Final Project Proposal – A Look into New York City 311 Data

Xiaolan Li and Bernard Cooper

Yeshiva University | KATZ SCHOOL

NYC OpenData

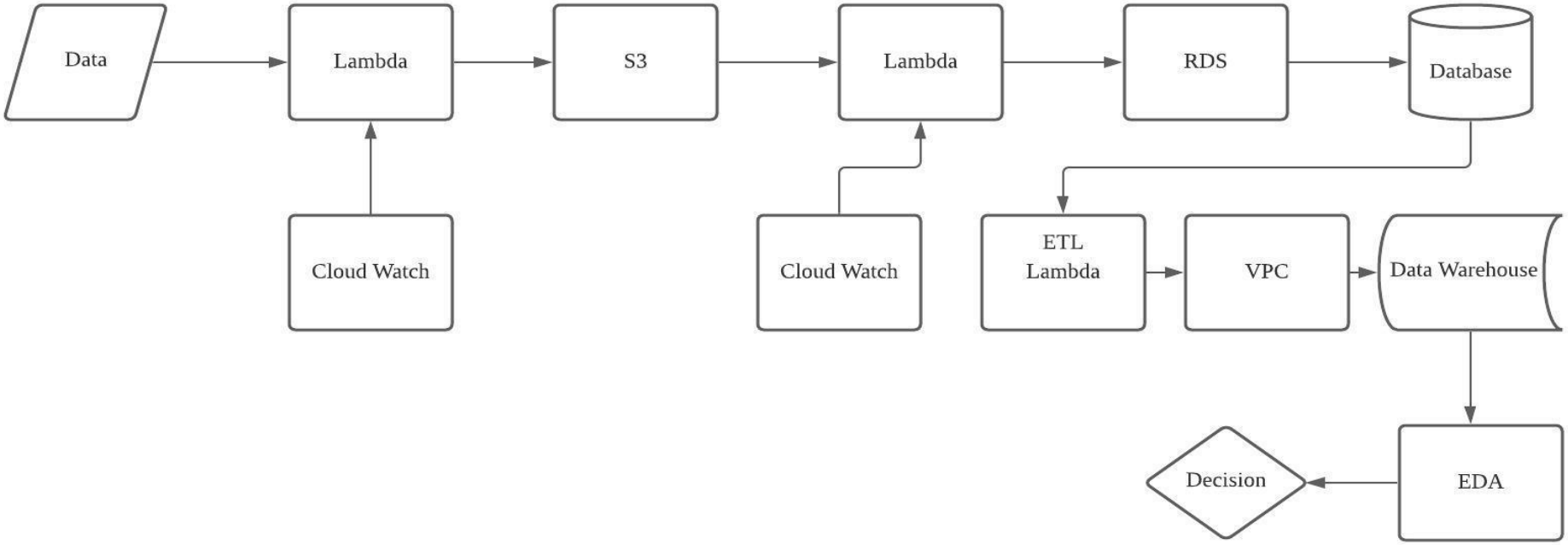
Description

The purpose of our project is to get a better analytical understanding of 311 calls in New York City by looking at 311 Data from 2010 until present day as well as looking at the Median Income Data for New York City.

The we are trying see if income has an impact on the amount of calls as well as the types of calls. We also want to check how different regions within the boroughs compare to each other and the borough as a whole.

NYC OpenData

Flow



NYC OpenData

Datasets

We are using the 311 Data From 2010 to Present from https://data.cityofnewyork.us/Social-Services/311-Service-Requests-from-2010-to-Present/erm2-nwe9 and the Median Income Data from https://data.cccnewyork.org/data/table/66/median-incomes#66/107/62/a/a.

The 311 Data contains data from over 1 million 311 calls over the past 11 years. The Median income Data has a breakdown of the Household Income in New York City which is broken down into sections like Borough, Community Districts, Zip Code, as well as which type of household in those areas.

NYC OpenData

Datasets

Unstructured Data - The median income based on boroughs and regions web scraping from html unstructured format.

Structured Data - The remaining file which will be downloaded as structured csv file from NYC Open Data.

NYC OpenData

Dataset Profile

Data set Profile		
Datatype	Structured Data	Unstructured Data
name	311_Service_Requests_from_2010_to_Present	median_income
size	560 MB	2.54 KB
number of records	1048575	59
frequency of updates	daily	one time
Number of fields	16	2
Extraction methods	(download) csv	(web scraping) html

NYC OpenData

Dataset Profile Cont.

Dimentional Table				
city	**agency**	**incident**	**location**	**date**
id	id	id	id	id
name	code	city_id	incident_id	incident_id
median_income	name	agency_id	latitude	created_year
		complaint_type	longitude	created_month
		descriptor	address_type	created_day
		description	incident_address	closed_year
		channel_type	zip	closed_month
			location_type	closed_day
			borough	

Fact Table
case
case_id
incident_id
city_id
agency_id
location_id
date_id

Roles and Responsibilities

Integrations

- Batch/Migration – flat file interface to your system (Bernard, Xiaolan)
- Real-time – integrated to your solution (Xiaolan)

Platform

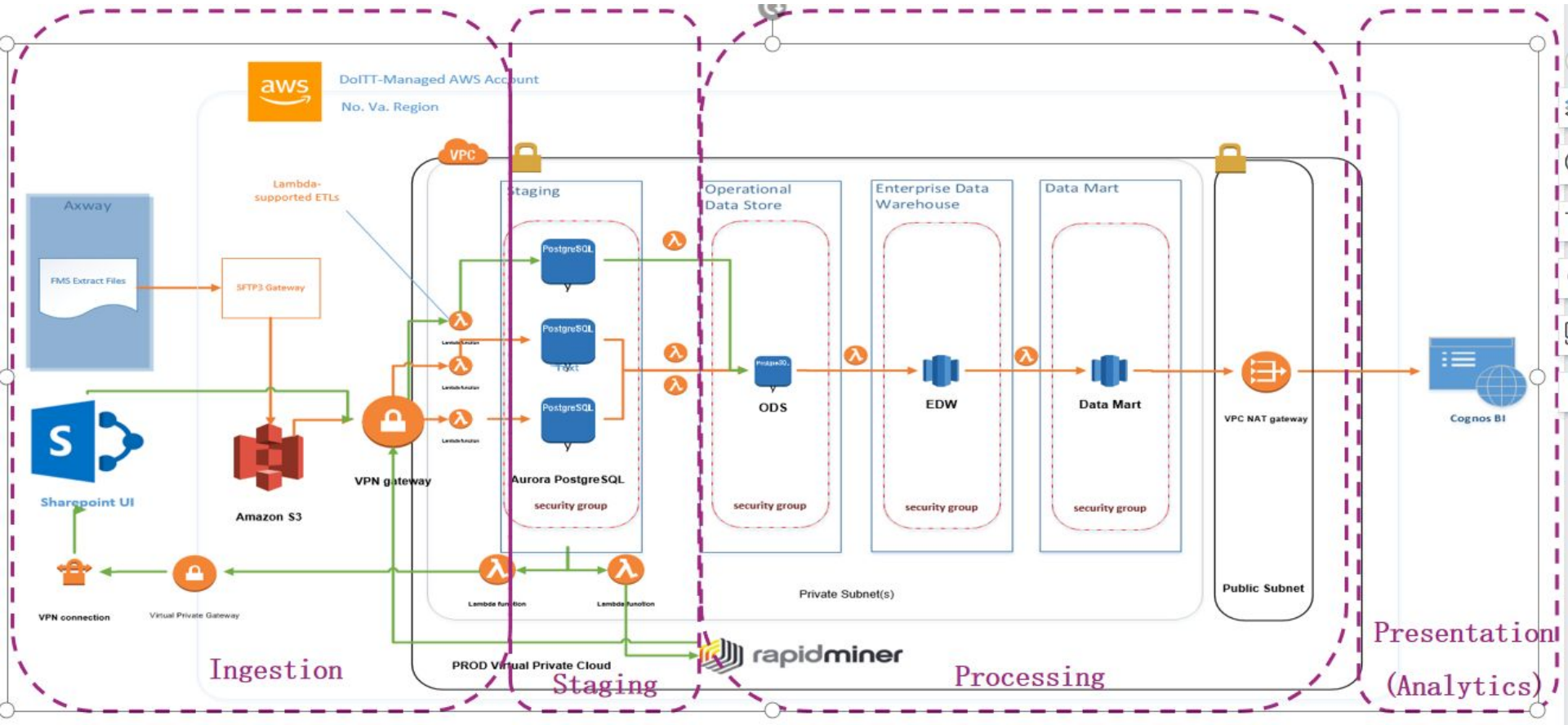
- Amazon Web Services (Bernard, Xiaolan)
- Any visualization software (Bernard)

Documentation

- Develop the necessary documentation outlined in each of the subsections (Bernard, Xiaolan)

High Level Timeline

(9 weeks totally)
Source- 5th week
Ingestion- 6th week
Staging- 7th, 8th, 9th week
Processing to Data Warehouse- 10th, 11th, 12th week
Visualization- 13th, 14th week


aws
DoITT-Managed AWS Account
No. Va. Region
VPC
Lambda-supported ETLs
Axway
FMS Extract Files
SFTP3 Gateway
Staging
Operational Data Store
Enterprise Data Warehouse
Data Mart
PostgreSQL
PostgreSQL
PostgreSQL
ODS
EDW
Data Mart
VPC NAT gateway
Cognos BI
Sharepoint UI
Amazon S3
VPN gateway
Aurora PostgreSQL
security group
security group
security group
security group
VPN connection
Virtual Private Gateway
Lambda function
Lambda function
Private Subnet(s)
Public Subnet
PROD Virtual Private Cloud
rapidminer
Ingestion
Staging
Processing
Presentation
(Analytics)