

Introduction to Big Data & Industry Use Cases

Mehal Sakthi M S

Analyst SDE

Why Are We Talking About **Big Data**?



How many of you use Instagram daily?



How much data does YouTube upload every minute?



How many Google searches per second?

500+ hours of video uploaded to YouTube every minute

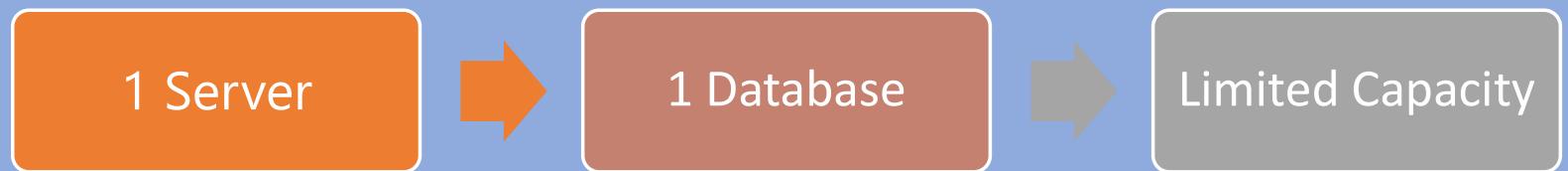
Millions of transactions per second globally

Billions of mobile users generating data daily

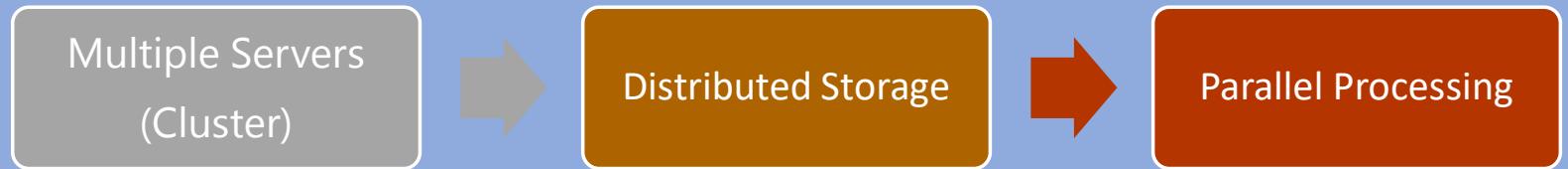
What is **Big Data**?

Extremely large datasets that cannot be processed using traditional databases and tools

Traditional Systems



Big Data Systems



5Vs of Big Data

Volume

- Terabytes → Petabytes → Exabytes

Velocity

- Real-time data (stock markets, UPI transactions)

Variety

- Structured, Semi-structured, Unstructured

Veracity

- Data accuracy and trustworthiness

Value

- Turning data into business decisions

Traditional vs Big Data Systems

Traditional Systems

GB scale

Single server
(Database mostly)

Structured data

Fixed Schema

Limited scalability

Batch Processing

Big Data Systems

TB–PB scale

Distributed clusters

All data types

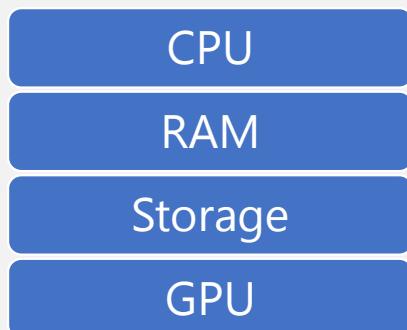
Dynamic Schema

Horizontally scalable

Batch and Streaming

Vertical Scaling - Traditional

Increasing the capacity of a single machine by adding more



Upgrading a server from
8GB RAM → 64GB RAM
4 Cores → 32 Cores

Advantages

- Simple to implement
- No change in application architecture
- Easy maintenance

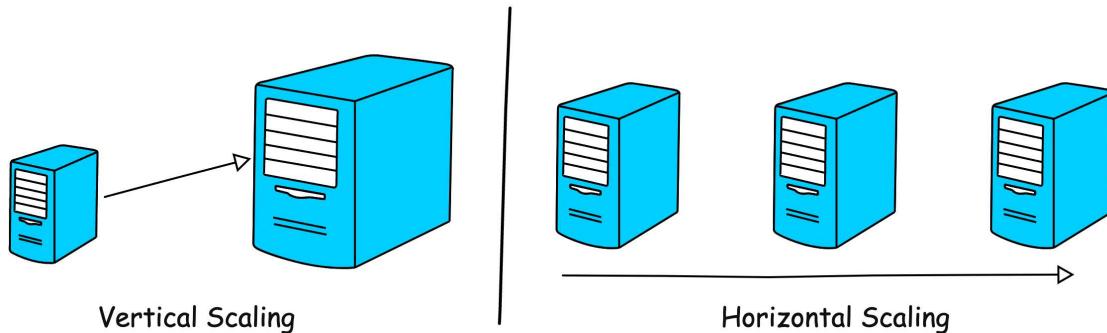
Limitations

- Hardware limit exists
- Expensive upgrades
- Single point of failure

Horizontal Scaling – Big Data

Adding more machines (nodes) to distribute workload

Instead of one powerful server
Use 10 smaller servers working together (Cluster).



Advantages

- Highly scalable
- Cost-effective
- Fault tolerant
- Handles massive data (TB–PB scale)

Limitations

- More complex architecture
- Requires distributed systems design

Why Big Data became Important

Explosion of
Internet

Smartphones

Social Media

IoT Devices

Cloud
Computing

Industry Use Case 1

Netflix Recommendation System

Data Collection at
Massive Scale

User Profiling

Recommendation
Algorithms

Real-Time
Personalization

Industry Use Case 2

Banking Fraud Detection Systems

Process millions
of records per
second

Low latency
decision systems

Distributed
architecture

High reliability

So Far...

What is Big Data

Upcoming Next..

**Who builds the systems and
who analyses them**

Thank You