# Introduction

This task is about transforming and presenting sensor timeseries data from a cultivation tank in the production of Active Pharmaceutical Ingredient (API). The production of API is run one batch at the time, and each batch is given a unique BatchID.

# Description of data

The data you have received consists of timeseries data from four sensors and including some metadata about the batches. The sensor data is stored in tab separated csv files and all data is collected at a one-minute rate. The four different sensors are: **400E\_Temp1, 400E\_Temp2, 400E\_PH1, 400E\_PH2.**

The batch metadata is stored in two separate files:

**batch\_info.csv** contains the StartDate and EndDate for each BatchID. The BatchID column contains two types of BatchID:

* Production batches that follow the naming convention XP400EYYYY where **X** is a letter from A-Z and **YYYY** is the trailing batch-number that resets when **X** changes (AP400E0101, AP400E0102, BP400E0101, etc.).
* Test-batches that follow the naming convention TEST\_YYYY

The time between batches is denoted as a NULL/NaN value in the BatchID column.

**batch\_phase.csv** contains StartDate, EndDate and the BatchPhase for each batch (however the BatchID is not included in this file). There’s two different BatchPhases:

* Preparation
  + Small period reserved for preparing the next batch
* Cultivation
  + The phase when the cell cultivation begins
  + This phase represents the period of sensor data that we’re interested in

# Task

Please create a small coding project where you:

* Create a table with aggregated information about all batches
  + Each row should contain the BatchID and BatchDuration (duration of the cultivation phase)
  + Include at least 4 additional rows with other aggregated values from the sensor data
  + The format of the resulting table is up to you, but you should be able to present it at the interview
* Compute and visualize the difference between the temperature (400E\_Temp2 – 400E\_temp1) and PH (400E\_PH2 – 400E\_PH2) sensors for the cultivation phase of one or more batches

You **can** store the data in a database, in a file or in memory, and manipulate it however you like before your calculations.

You **must** store your code in a publicly accessible git repository.