**Практична робота №9. Побудова моделі сховища даних**

**Мета: дослідити процес проектування сховищ**

**даних на основі моделі ROLAP.**

**Завдання**

1. Ознайомитися з теоретичними відомостями
2. Виконати збір вимог до сховища даних (продовження ЛР№10, тема діяльність коледжу) за наступним планом:

- Мета побудови сховища даних

- Опис організаційного середовища

- Опис корпоративної ЕR моделі

- Опис бізнес-вимог користувача

1. Визначити елементи моделі сховища даних

- Визначити виміри

- Визначити показники

- Визначити метрики

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, **DB3101Р**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**

Теоретичні відомості

Сховище даних - предметно-орієнтований, інтегрований, незмінуюваний, набір даних, що підтримує хронологію, організований для цілей підтримки прийняття рішень.

Можна виділити наступні архітектури таких систем:

* Система підтримки прийняття рішень (СППР) з фізичним (класичним) сховищем даних (СД);
* СППР з віртуальним СД;
* СППР з вітринами даних (ВД);
* СППР з фізичним СД і з ВД.

Всі дані в СД діляться на три основні категорії:

• детальні дані;

• агреговані дані;

• метадані.

Багатовимірне концептуальне уявлення (multi-dimensional conceptual view) - це множинна перспектива, що складається з декількох незалежних вимірювань, уздовж яких можуть бути проаналізовані певні сукупності даних. Одночасний аналіз по декількох вимірах визначається як багатовимірний аналіз. Кожен вимір може бути представлено у вигляді ієрархічної структури.

На перетинах осей вимірювань (Dimensions) розташовуються дані, що кількісно характеризують аналізовані факти, - метрики / заходи (Measures). Це можуть бути обсяги продажів, виражені в одиницях продукції або в грошовому вираженні, залишки на складі, витрати тощо.

****

Рис.1. Представлення даних у вигляді гіперкуба

Над таким гіперкубом можуть виконуватися наступні операції.

**Зріз** (Slice) - формування підмножини багатовимірного масиву даних, якому відповідає єдине значення одного або декількох елементів вимірів, що не входять в цю підмножину. Якщо розглядати термін "зріз" з позиції кінцевого користувача, то найбільш часто його роль грає двовимірна проекція куба.



Рис.2. Операцій зрізу

**Обертання** (Rotate) - зміна розташування вимірів, представлених у звіті або на відображуваній сторінці. Наприклад, операція обертання може полягати в перестановці місцями рядків і стовпців таблиці або переміщенні визначених вимірів в стовпці або рядки створюваного звіту, що дозволяє надавати йому зручного вид. Крім того, обертанням кубу даних є переміщення позатабличного виміру на місце вимірів, представлених на відображуваної сторінці, і навпаки (при цьому позатабличного вимір стає новим виміром рядка або виміром стовпця).



Рис.3. Операція обертання

**Консолідації** (Drill Up) і деталізації (Drill Down) - операції, які визначають перехід вгору у напрямку від детального (down) представлення даних до агрегованого (up) і навпаки, відповідно. Напрямок деталізації (узагальнення) може бути задано як по ієрархії окремих вимірювань, так і згідно іншим відносинам, встановленим в рамках вимірів або між вимірами.



Рис.4. Операція консолідацїі та деталізації

Виділяють три основних способи реалізації СД:

MOLAP - багатовимірний (multivariate) OLAP. Для реалізації багатовимірної моделі використовують багатовимірні БД;

ROLAP - реляційний (relational) OLAP. Для реалізації багатовимірної моделі використовують реляційні БД;

HOLAP - гібридний (hybrid) OLAP. Для реалізації багатовимірної моделі використовують і багатовимірні, і реляційні БД.

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 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. Зазвичай атрибути представляються текстовими полями з дискретними значеннями. Наприклад, габарити упаковки товару, запах товару.

**Вимір** (dimension) - це інтерпретація факту з деякою точки зору в реальному світі. Виміри, подібно атрибутам, містять текстові 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;

• **неаддитивну факти** (Non-additive facts). Факт називається неаддитивну, 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.

Схема "зірка" має одну таблицю фактів і кілька таблиць вимірів. Таблиці вимірів є денормалізованнимі.



Рис.4. Схема зірка

Схема "сніжинка" має одну таблицю фактів і кілька нормалізованих таблиць вимірів.

****

Рис.5. Схема сніжинка