Test: Python

# Introduction

This is related to IIoT. Industrial processes and associated sensors in a typical manufacturing production line generate huge volumes of data. The sampling rate varies from 5 to 100 milliseconds. Each data point is called a Tag. There are typically 1000 to 50,000 tags collected overall depending on the industry and the complexity of the process.

This task is to transform such data from row format (all tags in a single row) to a column format (one tag per row). In other words, from traditional SQL friendly to NoSQL friendly format.

Note: Here you can assume the tag is a sensor. tag\_name is sensor name and tag\_value is sensor reading.

# Task

Transform the data present in ‘[Sample Input](https://docs.google.com/spreadsheets/d/1MG3x99End5SEZU4p-EqkMbWnPR-v5-kuRwiLTItoLY4/edit?usp=sharing)’ CSV into ‘[Sample Output](https://docs.google.com/spreadsheets/d/1uzC9QtyBxfPfcL4MEaFikM1wBj690o7xCMsk8fxuIao/edit?usp=sharing)’ CSV using Python.

# Details

1. See comments in ‘Sample Input’ for column mapping (e.g., ItemId becomes tag\_\_id)
2. ‘Sample Input’ has all tags (sensor readings) captured in one row. Whereas in ‘Sample Output’ we need one tag per row. That’s why it is called a transformation task.
3. ‘Sample Input’ has data captured every 5 milliseconds. We need to aggregate it to 10 sec using MAX as the aggregation function.

# Submission Format

1. Preferably as a Jupyter Notebook
2. Upload the output CSV (as generated by the code) to the Git repository

# FAQ

*Q1. When converting from input format to output format, the logic of the tag\_values column in the output dataset is not clear.*

Input data is of high frequency (5 millisec sampling). We need to aggregate it to a 10 second interval using maximum as an aggregate function so that we reduce the no. of records to a manageable size.

In 10 seconds we collect a total 2000 samples (each at 5ms interval) for a tag (sensor) in the input data set. This is too much data to handle. We need to shrink it to just one sample per 10 sec by aggregating it.

For example, a tag (sensor) generates 2000 values (readings) in a given 10 sec interval (Sample Input). tag\_value is nothing but the maximum value (sensor reading) of that tag (sensor) in that interval (Sample Output).