

Базы данных, лекция 11

@mikhirurg

May 2020

1 Предпосылки возникновения NoSQL БД

1.1 Увеличение объемов данных

За 2013-18 год в хранилищах данных было сохранено больше информации, чем за всю предыдущую историю человечества.

- Закон удвоения количества данных за 2 года.
- "Закон Мура переключка с удвоением данных.
- К концу 2020 года совокупный объём данных превысит 40 зеттабайт.
- 40% этих данных относятся к сфере развлечений.
- Только 23% этих данных потенциально полезны
- Только 3% доступных данных активно используются
- Если в 2005 году объем автоматически генерируемых данных составлял 11%, то в 2020 уже 40%.

Например логи, автоматические бекапы и тд.

1.2 Взаимосвязь данных

- Прямые связи:
Страницы в Интернет, тэги, словари, онтологии и др.
- Косвенные связи:
Связь цифровых следов, группирование людей по схожести поведения.

1.3 Слабоструктурированная информация(пример)

- Параметры, общие для всей бытовой электроники.
- Параметры, общие для вида техники.
- Параметры, общие для конкретного класса этого вида техники.
- Параметры, характерные для конкретного бренда или модельного ряда.

1.4 Изменение в архитектуре информационных систем

- Многопользовательская архитектура с центральным узлом хранения и обработки данных
Мейнфреймы
- Многозвенные клиент-серверные архитектуры (1990-2000)
- Сервис-ориентированные архитектуры (настоящее время)

1.5 Общие характеристики NoSQL БД

- Отказ от использования баз SQL
- Неструктурированность схемы данных (schemaless)
- Использование агрегатов для представления данных
- Слабые ACID свойства

1.5.1 Использование агрегатов для представления данных

Данные в нормализованных отношениях	Данные в агрегатах
При обновлении сохраняется целостность данных Оптимальная производительность для широкого спектра запросов	Необходимо контролировать целостность данных Оптимизация только под определённый вид запросов
Неэффективность для распределённых хранилищ Низкая скорость чтения многосвязных данных Несоответствие объектной модели приложения	Большие возможности по оптимизации чтения данных в распределённых хранилищах Возможность существенно повысить скорость чтения некоторых агрегатов данных Естественное хранение объектов приложения, поддержка атомарности объектов

1.5.2 Слабые ACID свойства

- ACID: Atomicity, Consistency, Isolation, Durability –
Атомарность, Согласованность, Изоляция, Долговечность
- BASE: Basic Availability, Soft State, Eventual Consistency –
Базовая доступность, гибкое состояние, согласованность в конечном счёте.

Вспомним
свойства
транзакций