

**Я**ндекс

Яндекс

Yandex DataLens

# Опыт использования ClickHouse при построении BI-системы Yandex DataLens

Иван Москвин, CTO Yandex DataLens

# Содержание

- 01 | Пара слов про DataLens
- 02 | ClickHouse как аналитическое хранилище данных
- 03 | Формирование запросов по пользовательским формулам
- 04 | Загрузка аналитических данных в ClickHouse
- 05 | Геокодинг

01



# Про Yandex DataLens

Цель продукта

Принципиальная схема работы

**Заказы** Выручка Рекламные кампании Колл-центр Курьеры

Дата: 01.07.2019 - 31.08.2019

| Точка продаж | Количество заказов | Сумма заказа |
|--------------|--------------------|--------------|
| Охра         | 141                | 37 733,8     |
| Морковный    | 119                | 32 455       |
| Лайм         | 110                | 30 650,4     |
| Серебряный   | 107                | 28 048,2     |
| Оливковый    | 105                | 27 915,4     |
| Синий        | 101                | 30 891,8     |

Город: Казань

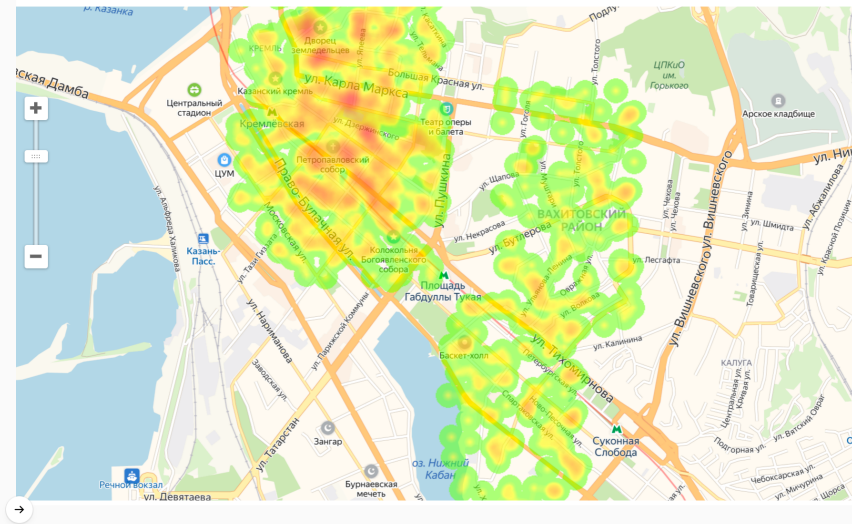
Точка продаж: —

Категория продукта: —

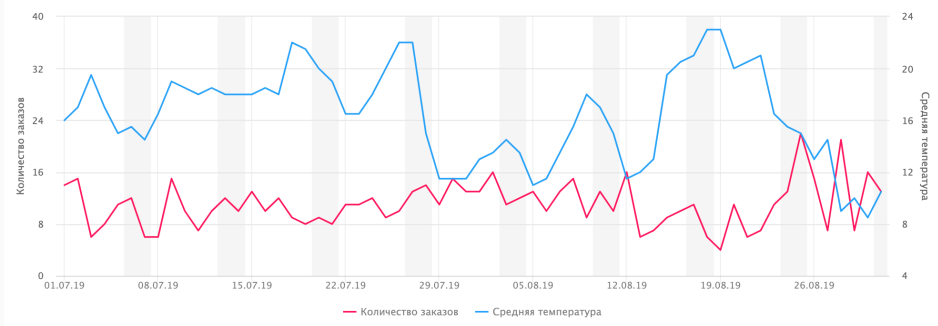
Продукт: —

Сегмент клиента: —

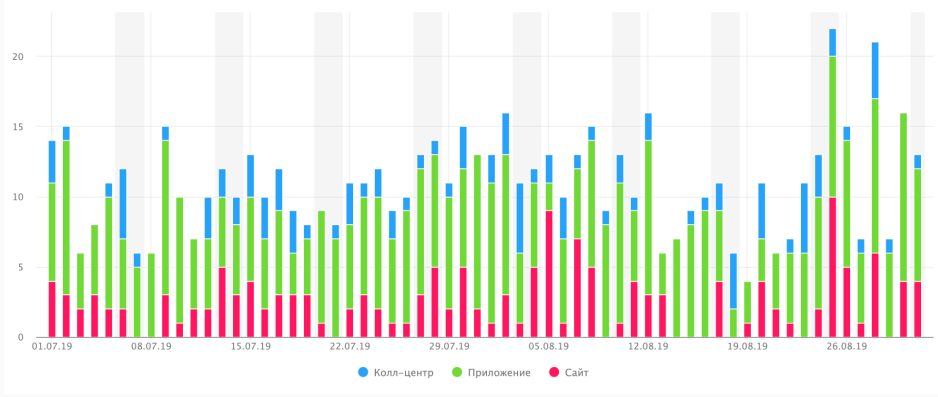
Тепловая карта заказов



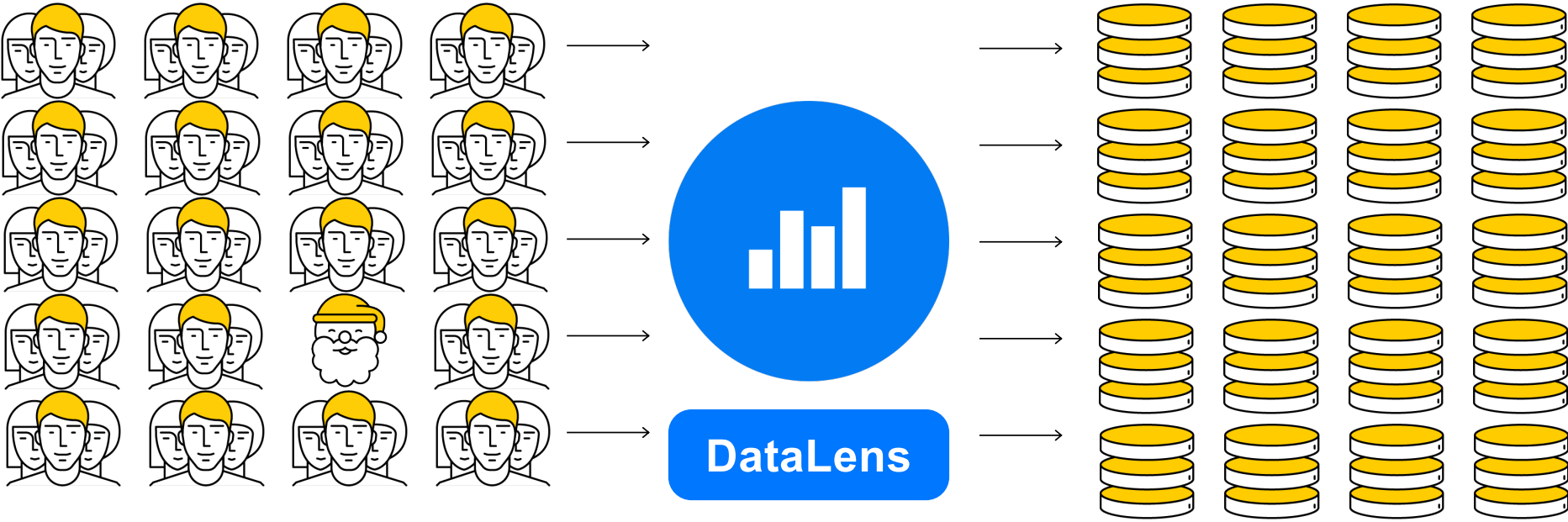
Динамика заказов (RKeeper) + Температура воздуха, °C



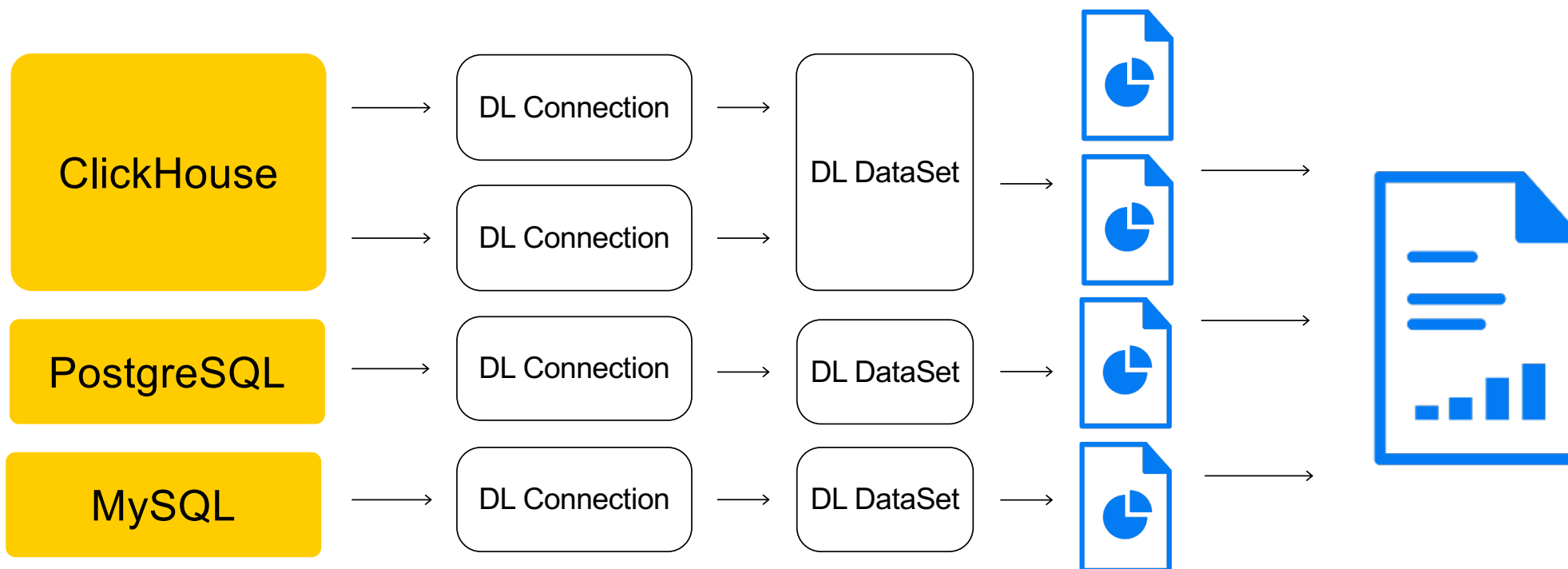
Заказы по каналам (Metrica, AppMetrica)



# Multitenant



# Принципиальная схема работы



02



# ClickHouse как аналитическое хранилище данных

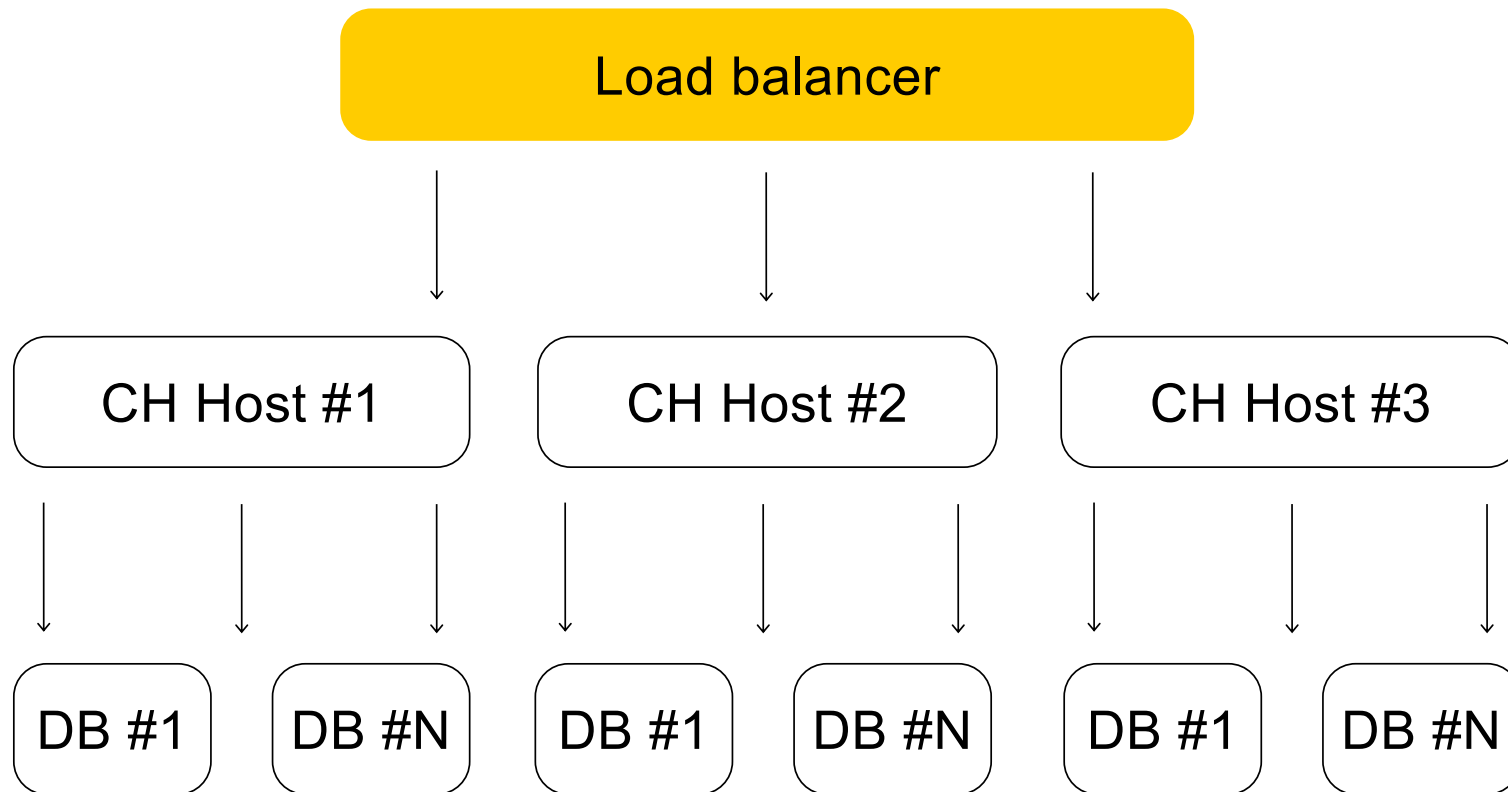
Multitenant кластера для хранения пользовательских данных

Материализация данных из удаленных источников данных

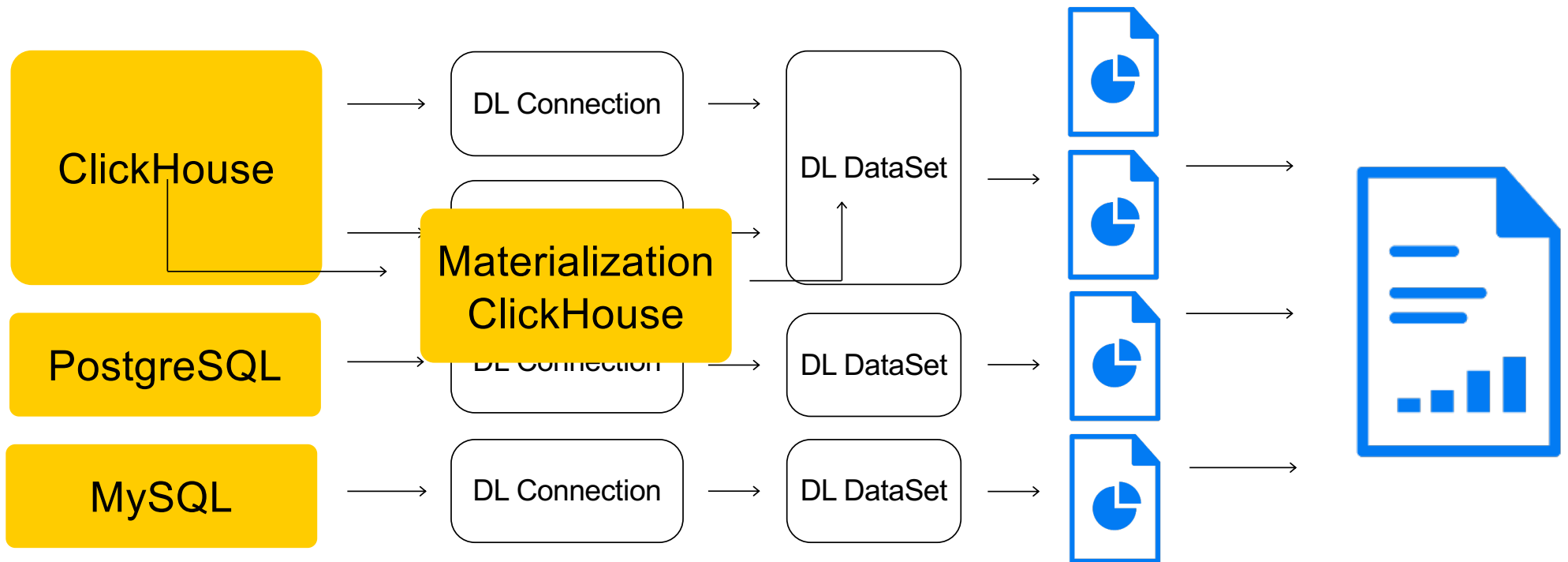
Особенности переноса данных и изоляции пользователей



# Отказоустойчивость



# Материализация



# 03



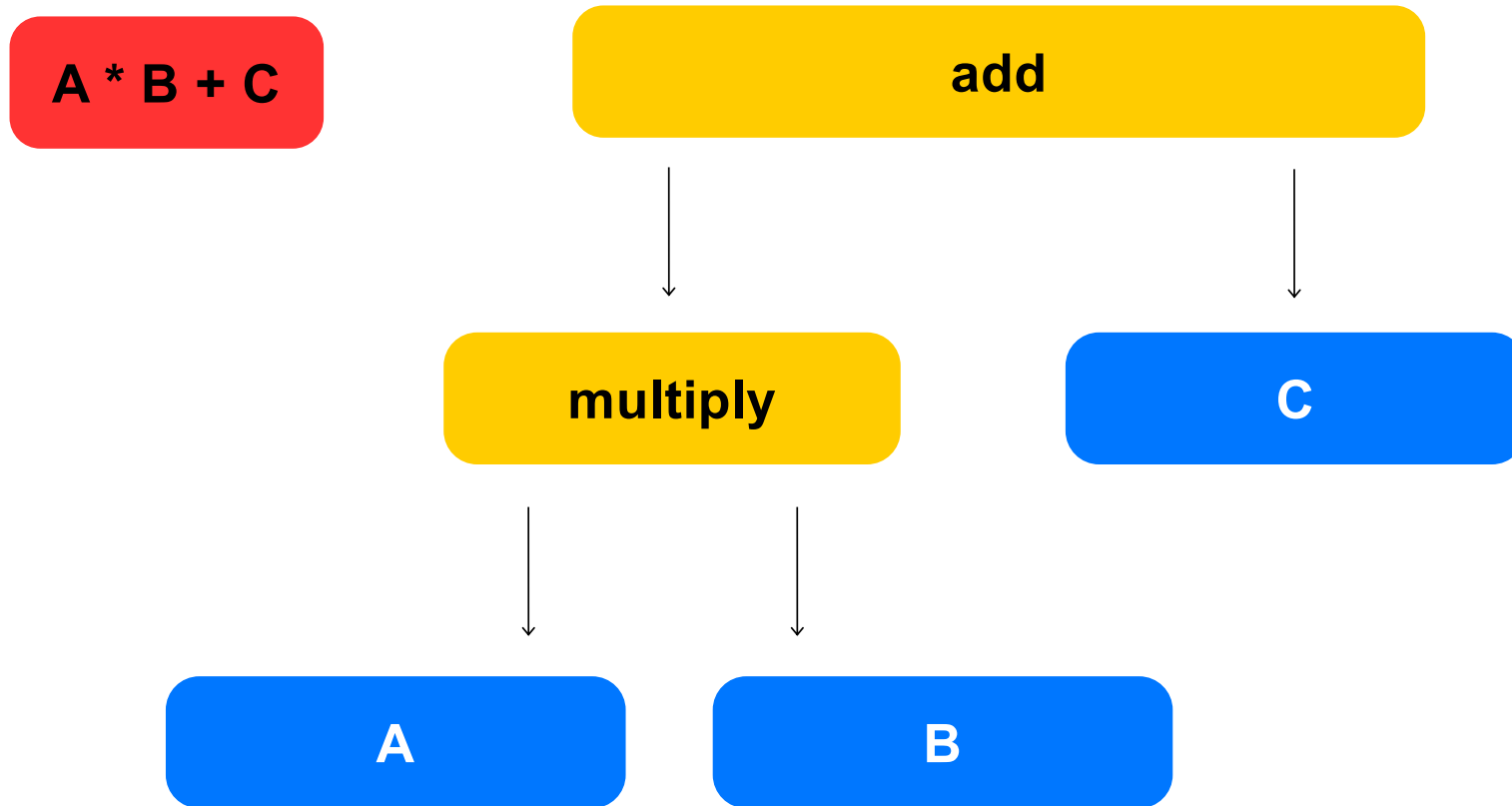
## **Формирование запросов по пользовательским формулам**

Разбор формул в AST дерево

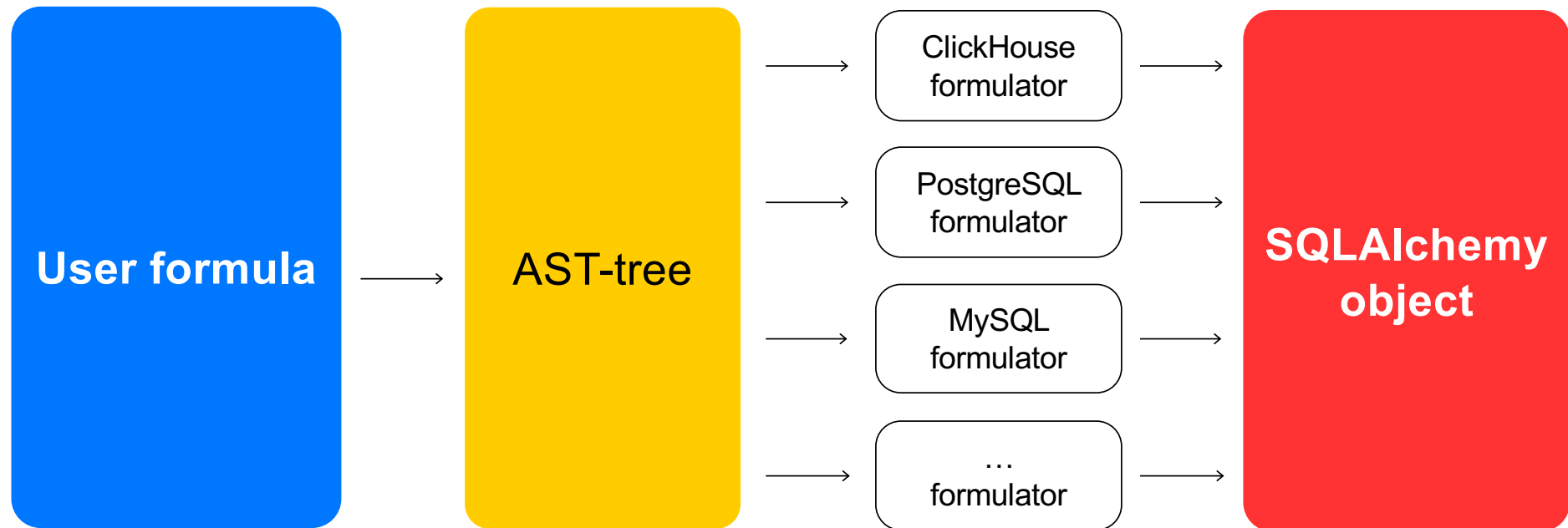
Преобразование AST в SELECT-запрос

Особенности использования SQLAlchemy

# Обработка формулы



# User-formula -> SQL query



<https://github.com/xzkostyan/clickhouse-sqlalchemy>

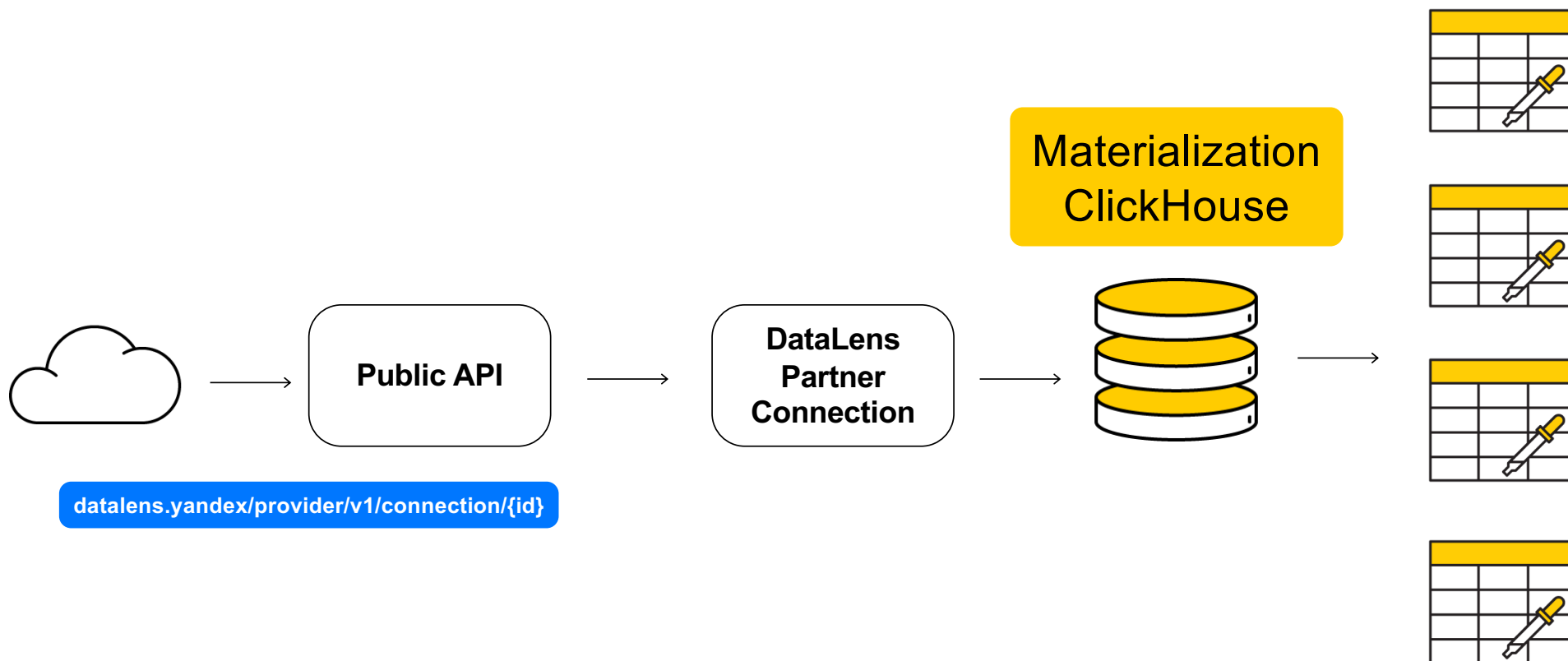
04



# Загрузка аналитических данных в ClickHouse

Принцип работы поставщиков данных  
Задержка репликации при создании таблиц

# Партнерские подключения



05



# Геокодинг

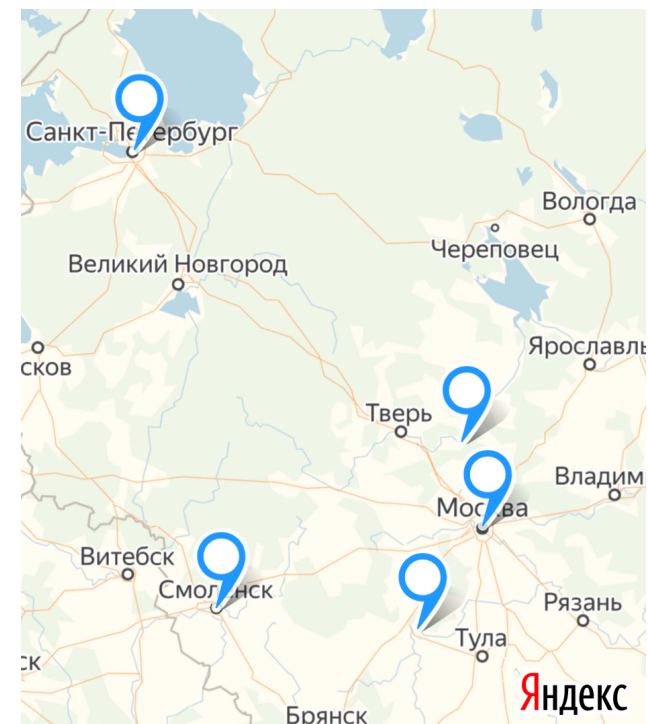
Что это такое и зачем это нужно

Принцип работы



# Geocode

| Столбец 1 | Столбец 2 | Адрес           | Столбец 3 |
|-----------|-----------|-----------------|-----------|
| Данные 1  | 10        | Москва          | 20        |
| Данные 2  | 10        | Санкт-Петербург | 20        |
| Данные 3  | 10        | Смоленск        | 20        |
| Данные 4  | 10        | Дубна           | 20        |
| Данные 5  | 40        | Калуга          | 80        |



# Geocache

| Адрес           | Координата       | Covert Time |
|-----------------|------------------|-------------|
| Москва          | 55.7522, 37.6156 | 2019-12-09  |
| Санкт-Петербург | 59.9386, 30.3141 | 2019-12-09  |
| Смоленск        | 54.7818, 32.0401 | 2019-12-09  |
| Дубна           | 56.7333, 37.1667 | 2019-12-09  |
| Калуга          | 54.5293, 36.2754 | 2019-12-09  |

PARTITION BY toDate(Convert Time)

Distinct(Адрес)



Geocode



Insert geocache

Update Old

Drop  
Outdated


Яндекс

Yandex DataLens

**Спасибо**

**Иван Москвин**

CTO Yandex DataLens

 [hans@yandex-team.ru](mailto:hans@yandex-team.ru)

 [@imoskvin](https://t.me/imoskvin)