# REQUIREMENT FROM (OUR SIMULATED) CUSTOMER

* Launching new data-driven campaign
* Main advertising channel: **YouTube**
* Initial questions to answer:
* "How to categorize videos, based on their comments and statistics"
* "What factors affect how popular a YouTube video will be"

# Goals and Success Criteria

## Data Ingestion

* Ingest data, one-offs and incrementally

## Data Lake

* Design and build a new Data Lake architecture

## AWS Cloud

* AWS as the cloud provider

## ETL Design

* Extract, transform and load data efficiently

## Scalability

* The data architecture should be scaled efficiently

## Reporting

* Build a Business Intelligence tier, incl. Dashboards

Steps:

1. Build IAM user, name “AarshMehtani\_JI”
2. Log in through IAM user.
3. Configure the aws through cmd.

A computer screen with white text

Description automatically generated

1. Get the data, and create s3 bucket, for our landing bucket. Copy the data to s3, using aws cli

**aws s3 cp . s3://de-youtube-raw-ap-south-1/youtube/raw\_statistics\_reference\_data/ --recursive --exclude "\*" --include "\*.json”**

****