01_Lecture_Exercise

May 20, 2025

1 Lecture Exercise 01 - Chem 273

1.1 Reading Files

1) Motivation

The goal of this exercise is to benchmark different tools for reading files of different formats. We also want to repeat some python coding as a warm-up for the course.

2) Preparation

In order to be able to measure runtime accurately, we will use a *decorator*:

[34]:
```
from my_timer import my_timer
```

In the next step we want to read the following data files: *Data_set_0.xlsx Data_set_0.csv Data_set_0.txt* All three files have the exact same content, but are of different formats. We now load the required libraries *pandas*, *dask* and *polars*:

[21]:
```
import pandas as pd
import dask.dataframe as dd
import polars as pl
#run pip install dask and/or pip install polars if needed!
```

3) Exercise

Write a short function using *def* that reads the data file of a given format and using a specific library. Apply the decorator via

[ ]:
```
#@my_timer
#def My_Function(input1, input2, ...)
```

What is the difference in time you measure? In order to obtain the same functionality of data frames you are used to by *pandas*, sometimes the data frame which has been generated using another library, such as *polars* or *dask*, has to be converted into a *pandas* data frame via:

[31]:
```
#df = pd.DataFrame(df)
```

How much time does the conversion require? Do we still gain time?