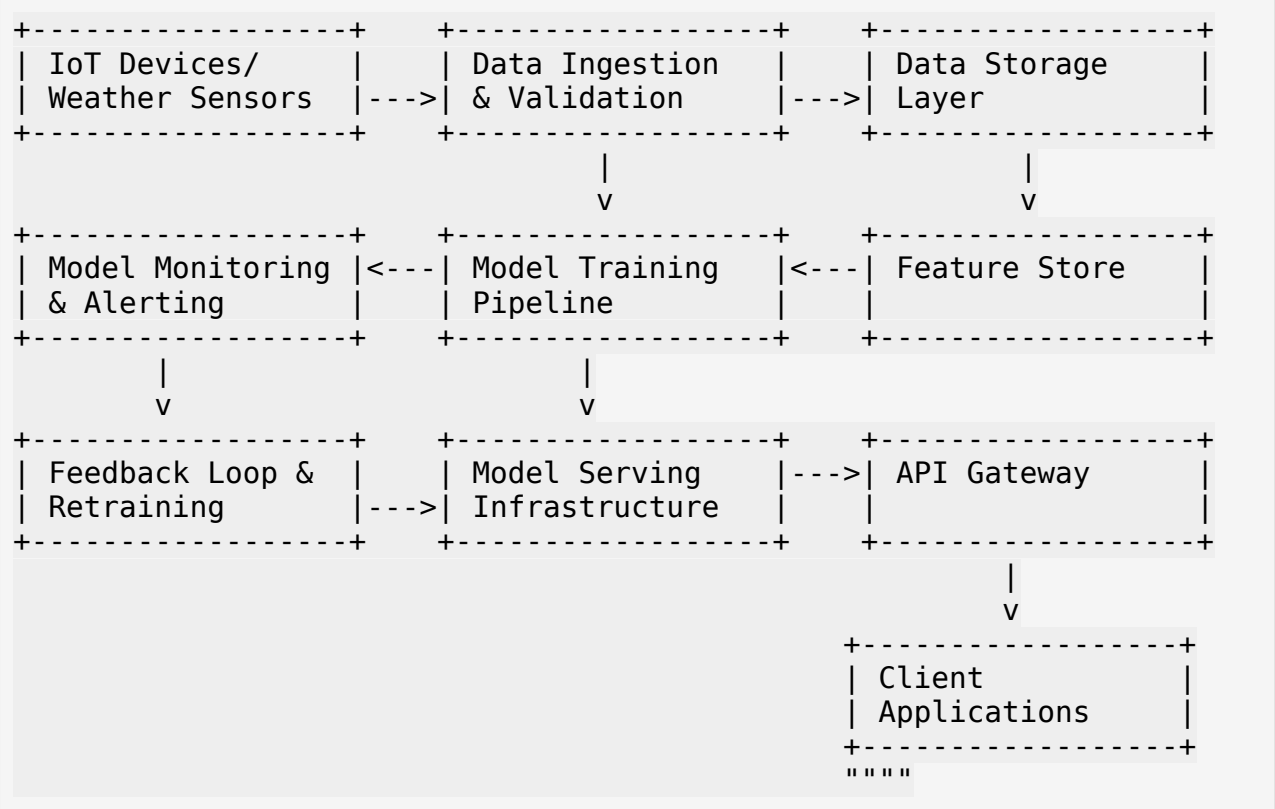
MLOps System Design for Smart Agriculture Weather Forecasting

System Architecture Diagram

"""



"""

Component Descriptions

1. IoT Devices/Weather Sensors

- **Description**: Network of weather sensors deployed across farms collecting real-time data at 1-minute intervals.
- **Functionality**: Captures temperature, humidity, wind speed, pressure, and cloud cover measurements.
- **Fault Tolerance**: Implements heartbeat mechanisms to detect sensor malfunctions.

2. Data Ingestion & Validation

- **Description**: System that receives, validates, and processes incoming sensor data.
- **Functionality**:

  - Performs data quality checks (range validation, missing value detection)
  - Identifies anomalous readings using statistical methods
  - Buffers data to handle connectivity issues
  - Implements retry mechanisms for failed transmissions
- **Error Handling**: Flags suspicious data for human review when needed

3. Data Storage Layer

- **Description**: Scalable database system for storing both raw and processed weather data.
- **Components**:
  - Time-series database for raw sensor readings
  - Data warehouse for aggregated historical data
  - Object storage for model artifacts and backups
- **Data Management**: Implements data retention policies and automated archiving

4. Feature Store

- **Description**: Centralized repository for storing, managing, and serving features.
- **Functionality**:
  - Creates and stores engineered features (e.g., rolling averages, cyclical encodings)
  - Ensures consistent feature transformations between training and inference
  - Caches frequently used features for performance
  - Manages feature versioning and lineage

5. Model Training Pipeline

- **Description**: Automated system for training, validating, and registering models.
- **Components**:
  - Data extraction and preprocessing
  - Hyperparameter optimization
  - Model validation and performance evaluation
  - Model registration and versioning
- **Scheduling**: Runs daily retraining jobs with new data

6. Model Serving Infrastructure

- **Description**: System for deploying and serving predictions from trained models.
- **Components**:
  - Model registry for versioning
  - Containerized model deployment
  - Load balancing for high availability
  - A/B testing infrastructure for model comparison
- **Performance**: Optimized for low-latency prediction serving

7. Model Monitoring & Alerting

- **Description**: Continuous monitoring system for model performance and data quality.
- **Functionality**:

  - Tracks prediction accuracy against ground truth
  - Monitors data drift and concept drift
  - Detects anomalies in model inputs and outputs
  - Triggers alerts when metrics exceed thresholds
- **Dashboards**: Visual monitoring of system health and model metrics

8. Feedback Loop & Retraining

- **Description**: Mechanism to incorporate actual outcomes into the model improvement process.
- **Functionality**:
  - Collects actual rain occurrence data
  - Evaluates prediction accuracy and updates training datasets
  - Triggers retraining when performance degrades
  - Implements champion-challenger model evaluation

9. API Gateway

- **Description**: Interface for external systems to access predictions and data.
- **Functionality**:
  - Authentication and authorization
  - Rate limiting and request validation
  - Response caching for improved performance
  - API versioning and documentation
- **Endpoints**: Provides daily and hourly prediction endpoints

10. Client Applications

- **Description**: User-facing applications that consume the predictions.
- **Types**:
  - Farmer dashboards showing 21-day forecasts
  - Mobile apps with alerts for upcoming rain
  - Integration with farm management systems
  - Weather notification services

Handling Malfunctioning Sensors

1. **Data Validation Checks**:
   - Range validation (values within expected physical limits)
   - Consistency checks (rapid changes detection)
   - Comparative validation against nearby sensors
2. **Missing Data Handling**:
   - Interpolation for short-term gaps
   - Historical averaging for longer gaps
   - Weighted averaging from nearby sensors
   - Clearly marking imputed values in reports
3. **Anomaly Detection**:

- Statistical methods (Z-score, IQR)
- Machine learning based outlier detection
- Seasonal decomposition for trend analysis

4. **Automated Recovery**:
   - Self-healing mechanisms for sensors
   - Automated recalibration routines
   - Remote sensor diagnostics
5. **Redundancy**:
   - Deployment of redundant sensors in critical areas
   - Backup data sources (public weather APIs)
   - Fallback prediction models for degraded data scenarios

This MLOps system ensures reliable, accurate, and timely weather predictions even with occasional sensor malfunctions, providing farmers with dependable information for their agricultural planning needs. """