

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

Date	03 November 2022
Team ID	Team -592110
Project Name	Wholesale Customer Segmentation
Maximum Marks	4 Marks

**Technical Architecture:**

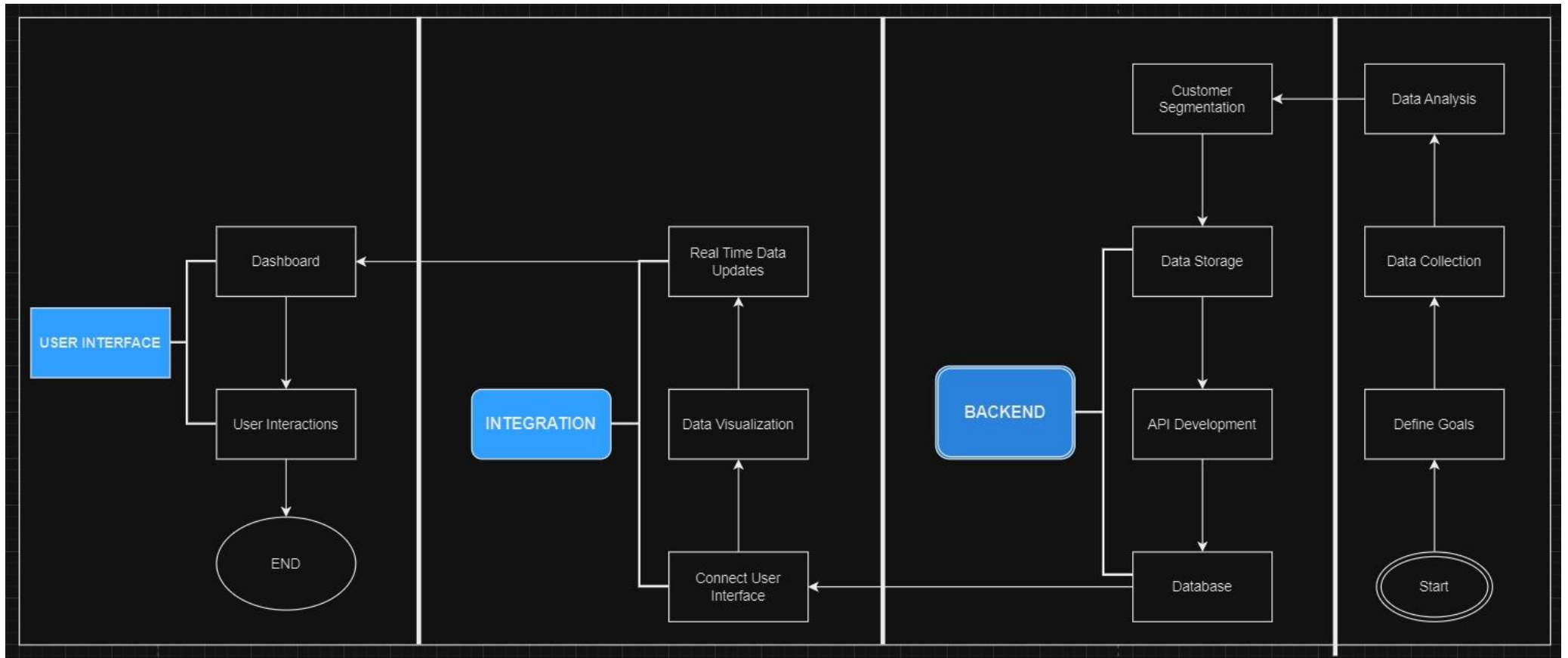
The Deliverable shall include the architectural diagram as below and the information as per the table1 & table

**2 Example: Order processing during pandemics for offline mode**

**Reference:** <https://developer.ibm.com/patterns/ai-powered-backend-system-for-order-processing-during-pandemics/>

**Guidelines:**

1. Include all the processes (As an application logic / Technology Block)
2. Provide infrastructural demarcation (Local / Cloud)
3. Indicate external interfaces (third party API's etc.)
4. Indicate Data Storage components / services
5. Indicate interface to machine learning models (if applicable)



**Table-1 : Components & Technologies:**

S.No	Component	Description	Technology
1.	Data Collection and Ingestion	Gather data from various sources, including sales records, customer databases, and transaction histories.	ETL Tools: Apache Nifi, Talend Data Sources: Relational Databases (e.g., PostgreSQL), Cloud Data Warehouses (e.g., AWS Redshift)
2.	Hyperparameter Tuning	Experiment with different K values to select the optimal number of clusters.	Hyperparameter Tuning: Grid Search, Random Search
3.	Data Visualization	Create interactive dashboards and reports to visualize customer segments.	Data Visualization: Tableau, Power BI Dashboard Development: Custom Web Applications (e.g., Flask, Django)
4.	File Storage/ Data	Storage of raw and processed data	databases (e.g., MongoDB) - Cloud storage (e.g., AWS S3)
5.	Frame Work	Used to Create a web Application, Integrating Frontend and Back End	Python Flask, Django etc
6.	Machine Learning	Implement the K-Means clustering algorithm for customer segmentation.	Machine Learning Frameworks: scikit-learn, TensorFlow, Keras
7.	Deployment(cloud)	Deploy the solution on cloud platforms (AWS, Azure, Google Cloud) for scalability and accessibility.	Cloud Services: AWS, Azure, Google Cloud Load Balancers and Auto-Scaling

**Table-2: Application Characteristics:**

S.No	Characteristics	Description	Technology
1	Data Preprocessing	Data cleaning, feature selection, normalization, encoding, PCA	Data Preprocessing: Python (pandas, scikit-learn), R Feature Engineering: Python.
2	Data Security & Compliance	Stringent measures to protect customer data and ensure compliance	Data Encryption, Access Control, Compliance Frameworks (e.g., GDPR)
3	Scalability	Easily adapt to changing data demands.	Auto-scaling mechanisms
4	Business Goals	Segment wholesale customers to tailor strategies	Technology used
5	Visualization Tools	Interactive dashboards, Tableau, Power BI	Data Visualization: Tableau, Power BI Dashboard Development: Custom Web Applications (e.g., Flask, Django)

**References:**

<https://ijsrst.com/paper/8152.pdf>

<https://www.kaggle.com/datasets/aggle6666/wholesale-customer-segmentation-dataset/discussion>

<https://www.aaarl.ca/post/customer-segmentation-for-wholesalers>

<https://rpubs.com/NikRoy/778968>

[https://www.researchgate.net/profile/Mehnaz-Tabassum/publication/323771551\\_Comparative\\_Performance\\_Of\\_Using\\_PCA\\_With\\_K-Means\\_And\\_Fuzzy\\_C\\_Means\\_Clustering\\_For\\_Customer\\_Segmentation/links/5aa9f8680f7e9b88266f6e35/Comparative-Performance-Of-Using-PCA-With-K-Means-And-Fuzzy-C-Means-Clustering-For-Customer-Segmentation.pdf](https://www.researchgate.net/profile/Mehnaz-Tabassum/publication/323771551_Comparative_Performance_Of_Using_PCA_With_K-Means_And_Fuzzy_C_Means_Clustering_For_Customer_Segmentation/links/5aa9f8680f7e9b88266f6e35/Comparative-Performance-Of-Using-PCA-With-K-Means-And-Fuzzy-C-Means-Clustering-For-Customer-Segmentation.pdf)