Microsoft Azure Global Infrastructure

Introduction

Microsoft Azure is a cloud platform that ranks among the global leading platforms with a huge international infrastructure and industry-level services. To address the rising demand of reliable, scalable and resilient cloud computing, Azure has developed a very distributed platform that entails Geographies, Azure Regions, Availability Zones and Data Centers.

This layered infrastructure gives organizations across industry or any region the ability to have low latency and high availability of their applications. From AI optimized areas to energy-sustainable data centers, Azure keeps pushing its infrastructure to accommodate the ever-increasing requirements of the modern enterprises.

# Geographies

An Azure Geography is a discrete market which typically contains one or more regions such as the United states, India or the European union and the purpose of these geographies is to ensure data residency, sovereignty, compliance, and resiliency.

## **Key Characteristics**

* **Data Residency & Compliance Preservation:-**

Geographies guarantee that the customer information does not breach the boundaries defined by politics or law.

* **Effective fault tolerant design:-**

Inside each geography, there are a number of regions that utilise the Azure backbone to provide disaster recovery and cross-region replication.

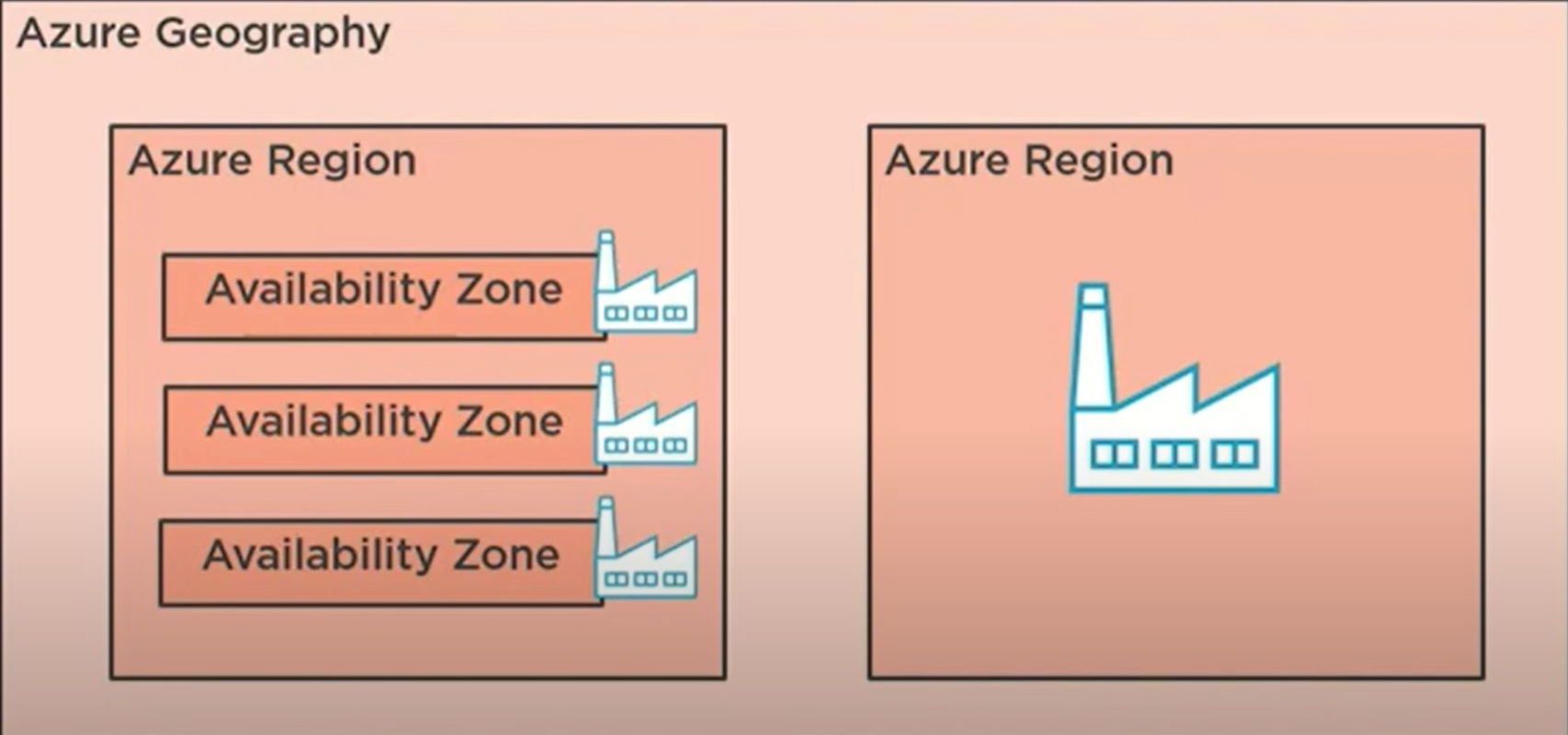
* **Customer-Focused structure:-**

It is designed according to the industry and government standards in various countries.

## **Structure**

Every single Azure Geography:

* Has two or more regions.
* Use of regional pair (It refers to pair one Azure region with another allowing replication) architecture to get high availability and fault tolerance.
* It contains regions with availability and non availability zones.
* Is associated with fast paced networking infrastructure.



## **Illustration**

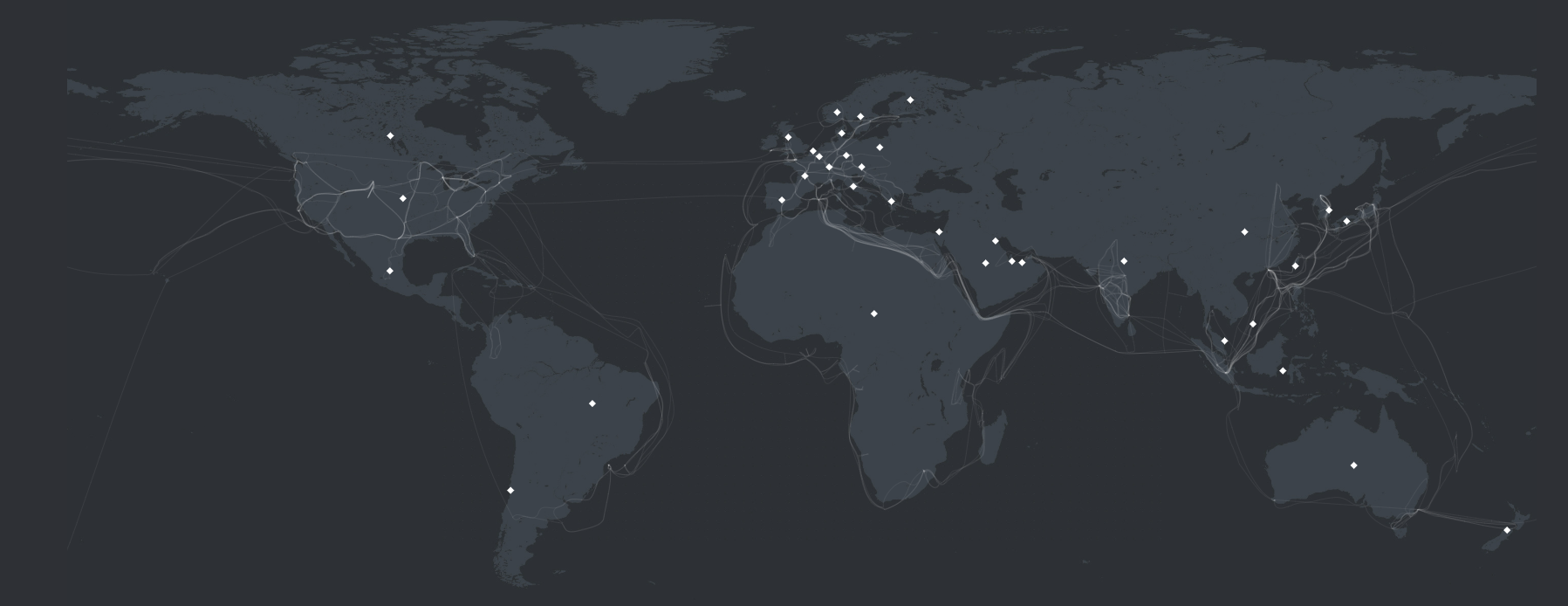
In the left side, there is displayed an Azure Region that is constituted of various Availability Zones. Every zone is an independent datacenter that has its own power, cooling, and networking. Such zones enable high availability but the application and data of one zone can be replicated in other zones in the same region.

Therefore, On the right side it indicates an area that is not composed of Availability Zones. There are some regions constructed as part of a single-zone, meaning there is no zone based separation but they have data centers.

These two kinds of regions are located in the same Azure Geography and therefore they abide by the same rule of compliance and data residence despite the fact that they can be internally not identical.

## **List of geographies around the world**

| **Geography** | **Representative Countries** |
| --- | --- |
| United States | USA(Multiple regions: East US, West US…) |
| Canada | Canada Central, Canada East |
| Korea | Korea Central, Korea South |
| Brazil | Brazil Southeast |
| Mexico | Mexico Central |
| Europe | Austria, France, Germany, Ireland, Netherlands, Norway, Poland, Switzerland, UK |
| United kingdom | UK South, UK West |
| India | Central India, South India, west India |
| China(21 Vianet) | Regions are operated independently for regulatory needs |
| Japan | Japan East, Japan West |
| Australia | Australia East, Southeast, Central |
| Southeast Asia | Singapore, Malaysia |
| Middle East | UAE North, UAE Central, Qatar |
| South Africa | South Africa North, South Africa West |
| Sweden | Sweden Central |
| Israel | Israel Central |



**Map Illustration**

This world map visualizes the Microsoft Azure infrastructure with the help of white diamond icons each representing one Azure Geography. A geography refers to a particular market (Asia Pacific or Europe) and a geography consists of one or more Azure regions. They are geographically located to provide data residency, compliance and disaster recovery needs.

There are five major geographies of azure in which azure is classified as:

* Americas
* Europe
* Asia Pacific
* Middle East
* Africa

Every white marker denotes a location in which Azure runs services in a compliant and secure perimeter. The geographies are capable of high availability, regional pairing, and local reach of cloud services worldwide.

# Azure Regions

An Azure Region is a specific geographic location, consisting of one or more interconnected data centers, which can deliver cloud services, including compute, storage, networking, AI and a lot more services. These locations are spread across the globe and they enable their customers to position their resources near their users while also meeting compliance, data residency and performance requirements.

Microsoft represents the largest presence of cloud infrastructure in the world due to the fact it has announced more than 60 Azure regions which helps to provide cloud services with low latency and high availability.

## **Key Features**

## **Global Network Backbone:**

Each region in Azure is connected to all the other regions using microsoft private global fiber backbone network of more than 175,000 miles around the globe. The following is supported in this network:

* Ultra low latency transfer of data.
* There is secure and private inter-regional communication.
* The data is overlapped between regions which makes backup and recovery possible.

This ultra-fast infrastructure guarantees that the Azure services are distributed in all parts across the world but still interconnected in a smooth way, providing uniform results even in distant continents.

## **Security and Compliance:**

Azure Regions are constructed in multi level security controls which includes physical, network and operational controls. Microsoft guarantees regional infrastructure compiles with over 100 of certifications worldwide including industry specific standards such as:-

* GDPR
* HIPAA
* ISO 27001
* SOC 1, 2 and 3
* FedRAMP(Government of the U.S.)

Every area assists clients to comply with the data residency laws, such that sensitive information remains within set parameters and is only available under legal frameworks.

## **AI optimized Infrastructure:**

Some regions of Azure have AI- and high-performance computing (HPC) specific infrastructure. Such areas are:

* Regions of Azure include high-performance GPU clusters such as NVIDIA which helps to handle AI workloads.
* Quicker and efficient training of models which helps in low latency and helps provide efficient training.
* The regions offer scalable compute capabilities to provide inference, deep learning and additional machine learning workloads.
* There are AI-optimized regions and that are:

East US, South Central of the US, Sweden Central.

* Numerous Azure regions supports specialized services, such as Azure OpenAI, Azure Machine Learning, Azure Cognitive Services, and Azure Synapse Analytics, so more Azure regions are bringing advanced solutions to fill the near-user gap so organizations can bring more of the industry-defining AI applications, large language models, and data pipelines to their users.

These areas make possible huge AI developments, such as natural language processing, generative AI models, and real-time analytics.

## **Availability of products by Regions:**

Not all services are available in every region by Azure. The deployment of services depends on infrastructure, rules and user demand.

Examples:

| **Service** | **Available Regions(Examples)** |
| --- | --- |
| Azure OpenAI Service | East US, West Europe, South Central US |
| Azure Kubernetes Service (AKS) | East US, Central India, UK South, West Europe |
| Azure Virtual Desktop | Southeast Asia, Japan East, North Europe |
| Azure Confidential Computing | East US, North Europe, West Europe, Central US |

Before deployment, customers have to check the available services in the region of choice.

## **Azure Pricing:**

Prices of azure services per region are different according to:

* Domestic operating expenses (electricity, cooling and manpower).
* The exchange rates and the taxes.
* Availability of resources in the Region.

This allows users to have the best costs as they can choose the regions with different pricing that works better with their workloads.

**Pricing Difference:**

| **Region** | **Approx price per hour(USD)** |
| --- | --- |
| East US | $0.021 |
| Central India | $0.023 |
| UK South | $0.025 |
| Japan East | $0.028 |
| Brazil South | $0.032 |

Although the difference in prices seems minor within one hour, it diverges drastically within hundreds of VMs or services over an extended deployment.

## **Region Pairs:**

Each Azure region has a similar region with a similar designated region. The regional pair for a particular region should be at least 300 miles away which ensures that the same disaster cannot affect both the regions. These pairs of region are being used in:

* Storage and Backup.
* Recovery plannings for any future disaster.
* The data is being replicated between the pair.
* In case of Maintenance schedules.

Examples of Regional pairs:

| **Region** | **Paired Region** |
| --- | --- |
| East US | West US |
| North Europe | West Europe |
| Australia East | Australia Southeast |
| Central India | South India |

Regional pairing is required so that the services are available in case of an emergency that may be caused by a natural disaster or in the event of large-scale failure of the region.

**Special Azure Region:**

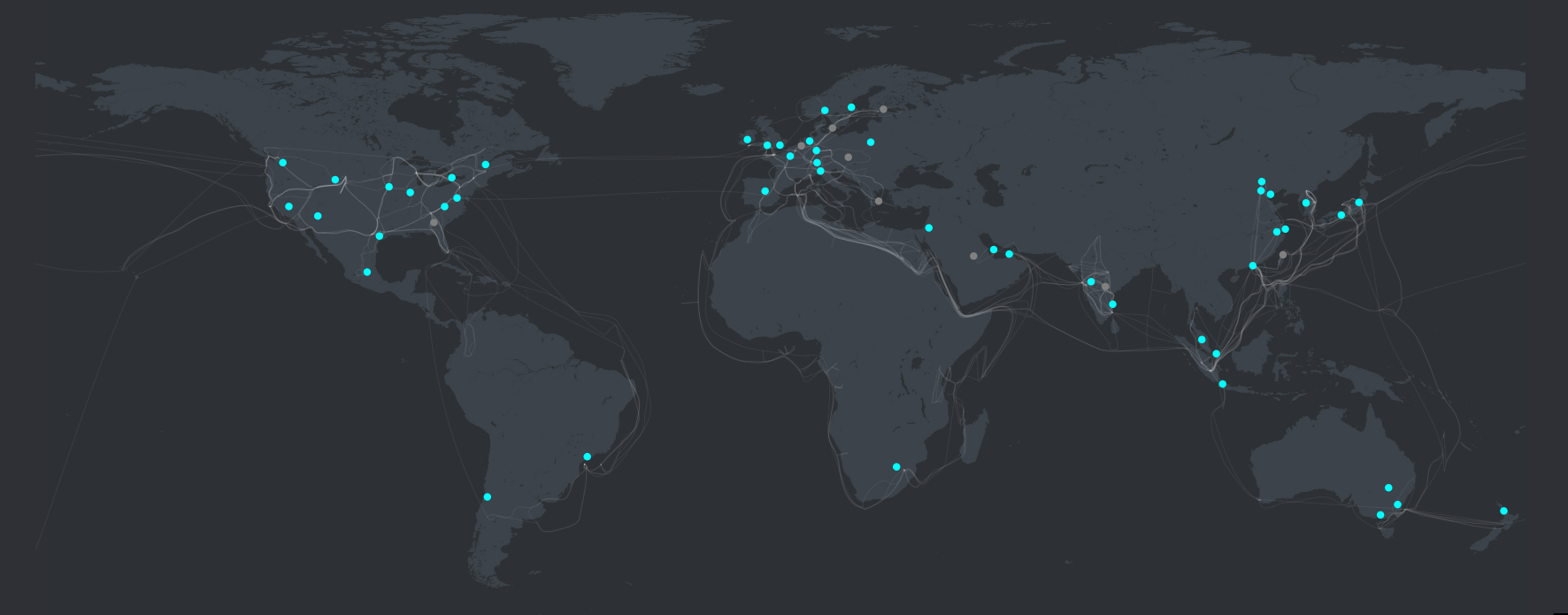
Azure has also come up with exclusive regions to fit specific legal or regulatory requirements of customers. In these areas, we obtain increased control, seclusion or obedience.

Kinds of special Azures Regions:

* Azure Government- With FedRAMP and DOD compliance it is made for US public sector
* Azure China (21Vianet)- Was run separately in order to satisfy Chinese legislation and rules.
* Azure Germany- It is built with Strict German data residency.

Such areas tend to be physically distant and keep distinct operation patterns to serve secure and sovereign workloads.

**Azure Regions in map:**



In this map:

Blue dots- It represents active azure regions.

Gray dots- It represents upcoming Azure regions.

# Availability Zones

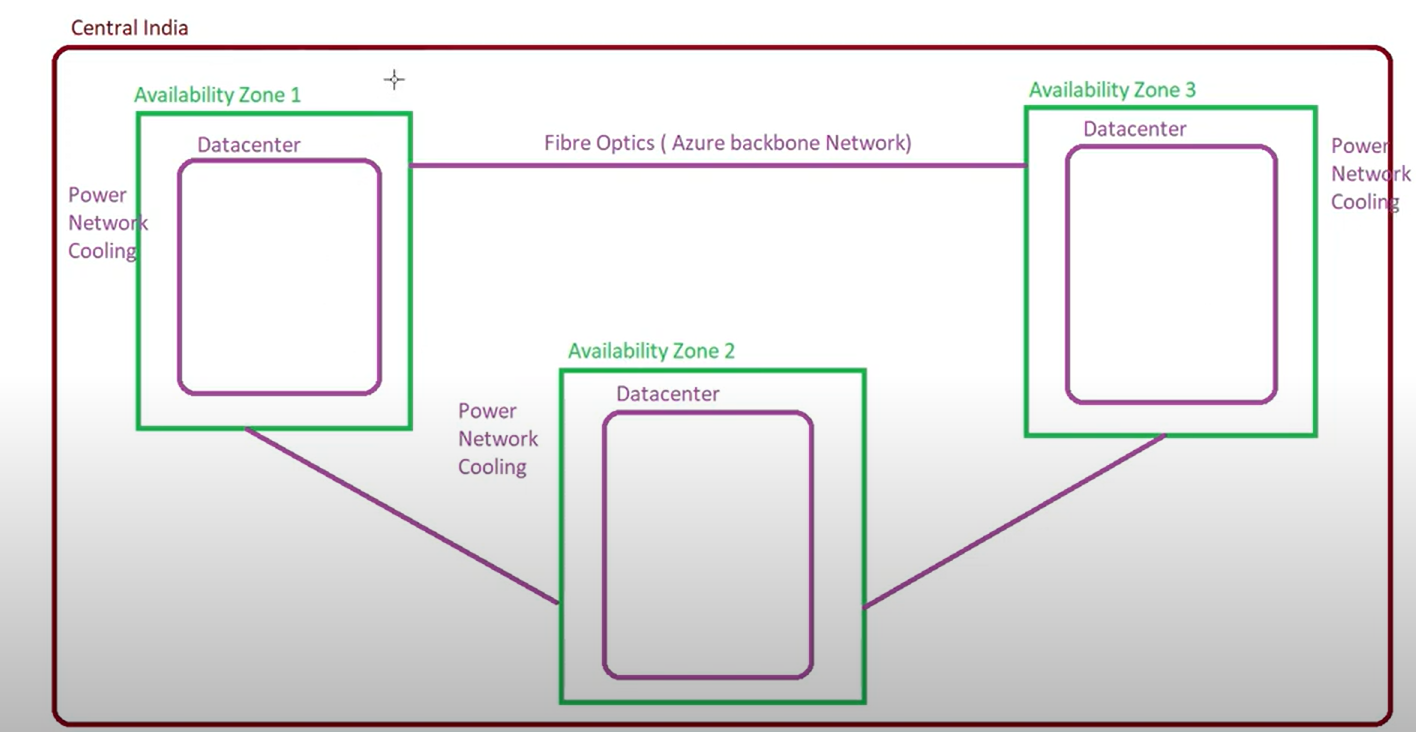
Availability zones are made up of data centres in an Azure region that are physically apart with each providing independent power, networking, and cooling. The zones are set up to guarantee high availability of applications as well as designed to provide fault isolation which means if one zone fails, the other zone will continue to operate.

There is high availability due to zones being interconnected using the private, high-speed fiber-optic backbone network that belongs to Microsoft. These areas are placed in strategic distances, to ensure that there is the least chance of failure that may be caused by natural catastrophe or power loss, and yet the peak performance in terms of latency is maintained.

**Key features**

* **High Availability:** Zone- redundant services are guaranteed having 99.99 percent uptime.
* **Redundancy:** It involves distribution of components into various zones.
* **Fault Isolation:** When one zone fails the others do not get any effect.
* **Low Latency:** Quick movement of information with the assistance of the private global network.
* **Business Continuity:** Applications will operate online in continuity even in failures.

**Structure**

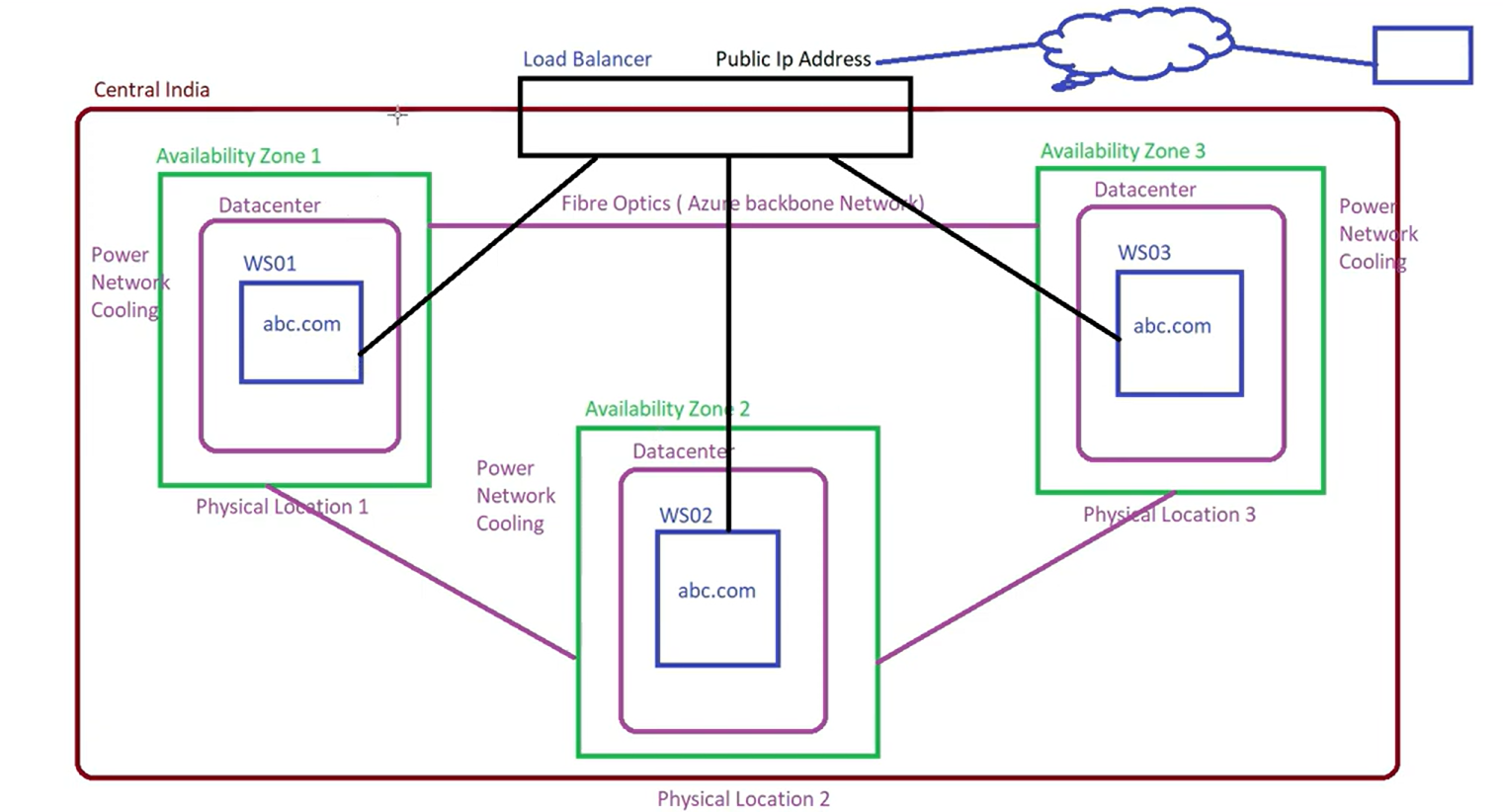


Availability Zones include:

* Data Centers
* Power supply
* Cooling
* Networking
* Fibre optic for azure backbone network

These zones are separated logically but they are physically close enough to provide fast synchronization with ensuring fault tolerance.

**Zone based Application deployment**



In this diagram, the means to reach high availability in Azure will be through placement of applications in Availability Zones in a single region, in this case, the Central India region.

**Diagram Illustration:**

1. **Three Availability Zones:**

* Each green box marked Availability Zone 1, 2 and 3 has its own Datacenter which has its own power, network and cooling infrastructure.
* The zones are deployed at different geographical locations (1, 2 and 3) to eliminate single points of failure.

1. **Application Deployment (abc.com):**

* The web app is the abc.com that is implemented on the three zones.
* The application is hosted on different web servers in every zone as follows: WS01, WS02, and WS03.
* The replicas guarantee the redundancy and load balance.

1. **Load Balancer:**

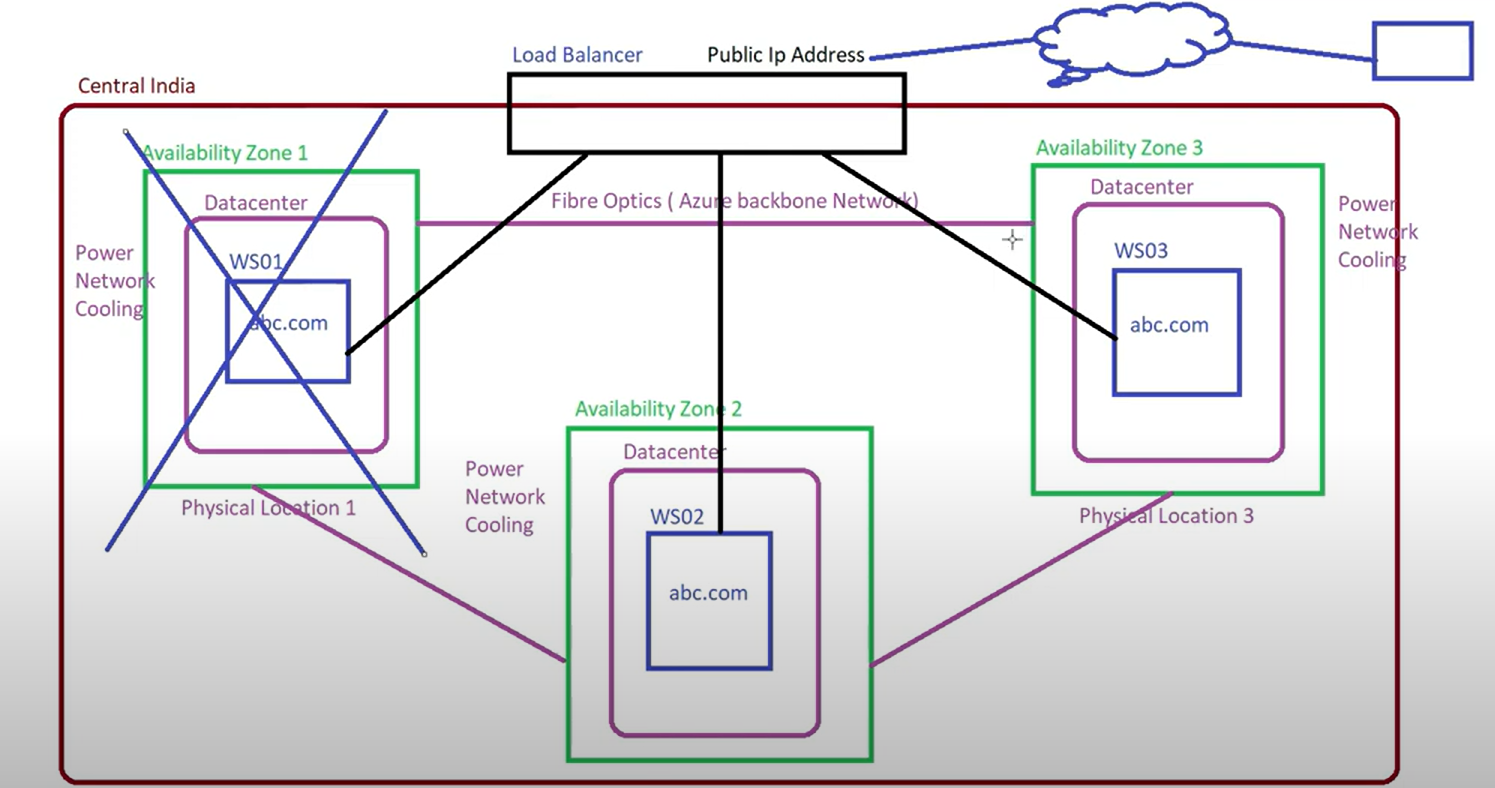
* In the top center, there is a Load Balancer that links to all the three zones.
* It possesses a Public IP Address with which external users communicate with.
* The Load Balancer directs the user traffic to one of the available replicas of the application, (WS01, WS02, WS03) on the basis of load, health and response time.

1. **Backbone: Fiber Optics:**

* These zones are linked through the integration of a low-latency fibre optic backbone of Azure that makes the internal communication fast.

**Failover Scenario**

If Availability Zone 1 has failed then the load balancer will redirect traffic to WS02 and WS03 in zones 2 and 3, with this there is no interruption and the app will remain live through its replicas.



# Data Centers

The Azure Data center forms the backbone of the Microsoft cloud platform that is developed to sustain high availability and security. There are 60+ Azure regions and each data center ensures that it is strategically placed, environmentally friendly as well as it should be protected with multi-layer security.

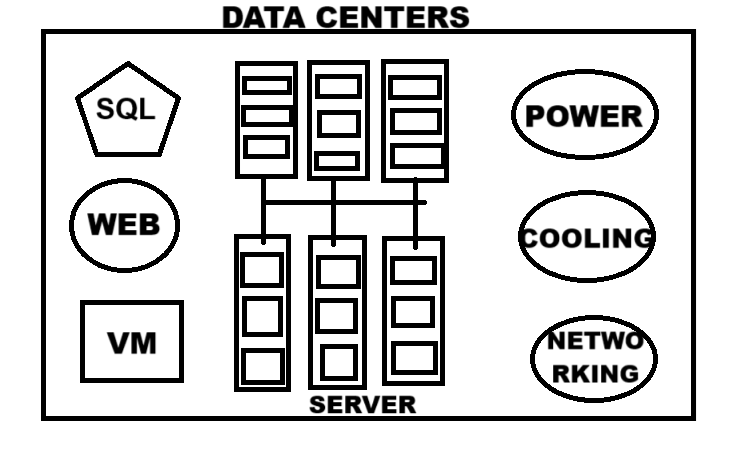
They are constructed to provide compute, storage and networking services in Azure with emphasis on redundancy, disaster recovery and low-latency access to global users.

# **Structure**

The Microsoft data centers have a very modular and layered architecture to provide redundancy and fault tolerance. They include:

* **Compute Layer:** It hosts applications, virtual machines (VMs) and SQL databases. These loads are taken by powerful physical servers.
* **Server Clusters:** Racks and Rows of servers helps to promote horizontal scaling. With this arrangement Microsoft can either increase or decrease capacity according to demand.
* **Power Systems:** Data centers are installed with their own uninterruptible power supplies (UPS) and backup diesel generators so that the operations remain continuous.
* **Cooling Systems:** The Air and water-based cooling systems ensure the best temperatures to the server environment to avoid overheating and breakdown of the hardware.
* **Networking:** It facilitates the smooth communication within racks due to fast and redundant network architecture and also provides connections to the outside world in other regions of the azure.

The illustration demonstrates the core layers of a Microsoft data center. It emphasizes the effectiveness of the various loads (SQL, Web, VMs) that are operated on servers. Scalable power, cooling and network are provided to these servers to guarantee 24X7 performance and availability.



**Key features**

* **Defense-in-Depth Security:**

Microsoft has a multi layered security system that has biometric locks, the use of perimeter fencing, smart card access and 24/7 surveillance. Logical security consists of: firewall, threat detection and guaranteeing end-to-end protection of the framework and information.

* **Threat & Vulnerability Risk Assessment(TVRA):**

To ensure that any risks that the company faces such as cyber threats, power disruptions or physical breakages, Microsoft has been carrying out regular TVRA to assess and reduce the effect of such risks. These assessments overall help to take any preventive measures before any failure or interruption occurs.

* **Protection of Data:**

Standard industry encrypted tools help to encrypt or protect data during both storage or transmission. Global compliance certification, such as ISO 27001 and GDPR helps Microsoft to provide high availability to its customers.

* **Business Continuity Disaster Recovery(BCDR):**

Geo redundant storage and data replication helps promote business continuity. This enables the services to operate continuously without interruptions or little interruptions in the event of a disaster.

* **Environmental protection:**

Fire, water leakage and thermal issues are also monitored using advanced systems. The AI-optimized cooling is environmentally friendly which supports Microsoft to achieve its vision of consuming 100 percent renewable energy in all its data centers by 2025.

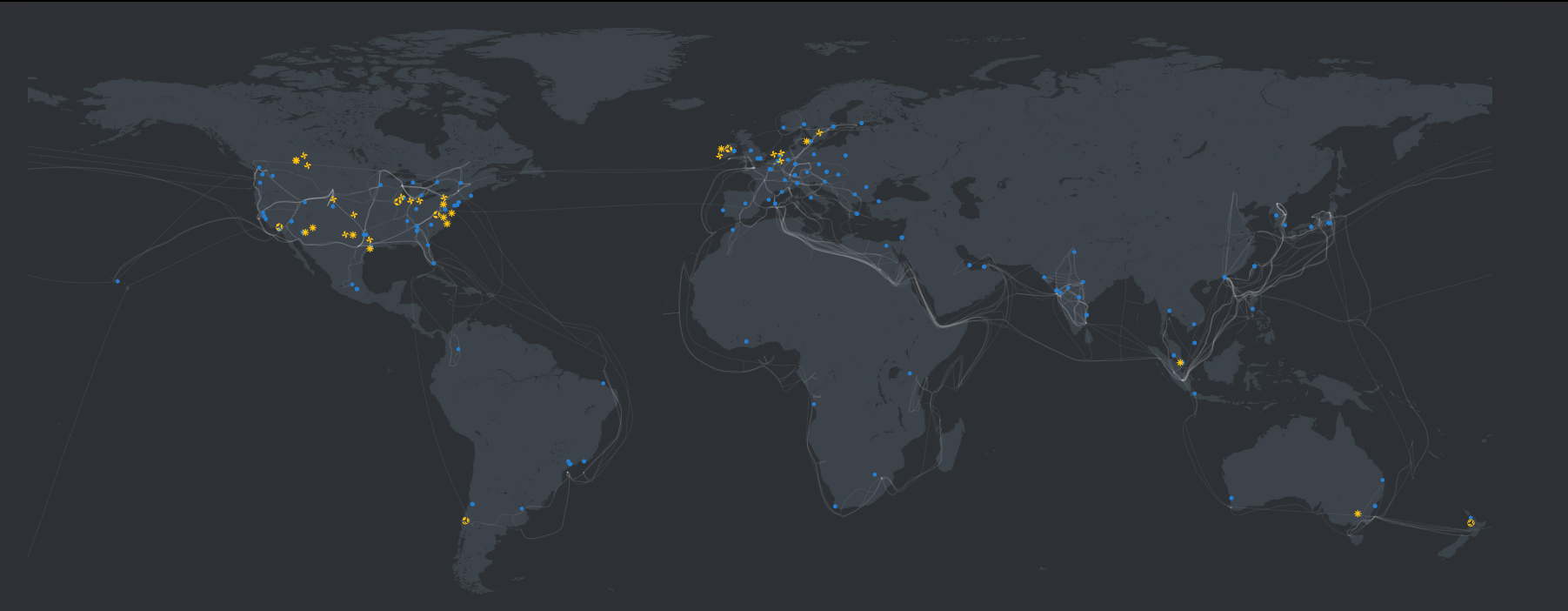
**Sustainability in Data centers**

Microsoft is determined to be carbon negative by the year 2030. It designed its data centers on the concept of sustainability using the following initiatives:

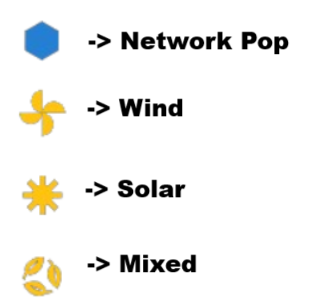
* **Renewable energy:** Data centers are powered through solar or wind energy or with a combination of the two depending on the availability in each locality.
* **Water-Positive Cooling:** Minimal water is used by microsoft for cooling oftenly using the outside air and AI based thermal regulation.
* **Circular Economy:** To minimize e-wastes and carbon footprints the process of recycle and reuse is used.

**Map Illustration:**

The map below illustrates Microsoft datacenter regions in the form of blue dots and highlights those with yellow icons that run on sustainable energy sources. Yellow icons mark places that use wind, solar, or a mix of renewable energy options."



**Representation:**



**Conclusion**

The global infrastructure of Microsoft Azure is designed strategically, to provide safe, high-performing, and reliable cloud services all over the globe. Azure promotes data residency, compliance, fault tolerance, low-latency access and ensures data residency within a properly defined hierarchy of Geographies, Regions, Availability Zones, and Data Centers applicable to organizations of all sizes.

The structure of Azure has a safe worldwide backbone network that is enhanced with AI-optimized regions and improved with multi-layered security. Additionally, the strategic approach to sustainability, AI-optimized regions and environment protection shows the willingness of Azure to become a responsible cloud platform and future-ready.

With this infrastructure, businesses are empowered to develop and grow their applications in any region across the world and still meet the most rigorous demands in terms of performance, security and safety.