Samantha Traversi
MDS 523
Dr. Dean Jensen
Data Warehousing Final Project

Table of Contents

Section 1 Requirements

Organization Description

Manos con Mayas is a fictional NGO whose mission is to provide preventative education, vaccinations, and treatment for Dengue, a vector-borne illness with a high prevalence in Central America. This NGO has partnered with four rural hospitals in Honduras and Nicaragua that serve residents of the surrounding area. The NGO itself provides funding specifically directed at reducing the harmful physical and socio-economic effects of Dengue with a three-pronged approach of prevention, education, and treatment. After a period of unstable funding resulting in personnel changes, Manos con Mayas has a refreshed focus to monitor and track educational initiatives and patients, improve reporting in order to present donors and investors with a plan that will allow for more funding for preventative measures and treatment in these specific hospitals.

Organization Problems

- The hospital records and public record of tracking of health and risks to the population due to these mosquito borne illnesses are unreliable and likely inaccurate, so the members of this NGO don't have a clear idea of the scope of their need.
- The individuals in the communities surrounding the hospital are largely misinformed about how to identify symptoms of mosquito-borne illnesses, so they often go untreated until there are severe health concerns, or the infected individuals meet their death.
- Administering widespread preventative health education and vaccinations has been difficult due to social, economic, and geographic barriers.
- Shifting management within the upper levels of the NGO has led to unstable direction at these hospitals and currently no projects have yielded measurable results.

Organization Requirements

Operational System

1. The system must allow for new people to be registered with their sixteen-digit government identification number.
2. Only one person can register as themselves and they can hold only one account.
3. The system must link family members who are allowed to receive medical information on behalf of the patient by adding their registration number and four-digit family code
4. The system must allow for updates to NULL date fields in the number of vaccinations.
5. The system must be able to create a view of people who received a vaccination before and after the six month period (meaning, view for those who have a vaccination overdue and a vaccination upcoming)
6. The system must be able to provide a list of individuals who completed one of the preventative education sessions in the last year.
7. The system must allow for updates to the date fields and the grade fields in Preventative Education.

8. The system must be able to provide a list of cases that had the most severe symptoms, including Severe Dengue and Death.
9. The system must be able to provide a list of hospitalizations due to Dengue.

Data Warehouse

1. Must have the ability to track administered vaccinations during a certain calendar period.
2. Must have the ability to track preventative education administered during a certain calendar period.
3. Must have the ability to provide a snapshot of suspected rate of infection for each illness.
4. Must be able to determine if there is a decrease in infections after a period of increased preventative education or vaccination administration.

Technical Requirements

· SQL Server 2017
· Visio 2017
· Visual Studio 2017

Data Requirements

Operational Database

Person Information

- The Person ID must be 16 characters in length, may not include letters or special characters
- The Family ID must be 16 characters in length, may not include letters or special characters
- The Relationship is a four-digit alphanumeric code: “Sp01” for Spouse 1, “Pa01” for Parent 1, “Sb01” for Sibling 1, and “Ch01” for Child 1. As the number increases for the number of siblings or children, etc., the number in the code increases likewise

Vaccination Information

- The Vaccination ID is a four-digit code of numbers, no special characters or letters, that tracks the number of vaccinations administered. This number is only applied upon the first vaccination and must not be changed
- The dates of all three vaccinations may be null, but cannot be changed once filled in

Preventative Education Information

- The Student ID is an auto-generated numeric code
- The date of completion field for both courses is in date format, it may be updated if the person retakes the course
- The grade field for both courses is in numeric format, it may be updated if the person retakes the course

Cases Information

- The Case number is numerical in the order that the case is reported
- The Date of first symptoms is a date for when the symptoms first presented, not when the report was made

Symptoms Information

- The Symptom Code is a three digit letter code that represents the symptoms presented when an individual contracts Dengue
- The Symptom Description should be a shorthand description which may include numbers or special characters

Treatments Information

- The Treatment Code is a three digit letter code that represents the standard levels of treatment for Dengue
- The Treatment Description should be a shorthand description which may include numbers or special characters

Data Warehouse

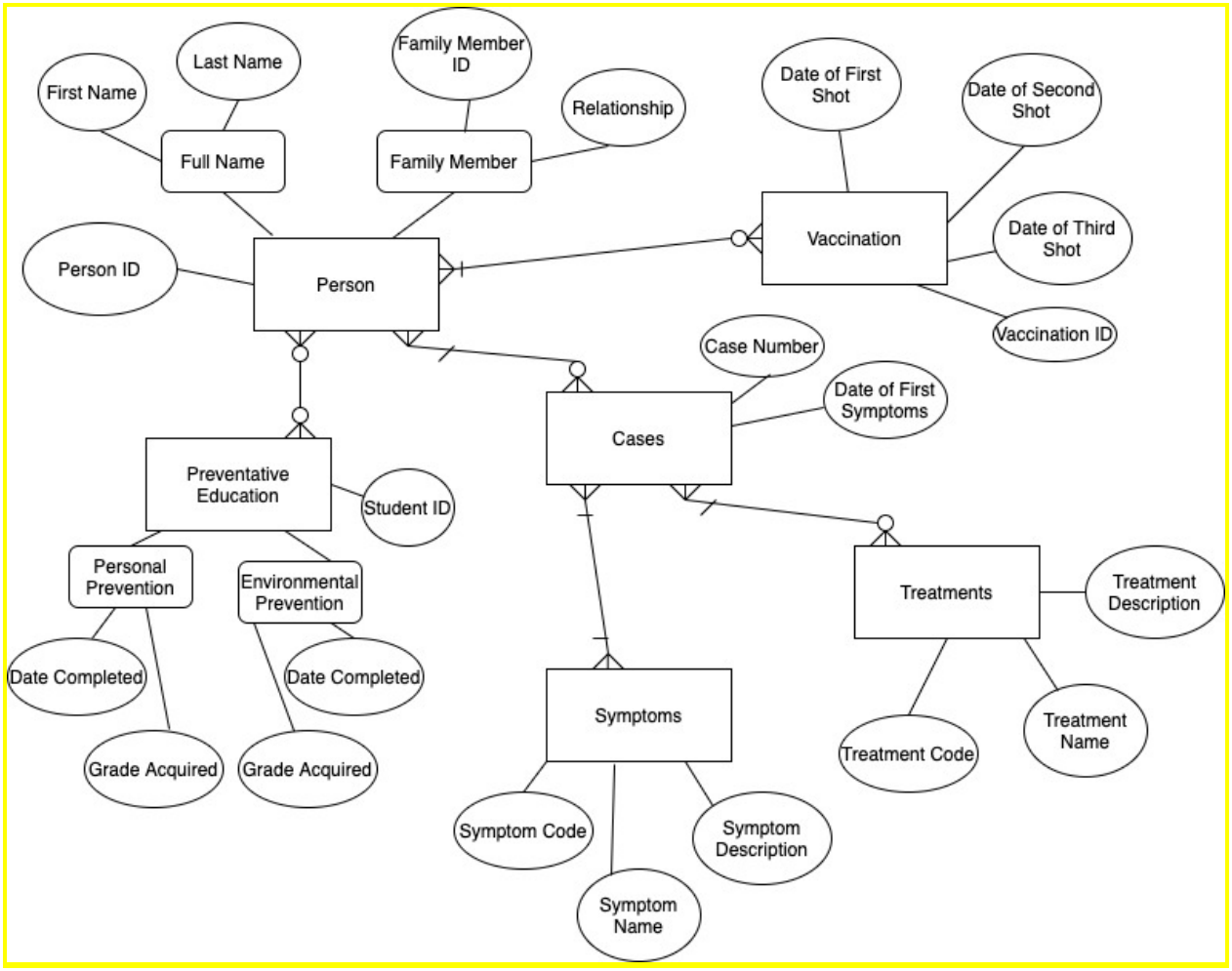
- Person should be the center fact table
- Cases, symptoms, and treatments should be included in one dimension table at atomic granularity
- Preventative Education should include a record of all logged dates of completion and grades acquired

Organization Rules

A person can have 0, 1, or many cases.
A case must have at least one or many persons associated with it.
A case must have at least one symptom.
A symptom must have at least one case.
A case can have 0, 1, or many treatments.
A treatment can have 0, 1, or many cases.
A person can have 0, 1, or many vaccinations.
A vaccination must have at least one or many persons associated with it.
A person can have 0, 1, or many preventative education instances.
A preventative education can have 0, 1, or many persons associated with it.

Section 2 Data Models

Conceptual Data Model



Conceptual Data Model Entity and Attribute Descriptions

There are six entities with several attributes per entity in this data model.

1. Person - This entity contains information about the individual who has been targeted for preventative education, started a vaccination series, or has been infected with Dengue.
    1.1. Person ID - Both the Honduran and Nicaraguan governments use a sixteen-digit numerical identification code which is unique. These serve as the primary key for the individuals.
    1.2. Full Name - This is the individual's first and last names combined.
    1.3. Family Members - These are members of the family who are allowed to receive medical information about the individual, and who benefit from an individual receiving preventative education.
2. Cases - This is a record of the times an individual has been infected with Dengue. There are four strains of Dengue, so an individual can be infected four separate times. After infection with one strain, the individual is immune to only that particular strain.

    2.1. Start Date - The date that the individual first noticed symptoms of Dengue.
3. Symptoms - These are the symptoms of Dengue that a person has. They could have one, several, or all symptoms.
    3.1. Symptom Code - This is a three-letter code that represents a symptom.
    3.2. Symptom Description - This is a short description of the symptoms common to Dengue infections.
4. Treatments - These are the treatments that a person infected with Dengue has received.
    4.1. Treatment Code - This is a three-letter code that represents a treatment.
    4.2. Treatment Description - This is a short description of the treatment provided to the person infected with Dengue.
5. Vaccinations - The vaccination for Dengue requires three separate doses, each administered six months apart.
    5.1. Date of First Shot - This is the date of the first dose.
    5.2. Date of Second Shot - This is the date of the second dose.
    5.3. Date of Third Shot - This is the date of the third dose.
6. Preventative Education - This information tracks preventative education administered to the individual.
    6.1. Personal Prevention - This is a course dedicated to personal prevention methods for dengue, it is split out into the date of completion and the grade acquired.
    6.2. Environmental Prevention - This is a course dedicated to environmental prevention methods for dengue, it is split out into the date of completion and the grade acquired.

Logical Data Model

	Person
PK	Person ID
	First Name
	Last Name
	Family ID
	Family Relationship

	Cases
PK	Case Number
FK	Person ID
	Date of Start
	Treatment Code
	Symptom Code

	Preventative Education
PK	Student Number
FK	Person ID
	Personal Prevention Date
	Personal Prevention Grade
	Environmental Prevention Date
	Environmental Prevention Grade

	Symptoms
PK	Symptom Code
	Symptom Name
	Symptom Description

	Vaccinations
PK	Vaccination Number
FK	Person ID
	Date First Shot
	Date Second Shot
	Date Third Shot

	Treatments
PK	Treatment Code
	Treatment Name
	Treatment Description

Logical Data Model Entity and Attribute Descriptions
The above Logical Diagram contains the same entities and attributes as the Conceptual Data Model with a few exceptions:

1. Composite Customer Name has been separated out into its components
2. Composite Family Member has been separated out into its components
3. Composite Preventative Education courses have been separated out into their components

Section 3 Physical Database Design

Tables Converted to Normal Form

- Entity and Attribute names have spaces removed and are shortened in some cases
- To note denormalization, original names are described first with final names matched with a hyphen
- Keys are denoted Primary Key (PK) or Foreign Key (FK)
- Each table in normal form is below the description

Entity/Attribute Name Updates

- Person - Person
    - Person ID - PersonID (PK)
    - First Name - FirstName
    - Last Name - LastName
    - Family Member ID - FamilyID
    - Family Relationship - FamilyRelationship

PersonID	FirstName	LastName	FamilyID	FamilyRelationship

- Vaccination - Vaccination
    - Vaccination ID - VaccinationID (PK)
    - Person ID - PersonID (FK)
    - Date of First Shot - DateFirstShot
    - Date of Second Shot - DateSecondShot
    - Date of Third Shot - DateThirdShot

VaccinationID	PersonID	DateFirstShot	DateSecondShot	DateThirdShot

- Preventative Education - PreventativeEducation
    - Student ID - StudentID (PK)
    - Person ID - PersonID (FK)
    - Personal Prevention Completion Date - PersonalPreventionDate
    - Personal Prevention Completion Grade - PersonalPreventionGrade
    - Environmental Prevention Completion Date - EnvironmentalPreventionDate
    - Environmental Prevention Completion Grade - EnvironmentalPreventionGrade

StudentID	PersonID	PersonalPrevention Date	PersonalPrevention Grade	EnvironmentalPrevention Date	EnvironmentalPrev entionGrade

- Cases - Cases
    - Case Number - CaseNumber (PK)
    - Person ID - PersonID (FK)
    - Date of Start of Symptoms - DateofStart
    - Symptom Code - SymptomCode
    - Treatment Code - TreatmentCode

CaseNumber	PersonID	DateofStart	SymptomCode	TreatmentCode

- Symptoms - Symptoms
    - Symptom Code - SymptomCode (PK)
    - Symptom Name - SymptomName
    - Symptom Description - SymptomDescription

SymptomCode	SymptomName	SymptomDescription

- Treatments - Treatments
    - Treatment Code - TreatmentCode (PK)
    - Treatment Name - TreatmentName
    - Treatment Description - TreatmentDescription

TreatmentCode	TreatmentName	TreatmentDescription

Normalized Tables with Imaginary Sample Data

Person

PersonID	FirstName	LastName	FamilyID	FamilyRelationship
0000111122223333	Bertha	Lempira	1111222233334444	Sp01
1111222233334444	Gregory	Lempira	0000111122223333	Sp01
2222333344445555	Estefany	Suyapa	3333444455556666	Ch01
3333444455556666	Paolo	Suyapa	2222333344445555	Pa01
4444555566667777	Geovanna	Suyapa	2222333344445555	Pa02
5555666677778888	Alejandro	Silva	NULL	NULL
6666777788889999	Pablo	Ernesto	NULL	NULL

Vaccination

VaccinationID	PersonID	DateFirstShot	DateSecondShot	DateThirdShot
1001	6666777788889999	2020-04-18	2020-10-22	2021-04-25
1002	2222333344445555	2020-10-12	2021-04-10	NULL
1003	3333444455556666	2020-10-12	2021-04-10	NULL
1004	4444555566667777	2020-10-12	2021-04-10	NULL
1005	0000111122223333	2021-05-21	NULL	NULL
1006	1111222233334444	2021-05-21	NULL	NULL

Preventative Education

StudentID	PersonID	PersonalPrevention Date	PersonalPrevention Grade	EnvironmentalPrevention Date	EnvironmentalPreventionGrade
2001	666677778 8889999	NULL	NULL	2021-01-28	80
2002	111122223 3334444	2021-02-13	95	NULL	NULL
2003	333344445 5556666	2021-02-13	88	NULL	NULL
2004	222233334 4445555	2021-02-13	92	NULL	NULL

Cases

CaseNumber	PersonID	DateofStart	SymptomCode	TreatmentCode
3001	2222333344445555	2020-11-19	SEV	HOS
3002	1111222233334444	2020-11-28	ACH	NULL
3003	5555666677778888	2020-12-02	SEV	HOS
3004	3333444455556666	2021-03-08	NAU	RNF

Symptoms

SymptomCode	SymptomName	SymptomDescription
NAU	Nausea	May or may not include vomiting
RAS	Rash	NULL
ACH	Aches	Body aches or behind eyes
SEV	SevereDengue	Loss of Conscious or high fever
DTH	Death	NULL

Treatments

TreatmentCode	TreatmentName	TreatmentDescription
PCM	Paracetamol	Also Acetaminophen
RNF	Rest and Fluids	NULL
HOS	Hospitalization	Intake for treatment

SQL Code to Create Tables in SQL Server

```
CREATE TABLE Person
 (PersonID CHAR (16) NOT NULL,
 FirstName VARCHAR (25) NOT NULL,
 LastName VARCHAR (25) NOT NULL,
 FamilyID CHAR (16),
 FamilyRelationship CHAR (4),
 PRIMARY KEY (PersonID)
 );


CREATE TABLE Vaccination
 (VaccinationID INT NOT NULL,
 PersonID CHAR (16) NOT NULL,
 DateFirstShot DATE,
 DateSecondShot DATE,
 DateThirdShot DATE,
 PRIMARY KEY (VaccinationID),
 FOREIGN KEY (PersonID) REFERENCES Person(PersonID)
 );


CREATE TABLE PreventativeEducation
 (StudentID INT NOT NULL,
 PersonID CHAR (16) NOT NULL,
 PersonalPreventionDate DATE,
 PersonalPreventionGrade NUMERIC (3,1),
 EnvironmentalPreventionDate DATE,
 EnvironmentalPreventionGrade NUMERIC (3,1),
 PRIMARY KEY (StudentID),
 FOREIGN KEY (PersonID) REFERENCES Person(PersonID)
 );


CREATE TABLE Cases
 (CaseNumber INT NOT NULL,
 PersonID CHAR (16) NOT NULL,
 DateofStart DATE  NOT NULL,
 SymptomCode CHAR (3) NOT NULL,
 TreatmentCode CHAR (3),
 PRIMARY KEY (CaseNumber),
```

```
 FOREIGN KEY (PersonID) REFERENCES Person(PersonID)
 );


CREATE TABLE Symptoms
 (SymptomCode CHAR (3) NOT NULL,
 SymptomName VARCHAR (3),
 SymptomDescription VARCHAR (3),
 PRIMARY KEY (SymptomCode),
 );


CREATE TABLE Treatment
 (TreatmentCode CHAR (3) NOT NULL,
 TreatmentName VARCHAR (3),
 TreatmentDescription VARCHAR (3),
 PRIMARY KEY (TreatmentCode),
 );
```

SQL Code to Load Tables in SQL Server

```
INSERT INTO Person VALUES ('0000111122223333', 'Bertha', 'Lempira',
'1111222233334444', 'SP01');
INSERT INTO Person VALUES ('1111222233334444', 'Gregory', 'Lempira',
'0000111122223333', 'SP01');
INSERT INTO Person VALUES ('2222333344445555', 'Estefany', 'Suyapa',
'3333444455556666', 'CH01');
INSERT INTO Person VALUES ('3333444455556666', 'Paolo', 'Suyapa', '2222333344445555',
'PA01');
INSERT INTO Person VALUES ('4444555566667777', 'Geovanna', 'Suyapa',
'2222333344445555', 'PA02');
INSERT INTO Person VALUES ('5555666677778888', 'Alejandro', 'Silva', NULL, NULL);
INSERT INTO Person VALUES ('6666777788889999', 'Pablo', 'Ernesto', NULL, NULL);


INSERT INTO Vaccination VALUES (1005, '0000111122223333', '2021-05-21', NULL, NULL);
INSERT INTO Vaccination VALUES (1006, '1111222233334444', '2021-05-21', NULL, NULL);
INSERT INTO Vaccination VALUES (1002, '2222333344445555', '2020-10-12', '2021-04-10',
NULL);
INSERT INTO Vaccination VALUES (1003, '3333444455556666', '2020-10-12', '2021-04-10',
NULL);
INSERT INTO Vaccination VALUES (1004, '4444555566667777', '2020-10-12', '2021-04-10',
NULL);
INSERT INTO Vaccination VALUES (1001, '6666777788889999', '2020-04-18', '2020-10-22',
'2021-04-25');


INSERT INTO PreventativeEducation VALUES (2001, '6666777788889999', NULL, NULL,
'2021-01-28', 80);
INSERT INTO PreventativeEducation VALUES (2002, '1111222233334444', '2021-02-13', 95,
NULL, NULL);
INSERT INTO PreventativeEducation VALUES (2003, '3333444455556666', '2021-02-13', 88,
NULL, NULL);
INSERT INTO PreventativeEducation VALUES (2004, '2222333344445555', '2021-02-13', 92,
NULL, NULL);
```

```
INSERT INTO Cases VALUES (3001, '2222333344445555', '2020-11-19', 'SEV', 'HOS');
INSERT INTO Cases VALUES (3002, '1111222233334444', '2020-11-28', 'ACH', NULL);
INSERT INTO Cases VALUES (3003, '5555666677778888', '2020-12-02', 'SEV', 'HOS');
INSERT INTO Cases VALUES (3004, '3333444455556666', '2021-03-08', 'NAU', 'RNF');
```

```
INSERT INTO Treatment VALUES ('PCM', 'Paracetamol', 'Also acetaminophen');
INSERT INTO Treatment VALUES ('RNF', 'Rest and Fluids', NULL);
INSERT INTO Treatment VALUES ('HOS', 'Hospitalization', 'Intake for treatment');
```

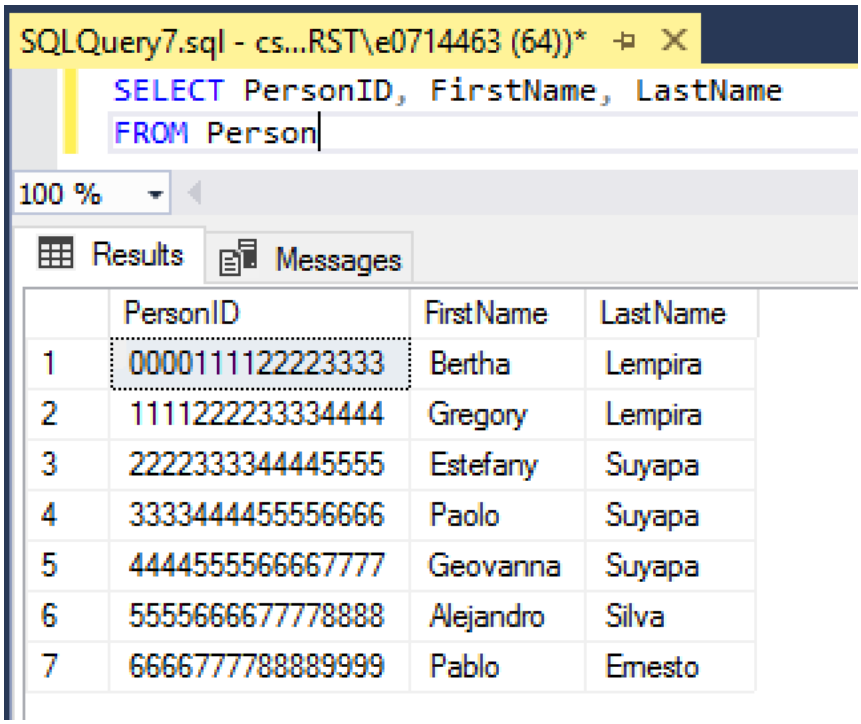
Section 4 Queries

Simple Queries for Select Organization Requirements

Organization Requirement #1 - Check for full name against Person ID

```
SELECT PersonID, FirstName, LastName
FROM Person
```

SQL Code with Results

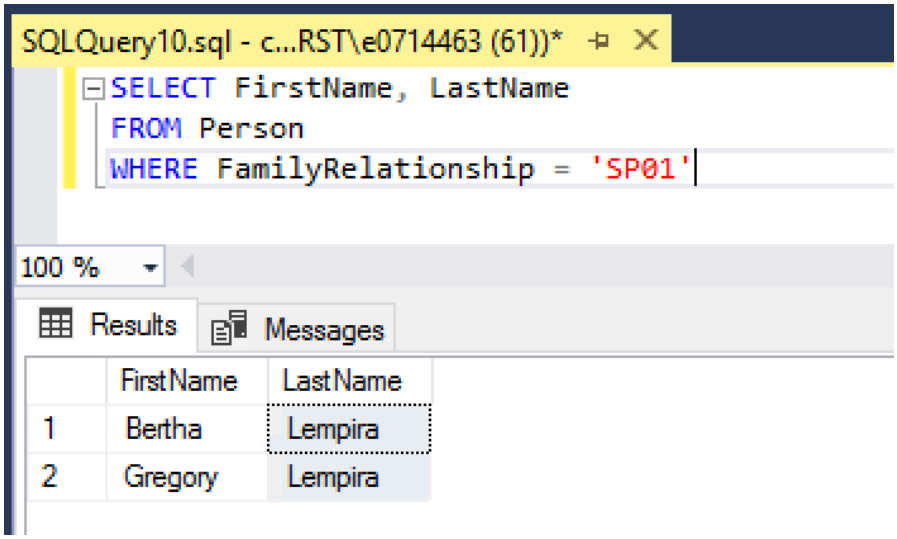


	PersonID	FirstName	LastName
1	0000111122223333	Bertha	Lempira
2	1111222233334444	Gregory	Lempira
3	2222333344445555	Estefany	Suyapa
4	3333444455556666	Paolo	Suyapa
5	4444555566667777	Geovanna	Suyapa
6	5555666677778888	Alejandro	Silva
7	6666777788889999	Pablo	Ernesto

Organization Requirement #3 - Look up family members who are linked to a person

```
SELECT FirstName, LastName
FROM Person
WHERE FamilyRelationship = 'SP01'
```

SQL Code with Results

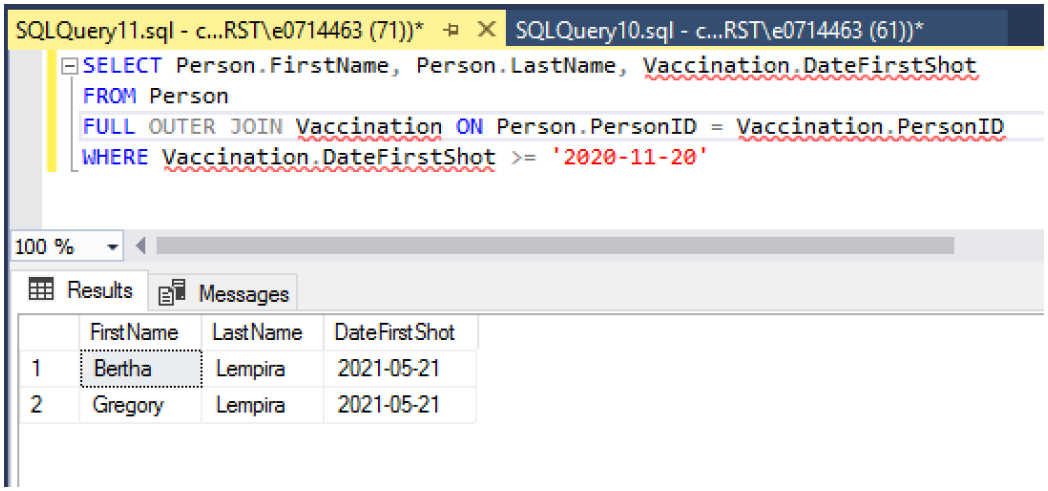


Complex Queries for Select Organization Requirements

Organization Requirement #5 - Provide a list of people who received a first vaccination less than six months in the past

```
SELECT Person.FirstName, Person.LastName, Vaccination.DateFirstShot
FROM Person
FULL OUTER JOIN Vaccination ON Person.PersonID = Vaccination.PersonID
WHERE Vaccination.DateFirstShot >= '2020-11-20'
```

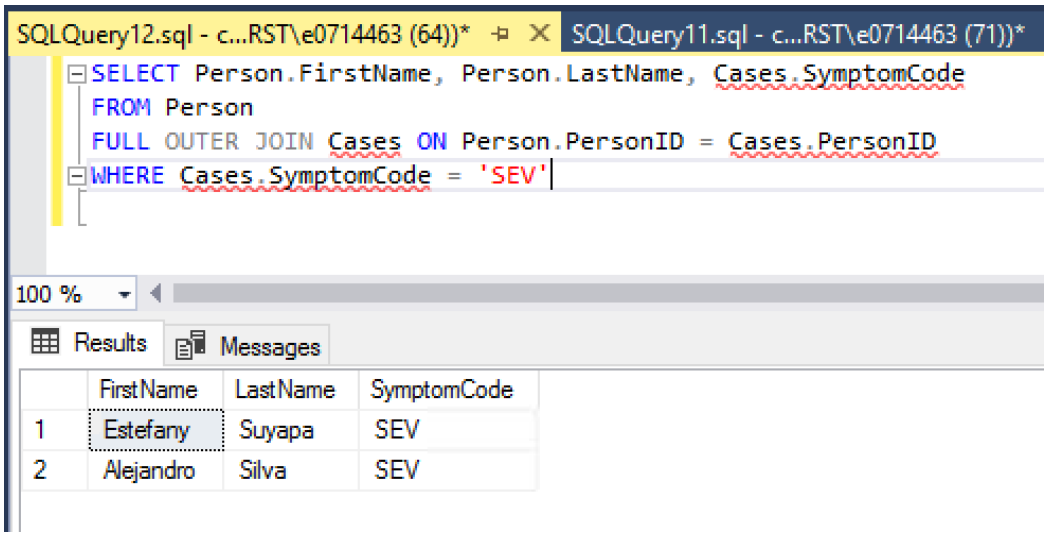
SQL Code with Results



Organization Requirement #8 - Provide a list of cases that had Severe Dengue

```
SELECT Person.FirstName, Person.LastName, Cases.SymptomCode
FROM Person
FULL OUTER JOIN Cases ON Person.PersonID = Cases.PersonID
WHERE Cases.SymptomCode = 'SEV'
```

SQL Code with Results



Section 5 Data Warehouse Design and Coding

Data Mart

For this particular organization, the cost and labor that would go into developing a data mart is not worth it. An NGO is not likely to even have a Data Warehouse, and if it weren't for the requirement within this project, the development of the Data Warehouse is unnecessary. It would be possible for all four hospitals to use the same database to log all the individuals who have received services or care.

Data Warehouse Design

- The Organization will need to pull snapshots of Person Information, as well as the information about Cases and preventative measures, Education and Vaccinations, over time to make predictive analysis
- Cases, symptoms, and treatments have been merged into one dimension table, "InfectionCases" since the repeated information about treatment and symptoms won't be a storage concern
- Person Information is a Dimension Table
- Preventative measures of education and vaccinations are Fact tables.
- The dimensional model below illustrates these relationships

Data Warehouse Dimensional Model

	InfectionCasesDimension
PK	CaseNumber
FK	PersonID
	DateofStart
	SymptomName
	SymptomDescription
	TreatmentName
	TreatmentDescription
	EducationFacts
PK	StudentID
FK	PersonID
	PersonalPreventionDate
	EnvironmentalPreventionDate

	PersonDimension
PK	PersonID
	FirstName
	LastName
	FamilyID
	FamilyRelationship
	VaccinationFacts
PK	VaccinationID
FK	PersonID
	DateFirstShot
	DateSecondShot
	DateThirdShot

Data Warehouse Coding

PersonDimension

```
SELECT PERSON.PersonID, Person.FirstName, Person.LastName,
Person.FamilyID, Person.FamilyRelationship
FROM ManosconMayas.dbo.Person
```

InfectionCasesDimension

```
SELECT Case.CaseNumber, Case.PersonID, Case.SymptomCode,
Case.TreatmentCode, Symptom.SymptomCode, Symptom.SymptomName,
Symptom.SymptomDescription, Treatment.TreatmentCode,
Treatment.TreatmentName, Treatment.TreatmentDescription
FROM ManosconMayas.dbo.Case
full join ManosconMayas.dbo.Symptom ON Case.SymptomCode =
Symptom.SymptomCode full join ManosconMaya.dbo.Case ON
Case.TreatmentCode = Treatment.TreatmentCode
```

PreventativeEducationFacts

```
SELECT PreventativeEducation.StudentID, PreventativeEducation.PersonID,
PreventativeEducation.PersonalPreventionDate ,
PreventativeEducation.EnvironmentalPreventionDate
INTO PreventativeEducationFacts
FROM ManosconMayas.dbo.PreventativeEducation
```

VaccinationFacts

```
SELECT Vaccination.VaccinationID, Vaccination.PersonID, Vaccination.DateFirstShot,
Vaccination.DateSecondShot, Vaccination.DateThirdShot
INTO VaccinationFacts
FROM ManosconMayas.dbo.Vaccination
```