Getting Started with

Cloud Fundamentals

The Journey: From Own Data Centers to Cloud

Goal: Understand the enterprise journey from managing data centers to adopting the cloud

The Starting Point: Challenges with managing data centers

- **Servers**: What a server really is

- **Data Centers**: Where servers live

- **The Pain**: Why managing data centers is difficult

The Solution – Cloud

- **The Benefits** – Why the world is moving to the Cloud
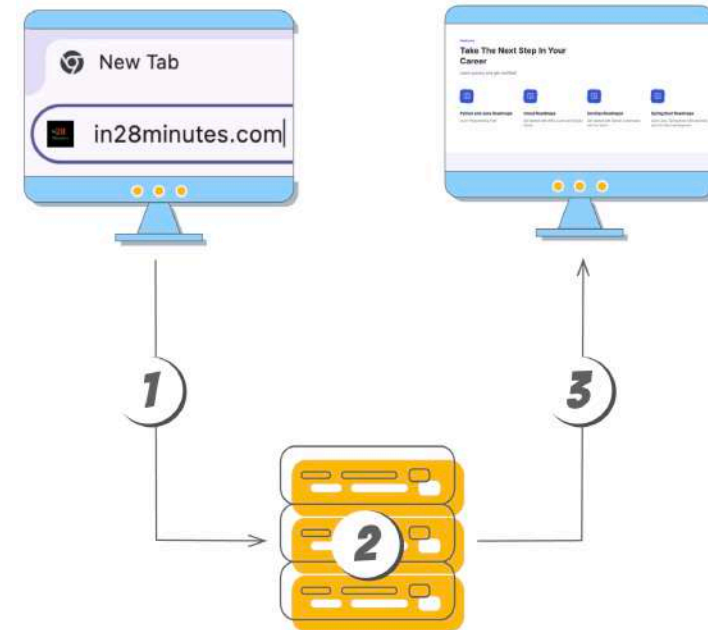
How does in28minutes.com work?

1: Browser - User types URL in the browser

- **Send Request** - Browser finds the server and sends a request to the server

2: Process Request - Server processes the request and prepares the response

3: Send Response - Server sends response back to the browser

- **User Display**: Browser shows the response to the user

What Is a Server?

The Core Concept: A server is a powerful computer designed to provide services

- **Server**: The provider (e.g., in28minutes web server)
- **Client**: The requester (e.g., browser)

What Does a Server Serve?:

- **Websites**: HTML pages, images, videos
- **Files**: Documents, photos, backups
- **Emails**: Sending and receiving messages

Summary: Servers provide services, and clients consume those services
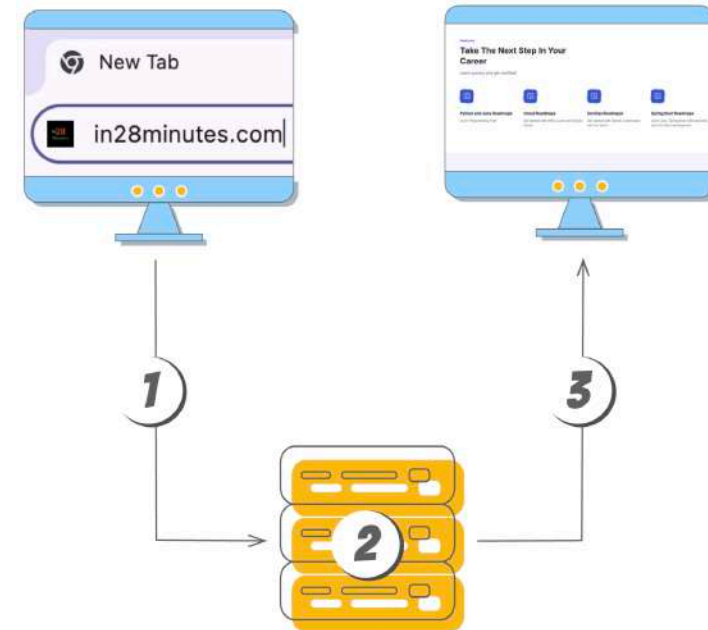
Can One Server Handle Everything?

Scenario: in28minutes.com goes viral

- **Traffic Spike**: Millions of users try to access it

Challenge 1: One server has limits

- **Capacity**: Can only handle a limited number of users

Challenge 2: What if the hardware fails?

- **Single Point of Failure**❗: If the server fails, the application goes offline

Conclusion: Relying on a single server is risky for most applications

The Enterprise Reality – 1000s of Servers

Scenario: Consider a large bank or a financial institution

- **Lots of Apps**: Websites, mobile applications, internal applications and others

- **Thousands of Servers**: Needed to run all kinds of applications reliably

The Challenge: Not recommended to manage thousands of servers in an office

- **Requirements**: You need massive space and power

- **Question**: Where to house this infrastructure?

What is a Data Center?

Basic Idea: A secure place for hosting servers

- **Massive Scale**: Houses hundreds, thousands, or even millions of servers

It is Not Just a Building: Needs specialized infrastructure

- **Power**: Uninterrupted power supply
- **Cooling**: Prevent servers from overheating
- **Networking**: High speed connectivity
- **Strong Security Needs**: Hosts key business systems

**The Challenge**: Building and managing data centers is expensive and complex

London

Data Center - A
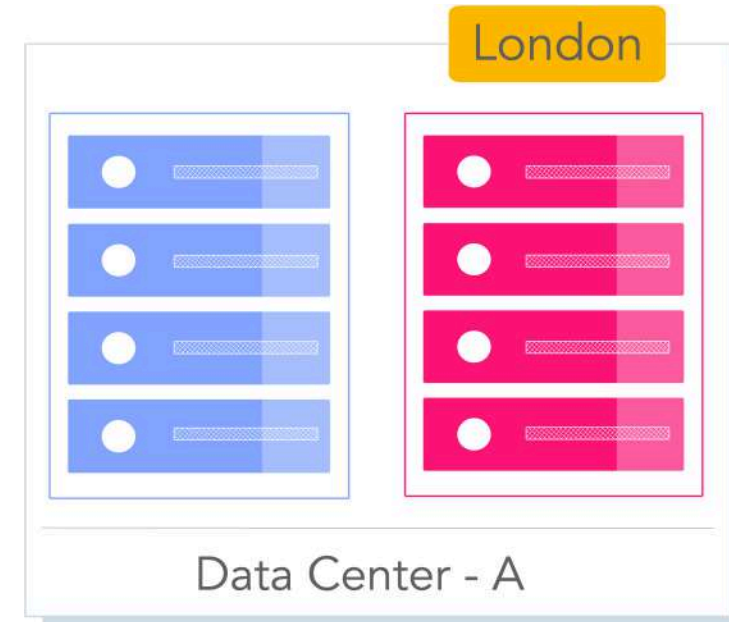
Example 1: Shopping Website

What Happens?: Online shopping demand is not constant

- **Peak Times**: Holidays, sales, weekends
- **Low demand**: During other times

Traditional Approach: Provision For Peak

- **Peak Load Provisioning**: Buy enough servers for peak (maximum) usage
- **Why?**: If you don't provision for peak, the site crashes when you need it most
- **Under Utilized Resources**: Most servers remain idle during periods of low demand
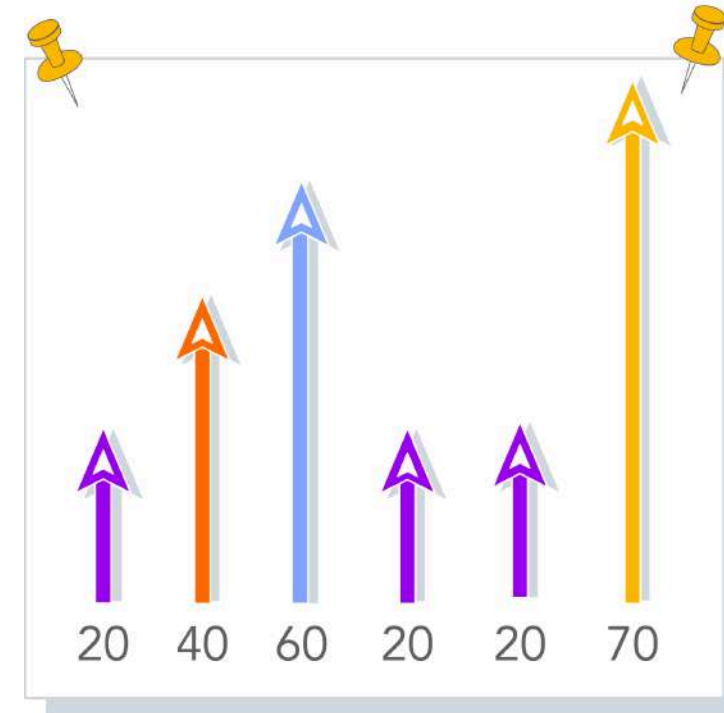
Example 2 : A Startup Succeeds

Scenario: Launches a new app that goes viral

The Problem: Unpredictable growth

- **Surge**: Users grow from 1000 to 1 million overnight
- **The Constraint**: Servers take time to buy and install

Traditional Approach: Buy servers in advance

- **Over-Provisioning**: Buy infrastructure in advance for future growth
- **Risk**: What if growth does not happen?
 - **Outcome**: Expensive hardware remains unused

15 30 15 35 18 50 40 80
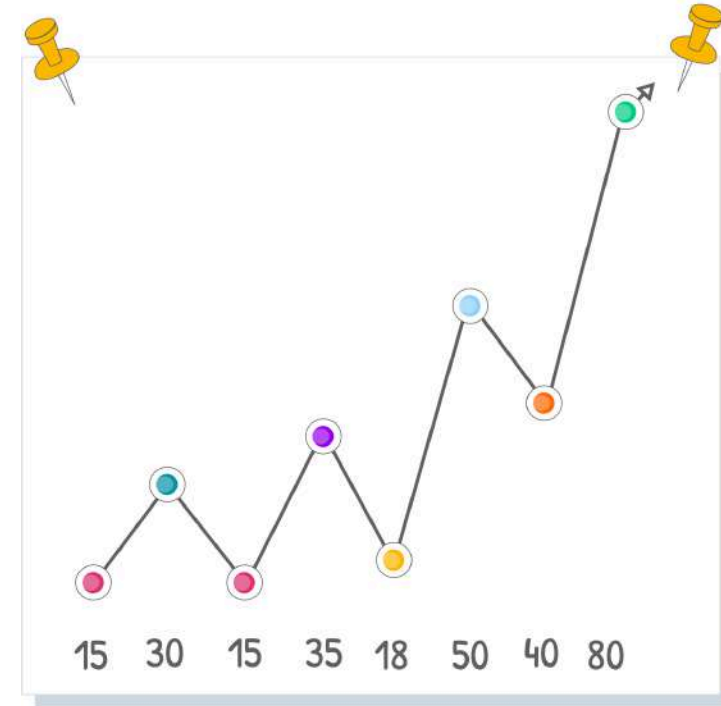
Example 3 : Enterprise Expansion

Scenario: A European company wants to serve Indian customers from a closer location

- **Current Infrastructure:** A Data Center in London

- **Goal**: Enable faster access for Indian users
  - Decrease Latency ❗

Traditional Approach: Set up a New Data Center in India

- **Time**: Takes months or years to set up

- **Cost**: Massive upfront investment
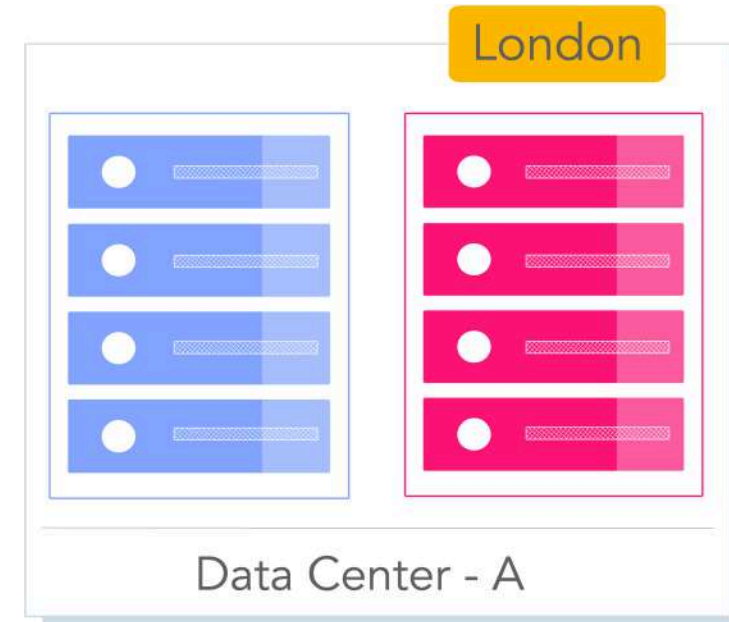
Challenges In Setting Up Own Data Centers

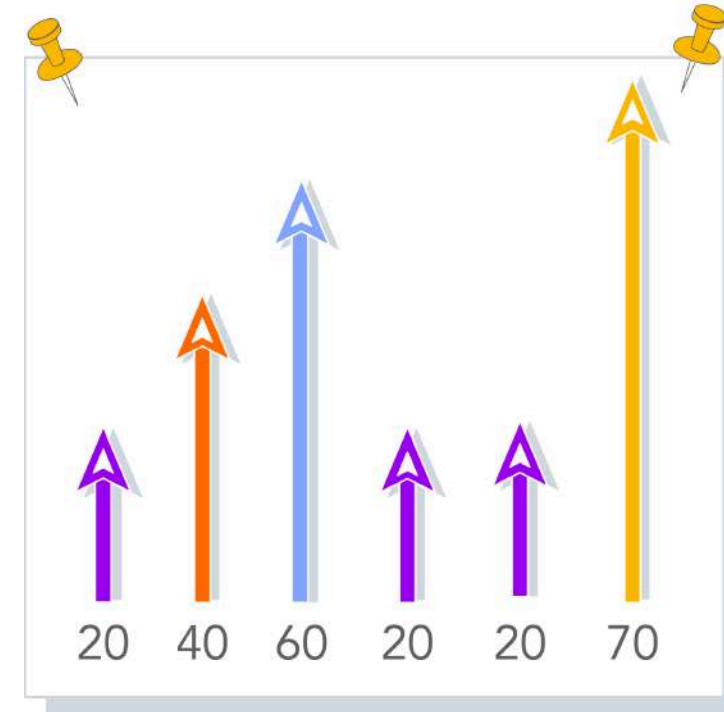**High Upfront Cost**: To Buy Servers

- **Slow Speed**: Takes weeks or months to set up

Needs a Dedicated Team: To maintain infrastructure

- **Think of a startup**: Can a startup afford it?

Guessing Game: You must predict traffic months in advance

**Geographic Limitations**: Difficult to expand into new locations quickly

The Cloud: A Silver Lining

Key Question: Why buy infrastructure when you can rent it?

- **Cloud**: Rent servers when you need them!

- **Remember**: Cloud is NOT magic - it's just someone else's infrastructure (servers) that you rent

Core Concepts:

- **On-Demand Access**: Rent resources (servers, storage, databases) instantly

- **Elasticity**❗ : Increase or decrease the number of resources (servers) based on demand

- **Pay-As-You-Go**❗ : Pay only for what you use
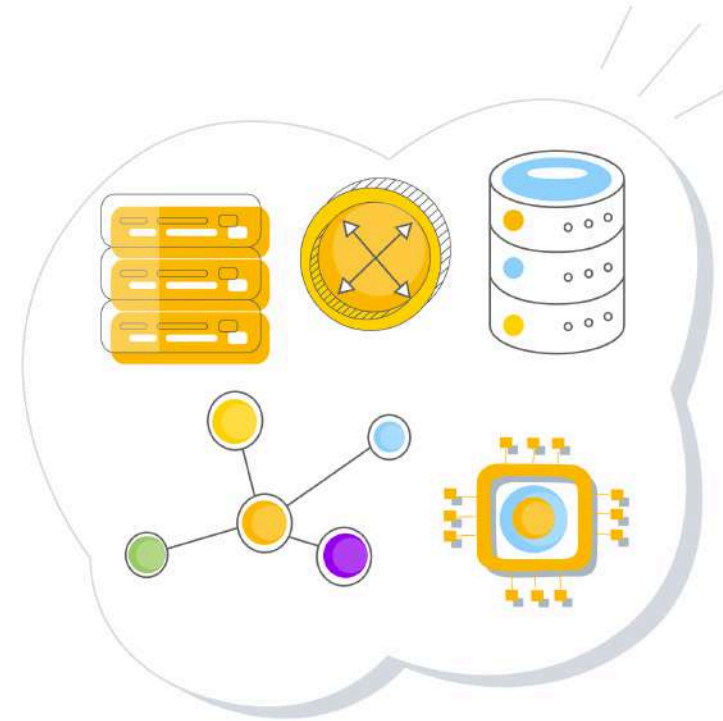
Cloud Advantage – Pay for Use

Scenario: A startup launching a new app

Traditional Way: Needs investment to set up a data center before launching

- **CapEx**❗ : Capital Expenditure or Capital Expense

The Cloud Solution: Rent and start small

- **Rent Servers**: From cloud provider
- **Pay for Use**: Costs only grow if the app grows
- **Lower Risk**: Fail without heavy investment
- **OpEx**❗ : Operational Expenditure or Variable Expense

15 30 15 35 18 50 40 80
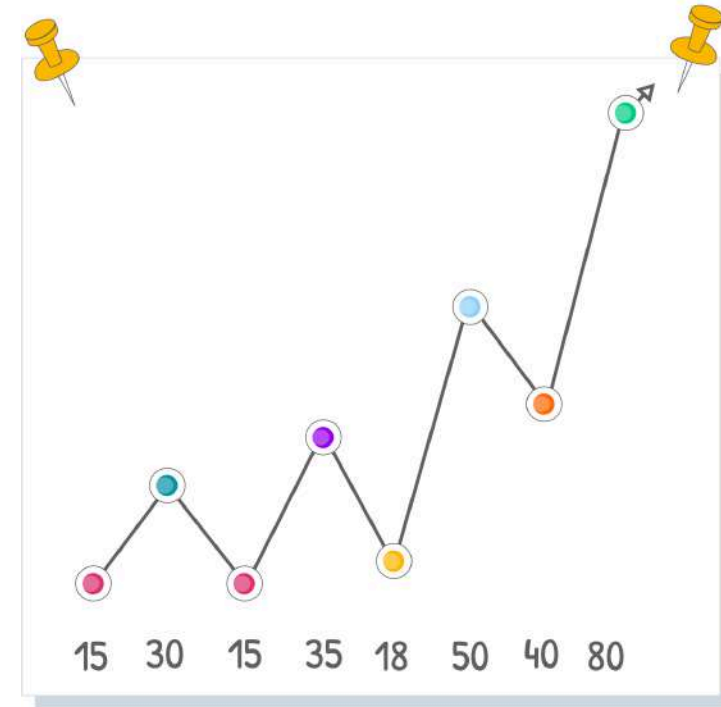
Cloud Advantage – Economies of Scale

Scenario: Who gets a better deal?

- **You**: Buying 10 hard disks OR
- **Your Cloud Provider**: Buying 10 million hard disks

Economies of Scale ❗ : Buying infrastructure in large quantities typically lowers the cost per unit

The Cloud Reality: Cloud providers purchase infrastructure in massive volumes

- **Result**: Lower cost per unit
- **Customer Benefit**: Lower prices passed on to users

New York

London

Cloud Advantage – Stop Guessing Capacity

Scenario: Applications and websites needs to predict traffic a few weeks ahead of time

- **Not easy**: Difficult to predict accurately
- **Consequence**: Over-provisioning or Under-provisioning

The Cloud Solution: Elasticity ❗

- **Scale on Demand**: Add more servers instantly when traffic peaks
- **Release Servers**: Release servers when traffic drops
- **Impact**: You do NOT need to pay for idle resources
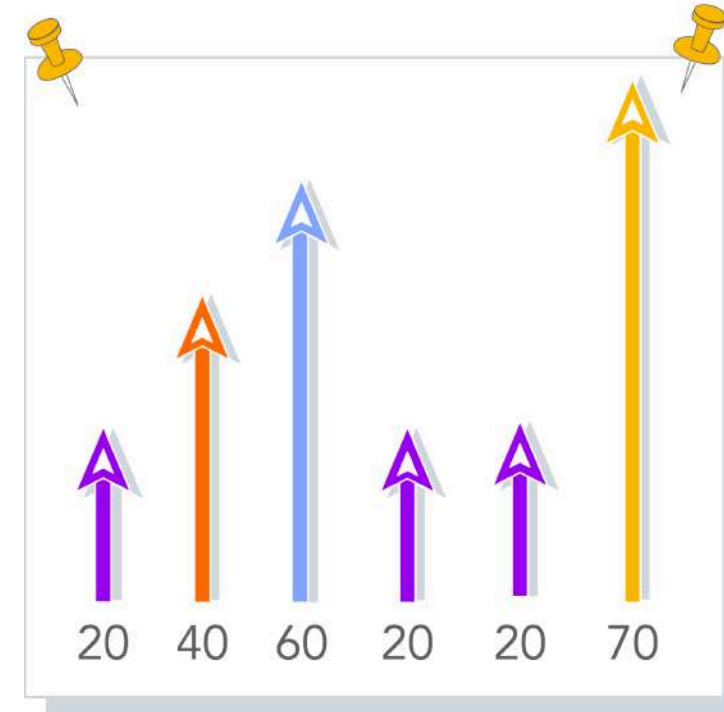
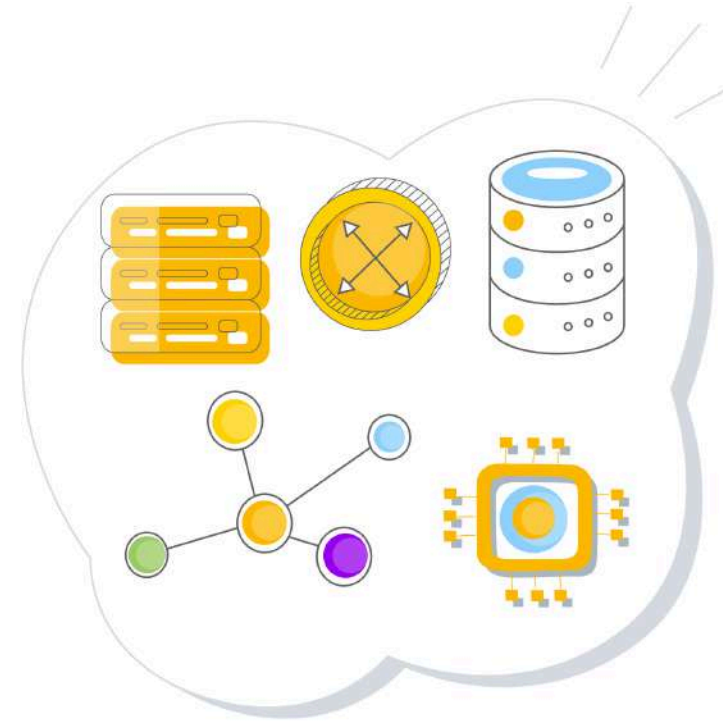# Cloud Advantage – Increased Speed and Agility

Scenario: Your team has a great idea for a new application

- **Traditional Approach**: "Wait 6 weeks for us to buy new servers"
- **Result**: Innovation slows down!

The Cloud Solution: "Click a button, get a server in minutes"

- **Agility**❗ : The ability to move fast, and respond to changes quickly
- **Faster Time to Market**: Launch features faster

Cloud Advantage – Go Global

Scenario: You have users around the world!

The Challenge: High latency for global users if your servers are only in one location

Cloud Solution: Deploy to multiple regions in minutes

- **Global Reach** ❗ : Expand to new markets around the world quickly

- **Better Experience**: Provide low latency (faster access) to users worldwide

- **Lower Cost**: No need to build data centers and manage infrastructure in multiple locations

Cloud Advantage – Avoid Heavy Lifting

Scenario: Insurance Company

- 30% on infrastructure, 70% on core business
- **Undifferentiated Heavy Lifting**❗ : Tasks that are hard but don't make you unique

The Cloud Solution: Cloud reduces infrastructure burden

- **Avoid Undifferentiated Heavy Lifting**: Spend more time on what makes your business unique
- **Example – Insurance Company**: Get better at calculating premiums or detecting fraud (using AI)
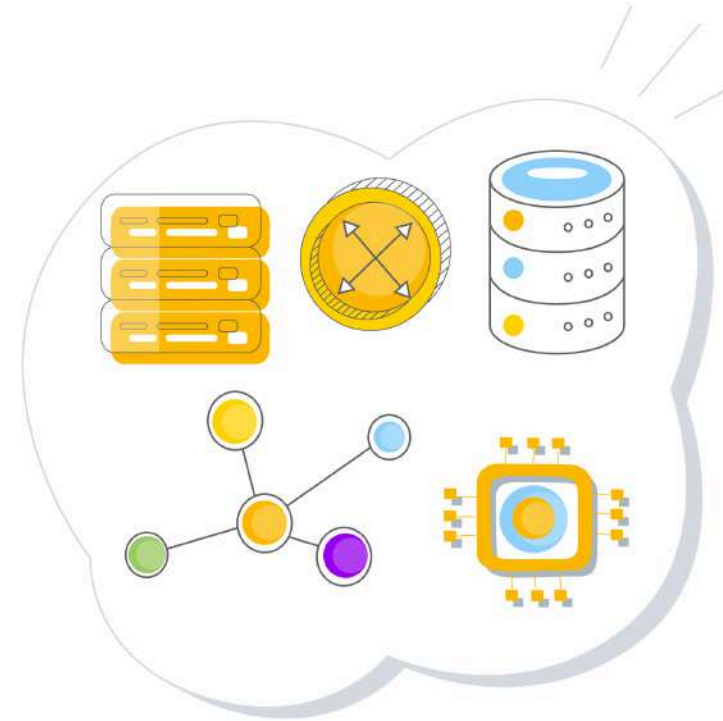
Cloud – Key Terminology to Remember (1 of 2)

Term	Description
Elasticity	Ability to automatically scale resources up or down based on demand
Agility	Ability to move fast and respond quickly to business needs Quickly deliver software Quickly provision infrastructure
Global Reach and Geo-Distribution	The ability to expand to new markets and deploy applications to data centers around the world quickly
Availability	Are applications available when users need them
Latency	Time taken for a request to travel between user and server Reduced by serving users from nearby locations

Cloud – Key Terminology to Remember (2 of 2)

Term	Description
Capital Expenditure (CapEx)	Large upfront investment - For example, in infrastructure like servers and data centers
Operational Expenditure (OpEx)	Ongoing cost for renting resources (or services) as needed
Pay As You Go	Trade capital expense (CapEx) for variable expense(OpEx). Pay only for the resources you actually use.
Economies of Scale	Buying resources in large quantities lowers the cost per unit
Avoid Undifferentiated Heavy Lifting	Spend more time on what makes your business unique Example: Insurance company focusing on calculating better premiums.

Cloud Concepts – A Few Scenarios (1 of 2)

Scenario	Cloud Concept
Netflix adds thousands of servers instantly on Friday nights to stream movies to millions, then releases them on Monday morning.	Elasticity - increase or decrease no of servers based on demand
A startup launches a new app with minimal cost and increases spending only as user adoption grows	Pay As You Go or Operational Expenditure
A product team provisions infrastructure in minutes to test new features and releases updates faster	Agility
A healthcare application ensures patients can access services 24x7, even during maintenance or failures	High Availability

Cloud Concepts – A Few Scenarios (2 of 2)

Scenario	Cloud Concept
A SaaS company deploys its application in multiple regions to serve customers worldwide with low latency	Global Reach and Geo-Distribution
A media streaming platform serves videos from locations closest to users for faster playback	Low Latency
A cloud provider buys hardware in massive quantities and offers lower prices to customers	Economies of Scale
An insurance company focuses on fraud detection and customer experience while the cloud provider manages infrastructure	Avoid Undifferentiated Heavy Lifting