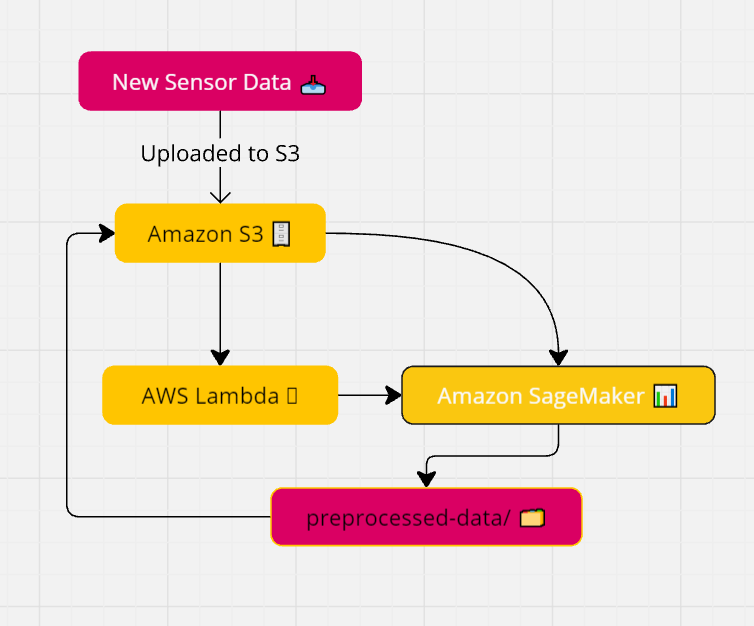
**AWS-Based Predictive Maintenance System - Week 1 Report**

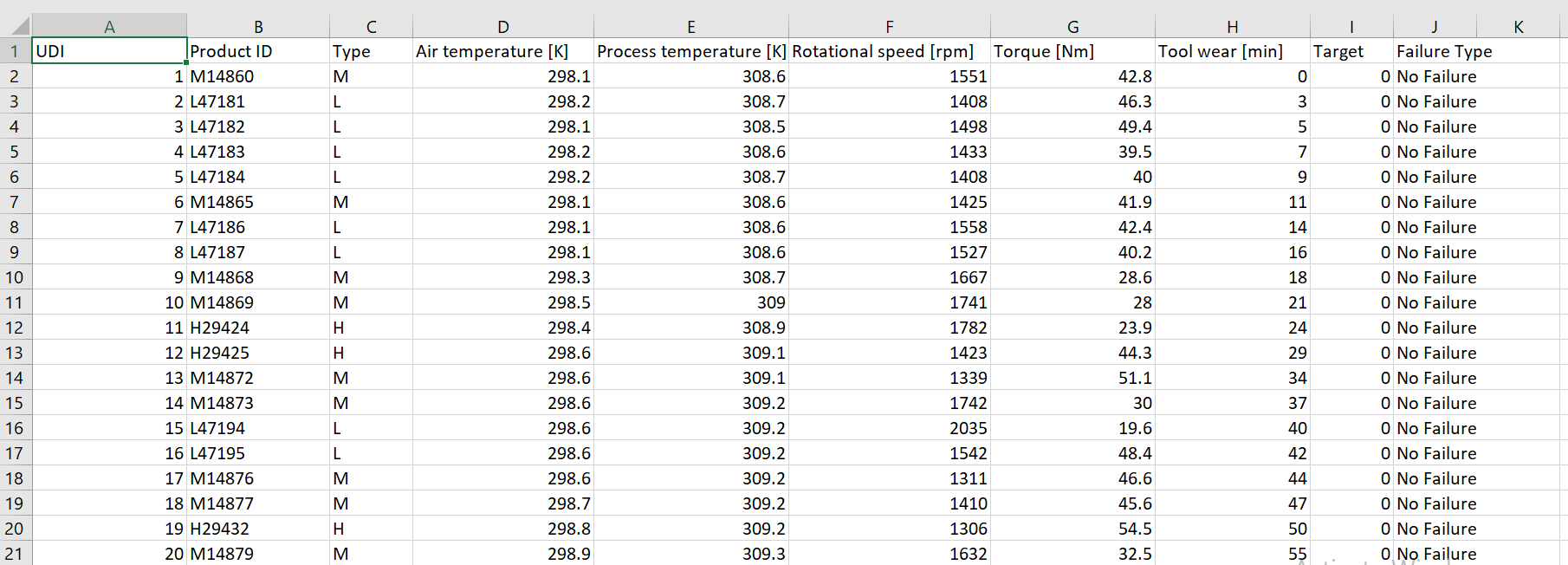
**Introduction**

In Week 1 of the AWS-Based Predictive Maintenance System project, the primary focus was setting up the infrastructure and preparing the sensor data for analysis. We are going to work on a predictive maintenance data file that we get from kaggle. We will be doing this projet using AWS free tier account . The following AWS services were used to support the system architecture and ensure data collection and storage for further model development and deployment. This report outlines the services used, why they were chosen, and the configurations implemented for each.

**Architecture diagram**

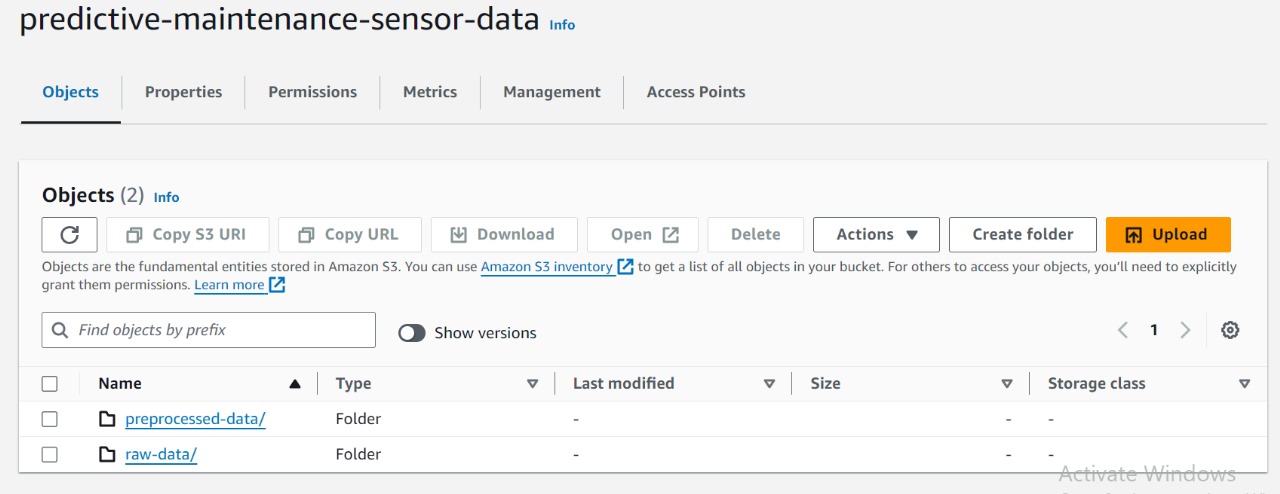


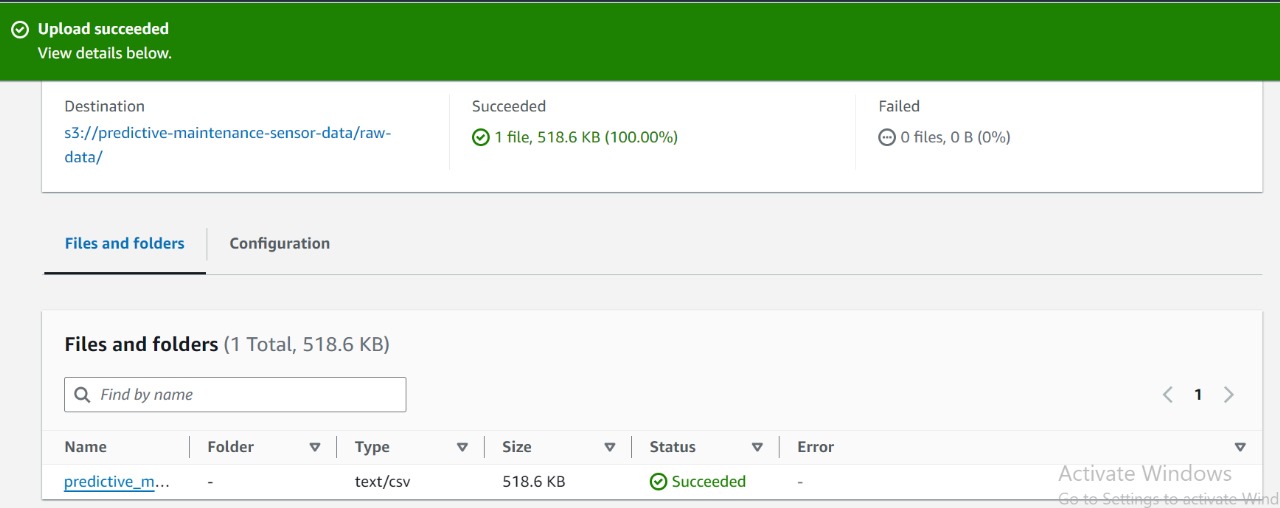
**Data set sample :**

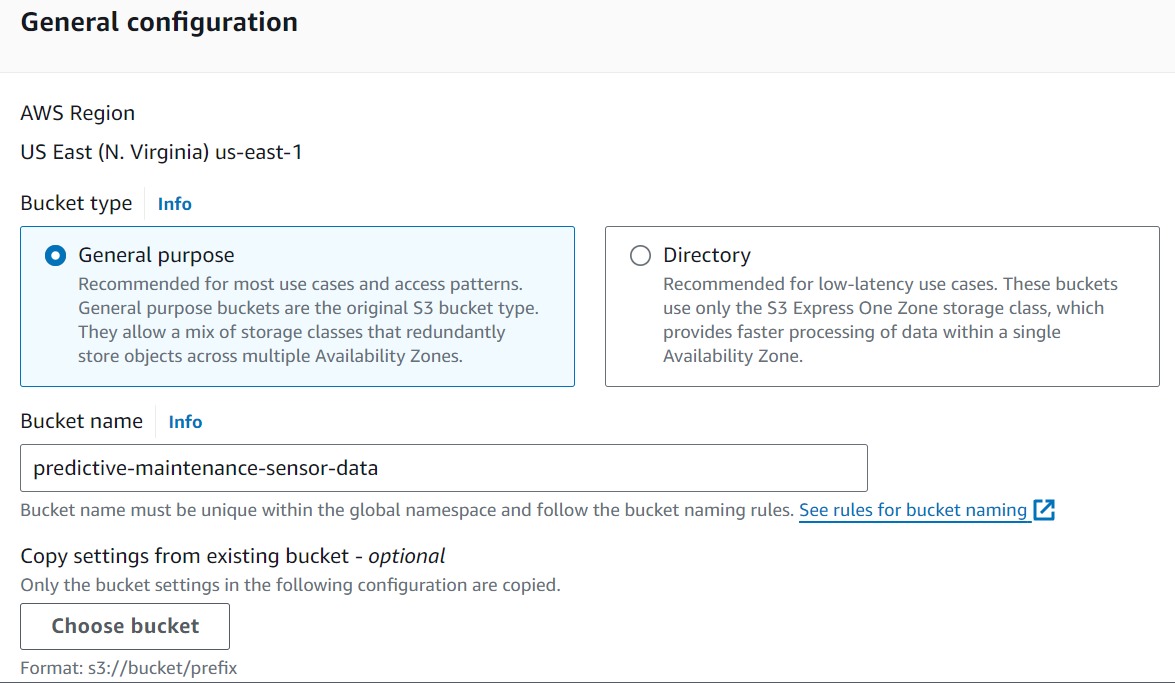
****

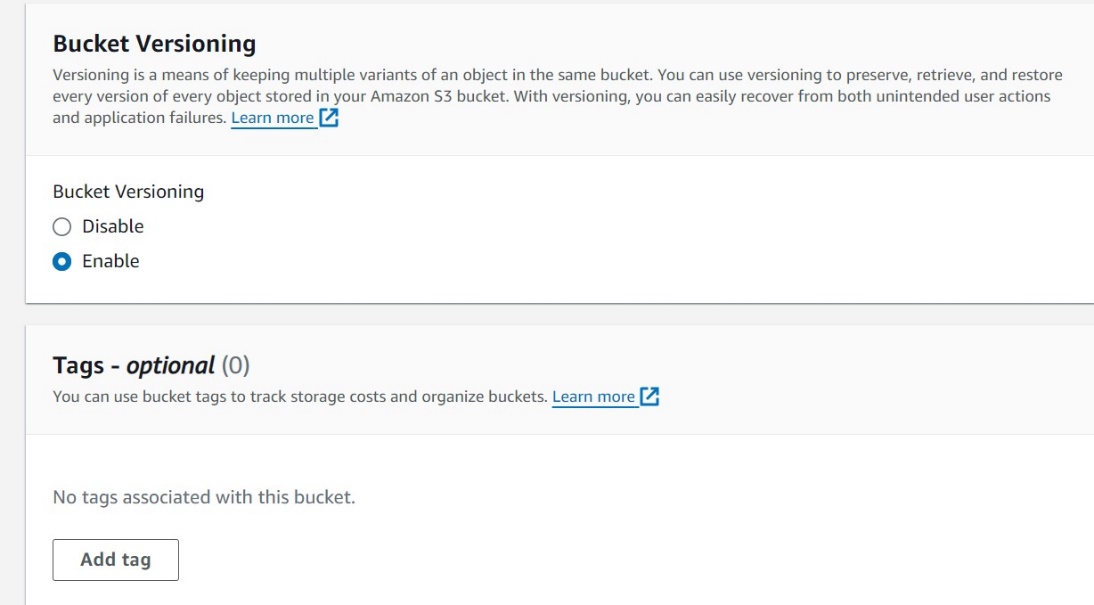
**Services Used**

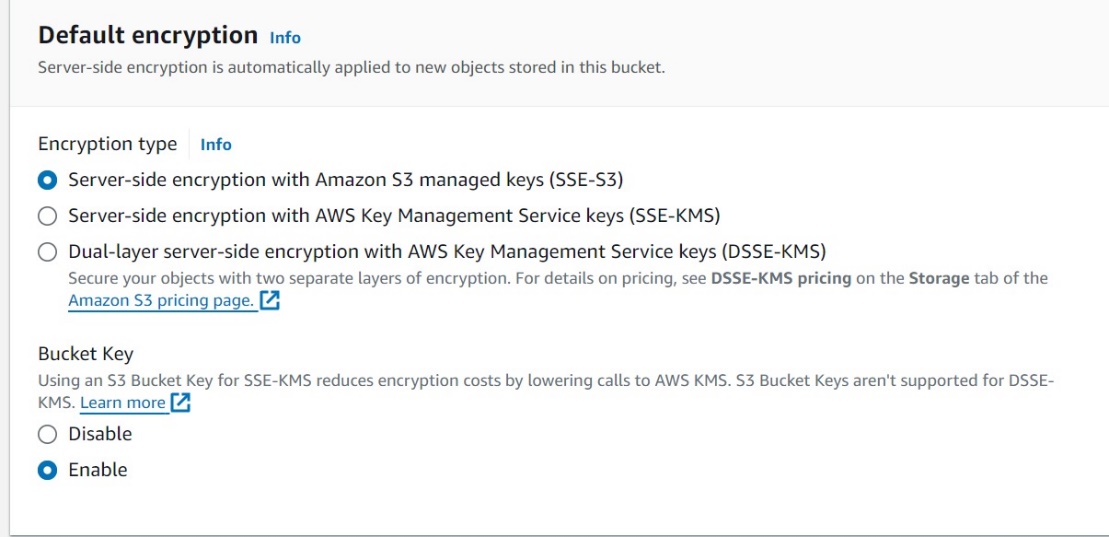
1. **Amazon S3 (Simple Storage Service)**
   * **Purpose**: S3 is used to store the collected sensor data, both raw and preprocessed. It provides scalable and secure storage that allows easy access for further processing by other AWS services such as SageMaker and EC2.
   * **Why Used**:
     + S3 is cost-effective, durable, and provides high availability for data storage.
     + It easily integrates with other AWS services for data processing and retrieval.
     + It allows versioning, ensuring that any changes to datasets can be tracked.
   * **Configuration**:
     + **Bucket Configuration**:
       - Created an S3 bucket called predictive-maintenance-sensor-data.
       - Enabled **versioning** to keep track of changes made to the dataset.
       - Configured the bucket with **server-side encryption** using AWS managed keys to secure the data.
     + **Folder Structure**:
       - Created two folders: raw-data/ for storing the raw sensor data, and preprocessed-data/ for storing the cleaned data.
     + **Permissions**:
       - Access to the bucket was restricted using IAM roles, allowing only the EC2 instance and SageMaker to access the data.



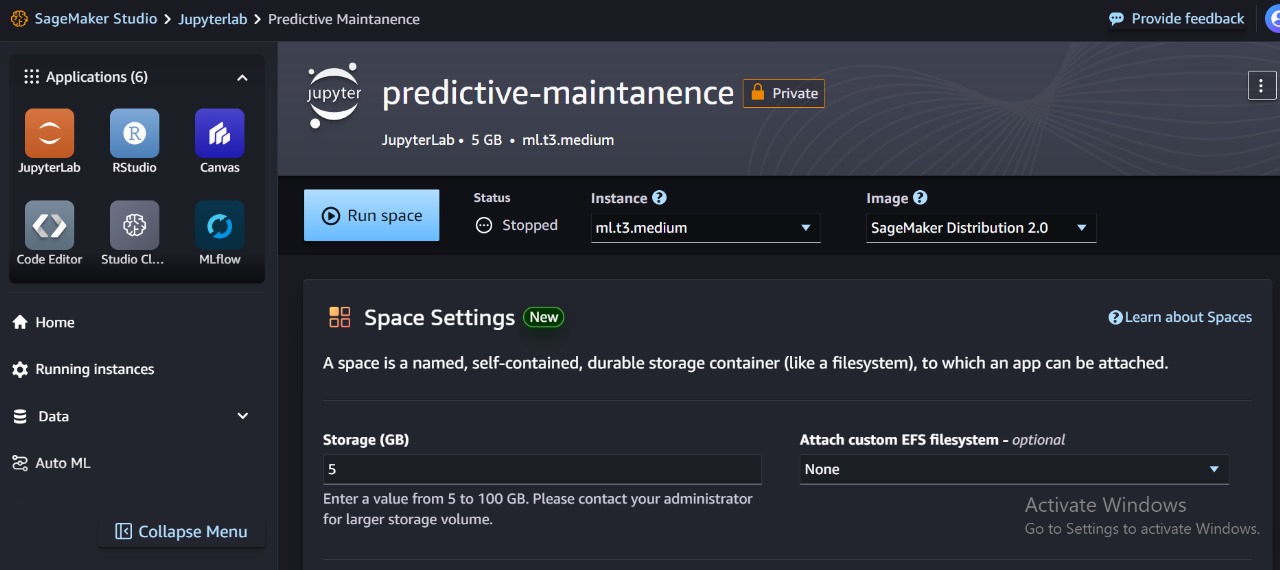


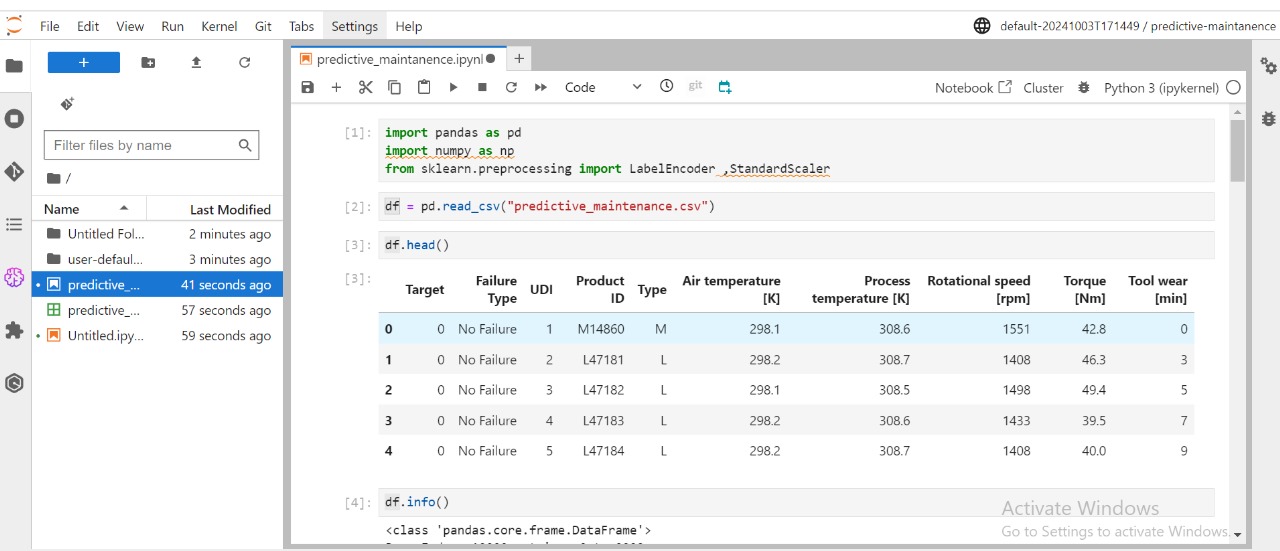






 **Amazon SageMaker**

* **Purpose**: SageMaker is used to preprocess the raw sensor data before it is used for training machine learning models. It provides a fully managed environment where you can write Python scripts for data cleaning, normalization, and feature extraction.
* **Why Used**:
  + SageMaker provides scalable computing resources, ideal for both small-scale and large-scale preprocessing tasks.
  + The Jupyter notebook instance in SageMaker allows you to interactively explore the data, clean it, and prepare it for model training.
  + It integrates directly with S3, allowing seamless retrieval and storage of datasets.
* **Configuration**:
  + **Notebook Instance**: Created a **ml.t3.medium** notebook instance, which provides 2 vCPUs and 4GB RAM, sufficient for preprocessing the sensor data.
  + **Data Access**: Configured SageMaker to read the sensor data stored in the **S3 bucket** (predictive-maintenance-sensor-data) and write the preprocessed data back into the same bucket.
  + **Libraries**: Pre-installed necessary libraries such as **Pandas** and **NumPy** for data preprocessing within the notebook.
  + **Security**: Configured SageMaker with an **IAM role** that has permission to access the S3 bucket and handle the sensor data securely. 



[preprocessing & data cleaning code.pdf](https://drive.google.com/file/d/186gsj0OhULKp7V6v8aUsvUv2GDe552VY/view?usp=sharing)[solving data imbalance and using logistec regression model](https://colab.research.google.com/drive/1FddhgsK64JOJDgNOZtiLBjEBCD6ajWOv?usp=drive_link)

**Conclusion**

For Week 1, we successfully set up the necessary AWS infrastructure for storing and processing sensor data. Amazon S3 was configured for scalable storage of both raw and preprocessed data, While SageMaker is used to do data cleaning and preprocessing using SageMaker studio. The infrastructure is ready for the next stage of model development, where AWS SageMaker will play a key role in building the predictive maintenance model.