**SQL-Mongo Project –**

**Spatial Data of US Wildfires**

BUAN 6320

Group #: 3

Anu Jose

Shylaja Vijayaraghavan

Geetika Yalamanchili

Anne Emmanuel

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Activity** | **Member 1** | **Member 2** | **Member 3** | **Member 4** |
| Prepared Data Model and Created Physical DB | x | x | x | x |
| Loaded Data into Database | x | x | x | x |
| Wrote SQL Queries | x | x | x | x |
| Prepared Mongo Database | x | x | x | x |
| Loaded data into Mongo DB | x | x | x | x |
| Wrote Mongo Queries | x | x | x | x |
| Prepared Report | x | x | x | x |
| Reviewed Report | x | x | x | x |

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**

No assumptions are made while making the model. All the records on the data set are kept intact. Some data has been mis-recorded which has affected the normalization.

Following fields were excluded since they presented Functional dependency in the tables:

* Cont\_DOY (Calculated Field – Can be obtained from Contained Date)
* Discovery\_DOY (Calculated Field – Can be obtained from Discovery Date)
* Fire\_Year (Calculated Field – Can be obtained from Discovery Date)

## **Notes About Data Entities and Relationships**

Following are the Table descriptions and relationships:

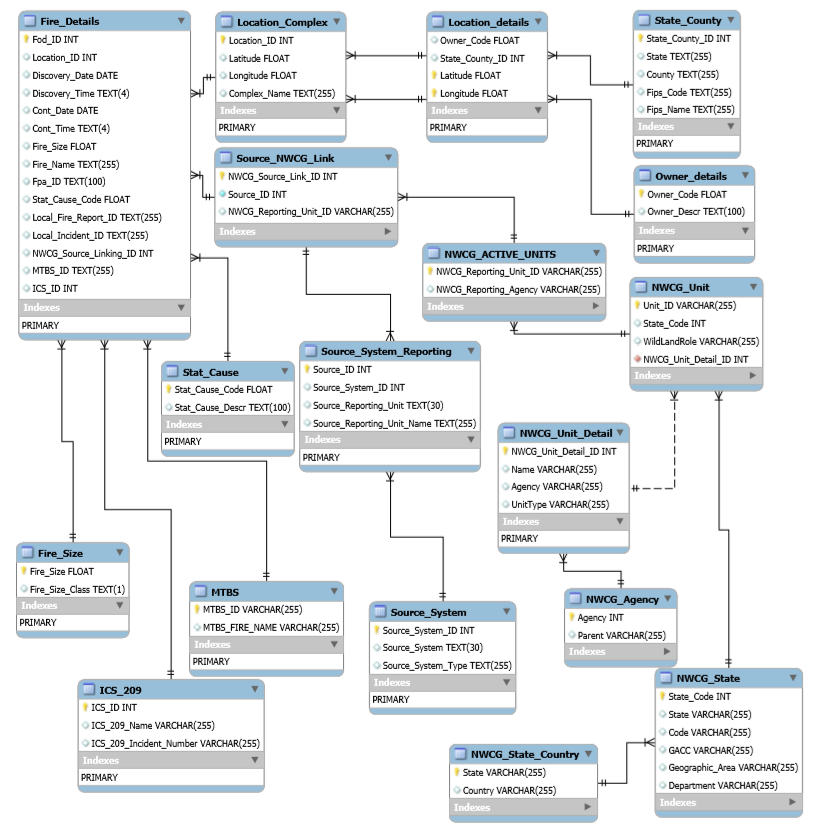
1. **Fires –** This is the main data table that includes the fire details. The Fire location, Fire Discovery details, Fire Contained details, Fire Size details, Fire Name, Statistical Cause of the fire, Local incident and report details, NWCG\_Source details, ICS\_209 and MTBS report details can be joined to this table to get all information about the fire
2. **Owner –** Describes the owner details of the location where fire occurred
3. **Location –** Describes the details of the location of fire including the latitude, longitude. State, Country and Owner details can be joined from the respective linked tables.
4. **State\_Country –** The details of the state and country where the fire occurred can be obtained from this table
5. **Location\_Complex –** The details of the location of fire including the complex name can be obtained here
6. **Fire\_Size –** The size of the fire along with the classification of the fire size is included
7. **Stat\_Cause –** The statistical cause of the fire is described in this table
8. **Source\_System –** Includes the Source system from which the fire record is drawn and also identifies the type of source system (Fed, Non Fed and Interagency)
9. **Source\_System\_Reporting –** Details of the Source Reporting unit
10. **Source\_NWCG\_Link –** Linking table of the Source database of fire with NWCG reporting database of fire
11. **ICS\_209 –** Details of the incident from ICS\_209 report
12. **MTBS –** Details of the incident from the MTBS perimeter dataset
13. **NWCG\_Active\_Units –** Details of the active NWCG units present in the database
14. **NWCG\_Unit –** Details of the NWCG unit including the WildLand role of the unit in the fire community along with the details of the unit mapped from the NWCG\_Unit\_Detail table
15. **NWCG\_Unit\_Detail –** Describes the name, agency and unit type of the NWCG\_Unit
16. **NWCG\_Agency –** Agency details and the Agency Sub group details are included in the table
17. **NWCG\_State\_Country –** State, Geographic are, GACC and department details are included
18. **NWCG\_State –** NWCG unit location – State and country included in the table

## **Reasons for the Model to be in 3NF**

The model presented in the 3NF form. It is said to be in 3NF since it satisfies the following conditions:

* Each table contains only a single value
* Each record is unique
* Primary key is defined for each table that is Unique and Not Null
* No transitive and functional dependencies

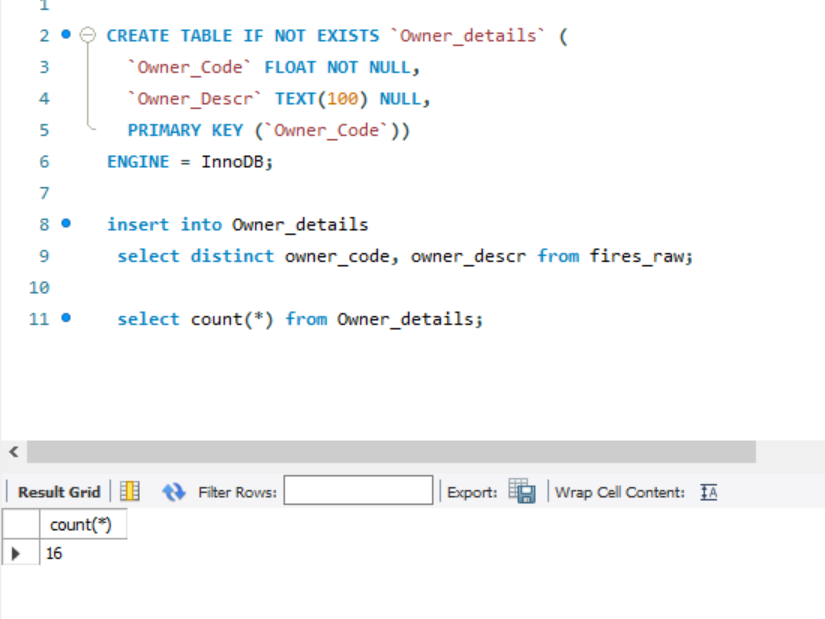
## **Entity-Relationship Diagram**



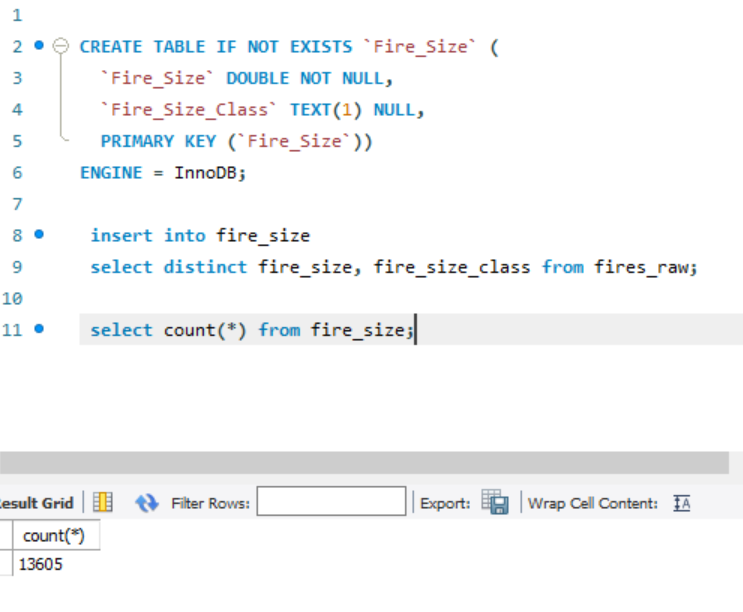
# **Physical Database**

## **Screen shot of Physical Database objects**

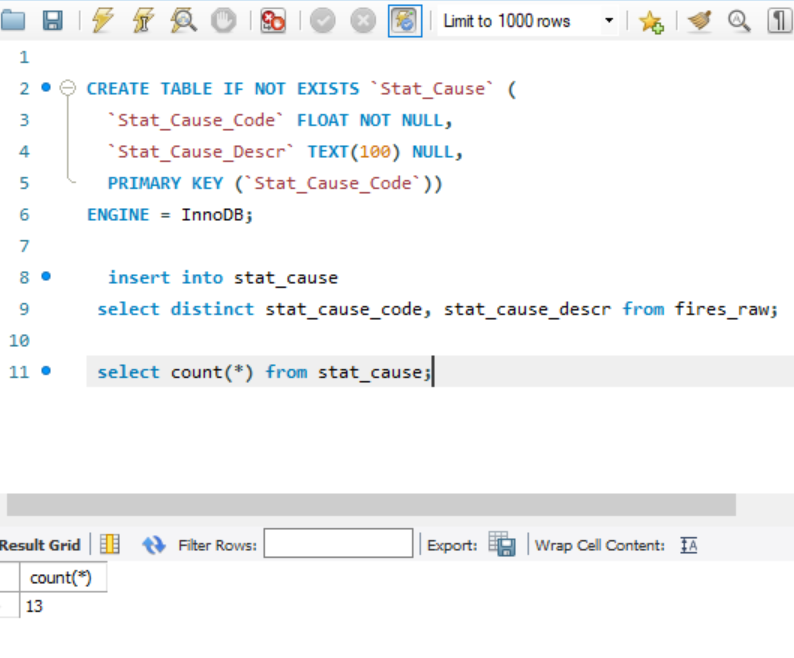
1. **Owner**



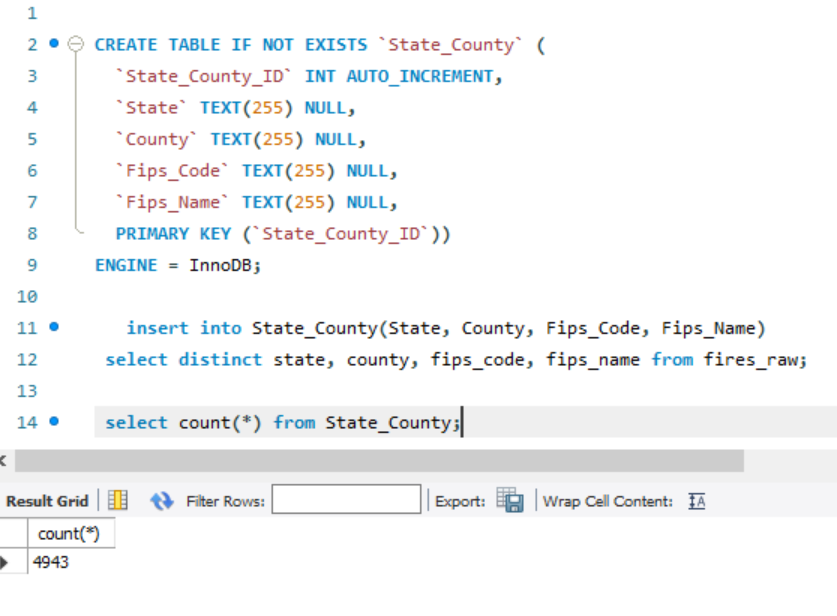
1. **Fire\_Size**



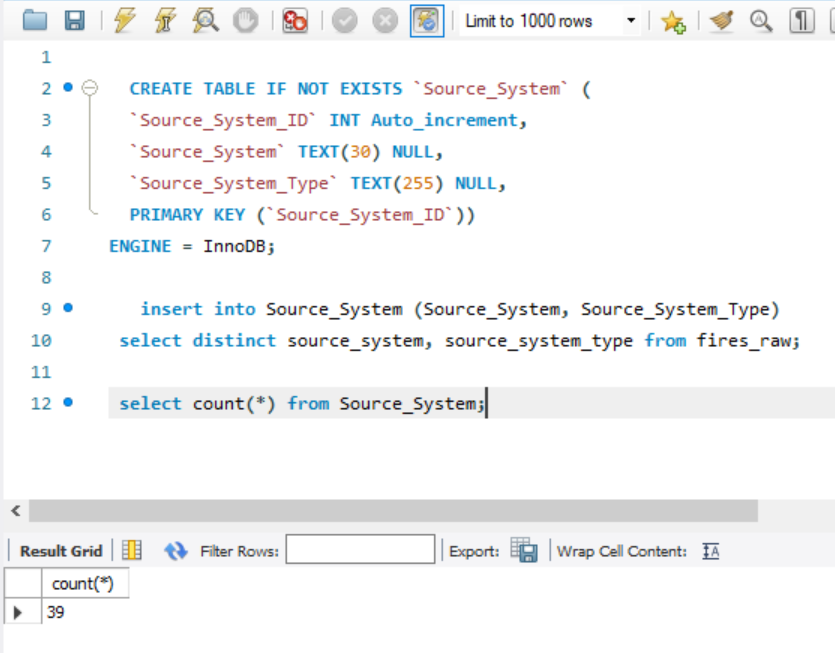
1. **Stat\_Cause**



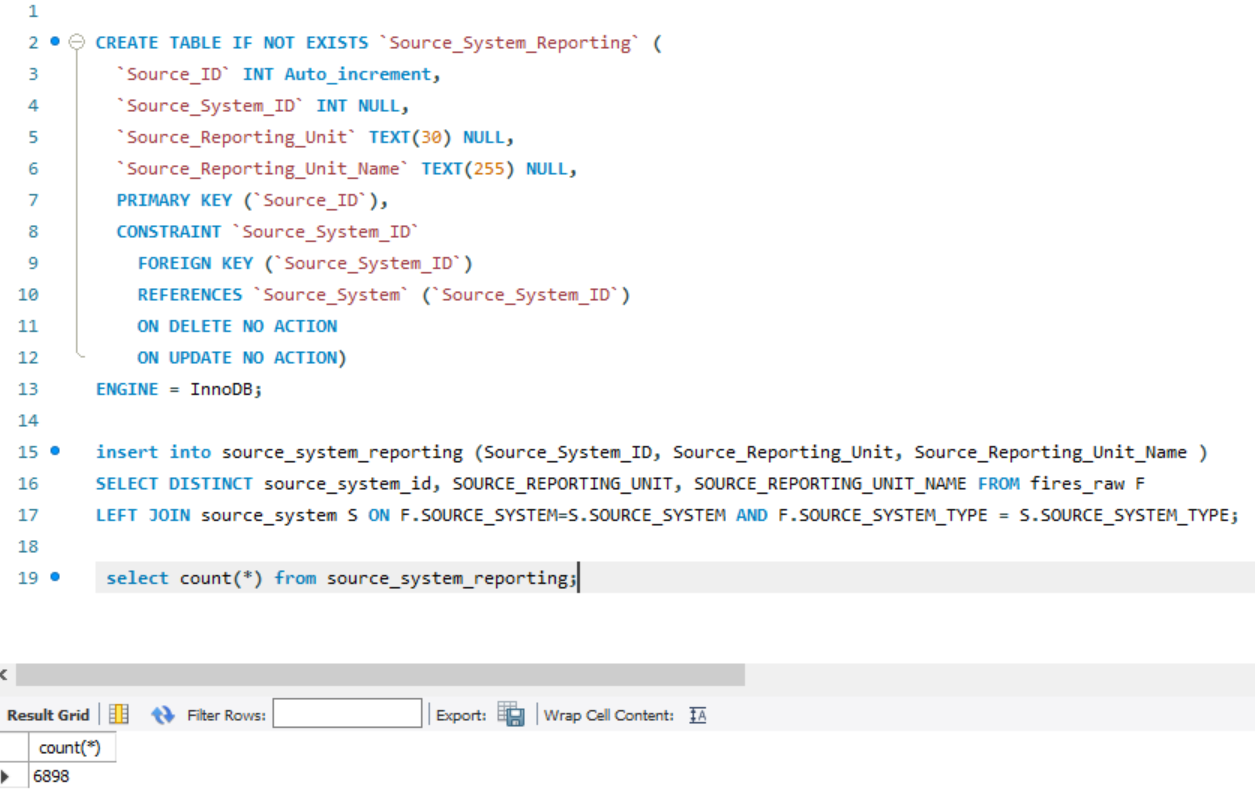
1. **State\_County**



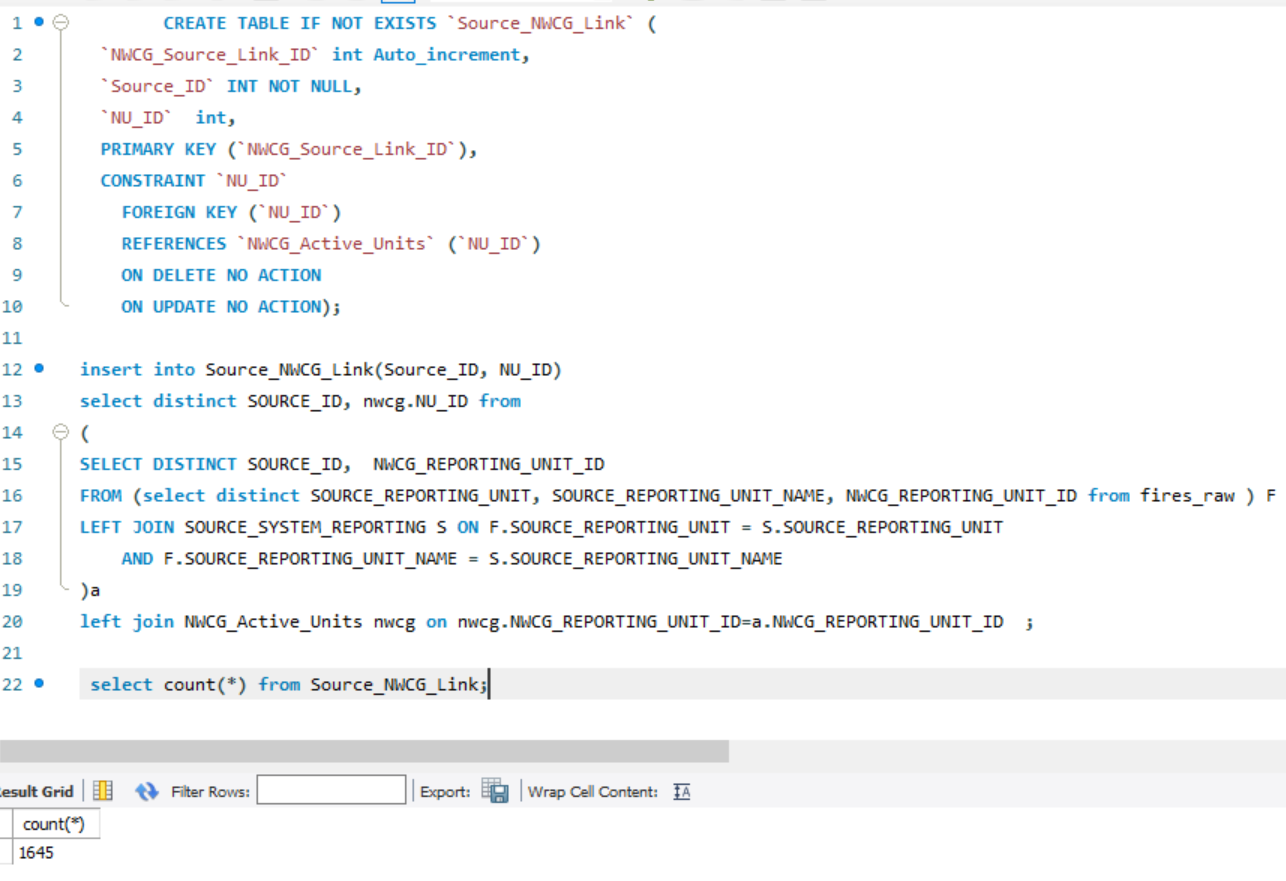
1. **Source\_System**



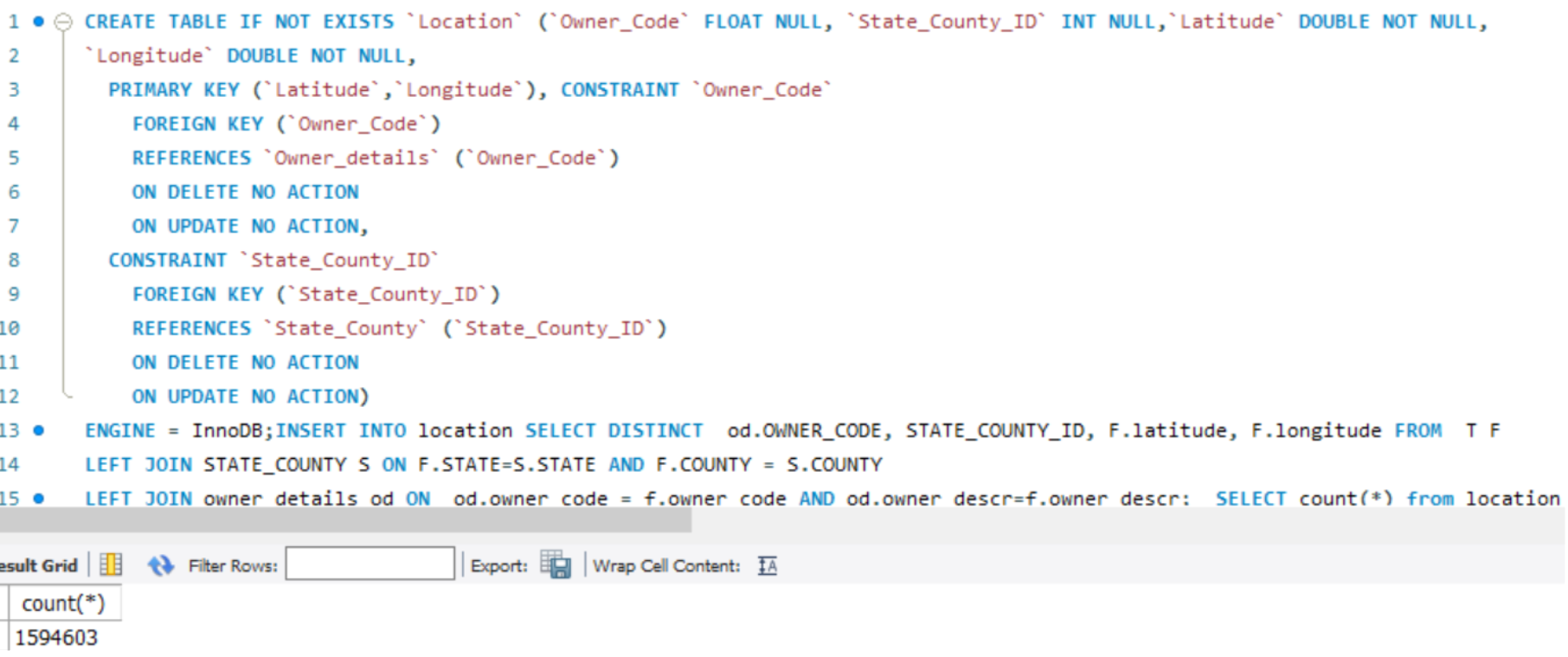
1. **Source\_System\_Reporting**



1. **Source\_NWCG\_Link**



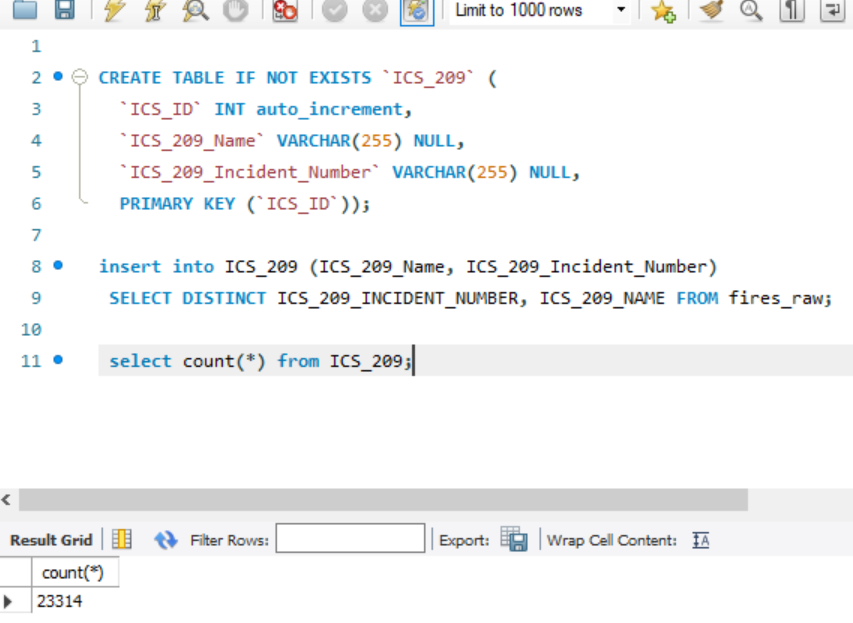
1. **Location**



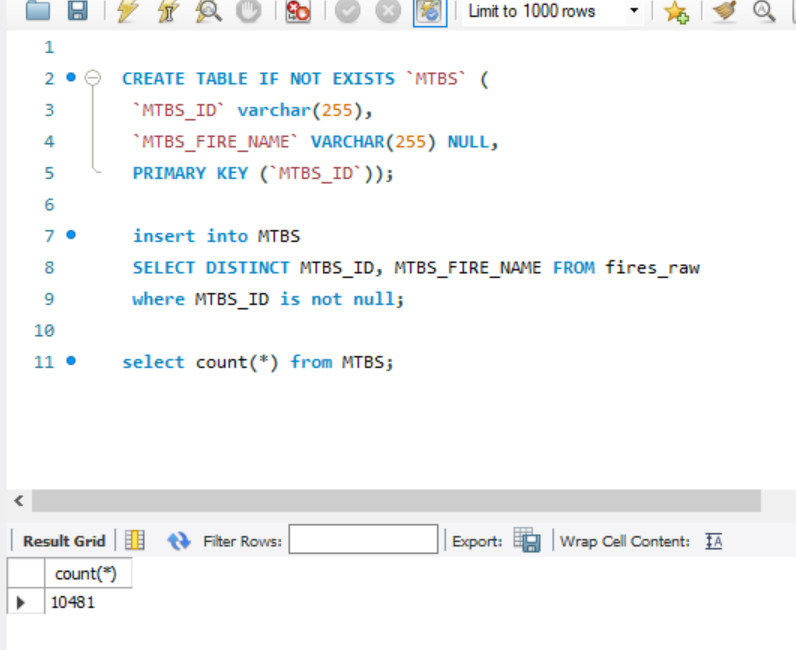
1. **Location\_Complex**



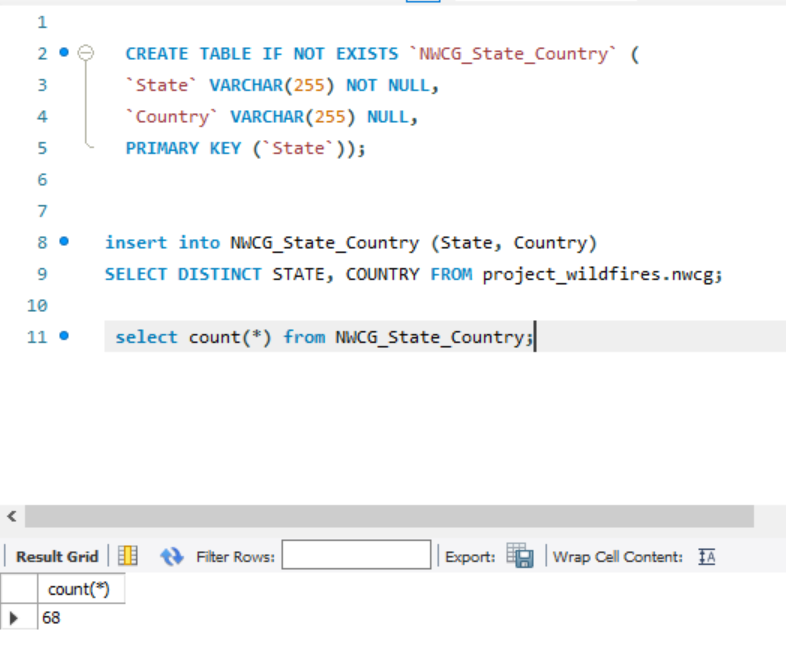
1. **ICS\_209**



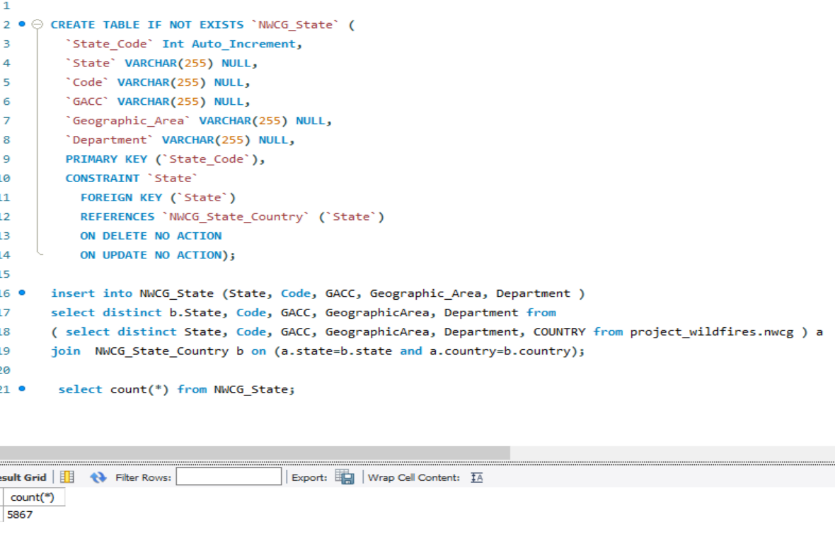
1. **MTBS**



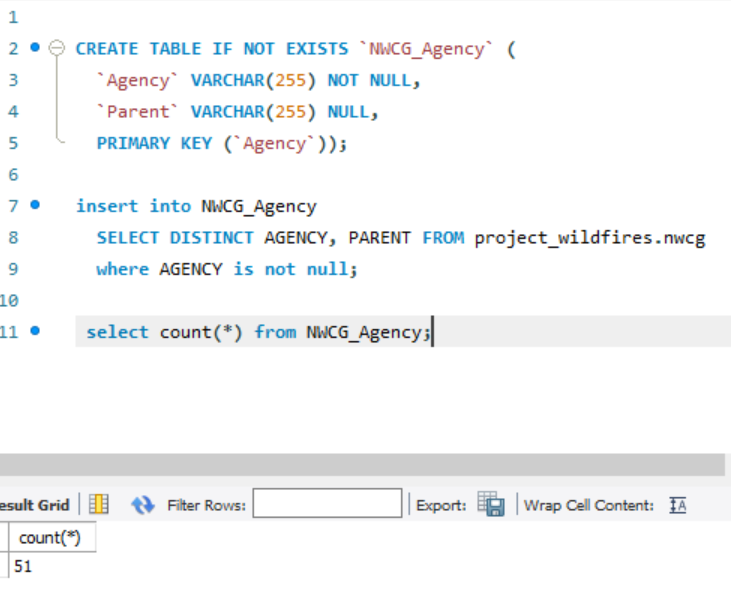
1. **NWCG\_State\_Country**



1. **NWCG\_State**



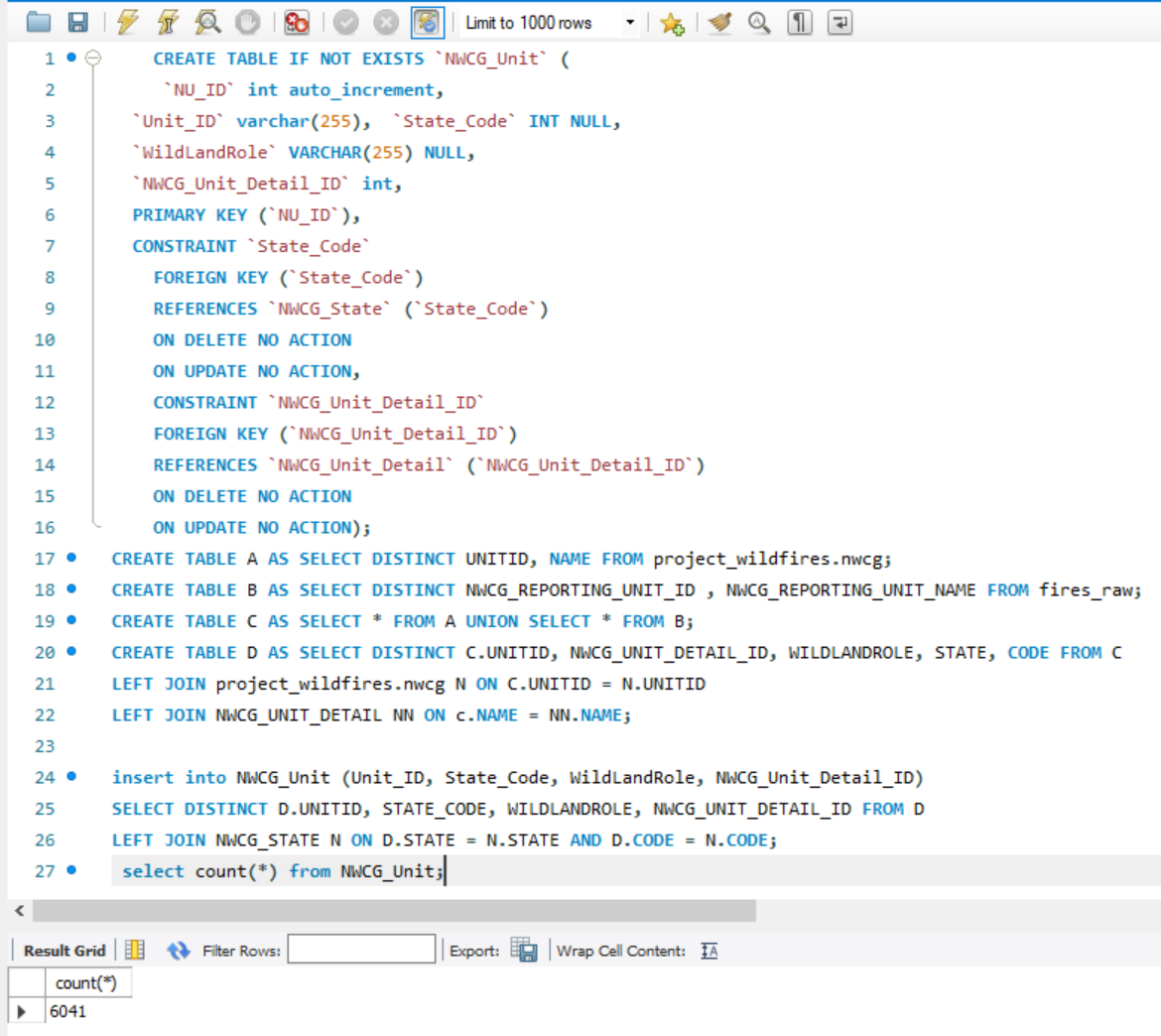
1. **NWCG\_Agency**



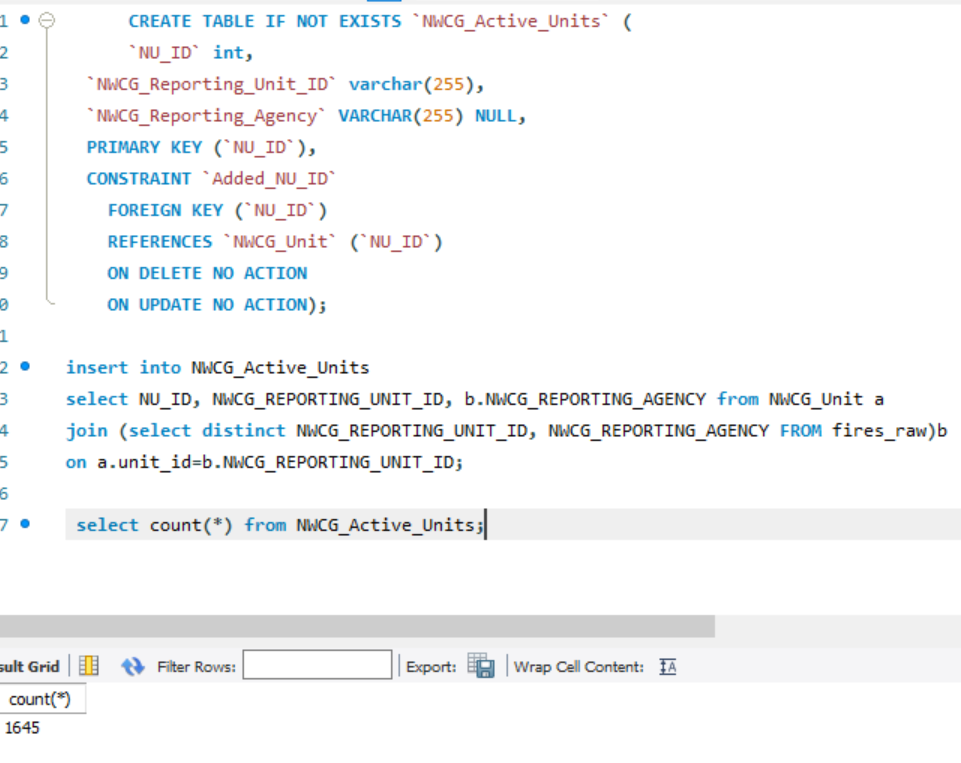
1. **NWCG\_Unit\_Detail**



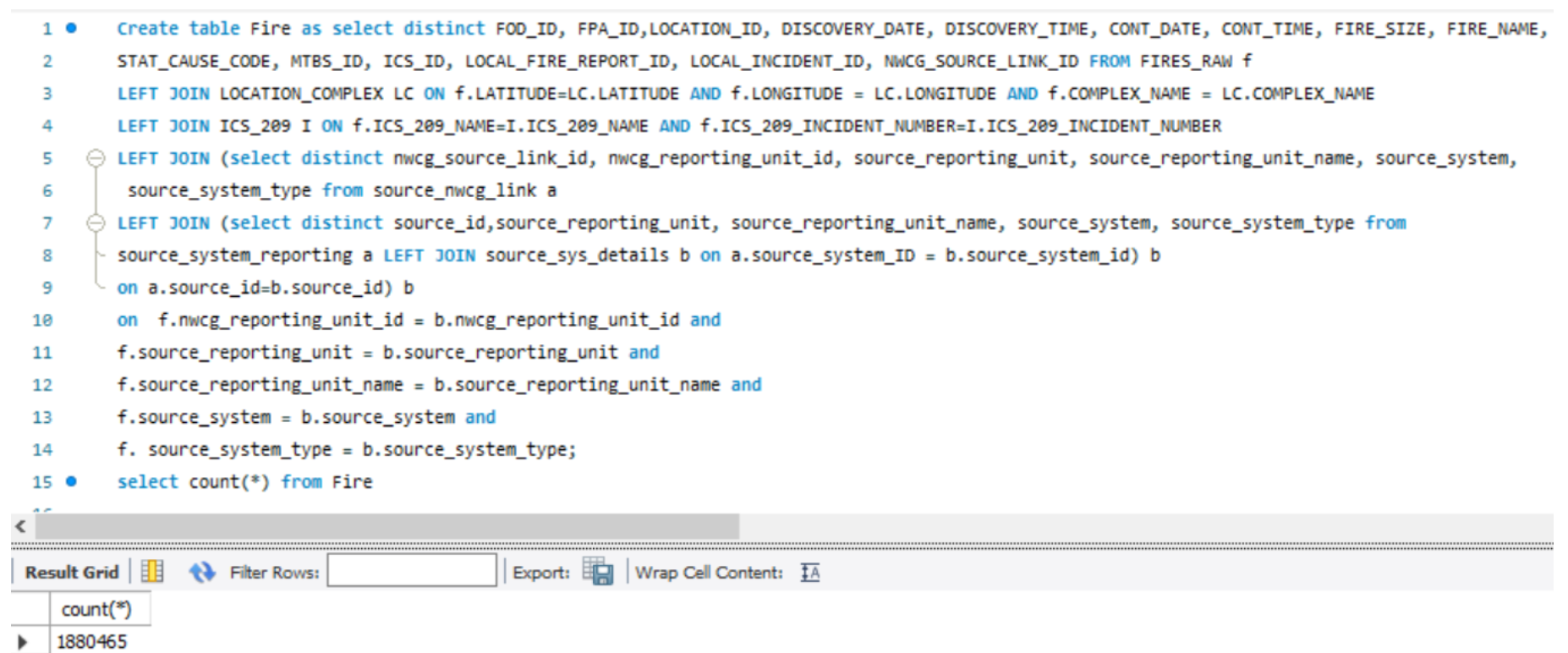
**16. NWCG\_Unit**



1. **NWCG\_Active\_Unit**



1. **Fires**



**Data in Database**

|  |  |  |  |
| --- | --- | --- | --- |
| **Table Name** | **Primary Key** | **Foreign Key** | **# of Rows in Table(mysql)** |
| Fire Details | FOD\_ID | Location\_ID; Fire Size; Stat\_Cause\_Code; NWCG\_Source\_Linking\_ID; MTBS\_ID; ICS\_ID | 1880645 |
| Owner | Owner Code | NA | 16 |
| Location | Latitude; Longitude | Owner\_Code ; State\_County\_ID | 15,94,382 |
| State\_County | State\_County\_ID | NA | 4,943 |
| Location\_Complex | Location\_ID | Latitude & Longitude | 15,69,862 |
| Fire\_Size | Fire\_Size | NA | 13,605 |
| Stat\_Cause | Stat\_Cause\_ID | NA | 13 |
| Source\_System | Source\_System\_ID | NA | 39 |
| Source\_System\_Reporting | Source\_ID | Source\_System\_ID | 6646 |
| Source\_NWCG\_Link | NWCG\_Source\_Link\_ID | Source\_ID; NWCG\_Reporting\_Unit\_ID | 7135 |
| ICS\_209 | ICS\_ID | NA | 23,314 |
| MTBS | MTBS\_ID | NA | 10,482 |
| NWCG\_Unit | NWCG\_Unit\_ID | NWCG\_Unit\_Detail\_ID; State\_Code | 6,039 |
| NWCG\_Unit\_Detail | NWCG\_Unit\_Detail\_ID | Agency | 5,802 |
| NWCG\_State | State\_Code | State | 5,867 |
| NWCG\_Agency | Agency | NA | 51 |
| NWCG\_State\_Country | State | NA | 68 |
| NWCG\_Active\_Units | NWCG\_Reporting\_Unit\_ID |  | 1,640 |

# **SQL Queries**

## **Assumptions:**

1. “Source\_Reporting\_Unit\_Name” is considered as forests with the assumption that each forest has one and only one agency to report the fires occurring the respective forest
2. “FOD\_ID” is considered as the fires under the assumption that each fire is reported only by one reporting unit

## **Query 1**

### **Question 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?

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

The state with maximum number of fires is expected to have maximum debris and thereby have higher chances of receiving the funds for enabling them to ensure debris removal. The state with more debris must have more funds to clean up, hence better chance of the state to win the share. # of Rows in Result: 1

### **Translation**

-- select state which has max no. of fires from fire table left joined with location complex table on location id of both fire and location complex tables and

left join on latitude and longitude from location and location complex table and left join on state county id from statecounty and location tables that is

grouped by state.

### **Clean-Up**

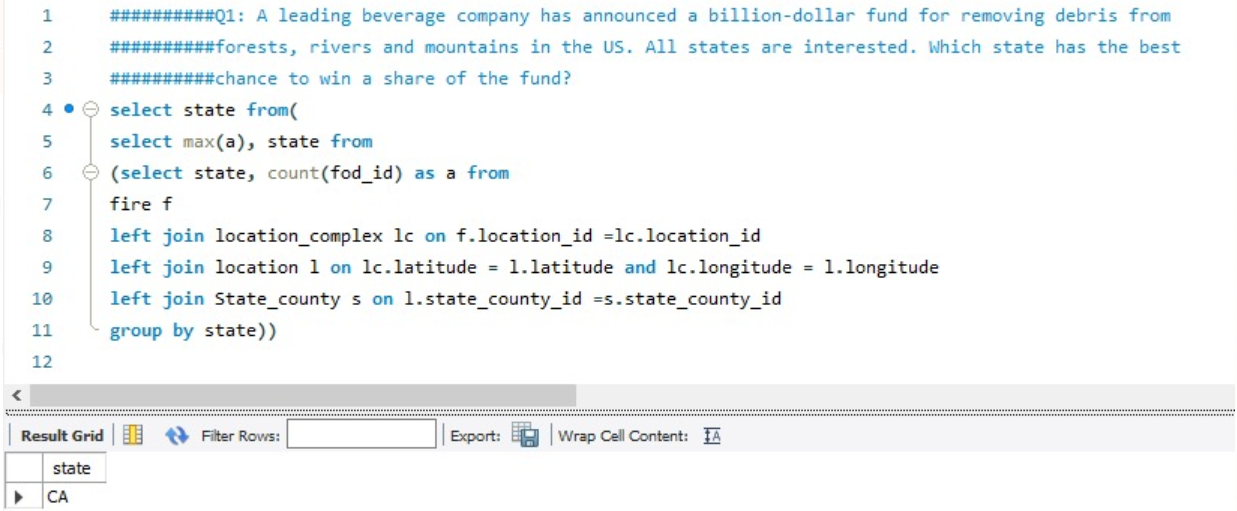
-- cleanup: select state from max no. of fires from fire left join location\_complex table on location id of fire and location complex,

left join on latitude and longitude from location and location complex and left join on state county id from state county and location, grouped by

state

### ***(Screenshot next page)***

### **Screen Shot of SQL Query and Results**



## **Query 2**

### **Question 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.

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

Georgia Forestry Commission, Red Lake Agency and New Jersey Forest Fire Service Division C are the three forests where the mentioned suggestion seems appropriate. These three forests have the maximum fires caused by the children. Since children have been the statistical cause for the fire, it is ideal to implement regulations on the visit of children into the forests under the supervision of the adults.

# of Rows in Result: 3

### **Translation**

-- select source reporting unit name as forest from fire table left joined with source nwcg link table on NWCG source link id from both the tables, left

join on source id from source nwcg link and source system reporting tables and left join on stat cause code from fire and stat cause details tables grouped by

source reporting unit name, ordered by count of fires(count of fod\_id) and limit by 3.

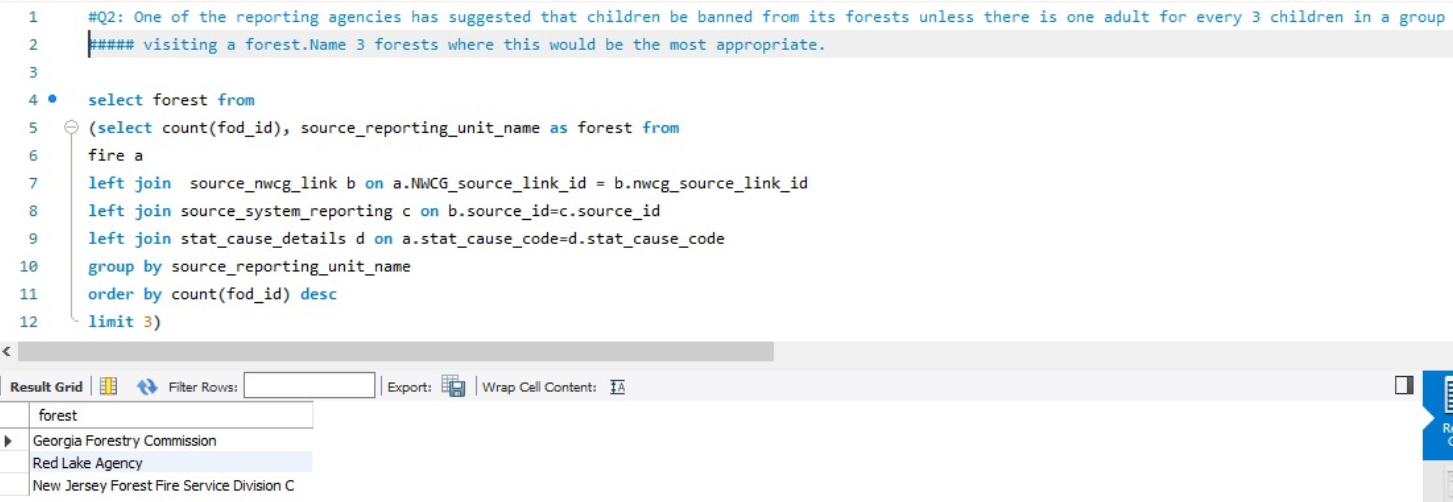
### **Clean-Up**

-- cleanup: select source reporting unit name as forest from fire left join with source nwcg link on NWCG\_source\_link\_id from fire and source nwcg link, left

join on source\_id from source nwcg link and source system reporting, left join on stat\_cause\_code from fire and stat\_cause\_details, grouped by

source\_reporting\_unit\_name, order by count of fod\_id, limit by 3.

### **Screen Shot of SQL Query and Results**



## **Query 3**

### **Question 3**

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

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

Fires Due to Natural Cause: 768996; Fires Due to Manmade Causes: 1111469. We do not have enough evidence to prove that nature and not human actions is to blame for most wildfires. The statistical cause of Nature caused fires were considered as Lightening, Miscellaneous and Missing/Unidentified while the rest causes were categorized as Human caused fires. Data clearly depicts that the statement cannot be supported unless any other Human caused fires is erroneously bucketed under Nature caused fires.

# of Rows in Result: 1

### **Translation**

-- select the count of fires caused by natural causes as count\_natural from fire table left joined with stat cause details table on stat\_cause\_code

where stat\_cause\_code is either 1,9 or 13, select the count of fires caused by mannmade causes as count\_manmade from fire table left joined with stat cause detals table on stat\_cause\_code where

stat\_cause\_code is neiter 1,9 or 13 and display count\_natural and count\_manmade.

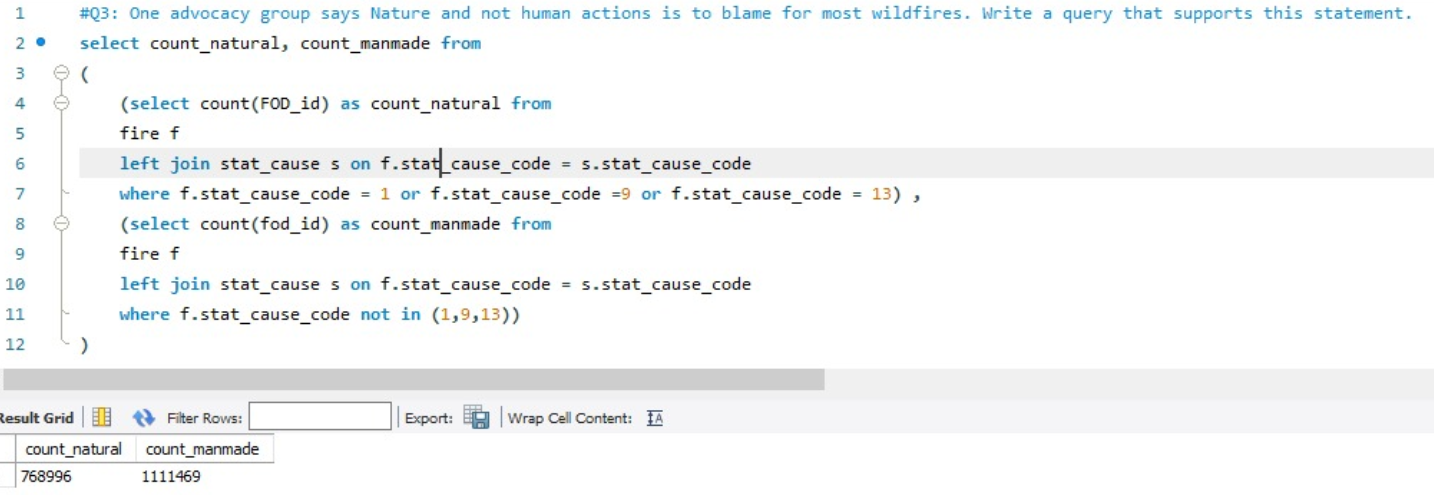
### **Clean-Up**

-- cleanup: select count(FOD\_ID) as count\_natural from fire left join with stat\_cause\_details on stat\_cause\_code

where stat\_cause\_code is (1,9 or 13), select count(FOD\_ID) as count\_manmade from fire left join with stat\_cause\_details on stat\_cause\_code where

stat\_cause\_code is not( 1,9 or 13), display count\_natural and count\_manmade.

### **Screen Shot of SQL Query and Results**



## **Query 4**

### **Question 5**

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

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

0 records returned. Since FOD ID is assumed to be the wildfire, the FOD ID is unique across the database and hence there is no chance for the one fire to be reported by multiple unit/agency. Hence result is 0.

# of Rows in Result: 0

### **Translation**

-- select FOD\_ID and count of distinct source\_reporting\_unit\_name as no\_of\_units from fire table left joined with source nwcg link table on nwcg\_sourcr\_link\_id

from both tables, left join on source\_id from source nwcg link and source system reporting tables grouped by FOD\_ID that have count of

distinct source\_reporting\_unit\_name greater than 1

### **Screen Shot of SQL Query and Results**

### 

## **Query 5**

### **Question 6**

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

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

778 Forests had more than one fire that lasted for more than 2 days

### **Translation**

-- select the no of rows from the result set obtained by selecting count of FOD\_ID as no\_of\_fires and source\_reporting\_unit\_name as forest

from another result set obtained by selecting FOD\_ID, source\_reporting\_unit\_name, difference between cont\_date and discovery\_date from fire table

left joined with source nwcg link on NWCG\_source\_link\_id from both tables, left joined on source\_id from source system reporting and

source nwcg link tables where difference between cont\_date and discovery\_date is greater than 2, grouped by source\_reporting\_unit\_name

and having count of FOD\_ID greater than 1

### **Screen Shot of SQL Query and Results**



## **Query 6**

### **Question 8**

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

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

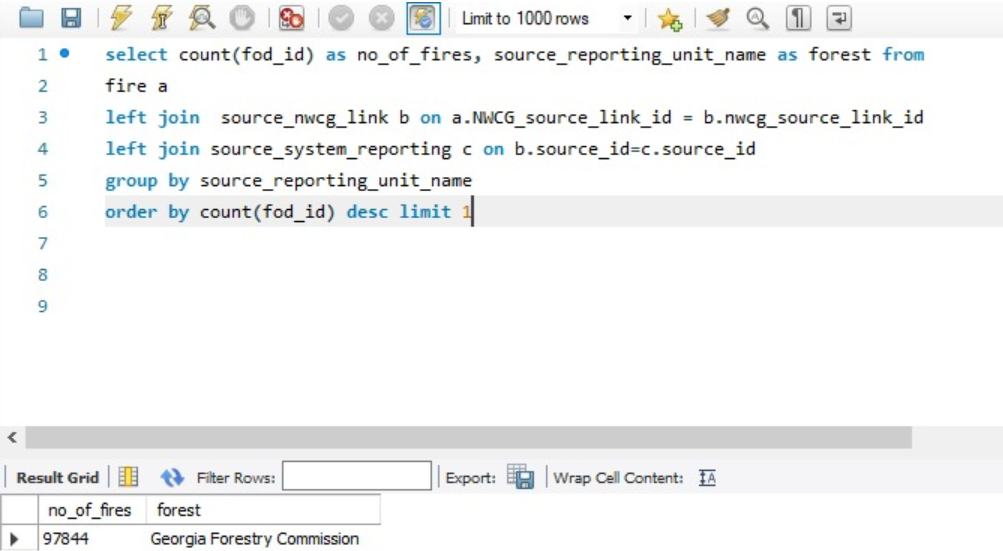
Georgia Forestry Commission has the maximum number of fires

### **Translation**

--select the source\_reporting\_unit\_name as forest and the corresponding count of fod\_id as no\_of\_fires from fire table left joined with source nwcg link table on NWCG\_source\_link\_id

left joined with source system reporting table on source\_id, grouped by source reporting unit\_name and ordered by the count of fod\_id and limit by 1

### **Screen Shot of SQL Query and Results**



# **Data Review for MongoDB**

## **Assumptions for Querying:**

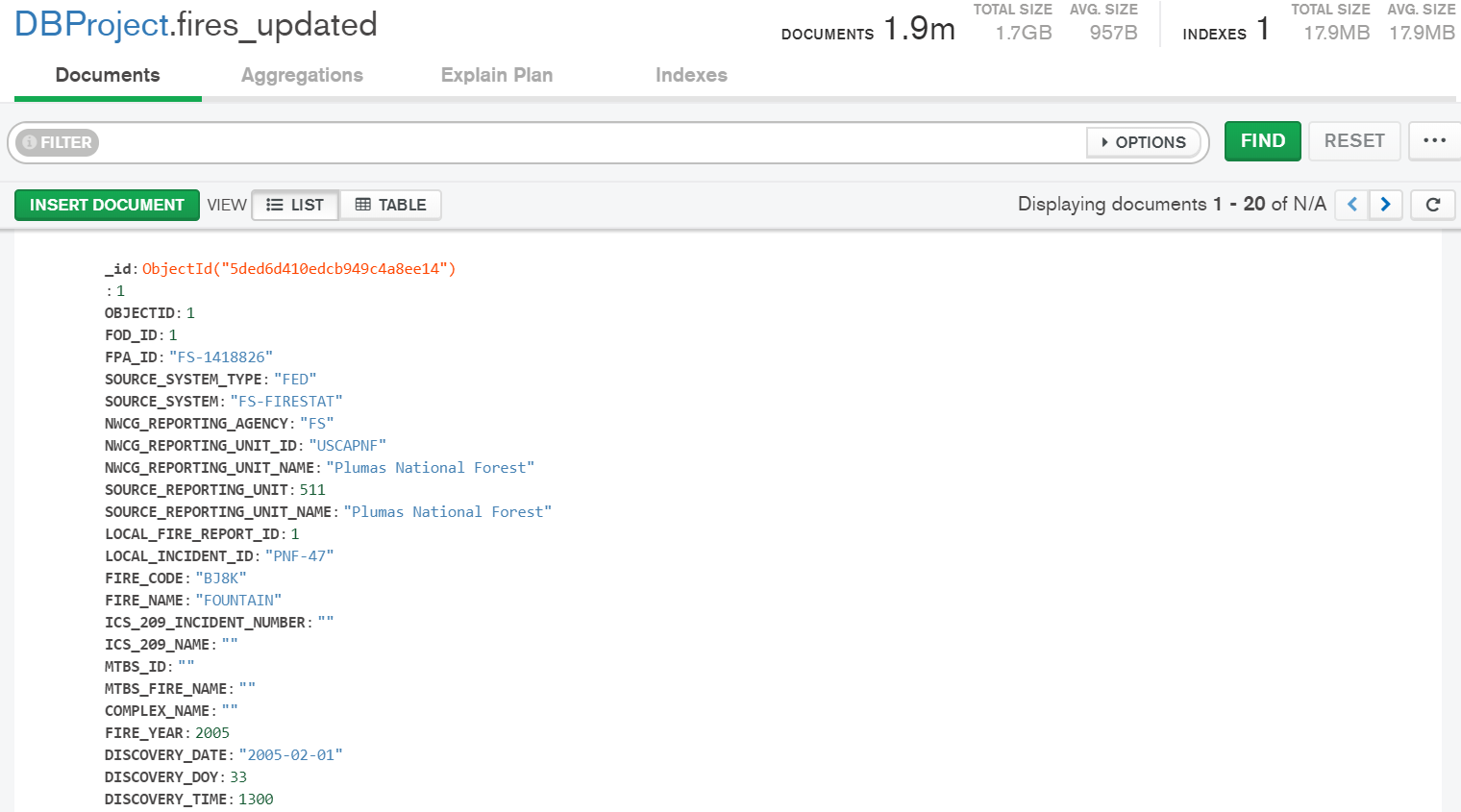
1. “Source\_Reporting\_Unit\_Name” is considered as forests with the assumption that each forest has one and only one agency to report the fires occurring the respective forest
2. “FOD\_ID” is considered as the fires under the assumption that each fire is reported only by one reporting unit

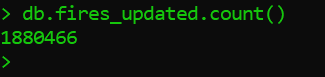
**Notes About Data Collections**

The un-normalized data of the Fires database and NWCG database is loaded.

# **Physical Mongo Database**

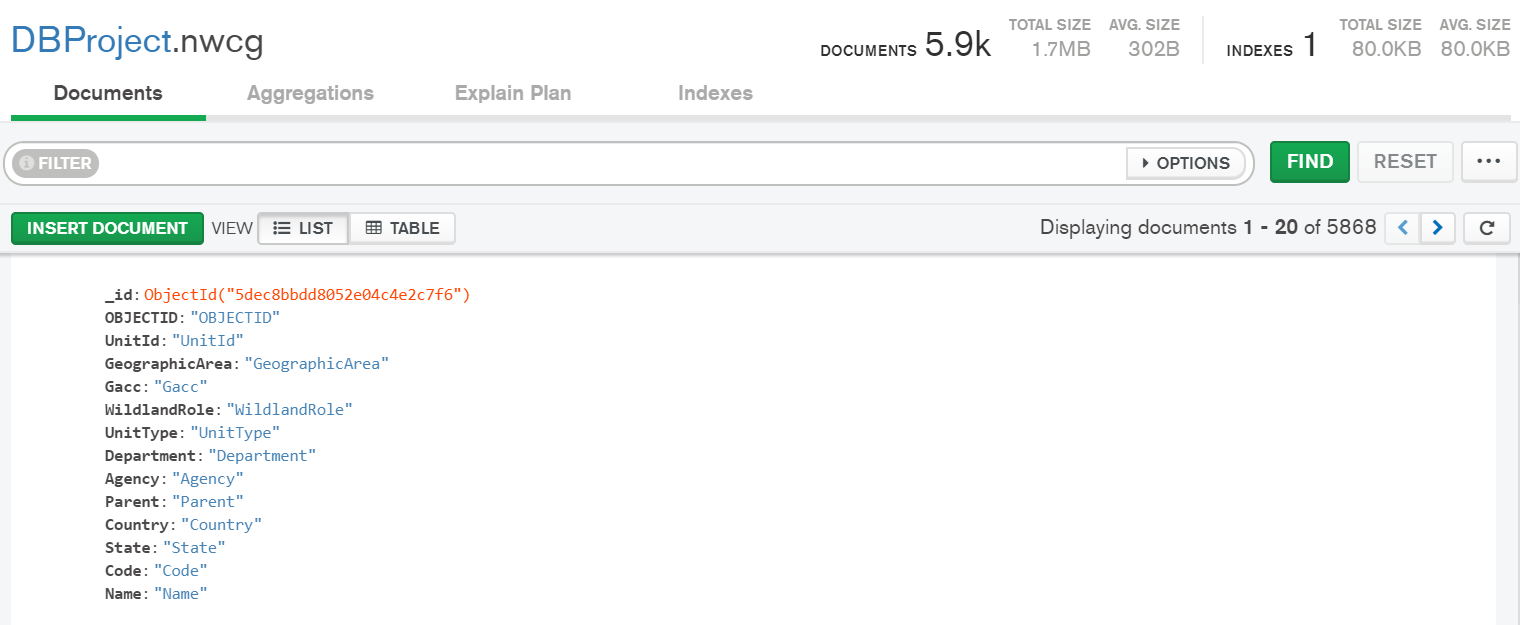
## **Screen shot of Physical Database objects (Database, Collections and Attributes)**

1. Fires Collection



## Fires Database is uploaded with 1.88million documents

1. NWCG Collection





NWCG database is uploaded with 5868 documents

## **Data in the Database**

|  |  |  |
| --- | --- | --- |
| **Collection Name** | **Relationshps With Other Collections (if any)** | **# of Documents in Collection** |
| Fires\_Updated | NA | 1880466 |
| NWCG | NA | 5868 |

## **Comment:** While uploading the databases, a null record was automatically uploaded for each of the databases. Hence the increment of the total documents by 1 for each.

# **MongoDB Queries/Code**

## **Query 1**

### **Question 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?

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

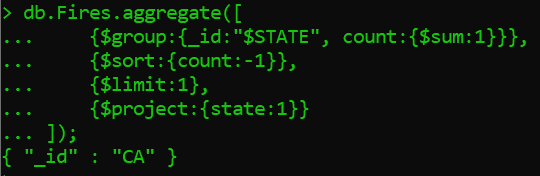
**Result:** California is the state that has the best chance to win a share of the fund.

**Result Transalation:** The result depicts the state with maximum fires which would ideally have the maximum debris and hence they should stand the best chance to win the fund (to have the positive impact on the state through cleaning up utilizing these funds)

### **Translation - Query**

* DBProject Database selected to Db variable
* Fires Collection used
* Aggeregate Pipeline
* Group the fires using “State” using $group
* Order the result using sort in descending order using $sort
* Display the State with maximum fires using $project and limit the result to 1 using $limit

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



## **Query 2**

### **Question 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.

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

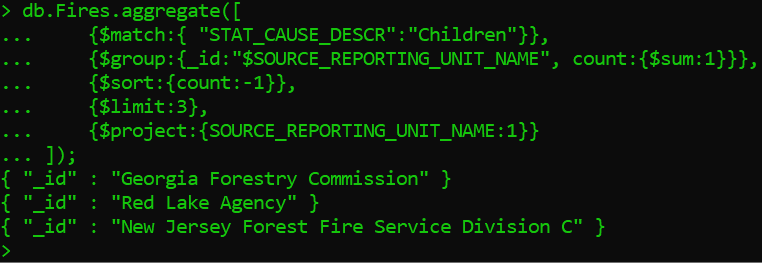
**Result:** Georgia Forestry Commission, Red Lake Agency and New Jersey Forest Fire Service Division C are the three forests where the mentioned suggestion seems appropriate.

**Result Translation:** These three forests have the maximum fires caused by the children. Since children have been the statistical cause for the fire, it is ideal to implement regulations on the visit of children into the forests under the supervision of the adults.

### **Translation - Query**

* DBProject Database selected to Db variable
* Fires Collection used
* Aggeregate Pipeline
* Filter Statistical Cause as Children
* Group the Source Reporting Unit Name of the resultant data using $group
* Order the result using sort in descending order using $sort
* Display the Source Reporting Unit Name using $project and limit the result to 1 using $limit

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



## **Query 3**

### **Question 3:**

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

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

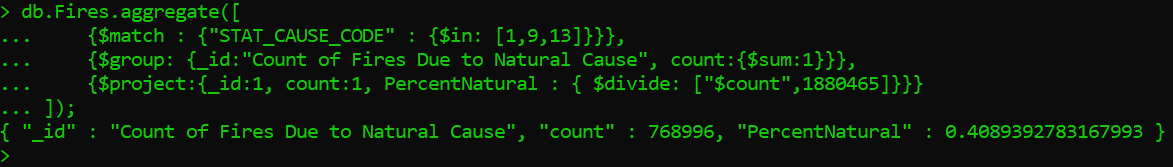
**Result:** Fires Due to Natural Cause: 768996; Percent of Nature caused Fires: 40.89%. We do not have enough evidence to prove that nature and not human actions is to blame for most wildfires.

**Result Translation:** The statistical cause of Nature caused fires were considered as Lightening, Miscellaneous and Missing/Unidentified while the rest causes were categorized as Human caused fires. Data clearly depicts that the statement cannot be supported unless any other Human caused fires is erroneously bucketed under Nature caused fires.

### **Translation**

* DBProject Database selected to Db variable
* Fires Collection used
* Aggeregate Pipeline
* Statistical Code – 1,9,13 which correspond to “Lightening”, “Miscellaneous” & “Missing/Unidentified” are filtered using $match
* Group the resultant data using $group
* Display the resultant data using $project along with the calculated field - Percentage of the Nature caused fires

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



## **Query 4**

### **Question 5**

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

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

**Result:** 0

**Result Translation:** Since FOD ID is assumed to be the wildfire, the FOD ID is unique across the database and hence there is no chance for the one fire to be reported by multiple unit/agency. Hence result is 0.

### **Translation**

* DBProject Database selected to Db variable
* Fires Collection used
* Aggeregate Pipeline
* Group the data by FOD\_ID and sum the Source Reporting Unit Name using $group
* Filter the Source Reporting unit greater than 1 using $match
* The length of the resultant data is returned

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



## **Query 5**

### **Question 6**

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

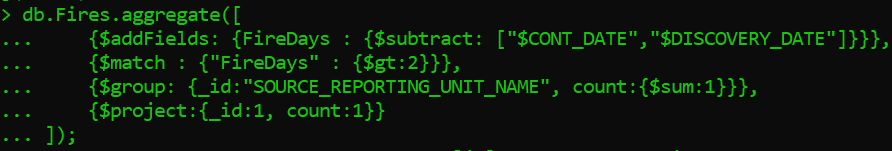
### **Notes/Comments About MongoDB Query/Code and Results (Include # of Documents in Result)**

778 Forests had more than one fire that lasted for more than 2 days

### **Translation**

* DBProject Database selected to Db variable
* Fires Collection used
* Aggeregate Pipeline
* Calculated Field FireDays created using $addFields (Subtract Discovery Date from Controlled Date)
* Filter the data set with FireDays > 2 using $match
* Group the data by fires on Source Reporting Unit Name using $group
* The length of the resultant data is returned

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



## **Query 6**

### **Question 8:**

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

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

**Result:** Georgia Forestry Commission has the maximum number of fires

### **Translation**

* DBProject Database selected to Db variable
* Fires Collection used
* Aggeregate Pipeline
* Group the resultant data on Source Reporting Unit Name using $group
* Order the result data using $sort in the descending order
* Display the resultant data – “Source Reporting Unit Name” field using $project limiting the result table to 1 record using $limit

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

