CS 487/587 Database Implementation
Winter 2021
Database Benchmarking Project
Part I - Data Generation & System Selection

Due: Tues February 2, midnight (end of the day on the 2nd), D2L

Team: Swetha Venkatesan, Sai Deepika Gooty

Project Options:

Option 2: Single System: Evaluate a single 'relational' system by changing system parameters and varying relation (data) size.

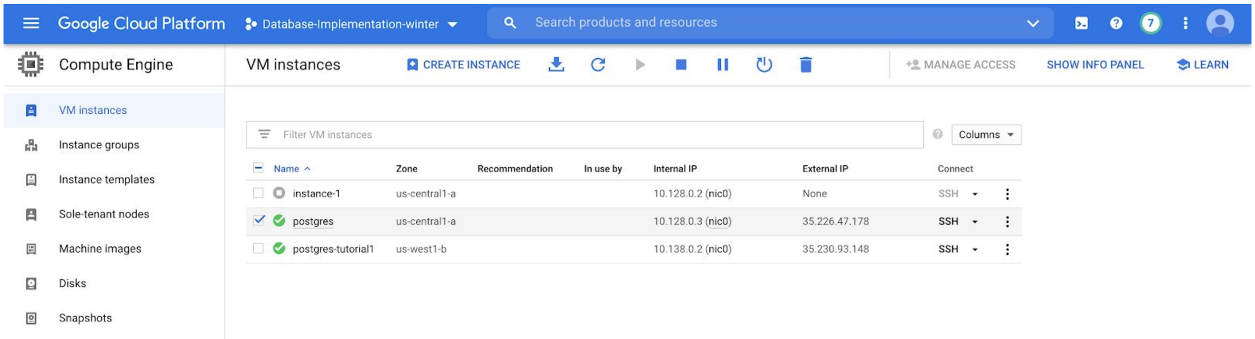
System Selection: PostgreSQL

Project Part I: Data Generation & System Selection

GitHub Repository - https://github.com/swet09/DB-Implementation-Project

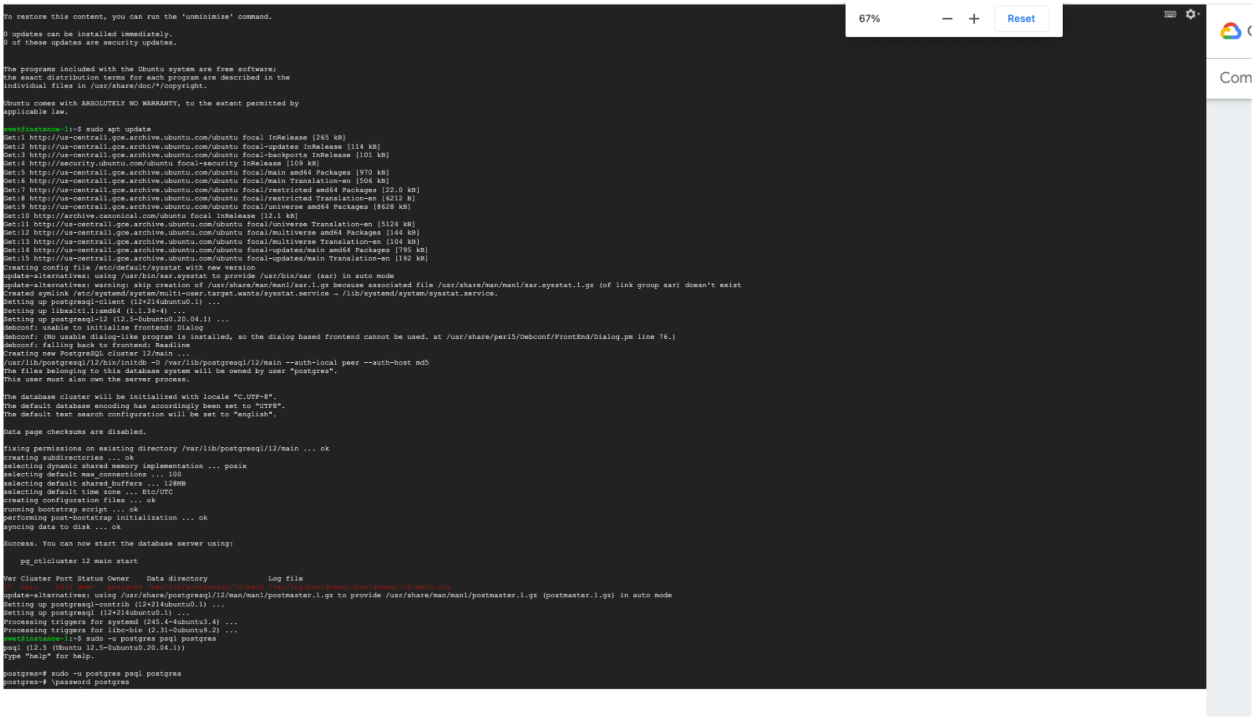
1. Screenshots or other mechanisms that demonstrate that you have successfully generated data and loaded that data into your system.

   Creation of VM



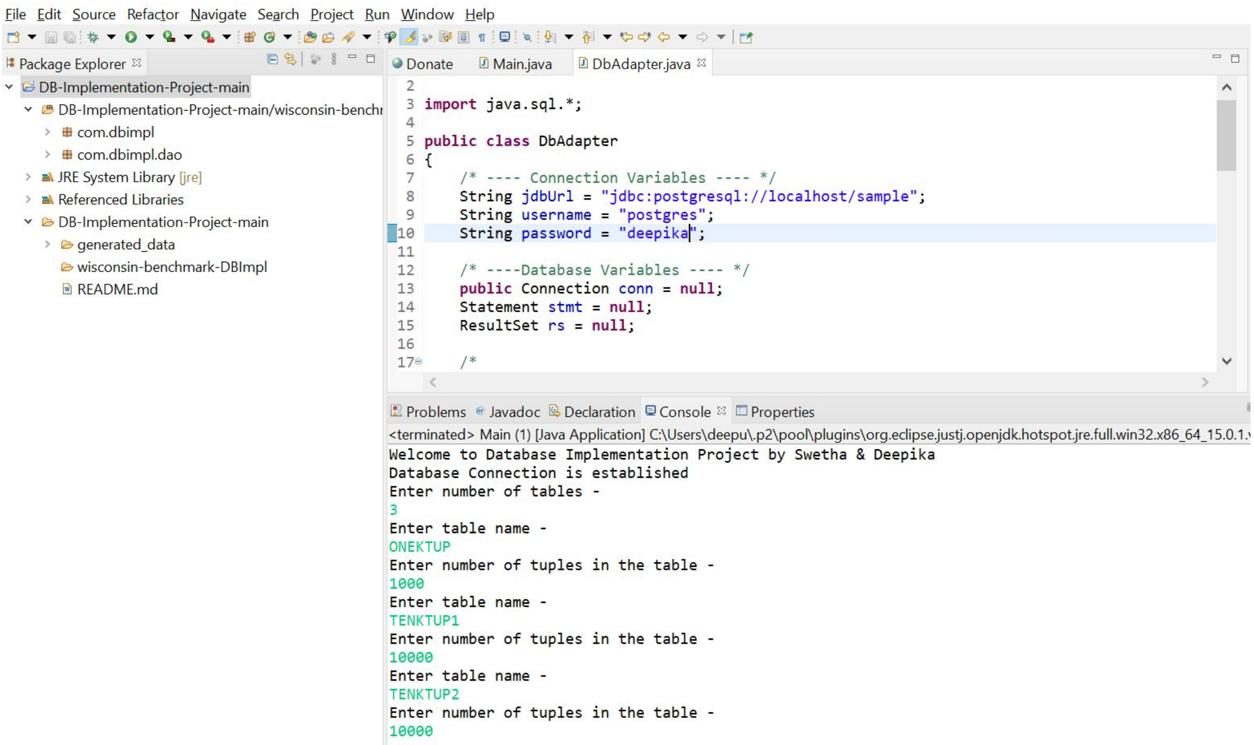
	Name	Zone	Recommendation	In use by	Internal IP	External IP	Connect
□	instance-1	us-central1-a			10.128.0.2 (nic0)	None	SSH
☑	postgres	us-central1-a			10.128.0.3 (nic0)	35.226.47.178	SSH
□	postgres-tutorial1	us-west1-b			10.138.0.2 (nic0)	35.230.93.148	SSH

Installation of Postgres



```
To restore this content, you can run the 'unminimize' command.

0 updates can be installed immediately.
0 of these updates are security updates.


The programs included with the Ubuntu system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/copyright.

Ubuntu comes with ABSOLUTELY NO WARRANTY, to the extent permitted by
applicable law.

swet@instance-1:~$ sudo apt update
Get:1 http://us-central1.gce.archive.ubuntu.com/ubuntu focal InRelease [265 kB]
Get:2 http://us-central1.gce.archive.ubuntu.com/ubuntu focal-updates InRelease [114 kB]
Get:3 http://us-central1.gce.archive.ubuntu.com/ubuntu focal-backports InRelease [101 kB]
Get:4 http://security.ubuntu.com/ubuntu focal-security InRelease [109 kB]
Get:5 http://us-central1.gce.archive.ubuntu.com/ubuntu focal/main amd64 Packages [970 kB]
Get:6 http://us-central1.gce.archive.ubuntu.com/ubuntu focal/main Translation-en [506 kB]
Get:7 http://us-central1.gce.archive.ubuntu.com/ubuntu focal/restricted amd64 Packages [22.0 kB]
Get:8 http://us-central1.gce.archive.ubuntu.com/ubuntu focal/restricted Translation-en [6212 B]
Get:9 http://us-central1.gce.archive.ubuntu.com/ubuntu focal/universe amd64 Packages [8628 kB]
Get:10 http://archive.canonical.com/ubuntu focal InRelease [12.1 kB]
Get:11 http://us-central1.gce.archive.ubuntu.com/ubuntu focal/universe Translation-en [5124 kB]
Get:12 http://us-central1.gce.archive.ubuntu.com/ubuntu focal/multiverse amd64 Packages [144 kB]
Get:13 http://us-central1.gce.archive.ubuntu.com/ubuntu focal/multiverse Translation-en [104 kB]
Get:14 http://us-central1.gce.archive.ubuntu.com/ubuntu focal-updates/main amd64 Packages [795 kB]
Get:15 http://us-central1.gce.archive.ubuntu.com/ubuntu focal-updates/main Translation-en [192 kB]
Creating config file /etc/default/sysstat with new version
update-alternatives: using /usr/bin/sar.sysstat to provide /usr/bin/sar (sar) in auto mode
update-alternatives: warning: skip creation of /usr/share/man/man1/sar.1.gz because associated file /usr/share/man/man1/sar.sysstat.1.gz (of link group sar) doesn't exist
Created symlink /etc/systemd/system/multi-user.target.wants/sysstat.service → /lib/systemd/system/sysstat.service.
Setting up postgresql-client (12+214ubuntu0.1) ...
Setting up libxslt1.1:amd64 (1.1.34-4) ...
Setting up postgresql-12 (12.5-0ubuntu0.20.04.1) ...
debconf: unable to initialize frontend: Dialog
debconf: (No usable dialog-like program is installed, so the dialog based frontend cannot be used. at /usr/share/perl5/Debconf/FrontEnd/Dialog.pm line 76.)
debconf: falling back to frontend: Readline
Creating new PostgreSQL cluster 12/main ...
/usr/lib/postgresql/12/bin/initdb -D /var/lib/postgresql/12/main --auth-local peer --auth-host md5
The files belonging to this database system will be owned by user "postgres".
This user must also own the server process.

The database cluster will be initialized with locale "C.UTF-8".
The default database encoding has accordingly been set to "UTF8".
The default text search configuration will be set to "english".

Data page checksums are disabled.

fixing permissions on existing directory /var/lib/postgresql/12/main ... ok
creating subdirectories ... ok
selecting dynamic shared memory implementation ... posix
selecting default max_connections ... 100
selecting default shared_buffers ... 128MB
selecting default time zone ... Etc/UTC
creating configuration files ... ok
running bootstrap script ... ok
performing post-bootstrap initialization ... ok
syncing data to disk ... ok

Success. You can now start the database server using:

    pg_ctlcluster 12 main start

Ver Cluster Port Status Owner    Data directory              Log file
12  main    5432 down   postgres /var/lib/postgresql/12/main /var/log/postgresql/postgresql-12-main.log
update-alternatives: using /usr/share/postgresql/12/man/man1/postmaster.1.gz to provide /usr/share/man/man1/postmaster.1.gz (postmaster.1.gz) in auto mode
Setting up postgresql-contrib (12+214ubuntu0.1) ...
Setting up postgresql (12+214ubuntu0.1) ...
Processing triggers for systemd (245.4-4ubuntu3.4) ...
Processing triggers for libc-bin (2.31-0ubuntu9.2) ...
swet@instance-1:~$ sudo -u postgres psql postgres
psql (12.5 (Ubuntu 12.5-0ubuntu0.20.04.1))
Type "help" for help.

postgres=# sudo -u postgres psql postgres
postgres-# \password postgres
```

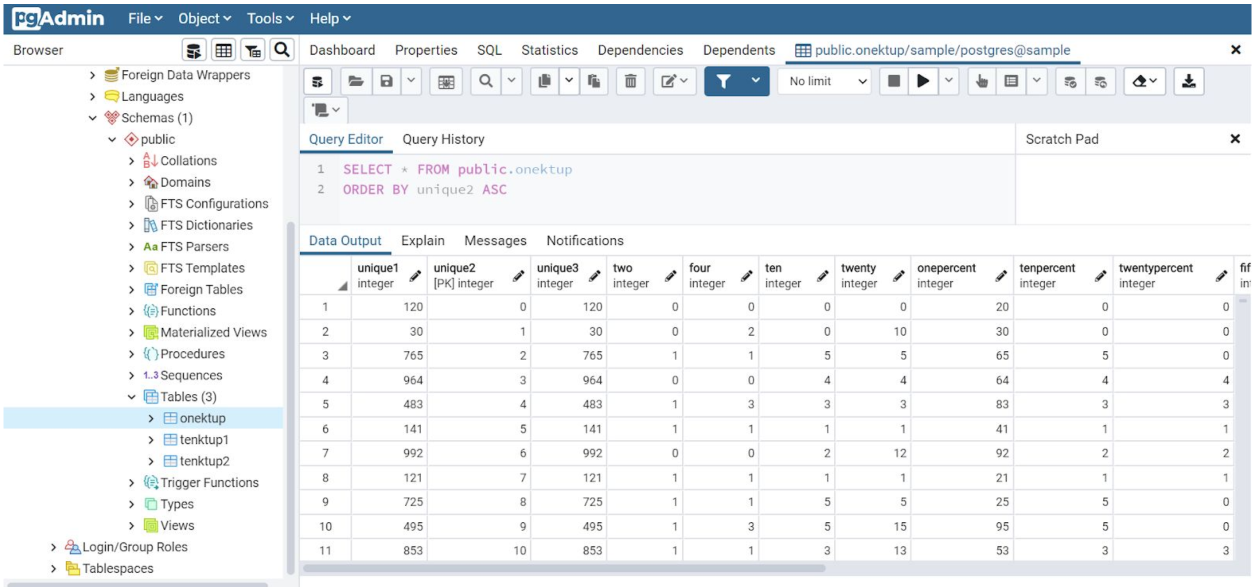
Running the Java Code


File Edit Source Refactor Navigate Search Project Run Window Help
Package Explorer
DB-Implementation-Project-main
DB-Implementation-Project-main/wisconsin-benchr
com.dbimpl
com.dbimpl.dao
JRE System Library [jre]
Referenced Libraries
DB-Implementation-Project-main
generated_data
wisconsin-benchmark-DBImpl
README.md
Donate
Main.java
DbAdapter.java
Problems
Javadoc
Declaration
Console
Properties

```java

import java.sql.*;

public class DbAdapter
{
    /* ---- Connection Variables ---- */
    String jdbcUrl = "jdbc:postgresql://localhost/sample";
    String username = "postgres";
    String password = "deepika";

    /* ----Database Variables ---- */
    public Connection conn = null;
    Statement stmt = null;
    ResultSet rs = null;

    /*
```

```
<terminated> Main (1) [Java Application] C:\Users\deepu\.p2\pool\plugins\org.eclipse.justj.openjdk.hotspot.jre.full.win32.x86_64_15.0.1.
Welcome to Database Implementation Project by Swetha & Deepika
Database Connection is established
Enter number of tables -
3
Enter table name -
ONEKTUP
Enter number of tuples in the table -
1000
Enter table name -
TENKTUP1
Enter number of tuples in the table -
10000
Enter table name -
TENKTUP2
Enter number of tuples in the table -
10000
```

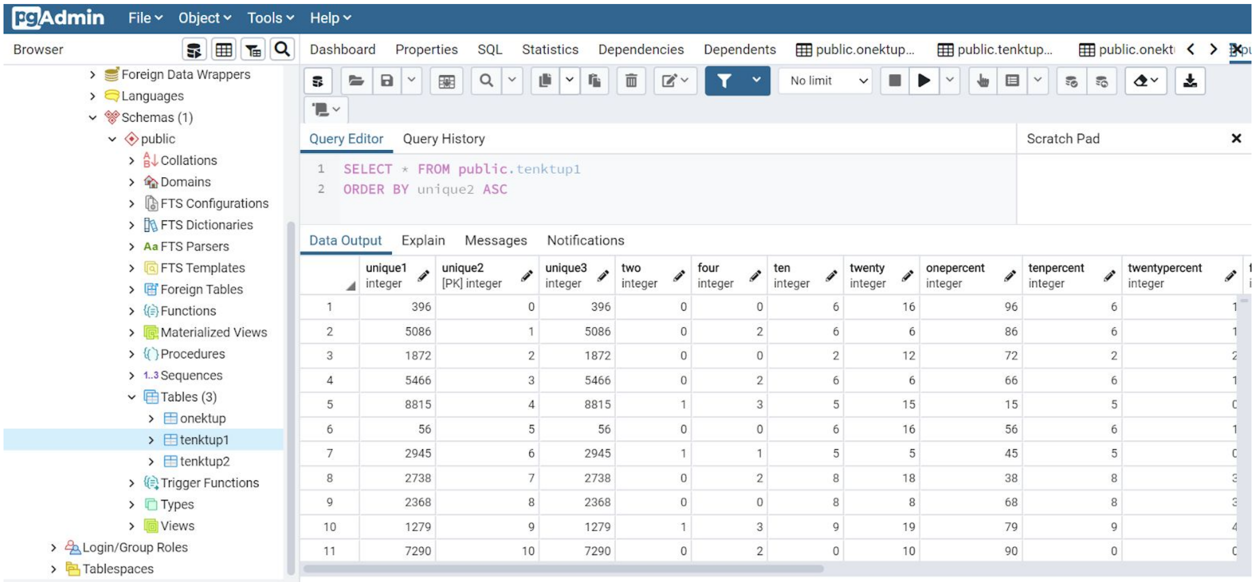
ONEKTUP



```
SELECT * FROM public.onektup
ORDER BY unique2 ASC
```

	unique1 integer	unique2 [PK] integer	unique3 integer	two integer	four integer	ten integer	twenty integer	onepercent integer	tenpercent integer	twentypercent integer
1	120	0	120	0	0	0	0	20	0	0
2	30	1	30	0	2	0	10	30	0	0
3	765	2	765	1	1	5	5	65	5	0
4	964	3	964	0	0	4	4	64	4	4
5	483	4	483	1	3	3	3	83	3	3
6	141	5	141	1	1	1	1	41	1	1
7	992	6	992	0	0	2	12	92	2	2
8	121	7	121	1	1	1	1	21	1	1
9	725	8	725	1	1	5	5	25	5	0
10	495	9	495	1	3	5	15	95	5	0
11	853	10	853	1	1	3	13	53	3	3

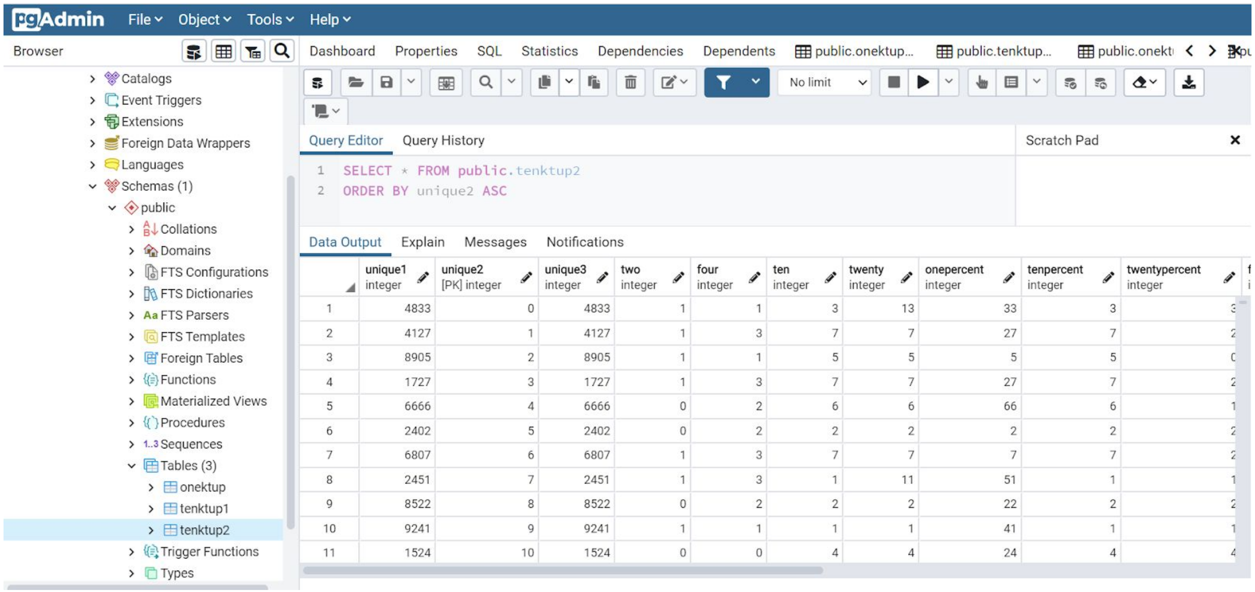
TENKTUP1



```
SELECT * FROM public.tenktup1
ORDER BY unique2 ASC
```

	unique1 integer	unique2 [PK] integer	unique3 integer	two integer	four integer	ten integer	twenty integer	onepercent integer	tenpercent integer	twentypercent integer
1	396	0	396	0	0	6	16	96	6	1
2	5086	1	5086	0	2	6	6	86	6	1
3	1872	2	1872	0	0	2	12	72	2	2
4	5466	3	5466	0	2	6	6	66	6	1
5	8815	4	8815	1	3	5	15	15	5	0
6	56	5	56	0	0	6	16	56	6	1
7	2945	6	2945	1	1	5	5	45	5	0
8	2738	7	2738	0	2	8	18	38	8	3
9	2368	8	2368	0	0	8	8	68	8	3
10	1279	9	1279	1	3	9	19	79	9	4
11	7290	10	7290	0	2	0	10	90	0	0

TENKTUP2



```
SELECT * FROM public.tenktup2
ORDER BY unique2 ASC
```

	unique1 integer	unique2 [PK] integer	unique3 integer	two integer	four integer	ten integer	twenty integer	onepercent integer	tenpercent integer	twentypercent integer
1	4833	0	4833	1	1	3	13	33	3	3
2	4127	1	4127	1	3	7	7	27	7	2
3	8905	2	8905	1	1	5	5	5	5	0
4	1727	3	1727	1	3	7	7	27	7	2
5	6666	4	6666	0	2	6	6	66	6	1
6	2402	5	2402	0	2	2	2	2	2	2
7	6807	6	6807	1	3	7	7	7	7	2
8	2451	7	2451	1	3	1	11	51	1	1
9	8522	8	8522	0	2	2	2	22	2	2
10	9241	9	9241	1	1	1	1	41	1	1
11	1524	10	1524	0	0	4	4	24	4	4

2. README-Part I. 1-2 pages including:
Link - https://github.com/swet09/DB-Implementation-Project#db-implementation-project

DB-Implementation-Project

Introduction

The project's aim is to evaluate a single relational system such as PostgreSQL by changing system parameters and varying relation or data size and through this project we have an increased understanding of database performance and improve programming experience with a RDBMS system such as PostgreSQL.

Brief description of your data generation process

The main objective of part 1 is to generate data based on the Wiscosin Benchmark specification. The original benchmark was composed of three relations (onektup, tenktup1 and tenktup2), each relation was composed of the thirteen integer attributes and three 52 byte string attributes. The three relations or data are generated using a Java. The 'Main.java' file consists of the main code for data generation and the 'DbAdapter.java' file in the dao folder consists of the Database connection configurations. The generated data files are placed in a folder named ‘generated_data’.

System we are working with and why

The database system we have chosen to work with is PostgreSQL and for data generation we have decided to write code in the programming language Java, because of our prior knowledge and experience with both PostgreSQL and Java is why we decided to go with them. We plan to develop benchmark comparisons using queries from Wisconsin Benchmark and evaluate against a progress system. We have decided to use a single database system as we want to keep the other parameters such as main memory, OS load and processor speed constant while testing the database against different parameter values and optimizer options.

Prerequisites

- Create a VM on GCP
- Configuring postgres(https://cloud.google.com/community/tutorials/setting-up-postgres)
- Allow remote ip to connect to postgres server

Validation that data is loaded into the system

After successful execution of the java code, data is generated as CSV files in ''generated_data' folder as a result of select queries from Postgres.

Lessons learned or issues encountered

The major learnings for this project part 1 was from the Wisconsin Benchmark paper. The paper provided great insight into designing a specification. The information provided by the Benchmark Relation Generator was very useful in generating data files. The important takeaways from the paper were, how the queries in the benchmark should test the performance of the major components of a relational database system and why the semantics and statistics of the underlying relations should be well understood so that it is easy to add new queries and to their behavior. Through this project we learned to generate relations in pgadmin and populate them through java. We did encounter a couple of issues, one was to allow remote ip to connect to postgres server. Another was to randomise the values of unique1. For which we implemented ArrayList and used Collections.shuffle to shuffle that ArrayList. Another issue we faced was while generating values for string4, we used switch case statement to implement it.