

Міністерство освіти і науки України
Державний університет «Житомирська політехніка»
Факультет інформаційно-комп'ютерних технологій
Кафедра інженерії програмного забезпечення

КУРСОВА РОБОТА
(ПОЯСНЮВАЛЬНА ЗАПИСКА)

з дисципліни: «Моделювання інформаційних процесів та системи обробки великих даних»

на тему:

**«Моделювання інформаційних процесів платформи
для спільного кіноперегляду»**

Виконав магістр 1-го курсу, групи ПЗм-24-1
спеціальності 121 «Інженерія програмного
забезпечення»

Костюченко Артем Валерійович
(прізвище, ім'я та по-батькові)

Керівник: к.т.н., доцент кафедри КН Сугоняк І.І.

Дата захисту: " ____ " _____ 2024 р.
Національна шкала _____
Кількість балів: _____
Оцінка: ECTS _____

Члени комісії

_____	<u>Інна СУГОНЯК</u>
(підпис)	(прізвище та ініціали)
_____	<u>Марина ГРАФ</u>
(підпис)	(прізвище та ініціали)
_____	_____
(підпис)	(прізвище та ініціали)

Житомир – 2024

ЗМІСТ

ВСТУП.....	3
РОЗДІЛ 1. МОДЕЛЬ ІНФОРМАЦІЙНИХ ПРОЦЕСІВ ПЛАТФОРМИ ДЛЯ СПІЛЬНОГО КІНОПЕРЕГЛЯДУ	4
1.1 Постановка задачі	4
1.2 Інформаційні процеси, обґрунтування вибору засобів моделювання та програмного середовища	4
1.3 Побудова моделей бізнес-процесів та потоків даних	7
Висновки до розділу 1	16
РОЗДІЛ 2. ОРГАНІЗАЦІЯ СХОВИЩА ДАНИХ ДЛЯ АНАЛІТИКИ ТА ETL ПРОЦЕСУ	18
2.1 Визначення даних для аналізу та розробка сховища даних	18
2.2 ETL процес та використання Azure Synapse Analytics для копіювання та трансформації даних.....	23
Висновки до розділу 2	33
РОЗДІЛ 3. ОРГАНІЗАЦІЯ ЗАСОБІВ ВІЗУАЛІЗАЦІЇ ТА АНАЛІЗУ ДАНИХ	34
3.1 Технологія Power BI та її основні можливості	34
3.2 Публікація звіту на прикладній дошці Power BI	34
Висновки до розділу 3	41
ВИСНОВКИ.....	42
ПЕРЕЛІК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ	43

					«Житомирська політехніка» 24.121.14.000			
Змн.	Арк.	№ докум.	Підпис	Дата				
Розроб.		Костюченко А.В.			Пояснювальна записка до курсової роботи		Літ.	Арк.
Керівник		Сугоняк І.І.						Аркушів
								2
								50
Реценз.					ФІКТ, гр. ІПЗм-24-1			
Затверд.		Граф М. С.						

ВСТУП

У сучасному світі з стрімким розвитком інформаційних технологій, де все більше людей потребують різних інтерактивних взаємодій під час перегляду різного відеоконтенту, веб-платформи набувають особливого значення в суспільстві. Інтерактивна взаємодія під час спільного перегляду, обмін враженнями та спілкування в режимі реального часу є ключовими характеристиками платформ, що забезпечують ці можливості незалежно від географічного розташування.

Актуальність теми полягає у вдосконаленні існуючих веб-платформ, які не надають достатньо функцій для задоволення потреб користувачів, шляхом впровадження більш інтерактивних можливостей та цікавих форматів взаємодії під час перегляду.

Метою курсової роботи є дослідження та розробка моделі інформаційних процесів платформи для спільного кіноперегляду, проведення аналізу ключових показників її функціонування та візуалізація отриманих результатів.

Поставлена мета передбачає вирішення таких завдань:

- Проведення аналізу вимог до платформи;
- Побудова діаграм IDEF0 та DFD;
- Моделювання діаграм бізнес-процесів;
- Побудова схем OLTP та OLAP;
- Створення сховища даних, його подальший аналіз та візуалізація;

Об'єктом дослідження є інформаційні процеси, необхідні для забезпечення функціонування платформи.

Предметом дослідження є процес моделювання системи із застосуванням діаграм IDEF0 та DFD, побудови діаграм BPMN, використання технологій OLTP та OLAP для проектування схеми, а також використання сучасних інструментів аналізу, таких як Power BI.

		Костюченко А.В.			«Житомирська політехніка» 24.121.14.000	Арк.
						3
Змн.	Арк.	№ докум.	Підпис	Дата		

РОЗДІЛ 1. МОДЕЛЬ ІНФОРМАЦІЙНИХ ПРОЦЕСІВ ПЛАТФОРМИ ДЛЯ СПІЛЬНОГО КІНОПЕРЕГЛЯДУ

1.1 Постановка задачі

Головною метою розробки платформи для спільного кіноперегляду є забезпечення ефективного аналізу та організації процесів взаємодії користувачів. Основним завданням є розробка функціональної платформи, що забезпечуватиме можливості для спільного перегляду відеоконтенту, обміну повідомленнями та впровадження інтерактивних функцій.

Основними етапами реалізації розробки та аналізу інформаційних процесів платформи для спільного кіноперегляду є:

1. Моделювання інформаційних процесів, які необхідні для функціонування платформи.
 - 1.1. Використання нотацій IDEF0 та DFD для моделювання функціональної структури системи.
 - 1.2. Побудова BPMN-діаграми, яка відображає основні процеси функціонування системи.
2. Проектування сховища даних та його візуалізація за допомогою OLTP і OLAP-систем.
 - 2.1. Формування даних таблиць у форматі CSV для створення бази даних.
 - 2.2. Завантаження даних у форматі CSV в Azure Synapse Analytics, створення бази даних та таблиць.
 - 2.3. Створення запитів для аналізу та обробки даних у сховищі.
3. Аналіз і візуалізація інформації з використанням функціоналу Power BI.

1.2 Інформаційні процеси, обґрунтування вибору засобів моделювання та програмного середовища

Для розробки функціональних моделей доцільно застосовувати нотації DFD і IDEF0, що дозволяють чітко формалізувати інформаційні потоки та процеси.

		Костюченко А.В.			«Житомирська політехніка» 24.121.14.000	Арк.
						4
Змн.	Арк.	№ докум.	Підпис	Дата		

Data Flow Diagram (DFD) або діаграма потоків даних відображає потік даних в інформаційних системах. Діаграма забезпечує графічне представлення потоку даних в системі, яке може бути зрозумілим як технічним, так і нетехнічним користувачам. Модель дозволяє інженерам-програмістам, замовникам і користувачам ефективно співпрацювати під час аналізу та специфікації вимог.

Головною метою діаграми потоків даних (DFD) є візуалізація руху інформації в системі для полегшення аналізу її функціональних компонентів і виявлення можливих покращень у процесах та взаємодії [1].

IDEF0 (ICAM DEFINITION for Function Modeling, де «ICAM» – аббревіатура від Integrated Computer Aided Manufacturing) є стандартом для моделювання функцій, який був розроблений у рамках проекту ICAM, спрямованого на підвищення ефективності виробничих процесів. Представляє собою інтегровану картину входів, управління, виходів і механізмів (ICOM) для декомпозиції функції з метою детального аналізу. Методологія IDEF0 є частиною сімейства методів IDEF, які охоплюють різні аспекти проектування й оптимізації систем. Сімейство IDEF створювалось для стандартизації та уніфікації підходів до моделювання в промислових і корпоративних системах. Якщо коротко, то це інструмент для візуалізації, аналізу й оптимізації функціональних аспектів системи [2].

Головною метою діаграми IDEF0 є надання повного, точного та адекватного опису системи, що дозволяє отримувати відповіді на конкретні питання щодо її функціонування з необхідною точністю.

Для побудови діаграм DFD та IDEF0 було обрано середовище моделювання Ramus.

Ramus – це програма для моделювання, яка дозволяє візуалізувати процеси, їхній взаємозв'язок та потоки даних у системах, забезпечуючи зручний інтерфейс для роботи з різними рівнями декомпозиції. До переваг можна віднести підтримку функціональних діаграм та структурованого підходу до аналізу і проектування систем. До недоліків можна віднести, не зручний інтерфейс, який без інструкції важко розібратися. Серед недоліків варто відзначити недостатньо інтуїтивний інтерфейс,

		Костюченко А.В.			«Житомирська політехніка» 24.121.14.000	Арк.
						5
Змн.	Арк.	№ докум.	Підпис	Дата		

що може ускладнювати його освоєння без попереднього ознайомлення з інструкціями.

Для опису бізнес-процесів платформи було використано нотацію BPMN 2.0.

Нотація моделювання бізнес-процесів (BPMN) — це стандартний спосіб візуалізації бізнес-процесів за допомогою зрозумілих графічних схем. Вона використовується для опису робочих процесів підприємств і допомагає всім учасникам бізнесу — від звичайних користувачів до розробників і аналітиків — легко зрозуміти, як працюють ці процеси. BPMN є зручною та поширеною мовою для аналізу та вдосконалення бізнесу [3].

Для побудови схеми бази даних доцільно застосовувати OLTP та OLAP підходи.

Онлайн-обробка транзакцій (OLTP) — це спосіб обробки даних у реальному часі, коли виконуються численні короткі операції, наприклад, внесення платежів або бронювання квитків. OLTP-системи забезпечують надійність і точність даних завдяки дотриманню правил ACID: кожна операція виконується повністю, дані завжди залишаються узгодженими, операції працюють незалежно одна від одної, а результати зберігаються навіть у разі збою [4].

Онлайн-аналітична обробка даних (OLAP) — це програми, які допомагають аналізувати дані, щоб приймати правильні бізнес-рішення. OLAP-системи дозволяють дивитися на дані з різних сторін, зазвичай у вигляді багатовимірних таблиць. Вони використовуються в сховищах даних і бізнес-аналітиці, щоб аналізувати тренди, прогнозувати фінанси та виконувати детальний аналіз інформації [5].

Основна відмінність OLAP від OLTP полягає в їхньому призначенні та використанні. OLAP призначена для комплексного аналізу даних і прийняття рішень шляхом обробки великих обсягів історичних даних, тоді як OLTP - для швидкого та ефективного управління щоденними транзакціями [6].

Для побудови та аналізу сховища даних можна використати сервіси Azure, які надають інструменти для інтеграції, управління та візуалізації даних.

		Костюченко А.В.			«Житомирська політехніка» 24.121.14.000	Арк.
						6
Змн.	Арк.	№ докум.	Підпис	Дата		

Azure Synapse Analytics — це хмарна платформа від Microsoft для роботи з великими даними, яка об’єднує всі необхідні інструменти для зберігання, обробки та аналізу даних в одному місці.

Azure Synapse є продовженням Azure SQL Data Warehouse, додаючи нові можливості, як аналіз великих даних, інтеграцію з Data Lake та використання серверних або серверлесс-ресурсів для запитів. Сервіс також має вбудовані інструменти, такі як ETL для підготовки даних, Power BI для візуалізації та Azure Machine Learning для прогнозів [7].

1.3 Побудова моделей бізнес-процесів та потоків даних

Для побудови діаграм DFD та IDEF0 було використано програму Ramus. Для DFD діаграми (рис 1.1.) було створено основний блок “Платформа спільного кіноперегляду”, додано зовнішні ресурси “Соціальні мережі”, “Постачальники контенту”, “Користувачі” та сховища даних “Список кімнат”, “Список користувачів”, “Список повідомлень”.

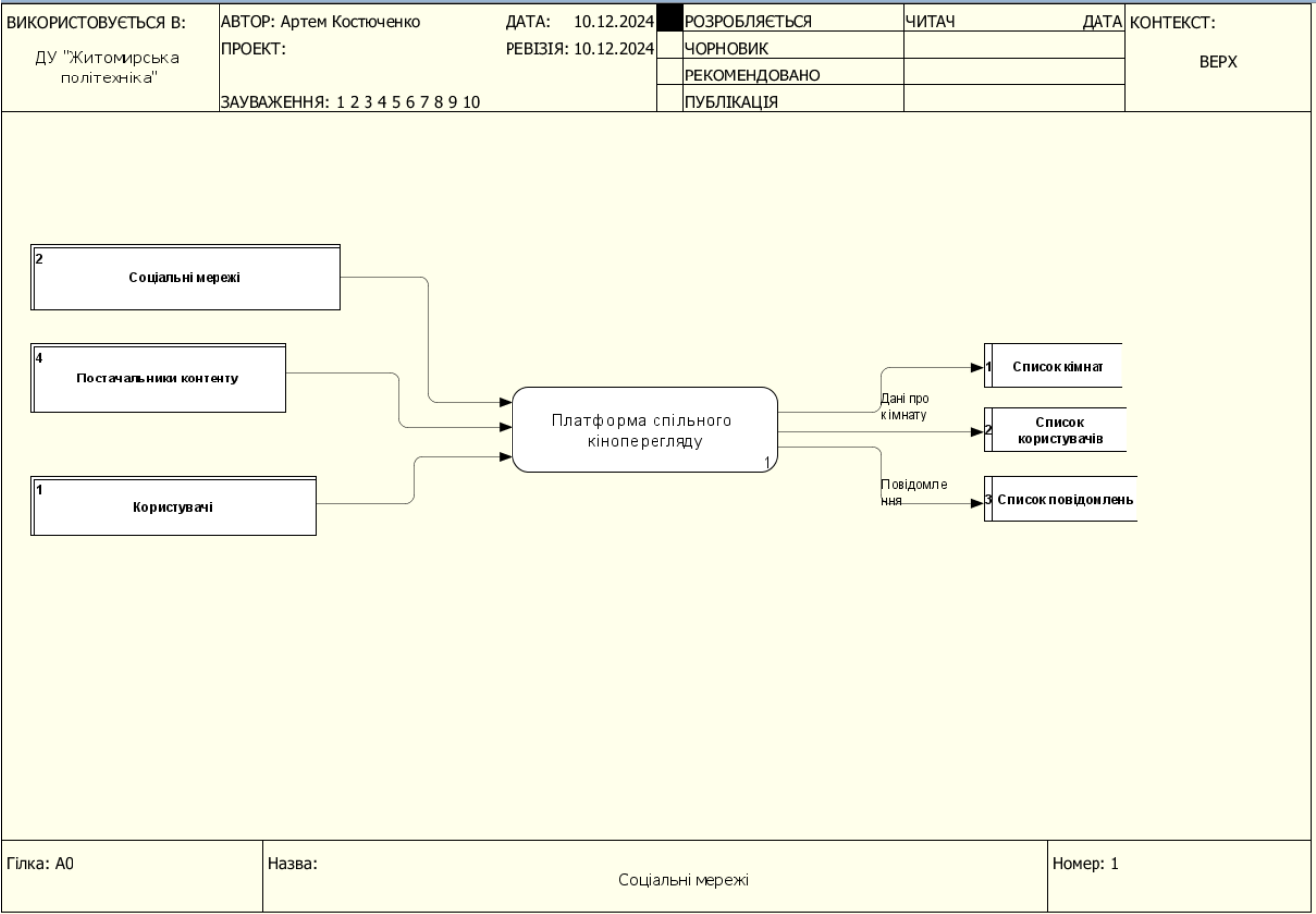


Рисунок 1.1 – Контекстна діаграма DFD моделі

Після створення контекстної діаграми було виконано декомпозицію діаграми з 4 функціональними блоками: “Реєстрація та авторизація“, “Пошук та вибір фільмів“, “Створення кімнати для перегляду“ та “Спільний перегляд фільму“.

На рис.1.2 зображено декомпозицію контекстної діаграми.

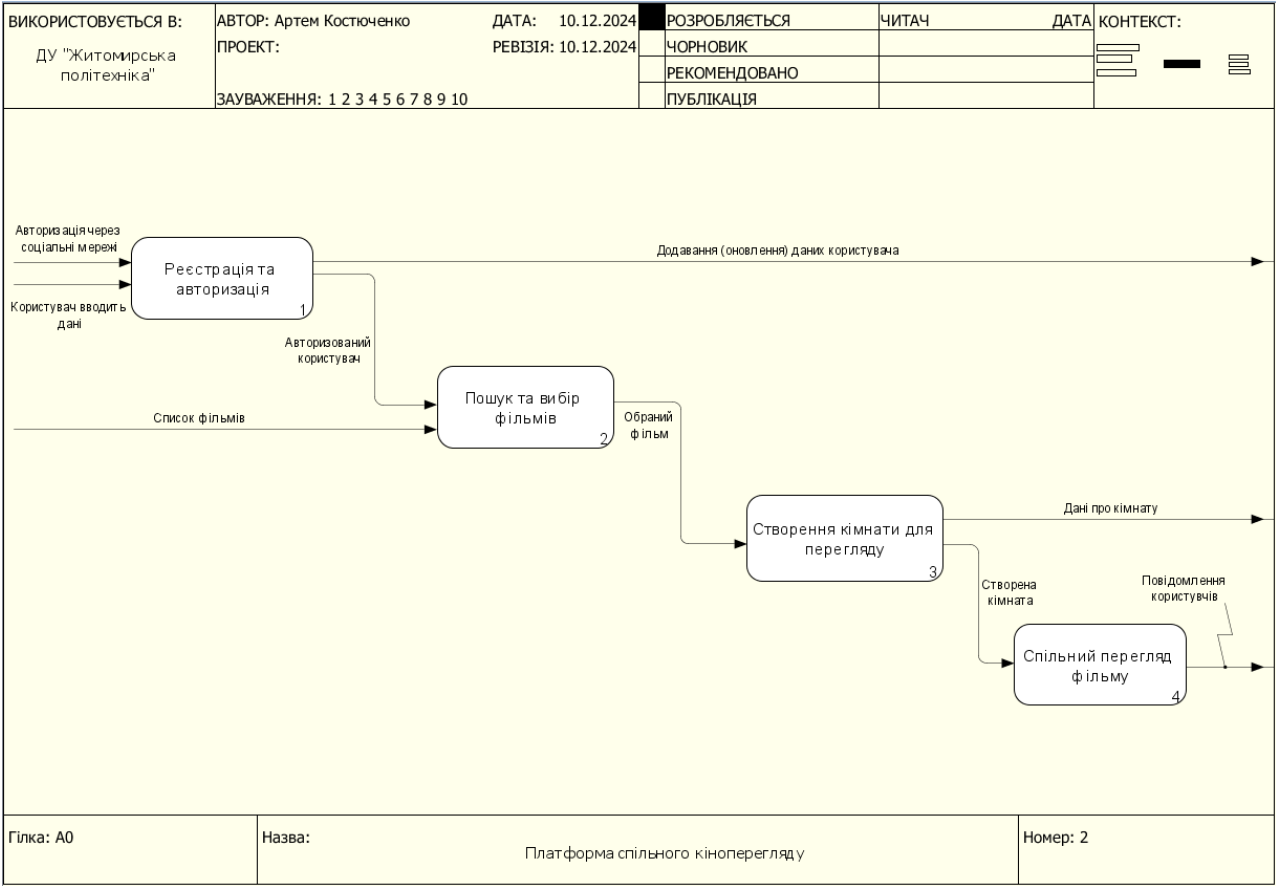


Рисунок 1.2 – Декомпозиція контекстної діаграми DFD

Було виконано декомпозицію кожного функціонального блоку.

Процес реєстрації та авторизації починається із введення даних користувачем або авторизації через соціальні мережі. Система перевіряє дані, реєструє або авторизує користувача та передає статус авторизованого користувача далі для доступу до функціональності платформи.

На рис.1.3 зображено декомпозицію роботи “Реєстрація та авторизація”.

Після авторизації користувач переходить до процесу пошуку та вибору фільмів. Він отримує доступ до списку фільмів, може переглядати, фільтрувати та вибирати бажаний фільм. Обраний фільм передається як результат цього процесу для подальшого створення кімнати для перегляду.

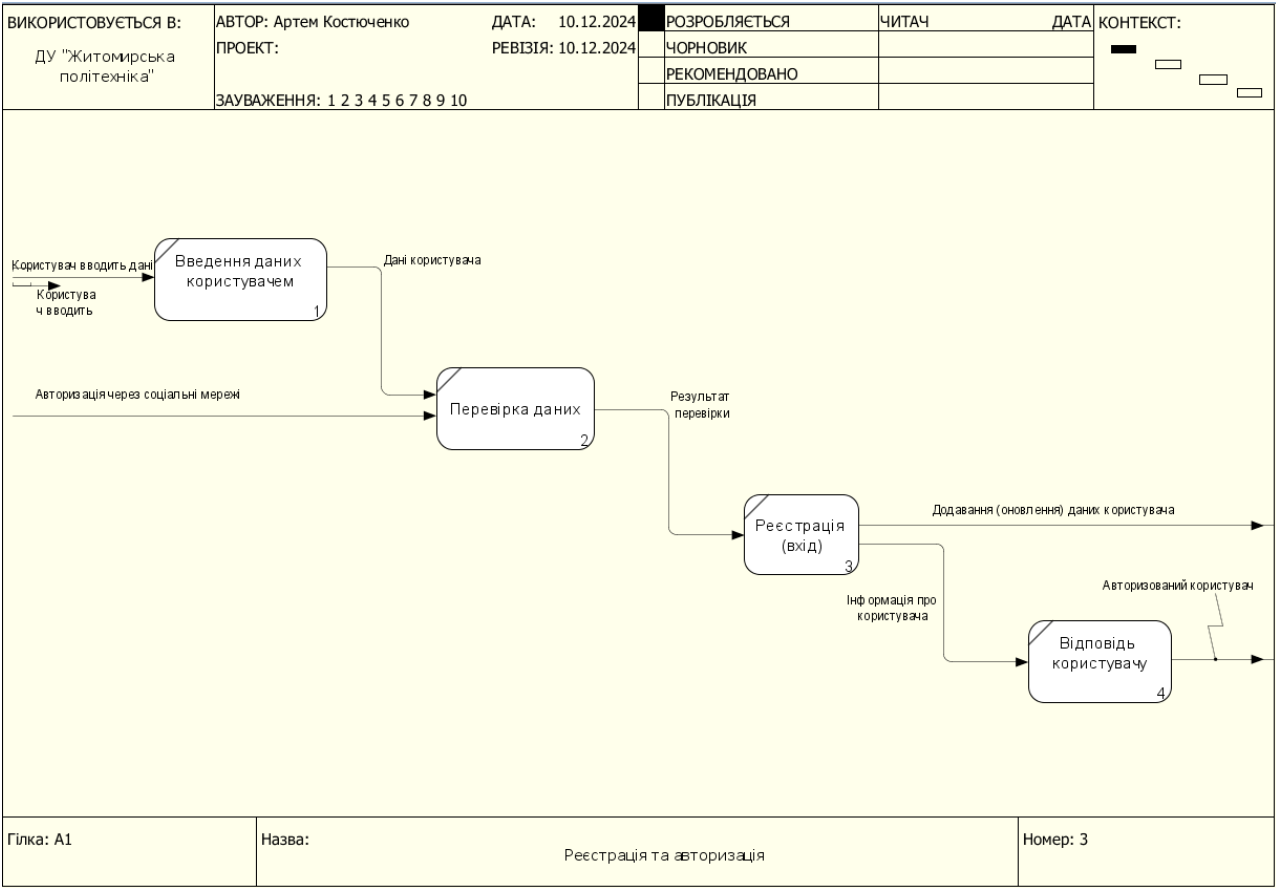


Рисунок 1.3 – Декомпозиція роботи “Реєстрація та авторизація”

На рис.1.4 зображено декомпозицію роботи “Пошук та вибір фільмів”.

Наступним процесом є створення кімнати для перегляду. На основі обраного фільму система створює віртуальну кімнату для спільного перегляду. У процесі формуються дані про кімнату, включаючи унікальний ідентифікатор, учасників та налаштування. Ці дані передаються у наступний процес.

На рис.1.5 зображено декомпозицію роботи “Створення кімнати для перегляду”.

Кінцевим процесом є спільний перегляд фільму. У створеній кімнаті користувач може розпочати спільний перегляд. Відбувається синхронізація відтворення для всіх учасників, а також надсилаються повідомлення користувачам, що підтверджують підключення до кімнати. Дані про кімнату та синхронізований перегляд передаються для подальшої обробки.

На рис.1.6 зображено декомпозицію роботи “Спільний перегляд фільму”.

ВИКОРИСТОВУЄТЬСЯ В: ДУ "Житомирська політехніка"	АВТОР: Артем Костюченко ПРОЕКТ: ЗАУВАЖЕННЯ: 1 2 3 4 5 6 7 8 9 10	ДАТА: 10.12.2024 РЕВІЗІЯ: 10.12.2024	<div> <div> <div>РОЗРОБЛЯЄТЬСЯ</div> <div>ЧОРНИК</div> <div>РЕКОМЕНДОВАНО</div> <div>ПУБЛІКАЦІЯ</div> </div> <div> <div>ЧИТАЧ</div> <div>ДАТА</div> </div> </div> <div> <div>КОНТЕКСТ:</div> <div> <div></div> <div></div> <div></div> <div></div> </div> </div>
---	--	---	--

```

graph LR
    Start([Користувач вводить дані]) --> Step1[/Введення даних користувачем 1/]
    Step1 -- "Дані користувача" --> Step2[/Перевірка даних 2/]
    Start2([Авторизація через соціальні мережі]) --> Step2
    Step2 -- "Результат перевірки" --> Step3[/Реєстрація (вихід) 3/]
    Step3 -- "Додавання (оновлення) даних користувача" --> Next1[ ]
    Step3 -- "Інформація про користувача" --> Step4[/Відповідь користувачу 4/]
    Step4 --> End([Авторизований користувач])
  
```

Гілка: A1	Назва: Реєстрація та авторизація	Номер: 3
-----------	----------------------------------	----------

Рисунок 1.4 – Декомпозиція роботи “Пошук та вибір фільмів”

ВИКОРИСТОВУЄТЬСЯ В: ДУ "Житомирська політехніка"	АВТОР: Артем Костюченко ПРОЕКТ: ЗАУВАЖЕННЯ: 1 2 3 4 5 6 7 8 9 10	ДАТА: 10.12.2024 РЕВІЗІЯ: 10.12.2024	<div> <div>РОЗРОБЛЯЄТЬСЯ</div> <div>ЧОРНИК</div> <div>РЕКОМЕНДОВАНО</div> <div>ПУБЛІКАЦІЯ</div> </div>	ЧИТАЧ ДАТА	<div> <div>КОНТЕКСТ:</div> <div> <div></div> <div></div> <div></div> <div></div> </div> </div>
---	--	---	--	---------------	--

```

graph TD
    Start([Користувач вводить дані]) --> Step1[Введення даних користувачем 1]
    Step1 -- "Дані користувача" --> Step2[Перевірка даних 2]
    Step2 -- "Авторизація через соціальні мережі" --> Step2
    Step2 -- "Результат перевірки" --> Step3[Реєстрація вхід 3]
    Step3 -- "Додавання оновлення даних користувача" --> End1([ ])
    Step3 -- "Інформація про користувача" --> Step4[Відповідь користувачу 4]
    Step4 -- "Авторизований користувач" --> End2([ ])
  
```

Гілка: A1	Назва: Реєстрація та авторизація	Номер: 3
-----------	-------------------------------------	----------

Рисунок 1.5 – Декомпозиція роботи “Створення кімнати для перегляду”

		Костюченко А.В.			«Житомирська політехніка» 24.121.14.000	Арк.
						10
Змн.	Арк.	№ докум.	Підпис	Дата		

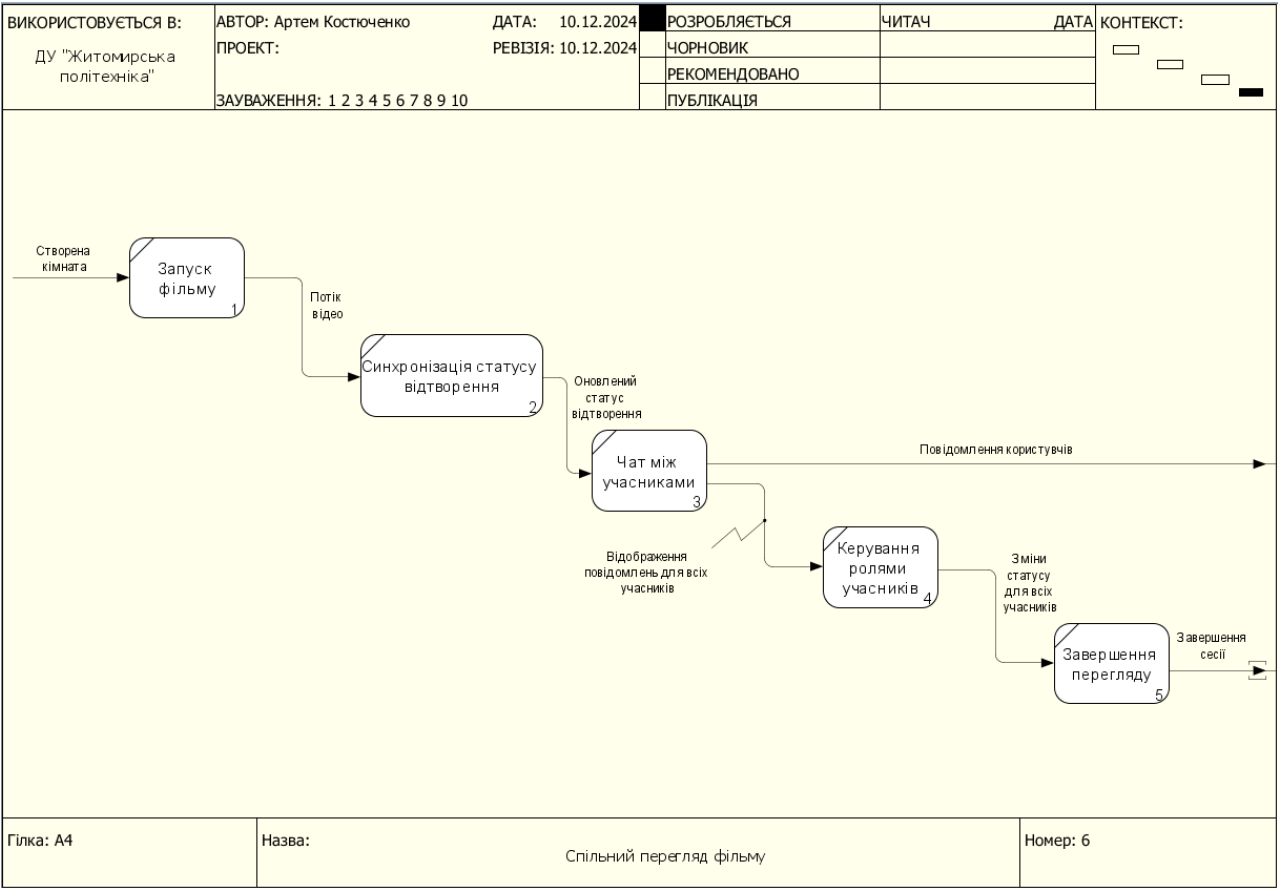


Рисунок 1.6 – Декомпозиція роботи “Спільний перегляд фільму”

Наступним кроком було розроблено IDEF0 діаграму.

Інтерфейси входу: “Дані користувачів”, “Обраний фільм”, інтерфейси управління: “Політика доступу“, “Налаштування кімнати“, інтерфейси механізму: “Сервер“, “База даних“, “API авторизації“ та інтерфейси виходу: “Створена кімната”, “Синхронізоване відтворення”, “Історія чату”.

На рис.1.7 зображено контекстну діаграму IDEF0 моделі.

Дана контекстна діаграма після декомпозиції складається з 4 функціональних блоків: “Реєстрація та авторизація“, “Пошук та вибір фільмів“, “Створення кімнати для перегляду“ та “Спільний перегляд фільму“.

Опис кожної декомпозиції функціонального блоку відповідає блокам DFD.

На рис.1.8 зображено декомпозицію контекстної діаграми.

На рис.1.9 зображено декомпозицію роботи “Реєстрація та авторизація”.

На рис.1.10 зображено декомпозицію роботи “Пошук та вибір фільмів”.

На рис.1.11 зображено декомпозицію роботи “Створення кімнати для перегляду”.

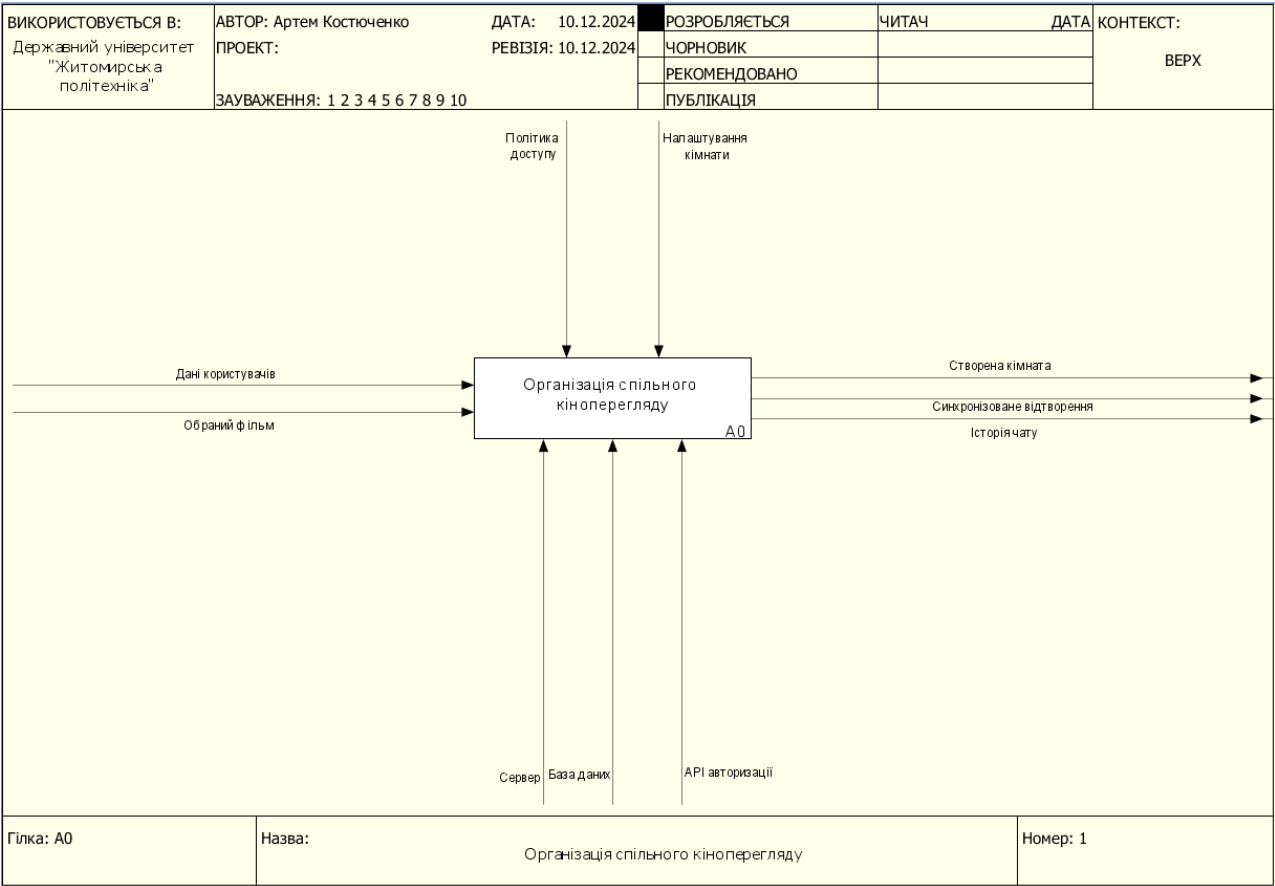


Рисунок 1.7 – Контекстна діаграма IDEF0 моделі

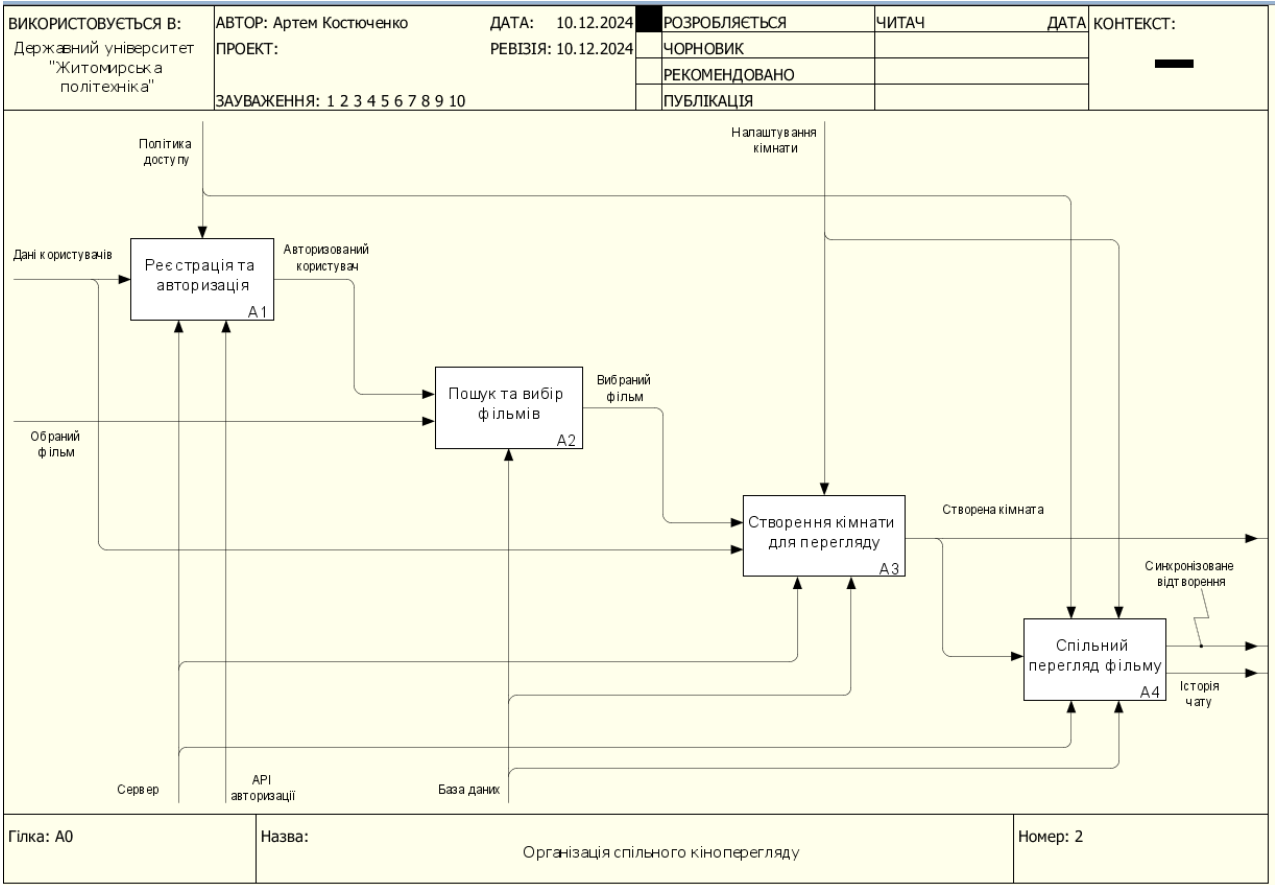


Рисунок 1.8 – Декомпозиція контекстної діаграми IDEF0

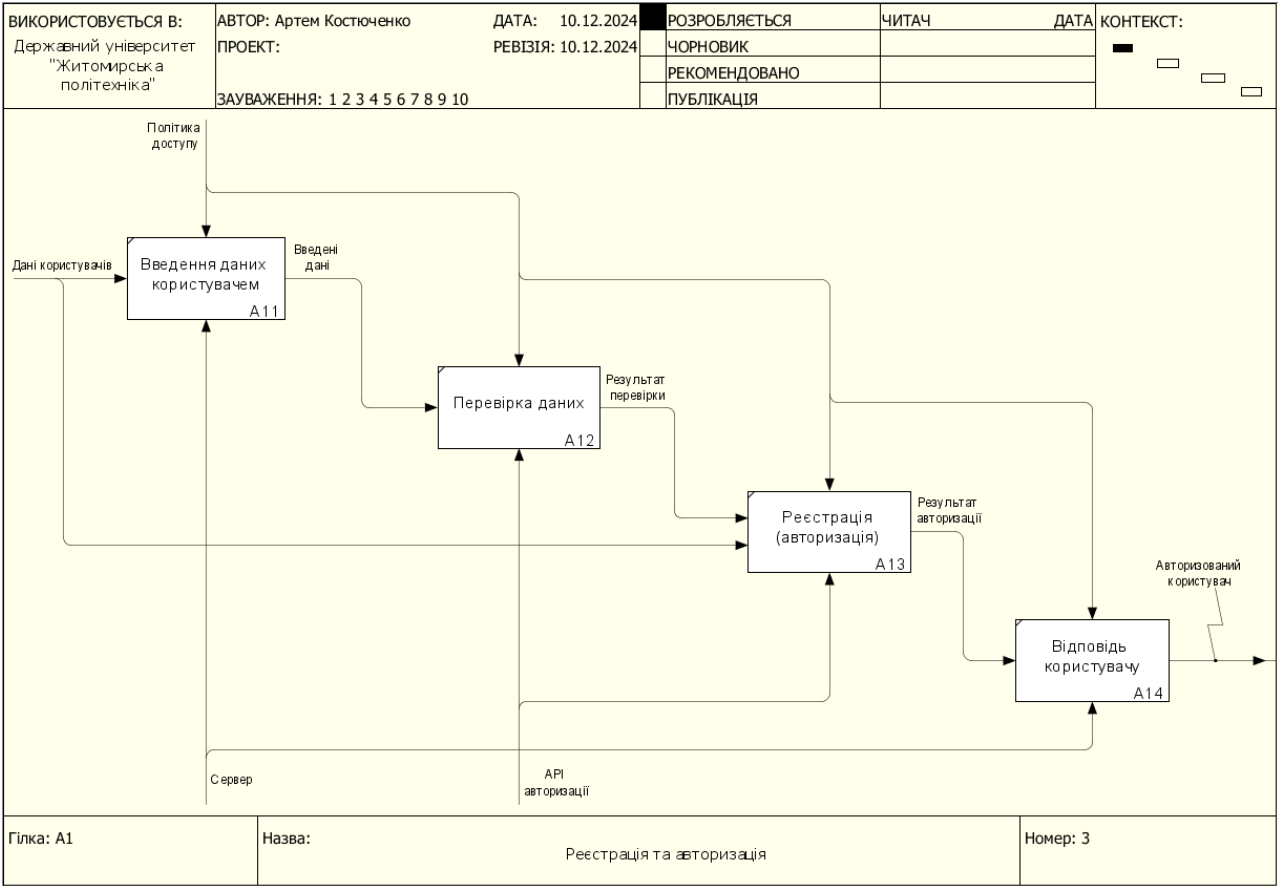


Рисунок 1.9 – Декомпозиція роботи “Реєстрація та авторизація”

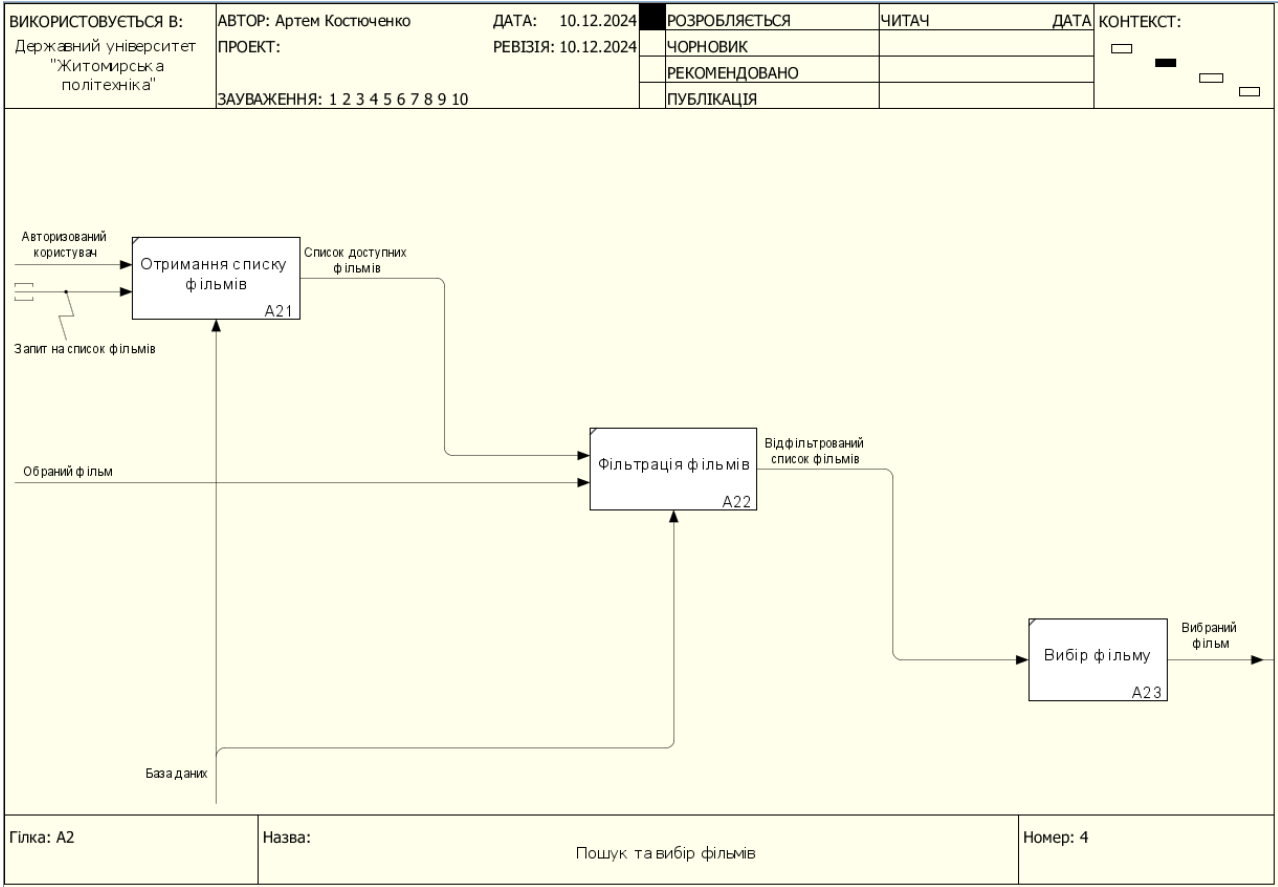


Рисунок 1.10 – Декомпозиція роботи “Пошук та вибір фільмів”

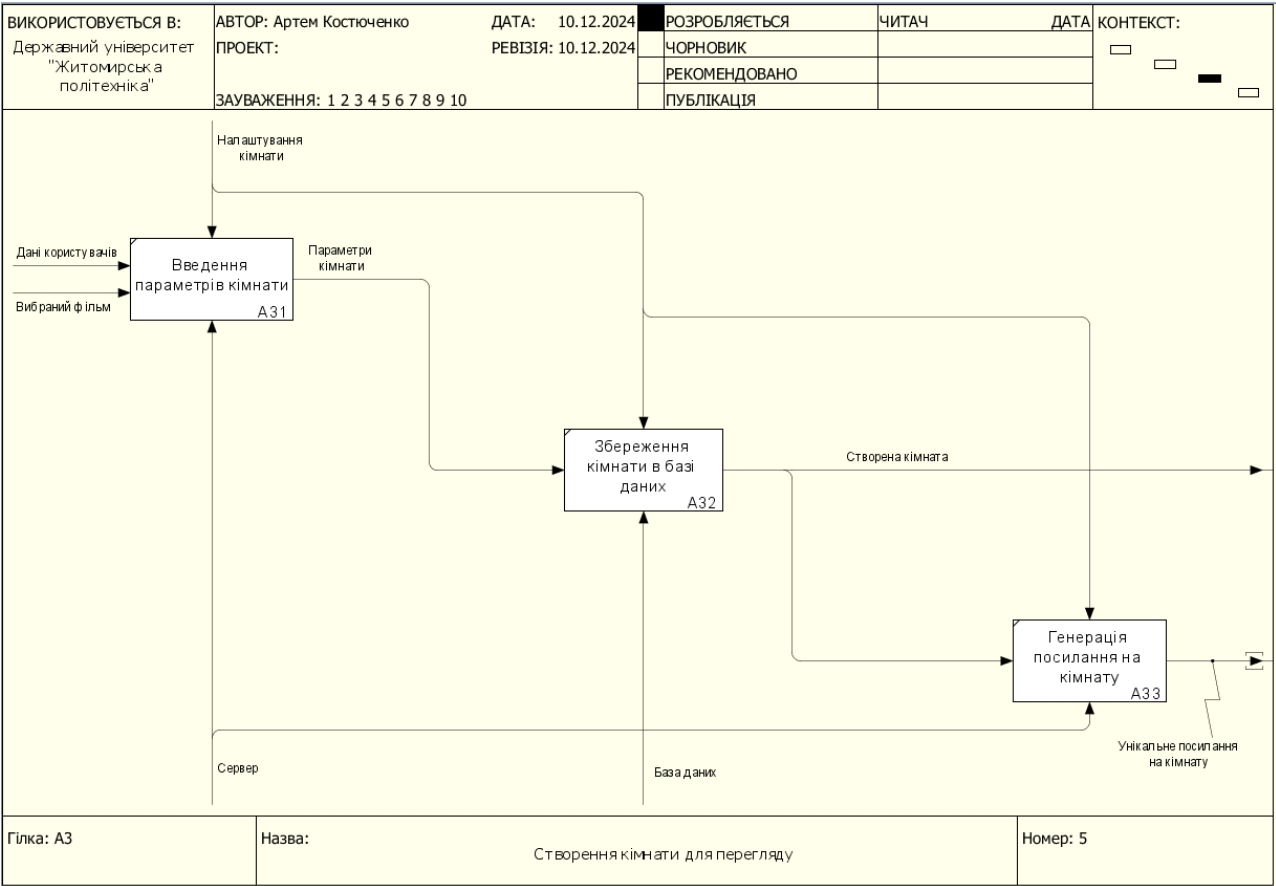


Рисунок 1.11 – Декомпозиція роботи “Створення кімнати для перегляду”

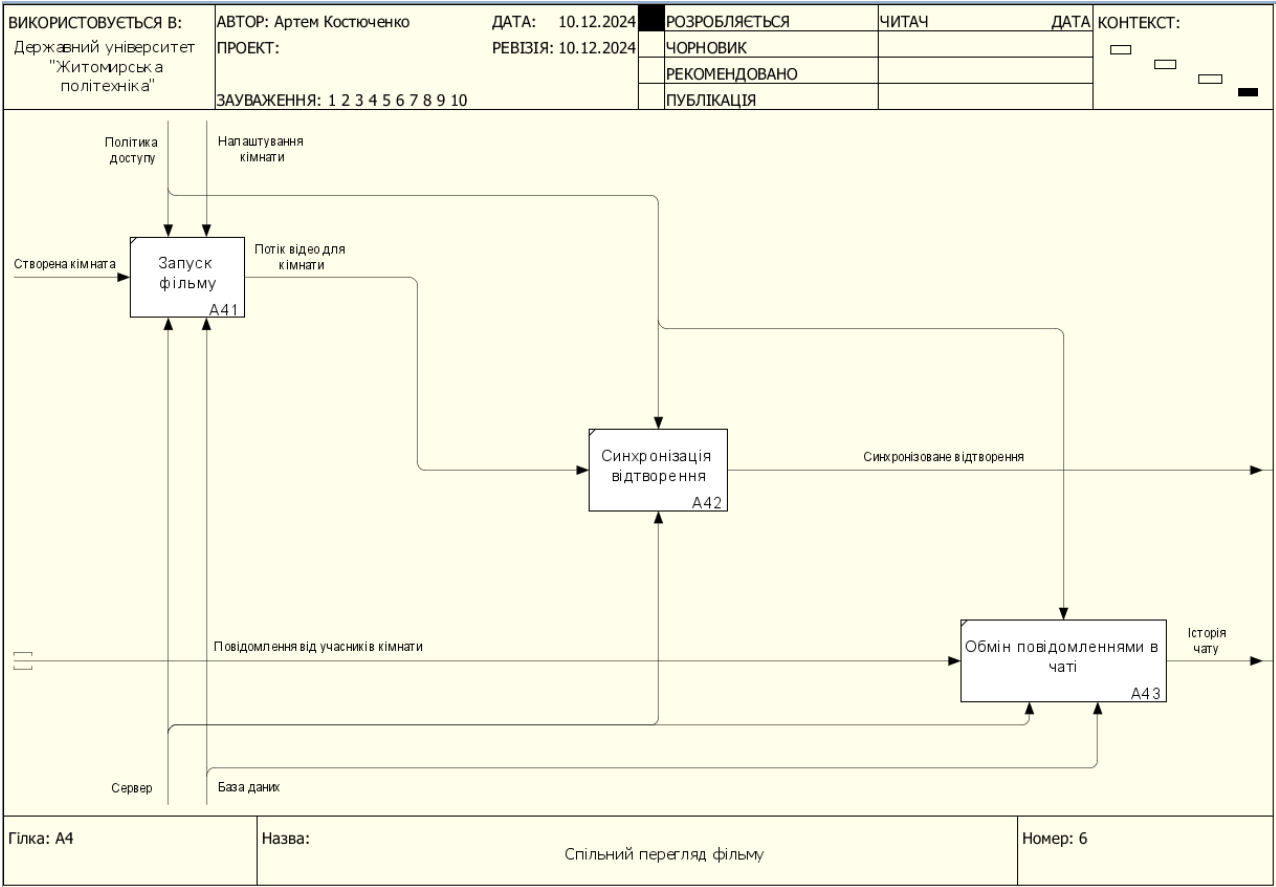


Рисунок 1.12 – Декомпозиція роботи “Спільний перегляд фільму”

На рис.1.12 зображено декомпозицію роботи “Спільний перегляд фільму”.

Кінцевим кроком було розроблено три BPMN діаграми процесів роботи системи.

Діаграма процесів “Створення кімнати” відбувається наступним чином: користувач відкриває головну сторінку та обирає фільм/серіал, далі він натискає на опцію "Створити кімнату" та заповнює форму, після заповнення користувач відправляє форму, система отримує дані та перевіряє їх валідацію, якщо дані є некоректними, система повідомляє користувача про невідповідність, у випадку, якщо дані правильні, система створює кімнату, після цього користувача перенаправляють на сторінку кімнати, система отримує відповідне повідомлення та відображає його користувачу і нарешті користувачу надсилається підтвердження про його приєднання до кімнати.

На рис.1.13 зображено діаграму процесу “Створення кімнати”.

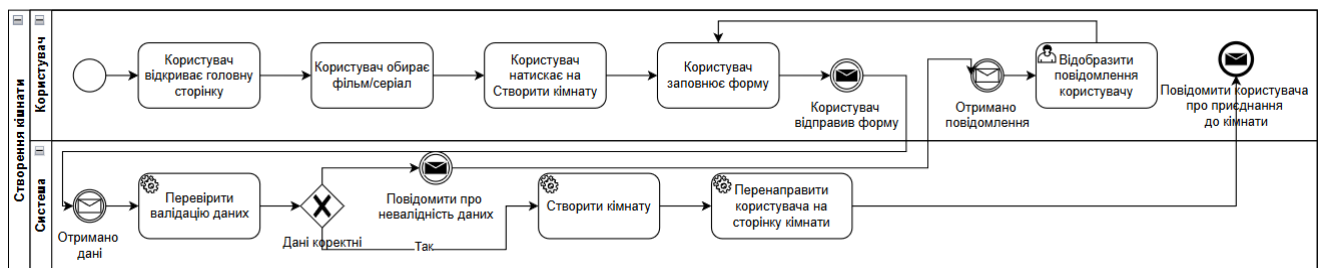


Рисунок 1.13 – Діаграма процесу “Створення кімнати”

Діаграма процесів “Під’єднання до кімнати” відбувається наступним чином: користувач отримує запрошувальне посилання від власника кімнати та відкриває його; система перевіряє існування посилання, якщо посилання є недійсним або протермінованим, користувачу надсилається повідомлення про це, якщо посилання є дійсним, система перевіряє існування кімнати, у разі, якщо кімната не існує, користувач отримує відповідне повідомлення про неіснуючу кімнату, якщо кімната існує, система підключає користувача до кімнати та надсилає повідомлення про успішне приєднання.

На рис.1.14 зображено діаграму процесу “Під’єднання до кімнати”.

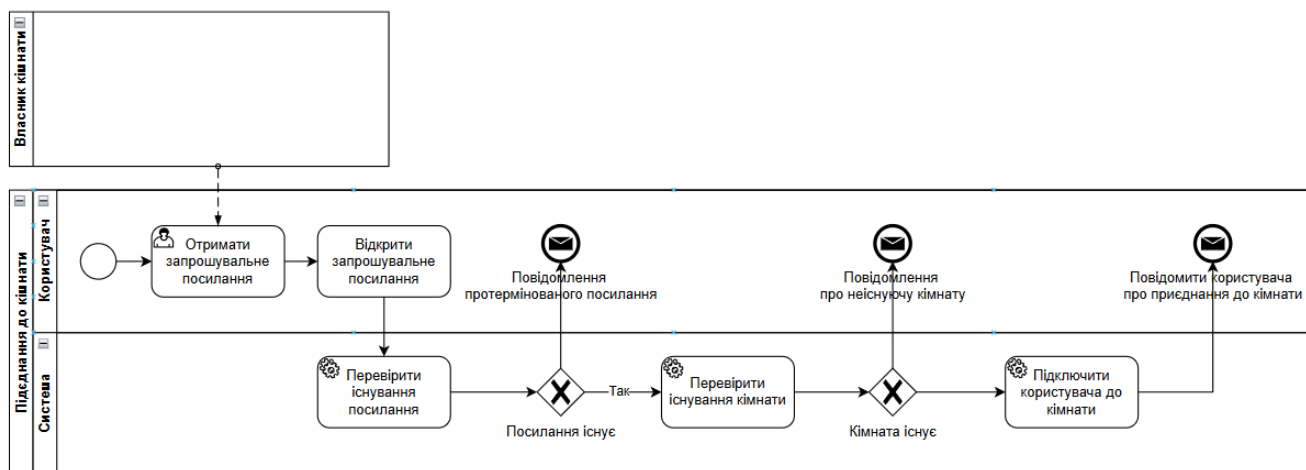


Рисунок 1.14 – Діаграма процесу “Під’єднання до кімнати”

Діаграма процесів “Синхронізація перегляду” відбувається наступним чином: користувач відкриває сторінку кімнати, після чого система перевіряє доступ до управління керуванням, якщо доступ існує, користувач може виконувати дії, такі як відтворення, пауза або перемотування; при виконанні цих дій генеруються відповідні події: "Play", "Pause" та "Seek"; система оброблює подію користувача, оновлює статус перегляду та зберігає поточний статус відтворення, після цього статус відтворення відправляється системою для подальшої синхронізації відтворення серед користувачів у кімнаті; процес завершується синхронізацією відтворення.

На рис.1.15 зображено діаграму процесу “Синхронізація перегляду”.

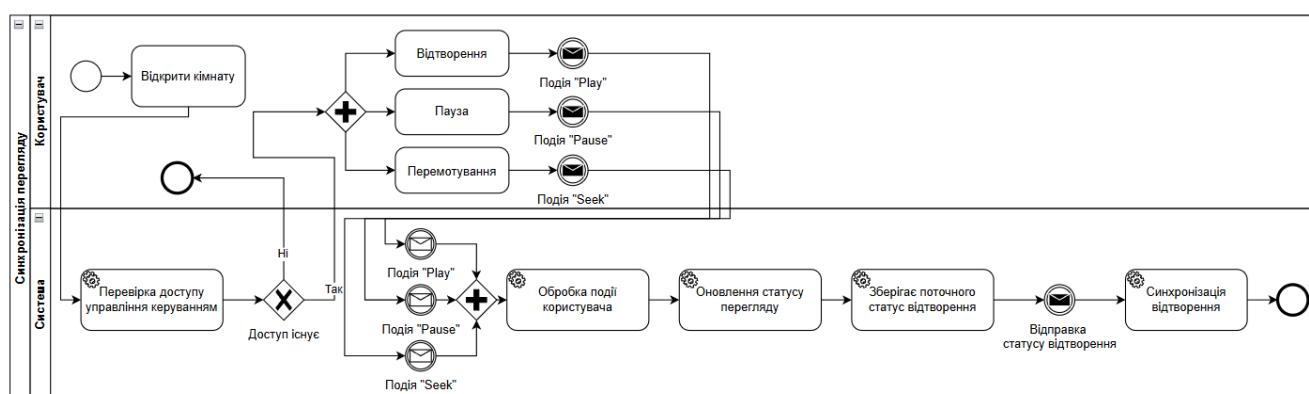


Рисунок 1.15 – Діаграма процесу “Синхронізація перегляду”

Висновки до розділу 1

У даному розділі було проведено моделювання інформаційних процесів платформи для спільного кіноперегляду. Перед початком моделювання було поставлено задачі, які потрібно виконати. Задачі були розбиті на основні етапи для зручності реалізації інформаційних процесів даної системи.

Додатково було проаналізовано та обґрунтовано різні методи та нотації для реалізації діаграм та програмні засоби для їх створення. Проведено моделювання процесів за допомогою моделей DFD та IDEF0, описано та відображено основні процеси роботи системи за допомогою нотації BPMN 2.0.

		Костюченко А.В.			«Житомирська політехніка» 24.121.14.000	Арк.
						17
Змн.	Арк.	№ докум.	Підпис	Дата		

РОЗДІЛ 2. ОРГАНІЗАЦІЯ СХОВИЩА ДАНИХ ДЛЯ АНАЛІТИКИ ТА ETL ПРОЦЕСУ

2.1 Визначення даних для аналізу та розробка сховища даних

Транзакційне сховище даних системи, яке буде активно використовуватися під час роботи платформи для спільного кіноперегляду, має бути спроектоване до того, як буде розроблено аналітичне сховище даних.

При створенні OLTP-схеми для цієї платформи необхідно враховувати важливі фактори, як нормалізація даних для мінімізації дублювання, забезпечення транзакційної ізоляції та максимізації ефективності обробки запитів при роботі з кімнатами, користувачами, чатом та відтворенням відеоконтенту.

Транзакційне сховище даних міститиме наступні таблиці, які мають забезпечувати роботу основного функціоналу платформи.

На табл.2.1 представлений опис таблиці “Users”, яка зберігає інформацію про користувачів.

Таблиця 2.1

Поле	Тип даних	Опис
user_id	INT (PK)	Унікальний ідентифікатор користувача
username	NVARCHAR(50)	Ім'я користувача
email	NVARCHAR(100)	Унікальна електронна адреса користувача
password	NVARCHAR(255)	Пароль користувача
role	NVARCHAR(20)	Роль користувача (за замовчуванням 'user')
date_created	DATETIME	Час створення користувача

На табл.2.2 представлений опис таблиці “Room”, яка містить інформацію про кімнати для спільного перегляду, створені користувачами.

Таблиця 2.2

Поле	Тип даних	Опис
room_id	INT (PK)	Унікальний ідентифікатор кімнати
room_name	NVARCHAR(100)	Назва кімнати
is_private	BOOLEAN	Позначка, чи є кімната приватною
created_at	DATETIME	Час створення кімнати
created_by	INT (FK)	Ідентифікатор користувача

На табл.2.3 представлений опис таблиці “TypesContent”, яка визначає типи контенту, такі як "фільм" чи "серіал", щоб категоризувати різні види відеоконтенту.

Таблиця 2.3

Поле	Тип даних	Опис
type_content_id	INT (PK)	Унікальний ідентифікатор типу контенту
name	NVARCHAR(50)	Назва типу контенту

На табл.2.4 представлений опис таблиці “VideoContent”, яка зберігає інформацію про відеоконтент, який можна переглядати в кімнатах.

Таблиця 2.4

Поле	Тип даних	Опис
video_id	INT (PK)	Унікальний ідентифікатор відеоконтенту
title	NVARCHAR(200)	Назва відеоконтенту
original_title	NVARCHAR(200)	Оригінальна назва відеоконтенту
IMDb	FLOAT	Оцінка на IMDb
description	TEXT	Опис відеоконтенту

type_content	INT (FK)	Ідентифікатор типу відеоконтенту
release_date	DATE	Дата випуску
duration	INT	Тривалість у секундах
preview_url	NVARCHAR(255)	Посилання основного фото
background_url	NVARCHAR(255)	Посилання заднього фону
trailer_link	NVARCHAR(255)	Посилання трейлера
is_serial	BOOLEAN	Позначка, чи є контент серіалом

На табл.2.5 представлений опис таблиці “PlaylistUrl”, яка містить посилання на відеоконтент в різних якостях.

Таблиця 2.5

Поле	Тип даних	Опис
playlist_id	INT (PK)	Унікальний ідентифікатор плейліста
video_id	INT (FK)	Ідентифікатор відеоконтенту
url	NVARCHAR(255)	Посилання відео
quality	NVARCHAR(50)	Якість відеоконтенту

На табл.2.6 представлений опис таблиці “RoomAccess”, яка відстежує доступ користувачів до конкретних кімнат і їхні права контролю відтворення.

Таблиця 2.6

Поле	Тип даних	Опис
access_id	INT (PK)	Унікальний ідентифікатор доступу
room_id	INT (FK)	Ідентифікатор кімнати

user_id	INT (FK)	Ідентифікатор користувача
can_control	BOOLEAN	Позначка, чи може користувач керувати відтворенням
is_blocked	BOOLEAN	Позначка, чи користувач заблокований
access_granted_at	DATETIME	Час надання доступу
blocked_at	DATETIME	Час блокування

На табл.2.7 представлений опис таблиці “PlaybackControl”, що зберігає інформацію про дії користувачів із відтворення контенту у конкретній кімнаті.

Таблиця 2.7

Поле	Тип даних	Опис
control_id	INT (PK)	Унікальний ідентифікатор дії управління відтворенням
room_id	INT (FK)	Ідентифікатор кімнати
user_id	INT (FK)	Ідентифікатор користувача
video_id	INT (FK)	Ідентифікатор відеоконтенту
seek	INT	Позиція відтворення у секундах
action	NVARCHAR(50)	Тип дії
action_at	DATETIME	Час виконання дії

На табл.2.8 представлений опис таблиці “Message”, яка зберігає інформацію про повідомлення користувачів у конкретній кімнаті.

Таблиця 2.8

Поле	Тип даних	Опис
message_id	INT (PK)	Унікальний ідентифікатор повідомлення
room_id	INT (FK)	Ідентифікатор кімнати
user_id	INT (FK)	Ідентифікатор користувача
text	TEXT	Текст повідомлення
sent_at	DATETIME	Час відправлення повідомлення
is_edited	BOOLEAN	Позначка, чи було відредаговано повідомлення
is_deleted	BOOLEAN	Позначка, чи було видалено повідомлення

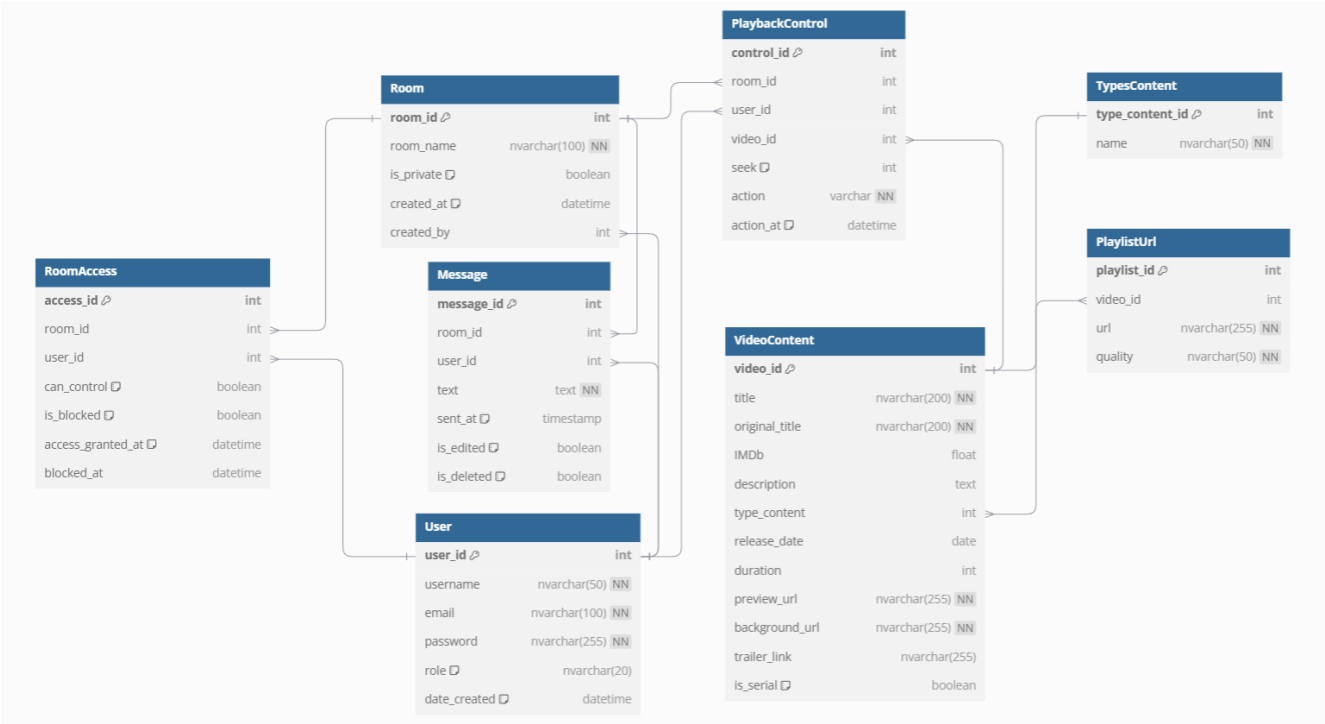


Рис. 2.1 – OLTP діаграма

На основі створених таблиць OLTP діаграми було створено OLAP діаграма, яка забезпечує можливість багатовимірного аналізу даних, що полегшує процес

створення звітів і виконання аналітичних запитів, а також консолідацію даних для прийняття стратегічних управлінських рішень.

Діаграма була розроблена у вигляді зіркоподібної схеми, яка є однією з найпоширеніших моделей структурування даних у сховищах. Її ключовою перевагою є спрощена архітектура, яка підвищує ефективність і швидкість виконання аналітичних операцій.

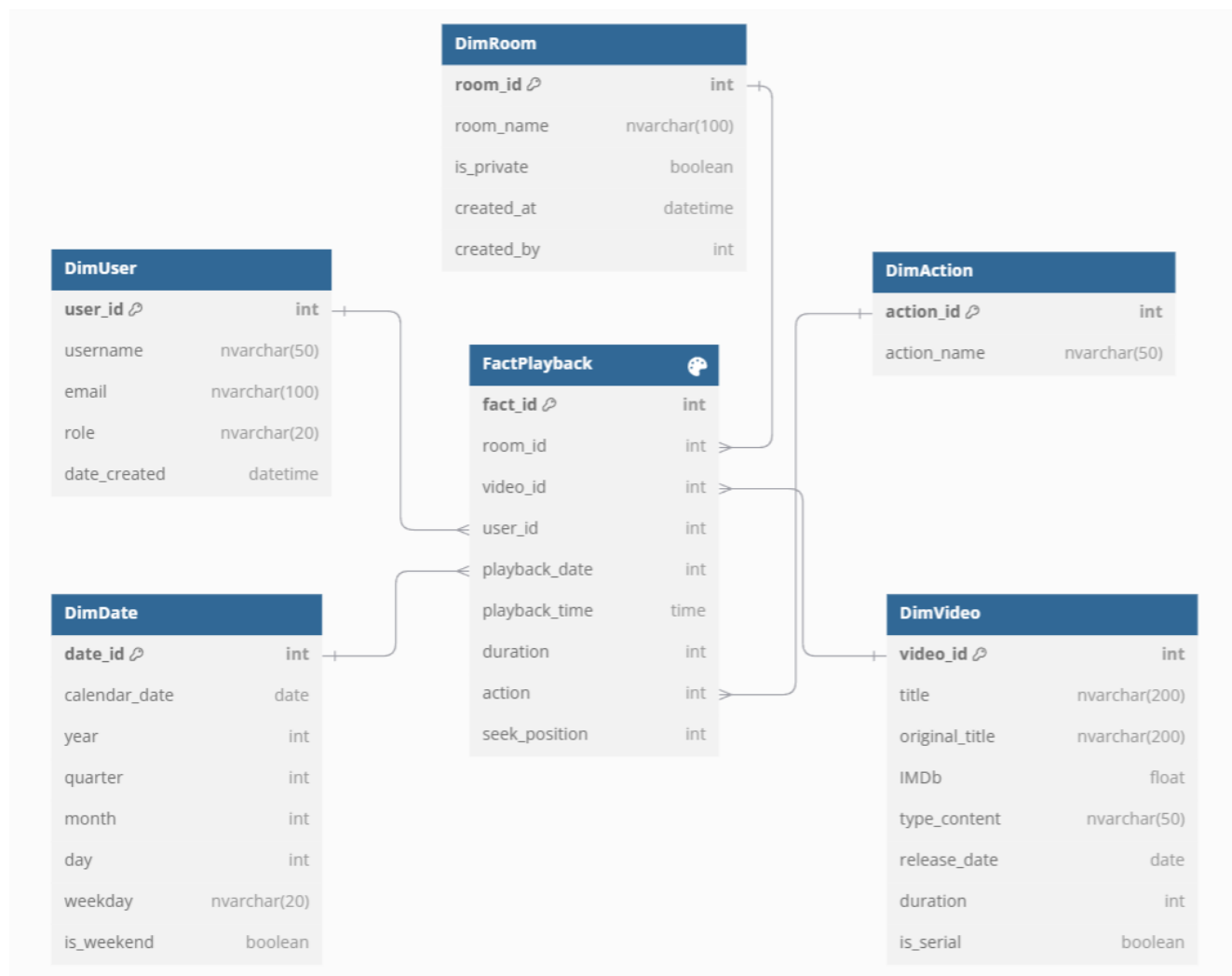


Рис. 2.2 - OLAP діаграма

2.2 ETL процес та використання Azure Synapse Analytics для копіювання та трансформації даних

Azure Synapse Analytics є інтегрованою платформою для обробки та аналізу великих даних. Вона об'єднує ETL-процеси, аналітику та візуалізацію в єдиному середовищі, що дозволяє інтегрувати дані з різних джерел (наприклад CSV файли),

трансформувати їх за допомогою SQL або Apache Spark, і зберігати для аналітичної обробки.

Початковим етапом трансформації даних у Azure Synapse Analytics є створення Azure Storage Account, яке слугуватиме сховищем для файлів із даними, що підлягають обробці. Для цього потрібно створити Storage Account в Azure.

На рис.2.3 представлено створення нового Storage Account.

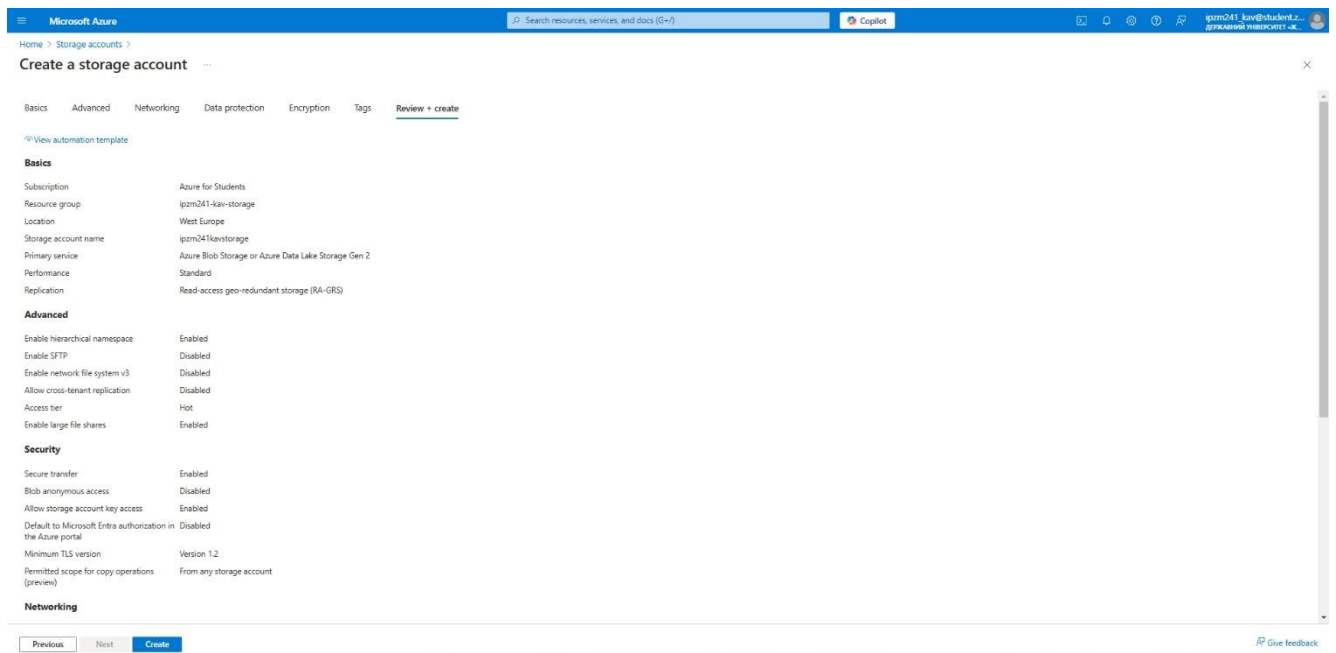


Рис. 2.3 - Створення сховища даних

Після створення сховища даних можна налаштувати новий робочий простір за допомогою сервісу Azure Synapse Analytics та підключити інтеграцію зі створеним сховищем даних.

На рис.2.4 представлено створення нового робочого простору за допомогою сервісу Azure Synapse Analytics.

Після створення робочого простору необхідно завантажити дані в сховище, для чого спершу було створено новий контейнер для зберігання даних, у межах якого для зручності було організовано папки відповідно до кожного набору даних.

На рис.2.5 представлено завантажений CSV-файл в сховище даних через Microsoft Azure Storage Explorer.

На рис.2.6 представлено підтвердження завантаженого набору даних.

		Костюченко А.В.			«Житомирська політехніка» 24.121.14.000	Арк.
						24
Змн.	Арк.	№ докум.	Підпис	Дата		

Microsoft Azure

Search resources, services, and docs (Ctrl+)

Copilot

ipzm241_kav@student...
STUDENT

Home >

Create Synapse workspace

Basics

Security

Networking

Tags

Review + create

Create a Synapse workspace to develop an enterprise analytics solution in just a few clicks.

Project details

Select the subscription to manage deployed resources and costs. Use resource groups like folders to organize and manage all of your resources.

Subscription *

Azure for Students

Resource group *

(New) ipzm241-kav-synapse

Create new

Managed resource group

Enter managed resource group name

Workspace details

Name your workspace, select a location, and choose a primary Data Lake Storage Gen2 file system to serve as the default location for logs and job output.

Workspace name *

ipzm241kavcourseworkspase

Region *

East US 2

Select Data Lake Storage Gen2 *

From subscription

Manually via URL

Account name *

ipzm241kavstorage

Create new

File system name *

data

Create new

Assign myself the Storage Blob Data Contributor role on the Data Lake Storage Gen2 account to interactively query it in the workspace.

We will automatically grant the workspace identity data access to the specified Data Lake Storage Gen2 account, using the Storage Blob Data Contributor role. To enable other users to use this storage account after you create your workspace, perform these tasks:

- Assign other users to the Contributor role on workspace
- Assign other users the appropriate Synapse SAS roles using Synapse Studio
- Assign yourself and other users to the Storage Blob Data Contributor role

Review + create

< Previous

Next: Security >

https://portal.azure.com/

Рис. 2.4 - Створення робочого простору Azure Synapse Analytics

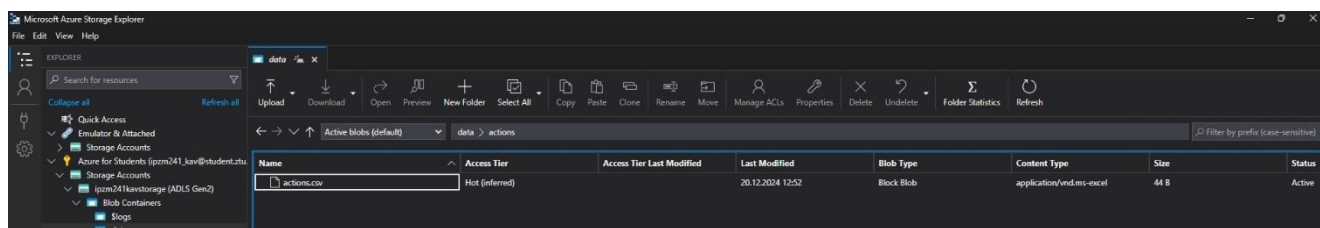


Рис. 2.5 - Завантажений CSV-файл в сховище даних

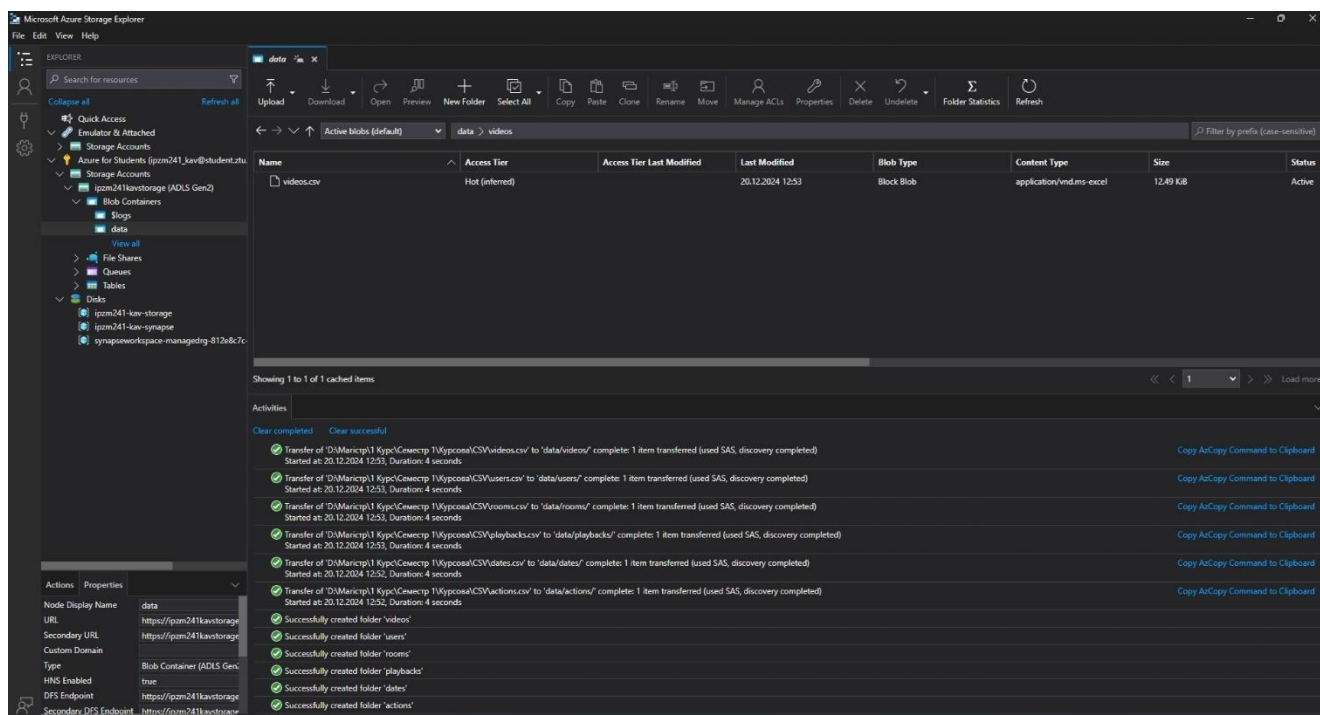


Рис. 2.6 - Завантажений набір даних

Після завантаження набору даних необхідно створити віддалений SQL пул, який слугуватиме платформою для створення таблиць, що формуватимуть аналітичне сховище даних.

На рис.2.7 зображено створення нового віддаленого SQL пулу.

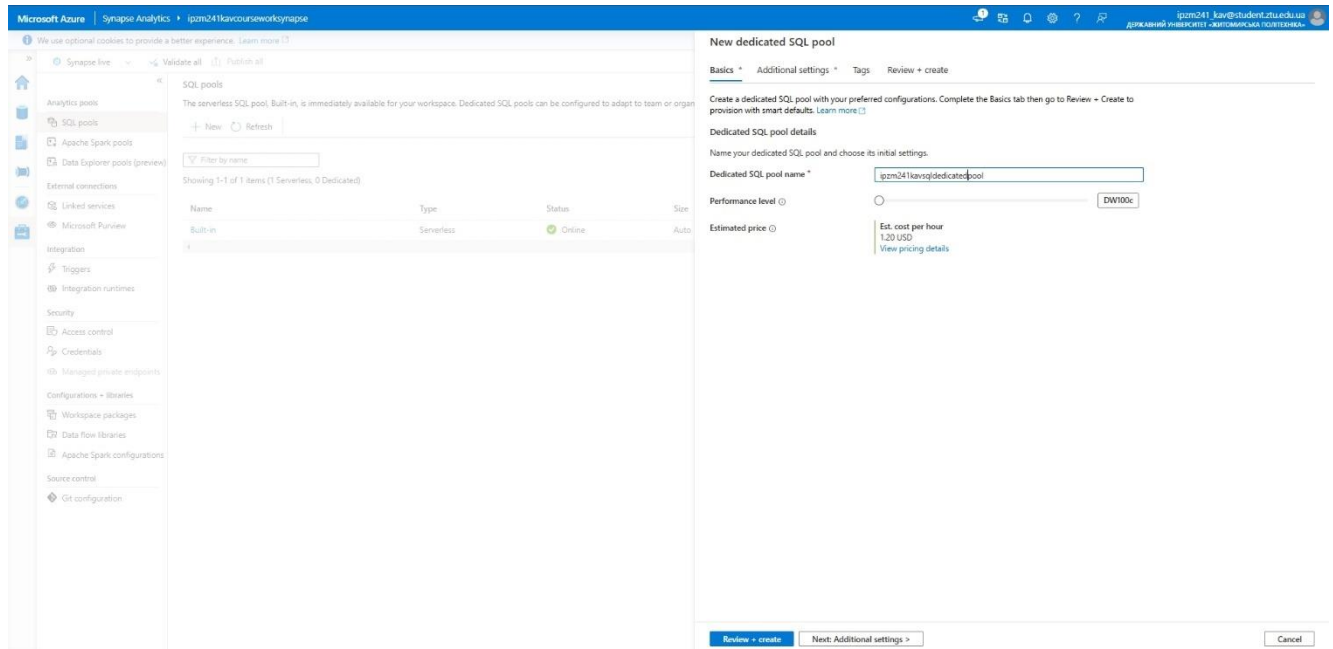


Рис. 2.7 - Створення SQL пулу

Після створення та успішної активації SQL-пулу необхідно виконати створення таблиць, використовуючи скрипт, представлений у Додатку А.

На рис.2.8 зображено виконання скрипта для створення аналітичних таблиць.

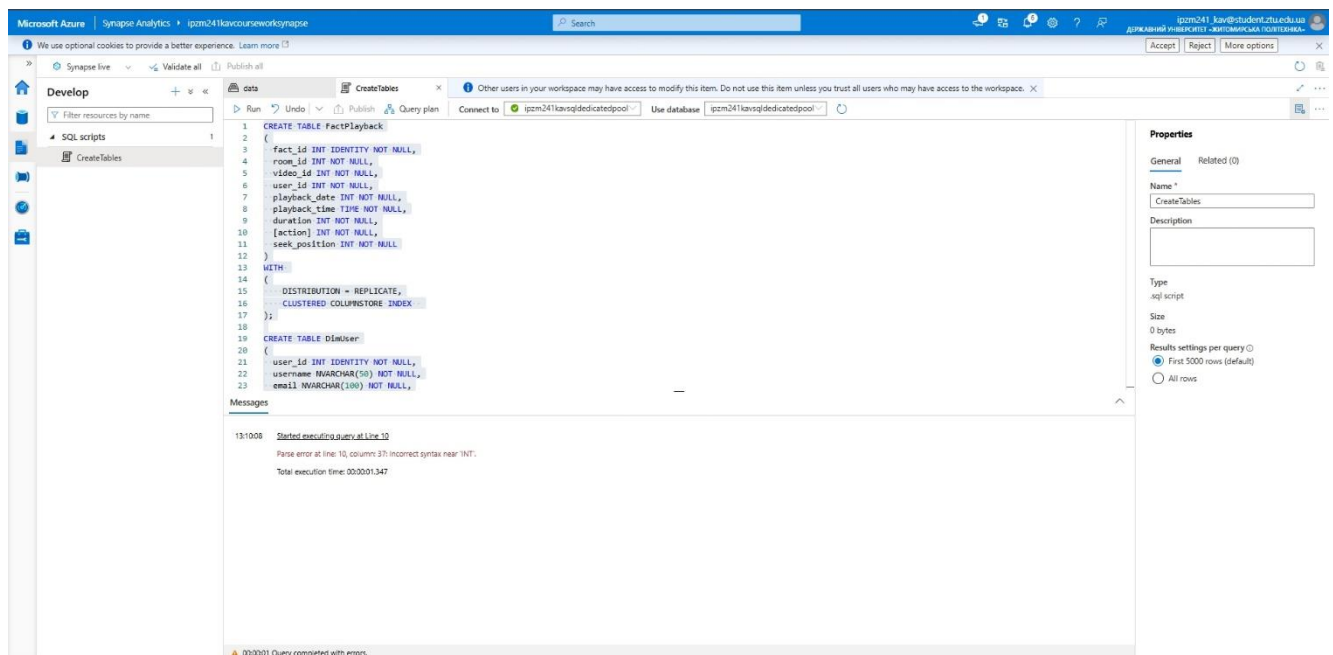


Рис. 2.8 - Створення таблиць

		Костюченко А.В.			«Житомирська політехніка» 24.121.14.000	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		26

На рис.2.9 зображено створені таблиці за допомогою SQL скрипта.

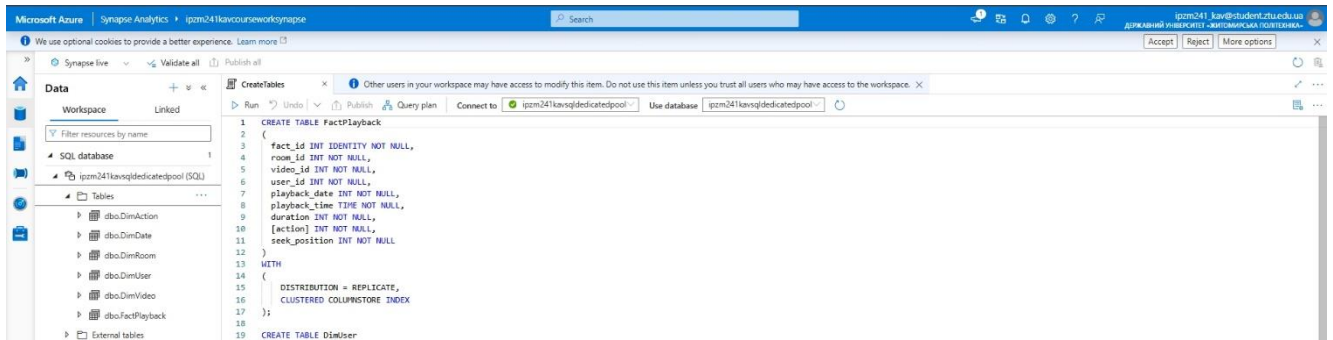


Рис. 2.9 - Створені таблиці

Після створення таблиць можна перейти до завантаження даних зі сховища файлів у таблиці. Для цього було створено та виконано SQL скрипт, представлений в Додатку А.

На рис.2.10 зображено виконання скрипта для завантаження даних у таблиці.

На рис.2.11 представлено перевірку завантажених даних у таблиці.

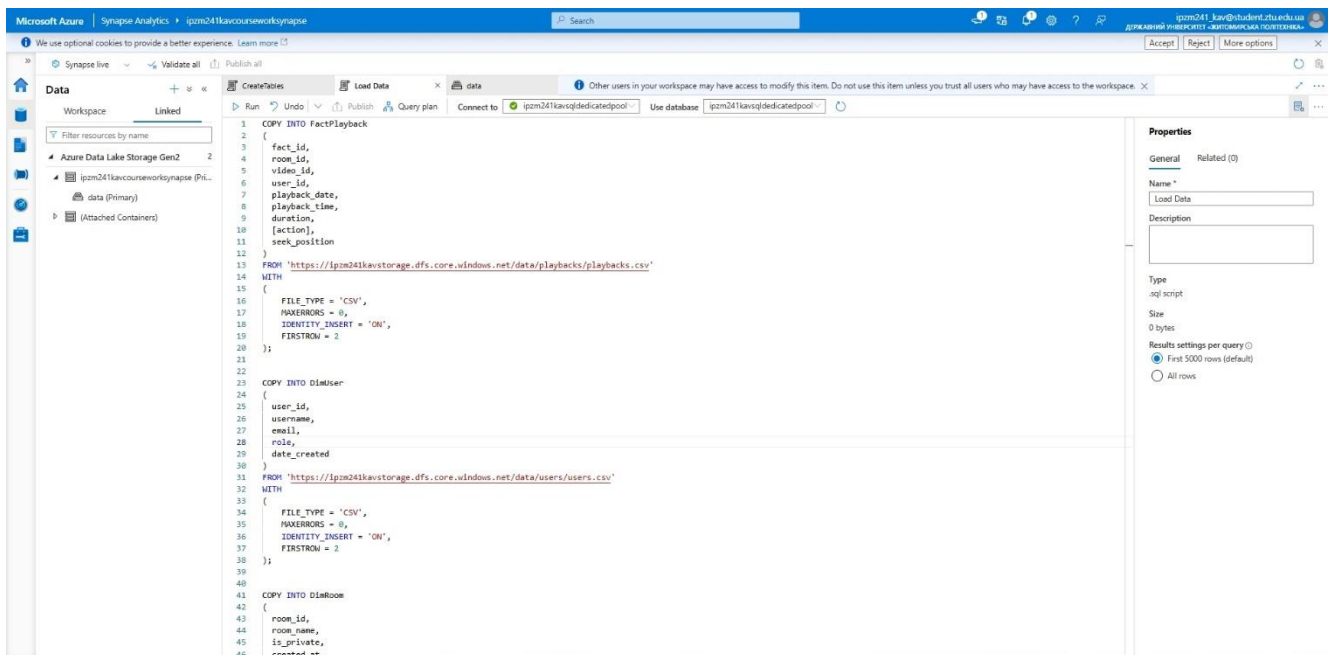


Рис. 2.10 - Завантаження даних у таблиці

Аналітичне сховище, яке містить дані для проведення аналітичного дослідження, було успішно створено. Наступним етапом є розробка дев'яти аналітичних SQL-запитів, спрямованих на аналіз ключових метрик платформи для спільного кіноперегляду, результати яких будуть візуалізовані для подальшого використання.

На рис.2.12 – 2.20 представлено аналітичні запити.

		Костюченко А.В.			«Житомирська політехніка» 24.121.14.000	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		27

Microsoft Azure

Synapse Analytics

ipzm241kavcourseworksynapse

Search

ipzm241_kav@student.uzhdu.edu.ua

ДЛЯ НАВЧАННЯ УНІВЕРСИТЕТУ - ЗАХОДИ НАС НА ПОШТОУ

Accept

Reject

More options

We use optional cookies to provide a better experience. Learn more

Synapse live

Validate all

Publish all

SQL script 1

Other users in your workspace may have access to modify this item. Do not use this item unless you trust all users who may have access to the workspace.

Workspace

Linked

Filter resources by name

SQL database

ipzm241kavsqldedicatedpool (SQL)

Tables

dbo.DimAction

dbo.DimDate

Columns

dbo.DimRoom

dbo.DimUser

dbo.DimVideo

Columns

dbo.FactPlayback

Columns

External tables

External resources

Views

Programmability

Schemas

Security

Create Tables

data

SQL script 1

Connect to ipzm241kavsqldedicatedpool

Use database ipzm241kavsqldedicatedpool

Run

Undo

Publish

Query plan

```
1 SELECT TOP (100) [video_id]
2 , [title]
3 , [original_title]
4 , [IMDb]
5 , [type_content]
6 , [release_date]
7 , [duration]
8 , [is_serial]
9 FROM [dbo].[DimVideo]
```

Results

Messages

View

Table

Chart

Export results

Search

video_id	title	original_title	IMDb	type_content	release_date	duration	is_serial
1	Той, що біжить лабіринтом	The Maze Runner	6.8	Movie	2014-09-19	144	False
2	Гра престолів	Game of Thrones	9.2	Series	2011-04-17	57	True
3	Крижане серце	Frozen	7.4	Cartoon	2013-11-27	102	False
5	Смислони	The Simpsons	8.7	CartoonSeries	1989-12-17	22	True
7	Друзі	Friends	8.9	Series	1994-09-22	22	True
4	Атака титанів	Attack on Titan	9	Anime	2013-04-06	24	True
6	Володар перснів: Хранителі Пел...	The Lord of the Rings: The Fellow...	8.8	Movie	2001-12-19	178	False
8	Віднесені привидами	Spirited Away	8.6	Anime	2001-07-20	125	False
9	Шрек	Shrek	7.9	Cartoon	2001-05-18	90	False
10	Наруто	Naruto	8.4	Anime	2002-10-03	23	True
12	Доктор Хаус	House M.D.	8.7	Series	2004-11-16	44	True
14	Один вдома	Home Alone	7.7	Movie	1990-11-16	103	False
11	Гаррі Поттер і Філософський ка...	Harry Potter and the Sorcerer's S...	7.6	Movie	2001-11-16	152	False
13	Той-реп'єр	Toy Story	8.3	Cartoon	1995-11-22	81	False

00:00:01 Query executed successfully.

Рис. 2.11 - Завантаженні дані у таблицю DimVideo

Microsoft Azure

Synapse Analytics

ipzm241kavcourseworksynapse

Search

ipzm241_kav@student.uzhdu.edu.ua

We use optional cookies to provide a better experience. Learn more

Synapse live

Validate all

Publish all

Accept

Reject

More options

Develop

Filter resources by name

SQL scripts

Analysis of user activity

Analysis of views by month

Analysis of weekend activity

Analyze IMDb ratings for popular v...

Average viewing duration

Create Tables

Load Data

Number of video plays per room

The most popular video by number ...

Viewing duration for each type of co...

data

Number of video pl...

Run

Undo

Publish

Query plan

Connect to ipzm241kavsqldedicatedpool

Use database ipzm241kavsqldedicatedpool

1 SELECT

2 dr.room_name,

3 COUNT(fp.fact_id) AS total_playbacks

4 FROM

5 FactPlayback fp

6 JOIN

7 DimRoom dr ON fp.room_id = dr.room_id

8 GROUP BY

9 dr.room_name

10 ORDER BY

11 total_playbacks DESC;

12

Results

Messages

View

Table

Chart

Export results

Search

room_name

total_playbacks

Cars cartoons

72

Horror Night Specials

59

Public comfort room - watching ...

53

Modern Love Stories

50

We watch the Cars cartoons

31

New Year's Eve movie

29

Kids Cartoon Hour

28

00:00:01 Query executed successfully

Рис. 2.12 - Кількість відтворень різного відеоконтенту по кімнатах

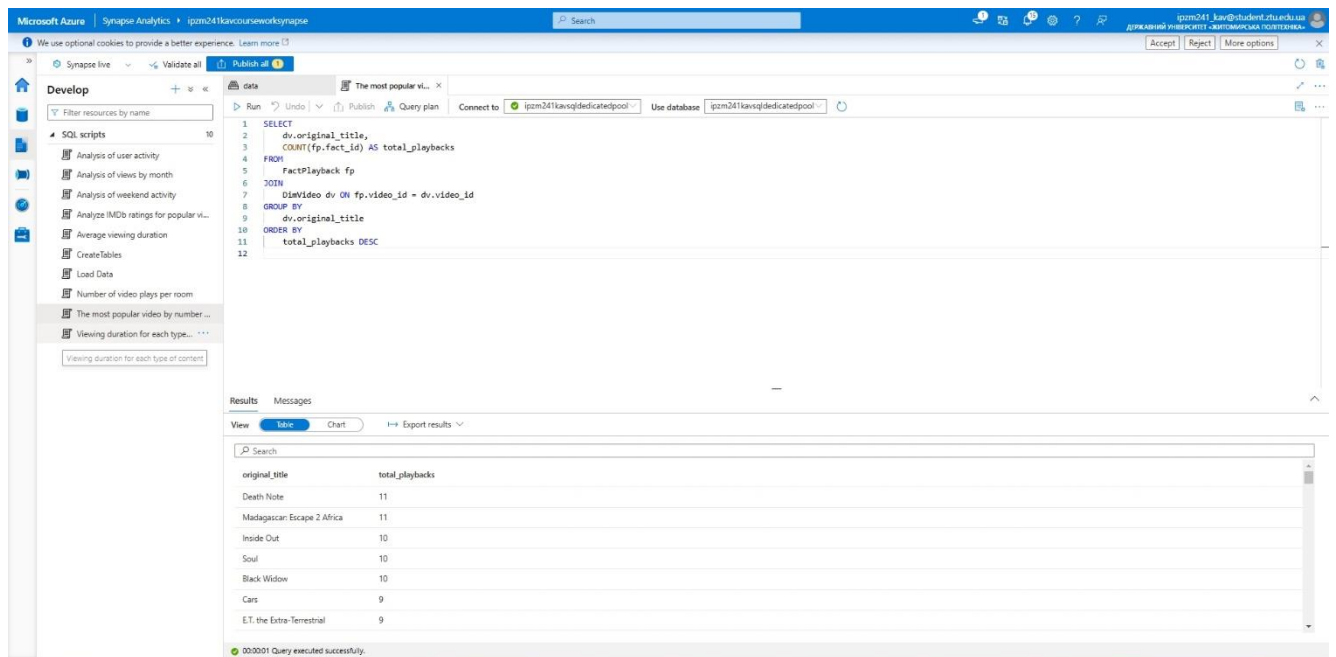


Рис. 2.13 - Найпопулярніший відеоконтент за кількістю відтворень

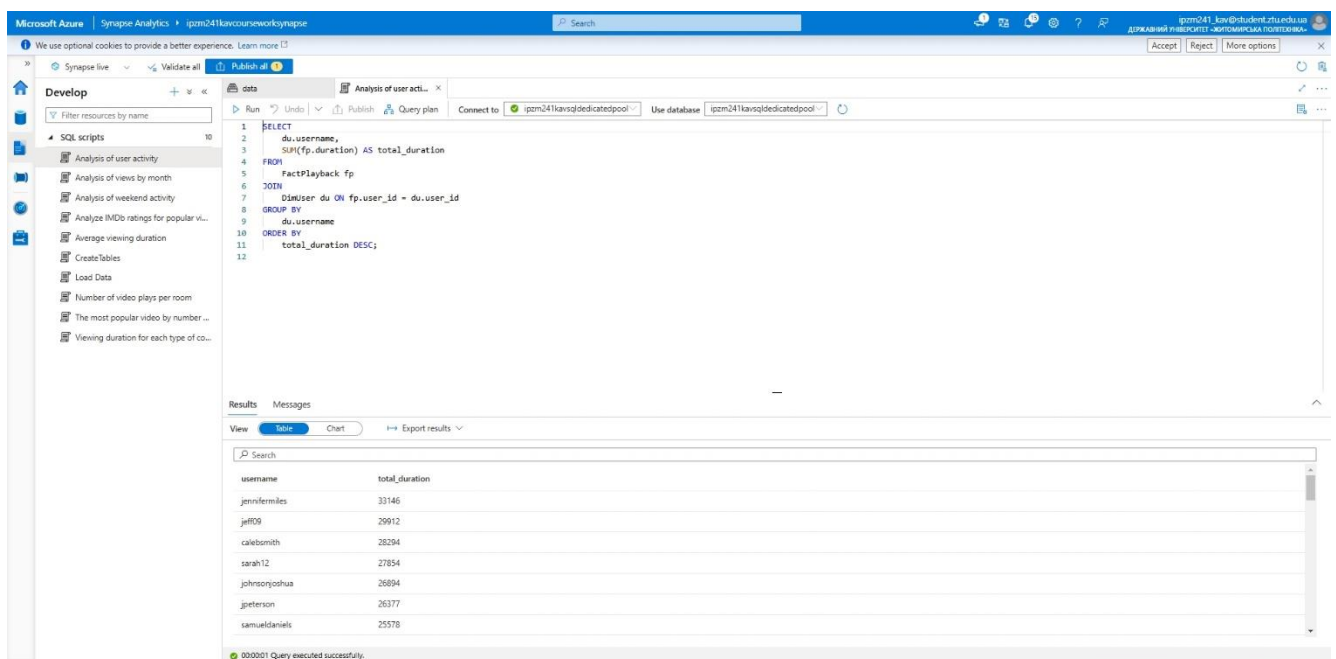


Рис. 2.14 - Аналіз активності користувачів

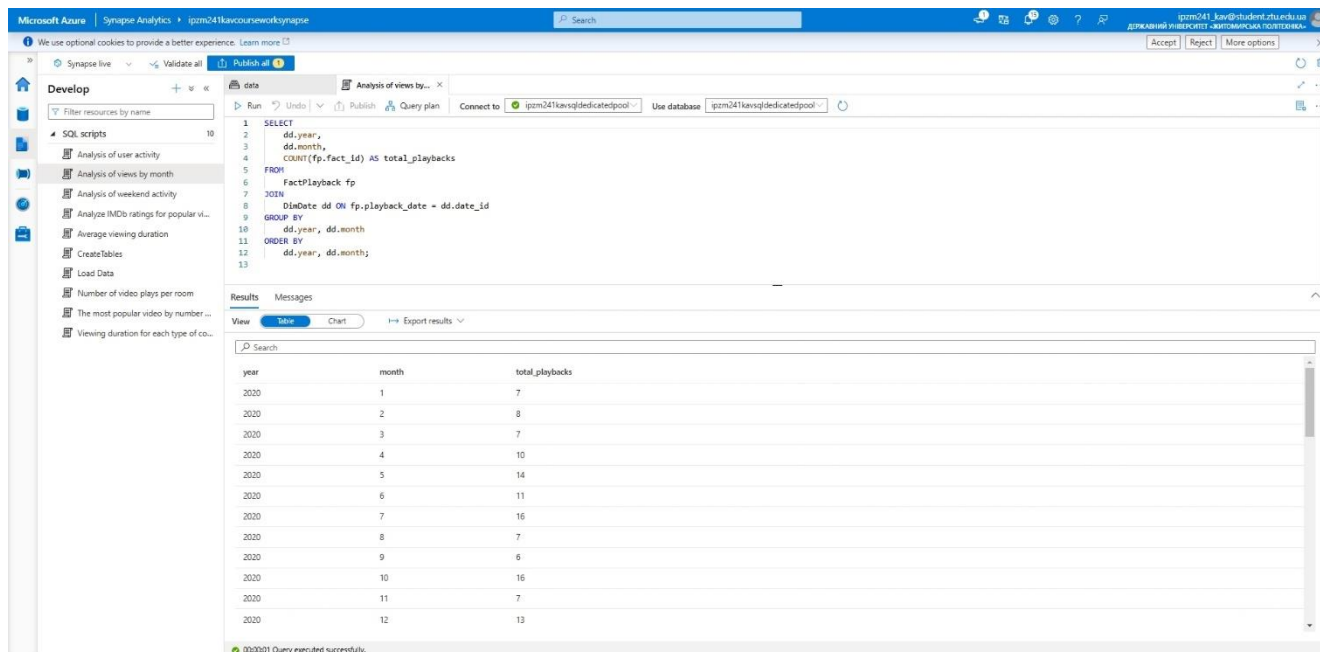


Рис. 2.15 - Аналіз переглядів в кімнатах по місяцях

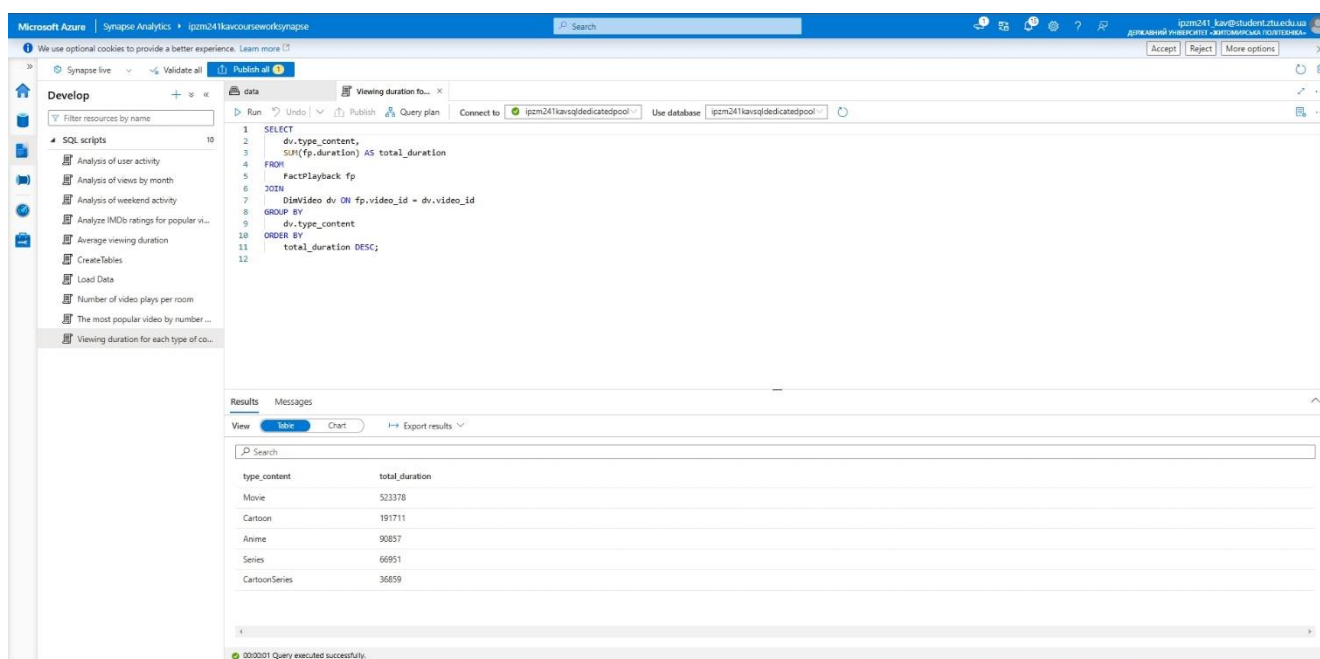


Рис. 2.16 - Тривалість перегляду для кожного типу відеоконтенту

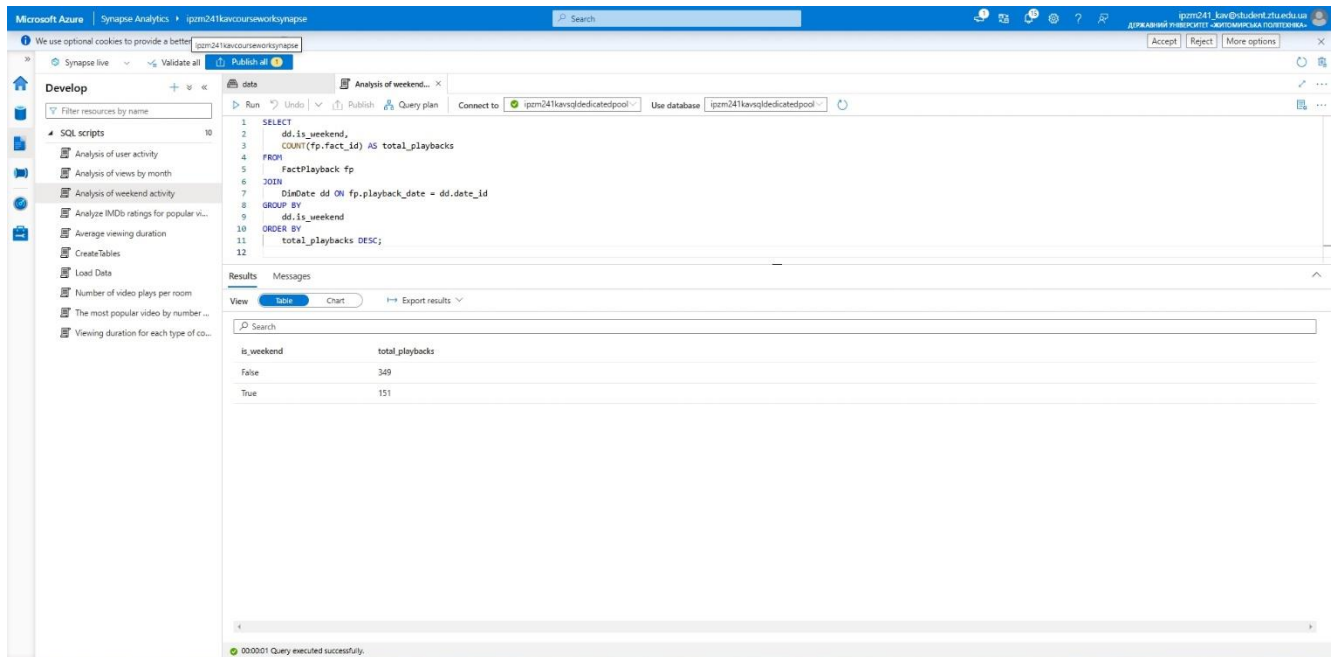


Рис. 2.17 - Аналіз активності кімнат у робочі та вихідні дні

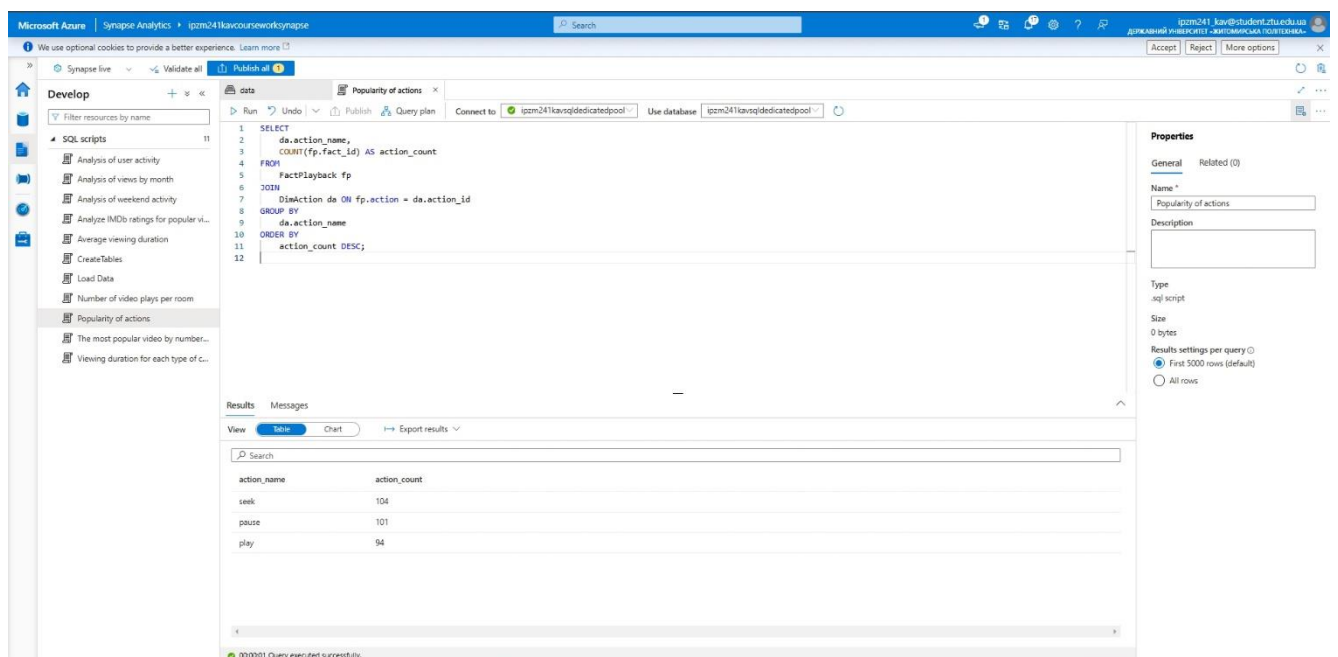


Рис. 2.18 - Популярність дій користувачів в кімнатах

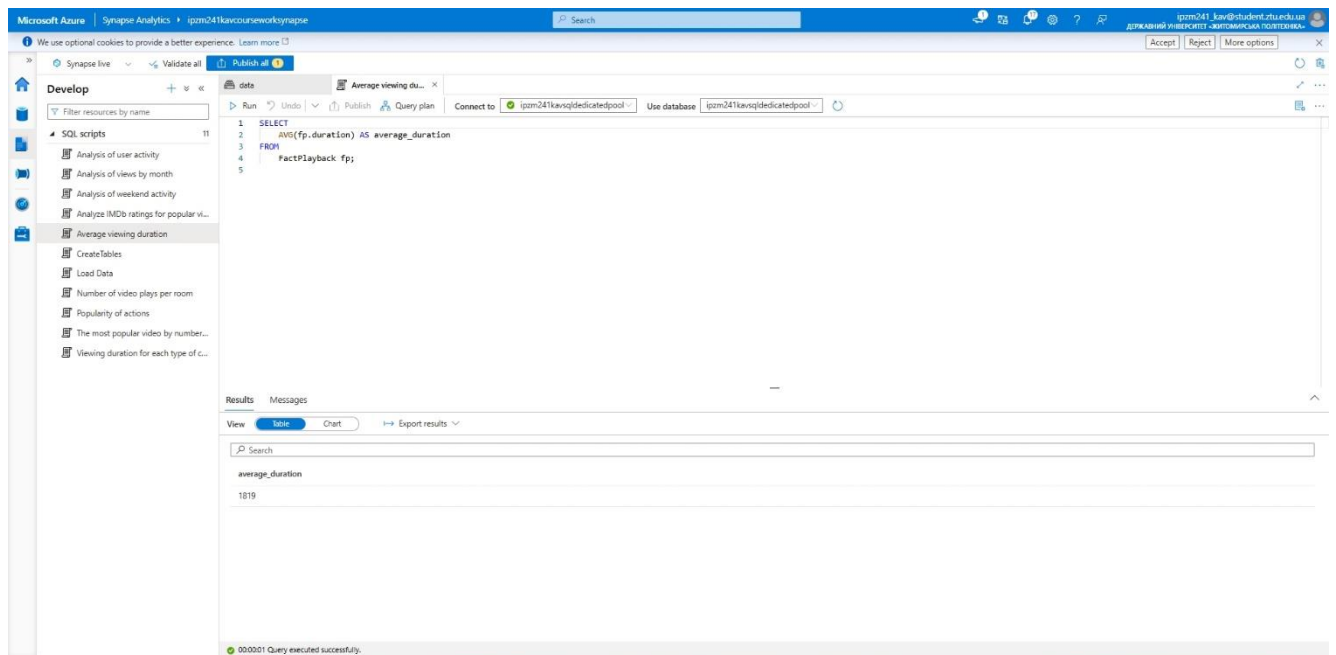


Рис. 2.19 - Середня тривалість перегляду в кімнатах за весь період часу

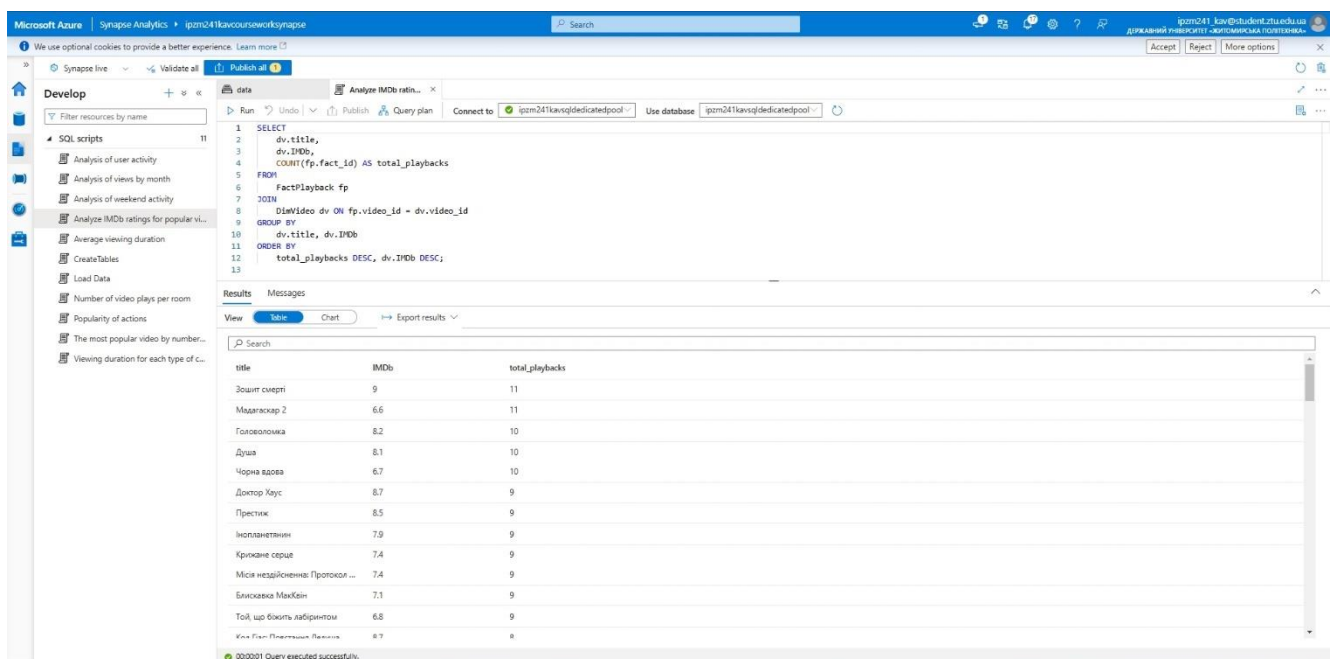


Рис. 2.20 - Аналіз IMDb-рейтингів для популярного відеоконтенту

Висновки до розділу 2

У даному розділі було здійснено проектування транзакційного та аналітичного сховищ даних для платформи спільного кіноперегляду. На основі транзакційного сховища було побудовано OLAP-діаграму у вигляді зіркоподібної схеми, яка забезпечує багатовимірний аналіз даних і полегшує виконання аналітичних запитів.

Для реалізації аналітичного сховища даних було використано платформу Azure Synapse Analytics. Покроково описано етапи створення Azure Storage Account, налаштування робочого простору, створення SQL-пулу, таблиць та завантаження даних у сховище.

Після побудови аналітичного сховища було розроблено дев'ять аналітичних SQL-запитів, спрямованих на аналіз ключових метрик роботи платформи. Всі результати були структуровані для подальшої візуалізації.

		Костюченко А.В.			«Житомирська політехніка» 24.121.14.000	Арк.
						33
Змн.	Арк.	№ докум.	Підпис	Дата		

РОЗДІЛ 3. ОРГАНІЗАЦІЯ ЗАСОБІВ ВІЗУАЛІЗАЦІЇ ТА АНАЛІЗУ ДАНИХ

3.1 Технологія Power BI та її основні можливості

Корисним інструментом для створення звітів і візуалізацій даних, які допомагають зрозуміти інформацію з кількох джерел, є Microsoft Power BI. Він дозволяє перетворювати складні таблиці та цифри на графіки, діаграми та карти, які легко вивчати. Цей інструмент підходить як для новачків, так і для досвідчених експертів. Інтегрований штучний інтелект Power BI дозволяє йому автоматично визначати важливу інформацію у ваших даних, щоб допомогти вам у прийнятті рішень.

Power BI Desktop програма для Windows, яка дозволяє створювати інтерактивні звіти та візуалізацію даних, поєднувати та перетворювати дані з різних джерел у прості для розуміння графіки, діаграми та інформаційні панелі. Power BI Desktop використовується для розробки звітів, які можна опублікувати в службі Power BI для подальшого аналізу та спільного використання.

Power BI Service — це хмарна платформа для публікації, спільного використання та спільної роботи над звітами та інформаційними панелями. Користувачі можуть взаємодіяти з даними в режимі реального часу, ділитися результатами з колегами та отримувати доступ до звітів з будь-якого пристрою.

Power BI також є частиною Microsoft Power Platform, яка включає Power Apps, Power Pages, Power Automate і Power Virtual Agents. Розроблені як «інструменти з низьким рівнем коду», ці програми допомагають компаніям аналізувати та візуалізувати дані, розробляти бізнес-рішення, автоматизувати процеси та створювати чат-ботів без використання коду [9].

3.2 Публікація звіту на прикладній дошці Power BI

Для візуалізації та аналізу даних, представлених у розділі 2, необхідно завантажити ці дані у Power BI. Для цього потрібно підключитися до Azure Synapse Analytics, вибрати створений SQL-пул та виконати авторизацію для забезпечення доступу до даних.

		Костюченко А.В.			«Житомирська політехніка» 24.121.14.000	Арк.
						34
Змн.	Арк.	№ докум.	Підпис	Дата		

На рис.3.1 зображено успішне підключення до Azure Synapse Analytics з вибором аналітичних таблиць.

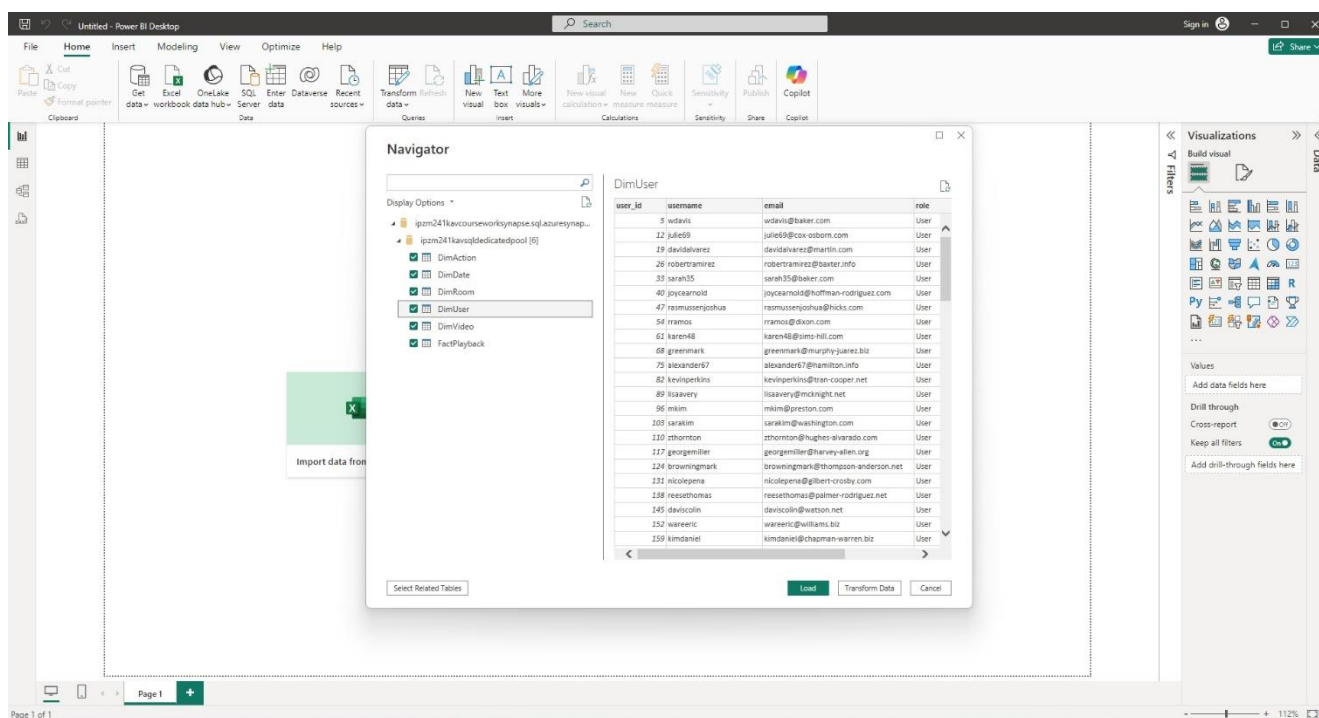


Рис. 3.1 – Вибір аналітичних таблиць для подальшого використання

Після вибору аналітичних таблиць, ці таблиці відображаються у правій частині інтерфейсу робочої панелі (рис. 3.2).

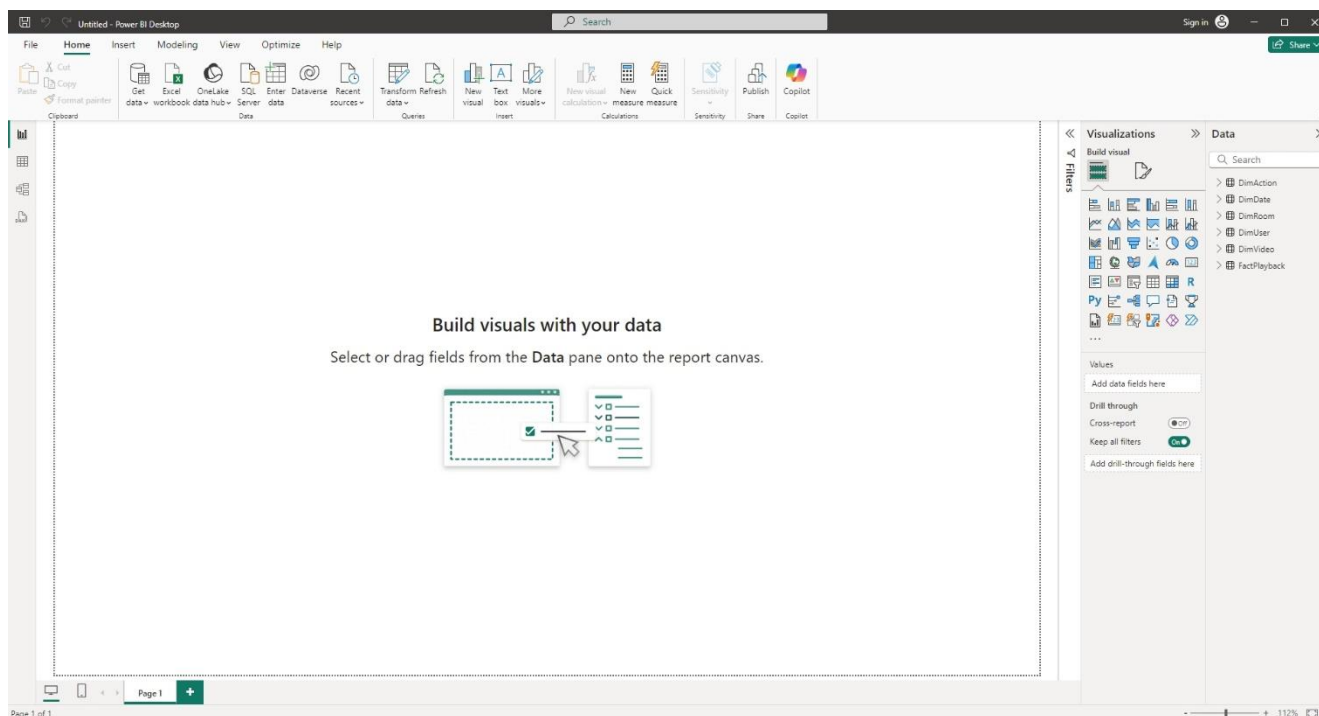


Рис. 3.2 – Таблиці з даними

Після завантаження таблиць із даними необхідно додатково перевірити зв'язки між таблицею фактів та таблицями вимірів, щоб забезпечити коректне застосування динамічних фільтрів для побудови графіків у подальшому аналізі.

На рис.3.3 представлено зв'язки між таблицею фактів та таблицями вимірів.

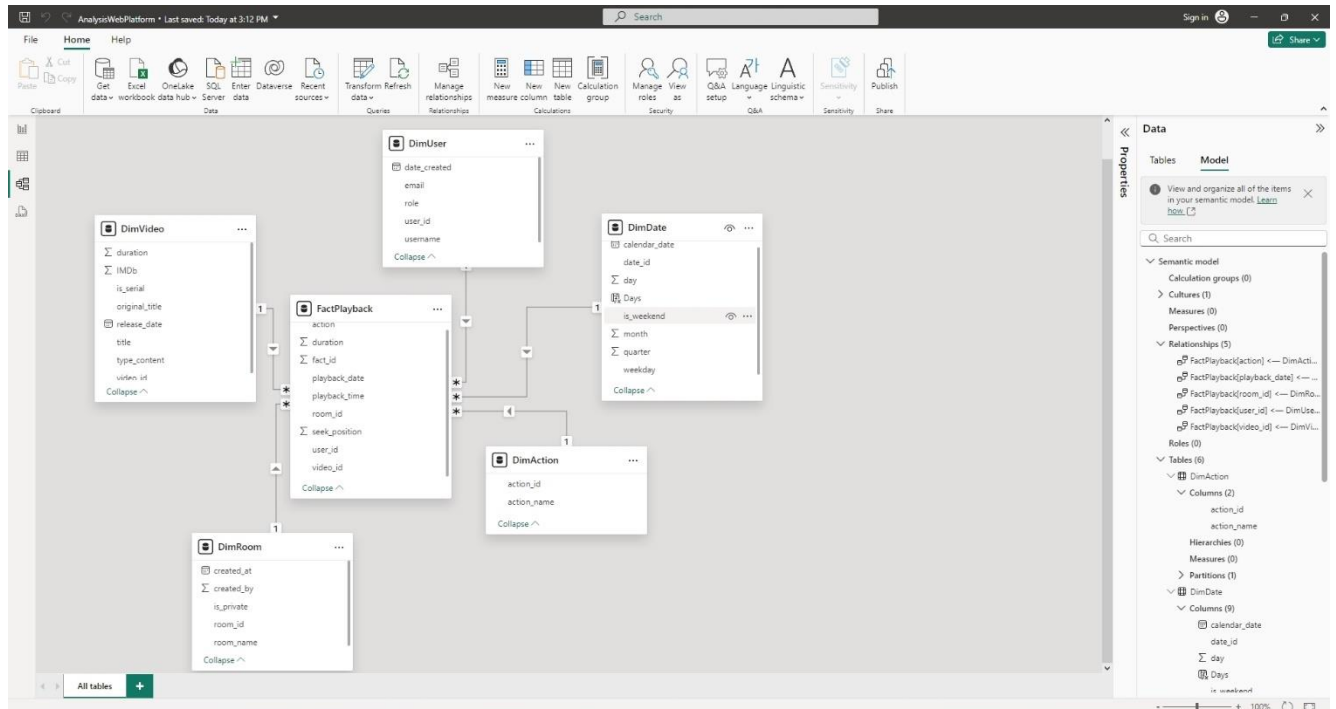


Рис. 3.3 – Зв'язки між таблицями

Створені метрики можна класифікувати на дві основні категорії: метрики активності користувачів та кімнат, які характеризують поведінку користувачів і взаємодію в рамках платформи, та метрики контенту та часу, які аналізують характеристики контенту і часові аспекти його споживання.

Для візуалізації метрик активності користувачів та кімнат були використані стовпчасті та кругові діаграми. Стовпчасті діаграми застосовані для відображення кількості відтворень відеоконтенту по кімнатах (рис.3.4) та аналізу активності користувачів (рис. 3.5). Кругові діаграми використовуються для представлення аналізу активності у робочі та вихідні дні (рис.3.6) та популярності дій користувачів (рис. 3.7).

Додатково було додано новий стовпець (рис.3.8) для кругової діаграми активності у робочі та вихідні дні для зручності відображення легенди з використанням DAX-формули.

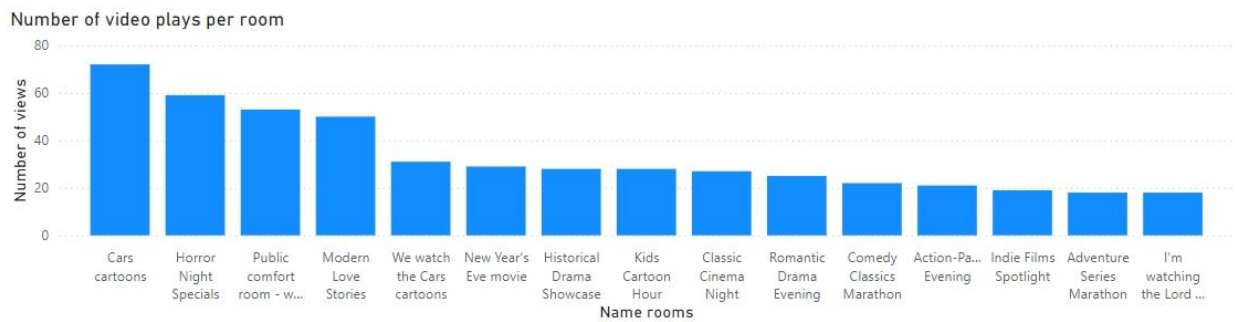


Рис. 3.4 – Кількість відтворень відеоконтенту по кімнатах

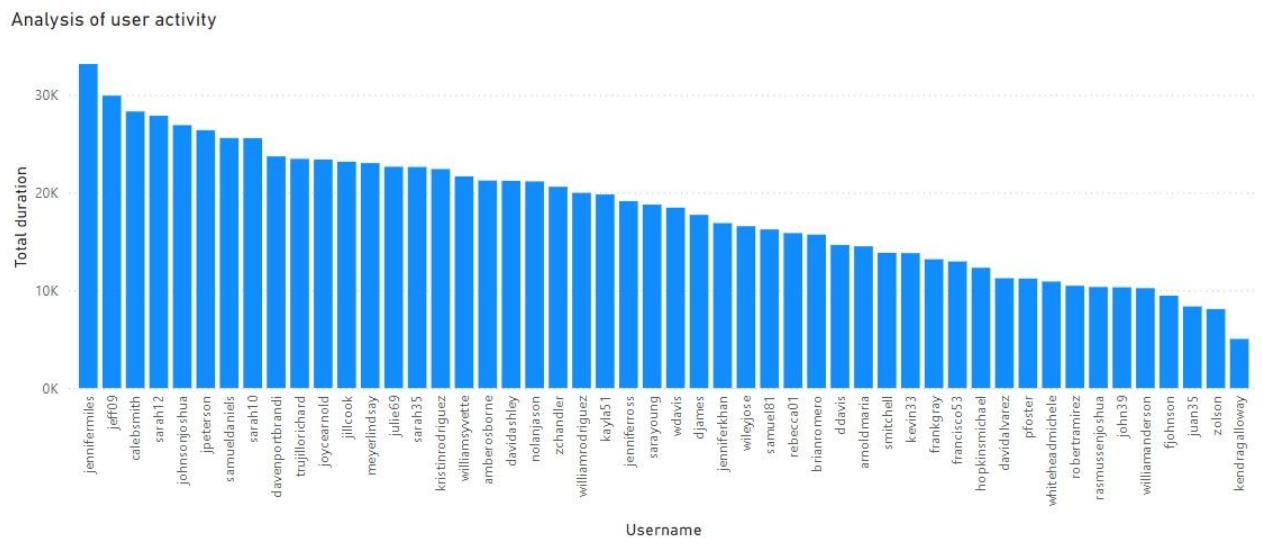


Рис. 3.5 – Аналіз активності користувачів

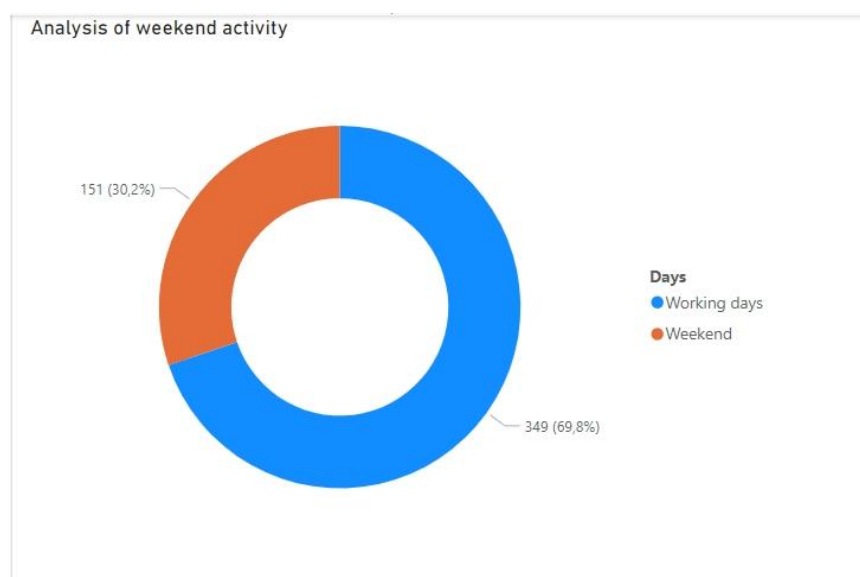


Рис. 3.6 – Аналіз активності у робочі та вихідні дні

A pie chart illustrating the distribution of three actions: seek, pause, and play. The chart is divided into three segments: a blue segment for 'seek' (104 occurrences, 34.78%), a dark blue segment for 'pause' (101 occurrences, 33.78%), and an orange segment for 'play' (94 occurrences, 31.44%). A legend on the right side of the chart identifies the colors for each action name.

Action name	Count	Percentage
seek	104	34,78%
pause	101	33,78%
play	94	31,44%

AnalysisWebPlatform • Last saved: Today at 3:12 PM

Search

File Home Insert Modeling View Optimize Help Table tools Column tools

Name Days

Date type Text

Format Text

\$ % 9 0 Auto

Summarization Don't summarize

Data category Uncategorized

Sort by column

Data groups

Manage relationships

New column

Structure

Formatting

Properties

1 Days = SWITCH(

2 TRUE(),

3 DimDate[is_weekend] = TRUE, "weekend",

4 DimDate[is_weekend] = FALSE, "working days"

5)

6)

7

151 (30.2%)

Days

Working days

Weekend

349 (69.8%)

Add data fields here

Filters on all pages

Add data fields here

Visualizations

Build visual

calendar_date

date_id

day

Days

is_weekend

month

quarter

weekday

year

DimRoom

DimUser

DimVideo

duration

IMDo

is_serial

original_title

release_date

title

type_content

video_id

Values

Add data fields here

Drill through

Cross-report

Keep all filters

Add drill-through fields here

Page 1

Page 2

Page 3

Page 4

Page 5

Page 6

Page 7

Page 8

Page 6 of 8

103%

Для візуалізації метрик контенту та часу були використані стовпчасті, лінійні та кругові діаграми. Стовпчасті діаграми застосовані для відображення найпопулярнішого відеоконтенту за кількістю відтворень (рис.3.9) та аналізу IMDb-рейтингів для популярного відеоконтенту (рис.3.10). Лінійна діаграма застосована для аналізу переглядів по місяцях (рис.3.11). Кругова діаграма використовуються для представлення тривалості перегляду для кожного типу відеоконтенту (рис.3.12).

The most popular video by number of plays

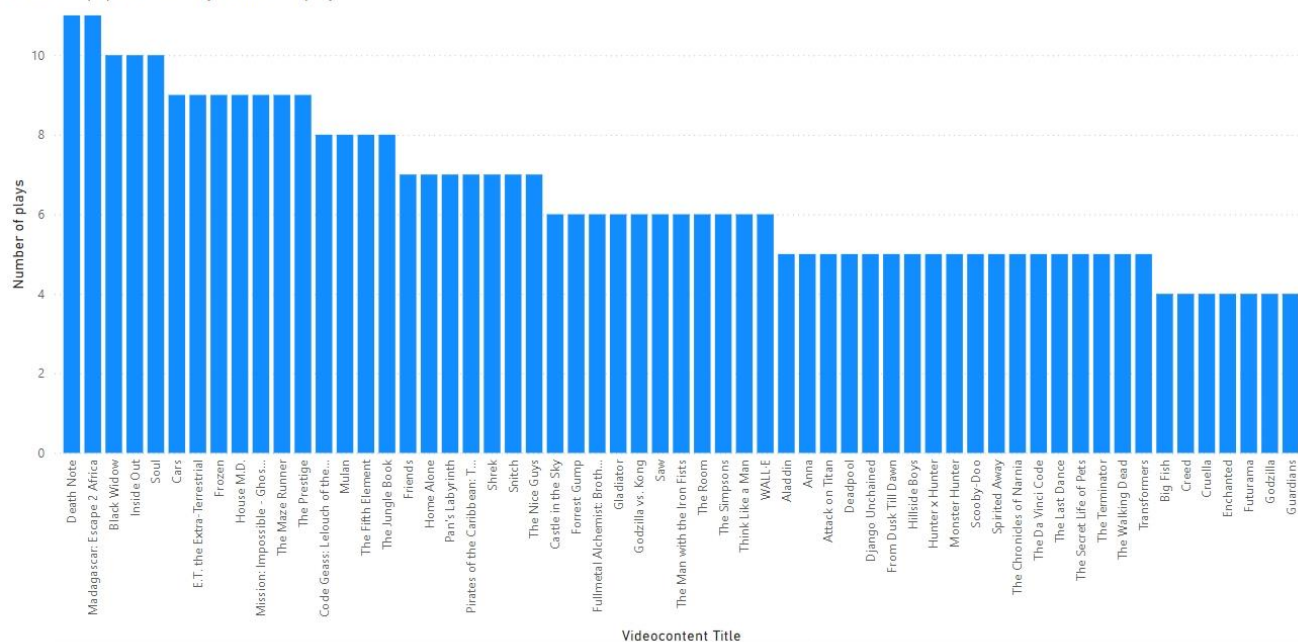


Рис. 3.9 – Найпопулярніший відеоконтент за кількістю відтворень

Analyze IMDb ratings for popular videos

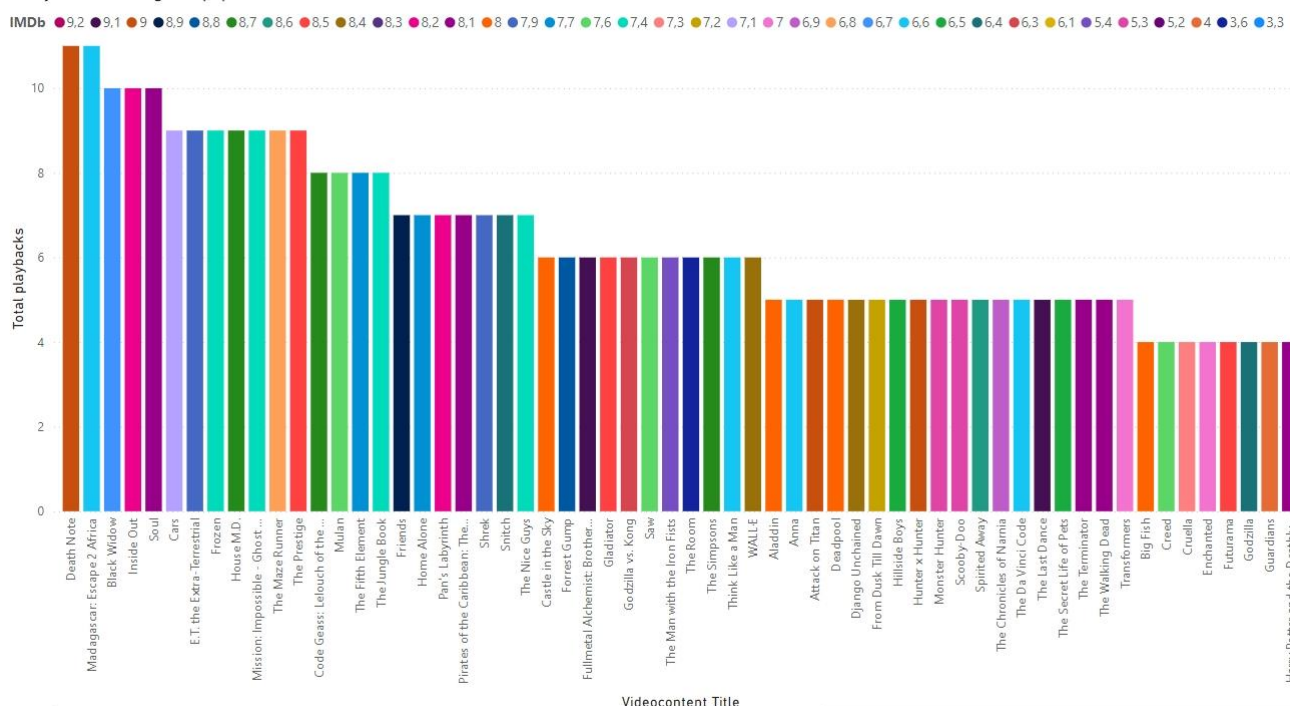


Рис. 3.10 – Аналіз IMDb-рейтингів для популярного відеоконтенту

Висновки до розділу 3

У даному розділі було здійснено візуалізацію та аналіз даних з Azure Synapse Analytics за допомогою платформи Power BI. Створено різноманітні діаграми для представлення метрик активності користувачів і кімнат, а також контенту і час.

Розроблені візуалізації сприяють отриманню більш глибокого розуміння функціонування платформи, дозволяючи виявляти тенденції, закономірності та потенційні проблемні області, що забезпечує можливість оцінювати загальну ефективність бізнес-процесів платформи, а також приймати зважені рішення на основі аналітичних даних.

Отримані результати дозволяють оперативно реагувати на проблеми, адаптувати бізнес-процеси до змін і нових вимог, а також розробляти стратегії для подальшого вдосконалення функціональності платформи та підвищення конкурентоспроможності.

		Костюченко А.В.			«Житомирська політехніка» 24.121.14.000	Арк.
						41
Змн.	Арк.	№ докум.	Підпис	Дата		

ВИСНОВКИ

У курсовій роботі було розглянуто та реалізовано інформаційні процеси платформи для спільного кіноперегляду, включаючи моделювання, проектування сховищ даних, розробку аналітичних запитів і візуалізацію результатів. На етапі моделювання було сформульовано та структуровано завдання, які необхідно виконати для забезпечення функціональності платформи. Проведено аналіз і обґрунтування застосування різних методів і нотацій для створення моделей DFD, IDEF0 і BPMN 2.0, що дозволило описати основні інформаційні процеси системи.

У рамках проектування сховищ даних було створено транзакційне та аналітичне сховища, побудовано OLAP-діаграму у вигляді зіркоподібної схеми для багатовимірного аналізу даних. З використанням платформи Azure Synapse Analytics реалізовано всі етапи створення аналітичного сховища, включаючи налаштування робочого простору, створення SQL-пулу, таблиць і завантаження даних.

Для забезпечення глибокого аналізу платформи розроблено аналітичні SQL-запити, що дозволяють досліджувати ключові метрики, такі як активність користувачів і кімнат, характеристики контенту та часові аспекти його споживання. Ці дані були використані для створення візуалізацій у Power BI для забезпечення зрозумілого та наочного представлення результатів аналізу.

Отримані результати дозволяють глибше зрозуміти роботу платформи, оцінити ефективність бізнес-процесів, визначити проблемні зони та сформулювати стратегії їх усунення. Проведені таким чином дослідження у курсовій роботі створюють основу для оптимізації роботи платформи, покращення її продуктивності та адаптації до змін у потребах користувачів і ринкових умов.

		Костюченко А.В.			«Житомирська політехніка» 24.121.14.000	Арк.
						42
Змн.	Арк.	№ докум.	Підпис	Дата		

ПЕРЕЛІК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. What is DFD(Data Flow Diagram) [Електронний ресурс]. – 2024. – Режим доступу до ресурсу: <https://www.geeksforgeeks.org/what-is-dfddata-flow-diagram/>.
2. IDEF0 Diagram [Електронний ресурс] – Режим доступу до ресурсу: <https://vitechcorp.com/resources/GENESYS/onlinehelp/desktop/Views/IDEF0.htm>.
3. What is BPMN [Електронний ресурс] – Режим доступу до ресурсу: <https://www.visual-paradigm.com/guide/bpmn/what-is-bpmn/>.
4. What Is OLTP [Електронний ресурс]. – 2023. – Режим доступу до ресурсу: <https://www.oracle.com/database/what-is-oltp/>.
5. OLAP (online analytical processing) [Електронний ресурс]. – 2023. – Режим доступу до ресурсу: <https://www.techtarget.com/searchdatamanagement/definition/OLAP>.
6. Difference Between OLAP and OLTP in DBMS [Електронний ресурс]. – 2024. – Режим доступу до ресурсу: <https://www.geeksforgeeks.org/difference-between-olap-and-oltp-in-dbms/>.
7. Azure Synapse Analytics [Електронний ресурс]. – 2023. – Режим доступу до ресурсу: <https://datascientest.com/en/azure-synapse-analytics-everything-you-need-to-know-about-microsofts-cloud-analytics-platform>.
8. What Is Power BI [Електронний ресурс]. – 2024. – Режим доступу до ресурсу: <https://www.coursera.org/articles/what-is-power-bi>.
9. Mastering Azure Analytics: Architecting in the Cloud with Azure Data Lake, HDInsight, and Spark 1st Edition, 2017. – 412 с.
10. Mastering Microsoft Power BI: Expert techniques for effective data analytics and business intelligence, 2018. – 638 с.

		Костюченко А.В.			«Житомирська політехніка» 24.121.14.000	Арк.
						43
Змн.	Арк.	№ докум.	Підпис	Дата		

ДОДАТКИ

		Костюченко А.В.			«Житомирська політехніка» 24.121.14.000	Арк.
						44
Змн.	Арк.	№ докум.	Підпис	Дата		

SQL скрипт створення таблиць

```
CREATE TABLE FactPlayback
(
    fact_id INT IDENTITY NOT NULL,
    room_id INT NOT NULL,
    video_id INT NOT NULL,
    user_id INT NOT NULL,
    playback_date INT NOT NULL,
    playback_time TIME NOT NULL,
    duration INT NOT NULL,
    [action] INT NOT NULL,
    seek_position INT NOT NULL
)
WITH
(
    DISTRIBUTION = REPLICATE,
    CLUSTERED COLUMNSTORE INDEX
);

CREATE TABLE DimUser
(
    user_id INT IDENTITY NOT NULL,
    username NVARCHAR(50) NOT NULL,
    email NVARCHAR(100) NOT NULL,
    role NVARCHAR(20) NOT NULL,
    date_created DATETIME NOT NULL
)
WITH
(
    DISTRIBUTION = REPLICATE,
    CLUSTERED COLUMNSTORE INDEX
);

CREATE TABLE DimRoom
(
    room_id INT IDENTITY NOT NULL,
    room_name NVARCHAR(100) NOT NULL,
    is_private BIT NOT NULL,
    created_at DATETIME NOT NULL,
```

```

        created_by INT NOT NULL
    )
WITH
(
    DISTRIBUTION = REPLICATE,
    CLUSTERED COLUMNSTORE INDEX
);

CREATE TABLE DimVideo
(
    video_id INT IDENTITY NOT NULL,
    title NVARCHAR(200) NOT NULL,
    original_title NVARCHAR(200),
    IMDb FLOAT NOT NULL,
    type_content NVARCHAR(50) NOT NULL,
    release_date DATE NOT NULL,
    duration INT NOT NULL,
    is_serial BIT NOT NULL
)
WITH
(
    DISTRIBUTION = REPLICATE,
    CLUSTERED COLUMNSTORE INDEX
);

CREATE TABLE DimDate
(
    date_id INT IDENTITY NOT NULL,
    calendar_date DATE NOT NULL,
    year INT NOT NULL,
    quarter INT NOT NULL,
    month INT NOT NULL,
    day INT NOT NULL,
    weekday NVARCHAR(20) NOT NULL,
    is_weekend BIT NOT NULL
)
WITH
(
    DISTRIBUTION = REPLICATE,
    CLUSTERED COLUMNSTORE INDEX

```

		Костюченко А.В.			«Житомирська політехніка» 24.121.14.000	Арк.
						46
Змн.	Арк.	№ докум.	Підпис	Дата		

```
);

CREATE TABLE DimAction
(
    action_id INT IDENTITY NOT NULL,
    action_name NVARCHAR(50) NOT NULL
)
WITH
(
    DISTRIBUTION = REPLICATE,
    CLUSTERED COLUMNSTORE INDEX
);
```

SQL скрипт завантаження даних в таблиці

```
COPY INTO FactPlayback
(
    fact_id,
    room_id,
    video_id,
    user_id,
    playback_date,
    playback_time,
    duration,
    [action],
    seek_position
)
FROM 'https://ipzm241kavstorage.dfs.core.windows.net/data/playbacks/playbacks.csv'
WITH
(
    FILE_TYPE = 'CSV',
    MAXERRORS = 0,
    IDENTITY_INSERT = 'ON',
    FIRSTROW = 2
);

COPY INTO DimUser
(
    user_id,
    username,
```

		Костюченко А.В.			«Житомирська політехніка» 24.121.14.000	Арк.
						47
Змн.	Арк.	№ докум.	Підпис	Дата		

```

email,
role,
date_created
)
FROM 'https://ipzm241kavstorage.dfs.core.windows.net/data/users/users.csv'
WITH
(
    FILE_TYPE = 'CSV',
    MAXERRORS = 0,
    IDENTITY_INSERT = 'ON',
    FIRSTROW = 2
);

COPY INTO DimRoom
(
    room_id,
    room_name,
    is_private,
    created_at,
    created_by
)
FROM 'https://ipzm241kavstorage.dfs.core.windows.net/data/rooms/rooms.csv'
WITH
(
    FILE_TYPE = 'CSV',
    MAXERRORS = 0,
    IDENTITY_INSERT = 'ON',
    FIRSTROW = 2
);

COPY INTO DimVideo
(
    video_id,
    title,
    original_title,
    IMDb,
    type_content,
    release_date,
    duration,

```

		Костюченко А.В.			«Житомирська політехніка» 24.121.14.000	Арк.
						48
Змн.	Арк.	№ докум.	Підпис	Дата		


```

    is_serial
)
FROM 'https://ipzm241kavstorage.dfs.core.windows.net/data/videos/videos.csv'
WITH
(
    FILE_TYPE = 'CSV',
    MAXERRORS = 0,
    IDENTITY_INSERT = 'ON',
    FIRSTROW = 2
);

COPY INTO DimDate
(
    date_id,
    calendar_date,
    year,
    quarter,
    month,
    day,
    weekday,
    is_weekend
)
FROM 'https://ipzm241kavstorage.dfs.core.windows.net/data/dates/dates.csv'
WITH
(
    FILE_TYPE = 'CSV',
    MAXERRORS = 0,
    IDENTITY_INSERT = 'ON',
    FIRSTROW = 2
);

COPY INTO DimAction
(
    action_id,
    action_name
)
FROM 'https://ipzm241kavstorage.dfs.core.windows.net/data/actions/actions.csv'
WITH
(

```

		Костюченко А.В.			«Житомирська політехніка» 24.121.14.000	Арк.
						49
Змн.	Арк.	№ докум.	Підпис	Дата		

```
FILE_TYPE = 'CSV',
MAXERRORS = 0,
IDENTITY_INSERT = 'ON',
FIRSTROW = 2
);
```