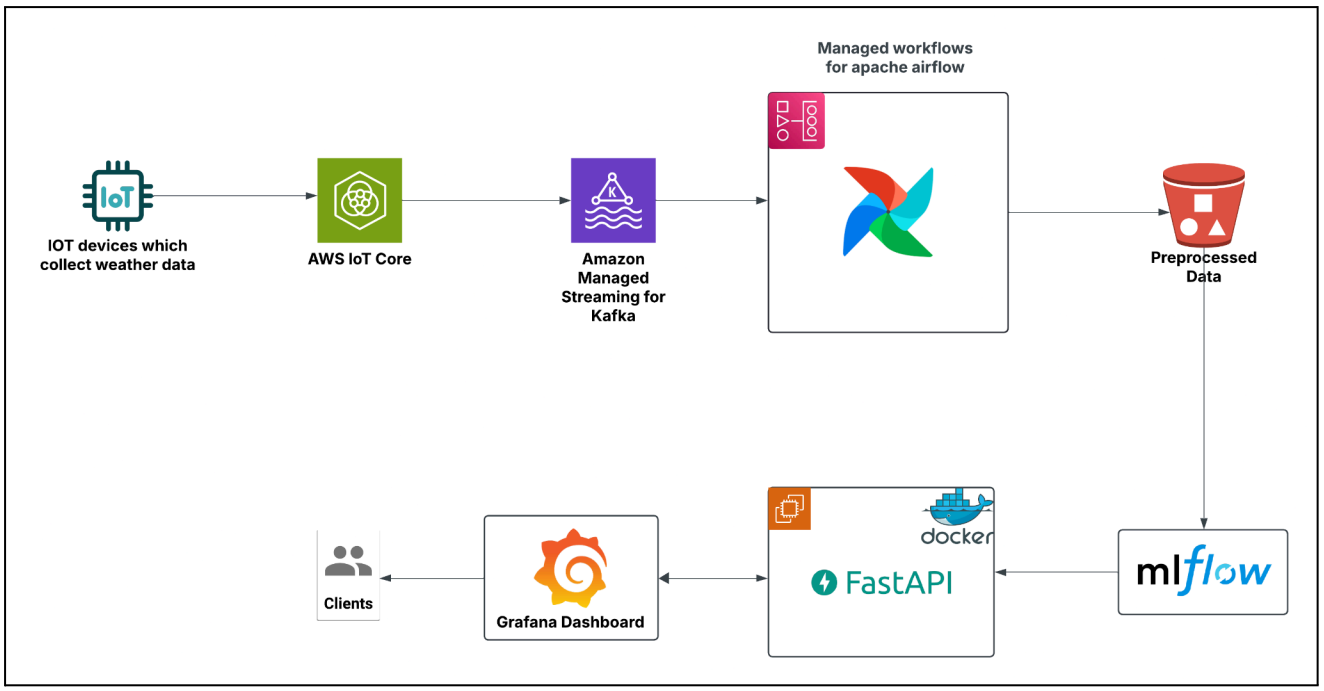
Task 1 : Weather Forecasting

Hybrid Architecture: AWS Services + Open-Source Technologies

To optimize cost, flexibility, and customization, we will combine AWS services with open-source technologies for data ingestion, processing, storage, machine learning, and visualization while ensuring scalability and fault tolerance.



Architecture Breakdown

Proposed architecture follows a structured pipeline, moving from data ingestion to processing, storage, ML modeling, API deployment, and visualization.

1. Data Ingestion	• IoT Devices: Collect real-time weather data (temperature, humidity, wind speed). • AWS IoT Core: Serves as a gateway for IoT devices, securely streaming data to Kafka.

2. Data Streaming & Processing	• **Amazon Managed Streaming for Kafka:**
	  ○ Handles real-time data ingestion.
	  ○ Ensures scalability and fault tolerance for IoT data streams.
	• **Apache Airflow (Managed Workflows for Apache Airflow):**
	  ○ Orchestrates data pipelines.
	  ○ Cleans, validates, and preprocesses incoming data before storing it.

3. Data Storage & Machine Learning	• **Amazon S3 (Preprocessed Data Storage):**
	  ○ Stores clean, preprocessed weather data for model training.
	• **MLflow:**
	  ○ Manages ML experiments, tracks model versions, and ensures reproducibility.
	  ○ Pulls preprocessed data from S3 for model training.

4. Model Deployment & API	• **FastAPI (Dockerized):**
	  ○ Deploys the trained model as a REST API.
	  ○ Provides real-time predictions for clients.

5. Visualization & User Interaction	• **Grafana Dashboard:**
	  ○ Provides users with real-time rain probability visualizations.
	  ○ Connects to FastAPI to fetch predictions.