**Step 1: Set Up EC2 Instances**

1. **Launch EC2 Instances:**
   * Launch 4 EC2 instances for training.
   * Launch 1 EC2 instance for predictions.
2. **Install Java on EC2 Instances:**
   * SSH into each EC2 instance.
   * Install Java on each instance:

sudo apt update sudo apt install default-jre

**Step 2: Install Apache Spark**

1. **Download and Install Spark on Training Instances:**

wget https://downloads.apache.org/spark/spark-3.2.0/spark-3.2.0-bin-hadoop3.2.tgz tar -xvf spark-3.2.0-bin-hadoop3.2.tgz

* + Configuring the Spark for standalone mode.

**Step 3: Develop and Train the Model**

1. **Write Spark Application in Java:**
   * Developing the Spark application in Java using Spark MLlib or Spark ML. Your application should read the dataset, preprocess the data, and train the ML model.
2. **Distribute Application:**
   * Distributing the Java application to the 4 training instances.
3. **Run Application in Parallel:**
   * Run the Spark application on each training instance in parallel. Spark will distribute the workload.

**Step 4: Save and Load the Model**

1. **Save Trained Model:**
   * Modifying Spark application to save the trained model. The process uses Spark ML's model-saving functionalities.

javaCopy code

model.save("s3://your-bucket/model");

1. **Load Model in Prediction Application:**
   * Here I wrote another Spark application for predictions on the prediction EC2 instance. Loaded the previously saved model.

javaCopy code

SparkSession spark = SparkSession.builder().appName("PredictionApp").getOrCreate(); PipelineModel model = PipelineModel.load("s3://your-bucket/model");

**Step 5: Implement Wine Quality Prediction**

1. **Implement Wine Quality Prediction:**
   * Modifying the Spark application and predicted new data or user input.
2. **Run Prediction Application:**
   * Starting the prediction Spark application on the prediction EC2 instance.