1)

Part 2

A diagram of a software development process

AI-generated content may be incorrect.

The system follows a modular approach with distinct layers for data ingestion, processing, model training/serving, monitoring, and visualization.

**Data Flow**

1. **Data Sources**
   * IoT sensors providing real-time data at 1-minute intervals
   * Weather APIs for supplementary data
   * Historical database for training and validation
2. **Data Ingestion Layer**
   * Kafka Streams for high-throughput real-time data collection
   * API Gateway to manage access and request limits
   * Data Validation to immediately flag erroneous readings
3. **Processing Layer**
   * Data Cleaning to handle missing values and outliers from malfunctioning sensors
   * Anomaly Detection to identify and handle sensor failures
   * Feature Engineering to prepare data for model training
4. **ML Pipeline**
   * Model Training using LSTM or XGBoost for time-series forecasting
   * Model Registry (MLflow) for versioning and experiment tracking
   * Inference Service to generate 21-day rainfall predictions
5. **Data Storage**
   * Time Series Database (InfluxDB) optimized for sensor readings and forecasts

6 **Monitoring System**

* Model Metrics tracking for accuracy and reliability
* Data Drift detection to identify changes in sensor behaviors