





## Run Time Models Experiment

### 1. Dataset:

Menu.csv from New York Public Library. [\[Menu\]](#)

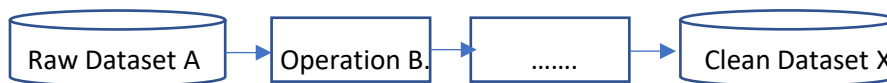
Data size: 3.2 MB

Data rows: 17545

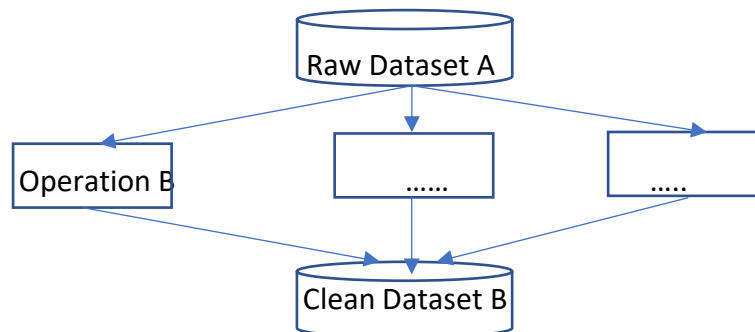
 Dish.csv	Sep 16, 2018 at 2:16 AM	26.6 MB
 Menu.csv	Sep 16, 2018 at 2:05 AM	3.2 MB
 Menuitem.csv	Sep 16, 2018 at 2:14 AM	118.5 MB
 MenuPage.csv	Sep 16, 2018 at 2:05 AM	4.7 MB

### 2. Run Time Model Methodology

#### 1. Sequential/Linear Run-time Model:



#### 2. Parallel Run-time Model:



### 3. Implementation [\[Github\]](#)

*Linear Run-time Model:*

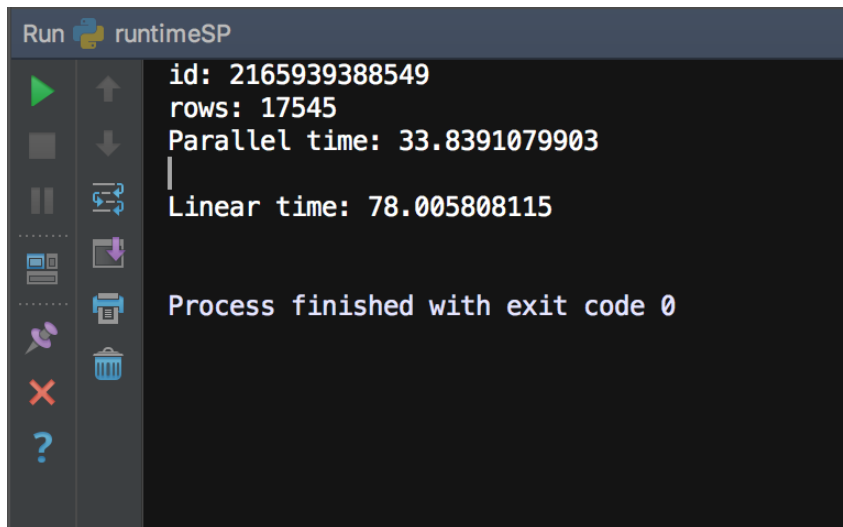
1. Read in csv file and set the project name
2. Create project with [OpenRefine-Client Library](#).
3. **Set start time timer here**

4. Traverse the dictionaries in the JSON format csv file. And pass parameters to operations in OpenRefine-Client Library.
5. **Set End time timer here**
6. Repeat step 4 for 1000 times. And get the difference of the End time and Start time.

*Parallel Run-time Model:*

1. Read in csv file and set the project name
2. Create project with [OpenRefine-Client Library](#).
3. Use [pool](#) to distribute processes (data parallelism)
4. **Set start time timer here**
5. Do the operations (B,C,...) with the processes.
6. **Set End time timer here**
7. Repeat step 5 for 1000 times. And get the difference of the End time and Start time.

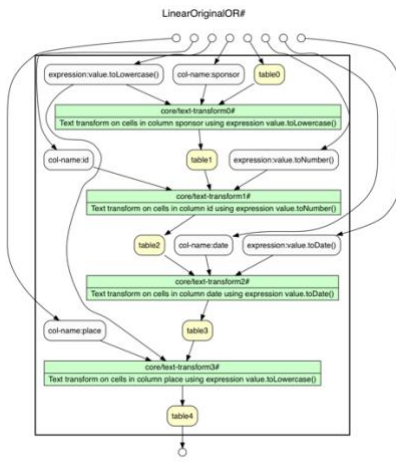
4. Results.



```
Run runtimeSP
id: 2165939388549
rows: 17545
Parallel time: 33.8391079903
|
Linear time: 78.005808115

Process finished with exit code 0
```

5. Conceptual Model with Yesworkflow



Parallel model:

