SQL-Mongo Project – Spatial Data of US Wildfires

BUAN 6320

Submitted By:
Rishi Bhatia

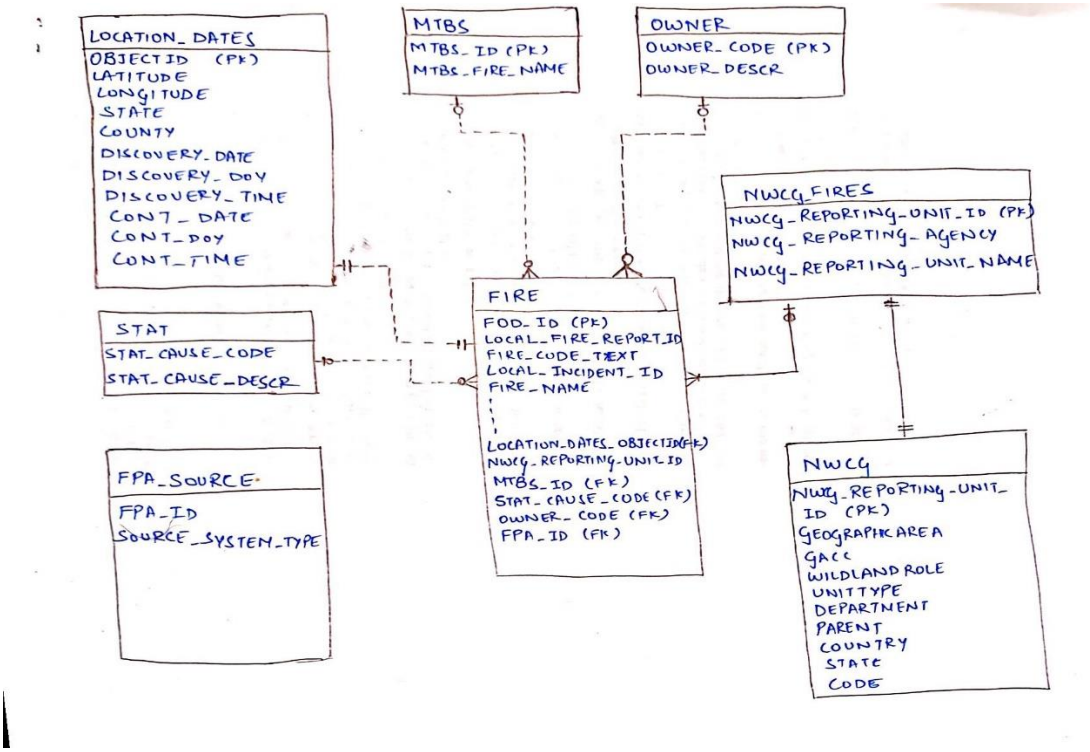
Contents

Data Model

Assumptions/Notes About Data Entities and Relationships

1. As FOD_ID is the global unique identifier and is non-null and unique, it is assumed to be the primary key of the main table. There are other foreign keys like STAT_CAUSE_ID, OWNER_CODE, MTBS_ID, NWCG_REPORTING_UNIT_ID, FPA_ID, OBJECTID which are the primary keys in other tables.
2. In Stat_Cause_Table, stat_cause_id is assumed to be the primary key as each key represent an individual stat_cause_descr, the stat_cause_id is unique and non_null and stat_cause_descr is directly dependant on the stat_cause_id.
3. In the Owner table, owner_code is assumed to be the primary key because it is non-null, unique and all the element of owner_descr depend directly on owner_code.
4. In the MTBS table, MTBS_ID directly depends on each unique MTBS_NAME. MTBS_ID is the primary key for this table because it is unique and non-null and every record of MTBS_Fire_Name directly depends on MTBS_ID.
5. The NWCG table is assumed to be a separate table because there are a lot of records in Unit_ID column, all of which are not used in the main table. NWCG table serves the purpose of all the records given by NWCG Reporting and all the columns from NWCG_FIRES, and NWCG are directly related to NWCG_REPORTING_Unit_ID. NWCG_REPORTING_Unit_ID, the primary key of the table is non-null and unique.
6. In the FPA_SOURCE table, FPA_ID is the primary key because all the other columns directly depend on the FPA_ID and the SOURCE_REPORTING_UNIT_ID cannot be a primary key because of its redundant values.

Entity-Relationship Diagram

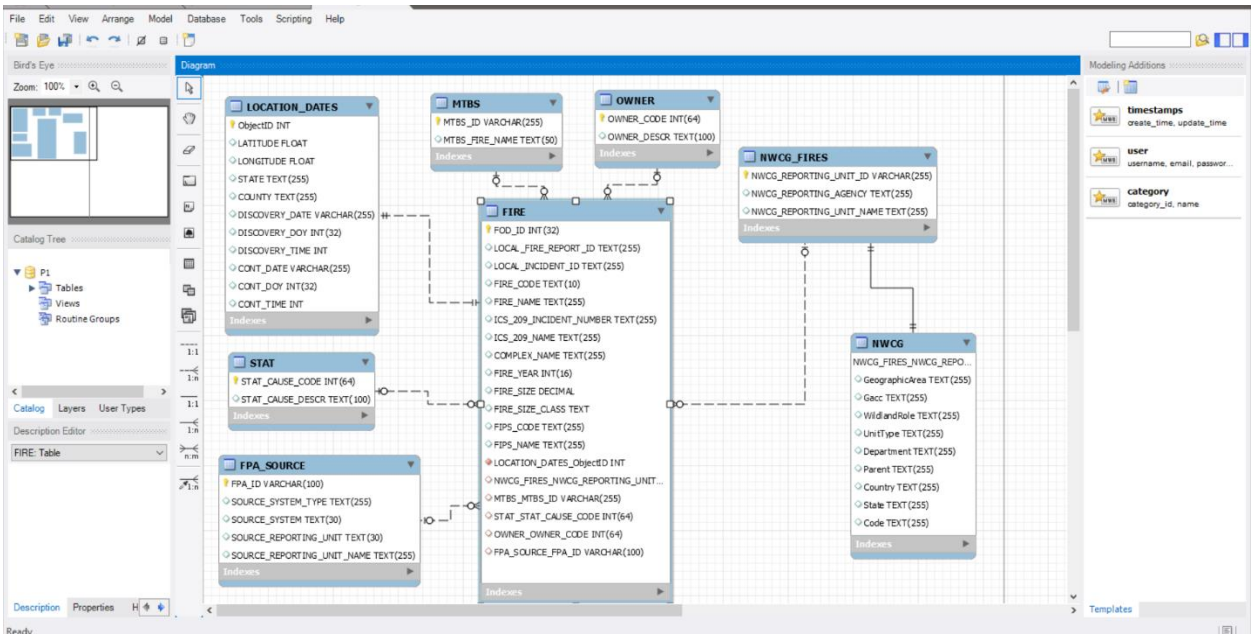


Physical Database

Assumptions/Notes About Data Set

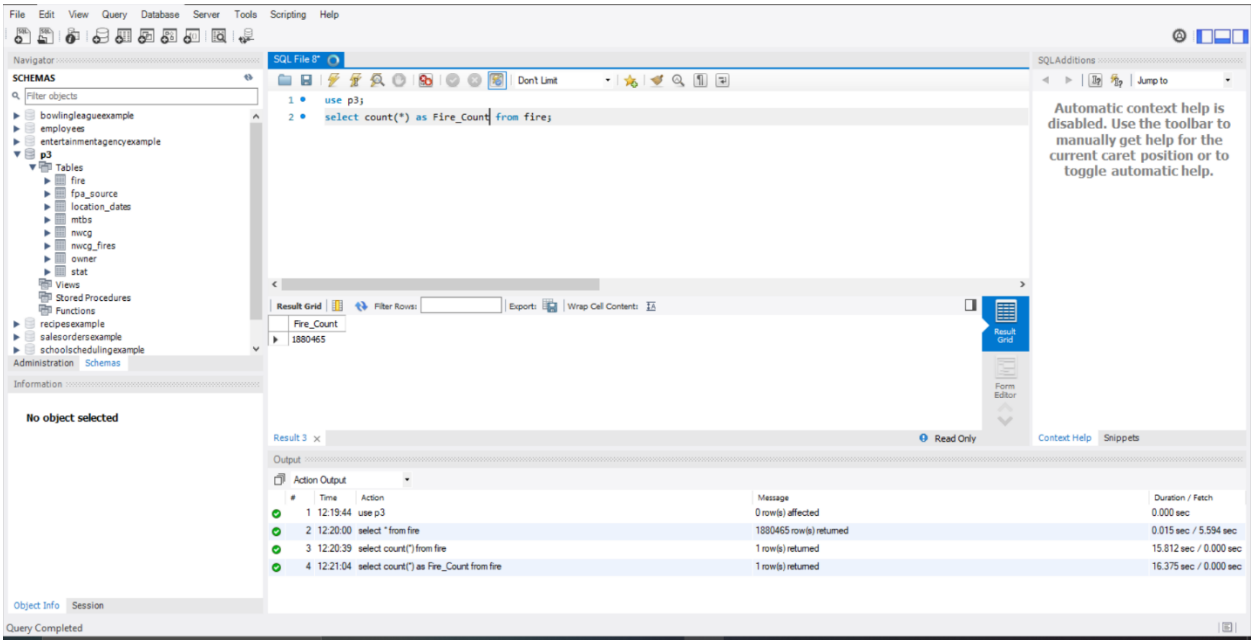
1. Shape column was removed because there was no relation of the column with the primary key and the data type was not of standard data types in SQL.
2. Source_Reporting_Unit could be a primary key in the Source Code table with the four columns namely Source_Reporting_Unit, Source_Reporting_Unit_Name, Source_System_Type, Source_System. But while examining the data, there were few duplicated values in Source_Reporting_Unit with different matching values in other columns, due to which this table was separated but with FPA_ID which showed the relationship between all the columns of the table to the primary key, FPA_ID
3. FIPS_CODE could be the primary key for the FIPS table with the columns namely, FIPS_CODE, FIPS_NAME. But after cross-examining the data, there were some duplicated values in FIPS_CODE, which mapped to two different FIPS_NAME. Thus, FIPS table could not be separated.
4. NWCG and NWCG_FIRES need to be separated because they all are connected with a common primary key, i.e. NWCG_REPORTING_UNIT_ID.
5. All the state, county, longitude, latitude, and dates of discovery with contained can be placed in one table that is LOCATION_DATES where the primary key is the ObjectID.

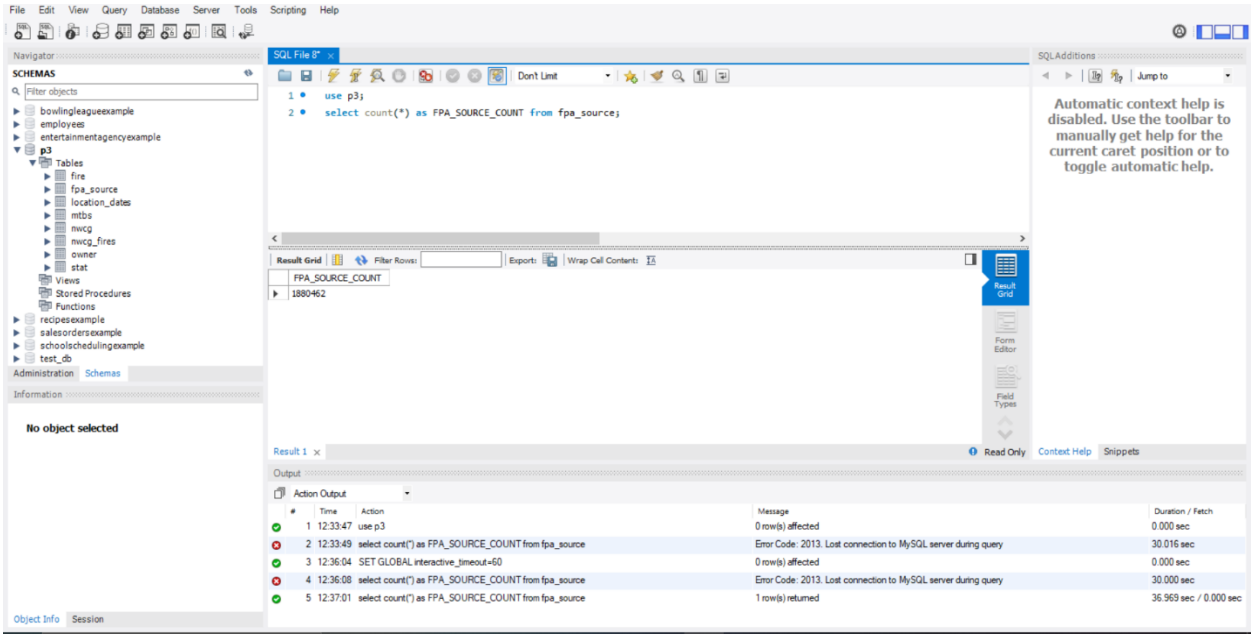
Screen shot of Physical Database objects

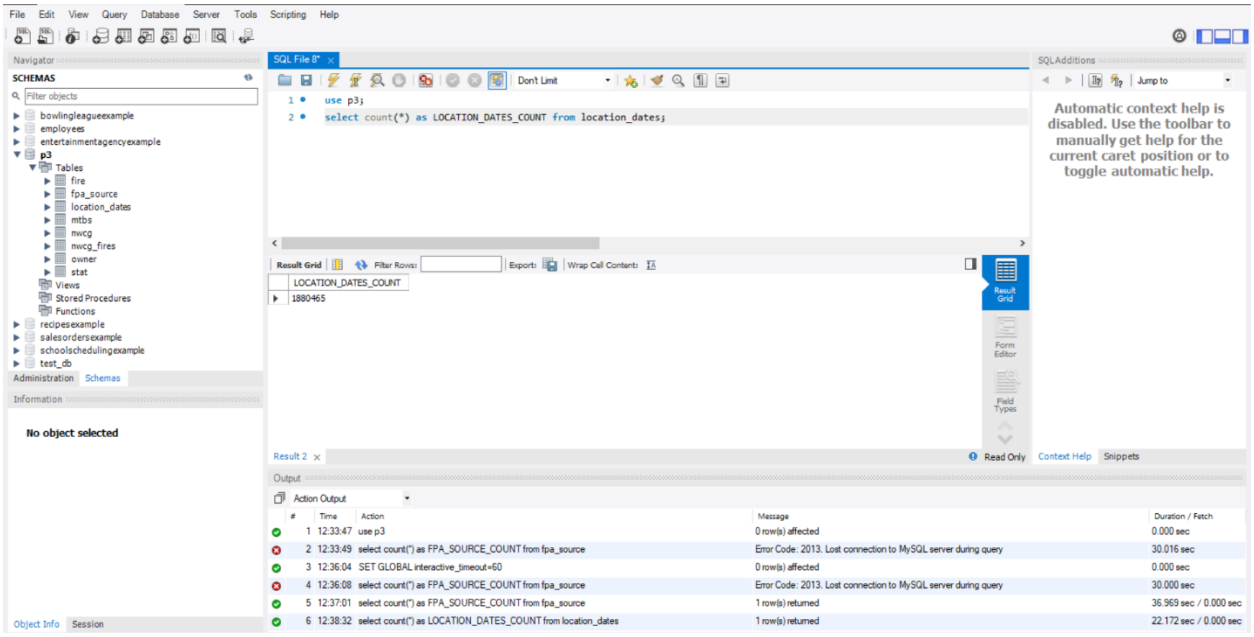


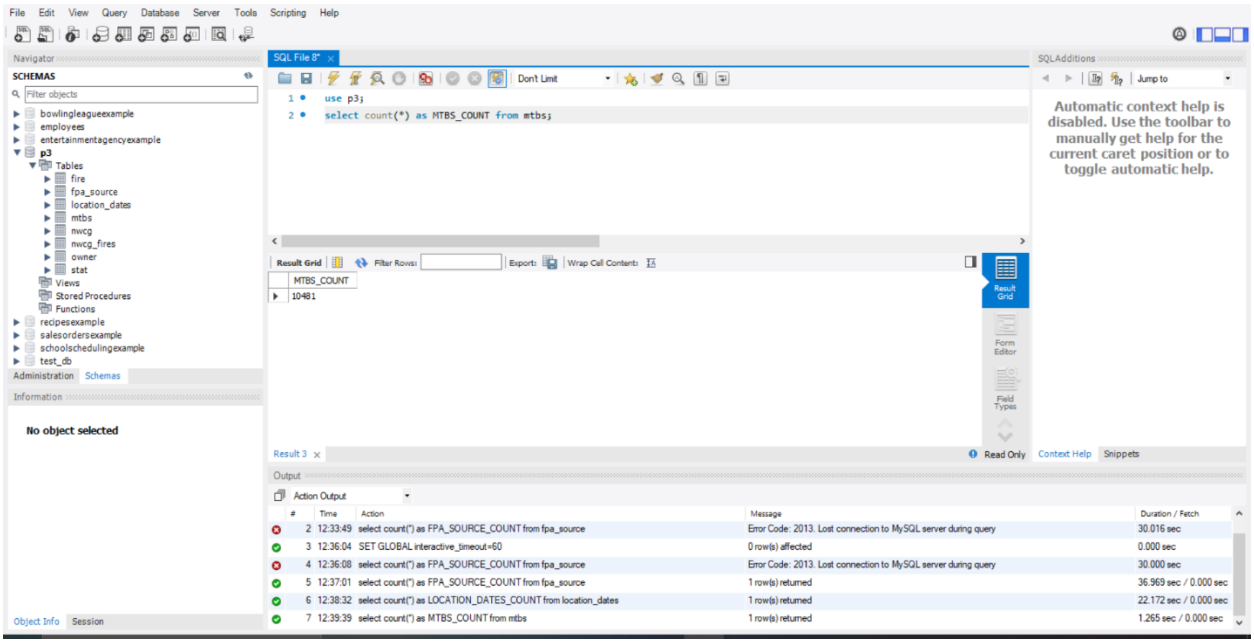
Data in the Database

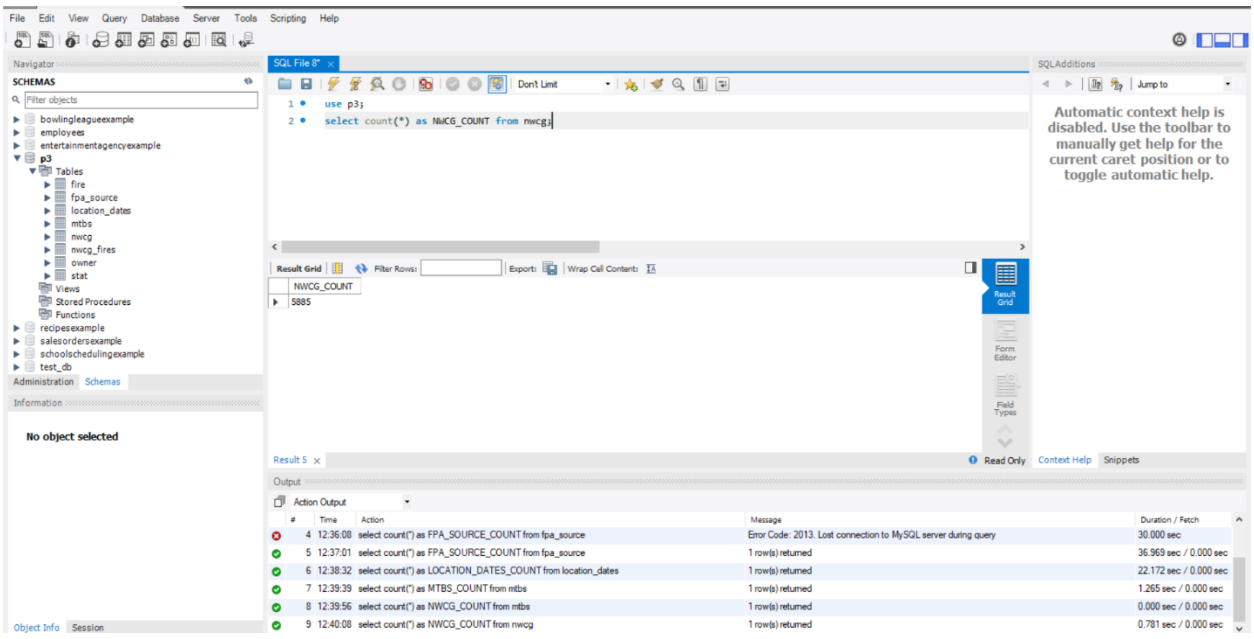
Table Name	Primary Key	Foreign Key	# of Rows in Table
FIRE	FOD_ID	LOCATION_DATE_OBJECTID NWCG_FIRES_NWCG_REPORTING_UNIT_ID MTBS_MTBS_ID STAT_STAT_CAUSE_CODE OWNER_OWNER_CODE FPA_SOURCE_FPA_ID	1880456
FPA_SOURCE	FPA_ID	-	1880462
LOCATION_DATES	OBJECT_ID	-	1880456
MTBS	MTBS_ID	-	10481
NWCG	NWCG_REPORTING_UNITID	NWCG_REPORTING_UNITID	5885
NWCG_FIRES	NWCG_REPORTING_UNIT_ID	-	5885
OWNER	OWNER_ID	-	16
STAT	STAT_CAUSE_CODE	-	13

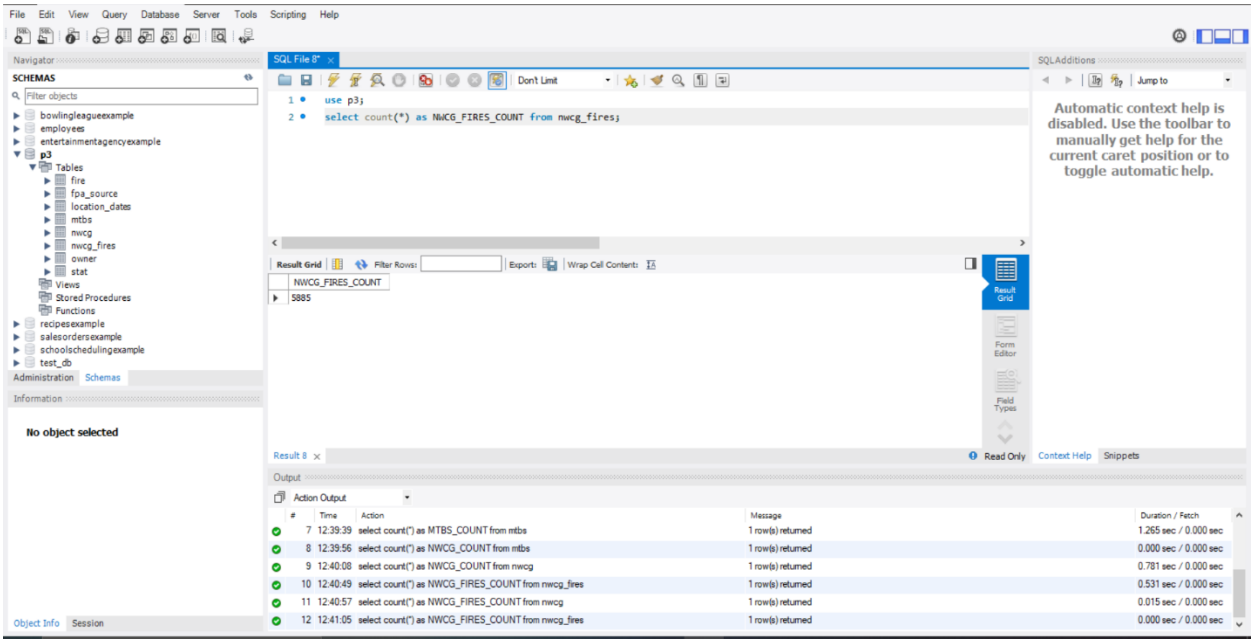


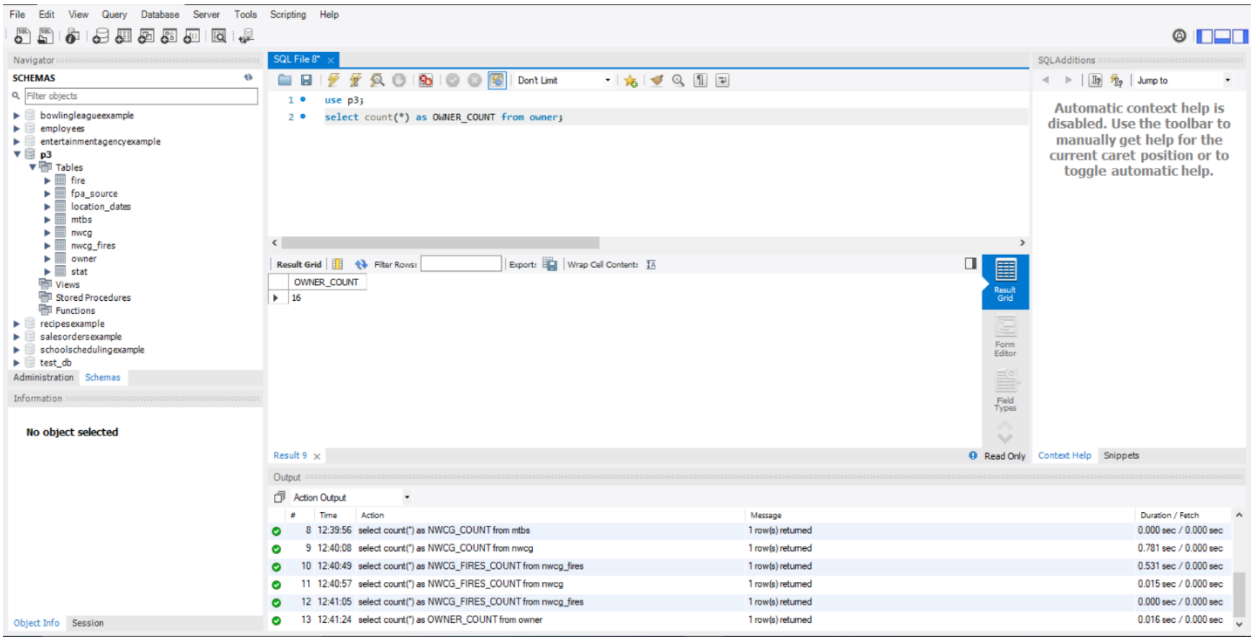

File Edit View Query Database Server Tools Scripting Help
Navigator
SCHEMAS
Filter objects
bowlingleagueexample
employees
entertainmentagencyexample
p3
Tables
fire
fpa_source
location_dates
mtbs
nwcg
nwcg_fires
owner
stat
Views
Stored Procedures
Functions
recipesexample
salesordersexample
schoolschedulingexample
test_db
Administration Schemas
Information
No object selected
Object Info Session
SQL File 8*
Don't Limit
1 use p3;
2 select count(*) as FPA_SOURCE_COUNT from fpa_source;
Result Grid Filter Rows: Export: Wrap Cell Content:
FPA_SOURCE_COUNT
1880462
Result 1
Read Only
SQLAdditions
Jump to
Automatic context help is disabled. Use the toolbar to manually get help for the current caret position or to toggle automatic help.
Context Help Snippets
Output
Action Output
Time Action Message Duration / Fetch
1 12:33:47 use p3 0 row(s) affected 0.000 sec
2 12:33:49 select count(*) as FPA_SOURCE_COUNT from fpa_source Error Code: 2013. Lost connection to MySQL server during query 30.016 sec
3 12:36:04 SET GLOBAL interactive_timeout=60 0 row(s) affected 0.000 sec
4 12:36:08 select count(*) as FPA_SOURCE_COUNT from fpa_source Error Code: 2013. Lost connection to MySQL server during query 30.000 sec
5 12:37:01 select count(*) as FPA_SOURCE_COUNT from fpa_source 1 row(s) returned 36.969 sec / 0.000 sec

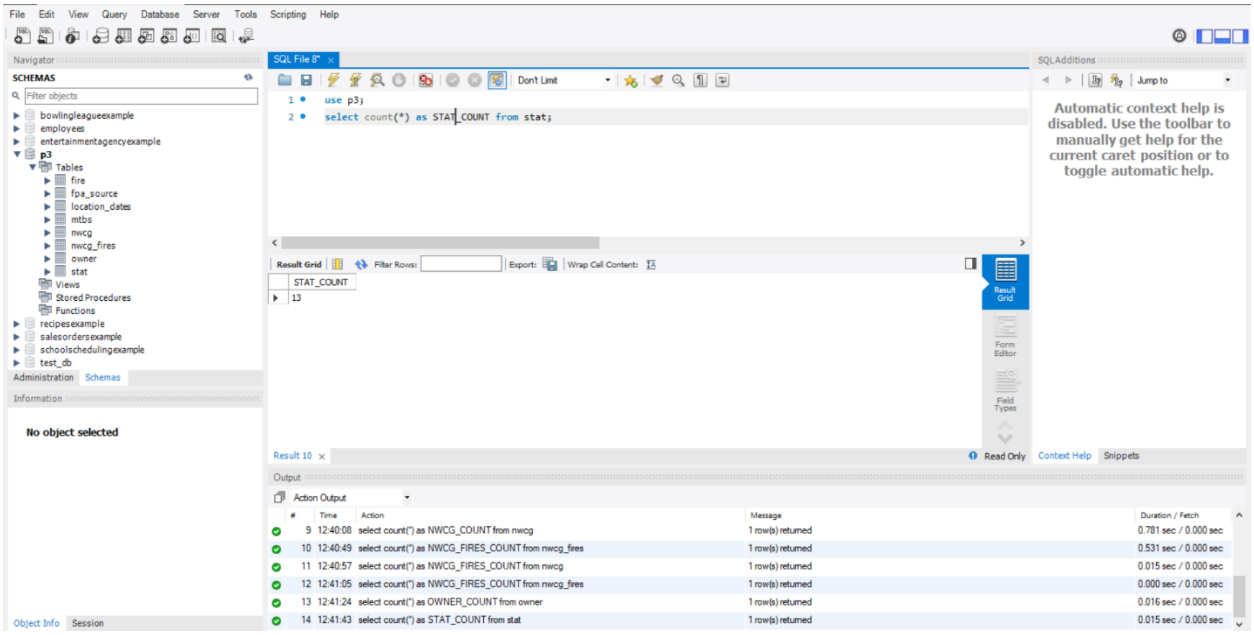

File Edit View Query Database Server Tools Scripting Help
Navigator
SCHEMAS
Filter objects
bowlingleagueexample
employees
entertainmentagencyexample
p3
Tables
fire
fpa_source
location_dates
mtbs
nwcg
nwcg_fires
owner
stat
Views
Stored Procedures
Functions
recipesexample
salesordersexample
schoolschedulingexample
test_db
Administration Schemas
Information
No object selected
Object Info Session
SQL File 8*
Don't Limit
1 use p3;
2 select count(*) as LOCATION_DATES_COUNT from location_dates;
Result Grid Filter Rows: Export: Wrap Cell Content:
LOCATION_DATES_COUNT
1880465
Result 2
Read Only
SQLAdditions
Jump to
Automatic context help is disabled. Use the toolbar to manually get help for the current caret position or to toggle automatic help.
Context Help Snippets
Output
Action Output
Time Action Message Duration / Fetch
1 12:33:47 use p3 0 row(s) affected 0.000 sec
2 12:33:49 select count(*) as FPA_SOURCE_COUNT from fpa_source Error Code: 2013. Lost connection to MySQL server during query 30.016 sec
3 12:36:04 SET GLOBAL interactive_timeout=60 0 row(s) affected 0.000 sec
4 12:36:08 select count(*) as FPA_SOURCE_COUNT from fpa_source Error Code: 2013. Lost connection to MySQL server during query 30.000 sec
5 12:37:01 select count(*) as FPA_SOURCE_COUNT from fpa_source 1 row(s) returned 36.969 sec / 0.000 sec
6 12:38:32 select count(*) as LOCATION_DATES_COUNT from location_dates 1 row(s) returned 22.172 sec / 0.000 sec


File Edit View Query Database Server Tools Scripting Help
SCHEMAS
bowlingleagueexample
employees
entertainmentagencyexample
p3
Tables
fire
fpa_source
location_dates
mtbs
nwcg
nwcg_fires
owner
stat
Views
Stored Procedures
Functions
recipesexample
salesordersexample
schoolschedulingexample
test_db
SQL File 8*
use p3;
select count(*) as MTBS_COUNT from mtbs;
MTBS_COUNT
10481
Automatic context help is disabled. Use the toolbar to manually get help for the current caret position or to toggle automatic help.
No object selected
Output
Action Output
2 12:33:49 select count(*) as FPA_SOURCE_COUNT from fpa_source Error Code: 2013. Lost connection to MySQL server during query 30.016 sec
3 12:36:04 SET GLOBAL interactive_timeout=60 0 row(s) affected 0.000 sec
4 12:36:08 select count(*) as FPA_SOURCE_COUNT from fpa_source Error Code: 2013. Lost connection to MySQL server during query 30.000 sec
5 12:37:01 select count(*) as FPA_SOURCE_COUNT from fpa_source 1 row(s) returned 36.969 sec / 0.000 sec
6 12:38:32 select count(*) as LOCATION_DATES_COUNT from location_dates 1 row(s) returned 22.172 sec / 0.000 sec
7 12:39:39 select count(*) as MTBS_COUNT from mtbs 1 row(s) returned 1.265 sec / 0.000 sec


File Edit View Query Database Server Tools Scripting Help
SCHEMAS
bowlingleagueexample
employees
entertainmentagencyexample
p3
Tables
fire
fpa_source
location_dates
mtbs
nwcg
nwcg_fires
owner
stat
Views
Stored Procedures
Functions
recipesexample
salesordersexample
schoolschedulingexample
test_db
SQL File 8*
use p3;
select count(*) as NWCG_COUNT from nwcg;
NWCG_COUNT
5885
Automatic context help is disabled. Use the toolbar to manually get help for the current caret position or to toggle automatic help.
No object selected
Output
Action Output
4 12:36:08 select count(*) as FPA_SOURCE_COUNT from fpa_source Error Code: 2013. Lost connection to MySQL server during query 30.000 sec
5 12:37:01 select count(*) as FPA_SOURCE_COUNT from fpa_source 1 row(s) returned 36.969 sec / 0.000 sec
6 12:38:32 select count(*) as LOCATION_DATES_COUNT from location_dates 1 row(s) returned 22.172 sec / 0.000 sec
7 12:39:39 select count(*) as MTBS_COUNT from mtbs 1 row(s) returned 1.265 sec / 0.000 sec
8 12:39:56 select count(*) as NWCG_COUNT from mtbs 1 row(s) returned 0.000 sec / 0.000 sec
9 12:40:08 select count(*) as NWCG_COUNT from nwcg 1 row(s) returned 0.781 sec / 0.000 sec


File Edit View Query Database Server Tools Scripting Help
Navigator
SCHEMAS
Filter objects
bowlingleagueexample
employees
entertainmentagencyexample
p3
Tables
fire
fpa_source
location_dates
mtbs
nwcg
nwcg_fires
owner
stat
Views
Stored Procedures
Functions
recipesexample
salesordersexample
schoolschedulingexample
test_db
Administration Schemas
Information
No object selected
Object Info Session
SQL File 8*
Don't Limit
1 use p3;
2 select count(*) as NWCG_FIRES_COUNT from nwcg_fires;
Result Grid Filter Rows: Export: Wrap Cell Content:
NWCG_FIRES_COUNT
5885
Result 8
Read Only
SQLAdditions
Jump to
Automatic context help is disabled. Use the toolbar to manually get help for the current caret position or to toggle automatic help.
Context Help Snippets
Output
Action Output
Time Action Message Duration / Fetch
7 12:39:39 select count(*) as MTBS_COUNT from mtbs 1 row(s) returned 1.265 sec / 0.000 sec
8 12:39:56 select count(*) as NWCG_COUNT from mtbs 1 row(s) returned 0.000 sec / 0.000 sec
9 12:40:08 select count(*) as NWCG_COUNT from nwcg 1 row(s) returned 0.781 sec / 0.000 sec
10 12:40:49 select count(*) as NWCG_FIRES_COUNT from nwcg_fires 1 row(s) returned 0.531 sec / 0.000 sec
11 12:40:57 select count(*) as NWCG_FIRES_COUNT from nwcg 1 row(s) returned 0.015 sec / 0.000 sec
12 12:41:05 select count(*) as NWCG_FIRES_COUNT from nwcg_fires 1 row(s) returned 0.000 sec / 0.000 sec


File Edit View Query Database Server Tools Scripting Help
Navigator
SCHEMAS
Filter objects
bowlingleagueexample
employees
entertainmentagencyexample
p3
Tables
fire
fpa_source
location_dates
mtbs
nwcg
nwcg_fires
owner
stat
Views
Stored Procedures
Functions
recipesexample
salesordersexample
schoolschedulingexample
test_db
Administration Schemas
Information
No object selected
Object Info Session
SQL File 8*
Don't Limit
1 use p3;
2 select count(*) as OWNER_COUNT from owner;
Result Grid Filter Rows: Export: Wrap Cell Content:
OWNER_COUNT
16
Result 9
Read Only
SQLAdditions
Jump to
Automatic context help is disabled. Use the toolbar to manually get help for the current caret position or to toggle automatic help.
Context Help Snippets
Output
Action Output
Time Action Message Duration / Fetch
8 12:39:56 select count(*) as NWCG_COUNT from mtbs 1 row(s) returned 0.000 sec / 0.000 sec
9 12:40:08 select count(*) as NWCG_COUNT from nwcg 1 row(s) returned 0.781 sec / 0.000 sec
10 12:40:49 select count(*) as NWCG_FIRES_COUNT from nwcg_fires 1 row(s) returned 0.531 sec / 0.000 sec
11 12:40:57 select count(*) as NWCG_FIRES_COUNT from nwcg 1 row(s) returned 0.015 sec / 0.000 sec
12 12:41:05 select count(*) as NWCG_FIRES_COUNT from nwcg_fires 1 row(s) returned 0.000 sec / 0.000 sec
13 12:41:24 select count(*) as OWNER_COUNT from owner 1 row(s) returned 0.016 sec / 0.000 sec


File Edit View Query Database Server Tools Scripting Help
Navigator
SCHEMAS
Filter objects
bowlingleagueexample
employees
entertainmentagencyexample
p3
Tables
fire
fpa_source
location_dates
mtbs
nwcg
nwcg_fires
owner
stat
Views
Stored Procedures
Functions
recipesexample
salesordersexample
schoolschedulingexample
test_db
Administration Schemas
Information
No object selected
Object Info Session
SQL File 8*
Don't Limit
1 use p3;
2 select count(*) as STAT_COUNT from stat;
Result Grid Filter Rows: Export: Wrap Cell Contents:
STAT_COUNT
13
Result Grid
Form Editor
Field Types
Result 10
Read Only
SQL Additions
Jump to
Automatic context help is disabled. Use the toolbar to manually get help for the current caret position or to toggle automatic help.
Context Help Snippets
Output
Action Output
Time Action Message Duration / Fetch
9 12:40:08 select count(*) as NWCG_COUNT from nwcg 1 row(s) returned 0.781 sec / 0.000 sec
10 12:40:49 select count(*) as NWCG_FIRES_COUNT from nwcg_fires 1 row(s) returned 0.531 sec / 0.000 sec
11 12:40:57 select count(*) as NWCG_FIRES_COUNT from nwcg 1 row(s) returned 0.015 sec / 0.000 sec
12 12:41:05 select count(*) as NWCG_FIRES_COUNT from nwcg_fires 1 row(s) returned 0.000 sec / 0.000 sec
13 12:41:24 select count(*) as OWNER_COUNT from owner 1 row(s) returned 0.016 sec / 0.000 sec
14 12:41:43 select count(*) as STAT_COUNT from stat 1 row(s) returned 0.015 sec / 0.000 sec

SQL Queries

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. Which state has the least chance to win a share of fund?

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

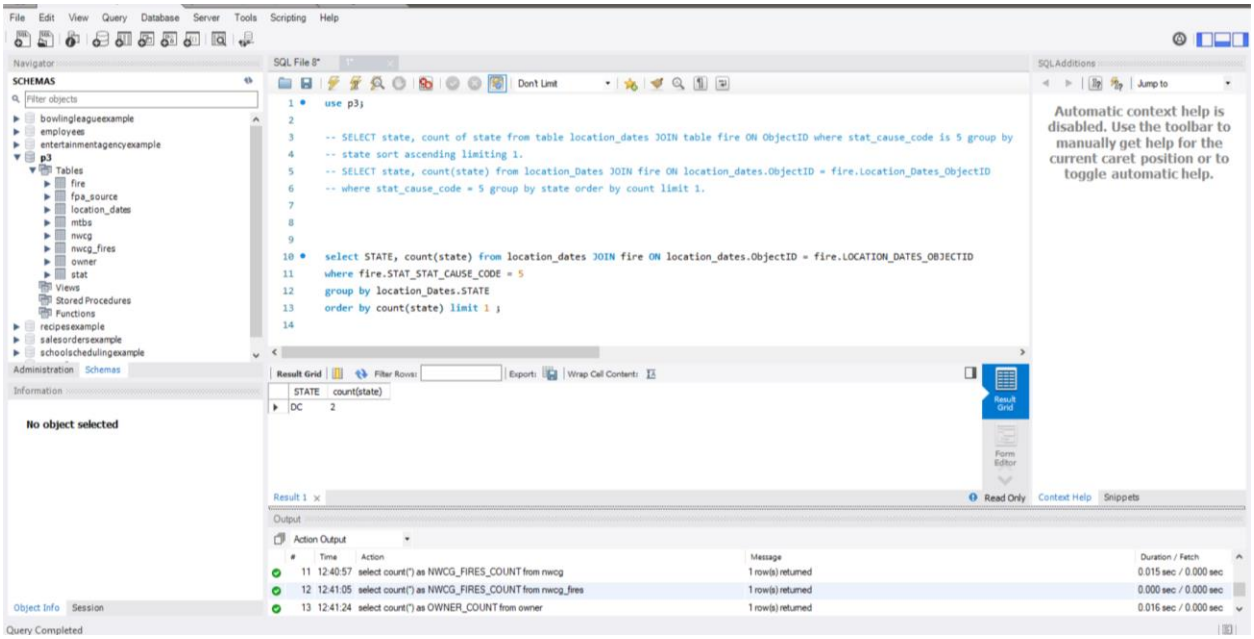
The state which had the least number of fires with stat_cause_code = 5, or 'debris burning' will have the least chance to win a share of fund. The result will contain 1 row stating the state having the least chance along with the count of debris burning.

Assumption: The fund shall be allocated to the state most affected state which have high number of debris burning in order to eliminate emergency.

Translation

SELECT state, count of state from table location_dates JOIN table fire ON ObjectID where stat_cause_code is 5 group by state sort ascending limiting

Screen Shot of SQL Query and Results



Query 2

Question 2

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

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

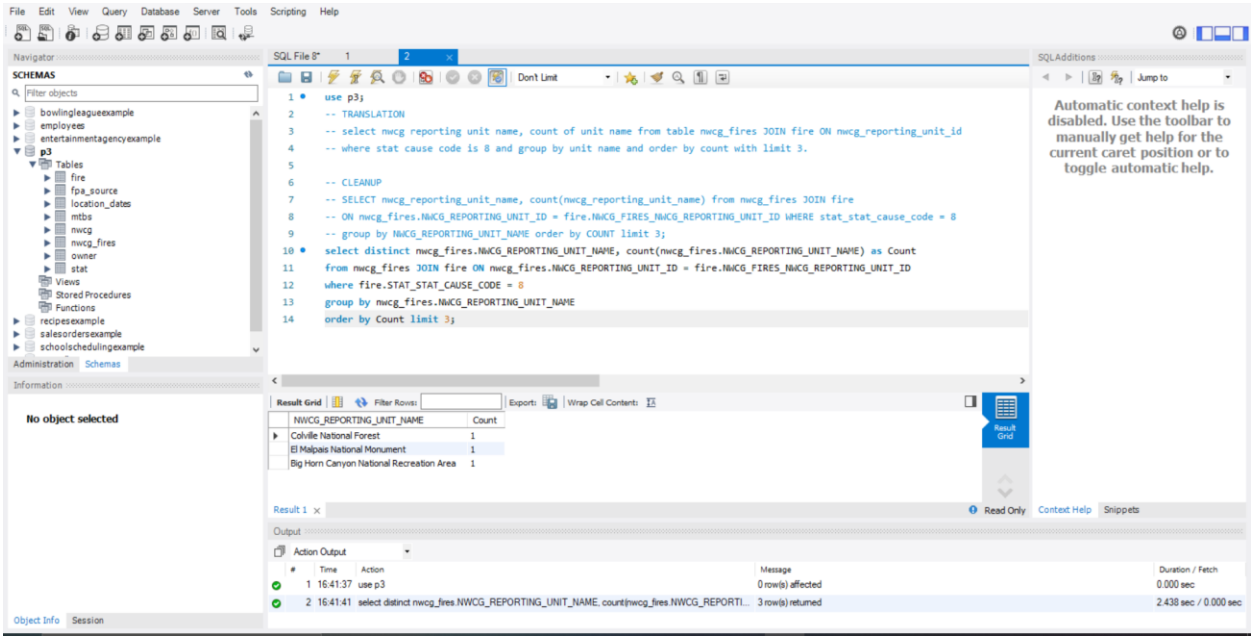
The forests which prohibits children to go alone are assumed to be the forests where stat_cause_code = 8, which means children.

ASSUMPTION: The forests where stat cause is children but the count is minimum is assumed to be least appropriate for 1 adult per 3 children because minimum occurrences means less threat.

Translation

select nwcg reporting unit name, count of unit name from table nwcg_fires JOIN fire ON nwcg_reporting_unit_id where stat cause code is 8 and group by unit name and order by count with limit 3.

Screen Shot of SQL Query and Results



Query 3

Question 3

One advocacy group says human actions and not Nature 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)

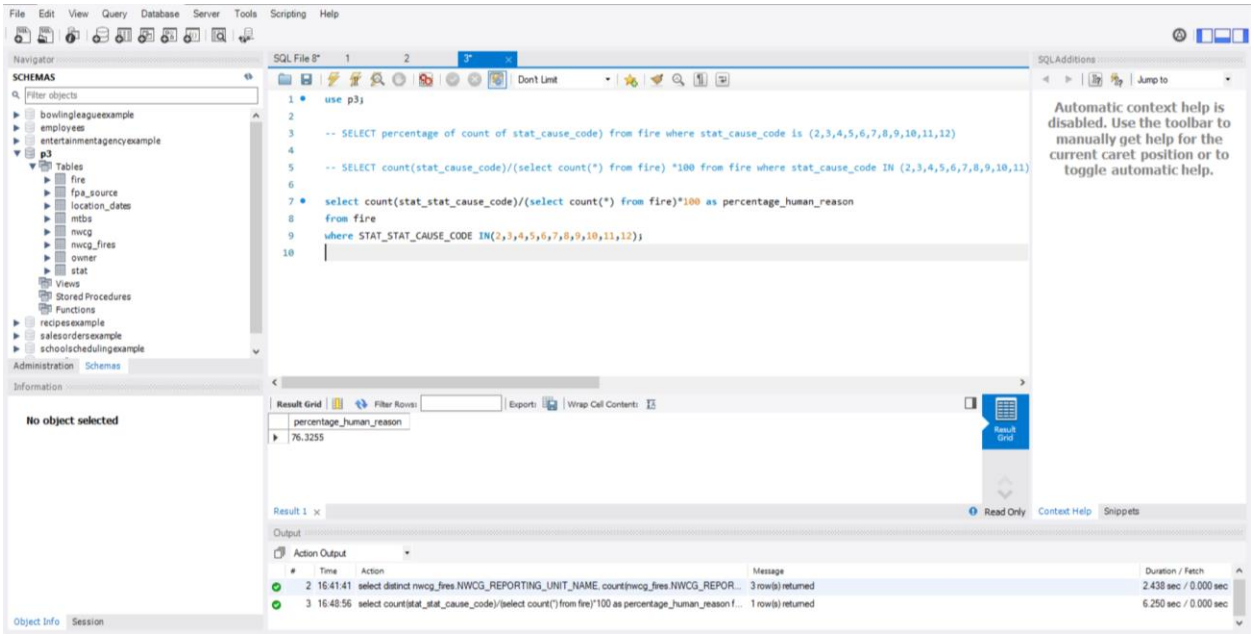
It is assumed that out of 13 stat causes only lightning (1) and Undefined/Missing (13) are the natural caused reasons. This also leads to assumption that Miscellaneous is a human action.

The result gives a percentage of count of human actions caused events and as the result is 76.3255 we can confidently support the statement that 'Human actions and not Nature is to blame for most wildfires'.

Translation

SELECT percentage of count of stat_cause_code) from fire where stat_cause_code is (2,3,4,5,6,7,8,9,10,11,12).

Screen Shot of SQL Query and Results



Query 4

Question 6

What were the forests that had no fires that lasted more than two days?

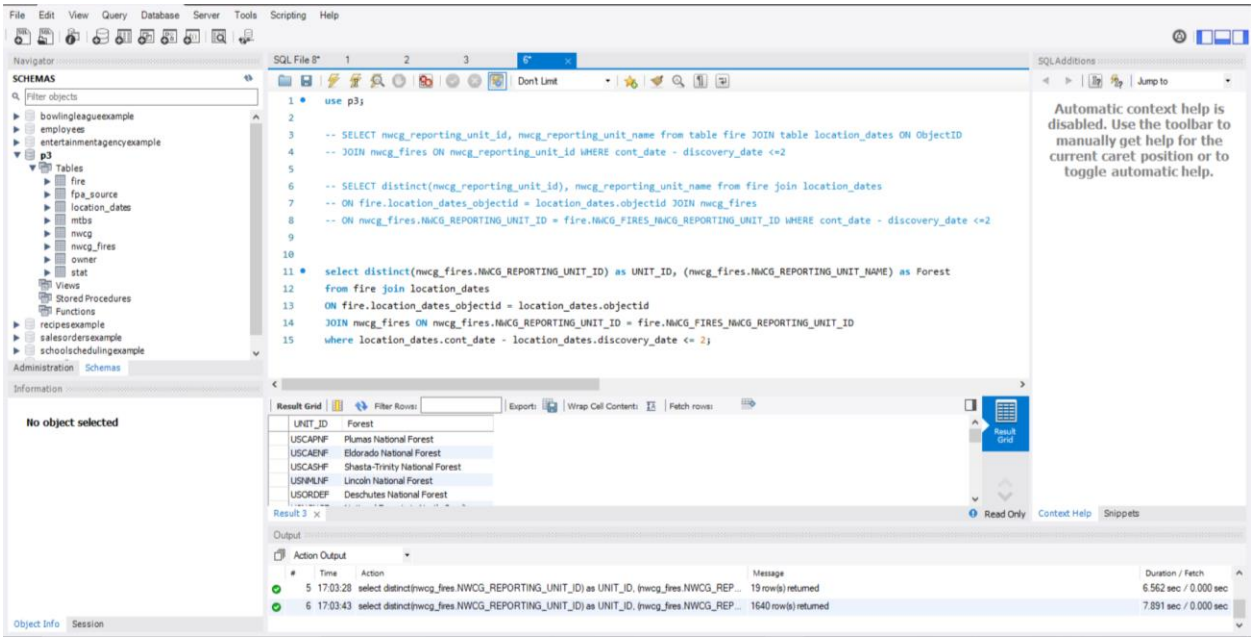
Notes/Comments About SQL Query and Results (Include # of Rows in Result)

To find the forests where no fire lasted more than two days assumption is made that the forest name lies in NWCG_REPORTING_UNIT_NAME. The result comprises of 1640 forests in which cont_date – discovery_date <= 2.

Translation

SELECT nwcg_reporting_unit_id, nwcg_reporting_unit_name from table fire JOIN table location_dates ON ObjectID JOIN nwcg_fires ON nwcg_reporting_unit_id WHERE cont_date - discovery_date <=2

Screen Shot of SQL Query and Results



Query 5

Question 7

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

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

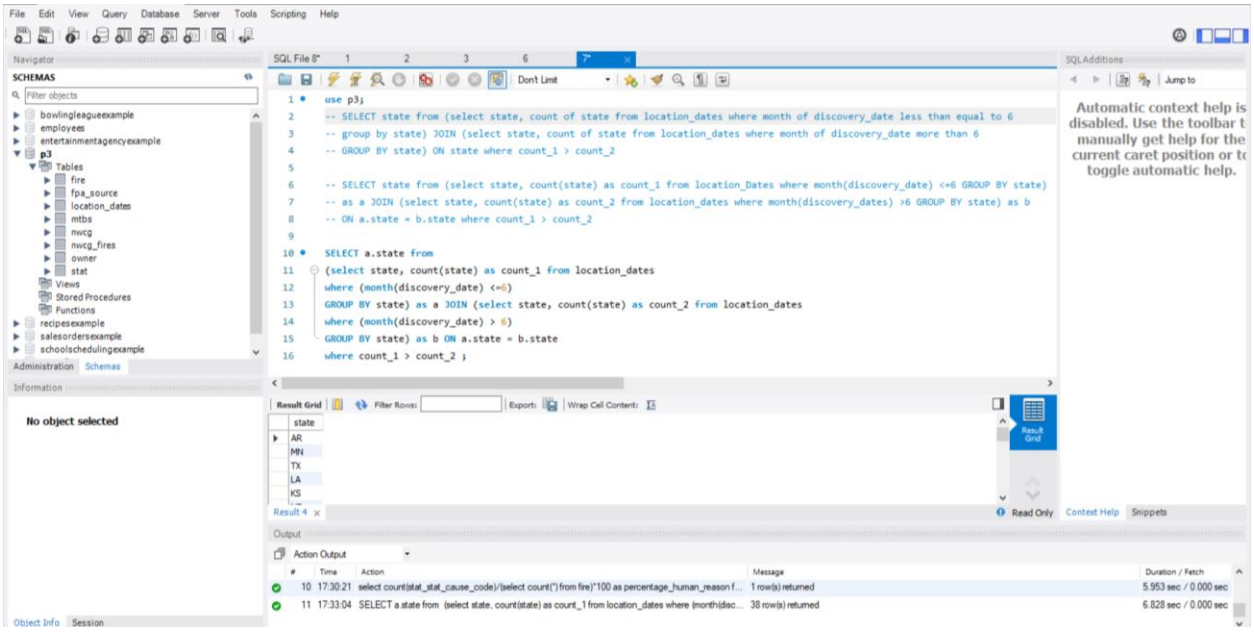
To compare the count of fires in each state by first half of the year to second half of the year, we make two subqueries delivering the count of first half and second half. If the count of the first half is more than the second half, we print the state.

The result comprises of 38 such states out of 52 total states where fires were more in first half of the year.

Translation

SELECT state from (select state, count of state from location_dates where month of discovery_date less than equal to 6 group by state) JOIN (select state, count of state from location_dates where month of discovery_date more than 6 GROUP BY state) ON state where count_1 > count_2

Screen Shot of SQL Query and Results



Query 6

Question 8

Which forest had the least number of fires?

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

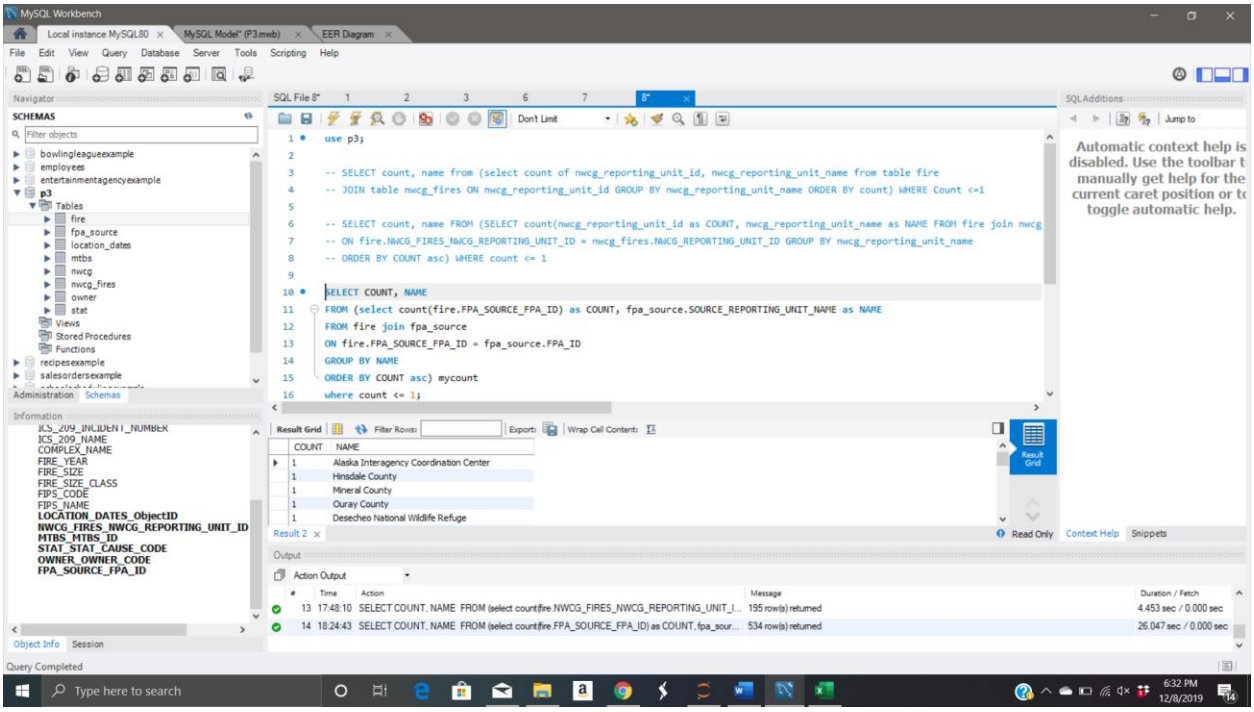
To find the forest with least number of fires, we assume that the forest names are given in the SOURCE_REPORTING_UNIT_NAME and also that while ordering it we assume the count <= 1 because in the main fire table, the least count of any forest reported is 1.

The result will publish the names of all the forests which have occurred 1 time in the mail table and they are thus considered to be the least in the count. There are 534 forest which only reported <= 1 fires.

Translation

select count, name from (select count of nwcg_reporting_unit_id, nwcg_reporting_unit_name from table fire JOIN table nwcg_fires ON nwcg_reporting_unit_id GROUP BY nwcg_reporting_unit_name ORDER BY count) WHERE Count <=1

Screen Shot of SQL Query and Results



Data Review for MongoDB

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

1. As FOD_ID is the global unique identifier and is non-null and unique, it is assumed to be the primary key of the main table. There are other foreign keys like STAT_CAUSE_ID, OWNER_CODE, MTBS_ID, NWCG_REPORTING_UNIT_ID, FPA_ID, OBJECTID which are the primary keys in other tables.
2. In Stat_Cause_Table, stat_cause_id is assumed to be the primary key as each key represent an individual stat_cause_descr, the stat_cause_id is unique and non_null and stat_cause_descr is directly dependant on the stat_cause_id.
3. In the Owner table, owner_code is assumed to be the primary key because it is non-null, unique and all the element of owner_descr depend directly on owner_code.
4. In the MTBS table, MTBS_ID directly depends on each unique MTBS_NAME. MTBS_ID is the primary key for this table because it is unique and non-null and every record of MTBS_Fire_Name directly depends on MTBS_ID.
5. The NWCG table is assumed to be a separate table because there are a lot of records in Unit_ID column, all of which are not used in the main table. NWCG table serves the purpose of all the records given by NWCG Reporting and all the columns from NWCG_FIRES, and NWCG are directly related to NWCG_REPORTING_Unit_ID. NWCG_REPORTING_Unit_ID, the primary key of the table is non-null and unique.
6. In the FPA_SOURCE table, FPA_ID is the primary key because all the other columns directly depend on the FPA_ID and the SOURCE_REPORTING_UNIT_ID cannot be a primary key because of its redundant values.
7. In the LOCATION_DATES table, OBJECTID is the primary key as all the other elements depends directly on the OBJECTID and OBJECTID is a non-null unique value column.

Physical Mongo Database

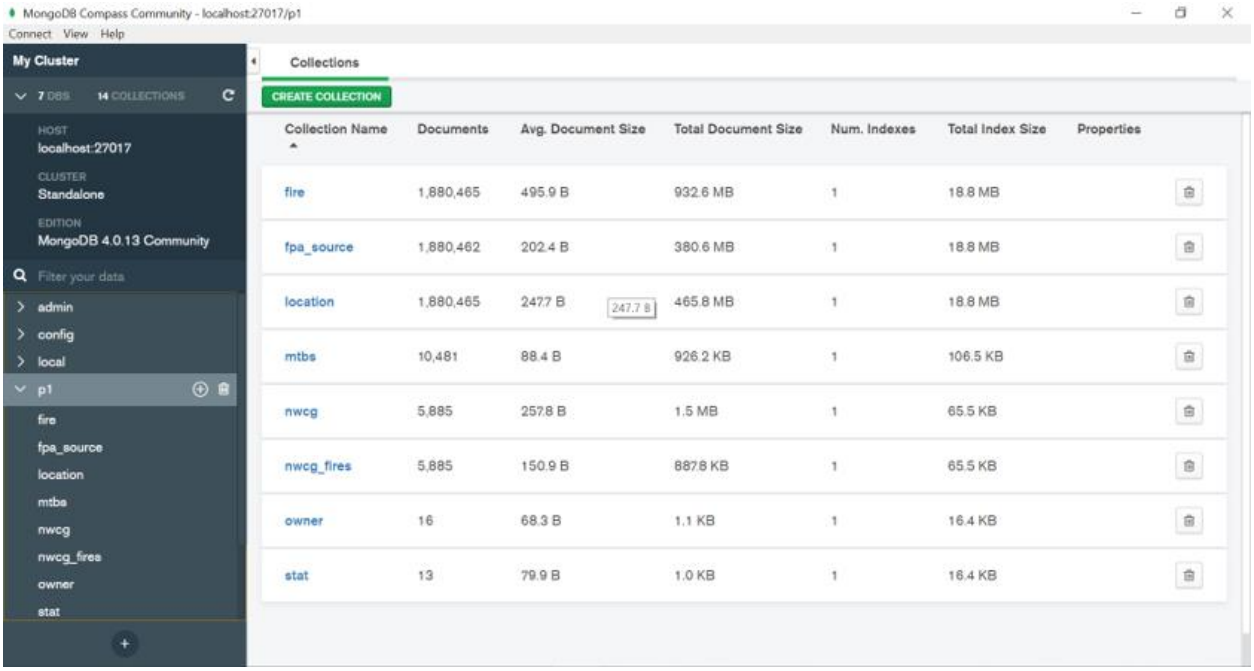
Assumptions/Notes About Data Set

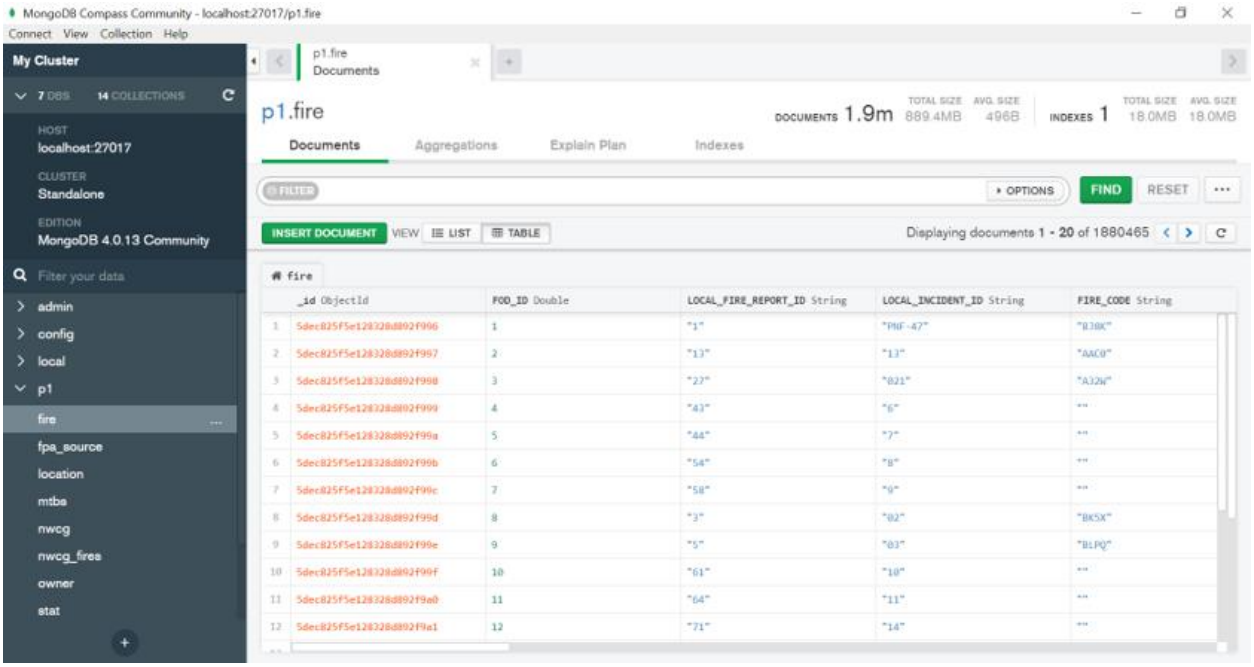
To organize the data effectively we divided the Fires table into 8 tables and joined the data with NWCG_UnitIDActive_20170109 table. The division was done to organise the data in 3rd Normalized Form.

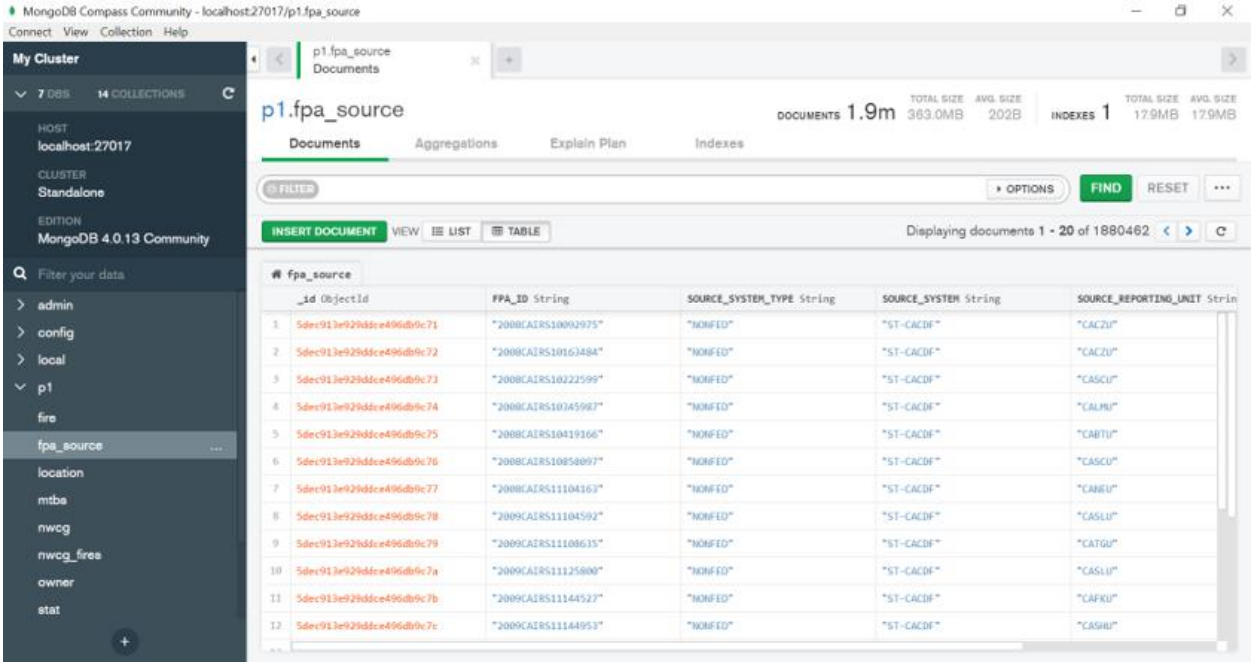
The assumptions were as follows:

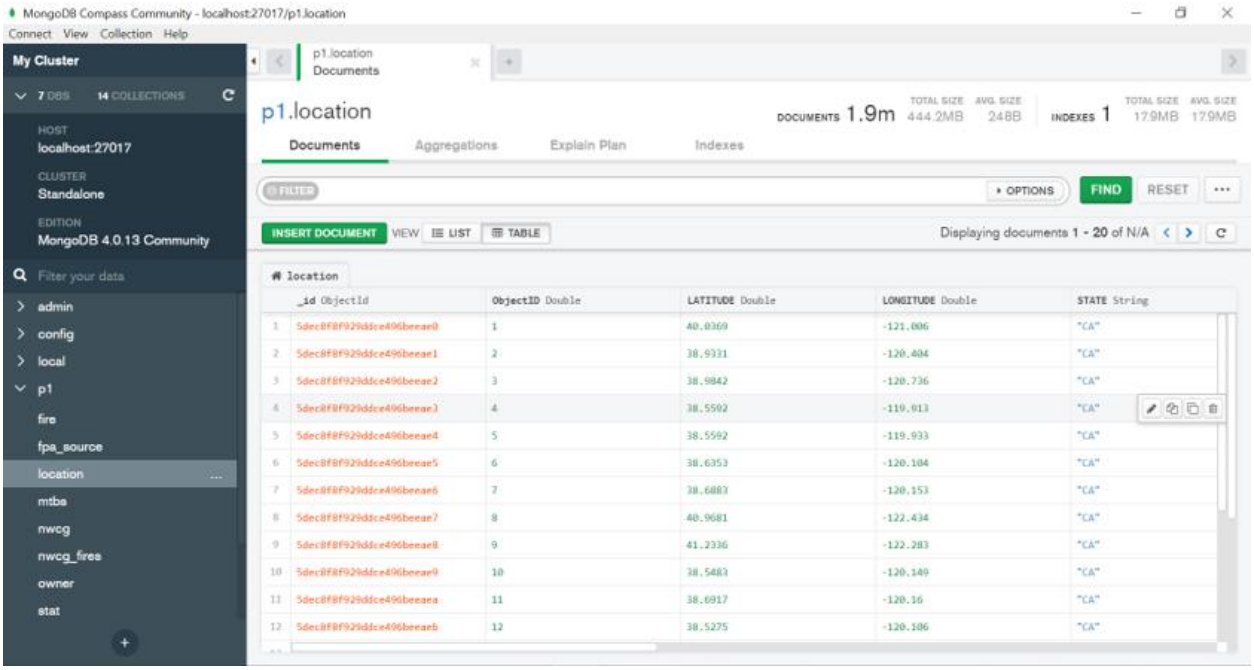
1. Data is provided only within Fires and NWCG-UnitIDActive_20170109 tables.
2. The collections are distributed to the best effort of reducing redundancy.

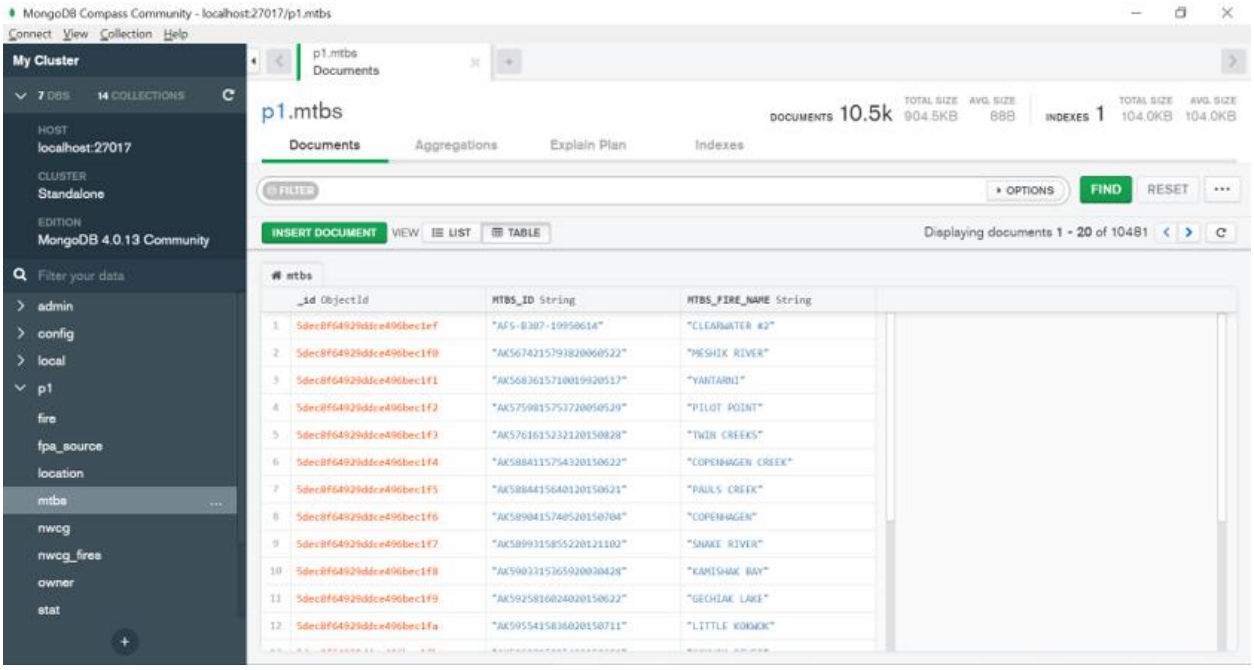
Screen shot of Physical Database objects (Database, Collections and Attributes)

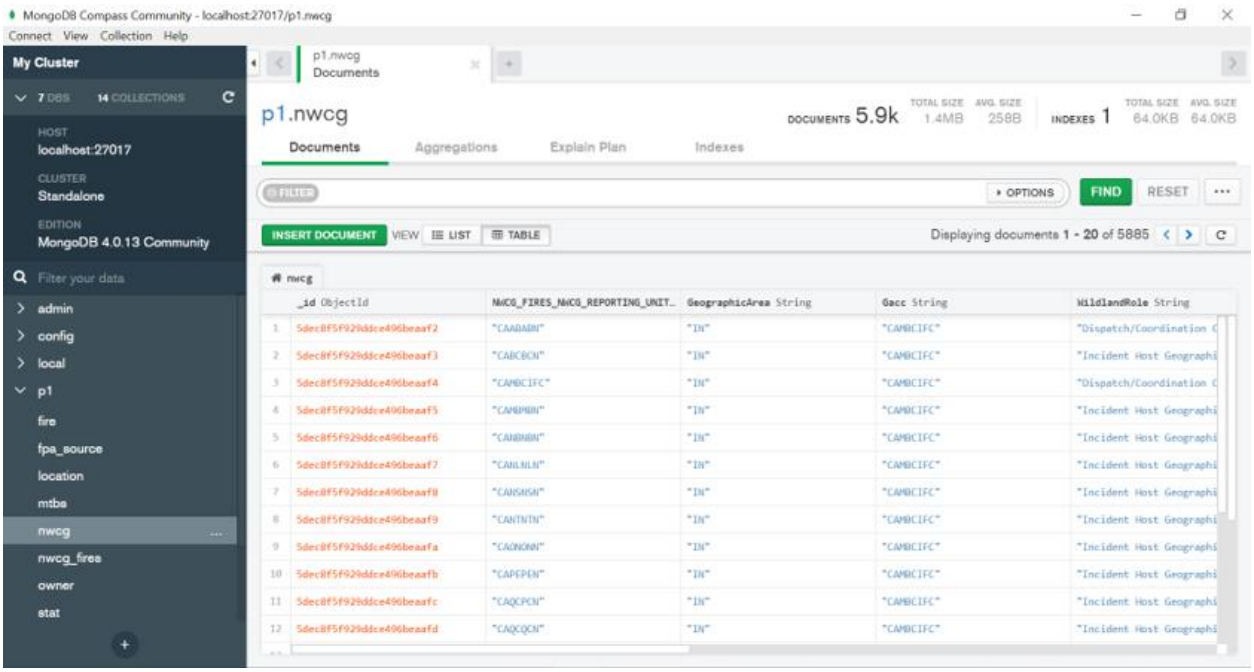


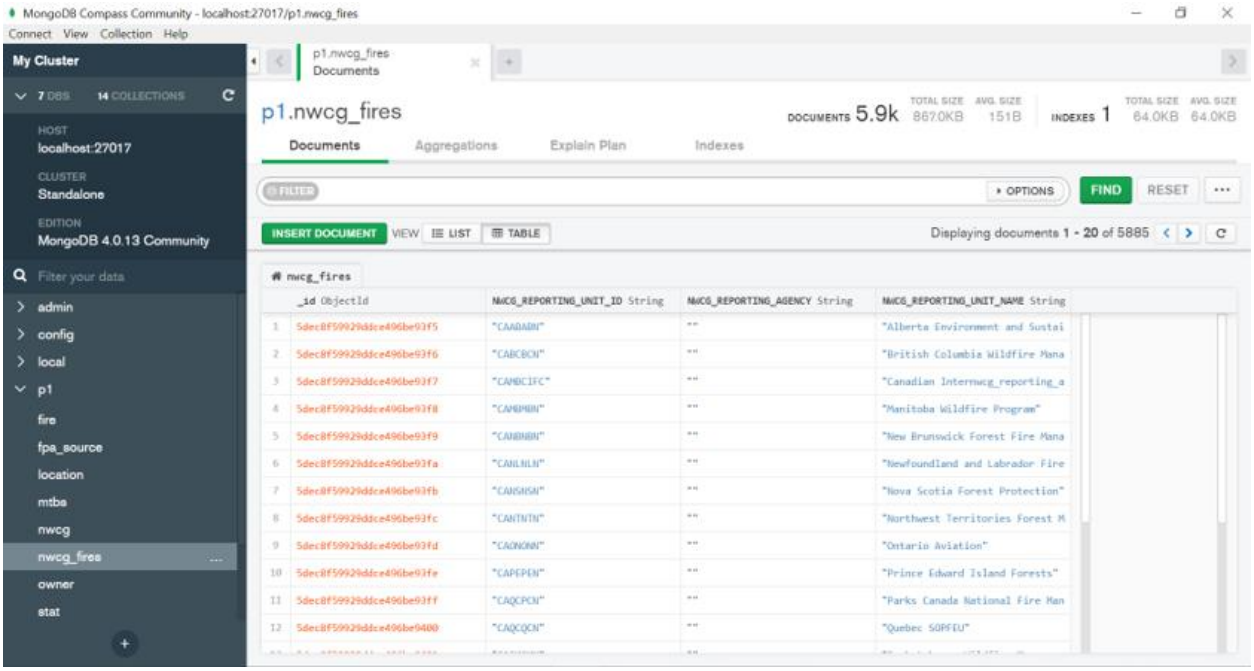

MongoDB Compass Community - localhost:27017/p1.fire
Connect View Collection Help
My Cluster
7 DBS 14 COLLECTIONS
HOST
localhost:27017
CLUSTER
Standalone
EDITION
MongoDB 4.0.13 Community
Filter your data
admin
config
local
p1
fire
fpa_source
location
mtbs
nwcg
nwcg_fires
owner
stat
p1.fire
Documents
p1.fire
DOCUMENTS 1.9m
TOTAL SIZE 889.4MB
AVG. SIZE 496B
INDEXES 1
TOTAL SIZE 18.0MB
AVG. SIZE 18.0MB
Documents
Aggregations
Explain Plan
Indexes
FILTER
OPTIONS
FIND
RESET
INSERT DOCUMENT
VIEW
LIST
TABLE
Displaying documents 1 - 20 of 1880465
fire
_id ObjectId
FOD_ID Double
LOCAL_FIRE_REPORT_ID String
LOCAL_INCIDENT_ID String
FIRE_CODE String

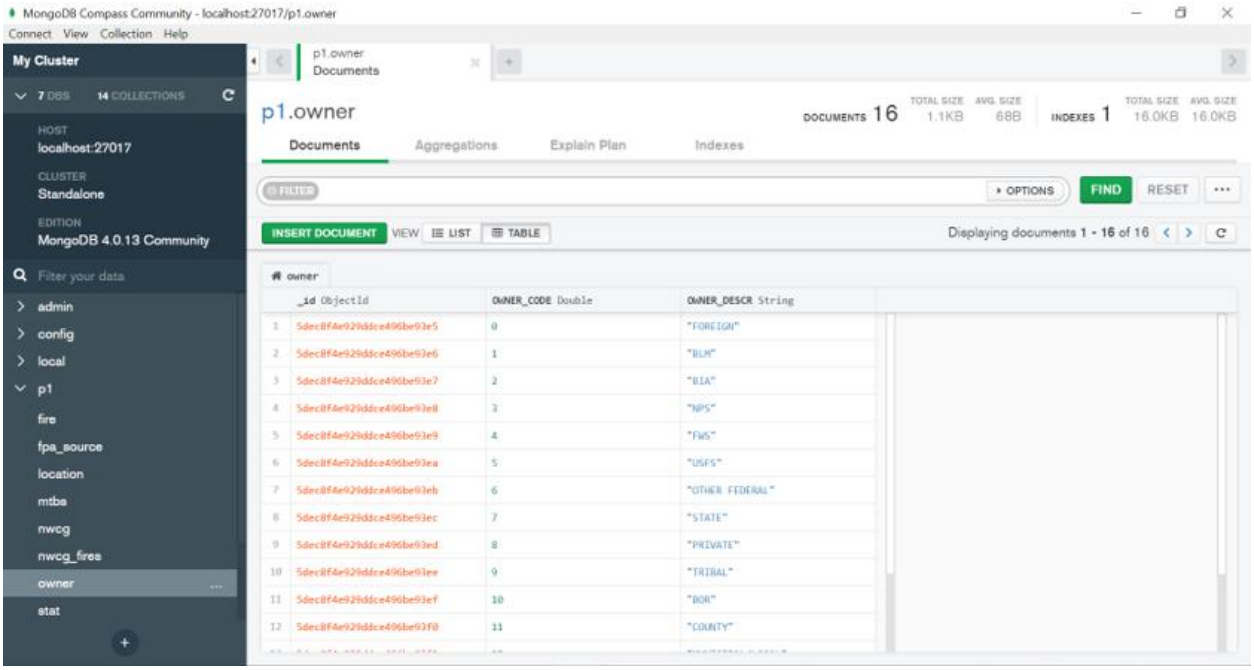

MongoDB Compass Community - localhost:27017/p1.fpa_source
Connect View Collection Help
My Cluster
7 DBS 14 COLLECTIONS
HOST
localhost:27017
CLUSTER
Standalone
EDITION
MongoDB 4.0.13 Community
Filter your data
admin
config
local
p1
fire
fpa_source
location
mtbs
nwcg
nwcg_fires
owner
stat
p1.fpa_source
Documents
p1.fpa_source
DOCUMENTS 1.9m
TOTAL SIZE 363.0MB
AVG. SIZE 202B
INDEXES 1
TOTAL SIZE 17.9MB
AVG. SIZE 17.9MB
Documents
Aggregations
Explain Plan
Indexes
FILTER
OPTIONS
FIND
RESET
INSERT DOCUMENT
VIEW
LIST
TABLE
Displaying documents 1 - 20 of 1880462
fpa_source
_id ObjectId
FPA_ID String
SOURCE_SYSTEM_TYPE String
SOURCE_SYSTEM String
SOURCE_REPORTING_UNIT Strin

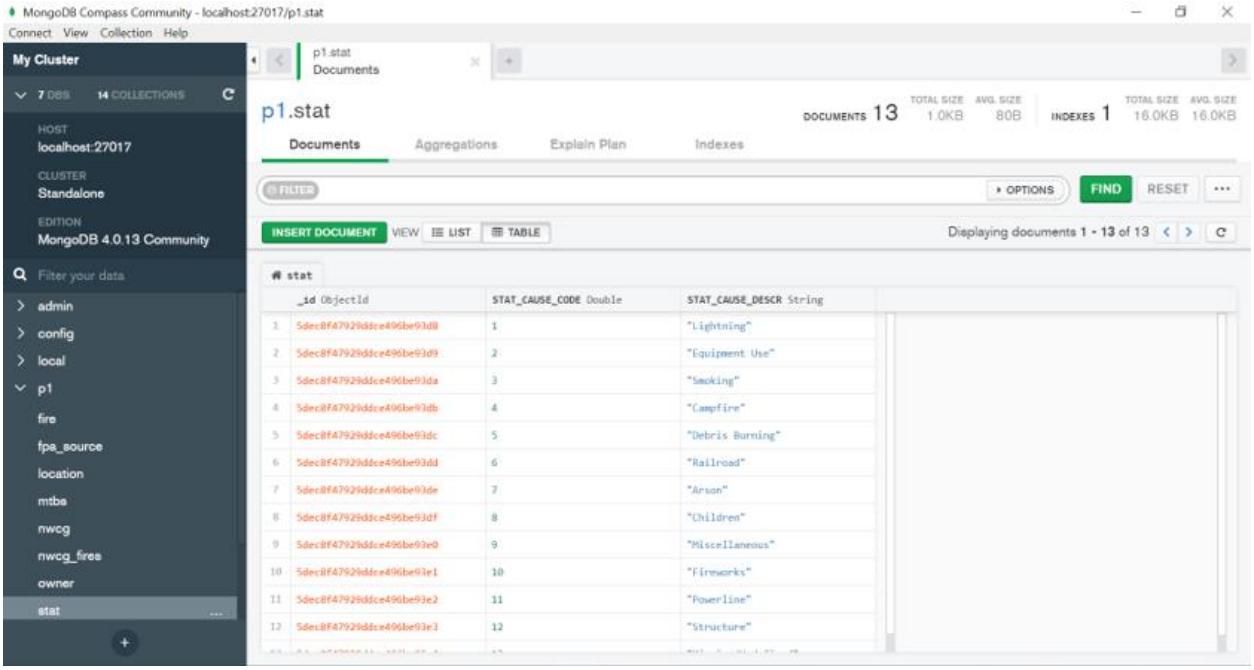

MongoDB Compass Community - localhost:27017/p1.location
Connect View Collection Help
My Cluster
7 DBS 14 COLLECTIONS
HOST
localhost:27017
CLUSTER
Standalone
EDITION
MongoDB 4.0.13 Community
Filter your data
admin
config
local
p1
fire
fpa_source
location
mtbs
nwcg
nwcg_fires
owner
stat
p1.location
Documents
p1.location
DOCUMENTS 1.9m
TOTAL SIZE 444.2MB
AVG. SIZE 248B
INDEXES 1
TOTAL SIZE 17.9MB
AVG. SIZE 17.9MB
Documents
Aggregations
Explain Plan
Indexes
FILTER
OPTIONS
FIND
RESET
INSERT DOCUMENT
VIEW
LIST
TABLE
Displaying documents 1 - 20 of N/A
location
_id ObjectId
ObjectID Double
LATITUDE Double
LONGITUDE Double
STATE String
1 5dec8f8f929ddce496beeae0 1 40.0369 -121.006 "CA"
2 5dec8f8f929ddce496beeae1 2 38.9331 -120.404 "CA"
3 5dec8f8f929ddce496beeae2 3 38.9842 -120.736 "CA"
4 5dec8f8f929ddce496beeae3 4 38.5592 -119.813 "CA"
5 5dec8f8f929ddce496beeae4 5 38.5592 -119.933 "CA"
6 5dec8f8f929ddce496beeae5 6 38.6353 -120.104 "CA"
7 5dec8f8f929ddce496beeae6 7 38.6883 -120.153 "CA"
8 5dec8f8f929ddce496beeae7 8 40.9681 -122.434 "CA"
9 5dec8f8f929ddce496beeae8 9 41.2336 -122.283 "CA"
10 5dec8f8f929ddce496beeae9 10 38.5483 -120.149 "CA"
11 5dec8f8f929ddce496beeaea 11 38.6917 -120.16 "CA"
12 5dec8f8f929ddce496beeaeb 12 38.5275 -120.106 "CA"


MongoDB Compass Community - localhost:27017/p1.mtbs
Connect View Collection Help
My Cluster
7 DBS 14 COLLECTIONS
HOST
localhost:27017
CLUSTER
Standalone
EDITION
MongoDB 4.0.13 Community
Filter your data
admin
config
local
p1
fire
fpa_source
location
mtbs
nwcg
nwcg_fires
owner
stat
p1.mtbs
Documents
p1.mtbs
DOCUMENTS 10.5k
TOTAL SIZE 904.5KB
AVG. SIZE 88B
INDEXES 1
TOTAL SIZE 104.0KB
AVG. SIZE 104.0KB
Documents
Aggregations
Explain Plan
Indexes
FILTER
OPTIONS
FIND
RESET
INSERT DOCUMENT
VIEW
LIST
TABLE
Displaying documents 1 - 20 of 10481
mtbs
_id ObjectId
MTBS_ID String
MTBS_FIRE_NAME String
1 5dec8f64929ddce496bec1ef "AFS-B307-19950614" "CLEARWATER #2"
2 5dec8f64929ddce496bec1f0 "AK567421579382006052 2" "MESHIK RIVER"
3 5dec8f64929ddce496bec1f1 "AK568361571001992051 7" "YANTARNI"
4 5dec8f64929ddce496bec1f2 "AK575981575372005052 9" "PILOT POINT"
5 5dec8f64929ddce496bec1f3 "AK576161523212015082 8" "TWIN CREEKS"
6 5dec8f64929ddce496bec1f4 "AK588411575432015062 2" "COPENHAGEN CREEK"
7 5dec8f64929ddce496bec1f5 "AK588441564012015062 1" "PAULS CREEK"
8 5dec8f64929ddce496bec1f6 "AK589041574052015070 4" "COPENHAGEN"
9 5dec8f64929ddce496bec1f7 "AK589931585522012110 2" "SNAKE RIVER"
10 5dec8f64929ddce496bec1f8 "AK590331536592002042 8" "KAMISHAK BAY"
11 5dec8f64929ddce496bec1f9 "AK592581602402015062 2" "GECHIAK LAKE"
12 5dec8f64929ddce496bec1fa "AK595541583602015071 1" "LITTLE KOKWOK"


MongoDB Compass Community - localhost:27017/p1.nwcg
Connect View Collection Help
My Cluster
7 DBS 14 COLLECTIONS
HOST
localhost:27017
CLUSTER
Standalone
EDITION
MongoDB 4.0.13 Community
Filter your data
admin
config
local
p1
fire
fpa_source
location
mtbs
nwcg
nwcg_fires
owner
stat
p1.nwcg
Documents
p1.nwcg
DOCUMENTS 5.9k
TOTAL SIZE 1.4MB
AVG. SIZE 258B
INDEXES 1
TOTAL SIZE 64.0KB
AVG. SIZE 64.0KB
Documents
Aggregations
Explain Plan
Indexes
FILTER
OPTIONS
FIND
RESET
INSERT DOCUMENT
VIEW
LIST
TABLE
Displaying documents 1 - 20 of 5885
nwcg
_id ObjectId
NWCG_FIRES_NWCG_REPORTING_UNIT...
GeographicArea String
Gacc String
WildlandRole String


MongoDB Compass Community - localhost:27017/p1.nwcg_fires
Connect View Collection Help
My Cluster
7 DBS 14 COLLECTIONS
HOST
localhost:27017
CLUSTER
Standalone
EDITION
MongoDB 4.0.13 Community
Filter your data
admin
config
local
p1
fire
fpa_source
location
mtbs
nwcg
nwcg_fires
owner
stat
p1.nwcg_fires
Documents
p1.nwcg_fires
DOCUMENTS 5.9k
TOTAL SIZE 867.0KB
AVG. SIZE 151B
INDEXES 1
TOTAL SIZE 64.0KB
AVG. SIZE 64.0KB
Documents
Aggregations
Explain Plan
Indexes
FILTER
OPTIONS
FIND
RESET
INSERT DOCUMENT
VIEW
LIST
TABLE
Displaying documents 1 - 20 of 5885
nwcg_fires
_id ObjectId
NWCG_REPORTING_UNIT_ID String
NWCG_REPORTING_AGENCY String
NWCG_REPORTING_UNIT_NAME String


MongoDB Compass Community - localhost:27017/p1.owner
Connect View Collection Help
My Cluster
7 DBS 14 COLLECTIONS
HOST
localhost:27017
CLUSTER
Standalone
EDITION
MongoDB 4.0.13 Community
Filter your data
admin
config
local
p1
fire
fpa_source
location
mtbs
nwcg
nwcg_fires
owner
stat
p1.owner
Documents
p1.owner
DOCUMENTS 16
TOTAL SIZE 1.1KB
AVG. SIZE 68B
INDEXES 1
TOTAL SIZE 16.0KB
AVG. SIZE 16.0KB
Documents
Aggregations
Explain Plan
Indexes
FILTER
OPTIONS
FIND
RESET
INSERT DOCUMENT
VIEW
LIST
TABLE
Displaying documents 1 - 16 of 16
owner
_id ObjectId
OWNER_CODE Double
OWNER_DESCR String
1 5dec8f4e929ddce496be93e5 0 "FOREIGN"
2 5dec8f4e929ddce496be93e6 1 "BLM"
3 5dec8f4e929ddce496be93e7 2 "BIA"
4 5dec8f4e929ddce496be93e8 3 "NPS"
5 5dec8f4e929ddce496be93e9 4 "FWS"
6 5dec8f4e929ddce496be93ea 5 "USFS"
7 5dec8f4e929ddce496be93eb 6 "OTHER FEDERAL"
8 5dec8f4e929ddce496be93ec 7 "STATE"
9 5dec8f4e929ddce496be93ed 8 "PRIVATE"
10 5dec8f4e929ddce496be93ee 9 "TRIBAL"
11 5dec8f4e929ddce496be93ef 10 "BOR"
12 5dec8f4e929ddce496be93f0 11 "COUNTY"


MongoDB Compass Community - localhost:27017/p1.stat
Connect View Collection Help
My Cluster
7 DBS 14 COLLECTIONS
HOST
localhost:27017
CLUSTER
Standalone
EDITION
MongoDB 4.0.13 Community
Filter your data
admin
config
local
p1
fire
fpa_source
location
mtbs
nwcg
nwcg_fires
owner
stat
p1.stat
Documents
p1.stat
DOCUMENTS 13
TOTAL SIZE 1.0KB
AVG. SIZE 80B
INDEXES 1
TOTAL SIZE 16.0KB
AVG. SIZE 16.0KB
Documents
Aggregations
Explain Plan
Indexes
FILTER
OPTIONS
FIND
RESET
INSERT DOCUMENT
VIEW
LIST
TABLE
Displaying documents 1 - 13 of 13
stat
_id ObjectId
STAT_CAUSE_CODE Double
STAT_CAUSE_DESCR String
1 5dec8f47929ddce496be93d8 1 "Lightning"
2 5dec8f47929ddce496be93d9 2 "Equipment Use"
3 5dec8f47929ddce496be93da 3 "Smoking"
4 5dec8f47929ddce496be93db 4 "Campfire"
5 5dec8f47929ddce496be93dc 5 "Debris Burning"
6 5dec8f47929ddce496be93dd 6 "Railroad"
7 5dec8f47929ddce496be93de 7 "Arson"
8 5dec8f47929ddce496be93df 8 "Children"
9 5dec8f47929ddce496be93e0 9 "Miscellaneous"
10 5dec8f47929ddce496be93e1 10 "Fireworks"
11 5dec8f47929ddce496be93e2 11 "Powerline"
12 5dec8f47929ddce496be93e3 12 "Structure"

Data in the Database

Collection Name	Relationshps With Other Collections (if any)	# of Documents in Collection
FIRE	-	1880465
FPA_SOURCE	-	1880462
LOCATION	-	1880465
MTBS	-	10481
NWCG	-	5885
NWCG_FIRES	-	5885
OWNER	-	16
STAT	-	13

MongoDB Queries/Code

Query 1

Question 8

Which forest had the least number of fires?

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

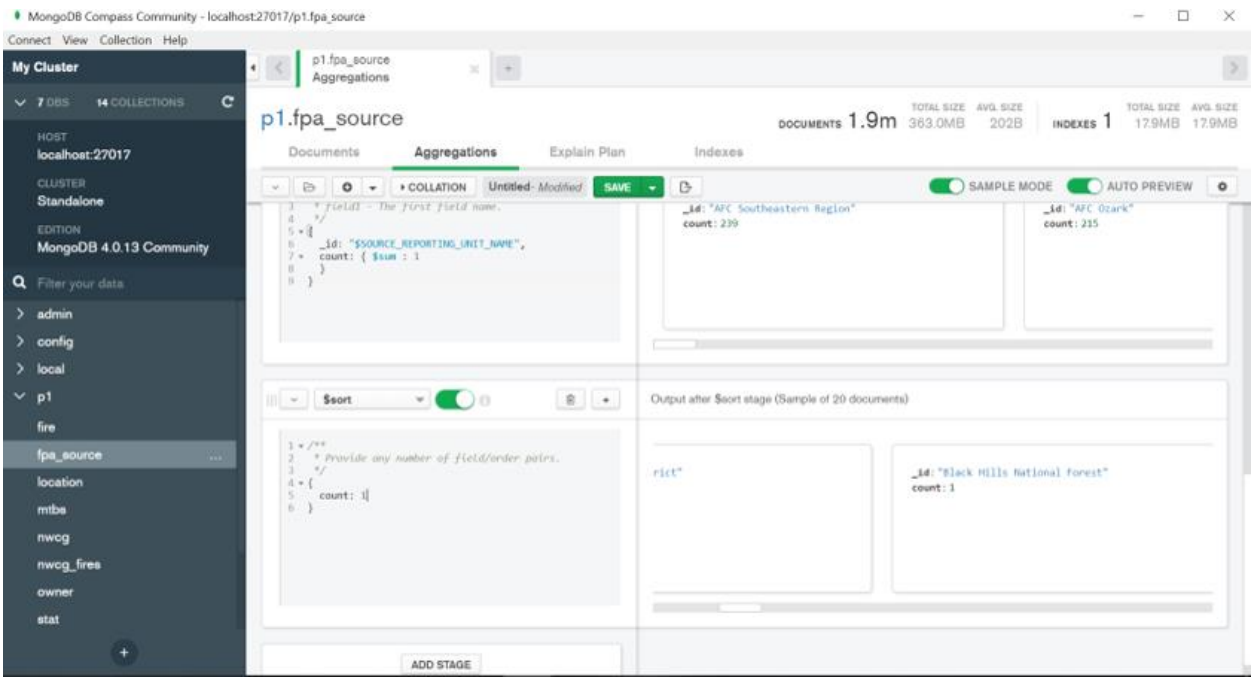
To find the name of forests with the least number of fires, we made an assumption that FPA_SOURCE_REPORTING_UNIT_NAME has all the forest names that are called in the main file.

Also, we assumed that count <=1 will give all the forest names that have the least occurrences in the main fire table as there can be no forest name with count 0 because that will not let it be in the main table. We do this by creating an aggregation pipeline and using lookup, group, sort functions in it.

Translation

Group SOURCE_REPORTING_UNIT_NAME along with the count and order it in ascending order.

Screen Shot of MongoDB Query/Code and Results



Query 2

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 least chance to win a share of the fund?

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

The state which had the least number of fires with stat_cause_code = 5, or 'debris burning' will have the least chance to win a share of fund. The result will contain 1 row stating the state having the least chance along with the count of debris burning.

Assumption: The fund shall be allocated to the state most affected state which have high number of debris burning in order to eliminate emergency and thus neglecting the ones least affected.

We use the $lookup function to associate location table with fire table, and then match the stat_stat_cause_code and group by state name with it's count.

Translation

Join the tables fire and locations using $lookup, define parameters for the lookup. Then, match the stat_stat_cause_code for debris burning. After this, group, the output using the state name and also display the count of state for occurences of stat_stat_cause_code.

Screen Shot of MongoDB Query/Code and Results

```
MongoDB shell version v4.0.13
connecting to: mongodb://127.0.0.1:27017/?gssapiServiceName=mongodb
Implicit session: session { "id" : UUID("08d2ae31-506e-4e17-bcb5-40253dbe68c3") }
MongoDB server version: 4.0.13
Server has startup warnings:
2019-12-08T19:43:22.084-0600 I CONTROL  [initandlisten]
2019-12-08T19:43:22.084-0600 I CONTROL  [initandlisten] ** WARNING: Access control is not enabled for the database.
2019-12-08T19:43:22.084-0600 I CONTROL  [initandlisten] **          Read and write access to data and configuration is unrestricted.
2019-12-08T19:43:22.084-0600 I CONTROL  [initandlisten]
---
Free Monitoring URL:
https://cloud.mongodb.com/freemonitoring/cluster/Q3ZFTYEJULVAGVRJTLZ6WDMEJI6BDDCS
---
> use p1;
switched to db p1
> db.location.aggregate([
...{ $lookup: {from: "fire",    localField: "ObjectID", foreignField: "location_dates_ObjectID", as: "fire" }
... },
...{ $match: { ["stat_stat_cause_code": 5] }
...},
...{ $group: { statename: "$state", "count": {$sum: 1}}
...}
...]).sort("count": 1);
{"STATE" : DC, "count(state)" : 2}
>
```

Query 3

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 least appropriate.

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

The forests which prohibits children to go alone are assumed to be the forests where stat_cause_code = 8, which means children.

ASSUMPTION: The forests where stat cause is children but the count is minimum is assumed to be least appropriate for 1 adult per 3 children because minimum occurrences means less threat to children

Translation

Create an aggregate pipeline using $lookup for nwcg_fires and fire tables. Then use $match to find the stat_stat_cause_code for children;ie:8. After this group the output, using NWCG_REPORTING_UNIT_NAME and it's count for stat_stat_cause_code.

Screen Shot of MongoDB Query/Code and Results

```
MongoDB shell version v4.0.13
connecting to: mongodb://127.0.0.1:27017/?gssapiServiceName=mongodb
Implicit session: session { "id" : UUID("08d2ae31-506e-4e17-bcb5-40253dbe68c3") }
MongoDB server version: 4.0.13
Server has startup warnings:
2019-12-08T19:43:22.084-0600 I CONTROL  [initandlisten]
2019-12-08T19:43:22.084-0600 I CONTROL  [initandlisten] ** WARNING: Access control is not enabled for the database.
2019-12-08T19:43:22.084-0600 I CONTROL  [initandlisten] **          Read and write access to data and configuration is unrestricted.
2019-12-08T19:43:22.084-0600 I CONTROL  [initandlisten]
---
Free Monitoring URL:
https://cloud.mongodb.com/freemonitoring/cluster/Q3ZFTYEJULVAGVRJTLZ6WDMEJI6BDDCS
---
> use p1;
switched to db p1
> db.nwcg_fires.aggregate([
...{ $lookup:{from: "fire", localField: "NWCG_REPORTING_UNIT_ID", foreignField: "NWCG_FIRES_NWCG_REPORTING_UNIT_ID",    as: "fire"  }
... },
...{$match: {"stat_stat_cause_code" : 8}
... },
...{ $group: {NWCG_REPORTING_UNIT_NAME: "$NWCG_REPORTING_UNIT_NAME", "count": {$sum: 1}}
... }
... ]).sort("count": 1), {$limit: 3};
{"NWCG_REPORTING_UNIT_NAME" : Colville National Forest, "Count" : 1}
{"NWCG_REPORTING_UNIT_NAME" : El Malpais National Monument, "Count" : 1}
{"NWCG_REPORTING_UNIT_NAME" : Big Horn Canyon National Recreation Area, "Count" : 1}
>
```

Query 4

Question 6

What were the forests that had no fires that lasted more than two days?

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

To find the forests where no fire lasted more than two days assumption is made that the forest name lies in NWCG_REPORTING_UNIT_NAME. The result comprises of 1640 forests in which cont_date – discovery_date <= 2.

Translation

This query also involves creating an aggregation pipeline, then using $lookup for fire and nwcg_fires tables. We need to use a second $lookup here for location and fire tables. Then match the difference of cont_date and discovery_date <= 2 and group the output using NWCG_REPORTING_UNIT_NAME.

Screen Shot of MongoDB Query/Code and Results

```
MongoDB shell version v4.0.13
connecting to: mongodb://127.0.0.1:27017/?gssapiServiceName=mongodb
Implicit session: session { "id" : UUID("08d2ae31-506e-4e17-bcb5-40253dbe68c3") }
MongoDB server version: 4.0.13
Server has startup warnings:
2019-12-08T19:43:22.084-0600 I CONTROL  [initandlisten]
2019-12-08T19:43:22.084-0600 I CONTROL  [initandlisten] ** WARNING: Access control is not enabled for the database.
2019-12-08T19:43:22.084-0600 I CONTROL  [initandlisten] **          Read and write access to data and configuration is unrestricted.
2019-12-08T19:43:22.084-0600 I CONTROL  [initandlisten]
---
Free Monitoring URL:
https://cloud.mongodb.com/freemonitoring/cluster/Q3ZFTYEJULVAGVRJTLZ6WDMEJI6BDDCS
---
> use p1;
switched to db p1
> db.fire.aggregate([
...{ $lookup: {from:"nwcg_fire", localField:"NWCG_FIRES_NWCG_REPORTING_UNIT_ID", foreignField:"NWCG_REPORTING_UNIT_ID", as:"nwcg_fire"}
... },
...{ $lookup: {from:"location", localField: "location_dates_objectid", foreignField: "objectid", as: "location" }
... },
...{ $match: {{$expr: {$lte:[{$subtract: ["$CONT_DATE","$DISCOVERY_DATE"] },2] }}}
... },
...{ $group: { "_id":  "$NWCG_REPORTING_UNIT_NAME"}}
...]);
{"UNIT_ID" : "USCAPNF", "Forest" : "Plumas National Forest"}
{"UNIT_ID" : "USCAENF", "Forest" : "Eldorado National Forest"}
{"UNIT_ID" : "USCASHF", "Forest" : "Shasta-Trinity National Forest"}
{"UNIT_ID" : "USNMLNF", "Forest" : "Lincoln National Forest"}
{"UNIT_ID" : "USORDEF", "Forest" : "Deschutes National Forest"}
{"UNIT_ID" : "USNCNCF", "Forest" : "National Forests in North Carolina"}
{"UNIT_ID" : "USWYMBF", "Forest" : "Medicine Bow National Forest"}
{"UNIT_ID" : "USORUMF", "Forest" : "Umatilla National Forest"}
{"UNIT_ID" : "USCOPSF", "Forest" : "Pike and San Isabel National Forest"}
{"UNIT_ID" : "USOROCF", "Forest" : "Ochoco National Forest"}
```

Query 5

Question 3

One advocacy group says human actions and not Nature 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)

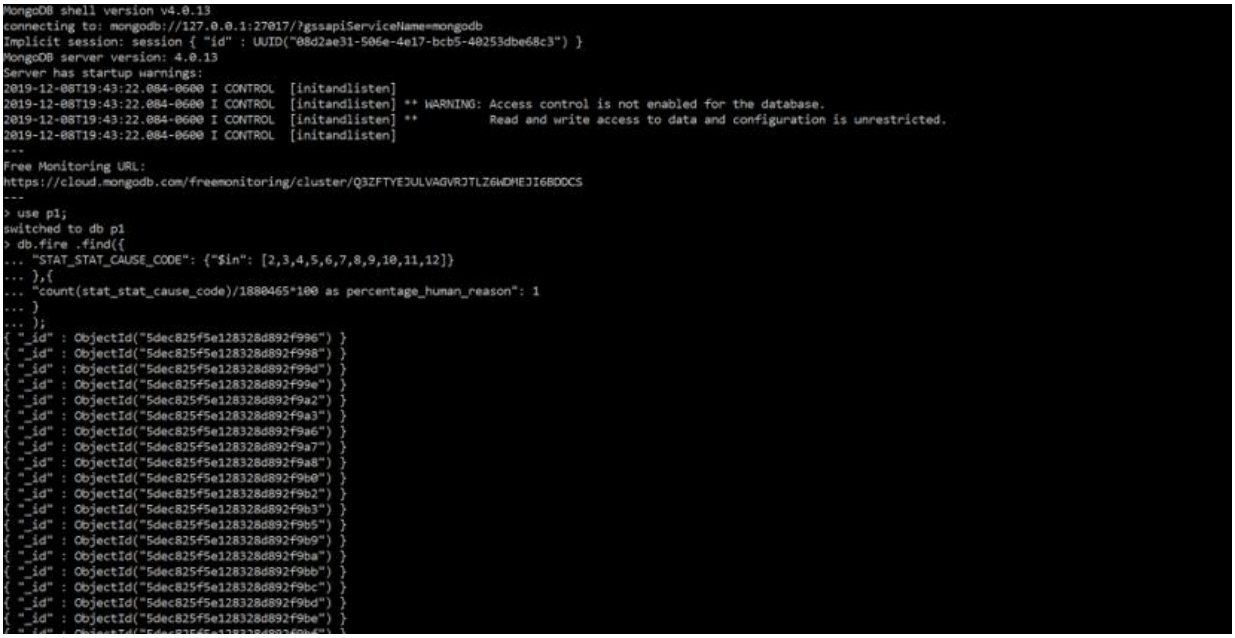
It is assumed that out of 13 stat causes only lightning (1) and Undefined/Missing (13) are the natural caused reasons. This also leads to assumption that Miscellaneous is a human action.

The result gives a percentage of count of human actions caused events and as the result is 76.3255 we can confidently support the statement that 'Human actions and not Nature is to blame for most wildfires'.

Translation

Use the db.collection.find() function to find the stat_cause_code that are relative to human causes and a percentage of that with respect to total number of records is found.

Screen Shot of MongoDB Query/Code and Results



```
MongoDB shell version v4.0.13
connecting to: mongodb://127.0.0.1:27017/?gssapiServiceName=mongodb
Implicit session: session { "id" : UUID("08d2ae31-506e-4e17-bcb5-40253dbe68c3") }
MongoDB server version: 4.0.13
Server has startup warnings:
2019-12-08T19:43:22.084-0600 I CONTROL  [initandlisten]
2019-12-08T19:43:22.084-0600 I CONTROL  [initandlisten] ** WARNING: Access control is not enabled for the database.
2019-12-08T19:43:22.084-0600 I CONTROL  [initandlisten] **          Read and write access to data and configuration is unrestricted.
2019-12-08T19:43:22.084-0600 I CONTROL  [initandlisten]
---
Free Monitoring URL:
https://cloud.mongodb.com/freemonitoring/cluster/Q3ZFTYEJULVAGVRJTLZ6WDMEJI6BODCS
---
> use p1;
switched to db p1
> db.fire .find({
... "STAT_STAT_CAUSE_CODE": {"$in": [2,3,4,5,6,7,8,9,10,11,12]}
... },{
... "count(stat_stat_cause_code)/1880465*100 as percentage_human_reason": 1
... }
... );
{ "_id" : ObjectId("5dec825f5e128328d892f996") }
{ "_id" : ObjectId("5dec825f5e128328d892f998") }
{ "_id" : ObjectId("5dec825f5e128328d892f99d") }
{ "_id" : ObjectId("5dec825f5e128328d892f99e") }
{ "_id" : ObjectId("5dec825f5e128328d892f9a2") }
{ "_id" : ObjectId("5dec825f5e128328d892f9a3") }
{ "_id" : ObjectId("5dec825f5e128328d892f9a6") }
{ "_id" : ObjectId("5dec825f5e128328d892f9a7") }
{ "_id" : ObjectId("5dec825f5e128328d892f9a8") }
{ "_id" : ObjectId("5dec825f5e128328d892f9b0") }
{ "_id" : ObjectId("5dec825f5e128328d892f9b2") }
{ "_id" : ObjectId("5dec825f5e128328d892f9b3") }
{ "_id" : ObjectId("5dec825f5e128328d892f9b5") }
{ "_id" : ObjectId("5dec825f5e128328d892f9b9") }
{ "_id" : ObjectId("5dec825f5e128328d892f9ba") }
{ "_id" : ObjectId("5dec825f5e128328d892f9bb") }
{ "_id" : ObjectId("5dec825f5e128328d892f9bc") }
{ "_id" : ObjectId("5dec825f5e128328d892f9bd") }
{ "_id" : ObjectId("5dec825f5e128328d892f9be") }
```

Query 6

Question 7

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

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

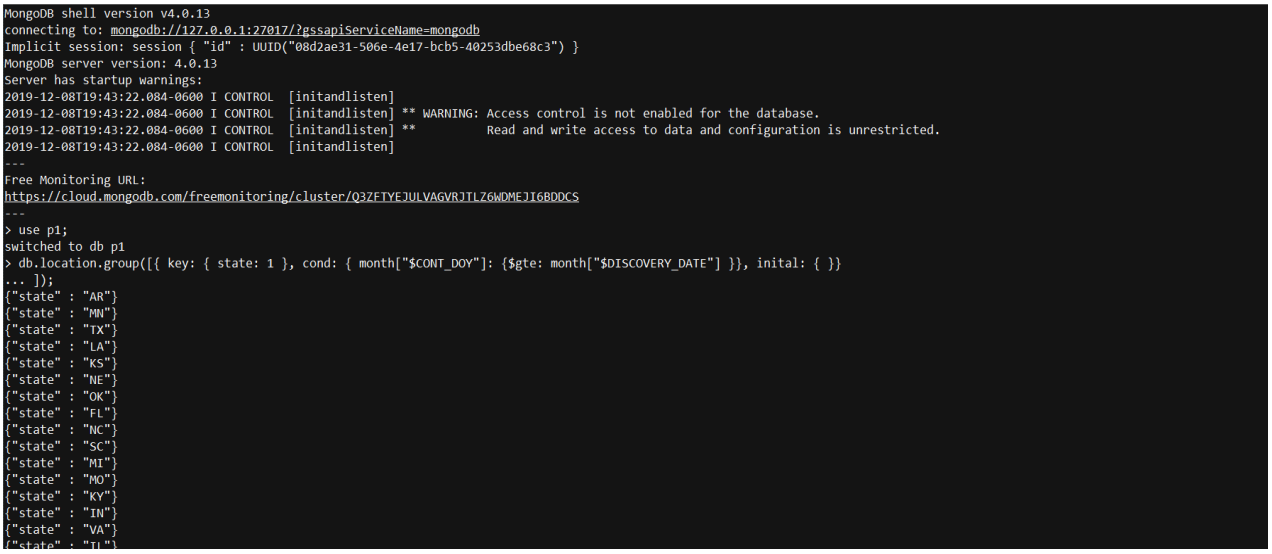
To compare the count of fires in each state by first half of the year to second half of the year, we make two subqueries delivering the count of first half and second half. If the count of the first half is more than the second half, we print the state.

The result comprises of 38 such states out of 52 total states where fires were more in first half of the year.

Translation

Using the db.collection.group function, we list out all the states where fires are more in the first half of a calendar year than the second half of the calendar year by defining a condition using $gte operator.

Screen Shot of MongoDB Query/Code and Results



```
MongoDB shell version v4.0.13
connecting to: mongodb://127.0.0.1:27017/?gssapiServiceName=mongodb
Implicit session: session { "id" : UUID("08d2ae31-506e-4e17-bcb5-40253dbe68c3") }
MongoDB server version: 4.0.13
Server has startup warnings:
2019-12-08T19:43:22.084-0600 I CONTROL  [initandlisten]
2019-12-08T19:43:22.084-0600 I CONTROL  [initandlisten] ** WARNING: Access control is not enabled for the database.
2019-12-08T19:43:22.084-0600 I CONTROL  [initandlisten] **          Read and write access to data and configuration is unrestricted.
2019-12-08T19:43:22.084-0600 I CONTROL  [initandlisten]
---
Free Monitoring URL:
https://cloud.mongodb.com/freemonitoring/cluster/Q3ZETYEJULVAGVRJTLZ6WDMEJI6BDDCS
---
> use p1;
switched to db p1
> db.location.group([{ key: { state: 1 }, cond: { month["$CONT_DOY"]: {$gte: month["$DISCOVERY_DATE"] }}, inital: { }}
... ]);
{"state" : "AR"}
{"state" : "MN"}
{"state" : "TX"}
{"state" : "LA"}
{"state" : "KS"}
{"state" : "NE"}
{"state" : "OK"}
{"state" : "FL"}
{"state" : "NC"}
{"state" : "SC"}
{"state" : "MI"}
{"state" : "MO"}
{"state" : "KY"}
{"state" : "IN"}
{"state" : "VA"}
{"state" : "IL"}
```