Group #: Tue4pm\_Group1

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

BUAN 6320

Group Members:

Aditi Kulkarni

Rupal Shrivastava

Sharannya Nair

Srushti Shah

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Activity** | Aditi | Rupal | Sharannya | Srushti |
| Prepared Data Model and Created Physical DB | x | x | x | x |
| Loaded Data into Database | x | x | x | x |
| 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 6

Physical Database 7

Assumptions/Notes About Data Set 7

Screen shot of Physical Database objects 8

Data in the Database 8

SQL Queries 19

Query 1 19

Question 19

Notes/Comments About SQL Query and Results (Include # of Rows in Result) 19

Translation 19

Screen Shot of SQL Query and Results 19

Query 2 20

Question 20

Notes/Comments About SQL Query and Results (Include # of Rows in Result) 20

Translation 20

Screen Shot of SQL Query and Results 20

Query 3 21

Question 21

Notes/Comments About SQL Query and Results (Include # of Rows in Result) 21

Translation 21

Screen Shot of SQL Query and Results 21

Query 4 22

Question 22

Notes/Comments About SQL Query and Results (Include # of Rows in Result) 22

Translation 22

Screen Shot of SQL Query and Results 22

Query 5 23

Question 23

Notes/Comments About SQL Query and Results (Include # of Rows in Result) 23

Translation 23

Screen Shot of SQL Query and Results 23

Query 6 24

Question 24

Notes/Comments About SQL Query and Results (Include # of Rows in Result) 24

Translation 24

Screen Shot of SQL Query and Results 24

Data Review for MongoDB 25

Assumptions/Notes About Data Collections, Attributes and Relationships between Collections 25

Physical Mongo Database 26

Assumptions/Notes About Data Set 26

Screen shot of Physical Database objects (Database, Collections and Attributes) 26

Data in the Database 26

MongoDB Queries/Code 27

Query 1 27

Question 27

Notes/Comments About MongoDB Query/Code and Results (Include # of Documents in Result) 27

Translation 27

Screen Shot of MongoDB Query/Code and Results 27

Query 2 28

Question 28

Notes/Comments About MongoDB Query/Code and Results (Include # of Documents in Result) 28

Translation 28

Screen Shot of MongoDB Query/Code and Results 28

Query 3 29

Question 29

Notes/Comments About MongoDB Query/Code and Results (Include # of Documents in Result) 29

Translation 29

Screen Shot of MongoDB Query/Code and Results 29

Query 4 30

Question 30

Notes/Comments About MongoDB Query/Code and Results (Include # of Documents in Result) 30

Translation 30

Screen Shot of MongoDB Query/Code and Results 30

Query 5 31

Question 31

Notes/Comments About MongoDB Query/Code and Results (Include # of Documents in Result) 31

Translation 31

Screen Shot of MongoDB Query/Code and Results 31

Query 6 32

Question 32

Notes/Comments About MongoDB Query/Code and Results (Include # of Documents in Result) 32

Translation 32

Screen Shot of MongoDB Query/Code and Results 32

# Data Model

## Assumptions/Notes About Data Entities and Relationships

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

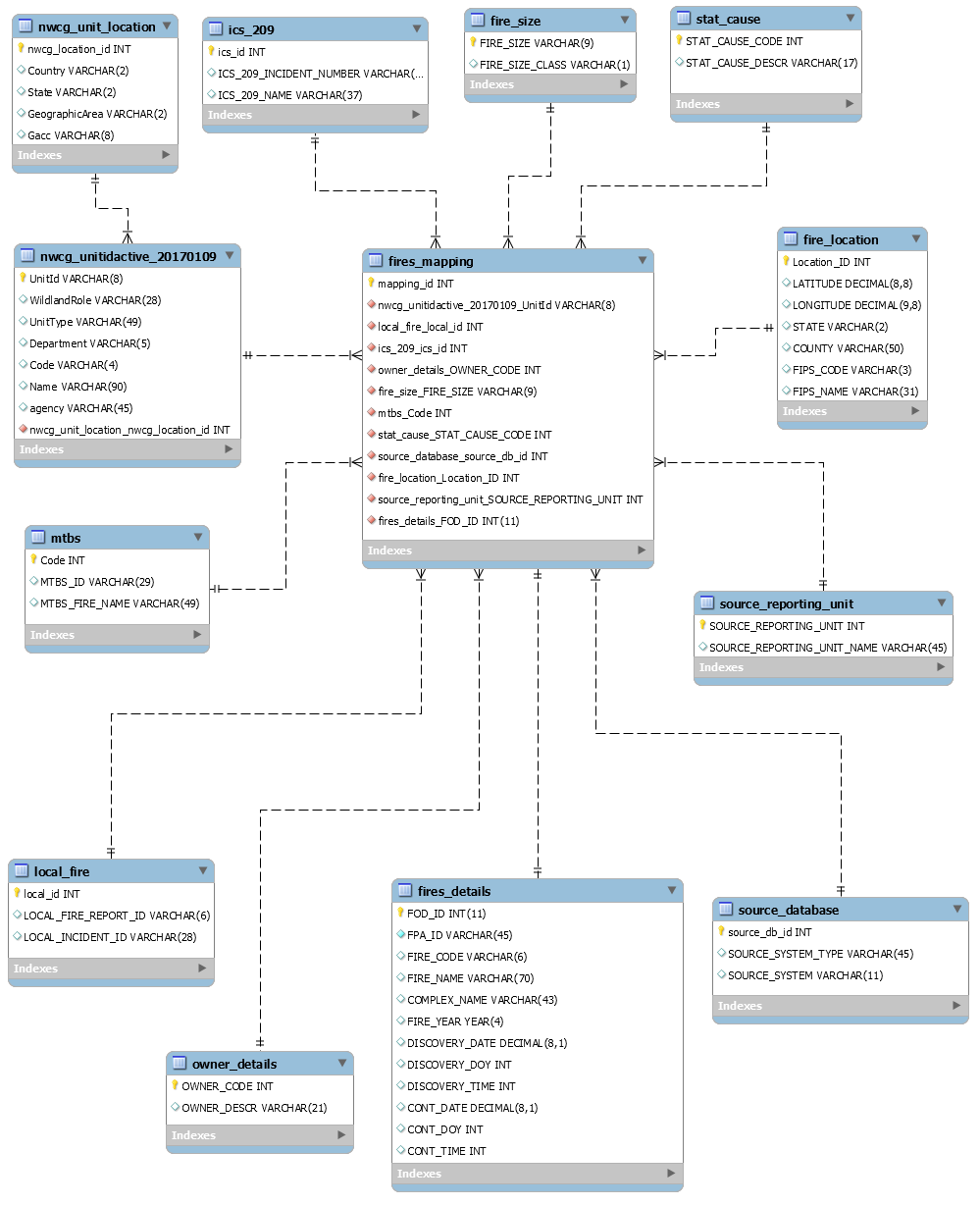
1. A source database can have one to many fire incidents
2. A fire incident can be taken only from one and only one source database
3. A location can have one-to-many fire incidents
4. A fire incident has one and only one location
5. One cause can give rise to many fires
6. A fire can have only one cause
7. There can be multiple fires of the same size
8. A fire can have one and only one size
9. An MTBS ID can have zero to many fire incidents
10. A fire incident can have one and only one MTBS ID
11. An ICS\_209 ID can have zero to many fire incidents
12. A fire incident can have one and only one ICS\_209 ID
13. A local incident ID can have one to many fire incidents
14. A fire incident can have one and only one local incident ID
15. An owner can manage one to many fire incidents
16. A fire incident can be managed by one and only one owner
17. A NWCG Agency can have one to many fire incidents
18. A fire incident can have one and only NWCG Agency
19. A location can have one to many NWCG Agencies
20. A NWCG Agency can be in one and only one location
21. One mapping id can be in one and only one FOD ID and vice versa

Include reasons why the data model is in 3NF.

Reasons why the data model is in 3NF:

1. There are no multipart or multi-value fields
2. Every table has a primary key
3. There is no partial dependency: every non-key column is functionally depending on the entire primary key
4. Each table has no partial dependency
5. There is no transitive dependency: every non-key column functionally depends only on the entire primary key of its table
6. The data model is in the 3NF form to reduce the duplication of data and ensuring referential integrity
7. The 3NF form also makes the data model more informative and reduces the need for restructuring over time
8. It also makes the data model neutral to different kinds of query statistics

## Entity-Relationship Diagram



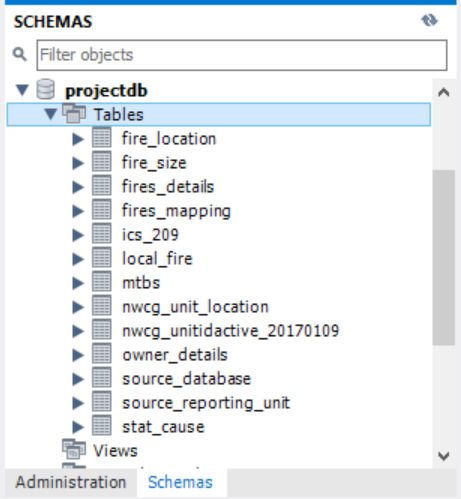
# Physical Database

## Assumptions/Notes About Data Set

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

1. A fire incident is taken only from the source database for the years 1992-2015.
2. Fire size class consists of the range fire size (A=greater than 0 but less than or equal to 0.25 acres, B=0.26-9.9 acres, C=10.0-99.9 acres, D=100-299 acres, E=300 to 999 acres, F=1000 to 4999 acres, and G=5000+ acres). Fire size is determined after the fire has taken place, so it has an identifying relationship with the fires table.
3. Majority of the null values are observed in mtbs\_id, mtbs\_fire\_name, ics\_209\_incident\_number, ics\_209\_name, and complex\_name.
4. The parent column in fires\_details table has null values due to which we have dropped these two columns.
5. shape column in the fires\_details table did not have relevant data due to which we have dropped the column.
6. The columns local\_fire\_report\_id and local\_incident\_id have null values. Hence, we created local\_id as the primary key of the table local\_fire.
7. Ics\_209\_incident\_number and ics\_209\_name have null values. Hence, we created ics\_id as the primary key of the table ics\_209.
8. Mtbs\_id and mtbs\_fire\_name have null values. Hence, we created mtbs\_code as the primary key of the table mtbs.
9. Source\_system\_type and source\_system do not have a distinct combination of the two variables. Hence, we created source\_db\_id as the primary key of the table source\_database.
10. Latitude and longitude do not have a distinct combination of the two variables. Hence, we created location\_id as the primary key of the table fire\_location.
11. Owner\_code has distinct and non-null values so; we have used it as the primary key for the owner\_details table.
12. Source\_reporting\_unit has distinct and non-null values so; we have used it as the primary key for the source\_reporting\_unit table.
13. Stat\_cause\_code has distinct and non-null values so; we have used it as the primary key for the stat\_cause table.
14. Fire\_size has distinct and non-null values so; we have used it as the primary key for fire\_size table.
15. Fod\_id has distinct and non-null values; hence we have used it as the primary key for the fire\_details table.
16. We have created NWCG\_location\_id for indicating the location in the nwcg\_unit\_location table, since the combination of country, state, geographic area, and gacc is not distinct.
17. NWCG\_unitidactive\_20170109 is mapped with the fire mapping table using the unit id present in both tables.

## Screen shot of Physical Database objects

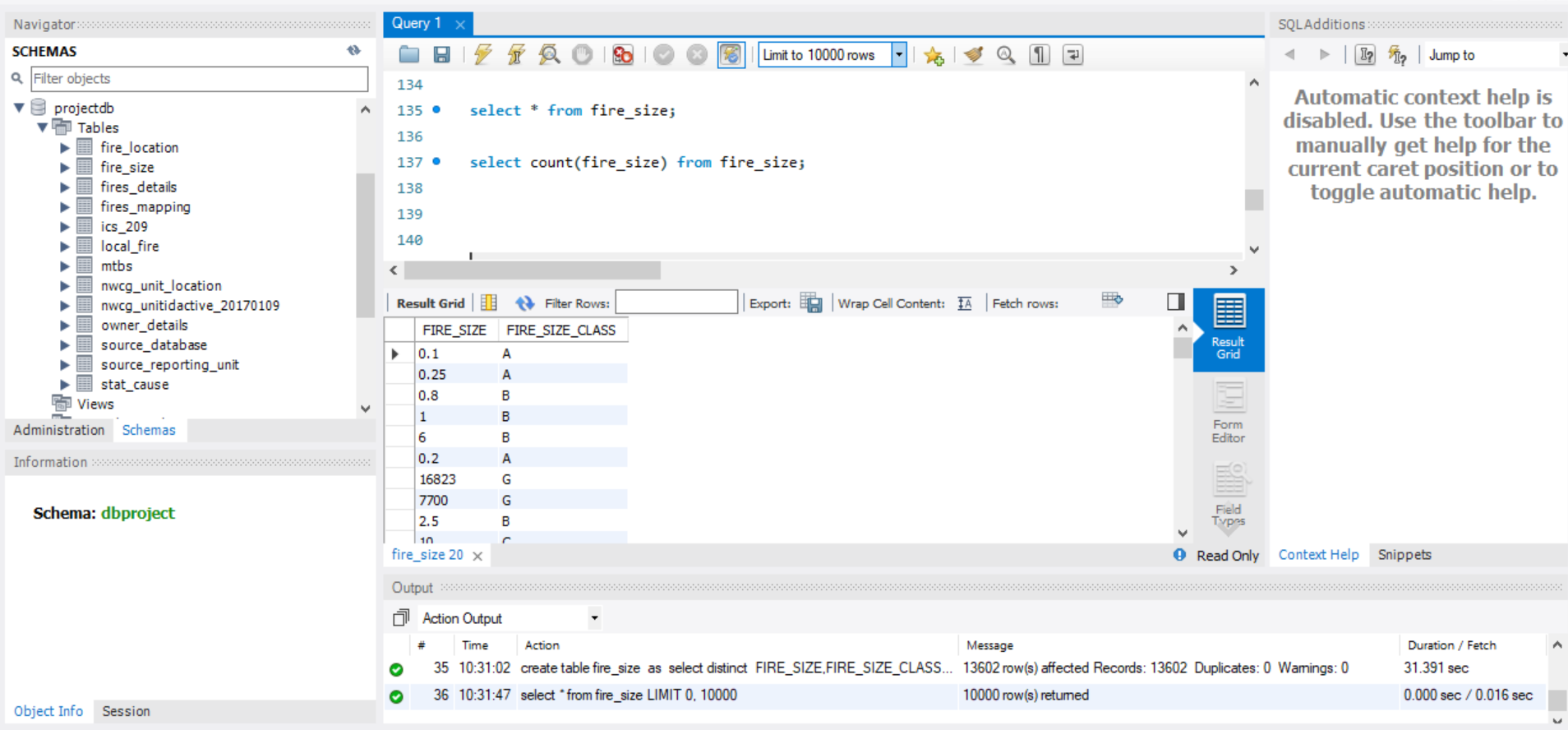


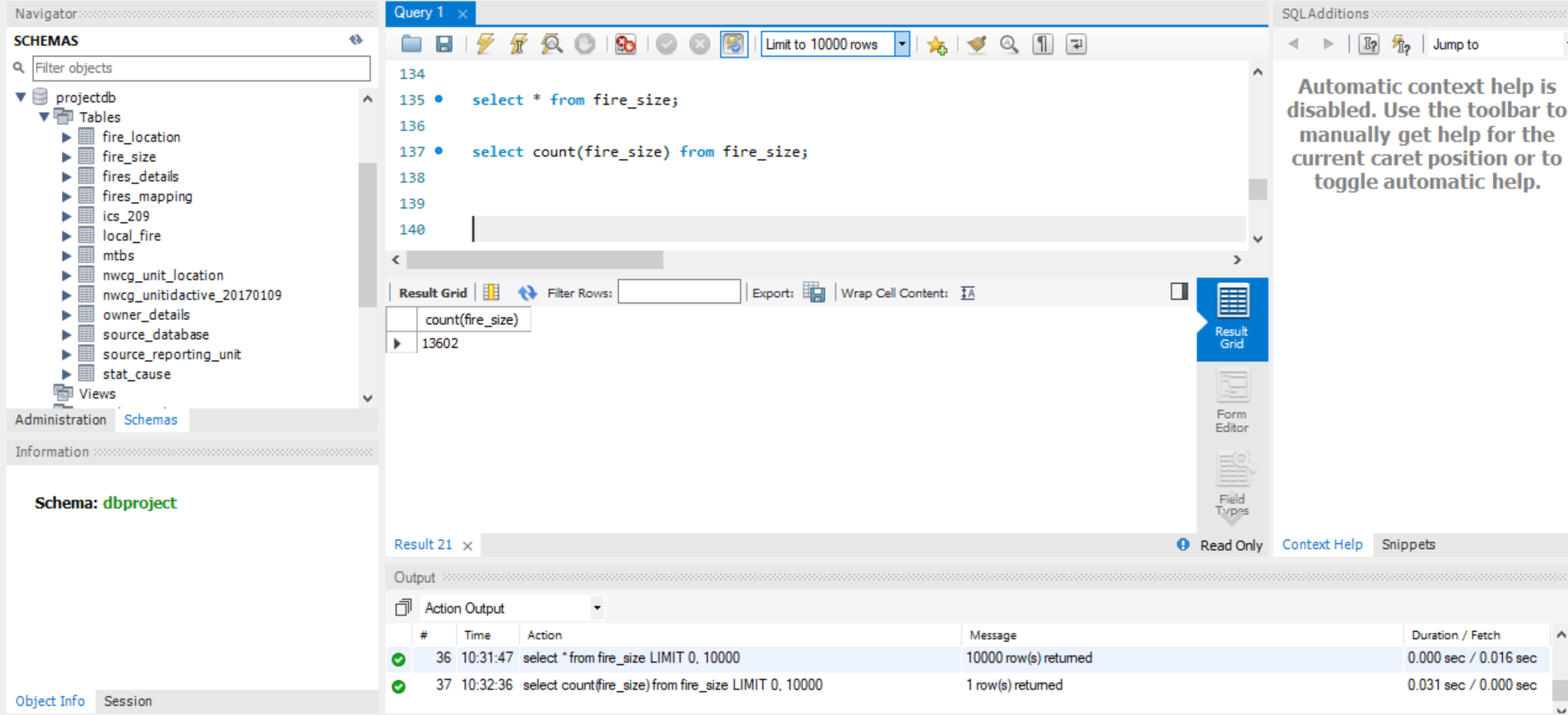
## Data in the Database

|  |  |  |  |
| --- | --- | --- | --- |
| **Table Name** | **Primary Key** | **Foreign Key** | **# of Rows in Table** |
| Fire\_size | Fire\_size | - | 13602 |
| Ics\_209 | Ics\_id | - | 23314 |
| Fire\_location | Location\_id | - | 1585927 |
| Stat\_cause | Stat\_cause\_code | - | 13 |
| Source\_reporting\_unit | Source\_reporting\_unit | - | 6646 |
| Source\_database | Source\_db\_id | - | 39 |
| Owner\_details | Owner\_code | - | 16 |
| Local\_fire | Local\_id | - | 624467 |
| Mtbs | code | - | 10482 |
| NWCG\_unit\_location | Nwcg\_location\_id | - | 86 |
| NWCG\_unitidactive\_20170109 | UnitId | Nwcg\_location\_id | 5867 |
| Fires\_details | FOD\_ID | - | 1880465 |
| Fires\_mapping | Mapping\_id | NWCG\_unitidactive\_20170109\_unitid, Local\_fire\_local\_id, ics\_209\_ Ics\_id, owner\_details\_owner\_code, fire\_size\_fire\_size,mtbs\_code, stat\_cause\_stat\_cause\_code, source\_database\_source\_db\_id, fire\_location\_location\_id, source\_reporting\_unit\_source\_reporting\_unit, fires\_details\_fod\_id | 1880465 |

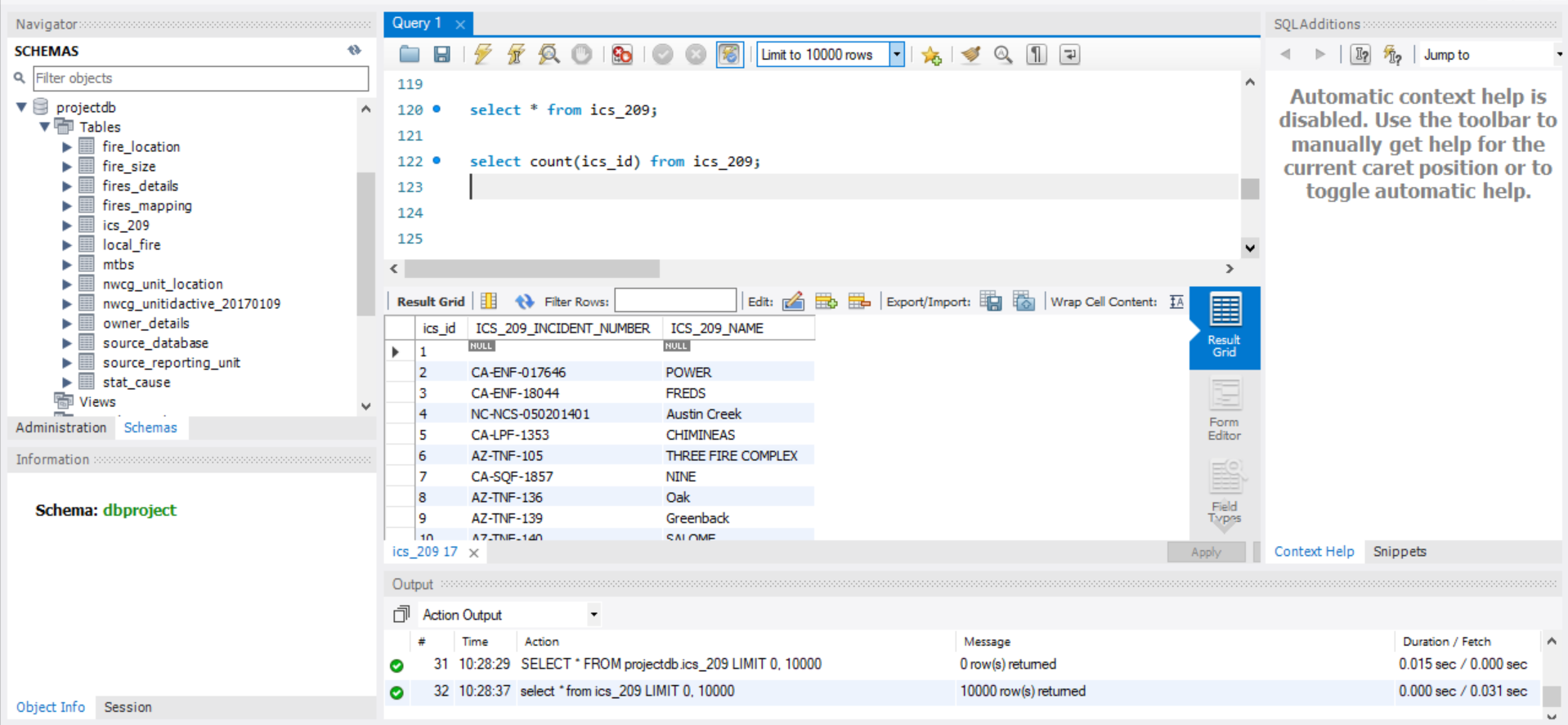
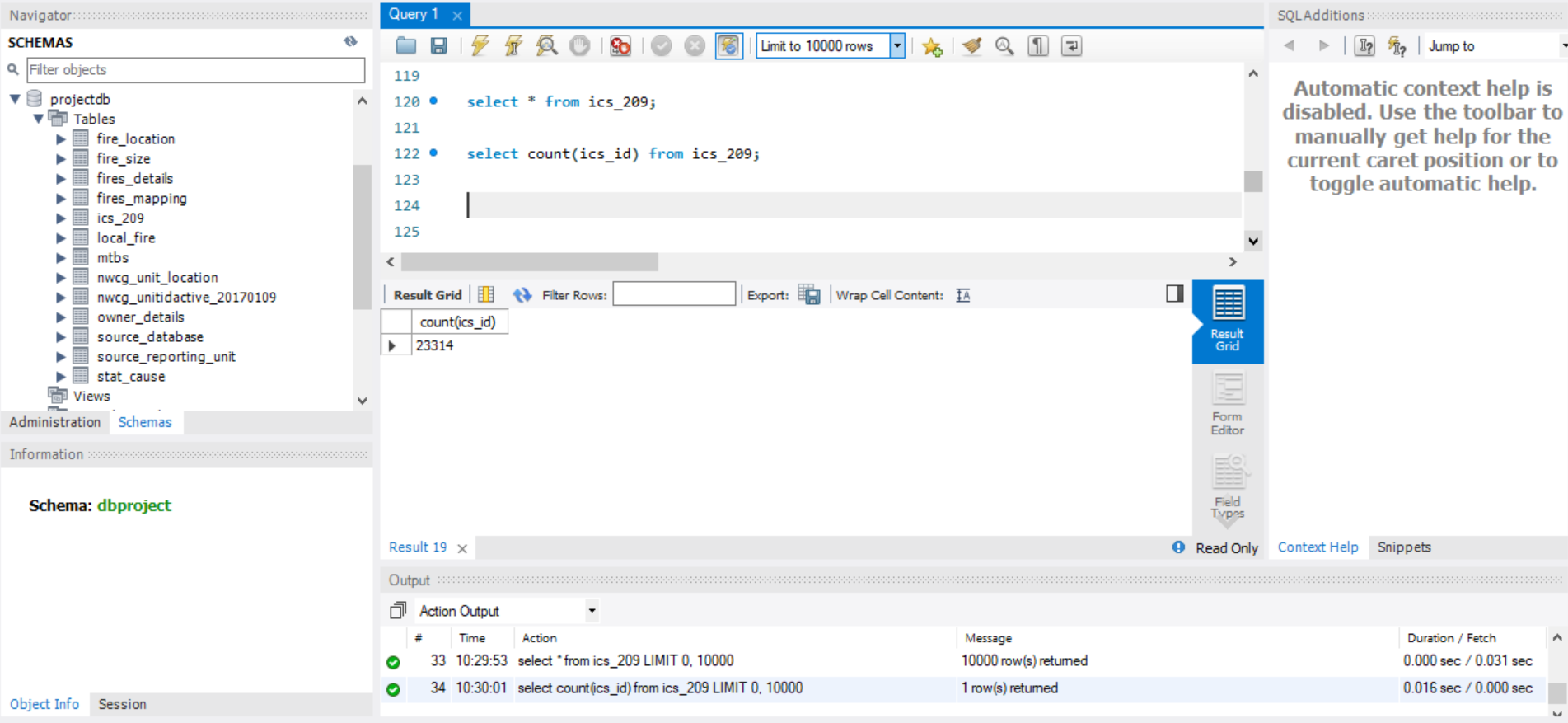
## Screenshot of the data in database

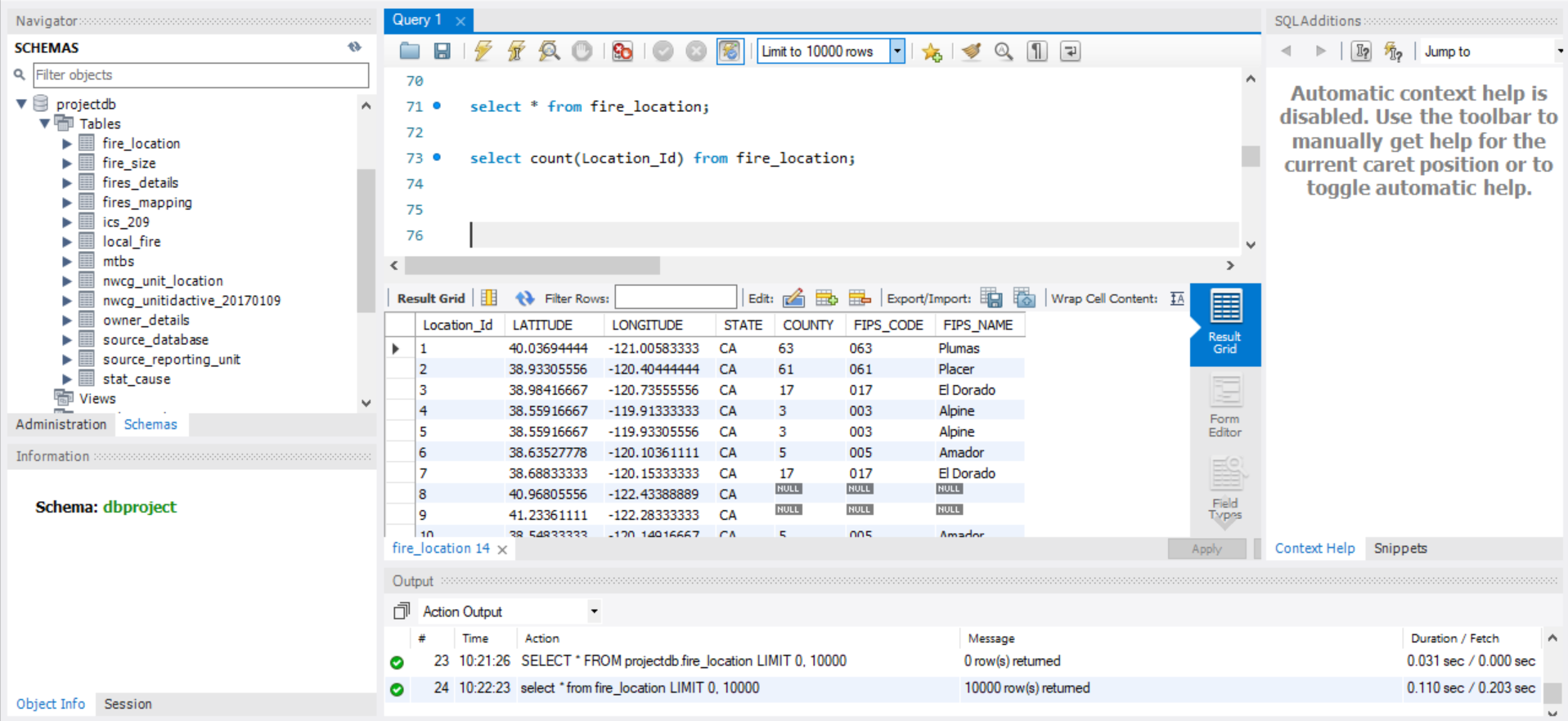
**Fire\_size**

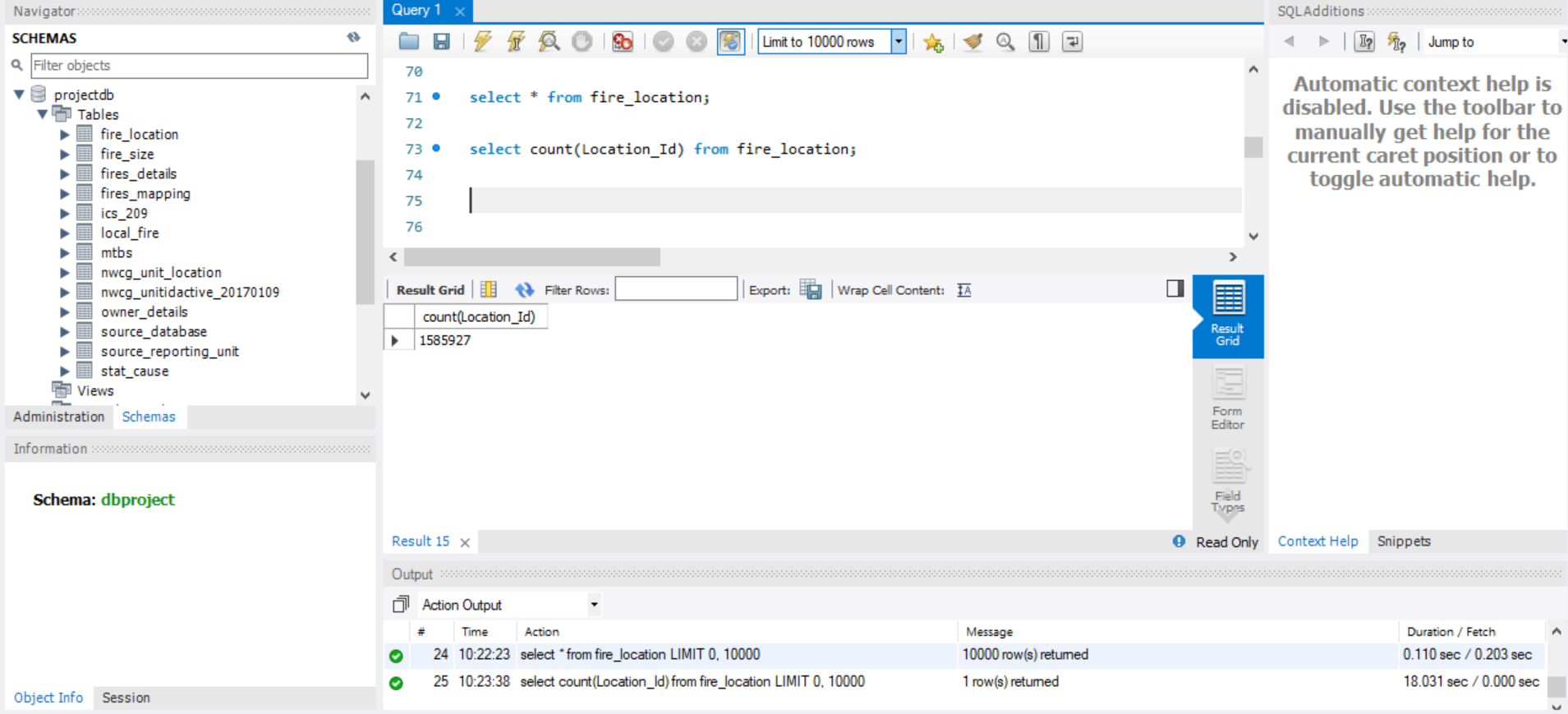




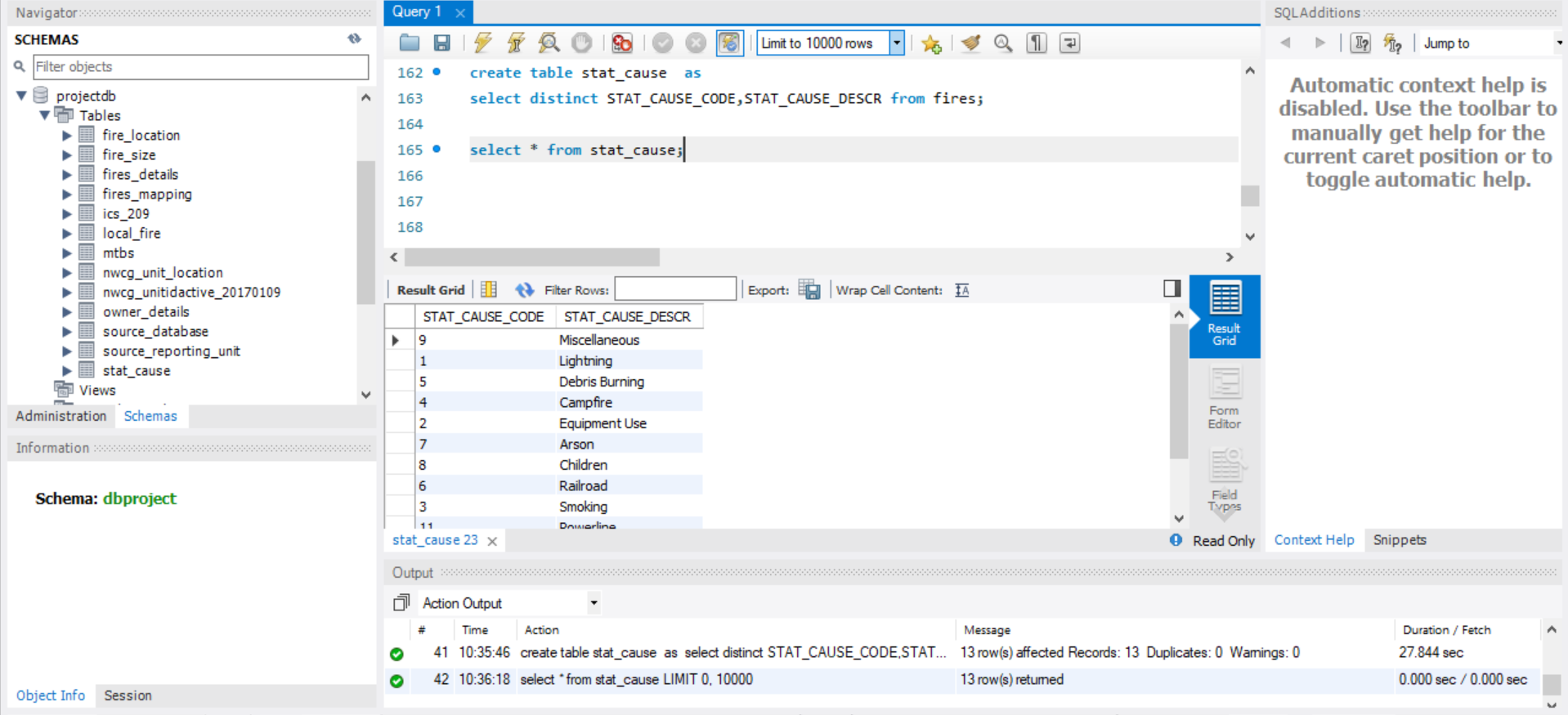
**ICS\_209**

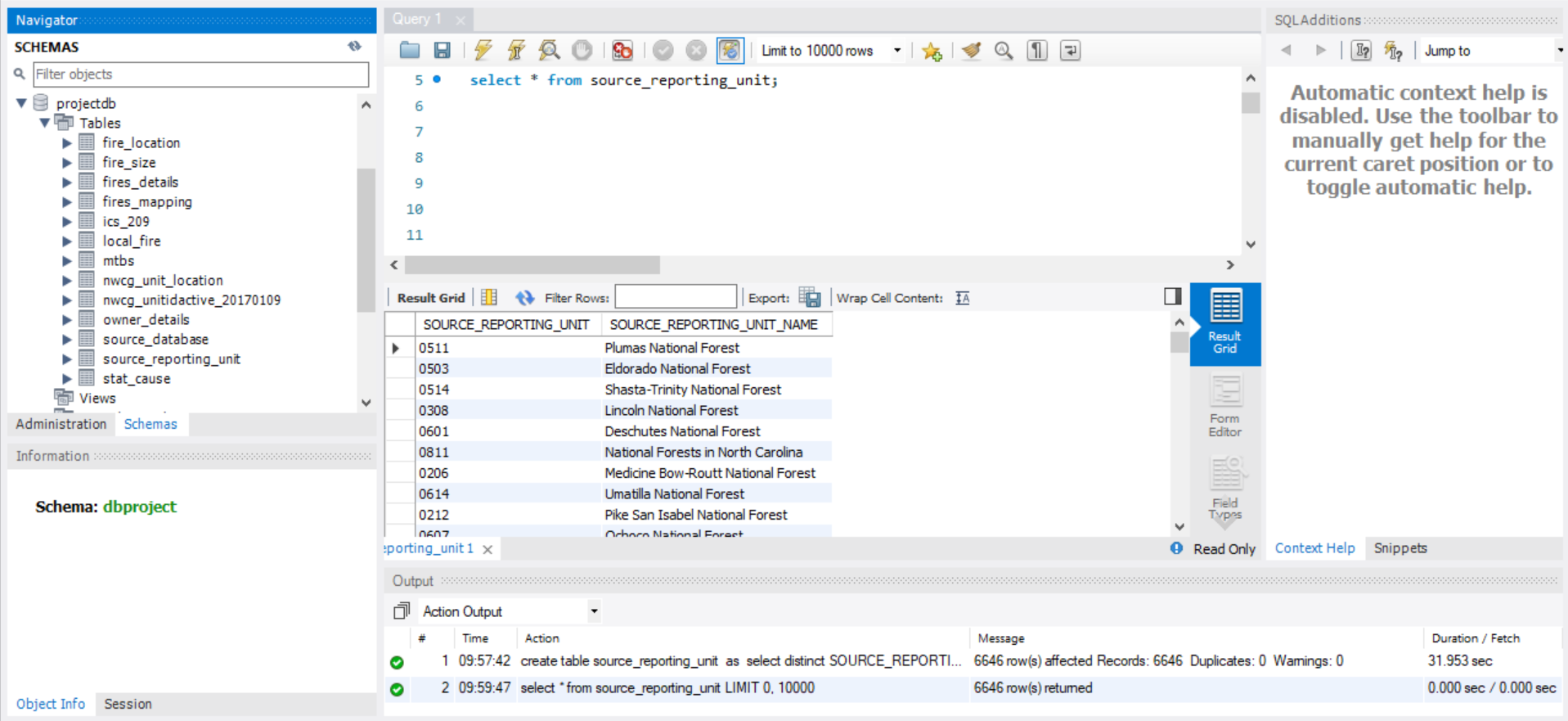
**Fire\_location**



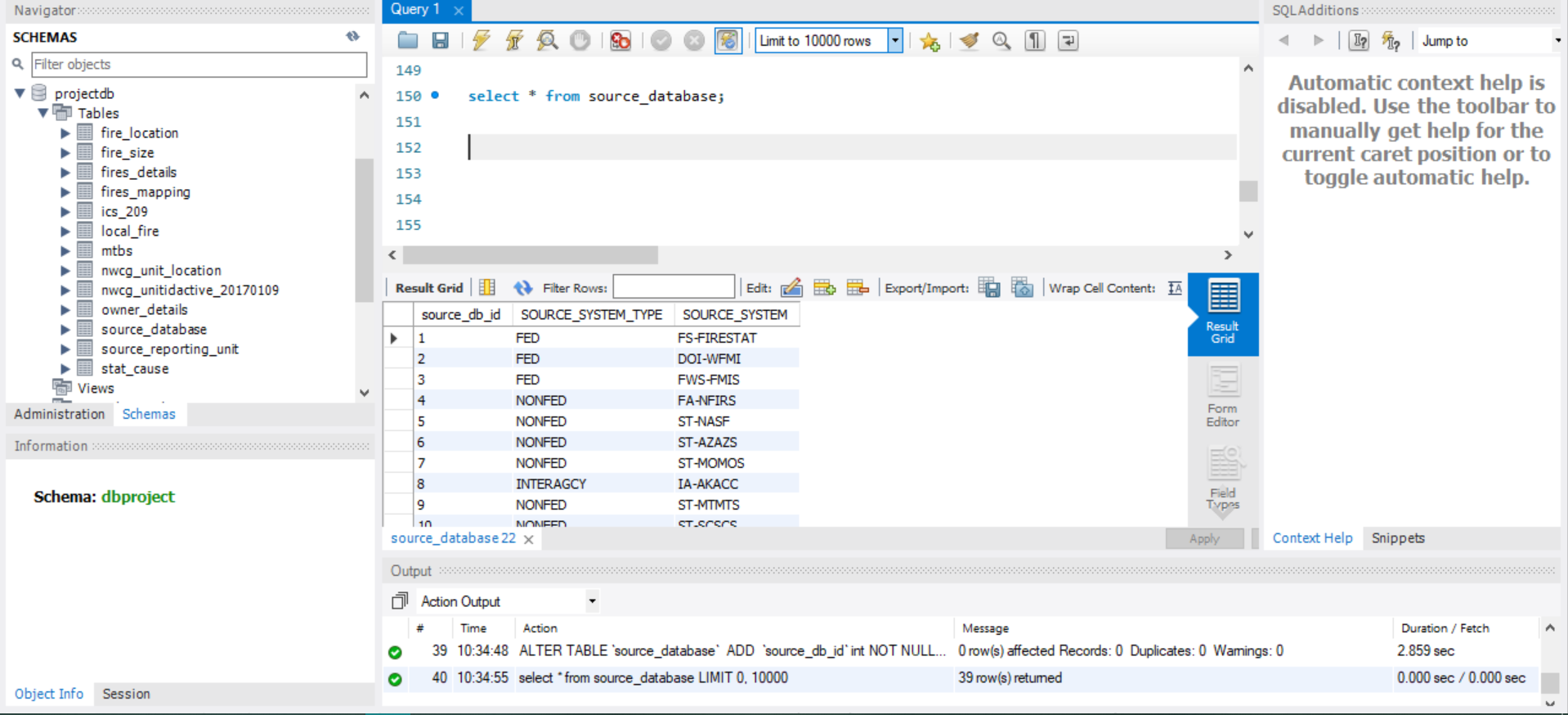
**Stat\_Cause**



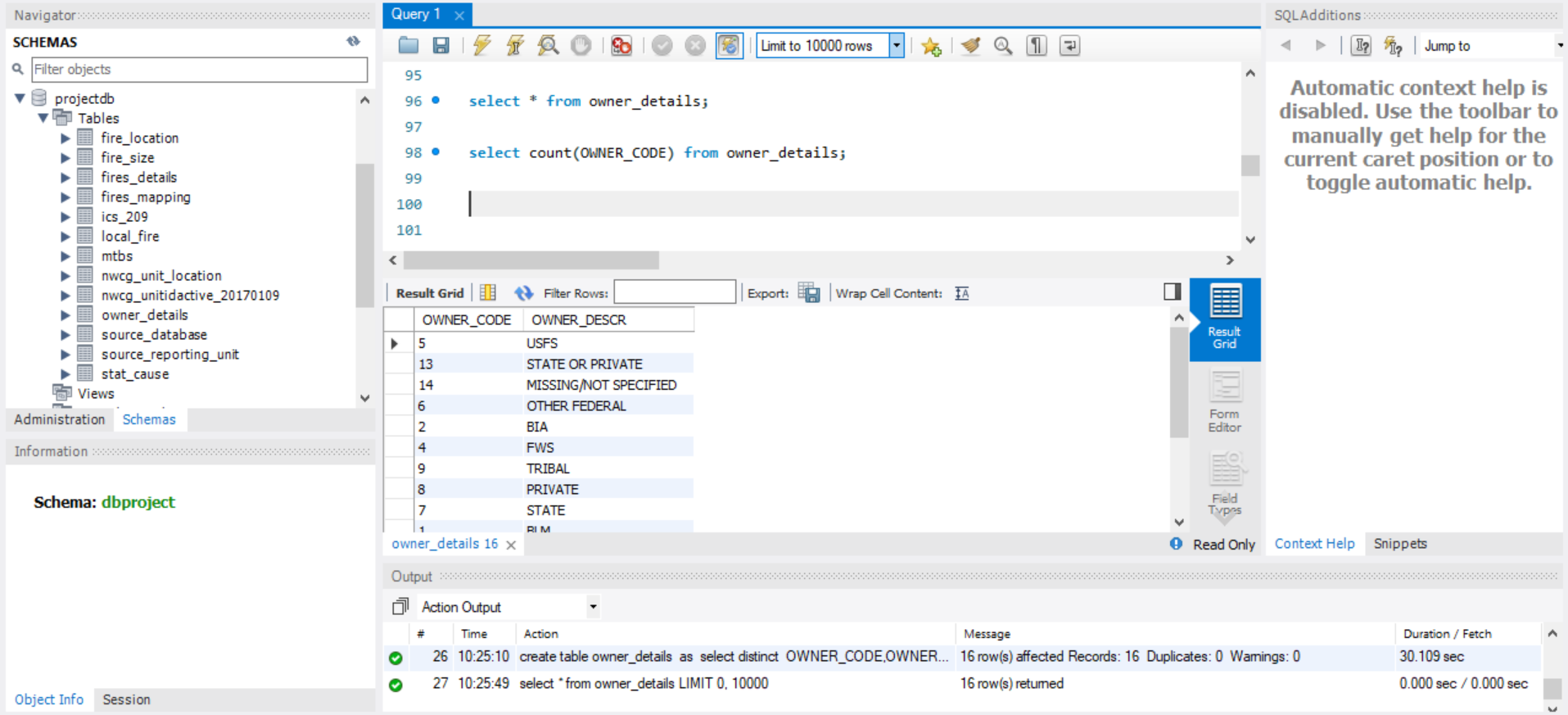
**Source\_reporting\_unit**



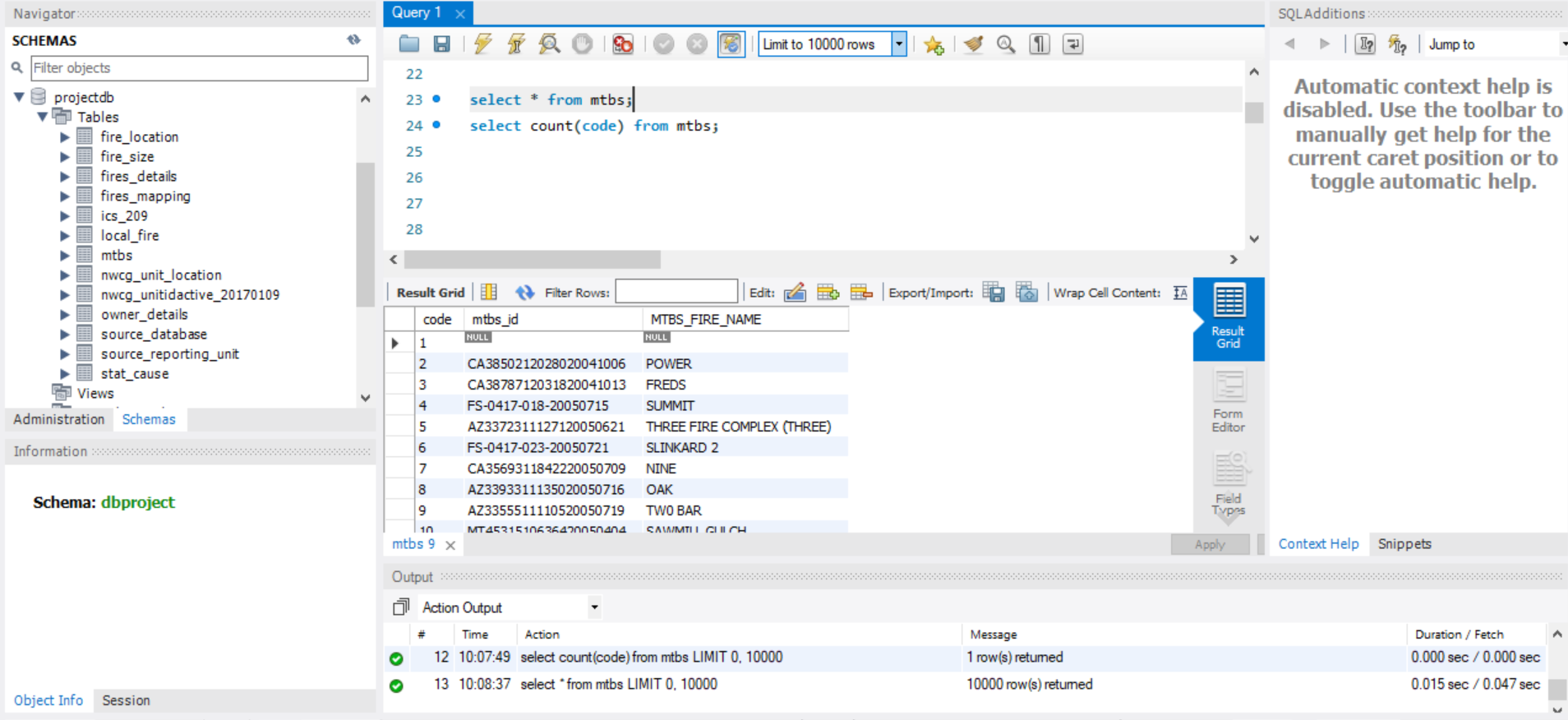
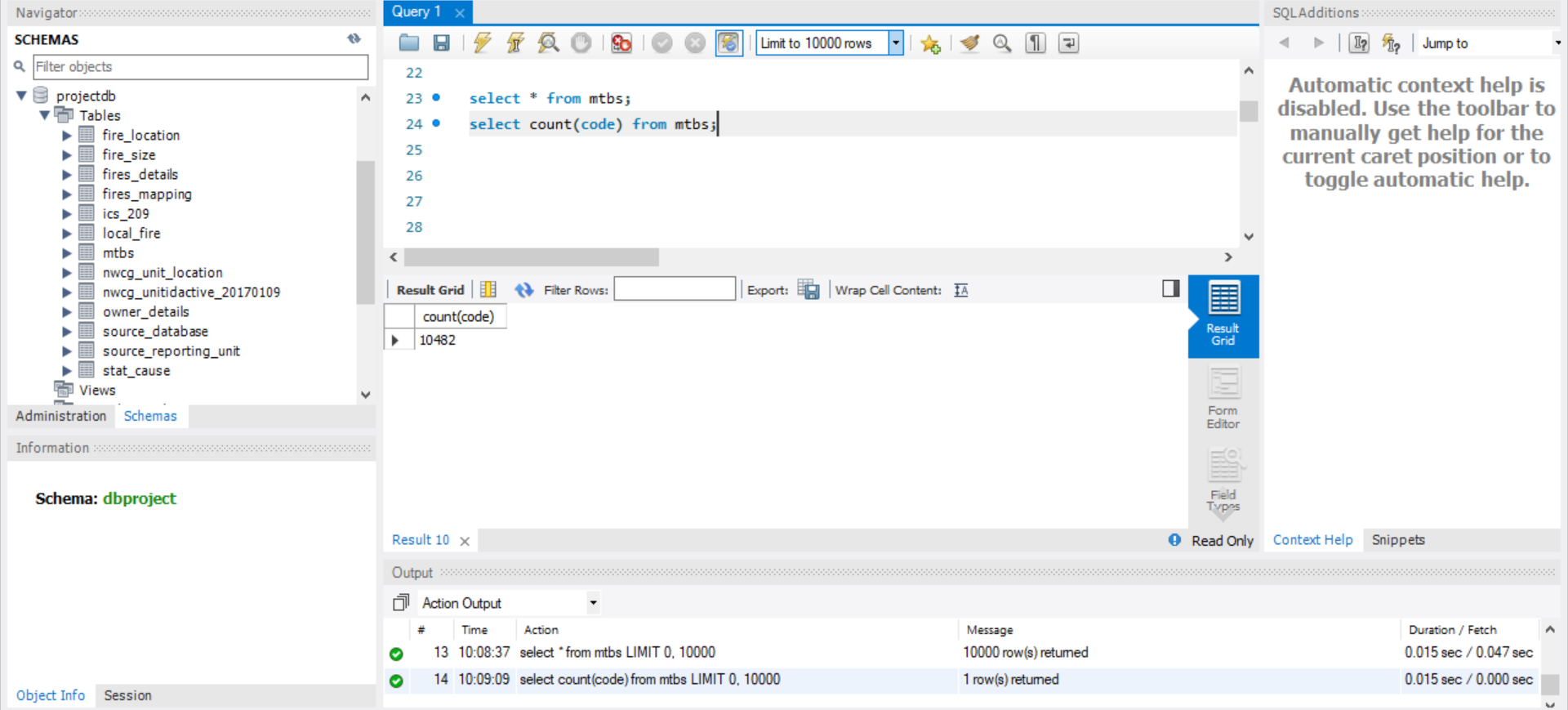
**Source\_database**



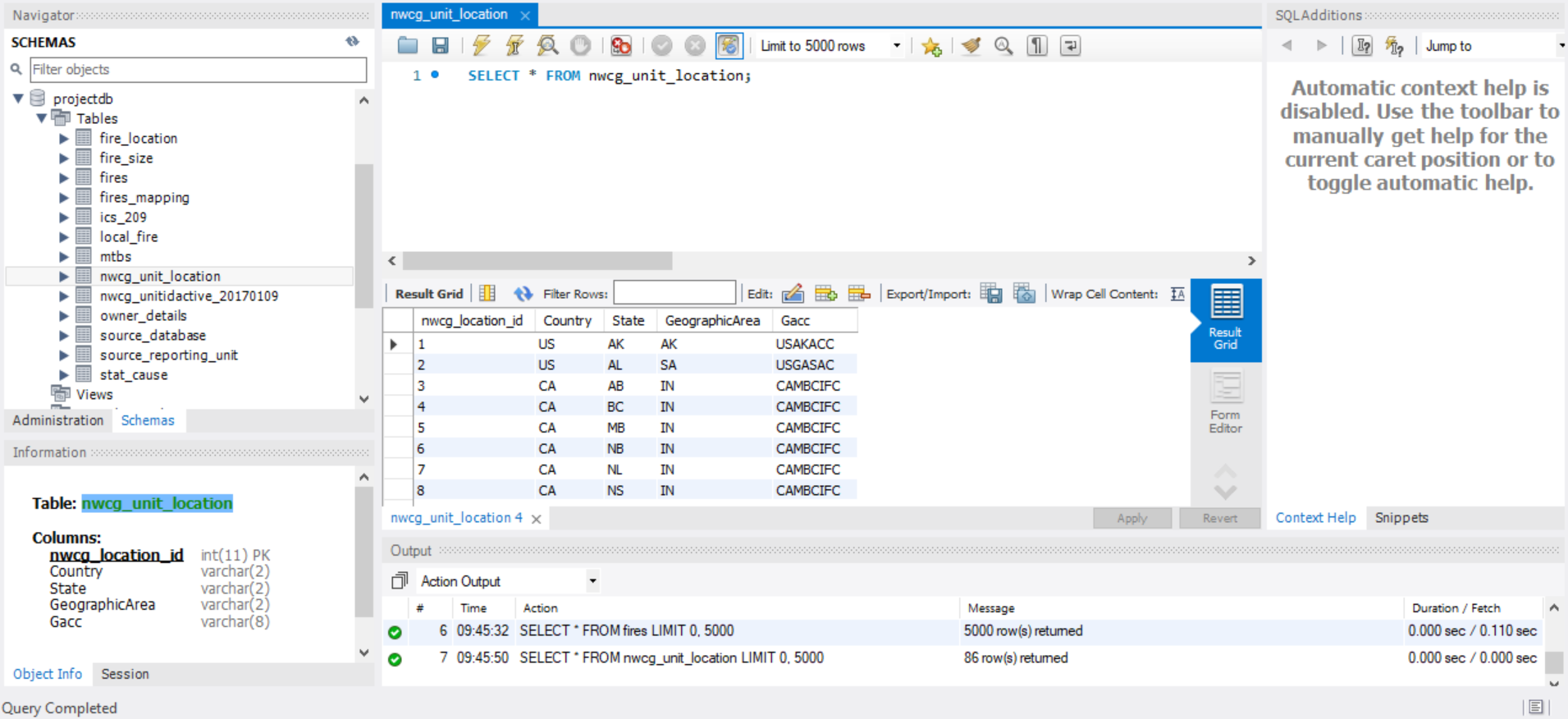
**Owner\_details**

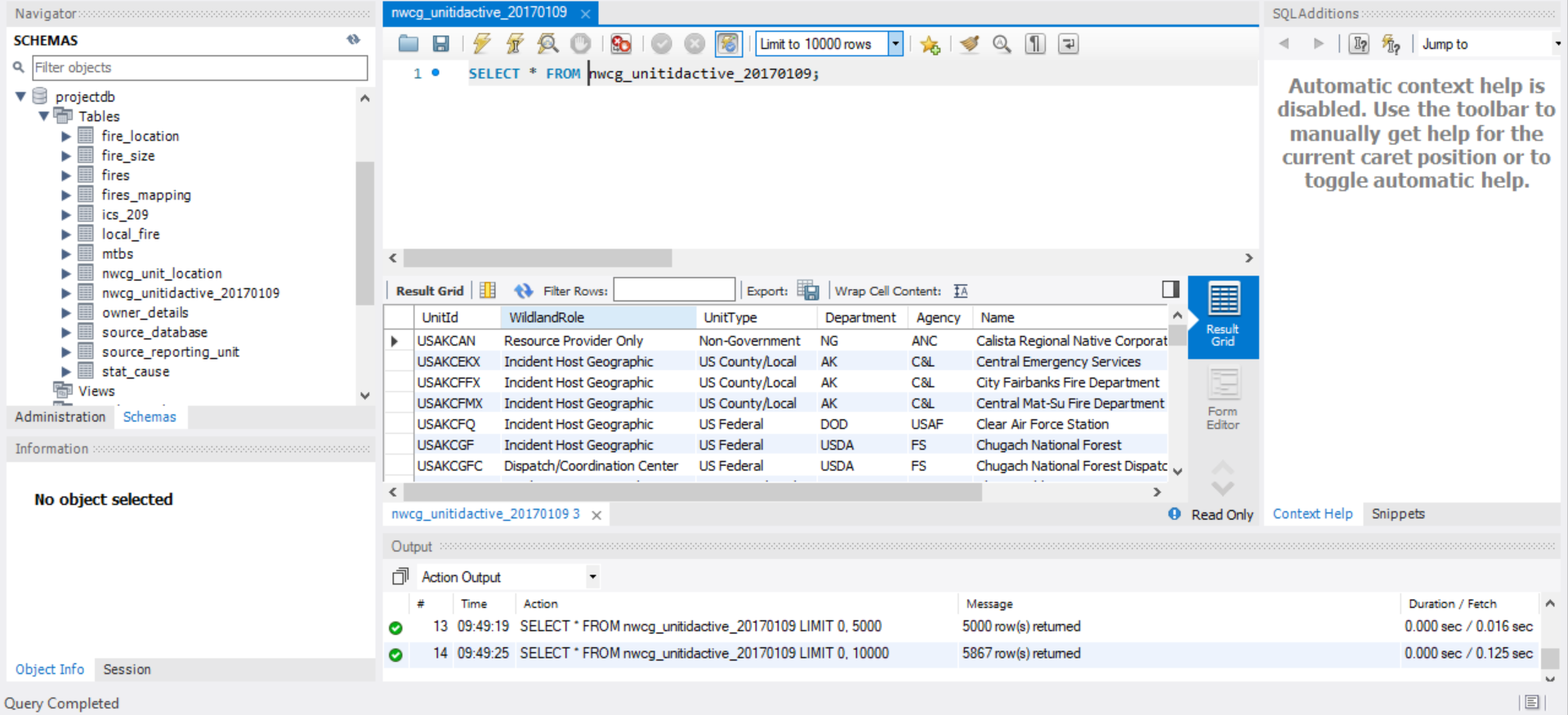


## **Local\_fire**

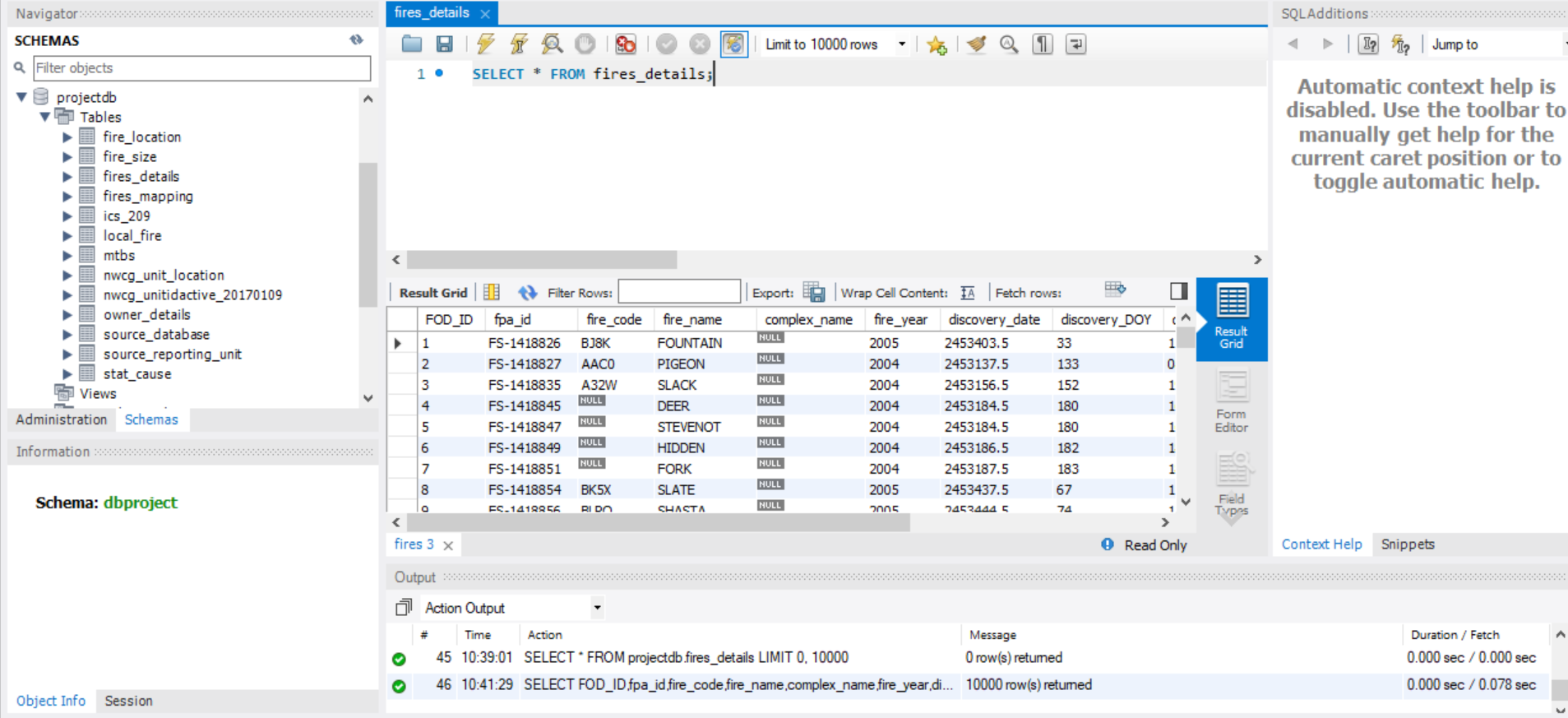
**MTBS** 

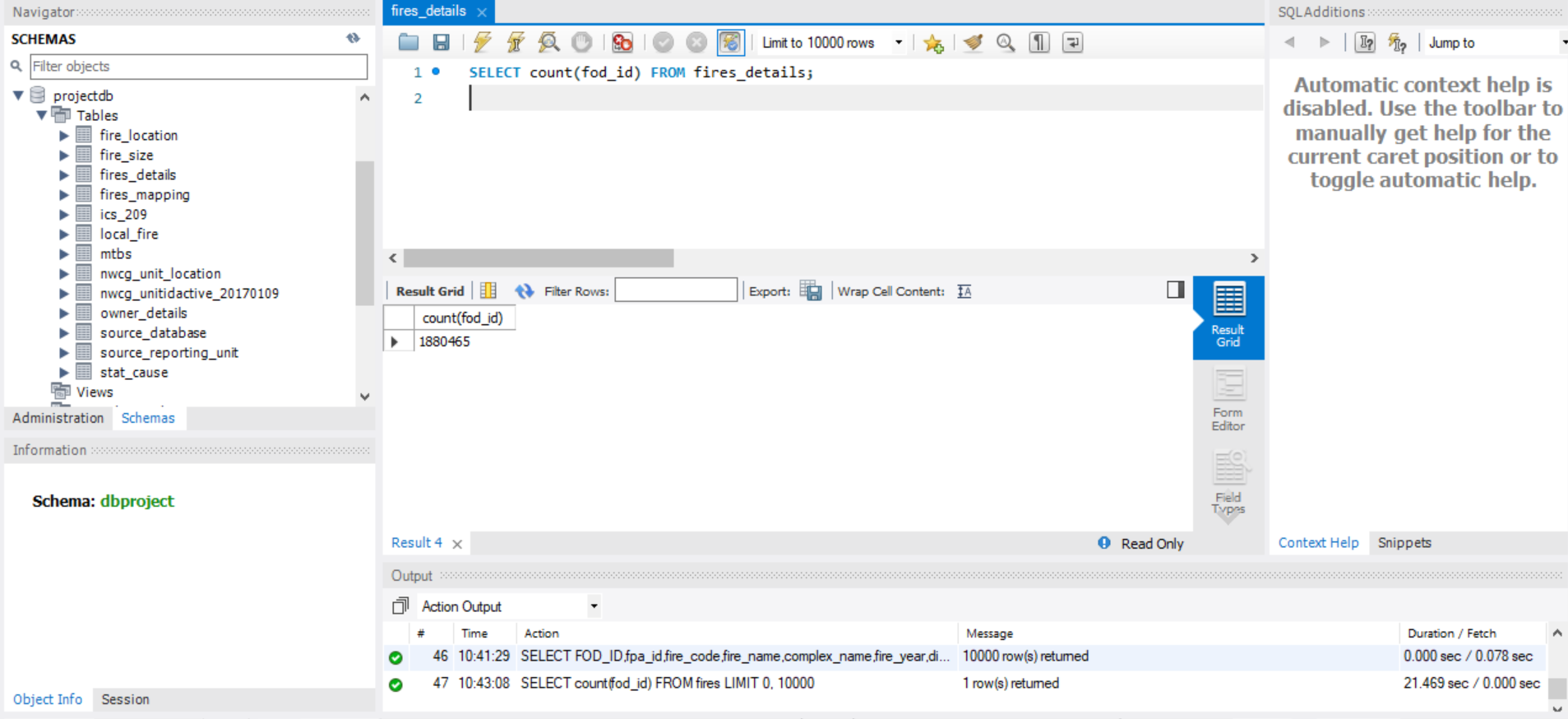
**NWCG\_Unit\_location**



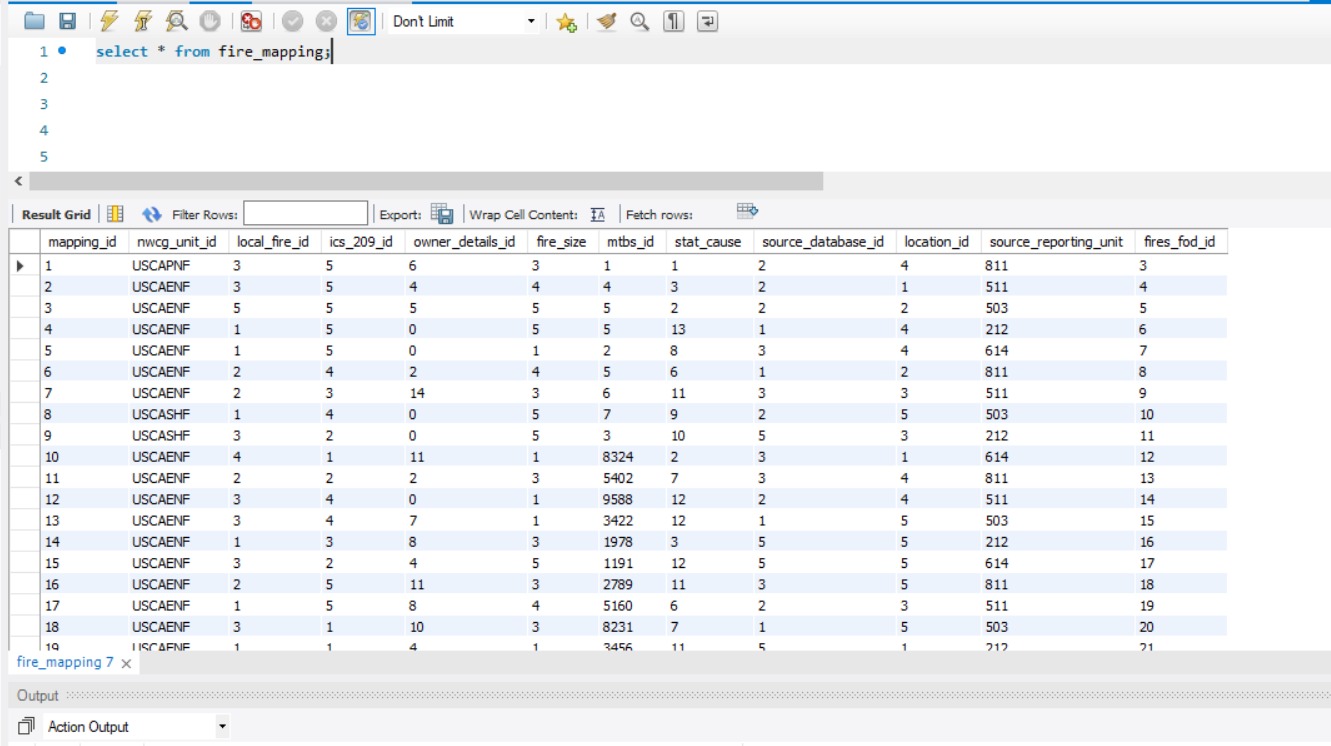
**NWCG\_unitidactive\_20170109**

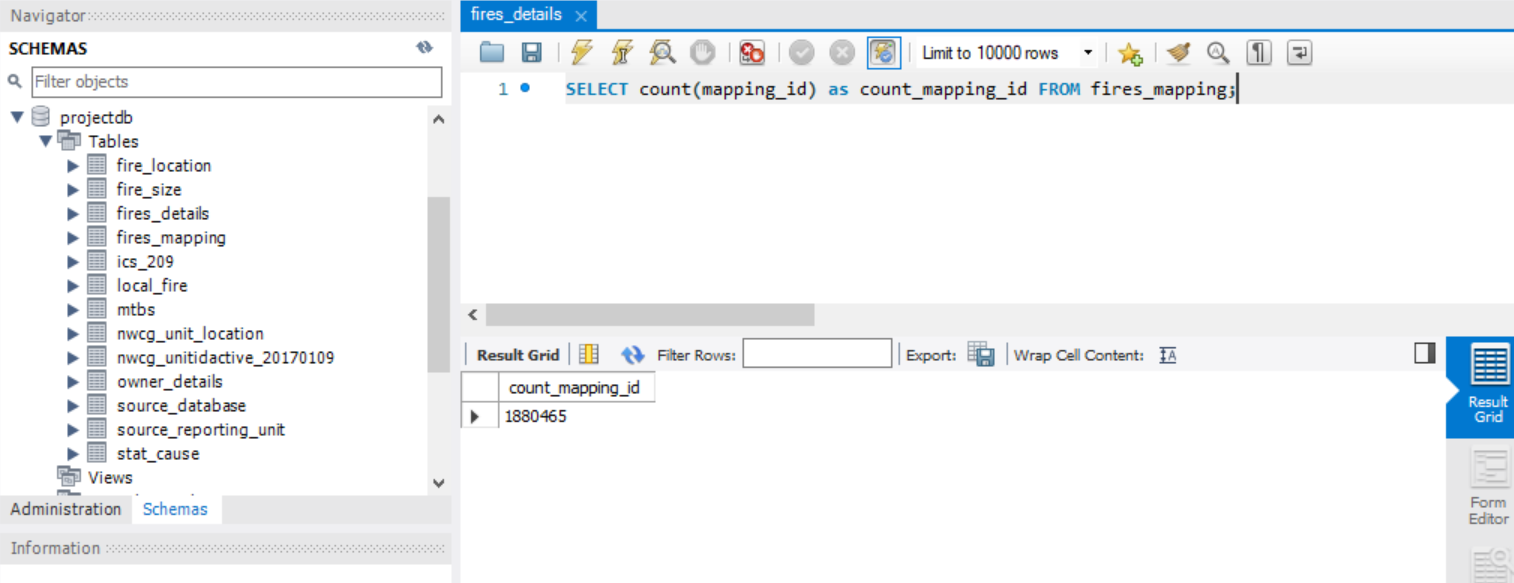
**Fire\_details**





**Fire\_mapping**

****



# 

# SQL Queries

## Query 1

### Question

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)

Notes: More the number of fires caused by debris burning in a state, higher the chance to win the fund.

Number of rows in result: 1

The state of Georgia (GA) has the best chance to win the fund as it has the highest number of fires caused by debris burning.

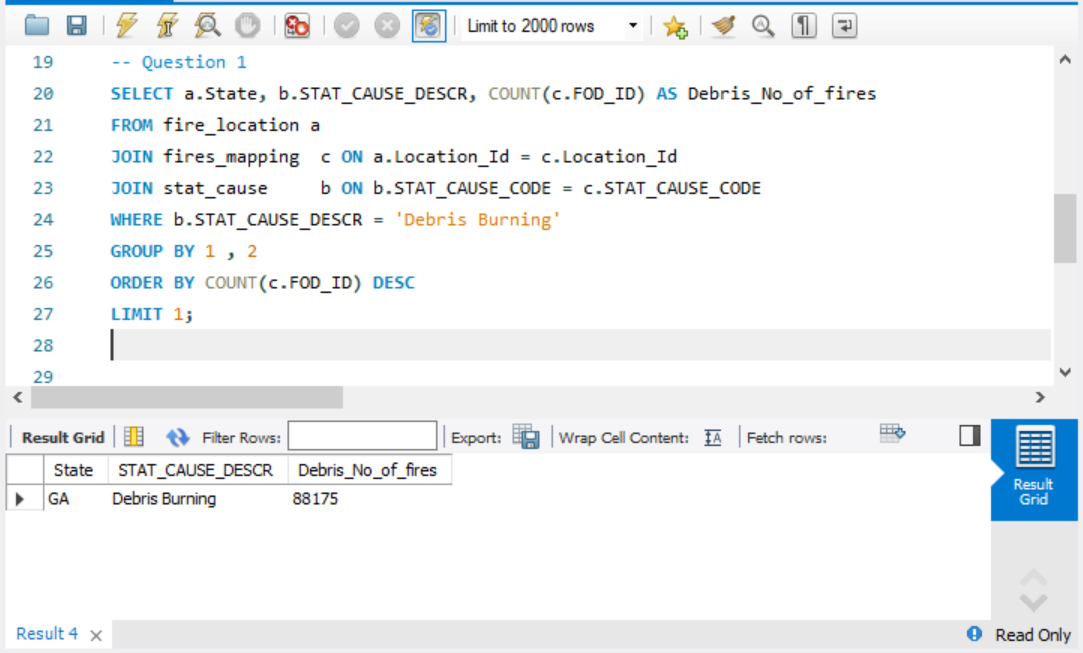
### Translation

Show me the state from fire\_location table with maximum number of fires caused by debris burning

### Clean up

Select state, STAT\_CAUSE\_DESCR, COUNT(FOD\_ID) from fire\_location where STAT\_CAUSE\_DESCR is Debris Burning, group by state, order by COUNT(FOD\_ID) in descending order, limit number of rows in result to 1

### Screen Shot of SQL Query and Results



## Query 2

### Question

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)

Notes: The restriction/ban would be more appropriate where the number of fires caused by children are high.

Number of rows in result: 3

Georgia Forestry Commission, Red Lake Agency and North Carolina Division of Forest Resources has the highest number of fires caused by children respectively. Hence, the ban would be more appropriate in these forests.

### Translation

List the top 3 forests from source\_reporting\_unit where the number of fires caused by children are the highest.

### Clean up

Select SOURCE\_REPORTING\_UNIT\_NAME, STAT\_CAUSE\_DESCR, COUNT(FOD\_ID) from source\_reporting\_unit where stat\_cause\_descr is children, order by count(FOD\_ID) in descending order, limit number of rows in result to 3

### Screen Shot of SQL Query and Results



## Query 3

### Question

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)

Notes: Natural cause of fire is Lightening while all the other causes of fire are manmade.

Number of rows in result: 1

The result obtained shows that fires caused due to human actions are more than those caused by nature. Hence, the claim made by the advocacy group cannot be supported based on data.

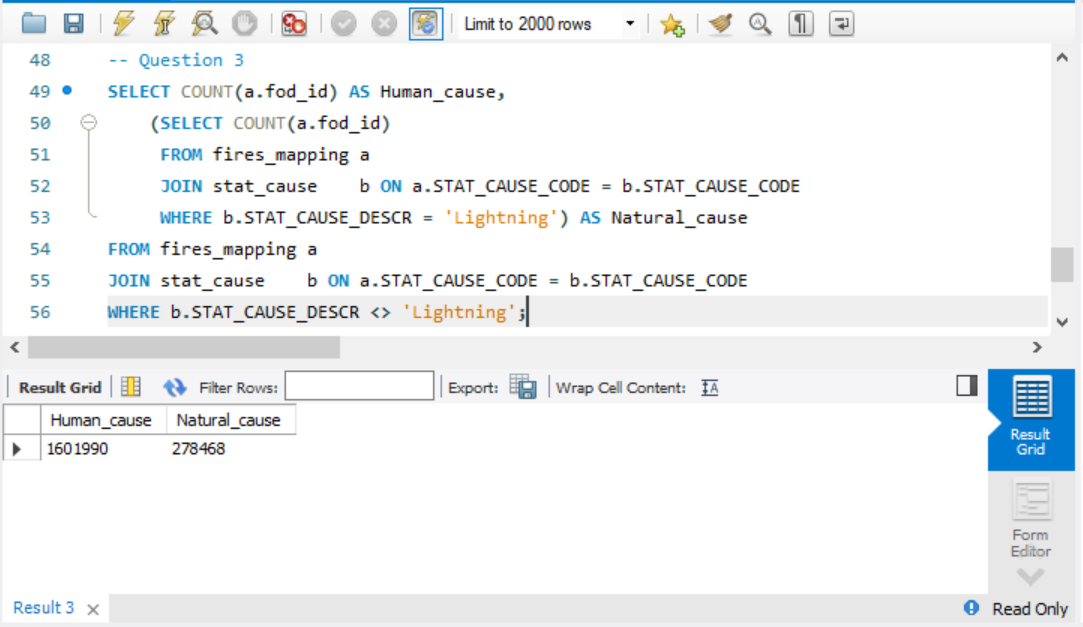
### Translation

Show me the number of fires from stat\_cause table where stat\_cause\_descr is equal to Lightening, the number of fires from stat\_cause table where stat\_cause\_descr is not equal to Lightening

### Clean up

Select count(FOD\_ID) from fires\_mapping join stat\_cause where stat\_cause\_descr is not equal to lightening as Human\_cause, count(FOD\_ID) where stat\_cause\_descr is equal to lightening as Natural\_cause

### Screen Shot of SQL Query and Results



## Query 4

### Question

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

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

Number of rows in result: 1

79357 wildfires were reported by more than one unit/agency.

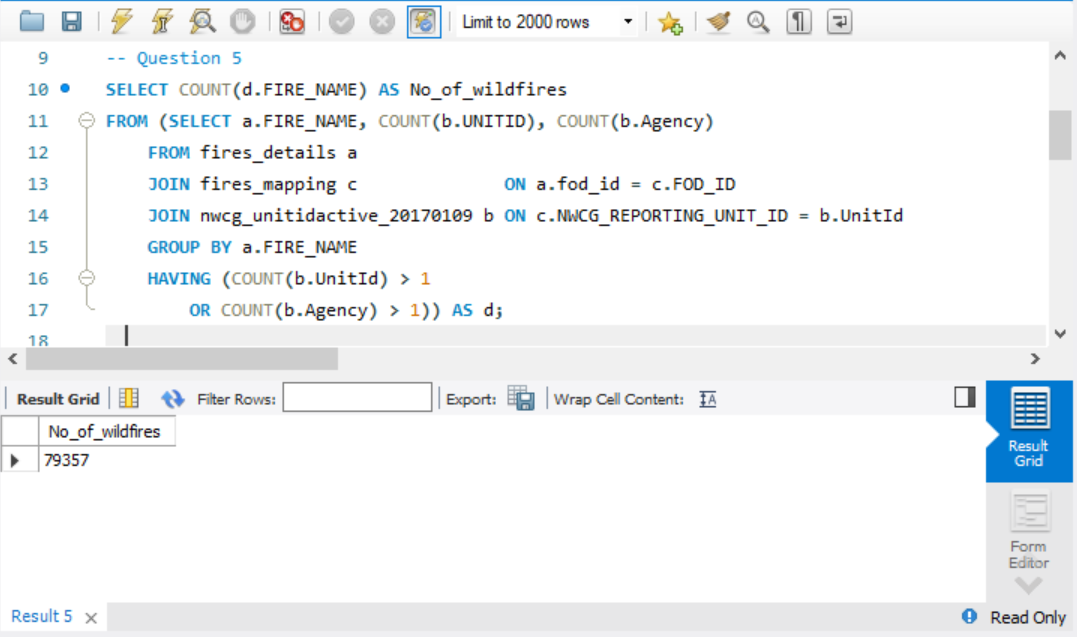
### Translation

Give me the count of fires from fires\_details table grouped by FIRE\_NAME where count of unitID is greater than 1 or count of Agency is greater than 1

### Clean up

Select count(FIRE\_NAME) from fires\_details join nwcg\_unitidactive\_20170109 group by FIRE\_NAME where count(unitID) >1 or count(Agency) > 1

### Screen Shot of SQL Query and Results



## Query 5

### Question

Which state had more fires in the second half of a calendar year than the first half of the calendar year?

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

Notes: Fires discovered in first half of the year are obtained when the discovery\_doy (day of the year on which the fire was discovered) is less than or equal to 183. Fires discovered in second half of the year are obtained when the discovery\_doy (day of the year on which the fire was discovered) is greater than 183.

Number of rows in result: 12

12 states had more fires in the second half of the calendar year than first half of the calendar year.

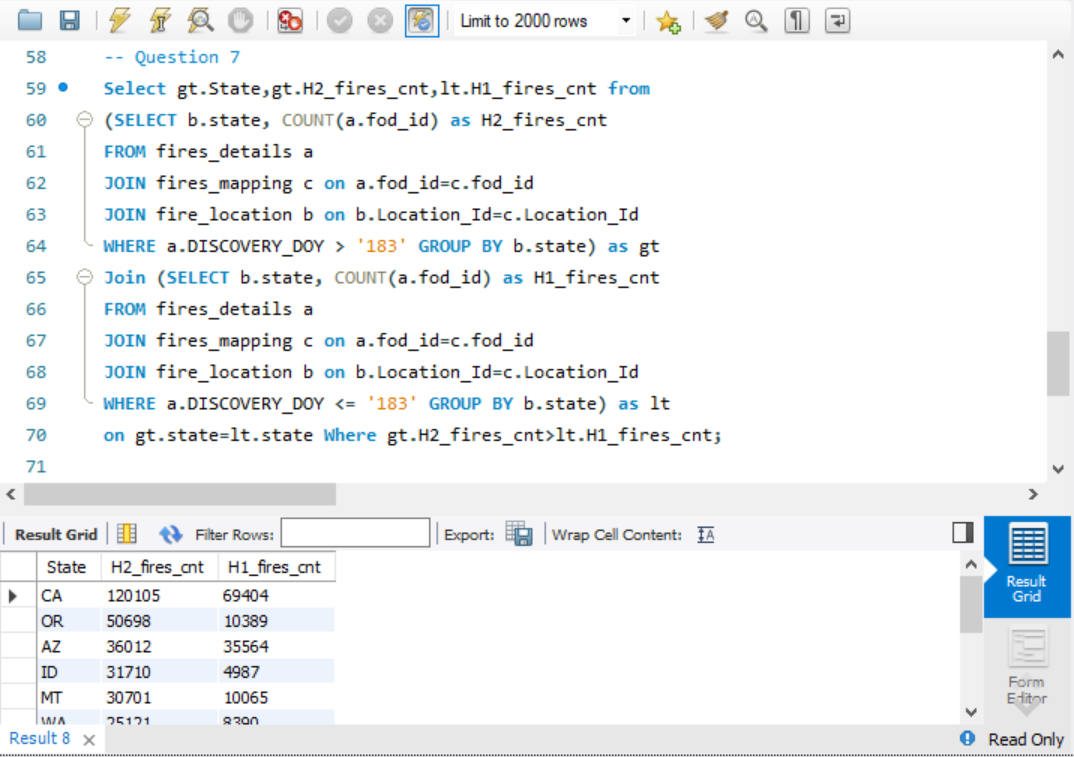
### Translation

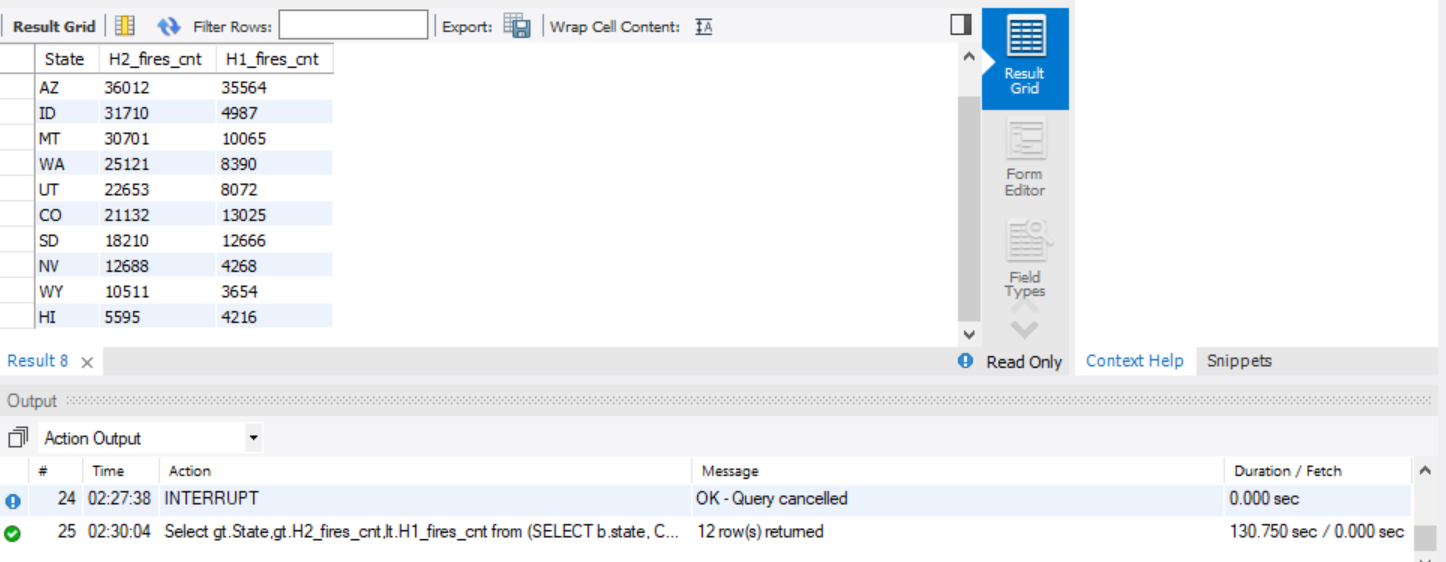
Give me the states where count(FOD\_ID) for DISCOVERY\_DOY > 183 is more than count(FOD\_ID) for DISCOVERY\_DOY <= 183 from fires\_details table grouped by states

### Clean up

Select state, (count(FOD\_ID) having DISCOVERY\_DOY >183) as H2\_fires\_cnt, (count(FOD\_ID) having DISCOVERY\_DOY <= 183) as H1\_fires\_cnt from fires\_details group by state where H2\_fires\_cnt > H1\_fires\_cnt

### Screen Shot of SQL Query and Results





## Query 6

### Question

Which forest had the most number of fires?

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

Notes: Source reporting unit name is same as the forest name.

Number of rows in result: 1

Georgia Forestry Commission had the most number of fires.

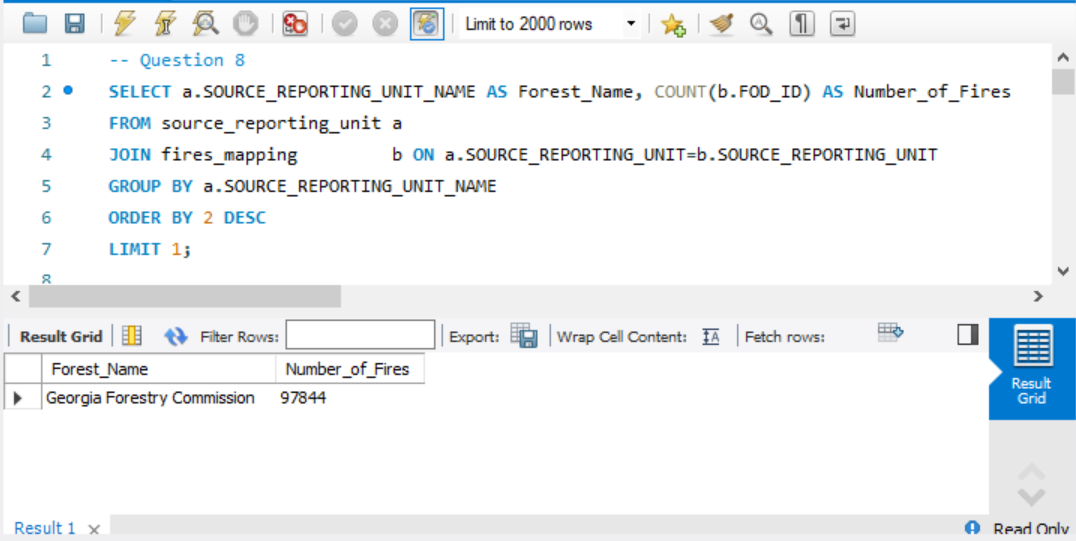
### Translation

Show me the source reporting unit name which has the highest count of fires, from source reporting unit table grouped by source reporting unit name.

### Clean up

Select SOURCE\_REPORTING\_UNIT\_NAME, count of FOD\_ID from source\_reporting\_unit group by SOURCE\_REPORTING\_UNIT\_NAME, order by count(FOD\_ID) in descending order, limit 1

### Screen Shot of SQL Query and Results



# Data Review for MongoDB

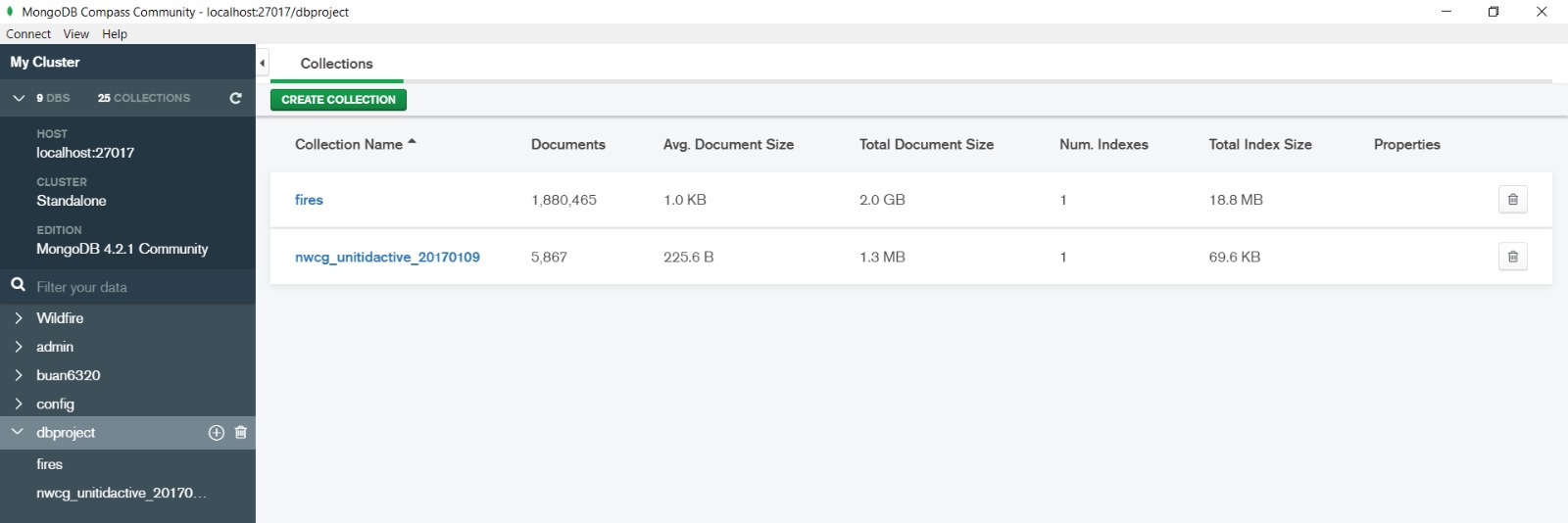
## Assumptions/Notes About Data Collections, Attributes and Relationships between Collections

As MongoDB is not a relational database system, it does not require to be in 3NF as in SQL. We have imported ‘fires’ data as a whole collection without splitting it up into different collections. We have also imported ‘nwcg\_unitidactive\_20170109’ as a separate collection in our database.

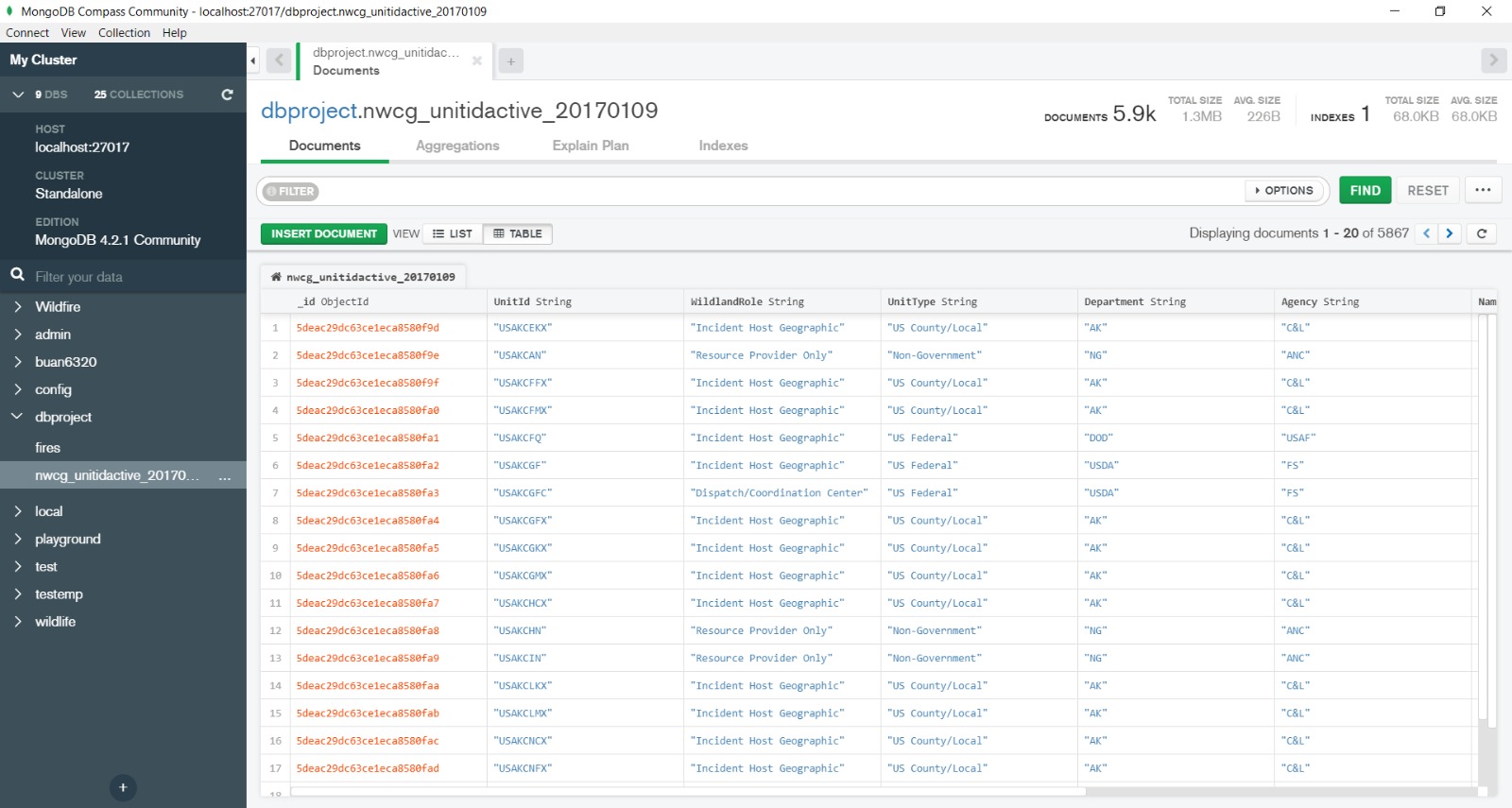
# Physical Mongo Database

## Assumptions/Notes About Data Set

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



## 



## Data in the Database

|  |  |  |
| --- | --- | --- |
| **Collection Name** | **Relationships With Other Collections (if any)** | **# of Documents in Collection** |
| fires | - | 1880465 |
| nwcg\_unitidactive\_20170109 | - | 5867 |

# MongoDB Queries/Code

## Query 1

### Question

### 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)

Number of documents in result: 1

GA has the best chance to win the fund.

### Translation

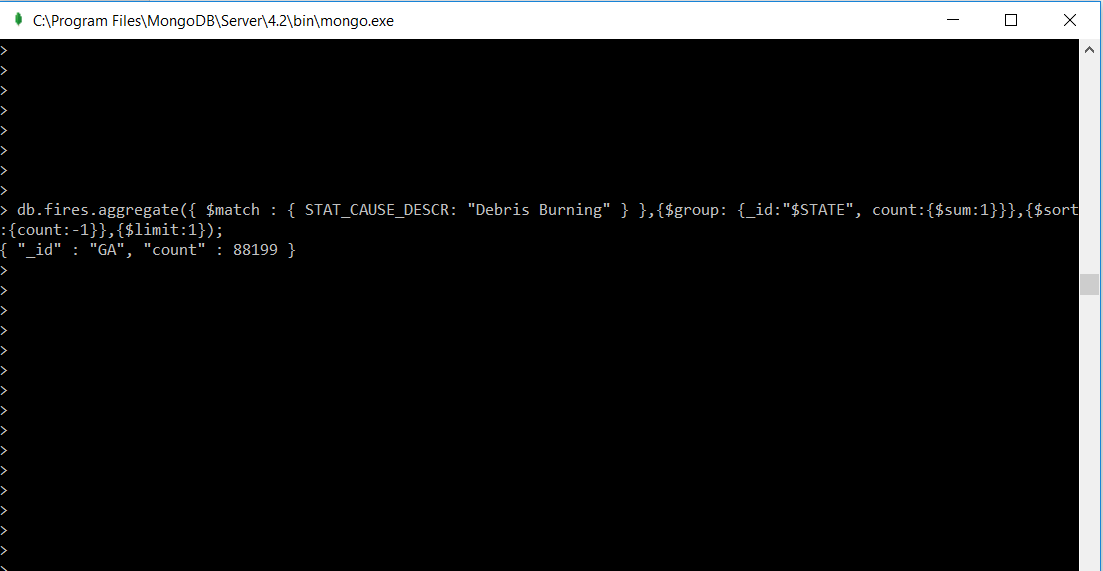
Show the state from the fires collection which has the maximum number of fires caused due to debris burning.

Apply aggregate function to Fires collection. Group the records by State and count the number of fires where the STAT\_CAUSE\_DESCR match Debris Burning. Sort the result in descending order of count. Limit the number of documents to 1.

Query

db.fires.aggregate({ $match : { STAT\_CAUSE\_DESCR: "Debris Burning" } },{$group: {\_id:"$STATE", count:{$sum:1}}},{$sort:{count:-1}},{$limit:1});

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



## Query 2

### Question

### 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)

Number of documents in result: 3

Georgia forestry Commission, Red Lake agency and New Jersey Forest Fire Division are the top 3 forests where the ban would be most appropriate.

### Translation

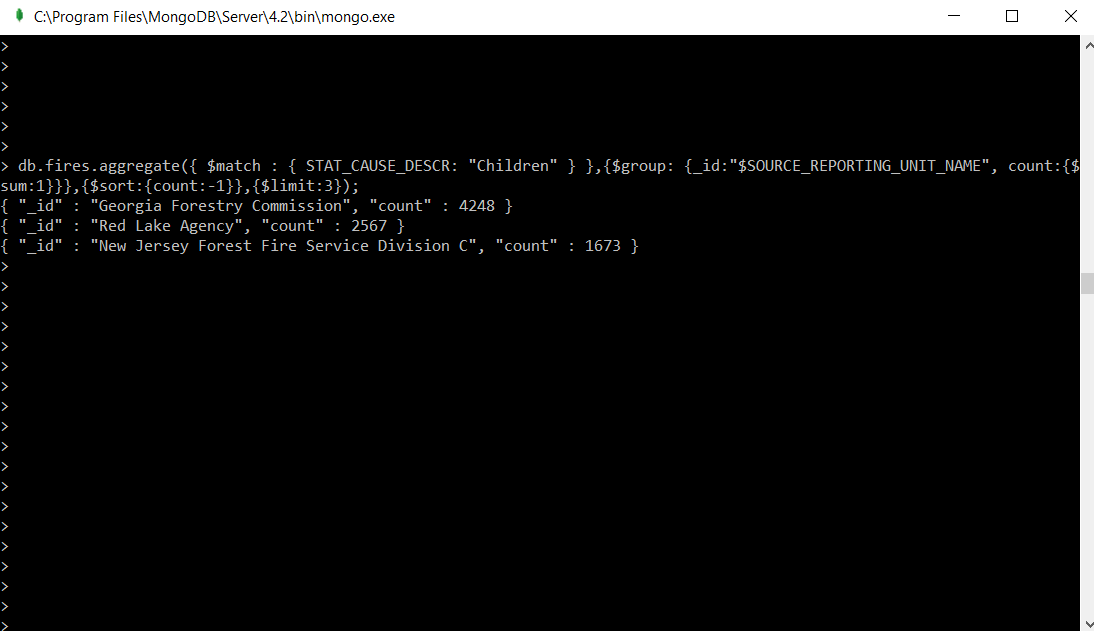
Show the top 3 states from the fires collection where STAT\_CAUSE\_DESCR of fires match Children.

Apply aggregate function to Fires collection. Group the records by SOURCE\_REPORTING\_UNIT\_NAME and count the number of fires where the STAT\_CAUSE\_DESCR match Children. Sort the results in descending order of count. Limit the number of documents to 3.

Query

db.fires.aggregate({ $match : { STAT\_CAUSE\_DESCR: "Children" } },{$group: {\_id:"$SOURCE\_REPORTING\_UNIT\_NAME", count:{$sum:1}}},{$sort:{count:-1}},{$limit:3});

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



## 

## Query 3

### Question

### 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)

Notes: Natural cause of fire is Lightening while all the other causes of fire are manmade.

Number of documents in result: 2

Number of fires caused by nature: 278468.

Number of fires caused by humans: 1601997

This does not support the claim that nature and not humans are responsible for most wildfires.

### Translation

Give the count of fires caused by lightening and count of fires caused by humans from Fires collection.

Apply aggregate function to Fires collection. Group the records by STAT\_CAUSE\_DESCR and count the number of fires where the STAT\_CAUSE\_DESCR matches Lightening as Nature else categorize as Human.

Query

db.fires.aggregate([

{

$project:

{

"STAT\_CAUSE\_DESCR":1,

"Brief":

{

$switch:

{

branches: [

{ case: { $eq : [ "$STAT\_CAUSE\_DESCR","Lightning"]}, then: "Nature"}

],

default: "Human"

}

}

}

},

{$group: {\_id: "$Brief", count: {$sum:1}}}

])

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



## Query 4

### Question

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

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

Number of documents in result: 1

Number of wildfires reported by more than one unit/agency: 76281

### Translation

Give the count of fires from Fires collection where the SOURCE\_REPORTING\_UNIT is greater than 1.

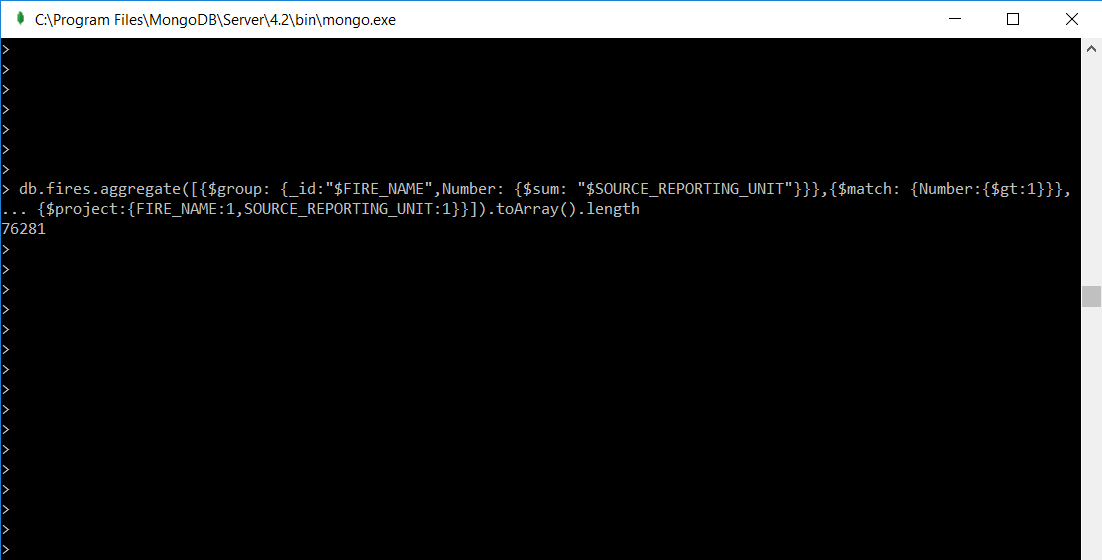
Apply aggregate function to Fires collection. Group the records by FIRE\_NAME having count of SOURCE\_REPORTING\_UNIT greater than 1. Project the count of fires.

Query

db.fires.aggregate([{$group: {\_id:"$FIRE\_NAME",Number: {$sum: "$SOURCE\_REPORTING\_UNIT"}}},{$match: {Number:{$gt:1}}},

{$project:{FIRE\_NAME:1,SOURCE\_REPORTING\_UNIT:1}}]).toArray().length

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



## Query 5

### Question

### Which state had more fires in the second half of a calendar year than the first half of the calendar year?

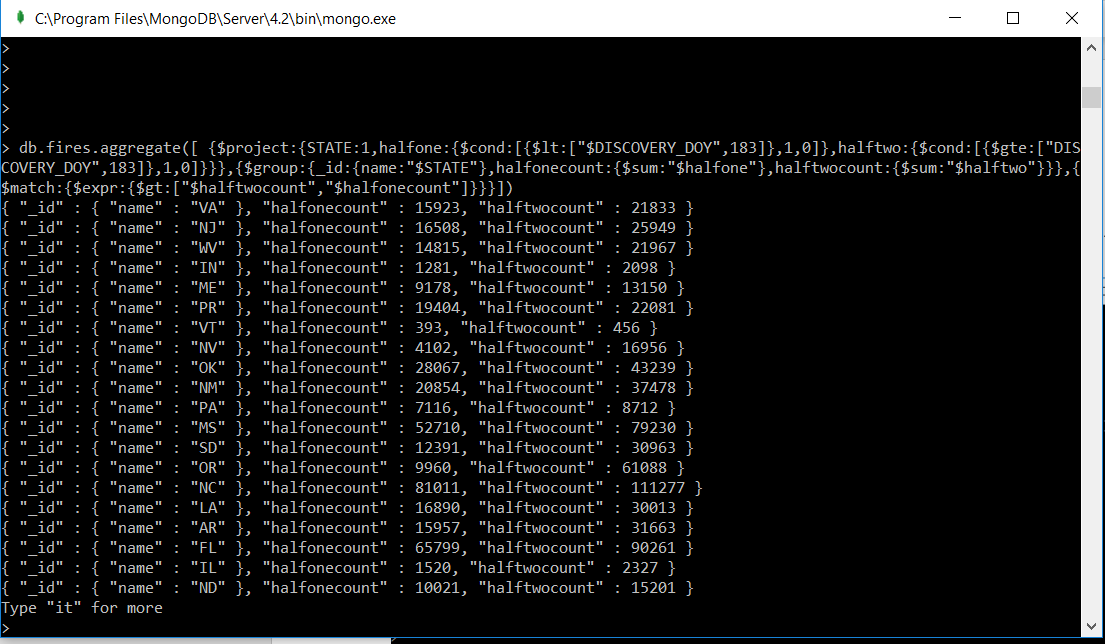
### Notes/Comments About MongoDB Query/Code and Results (Include # of Documents in Result)

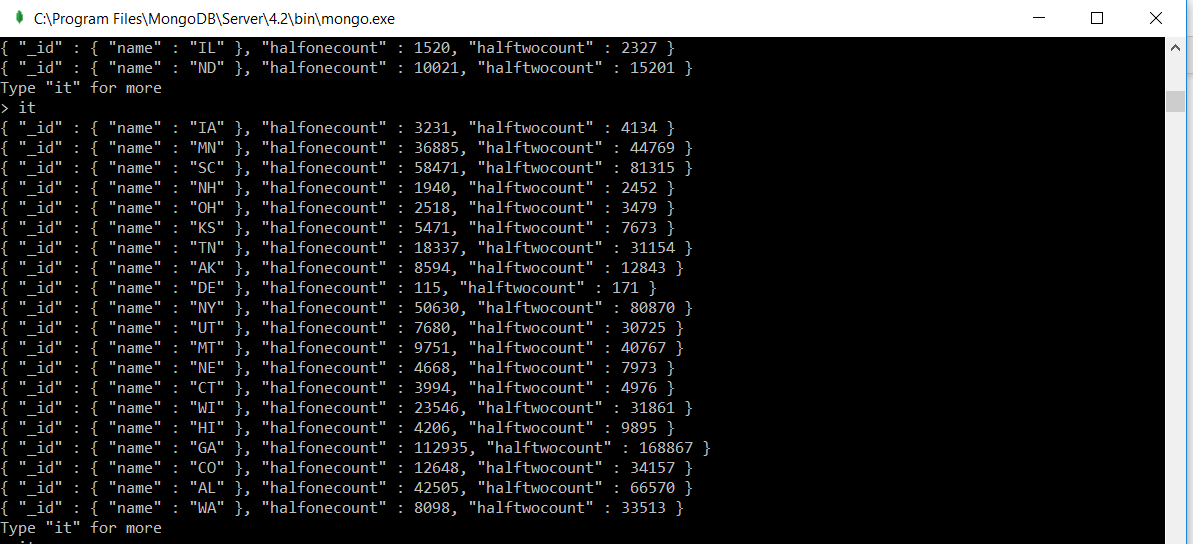
### Translation

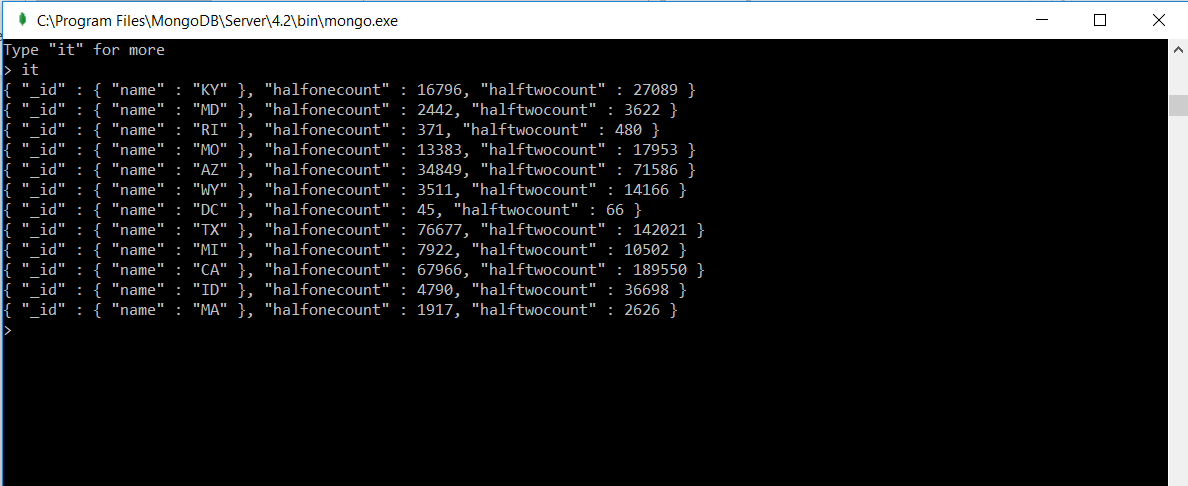
Query

db.fires.aggregate([ {$project:{STATE:1,halfone:{$cond:[{$lt:["$DISCOVERY\_DOY",183]},1,0]},halftwo:{$cond:[{$gte:["DISCOVERY\_DOY",183]},1,0]}}},{$group:{\_id:{name:"$STATE"},halfonecount:{$sum:"$halfone"},halftwocount:{$sum:"$halftwo"}}},{$match:{$expr:{$gt:["$halftwocount","$halfonecount"]}}}])

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







## Query 6

### Question

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

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

Number of documents in result: 1

Georgia Forestry Commission has the maximum number of fires reported.

### Translation

Show the SOURCE\_REPORTING\_UNIT\_NAME which has the most number of fires.

Apply aggregate to fires collection. Group by SOURCE\_REPORTING\_UNIT\_NAME, count the number of fires for each unit name. Sort the results in descending order. Limit the number of documents in result to 1.

Query

db.fires.aggregate({$group: {\_id:"$SOURCE\_REPORTING\_UNIT\_NAME", count:{$sum:1}}},{$sort:{count:-1}},{$limit:1});

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

