**Лабораторна робота №10. Побудова постреляційної, багатовимірної і об'єктної моделі даних**

**Мета:** Отримати навички побудови багатовимірної моделі даних

Завдання на лабораторну роботу

1.Ознайомитися з теоретичними відомостями.

2. За допомогою CASE засобів для проектування визначити структуру БД з інформаційним наповненням фактичними даними щодо діяльності коледжу.

З. Спроектувати на основі топології «Зірка» сховище даних (Data WereHouse) за індивідуальним варіантом.

3. Визначити послідовність дій для процесу завантаження даних (послідовність заповнення таблиць).

4.Оформити результати у вигляді звіту.

Результати надсилати на електронну адресу викладача [t.i.lumpova@gmail.com](mailto:t.i.lumpova@gmail.com)

Файл повинен мати назву в такому форматі:

**DB<Номер групи><Номер лекції / практичної / лабораторної [літера позначення типу роботи L – лекція, P – практична, R – лабораторна]<Прізвище англійською>**. Наприклад, **DB3101R**buts.doc.

Не копіюйте фрагментів з різних інформаційних джерел, подумайте і викладіть свою точку зору. При наявності робіт -"близнюків" відповідь буде зараховуватися першому за часом надсилання.

**Строк виконання цієї роботи ІПЗ-31 - 16.11.2022**

**Контрольні запитання.**

1. Як визначити багатовимірне моделювання?
2. Надайте пояснення терміну "факт" відповідно до сховища даних.
3. Що таке атрибут, вимір, метрика або показник стосовно сховища даних?

**Теоретичні відомості**

**Сховище даних** (англ. *data warehouse*) — предметно орієнтований, інтегрований, незмінний набір даних, що підтримує хронологію і здатний бути комплексним джерелом достовірної інформації для оперативного аналізу та прийняття рішень. В основі концепції сховища даних (СД) лежить розподіл інформації, що використовують в системах оперативної обробки даних (OLTP) і в системах підтримки прийняття рішень (СППР). Такий розподіл дозволяє оптимізувати як структури даних оперативного зберігання для виконання операцій введення, модифікації, знищення та пошуку, так і структури даних, що використовуються для аналізу. В СППР ці два типи даних називаються відповідно оперативними джерелами даних (ОДД) та сховищем даних. Для створення СД важливим є використання багатовимірного моделювання.

Багатовимірне моделювання є методом моделювання та візуалізації таких даних як множини числових або лінгвістичних показників або параметрів (measures), які описують загальні аспекти діяльності організації.

Метод багатовимірного моделювання базується на таких основних поняттях: факти, атрибути, виміри, параметри (метрики), ієрархія, гранулювання.

Факт (fact) - це набір пов'язаних елементів даних, що містять метрики і описові дані. Кожен факт звичайно являє елемент даних, чисельно описує діяльність організації, бізнес-операцію чи подію, що може бути використаний для аналізу діяльності організації або бізнес-процесів. У сховищі даних (СД) факти зберігаються в базових таблицях реляційної БД. Наприклад, вартість товару, кількість одиниць товару і т.д.

Атрибут (Аttribute) - це опис характеристики реального об'єкта предметної області. Як правило, атрибут містить заздалегідь відоме значення, що характеризує факт. Атрибути можуть представляються текстовими полями з дискретними значеннями (наприклад, габарити упаковки товару, запах товару) або у вигляді кодових значень, прив’язаних до відповідного довідника.

Вимір (dimension) - це інтерпретація факту з деякою точки зору в реальному світі. Виміри, подібно атрибутам, містять текстові або числові значення, які сильно пов'язані за змістом між собою. Зазвичай виміри це осі багатовимірного простору, точками якого є пов'язані з ними факти. У багатовимірної моделі кожен факт пов'язаний з однією або декількома осями. Виміри зазвичай представляють нечислові, лінгвістичні змінні, такі як філії організації, співробітники організації, покупці і т.д.

Параметр, метрика або показник (measure) - це числова характеристика факту, який визначає ефективність діяльності або бізнес-дії організації з погляду вимірювання. Як правило, метрика містить заздалегідь не відоме значення характеристики факту. Конкретні значення метрики описуються за допомогою змінних. Наприклад, нехай метрикою є чисельне вираження продажів товару в грошах, кількість проданих одиниць товару тощо. Метрика визначається за допомогою комбінації елементів виміру, і, таким чином, являє факт.

Гранулювання (Granularity) - це рівень деталізації даних, що зберігаються в СД. Наприклад, щоденні обсяги продажів.

З погляду взаємозв'язку вимірювань і фактів останні можна розбити на наступні класи:

• адитивні факти (Additive facts). Факт називається адитивним, якщо його має сенс використовувати з будь-якими вимірами для виконання операцій підсумовування з метою отримання будь-якого значимого результату. Наприклад, дискретні числові показники активності діяльності, такі як кількість продажів, обсяг продажів тощо;

• полуаддітівние факти (Semiadditive facts). Факт називається полуаддітівним, якщо його має сенс використовувати спільно з деякими вимірами для виконання операцій підсумовування з метою отримання будь-якого значимого результату. Наприклад, числові показники інтенсивності, такі як залишок на рахунку, рівень запасів на складі тощо;

• неаддитивні факти (Non-additive facts). Факт називається неаддитивним, якщо його не має сенсу використовувати спільно з яким-небудь виміром для виконання операцій підсумовування з метою отримання будь-якого значимого результату. Наприклад, вимірювання кімнатної температури;

• числові заходи інтенсивності (Numerical Measures of Intensity). Факт називається числовий мірою інтенсивності, якщо він, будучи неаддитивним за часом, допускає агрегацію і підсумовування по деякому числу часових періодів. Наприклад, залишок на рахунку.

Таблиці фактів поділяють на три основні категорії, залежно від рівня деталізації фактів (гранулированности).

• Транзакційна таблиця фактів. У такій таблиці фактів зберігають факти, які фіксують певні події (транзакції). Це факти, що описують кожну подію бізнесу. Наприклад, продажі товару.

• Таблиця фактів періодичних моментальних знімків У такій таблиці збирають факти, що фіксують поточний стан певного напряму бізнесу. Це факти, які описують поточний стан певного напряму бізнесу для будь-якої комбінації значень вимірів за даний період часу. Наприклад, продажі організації на певну дату (щоденно).

• Таблиця фактів кумулятивних моментальних знімків У такій таблиці збирають факти, що фіксують деякий підсумкове стан певного напряму бізнесу на поточний момент часу. Це факти, які описують проміжні підсумки діяльності організації за певним напрямом бізнесу для будь-якої комбінації значень вимірів за даний період часу. Наприклад, продажі цього року на певну дату.

Основними характеристиками таблиці вимірів є наступні.

1. Таблиці вимірів містять дані про деталізації фактів.

2. Таблиці вимірів містять описову інформацію про числових значеннях в таблиці фактів, тобто вони містять атрибути фактів.

3. Як правило, денормалізовані таблиці вимірів містять велику кількість полів.

4. Таблиці вимірів містять зазвичай значно менше рядків, ніж таблиці фактів.

5. Атрибути таблиць вимірів зазвичай використовуються при візуалізації даних у звітах і запитах.

Схематично подати факт в СД можна так (рис.1):



Рис.1. Багатовимірна організація даних у сховищі

Схема "зірка" (рис.2) має одну таблицю фактів і кілька таблиць вимірів. Таблиці вимірів є денормалізованнимі.

Схема "сніжинка" (рис.3) має одну таблицю фактів і кілька нормалізованих таблиць вимірів.



Рис.2. Схема зірка

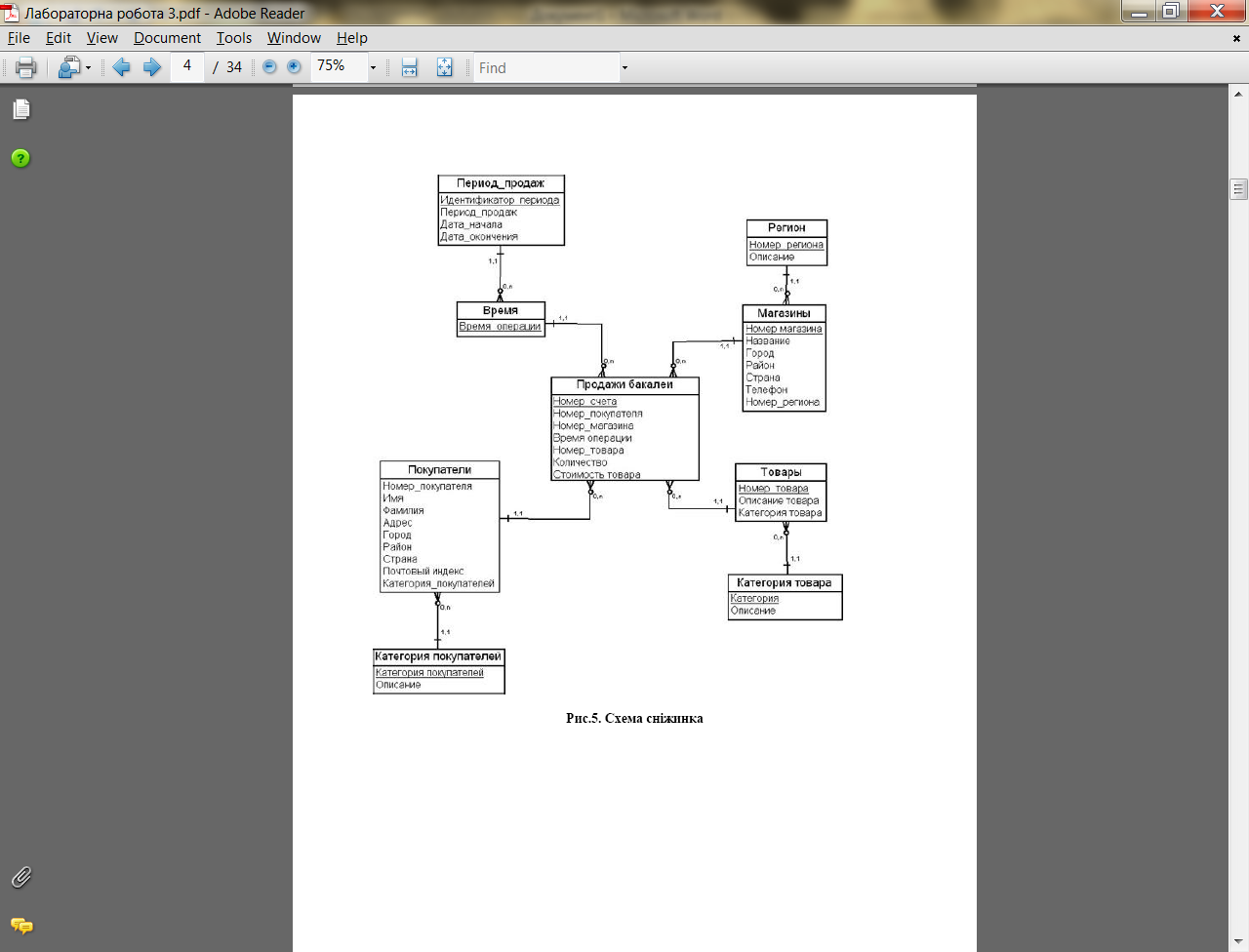


Рис.3. Схема сніжинка

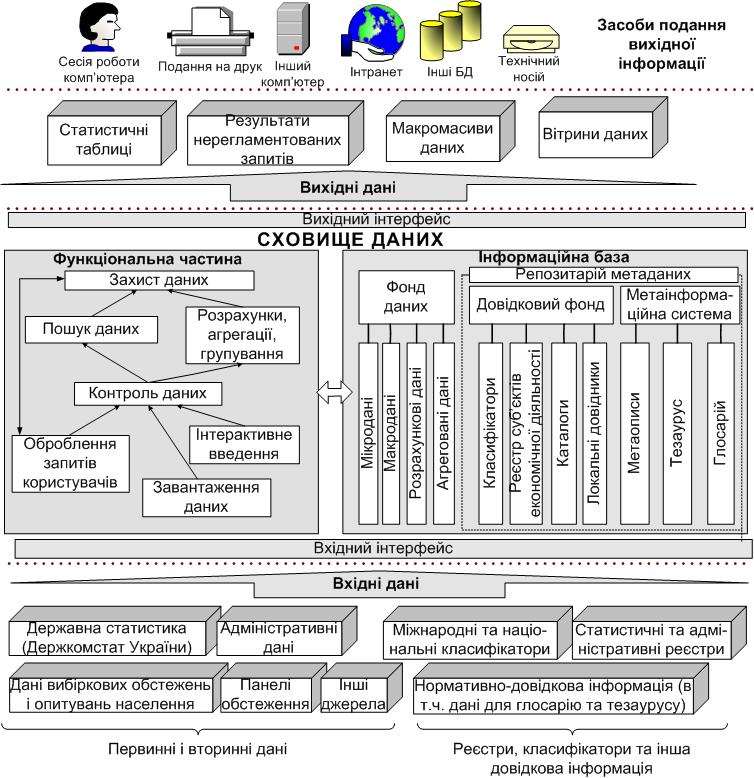


Рис. 4. Приклад структурно-функціональної моделі інформаційної системи зі сховищем статистичних даних