

BỘ MÔN HỆ THỐNG THÔNG TIN – KHOA CÔNG NGHỆ THÔNG TIN
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN - ĐẠI HỌC QUỐC GIA TP.HCM

XÂY DỰNG VÀ KHAI THÁC KHO DỮ LIỆU

GV phụ trách: Hồ Thị Hoàng Vy
Tiết Gia Hồng

ĐỒ ÁN MÔN HỌC - HỆ THỐNG THÔNG TIN PHỤC VỤ TRÍ
TUỆ KINH DOANH
HỌC KỲ I – NĂM HỌC 2022-2023

MỤC LỤC

A.	THÔNG TIN NHÓM	1
I.	Thành viên	1
II.	Phân công	1
B.	NỘI DUNG	3
I.	Mục tiêu	3
II.	Data warehouse	3
1.	Phân tích dữ liệu nguồn	3
2.	Thiết kế data warehouse	3
a)	NDS	4
b)	DDS	5
III.	Khai thác dữ liệu	8
1.	Yêu cầu 1	8
2.	Yêu cầu 2	9
3.	Yêu cầu 3	11
4.	Yêu cầu 4	12
5.	Yêu cầu 5	14
6.	Yêu cầu 6	17
7.	Yêu cầu 7	17
8.	Yêu cầu 9	18
IV.	Tạo Job tự động định kỳ thực hiện ETL	19
1.	Các bước thực hiện	19
2.	Kết quả	23
V.	Data Mining	23
1.	Mục tiêu	23
2.	Các bước thực hiện	23
3.	Kết quả	24
C.	THAM KHẢO	24

A. THÔNG TIN NHÓM

I. Thành viên

Mã nhóm	MSSV	Họ tên	Ghi chú
CQ2019-BI#17	19120148	Lê Huỳnh Minh Tuấn	Nhóm trưởng
	19120626	Hồ Minh Quân	
	19120687	Ngô Quốc Toại	
	19120692	Phạm Minh Trí	

II. Phân công

Họ tên	Công việc	Hoàn thành	Đánh giá
Lê Huỳnh Minh Tuấn	<ul style="list-style-type: none">- Phân tích thiết kế Data warehouse, ETL- Triển khai Data warehouse, ETL- Thiết kế và xây dựng Cube- Viết truy vấn MDX để khai thác dữ liệu- Viết truy vấn SQL để làm report	100%	10/10
Hồ Minh Quân	<ul style="list-style-type: none">- Phân tích dữ liệu nguồn (EDA)- Trực quan hoá dữ liệu bằng graph và chart- Data Mining- Viết báo cáo	100%	10/10
Ngô Quốc Toại	<ul style="list-style-type: none">- Phân tích thiết kế Data warehouse, ETL- Tìm hiểu OLAP, SSAS	100%	10/10

	<ul style="list-style-type: none"> - Trực quan hoá dữ liệu bằng regional map - Viết báo cáo 		
Phạm Minh Trí	<ul style="list-style-type: none"> - Phân tích dữ liệu nguồn (EDA) - Phân tích thiết kế Data warehouse, ETL - Triển khai Data warehouse, ETL - Viết truy vấn MDX để khai thác dữ liệu - Tạo Job định kỳ tự động thực hiện ETL 	100%	10/10

B. NỘI DUNG

I. Mục tiêu

- Thiết kế data warehouse để hỗ trợ thống kê phân tích tình hình dịch covid ở Canada.
- Xây dựng và triển khai OLAP dựa trên data warehouse đã thiết kế để khai phá các “insight” trong dữ liệu.

II. Data warehouse

1. Phân tích dữ liệu nguồn

Ta có bộ dữ liệu gồm 6 file. Dựa vào đó ta thấy có 6 đối tượng chính sau đây:

- Public health unit (PHU)
- PHU city
- PHU group
- Thông tin các ca nhiễm (case) ở các PHU
- Thông tin tiêm chủng ở các PHU (vaccine by age PHU)
- Thông các đợt bùng phát dịch (ongoing outbreak PHU)

Trong đó Một PHU Group chứa nhiều PHU city, Mỗi PHU city có một public health unit.

Các vấn đề về dữ liệu

Trong 6 file đã cho, 5 file thể hiện các thông tin ở Ontario ngoại trừ file *Compiled_COVID-19_Case_Deatails* thể hiện thông tin các ca nhiễm vừa ở Ontario vừa ở các tỉnh bang khác. Trong đây có chứa tên các PHU ngoài Ontario, nhưng các thông tin chi tiết khác như PHU city và PHU group của chúng đều không có.

Một số cách đặt tên không được thống nhất (tên các PHU, age group, outbreak group,...).

Thiếu thông tin natural key và nguồn cụ thể.

2. Thiết kế data warehouse

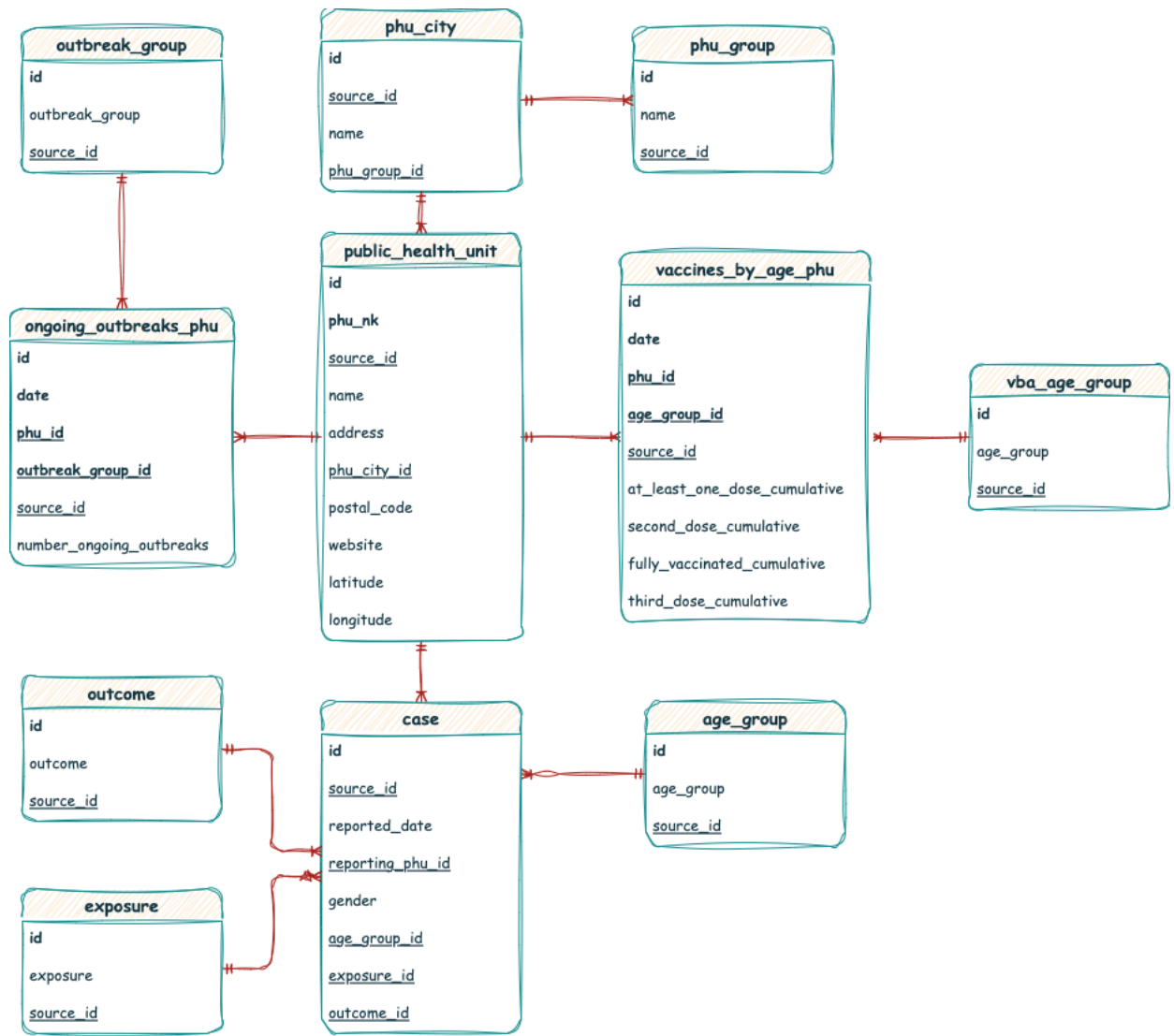
Data warehouse được thiết kế bao gồm NDS và DDS.

a) NDS

Dựa vào 6 đối tượng chính, ta có 6 bảng chính bao gồm: public_health_unit, phu_city, phu_group, vaccines_by_age_phu, ongoing_outbreaks_phu và case.

Các bảng outbreak_group, outcome, exposure, age_group và vba_age_group được tách ra từ những cột có kiểu dữ liệu *categorical* trong các bảng chính.

Có một PHU group đại diện cho các PHU group ngoài Ontario, có id là 0 và name là “PHU groups outside Ontario”. Một PHU city đại diện cho các PHU city ngoài Ontario, có id là 0 và name là “PHU cities outside Ontario” và phu_group_id là 0 (dummy value).



Bảng	Mô tả
phu_group	Thông tin các PHU group. phu_group có id là 0 sẽ đại diện các các PHU group ngoài Ontario
phu_city	Thông tin các PHU city. phu_city có id là 0 sẽ đại diện cho các city ngoài Ontario, và sẽ có phu_group_id tương ứng là 0
public_health_unit	Thông tin chi tiết một PHU. Các PHU ngoài Ontario sẽ có phu_city_id là 0 và các thông tin còn lại null.
case	Thông tin các ca nhiễm ở từng PHU
out_come	Các loại kết quả điều trị
exposure	Các dạng phơi nhiễm
age_group	Các nhóm tuổi
ongoing_outbreaks_phu	Thông tin các đợt bùng phát ở các PHU
outbreak_group	Các loại bùng phát
vaccines_by_age_phu	Thông tin tiêm chủng của từng PHU theo ngày
vba_age_group	Nhóm tuổi của vaccines_by_age_phu

b) DDS

Dựa vào các yêu cầu 1, 2, 3, 4 ta có:

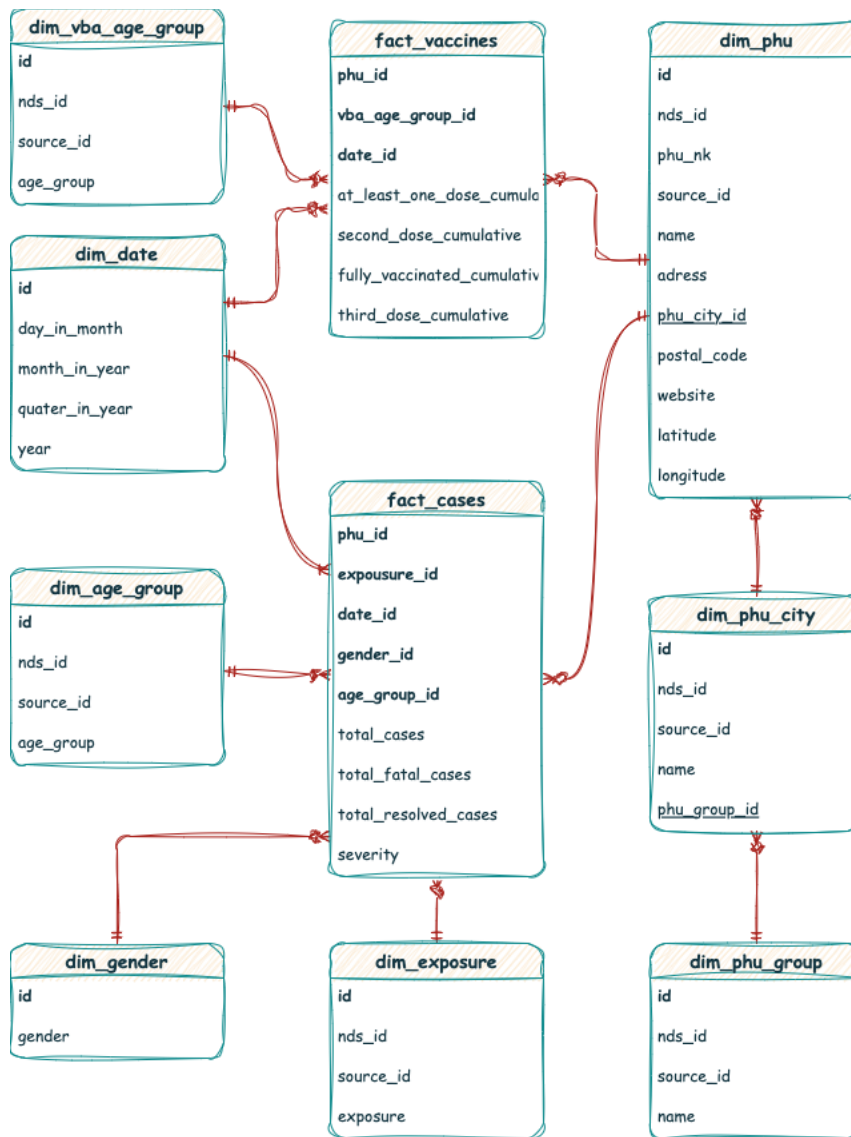
- Dimension:
 - Ở đâu? (PHU, PHU city, PHU group)
 - Khi nào? (Ngày, tháng, quý, năm)
 - Nhóm tuổi nào?
 - Giới tính gì?
 - Kiểu phơi nhiễm là gì?
- Measure:
 - Số người nhiễm (quantity – additive measure)

- Số ca tử vong (quantity – additive measure)
- Số ca hồi phục (quantity – additive measure)
- Mức độ nghiêm trọng (calculated measure – non-additive measure)

Dựa vào yêu cầu 5, ta có:

- Dimension:
 - Ở đâu? (PHU, PHU city, PHU group)
 - Khi nào? (ngày, tháng, quý, năm)
- Measure:
 - Số người đã tiêm vaccine (quantity – additive measure)

Yêu cầu 6 không thể thực hiện được vì trong bảng case (cho biết số ca nhiễm) không cho biết thông tin nhóm bùng phát. Ngược lại bảng ongoing_outbreaks_phu (cho biết thông tin nhóm bùng phát) không cho biết thông tin số ca nhiễm.



Bảng	Mô tả
dim_date	Dimension ngày, tháng, quý, năm
dim_gender	Dimension giới tính
dim_age_group	Dimension nhóm tuổi
dim_phu_group	Dimension geography (PHU group)
dim_phu_city	Dimension geography (PHU city)
dim_phu	Dimension geography (PHU)
dim_vba_age_group	Dimension nhóm tuổi, theo format của vaccine by age PHU

fact_vaccines	Fact table, thống kê số lượng vaccine theo PHU, vba age group và theo thời gian
fact_cases	Fact table, thống kê số ca nhiễm, tử vong, hồi phục.

III. Khai thác dữ liệu

Tuỳ theo yêu cầu thống kê nhóm sẽ phân tích và triển khai OLAP bằng việc xây dựng Cube và viết truy vấn MDX trên Cube đó, với dữ liệu khai thác được nhóm trực quan hoá bằng Excel thông qua Pivot Table và Pivot Chart. Đối với yêu cầu phân tích không thực hiện được bằng OLAP, nhóm sẽ dùng truy vấn SQL.

1. Yêu cầu 1

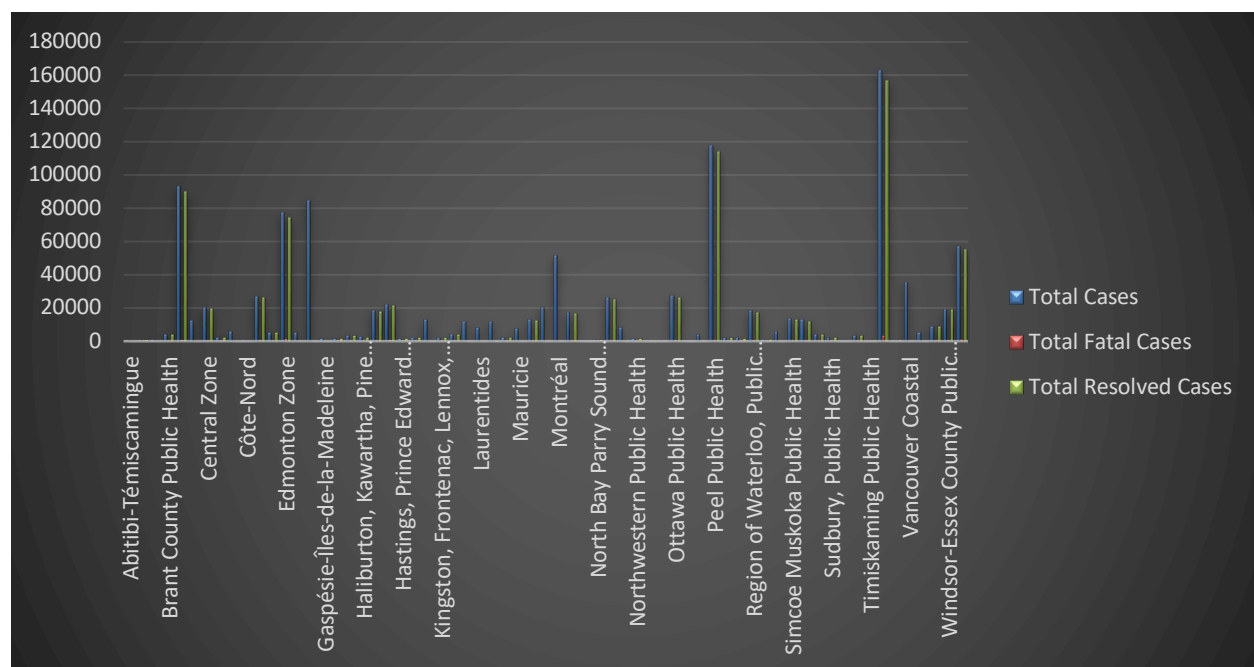
*Thống kê **Số ca nhiễm, số ca tử vong, số ca phục hồi** của dịch Covid-19 theo từng **PHU** trong từng năm.*

*MDX

		Total Cases	Total Fatal Cases	Total Resolved Cases
York Region Public Health Services	2020	23577	723	22849
York Region Public Health Services	2021	32767	268	31935
Toronto Public Health	2020	64514	2254	62259
Toronto Public Health	2021	98157	1113	94136
Peterborough Public Health	2020	516	7	509
Peterborough Public Health	2021	1108	16	1046
Haliburton, Kawartha, Pine Ridge Public Health	2020	829	48	781
Haliburton, Kawartha, Pine Ridge Public Health	2021	1462	35	1344
Durham Region Health Department	2020	10393	423	9970
Durham Region Health Department	2021	16421	109	15926
Wellington-Dufferin-Guelph Public Health	2020	3558	92	3466
Wellington-Dufferin-Guelph Public Health	2021	5073	59	4834
Leeds, Grenville, Lanark Public Health	2020	1082	103	979
Leeds, Grenville, Lanark Public Health	2021	1025	6	1012
Eastern Ontario Public Health	2020	1955	64	1891
Eastern Ontario Public Health	2021	2950	64	2841
Ottawa Public Health	2020	10642	401	10241
Ottawa Public Health	2021	16538	173	15749
Renfrew County, Public Health	2020	320	3	317
Renfrew County, Public Health	2021	452	7	424
Chatham-Kent Public Health	2020	1175	6	1169

*Report

Year	All		
Phu Name	Total Cases	Total Fatal Cases	Total Resolved Cases
Abitibi-Témiscamingue	268	0	0
Algoma Public Public Health	429	6	417
Bas-Saint-Laurent	880	0	0
Brant County Public Health	3941	27	3803
Calgary Zone	92712	678	89451
Capitale-Nationale	11989	0	0
Central Zone	20297	157	19352
Chatham-Kent Public Health	2210	23	2171
Chaudière-Appalaches	5340	0	0
Côte-Nord	209	0	0
Durham Region Health Department	26814	532	25896
Eastern Ontario Public Health	4905	128	4732
Edmonton Zone	76924	1085	74192
Estrie	4693	0	0
Fraser	84315	0	0
Gaspésie-Îles-de-la-Madeleine	1372	0	0
Grey Bruce Public Health	1433	7	1415
Haldimand-Norfolk Public Health	3078	87	2945
Haliburton, Kawartha, Pine Ridge Public Health	2291	83	2125



2. Yêu cầu 2

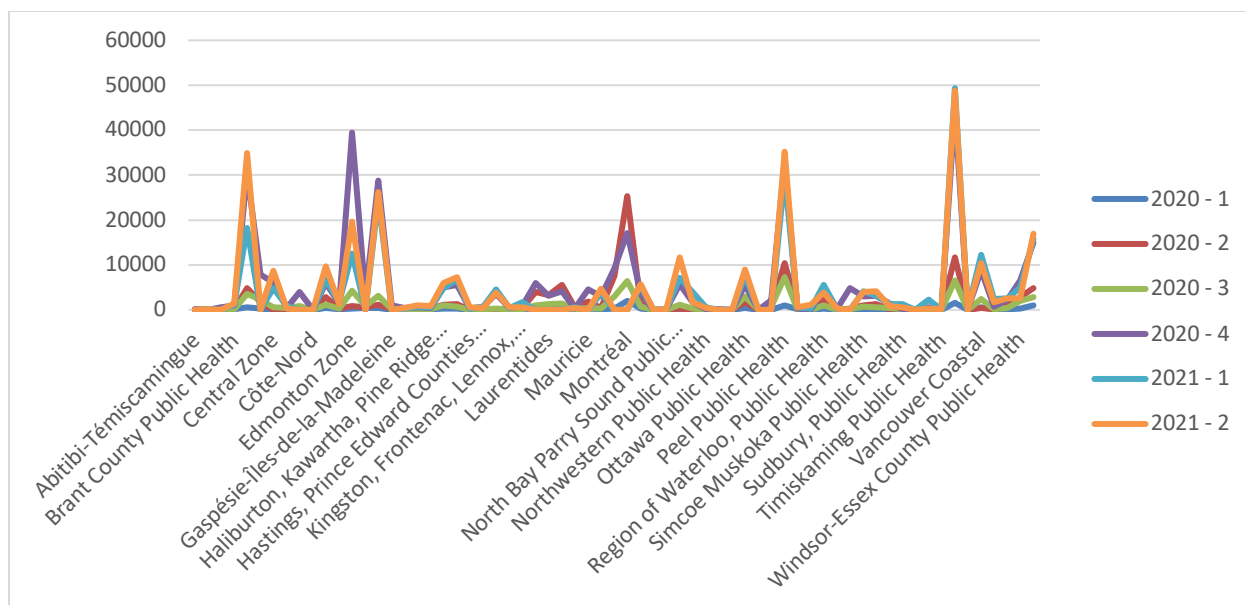
Thống kê **Mức Độ Nghiêm Trọng** (tiêu chí nghiêm trọng sinh viên tự định nghĩa) của dịch Covid-19 theo **PHU** và theo các **Quý trong từng năm**.

*MDX

	2020	2020	2020	2020	2021	2021	2021	2021	2022	2022	2022	2022
	1	2	3	4	1	2	3	4	1	2	3	4
1	1	1	1	1	1	1	1	1	1	1	1	1
York Region Public Health Services	5	6	6	6	6	6	1	1	1	1	1	1
Toronto Public Health	6	6	6	6	6	6	1	1	1	1	1	1
Peterborough Public Health	3	4	3	4	5	5	1	1	1	1	1	1
Haliburton, Kawartha, Pine Ridge Public Health	4	4	3	4	5	5	1	1	1	1	1	1
Durham Region Health Department	4	6	6	6	6	6	1	1	1	1	1	1
Wellington-Dufferin-Guelph Public Health	4	5	4	6	6	6	1	1	1	1	1	1
Leeds, Grenville, Lanark Public Health	4	5	3	4	5	4	1	1	1	1	1	1
Eastern Ontario Public Health	3	4	4	6	6	6	1	1	1	1	1	1
Ottawa Public Health	4	6	6	6	6	6	1	1	1	1	1	1
Renfrew County, Public Health	2	2	3	4	4	4	1	1	1	1	1	1
Chatham-Kent Public Health	2	4	4	4	5	4	1	1	1	1	1	1
Lambton Public Health	4	4	4	5	6	5	1	1	1	1	1	1
Windsor-Essex County Public Health	4	6	6	6	6	6	1	1	1	1	1	1
Brant County Public Health	3	4	4	5	6	6	1	1	1	1	1	1
Hamilton Public Health Services	4	6	5	6	6	6	1	1	1	1	1	1
Halton Region Health Department	4	6	5	6	6	6	1	1	1	1	1	1
Haldimand-Norfolk Public Health	4	5	3	5	5	5	1	1	1	1	1	1
Niagara Region Public Health Department	4	6	5	6	6	6	1	1	1	1	1	1
Peel Public Health	5	6	6	6	6	6	1	1	1	1	1	1
Northwestern Public Health	2	3	2	4	5	4	1	1	1	1	1	1
Timiskaming Public Health	1	2	1	3	2	3	1	1	1	1	1	1

*Report

Total Cases	Year ▼							
	2020				2021		Tổng Cuối	
Phu Name ▼	1	2	3	4	1	2		
Abitibi-Témiscamingue	34	138	28	68	0	0	268	
Algoma Public Public Health	14	34	20	55	150	156	429	
Bas-Saint-Laurent	13	47	270	550	0	0	880	
Brant County Public Health	72	185	154	923	1186	1421	3941	
Calgary Zone	552	4795	3579	30662	18238	34886	92712	
Capitale-Nationale	213	1632	2292	7852	0	0	11989	
Central Zone	48	44	579	6086	4892	8648	20297	
Chatham-Kent Public Health	24	285	408	458	819	216	2210	
Chaudière-Appalaches	117	404	802	4017	0	0	5340	
Côte-Nord	30	89	32	58	0	0	209	
Durham Region Health Department	408	2880	1082	6023	6750	9671	26814	
Eastern Ontario Public Health	77	242	255	1381	1748	1202	4905	
Edmonton Zone	206	831	4286	39414	12505	19682	76924	
Estrie	365	595	810	2923	0	0	4693	
Fraser	375	1164	3186	28843	24587	26160	84315	



3. Yêu cầu 3

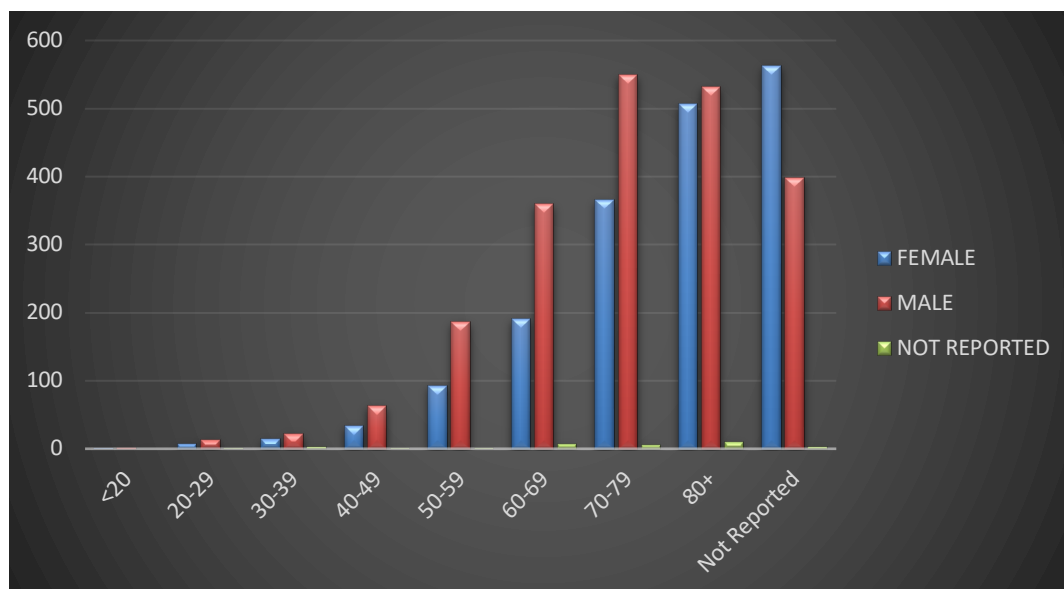
Thống kê tổng số người tử vong theo **Giới Tính và Nhóm Tuổi** theo các năm.

*MDX

		2020	2021
FEMALE	<20	1	1
FEMALE	20-29	8	7
FEMALE	30-39	4	14
FEMALE	40-49	25	34
FEMALE	50-59	102	93
FEMALE	60-69	267	191
FEMALE	70-79	597	366
FEMALE	80+	1571	507
FEMALE	Not Reported	1667	563
MALE	<20	0	2
MALE	20-29	7	13
MALE	30-39	20	22
MALE	40-49	50	63
MALE	50-59	167	186
MALE	60-69	503	360
MALE	70-79	964	549
MALE	80+	1317	532
MALE	Not Reported	1100	399
NOT REPORTED	20-29	0	1

*Report

Year	2021			
Total Fatal Cases	Gender			
Age Group	FEMALE	MALE	NOT REPORTED	Tổng Cuối
<20	1	2	0	3
20-29	7	13	1	21
30-39	14	22	3	39
40-49	34	63	1	98
50-59	93	186	1	280
60-69	191	360	6	557
70-79	366	549	5	920
80+	507	532	9	1048
Not Reported	563	399	3	965
Tổng Cuối	1776	2126	29	3931



4. Yêu cầu 4

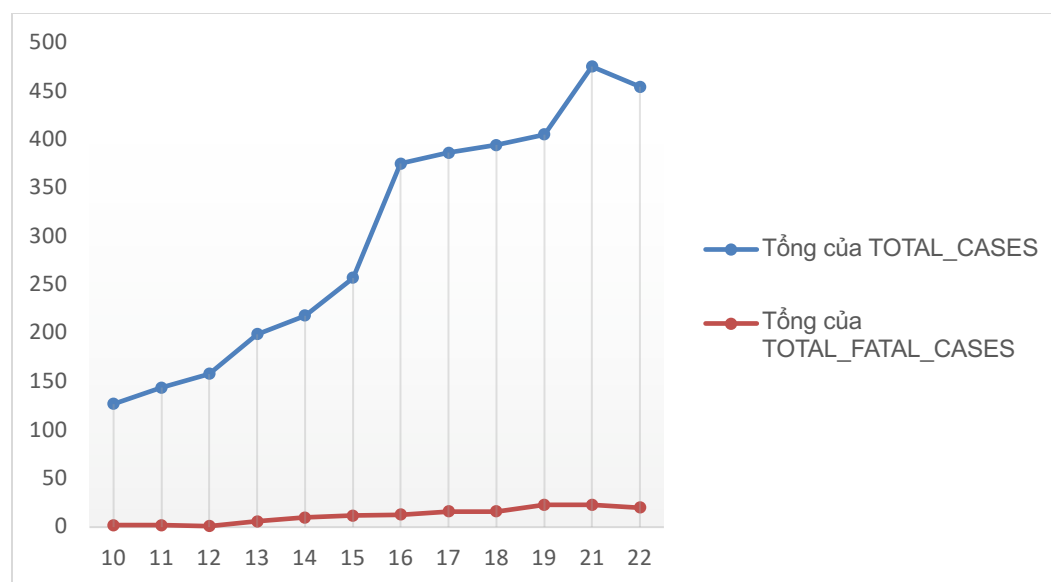
Thống kê số ca nhiễm, tử vong theo **Mức Độ Nghiêm Trọng** theo **Ngày Trong Tháng** của các năm.

*MDX

Messages			Results		
			Serious_Level	Total Cases	Total Fatal Cases
2020	1	1	1	3	0
2020	1	10	1	1	0
2020	1	11	1	(null)	(null)
2020	1	12	1	(null)	(null)
2020	1	13	1	1	0
2020	1	14	1	(null)	(null)
2020	1	15	1	(null)	(null)
2020	1	16	1	1	0
2020	1	17	1	(null)	(null)
2020	1	18	1	(null)	(null)
2020	1	19	1	(null)	(null)
2020	1	2	1	(null)	(null)
2020	1	20	1	(null)	(null)
2020	1	21	1	1	0
2020	1	22	1	2	0
2020	1	23	1	(null)	(null)
2020	1	24	1	1	0
2020	1	25	1	1	0
2020	1	26	1	(null)	(null)

*Report

SERIOUS_LEVEL	4	
YEAR	2020	
MONTH	3	
Ngày	Tổng của TOTAL_CASES	Tổng của TOTAL_FATAL_CASES
10	127	2
11	144	2
12	158	1
13	199	6
14	218	10
15	257	12
16	375	13
17	386	16
18	394	16
19	405	23
21	475	23
22	454	20
Tổng Cuối	3592	144



5. Yêu cầu 5

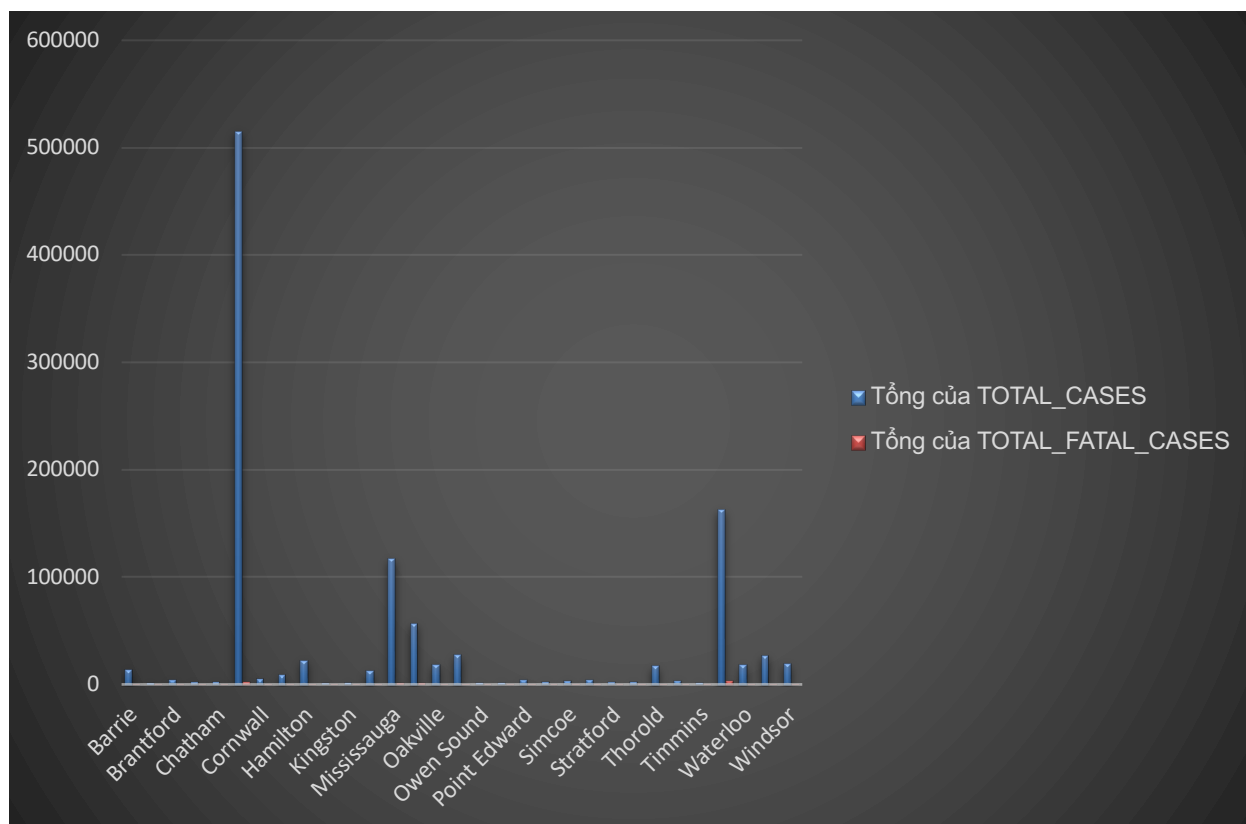
Thống kê số ca nhiễm, tử vong theo **Mức Độ Nghiêm Trọng**, khu vực (PHU_Group, City).

*MDX

		Serious_Level	Total Cases	Total Fatal Cases
		1	(null)	(null)
Central	Newmarket	6	56344	991
Central	Toronto	6	162671	3367
Central East	Peterborough	6	1624	23
Central East	Port Hope	6	2291	83
Central East	Whitby	6	26814	532
Central West	Guelph	6	8631	151
Champlain	Brockville	6	2107	109
Champlain	Cornwall	6	4905	128
Champlain	Ottawa	6	27180	574
Champlain	Pembroke	5	772	10
Erie St. Clair	Chatham	6	2210	23
Erie St. Clair	Point Edward	6	3819	85
Erie St. Clair	Windsor	6	19153	501
Hamilton Niagara Haldimand Brant	Brantford	6	3941	27
Hamilton Niagara Haldimand Brant	Hamilton	6	21833	433
Hamilton Niagara Haldimand Brant	Oakville	6	18244	251

*Report

Group Name	(Tất cả)	<input type="button" value="▼"/>
SEVERITY	6	<input type="button" value="▼"/>
City Name	Tổng của TOTAL_CASES	Tổng của TOTAL_FATAL_CASES
Barrie	13027	285
Belleville	1176	16
Brantford	3941	27
Brockville	2107	109
Chatham	2210	23
Cities outside Ontario	514814	2231
Cornwall	4905	128
Guelph	8631	151
Hamilton	21833	433
Kenora	1127	6
Kingston	1661	3
London	12319	220
Mississauga	116977	1058
Newmarket	56344	991
Oakville	18244	251
Ottawa	27180	574
Owen Sound	1433	7
Peterborough	1624	23
Point Edward	3819	85

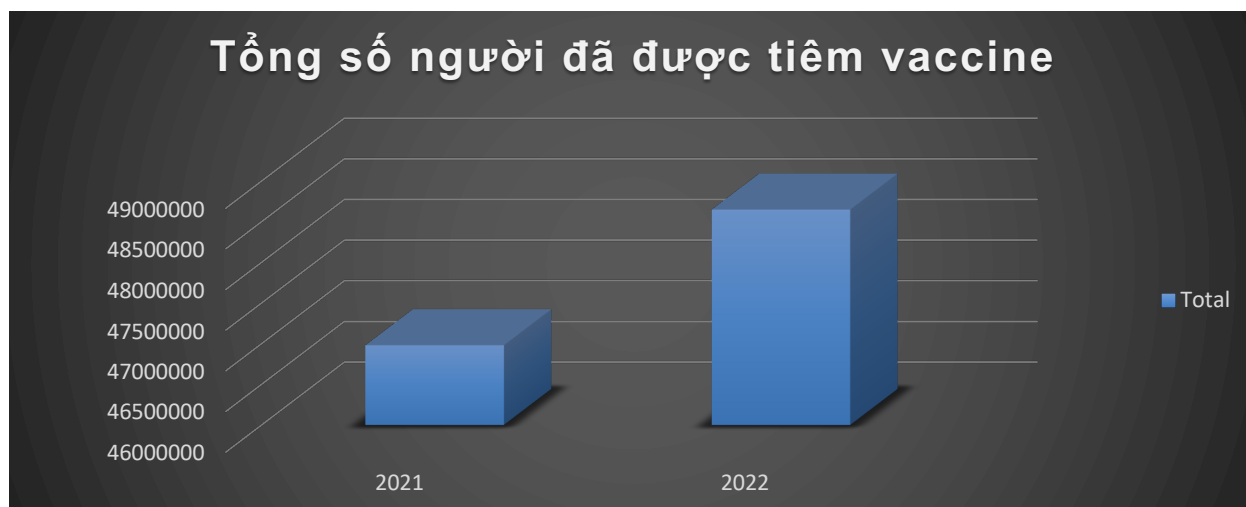


Thống kê **Số người đã được tiêm vaccin** trong các năm.

*SQL

	year	total
1	2021	46981455
2	2022	48649977

*Report



6. Yêu cầu 6

Thống kê số ca nhiễm theo **Mức Độ Nghiêm Trọng, nhóm bùng phát** của từng khu vực trong các năm.

- ⇒ không thể thực hiện được vì trong bảng case (cho biết số ca nhiễm) không cho biết thông tin nhóm bùng phát. Ngược lại bảng ongoing_outbreaks_phu (cho biết thông tin nhóm bùng phát) không cho biết thông tin số ca nhiễm.

7. Yêu cầu 7

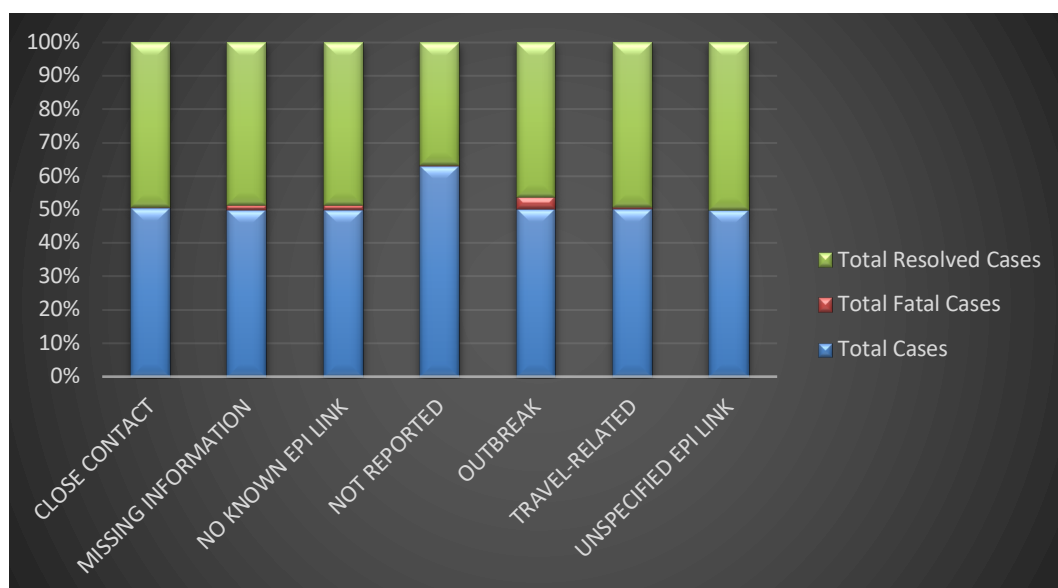
Thống kê **Số ca nhiễm, số ca tử vong** theo từng **loại phơi nhiễm** trong từng năm.

*MDX

	Total Cases	Total Fatal Cases	Total Resolved Cases
All	1077075	12339	761783
CLOSE CONTACT	248652	1212	243040
MISSING INFORMATION	254	7	247
NO KNOWN EPI LINK	6381	162	6219
NOT REPORTED	721779	4258	419802
OUTBREAK	89588	6582	82262
TRAVEL-RELATED	10288	118	10080
UNSPECIFIED EPI LINK	133	0	133

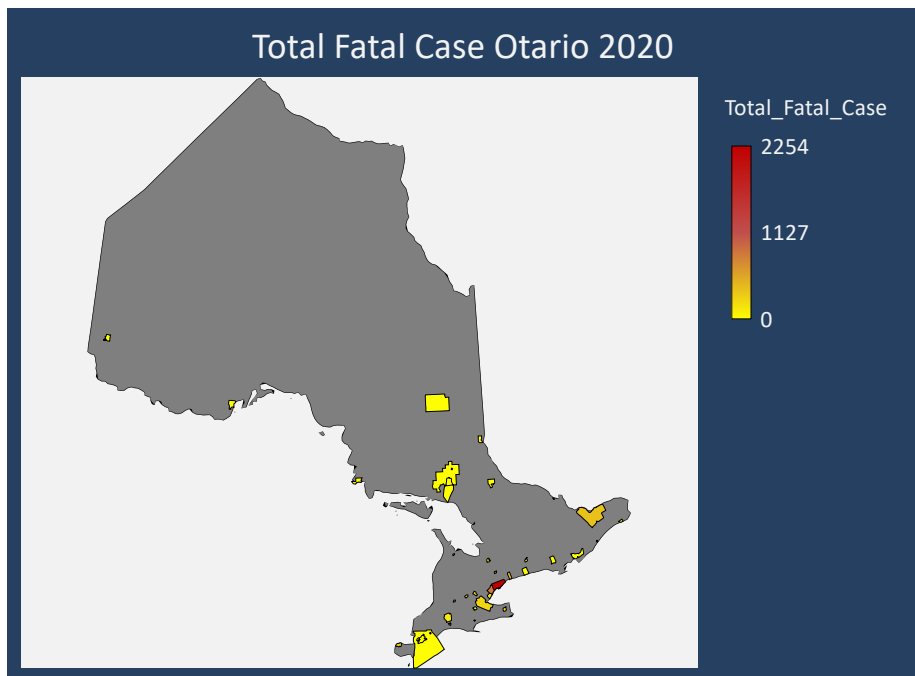
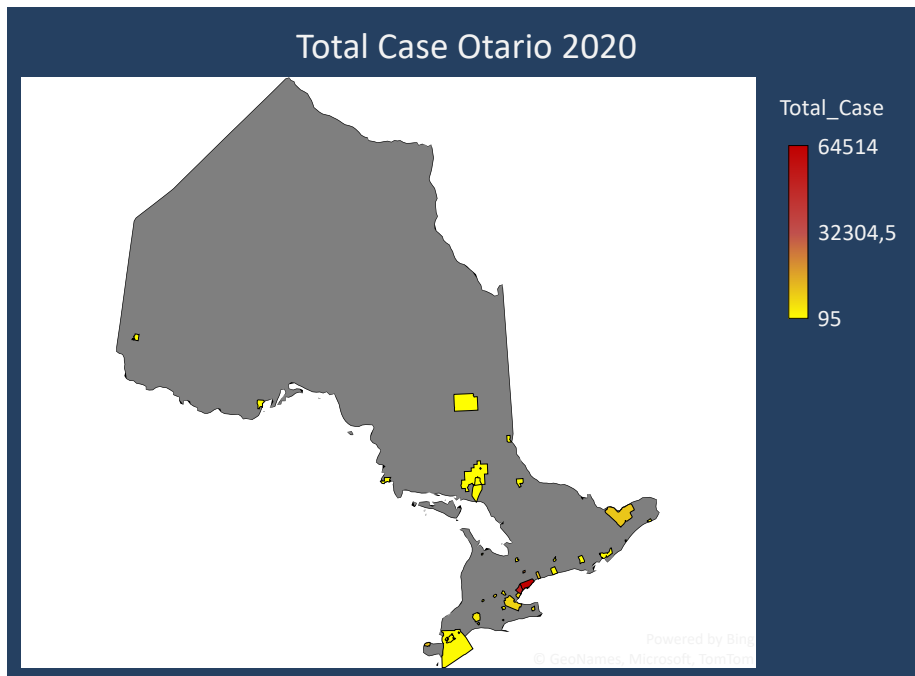
*Report

Exposure	☑ Total Cases	Total Fatal Cases	Total Resolved Cases
CLOSE CONTACT	248652	1212	243040
MISSING INFORMATION	254	7	247
NO KNOWN EPI LINK	6381	162	6219
NOT REPORTED	721779	4258	419802
OUTBREAK	89588	6582	82262
TRAVEL-RELATED	10288	118	10080
UNSPECIFIED EPI LINK	133	0	133
Tổng Cuối	1077075	12339	761783



8. Yêu cầu 9

Dùng **regional map** để biểu diễn trực quan (bằng màu sắc) số lượng ca nhiễm, số ca tử vong ở các city trong năm.



IV. Tạo Job tự động định kỳ thực hiện ETL

1. Các bước thực hiện

- Cài đặt Integration Services trong SQL Server Setup.

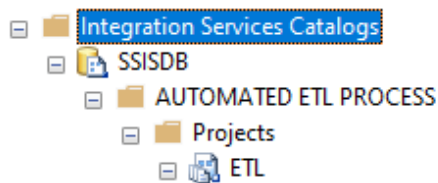
☒ Integration Services
☒ Scale Out Master
☒ Scale Out Worker
☐ Client Tools Background Compatibility

Select All Unselect All

Instance root directory: C:\Program Files\Microsoft SQL Server\ ...
 Shared feature directory: C:\Program Files\Microsoft SQL Server\ ...
 Shared feature directory (x86): C:\Program Files (x86)\Microsoft SQL Server\ ...

< Back Next > Cancel

- Tạo Integration Services Catalog.



- Deploy các package ETL.

Integration Services Deployment Wizard

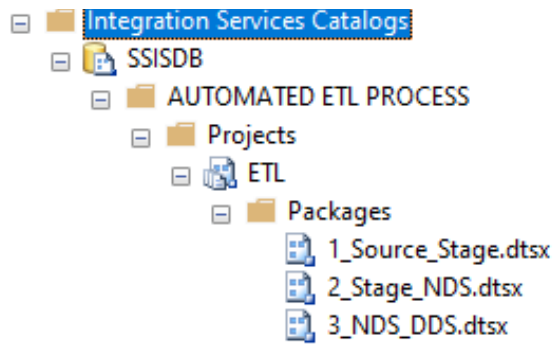
Select Destination

Introduction
 Select Source
 Select Deployment Target
Select Destination
 Review
 Results

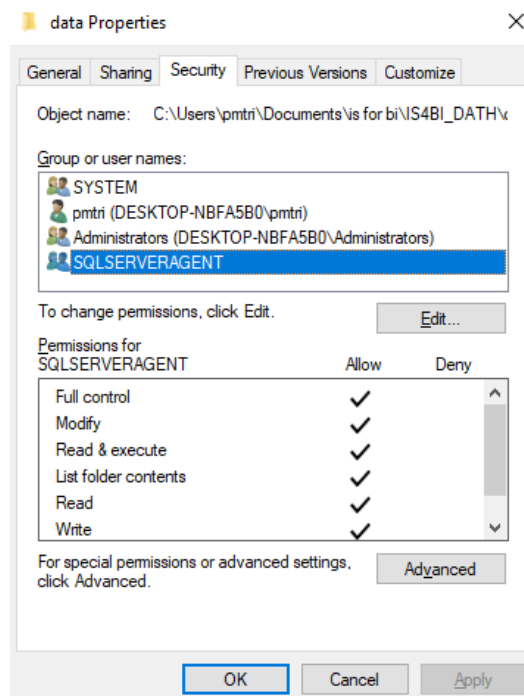
Enter the destination server name and where the project will be located in the Integration Services catalog.

Server name: Browse...
 Authentication: Windows Authentication Connect
 User name:
 Password:
 Path: Browse...

+ Kết quả sau khi deploy.



- Cấp quyền access cho account thực thi job với các source file cũng như các database.



- Chạy Service SQL Server Agent.

SQL Server (MSSQLSERVER)	Provides sto...	Running	Automatic	NT Service\MSSQLSERVER
SQL Server Agent (MSSQLSERVER)	Executes jo...	Running	Manual	NT SERVICE\SQLSERVERA...

- Tạo JOB chạy tự động theo lịch trong SSMS.

Job Properties - ETL_JOB

Select a page: General, Steps, Schedules, Alerts, Notifications, Targets

Connection: Server: DESKTOP-NBFA5B0, Connection: DESKTOP-NBFA5B0\pmtri, View connection properties

Progress: Ready

Name: ETL_JOB, Owner: DESKTOP-NBFA5B0\pmtri, Category: [Uncategorized (Local)], Description: No description available.

Enabled: ☒ Enabled, Source: , Created: 1/11/2023 3:44:19 PM, Last modified: 1/11/2023 4:00:06 PM, Last executed: 1/12/2023 4:13:00 PM, View Job History

OK Cancel

Job Properties - ETL_JOB

Select a page: General, Steps, Schedules, Alerts, Notifications, Targets

Connection: Server: DESKTOP-NBFA5B0, Connection: DESKTOP-NBFA5B0\pmtri, View connection properties

Progress: Ready

Job step list:

Step	Name	Type	On Success	On Failure
1	Step1_Source_Stage	SQL Server Integrat...	Go to the next step	Quit the job reporting failure
2	Step2_Stage_NDS	SQL Server Integrat...	Go to the next step	Quit the job reporting failure
3	Step3_NDS_DDS	SQL Server Integrat...	Quit the job reportin...	Quit the job reporting failure

Move step: Start step: 1.Step1_Source_Stage, New..., Insert..., Edit, Delete

Activate Windows, Go to Settings to activate Windows, OK, Cancel

Schedule list:

ID	Name	Enabled	Description	Jobs...
13	ETL_SCH	Yes	Occurs every day at 4:13:00 PM. Schedule will be used between 1/11/2023 and 1/14/...	View

Jobs Referencing a Schedule

Schedule: ETL_SCH

Select jobs that reference this schedule:

Selected	Name	Enabled	Category
<input checked="" type="checkbox"/>	ETL_JOB	<input checked="" type="checkbox"/>	[Uncategorized (Local)]

2. Kết quả

Log file summary: No filter applied				
Date ▾	Step ID	Server	Job Name	Step Name
☐ ✓ 1/12/2023 4:13:00 PM		DESKTOP-NBFA5B0	ETL JOB	
✓ 1/12/2023 4:13:53 ...	3	DESKTOP-NBFA5B0	ETL JOB	Step3_NDS_DDS
✓ 1/12/2023 4:13:15 ...	2	DESKTOP-NBFA5B0	ETL JOB	Step2_Stage_NDS
✓ 1/12/2023 4:13:01 ...	1	DESKTOP-NBFA5B0	ETL JOB	Step1_Source_Stage
☐ ✓ 1/11/2023 4:13:00 PM		DESKTOP-NBFA5B0	ETL JOB	
✓ 1/11/2023 4:13:53 ...	3	DESKTOP-NBFA5B0	ETL JOB	Step3_NDS_DDS
✓ 1/11/2023 4:13:12 ...	2	DESKTOP-NBFA5B0	ETL JOB	Step2_Stage_NDS
✓ 1/11/2023 4:13:00 ...	1	DESKTOP-NBFA5B0	ETL JOB	Step1_Source_Stage
☐ ✓ 1/11/2023 4:07:04 PM		DESKTOP-NBFA5B0	ETL JOB	
✓ 1/11/2023 4:07:53 ...	3	DESKTOP-NBFA5B0	ETL JOB	Step3_NDS_DDS
✓ 1/11/2023 4:07:15 ...	2	DESKTOP-NBFA5B0	ETL JOB	Step2_Stage_NDS
✓ 1/11/2023 4:07:04 ...	1	DESKTOP-NBFA5B0	ETL JOB	Step1_Source_Stage

V. Data Mining

1. Mục tiêu

Dự đoán số ca nhiễm của PHU mã 162 theo từng ngày (do đây là PHU có nhiều dòng dữ liệu nhất được ghi nhận trong bảng fact) bằng mô hình Time Series

2. Các bước thực hiện

- Bước 1: Tạo bảng tổng kết số ca nhiễm được ghi nhận ở PHU mã 162 theo ngày (ở đây nhóm sử dụng Python để làm).

Bảng tổng kết gồm 2 thuộc tính:

+ Date: Ngày ghi nhận (khóa chính).

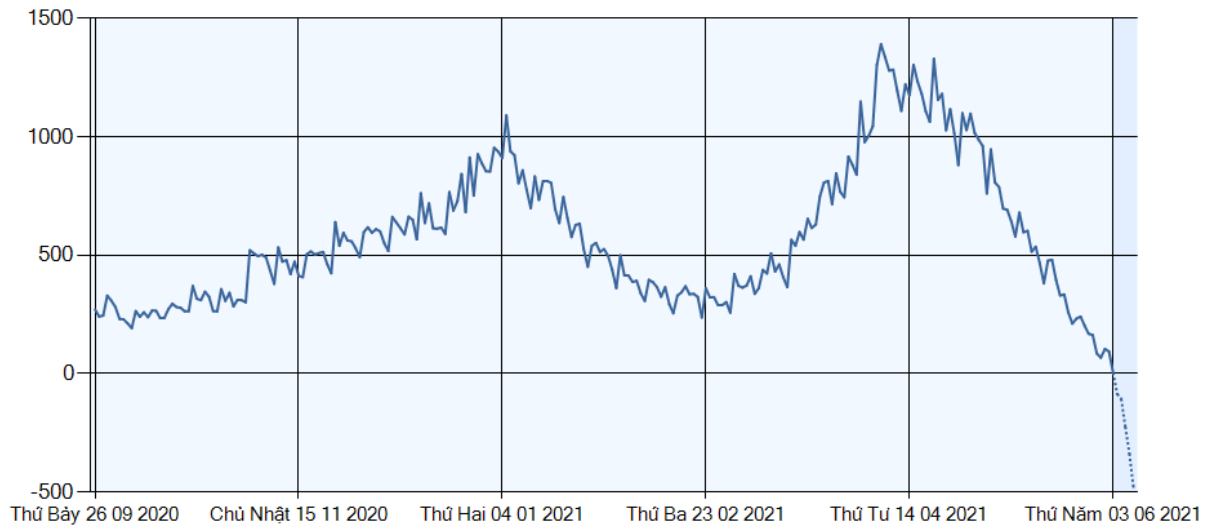
+ Total Cases: tổng số ca nhiễm được ghi nhận trong ngày đó.

- Bước 2: Đưa bảng tổng kết vào train tại mô hình Time Series.

Nhóm đã đưa thuộc tính Date làm Key Date, thuộc tính Total Cases được dùng để vừa làm input cho việc training mô hình, vừa làm thuộc tính để dự đoán.

- Bước 3: Chạy mô hình và xem kết quả được visualize trong SSAS.

3. Kết quả



C. THAM KHẢO

1. [Building a Data Warehouse: With Examples in SQL Server](#)
2. [Creating a date dimension or calendar table in SQL Server](#)
3. [Schedule SSIS Package in SQL Server Agent](#)
4. [How to Plot Cities on a Map in Excel \(2 Easy Methods\)](#)