

AVAILABILITY REGIONS AND ZONES FOR AWS, AZURE & GCP



Everyone's upgrading to [cloud computing](#). Cloud resources, of course, are still physical devices located somewhere on planet Earth. So even though your resources are stored in the cloud, you will still have to choose *where* they will be used. Selecting your Region and Availability Zone are about the first option when setting up any cloud resource.

What are availability regions and zones?

Availability regions are the geographic locations of the cloud data centers. Different regions offer different service qualities in terms of latency, solutions portfolios, and costs. For the large providers, their availability regions exist across the globe.

The **availability zone** refers to an isolated [data center](#) within a single region. Each availability zone includes multiple data centers, and no single data center is shared between multiple availability zones.

Of the three major providers, [Amazon's regions and zones](#) are the most developed. Microsoft is comparable. Google is the newest to the scene, but it really isn't that far behind (and certainly isn't far behind in any way that makes it less of a viable choice). Under most circumstances, each provider covers the same major areas:

- China
- North America
- Southeast Asia
- East Asia
- Europe

There are only a few exceptions where Azure and AWS cover different locations, such as South Africa, that Google has yet to reach.

Amazon Web Services

Here are the AWS regions [around the globe](#):



Azure regions and zones

Azure covers [these regions](#):



Google Cloud Provider

[GCP offers](#) 24 regions with 73 total zones. There are three zones per region, with the US-Central1 region having 4 zones.



Comparing availability zones

Each region has availability zones, and each zone has its own data center. Each data center has its own hardware. Regions can be known for being good at some things and bad at others. For example, AWS East is known for having more downtime than AWS West. The reason for this comes down to the hardware at and usage of an availability zone. The data centers in each zone can consist of different hardware.

There are two key features among zones:

1. The number of people using it compared to how many people the zone supports
2. The available hardware in the zone

Based on these two things, the user can gauge their experience on:

- Zone downtime
- Zone latency
- Resource availability (i.e.; a task needs to run on high-memory SSDs)

An example of the resources available in one zone versus another on GCP are:

Zone: South America East1-B

southamerica-east1-b

- Available [CPU platforms](#):
 - Intel Xeon E5 v4 (Broadwell) platform (default)
 - Intel Xeon (Skylake) platform
- [E2 machine types](#) up to 16 vCPUs and 128 GB of memory
- [N1 machine types](#) with up to 96 vCPUs when using the Skylake platform
- [M1 ultramem memory-optimized machine types](#) with up to 160 vCPUs and 3.75 TB of memory
- [Local SSDs](#)
- [Sole-tenant nodes](#)

Zone: Europe West4-C

europa-west4-c

- Available [CPU platforms](#):
 - Intel Xeon E5 v4 (Broadwell) platform
 - Intel Xeon (Skylake) platform (default)
 - Intel Xeon (Cascade Lake) platform
 - AMD EPYC Rome platform (default for N2D machines)
- [E2 machine types](#) up to 16 vCPUs and 128 GB of memory
- [N2 machine types](#) with up to 80 vCPUs and 640 GB of memory
- [N2D machine types](#) with up to 224 vCPUs and 896 GB of memory
- [N1 machine types](#) with up to 96 vCPUs when using the Skylake platform
- [M2 ultramem machine types](#) with up to 416 vCPUs and 11.5 TB of memory
- [M1 ultramem and megamem memory-optimized machine types](#) with up to 160 vCPUs and 3.75 TB of memory
- [C2 machine types](#) with up to 60 vCPUs and 240 GB of memory
- [Local SSDs](#)
- [GPUs](#)
- [Sole-tenant nodes](#)

Global vs regional vs zonal resources

The length of the list says nothing, but the resources themselves say a lot more. Some items are standard among all zones, some resources are old and get phased out, some are new, advanced features only available in some zones.

When designing your [cloud infrastructure](#), you will want to define what tasks need to be performed where. By keeping your computation local and doing as little cross-regional operations as possible, you isolate your system from hardware and infrastructure failures.

According to GCP: "Resources that live in a zone, such as [virtual machine instances](#) or zonal [persistent disks](#), are referred to as zonal resources. Other resources, like [static external IP addresses](#), are regional. Regional resources can be used by any resources in that region, regardless of zone, while zonal resources can only be used by other resources in the same zone.

"For example, to attach a zonal persistent disk to an instance, both resources must be in the same zone. Similarly, if you want to assign a static IP address to an instance, the instance must be in the same region as the static IP address."

You can see the different equipment available at each Availability Zone/ Region.

- [AWS](#)
- [Azure](#)
- [Google](#)

These subsections outline which resources apply at which level:

Global Resources

- Addresses: IP address list
- Images: Container images
- Snapshots: Persistent Disk Snapshots
- Instance Templates: Templates to create a VM
- Cloud Interconnects: On-Premise Network to Cloud Network connection
- Cloud Interconnect Locations: Physical connection point to Cloud Network
- VPC Network: A VPC Network
- Firewalls: Firewalls are on a single VPC and packets reach them from other networks
- Routes: Network Routes
- Global Operations: Operations can be of any type

Regional Resources

- Addresses
- Cloud Interconnect Attachments
- Subnets
- Regional Managed Instance Groups
- Regional Persistent Disks
- Regional Operations

Zonal Resources

- Instances
- Persistent Disks
- Machine Types
- Zonal Managed Instance Groups

- Per-zone Operations

Choosing an availability zone

Before choosing a provider, you will need to look at where your company does business and how it does business. A few questions to consider when selecting a cloud provider are:

- Where does your company do business?
- Can data be stored in a single place for remote offices or must data be shared among offices across regions?
- Will data need to get passed from one zone to another?
- Are data retrieval or computation times important?

Finally, when choosing which zone will be best for you, consider these [four criteria](#) to consider:

1. Latency and proximity

General rule of thumb—opt for the closest zone for lower latency. Check Stack Overflow and other forums to see what others might be saying about any zone's latency being better than another.

2. Cost

This is a task for the accountants. The differences between each are in the tenths of a penny, and the resources vary from provider to provider. Meaning, you're not always comparing apples to apples.

[Here's a brief comparison](#) for Google prices vs AWS prices for CPU and storage, for instance:

Instance Parameters	Google Cloud Per-Hour Price	AWS Per-Hour Price
On-Demand / Linux / General Purpose / 2 CPUs (Memory: AWS 8 GB / Google Cloud 7.5 GB)	\$0.107	\$0.100
On-Demand / Linux / Compute Optimized / 2 CPUs (Memory: AWS 3.75 GB / Google Cloud 1.8 GB)	\$0.813	\$0.100
On-Demand / Linux / Memory Optimized / 2 CPUs (Memory: AWS 15.25 GB / Google Cloud 13 GB)	\$0.134	\$0.133

Instance Parameters	Google Cloud Per-Hour Price	AWS Per-Hour Price
Reserved 1 Year / Linux / General Purpose / 4 CPUs (Memory: AWS 16 GB / Google Cloud 15 GB)	\$0.128	\$0.123
Reserved 1 Year / Linux / Compute Optimized / 4 CPUs (Memory: AWS 8 GB / Google Cloud 3.6 GB)	\$0.095	\$0.107
Reserved 1 Year / Linux / Memory Optimized / 4 CPUs (Memory: AWS 32 GB / Google Cloud 26 GB)	\$0.159	\$0.159

Here is a flat comparison of storage prices per GB across the Amazon S3 and Google Cloud Storage:

Storage Parameters	Google Cloud Per-GB Price	AWS Per-GB Price
Frequent Access / First 50 TB	\$0.026	\$0.230
Frequent Access / 51-500 TB	\$0.026	\$0.220
Infrequent Access	\$0.010	\$0.0125
Archive Storage	\$0.070	\$0.0040

3. Regulatory compliance and security

Each zone exists in a different country, and each country has different laws regarding data safety and protection. Some regions may prohibit the transfer of data between regions, which could affect how you design your infrastructure. There can be significant penalties for breaking these data compliance laws.

4. Service level agreements

You'll want the right parameters for better service. Check the [service level agreement \(SLA\)](#) for each provider.

For example, [AWS' General Service Commitment](#) for Compute on EC2 instances: "AWS will use commercially reasonable efforts to make the Included Services each available for each AWS region with a Monthly Uptime Percentage of at least 99.99%, in each case during any monthly billing cycle (the "Service Commitment"). In the event any of the Included Services do not meet the Service Commitment, you will be eligible to receive a Service Credit as described below."

Other service level agreements available here:

- [AWS](#)
- [Azure](#)

- [GCP](#)

Additional resources

For more on navigating cloud complexity, browse the [BMC Multi-Cloud Blog](#) or read these articles:

- [SaaS vs PaaS vs IaaS: What's The Difference and How To Choose](#)
- [Why Migrate to the Cloud](#)
- [What are the Hidden Costs of Cloud Adoption?](#)
- [Change Management in the Cloud](#)
- [Power Outages at Public Cloud Data Centers: How To Mitigate Risks](#)
- [What is a Hyperscale Data Center?](#)