**Базы данных и SQL**

База данных (БД) – хранилище данных, в котором данные могут быть взаимосвязаны и структурированы определенным образом (т.е. БД – сами данные).

Система управления базой данных (СУБД) – инструмент для управления данными в БД. Отвечает за поддержку языка для запросов (обеспечивает интерфейс для взаимодействия с данными), за механизм хранения и извлечения данных, за оптимизацию процессов извлечения данных и др.

**Виды БД:**

*Без использования СУБД*

|  |  |
| --- | --- |
| Файловые | Данные хранятся в виде текста в файле. |
| Иерархические | Имеет древовидную структуру (узел может иметь только одного родителя) Пример: файловая система. |
| Сетевые | Имеет структуру графа (узел может иметь более одного родителя) |

*С использованием СУБД*

|  |  |
| --- | --- |
| Реляционные | Структура в виде таблиц, связанных между собой |
| NoSQL | Общее название для нереляционных БД. |

**Типы СУБД:**

**NoSQL** (not only SQL) – Нереляционные СУБД.

1. Используют модель данных отличную от табличной,
2. Используется язык отличный от SQL.
3. Отсутствие жесткой схемы данных (Следствие отсутствия схемы — эффективность работы с разреженными (sparse) данными. Если в одном документе есть поле date\_published, а во втором — нет, значит никакого пустого поля date\_published для второго создано не будет)
4. Горизонтально масштабируемы – данные легко распределяются между множеством серверов (шардирование)
5. Жертвуют консистентностью в пользу доступности и масштабируемости. Применяется модель BASE.

1) Ключ-значение – база данных организована в виде словаря (пара ключ-значение). Подходит для кэширования. [Redis]

Высокая скорость чтения и записи за счет хранения в оперативной памяти, легкое масштабирование. Не подходит для сложных структур (сущности с разными типами данных) и сложных запросов.

2) Документ-ориентированная – данные хранятся в виде документов, где каждый документ – автономный набор данных [MongoDB].

3) Колонно-ориентированные – хранят данные по колонкам, что полезно для аналитических запросов [HBase, Cassandra].

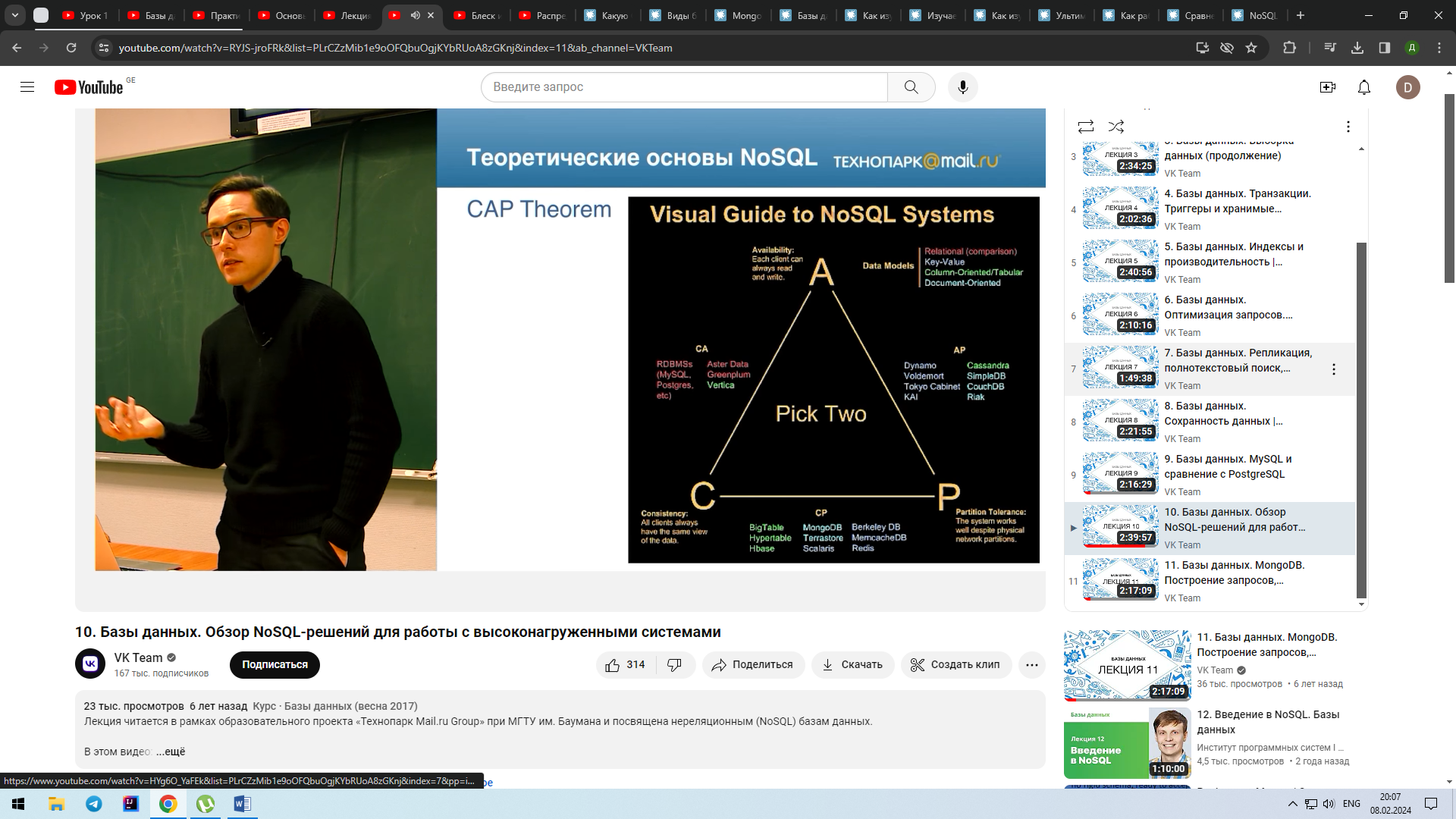
4) Графовые – имеют структуру графа, необходимы в ситуации, когда нужно отразить большое кол-во связей между сущностями. [Neo4j, AllegroGraph]

5) Поисковые – для осуществления поиска по большому объему неструктурированных данных [Elasticsearch].

**Реляционные**

1. Модель данных – данные организованы в виде отношений (таблиц)
2. Использование языка запросов SQL (операции над данными по законам реляционной алгебры)
3. Жесткая схема
4. Имеют сложности с шардированием, но реплицируемы.
5. Жертвуют доступностью и устойчивостью к разделению в пользу согласованности данных (ACID)
6. Хранение данных построчное (чаще всего) или колоночное.
   1. Построчное хранение данных – быстрая вставка, но накладные расходы при чтении данных.
   2. Колоночное – эффективная работа с аналитическими запросами [ClickHouse].

[Про типы СУБД.](https://habr.com/ru/companies/amvera/articles/754702/)



**CAP теорема –** распределенная система не может в полной мере обеспечивать свойства consistency (согласованность), availability (доступность), partition tolerance (устойчивость к разделению). Может быть сосредоточена в полной мере на 2х свойствах.

Consistency – Гарантирует, что запрос к системе выдаст последнюю версию изменений, даже если изменение было произведено на другом узле – актуальность данных.

Availability - Каждый запрос завершается успешным ответом.

Partition tolerance – Система продолжает функционировать при наличии разделения/разрыва сети между узлами

CA системы – такая система подразумевает, что если выходит из строя один из компонентов системы, то она перестает работать

CP – система не гарантирует, что ответ от нее будет получен в разумные сроки (пока запрашиваемые данные не будут синхронизованы на всех узлах, клиент будет ждать ответ на запрос)

AP – система не гарантирует консистентность данных, такие системы не подразумевают того, что данные будут всегда противоречивы, они говорят о том, что данные не будут констстентны в каждый момент времени

Данная «теорема» является условной, не является теоремой в математическом понимании и содержит много противоречий [1](https://habr.com/ru/articles/231703/), [2](https://habr.com/ru/articles/322276/), [3](https://habr.com/ru/articles/328792/), [4](https://habr.com/ru/articles/258145/)

**BASE** – альтернативный подход к построению распределенных систем или свойства распределенных систем, в которых Consistency и Availability выполнены не в полной мере.

Basically Available (Базовая доступность) – на любой запрос будет дан ответ, но он может отражать неконсистентные данные.

Soft state (Неустойчивое состояние) – данные в системе могут меняться, даже если изменений в системе нет

Eventual consistency (согласованность в конечном итоге) – при отсутствии изменений в системе, в конечном итоге, все данные будут согласованы.

**OLAP/OLTP –** подходы для обработки данных в БД

*OLAP* **(**Online Analytical Processing**) –** Предназначен для аналитической обработки данных: анализ и агрегация больших объемов данных, заточен под чтение, а не под изменение данных.

*OLTP*(Online Transaction Processing) – Предназначен для обработки транзакций в реальном времени.