# Lab Report 6

## Databases - CSC 3320 121

### Hudson Arney

### 27 March 2024

What have you learned from the lab?

* I learned how to set up an SQL Server from scratch.
* How to create a table and import data into SQL.
* How the tables can work together to create an optimally functional database.

What caused issues?

* It was difficult to keep track of all the steps and documentation. I feel like I did to much.

What went well or surprised you?

* Even though there were a lot of firsts in this lab it went mostly smooth (although plugging and chugging isn’t the hardest activity)
* Even though we didn’t have to come up with a lot of our own ideas for this lab, I was still able to learn quite a bit and was a good intro to using SQL for storing data.

Data Loading – Gene\_Info

Innodb

Query: 'C:/Users/arneyh/OneDrive - Milwaukee School of Engineering/Current Classes/CSC 3320 - Databases/Labs/Lab 1/gene\_info/gene\_info' into table genomics.gene\_info\_innodb fields terminated by '\t' lines terminated by '\n' IGNORE 1 LINES;

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

Total Rows = '4354758'

Myisam

Query: 'C:/Users/arneyh/OneDrive - Milwaukee School of Engineering/Current Classes/CSC 3320 - Databases/Labs/Lab 1/gene\_info/gene\_info' into table genomics.gene\_info\_myisam fields terminated by '\t' lines terminated by '\n' IGNORE 1 LINES;

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

Total Rows = '4354758'

Memory

Query: Load Data local Infile 'C:/Users/arneyh/OneDrive - Milwaukee School of Engineering/Current Classes/CSC 3320 - Databases/Labs/Lab 1/gene\_info/gene\_info50000.csv' into table genomics.gene\_info\_memory fields terminated by ',' lines terminated by '\n' IGNORE 1 LINES;

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

Total Rows = ‘412’

Data Loading – GenePubMed

Innodb

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

Total Rows = '11777868'

Myisam

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

Total Rows = '11777868'

Memory

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

Total Rows = 420661 # without memory allocation set

**QUERIES:**

Load Data local Infile

'C:/Users/arneyh/OneDrive - Milwaukee School of Engineering/Current Classes/CSC 3320 - Databases/Labs/Lab 1/gene2pubmed/gene2pubmed' into table

genomics.gene2pubmed\_innodb fields terminated by '\t' lines terminated by '\n' IGNORE 1 LINES;

Load Data local Infile

'C:/Users/arneyh/OneDrive - Milwaukee School of Engineering/Current Classes/CSC 3320 - Databases/Labs/Lab 1/gene2pubmed/gene2pubmed' into table

genomics.gene2pubmed\_myisam fields terminated by '\t' lines terminated by '\n' IGNORE 1 LINES;

Load Data local Infile

'C:/Users/arneyh/OneDrive - Milwaukee School of Engineering/Current Classes/CSC 3320 - Databases/Labs/Lab 1/gene2pubmed/gene2pubmed' into table

genomics.gene2pubmed\_memory fields terminated by '\t' lines terminated by '\n' IGNORE 1 LINES;

Final Counts

Innodb

A screenshot of a computer

Description automatically generated

A screenshot of a computer program

Description automatically generated

Myisam

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

Memory

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

Question 1

Innodb

A screenshot of a computer

Description automatically generated

Myisam



Memory



Profiles

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

InnoDB

A screenshot of a computer

Description automatically generated

MyISAM

A screenshot of a computer

Description automatically generated

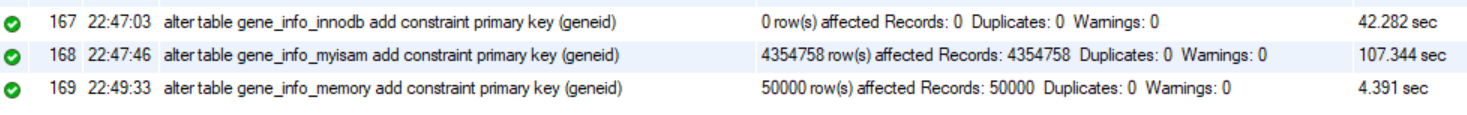
MEMORY

A screenshot of a computer

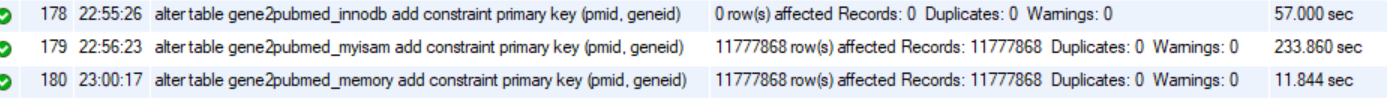
Description automatically generated

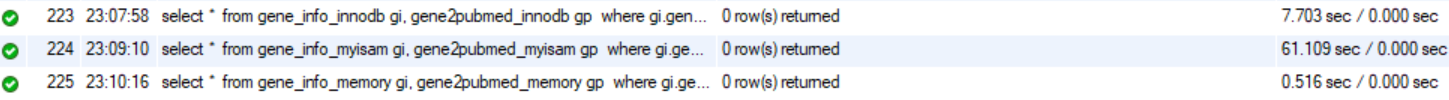
Add Primary Key Times

Gene\_Info



GenePubMed



Times after  


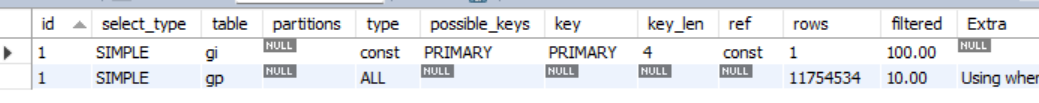
^Show Profiles

A close up of text

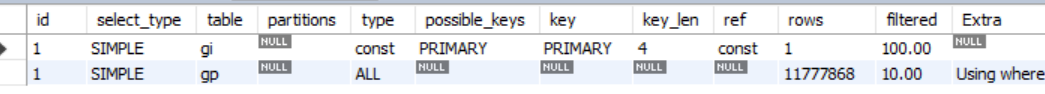
Description automatically generated

Explain Select

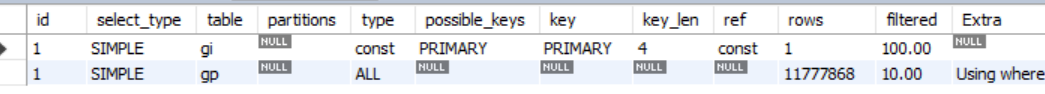
InnoDB



MyISAM



Memory



The PK may not be used for gene2pubmed because the table doesn't have a suitable index defined on the geneid column. In this case MySQL would resort to a full table scan of every row in the table to find matching records.

Innodb

A screenshot of a computer

Description automatically generated

Myisam

A screenshot of a computer

Description automatically generated

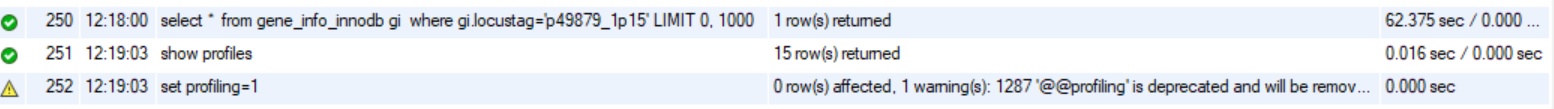
Memory

A screenshot of a computer

Description automatically generated

For all tables it is just using the PK for all of them

Question 2



A screenshot of a computer

Description automatically generated

Question 3

Question 4

Question 5

Innodb

Myisam

Memory

# create database genomics;

/\*

# record variables

SHOW VARIABLES LIKE 'max\_heap\_table\_size';

# 16777216

SHOW VARIABLES LIKE 'tmp\_table\_size';

# 87031808

select @@max\_heap\_table\_size;

# 16777216

select @@tmp\_table\_size;

# 87031808

# set global heap size to 2G

set @@max\_heap\_table\_size=1024 \* 1024 \* 1024 \* 2;

# 2147483648

set @@tmp\_table\_size=1024 \* 1024 \* 1024 \* 2;

#verify

select @@max\_heap\_table\_size;

# 2147483648

select @@tmp\_table\_size;

# 2147483648

\*/

/\*

create table if not exists gene\_info\_innodb (

tax\_id int,

GeneID int,

Symbol varchar(48),

LocusTag varchar(48),

Synonyms varchar(1000),

dbXrefs varchar(512),

chromosome varchar(48),

map\_location varchar(48),

description varchar(4000),

type\_of\_gene varchar(48),

Symbol\_from\_nomenclature\_authority varchar(64),

Full\_name\_from\_nomenclature\_authority varchar(256),

Nomenclature\_status varchar(24),

Other\_designations varchar(4000),

Modification\_date varchar(24)) ENGINE=INNODB;

create table if not exists gene\_info\_myisam (

tax\_id int,

GeneID int,

Symbol varchar(48),

LocusTag varchar(48),

Synonyms varchar(1000),

dbXrefs varchar(512),

chromosome varchar(48),

map\_location varchar(48),

description varchar(4000),

type\_of\_gene varchar(48),

Symbol\_from\_nomenclature\_authority varchar(64),

Full\_name\_from\_nomenclature\_authority varchar(256),

Nomenclature\_status varchar(24),

Other\_designations varchar(4000),

Modification\_date varchar(24)) ENGINE=MYISAM;

create table if not exists gene\_info\_memory (

tax\_id int,

GeneID int,

Symbol varchar(48),

LocusTag varchar(48),

Synonyms varchar(1000),

dbXrefs varchar(512),

chromosome varchar(48),

map\_location varchar(48),

description varchar(4000),

type\_of\_gene varchar(48),

Symbol\_from\_nomenclature\_authority varchar(64),

Full\_name\_from\_nomenclature\_authority varchar(256),

Nomenclature\_status varchar(24),

Other\_designations varchar(4000),

Modification\_date varchar(24)) ENGINE=MEMORY;

\*/

# verify

#desc gene\_info\_myisam;

# select \* from gene\_info\_innodb order by tax\_id;

#drop table if exists gene2pubmed;

/\*

create table gene2pubmed\_innodb (

tax\_id int,

GeneID int,

PMID int) ENGINE=INNODB;

\*/

/\*

create table if not exists gene2pubmed\_myisam (

tax\_id int,

GeneID int,

PMID int) ENGINE=MYISAM;

\*/

/\*

create table if not exists gene2pubmed\_memory (

tax\_id int,

GeneID int,

PMID int) ENGINE=MEMORY;

Load Data local Infile

'C:/Users/arneyh/OneDrive - Milwaukee School of Engineering/Current Classes/CSC 3320 - Databases/Labs/Lab 1/gene2pubmed/gene2pubmed' into table

genomics.gene2pubmed\_innodb fields terminated by '\t' lines terminated by '\n' IGNORE 1 LINES;

\*/

/\*

Load Data local Infile

'C:/Users/arneyh/OneDrive - Milwaukee School of Engineering/Current Classes/CSC 3320 - Databases/Labs/Lab 1/gene2pubmed/gene2pubmed' into table

genomics.gene2pubmed\_myisam fields terminated by '\t' lines terminated by '\n' IGNORE 1 LINES;

\*/

/\*

Load Data local Infile

'C:/Users/arneyh/OneDrive - Milwaukee School of Engineering/Current Classes/CSC 3320 - Databases/Labs/Lab 1/gene2pubmed/gene2pubmed' into table

genomics.gene2pubmed\_memory fields terminated by '\t' lines terminated by '\n' IGNORE 1 LINES;

# this takes a while

\*/

# note the number of records loaded

# verify

#desc gene2pubmed\_memory;

#select \* from gene2pubmed\_myisam limit 100;

# note row count

#select count(\*) from gene\_info\_memory;

#select count(\*) from gene2pubmed\_innodb;

#select count(\*) from gene2pubmed\_myisam;

#select count(\*) from gene2pubmed\_memory;

#####################################################

# Q1 - join

# first, set profiling on

#set profiling=0; # off

#set profiling=1; # on

/\*

select \*

from gene\_info\_innodb gi, gene2pubmed\_innodb gp

where gi.geneid=gp.geneid

and gi.geneid=4126706;

\*/

/\*

select \*

from gene\_info\_myisam gi, gene2pubmed\_myisam gp

where gi.geneid=gp.geneid

and gi.geneid=4126706;

\*/

/\*

select \*

from gene\_info\_memory gi, gene2pubmed\_memory gp

where gi.geneid=gp.geneid

and gi.geneid=4126706;

\*/

# list each query with duration

# show profiles;

# note execution time, you should compare it to the final time for Q1 in your observations.

/\*

# analyze query plan

explain select \*

from gene\_info\_innodb gi, gene2pubmed\_innodb gp

where gi.geneid=gp.geneid

and gi.geneid=4126706; # 1047684

\*/

/\*

explain select \*

from gene\_info\_myisam gi, gene2pubmed\_myisam gp

where gi.geneid=gp.geneid

and gi.geneid=4126706; # 1047684\*/

/\*

explain select \*

from gene\_info\_memory gi, gene2pubmed\_memory gp

where gi.geneid=gp.geneid

and gi.geneid=4126706; # 1047684

\*/

# note what keys are being used and the number of rows accessed from each table

/\*

#lets try adding primary keys

alter table gene\_info\_innodb add constraint primary key (geneid);

alter table gene\_info\_myisam add constraint primary key (geneid);

alter table gene\_info\_memory add constraint primary key (geneid);

\*/

/\*

alter table gene2pubmed\_innodb add constraint primary key (pmid, geneid);

alter table gene2pubmed\_myisam add constraint primary key (pmid, geneid);

alter table gene2pubmed\_memory add constraint primary key (pmid, geneid);

\*/

# note the time to create each primary key

# re-execute the same query

#set profiling=0;

#set profiling=1;

/\*

select \*

from gene\_info\_innodb gi, gene2pubmed\_innodb gp

where gi.geneid=gp.geneid

and gi.geneid=4126706;

\*/

/\*

select \*

from gene\_info\_myisam gi, gene2pubmed\_myisam gp

where gi.geneid=gp.geneid

and gi.geneid=4126706;

\*/

/\*

select \*

from gene\_info\_memory gi, gene2pubmed\_memory gp

where gi.geneid=gp.geneid

and gi.geneid=4126706;

\*/

#show profiles;

# note execution time (but this doesn't go in table for Q1)

# you should compare it to the final time for Q1 in your observations

# analyze query plan

# For InnoDB tables

/\*

EXPLAIN SELECT \*

FROM gene\_info\_innodb gi, gene2pubmed\_innodb gp

WHERE gi.geneid = gp.geneid

AND gi.geneid = 4126706;

# For MyISAM tables

EXPLAIN SELECT \*

FROM gene\_info\_myisam gi, gene2pubmed\_myisam gp

WHERE gi.geneid = gp.geneid

AND gi.geneid = 4126706;

# For MEMORY tables

EXPLAIN SELECT \*

FROM gene\_info\_memory gi, gene2pubmed\_memory gp

WHERE gi.geneid = gp.geneid

AND gi.geneid = 4126706;

\*/

/\*

-- For gene2pubmed\_innodb

alter table gene2pubmed\_innodb drop primary key;

alter table gene2pubmed\_innodb add constraint primary key (geneid, pmid);

set profiling=1;

select \*

from gene\_info\_innodb gi, gene2pubmed\_innodb gp

where gi.geneid=gp.geneid

and gi.geneid=4126706;

show profiles;

explain select \*

from gene\_info\_innodb gi, gene2pubmed\_innodb gp

where gi.geneid=gp.geneid

and gi.geneid=4126706;

-- For gene2pubmed\_myisam

alter table gene2pubmed\_myisam drop primary key;

alter table gene2pubmed\_myisam add constraint primary key (geneid, pmid);

set profiling=1;

select \*

from gene\_info\_myisam gi, gene2pubmed\_myisam gp

where gi.geneid=gp.geneid

and gi.geneid=4126706;

show profiles;

explain select \*

from gene\_info\_myisam gi, gene2pubmed\_myisam gp

where gi.geneid=gp.geneid

and gi.geneid=4126706;

-- For gene2pubmed\_memory

alter table gene2pubmed\_memory drop primary key;

alter table gene2pubmed\_memory add constraint primary key (geneid, pmid);

set profiling=1;

select \*

from gene\_info\_memory gi, gene2pubmed\_memory gp

where gi.geneid=gp.geneid

and gi.geneid=4126706;

show profiles;

explain select \*

from gene\_info\_memory gi, gene2pubmed\_memory gp

where gi.geneid=gp.geneid

and gi.geneid=4126706;

\*/

# record what keys are being used and the number of rows accessed from each table

# record your observations! Why is the PK not being used for gene2pubmed??

# lets re-create the gene2pubmed PK based on join criteria.

/\*

alter table gene2pubmed drop primary key;

alter table gene2pubmed add constraint primary key (geneid, pmid);

# re-execute the same query

set profiling=1;

-----

select \*

from gene\_info gi, gene2pubmed gp

where gi.geneid=gp.geneid

and gi.geneid=4126706;

show profiles;

# note execution time (record this time for the table for Q1)

# analyze query plan

explain select \*

from gene\_info gi, gene2pubmed gp

where gi.geneid=gp.geneid

and gi.geneid=4126706;

# note what keys are being used and the number of rows accessed from each table

# record your observations

\*/

####################################################################

# Q2 - restriction

#set profiling=0;

-- For gene\_info\_innodb

set profiling=1;

select \*

from gene\_info\_innodb gi

where gi.locustag='p49879\_1p15';

show profiles;

-- For gene\_info\_myisam

set profiling=1;

select \*

from gene\_info\_myisam gi

where gi.locustag='p49879\_1p15';

show profiles;

-- For gene\_info\_memory

set profiling=1;

select \*

from gene\_info\_memory gi

where gi.locustag='p49879\_1p15';

show profiles;

# note execution time, you should compare it to the final time for Q2 in your observations.

/\*

# analyze query plan

explain select \*

from gene\_info gi

where gi.locustag='p49879\_1p15';

# note what keys are being used and the number of rows accessed from each table

# create index

create index gene\_info\_locustag on gene\_info( locustag );

# note this time

# re-execute the query

set profiling=0;

set profiling=1;

select \*

from gene\_info gi

where gi.locustag='p49879\_1p15';

show profile;

# note execution time (record this time for the table for Q2)

# analyze query plan

explain select \*

from gene\_info gi

where gi.locustag='p49879\_1p15';

# note what keys are being used and the number of rows accessed from each table

# record your observations

####################################################################

# Q3 - range query

set profiling=0;

set profiling=1;

select \*

from gene\_info

where geneid between '5961931' and '5999886';

show profiles;

# note execution time (record this time for the table for Q3)

# analyze query plan

explain select \*

from gene\_info

where geneid between '5961931' and '5999886';

# note what keys are being used and the number of rows accessed from each table

# record your observations

####################################################################

# Q4 - insert

set profiling=0;

set profiling=1;

insert into gene2pubmed values (9606, 5555, 6666);

show profiles;

# note execution time (record this time for the table for Q4)

# analyze query plan

explain insert into gene2pubmed values (9606, 5555, 6666);

# note what keys are being used and the number of rows accessed from each table

# record your observations

####################################################################

# Q5 - update

set profiling=0;

set profiling=1;

update gene\_info set locustag='No Locus Tag' where locustag='-';

show profiles;

# note execution time (record this time for the table for Q5)

# analyze query plan

explain update gene\_info set locustag='No Locus Tag' where locustag='-';

# note what keys are being used and the number of rows accessed from each table

# record your observations

# make sure to drop both tables before moving on to recreate the tables with a different

# indexing method