

# San Francisco Fire Calls Case Study

## Case Scenario

### Background

#### **San Francisco Fire Department (SFFD)**

The San Francisco Fire Department (SFFD) is the agency responsible for providing fire protection and emergency medical services to the city of San Francisco, California. Here are some key points about the San Francisco Fire Department:

#### **History:**

The SFFD has a rich history dating back to the mid-19th century. It was officially established in 1866, making it one of the oldest fire departments on the West Coast.

#### **Services:**

The primary responsibilities of the SFFD include **firefighting, emergency medical services**, hazardous materials response, and disaster response. The department is also involved in fire prevention activities, public education, and community outreach.

#### **Fire Stations:**

The SFFD operates multiple fire stations strategically located throughout the city to ensure prompt response to emergencies. Each station is equipped with firefighting and emergency medical equipment.

#### **Personnel:**

The department is staffed by highly trained firefighters, paramedics, emergency medical technicians (EMTs), and support personnel. These individuals undergo rigorous training to handle a wide range of emergencies.

#### **Specialized Units:**

In addition to traditional firefighting services, the SFFD has specialized units for tasks such as water rescue, high-angle rescue, and urban search and rescue. These units are trained to respond to various types of emergencies.

### **Community Involvement:**

The SFFD is actively involved in community outreach and education programs. These initiatives aim to raise awareness about fire safety, disaster preparedness, and other important topics.

### **Emergency Medical Services (EMS):**

The SFFD provides emergency medical services, responding to medical emergencies and transporting patients to hospitals. Many firefighters are trained as paramedics or EMTs to provide advanced life support.

### **Mutual Aid:**

The SFFD collaborates with neighboring fire departments and agencies through mutual aid agreements. This allows for a coordinated response to large-scale emergencies or disasters.

**Technology and Equipment:** The department utilizes modern technology and equipment to enhance its firefighting and rescue capabilities. This includes advanced firefighting apparatus, communication systems, and medical equipment.

### **Fire Prevention:**

The SFFD is actively involved in fire prevention efforts, conducting inspections, enforcing fire codes, and educating the public on ways to reduce fire risks.

Source :- <https://sf-fire.org/>

<https://data.sfgov.org/Public-Safety/Fire-Department-Calls-for-Service/nuek-vuh3>

## Challenges

- San Francisco Fire department (SFFD) SFFD maintains large datasets for different services offered by them. Few datasets are having 60+ columns with millions of records.
- **Fire Calls-For-Service** datasets with 28 columns, includes all fire units responses to calls.
  - Each record includes the **call number**, incident number, address, unit identifier, call type, and disposition.
  - All relevant time intervals are also included. Because this dataset is based on responses, and since most calls involved multiple units, there are multiple records for each call number.
  - Addresses are associated with a block number, intersection or call box, not a specific address.
- **Fire Incidents** data sets with 60+ columns, includes a summary of each (**non-medical**) incident to which the SF Fire Department responded.
  - Each incident record includes the call number, incident number, address, number and type of each unit responding, call type (as determined by dispatch), prime situation (field observation), actions taken, and property loss.
- There is a fire station in San Francisco which responds to the fire alarm generated at different locations at different times. All these are being recorded with the actual priority and the type of fire alarm which has been raised. The fire station then responds to those alarm updating the requirement of ALS (ADVANCED LIFE SUPPORT).
- In raw datasets, few attributes having NULL values for hundreds of records.

- SFFD is facing challenges to create report and visualize it by combining the **Fire calls** and **Fire incidents** datasets using the raw datasets available in their collected data sources.

## **Business Need**

- The requirement is to **stage, process** and **store** the data of all SFFD Fire dataset irrespective of their sources and formats and transform it to the analytical needs.
- Finally build **reports/visualizations** which provide actionable insights.
- Most of this is needed in real time. There is an increase in demand for stream processing data as well due to increasing Fire alarm.
- The requirement is to ingest and process the data at very high frequencies.
- Finally, none of the data should be discarded. In fact, it should be preserved for enabling use-cases related to regulatory compliance, public safety, insurance, targeted ads/promotions/offers etc.,
- SFFD wants Accenture to build a common platform to store all data related to Fire calls and Fire incidents in order to analyze the data for better public services insights like Fire Prevention, Community Involvement, Emergency Medical Services (EMS), Mutual Aid.

## **Proposed Solution**

- Considering the Volume, Velocity and Veracity aspects of the data, **Accenture** has decided to build a Data Lake using Databricks Lake House Platform which is going to be a single store for raw intermittent and processed data.
- Accenture has analyzed various available options and delivered some PoC's. In a one-year contract with SFFD, Accenture will build a Spark based Data Lake augmented by a Cloud based Data Lake on Azure.
- The solution will implement the latest features and concepts of Databricks Lake House Platform

- Fire department decided to analysis the fire calls data which help them to understand how effectively the SF fire station performing on service duty.

## Attributes and Sample Data

**sf-fire-calls** datasets - Total 28 columns ( sf-fire-calls.csv )

Call Number,Unit ID,Incident Number,CallType,Call Date,Watch Date,Call Final Disposition,Available DtTm,Address,City,Zipcode of Incident,Battalion,Station Area,Box,OrigPriority,Priority,Final Priority,ALS Unit,Call Type Group,NumAlarms,UnitType,Unit sequence in call dispatch,Fire Prevention District,Supervisor District,Neighborhood,Location,RowID,Delay

Column Name	Data Type
Call Number	integer
Unit ID	string
Incident Number	integer
CallType	string
Call Date	date
Watch Date	date
Call Final Disposition	string
Available DtTm	string
Address	string
City	string
Zipcode of Incident	integer
Battalion	string
Station Area	string
Box	string
OrigPriority	string
Priority	string
Final Priority	integer
ALS Unit	boolean
Call Type Group	string
NumAlarms	integer
UnitType	string
Unit sequence in call dispatch	integer
Fire Prevention District	string
Supervisor District	string
Neighborhood	string
Location	string
RowID	string
Delay	double

**sf-fire-incidents datasets - Total 63 columns ( sf-fire-incidents.csv )**

Column Name	Data Type
-------------	-----------

Incident Number	integer
Exposure Number	integer
Address	string
Incident Date	date
Call Number	integer
Alarm DtTm	string
Arrival DtTm	string
Close DtTm	string
City	string
Zipcode	integer
Battalion	string
Station Area	integer
Box	string
Suppression Units	integer
Suppression Personnel	integer
EMS Units	integer
EMS Personnel	integer
Other Units	integer
Other Personnel	integer
First Unit On Scene	string
Estimated Property Loss	integer
Estimated Contents Loss	integer
Fire Fatalities	integer
Fire Injuries	integer
Civilian Fatalities	integer
Civilian Injuries	integer
Number of Alarms	string
Primary Situation	string
Mutual Aid	string
Action Taken Primary	string
Action Taken Secondary	string
Action Taken Other	string
Detector Alerted Occupants	string
Property Use	string
Area of Fire Origin	string
Ignition Cause	string
Ignition Factor Primary	string
Ignition Factor Secondary	string
Heat Source	string
Item First Ignited	string
Human Factors Associated with Ignition	string
Structure Type	string
Structure Status	string
Floor of Fire Origin	integer
Fire Spread	string
No Flame Spread	string
Number of floors with minimum damage	integer
Number of floors with significant damage	integer
Number of floors with heavy damage	integer
Number of floors with extreme damage	string
Detectors Present	string
Detector Type	string
Detector Operation	string
Detector Effectiveness	string
Detector Failure Reason	string
Automatic Extinguishing System Present	string
Automatic Extinguishing System Type	string
Automatic Extinguishing System Performance	string
Automatic Extinguishing System Failure Reason	string
Number of Sprinkler Heads Operating	integer
Supervisor District	integer
Neighborhood District	string
Location	string

## Relation between the datasets

**sf-fire-calls** datasets ( sf-fire-calls.csv )

Column Name
Call Number
Unit ID
Incident Number
CallType
Call Date
Watch Date
Call Final Disposition
Available DtTm
Address
City
Zipcode of Incident
Battalion
Station Area
Box
OrigPriority
Priority
Final Priority
ALS Unit
Call Type Group
NumAlarms
UnitType
Unit sequence in call dispatch
Fire Prevention District
Supervisor District
Neighborhood
Location
RowID
Delay

**sf-fire-incidents** datasets ( sf-fire-incidents.csv )

Column Name	Data Type
Incident Number	integer
Exposure Number	integer
Address	string
Incident Date	date
Call Number	integer
Alarm DtTm	string
Arrival DtTm	string
Close DtTm	string
City	string
Zipcode	integer
Battalion	string
Station Area	integer
Box	string
Suppression Units	integer
Suppression Personnel	integer
EMS Units	integer
EMS Personnel	integer
Other Units	integer
Other Personnel	integer
First Unit On Scene	string
Estimated Property Loss	integer
Estimated Contents Loss	integer
Fire Fatalities	integer
Fire Injuries	integer
Civilian Fatalities	integer
Civilian Injuries	integer
Number of Alarms	string
Primary Situation	string
Mutual Aid	string
Action Taken Primary	string
Action Taken Secondary	string
Action Taken Other	string
Detector Alerted Occupants	string
Property Use	string
Area of Fire Origin	string
Ignition Cause	string
Ignition Factor Primary	string
Ignition Factor Secondary	string
Heat Source	string
Item First Ignited	string
Human Factors Associated with Ignition	string
Structure Type	string
Structure Status	string
Floor of Fire Origin	integer
Fire Spread	string
No Flame Spread	string
Number of floors with minimum damage	integer
Number of floors with significant damage	integer
Number of floors with heavy damage	integer
Number of floors with extreme damage	string

## Location of Source Data

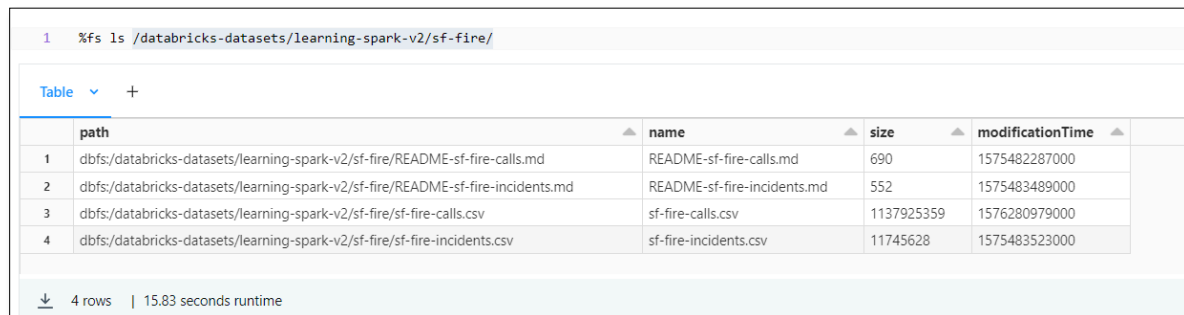
Location in Databricks repository -



**dbfs:/databricks-datasets/learning-spark-v2/sf-fire**

**dbfs:/databricks-datasets/learning-spark-v2/sf-fire/sf-fire-calls.csv**

**dbfs:/databricks-datasets/learning-spark-v2/sf-fire/sf-fire-incidents.csv**



	path	name	size	modificationTime
1	dbfs:/databricks-datasets/learning-spark-v2/sf-fire/README-sf-fire-calls.md	README-sf-fire-calls.md	690	1575482287000
2	dbfs:/databricks-datasets/learning-spark-v2/sf-fire/README-sf-fire-incidents.md	README-sf-fire-incidents.md	552	1575483489000
3	dbfs:/databricks-datasets/learning-spark-v2/sf-fire/sf-fire-calls.csv	sf-fire-calls.csv	1137925359	1576280979000
4	dbfs:/databricks-datasets/learning-spark-v2/sf-fire/sf-fire-incidents.csv	sf-fire-incidents.csv	11745628	1575483523000

**Location convention of uploaded data in Databricks File System (DBFS) -**

**(For batch data processing)**

**dbfs:/FileStore/tables/<CUSTOMIZED LOCATION>**

**Landing location of Source Data in DBFS**

dbfs:/FileStore/tables/sfdatasets/landing/sffirecalls

dbfs:/FileStore/tables/sfdatasets/landing/sffireincidents

**Stage1 Location of Data after ETL/Cleaning Operations from Landing Location:**

dbfs:/FileStore/tables/sfdatasets/stage1\_spark/sffirecalls

dbfs:/FileStore/tables/sfdatasets/stage1\_spark/sffireincidents

### **Merged Data location (Stage2)**

dbfs:/FileStore/tables/sfdatasets/stage2\_spark/

### **Processed/Output Files locations (Stage3)**

dbfs:/FileStore/tables/sfdatasets/stage3\_spark/

### **FILE LOCATION FOR STRUCTURED STREAMING**

dbfs:/FileStore/tables/sfdatasets/streaming/sffirecalls

### **COMMANDS TO WORK WITH FILES IN DBFS**

<https://docs.databricks.com/dbfs/index.html>

<https://docs.databricks.com/files/index.html>

### **Data Processing using Relational Entities in Databricks**

<b>sf-fire-calls.csv</b>	
<b>Column Name</b>	<b>Column Index</b>
Call Number	Attribute-0

Unit ID	Attribute-1
Incident Number	Attribute-2
CallType	Attribute-3
Call Date	Attribute-4
Watch Date	Attribute-5
Call Final Disposition	Attribute-6
Available DtTm	Attribute-7
Address	Attribute-8
City	Attribute-9
Zipcode of Incident	Attribute-10
Battalion	Attribute-11
Station Area	Attribute-12
Box	Attribute-13
OrigPriority	Attribute-14
Priority	Attribute-15
Final Priority	Attribute-16
ALS Unit	Attribute-17
Call Type Group	Attribute-18
NumAlarms	Attribute-19
UnitType	Attribute-20
Unit sequence in call dispatch	Attribute-21
Fire Prevention District	Attribute-22
Supervisor District	Attribute-23
Neighborhood	Attribute-24
Location	Attribute-25
RowID	Attribute-26
Delay	Attribute-27

**sf-fire-incidents.csv**

Column Name	Column Index		Column Name	Column Index
Incident Number	Attribute-0		Ignition Factor Primary	Attribute-36
Exposure Number	Attribute-1		Ignition Factor Secondary	Attribute-37
Address	Attribute-2		Heat Source	Attribute-38
Incident Date	Attribute-3		Item First Ignited	Attribute-39
Call Number	Attribute-4		Human Factors Associated with Ignition	Attribute-40
Alarm DtTm	Attribute-5		Structure Type	Attribute-41
Arrival DtTm	Attribute-6		Structure Status	Attribute-42
Close DtTm	Attribute-7		Floor of Fire Origin	Attribute-43
City	Attribute-8		Fire Spread	Attribute-44
Zipcode	Attribute-9		No Flame Spead	Attribute-45
Battalion	Attribute-10		Number of floors with minimum damage	Attribute-46
Station Area	Attribute-11		Number of floors with significant damage	Attribute-47
Box	Attribute-12		Number of floors with heavy damage	Attribute-48
Suppression Units	Attribute-13		Number of floors with extreme damage	Attribute-49
Suppression Personnel	Attribute-14		Detectors Present	Attribute-50
EMS Units	Attribute-15		Detector Type	Attribute-51
EMS Personnel	Attribute-16		Detector Operation	Attribute-52
Other Units	Attribute-17		Detector Effectiveness	Attribute-53
Other Personnel	Attribute-18		Detector Failure Reason	Attribute-54
First Unit On Scene	Attribute-19		Automatic Extinguishing System Present	Attribute-55
Estimated Property Loss	Attribute-20		Automatic Extinguishing Sytem Type	Attribute-56
Estimated Contents Loss	Attribute-21		Automatic Extinguishing Sytem Perfomance	Attribute-57
Fire Fatalities	Attribute-22		Automatic Extinguishing Sytem Failure Reason	Attribute-58
Fire Injuries	Attribute-23		Number of Sprinkler Heads Operating	Attribute-59
Civilian Fatalities	Attribute-24		Supervisor District	Attribute-60
Civilian Injuries	Attribute-25		Neighborhood District	Attribute-61
Number of Alarms	Attribute-26		Location	Attribute-62
Primary Situation	Attribute-27			
Mutual Aid	Attribute-28			
Action Taken Primary	Attribute-29			
Action Taken Secondary	Attribute-30			
Action Taken Other	Attribute-31			
Detector Alerted Occupants	Attribute-32			
Property Use	Attribute-33			
Area of Fire Origin	Attribute-34			
Ignition Cause	Attribute-35			

## Task:1 Solve using Spark SQL API

Use Spark SQL API to produce outputs as per below problem statements:-

1. Explore Fire Call dataset.
2. Explore Fire incident dataset.
3. Create databricks data engineer workspace with Cluster.

### Data Processing using Delta tables.

1. **Check the source files** using dbutils or %fs commands
2. **Create Separate Single Notebook for All SQL Operations**
3. Create Temporary view from source file –

dbfs:/databricks-datasets/learning-spark-v2/sf-fire/sf-fire-calls.csv using the schema given below.

CallNumber, Integer,

UnitID, String,

IncidentNumber, Integer,

CallType, String,

CallDate, String,

WatchDate, String,

CallFinalDisposition, String,

AvailableDtTm, String,

Address, String,

City, String,

Zipcode, Integer,

Battalion, String,

StationArea, String,

Box, StringType,

OriginalPriority, String,

Priority, String,  
FinalPriority, Integer,  
ALSUnit, Boolean,  
CallTypeGroup, String,  
NumAlarms, Integer,  
UnitType, String,  
UnitSequenceInCallDispatch', Integer,  
FirePreventionDistrict, String,  
SupervisorDistrict, String,  
Neighborhood, String,  
Location, String,  
RowID, String,  
Delay, Float

4. Check the schema.
5. Count the number of Records.
6. Create a delta table from the previous temporary view.
7. Create Temporary view from source file –

dbfs:/databricks-datasets/learning-spark-v2/sf-fire/sf-fire-incidents.csv using the schema given below.

IncidentNumber, Integer,  
ExposureNumber, Integer,  
Address, String,  
IncidentDate, Date,  
CallNumber, Integer,  
AlarmDtTm, String,

ArrivalDtTm, String,  
CloseDtTm, String,  
City, String,  
Zipcode, Integer,  
Battalion, String,  
StationArea, Integer,  
Box, String,  
SuppressionUnits, Integer,  
SuppressionPersonnel, Integer,  
EMSUnits, Integer,  
EMSPersonnel, Integer,  
OtherUnits, Integer,  
OtherPersonnel, Integer,  
FirstUnitOnScene, String,  
EstimatedPropertyLoss, Integer,  
EstimatedContentsLoss, Integer,  
FireFatalities, Integer,  
FireInjuries, Integer,  
CivilianFatalities, Integer,  
CivilianInjuries, Integer,  
NumberofAlarms, String,  
PrimarySituation, String,  
MutualAid, String,  
ActionTakenPrimary, String,  
ActionTakenSecondary, String,  
ActionTakenOther, String,

DetectorAlertedOccupants, String,  
PropertyUse, String,  
AreaofFireOrigin, String,  
IgnitionCause, String,  
IgnitionFactorPrimary, String,  
IgnitionFactorSecondary, String,  
HeatSource, String,  
ItemFirstIgnited, String,  
HumanFactorsAssociatedwithIgnition, String,  
StructureType, String,  
StructureStatus, String,  
FloorofFireOrigin, Integer,  
FireSpread, String,  
NoFlameSpead, String,  
NumberOfFloorsWithMinimumDamage, Integer,  
NumberOfFloorsWithSignificantDamage, Integer,  
NumberOfFloorsWithHeavyDamage, Integer,  
NumberOfFloorsWithExtremeDamage, String,  
DetectorsPresent, String,  
DetectorType, String,  
DetectorOperation, String,  
DetectorEffectiveness, String,  
DetectorFailureReason, String,  
AutomaticExtinguishingSystemPresent, String,  
AutomaticExtinguishingSytemType, String,  
AutomaticExtinguishingSytemPerfomance, String,



AutomaticExtinguishingSystemFailureReason, String,  
umberofSprinklerHeadsOperating, Integer,  
SupervisorDistrict, String,  
NeighborhoodDistrict, String,  
Location, String

8. Check the schema.
9. Count the number of Records.
10. Create a delta table from previous temporary view.
11. Create a delta table using file\_read

**Use Spark SQL API on created delta tables to produce outputs as per below problem statements: -**

**Note: - Create SQL notebook for all the solutions**

1. Calculate the average delay of fire calls for each call type?
2. Display times/months where they receive more or less calls?
3. What are the top 10 incident types that the San Francisco Fire Department responds to?
4. How many medical incidents and structure fires are there in each neighborhood?
5. What's the average response time for each type fire calls?
6. Which category of fire incidents have historically been the most common in San Francisco?
7. Which month of the year 2002 reported the most number of calls?
8. How many calls were made whose original priority and final priority were same ?
9. How many calls had a delay more than 4.0?
10. What is the average delay in priority 1?
11. How many calls were marked in the neighborhood of bernal heights?

12. How many medical incidents were reported in month of November?
13. How many fires alarm were generated with unit id 91 in month of November?
14. What is the total of alarms generated in March?
15. What is number of alarms reported in fire station with zip code 94104?

Above queries are just indicative ones, obtain interesting insights for the given datasets

## Task2: Data Processing using Databricks Auto Loader and Structured Streaming

### FILE LOCATION FOR STRUCTURED STREAMING

```
dbfs:/FileStore/tables/sfdatasets/streaming/sffirecalls
```

```
input_path = "dbfs:/FileStore/tables/sfdatasets/streaming/sffirecalls/landing"
```













```
checkpoint_location = "dbfs:/FileStore/tables/sfdatasets/streaming/sffirecalls/landing/checkpoint"
```

```
output_path = "dbfs:/FileStore/tables/sfdatasets /streaming/sffirecalls/landing/output"
```

### SF Fire Calls file list –

sf-fire-calls\_0.csv, sf-fire-calls\_1.csv, sf-fire-calls\_2.csv, sf-fire-calls\_3.csv, sf-fire-calls\_4.csv

sf-fire-calls\_5.csv

 sf-fire-calls_0		12/14/2023 6:11 PM	Microsoft Excel Co...	11,793 KB
 sf-fire-calls_1		12/14/2023 6:11 PM	Microsoft Excel Co...	11,790 KB
 sf-fire-calls_2		12/14/2023 6:12 PM	Microsoft Excel Co...	11,792 KB
 sf-fire-calls_3		12/14/2023 6:12 PM	Microsoft Excel Co...	11,790 KB
 sf-fire-calls_4		12/14/2023 6:13 PM	Microsoft Excel Co...	11,788 KB
 sf-fire-calls_5		12/14/2023 6:13 PM	Microsoft Excel Co...	11,787 KB

**Note :- Create separate notebook for Auto Loader solution**

1. **Upload the source files** (csv files shared separately with this case study document) –

**As the respective landing locations as:**

`dbfs:/FileStore/tables/sfdatasets/landing/sffirecalls`

`dbfs:/FileStore/tables/sfdatasets/landing/sffireincidents`

2. **Create Single Notebook for All Streaming and Auto Loader Operations**

**3. Stream the data and implement:**

- a. Update report with incremental count of Fire Calls on Different Call Type from firecalls datasets.**
- b. get the updated report with incremental Average Delay per call type per City from firecalls datasets.**
- c. Incrementally get 5 more insights each to process the data incrementally**

### Task3: Use Spark SQL API to produce outputs as per below problem statements: -

1. **Upload the source files** (csv files shared separately with this case study document) –

**As the respective landing locations as:**

`dbfs:/FileStore/tables/sfdatasets/landing/sffirecalls`

`dbfs:/FileStore/tables/sfdatasets/landing/sffireincidents`

2. **Create a Separate Notebook for each DLT Pipeline Code.**

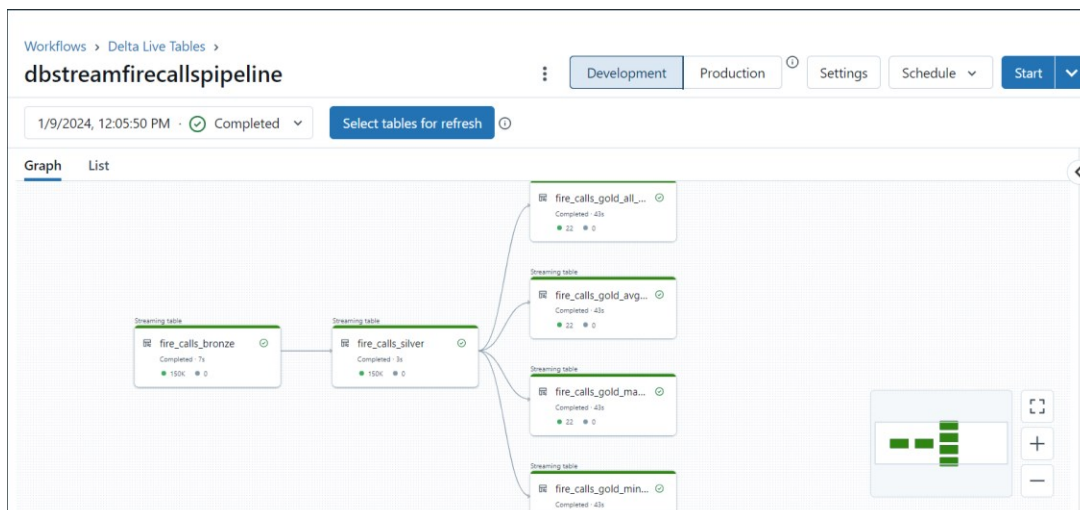
3. Create a (Raw/Bronze) delta live table from source file/s –

`dbfs:/FileStore/tables/sfdatasets/landing/sffirecalls`

4. Create another (Raw/Bronze) delta live table from source file –

`dbfs:/FileStore/tables/sfdatasets/landing/sffireincidents`

5. Create a pipeline similar to the one shown below



6. Update report with incremental count of Fire Calls on Different Call Type from firecalls datasets.
7. Build a DLT Pipeline to Find out the Zipcode and City where Total fire injuries are in positive number (greater than 0).
8. **Create a similar DLT pipeline and 4-5 gold tables with aggregates**
9. Create a job and execute it which includes:
  - a. Loading of dataset
  - b. DLT pipeline
  - c. Output execution
10. Use the silver tables to implement a dashboard with at least 6 charts

**THANK YOU**