



# **SQL-Mongo Project**

## **Spatial Data of US Wildfires**

By  
Sinduja Senthil Kumar

Activity	Status
Prepared Data Model and Created Physical DB	✓ <input type="checkbox"/>
Loaded Data into Database	✓
Wrote SQL Queries	✓
Prepared Mongo Database	✓
Loaded data into Mongo DB	✓
Wrote Mongo Queries	✓
Prepared Report	✓
Reviewed Report	✓

## Contents

Data Model.....	5
Assumptions/Notes About Data Entities and Relationships.....	5
Entity-Relationship Diagram.....	7
Physical Database.....	8
Assumptions/Notes About Data Set.....	8
Screen shot of Physical Database objects.....	9
Data in the Database .....	25
SQL Queries .....	27
Query 1.....	27
Question: .....	27
Assumption:.....	27
Translation:.....	27
Cleanup:.....	27
Result:.....	27
Screen Shot of SQL Query and Results.....	28
Query 2.....	29
Question: .....	29
Notes/Comments About SQL Query and Results (Include # of Rows in Result) .....	29
Assumption:.....	29
Translation:.....	29
Cleanup:.....	29
Result:.....	29
Screen Shot of SQL Query and Results.....	30
Query 3.....	32
Question .....	32
Assumption:.....	32
Translation .....	32
Cleanup.....	32
Result:.....	32
Screen Shot of SQL Query and Results.....	33

Query 4.....	34
Question: .....	34
Assumption:.....	34
Translation:.....	34
Cleanup:.....	34
Result:.....	34
Screen Shot of SQL Query and Results.....	35
Query 5.....	36
Question: .....	36
Assumption:.....	36
Translation:.....	36
Cleanup:.....	36
Result:.....	36
Screen Shot of SQL Query and Results.....	37
Query 6.....	38
Question: .....	38
Assumption:.....	38
Translation .....	38
Cleanup.....	38
Result:.....	38
Screen Shot of SQL Query and Results.....	39
Data Review for MongoDB .....	41
Assumptions/Notes About Data Collections, Attributes and Relationships between Collections.....	41
Screen shot of Physical Database objects (Database, Collections and Attributes) .....	42
Data in the Database .....	43
MongoDB Queries/Code.....	44
Query 1.....	44
Question: .....	44
Translation .....	44
Cleanup:.....	44
Screen Shot of MongoDB Query/Code and Results .....	44
Notes/Comments About MongoDB Query/Code and Results (Include # of Documents in Result)...	44
Query 2.....	45

Question: .....	45
Translation:.....	45
Cleanup:.....	45
Screen Shot of MongoDB Query/Code and Results .....	45
Notes/Comments About MongoDB Query/Code and Results (Include # of Documents in Result)...	45
Query 3.....	46
Question: .....	46
Translation:.....	46
Screen Shot of MongoDB Query/Code and Results .....	46
Notes/Comments About MongoDB Query/Code and Results (Include # of Documents in Result)...	46
Query 4.....	47
Question: .....	47
Translation:.....	47
Screen Shot of MongoDB Query/Code and Results .....	47
Notes/Comments About MongoDB Query/Code and Results (Include # of Documents in Result)...	47
Query 5.....	48
Question: .....	48
Translation:.....	48
Screen Shot of MongoDB Query/Code and Results .....	48
Notes/Comments About MongoDB Query/Code and Results (Include # of Documents in Result)...	48
Query 6.....	49
Question: .....	49
Translation:.....	49
Screen Shot of MongoDB Query/Code and Results .....	49
Notes/Comments About MongoDB Query/Code and Results (Include # of Documents in Result)...	49

## Data Model

### Assumptions/Notes About Data Entities and Relationships

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

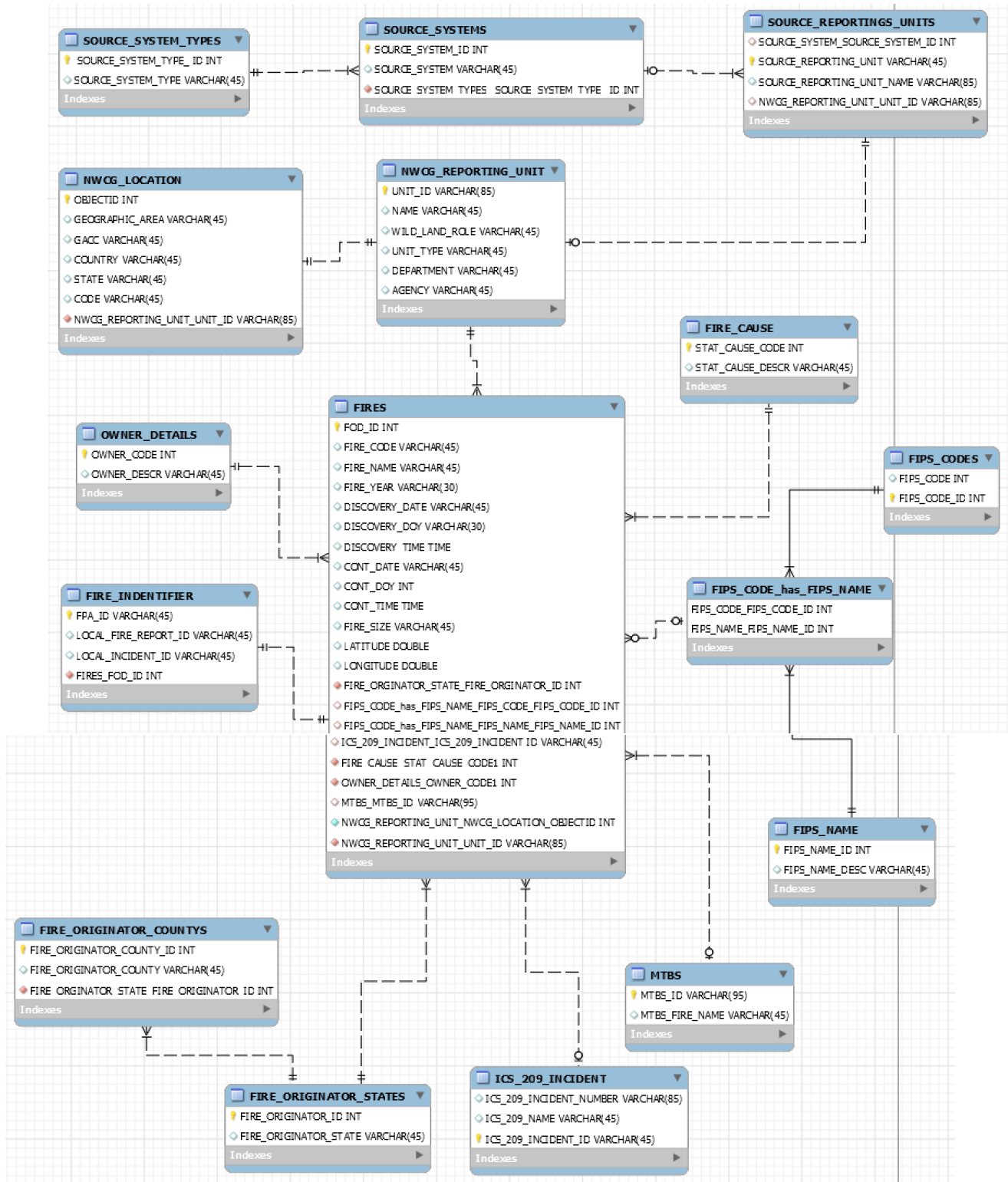
Assumptions about data entities and their relationships with each other:

1. There were 3 source system types (type of the source database) and for each type there were multiple source systems (Name/Identifier of the source database) that were linked. Hence separate tables as SOURCE\_SYSTEM\_TYPES and SOURCE\_SYSTEMS were used with generated primary key and corresponding relationship was given between the tables
2. Similarly, one source system was linked to many source reporting units (agency preparing the fire reports) and hence one to many relationships was given
3. The FIRE\_SIZE\_CLASS field can be calculated using the FIRE\_SIZE field while querying and hence it is dropped
4. The Julian dates in the FIRES table under DISCOVERY\_DATE and CONT\_DATE is converted and stored as normal dates as it would be easy for understanding as well as for querying
5. The county and the states column from the FIRES table are made as separate tables and are not added to owner's table because it is assumed that a owner can stay in a different place other than the fire location and manage the land
6. It is assumed that the state and county is not an redundant data though the latitude and longitude of the fire is mentioned separately because although the latitude and longitude gives the fire location, they may not necessarily give the point of origin of fire which might be given by the state and county
7. Separate table for FIPS name and FIPS code was generated because one FIPS name contained many FIPS code and one FIPS code contained many FIPS name resulting in the bridging table with one to one relationship
8. The parent field in the NWCG table didn't contain any values and hence it was dropped
9. Though the SOURCE\_REPORTING\_UNIT\_NAME is identical as the NWCG\_REPORTING\_UNIT\_NAME field and can be mapped one to one, there were certain values which were different. These differences were assumed to be manual typing errors and hence they were linked with one-to-one relationship
10. It is assumed that fields related to fire like the DISCOVERY DATE/DAY/TIME, CONTROLLED DATE/DAY/TIME, FIRE SIZE, LATITUDE, LONGITUDE etc., are the most important details that are required when querying about the fire and also this information seems mandatory to understand about the fire and hence they are grouped in a single table with the Global identifier as the primary key

The data model is in 3NF because of the following reasons:

1. Every table has its own primary key. The primary key was generated for tables like SOURCE\_SYSTEM\_TYPES, FIRE\_ORIGINATOR\_COUNTY etc., where there were no unique key columns. It is ensured that the values in the primary key column are unique and are not null.
2. It is ensured that there are no multi part or multi valued fields in any of the tables.  
Example: In table SOURCE\_REPORTING\_UNITS the attribute SOURCE\_REPORTING\_UNIT\_NAME had multi valued fields which were corrected  
Due to these reasons all tables are said to be in first Normal Form
3. It is ensured that every column other than the primary key is independent on the both the primary key columns when composite primary key is used. Thus, there is no partial dependency.  
Due to this reason all tables are said to be in second Normal Form
4. It is ensured that none of the tables have transitive dependency. That is, all the non key columns were purely made to depend on the primary key field, and they can cannot be derived or calculated using any other field. Example: The FIRE\_SIZE\_CLASS field is eliminated as it can be derived from the FIRE\_SIZE field which is non-primary key column  
Hence all tables are in third Normal form

## 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.

FPA\_FOD\_20170508 Database:

1. Though this database contains many tables, only the Fires and NWCG\_UnitIDActive\_20170109 tables was taken into consideration because,

- Tables such as Elementary Geometries, idx\_Fires\_Shape, KNN and Spatial Index were virtual tables and hence did not contain any information
- Tables such as geometry\_columns, geometry\_columns\_auth, geometry\_columns\_field\_infos, geometry\_columns\_statistics, geometry\_columns\_time, contain very less information ie only one row is present per table which is not relevant
- Tables such as idx\_Fires\_shape\_Node, idx\_Fires\_shape\_parent, idx\_Fires\_shape\_Rowid contain details about parent node columns which was also present in the NWCG table. However the parent column in the NWCG table itself is completely NULL which indicates that these data here will not be useful
- Tables such as spatial\_ref\_sys, spatial\_ref\_sys\_aux, spatialite\_history contains several rows with enough information. However, these tables are not related with Fires table and NWCG table which are our main tables. And so, we are not considering these tables
- Tables like views\_geometry\_columns, views\_geometry\_columns\_auth, views\_geometry\_columns\_field\_infos, views\_geometry\_columns\_statistics contain no data. Therefore, we are not considering these tables

2. Since the data dictionary indicates that NWCG\_UnitIDActive\_20170109 is used as the source for populating the following fields in the Fires table: NWCG\_REPORTING\_AGENCY, NWCG\_REPORTING\_UNIT\_ID, and NWCG\_REPORTING\_UNIT\_NAME, we are tempted to believe that the values from these columns present in the FIRES table would be a subset of NWCG\_UnitIDActive\_20170109 table. But however, there are values in NWCG\_REPORTING\_UNIT\_NAME tables which were not present in the main table

It is assumed that it could have happened because the NWCG\_UnitIDActive\_20170109 table contains all NWCG identifiers for agency units that were active/valid as of 9 January 2017. So, the identifiers for agency units which was present in the fires table could have been present before 2017

3. The FOD\_ID = Global unique identifier, though contained unique values to be used as a primary key, it skipped a value in between. (Which was 74)

## Screen shot of Physical Database objects

### 1. SOURCE\_SYSTEM\_TYPES

The screenshot displays a database query tool interface. At the top, a tab labeled "source\_system\_types" is active. Below the tab, a toolbar contains various icons for file operations, execution, and navigation. The query editor shows the following SQL statement:

```
1 • SELECT * FROM mydb.source_system_types;
```

Below the query editor, a "Result Grid" section displays the query results. The grid has two columns: "SOURCE\_SYSTEM\_TYPE\_ID" and "SOURCE\_SYSTEM\_TYPE". The results are as follows:

SOURCE_SYSTEM_TYPE_ID	SOURCE_SYSTEM_TYPE
1	FED
2	NONFED
3	INTERAGCY
NULL	NULL

At the bottom of the interface, an "Output" section is visible, showing the execution details of the query:

#	Time	Action	Message
1	09:36:23	SELECT * FROM mydb.source_system_types	3 row(s) returned

## 2. SOURCE\_SYSTEMS

Query 1 source\_systems x

1 • **SELECT \* FROM mydb.source\_systems;**

Result Grid | Filter Rows: | Edit: | Export/Import: | Wrap Cell Content

	SOURCE_SYSTEM_ID	SOURCE_SYSTEM	SOURCE_SYSTEM_TYPES SOURCE_SYSTEM_TYPE_ID
▶	1	FS-FIRESTAT	1
	2	DOI-WFMI	1
	3	FWS-FMIS	1
	4	FA-NFIRS	2
	5	ST-NASF	2
	6	ST-AZAZS	2
	7	ST-MOMOS	2
	8	IA-AKACC	3
	9	ST-MTMTS	2
	10	ST-SCSCS	2
	11	ST-COCOS	2
	12	ST-MEMES	2

source\_systems 1 x

Output

Action Output

#	Time	Action	Message
✓ 1	09:34:03	SELECT * FROM mydb.source_systems	31 row(s) returned

### 3. SOURCE\_REPORTING\_UNITS

The screenshot shows a SQL IDE interface with a query editor at the top and a results pane below. The query editor contains the following SQL statement:

```
1 • SELECT * FROM testemp.source_reportings_units;
```

The results pane displays a table with the following columns: `SOURCE_SYSTEM_SOURCE_SYSTEM_ID`, `SOURCE_REPORTING_UNIT`, `SOURCE_REPORTING_UNIT_NAME`, and `NWCG_REPORTING_UNIT_UNIT_ID`. The table contains 14 rows of data, representing various national forests and their reporting units.

SOURCE_SYSTEM_SOURCE_SYSTEM_ID	SOURCE_REPORTING_UNIT	SOURCE_REPORTING_UNIT_NAME	NWCG_REPORTING_UNIT_UNIT_ID
1	0419C	Wasatch-Cache-Uinta National Forest	USUTUWF
1	0610A	Rogue River-Siskiyou National Forest	USORRSF
1	1002	Tongass-Stikine National Forest	USAKTNF
1	1003	Tongass-Chatham National Forest	USAKTNF
1	1004	Chugach National Forest	USAKCGF
1	1005	Tongass National Forest	USAKTNF
1	102	Beaverhead-Deerlodge National Forest	USMTBDF
1	103	Bitterroot National Forest	USMTBRF
1	104	Idaho Panhandle National Forest	USIDIPF
1	105	Clearwater National Forest	USIDNCF
1	108	Custer National Forest	USMTCGF
1	109	Deerlodge National Forest	USMTBDF
1	110	Flathead National Forest	USMTNEF

The output pane at the bottom shows the execution of the query, indicating that 1404 row(s) were returned.

Output

Action Output

#	Time	Action	Message
1	16:59:14	SELECT * FROM testemp.source_reportings_units	1404 row(s) returned

#### 4. NWGC\_LOCATION

The screenshot shows a SQL IDE interface with multiple tabs at the top: 'SQL File 3', 'SQL File 3', 'SQL File 4', 'source\_reportings\_units', and 'nwgc\_location'. The active tab is 'nwgc\_location', which contains the SQL query: `1 • SELECT * FROM testemp.nwgc_location;`

Below the query editor is a toolbar with icons for saving, running, and other functions. The 'Result Grid' tab is selected, displaying a table of results. The table has 8 columns: OBJECTID, GEOGRAPHIC\_AREA, GACC, COUNTRY, STATE, CODE, and NWGC\_REPORTING\_UNIT\_UNIT\_ID. The results show 13 rows of data, all with GEOGRAPHIC\_AREA 'AK' and COUNTRY 'US'.

At the bottom, the 'Output' tab is selected, showing a log of actions. Two actions are listed: a successful query execution on 'source\_reportings\_units' at 16:59:14 returning 1404 rows, and a successful query execution on 'nwgc\_location' at 17:03:19 returning 5866 rows.

	OBJECTID	GEOGRAPHIC_AREA	GACC	COUNTRY	STATE	CODE	NWGC_REPORTING_UNIT_UNIT_ID
▶	1	AK	USAKACC	US	AK	CAN	USAKCAN
	2	AK	USAKACC	US	AK	CEKX	USAKCEKX
	3	AK	USAKACC	US	AK	CFFX	USAKCFFX
	4	AK	USAKACC	US	AK	CFMX	USAKCFMX
	5	AK	USAKACC	US	AK	CFQ	USAKCFQ
	6	AK	USAKACC	US	AK	CGF	USAKCGF
	7	AK	USAKACC	US	AK	CGFC	USAKCGFC
	8	AK	USAKACC	US	AK	CGFX	USAKCGFX
	9	AK	USAKACC	US	AK	CGKX	USAKCGKX
	10	AK	USAKACC	US	AK	CGMX	USAKCGMX
	11	AK	USAKACC	US	AK	CHCX	USAKCHCX
	12	AK	USAKACC	US	AK	CHN	USAKCHN
	13	AK	USAKACC	US	AK	CTN	USAKCTN

#	Time	Action	Message
✓ 1	16:59:14	SELECT * FROM testemp.source_reportings_units	1404 row(s) returned
✓ 2	17:03:19	SELECT * FROM testemp.nwgc_location	5866 row(s) returned

## 5. NWGC\_REPORTING\_UNIT

SQL File 3\* SQL File 3\* SQL File 4\* nwgc\_reporting\_unit x

1 • SELECT \* FROM testemp.nwgc\_reporting\_unit;

Result Grid Filter Rows: Edit: Export/Import: Wrap Cell Content: Fetch rows:

UNIT_ID	NAME	WILD_LAND_ROLE	UNIT_TYPE	DEPARTMENT	AGENCY
USCAXIM	Imperial County Operational Area	Incident Host Geographic	US State	CA	OES
USCAXIMC	Imperial County Operational Area Coordination	Dispatch/Coordination Center	US State	CA	OES
USCAXIN	Inyo County Operational Area	Incident Host Geographic	US State	CA	OES
USCAXINC	Inyo County Operational Area Coordination	Dispatch/Coordination Center	US State	CA	OES
USCAXKE	Kern County Operational Area	Incident Host Geographic	US State	CA	OES
USCAXKEC	Kern County Operational Area Coordination	Dispatch/Coordination Center	US State	CA	OES
USCAXKI	Kings County Operational Area	Incident Host Geographic	US State	CA	OES
USCAXKIC	Kings County Operational Area Coordination	Dispatch/Coordination Center	US State	CA	OES
USCAXLA	Los Angeles County Operational Area "A"	Incident Host Geographic	US State	CA	OES
USCAXLAC	Los Angeles Operational Area "A" Coordination	Dispatch/Coordination Center	US State	CA	OES
USCAXLB	Los Angeles County Operational Area "B"	Incident Host Geographic	US State	CA	OES
USCAXLBC	Los Angeles Operational Area "B" Coordination	Dispatch/Coordination Center	US State	CA	OES
USCAXLC	Los Angeles County Operational Area "C"	Incident Host Geographic	US State	CA	OES

porting\_unit 1 x Apply

Output

Action Output

#	Time	Action	Message	Duration /
✓ 1	16:59:14	SELECT * FROM testemp.source_reportings_units	1404 row(s) returned	0.000 sec
✓ 2	17:03:19	SELECT * FROM testemp.nwgc_location	5866 row(s) returned	0.047 sec
✓ 3	17:04:37	SELECT * FROM testemp.nwgc_reporting_unit	5866 row(s) returned	0.015 sec

## 6. OWNER\_DETAILS

The screenshot shows a SQL IDE interface with multiple tabs. The active tab is 'owner\_details'. The SQL editor contains the query: `1 • SELECT * FROM testemp.owner_details;`. Below the editor, the 'Result Grid' displays the query results. The grid has two columns: 'OWNER\_CODE' and 'OWNER\_DESCR'. The results are as follows:

	OWNER_CODE	OWNER_DESCR
▶	0	FOREIGN
	1	BLM
	2	BIA
	3	NPS
	4	FWS
	5	USFS
	6	OTHER FEDERAL
	7	STATE
	8	PRIVATE
	9	TRIBAL
	10	BOR
	11	COUNTY
	12	MUNICIPAL A.O.

Below the result grid, the 'Output' pane shows the 'Action Output' log. It contains four entries, each with a green checkmark icon, indicating successful execution:

#	Time	Action	Message
✓ 1	16:59:14	SELECT * FROM testemp.source_reportings_units	1404 row(s) returned
✓ 2	17:03:19	SELECT * FROM testemp.nwcg_location	5866 row(s) returned
✓ 3	17:04:37	SELECT * FROM testemp.nwcg_reporting_unit	5866 row(s) returned
✓ 4	17:07:40	SELECT * FROM testemp.owner_details	15 row(s) returned

## 7. FIRES

SQL File 3\* fires

1 • `SELECT * FROM testemp.fires;`

Result Grid

	FOD_ID	FIRE_CODE	FIRE_NAME	FIRE_YEAR	DISCOVERY_DATE	DISCOVERY_DOY	DISCOVERY_TIME	CONT_DATE	CONT_DOY	CONT_TIME	FIRE_SIZE
	32864	EL0N	KITSON	2008	9/24/2008	268	00:14:00	10/2/2008	276	00:09:30	804
	32865	EL4X	629	2008	9/26/2008	270	00:12:52	9/26/2008	270	00:18:00	0.1
	32866	ED4L	TELEPHONE	2008	7/28/2008	210	00:09:55	7/28/2008	210	00:12:00	3
	32867	DZ8H	DRY	2008	9/30/2008	274	00:11:22	9/30/2008	274	00:16:00	0.1
	32868	DZ95	INC 248	2008	8/20/2008	233	00:12:00	8/21/2008	234	00:18:00	0.5
	32869	EMY2	TENNANT	2008	10/7/2008	281	00:06:59	10/7/2008	281	00:17:30	0.25
	32870	DZ8H	BUCK	2008	9/10/2008	254	00:09:27	9/11/2008	255	00:13:00	0.1
	32871	DZ98	PAYNE	2008	7/9/2008	191	00:14:03	7/9/2008	191	00:15:30	2.5

fires 1 x

Output

Action Output

#	Time	Action	Message	Duration / F
✓ 1	16:59:14	SELECT * FROM testemp.source_reportings_units	1404 row(s) returned	0.000 sec /
✓ 2	17:03:19	SELECT * FROM testemp.nwcg_location	5866 row(s) returned	0.047 sec /
✓ 3	17:04:37	SELECT * FROM testemp.nwcg_reporting_unit	5866 row(s) returned	0.015 sec /
✓ 4	17:07:40	SELECT * FROM testemp.owner_details	15 row(s) returned	0.062 sec /
✓ 5	17:11:10	SELECT * FROM testemp.fires	1880465 row(s) returned	0.032 sec /



## 8. FIRE\_CAUSE

SQL File 3\* fire\_cause x

1 • `SELECT * FROM testemp.fire_cause;`

Result Grid

	STAT_CAUSE_CODE	STAT_CAUSE_DESCR
▶ 1		Lightning
2		Equipment Use
3		Smoking
4		Campfire
5		Debris Burning
6		Railroad
7		Arson
8		Children
9		Miscellaneous
10		Fireworks

fire\_cause 1 x

Output

Action Output

#	Time	Action	Message
✓ 1	16:59:14	SELECT * FROM testemp.source_reportings_units	1404 row(s) returned
✓ 2	17:03:19	SELECT * FROM testemp.nwgcg_location	5866 row(s) returned
✓ 3	17:04:37	SELECT * FROM testemp.nwgcg_reporting_unit	5866 row(s) returned
✓ 4	17:07:40	SELECT * FROM testemp.owner_details	15 row(s) returned
✓ 5	17:11:10	SELECT * FROM testemp.fires	1880465 row(s) returned
✓ 6	17:15:50	SELECT * FROM testemp.fire_cause	13 row(s) returned

## 9. FIRE\_IDENTIFIER

1 • `SELECT * FROM testemp.fire_identifier;`

Result Grid | Filter Rows: | Edit: | Export/Import:

FPA_ID	LOCAL_FIRE_REPORT_ID	LOCAL_INCIDENT_ID	FIRES_FOD_ID
2008CAIRS10092975	129396	5238	300313922
2008CAIRS10163484	131984	5581	300343761
2008CAIRS10222599	109872	2548	300331400
2008CAIRS10345987	134347	2725	300346248
2008CAIRS10419166	127648	7089	300338871
2008CAIRS10858097	134214	3094	300334020
2008CAIRS11104163	119089	18862	300321493
2009CAIRS11104592	238695	103	300311735
2009CAIRS11108635	235873	212	300344046
2009CAIRS11125800	215855	233	300341663
2009CAIRS11144527	220011	681	300311697
2009CAIRS11144953	211055	250	300341674
2009CAIRS11144959	236926	239	300331694
2009CAIRS11154620	226048	447	300314217

re\_identifier 1 x

Output

Action Output

#	Time	Action	Message
155	19:29:49	SELECT distinct fire_originator_states.fire_originato...	Error Code: 1054. Unknown c
156	19:29:57	SELECT distinct fire_originator_states.fire_originato...	1 row(s) returned
157	19:34:06	SELECT source_reporting_unit_name COUNT(*) F...	Error Code: 1064. You have a
158	19:35:12	SELECT source_reporting_unit_name, COUNT(*) ...	1010 row(s) returned
159	19:37:08	SELECT source_reporting_unit_name, COUNT(*) ...	1 row(s) returned
160	17:47:20	SELECT * FROM testemp.fire_identifier	1879442 row(s) returned

## 10. FIPS\_CODES

The screenshot shows a SQL IDE window titled "SQL File 3\*" with a tab for "fips\_codes". The query editor contains the SQL statement: `SELECT * FROM testemp.fips_codes;`. Below the editor, the "Result Grid" displays the query results. The grid has two columns: "FIPS\_CODE" and "FIPS\_CODE\_ID". The results are as follows:

	FIPS_CODE	FIPS_CODE_ID
▶	63	1
	61	2
	17	3
	3	4
	5	5
	27	6
	21	7
	113	8
	11	9
	9	10
	69	11
	37	12
	33	13

Below the result grid, the "Output" pane shows the "Action Output" for the query execution:

#	Time	Action	Message
✓ 1	17:41:23	SELECT * FROM testemp.fips_codes	277 row(s) returned
✓ 2	17:43:54	SELECT * FROM testemp.fire_cause	13 row(s) returned
✓ 3	17:46:32	SELECT * FROM testemp.fips_codes	277 row(s) returned

## 11. FIPS\_NAME

SQL File 3\* fips\_codes fips\_name x

1 • `SELECT * FROM testemp.fips_name;`

Result Grid

	FIPS_NAME_ID	FIPS_NAME_DESC
▶	1	Plumas
	2	Placer
	3	El Dorado
	4	Alpine
	5	Amador
	6	Lincoln
	7	Deschutes
	8	Buncombe
	9	Macon
	10	Caldwell
	11	Avery
	12	Converse
	13	Wheeler

fips\_name 1 x

Output

Action Output

#	Time	Action	Message
✓ 1	17:41:23	SELECT * FROM testemp.fips_codes	277 row(s) returned
✓ 2	17:43:54	SELECT * FROM testemp.fire_cause	13 row(s) returned
✓ 3	17:46:32	SELECT * FROM testemp.fips_codes	277 row(s) returned
✓ 4	17:47:38	SELECT * FROM testemp.fips_name	1425 row(s) returned

## 12. FIPS\_CODE\_has\_FIPS\_NAME

SQL File 3\* fips\_code\_has\_fips\_name x

1 • `SELECT * FROM testemp.fips_code_has_fips_name;`

Result Grid Filter Rows: Edit: Export/Import: Wrap Cell Content

	FIPS_CODE_FIPS_CODE_ID	FIPS_NAME_FIPS_NAME_ID
▶	1	1
	2	2
	3	3
	4	4
	5	5
	1	6
	2	6
	3	6
	6	6
	8	6
	11	6
	14	6
	18	6
	-	-

as\_fips\_name 1 x

Output

Action Output

	#	Time	Action	Message
✓	1	17:41:23	SELECT * FROM testemp.fips_codes	277 row(s) returned
✓	2	17:43:54	SELECT * FROM testemp.fire_cause	13 row(s) returned
✓	3	17:46:32	SELECT * FROM testemp.fips_codes	277 row(s) returned
✓	4	17:47:38	SELECT * FROM testemp.fips_name	1425 row(s) returned
✓	5	17:48:15	SELECT * FROM testemp.fips_code_has_fips_name	2120 row(s) returned

### 13. MTBS

SQL File 3\* mtbs

1 • `SELECT * FROM testemp.mtbs;`

Result Grid

MTBS_ID	MTBS_FIRE_NAME
AK5674215793820060522	MESHIK RIVER
AK5759815753720050529	PILOT POINT
AK5903315365920030428	KAMISHAK BAY
AK5975315141520050429	TRACY AVE
AK5978615491320030523	ILIAMNA
AK6002415119720070619	CARIBOU HILLS
AK6011915081019940830	WINDY POINT
AK6012815061920040814	GLACIER CREEK
AK6015115098020050711	FOX CREEK WFU
AK6037615081420090630	SHANTA CREEK
AK6038615043820070729	SWAN LAKE
AK6040215016520050706	IRISH CHANNEL
AK6041514945920010625	KENAI LAKE

mtbs 1 x

Output

Action Output

#	Time	Action	Message
1	17:41:23	SELECT * FROM testemp.fips_codes	277 row(s) returned
2	17:43:54	SELECT * FROM testemp.fire_cause	13 row(s) returned
3	17:46:32	SELECT * FROM testemp.fips_codes	277 row(s) returned
4	17:47:38	SELECT * FROM testemp.fips_name	1425 row(s) returned
5	17:48:15	SELECT * FROM testemp.fips_code_has_fips_name	2120 row(s) returned
6	17:49:25	SELECT * FROM testemp.mtbs	6655 row(s) returned

#### 14. FIRE\_ORIGINATOR\_COUNTYS

SQL File 3\* fire\_originator\_countys x

1 • `SELECT * FROM testemp.fire_originator_countys;`

Result Grid Filter Rows: Edit: Export/Import: Wrap Cell Content

	FIRE_ORIGINATOR_COUNTY_ID	FIRE_ORIGINATOR_COUNTY	FIRE_ORGINATOR_STATE_FIRE_ORIGINATOR_IC
▶	1	1	10
	2	1	8
	3	1	2
	4	1	14
	5	1	9
	6	1	5
	7	1	39
	8	1	18
	9	1	3
	10	1	24
	11	1	41
	12	1	12
	13	2	3
	...	...	...

ator\_countys1 x

Output

Action Output

#	Time	Action	Message
✓	1 17:41:23	SELECT * FROM testemp.fips_codes	277 row(s) returned
✓	2 17:43:54	SELECT * FROM testemp.fire_cause	13 row(s) returned
✓	3 17:46:32	SELECT * FROM testemp.fips_codes	277 row(s) returned
✓	4 17:47:38	SELECT * FROM testemp.fips_name	1425 row(s) returned
✓	5 17:48:15	SELECT * FROM testemp.fips_code_has_fips_name	2120 row(s) returned
✓	6 17:49:25	SELECT * FROM testemp.mtbs	6655 row(s) returned
✓	7 17:49:58	SELECT * FROM testemp.mtbs	6655 row(s) returned
✓	8 17:50:41	SELECT * FROM testemp.fire_originator_countys	2891 row(s) returned

## 15. FIRE\_ORIGINATOR\_STATES

SQL File 3\*    fire\_originator\_countys    **fire\_originator\_states** x

1 • `SELECT * FROM testemp.fire_originator_states;`

Result Grid | Filter Rows: | Edit: | Export/Import: | Wrap Cell C

	FIRE_ORIGINATOR_ID	FIRE_ORIGINATOR_STATE
▶ 1		CA
2		NM
3		OR
4		NC
5		WY
6		CO
7		WA
8		MT
9		UT
10		AZ
11		SD
12		AR
13		NV
...		..

ator\_states 1 x

Output :

Action Output

#	Time	Action	Message
✓ 2	17:43:54	SELECT * FROM testemp.fire_cause	13 row(s) returned
✓ 3	17:46:32	SELECT * FROM testemp.fips_codes	277 row(s) returned
✓ 4	17:47:38	SELECT * FROM testemp.fips_name	1425 row(s) returned
✓ 5	17:48:15	SELECT * FROM testemp.fips_code_has_fips_name	2120 row(s) returned
✓ 6	17:49:25	SELECT * FROM testemp.mtbs	6655 row(s) returned
✓ 7	17:49:58	SELECT * FROM testemp.mtbs	6655 row(s) returned
✓ 8	17:50:41	SELECT * FROM testemp.fire_originator_countys	2891 row(s) returned
✓ 9	17:51:59	SELECT * FROM testemp.fire_originator_states	52 row(s) returned



## 16. ICS\_209\_INCIDENT

SQL File 3\* ics\_209\_incident x

1 • `SELECT * FROM testemp.ics_209_incident;`

Result Grid

ICS_209_INCIDENT_NUMBER	ICS_209_NAME	ICS_209_INCIDENT_ID
TX-TXS-88041	Cotulla	10000
TX-TXS-88051	Laureles	10001
TX-TXS-88053	Old West	10002
TX-TXS-88057	Hopewell Road	10003
TX-TXS-88048	Crane	10004
TX-TXS-88047	Briggs	10005
TX-TXS-88056	Carta Valley Fire	10006
TX-TXS-88042	Kuykendall Ranch	10007
TX-TXS-88037	Kerr County C...	10008
TX-TXS-88055	Garvens Store ...	10009
ID-IPF-04021	Fawntooth	1001
TX-TXS-88043	Powers	10010
TX-TXS-88044	Cottonfield	10011

209\_incident 1 x

Output

Action Output

#	Time	Action	Message
✓ 3	17:46:32	SELECT * FROM testemp.fips_codes	277 row(s) returned
✓ 4	17:47:38	SELECT * FROM testemp.fips_name	1425 row(s) returned
✓ 5	17:48:15	SELECT * FROM testemp.fips_code_has_fips_name	2120 row(s) returned
✓ 6	17:49:25	SELECT * FROM testemp.mtbs	6655 row(s) returned
✓ 7	17:49:58	SELECT * FROM testemp.mtbs	6655 row(s) returned
✓ 8	17:50:41	SELECT * FROM testemp.fire_originator_countys	2891 row(s) returned
✓ 9	17:51:59	SELECT * FROM testemp.fire_originator_states	52 row(s) returned
✓ 10	17:53:22	SELECT * FROM testemp.ics_209_incident	10875 row(s) returned

## Data in the Database

SNO.	Table Name	Primary Key	Foreign Key	# of Rows in Table
1	SOURCE_SYSTEM_TYPES	SOURCE_SYSTEM_TYPE_ID	-	3
2	SOURCE_SYSTEMS	SOURCE_SYSTEM_ID	SOURCE_SYSTEM_TYPE_ID	31
3	SOURCE_REPORTING_UNITS	SOURCE_REPORTING_UNIT	SOURCE_SYSTEM_ID, UNIT_ID(NWGC_REPORTING_UNIT)	1404
4	NWGC_LOCATION	OBJECTID	UNIT_ID(NWGC_REPORTING_UNIT)-	5866
5	NWGC_REPORTING_UNIT	UNIT_ID		5866
6	OWNER_DETAILS	OWNER_CODE	-	15
7	FIRES	FOD_ID	FIRE_ORIGINATOR_ID, FIPS_CODE_ID, FIPS_NAME_ID, ICS_209_INCIDENT_ID, STAT_CAUSE_CODE, OWNER_CODE, MTBS_ID, OBJECTID (NWGC_LOCATION)  UNIT_ID (NWGC_REPORTING_UNIT)	1880465

8	FIRE_CAUSE	STAT_CAUSE_CODE	-	13
9	FIRE_IDENTIFIER	FPA_ID	FOD_ID	1879442
10	FIPS_CODE	FIPS_CODE_ID	-	277
11	FIPS_NAME	FIPS_NAME_ID	-	1425
12	FIPS_CODE_has_FIPS_NAME	FIPS_CODE_ID FIPS_NAME_ID		2120
13	MTBS	MTBS_ID	-	6655
14	FIRE_ORIGINATOR_COUNTYS	FIRE_ORIGINATOR_COUNTY_ID	FIRE_ORIGINATOR_ID (From FIRE_ORIGINATOR_STATE)	2891
15	FIRE_ORIGINATOR_STATE	FIRE_ORIGINATOR_ID	-	52
16	ICS_209_INCIDENT	ICS_209_INCIDENT_ID	-	10875

## 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. All states are interested. Which state has the least chance to win a share of the fund?**

#### Assumption:

It is assumed that the company is going to fund the states to remove debris in order to prevent fires in the future. Hence the company is looking for the states that have the highest amount of fires due to debris. The question is about which state is having the least chance to receive the funding, i.e, the state with the least number of fires caused by debris will have the least chance to win the fund. So, we are considering the states that had least fires caused by debris. Since many states have the same number of fires due to debris, we are randomly picking one as our answer.

#### Translation:

```
SELECT FIRE_ORIGINATOR_STATE from fires table
SUM of fire_size as firesize then
INNER JOIN NWCG_Reporting_Unit ON nwcg_reporting_unit.unit_id and
INNER JOIN source_reportings_units ON nwcg_reporting_unit.unit_id and
INNER JOIN fire_originator_states ON FIRE_ORIGINATOR_STATE_FIRE_ORIGINATOR_ID and
INNER JOIN fire_cause on fire_cause.stat_cause_code and
WHERE stat_cause_code= '5' then
GROUP BY source_reporting_unit_name and then
ORDER BY firesize ASC and limit 1
```

#### Cleanup:

```
SELECT FIRE_ORIGINATOR_STATE from fires
INNER JOIN NWCG_Reporting_Unit ON nwcg_reporting_unit.unit_id
INNER JOIN source_reportings_units ON nwcg_reporting_unit.unit_id
INNER JOIN fire_originator_states ON FIRE_ORIGINATOR_STATE_FIRE_ORIGINATOR_ID
INNER JOIN fire_cause on fire_cause.stat_cause_code
where stat_cause_code= '5'
GROUP BY source_reporting_unit_name
ORDER BY COUNT(*) ASC
limit 1
```

#### Result:

Though there were many states (around 697 states) that had lower fires caused because of debris they all have the least chance of getting selected for the fund. However, we have randomly picked only one among the states that have the least chance of being selected

## Screen Shot of SQL Query and Results

```
Query 1  SQL File 2* x
-- 1. Translation:
-- SELECT FIRE_ORIGINATOR_STATE from fires table
-- SUM of fire_size as firesize then
-- INNER JOIN NWCG_Reporting_Unit ON nwcg_reporting_unit.unit_id and
-- INNER JOIN source_reportings_units ON nwcg_reporting_unit.unit_id and
-- INNER JOIN fire_originator_states ON FIRE_ORIGINATOR_STATE_FIRE_ORIGINATOR_ID and
-- INNER JOIN fire_cause on fire_cause.stat_cause_code and
-- WHERE stat_cause_code= '5' then
-- GROUP BY source_reporting_unit_name and then
-- ORDER BY firesize ASC and limit 1
-- Cleanup:
-- SELECT FIRE_ORIGINATOR_STATE
-- from fires
-- SUM (fire_size) as firesize
-- INNER JOIN NWCG_Reporting_Unit ON nwcg_reporting_unit.unit_id
-- INNER JOIN source_reportings_units ON nwcg_reporting_unit.unit_id
-- INNER JOIN fire_originator_states ON FIRE_ORIGINATOR_STATE_FIRE_ORIGINATOR_ID
-- INNER JOIN fire_cause on fire_cause.stat_cause_code
-- WHERE stat_cause_code= '5'
-- GROUP BY source_reporting_unit_name
```

```
Query 1  SQL File 2*
-- ORDER BY firesize ASC
-- limit 1
SELECT
FIRE_ORIGINATOR_STATE,
sum(fire_size) as totalsize
FROM
fires
INNER JOIN NWCG_Reporting_Unit ON replace(fires.nwcg_reporting_unit_unit_id, '\r','') = nwcg_reporting_unit.unit_id
INNER JOIN source_reportings_units ON source_reportings_units.NWCG_REPORTING_UNIT_UNIT_ID = nwcg_reporting_unit.unit_id
INNER JOIN fire_originator_states ON fire_originator_id = FIRE_ORIGINATOR_STATE_FIRE_ORIGINATOR_ID
INNER JOIN fire_cause on fires.FIRE_CAUSE_STAT_CAUSE_CODE1 = fire_cause.stat_cause_code
where stat_cause_code= '5'
GROUP BY
source_reporting_unit_name
ORDER BY
totalsize ASC
limit 1
```

Result Grid

FIRE_ORIGINATOR_STATE	totalsize
WA	0.1

Result 4 x

Output

Action Output

#	Time	Action	Message
4	21:46:41	SELECT FIRE_ORIGINATOR_STATE, sum(fire_size) as totalsize FROM fires INNER JOIN NWCG_Rep...	1 row(s) returned

## Query 2

### Question:

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

**(3 forests where fire due to children is least prevalent)**

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

#### Assumption:

We need to find 3 forests that have the least fires that are caused due to the activities of children, since the agencies are suggesting that children should be banned from the forests that have the highest rate of fires due to children. After running the query, there were more than 3 forests where this would apply, so we used the "limit 3" query to only show 3 forests where this is applicable. However, in figure 1 we have run the query to show that there were 63 forests which complies to this, but we are choosing only 3 randomly. Also, we are using the Source Reporting Unit Name as the Forest Name because there is no actual forest name that is available in the given dataset.

#### Translation:

Select source\_reporting\_unit\_name from fires table and  
Inner join NWCG\_Reporting\_Unit table on Unit\_ID and  
Inner join fire\_cause table on STAT\_CAUSE\_CODE column  
Where the stat\_cause\_code= '8' and SOURCE\_REPORTING\_UNIT\_NAME like '%forest' then  
Group by the source\_reporting\_unit\_name and  
Order by COUNT(\*) in Asc and then  
Limit 3 rows

#### Cleanup:

Select source\_reporting\_unit\_name from fires  
Inner join NWCG\_Reporting\_Unit on Unit\_ID  
Inner join fire\_cause table on STAT\_CAUSE\_CODE  
Where stat\_cause\_code= '8' and SOURCE\_REPORTING\_UNIT\_NAME like '%forest'  
Group by source\_reporting\_unit\_name  
Order by COUNT(\*) Asc  
Limit 3

#### Result:

The 3 forests that were randomly selected among 63 are

1. Bridger-Teton National Forest
2. Pike San Isabel National Forest
3. Wallowa- Whitman National Forest

No. of rows = 3

## Screen Shot of SQL Query and Results

```
SQL File 3*  ics_209_incident  SQL File 4*  x
[Icons]  Don't Limit  [Icons]
1  -- 2. One of the reporting agencies has suggested that children be banned from its forests unless there is one adult for
2  -- every 3 children in a group visiting a forest. Name 3 forests where this would be the least appropriate.
3  -- (3 forests where fire due to children is least prevalent)
4  -- Translation:
5  -- Select source_reporting_unit_name from fires table and
6  -- Inner join NWCG_Reporting_Unit table on Unit_ID and
7  -- Inner join fire_cause table on STAT_CAUSE_CODE column
8  -- Where the stat_cause_code= '8' then
9  -- Group by the source_reporting_unit_name and
10 -- Order by COUNT(*) in Asc and then
11 -- Limit 3 rows
12 -- Cleanup :
13 -- Select source_reporting_unit_name from fires
14 -- Inner join NWCG_Reporting_Unit on Unit_ID
15 -- Inner join fire_cause table on STAT_CAUSE_CODE
16 -- Where stat_cause_code= '8'
17 -- Group by source_reporting_unit_name
18 -- Limit 3
19 • SELECT source_reporting_unit_name,
20 COUNT(*)
21 FROM fires
22 INNER JOIN NWCG_Reporting_Unit ON replace(fires.nwcg_reporting_unit_unit_id, '\r','') = nwcg_reporting_unit.unit_id
23 INNER JOIN source_reportings_units ON source_reportings_units.NWCG_REPORTING_UNIT_UNIT_ID = nwcg_reporting_unit.unit_id
```

```
24 INNER JOIN fire_cause on fires.FIRE_CAUSE_STAT_CAUSE_CODE1 = fire_cause.stat_cause_code
25 where stat_cause_code= '8' and SOURCE_REPORTING_UNIT_NAME like '%forest'
26 GROUP BY source_reporting_unit_name
27 ORDER BY COUNT(*) ASC
28 Limit 3
29
```

< [Progress Bar]

Result Grid [Icons] Filter Rows: [Text Box] Export: [Icons] Wrap Cell Content: [Icon]

	source_reporting_unit_name	COUNT(*)
▶	Bridger-Teton National Forest	1
	Pike San Isabel National Forest	1
	Wallowa-Whitman National Forest	1

Result 1 x [Progress Bar]

Output [Dotted Line]

[Icon] Action Output [Dropdown]

#	Time	Action	Message
✓ 1	19:38:36	SELECT source_reporting_unit_name, COUNT(*) FROM fires INNER JOIN NWCG_Re...	3 row(s) returned

Figure 1:

```
22
23 • SELECT source_reporting_unit_name,
24 COUNT(*)
25 FROM fires
26 INNER JOIN NWCG_Reporting_Unit ON replace(fires.nwcg_reporting_unit_unit_id, '\r','') = nwcg_reporting_unit.unit_id
27 INNER JOIN source_reportings_units ON source_reportings_units.NWCG_REPORTING_UNIT_UNIT_ID = nwcg_reporting_unit.unit_id
28 INNER JOIN fire_cause ON fires.FIRE_CAUSE_STAT_CAUSE_CODE1 = fire_cause.stat_cause_code
29 where stat_cause_code= '8'
30 GROUP BY source_reporting_unit_name
31 HAVING COUNT(*) = '1'
32 ORDER BY COUNT(*) ASC
```

Result Grid

source_reporting_unit_name	COUNT(*)
Six Rivers National Forest	1
Shoshone Field Office	1
Chesapeake & Ohio Canal National Historic Park	1
Bois Forte Tribe	1
Shawnee National Forest	1
Puget Sound Agency	1
Philomath Unit	1
South Idaho District	1
Chesapeake & Ohio Canal National Historical Park	1
Round Valley Tribe	1

Output

Action Output

#	Time	Action	Message	Duration / F
5	18:57:39	SELECT source_reporting_unit_name COUNT(*) FROM fires INNER JOIN ...	Error Code: 1064. You have an error in your SQL syntax; check the manual...	0.000 sec
6	18:57:52	SELECT source_reporting_unit_name, COUNT(*) FROM fires INNER JOIN...	472 row(s) returned	2.469 sec /
7	18:58:31	SELECT source_reporting_unit_name, COUNT(*) FROM fires INNER JOIN...	63 row(s) returned	2.547 sec /



### 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.**

#### Assumption:

We need to check whether the assumption of the advocacy group that “human actions are reason for forest fires and not the nature” is right. In the fire cause table, there are several causes of fires that are given. We are considering the nature causes to be lightening and debris, and all other causes are due to human activities. We are also considering that the miscellaneous and undefined fields are also human activity. If the count of fires is more for human related than nature related, then the advocacy group is proven right.

#### Translation:

```
Select Case When stat_cause_code='1' or stat_cause_code = '5' then 'NATURE' ELSE as 'HUMAN' then
count(*) as Fire_count
from the fires table
Inner join fire_cause table on stat_cause_code then
Group by Cause_Type and then
order by Fire_count in desc
```

#### Cleanup:

```
Select Case When stat_cause_code='1' or stat_cause_code = '5' then 'NATURE' ELSE as 'HUMAN'
count(*) as Fire_count
from fires
Inner join fire_cause table on stat_cause_code
Group by Cause_Type
order by Fire_count desc
```

#### Result:

From the result we see that though the fires because of humans is more than the number of fires caused from nature, we cannot completely assume that all fires are caused by humans only and not by nature.

Fires Caused:

By humans = 583629

By Nature = 329853

## Screen Shot of SQL Query and Results

SQL File 3\* Query 2 - Children\* SQL File 5\* x SQL File 6

Don't Limit

```
1 -- 3. One advocacy group says human actions and not Nature is to blame for most wildfires. Write a query that supports this statement
2 -- Translation:
3 -- Select Case When stat_cause_code='1' or stat_cause_code = '5' then 'NATURE' ELSE as 'HUMAN' then
4 -- count(*) as Fire_count
5 -- from the fires table
6 -- Inner join fire_cause table on stat_cause_code then
7 -- Group by Cause_Type and then
8
9 -- Cleanup :
10 -- Select Case When stat_cause_code='1' or stat_cause_code = '5' then 'NATURE' ELSE as 'HUMAN'
11 -- count(*) as Fire_count
12 -- from fires
13 -- Inner join fire_cause table on stat_cause_code
14 -- Group by Cause_Type
15 -- order by Fire_count desc
16
17 • SELECT CASE WHEN stat_cause_code='1' or stat_cause_code = '5' then 'NATURE' ELSE 'HUMAN' END AS Cause_Type,
```

```
18 count(*) as Fire_count
19 FROM fires
20 INNER JOIN fire_cause ON fire_cause.stat_cause_code = fires.fire_cause_stat_cause_code1
21 GROUP BY Cause_Type
22 order by Fire_count desc
23
24
```

<

Result Grid | Filter Rows: | Export: | Wrap Cell Content: |

	Cause_Type	Fire_count
▶	HUMAN	583629
	NATURE	329853

Result 3 x

Output

Action Output

#	Time	Action	Message
✓ 9	19:38:44	SELECT CASE WHEN stat_cause_code='1' or STAT_CAUSE_CODE = '5' then 'NA...	13 row(s) returned
✓ 10	19:50:41	SELECT CASE WHEN stat_cause_code='1' or stat_cause_code = '5' then 'NATUR...	2 row(s) returned
✓ 11	20:13:28	SELECT CASE WHEN stat_cause_code='1' or stat_cause_code = '5' then 'NATUR...	2 row(s) returned

## Query 4

### Question:

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

### Assumption:

We need to find which state had more fires in the first 182 days (first half) of the year when compared to the second 182 days (second half) of the year. In this case, we are assuming that 182 days is the exact half of the entire calendar year, so we are using the “day of year” field to accomplish this calculation.

### Translation:

```
Select distinct fire_originator_state and sum(fist_half-sec_half) Difference
Then do subquery, where
Select FIRE_ORGINATOR_ID, sum( CASE when discovery_doy > 182 then '1' else '0' end) as FIST_HALF,
then
sum( CASE when discovery_doy < 182 then '1' else '0' end) as sec_half
from fires table then
group by 1)
fires INNER JOIN fire_originator_states table on FIRE_ORGINATOR_ID
where fist_half > sec_half then
group by 1 and
order by Difference in descending and finally
limit 1
```

### Cleanup:

```
Select distinct fire_originator_state, sum(fist_half-sec_half) Difference
From
(Select FIRE_ORGINATOR_ID, sum( CASE when discovery_doy > 182 then '1' else '0' end) as FIST_HALF,
sum( CASE when discovery_doy < 182 then '1' else '0' end) as sec_half
from fires group by 1)
fires INNER JOIN fire_originator_states on FIRE_ORGINATOR_ID
where fist_half > sec_half
group by 1
order by Difference DESC
limit 1
```

### Result:

After getting the states based on the condition that “fires in the first half of a calendar year greater than the second half of the calendar year”, we sort them in descending order and then pick 1.

From the result we see that California has more no. of fires in the first half of the calendar year than in the second half when compared to the other states.

## Screen Shot of SQL Query and Results

```
SQL File 3* Query 2 - Children* Query 3 - Human_Nature SQL File 6* x
1 -- 7. Which state had more fires in the first half of a calendar year than the second half of the calendar year?
2 -- Translation:
3 -- Select distinct fire_originator_state and sum(first_half -sec_half) Difference
4 -- Then do subquery, where
5 -- Select FIRE_ORGINATOR_ID, sum( CASE when discovery_doy > 182 then '1' else '0' end) as FIST_HALF, then
6 -- sum( CASE when discovery_doy < 182 then '1' else '0' end) as sec_half
7 -- from fires table then
8 -- group by 1)
9 -- fires INNER JOIN fire_originator_states table on FIRE_ORGINATOR_ID
10 -- where fist_half > sec_half then
11 -- group by 1 and
12 -- order by Difference in descending and finally
13 -- limit 1
14
15 -- Cleanup :
16 -- Select distinct fire_originator_state, sum(first_half -sec_half) Difference
17 -- From
18 -- (Select FIRE_ORGINATOR_ID, sum( CASE when discovery_doy > 182 then '1' else '0' end) as FIST_HALF,
19 -- sum( CASE when discovery_doy < 182 then '1' else '0' end) as sec_half
20 -- from fires group by 1)
21 -- fires INNER JOIN fire_originator_states on FIRE_ORGINATOR_ID
22 -- where fist_half > sec_half
23 -- group by 1
```

```
24 -- order by Difference DESC
25 -- limit 1
26
27
28 • SELECT distinct fire_originator_states.fire_originator_state, sum(first_half -sec_half) Difference from
29 (Select FIRE_ORGINATOR_STATE_FIRE_ORGINATOR_ID, sum( CASE when discovery_doy > 182 then '1' else '0' end) FIST_HALF,
30 sum( CASE when discovery_doy < 182 then '1' else '0' end) sec_half from fires group by 1) fires
31 INNER JOIN fire_originator_states
32 ON fires.FIRE_ORGINATOR_STATE_FIRE_ORGINATOR_ID= fire_originator_states.fire_originator_id
33 where fist_half > sec_half group by 1 order by Difference desc
34 limit 1
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: |

fire_originator_state	Difference
CA	54760

Result 6 x

Output

Action Output

#	Time	Action	Message
18	20:37:53	SELECT distinct fire_originator_states.fire_originator_state, sum(first_half -sec_half) c...	12 row(s) returned
19	20:38:33	SELECT distinct fire_originator_states.fire_originator_state, sum(first_half -sec_half) c...	1 row(s) returned
20	20:40:41	SELECT distinct fire_originator_states.fire_originator_state, sum(first_half -sec_half) Di...	1 row(s) returned

## Query 5

### Question:

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

### Assumption:

We need to find the forests that had fires that lasted 2 days or less. Therefore, we are assuming that this is applicable with the fires that are lasting for less than 3 days, but not 3 days. We are considering “discovery\_doy” as the day that the fire was discovered and use that as the first day of the fire. We are using the “cont\_doy” as the last day of the fire, so the difference between these two days will give the total length of the fire. We are using the Source Reporting Unit Name as the Forest Name because there is no actual forest name that is available in the given dataset

### Translation:

```
Select distinct SOURCE_REPORTING_UNIT_NAME column
From
From here subquery:
(Select all from fires tables where cont_doy - discovery_doy <3)
Fires
INNER JOIN NWCG_Reporting_Unit table ON unit_id column and replace
(fires.nwcg_reporting_unit_unit_id, '\r', '')
INNER JOIN source_reportings_units table ON unit_id
```

### Cleanup:

```
Select distinct SOURCE_REPORTING_UNIT_NAME
From
(Select * from fires where cont_doy - discovery_doy <3)
Fires
INNER JOIN NWCG_Reporting_Unit table ON unit_id
Replace (fires.nwcg_reporting_unit_unit_id, '\r', '')
INNER JOIN source_reportings_units ON unit_id
```

### Result:

From the result we can see that there were 1387 forests had fires which lasted for less than 3 days

## Screen Shot of SQL Query and Results

```
SQL File 3* Query 2 - Children* Query 3 - Human_Nature Query 7 - Mores fires in first cale... Query 6 - fires less than 3 days* x
1 -- 6. What were the forests that had no fires that lasted more than two days?
2 -- Translation:
3 -- Select distinct SOURCE_REPORTING_UNIT_NAME column
4 -- From
5 -- From here subquery:
6 -- (Select all from fires tables where cont_doy - discovery_doy <3)
7 -- Fires
8 -- INNER JOIN NWCG_Reporting_Unit table ON unit_id column and replace (fires.nwcg_reporting_unit_unit_id, '\r','')
9 -- INNER JOIN source_reportings_units table ON unit_id
10
11 -- Cleanup :
12 -- Select distinct SOURCE_REPORTING_UNIT_NAME
13 -- From
14 -- (Select * from fires where cont_doy - discovery_doy <3)
15 -- Fires
16 -- INNER JOIN NWCG_Reporting_Unit table ON unit_id
17 -- Replace (fires.nwcg_reporting_unit_unit_id, '\r','')
18 -- INNER JOIN source_reportings_units ON unit_id
19
20 • SELECT
21     distinct source_reportings_units.SOURCE_REPORTING_UNIT_NAME
22     from
```

```
23 (Select * from fires where cont_doy - discovery_doy <3) Fires
24 INNER JOIN NWCG_Reporting_Unit ON nwcg_reporting_unit.unit_id = replace(fires.nwcg_reporting_unit_unit_id, '\r','')
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows:

SOURCE_REPORTING_UNIT_NAME
Plumas National Forest
Eldorado National Forest
Shasta-Trinity National Forest
Lincoln National Forest
Deschutes National Forest
National Forests in North Carolina
Medicine Bow-Routt National Forest
Umatilla National Forest
Pike San Isabel National Forest
Ochoco National Forest
Fremont-Winema National Forest
Winema National Forest
Bighorn National Forest
Kootenai National Forest
Wacatch-Cache-I linta National For...

Result 1 x

Output

Action Output

#	Time	Action	Message
19	20:38:33	SELECT distinct fire_originator_states.fire_originator_state, sum(fist_half - sec_half) c...	1 row(s) returned
20	20:40:41	SELECT distinct fire_originator_states.fire_originator_state, sum(fist_half - sec_half) Di...	1 row(s) returned
21	21:09:59	SELECT distinct source_reportings_units.SOURCE_REPORTING_UNIT_NAME S...	1387 row(s) returned

## Query 6

### Question:

#### 8. Which forest had the least number of fires?

### Assumption:

We need to find the one forest that had the least number of fires. After running the query, there were 82 rows (forests) with the count as 1 fire as shown in figure 2. So, we had to limit the result to one forest, and got the result. We are using the Source Reporting Unit Name as the Forest Name because there is no actual forest name that is available in the given dataset.

### Translation:

```
Select source_reporting_unit_name and COUNT(*)  
FROM the fires table  
INNER JOIN NWCG_Reporting_Unit table ON unit_id as nwcg_reporting_unit.unit_id  
INNER JOIN source_reportings_units table ON UNIT_ID then  
GROUP BY the source_reporting_unit_name  
ORDER BY COUNT(FOD_ID) ASC then  
Limit it by 1
```

### Cleanup:

```
Select source_reporting_unit_name, COUNT(*)  
FROM fires  
INNER JOIN NWCG_Reporting_Unit table ON unit_id as nwcg_reporting_unit.unit_id  
INNER JOIN source_reportings_units table ON UNIT_ID  
GROUP BY source_reporting_unit_name  
ORDER BY COUNT(FOD_ID) ASC  
limit 1
```

### Result:

From the 82 forests that had only one fire we have randomly selected one for our query

## Screen Shot of SQL Query and Results

```
SQL File 3* Query 8 - Forests with least fires x
-- 8. Which forest had the least number of fires?
-- Translation:
-- Select source_reporting_unit_name and COUNT(*)
-- FROM the fires table
-- INNER JOIN NWCG_Reporting_Unit table ON unit_id as nwcg_reporting_unit.unit_id
-- INNER JOIN source_reportings_units table ON UNIT_ID then
-- GROUP BY the source_reporting_unit_name
-- ORDER BY COUNT(FOD_ID) ASC then
-- Limit it by 1

-- Cleanup:
-- Select source_reporting_unit_name, COUNT(*)
-- FROM fires
-- INNER JOIN NWCG_Reporting_Unit table ON unit_id as nwcg_reporting_unit.unit_id
-- INNER JOIN source_reportings_units table ON UNIT_ID
-- GROUP BY source_reporting_unit_name
-- ORDER BY COUNT(FOD_ID) ASC
-- limit 1

• SELECT source_reporting_unit_name, COUNT(FOD_ID)
FROM fires
```

```
22 INNER JOIN NWCG_Reporting_Unit ON replace(fires.nwcg_reporting_unit_unit_id, '\r','') = nwcg_reporting_unit.unit_id
23 INNER JOIN source_reportings_units ON source_reportings_units.NWCG_REPORTING_UNIT_UNIT_ID = nwcg_reporting_unit.unit_id
24 GROUP BY source_reporting_unit_name
25 ORDER BY COUNT(FOD_ID) ASC
26 Limit 1
27
28
```

Result Grid

source_reporting_unit_name	COUNT(FOD_ID)
Russell Cave National Monument	1

Result 3 x

Output

Action Output

#	Time	Action	Message
✓ 23	22:01:19	SELECT source_reporting_unit_name, COUNT(FOD_ID) FROM fires INNER JOIN N...	1 row(s) returned
✓ 24	22:03:16	SELECT source_reporting_unit_name, COUNT(FOD_ID) FROM fires INNER JOIN N...	82 row(s) returned
✓ 25	22:06:56	SELECT source_reporting_unit_name, COUNT(FOD_ID) FROM fires INNER JOIN N...	1 row(s) returned



Figure 2:

SQL File 3\* Query 2 - Children\* Query 3 - Human\_Nature Query 7 - Mores fires in first cale... Query 6 - fires less than 3 days\* SQL File 9\* x

Don't Limit

```

1  -- 8. Which forest had the least number of fires?
2  -- Translation:
3  -- Select source_reporting_unit_name and COUNT(*)
4  -- FROM the fires table
5  -- INNER JOIN NWCG_Reporting_Unit table ON unit_id as nwcg_reporting_unit.unit_id
6  -- INNER JOIN source_reportings_units table ON UNIT_ID then
7  -- GROUP BY the source_reporting_unit_name
8  -- ORDER BY COUNT(FOD_ID) ASC then
9  -- Limit it by 1
10

```

Result Grid

source_reporting_unit_name	COUNT(FOD_ID)
Russell Cave National Monument	1
Banks Lake National Wildlife Refuge	1
Huleia National Wildlife Refuge	1
Effigy Mounds National Monument	1
Middle Mississippi National Wildlife Refuge	1
Thomas Stone National Historic Site	1
Humboldt Fire Protection District	1
Goodwin Volunteer Fire Department	1
Hill City Volunteer Fire Department	1
Lodgepole Volunteer Fire Department	1
Silver City Volunteer Fire Department	1
Lake Isom National Wildlife Refuge	1
Chamizal National Monument	1
Timpanogos Cave National Monument	1
Hinsdale County	1

Result 2 x

Output

Action Output

#	Time	Action	Message
✓ 22	21:12:36	SELECT distinct source_reportings_units.SOURCE_REPORTING_UNIT_NAME ...	1387 row(s) returned
✓ 23	22:01:19	SELECT source_reporting_unit_name, COUNT(FOD_ID) FROM fires INNER JOIN N...	1 row(s) returned
✓ 24	22:03:16	SELECT source_reporting_unit_name, COUNT(FOD_ID) FROM fires INNER JOIN N...	82 row(s) returned

## Data Review for MongoDB

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

In mongo DB, the database contains Collections and documents. The database name used here is "Wildfires". The collections are similar to tables in MySQL. This database contains only two tables and they are "fires" and "nwcg". The documents are similar to rows. The collection name "fires" contains 1880465 rows and the collection named nwcg contains 5867 rows.

Since MongoDB is a document-oriented database, no schema or structure is mandatory. Hence the tables are not normalized, and no relationship is created between them.

## Physical Mongo Database

### Assumptions/Notes About Data Set

The dataset "FPA\_FOD\_20170508" extracted from SQL Lite had many tables and were not in normalized form. The various tables that were present are Fires, NWCG, Elementary Geometries, idx\_Fires\_Shape, KNN, Spatial

Index geometry\_columns, geometry\_columns\_auth, geometry\_columns\_field\_infos, geometry\_columns\_statistics, geometry\_columns\_time, idx\_Fires\_shape\_Node, idx\_Fires\_shape\_parent, idx\_Fires\_shape\_Rowid, spatial\_ref\_sys, spatial\_ref\_sys\_aux, spatialite\_history, views\_geometry\_columns, views\_geometry\_columns\_auth, views\_geometry\_columns\_feild\_infos, views\_geometry\_columns\_statistics. But we have only exported the Fires and NWCG tables from SQL Lite to csv or json and imported it into MongoDB because of the following reasons.

- Elementary Geometries, idx\_Fires\_Shape, KNN and Spatial Index were virtual tables and hence did not contain any information
- Tables as geometry\_columns, geometry\_columns\_auth, geometry\_columns\_field\_infos, geometry\_columns\_statistics, geometry\_columns\_time contain very less information i.e. only one row is present per table which is not relevant
- Tables such as idx\_Fires\_shape\_Node, idx\_Fires\_shape\_parent, idx\_Fires\_shape\_Rowid contain details about parent node columns which was also present in the NWCG table. However, the parent column in the NWCG table itself is completely NULL which indicates that these data here will not be useful
- Tables such as spatial\_ref\_sys, spatial\_ref\_sys\_aux, spatialite\_history contains several rows with enough information. However, these tables are not related with Fires table and NWCG table which are our main tables. And so, we are not considering these tables
- Tables like views\_geometry\_columns, views\_geometry\_columns\_auth, views\_geometry\_columns\_feild\_infos, views\_geometry\_columns\_statistics contain no data. Therefore, we are not considering these tables

The two table Fires and NWCG were imported into Mongo DB directly without into normalization as Mongo DB doesn't required fixed structure or schema.

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

Database Name: Wildfires

Collections: Fires, NWCG

### 1. Screenshot of database with collections and attributes

MongoDB Compass Community - localhost:27017/wildfires

Connect View Help

My Cluster

- 5 DBS
- 5 COLLECTIONS

HOST: localhost:27017

CLUSTER: Standalone

EDITION: MongoDB 4.2.1 Community

Filter your data

- > admin
- > config
- > local
- > wildfires
- fires
- nwcg

Collections

CREATE COLLECTION

Collection Name	Documents	Avg. Document Size	Total Document Size	Num. Indexes	Total Index Size	Properties
fires	1,880,465	930.6 B	1.7 GB	1	18.8 MB	
nwcg	5,867	290.2 B	1.7 MB	1	81.9 KB	

### 2. Screenshot of collection "fires"

MongoDB Compass Community - localhost:27017/wildfires.fires

Connect View Collection Help

My Cluster

- 5 DBS
- 5 COLLECTIONS

HOST: localhost:27017

CLUSTER: Standalone

EDITION: MongoDB 4.2.1 Community

Filter your data

- > admin
- > config
- > local
- > wildfires
- fires
- nwcg

wildfires.fires Documents

DOCUMENTS 1.9m TOTAL SIZE 1.8GB AVG. SIZE 931B 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

#	fires	_id ObjectId	FID_ID Int32	FPA_ID String	SOURCE_SYSTEM_TYPE String	SOURCE_SYSTEM String	NWCG_REPORTING_AGENCY String
1		Sdecac3c09591153dbaaf023	2	"FS-1418827"	"FED"	"FS-FIRESTAT"	"FS"
2		Sdecac3c09591153dbaaf024	1	"FS-1418826"	"FED"	"FS-FIRESTAT"	"FS"
3		Sdecac3c09591153dbaaf025	3	"FS-1418835"	"FED"	"FS-FIRESTAT"	"FS"
4		Sdecac3c09591153dbaaf026	4	"FS-1418845"	"FED"	"FS-FIRESTAT"	"FS"
5		Sdecac3c09591153dbaaf027	5	"FS-1418847"	"FED"	"FS-FIRESTAT"	"FS"
6		Sdecac3c09591153dbaaf028	6	"FS-1418849"	"FED"	"FS-FIRESTAT"	"FS"
7		Sdecac3c09591153dbaaf029	7	"FS-1418851"	"FED"	"FS-FIRESTAT"	"FS"
8		Sdecac3c09591153dbaaf02a	8	"FS-1418854"	"FED"	"FS-FIRESTAT"	"FS"
9		Sdecac3c09591153dbaaf02b	9	"FS-1418856"	"FED"	"FS-FIRESTAT"	"FS"
10		Sdecac3c09591153dbaaf02c	10	"FS-1418859"	"FED"	"FS-FIRESTAT"	"FS"
11		Sdecac3c09591153dbaaf02d	12	"FS-1418863"	"FED"	"FS-FIRESTAT"	"FS"
12		Sdecac3c09591153dbaaf02e	11	"FS-1418861"	"FED"	"FS-FIRESTAT"	"FS"
13		Sdecac3c09591153dbaaf02f	14	"FS-1418872"	"FED"	"FS-FIRESTAT"	"FS"
14		Sdecac3c09591153dbaaf030	13	"FS-1418865"	"FED"	"FS-FIRESTAT"	"FS"
15		Sdecac3c09591153dbaaf031	16	"FS-1418876"	"FED"	"FS-FIRESTAT"	"FS"
16		Sdecac3c09591153dbaaf032	17	"FS-1418878"	"FED"	"FS-FIRESTAT"	"FS"
17		Sdecac3c09591153dbaaf033	19	"FS-1418884"	"FED"	"FS-FIRESTAT"	"FS"



## MongoDB Queries/Code

### Query 1

Question:

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

### Translation

Select documents from the Fires collection and use the aggregate function such as \$match

To match the condition Stat\_Cause\_Desc to Debris burning, then

Use \$group to group the documents based on state

And then calculate the sum of fire size for each corresponding groups, and

then \$sort to return the documents according to the sum of fire size in ascending order, and then

Set \$limit as 1 to display 1 document

### Cleanup:

Use db.fires.aggregate

\$match Stat\_Cause\_Desc = "Debris Burning"

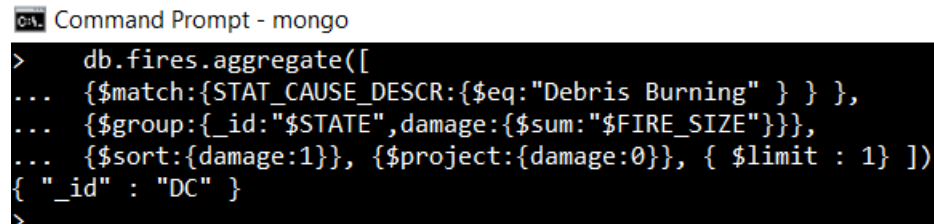
\$group "\$State", damage: { \$sum : "\$fire size" }

\$sort : damage :1

\$project : damage :0

\$limit 1

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



```
Command Prompt - mongo
> db.fires.aggregate([
...   {$match:{STAT_CAUSE_DESC:{$eq:"Debris Burning" } } },
...   {$group: {_id:"$STATE", damage:{$sum:"$FIRE_SIZE"}}},
...   {$sort:{damage:1}}, {$project:{damage:0}}, { $limit : 1} ])
{ "_id" : "DC" }
>
```

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

The leading beverage company is going to fund the states that has higher fires caused because of debris. Hence all states that had fires caused by debris are grouped and their fire size is calculated. The state which has the lease fire size will have the lowest chance to be picked.

**Result:** No. Of documents = 1

**Assumption:** Though there were many states that had least number of fires, one was picked randomly for the sake of answering to this query

## Query 2

### Question:

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

### Translation:

Select documents from the Fires collection and use the aggregate function such as \$match

To check the condition Stat\_Cause\_Code as 8, then

Use \$group to group the documents based on reporting unit id as SOURCE\_REPORTING\_UNIT\_NAME

And then calculate the count of fires for each corresponding groups, and

then \$sort to return the documents according to the count of fires in ascending order, and then

Set \$limit as 3 for produce 3 documents.

### Cleanup:

Use db.fires.aggregate

\$match Stat\_Cause\_Code= 8

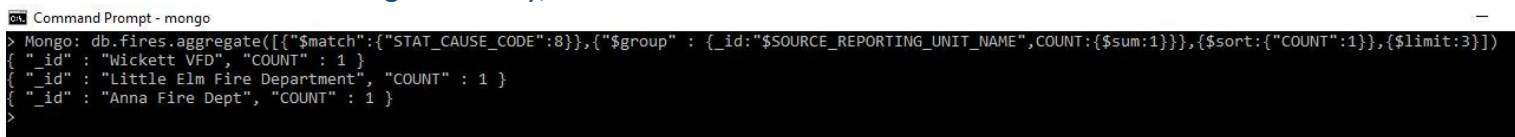
\$group SOURCE\_REPORTING\_UNIT\_NAME

COUNT: \$SUM,

\$sort Firecount in ascending,

\$limit 3

## Screen Shot of MongoDB Query/Code and Results



```
Command Prompt - mongo
> Mongo: db.fires.aggregate([{"$match":{"STAT_CAUSE_CODE":8}},{"$group":{"_id":"SOURCE_REPORTING_UNIT_NAME","COUNT":{"$sum:1}}},{"$sort":{"COUNT":1}},{"limit:3}])
{ "_id" : "Wickett VFD", "COUNT" : 1 }
{ "_id" : "Little Elm Fire Department", "COUNT" : 1 }
{ "_id" : "Anna Fire Dept", "COUNT" : 1 }
>
```

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

We need to find 3 forests that have the least fires that are caused due to the activities of children, since the agencies are suggesting that children should be banned from the forests that have the highest rate of fires due to children. Though there were many forests that satisfied the condition, since the question specifically asks us to pick only 3, we have randomly picked 3 forest from among all the forest that are in alignment with this condition.

The forests where this condition would be least appropriate are displayed in output

**Result :** No. of documents = 3

**Assumption:** We are using source reporting unit name as the forest name because there is no actual forest name that is available in the given dataset

### 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.**

#### Translation:

Select documents from the Fires collection and use the aggregate function \$project  
To display the condition Stat\_Cause\_Code, then  
Use \$switch to group the documents based on branches Nature and human  
And then use \$group to display counts based on groups

#### Cleanup:

Use db.fires.aggregate

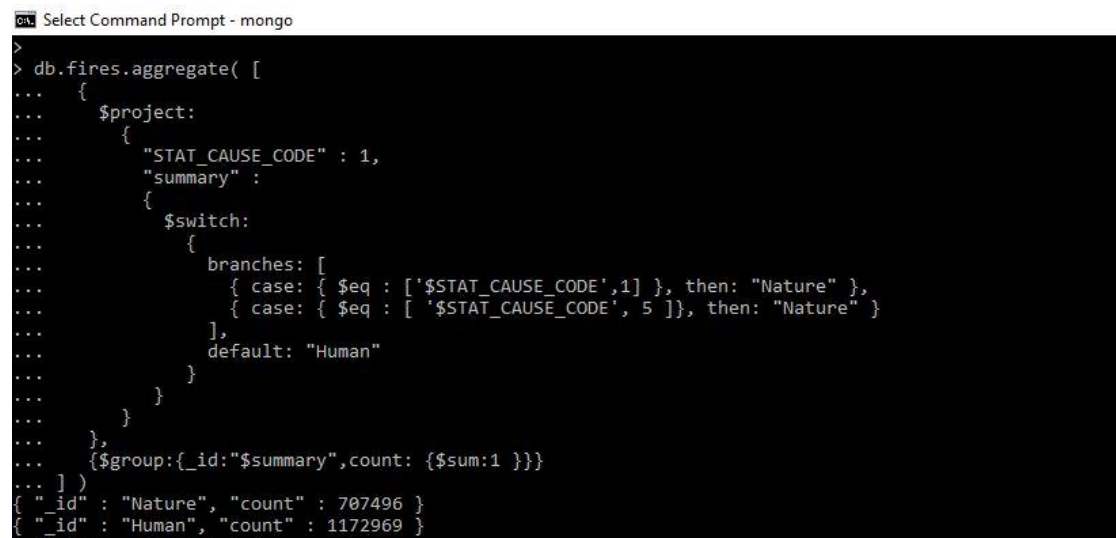
\$project Stat\_Cause\_Code=1, "summary":

\$switch branch '\$\$STAT\_CAUSE\_CODE', 1 then "Nature", '\$\$STAT\_CAUSE\_CODE', 5 then "Nature"

Default: "Human"

\$group: "\$summary", count: \$sum:1

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



```
Select Command Prompt - mongo
>
> db.fires.aggregate( [
...   {
...     $project:
...     {
...       "STAT_CAUSE_CODE" : 1,
...       "summary" :
...       {
...         $switch:
...         {
...           branches: [
...             { case: { $eq : ['$$STAT_CAUSE_CODE',1] }, then: "Nature" },
...             { case: { $eq : [ '$$STAT_CAUSE_CODE', 5 ]}, then: "Nature" }
...           ],
...           default: "Human"
...         }
...       }
...     }
...   },
...   { $group: { _id: "summary", count: { $sum: 1 } } }
... ] )
{ "_id" : "Nature", "count" : 707496 }
{ "_id" : "Human", "count" : 1172969 }
```

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

The total number of fires caused by humans is 1172969 and the total number of fires caused by nature is 707496. This proves that human cause more fires than nature and hence this query supports the statement made by the advocacy group

**Assumptions:** We are considering the nature causes to be lightening and debris, and all other causes are due to human activities. We are also considering that the miscellaneous and undefined fields are also human activity

## Query 4

### Question:

**5. How many wildfires were not reported by more than one unit/agency? (wildfires that were reported by only one agency?)**

### Translation:

Select the fires collection and use the aggregate function \$project

To show the documents with FIRE\_NAME and SOURCE\_REPORTING\_UNIT then we

Use \$group to group the documents according to FIRE\_NAME and sum of SOURCE\_REPORTING\_UNIT and then use \$match for to set the condition, where, No.\_of\_units\_reported is less than or equal to 1, then we use the function toArray().length to find the count

### Cleanup:

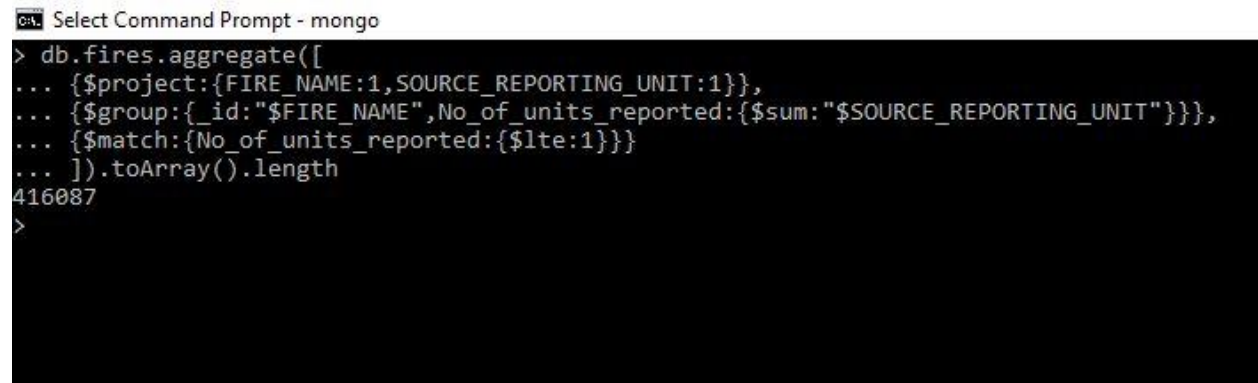
Use db.fires.aggregate

\$project FIRE\_NAME and SOURCE\_REPORTING\_UNIT,

\$group FIRE\_NAME and total SOURCE\_REPORTING\_UNIT,

\$match No\_of\_units\_reported <= 1, toArray().length

## Screen Shot of MongoDB Query/Code and Results

A screenshot of a terminal window titled "C:\> Select Command Prompt - mongo". The terminal shows a MongoDB aggregate query being executed. The query is: > db.fires.aggregate([ ... {\$project:{FIRE\_NAME:1,SOURCE\_REPORTING\_UNIT:1}}, ... {\$group:{\_id:"\$FIRE\_NAME",No\_of\_units\_reported:{\$sum:"\$SOURCE\_REPORTING\_UNIT"}}}, ... {\$match:{No\_of\_units\_reported:{\$lte:1}}} ... ]).toArray().length. The result of the query is 416087, followed by a prompt >.

```
C:\> Select Command Prompt - mongo
> db.fires.aggregate([
... {$project:{FIRE_NAME:1,SOURCE_REPORTING_UNIT:1}},
... {$group:{_id:"$FIRE_NAME",No_of_units_reported:{$sum:"$SOURCE_REPORTING_UNIT"}}},
... {$match:{No_of_units_reported:{$lte:1}}}
... ]).toArray().length
416087
>
```

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

The wildfires that were not reported by more than one agency or that was reported by only one agency were around 416087

**Result :** No. of documents = 416087



## Query 5

### Question:

**8. Which forest had the least number of fires?**

### Translation:

Select documents from the fires collection and use the aggregate function \$group

Use \$group based on "Source\_Reporting\_Unit\_Name" and count the no. of fires, then

Use \$sort to arrange the documents based on count values in ascending order

And then use \$limit to display forest with least no. of fires

### Cleanup:

Use db.fires.aggregate

\$group: "\$Source\_Reporting\_Unit\_Name", Count: \$sum:1

\$Sort: Count:1

\$limit : 1

## Screen Shot of MongoDB Query/Code and Results

Command Prompt - mongo

```
> Mongo: db.fires.aggregate([{"$group" : {_id:"$SOURCE_REPORTING_UNIT_NAME",COUNT:{$sum:1}}},{ $sort:{COUNT:1}},{ $limit:1}])
{ "_id" : "SAM D. HAMILTON NOXUBEE NWR", "COUNT" : 1 }
>
```

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

There are many forests (around 4441 forests - as in figure A below) that had least no. of fires but for the sake of answering the query we have used the limit function to limit the answer to one

**Assumption:** We are using source reporting unit name as the forest name because there is no actual forest name that is available in the given dataset

**Figure A**

Select Command Prompt - mongo

```
> db.fires.aggregate([{"$group: {_id:"$SOURCE_REPORTING_UNIT_NAME",COUNT:{$sum:1}}},{ $sort:{COUNT:1}}]).toArray().length
4441
>
```

## Query 6

### Question:

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

### Translation:

Select the documents from fires collection and use the aggregate function \$addfields, Using this, we calculate date\_diff which is obtained by subtracting Cont\_Date and Discovery\_Date Then we use \$match to match the condition where date\_diff is less than or equal to 2 Then we use \$project to show source\_reporting\_unit\_name Finally group the documents based on source\_reporting\_unit\_name using \$group and then find the total number of fires using Count function, then we use the function toArray().length to find the count

### Cleanup:

Use db.fires.aggregate

\$addFields : date\_diff : {subtract : ["\$CONT\_DATE", "\$DISCOVERY\_DATE"]}

\$match : date\_diff : \$lt : 3

\$project : {Source\_Reporting\_Unit\_Name":1}

\$group: "\$Source\_Reporting\_Unit\_Name", Firescount: \$sum:1

\$project : {Source\_Reporting\_Unit\_Name":1, Firescount:1 }

toArray().length

## Screen Shot of MongoDB Query/Code and Results

```
Command Prompt - mongo
> db.fires.aggregate([
...  {$addFields:{date_diff: { $subtract: ["$CONT_DATE", "$DISCOVERY_DATE"]}},
...  {$match:{date_diff: { $lt: 3}}},
...  {$project:{SOURCE_REPORTING_UNIT_NAME : 1}},
...
...  {$group: { _id:{SOURCE_REPORTING_UNIT_NAME: "$SOURCE_REPORTING_UNIT_NAME"}, Firescount:{ $sum:1}},
...  {$project:{Firescount:1,SOURCE_REPORTING_UNIT_NAME : 1}}
...  ]).toArray().length
4400
>
```

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

To identify how many days a fire existed we use the subtracted value of fire discovery date and fire contained date. Here the fire contained date is the day when the fire was controlled or put out.

**Assumption:** We are using source reporting unit name as the forest name because there is no actual forest name that is available in the given dataset

**Result :** No. of documents = 4400