Group #: 3

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

BUAN 6320

Group Members

**Siddhartha Bingi**

**Alekhya Ravilla**

**Sparsh Sharma**

**Jerin Jose**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Activity** | **Siddhartha Bingi** | **Alekhya Ravilla** | **Sparsh Sharma** | **Jerin Jose** |
| 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 | X | X |  |  |
| Reviewed Report |  |  | 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/Notes About Data Entities and Relationships

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

**Overview of Data Entities and their Relationships**

|  |  |  |  |
| --- | --- | --- | --- |
| **Fact Table (Child)** | **Dimension Table (Parent)** | **Relationship** | **Assumption** |
| fire\_table  FOD\_ID (PK) | - | - |  |
| FPA\_ID (FK, NOT NULL) | source\_system\_table  FPA\_ID (PK) | One to One | For each Fire record, there is only one FPA entry and for each FPA entry there is only one Fire record |
| UnitId (FK) | nwcg\_agency\_identifier\_table  UnitId (PK) | Zero to Many | For each Fire record, there is zero or one NWCG UnitId and for each NWCG UnitId there is at least one Fire record |
| STAT\_CAUSE\_CODE (FK, NOT NULL) | Stat\_cause\_table  STAT\_CAUSE\_CODE(PK) | One to Many | For each Fire record, there is only one STAT\_CAUSE entry and for each STAT\_CAUSE entry there is at least one Fire record |
| LATITUDE (FK, NOT NULL)  LONGITUDE (FK, NOT NULL) | fire\_location\_table  LATITUDE, LONGITUDE (Composite PK) | One to Many | For each Fire record, there is only one LATITUDE/LONGITUDE (location) entry and for each LATITUDE/LONGITUDE entry there is at least one Fire record |
| MTBS\_ID (FK) | mtbs\_table  MTBS\_ID (PK) | Zero to Many | For each Fire record, there is zero or one MTBS entry and for each MTBS entry there is at least one Fire record |
| SOURCE\_REPORTING\_UNIT (FK, NOT NULL ) | source\_reporting\_table  SOURCE\_REPORTING\_UNIT (PK) | One to Many | For each Fire record, there is only one SOURCE\_REPORTING\_UNIT entry and for each FPA entry there is at least one Fire record |
| OWNER\_CODE (FK, NOT NULL) | owner\_table  OWNER\_CODE (PK) | One to Many | For each Fire record, there is only one OWNER entry and for each OWNER entry there is at least one Fire record |
| ICS\_209\_INCIDENT\_NUMBER (FK) | incident\_identifier\_table  ICS\_209\_INCIDENT\_NUMBER (PK) | Zero to Many | For each Fire record, there is zero or one ICS entry and for each ICS entry there is at least one Fire record |

|  |  |  |  |
| --- | --- | --- | --- |
| **Child Table** | **Parent Table** | **Relationship** | **Assumptions** |
| nwcg\_agency\_identifier\_table  UnitId (PK)  State (FK, NOT NULL)  Gacc (FK, NOT NULL) | nwcg\_location\_table  State (PK)  Gacc (PK) | One to Many | For each NWCG unit record, there is only one location entry and for each location there is at least one NWCG unit |
| fire\_location\_table  LATITUDE (PK)  LONGITUDE (PK)  FIPS\_CODE (FK)  FIPS\_NAME (FK) | fips\_table  FIPS\_CODE (PK)  FIPS\_NAME (PK) | Zero to Many | For each Location, there is zero or one FIPS record and for each FIPS record, there is at least one Location |

**How the Data Model has been built**

The given data has been split and grouped into logical entities such that each entity unambiguously describes its own atomic properties and any complex properties have been further split and grouped into separate child entities such that only the primary key columns are included as Foreign keys in parent entity.

1. The database has been designed using snowflake schema such that fire\_table is the fact table and other tables are the dimension tables.
2. Each entry in fire\_table can be uniquely identified by FOD-ID and all the other references to dimension tables are represented using Foreign Keys.
3. All the individual measurements of fire\_table are defined as attributes and complex units are separated into dimension tables.
4. List of dimension tables: source\_system\_table, source\_reporting\_table, nwcg\_agency\_identifier\_table, nwcg\_location\_table, incident\_identifier\_table, owner\_table, mtbs\_table, fire\_location\_table, fips\_table, stat\_cause\_table

**1st Normal Form:**

Rule 1: Each table cell should contain a single value

Rule 2: Each record needs to be unique

* fire\_table has been checked for multi-values and found none. Hence rule 1 is satisfied.
* FOD\_ID is uniquely identifies each row - hence made Primary key

**2nd Normal Form:**

Rule 1: Must be in 1st Normal Form

Rule 2: No Partial dependencies

* Model is already in 1ST Normal Form hence Rule 1 satisfied
* Columns SOURCE\_SYSTEM\_TYPE, SOURCE\_SYSTEM partially depend on FPA\_ID. Hence separated into source\_system\_table with FPA\_ID as PK and made FK in fire\_table
* Columns Agency, Wildlandrole, Unittype, Department, Code, Name, Parent, State, Gacc, GeographicArea, Country partially depend on UnitId. Hence separated into nwcg\_identifier\_table with UnitId as PK. But columns GeographicArea, Country are partially dependent on State and Gacc columns, hence further broke down the entity to form a new entity – nwcg\_location\_table with columns Gacc, GeographicArea, State, Country – State and Gacc made Composite PK and made FK(s) in nwcg\_identifier\_table.
* Column STAT\_CAUSE\_DESCR partially depends on STAT\_CAUSE\_CODE hence separated into stat\_cause\_table with STAT\_CAUSE\_CODE as PK.
* Columns STATE, FIPS\_CODE, FIPS\_NAME, COUNTY partially depend on LATITUDE and LONGITUDE columns hence separating them into fire\_location\_table with LATITUDE and LONGITUDE as PK. Further COUNTY partially depends on FIPS\_CODE and FIPS\_NAME, hence separating them into fips\_table with FIPS\_CODE and FIPS\_NAME as PK and FK in fire\_location\_table.
* Columns MTBS\_FIRE\_NAME partially depends on MTBS\_ID hence separating them into mtbs\_table with MTBS\_ID as PK
* Columns OWNER\_DESCR partially depends on OWNER\_CODE hence separating them into owner\_table with OWNER\_ID as PK
* Columns SOURCE\_REPORTING\_UNIT\_NAME partially depends on SOURCE\_REPORTING\_UNIT hence separating them into source\_reporting table with SOURCE\_REPORTING\_UNIT as PK
* Columns ICS\_209\_NAME partially depends on ICS\_209\_INCIDENT\_NUMBER hence separating them into incident\_identifier\_table with ICS\_209\_INCIDENT\_NUMBER as PK
* After performing the above steps, the entire model is in 2nd Normal Form

**3rd Normal Form**

Rule 1: Must be in 2nd Normal Form

Rule 2: No Transitive Dependencies

* The model is already in 2nd Normal Form satisfying Rule 1
* Columns NWCG\_REPORTING\_UNIT\_ID, NWCG\_REPORTING\_AGENCY and NWCG\_REPORTING\_UNIT\_NAME have transitive dependency on UnitId, Agency and Name. Hence removed these columns to satisfy Rule 2. With this, the model is in 3rd Normal Form

**Include reasons why the data model is in 3NF**

* Each table defines unambiguously specific functionality
* Every table has a primary key that uniquely identifies a row and each table contains only atomic values and refers to complex/group values through foreign keys. These complex/ group values are further divided into tables with appropriate Primary Keys
* Every column in each table is completely dependent on primary key after removing all the partial dependencies as explained in the earlier section
* All the transitive dependencies have been removed as explained in the earlier section

## Entity-Relationship Diagram

A screenshot of a social media post

Description automatically generated

# Physical Database

## Assumptions/Notes About Data Set

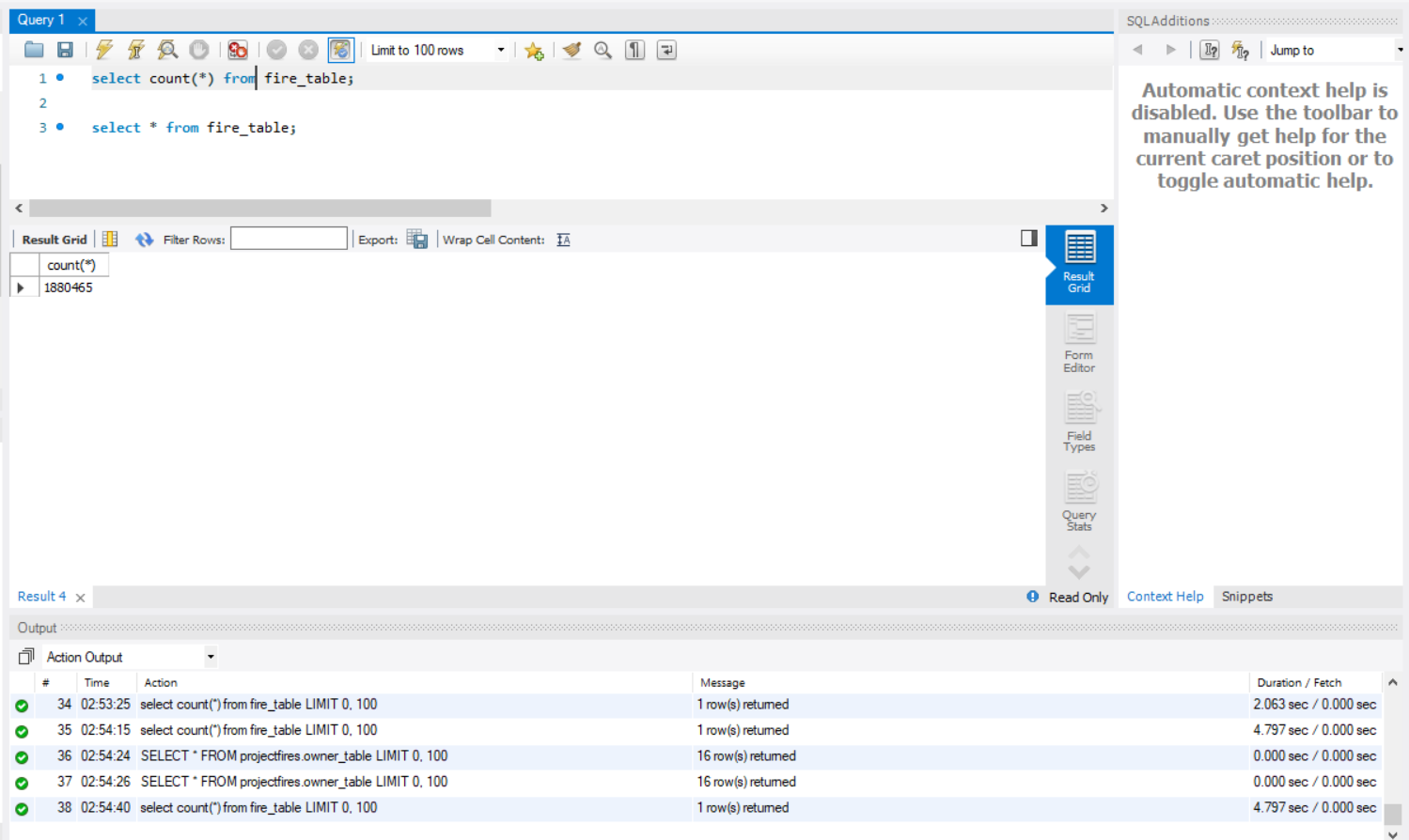
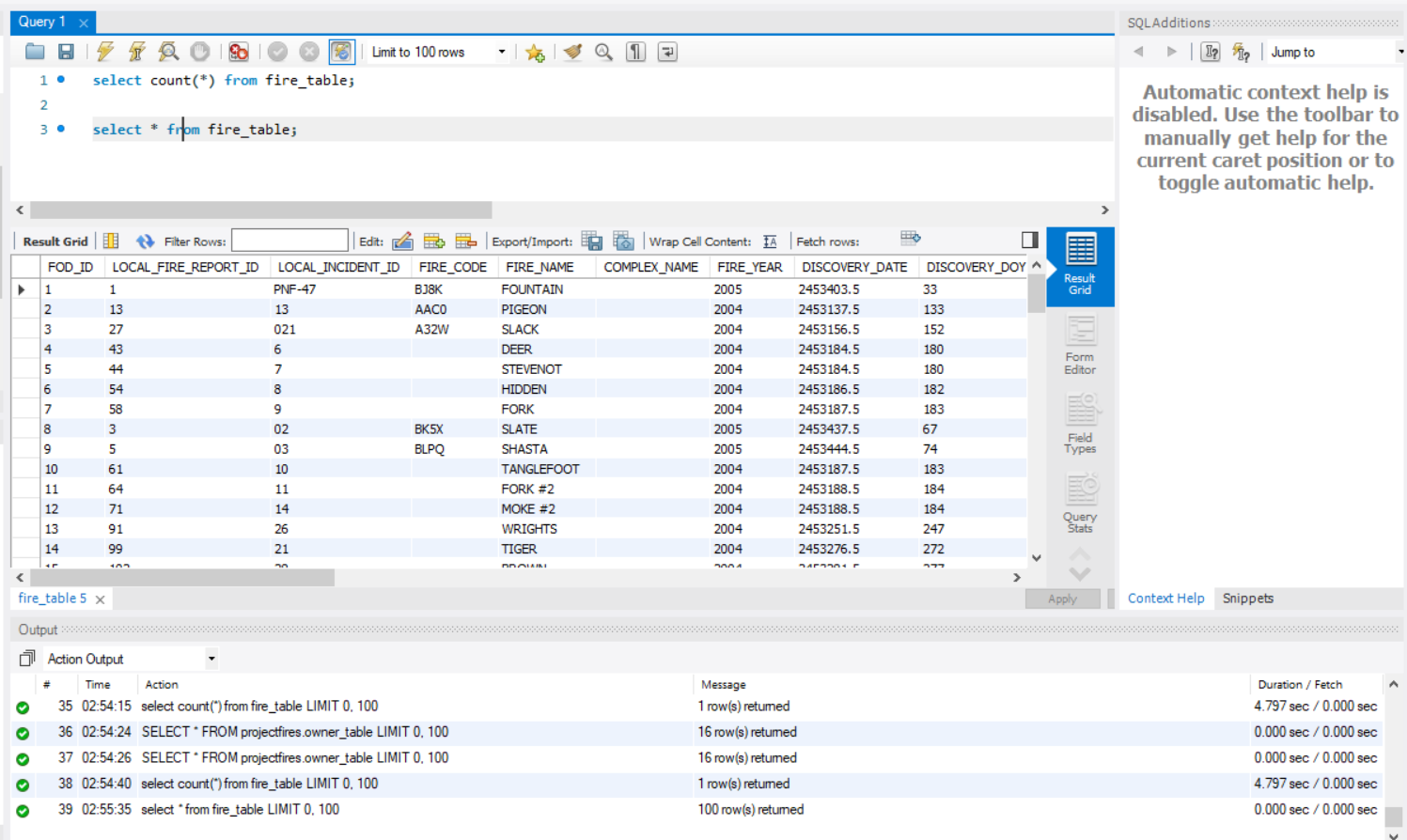
Include any assumptions made about data such as empty fields, sparse data, bad data, etc.

Data Handling/ Cleanup

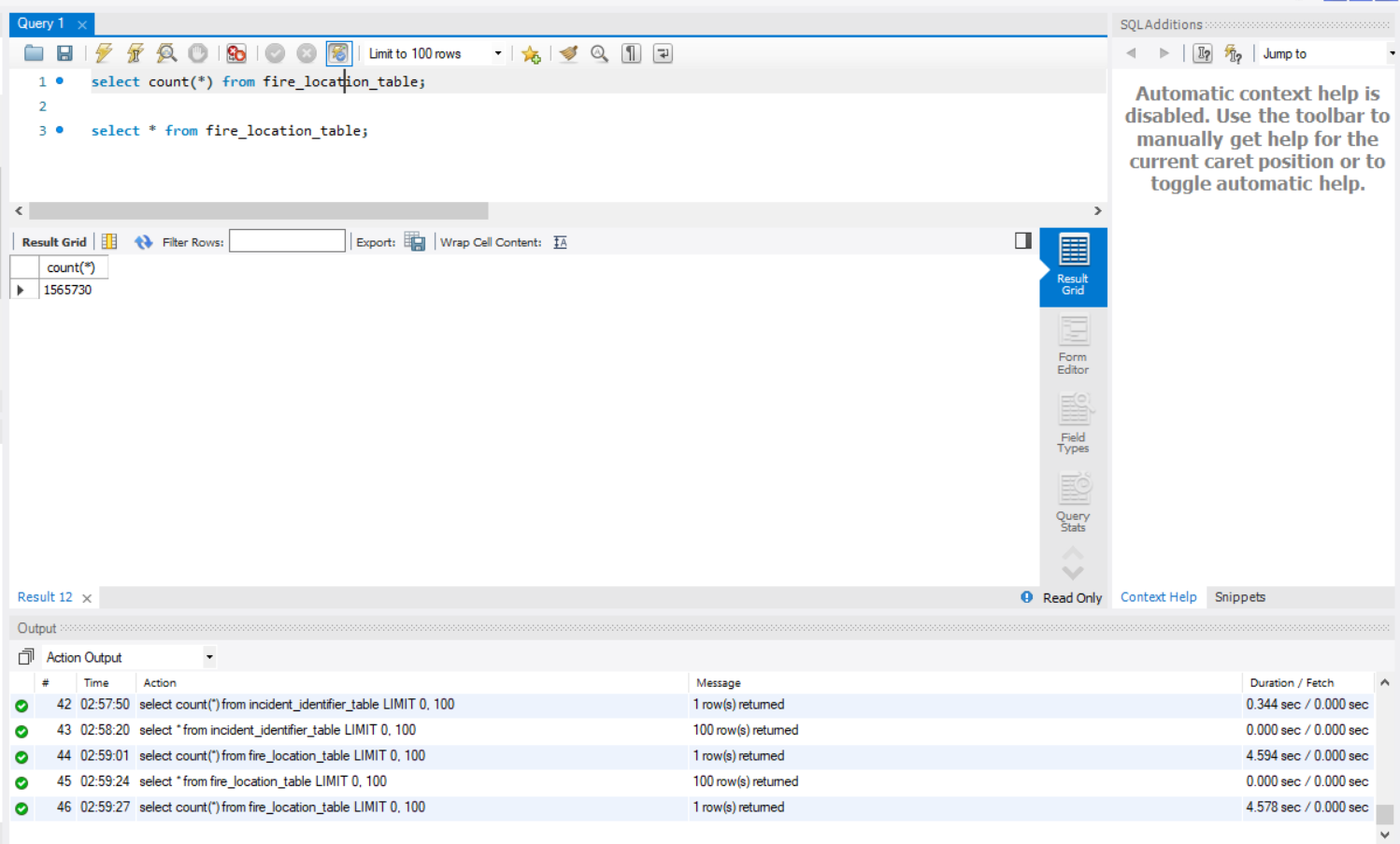
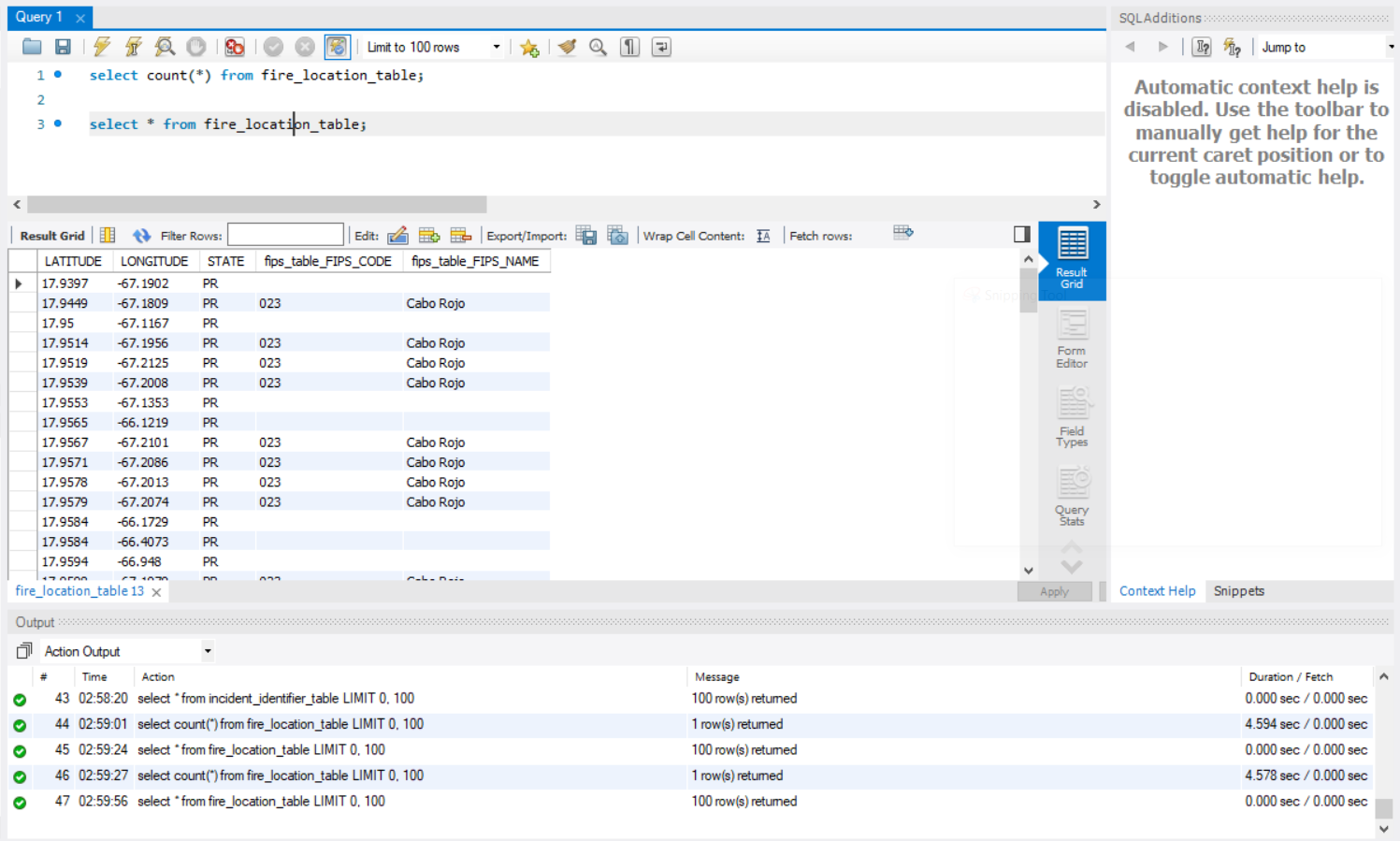
* The redundant columns in source data have been merged. UnitId and NWCG\_REPORTING\_UNIT\_ID, Agency and NWCG\_REPORTING\_AGENCY, NWCG\_REPORTING\_UNIT\_NAME and Name are such cases
* The source data has bad data in the form: lower/upper case letters interchangeably used. Such rows are grouped by code/id and merged into one record. One such case is “National Forests in Texas” and “National Forests IN Texas”
* Redundant data has been removed to retain the 3NF form of data model. Like in case of FIPS\_TABLE, for unique FIPS\_NAME, FIPS\_CODE, there were multiple COUNTY values
* Missing Values and duplicate data have been removed

## Screen shot of Physical Database objects

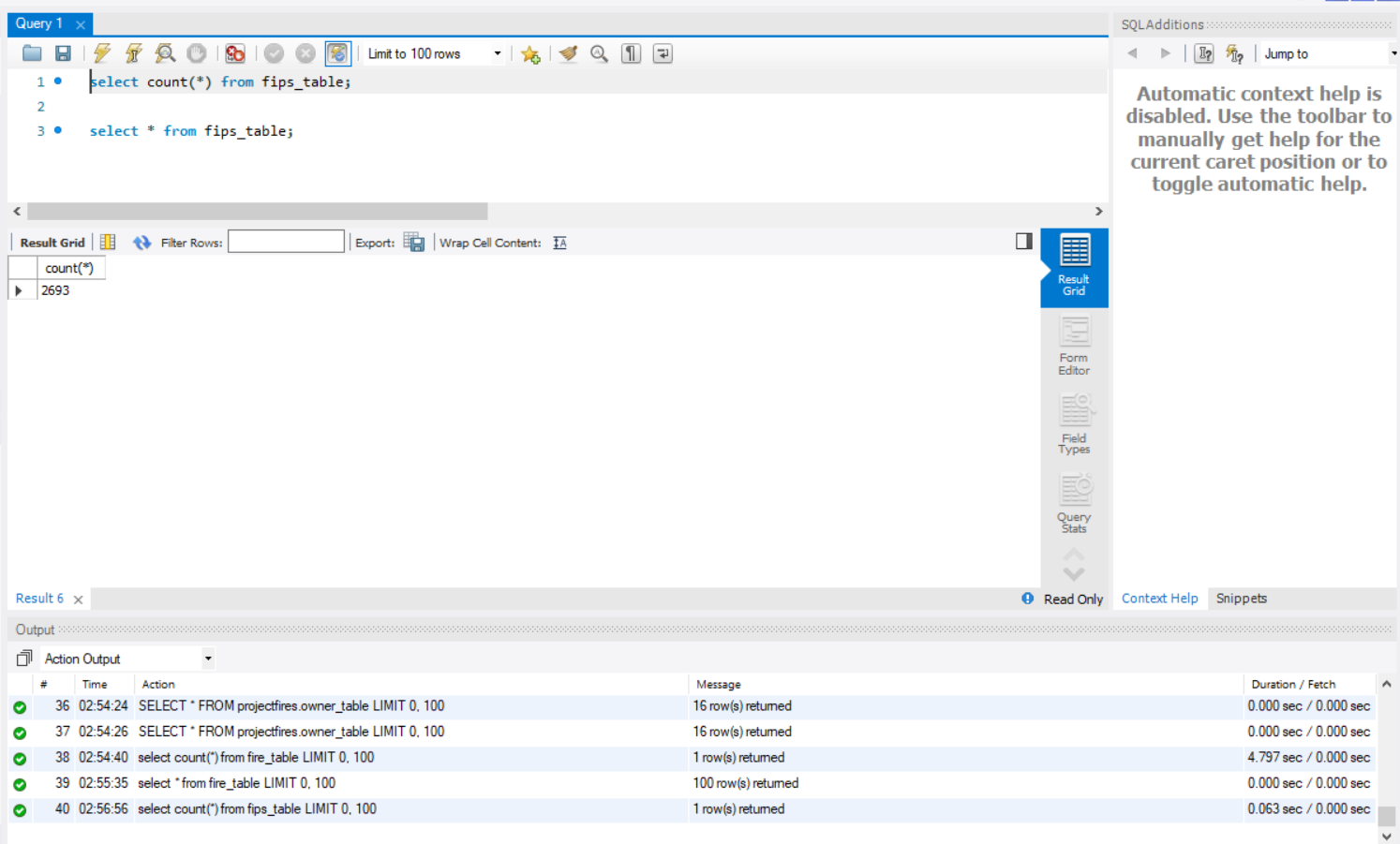
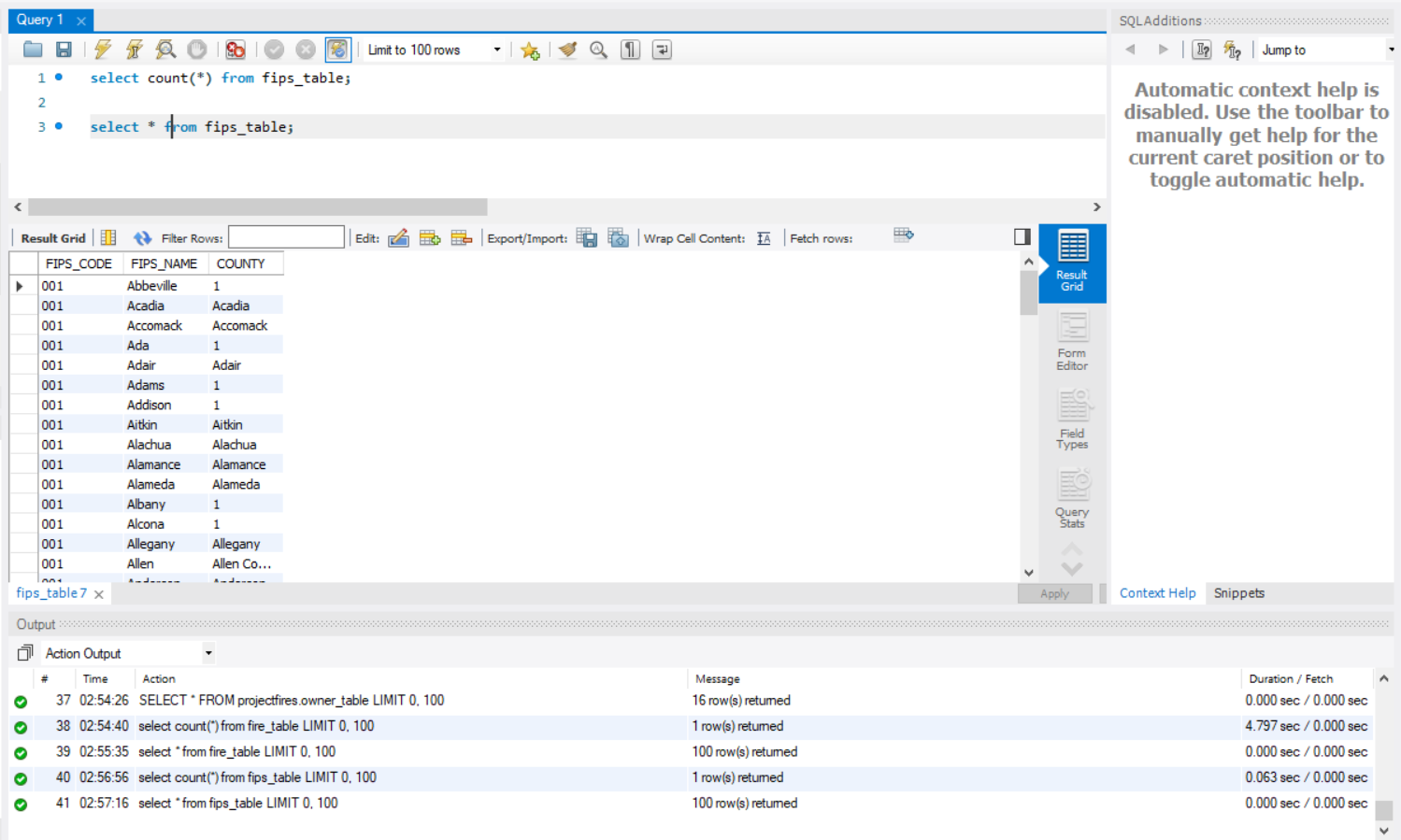
FIRE\_TABLE:



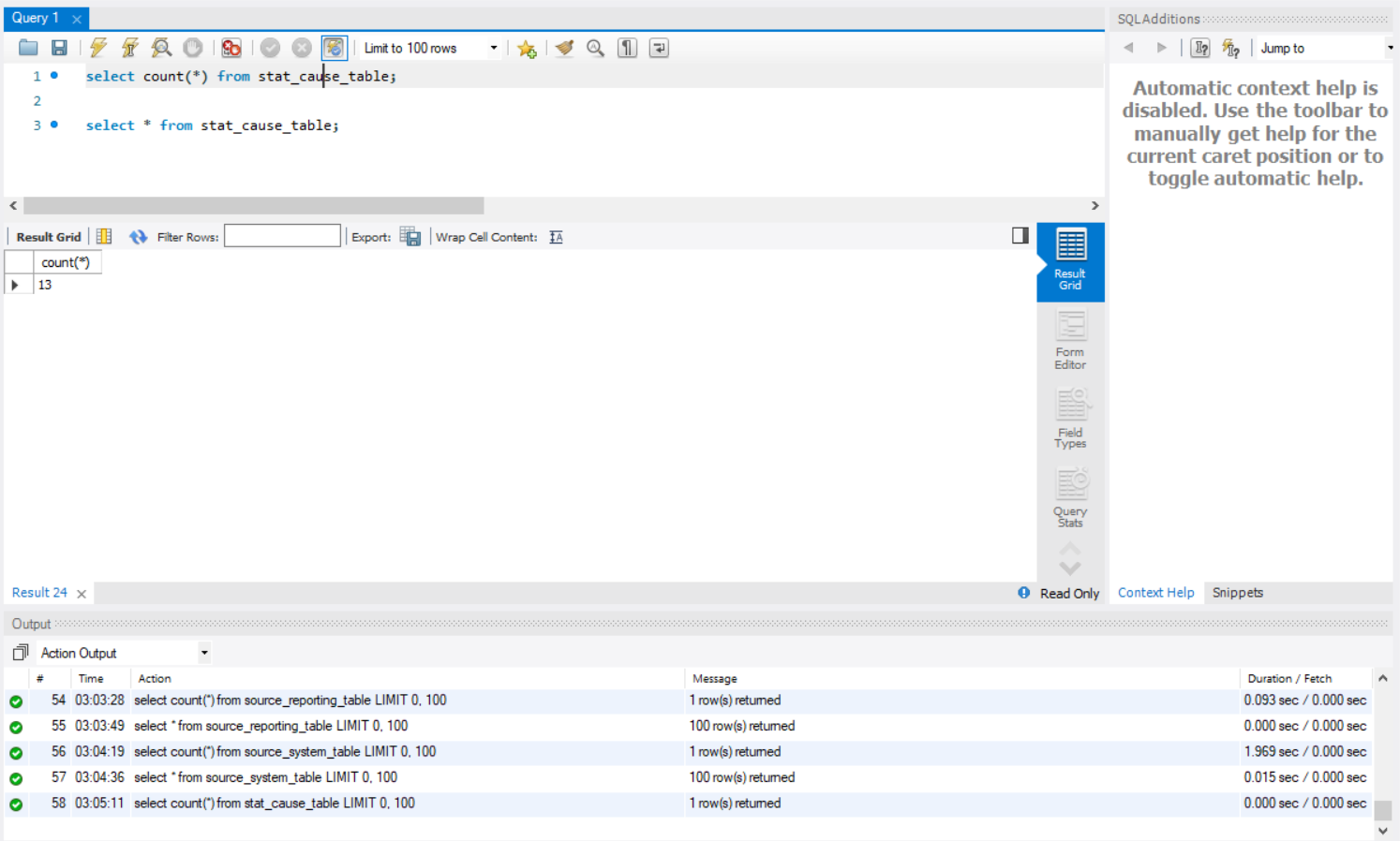
FIRE\_LOCATION\_TABLE:



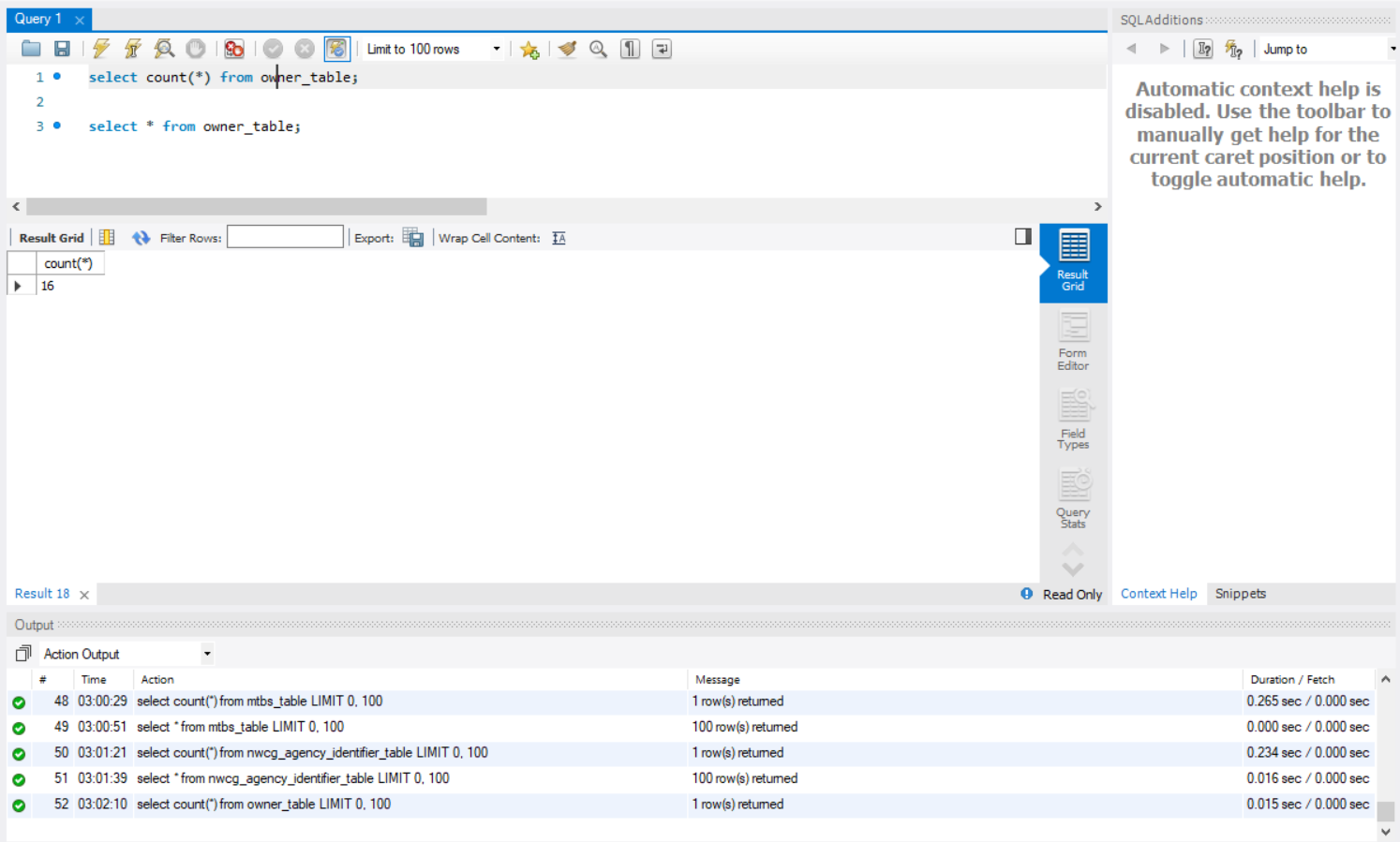
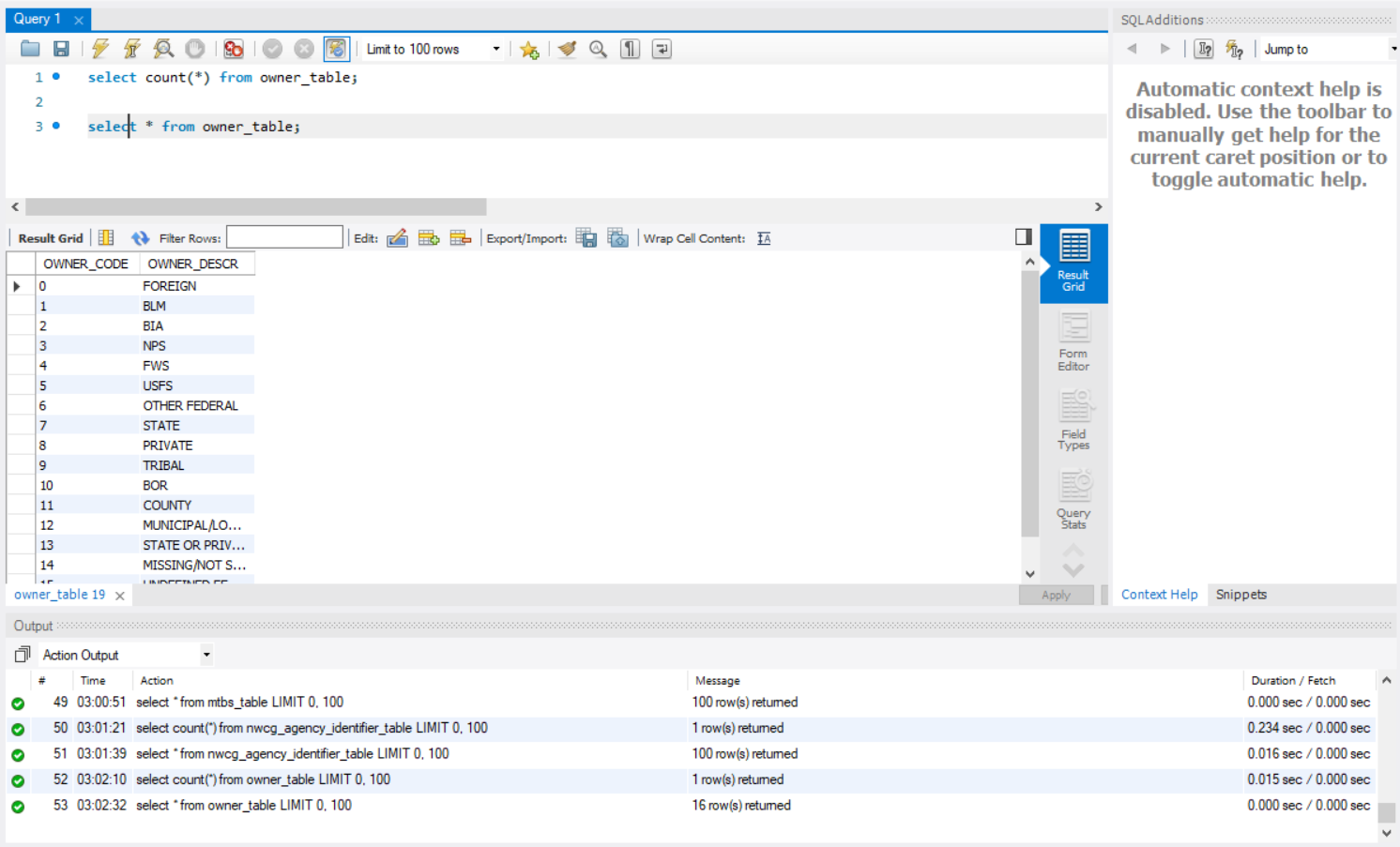
FIPS\_TABLE:



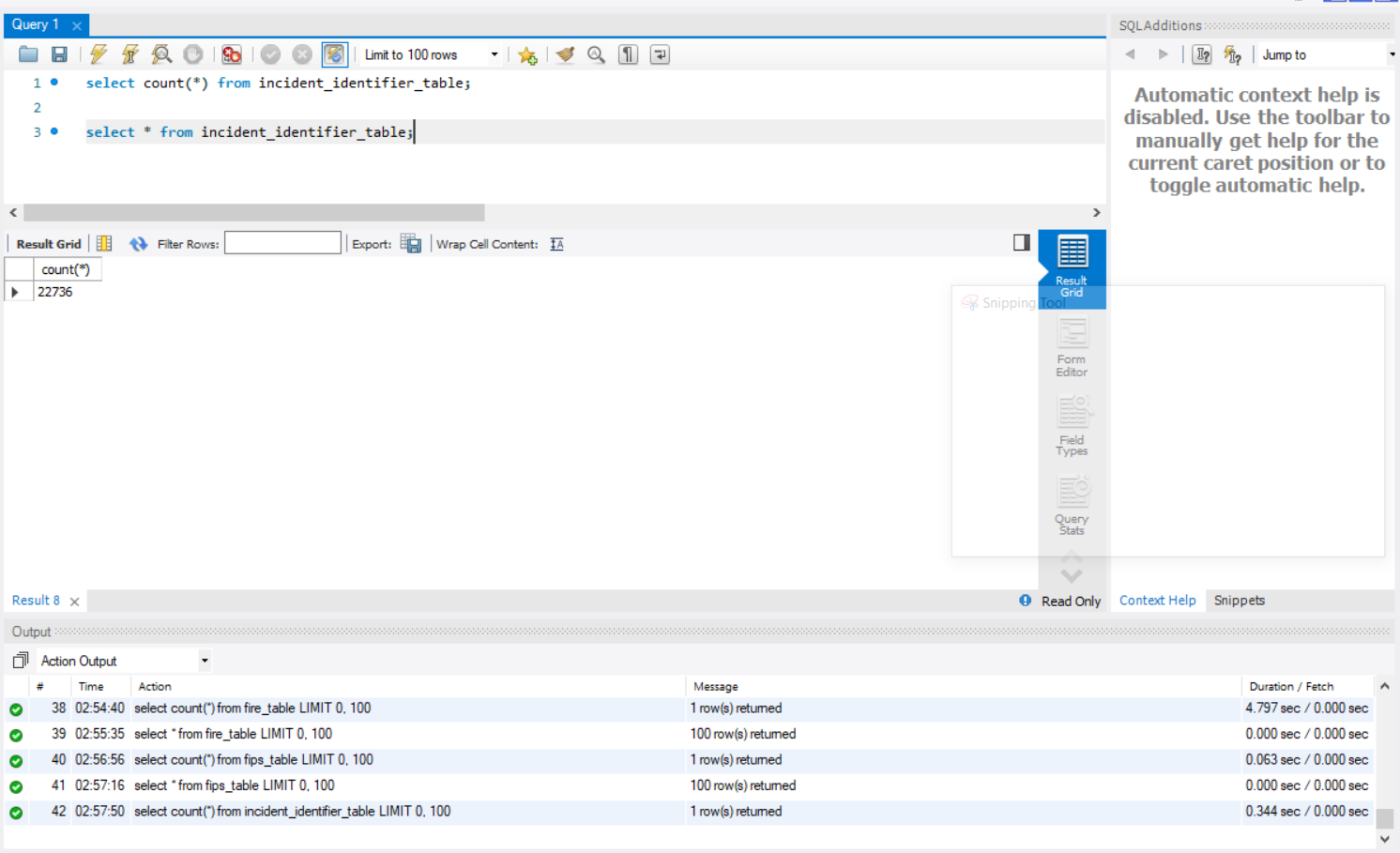
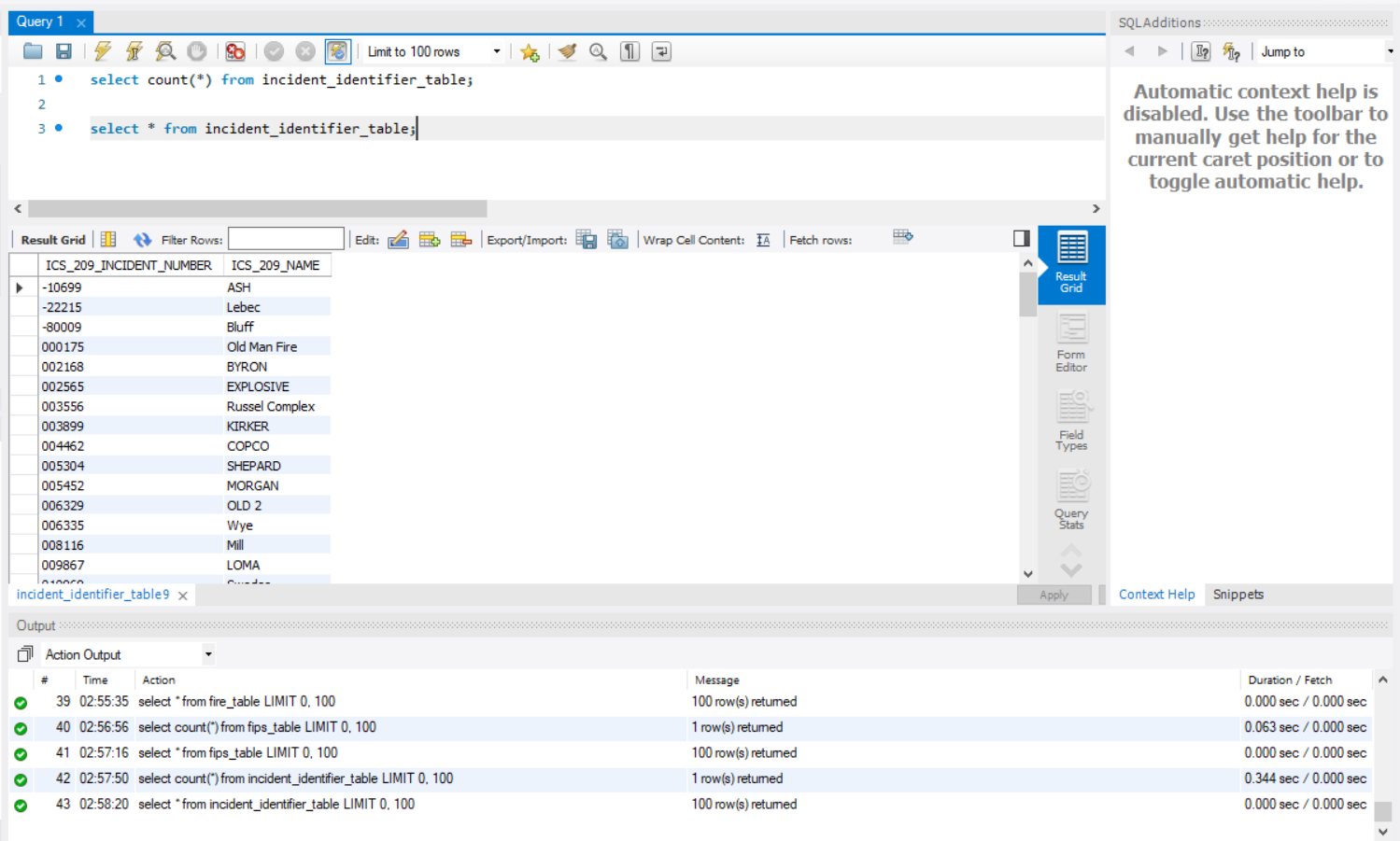
STAT\_CAUSE\_TABLE:



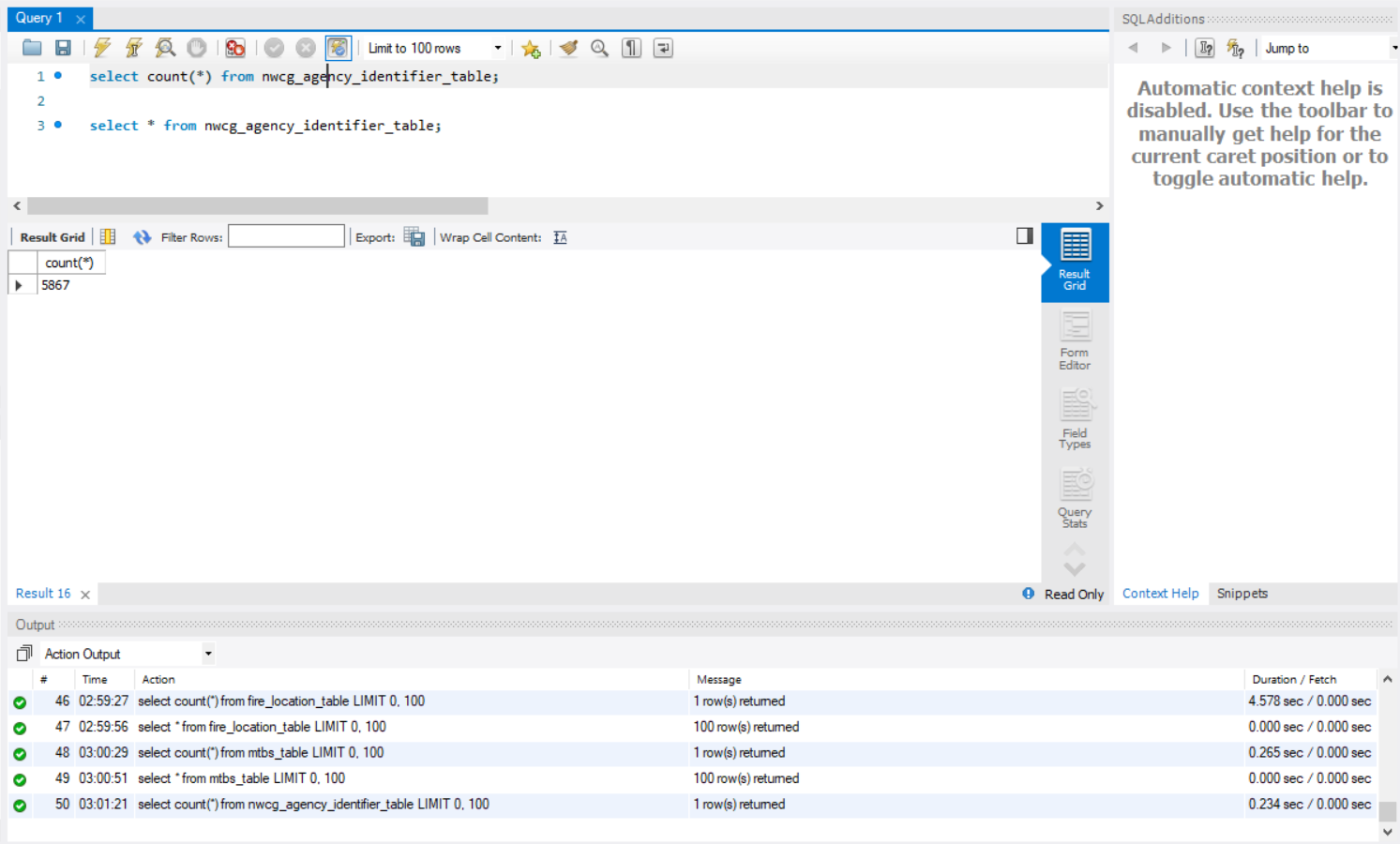
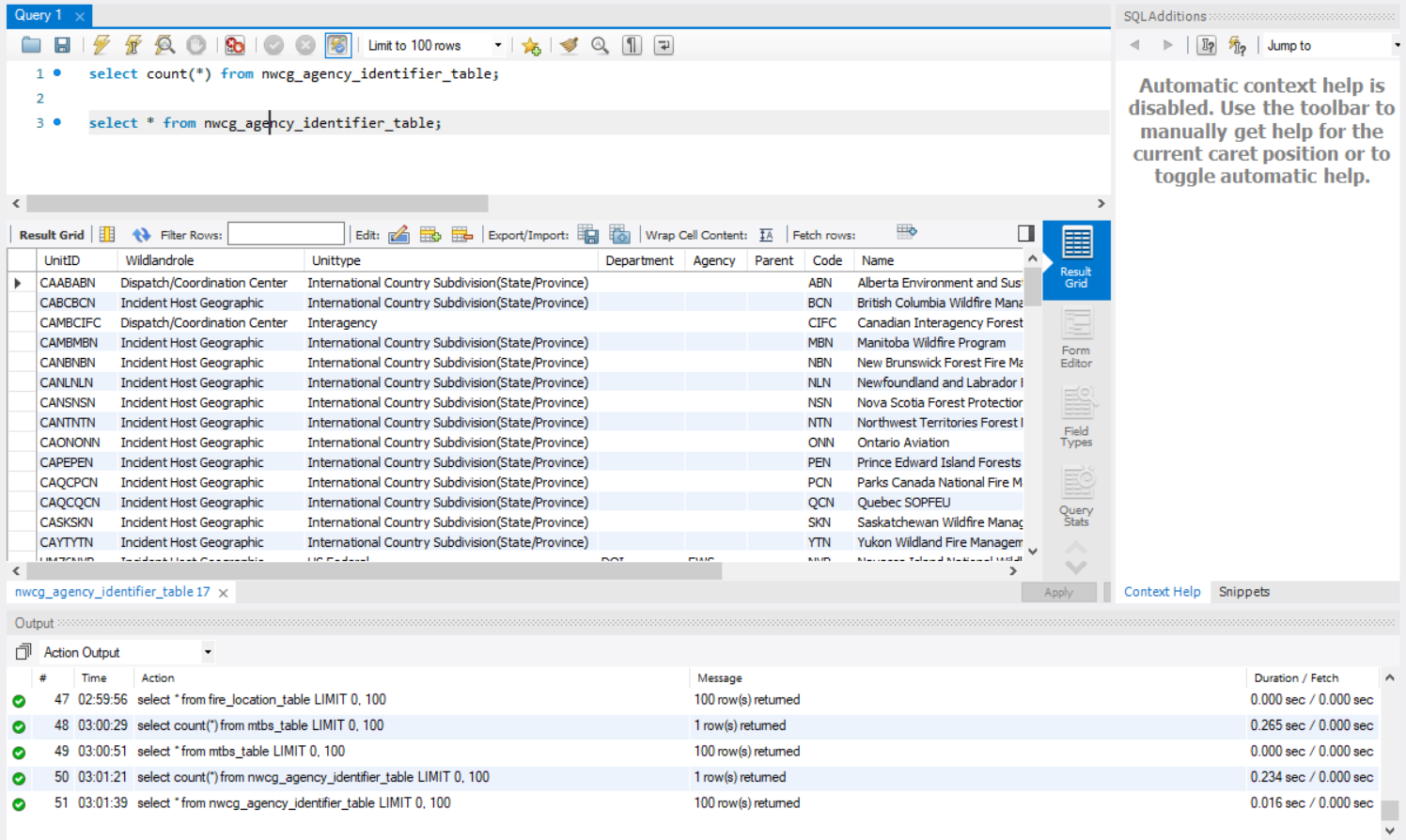
OWNER\_TABLE:



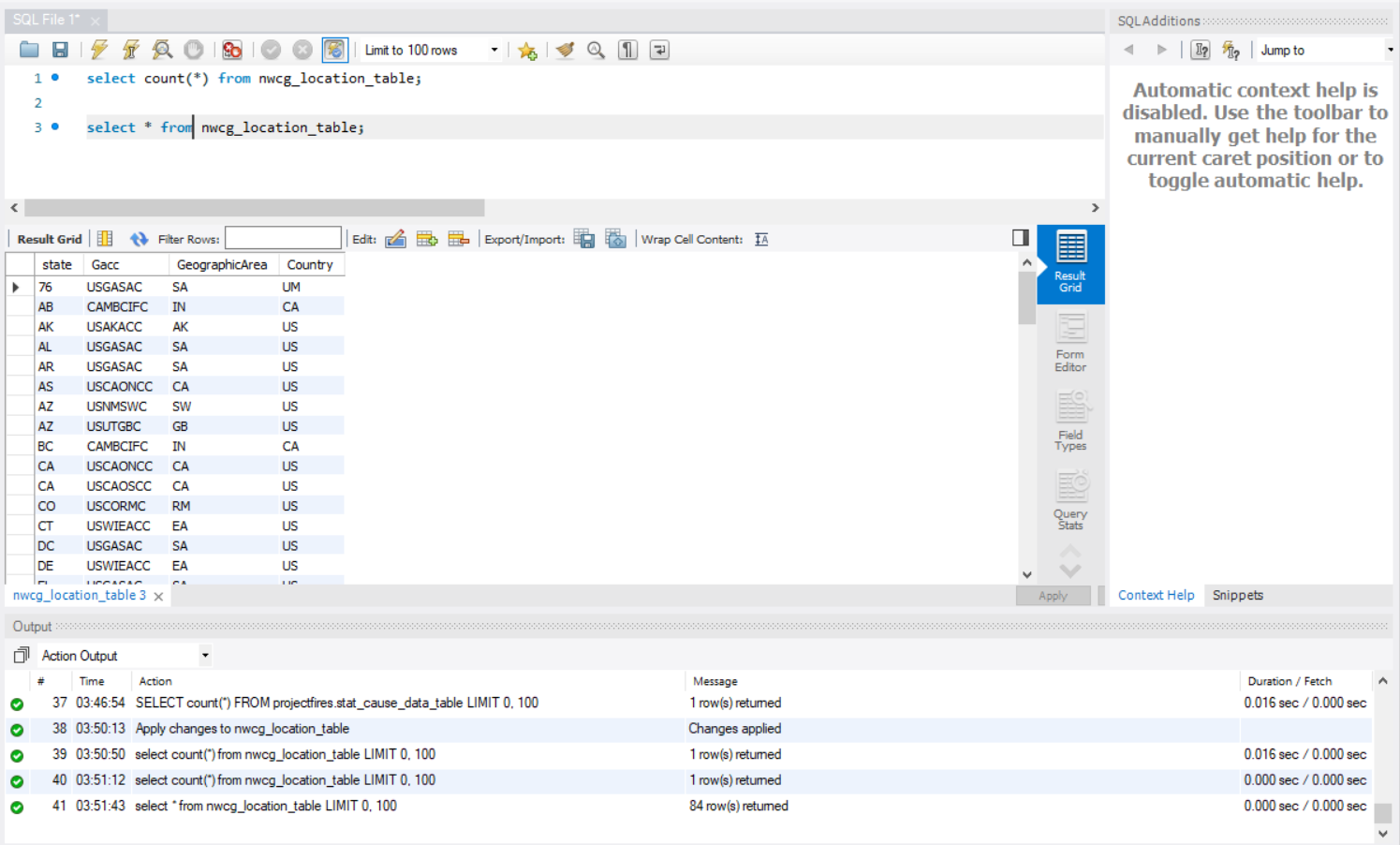
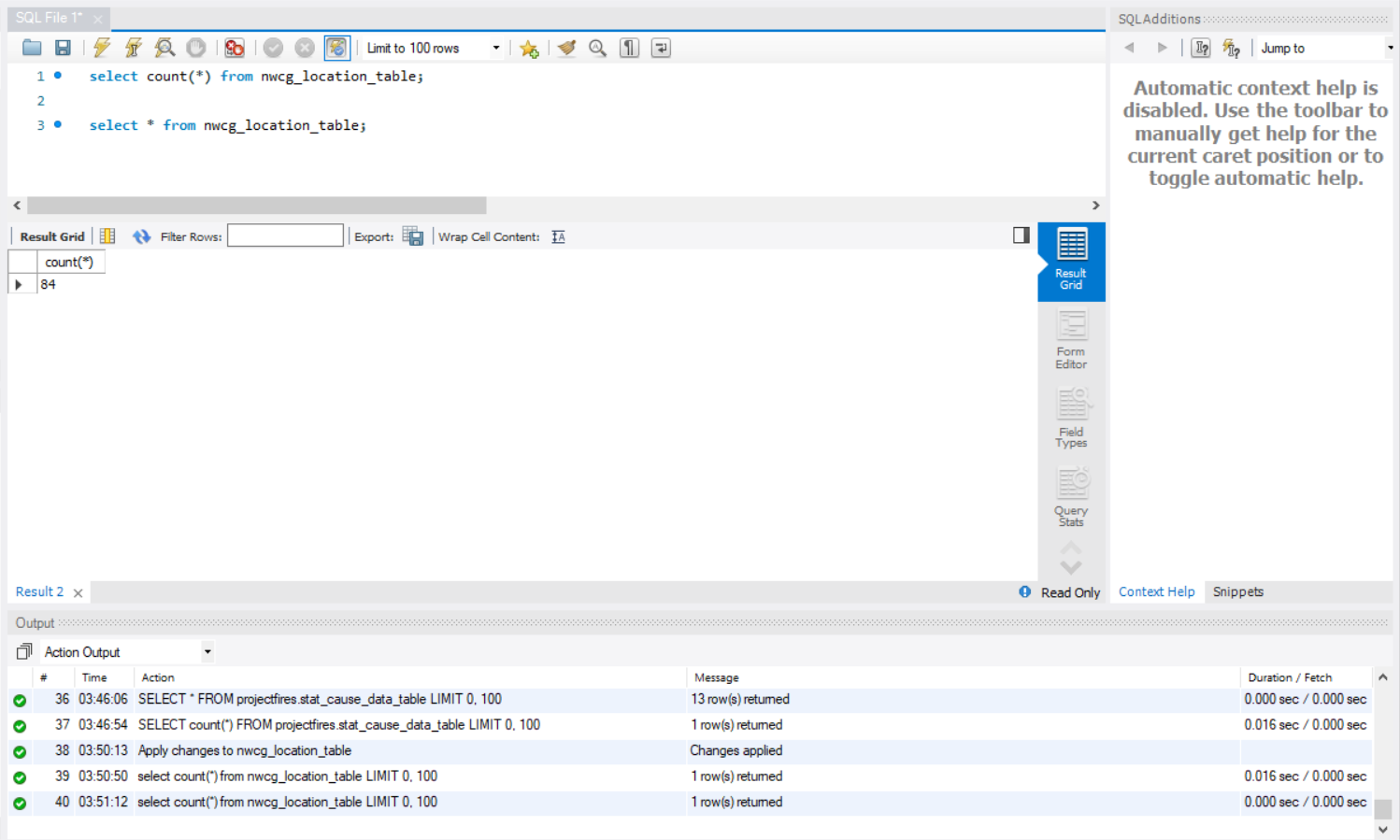
INCIDENT\_IDENTIFIER\_TABLE:



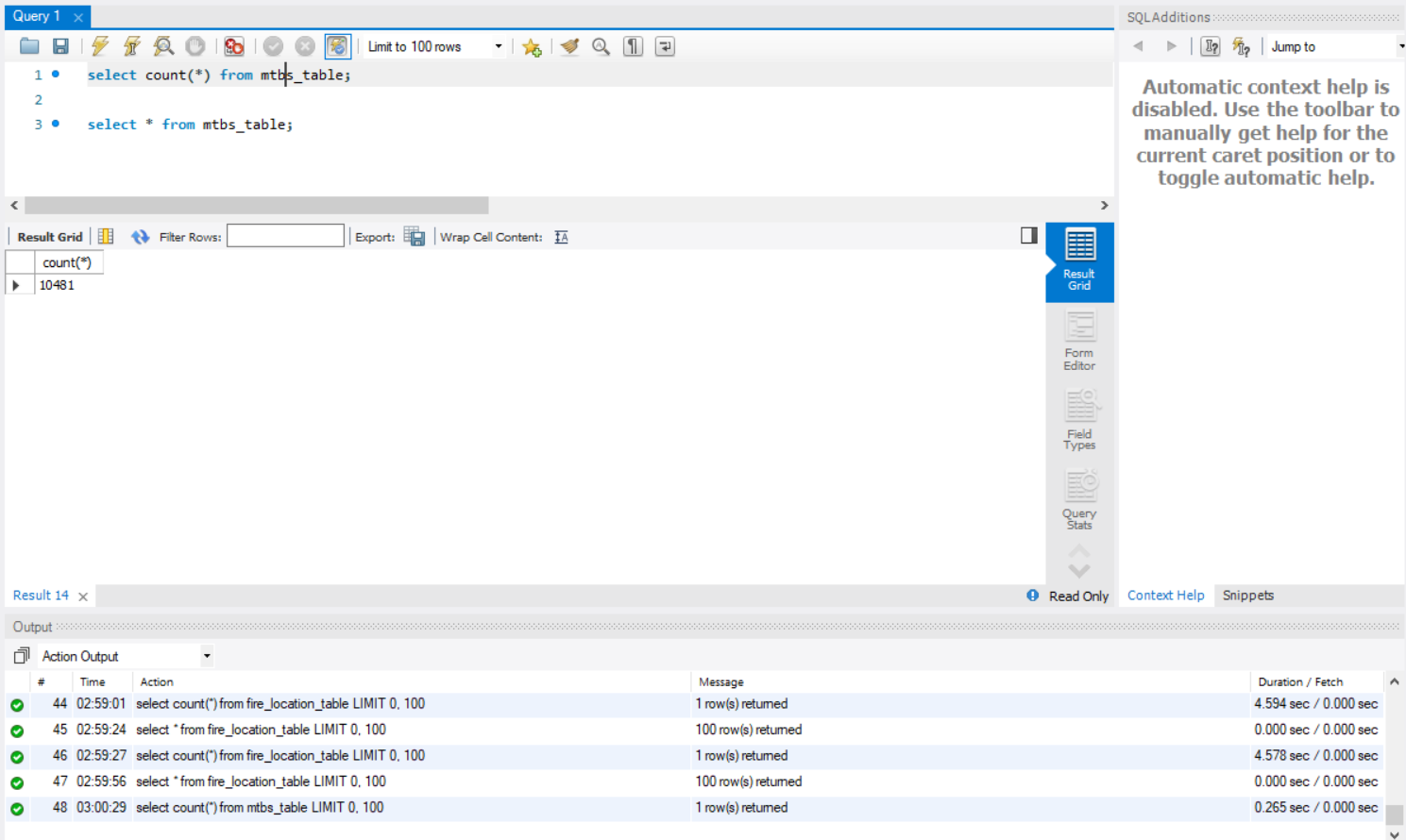
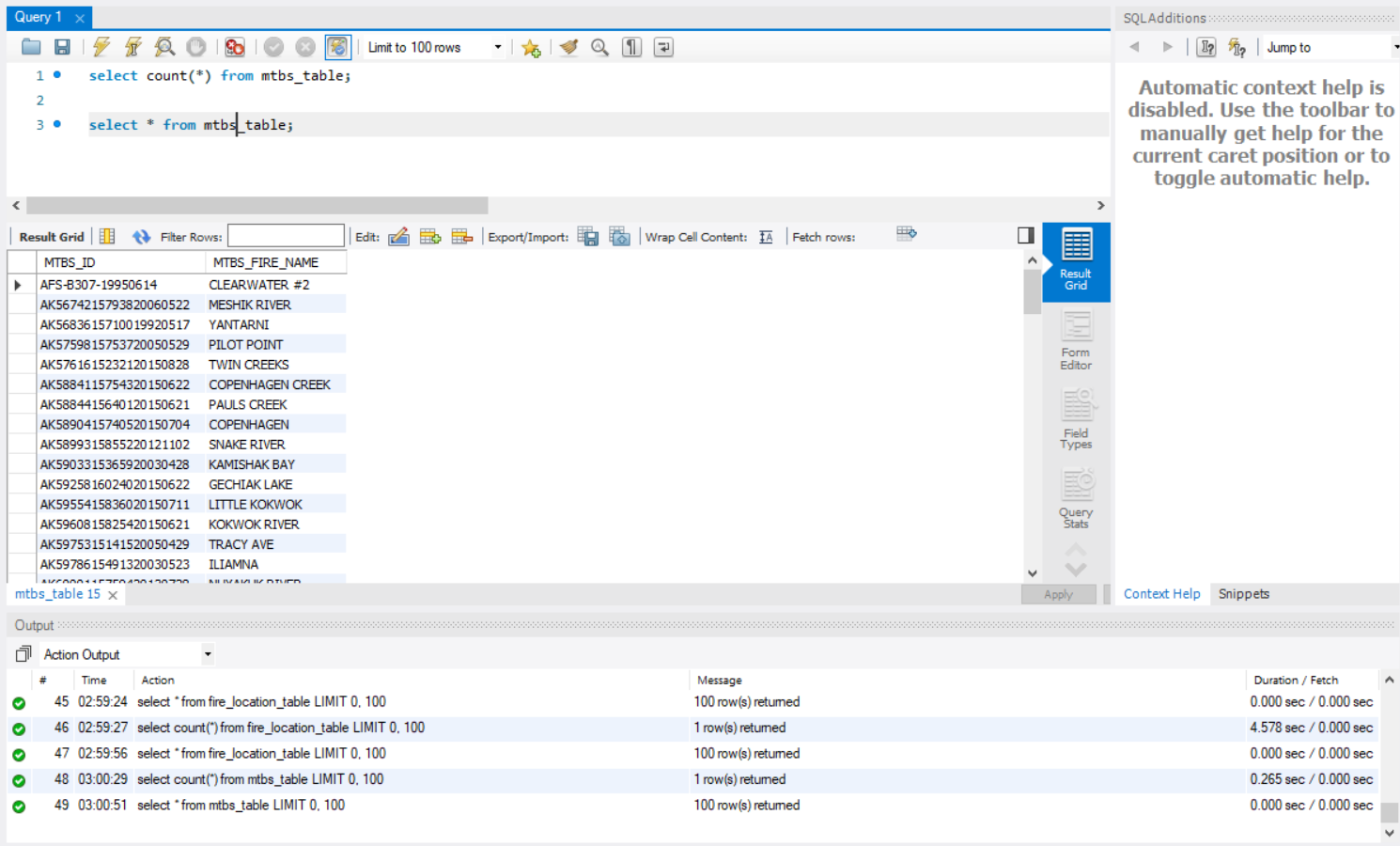
NWCG\_AGENCY\_IDENTIFIER\_TABLE:



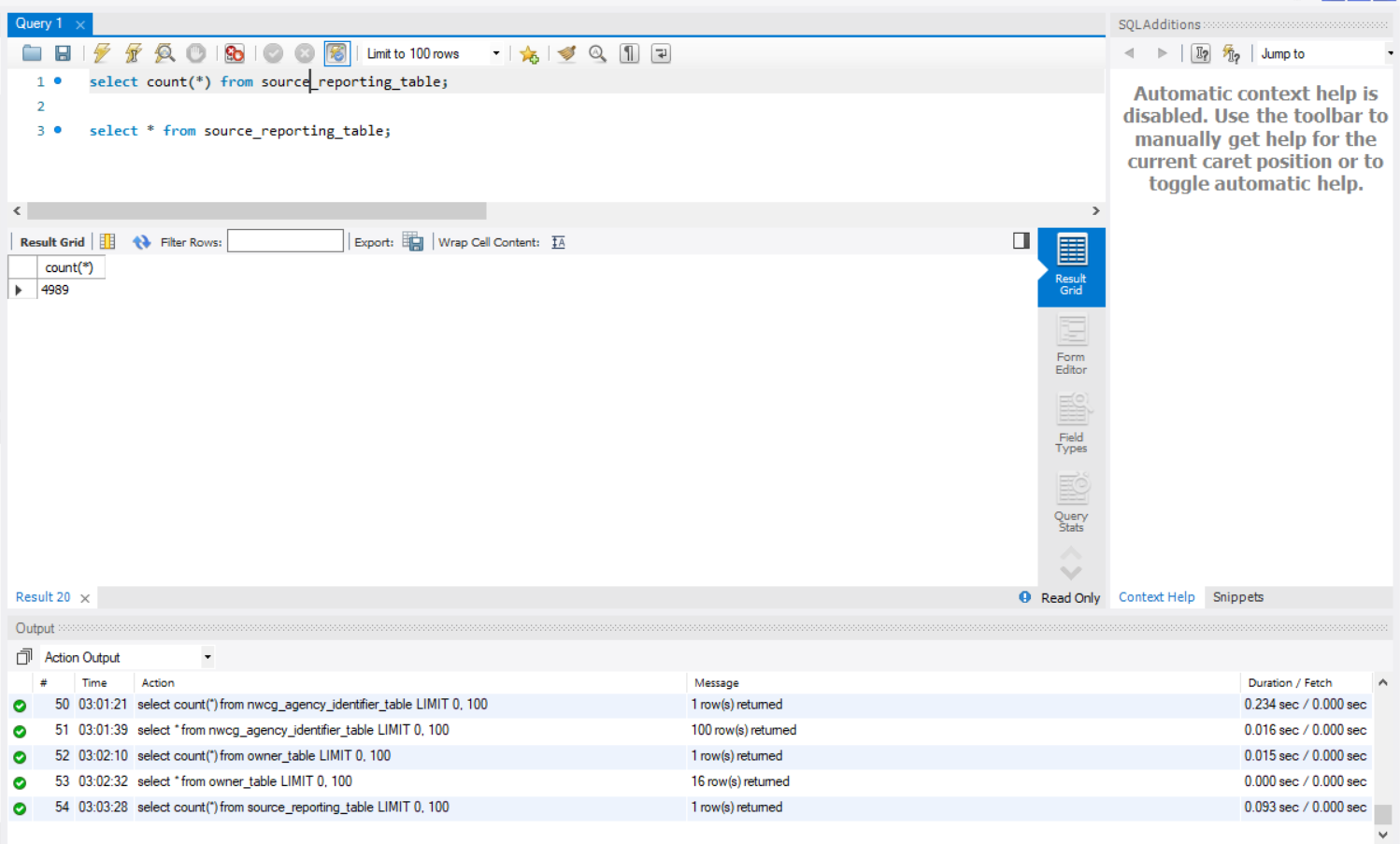
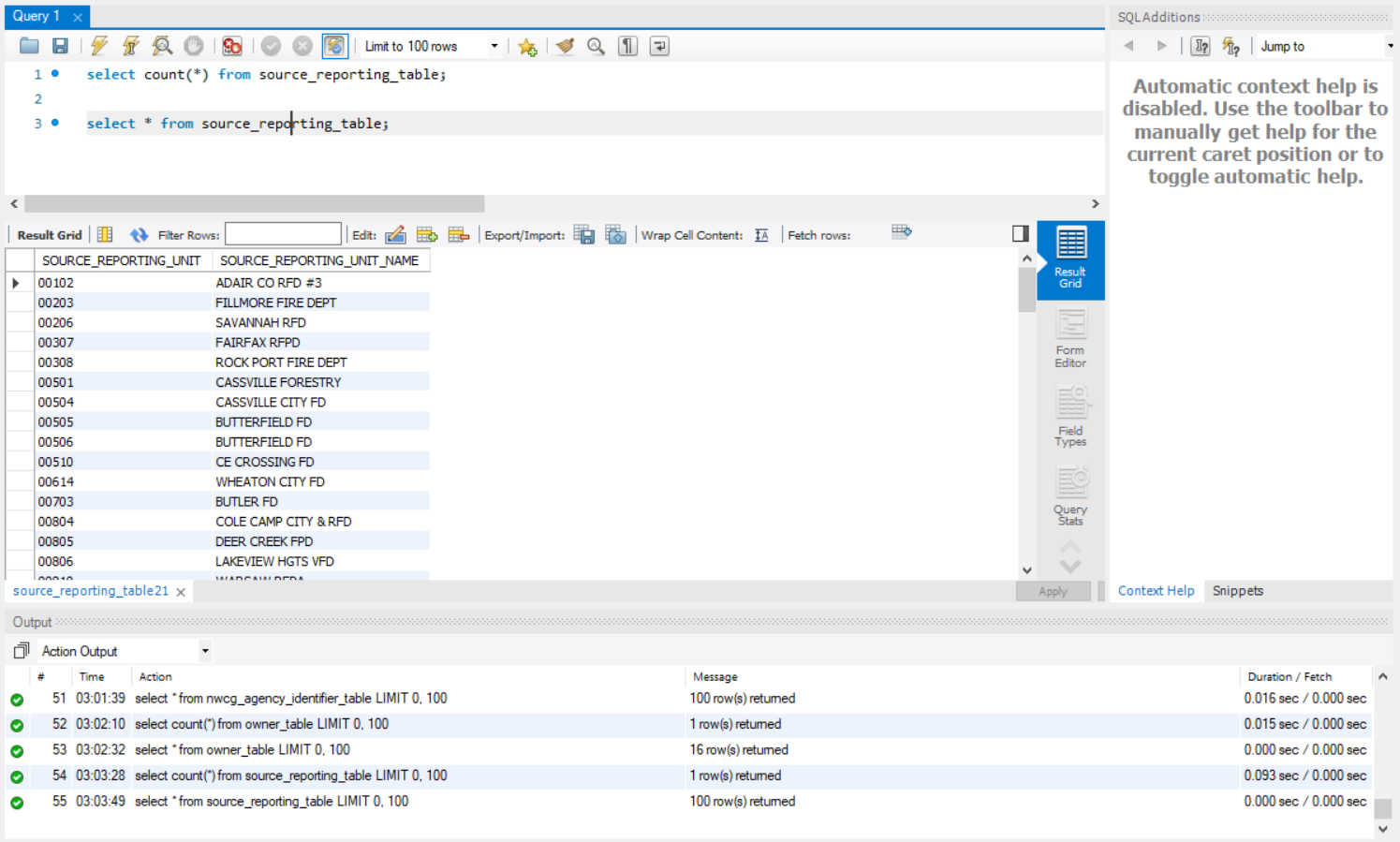
NWCG\_LOCATION\_TABLE:



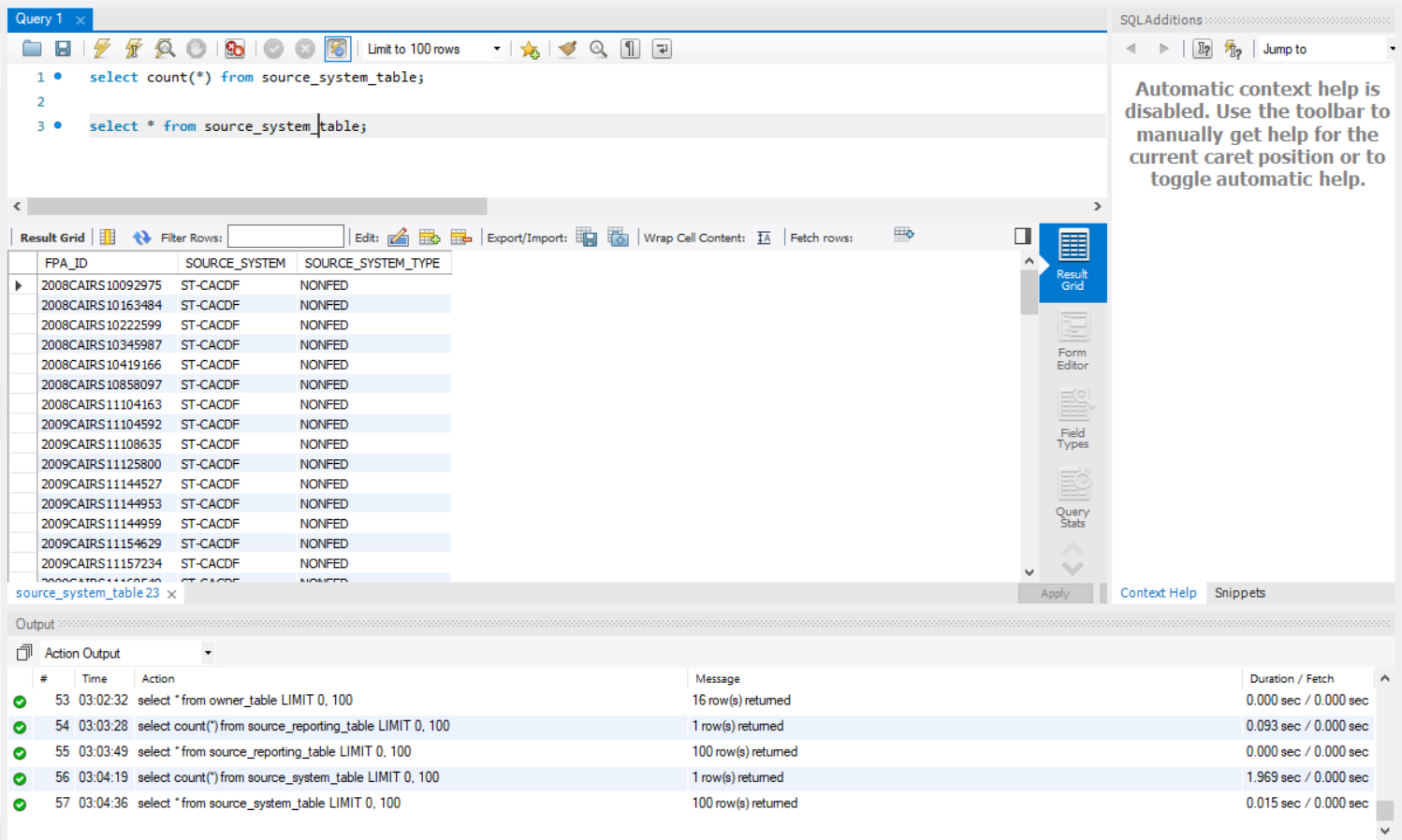
MTBS\_TABLE:



SOURCE\_REPORTING\_TABLE:



SOURCE\_SYSTEM\_TABLE:



## Data in the Database

|  |  |  |  |
| --- | --- | --- | --- |
| **Table Name** | **Primary Key** | **Foreign Key** | **# of Rows in Table** |
| FIRE\_TABLE | FOD\_ID | FPA\_ID  SOURCE\_REPORTING\_UNIT  MTBS\_ID  NWCG\_UnitID  (STATE,Gacc)  ICS\_209\_INCIDENT\_NUMBER  OWNER\_CODE,  STAT\_CAUSE\_CODE  (FIPS\_CODE,FIPS\_NAME)  (LATITUDE,LONGITUDE) | 1880465 |
| FIRE\_LOCATION\_TABLE | (Latitude, Longitude) | (FIPS\_CODE, FIPS\_NAME) | 1565730 |
| FIPS\_TABLE | (FIPS\_CODE, FIPS\_NAME) |  | 2693 |
| STAT\_CAUSE\_TABLE | STAT\_CAUSE\_CODE |  | 13 |
| OWNER\_TABLE | OWNER\_CODE |  | 16 |
| INCIDENT\_IDENTIFIER\_TABLE | ICS\_209\_INCIDENT\_NUMBER |  | 22736 |
| NWCG\_AGENCY\_IDENTIFIER\_TABLE | UnitID |  | 5867 |
| NWCG\_LOCATION\_TABLE | (state,Gacc) |  | 84 |
| MTBS\_TABLE | MTBS\_ID |  | 10481 |
| SOURCE\_REPORTING\_TABLE | SOURCE\_REPORTING\_UNIT |  | 4989 |
| SOURCE\_SYSTEM\_TABLE | SOURCE\_SYSTEM\_TABLE |  | 1880462 |

# 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 2 states have the least chance to win a share of the fund?

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

# of rows in the result: 2

### Translation

TRANSLATION: select the states from fire\_table joined with fire\_location\_table on LATITUDE and LONGITUDE joined with stat\_cause\_table on STAT\_CAUSE\_CODE where stat\_cause\_table STAT\_CAUSE-DESCR equals Debris Burning, grouping the result by STATE, ordering it by count of STATE in ascending order and displaying the top 2 results.

CLEAN UP: SELECT STATE FROM fire\_table AS ft JOIN fire\_location\_table AS flt

ON ft.location\_LATITUDE =flt.LONGITUE AND ft.location\_LONGITUDE=flt.LONGITUDE

JOIN stat\_cause\_table AS sct ON ft.stat\_cause\_table\_STAT\_CAUSE\_CODE= sct.STAT\_CAUSE\_CODE

WHERE sct.STAT\_CAUSE\_CODE=’Debris Burning’

GROUP BY flt.STATE

ORDER BY COUNT (flt.STATE)

LIMIT 2;

### 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 4 children in a group visiting a forest. Name top 5 forests where this would be the least appropriate.

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

# of rows in the result: 5

### Translation

TRANSLATION: select SOURCE\_REPORTING\_UNIT\_NAME from fire\_table joined with source\_reporting\_table on SOURCE\_REPORTING\_UNIT joined with stat\_cause\_table on STAT\_CAUSE\_CODE where stat\_cause\_STAT\_CAUSE\_CODE equals 8 and source\_reporting\_table\_ SOURCE\_REPORTING\_UNIT\_NAME ends with FOREST, group by SOURCE\_REPORTING\_UNIT\_NAME, order it by COUNT(SOURCE\_REPORTING\_UNIT\_NAME), SOURCE\_REPORTING\_UNIT\_NAME in ascending order and display the top 5 results.

CLEANUP: SELECT SOURCE\_REPORTING\_UNIT\_NAME FROM

fire\_table AS ft JOIN source\_reporting\_table AS srt ON ft.source\_reporting\_table\_SOURCE\_REPORTING\_UNIT= srt.SOURCE\_REPORTING\_UNIT

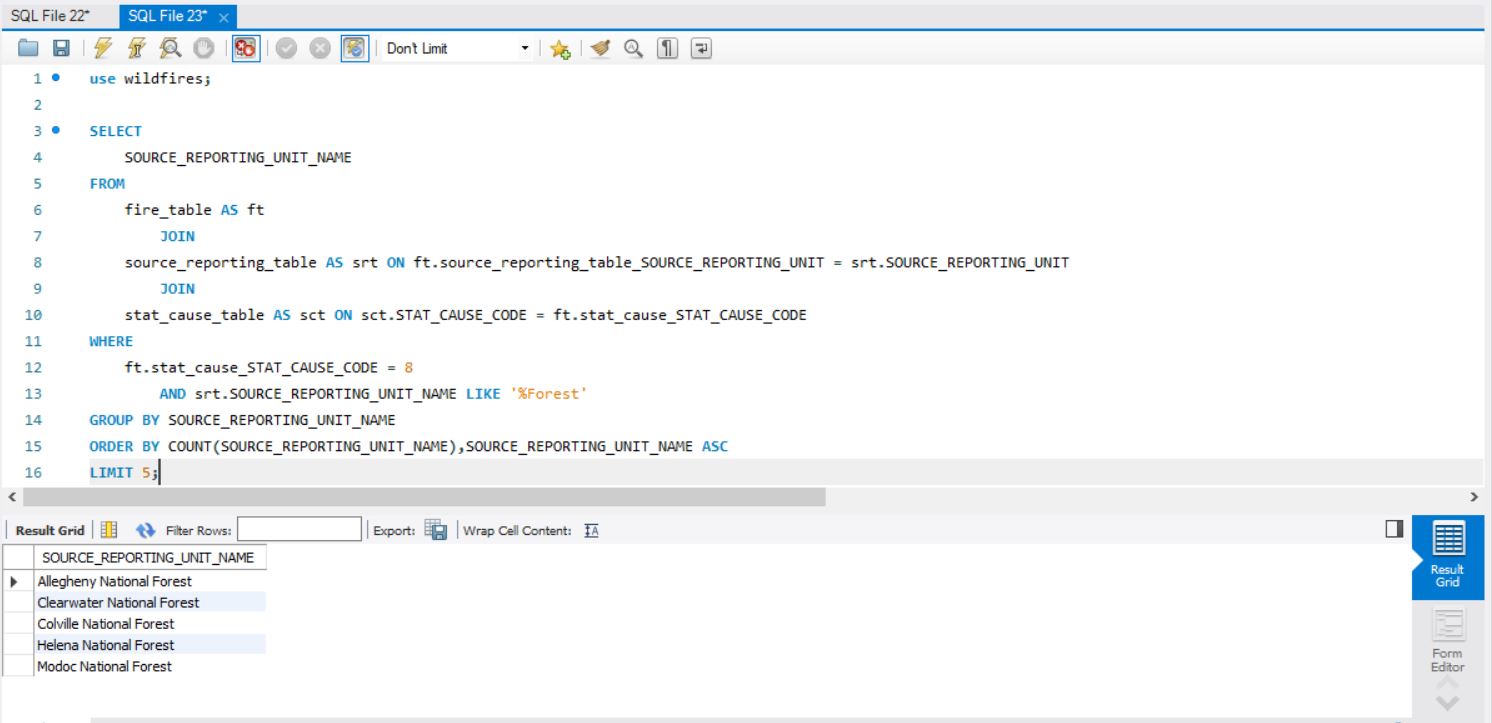
JOIN stat\_cause\_table AS sct ON sct.STAT\_CAUSE\_CODE= ft.stat\_cause\_STAT\_CAUSE\_CODE

WHERE ft.stat\_cause\_STAT\_CAUSE\_CODE=8 AND

srt. SOURCE\_REPORTING\_UNIT\_NAME LIKE ‘%Forest’ GROUP BY SOURCE\_REPORTING\_UNIT\_NAME ORDER BY COUNT(SOURCE\_REPORTING\_UNIT\_NAME), SOURCE\_REPORTING\_UNIT\_NAME ASC

LIMIT 5;

### Screen Shot of SQL Query and Results



## Query 3

### Question

One advocacy group says human actions and nature are equally to blame for most wildfires. Write a query that can help determine the truth of this statement.

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

The query results show that the count of damages due to natural causes is 282247 and the count of damages due to human actions is 1107167. Thus, the group’s claim that both human and natural causes are equally responsible for wildfires is incorrect.

### Translation

TRANSLATION: select fire\_cause, count of the cause from fire\_table joined with stat\_cause\_table on STAT\_CAUSE\_CODE where stat\_cause\_code table STAT\_CAUSE\_DESCR is neither ‘Miscellaneous’ nor ‘Missing/Undefined’ with result stored as ‘fire\_type’, and if STAT\_CAUSE\_DESCR equals ‘Lightning’, then it is assigned ‘Nature’ otherwise it is assigned ‘Human’ and the result is stored in the column fire\_cause.

CLEANUP: SELECT fire\_cause, COUNT (\*) AS count FROM

(SELECT

IF (STAT\_CAUSE\_DESCR= ‘Lightning’, ’Nature’, ‘Human’) AS fire\_cause FROM

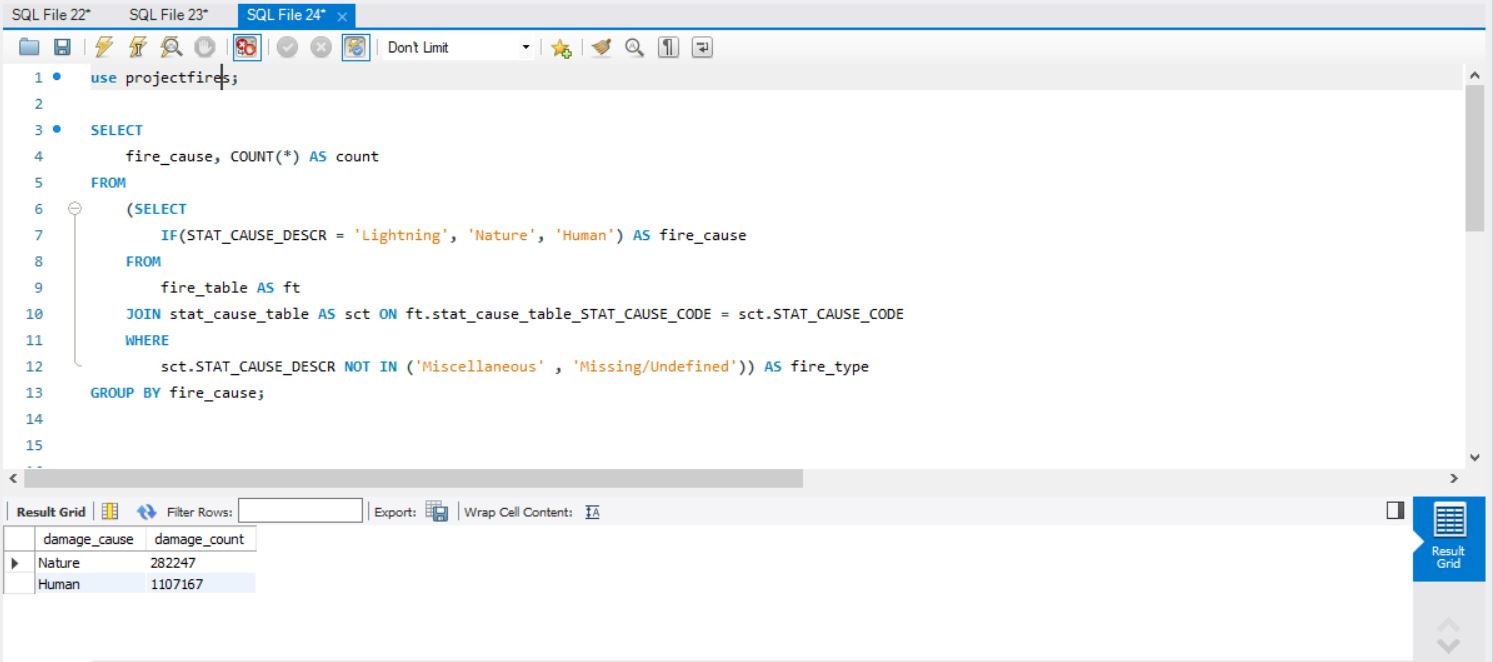
fire\_table AS ft JOIN stat\_cause\_table AS sct ON

ft.stat\_cause\_table\_STAT\_CAUSE\_CODE = sct. STAT\_CAUSE\_CODE

WHERE sct.STAT\_CAUSE\_DESCR NOT IN (‘Miscellaneous’, ’Missing/Undefined’)) AS fire\_type

GROUP BY fire\_cause;

### Screen Shot of SQL Query and Results



## Query 4

### Question

How many wildfires were reported by at least two units/agencies?

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

The query returns 76634 wildfires.

### Translation

TRANSLATION: select total count as fire\_count from select fire\_table FIRE\_NAME, COUNT(SOURCE\_REPORTING\_UNIT) AS reported\_units FROM fire\_table joined with source\_reporting\_table on SOURCE\_REPORTING\_UNIT group by FIRE\_NAME havingreported\_units greater than or equal to 2

CLEANUP: SELECT COUNT(\*) AS fire\_count

FROM

(SELECT ft.FIRE\_NAME, COUNT(SOURCE\_REPORTING\_UNIT) AS reported\_units

FROM fire\_table AS ft

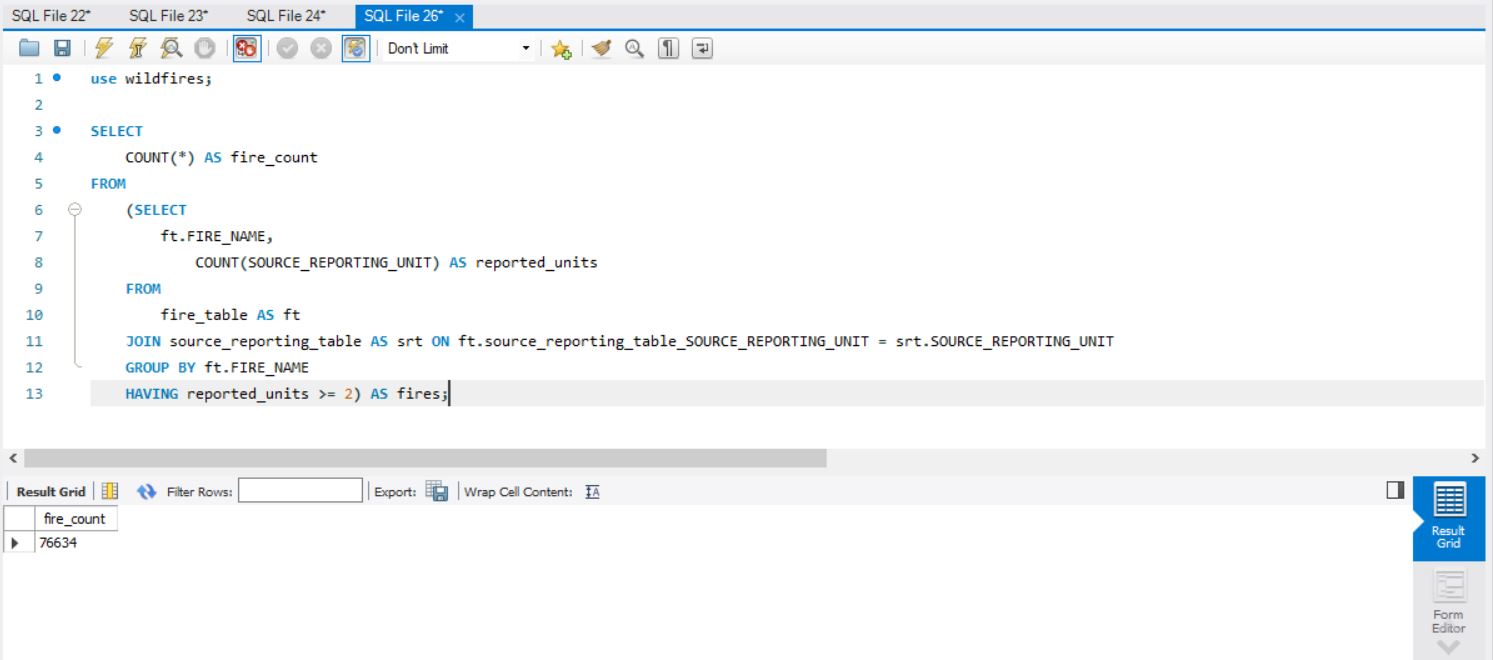
JOIN source\_reporting\_table AS srt ON ft.source\_reporting\_table\_ SOURCE\_REPORTING\_UNIT=

srt. SOURCE\_REPORTING\_UNIT

GROUP BY ft.FIRE\_NAME

HAVING reported\_units>=2) as fires;

### Screen Shot of SQL Query and Results



## Query 5

### Question

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

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

# of rows in the result: 9

### Translation

TRANSLATION: select SOURCE\_REPORTING\_UNIT\_NAME from fire\_table joined with source\_reporting\_table on SOURCE\_REPORTING\_UNIT where the difference between fire\_table CONT\_DATE and fire\_table DISCOVERY\_DATE is greater than 2 and SOURCE\_REPORTING\_UNIT\_NAME contains ‘Forest’, group the result by SOURCE\_REPORTING\_UNIT\_NAME having count of SOURCE\_REPORTING\_UNIT\_NAME equal to 1

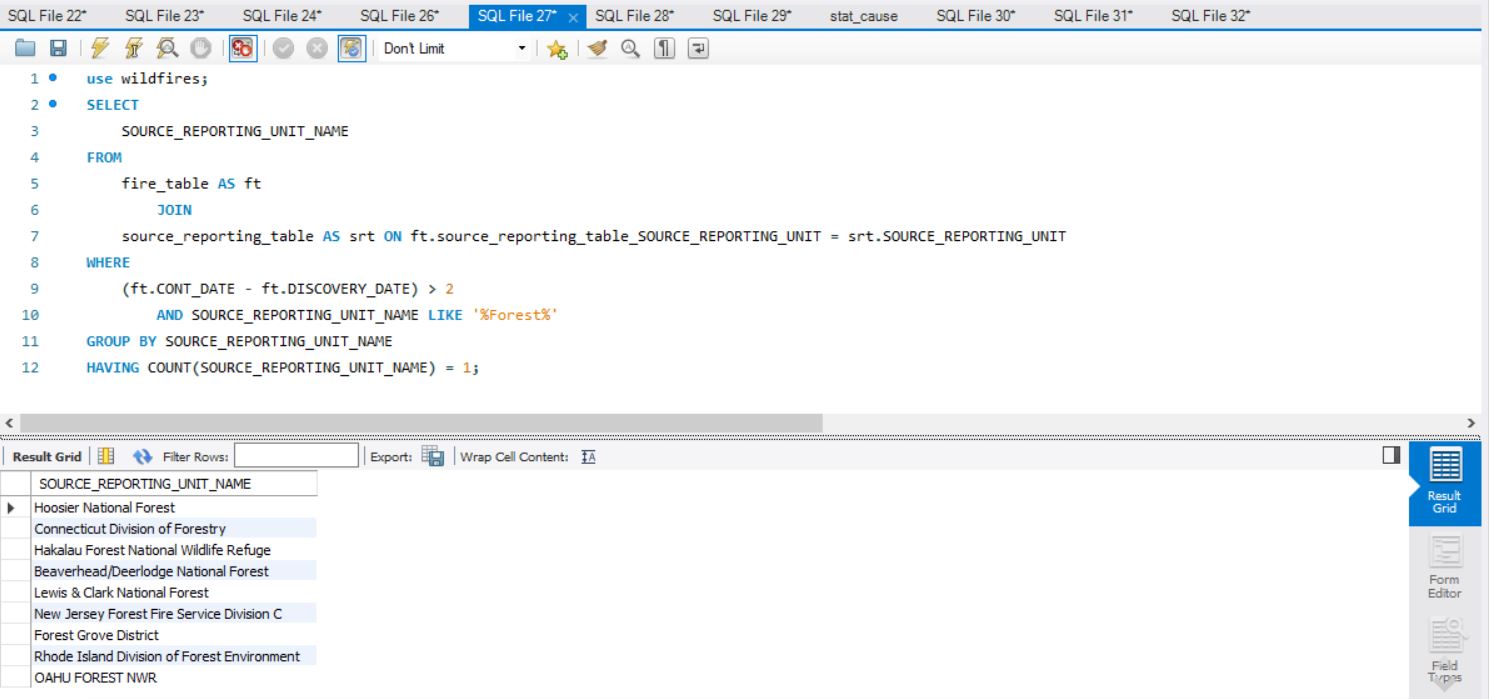
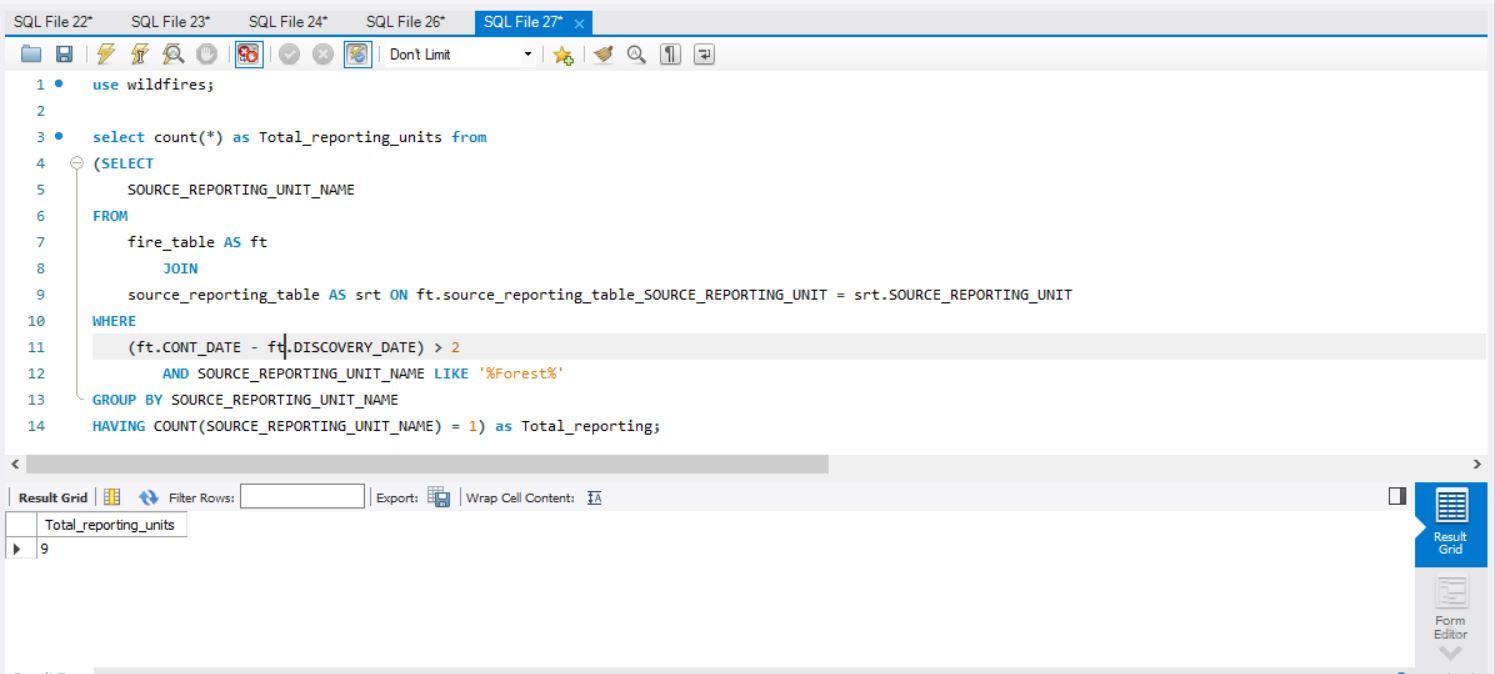
CLEAN-UP: SELECT SOURCE\_REPORTING\_UNIT\_NAME from fire\_table ft JOIN source\_reporting\_table srt ON ft.SOURCE\_REPORTING\_UNIT=srt. ft.SOURCE\_REPORTING\_UNIT

WHERE (ft.CONT\_DATE- ft.DISCOVERY\_DATE) >2 AND

SOURCE\_REPORTING\_UNIT\_NAME LIKE ‘%Forest%’,

GROUP BY SOURCE\_REPORTING\_UNIT\_NAME

HAVING COUNT(SOURCE\_REPORTING\_UNIT\_NAME)= 1;



### Screen Shot of SQL Query and Results

## Query 6

### Question

Which state had fires only in the second half of the calendar years?

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

There is no state where fires occurred only in the second half of the calendar years. So, no row is returned.

### Translation

TRANSLATION: select STATE from fire\_table joined with fire\_location\_table on fire\_table location\_LATITUDE equals fire\_location\_table LATITUDE and fire\_table location\_LONGITUDE equals fire\_location\_table LONGITUDE where DISCOVERY\_DOY is greater than 182 and fire\_location\_table STATE is not in select STATE from fire\_table joined with fire\_location\_table fire\_table location\_LATITUDE equals fire\_location\_table LATITUDE and fire\_table location\_LONGITUDE equals fire\_location\_table LONGITUDE where DISCOVERY\_DOY is less than 182

CLEAN-UP: select STATE FROM fire\_table AS ft JOIN fire\_location\_table AS flt

ON ft.location\_LATITUDE=flt.LATITUDE AND ft.location\_LONGITUDE=flt.LONGITUDE

WHERE DISCOVERY\_DOY>182

AND flt.STATE NOT IN

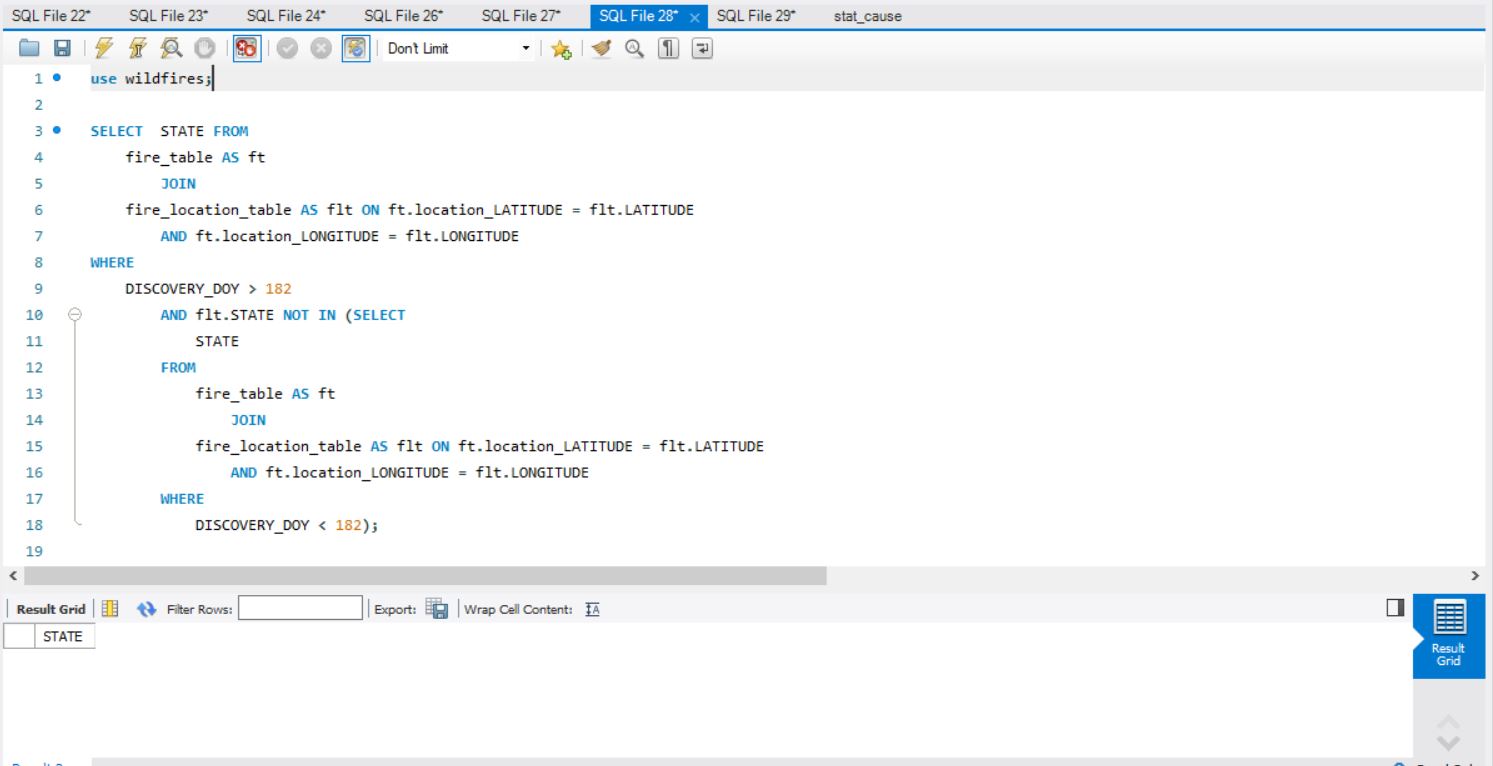
(SELECT STATE FROM

fire\_table AS ft JOIN fire\_location\_table AS flt

ON ft.location\_LATITUDE=flt.LATITUDE AND ft.location\_LONGITUDE=flt.LONGITUDE

WHERE DISCOVERY\_DOY<182);

### Screen Shot of SQL Query and Results



# Data Review for MongoDB

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

* The given data has been loaded into MongoDB in two collections: *fires* and *nwcg.*
* These collections in turn consist of documents that contain attributes in *field: value* pairs or embedded documents.
* The ‘\_id’ field is the primary key in each document.
* Both *fires* and *nwcg* collections have a common field: ‘unitID’.
* The following fields contain ‘integer’ data type:

1. FOD\_ID
2. FIRE\_YEAR
3. DISCOVERY\_DOY
4. STAT\_CAUSE\_CODE
5. OWNER\_CODE

* The following fields contain ‘float’ data type:

1. CONT\_DOY
2. FIRE\_SIZE
3. LATITUDE
4. LONGITUDE
5. FIPS\_CODE

* ‘CONT\_DATE and ‘DISCOVERY\_DATE’ contain data in ‘date’ format.

# Physical Mongo Database

## Assumptions/Notes About Data Set

* A new document STAT\_PROP\_CAUSE has been created that contains the embedded documents STAT\_CAUSE\_DESCR (data type: String) and STAT\_CAUSE\_CODE (data type:Integer).
* Missing values and duplicate data have been removed.

## 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 |  | 1880465 |
| nwcg |  | 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 2 states have the least chance to win a share of the fund?

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

The least chance to win a share can be interpreted in two ways:

1. Those states that didn’t suffer a forest fire due to debris will have no/least chance to win a share of the fund.
   1. As per Query1 – Part 1 results, every state suffered forest fire due to debris
2. Those states that suffered minimum in-terms of FIRE\_SIZE will have least chance to win a share of the fund.

Results:

States – DC , RI

Count - 2

### Translation

Filter (MATCH) all the forest fires due to debris burning

Then GROUP by STATE and compute the SUM of FIRE\_SIZE as TOTAL\_FIRE\_SIZE

Then SORT the results in ascending order based on SUM of FIRE\_SIZE (TOTAL\_FIRE\_SIZE)

PROJECT the STATE

And LIMIT the result set to top 2 states

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

Part 1

A screen shot of a clock

Description automatically generated

Part 2 – Actual Query with resultsA close up of a computer

Description automatically generated

## Query 2

### Question

One of the reporting agencies has suggested that children be banned from its forests unless there is one adult for every 4 children in a group visiting a forest. Name top 5 forests where this would be the least appropriate

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

For all those forests which had least no of forest fires due to Children, the reporting agencies suggestion would be least appropriate

As per data set, fire due to Children is given a code value 8. The query uses STAT\_CAUSE\_CODE 8 to identify fires due to Children

Result:

"Allegheny National Forest"

"Clearwater National Forest"

"Colville National Forest"

"Custer Gallatin National Forest"

"Custer National Forest"

**Count**: 5

### Translation

Filter (MATCH) all the source reporting units which are Forest(s) and all the forest fires due to Children (STAT\_CAUSE\_CODE - 8)

Then GROUP by SOURCE\_REPORTING\_UNIT\_NAME and compute the COUNT of fires as count

Then SORT the results in ascending order based on the count

PROJECT the SOURCE\_REPORTING\_UNIT\_NAME

And LIMIT the result set to top 5 Forests

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

A picture containing object, clock

Description automatically generated

## Query 3

### Question

One advocacy group says human actions and nature are equally to blame for most wildfires. Write a query that can help determine the truth of this statement.

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

All the forest fires occurred due to “Lightening” are considered as caused by Nature, all the others [such as 'Debris Burning', 'Campfire', 'Equipment Use', 'Arson', 'Children', 'Railroad', 'Smoking', 'Powerline', 'Structure', 'Fireworks'] except Miscellaneous OR Missing/Undefined are categorized into human actions.

Result:

"Nature", "count" : 278468

"Human", "count" : 1111469

"Miscellaneous/Unknown", "count" : 490528

### Translation

Summarize all the forest fires due to “Lightening” as “Nature”, all the forest fires due to “Miscellaneous” OR “Missing/Undefined” as "Miscellaneous/Unknown" and all the others as “Human”

GROUP by Summary (Nature [OR] Human [OR] Miscellaneous/Unknown) and COUNT the values in each group

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

A clock sitting in the dark

Description automatically generated

## Query 4

### Question

How many wildfires were reported by at least two units/agencies?

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

Each fire identified by FIRE\_NAME reported by two or more SOURCE\_REPORTING\_UNIT(s) must be retrieved

**RESULT**

Count: 50954

### Translation

PROJECT FIRE\_NAME and SOURCE\_REPORTING\_UNIT

Then GROUP by FIRE\_NAME and COUNT the no of different SOURCE\_REPORTING\_UNITs into unitsReported

Then MATCH all the fires which have unitsReported greater than or equal to two

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

A picture containing object, clock, indoor

Description automatically generated

## Query 5

### Question

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

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

The duration between DISCOVERY\_DATE and CONT\_DATE tells us the no od days the fire lasted. All those forests which had a forest fire for 2 days or more are retrieved

**RESULT**

{ "\_id" : { "SOURCE\_REPORTING\_UNIT\_NAME" : "Rhode Island Division of Forest Environment" }, "count" : 1 }

{ "\_id" : { "SOURCE\_REPORTING\_UNIT\_NAME" : "Hakalau Forest National Wildlife Refuge" }, "count" : 1 }

{ "\_id" : { "SOURCE\_REPORTING\_UNIT\_NAME" : "Connecticut Division of Forestry" }, "count" : 1 }

{ "\_id" : { "SOURCE\_REPORTING\_UNIT\_NAME" : "Pike and San Isabel National Forest" }, "count" : 1 }

{ "\_id" : { "SOURCE\_REPORTING\_UNIT\_NAME" : "Ozark-St. Francis Nat Forests" }, "count" : 1 }

{ "\_id" : { "SOURCE\_REPORTING\_UNIT\_NAME" : "Division of Forestry and Wildlife" }, "count" : 1 }

{ "\_id" : { "SOURCE\_REPORTING\_UNIT\_NAME" : "Ozark & St. Francis National Forests" }, "count" : 1 }

{ "\_id" : { "SOURCE\_REPORTING\_UNIT\_NAME" : "New Jersey Forest Fire Service Division C" }, "count" : 1 }

{ "\_id" : { "SOURCE\_REPORTING\_UNIT\_NAME" : "Forest Grove District" }, "count" : 1 }

{ "\_id" : { "SOURCE\_REPORTING\_UNIT\_NAME" : "Hoosier National Forest" }, "count" : 1 }

Count: 10

### Translation

Compute difference between DISCOVERY\_DATE and CONT\_DATE in terms of days into a new field “duration”

Then MATCH those fires if the “duration” is greater than 2 days

Then Match those fires if SOURCE\_REPORTING\_UNIT\_NAME contains “Forest”

Then GROUP by SOURCE\_REPORTING\_UNIT\_NAME

Then MATCH those fires which had a total one fire for a SOURCE\_REPORTING\_UNIT\_NAME

Then PROJECT SOURCE\_REPORTING\_UNIT\_NAME and count

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

### A screen shot of a computer Description automatically generatedQuery 6

### Question

Which state had fires only in the second half of the calendar years?

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

Identify all the states which had fires in the second half of the calendar years i.e., DISCOVERY\_DOY greater than 182 days

**RESULT:**

NONE

Count: 0

### Translation

### Compute whether the fire occurred in second half by comparing if DISCOVERY\_DOY is greater than 182 and add a new field secondHalf

### GROUP by STATE and unique values for secondHalf for each state

### Then MATCH all the fires with only secondHalf with value [true]

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

A clock sitting in the dark

Description automatically generated