

Всю открытую информацию о внедрении КХ в компании я нашел тут <https://clickhouse.com/docs/en/about-us/adopters>.

1. OK.RU

Индустрия: соцсеть.

Кол-во серверов: 72

Объем данных: 810 TB compressed, 50bn rows/day, 1.5 TB/day

Ссылка: <https://assets.ctfassets.net/oxjq45e8ilak/4JPHkbJenLgZhBGGyyonFP/57472ec6987003ec4078d0941740703b/ClickHouse.pdf>

Особенности: Используют как OLAP хранилище данных, на 2021 год данных у них 810 тб. Кафкой наполняют таблицы, Спарком рассчитывают.

2. SBER

Индустрия: банк.

Кол-во серверов: 128

Объем данных: >1 PB

Ссылка: <https://career.habr.com/vacancies/1000073536>, <https://rabota.sber.ru/search/4327580/>

Особенности: Кафкой наполняется, GreenPlum рассчитывают витрины, Apache Superset визуализируют.

3. Yandex Metrica

Индустрия: веб аналитика.

Кол-во серверов: 630 серверов в одном кластере, 360 в другом, 1862 серверов в одном отделе.

Объем данных: 133 PiB / 8.31 PiB / 120 trillion records

Ссылка: <https://presentations.clickhouse.com/meetup40/introduction/#13>

Особенности: Думаю не ошибусь, если скажу, что это самый большой кластер клика из когда либо собранных.

Именно для Яндекс метрики придумали ClickHouse. Кстати имя БД придумали довольно легко, если перевести то дословно получится "Дом кликов". т.к. Яндекс метрика это веб аналитика, 99% данных которые они собирают - это клики юзеров по странице.