Automated Social Media Data Ingestion Pipeline

DSCI-6007-01

Team 1

OUR TEAM

Hima Sai Kuruba
Data visualization and
Story Telling

Srikanth Thota
Data processing and data handling

Diwas Chaulagain
Team leader and data source arrangements

BUSINESS PROBLEM 🔍

We are the data engineering team of a marketing company which looks after the brand images of large companies. Pfizer, a large pharmaceutical company, wants to know how positively people think about them. So, our company has the task to determine how people think about them and doing what actions causes the users to think positively or negatively. Our company has a team of data scientists whose work is to do modelling and predictions. We as data engineers, will only focus on the part of data gathering, ingestion, storage, processing and visualization. Doing so we as data engineers are also focusing on data's excellence, reliability, security , efficiency, cost-optimization and sustainability.

OUR PROPOSED SOLUTION

A robust data engineering pipeline that takes in the data from the API to securely store and process the data

Analyze the data to provide insights to the company about the positives and negatives about the opinion of people towards the company
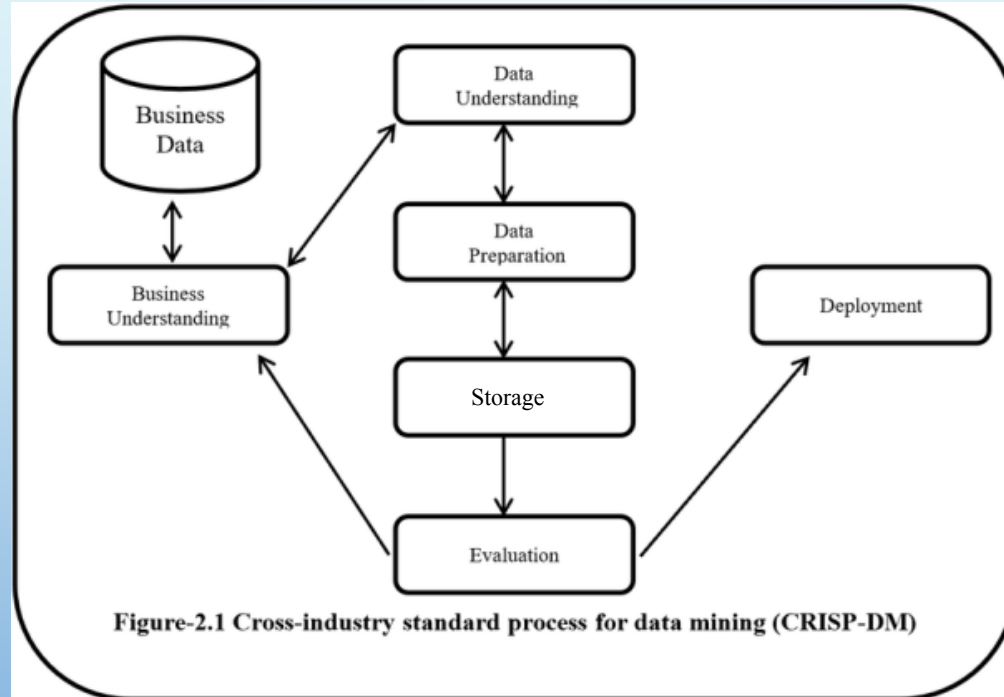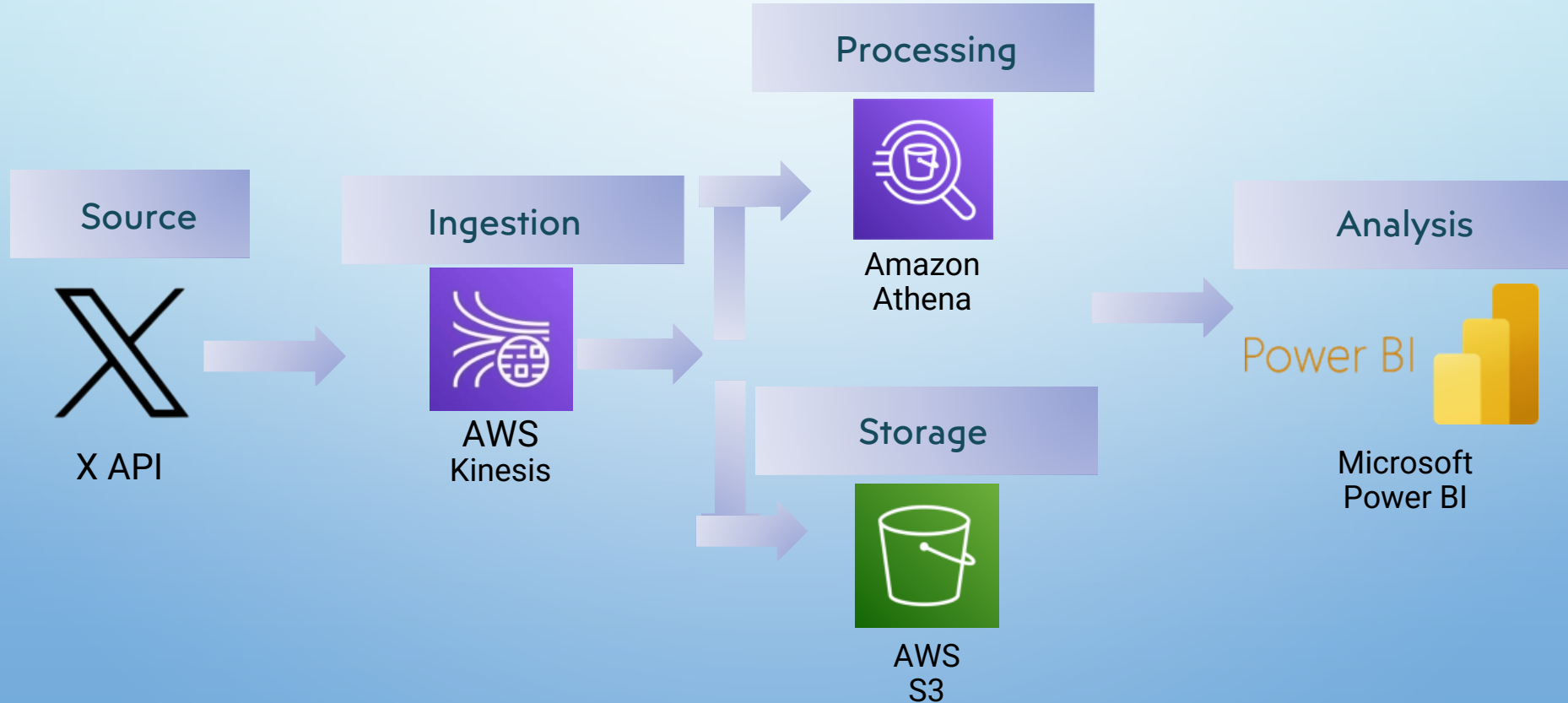
DATA SOURCE

X API

https://developer.twitter.com/en/portal/dashboard

CRISP – DM METHODOLOGY

Figure-2.1 Cross-industry standard process for data mining (CRISP-DM)

Source: data-science-blog.com

TOOLS AND FLOW OF PROJECT

Source

X API

Ingestion

AWS Kinesis

Processing

Amazon Athena

Storage

AWS S3

Analysis

Power BI

Microsoft Power BI

SUMMARY AND RECAP

❖ Our team has been tasked to prepare a robust data engineering pipeline for brand image analysis.

❖ We have got the X API as the data source and we have decided to use AWS tools like Kinesis, Athena and S3 along with Power BI for data analysis.

❖ All our team members have been assigned specific roles which will be followed to achieve the desired results.