</talentlabs>

CHAPTER 2

Cloud Introduction

</talentlabs>

AGENDA

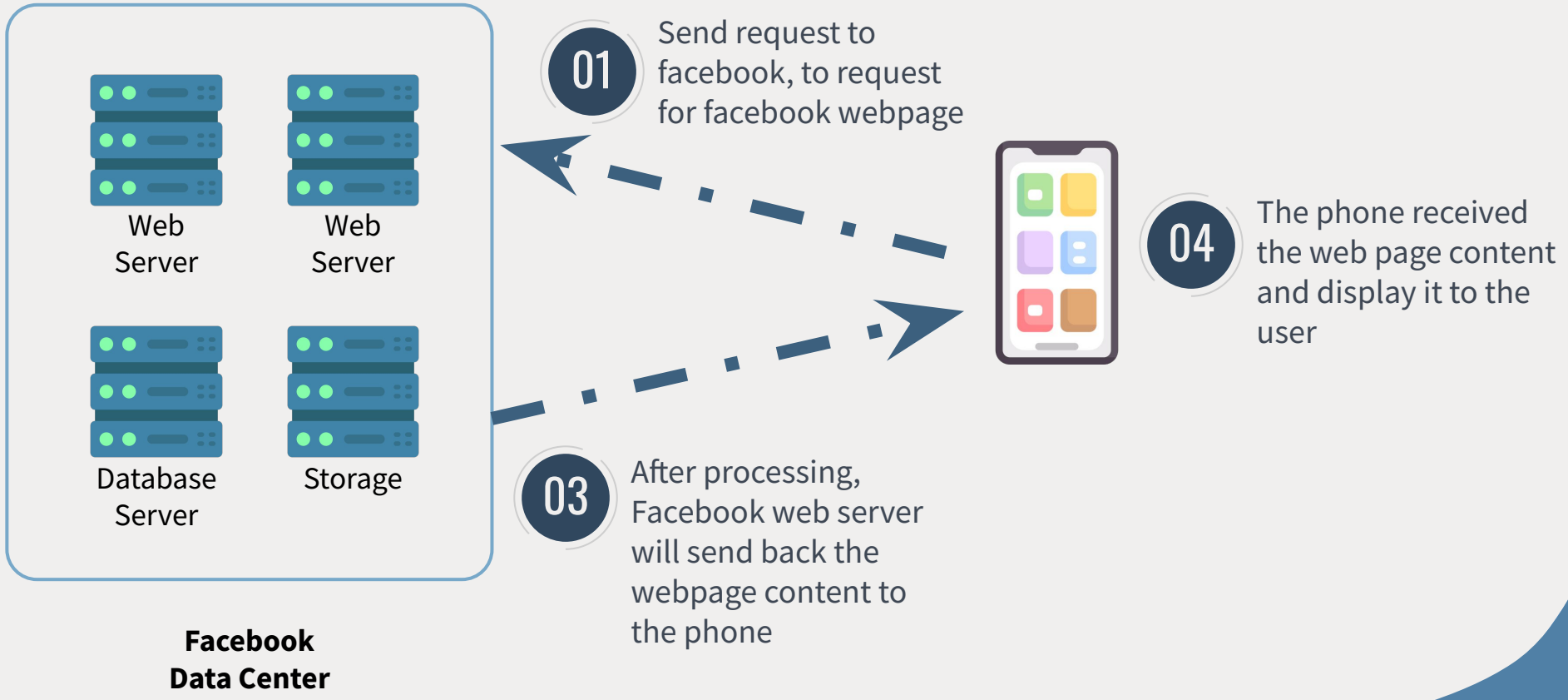
- Servers & Cloud
- Virtual Machine in the Cloud
- Storage in the Cloud
- Database in the Cloud
- Cloud Case Study

talentlabs

Servers and Cloud

</talentlabs>

How do we get access to Facebook?



</talentlabs>

Learnings from the Diagram

01

The need of servers

All apps or web pages need to be served from a special type of computer call “servers”

02

Various types of servers

There are different types of servers, like **web servers, database servers, and storage servers**

03

Data centers are costly

To serve users, you will need **a lot of servers** in the data center

What are servers?

Servers is essentially computers with network features.

Characteristics

- Durable hardware
- Reliable and fail-safe setup
- Enhanced performance
- With **web server software** installed, e.g.
  - Windows Server
  - NGINX
  - Apache Web Server

NGINX

APACHE HTTP SERVER PROJECT

</talentlabs>

Buy or Rent?

In the old days, most companies buy and set up their own servers

Nowadays, people want to rent servers instead, for cost saving.

Cloud

- Cloud is basically a server farm for people to rent all sorts of servers

- Additional services provided by Cloud:
  - Security
  - Fast network
  - 99.99% uptime
  - Back-up and fail-safe
  - Optimized servers for different purposes

</talentlabs>

Why Cloud?

01 No upfront cost

02 On-demand

03 Minimal Setup

</talentlabs>

Major Cloud Providers

by Google

by Amazon

by Microsoft

</talentlabs>

Major Cloud Providers

by Google

by Amazon

highest market share worldwide

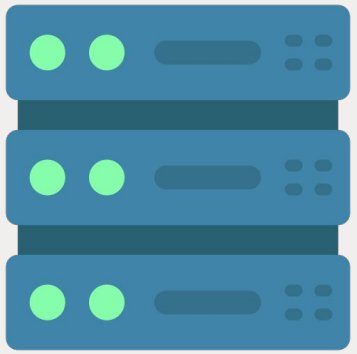
by Microsoft

</talentlabs>

Common Cloud Offerings

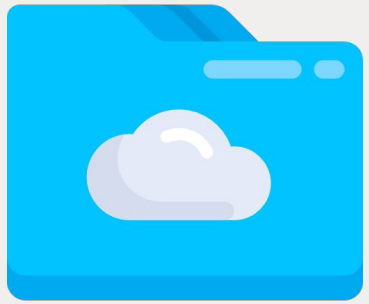
</talentlabs>

3 Common Types of Common Cloud Offerings



Virtual Machine
(Basic Virtual Server)

For running any applications
(web server, scheduled job,
data processing job etc.)



Object Storage
(File Storage)

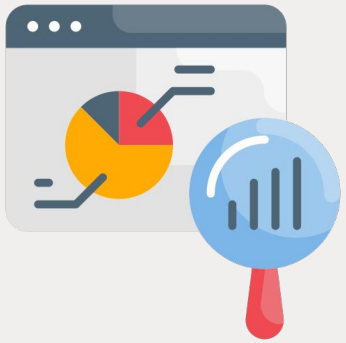
For storing file and object data
(images, documents etc.)

Database

For storing application data
(user database, app data etc.)

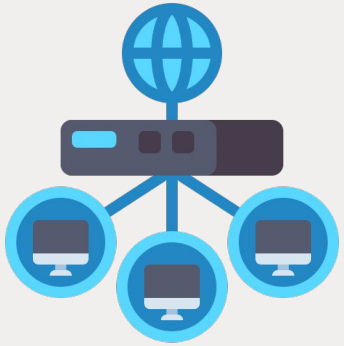
</talentlabs>

Additional Offerings



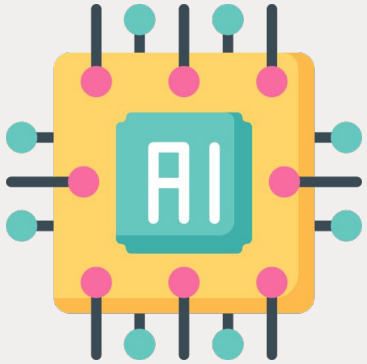
Monitoring

For monitoring all your servers



Load balancing and Scaling

For making sure your servers can handle unlimited traffic



AI and Big Data Engine

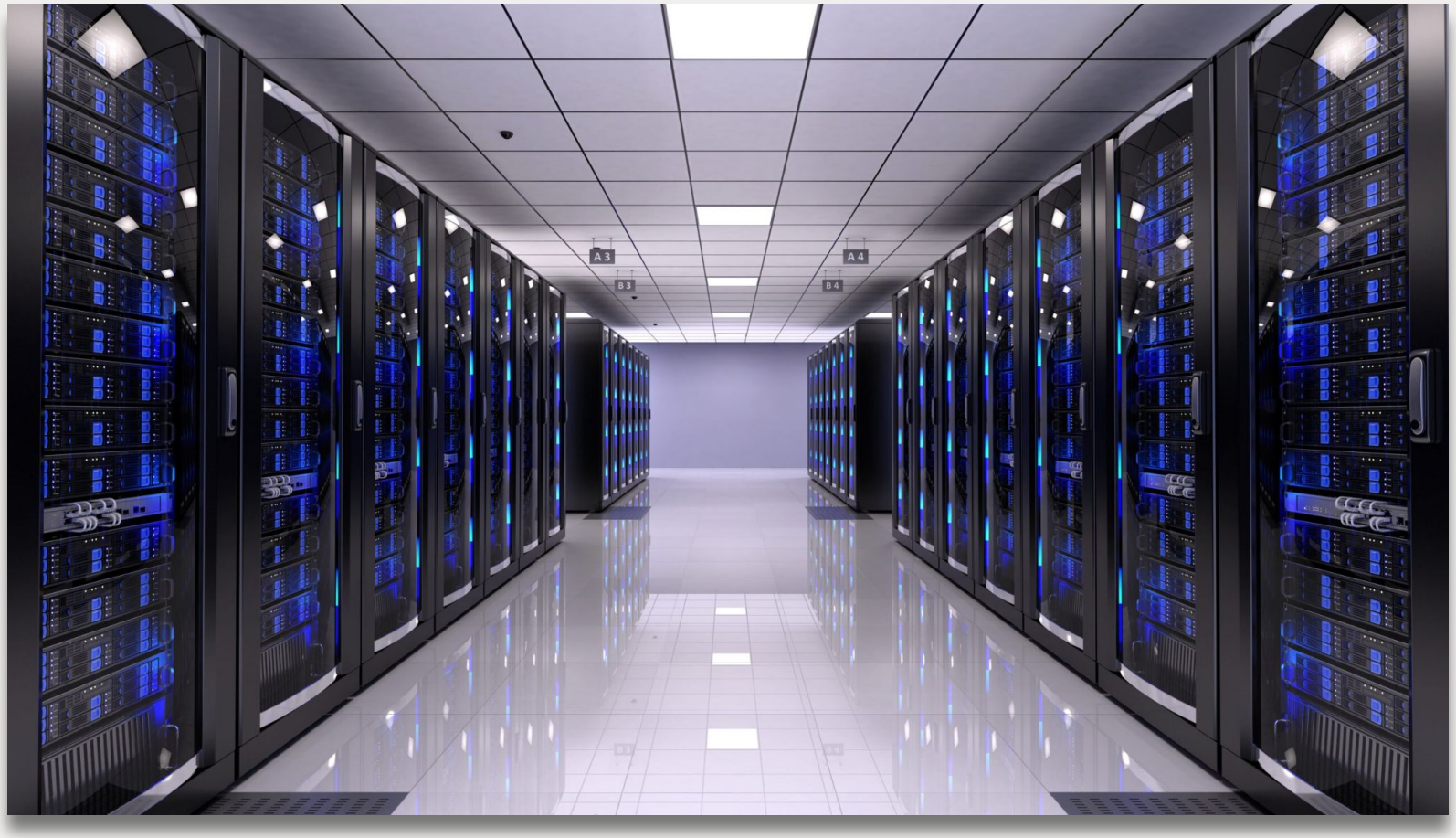
For digging out insights from your data

</talentlabs>

Virtual Machines in the Cloud

</talentlabs>

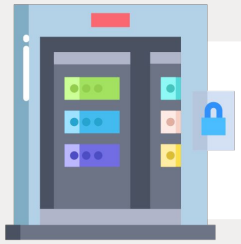
Traditional Server Room



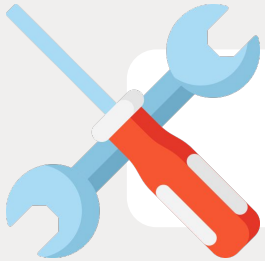
</talentlabs>

Downside of Traditional Server

Expensive



Space Consuming

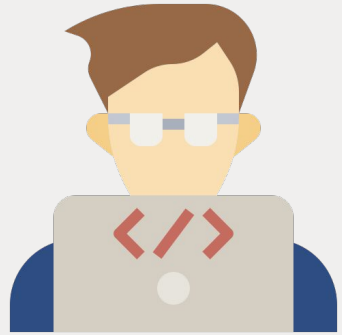


Maintenance

Virtual Machines

Traditional Servers sitting in Cloud Provider Data Center

Remote Login with Internet



Developer sitting anywhere

Elastic Compute Cloud (EC2)

Translation in Human Language

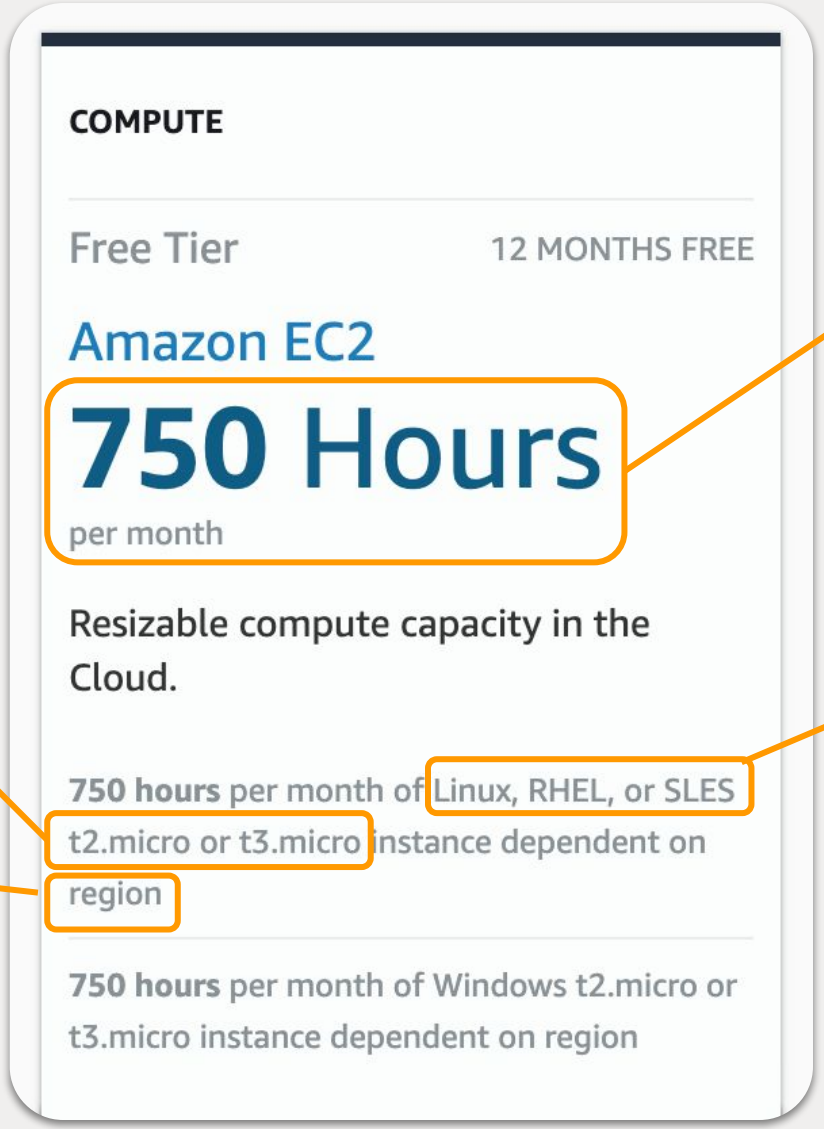
- The simple basic computer server that you can install anything on it.
- aka Virtual Machine

Usage

To run your web application server (can be frontend or backend or both)

</talentlabs>

AWS EC2 for Free



There are only 720 hours a month...

Operating System (Only Linux is free, so you know why we need to learn about command line)

You can pick the the machine you want (different computing power)

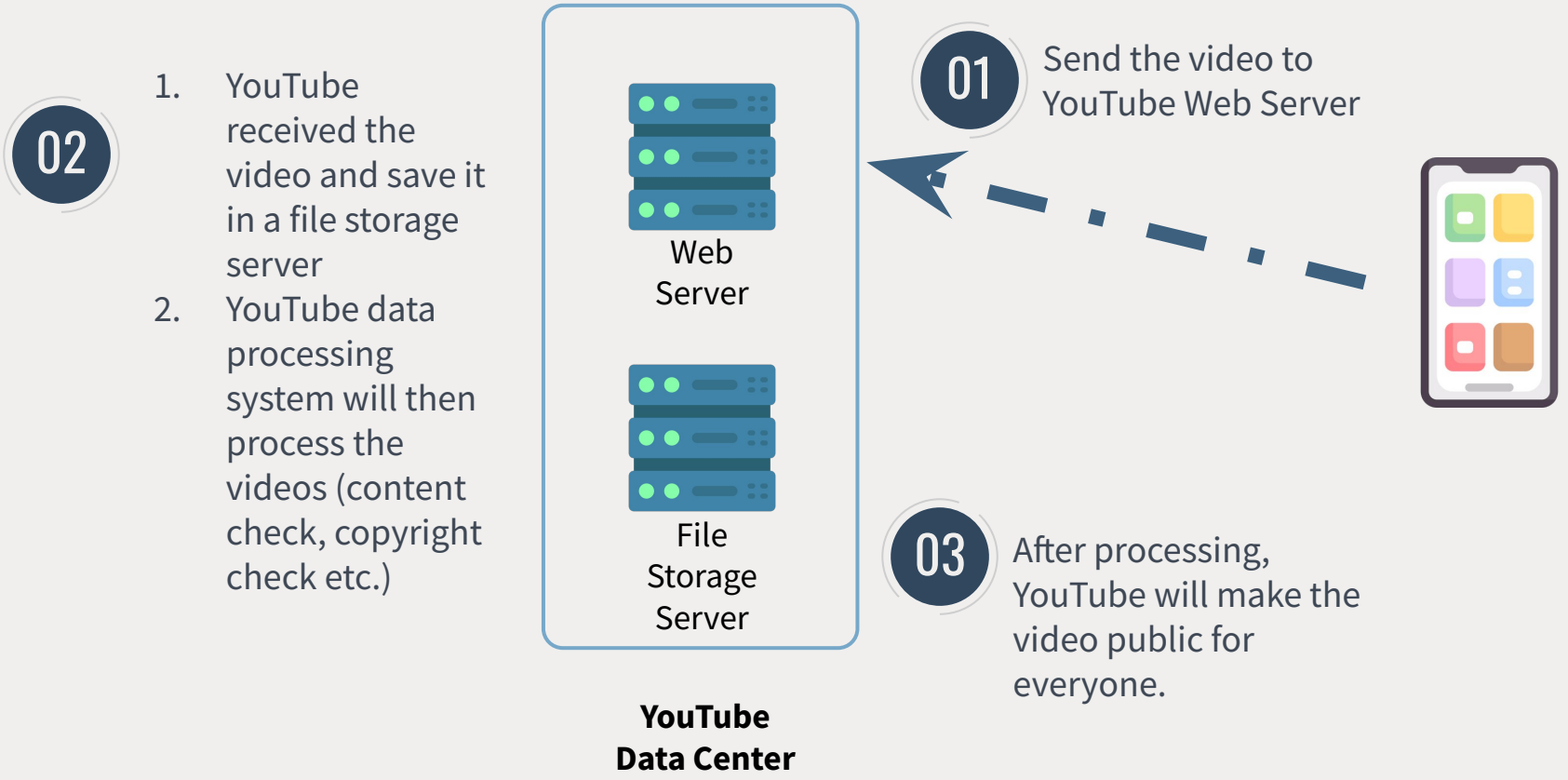
You can pick the location of your server

</talentlabs>

Object Storage in the Cloud

</talentlabs>

How do we upload video to YouTube?



</talentlabs>

Simple Storage Service (S3)

Translation in Human Language

- Storage space for you to store any static files and serve them to your users.
- Each file would have a separate URL/web address for people to access

Usage

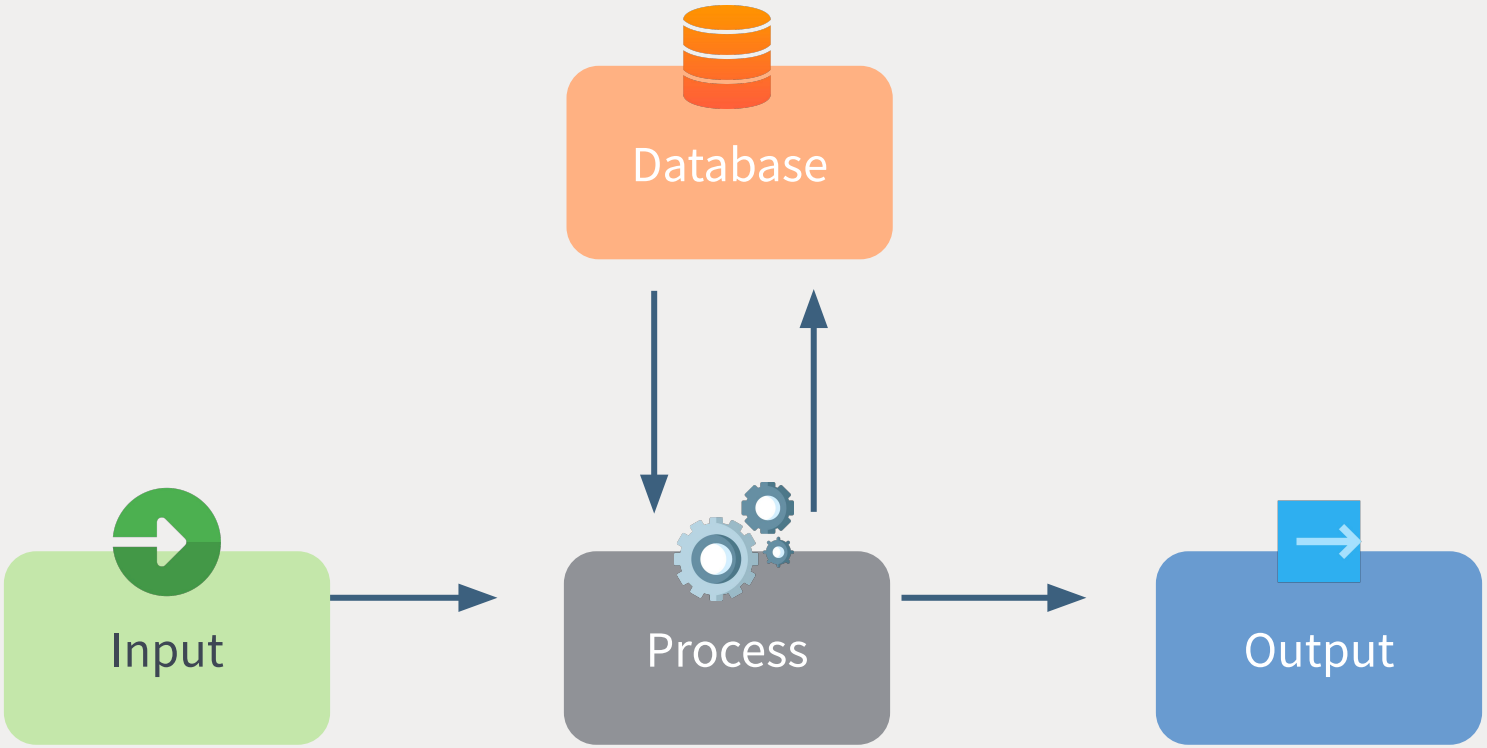
To store the static files of your application (e.g. images, css files, js files, files for download, files uploaded by users)

</talentlabs>

Database in the Cloud

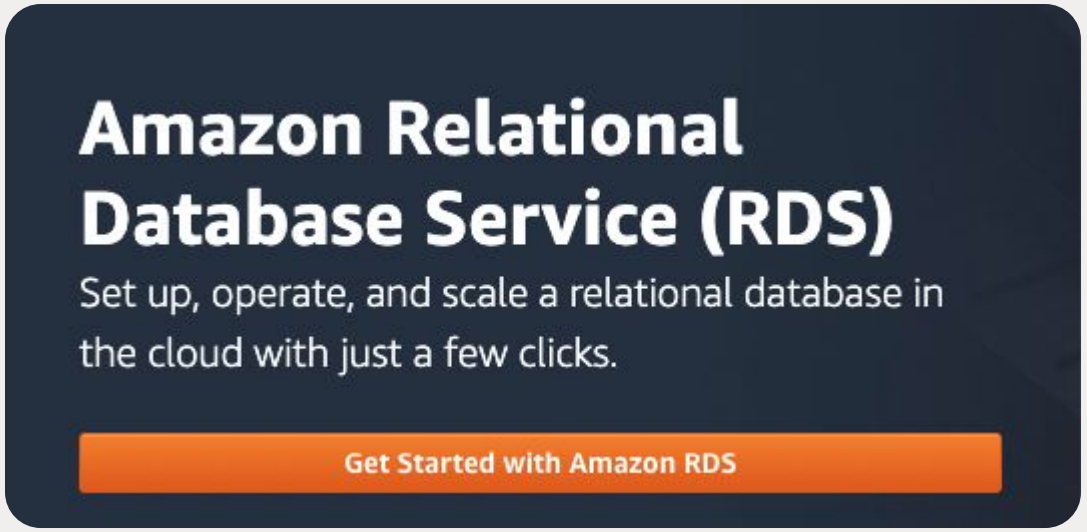
</talentlabs>

Databases Review



</talentlabs>

Relational Database Service (RDS)



Translation in Human Language

- Cloud Database without minimal installation and setup.
- Also, it provides scaling, back-up, failsafe, security settings.
- Supports all major databases including MySQL, PostgreSQL, Oracle etc.

Usage

To store application data, user data and other data in a structured way

</talentlabs>

Case Study

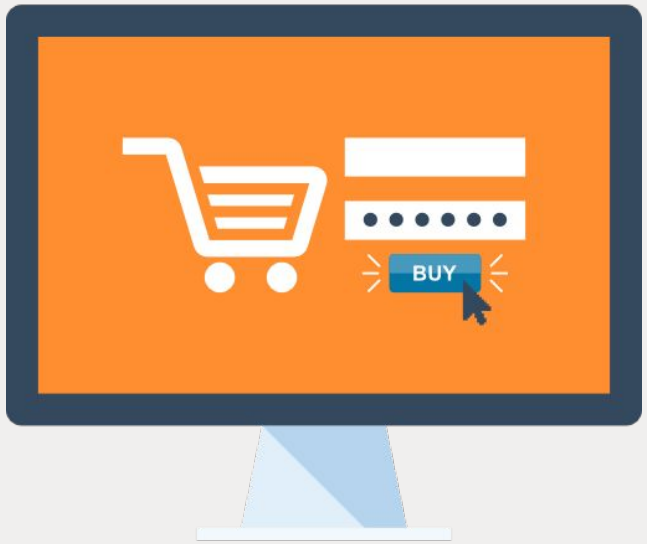
</talentlabs>

</talentlabs>

Jamming eCommerce Platform

ACME Online Shopping Mall

- Online shopping website which offer deals every weekend. Deals are sold at extreme low price at a limited time and quota.
- Over a million users will be rushing in 10 minutes, fighting for the massive discount.
- The website will be down as there are too many users browsing the website at one time (max. 0.5 million in parallel)
- On a normal day, there are only around 10000 people are shopping on the site at the peak hour.



Challenge

Setting up the server to handle 1 million users would be expensive. Also, that huge traffic only happen once a week. Other days, most of the servers are idle.

</talentlabs>

Solution for ACME Online Shopping Mall

Solution

- Set up the online store by using servers(virtual machines) from Cloud providers such as AWS
- Setup auto scale up (i.e. rent more servers) for excess number of users
- Setup auto scale up (i.e. rent less servers) for less number of users

Benefits

- Cloud providers are charging with paid-as-you-go model, so ACME doesn't need to pay for all the servers for whole month.
- If ACME need even more servers in the future (e.g. 2 million of users), they can just rent more virtual machines with just a few clicks.