

Міністерство освіти і науки України
Національний університет «Львівська політехніка»

Інститут прикладної математики та фундаментальних наук
Кафедра прикладної математики



Курсова робота

з дисципліни «Робота з великими базами даних»

Виконала : студентка
групи ПМ-43 Кос
В.П.

Перевірив:
Любінський Б.Б.

Львів 2026

АНОТАЦІЯ

У даній курсовій роботі розглянуто процес проектування та реалізації інформаційно-аналітичної системи для сфери кабельного телебачення з використанням сучасних технологій керування базами даних і бізнес-аналітики. Основна увага приділяється побудові операційної бази даних, сховища даних (Data Warehouse), реалізації ETL-процесів, створенню OLAP-куба та розробці аналітичних звітів.

У роботі виконано аналіз предметної області, визначено ключові бізнес-процеси та інформаційні потреби користувачів. На основі цього спроектовано структуру бази даних із дотриманням принципів нормалізації та цілісності даних. Для забезпечення аналітичних задач реалізовано схему сховища даних типу «зірка», що дозволяє ефективно виконувати агрегації та багатовимірний аналіз.

Практична частина курсової роботи включає генерацію великих обсягів реалістичних даних, налаштування ETL-процесів за допомогою SQL Server Integration Services, побудову OLAP-куба в SQL Server Analysis Services та створення аналітичних звітів у SQL Server Reporting Services. Отримана система дозволяє здійснювати аналіз популярності кінофільмів, фінансових показників, поведінки абонентів і динаміки абонентської бази.

Результати роботи можуть бути використані як основа для побудови повноцінних інформаційно-аналітичних систем у сфері телекомунікацій та цифрових сервісів

ABSTRACT

This course paper examines the process of designing and implementing an information and analytical system for the cable television industry using modern database management and business analytics technologies. The main focus is on building an operational database, a data warehouse, implementing ETL processes, creating an OLAP cube, and developing analytical reports.

The work analyzes the subject area, identifies key business processes and user information needs. Based on this, a database structure was designed in accordance with the principles of normalization and data integrity. To ensure analytical tasks, a star-type data warehouse scheme was implemented, which allows for effective aggregation and multidimensional analysis.

The practical part of the coursework includes the generation of large volumes of realistic data, the configuration of ETL processes using SQL Server Integration Services, the construction of an OLAP cube in SQL Server Analysis Services, and the creation of analytical reports in SQL Server Reporting Services. The resulting system allows for the analysis of movie popularity, financial indicators, subscriber behavior, and subscriber base dynamics.

The results of the work can be used as a basis for building full-fledged information and analytical systems in the field of telecommunications and digital services.

ЗМІСТ :

РОЗДІЛ 1 : АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ.....	8
1.1 Загальна характеристика предметної області.....	8
1.2 Опис основних сутностей предметної області.....	8
1.3 Основні бізнес-процеси предметної області.....	10
1.4 Інформаційні потреби користувачів системи.....	11
1.5 Функціональні вимоги до системи.....	11
1.6 Бізнес-правила та обмеження.....	12
1.7 Висновки до розділу.....	12
РОЗДІЛ 2 : ПРОЄКТУВАННЯ БАЗИ ДАНИХ.....	13
2.1 Мета та етапи проєктування бази даних.....	13
2.2 Концептуальне проєктування бази даних.....	13
2.3 Логічне проєктування та нормалізація даних.....	14
2.4 Проєктування операційної бази даних (OLTP).....	15
2.5 Обґрунтування використання сховища даних.....	16
2.6 Проєктування сховища даних за схемою «зірка».....	16
2.7 Проєктування ключів та використання сурогатних ідентифікаторів... 16	
2.9 Фізичне проєктування бази даних.....	17
2.10 Проєктування індексів та оптимізація продуктивності.....	18
2.11 Забезпечення цілісності та контролю якості даних.....	18
2.12 Висновки до розділу.....	19
Розділ 3: Реалізація ETL-процесів (SQL Server Integration Services).....	20
3.1 Загальна характеристика ETL-процесів.....	20
3.2 Створення та налаштування SSIS-проєкту.....	20
3.3 Проєктування сховища даних для ETL.....	21
3.4 Витягування даних (Extract).....	21
3.5 Трансформація даних (Transform).....	22
3.5.1 Очищення даних (Data Cleansing).....	22
3.5.2 Конвертація типів даних (Data Conversion).....	22
3.5.3 Похідні колонки (Derived Column).....	22
3.5.4 Пошук відповідностей (Lookup).....	23
3.5.5 Умовний розподіл даних (Conditional Split).....	23
3.6 Завантаження даних (Load).....	23

3.7 Контроль якості та тестування ETL-процесів.....	24
РОЗДІЛ 4 : ПОБУДОВА OLAP-КУБА ТА АНАЛІТИЧНІ ЗВІТИ.....	25
4.1 Призначення OLAP-технологій в інформаційно-аналітичних системах.....	25
4.2 Архітектура аналітичної підсистеми.....	25
4.3 Проєктування OLAP-куба.....	26
4.4 Виміри OLAP-куба.....	26
4.5 Міри та групи мір.....	27
4.6 Обробка та оптимізація OLAP-куба.....	28
4.7 Аналітичні звіти як завершальний рівень системи.....	28
4.8 Типи аналітичних звітів та їх призначення.....	28
4.9 Параметризація та інтерактивність аналітичних звітів.....	29
4.10 Практичне використання OLAP-куба та звітів.....	30
4.11 Інтеграція аналітичної підсистеми в загальну архітектуру.....	31
4.12 Переваги використання OLAP і звітів у курсовій роботі.....	31
4.13 Висновки до розділу 4.....	31
ВИСНОВКИ.....	33
СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ :	35

ВСТУП :

В умовах стрімкого розвитку інформаційних технологій та цифрових сервісів особливої актуальності набуває питання ефективного зберігання, обробки та аналізу великих обсягів даних. Компанії, що працюють у сфері кабельного телебачення, щоденно оперують значною кількістю інформації, пов'язаної з абонентами, контентом, фінансовими операціями та підключеннями каналів. Для прийняття обґрунтованих управлінських рішень необхідні сучасні інструменти аналітики, які дозволяють перетворювати великі масиви даних у корисну бізнес-інформацію.

Традиційні операційні бази даних (OLTP) орієнтовані переважно на швидке виконання транзакцій та зберігання актуального стану даних, однак вони є недостатньо ефективними для складного аналітичного аналізу, побудови звітів та прогнозування. Саме тому для підтримки аналітичних задач широко застосовуються сховища даних (Data Warehouse), OLAP-технології та спеціалізовані інструменти бізнес-аналітики.

Актуальність даної курсової роботи зумовлена необхідністю комплексного підходу до проєктування інформаційно-аналітичних систем, які охоплюють усі етапи роботи з даними: від збору та очищення до багатовимірного аналізу і візуалізації результатів. Використання таких технологій, як SQL Server Integration Services, SQL Server Analysis Services та SQL Server Reporting Services, дозволяє створити ефективну та масштабовану систему аналітики.

Метою курсової роботи є проєктування та реалізація інформаційно-аналітичної системи для сфери кабельного телебачення на основі сховища даних та OLAP-технологій, яка забезпечує аналіз ключових показників діяльності компанії.

Для досягнення поставленої мети в роботі необхідно вирішити такі завдання:

- виконати аналіз предметної області та визначити основні бізнес-процеси;
- спроектувати структуру операційної бази даних та сховища даних;
- реалізувати генерацію великих обсягів реалістичних тестових даних;
- налаштувати ETL-процеси для завантаження даних у сховище;
- побудувати OLAP-куб для багатовимірного аналізу;
- розробити аналітичні звіти для підтримки управлінських рішень.

Об'єктом дослідження є процеси зберігання та аналізу даних у системі кабельного телебачення.

Предметом дослідження є методи та засоби проектування баз даних, сховищ даних, ETL-процесів і аналітичних звітів.

РОЗДІЛ 1 : АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ

1.1 Загальна характеристика предметної області

Предметною областю даної курсової роботи є діяльність підприємства у сфері кабельного телебачення, яке надає послуги доступу до телевізійного контенту для кінцевих споживачів — абонентів. Такі підприємства працюють у середовищі з високою конкуренцією, значною кількістю клієнтів та великими обсягами інформації, що постійно генерується в процесі їхньої діяльності.

Основними напрямками роботи компанії кабельного телебачення є підключення та обслуговування абонентів, управління каталогом кінофільмів і телевізійних каналів, обробка замовлень на перегляд контенту, формування рахунків, приймання платежів і контроль за фінансовими показниками. Кожен із цих напрямів супроводжується створенням та обробкою значних масивів даних, які необхідно не лише зберігати, а й аналізувати для підтримки управлінських рішень.

У сучасних умовах ефективна діяльність таких компаній неможлива без використання інформаційно-аналітичних систем, які дозволяють оперативно отримувати узагальнену інформацію про популярність контенту, фінансові результати, поведінку абонентів та динаміку розвитку клієнтської бази. Саме тому актуальним є впровадження сховищ даних та OLAP-технологій, що забезпечують багатовимірний аналіз і підтримку стратегічного планування.

1.2 Опис основних сутностей предметної області

У процесі аналізу предметної області було визначено ключові сутності, які характеризують діяльність компанії кабельного телебачення.

Абонент є основною сутністю системи. Абонент характеризується унікальним ідентифікатором, персональними даними, датою підключення, статусом обслуговування, а також історією змін своїх характеристик. Дані про абонентів використовуються для аналізу клієнтської бази, сегментації споживачів та оцінки динаміки підключень і відключень.

Кінофільм є одиницею контенту, що пропонується абонентам. Для кожного кінофільму зберігається інформація про назву, жанр, рік випуску та інші атрибути. Аналіз даних про фільми дозволяє визначати популярність контенту, виявляти тренди та формувати рекомендації.

Замовлення на перегляд фільму відображає факт споживання контенту абонентом. Воно пов'язує абонента, фільм та дату замовлення і містить кількісні та вартісні показники. Саме на основі цих даних формується аналітика щодо популярності фільмів і споживчої орієнтації клієнтів.

Рахунок (інвойс) є фінансовою сутністю, яка відображає зобов'язання абонента щодо оплати наданих послуг. Рахунок містить загальну суму, суму оплачених коштів та заборгованість. Аналіз рахунків дозволяє оцінювати фінансовий стан компанії та платіжну дисципліну абонентів.

Платіж відображає факт надходження коштів від абонента. Платежі пов'язані з рахунками та використовуються для аналізу грошових потоків і контролю заборгованостей.

Група каналів є додатковою сутністю, що дозволяє класифікувати телевізійний контент за тематикою або типом пакету.

1.3 Основні бізнес-процеси предметної області

На основі аналізу діяльності компанії кабельного телебачення було виділено такі основні бізнес-процеси.

Першим бізнес-процесом є **реєстрація та обслуговування абонентів**. Він включає підключення нових клієнтів, зміну їхніх параметрів, оновлення статусів та зберігання історії змін. Цей процес є критично важливим для аналізу динаміки абонентської бази.

Другим бізнес-процесом є **управління контентом**, що охоплює додавання нових фільмів, класифікацію їх за жанрами та групами, а також відстеження їхньої популярності серед абонентів.

Третім бізнес-процесом є **оформлення замовлень на перегляд контенту**. У межах цього процесу фіксується інформація про те, який абонент, який фільм і коли переглядав. Дані цього процесу є основою для аналітики споживчих уподобань.

Четвертим бізнес-процесом є **формування рахунків та облік платежів**. Він включає розрахунок вартості послуг, контроль оплат та заборгованостей, а також аналіз фінансових показників.

П'ятим бізнес-процесом є **аналітична обробка даних**, яка не впливає безпосередньо на операційну діяльність, але є необхідною для прийняття управлінських рішень.

1.4 Інформаційні потреби користувачів системи

Користувачами інформаційно-аналітичної системи є керівники компанії, аналітики та менеджери. Для них важливо мати доступ до узагальненої та деталізованої інформації у зручному вигляді.

Основними інформаційними потребами є:

- аналіз популярності кінофільмів і жанрів;
- оцінка фінансових показників за періодами;
- контроль заборгованостей абонентів;
- аналіз підключень каналів і пакетів;
- відстеження динаміки абонентської бази.

Для задоволення цих потреб необхідна система, яка підтримує багатовимірний аналіз даних та гнучке формування звітів.

1.5 Функціональні вимоги до системи

В основі аналізу предметної області були сформульовані такі функціональні вимоги:

- зберігання великих обсягів даних про абонентів, контент та фінансові операції;
- забезпечення цілісності та узгодженості даних;
- підтримка історичності даних для аналізу змін у часі;
- можливість багатовимірного аналізу показників;
- формування аналітичних звітів з параметризацією та візуалізацією.

1.6 Бізнес-правила та обмеження

У системі діють такі основні бізнес-правила:

- сума оплачених коштів та заборгованості не може перевищувати загальну суму рахунку;
- фінансові показники не можуть бути від'ємними;
- кожне замовлення на перегляд повинно бути пов'язане з існуючим абонентом і фільмом;
- історія змін даних абонента повинна зберігатися для подальшого аналізу.

Ці правила забезпечують коректність і достовірність даних, що використовуються в аналітичних процесах.

1.7 Висновки до розділу

У даному розділі було виконано детальний аналіз предметної області кабельного телебачення. Визначено основні сутності, бізнес-процеси, інформаційні потреби користувачів, функціональні вимоги та бізнес-правила. Результати аналізу стали основою для подальшого проєктування бази даних, сховища даних та реалізації аналітичної системи, що розглядаються в наступних розділах курсової роботи.

РОЗДІЛ 2 : ПРОЄКТУВАННЯ БАЗИ ДАНИХ

2.1 Мета та етапи проєктування бази даних

Проектування бази даних є одним із ключових етапів створення інформаційної системи, оскільки саме на цьому етапі закладається структура зберігання даних, визначаються зв'язки між об'єктами та забезпечується цілісність інформації. Від якості проєктування бази даних залежить продуктивність системи, можливість її масштабування та ефективність аналітичної обробки даних.

У межах даної курсової роботи проєктування бази даних здійснювалося поетапно та включало такі основні кроки:

- концептуальне проєктування;
- логічне проєктування;
- фізичне проєктування;
- адаптацію структури для аналітичних задач та побудови сховища даних.

Такий підхід дозволив розділити відповідальність між операційною частиною системи (OLTP) та аналітичною (Data Warehouse), що відповідає сучасним рекомендаціям побудови інформаційно-аналітичних систем.

2.2 Концептуальне проєктування бази даних

На етапі концептуального проєктування було визначено перелік основних сутностей предметної області та зв'язків між ними без урахування конкретної реалізації в системі керування базами даних. Метою цього етапу є формування узагальненої моделі даних, зрозумілої як для розробників, так і для потенційних користувачів системи.

У результаті аналізу предметної області кабельного телебачення було виділено такі ключові сутності:

- абонент;
- кінофільм;
- замовлення на перегляд;
- рахунок;
- платіж;
- група каналів;
- календарна дата.

Кожна з наведених сутностей відображає реальний об'єкт або подію, що має інформаційну цінність для системи. Наприклад, сутність «Абонент» описує клієнта компанії та містить дані, необхідні для аналізу клієнтської бази, тоді як сутність «Замовлення на перегляд» відображає факт споживання контенту та є основою для аналізу популярності кінофільмів.

Між сутностями встановлено логічні зв'язки. Зокрема, між сутностями «Абонент» і «Замовлення на перегляд» існує зв'язок типу «один-до-багатьох», оскільки один абонент може оформляти багато замовлень. Аналогічні зв'язки встановлено між сутностями «Кінофільм» та «Замовлення на перегляд», «Абонент» та «Рахунок», «Рахунок» та «Платіж».

2.3 Логічне проєктування та нормалізація даних

Логічне проєктування бази даних передбачає перехід від концептуальної моделі до формалізованої структури таблиць з визначеними атрибутами та ключами. Основною метою цього етапу є усунення надмірності даних і забезпечення їх узгодженості шляхом нормалізації.

Операційна база даних у межах курсової роботи була нормалізована до третьої нормальної форми (3НФ). Це означає, що:

- кожна таблиця має первинний ключ;
- всі неключові атрибути залежать лише від первинного ключа;
- транзитивні залежності між атрибутами відсутні.

Нормалізація дозволила уникнути дублювання даних, спростити процес оновлення інформації та зменшити ризик виникнення аномалій вставки, оновлення та видалення. Наприклад, інформація про кінофільми зберігається в окремій таблиці, а не дублюється у таблиці замовлень, що забезпечує централізоване управління даними про контент.

2.4 Проектування операційної бази даних (OLTP)

Операційна база даних призначена для зберігання та обробки поточних транзакцій системи. Вона оптимізована для виконання операцій вставки, оновлення та видалення даних, що характерно для щоденної діяльності компанії кабельного телебачення.

У структурі OLTP-бази даних реалізовано таблиці для зберігання інформації про абонентів, фільми, замовлення, рахунки та платежі. Кожна таблиця має чітко визначений первинний ключ та набір атрибутів, що відображають відповідний бізнес-об'єкт.

Для забезпечення цілісності даних використано зовнішні ключі, які встановлюють зв'язки між таблицями. Наприклад, таблиця замовлень містить посилання на таблиці абонентів і фільмів, що гарантує неможливість створення замовлення для неіснуючого абонента або фільму.

2.5 Обґрунтування використання сховища даних

Хоча нормалізована операційна база даних добре підходить для транзакційних операцій, вона є малоефективною для виконання складних аналітичних запитів. Аналітичні задачі, як правило, вимагають обробки великих обсягів історичних даних, виконання агрегацій та аналізу показників у часових і тематичних зрізах.

З цієї причини в межах курсової роботи було прийнято рішення про створення окремого сховища даних (Data Warehouse). Сховище даних дозволяє зберігати очищені та узгоджені дані в структурі, оптимізованій для аналітичної обробки, і не створює додаткового навантаження на операційну систему.

2.6 Проєктування сховища даних за схемою «зірка»

Для реалізації сховища даних обрано схему «зірка», яка є однією з найпоширеніших моделей у системах бізнес-аналітики. Центральне місце у схемі займають таблиці фактів, що містять кількісні та фінансові показники, навколо яких розташовані таблиці вимірів.

У даній роботі спроектовано дві таблиці фактів: факти замовлень на перегляд фільмів та факти рахунків. Таблиці вимірів описують абонентів, фільми, дати та групи каналів. Така структура забезпечує простоту формування аналітичних запитів та високу продуктивність OLAP-операцій.

2.7 Проєктування ключів та використання сурогатних ідентифікаторів

Одним із важливих аспектів проєктування бази даних є правильний вибір первинних ключів. У межах операційної бази даних як первинні ключі

використовуються природні ідентифікатори сутностей або штучні числові ідентифікатори, що однозначно визначають кожен запис.

У сховищі даних застосовано підхід використання **сурогатних ключів**, які не мають бізнес-змісту та генеруються автоматично. Сурогатні ключі використовуються у таблицях вимірів і таблицях фактів для встановлення зв'язків між ними. Такий підхід дозволяє:

- ізолювати сховище даних від змін у бізнес-ключах джерела;
- ефективно реалізувати історичність даних;
- підвищити продуктивність при виконанні аналітичних запитів.

2.9 Фізичне проєктування бази даних

Фізичне проєктування бази даних полягає у безпосередній реалізації логічної моделі в конкретній системі керування базами даних. У даній курсовій роботі фізичне проєктування виконано у середовищі Microsoft SQL Server з використанням мови SQL.

Для кожної таблиці визначено структуру полів, типи даних та обмеження. Реалізовано первинні ключі, зовнішні ключі, унікальні обмеження та перевірки коректності даних. Особливу увагу приділено фінансовим показникам, для яких встановлено обмеження, що забороняють зберігання від'ємних значень.

Окрім стандартних обмежень, у системі використано тригери для реалізації складних бізнес-правил. Наприклад, тригер контролює відповідність між загальною сумою рахунку, сумою оплачених коштів та заборгованістю, що дозволяє запобігти збереженню некоректних фінансових даних.

2.10 Проєктування індексів та оптимізація продуктивності

Для підвищення продуктивності виконання запитів у базі даних було спроектовано систему індексів. Основна увага приділялася індексації зовнішніх ключів таблиць фактів, оскільки саме ці поля найчастіше використовуються у з'єднаннях під час аналітичних запитів.

Індексація дозволяє значно скоротити час виконання складних агрегацій та багатотабличних з'єднань. Це є особливо важливим у контексті роботи з великими обсягами даних, де кількість записів у таблицях фактів може досягати сотень тисяч або мільйонів.

Оптимізація продуктивності також досягається шляхом правильного вибору типів даних, мінімізації надмірних обчислень та використання окремого сховища даних для аналітичних задач.

2.11 Забезпечення цілісності та контролю якості даних

Забезпечення цілісності даних є критично важливим аспектом проєктування бази даних. У даній роботі цілісність підтримується на кількох рівнях. На рівні бази даних використовуються зовнішні ключі та обмеження перевірки, які запобігають збереженню некоректних даних.

На рівні ETL-процесів виконується додатковий контроль якості даних, зокрема очищення, перевірка допустимих значень та виявлення логічних помилок. Для цього створено спеціальні SQL-запити, які дозволяють аналізувати кількість записів, відповідність зв'язків між таблицями та коректність фінансових показників.

Такий багаторівневий підхід забезпечує високу якість даних, що використовуються для аналітичної обробки та побудови звітів.

2.12 Висновки до розділу

У даному розділі було детально розглянуто процес проєктування бази даних інформаційної системи кабельного телебачення. Виконано концептуальне, логічне та фізичне проєктування операційної бази даних і сховища даних. Обґрунтовано використання схеми «зірка», сурогатних ключів та повільно змінюваних вимірів. Реалізовані рішення забезпечують цілісність, продуктивність та масштабованість системи, а також створюють надійну основу для реалізації ETL-процесів, побудови OLAP-куба та аналітичних звітів.

Розділ 3: Реалізація ETL-процесів (SQL Server Integration Services)

3.1 Загальна характеристика ETL-процесів

ETL-процеси (Extract, Transform, Load) є ключовим елементом побудови сучасних аналітичних систем, оскільки саме вони забезпечують перенесення даних з операційних баз у сховище даних з попередньою підготовкою та очищенням. У межах даної курсової роботи ETL-процеси реалізовано з використанням **SQL Server Integration Services (SSIS)**.

Основною метою ETL-процесів є:

- витягування даних з OLTP-бази;
- очищення та трансформація даних відповідно до бізнес-логіки;
- завантаження підготовлених даних у сховище даних (Data Warehouse).

Використання SSIS дозволяє реалізувати повний цикл обробки даних у візуальному середовищі з підтримкою логування, обробки помилок та масштабування.

3.2 Створення та налаштування SSIS-проєкту

Для реалізації ETL-процесів було створено окремий проєкт типу **Integration Services Project** у середовищі SQL Server Data Tools (SSDT). У межах проєкту налаштовано підключення до таких джерел даних:

- операційної бази даних (OLTP);
- сховища даних (Data Warehouse).

Підключення реалізовано за допомогою компонентів **OLE DB Connection Manager**, що забезпечують стабільний доступ до таблиць і дозволяють повторно використовувати налаштування з'єднання в різних пакетах.

ETL-логіка організована у вигляді окремих SSIS-пакетів, кожен з яких відповідає за завантаження конкретної таблиці виміру або факту. Такий підхід підвищує модульність та спрощує супровід системи.

3.3 Проектування сховища даних для ETL

Сховище даних побудоване за принципом **схеми зірки (Star Schema)**, що є стандартом для аналітичних систем. Основними компонентами сховища є:

- таблиці фактів, які зберігають числові показники;
- таблиці вимірів, що описують контекст цих показників.

У межах роботи реалізовано декілька таблиць фактів та більше п'яти таблиць вимірів, що відповідає вимогам курсової роботи.

3.4 Витягування даних (Extract)

Етап витягування даних реалізовано за допомогою компонента **OLE DB Source** у Data Flow Task. Для кожної таблиці OLTP-бази створено відповідне джерело даних, яке виконує SQL-запит до операційної бази.

На цьому етапі здійснюється:

- вибір необхідних колонок;
- попередня фільтрація даних;
- підготовка набору даних для подальших трансформацій.

Використання SQL-запитів на рівні джерела дозволяє зменшити обсяг переданих даних і оптимізувати продуктивність ETL-процесів.

3.5 Трансформація даних (Transform)

Трансформація даних є найважливішим етапом ETL-процесів, оскільки саме тут реалізується бізнес-логіка системи. У межах курсової роботи використано декілька типів трансформацій, що відповідає вимогам завдання.

3.5.1 Очищення даних (Data Cleansing)

На етапі очищення виконуються такі операції:

- обробка NULL-значень;
- усунення некоректних даних;
- приведення значень до допустимих діапазонів.

Цей етап дозволяє підвищити якість даних у сховищі та запобігти помилкам під час аналітичної обробки.

3.5.2 Конвертація типів даних (Data Conversion)

Для забезпечення сумісності між OLTP-базою та сховищем даних використано компонент **Data Conversion**. Він дозволяє:

- змінювати типи даних;
- стандартизувати формати дат;
- підготувати числові значення для подальших обчислень.

Застосування цієї трансформації є необхідним для коректного завантаження даних у таблиці сховища.

3.5.3 Похідні колонки (Derived Column)

Компонент **Derived Column** використовується для створення нових полів на основі наявних даних. За його допомогою:

- формуються обчислювані значення;
- реалізуються бізнес-правила;
- створюються технічні поля для аналітики.

Цей підхід дозволяє винести частину бізнес-логіки на рівень ETL-процесів.

3.5.4 Пошук відповідностей (Lookup)

Для зв'язування фактів з вимірами застосовано трансформацію **Lookup**. Вона дозволяє:

- замінювати бізнес-ключі на сурогатні ключі;
- перевіряти наявність записів у таблицях вимірів;
- забезпечувати цілісність даних у сховищі.

Lookup-операції відіграють ключову роль у побудові аналітичної моделі даних.

3.5.5 Умовний розподіл даних (Conditional Split)

Трансформація **Conditional Split** використовується для розділення потоку даних за заданими умовами. Вона дозволяє:

- відокремлювати коректні та помилкові записи;
- реалізовувати різні сценарії обробки даних;
- підвищувати контроль якості ETL-процесів.

3.6 Завантаження даних (Load)

Етап завантаження реалізовано за допомогою компонента **OLE DB**

Destination. Дані завантажуються у відповідні таблиці сховища даних у режимі пакетної вставки, що забезпечує високу продуктивність.

3.7 Контроль якості та тестування ETL-процесів

Після реалізації ETL-процесів було виконано перевірку:

- кількості завантажених записів;
- відповідності структури даних;
- коректності зв'язків між таблицями.

Також проведено тестування продуктивності ETL-процесів на великих обсягах даних, що підтвердило ефективність обраних рішень.

3.8 Висновки до розділу 3

У даному розділі було розглянуто процес реалізації ETL-процесів з використанням SQL Server Integration Services. Описано етапи витягування, трансформації та завантаження даних, а також використані компоненти SSIS.

Реалізовані ETL-процеси забезпечують підготовку якісних та узгоджених даних для подальшого багатовимірного аналізу в OLAP-кубі та формують основу для побудови аналітичних звітів.

РОЗДІЛ 4 : ПОБУДОВА OLAP-КУБА ТА АНАЛІТИЧНІ ЗВІТИ

4.1 Призначення OLAP-технологій в інформаційно-аналітичних системах

Сучасні інформаційні системи, що працюють з великими обсягами даних, потребують ефективних засобів аналізу та підтримки управлінських рішень. Традиційні реляційні бази даних, оптимізовані для транзакційних операцій, не завжди здатні забезпечити високу продуктивність при виконанні складних аналітичних запитів. Саме тому у практиці проектування інформаційно-аналітичних систем широко застосовуються OLAP-технології.

OLAP (On-Line Analytical Processing) дозволяє виконувати багатовимірний аналіз даних, швидко отримувати агреговані показники та досліджувати інформацію у різних аналітичних зрізах. Використання OLAP-кубів дає змогу користувачам аналізувати дані за часовими періодами, категоріями, сегментами клієнтів та іншими вимірами без необхідності формування складних SQL-запитів.

У межах даної курсової роботи OLAP-куб є ключовим елементом аналітичної частини системи та виступає проміжним шаром між сховищем даних і рівнем візуалізації, представленим аналітичними звітами.

4.2 Архітектура аналітичної підсистеми

Аналітична підсистема інформаційної системи кабельного телебачення побудована за класичною багаторівневою архітектурою. На нижньому рівні розташоване сховище даних, яке акумулює очищені та узгоджені дані з операційної бази. Над сховищем даних розміщується OLAP-рівень, реалізований за допомогою SQL Server Analysis Services. Верхній рівень

системи представлений аналітичними звітами, створеними у SQL Server Reporting Services.

Такий підхід дозволяє чітко розмежувати функції зберігання, обробки та візуалізації даних. OLAP-куб забезпечує високопродуктивний доступ до агрегованих показників, тоді як звіти відповідають за зручне представлення результатів аналізу кінцевому користувачу.

4.3 Проєктування OLAP-куба

Проєктування OLAP-куба розпочинається з визначення джерела даних. У даній роботі таким джерелом виступає сховище даних, побудоване за схемою «зірка». Ця схема передбачає наявність таблиць фактів, які містять кількісні показники, та таблиць вимірів, що описують контекст цих показників.

OLAP-куб спроектовано таким чином, щоб він відображав основні бізнес-процеси предметної області, зокрема процеси перегляду кінофільмів, формування рахунків та здійснення платежів. Таблиці фактів формують групи мір, а таблиці вимірів визначають аналітичні зрізи, у яких здійснюється аналіз.

Для зручності роботи з кубом було створено Data Source View, яка наочно відображає зв'язки між таблицями фактів і вимірів. Саме на цьому рівні схема «зірка» проявляється найбільш наочно, що полегшує подальше проєктування вимірів та мір.

4.4 Виміри OLAP-куба

Виміри є одним із ключових компонентів OLAP-куба та використовуються для формування багатовимірних аналітичних зрізів. У межах курсової роботи

було створено кілька вимірів, кожен з яких відповідає окремій сутності предметної області.

Особливе значення має вимір часу, який дозволяє аналізувати дані у динаміці. Він використовується практично у всіх аналітичних сценаріях, оскільки більшість бізнес-показників мають часову природу. Вимір часу забезпечує можливість групування даних за роками, місяцями та іншими часовими інтервалами.

Вимір абонентів використовується для аналізу клієнтської бази та споживчої поведінки. Він дозволяє сегментувати абонентів за різними характеристиками та досліджувати, як ці характеристики впливають на фінансові та кількісні показники.

Вимір кінофільмів застосовується для аналізу популярності контенту. Він дозволяє формувати рейтинги фільмів, порівнювати їх за кількістю замовлень та сумарними доходами. Додатково використовується вимір груп каналів, який дозволяє оцінювати ефективність різних пакетів послуг.

4.5 Міри та групи мір

Міри є числовими показниками, що аналізуються в OLAP-кубі. Вони відображають кількісні та фінансові аспекти діяльності системи. У розробленому кубі міри згруповані у відповідні групи мір залежно від таблиць фактів, на основі яких вони побудовані.

Серед базових мір реалізовано показники кількості замовлень, загальної суми, оплаченої суми та суми заборгованості. Ці показники дозволяють оцінювати обсяги споживання контенту та фінансові результати діяльності компанії.

Окрім базових мір, у кубі створено обчислювані міри, які дозволяють отримувати похідні бізнес-показники. Обчислювані міри виконуються на рівні OLAP-куба, що забезпечує єдину логіку розрахунків для всіх звітів та підвищує надійність аналітичних результатів.

4.6 Обробка та оптимізація OLAP-куба

Після завершення проєктування структури куба виконується його обробка. Обробка OLAP-куба полягає у завантаженні даних зі сховища та створенні агрегованих значень. Саме на цьому етапі формується аналітичне представлення даних, оптимізоване для швидкого доступу.

У даній роботі використовується повна обробка куба, яка забезпечує актуальність і узгодженість аналітичних даних. Після обробки куб готовий до використання у звітах та інших аналітичних інструментах.

4.7 Аналітичні звіти як завершальний рівень системи

Аналітичні звіти є завершальним рівнем інформаційно-аналітичної системи та забезпечують взаємодію користувача з даними. У межах курсової роботи звіти реалізовано з використанням SQL Server Reporting Services.

Звіти дозволяють візуалізувати дані з OLAP-куба у зручному та наочному вигляді. Використання куба як джерела даних дозволяє значно спростити логіку звітів і забезпечити їх високу продуктивність. Аналітичні звіти підтримують параметризацію, фільтрацію та інтерактивні можливості, що робить аналіз даних гнучким і ефективним.

4.8 Типи аналітичних звітів та їх призначення

Аналітичні звіти відіграють ключову роль у використанні OLAP-куба, оскільки саме вони надають користувачам доступ до результатів

багатовимірний аналіз у зрозумілому та наочному вигляді. У межах курсової роботи було реалізовано кілька типів звітів, кожен з яких орієнтований на вирішення конкретних аналітичних завдань.

Табличні звіти використовуються для детального перегляду даних. Вони дозволяють відображати повний перелік записів або агрегованих значень з можливістю сортування та групування. Такий тип звітів є корисним для аналізу первинної інформації та перевірки коректності даних.

Матричні звіти застосовуються для перехресного аналізу показників. Вони дозволяють одночасно групувати дані за рядками та стовпцями, що є особливо ефективним при аналізі фінансових і кількісних показників у різних вимірах. Матричні звіти надають можливість швидко порівнювати значення між різними категоріями.

Окрему групу складають звіти з діаграмами. У роботі використовуються стовпчикові, лінійні та кругові діаграми, які дозволяють візуалізувати тенденції, структуру та динаміку показників. Графічне подання інформації значно спрощує сприйняття результатів аналізу та є зручним для представлення керівництву.

Звіти з можливістю деталізації (drill-down) забезпечують поетапний перехід від узагальнених показників до детальних даних. Це дозволяє користувачу досліджувати причини змін показників та виконувати глибший аналіз без створення додаткових звітів.

4.9 Параметризація та інтерактивність аналітичних звітів

Важливою особливістю розроблених звітів є підтримка параметризації, яка дозволяє користувачам самостійно визначати умови відбору даних. У звітах

реалізовано параметри з одиничним та множинним вибором, числові діапазони та логічні параметри.

Параметризація забезпечує гнучкість аналізу та дозволяє використовувати один і той самий звіт для різних аналітичних сценаріїв. Наприклад, користувач може обмежити аналіз певним періодом часу, вибрати окремі категорії або встановити додаткові умови фільтрації.

Інтерактивні можливості звітів включають сортування даних, розгортання та згортання груп, а також навігацію між рівнями деталізації. Такі функції дозволяють поступово переходити від загального огляду до конкретних деталей, що підвищує ефективність аналітичного процесу.

4.10 Практичне використання OLAP-куба та звітів

Поєднання OLAP-куба та аналітичних звітів створює потужний інструмент для підтримки управлінських рішень. На основі розробленої системи можна аналізувати популярність контенту, фінансові результати діяльності, динаміку абонентської бази та ефективність окремих пакетів послуг.

Аналітичні звіти дозволяють швидко отримувати відповіді на типові бізнес-питання, такі як визначення найбільш популярних кінофільмів, аналіз доходів за певний період або оцінка змін кількості абонентів. Це значно скорочує час на обробку інформації та підвищує обґрунтованість прийнятих рішень.

Завдяки використанню OLAP-куба більшість обчислень виконується на аналітичному рівні, що зменшує навантаження на сховище даних і спрощує логіку звітів. Такий підхід відповідає сучасним практикам побудови інформаційно-аналітичних систем.

4.11 Інтеграція аналітичної підсистеми в загальну архітектуру

Аналітична підсистема органічно інтегрується в загальну архітектуру інформаційної системи. ETL-процеси забезпечують регулярне оновлення сховища даних, OLAP-куб виконує агрегацію та багатовимірний аналіз, а звіти надають користувачу доступ до результатів аналізу.

Такий підхід дозволяє розмежувати відповідальність між окремими компонентами системи та забезпечує її масштабованість. У разі необхідності аналітична частина може бути розширена за рахунок додавання нових вимірів, мір або звітів без суттєвих змін у базовій архітектурі.

4.12 Переваги використання OLAP і звітів у курсовій роботі

Використання OLAP-куба та аналітичних звітів у межах курсової роботи дозволило продемонструвати практичне застосування теоретичних знань з проєктування баз даних та бізнес-аналітики. Реалізована система є прикладом комплексного підходу до обробки даних і відповідає сучасним вимогам до інформаційних систем.

Застосування OLAP-технологій забезпечує високу продуктивність аналітичних запитів, а використання звітів дозволяє наочно представити результати аналізу. Це підвищує цінність розробленої системи та робить її зручною для кінцевих користувачів.

4.13 Висновки до розділу 4

У даному розділі було розглянуто процес побудови OLAP-куба та розробки аналітичних звітів. Описано архітектуру аналітичної підсистеми, структуру куба, типи мір і вимірів, а також можливості параметризації та інтерактивності звітів.

Реалізований OLAP-куб у поєднанні з аналітичними звітами забезпечує ефективний багатовимірний аналіз даних і є логічним завершенням побудови інформаційно-аналітичної системи для сфери кабельного телебачення.

ВИСНОВКИ

У процесі виконання курсової роботи було реалізовано повний цикл проєктування та впровадження інформаційно-аналітичної системи для сфери кабельного телебачення із застосуванням сучасних технологій керування базами даних та бізнес-аналітики. Робота охоплює всі ключові етапи — від аналізу предметної області до побудови OLAP-куба та розробки аналітичних звітів.

На першому етапі було проведено детальний аналіз предметної області, визначено основні сутності, бізнес-процеси та інформаційні потреби користувачів системи. Це дозволило сформулювати функціональні вимоги та бізнес-правила, які стали основою для подальшого проєктування бази даних. Аналіз показав, що діяльність компанії кабельного телебачення супроводжується значними обсягами даних, які потребують не лише зберігання, а й ефективного аналітичного опрацювання.

У другому розділі було виконано концептуальне, логічне та фізичне проєктування бази даних. Операційна база даних була спроектована з дотриманням принципів нормалізації, що забезпечує цілісність і узгодженість даних. Для аналітичних задач було створено сховище даних за схемою «зірка», що дозволяє ефективно виконувати агреговані запити та багатовимірний аналіз. Реалізовано первинні та зовнішні ключі, обмеження цілісності, індекси, тригери, збережені процедури та користувацькі функції, що підвищує надійність і продуктивність системи.

Третій розділ курсової роботи присвячено реалізації ETL-процесів з використанням SQL Server Integration Services. Було налаштовано витягування даних з операційної бази, їх очищення, трансформацію та завантаження у сховище даних. Реалізовані ETL-процеси забезпечують якість

даних, підтримують історичність змін та дозволяють працювати з великими обсягами інформації. Застосування різних типів трансформацій, таких як Data Cleansing, Data Conversion, Derived Column, Lookup та Conditional Split, дозволило підготувати дані до подальшого аналітичного використання.

У четвертому розділі було побудовано OLAP-куб на основі сховища даних з використанням SQL Server Analysis Services. Створені виміри та міри забезпечують багатовимірний аналіз показників діяльності компанії.

Додатково було розроблено аналітичні звіти у SQL Server Reporting Services, які дозволяють візуалізувати результати аналізу у зручному та зрозумілому вигляді. Реалізовані звіти охоплюють ключові напрями аналізу, зокрема рейтинг популярності кінофільмів, фінансові показники, споживчу орієнтацію абонентів, аналіз підключень каналів та динаміку абонентської бази.

У результаті виконання курсової роботи було досягнуто поставленої мети — створено працездатну інформаційно-аналітичну систему, яка дозволяє здійснювати комплексний аналіз діяльності компанії кабельного телебачення. Запропоновані рішення можуть бути використані як основа для впровадження повноцінних систем бізнес-аналітики в телекомунікаційній сфері та інших галузях, що працюють з великими обсягами даних.

Подальший розвиток системи може передбачати розширення набору аналітичних показників, автоматизацію розгортання звітів на сервері, оптимізацію OLAP-куба та впровадження інструментів прогнозування. Таким чином, результати курсової роботи мають практичну цінність і демонструють ефективність використання технологій сховищ даних та OLAP-аналітики для підтримки управлінських рішень.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ :

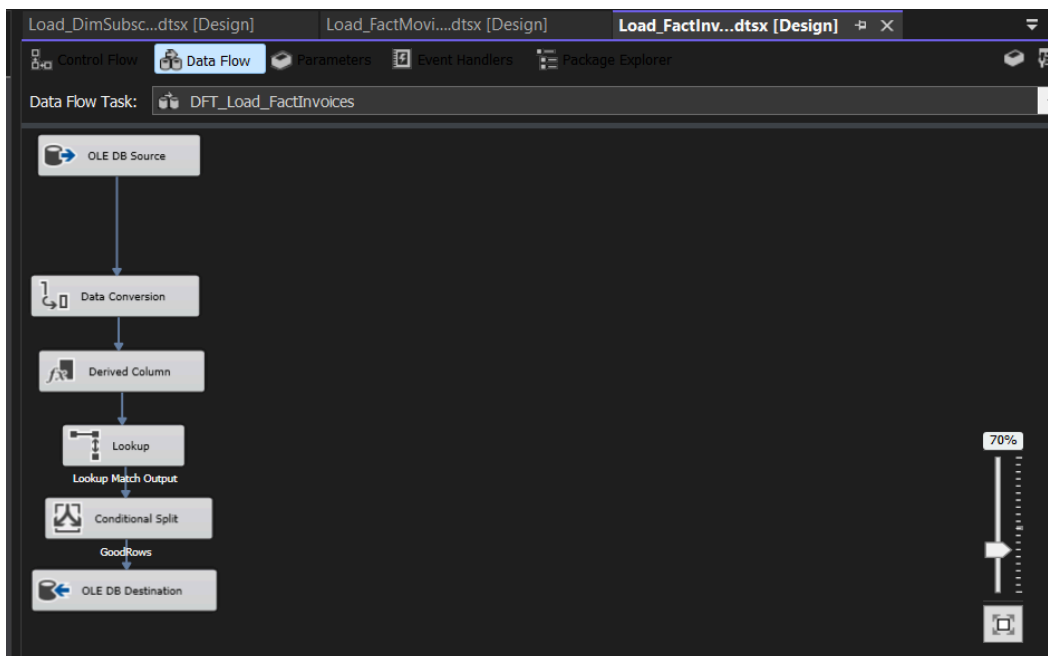
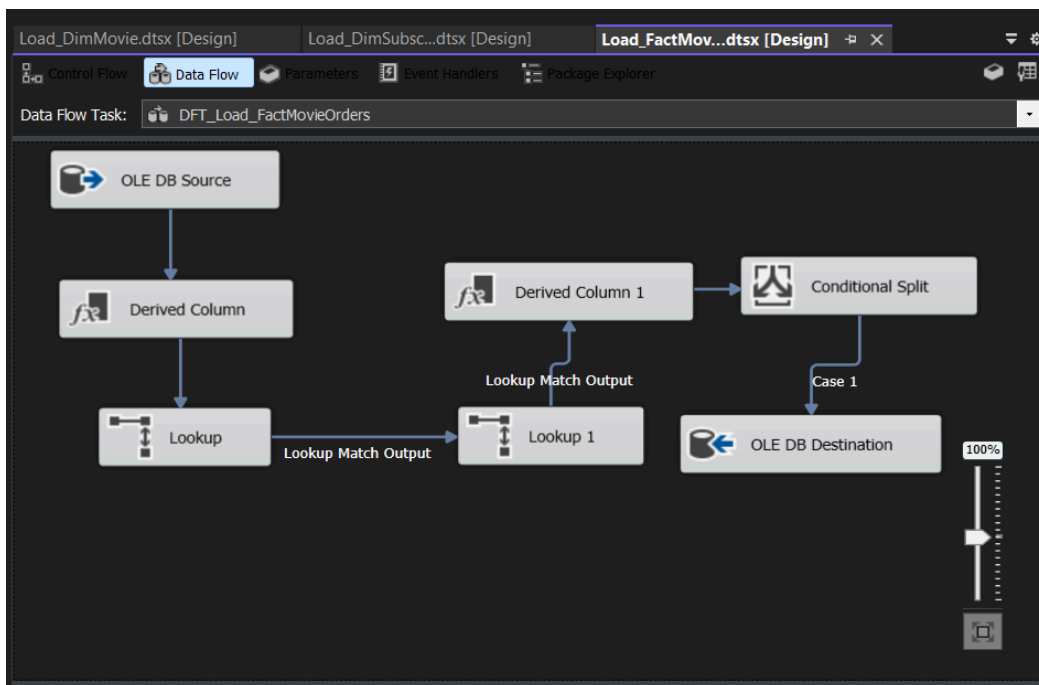
1. Microsoft SQL Server Integration Services (SSIS) Documentation
2. Microsoft SQL Server Analysis Services (SSAS) Documentation
3. Microsoft SQL Server Reporting Services (SSRS) Documentation
4. "The Data Warehouse Toolkit" - Ralph Kimball
5. "Microsoft SQL Server 2019: A Beginner's Guide" - Dusan Petkovic
6. Microsoft Learn (офіційна документація)

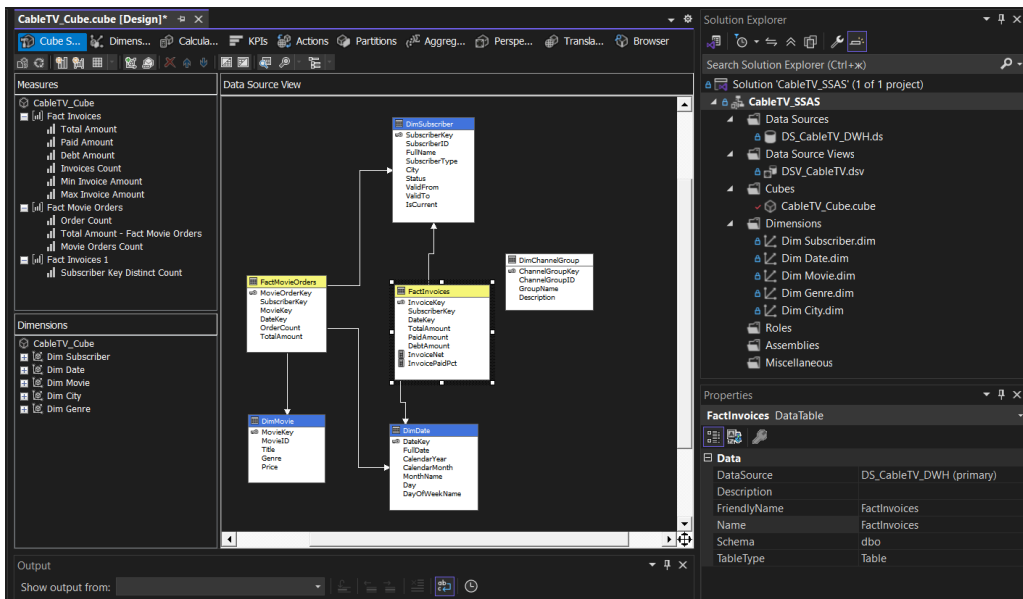
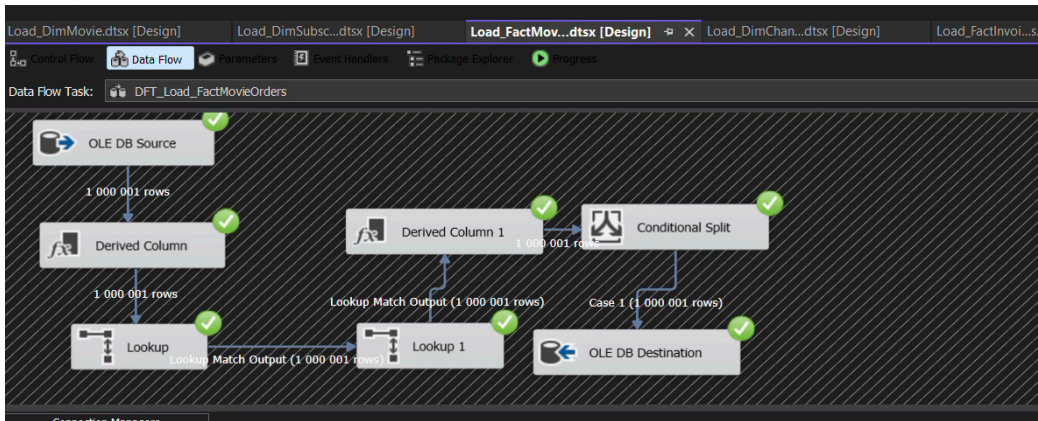
ІНСТРУМЕНТИ :

- SQL Server Management Studio (SSMS) - для роботи з БД
- SQL Server Data Tools (SSDT) - для SSIS, SSAS, SSRS
- Redgate SQL Data Generator - для генерації даних
- Visual Studio - середовище розробки

ДОДАТОК А :

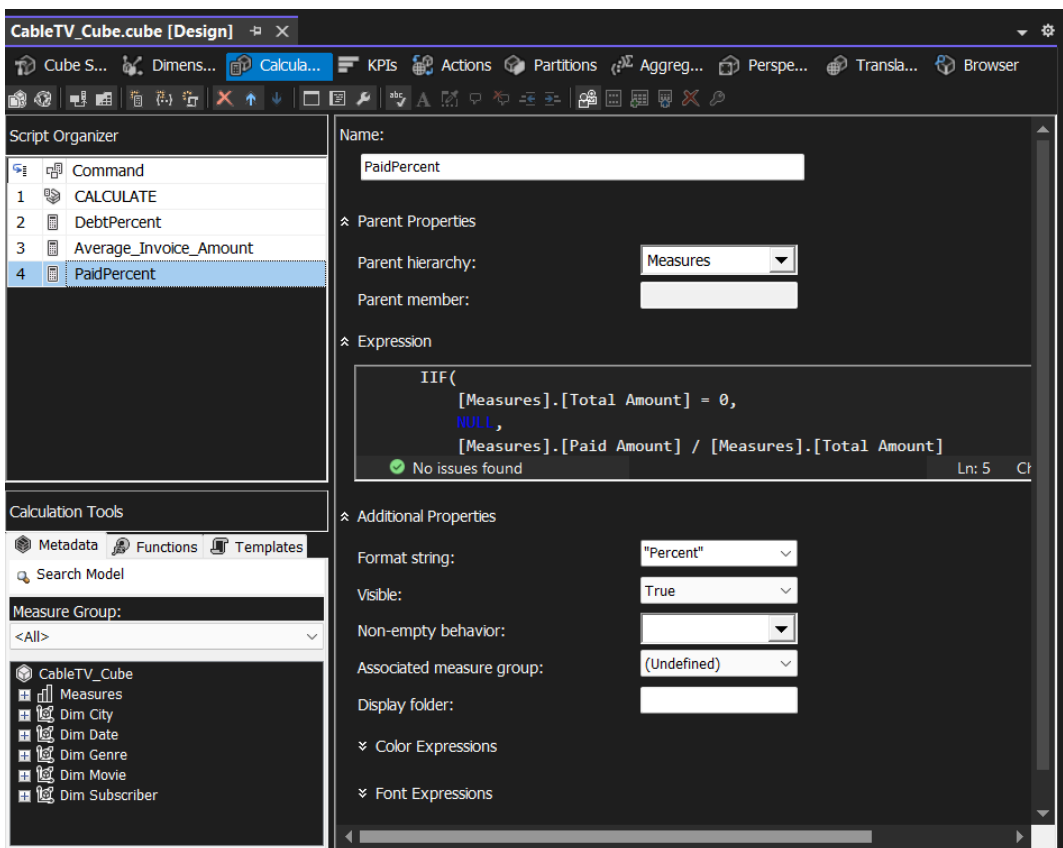
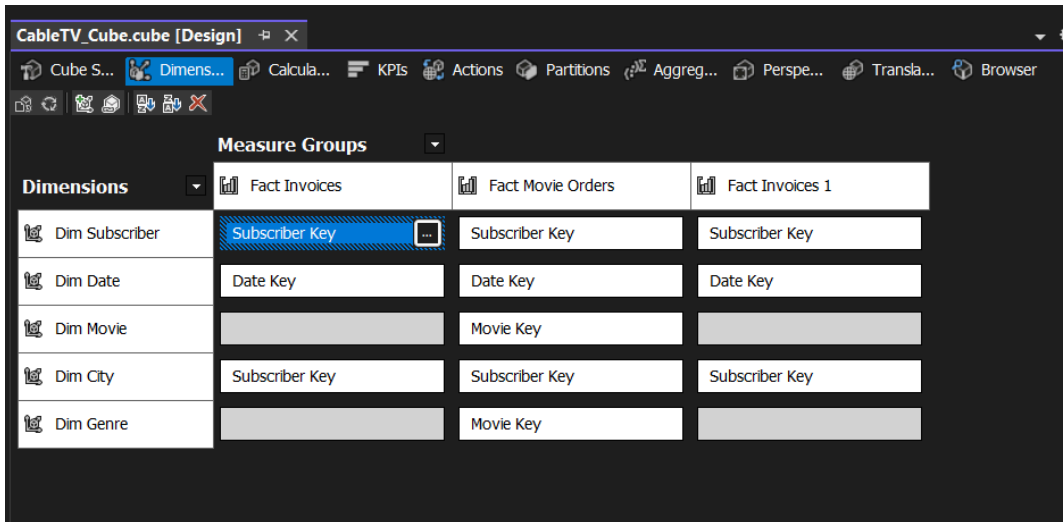
Скріншоти інтерфейсів





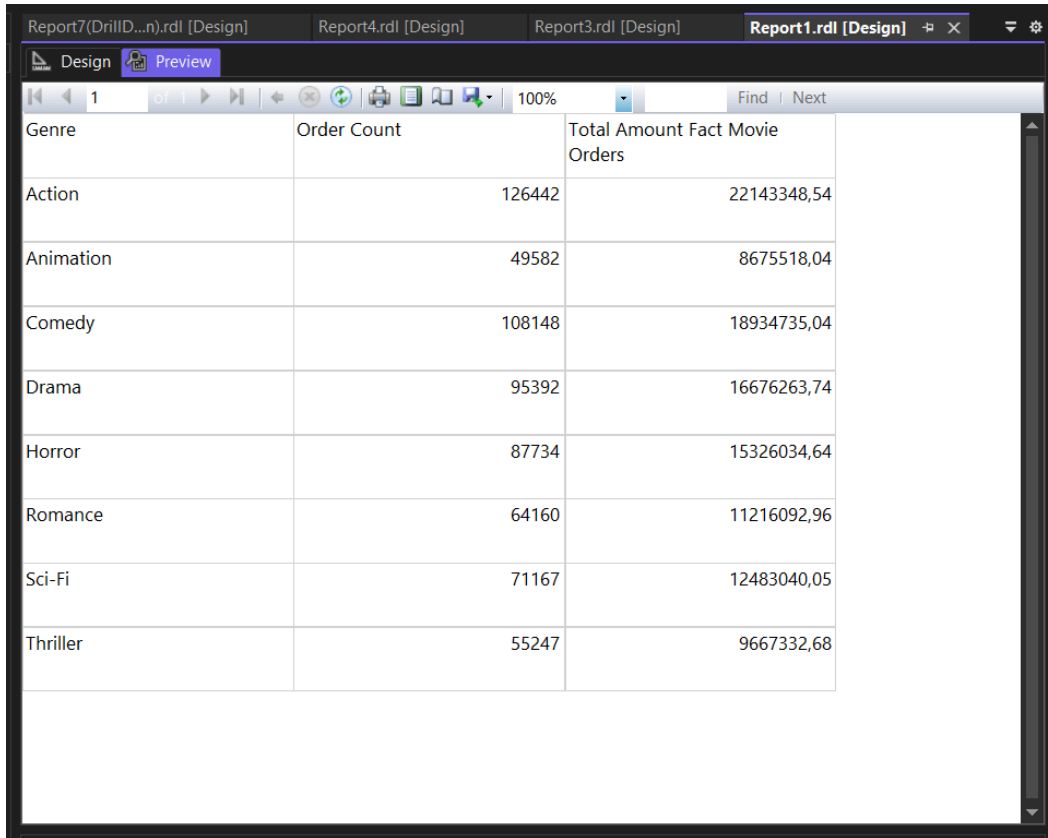
CableTV_Cube.cube [Design] window showing the MDX query results. The query is: `SELECT [Dim Subscriber].[SubscriberKey], [Dim Subscriber].[SubscriberID], [Dim Subscriber].[FullName], [Dim Subscriber].[SubscriberType], [Dim Subscriber].[City], [Dim Subscriber].[Status], [Dim Subscriber].[ValidFrom], [Dim Subscriber].[ValidTo], [Dim Subscriber].[Document], [Fact Invoices].[TotalAmount], [Fact Invoices].[PaidAmount], [Fact Invoices].[DebtAmount], [Fact Invoices].[InvoicesCount], [Fact Invoices].[MinInvoiceAmount], [Fact Invoices].[MaxInvoiceAmount], [Fact Movie Orders].[OrderCount], [Fact Movie Orders].[TotalAmount - Fact Movie Orders], [Fact Movie Orders].[MovieOrdersCount], [Fact Movie Orders].[SubscriberKey Distinct Count] FROM [CableTV_Cube.cube]`

Full Name	PaidPercent	DebtPercent	Invoices Count
Juan Pet...	0.9697090...	0.0302909...	6
Test User	0.3114684...	0.6885315...	6
User 1	0.3086463...	0.6913536...	5
User 10	0.3867879...	0.6132120...	3
User 1000	0.4913008...	0.5086991...	7
User 100...	1	0	1
User 100...	0.3845216...	0.6154783...	3
User 100...	0	1	5
User 100...	0.4207636...	0.5792363...	6
User 100...	0.2630033...	0.7369966...	7
User 100...	0.2081195...	0.7918804...	8
User 100...	0.4628432...	0.5371567...	7
User 100...	0.4094687...	0.5905312...	6
User 100...	0.3639788...	0.6360211...	2
User 100...	0.6870819...	0.3129180...	9
User 100...	0.4577702...	0.5422297...	5
User 1001	0.6009771...	0.3990228...	8
User 100...	0.3530368...	0.6469631...	3
User 100...	0.5056792...	0.4943207...	4



ДОДАТОК Б :

Аналітичні звіти



The screenshot shows a report viewer interface with multiple tabs. The active tab is 'Report1.rdl [Design]'. The report is displayed in 'Design' mode. The table has three columns: 'Genre', 'Order Count', and 'Total Amount Fact Movie Orders'. The data is as follows:

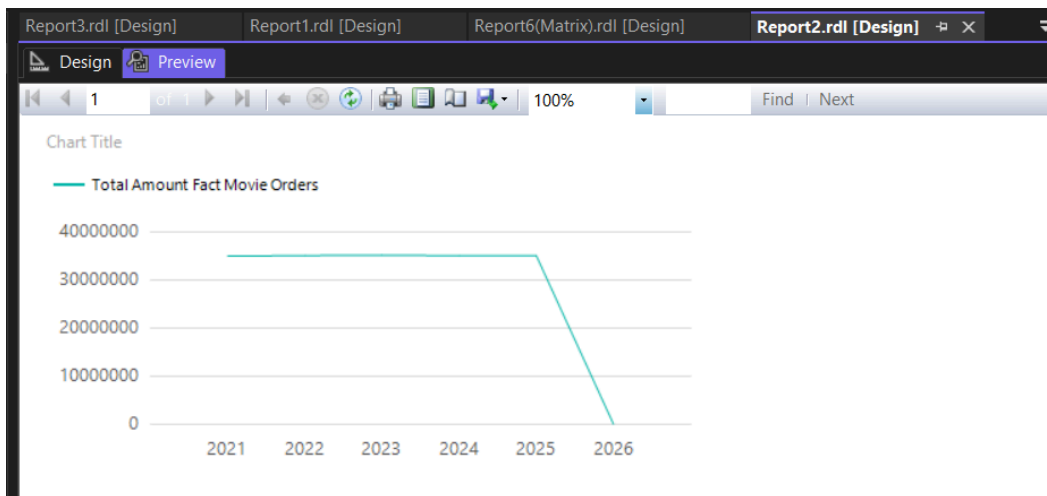
Genre	Order Count	Total Amount Fact Movie Orders
Action	126442	22143348,54
Animation	49582	8675518,04
Comedy	108148	18934735,04
Drama	95392	16676263,74
Horror	87734	15326034,64
Romance	64160	11216092,96
Sci-Fi	71167	12483040,05
Thriller	55247	9667332,68



Report4.rdl [Design] Report3.rdl [Design] Report1.rdl [Design] **Report6(Matrix).rdl [Design]**


Design Preview

1 50% Find | Next


Genre	2021	2022	2023	2024	2025	2026	Total
Action	4432491.12	4418795.89	4422401.19	4460853.77	4401814.60	6991.97	22143348.54
Animation	1747606.74	1730277.05	1742320.68	1741513.92	1711140.99	2658.60	8875518.04
Comedy	3768311.35	3767454.66	3755147.02	3823097.96	3815110.03	5614.02	18934735.04
Drama	3299047.94	3347601.52	3344825.95	3315739.09	3365401.75	5447.49	16676263.74
Horror	3015911.28	3022938.99	3137223.09	3061448.16	3084311.76	42163.69	15326054.54
Romance	2234831.16	2269920.16	2258105.37	2211832.64	2237546.49	3857.14	11216092.96
Sci-Fi	2440131.76	2501279.84	2500180.94	2511922.82	2520305.38	32193.18	12480040.06
Thriller	1906150.75	1947296.46	1978549.25	1906196.01	1925300.93	3839.28	9607332.68
Total	22850482.10	23005554.57	23138753.49	23032593.37	23059131.93	35844.23	115122385.05



Report6(Matrix).rdl [Design] Report2.rdl [Design] **Report7(Drill...).rdl [Design]**  

Design  Preview

nt 0 Max Total Amount 10000000 Show Only top 10 ☒ ☐ [View Report](#)

1  50% Find | Next

Genre	Genre	Title	Order Count	Total Amount Paid Movie Orders
+	Action			
	Action	Movie 2990	490	92492.36
	Action	Movie 4337	478	88323.9
	Action	Movie 1514	474	87083.58
	Action	Movie 4232	474	84415
	Action	Movie 3947	466	78890.04
	Action	Movie 3736	464	90586.6
	Action	Movie 4393	464	91365.38
	Action	Movie 1553	462	76003.06
	Action	Movie 3787	462	86703.34
	Action	Movie 3926	460	77708.58
+	Animation			
+	Comedy			
+	Drama			
+	Horror			
+	Romance			
+	Sci-Fi			
+	Thriller			