

Readme File

Database Project Fall 2019
Sajan Bang(sb6632), Hitesh Patel(hlp276)

The project folder consists of the following files:

- 1) main.py
- 2) inputFile.txt
- 3) Sales1.txt
- 4) Sales2.txt
- 5) Reprozip
- 6) Readme file

- For taking the input command and manipulating those command to call respective functions we used regex, our regex removed all the special characters except the ones that are needed for general arithmetic operators(+,-,*,/) and comparison operators(>,<,==,<=,>=,!=).
- For executing the functions of the project, we used following libraries and packages from python:
 - Import numpy as np
 - Import re
 - Import csv
 - Import operators
 - Import itertools
 - From Btrees.ODBTree import Btree
- For comparison and other operations we have used operators which can perform operations like >, <, ==, <=, >=, !=.
- For performing the sort function we have used np.lexsort function.
- We also used np.cumsum for performing moving average and moving sum functionality.
- For implementing B-Trees we have used the following library:
 - Btrees - <https://pypi.org/project/BTrees/>
 - Installation : pip install Btrees
- Hash structure is implemented using the python dictionary.
- We have used python numpy library to generate ndarray's for storing the data from text files in a structured-format.
- Name of the functions in the project:
 - Main()
 - Inputfromfile()
 - Commandfile()
 - Select()
 - Project()
 - Avg()

- [Sum\(\)](#)
- [Count\(\)](#)
- [Sumgroup\(\)](#)
- [Avggroup\(\)](#)
- [Countgroup\(\)](#)
- [Join\(\)](#)
- [Operator\(\)](#)
- [Condition\(\)](#)
- [Sort\(\)](#)
- [Movavg\(\)](#)
- [Movsum\(\)](#)
- [Btree\(\)](#)
- [Hash\(\)](#)
- [Concat\(\)](#)
- [Outputtofile\(\)](#)

- **Output format:** We have printed the output to text files. The row elements in the text files are tab delimited and every row is separated by a new line. Headers have been printed as well wherever necessary.
- **Time** taken in executing the query is display on the terminal, also for reference find below attached screenshot of timings.

```
[SAJANS-MacBook-Pro:Desktop sajan$ cd sb6632_hlp276_project
[SAJANS-MacBook-Pro:sb6632_hlp276_project sajan$ python main.py
Final Compute time for inputfromfile in Seconds= 0.014625072479248047
Final Compute time for select in Seconds= 0.021689891815185547
Final Compute time for project in Seconds= 0.012629032135009766
Final Compute time for avg in Seconds= 0.013347148895263672
Final Compute time for sumgroup in Seconds= 0.042428016662597656
Final Compute time for sumgroup in Seconds= 0.018548965454101562
Final Compute time for avggroup in Seconds= 0.017786026000976562
Final Compute time for inputfromfile in Seconds= 1.5500729084014893
Final Compute time for join in Seconds= 102.17553782463074
Final Compute time for sort in Seconds= 0.09949207305908203
Final Compute time for sort in Seconds= 0.09667611122131348
Final Compute time for movavg in Seconds= 0.10829281806945801
Final Compute time for movsum in Seconds= 0.11632084846496582
Final Compute time for select in Seconds= 0.014198780059814453
Final Compute time for Btree in Seconds= 0.02774500846862793
Final Compute time for select in Seconds= 0.014068841934204102
Final Compute time for Hash in Seconds= 0.055886030197143555
SAJANS-MacBook-Pro:sb6632_hlp276_project sajan$ ]
```

Note:

While implementing “`T1 := join(R1, S, (R1.qty > S.Q) and (R1.saleid = S.saleid))`”, our system was taking a lot of time so we didn’t got the end result as we slightly changed the query from “`>`” to “`=`” operation but sometimes it was giving end result and sometime system was crashing due to high computation. But

we need file that will be formed after executing this query for the other queries after this. So please use your T1.txt file so that other queries perform perfectly, which are working correctly for us. We are reading the file as tab separated and with header so please give that T1.txt file in the desired way (which I believe are there).

T1.txt is required for the following queries:

- $T2 := \text{sort}(T1, S_C)$
- $T2\text{prime} := \text{sort}(T1, R1_time, S_C)$

These are the directly related queries whereas text files from these above-mentioned queries are also required for some other queries given for the project.