Algonquin College

CST2112

Assignment 5

National Address Register Dataset Normalization and ERD Design

<u>GROUP - 9</u>

Shara Khandakar

Anam Vakil

Maksuda E Elahi

Murk Asad

Submitted to: Andreas Gausrab

Table of Contents

1. Introduction

The goal of this assignment was to transform the raw GA5 National Address Register (NAR) dataset into a properly normalized relational database. The raw data was provided as multiple CSV files containing complex address components such as civic information, street names, municipality descriptors, province codes, legal land descriptions (DLS), and mailing address fields.

To prepare the dataset for reliable querying and long-term storage, the files were ingested into a staging table (GA5_Address) and then transformed into a fully normalized Third Normal Form (3NF) schema. The resulting model separates distinct real-world entities (street, municipality, province, geography, and mailing address) and links them through foreign key relationships to a centralized GA5_AddressNormalized table that represents each physical address.

2. Identifying Entities and Relationships

An initial review of the raw NAR data revealed that many attributes were highly repetitive across rows:

- Street names and street types appeared thousands of times.
- Municipality names and subdivision codes were repeated for all addresses within the same region.
- Province codes were duplicated across the entire dataset.
- Geography fields (BG_DLS_*) described legal land descriptions shared by multiple addresses.
- Mailing address attributes repeated where postal routing was consistent.

From these patterns, the following entities were identified:

1. GA5_Street – Official street name, type, and direction
2. GA5_Municipality – Census Subdivision (CSD) names and municipal type codes
3. GA5_Province – Province code and mailing abbreviation
4. GA5_Geography – Legal land description fields and coordinate values
5. GA5_MailingAddress – Mailing street, direction, municipality name, and postal code

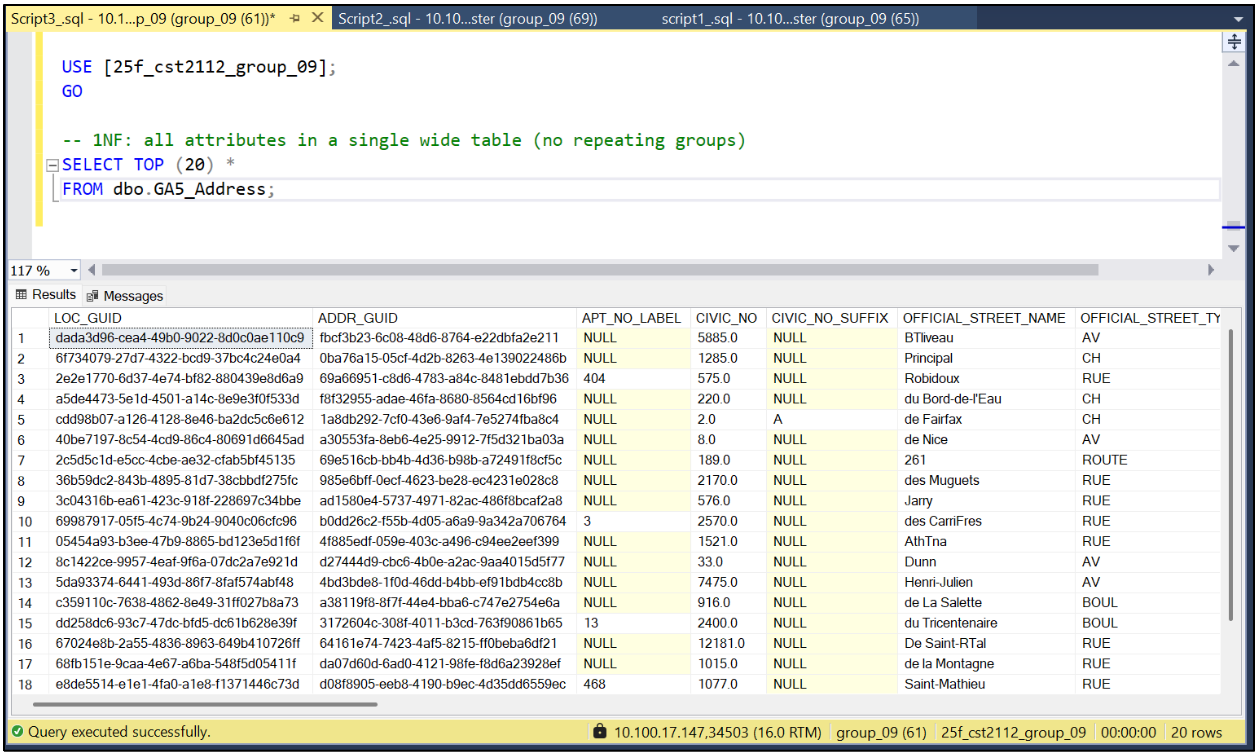
6. GA5_AddressNormalized – Central entity representing one physical address, with foreign keys to the above tables

The GA5_AddressNormalized table serves as the hub of the design. It contains address-specific attributes such as civic number, unit label, and unique identifiers, while referencing the supporting entity tables through foreign keys. This ensures consistency, avoids repetition, and supports efficient querying of address information.

3. Normalization Process

3.1 First Normal Form (1NF)

The original sourced NAR data was loaded into a staging table GA5_Address. In this structure, all data values were atomic (no repeating groups or multi-valued cells), which satisfies the definition of 1NF. However, this table still contained significant redundancy: street names, municipalities, provinces, and other attributes were repeated for every address row. This meant that updating a single attribute (such as a street name or municipality code) would require changes in multiple rows, increasing the risk of data inconsistencies.



Script3_.sql - 10.1...p_09 (group_09 (61))*
Script2_.sql - 10.10...ster (group_09 (69))
script1_.sql - 10.10...ster (group_09 (65))
117 %
Results
Messages

```
USE [25f_cst2112_group_09];
GO

-- 1NF: all attributes in a single wide table (no repeating groups)
SELECT TOP (20) *
FROM dbo.GA5_Address;
```

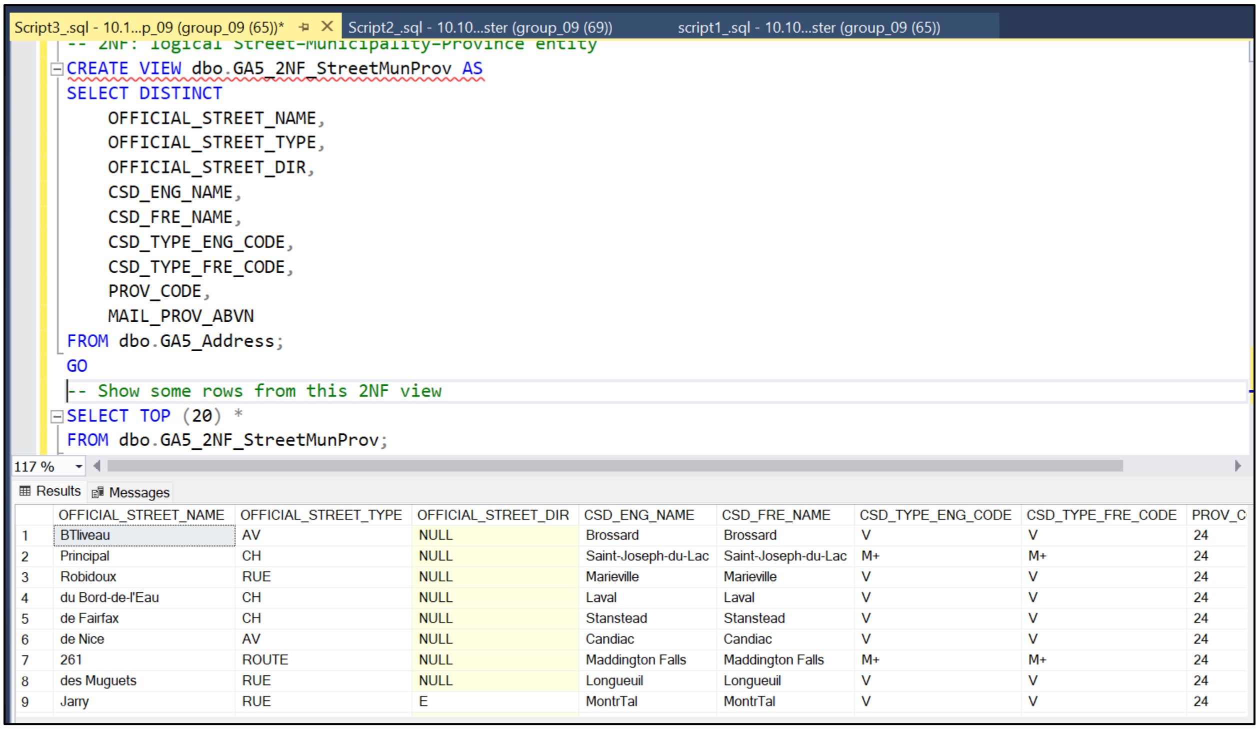
	LOC_GUID	ADDR_GUID	APT_NO_LABEL	CIVIC_NO	CIVIC_NO_SUFFIX	OFFICIAL_STREET_NAME	OFFICIAL_STREET_TY
1	dada3d96-cea4-49b0-9022-8d0c0ae110c9	fbcf3b23-6c08-48d6-8764-e22dbfa2e211	NULL	5885.0	NULL	BTliveau	AV
2	6f734079-27d7-4322-bcd9-37bc4c24e0a4	0ba76a15-05cf-4d2b-8263-4e139022486b	NULL	1285.0	NULL	Principal	CH
3	2e2e1770-6d37-4e74-bf82-880439e8d6a9	69a66951-c8d6-4783-a84c-8481ebdd7b36	404	575.0	NULL	Robidoux	RUE
4	a5de4473-5e1d-4501-a14c-8e9e3f0f533d	f8f32955-adae-46fa-8680-8564cd16bf96	NULL	220.0	NULL	du Bord-de-l'Eau	CH
5	cdd98b07-a126-4128-8e46-ba2dc5c6e612	1a8db292-7cf0-43e6-9af4-7e5274fba8c4	NULL	2.0	A	de Fairfax	CH
6	40be7197-8c54-4cd9-86c4-80691d6645ad	a30553fa-8eb6-4e25-9912-7f5d321ba03a	NULL	8.0	NULL	de Nice	AV
7	2c5d5c1d-e5cc-4cbe-ae32-cfab5bf45135	69e516cb-bb4b-4d36-b98b-a72491f8cf5c	NULL	189.0	NULL	261	ROUTE
8	36b59dc2-843b-4895-81d7-38cbbdf275fc	985e6bff-0ecf-4623-be28-ec4231e028c8	NULL	2170.0	NULL	des Muguets	RUE
9	3c04316b-ea61-423c-918f-228697c34bbe	ad1580e4-5737-4971-82ac-486f8bcaf2a8	NULL	576.0	NULL	Jarry	RUE
10	69987917-05f5-4c74-9b24-9040c06cfc96	b0dd26c2-f55b-4d05-a6a9-9a342a706764	3	2570.0	NULL	des CarriFres	RUE
11	05454a93-b3ee-47b9-8865-bd123e5d1f6f	4f885edf-059e-403c-a496-c94ee2eef399	NULL	1521.0	NULL	AthTna	RUE
12	8c1422ce-9957-4eaf-9f6a-07dc2a7e921d	d27444d9-cbc6-4b0e-a2ac-9aa4015d5f77	NULL	33.0	NULL	Dunn	AV
13	5da93374-6441-493d-86f7-8faf574abf48	4bd3bde8-1f0d-46dd-b4bb-ef91bdb4cc8b	NULL	7475.0	NULL	Henri-Julien	AV
14	c359110c-7638-4862-8e49-31ff027b8a73	a38119f8-8f7f-44e4-bba6-c747e2754e6a	NULL	916.0	NULL	de La Salette	BOUL
15	dd258dc6-93c7-47dc-bfd5-dc61b628e39f	3172604c-308f-4011-b3cd-763f90861b65	13	2400.0	NULL	du Tricentenaire	BOUL
16	67024e8b-2a55-4836-8963-649b410726ff	64161e74-7423-4af5-8215-ff0beba6df21	NULL	12181.0	NULL	De Saint-RTal	RUE
17	68fb151e-9caa-4e67-a6ba-548f5d05411f	da07d60d-6ad0-4121-98fe-f8d6a23928ef	NULL	1015.0	NULL	de la Montagne	RUE
18	e8de5514-e1e1-4fa0-a1e8-f1371446c73d	d08f8905-eeb8-4190-b9ec-4d35dd6559ec	468	1077.0	NULL	Saint-Mathieu	RUE

Query executed successfully. | 10.100.17.147,34503 (16.0 RTM) | group_09 (61) | 25f_cst2112_group_09 | 00:00:00 | 20 rows

Figure1: 1NF GA5_Address (all attributes, still redundant)

3.2 Second Normal Form (2NF)

To remove partial dependencies, logical groupings of attributes were identified that did not depend on the full address key. For example, street details, municipality details, and province information were repeated across multiple addresses. A 2NF representation was created by projecting these attributes into a grouped structure using a SELECT DISTINCT view, separating them from dwelling-specific elements such as civic numbers and unit labels. This eliminated partial dependencies by ensuring that attributes like street name or municipality type did not depend on the entire address record..



Script3_.sql - 10.1...p_09 (group_09 (65))*
Script2_.sql - 10.10...ster (group_09 (69))
script1_.sql - 10.10...ster (group_09 (65))
117 %
Results
Messages

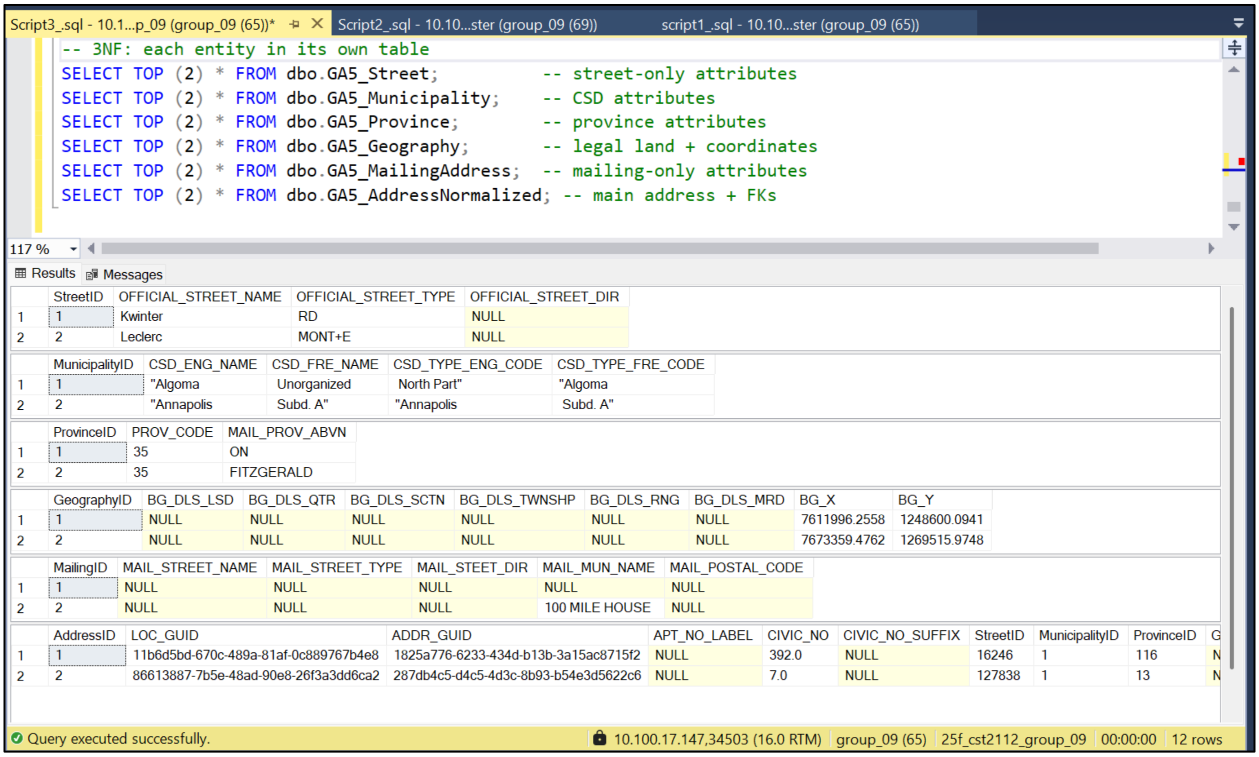
```sql
-- 2NF: logical Street-Municipality-Province entity
CREATE VIEW dbo.GA5_2NF_StreetMunProv AS
SELECT DISTINCT
    OFFICIAL_STREET_NAME,
    OFFICIAL_STREET_TYPE,
    OFFICIAL_STREET_DIR,
    CSD_ENG_NAME,
    CSD_FRE_NAME,
    CSD_TYPE_ENG_CODE,
    CSD_TYPE_FRE_CODE,
    PROV_CODE,
    MAIL_PROV_ABVN
FROM dbo.GA5_Address;
GO
-- Show some rows from this 2NF view
SELECT TOP (20) *
FROM dbo.GA5_2NF_StreetMunProv;
```

	OFFICIAL_STREET_NAME	OFFICIAL_STREET_TYPE	OFFICIAL_STREET_DIR	CSD_ENG_NAME	CSD_FRE_NAME	CSD_TYPE_ENG_CODE	CSD_TYPE_FRE_CODE	PROV_C
1	BTliveau	AV	NULL	Brossard	Brossard	V	V	24
2	Principal	CH	NULL	Saint-Joseph-du-Lac	Saint-Joseph-du-Lac	M+	M+	24
3	Robidoux	RUE	NULL	Marieville	Marieville	V	V	24
4	du Bord-de-l'Eau	CH	NULL	Laval	Laval	V	V	24
5	de Fairfax	CH	NULL	Stanstead	Stanstead	V	V	24
6	de Nice	AV	NULL	Candiac	Candiac	V	V	24
7	261	ROUTE	NULL	Maddington Falls	Maddington Falls	M+	M+	24
8	des Muguets	RUE	NULL	Longueuil	Longueuil	V	V	24
9	Jarry	RUE	E	MontrTal	MontrTal	V	V	24

Figure 2: 2NF – Street–Municipality–Province grouped (partial dependencies removed from full address)

3.3 Third Normal Form (3NF)

The final design decomposed the address data into separate tables, each representing a single entity: GA5_Street, GA5_Municipality, GA5_Province, GA5_Geography, GA5_MailingAddress, and GA5_AddressNormalized. Each table contains only attributes that depend on its primary key. All transitive dependencies were removed — for example, province abbreviations are stored in GA5_Province and not in the address table. The central GA5_AddressNormalized table contains address-specific attributes and foreign keys linking to the normalized entity tables. This structure eliminates redundancy, maintains data consistency, and enables lossless reconstruction of the original address via joins



Script3_.sql - 10.1...p_09 (group_09 (65))*
Script2_.sql - 10.10...ster (group_09 (69))
script1_.sql - 10.10...ster (group_09 (65))
117 %
Results
Messages

```sql
-- 3NF: each entity in its own table
SELECT TOP (2) * FROM dbo.GA5_Street;          -- street-only attributes
SELECT TOP (2) * FROM dbo.GA5_Municipality;    -- CSD attributes
SELECT TOP (2) * FROM dbo.GA5_Province;        -- province attributes
SELECT TOP (2) * FROM dbo.GA5_Geography;       -- legal land + coordinates
SELECT TOP (2) * FROM dbo.GA5_MailingAddress;  -- mailing-only attributes
SELECT TOP (2) * FROM dbo.GA5_AddressNormalized; -- main address + FKs
```

	StreetID	OFFICIAL_STREET_NAME	OFFICIAL_STREET_TYPE	OFFICIAL_STREET_DIR
1	1	Kwinter	RD	NULL
2	2	Leclerc	MONT+E	NULL

	MunicipalityID	CSD_ENG_NAME	CSD_FRE_NAME	CSD_TYPE_ENG_CODE	CSD_TYPE_FRE_CODE
1	1	"Algoma	Unorganized	North Part"	"Algoma
2	2	"Annapolis	Subd. A"	"Annapolis	Subd. A"

	ProvinceID	PROV_CODE	MAIL_PROV_ABVN
1	1	35	ON
2	2	35	FITZGERALD

	GeographyID	BG_DLS_LSD	BG_DLS_QTR	BG_DLS_SCTN	BG_DLS_TWNSHP	BG_DLS_RNG	BG_DLS_MRD	BG_X	BG_Y
1	1	NULL	NULL	NULL	NULL	NULL	NULL	7611996.2558	1248600.0941
2	2	NULL	NULL	NULL	NULL	NULL	NULL	7673359.4762	1269515.9748

	MailingID	MAIL_STREET_NAME	MAIL_STREET_TYPE	MAIL_STEET_DIR	MAIL_MUN_NAME	MAIL_POSTAL_CODE
1	1	NULL	NULL	NULL	NULL	NULL
2	2	NULL	NULL	NULL	100 MILE HOUSE	NULL

	AddressID	LOC_GUID	ADDR_GUID	APT_NO_LABEL	CIVIC_NO	CIVIC_NO_SUFFIX	StreetID	MunicipalityID	ProvinceID	G
1	1	11b6d5bd-670c-489a-81af-0c889767b4e8	1825a776-6233-434d-b13b-3a15ac8715f2	NULL	392.0	NULL	16246	1	116	N
2	2	86613887-7b5e-48ad-90e8-26f3a3dd6ca2	287db4c5-d4c5-4d3c-8b93-b54e3d5622c6	NULL	7.0	NULL	127838	1	13	N

```
Query executed successfully.    10.100.17.147,34503 (16.0 RTM) | group_09 (65) | 25f_cst2112_group_09 | 00:00:00 | 12 rows
```

Figure 3: Final Tables converted into 3NF

Summary:

The final 3NF schema ensures that each real-world concept is stored once, all non-key attributes depend only on the key of their table, and the original address can be reconstructed without data loss using foreign key relationships.

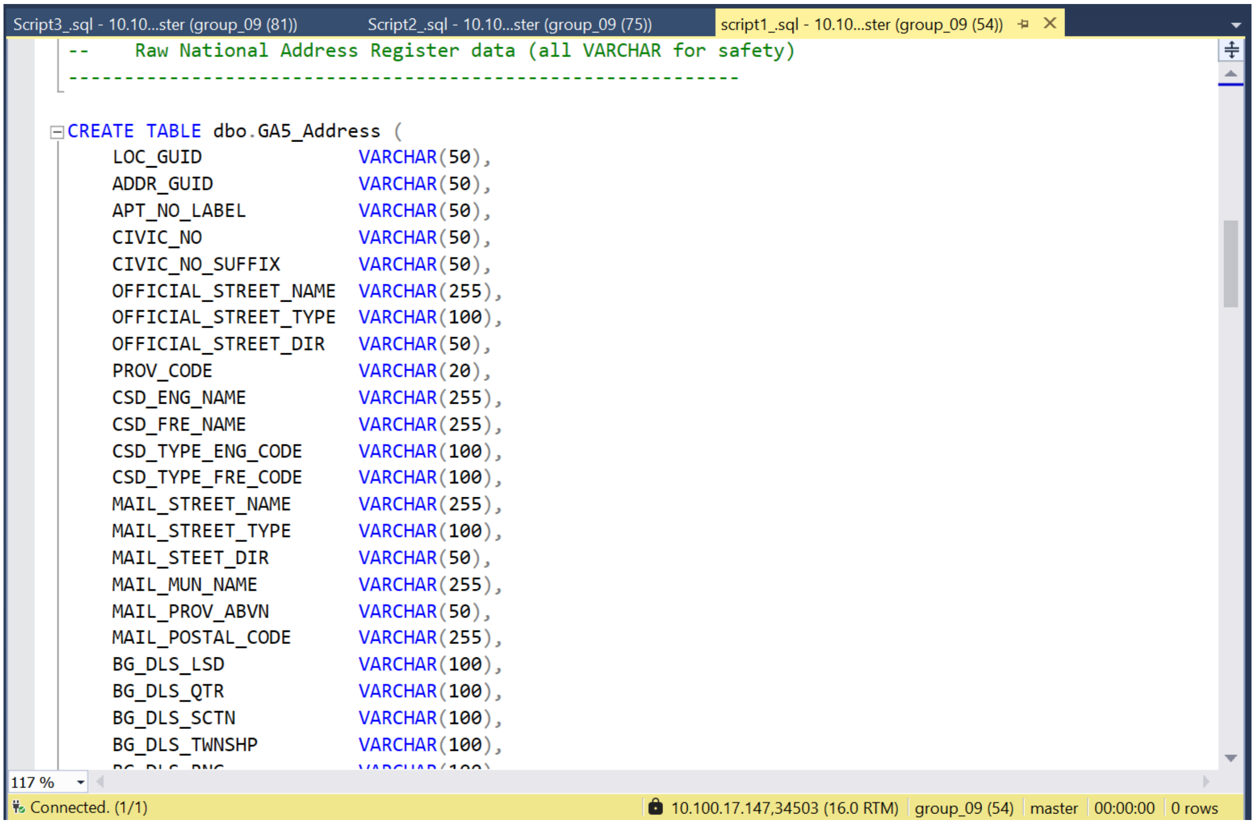
The series of figures demonstrate the transition from a single denormalized table to a fully normalized 3NF schema where each real-world entity is stored once, removing redundancy and preserving data integrity.

4. Implementation Summary

Three T-SQL scripts were created to automate the database pipeline.

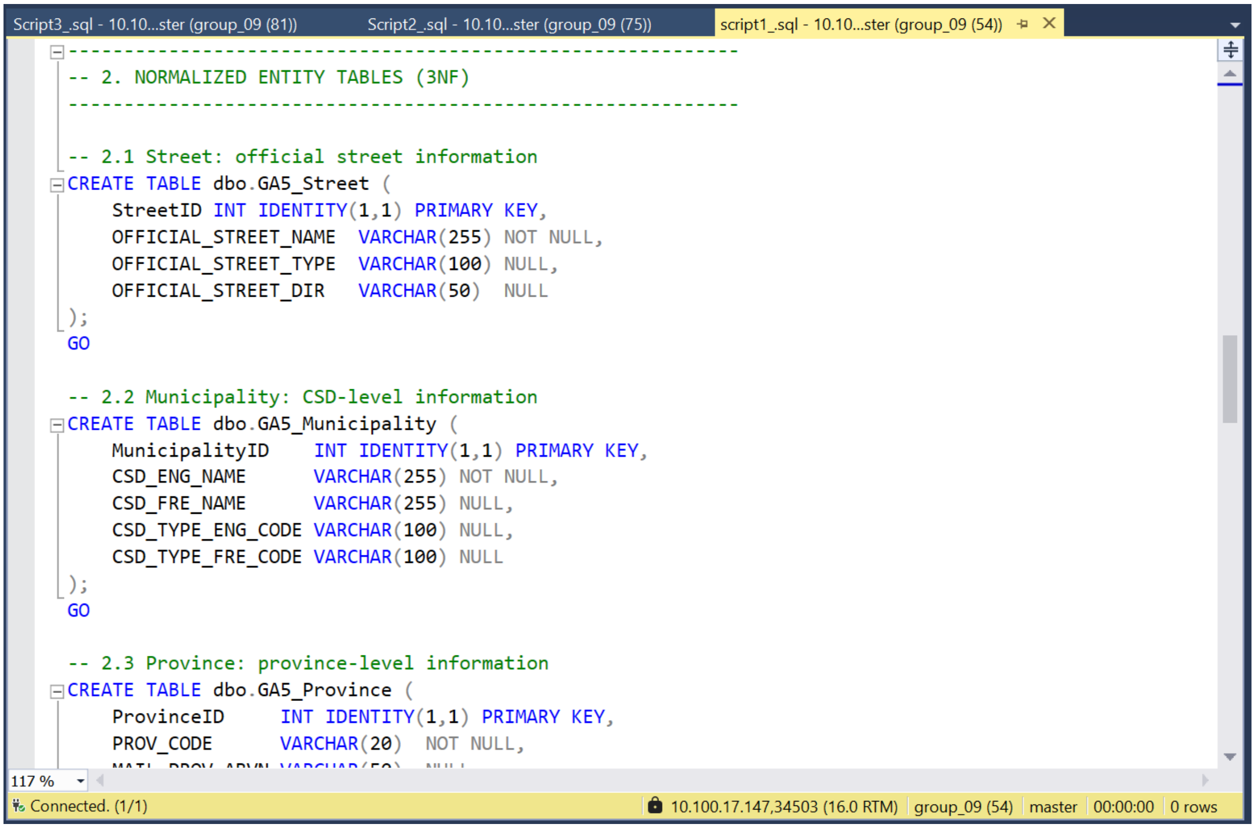
Script 1 – Table Creation

Created the staging table and all normalized tables, including primary keys, identity columns, and foreign key constraints.



Script3_.sql - 10.10...ster (group_09 (81))
Script2_.sql - 10.10...ster (group_09 (75))
script1_.sql - 10.10...ster (group_09 (54))
117 %
Connected. (1/1)
10.100.17.147,34503 (16.0 RTM)
group_09 (54)
master
00:00:00
0 rows

```sql
--    Raw National Address Register data (all VARCHAR for safety)
------------------------------------------------------------

CREATE TABLE dbo.GA5_Address (
    LOC_GUID                VARCHAR(50),
    ADDR_GUID               VARCHAR(50),
    APT_NO_LABEL            VARCHAR(50),
    CIVIC_NO                VARCHAR(50),
    CIVIC_NO_SUFFIX         VARCHAR(50),
    OFFICIAL_STREET_NAME    VARCHAR(255),
    OFFICIAL_STREET_TYPE    VARCHAR(100),
    OFFICIAL_STREET_DIR     VARCHAR(50),
    PROV_CODE               VARCHAR(20),
    CSD_ENG_NAME            VARCHAR(255),
    CSD_FRE_NAME            VARCHAR(255),
    CSD_TYPE_ENG_CODE       VARCHAR(100),
    CSD_TYPE_FRE_CODE       VARCHAR(100),
    MAIL_STREET_NAME        VARCHAR(255),
    MAIL_STREET_TYPE        VARCHAR(100),
    MAIL_STEET_DIR          VARCHAR(50),
    MAIL_MUN_NAME           VARCHAR(255),
    MAIL_PROV_ABVN          VARCHAR(50),
    MAIL_POSTAL_CODE        VARCHAR(255),
    BG_DLS_LSD              VARCHAR(100),
    BG_DLS_QTR              VARCHAR(100),
    BG_DLS_SCTN             VARCHAR(100),
    BG_DLS_TWNSHP           VARCHAR(100),
    BG_DLS_RNG              VARCHAR(100),
```

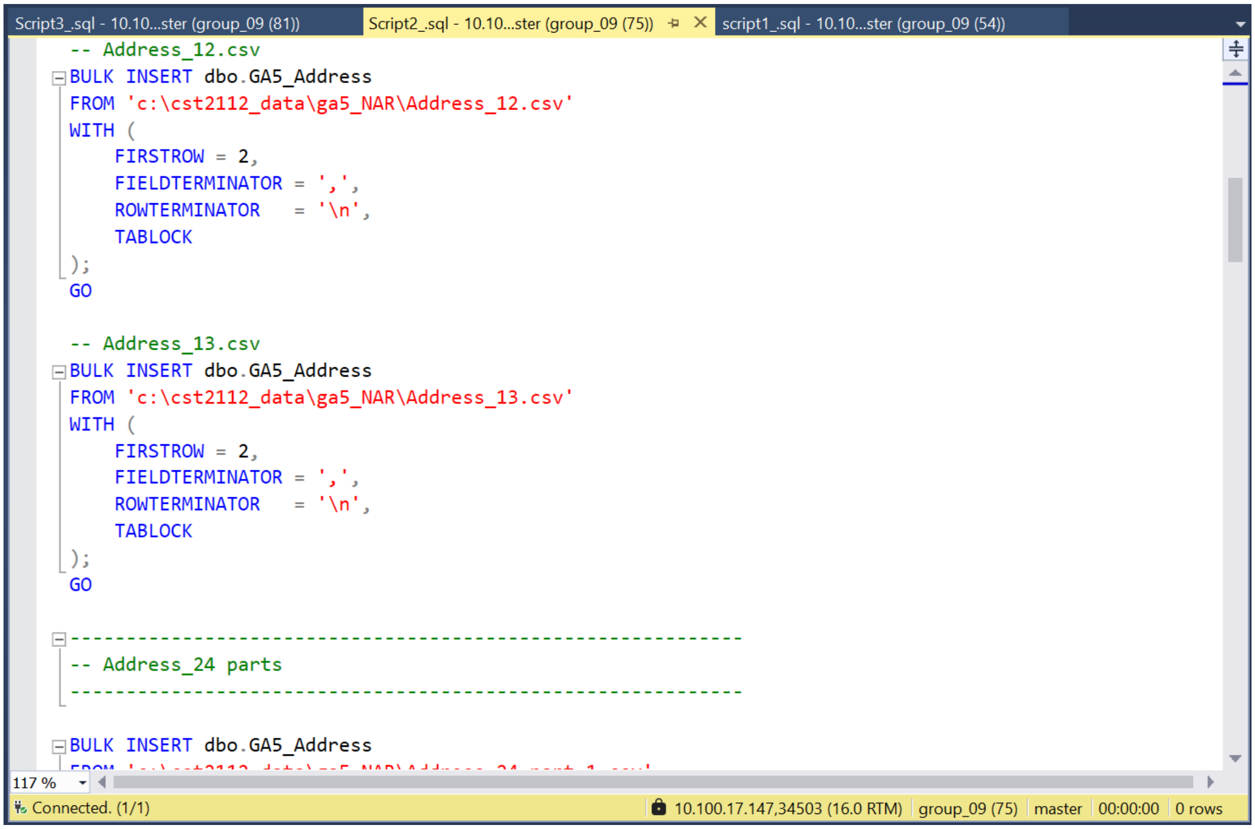


Script3_.sql - 10.10...ster (group_09 (81))
Script2_.sql - 10.10...ster (group_09 (75))
script1_.sql - 10.10...ster (group_09 (54))
117 %
Connected. (1/1)
10.100.17.147,34503 (16.0 RTM)
group_09 (54)
master
00:00:00
0 rows

```sql
------------------------------------------------------------
-- 2. NORMALIZED ENTITY TABLES (3NF)
------------------------------------------------------------

-- 2.1 Street: official street information
CREATE TABLE dbo.GA5_Street (
    StreetID INT IDENTITY(1,1) PRIMARY KEY,
    OFFICIAL_STREET_NAME  VARCHAR(255) NOT NULL,
    OFFICIAL_STREET_TYPE  VARCHAR(100) NULL,
    OFFICIAL_STREET_DIR   VARCHAR(50)  NULL
);
GO

-- 2.2 Municipality: CSD-level information
CREATE TABLE dbo.GA5_Municipality (
    MunicipalityID    INT IDENTITY(1,1) PRIMARY KEY,
    CSD_ENG_NAME      VARCHAR(255) NOT NULL,
    CSD_FRE_NAME      VARCHAR(255) NULL,
    CSD_TYPE_ENG_CODE VARCHAR(100) NULL,
    CSD_TYPE_FRE_CODE VARCHAR(100) NULL
);
GO

-- 2.3 Province: province-level information
CREATE TABLE dbo.GA5_Province (
    ProvinceID     INT IDENTITY(1,1) PRIMARY KEY,
    PROV_CODE      VARCHAR(20)  NOT NULL,
    MAIL_PROV_ABVN VARCHAR(50)  NULL
```

Script 2 – Data Loading

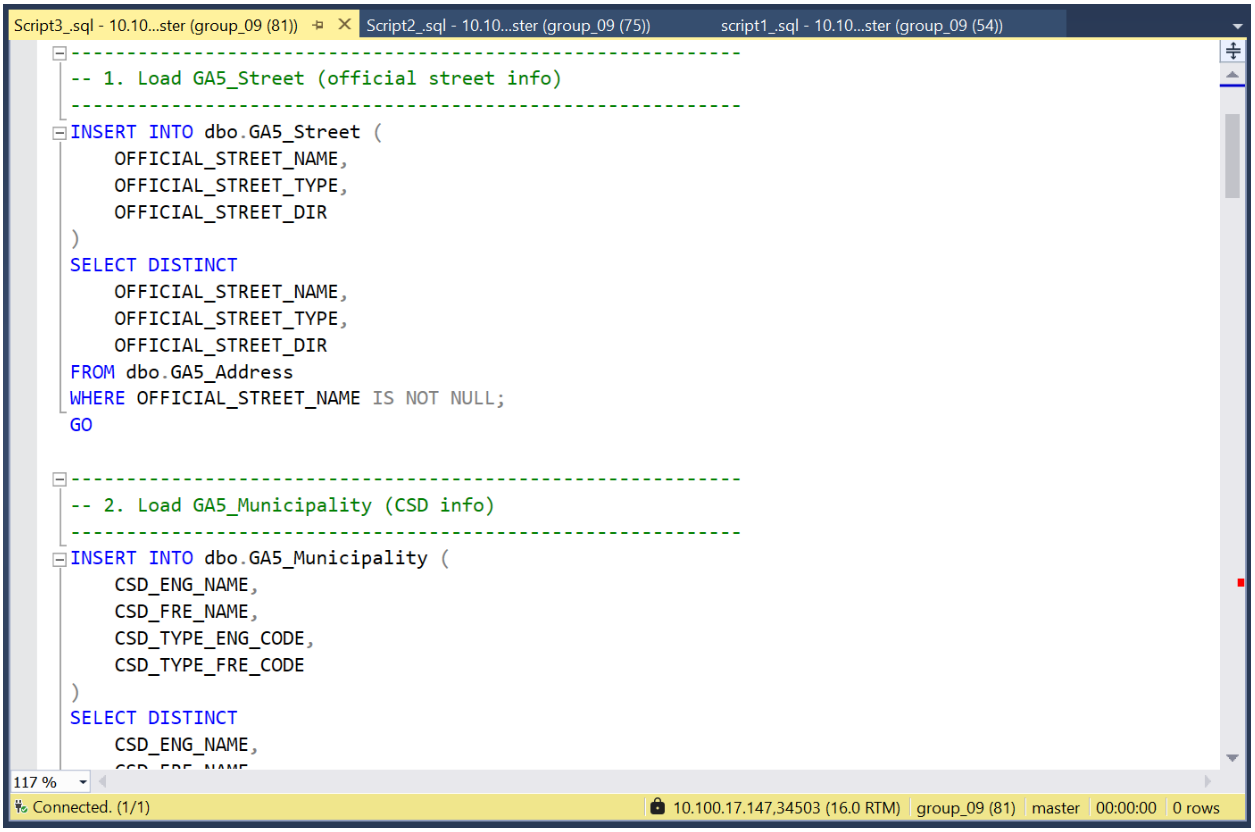
Used BULK INSERT to load every NAR CSV file into the staging table (GA5_Address). This step ensured that the raw files were ingested efficiently and consistently without manual intervention.


Script3_.sql - 10.10...ster (group_09 (81))
Script2_.sql - 10.10...ster (group_09 (75))
script1_.sql - 10.10...ster (group_09 (54))
117 %
Connected. (1/1)
10.100.17.147,34503 (16.0 RTM)
group_09 (75)
master
00:00:00
0 rows

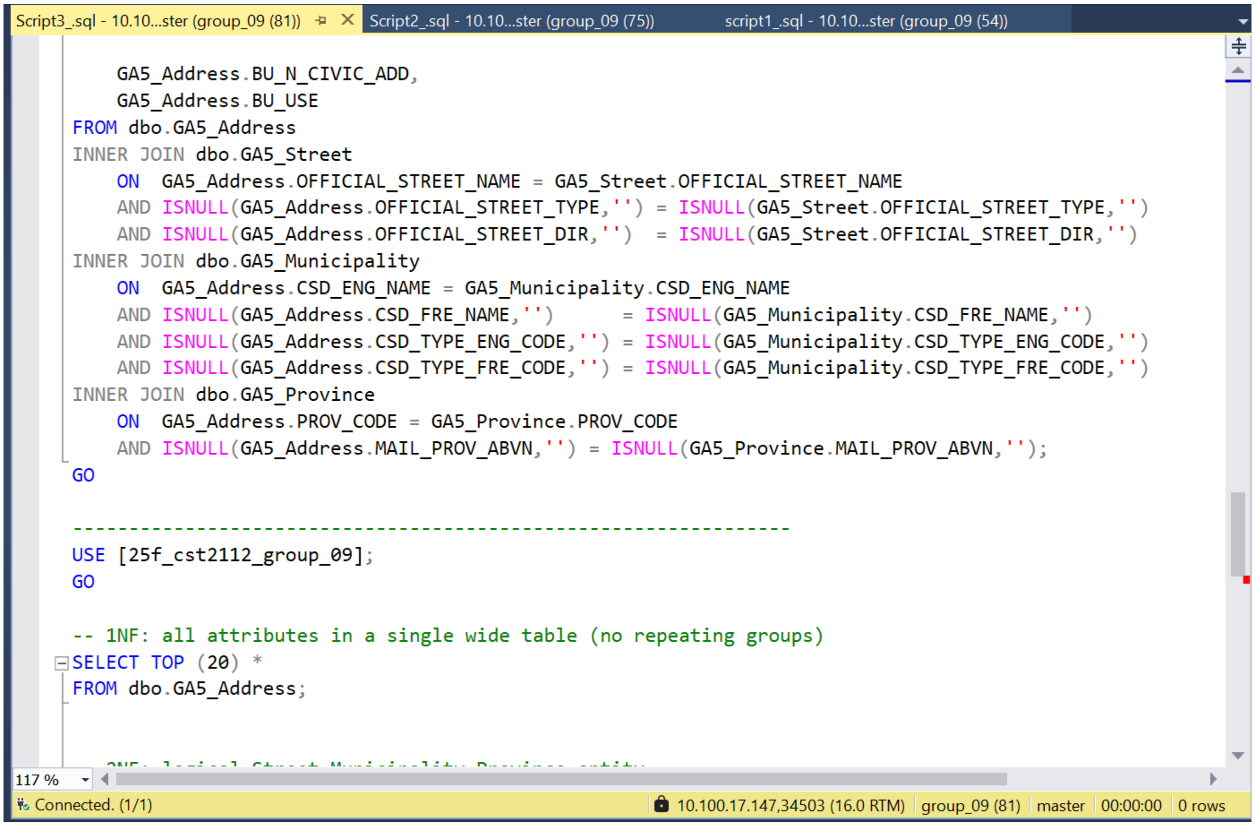
```
-- Address_12.csv
BULK INSERT dbo.GA5_Address
FROM 'c:\cst2112_data\ga5_NAR\Address_12.csv'
WITH (
    FIRSTROW = 2,
    FIELDTERMINATOR = ',',
    ROWTERMINATOR   = '\n',
    TABLOCK
);
GO

-- Address_13.csv
BULK INSERT dbo.GA5_Address
FROM 'c:\cst2112_data\ga5_NAR\Address_13.csv'
WITH (
    FIRSTROW = 2,
    FIELDTERMINATOR = ',',
    ROWTERMINATOR   = '\n',
    TABLOCK
);
GO

------------------------------------------------------------
-- Address_24 parts
------------------------------------------------------------

BULK INSERT dbo.GA5_Address
FROM 'c:\cst2112_data\ga5_NAR\Address_24_part_1.csv'
```

Script 3 – Data Transformation

- Populated lookup tables using INSERT…SELECT DISTINCT.
- Joined the staging data to the lookup entities using exact attribute matching.
- Filled the GA5_AddressNormalized table with foreign key references to each entity.


Script3_.sql - 10.10...ster (group_09 (81))
Script2_.sql - 10.10...ster (group_09 (75))
script1_.sql - 10.10...ster (group_09 (54))
117 %
Connected. (1/1)
10.100.17.147,34503 (16.0 RTM)
group_09 (81)
master
00:00:00
0 rows

```sql
------------------------------------------------------------
-- 1. Load GA5_Street (official street info)
------------------------------------------------------------
INSERT INTO dbo.GA5_Street (
    OFFICIAL_STREET_NAME,
    OFFICIAL_STREET_TYPE,
    OFFICIAL_STREET_DIR
)
SELECT DISTINCT
    OFFICIAL_STREET_NAME,
    OFFICIAL_STREET_TYPE,
    OFFICIAL_STREET_DIR
FROM dbo.GA5_Address
WHERE OFFICIAL_STREET_NAME IS NOT NULL;
GO

------------------------------------------------------------
-- 2. Load GA5_Municipality (CSD info)
------------------------------------------------------------
INSERT INTO dbo.GA5_Municipality (
    CSD_ENG_NAME,
    CSD_FRE_NAME,
    CSD_TYPE_ENG_CODE,
    CSD_TYPE_FRE_CODE
)
SELECT DISTINCT
    CSD_ENG_NAME,
    CSD_FRE_NAME
```


Script3_.sql - 10.10...ster (group_09 (81))
Script2_.sql - 10.10...ster (group_09 (75))
script1_.sql - 10.10...ster (group_09 (54))
117 %
Connected. (1/1)
10.100.17.147,34503 (16.0 RTM)
group_09 (81)
master
00:00:00
0 rows

```sql
    GA5_Address.BU_N_CIVIC_ADD,
    GA5_Address.BU_USE
FROM dbo.GA5_Address
INNER JOIN dbo.GA5_Street
    ON  GA5_Address.OFFICIAL_STREET_NAME = GA5_Street.OFFICIAL_STREET_NAME
    AND ISNULL(GA5_Address.OFFICIAL_STREET_TYPE,'') = ISNULL(GA5_Street.OFFICIAL_STREET_TYPE,'')
    AND ISNULL(GA5_Address.OFFICIAL_STREET_DIR,'')  = ISNULL(GA5_Street.OFFICIAL_STREET_DIR,'')
INNER JOIN dbo.GA5_Municipality
    ON  GA5_Address.CSD_ENG_NAME = GA5_Municipality.CSD_ENG_NAME
    AND ISNULL(GA5_Address.CSD_FRE_NAME,'')      = ISNULL(GA5_Municipality.CSD_FRE_NAME,'')
    AND ISNULL(GA5_Address.CSD_TYPE_ENG_CODE,'') = ISNULL(GA5_Municipality.CSD_TYPE_ENG_CODE,'')
    AND ISNULL(GA5_Address.CSD_TYPE_FRE_CODE,'') = ISNULL(GA5_Municipality.CSD_TYPE_FRE_CODE,'')
INNER JOIN dbo.GA5_Province
    ON  GA5_Address.PROV_CODE = GA5_Province.PROV_CODE
    AND ISNULL(GA5_Address.MAIL_PROV_ABVN,'') = ISNULL(GA5_Province.MAIL_PROV_ABVN,'');
GO

----------------------------------------------------------------
USE [25f_cst2112_group_09];
GO

-- 1NF: all attributes in a single wide table (no repeating groups)
SELECT TOP (20) *
FROM dbo.GA5_Address;


    2NF: logical Street Municipality Province entity
```

5. Challenges and Lessons Learned

Challenge 1 – Extremely Large Data Volume

The dataset contained millions of records. Lookup tables such as GA5_Street and GA5_Geography had tens of thousands to millions of distinct values. Efficient joins and appropriate indexing were required to avoid performance bottlenecks.

Challenge 2 – Matching Lookup Values

Some fields contained NULL values or minor spelling inconsistencies. These were addressed using:

- ISNULL() comparisons during joins
- LEFT JOIN for optional relationships instead of INNER JOIN

Challenge 3 – Normalization vs. Analytical Schemas

Initially, there was uncertainty between designing a 3NF relational model versus a dimensional star schema. After clarifying this doubt during the lecture, the focus was therefore on eliminating redundancy and ensuring functional dependency, not on analytical reporting measures.

Challenge 4 – Foreign Key Load Order

Foreign key dependencies required careful load sequencing. When reloading data, tables were cleared and repopulated in the following order:

1. GA5_AddressNormalized
2. GA5_MailingAddress
3. GA5_Geography
4. GA5_Province
5. GA5_Municipality
6. GA5_Street

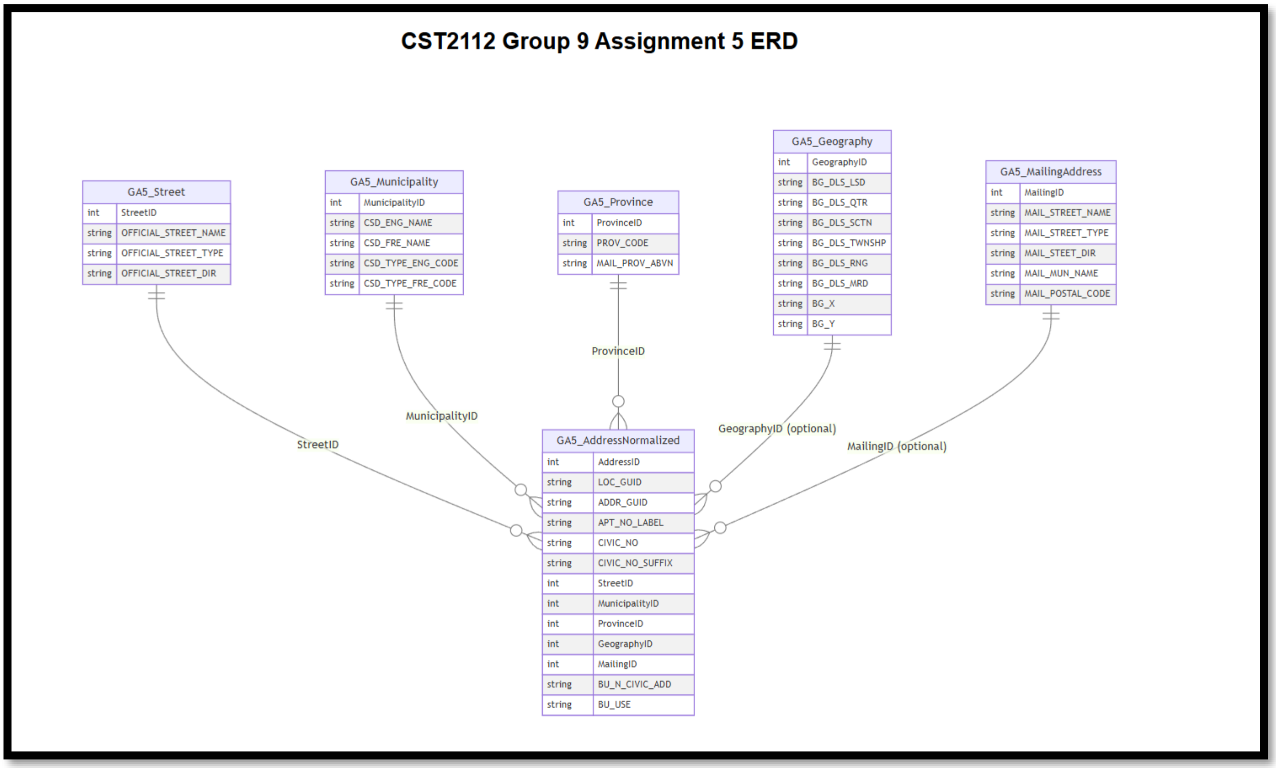
This ordering prevented constraint violations and was incorporated into Script 3.

6. Final ERD

The final ERD (included in the submission) demonstrates:

- Five lookup entities (Street, Municipality, Province, Geography, MailingAddress)
- One central address entity (GA5_AddressNormalized)
- One-to-many relationships from lookup entities to addresses
- Optional relationships for geography and mailing using nullable foreign keys

This accurately represents the final 3NF schema implemented in SQL Server.



GA5_Street	
int	StreetID
string	OFFICIAL_STREET_NAME
string	OFFICIAL_STREET_TYPE
string	OFFICIAL_STREET_DIR

GA5_Municipality	
int	MunicipalityID
string	CSD_ENG_NAME
string	CSD_FRE_NAME
string	CSD_TYPE_ENG_CODE
string	CSD_TYPE_FRE_CODE

GA5_Province	
int	ProvinceID
string	PROV_CODE
string	MAIL_PROV_ABVN

GA5_Geography	
int	GeographyID
string	BG_DLS_LSD
string	BG_DLS_QTR
string	BG_DLS_SCTN
string	BG_DLS_TWNSHP
string	BG_DLS_RNG
string	BG_DLS_MRD
string	BG_X
string	BG_Y

GA5_MailingAddress	
int	MailingID
string	MAIL_STREET_NAME
string	MAIL_STREET_TYPE
string	MAIL_STEET_DIR
string	MAIL_MUN_NAME
string	MAIL_POSTAL_CODE

GA5_AddressNormalized	
int	AddressID
string	LOC_GUID
string	ADDR_GUID
string	APT_NO_LABEL
string	CIVIC_NO
string	CIVIC_NO_SUFFIX
int	StreetID
int	MunicipalityID
int	ProvinceID
int	GeographyID
int	MailingID
string	BU_N_CIVIC_ADD
string	BU_USE

7. Conclusion

The normalization process transformed a duplicated, semi-structured CSV dataset into a clean relational database that supports reliable querying and avoids update anomalies.

Key outcomes included:

- Identifying real-world entities and functional dependencies
- Separating repeating attributes into lookup tables

- Enforcing referential integrity through foreign keys
- Automating creation, loading, and population using T-SQL
- Producing a fully normalized 3NF structure consistent with address data management principles

This project strengthened my understanding of database normalization, staging pipelines, and T-SQL scripting, and provided practical experience working with large-scale government datasets.

.