Principles in Cloud Architecture Design - Performance

Cloud Infrastructure Engineering

**Nanyang Technological University
& Skills Union - 2022/2023**

Course Content

- Quick Check-In
- Dive into what Performance means and the best practices in terms of Performance
- Build architectures that have strong performance features

Time	What	How or Why
7:15pm - 7:40pm	Part 1 - Presentation	Performance Principles
7:40pm - 8:00pm	Part 2 - Presentation	Best Practices for Performance
8:00pm - 8:10pm	Break	
8:10pm - 8:30pm	Part 3 - Presentation	Best Practices for Performance
8:30pm - 8:50pm	Part 4 - Activity	
8:50pm - 10:00pm	Summary & Assignments	

Recap

- Reliability Design Principles
    - Automatically recover from failure, Test recovery procedures, Scale horizontally to increase aggregate workload availability, Stop guessing capacity, Manage change in automation
- Best Practices
    - Strong Foundations, Workload Architecture, Change Management, Failure Management

Self Study Check-In

Q1) How would you deploy your application such that you can achieve the best performance?

Overview

Overview

This module focuses on applying the principles of the **performance efficiency** pillar to your workloads.

In traditional, on-premises environments, achieving high and lasting performance is challenging.

Using the principles in this module will help you **build architectures on AWS that efficiently deliver sustained performance over time**.

Overview

This module is intended for those in technology roles, such as chief technology officers (CTOs), architects, developers, and operations team members.

After this, you'll understand AWS best practices and strategies to use when designing a performant cloud architecture.

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

Performance Principles

Performance Principles Summary

- Democratize advanced technologies
- Go global in minutes
- Use serverless architectures
- Experiment more often
- Consider mechanical sympathy

Democratize Advanced Technologies

Make advanced technology implementation **easier for your team by delegating complex tasks to your cloud vendor**.

Rather than asking your IT team to learn about hosting and running a new technology, **consider consuming the technology as a service.**

Democratize Advanced Technologies

For example, NoSQL databases, media transcoding, and machine learning are all technologies that require specialized expertise.

In the cloud, these technologies become **services that your team can consume**, allowing your team to focus on product development rather than resource provisioning and management.
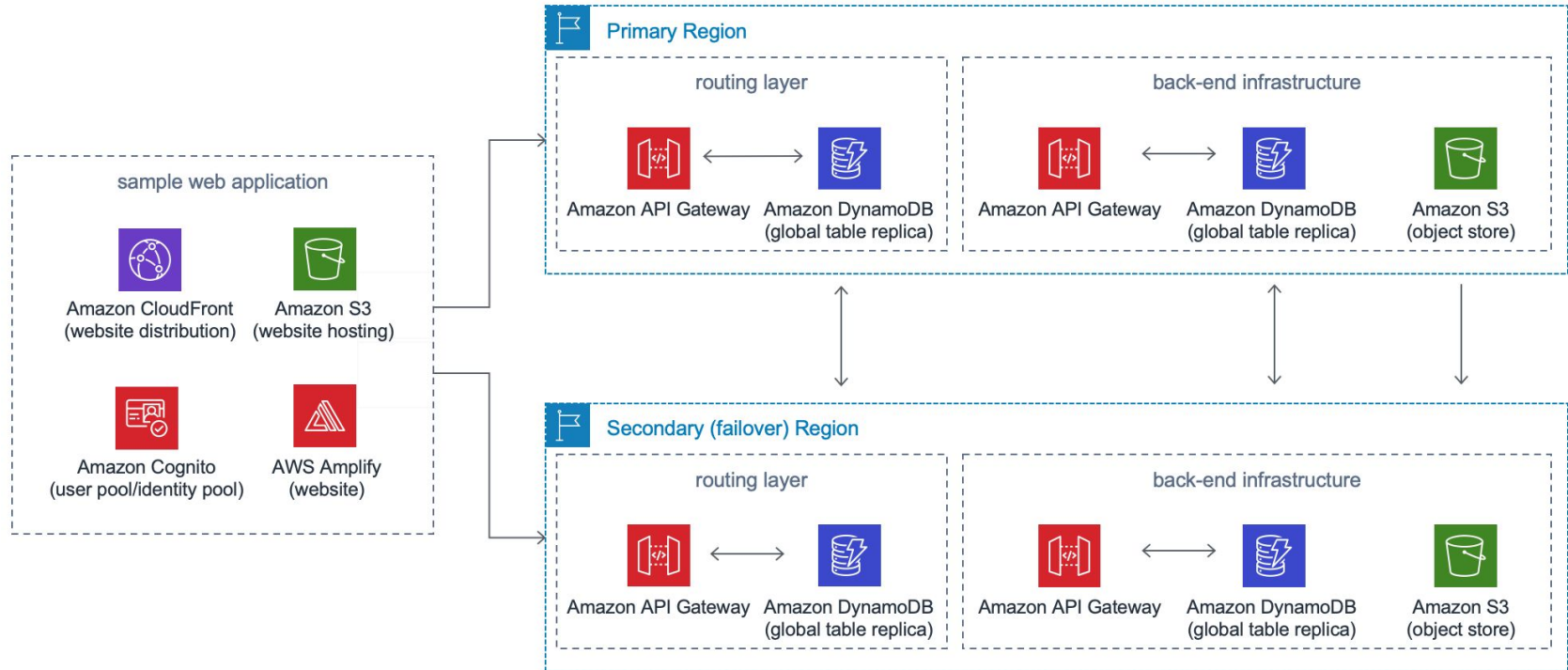
Democratize Advanced Technologies
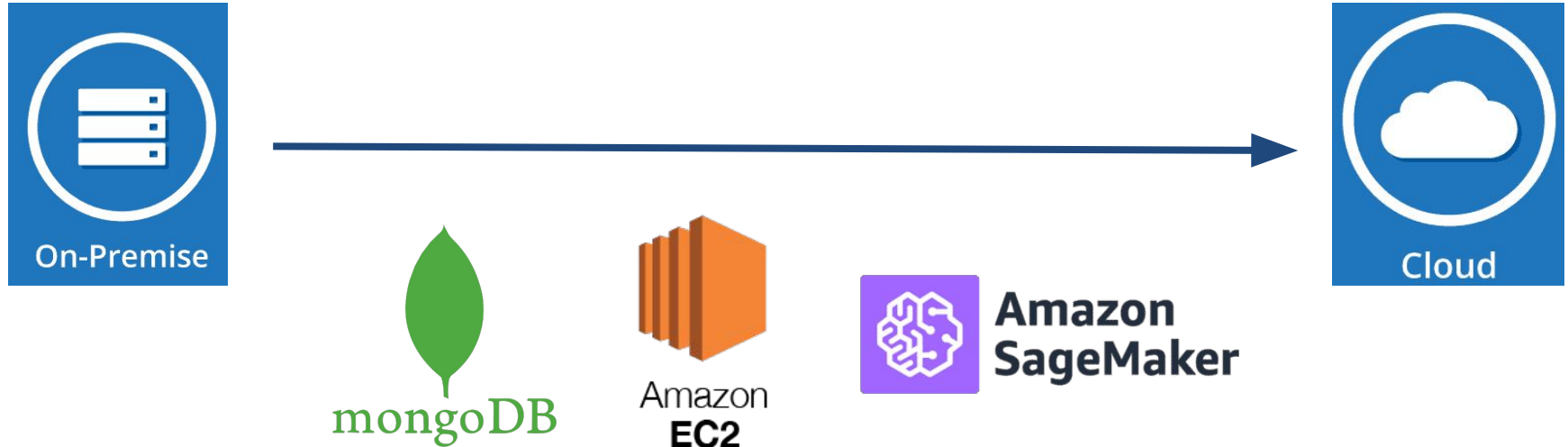
Go Global In Minutes

Deploying your workload in **multiple AWS Regions** around the world allows you to provide **lower latency** and a **better experience** for your customers at minimal cost.
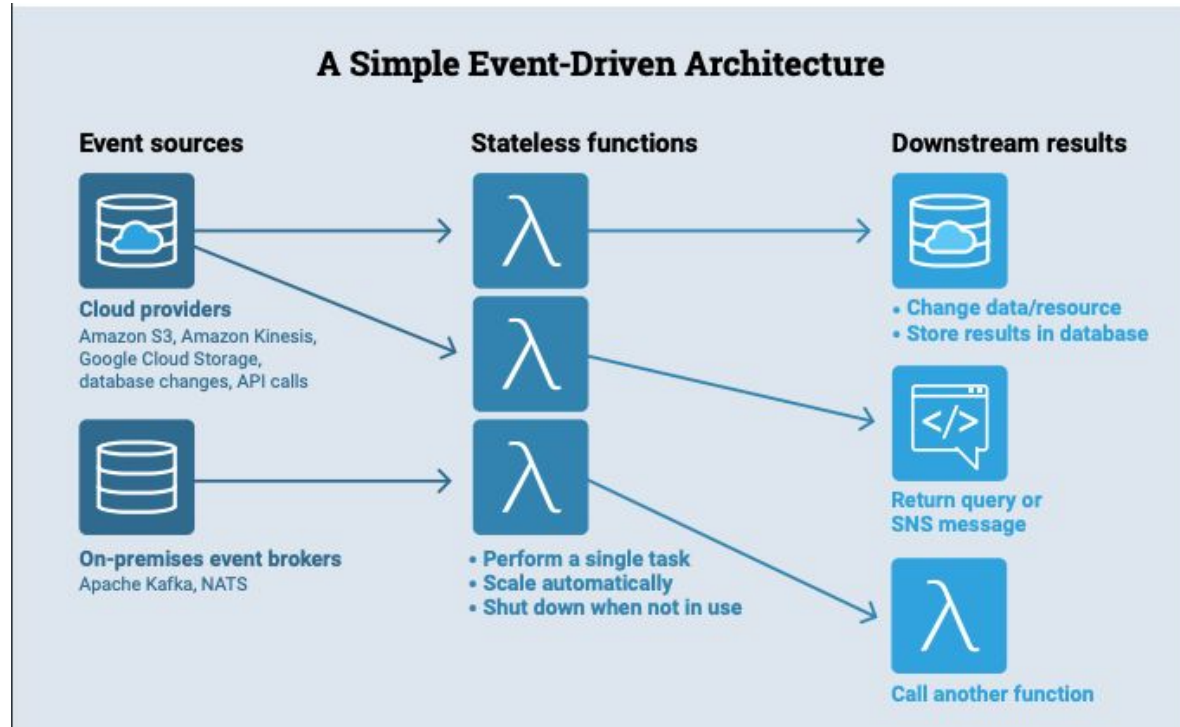
Go Global In Minutes

Use Serverless Architecture

Serverless architectures **remove the need for you to run and maintain physical servers** for traditional compute activities.

For example, serverless storage services can act as **static websites** (removing the need for web servers) and event services can **host code.**

This **removes the operational burden** of managing physical servers, and can lower transaction costs because managed services operate at cloud scale.

Use Serverless Architecture

A Simple Event-Driven Architecture

Event sources

Cloud providers
Amazon S3, Amazon Kinesis,
Google Cloud Storage,
database changes, API calls

On-premises event brokers
Apache Kafka, NATS

Stateless functions

- Perform a single task
- Scale automatically
- Shut down when not in use

Downstream results

- Change data/resource
- Store results in database

Return query or
SNS message

Call another function

Experiment More Often

With virtual and automatable resources, you can quickly **carry out comparative testing using different types of instances, storage, or configurations**.

Experiment More Often
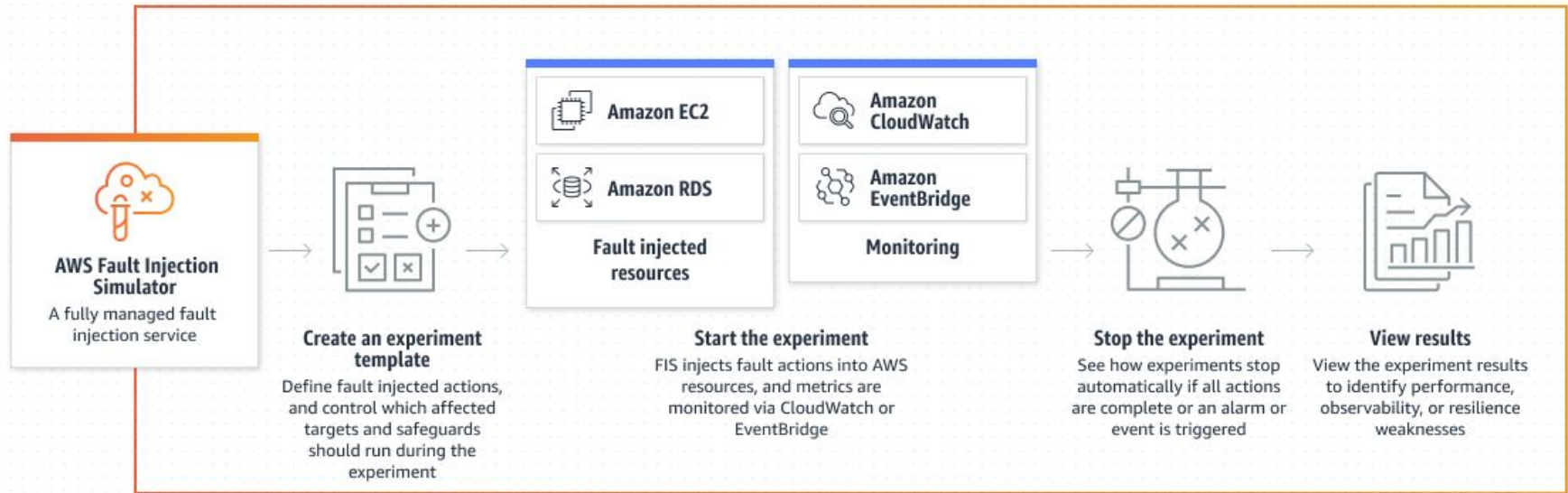
Consider Mechanical Sympathy

Understand how **cloud services are consumed** and always **use the technology approach that aligns best with your workload goals**.

For example, consider data access patterns when you select database or storage approaches.

Performance Principles Summary

- Democratize advanced technologies
- Go global in minutes
- Use serverless architectures
- Experiment more often
- Consider mechanical sympathy

Activity - Understanding Performance

In this activity, gather into your own group and each group should take on one or two research problem.

Ensure all research problems are taken and presented by the end of this section.

Activity – Understanding Performance

- Democratize advanced technologies
- Go global in minutes
- Use serverless architectures
- Experiment more often
- Consider mechanical sympathy

Best Practices For Performance

Summary

There are four best practice areas for performance in the cloud:

- Selection
- Review
- Monitoring
- Tradeoffs

Selection

The optimal solution for a particular workload varies, and solutions often combine multiple approaches.

Well-architected workloads use multiple solutions and enable different features to improve performance.

Selection

AWS resources are **available in many types and configurations**, which makes it easier to find an approach that closely matches your workload needs.

You can also find options that are not easily achievable with on-premises infrastructure. For example, a managed service such as **Amazon DynamoDB provides a fully managed NoSQL database with single-digit millisecond latency at any scale**.

Selection

Use a **data-driven approach** to select the patterns and implementation for your architecture and achieve a cost effective solution.

Your architecture will likely combine a number of different architectural approaches (for example, event-driven, ETL, or pipeline).

The implementation of your architecture will use the AWS services that are **specific to the optimization of your architecture's performance**.

Think mainly around four main resource types - **compute, storage, database, and network**.

Review

Cloud technologies are **rapidly evolving** and you must ensure that workload components are **using the latest technologies and approaches to continually improve performance**.

You must continually **evaluate and consider changes** to your workload components to ensure you are meeting its **performance and cost objectives**.

Review

New technologies, such as machine learning and artificial intelligence (AI), can allow you to re-imagine customer experiences and innovate across all of your business workloads.

Take advantage of the continual innovation at AWS driven by customer need.

AWS releases new Regions, edge locations, services, and features regularly.

Any of these releases could **positively improve the performance efficiency** of your architecture.

Monitoring

After you implement your workload, you must monitor its performance so that you can remediate any issues before they impact your customers.

Monitoring metrics should be used to raise alarms when thresholds are breached.

Monitoring

Amazon CloudWatch is a monitoring and observability service that provides you with data and actionable insights to monitor your workload, respond to system-wide performance changes, optimize resource utilization, and get a unified view of operational health.
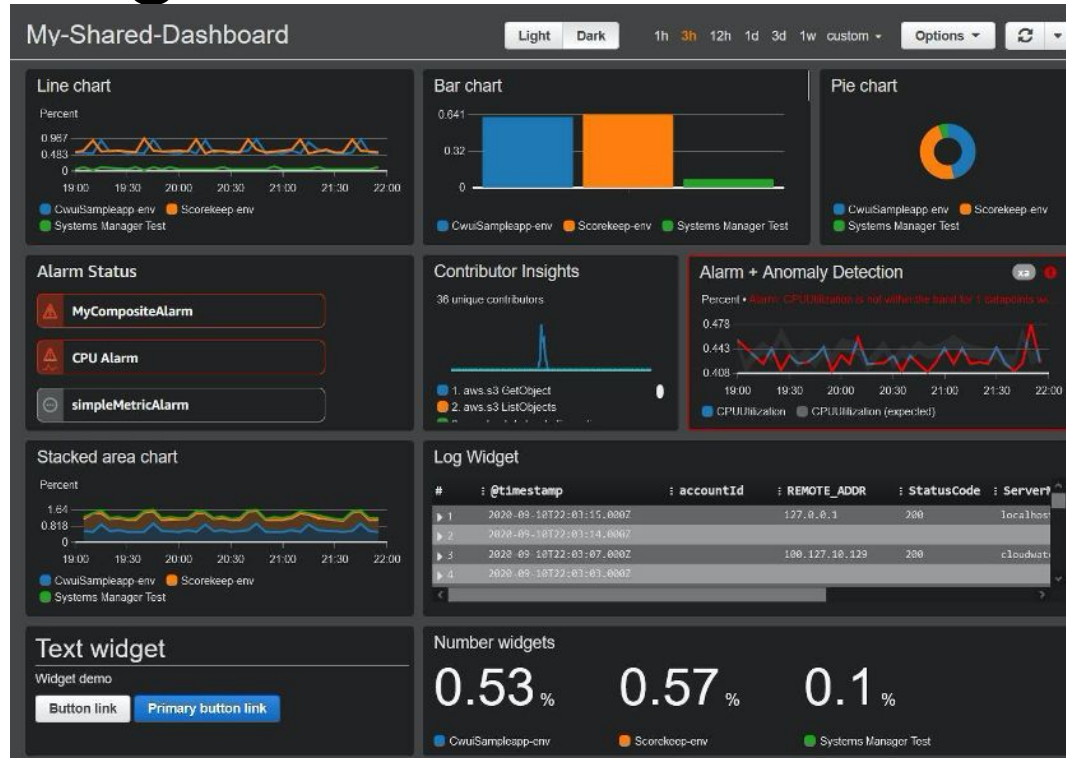
Monitoring

CloudWatch **collects monitoring and operational data** in the form of **logs, metrics, and events from workloads** that run on AWS and on-premises servers.

AWS X-Ray helps developers **analyze and debug production, distributed applications**. With AWS X-Ray, you can glean insights into how your **application is performing** and **discover root causes** and **identify performance bottlenecks.**

You can use these insights to react quickly and keep your workload running smoothly.

Monitoring

Monitoring

NANYANG TECHNOLOGICAL UNIVERSITY | SINGAPORE

Tradeoffs

When you architect solutions, think about **tradeoffs** to **ensure an optimal approach.**

Depending on your situation, you could **trade consistency, durability, and space for time or latency, to deliver higher performance**.

Tradeoffs

Using AWS, you can go global in minutes and deploy resources in multiple locations across the globe to be closer to your end users.

You can also **dynamically add read-only replicas** to information stores (such as database systems) to **reduce the load on the primary database**.

Summary

There are four best practice areas for performance in the cloud:

- Selection
- Review
- Monitoring
- Tradeoffs

Activity - Understanding Performance

In this activity, gather into your own group and each group should take on one or two research problem.

Ensure all research problems are taken and presented by the end of this section.

Activity - Understanding Performance

- How do you select the best performing architecture?
- How do you monitor your resources to ensure they are performing?
- How do you evolve your workload to take advantage of new releases?
- How do you use tradeoffs to improve performance?

Activity

Learner:

- Clean up AWS.
- Remove/delete/terminate all service/ resources that created.

Instructor

- Clean up AWS.
- Remove/delete/terminate all service/ resources that created.
- Check the AWS account after learner clean up.

What's Next?