

Міністерство освіти і науки України
Національний університет “Львівська політехніка”
Інститут прикладної математики та фундаментальних наук
Кафедра прикладної математики



Звіт
про виконання курсової роботи
з курсу “ Робота з великими базами даних”

Виконала:

Студентка групи ПМ-42

Карпин О. І.

Перевірив:

Любінський Б. Б.

Львів-2025 р.

Розроблено інформаційно-аналітичну систему на базі Microsoft SQL Server з використанням технологій SSIS, SSAS та SSRS, що забезпечує реалізацію ETL-процесів, побудову сховища даних, OLAP-куба та формування аналітичних звітів для підтримки прийняття управлінських рішень у сфері організації наукових конференцій.

An information and analytical system based on Microsoft SQL Server has been developed using SSIS, SSAS, and SSRS technologies, which ensures the implementation of ETL processes, the construction of a data warehouse, an OLAP cube, and the generation of analytical reports to support management decision-making in the field of organizing scientific conferences.

Зміст:

| | |
|---|----|
| Вступ..... | 1 |
| Аналіз предметної області..... | 4 |
| Проектування бази даних..... | 6 |
| Реалізація ETL-процесів..... | 10 |
| Побудова OLAP-куба та аналітичні звіти..... | 19 |
| Висновок..... | 29 |
| Список використаних джерел..... | 30 |
| Додатки..... | 31 |

Вступ

У сучасних умовах стрімкого розвитку інформаційних технологій та цифровізації бізнес-процесів особливого значення набуває ефективна робота з великими та надвеликими обсягами даних. Організації різних сфер діяльності накопичують значні масиви інформації, аналіз яких є необхідним для прийняття обґрунтovanих управлінських рішень, прогнозування тенденцій та оцінки ефективності діяльності.

Надвеликі бази даних характеризуються значним обсягом, різноманіттям структур та високими вимогами до швидкодії обробки інформації. Традиційні підходи до зберігання та аналізу даних стають недостатньо ефективними, що зумовлює необхідність використання сучасних технологій Data Warehouse, ETL-процесів та OLAP-аналітики.

Метою даної курсової роботи є розробка та аналіз надвеликої бази даних для обраної предметної області з використанням технологій Microsoft SQL Server, а саме SQL Server Integration Services (SSIS), SQL Server Analysis Services (SSAS) та SQL Server Reporting Services (SSRS). У межах роботи реалізовано повний цикл обробки даних: від проєктування бази даних та наповнення її великими обсягами інформації до побудови багатовимірного OLAP-куба та створення аналітичних звітів.

Великі дані (Big Data) – позначення структурованих и неструктурзованих даних величезних обсягів і значного розмаїття, що піддаються ефективній обробці програмних інструментів, які горизонтально масштабуються та з'явилися у кінці 2000-х років, і альтернативних традиційних систем управління базами даних і рішенням класу рішень Business Intelligence.

Технології і тенденції роботи з Big Data:

Початково у сукупність підходів і технологій включались засоби масовопаралельної обробки невизначено-структурзованих даних, такі як СУБД NoSQL, алгоритми MapReduce і засоби проекту Hadoop. У подальшому до технологій великих даних почали відносити й інші рішення, що забезпечують схожі за характеристиками можливості обробки надвеликих масивів даних, а також деякі апаратні засоби.

MapReduce — модель розподілених обчислювань у комп’ютерних кластерах, представлена компанією Google. Згідно з цією моделлю, додаток розділяється на значну кількість однакових елементарних завдань, що виконуються на вузлах кластера і потім, природнім шляхом зводяться у кінцевий результат.

SQL - мова структурованих запитів, що дозволяє працювати з базами даних.

За допомогою SQL можна створювати і модифікувати дані, а управлінням масиву даних займається відповідна система управління базами даних.

NoSQL (Not Only SQL, не лише SQL) — загальний термін для різних нереляційних баз даних і сховищ, не означає якусь конкретну технологію чи продукт. Звичайні реляційні бази даних добре підходять для досить швидких і однотипних запитів, а на складних і гнучко побудованих запитах, характерних для великих даних, навантаження перевищує розумні межі і використання СУБД стає неефективним. Основні риси: – базова доступність - запити гарантовано завершується (успішно чи безуспішно); – гнучкий стан - стан системи може змінюватися з часом, навіть без введення нових даних, для досягнення узгодження даних; – узгодженість в кінцевому рахунку - дані можуть бути деякий час неузгодженими, але приходять до узгодження через деякий час.

Hadoop — проект фонду Apache Software Foundation, набір утилітів, бібліотек і фреймворків, що вільно розповсюджується, для розробки і виконання розподілених програм, які працюють на кластерах із сотень і тисяч вузлів. Вважається однією з основоположних технологій більшості даних. Використовується для реалізації пошукових і контекстних механізмів багатьох високонавантажених веб-сайтів, у тому числі, для Yahoo! та Facebook. Розроблено на Java в рамках обчислювальної парадигми MapReduce, згідно з якою додаток розділяється на велику кількість однакових елементарних завдань, здійснених на вузлах кластера і природним чином приводяться в кінцевий результат.

У даній роботі використано такі компоненти Microsoft SQL Server:

SQL Server Integration Services (SSIS) — для реалізації процесів Extract, Transform, Load (ETL), очищення, трансформації та завантаження великих обсягів даних у сховище даних.

SQL Server Analysis Services (SSAS) — для побудови багатовимірного OLAP-куба, створення вимірів, мір, обчислюваних показників та реалізації MDX-запитів і KPI.

SQL Server Reporting Services (SSRS) — для створення аналітичних звітів різних типів (табличних, матричних, графічних, дашбордів) з підтримкою параметризації, drill-down та інтерактивних можливостей.

Застосування даних технологій дозволяє побудувати масштабовану, надійну та ефективну систему аналітичної обробки даних, що повністю відповідає вимогам до роботи з надвеликими базами даних.

Аналіз предметної області

"Конференції"

Основні сутності: Конференції, Секції, Виступи, Виступаючі, Обладнання, Приміщення.

Специфічні вимоги:

- Мінімум 1 000 конференцій
- Мінімум 100 000 виступів
- Історія за 10 років

Обов'язкові звіти:

1. Розклад конференції
2. Список учасників з місцем праці
3. Потреба в обладнанні
4. Статистика виступів по науковим ступеням
5. Аналіз популярності тематик

Предметна область «Конференції» охоплює процеси організації та проведення наукових заходів, що включають управління конференціями, секціями, виступами, виступаючими, приміщеннями та обладнанням. У межах цієї області формується значний обсяг даних, які накопичуються протягом тривалого часу та мають виражений часовий характер.

Дані предметної області є взаємопов'язаними та ієрархічними: кожна конференція складається з кількох секцій, у межах яких проводяться виступи. Кожен виступ пов'язаний із конкретним виступаючим, тематикою, приміщенням та часом проведення. Окрім роль відіграє інформація про наукові ступені учасників і використане обладнання, що є важливим для аналітичних досліджень.

Для даної предметної області характерна потреба у багатовимірному аналізі, зокрема аналізі розкладів конференцій, складу учасників, навантаження на приміщення та обладнання, а також популярності наукових тематик і статистики виступів за науковими ступенями. Операційні бази даних не забезпечують достатньої ефективності для виконання таких аналітичних запитів.

З огляду на це, використання сховища даних є доцільним, оскільки воно дозволяє зберігати історичні дані, виконувати агрегацію інформації та забезпечувати швидке формування аналітичних звітів для підтримки прийняття управлінських рішень.

Структура бази даних:

Операційна база даних conferences_db:

Операційна база даних містить таблиці, що зберігають детальну інформацію про наукові конференції, учасників та виступи.

Таблиця conference

Містить загальну інформацію про наукові конференції: назву конференції, дати проведення та місто, у якому вона відбувається.

Таблиця sections

Призначена для зберігання інформації про секції конференцій. Кожна секція пов'язана з певною конференцією та має власну назву.

Таблиця topics

Містить перелік наукових тематик, до яких належать виступи на конференціях.

Таблиця speakers

Зберігає інформацію про виступаючих, зокрема прізвище та ім'я, науковий ступінь і місце роботи.

Таблиця rooms

Містить дані про приміщення, у яких проводяться виступи, включаючи назву та місткість аудиторії.

Таблиця equipment

Призначена для зберігання інформації про наявне обладнання.

Таблиця room_equipment

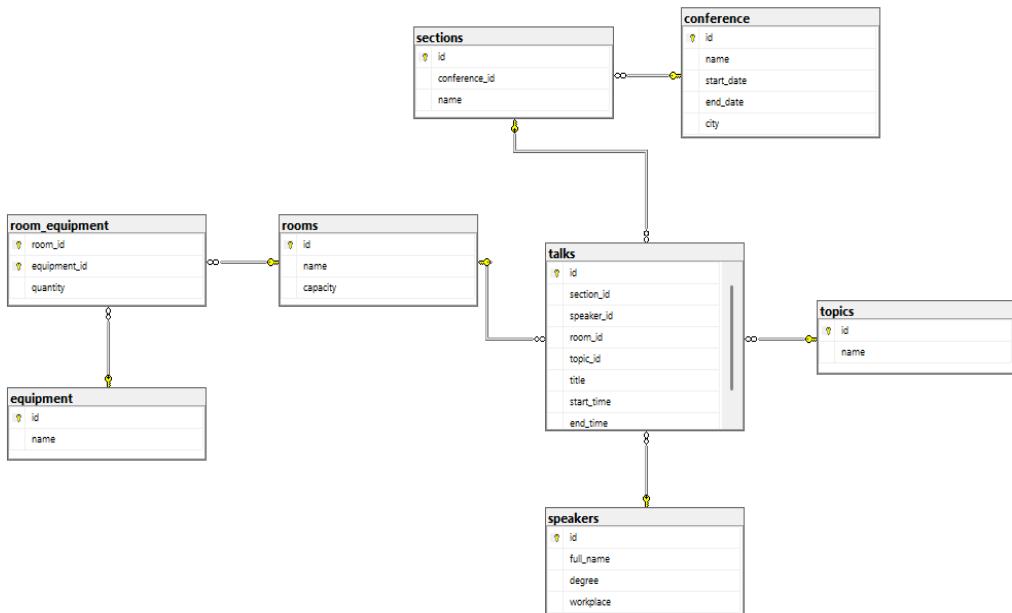
Реалізує зв'язок між приміщеннями та обладнанням і містить інформацію про кількість одиниць обладнання в кожному приміщенні.

Таблиця talks

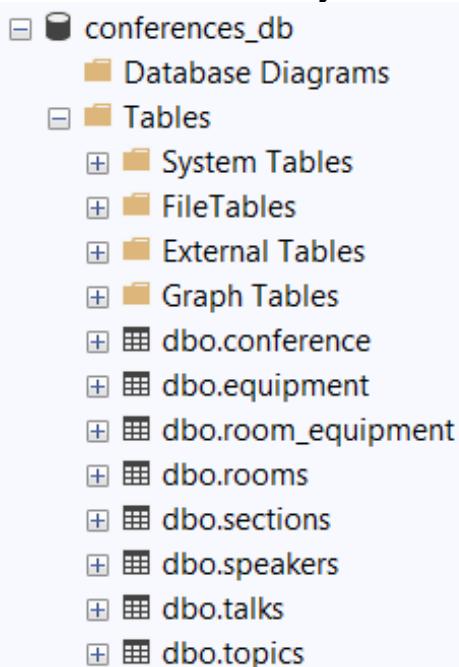
Основна таблиця операційної бази, яка містить інформацію про виступи: секцію, виступаючого, приміщення, тематику, назву виступу, дату та час початку і завершення.

Проектування бази даних

Схема бази даних:



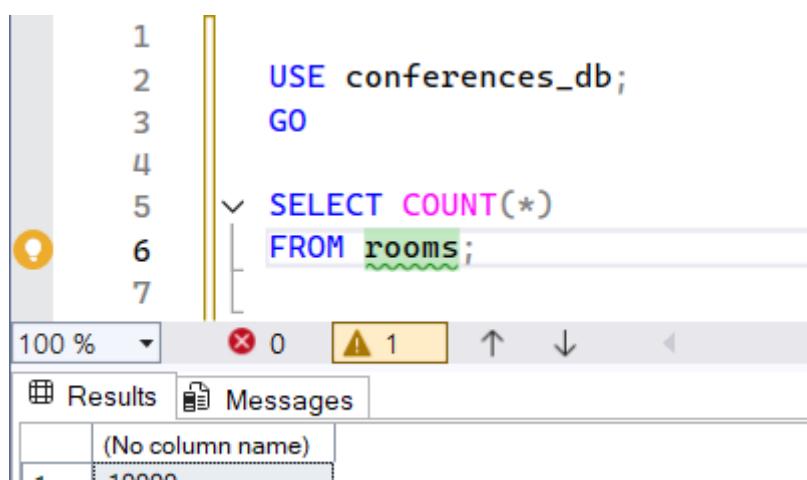
Проектування бази даних виконано з використанням SQL-скриптів, які визначають структуру таблиць, їхні атрибути, первинні та зовнішні ключі. У процесі проектування було виділено основні сутності предметної області та встановлено логічні зв'язки між ними, що забезпечує цілісність і узгодженість даних. Структура бази даних оптимізована для подальшого завантаження даних у сховище даних та виконання аналітичної обробки.



Генерація даних та їх запис:

Для наповнення бази даних було використано інструмент **Redgate SQL Data Generator**, який дозволяє автоматично створювати великі обсяги тестових даних. Застосування даного інструменту дало змогу швидко згенерувати реалістичні та взаємоузгоджені дані з дотриманням усіх зв'язків між таблицями.

Використання Redgate SQL Data Generator забезпечило формування необхідних обсягів даних для подальшого тестування ETL-процесів, побудови сховища даних, OLAP-куба та аналітичних звітів, що неможливо ефективно виконати за допомогою ручного введення даних.

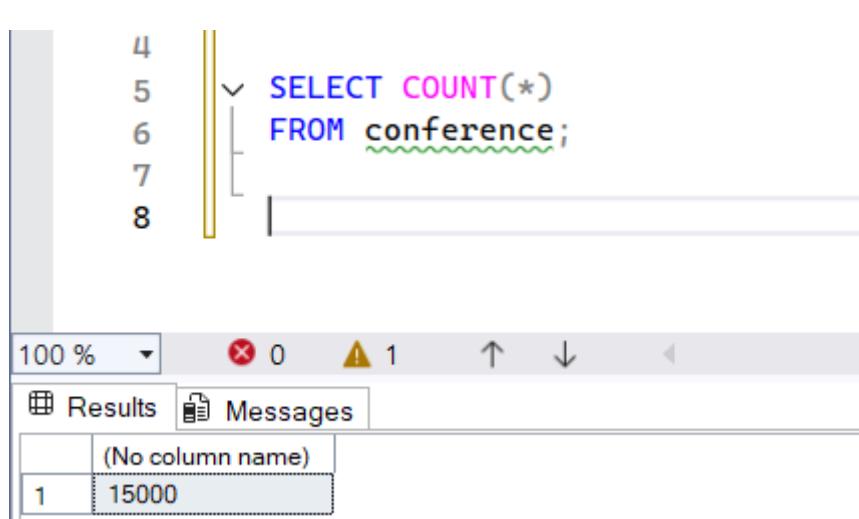


1
2 USE conferences_db;
3 GO
4
5 SELECT COUNT(*)
6 FROM rooms;
7

100 % ✖ 0 ⚠ 1 ↑ ↓ ◀

Results Messages

| | (No column name) |
|---|------------------|
| 1 | 10000 |



4
5 SELECT COUNT(*)
6 FROM conference;
7
8

100 % ✖ 0 ⚠ 1 ↑ ↓ ◀

Results Messages

| | (No column name) |
|---|------------------|
| 1 | 15000 |

```
1 USE conferences_db;
2 GO
3
4
5
6
7
```

100 % 0 1 ↑ ↓

Results Messages

| | (No column name) |
|---|------------------|
| 1 | 2000 |

```
1
2
3
4
5
6
7
```

USE conferences_db;
GO

SELECT COUNT(*)
FROM room_equipment;

100 % 0 1 ↑ ↓

Results Messages

| | (No column name) |
|---|------------------|
| 1 | 10000 |

```
1
2
3
4
5
6
7
```

USE conferences_db;
GO

SELECT COUNT(*)
FROM sections;

100 % 0 1 ↑ ↓

Results Messages

| | (No column name) |
|---|------------------|
| 1 | 20000 |

```
1 USE conferences_db;
2 GO
3
4
5   ▾ SELECT COUNT(*)
6     FROM speakers;
7
```

100 % 0 1 ↑ ↓

Results Messages

| | (No column name) |
|---|------------------|
| 1 | 40000 |

```
SQLQuery1.sq...db (sa (60))*
1 USE conferences_db;
2 GO
3
4
5   ▾ SELECT COUNT(*)
6     FROM talks;
7
```

100 % 0 1 ↑ ↓

Results Messages

| | (No column name) |
|---|------------------|
| 1 | 600000 |

```
1 USE conferences_db;
2 GO
3
4
5   ▾ SELECT COUNT(*)
6     FROM topics;
7
```

100 % 0 1 ↑ ↓

Results Messages

| | (No column name) |
|---|------------------|
| 1 | 50000 |

Бачимо, що всі дані успішно згенерувались і були записані у таблиці.

Реалізація ETL-процесів

Побудова Data Warehouse: Сховище даних conferences_dw:

Таблиця Dim_Conference

Таблиця виміру, яка містить узагальнену інформацію про конференції: назву, дати проведення та місто.

Таблиця Dim_Section

Таблиця виміру, що зберігає інформацію про секції конференцій та їх належність до відповідної конференції.

Таблиця Dim_Speaker

Містить інформацію про виступаючих, включаючи ім'я, науковий ступінь та місце роботи.

Таблиця Dim_Room

Таблиця виміру, що містить дані про приміщення та їх місткість.

Таблиця Dim_Equipment

Зберігає перелік обладнання, яке використовується під час проведення конференцій.

Таблиця Dim_Topic

Містить інформацію про наукові тематики виступів.

Таблиця Dim_Time

Часовий вимір, який використовується для аналізу даних у часовому розрізі. Містить інформацію про дату, рік, квартал, місяць, день тижня та додаткові часові характеристики.

Таблиця Fact_Talks

Таблиця фактів, яка містить кількісні показники виступів, зокрема тривалість виступів, кількість слухачів, а також зовнішні ключі до відповідних вимірів.

Таблиця Fact_Equipment_Usage

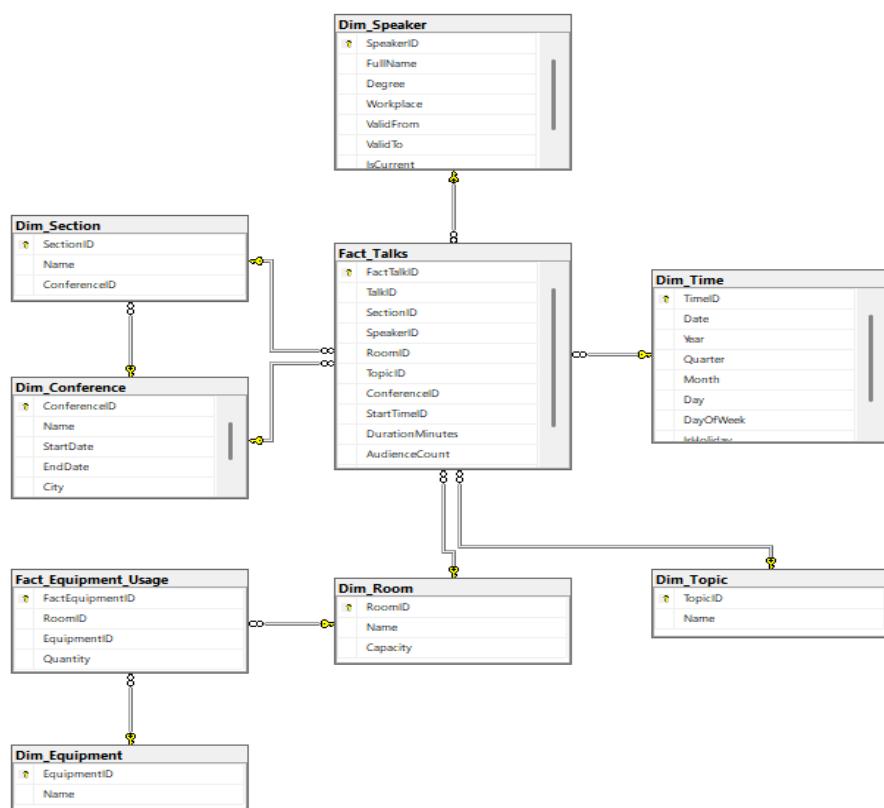
Таблиця фактів, яка містить інформацію про використання обладнання під

час проведення виступів.

- + dbo.Dim_Conference
- + dbo.Dim_Equipment
- + dbo.Dim_Room
- + dbo.Dim_Section
- + dbo.Dim_Speaker
- + dbo.Dim_Time
- + dbo.Dim_Topic
- + dbo.Fact_Equipment_Usage
- + dbo.Fact_Talks

Проектування бази даних виконано шляхом написання та виконання SQL-скриптів, які визначають структуру таблиць сховища даних. За допомогою скриптів було створено таблиці вимірів і таблиці фактів, а також задано первинні та зовнішні ключі для забезпечення цілісності даних.

Схема бази даних:

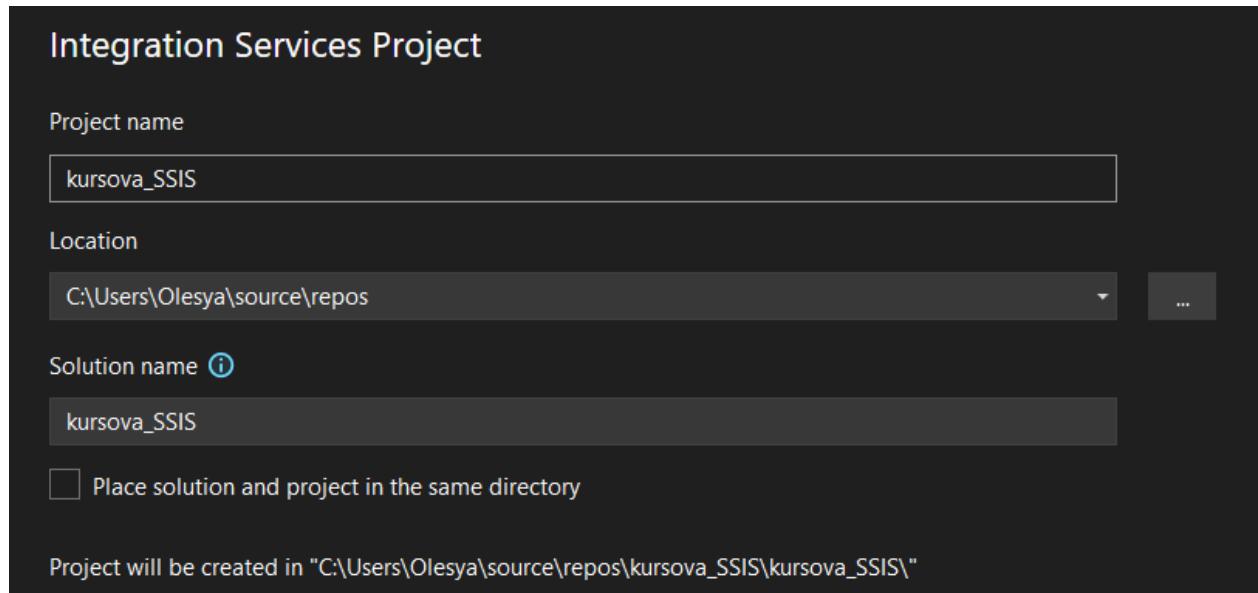


Реалізація ETL-процесів:

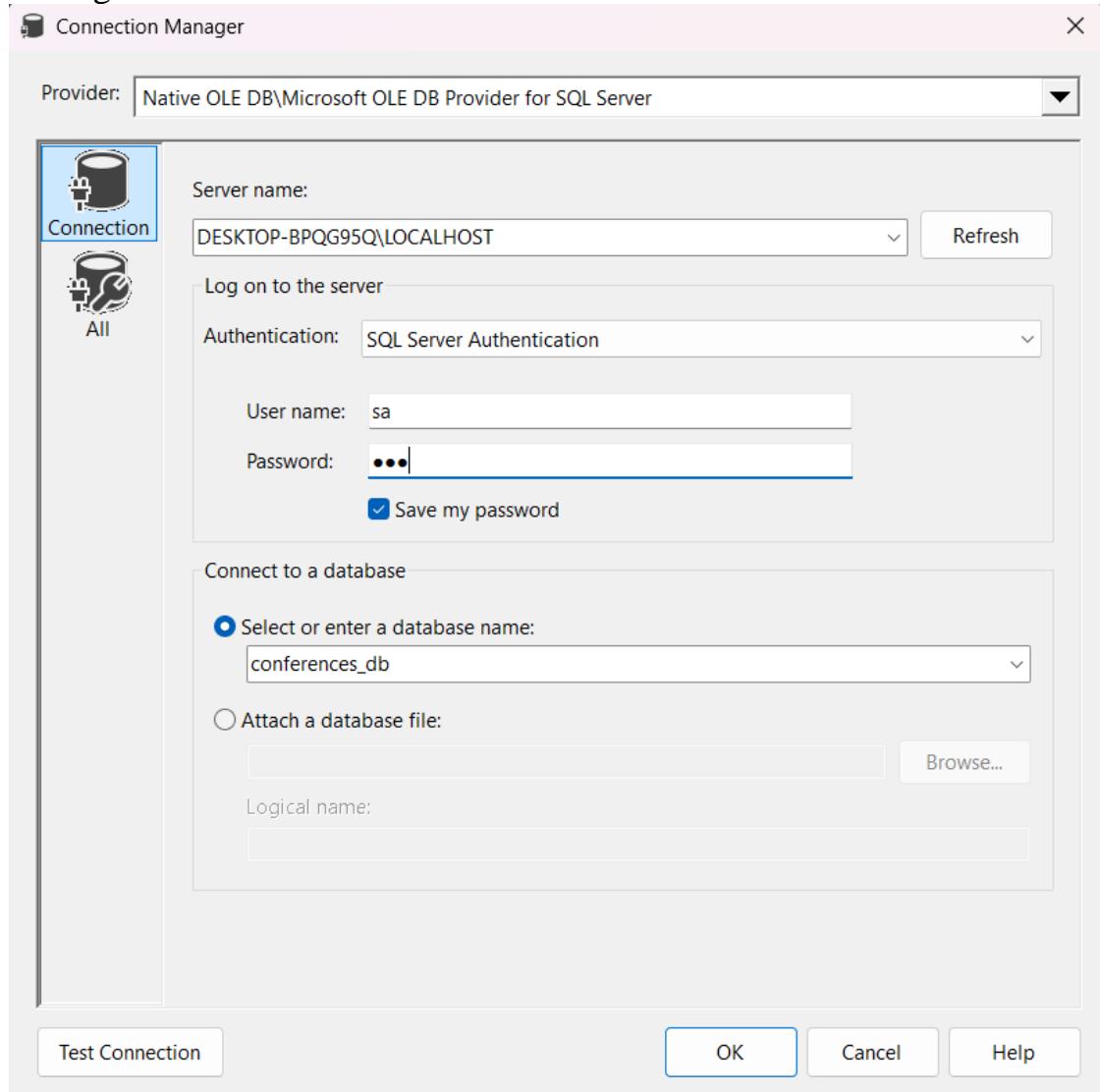
Для забезпечення коректного та ефективного завантаження даних із операційної бази даних у сховище даних було реалізовано ETL-процеси з використанням технології SSIS.

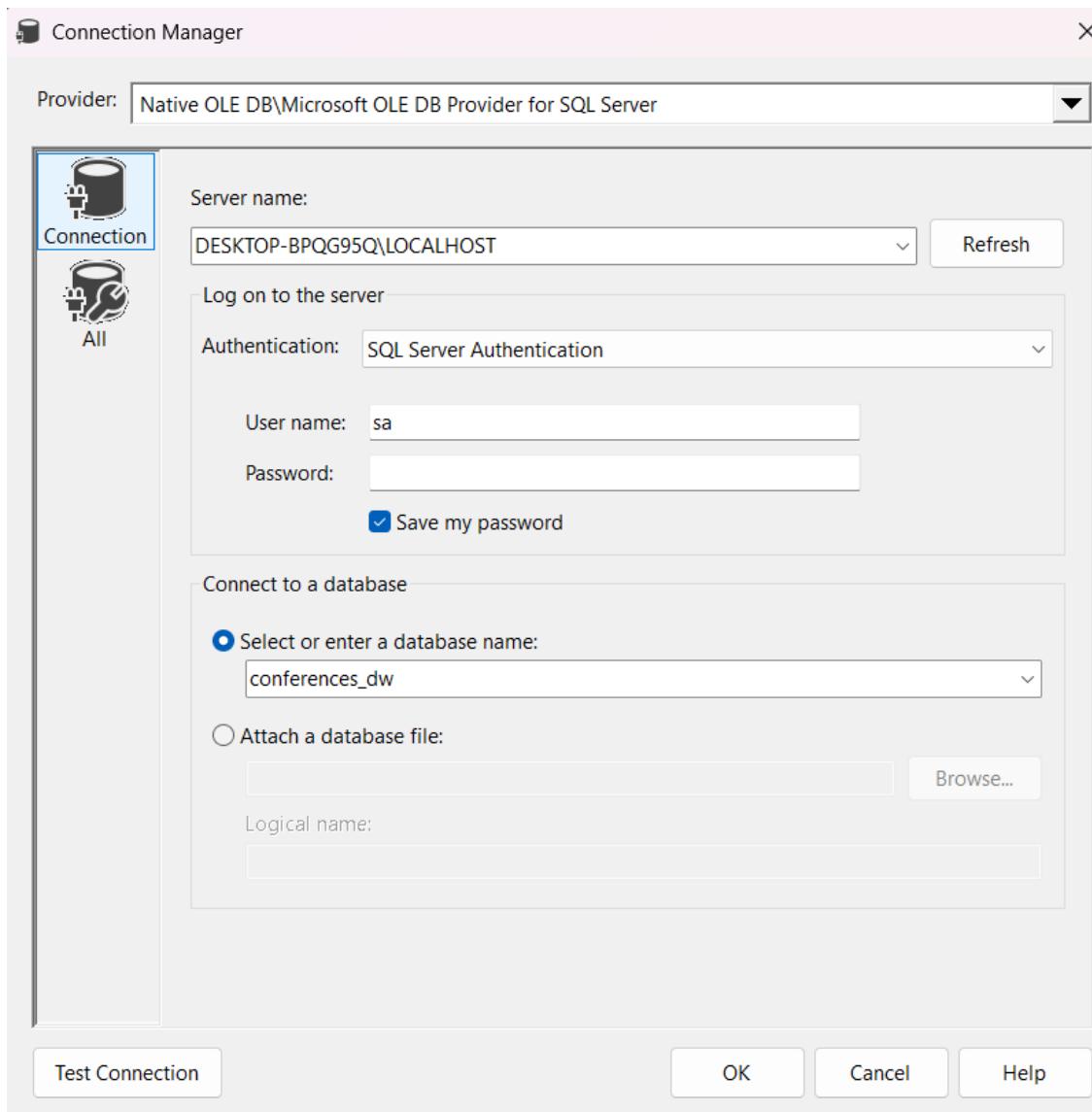
В середовищі Visual Studio, створено новий SSIS-проект під назвою

kursova_SSIS.

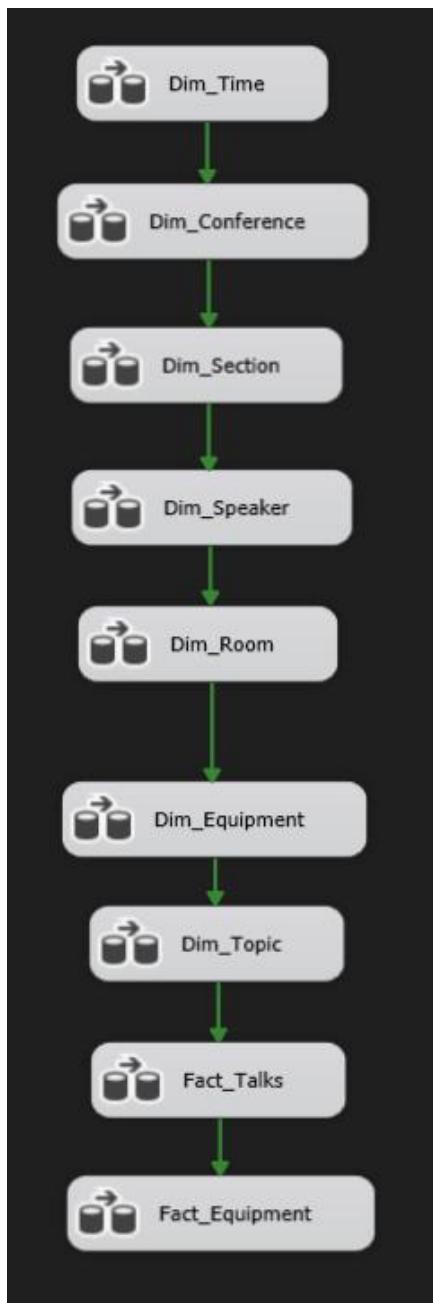


Підключено до бази даних та сховища даних за допомогою Connection Manager:

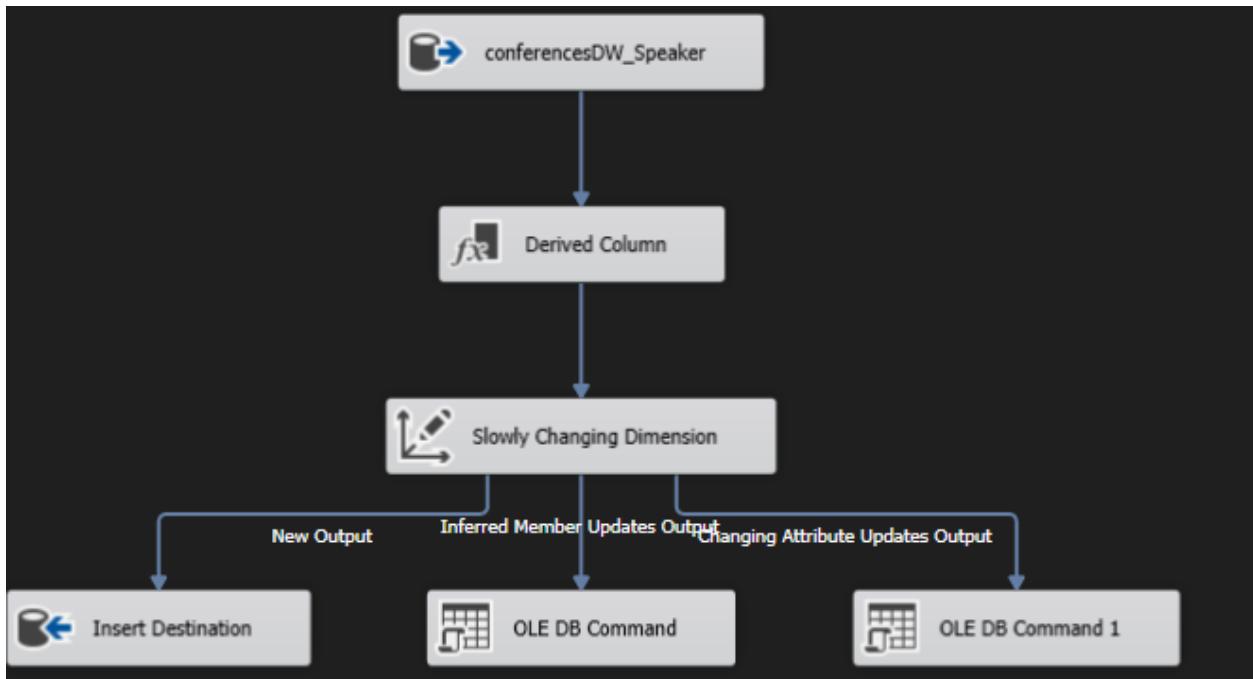




На етапі витягування дані отримуються з операційної бази даних conferences_db, яка містить інформацію про конференції, секції, виступи, виступаючих, приміщення, тематики та обладнання. Для кожної таблиці джерела було створено окремі Data Flow Task, у яких використано компонент OLE DB Source.



В Dim_Speaker Data Flow додано компонент Slowly Changing Dimension, в якому ID виступаючого було обрано як бізнес ключ. Решту полів таблиці обрано для SCD Type-1. Це дозволяє оновлювати атрибути без збереження історії попередніх значень.



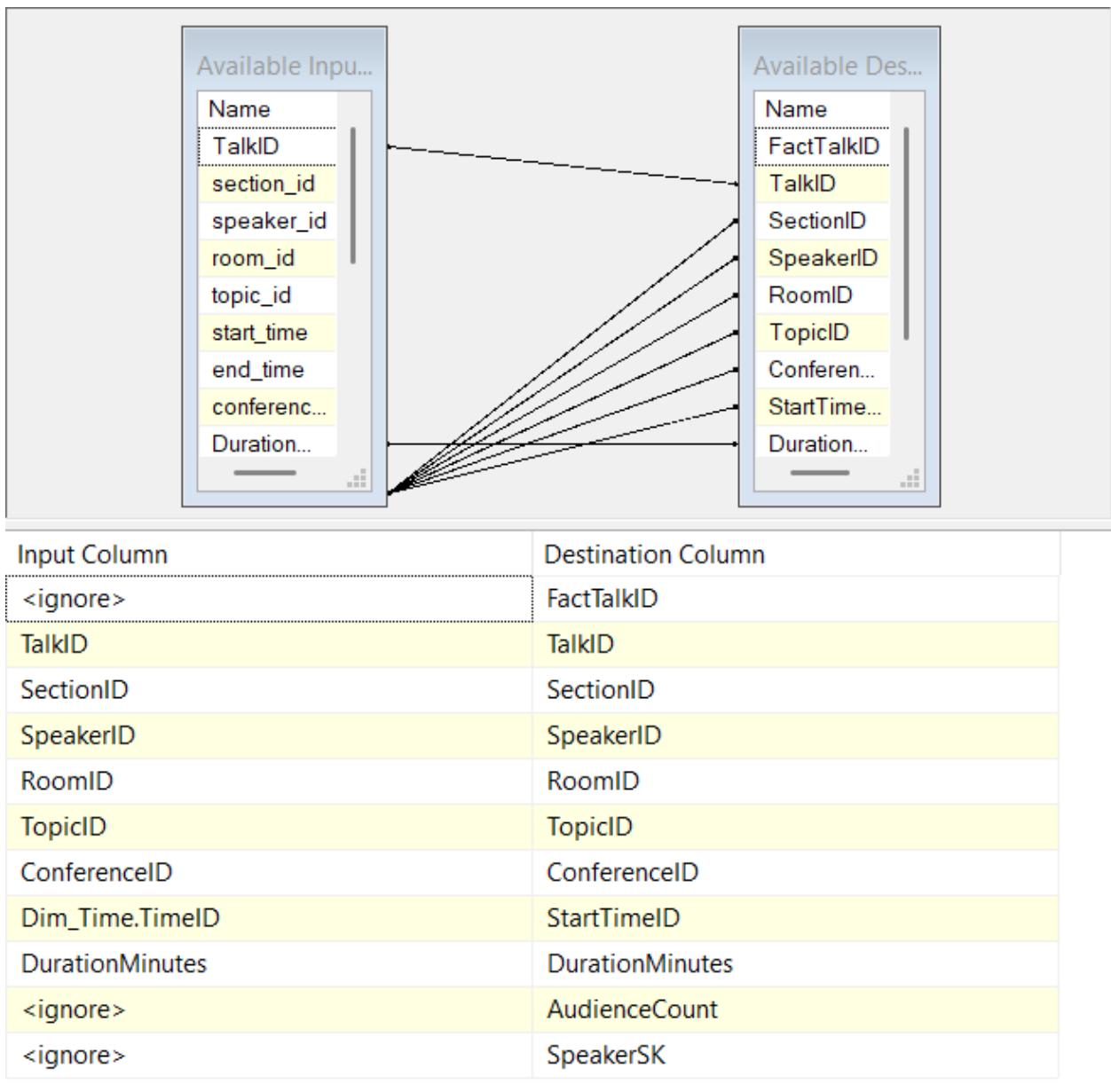
Всередині Fact_Talks Data Flow створено OLE DB Source з sql-скриптом як джерело даних, який посилається на потрібні для таблиці поля:

```

SELECT
    t.id AS TalkID,
    t.section_id,
    t.speaker_id,
    t.room_id,
    t.topic_id,
    t.start_time,
    t.end_time,
    s.conference_id
FROM talks t
JOIN sections s ON t.section_id = s.id

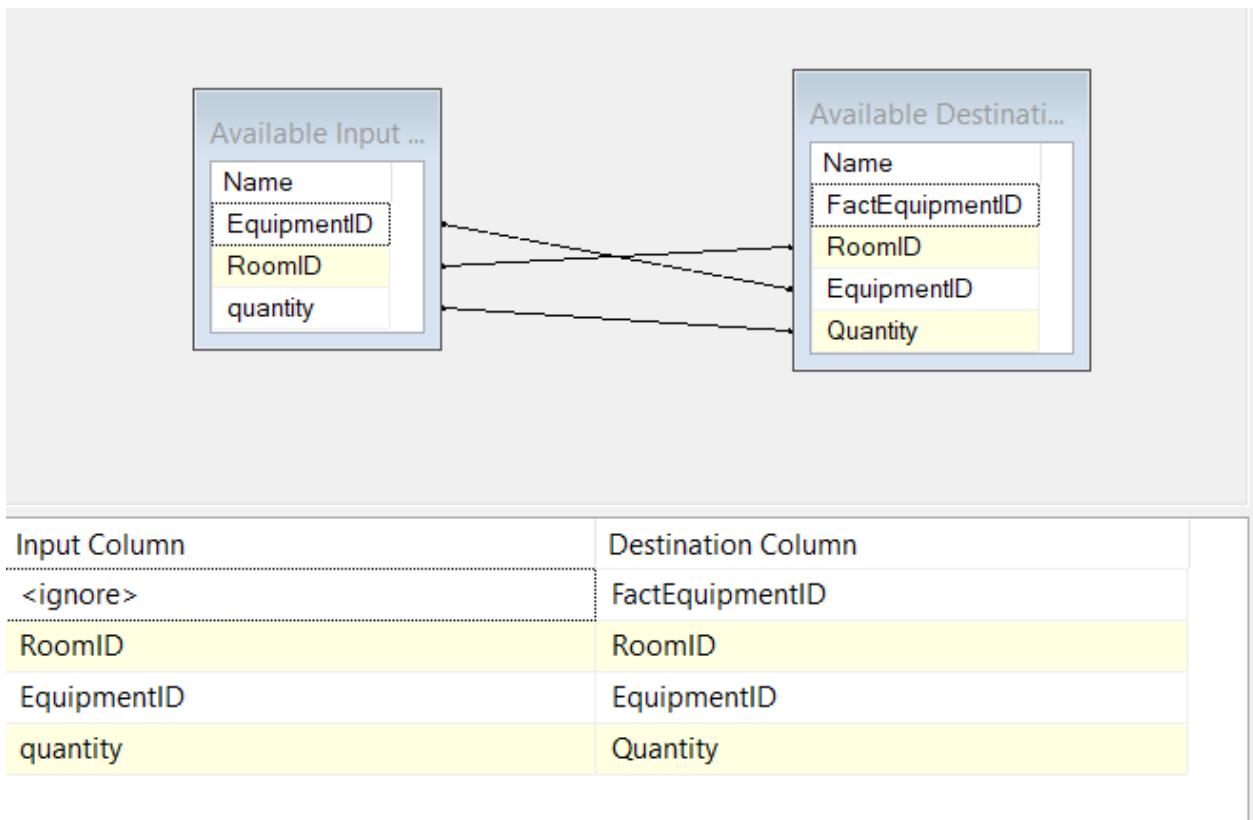
```

Пізніше було створено Lookup блоки. З їх допомогою виконано підстановку ключів вимірів із відповідних таблиць. Зв'язки, що вийшли в результаті:

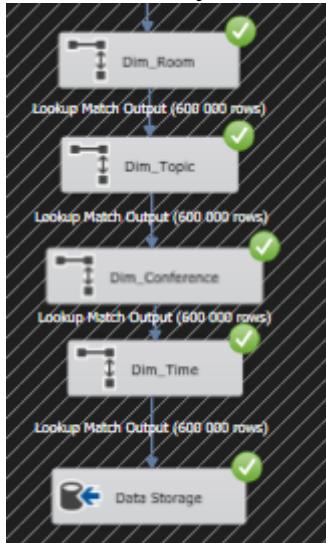


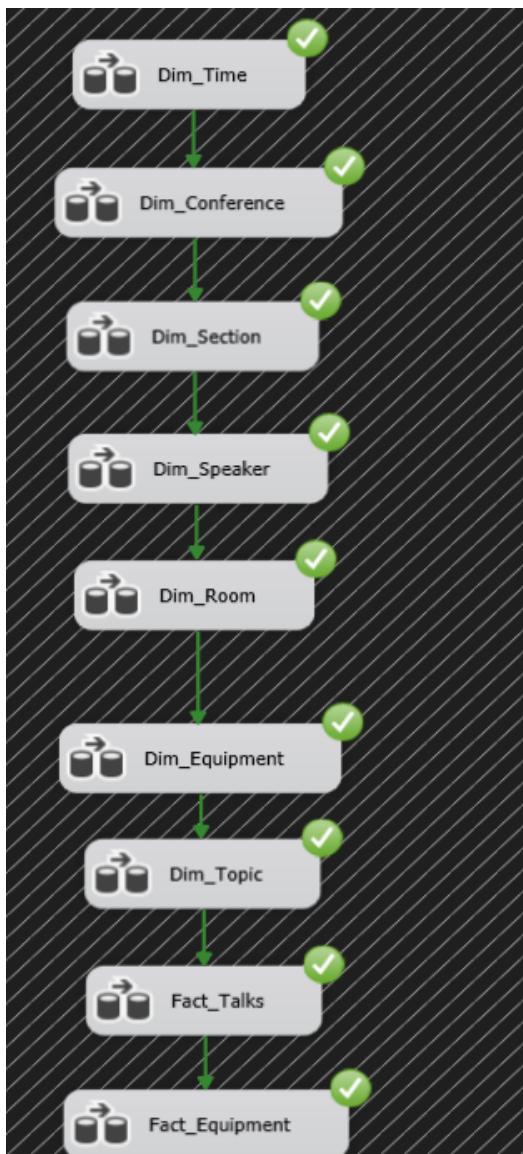
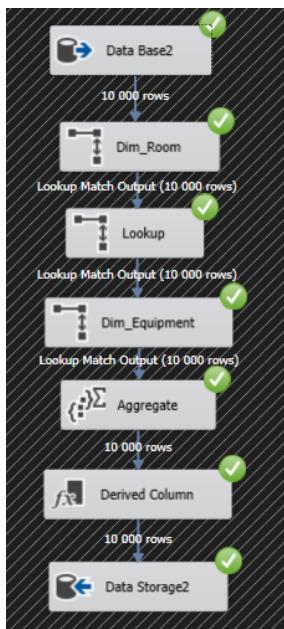
Всередині Fact_Equipment Data Flow також створено OLE DB Source з sql-скриптом як джерело даних, який посилається на таблицю room_equipment.

Пізніше було створено Lookup блоки. З їх допомогою виконано підстановку ключів вимірів із відповідних таблиць. Зв'язки, що вийшли в результаті:



Пізніше запускаємо проект і бачимо що він успішно виконався:

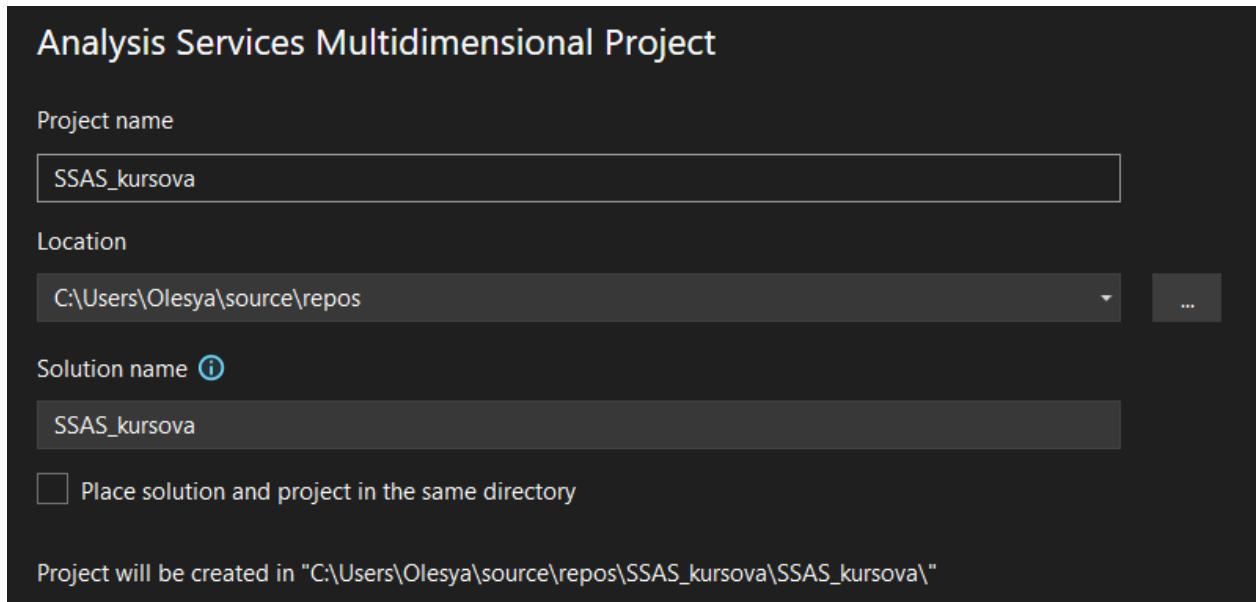




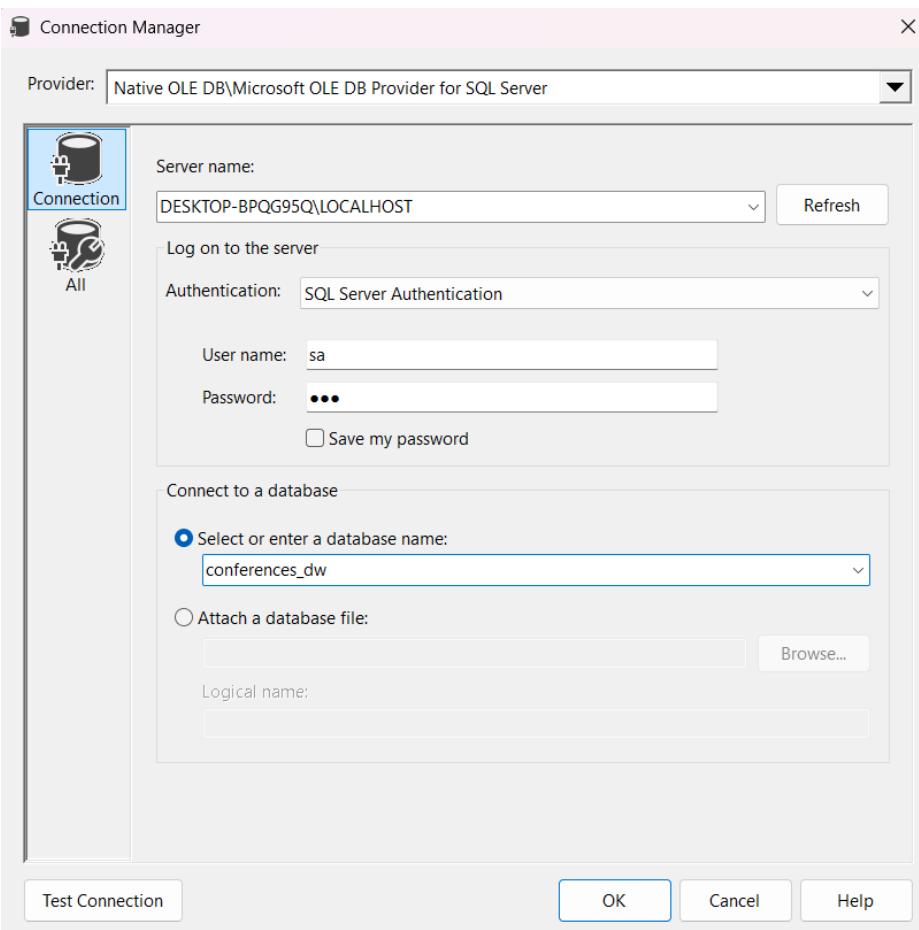
Дані записані в таблиці.

Побудова OLAP-куба

Спочатку створено Analysis Services Multidimensional проект в середовищі Visual Studio, який названо SSAS_kursova:

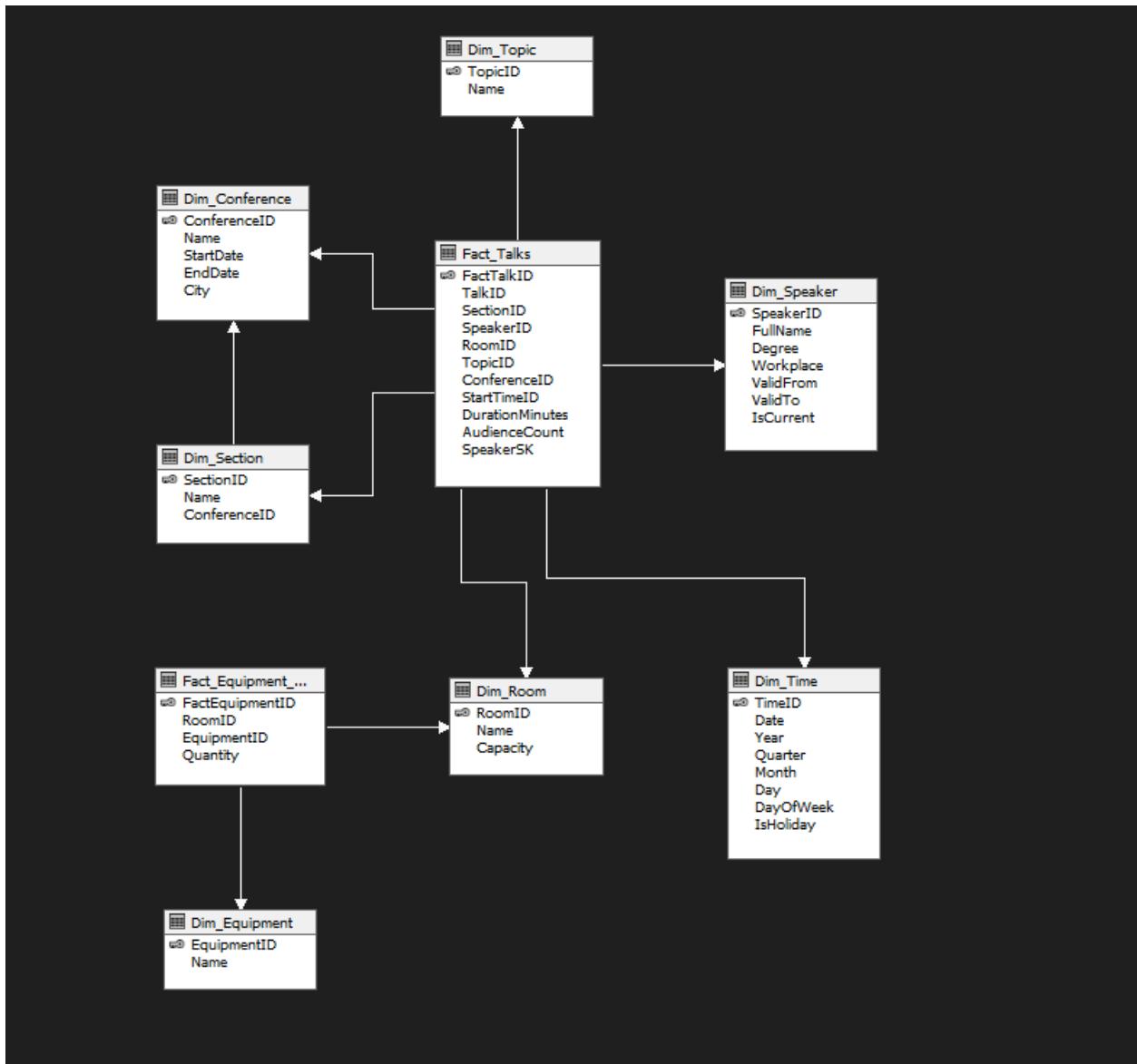


В Data Sources під'єднаємо сховище даних. Обираємо New Data Source. Вказуємо наш SQL-server, ім'я сховища та вводимо логін та пароль користувача:



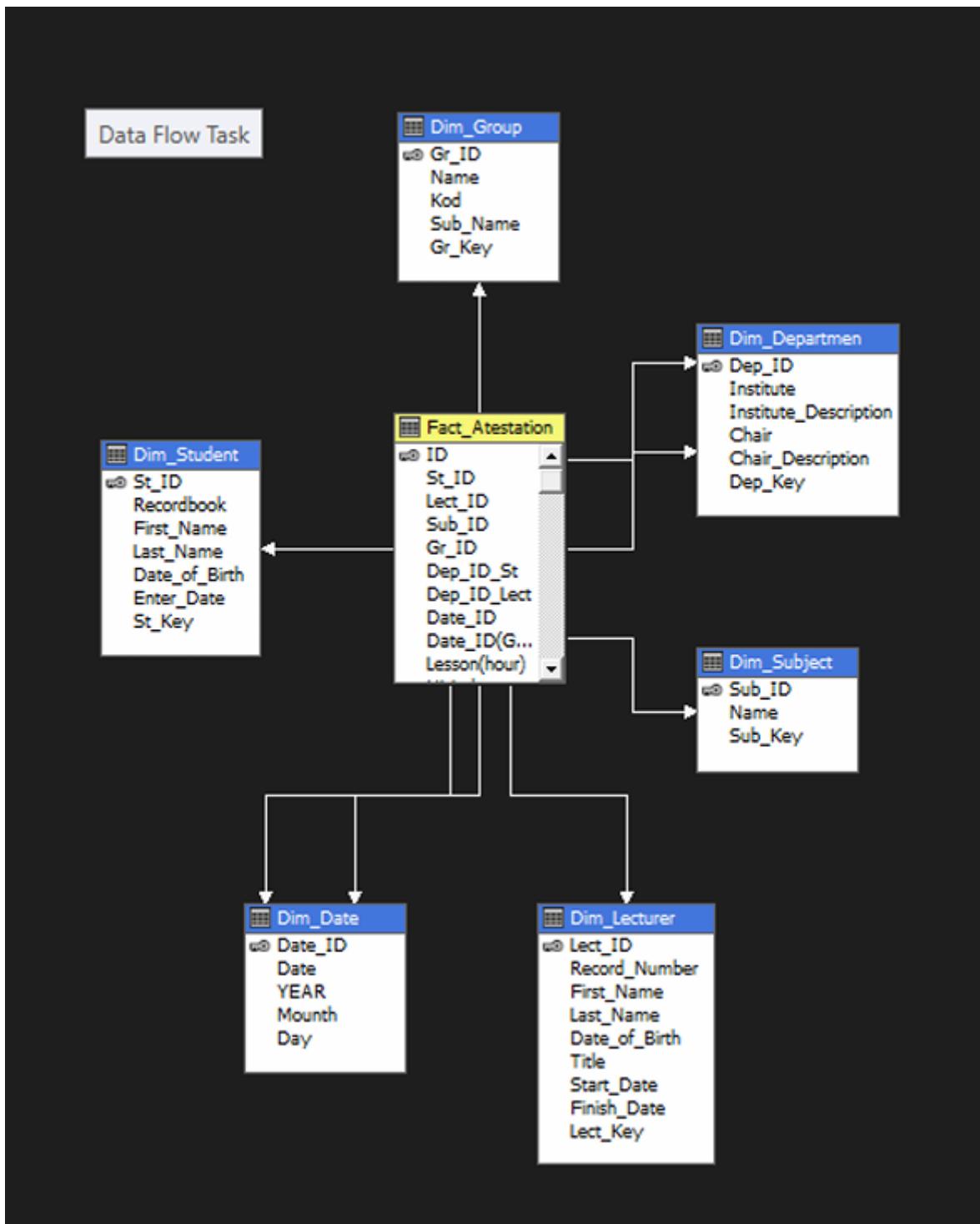
У вкладці Data Views обираємо New Data View та додаємо всі наші таблиці. Після чого, у вкладці Dimensions, обираємо New Dimension і додаємо таблиці вимірів.

Структуру сховища даних:

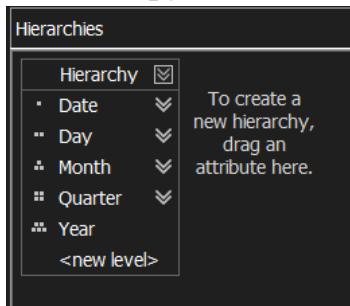


Після цього створюємо OLAP-куб у середовищі SQL Server Analysis Services (SSAS). Оибраємо опцію New Cube, після чого в налаштуваннях вибираємо “Використати існуючі таблиці”. Ставимо галочку біля таблиць фактів, що слугуватимуть групами мір (Measures) у кубі.

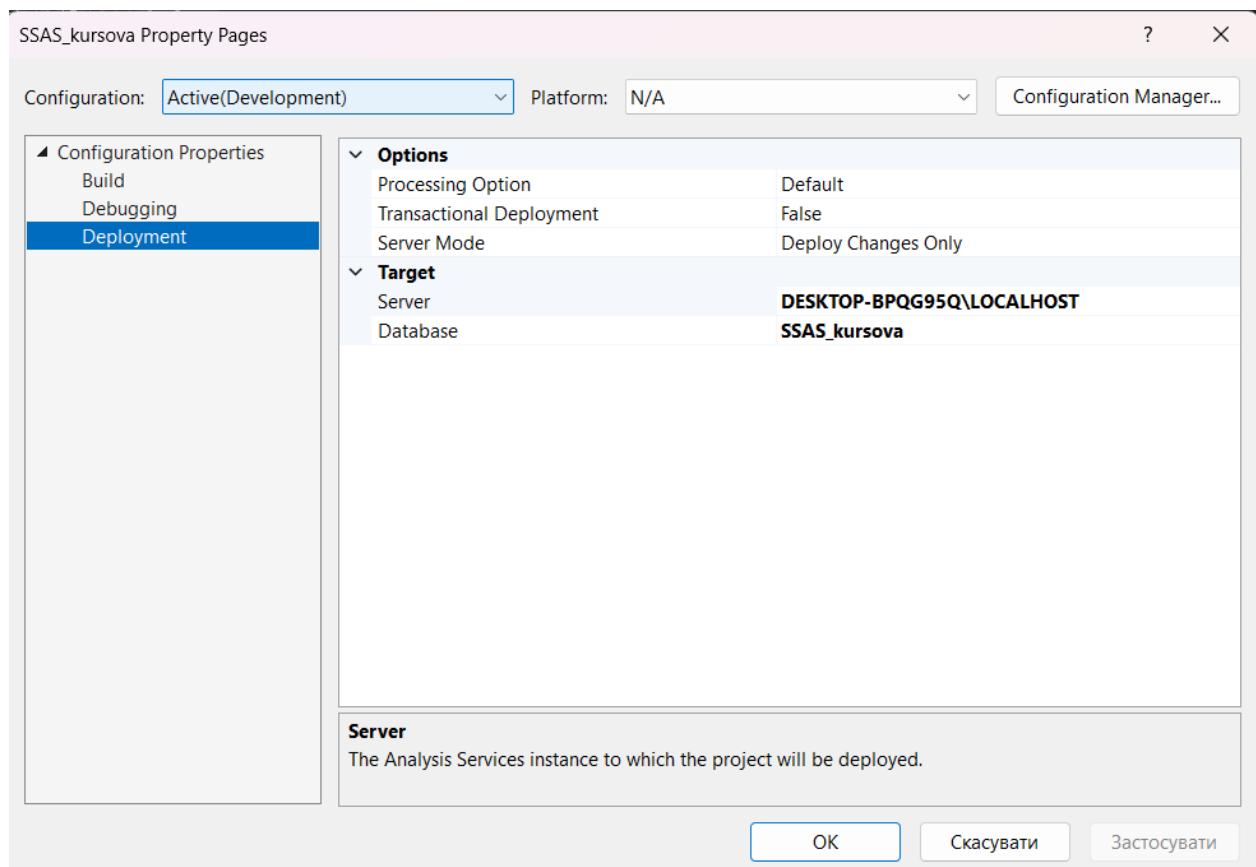
Створений куб:



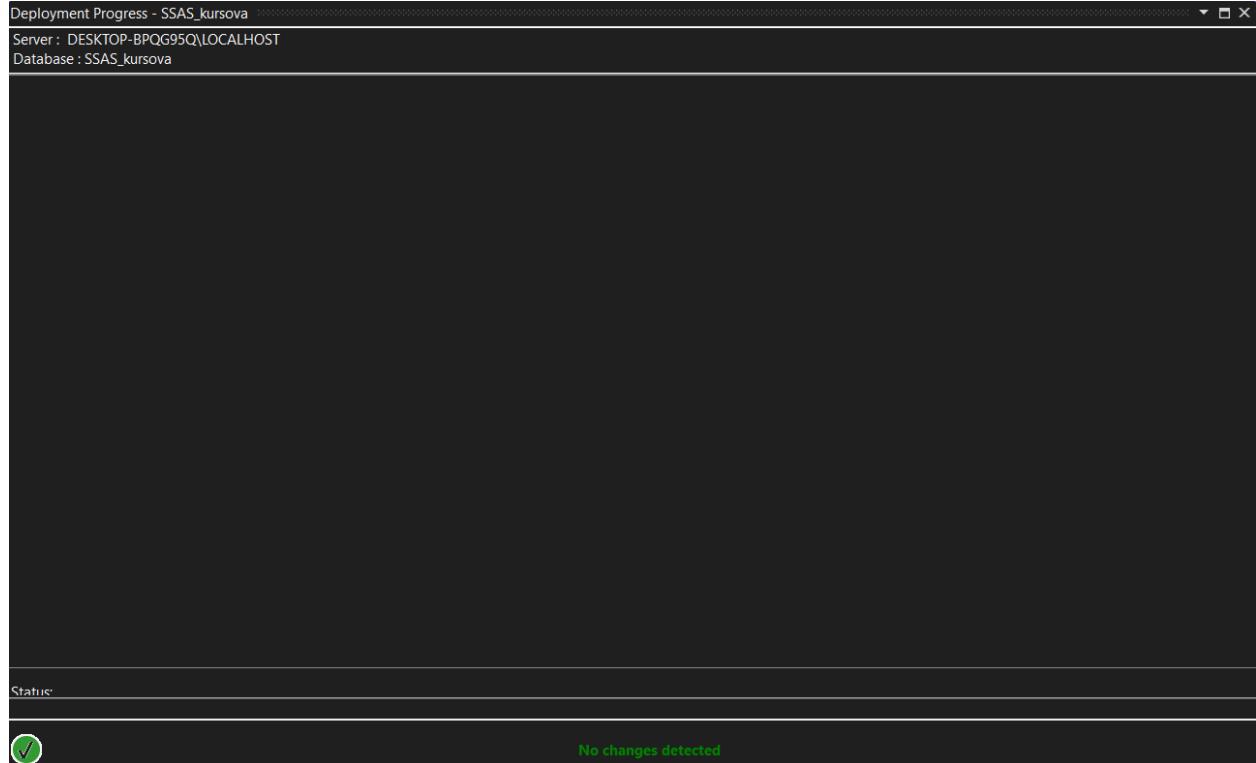
Для виміру Dim_Date налаштовуємо ієархію:



Тепер куб можна деплоїти, але перед цим нам треба вказати шлях деплою:

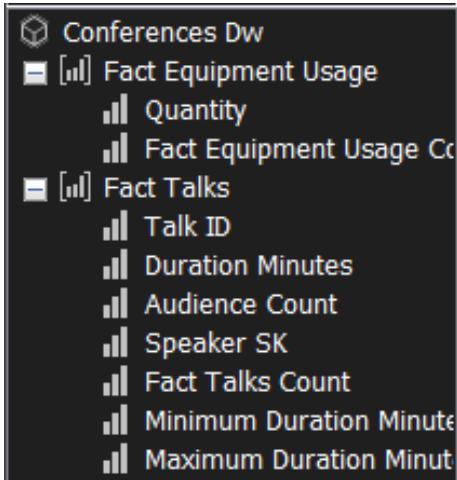


Після цього натискаємо Deploy і чекаємо:



Бачимо що деплой успішний і без помилок.

У вкладці Measures було автоматично створено базові міри і деякі були створені додатково:



Вони будуть потрібні для аналітичних звітів.

Аналітичні звіти:

Для побудови звітів використовуватимемо програму Power BI.

Спочатку підключаємось до куба в режимі реального часу.

SQL Server Analysis Services database

Server ①

DESKTOP-BPQG95Q\LOCALHOST

Database (optional)

SSAS_kursova

Import

Connect live

▷ MDX or DAX query (optional)

OK

Cancel

Після цього можемо приступити до виконання аналітичних звітів.

1) Розклад конференцій:

| Conference ID | Name | Start Date | End Date | City |
|---------------|--|------------|------------|--------------|
| 5518 | ACM International Conference on Web Search and Data Mining | 2025-12-07 | 2025-12-29 | San Antonio |
| 11409 | International Workshop on Machine Learning Applications | 2025-12-07 | 2025-12-27 | Los Angeles |
| 11510 | World Conference on Engineering and Technology | 2025-12-03 | 2025-12-29 | Washington |
| 679 | International Symposium on Distributed Computing | 2025-12-02 | 2025-12-24 | Detroit |
| 251 | European Conference on Computer Science | 2025-11-30 | 2025-12-29 | Kansas |
| 972 | IEEE International Conference on Robotics and Automation | 2025-11-30 | 2025-12-24 | Columbus |
| 2634 | International Symposium on Distributed Computing | 2025-11-29 | 2025-12-20 | Philadelphia |
| 5609 | World Conference on Artificial Intelligence | 2025-11-28 | 2025-12-24 | Seattle |
| 12118 | ACM International Conference on Web Search and Data Mining | 2025-11-28 | 2025-12-21 | San Antonio |
| 8349 | World Congress on Information Technology | 2025-11-27 | 2025-12-24 | New Orleans |
| 11028 | European Conference on Software Architecture | 2025-11-27 | 2025-12-25 | Denver |
| 12879 | International Conference on Applied Mathematics | 2025-11-27 | 2025-12-27 | St. Louis |
| 1167 | International Symposium on Distributed Computing | 2025-11-25 | 2025-12-16 | Philadelphia |
| 3891 | International Symposium on Distributed Computing | 2025-11-25 | 2025-12-17 | San Diego |
| 12338 | International Conference on Software Engineering | 2025-11-25 | 2025-12-28 | Minneapolis |
| 12678 | ACM Conference on Human Factors in Computing Systems | 2025-11-25 | 2025-12-18 | San Antonio |
| 13521 | International Conference on Cyber Security | 2025-11-25 | 2025-12-26 | Albuquerque |
| 7767 | International Conference on Software Engineering | 2025-11-24 | 2025-12-27 | Minneapolis |
| 9421 | International Conference on Information Systems | 2025-11-24 | 2025-12-26 | Pittsburgh |
| 12266 | International Symposium on Theoretical Computer Science | 2025-11-24 | 2025-12-15 | Houston |
| 4071 | International Conference on Computer Science | 2025-11-23 | 2025-12-28 | Toledo |
| 11126 | World Conference on Artificial Intelligence | 2025-11-23 | 2025-12-19 | El Paso |
| 14610 | ACM Conference on Human Factors in Computing Systems | 2025-11-23 | 2025-12-16 | Baltimore |
| 12128 | International Conference on Information Systems | 2025-11-22 | 2025-12-24 | Pittsburgh |
| 73 | International Conference on Artificial Intelligence | 2025-11-21 | 2025-12-24 | Honolulu |
| 4833 | ACM Symposium on Operating Systems Principles | 2025-11-21 | 2025-12-13 | Detroit |
| 6540 | ACM International Conference on Web Search and Data Mining | 2025-11-21 | 2025-12-29 | Louisville |
| 11806 | European Conference on Computer Science | 2025-11-21 | 2025-12-20 | Kansas |
| 13684 | ACM Conference on Human Factors in Computing Systems | 2025-11-21 | 2025-12-14 | San Jose |
| 14981 | European Conference on Data Engineering | 2025-11-21 | 2025-12-20 | Fort Worth |
| 3943 | ACM Conference on Human Factors in Computing Systems | 2025-11-20 | 2025-12-13 | San Jose |
| 7048 | IEEE International Conference on Computer Vision | 2025-11-20 | 2025-12-15 | Memphis |

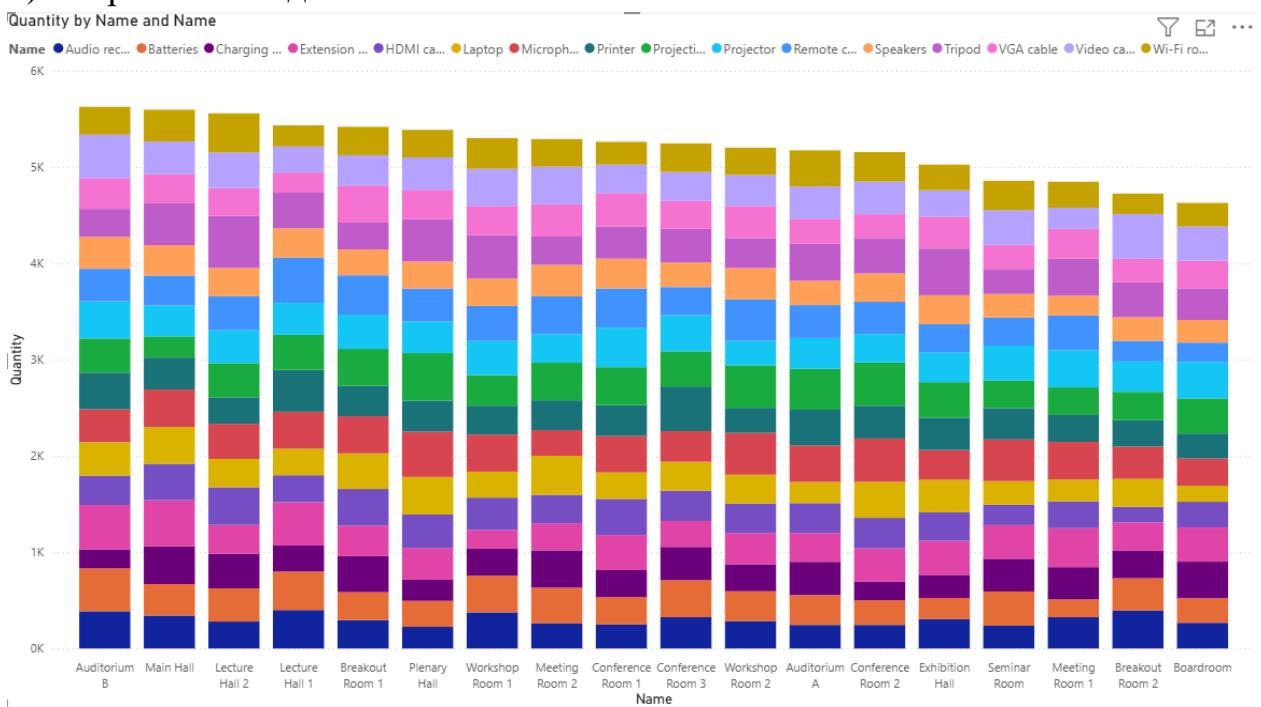
Цей звіт містить ID конференції, дату початку та кінця конференції, а також місто проведення. Дані посортовані за датою початку конференцій.

2) Список учасників з місцем праці:

| Speaker ID | Full Name | Conference ID | Name | Date | Workplace |
|------------|--------------------|---------------|---|------------|---|
| 1 | Jeanine Walton | 727 | International Workshop on Cloud and Edge Computing | 2024-02-07 | Harvard University |
| 2 | Claire Farrell | 818 | World Conference on Data Science | 2022-11-29 | Massachusetts Institute of Technology (MIT) |
| 3 | Valerie Huynh | 769 | ACM Symposium on Operating Systems Principles | 2025-07-29 | Technical University of Munich |
| 4 | Roy Massey | 559 | World Conference on Engineering and Technology | 2021-12-11 | Massachusetts Institute of Technology (MIT) |
| 5 | Forrest Baxter | 207 | European Conference on Software Architecture | 2024-07-12 | University of Oxford |
| 6 | Gerard Drake | 1004 | International Conference on Artificial Intelligence | 2023-09-13 | Carnegie Mellon University |
| 7 | Gilberto Woodard | 907 | World Conference on Data Science | 2017-04-27 | National University of Singapore (NUS) |
| 8 | Roberta Carroll | 443 | ACM International Conference on Knowledge Discovery and Data Mining | 2016-09-29 | University of California, Berkeley |
| 9 | Randal Anderson | 978 | International Conference on Cyber Security | 2021-01-12 | Google Research |
| 10 | Max Peters | 274 | International Conference on Artificial Intelligence | 2020-08-13 | Tsinghua University |
| 11 | Roland Matthews | 292 | World Conference on Artificial Intelligence | 2017-05-14 | University of California, Berkeley |
| 12 | Lee Sanchez | 468 | ACM Conference on Human Factors in Computing Systems | 2021-01-13 | Stanford University |
| 13 | Roderick Little | 633 | International Symposium on Theoretical Computer Science | 2017-05-01 | Peking University |
| 14 | Ty Mosley | 470 | European Conference on Artificial Intelligence | 2016-03-09 | Seoul National University |
| 15 | Donna Norris | 983 | IEEE International Conference on Computer Vision | 2018-08-19 | Amazon Web Services (AWS) |
| 16 | Yvonne Reeves | 31 | World Conference on Artificial Intelligence | 2018-03-07 | Imperial College London |
| 17 | Bridgette Hardy | 863 | International Conference on Big Data Analytics | 2023-12-08 | Tsinghua University |
| 18 | Daphne Woodard | 996 | IEEE International Conference on Communications | 2020-12-10 | ETH Zurich |
| 19 | Damian Horton | 678 | International Conference on Machine Learning | 2017-01-26 | Tsinghua University |
| 20 | Andrea Eaton | 315 | World Congress on Information Technology | 2020-11-30 | Facebook AI Research (FAIR) |
| 21 | Lewis Rowe | 817 | European Conference on Artificial Intelligence | 2021-10-06 | Carnegie Mellon University |
| 22 | Dwayne Arias | 849 | International Conference on Big Data Analytics | 2020-05-24 | Seoul National University |
| 23 | Glen Greer | 992 | International Conference on Big Data Analytics | 2022-05-30 | Microsoft Research |
| 24 | Garrett Preston | 33 | IEEE International Conference on Communications | 2020-05-29 | Seoul National University |
| 25 | Marco Rogers | 700 | International Conference on Information Systems | 2016-06-20 | Google Research |
| 26 | Kate Ellison | 527 | European Conference on Data Engineering | 2015-08-28 | University of California, Berkeley |
| 27 | Anna Frost | 935 | World Conference on Artificial Intelligence | 2020-09-30 | Carnegie Mellon University |
| 28 | Christine Cisneros | 688 | International Conference on Machine Learning | 2018-02-25 | Amazon Web Services (AWS) |
| 29 | Trent Weber | 547 | International Workshop on Machine Learning Applications | 2025-06-04 | University of California, Berkeley |
| 30 | Alisha Key | 82 | International Conference on Applied Mathematics | 2014-12-23 | Amazon Web Services (AWS) |
| 31 | Marisol Villanueva | 188 | International Symposium on Distributed Computing | 2022-07-17 | Imperial College London |
| 32 | Elisabeth Duncan | 454 | European Conference on Computer Science | 2017-03-18 | IBM Research |

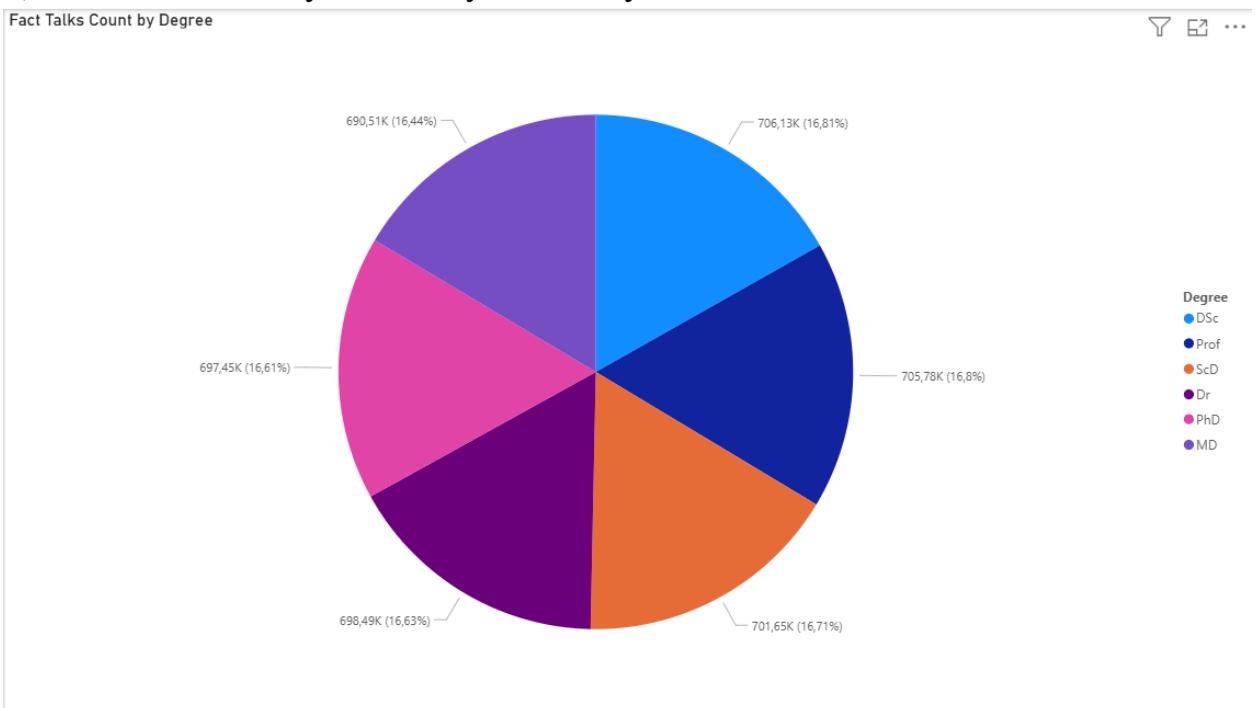
Цей звіт містить ID учасника, ім'я та прізвище, місце роботи. Також місить ID, назву та дату конференції, в якій вони беруть участь.

3) Потреба в обладнанні:



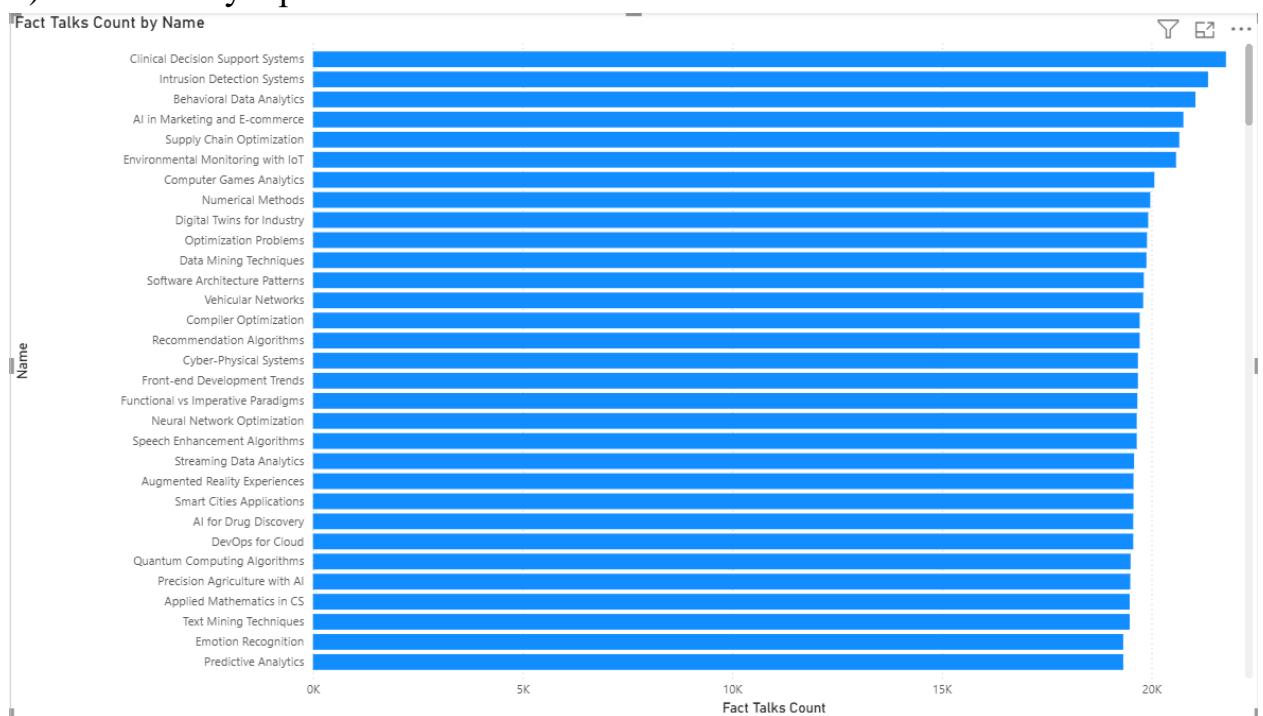
Цей звіт показує потребу в обладнанні під час виступів у приміщеннях протягом усіх конференцій. А також містить назви приміщень та назви обладнань.

4) Статистика виступів по науковим ступеням:



Цей звіт показує статистику виступів по науковим ступеням виступаючих. На цьому графіку ми можемо побачити назви наукових ступенів та кількість виступів для кожного.

5) Аналіз популярності тематик:



Цей звіт показує популярність тематик у виступах. На цьому графіку ми можемо побачити назви тем та кількість виступів з нею. Графік

посортований за популярністю тем, а саме від найпопулярніших до менш популярних.

Висновок:

У ході виконання курсової роботи було досліджено сучасні підходи до зберігання, обробки та аналізу надвеликих обсягів даних. Розглянуто основні концепції побудови сховищ даних, багатовимірного моделювання та аналітичної обробки інформації з використанням OLAP-технологій.

У рамках практичної частини роботи розроблено сховище даних для предметної області проведення наукових конференцій. Виконано проектування структури бази даних із використанням багатовимірної моделі, що забезпечує ефективний аналіз інформації за різними вимірами, зокрема конференціями, секціями, виступаючими, тематикою та часом.

Реалізовано ETL-процеси з використанням технологій SQL Server Integration Services, які забезпечують витягування, очищення, трансформацію та завантаження даних у сховище даних. У процесі розробки ETL було застосовано різні типи трансформацій, що дозволило підвищити якість та цілісність даних.

На основі сховища даних побудовано багатовимірний OLAP-куб за допомогою SQL Server Analysis Services. У кубі реалізовано виміри, міри, обчислювані показники, часову ієархію та аналітичні обчислення, що дозволяють виконувати багатовимірний аналіз даних, включаючи аналіз динаміки показників, рейтинги та агреговані статистичні дані.

З використанням SQL Server Reporting Services розроблено комплекс аналітичних звітів різних типів.

Отримані результати підтверджують доцільність використання платформи Microsoft SQL Server для побудови інформаційно-аналітичних систем та аналізу надвеликих обсягів даних.

Посилання на репозиторій:

https://github.com/Olesia-Karpyn/Kursova_NVBD_Variant_08_Karpyn_Olesia.git

Використані джерела:

https://e-tk.lntu.edu.ua/pluginfile.php/20333/mod_resource/content/0/%D0%A2%D0%B5%D0%BC%D0%B0%2016.%20%D0%A2%D0%B5%D1%85%D0%BD%D0%BE%D0%BB%D0%BE%D0%B3%D1%96%D1%96%CC%88%20%D0%BE%D0%B1%D1%80%D0%BE%D0%B1%D0%BA%D0%B8%20%D0%B2%D0%BD%D0%BB%D0%BA%D0%B8%D1%85%20%D0%B4%D0%B0%D0%BD%D0%BB%D0%BA%D1%85%20%28Big%20Data%29.pdf

Додатки:

SQL-скрипт для бази даних conferences_db:

USE conferences_db;

GO

```
CREATE TABLE conferences (
```

```
    id INT PRIMARY KEY,  
    name NVARCHAR(255) NOT NULL,  
    start_date DATE NOT NULL,  
    end_date DATE NOT NULL,  
    city NVARCHAR(100) NOT NULL
```

```
);
```

```
CREATE TABLE sections (
```

```
    id INT PRIMARY KEY,  
    conference_id INT NOT NULL,  
    name NVARCHAR(255) NOT NULL,  
    CONSTRAINT fk_sections_conference  
        FOREIGN KEY (conference_id)  
        REFERENCES conferences(id)  
        ON DELETE CASCADE
```

```
);
```

```
CREATE TABLE topics (
```

```
    id INT PRIMARY KEY,  
    name NVARCHAR(100) NOT NULL
```

```
);
```

```
CREATE TABLE speakers (
    id INT PRIMARY KEY,
    full_name NVARCHAR(255) NOT NULL,
    degree NVARCHAR(50) NOT NULL,
    workplace NVARCHAR(255) NOT NULL
);
```

```
CREATE TABLE rooms (
    id INT PRIMARY KEY,
    name NVARCHAR(100) NOT NULL,
    capacity INT NOT NULL
);
```

```
CREATE TABLE equipment (
    id INT PRIMARY KEY,
    name NVARCHAR(100) NOT NULL
);
```

```
CREATE TABLE room_equipment (
    room_id INT NOT NULL,
    equipment_id INT NOT NULL,
    quantity INT NULL,
    CONSTRAINT pk_room_equipment PRIMARY KEY (room_id,
    equipment_id),
    CONSTRAINT fk_re_room
        FOREIGN KEY (room_id)
        REFERENCES rooms(id)
        ON DELETE CASCADE,
    CONSTRAINT fk_re_equipment
```

```
    FOREIGN KEY (equipment_id)
    REFERENCES equipment(id)
    ON DELETE CASCADE
);
```

```
CREATE TABLE talks (
    id INT PRIMARY KEY,
    section_id INT NOT NULL,
    speaker_id INT NOT NULL,
    room_id INT NOT NULL,
    topic_id INT NOT NULL,
    title NVARCHAR(255) NOT NULL,
    start_time DATETIME NOT NULL,
    end_time DATETIME NOT NULL,
    CONSTRAINT fk_talks_section
        FOREIGN KEY (section_id)
        REFERENCES sections(id)
        ON DELETE CASCADE,
    CONSTRAINT fk_talks_speaker
        FOREIGN KEY (speaker_id)
        REFERENCES speakers(id)
        ON DELETE CASCADE,
    CONSTRAINT fk_talks_room
        FOREIGN KEY (room_id)
        REFERENCES rooms(id)
        ON DELETE CASCADE,
    CONSTRAINT fk_talks_topic
        FOREIGN KEY (topic_id)
```

```
    REFERENCES topics(id)
    ON DELETE CASCADE
);
```

SQL-скрипт для базы данных conferences_dw:

```
CREATE DATABASE conferences_dw;
```

```
GO
```

```
USE conferences_dw;
```

```
GO
```

```
CREATE TABLE Dim_Conference (
    ConferenceID INT PRIMARY KEY,
    Name NVARCHAR(255) NOT NULL,
    StartDate DATE NOT NULL,
    EndDate DATE NOT NULL,
    City NVARCHAR(100) NOT NULL
);
```

```
CREATE TABLE Dim_Section (
    SectionID INT PRIMARY KEY,
    Name NVARCHAR(255) NOT NULL,
    ConferenceID INT NOT NULL,
    FOREIGN KEY (ConferenceID) REFERENCES
    Dim_Conference(ConferenceID)
);
```

```
CREATE TABLE Dim_Speaker (
    SpeakerID INT PRIMARY KEY,
    FullName NVARCHAR(255) NOT NULL,
```

```
Degree NVARCHAR(50),  
Workplace NVARCHAR(255)  
);
```

```
CREATE TABLE Dim_Room (  
    RoomID INT PRIMARY KEY,  
    Name NVARCHAR(100) NOT NULL,  
    Capacity INT NOT NULL  
);
```

```
CREATE TABLE Dim_Equipment (  
    EquipmentID INT PRIMARY KEY,  
    Name NVARCHAR(100) NOT NULL  
);
```

```
CREATE TABLE Dim_Topic (  
    TopicID INT PRIMARY KEY,  
    Name NVARCHAR(100) NOT NULL  
);
```

```
CREATE TABLE Dim_Time (  
    TimeID INT PRIMARY KEY,  
    [Date] DATE NOT NULL,  
    [Year] INT NOT NULL,  
    Quarter INT NOT NULL,  
    Month INT NOT NULL,  
    Day INT NOT NULL,  
    DayOfWeek INT NOT NULL,
```

```
    IsHoliday BIT NOT NULL DEFAULT 0  
);
```

```
CREATE TABLE Fact_Talks (  
    FactTalkID INT PRIMARY KEY IDENTITY(1,1),  
    TalkID INT NOT NULL,  
    SectionID INT NOT NULL,  
    SpeakerID INT NOT NULL,  
    RoomID INT NOT NULL,  
    TopicID INT NOT NULL,  
    ConferenceID INT NOT NULL,  
    StartTimeID INT NOT NULL,  
    DurationMinutes INT NOT NULL,  
    AudienceCount INT NULL,  
    FOREIGN KEY (SectionID) REFERENCES Dim_Section(SectionID),  
    FOREIGN KEY (SpeakerID) REFERENCES Dim_Speaker(SpeakerID),  
    FOREIGN KEY (RoomID) REFERENCES Dim_Room(RoomID),  
    FOREIGN KEY (TopicID) REFERENCES Dim_Topic(TopicID),  
    FOREIGN KEY (ConferenceID) REFERENCES  
        Dim_Conference(ConferenceID),  
    FOREIGN KEY (StartTimeID) REFERENCES Dim_Time(TimeID)  
);
```