DHVANILKUMAR PATEL

NATURE PICK

DATA ARCHITECT

2024

Contents

INTRODUCTION

Nature Pick's is dedicated to producing and delivering fresh, high-quality products to our customers. By integrating advanced IoT devices into our farming processes, we ensure efficiency, sustainability, and exceptional quality from the field to your hands. Our products are available both in-store and through our user-friendly mobile application, providing convenience and accessibility to meet the needs of our valued customers.

MISSION STATEMENT

Nature Pick's is dedicated to revolutionizing the agricultural industry by harnessing cutting-edge technology and sustainable practices. Our goal is to provide customers with unparalleled freshness and quality, ensuring a seamless shopping experience both online and in-store. We strive to be leaders in eco-friendly farming, enhancing efficiency and promoting a healthier planet for future generations.

OBJECTIVE

- Quality Assurance: Ensure the highest quality of products from farm to table using IoT devices for real-time monitoring and management.

- Sustainability: Implement eco-friendly farming practices to minimize environmental impact and promote long-term sustainability.

- Customer Accessibility: Provide seamless access to our products through both physical stores and a user-friendly mobile application.

- Innovation: Continuously integrate cutting-edge technology like machine learning and data analytics to optimize operations and enhance customer experience.

- Market Expansion: Expand our reach through strategic marketing and partnerships to deliver our products to a broader audience.

DATA SOURCES
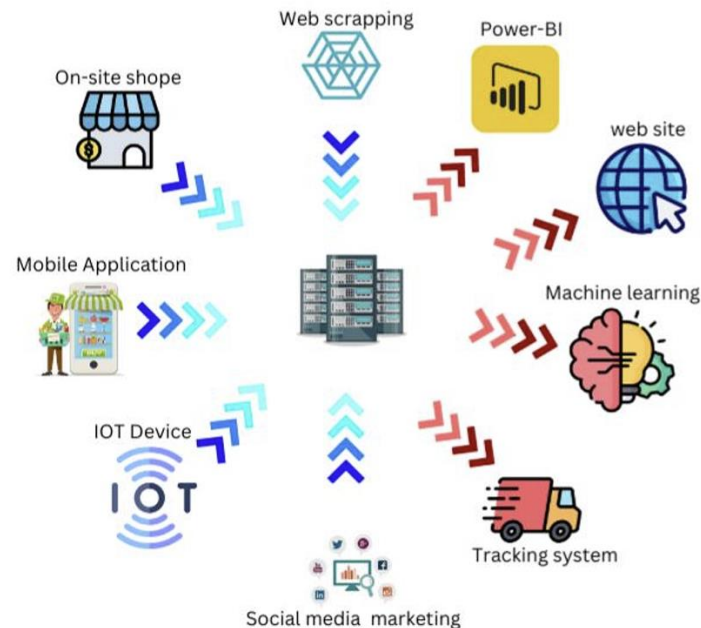
1.Mobile Application

In the mobile application, the sale of farming products like vegetables and fruits is managed in a structured format using an SQL database. The data is stored in well-defined tables such as:

Products Table: Stores product details like product ID, name, category, price, and stock levels.

Orders Table: Tracks order details, including order ID, customer ID, product ID, quantity, total price, and order status.

Customers Table: Contains customer information like customer ID, name, contact details, and address.

Transactions Table: Logs payment details such as transaction ID, payment method, and timestamps.

This structured approach ensures easy querying, updating, and reporting for smooth operations and analytics.

2.on-site system

This system generates SQL data in which it consist of these type of data

Product Information System: Stores product details like name, description, price, images, and specifications.

Customer Relationship Management (CRM) System: Contains customer information, purchase history, and preferences.

Inventory Management System: Tracks stock levels, order history, and supplier information.

Enterprise Resource Planning (ERP) System: Integrates various business functions, including finance, HR, and supply chain.

3.IOT data

1. Sensor Data:

Time-series Data: Continuous measurements of soil moisture and sunlight levels over time.

Event-based Data: Triggered by specific events, such as sudden changes in environmental conditions or sensor failures.

2. System and Logistics Data:

Device Status Data: Information about the health and performance of IoT devices, including battery levels, connectivity status, and sensor readings.

Location Data: GPS coordinates of products and delivery vehicles, providing real-time tracking and monitoring.

Product Information: Details about products, including SKU, quantity, and delivery status.

Log Data: System logs recording events like device activations, data transmissions, and error messages.

Data Format:

While some data, like sensor readings and device status, can be structured in a tabular format, other data, such as log messages and system events, may be semi-structured or unstructured. This semi-structured nature often involves hierarchical data formats like JSON or XML, which can accommodate varying data structures and complexities.

4.web scrapping:

Web scraping extracts structured product data from e-commerce sites like Blinkit and JioMart. Tools identify specific HTML elements to pinpoint desired information, such as product names, prices, and reviews. While challenges like dynamic content and anti-scraping measures exist, techniques like browser automation and proxy rotation help overcome them. This process provides valuable insights into market trends and consumer behavior.

5.social media data

Social media platforms like Facebook and Instagram provide APIs to access valuable marketing data. This data, often in JSON format, includes insights on ad performance, audience demographics, post engagement, and campaign metrics. By leveraging these APIs, businesses can gain deeper insights into their social media strategies, optimize campaigns, and make data-driven decisions to improve marketing ROI.

Outcomes

1.Power BI: Power BI is a powerful tool for your farming app, enabling you to visualize and analyze data related to product sales, inventory management, and tracking systems. It allows you to monitor sales performance by showcasing total revenue, top-selling products, and seasonal

trends through interactive dashboards. Customer insights, such as buying patterns, regional demand, and feedback, can help you enhance customer engagement and personalize offerings. Power BI also aids in inventory management by providing real-time visibility into stock levels, low-stock alerts, and inventory turnover rates, ensuring efficient stock handling and reducing waste. Additionally, tracking system data, such as delivery statuses and logistics performance, can be visualized to improve operations and optimize routes. With these insights, Power BI empowers you to make data-driven decisions for growing your farming business effectively.

2.website:in website we will display our top trending products most selling products. Moreover, we give some coupon code and data according to user trends.

3.Machine Learning: In the farming app, machine learning plays a crucial role in providing product suggestions and optimizing pricing strategies. By comparing the app's product prices with competitor websites through web-scraped data, machine learning algorithms analyze market trends and suggest competitive pricing to attract customers while maintaining profitability. Techniques like RFM (Recency, Frequency, Monetary) analysis are used to segment customers based on their purchasing behavior, identifying high-value customers and tailoring product suggestions to their preferences. Additionally, Market Basket Analysis helps identify product combinations frequently bought together, enabling cross-selling and upselling opportunities. These data-driven approaches enhance customer satisfaction, boost sales, and ensure your farming products remain competitively priced in the market.

LAKEHOUSE ARCHITECTURE
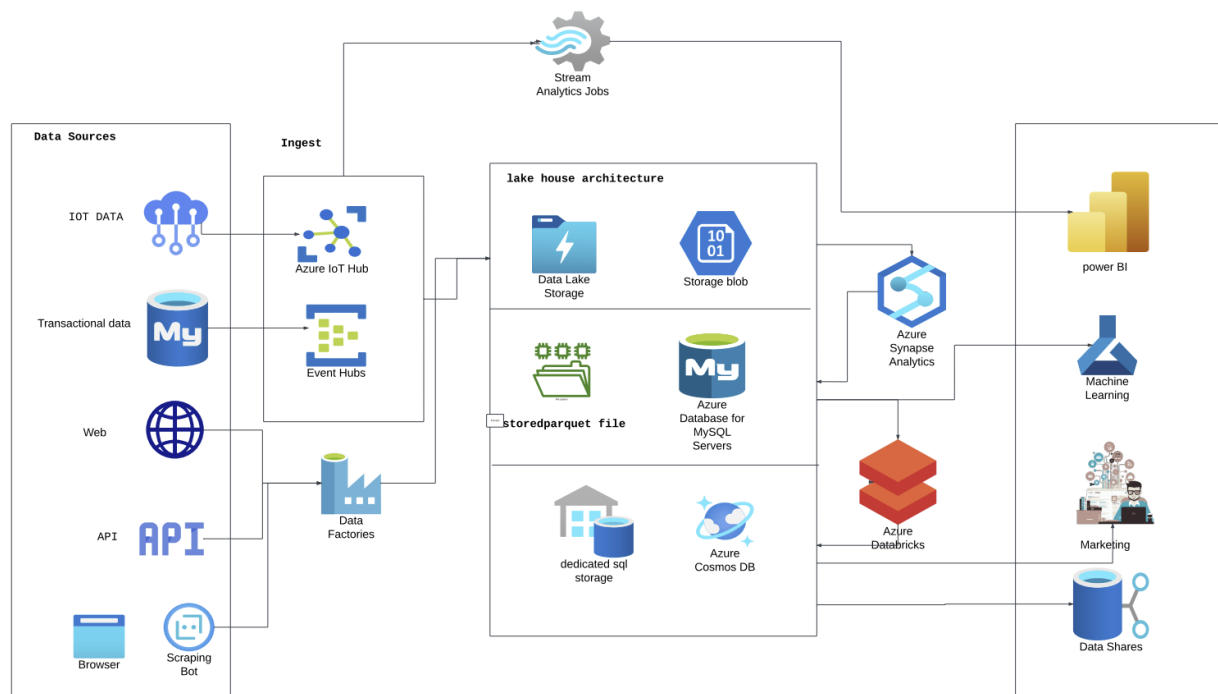
Data Sources and Ingestion

1. **Data Sources**:

   - **IoT Data**: Real-time data from IoT devices (likely used for monitoring farming operations in your case).

   - **Transactional Data**: Operational data generated by business transactions.

   - **Web Scraping Bots**: Collecting data from online sources, such as competitor prices or market trends.

   - **API**: External or internal systems sharing data via API calls.

2. **Data Ingestion**:

   - Data from the sources flows into the system through tools like **Azure IoT Hub** and **Event Hubs**, designed for capturing real-time data streams.

Storage

- **Data Lake Storage**: Acts as the central repository to store raw and processed data in various formats, such as **parquet files** for efficient querying.

- **Azure Blob Storage**: Stores unstructured data.

- **Azure Cosmos DB**: Manages NoSQL data for high-performance applications.

- **Azure Database for MySQL Servers**: Stores relational data.

Processing

- **Azure Synapse Analytics**: Used for data warehousing and advanced analytics.

- **Azure Databricks**: Performs scalable data processing, big data analytics, and supports machine learning workflows.

- **Stream Analytics Jobs**: Processes and analyzes real-time data streams.

- **Azure Data Factory**: Orchestrates data movement and transformation.

Analytics and Usage

- **Machine Learning**: Supports AI/ML models, likely for predictive analytics (e.g., price recommendations or crop predictions).

- **Power BI**: Visualizes data and provides actionable insights for end users.

- **Data Shares**: Enables collaboration and sharing of data with external stakeholders.

- **Marketing**: Likely uses analytics for targeted campaigns.

PIPELINE STRATEGY
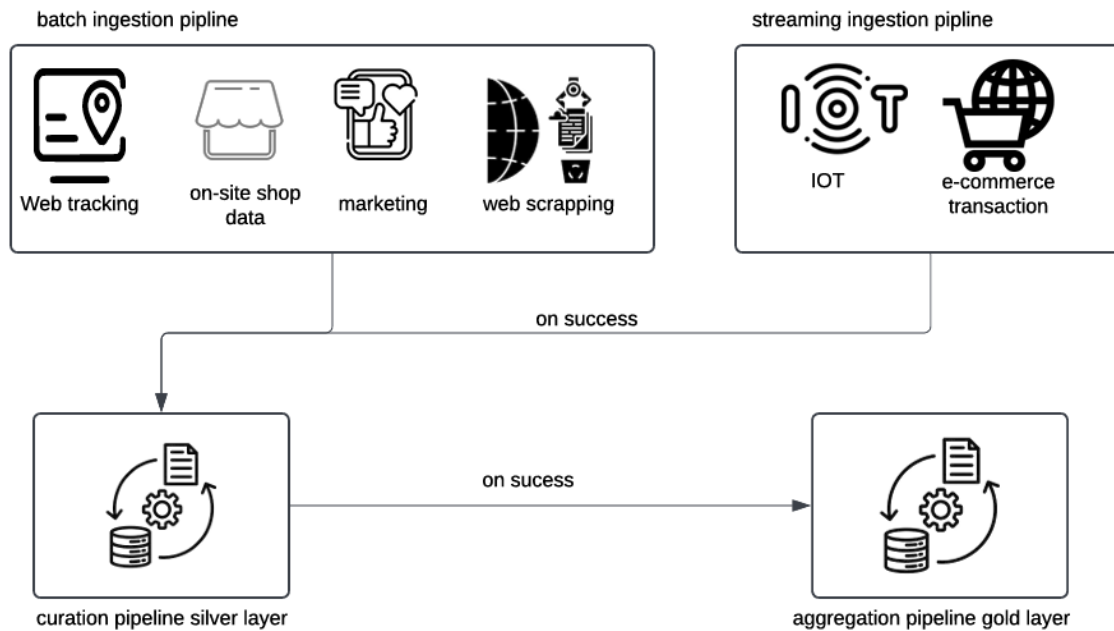
1. Batch Ingestion Pipeline

- **Data Sources**:

    - **Web Tracking**: Collects user behavior and tracking data from the web.

    - **On-Site Shop Data**: Gathers sales, inventory, and operational data from physical stores.

    - **Marketing**: Includes data from campaigns, customer engagement, and social media.

    - **Web Scraping**: Extracts data from other online sources, such as competitor pricing or product information.

- **Processing**:

    - Data from these batch sources is ingested periodically and sent to the **Curation Pipeline Silver Layer**.

2. Streaming Ingestion Pipeline

- **Data Sources**:

    - **IoT**: Real-time data from IoT devices, such as sensors used in farming or store environments.

-      o   **E-Commerce Transactions**: Live transactional data from the company's mobile app or online store.

- **Processing**:

  - o   Real-time data is ingested continuously and also sent to the **Curation Pipeline Silver Layer**.

3. Curation Pipeline Silver Layer

- **Purpose**:

  - o   This stage processes, cleanses, and normalizes the raw data from both batch and streaming pipelines.

  - o   Prepares the data for further analysis by organizing it into a semi-structured format.

- **Output**:

  - o   Once processing is successful, the data moves to the **Aggregation Pipeline Gold Layer**.

4. Aggregation Pipeline Gold Layer

- **Purpose**:

  - o   Aggregates and structures data into a highly curated, analytics-ready format.

  - o   Supports business intelligence, advanced analytics, and machine learning applications.