

# Программа Перезапуск

## Модуль SQL.

### Занятие 8

Преподаватель: Марат Гарафутдинов

# Что смотрели на предыдущих занятиях

- Почему SQL нужен
- Реляционные таблицы
- Базовые операции
  - SELECT, INSERT, UPDATE, DELETE
- Операторы
  - WHERE, GROUP BY, AND, OR, NOT
- Операции JOIN
- Подзапросы
- Конструкции CASE WHEN
- Операторы ANY, ALL, IS NOT NULL
- Конструкции LIKE
- Конструкции CREATE TABLE & VIEW
- Оконные функции (ROW\_NUMBER, RANK, DENSE\_RANK)
- Индексы
- Виды баз данных

# Оператор Having

Оператор SQL HAVING является указателем на результат выполнения агрегатных функций.

Агрегатной функцией в языке SQL называется функция, возвращающая какое-либо одно значение по набору значений столбца. Такими функциями являются: COUNT(), MIN(), MAX(), AVG(), SUM().



# BigData: Hadoop

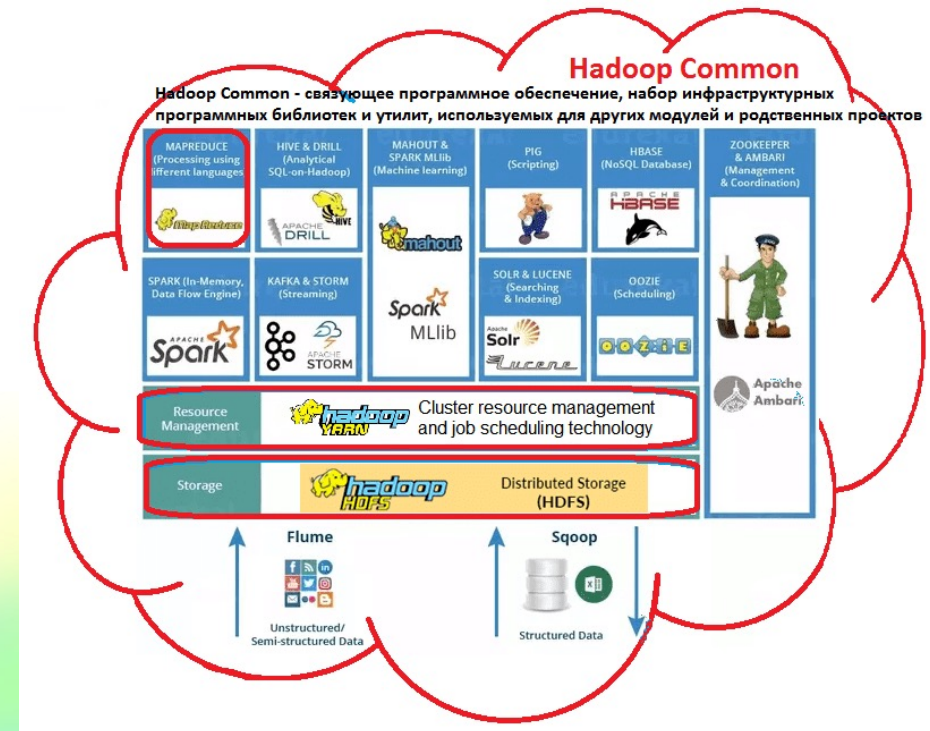
**Hadoop** — проект фонда Apache Software Foundation, свободно распространяемый набор утилит, библиотек и фреймворк для разработки и выполнения распределённых программ, работающих на кластерах из сотен и тысяч узлов.

**Hadoop Common** – набор инфраструктурных программных библиотек и утилит, которые используются в других решениях и родственных проектах, в частности, для управления распределёнными файлами и создания необходимой инфраструктуры

**HDFS** – распределённая файловая система, Hadoop Distributed File System – технология хранения файлов на различных серверах данных (узлах, DataNodes), адреса которых находятся на специальном сервере имен (мастере, NameNode)

**YARN** – система планирования заданий и управления кластером (Yet Another Resource Negotiator), которую также называют MapReduce 2.0 (MRv2) – набор системных программ (демонов), обеспечивающих совместное использование, масштабирование и надежность работы распределённых приложений

**Hadoop MapReduce** – платформа программирования и выполнения распределённых MapReduce-вычислений, с использованием большого количества компьютеров (узлов, nodes), образующих кластер.



<https://stepik.org/course/150/promo>

<https://www.coursera.org/learn/big-data-essentials>

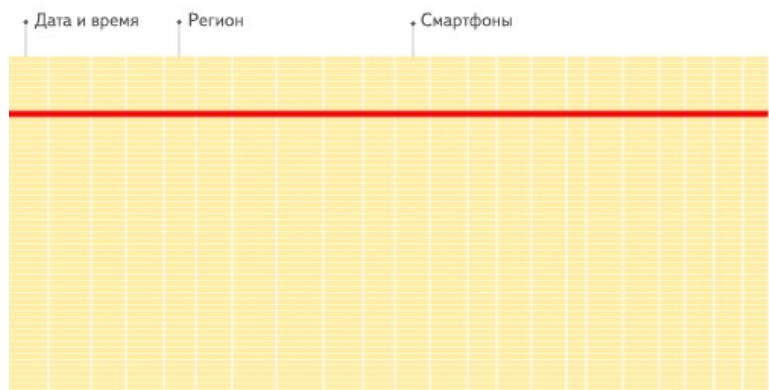
# BigData: Clickhouse (by Yandex)

**ClickHouse** — это колоночная аналитическая СУБД с открытым кодом, позволяющая выполнять аналитические запросы в режиме реального времени на структурированных больших данных, разрабатываемая компанией Яндекс.

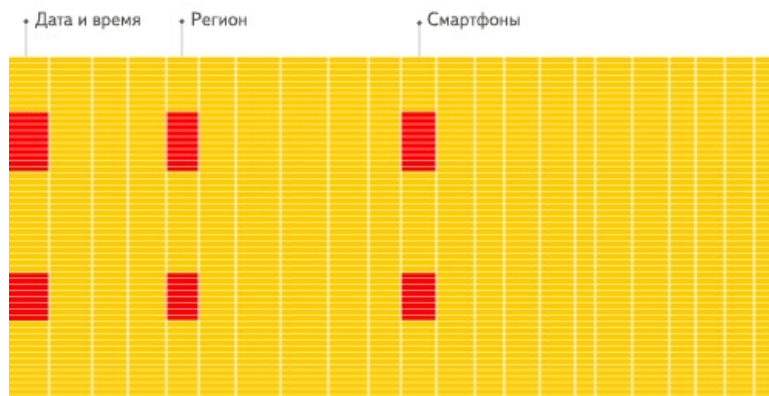
ClickHouse использует собственный диалект SQL близкий к стандартному, но содержащий различные расширения: массивы и вложенные структуры данных, функции высшего порядка, вероятностные структуры, функции для работы с URI, возможность для работы с внешними key-value хранилищами («словарями»), специализированные агрегатные функции, функциональности для семплирования, приблизительных вычислений, возможность создания хранимых представлений с агрегацией, наполнения таблицы из потока сообщений Apache Kafka и т. д.

# BigData: Clickhouse (by Yandex)

## Строковые СУБД



## Колоночные (столбцовые) СУБД



Ключевые особенности OLAP сценария работы

- подавляющее большинство запросов - на чтение;
- данные обновляются достаточно большими пачками (> 1000 строк), а не по одной строке, или не обновляются вообще;
- данные добавляются в БД, но не изменяются;
- при чтении, вынимается достаточно большое количество строк из БД, но только небольшое подмножество столбцов;
- таблицы являются «широкими», то есть, содержат большое количество столбцов;
- запросы идут сравнительно редко (обычно не более сотни в секунду на сервер);
- при выполнении простых запросов, допустимы задержки в районе 50 мс;

# Linux: Работаем с Google Cloud Shell и PostgreSQL

<https://console.cloud.google.com/>

Команда: `psql -h 52.157.159.24 -Ustudent0 sql_ex_for_student0`

Далее вводим пароль



# Демонстрация и решение упражнений