

# Data analytics -chapter 1

## module 2

**why we need effective cloud migration ke steps:**




### **Example: Company Migrating to AWS**

#### **Situation:**

Ek E-commerce company ke paas apna on-premises data center hai jisme:

- Website host hai
- Customer database store hai
- Billing system chal raha hai

Lekin ab unhe problems aa rahi hain:

- Traffic sudden increase hone par server crash ho jaata hai  

- Infrastructure maintain karne ka kharcha zyada hai
- Global customers ke liye speed slow hai

#### **Steps for Effective Cloud Migration**

##### **1. Assessment & Planning (Current System ka Analysis karna)**

- Pehle apne existing IT infrastructure, applications, databases, aur workloads ka analysis karo.
- Decide karo ki kaunsi cheez migrate karni hai aur kaunsi on-premises rakhni hai.
- Cost, performance, aur compliance check karo.

##### **2. Choose Right Cloud Model (Public, Private, Hybrid, Multi-Cloud)**

- Business ki zarurat ke hisaab se model select karo.
- Example: Agar high security chahiye → Private Cloud, scalability chahiye → Public Cloud.

### 3. Select Migration Strategy (6 R's of Migration)

- **Rehost (Lift & Shift)** – As-is move karo.
- **Replatform** – Thoda modify karke shift karo.
- **Refactor** – Pura redesign karke migrate karo.
- **Repurchase** – New SaaS solution adopt karo.
- **Retire** – Unwanted apps ko hata do.
- **Retain** – Kuch cheezein abhi ke liye on-premises hi rakho.

### 4. Data Migration Planning

- Backup banao.
- Secure aur encrypted channels use karo.
- Phased migration karo taaki **downtime** kam ho.

#### **Note: What is Downtime?**

Downtime ka matlab hai jab system, website, ya application temporarily unavailable ho jaye (band ho jaye) aur users usse access na kar paaye.

### 5. Application Migration & Testing (deploy)

- Applications ko step-by-step migrate karo.
- Testing karo – performance, security, aur integration ke liye.

### 6. Optimization & Security

- Resources ko monitor karo aur optimize karo (cost cutting).
- Cloud-native security tools aur IAM (Identity & Access Management) apply karo.

### 7. Training & Change Management

- Team ko new cloud tools aur workflows sikhana zaroori hai.
- Documentation aur guidelines ready rakho.

### 8. Monitoring & Continuous Improvement

- Migration ke baad monitoring tools lagao.
- Performance aur cost optimize karte raho.
- Feedback loop maintain karo.
- 📌 **Isko yaad rakhne ka simple mnemonic:**  
APSMOTCM **A**ssess  
• **P**lan Cloud Model , **S**elect Strategy , **M**igrate Data , **O**ptimize  
• **T**est **C**hange Management **M**onitor

Migration path phases :

Assess , plan , deploy , optimize

**Cloud deployment model :** public cloud , private cloud , hybrid cloud

▪ **Public Cloud :**

- Yeh ek kiraye ke ghar (rented apartment) jaisa hai. Jaise ek building mein bahut saare log alag-alag flats mein rehte hain, waise hi public cloud mein ek hi server par bahut si companies apna data rakhti hain.
- Saara maintenance aur management (jaise security, updates) cloud provider (jaise Google Cloud, AWS) dekhta hai. Aapko bas apna "rent" dena hai aur use karna hai.
- Fayda: Flexible hai, sasta padta hai, aur aap jitna use karoge utna hi paisa doge (scalable).
- Nuksaan: Security aur control thoda kam hota hai kyunki sab kuch shared hai.

**Benefits:** Cost-efficient • Scalable • No maintenance • Reliable • Fast and easy to deploy • Innovative

**Example tools:** Google Cloud Platform • Amazon Web Services • Microsoft Azure

▪ **Private Cloud :**

- Yeh aapke apne ghar (owned house) jaisa hai. Sab kuch aapka hai aur aapke control mein hai.
- Saare resources (servers, storage) sirf ek hi company ke liye hote hain. Company isko khud manage karti hai ya kisi third-party se karwati hai, lekin rehta yeh private hi hai.
- Fayda: Full control aur high security milti hai. Sensitive data jaise banking ya health records ke liye yeh best hai.
- Nuksaan: Bahut mehenga (expensive) hota hai aur isko manage karna ek badi zimmedari hai.

**Benefits:** Secure • Compliant with internal regulations • Consistent and predictable performance

**Example tools:** HPE Greenlake • Amazon Virtual Private Cloud • VMWare

▪

## Hybrid Cloud :

- Yeh "best of both worlds" wala option hai. Matlab, apna ghar bhi hai aur zaroorat padne par ek rented storage unit bhi le rakha hai.
- Ismein company apna zaroori aur sensitive data private cloud mein rakhti hai (apne ghar mein), aur baaki applications ya kam zaroori data ke liye public cloud ka istemaal karti hai (kiryae ka storage).
- Fayda: Security bhi mil jaati hai aur public cloud ki flexibility aur kam kharche ka fayda bhi utha sakte hain.

**Benefits:** Strong computing power • Uses modern innovations • Secure • Compliant with internal regulations • Fast performance • Flexible operations

**Example tools:** Google Anthos • Microsoft Azure Arc • AWS Outposts

## Managed cloud services:

Managed cloud services ka seedha sa matlab hai ki aap apni cloud infrastructure (jaise servers, storage, ya database) ki management aur maintenance ka kaam ek third-party expert company ko de dete ho.

Eg,, Soch lijiye aapne ek car khareedi (yeh aapka cloud server hai), lekin uski servicing, safai, aur maintenance ke liye aapne ek company hire kar li hai jo yeh sab tension leti hai. Aap bas car chalte ho. Yahan, woh company aapki managed service provider hai.

**Cloud data management :** 1. data integration , 2. data ingestion

## ☁ Cloud Data Management

Cloud Data Management ka matlab hai cloud par maujood apne saare data ko manage karna—yani usko store karna, सुरक्षित rakhna, organize karna, aur access karna.

Soch lijiye aapka saara important data aapke local computer ya office server par nahi, balki internet par Google, Amazon, ya Microsoft ke powerful data centers (yani cloud) mein rakha hai. Uss data ki poori dekhbhaal ko hi Cloud Data Management kehte hain.

## **data integration :**

Data Integration ka matlab hai **alag-alag sources (jagahon) se data ko ikattha karke ek jagah par jodna** taaki aapko ek saaf aur complete view mil sake.

Aapka data alag-alag jagah ho sakta hai—kuch sales ka data ek software mein, customer ki jaankari doosre software mein, aur marketing ka data social media par. Data integration in sab tukdon ko jodkar ek poori tasveer banata hai.

sko karne ka ek popular tareeka hai **ETL (Extract, Transform, Load)**, **ELT (Extract, Load, Transform)**

### **ETL(Extract, Transform, Load) :**

**Extract (Nikaalna):** Alag-alag sources se data ko nikaalna.

**Transform (Badalna):** Data ko saaf karna, aaps mein milane ke liye ek common format mein laana.

**Load (Daalna):** Uss saaf aur format kiye hue data ko ek central jagah, jaise ki Data Warehouse, mein daal dena.

**WHEN WE USE :** Structured Data (Tables), chhota data, jab privacy important ho.

### **ELT (Extract, Load, Transform) :**

**Extract (Nikaalna):** Aap alag-alag jagahon (Data Source) se saara samaan (Raw Data) order karte ho.

**Load (Daalna):** Delivery wala saara kachcha samaan (Raw Data) jaisa ka taisa aapke ghar (Data Lake/Warehouse) par seedha deliver kar deta hai. Yahan data bahut tezi se load ho jaata hai kyunki us par koi processing nahi hui.

**Transform (Badalna/Pakana):** Ab jab aapko zaroorat padti hai, aap apni kitchen (Data Lake) ke andar apne hisaab se samaan nikaalte ho aur use pakate (Transform) ho. Transformation ka kaam destination ke andar hi hota hai.

**When we use :** Big Data, Unstructured Data (images, videos, text), jab speed zaroori ho.

## Data ingestion : Data Ingestion

Data Ingestion ka seedha sa matlab hai alag-alag sources (jagahon) se data ko collect karke, ek central storage system (jaise ki Data Lake ya Data Warehouse) ke andar laana.

Yeh kisi bhi data pipeline ka sabse pehla step hota hai.

Ek analogy se samjho: Soch lijiye aapke shehar ka ek bada Mail Sorting Center (post office) hai. Wahan par alag-alag jagahon se—ghar, letter box, doosre shehron se—chitthiyon aur parcels (yeh aapka Data hai) aate hain.

Mail center ka sabse pehla kaam unni saari chitthiyon aur parcels ko collect karke building ke andar laana hai. Bas isi process ko, yani data ko bahar se system ke andar laane ko, Data Ingestion kehte hain.

Data ko sort karna, padhna, ya सही पते पर bhejna (yani Transform ya Process karna) yeh sab kaam *baad* mein hota hai. Pehla kaam sirf data ko 'ingest' karna yaani andar lena hai.

### Data Ingestion Ke Types (Yeh kaise hota hai?)

Data ko system mein laane ke do mukhya tareeke hain:

#### 1. Batch Ingestion (Ek Saath Group Mein)

- Ismein data ko groups ya 'batches' mein ikattha kiya jaata hai aur phir ek schedule ke hisaab se (jaise har ghante, ya din ke end mein) ek saath system mein laya jaata hai.
- **Example:** Har raat 12 baje, poore din ki sales ka data ek saath database mein upload karna.
- **Analogy:** Ek school bus jo subah ek fixed time par saare bachchon ko ek saath school laati hai.

#### 2. Real-time Ingestion (Foran/Live)

- Ismein jaise hi data kisi source par create hota hai, use foran (milliseconds mein) system ke andar bhej diya jaata hai. Ise **Streaming Data** bhi kehte hain.
- **Example:** Cricket match ka live score, stock market ke live prices, ya aapki social media feed par aane wale live updates.
- **Analogy:** Ek taxi jo aapko book karte hi turant lene aa jaati hai, schedule ka intezaar nahi karti.

## ETL/ELT Se Kya Connection Hai?

Yaad hai humne ETL (Extract, Transform, Load) aur ELT (Extract, Load, Transform) padha tha?

**Data Ingestion** us process ka "**E**" yaani "**Extract**" wala hissa hai. Yeh data ko uske source se nikaalkar system tak laane ka kaam karta hai. Uske baad 'T' (Transform) aur 'L' (Load) ke steps aate hain.

## How to access data :

- **APIs (Application Programming Interfaces):** Yeh do applications ko aapas mein baat karne aur data exchange karne ka tareeka hai.
- **SQL (Structured Query Language):** Isse database se data nikaala (query) jaata hai.
- **Ingestion Tools (jaise Pub/Sub):** Yeh special services hain jo real-time data ko handle karne ke liye banayi gayi hain.
- **Business Intelligence Solutions (jaise Looker, Jupyter Notebooks):** Yeh tools data ko access karke use analyze aur visualize karne mein madad karte hain.

## Cost in cloud services : cloud cost manage

### A)Resource Provisioning (Resources Ko Arrange Karna):

Yeh cloud costs ko manage karne ka pehla aur sabse important tareeka hai. app (user) apne kaam ke liye zaroori **software aur hardware (jaise server, CPU, RAM) chunte hain**, aur cloud provider (jaise Google) unhe aapke liye set up aur manage karta hai.

Iske teen **tareeke (delivery models)** hain:

1. **Advance Provisioning:** Aap provider ke saath pehle se ek contract sign karte hain aur ek **fixed price** dete hain. Provider aapke liye utne resources hamesha taiyaar rakhta hai. Yeh ek party ke liye hall pehle se book karne jaisa hai.
2. **Dynamic Provisioning:** Yahan resources aapki **zaroorat ke hisaab se automatically kam ya zyada** hote hain. Agar traffic badh gaya to resources badh jaayenge, traffic kam hua to ghat jaayenge. Aap sirf utna hi pay karte hain jitna use karte hain. Ise "pay-as-you-go" bhi kehte hain.

3. **Self-Provisioning:** Aap ek website ya online portal par jaakar **khud hi apne liye resources select karke khareedte hain**, aur woh kuch hi minutes ya ghanton mein aapko mil jaate hain. Yeh online shopping jaisa hai.

## B) Storage Costs (Data Store Karne Ka Kharcha):

- Data Storage (Kitna data store kiya):

Types of data :

**Coldline Storage:** Aise data ke liye jo aap 3 mahine mein shayad ek baar access karte hain.

**Archive Storage:** Yeh sirf data backup ke liye hota hai. Yeh sabse sasta storage option hai.

- Data Processing (Data ko saaf karna):
- Network Use (Data ko move karna):

## C) Running Queries (Data Se Sawal Puchne Ka Kharcha):

Jab aap koi query chalate hain, to cloud provider aapse uss data ke liye charge nahi karta jo result mein aaya, balki uss **poore data ke liye charge karta hai jisko query ne scan kiya** ya process kiya.

**Google big query pricing :** on-demand pricing , capacity pricing

## Cloud in different industries :

Education , medical , manufacturing , transportation industries

**Cloud architecture :** Cloud Architecture Kya Hai? (The Shopping Mall Analogy)

Aap mall jaate hain, wahan aapko har tarah ki dukaan (services) milti hai— kapde, khana, electronics. Aapko jo chahiye, aap use karte hain. Lekin aapko mall ki **bijli, paani, maintenance, ya security ki tension lene ki zaroorat nahi hoti**. Woh sab mall management (Cloud Provider) dekhta hai.

Theek isi tarah, **Cloud Architecture woh blueprint ya design hai jiske upar poora cloud bana hota hai**. Yeh batata hai ki alag-alag components (jaise storage, servers, software) kaise aapas mein judkar kaam karenge taaki companies (users) un services ko use kar sakein.



**Cloud architecture ke do main hisse hote hain:** Frontend aur Backend.

**Frontend:**

- Yeh architecture ka woh hissa hai jise user seedha dekhta hai aur interact karta hai. Yeh "parde ke aage" ka part hai.
- Example: Jab aap apne phone par web browser kholkar ek Google Doc edit karte hain, toh aapka phone, web browser, aur Google Docs ka app—yeh sab frontend components hain.

**Backend Platform:**

Application, Service , Runtime Cloud , Storage , Infrastructure(cpu,gpu), Security

## **Google Cloud Architecture Framework Explained**

Yeh document **Google Cloud Architecture Framework** ke baare mein hai. Aap ise ek mazboot building banane ke liye

**blueprint** ya **naksha** samajh sakte hain. Yeh framework cloud professionals ko aise cloud systems design karne aur chalaane mein madad karta hai jo **surakshit, kushal (efficient), bharosemand (resilient), high-performing, aur saste (cost-effective)** hon.

Is framework ke chhah (6) mukhya hisse hain, jinhe "pillars" ya stambh kaha jaata hai:

1. **System Design (Buniyad):** Yeh framework ka foundation ya buniyad hai. Ismein aap cloud system ki poori planning karte hain—uska architecture, components, aur data kaisa hoga.
2. **Operational Excellence (Kushal Sanchalan):** Iska matlab hai ki aap apne cloud ke kaam ko kitne behtar tareeke se deploy, operate, monitor, aur manage karte hain.
3. **Security, Privacy, and Compliance (Suraksha aur Niyamon Ka Paalan):** Yeh sunishchit karta hai ki aapka data cloud par surakshit hai, aapki jaankari private hai, aur aapka design company ke niyamon ke anusaar hai.
4. **Reliability (Bharosemandi):** Iska matlab hai ki aapka system itna mazboot ho ki woh cloud par apne saare kaam (workloads) ko aasaani se sambhal sake.

5. **Cost Optimization (Laagat Mein Bachat):** Iska matlab hai ki aap cloud par kharch kiye gaye paise ka poora fayda uthayein aur laagat ko kam se kam rakhein.
6. **Performance Optimization (Behtareen Performance):** Iska matlab hai ki aap apne cloud resources ko lagataar behtar banate rahein taaki system hamesha best performance de.

### **Key Strategies for Cloud Cost Optimization:**

Rightsizing , Auto-scaling , Reserved Instances

**Rightsizing:** Yeh apne computing resources, jaise ki processing power aur storage, ko kisi application ya workload ki bilkul aavashyak zarooraton (exact needs) ke hisaab se adjust karne ka process ha