Multiway Join - Usage Guide

- Information
- System Requirements
- Software Requirements
- Algorithm
 - Relation Overview
 - Aligned Tables
 - Implementation
- Project Setup
 - MySQL
 - Python
 - pip
 - flask
- Project Execution
- Results
 - Tabular Result

Information

Student Name: Avinash Sorab | DAWG Tag: siu854709544 | Professor: Dr. Wen-Chi Hou

System Requirements

The project was developed and tested on a laptop that had

RAM: 4GBHDD: 320GB

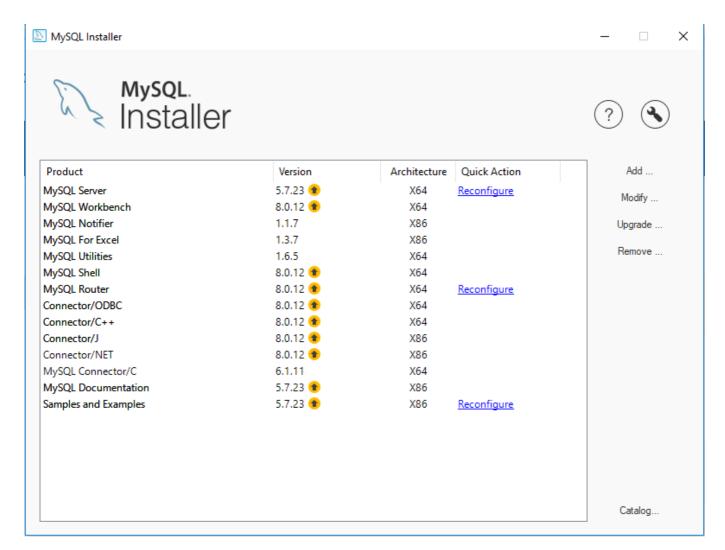
Processor: Core i5OS: Windows 10 Pro

Software Requirements

- Visual Studio Code A light weight code editor that has command line integrated to it. Easy to write
 code, debugging as it enables us to install intellisense.
 https://code.visualstudio.com/
- MySQL Installer 5.7.23 (Community Version) https://downloads.mysql.com/archives/installer/

Choose the product version from latest to 5.7.23

It'll let you install all the below softwares



Make sure you've all these softwares installed below

Microsoft Visual C++ 2010 Redistributable (x86)

Microsoft Visual C++ 2010 Redistributable (x64)

Microsoft Visual C++ 2013 Redistributable (x86)

Microsoft Visual C++ 2013 Redistributable (x64)

Microsoft Visual C++ 2015 Redistributable (x86)

Microsoft Visual C++ 2015 Redistributable (x64)

to support all the softwares present in the installer.

Python 3.7.0

https://www.python.org/

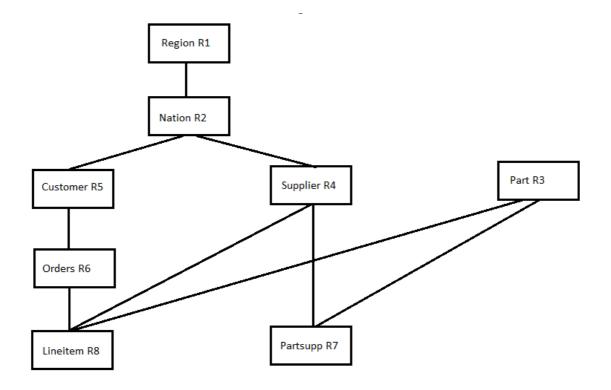
Git 2.19.2.windows.1

https://git-scm.com/download/win

Algorithm

Relation Overview

The TPCH benchmark relation used for this is as below



Aligned Tables

There are 11 aligned tables possible in this

- ALIGNED_REGION (R1)
- ALIGNED_NATION (R1 -> R2)
- ALIGNED_PART (R3)
- ALIGNED SUPPLIER (R2 -> R4)
- ALIGNED_CUSTOMER (R2 -> R5)
- ALIGNED_ORDERS (R5 -> R6)
- ALIGNED_PARTSUPP (R4 -> R7, R3 -> R7)
- ALIGNED_LINEITEM (R4 -> R8, R3 -> R8)

Implementation

- Create aligned tables
- Allow the user to select the relation he wants to create join on
- Ex: if the user selects REGION -> NATION -> SUPPLIER, for every single tuple of region, select all the nations and following suppliers which belong to that region (DFS)
- Send it through the log_linear_result function that compares the self_sid of one table with the parent_sid of another table,
- In the below example it compares r1_sid from aligned_region table with r1_sid from aligned_nation table and r2_sid from aligned_nation is compared with r2_sid in aligned_supplier, if they are equal, they'll be logged into a file

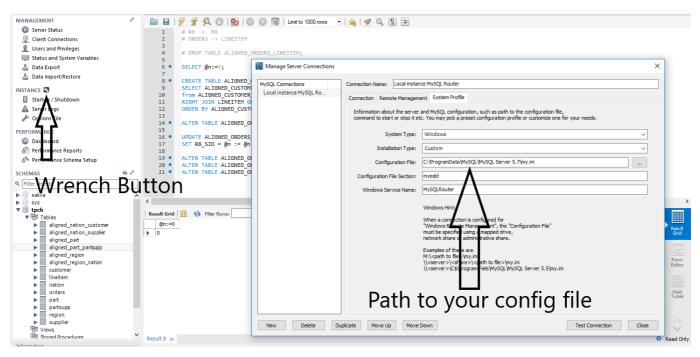
```
for tuple1 in aligned_region:
   for tuple2 in aligned_nation:
     for tuple3 in aligned_supplier:
```

- For Divergent Graphs, such as REGION -> NATION -> (SUPPLIER, CUSTOMER), Identify the shortest branch,
- Implement linear join for the shortest branch, then implement linear join for another branch and store the result
- Output the crossproduct of both the result.

Project Setup

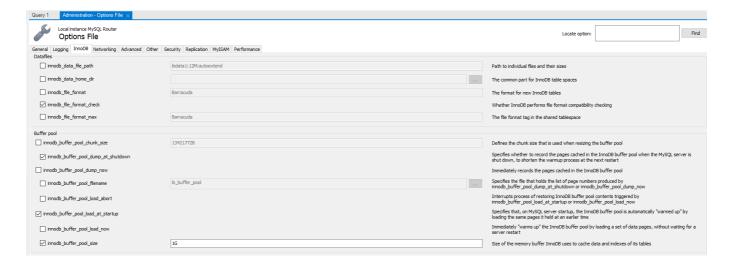
MySQL

- Open MySQL WorkBench
- From the Main Menu -> Server -> Options File, if the file doesn't load then click on the Wrench button beside Instance in the left pane
- A window pops up

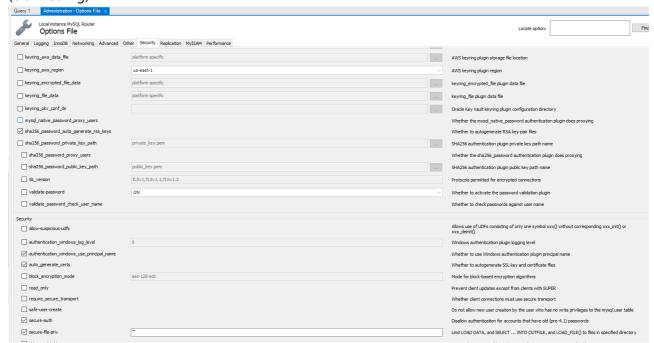


After this you should be able to open the options file

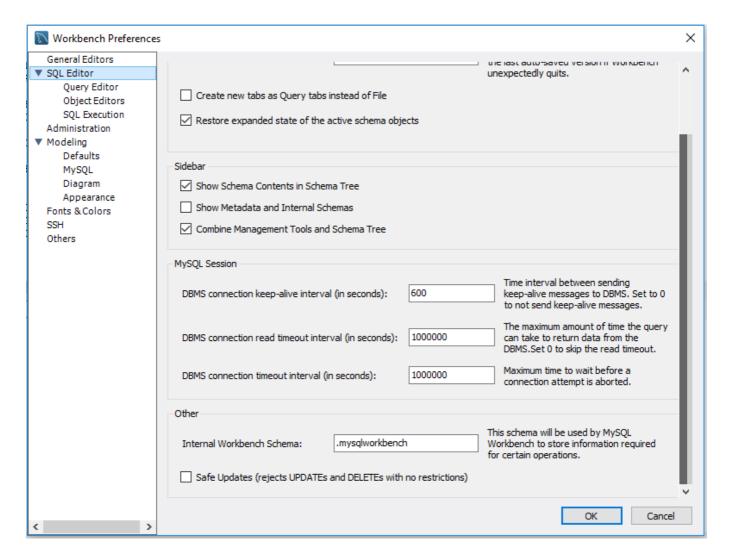
Increase the InnoDB Buffer Pool Size from 8M to 1G by going to Main Menu -> Server -> Options File InnoDB tab -> Check innodb_buffer_pool_size and change it to 1G



Go to Security Tab, scroll down till you find secure-file-priv, check it if it's unchecked and set it to ""
(blank string)



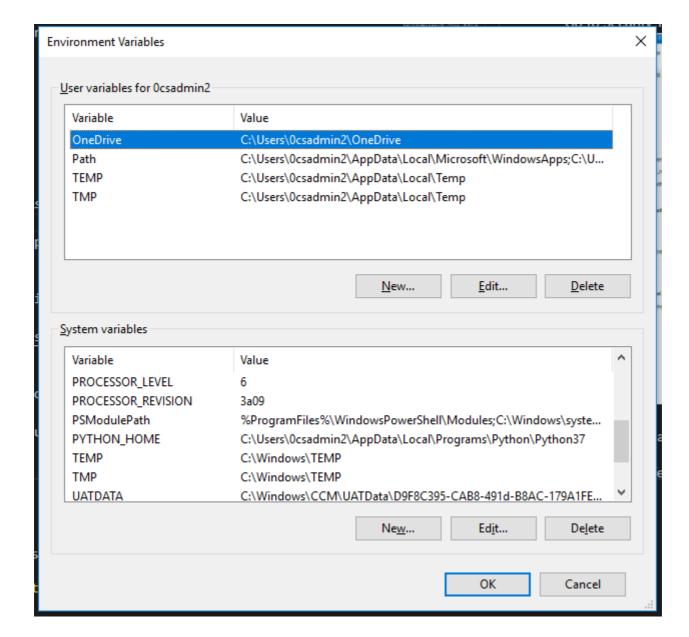
- Click on Apply and then close the window
- Disable the Safe Updates by going to Edit -> Preferences -> SQL Editor and scroll down and you'll find an option, uncheck it.



Python

After you install python, add python to the ENVIROMENT_VARIABLES in your windows system.

• On your Windows machine, go to System Properties -> Advanced Tab -> Environment Variables



You can verify this by opening cmd from anywhere and type python to see if python command line opens up.

You need pip to install most of the python tools

pip

- You can download the code get-pip.py from the below link. https://bootstrap.pypa.io/get-pip.py and run it on python command line
- Navigate to the location where you've downloaded get-pip.py file.
- > python get-pip.py
 - It will download and install pip
 - Put python and pip executables into a common location

C:\Python\Scripts

and add this to the environment variables list so that pip can be run from anywhere

Verify by typing

```
> pip freeze
```

from anywhere in the window.

flask

• Install flask, the GUI tool that's required to get inputs from the User and show aligned relations

```
> pip install flask
```

• Clone the project from the below link into your system.

```
> git clone https://github.com/avinashsp93/multiwayjoin.git
```

- Open the folder in Visual Studio Code, Press Ctrl + `
- Install some important plugins required

```
> pip3 install pandas
> pip3 install mysql-connector-python-rf
```

• Navigate to the .\scripts folder

```
> cd .\scripts
```

• Set the FLASK_APP environment variable

```
> $env:FLASK_APP="main"
```

Project Execution

- Navigate to your project where you've cloned, using commandline
- Run the below command from the commandline

```
> flask run
```

You should be getting below output

```
PS C:\Users\0csadmin2\Desktop\multiwayjoin\scripts> flask run

* Serving Flask app "main"

* Environment: production
WARNING: Do not use the development server in a production environment.
Use a production WSGI server instead.

* Debug mode: off

* Running on http://127.0.0.1:5000/ (Press CTRL+C to quit)
```

Click on the link http://127.0.0.1:5000/ and open it in chrome browser

• At this state if you refresh the server, you'll be getting a new tkinter window where the user can specify inputs

Input



- The highlighted cyan color indicates the aligned relation that's going to be created
- Click on Generate Result button and head back to command line

Output

```
Process time: 16623.3985

Number of rows 1197064

SELECT * FROM ALIGNED_REGION_NATION WHERE R1_SID >= 2 AND R1_SID <= 2 ORDER BY R2_SID

SELECT * FROM ALIGNED_NATION_CUSTOMER WHERE R2_SID >= 6 AND R2_SID <= 10 ORDER BY R5_SID

SELECT * FROM ALIGNED_CUSTOMER_ORDERS WHERE R5_SID >= 29765 AND R5_SID <= 59716 ORDER BY R6_SID

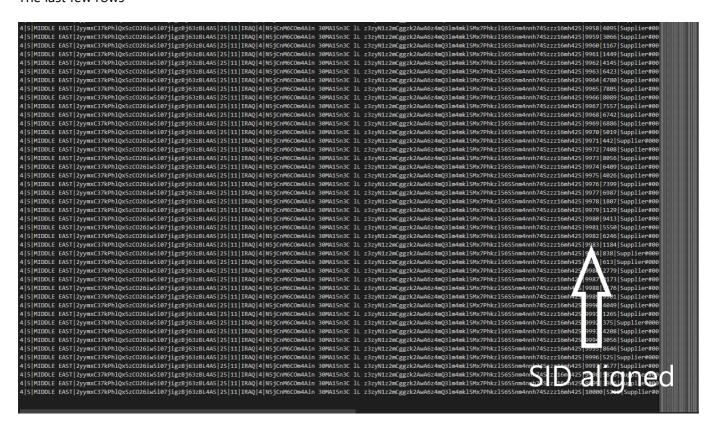
SELECT * FROM ALIGNED_ORDERS_LINEITEM WHERE R6_SID >= 299121 AND R6_SID <= 597076 ORDER BY R8_SID
```

Print File

The first few rows



The last few rows



Results

Tabular Result

Relation	Time	File Size	Number of Rows
R -> N	0.02sec	5kB	25

Relation	Time	File Size	Number of Rows
R -> N -> S	0.5299sec	3235kB	10000
R -> N -> C	8.1313sec	51846kB	150000
R -> N -> C -> O	1991.81sec	698625kB	1500000
R -> N -> S -> PS	113.0356sec	380113kB	800000
R -> N -> C -> O -> L	83487.59sec	2543871kB	6000000
N -> S	0.5097sec	2302kB	10000
N -> C	7.1866sec	37855kB	150000
N -> S -> PS	49.8817sec	305503kB	800000
N -> C -> O	492.17sec	558689kB	1500000
S -> PS	433.3983sec	2350kB	800000