Лекция 1

Введение в базы данных. Модели данных. Реляционные базы данных.

СУБД

Цель

Сформировать у студентов базовое понимание концепции баз данных, познакомить с различными моделями данных, дать фундаментальные знания о реляционных базах данных и системах управления базами данных (СУБД).

Основные вопросы

1. Понятие и назначение баз данных
2. Эволюция моделей данных: иерархическая, сетевая, реляционная
3. Основные концепции реляционной модели данных
4. Компоненты и функции СУБД
5. Преимущества и недостатки различных подходов к организации данных

Введение в базы данных

База данных (БД) - это организованная совокупность структурированных данных, хранящихся в памяти вычислительной системы и отображающих состояние объектов и их взаимосвязей в рассматриваемой предметной области.

Основные характеристики баз данных:

- Постоянство хранения данных;
- Структурированность и организация;
- Независимость от прикладных программ;
- Многопользовательский доступ с контролем параллелизма;
- Защита от несанкционированного доступа;
- Обеспечение целостности и восстановления данных;

Историческая эволюция моделей данных

Иерархическая модель организует данные в структуру дерева, где каждый узел имеет только одного предка. Эта модель была популярна в 1960-1970-х годах.

Основные особенности:

- Строгая древовидная структура
- Связи "один-ко-многим"
- Быстрый доступ к данным по известному пути
- Сложности при изменении структуры
- Пример: IMS (Information Management System) от IBM

Недостатки:

- Сложность представления сложных связей
- Избыточность данных
- Жесткость структуры

Сетевая модель данных

Сетевая модель расширяет иерархическую, позволяя узлам иметь несколько предков. Разработана в рамках конференции CODASYL.

Основные особенности:

- Графовая структура данных
- Связи "многие-ко-многим"
- Более гибкая, чем иерархическая модель
- Сложность в проектировании и использовании
- Пример: IDMS (Integrated Database Management System)

Реляционная модель данных

Предложена Эдгаром Коддом в 1970 году в статье "A Relational Model of Data for Large Shared Data Banks". Стала доминирующей моделью на десятилетия вперед.

Основные концепции:

- Данные представляются в виде таблиц (отношений)
- Строки (кортежи) и столбцы (атрибуты)
- Математическая основа - теория множеств и реляционная алгебра
- Декларативный язык запросов SQL

Отношение (Relation) - фундаментальное понятие, представляющее собой таблицу со следующими свойствами:

- Каждая строка уникальна;
- Порядок строк не определен;
- Все значения в столбце однородны;
- Имена столбцов уникальны в пределах таблицы;

Схема отношения - описание структуры отношения, включающее:

- Имена атрибутов;
- Типы данных атрибутов;
- Ограничения целостности;

Ключи:

- Первичный ключ - уникальный идентификатор кортежа;
- Внешний ключ - атрибут, ссылающийся на первичный ключ другого отношения;

Преимущества реляционной модели:

- Простота и математическая обоснованность
- Независимость данных
- Гибкость при изменении структуры
- Мощный язык запросов SQL
- Широкое распространение и поддержка

Системы управления базами данных (СУБД)

СУБД - комплекс программных средств, предназначенных для создания, ведения и использования баз данных.

Основные функции СУБД:

Управление данными:

- Создание, изменение и удаление структур данных;
- Манипулирование данными (вставка, обновление, удаление);
- Оптимизация выполнения запросов;

Управление транзакциями:

- ACID-свойства:
  - Атомарность - транзакция выполняется полностью или не выполняется вовсе;
  - Согласованность - транзакция переводит базу данных из одного согласованного состояния в другое;
  - Изолированность - параллельные транзакции не мешают друг другу;
  - Долговечность - результаты завершенной транзакции сохраняются после сбоев;

Управление параллельным доступом:

- Механизмы блокировок;
- Протоколы управления параллелизмом;
- Разрешение конфликтов;

Восстановление после сбоев:

- Журналирование изменений;
- Резервное копирование;
- Восстановление после аппаратных и программных сбоев;

Архитектура СУБД

Трехуровневая архитектура ANSI/SPARC:

1. Внутренний уровень - физическое представление данных на носителе
2. Концептуальный уровень - логическая структура всей БД
3. Внешний уровень - представления данных для конкретных пользователей

Современные тенденции

NoSQL базы данных - альтернатива реляционным БД для специфических задач:

- Документоориентированные (MongoDB)
- Ключ-значение (Redis)
- Колоночные (Cassandra)
- Графовые (Neo4j)

NewSQL - системы, сочетающие масштабируемость NoSQL и ACID-гарантии реляционных СУБД.

Контрольные вопросы

1. Что такое база данных и каковы ее основные характеристики?
2. В чем основные различия между иерархической, сетевой и реляционной моделями данных?
3. Перечислите основные понятия реляционной модели данных.
4. Каковы основные функции СУБД?
5. Что такое ACID-свойства транзакций? Раскройте каждое из них.
6. Какие преимущества реляционной модели способствовали ее широкому распространению?

Литература

1. Дейт К. Дж. Введение в системы баз данных. - 8-е изд. - Глава 1-2.
2. Коннолли Т., Бегг К. Базы данных: проектирование, реализация и сопровождение. - 3-е изд. - Глава 1-2.
3. Кузнецов С.Д. Основы баз данных. - 2-е изд. - Глава 1-2.
4. Хомоненко А.Д., Цыганков В.М., Мальцев М.Г. Базы данных. - Глава 1.