Project Design Phase-II
Technology Stack (Architecture & Stack)

Date	31 January 3035
Team ID	LTVIP2025TMID49338
Project Name	iRevolution: a data-driven exploration of apple's iphone impact in India using tableau
Maximum Marks	4 Marks

Technical Architecture:

The Deliverable shall include the architectural diagram as below and the information as per the table1 & table 2

Example: Order processing during pandemics for offline mode

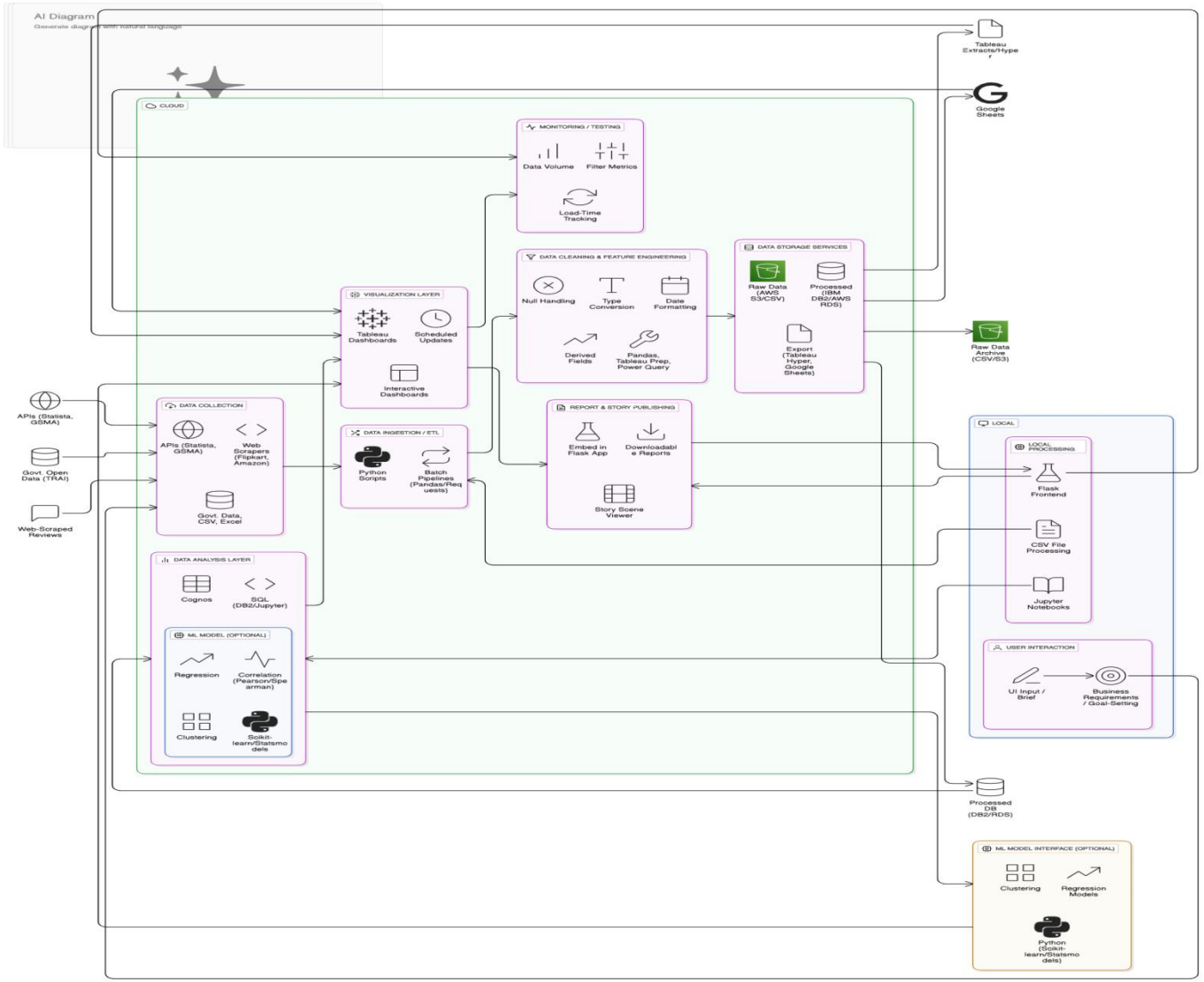
AI Diagram
Generate diagram with natural language
CLOUD
MONITORING / TESTING
Data Volume
Filter Metrics
Load-Time Tracking
DATA CLEANING & FEATURE ENGINEERING
Null Handling
Type Conversion
Date Formatting
Derived Fields
Pandas, Tableau Prep, Power Query
DATA STORAGE SERVICES
Raw Data (AWS S3/CSV)
Processed (IBM DB2/AWS RDS)
Export (Tableau Hyper, Google Sheets)
VISUALIZATION LAYER
Tableau Dashboards
Scheduled Updates
Interactive Dashboards
DATA COLLECTION
APIs (Statista, GSMA)
Web Scrapers (Flipkart, Amazon)
Govt. Data, CSV, Excel
DATA INGESTION / ETL
Python Scripts
Batch Pipelines (Pandas/Requests)
REPORT & STORY PUBLISHING
Embed in Flask App
Downloadable Reports
Story Scene Viewer
DATA ANALYSIS LAYER
Cognos
SQL (DB2/Jupyter)
ML MODEL (OPTIONAL)
Regression
Correlation (Pearson/Spearman)
Clustering
Scikit-learn/Statsmodels
APIs (Statista, GSMA)
Govt. Open Data (TRAI)
Web-Scraped Reviews
Tableau Extracts/Hyper
Google Sheets
Raw Data Archive (CSV/S3)
LOCAL
LOCAL PROCESSING
Flask Frontend
CSV File Processing
Jupyter Notebooks
USER INTERACTION
UI Input / Brief
Business Requirements / Goal-Setting
Processed DB (DB2/RDS)
ML MODEL INTERFACE (OPTIONAL)
Clustering
Regression Models
Python (Scikit-learn/Statsmodels)

Table-1 : Components & Technologies:

S.No	Component	Description	Technology
1.	Data Collection	Ingest data from APIs, web scraping, and cloud storage	APIs, Web Crawlers, Cloud Buckets (GCS, S3)
2.	Data Ingestion / ETL	Data Ingestion / ETLExtract, transform, load data into the system	Apache NiFi, Airflow, custom ETL scripts
3.	Data Storage (Raw)	Store raw collected data	Cloud Storage, BigQuery, PostgreSQL
4.	Data Processing	Clean and transform data for modeling	PySpark, Pandas, Dataflow
5.	Feature Engineering	Create meaningful features for ML models	Python (pandas/sklearn), Feature Store
6.	Machine Learning Training	Train ML models with processed data	scikit-learn, TensorFlow, PyTorch, Vertex AI
7.	Model Registry / Tracking	Track experiments and models	MLflow, Vertex AI, TensorBoard
8.	Model Deployment	Serve models for inference	Vertex AI, Flask API, FastAPI, Kubernetes
9.	Visualization Layer	Create dashboards and visualizations	Tableau, Power BI, Looker
10.	Local Environment	Jupyter Notebooks and local model testing	Jupyter, Python (locally run)
11.	User Interaction	User interface for model feedback & results	UI/UX Frontend, Web App, Feedback APIs

Table-2: Application Characteristics:

S.No	Characteristics	Description	Technology
1.	Scalable Infrastructure	Supports scaling for data, users, and ML models using cloud infrastructure	Google Cloud Platform (GCP), Kubernetes
2.	Modular Design	Divided into independent modules like ETL, modeling, visualization	Microservices Architecture, REST APIs
3.	Real-time & Batch Processing	Handles both streaming and batch data workflows	Apache Beam, Dataflow, BigQuery
4.	End-to-End Automation	From data ingestion to model serving and monitoring	Airflow, MLflow, CI/CD Pipelines
5.	Secure & Compliant	.Includes secure access, data encryption, and audit logging	IAM, VPC, Cloud KMS, Compliance tools