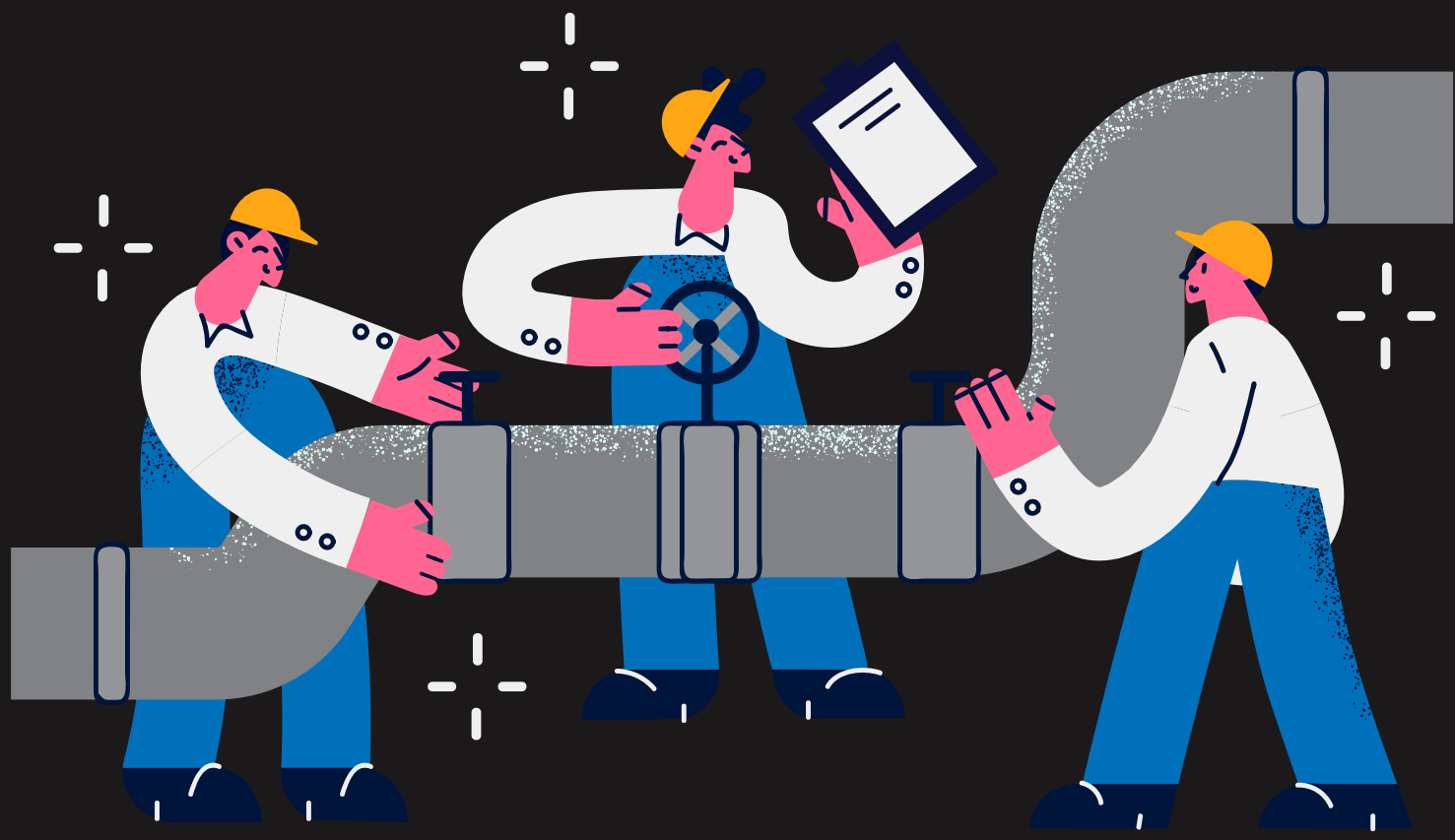
DATA PIPELINE



SIMPLIFIED

BY NISCHAY THAPA

Imagine a retail company that wants to analyse its sales data to understand customer behaviours.

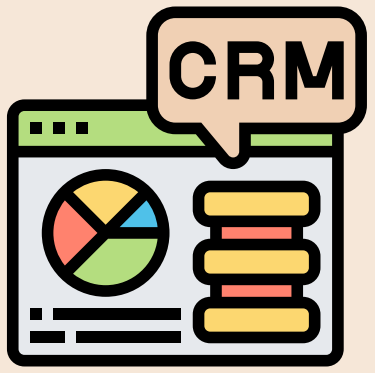
The company collects data and stores it from different systems

POS System

Website

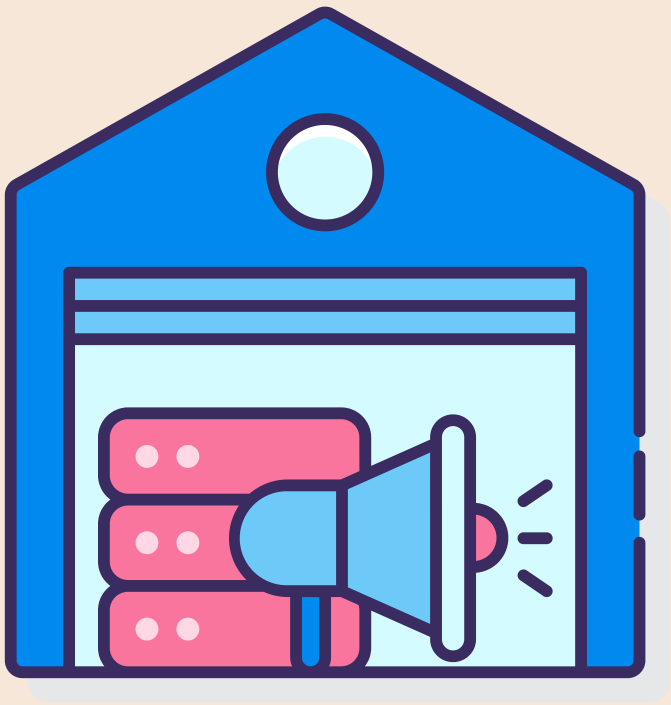
Social Media

CRM System



These are known as 'Sources'

For analysis, they would like to collect and store all this data at one place.



Data Warehouse

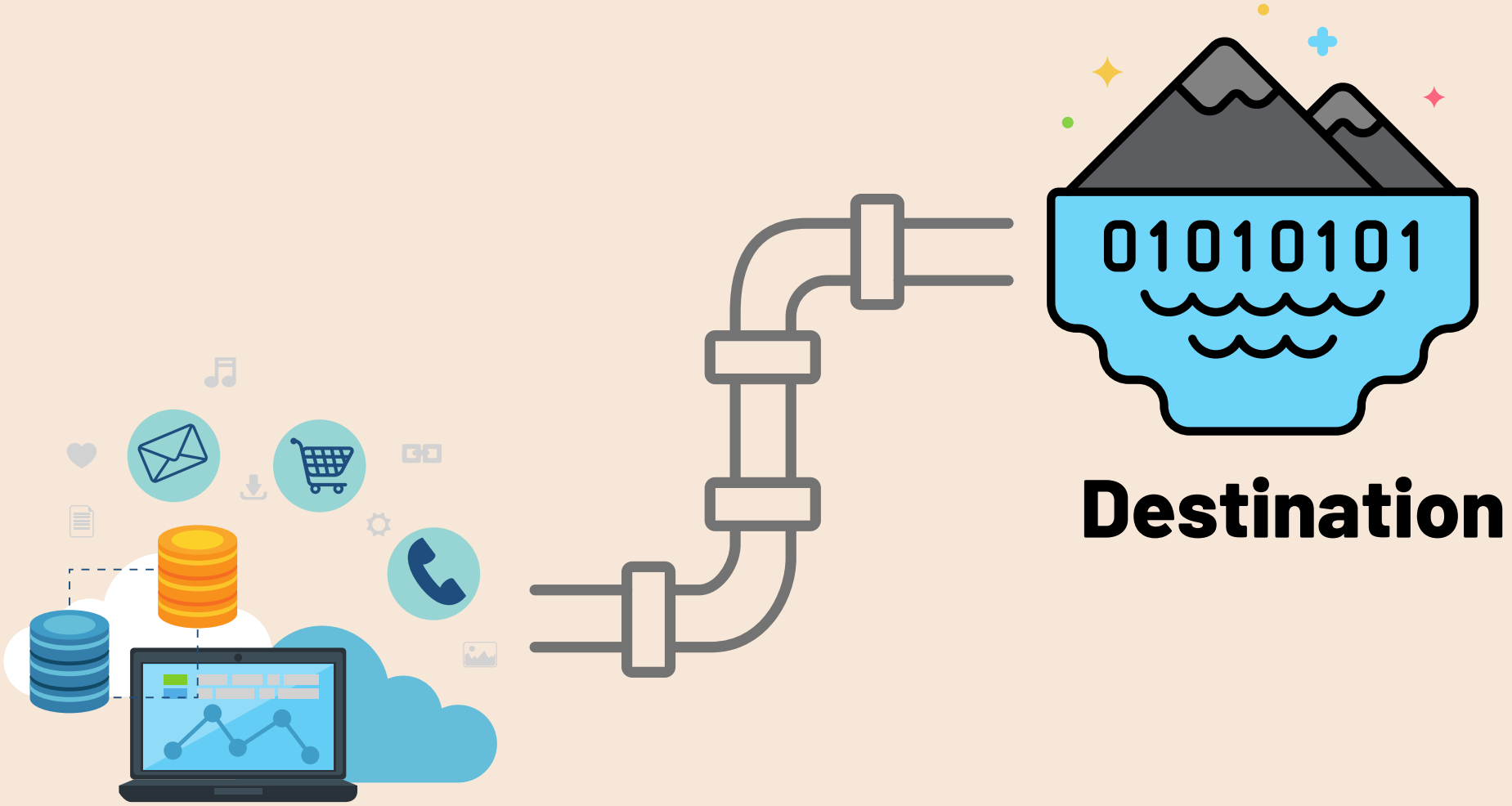
Data Lake

These are known as 'Destinations'

How will they build the connection?

Data Pipeline

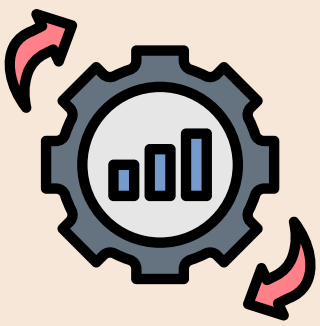
A data pipeline is a series of automated processes that move data from one system or stage to another.



The processes in a data pipeline can include

Extraction

Validation



Transformation

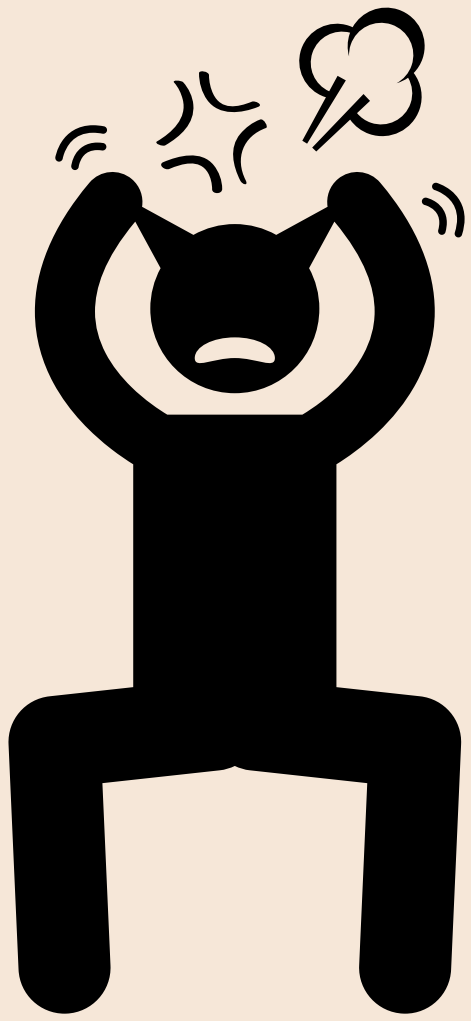
Loading

Quality Checks

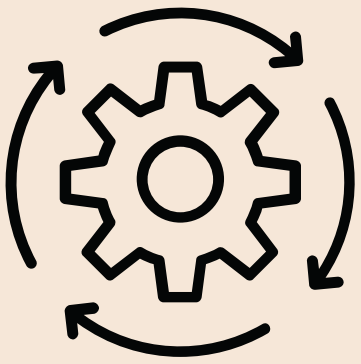
Monitoring

Without a data pipeline,

The company will have to **manually transfer data**, perform **multiple** extracts and transformations on the **same data,** which makes it **difficult** to track changes, and becomes **time-consuming** leading to **poor data quality** and **insights.**

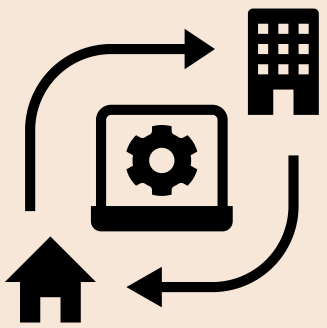


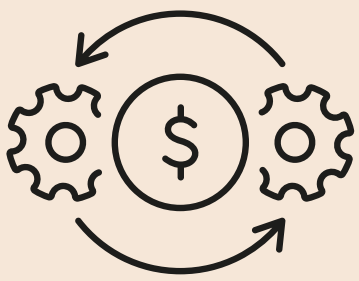
With data pipelines, they can



Automate data flows

Have flexible integration



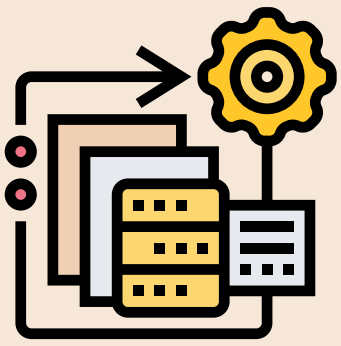


Be cost-effective

Produce better insights

Maintain data consistency

Common types of data pipeline

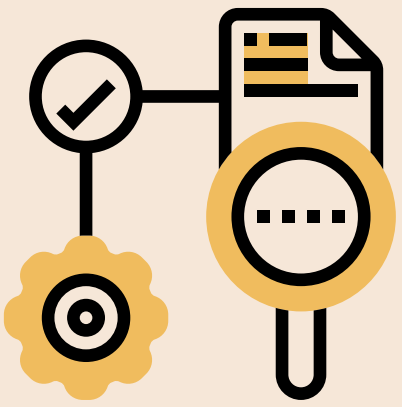


Batch Data Pipeline:

- Processes data in large chunks
- At specific intervals.
- Used for non-time-sensitive data

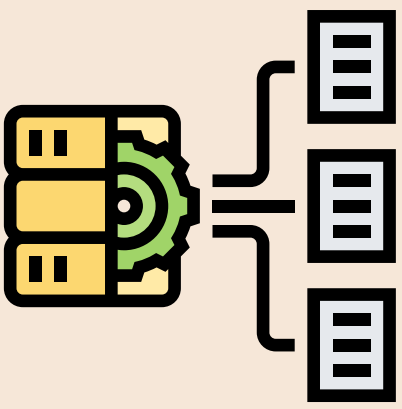
Streaming Data Pipeline:

- Processes data in real-time
- Commonly used for time-sensitive data
    - financial transactions
    - social media feeds.



ETL Pipeline:

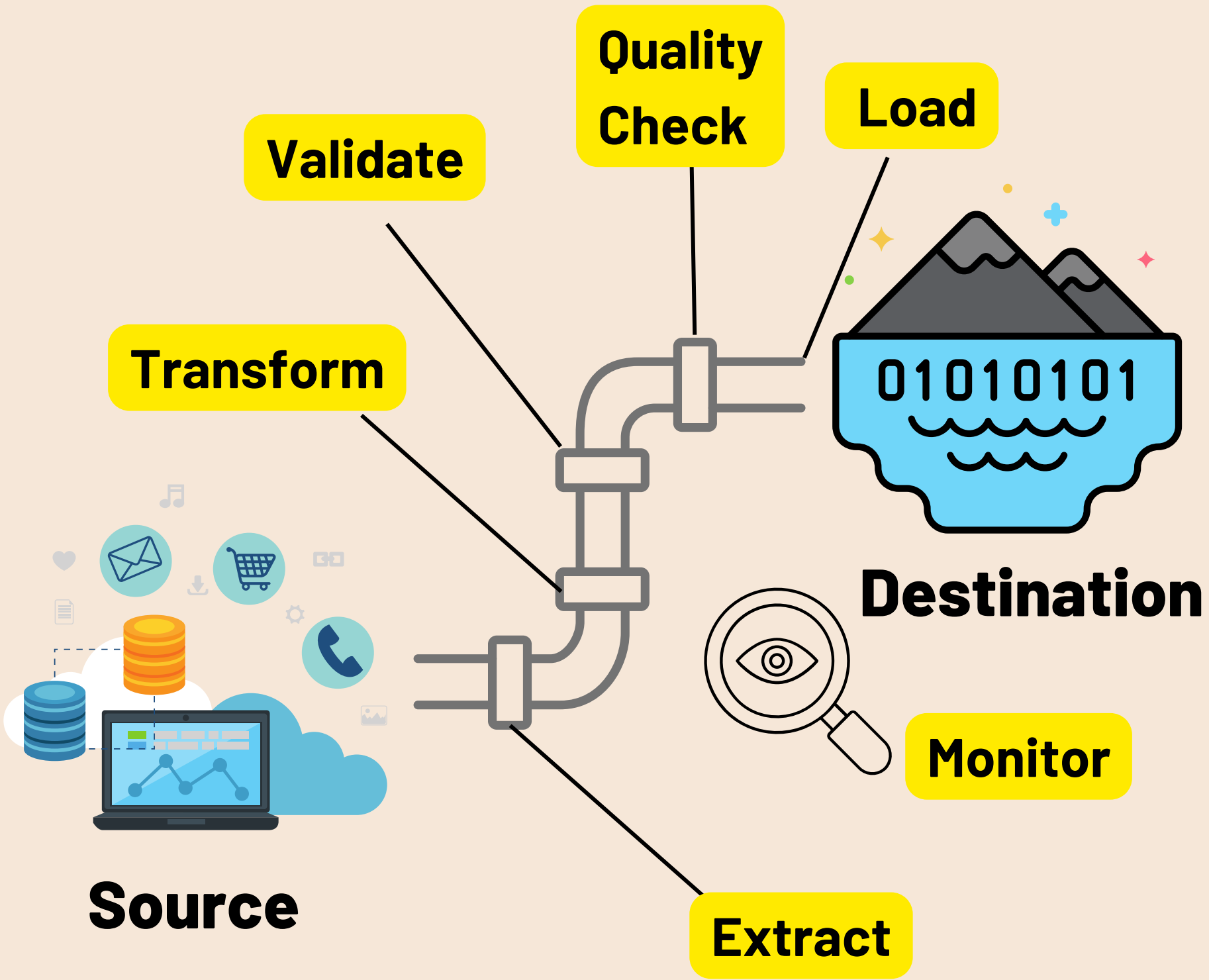
- **E**xtracts data from various sources
- **T**ransforms it
- **L**oads it into the destination system.



ELT Pipeline:

- Extracts data from various sources
- Loads it into a destination system
- Transforms it

Data pipelines are an efficient means for managing and processing data enabling automation, improved governance, and providing accurate insights to inform decision-making.



RESOURCES

Data Pipeline | IBM

Design a data pipeline

Build a data pipeline | GCP

Build a data pipeline | Azure

Build a data pipeline | AWS

Click them to find out more!