Dylan Phoutthavong

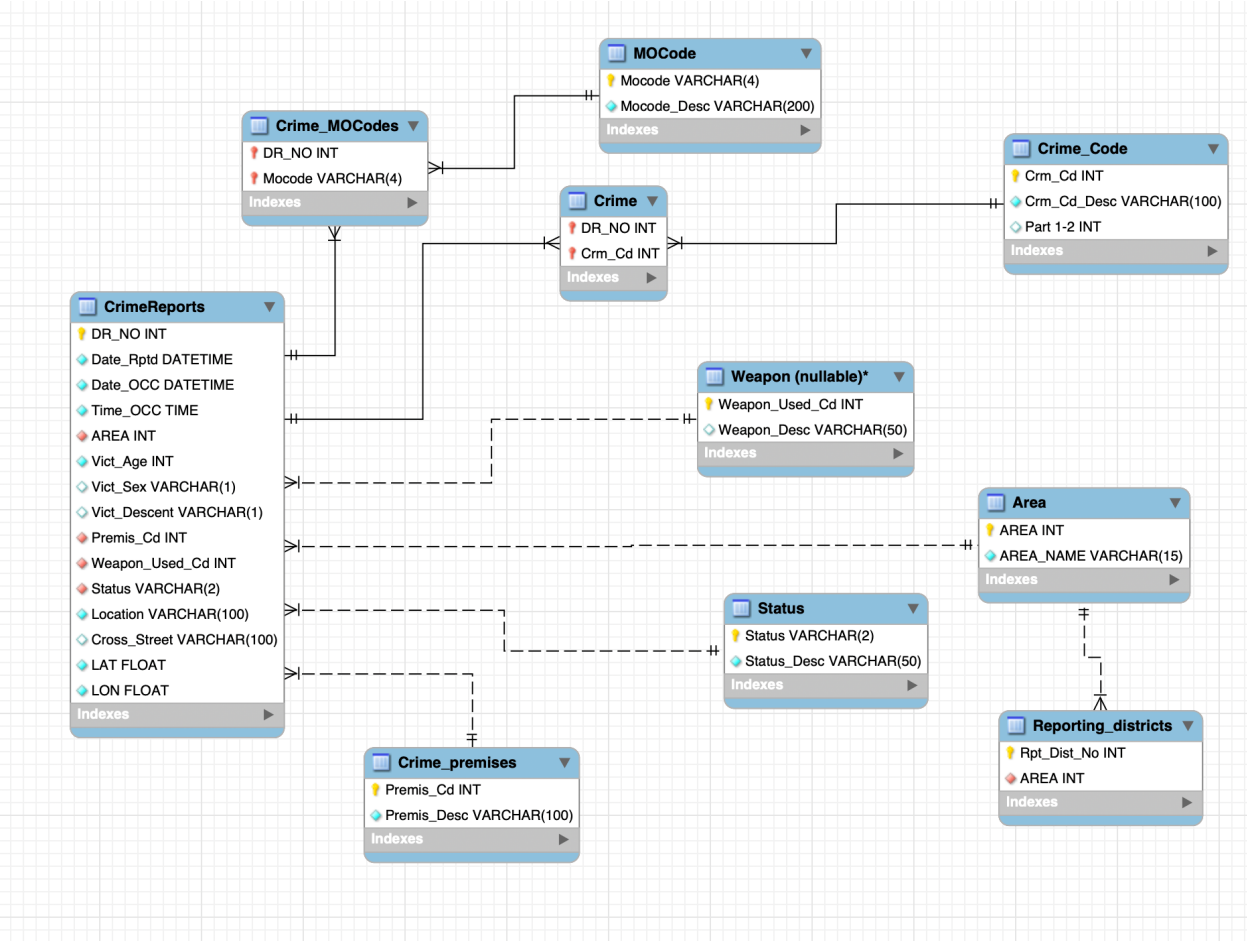
April 21st, 2025

CSCI 3287

PRJ 1

1 & 2) Normalization Process & Database Schema

3. ER Diagram



4. List of Assumptions

1. Each crime report (DR_NO) can have multiple crimes and multiple MOCodes.
2. Victim information (age, sex, descent) is only available at the report level, not in a separate table.
3. The Weapon_Used_Cd may be null, since not all crimes involve a.
4. All lookup tables (Weapon, Status, Premises, Area, Crime_Code, MOCODE_Description) are fully normalized and contain no redundant dependencies.
5. MOCODE values use VARCHAR(4) due to leading zeros found in the source CSV.
6. Area names are limited in length as per dataset; VARCHAR(10) is sufficient.
7. Reporting districts are uniquely identified by Rpt_Dist_No and belong to a single Area.

8. Crime table uses a composite key including Crime_Seriousness to preserve all Crm_Cd_1–4 values.
9. MOCODE_Description is treated as a reference table and is linked via the bridge table Crime_MOCodes.
10. Date and time fields are retained in standard DATETIME and TIME formats.

5. SQL Scripts PLAN for 6 Queries

Query	Description	Tables Used	Fields/Columns Involved	SQL Set Type
1	Total incidents per area	CrimeReports, Area	AREA_NAME, COUNT(DR_NO)	INNER JOIN, GROUP BY
2	Crimes and weapons used	CrimeReports, Crime, Crime_Code, Weapon	Crm_Cd_Desc, Weapon_Desc	LEFT OUTER JOIN
3	MO codes used in reports but not described	Crime_MOCodes, MOCODE_Description	Mocode	SET THEORY: EXCEPT / NOT IN
4	Number of crimes per premises type	CrimeReports, Crime_premises	Premis_Desc, COUNT(*)	INNER JOIN, GROUP BY
5	Top 5 most frequent crime types	Crime, Crime_Code	Crm_Cd_Desc, COUNT(*)	INNER JOIN, ORDER BY, LIMIT
6	Crime count by reporting district	CrimeReports, Reporting_districts	Rpt_Dist_No, COUNT(*)	INNER JOIN, GROUP BY