

# Базы данных, лекция 10

@mikhirurg

April 2020

## 1 Распределённая база данных

"Узкое место" современных информационных систем и баз данных - это память. За все время развития информационных технологий в последние 20-30 лет мы не видели взрывного роста скорости доступа к данным со стороны носителей информации. Носители информации с быстрым доступом имеют высокую стоимость. В результате этого появилась идея распределения данных по нескольким носителям информации.

*Распределённая БД* - это набор логически связанных между собой разделяемых данных и их описаний, которые физически распределены по нескольким вычислительным узлам некоторой сети.

Такая база данных, в отличие от распределённой файловой системы, имеет высокоуровневый доступ к данным и модель данных.

### 1.1 Фрагментация данных

- Горизонтальное фрагментирование  
*фрагментация по хронологическому порядку появления данных.*
- Вертикальное фрагментирование  
*Пример: таблица пользователей, фрагментация по столбцам. какая-то информация о пользователе требуется реже другой.*
- Репликация фрагментов  
Необходимо поддерживать актуальность копий

### 1.2 Стратегии размещения данных

- Раздельное (фрагментированное) размещение.  
*Потеря связи с одним узлом, нарушает работу всей системы  
Низкая скорость доступа*
- Размещение с полной репликацией  
*Копирование всей базы данных во всех узлах.*
- Размещение с выборочной репликацией  
*Необходима оптимальная стратегия размещения данных*

### 1.3 Распределённая СУБД

Распределённая СУБД - это комплекс программ, предназначенный для управления распределённой БД и позволяющий сделать распределённость информации "прозрачной" для конечного пользователя.

Виды прозрачности:

- Прозрачность фрагментации
- Прозрачность расположения фрагмента
- Прозрачность количества реплик фрагмента
- Прозрачность контроля доступа

Гомогенные и гетерогенные распределённые СУБД: разделение по модели хранения данных в узлах.

#### 1.4 12 правил распределённой БД от К. Дейта



Рис. 1: Кристофер Дейт

1. Локальная автономность
2. Отсутствие опоры на центральный узел  
*Не должно быть одного узла, который бы распределял данные между другими.*
3. Непрерывное функционирование  
*Не должно быть плановых остановок системы.*
4. Независимость от расположения
5. Независимость от фрагментации
6. Независимость от репликации
7. Обработка распределённых запросов  
*Работа с несколькими узлами*
8. Обработка распределённых транзакций
9. Независимость от типа оборудования
10. Независимость от сетевой архитектуры
11. Независимость от операционной системы
12. Независимость от типа СУБД

## 1.5 Распределённые запросы

Обработка распределённого запроса:

- Определение фрагмента
- Определение реплики фрагмента
- Определение местоположений постороения временных структур данных и маршрутизация данных

Распределённые транзакции. Двухфазное завершения транзакций. Необходимо получить подтверждение от узлов о благополучном завершении транзакции.

## 1.6 Преимущества распределённых БД

- Отражение структуры организации
- Разделяемость и локальная автономность
- Повышение доступности данных
- Повышение надёжности
- Повышение производительности
- Модульность системы

## 1.7 Недостатки распределённых БД

- Повышение сложности
- Увеличение стоимости
- Сложности в обеспечении защиты
- Сложности в обеспечении целостности данных
- Высокая стоимость стандартизации и поддержки стандартов