



Реляционные базы, история и место в современном мире



Проверить, идет ли запись

Меня хорошо видно && слышно?



Ставим "+", если все хорошо
"-", если есть проблемы



Тема вебинара

Реляционные базы, история и место в современном мире



Ведущий разработчик PostgreSQL/Greenplum в Сбере

Специалист в области разработки и проектирования витрин данных в PostgreSQL/Greenplum, а также в области разработки хранимых процедур в таких СУБД как PostgreSQL/Greenplum, Oracle, MS SQL Server.

Правила вебинара



Активно
участвуем



Off-topic обсуждаем
в учебной группе
<https://t.me/+p-vcSnf6EtVjMzli>



Задаем вопрос
в чат или голосом



Вопросы вижу в чате,
могу ответить не сразу

Маршрут вебинара

Знакомство

1. Где мы были и где сейчас

2. NoSQL vs RDBMS

3. Тренды

4. Обзор GCP & YC

OLAP/OLTP/DWH

Масштабирование/backup/миграция

Рефлексия. ДЗ

Цели вебинара

К концу занятия вы сможете

1. Верхнеуровненно отличать SQL и NoSQL
2. Представлять сложность выбора той или иной технологии СУБД
3. Чуть лучше понимать, чем мы будем заниматься на этом курсе

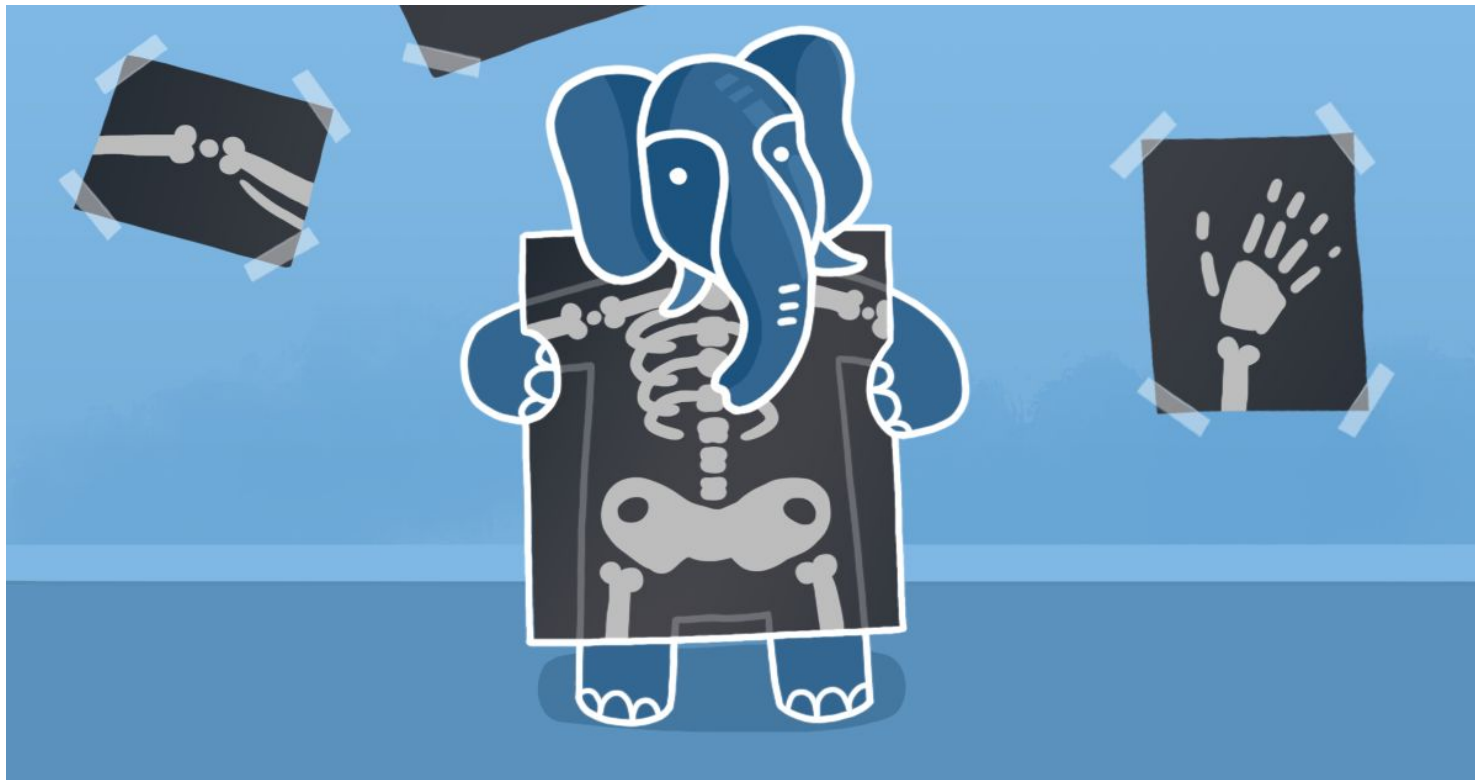
Смысл

Зачем вам это уметь

1. Осознанно развивать востребованные навыки
2. с точки зрения разных технологий СУБД
3. и облачных сервисов



PostgreSQL



Где мы были и где сейчас

История. Когда деревья были высокими

Выбор конца прошлого века

The Oracle logo is centered within a large red rectangular box. The word "ORACLE" is written in a white, bold, sans-serif typeface. A registered trademark symbol (®) is positioned to the upper right of the final letter 'E'.

ORACLE®

Нужные навыки для DBA

- установить Oracle Server из дистрибутива(Solaris)
- настроить Oracle Server
 - init.ora
 - listener.ora
- создать
 - табличные пространства
 - схемы
 - пользователей
- настроить резервное копирование и восстановление
- настроить мониторинг
- бороться с неоптимальными запросами

Чего не ожидали от DBA

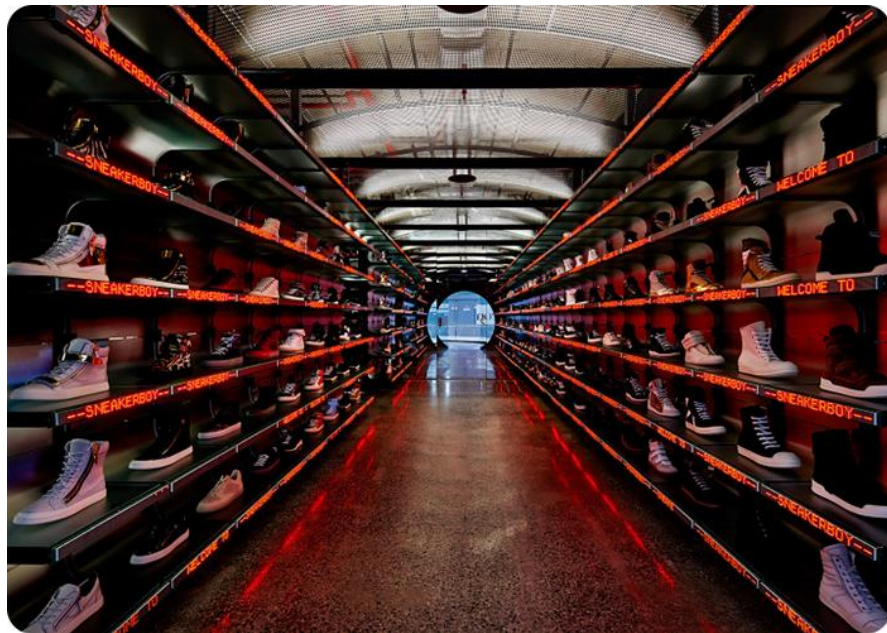
- выбор оптимальной СУБД или другой технологии хранения и обработки данных
- высокая доступность и горизонтальная масштабируемость, соответственно репликация и шардирование, балансировщики нагрузки
- docker, k8s
- OLTP, OLAP, оптимизации, BI, витрины данных
- знание предметной области и прикладной архитектуры
- хорошее знание инфраструктуры
- знание ОС
- загрузка и выгрузка данных - ETL
- миграция данных между разными СУБД
- облаков разных всё больше и больше - **какие вы знаете?**

Где мы были и где сейчас

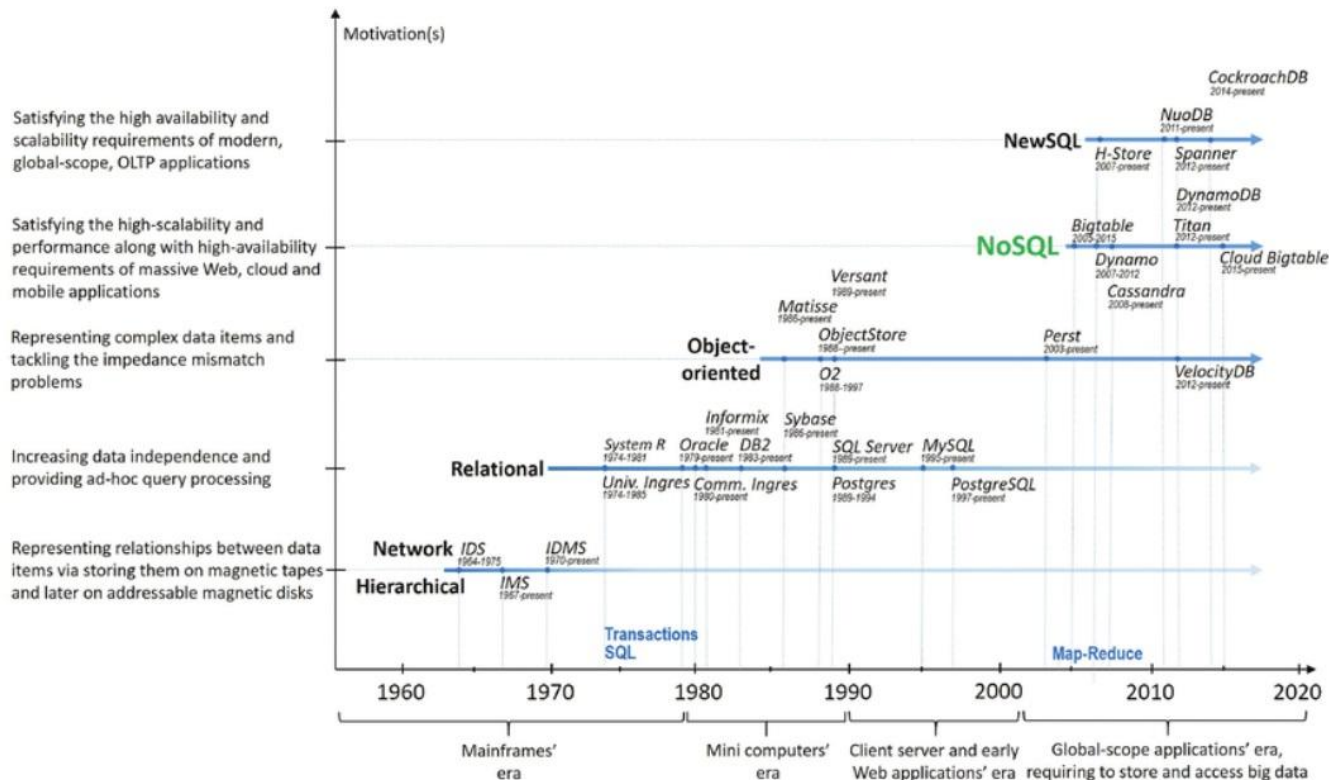
Раньше



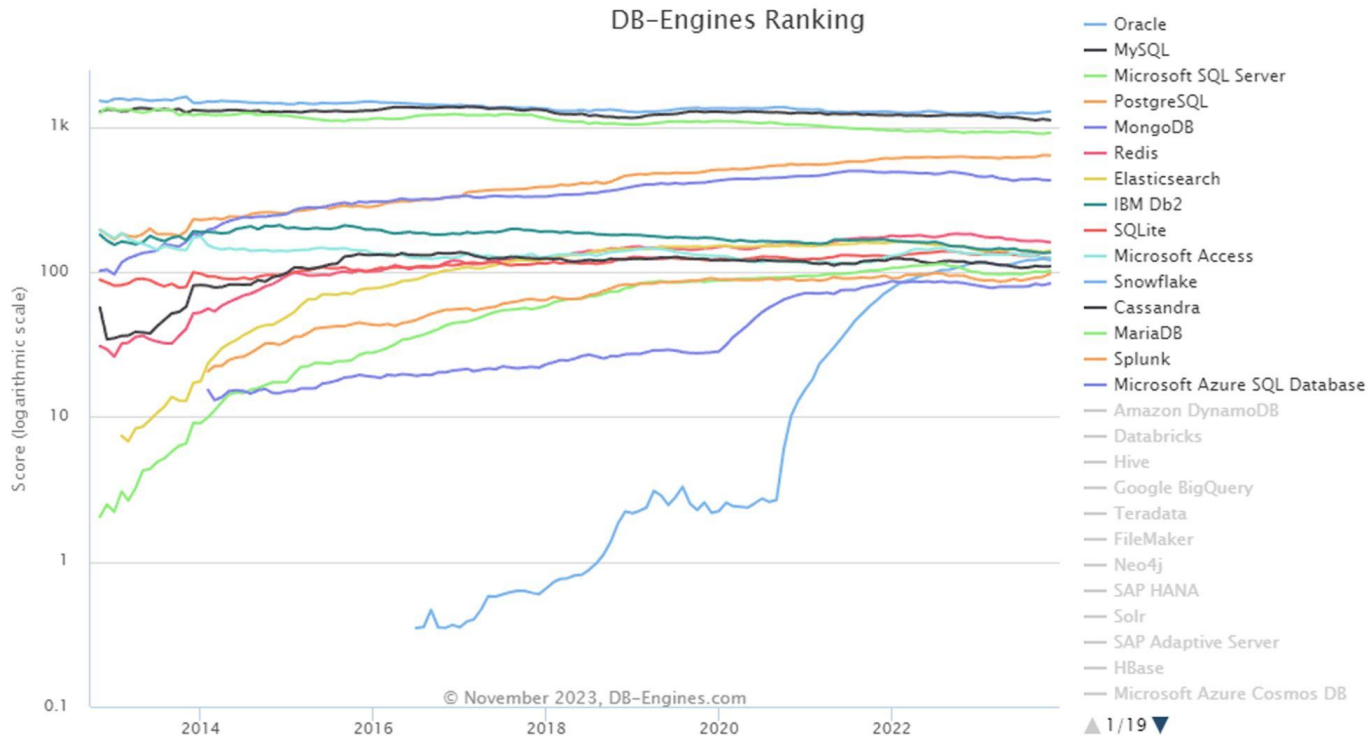
Сейчас



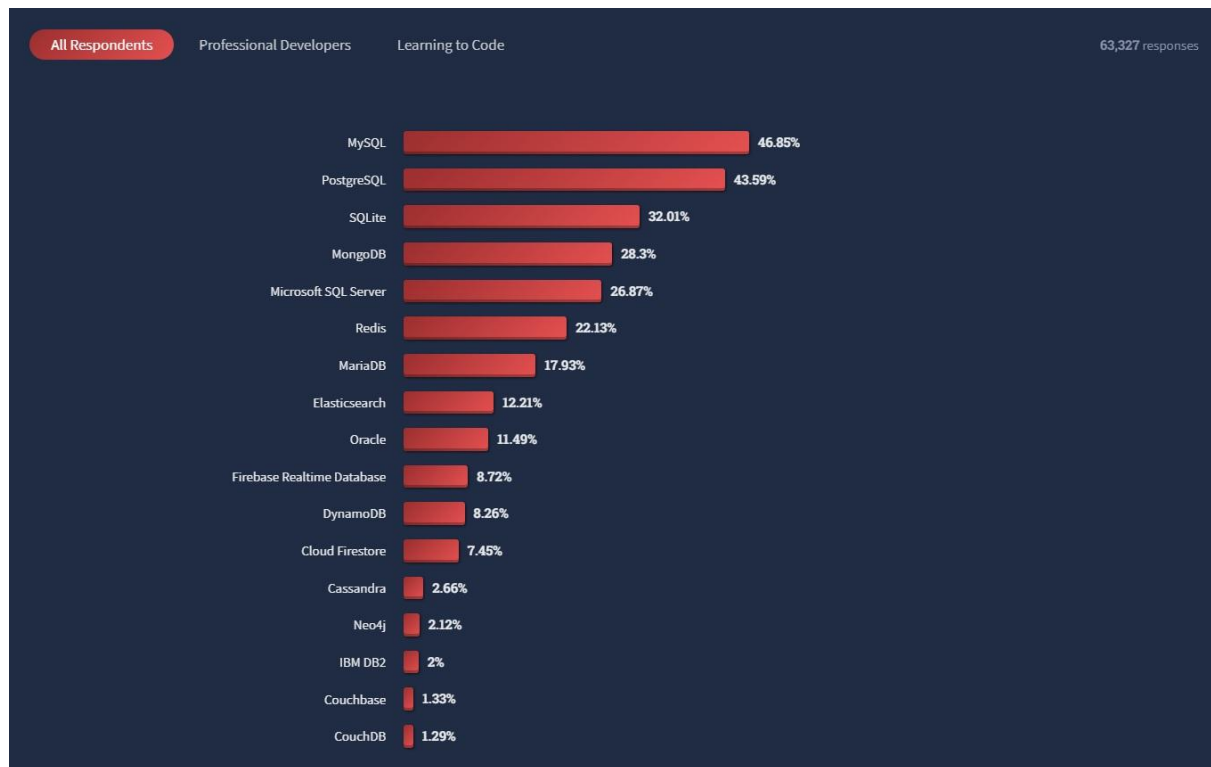
История



Что выбрать?



Top 10 DB in 2020



Тенденции

<https://statisticsanddata.org/data/the-most-popular-databases-2006-2023/>

Most Popular Databases 2023

Rank	Name	February 2023	Last month	Last year
1.	Oracle	1,247.52	2.35	-9.31
2.	MySQL	1,195.45	-16.51	-19.23
3.	Microsoft SQL Server	929.09	9.70	-19.96
4.	PostgreSQL	616.50	1.65	7.12
5.	MongoDB	452.77	-2.42	-35.88
6.	Redis	173.83	-3.72	-1.96



SQL

SQL

- способ взаимодействия SQL
- внутри базы данных
- ACID
- no scale

Например

- Oracle Database
- Microsoft SQL
- Maria DB
- MySQL
- PostgreSQL



NoSQL

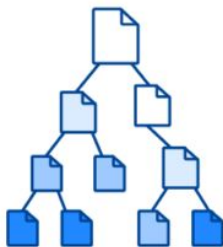
NoSQL

- способ взаимодействия поиск по ключу/ам
- внутри базы key-value, или json, или граф, или...
- Обычно нет ACID
- scale

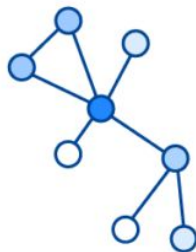
Например

- MongoDB
- Cassandra
- Elasticsearch
- Google Datastore
- Amazon DynamoDB

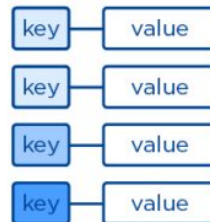
Document



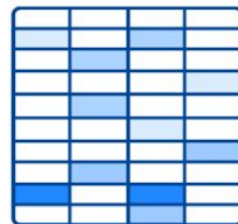
Graph



Key-Value



Wide-column



NewSQL

NewSQL

- способ взаимодействия SQL
- внутри базы таблицы и иногда JSON
- ACID
- scale

Например

- Amazon Aurora
- Google Spanner
- CockroachDB
- Yandex Managed Service for YDB

PostgreSQL

PostgreSQL

- SQL
- с элементами NoSQL (jsonb)
- с элементами NewSQL (scaleable - репликация сторонними и встроенными инструментами)
- ACID



Что дальше?

Отчет аналитического ресурса Gartner: Будущее систем управления базами данных - за облаками!

Что это значит для ИТ-сообщества?

1. Во-первых, это подтверждает, что все больше организаций конечных пользователей развертывают системы и приложения в облаке, включая замену локальных систем SaaS. Это может быть движущей силой этих изменений, а может и не быть.
2. Во-вторых, организации, которые хотят воспользоваться преимуществами новых инноваций в СУБД, переходят в облако. Это происходит только там или, по крайней мере, в первую очередь в облаке. Однако появляется все больше инноваций, которые никогда не будут внедряться локально, даже если у поставщика есть локальные продукты.
3. В-третьих, новые модели ценообразования, отказ от капитальных затрат в пользу операционных расходов и использование подхода "оплата по мере поступления", по-видимому, стимулируют переход в облако.

Перевод статьи:

<https://blogs.gartner.com/adam-ronthal/2019/06/23/future-database-management-systems-cloud/>

Где мы сейчас?

- в облаке (около 70% решений) или на пути к нему
- выбор технологий СУБД, даже в формате on premises - огромный
- к этому добавляется то что есть в облаке в виде DBaaS (Database-as-a-service)

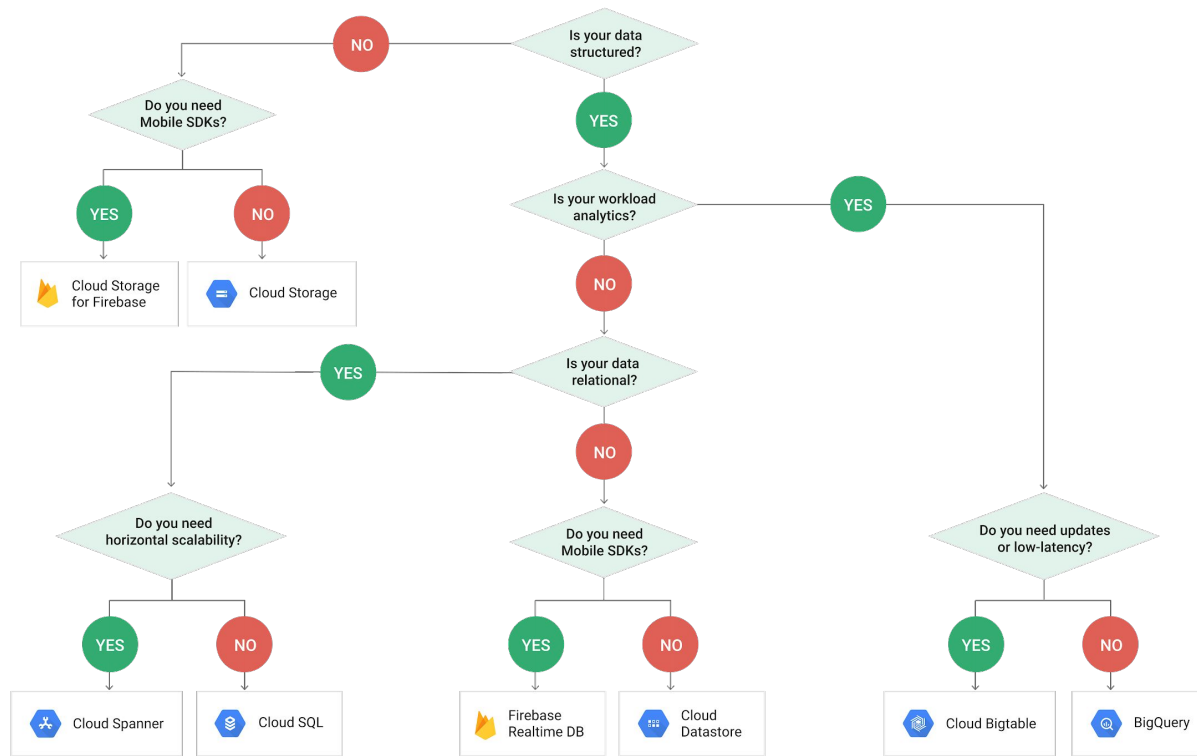
ограничения:

- навыки
- цена
- ряд ограничений в случае DBaaS (на примере su)

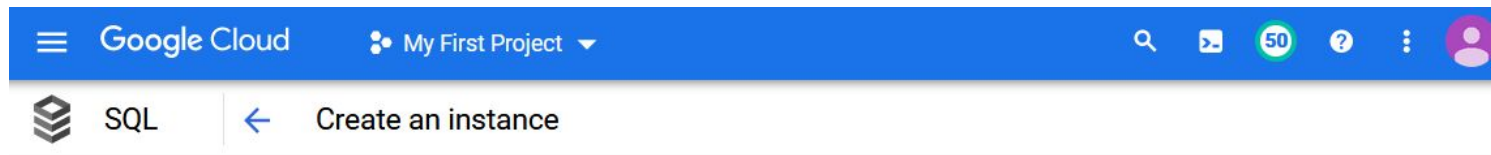


Обзор GCP & YC.

на примере GCP



на примере GCP



Choose your database engine



MySQL

Versions: 8.0, 5.7, 5.6

[Choose MySQL](#)



PostgreSQL

Versions: 14, 13, 12, 11, 10, 9.6

[Choose PostgreSQL](#)



SQL Server

Versions: 2019, 2017

[Choose SQL Server](#)



Итог

выбор

- даже если смотреть только на GCP
- и только на managed services и SaaS
- выбор РСУБД только на 4-м шаге (а там Spanner рядом)
- а выбор PostgreSQL на 5-м (а рядом MySQL и SQL Server)

в условиях ограничений и соблазнов

- а хочется то Spanner!
- но как же дорого!
- а можете сделать так же как Spanner, но за цену Cloud SQL?
- ну вы же профессионалы



Вопросы?



Ставим “+”,
если вопросы есть



Ставим “-”,
если вопросов нет

Минитест

<https://forms.gle/BScK9bCCaqsHNz418>

3-5 минут

**Что еще кроме того мы
должны учитывать**

OLTP && OLAP

Какая у нас нагрузка?

OnLine Transaction Processing - много быстрых маленьких запросов

```
select name from users where id = '777';
```

OnLine Analytical Processing - редкие огромные аналитические запросы

```
select payment_type, round(sum(tips)/sum(tips+fare)*100) tips_percent, count(*) from  
bigquery-public-data.chicago_taxi_trips.taxi_trips group by payment_type
```

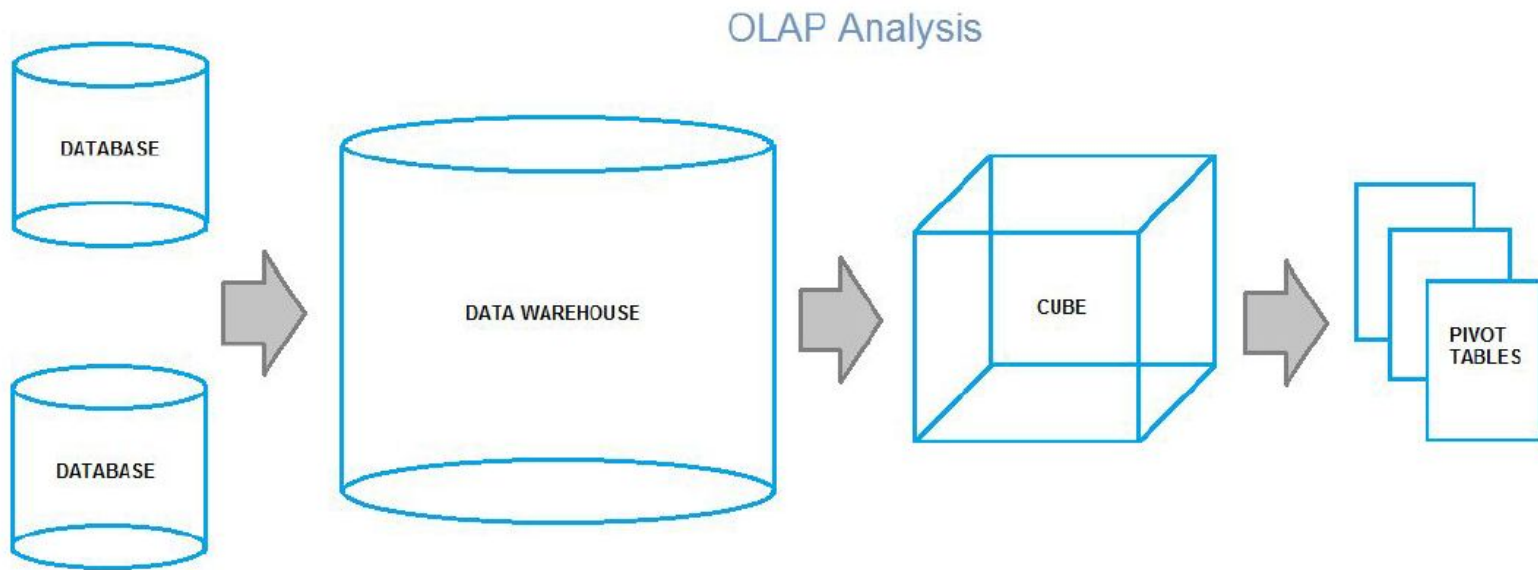
OLTP и OLAP

DWH

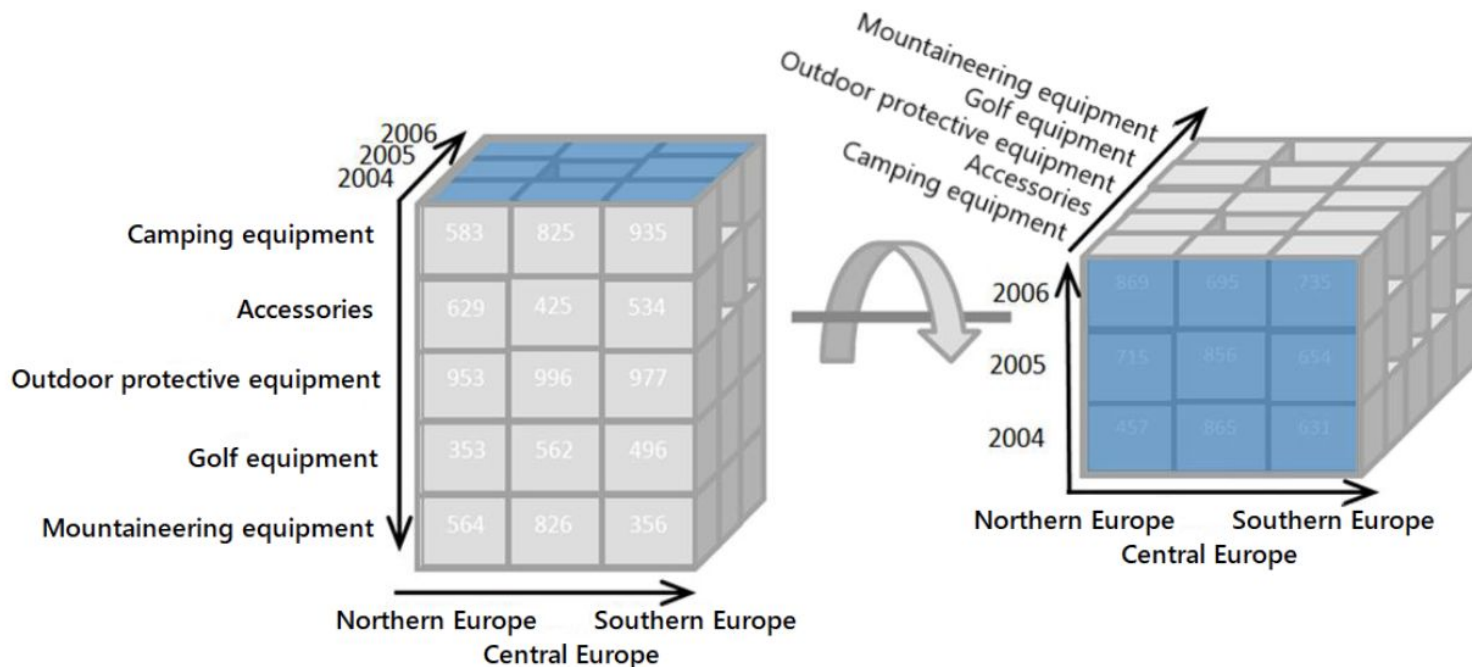
DWH - Data WareHouse

- Получает данные посредством ETL <https://ru.wikipedia.org/wiki/ETL> из многих источников
- Часто денормализована [Денормализация](#)
- Используется для аналитики и систем принятия решений
- Требуется много индексов
- Оперирует данными за несколько лет
- Обычно сотни терабайт
- Скорость обработки запроса может измеряться часами
- Использует OLAP (<https://ru.wikipedia.org/wiki/OLAP>)
 - Технология для построения аналитических отчетов

OLAP - Online Analytical Processing



OLAP - Online Analytical Processing



Примеры

- OLTP
 - магазин, биржа, биллинг в телекоме, банковская система
 - [ERP](#), [CRM](#)
 - Витрины товаров, каталоги
- DWH (на основе PostgreSQL - GreenPlum <https://greenplum.org/> , ArenaData <https://arenadata.tech/>)
 - аналитика вышеперечисленных систем
 - системы логирования
- БД для микросервисов
- Кэш системы (redis, tarantool)
- Системы управления конфигурациями (etcd, consul)

Характеристики при выборе СУБД

Консистентность

- Консистентность реплик
- Объем данных
- Размер транзакций (длительность, объем данных,

кол-во одновременных транзакций)

- Кол-во одновременных соединений/сессий
- Соотношение операций чтения/записи (сколько нужно реплик на чтение)
- Масштабирование
- Шардинг (в ванильной версии из коробки нет, но есть Citus)
- Партиционирование

Масштабирование

Репликация

- Журналы (REDO logs (MySQL), WAL (Postgres), etc..)
- Репликация
 - Процесс синхронизации данных путем доставки журналов
- на другие ноды
- Задачи: балансировка чтения(/записи?), fault tolerance, Master-Slave
- Проблемы: организация автоматического переключения
- Разновидности:
 - Асинхронная
 - Синхронная
 - Полусинхронная
 - Групповая
- Шардирование

Шардирование

Таблицы БД в схеме

Customers

Customer	Name
123	Mary
456	John
999	Peter

Orders

Order	Customer
4001	123
4002	456
4003	999
4004	456
4005	456

LineItems

Customer	Order	Line
123	4001	40011
999	4003	40012
123	4001	40013
456	4004	40014
999	4003	40015
999	4003	40016

Products

SKU	Product
100	Coil
101	Piston
102	Belt

Sharded Tables

Shard 1



Shard 2



Shard N



Таблицы справочников дублируются по всем БД

Кластеризация

- Кластеризация
 - Полноценная масштабируемая, отказоустойчивая конфигурация
- MySQL: Galera cluster, InnoDB cluster, NDB cluster, eXtra DB cluster
- PostgreSQL: Patroni, Slony, Stolon, CockroachDB, Greenplum и др
- Oracle: RAC (требует общего хранилища)
- MSSQL: AlwaysON
- а там уже Vertica, DBaaS, BigQuery и т.д. на подходе

Резервное копирование

Бэкапы



Миграция

Миграция

- Миграция, в данном контексте, — обновление структуры базы данных
- от одной версии до другой
- Версионирование БД
- Средства
 - Liquibase
 - Alembic
 - Flyway

Погнали в облака



Практика

1. Облака:
cli GCloud
<https://cloud.google.com/sdk/docs/downloads-apt-get>
2. cli ЯО
<https://cloud.yandex.ru/docs/cli/quickstart>
3. Какие есть варианты без облаков:
VirtualBox
<https://www.virtualbox.org/wiki/Downloads>
4. Docker
<https://www.docker.com/>
5. On-premise
<https://www.postgresql.org/download/>



Программа курса

Программа курса

<https://otus.ru/lessons/postgresql-dba>

Студенту, набравшему наибольшее количество баллов за домашние работы, мы подарим книгу в .pdf руководителя курса Евгения Аристова

["PostgreSQL 14. Оптимизация, Kubernetes, кластера, облака."](#)

**Заполните, пожалуйста,
опрос о занятии
по ссылке <https://otus.ru/polls/79270/>**

Рефлексия

Результаты опроса. Рефлексия

https://docs.google.com/forms/d/1CFnVKldNSO_mtSjHrpZ00TTaC-lwbtXV1bClhOqM85U/edit#responses



С какими впечатлениями уходите с вебинара?



Как будете применять на практике то, что узнали на вебинаре?

*** Напоминаю, что сегодня вводная лекция по типу 1 сентября**

ДЗ



Домашнее задание

1. нет его 😊
2. Изучаем гитхаб и маркдаун
3. Есть GUI для гитхаба <https://desktop.github.com/>
4. Пример маркдауна <https://github.com/aeuge/postgres16book>

Спасибо за внимание!

Приходите на следующие вебинары



Ведущий разработчик PostgreSQL/Greenplum в Сбере

Специалист в области разработки и проектировании витрин данных в PostgreSQL/Greenplum, а также в области разработки хранимых процедур в таких СУБД как PostgreSQL/Greenplum, Oracle, MS SQL Server.