Jinming Chen

Benjamin Freund

Ziyang Guo

Randy Leon

Alejandra Zapata

Porfessor Brandon Chiazza

DAV 6100 Information Architecture

**Hate Crimes in NYC – Data Warehouse**

**Basic Use Cases / User Stories for Analytics**

*Basic Use Case #1*:

Captain Raymond Holt is a captain at the New York City’s 99th police precinct. He is in charge of many detectives and officers at the precinct. Seeing as the 99th precinct covers a large portion of Brooklyn, Captain Holt needs to obtain, analyze, and understand data from numerous sources and systems. He wants to look at crimes recorded throughout Brooklyn and determine if there are trends amongst the crimes that criminals commit or their motives behind those crimes.

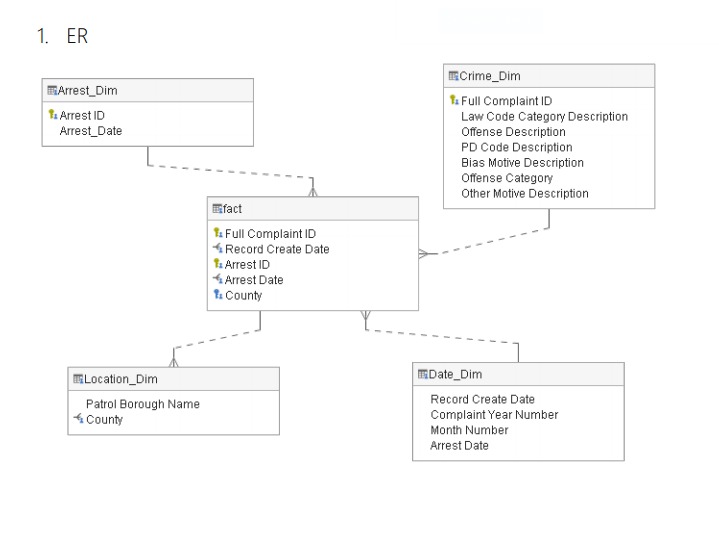
*Basic Use Case #2*:

Agent Jimmy Woo works for the Federal Bureau of Investigation. The crimes Agent Woo deals with at the FBI are typically on a more global scale. Because of this reality, he often hears about crimes and threats over the news before his superior agents can relay them to him. Consequently, Agent Woo wants a place where he can track both crime records and various news sources, so he can always stay on top of leads and tips to help him solve his cases.

*Basic Use Case #3*:

Brandon Chiazza has been the CTO at the City of New York’s Mayor’s Office of Contract Services since 2016. As an avid data enthusiast, Mr. Chiazza always wants to learn more about the city he serves, so he can better help the people within the city and the city at large. He heard about increasing crime rates throughout NYC, and he was curious how he could best assist NYC’s mayor in tackling this issue. To do this, Mr. Chiazza is looking for a centralized place where he can view all of the violent crimes that have been occurring in NYC over the past couple of years.

**Proposed Dimension Model ERD**

****

**Conceptual AWS Architecture Diagram**

****

**ETL Steps**

*E: Extract*

* Logical data map: The data was obtained in a structured an unstructured process.
* Judge Data: The data profiling was done to the data and the team is able to fill in the missing gaps for the dimensional model.
* Extract: By profiling the data we are able to understand business use cases and organized the data into a snowflake schema. We create the dimension and fact tables and extract the data.

*T: Transform*

* Control: We connect the data on the primary and surrogate keys and find any information mis formatted or missing data so that our dimensional model works properly.
  + We also perform different test queries as a control that our data and model is successful.
* Integration: We have a well formatted and working dimensional model.

*L: Load*

* Time variance: We add a type 2 dimension to track changes in the data.
* Keys
* Result: Our dimension model facts and dimensions are ready.