AWS REGIONS AND AVAILABILITY ZONES: AN INTRODUCTION

Amazon Web Services (AWS) pioneered the modern-day cloud computing IT service delivery model. Amazon's success with delivering cloud resources as a service comes down to the ease of accessibility and availability of the technology for end-users.

Over the years AWS has expanded rapidly, expanding the cloud service across different geographic locations around the world. In fact, the vast AWS data center footprint is a key value proposition that sets it apart from leading competitors like Microsoft Azure and Google Cloud. Amazon describes its cloud infrastructure locations as AWS Regions and Availability Zones.

This brief primer describes the terms AWS Regions and AWS Availability Zones and explores what it means for CIOs.

What is an AWS region?

AWS Region refers to the geographic location of the AWS [data center](#). The AWS cloud infrastructure currently spans across 18 locations globally, with plans to announce four new regions in Asia and Europe in the near future. Each AWS Region may offer different service quality in terms of latency, solutions portfolio, and cost based on its geographic location and distance from customer sites.

Learn about regions for more cloud providers in [Availability Regions & Zones for AWS, Azure & GCP](#).

What is an availability zone?

AWS Availability Zone refers to the isolated data center infrastructure within a single region. In total, AWS cloud spans 55 Availability zones, with plans to announce 12 more Availability Zones in the near future.

Multiple AWS Availability Zones are connected through low latency and high throughput networking channels to maximize service [availability and redundancy](#) across the single and multiple data center facilities within a single Region. Each Availability Zone typically includes multiple data centers. No single data center is shared between multiple Availability Zones.

What is AWS local region?

AWS Local Region refers to a single isolated AWS data center infrastructure designed to complement an existing AWS Region but at the same time is also isolated from that region. Currently AWS operates one AWS Local Region in Osaka, Japan, that contains one Availability Zone and is used in conjunction with the Asia Pacific (Tokyo) Region as described [here](#).

Up-to-date information on the location and count of AWS Regions, Availability Zones and AWS Local Region can be found on the AWS resource [here](#).

How to chose AWS Regions and Availability Zones

The choice for AWS Regions and Availability zones may come as a tradeoff for organizations based on a range of factors, including:

Latency & Proximity

The distance between the cloud deployments and end-users is a key factor that determines the latency and network performance of the cloud service. The performance is further affected when the cloud solution is integrated with on-premise legacy technologies and apps as part of a hybrid cloud strategy considering the varying data transfer and network performance capabilities. Finally, selecting the AWS Region closest to customer or end-user proximity ensures best user experience, usually as the least expensive option as compared to AWS Regions at distant geographic locations.

Amazon also offers services such as the [AWS Route 53](#) service to automatically direct global network traffic through the most optimal channels to maximize availability and performance.

Cost

The price across AWS Regions and Availability Zones varies due to the different [CapEx, OpEx](#), and regulations across different geographic locations. Organizations may need to identify the optimal tradeoff between the cost and other factors for choosing AWS Regions including service catalog, latency, network performance, and regulatory compliance limitations, among others.

AWS offers a simple [cost calculator](#) to estimate the expected costs of the service based on AWS Regions.

Service Catalog

Amazon offers a vast portfolio of cloud-based solutions spanning across AWS Regions. While the most popular AWS services are available across all AWS Regions, not every location offers all of the AWS product services. Follow the AWS Region Table to learn about the availability of AWS product services in every Region.

The decision for an AWS Region should be based on the current and future AWS needs of the IT workloads, as any change in workload requirements may necessitate investments toward AWS services from additional AWS Regions.

Regulatory Compliance & Security

Tightly regulated industries may be required to process sensitive end-user in specific geographic locations. Cloud computing makes it easy for organizations to transfer, store, and process information in data centers at distant locations, which may be considered as a violation of compliance regulations, which may lead to costly lawsuits and damages to the brand reputation. Similarly, organizations may also be obliged to distribute workloads across multiple geographically disparate cloud data centers to ensure high availability and security standards of sensitive business information and IT-enabled services.

Service Level Agreements

In order to meet the desired standards of IT service availability and performance, AWS commits to Service Level Agreements (SLAs) for parameters including service performance, uptime and support, among others. The cost structure of SLA standards may also vary across different AWS Regions based on the availability, performance, and product requirements of AWS customers.

This AWS documentation outlines the limitations for SLA commitment that must be considered when selecting AWS Regions and Availability Zones. In event of an outage, the rebate is offered only for the service outage beyond the SLA, so it is up to customers to introduce redundancies and maximize the service uptime capabilities by maintaining a cloud architecture distributed across multiple AWS Availability Zones and Regions.

The Green Factor

Many organizations are pushing to achieve carbon neutrality and adopting environmentally friendly business practices. Amazon is no different as it aims to operate a net carbon neutral global infrastructure by contributing toward various renewable energy and data center projects as described here. In effect, this means that only a proportion of AWS Regions are actually powered by renewable energy sources.

The location of AWS data centers is therefore one criteria to consider in your annual sustainability report that could serve as a competitive differentiation toward your sustainability efforts and making this world a better place.

Considering your cloud strategy

Fortunately, AWS offers a range with AWS Regions and Availability Zones that can fit well in your IT

strategy whether you focus on service uptime, product portfolio, availability, compliance or even the sustainability factors. However, AWS may not always be the right answer as DropBox figured out for their IT strategy.

Regardless, CIOs need to make well-informed decisions pertaining to AWS Regions, Availability Zones, and the cloud service at large, with considerations for both the near-term and long-term implications of their IT investments.

Additional resources

For more on this topic, check out these resources:

- BMC Multi-Cloud Blog
- 3 Essential Steps for Migrating to AWS, part of our AWS Guide
- Getting Started With a Multi-Cloud Strategy
- Key Facets of a Smart Cloud Migration Strategy