**SQL-Mongo Project – Spatial Data of US Wildfires**

BUAN 6320

Submitted by:

***Prem Kumar Pulluri***

|  |
| --- |
| **Activity** |
| Prepared Data Model and Created Physical DB |
| Loaded Data into Database |
| Wrote SQL Queries |
| Prepared Mongo Database |
| Loaded data into Mongo DB |
| Wrote Mongo Queries |
| Prepared Report |
| Reviewed Report |

Contents

Data Model 5

Assumptions/Notes About Data Entities and Relationships 5

Entity-Relationship Diagram 5

Physical Database 6

Assumptions/Notes About Data Set 6

Screen shot of Physical Database objects 6

Data in the Database 6

SQL Queries 7

Query 1 7

Question 7

Notes/Comments About SQL Query and Results (Include # of Rows in Result) 7

Translation 7

Screen Shot of SQL Query and Results 7

Query 2 8

Question 8

Notes/Comments About SQL Query and Results (Include # of Rows in Result) 8

Translation 8

Screen Shot of SQL Query and Results 8

Query 3 9

Question 9

Notes/Comments About SQL Query and Results (Include # of Rows in Result) 9

Translation 9

Screen Shot of SQL Query and Results 9

Query 4 10

Question 10

Notes/Comments About SQL Query and Results (Include # of Rows in Result) 10

Translation 10

Screen Shot of SQL Query and Results 10

Query 5 11

Question 11

Notes/Comments About SQL Query and Results (Include # of Rows in Result) 11

Translation 11

Screen Shot of SQL Query and Results 11

Query 6 12

Question 12

Notes/Comments About SQL Query and Results (Include # of Rows in Result) 12

Translation 12

Screen Shot of SQL Query and Results 12

Data Review for MongoDB 13

Assumptions/Notes About Data Collections, Attributes and Relationships between Collections 13

Physical Mongo Database 14

Assumptions/Notes About Data Set 14

Screen shot of Physical Database objects (Database, Collections and Attributes) 14

Data in the Database 14

MongoDB Queries/Code 15

Query 1 15

Question 15

Notes/Comments About MongoDB Query/Code and Results (Include # of Documents in Result) 15

Translation 15

Screen Shot of MongoDB Query/Code and Results 15

Query 2 16

Question 16

Notes/Comments About MongoDB Query/Code and Results (Include # of Documents in Result) 16

Translation 16

Screen Shot of MongoDB Query/Code and Results 16

Query 3 17

Question 17

Notes/Comments About MongoDB Query/Code and Results (Include # of Documents in Result) 17

Translation 17

Screen Shot of MongoDB Query/Code and Results 17

Query 4 18

Question 18

Notes/Comments About MongoDB Query/Code and Results (Include # of Documents in Result) 18

Translation 18

Screen Shot of MongoDB Query/Code and Results 18

Query 5 19

Question 19

Notes/Comments About MongoDB Query/Code and Results (Include # of Documents in Result) 19

Translation 19

Screen Shot of MongoDB Query/Code and Results 19

Query 6 20

Question 20

Notes/Comments About MongoDB Query/Code and Results (Include # of Documents in Result) 20

Translation 20

Screen Shot of MongoDB Query/Code and Results 20

# Data Model

## Assumptions/Notes About Data Entities and Relationships

Include assumptions about data entities and their relationships with each other.

1. The relationship between the primary key in the above-mentioned tables with the **fire\_table** main table is **one to many relationships** other than Source\_System\_info as the FPA\_ID, which is primary key has single record mapping with the foreign key, all other tables have multiple values repeated in foreign key column.
2. The **nwcg\_agency\_identifier\_table** has the primary key unitID which has a **one to many relationships** to the **Fire\_data\_Table**.
3. nwcg\_location\_info has one to many relationship with **nwcg\_agency\_identifier\_table** with State and Gacc Columns as primary composite key.
4. stat\_cause\_table has one to many relationship with fire\_table with stat\_cause\_code as primary key.
5. location\_info has one to many relationship with fire\_table with Latitude and Longitude as primary composite key.
6. fips\_info has one to many relationship with location\_info with FIPS\_CODE and FIPS\_NAME as composite primary key.
7. mtbs\_info has one to many relationship with fire\_table with MTBS\_ID as primary key.
8. Source\_reporting\_info has one to many relationship with fire\_table with Source\_Reporting\_Unit as the primary key.
9. Owner\_info has one to many relationship with fire\_table with Owner\_ID as the primary key.
10. Ics\_209\_info has one to many relationships with fire\_table with ICS\_209\_Incident\_number as the primary key with Non Identifying relationship.
11. Each owner may own land on many locations, while every single location can have only one land owner.
12. Many fire incidents may occur due to one statistical cause, each fire incident is caused by only one statistical cause.
13. Each MTBS identifier may identify many fire incidents, each fire is identified by only one MTBS identifier.
14. Each ICS-209 report identifier may identify many fire incidents, each fire is identified by only one MTBS identifier.
15. Each FIPS may represent many locations, each location is represented by one FIPS.
16. On one location many times a fire may occur, one fire would occur only on one location
17. Each location may have many NWCG Units, Each NWCG unit would be on only one location.
18. For each Active NWCG unit many fire incidents may occur, for each fire there was one unique NWCG unit active.
19. Each source reporting unit may prepare many fire reports, each fire report is prepared by only one source reporting unit.

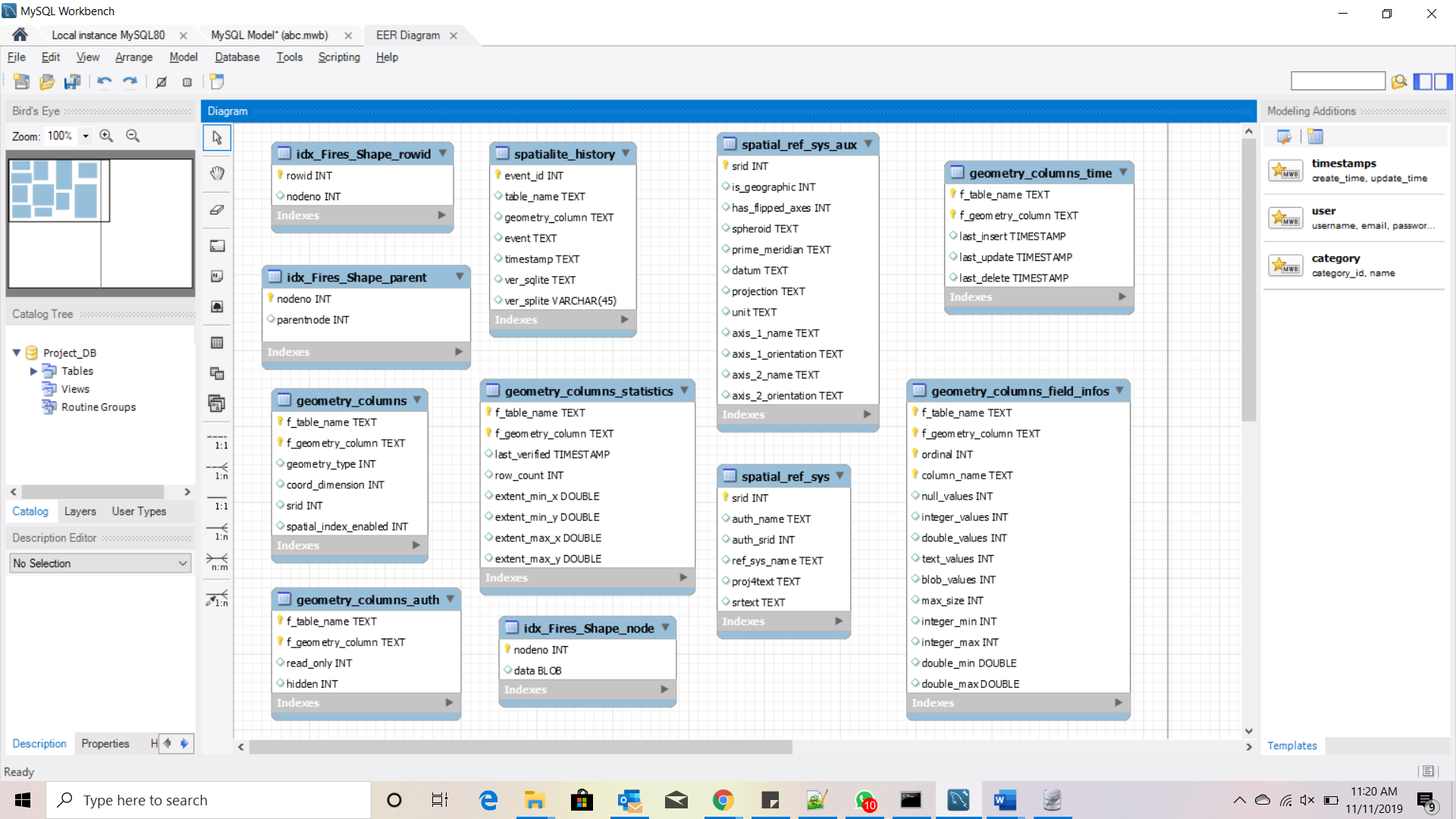
Include reasons why the data model is in 3NF.

* Each table has its own functionality and serves its own purpose.
* Every table has a primary key and each table contains only atomic values instead of group values which satisfies the condition of 1NF.
* Every column in each table is dependent on a single primary key and not on a composite key which satisfies the condition of 2NF.
* There are no transitive functional dependencies between any column in table which satisfies the condition of 3NF

## Entity-Relationship Diagram

## ER DIAGRAM

## 



# Physical Database

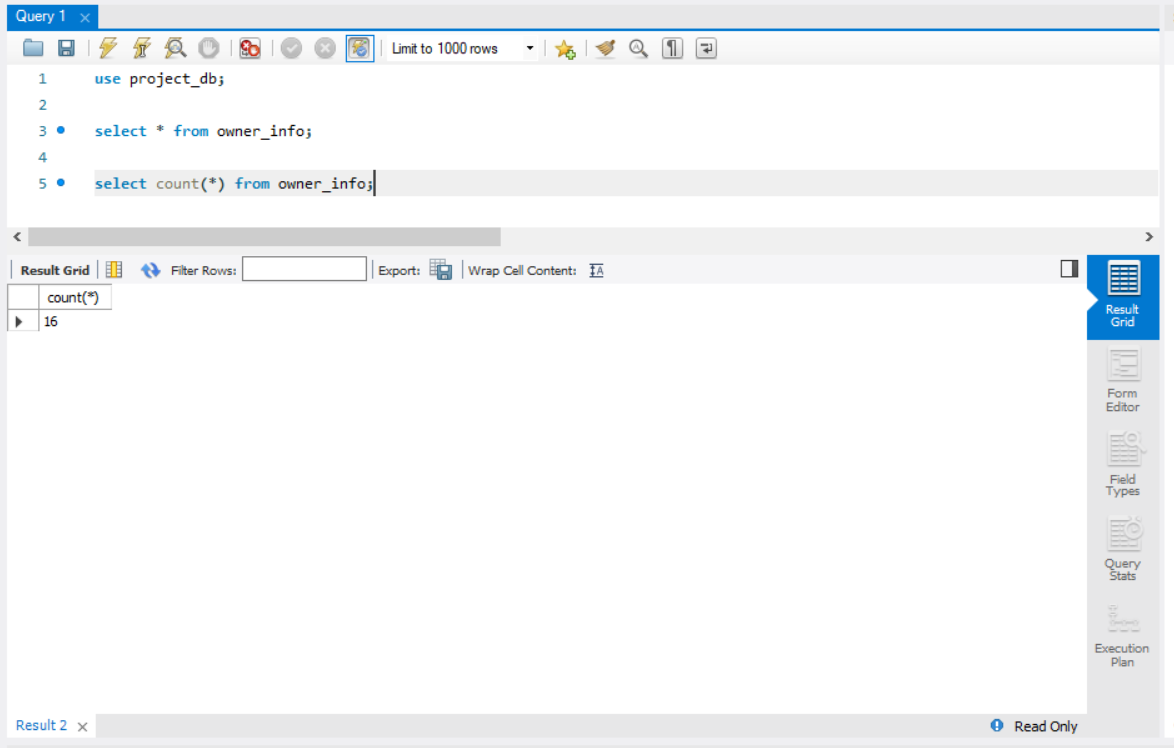
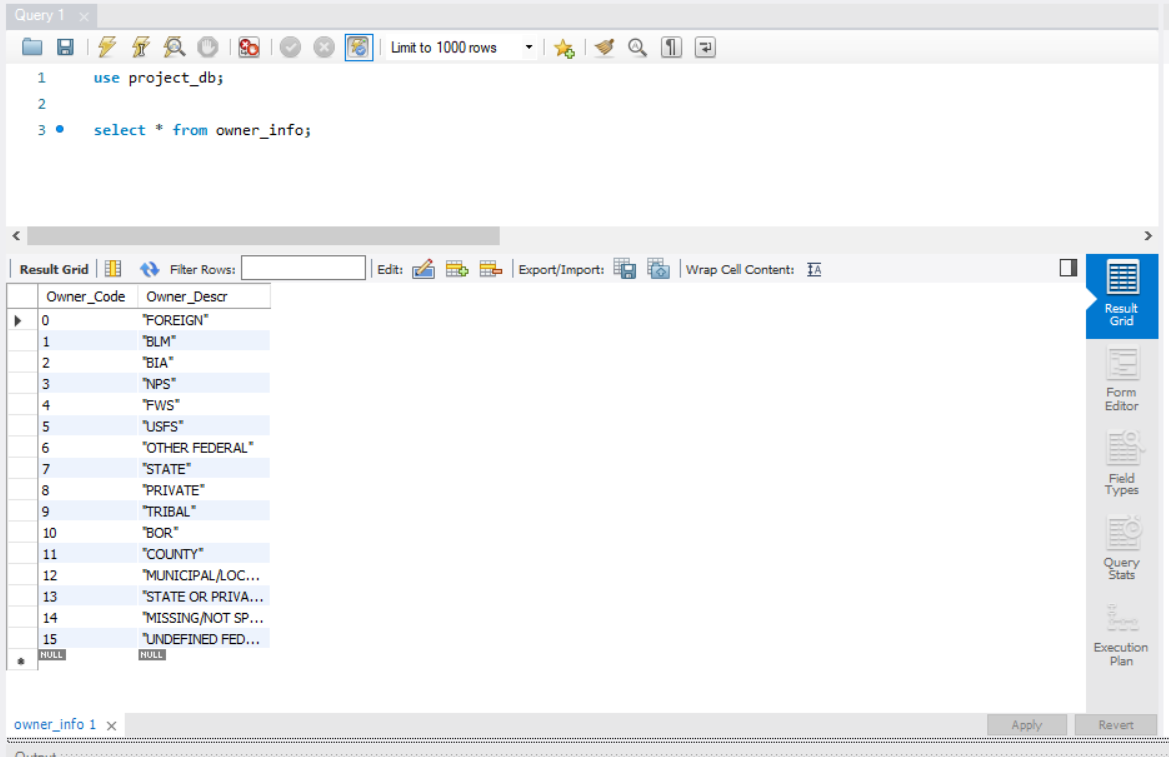
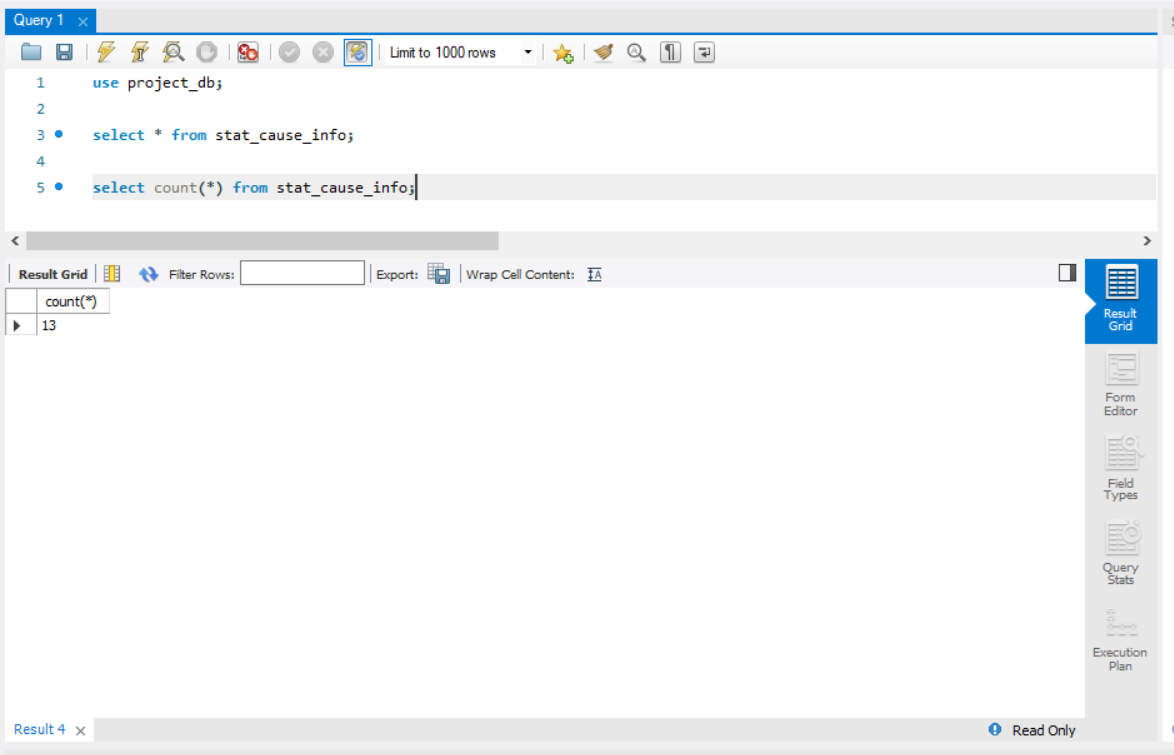
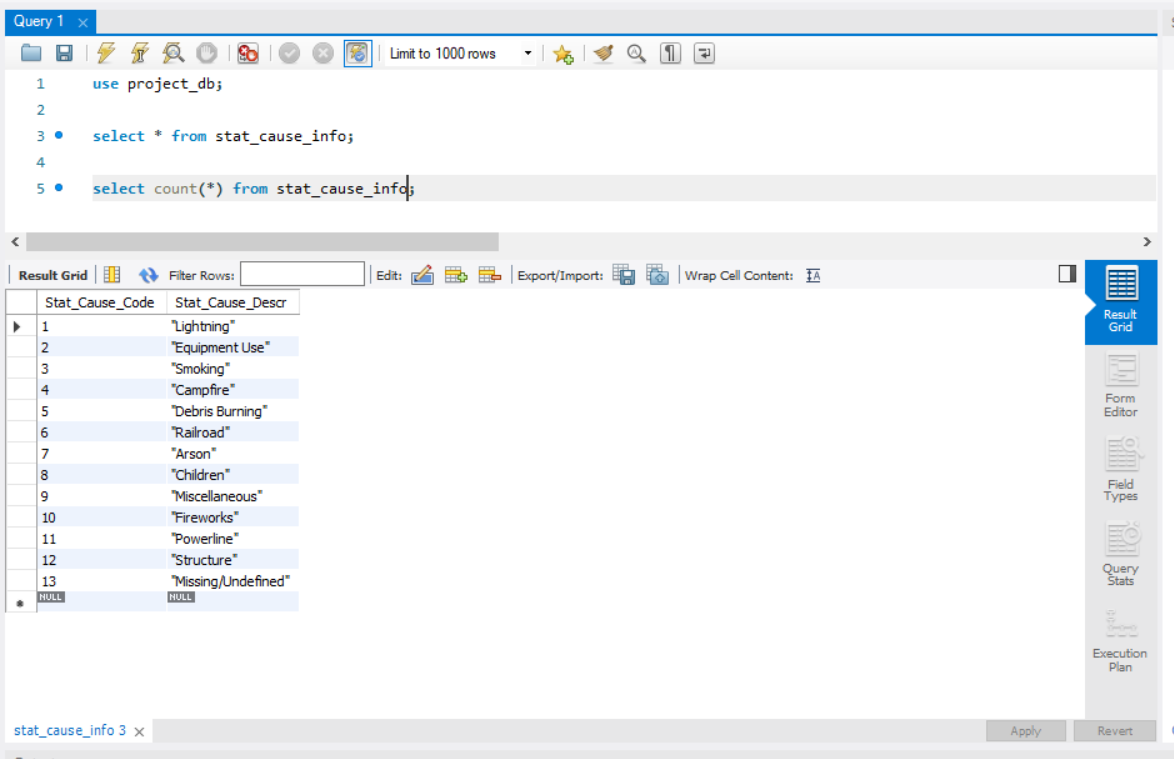
## Assumptions/Notes About Data Set

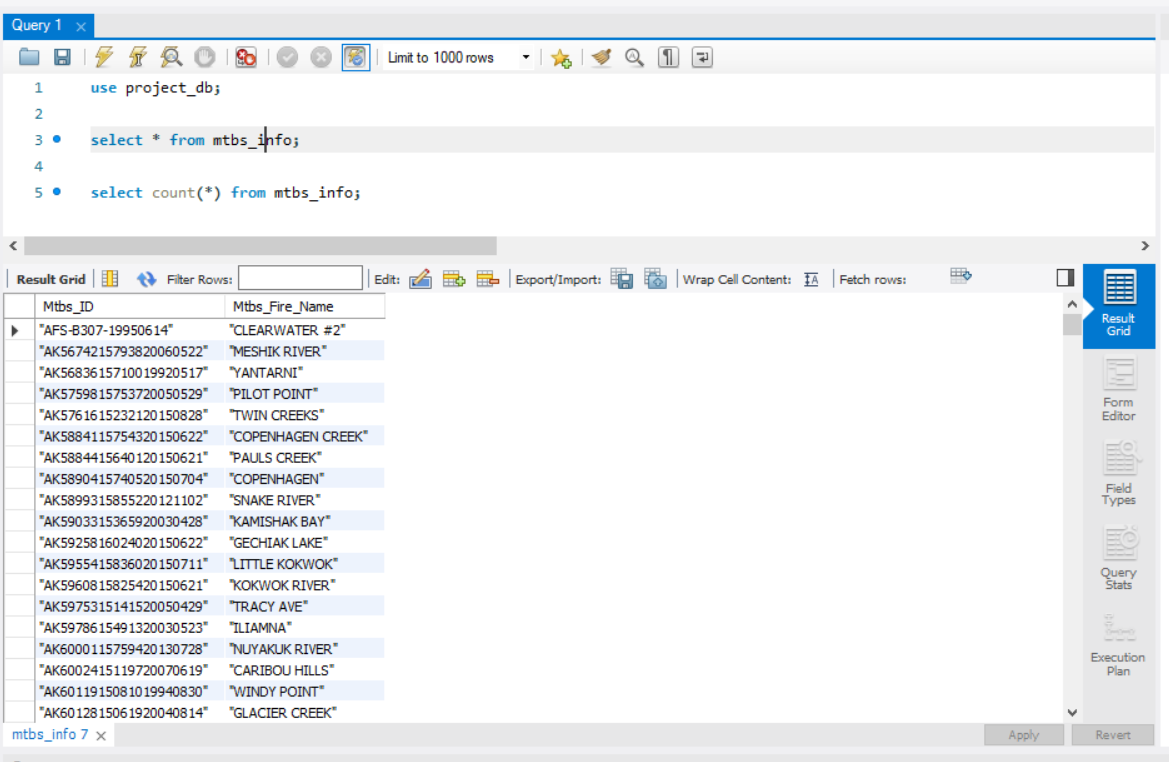
Include any assumptions made about data such as empty fields, sparse data, bad data, etc.

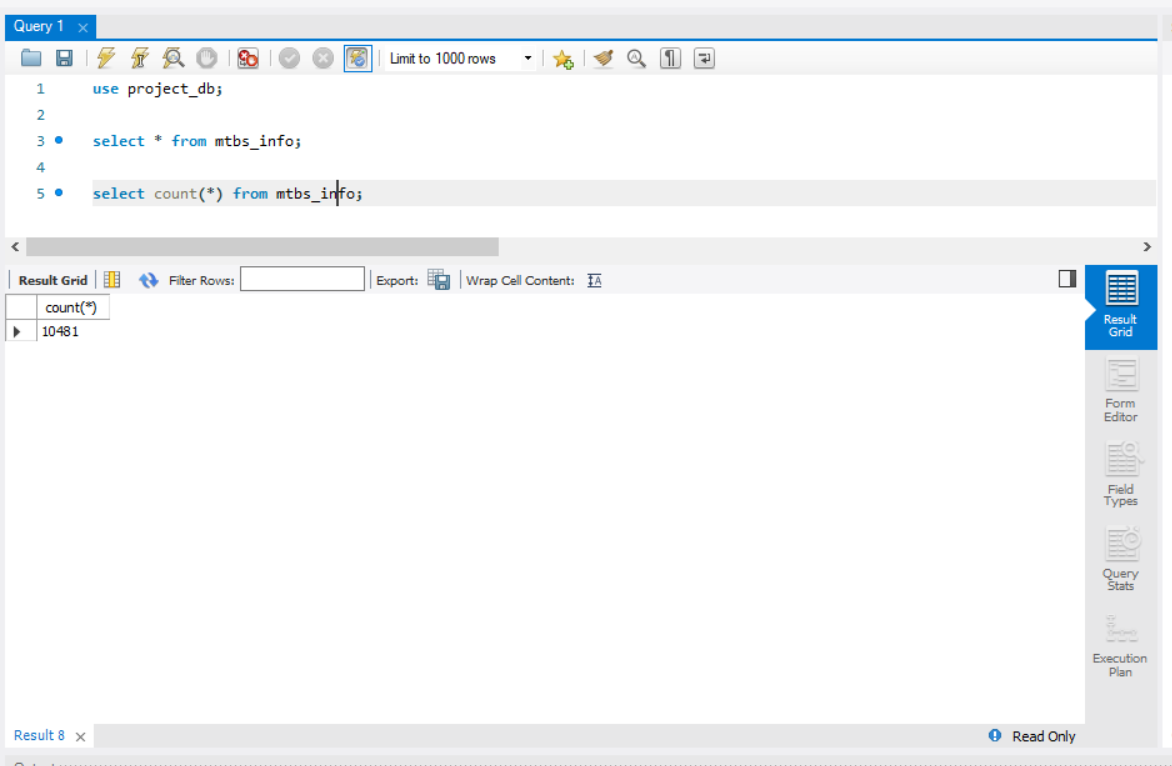
**Bad Data**:

1. The column county in the original fire table does not have a **normalized entry** for example the county name harding is entered as Harding or Harding-county or 39 in different rows because of which we are unable to take the combination of state and county as unique values
2. The combination of columns fips\_code,state should be unique value but due to no normalized entry of fips\_name we are unable to use the combination as a primary key.

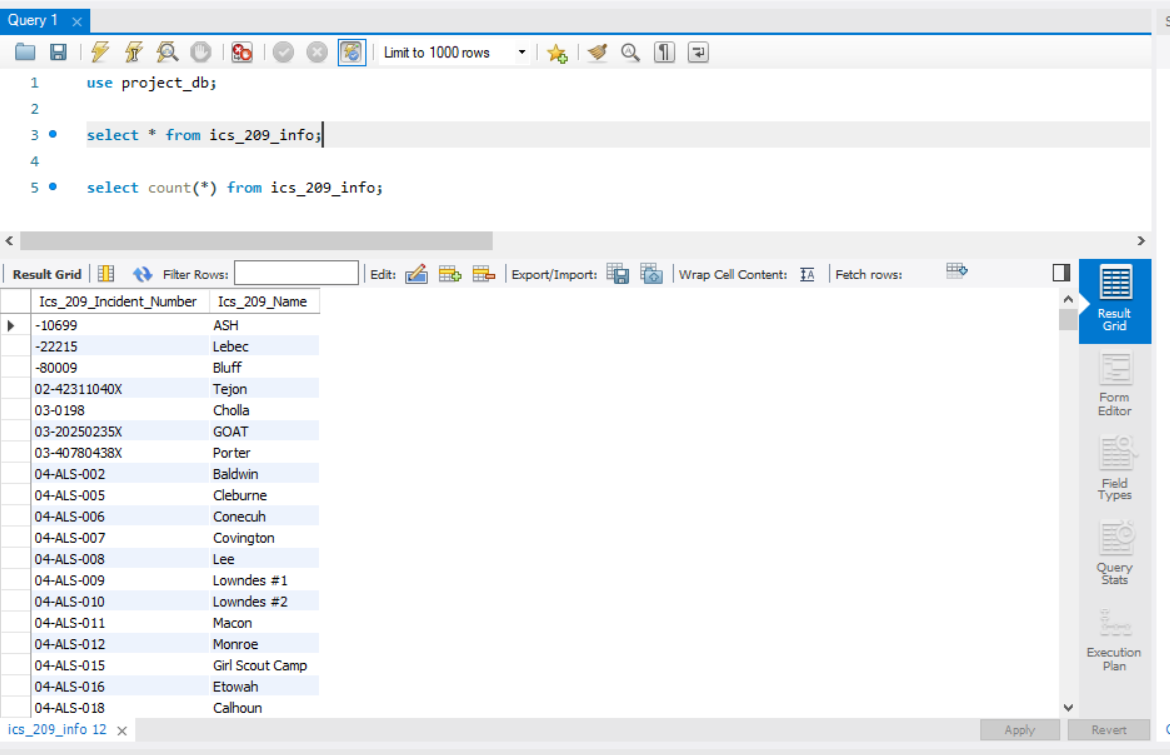
## Screen shot of Physical Database objects

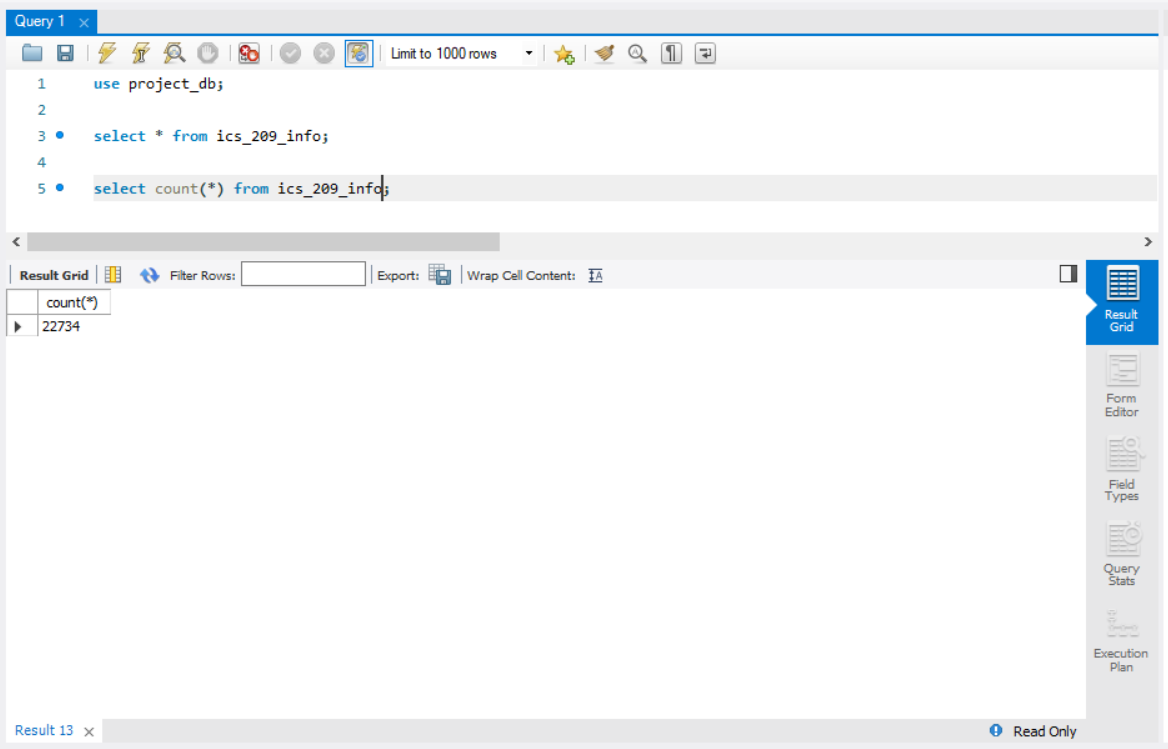
* owner\_info
* stat\_cause\_info
* mtbs\_info



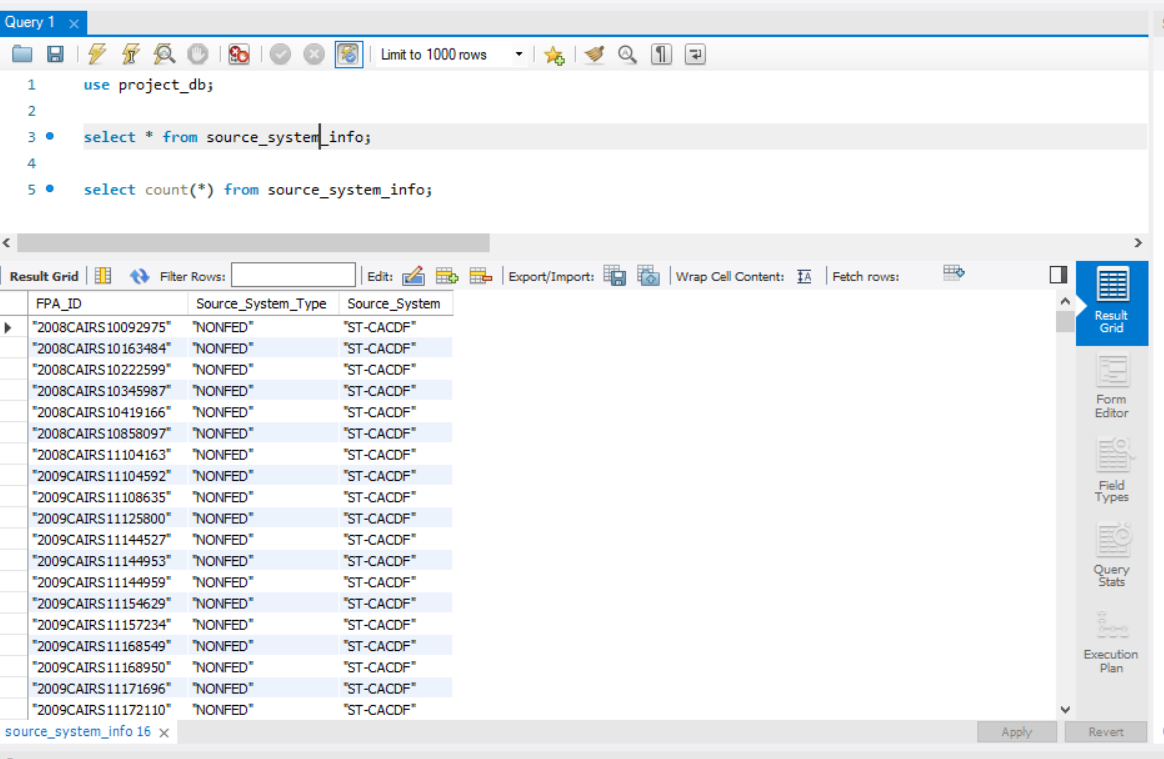
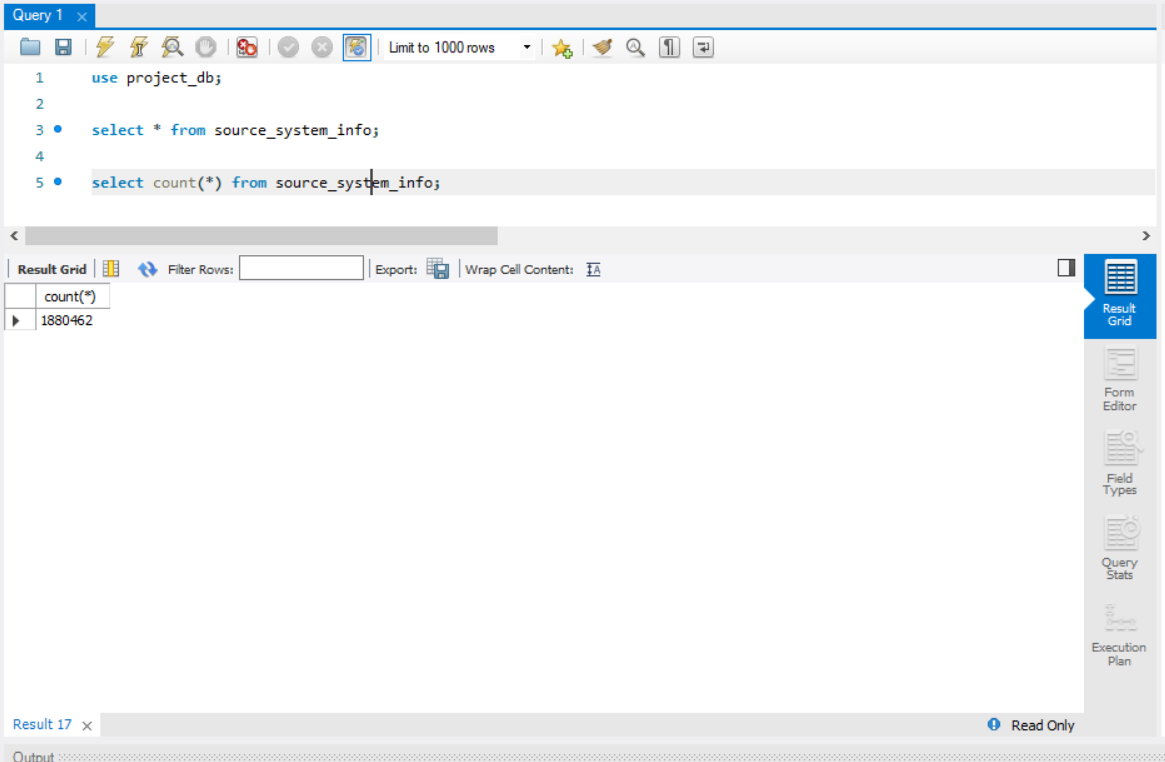


* ics\_209\_info

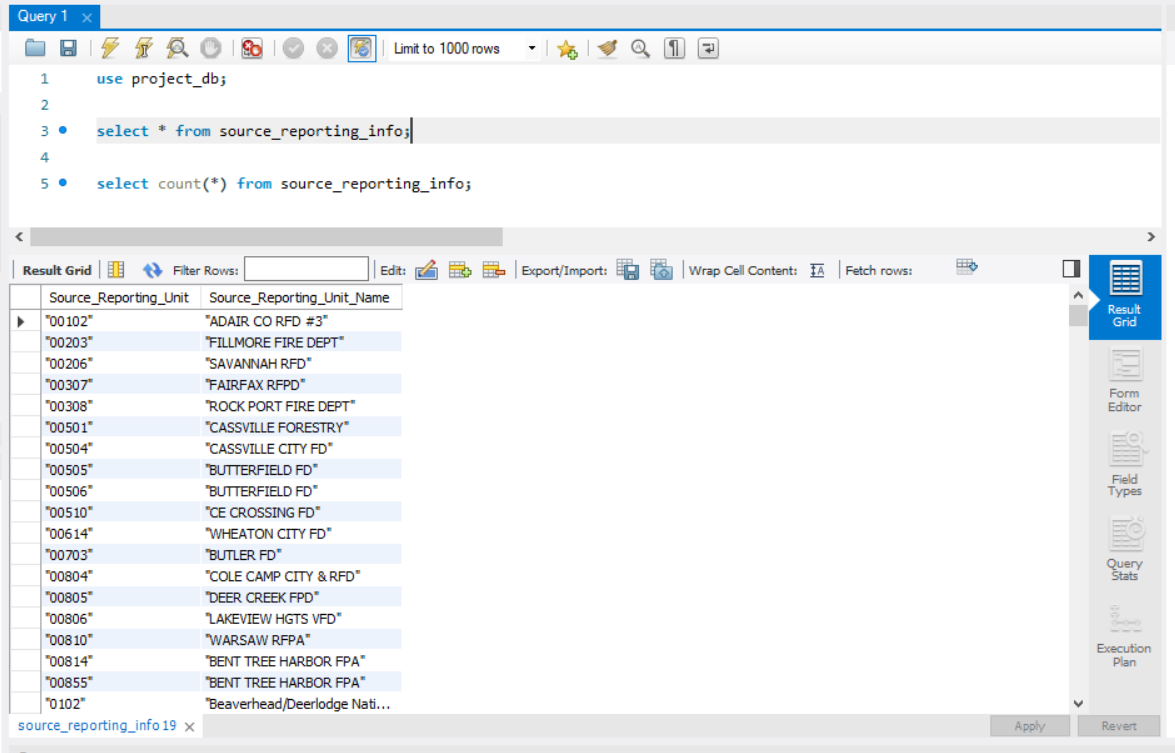
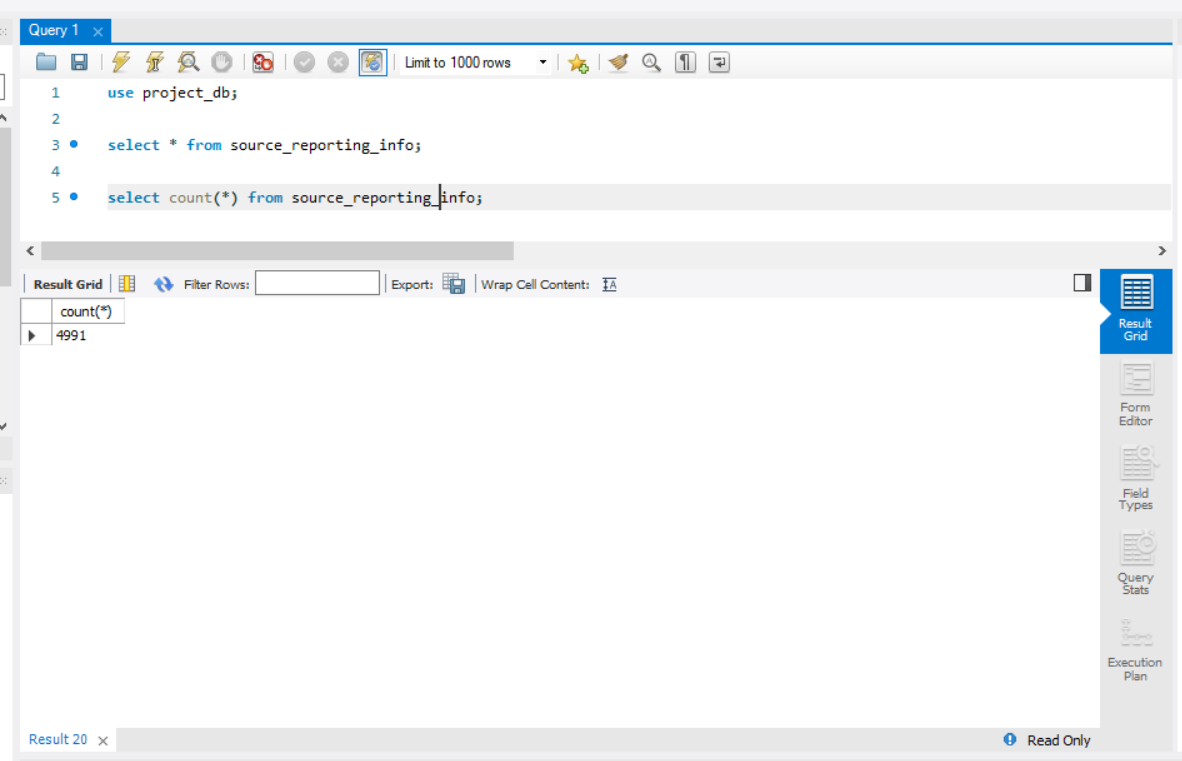




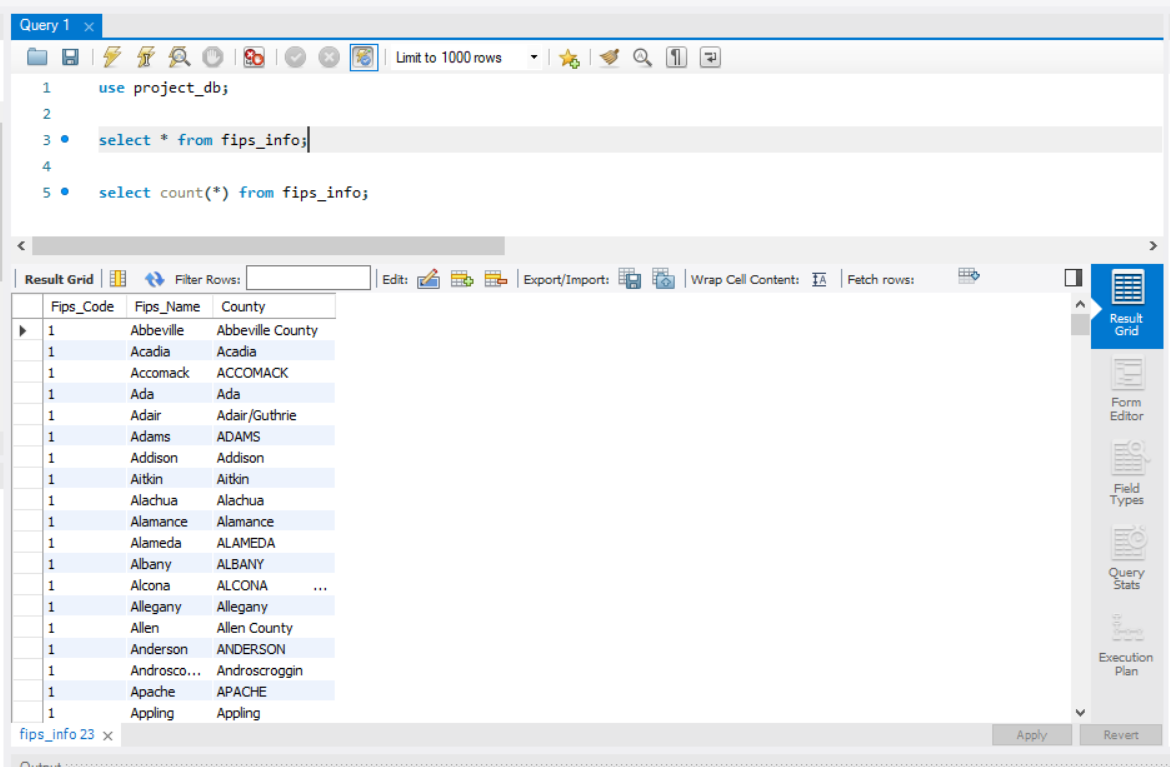
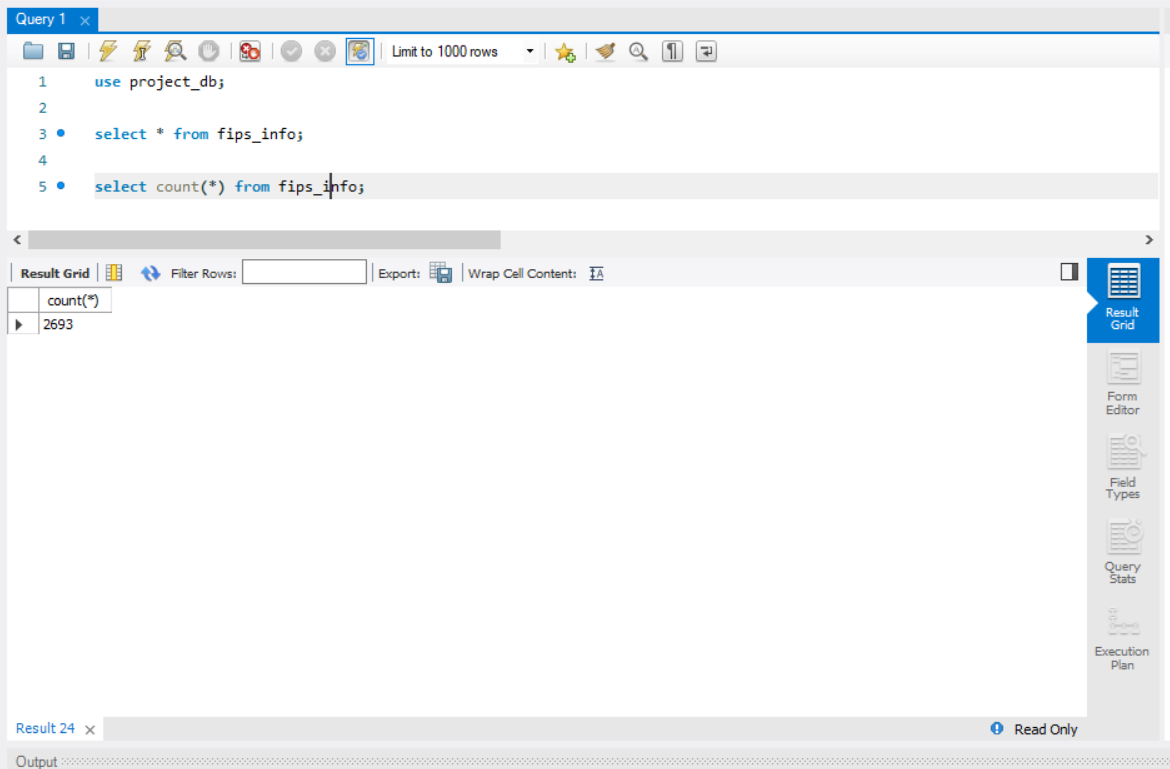
* source\_system\_info

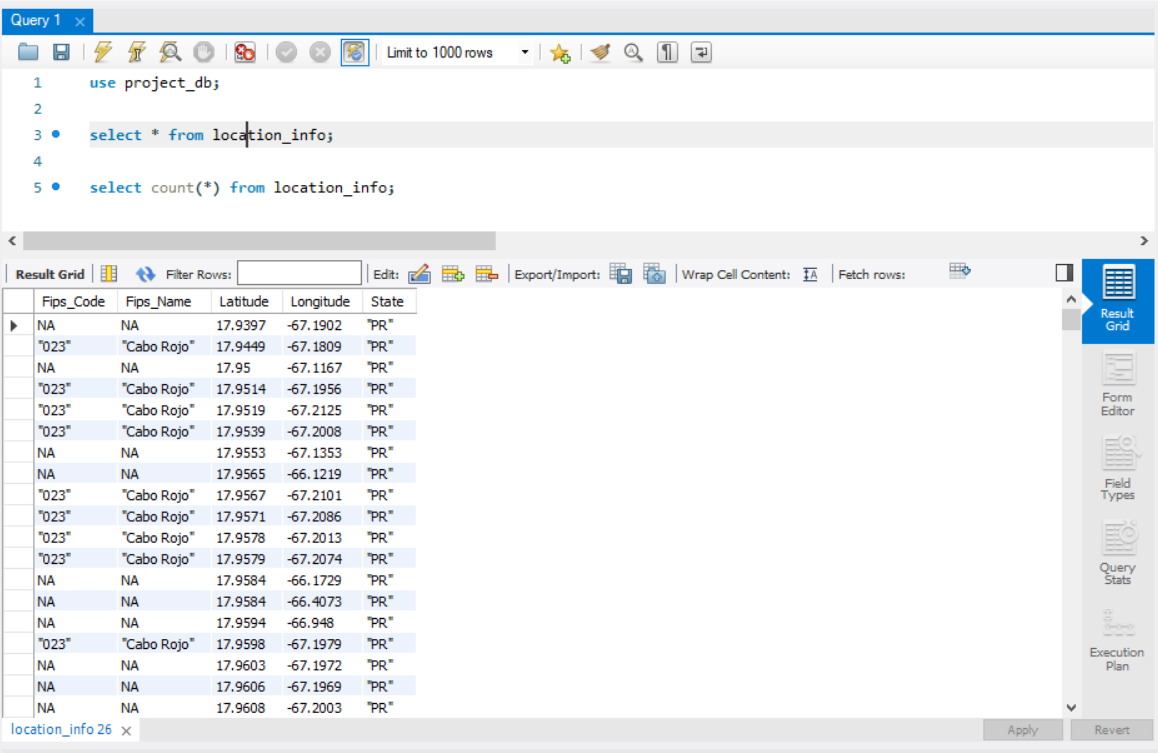
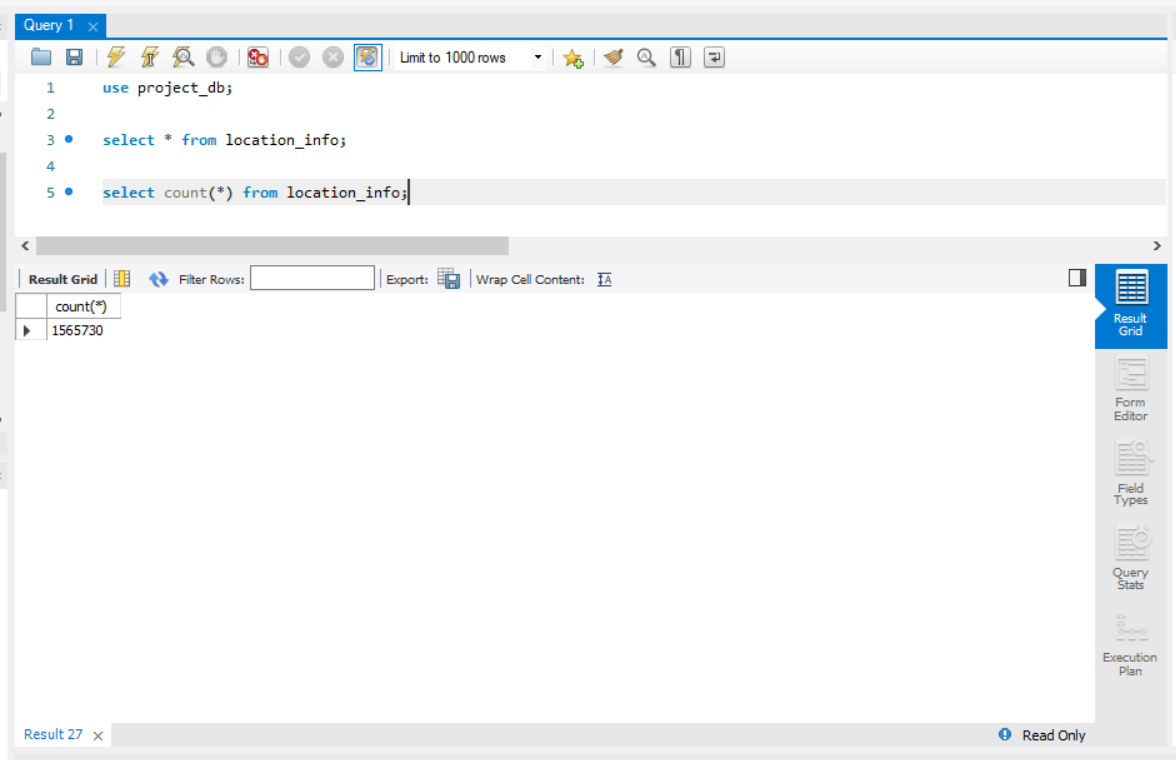
* source\_reporting\_info

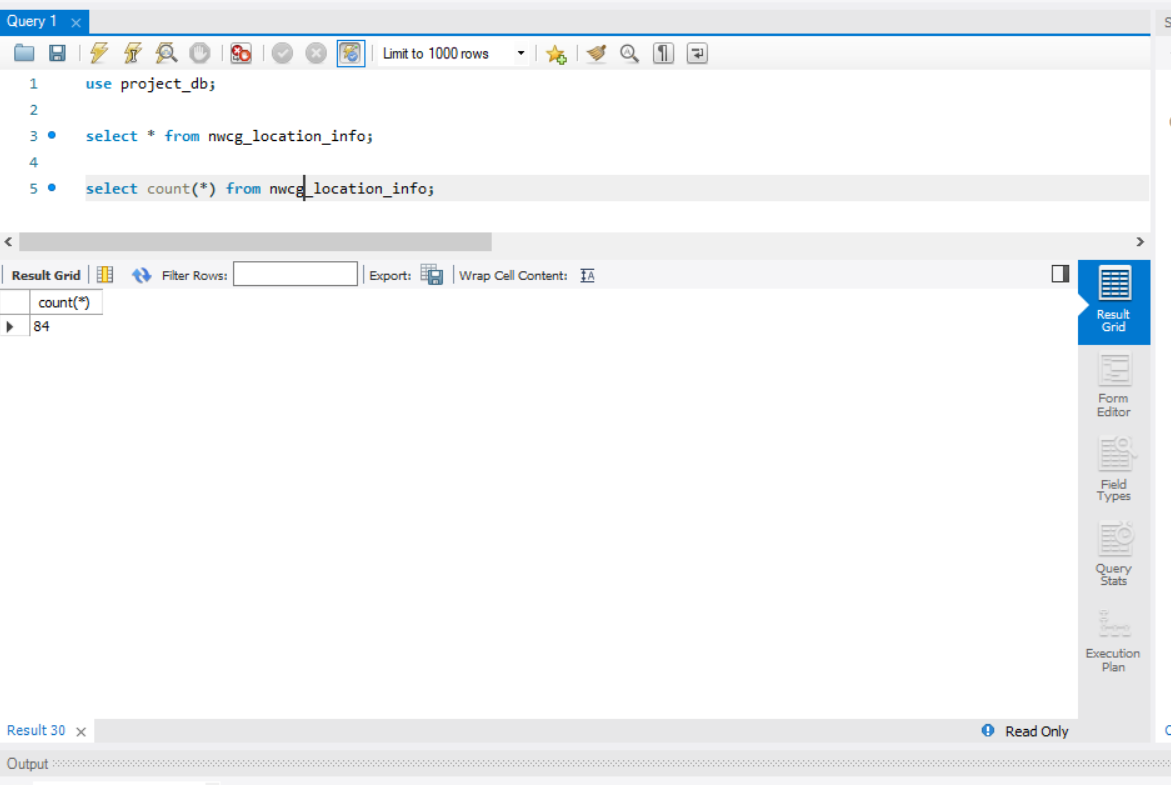
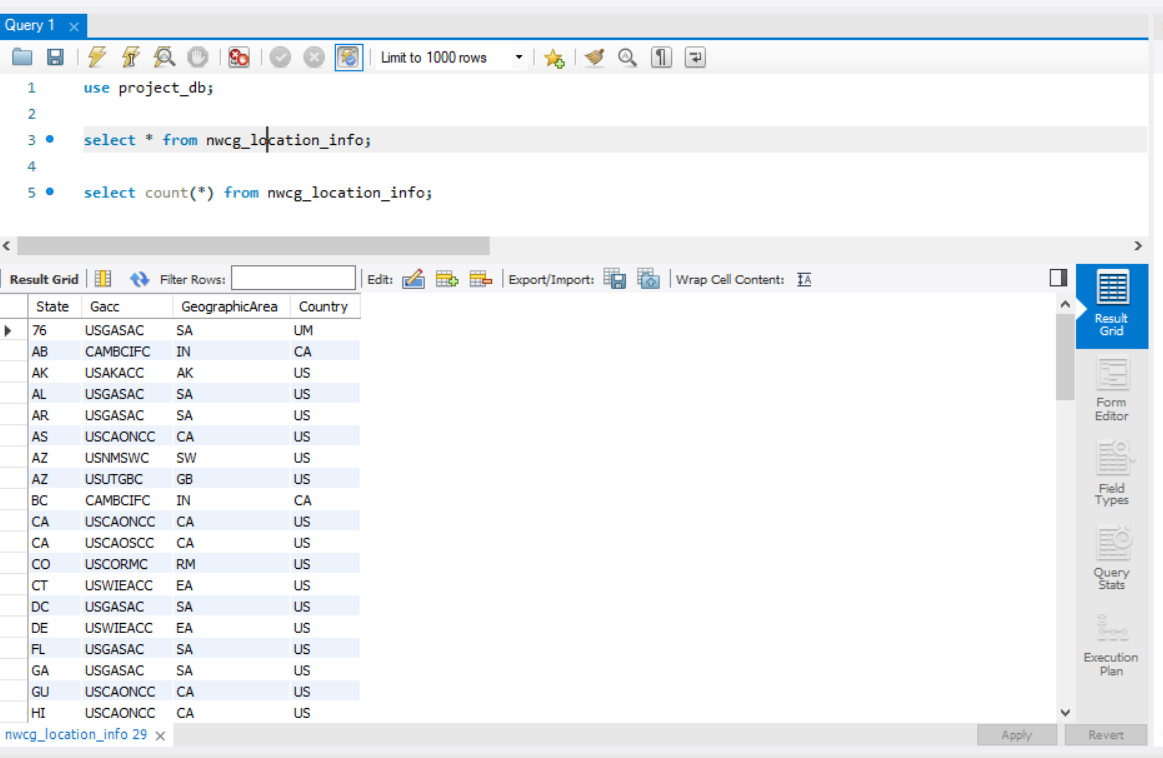
* fips\_info

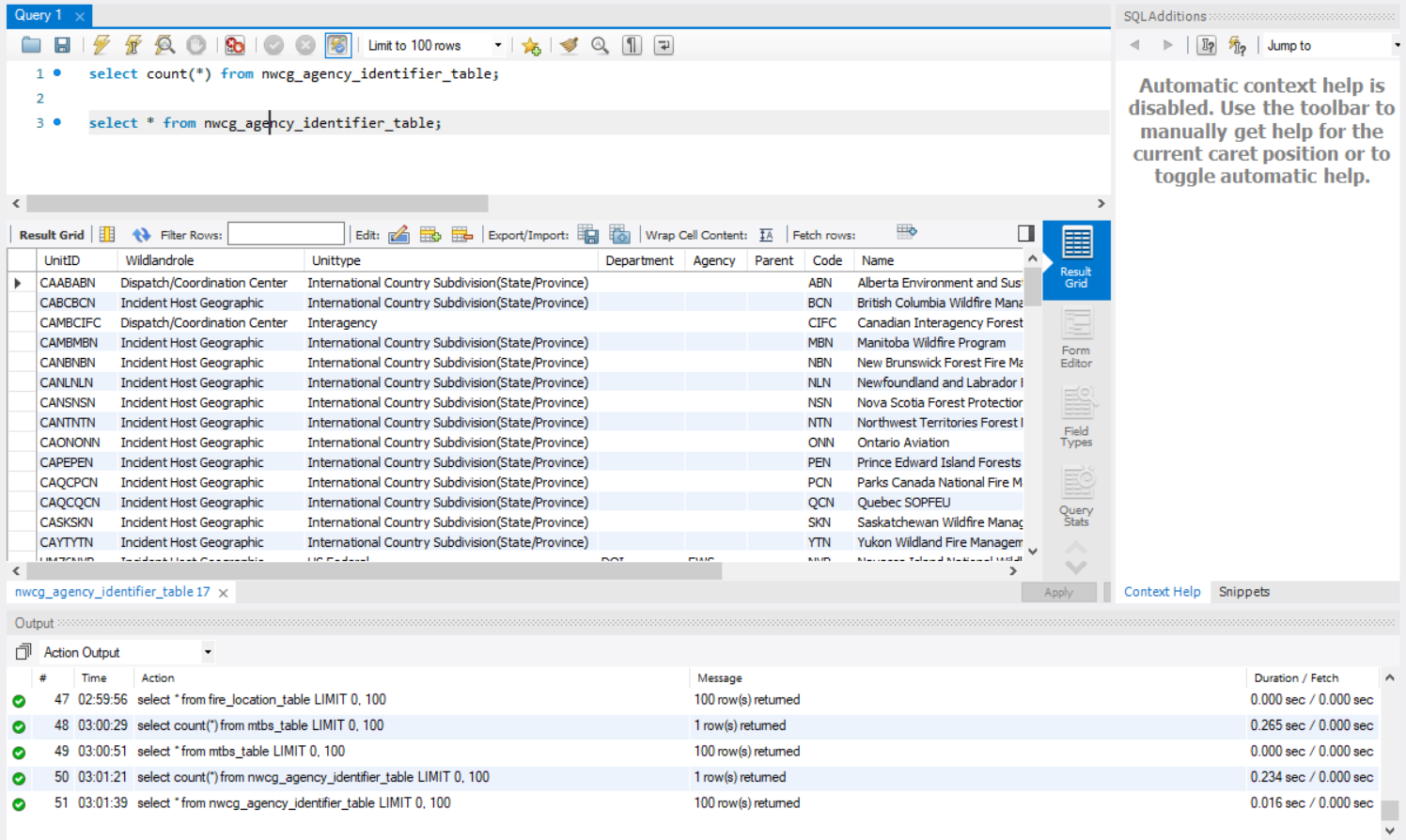
* location\_info

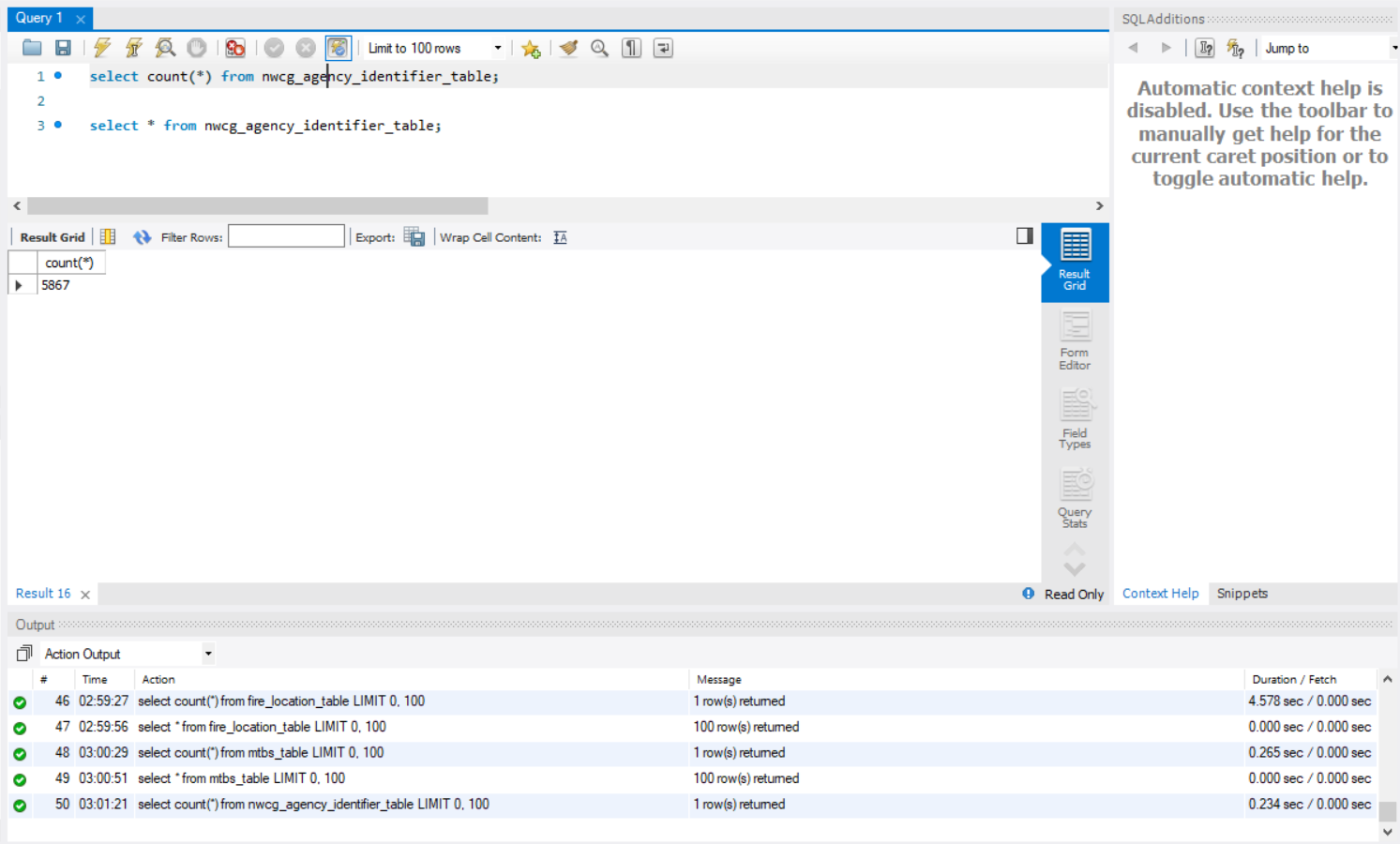
 

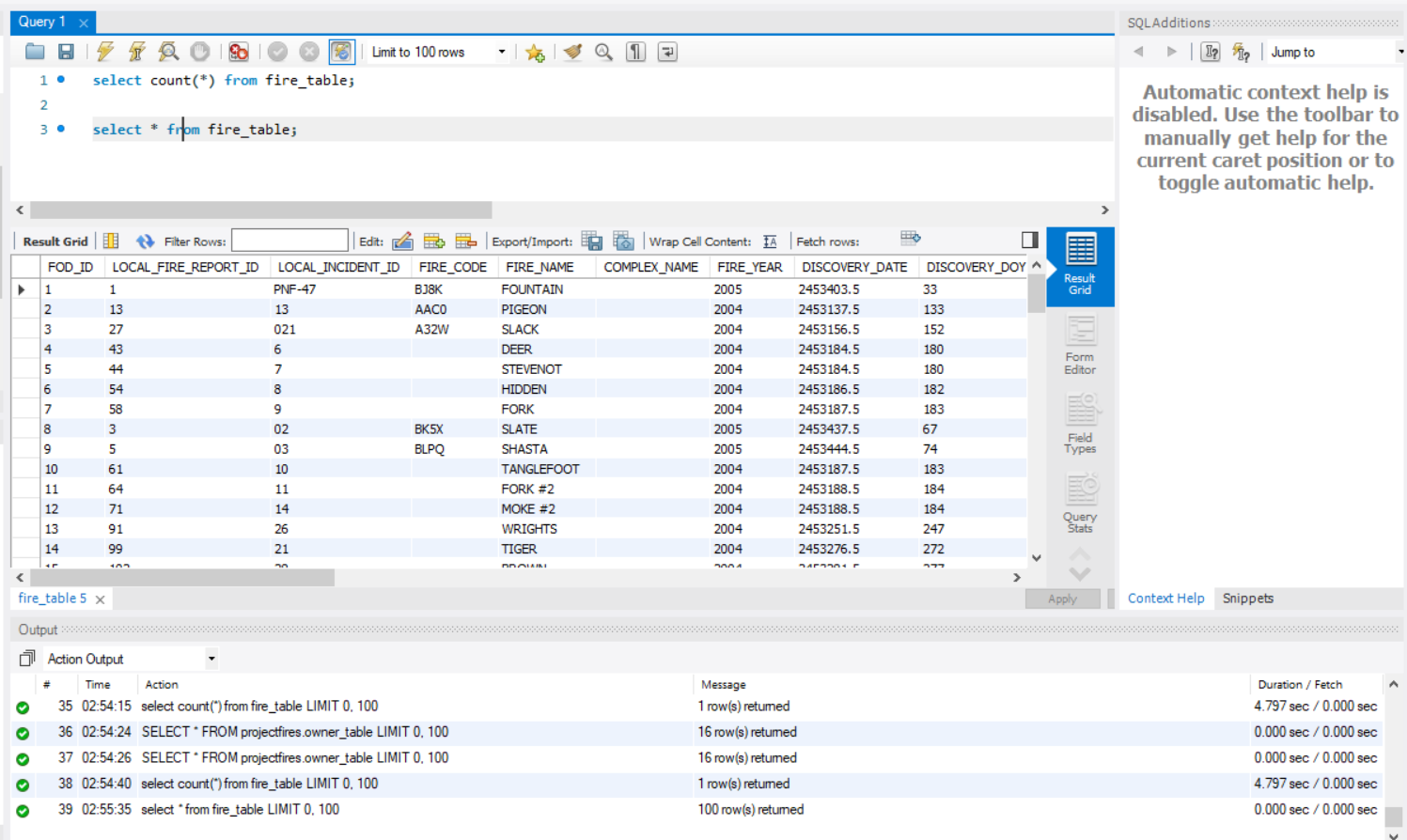
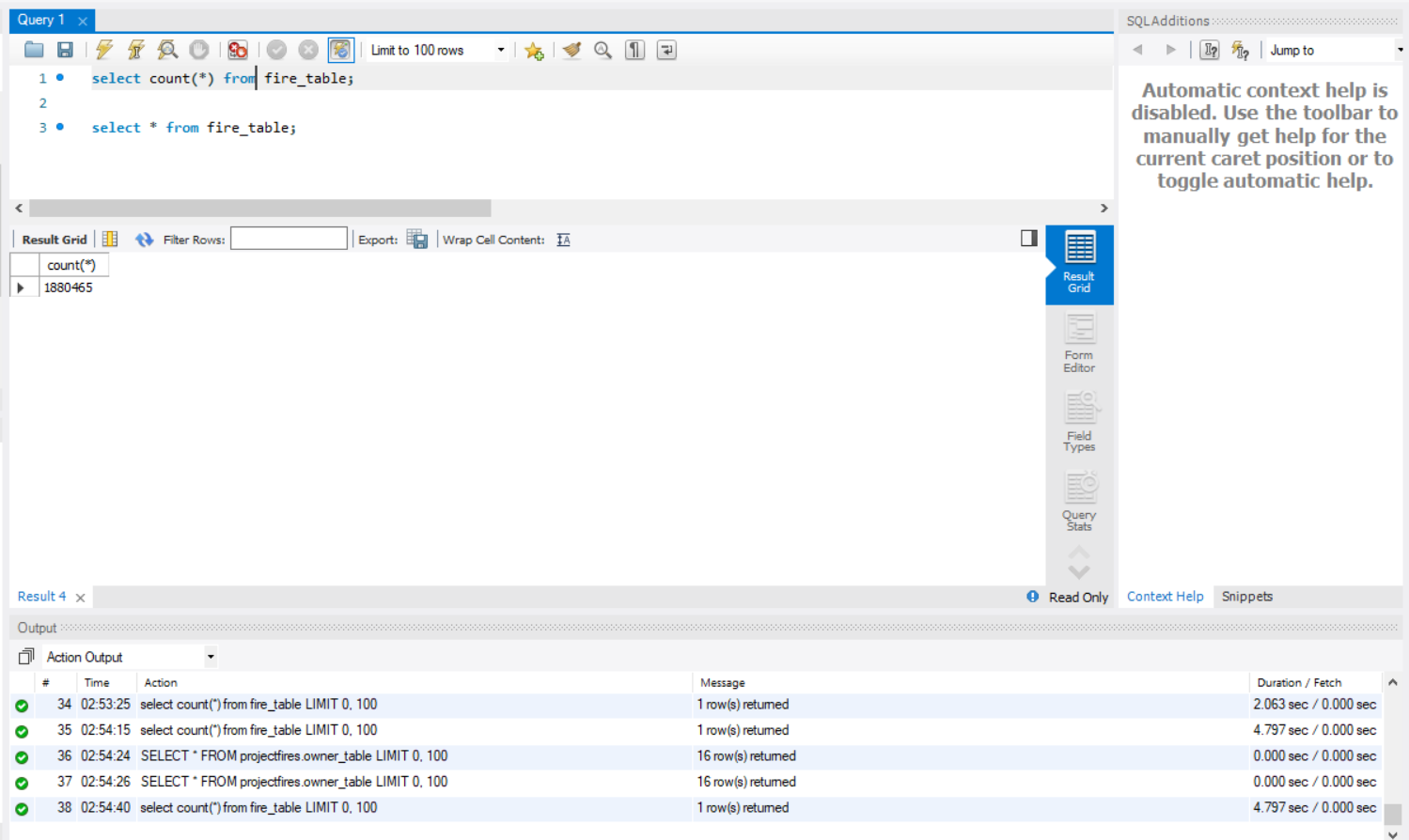
* nwcg\_location\_info



* nwcg\_agency\_identifier\_table





* fire\_table  

## Data in the Database

|  |  |  |  |
| --- | --- | --- | --- |
| **Table Name** | **Primary Key** | **Foreign Key** | **# of Rows in Table** |
| stat\_cause\_info | Stat\_Cause\_Code |  | 13 |
| owner\_info | Owner\_Code |  | 16 |
| ics\_209\_info | Ics\_209\_Incident\_Index |  | 22734 |
| mtbs\_info | Mtbs\_ID |  | 10481 |
| source\_reporting\_info | Source\_Reporting\_Unit |  | 4991 |
| source\_system\_info | FPA\_ID |  | 1880462 |
| location\_info | Latitude, Longitude | fips\_info\_Fips\_Code  fips\_info\_Fips\_Name | 1565730 |
| fips\_info | Fips\_Code, Fips\_Name |  | 2693 |
| nwcg\_location\_info | State,Gacc |  | 84 |
| nwcg\_agency\_identifier\_table | UnitID, Agency | nwcg\_location\_info\_State  nwcg\_location\_info\_Gacc | 5867 |
| fires\_table | FOD\_ID, Fire\_Size | ics\_209\_info\_Ics\_209\_Incident\_Number  source\_system\_info\_FPA\_ID  nwcg\_agency\_identifier\_table\_UnitID  nwcg\_agency\_identifier\_table\_Agency  source\_reporting\_info\_Source\_Reporting\_Unit  owner\_info\_Owner\_Code  mtbs\_info\_Mtbs\_ID  stat\_cause\_info\_Stat\_Cause\_Code  location\_info\_Latitude  location\_info\_Longitude | 1880465 |

## SQL Queries

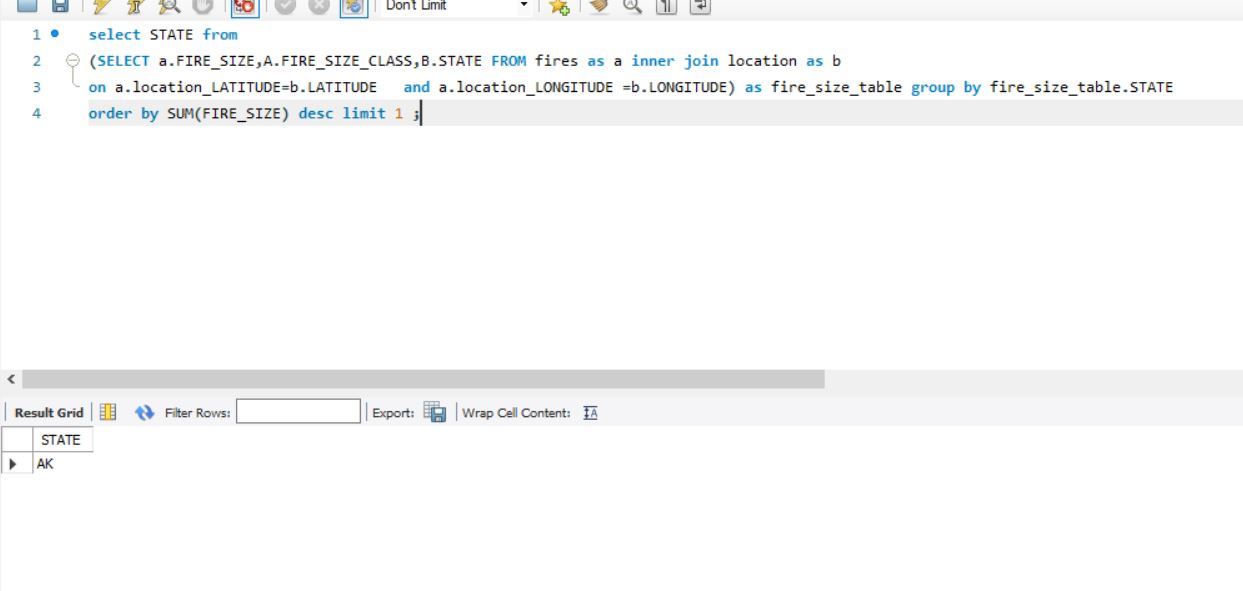
## Query 1

### A leading beverage company has announced a billion-dollar fund for removing debris from forests, rivers and mountains in the US. All states are interested. Which state has the best chance to win a share of the fund?

### As all the states are interested in the billion-dollar fund we group by state, but the state with which has highest number of fires will receive the fund. AK has the highest fires.

### Translation : Select states from location by joining it with fires on latitude and longitude and group by state in descending order.

### Screen Shot of SQL Query and Results:



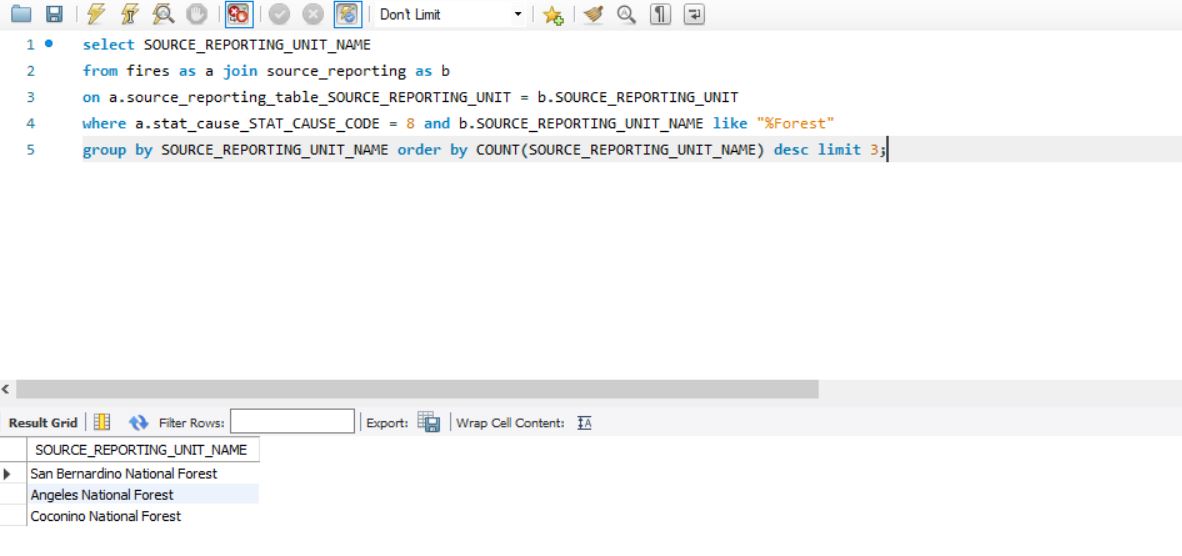
## Query 2

### One of the reporting agencies has suggested that children be banned from its forests unless there is one adult for every 3 children in a group visiting a forest. Name 3 forests where this would be the most appropriate.

### As the reporting agencies has suggested every 3 children should be accompanied by one adult, we group by source reporting unit in descending order and selecting first three forests will give us 3 rows.

### Translation: select source \_reporting\_unit\_name from source\_reporting table by joining on source reporting unit where stat\_cause\_code is 8 and source\_reporting\_unit name is like forest and grouping them by source \_reporting\_unit\_name in descending order.

### Screen Shot of SQL Query and Results:



## Query 3

### One advocacy group says human actions and not Nature is to blame for most wildfires. Write a query that supports this statement.

### As the statement states human actions are to be blamed for most wildfires, we need the count the number of fires that are caused by human and nature. If the count of fires caused by human are more this proves the above statement is true. 2 rows are retrieved.

### Translation: select cause and count of fires from fires table where the stat\_cause\_descr is lightning, nature and human and group by cause.

### Screen Shot of SQL Query and Results:



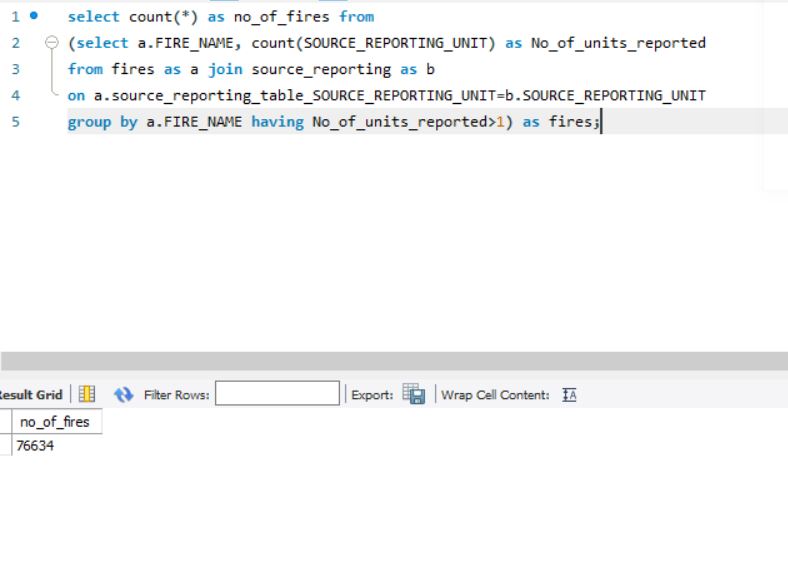
## Query 4

### How many wildfires were reported by more than one unit/agency?

### Retrieve the number of wildfires reported in more than one unit/agency. One row is retrieved.

### Translation : select count of source\_reporting\_name from fires and source\_reporting table by them on source \_reporting\_unit group by fire\_name and having no of units greater than 1.

### Screen Shot of SQL Query and Results:



## Query 5

### What were the forests that had more than one fire that lasted more than two days?

### Retrieve the forests where the fires lasted for more than 2 days. 128 rows were retrieved.

### Translation: select source reporting unit name from source\_reporting table by joining it with fires on source reporting unit where the difference between controlled date and discovery date is greater than 2 and source reporting unit name like forest and grouping by source reporting unit having count of source reporting unit name is greater than 1.

### Screen Shot of SQL Query and Results:

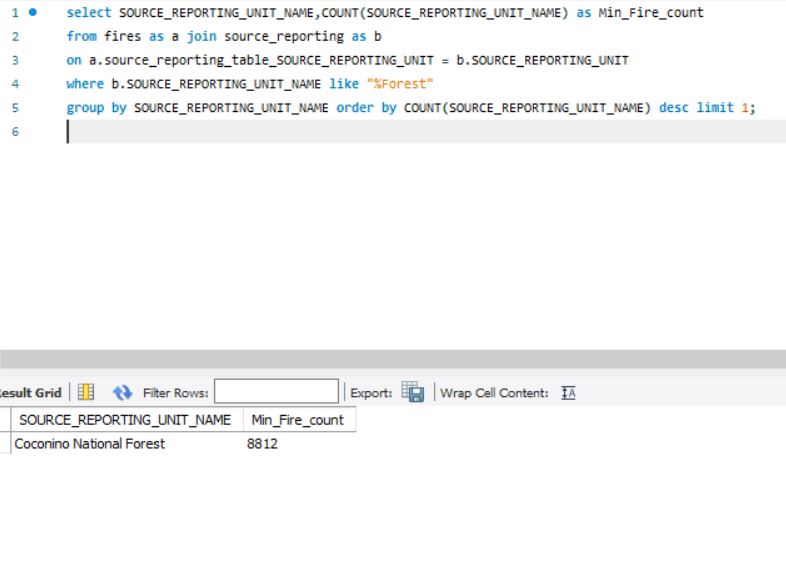
## Query 6

### Which forest had the most number of fires?

### Report the forest which has most number of fires. One row is retrieved

### Translation: select source\_reporting\_unit\_name from source\_reporting\_unit table by joining it with fires table on soure\_reporting\_unit where source\_reporting\_unit\_name like forest and grouping it by source\_reporting\_unit\_name in descending order.

### Screen Shot of SQL Query and Results:



# Data Review for MongoDB

## Assumptions/Notes About Data Collections, Attributes and Relationships between Collections

The wildfires database in MongoDB contains 2 collections:

1)Fires

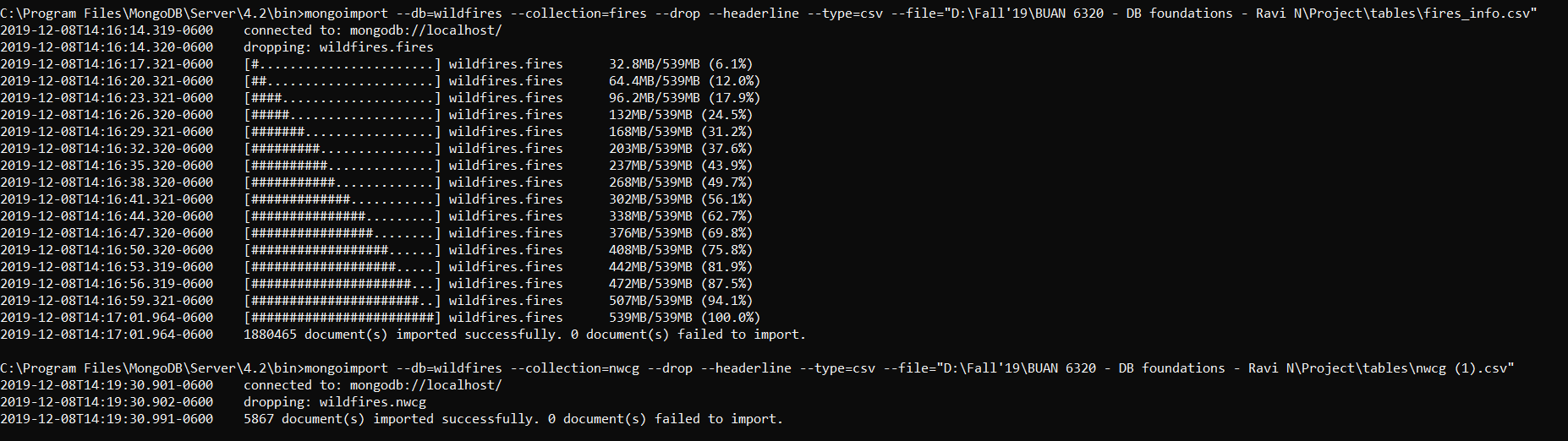
2) NWCG

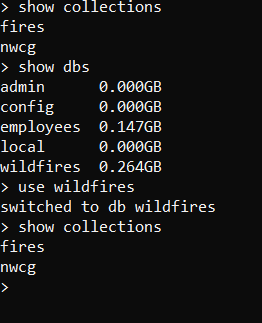
Relationship: NWCG\_REPORTING\_UNIT\_ID  field in Fires collection is linked with UnitId field in NWCG Collection

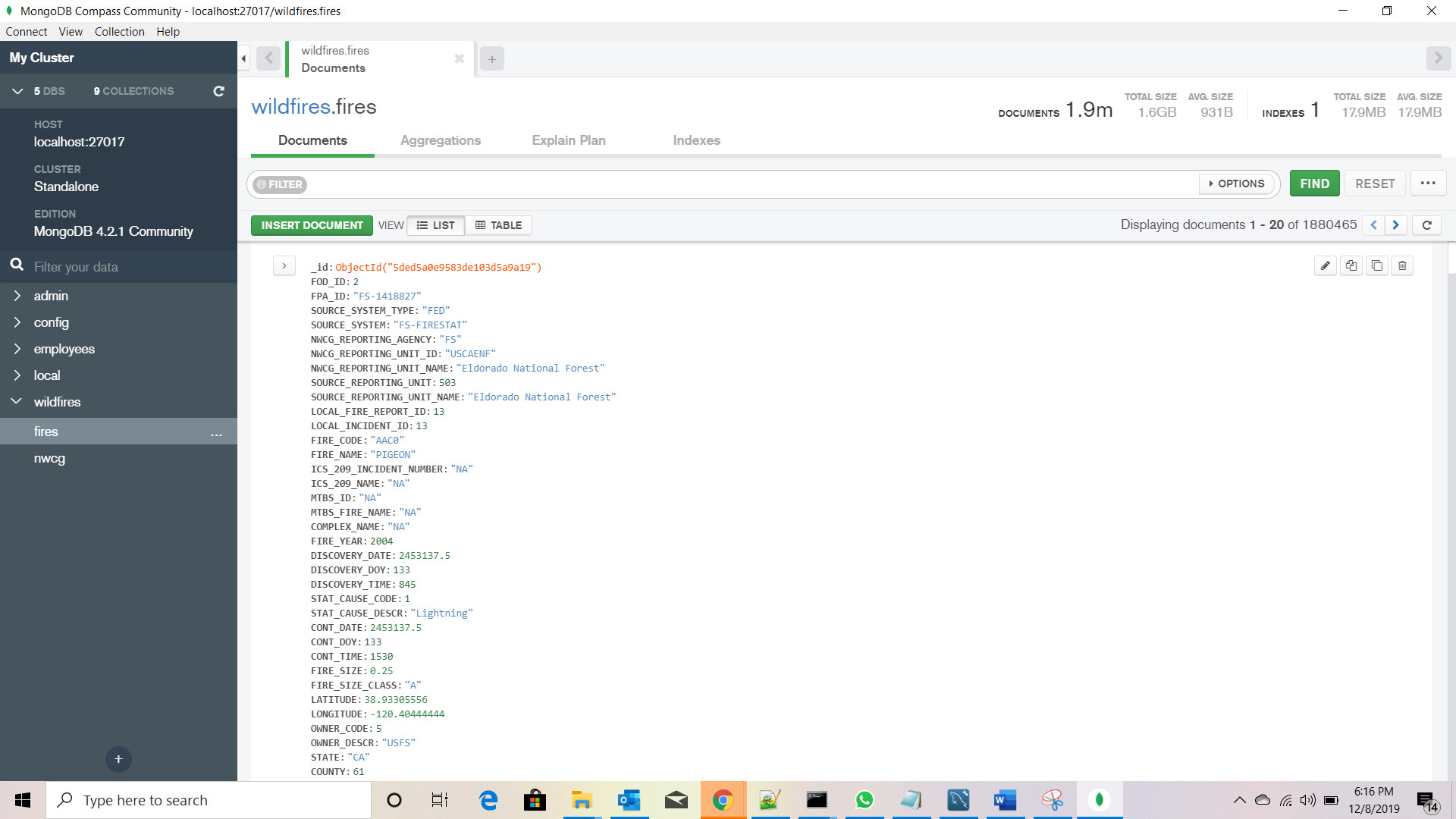
# Physical Mongo Database

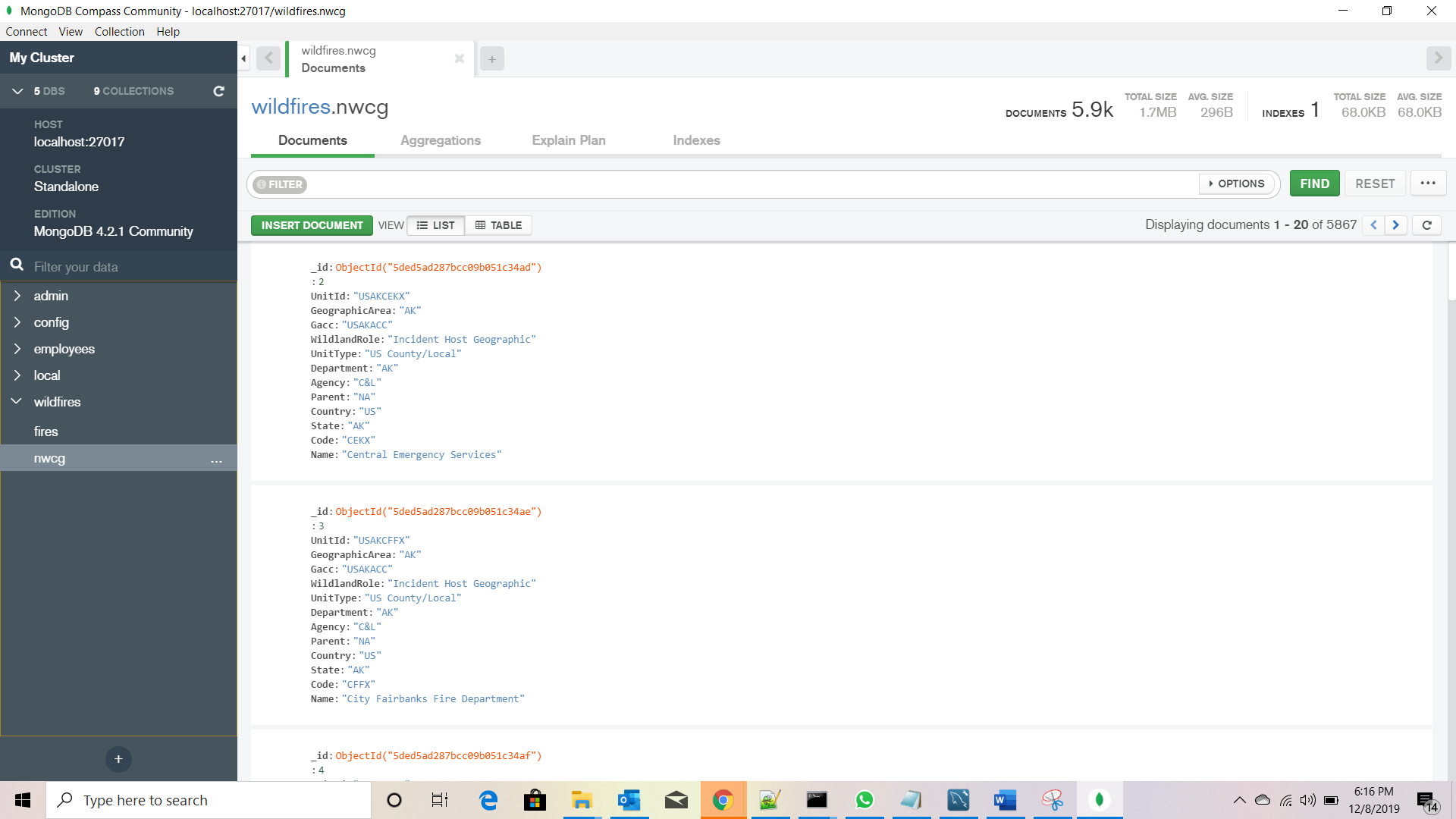
## Assumptions/Notes About Data Set

## Screen shot of Physical Database objects (Database, Collections and Attributes)









## Data in the Database

|  |  |  |
| --- | --- | --- |
| **Collection Name** | **Relationshps With Other Collections (if any)** | **# of Documents in Collection** |
| Fires | NWCG\_REPORTING\_UNIT\_ID  field in Fires is linked with UnitId field in NWCG Collection | 1880465 |
| NWCG | NWCG\_REPORTING\_UNIT\_ID  field in Fires is linked with UnitId field in NWCG Collection | 5867 |

# MongoDB Queries/Code

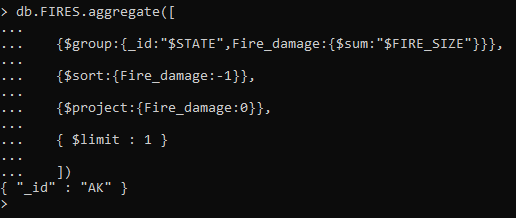
## Query 1

### A leading beverage company has announced a billion-dollar fund for removing debris from forests, rivers and mountains in the US. All states are interested. Which state has the best chance to win a share of the fund?

### As all the states are interested in the billion-dollar fund we group by state, but the state with which has highest number of fires will receive the fund. AK has the highest fires.

### Translation: Project the highest fire\_damage grouping by state in descending order.

### Screen Shot of MongoDB Query/Code and Results:



## Query 2

### One of the reporting agencies has suggested that children be banned from its forests unless there is one adult for every 3 children in a group visiting a forest. Name 3 forests where this would be the most appropriate.

### As the reporting agencies has suggested every 3 children should be accompanied by one adult, we group by source reporting unit in descending order and selecting first three forests will give us 3 documents.

### Translation: Report the source\_reporting\_unit\_name where the source\_reporting\_unit\_name contains forest and stat\_cause\_code is 8 by sorting them in descending order.

### Screen Shot of MongoDB Query/Code and Results:



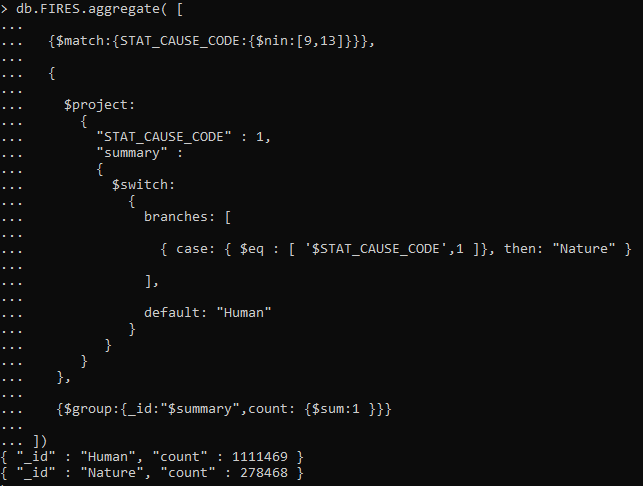
## Query 3

### One advocacy group says human actions and not Nature is to blame for most wildfires. Write a query that supports this statement.

### As the statement states human actions are to be blamed for most wildfires, we need the count the number of fires that are caused by human and nature. If the count of fires caused by human are more this proves the above statement is true. 2 documents are retrieved.

### Translation: Projecting the summary which contains human and nature. If the stat\_cause\_code is 1 then its nature else human based on which the count of each stat\_cause\_code is projected.

### Screen Shot of MongoDB Query/Code and Results:



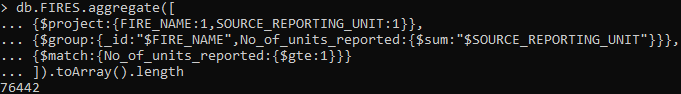
## Query 4

### How many wildfires were reported by more than one unit/agency?

### Retrieve the number of wildfires reported in more than one unit/agency. One document is retrieved.

### Translation: Projecting the fire\_name and the source\_reporting\_unit where the no\_of\_units\_reported is greater than 1.

### Screen Shot of MongoDB Query/Code and Results:



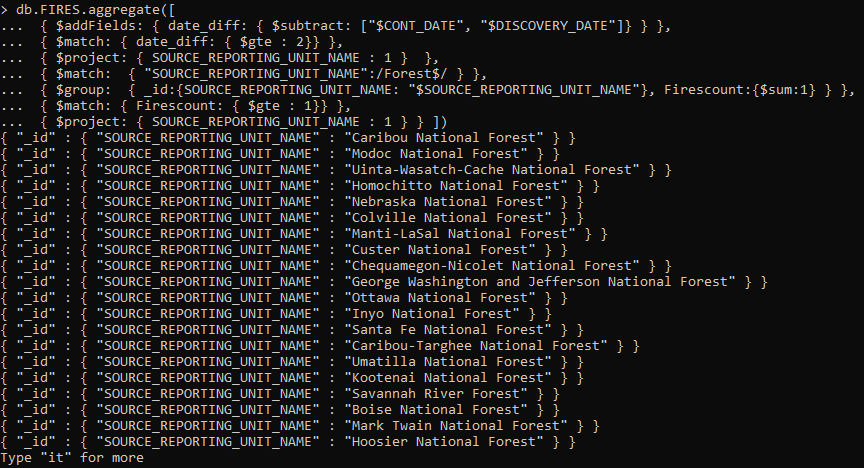
## Query 5

### What were the forests that had more than one fire that lasted more than two days?

### Retrieve the forests where the fires lasted for more than 2 days. 128 documents were retrieved.

### Translation: Projecting the source\_unit\_reporting\_name where the difference between controlled date and discovery date is greater than 2 and source\_unit\_reporting\_name contains “forest“ and firecount is greater than 1.

### Screen Shot of MongoDB Query/Code and Results:



## Query 6

### Which forest had the most number of fires?

### Report the forest which has most number of fires. One document is retrieved

### Translation: Project nwcg\_reporting\_unit\_name which contains “ forest“ by sorting in descending order.

### Screen Shot of MongoDB Query/Code and Results:

