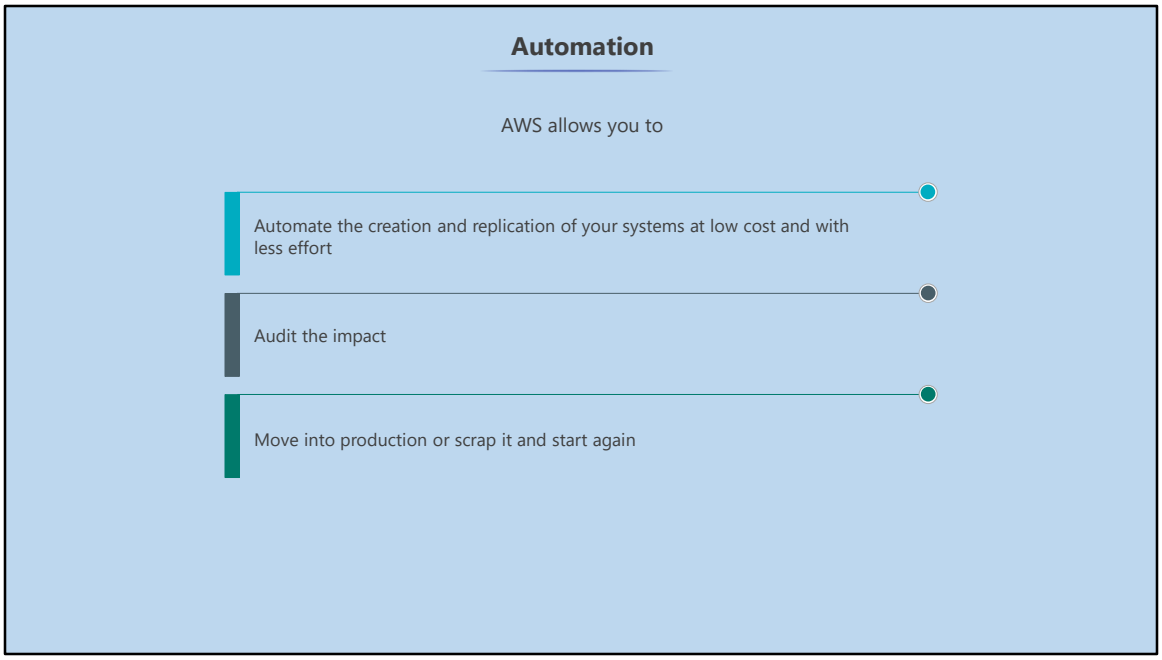
Traditionally, in a non-cloud environment it is very difficult to completely test new products or services in development as it is cost prohibitive or resources are unavailable. Projects go wrong as soon as the product goes live due to insufficient testing.

With the cloud, you can create duplicate environments on demand to test the new product. After completing the testing, you can shut them down and pay only for the time the test environment was up and running.

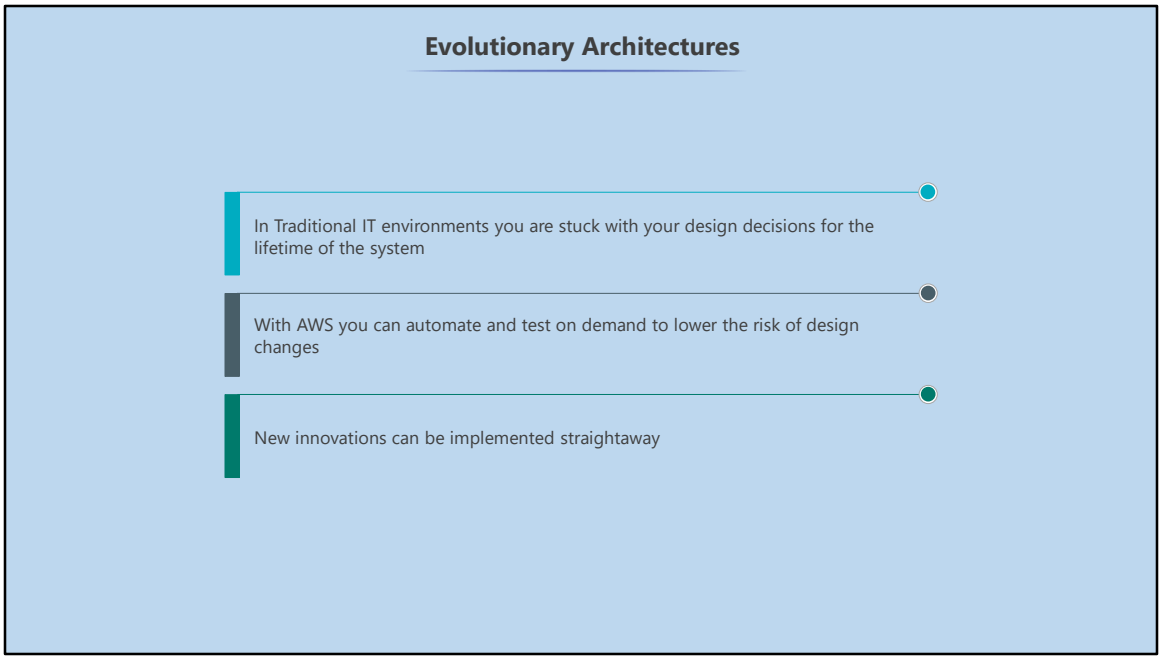


With AWS you can automate the creation of test environments that emulate your production configurations, which means you can carry out testing easily against comparable environments.

You can remove testing bottlenecks when various teams line up to use your test environments. With AWS you can create as many test environments as you need in a cost efficient way and ensure that all production changes have been sufficiently tested.

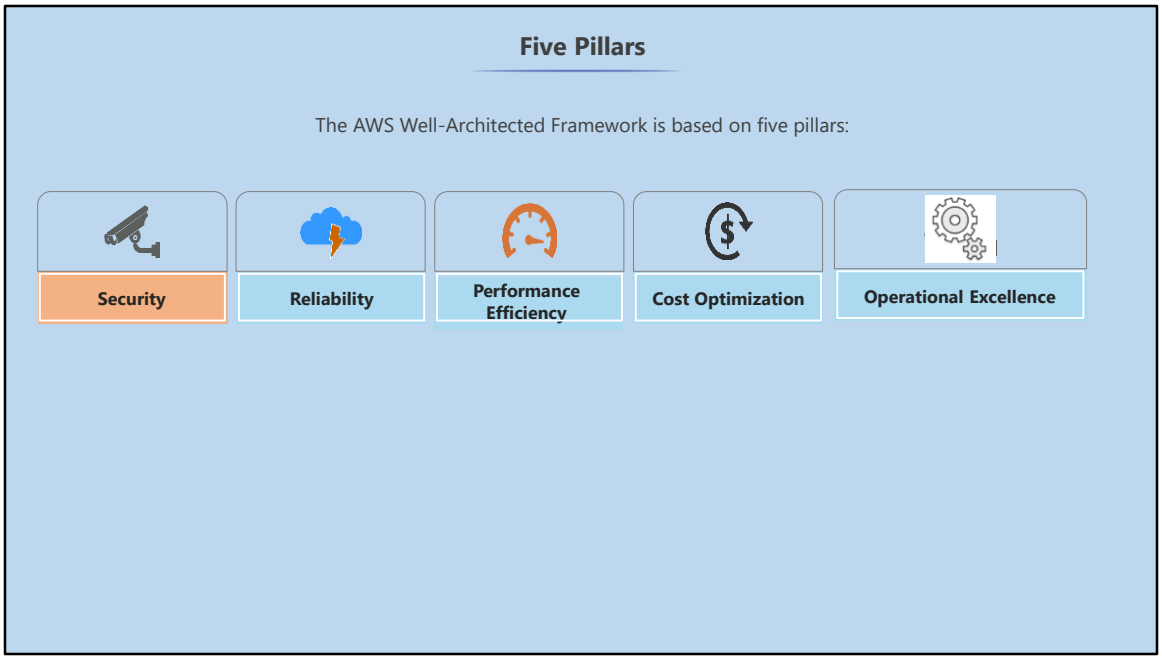


You can automate the creation and replication of systems at low cost and with minimal effort to test changes, audit the impact, and revert if necessary.



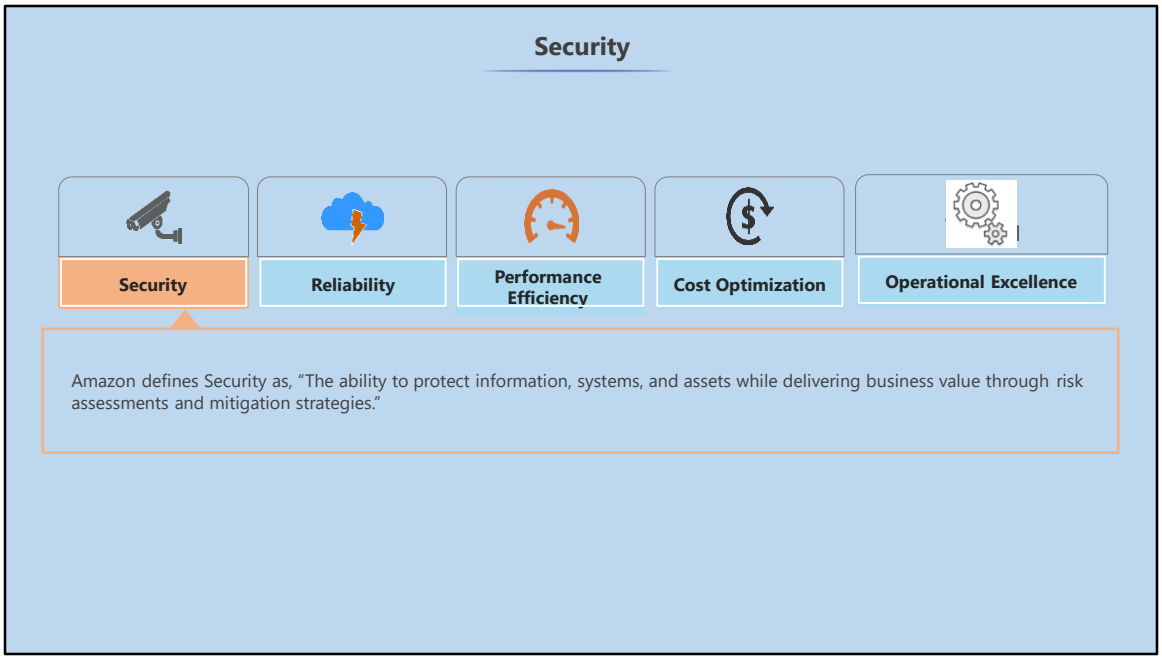
In traditional environments, architectural decisions are usually implemented as a static, one-time event during the lifetime of the system.

The cloud provides the capability to automate and test on demand. This lowers the risk of design changes and makes implementation of innovations and features easy, which means you aren't stuck with last year's technology.

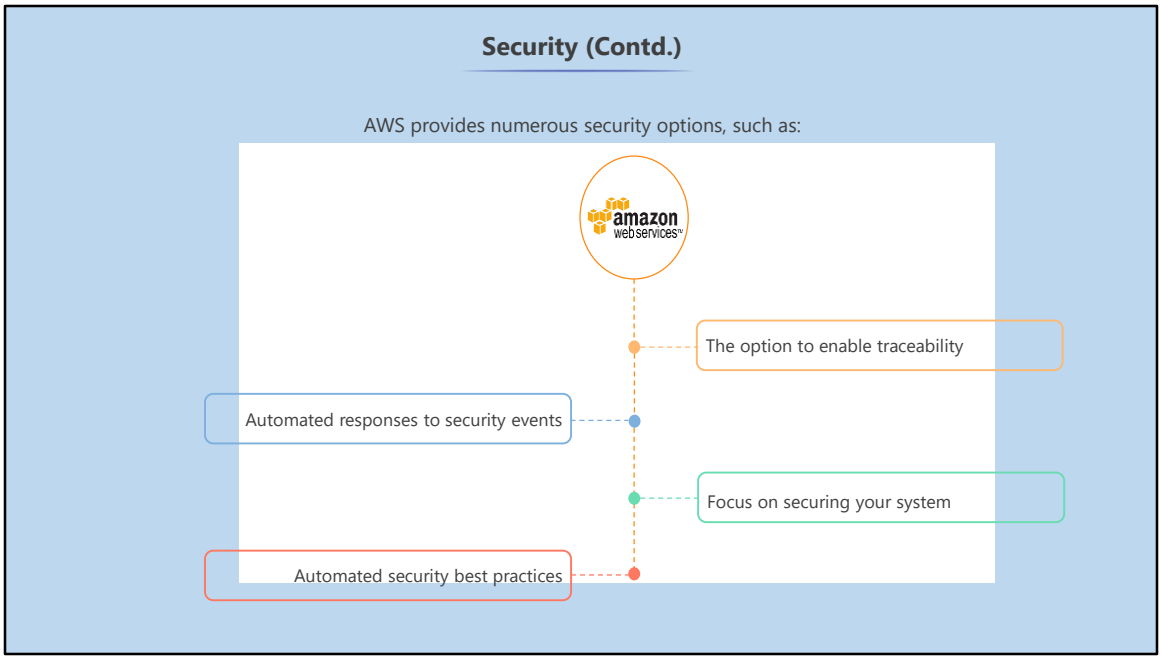


The “Well-Architected Framework” is based on four pillars:

1. Security
2. Reliability
3. Performance efficiency
4. Cost optimization

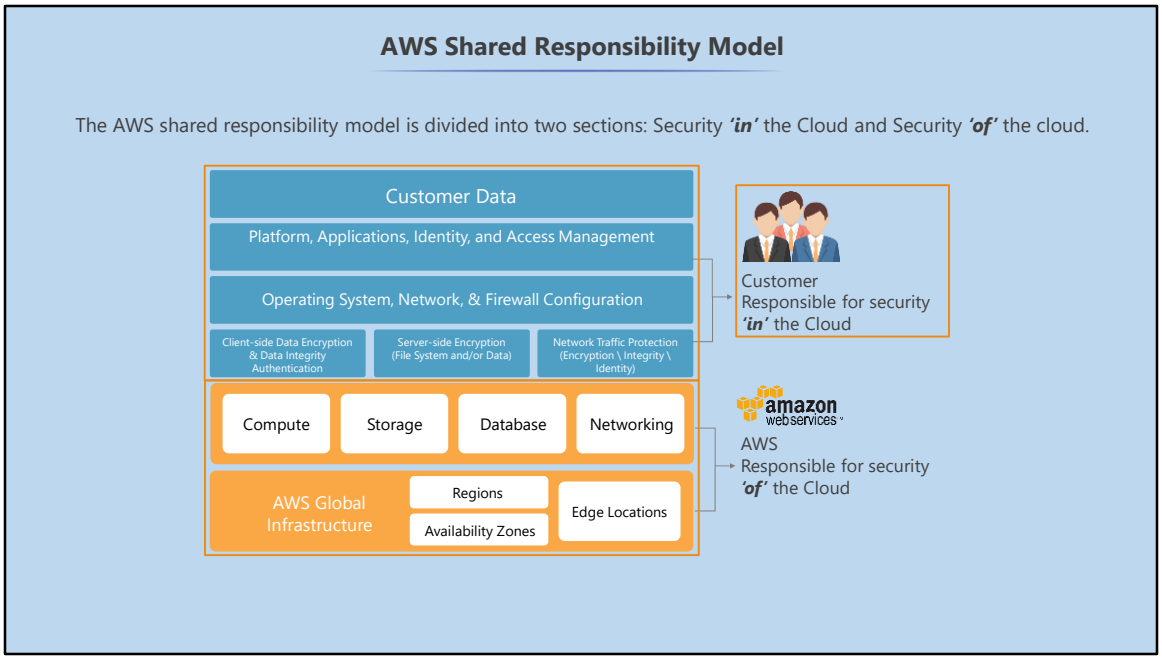


Security is the ability to protect information, systems, and assets, while delivering business value through risk assessments and mitigation strategies.



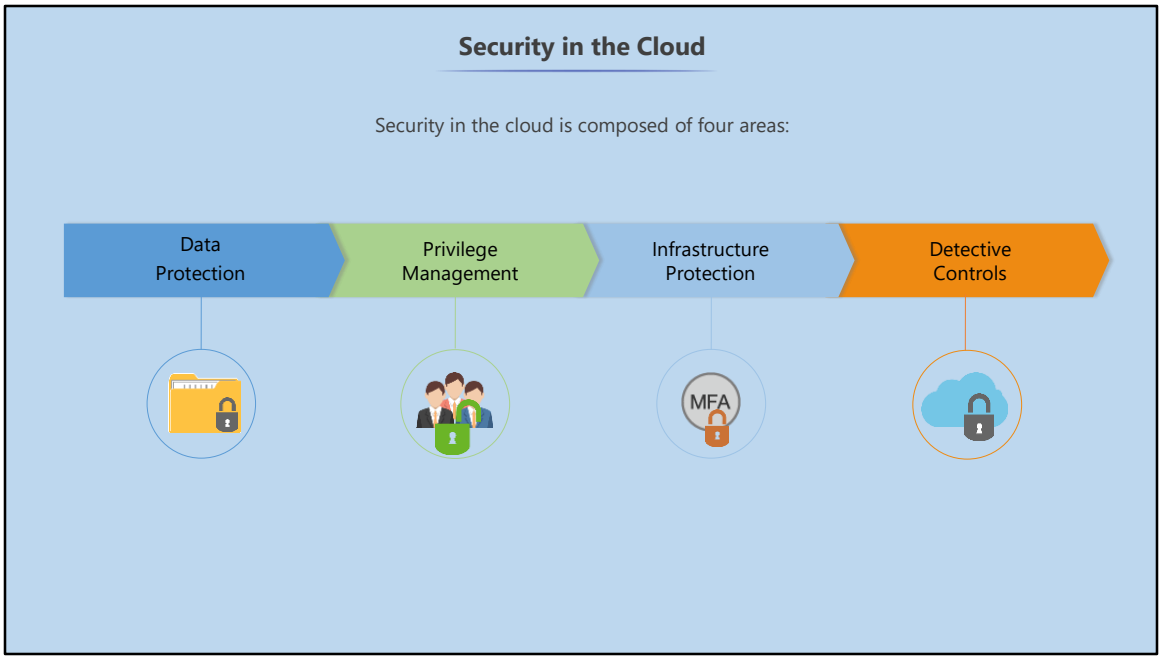
Apply security at all layers: Security should be thorough. For example, instead of firewalls at the edge of the infrastructure, use firewalls and other security controls on all of the resources. Another example is the security groups on EC2 instances.

- Enable traceability: Log and audit all changes to the AWS environment.
- Automate responses to security events: Monitor and automatically trigger alerts.
- Focus on securing your system: The AWS Shared Responsibility Model allows you to secure your applications, data, and operating systems, while AWS secures your infrastructure and services.
- Automate security best practices: AWS allows you to create and save a custom baseline image of a virtual server, and use that image automatically on each new server you launch. This ensures all new resources adhere to your security standards.



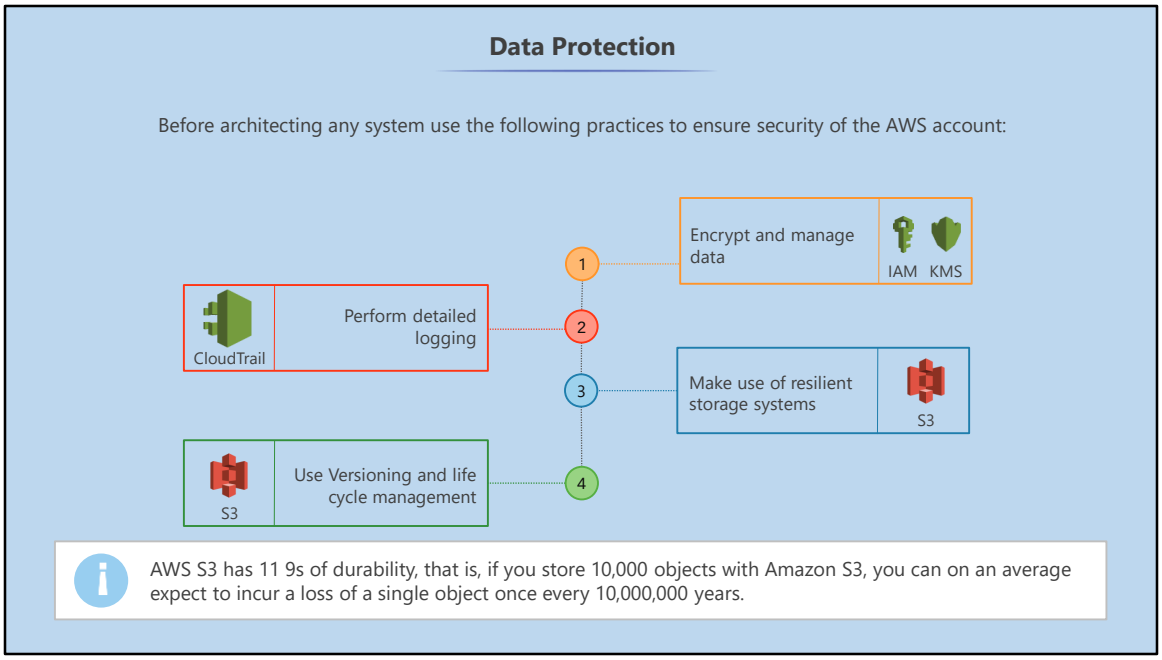
The AWS shared responsibility model is divided into "Security ***'in'*** the Cloud" and "Security '***of'*** the cloud." The cloud service provider or AWS implements and operates security that relates to the services it provides. This is referred to as the "Security '***of'*** the cloud."

The customer implements and operates, the security of customer content and applications that make use of AWS services –"Security ***'in'*** the cloud."



Security in the cloud is composed of four areas:

1. Data protection
2. Privilege management
3. Infrastructure protection
4. Detective controls

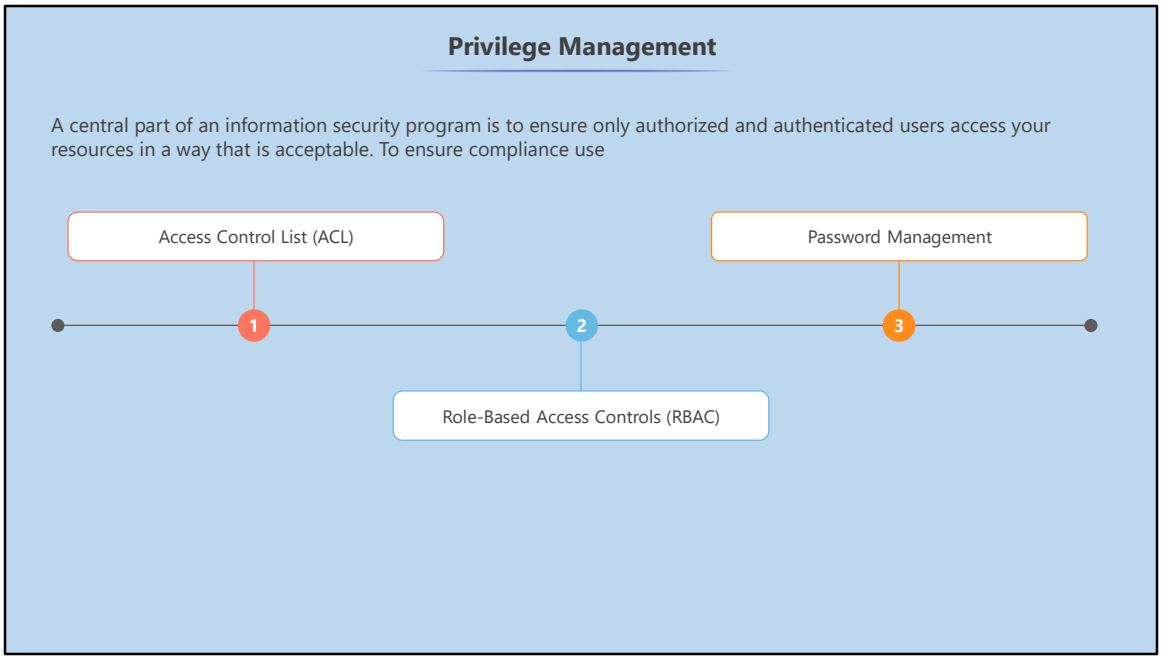


Before architecting any system use the following practices to ensure security of the AWS account:

- Categorize data based on levels of sensitivity
- Grant least privilege while allowing users to perform their work
- Encrypt sensitive data

AWS provides a number of tools to ensure security such as:

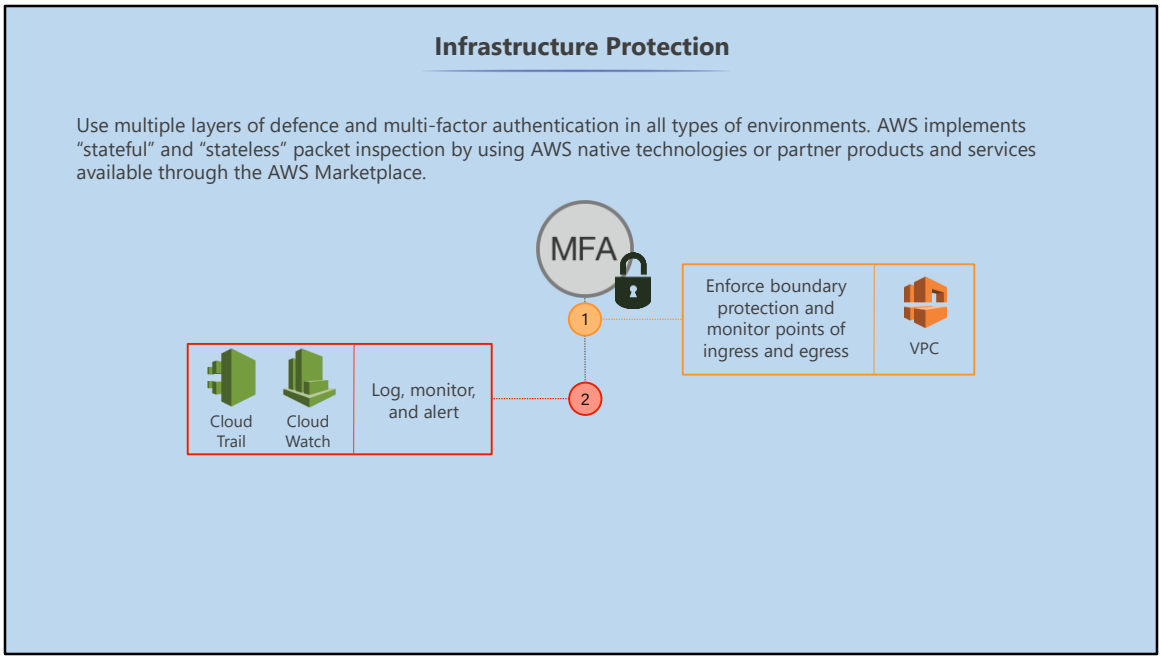
- IAM and KMS allow easy data encryption and access key management.
- CloudTrail allows you to perform detailed Logging.
- Amazon S3 is a resilient storage systems that offers 11 9s of durability. This means if you store 10,000 objects with Amazon S3, you can on an average expect to incur a loss of a single object once every 10,000,000 years.
- Versioning and Life Cycle Management protect against accidental deletes or overwrites.



A central part of an information security program is to ensure that only authorized and authenticated users can access your resources in a way that is acceptable. For example, if users need access to a resource, don't just give them admin or read write permissions when they only need read-only access.

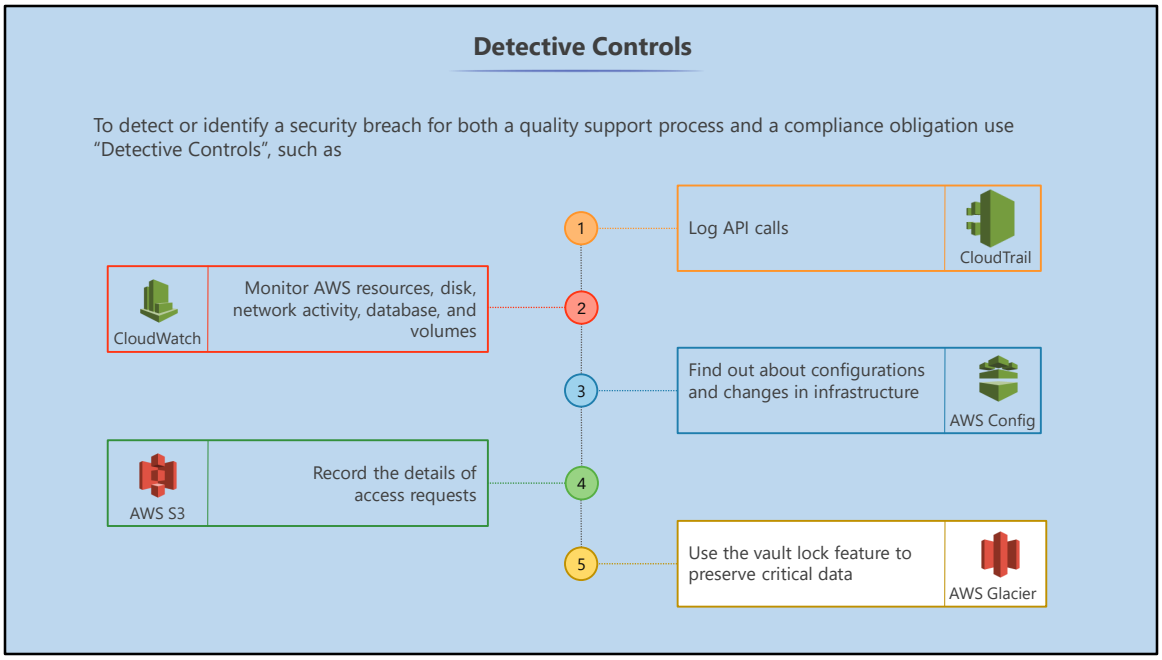
Access Control List (ACL) should be maintained to list access permissions attached to an object.
Role-Based Access Controls (RBAC) define the permissions for a particular end user's role or function. Password management includes complexity requirements and change intervals.

The AWS Identity and Access Management (IAM) service is the primary service used to control access to AWS services and resources for users. Apply granular policies, which assign permissions to a user, group, role, or resource. Ensure strong password practices and use federation with the existing directory service.



To meet best practices and industry or regulatory obligations use multiple layers of defense and multi-factor authentication in all types of environments.
In AWS, implement "stateful" and "stateless" packet inspection by using AWS native technologies or partner products and services available through the AWS Marketplace.

The Amazon Virtual Private Cloud (VPC) enforces boundary protection and monitors points of ingress and egress. Use CloudTrail and CloudWatch for comprehensive logging, monitoring, and alerting.

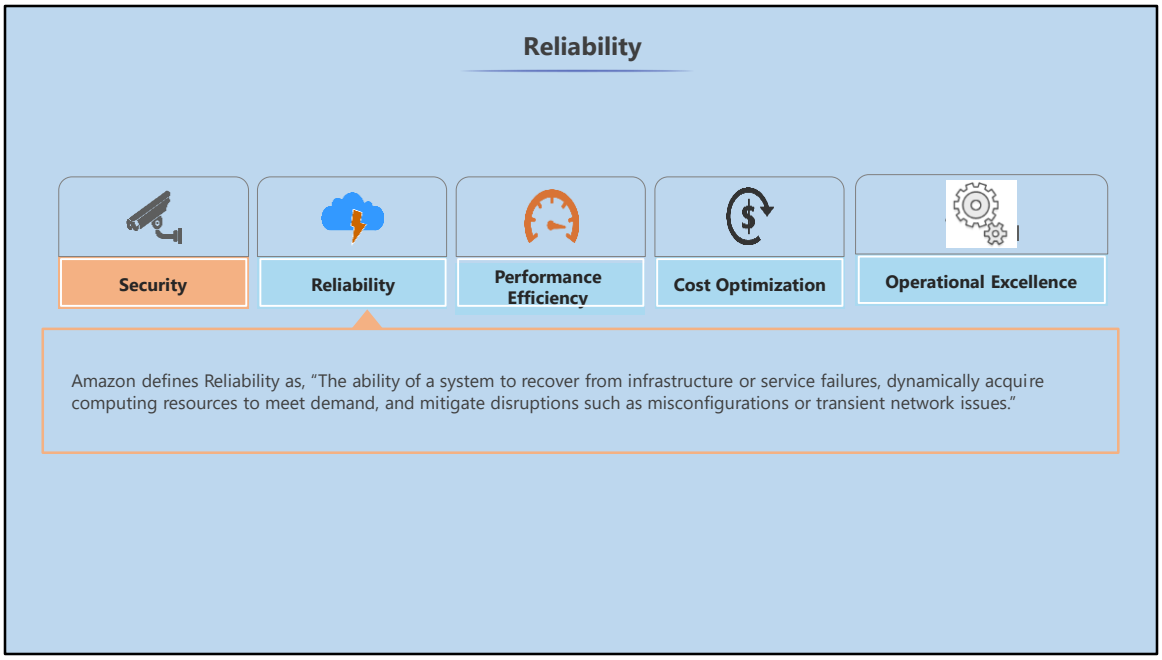


Use "Detective Controls" to detect or identify a security breach for both a quality support process and a compliance obligation.

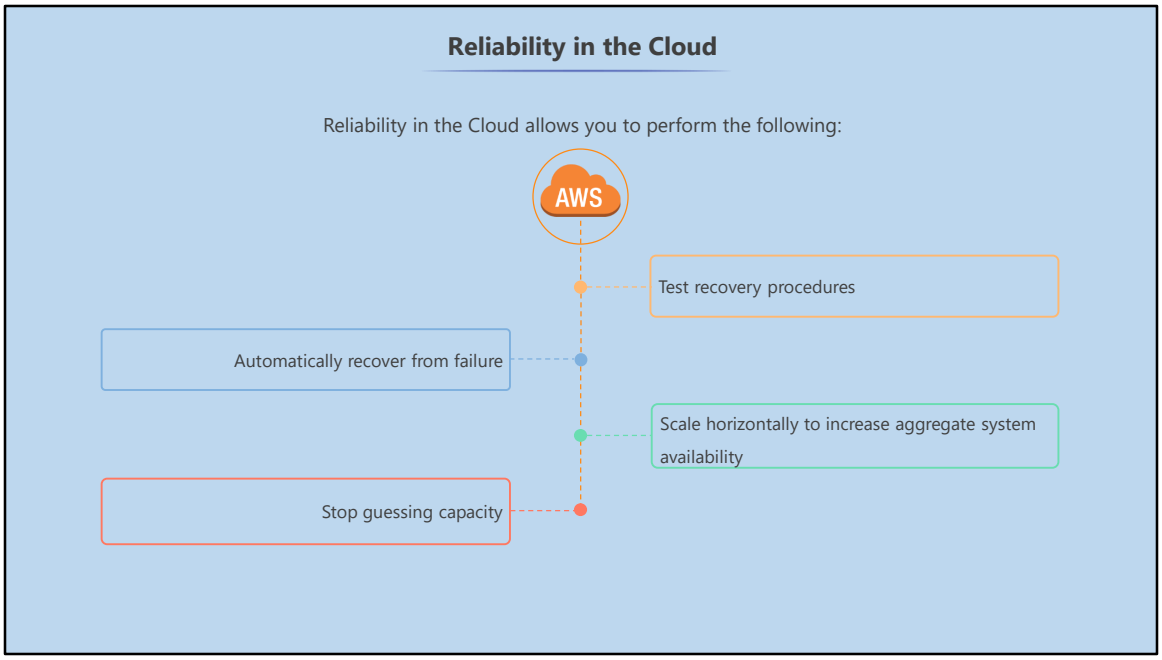
A few of the "Detective Controls" tools are:

- AWS CloudTrail–A web service that logs API calls, including the identity of the call, the time of the call, source IP address, parameters, and response elements.
- Amazon CloudWatch–A monitoring service for AWS resources that logs aspects such as Amazon Elastic Compute Cloud (EC2), CPU disk, and network activity, Amazon Relational Database Service (RDS), database instances, and Amazon Elastic Block Store (EBS) volumes. It offers the ability to provide alarms on these and other metrics.
- AWS Config–An inventory and configuration history service that provides information about the configurations and changes in infrastructure over time.
- Amazon Simple Storage Service (S3)–Amazon S3 data access auditing allows configuring Amazon S3 buckets to record the details of access requests, including the type, resource, date, and time.
- Amazon Glacier–Uses the vault lock feature to preserve critical data with

compliance controls designed to support auditable long-term retention.

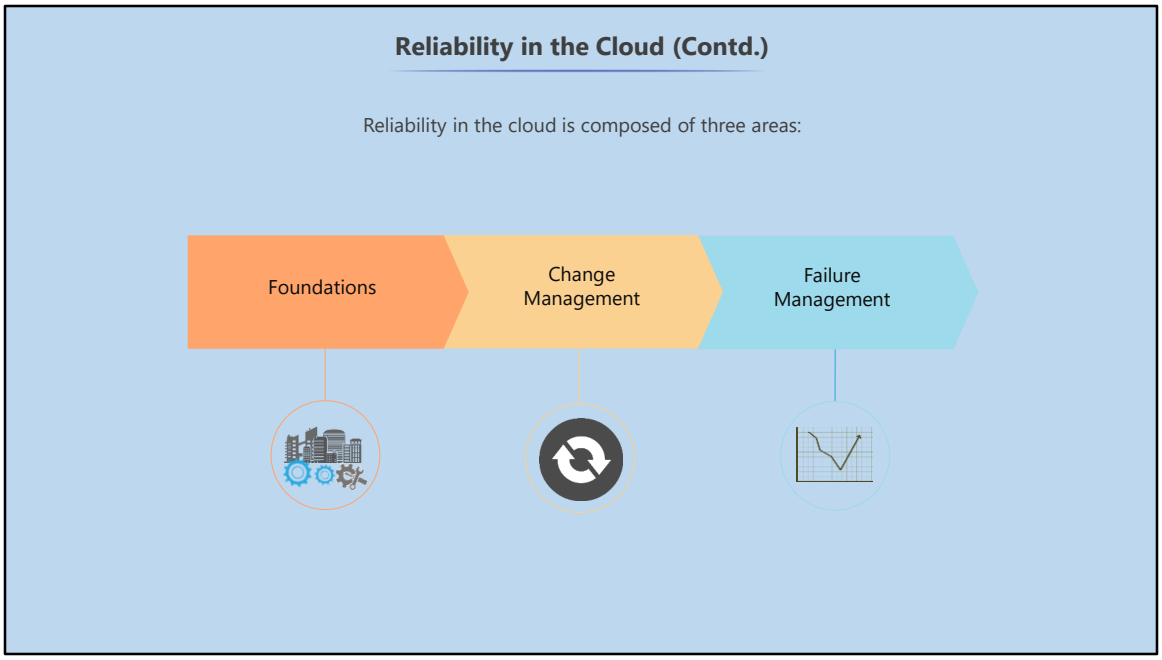


Reliability is the ability of a system to recover from infrastructure or service failures, dynamically acquire computing resources to meet demand, and mitigate disruptions such as misconfigurations or transient network issues.



Reliability in the Cloud allows you to

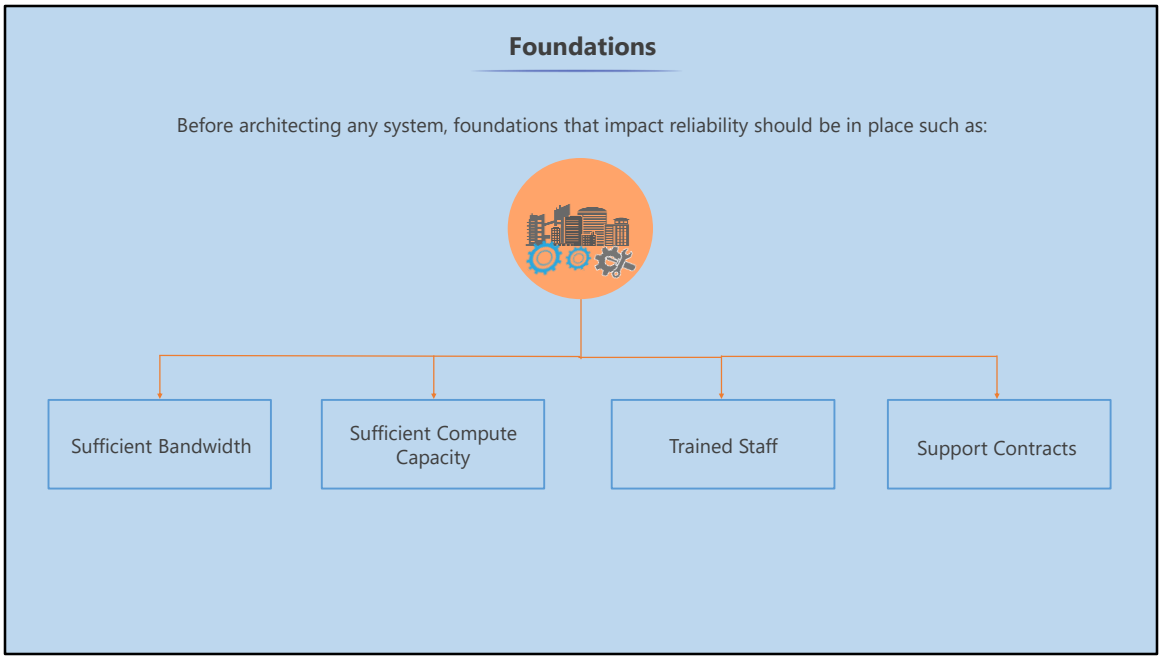
Test recovery procedures: These procedures help test how the system fails and allows you to validate the recovery strategy.
Automatically recover from failure: Monitor a system for key performance indicators (KPIs) and automatically trigger an automated recovery process when a threshold is breached.
Scale horizontally to increase aggregate system availability: Use multiple small resources instead of one large resource to reduce single points of failure.
Stop guessing capacity: “Lack of memory or CPU” as a failure for a resource is a thing of the past as the cloud allows you to scale as and when required.



Reliability in the cloud is composed of three areas:

1. Foundations
2. Change management
3. Failure management

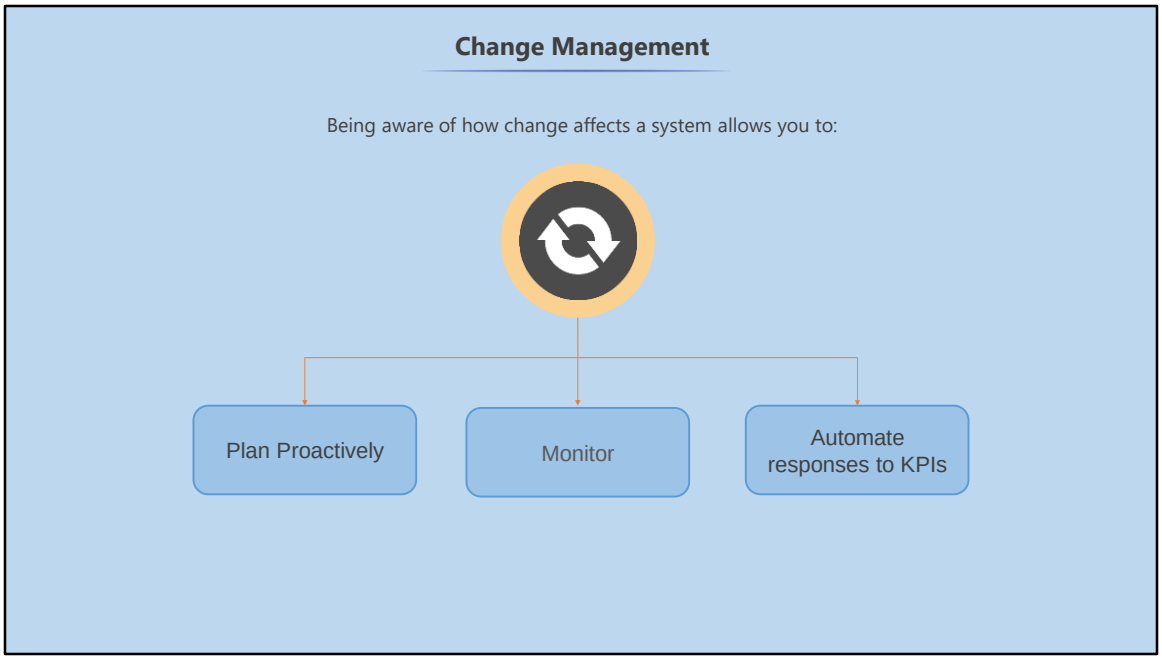
To achieve reliability, a system must have a well-planned foundation and monitoring in place, plus systems for handling changes in demand or requirements.



Before architecting any system, foundations that impact reliability should be in place such as:

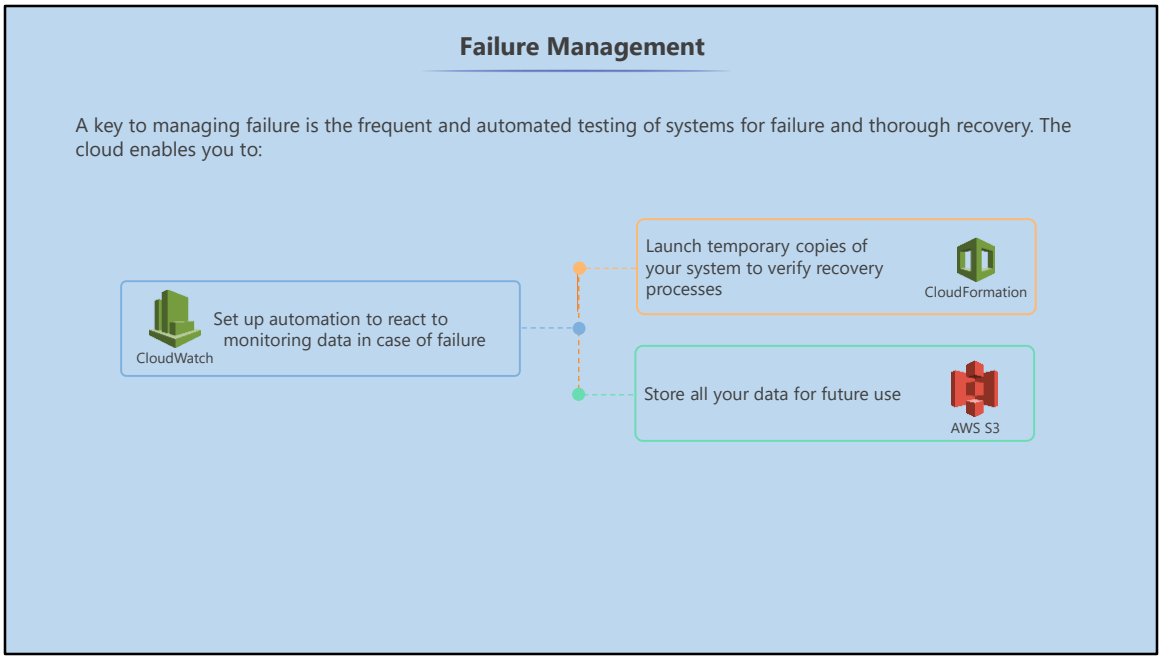
- Sufficient network bandwidth for the data center
- Sufficient compute capacity
- Well-trained staff
- Support Contracts, in case things go wrong
- Sufficient compute capacity

With the AWS Shared Responsibility model, network bandwidth and compute capacity is taken care of. AWS offers a range of support contracts to help you with issues and a range of training courses for your staff.



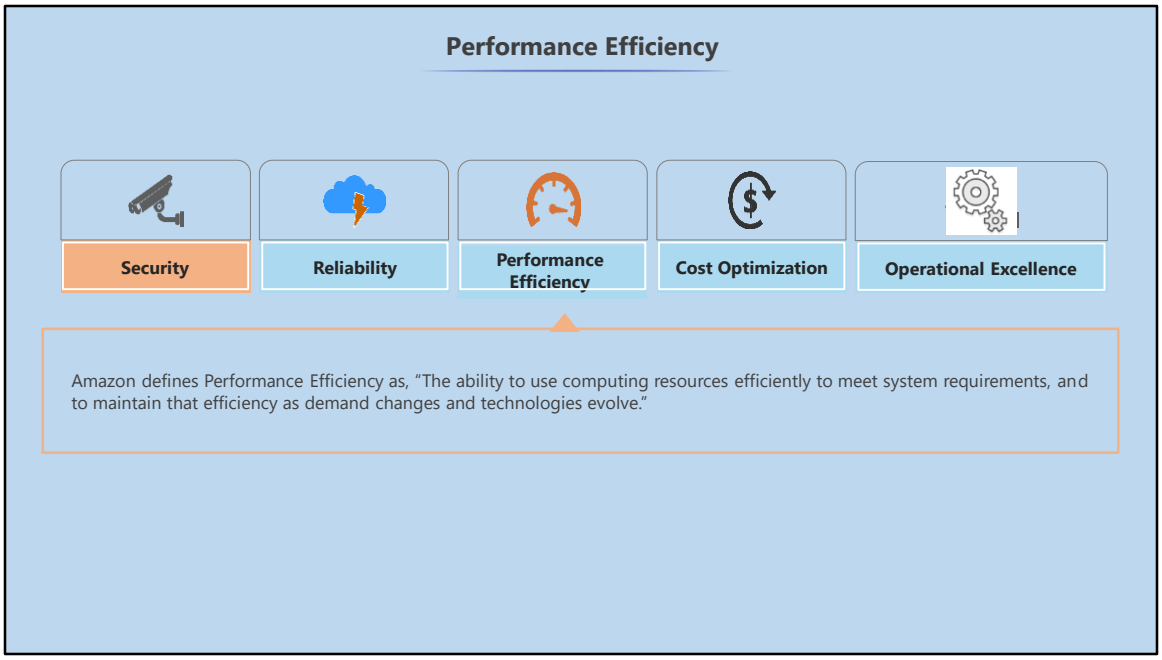
Being aware of how change affects a system allows you to plan proactively, and monitoring allows you to quickly identify trends that could lead to capacity issues or SLA breaches.

With AWS, you can monitor the behavior of a system and automate the response to KPIs, for example, adding additional servers as a system gains more users. You can control who has permission to make system changes and audit the history of these changes.

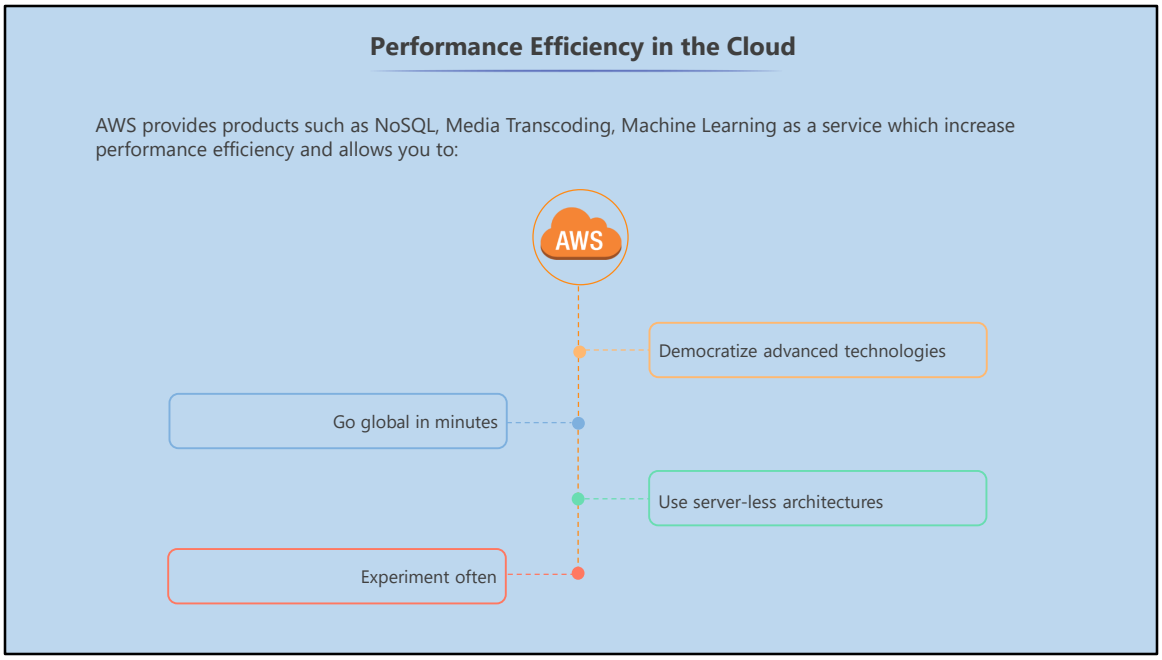


Failures will occur. It's a good practice to understand why things failed and prevent them from happening again. A key to managing failure is frequent and automated testing of systems for failure and thorough recovery (ideally on a regular schedule and also triggered after significant system changes).
The cloud enables you to launch temporary copies of a whole system at low cost and use automated testing to verify full recovery processes.
Set up automation to react to data that indicates a failure.

"Actively track KPIs, such as the recovery time objective or RTO and recovery point objective or RPO. RTO refers to the duration of time and a service level within which a business process must be restored after a disaster to avoid unacceptable consequences associated with a break in continuity. RPO refers to the age of files that must be recovered from backup storage for normal operations to resume if a computer, system, or network goes down as a result of a hardware, program, or communications failure."

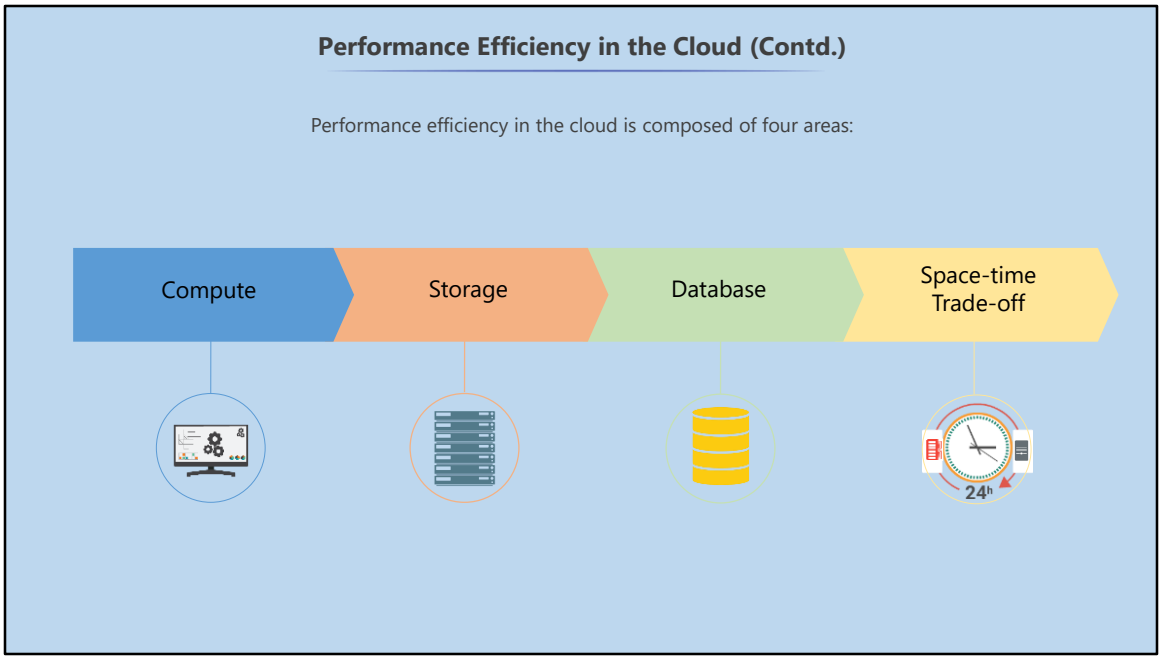


Performance Efficiency is the ability to use computing resources efficiently to meet system requirements, and maintain that efficiency as demand changes and technologies evolve.



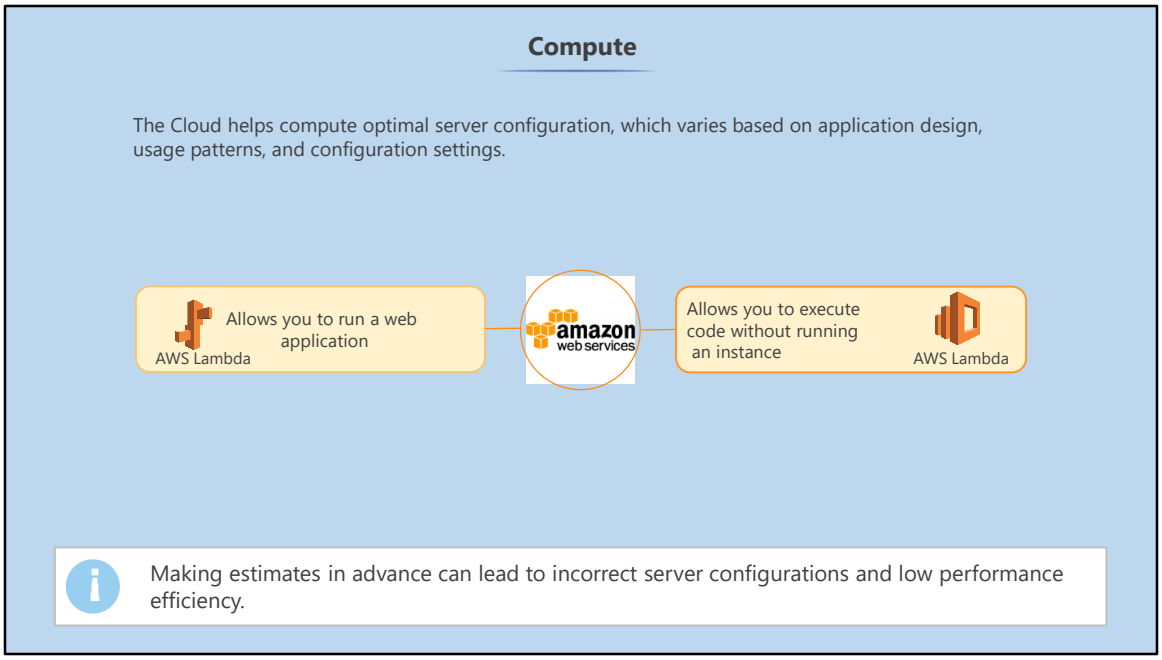
AWS provides products such as NoSQL, media transcoding, machine learning as a service which increase performance efficiency.

- Go global in minutes: Easily deploy your system in multiple regions around the world with just a few clicks, which leads to lower latency for your customers.
- Use server-less architectures: In the cloud, server-less architectures remove the need to run and maintain servers to carry out traditional compute activities. For example, storage services can act as static websites, removing the need for web servers. You can use services such as Lambda or Elastic Beanstalk.
- Experiment often: Virtual and automatable resources help to carry out comparative testing using different types of instances, storage, or configurations.

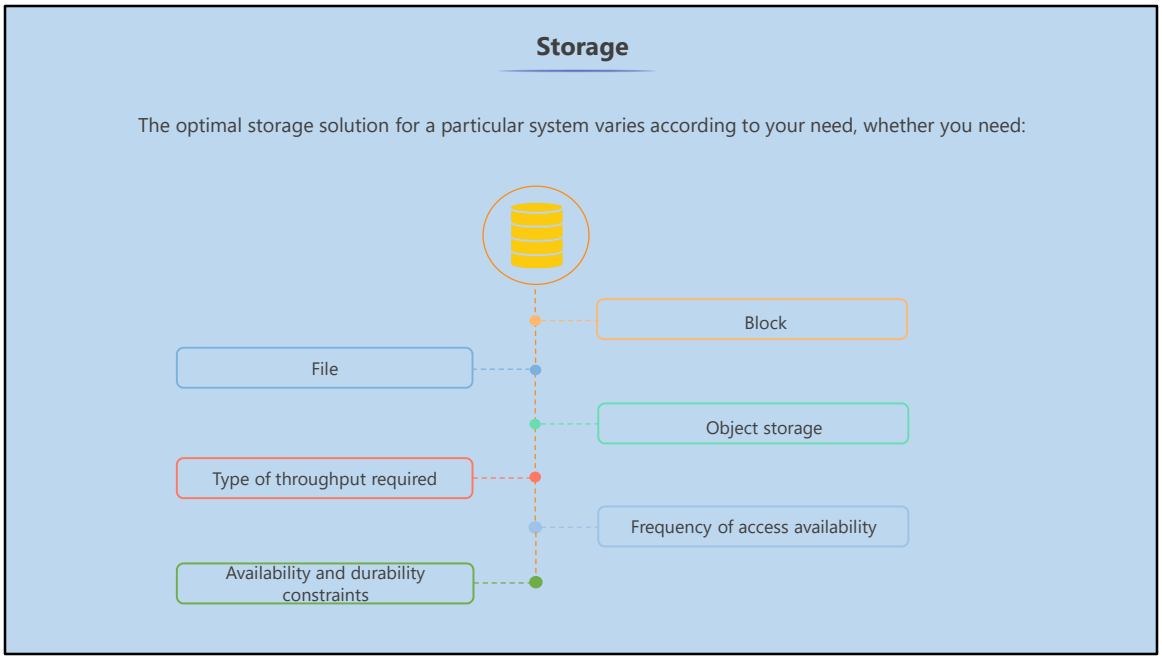


Performance efficiency in the cloud is composed of four areas:

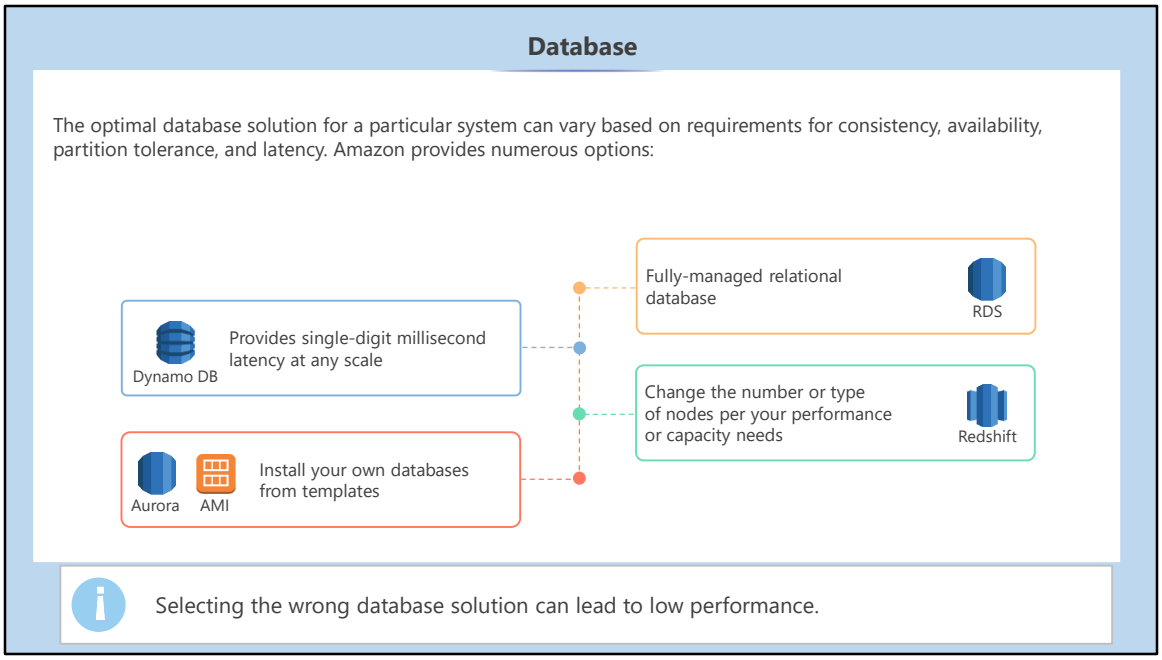
1. Compute
2. Storage
3. Database
4. Space-time trade-off



Finding the optimal server configuration for a product in the cloud will vary based on application design, usage patterns, and configuration settings.
Traditional IT architecture involves estimating in advance what can lead to incorrect server configurations and lower performance efficiency.
AWS is virtualized so you can quickly change the server configuration to increase its performance efficiency.
It may be optimal to run serverless computing. For example, AWS Lambda allows you to execute code without running an instance.
From an operational standpoint, monitoring should be in place to notify you of any degradation in performance.

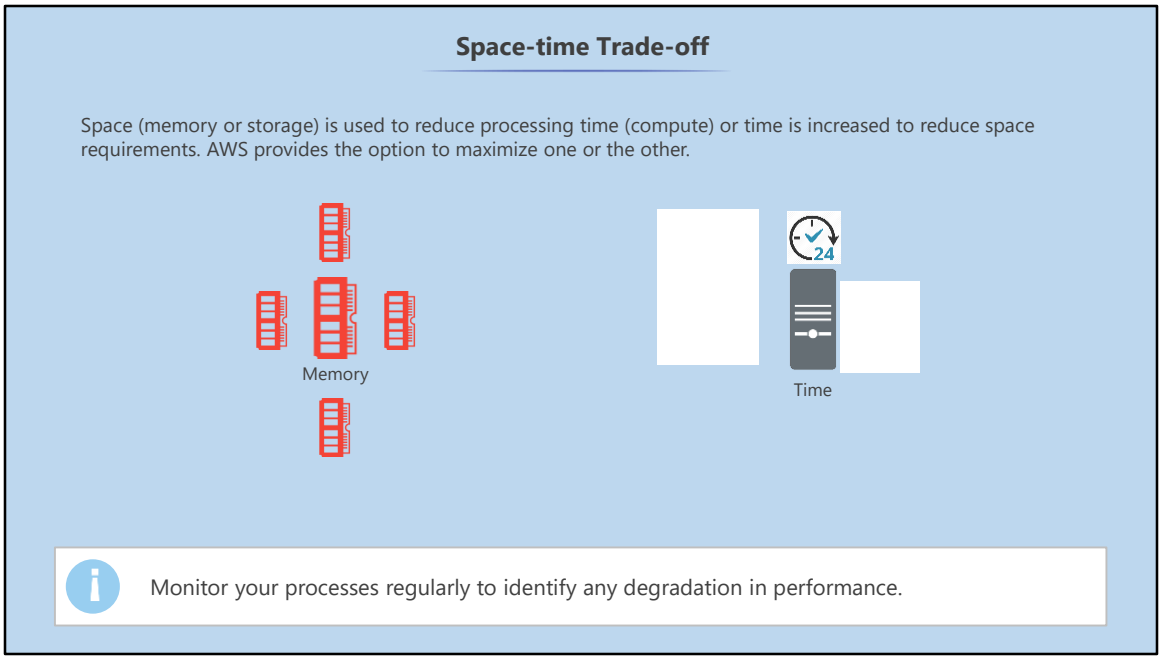


The optimal storage solution for a particular system varies according to your need, whether you need block, file, or object storage, the type of throughput required, and frequency of access, availability, and durability constraints.
Well-architected systems use multiple storage solutions and enable different features to improve performance. Storage is virtualized and is available in different types. This makes it easier to match storage methods more closely with your needs, and also offers storage options that are not easily achievable with on-premises infrastructure. For example, Amazon S3 is designed for 11 9s of durability. You can also change from using magnetic hard drives (HDDs) to solid state drives (SSDs), and easily move virtual drives from one instance to another in seconds.



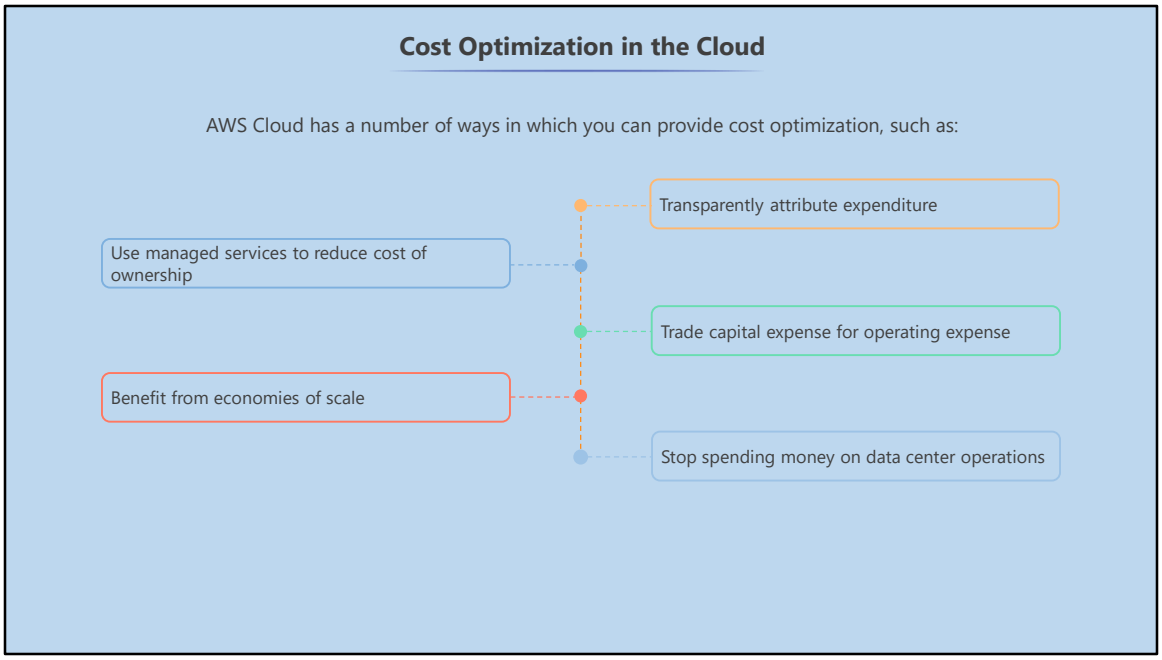
The optimal database solution for a particular system can vary based on requirements for consistency, availability, partition tolerance, and latency. Many systems use different database solutions for various sub-systems and enable different features to improve performance. Selecting the wrong database solution and features for a system can lead to lower performance efficiency.
The Amazon Relational Database Service (RDS) provides a fully managed relational database and the Amazon DynamoDB is a fully managed NoSQL database that provides single-digit millisecond latency at any scale.

Amazon Redshift is a managed petabyte-scale data warehouse that allows you to change the number or type of nodes per your performance or capacity needs.
Use Aurora or AMI to install your own databases from templates.



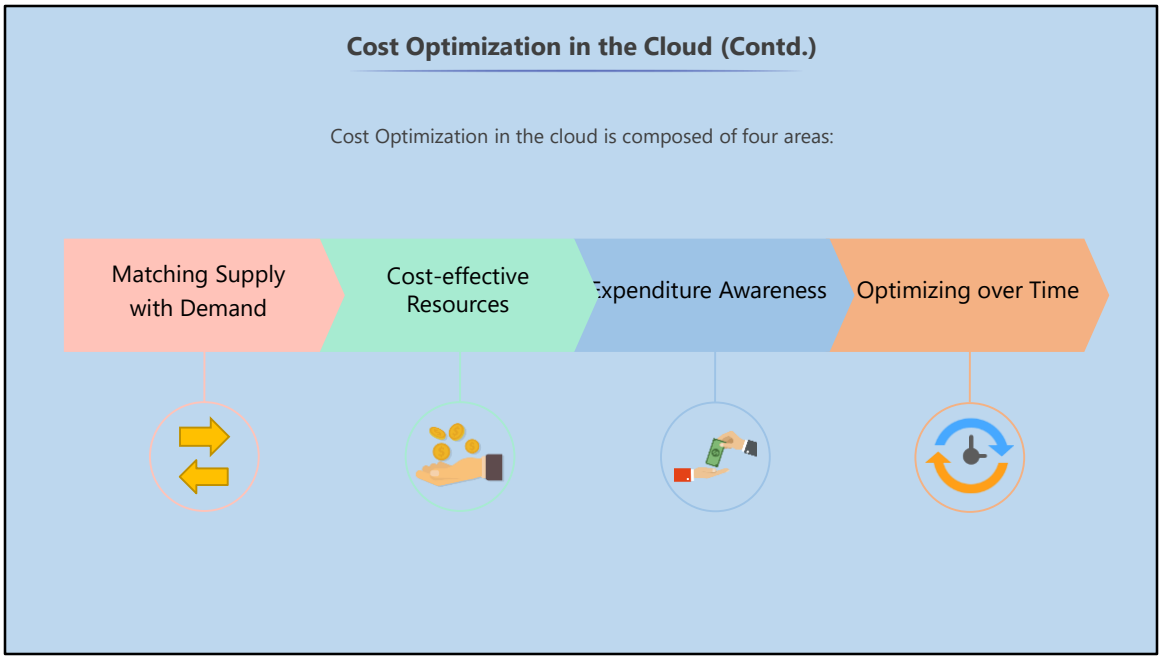
Space (memory or storage) is used to reduce processing time (compute) or time is increased to reduce space requirements.
AWS provides the option to maximize one or the other. For example, you can launch systems globally to be closer to your end users to reduce latency (time) or you could use Amazon ElastiCache as a cache to improve performance efficiency. You need to have monitoring in place to notify you of any degradation in performance.

Cost optimization is the ability to avoid or eliminate unneeded cost or suboptimal resources.



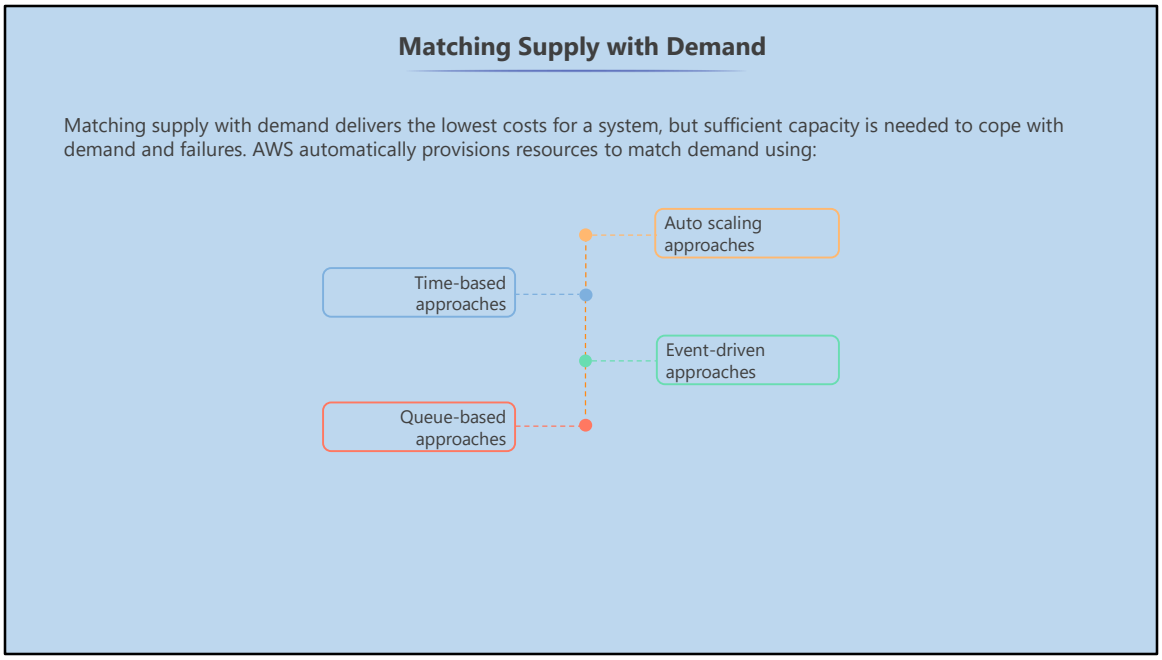
AWS Cloud helps you provide cost optimization in many ways, such as:

- Transparently attribute expenditure: The cloud makes it easier to identify the cost of a system and attribute it to individual business owners.
- Use managed services to reduce cost of ownership: In the cloud, managed services reduce the operational burden of maintaining servers for tasks such as sending email or managing databases.
- Trade capital expense for operating expense: Instead of investing heavily in data centers and servers before you know how you're going to use them, pay only for the computing resources you consume and when you consume them.
- Benefit from economies of scale: By using cloud computing, you can achieve a lower variable cost than you could on your own.
- Stop spending money on data center operations: AWS does the heavy lifting of racking, stacking, and powering servers, so you can focus on your customers and business projects rather than on the IT infrastructure.



Cost Optimization in the cloud is composed of four areas:

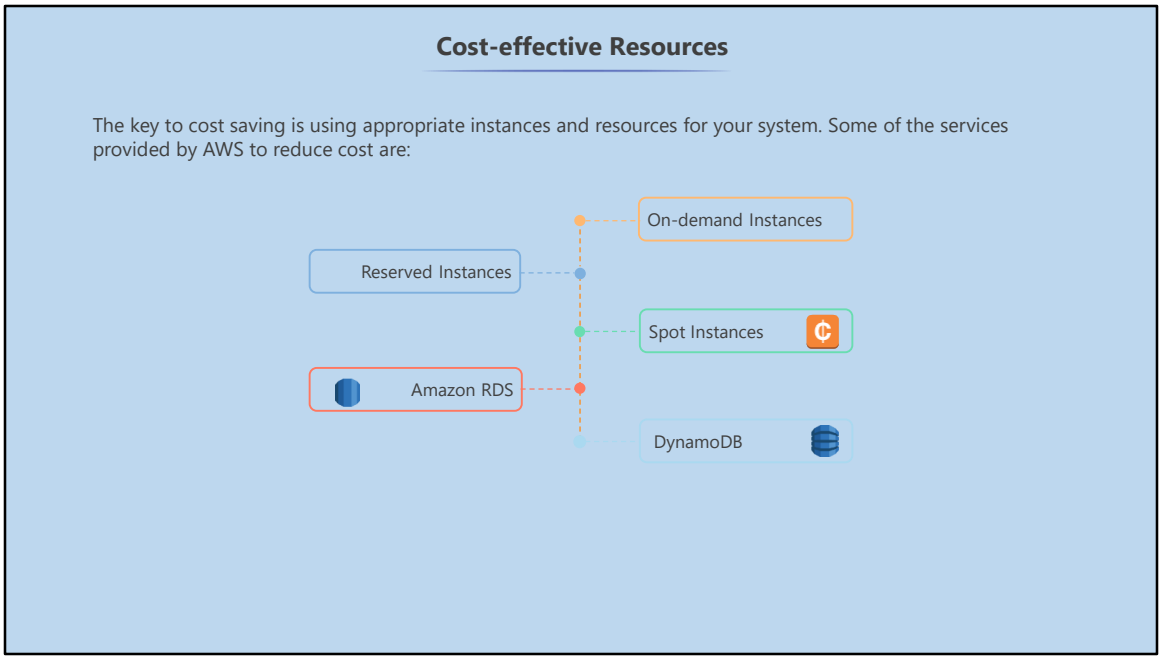
1. **Matching supply with demand**
2. Cost-effective resources
3. Expenditure awareness
4. Optimizing over time



Matching supply with demand delivers the lowest costs for a system, but sufficient capacity is needed to cope with demand and failures.

AWS automatically provisions resources to match demand. Auto Scaling and time-based, event-driven, and queue-based approaches allow you to add and remove resources as needed.

Monitoring tools and regular benchmarking can help you achieve much greater utilization of resources.



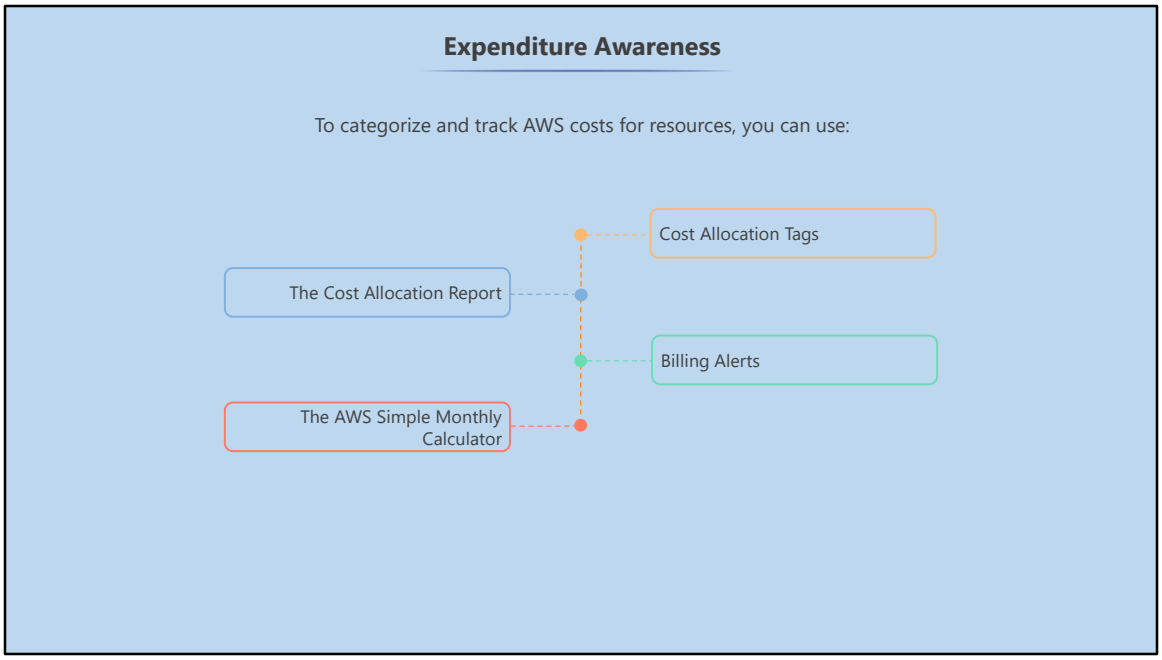
The key to cost saving is using appropriate instances and resources for your system. For example, a reporting process might take five hours to run on a smaller server, but a larger server, which is twice as expensive, can do it in one hour. Both jobs give you the same outcome, but the smaller server will incur more cost over time.

On-Demand instances allow you to pay for compute capacity by the hour, with no minimum commitments required.

Reserved Instances (RIs) allow you to reserve capacity and offer savings of up to 75 percent on-demand pricing.

With Spot instances, you can bid on unused Amazon EC2 capacity at significant discounts.

You can also take advantage of managed AWS services, such as Amazon RDS, Amazon Elastic MapReduce (EMR), and Amazon DynamoDB, which can lower costs.

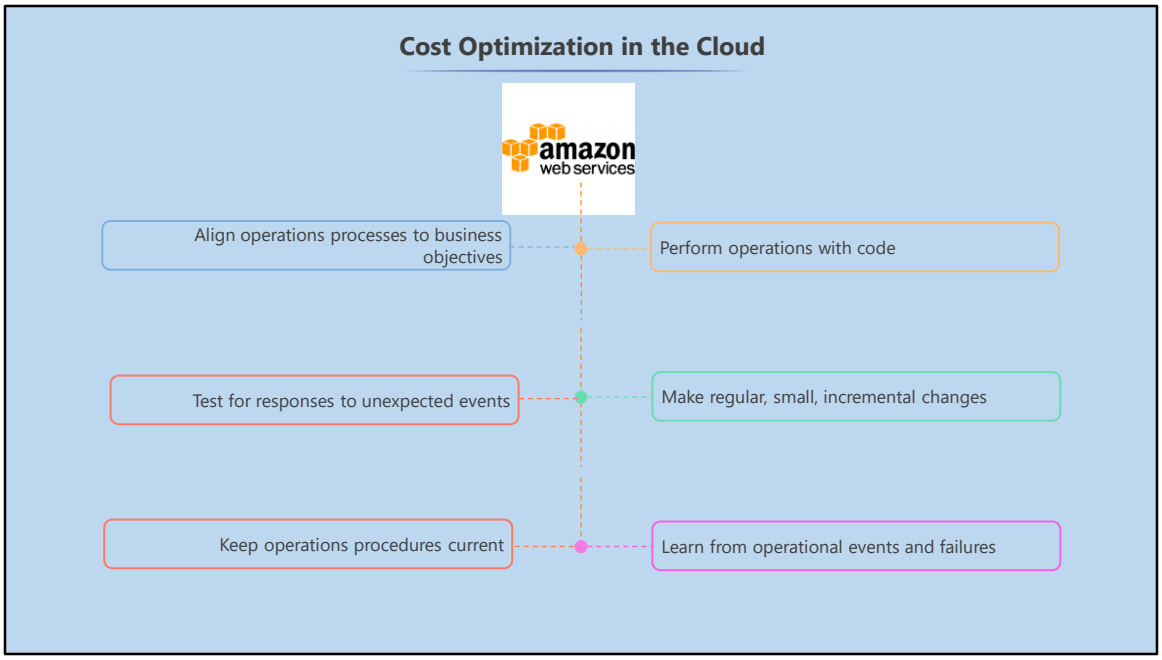


Use cost allocation tags to categorize and track AWS costs for AWS resources.

- AWS generates a cost allocation report with usage and costs aggregated by your tags. You can set up tags for each of your business categories units.
- This increased visibility of costs makes it easier to identify resources or projects that are no longer generating value and should be decommissioned.
- Billing alerts can be set up to notify predicted overspending.
- The AWS Simple Monthly Calculator allows you to calculate data transfer costs.

AWS always releases new products and services; therefore, it helps to reassess your existing setup to see if it is the most cost effective. For example, running an Amazon RDS database might be cheaper than running your own database on Amazon EC2. Or, Lambda might be more cost effective than EC2.

Cost optimization is the ability to avoid or eliminate unneeded cost or suboptimal resources.



Perform operations with code: When there are common repetitive processes or procedures, use automation. For example, consider automating configuration management, changes, and responses to events.

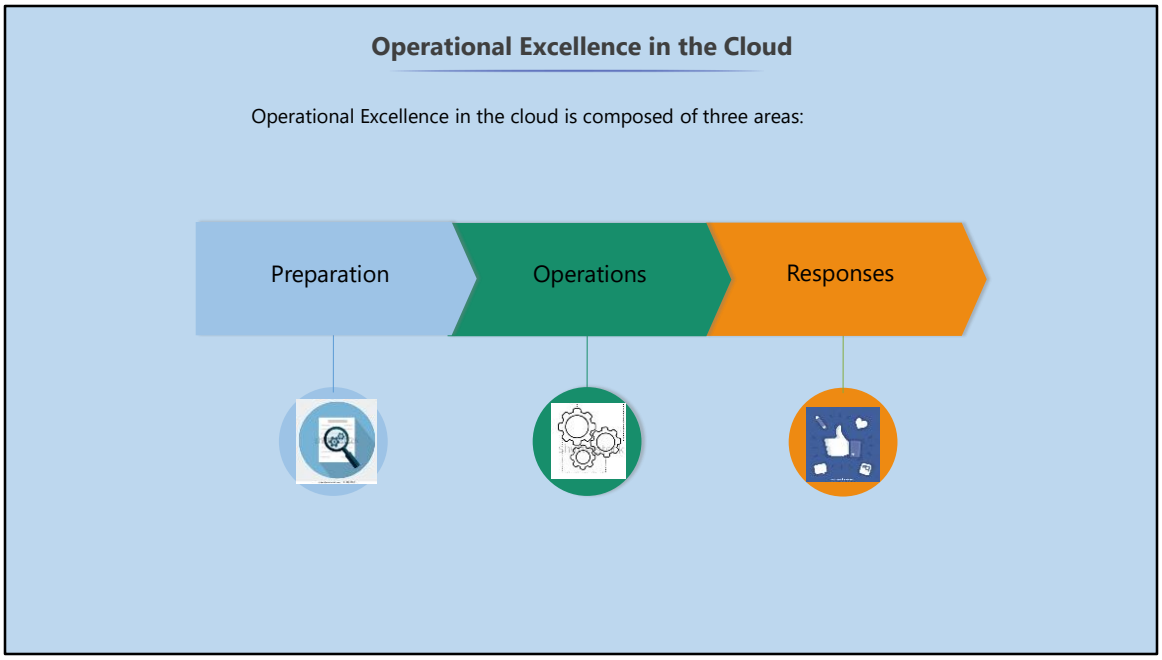
Align operations processes to business objectives: Collect metrics that indicate operational excellence in meeting business objectives. The goal should be to reduce the signal to noise ratio in metrics, so operational monitoring and responses are targeted to support business-critical needs. Collecting metrics that are unnecessary will prevent effective responses to unexpected operational events by complicating monitoring and response.

Make regular, small, incremental changes: Workloads should be designed to allow components to be updated regularly. Changes should be done in small increments, not large batches, and should be able to be rolled back without affecting operations.

Test for responses to unexpected events: Workloads should be tested for component failures and other unexpected operational events. It is important to test and understand procedures for responding to operational events, so that they are followed when operational events occur.

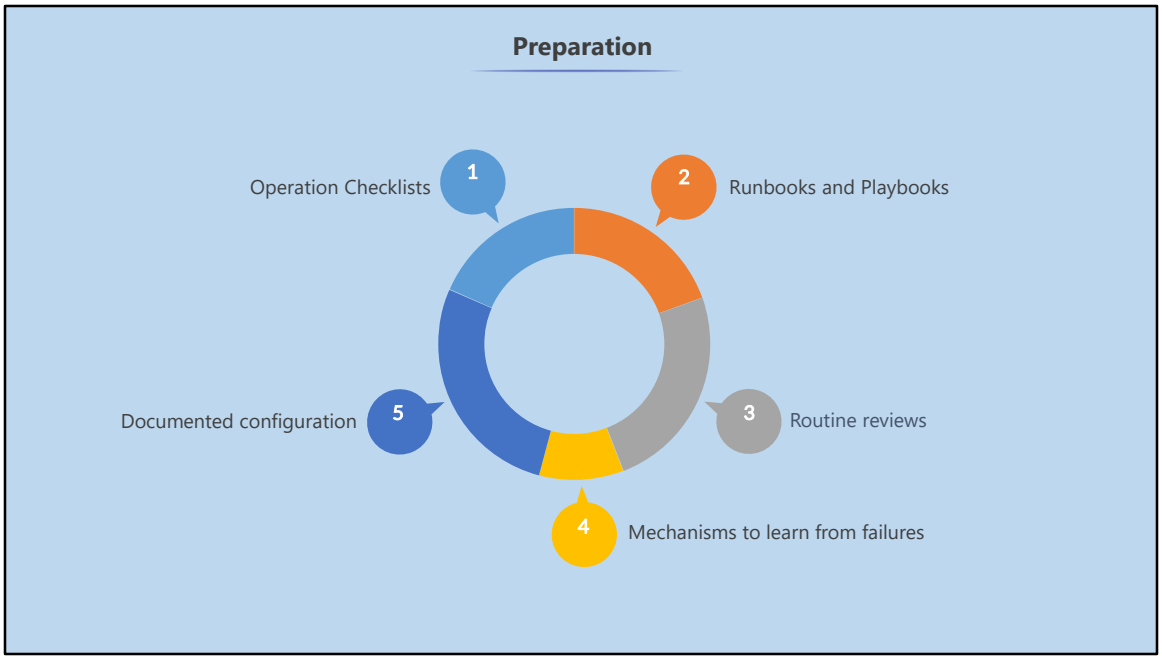
Learn from operational events and failures: Processes should be in place so that all types of operational events and failures are captured, reviewed, and then used for improvements. Regular cross-functional operations reviews should result in process improvements that drive operational excellence.

Keep operations procedures current: Process and procedure guides should be adapted as environments and operations evolve. This includes updating regular operations runbooks (standard operations procedures), as well as playbooks (response plans for unexpected operational events or production failures). Guidance and learnings for operations should be shared between teams to prevent repeated mistakes. Consider using a wiki or an internal knowledge base for this information. The information that should be evaluated includes operations metrics, unexpected anomalies, failed deployments, system failures, and ineffective or inappropriate responses to failures. System and architecture documentation should also be captured and updated using automation as environments and operations evolve



Operational Excellence in the cloud is composed of three areas:

1. Preparation
2. Operations
3. Responses



Operations checklists will ensure that workloads are ready for production operation, and prevent unintentional production promotion without effective preparation.

Runbooks & Playbooks: Operations teams should be able to perform normal daily tasks using runbooks, as well as guidance for responding to unexpected operational events (playbooks).

Routine reviews of business cycle events that can drive changes in operations should also be performed. All runbooks and playbooks should be fully tested so that gaps or challenges can be identified and any potential risk can be mitigated.

Mechanisms to track and learn from failures should be in place.

Environments, architecture, and the configuration parameters for resources within them, should be documented in a way that allows components to be easily identified for tracking and troubleshooting. Changes to configuration should also be trackable and automated.

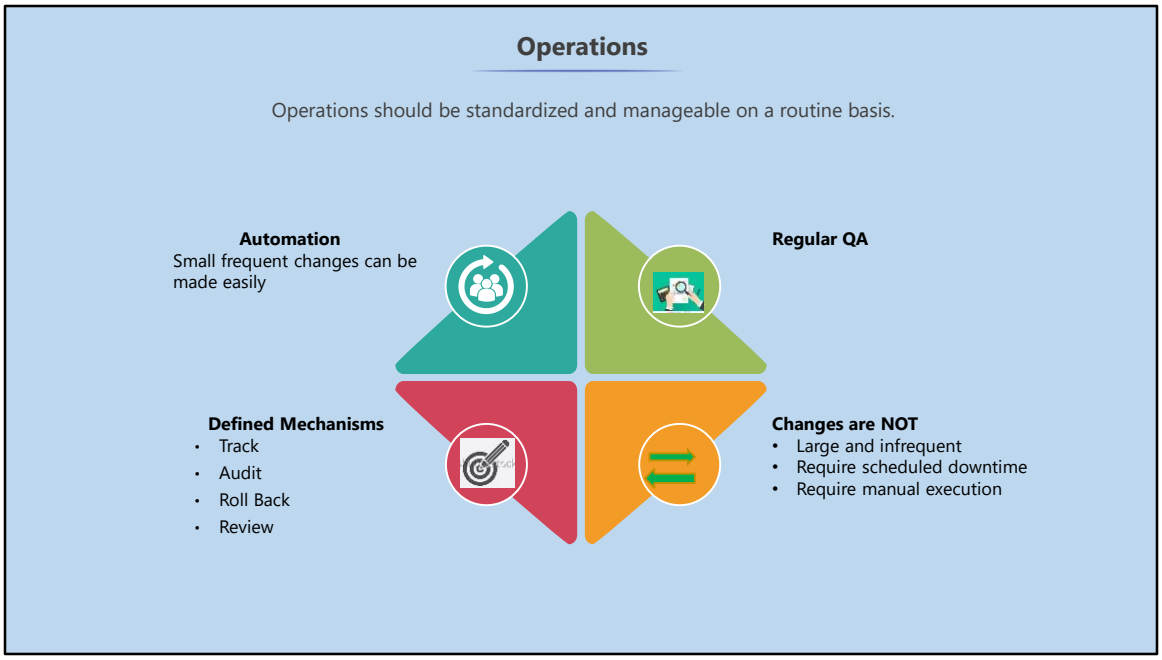
Preparation (Contd.)

- AWS services such as AWS CloudFormation can be used to ensure that environments contain all required resources when deployed in production, which reduces the opportunity for human error.
- Implementing Auto Scaling will allow workloads to automatically respond when business-related events affect operational needs.
- Services like AWS Config with the AWS Config rules feature create mechanisms to automatically track and respond to changes in your AWS workloads and environments.
- AWS Trusted advisor can highlight any areas for improvement.

AWS services such as AWS CloudFormation can be used to ensure that environments contain all required resources when deployed in production, which reduces the opportunity for human error.

Implementing Auto Scaling will allow workloads to automatically respond when business-related events affect operational needs.

Services like AWS Config with the AWS Config rules feature create mechanisms to automatically track and respond to changes in your AWS workloads and environments.
AWS Trusted advisor can highlight any areas for improvement.



Operations should be standardized and manageable on a routine basis.

Automation is key so that small frequent changes can be made easily.

As well as regular quality assurance testing,

and defined mechanisms to track, audit, roll back, and review changes.

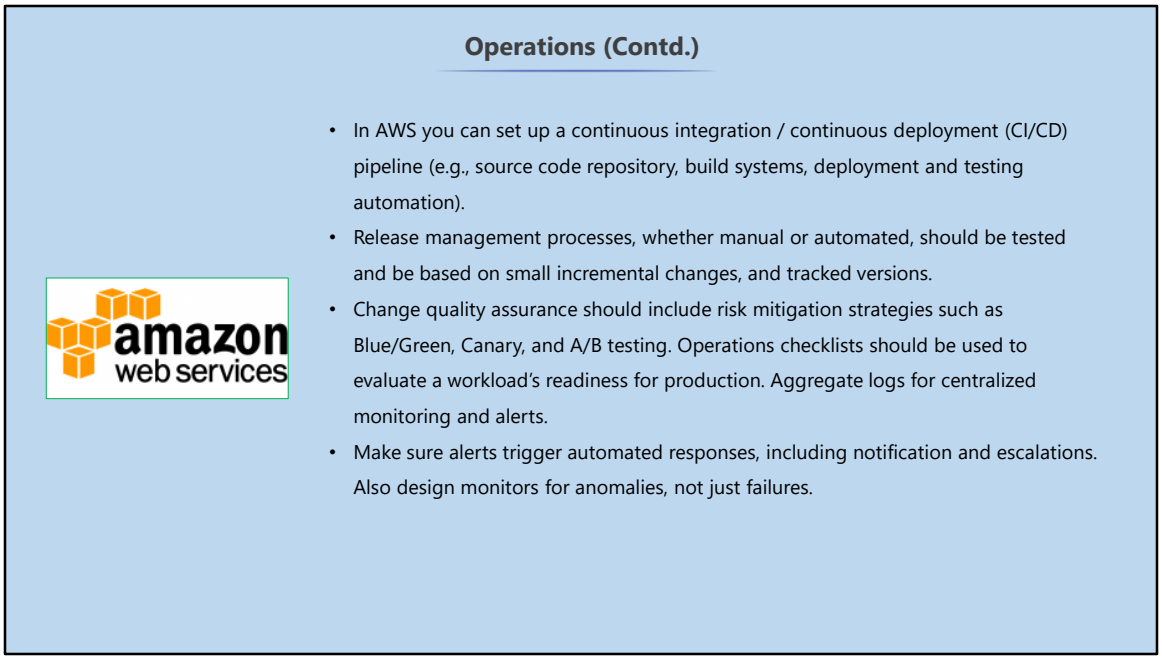
Changes should not be large and infrequent, they should not require scheduled downtime, and they should not require manual execution.

In AWS you can set up a continuous integration / continuous deployment (CI/CD) pipeline (e.g., source code repository, build systems, deployment and testing automation).

Release management processes, whether manual or automated, should be tested and be based on small incremental changes, and tracked versions. You should be able to revert changes that introduce operational issues without causing operational impact.

Change quality assurance should include risk mitigation strategies such as Blue/Green, Canary, and A/B testing. Operations checklists should be used to evaluate a workload's readiness for production. Aggregate logs for centralized monitoring and alerts.

Make sure alerts trigger automated responses, including notification and escalations. Also design monitors for anomalies, not just failures. Amazon Web Services

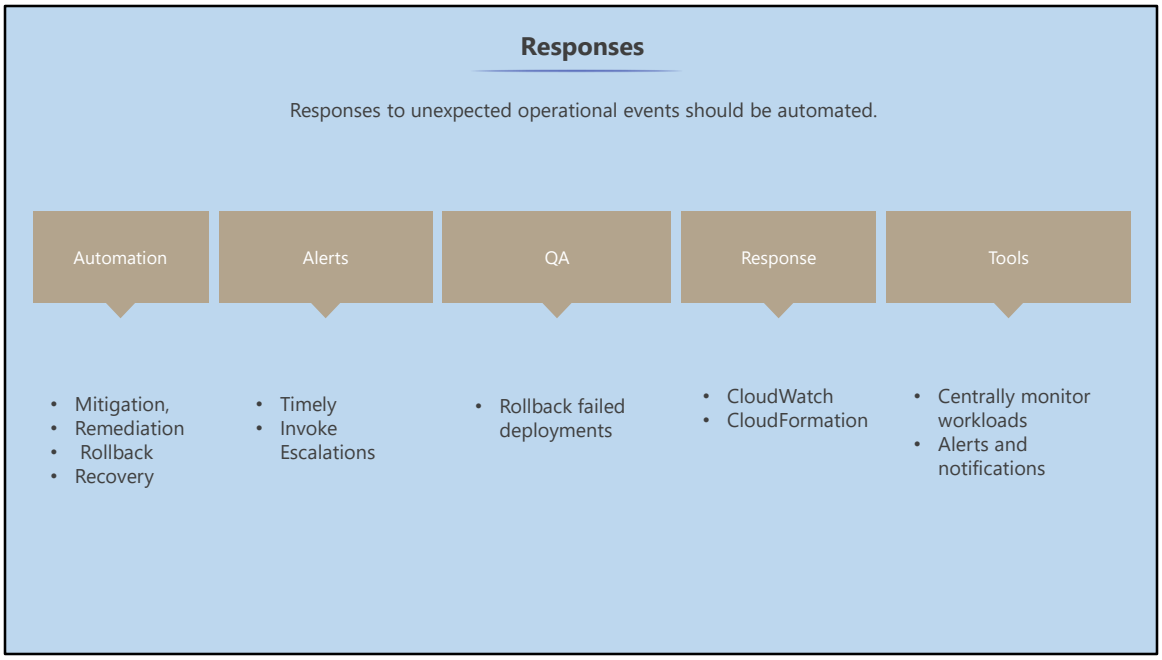


In AWS you can set up a continuous integration / continuous deployment (CI/CD) pipeline (e.g., source code repository, build systems, deployment and testing automation).

Release management processes, whether manual or automated, should be tested and be based on small incremental changes, and tracked versions. You should be able to revert changes that introduce operational issues without causing operational impact.

Change quality assurance should include risk mitigation strategies such as Blue/Green, Canary, and A/B testing. Operations checklists should be used to evaluate a workload's readiness for production. Aggregate logs for centralized monitoring and alerts.

Make sure alerts trigger automated responses, including notification and escalations. Also design monitors for anomalies, not just failures.



Responses to unexpected operational events should be automated.

This is not just for alerting, but also for mitigation, remediation, rollback, and recovery.

Alerts should be timely, and should invoke escalations when responses are not adequate to mitigate the impact of operational events.

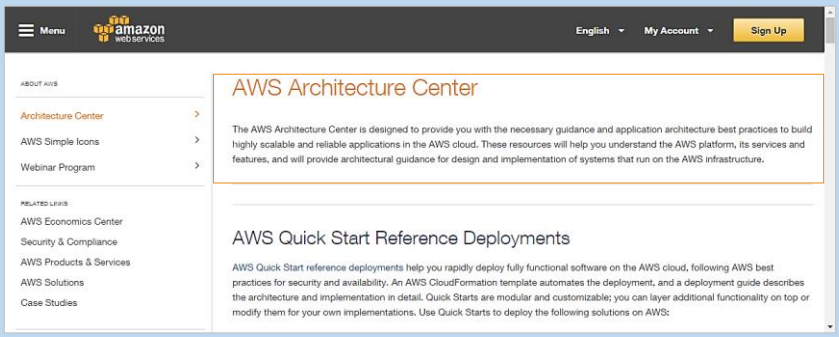
Quality assurance mechanisms should be in place to automatically roll back failed deployments.

To ensure both appropriate alerting and notification in response to unplanned operational events, as well as automated responses you can use CloudWatch and CloudFormation.

Tools should also be in place to centrally monitor workloads and create effective alerts and notifications based on all available logs and metrics.

Accessing More Information

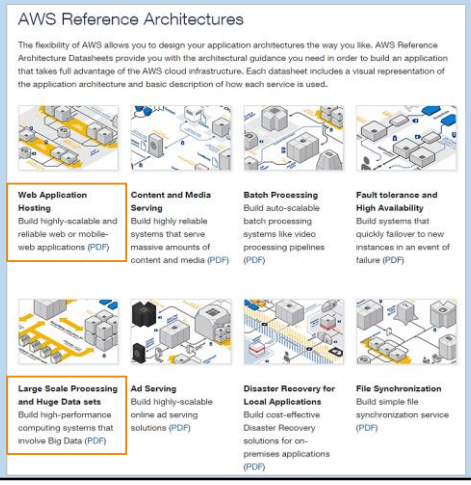
The AWS Architecture Center provides guidance and application architecture best practices to build highly scalable and reliable applications in the AWS cloud.



The AWS Architecture Center is a resource that provides guidance and application architecture best practices to build highly scalable and reliable applications in the AWS cloud.

AWS Reference Architectures

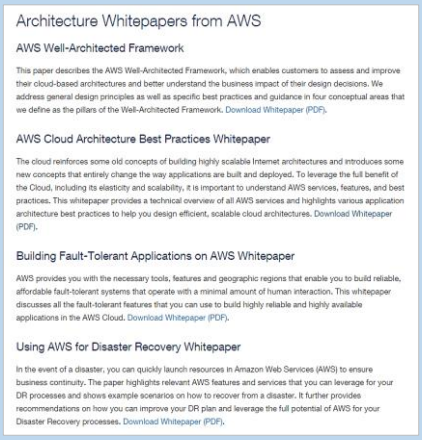
The AWS Reference Architecture Datasheets provide architectural guidance to build an application on the AWS cloud infrastructure.



AWS Reference Architecture Datasheets provide architectural guidance to build an application on the AWS cloud infrastructure. Datasheets include a basic description of how each service is used plus visual representations of the application architecture.

AWS Whitepapers

The technical AWS whitepapers cover all AWS related topics such as architecture, security, and economics.



Architecture Whitepapers from AWS

AWS Well-Architected Framework

This paper describes the AWS Well-Architected Framework, which enables customers to assess and improve their cloud-based architectures and better understand the business impact of their design decisions. We address general design principles as well as specific best practices and guidance in four conceptual areas that we define as the pillars of the Well-Architected Framework. Download Whitepaper (PDF).

AWS Cloud Architecture Best Practices Whitepaper

The cloud reinforces some old concepts of building highly scalable Internet architectures and introduces some new concepts that entirely change the way applications are built and deployed. To leverage the full benefit of the Cloud, including its elasticity and scalability, it is important to understand AWS services, features, and best practices. This whitepaper provides a technical overview of all AWS services and highlights various application architecture best practices to help you design efficient, scalable cloud architectures. Download Whitepaper (PDF).

Building Fault-Tolerant Applications on AWS Whitepaper

AWS provides you with the necessary tools, features and geographic regions that enable you to build reliable, affordable fault-tolerant systems that operate with a minimal amount of human interaction. This whitepaper discusses all the fault-tolerant features that you can use to build highly reliable and highly available applications in the AWS Cloud. Download Whitepaper (PDF).

Using AWS for Disaster Recovery Whitepaper

In the event of a disaster, you can quickly launch resources in Amazon Web Services (AWS) to ensure business continuity. The paper highlights relevant AWS features and services that you can leverage for your DR processes and shows example scenarios on how to recover from a disaster. It further provides recommendations on how you can improve your DR plan and leverage the full potential of AWS for your Disaster Recovery processes. Download Whitepaper (PDF).

There is a comprehensive list of technical AWS whitepapers that cover all AWS related topics such as architecture, security, and economics. New whitepapers are being released all the time and are written by some of the most knowledgeable AWS team or community.

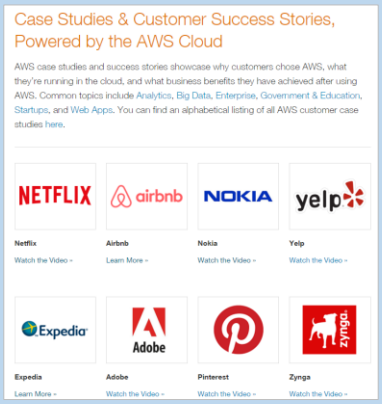
AWS Quick Start Reference Deployments

You can rapidly deploy a fully functional environment for a number of enterprise software applications using the AWS CloudFormation templates.

. Using AWS CloudFormation templates, you can rapidly deploy a fully functional environment for a variety of enterprise software applications. Supplementary deployment guides describe the architecture and implementation in detail.
In this course we will deploy things manually, but with cloud formation templates you can do several hours of work just with the click of a button.
Examples are Sharepoint, RD Gateway, SQL Server on WSFC.

Case Studies

AWS maintains a large list of case studies and success stories from their clients. These case studies can be used to explain how, and why some of the largest and most successful companies use AWS for their business.



AWS maintains a large list of case studies and success stories from their clients. You can check how and why some of the largest and most successful companies use AWS for their business.

Knowledge Check

1 **The AWS Well-Architected Framework is designed to help you _____.**

Stop guessing your capacity needs

Test systems at production scale

Lower the risk of architecture change

Increase the amount of administration required

1 **The AWS Well-Architected Framework is designed to help you _____.**

Stop guessing your capacity needs

Test systems at production scale

Lower the risk of architecture change

Increase the amount of administration required

a), b), and c)

The AWS Well-Architected Framework is designed to help you understand the pros and cons of the decisions you make while building systems on AWS. So you can stop guessing your capacity needs, test systems at production scale, lower the risk of architecture change, automate to make architectural experimentation easier, and allow for evolutionary architectures.

2 **The AWS Well-Architected Framework is based on which of the following five pillars:**

Security, Reliability, Performance Efficiency, Cost Optimization, and Operational Excellence

Security, Redundancy, Performance Efficiency, Cost Optimization, and Operational Excellence

Security, Reliability, Environmental Efficiency, Cost Optimization, and Operational Excellence

Security, Redundancy, Performance Efficiency, Resource Optimization, and Operational Excellence

2 **The AWS Well-Architected Framework is based on which of the following five pillars:**

Security, Reliability, Performance Efficiency, Cost Optimization, and Operational Excellence

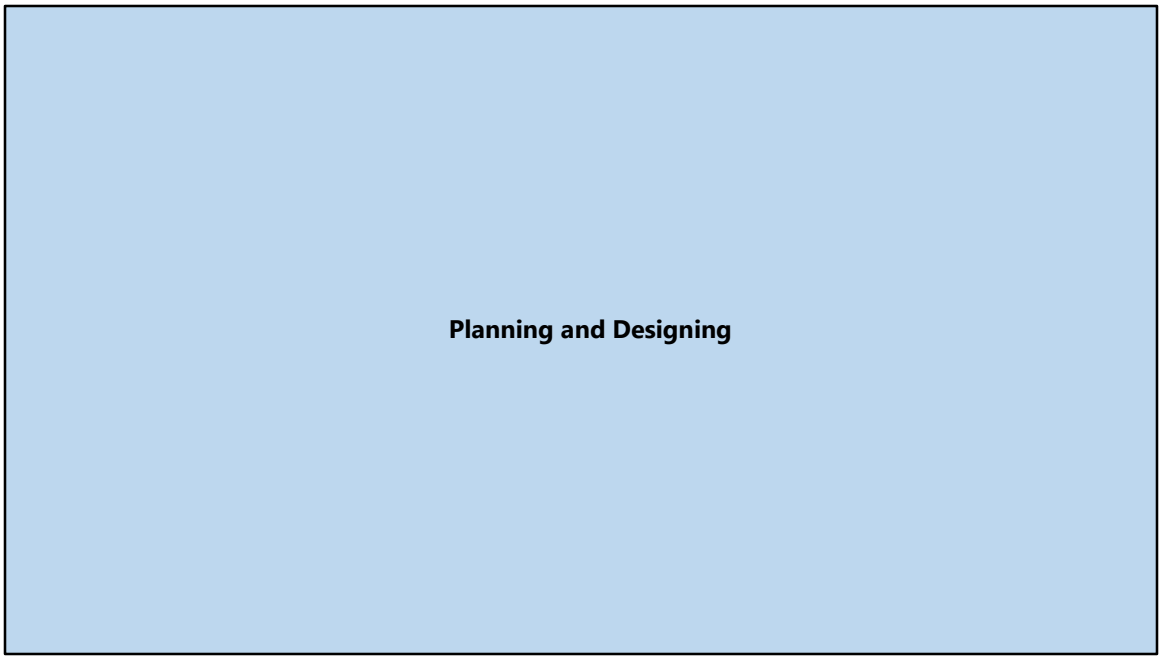
Security, Redundancy, Performance Efficiency, Cost Optimization, and Operational Excellence

Security, Reliability, Environmental Efficiency, Cost Optimization, and Operational Excellence

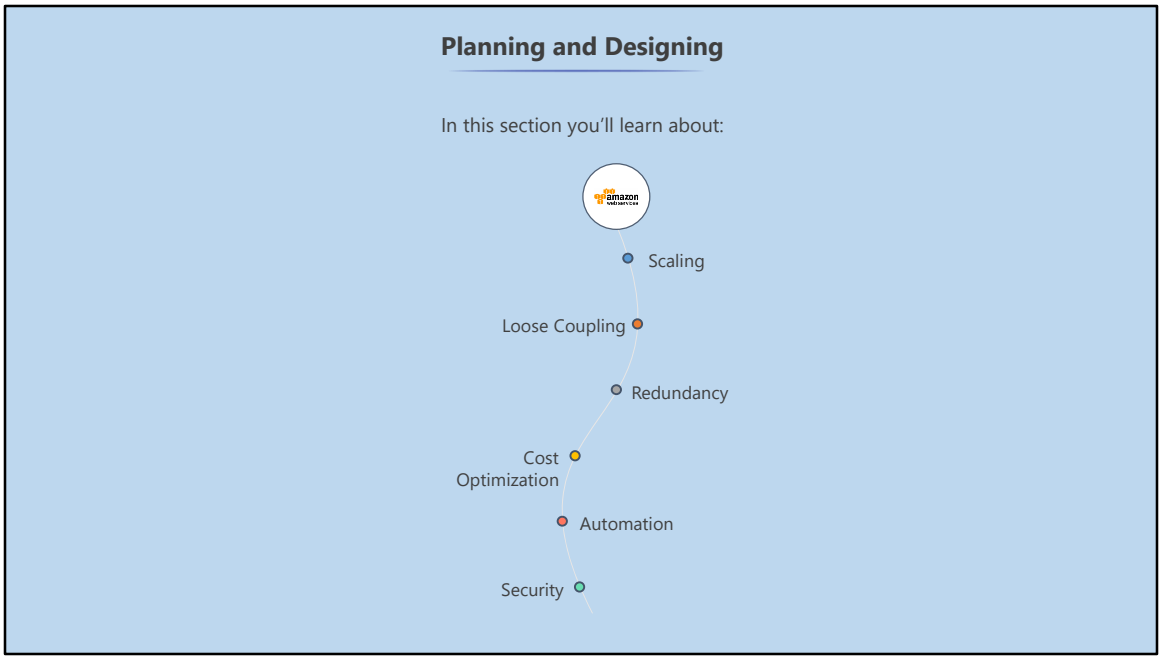
Security, Redundancy, Performance Efficiency, Resource Optimization, and Operational Excellence

a

The AWS Well-Architected Framework is based on Security, Reliability, Performance Efficiency, Cost Optimization, and Operational Excellence.



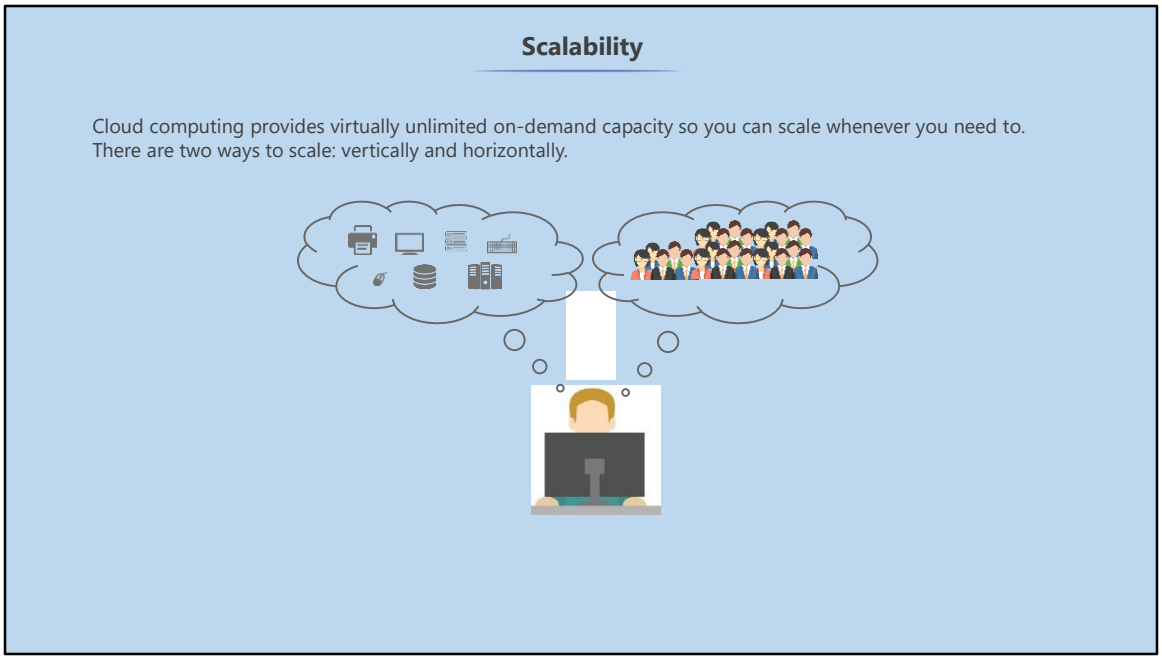
This section covers the principles involved in planning and designing cloud infrastructure.



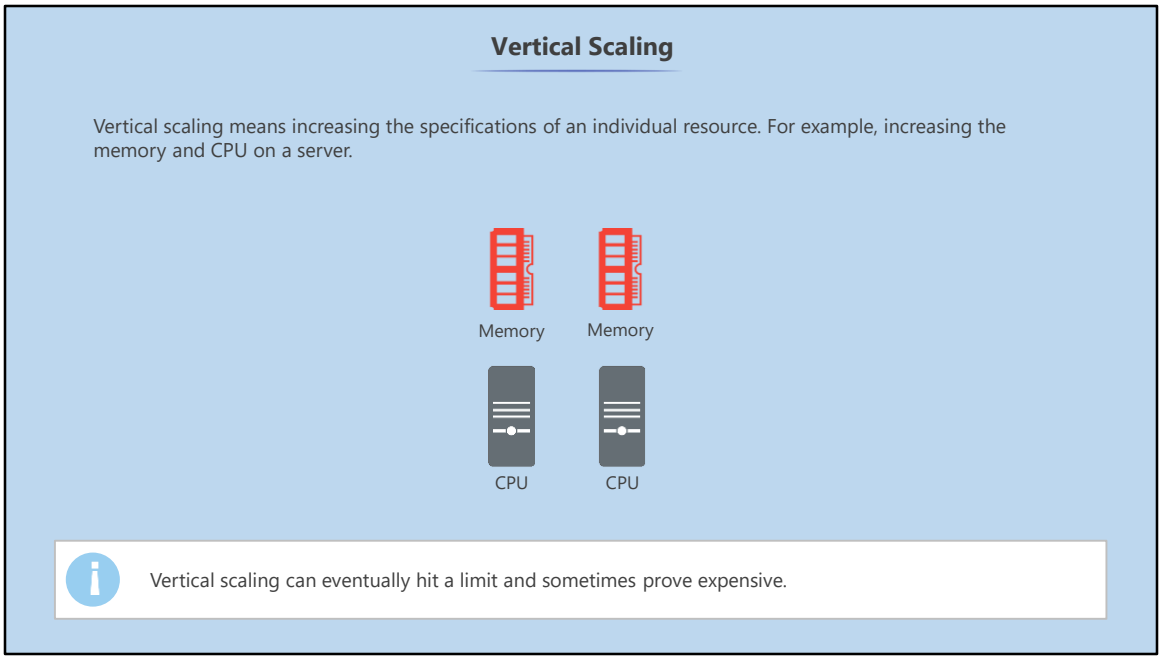
You will learn how the AWS Well-Architected Framework is used in planning and designing.

We will cover:

- Scaling
- Loose Coupling
- Redundancy
- Cost Optimization
- Automation
- Security



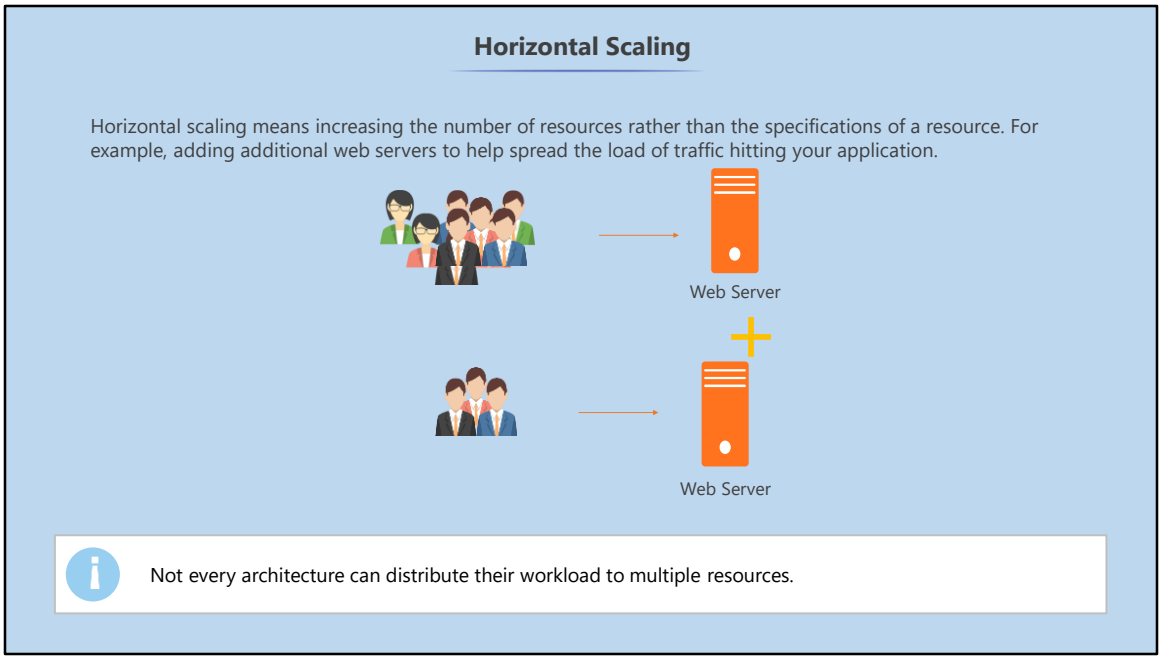
Cloud computing provides virtually unlimited on-demand capacity so you can scale when required. To do this your design needs to be able to seamlessly take advantage of these resources.
There are two ways to scale–vertically and horizontally.



Vertical scaling means increasing the specifications of an individual resource. For example, increasing the memory and CPU on a server.

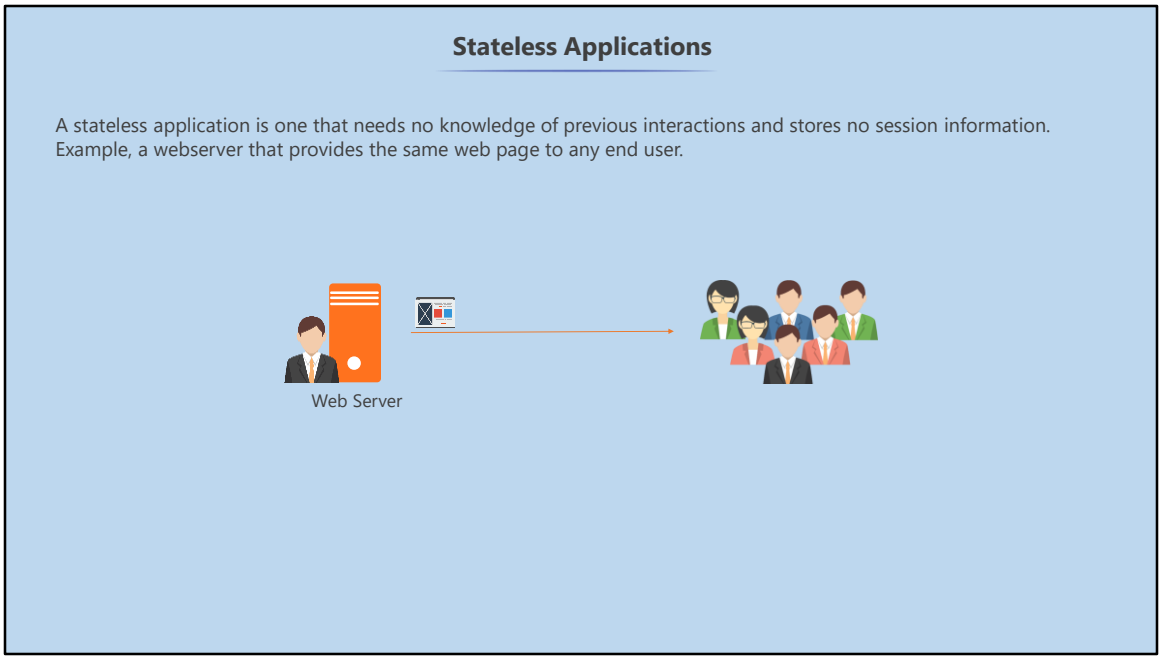
This is easily achieved with AWS, restart your virtual server to resize to a larger or smaller instance type.

The advantage of this approach is that it is very easy and requires little thought in the short term; however, this type of scaling can eventually hit a limit and can sometimes be cost inefficient.



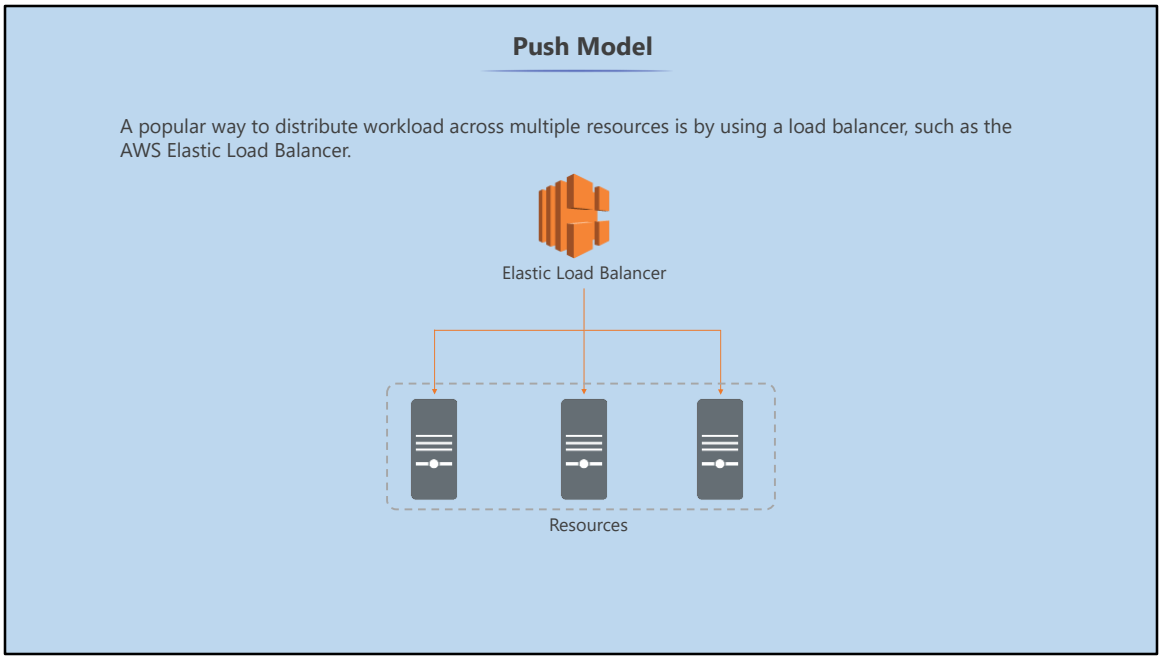
Horizontal scaling means increasing the number of resources rather than the specifications of a resource. For example, adding additional web servers to help spread the load of traffic hitting your application.
This is a great way to build applications that leverage the elasticity of cloud computing.

However, not all architectures can distribute their workload to multiple resources.



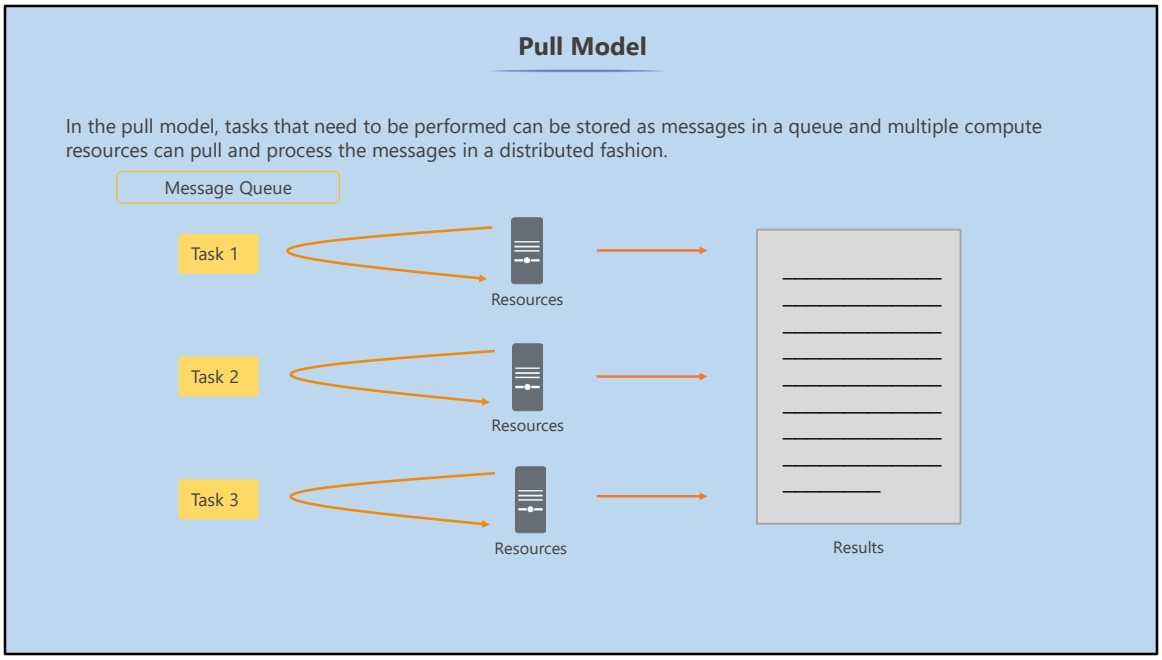
A stateless application is one that needs no knowledge of previous interactions and stores no session information. An example of this would be a webserver that provides the same web page to any end user.

Stateless applications scale horizontally as a request from an end user can be serviced by any of the available compute resources. Based on capacity requirements, resources can be added or removed. Stateless applications do not need to be aware of their peers, all that is required is a way to distribute the end user sessions.



A load balancer, such as the AWS Elastic Load Balancer, is a popular way to distribute workload across multiple resources.

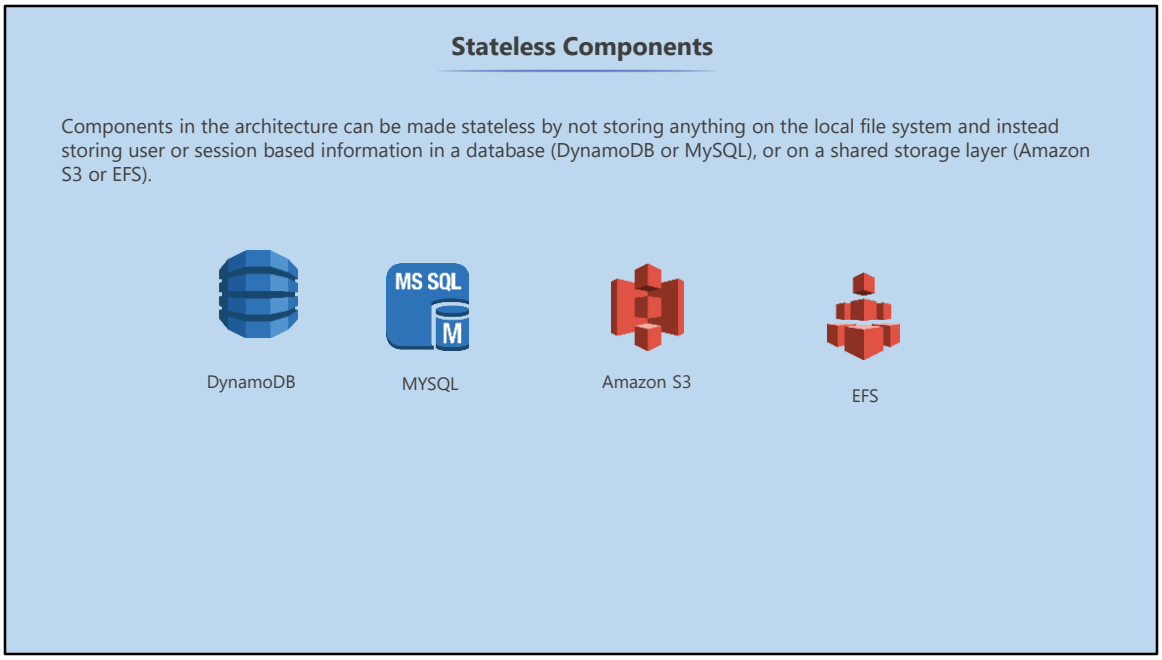
An alternative, but less recommended approach, is to implement a DNS round robin using Amazon Route 53 where DNS responses return an IP address from a list of valid hosts in a round robin fashion. This option is easy to set up, but can cause issues due to cached DNS records which are outside the control of Route 53 and might not always respect your set configuration settings.



In the pull model, tasks that need to be performed can be stored as messages in a queue and multiple compute resources can pull and process the messages in a distributed fashion.

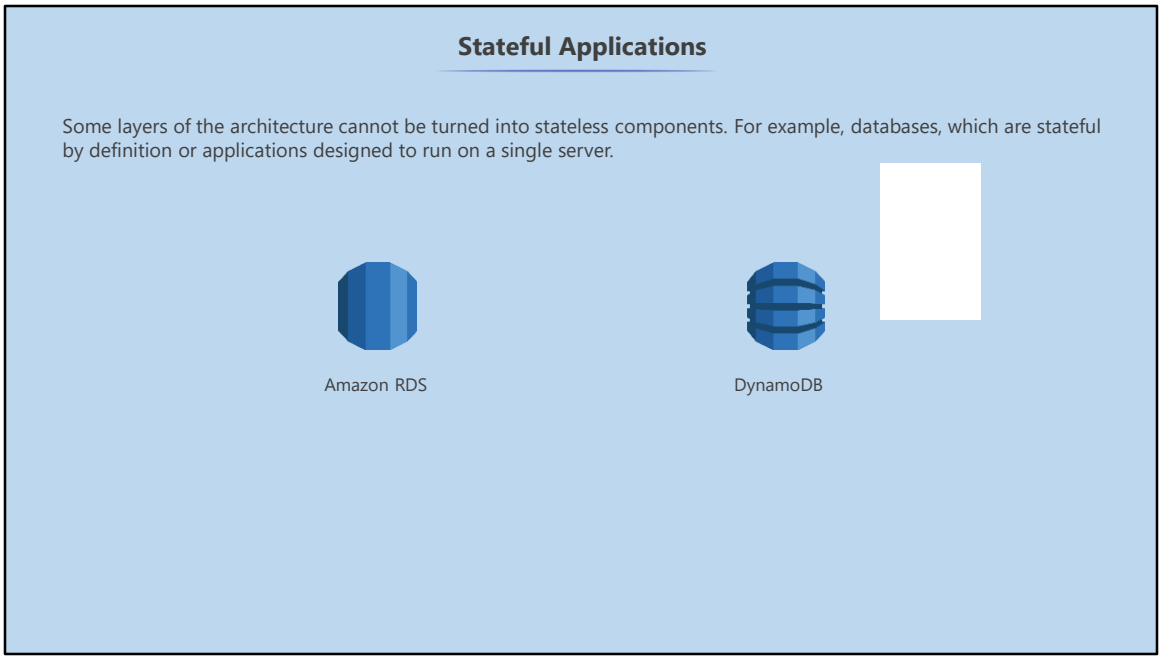
Examples of this are big data processing scenarios where many servers work to analyze data and return result. Or, media file conversion processes where multiple servers convert the files as they arrive.

Amazon SQS or Amazon Kinesis are services that can provide a pull model load balancing.



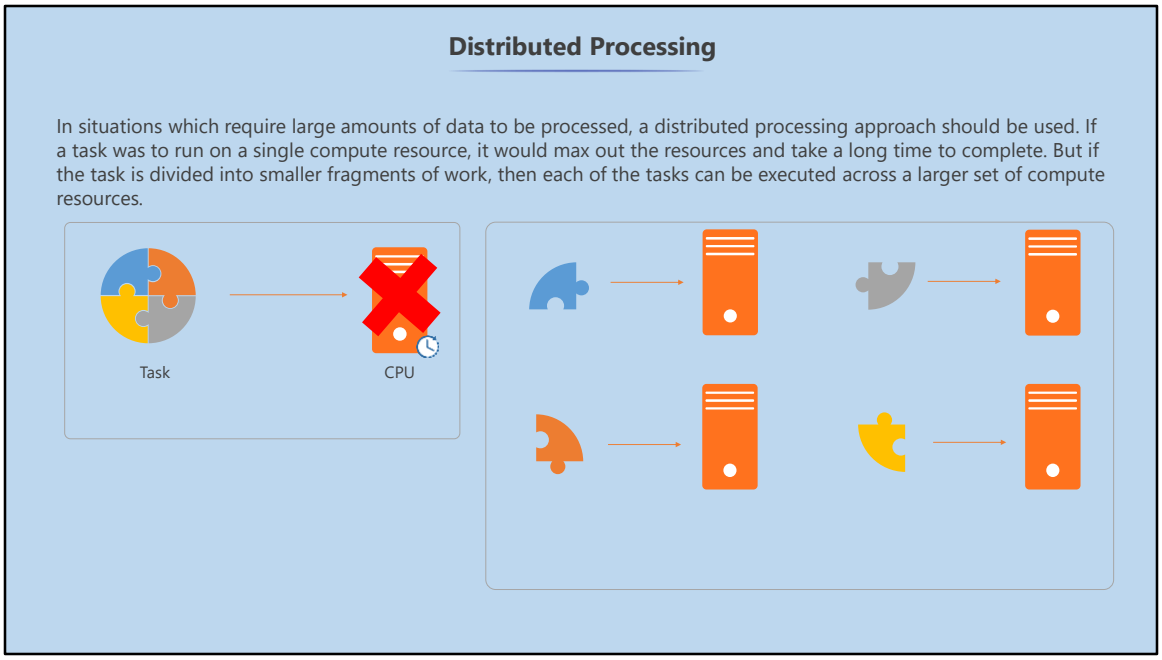
Most applications need to maintain some kind of state information. For example, an automated multi-step process should track previous activity to decide what its next action should be.

Components in your architecture can be made stateless by not storing anything on the local file system and instead storing user or session based information in a database (DynamoDB or MySQL), or on a shared storage layer (Amazon S3 or EFS). Additionally, Amazon SWF can be used to centrally store execution history and make the workloads stateless.



Some layers of your architecture cannot be turned into stateless components. For example, databases, which are stateful by definition or applications designed to run on a single server.

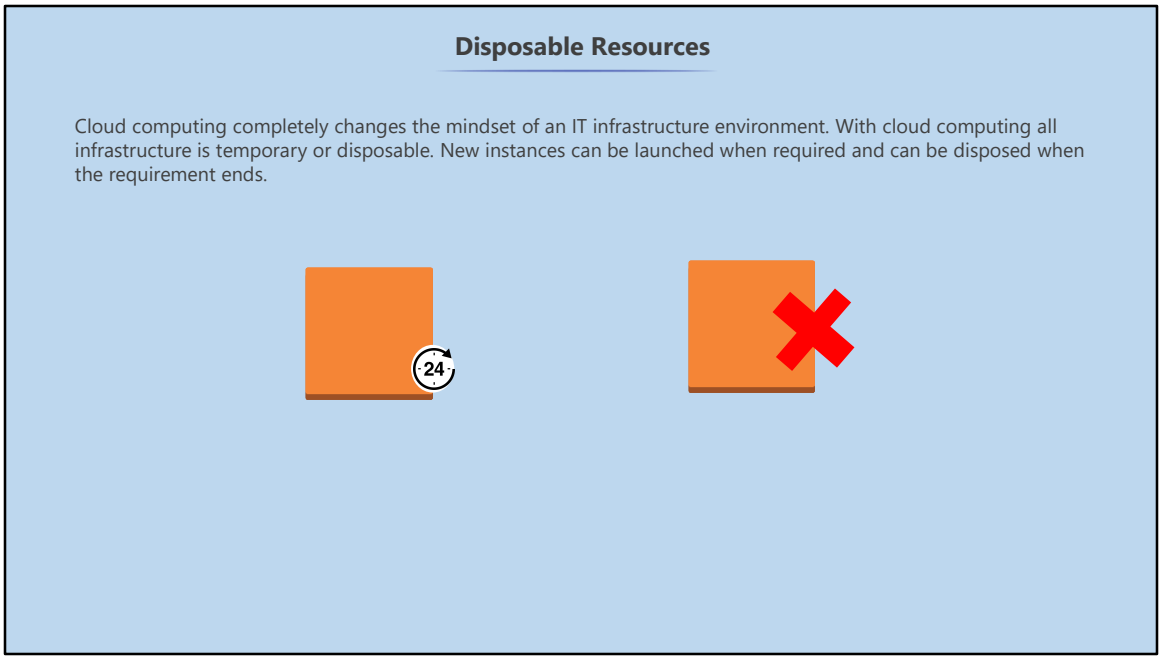
Providing multiple users a consistent view of the database or application is much simpler when components are not distributed.
Distributing the load is possible through the use of “session affinity”. This means all transactions of a session are bound to a specific compute resource; however, this means that existing sessions would not utilize newly introduced compute nodes. And if a node is shutdown or becomes unavailable, users bound to that session will be disconnected.



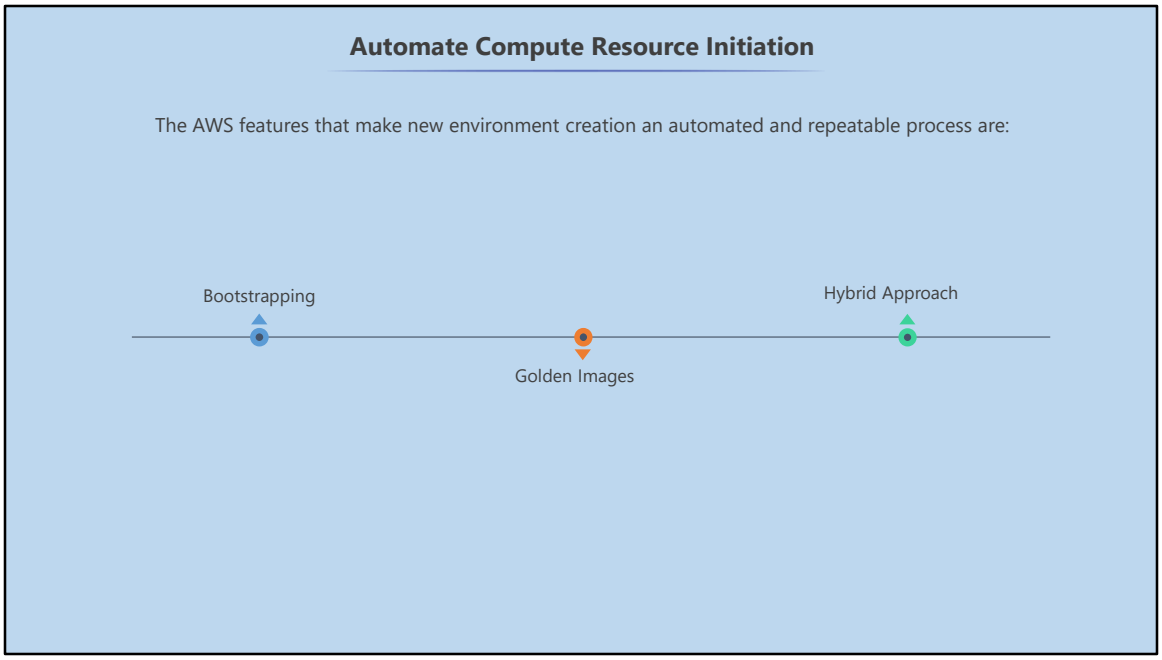
In situations which require large amounts of data to be processed, it is beneficial to see if a distributed processing approach can be used. Imagine a task that requires a huge amount of data processing. If it were to run on a single compute resource, it would max out the resources and would take a long time to complete. But if you can divide the task into smaller fragments of work, then each of the tasks can be executed across a larger set of compute resources.

For example, AWS Elastic MapReduce service allows Hadoop workloads to run across multiple EC2 instances.

Amazon Kinesis allows multiple shards of data to run on EC2 or Lambda resources for real-time processing of streaming data.

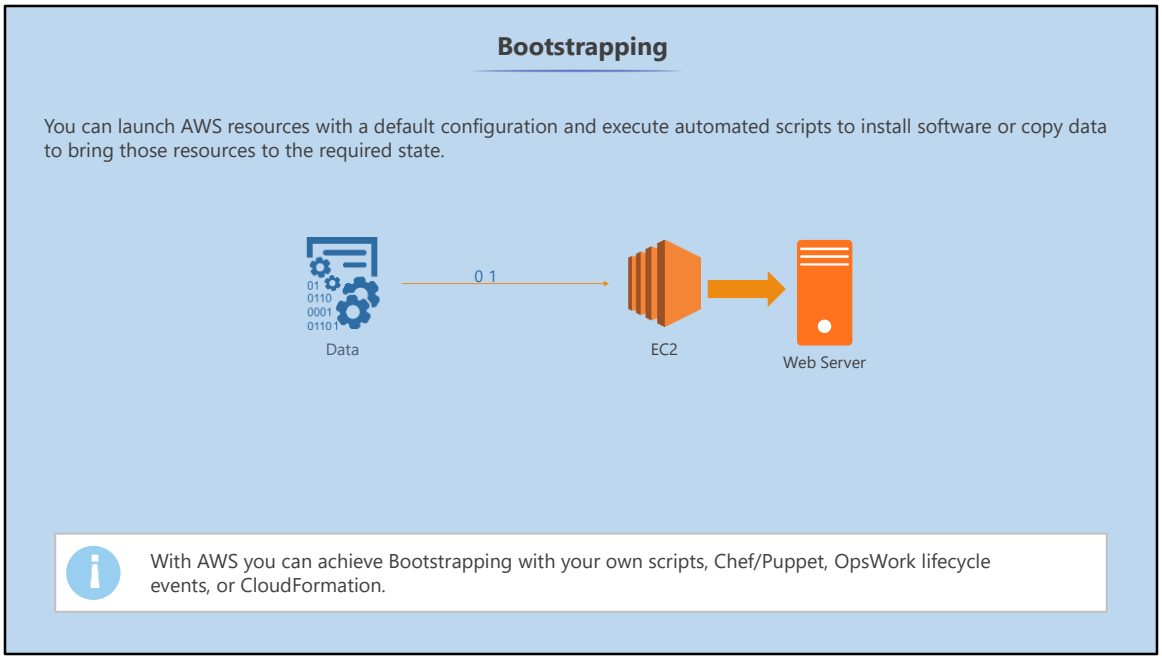


Cloud computing completely changes the mindset of an IT infrastructure environment. Traditional environments involve purchasing hardware upfront and require manual configuration of software, network, and so on. With cloud computing all infrastructure is temporary or disposable. You can launch new instances when you need them and can dispose them when the requirement ends.



The AWS features that make new environment creation an automated and repeatable process are:

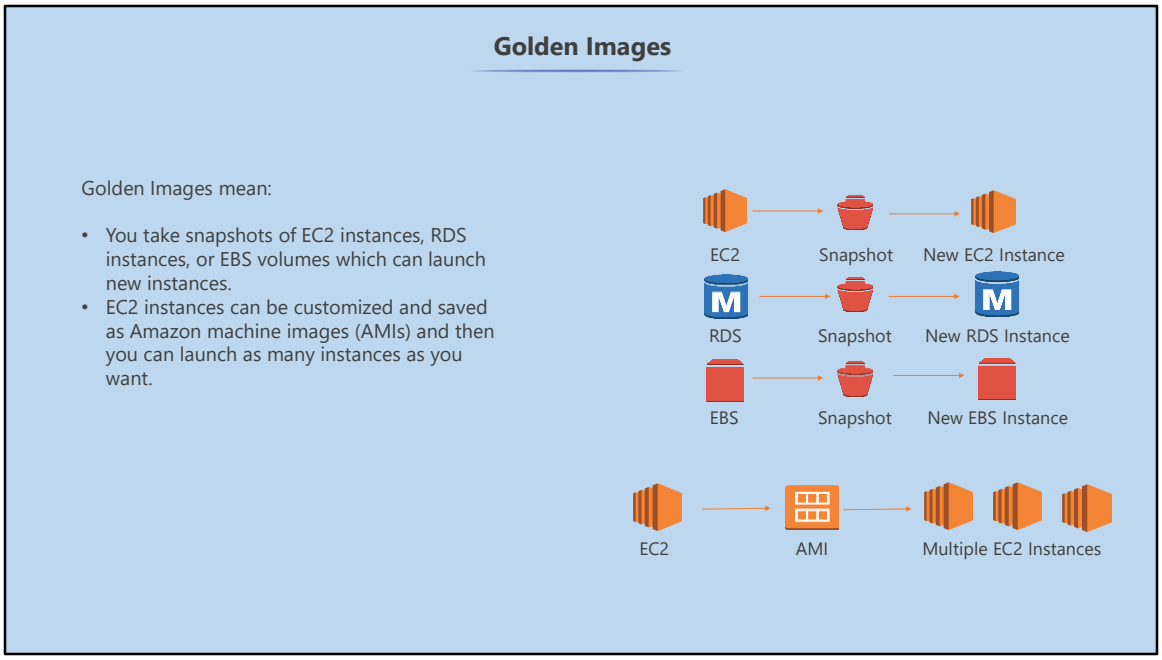
- Bootstrapping
- Golden Images
- Hybrid Approach



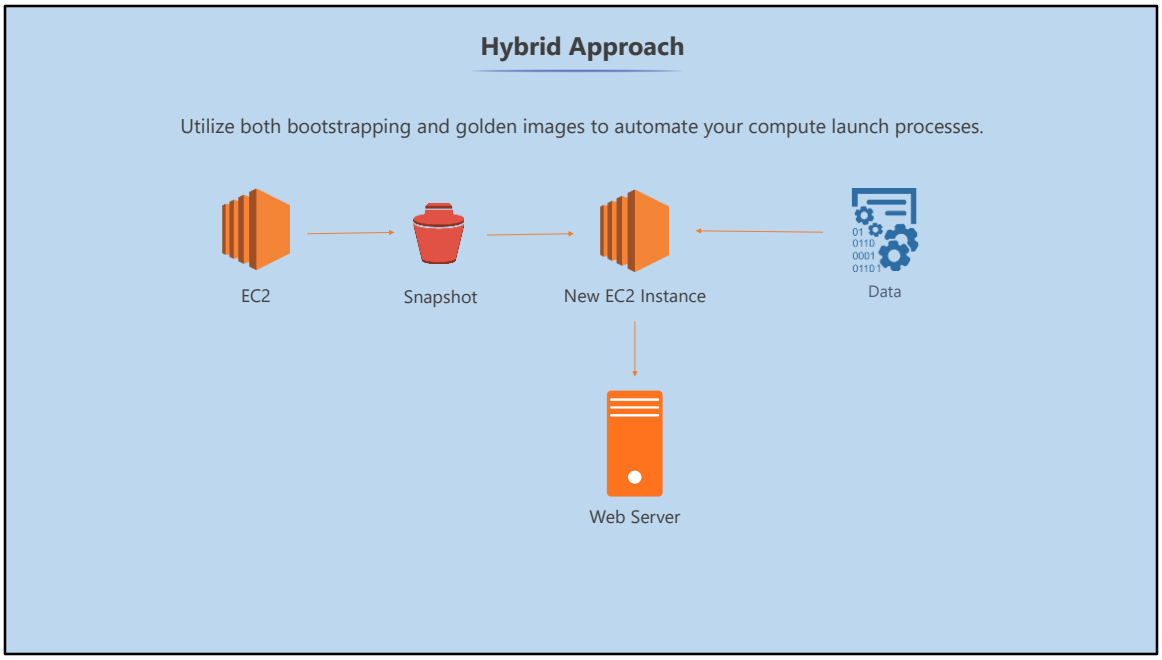
Newly launched AWS resources start with a default configuration, on which you can execute automated scripts to install software or copy data to bring that resources to the required state.

Scripts can be parameterized so that different environments (for example, production or development) can be initiated easily.

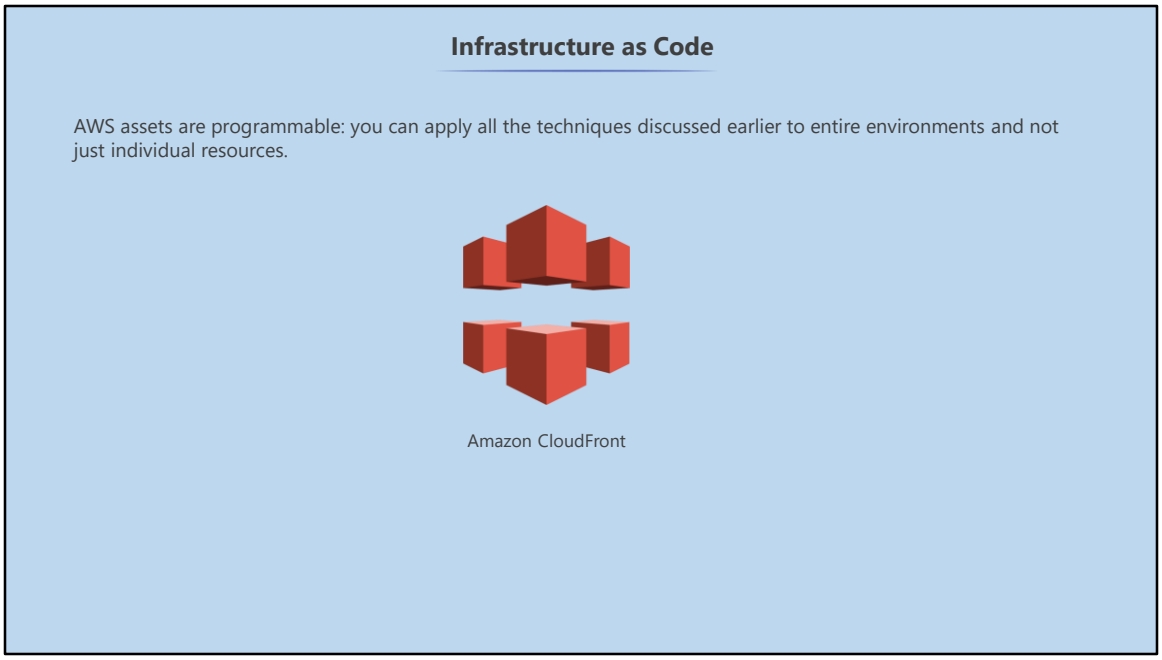
With AWS this can be achieved with your own scripts such as: Chef/Puppet, OpsWork Lifecycle Events, or CloudFormation.



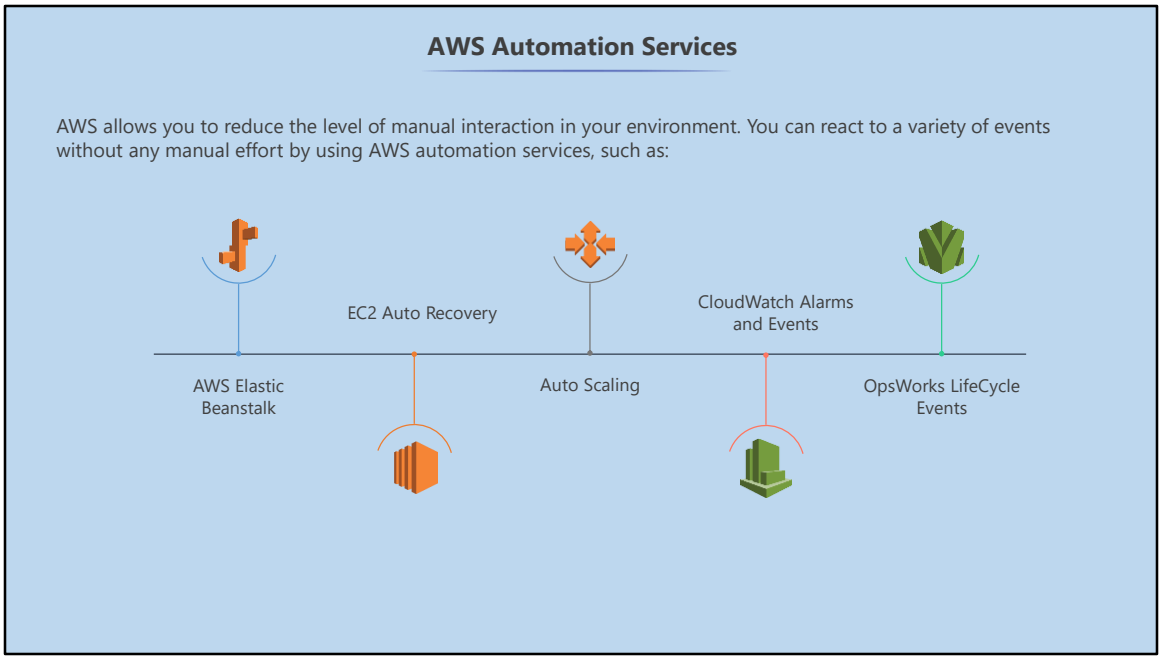
You can take snapshots of EC2 instances, RDS instances, or EBS volumes which can launch new instances.
EC2 instances can be customized and saved as AMIs and then you can launch as many instances as you want.
If you own an on-premise virtualized environment, you can use AWS VM Import/Export to create your own AWS AMIs.



Use both approaches to launch new instances; form a golden image and then bootstrap to configure other actions.

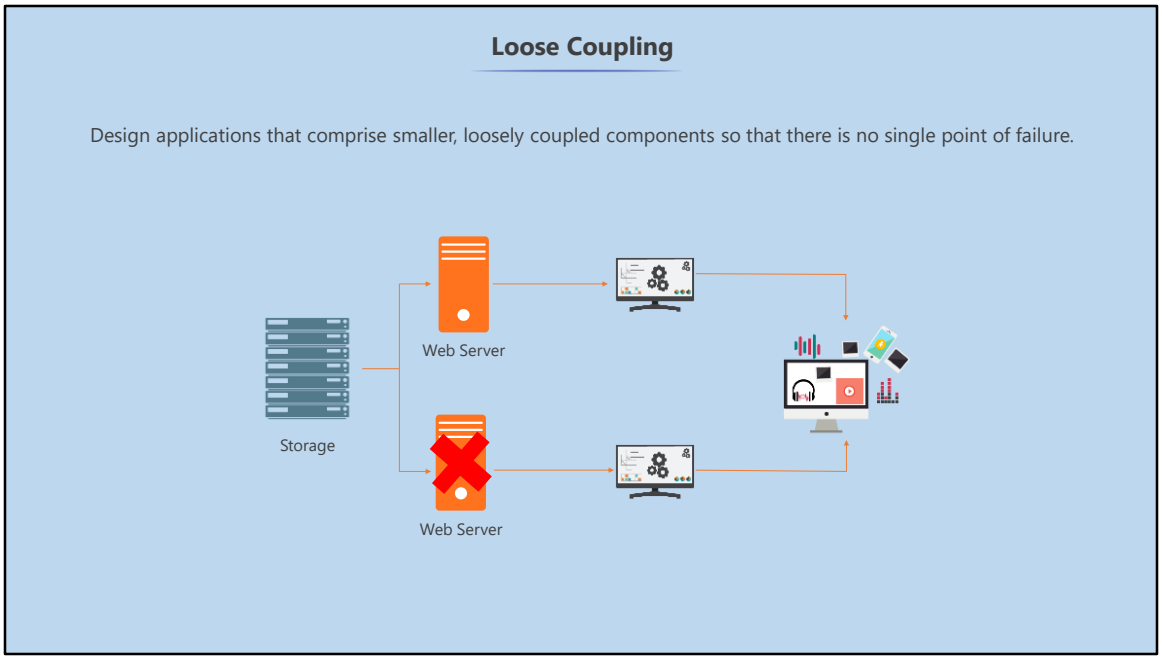


AWS assets are programmable—these techniques and principles can be applied to entire environments and not just individual resources. This concept is called Infrastructure as Code.

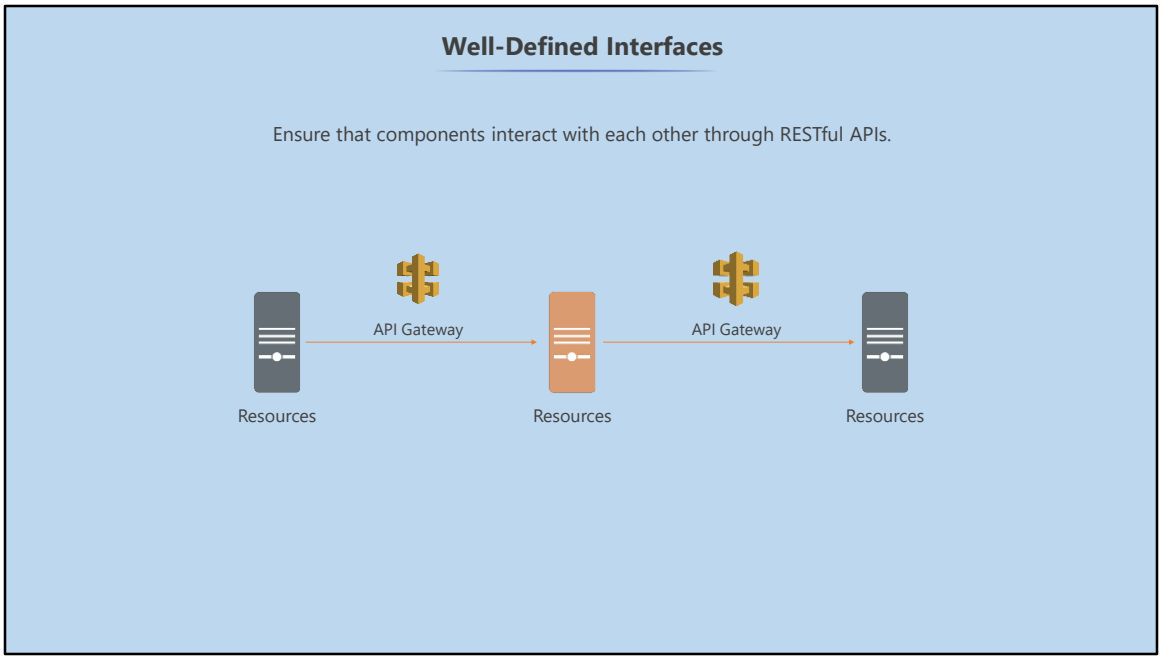


AWS allows reducing the level of manual interaction in your environment. Using automation you can react to a variety of events without any manual task.

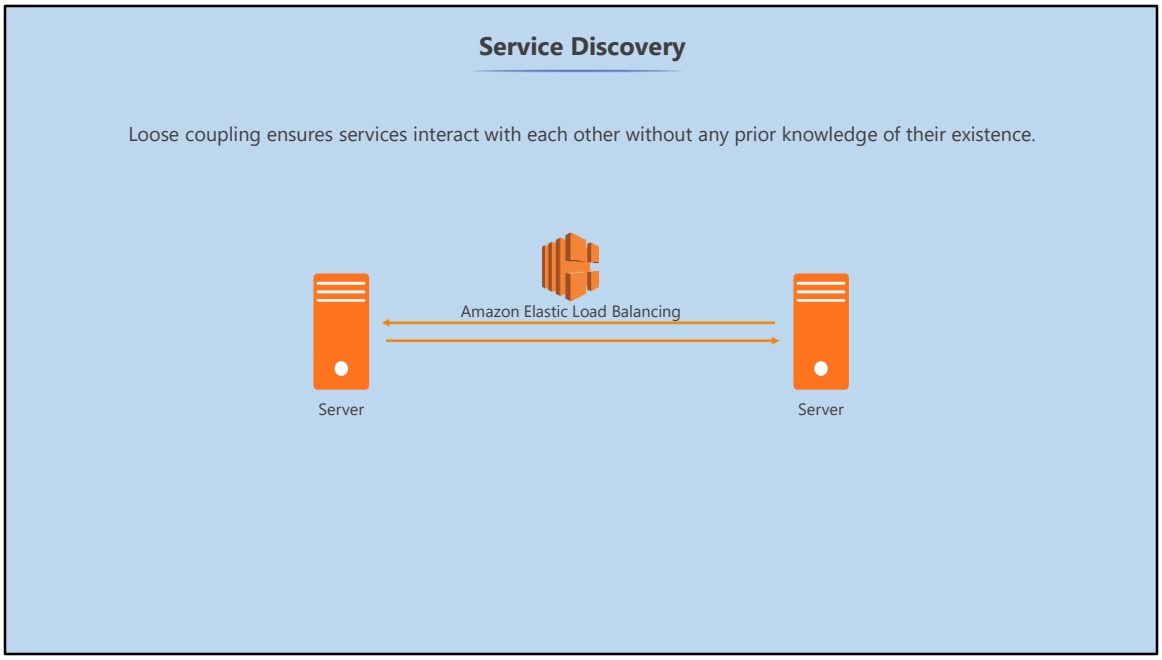
AWS Elastic Beanstalk–Developers can upload their code and let the Beanstalk service handle resource provisioning, auto scaling, load balancing, and monitoring.
EC2 Auto Recovery–In some situations use Auto Recovery to automatically recover an EC2 instance if it becomes impaired. The recovered instance is identical to the original (ID, IP, EIP, metadata).
Auto Scaling–Automatically adds resources to cope with demand spikes and remove during quiet times.
CloudWatch Alarms and Events–Cloudwatch can send SNS notifications to alert when a particular event occurs. The notifications can be used to perform follow up actions, like running a Lambda function.
OpsWorks Lifecycle Events–Continually update your instances' configuration to adapt to environment changes. For example, if a new database instance is added to your environment then an event can automatically trigger a Chef recipe that points existing applications to the new server.



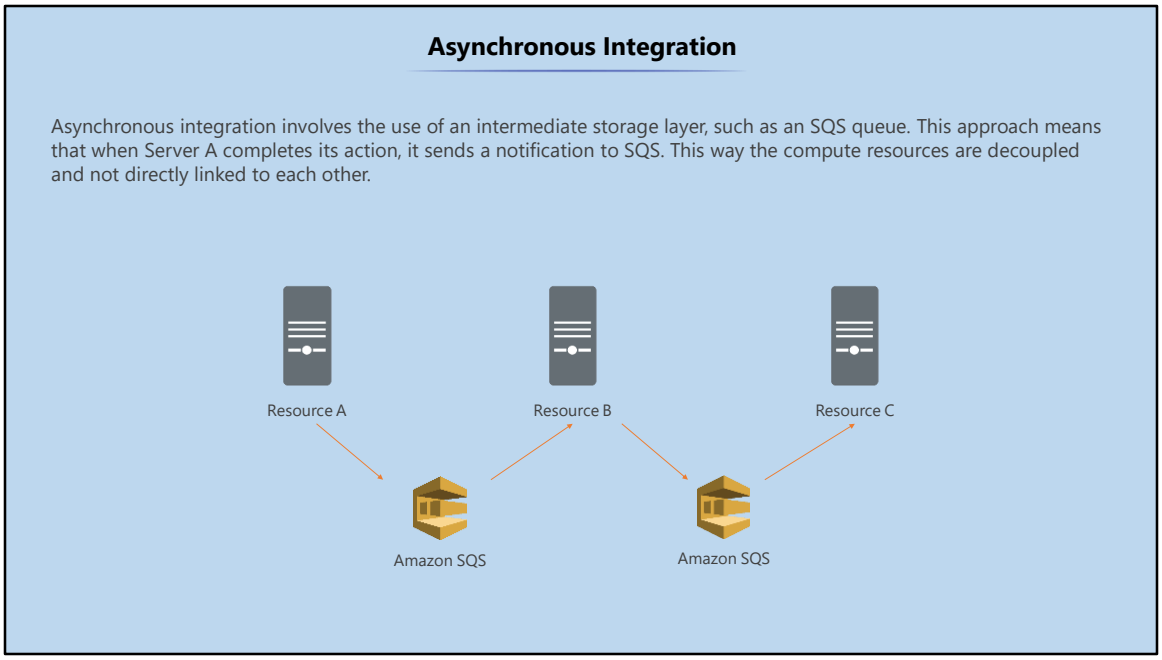
Applications should be designed so that they are broken into smaller, loosely coupled components. The desired outcome is that a failure in one component should not cause other components to fail.



Ensure that all components interact with each other through specific, technology-agnostic interfaces. For example RESTful APIs, will result in the ability to modify resources without affecting other components. Amazon API Gateway is a fully managed service for APIs.

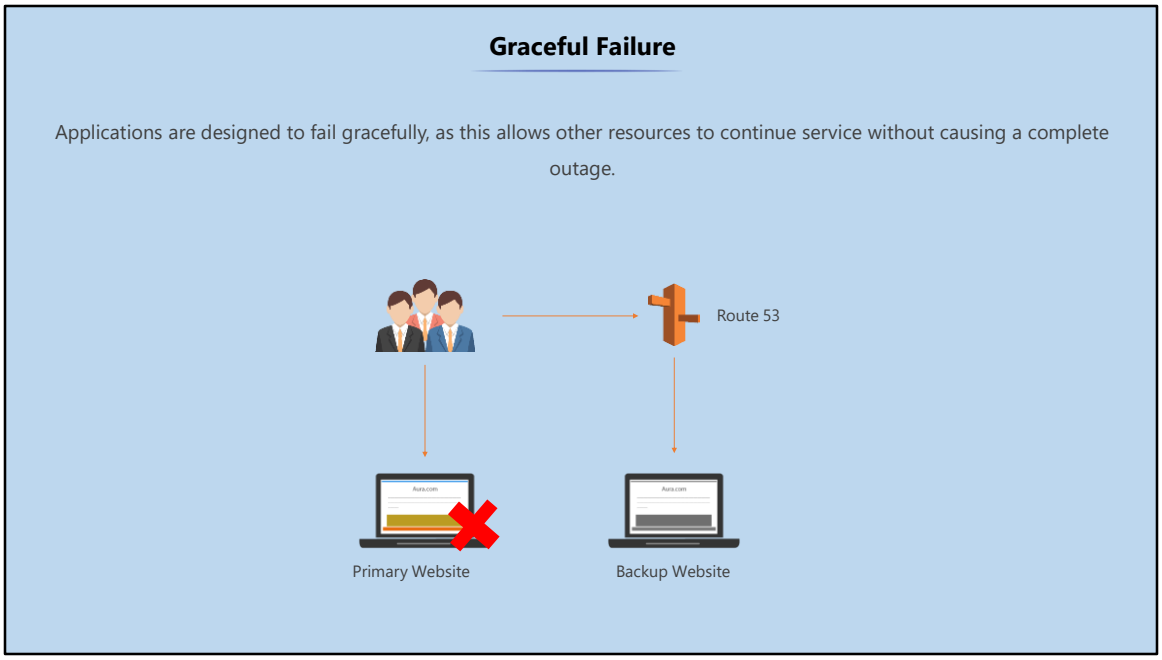


Loose Coupling ensures services to interact with each other without any prior knowledge of their existence. For example, if two servers communicate on hardcoded IP addresses, then the addition or modification of resources will result in manual interaction to update the configuration.
Loosely coupled infrastructure is an essential ingredient in making the most of cloud computing elasticity. The easiest way to achieve this is to use the Elastic Load Balancing service.

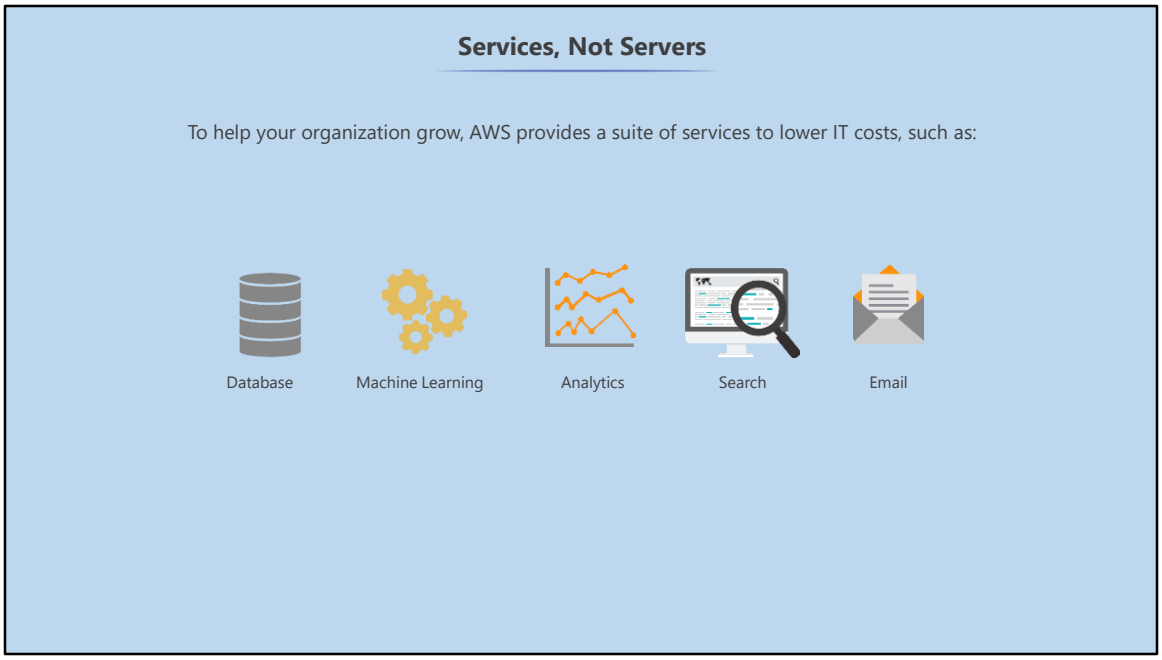


Instead of synchronous integration, where Server A completes an action and passes it to Server B, which then passes it to Server C, Asynchronous integration involves the use of an intermediate storage layer, such as an SQS queue.

This approach means that when Server A completes its action, it sends a notification to SQS. This way the compute resources are decoupled and not directly linked to each other. Resources can easily be added or removed without the need to update configuration of the existing compute resources.



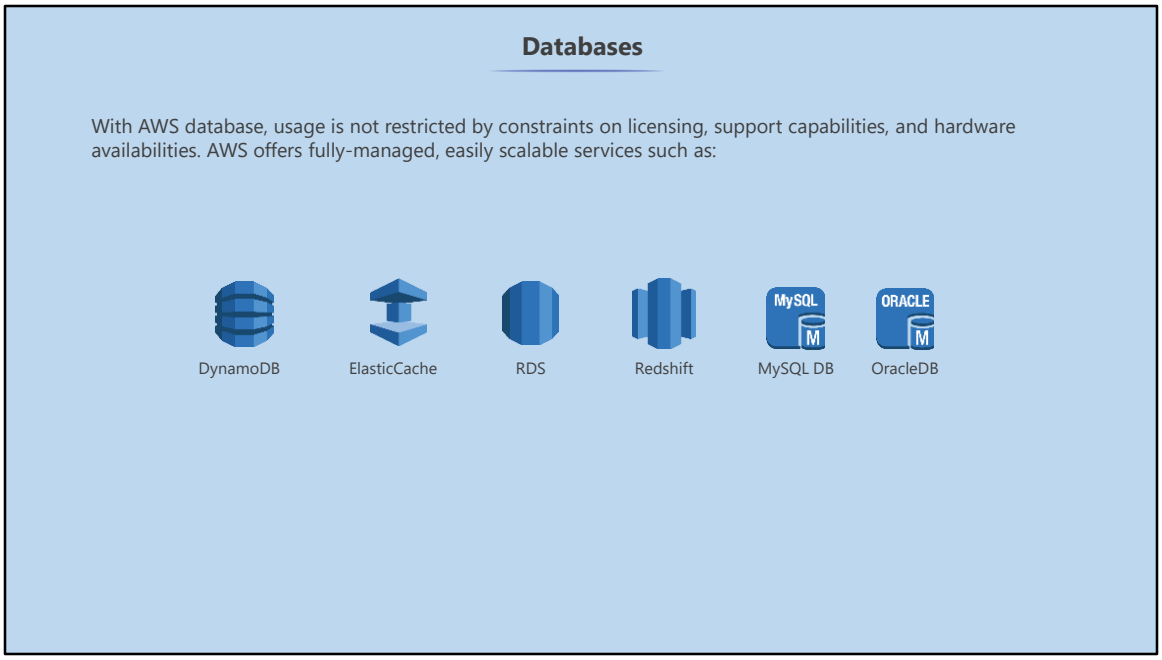
Designing applications to fail gracefully allows other resources to continue the service without causing a complete outage. For example, if your primary website fails, the Route 53 DNS failover feature points your users to a backup site hosted as a static website on Amazon S3. The idea is to continue to provide a service, even if it is not the full experience.



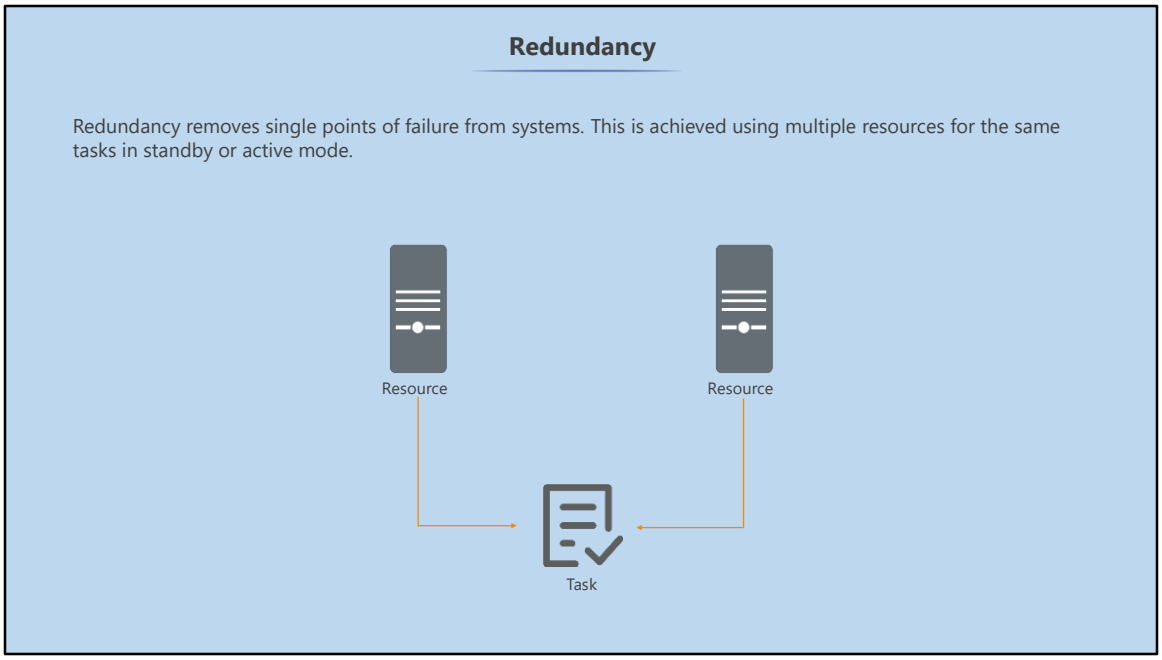
AWS provides multiple services that are fully managed: databases, machine learning, analytics, search, and email to lessen the burden on your IT infrastructure department.

With RDS databases you don't need to worry about backups and fault tolerance. S3 allows you to store as much data as you need without concerns about capacity management.

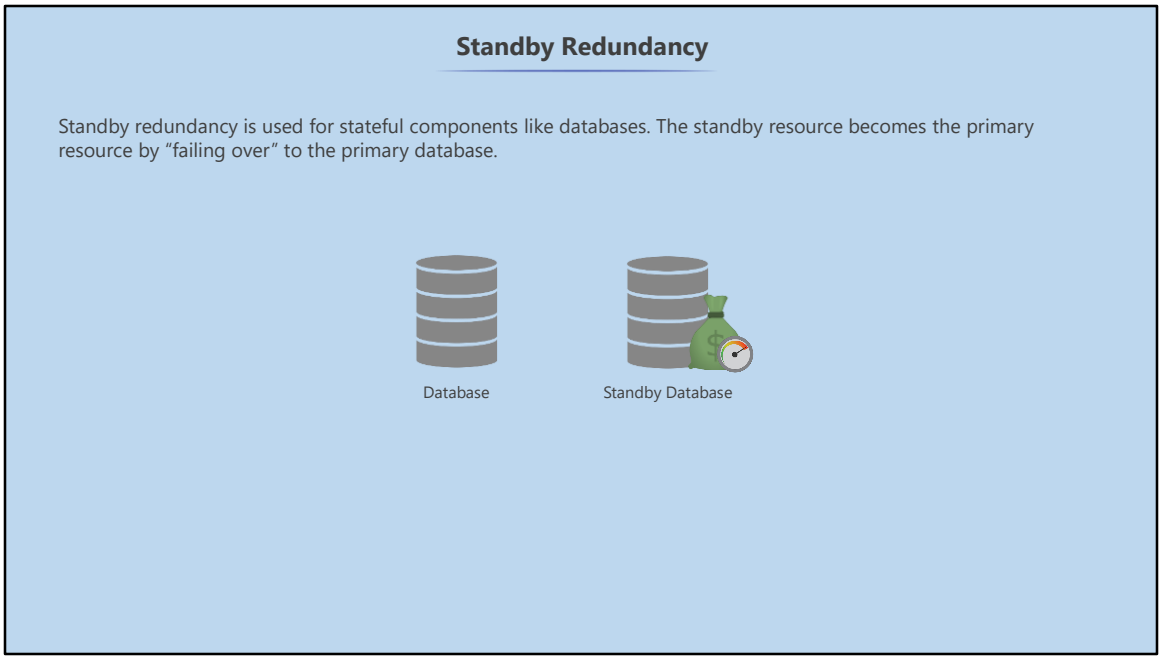
Using Serverless Architectures reduce operational complexity of applications. Tools such as IoT, Lambda, S3, and Beanstalk allow applications to run without paying for any server infrastructure.



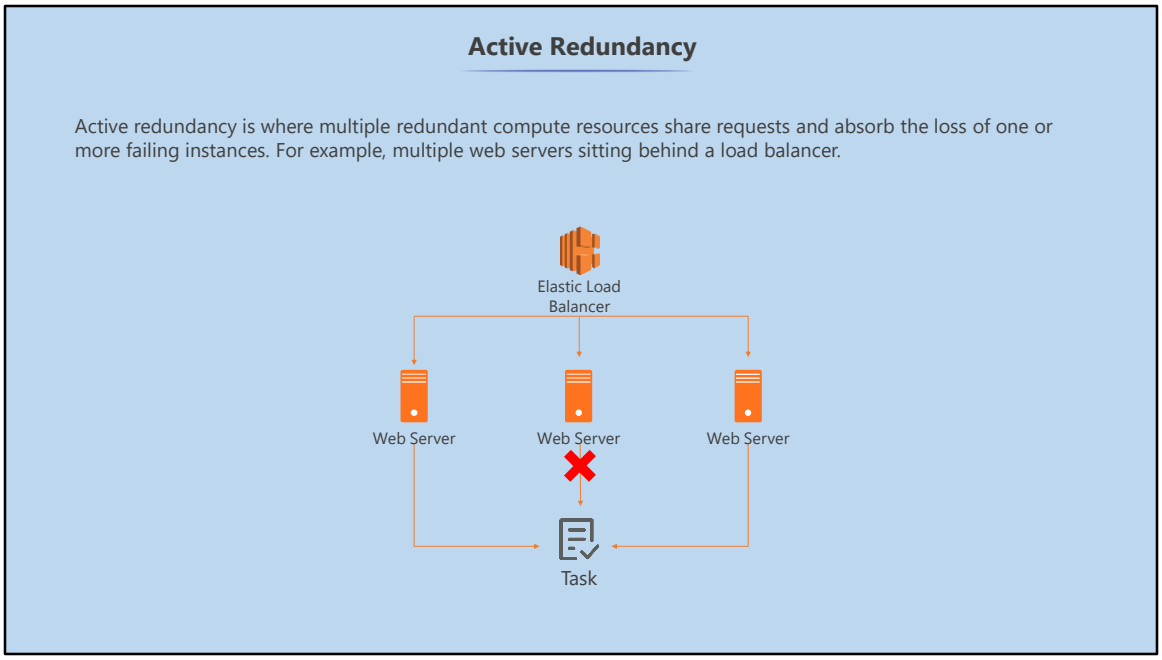
In traditional IT environments, the type of database used is restricted by constraints on licensing, support capabilities, and hardware availability.
AWS provides multiple managed database server options that remove these constraints, which means you can choose the right database for your needs rather than making use of what is available to you.
AWS offers fully managed, easily scalable services for Relational Databases, NoSQL, Data Warehouse, and search functions that provide fully managed, easily scalable services. This create synchronously replicated standby instance in a different Availability Zone, and automatically failover without the need for manual intervention.



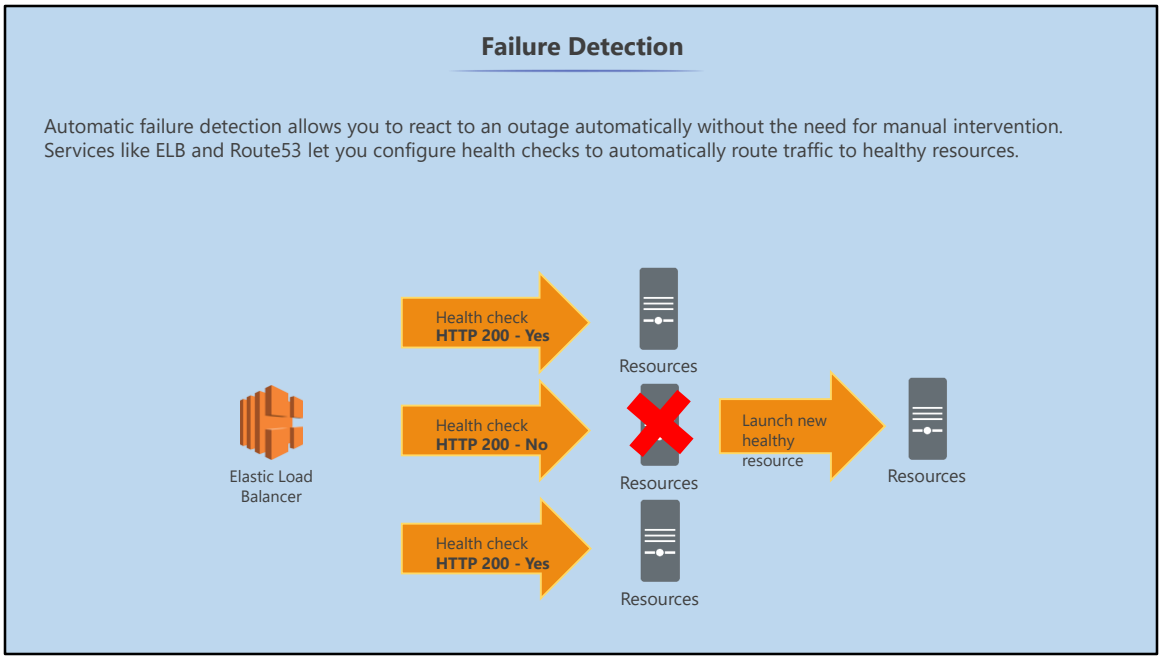
A system is considered highly available if it withstands the failure of individual or multiple components (for example, hard disks, servers, and network links). Introducing redundancy removes single points of failure from systems. This is achieved using multiple resources for the same tasks in standby or active mode.



Standby redundancy is used for stateful components like databases. The standby resource becomes the primary resource by "failing over" to the primary database. The Standby resource is brought online when required to save cost or it can be run to speed up the failover process and minimize the outage.

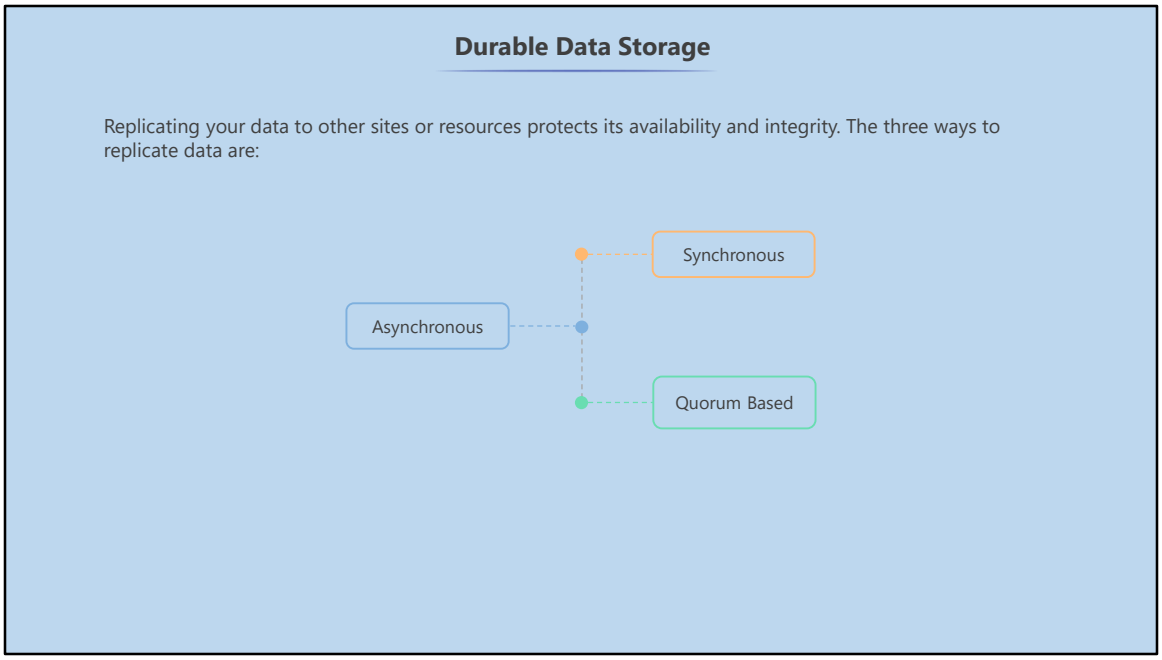


Active redundancy is where multiple redundant compute resources share requests and absorb the loss of one or more failing instances. For example, multiple web servers sitting behind a load balancer.



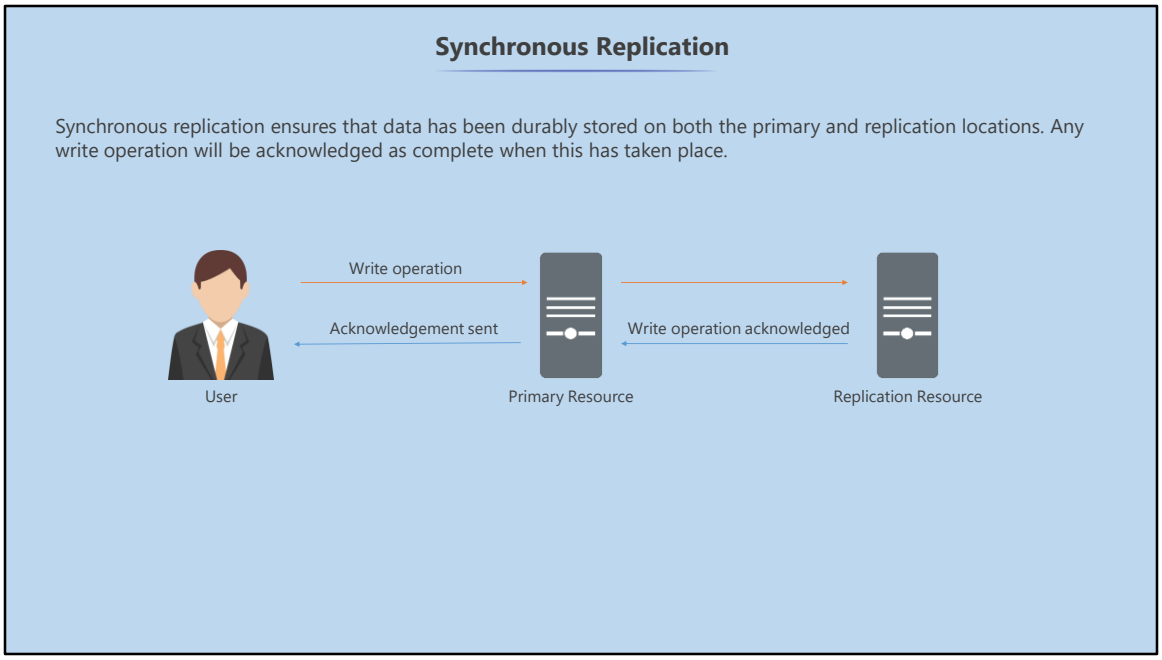
Automatic failure detection allows you to react to an outage automatically without the need for manual intervention.

Services like ELB and Route53 let you configure health checks to automatically route traffic to healthy resources. For example, setting up a health check to test for HTTP 200 response for a simple request would allow AWS to know if a node is healthy or not. Services such as OpsWorks, ELB, and EC2 Auto-recovery let you replace unhealthy resources.

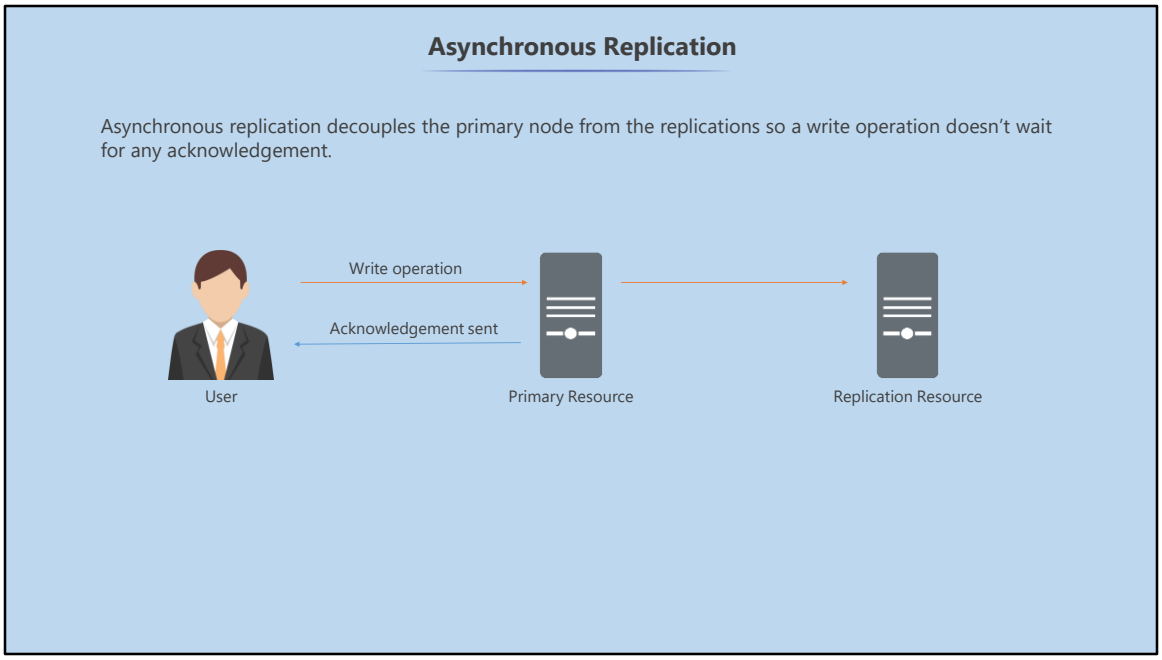


Replicating data to other sites or resources protect its availability and integrity.
There are three ways to replicate your data:

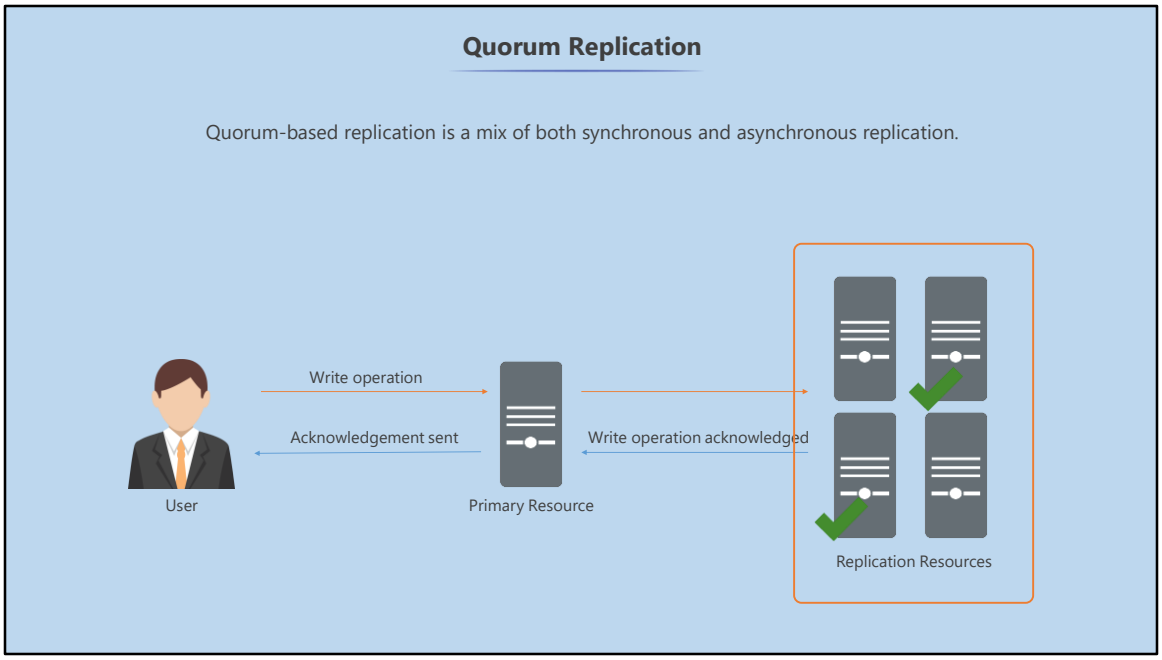
1. Synchronous
2. Asynchronous
3. Quorum based



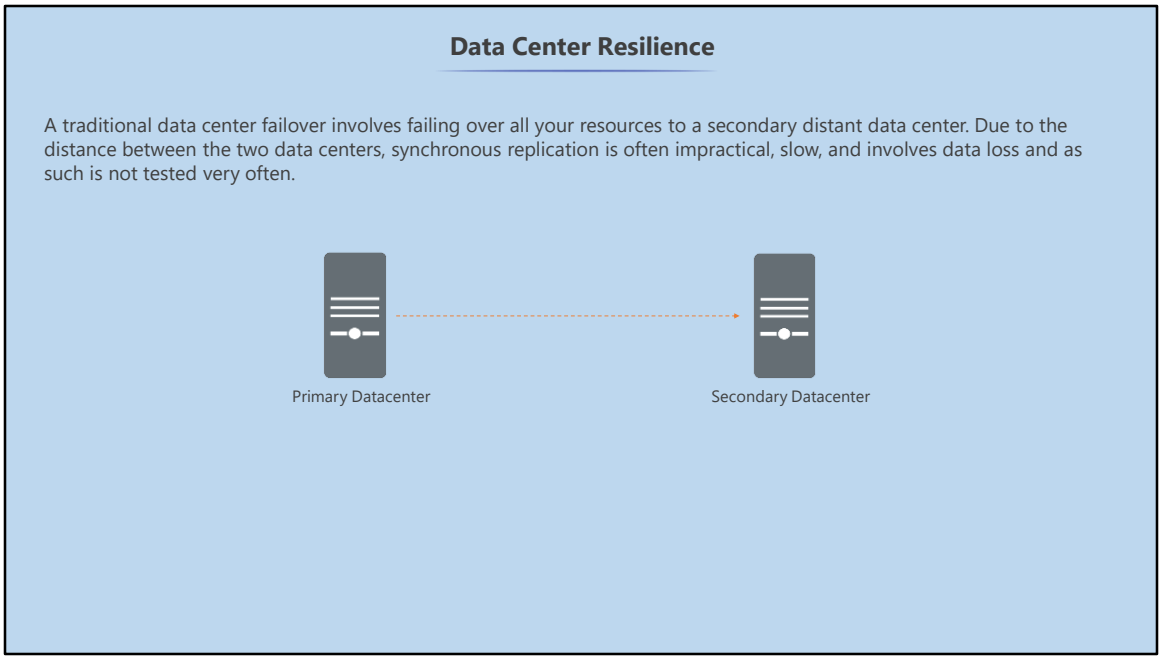
Synchronous replication ensures that data has been durably stored on both the primary and replication locations. Any write operation will be acknowledged as complete when this has taken place. This should be used on low latency network connections when absolutely necessary to ensure maximum performance.



Asynchronous replication decouples the primary node from the replications so a write operation doesn't wait for any acknowledgement. This can introduce a replication lag, but does not impact performance.

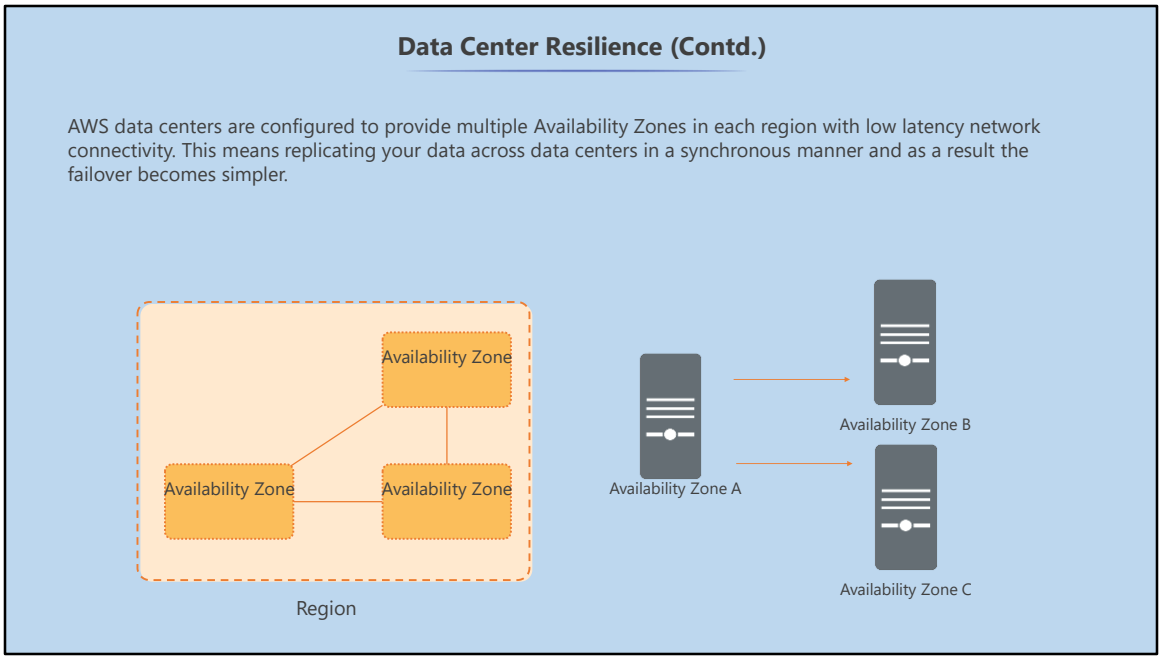


Quorum-based replication is a mix of both synchronous and asynchronous replication. Replication to multiple nodes is controlled by defining the minimum number of nodes that must participate in a successful write operation.

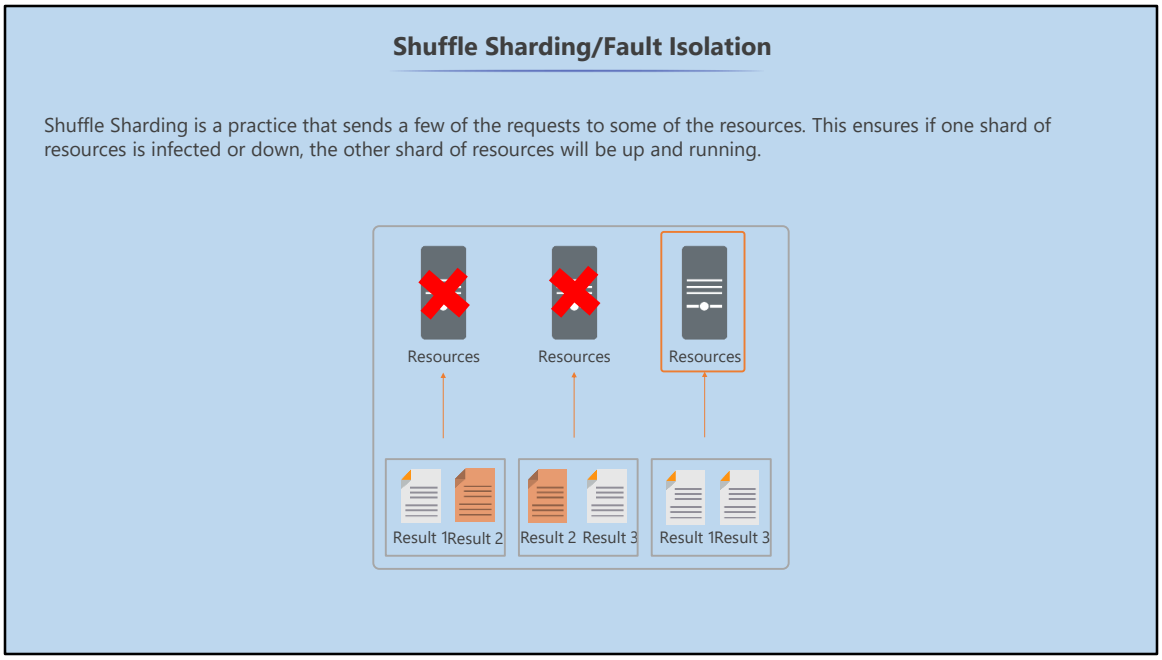


A traditional data center failover scenario involves failing over all your resources to a secondary distant data center. Due to the distance between the two data centers, synchronous replication is often impractical, slow, and involves data loss and as such is not tested very often. However, it does protect against low probability scenarios, such as a natural catastrophe.

A more likely scenario is a shorter interruption in a data center, which is not predicted to be too long. In this situation your choice is to sit it out and wait or initiate the complex failover procedure.

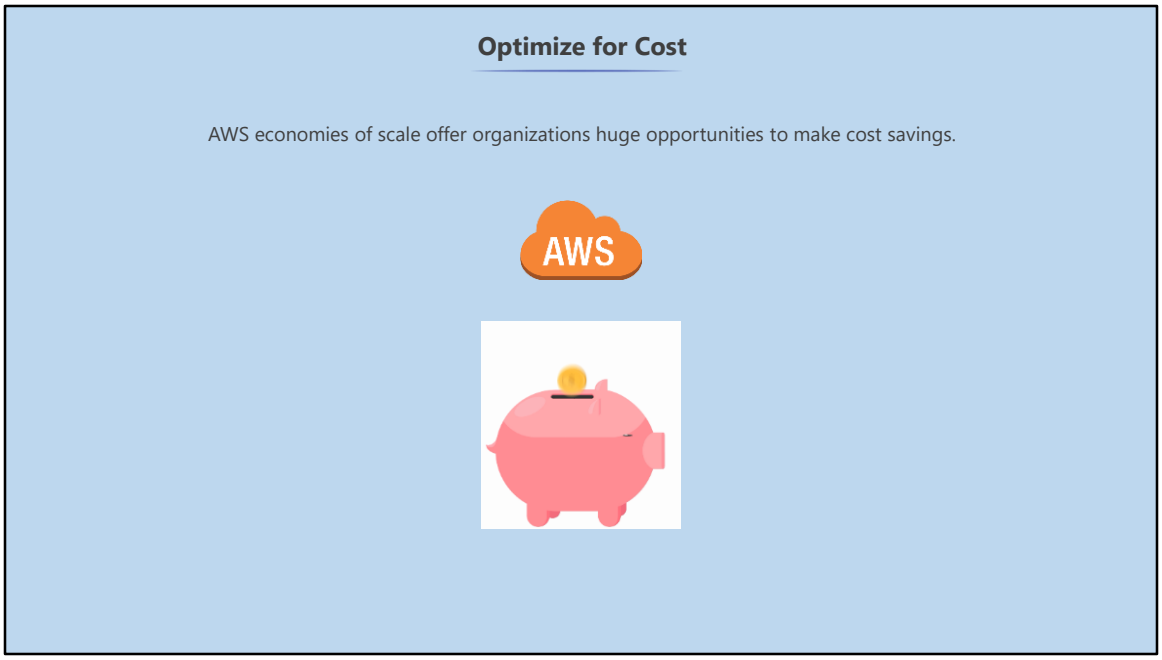


AWS data centers are configured to provide multiple Availability Zones in each region with low latency network connectivity. This means replicating your data across data centers in a synchronous manner and as a result the failover becomes simpler. This is provided by many AWS services such as RDS and S3.



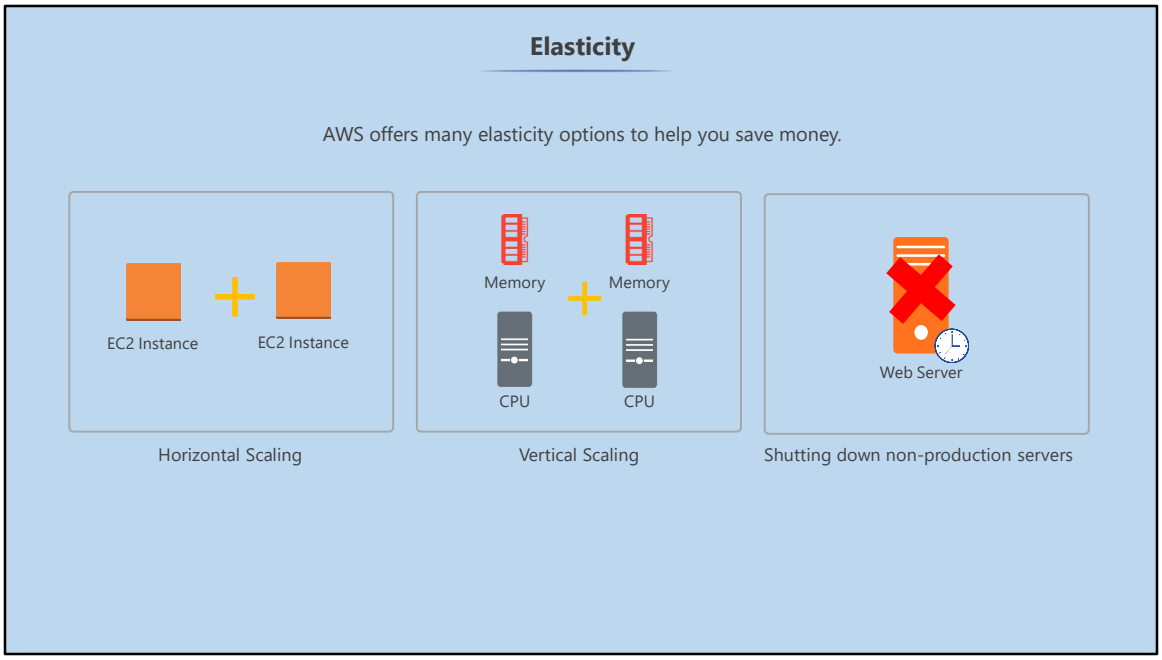
Active redundancy is great for disaster recovery situations or balancing traffic. But what happens if all the requests to your resources were harmful, that is, the request caused a bug resulting in multiple instance crashes?

To avoid this a practice called Shuffle Sharding is used. Shuffle Sharding sends a few of the requests to some of the resources. This ensures if one shard of resources is infected or down, the other shard of resources will be up and running.

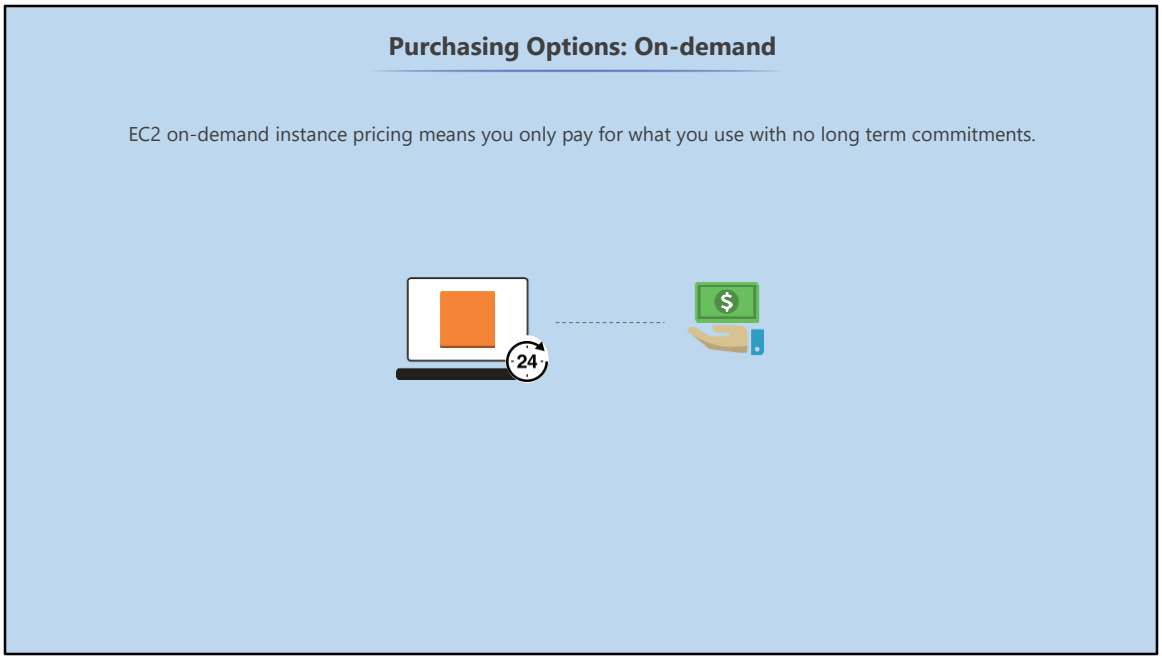


AWS economies of scale offers organizations cost saving opportunities. AWS capabilities provide scope to create cost-optimized cloud architectures.

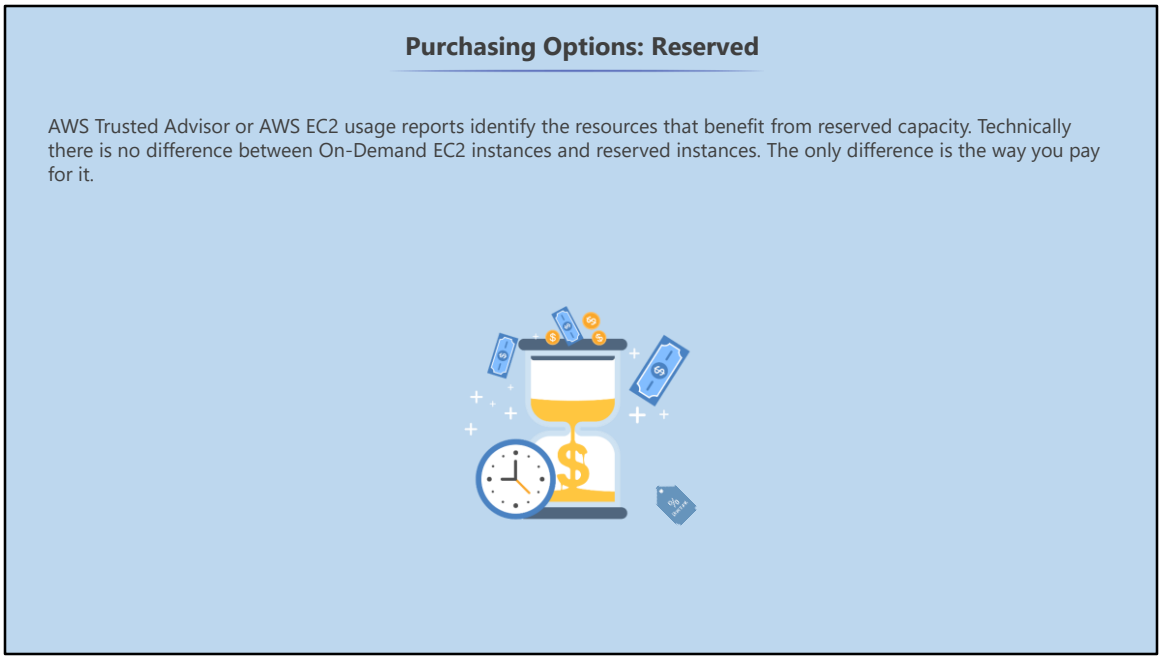
In contrast to traditional IT infrastructure, cloud computing allows you to select the most cost effective resource and configuration to fit your requirements. EC2, RDS, Redshift, and Elastic Search provide a wide variety of instance types to choose from. Benchmarking helps determine the optimal configuration for your workload. In some cases, it might be optimal to select many of the cheapest instance type available, in others it might be beneficial to select fewer instances of a larger instance type.
AWS storage offers the same level of right sizing opportunities. Amazon S3 offers different storage classes to suit your needs and EBS comes in a variety of volume types.



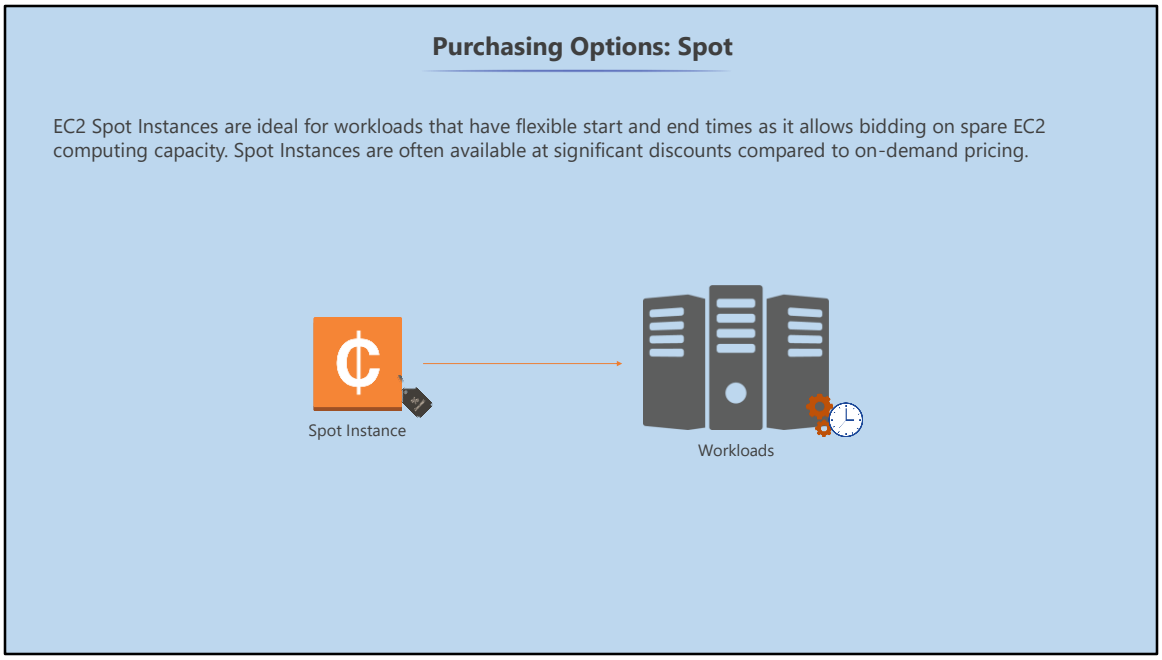
AWS offers many elasticity options that save money. You can auto scale EC2 workloads horizontally, up and down as required.
Turn off non-production workloads when not required or use Lambda compute workloads so you don't pay for idle resources.



EC2 on-demand instance pricing means you only pay for what you use with no long term commitments; however, there are more ways to pay for EC2 instances.



Committing to a defined period, between 12-36 months, gives discounted hourly rates compared to the on-demand pricing.
AWS Trusted Advisor or AWS EC2 usage reports identify the resources that benefit from reserved capacity. There is technically no difference between On-Demand EC2 instances and reserved instances, the only difference is the way you pay for it.
This concept also exists for Redshift, RDS, DynamoDB, and CloudFront.

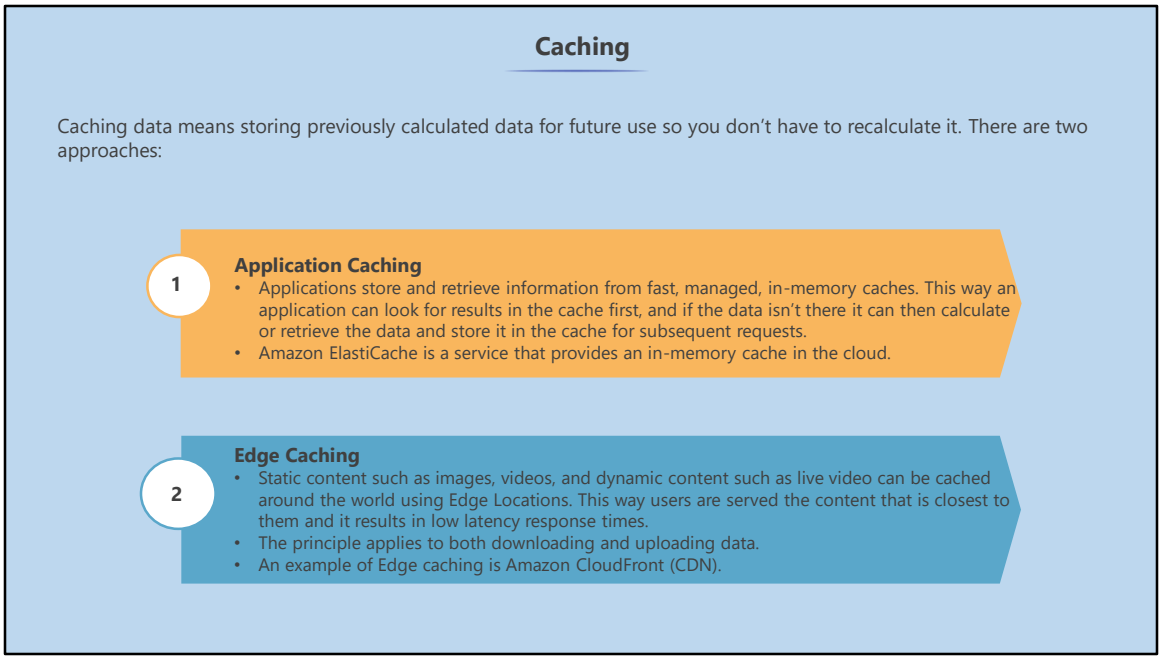


EC2 Spot Instances are ideal for workloads that have flexible start and end times as it allows bidding on spare EC2 computing capacity. Spot Instances are often available at significant discounts compared to on-demand pricing.

Your EC2 spot instance is launched when your bid exceeds the current spot market price and will continue to run until you terminate it, or the spot market price exceeds your bid. When the latter happens, your instance is terminated and you will not be charged for the partial hour that it was running.

There are three strategies ~~to use~~ for Spot instances:

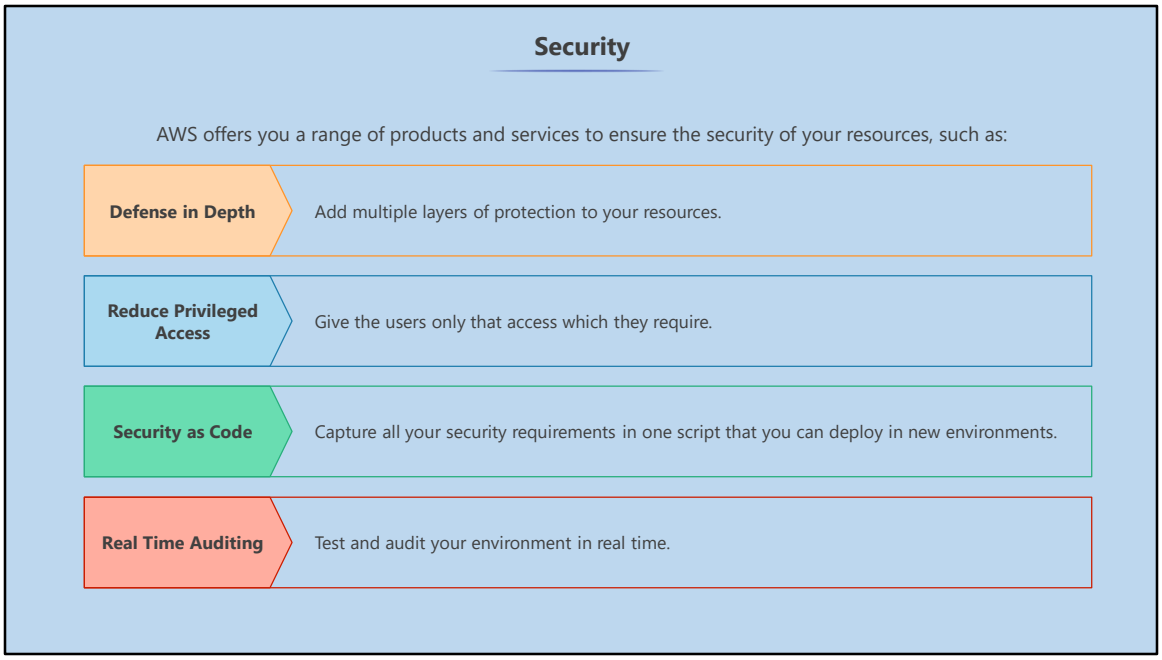
- Bidding: Spot instances are ideal for workloads that can tolerate interruption; however you could bid much higher than the spot market price for as long as the instance runs. Thus, if the market price spikes occasionally, you still make a saving. Spot prices fluctuate depending on Availability Zone and instance type. So design your apps to be flexible about instance types or Availability Zone and bid on all instance types that meet your requirements.
- Mix: AWS recommends designing applications that use a mixture of reserved, on-demand, and spot instances. This way you can combine predictable capacity with "opportunistic" access to cheaper additional compute resources.
- Spot Blocks: AWS allows you to bid for fixed duration spot instances. These have a different hourly pricing, but allow you to define a duration requirement. If your bid is accepted, your instance will run until you terminate it or the duration ends.



Caching data means storing previously calculated data for future use so you don't have to recalculate it. There are two approaches:

Application Caching
Design applications to store and retrieve information from fast, managed, in-memory caches. This way an application can look for results in the cache first, and if the data isn't there it can then calculate or retrieve the data and store it in the cache for subsequent requests. This approach improves user response time and reduces load on your resources.
Amazon ElastiCache is a service that provides an in-memory cache in the cloud. It is suitable for use with web services.

Edge Caching
Using Amazon CloudFront (CDN) you can cache static content such as images, videos, and dynamic content such as live video around the world using Edge Locations. This way users are served the content that is closest to them and it results in low latency response times. The principle applies to both downloading and uploading of data.



AWS offers you a range of products and services to ensure the security of your resources, such as:

- Defense in Depth: Add multiple layers of protection to your resources.
- Reduce Privileged Access: Give the users only that access which they require.
- Security as Code: Capture all your security requirements in one script that you can deploy in new environments.
- Real Time Auditing: Test and audit your environment in real time.

Knowledge Check

1 **A Stateless application is one that _____.**

requires knowledge of previous interactions but stores no session information

needs no knowledge of previous interactions and stores no session information

requires knowledge of previous interactions and stores session information

needs no knowledge of previous interactions but stores session information

1 **A Stateless application is one that _____.**

requires knowledge of previous interactions but stores no session information

needs no knowledge of previous interactions and stores no session information

requires knowledge of previous interactions and stores session information

needs no knowledge of previous interactions but stores session information

b

A stateless application is one that needs no knowledge of previous interactions and stores no session information. For example, a webserver that provides the same web page to any end user.

2 **Loose Coupling is desirable because _____.**

it reduces the cost of your AWS resources

it stores previously calculated data for future use

it means the failure of one or more resources does not result in a service outage

it assists you to select resource and configuration to fit your requirements

2 **Loose Coupling is desirable because _____.**

it reduces the cost of your AWS resources

it stores previously calculated data for future use

it means the failure of one or more resources does not result in a service outage

it assists you to select resource and configuration to fit your requirements

c

Applications should be designed so that they can be broken into smaller, loosely coupled components. The desired outcome is that a failure in one component should not cause other components to fail.

3 **The three EC2 purchasing options that make cloud computing unique are _____.**

On-Request, Auction, and Reserved pricing

On-Demand, Spot, and Permanent pricing

On-Request, Local, and Permanent pricing

On-Demand, Spot, and Reserved pricing

3 **The three EC2 purchasing options that make cloud computing unique are _____.**

On-Request, Auction, and Reserved pricing

On-Demand, Spot, and Permanent pricing

On-Request, Local, and Permanent pricing

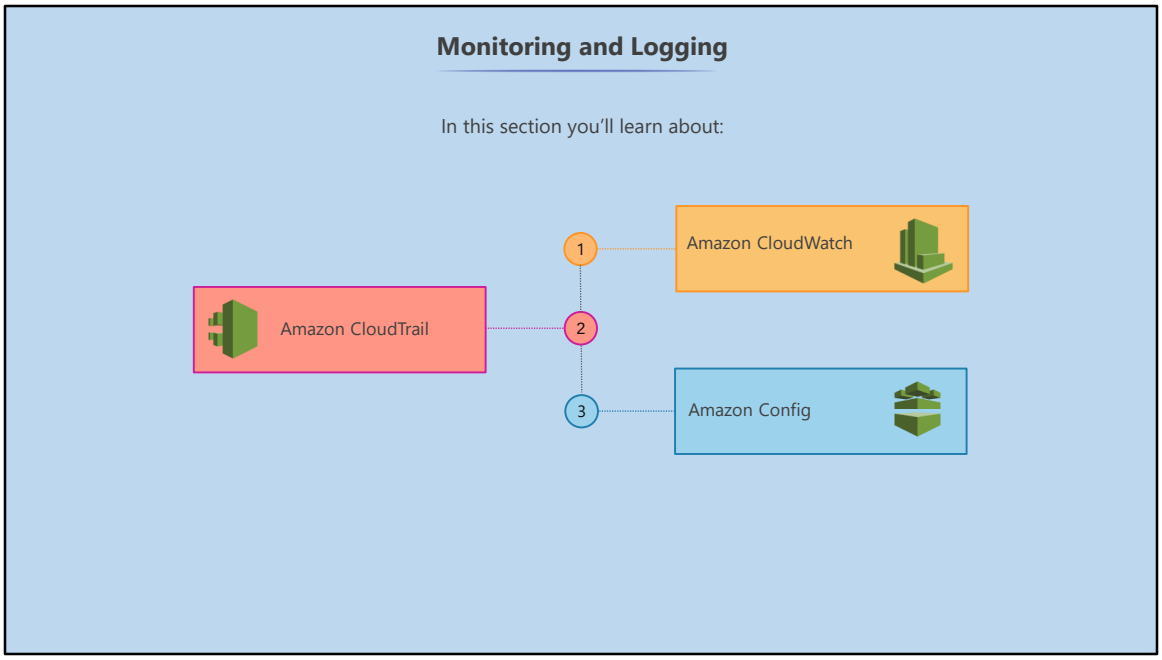
On-Demand, Spot, and Reserved pricing

d

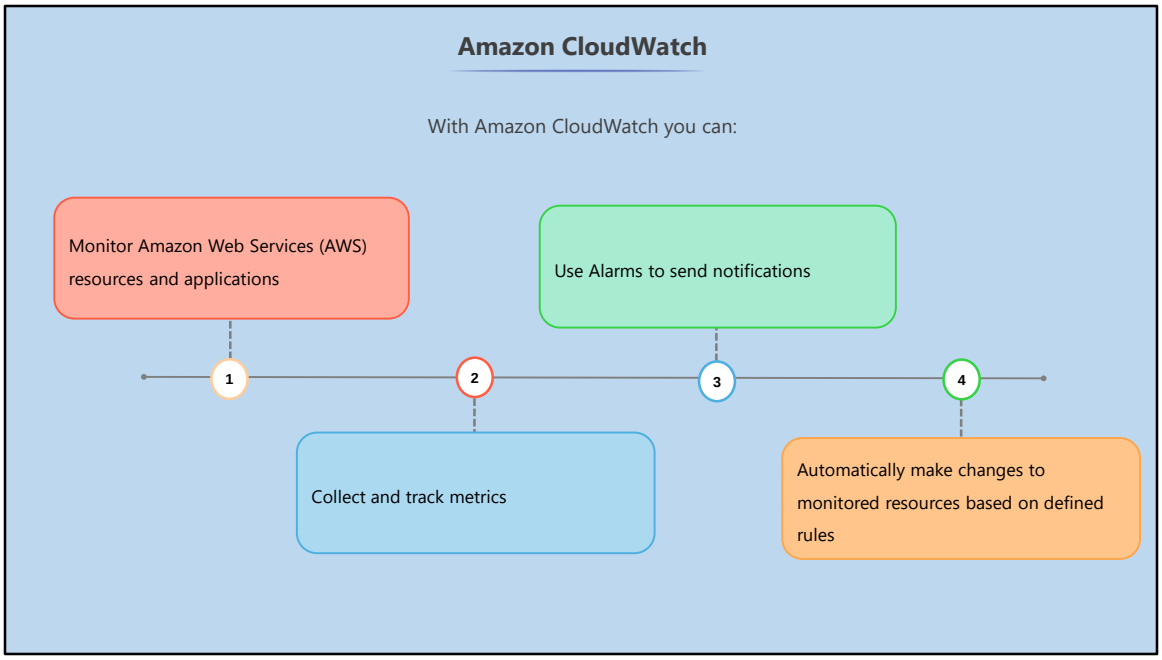
EC2 on-demand instance pricing means you only pay for what you use with no long term commitments. Reserved enables you to commit to a defined period of 12-36 months to receive significantly discounted hourly rates compared to on-demand pricing. EC2 Spot Instances are ideal for workloads that have flexible start and end time as you are allowed to bid on spare EC2 computing capacity.

Monitoring and Logging

Overview of the tools to enable AWS monitoring and logging

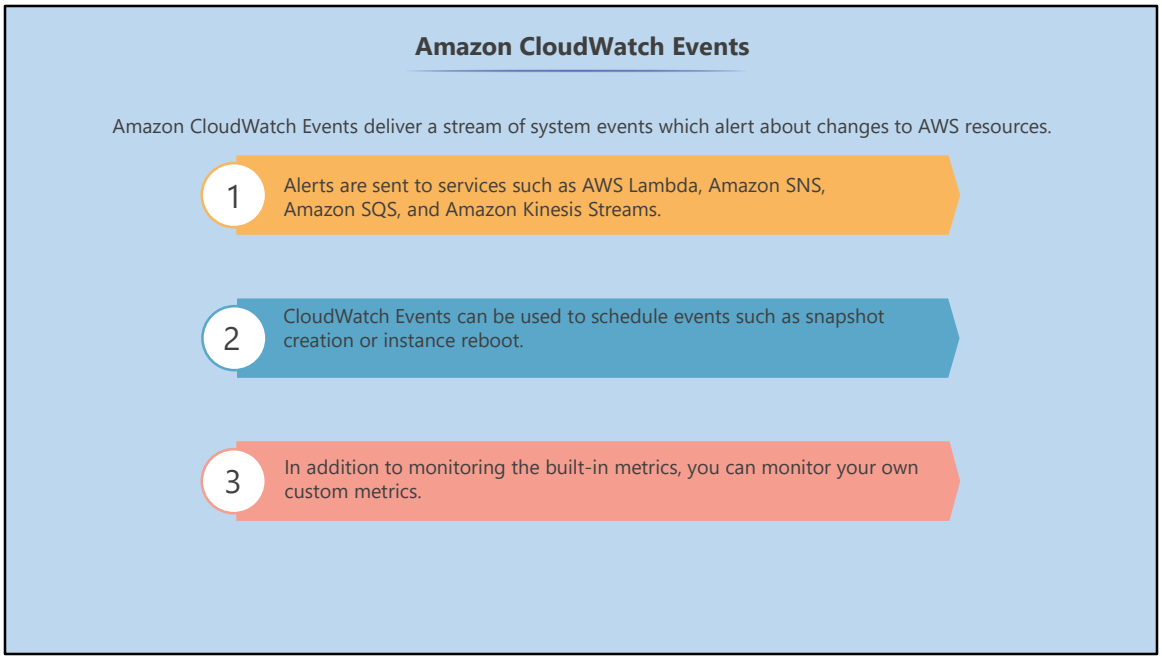


In this section you will learn about Amazon CloudWatch, Amazon CloudTrail, and Amazon Config.



Amazon CloudWatch:

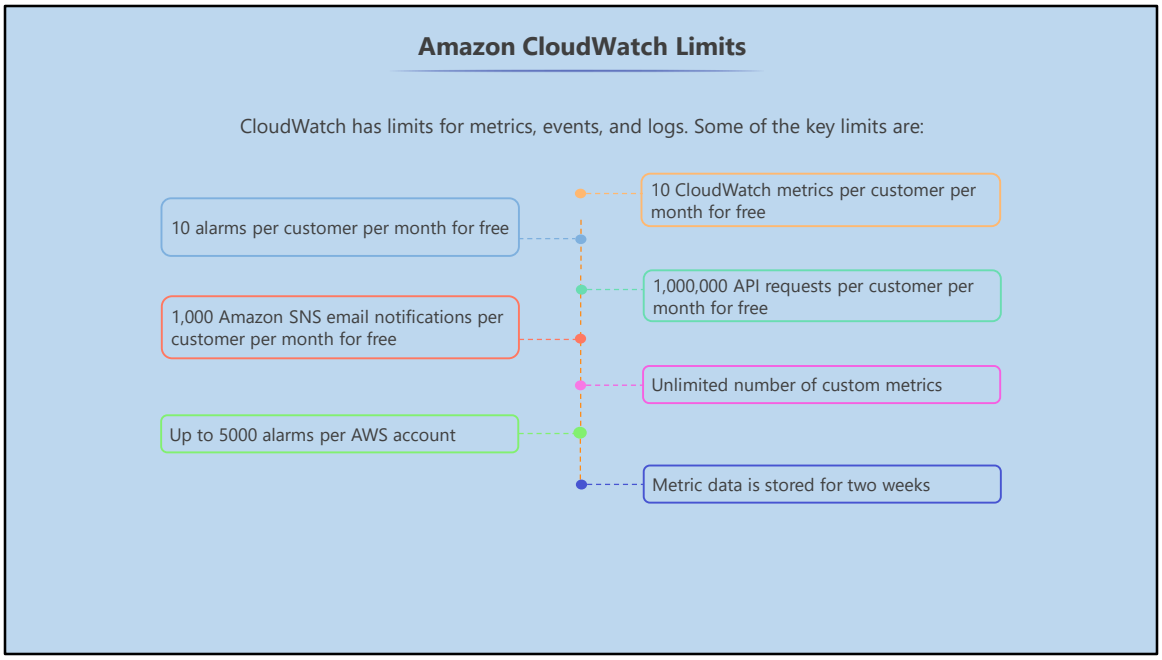
- Monitors your Amazon Web Services (AWS) resources and the applications for a particular region in real-time
- Collects and tracks metrics
- Uses alarms to send notifications
- Automatically makes changes to monitored resources based on defined rules



Amazon CloudWatch Events deliver a stream of system events which alert about changes to AWS resources. Alerts are sent to services such as AWS Lambda, Amazon SNS, Amazon SQS, and Amazon Kinesis Streams. For example, you can monitor the CPU usage, disk reads and writes of your Amazon Elastic Compute Cloud (Amazon EC2) instances and use this data to determine whether you should launch additional instances to handle the increased load. You can use this data to stop under-used instances to save money.

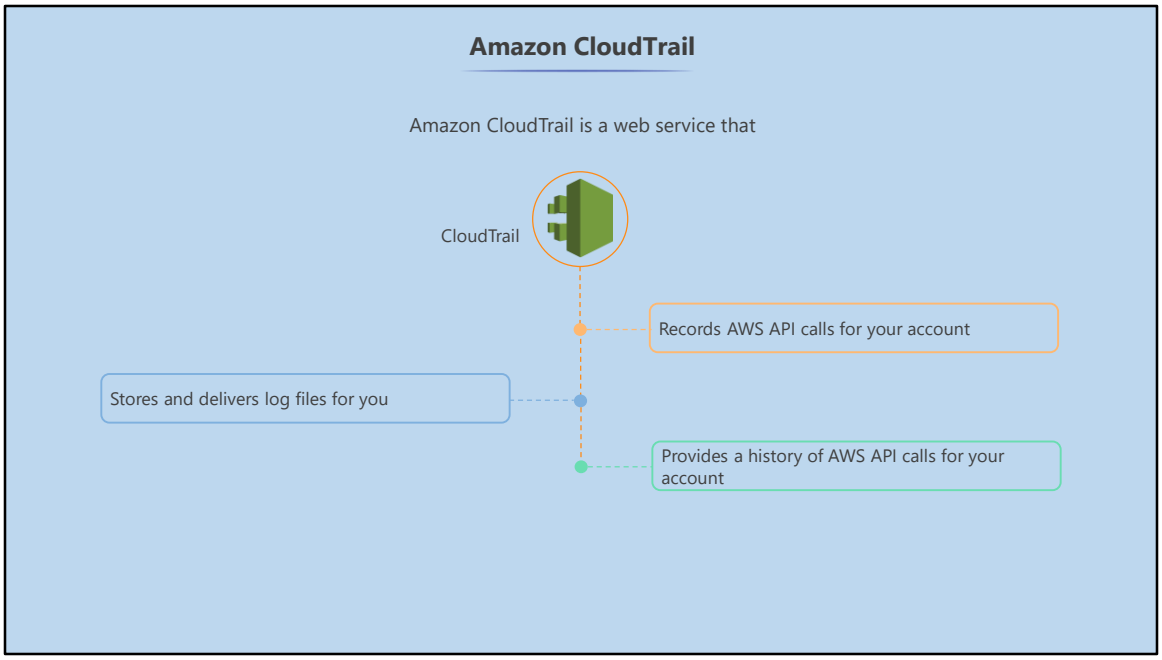
You can also use CloudWatch Events to schedule events such as snapshot creation or instance reboot. In addition to monitoring the built-in metrics that come with AWS, you can monitor your own custom metrics to enable system-wide visibility into resource utilization, application performance, operational health, and even cost.

Use Amazon CloudWatch Logs to monitor, store, and access application or system log files from Amazon EC2 instances, AWS CloudTrail, or other sources. You can then retrieve the associated log data from CloudWatch Logs.



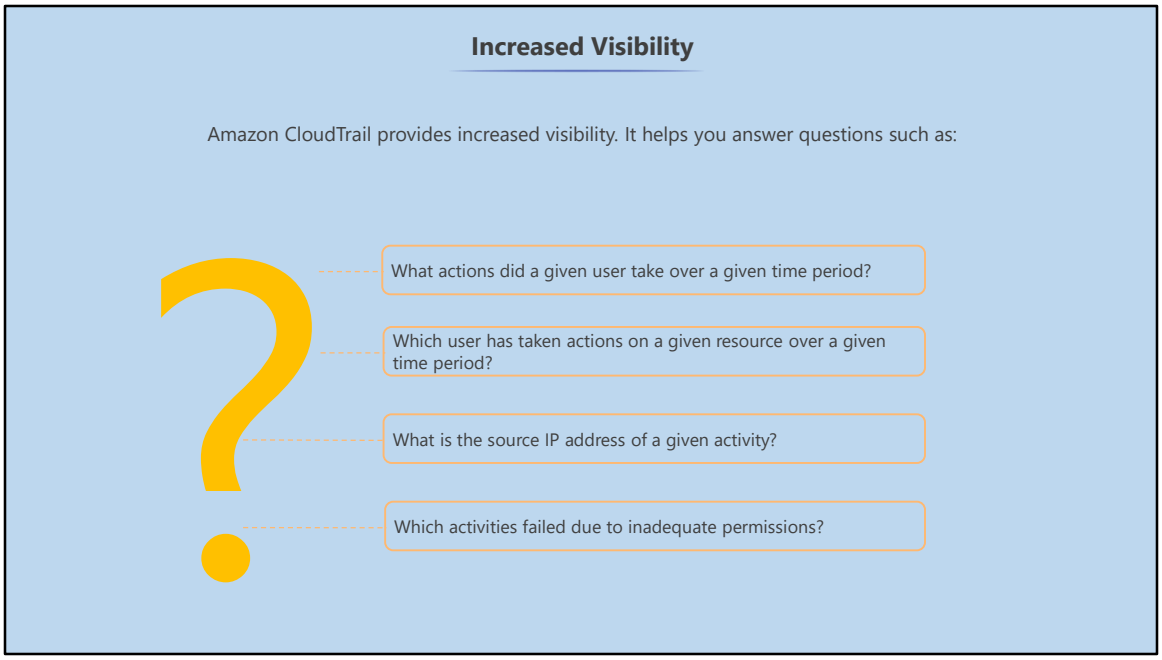
CloudWatch has limits for metrics, events, and logs. Some of the key limits are:

- 10 CloudWatch metrics—10 alarms, 1,000,000 API requests, and 1,000 Amazon SNS email notifications per customer per month for free.
- There is no limit on the number of custom metrics you create.
- You can create up to 5000 alarms per AWS account.
- Metric data is kept for two weeks.



Amazon CloudTrail is a web service that records AWS API calls for your account. The recorded information includes the identity of the API caller, the time of the API call, the source IP address of the API caller, the request parameters, and the response elements returned by the AWS service. It also stores and delivers log files to you.

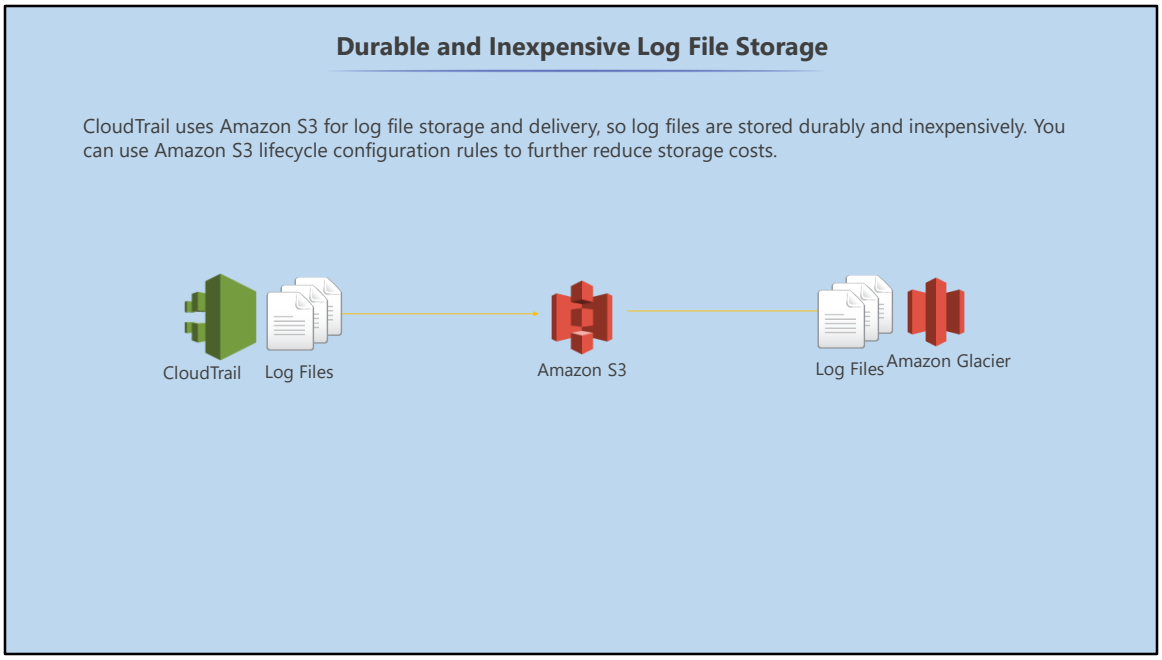
It provides history of AWS API calls for your account including API calls made via the AWS Management Console, AWS SDKs, command line tools, and higher-level AWS services (such as AWS CloudFormation).



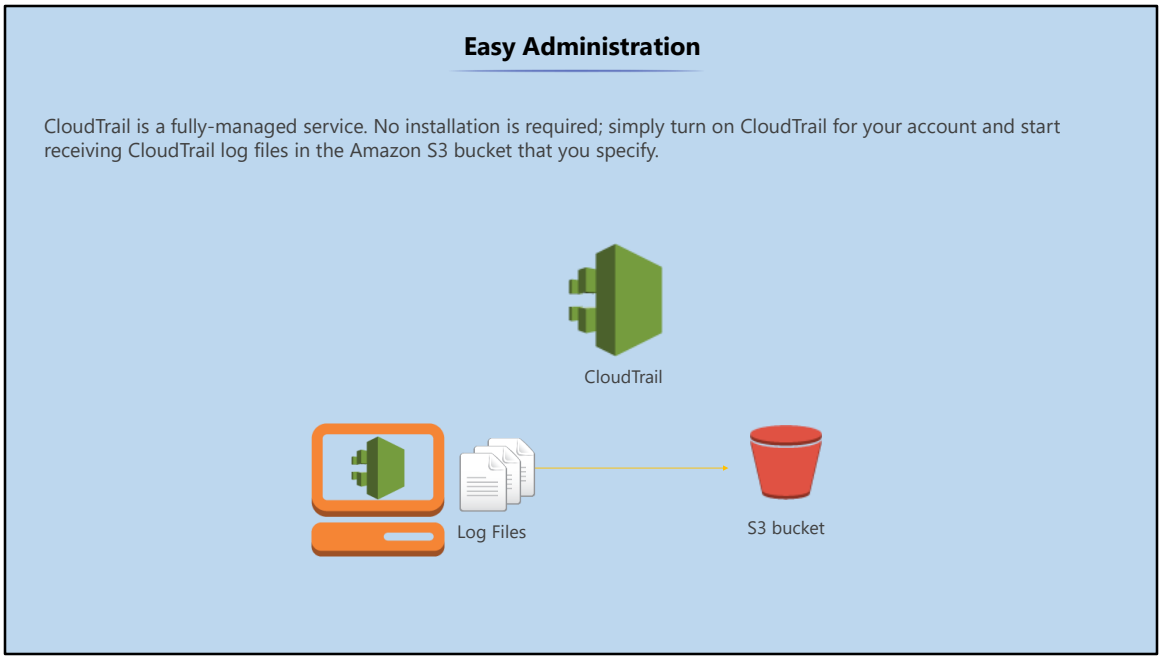
CloudTrail provides increased visibility. It helps you answer questions such as:

- What actions did a given user take over a given time period?
- For a given resource, which user has taken actions on it over a given time period?
- What is the source IP address of a given activity?
- Which activities failed due to inadequate permissions?

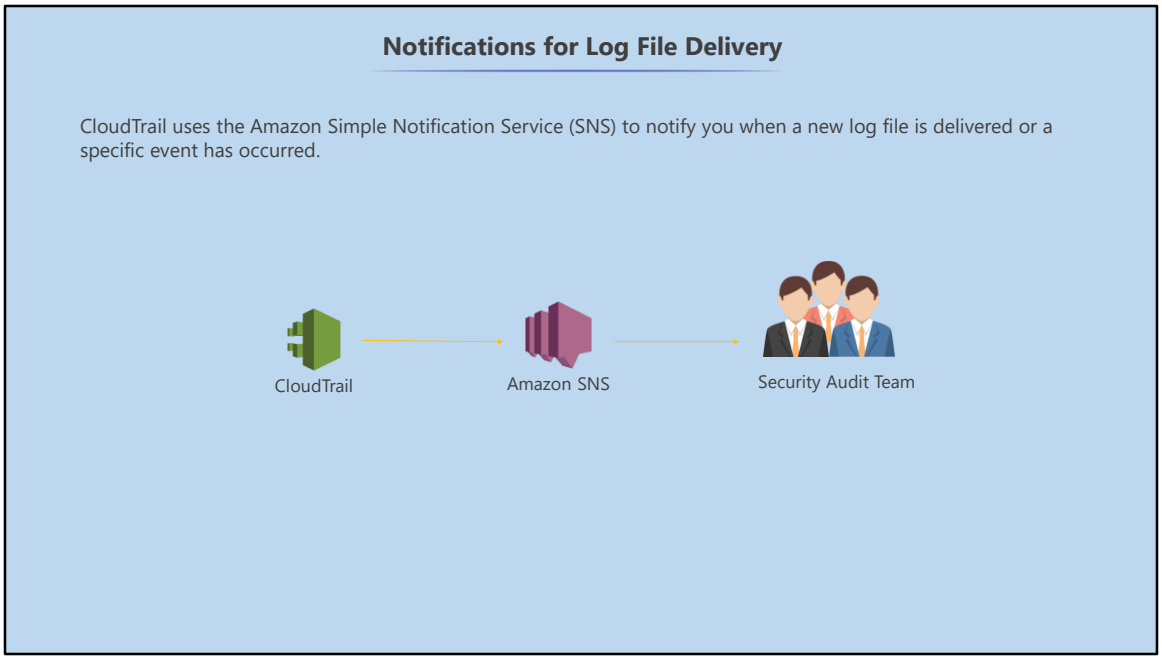
CloudTrail provides answers when you are trying to solve a problem.



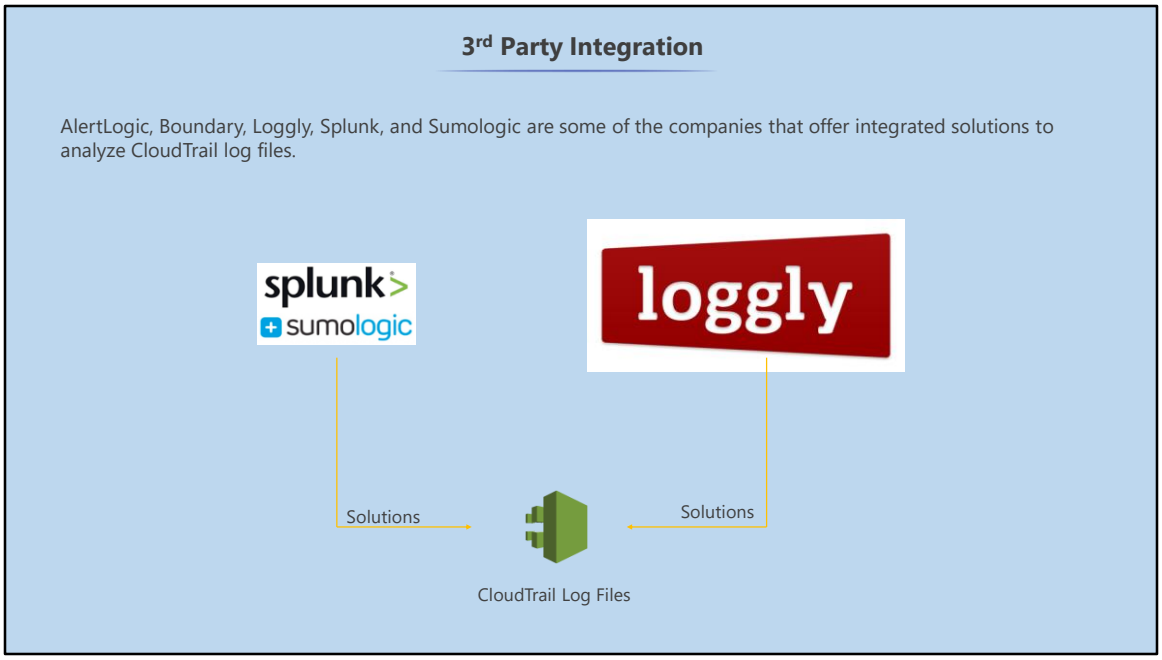
CloudTrail uses Amazon S3 for log file storage and delivery, so log files are stored durably and inexpensively.
Use Amazon S3 lifecycle configuration rules to further reduce storage costs. For example, you can define rules to automatically delete old log files or archive them to Amazon Glacier for additional savings.



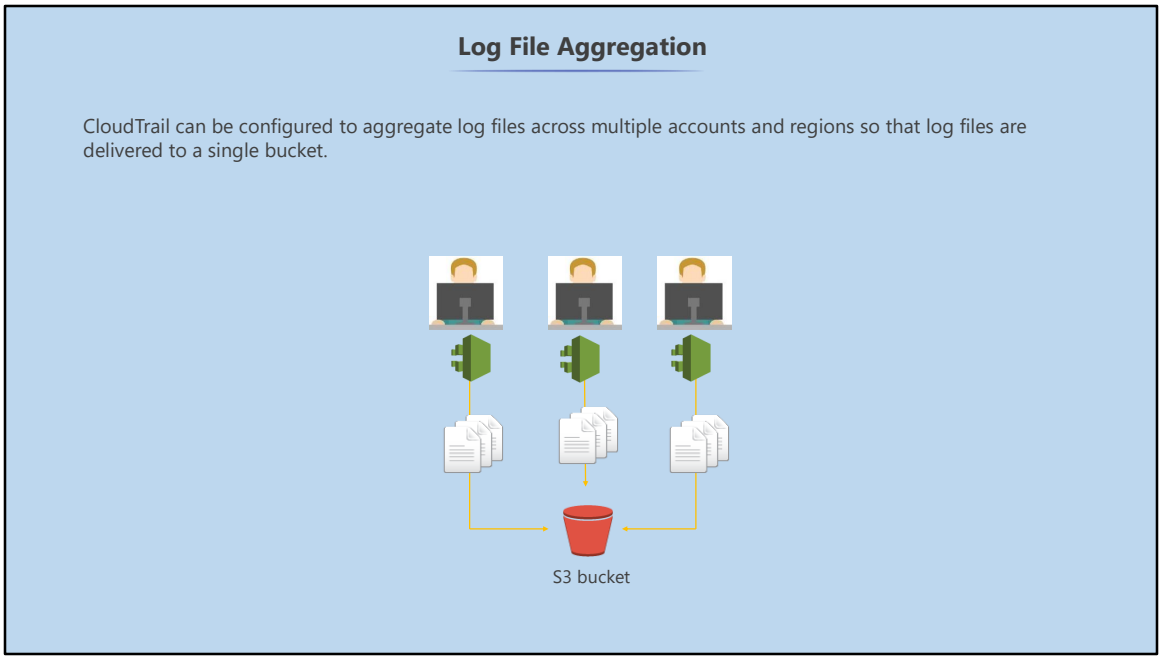
CloudTrail is a fully managed service. No installation is required. In the AWS Management Console turn on CloudTrail for your account and start receiving CloudTrail log files in the Amazon Simple Storage Service (Amazon S3) bucket that you specify.



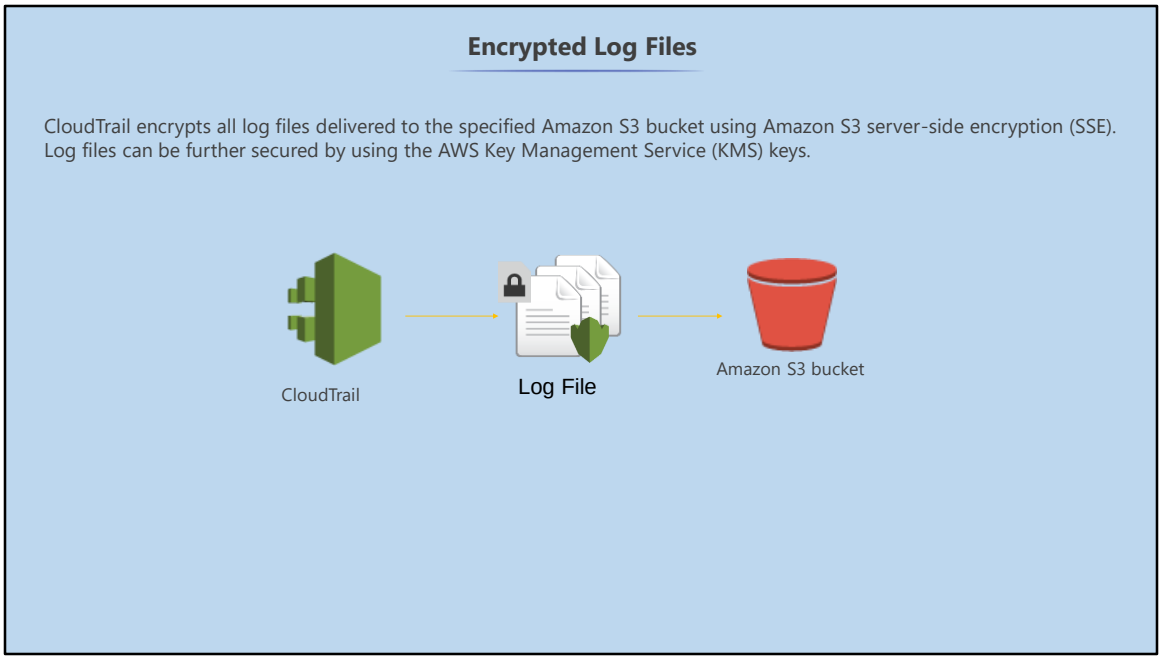
CloudTrail uses the Amazon Simple Notification Service (SNS) to notify you when a new log file is delivered or a specific event has occurred.



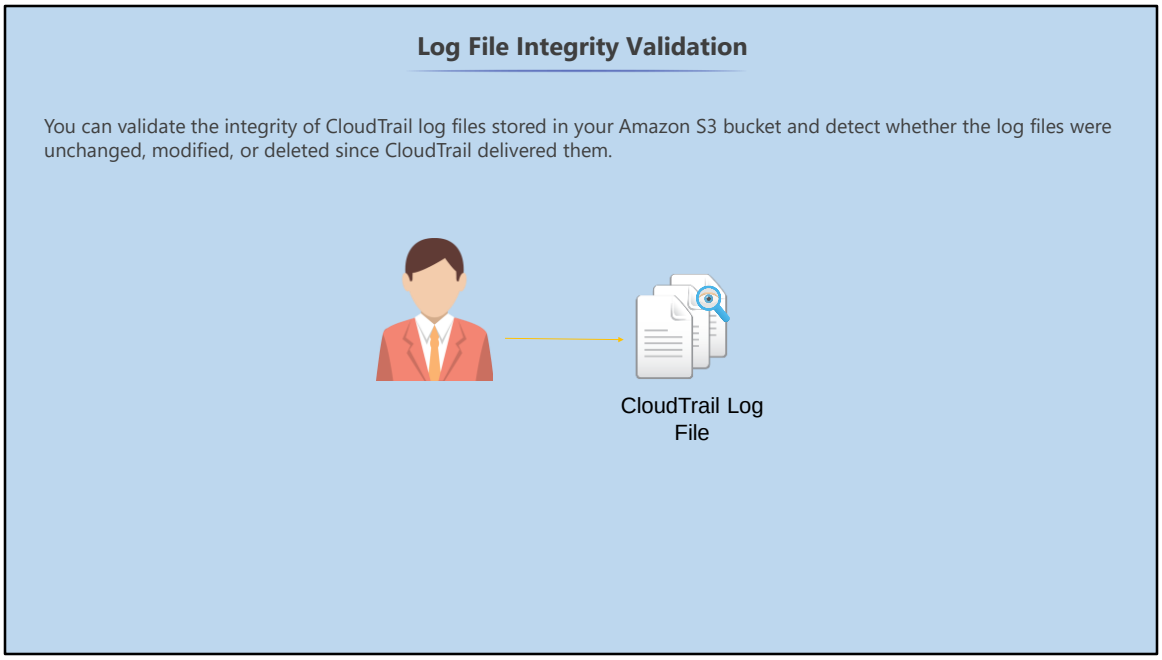
Companies such as AlertLogic, Boundary, Loggly, Splunk, and Sumologic offer integrated solutions to analyze CloudTrail log files.



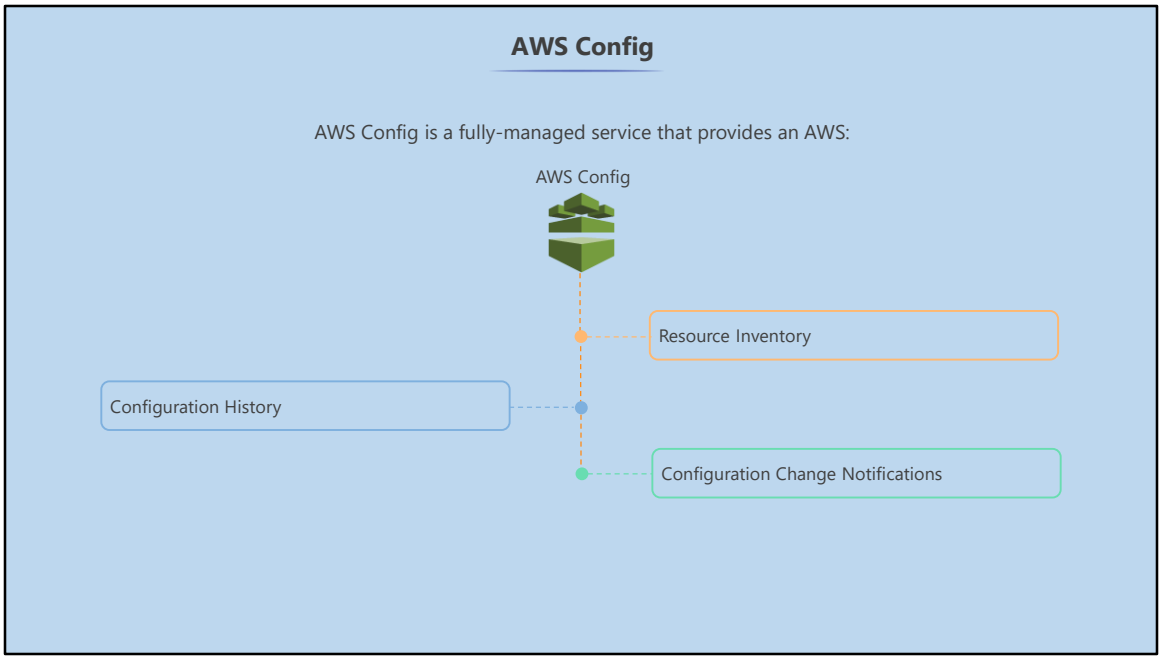
CloudTrail can be configured to aggregate log files across multiple accounts and regions so that log files are delivered to a single bucket.



CloudTrail by default, encrypts all log files delivered to the specified Amazon S3 bucket using Amazon S3 server-side encryption (SSE). You can further secure log files by using the AWS Key Management Service (KMS) keys.



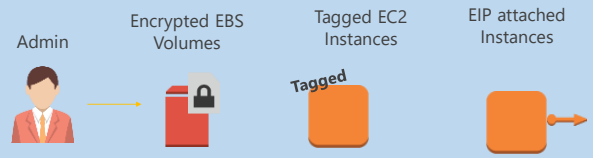
You can validate the integrity of CloudTrail log files stored in your Amazon S3 bucket and detect whether the log files were unchanged, modified, or deleted since CloudTrail delivered them to your Amazon S3 bucket.



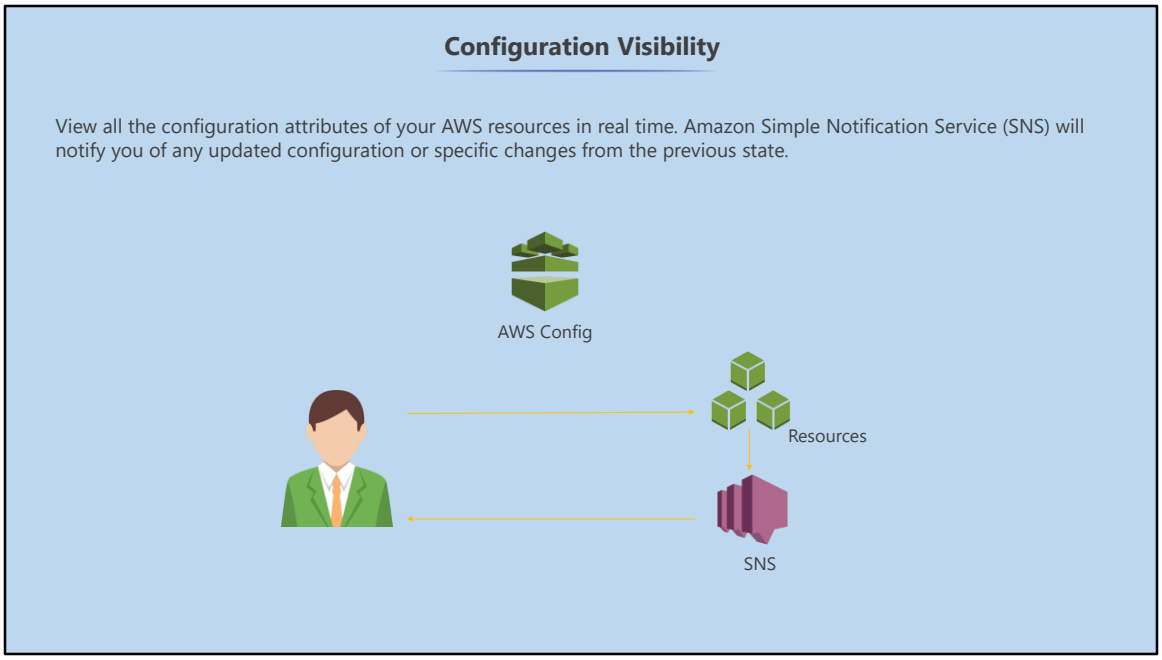
AWS Config is a fully managed service that provides an AWS resource inventory, configuration history, and configuration change notifications.
Config Rules enable you to create rules that automatically check the configuration of AWS resources recorded by AWS Config.

AWS Config (Contd.)

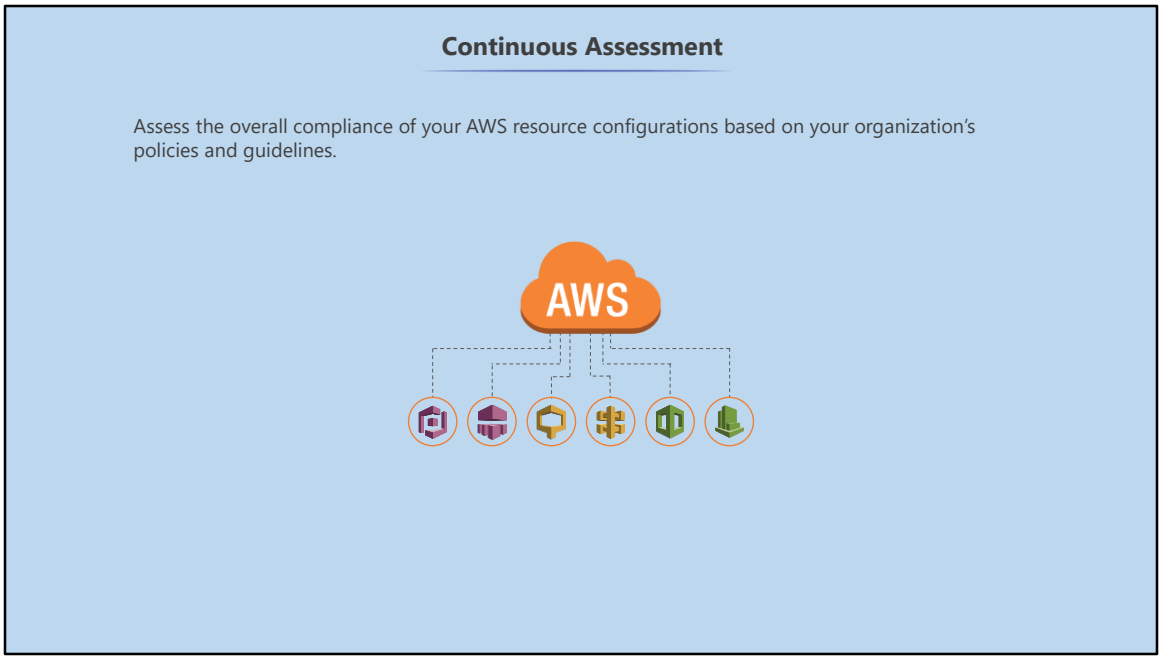
Using Config Rules, an IT Administrator can quickly determine when and how a resource went out of compliance. For example, it ensures EBS volumes are encrypted, EC2 instances are tagged, and Elastic IP addresses (EIPs) are attached to instances.



Using Config Rules, an IT Administrator can quickly determine when and how a resource went out of compliance. For example, ensure EBS volumes are encrypted, EC2 instances are properly tagged, and Elastic IP addresses (EIPs) are attached to instances.



View all the configuration attributes of your AWS resources in real time. Amazon Simple Notification Service (SNS) will notify you of any updated configuration or specific changes from the previous state.



You can easily assess the overall compliance of your AWS resource configurations based on your organization's policies and guidelines.

AWS Config Rules provide a visual dashboard with lists, charts, and graphs to help quickly spot non-compliant resources and take appropriate action. AWS config is also compatible with 3rd party applications such as Splunk.

Knowledge Check

1 **What services assist you with the monitoring and logging of your cloud environment?**

Amazon CloudFront, Amazon CloudFormation, and Amazon Trusted Advisor

Amazon CloudWatch, Amazon CloudFormation, and Amazon CloudTrail

Amazon CloudWatch, Amazon CloudTrail, and Amazon Config

Amazon CloudFront, Amazon CloudFormation, and Amazon Trusted Advisor

1 **What services assist you with the monitoring and logging of your cloud environment?**

Amazon CloudFront, Amazon CloudFormation, and Amazon Trusted Advisor

Amazon CloudWatch, Amazon CloudFormation, and Amazon CloudTrail

Amazon CloudWatch, Amazon CloudTrail, and Amazon Config

Amazon CloudFront, Amazon CloudFormation, and Amazon Trusted Advisor

c

Amazon CloudWatch, Amazon CloudTrail, and Amazon Config are the managed services that provide monitoring and logging of your cloud environment.

2 **Which tool would you use to monitor AWS resource and performance utilization?**

Amazon CloudTrail

Amazon CloudWatch

Amazon Config

Amazon CloudFront

2 **Which tool would you use to monitor AWS resource and performance utilization?**

Amazon CloudTrail

Amazon CloudWatch

Amazon Config

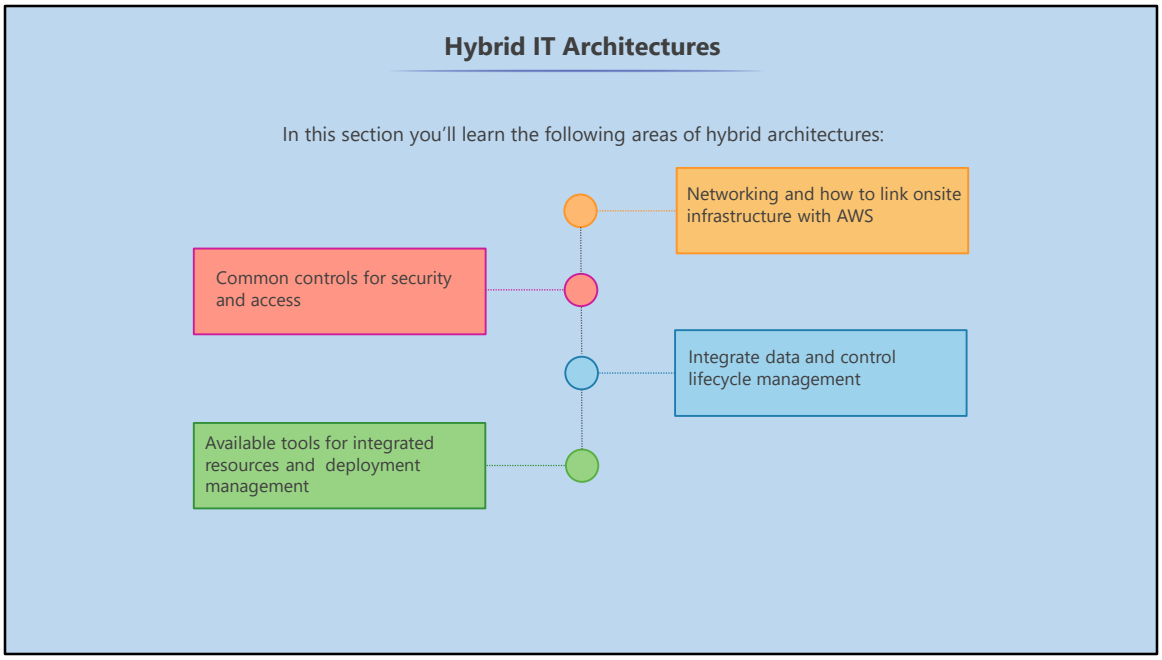
Amazon CloudFront

b

Amazon CloudWatch monitors your Amazon Web Service (AWS) resources and the applications in real-time in a particular region. Amazon CloudTrail records AWS API calls for your account. Amazon Config reports on configuration changes made to your AWS resources. Amazon CloudFront is the Amazon CDN service.

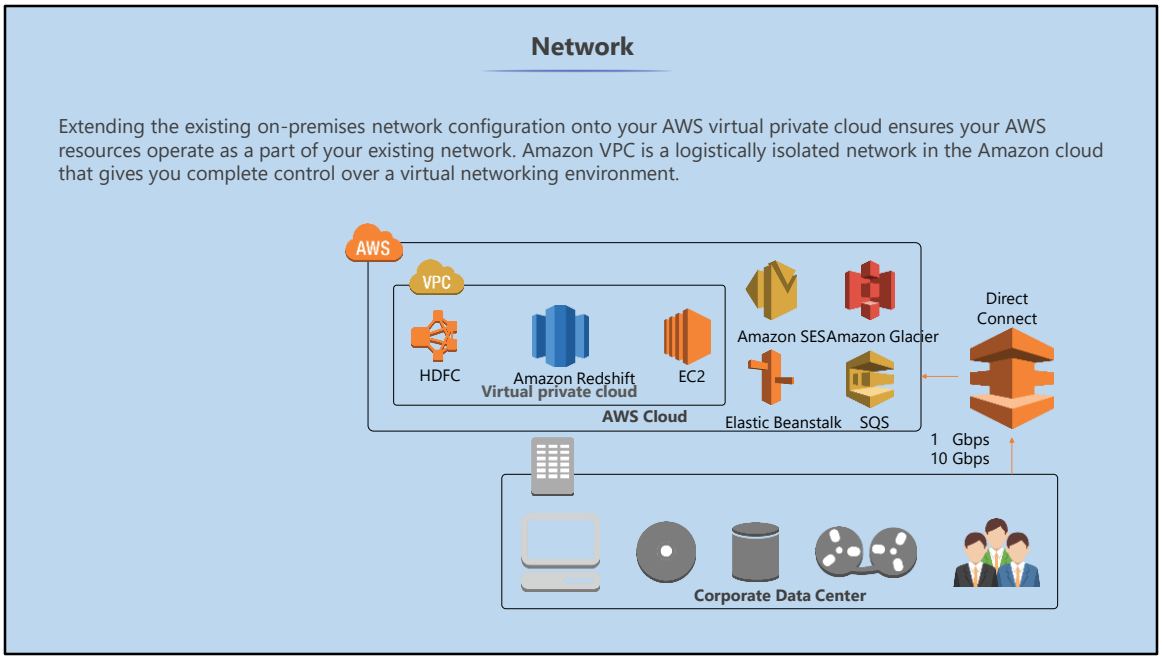
Hybrid IT architectures

This topic provides an overview of the tools and functionality available to run hybrid cloud architectures.



In this section you'll learn the following areas of hybrid architectures:

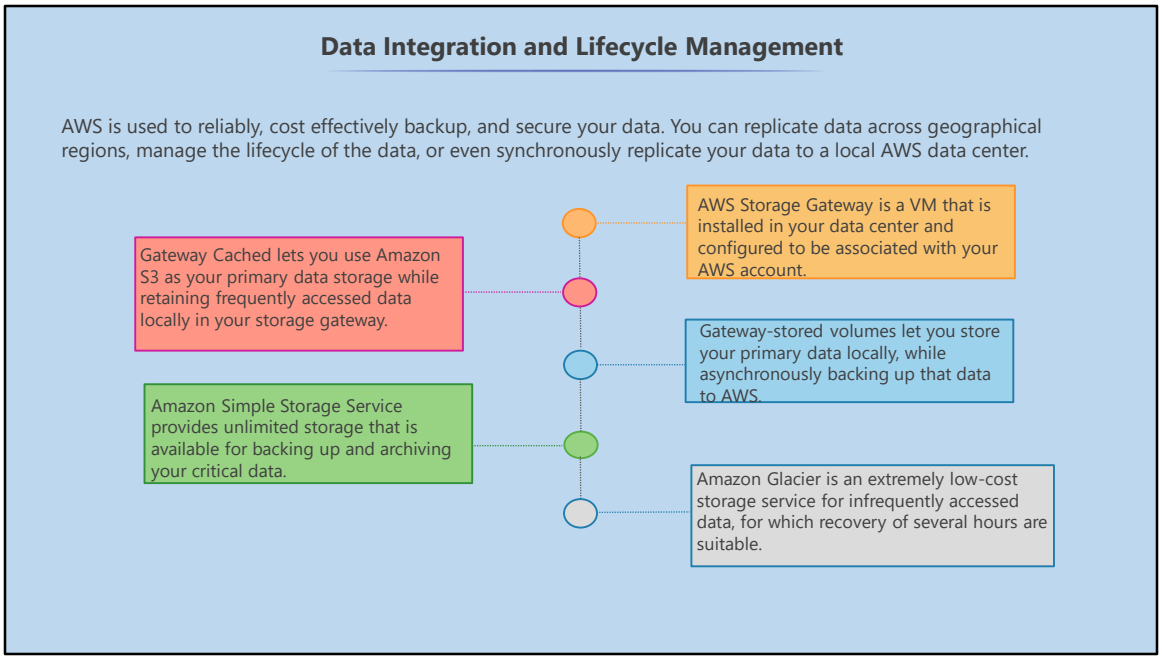
- Networking and how to link your onsite infrastructure with AWS
- How to integrate your data and control lifecycle management
- The common controls for security and access
- The tools available for integrated resource and deployment management



Extending the existing on-premises network configuration onto your AWS virtual private cloud ensures that your AWS resources will operate as a part of your existing network.

Amazon VPC is a logistically isolated network in the Amazon cloud that gives you complete control over a virtual networking environment. You can choose your own IP address range, subnets, route tables, network gateways, and more.

Rather than use Internet–based connections between your onsite resources and the Amazon cloud, AWS Direct Connect lets you establish a dedicated network connection between the two to provide lower costs and a higher level of service.



Use AWS to reliably, cost effectively backup, and secure your data.
AWS allows you to replicate data across geographical regions, manage the lifecycle of the data, or even synchronously replicate your data to a local AWS datacenter.

AWS Storage Gateway is a VM that is installed in your datacenter and configured to be associated with your AWS account. Once it's set up you can use the AWS Management Console to create volumes that are either Gateway-Cached or Gateway-Stored storage volumes which are mounted on the on-premise applications.

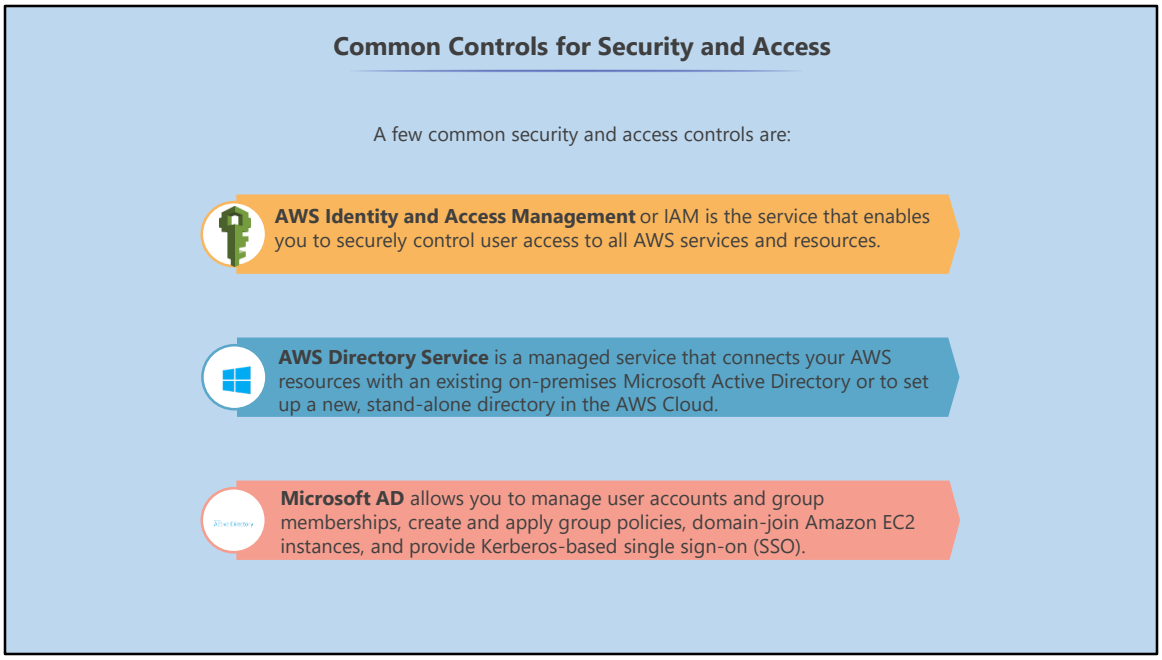
Gateway Cached lets you use Amazon Simple Storage Service (Amazon S3) as your primary data storage while retaining frequently accessed data locally in your storage gateway.

Gateway-stored volumes let you store your primary data locally, while asynchronously backing up that data to AWS.

Amazon Simple Storage Service

Provides unlimited storage that is available for backing up and archiving your critical data. Data is stored in multiple facilities, on multiple devices within each facility to provide redundancy.

Amazon Glacier is extremely low-cost storage service for infrequently accessed data, for which recovery intervals of several hours are suitable.



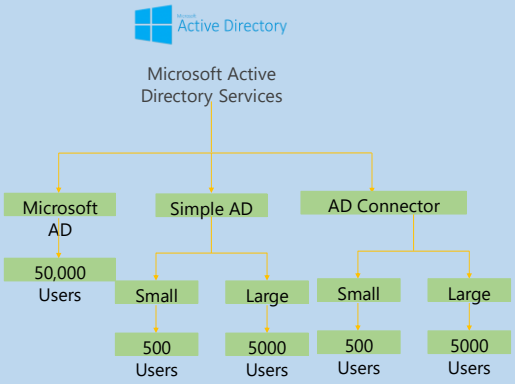
AWS Identity and Access Management or IAM is the service that enables you to securely control user access to all AWS services and resources.

AWS Directory Service is a managed service that connect your AWS resources with an existing on-premises Microsoft Active Directory or to set up a new, stand-alone directory in the AWS Cloud.

Microsoft AD allows you to manage user accounts and group memberships, create and apply group policies, domain-join Amazon EC2 instances, and provide Kerberos-based single sign-on (SSO).

Common Controls for Security and Access (Contd.)

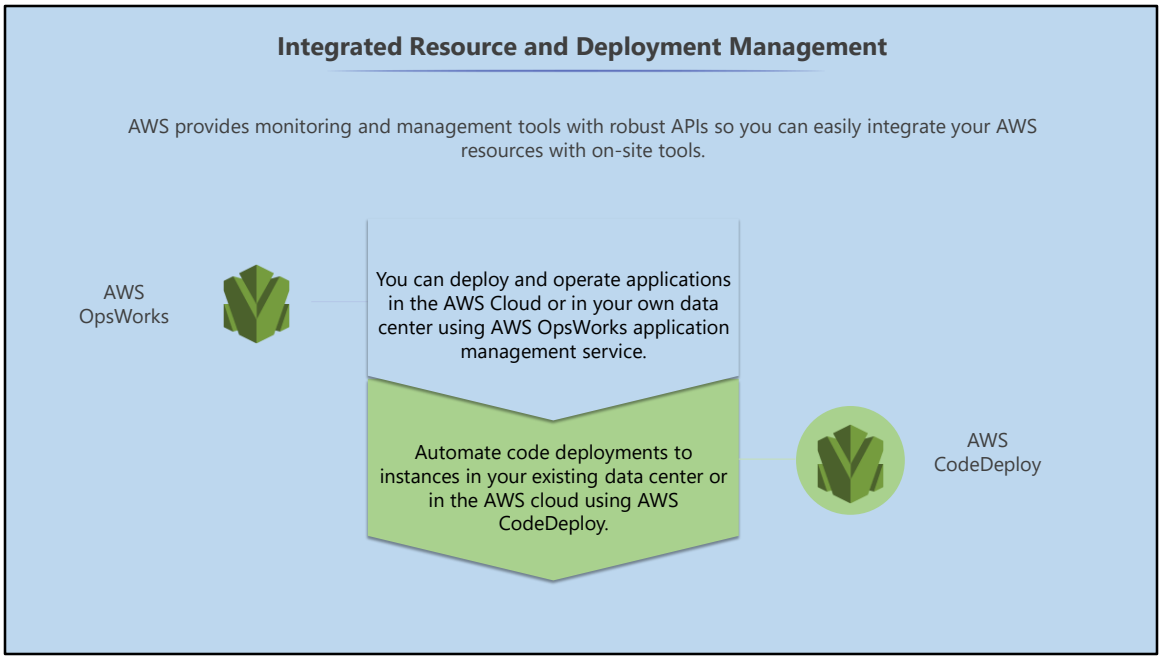
- Microsoft AD is a full blown managed Microsoft Active Directory service that supports up to 50,000 users.
- Simple AD is a stand-alone, managed directory that is available in two sizes; small and large.
- AD Connector is a directory gateway that allows you to proxy directory requests to your on-premises Microsoft Active Directory. AD Connector comes in two sizes; small and large.



Microsoft AD is full blown managed Microsoft Active Directory service that supports up to 50,000 users.

Simple AD is a stand-alone, managed directory that is available in two sizes; small and large. A small Simple AD supports up to 500 users. A large Simple AD supports up to 5,000 users.

AD Connector is a directory gateway that allows you to proxy directory requests to your on-premises Microsoft Active Directory. AD Connector comes in two sizes; small and large. A small AD Connector is designed for smaller organizations of up to 500 users. A large AD Connector is designed for larger organizations of up to 5,000 users.



AWS provides monitoring and management tools with robust APIs so you can easily integrate your AWS resources with on-site tools. Most major software vendors already support AWS, for example, Microsoft, VMWare, and BMC.

AWS OpsWorks deploy and operate applications in the AWS Cloud or in your data center using the AWS OpsWorks application management service. You can use templates or build your own tasks to specify an applications architecture.

AWS CodeDeploy automates code deployments to instances in your existing data center or in the AWS cloud using AWS CodeDeploy. AWS CodeDeploy simplifies the code deployment process to your instances.

Knowledge Check

1 **Which of the following is NOT a service used directly in hybrid architectures?**

AWS Direct Connect

AWS Direct Connect

Amazon Config

AWS Directory Service

1 **Which of the following is NOT a service used directly in hybrid architectures?**

AWS Direct Connect

AWS Direct Connect

Amazon Config

AWS Directory Service

c

AWS Storage Gateway is used to store your data on the cloud via your data center. AWS Direct Connect lets you establish a dedicated network connection between your onsite premises and AWS. AWS Directory Service is a managed service that allows you to connect your AWS resources with an existing on-premises Microsoft Active Directory or to set up a new, stand-alone directory in the AWS Cloud.

2 **Which Storage Gateway option would you choose if you wanted to use solely AWS storage?**

Gateway Cached

Gateway S3

Gateway Stored

Gateway Remote

2 **Which Storage Gateway option would you choose if you wanted to use solely AWS storage?**

Gateway Cached

Gateway S3

Gateway Stored

Gateway Remote

a

Gateway Cached lets you use Amazon Simple Storage Service (Amazon S3) as your primary data storage while retaining frequently accessed data locally in your storage gateway. Gateway Stored volumes let you store your primary data locally, while asynchronously backing up that data to AWS.

Designing Hybrid Storage

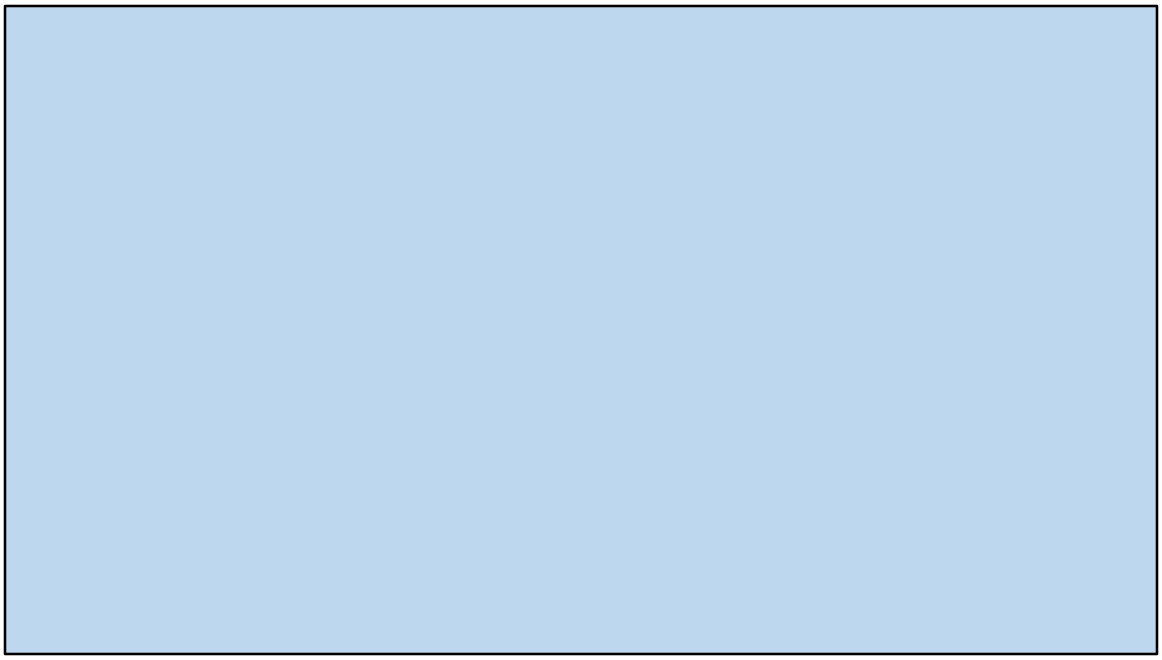
You have been hired by a medium sized law firm. They have an aging storage solution which they want to replace, but they do not want to purchase any hardware.

They store several terabytes of data, comprising documents and images. A lot of the data is legacy data, which is rarely accessed and can be archived. However, the most recent files need to be available instantly.

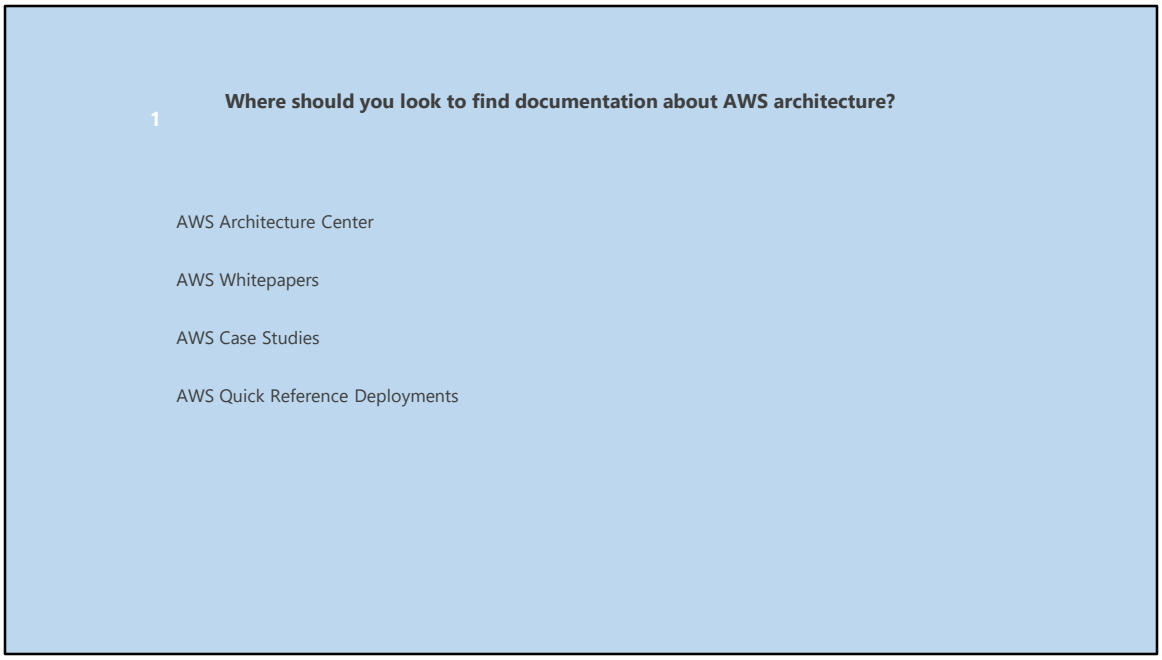
You have been asked to provide a basic, high-level plan for a hybrid storage solution using the client's existing data center and AWS.

Detail the products and services you would use and sketch out a basic plan of the infrastructure.

- AWS Framework helps you to understand the pros and cons of decisions you make while building systems on AWS.
- The AWS Framework is based on four pillars: Security, Reliability, Performance efficiency, and Cost Optimization.
- Cloud computing helps achieve optimal server configuration by providing various features.
- You can configure Amazon CloudWatch, CloudTrail, and AWS Config to provide you with alerts and notifications.
- Hybrid technologies helps you link your existing on-premises network configuration onto your AWS virtual private cloud.



- AWS Well-Architected Framework helps you to understand the pros and cons of decisions you make while building systems on AWS.
- The AWS Well-Architected Framework is based on four pillars: Security, Reliability, Performance efficiency, and Cost Optimization.
- Cloud computing helps achieve optimal server configuration by providing various features.
- You can configure Amazon CloudWatch, CloudTrail, and AWS Config to provide you with alerts and notifications.
- With hybrid technologies you can link your existing on-premises network configuration onto your AWS virtual private cloud.



- AWS Well-Architected Framework helps you to understand the pros and cons of decisions you make while building systems on AWS.
- The AWS Well-Architected Framework is based on four pillars: Security, Reliability, Performance efficiency, and Cost Optimization.
- Cloud computing helps achieve optimal server configuration by providing various features.
- You can configure Amazon CloudWatch, CloudTrail, and AWS Config to provide you with alerts and notifications.
- With hybrid technologies you can link your existing on-premises network configuration onto your AWS virtual private cloud.

1 **Where should you look to find documentation about AWS architecture?**

AWS Architecture Center

AWS Whitepapers

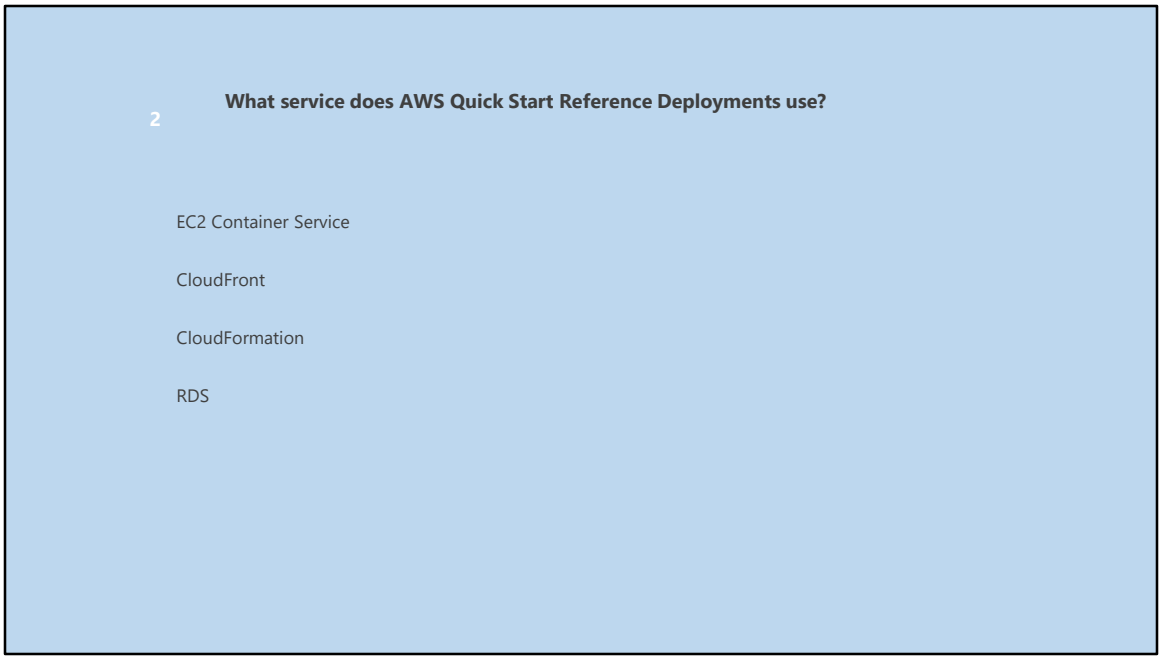
AWS Case Studies

AWS Quick Reference Deployments

a, b, and c

AWS Architecture Center, AWS Whitepapers, and AWS Case Studies will provide you information about AWS architecture.

- AWS Well-Architected Framework helps you to understand the pros and cons of decisions you make while building systems on AWS.
- The AWS Well-Architected Framework is based on four pillars: Security, Reliability, Performance efficiency, and Cost Optimization.
- Cloud computing helps achieve optimal server configuration by providing various features.
- You can configure Amazon CloudWatch, CloudTrail, and AWS Config to provide you with alerts and notifications.
- With hybrid technologies you can link your existing on-premises network configuration onto your AWS virtual private cloud.



- AWS Well-Architected Framework helps you to understand the pros and cons of decisions you make while building systems on AWS.
- The AWS Well-Architected Framework is based on four pillars: Security, Reliability, Performance efficiency, and Cost Optimization.
- Cloud computing helps achieve optimal server configuration by providing various features.
- You can configure Amazon CloudWatch, CloudTrail, and AWS Config to provide you with alerts and notifications.
- With hybrid technologies you can link your existing on-premises network configuration onto your AWS virtual private cloud.

2 What service does AWS Quick Start Reference Deployments use?

EC2 Container Service

CloudFront

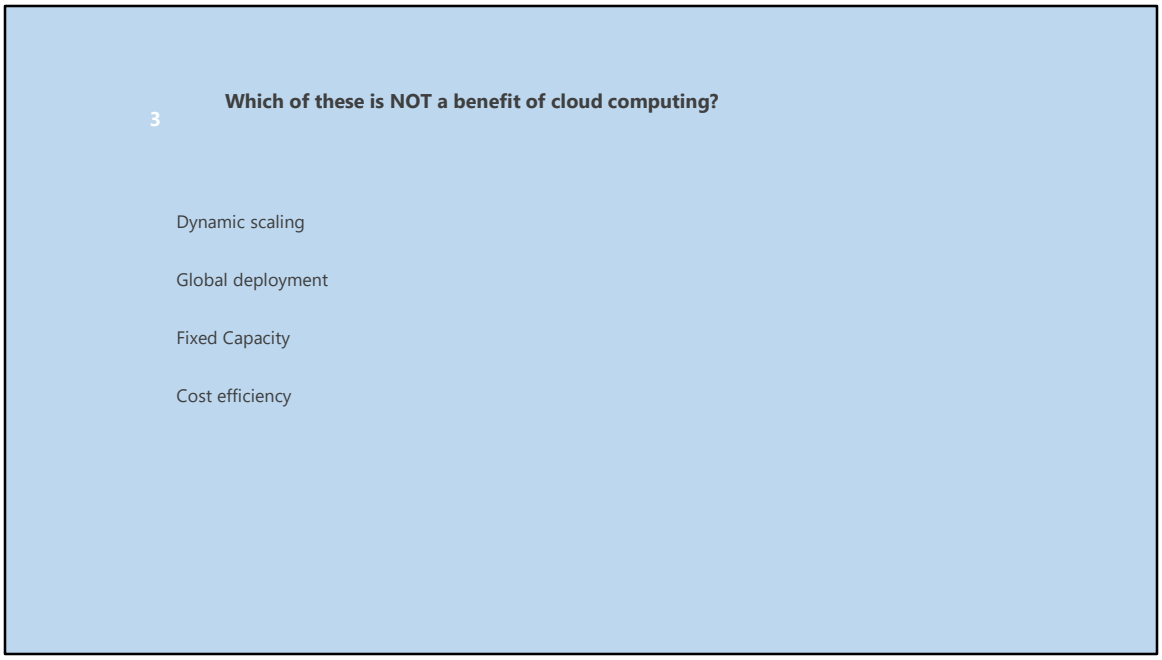
CloudFormation

RDS

c

CloudFormation is used to rapidly deploy a fully functioning environment for a variety of enterprise software applications.

- AWS Well-Architected Framework helps you to understand the pros and cons of decisions you make while building systems on AWS.
- The AWS Well-Architected Framework is based on four pillars: Security, Reliability, Performance efficiency, and Cost Optimization.
- Cloud computing helps achieve optimal server configuration by providing various features.
- You can configure Amazon CloudWatch, CloudTrail, and AWS Config to provide you with alerts and notifications.
- With hybrid technologies you can link your existing on-premises network configuration onto your AWS virtual private cloud.



- ~~AWS Well-Architected Framework helps you to understand the pros and cons of decisions you make while building systems on AWS.~~
- ~~The AWS Well-Architected Framework is based on four pillars: Security, Reliability, Performance efficiency, and Cost Optimization.~~
- ~~Cloud computing helps achieve optimal server configuration by providing various features.~~
- ~~You can configure Amazon CloudWatch, CloudTrail, and AWS Config to provide you with alerts and notifications.~~
- ~~With hybrid technologies you can link your existing on-premises network configuration onto your AWS virtual private cloud~~.

3 **Which of these is NOT a benefit of cloud computing?**

Dynamic scaling

Global deployment

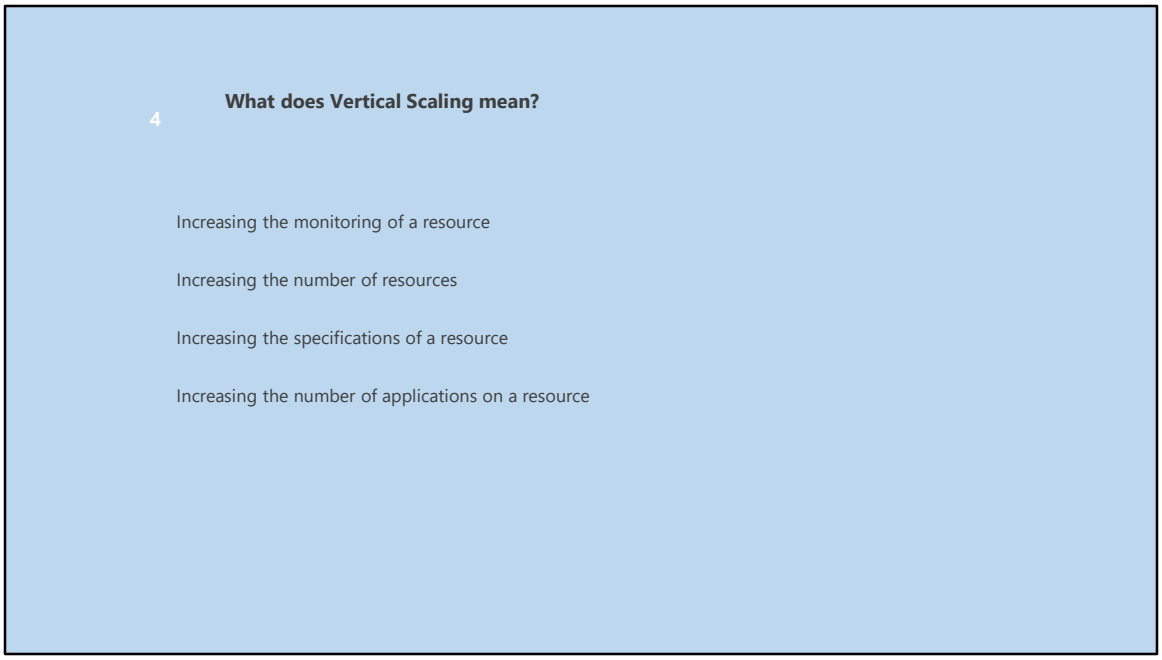
Fixed Capacity

Cost efficiency

c

Fixed capacity is associated with traditional IT infrastructure. With AWS you don't need to worry about provisioning capacity based on estimates.

- AWS Well-Architected Framework helps you to understand the pros and cons of decisions you make while building systems on AWS.
- The AWS Well-Architected Framework is based on four pillars: Security, Reliability, Performance efficiency, and Cost Optimization.
- Cloud computing helps achieve optimal server configuration by providing various features.
- You can configure Amazon CloudWatch, CloudTrail, and AWS Config to provide you with alerts and notifications.
- With hybrid technologies you can link your existing on-premises network configuration onto your AWS virtual private cloud.



- AWS Well-Architected Framework helps you to understand the pros and cons of decisions you make while building systems on AWS.
- The AWS Well-Architected Framework is based on four pillars: Security, Reliability, Performance efficiency, and Cost Optimization.
- Cloud computing helps achieve optimal server configuration by providing various features.
- You can configure Amazon CloudWatch, CloudTrail, and AWS Config to provide you with alerts and notifications.
- With hybrid technologies you can link your existing on-premises network configuration onto your AWS virtual private cloud.

4 **What does Vertical Scaling mean?**

Increasing the monitoring of a resource

Increasing the number of resources

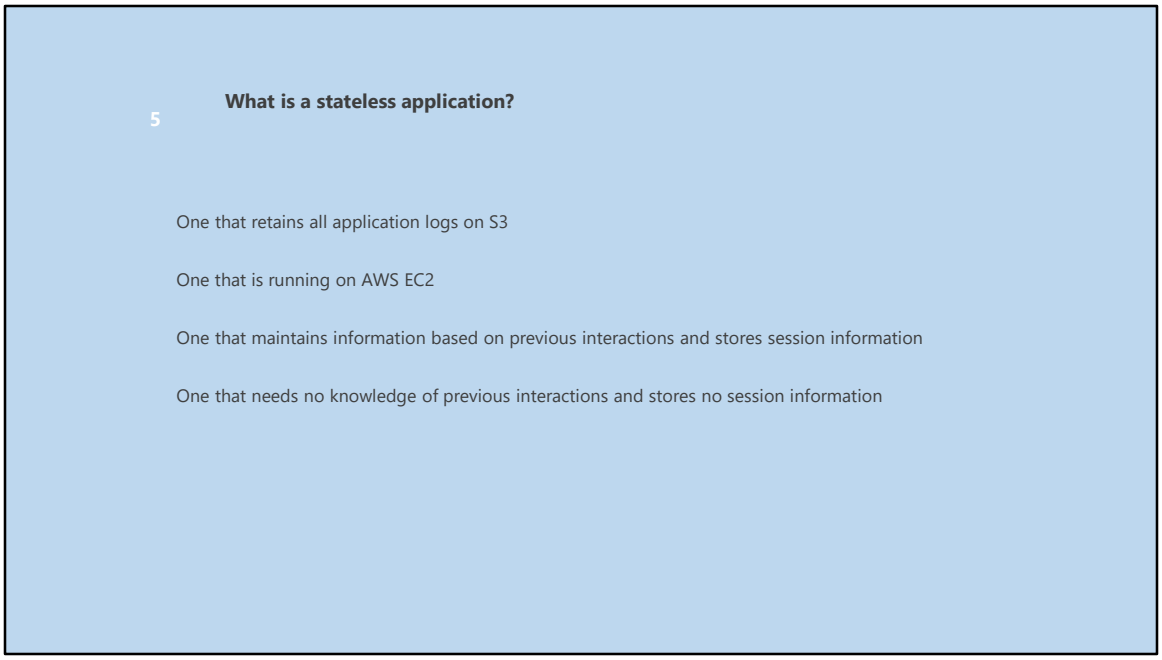
Increasing the specifications of a resource

Increasing the number of applications on a resource

c

Vertical Scaling means increasing the specifications of a resource, for example increasing the memory and CPU.

- AWS Well-Architected Framework helps you to understand the pros and cons of decisions you make while building systems on AWS.
- The AWS Well-Architected Framework is based on four pillars: Security, Reliability, Performance efficiency, and Cost Optimization.
- Cloud computing helps achieve optimal server configuration by providing various features.
- You can configure Amazon CloudWatch, CloudTrail, and AWS Config to provide you with alerts and notifications.
- With hybrid technologies you can link your existing on-premises network configuration onto your AWS virtual private cloud.



- AWS Well-Architected Framework helps you to understand the pros and cons of decisions you make while building systems on AWS.
- The AWS Well-Architected Framework is based on four pillars: Security, Reliability, Performance efficiency, and Cost Optimization.
- Cloud computing helps achieve optimal server configuration by providing various features.
- You can configure Amazon CloudWatch, CloudTrail, and AWS Config to provide you with alerts and notifications.
- With hybrid technologies you can link your existing on-premises network configuration onto your AWS virtual private cloud.

5 **What is a stateless application?**

One that retains all application logs on S3

One that is running on AWS EC2

One that maintains information based on previous interactions and stores session information

One that needs no knowledge of previous interactions and stores no session information

d

A stateless application is one that needs no knowledge of previous interactions and stores no session information. An example of this would be a webserver that provides the same web page to any end user.

- AWS Well-Architected Framework helps you to understand the pros and cons of decisions you make while building systems on AWS.
- The AWS Well-Architected Framework is based on four pillars: Security, Reliability, Performance efficiency, and Cost Optimization.
- Cloud computing helps achieve optimal server configuration by providing various features.
- You can configure Amazon CloudWatch, CloudTrail, and AWS Config to provide you with alerts and notifications.
- With hybrid technologies you can link your existing on-premises network configuration onto your AWS virtual private cloud.

6 **Which one is an example of a Push Model distributing load to multiple nodes?**

AWS Elastic Load Balancer

AWS SQS

Amazon Kinesis

Storage Gateway

- AWS Well-Architected Framework helps you to understand the pros and cons of decisions you make while building systems on AWS.
- The AWS Well-Architected Framework is based on four pillars: Security, Reliability, Performance efficiency, and Cost Optimization.
- Cloud computing helps achieve optimal server configuration by providing various features.
- You can configure Amazon CloudWatch, CloudTrail, and AWS Config to provide you with alerts and notifications.
- With hybrid technologies you can link your existing on-premises network configuration onto your AWS virtual private cloud.

6 **Which one is an example of a Push Model distributing load to multiple nodes?**

AWS Elastic Load Balancer

AWS SQS

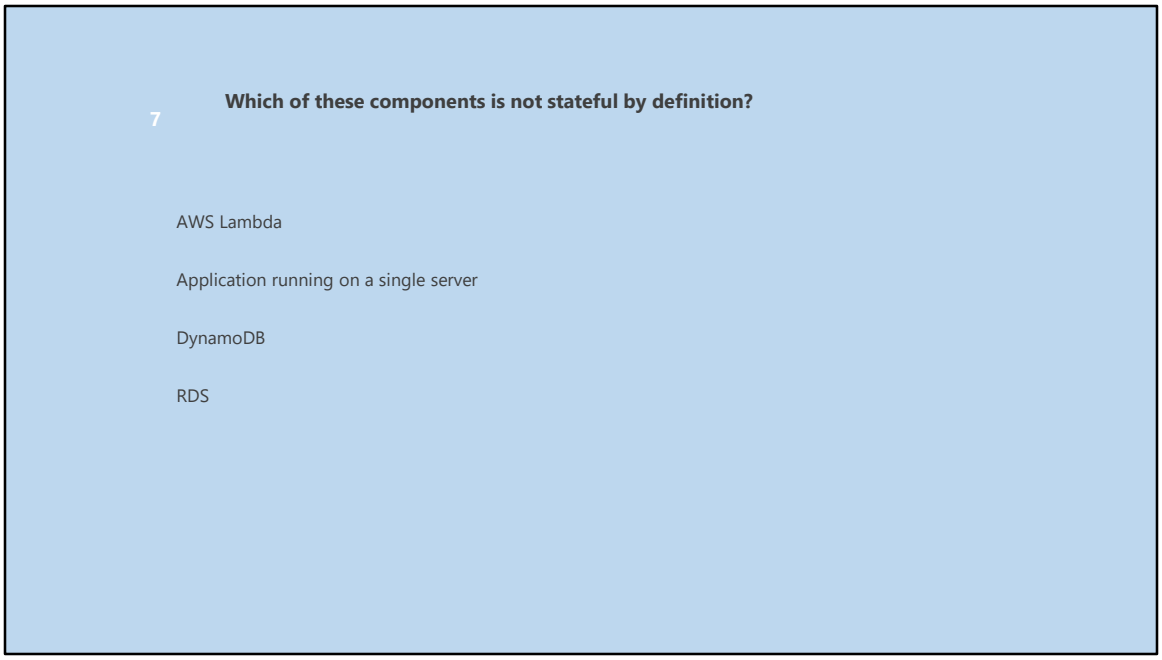
Amazon Kinesis

Storage Gateway

a

A load balancer, such as the AWS Elastic Load Balancer, is a popular way to distribute a workload across multiple resources.

- AWS Well-Architected Framework helps you to understand the pros and cons of decisions you make while building systems on AWS.
- The AWS Well-Architected Framework is based on four pillars: Security, Reliability, Performance efficiency, and Cost Optimization.
- Cloud computing helps achieve optimal server configuration by providing various features.
- You can configure Amazon CloudWatch, CloudTrail, and AWS Config to provide you with alerts and notifications.
- With hybrid technologies you can link your existing on-premises network configuration onto your AWS virtual private cloud.



- AWS Well-Architected Framework helps you to understand the pros and cons of decisions you make while building systems on AWS.
- The AWS Well-Architected Framework is based on four pillars: Security, Reliability, Performance efficiency, and Cost Optimization.
- Cloud computing helps achieve optimal server configuration by providing various features.
- You can configure Amazon CloudWatch, CloudTrail, and AWS Config to provide you with alerts and notifications.
- With hybrid technologies you can link your existing on-premises network configuration onto your AWS virtual private cloud.

7 **Which of these components is not stateful by definition?**

AWS Lambda

Application running on a single server

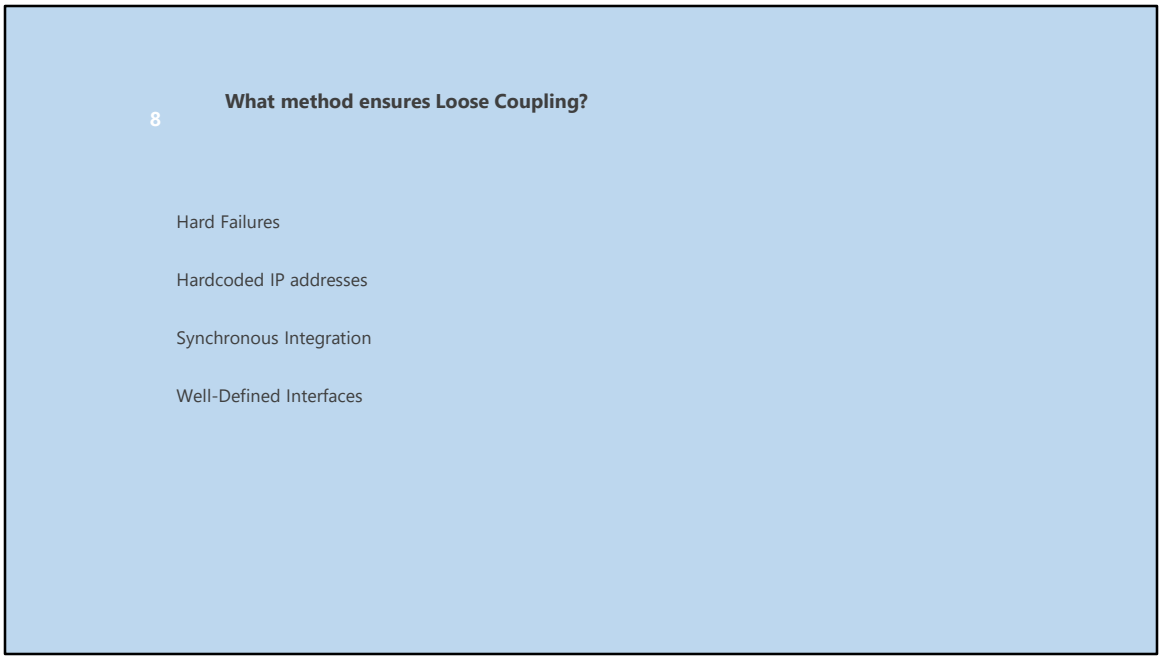
DynamoDB

RDS

a

Databases are stateful by definition as they store and retain data. Lambda uses a stateless programming model.

- AWS Well-Architected Framework helps you to understand the pros and cons of decisions you make while building systems on AWS.
- The AWS Well-Architected Framework is based on four pillars: Security, Reliability, Performance efficiency, and Cost Optimization.
- Cloud computing helps achieve optimal server configuration by providing various features.
- You can configure Amazon CloudWatch, CloudTrail, and AWS Config to provide you with alerts and notifications.
- With hybrid technologies you can link your existing on-premises network configuration onto your AWS virtual private cloud.



- AWS Well-Architected Framework helps you to understand the pros and cons of decisions you make while building systems on AWS.
- The AWS Well-Architected Framework is based on four pillars: Security, Reliability, Performance efficiency, and Cost Optimization.
- Cloud computing helps achieve optimal server configuration by providing various features.
- You can configure Amazon CloudWatch, CloudTrail, and AWS Config to provide you with alerts and notifications.
- With hybrid technologies you can link your existing on-premises network configuration onto your AWS virtual private cloud.

What method ensures Loose Coupling?

8

Hard Failures

Hardcoded IP addresses

Synchronous Integration

Well-Defined Interfaces

d

Ensure that all components only interact with each other through specific, technology-agnostic interfaces, for example RESTful APIs, will result in being able to modify resources without affecting other components.

- AWS Well-Architected Framework helps you to understand the pros and cons of decisions you make while building systems on AWS.
- The AWS Well-Architected Framework is based on four pillars: Security, Reliability, Performance efficiency, and Cost Optimization.
- Cloud computing helps achieve optimal server configuration by providing various features.
- You can configure Amazon CloudWatch, CloudTrail, and AWS Config to provide you with alerts and notifications.
- With hybrid technologies you can link your existing on-premises network configuration onto your AWS virtual private cloud.