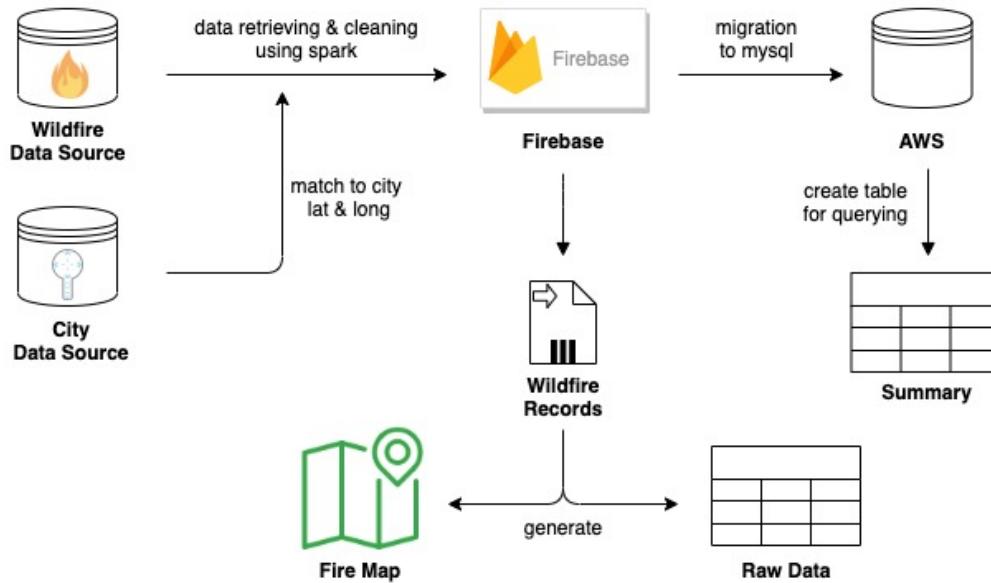


Wildfire Monitoring System

Project Description

2020's fire season has already seen above-normal activity. It has been a brutal summer for firefighters and thousands of residents in California. While weather conditions and air quality have continued to get worse, firefighters are monitoring the shape, size, rate of perimeter increase, and rate of area growth of a wind-driven wildland fire closely. This project is to better share maps and data related to wildland fire activities across California, helping utility companies predict wildfire risk and taking proactive measures to save lives and property.

System Design / Architecture



Data Sets

1) Wildland Fire Open Data

The data set describes the location, spreading, cause, etc. of current wildfire all over the United States. It is updated every 5 minutes.

2) Latitude and Longitude of All US Cities

The data set includes location information of all the cities in the US (e.g. city, country, state, latitude, longitude, population, timezone, zip code, etc.)

Data Problems

• Data Cleaning

The wildland fire open data provides data in json format. We request the json data from URL and import it to Spark to preprocess the data. In our project, we only focused on the wildfire instance in California, so we selected the instance whose POOState is ‘CA’. Raw data has tons of attributes. We only chose a small number of attributes that are meaningful and have less or none missing values to describe the wildfire incidents in our database.

<input checked="" type="checkbox"/> OBJECTID	<input checked="" type="checkbox"/> ABCDMisc	<input checked="" type="checkbox"/> ADSPermissionState	<input checked="" type="checkbox"/> CalculatedAcres	<input checked="" type="checkbox"/> ContainmentDateTime	<input checked="" type="checkbox"/> ControlDateTime	<input checked="" type="checkbox"/> CreatedOnDateTime	<input checked="" type="checkbox"/> DailyAcres	<input checked="" type="checkbox"/> DiscoveryAcres	<input checked="" type="checkbox"/> DispatchCenterID	<input checked="" type="checkbox"/> FinalFireReportApprovedByTitle	<input checked="" type="checkbox"/> FinalFireReportApprovedByUnit	<input checked="" type="checkbox"/> FinalFireReportApprovedDate	<input checked="" type="checkbox"/> FireBehaviorGeneral	<input checked="" type="checkbox"/> FireBehaviorGeneral1	<input checked="" type="checkbox"/> FireBehaviorGeneral2	<input checked="" type="checkbox"/> FireBehaviorGeneral3	<input checked="" type="checkbox"/> FireCause	<input checked="" type="checkbox"/> FireCauseGeneral	<input checked="" type="checkbox"/> FireCauseSpecific	<input checked="" type="checkbox"/> FireCode	<input checked="" type="checkbox"/> FireDepartmentID	<input checked="" type="checkbox"/> FireDiscoveryDateTime	<input checked="" type="checkbox"/> FireMgmtComplexity	<input checked="" type="checkbox"/> FireOutDateTime	<input checked="" type="checkbox"/> FireStrategyConfinePercent	<input checked="" type="checkbox"/> FireStrategyFullSuppPercent	<input checked="" type="checkbox"/> FireStrategyMonitorPercent	<input checked="" type="checkbox"/> FireStrategyPointZonePercent	<input checked="" type="checkbox"/> FSJobCode	<input checked="" type="checkbox"/> FSOVERRIDECode	<input checked="" type="checkbox"/> GACC	<input checked="" type="checkbox"/> ICS209ReportDateTime	<input checked="" type="checkbox"/> ICS209ReportForTimePeriodFrom	<input checked="" type="checkbox"/> ICS209ReportForTimePeriodTo	<input checked="" type="checkbox"/> ICS209ReportStatus	<input checked="" type="checkbox"/> IncidentManagementOrganization	<input checked="" type="checkbox"/> IncidentName	<input checked="" type="checkbox"/> IncidentShortDescription	<input checked="" type="checkbox"/> IncidentTypeCategory	<input checked="" type="checkbox"/> IncidentTypeKind	<input checked="" type="checkbox"/> InitialLatitude	<input checked="" type="checkbox"/> InitialLongitude	<input checked="" type="checkbox"/> InitialResponseAcres	<input checked="" type="checkbox"/> InitialResponseDateTime	<input checked="" type="checkbox"/> IrwinID	<input checked="" type="checkbox"/> IsFireCauseInvestigated	<input checked="" type="checkbox"/> IsFireCodeRequested	<input checked="" type="checkbox"/> IsFSAssisted	<input checked="" type="checkbox"/> IsMultiJurisdictional
<input checked="" type="checkbox"/> IsQuarantined	<input checked="" type="checkbox"/> IsReimbursable	<input checked="" type="checkbox"/> IsTrespass	<input checked="" type="checkbox"/> IsUnifiedCommand	<input checked="" type="checkbox"/> IsValid	<input checked="" type="checkbox"/> LocalIncidentIdentifier	<input checked="" type="checkbox"/> ModifiedOnDateTime	<input checked="" type="checkbox"/> PercentContained	<input checked="" type="checkbox"/> PercentPerimeterToBeContained	<input checked="" type="checkbox"/> POOCity	<input checked="" type="checkbox"/> POOCounty	<input checked="" type="checkbox"/> POODispatchCenterID	<input checked="" type="checkbox"/> POOFips	<input checked="" type="checkbox"/> POOJurisdictionalAgency	<input checked="" type="checkbox"/> POOJurisdictionalUnit	<input checked="" type="checkbox"/> POOJurisdictionalUnitParentUnit	<input checked="" type="checkbox"/> POOLandownerCategory	<input checked="" type="checkbox"/> POOLandownerKind	<input checked="" type="checkbox"/> POOLegalDescPrincipalMeridian	<input checked="" type="checkbox"/> POOLegalDescQtr	<input checked="" type="checkbox"/> POOLegalDescQtrQtr	<input checked="" type="checkbox"/> POOLegalDescRange	<input checked="" type="checkbox"/> POOLegalDescSection	<input checked="" type="checkbox"/> POOLegalDescTownship	<input checked="" type="checkbox"/> POOPredictiveServiceAreaID	<input checked="" type="checkbox"/> POOProtectingAgency	<input checked="" type="checkbox"/> POOProtectingUnit	<input checked="" type="checkbox"/> POOState	<input checked="" type="checkbox"/> PredominantFuelGroup	<input checked="" type="checkbox"/> PredominantFuelModel	<input checked="" type="checkbox"/> PrimaryFuelModel	<input checked="" type="checkbox"/> SecondaryFuelModel	<input checked="" type="checkbox"/> TotalIncidentPersonnel	<input checked="" type="checkbox"/> UniqueFireIdentifier	<input checked="" type="checkbox"/> WFDSSDecisionStatus															
<input checked="" type="checkbox"/> CreatedBySystem	<input checked="" type="checkbox"/> ModifiedBySystem	<input checked="" type="checkbox"/> CreatedOn	<input checked="" type="checkbox"/> ModifiedOn	<input checked="" type="checkbox"/> IsDispatchComplete	<input checked="" type="checkbox"/> OrganizationalAssessment	<input checked="" type="checkbox"/> StrategicDecisionPublishDate	<input checked="" type="checkbox"/> GlobalID	<input checked="" type="checkbox"/> ModifiedOnDateTime_dt	<input checked="" type="checkbox"/> CreatedOnDateTime_dt	The fields to be included in the returned result set.																																							

We chose below attributes:

- **UniqueFireIdentifier:** Unique identifier assigned to each wildland fire. yyyy = calendar year, SSUUUU = POO protecting unit identifier (5 or 6 characters), xxxxxx = local incident identifier (6 to 10 characters)
- **FireDiscoveryDateTime:** The date and time a fire was reported as discovered or confirmed to exist. May also be the start date for reporting purposes.
- **InitialLatitude:** The latitude location of the initial reported point of origin specified in decimal degrees.
- **InitialLongitude:** The longitude location of the initial reported point of origin specified in decimal degrees.
- **POOCity:** The closest city to the incident point of origin.
- **POOCountry:** The County Name identifying the county or equivalent entity at point of origin designated at the time of collection.
- **POOState:** The State alpha code identifying the state or equivalent entity at point of origin.
- **FireCause:** Broad classification of the reason the fire occurred identified as human, natural or unknown.
- **DiscoveryAcres:** An estimate of acres burning upon the discovery of the fire. More specifically when the fire is first reported by the first person that calls in the fire. The estimate should include the

number of acres within the current perimeter of a specific, individual incident, including unburned and unburnable islands.

- **DailyAcres:** A measure of acres reported for a fire. More specifically, the number of acres within the current perimeter of a specific, individual incident, including unburned and unburnable islands. The minimum size must be 0.1.
- **InitialResponseAcres:** An estimate of acres burning at the time of initial response. More specifically when the IC arrives and performs initial size up. The minimum size must be 0.1. The estimate should include the number of acres within the current perimeter of a specific, individual incident, including unburned and unburnable islands.
- **PercentContained:** Indicates the percent of incident areas that are no longer active. Reference definition in fire line handbook when developing standard.
- **ContainmentDateTime:** The date and time a wildfire was declared contained.
- **ControlDateTime:** The date and time a wildfire was declared under control.

• Data Integration

Simplemaps open data provides data in csv format. We only selected cities in California with features: country, state, latitude, longitude, and population. We used spark to integrate the json file from wildland fire open dataset and the csv file from city data set. For those incidents that have missing value on city attributes in wildland fire dataset, we use the city dataset to map incident to the closest city with longitude and latitude.

city	county_name	state_id	lat	lng	population
Los Angeles	Los Angeles	CA	34.1139	-118.4068	12750807
San Francisco	San Francisco	CA	37.7562	-122.443	3592294
San Diego	San Diego	CA	32.8312	-117.1225	3220118
Riverside	Riverside	CA	33.9381	-117.3948	2107852
Sacramento	Sacramento	CA	38.5667	-121.4683	1898019
San Jose	Santa Clara	CA	37.3019	-121.8486	1798103
Fresno	Fresno	CA	36.7831	-119.7941	703498
Concord	Contra Costa	CA	37.9722	-122.0016	652308
Mission Viejo	Orange	CA	33.6095	-117.655	590249
Bakersfield	Kern	CA	35.3529	-119.0359	579295
Murrieta	Riverside	CA	33.572	-117.1909	496046
Long Beach	Los Angeles	CA	33.7981	-118.1675	462628
Oakland	Alameda	CA	37.7903	-122.2165	433031

• Data Aggregation:

We aggregated the fire data by city to summarize the active fire conditions in California.

```

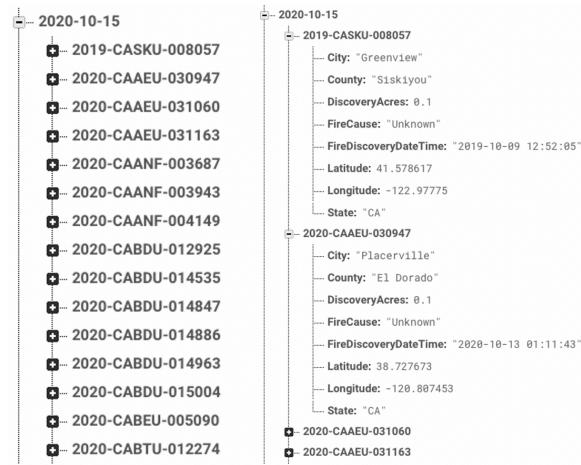
1  [[["Los Angeles", 33], ["Shasta", 5], ["El Dorado", 4], ["Riverside", 4], ["Tulare", 4], ["Sonoma", 4],
   ["Mendocino", 4],
2  ["Tuolumne", 3], ["Monterey", 3], ["Trinity", 3], ["Siskiyou", 3], ["Glenn", 3], ["Mariposa", 2], ["Fresno",
   2],
3  ["Plumas", 2], ["Yuba", 2], ["Lake", 2], ["Napa", 2], ["Humboldt", 2], ["San Diego", 2], ["Butte", 2], ["San
   Benito",
4  2], ["Kings", 1], ["Madera", 1], ["Placer", 1], ["Orange", 1], ["Santa Clara", 1], ["San Mateo", 1],
   ["Calaveras", 1],
5  ["San Bernardino", 1]]

```

- **Data Storage:**

1. Source->Firebase Storage for front end query.

As the wildfire data is updated to current wildfire instances status, we built a real-time database to keep track of the wildfire data everyday with integration of city data. It would be stored under the node of the current date and each wildfire instance would be stored under the node of its uniqueID, so that users could query the firebase database by a specific date or attributes from the front end.



2. Firebase Storage->AWS for archive and summary table query.

We flattened the nested data to relational data structure and then uploaded to AWS RDS for aggregation.

aws-rds-wildfire

Administration Schemas [create_table*](#) [SQL File 1*](#) [SQL File 4*](#)

Limit to 1000 rows

1 `select * from instances;`

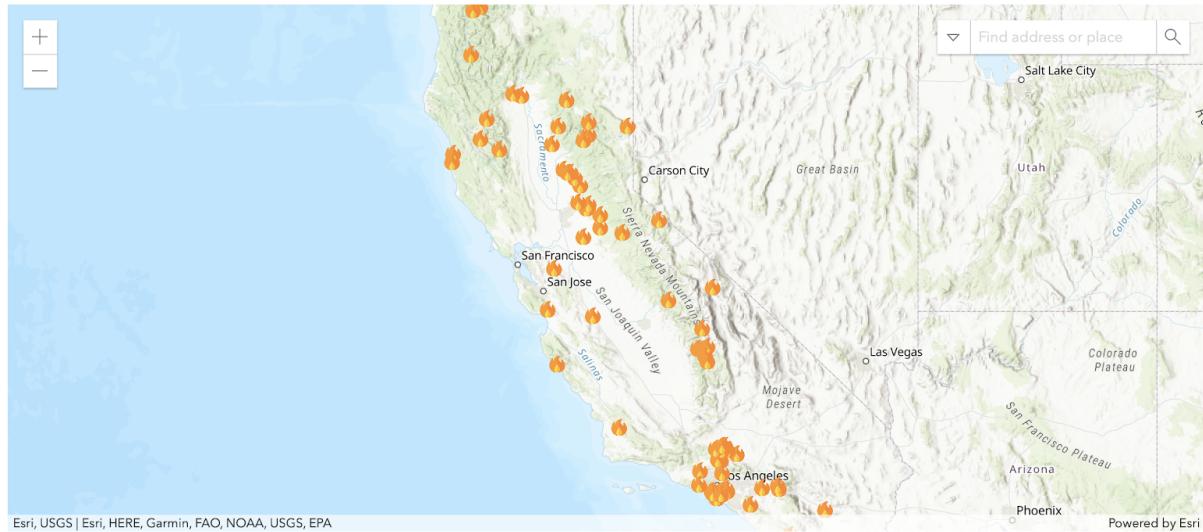
Instance_ID	Incident_Name	City	County	State	record_Date	Latitude	Longitude	Discovery_Acres	Daily_Acres	Initial_Response_Acres	Fire_Cause	Fire_Discovery_DateTime	Per
2019-CASKU-008057	Unknown	Greenview	Siskiyou	CA	2020-10-15	41.578617	-122.97778	0.1	.99	.99	Unknown	2019-10-09 12:52:05	.99
2020-CAAEU-030947	Unknown	El Dorado	CA	2020-10-15	38.727673	-120.807453	0.1	.99	.99	Unknown	2020-10-13 03:14:00	.99	
2020-CAAEU-031060	Unknown	Folsom	Sacramento	CA	2020-10-15	38.495754	-121.1466	0.1	.99	.99	Unknown	2020-10-14 03:31:00	.99
2020-CAAEU-031163	Unknown	Ione	Amador	CA	2020-10-15	30.506396	-120.944227	0.1	.99	.99	Unknown	2020-10-15 01:43:18	.99
2020-CAANF-030867	Unknown	Altadena	Los Angeles	CA	2020-10-15	34.26779	-118.1412	1	115796	.99	Unknown	2020-09-06 12:21:00	.99
2020-CAANF-030943	Unknown	Green Valley	Los Angeles	CA	2020-10-15	34.56246	-118.3999	230	230	.99	Unknown	2020-09-28 14:51:00	10x
2020-CAANF-030416	Unknown	Palmdale	Los Angeles	CA	2020-10-15	34.562465	-118.0735	0.5	0.5	.99	Unknown	2020-10-05 10:46:00	.99
2020-CAANF-030429	Unknown	Oak Glen	San Bernardino	CA	2020-10-15	34.652719	-117.05209	1	22744	.99	Human	2020-09-25 10:42:00	.95
2020-CABDU-014535	Unknown	Lake Arrowhead	San Bernardino	CA	2020-10-15	34.34375	-117.221783	0.1	.99	.99	Unknown	2020-10-06 14:20:48	.99
2020-CABDU-014847	Unknown	Mentone	San Bernardino	CA	2020-10-15	34.073445	-117.097923	0.1	.99	.99	Unknown	2020-10-12 16:16:17	.99
2020-CABDU-014893	Unknown	Wrightwood	San Bernardino	CA	2020-10-15	34.427923	-117.642477	0.1	.99	.99	Unknown	2020-10-13 12:38:33	.99
► 2020-CABDU-014863	Unknown	Mentone	San Bernardino	CA	2020-10-15	34.013581	-117.137756	0.1	.99	.99	Unknown	2020-10-14 20:51:54	.99
2020-CABDU-015004	Unknown	Oak Hills	San Bernardino	CA	2020-10-15	34.428712	-117.434426	0.1	.99	.99	Unknown	2020-10-15 19:43:37	.99
2020-CAEBU-005090	Unknown	San Miguel	San Luis Obispo	CA	2020-10-15	36.183316	-120.699783	0.1	.99	.99	Unknown	2020-10-14 17:43:43	.99
2020-CABTU-012274	Unknown	Paradise	Butte	CA	2020-10-15	39.705164	-121.576222	0.1	.99	.99	Unknown	2020-10-14 09:18:34	.99
2020-CABTU-012308	Unknown	Oroville	Butte	CA	2020-10-15	39.519851	-121.571591	0.1	.99	.99	Unknown	2020-10-14 12:46:14	.99
2020-CACDD-015245	Unknown	Potrero	San Diego	CA	2020-10-15	32.57394	-116.760465	1	1	.99	Unknown	2020-10-14 07:21:19	.99
2020-CACDD-015259	Unknown	Highland	San Bernardino	CA	2020-10-15	34.089116	-117.1851	1	0.1	.99	Unknown	2020-10-14 13:44:52	.99
2020-CACNF-002509	Unknown	Coronita	Riverside	CA	2020-10-15	33.83546	-117.6094	1	1	.99	Unknown	2020-08-13 16:40:00	.99
2020-CAFU-006758	Unknown	Lompoc	Santa Cruz	CA	2020-10-15	37.183509	-120.068655	0.1	.99	.99	Unknown	2020-10-15 03:22:00	.99
2020-CAFU-006760	Unknown	Kingsgate	Nevada	CA	2020-10-15	38.000001	-120.068655	0.1	1670	.99	Unknown	2020-10-15 03:22:00	.95
2020-CAFUH-006772	Unknown	Kidman	Modoc	CA	2020-10-15	40.672408	-120.202211	0.1	.99	.99	Unknown	2020-10-14 09:23:04	.99
2020-CAFUH-015372	Unknown	Squaw Valley	Placer	CA	2020-10-15	36.738085	-119.083441	0.1	.99	.99	Unknown	2020-10-13 09:21:46	.99
2020-CAHUJ-008729	Unknown	Alderpoint	Humboldt	CA	2020-10-15	40.14848	-123.827078	0.1	.99	.99	Unknown	2020-10-13 18:23:01	.99
2020-CAHUJ-008787	Unknown	McKinleyville	Humboldt	CA	2020-10-15	40.161083	-124.069075	0.1	.99	.99	Unknown	2020-10-15 18:28:05	.99
2020-CAKNF-006332	Unknown	Coffee Creek	Trinity	CA	2020-10-15	41.137448	-122.974418	0.1	0.3	.99	Unknown	2020-08-16 06:54:33	.99
2020-CAKNF-007000	Unknown	Trinity Village	Trinity	CA	2020-10-15	41.249283	-123.3241	0.1	6.1	.99	Human	2020-09-27 22:30:00	.99
► 2020-CAKNF-011793	Unknown	Hannan Ranch	Siskiyou	CA	2020-10-15	41.856026	-123.363564	0.1	116616	.99	Unknown	2020-09-28 04:08:41	.99

Object Info Session

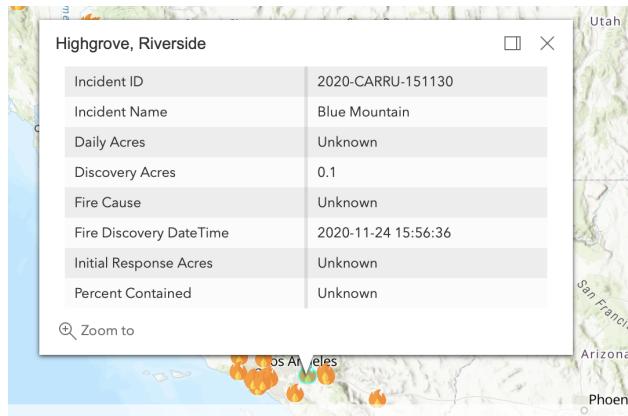
Schema: **wildfire**

- **Data Visualization:**

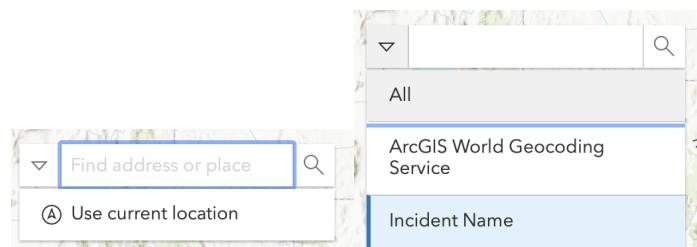
1. Create a feature layer in the map view and feed the fire data into the layer source.



2. Add a popup to show information and a search module. The detailed information would pop up when the user clicks on the wildfire icon.



3. Users could search a location to find nearby wildfires or search an incident name to find its location.



4. Raw table allows users to explore clean data retrieved from the source.

Raw Data										
#	IncidentName	IncidentID	DailyAcres	DiscoveryAcres	FireCause	FireDiscoveryDateTime	InitialResponseAcres	City	County	
1	Luneman	2020-CAAEU-035570	Unknown	0.1	Unknown	2020-11-26 12:23:38	Unknown	Shingle Springs	El Dorado	
2	View Terrace St / Pioneer	2020-CAAEU-035588	Unknown	0.1	Unknown	2020-11-26 15:09:24	Unknown	Mokelumne Hill	Amador	
3	American	2020-CAAEU-035665	Unknown	0.1	Unknown	2020-11-27 13:33:25	Unknown	Fiddletown	Amador	
4	Alameda	2020-CAAEU-035764	Unknown	0.1	Unknown	2020-11-28 11:19:04	Unknown	Camino	El Dorado	
5	Bobcat	2020-CAANF-003687	115997	1	Unknown	2020-09-06 12:21:00	Unknown	Azusa	Los Angeles	

Raw Data from [The National Interagency Fire Center \(NIFC\)](#)

5. Summary table shows the aggregation result using ‘group by city’.

Summarize

#	County	number of fire
1	Los Angeles	45
2	Humboldt	29
3	Siskiyou	11
4	Riverside	6
5	Tulare	5
6	Shasta	5
7	Monterey	3
8	Plumas	2
9	Tehama	2
10	San Luis Obispo	2
11	El Dorado	2
12	Nevada	2

Unique Features

1. Intuitive map visualization and flexible searching of everyday fire data in California.
2. User friendly front end experience and intuitive UI design gives users full control on data searching and exploration.

Implementation

Data Resource:

Our raw data are from 2 data resources. One is in json file format, the other in csv file format. We used Spark RDD to build a pipeline to preprocess and integrate two files.

Database:

We used two databases to store our data. One is a relational database AWS RDS and the other is Firebase Realtime Database. We maintained these two databases everyday to keep updated to the latest wildfire status.

Data visualization:

We used ArcGIS (Geographic Information System) web API to draw our map. It shows the most current active wildfires in California. It has a popup table that shows up with the related attributes of chosen fire, a search box that allows users to search for individual fire and locate it with its name.

Searching and Exploring:

We created a table that presents all the raw data retrieved from the firebase. A date input and a search box allows users to explore raw data. These ‘on change’ functions will update the map and the table on a chosen day or with a certain value in real time. Users could search a location to find nearby wildfires or search an incident name to find its location.

Aggregation:

We set up a local server to query the data from RDS, and filter the fires whose Percent Contained are smaller than 100 percent, which means fires are not completely controlled. At the same time we query ‘group by city’ data from SQL by date. In this way we can show how many fires are in one city. We use Network Address Translator to allow users to access our back-end. Due to GitHub page security technical restrictions, we are not able to connect our back end, resulting the failure of updating the summary table in real-time.

Links

Github Repo: [link](#)

Github Page: [link](#)

Presentation: [link](#)

Team Members & Tasks

Jiaying Wang (Current: Applied Data Science, Undergrad: Biotechnology & Statistics):

- LinkedIn: <https://www.linkedin.com/in/jiaying-wang-161694176>
- Skills: Data Visualization, Data Analysis, Machine Learning, NLP, Spark
- Work Accomplished:
 1. Built up a pipeline with Spark to preprocess and integrate two different data structures and uploaded to the firebase database.
 2. Maintained the firebase database, synchronized it to AWS RDS and updated them everyday.
 3. Implemented the search function on our HTML web page to navigate raw data.

Yanan Lin (Current: Computer Science, Undergrad: Software Engineering):

- LinkedIn: <https://linkedin.com/in/yanan-lin>
- Skills: Full Stack, Computer Graphics
- Work Accomplished:
 1. Engineered HTML web page and basic functionality, designed front end UI, and published to GitHub.
 2. Created GIS map module to receive and visualize wildfire geographic data.
 3. Generated table to present raw data.

Yao Xiao (Current: Applied Data Science, Undergrad: Theoretical and Applied Mechanics):

- LinkedIn: <https://www.linkedin.com/in/yao-xiao-a63004197/>
- Skills: Deep Learning, Web Crawler, Data Mining
- Work Accomplished:
 1. Set up the back end and connect to RDS to complete the aggregation part.
 2. Generated table to present the summary data.