



WESTERN UNIVERSITY

# DATA DRIVEN ANALYSIS FOR OPTIMIZATION OF HOSPITAL OPERATIONS

PRESENTED BY  
GROUP 14





# Agenda

---

INTRODUCTION

DIMENSIONAL MODEL

---

DATASET DESCRIPTION

ETL PROCESS

---

PROJECT WORKFLOW

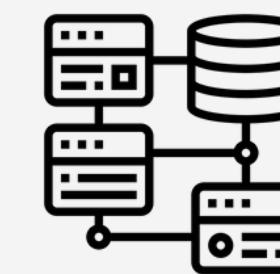
MULTI-DIMENSIONAL CUBE

---

RELATIONAL DATABASE

REFERENCES

# INTRODUCTION



Build relational database model



Create dimensional model tables



Use ETL for data loading



Build multi-dimensional cube



Analyze data with MDX using cube

# Business Requirements

The requirements we will analyze from this data warehousing project



## Patient Related

Measure the total time spent  
by a patient in the hospital



## Nurse Related

Measure the number of patients  
handled by a nurse in the hospital

# Requirements Analysis



## Analysis # 1

### Staffing Decisions

Accurately hire new people to avoid procedure delays.

## Analysis # 2

### Resource Allocation

Deploy under-utilized nurses to appropriate departments

## Analysis # 3

### Promotional Analysis

Accordingly promote nurses to senior positions.

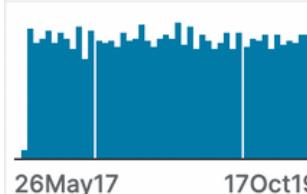
# Dataset Description

**Significant Well Being Dataset- A dataset about tracking patient progress using intake management system**

**well\_being\_dataset.csv** (24.05 MB)

Detail   Compact   Column

27 of 27 columns ▾

| CompletedDateTi...  | # TimeTo_Completed | Patient                    | Priority   | Referral_Category                             |
|---|--------------------|----------------------------|--|---|
|  | 0 170k             | New<br>Active<br>Other (1) | 63% Time Sensitive<br>37% Urgent<br>0% Other (18774) | IV 22%<br>Wound Care 17%<br>Other (63844) 60% |
| 2017-12-19<br>13:24:47.000  | 168                | Active                     | Time Sensitive                                       | Palliative                                    |
| 2017-12-19<br>11:27:10.000  | 47                 | New                        | Time Sensitive                                       | Special Case                                  |
| 2017-12-19<br>13:24:42.000  | 163                | Active                     | Time Sensitive                                       | IV  |
| 2017-12-19<br>11:38:11.000  | 54                 | New                        | Urgent   | Special Case                                  |
| 2017-12-28<br>13:55:55.000  | 13146              | New                        | Medium   | LTC Placement                                 |
| 2017-12-19<br>15:27:43.000  | 254                | New                        | Urgent   | Post Ortho                                    |
| 2017-12-19<br>12:32:07.000  | 77                 | Active                     | Time Sensitive                                       | IV  |
| 2017-12-19<br>14:35:10.000  | 54                 | New                        | TBD  | Special Case                                  |

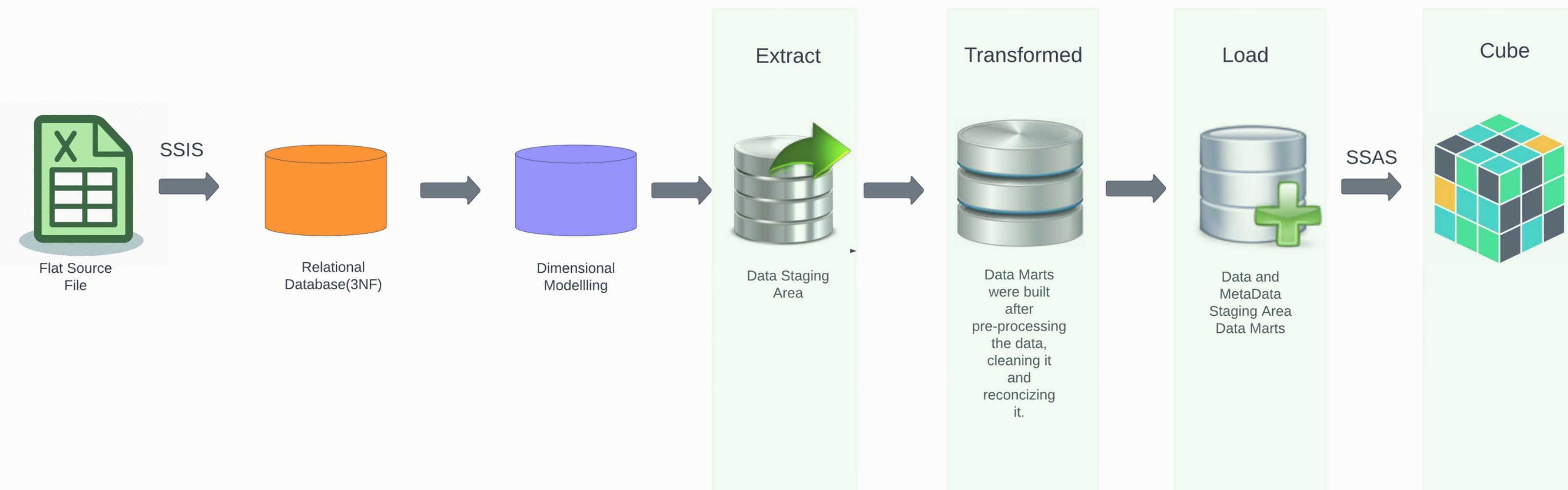
Source:  
**Kaggle.com**

Number of Records:  
**10,000**

Number of Attributes  
**27**

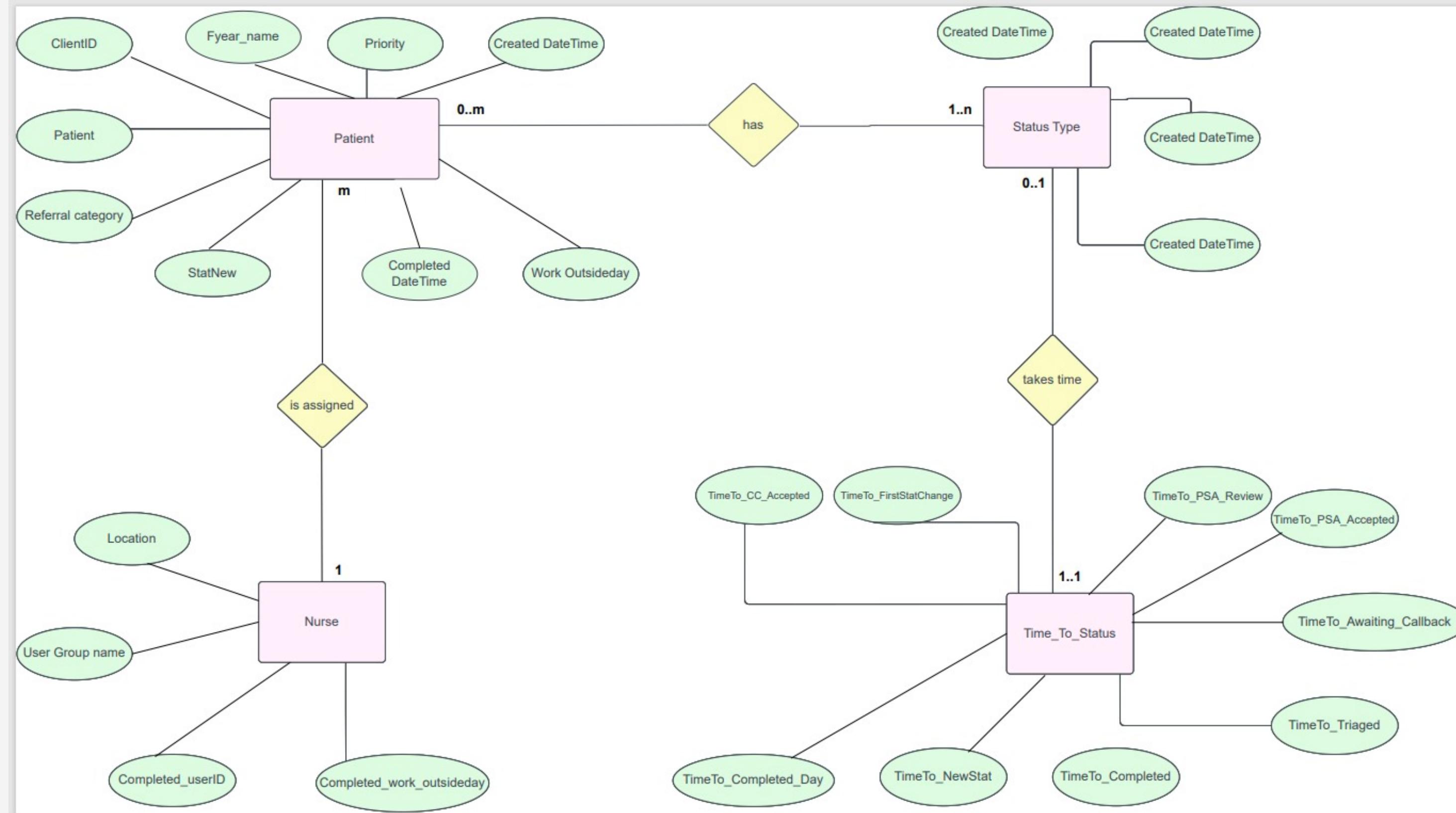
# Date Warehousing Workflow

In order to meet the business requirements for the data warehousing project, this overflow was followed



# Relational Database Model

To reduce duplication and increase integrity for the Referral Intake Management System, a non-relational dataset was reorganized into relational database tables.



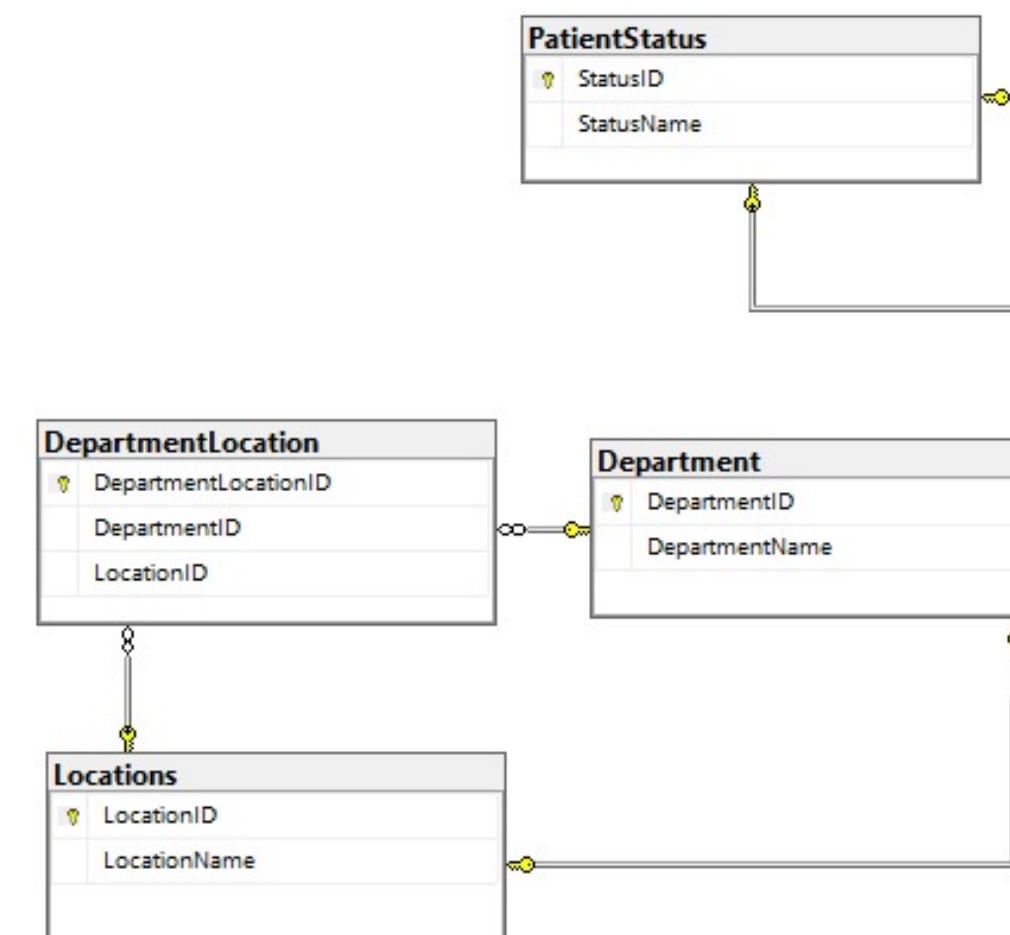
Entity Relationship Model

# Relational Database Model

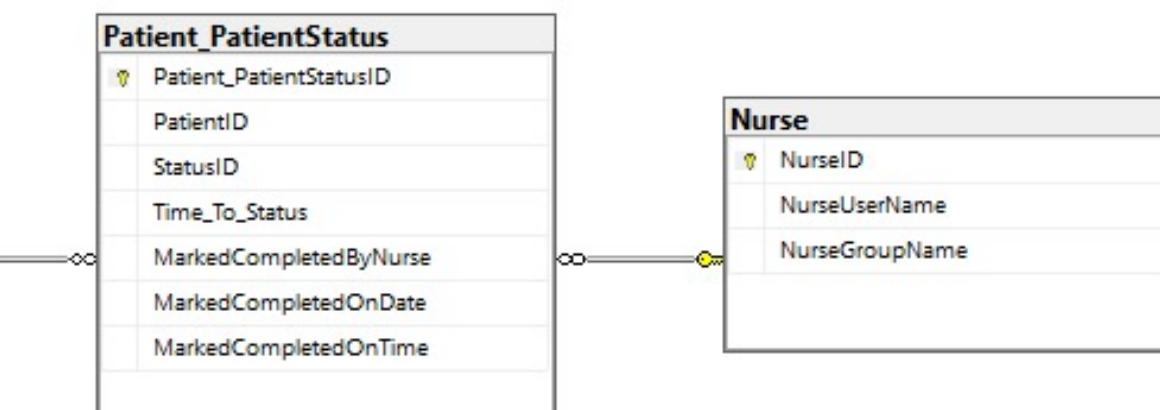
By removing anomalies and breaking interdependent connections into separate entities for better data organisation, the normalisation method was used to turn the main flat file into a 3NF.



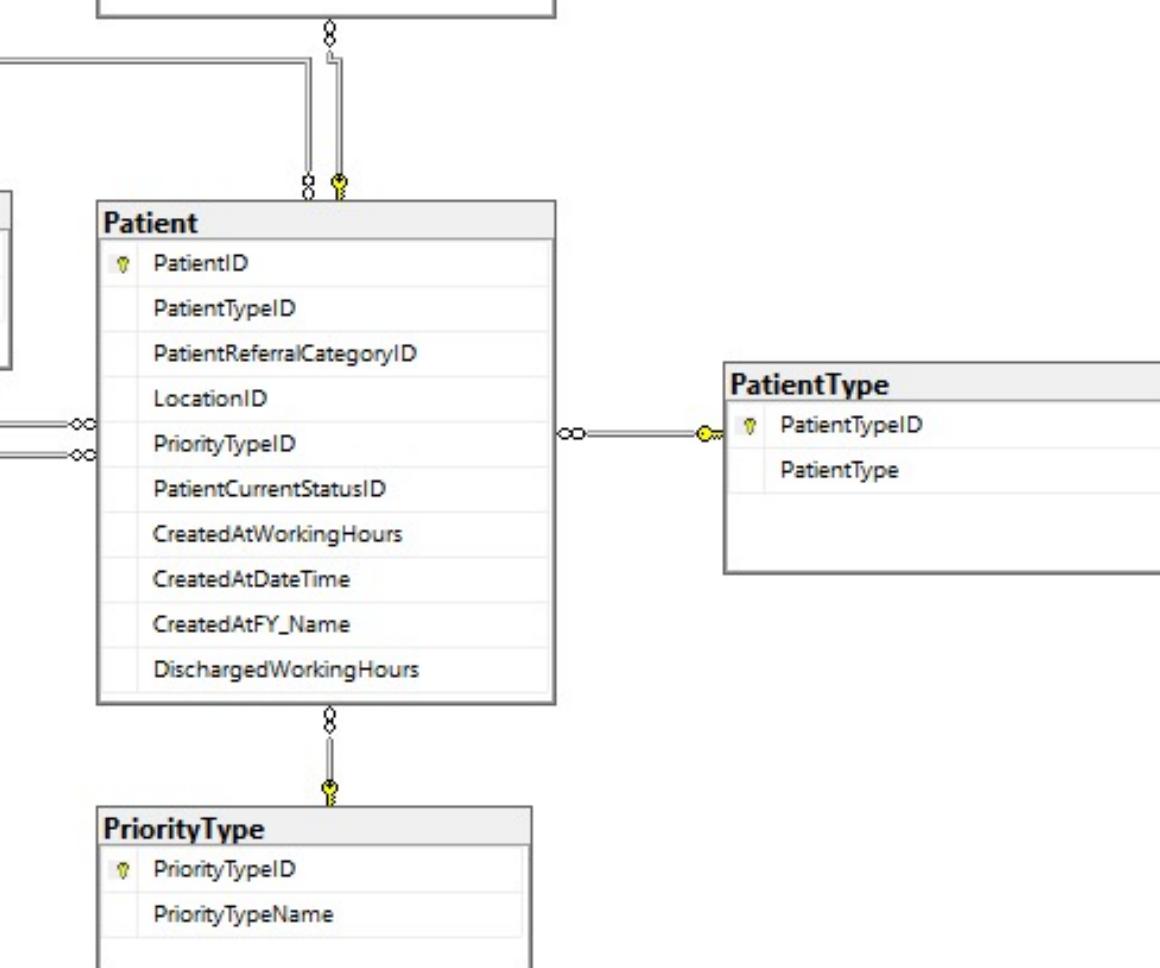
1st Normal Form



2nd Normal Form

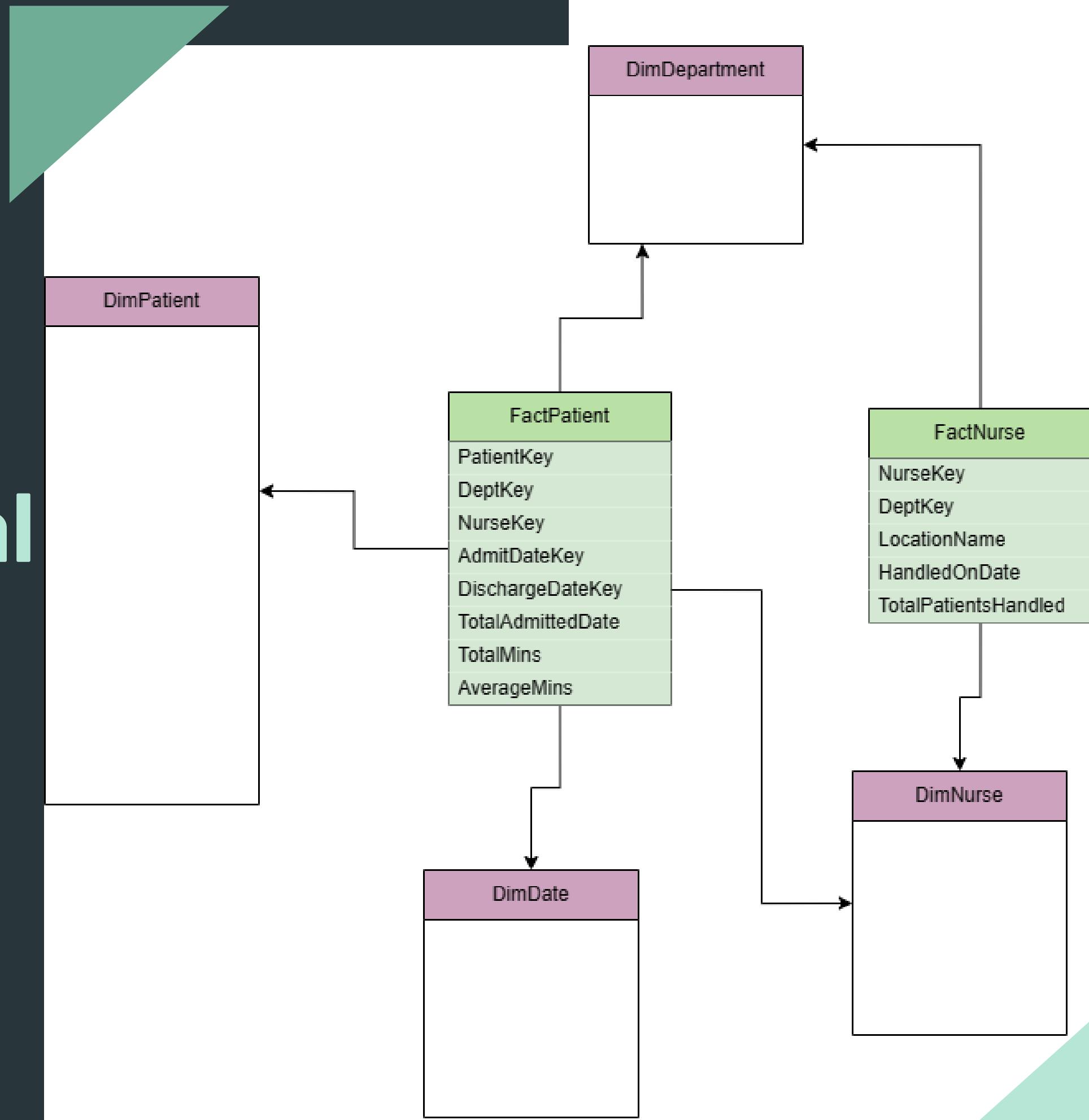


3rd Normal Form



Final Relational Model

# Dimensional Model



SCHEMA  
CONSTELLATION  
FACT TABLES  
TWO  
DIM. TABLES  
FOUR

# Fact Tables

| FactPatient       |
|-------------------|
| PatientKey        |
| DeptKey           |
| NurseKey          |
| AdmitDateKey      |
| DischargeDateKey  |
| TotalAdmittedDays |
| TotalMins         |
| AverageMins       |

**NAME: FACT PATIENT**

**NO. OF MEASUREMENTS: 8**

## **PROPERTIES:**

1. Dates of admission and release, duration.
2. Nurse and department data.
3. Average/total length of stay.
4. Analysis of patient outcomes.

# Fact Tables

## FactNurse

NurseKey

DeptKey

LocationName

HandledOnDate

TotalPatientsHandled

**NAME: FACT NURSE**

**NO. OF MEASUREMENTS: 5**

## PROPERTIES:

1. Details about the nurse and the dept.
2. Site of patient handling.
3. Date and the number of patients served.
4. Performance evaluation of nurses.

# Dimension Tables

| DimDate       |
|---------------|
| DateKey       |
| DateValue     |
| Year          |
| Month         |
| Day           |
| Quarter       |
| StartOfMonth  |
| EndOfMonth    |
| MonthName     |
| DayOfWeekName |

**NAME: DIM DATE**

**NO. OF ATTRIBUTES: 10**

**SCD TYPE: TYPE 0**

**PROPERTIES:**

1. Detailed dates are shown in the table.
2. Dates for the beginning and end of the month are included.
3. Weekday and month name have been added.

# Dimension Tables

| DimPatient          |
|---------------------|
| PatientKey          |
| PatientUsername     |
| PatientType         |
| PriorityTypeName    |
| CreatedAtIndexTime  |
| CreatedAtIndexHours |
| CreatedAtIndexName  |
| PatientStatusName   |
| DepartmentName      |
| LocationName        |
| ...                 |

**NAME: DIM PATIENT**

**NO. OF ATTRIBUTES: 17**

**SCD TYPE: TYPE 2**

**PROPERTIES:**

1. Patient characteristics and information.
2. Holds information about periods of patient record creation and completion.
3. Nurse and status information.

# Dimension Tables

## DimNurse

NurseKey

NurseUserName

NurseGroupName

StartDate

EndDate

**NAME: DIM NURSE**

**NO. OF ATTRIBUTES: 5**

**SCD TYPE: TYPE 2**

**PROPERTIES:**

1. Statistics on nurse employment.
2. Monitoring of nurse employment (DOJ and LWD).
3. Group classification of nurses.

# Dimension Tables

| DimDepartment  |
|----------------|
| DeptKey        |
| DepartmentName |
| LocationName   |

**NAME: DIM DEPARTMENT**

**NO. OF ATTRIBUTES: 3**

**SCD TYPE: TYPE 1**

**PROPERTIES:**

1. Details about the department name.
2. Site of department.

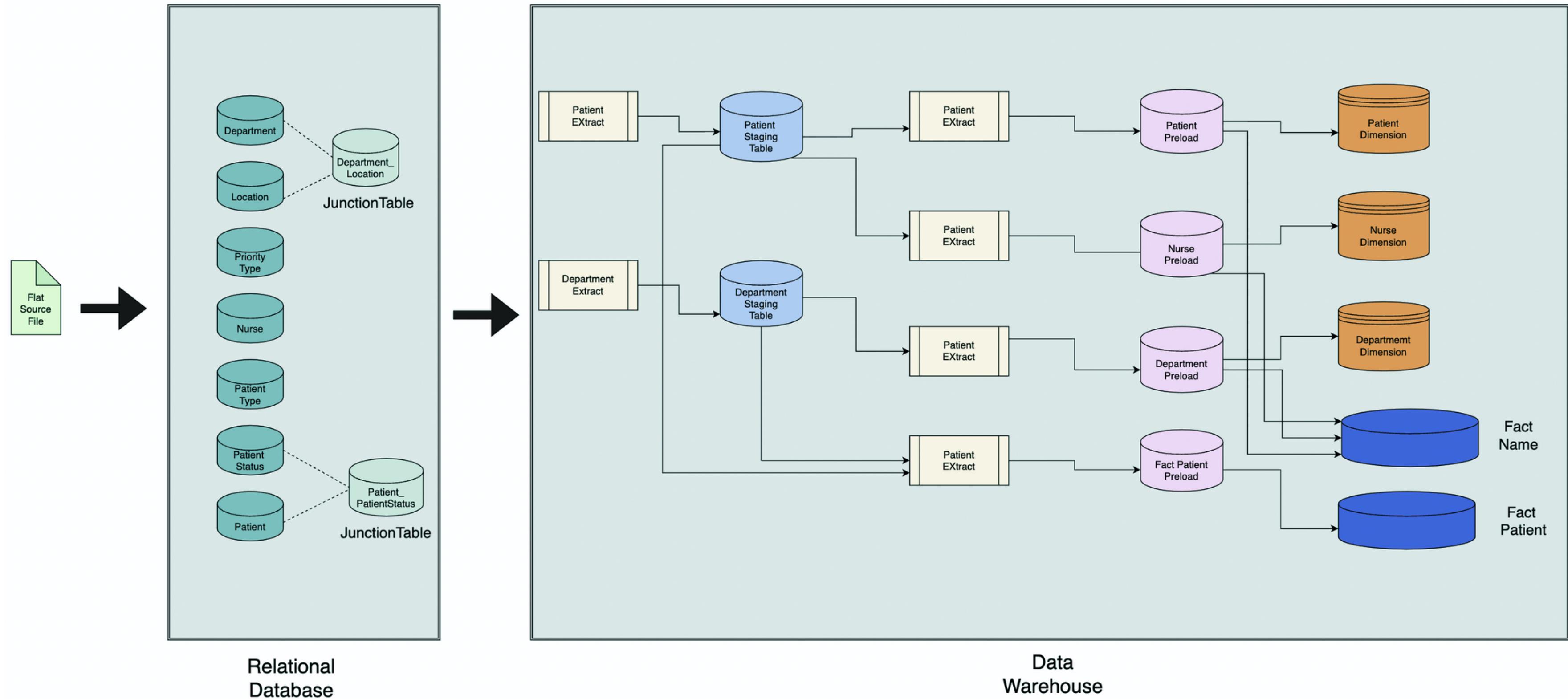
# Extract Transform Load- SSIS

A high level overview of Extract Transform Load

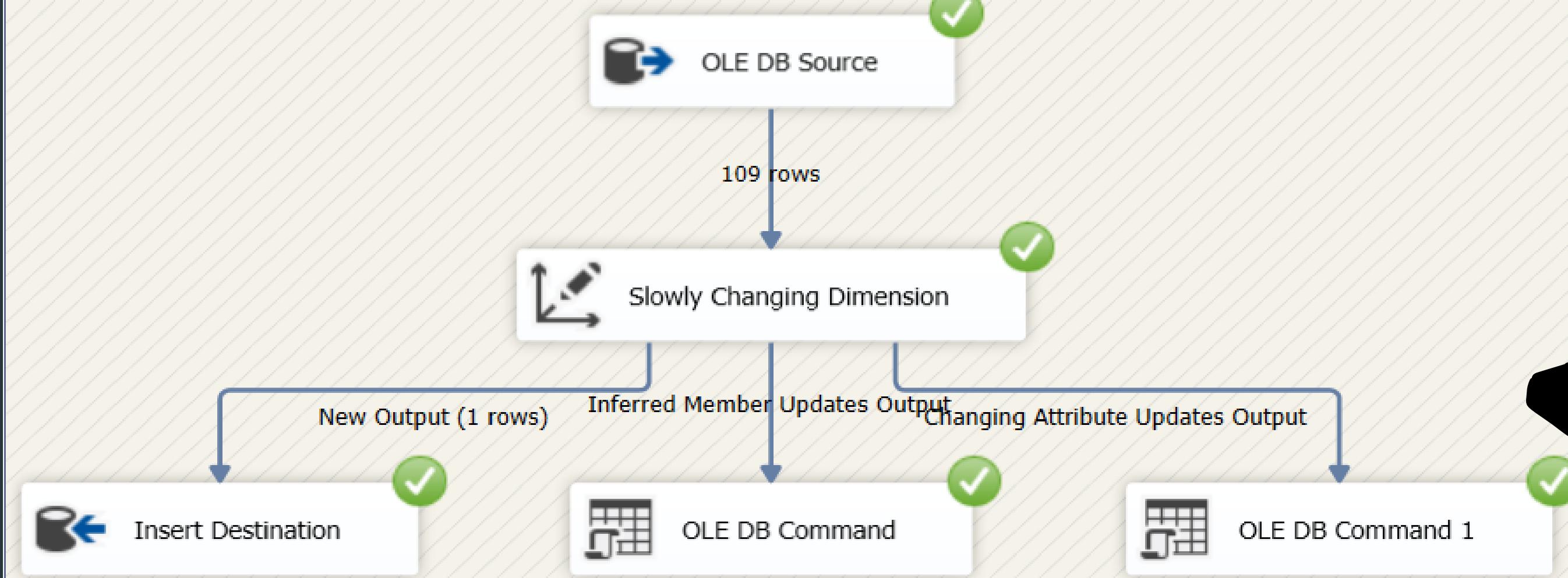
| METRIC      | EXTRACT                  | TRANSFORM               | LOAD                          |
|-------------|--------------------------|-------------------------|-------------------------------|
| PURPOSE     | <b>Get Data</b>          | <b>Alter Data</b>       | <b>Place Data</b>             |
| INPUT DATA  | <b>Database (Source)</b> | <b>Staging Table</b>    | <b>Staging Table</b>          |
| OUTPUT DATA | <b>Staging Table</b>     | <b>Preload Table</b>    | <b>Dimension + Fact Table</b> |
| TOOLS       | <b>SQL Query</b>         | <b>SSIS + SQL Query</b> | <b>SSIS + SQL Query</b>       |

# Extract Transform Load- SSIS

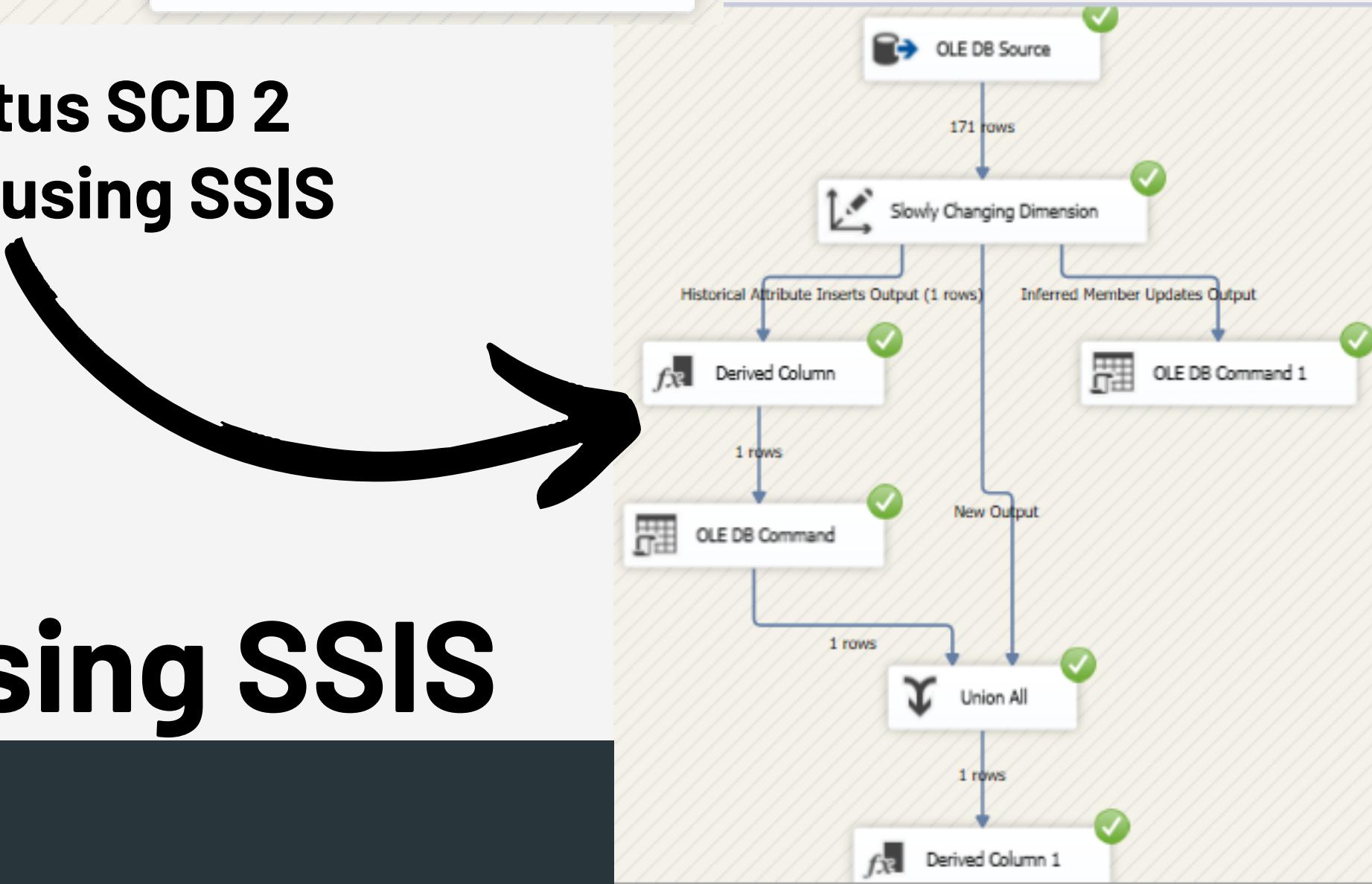
A high level overview of performing ETL using SSIS packages



## DimDepartment SCD 1 handling using SSIS

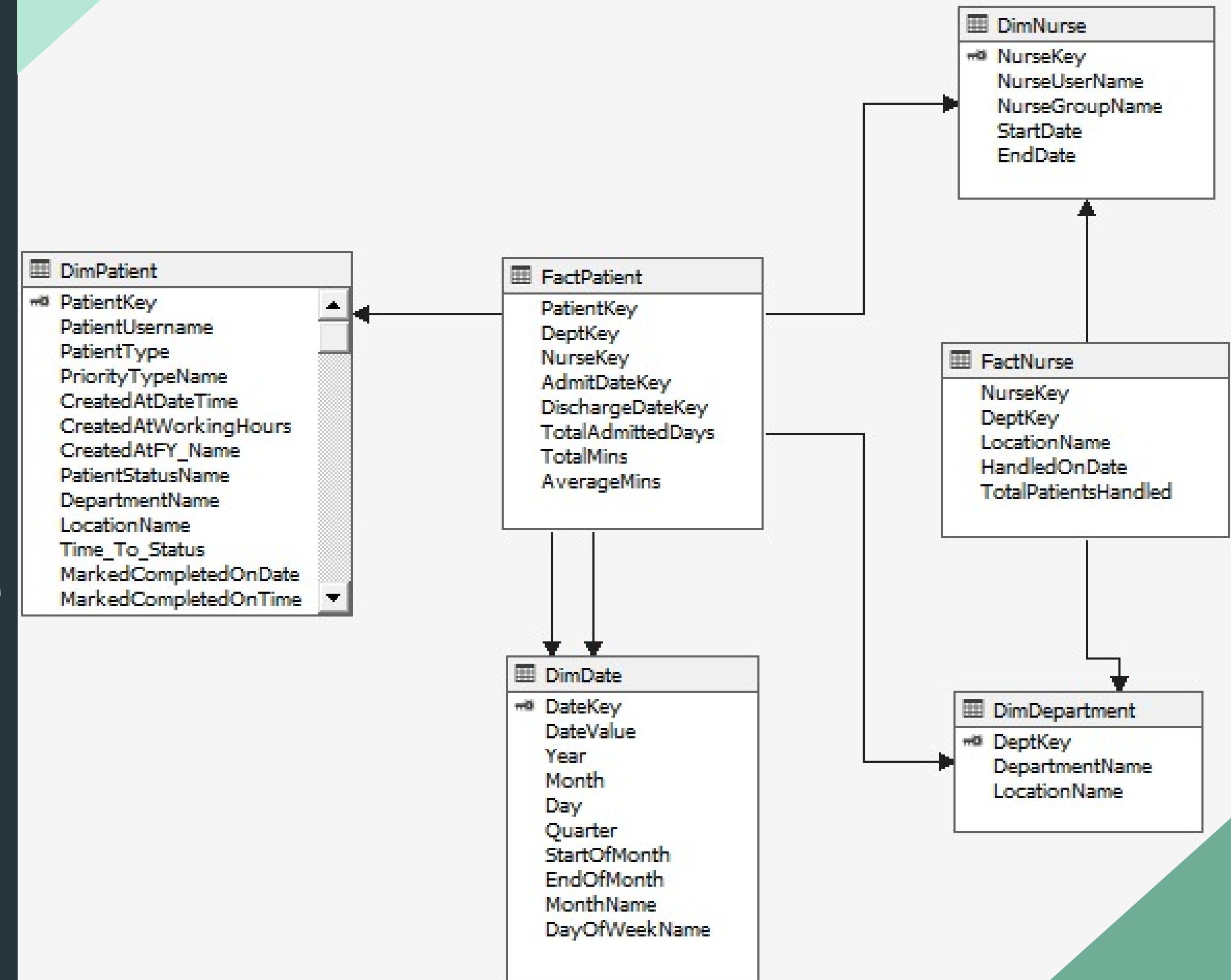


## DimStatus SCD 2 handling using SSIS



# Screenshots of ETL using SSIS

# Multi-dimensional OLAP Cube



- 1. Total Patients Handled by the nurse. (Total Patients Handled)
- 2. Average Minutes Patient has spent in the hospital. (Average Mins)
- 3. Total Minutes spent by the patient in the hospital. (Total Mins)
- 4. Total admitted days of the patient until discharge. (Total Admitted Days)

# Hierarchy

Data may be organized and analysed at various degrees of granularity using hierarchies, which are logical arrangements of data properties in a tree-like structure.

**NURSE HIERARCHY**  
A hierarchy for studying data on nurses based on nurse characteristics.

**DEPARTMENT HIERARCHY**  
A hierarchy for analyzing department-related data based on department attributes.

**CALENDAR HIERARCHY**  
A hierarchy for analyzing time-related data based on calendar attributes.

# MDX Query

A query language called MDX (Multidimensional Expressions) is used to get information out of OLAP (Online Analytical Processing) databases.

## PATIENT CUBE QUERY

```
(SELECT NON EMPTY { [Measures].[Patient Key], [Measures].[Total Admitted Days], [Measures].[Average Mins], [Measures].[Total Mins] } ON COLUMNS, NON EMPTY { ([Dim Nurse].[Nurse Key].[Nurse Key].ALLMEMBERS * [Dim Nurse].[Nurse User Name].[Nurse User Name].ALLMEMBERS * [Dim Department].[Dept Key].[Dept Key].ALLMEMBERS * [Dim Department].[Location Name].[Location Name].ALLMEMBERS ) } DIMENSION PROPERTIES MEMBER_CAPTION, MEMBER_UNIQUE_NAME ON ROWS FROM [RIMS DW] CELL PROPERTIES VALUE, BACK_COLOR, FORE_COLOR, FORMATTED_VALUE, FORMAT_STRING, FONT_NAME, FONT_SIZE, FONT_FLAGS).
```

## NURSE CUBE QUERY

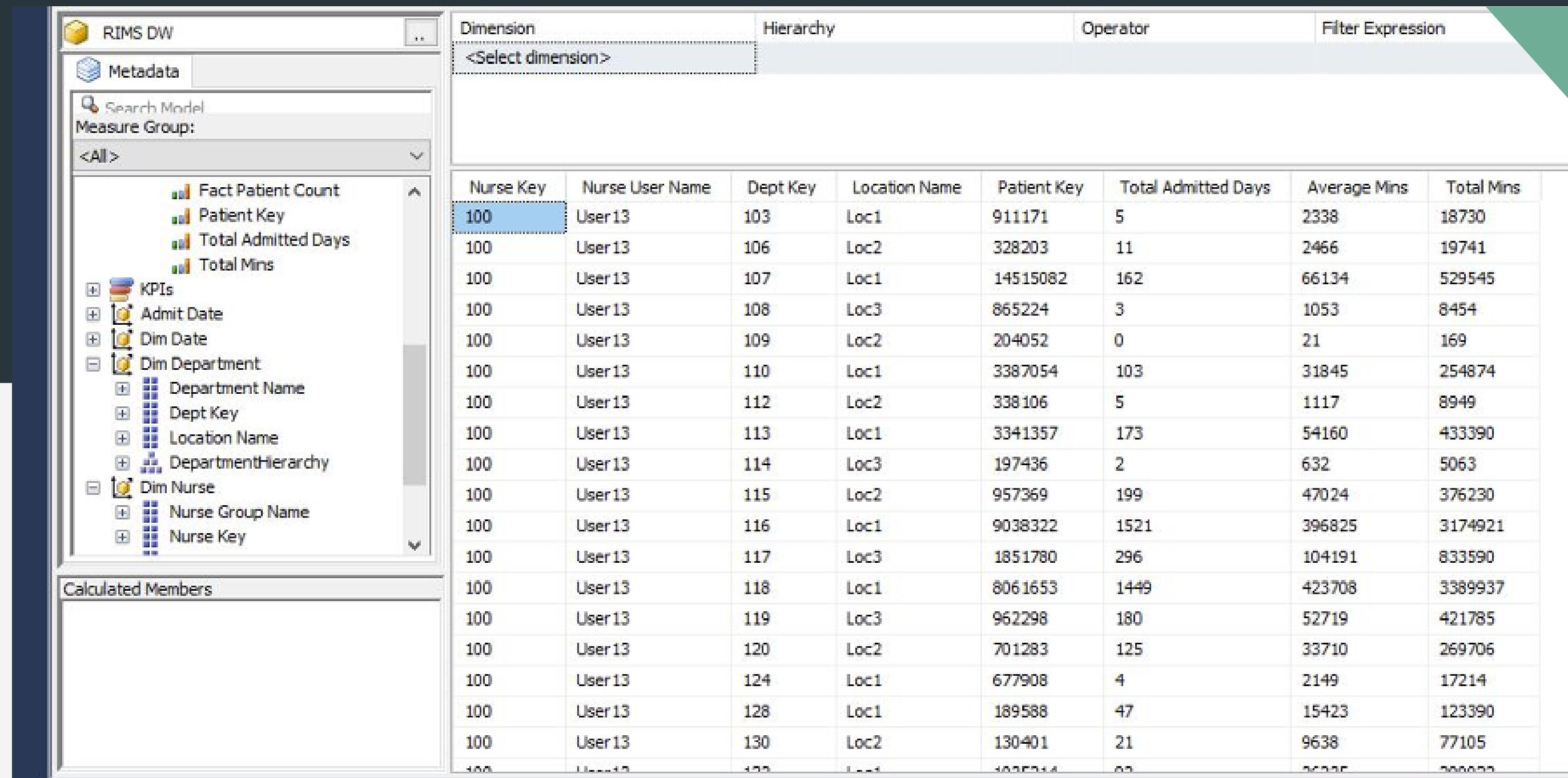
```
SELECT NON EMPTY { [Measures].[Total Patients Handled] } ON COLUMNS, NON EMPTY { ([Dim Department].[Department Name].[Department Name].ALLMEMBERS * [Dim Department].[Location Name].[Location Name].ALLMEMBERS * [Dim Nurse].[NurseHierarchy].[Nurse User Name].ALLMEMBERS ) } DIMENSION PROPERTIES MEMBER_CAPTION, MEMBER_UNIQUE_NAME ON ROWS FROM [RIMS DW] CELL PROPERTIES VALUE, BACK_COLOR, FORE_COLOR, FORMATTED_VALUE, FORMAT_STRING, FONT_NAME, FONT_SIZE, FONT_FLAGS
```

# Multidimensional Cube- Nurse Cube

| Location Name | Department Name | Nurse Group Name         | Nurse User Name | Total Patients Handled |
|---------------|-----------------|--------------------------|-----------------|------------------------|
| Loc1          | ADP/SH/AL       | IOCC Care Coordinators   | User131         | 16                     |
| Loc1          | ADP/SH/AL       | Other N/A                | User107         | 8                      |
| Loc1          | ADP/SH/AL       | Other N/A                | User133         | 8                      |
| Loc1          | ADP/SH/AL       | Other N/A                | User174         | 8                      |
| Loc1          | ADP/SH/AL       | Other N/A                | User29          | 8                      |
| Loc1          | ADP/SH/AL       | Other N/A                | User58          | 8                      |
| Loc1          | ADP/SH/AL       | Other N/A                | User89          | 8                      |
| Loc1          | Aredia/Zometa   | ICR PSA                  | User52          | 16                     |
| Loc1          | Aredia/Zometa   | Intake Care Coordinators | User105         | 216                    |
| Loc1          | Aredia/Zometa   | Intake Care Coordinators | User109         | 16                     |
| Loc1          | Aredia/Zometa   | Intake Care Coordinators | User116         | 8                      |
| Loc1          | Aredia/Zometa   | Intake Care Coordinators | User123         | 56                     |
| Loc1          | Aredia/Zometa   | Intake Care Coordinators | User137         | 40                     |
| Loc1          | Aredia/Zometa   | Intake Care Coordinators | User140         | 40                     |
| Loc1          | Aredia/Zometa   | Intake Care Coordinators | User143         | 8                      |
| Loc1          | Aredia/Zometa   | Intake Care Coordinators | User155         | 824                    |
| Loc1          | Aredia/Zometa   | Intake Care Coordinators | User162         | 16                     |
| Loc1          | Aredia/Zometa   | Intake Care Coordinators | User168         | 8                      |
|               |                 |                          |                 | 824                    |

The "Nurse" cube's measures and dimension members are picked out and shown in rows and columns, respectively, by this MDX query, which also defines attributes for each cell of the resultant table.

# Multidimensional Cube- Patient Cube



The screenshot shows the RIMS DW application interface. On the left, the 'Metadata' pane displays the 'Measure Group' <All> with various measures: Fact Patient Count, Patient Key, Total Admitted Days, Total Mins, KPIs, Admit Date, Dim Date, Dim Department, Department Name, Dept Key, Location Name, DepartmentHierarchy, and Dim Nurse, Nurse Group Name, and Nurse Key. The main area shows a grid of data with the following columns: Nurse Key, Nurse User Name, Dept Key, Location Name, Patient Key, Total Admitted Days, Average Mins, and Total Mins. The data is organized by Nurse Key (100), Nurse User Name (User13), Dept Key (103, 106, 107, 108, 109, 110, 112, 113, 114, 115, 116, 117, 118, 119, 120, 124, 128, 130), Location Name (Loc1, Loc2, Loc3), Patient Key (911171, 328203, 14515082, 865224, 204052, 3387054, 338106, 3341357, 197436, 957369, 9038322, 1851780, 8061653, 962298, 701283, 677908, 189588, 130401), Total Admitted Days (5, 11, 162, 3, 0, 103, 5, 173, 2, 199, 1521, 296, 1449, 180, 125, 4, 47, 21), Average Mins (2338, 2466, 66134, 1053, 21, 31845, 1117, 54160, 632, 47024, 396825, 104191, 423708, 52719, 33710, 2149, 15423, 9638), and Total Mins (18730, 19741, 529545, 8454, 169, 254874, 8949, 433390, 5063, 376230, 3174921, 833590, 3389937, 421785, 269706, 17214, 123390, 77105, 600000).

| Dimension          |                 | Hierarchy |               | Operator    |                     | Filter Expression |            |
|--------------------|-----------------|-----------|---------------|-------------|---------------------|-------------------|------------|
| <Select dimension> |                 |           |               |             |                     |                   |            |
| Nurse Key          | Nurse User Name | Dept Key  | Location Name | Patient Key | Total Admitted Days | Average Mins      | Total Mins |
| 100                | User13          | 103       | Loc1          | 911171      | 5                   | 2338              | 18730      |
| 100                | User13          | 106       | Loc2          | 328203      | 11                  | 2466              | 19741      |
| 100                | User13          | 107       | Loc1          | 14515082    | 162                 | 66134             | 529545     |
| 100                | User13          | 108       | Loc3          | 865224      | 3                   | 1053              | 8454       |
| 100                | User13          | 109       | Loc2          | 204052      | 0                   | 21                | 169        |
| 100                | User13          | 110       | Loc1          | 3387054     | 103                 | 31845             | 254874     |
| 100                | User13          | 112       | Loc2          | 338106      | 5                   | 1117              | 8949       |
| 100                | User13          | 113       | Loc1          | 3341357     | 173                 | 54160             | 433390     |
| 100                | User13          | 114       | Loc3          | 197436      | 2                   | 632               | 5063       |
| 100                | User13          | 115       | Loc2          | 957369      | 199                 | 47024             | 376230     |
| 100                | User13          | 116       | Loc1          | 9038322     | 1521                | 396825            | 3174921    |
| 100                | User13          | 117       | Loc3          | 1851780     | 296                 | 104191            | 833590     |
| 100                | User13          | 118       | Loc1          | 8061653     | 1449                | 423708            | 3389937    |
| 100                | User13          | 119       | Loc3          | 962298      | 180                 | 52719             | 421785     |
| 100                | User13          | 120       | Loc2          | 701283      | 125                 | 33710             | 269706     |
| 100                | User13          | 124       | Loc1          | 677908      | 4                   | 2149              | 17214      |
| 100                | User13          | 128       | Loc1          | 189588      | 47                  | 15423             | 123390     |
| 100                | User13          | 130       | Loc2          | 130401      | 21                  | 9638              | 77105      |
| 100                | User13          | 132       | Loc1          | 1000000     | 0                   | 0                 | 000000     |

This MDX query retrieves the "Patients Details" measure from the "Patient" cube and shows it for each combination of the department name, location name, and nurse user name in rows and columns, respectively. It also defines characteristics for each cell of the resultant table.

# Data Mining via Machine Learning

## Objective

### **What are we predicting?**

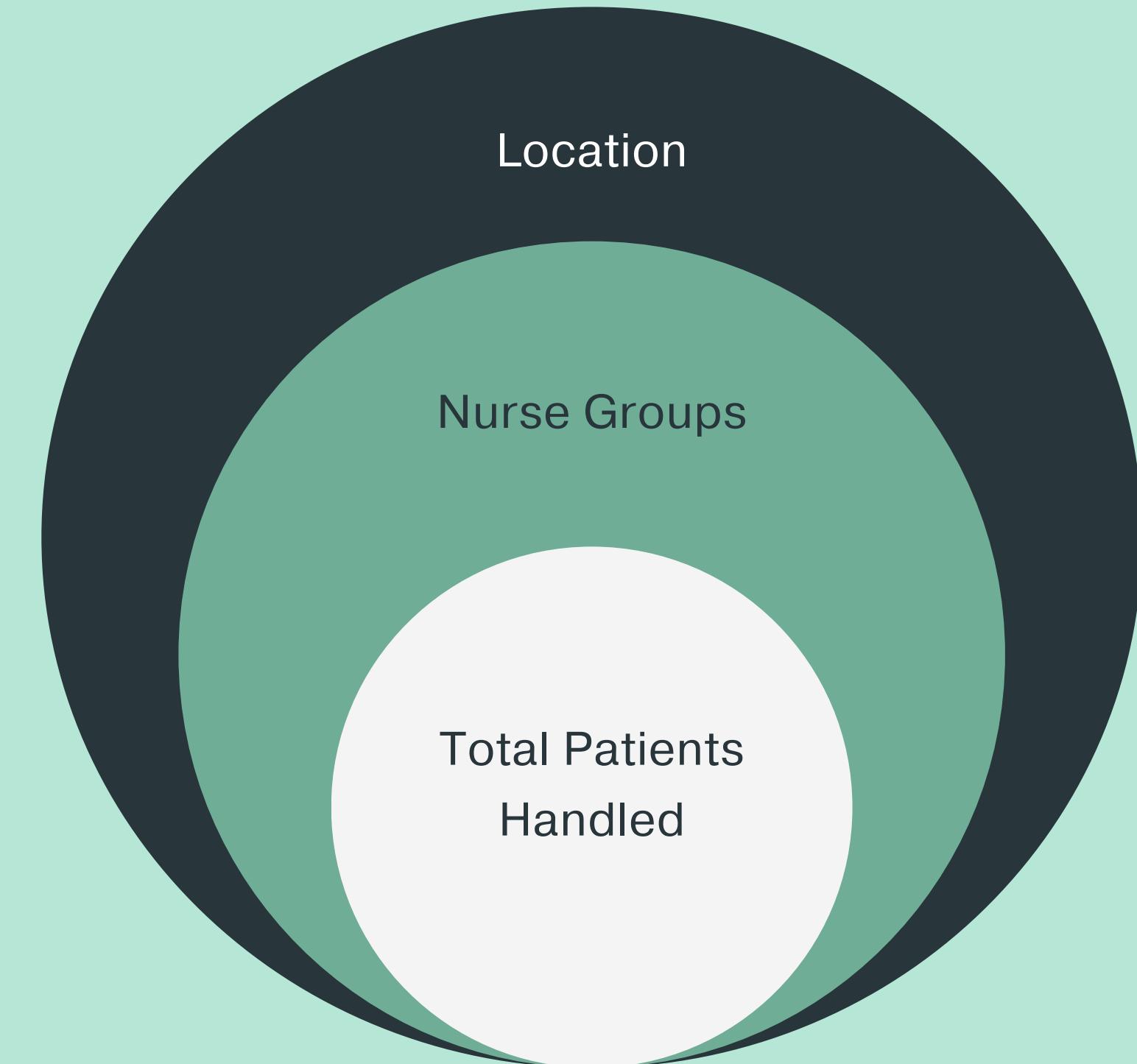
The number of patients a new nurse, of a specialization group, deployed at a particular location, would be in charge of.

### **How is this useful?**

To aid resource allocation for proper utilization of nurse workforce.

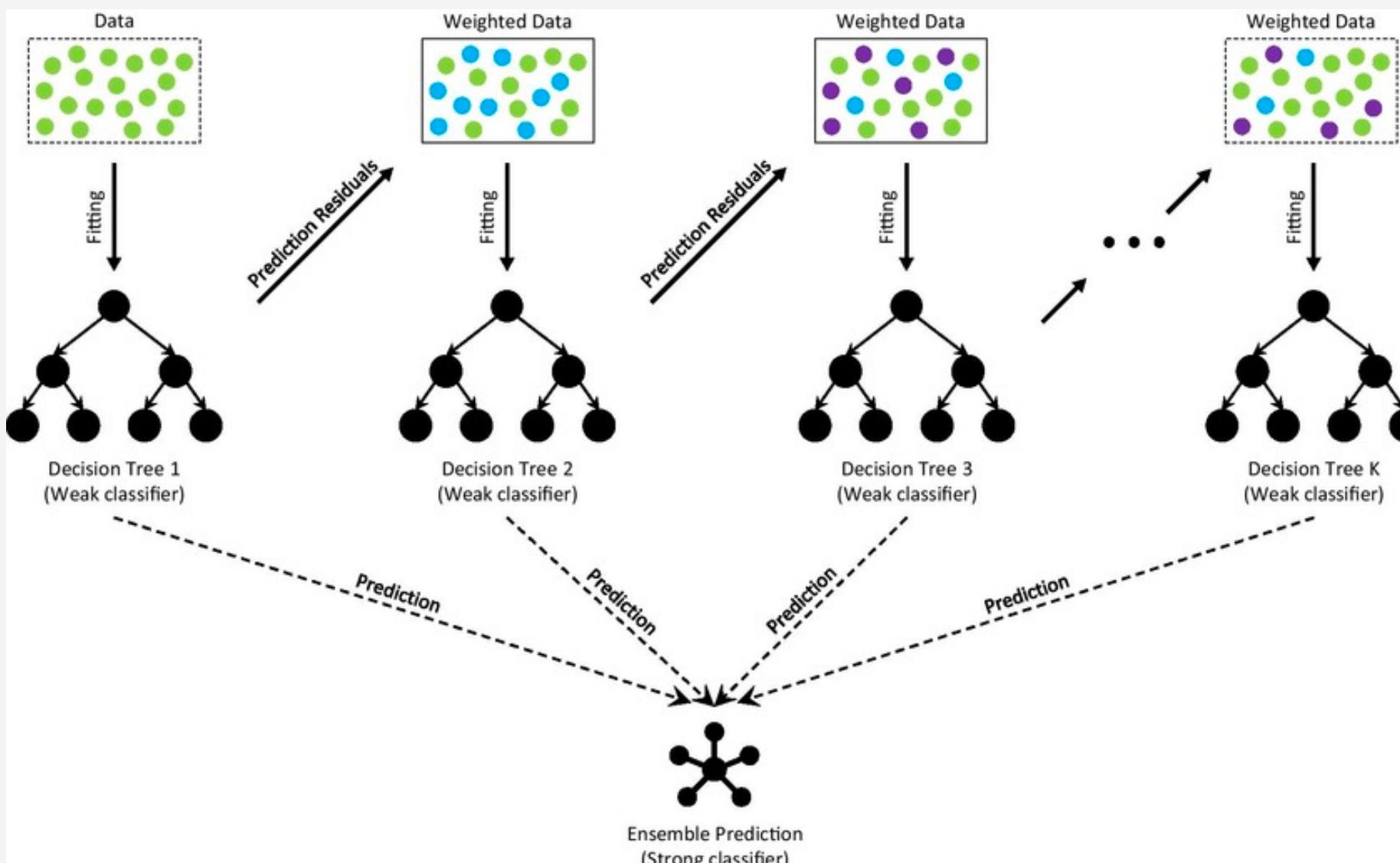
### **Real-Life implementation**

Centralized nurse scheduling improves staffing and patient care.



# Data Mining via Machine Learning

Applied one of Microsoft's machine learning algorithms to the chosen dataset after creating a data mining structure.



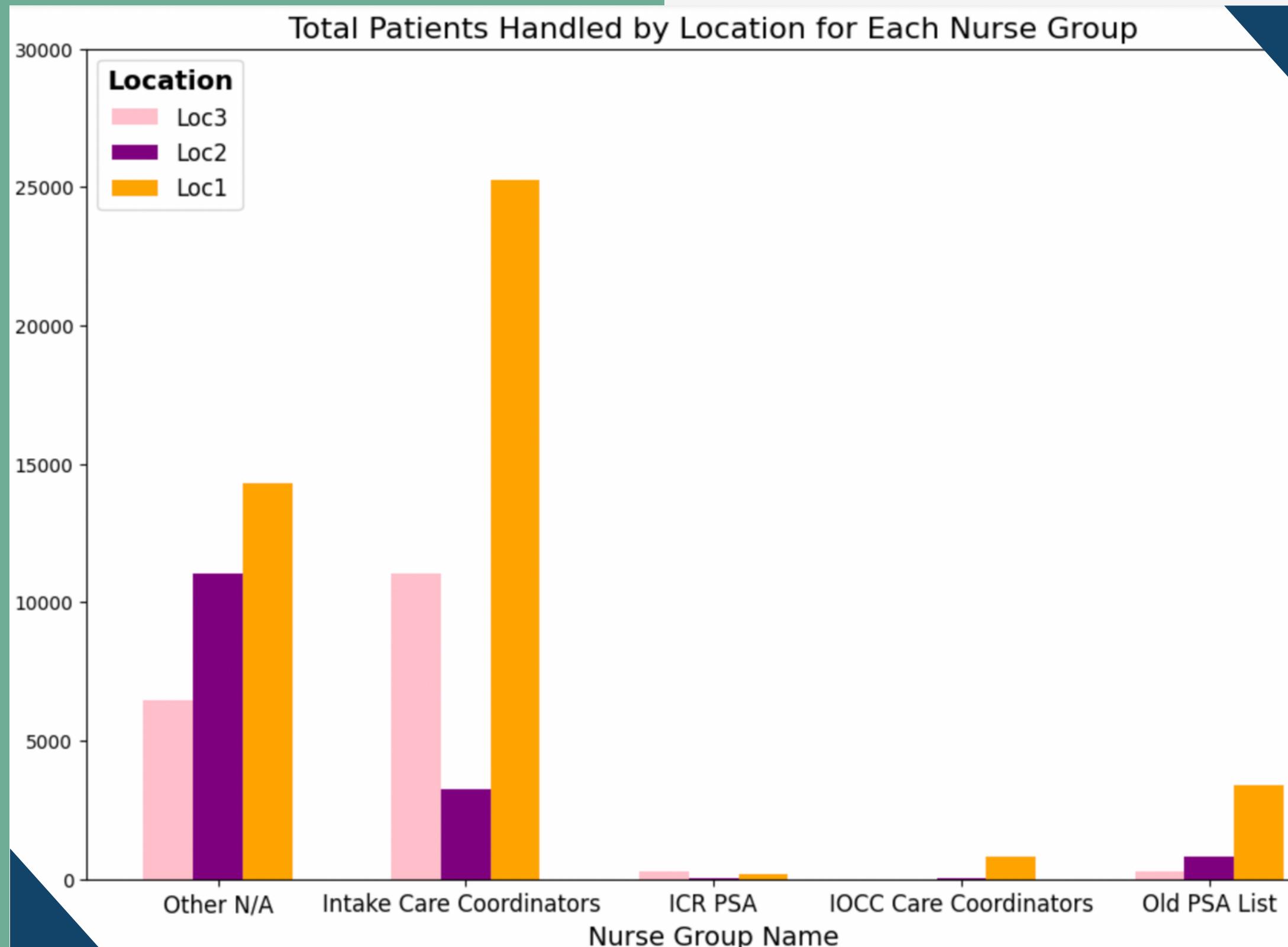
Multi-Class:  
**Regressor**

Relationship:  
**Non-Linear**

Algorithm:  
**Gradient Boosting Regression**

# Stacked Bar Chart

The graph visualizes the total number of patients handled by each nurse group at each location.

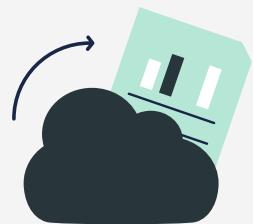


# Resources

[https://www.kimballgroup.com/data-warehouse-business-intelligence-resources/.](https://www.kimballgroup.com/data-warehouse-business-intelligence-resources/)



[https://www.tutorialspoint.com/dbms/er\\_diagram.htm](https://www.tutorialspoint.com/dbms/er_diagram.htm)



[https://www.sas.com/en\\_ca/insights/data-management/what-is-etl.html](https://www.sas.com/en_ca/insights/data-management/what-is-etl.html)



<https://learn.microsoft.com/en-us/sql/integration-services/sql-server-integration-services?view=sql-server-ver16>



<https://learn.microsoft.com/en-us/analysis-services/ssas-overview?view=asallproducts-allversions>

Thank  
You