

Lab 1 Report

What I've learned

- I learned how to use MySQL Workbench and how to use MySQL in the terminal, how to make tables, how to import data into the tables, how to make tables fit the data, and how to use primary keys and indices.

How it could be improved

- I would like to not have to record the output of what I'm doing. I'd rather just write down observations I had.

Major Steps

- Viewing data in Excel
 - There are 1000 rows and 15 columns. Most of the time, columns E-H and K-N are blank. The first row seems to be like a template because it says "NEWENTRY" and most of the columns are blank.
- Creating a database table
 - Output:
0 row(s) affected 0.015 sec
 - After executing the SQL and refreshing the tab on the left, the table "gene_info" appeared with all of the columns that were specified.
- Loading data into the table
 - Output:
Query OK, 1000 rows affected, 1103 warnings (0.10 sec)
Records: 1000 Deleted: 0 Skipped: 0 Warnings: 1103
 - The table was populated with 1000 rows, but some of the strings in the 'type_of_gene' and 'other_designations' columns were cut off.
- Show warnings
 - Output:

```
+-----+-----+-----+-----+
| Level | Code | Message                                     |
+-----+-----+-----+-----+
| Warning|1265|Data truncated for column 'type_of_gene' at row 2|
| Warning|1265|Data truncated for column 'type_of_gene' at row 3|
| Warning|1265|Data truncated for column 'type_of_gene' at row 4|
...
```

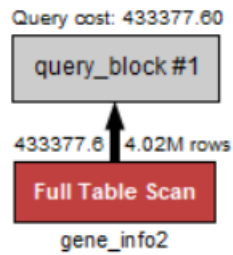
```
|Warning|1265|Data truncated for column 'type_of_gene' at row  
922|
```

```
|Warning|1265|Data truncated for column 'type_of_gene' at row  
923|
```

```
+-----+-----+-----+-----+-----+-----+-----+-----+  
1024 rows in set (0.00 sec)
```

- They are all data truncated warnings for columns 'type_of_gene' and 'other_designations'.
- New table with larger fields
 - Output:
0 row(s) affected 0.094 sec
 - Again, after executing the SQL and refreshing the tab on the left, the table "gene_info" appeared with all of the columns that were specified. The SQL gave an increase in size to the 'type_of_gene' and 'other_designations' columns.
- Reload entire dataset into new table
 - Output:
Query OK, 4354758 rows affected (1 min 52.62 sec)
Records: 4354758 Deleted: 0 Skipped: 0 Warnings: 0
 - It took a long time, but the loading seemed to be successful. It ignored the first row, as it was told.
- Check for warnings and verify the length
 - Output:
Empty set (0.00 sec)
+-----+
| count(*) |
+-----+
| 4354758 |
+-----+
1 row in set (2 min 44.60 sec)
 - There were no warnings and the length of the table is correct.
- Creating the working table
 - Output:
0 row(s) affected 0.094 sec
Query OK, 4354758 rows affected (1 min 41.63 sec)
Records: 4354758 Duplicates: 0 Warnings: 0
 - I successfully created the new table and then inserted the necessary columns from every row of the original table, where GenelD was not blank, into it. I then dropped the original table.
- Primary keys and indices
 - Output:

- Selecting specific gene:
1 row in set (7.94 sec)
- Explain:



```

{
  "query_block": {
    "select_id": 1,
    "cost_info": {
      "query_cost": "433377.60"
    },
  },
  "table": {
    "table_name": "gene_info2",
    "access_type": "ALL",
    "rows_examined_per_scan": 4020816,
    "rows_produced_per_join": 402081,
    "filtered": "10.00",
    "cost_info": {
      "read_cost": "393169.44",
      "eval_cost": "40208.16",
      "prefix_cost": "433377.60",
      "data_read_per_join": "7G"
    },
  },
  "used_columns": [
    "tax_id",
    "GeneID",
    "Symbol",
    "LocusTag",
    "Synonyms",
    "description",
    "type_of_gene",
    "Modification_date"
  ],
  "attached_condition": "(`genomics`.`gene_info2`.`GeneID` = 4226999)"
}

```

- Adding primary keys:
0 row(s) affected Records: 0 Duplicates: 0 Warnings: 0
38.750 sec
- Selecting specific gene again:
1 row(s) returned 0.031 sec / 0.000 sec
- Creating index on *tax_id*:
0 row(s) affected Records: 0
Duplicates: 0 Warnings: 0 18.719 sec
- Creating gene2pubmed table:
0 row(s) affected 0.063 sec
- Loading data into gene2pubmed table:
Query OK, 11777869 rows affected, 3 warnings (2 min 20.66 sec)
Records: 11777869 Deleted: 0 Skipped: 0 Warnings: 3
- After adding the primary keys, selecting rows from the large table were obviously exponentially faster.

Entity Relation Diagram

