CONCEPTUAL DESIGN AND IMPLEMENTATION OF A DATAWAREHOUSE/MART SYSTEM

CASE STUDY: **NORTH-WEST YORKSHIRE ELDERLY SOCIAL & HEALTHCARE**

AUTHOR: Victoria Chika Atasie

INTRODUCTION

The aim of this case study of the North and West Yorkshire Elderly and Social Health Care System report is to implement a centralized database solution that would assist in providing a solution that is associated with the integration of data from the West Yorkshire and North Yorkshire health and social care datasets whilst implementing the Extract, Transform and Load (ETL) processes and maintenance. To develop a comprehensive and real time view of bed occupancy rate which is considered the key performance indicator and optimizing the bed resources to meet patient's needs, create a data mart/data warehouse and enable the business owner who is the stakeholder in making a better data driven decision in providing quality care, improve the business reputation and the effectiveness of care homes.

See the detailed case study, all evidence codes, screenshot of codes running successfully would be in the Appendix.

DISCLAIMER: The data used for this case study project has been anonymized to ensure no individual's identity can be determined. All sensitive and personal identifiable information has been removed in accordance with data privacy law and ethical guidelines. This project is for educational and informational purposes. Any resemblance to real persons or organizations is purely coincidental. The analysis presented is based on information available at the time of writing.

PROPOSED CENTRALIZED DATABASE SOLUTION, AIMS AND OBJECTIVES

As a BI Consultant, I would be implementing a centralized database solution that would help resolve the problem in integrating data from West Yorkshire and North Yorkshire health and social care datasets whilst combining the bed occupancy data from both care homes. The Aim would create a comprehensive and real time view of the bed occupancy rate and considering the bed resources optimization allocation to meet patient's needs and assist the business owner in making better data driven decision in the provision of quality care, enhance their business, reputation and increased

demand for care home services. **The Main Aim** is to understand dimensions in a star schema and fact table which would aid in indicating how data can be handled from multiple sources of data.

Objectives: The objective would support the optimization and utilization of beds leading to high bed occupancy rates and minimizing the possible bed vacancy.

Improve Quality of Care whilst maintaining bed occupancy in care homes

Effective costing in managing resource allocation with the bed occupancy rate whilst reducing expenses.

Enhancing patients' satisfaction whilst creating positive satisfaction and long-term occupancy.

STAKEHOLDER

The stakeholder for the project would be **North Yorkshire and West Yorkshire Care Homes Business Owner**. These are owners of the care homes and run the overall business strategy and are the people who require a data warehouse to be implemented and help them in better decision-making.

The relevant business question that would be answered for the business owner or investor to aid the implementation of the proposed centralized database solution would include:

-Does the bed occupancy rate vary from West Yorkshire and North Yorkshire Care homes per year

-What is the bed occupancy rate in West Yorkshire and North Yorkshire Care homes per year?

-What are the trends and patterns in the bed occupancy in the West Yorkshire and North Yorkshire Care homes?

STRATEGY, EXTERNAL AND INTERNAL BUSINESS DRIVERS

The Strategy would assist in the implementation of the centralized database solution by integrating the health and social care data in the optimization of the bed occupancy whilst enhancing the quality of care provided as a business and creating a real time system of the bed occupancy rates and give insight for an informed decision making.

The external business driver could involve meeting up the business intelligence solution to the standard compliance requirements and regulation in the health and social care data privacy for the data integration.

The internal business driver could involve maintaining a high level of bed occupancy rate for revenue generated growth as more of available bed can lead to financial loses to the business owners whilst aiming at using the data insight to a better decision making.

Short/long term strategy: The short term may involve improving decision making and optimizing bed occupancy whilst implementing centralized database solution and integrating data.

The Long-term strategy would enhance the possibility of maintaining high bed occupancy rates leading to the success of the business for health and care homes.

KEY PERFORMANCE INDICATOR

The key performance indicator is **Bed Occupancy Rate**. This would help in setting the right indicator in terms of performance and give insight into what is measurable and contribute to decision making.

REPORTS

The stated reports would help in tracking and meet the expected Key performance indicators.

The reports required to support business questions would involve measures such as

The Percentage of Bed occupancy in the North Yorkshire and West Yorkshire Care Homes per year?

Wha is the Total number of Bed Occupied for 2023 per North Yorkshire and West Yorkshire Care Homes?

The Total number of Available Bed per month per year in the North Yorkshire and West Yorkshire Care Homes?

?

The Total number of Bed Available per year in the North Yorkshire and West Yorkshire Care Homes?

?

The relevant business question that would be answered for the business owner to aid the implementation of the proposed centralized database solution would include:

-Does the bed occupancy rate vary from West Yorkshire and North Yorkshire Care homes per year?

-What is the bed occupancy rate in West Yorkshire and North Yorkshire Care homes per year?

-What are the trends and patterns in the bed occupancy in the West Yorkshire and North Yorkshire Care homes?

DATA WAREHOUSE ARCHITECTURE, TECHNOLOGY, METHODOLOGY, AND IMPLEMENTATION.

The architecture for this case study would be the Three-tier Architecture and Dependent Datamart would be created. The data sources would be derived from North Yorkshire and West Yorkshire datasets, the data derived from these sources would be extracted, transformed and loaded (ETL) which would involve data integration, data cleaning and checking the data quality and consistency into the DataMart. In Datamart the data integrated would be stored for which can be used for analysis leading to generating reports for the Business Owner.

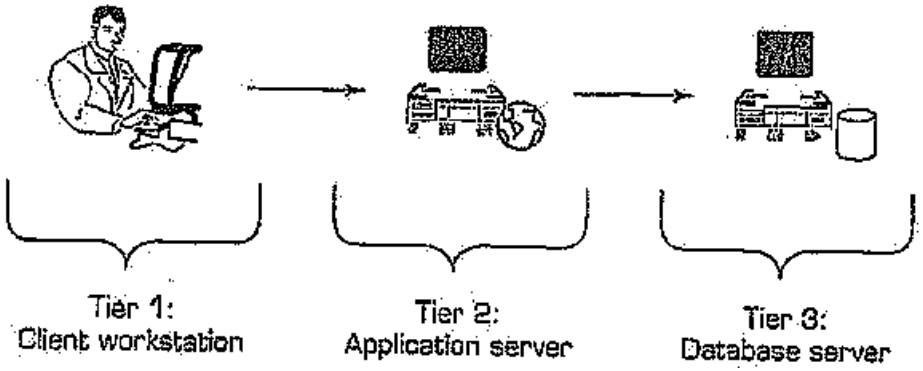


Figure 1:Three Tier Architecture Data Warehouse (Source: From Decision Support and Business Intelligence System - Efraim Turban 2011)

Operational systems in the three tier involves software for data acquisition and data in one tier which can be referred to as the server, the data warehouse forms another tier and the third tier includes the application server and the client(Efraim

Turban).The data are processed twice and placed in the multi-dimensional database from the warehouse which makes it easy for multidimensional presentation and analysis and can be repeated in the data mart.

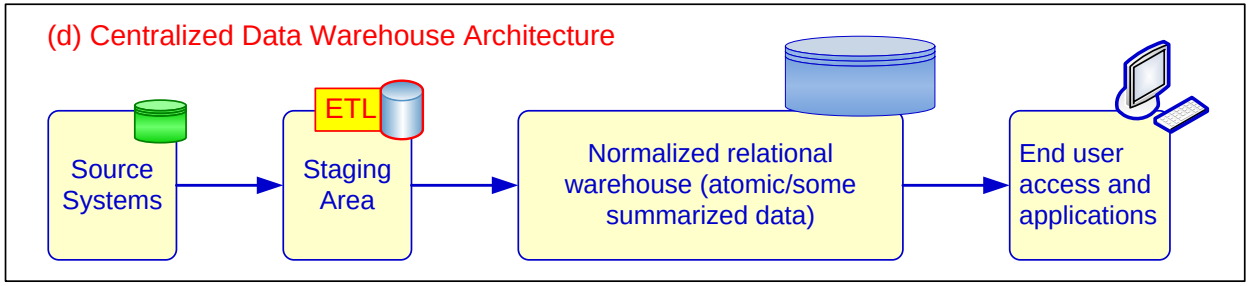


Figure 2: Centralized Data Warehouse Architecture (Source: The DW Design and Architecture Slides)

TECHNOLOGY

The technology that can be used are various Database Management System such as Oracle, Microsoft SQL Server which can be used for the storage of data, Business Intelligence tools such as Tableau for data visualization to the business owner.

METHODOLOGY

The methodology for the software development life cycle would include the Planning Phase where the requirement ,objectives and scope would be initiated, Gather the data from the data sources which would be Analysis phase, consider the architecture to be used in the Design phase, perform the extraction, transformation and loading processes and ensuring data quality checks and accuracy, performance would be tested in Testing phase, conduct implementation for the Deployment phase and maintenance would be established.

IMPLEMENTATION

The implementation approach would include the Extraction from the data sources, Transformed and Loaded into the DataMart. Data accuracy would be considered, performing data cleaning and consistency to enable analysis and reporting.

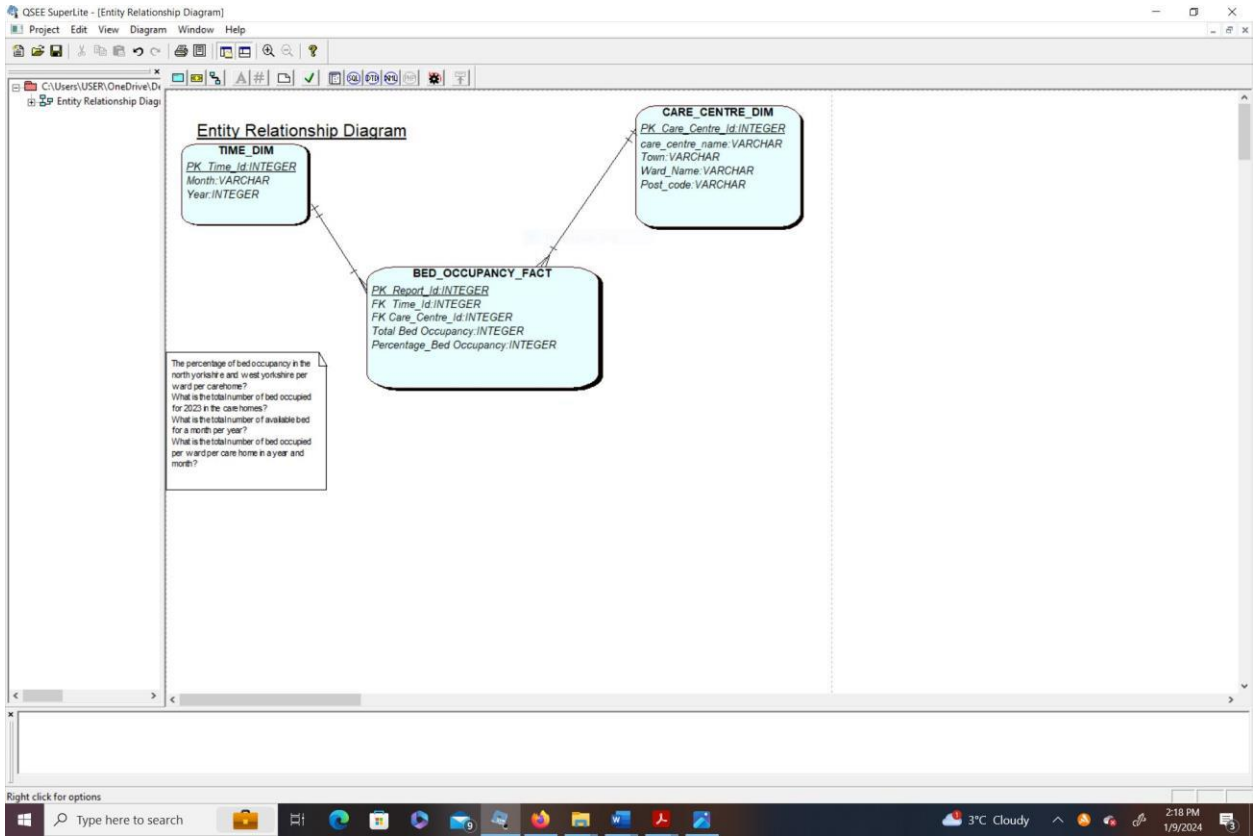
MAINTENANCE

The maintenance for centralized database solution of this case study would include constant data updating, queries monitoring for better performance, the data integration to ensure data accuracy ,monitoring the ensuring errors are resolved in the DataMart, providing adequate communication with business owner as the stakeholder and providing updates of reports ,data security would be implemented, ensure easy access for the business owner or other users and ensuring the Datamart or Datawarehouse capacity for future voluminous data.

STAR SCHEMA AND DATA DICTIONARY

Star schema involves dimensional design for relational database are categorized as columns of the dimension table whilst the fact table contains facts that are stored as columns (Christopher Adams ,2010)

The star schema for the decision support system for supporting the key performance indicator comprises of the Time Dimension, the bed occupancy fact, the care center dimension, and the measures. The aggregated measures would be used to provide reports and querying of its data.


QSEE SuperLite - [Entity Relationship Diagram]
Project Edit View Diagram Window Help
Entity Relationship Diagram
TIME_DIM
PK_Time_Id:INTEGER
Month:VARCHAR
Year:INTEGER
CARE_CENTRE_DIM
PK_Care_Centre_Id:INTEGER
care_centre_name:VARCHAR
Town:VARCHAR
Ward_Name:VARCHAR
Post_code:VARCHAR
BED_OCCUPANCY_FACT
PK_Report_Id:INTEGER
FK_Time_Id:INTEGER
FK Care_Centre_Id:INTEGER
Total Bed Occupancy:INTEGER
Percentage_Bed Occupancy:INTEGER
The percentage of bed occupancy in the north yorkshire and west yorkshire per ward per carehome?
What is the total number of bed occupied for 2023 in the care homes?
What is the total number of available bed for a month per year?
What is the total number of bed occupied per ward per care home in a year and month?
Right click for options
Type here to search
3°C Cloudy
2:18 PM
1/9/2024

SOURCE 1: WYR_CARE CENTRE (care_Id, **Care_centre_name**, address, **postcode, town**, email)

WYR_Bed (ward_no**,** bed_no**, bed_status**)

WYR_RESERVATION (Reservation_Id, **Admission_Date**)

WYR_Bed Assigned(reservation_Id,bed_no)

WYR_WARD (**WARD_NO**, **WARD_NAME**, **CARE_ID**)

SOURCE 2: NYR_CARE CENTRE (care_centre_Id, **Care_centre_name**, address, **postcode, town**, email)

NYR_BED (bed_id, **bed_status**, ward_id, bed_type)

NYR_ADMISSION (Admission_no ,**Admission_date**)

NYR_WARD (**WARD_ID**, **WARD_NAME, CARE_CENTRE_ID**)

The lowest level of granularity for each Dimension is as follows:

Time Dimension:(Time_Id,Month,Year)The Time_Id would be the lowest level of granularity

Care_Centre Dimension:(Care_Centre_Id,Care_Centre_Name,Postcode,Town)The Care_Centre_Id would be the lowest level of granularity.

DATA DICTIONARY DOCUMENTED

SS Definitions				Mapping/ Data Source			
Dimension	**Attribute Name**	**Data Type**	**Key**	**Data Sources**	**Data Sources Type**	**Definition**:	**Notes**
Time_Dim	time_id	NUMBER	Primary	none	NUMBER	time_id_SEQ This is the unique Identifier of time, e.g. 1, 2, 3,	This is a surrogate key, that will be populated through Sequence, generated at the staging area.
	Month	NUMBER	No	Admission_Date	DATE	This is the Month extracted from data source columns, populated in time dimension as January, March, April, November	Data would be pulled from two tables, making sure there are no duplicates
	Year	NUMBER	No	AdmissionDate	DATE	This is the Year extracted from data source columns, populated in time dimension as 2022, 2023	Data would be pulled from two tables, making sure there are no duplicates
Care_Centre_Dim	Care_Centre_id	NUMBER	Primary	Care_Centre_Id	NUMBER	This is the unique Identifier of the care_centre_Id e.g., 1, 2, 3	Will pull in data from two tables (WYR&NYR), and ensure

							there are no duplicates
	Care_centre_name	VARCHAR	None	Care_centre_name	VARCHAR	This is the name of the care centre to be used e.g Oscar Care home, Wellbeing	Will pull in data from two tables, and ensure there are no duplicates
	Address	VARCHAR	None	Address	VARCHAR	This would be the address of the care_centre e.g. Meifod Isaf Cottage	Will pull in data from two tables, and ensure there no duplicates

	Town	VARCHAR	None	Town	VARCHAR	This would be the town description of the care_centre	Will pull in data from two tables, and ensure there no duplicates
	Postcode	VARCHAR	None	Postcode	VARCHAR	This would be the postcode description of the care_centre	Will pull in data from two tables, and ensure there no duplicates
Bed_Occupancy_Fact	Report_Id	NUMBER	Primary	Report_SEQ	NUMBER	This is the unique identifier of Bed_Occupancy_fact table, e.g. 1, 2, 3…	This is a Surrogate key, that will be populated through Sequence
	Time_Id	NUMBER	Foreign Key	Time_dim.time_id	NUMBER	The values are linked to measure and Time_Dim, 1, 2, 3	This is the foreign key in the fact table.
	Care_Centre_Id	NUMBER	Foreign Key	Care_Centre_Dim, Care_Centre_Id	NUMBER	The values here are linked to measure and Care_Centre_Dim, 1, 2, 3	This is the foreign key in the fact table.
	Percentage of bed occupancy rate	NUMBER	None	This is a Measure so might require temporary tables to hold dates while working out	NUMBER	pseudocode: Occupied Bed divided by Total Bed Multiply by 100	Measure of the percentage of bed occupancy for the report, e.g. 60%
	Total number of bed occupancy	NUMBER	None	This is a Measure so might require temporary	NUMBER	Count(*) where Bed_satus	Measure of the number of Bed

				tables to hold data with calculation		=Occupied and Year =2023	occupied e.g. 45,
	Total number of available bed	NUMBER	None	This is a Measure so might require temporary tables to hold data with calculation	NUMBER	Pseudocode: Count(*) where bed_status = Available and Month=November_	Measure of available bed per Month,Year

EXTRACT, TRANSFORM AND LOAD DATA DICTIONARY

Extract, Transform and Load is the process of data moving from one system to another which can also include integration, presentations and movement of data (Christopher Adams 2010). Extract, Load and Transform (ETL) processes are important in the component of data integration which enables data to be extracted and transformed from various sources of data to a final location referred to as a data mart or data warehouse. Data from its sources would possibly have “inconsistence and redundancies” which is required to be solved to enable an organisation to have a collective view of their data (Calvanese, D.et al,2001).

Extract					Transform			Load			
Database Source (s)	**Field Attribute**	**Data Type**	**Key**	**Table Name**	**Data Quality Check**	**Data quality Issues**	**Action Note**	**SS table**	**SS Column/Data Type**	**Key**	**Notes**
Definition:	Care_Centre_name represents the names of various care homes in the NYR AND WYR Care_Centre.										
Notes:	Care_Centre_name columns exist in both data sources										
NYR_TABLE		Varchar	No	NYR_Care_Centre Table	Consistency	None	None	**Care_Centre_Dim**	Care_Centre_Id /NUMBER	PK	Primary key extract from NYR And WYR Care_Centr introduced i the staging area.
WYR_TABLE	**Care_centre_name**	Varchar	No	WYR_ Care_Centre Table	Consistency	None	None	**Care_Centre_Dim**	Care_centre_name /VARCHAR		Insert value from the staging area table values from tables NYR And

											WYR Care_Centr
Definition:	Post Code represent the location of the Carehomes in NYR and WYR Care Centre										
Notes:	Post Code exists in both data sources.							**SS Table**	**SS Column/Data Type**	**Key**	**Notes**
NYR_TABLE	**Postcode**	Varchar	No	NYR_Care_Centre Table	Consistency	None	None	**Care_Centre_Dim**	Care_Centre_Id /NUMBER		Primary key extract from NYR And WYR Care_Centr introduced i the staging area.
WYR_TABLE		Varchar	No	WYR_Care_Centre_Table	Consistency	None	None	**Care_Centre_Dim**	Postcode/Varchar		Insert value from the staging area table NYR And WYR Care_Centr

Definition:	Town represent the location of the Care homes in NYR and WYR Care Centre.									
Notes:	Town column exists in both data sources.							**SS Table**	**SS Column/Data Type**	**Key**
NYR_TABLE	**Town**	Varchar	No	NYR_Care_Centre Table	Consistency	None	None	**Care_Centre_Dim**	Care_Centre_Id/NUMBER	
WYR_TABLE		Varchar	No	WYR_Care_Centre_Table	Consistency	None	None	**Care_Centre_Dim**	Town/Varchar	

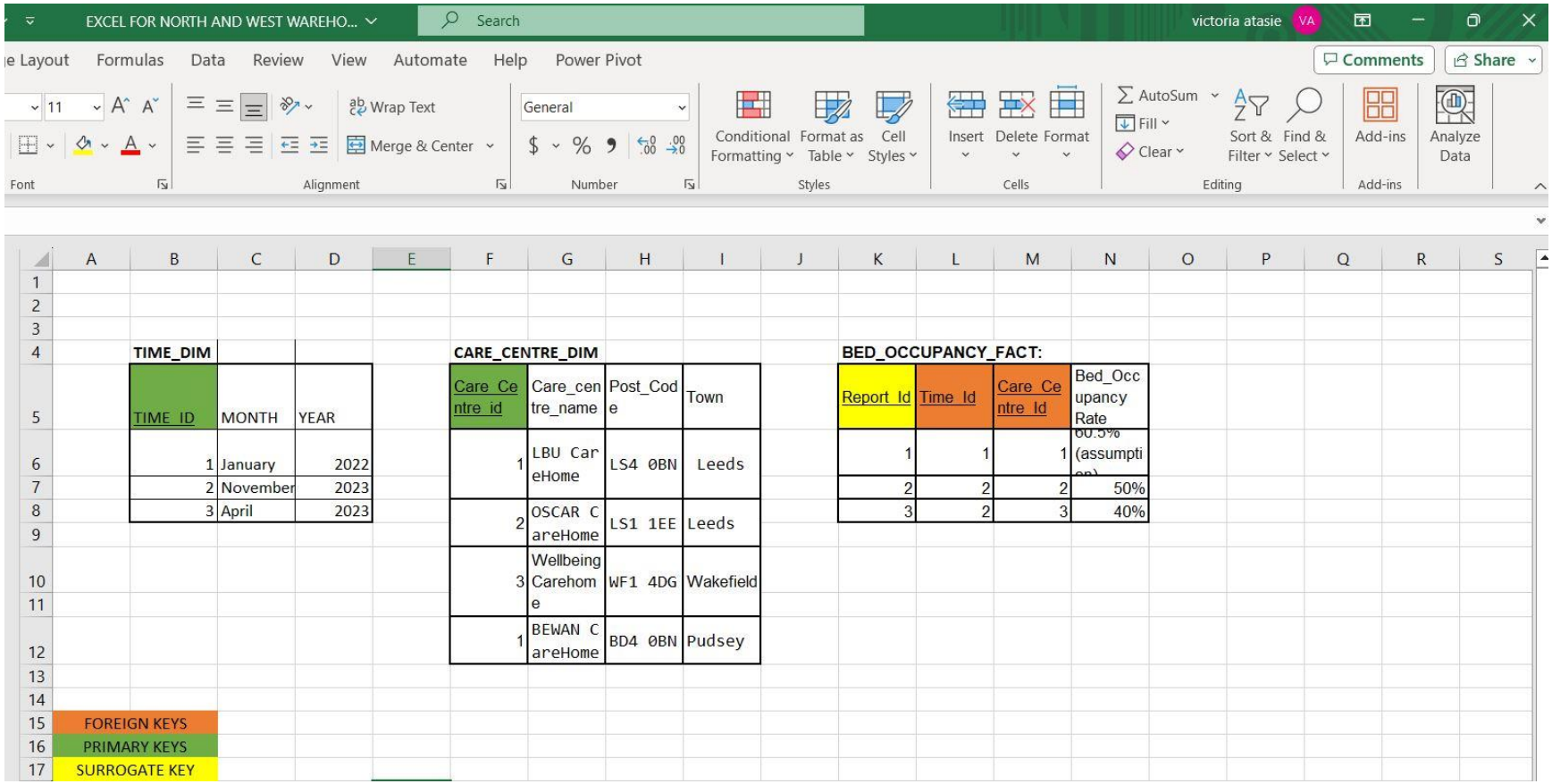
Definition:	Bed status represents the bed occupied or bed available in the care homes of NYR and WYR Care Centre.										
Notes:	Bed status column exist in both data sources.							**SS Table**	**SS Column/Data Type**	**Key**	**Notes**
NYR_TABLE	**Bed_Status**	Varchar	No	NYR_Bed	Consistency	None	None	**Bed_Occupancy_Fact**	Report_Id/NUMBER	Surrogate Key	Surrogate key is populated using sequence which is introduced in the staging area.
WYR_TABLE		Varchar	No	WYR_Bed	Consistency	None	None	**Bed_Occupancy_Fact**	Pecentage_Bed_Occupancy_rate/Number Total Number of Bed/Number Total Number of Available Bed per Month/Number Total Number of bed available per Year/Number		The bed status is used as the attribute in calculating the measure from the fact table.
Definition:	Bed status represents the bed occupied or bed available in the care homes of NYR and WYR Care Centre.										

Notes:	Bed status column exist in both data sources.							**SS Table**	**SS Column/Data Type**	**Key**	**Notes**
NYR_TABLE	**Bed_Status**	Varchar	No	NYR_Bed	Consistency	None	None	**Bed_Occupancy_Fact**	Report_Id/NUMBER	Surrogate Key	Surrogate key is populated using sequence which is introduced in the staging area.
WYR_TABLE		Varchar	No	WYR_Bed	Consistency	None	None	**Bed_Occupancy_Fact**	Pecentage_Bed_Occupancy_rate/Number Total Number of Bed/Number Total Number of Available Bed per Month/Number Total Number of bed available per Year/Number		The bed status is used as the attribute in calculating the measure from the fact table.
Definition:	Admission_Date represents the admission date when patient is admitted leading to the bed occupied determined.										
Notes:	Admission_Date columns exist in both data sources.							**SS Table**	**SS Column/Data Type**	**Key**	**Notes**

NYR_TABLE	**Admission_Date**	Date	No	NYR_Admission Table	Consistency	NYR_Admission data formatdiffers from WYR_Reservation data format.	Extract Year and Month from admission date column from NYR_Admision	**Time_Dim**	Time_Id/NUMBER	Surrogate Key	Surrogate key is populated using sequence which is introduced in the staging area.
									Month/Varchar Year/Number		

WYR_TABLE		Date	No	WYR_Reservation Table	Inconsistency	Missing Values of the Admission Date in the WYR_Table	Extract Year and Month from admission date column in the WYR_Reservation.	**Time_Dim**	Month/Varchar Year/Number		Insert value from the staging area table NYR_Admission And WYR_Reservation and cleaned and transformed.

SAMPLE DATA FOR STAR SCHEMA TABLE



TIME_DIM

TIME_ID	MONTH	YEAR
1	January	2022
2	November	2023
3	April	2023

CARE_CENTRE_DIM

Care_Centre_id	Care_centre_name	Post_Code	Town
1	LBU CareHome	LS4 0BN	Leeds
2	OSCAR CareHome	LS1 1EE	Leeds
3	Wellbeing Carehome	WF1 4DG	Wakefield
1	BEWAN CareHome	BD4 0BN	Pudsey

BED_OCCUPANCY_FACT:

Report_Id	Time_Id	Care_Centre_Id	Bed_Occupancy Rate
1	1	1	60.5% (assumption)
2	2	2	50%
3	2	3	40%

FOREIGN KEYS
PRIMARY KEYS
SURROGATE KEY

Figure 3: EXCEL SAMPLE DATA

DATA INTEGRATION(ETL)AND MAINTENANCE

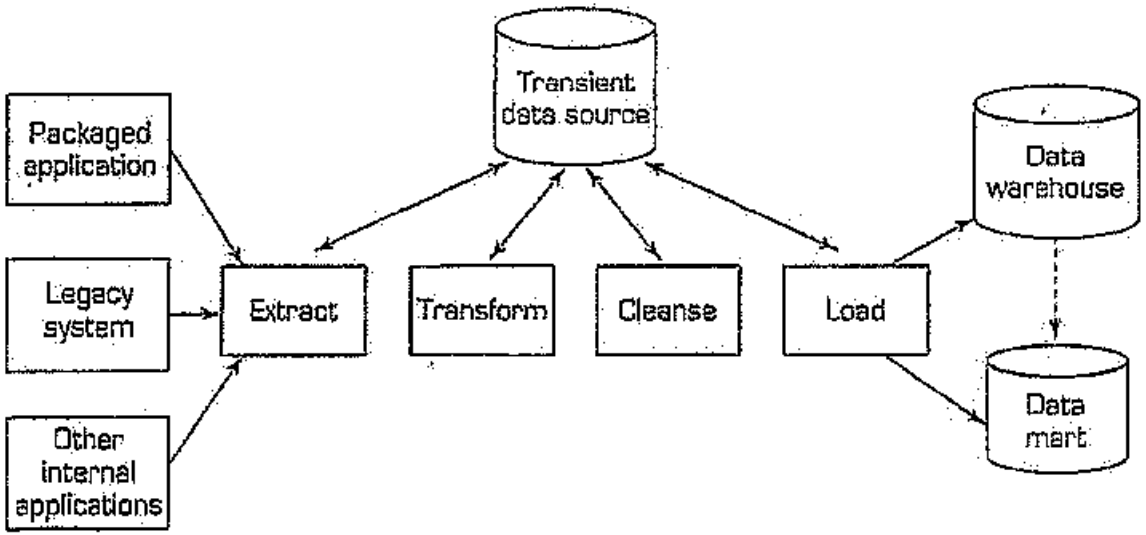
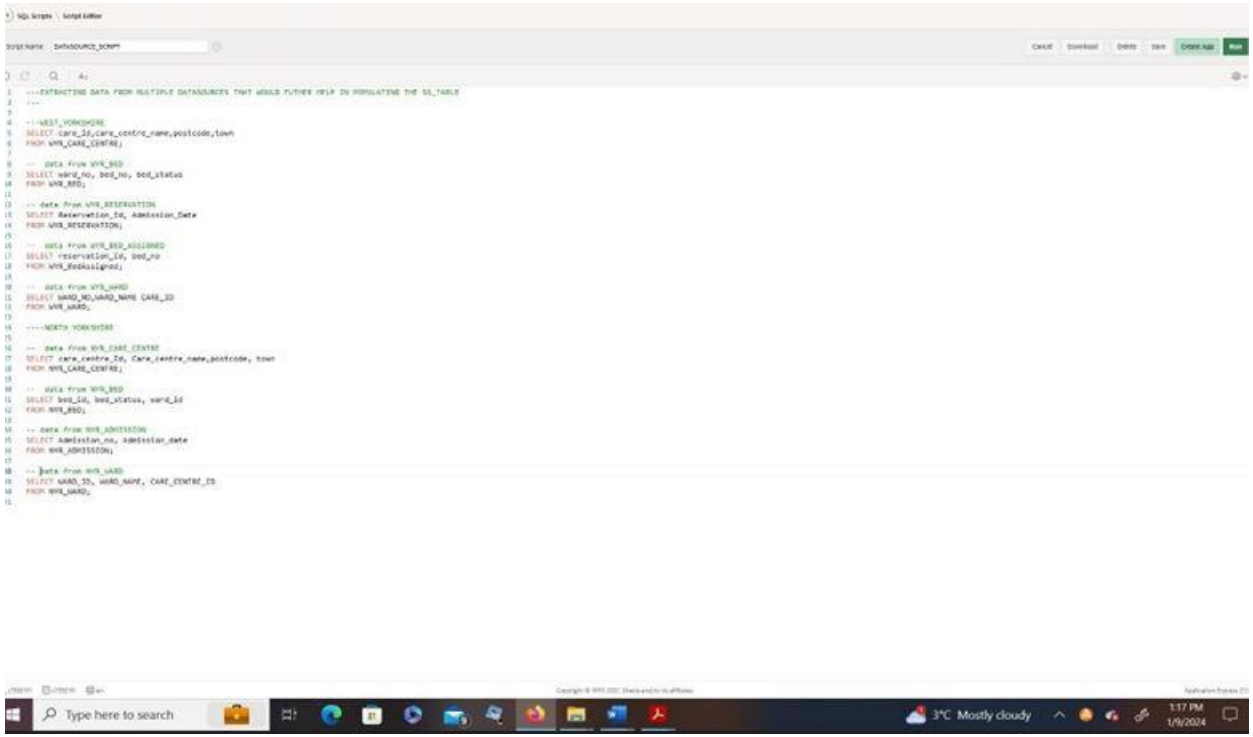


Figure 4: ETL PROCESS:(Source: From Decision Support and Business Intelligence System -Efraim Turban,2011)

Original data sets that are used for star schema includes the table and columns highlighted and would be used in the staging area. This phase of the ETL process involved the extraction of data from the West Yorkshire and North Yorkshire datasets as a multiple source of data which would help contribute to the analytical report process.

In this screenshot below I explore the data for the West Yorkshire and North Yorkshire data using SELECT, FROM STATEMENT to have a view of data for extraction process.



STAR SCHEMA IMPLEMENTATION

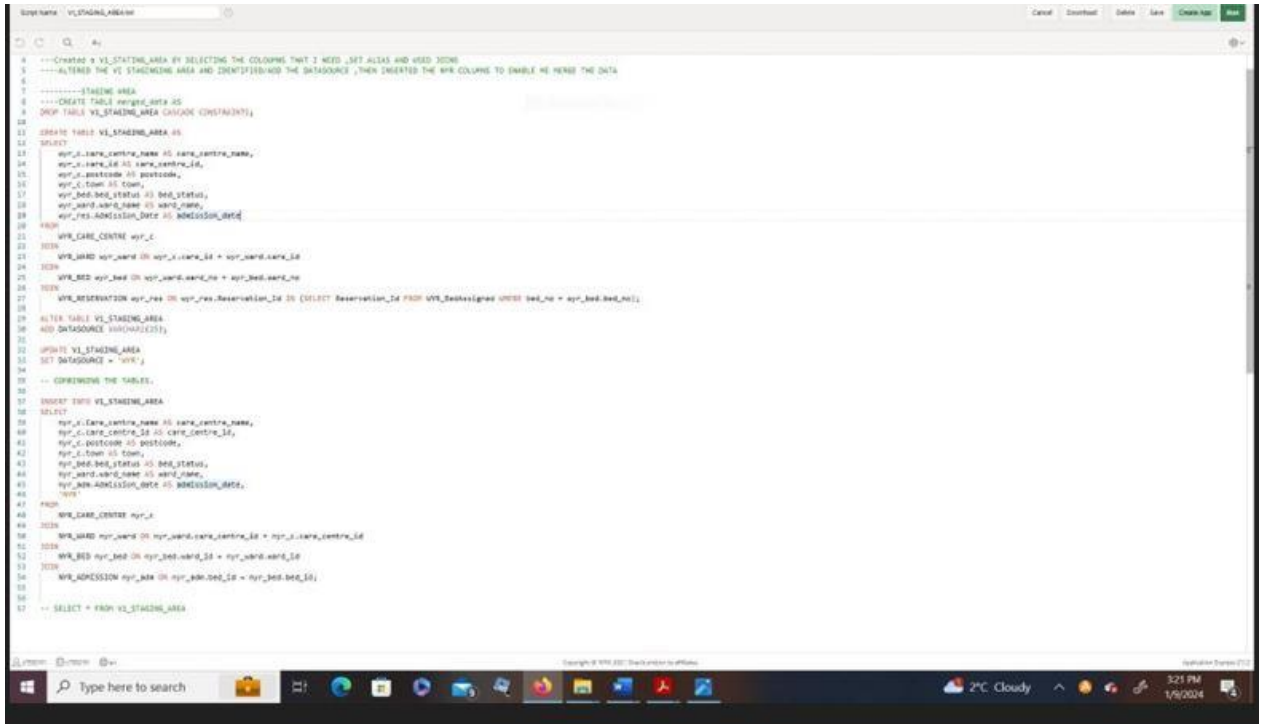
Star schema includes dimensional design for relational database are categorized as columns of the dimension table and the fact table contains facts that are stored as columns (Christopher Adams ,2010). A data warehouse design involves the conceptualization of dimensional modelling, and a dimensional modelling is the “retrieval-based system” which allows a huge amount of query accessibility. The fact table contains the features required for analytical process for decision, foreign keys are connected to the dimensional tables and attributes are used for querying a report (Turban, E., 2011).

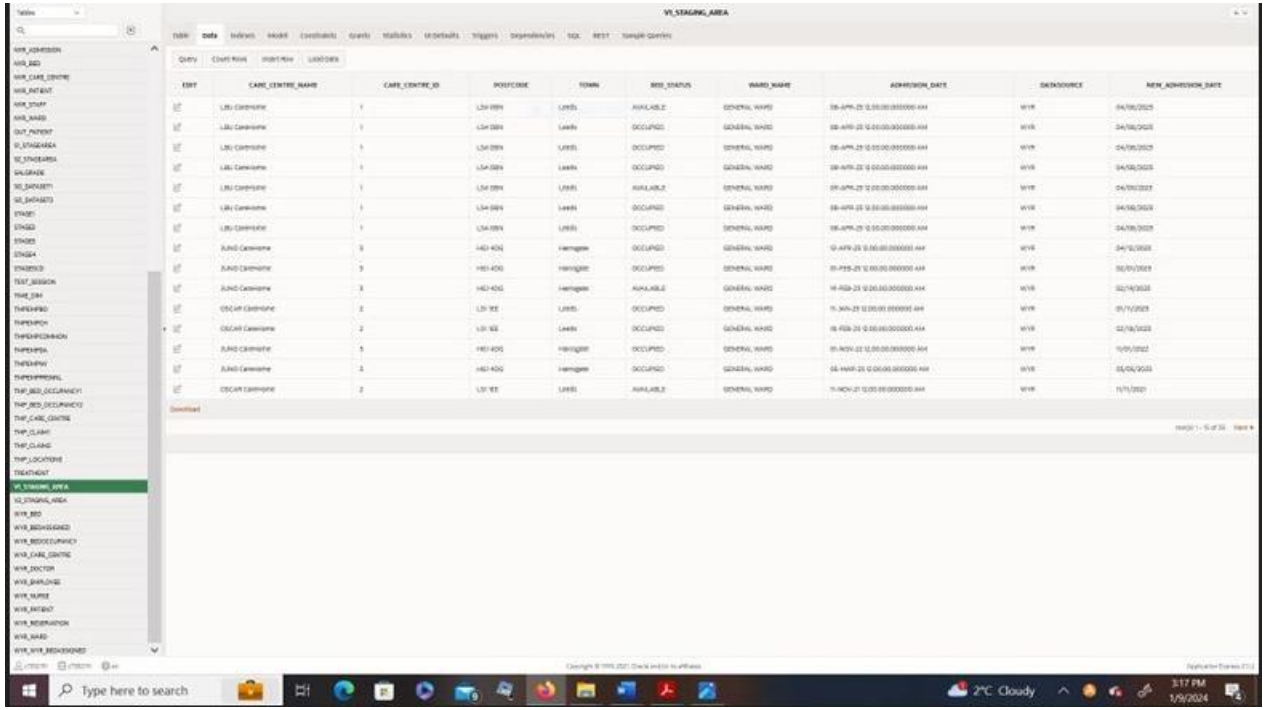
The star schema above was designed, and I proceeded to forward engineer my star schema and implemented sequence for my surrogate key. The measures in the Bed_Occupancy_Fact table would be percentage of bed occupied which would be drilled down to per ward per care centre and total bed occupied per ward per care centre. The evidence code would be showed alongside the Load process below.

The Time_Dim contains the Time_Id Month and Year and the Care_centre_Dim contains the Care_centre_Id, care_centre_name, town, ward_name and post_code.

STAGING AREA

The staging area approach was implemented as V1_staging_area to serve as an intermediate stage of combining the data extracted from multiple sources before the load process in the star schema.


Cancel
Download
Delete
Save
Create App
Run
---Created a V1_STAGING_AREA BY SELECTING THE COLUMNS THAT I NEED ,SET ALIAS AND USED JOINS
----ALTERED THE V1 STAGEING AREA AND IDENTIFIED/ADD THE DATASOURCE ,THEN INSERTED THE NYR COLUMNS TO ENABLE ME MERGE THE DATA
---------STAGING AREA
----CREATE TABLE merged_data AS
DROP TABLE V1_STAGING_AREA CASCADE CONSTRAINTS;
CREATE TABLE V1_STAGING_AREA AS
SELECT
nyr_c.care_centre_name AS care_centre_name,
nyr_c.care_id AS care_centre_id,
nyr_c.postcode AS postcode,
nyr_c.town AS town,
nyr_bed.bed_status AS bed_status,
nyr_ward.ward_name AS ward_name,
nyr_res.Admission_Date AS admission_date
FROM
NYR_CARE_CENTRE nyr_c
JOIN
NYR_WARD nyr_ward ON nyr_c.care_id = nyr_ward.care_id
JOIN
NYR_BED nyr_bed ON nyr_ward.ward_no = nyr_bed.ward_no
JOIN
NYR_RESERVATION nyr_res ON nyr_res.Reservation_Id IN (SELECT Reservation_Id FROM NYR_Bedsassigned WHERE bed_no = nyr_bed.bed_no);
ALTER TABLE V1_STAGING_AREA
ADD DATASOURCE VARCHAR(255);
UPDATE V1_STAGING_AREA
SET DATASOURCE = 'NYR';
-- COMBINING THE TABLES.
INSERT INTO V1_STAGING_AREA
SELECT
nyr_c.Care_centre_name AS care_centre_name,
nyr_c.care_centre_id AS care_centre_id,
nyr_c.postcode AS postcode,
nyr_c.town AS town,
nyr_bed.bed_status AS bed_status,
nyr_ward.ward_name AS ward_name,
nyr_adm.Admission_date AS admission_date,
'NYR'
FROM
NYR_CARE_CENTRE nyr_c
JOIN
NYR_WARD nyr_ward ON nyr_ward.care_centre_id = nyr_c.care_centre_id
JOIN
NYR_BED nyr_bed ON nyr_bed.ward_id = nyr_ward.ward_id
JOIN
NYR_ADMISSION nyr_adm ON nyr_adm.bed_id = nyr_bed.bed_id;
-- SELECT * FROM V1_STAGING_AREA
Type here to search
2°C Cloudy
3:21 PM
1/9/2024



Further discussion, the V1_staging_are was created and selected columns care_centre_name, centre_centre_id, postcode,town,bed_status,ward_name,admission_date and used a JOIN based on the relationship between the tables ,altered and updated with a datasource and perform an insert and select statement into the V1_staging area.

ALTER TABLE V1_STAGING_AREA ADD NEW_ADMISSION_DATE DATE; This was used to add a new_admission_date column considering the format of the initial admission_date column by separating the date and time using the code below.

```
UPDATE V1_STAGING_AREA
SET NEW_ADMISSION_DATE = TO_DATE(TO_CHAR(ADMISSION_DATE, 'DD-MON-YY'), 'DD-MON-YY');
```

TRANSFORMATION, QUALITY CHECKS AND TEMPORAL TABLE

The transformation involves data undergoing cleaning, data validation, normalization, ensuring accuracy and consistency of data to avoid data quality issues and maintain data integrity.

Further discussion, missing data was checked using SELECT *
FROM V1_STAGING_AREA WHERE CARE_CENTRE_NAME IS NULL OR POSTCODE IS NULL OR TOWN IS NULL OR BED_STATUS IS NULL OR WARD_NAME IS NULL OR ADMISSION_DATE IS NULL.
The Admission_date appeared that there were missing values and other showed no missing values.
DELETE FROM V1_STAGING_AREA WHERE ADMISSION_DATE IS NULL;was used to delete null values.

```
-- Code was used in checking for Duplicates values
--SELECT CARE_CENTRE_NAME, CARE_CENTRE_ID, POSTCODE,
TOWN, BED_STATUS, WARD_NAME, ADMISSION_DATE, COUNT(*)
--FROM V1_STAGING_AREA
--GROUP BY CARE_CENTRE_NAME,
CARE_CENTRE_ID, POSTCODE, TOWN, BED_STATUS, WARD_NAME, ADMISSION_DATE
--HAVING COUNT(*) > 1;
```

Consistency of data was checked such as :
Checking for Consistency in BED_STATUS

```
--SELECT DISTINCT BED_STATUS
--FROM V1_STAGING_AREA;

-- Check for Consistency in WARD_NAME
--SELECT DISTINCT WARD_NAME
--FROM V1_STAGING_AREA;
```

---Quality checks were also done, updating bed status for integrity, standardizations and all lowercase occupied to uppercase occupied

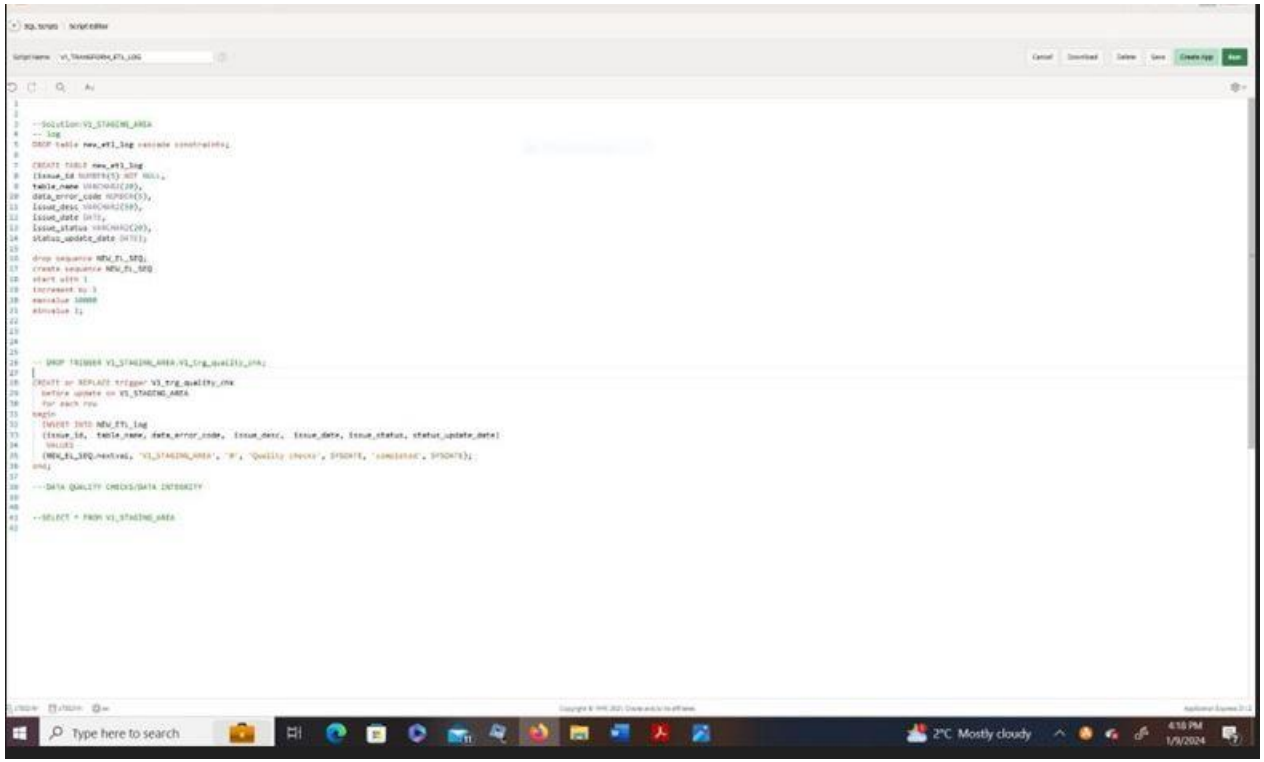
```
UPDATE V1_STAGING_AREA
SET BED_STATUS = 'OCCUPIED'
WHERE UPPER(BED_STATUS) = 'OCCUPIED' OR BED_STATUS = 'Occupied';
----Updated rows with 'NOT OCCUPIED' to 'Available'
UPDATE V1_STAGING_AREA
```

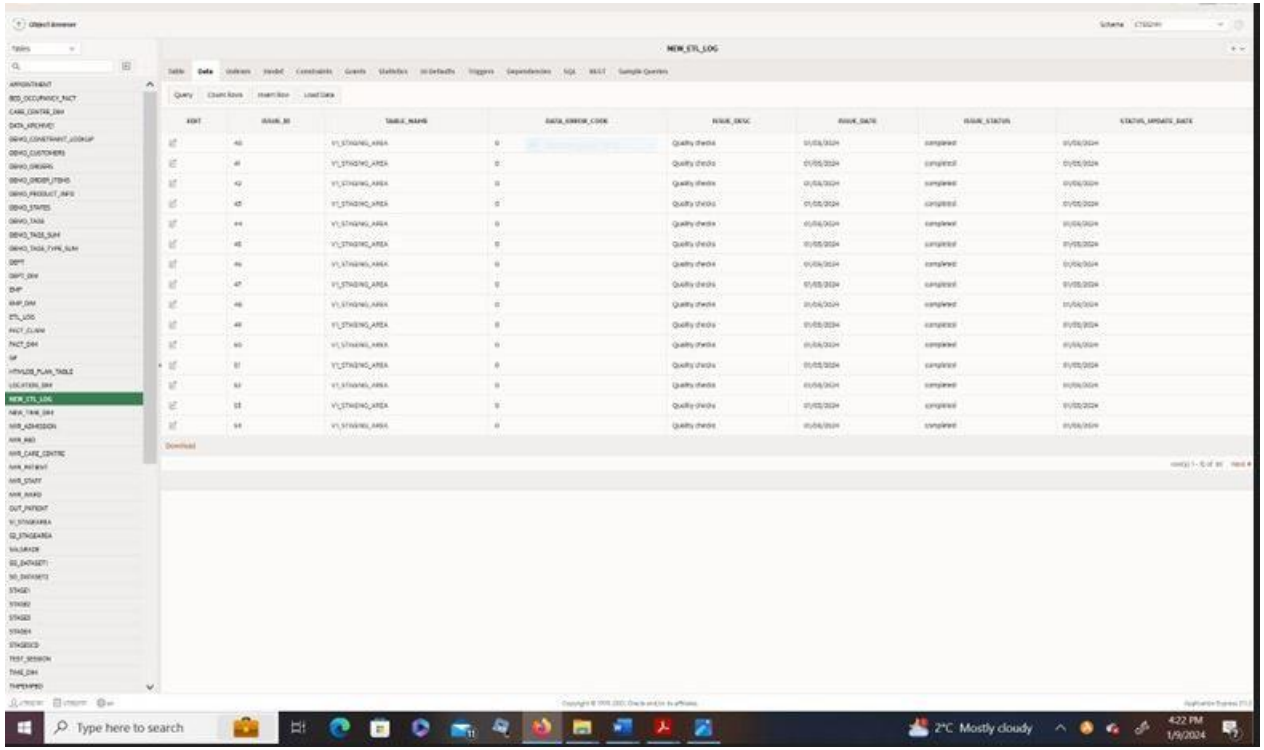
```
SET BED_STATUS = 'AVAILABLE'
WHERE UPPER(BED_STATUS) = 'NOT OCCUPIED';
----------Updated lowercase 'Available' to 'AVAILABLE'
UPDATE V1_STAGING_AREA
SET BED_STATUS = 'AVAILABLE'
WHERE UPPER(BED_STATUS) = 'AVAILABLE' OR BED_STATUS = 'Available';

UPDATE V1_STAGING_AREA
SET WARD_NAME = 'GENERAL WARD'
WHERE UPPER(WARD_NAME) = 'GENERAL CARE';
```

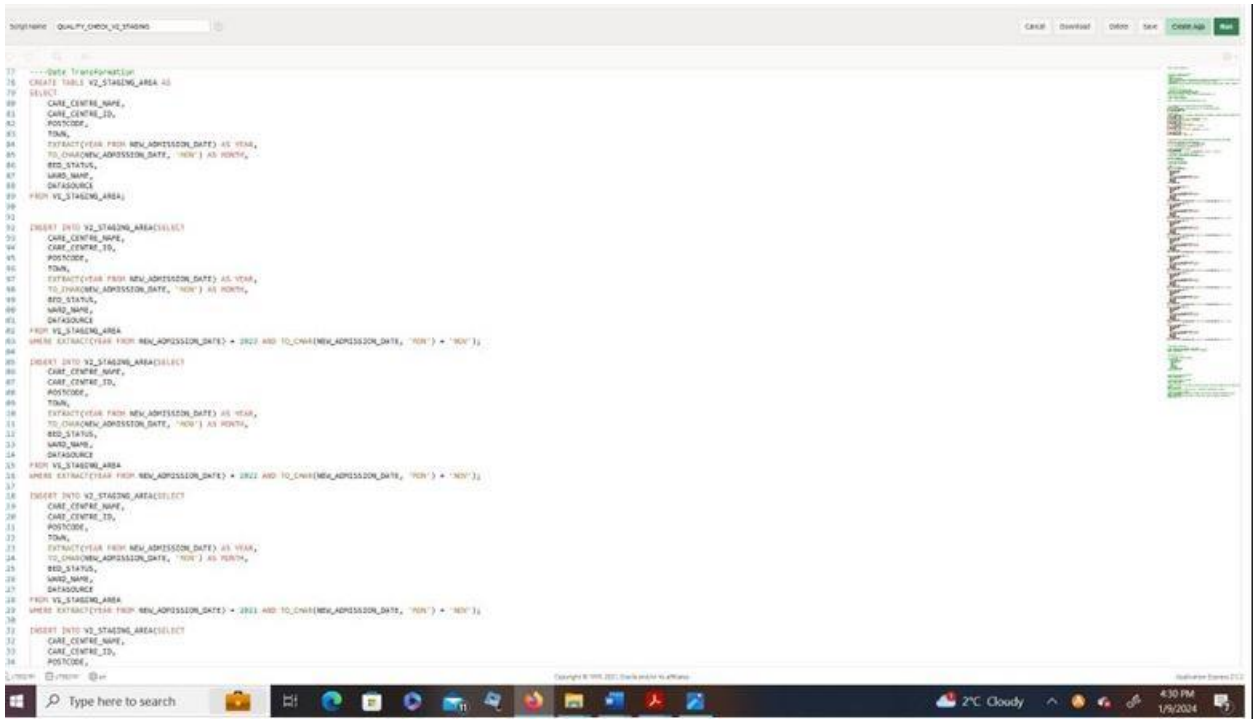
An ETL log was created to log in changes made in the data, created a trigger and

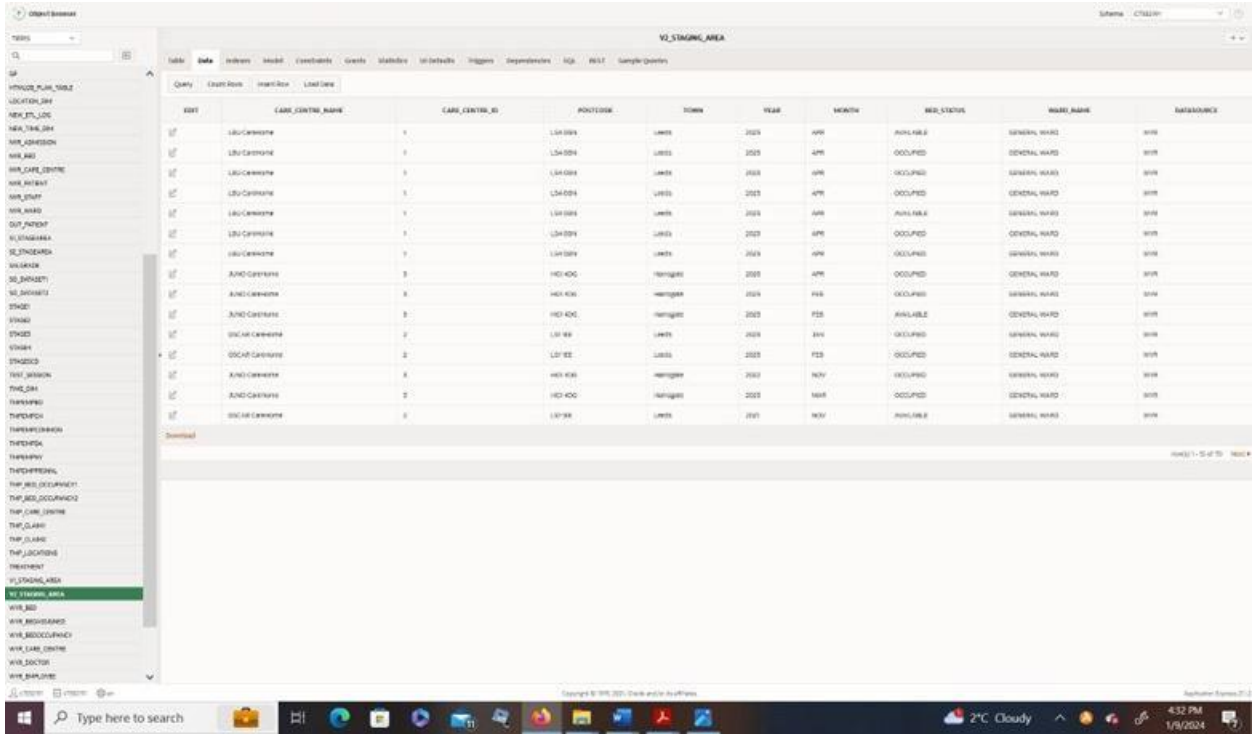
NEW_ETL_LOG


--Solution:V1_STAGING_AREA
-- log
DROP table new_etl_log cascade constraints;
CREATE TABLE new_etl_log
(issue_id NUMBER(5) NOT NULL,
table_name VARCHAR2(20),
data_error_code NUMBER(5),
issue_desc VARCHAR2(50),
issue_date DATE,
issue_status VARCHAR2(20),
status_update_date DATE);
drop sequence NEW_ETL_SEQ;
create sequence NEW_ETL_SEQ
start with 1
increment by 1
maxvalue 10000
minvalue 1;
-- DROP TRIGGER V1_STAGING_AREA.V1_trg_quality_chk;
CREATE or REPLACE trigger V1_trg_quality_chk
before update on V1_STAGING_AREA
for each row
begin
INSERT INTO NEW_ETL_log
(issue_id, table_name, data_error_code, issue_desc, issue_date, issue_status, status_update_date)
VALUES
(NEW_ETL_SEQ.nextval, 'V1_STAGING_AREA', '0', 'Quality checks', SYSDATE, 'completed', SYSDATE);
end;
---DATA QUALITY CHECKS/DATA INTEGRITY
--SELECT * FROM V1_STAGING_AREA
Type here to search
2°C Mostly cloudy
4:16 PM
1/9/2024



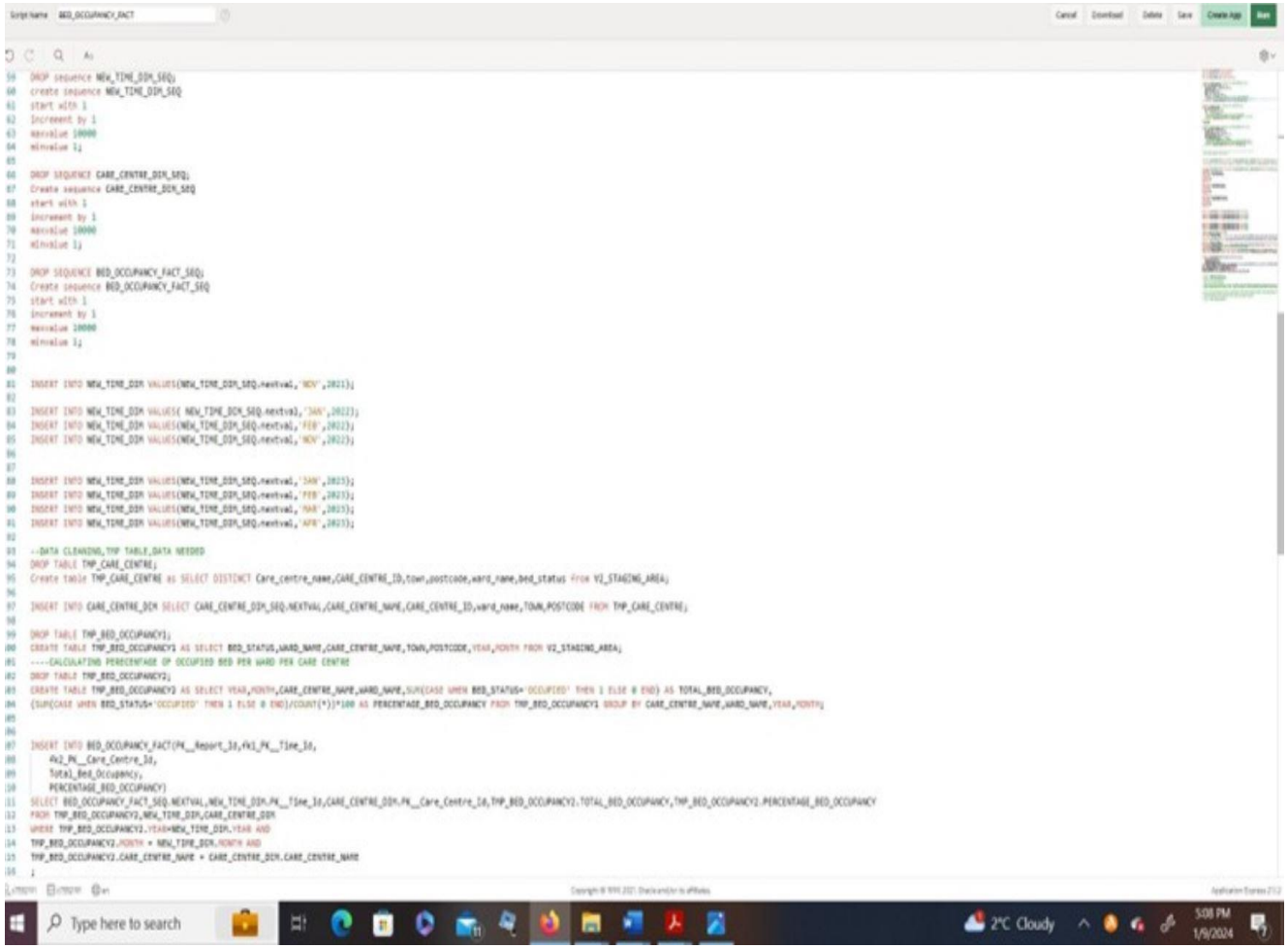
Transformation was done on the New_Admission_Date to extract Month and Year into separate columns and created a V2_Staging_Area .

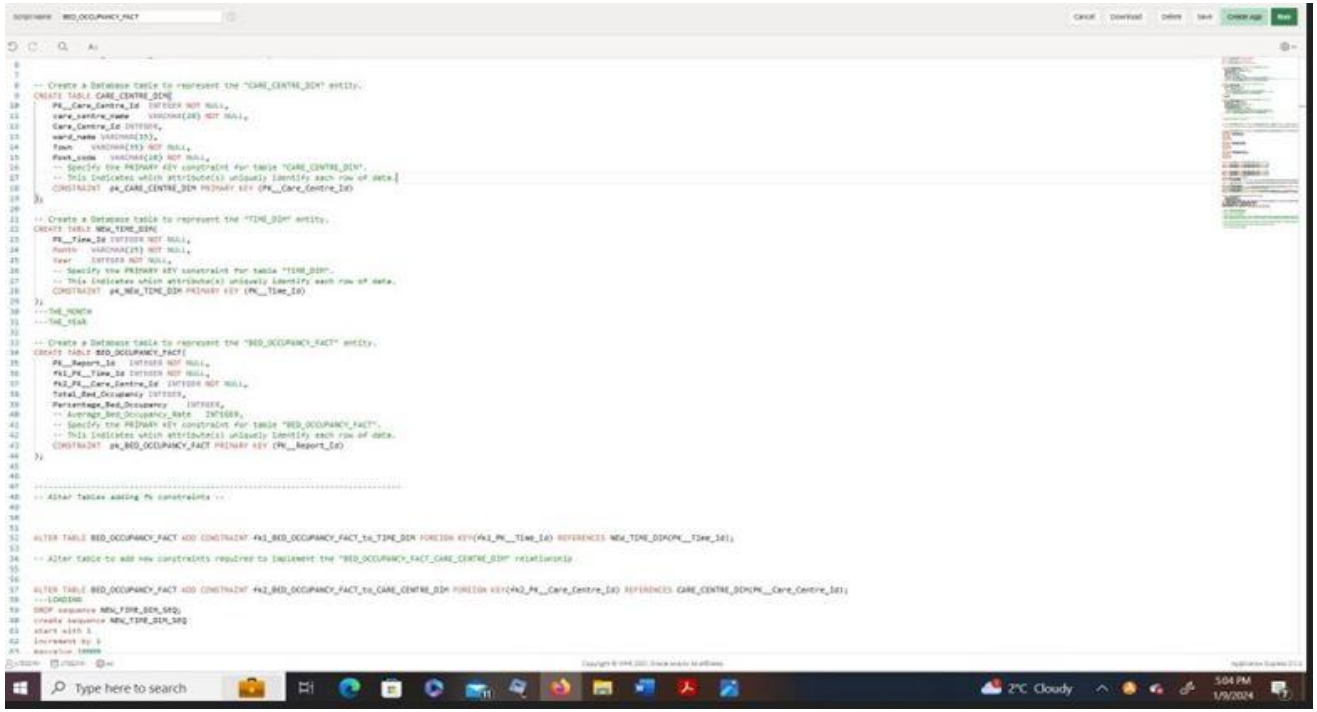

Cancel
Download
Delete
Save
Create App
Run
---Data Transformation
CREATE TABLE V2_STAGING_AREA AS
SELECT
CARE_CENTRE_NAME,
CARE_CENTRE_ID,
POSTCODE,
TOWN,
EXTRACT(YEAR FROM NEW_ADMISSION_DATE) AS YEAR,
TO_CHAR(NEW_ADMISSION_DATE, 'MON') AS MONTH,
BED_STATUS,
WARD_NAME,
DATASOURCE
FROM V1_STAGING_AREA;
INSERT INTO V2_STAGING_AREA(SELECT
CARE_CENTRE_NAME,
CARE_CENTRE_ID,
POSTCODE,
TOWN,
EXTRACT(YEAR FROM NEW_ADMISSION_DATE) AS YEAR,
TO_CHAR(NEW_ADMISSION_DATE, 'MON') AS MONTH,
BED_STATUS,
WARD_NAME,
DATASOURCE
FROM V1_STAGING_AREA
WHERE EXTRACT(YEAR FROM NEW_ADMISSION_DATE) = 2022 AND TO_CHAR(NEW_ADMISSION_DATE, 'MON') = 'NOV');
INSERT INTO V2_STAGING_AREA(SELECT
CARE_CENTRE_NAME,
CARE_CENTRE_ID,
POSTCODE,
TOWN,
EXTRACT(YEAR FROM NEW_ADMISSION_DATE) AS YEAR,
TO_CHAR(NEW_ADMISSION_DATE, 'MON') AS MONTH,
BED_STATUS,
WARD_NAME,
DATASOURCE
FROM V1_STAGING_AREA
WHERE EXTRACT(YEAR FROM NEW_ADMISSION_DATE) = 2021 AND TO_CHAR(NEW_ADMISSION_DATE, 'MON') = 'NOV');
INSERT INTO V2_STAGING_AREA(SELECT
CARE_CENTRE_NAME,
CARE_CENTRE_ID,
POSTCODE,
TOWN,
EXTRACT(YEAR FROM NEW_ADMISSION_DATE) AS YEAR,
TO_CHAR(NEW_ADMISSION_DATE, 'MON') AS MONTH,
BED_STATUS,
WARD_NAME,
DATASOURCE
FROM V1_STAGING_AREA
WHERE EXTRACT(YEAR FROM NEW_ADMISSION_DATE) = 2021 AND TO_CHAR(NEW_ADMISSION_DATE, 'MON') = 'NOV');
INSERT INTO V2_STAGING_AREA(SELECT
CARE_CENTRE_NAME,
CARE_CENTRE_ID,
POSTCODE,
Type here to search
2°C Cloudy
4:30 PM
1/9/2024

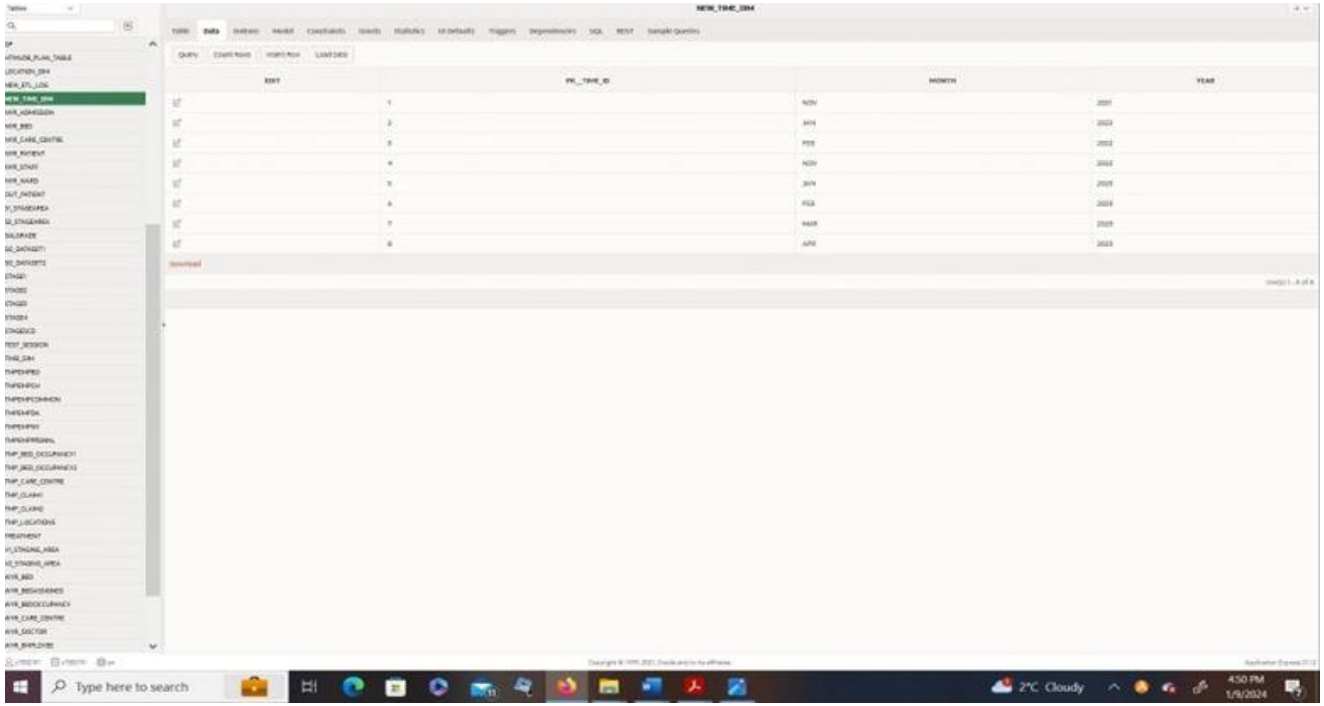

V2_STAGING_AREA
Type here to search
2°C Cloudy
4:32 PM
1/9/2024

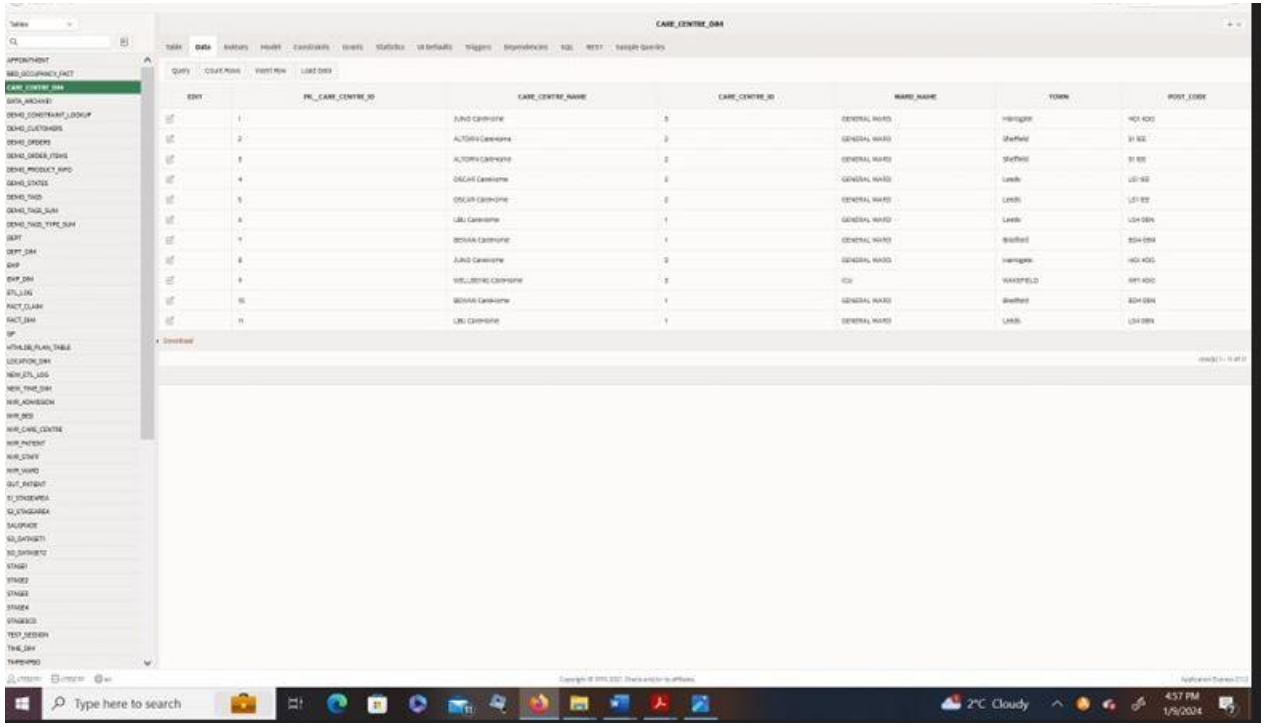
LOADING AND TEMPORAL TABLE

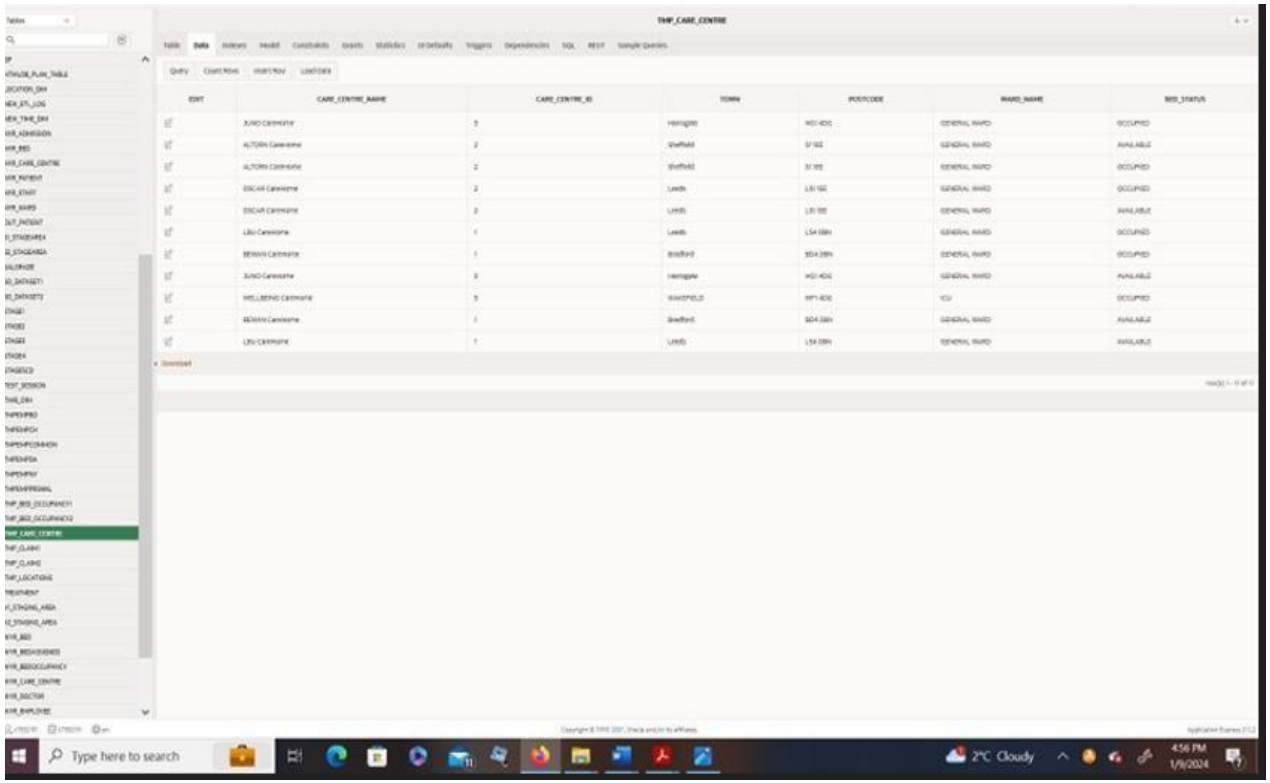
The loading phase of this ETL process would involve populating the dimension table (Time dimension), the fact table (Bed_Occupancy_fact),the surrogate keys are defined, the measures in the fact table would be calculated, sequences created, temporal table would be created, and insertion into the bed_occupancy_fact.

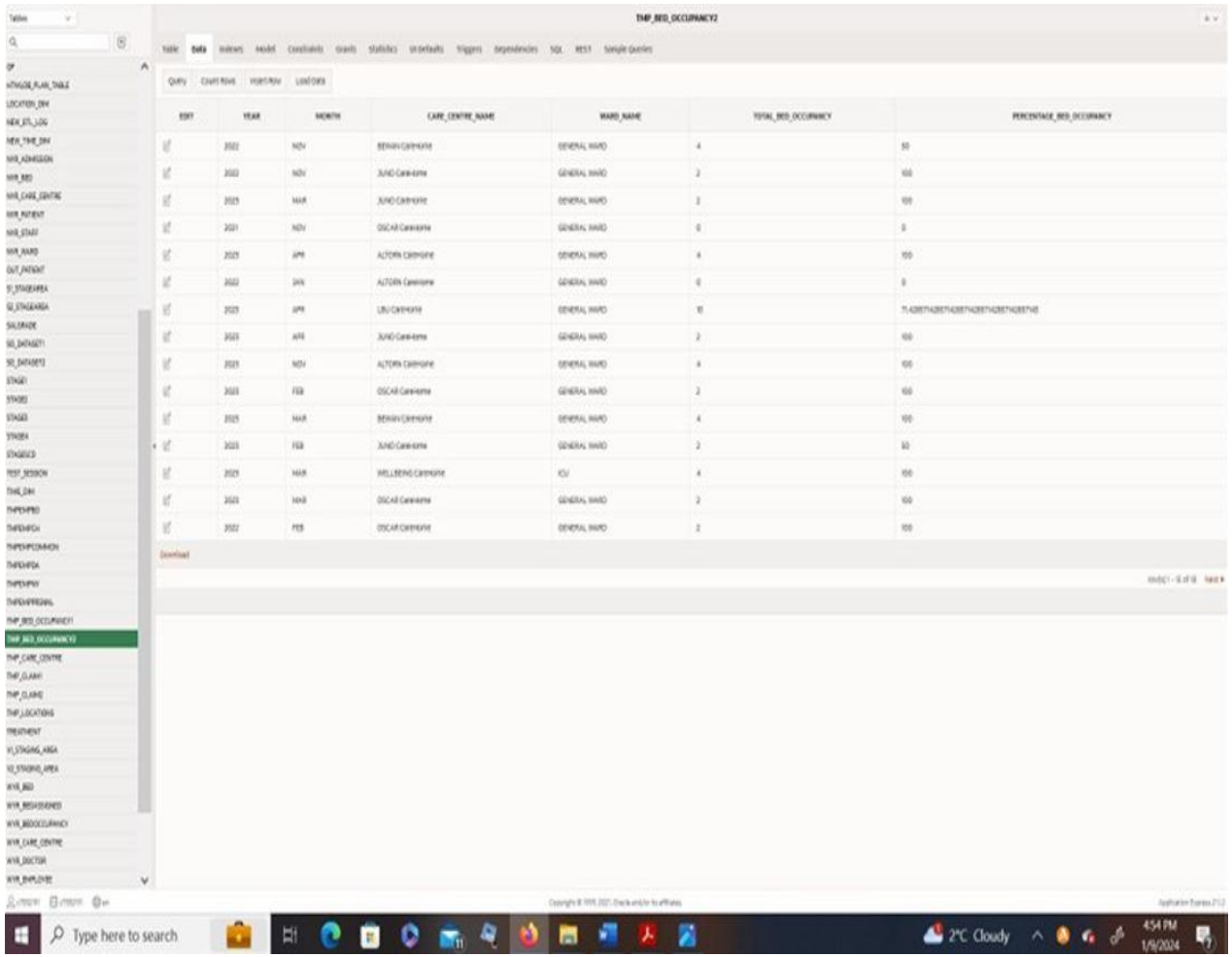


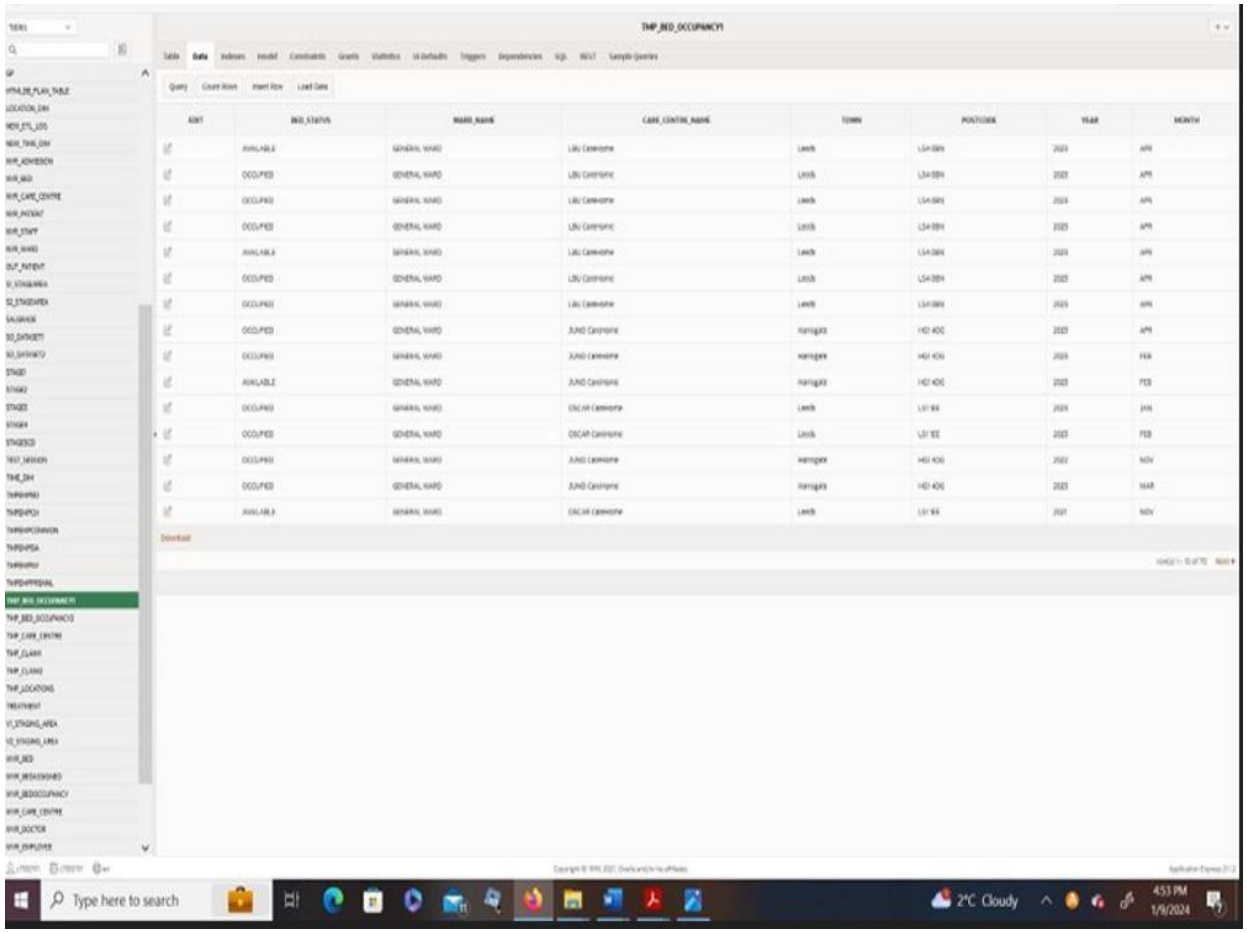

-- Create a Database table to represent the "CARE_CENTRE_DIM" entity.
CREATE TABLE CARE_CENTRE_DIM(
-- Specify the PRIMARY KEY constraint for table "CARE_CENTRE_DIM".
-- This indicates which attribute(s) uniquely identify each row of data.
-- Create a Database table to represent the "TIME_DIM" entity.
CREATE TABLE NEW_TIME_DIM(
-- Specify the PRIMARY KEY constraint for table "TIME_DIM".
-- This indicates which attribute(s) uniquely identify each row of data.
-- Create a Database table to represent the "BED_OCCUPANCY_FACT" entity.
CREATE TABLE BED_OCCUPANCY_FACT(
-- Specify the PRIMARY KEY constraint for table "BED_OCCUPANCY_FACT".
-- This indicates which attribute(s) uniquely identify each row of data.
-- Alter Tables adding Fk constraints --
-- Alter table to add new constraints required to implement the "BED_OCCUPANCY_FACT_CARE_CENTRE_DIM" relationship
Type here to search
2°C Cloudy
5:04 PM
1/9/2024


NEW_TIME_DIM
Data
Query
Count Rows
Insert Row
Load CSV
EDIT
PK_TIME_ID
MONTH
YEAR
Download
Type here to search
2°C Cloudy
4:50 PM
1/9/2024

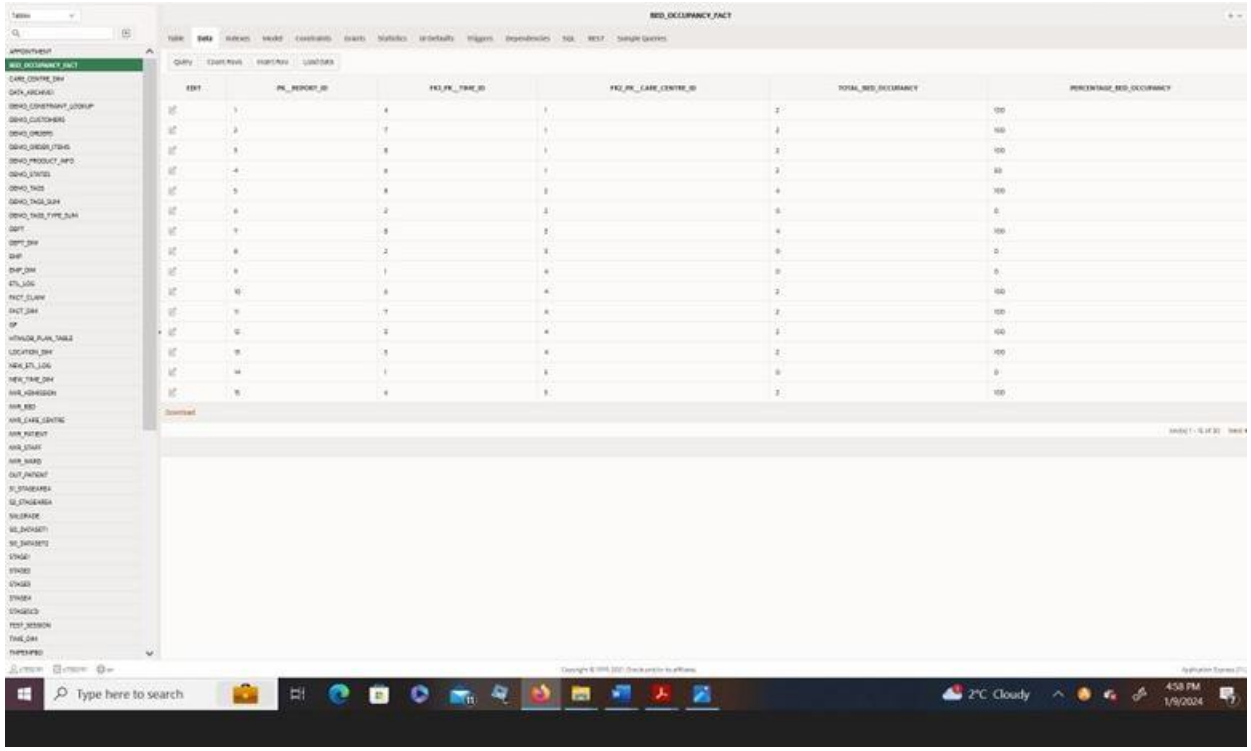

CARE_CENTRE_DIM
Data
Type here to search
2°C Cloudy
4:57 PM
1/9/2024


TMP_CARE_CENTRE
Table
Data
Indexes
Model
Constraints
Grants
Statistics
Triggers
Dependencies
SQL
REST
Sample Queries
Query
Count Rows
Insert Row
Load Data
EDIT
CARE_CENTRE_NAME
CARE_CENTRE_ID
TOWN
POSTCODE
WARD_NAME
BED_STATUS
Download
Type here to search
2°C Cloudy
4:56 PM
1/9/2024


TMP_BED_OCCUPANCY2
Table
Data
Indexes
Model
Constraints
Grants
Statistics
UI defaults
Triggers
Dependencies
SQL
REST
Sample Queries
Query
Count Rows
Insert Row
Load Data
EDIT
YEAR
MONTH
CARE_CENTRE_NAME
WARD_NAME
TOTAL_BED_OCCUPANCY
PERCENTAGE_BED_OCCUPANCY
Download
TMP_BED_OCCUPANCY1
TMP_BED_OCCUPANCY2
TMP_CARE_CENTRE
TMP_CLAIM
TMP_CLAIMS
TMP_LOCATIONS
TREATMENT
V_STAGING_AREA
V2_STAGING_AREA
WVR_BED
WVR_BEDOCCUPANCY
WVR_CARE_CENTRE
WVR_DOCTOR
Type here to search
2°C Cloudy
4:54 PM
1/9/2024


TMP_BED_OCCUPANCY1
Table
Data
Indexes
Model
Constraints
Grants
Statistics
Id Defaults
Triggers
Dependencies
SQL
REST
Sample Queries
Query
Count Rows
Insert Row
Load Data
EDIT
BED_STATUS
WARD_NAME
CARE_CENTRE_NAME
TOWN
POSTCODE
YEAR
MONTH
Download
Type here to search
2°C Cloudy
4:53 PM
1/9/2024

REPORT



Further discussion: I inserted into New_Time_Dim sequence,month and year individually ,progressed to create a Temporal Care Centre using Distinct to select columns ensuring no duplicate values are inserted from the V2_Staging_Area.I inserted from the TMP_CARE_CENTRE INTO THE CARE_CENTRE_DIM, created the TMP_BEDOCCUPANCY1 AND TMP_BEDOCCUPANCY2,performed the calculations for percentage of bed occupied per ward per carehome and total bed occupied per ward per carehome and loaded into the star schema.

Slowing Changing Dimension (SCD)

KIMBALL TYPE1 which involves altering or changing the value that is wrong to a value that is corrected (Kimball,2013)

A	B	C	D	E	F	G	H
CARE_CENTRE_ID		CARE_CENTRE_NAME		TOWN	EMAIL	PHONE	POSTCODE
1		BEWAN CAREHOME		BRADFORD	BEWANCAREHOME@GMAIL.COM	(088)224-34125	S1 1EE
CARE_CENTRE_ID		CARE_CENTRE_NAME		TOWN	EMAIL	PHONE	POSTCODE
260		BEWAN CAREHOME		BRADFORD	BEWANCAREHOME@GMAIL.COM	(088)244-34125	BD1 ZBN

Figure 5: Sample Data Type 1

KIMBALL TYPE 2(SECOND APPROACH) which involves creating a new row added to the previous information.

CARE_CENTRE_ID	CARE_CENTRE_NAME	TOWN	EMAIL	PHONE	POSTCODE
1	BEWAN CAREHOME	BRADFORD	BEWANCAREHOME@GMAIL.COM	(088)224-34125	S1 1EE
1	BEWAN CAREHOME	BRADFORD	BEWANCAREHOME@GMAIL.COM	(088)244-34125	BD1 ZBN

Figure 6: Sample Data Type 2

The type 3 and a final approach would be type 3. It involves an approach to add a new column to the previous column with the postcode.

CARE_CENTRE_ID	CARE_CENTRE_NAME	TOWN	EMAIL	PHONE	POSTCODE	PRIOR POSTCODE
1	BEWAN CAREHOME	BRADFORD	BEWANCAREHOME@GMAIL.COM	(088)224-34125	BD1 ZBN	S1 1EE

Figure 7: Sample Data Type 3

APPENDIXES:

```
---------STAGING AREA
----CREATE TABLE merged_data AS
DROP TABLE V1_STAGING_AREA CASCADE CONSTRAINTS;

CREATE TABLE V1_STAGING_AREA AS
```

```
SELECT
    wyr_c.care_centre_name AS care_centre_name,
    wyr_c.care_id AS care_centre_id,
    wyr_c.postcode AS postcode,
    wyr_c.town AS town,
    wyr_bed.bed_status AS bed_status,
    wyr_ward.ward_name AS ward_name,
    wyr_res.Admission_Date AS admission_date
FROM
    WYR_CARE_CENTRE wyr_c
JOIN
    WYR_WARD wyr_ward ON wyr_c.care_id = wyr_ward.care_id
JOIN
    WYR_BED wyr_bed ON wyr_ward.ward_no = wyr_bed.ward_no
JOIN
    WYR_RESERVATION wyr_res ON wyr_res.Reservation_Id IN (SELECT Reservation_Id FROM WYR_BedAssigned W
HERE bed_no = wyr_bed.bed_no);

ALTER TABLE V1_STAGING_AREA
ADD DATASOURCE VARCHAR2(25);

UPDATE V1_STAGING_AREA
SET DATASOURCE = 'WYR';

-- COMBINNING THE TABLES.

INSERT INTO V1_STAGING_AREA
SELECT
    nyr_c.Care_centre_name AS care_centre_name,
    nyr_c.care_centre_id AS care_centre_id,
    nyr_c.postcode AS postcode,
    nyr_c.town AS town,
```

```
    nyr_bed.bed_status AS bed_status,
    nyr_ward.ward_name AS ward_name,
    nyr_adm.Admission_date AS admission_date,
    'NYR'
FROM
    NYR_CARE_CENTRE nyr_c
JOIN
    NYR_WARD nyr_ward ON nyr_ward.care_centre_id = nyr_c.care_centre_id
JOIN
    NYR_BED nyr_bed ON nyr_bed.ward_id = nyr_ward.ward_id
JOIN
    NYR_ADMISSION nyr_adm ON nyr_adm.bed_id = nyr_bed.bed_id;


-- SELECT * FROM V1_STAGING_AREA


- ETL_log
DROP table new_etl_log cascade constraints;

CREATE TABLE new_etl_log
(issue_id NUMBER(5) NOT NULL,
table_name VARCHAR2(20),
data_error_code NUMBER(5),
issue_desc VARCHAR2(50),
issue_date DATE,
issue_status VARCHAR2(20),
status_update_date DATE);

drop sequence NEW_EL_SEQ;
create sequence NEW_EL_SEQ
start with 1
increment by 1
```

```
maxvalue 10000
minvalue 1;



-- DROP TRIGGER V1_STAGING_AREA.V1_trg_quality_chk;

CREATE or REPLACE trigger V1_trg_quality_chk
  before update on V1_STAGING_AREA
  for each row
begin
  INSERT INTO NEW_ETL_log
  (issue_id,  table_name, data_error_code,  issue_desc,  issue_date, issue_status, status_update_date)
   VALUES
  (NEW_EL_SEQ.nextval, 'V1_STAGING_AREA', '0', 'Quality checks', SYSDATE, 'completed', SYSDATE);
end;



DELETE FROM V1_STAGING_AREA
WHERE ADMISSION_DATE IS NULL;

---QUALITY CHECKS
---UPDATING THE BED STATUS FOR DATA INTEGRITY,
STANDARDLIZING DATA CHANGING ALL LOWERCASE OCCUPIED TO UPPERCASE Occupied
UPDATE V1_STAGING_AREA
SET BED_STATUS = 'OCCUPIED'
WHERE UPPER(BED_STATUS) = 'OCCUPIED' OR BED_STATUS = 'Occupied';
----Updated rows with 'NOT OCCUPIED' to 'Available'
UPDATE V1_STAGING_AREA
SET BED_STATUS = 'AVAILABLE'
```

```
WHERE UPPER(BED_STATUS) = 'NOT OCCUPIED';
----------Updated lowercase 'Available' to 'AVAILABLE'
UPDATE V1_STAGING_AREA
SET BED_STATUS = 'AVAILABLE'
WHERE UPPER(BED_STATUS) = 'AVAILABLE' OR BED_STATUS = 'Available';

UPDATE V1_STAGING_AREA
SET WARD_NAME = 'GENERAL WARD'
WHERE UPPER(WARD_NAME) = 'GENERAL CARE';



---
ADMISSION DATE WAS CLEANED BY SEPERATING DATE AND TIME USING TO_DATE TO CONVERT TO DATE FO
RMATE

---ADDED A NEW COLUMN FOR THE DATE EXCLUDING TIME
ALTER TABLE V1_STAGING_AREA DROP COLUMN NEW_ADMISSION_DATE;

ALTER TABLE V1_STAGING_AREA ADD NEW_ADMISSION_DATE DATE;

----UPDATE MY NEW COLUMN DATE FORMAT
UPDATE V1_STAGING_AREA
SET NEW_ADMISSION_DATE = TO_DATE(TO_CHAR(ADMISSION_DATE, 'DD-MON-YY'), 'DD-MON-YY');

----FOR DROPPING THE REAL ADMISSION DATE COLUMN
-- ALTER TABLE V1_STAGING_AREA DROP COLUMN ADMISSION_DATE;

---DROP V2_STAGING_AREA
DROP TABLE V2_STAGING_AREA;

---CREATE TABLE V2_STAGING AREA;
---DELETE FROM V2_STAGING_AREA;
```

```
-- #DONE
----Date Transformation
CREATE TABLE V2_STAGING_AREA AS
SELECT
    CARE_CENTRE_NAME,
    CARE_CENTRE_ID,
    POSTCODE,
    TOWN,
    EXTRACT(YEAR FROM NEW_ADMISSION_DATE) AS YEAR,
    TO_CHAR(NEW_ADMISSION_DATE, 'MON') AS MONTH,
    BED_STATUS,
    WARD_NAME,
    DATASOURCE
FROM V1_STAGING_AREA;


INSERT INTO V2_STAGING_AREA(SELECT
    CARE_CENTRE_NAME,
    CARE_CENTRE_ID,
    POSTCODE,
    TOWN,
    EXTRACT(YEAR FROM NEW_ADMISSION_DATE) AS YEAR,
    TO_CHAR(NEW_ADMISSION_DATE, 'MON') AS MONTH,
    BED_STATUS,
    WARD_NAME,
    DATASOURCE
FROM V1_STAGING_AREA
WHERE EXTRACT(YEAR FROM NEW_ADMISSION_DATE) = 2023 AND TO_CHAR(NEW_ADMISSION_DATE, 'MON')
= 'NOV');

INSERT INTO V2_STAGING_AREA(SELECT
    CARE_CENTRE_NAME,
```

```
    CARE_CENTRE_ID,
    POSTCODE,
    TOWN,
    EXTRACT(YEAR FROM NEW_ADMISSION_DATE) AS YEAR,
    TO_CHAR(NEW_ADMISSION_DATE, 'MON') AS MONTH,
    BED_STATUS,
    WARD_NAME,
    DATASOURCE
FROM V1_STAGING_AREA
WHERE EXTRACT(YEAR FROM NEW_ADMISSION_DATE) = 2022 AND TO_CHAR(NEW_ADMISSION_DATE, 'MON')
= 'NOV');

INSERT INTO V2_STAGING_AREA(SELECT
    CARE_CENTRE_NAME,
    CARE_CENTRE_ID,
    POSTCODE,
    TOWN,
    EXTRACT(YEAR FROM NEW_ADMISSION_DATE) AS YEAR,
    TO_CHAR(NEW_ADMISSION_DATE, 'MON') AS MONTH,
    BED_STATUS,
    WARD_NAME,
    DATASOURCE
FROM V1_STAGING_AREA
WHERE EXTRACT(YEAR FROM NEW_ADMISSION_DATE) = 2021 AND TO_CHAR(NEW_ADMISSION_DATE, 'MON')
= 'NOV');

INSERT INTO V2_STAGING_AREA(SELECT
    CARE_CENTRE_NAME,
    CARE_CENTRE_ID,
    POSTCODE,
    TOWN,
    EXTRACT(YEAR FROM NEW_ADMISSION_DATE) AS YEAR,
```

```
    TO_CHAR(NEW_ADMISSION_DATE, 'MON') AS MONTH,
    BED_STATUS,
    WARD_NAME,
    DATASOURCE
FROM V1_STAGING_AREA
WHERE EXTRACT(YEAR FROM NEW_ADMISSION_DATE) = 2022 AND TO_CHAR(NEW_ADMISSION_DATE, 'MON')
= 'JAN');


INSERT INTO V2_STAGING_AREA(SELECT
    CARE_CENTRE_NAME,
    CARE_CENTRE_ID,
    POSTCODE,
    TOWN,
    EXTRACT(YEAR FROM NEW_ADMISSION_DATE) AS YEAR,
    TO_CHAR(NEW_ADMISSION_DATE, 'MON') AS MONTH,
    BED_STATUS,
    WARD_NAME,
    DATASOURCE
FROM V1_STAGING_AREA
WHERE EXTRACT(YEAR FROM NEW_ADMISSION_DATE) = 2023 AND TO_CHAR(NEW_ADMISSION_DATE, 'MON')
= 'JAN');

INSERT INTO V2_STAGING_AREA(SELECT
    CARE_CENTRE_NAME,
    CARE_CENTRE_ID,
    POSTCODE,
    TOWN,
    EXTRACT(YEAR FROM NEW_ADMISSION_DATE) AS YEAR,
    TO_CHAR(NEW_ADMISSION_DATE, 'MON') AS MONTH,
    BED_STATUS,
    WARD_NAME,
    DATASOURCE
```

```
FROM V1_STAGING_AREA
WHERE EXTRACT(YEAR FROM NEW_ADMISSION_DATE) = 2023 AND TO_CHAR(NEW_ADMISSION_DATE, 'MON')
= 'FEB');

INSERT INTO V2_STAGING_AREA(SELECT
    CARE_CENTRE_NAME,
    CARE_CENTRE_ID,
    POSTCODE,
    TOWN,
    EXTRACT(YEAR FROM NEW_ADMISSION_DATE) AS YEAR,
    TO_CHAR(NEW_ADMISSION_DATE, 'MON') AS MONTH,
    BED_STATUS,
    WARD_NAME,
    DATASOURCE
FROM V1_STAGING_AREA
WHERE EXTRACT(YEAR FROM NEW_ADMISSION_DATE) = 2022 AND TO_CHAR(NEW_ADMISSION_DATE, 'MON')
= 'FEB');

INSERT INTO V2_STAGING_AREA(SELECT
    CARE_CENTRE_NAME,
    CARE_CENTRE_ID,
    POSTCODE,
    TOWN,
    EXTRACT(YEAR FROM NEW_ADMISSION_DATE) AS YEAR,
    TO_CHAR(NEW_ADMISSION_DATE, 'MON') AS MONTH,
    BED_STATUS,
    WARD_NAME,
    DATASOURCE
FROM V1_STAGING_AREA
WHERE EXTRACT(YEAR FROM NEW_ADMISSION_DATE) = 2023 AND TO_CHAR(NEW_ADMISSION_DATE, 'MON')
= 'APR');
---DONE
```

```
INSERT INTO V2_STAGING_AREA(SELECT
    CARE_CENTRE_NAME,
    CARE_CENTRE_ID,
    POSTCODE,
    TOWN,
    EXTRACT(YEAR FROM NEW_ADMISSION_DATE) AS YEAR,
    TO_CHAR(NEW_ADMISSION_DATE, 'MON') AS MONTH,
    BED_STATUS,
    WARD_NAME,
    DATASOURCE
FROM V1_STAGING_AREA
WHERE EXTRACT(YEAR FROM NEW_ADMISSION_DATE) = 2023 AND TO_CHAR(NEW_ADMISSION_DATE, 'MON')
= 'MAR');



DROP TABLE CARE_CENTRE_DIM CASCADE CONSTRAINTS;

DROP TABLE NEW_TIME_DIM CASCADE CONSTRAINTS;

DROP TABLE BED_OCCUPANCY_FACT CASCADE CONSTRAINTS;


-- Create a Database table to represent the "CARE_CENTRE_DIM" entity.
CREATE TABLE CARE_CENTRE_DIM(
    PK__Care_Centre_Id  INTEGER NOT NULL,
    care_centre_name    VARCHAR(20) NOT NULL,
    Care_Centre_Id INTEGER,
    ward_name VARCHAR(35),
    Town    VARCHAR(35) NOT NULL,
    Post_code   VARCHAR(10) NOT NULL,
```

```
    -- Specify the PRIMARY KEY constraint for table "CARE_CENTRE_DIM".
    -- This indicates which attribute(s) uniquely identify each row of data.
    CONSTRAINT  pk_CARE_CENTRE_DIM PRIMARY KEY (PK__Care_Centre_Id)
);

-- Create a Database table to represent the "TIME_DIM" entity.
CREATE TABLE NEW_TIME_DIM(
    PK__Time_Id INTEGER NOT NULL,
    Month   VARCHAR(25) NOT NULL,
    Year    INTEGER NOT NULL,
    -- Specify the PRIMARY KEY constraint for table "TIME_DIM".
    -- This indicates which attribute(s) uniquely identify each row of data.
    CONSTRAINT  pk_NEW_TIME_DIM PRIMARY KEY (PK__Time_Id)
);
---THE_MONTH
---THE_YEAR

-- Create a Database table to represent the "BED_OCCUPANCY_FACT" entity.
CREATE TABLE BED_OCCUPANCY_FACT(
    PK__Report_Id   INTEGER NOT NULL,
    fk1_PK__Time_Id INTEGER NOT NULL,
    fk2_PK__Care_Centre_Id  INTEGER NOT NULL,
    Total_Bed_Occupancy INTEGER,
    Percentage_Bed_Occupancy    INTEGER,
    -- Average_Bed_Occupancy_Rate   INTEGER,
    -- Specify the PRIMARY KEY constraint for table "BED_OCCUPANCY_FACT".
    -- This indicates which attribute(s) uniquely identify each row of data.
    CONSTRAINT  pk_BED_OCCUPANCY_FACT PRIMARY KEY (PK__Report_Id)
);


-------------------------------------------------------------------------
-- Alter Tables adding fk constraints --
```

```
ALTER TABLE BED_OCCUPANCY_FACT ADD CONSTRAINT fk1_BED_OCCUPANCY_FACT_to_TIME_DIM FOREIGN
 KEY(fk1_PK__Time_Id) REFERENCES NEW_TIME_DIM(PK__Time_Id);

-
- Alter table to add new constraints required to implement the "BED_OCCUPANCY_FACT_CARE_CENTRE_DIM" relation
ship


ALTER TABLE BED_OCCUPANCY_FACT ADD CONSTRAINT fk2_BED_OCCUPANCY_FACT_to_CARE_CENTRE_DIM
 FOREIGN KEY(fk2_PK__Care_Centre_Id) REFERENCES CARE_CENTRE_DIM(PK__Care_Centre_Id);
---LOADING
DROP sequence NEW_TIME_DIM_SEQ;
create sequence NEW_TIME_DIM_SEQ
start with 1
increment by 1
maxvalue 10000
minvalue 1;

DROP SEQUENCE CARE_CENTRE_DIM_SEQ;
Create sequence CARE_CENTRE_DIM_SEQ
start with 1
increment by 1
maxvalue 10000
minvalue 1;

DROP SEQUENCE BED_OCCUPANCY_FACT_SEQ;
Create sequence BED_OCCUPANCY_FACT_SEQ
start with 1
increment by 1
maxvalue 10000
```

```
minvalue 1;


INSERT INTO NEW_TIME_DIM VALUES(NEW_TIME_DIM_SEQ.nextval,'NOV',2021);

INSERT INTO NEW_TIME_DIM VALUES( NEW_TIME_DIM_SEQ.nextval,'JAN',2022);
INSERT INTO NEW_TIME_DIM VALUES(NEW_TIME_DIM_SEQ.nextval,'FEB',2022);
INSERT INTO NEW_TIME_DIM VALUES(NEW_TIME_DIM_SEQ.nextval,'NOV',2022);


INSERT INTO NEW_TIME_DIM VALUES(NEW_TIME_DIM_SEQ.nextval,'JAN',2023);
INSERT INTO NEW_TIME_DIM VALUES(NEW_TIME_DIM_SEQ.nextval,'FEB',2023);
INSERT INTO NEW_TIME_DIM VALUES(NEW_TIME_DIM_SEQ.nextval,'MAR',2023);
INSERT INTO NEW_TIME_DIM VALUES(NEW_TIME_DIM_SEQ.nextval,'APR',2023);

--DATA CLEANING,TMP TABLE,DATA USED
DROP TABLE TMP_CARE_CENTRE;
Create table TMP_CARE_CENTRE as SELECT DISTINCT Care_centre_name,CARE_CENTRE_ID,town,postcode,ward
_name,bed_status from V2_STAGING_AREA;

INSERT INTO CARE_CENTRE_DIM SELECT CARE_CENTRE_DIM_SEQ.NEXTVAL,CARE_CENTRE_NAME,CARE_C
ENTRE_ID,ward_name,TOWN,POSTCODE FROM TMP_CARE_CENTRE;

DROP TABLE TMP_BED_OCCUPANCY1;
CREATE TABLE TMP_BED_OCCUPANCY1 AS SELECT BED_STATUS,WARD_NAME,CARE_CENTRE_NAME,TOWN,
POSTCODE,YEAR,MONTH FROM V2_STAGING_AREA;
----CALCULATING PERECENTAGE OF OCCUPIED BED PER WARD PER CARE CENTRE,TOTAL BED OCCUPIED
DROP TABLE TMP_BED_OCCUPANCY2;
CREATE TABLE TMP_BED_OCCUPANCY2 AS SELECT YEAR,MONTH,CARE_CENTRE_NAME,WARD_NAME,SUM(C
ASE WHEN BED_STATUS='OCCUPIED' THEN 1 ELSE 0 END) AS TOTAL_BED_OCCUPANCY,
(SUM(CASE WHEN BED_STATUS='OCCUPIED' THEN 1 ELSE 0 END)/COUNT(*))*100 AS PERCENTAGE_BED_OCC
UPANCY FROM TMP_BED_OCCUPANCY1 GROUP BY CARE_CENTRE_NAME,WARD_NAME,YEAR,MONTH;
```

```
INSERT INTO BED_OCCUPANCY_FACT(PK__Report_Id,fk1_PK__Time_Id,
   fk2_PK__Care_Centre_Id,
   Total_Bed_Occupancy,
   PERCENTAGE_BED_OCCUPANCY)
SELECT BED_OCCUPANCY_FACT_SEQ.NEXTVAL,NEW_TIME_DIM.PK__Time_Id,CARE_CENTRE_DIM.PK__Care_
Centre_Id,TMP_BED_OCCUPANCY2.TOTAL_BED_OCCUPANCY,TMP_BED_OCCUPANCY2.PERCENTAGE_BED_OC
CUPANCY
FROM TMP_BED_OCCUPANCY2,NEW_TIME_DIM,CARE_CENTRE_DIM
WHERE TMP_BED_OCCUPANCY2.YEAR=NEW_TIME_DIM.YEAR AND
TMP_BED_OCCUPANCY2.MONTH = NEW_TIME_DIM.MONTH AND
TMP_BED_OCCUPANCY2.CARE_CENTRE_NAME = CARE_CENTRE_DIM.CARE_CENTRE_NAME
;
```

SHORT CASE DUTY DESCRIPTION

North-West, Yorkshire Elderly Social & Health Care

North and West Yorkshire CCG (Clinical Commission Group) are requesting a health and social care system that integrates data from care homes and social care. The Leeds City Council (LCC) operates 6 Care Homes (3 NYR and 3 WYR) that provide additional support for elderly patients during their recovery period. The role of care homes is to act as an extended wing of a hospital. While LCC collaborate with over 200 elderly care providers, these six are part of the test case to produce a system to improve elderly patients experience and provide high quality care recovery period support.

There are two objectives with the system. One – to review Care Homes effectiveness, in terms of the recovery period and bed occupancy, second – support for doctors and social care services, where doctors *need to see the social care services provided to a patient, and social workers and service providers need to have access to the health care record.*

Giving the limits of the social care data (in this case study), there is more value for Social Carers to see patient data, then for doctors and nurses to see Social care systems data.

Note that while all elderly is seen as patients of the Care Homes, they may not necessarily be in the Social Care database.

As a BI consultant, you are expected to implement a solution that integrates WYR and NYR health data, as well as social care case data. The structure of the elderly health care system and social care system can be found below.

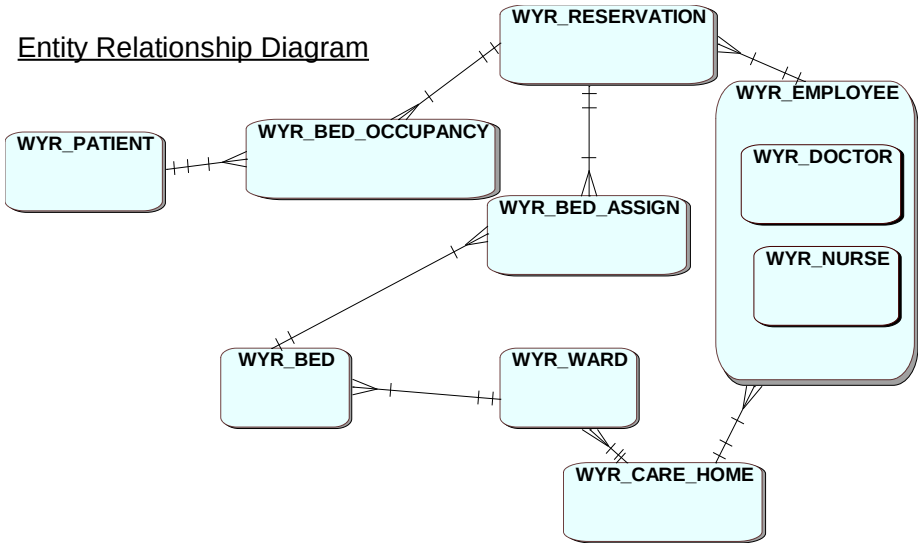
The structure of the health care system for the elderly

Below are details for the care homes that specialise in the care of elderly patients. At present, there are six "care centres" in North and West Yorkshire, that provide care to elderly patients during their recovery period.

Here's a sample of the information that could be stored:

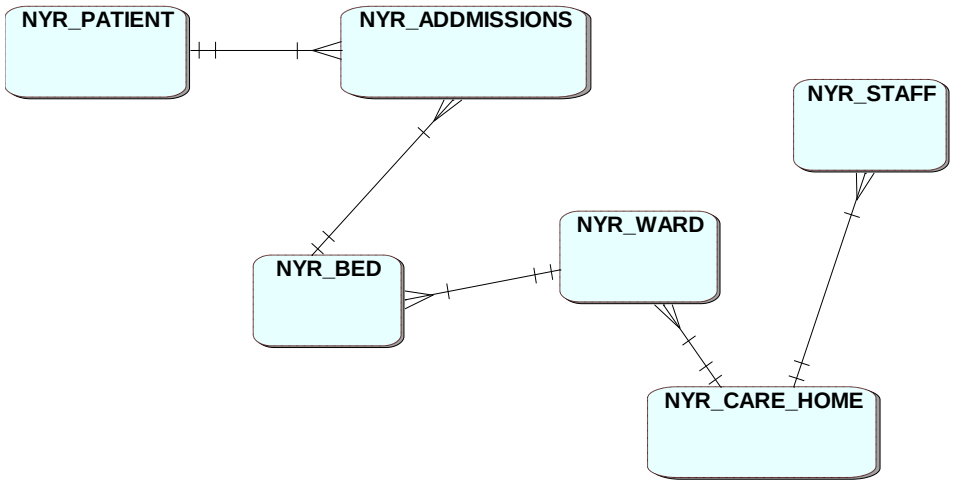
- Patient name, date of birth, address and contact details.
- Patient ID, which makes sure your records are linked to your other records.
- Details of health conditions and illnesses
- Medicines and other treatments
- Stays on the ward/bed, long or short term
- Records of contact you've had with doctors and other health and care workers.

West Yorkshire (WYR) - Relational db design (ERD)



North Yorkshire (NYR) – Relational db design (ERD)

Entity Relationship Diagram



The Social Care System's partial structure can be found below:

Let's assume that the majority of the databases here are relational. Note that not all elderly people are in social care, hence, why there are many fewer records in the SCS system. It is assumed that Leeds City Council Elderly Social Health organizes its data in the following manner.

Client Table:

Name	Address	ClientID
Joy Ray	4 Church road, Bradford	1
Paula M. Condit	15 Quay Street, NAVESTOCK SIDE, CM14 1JQ	1

Referral Table

Date	Reason	Outcome	Source	CaseID
Date when the referral was made	Why need ASC services?	Decision	Who made the referral?	Internal Case ID
10/5/2023	Referral for Health Care	Referral to Elderly Care	Doctor	CID001
16/5/2023	Referral for Health Care	Referral to Elderly Care	Doctor	CID002

1/5/2023	Referral for Health Care	Referral to Elderly Care	Doctor	CID003

Reason examples:
- Need for Assessment - Referral for Mental Health - Referral for Health Care – Signposting.

Outcome examples: Referral to Elderly Care, Referral to Mental Care, Referral to Health Sector.

Source examples: GP, Hospital, Police, Neighbour, Social worker, Family, etc.

Assessment Table

ClientID	CaseID	Date	Outcome	Source
Internal Client ID	Internal Case ID	When Assessment was made	Type of Service	Who made assessment ?
WYR_1	CID001	01/05/2023	Referral to Elderly Care	GP
NYR_1	CID002	01/05/2023	Referral to Elderly Care	GP

REFERENCE/BIBLIOGRAPHY

Turban, E., 2011. From *Decision support and business intelligence systems*. Pearson Education India.

Adams, C.2010: Star schema the complete reference, McGraw Hill Professional (2010) pp.34

Calvanese, D., De Giacomo, G., Lenzerini, M., Nardi, D. and Rosati, R., 2001. Data integration in data warehousing. *International Journal of Cooperative Information Systems*, *10*(03), pp.237-271.

Freeman, R.E., Harrison, J.S. and Zyglidopoulos, S., 2018. *Stakeholder theory: Concepts and strategies*. Cambridge University Press.

Floridi, L. and Taddeo, M., 2016. What is data ethics?. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *374*(2083), p.20160360

Gruschka, N., Mavroeidis, V., Vishi, K. and Jensen, M., 2018, December. Privacy issues and data protection in big data: a case study analysis under GDPR. In *2018 IEEE International Conference on Big Data (Big Data)* (pp. 5027-5033). IEEE.

Kimball, R., Ross, M., Thorthwaite, W., Becker, B. and Mundy, J., 2008. *The data warehouse lifecycle toolkit*. John Wiley & Sons.

Kimball, R., Ross, M., Becker, B. and Mundy, J. (2013). The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling.Wiley

Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A., Brewer, D.,Christakis, N., Contractor,N., Fowler, J.,Gutmann, M., Jebara, T., King, G., Macy, M., Roy, D., & Van Alstyne, M. (2009).'Computational Social Science'. Science vol. 323, pp. 721-3

Lashmar, P. (2017). No More Sources? :The impact of Snowden's revelations on journalists and their confidential sources. Journalism Practice, 11(6), 665–688. https://doi.org/10.1080/17512786.2016.1179587

Mason, R. O. (1986). Four ethical issues of the information age. MIS Quarterly, 10(1), 5–12.

Malinowski, E. and Zimányi, E., 2004. OLAP hierarchies: A conceptual perspective. In *Advanced Information Systems Engineering: 16th International Conference, CAiSE 2004, Riga, Latvia, June 7-11, 2004. Proceedings 16* (pp. 477-491). Springer Berlin Heidelberg.

N.K. Karthikeyan, Saravana Balaji. B and Revathy, S. (2013) From Data Warehouse to Streaming Warehouses: A survey on the Challenges for Real-Time Data Warehousing and Available Solution. International Journal of Computer Applications (0975-8887) Volume 81-No2,November 2013.