



### Знакомство с Clickhouse

В этой таблице вы можете увидеть, что наборы значений полей лежат по строчкам. На уроке были показаны файлы, одноименные с наборами полей, создающиеся при вставке данных.

**Структура строковой БД отличается от привычной столбцовой:**

Строка:	#0	#1	#2	#N
WatchID:	89354350662	90329509958	89953706054	...
JavaEnable:	1	0	1	...
Title:	Investor Relations	Contact us	Mission	...
GoodEvent:	1	1	1	...
EventTime:	2016-05-18 05:19:20	2016-05-18 08:10:20	2016-05-18 07:38:00	...

#### Clickhouse плюсы и минусы

При выборе ClickHouse для хранения данных следует учитывать несколько важных моментов. Одним из главных преимуществ ClickHouse является то, что он поддерживает SQL-подобный язык запросов. Это делает его понятным и удобным для большинства разработчиков и аналитиков данных, которые уже знакомы с языком SQL. Кроме того, ClickHouse имеет широкий набор типов данных, поддерживая различные форматы, такие как числа, строки, даты и времена, а также массивы и структуры данных. Еще одним важным преимуществом ClickHouse является возможность компиляции запросов. Это позволяет оптимизировать производительность выполнения запросов, ускоряя их исполнение. ClickHouse также использует векторную обработку запросов, что означает, что вместо обработки данных по одному элементу, он обрабатывает их в виде векторов. Это значительно увеличивает производительность системы.

Кроме того, ClickHouse поддерживает компрессию данных. Это позволяет существенно сократить объем хранимых данных, что имеет большое значение при работе с большими объемами данных. ClickHouse также эффективно работает с большими данными, особенно если речь идет о массовой агрегации и фильтрации данных.

Важными функциями ClickHouse являются разреженные индексы, которые позволяют эффективно работать с разреженными данными и способствуют снижению потребления ресурсов. ClickHouse также поддерживает репликацию данных, что обеспечивает отказоустойчивость и повышает доступность системы. Кроме того, ClickHouse поддерживает шардирование, позволяя распределить данные и нагрузку на несколько узлов, что способствует масштабируемости системы.

К сожалению, у ClickHouse есть и некоторые ограничения и недостатки. Например, он не поддерживает транзакции, что может быть проблематично в некоторых сценариях работы с данными, особенно если требуется гарантированная целостность данных. Также массовое изменение данных может быть достаточно трудоемким процессом в ClickHouse.



В целом, ClickHouse является мощной и эффективной системой для хранения и анализа данных. Однако перед его выбором необходимо тщательно рассмотреть особенности и требования вашего проекта, чтобы убедиться, что ClickHouse является оптимальным решением для ваших задач.

#### Плюсы:

- SQL-подобие
- Разнообразие хранимых типов данных
- Компиляция запросов
- Векторная обработка запросов
- Компрессия данных
- Работа с большими данными
- Разреженный индекс
- Репликация
- Шардирование
- etc

#### Минусы:

- Транзакции
- Массовое изменение данных

### Выводы, влияющие на работу с clickhouse

Первый важный вывод - на каждую вставку данных создается свой кусок данных. ClickHouse использует сжатие и компрессию данных для оптимизации производительности, поэтому данные хранятся не в виде единого файла, а разбиваются на небольшие куски. Это позволяет быстро и эффективно обращаться к нужным данным при выполнении запросов.

Второй вывод связан с процессом вставки данных. Каждая вставка в ClickHouse приводит к созданию не менее одного файла на каждую колонку в таблице. То есть, если у вас есть таблица с несколькими колонками, каждая вставка будет добавлять файлы для каждой из этих колонок. Необходимо учитывать это при проектировании схемы таблицы и решении о типе данных для каждой колонки.

Еще один важный вывод связан с процессом мержа данных. ClickHouse автоматически объединяет различные части данных в фоновом режиме для оптимизации производительности. Однако, частые мелкие вставки могут негативно сказываться на производительности системы. Рекомендуется проводить вставку данных хотя бы раз в секунду, чтобы сократить количество мелких кусков данных и улучшить производительность.

Еще один рекомендуемый подход - вставлять большие куски данных за раз, примерно около 1 миллиона строк. Большие вставки позволяют эффективно использовать преимущества компрессии данных и ускоряют процесс вставки.

Наконец, стоит упомянуть о команде "optimize table <имя\_таблицы> final". Эта команда используется для проведения внеплановых мерджей данных в таблице. Рекомендуется использовать эту команду, когда таблица не используется или используется минимально, чтобы избежать негативного влияния на производительность системы.

В итоге, при работе с ClickHouse необходимо учитывать возможные негативные последствия мелких и частых вставок данных. Рекомендуется проводить вставки большими кусками данных, оптимизировать таблицы, и проектировать схему таблицы с учетом особенностей ClickHouse. Это позволит эффективно использовать преимущества системы и обеспечить высокую производительность работы с данными.