



# **DATA WAREHOUSE MODELS AND APPROACHES- 17301**

INDIVIDUAL CASE STUDY ASSESSMENT

**TOPIC:**

**IMPLEMENTATION OF DATAWAREHOUSE/MART SYSTEM**

**CASE STUDY: North-West Yorkshire Elderly Social & HealthCare**

School of Built Environment, Engineering and Computing

JANUARY 2024

Student Author

Victoria Chika Atasie (Data Science, 77332191)

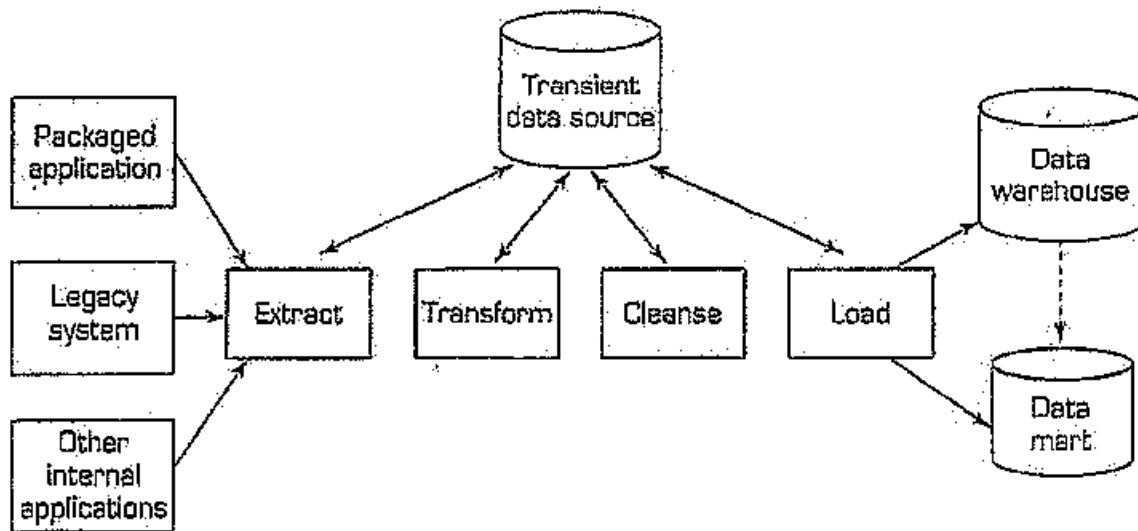
## INTRODUCTION

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.

All evidence codes, **screenshot of codes running successfully would be in the Appendix.**

## TASK 1-DATA INTEGRATION(ETL)AND MAINTENANCE

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).Bill Inmon, the father of Data warehousing, referred to data warehouse as the," A Data Warehouse is a subject oriented, integrated, time variant, non-volatile collection of data"( N.K. Karthikeyan, 2013).



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.

**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**)

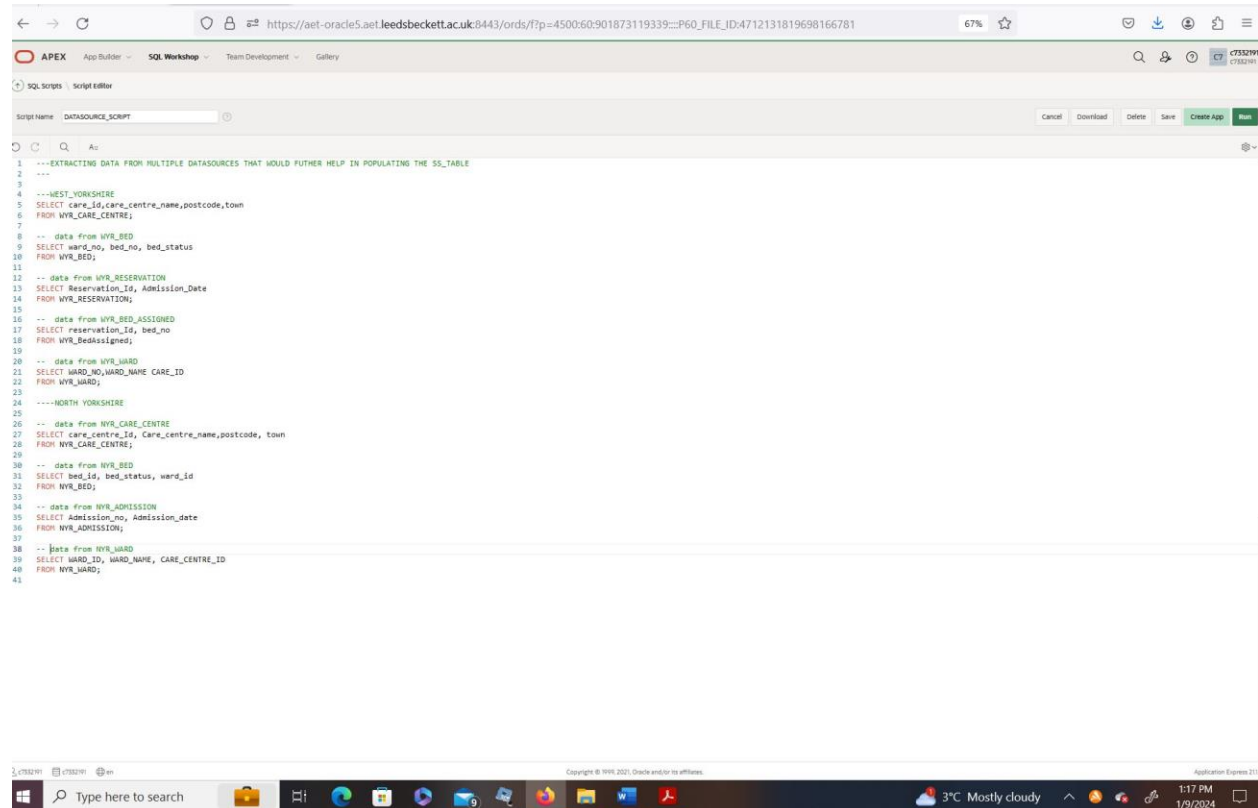
**SOURCE 2:** NYR\_CARE CENTRE (care\_centre\_Id, **Care\_centre\_name**, address, **postcode**, **town**, email)

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**)

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.



The screenshot displays the APEX SQL Workshop interface. The browser address bar shows the URL: [https://aet-oracle5.aet.leedsbeckett.ac.uk:8443/ords/f?p=4500:60:901873119339::P60\\_FILE\\_ID:4712131819698166781](https://aet-oracle5.aet.leedsbeckett.ac.uk:8443/ords/f?p=4500:60:901873119339::P60_FILE_ID:4712131819698166781). The interface includes a top navigation bar with 'APEX', 'App Builder', 'SQL Workshop', 'Team Development', and 'Gallery'. Below this, there's a 'SQL scripts' section with a 'script editor' tab. The main area contains a SQL script with the following content:

```
1  ---EXTRACTING DATA FROM MULTIPLE DATASOURCES THAT WOULD FURTHER HELP IN POPULATING THE SS_TABLE
2  ---
3
4  ---WEST YORKSHIRE
5  SELECT care_id,care_centre_name,postcode,town
6  FROM WVR_CARE_CENTRE;
7
8  -- data from WVR_BED
9  SELECT ward_no, bed_no, bed_status
10 FROM WVR_BED;
11
12 -- data from WVR_RESERVATION
13 SELECT Reservation_Id, Admission_Date
14 FROM WVR_RESERVATION;
15
16 -- data from WVR_BED_ASSIGNED
17 SELECT reservation_id, bed_no
18 FROM WVR_BedAssigned;
19
20 -- data from WVR_WARD
21 SELECT WARD_NO,WARD_NAME CARE_ID
22 FROM WVR_WARD;
23
24 ---NORTH YORKSHIRE
25
26 -- data from WVR_CARE_CENTRE
27 SELECT care_centre_id, Care_centre_name,postcode, town
28 FROM WVR_CARE_CENTRE;
29
30 -- data from WVR_BED
31 SELECT bed_id, bed_status, ward_id
32 FROM WVR_BED;
33
34 -- data from WVR_ADMISSION
35 SELECT Admission_No, Admission_date
36 FROM WVR_ADMISSION;
37
38 -- data from WVR_WARD
39 SELECT WARD_ID, WARD_NAME, CARE_CENTRE_ID
40 FROM WVR_WARD;
41
```

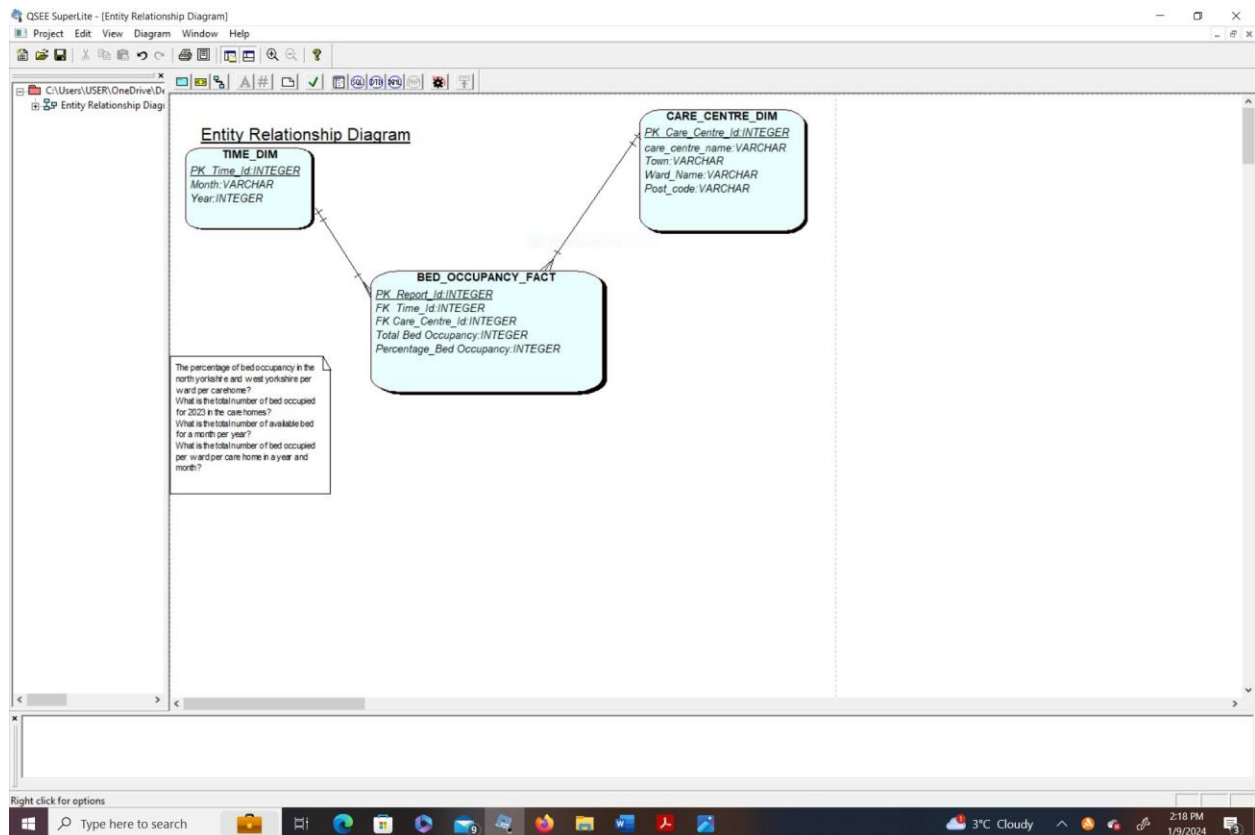
The bottom of the screenshot shows a Windows taskbar with various application icons, a search bar, and system information including the date and time (1:17 PM, 1/9/2024).

## STAR SCHEMA IMPELNTATION

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 below 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 along side 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.

Browser address bar: [https://aet-oracle5.aet.leedsbeckett.ac.uk:8443/ords/f?p=4500:60:901873119339:::P60\\_FILE\\_ID:4720599512719102652](https://aet-oracle5.aet.leedsbeckett.ac.uk:8443/ords/f?p=4500:60:901873119339:::P60_FILE_ID:4720599512719102652)

APEX SQL Workshop - Team Development - Gallery

Script Name: V1\_STAGING\_AREA.DML

```
4 ---Created a V1_STAGING_AREA BY SELECTING THE COLOUMNS THAT I NEED ,SET ALIAS AND USED JOINING
5 ---ALTERED THE V1 STAGING AREA AND IDENTIFIED/ADD THE DATASOURCE ,THEN INSERTED THE WVR COLUMNS TO ENABLE ME MERGE THE DATA
6
7 -----STAGING AREA
8 -----CREATE TABLE merge_data AS
9 DROP TABLE V1_STAGING_AREA CASCADE CONSTRAINTS;
10
11 CREATE TABLE V1_STAGING_AREA AS
12 SELECT
13   wvr_c.care_centre_name AS care_centre_name,
14   wvr_c.care_id AS care_centre_id,
15   wvr_c.postcode AS postcode,
16   wvr_c.town AS town,
17   wvr_bed.bed_status AS bed_status,
18   wvr_ward.ward_name AS ward_name,
19   wvr_res.admission_date AS admission_date
20 FROM
21   WVR_CARE_CENTRE wvr_c
22 JOIN
23   WVR_WARD wvr_ward ON wvr_c.care_id = wvr_ward.care_id
24 JOIN
25   WVR_BED wvr_bed ON wvr_ward.ward_no = wvr_bed.ward_no
26 JOIN
27   WVR_RESERVATION wvr_res ON wvr_res.Reservation_Id IN (SELECT Reservation_Id FROM WVR_Bedassigned WHERE bed_no = wvr_bed.bed_no);
28
29 ALTER TABLE V1_STAGING_AREA
30 ADD DATASOURCE VARCHAR2(25);
31
32 UPDATE V1_STAGING_AREA
33 SET DATASOURCE = 'WVR';
34
35 --- COMBINING THE TABLES.
36
37 INSERT INTO V1_STAGING_AREA
38 SELECT
39   nyr_c.care_centre_name AS care_centre_name,
40   nyr_c.care_centre_id AS care_centre_id,
41   nyr_c.postcode AS postcode,
42   nyr_c.town AS town,
43   nyr_bed.bed_status AS bed_status,
44   nyr_ward.ward_name AS ward_name,
45   nyr_adm.admission_date AS admission_date,
46   'NYR'
47 FROM
48   WVR_CARE_CENTRE nyr_c
49 JOIN
50   WVR_WARD nyr_ward ON nyr_ward.care_centre_id = nyr_c.care_centre_id
51 JOIN
52   WVR_BED nyr_bed ON nyr_bed.ward_id = nyr_ward.ward_id
53 JOIN
54   WVR_ADMISSION nyr_adm ON nyr_adm.bed_id = nyr_bed.bed_id;
55
56
57 -- SELECT * FROM V1_STAGING_AREA
```

Windows taskbar: 2°C Cloudy, 3:21 PM, 1/9/2024

Browser address bar: <https://aet-oracle5.aet.leedsbeckett.ac.uk:8443/ords/f?p=4500:1001:901873119339:::>

APEX SQL Workshop - Team Development - Gallery

Object Browser

Schema: C7352791

Table: V1\_STAGING\_AREA

Table	Count	Rows	Insert Row	Load Data
NYR_ADMISSION				
NYR_BED				
NYR_CARE_CENTRE				
NYR_PATIENT				
NYR_STAFF				
NYR_WARD				
OUT_PATIENT				
S1_STAGINGAREA				
S2_STAGINGAREA				
SALGRADE				
SO_DATASET1				
SO_DATASET2				
STAGE1				
STAGE2				
STAGE3				
STAGE4				
STAGE5C				
TEST_SESSION				
TIME_DIM				
TIMEPMPO				
TIMEPMPOH				
TIMEPMPOCOMMON				
TIMEPMPOA				
TIMEPMPOH				
TIMEPMPOH2				
TIMEPMPOH3				
TIMEPMPOH4				
TIMEPMPOH5				
TIMEPMPOH6				
TIMEPMPOH7				
TIMEPMPOH8				
TIMEPMPOH9				
TIMEPMPOH10				
TIMEPMPOH11				
TIMEPMPOH12				
TIMEPMPOH13				
TIMEPMPOH14				
TIMEPMPOH15				
TIMEPMPOH16				
TIMEPMPOH17				
TIMEPMPOH18				
TIMEPMPOH19				
TIMEPMPOH20				
TIMEPMPOH21				
TIMEPMPOH22				
TIMEPMPOH23				
TIMEPMPOH24				
TIMEPMPOH25				
TIMEPMPOH26				
TIMEPMPOH27				
TIMEPMPOH28				
TIMEPMPOH29				
TIMEPMPOH30				
TIMEPMPOH31				
TIMEPMPOH32				
TIMEPMPOH33				
TIMEPMPOH34				
TIMEPMPOH35				
TIMEPMPOH36				
TIMEPMPOH37				
TIMEPMPOH38				
TIMEPMPOH39				
TIMEPMPOH40				
TIMEPMPOH41				
TIMEPMPOH42				
TIMEPMPOH43				
TIMEPMPOH44				
TIMEPMPOH45				
TIMEPMPOH46				
TIMEPMPOH47				
TIMEPMPOH48				
TIMEPMPOH49				
TIMEPMPOH50				
TIMEPMPOH51				
TIMEPMPOH52				
TIMEPMPOH53				
TIMEPMPOH54				
TIMEPMPOH55				
TIMEPMPOH56				
TIMEPMPOH57				
TIMEPMPOH58				
TIMEPMPOH59				
TIMEPMPOH60				
TIMEPMPOH61				
TIMEPMPOH62				
TIMEPMPOH63				
TIMEPMPOH64				
TIMEPMPOH65				
TIMEPMPOH66				
TIMEPMPOH67				
TIMEPMPOH68				
TIMEPMPOH69				
TIMEPMPOH70				
TIMEPMPOH71				
TIMEPMPOH72				
TIMEPMPOH73				
TIMEPMPOH74				
TIMEPMPOH75				
TIMEPMPOH76				
TIMEPMPOH77				
TIMEPMPOH78				
TIMEPMPOH79				
TIMEPMPOH80				
TIMEPMPOH81				
TIMEPMPOH82				
TIMEPMPOH83				
TIMEPMPOH84				
TIMEPMPOH85				
TIMEPMPOH86				
TIMEPMPOH87				
TIMEPMPOH88				
TIMEPMPOH89				
TIMEPMPOH90				
TIMEPMPOH91				
TIMEPMPOH92				
TIMEPMPOH93				
TIMEPMPOH94				
TIMEPMPOH95				
TIMEPMPOH96				
TIMEPMPOH97				
TIMEPMPOH98				
TIMEPMPOH99				
TIMEPMPOH100				
TIMEPMPOH101				
TIMEPMPOH102				
TIMEPMPOH103				
TIMEPMPOH104				
TIMEPMPOH105				
TIMEPMPOH106				
TIMEPMPOH107				
TIMEPMPOH108				
TIMEPMPOH109				
TIMEPMPOH110				
TIMEPMPOH111				
TIMEPMPOH112				
TIMEPMPOH113				
TIMEPMPOH114				
TIMEPMPOH115				
TIMEPMPOH116				
TIMEPMPOH117				
TIMEPMPOH118				
TIMEPMPOH119				
TIMEPMPOH120				
TIMEPMPOH121				
TIMEPMPOH122				
TIMEPMPOH123				
TIMEPMPOH124				
TIMEPMPOH125				
TIMEPMPOH126				
TIMEPMPOH127				
TIMEPMPOH128				
TIMEPMPOH129				
TIMEPMPOH130				
TIMEPMPOH131				
TIMEPMPOH132				
TIMEPMPOH133				
TIMEPMPOH134				
TIMEPMPOH135				
TIMEPMPOH136				
TIMEPMPOH137				
TIMEPMPOH138				
TIMEPMPOH139				
TIMEPMPOH140				
TIMEPMPOH141				
TIMEPMPOH142				
TIMEPMPOH143				
TIMEPMPOH144				
TIMEPMPOH145				
TIMEPMPOH146				
TIMEPMPOH147				
TIMEPMPOH148				
TIMEPMPOH149				
TIMEPMPOH150				
TIMEPMPOH151				
TIMEPMPOH152				
TIMEPMPOH153				
TIMEPMPOH154				
TIMEPMPOH155				
TIMEPMPOH156				
TIMEPMPOH157				
TIMEPMPOH158				
TIMEPMPOH159				
TIMEPMPOH160				
TIMEPMPOH161				
TIMEPMPOH162				
TIMEPMPOH163				
TIMEPMPOH164				
TIMEPMPOH165				
TIMEPMPOH166				
TIMEPMPOH167				
TIMEPMPOH168				
TIMEPMPOH169				
TIMEPMPOH170				
TIMEPMPOH171				
TIMEPMPOH172				
TIMEPMPOH173				
TIMEPMPOH174				
TIMEPMPOH175				
TIMEPMPOH176				
TIMEPMPOH177				
TIMEPMPOH178				
TIMEPMPOH179				
TIMEPMPOH180				
TIMEPMPOH181				
TIMEPMPOH182				
TIMEPMPOH183				
TIMEPMPOH184				
TIMEPMPOH185				
TIMEPMPOH186				
TIMEPMPOH187				
TIMEPMPOH188				
TIMEPMPOH189				
TIMEPMPOH190				
TIMEPMPOH191				
TIMEPMPOH192				
TIMEPMPOH193				
TIMEPMPOH194				
TIMEPMPOH195				
TIMEPMPOH196				
TIMEPMPOH197				
TIMEPMPOH198				
TIMEPMPOH199				
TIMEPMPOH200				
TIMEPMPOH201				
TIMEPMPOH202				
TIMEPMPOH203				
TIMEPMPOH204				
TIMEPMPOH205				
TIMEPMPOH206				
TIMEPMPOH207				
TIMEPMPOH208				
TIMEPMPOH209				
TIMEPMPOH210				
TIMEPMPOH211				
TIMEPMPOH212				
TIMEPMPOH213				
TIMEPMPOH214				
TIMEPMPOH215				
TIMEPMPOH216				
TIMEPMPOH217				
TIMEPMPOH218				
TIMEPMPOH219				
TIMEPMPOH220				
TIMEPMPOH221				
TIMEPMPOH222				
TIMEPMPOH223				
TIMEPMPOH224				
TIMEPMPOH225				
TIMEPMPOH226				
TIMEPMPOH227				
TIMEPMPOH228				
TIMEPMPOH229				
TIMEPMPOH230				
TIMEPMPOH231				
TIMEPMPOH232				
TIMEPMPOH233				
TIMEPMPOH234				
TIMEPMPOH235				
TIMEPMPOH236				
TIMEPMPOH237				
TIMEPMPOH238				
TIMEPMPOH239				
TIMEPMPOH240				
TIMEPMPOH241				
TIMEPMPOH242				
TIMEPMPOH243				
TIMEPMPOH244				
TIMEPMPOH245				
TIMEPMPOH246				
TIMEPMPOH247				
TIMEPMPOH248				
TIMEPMPOH249				
TIMEPMPOH250				
TIMEPMPOH251				
TIMEPMPOH252				
TIMEPMPOH253				
TIMEPMPOH254				
TIMEPMPOH255				
TIMEPMPOH256				
TIMEPMPOH257				
TIMEPMPOH258				
TIMEPMPOH259				
TIMEPMPOH260				
TIMEPMPOH261				
TIMEPMPOH262				
TIMEPMPOH263				
TIMEPMPOH264				
TIMEPMPOH265				
TIMEPMPOH266				
TIMEPMPOH267				
TIMEPMPOH268				
TIMEPMPOH269				
TIMEPMPOH270				
TIMEPMPOH271				
TIMEPMPOH272				
TIMEPMPOH273				
TIMEPMPOH274				
TIMEPMPOH275				
TIMEPMPOH276				
TIMEPMPOH277				
TIMEPMPOH278				
TIMEPMPOH279				
TIMEPMPOH280				
TIMEPMPOH281				
TIMEPMPOH282				
TIMEPMPOH283				
TIMEPMPOH284				
TIMEPMPOH285				
TIMEPMPOH286				
TIMEPMPOH287				
TIMEPMPOH288				
TIMEPMPOH289				
TIMEPMPOH290				
TIMEPMPOH291				
TIMEPMPOH292				
TIMEPMPOH293				
TIMEPMPOH294				
TIMEPMPOH295				
TIMEPMPOH296				
TIMEPMPOH297				
TIMEPMPOH298				
TIMEPMPOH299				
TIMEPMPOH300				
TIMEPMPOH301				
TIMEPMPOH302				
TIMEPMPOH303				
TIMEPMPOH304				
TIMEPMPOH305				
TIMEPMPOH306				
TIMEPMPOH307				
TIMEPMPOH308				
TIMEPMPOH309				
TIMEPMPOH310				
TIMEPMPOH311				
TIMEPMPOH312				
TIMEPMPOH313				
TIMEPMPOH314				
TIMEPMPOH315				
TIMEPMPOH316				
TIMEPMPOH317				
TIMEPMPOH318				
TIMEPMPOH319				
TIMEPMPOH320				
TIMEPMPOH321				
TIMEPMPOH322				
TIMEPMPOH323				
TIMEPMPOH324				
TIMEPMPOH325				
TIMEPMPOH326				
TIMEPMPOH327				
TIMEPMPOH328				
TIMEPMPOH329				
TIMEPMPOH330				
TIMEPMPOH331				
TIMEPMPOH332				
TIMEPMPOH333				

Further discussion, the V1\_staging\_area 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

--**ELECT** **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

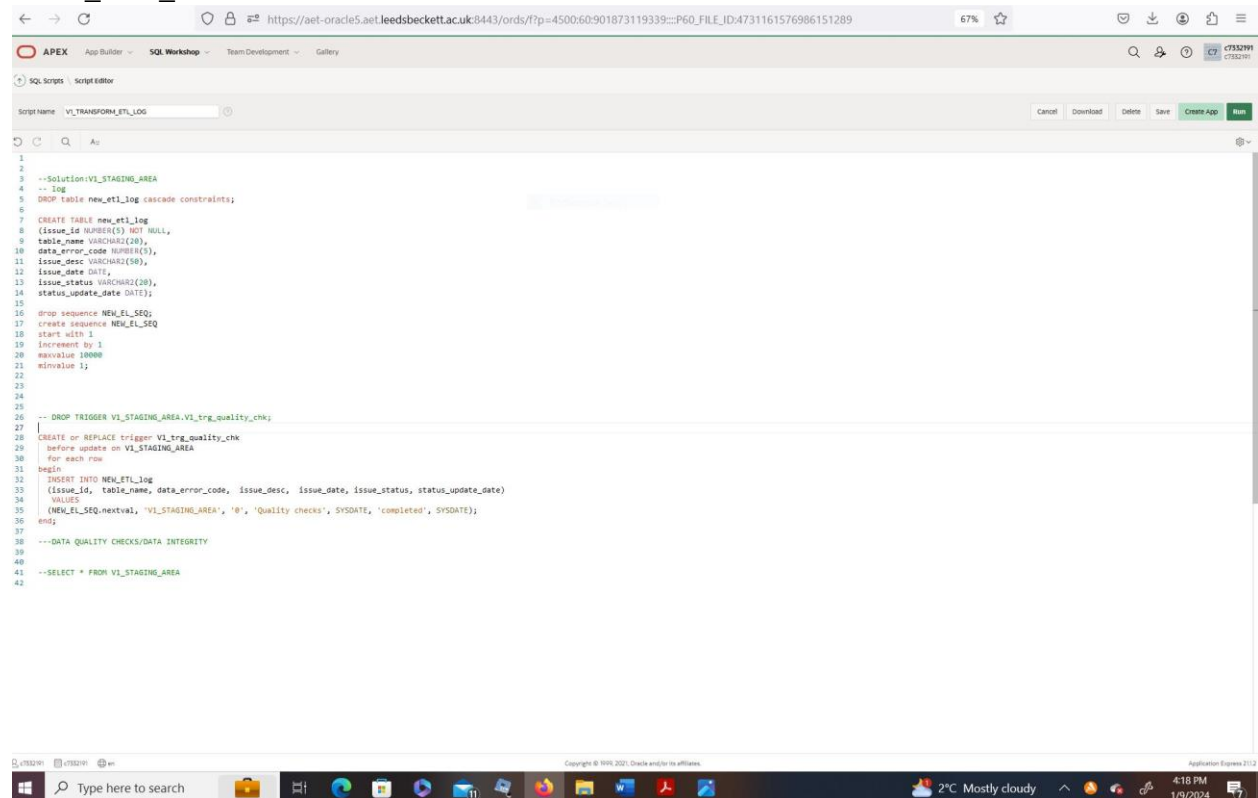
```

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



The screenshot shows the APEX SQL Workshop interface. The browser address bar displays the URL: [https://aet-oracle5.aet.leedsbeckett.ac.uk:8443/ords/?p=4500:60:901873119339:::P60\\_FILE\\_ID:4731161576986151289](https://aet-oracle5.aet.leedsbeckett.ac.uk:8443/ords/?p=4500:60:901873119339:::P60_FILE_ID:4731161576986151289). The page title is "APEX SQL Workshop". The "Script Name" field is set to "V1\_TRANSFORM\_ETL\_LOG". The script content is as follows:

```

1
2
3 --Solution:V1_STAGING_AREA
4 -- log
5 DROP table new_etl_log cascade constraints;
6
7 CREATE TABLE new_etl_log
8 (issue_id NUMBER(5) NOT NULL,
9 table_name VARCHAR2(20),
10 data_error_code NUMBER(3),
11 issue_desc VARCHAR2(50),
12 issue_date DATE,
13 issue_status VARCHAR2(20),
14 status_update_date DATE);
15
16 drop sequence NEW_ETL_SEQ;
17 create sequence NEW_ETL_SEQ
18 start with 1
19 increment by 1
20 maxvalue 10000
21 minvalue 1;
22
23
24
25
26 -- DROP TRIGGER V1_STAGING_AREA.V1_trg_quality_chk;
27 |
28 CREATE or REPLACE trigger V1_trg_quality_chk
29 before update on V1_STAGING_AREA
30 for each row
31 begin
32 INSERT INTO NEW_ETL_log
33 (issue_id, table_name, data_error_code, issue_desc, issue_date, issue_status, status_update_date)
34 VALUES
35 (NEW_ETL_SEQ.nextval, 'V1_STAGING_AREA', '0', 'Quality checks', SYSDATE, 'completed', SYSDATE);
36 end;
37
38 ---DATA QUALITY CHECKS/DATA INTEGRITY
39
40
41 --SELECT * FROM V1_STAGING_AREA
42

```

The bottom of the screenshot shows a Windows taskbar with the date and time as 4:18 PM on 1/9/2024, and a weather widget showing 2°C Mostly cloudy.



← → ↺ https://aet-oracle5.aet.leedsbeckett.ac.uk:8443/ords/?p=4500:1001:901873119339::: 67% ☆

APEX App Builder SQL Workshop Team Development Gallery

Object browser Schema: CT332191

Tables

Table Data Indices Model Constraints Grants Statistics UI Defaults Triggers Dependencies SQL REST Sample Queries

EDT	ISSUE_ID	TABLE_NAME	DATA_ERROR_CODE	ISSUE_DESC	ISSUE_DATE	ISSUE_STATUS	STATUS_UPDATE_DATE
40		V1_STAGING_AREA	0	Quality checks	01/08/2024	completed	01/08/2024
41		V1_STAGING_AREA	0	Quality checks	01/08/2024	completed	01/08/2024
42		V1_STAGING_AREA	0	Quality checks	01/08/2024	completed	01/08/2024
43		V1_STAGING_AREA	0	Quality checks	01/08/2024	completed	01/08/2024
44		V1_STAGING_AREA	0	Quality checks	01/08/2024	completed	01/08/2024
45		V1_STAGING_AREA	0	Quality checks	01/08/2024	completed	01/08/2024
46		V1_STAGING_AREA	0	Quality checks	01/08/2024	completed	01/08/2024
47		V1_STAGING_AREA	0	Quality checks	01/08/2024	completed	01/08/2024
48		V1_STAGING_AREA	0	Quality checks	01/08/2024	completed	01/08/2024
49		V1_STAGING_AREA	0	Quality checks	01/08/2024	completed	01/08/2024
50		V1_STAGING_AREA	0	Quality checks	01/08/2024	completed	01/08/2024
51		V1_STAGING_AREA	0	Quality checks	01/08/2024	completed	01/08/2024
52		V1_STAGING_AREA	0	Quality checks	01/08/2024	completed	01/08/2024
53		V1_STAGING_AREA	0	Quality checks	01/08/2024	completed	01/08/2024
54		V1_STAGING_AREA	0	Quality checks	01/08/2024	completed	01/08/2024

Download

row(s) 1 - 5 of 51 Next

Copyright © 1995-2021, Oracle and/or its affiliates. Application Express 21.2

2°C Mostly cloudy 4:22 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 .

APEX App Builder SQL Workshop Team Development Gallery

SQL Scripts Script editor

Script Name: QUALITY\_CHECK\_V2\_STAGING

Cancel Download Delete Save Create Add Run

```

77 ---Data Transformation
78 CREATE TABLE V2_STAGING_AREA AS
79 SELECT
80     CARE_CENTRE_NAME,
81     CARE_CENTRE_ID,
82     POSTCODE,
83     TOWN,
84     EXTRACT(YEAR FROM NEW_ADMISSION_DATE) AS YEAR,
85     TO_CHAR(NEW_ADMISSION_DATE, 'MON') AS MONTH,
86     BED_STATUS,
87     WARD_NAME,
88     DATASOURCE
89 FROM V1_STAGING_AREA;
90
91
92 INSERT INTO V2_STAGING_AREA(SELECT
93     CARE_CENTRE_NAME,
94     CARE_CENTRE_ID,
95     POSTCODE,
96     TOWN,
97     EXTRACT(YEAR FROM NEW_ADMISSION_DATE) AS YEAR,
98     TO_CHAR(NEW_ADMISSION_DATE, 'MON') AS MONTH,
99     BED_STATUS,
100     WARD_NAME,
101     DATASOURCE
102 FROM V1_STAGING_AREA
103 WHERE EXTRACT(YEAR FROM NEW_ADMISSION_DATE) = 2023 AND TO_CHAR(NEW_ADMISSION_DATE, 'MON') = 'NOV');
104
105 INSERT INTO V2_STAGING_AREA(SELECT
106     CARE_CENTRE_NAME,
107     CARE_CENTRE_ID,
108     POSTCODE,
109     TOWN,
110     EXTRACT(YEAR FROM NEW_ADMISSION_DATE) AS YEAR,
111     TO_CHAR(NEW_ADMISSION_DATE, 'MON') AS MONTH,
112     BED_STATUS,
113     WARD_NAME,
114     DATASOURCE
115 FROM V1_STAGING_AREA
116 WHERE EXTRACT(YEAR FROM NEW_ADMISSION_DATE) = 2022 AND TO_CHAR(NEW_ADMISSION_DATE, 'MON') = 'NOV');
117
118 INSERT INTO V2_STAGING_AREA(SELECT
119     CARE_CENTRE_NAME,
120     CARE_CENTRE_ID,
121     POSTCODE,
122     TOWN,
123     EXTRACT(YEAR FROM NEW_ADMISSION_DATE) AS YEAR,
124     TO_CHAR(NEW_ADMISSION_DATE, 'MON') AS MONTH,
125     BED_STATUS,
126     WARD_NAME,
127     DATASOURCE
128 FROM V1_STAGING_AREA
129 WHERE EXTRACT(YEAR FROM NEW_ADMISSION_DATE) = 2021 AND TO_CHAR(NEW_ADMISSION_DATE, 'MON') = 'NOV');
130
131 INSERT INTO V2_STAGING_AREA(SELECT
132     CARE_CENTRE_NAME,
133     CARE_CENTRE_ID,
134     POSTCODE,

```

Copyright © 1995-2021, Oracle and/or its affiliates. Application Express 21.2

2°C Cloudy 4:30 PM 1/9/2024

<https://aet-oracle5.aet.leedsbeckett.ac.uk/8443/ords/f?p=4500:1001901873119339::>
67%

**APEX** App Builder **SQL Workshop** Team Development Gallery

**Object Browser** Schema C7352191

**V2\_STAGING\_AREA**

**Table** **Data** **Indexes** **Model** **Constraints** **Grants** **Statistics** **UI Defaults** **Triggers** **Dependencies** **SQL** **REST** **Sample Queries**

**Query** **Count Rows** **Insert Row** **Load Data**

EDIT	CARE_CENTRE_NAME	CARE_CENTRE_ID	POSTCODE	TOWN	YEAR	MONTH	BED_STATUS	WARD_NAME	DATASOURCE
	LIBU Carehome	1	L54 0BH	Leeds	2023	APR	AVAILABLE	GENERAL WARD	WYR
	LIBU Carehome	1	L54 0BH	Leeds	2023	APR	OCCUPIED	GENERAL WARD	WYR
	LIBU Carehome	1	L54 0BH	Leeds	2023	APR	OCCUPIED	GENERAL WARD	WYR
	LIBU Carehome	1	L54 0BH	Leeds	2023	APR	OCCUPIED	GENERAL WARD	WYR
	LIBU Carehome	1	L54 0BH	Leeds	2023	APR	AVAILABLE	GENERAL WARD	WYR
	LIBU Carehome	1	L54 0BH	Leeds	2023	APR	OCCUPIED	GENERAL WARD	WYR
	JUNO Carehome	3	HQ1 4DG	Harrogate	2023	APR	OCCUPIED	GENERAL WARD	WYR
	JUNO Carehome	3	HQ1 4DG	Harrogate	2023	FEB	OCCUPIED	GENERAL WARD	WYR
	JUNO Carehome	3	HQ1 4DG	Harrogate	2023	FEB	AVAILABLE	GENERAL WARD	WYR
	OSCAR Carehome	2	L31 1EE	Leeds	2023	JAN	OCCUPIED	GENERAL WARD	WYR
	OSCAR Carehome	2	L31 1EE	Leeds	2023	FEB	OCCUPIED	GENERAL WARD	WYR
	JUNO Carehome	3	HQ1 4DG	Harrogate	2023	NOV	OCCUPIED	GENERAL WARD	WYR
	JUNO Carehome	3	HQ1 4DG	Harrogate	2023	MAR	OCCUPIED	GENERAL WARD	WYR
	OSCAR Carehome	2	L31 1EE	Leeds	2021	NOV	AVAILABLE	GENERAL WARD	WYR

[Download](#)

(row) 1 - 15 of 70

Copyright © 1999, 2021, Oracle and/or its affiliates.

Application Express 2013

```
→ ↻ 🔍 https://aet-oracle5.aet.leedsbeckett.ac.uk:8443/ords/f?p=4500:60:901871319339::P60_FILE_ID:4652496021584759484 67% ☆
APEX App Builder SQL Workshop Team Development Gallery
SQL Scripts Script Editor
Script Name BED_OCCUPANCY_FACT
Cancel Download Delete Save Create App Run
6
7
8 -- Create a Database table to represent the "CARE_CENTRE_DIN" entity.
9 CREATE TABLE CARE_CENTRE_DIN
10 PK_Care_Centre_Id INTEGER NOT NULL,
11 care_centre_name VARCHAR(20) NOT NULL,
12 Care_Centre_Id INTEGER,
13 ward_name VARCHAR(35),
14 Team VARCHAR(25) NOT NULL,
15 Post_code VARCHAR(10) NOT NULL,
16 -- Specify the PRIMARY KEY constraint for table "CARE_CENTRE_DIN".
17 -- This indicates which attribute(s) uniquely identify each row of data.
18 CONSTRAINT pk_CARE_CENTRE_DIN PRIMARY KEY (PK_Care_Centre_Id)
19
20
21 -- Create a Database table to represent the "TIME_DIN" entity.
22 CREATE TABLE NEW_TIME_DIN
23 PK_Time_Id INTEGER NOT NULL,
24 Month VARCHAR(25) NOT NULL,
25 Year INTEGER NOT NULL,
26 -- Specify the PRIMARY KEY constraint for table "TIME_DIN".
27 -- This indicates which attribute(s) uniquely identify each row of data.
28 CONSTRAINT pk_NEW_TIME_DIN PRIMARY KEY (PK_Time_Id)
29
30
31 --THE_MONTH
32 --THE_YEAR
33
34 -- Create a Database table to represent the "BED_OCCUPANCY_FACT" entity.
35 CREATE TABLE BED_OCCUPANCY_FACT
36 PK_Report_Id INTEGER NOT NULL,
37 fk1_PK_Time_Id INTEGER NOT NULL,
38 fk2_PK_Care_Centre_Id INTEGER NOT NULL,
39 Total_Bed_Occupancy INTEGER,
40 Percentage_Bed_Occupancy INTEGER,
41 -- Average_Bed_Occupancy_Rate INTEGER,
42 -- Specify the PRIMARY KEY constraint for table "BED_OCCUPANCY_FACT".
43 -- This indicates which attribute(s) uniquely identify each row of data.
44 CONSTRAINT pk_BED_OCCUPANCY_FACT PRIMARY KEY (PK_Report_Id)
45
46
47 -----
48 -- Alter Tables adding fk constraints --
49
50
51
52 ALTER TABLE BED_OCCUPANCY_FACT ADD CONSTRAINT fk1_BED_OCCUPANCY_FACT_to_TIME_DIN FOREIGN KEY(fk1_PK_Time_Id) REFERENCES NEW_TIME_DIN(PK_Time_Id);
53
54 -- Alter table to add new constraints required to implement the "BED_OCCUPANCY_FACT,CARE_CENTRE_DIN" relationship
55
56
57 ALTER TABLE BED_OCCUPANCY_FACT ADD CONSTRAINT fk2_BED_OCCUPANCY_FACT_to_CARE_CENTRE_DIN FOREIGN KEY(fk2_PK_Care_Centre_Id) REFERENCES CARE_CENTRE_DIN(PK_Care_Centre_Id);
58
59 ---LOADING
60 DROP SEQUENCE NEW_TIME_DIN_SEQ;
61 create sequence NEW_TIME_DIN_SEQ
62 start with 1
63 increment by 1
64 maxvalue 100000
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
88
```

Browser address bar: <https://aet-oracle5.aet.leedsbeckett.ac.uk:8443/ords/f?p=4500:1001:901873119339::>

Application: APEX SQL Workshop

Object Browser: **NEW\_TIME\_DIM**

PK_TIME_ID	MONTH	YEAR
1	NOV	2021
2	JAN	2022
3	FEB	2022
4	NOV	2022
5	JAN	2023
6	FEB	2023
7	MAR	2023
8	APR	2023

Download

row(s) 1 - 8 of 8

Copyright © 1999, 2021, Oracle and/or its affiliates. Application Express 21.2

Browser address bar: <https://aet-oracle5.aet.leedsbeckett.ac.uk:8443/ords/f?p=4500:1001:901873119339::>

Application: APEX SQL Workshop

Object Browser: **CARE\_CENTRE\_DIM**

PK_CARE_CENTRE_ID	CARE_CENTRE_NAME	CARE_CENTRE_ID	WARD_NAME	TOWN	POST_CODE
1	JUNO Carehome	3	GENERAL WARD	Harrigate	HQ1 4DG
2	ALTORN Carehome	2	GENERAL WARD	Sheffield	S1 3EE
3	ALTORN Carehome	2	GENERAL WARD	Sheffield	S1 3EE
4	OSCAR Carehome	2	GENERAL WARD	Leeds	LS1 3EE
5	OSCAR Carehome	2	GENERAL WARD	Leeds	LS1 3EE
6	LIBU Carehome	1	GENERAL WARD	Leeds	LS4 0BN
7	BEHAN Carehome	1	GENERAL WARD	Bradford	BD4 0BN
8	JUNO Carehome	3	GENERAL WARD	Harrigate	HQ1 4DG
9	WELLBEING Carehome	3	ICU	WAKEFIELD	WF1 4DS
10	BEHAN Carehome	1	GENERAL WARD	Bradford	BD4 0BN
11	LIBU Carehome	1	GENERAL WARD	Leeds	LS4 0BN

Download

row(s) 1 - 11 of 11

Copyright © 1999, 2021, Oracle and/or its affiliates. Application Express 21.2

[illegible]

Windows taskbar showing search bar, taskbar icons (File Explorer, Microsoft Edge, etc.), and system tray (clock, weather, network, volume).

## REPORT:

EBIT	PK1_REPORT_ID	PK1_PK_TIME_ID	PK2_PK_CARE_CENTRE_ID	TOTAL_BED_OCCUPANCY	PERCENTAGE_BED_OCCUPANCY
1	1	4	1	2	100
2	2	7	1	2	100
3	3	8	1	2	100
4	4	8	1	2	100
5	5	8	2	4	100
6	6	2	2	0	0
7	7	8	3	4	100
8	8	2	3	0	0
9	9	1	4	0	0
10	10	6	4	2	100
11	11	7	4	2	100
12	12	3	4	2	100
13	13	5	4	2	100
14	14	1	5	0	0
15	15	6	5	2	100

Further discussion: I insert 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.

The Slowing Changing Dimension suitable for maintenance would be the type 2 slowing dimension where both the old and the new record would be present in the Care\_centre\_dimension. The new added column would contain its primary key and effective date column would also be present.

## **TASK 2-OLAP**

Business Intelligence (BI) Systems depends on “integrated, consistent and certified information repository “referred as the Data Warehouse (D W) that is updated regularly with day-to-day operations of an organisation. In decision making process the data that are analyzed are stored in the data warehouse in multidimensional cubes and the cubes are queried by the decision makers which correspond to the online analytical processing paradigms (Djiroun, R. et al.,2019)

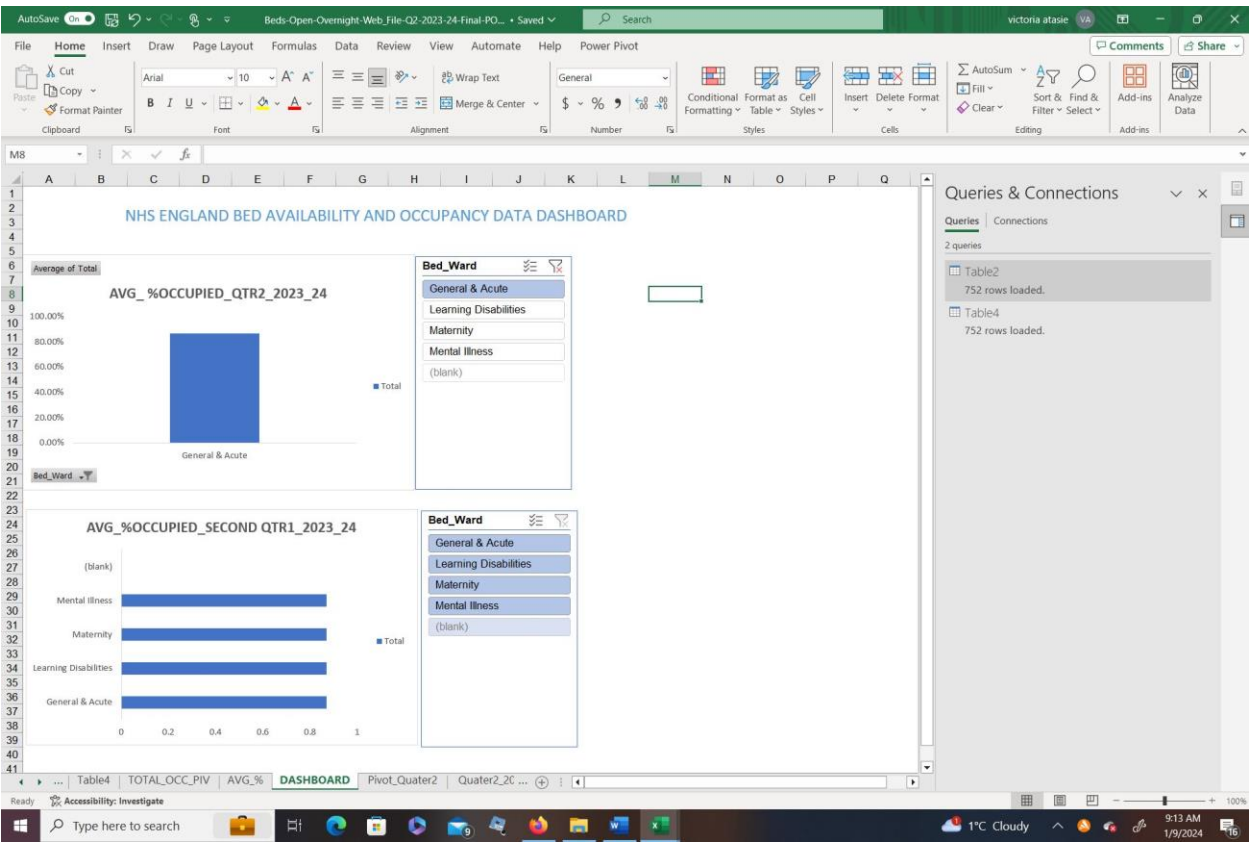
Data warehousing and On-Line analytical processing (OLAP) are considered important elements of decision-making processes and had continuously become a relevant centre of attention in the database industry (Chaudhuri, S. and Dayal, U., 1997). The Online Analytical Processing tool enables decision-making by granting it users the chance to analyse huge amount of historical data using the roll-up or drill down operations. This requires a hierarchical order to make ready automatic calculations although there are different kind of compound hierarchies evolving in the real world are not resolved with online analytical processing implementation (Malinowski, E. and Zimányi, E., 2004).

The additional dataset for NHS Bed Availability and Occupancy Data would be used visualizing the NHS organization in England, Quarter 1 and Quarter 2 for the average percentage of bed occupancy.

Power query, Pivot table and Pivot chart in Excel were used to make visualizations.

The Column and Bar Chart where preferred for visualizing the average percentage bed occupied for each quarter to compare the related categories and a slicer was inserted.





### **TASK 3: FINDINGS AND REFLECTION, SOCIAL AND ETHICAL IMPLICATIONS, DATA GOVERNANCE, AND AS A DATA SCIENTIST**

(Choice of the implementation for the centralised db solution, reconsider other options)

The choice of implementing the centralised database solution as a BI consultant that would resolve the issues associated with integrating data from the West Yorkshire and North Yorkshire health and social care datasets revolves around creating a comprehensive and real time view of the bed occupancy rate, optimizing allocation of bed resources 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 increase demand for care home services.

In considering the findings, some care homes in the North Yorkshire and West Yorkshire care homes may be experiencing a high level of bed occupancy rate and a lack of bed availability can have its consequences in the health and social care sector. The findings involve the business objectives such as the effectiveness of care homes, the optimization and utilization of beds leading to high bed occupancy rates and whilst minimizing the possible bed vacancy, improving quality of care, an effective cost of managing resources allocation with the rate of bed occupancy ensuring a reduction in expenses and enhancing patients' satisfaction whilst creating positive satisfaction and long-term occupancy.

The identification of the business stakeholder, which is the business owner, would also ensure that the implementation of the centralized database solution meets the needs of the business.

The methodology and processes in implementing the centralized database solution would follow the Kimball lifecycle approach in planning which includes setting the framework of managing the project, defining the scope and objectives of the implementation of the database solution. The business requirements would be identified such as the key performance indicators and various factors that affects the business needs, the identification of useful technology to aid the process, designing a dimensional model that relates to the business requirements and measures such as star schemas and based on the dimensional model created , extraction of data from multiple sources, cleaning the data ensuring data integrity and perform transformation and loading (ETL) processes would be considered, business intelligence reports would be implemented, deployment and continuous maintenance would be considered in the implementation of the database solution (Kimball, R et al., 2008)

The architecture considered as the most appropriate to implement for case study organization is the Three-tier Architecture and Dependent DataMart. The advantage of considering the three-tier architecture includes the separating of Datawarehouse, application server and the client layers contributing performance optimization, enabling multi-dimensional database analysis of data and presentation to provide insights for a better decision making (Efraim Turban, 2011). The

dependent DataMart incorporate the use of the single enterprise-wide data model as a concept whilst the end user views the same data which can promote consistency(Efraim Turban,2011).

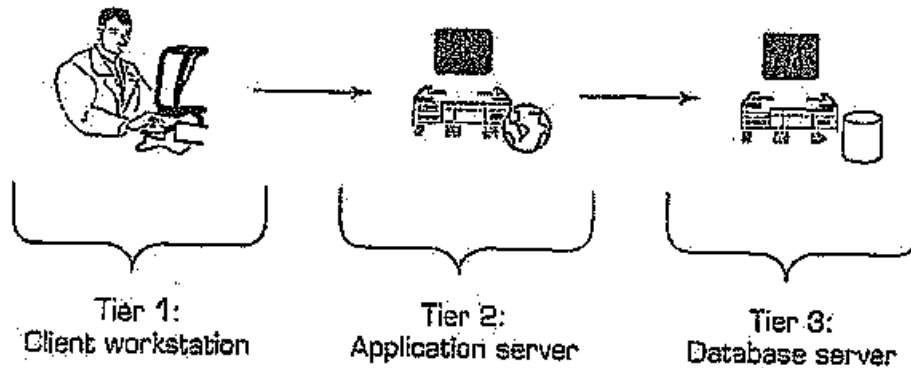


Figure: Three Tier Architecture Data Warehouse (Source: From Decision Support and Business Intelligence System -Efraim Turban)

In conclusion, considering other options would be a centralized data warehouse architecture. The centralized data warehouse architecture creates access to users to every data in the warehouse rather than limiting to DataMart only and can also assist in data management and administrative procedures. (Efraim Turban,2011)

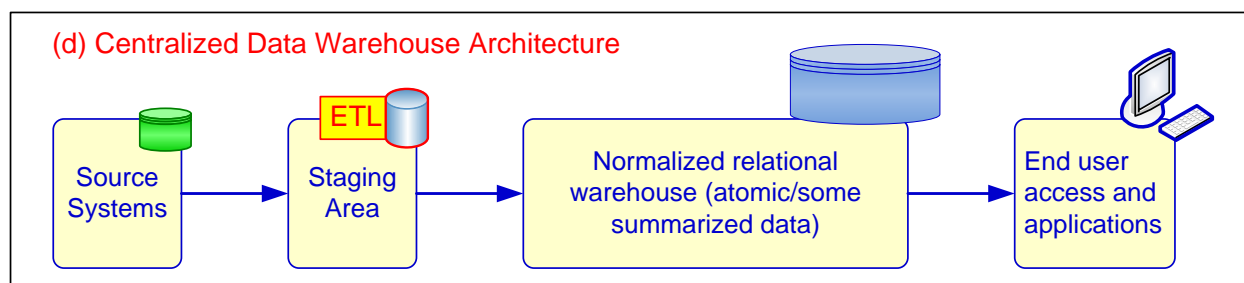


Figure : Centralized Data Warehouse Architecture (Source: The DW Design and Architecture Slides)

## SOCIAL AND ETHICAL IMPLICATIONS AND ROLE OF GOVERNANCE

In a data driven world, developing a database solution is as important as considering social and ethical issues, data governance associated with the implementation and data privacy should be considered in data collection. Floridi, L. and Taddeo, M. (2016) highlights that data ethics involves various ethical issues in the collection or gathering and vast number of datasets for analyses.

Evaluating the current and potential social and ethical implications around the implementation of the centralised system and the role of data governance in the health sector that is associated with care of patients, occupancy of bed can be seen from the aspect of privacy issues with the health and social dataset of the case study organization. The privacy and protection of sensitive data is paramount with patient whilst maintaining trust with their private information such as patient's names, details of health conditions, the particular ward/care home that patients resides and so on. The PAPA framework Mason 1986 emphasizes the four ethical issues in society which are "privacy, accuracy, property and accessibility" (Young, J. et al., 2020). The privacy of the PAPA framework highlights what information of individual can be reviewed to other people, how can individual privacy can be protected, what information can individuals make private to themselves. Accuracy was stressed on ensuring how accurate information are utilized and the error consequences associated with it, Property emphasis on ownership of the information, how appropriate is it for information to be trade off, accessibility highlights on what information is accessible and obtainable (Mason, 1986). Relating PAPA framework in the implementation of the database solution privacy policies, obtaining consents, creating encryption would be considered to protect various identities in the datasets. Data validation to address inaccurate data that might affect better decisions, information ownership of health and social datasets would be adhere to, gaining authorized access to enable protection over information that are sensitive and ensuring that health and social carers can have access to system data following ethical standards.

Big PAPA which stands as an extension of the PAPA framework addressing issues with data in term of "behavioral surveillance, interpretation, and governance" on how individuals can be aware before sharing vital information, how ethical issues associated with data be controlled, how can substandard analysis be identified (Young, J. et al., 2020). The BigPAPA concept relating to the implementation of the centralised database solution which involves data of health and social care, ethical considerations would be put in place to address security issues and data are analysed properly and its advantages are distributed among involved parties.

In conclusion the EU General Data Protection Regulation (GDPR) which took effect in May 2018, is stated to be important to every organisation which regulates how personal data is stored, collected, and processed (Gruschka, N. et al 2018) The centralised data store should comply with the GDPR regulations ensuring individual data are protected. According to Freeman, R. et al., (2018), the theory of stakeholder involves stakeholders as group of individuals who affects or have impacts in the various activities of an organization. The major stakeholder here is the business owner and other stakeholders' interest should be considered in the ethical considerations, constant engagement and addressing issues in the implementation of the database solution.

### **Reflect upon your own understanding of the importance as a data scientist.**

I recognize the importance as a data scientist in contributing to the decision-making processes and considering social and ethical implications and the role of governance. Data Science creates

a vast opportunity in improving public and private lives and social environments as well, there is an increasing utilisation of data of recent which can be associated with its ethical issues. These issues can be resolved properly as data science advances. It should be done to protect human rights and not considering ethical issues may result in negative implications and social exclusion (Floridi, L. and Taddeo, M., 2016). Adhering to ethical standards in considering data privacy, maintaining integrity, fairness and transparency is imperative and it's an ethical obligation as a data scientist to implement security measures to protect sensitive data to avoid the consequences of breaching individual privacy.

A substandard data quality can liable have an essential social and economic effects (Wang, R et al.,1996) The importance of data governance practice can assist in ensuring data is secure, conducting quality checks to maintain data accuracy, consistency, and data validations. This practice can help organisations maintain data integrity, correctness, and accuracy of data in their database, addressing bias which can lead to excellent analytical results and better decision-making. Compliance with the general data protection regulation GDPR that secure and protect personal data is an important practice. Stakeholder engagement and regulatory bodies are important in understanding the use of data which would help ensure database solutions adhere to ethical standards.

In conclusion, the importance of identifying social and ethical issues, considering data governance is of great importance when implementing a centralised database solution for the case study organization as a data scientist or business intelligence consultant by ensuring that organization make careful consideration related to ethical issues and data scientist or IT personnel and everyone in the organization ensure that data is always protected and secured.

TASK 4:

EVIDENCE OF COMPLETED TUTORIAL ON DM DESIGN AND IMPLEMENTATION

Object browser

Tables

APPOINTMENT

BED\_OCCUPANCY\_FACT

CARE\_CENTRE\_DIM

DATA\_ARCHIVE1

DEMO\_CONTRACTANT\_LOOKUP

DEMO\_CUSTOMERS

DEMO\_ORDERS

DEMO\_ORDER\_ITEMS

DEMO\_PRODUCT\_INFO

DEMO\_STATES

DEMO\_TAGS

DEMO\_TAG\_SUM

DEMO\_TAG\_TYPE\_SUM

DEPT

DEPT\_DIM

EMP

EMP\_DIM

EMP\_DIM

EMP\_DIM

ETL\_LOG

**FACT\_CLAIM**

GP

HTMLDB\_PLAN\_TABLE

LOCATION\_DIM

NEW\_ETL\_LOG

NEW\_TIME\_DIM

NIV\_ADMISSION

NIV\_BED

NIV\_CARE\_CENTRE

NIV\_PATIENT

NIV\_STAFF

NIV\_WARD

OUT\_PATIENT

S1\_STUDYAREA

S2\_STUDYAREA

SALGRADE

SO\_DATASET1

SO\_DATASET2

STAGE1

STAGE2

STAGE3

STAGE4

STAGECD

TEST\_SESSION

TIME\_DIM

TIMEEMPED

FACT\_CLAIM

Table

Date

Indexes

Model

Constraints

Grants

Statistics

UI Defaults

Triggers

Dependencies

SQL

REST

Sample Queries

Query

Count Rows

Insert Row

Load Data

EDIT	REPORT_ID	PRI_TIME_ID	PRI_LOCATION_ID	TOTAL_CLAIM_LOCATION
1	2	7	775	
2	3	7	846	
3	1	7	385637	
4	1	8	720852	
5	2	8	990	
6	2	9	792	
7	4	9	809	
8	1	9	102945	
9	2	10	1891	
10	1	10	187652	
11	1	11	31480	
12	1	12	898	

Download

rows 1 - 12 of 12

EVIDENCE OF COMPLETED TUTORIAL SCD SCOTT

← → ↺ https://aet-oracle5.aet.leedsbeckett.ac.uk/ords/f?p=4500:1001:6573464408761::: 67% ☆

APEX App Builder SQL Workshop Team Development Gallery

Object Browser Schema: CT352101

Tables

Table Data Indexes Model Constraints Grants Statistics UI Defaults Triggers Dependencies SQL REST Sample Queries

Query Count Rows Insert Row Load Data

EDIT	EMPNO	ENAME	DEPTNO	DNAME	DATE_ST
	7994	MILLER	10	ACCOUNTING	01/09/2024
	7839	KING	10	ACCOUNTING	01/09/2024
	7782	CLARK	10	ACCOUNTING	01/09/2024
	7369	SMITH	20	RESEARCH	01/09/2024
	7902	FORD	20	RESEARCH	01/09/2024
	7876	ADAMS	20	RESEARCH	01/09/2024
	7788	SCOTT	20	RESEARCH	01/09/2024
	7566	JONES	20	RESEARCH	01/09/2024
	8899	JOHN	30	SALES	01/09/2024
	1000	VISKY	30	SALES	01/09/2024
	7844	TURNER	30	SALES	01/09/2024
	1008	VISKY	30	SALES	01/09/2024
	7521	WARD	30	SALES	01/09/2024
	7698	BLAKE	30	SALES	01/09/2024
	7654	MARTIN	30	SALES	01/09/2024

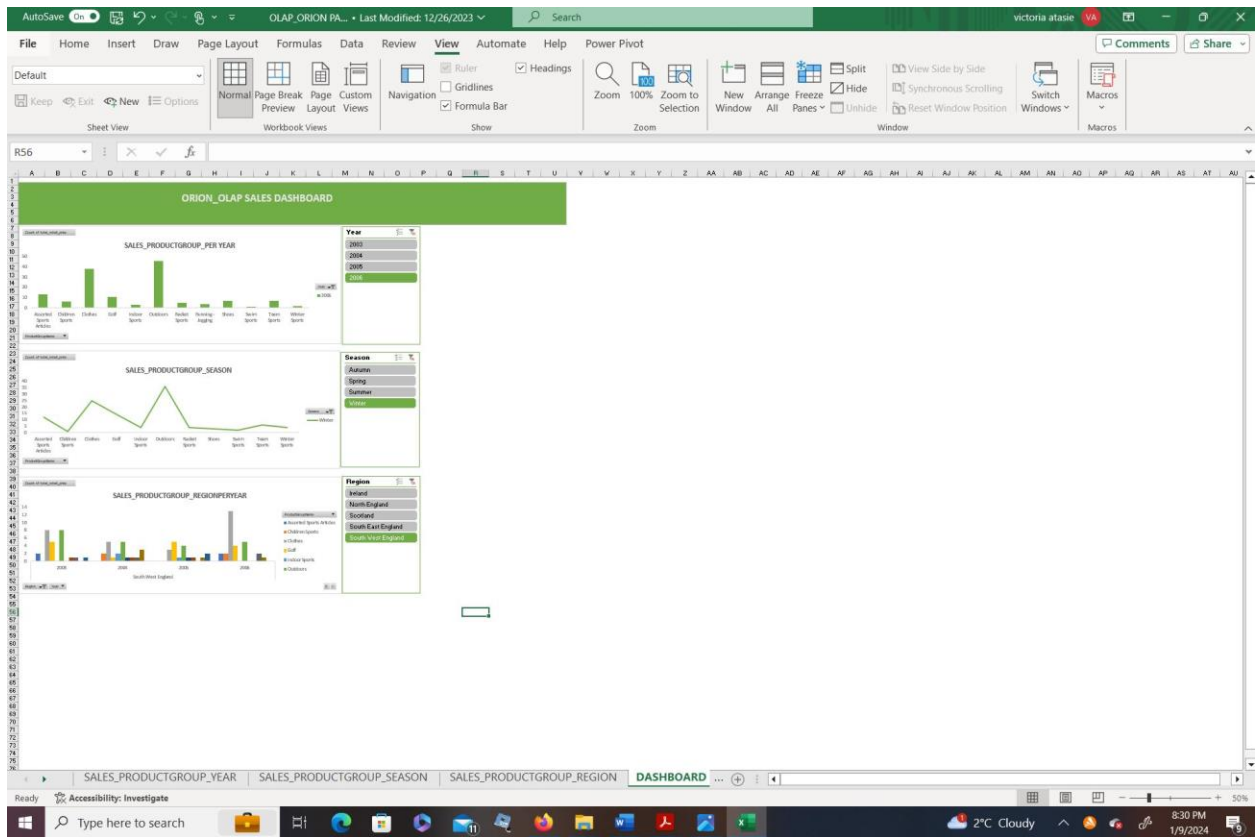
Download

ROW(S) 1 - 15 of 17

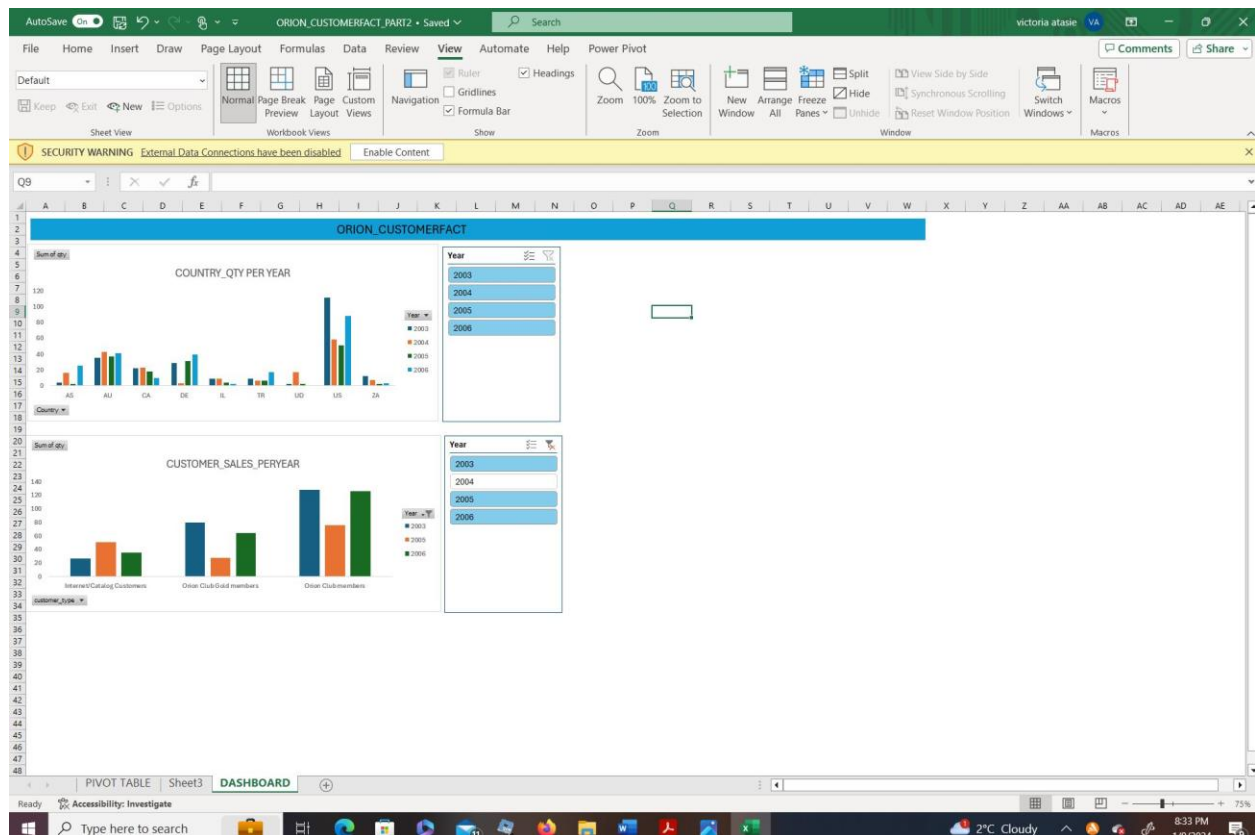
Copyright © 1996-2021 Oracle and/or its affiliates. | Notification Expires 21/2

Windows Taskbar: Type here to search, 2°C Cloudy, 8:24 PM 1/9/2024

**EVIDENCE OF COMPLETED TUTORIAL ON OLAP**







## BIG PAPA FRAMEWORK PAPER

The Big PAPA paper emphasizes on the Richard Mason's in the 1986 as regards the Papa schema that was provided for the four possible ethical concerns that included "privacy, accuracy, property and accessibility" shorten to PAPA (Young, J. et al., 2020). In addition to the previously stated four possible ethical concerns, three more ethical concerns were included as an expansion to the PAPA schema for big data which are viably associated to the initial Mason's PAPA schema that includes governing, behavioural monitoring and the interpretation which lead to the founding of BIG PAPA.

Young, J. et al. (2020) stated that the importance of considering ethical issues as related to big data and not disregarding the Mason initial framework as they are useful for research and practice.

Mason's PAPA Framework discussed the major four aspects of concerns which are privacy, accuracy, property, and accessibility.

Mason's discussion on privacy focuses on the idea that every individual should be free from disclosing their private information to others. Florida legislature's effort was cited to investigate vacant properties and for the Chicago computer centres to investigate employee debts as two

instances of issues with privacy. Mason proceeded to discuss the risk of divulging information without authorization by users and stated that information can be abused which can be used for blackmailing and encouraged the value of privacy and security of personal data (Young, J.et al.,2020)

Mason emphasized how imperative it is for information to be accurate involving large datasets in processes of transaction and forecasting weather. The consequence of the error was highlighted in a situation of bank refuse to accept mortgage payment because of computer system not displaying it being paid for. Mason stated that information systems require proper development and testing to avoid mistakes reoccurring (Young, J.et al.,2020)

Property was stated focuses on artificial intelligence and communication bandwidth that revolves about extracting human thought and creativity whilst incorporating it into machines. Accessibility was stated to be essential for organization and individual, literacy, Intellectual skills, technology, and information. Mason talks about the necessity of accessing stored information in modern databases (Young, J.et al.,2020)

The BIG aspect of the PAPA on the seven(7) issues such as “Behavioural surveillance, Interpretation, Governance, Privacy, Accessibility, Accuracy, Property “that contributed to the founding of Big PAPA ,the societal impact of Big Data with relative questions such how can a person’s freedom be protected when there is a monitoring of behaviour, how can people be aware of analysis that is substandard ,what restriction are available for issues associated with big data, how can information be detected as being accurate, who take ownership of transmission of information ,how accessible is information allowed for companies or individuals that requires them and many others questions(Young, J.et al.,2020 )

“Behavioural surveillance” can have effects on behaviour which can expose the activities such location, interaction, and communication of individuals (Lashmar, 2017; Richards, 2012; Stoycheff,2016)

“Interpretation” was based on the importance of avoiding models that are defective because of inaccurate or insufficient data, reducing the possibility of incorrect conclusions and educating others to identify analyses that are substandard. (Young, J.et al.,2020)

According to Smith, Milberg, & Burke (1996), “Governance” involves making sure procedures are available to deal with ethical issues. People should also be shielded from unethical behaviour (Mason,1986)

Conclusively, the Paper emphasized the consideration of the advancement of modern technology such as social media platforms, artificial intelligence and prediction, health electronic records, personal computer, smartphones, search engines with respect to the Mason ‘Papa published in 1986 and its impact on data in deriving information (Young, J.et al.,2020)

## **SIX PROVOCATIONS FOR BIG DATA**

The six provocations can be related to the potential issues as regards the current development of Artificial Intelligence impact on the society is stated below:

### **1. Automating Research Changes the Definition of Knowledge:**

Henry Ford created a mass production of systems for manufacturing with expert machines and goods of standards in the 20<sup>th</sup> century which led to the dominance of assembly lines and automation accepted in manufacturing industries (Baca 2004). Big Data was described as a huge amount of data and analysing it with the required tools and in computerized ways in research and thoughts (Burkholder 1992). Big Data created a thorough way of how we envision, offering an unheard-of breadth, depth, scale in research (Lazer et al, 2009). “Petabyte Age” was praised by Chris Anderson, who stated that the enormous amount of data and mathematics have replaced the possibility of other tools (Anderson, C. 2008)

It is imperative to take note of the inherent shortcomings of machines and how they can possibly alter the definition of learning, new opportunities, and constraints of the systems (Boyd, D. and Crawford, K., 2011). Related issues in the development of Artificial Intelligence can be envisaged in the aspect of the algorithms of machine learning aiding analysis of data and decisions.

### **2. Claims to Objectivity and Accuracy are Misleading and Bigger Data are Not Always Better (3)**

The problem with the assumption that utilizing Big Data will always produce accurate results. It draws attention to how the opinion of working with Big Data can be when handling data from social media (Boyd, D. and Crawford, K., 2011). Bigger data are not always better emphasizing that researcher utilize quantitative techniques for statistical significance whilst some researchers assumes that individual who embraces data that the methods in social sciences has been resolved and that bigger data equates to higher quality and usefulness (Boyd, D. and Crawford, K., 2011). Related issues in the development of Artificial Intelligence can be envisaged in the interpretation of data ,accuracy of data and predictions.

### **4. Not All Data Are Equivalent**

“Not all data are equivalent” addresses the perception that a small amount of data can be improved better with the use of large data during analysis with the assumption that data are exchangeable (Boyd, D. and Crawford, K., 2011). Envisaging the issues in the development of Artificial Intelligence can be attributed to machine learning models training on the assumption of huge datasets compared to smaller datasets.

### **5. Just Because it is Accessible Doesn't Make It Ethical.**

Accessible data does not mean it is ethical for use by every individual (boyd & Marwick, 2011). This provocation addresses the various challenges associated with ethical approval of data as Big Data is becoming an important area and how important accountability is to be considered (Boyd, D. and Crawford, K., 2011). In relating this similar provocation in envisaging the issues in development of Artificial Intelligence requires data which may have issues with ethical approval.

#### 6. Limited Access to Big Data Creates New Digital Divides

A growing number of studies utilize data from social media which concerns the context, its limitations, access, and purpose (Boyd, D. and Crawford, K., 2011). The emergence of big data is seen as an innovative kind of “digital divide” initiated by the existence of Big Data “ecosystems” (Boyd, D. and Crawford, K., 2011). In this context, the development of Artificial Intelligence models may have issues with constraint access to data.

#### REFERENCES:

Anderson, C. (2008) ‘The End of Theory, Will the Data Deluge Makes the Scientific Method Obsolete?’, Edge, <[http://www.edge.org/3rd\\_culture/anderson08/anderson08\\_index.html](http://www.edge.org/3rd_culture/anderson08/anderson08_index.html)>. [25 July 2011]

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

Boyd, d. and Marwick, A. (2011) ‘Social Privacy in Networked Publics: Teens’ Attitudes, Practices, and Strategies: Paper given at Oxford Internet Institute Decade in Time Conference. Oxford, England.

Boyd, D. and Crawford, K., 2011, September. Six provocations for big data. In *A decade in internet time: Symposium on the dynamics of the internet and society*. pp. 1-17

Baca, G. (2004) ‘Legends of Fordism: Between Myth, History, and Foregone Conclusions’, *Social Analysis*, vol. 48, no.3, pp. 169-178.

Burkholder, L, ed. (1992) *Philosophy and the Computer*, Boulder, San Francisco, and Oxford: Westview Press

Chaudhuri, S. and Dayal, U., 1997. An overview of data warehousing and OLAP technology. *ACM Sigmod record*, 26(1), pp.65-74.

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.

Djiroun, R., Boukhalfa, K. and Alimazighi, Z., 2019. Designing data cubes in OLAP systems: a decision makers' requirements-based approach. *Cluster Computing*, 22, pp.783-803.

Design-Data Warehouse Architecture Slide. The Star Schema. Datawarehouse Models and Approaches 17301. Leeds Beckett University

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.

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.

Richards, N. M. (2012). The Dangers of Surveillance. Harvard Law Review, 126(7), 1934–1965.

Stoycheff, E. (2016). Under surveillance: Examining Facebook’s spiral of silence effects in the wake of NSA internet monitoring. Journalism & Mass Communication Quarterly, 93(2), 296–311.

Smith, H. J., Milberg, S. J., & Burke, S. J. (1996). Information Privacy: Measuring Individuals’ Concerns about Organizational Practices. MIS Quarterly, 20(2), 167–196.  
<https://doi.org/10.2307/249477>

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

Wang, R.Y. and Strong, D.M., 1996. Beyond accuracy: What data quality means to data consumers. *Journal of management information systems*, 12(4), pp.5-33.

Young, J., Smith, T.J. and Zheng, S.H., 2020. Call me BIG PAPA: An extension of Mason’s information ethics framework to big data. *Journal of the Midwest Association for Information Systems (JMWAIIS)*, 2020(2), p.3.

<https://doi.org/10.1177/1077699016630255>

## 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 F
ROM WYR_BedAssigned WHERE 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, 'comple  
ted', 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 T
O DATE FORMATE

---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" relationship

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_CENTRE_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 PERCENTAGE 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(CASE 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_OCCUPANCY 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_OCCUPANCY
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
;
```

**APEX** App Builder - SQL Workshop - Team Development - Gallery

SQL Scripts Results

Script: **V1\_STAGING\_AREA.sql** Status: **Complete**

View: ☐ Detail ☒ Summary Rows: 15

Number	%	Elapsed	Statement	Feedback	Rows
1		0:05	DROP TABLE V1_STAGING_AREA CASCADE CONSTRAINTS	Table dropped.	0
2		0:05	CREATE TABLE V1_STAGING_AREA AS SELECT my_c_cust_id FROM	Table created.	0
3		0:02	ALTER TABLE V1_STAGING_AREA ADD DATA_SOURCE VARCHAR(255)	Table altered.	0
4		0:00	UPDATE V1_STAGING_AREA SET DATA_SOURCE = 'WY'	10 row(s) updated.	10
5		0:02	INSERT INTO V1_STAGING_AREA SELECT my_c_cust_id, my_c	10 row(s) inserted.	10

Download

5 Statements Processed Successful With Errors 0 With Errors

row(s) 1 - 5 of 5



← → ↺ <https://aet-oracle5.aet.leedsbeckett.ac.uk:8443/ords/f?p=4500:1225:901873119339::> 67% ☆

APEX App Builder SQL Workshop Team Development Gallery

SQL Scripts Results

Script: QUALITY\_CHECK\_V2\_STAGING Status: Complete

View: Detail Summary Rows 15 Go

Number	Elapsed	Statement	Feedback	Rows
1	0.01	DELETE FROM V1_STAGING_AREA WHERE ADMISSION_DATE IS NULL	0 row(s) deleted.	0
2	0.02	UPDATE V1_STAGING_AREA SET BED_STATUS = 'OCCUPIED' WHERE LPP	27 row(s) updated.	27
3	0.01	UPDATE V1_STAGING_AREA SET BED_STATUS = 'AVAILABLE' WHERE UP	0 row(s) updated.	0
4	0.01	UPDATE V1_STAGING_AREA SET BED_STATUS = 'AVAILABLE' WHERE UP	8 row(s) updated.	8
5	0.00	UPDATE V1_STAGING_AREA SET WARD_NAME = 'GENERAL WARD' WHERE	0 row(s) updated.	0
6	0.04	ALTER TABLE V1_STAGING_AREA DROP COLUMN NEW_ADMISSION_DATE	Table altered.	0
7	0.03	ALTER TABLE V1_STAGING_AREA ADD NEW_ADMISSION_DATE DATE	Table altered.	0
8	0.04	UPDATE V1_STAGING_AREA SET NEW_ADMISSION_DATE = TO_DATE(TO_C	35 row(s) updated.	35
9	0.04	DROP TABLE V2_STAGING_AREA	Table dropped.	0
10	0.37	CREATE TABLE V2_STAGING_AREA AS SELECT CARE_CENTRE_NAME,	Table created.	0
11	0.00	INSERT INTO V2_STAGING_AREA SELECT CARE_CENTRE_NAME,	4 row(s) inserted.	4
12	0.00	INSERT INTO V2_STAGING_AREA SELECT CARE_CENTRE_NAME,	5 row(s) inserted.	5
13	0.00	INSERT INTO V2_STAGING_AREA SELECT CARE_CENTRE_NAME,	1 row(s) inserted.	1
14	0.01	INSERT INTO V2_STAGING_AREA SELECT CARE_CENTRE_NAME,	2 row(s) inserted.	2
15	0.00	INSERT INTO V2_STAGING_AREA SELECT CARE_CENTRE_NAME,	3 row(s) inserted.	3

Download

19 Statements Processed 19 Successful 0 With Errors

Copyright © 1999-2021 Oracle and/or its affiliates. Application Express 21.2

← → ↺ <https://aet-oracle5.aet.leedsbeckett.ac.uk:8443/ords/f?p=4500:1225:901873119339::> 67% ☆

APEX App Builder SQL Workshop Team Development Gallery

SQL Scripts Results

Script: V1\_TRANSFORM\_ETL\_LOG Status: Complete

View: Detail Summary Rows 15 Go

Number	Elapsed	Statement	Feedback	Rows
1	0.01	DROP table new_etl_log cascade constraints	Table dropped.	
2	0.00	CREATE TABLE new_etl_log (source_id NUMBER(5) NOT NULL, id)	Table created.	
3	0.01	drop sequence NEW_ETL_SEQ	Sequence dropped.	
4	0.00	create sequence NEW_ETL_SEQ start with 1 increment by 1 minval	Sequence created.	
5	0.02	CREATE or REPLACE TRIGGER V1_ETL_qualitychk before update	Trigger created.	

Download

5 Statements Processed 5 Successful 0 With Errors

Copyright © 1999-2021 Oracle and/or its affiliates. Application Express 21.2



