

Організація баз даних

Автор

Голуб Б.Л.

ЛЕКЦІЯ 2

Основні концепції реляційної бази даних

Модель даних

Модель даних - теорія подання та обробки даних в системі управління базами даних (СКБД), яка включає, щонайменше, три аспекти:

- аспект *структури*: методи опису типів і логічних структур даних в базі даних;
- аспект *маніпуляції*: методи маніпулювання даними;
- аспект *цілісності*: методи опису і підтримки цілісності бази даних.

Основи реляційної моделі

Реляційна модель даних (РМД) - логічна модель даних, прикладна теорія побудови баз даних, яка є додатком до завдань обробки даних таких розділів математики, як теорії множин і логіки першого порядку.

На реляційної моделі даних будуються *реляційні бази даних*.

Концепції реляційної моделі

- У будь-якій *реляційній базі даних* користувач сприймає базу даних як деяку сукупність *таблиць*.
- Кожна таблиця в реляційній моделі називається *відношенням*.
- Відношення має вигляд двомірної таблиці, рядки якої – окремі *записи*, а стовпчики з назвами – *атрибути*.
- Сукупність значень, на якому визначаються атрибути, називається *доменом*.
- Кількість атрибутів у відношенні отримала назву *ступеня* відношення.
- Кожний рядок відношення, тобто кожний рядок таблиці, має назву *кортеж*. Кортежі можуть бути записані в довільному порядку у відношенні, але від цього структура відношення не змінюється.
- Кількість кортежів у відношенні має назву *кардинальність*.

Перші два правила Кодда

- *Фундаментальне правило* (Foundation Rule)

Реляційна СУБД повинна бути здатна повністю керувати базою даних, використовуючи зв'язки між даними.

- *Інформаційне правило* (Information Rule)

Інформація повинна бути представлена у вигляді даних, що зберігаються в осередках. Дані, що зберігаються в комірках, повинні бути атомарним. Порядок рядків в реляційній таблиці не повинен впливати на зміст даних.

Правило інформації. Вся інформація в базі даних повинна бути представлена виключно на логічному рівні і лише одним способом - у вигляді значень, що містяться в таблицях. Фактично це неформальне визначення реляційної бази даних.

Нормалізація даних

Головний постулат БД – дані не повинні бути *надлишковими*, тобто необхідно зводити до мінімуму повторення даних.

Нормалізація – це процес скорочення повторень інформації в базі даних. Нормалізуються в БД не лише дані, а й імена.

Розрізняють декілька етапів процесу нормалізації. У результаті виконання кожного з них створюється, так звана, *нормальна форма*. Говорять, що нормальна форма – це міра глибини, до якої повинна бути виконана нормалізація БД.

Перша нормальна форма (1NF)

Метою першої нормальної форми є розділення БД на логічні одиниці – *таблиці*.

Після створення таблиць для більшості з них будуть визначені *ключові поля*.

Іншими словами, для переходу до першої нормальної форми БД повинна бути розбита на декілька логічних одиниць, в кожній з них визначений ключ і відсутні групи, що повторюються.

Друга нормальна форма (2NF)

Метою другої нормальної форми є *відокремлення* даних, які майже *не залежать від ключового поля*, і розміщення таких даних в окрему таблицю.

Іншими словами, друга нормальна форма отримується з першої шляхом подальшого розділення таблиць на менші за розміром.

Третя нормальна форма (3NF)

Метою третьої нормальної форми є видалення із таблиць даних, які не залежать від ключа.

У теорії реляційних баз даних за звичай виділяються ще такі нормальні форми:

- *нормальна форма Бойса-Кодда* (BCNF);
- *четверта нормальна форма* (4NF);
- *п'ята нормальна форма*, або нормальна форма проекції-об'єднання (5NF або PJ/NF).

Основні властивості нормальних форм

- кожна наступна нормальна форма в деякому розумінні *краще* попередньої;
- при переході до наступної нормальної форми властивості попередніх нормальних форм *зберігаються*.

Переваги нормалізації

Нормалізація має цілий ряд переваг:

- *краща загальна організація* БД;
- *скорочення кількості непотрібних повторень* БД;
- *погодження даних* усередині таблиць;
- *більш гнучка структура* БД;
- *ефективні можливості забезпечення безпеки та надійності* БД.

Недоліки нормалізації

Нормалізація має один суттєвий недолік — сповільнення роботи самої БД.

Чим більше невеликих таблиць пов'язані між собою, тим довший шлях необхідно подолати для пошуку тої чи іншої інформації.

Процес денормалізації

Денормалізація – це процес модифікації структури таблиць нормалізованої БД *з метою підвищення продуктивності* за рахунок дозволу деякого керованого надлишку даних.

Денормалізація передбачає *об'єднання* деяких таблиць та створення таблиць з даними, що дублюються, з метою зменшення кількості зв'язаних таблиць, що використовуються при пошуку даних. При цьому треба враховувати, що отримавши при збільшенні швидкості роботи БД, втрачається простота управління такими даними.