



SQL-Mongo Project – Spatial Data of US Wildfires

BUAN 6320

Chirag Hamirani
Sudheer Akkiralala
Vinay Krishna Vuppala
R Vidya Reddy

Group 5

Activity	Chirag	Sudheer	Vinay	Vidya
Prepared Data Model and Created Physical DB	x	x	x	x
Loaded Data into Database		x		
Write SQL Queries	x	x		
Prepared Mongo Database	x			
Loaded data into Mongo DB	x			
Wrote Mongo Queries	x			
Prepared Report	x		x	x
Reviewed Report	x			x

Contents

Data Model	3
Assumptions/Notes About Data Entities and Relationships.....	3
Entity-Relationship Diagram	4
Physical Database	5
Assumptions/Notes About Data Set	7
Screen shot of Physical Database objects.....	7
Data in the Database.....	11
SQL Queries.....	12
Query 1.....	12
Question.....	12
Notes/Comments About SQL Query and Results (Include # of Rows in Result).....	12
Translation	12
Screen Shot of SQL Query and Results.....	12
Query 2.....	13
Question.....	13
Notes/Comments About SQL Query and Results (Include # of Rows in Result).....	13
Translation	13
Screen Shot of SQL Query and Results.....	13
Query 3.....	14
Question.....	14
Notes/Comments About SQL Query and Results (Include # of Rows in Result).....	14
Translation	14
Screen Shot of SQL Query and Results.....	14
Query 4.....	15
Question.....	15
Notes/Comments About SQL Query and Results (Include # of Rows in Result).....	15
Translation	15
Screen Shot of SQL Query and Results.....	15
Query 5.....	16
Question.....	16
Notes/Comments About SQL Query and Results (Include # of Rows in Result).....	16
Translation	16

Screen Shot of SQL Query and Results.....	16
Query 6.....	17
Question.....	17
Notes/Comments About SQL Query and Results (Include # of Rows in Result).....	17
Translation	17
Screen Shot of SQL Query and Results.....	17
Data Review for MongoDB.....	18
Assumptions/Notes About Data Collections, Attributes and Relationships between Collections	18
Physical Mongo Database	19
Assumptions/Notes About Data Set	19
Screen shot of Physical Database objects (Database, Collections and Attributes)	19
Data in the Database.....	24
MongoDB Queries/Code	25
Query 1.....	25
Question.....	25
Notes/Comments About MongoDB Query/Code and Results (Include # of Documents in Result) ...	25
Translation	25
Screen Shot of MongoDB Query/Code and Results.....	25
Query 2.....	26
Question.....	26
Notes/Comments About MongoDB Query/Code and Results (Include # of Documents in Result) ...	26
Translation	26
Screen Shot of MongoDB Query/Code and Results.....	26
Query 3.....	27
Question.....	27
Notes/Comments About MongoDB Query/Code and Results (Include # of Documents in Result) ...	27
Translation	27
Screen Shot of MongoDB Query/Code and Results.....	27
Query 4.....	28
Question.....	28
Notes/Comments About MongoDB Query/Code and Results (Include # of Documents in Result) ...	28
Translation	28
Screen Shot of MongoDB Query/Code and Results.....	28

Query 5.....	29
Question.....	29
Notes/Comments About MongoDB Query/Code and Results (Include # of Documents in Result) ...	29
Translation	29
Screen Shot of MongoDB Query/Code and Results.....	29
Query 6.....	30
Question.....	30
Notes/Comments About MongoDB Query/Code and Results (Include # of Documents in Result) ...	30
Translation	30
Screen Shot of MongoDB Query/Code and Results.....	30

Data Model

Assumptions/Notes About Data Entities and Relationships

Functional Requirements:

Functionally our project has two modules:

- Incidents being the Fires
- Incident reporting being the Reporting

Technical Requirements:

The data model has four main entities: the fires, the owners on whose land the fire occurred, the report on the fire incident and the agency who reported it.

The detailed information about the 4 main entities are as follows:

1.) Key Entity 1: Fires

The fires table contains the details about the incident like

- Discovery date and time
- Containment date and time
- Latitude and longitude
- State and County
- FIPS code and name

Relationships:

- One or many Fires can occur in one Owner's property
- One or many Fires can occur under each FireClass
- One or many Fires can happen due the same cause
- One fire can be reported only once

Assumptions:

- Multiple Fires can occur in the same property
- Multiple Fires can occur at the same time in different properties
- Each cause can have many fires categorized under it

2.) Key Entity 2: Owner

The Owner table has

- Details about the primary owner at the time of the fire
- Details about the person managing the property at the time of the fire

Relationships:

- One or many Fires can occur in the property owned by/ managed by the same person

3.) Key Entity 3: Reporting

The Reporting incidents have

- Source system and type
- Local fire report ID
- Local incident ID
- Fire code and name
- Reporting agency who reported the incident

Relationships:

- One to many reports can be there under one MTBS
- One to many reports can be there under one ICS identifier
- There can only to one report for one fire

Assumptions:

- Many fire reports can come under one complex
- The agency name is local but its recognized on a globally to back track the fire reporting

4.) Key Entity 4: Agency

Agency table has

- Agency name
- Department
- Wild and Role
- Geographic area where the unit is location
- County and State

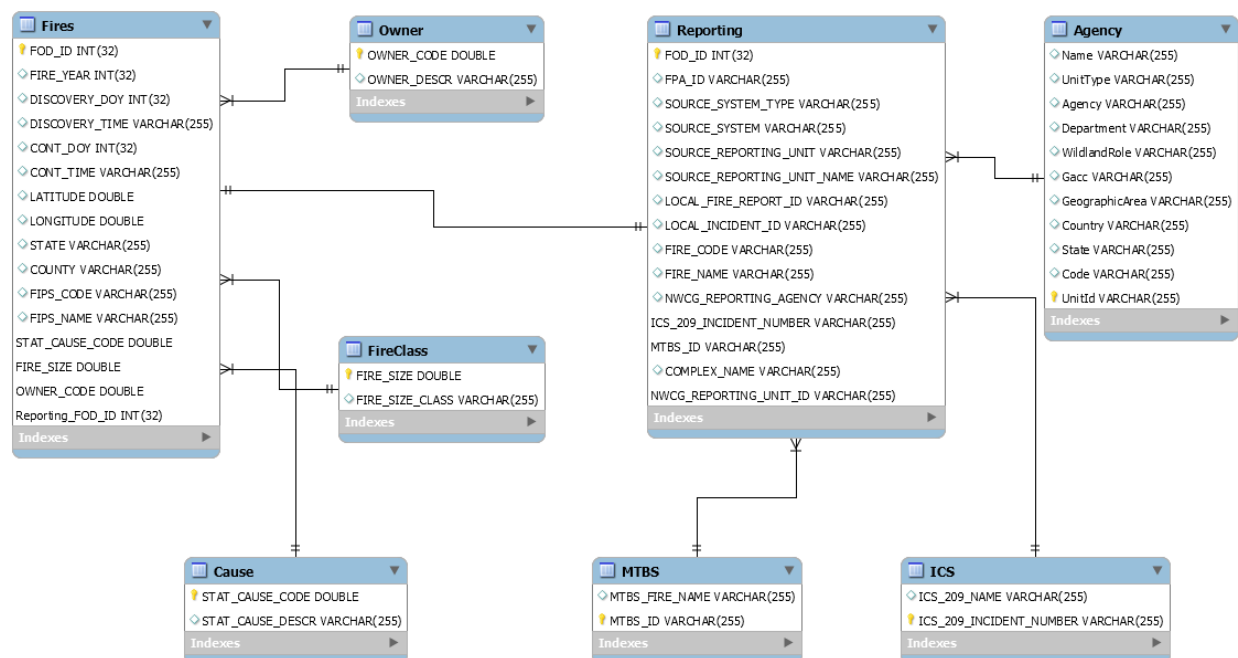
Relationships:

- Agency can have one to many reports
- Agency can report one or many incidents

Other than the four main entities, the data model has four more tables which are:

- 1.) FireClass: FireClass table has details about the size of the fire and what class it comes under depending on the perimeter of the fire.
- 2.) Cause: Cause table has details about the cause code of the fire and the cause description.
- 3.) ICS: ICS table has incident identifier name and number that is present in the report.
- 4.) MTBS: MTBS table has details about the trends in severity of the fires which are obtained from various aspects obtained from the wildlife fire reports.

Entity-Relationship Diagram



Physical Database

Assumptions/Notes About Data Set

- The shapes column was not used in our data model as it has only null values and it doesn't impose any significance.
- FIPS Name and County is the same.
- FIPS Code and County Code is the same.

Screen shot of Physical Database objects

Fires Table

The screenshot shows the MySQL Workbench interface. On the left, the 'SCHEMAS' pane displays the 'Project1' database structure, including tables like Agency, Cause, FireClass, Fires, ICS, MTBS, Owner, and Source. The 'Fires' table is selected, and its structure is shown in the 'Table: Fires' pane. The table has columns: FID_ID, FIRE_YEAR, OWNER_CODE, DISCOVERY_DOY, DISCOVERY_TIME, STAT_CAUSE_CODE, CONT_DOY, CONT_TIME, FIRE_SIZE, LATITUDE, LONGITUDE, STATE, COUNTY, FIPS_CODE, and FIPS_NAME. The 'Columns' pane lists the data types and constraints for each column.

The 'Query' pane shows the following SQL query:

```
1 use Project1;
2 select * from Fires limit 1359914;
3
```

The 'Result Grid' pane displays the query results. The first row is highlighted, showing the following data:

FID_ID	FIRE_YEAR	OWNER_CODE	DISCOVERY_DOY	DISCOVERY_TIME	STAT_CAUSE_CODE	CONT_DOY	CONT_TIME	FIRE_SIZE	LATITUDE	LONGITUDE	STATE	COUNTY	FIPS_CODE	FIPS_NAME
3	2004	13	152	1921	5	152	2024	0.1	38.98416667	-120.73555556	CA	17	017	El Dorado

The 'Action Output' pane shows the execution of the query, indicating that 1359914 rows were returned.

Owner Table

The screenshot shows the MySQL Workbench interface with the 'Owner' table selected in the Schemas pane. The table structure is as follows:

OWNER_CODE	OWNER_DESCR
2	BIA
3	NPS
4	FWS
5	USFS
6	OTHER FEDERAL
7	STATE
8	PRIVATE
9	TRIBAL
10	BOF
11	COUNTY
12	MUNICIPAL/LOC...
13	STATE OR PRIV...
14	MISSING/NOT S...
15	UNDEFINED FE...

The query results show 16 rows returned for the query: `select * from Owner;`

Cause Table

The screenshot shows the MySQL Workbench interface with the 'Cause' table selected in the Schemas pane. The table structure is as follows:

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
11	Powerline
12	Structure
13	Missing/Undefined

The query results show 13 rows returned for the query: `select * from Cause;`

FireClass Table

MySQL Workbench interface showing the FireClass table structure and a query result.

Table: FireClass

Columns:

- FIRE_SIZE (double PK)
- FIRE_SIZE_CLASS (varchar(255))

Query 7:

```
1 use Project1;
2 select * from FireClass;
3
```

Result Grid:

FIRE_SIZE	FIRE_SIZE_CLASS
0.0001	A
0.0002	A
0.00022	A
0.00034	A
0.0004	A
0.000459	A
0.0008	A
0.0009	A
0.001	A
0.00138	A
0.00159	A
0.0016	A
0.002	A
0.0025	A
...	...
FireClass 8	

Action Output:

Time	Action	Response	Duration / Fetch Time
01:17:43	select * from FireClass LIMIT 0, 50000	13605 row(s) returned	0.0017 sec / 0.013 sec

Reporting Table

MySQL Workbench interface showing the Reporting table structure and a query result.

Table: Reporting

Columns:

- FOD_ID (int(32) PK)
- FPA_ID (varchar(255))
- SOURCE_SYSTEM_TYPE (varchar(255))
- SOURCE_SYSTEM (varchar(255))
- SOURCE_REPORTING_UNIT (varchar(255))
- SOURCE_REPORTING_UNIT_NAME (varchar(255))
- LOCAL_FIRE_REPORT_ID (varchar(255))
- LOCAL_INCIDENT_ID (varchar(255))
- FIRE_CODE (varchar(255))
- FIRE_NAME (varchar(255))
- NWCO_REPORTING_UNIT_ID (varchar(255))
- ICS_209_INCIDENT_NUMBER (varchar(255))
- MTBS_ID (varchar(255))
- COMPLEX_NAME (varchar(255))

Query 7:

```
1 use Project1;
2 select * from Reporting limit 1337874;
3
```

Result Grid:

FOD_ID	FPA_ID	SOURCE_SYSTEM_TYPE	SOURCE_SYSTEM	SOURCE_REPORTING_UNIT	SOURCE_REPORTING_UNIT_NAME	LOCAL_FIRE_REPORT_ID	LOCAL_INCIDENT_ID
1	FS-1418826	FED	FS-FIRESTAT	0511	Pumas National Forest	1	PNF-47
2	FS-1418827	FED	FS-FIRESTAT	0503	Eldorado National Forest	13	13
3	FS-1418835	FED	FS-FIRESTAT	0503	Eldorado National Forest	27	021
4	FS-1418845	FED	FS-FIRESTAT	0503	Eldorado National Forest	43	6
5	FS-1418847	FED	FS-FIRESTAT	0503	Eldorado National Forest	44	7
6	FS-1418849	FED	FS-FIRESTAT	0503	Eldorado National Forest	54	8
7	FS-1418851	FED	FS-FIRESTAT	0503	Eldorado National Forest	58	9
8	FS-1418854	FED	FS-FIRESTAT	0514	Shasta-Trinity National Forest	3	02
9	FS-1418856	FED	FS-FIRESTAT	0514	Shasta-Trinity National Forest	5	03
10	FS-1418859	FED	FS-FIRESTAT	0503	Eldorado National Forest	61	10
11	FS-1418861	FED	FS-FIRESTAT	0503	Eldorado National Forest	64	11
12	FS-1418863	FED	FS-FIRESTAT	0503	Eldorado National Forest	71	14
13	FS-1418865	FED	FS-FIRESTAT	0503	Eldorado National Forest	91	26
14	FS-1418872	FED	FS-FIRESTAT	0503	Eldorado National Forest	99	21
15	FS-1418874	FED	FS-FIRESTAT	0503	Eldorado National Forest	102	28

Action Output:

Time	Action	Response	Duration / Fetch Time
01:27:39	select * from Reporting limit 1337874	1337874 row(s) returned	0.0017 sec / 2.141 sec

Agency Table

MySQL Workbench

Administration Schemas Query 7

Limit to 50000 rows

1 use Project1;
2 select * from Agency;
3

100% 1-3

Result Grid

UnitId	Name	UnitType	Agency	Department	WildlandRole	Gacc	GeographicArea	Country	State	Code
USAKAKS	Central Office	US State	DVF	AK	Incident Host Geographic	USAKACC	AK	US	AK	AKS
USAKALN	Alut Regional Native Corporation	Non-Government	ANC	NG	Resource Provider Only	USAKACC	AK	US	AK	ALN
USAKAMMX	Anchorage Municipality	US County/Local	C&L	AK	Incident Host Geographic	USAKACC	AK	US	AK	AMMX
USAKAMR	Alaska Maritime National Wildlife Refuge	US Federal	FWS	DOI	Incident Host Geographic	USAKACC	AK	US	AK	AMR
USAKAMRO	Aviation Management - Alaska Regional Office	US Federal	NBC	DOI	Incident Host Geographic	USAKACC	AK	US	AK	AMRO
USAKANP	Anchorage District Office	US Federal	BLM	DOI	Incident Host Geographic	USAKACC	AK	US	AK	AND
USAKANP	Aniakchak National Monument & Preserve	US Federal	NPS	DOI	Incident Host Geographic	USAKACC	AK	US	AK	ANP
USAKAOP	Alaska Regional Office	US Federal	NPS	DOI	Incident Host Geographic	USAKACC	AK	US	AK	AOP
USAKAPIK	Anchor Point Volunteer Fire Department	US County/Local	C&L	AK	Incident Host Geographic	USAKACC	AK	US	AK	APIK
USAKAPR	Alaska Peninsula National Wildlife Refuge	US Federal	FWS	DOI	Incident Host Geographic	USAKACC	AK	US	AK	APR
USAKARD	Arctic Field Office	US Federal	BLM	DOI	Incident Host Geographic	USAKACC	AK	US	AK	ARD
USAKARR	Arctic National Wildlife Refuge	US Federal	FWS	DOI	Incident Host Geographic	USAKACC	AK	US	AK	ARR
USAKASN	Arctic Slope Regional Native Corporation	Non-Government	ANC	NG	Resource Provider Only	USAKACC	AK	US	AK	ASN
USAKASO	Alaska State Office	US Federal	BLM	DOI	Incident Host Geographic	USAKACC	AK	US	AK	ASO
Agency 5										

Object Info Session

Table: Agency

Columns:

Column	Field Type
UnitId	varchar(255) PK
Name	varchar(255)
UnitType	varchar(255)
Agency	varchar(255)
Department	varchar(255)
WildlandRole	varchar(255)
Gacc	varchar(255)
GeographicArea	varchar(255)
Country	varchar(255)
State	varchar(255)
Code	varchar(255)

Action Output

Time	Action	Response	Duration / Fetch Time
01:15:45	select * from Agency LIMIT 0, 50000	5867 row(s) returned	0.00094 sec / 0.018...

Query Completed

ICS Table

MySQL Workbench

Administration Schemas Query 7

Limit to 50000 rows

1 use Project1;
2 select * from ICS;
3

100% 1-2

Result Grid

ICS_209_INCIDENT_NUMBER	ICS_209_NAME
-10699	ASH
-22215	Lebec
-80009	Bluff
000175	Old Man Fire
002168	BYRON
002565	EXPLOSIVE
003556	Russel Complex
003699	KIRKER
004462	COPOCO
005304	SHEPARD
005452	MORGAN
006329	OLD 2
006335	Wye
008116	Mit
009867	LOMA

Object Info Session

Table: ICS

Columns:

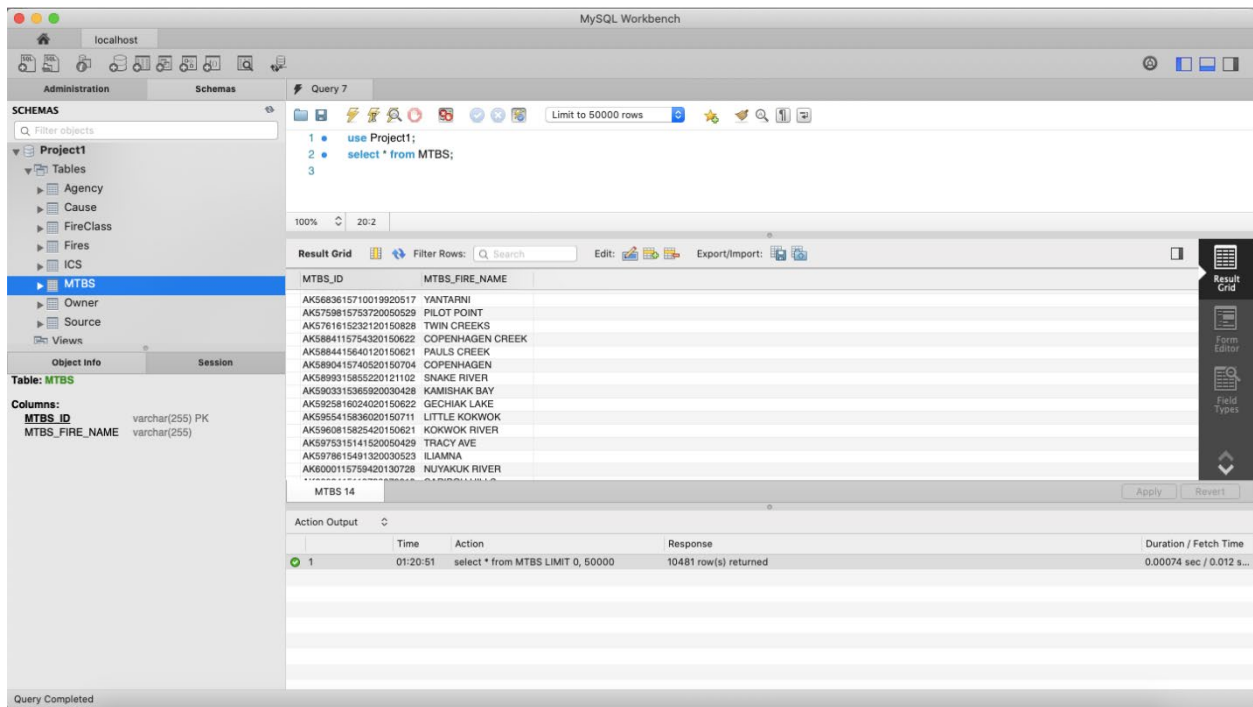
Column	Field Type
ICS_209_INCIDENT_NUMBER	varchar(255) PK
ICS_209_NAME	varchar(255)

Action Output

Time	Action	Response	Duration / Fetch Time
01:19:59	select * from ICS LIMIT 0, 50000	22736 row(s) returned	0.0013 sec / 0.019 sec

Query Completed

MTBS Table



Data in the Database

Table Name	Primary Key	Foreign Key	# of Rows in Table
Fires	FOD_ID	STAT_CAUSE_CODE FIRE_SIZE OWNER_CODE REPORTING_FOD_ID	1359914
Owner	OWNER_CODE		16
FireClass	FIRE_SIZE		13605
Cause	STAT_CAUSE_CODE		13
Reporting	FOD_ID	ICS_209_INCIDENT_NUMBER MTBS_ID NWCG_REPORTING_UNIT_ID	1337874
Agency	UnitID		5867
MTBS	MTBS_ID		10481
ICS	ICS_209_INCIDENT_NUMBER		22736

SQL Queries

Query 1

Question : A leading beverage company has announced a billion-dollar fund for removing debris from forests, rivers and mountains in the US. All states are interested. Which 2 states have the least chance to win a share of the fund?

Notes:

of rows in result: 2

Translation: States with less count of fires caused by debris burning have less chance of the fund.

Cleanup: select state, count of fires from fires join cause where cause description is 'Debris Burning' sorting by count of fires ascending order and limit 2.

The screenshot shows a SQL query editor with the following code:

```
8 • use Project1;
9 • select f.STATE
10 from Fires f join Cause c on f.STAT_CAUSE_CODE = c.STAT_CAUSE_CODE
11 where c.STAT_CAUSE_DESCR in ('Debris Burning')
12 group by f.STATE
13 ORDER BY count(f.FOD_ID) ASC
14 limit 2;
15
```

Below the query editor is a toolbar with options: Filter Rows, Export, Wrap Cell Content, and Fetch rows. To the right of the toolbar is a vertical sidebar with buttons for Result Grid, Form Editor, and a Read Only button.

The Result Grid shows the following data:

STATE
DC
RI

At the bottom of the interface, there is a tab labeled "Result 2" and a "Read Only" button.

Interpretations:

The 2 states which have the least chance to win a share of the fund are District of Columbia and Rhode Island as these have been the least successful in removing debris from forests, rivers and mountains among all the states.

Query 2

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

Notes:

- Assumption 1: forests with most fires caused by other causes apart from children will be inappropriate places.
- Assumption 2: higher the non-children fire count, less appropriate they are.

of rows in result: 5

Translation: select the forest where the fires are not caused by children and does not require the ban.

Cleanup: select Forest from Reporting, Fires and Cause where cause description is not children group by forest order by count of fires limit 5.

```

25 • use Project1;
26 • select distinct r.SOURCE_REPORTING_UNIT_NAME as 'Forest'
27 from Reporting r join Fires f1 on r.FOD_ID = f1.FOD_ID
28 join Cause c on f1.STAT_CAUSE_CODE = c.STAT_CAUSE_CODE
29 where c.STAT_CAUSE_DESCR not like 'Children'
30 group by Forest
31 order by count(r.fod_id) desc
32 limit 5;

```

Forest
Fire Department of New York
South Carolina Forestry Commission
Mississippi Forestry Commission
Texas Forest Service
Georgia Forestry Commission

Interpretations:

The top 5 forests where it is safe to allow children or in other words it would be least appropriate to ban children are the ones that are reported by Fire Department of New York, South Carolina Forestry Commission, Mississippi Forestry Commission, Texas Forest Service and Georgia Forestry Commission. These 5 forests are the ones that do not have any fires caused by children in the past in the fires database. Hence, these places are safe for children to hike, camp, etc.

Query 3

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

Notes:

- Assumption 1: Natural cause is lightning alone.
- Assumption 2: Human cause is everything apart from lightening, misc and undefined.

of rows in result: 1

Translation: Compare fires caused by lightning and other human reasons.

Clean up: select count of fires from fires join causes where cause description is 'Lightning' grouped by cause description displayed as 'Natural cause' and sum of counts as 'Human caused' from count, cause description from fires join cause group by cause description having cause description not in 'Lightning', 'Miscellaneous', 'Missing/Undefined'.

```

45 • use project1;
46 • select (select count(f.FOD_ID)
47   from fires f join cause c on f.STAT_CAUSE_CODE = c.STAT_CAUSE_CODE
48  where c.STAT_CAUSE_DESCR = 'Lightning'
49   group by f.STAT_CAUSE_CODE
50 ) as 'Natural Caused', sum(count2) as 'Human Caused'
51 from (
52   select count(f.FOD_ID) as 'count2', c.STAT_CAUSE_DESCR
53   from fires f join cause c on f.STAT_CAUSE_CODE = c.STAT_CAUSE_CODE
54   group by f.STAT_CAUSE_CODE having c.STAT_CAUSE_DESCR not in ('Lightning','Miscellaneous','Missing/Undefined')
55 ) as counts;

```

Natural Caused	Human Caused
222058	809068

Result 6 x Read Only

Interpretations:

We can see from the results that the fires caused by human actions are almost four times the fires caused by nature. The advocacy group was wrong to say that human actions and nature are equally to blame for wildfires. Human actions are more responsible for causing wildfires and necessary precautions have to be taken to avoid more forest fires in the future.

Query 4

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

Notes:

of rows in result: 132

Translation: select forest from reporting join fires where contained day minus discovered day is greater than 2 group by forest having count = 1.

Clean up: select source reporting unit name from reporting join fires where contained day - discovered day > 2 group by source reporting unit name having count = 1.

The screenshot shows a SQL query editor window with a toolbar at the top. The query is as follows:

```
69 • use project1;  
70 • select r.SOURCE_REPORTING_UNIT_NAME as 'Forest'  
71   from reporting r join fires f on r.fod_id = f.fod_id  
72   where f.CONT_DOY - f.DISCOVERY_DOY > 2  
73   group by r.SOURCE_REPORTING_UNIT_NAME  
74   having count(f.FOD_ID) = 1;  
75
```

Below the query editor is a 'Result Grid' section. It includes a 'Filter Rows' input field, an 'Export' button, and a 'Wrap Cell Content' checkbox. The results are displayed in a table with one column, 'Forest'. The table lists the following forests:

Forest
Hoosier National Forest
Crow Creek Agency
Minnesota Agency
Northern Idaho Agency
Umatilla Agency
Central California Agency
Channel Islands National Park
Pinnacles National Monument
Miccosukee Tribe
Lower Brule Agency

On the right side of the result grid, there are buttons for 'Result Grid', 'Form Editor', and 'Field Types'. At the bottom right, there is a 'Read Only' status indicator.

Interpretations:

The forests that had only one fire that lasted more than 2 days are the ones listed in the results.

Query 5

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

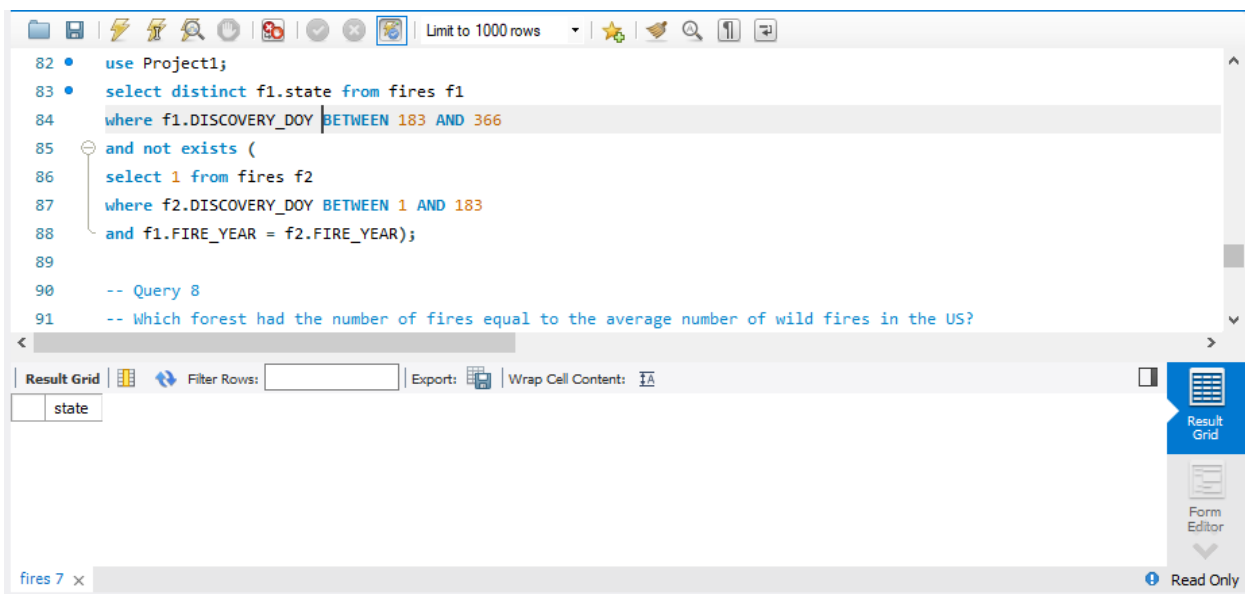
Notes:

- Assumption 1: For non-leap years the second half of the year is from 183 to 365.
- Assumption2: We have to check for states with fires in second half of the year, where they dint have fires in same year's first half

of rows in result: 0

Translation: select states from Fires with discovery date in second half of year only.

Cleanup: select state from Fires where Discovery Day of Year is between 183 and 366 and state is not in select state from fires where discovery say of the year is between 1 and 183.



```
82 • use Project1;
83 • select distinct f1.state from fires f1
84   where f1.DISCOVERY_DOY BETWEEN 183 AND 366
85   and not exists (
86     select 1 from fires f2
87     where f2.DISCOVERY_DOY BETWEEN 1 AND 183
88     and f1.FIRE_YEAR = f2.FIRE_YEAR);
89
90 -- Query 8
91 -- Which forest had the number of fires equal to the average number of wild fires in the US?
```

The screenshot shows a SQL query editor with a toolbar at the top. The query is written in a syntax-highlighted language. Below the query editor, there is a 'Result Grid' tab and a 'Filter Rows' input field. The 'Result Grid' tab is active, showing a single column header 'state'. The 'Filter Rows' input field is empty. The 'Export' button is visible next to the 'Filter Rows' input field. The 'Wrap Cell Content' button is also visible. The 'Result Grid' tab is highlighted in blue. The 'Form Editor' tab is visible below the 'Result Grid' tab. The 'Read Only' status is indicated at the bottom right.

Interpretations:

There are no states which have had fires only in the second half of the calendar year.

Query 6

Question: Which forest had the number of fires equal to the average number of wild fires in the US?

Notes:

- Assumption 1: All fires in given table are from US.
- Assumption 2: Since there so many records, we will not find an exact match for average value.
- Assumption 3: For business, let's find forests with above average count.

of rows in result: 360

Translation: Select forests in source reporting unit name from reporting with sum of fires reported equal to average number of fires in the us.

Clean up: Select source reporting unit name from reporting where sum of fires reported equal to select average of fires from reporting.

```

99 • select SOURCE_REPORTING_UNIT_NAME
100   from reporting
101  group by SOURCE_REPORTING_UNIT_NAME
102  having count(*) > (
103    select avg(count)
104    from ( select count(*) as count
105          from reporting
106          group by SOURCE_REPORTING_UNIT_NAME
107        ) as counts
108  );

```

Result Grid

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

reporting 10 x

Interpretations:

The forests that have fires which are above the average number of fires occurring in the US are the ones listed above in the results. These forests have a higher danger of catching fire or have had higher occurrences of fires than the other forests in the US which means these forests are not as safe to visit. It is advisable to not pay visits to these and find alternative forests instead. It is also advisable to evacuate any animals that might be in danger or take precautions to provide them safe habitats in case these forests have high animal population.

Data Review for MongoDB

Attributes and Relationships between Collections

While functional and technical requirements have been kept same for mongo database as well, there are few notes about data migration which will explain the collections and data in it.

Notes About Collections & Data Migration

1. Collections in mongo are constructed with same structure as that of sql relational database for ease of understanding
2. Every table had data extracted from sql into csv files, which were then imported in mongo using mongo compass
3. All the relational tables are made into collections with same functional value
4. While migrating, fires and reporting tables had over 13 million records, these were split into over 20 files each to facilitate data load
5. All the data load happened flawlessly, except one file in reporting collection, during which connection to database was lost. The file was then reimported, which may have led to certain duplicate records. However, querying is done to avoid any such duplicates affecting the results
6. Since the collections and tables have same functional values, the results are expected to reflect similar conclusions. Any difference can open opportunity to discussion of ways in which both software function and error in querying.

Physical Mongo Database

Assumptions/Notes About Data Set

- The shapes field was not used as it has only null values and it doesn't impose any significance.
- FIPS Name field and County field are the same.
- FIPS Code field and County field Code are the same.

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

Project Database

MongoDB Compass - localhost:27017
Connect View Help

My Cluster
5 DBS 16 COLLECTIONS
HOST: localhost:27017
CLUSTER: Standalone
EDITION: MongoDB 4.2.1 Community

Filter your data

- > admin
- > config
- > employees
- > local
- ▼ project_fires
 - agency
 - cause
 - fireclass
 - fires
 - ics
 - mtbs
 - owners
 - reporting

Collections
CREATE COLLECTION

Collection Name	Documents	Avg. Document Size	Total Document Size	Num. Indexes	Total Index Size	Properties
agency	5,868	281.3 B	1.7 MB	1	69.6 KB	
cause	14	77.4 B	1.1 KB	1	20.5 KB	
fireclass	13,607	65.7 B	893.4 KB	1	135.2 KB	
fires	1,359,947	330.9 B	450.1 MB	1	13.5 MB	
ics	22,737	94.7 B	2.2 MB	1	217.1 KB	
mtbs	10,482	89.3 B	936.5 KB	1	110.6 KB	
owners	17	66.2 B	1.1 KB	1	20.5 KB	
reporting	1,292,027	438.8 B	566.9 MB	1	12.9 MB	

Agency Collection

MongoDB Compass - localhost:27017/project_fires.agency

Connect View Collection Help

My Cluster

- 5 DBS 16 COLLECTIONS
- HOST: localhost:27017
- CLUSTER: Standalone
- EDITION: MongoDB 4.2.1 Community

Filter your data

- admin
- config
- employees
- local
- project_fires
 - agency
 - cause
 - fireclass
 - fires
 - ics
 - mtbs
 - owners
 - reporting

project_fires.agency

DOCUMENTS 5.9k TOTAL SIZE 1.6MB AVG. SIZE 261B INDEXES 1 TOTAL SIZE 68.0KB AVG. SIZE 68.0KB

Documents Aggregations Schema Explain Plan Indexes Validation

Filter

INSERT DOCUMENT VIEW LIST TABLE

Displaying documents 1 - 20 of 5668

#	agency	_id ObjectId	UnitId String	Name String	Unittype String	Agency String	Departme
1		5dec9d4e3dda2f573c867cd7	"CAABABIN"	"Alberta Environment and Susta	"International Country Subdivi	"NULL"	"NULL"
2		5dec9d4e3dda2f573c867cd8	"CABCBIN"	"British Columbia Wildfire Man	"International Country Subdivi	"NULL"	"NULL"
3		5dec9d4e3dda2f573c867cd9	"CANBCTFC"	"Canadian Interagency Forest F	"Interagency"	"NULL"	"NULL"
4		5dec9d4e3dda2f573c867cda	"CAMBMBIN"	"Manitoba Wildfire Program"	"International Country Subdivi	"NULL"	"NULL"
5		5dec9d4e3dda2f573c867cdb	"CANMBIN"	"New Brunswick Forest Fire Man	"International Country Subdivi	"NULL"	"NULL"
6		5dec9d4e3dda2f573c867cdc	"CANLNLIN"	"Newfoundland and Labrador Fir	"International Country Subdivi	"NULL"	"NULL"
7		5dec9d4e3dda2f573c867cdd	"CANNSIN"	"Nova Scotia Forest Protection	"International Country Subdivi	"NULL"	"NULL"
8		5dec9d4e3dda2f573c867cde	"CANTNTIN"	"Northwest Territories Forest	"International Country Subdivi	"NULL"	"NULL"
9		5dec9d4e3dda2f573c867cdf	"CAONONIN"	"Ontario Aviation"	"International Country Subdivi	"NULL"	"NULL"
10		5dec9d4e3dda2f573c867ce0	"CAPEPEN"	"Prince Edward Island Forests"	"International Country Subdivi	"NULL"	"NULL"
11		5dec9d4e3dda2f573c867ce1	"CAQCPCIN"	"Parks Canada National Fire Ma	"International Country Subdivi	"NULL"	"NULL"
12		5dec9d4e3dda2f573c867ce2	"CAQCQCIN"	"Quebec SOPFEU"	"International Country Subdivi	"NULL"	"NULL"
13		5dec9d4e3dda2f573c867ce3	"CASKSKIN"	"Saskatchewan Wildfire Managem	"International Country Subdivi	"NULL"	"NULL"
14		5dec9d4e3dda2f573c867ce4	"CAYTYTIN"	"Yukon Wildland Fire Managemen	"International Country Subdivi	"NULL"	"NULL"

Cause Collection

MongoDB Compass - localhost:27017/project_fires.cause

Connect View Collection Help

My Cluster

- 5 DBS 16 COLLECTIONS
- HOST: localhost:27017
- CLUSTER: Standalone
- EDITION: MongoDB 4.2.1 Community

Filter your data

- admin
- config
- employees
- local
- project_fires
 - agency
 - cause
 - fireclass
 - fires
 - ics
 - mtbs
 - owners
 - reporting

project_fires.cause

DOCUMENTS 14 TOTAL SIZE 1.1KB AVG. SIZE 77B INDEXES 1 TOTAL SIZE 20.0KB AVG. SIZE 20.0KB

Documents Aggregations Schema Explain Plan Indexes Validation

Filter

INSERT DOCUMENT VIEW LIST TABLE

Displaying documents 1 - 14 of 14

#	cause	_id ObjectId	STAT_CAUSE_CODE String	STAT_CAUSE_DESCR Mixed
1		5dec9d803dda2f573c8693c3	"1"	"Lightning"
2		5dec9d803dda2f573c8693c4	"2"	"Equipment Use"
3		5dec9d803dda2f573c8693c5	"3"	"Smoking"
4		5dec9d803dda2f573c8693c6	"4"	"Campfire"
5		5dec9d803dda2f573c8693c7	"5"	"Debris Burning"
6		5dec9d803dda2f573c8693c8	"6"	"Railroad"
7		5dec9d803dda2f573c8693c9	"7"	"Arson"
8		5dec9d803dda2f573c8693ca	"8"	"Children"
9		5dec9d803dda2f573c8693cb	"9"	"Miscellaneous"
10		5dec9d803dda2f573c8693cc	"10"	"Fireworks"
11		5dec9d803dda2f573c8693cd	"11"	"Powerline"
12		5dec9d803dda2f573c8693ce	"12"	"Structure"
13		5dec9d803dda2f573c8693cf	"13"	"Missing/Undefined"
14		5dec9d803dda2f573c8693d0	""	null

Fireclass Collection

MongoDB Compass - localhost:27017/project_fires.fireclass

Connect View Collection Help

My Cluster

5 DBS 16 COLLECTIONS

HOST localhost:27017

CLUSTER Standalone

EDITION MongoDB 4.2.1 Community

Filter your data

admin

config

employees

local

project_fires

agency

cause

fireclass

fires

ics

mtbs

owners

reporting

project_fires.fireclass Documents

DOCUMENTS 13.6k TOTAL SIZE 872.5KB AVG. SIZE 66B INDEXES 1 TOTAL SIZE 132.0KB AVG. SIZE 132.0KB

Documents Aggregations Schema Explain Plan Indexes Validation

FILTER

INSERT DOCUMENT VIEW LIST TABLE

Displaying documents 1 - 20 of 13607

#	_id ObjectId	FIRE_SIZE String	FIRE_SIZE_CLASS String
1	5dec9db33dda2f573c8693d1	"0.00001"	"A"
2	5dec9db33dda2f573c8693d2	"0.00009"	"A"
3	5dec9db33dda2f573c8693d3	"0.0001"	"A"
4	5dec9db33dda2f573c8693d4	"0.0002"	"A"
5	5dec9db33dda2f573c8693d5	"0.00022"	"A"
6	5dec9db33dda2f573c8693d6	"0.00034"	"A"
7	5dec9db33dda2f573c8693d7	"0.0004"	"A"
8	5dec9db33dda2f573c8693d8	"0.000459"	"A"
9	5dec9db33dda2f573c8693d9	"0.0008"	"A"
10	5dec9db33dda2f573c8693da	"0.0009"	"A"
11	5dec9db33dda2f573c8693db	"0.001"	"A"
12	5dec9db33dda2f573c8693dc	"0.00136"	"A"
13	5dec9db33dda2f573c8693dd	"0.00159"	"A"
14	5dec9db33dda2f573c8693de	"0.0016"	"A"

Type here to search

6:24 PM 12/8/2019

Fires Collection

MongoDB Compass - localhost:27017/project_fires.fires

Connect View Collection Help

My Cluster

5 DBS 16 COLLECTIONS

HOST localhost:27017

CLUSTER Standalone

EDITION MongoDB 4.2.1 Community

Filter your data

admin

config

employees

local

project_fires

agency

cause

fireclass

fires

ics

mtbs

owners

reporting

project_fires.fires Documents

DOCUMENTS 1.4m TOTAL SIZE 429.2MB AVG. SIZE 331B INDEXES 1 TOTAL SIZE 12.9MB AVG. SIZE 12.9MB

Documents Aggregations Schema Explain Plan Indexes Validation

FILTER

INSERT DOCUMENT VIEW LIST TABLE

Displaying documents 1 - 20 of 1359947

#	FID_ID String	FIRE_YEAR String	OWNER_CODE String	DISCOVERY_DOY String	DISCOVERY_TIME String	STAT_CAUSE
1	"1"	"2005"	"5"	"33"	"1300"	"9"
2	"2"	"2004"	"5"	"133"	"0845"	"1"
3	"3"	"2004"	"13"	"152"	"1921"	"5"
4	"4"	"2004"	"5"	"100"	"1600"	"1"
5	"5"	"2004"	"5"	"100"	"1600"	"1"
6	"6"	"2004"	"5"	"102"	"1800"	"1"
7	"7"	"2004"	"5"	"103"	"1800"	"1"
8	"8"	"2005"	"13"	"67"	"1300"	"5"
9	"9"	"2005"	"13"	"74"	"1200"	"5"
10	"10"	"2004"	"5"	"103"	"1800"	"1"
11	"11"	"2004"	"5"	"104"	"1830"	"1"
12	"12"	"2004"	"5"	"104"	"1730"	"1"
13	"13"	"2004"	"5"	"247"	"1600"	"1"
14	"14"	"2004"	"14"	"272"	"1200"	"1"

Type here to search

6:24 PM 12/8/2019

ICS collection

MongoDB Compass - localhost:27017/project_fires.ics

Connect View Collection Help

My Cluster

- 5 DBS 16 COLLECTIONS
- HOST localhost:27017
- CLUSTER Standalone
- EDITION MongoDB 4.2.1 Community

Filter your data

- admin
- config
- employees
- local
- project_fires
 - agency
 - cause
 - fireclass
 - fires
 - ics
 - mtbs
 - owners
 - reporting

project_fires.ics

DOCUMENTS 22.7k TOTAL SIZE 2.1MB AVG. SIZE 95B INDEXES 1 TOTAL SIZE 212.0KB AVG. SIZE 212.0KB

Documents Aggregations Schema Explain Plan Indexes Validation

FILTER OPTIONS FIND RESET

INSERT DOCUMENT VIEW LIST TABLE

Displaying documents 1 - 20 of 22737

	_id ObjectId	ICS_209_INCIDENT_NUMBER String	ICS_209_NAME String
1	5dec9d083dda2f573c86c8f8	"-10699"	"ASH"
2	5dec9d083dda2f573c86c8f9	"-22215"	"Lebec"
3	5dec9d083dda2f573c86c8fa	"-80009"	"Bluff"
4	5dec9d083dda2f573c86c8fb	"000175"	"Old Man Fire"
5	5dec9d083dda2f573c86c8fc	"002168"	"BYRON"
6	5dec9d083dda2f573c86c8fd	"002565"	"EXPLOSIVE"
7	5dec9d083dda2f573c86c8fe	"003556"	"Russel Complex"
8	5dec9d083dda2f573c86c8ff	"003899"	"KIRKER"
9	5dec9d083dda2f573c86c900	"004462"	"COPCO"
10	5dec9d083dda2f573c86c901	"005304"	"SHEPARD"
11	5dec9d083dda2f573c86c902	"005452"	"MORGAN"
12	5dec9d083dda2f573c86c903	"006329"	"OLD 2"
13	5dec9d083dda2f573c86c904	"006335"	"Wye"
14	5dec9d083dda2f573c86c905	"008116"	"Hill"

MTBS Collection

MongoDB Compass - localhost:27017/project_fires.mtbs

Connect View Collection Help

My Cluster

- 5 DBS 16 COLLECTIONS
- HOST localhost:27017
- CLUSTER Standalone
- EDITION MongoDB 4.2.1 Community

Filter your data

- admin
- config
- employees
- local
- project_fires
 - agency
 - cause
 - fireclass
 - fires
 - ics
 - mtbs
 - owners
 - reporting

project_fires.mtbs

DOCUMENTS 10.5k TOTAL SIZE 914.6KB AVG. SIZE 89B INDEXES 1 TOTAL SIZE 108.0KB AVG. SIZE 108.0KB

Documents Aggregations Schema Explain Plan Indexes Validation

FILTER OPTIONS FIND RESET

INSERT DOCUMENT VIEW LIST TABLE

Displaying documents 1 - 20 of 10482

	<pre>{ "_id": ObjectId("5dec9deb3dda2f573c8721c9"), "MTBS_ID": "AFS-8307-19950614", "MTBS_FIRE_NAME": "CLEARWATER #2" }</pre>
	<pre>{ "_id": ObjectId("5dec9deb3dda2f573c8721ca"), "MTBS_ID": "AKS674215793820060522", "MTBS_FIRE_NAME": "MESHNIK RIVER" }</pre>
	<pre>{ "_id": ObjectId("5dec9deb3dda2f573c8721cb"), "MTBS_ID": "AKS683615710019920517", "MTBS_FIRE_NAME": "YANTARNI" }</pre>
	<pre>{ "_id": ObjectId("5dec9deb3dda2f573c8721cc"), "MTBS_ID": "AKS75981573720060529", "MTBS_FIRE_NAME": "PILOT POINT" }</pre>
	<pre>{ "_id": ObjectId("5dec9deb3dda2f573c8721cd"), "MTBS_ID": "AKS761615232120150028", "MTBS_FIRE_NAME": "TWIN CREEKS" }</pre>

Owners Collection

MongoDB Compass - localhost:27017/project_fires.owners

Connect View Collection Help

My Cluster

5 DBS 16 COLLECTIONS

HOST localhost:27017

CLUSTER Standalone

EDITION MongoDB 4.2.1 Community

Filter your data

admin

config

employees

local

project_fires

agency

cause

fireclass

fires

ics

mtbs

owners

reporting

project_fires.owners

Documents

Aggregations Schema Explain Plan Indexes Validation

17 DOCUMENTS 1.1KB TOTAL SIZE 66B AVG. SIZE 1 INDEXES 20.0KB TOTAL SIZE 20.0KB AVG. SIZE

INSERT DOCUMENT VIEW LIST TABLE

Displaying documents 1 - 17 of 17

_id	Objectid	OWNER_CODE	String	OWNER_DESCR	Mixed
1	5dec9dfe3dda2f573c874abb	"0"		"FOREIGN"	
2	5dec9dfe3dda2f573c874abc	"1"		"BLN"	
3	5dec9dfe3dda2f573c874abd	"2"		"BEA"	
4	5dec9dfe3dda2f573c874abe	"3"		"NPS"	
5	5dec9dfe3dda2f573c874abf	"4"		"FWS"	
6	5dec9dfe3dda2f573c874ac0	"5"		"USFS"	
7	5dec9dfe3dda2f573c874ac1	"6"		"OTHER FEDERAL"	
8	5dec9dfe3dda2f573c874ac2	"7"		"STATE"	
9	5dec9dfe3dda2f573c874ac3	"8"		"PRIVATE"	
10	5dec9dfe3dda2f573c874ac4	"9"		"TRIBAL"	
11	5dec9dfe3dda2f573c874ac5	"10"		"BOR"	
12	5dec9dfe3dda2f573c874ac6	"11"		"COUNTY"	
13	5dec9dfe3dda2f573c874ac7	"12"		"MUNICIPAL/LOCAL"	
14	5dec9dfe3dda2f573c874ac8	"13"		"STATE OR PRIVATE"	

Reporting Collection

MongoDB Compass - localhost:27017/project_fires.reporting

Connect View Collection Help

My Cluster

5 DBS 16 COLLECTIONS

HOST localhost:27017

CLUSTER Standalone

EDITION MongoDB 4.2.1 Community

Filter your data

admin

config

employees

local

project_fires

agency

cause

fireclass

fires

ics

mtbs

owners

reporting

project_fires.reporting

Documents

Aggregations Schema Explain Plan Indexes Validation

1.3m DOCUMENTS 540.7MB TOTAL SIZE 439B AVG. SIZE 1 INDEXES 12.3MB TOTAL SIZE 12.3MB AVG. SIZE

INSERT DOCUMENT VIEW LIST TABLE

Displaying documents 1 - 20 of 1292027

_id	Objectid	FID_ID	String	FPA_ID	String	SOURCE_SYSTEM_TYPE	String	SOURCE_SYSTEM	String	SOURCE_R
1	5dec9e423dda2f573c874acc	"1"		"FS-1418826"		"FED"		"FS-FIRESTAT"		"0511"
2	5dec9e423dda2f573c874acd	"2"		"FS-1418827"		"FED"		"FS-FIRESTAT"		"0503"
3	5dec9e423dda2f573c874ace	"3"		"FS-1418835"		"FED"		"FS-FIRESTAT"		"0503"
4	5dec9e423dda2f573c874acf	"4"		"FS-1418845"		"FED"		"FS-FIRESTAT"		"0503"
5	5dec9e423dda2f573c874ad0	"5"		"FS-1418847"		"FED"		"FS-FIRESTAT"		"0503"
6	5dec9e423dda2f573c874ad1	"6"		"FS-1418849"		"FED"		"FS-FIRESTAT"		"0503"
7	5dec9e423dda2f573c874ad2	"7"		"FS-1418851"		"FED"		"FS-FIRESTAT"		"0503"
8	5dec9e423dda2f573c874ad3	"8"		"FS-1418854"		"FED"		"FS-FIRESTAT"		"0514"
9	5dec9e423dda2f573c874ad4	"9"		"FS-1418856"		"FED"		"FS-FIRESTAT"		"0514"
10	5dec9e423dda2f573c874ad5	"10"		"FS-1418859"		"FED"		"FS-FIRESTAT"		"0503"
11	5dec9e423dda2f573c874ad6	"11"		"FS-1418861"		"FED"		"FS-FIRESTAT"		"0503"
12	5dec9e423dda2f573c874ad7	"12"		"FS-1418863"		"FED"		"FS-FIRESTAT"		"0503"
13	5dec9e423dda2f573c874ad8	"13"		"FS-1418865"		"FED"		"FS-FIRESTAT"		"0503"
14	5dec9e423dda2f573c874ad9	"14"		"FS-1418872"		"FED"		"FS-FIRESTAT"		"0503"

Data in the Database

Collection Name	Relationships with Other Collections (if any)	# of Documents in Collection
Fires	Owner collection (Owner_Code) FireClass collection (Fire_Size) Cause Collection (Stat_Cause) Reporting Collection (FOD_ID)	1359914
Owner	Fire Collection (Owner_Code)	16
FireClass	Fire Collection (Fire_Size)	13605
Cause	Fire Collection (Stat_Cause)	13
Reporting	Agency Collection (UnitId) MTBS Collection (MTBS_ID) ICS Collection (ICS_Incident_Number) Fire Collection (FOD_ID)	1337874
Agency	Reporting Collection (UnitId)	5867
MTBS	Reporting Collection (MTBS_ID)	10481
ICS	Reporting Collection (ICS_Incident_Number)	22736

MongoDB Queries/Code

Query 1

Question : A leading beverage company has announced a billion-dollar fund for removing debris from forests, rivers and mountains in the US. All states are interested. Which 2 states have the least chance to win a share of the fund?

Notes:

of rows in result: 2

Translation: Match cause code to '5' for Debris, Group by State and generate count. Project state, sort by count, limit at 2

Screen Shot of MongoDB Query/Code and Results

The screenshot shows the MongoDB Query and Results interface. The top section displays the query for the `$match` stage, which filters documents where `STAT_CAUSE_CODE` is '5'. The output shows two sample documents with their fields: `_id`, `FOD_ID`, `FIRE_YEAR`, `OWNER_CODE`, `DISCOVERY_DOY`, `DISCOVERY_TIME`, `STAT_CAUSE_CODE`, `CONT_DOY`, and `CONT_TIME`.

The bottom section displays the query for the `$group` stage, which groups documents by state (`$STATE`) and calculates the count. The output shows two sample documents with their fields: `_id` and `count`.

The screenshot shows the MongoDB Atlas interface. The top section displays the aggregation pipeline, which includes a `project` stage to project the state (`$_id:1`) and a `sort` stage to sort by count (`{count:1}`). The results show two documents with the state (`_id`) and count (`count`).

State	Count
DC	4642
PR	2520

Interpretations:

The 2 states which have the least chance to win a share of the fund are DC & PR as these have been the least successful in removing debris from forests, rivers and mountains among all the states.

Query 2

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

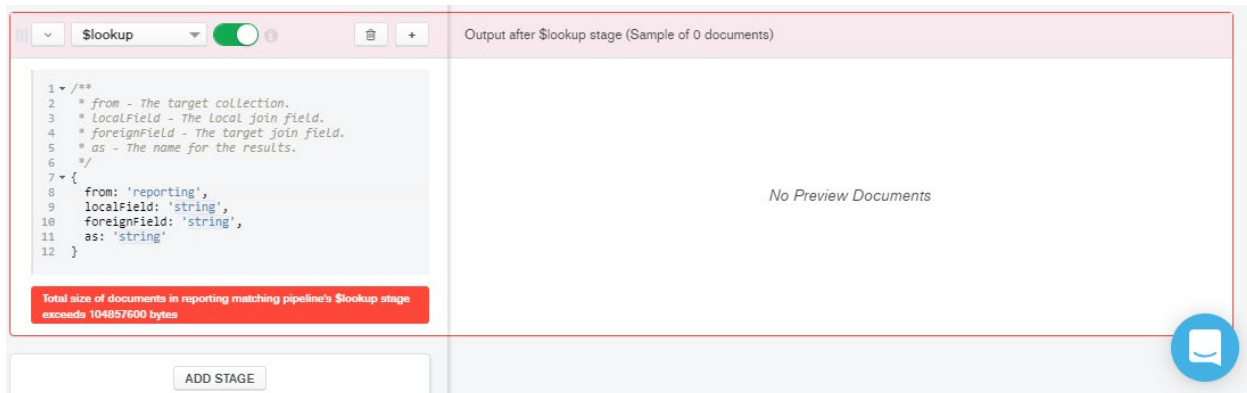
Notes:

- Assumption 1: forests with most fires caused by other causes apart from children will be inappropriate places.
- Assumption 2: higher the non-children fire count, lesser appropriate there are.

of rows in result: NA

Translation: Match cause code as '8' for children, lookup reporting collection with FOD_ID, group by forests and generate count. Project Forest, sort by count desc, limit 5

Screen Shot of MongoDB Query/Code and Results



Interpretations:

Compass throws error for lookup between fires and reporting table, saying the aggregation pipeline exceeds 104mb. This has no possible work around without altering database structure.

Query 3

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

Notes:

- Assumption 1: Natural cause is lightening alone.
- Assumption 2: Human cause is everything apart from lightening, misc and undefined.

of rows in result:

Translation: 1

Screen Shot of MongoDB Query/Code and Results

The screenshot displays the MongoDB Query and Results interface. The top section shows the query for the `$group` stage, which groups documents by `STAT_CAUSE_CODE` and calculates the count. The output shows two groups: one for `"6"` with a count of 28570, and another for `"5"` with a count of 296178.

The bottom section shows the query for the `$lookup` stage, which performs a left outer join between the `cause` collection and the `STAT_CAUSE_CODE` field. The output shows two documents: one for `"7"` with a count of 9867, and another for `"6"` with a count of 490. The `Cause_Descr` field is an array containing the details of the cause.

The bottom of the interface shows the query editor with the following query:

```
1 //**
2 * from - The target collection.
3 * localField - The local join field.
4 * foreignField - The target join field.
5 * as - The name for the results.
6 */
7 {
8   from: 'cause',
9   localField: '_id',
10  foreignField: 'STAT_CAUSE_CODE',
11  as: 'Cause_Descr'
12 }
```

The query is executed, and the results are displayed in a table view. The first document is shown, with `_id: "1"`, `count: 223903`, and `Cause_Descr` as an array containing one object with `STAT_CAUSE_DESCR: "Lightning"`.

Interpretations:

We can see from the results that the fires caused by human actions are almost four times the fires caused by nature. The advocacy group was wrong to say that human actions and nature are equally to blame for wildfires. Human actions are more responsible for causing wildfires and necessary precautions have to be taken to avoid more forest fires in the future.

Query 4

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

Notes:

of rows in result: NA

Translation: Project 'Diff' as subtract between int values of discovery and contained day. Match diff with greater than 2. Lookup fires FOD_IDs in reporting collection. Group by forests and generate count. Match count to 1 for output.

Screen Shot of MongoDB Query/Code and Results

The screenshot displays the MongoDB Compass interface with a query pipeline consisting of three stages: \$project, \$match, and \$lookup. The \$lookup stage is highlighted in red, indicating an error.

\$project stage:

```
1 /**
2  * specifications - The fields to
3  * include or exclude.
4  */
5 {
6   diff : {
7     $subtract : {
8       parseInt(data."$CONT_DOY"),
9       parseInt(data."$DISCOVERY_DOY")
10    }
11  }
12 }
```

\$match stage:

```
1 /**
2  * query - The query in MQL.
3  */
4 {
5   diff : { $gt : 2 }
6 }
```

\$lookup stage:

```
1 /**
2  * from - The target collection.
3  * localField - The local join field.
4  * foreignField - The target join field.
5  * as - The name for the results.
6  */
7 {
8   from: 'reporting',
9   localField: 'string',
10  foreignField: 'string',
11  as: 'string'
12 }
```

Output after \$lookup stage (Sample of 0 documents):

No Preview Documents

Error Message: Total size of documents in reporting matching pipeline's \$lookup stage exceeds 104857600 bytes

Interpretations:

Compass throws error for lookup between fires and reporting table, saying the aggregation pipeline exceeds 104mb. This has no possible work around without altering database structure.

Query 5

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

Notes:

- Assumption 1: For non-leap years the second half of the year is from 183 to 365.

of rows in result: 0

Translation:

Screen Shot of MongoDB Query/Code and Results

The screenshot shows the MongoDB Query and Results interface for the `project_fires.fires` collection. The query is:

```
1 /**
2  * query - The query in MQL.
3  */
4 {
5   $match: {
6     DISCOVERY_DOY: {
7       $gt: "183"
8     }
9   }
10 }
```

The output shows two sample documents (though the result set is empty):

```
{ "_id": "ObjectId('5dec9cc13dda2f573c85b984')", "FOD_ID": "1", "FIRE_YEAR": "2005", "OWNER_CODE": "5", "DISCOVERY_DOY": "33", "DISCOVERY_TIME": "1300", "STAT_CAUSE_CODE": "0" }
{ "_id": "ObjectId('5dec9cc13dda2f573c85b98b')", "FOD_ID": "8", "FIRE_YEAR": "2005", "OWNER_CODE": "13", "DISCOVERY_DOY": "67", "DISCOVERY_TIME": "1300", "STAT_CAUSE_CODE": "0" }
```

The interface also shows the collection statistics:

DOCUMENTS	TOTAL SIZE	AVG. SIZE	INDEXES	TOTAL SIZE	AVG. SIZE
1.4m	429.2MB	331B	1	12.9MB	12.9MB

The query execution options are:

- MAXTIMES: 5000
- SKIP: 0
- LIMIT: 0

The status bar indicates: Displaying documents 0 - 0 of N/A.

Interpretations:

There are no states which have had fires only in the second half of the calendar year.

Query 6

Question: Which forest had the number of fires equal to the average number of wild fires in the US?

Notes:

- Assumption 1: All fires in given table are from US.
- Assumption 2: Since there so many records, we will not exact match for average value.
- Assumption 3: For business, let's find forests with above average count.

of rows in result: 251

Translation: Group by forest and generate count. Group by null to generate average of count. Match by count greater than average.

Screen Shot of MongoDB Query/Code and Results

The screenshot displays the MongoDB Query and Results interface. It shows two aggregation stages and a final table view.

Stage 1: Aggregation

```
1 {
2   _id: '$SOURCE_REPORTING_UNIT_NAME',
3   Count: {
4     $sum: 1
5   }
6 }
```

Output after \$group stage (Sample of 20 documents):

```
{ "_id": "Cheyenne River Agency", "Count": 1802 }
{ "_id": "Fremont County", "Count": 58 }
```

Stage 2: Aggregation

```
1 {
2   _id: null,
3   average: {
4     $avg: '$Count'
5   }
6 }
```

Output after \$group stage (Sample of 1 document):

```
{ "_id": null, "average": 1122.5256298870547 }
```

Table View:

VIEW LIST TABLE

reporting_g.forest_count

	_id String	Count Int32
1	"Rosebud Agency"	3719
2	"Gila National Forest"	4332
3	"MSS Southwest District"	2760
4	"Phoenix District Office"	1386
5	"Black River Falls District"	1212
6	"Modoc National Forest"	1975
7	"Florida Forest Service"	2152
8	"Turtle Mountain Agency"	3119
9	"National Forests in Alabama"	1385
10	"Zuni Agency"	1309
11	"Central Land Office"	1439

Displaying documents 1 - 20 of 251

Interpretations:

The forests that have fires which are above the average number of fires occurring in the US are the ones listed above in the results. These forests have a higher danger of catching fire or have had higher occurrences of fires than the other forests in the US which means these forests are not as safe to visit. It is advisable to not pay visits to these and find alternative forests instead. It is also advisable to evacuate any animals that might be in danger or take precautions to provide them safe habitats in case these forests have high animal population.