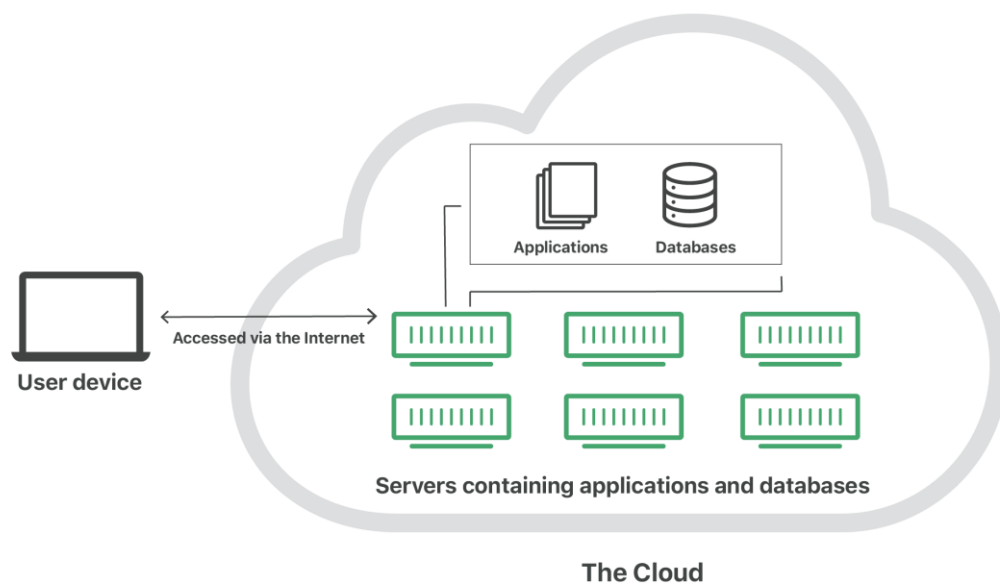


Cloud:

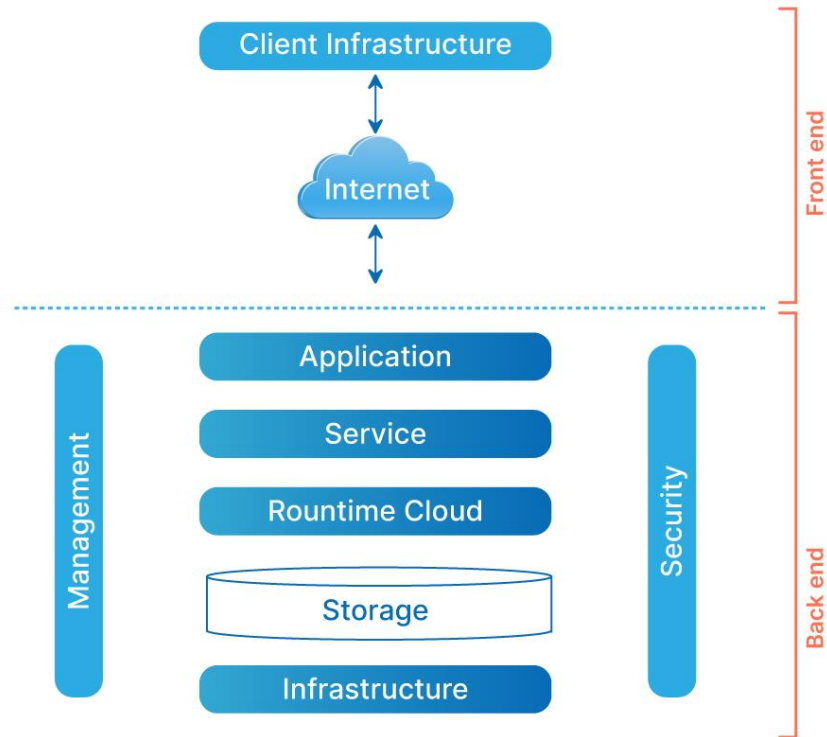
Cloud computing is a technology that enables you to store and access data and software applications over the Internet, rather than using your local device's hard drive or local server. It involves using remote servers hosted on the internet for storing, managing, and processing data. These resources are delivered as services, such as servers, storage, databases, networking, software, analytics, and more. The main advantages include cost savings, scalability, reliability, and on-demand accessibility



Core Architecture

- Front End: User devices (like computers, smartphones, browsers) interact with the cloud via thin or fat clients.
- Back End: Data centers with servers and storage where the actual processing and storage happen
- Network: The Internet (and sometimes an intranet), enabling users globally to connect to cloud services

ARCHITECTURE OF CLOUD COMPUTING



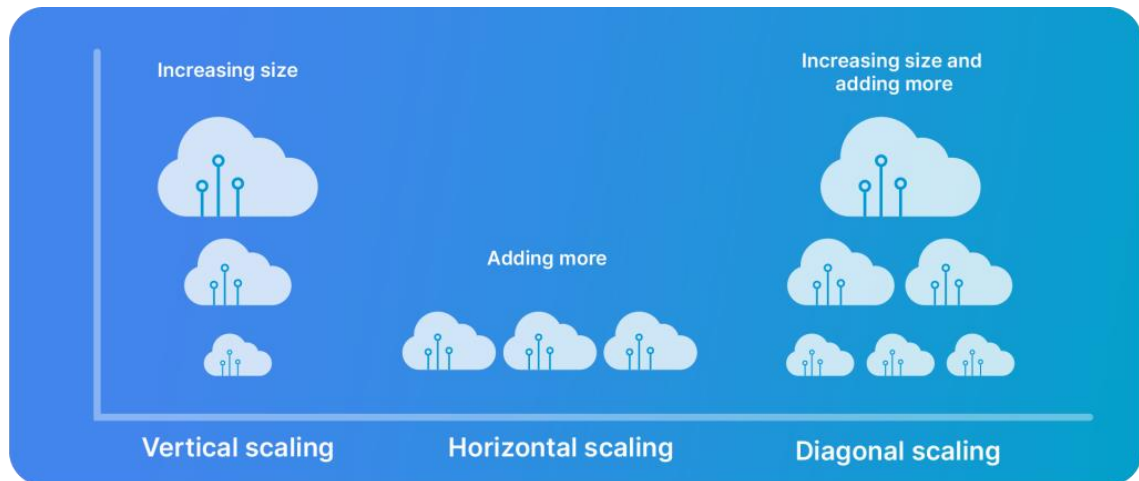
Cloud Scalability:

The cloud scalability is the system's capability to continuously increase (or decrease) its resources to handle growing (or shrinking) workloads, typically over a longer-term or persistent basis

Cloud scalability refers to the ability of a cloud computing system to dynamically adjust its resources (like computing power, storage, or network capacity) to handle changing workload demands. This means it can scale up (increase resources) to handle increased traffic or scale down (decrease resources) when demand decreases, ensuring optimal performance and cost-efficiency

Types of Scalability:

Type	Description	Example
Vertical (Scale-Up)	Increase power (CPU, RAM, storage) of an existing server	Upgrading a server in the cloud to have more memory/CPU
Horizontal (Scale-Out)	Add more instances of servers/resources to work in parallel	Adding more web servers to distribute traffic
Diagonal	Combination of scaling both vertically and horizontally	Add more servers and also upgrade their hardware



How Does it Works:

- **Predictable Growth:** When a business expects its customer base to double yearly, it plans to scale up storage and processing continuously
- **Persistent Workload:** Ideal for applications with steady increases, like expanding a database as your customer records grow.
- **Manual or Semi-Automated:** Scalability can be handled according to projected business growth, unlike the rapid automatic adjustments of elasticity.

Cloud elasticity:

Cloud elasticity refers to the ability of a cloud computing system to automatically adjust its resources (like computing power, memory, and storage) in response to changing demands. This dynamic scaling allows businesses to efficiently handle fluctuating workloads, provisioning resources when needed and releasing them when they are not, optimizing both performance and cost.

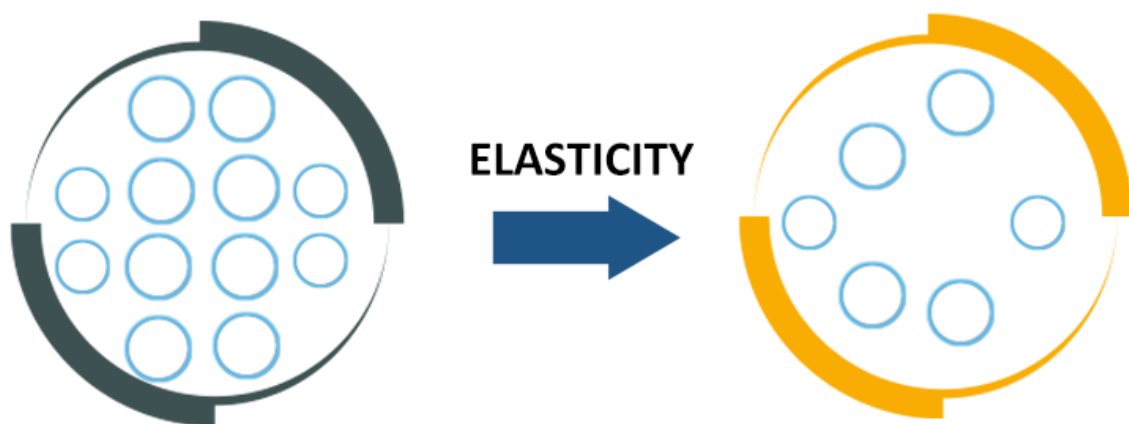
Cloud elasticity refers to the automatic ability of a system to adjust its resources (like CPU, memory, or storage) in real time—expanding or shrinking as workload demands change suddenly. This means if a service suddenly gets a spike in usage, the cloud instantly provisions more resources, and scales them back when the demand drops. Elasticity ensures that fluctuations in load are managed efficiently and cost-effectively, without any need for manual intervention

How Cloud Elasticity Works

- **Automated Resource Provisioning:** The system monitors application load and resource use; when a spike is detected, it automatically adds resources (e.g., virtual machines or storage). When demand decreases, those resources are released.
- **Cost-Efficiency:** Organizations pay only for what they use, which optimizes cloud spend.
- **Real-Time Scaling:** The response to changes happens rapidly, crucial for things like online shopping sites during sales or web applications with unpredictable traffic.
- **Typical Use Cases:** Short-term spikes in usage (festival sales, breaking news sites, viral social media trends).

Example

An online retailer sees a spike in shoppers during a holiday sale. Elasticity means their cloud service can automatically provide more server power for those days and reduce again afterward, so they don't pay for unused resources year-round



Elasticity VS Scalability:

Aspect	Elasticity	Scalability
Nature	Real-time, rapid, automatic adjustment	Planned, gradual long-term growth or downsizing
Scenario	Short-term, unpredictable workload fluctuations	Long-term, predictable growth
Cost Model	Pay-per-use, cost-efficient for fluctuating needs	May involve upfront planning, cost-effective growth
Control	Automated by system	User or admin initiated
Example	E-commerce sales rush	Growing user base or database size

Scalable and Elastic Tax Submission Portal Project Plan

1. On-Premises Solution

Timeline & Server Planning

- TaxSmile – Tax Submission Portal
 - Deployment: On-Premises
 - Initial Infrastructure: 2 Servers
 - Go-Live Month: February

- Scaling Up
 - Load Testing & Stress Testing: March–April
 - Additional Servers: 3 more servers (after testing, total 5 servers)

Implementation Steps

1. Approvals from Stakeholders
 - Directors, COO, CTO
2. Place the Order
 - Vendor: Germany
3. Retrofit the Servers
 - Network Engineers: 24/7 Coverage
 - Teams: 3 operational teams + 1 support team
4. Rent for the Room
 - Arrange appropriate physical space
5. Power Supply
 - Ensure robust and reliable power arrangement

Traffic Considerations

- April–August: Anticipate decline in portal traffic during these months

2. Cloud Solution Alternative

- TaxSmile – Tax Submission Portal
 - Deployment: Cloud Solution (e.g., Azure)
 - Initial Infrastructure: 2 Servers
 - Go-Live Month: February

Rapid Configuration Example

- Azure Portal Steps:
 - Add a text box: 2 steps
 - Save the configuration: 1 minute

3. Notes on Scalability & Elasticity

- On-premises: Scaling up requires significant planning, approval, procurement, and physical setup; scaling down is difficult if traffic drops.
- Cloud: Resources (servers, storage) can be provisioned, scaled, or removed virtually instantly—critical for cost control when traffic varies seasonally.

4. Summary Table

Approach	Key Steps & Features	Timeline	Scaling Details
On-Premises	Approvals, ordering, retrofitting, rooms, power supply	Feb–Apr	2 servers (Feb), +3 after load test (Mar–Apr)
	Multiple operational/support teams needed		Hard to downscale if traffic shrinks (Apr–Aug)
Cloud Solution	Configure servers via Azure portal, rapid setup	February onward	2 servers, easy to add or remove resources instantly
	Text box, save button—1 minute setup		Elastic, cost-effective, quick

Cloud Growths:

