

Crime Database System

Chaeun Ryu (cr3413) Eesun Moon (em3907)

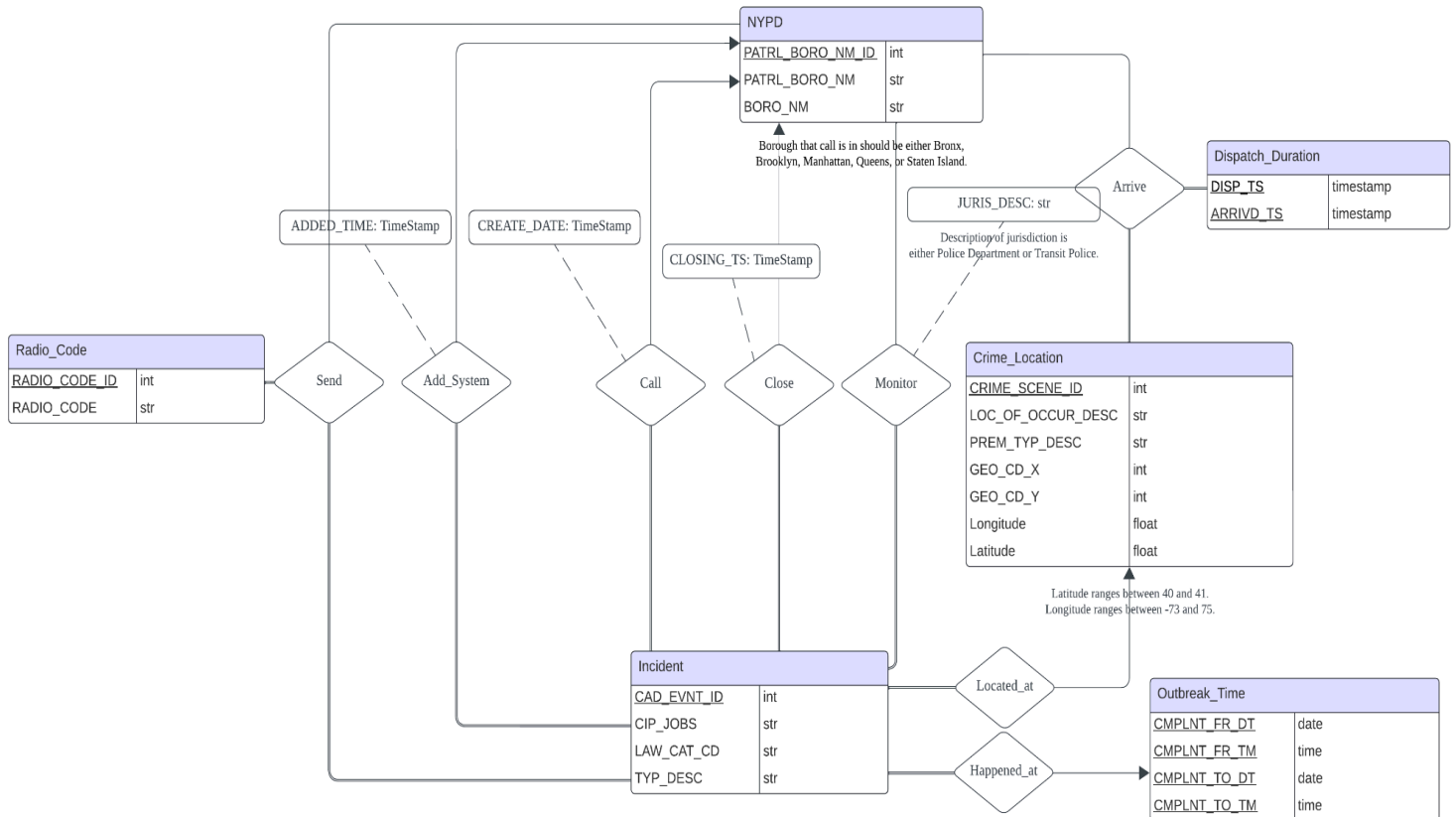
Description of Application

This database application is designed to manage and analyze crime incidents handled by the New York Police Department(NYPD).

The system incorporates six main entities: **NYPD**, **Dispatch_Duration**, **Incident**, **Outbreak_time**, **Location**, and **Radio_Code**. **NYPD** represents patrol units assigned to boroughs (Bronx, Brooklyn, Manhattan, Queens, and Staten Island) as shown in the plain-english text description in the E/R diagram, each identified by unique patrol borough IDs. **Dispatch_Duration** logs timestamps of when NYPD units were dispatched and arrived at crime scenes. **Incident** records crime events, including event ID, crime in progress, level of offense (felony, misdemeanor, or violation), and the description of the crime. **Outbreak_time** captures temporal information, such as the exact and ending date and time of incidents. **Location** stores spatial data such as location (i.e., inside, opposite of, front of, or rear of), premises (i.e., grocery store, resident, or street), geospatial coordinates (X and Y midblock coordinates), latitude, and longitude. Latitude and longitude of New York City have specific ranges, denoted in the plain text english description in the E/R diagram. **Radio_Code** records communication codes used during incidents.

The system also establishes eight relationships: **Call**, **Add_System**, **Close**, **Send**, **Monitor**, **Arrive**, **Located_at**, and **Happened_at**. The three relationships, **Call**, **Add_System**, and **Close**, represent when the incidents are reported, added to the system, and closed, respectively. For these three relationships, these three events only occur once per incident to the NYPD, necessitating both key and participation constraints to convey: 'Each incident has to be called, added to system, or closed by/to the NYPD exactly one time.' For the **Send** relationship, the radio code can be sent at least once depending on the status of the crime. For example, if there is new information that needs to be shared or a new police department to be shared with about the crime, there will be a new radio code sent which enables multiple radio codes to be sent for a specific incident. Likewise, for the **Monitor** relationship, the particular incident can be monitored at least one time depending on their needs by the department in charge, which is denoted by the jurisdiction description. The jurisdiction description indicates that the incident should be monitored by either police department or transit police as given in the plain English text description. For the **Arrive** relationship, crime location is arrived at by at least one department from the NYPD which explains the participation constraint. For instance, if the incident is so serious that more than one NYPD unit should be dispatched at the crime scene, there can be multiple timestamps of arrival to the crime location. Likewise, when the NYPD is dispatched multiple times, there will be multiple dispatch durations so there is a participation constraint for **Dispatch_Duration**, as well. Lastly, **Located_at** and **Happened_at** are relationships explaining where the incident was located and when it occurred, respectively. Each incident has to occur at one specific place or timestamp, represented by using both key and participation constraints.

ER Diagram



SQL Schema

```
CREATE TABLE NYPD (
    PATRL_BORO_NM_ID INTEGER,
    PATRL_BORO_NM CHAR(3),
    BORO_NM CHAR(20),
    PRIMARY KEY(PATRL_BORO_NM_ID)
)
```

```
CREATE TABLE Incident (
    CAD_EVT_ID INTEGER,
    CIP_JOBS CHAR(20),
    LAW_CAT_CD CHAR(15),
    TYP_DESC CHAR(100),
    PRIMARY KEY(CAD_EVT_ID)
)
```

```
CREATE TABLE Monitor (  
    CAD_EVNT_ID INTEGER,  
    PATRL_BORO_NM_ID INTEGER,  
    JURIS_DESC CHAR(20),  
    PRIMARY KEY (CAD_EVNT_ID, PATRL_BORO_NM_ID),  
    FOREIGN KEY (CAD_EVNT_ID) REFERENCES Incident,  
    FOREIGN KEY (PATRL_BORO_NM_ID) REFERENCES NYPD  
)  
## We cannot capture such participation constraints (yet) in SQL
```

```
CREATE TABLE Call (  
    CAD_EVNT_ID INTEGER,  
    PATRL_BORO_NM_ID INTEGER NOT NULL,  
    CREATE_DATE TIMESTAMP,  
    PRIMARY KEY (CAD_EVNT_ID),  
    FOREIGN KEY (CAD_EVNT_ID) REFERENCES Incident,  
    FOREIGN KEY (PATRL_BORO_NM_ID) REFERENCES NYPD  
)  
## We cannot capture such participation constraints (yet) in SQL
```

```
CREATE TABLE Close (  
    CAD_EVNT_ID INTEGER,  
    PATRL_BORO_NM_ID INTEGER NOT NULL,  
    CLOSING_TS TIMESTAMP,  
    PRIMARY KEY (CAD_EVNT_ID),  
    FOREIGN KEY (CAD_EVNT_ID) REFERENCES Incident,  
    FOREIGN KEY (PATRL_BORO_NM_ID) REFERENCES NYPD  
)  
## We cannot capture such participation constraints (yet) in SQL
```

```
CREATE TABLE Add_System (  
    CAD_EVNT_ID INTEGER,  
    PATRL_BORO_NM_ID INTEGER NOT NULL,  
    ADDED_TIME TIMESTAMP,  
    PRIMARY KEY(CAD_EVNT_ID),  
    FOREIGN KEY (CAD_EVNT_ID) REFERENCES Incident,  
    FOREIGN KEY (PATRL_BORO_NM_ID) REFERENCES NYPD  
)
```

```
CREATE TABLE Radio_Code (  
    RADIO_CODE_ID INTEGER,  
    RADIO_CODE CHAR(5),  
    PRIMARY KEY(RADIO_CODE_ID)  
)
```

```
CREATE TABLE Send (  
    CAD_EVNT_ID INTEGER,  
    PATRL_BORO_NM_ID INTEGER,  
    RADIO_CODE_ID INTEGER,  
    PRIMARY KEY(CAD_EVNT_ID, PATRL_BORO_NM_ID, RADIO_CODE_ID),  
    FOREIGN KEY (CAD_EVNT_ID) REFERENCES Incident,  
    FOREIGN KEY (PATRL_BORO_NM_ID) REFERENCES NYPD,  
    FOREIGN KEY (RADIO_CODE_ID) REFERENCES Radio_Code  
)
```

We cannot capture such participation constraints (yet) in SQL

```
CREATE TABLE Outbreak_Time (  
    CMPLNT_FR_DT DATE,  
    CMPLNT_FR_TM TIME,  
    CMPLNT_TO_DT DATE,  
    CMPLNT_TO_TM TIME,  
    PRIMARY KEY(CMPLNT_FR_DT, CMPLNT_FR_TM, CMPLNT_TO_DT, CMPLNT_TO_TM)  
)
```

```
CREATE TABLE Happened_at (  
    CAD_EVNT_ID INTEGER,  
    CMPLNT_FR_DT DATE NOT NULL,  
    CMPLNT_FR_TM TIME NOT NULL,  
    CMPLNT_TO_DT DATE NOT NULL,  
    CMPLNT_TO_TM TIME NOT NULL,  
    PRIMARY KEY (CAD_EVNT_ID),  
    FOREIGN KEY (CAD_EVNT_ID) REFERENCES Incident,  
    FOREIGN KEY (CMPLNT_FR_DT, CMPLNT_FR_TM, CMPLNT_TO_DT, CMPLNT_TO_TM)  
    REFERENCES Outbreak_Time  
)
```

```

CREATE TABLE Crime_Location (
    CRIME_SCENE_ID INTEGER,
    LOC_OF_OCCUR_DESC CHAR(10),
    PREM_TYP_DESC CHAR(20),
    GEO_CD_X INTEGER,
    GEO_CD_Y INTEGER,
    Longitude REAL,
    Latitude REAL,
    PRIMARY KEY(CRIME_SCENE_ID)
)

CREATE TABLE Located_at (
    CAD_EVNT_ID INTEGER,
    CRIME_SCENE_ID INTEGER NOT NULL,
    PRIMARY KEY (CAD_EVNT_ID),
    FOREIGN KEY (CAD_EVNT_ID) REFERENCES Incident,
    FOREIGN KEY (CRIME_SCENE_ID) REFERENCES Crime_Location
)

CREATE TABLE Dispatch_Duration (
    DISP_TS TIMESTAMP,
    ARRIVD_TS TIMESTAMP,
    PRIMARY KEY(DISP_TS, ARRIVD_TS)
)

CREATE TABLE Arrive (
    CRIME_SCENE_ID INTEGER,
    PATRL_BORO_NM_ID INTEGER,
    DISP_TS TIMESTAMP,
    ARRIVD_TS TIMESTAMP,
    PRIMARY KEY (CRIME_SCENE_ID, PATRL_BORO_NM_ID, DISP_TS, ARRIVD_TS),
    FOREIGN KEY (CRIME_SCENE_ID) REFERENCES Crime_Location,
    FOREIGN KEY (PATRL_BORO_NM_ID) REFERENCES NYPD,
    FOREIGN KEY (DISP_TS, ARRIVD_TS) REFERENCES Dispatch_Duration
)
## We cannot capture such participation constraints (yet) in SQL

```

Data Plan

We plan to collect real-world datasets acquired from the New York City Police Department and NYC Open Data. We are planning to do crawling to collect additional information from the New York City Police Department website and the data from NYC Open Data is accessible in a csv format. The data includes information about victims, suspects, where the crime took place, when it was committed, jurisdiction allocated for the incident, and etc.

Plan for Part 3: Expanded-Design Option

We plan to expand the design of the crime database by focusing on the people that were involved in the crime, which is approved by the instructor. In part I, the database held basic information about the crime such as location and time. In the expansion, we will add multiple stakeholders of the incident including police officers, suspects, victims, medical facilities, and court of law which makes our database more complex. We are able to expand the database with regard to the people involved as we have already collected data that is consistent with the original dataset. To be specific, we collected additional information regarding the stakeholders involved with respect to each unique incident from the data used in Part 1 and merged multiple datasets on multiple keys to create the final database. Age, race, and gender about the suspect or victim, the hierarchical structure of police departments, and which police department was in charge of the crime are examples we were able to collect as additional information. We believed that this is an essential expansion to our crime database, because we could also gather insights about demographics related to crime, such as the tendency of criminals, suspects, and average arriving time depending on the type of police officer dispatched.

Contingency Plan

As part of the contingency plan, instead of concentrating on time-based relationships, we will focus on the core relationships between patrol, precinct, and crime incidents, as these are essential for managing the basic crime-reporting process.

References of Data

- [NYPD Calls for Service \(Year to Date\)](#)
- [NYPD Complaint Data Historic](#)
- [Information about precincts](#)
- [Information about transit districts](#)