**7BUIS010W Data Warehousing and OLAP – (2021/22)**

**University of Westminster**

Department of Computer Science

|  |
| --- |
| **7BUIS010W Data Warehousing and OLAP – Coursework (2021/22)** |

**0W 7BUIS0** Module leader- Dr Panagiotis Chountas

Malaya Kumar Mishra

Student ID-w1852625  **Data Warehousing and OLAP – Coursework (2021/22) Data Warehousing and OLAP – Coursework (2021/22)Data Warehousing OLAP – Coursework (2021/22)**

Contents

[Data Model 3](#_Toc99012458)

[Attribute Tree 4](#_Toc99012459)

[Dimensional Fact Model 4](#_Toc99012460)

[Mapping DFM Model to Logical Model 4](#_Toc99012461)

[Schema 4](#_Toc99012462)

[Staging 4](#_Toc99012463)

[DWH 4](#_Toc99012464)

[Design Considerations 5](#_Toc99012465)

[DFM Matrix 5](#_Toc99012466)

[Materialised Views 5](#_Toc99012467)

[Justification 5](#_Toc99012468)

[Extract, Load & Transform 6](#_Toc99012469)

[Reports 8](#_Toc99012470)

[Reference List 11](#_Toc99012471)

# Data Model

Diagram

Description automatically generated

# Attribute Tree

* 1. Free hand diagram

****

# Dimensional Fact Model

* 1. Free hand diagram

****

# Mapping DFM Model to Logical Model

* 1. Free hand diagram



# Schema

# Staging



# DWH



# Design Considerations

# DFM Matrix

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **virus** | **location** | **time** | **vaccine** | **target\_group** |  |
| **pandemic** | ***1*** | ***1*** | ***1*** |  |  | 3 |
| **Vaccination** |  | ***1*** | ***1*** |  | ***1*** | 3 |
| **ICU** |  | ***1*** | ***1*** |  |  | 2 |
|  | 1 | 3 | 3 | 0 | 1 |  |

# Materialised Views

The current version handles views composed of selections, joins and a final group-by..As a result of which ,we are able to get a summarized view of the problem statement so easily and quickly.

# Justification

One of the most important decisions in designing a DW is selecting views to materialize for the purpose of efficiently supporting decision making. Queries represent the most  
important input of the problem of selecting materialized views. In this context of materialized view queries are expressed using algorithm Vaccination, target\_ group, cases and ICU the appropriate combination of queries to get convenient materialized views with more efficient way and quick response time for aggregation queries over large tables ,used by European Center for Dieses Prevention and Control . Because 5-6 combination of important parameters memory storage space, cost, and query processing time makes the query processing faster and their occurrence, play an role to get the required accuracy in results with faster response time.

# Extract, Load & Transform

#truncate table dwh.virus;

INSERT INTO dwh.virus

SELECT

@a:=@a + 1 v\_id,

variant AS name,

NULL AS family,

NULL AS indicator\_rate,

percent\_variant AS type

FROM

staging.source\_variant,

(SELECT @a:=0) AS a;

truncate table dwh.location;

INSERT INTO dwh.location

SELECT

@a:=@a + 1 l\_id,

countriesAndTerritories AS country,

countryterritoryCode AS region,

popData2020 AS population

FROM

staging.source\_pandemic,

(SELECT @a:=0) AS a;

truncate table dwh.time;

INSERT INTO dwh.time

SELECT

@a:=@a + 1 t\_id,

year AS year,

month AS month,

day AS date,

dateRep AS start\_date

FROM

staging.source\_pandemic,

(SELECT @a:=0) AS a;

#truncate table dwh.vaccination;

INSERT into dwh.vaccination

SELECT

NumberDosesReceived AS doses\_received,

FirstDose AS first\_dose,

SecondDose AS second\_dose,

DoseAdditional1 AS third\_dose,

vac\_id,

t\_id,

l\_id,

trg\_id

FROM

staging.source\_vaccination AS sv,

dwh.vaccine AS dv,

staging.source\_location AS l,

dwh.time t,

dwh.location dl,

staging.source\_pandemic AS p,

dwh.target\_group as g

WHERE

sv.Vaccine = dv.name

AND sv.YearWeekISO = l.year\_week

AND p.dateRep = t.start\_date

AND CONCAT(p.year, '-', p.month, '-', p.day) = l.date

AND dl.country = p.countriesAndTerritories

AND dl.region = p.countryterritoryCode

AND g.age\_range = sv.TargetGroup

;

#truncate table dwh.pandemic;

INSERT into dwh.pandemic

SELECT distinct

cases AS current\_cases,

cases AS new\_cases,

cases-deaths AS recovered,

deaths AS deaths,

l\_id,

t\_id,

v\_id

FROM

staging.source\_pandemic AS sp,

dwh.virus AS v ,

dwh.time AS t,

dwh.location AS dl,

staging.source\_variant as sv

WHERE

sp.dateRep = t.start\_date

AND CONCAT(sp.year, '-', sp.month, '-', sp.day) = CONCAT(t.year, '-', t.month, '-', t.date)

AND dl.country = sp.countriesAndTerritories

AND dl.region = sp.countryterritoryCode

AND sv.variant = v.name

AND sp.geoId = sv.country\_code

;

#truncate table dwh.icu;

INSERT into dwh.icu

SELECT

indicator AS indicator,

source,

value as icu\_cases,

DoseAdditional1 AS third\_dose,

t\_id,

l\_id

FROM

staging.source\_vaccination AS sv,

dwh.vaccine AS dv,

staging.source\_location AS l,

dwh.time t,

dwh.location dl,

staging.source\_pandemic AS p,

dwh.target\_group as g

WHERE

sv.Vaccine = dv.name

AND sv.YearWeekISO = l.year\_week

AND p.dateRep = t.start\_date

AND CONCAT(p.year, '-', p.month, '-', p.day) = l.date

AND dl.country = p.countriesAndTerritories

AND dl.region = p.countryterritoryCode

AND g.age\_range = sv.TargetGroup

;

CREATE INDEX idx\_pandemic\_l\_t\_id

ON dwh.pandemic (l\_id,t\_id);

CREATE INDEX idx\_vaccination\_l\_t\_id

ON dwh.vaccination (l\_id,t\_id);

CREATE INDEX idx\_vaccination\_l\_t\_id

ON dwh.vaccination (l\_id,t\_id);

# Reports

CREATE TABLE Cases\_view AS SELECT current\_cases,

deaths,

icu\_cases,

region,

month,

name,

age\_range FROM

dwh.pandemic AS p,

dwh.vaccination AS v,

dwh.time AS t,

dwh.location AS l,

dwh.target\_group g,

dwh.vaccine va,

dwh.icu AS i

WHERE

p.t\_id = v.t\_id AND p.l\_id = v.l\_id

AND p.t\_id = i.t\_id

AND p.l\_id = i.l\_id

AND p.t\_id = t.t\_id

AND p.l\_id = l.l\_id

AND v.trg\_id = g.trg\_id

AND v.vac\_id = va.vac\_id;

Table

Description automatically generated

CREATE TABLE vaccination\_view

AS

SELECT

l.country,

year\_week,

SUM(current\_cases) AS total\_cases,

SUM(deaths) AS total\_deaths,

SUM(icu\_cases) AS total\_icu\_cases,

SUM(name) AS total\_vaccinated,

age\_range

FROM

dwh.pandemic AS p,

dwh.vaccination AS v,

dwh.time AS t,

dwh.location AS l,

staging.source\_location AS sl,

dwh.target\_group g,

dwh.vaccine va,

dwh.icu as i

WHERE

p.t\_id = v.t\_id AND p.l\_id = v.l\_id

AND p.t\_id = t.t\_id

AND p.l\_id = l.l\_id

AND p.t\_id = i.t\_id

AND p.l\_id = i.l\_id

AND v.trg\_id = g.trg\_id

AND v.vac\_id = va.vac\_id

AND sl.date = CONCAT(t.year, '-', t.month, '-', t.date)

and p.l\_id = 1

GROUP BY l.country , year\_week , age\_range;

Graphical user interface, table

Description automatically generated

CREATE TABLE Target\_Group\_view AS SELECT DISTINCT l.country,

t.year,

SUM(current\_cases) AS total\_cases,

SUM(deaths) AS total\_deaths,

SUM(icu\_cases) AS total\_icu\_cases,

SUM(name) AS total\_vaccinated,

age\_range FROM

dwh.pandemic AS p,

dwh.vaccination AS v,

dwh.time AS t,

dwh.location AS l,

staging.source\_location AS sl,

dwh.target\_group g,

dwh.vaccine va,

dwh.icu AS i

WHERE

p.t\_id = v.t\_id AND p.l\_id = v.l\_id

AND p.t\_id = t.t\_id

AND p.l\_id = l.l\_id

AND p.t\_id = i.t\_id

AND p.l\_id = i.l\_id

AND v.trg\_id = g.trg\_id

AND v.vac\_id = va.vac\_id

AND sl.date = CONCAT(t.year, '-', t.month, '-', t.date)

AND (age\_range = 'Age60\_69'

OR age\_range = 'Age70\_79'

OR age\_range = 'Age80+')

GROUP BY l.country , t.year

;Graphical user interface, application

Description automatically generated

# Reference List

Cuzzocrea, A., Song, I.-Y., and Davis, K. C. Analytics over Large-Scale Multidimensional Data: The Big Data Revolution! Proc. of ACM DOLAP, 2011..

Nalini, T., Kumaravel, A., Rangarajan, K.,: An Efficient I-Mine Algorithm For Materialized Views In A Data Warehouse Environment, IJCSI, International Journal Of Computer Science Issues, Vol. 8, Issue 5, No.1 (2011). Damiani, M.L., and Spaccapietra S. Spatial Data Warehouse Modeling. Processing and Managing Complex Data for Decision Support, Darmont, J. & Boussaid, O. (Eds), IDEA Group Publishing, 2006, 1--27.

Bimonte, S., Tchounikine, A., and Miquel, M. Towards a Spatial Multidimensional Model. In Proceedings of the Data Warehousing and OLAP Conference (DOLAP'05), 2005, 39--46.

Kimball, R. (1996) The Data Warehouse Toolkit: Practical Techniques for Building Dimensional Data Warehouses, New York: John Wiley & Sons..

Chaudhuri, S & Dayal, U 1997, An Overview of Data Warehousing and OLAP Technology, SIGMOD Record 26:1, Mar, pp 65-74.