

# ClickHouse

Александр Смирнов

17.09.2020

# ClickHouse

- ▶ Распределенная аналитическая столбцовая (column-oriented) СУБД

# Почему column-oriented?

- ▶ В обычной row-oriented СУБД данные хранятся по строкам
- ▶ Хорошо подходит для обычных сценариев
  - ▶ изменить поле у пользователя
- ▶ Плохо подходит для задачи аналитики
  - ▶ пусть много столбцов, e.g. время, браузер, модель телефона
  - ▶ строки — события, e.g. просмотр страницы
  - ▶ нужен отчет по моделям телефона

# Почему column-oriented?

- ▶ Подходит система где данные хранятся по столбцам

# История

- ▶ Yandex.Metrika
- ▶ Данные по действиям пользователей на сайтах
  - ▶ как сохранить данные так, чтобы быстро строить отчеты
  - ▶ изначально предагрегировали данные, но пользователи хотели больше вариативности
- ▶ Пробовали разные столбцовые субд, не подошло

# Распространение

- ▶ Open-source
- ▶ Есть своя небольшая ниша
- ▶ В России
  - ▶ Яндекс
  - ▶ Mail
  - ▶ СКБ Контур
  - ▶ Rambler
- ▶ За рубежом
  - ▶ Cloudfare
  - ▶ Wikimedia
  - ▶ Nvidia

# Преимущества

- ▶ Скорость
- ▶ Гибкость
  - ▶ доступны все данные, что хотим то и делаем с ними
- ▶ Масштабируемость
  - ▶ распределенное хранение данных
- ▶ Цена
  - ▶ цена разработчиков, не дорого

# Недостатки

- ▶ Ограниченная поддержка JOIN
- ▶ Нет UPDATE и DELETE
- ▶ Слабая совместимость со стандартом SQL