DATA LAKE

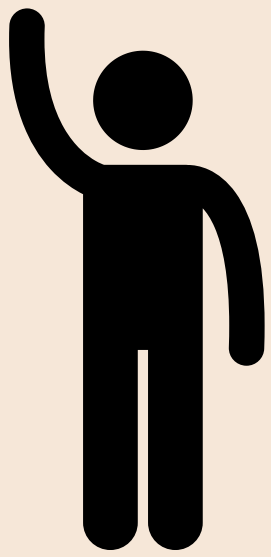
SIMPLIFIED

BY NISCHAY THAPA

HAVE YOU STRUGGLED TO PRODUCE INSIGHTS BECAUSE YOUR DATA IS EVERYWHERE?

LOOK AT ROB'S PROBLEM AND SEE HOW HE FIXES IT!

ROB owns a large restaurant chain



His goal is to implement an **analytics system** so he can get actionable **insights** and **improve** his restaurant's **performance**.

ROB already has an IT infrastructure that collects data from

POS System

Reservation

Social Media

Kitchen and Inventory

Customer data

coming from 5 different venues..

Based on the data he collects, ROB wants to:

Build better customer relationships

Innovate items in the menu

Offer deals

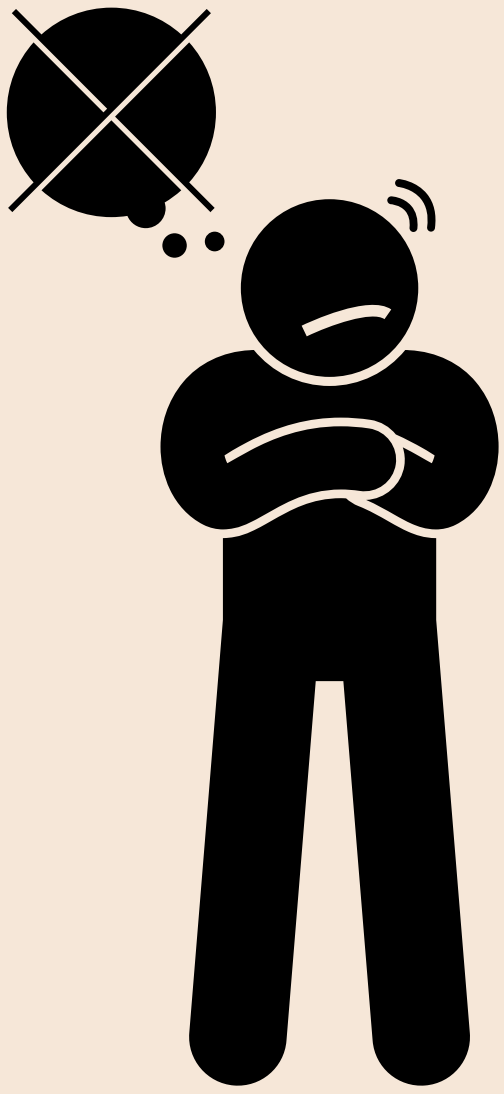
Promote business

Store data for future analysis

BUT ROB's business is growing FAST

And he’s finding it difficult to **store** and **manage** massive amounts of data it’s creating.

All his restaurant's data is stored in file systems or databases and in different formats (JSON, CSV, Avro, Parquet)



and ROB is unable to find the right solution to bring all the data into one place.

ROB decides to talk to his friend JANE, who is a data engineer in a large consulting firm, about his data challenges.

JANE recommends that ROB should set up a data lake architecture in the cloud where he can store and manage all of his data in one place.

JANE's solution helps as :

- ROB would not have to **WORRY** about the volume of the data as the platform can scale up or down as needed

- It would be more **cost-effective** than others types of data storage and processing systems because it will allow him to store data in its **raw**, **unprocessed** form and only pay for the resources he uses.

- It will be **flexible** and will handle a wide **variety** of **data types** and ingestion patterns. ROB can store data in its **original format**, without transforming or aggregation, and can **access** and **process** the data as needed.

- It can integrate with various tools and systems, making it easy to **incorporate data** from different **sources** and use it for various purposes.

- This will allow ROB and his other IT teams to **access** and **use** data, enabling better **collaboration** and **decision-making.**

By using DATA LAKE like ROB
You can:

Store large amounts of data without incurring high costs.

Scale to accommodate a large volume of data.

Have the flexibility to choose how you want to analyse and process.

Apply real-time data processing, allowing you to do in flight processing and analytics

Configure fine-grained access controls and data encryption to ensure the security of your data.

A data lake is at the **heart** of every company's effort to maximise the value of big data.

It is an essential aspect of computing because it allows us to **store**, **organise**, and **access** data as needed.

Before you think of producing any **value** from your data, knowing where and how they are being **stored** is imperative.

RESOURCES

What is a data lake? By AWS

Azure data lake

Accelerated data lake

Serverless data lake framework

Modern data architecture

Click them to find out more!