




ОНЛАЙН-ОБРАЗОВАНИЕ

Онлайн-образование



Меня хорошо видно && слышно?

Ставьте  , если все хорошо
Напишите в чат, если есть проблемы
заодно проверяем, включена ли запись занятия

Включил Юджин запись ли ты





Введение в NoSQL. RDBMS vs NoSQL



Аристов Евгений

telegram @AEugene

<https://aristov.tech>

Преподаватель



Евгений Аристов

Более 20 лет занимаюсь разработкой ПО на Java/Spring, C#, PHP.
Архитектор баз данных PostgreSQL, MongoDB, MSSQL, Oracle, MySQL, MariaDB, Couchbase и др.

Деплой БД как on-premise Google Cloud Platform, AWS, Azure, Yandex Cloud, так и Kubernetes, DBaaS, MultiCloud.

За это время было спроектировано и разработано более ста проектов для сетей магазинов, фитнес-центров, отелей, финансовом секторе.

Огромный опыт построения и эксплуатации систем виртуализации VmWare ESXi, Hyper-V.

Автор книг по PostgreSQL. Новинка ["PostgreSQL 14. Оптимизация, Kubernetes, кластера, облака."](#)

Правила вебинара



Активно участвуем



Задаем вопрос в чат



Вопросы вижу в чате, могу ответить не сразу

Дальнейшее обсуждение и оффтоп в Slack

Маршрут вебинара

1. Где мы были и где сейчас

3. Тренды

5. Практика

2. NoSQL vs RDBMS

4. GCP



Цели вебинара | После занятия вы сможете

1 Отличать SQL и NoSQL

2 Представлять сложность выбора той или иной технологии СУБД

3 Чуть лучше понимать, чем мы будем заниматься на этом курсе

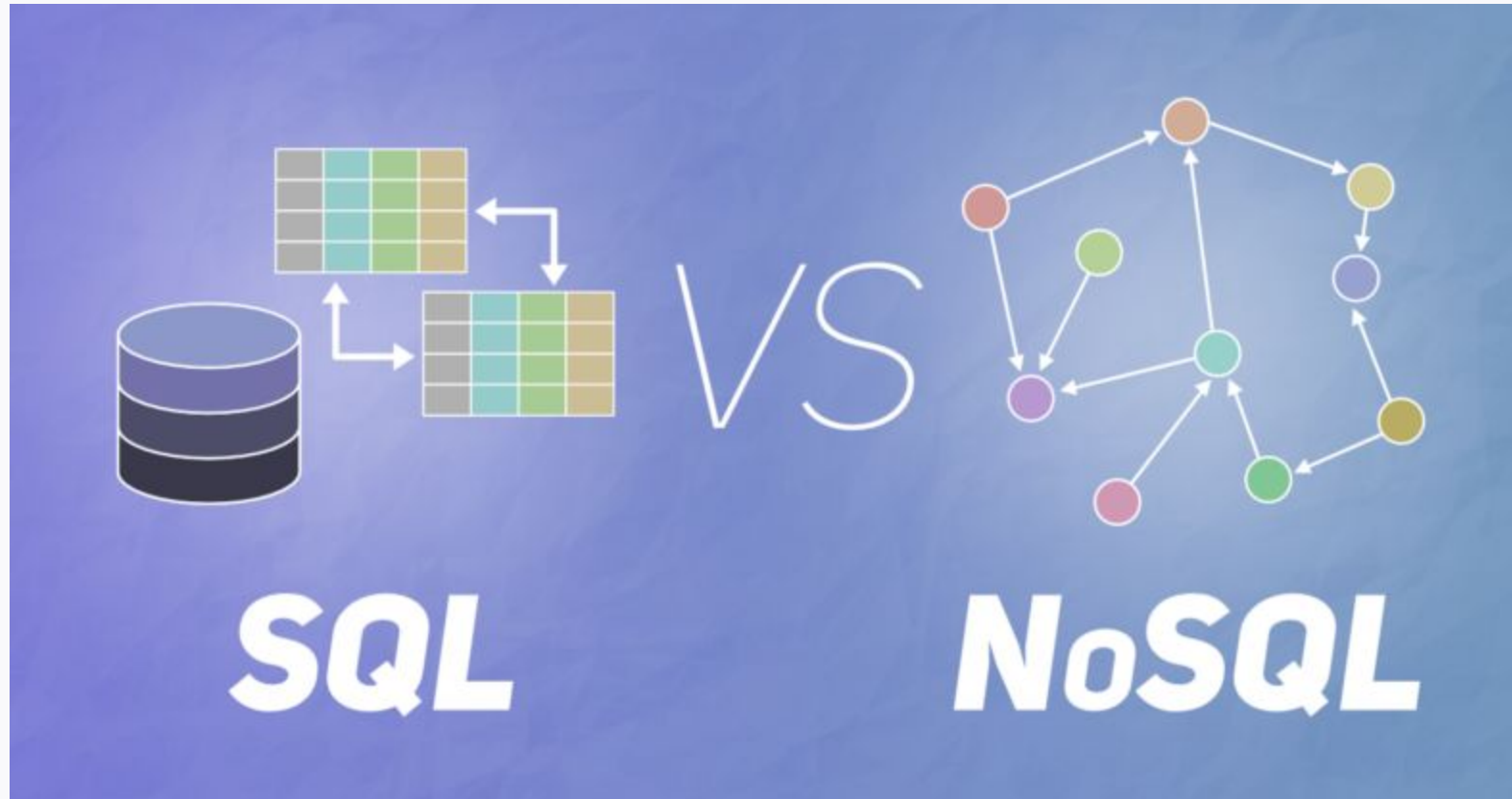
СМЫСЛ | Зачем вам это уметь

1 Осознанно развивать востребованные навыки

2 с точки зрения разных технологий СУБД

3 и облачных сервисов

SQL vs NoSQL



The image features a blue-toned aerial photograph of a dense urban skyline, likely New York City, with numerous skyscrapers. A semi-transparent blue band with a white geometric network pattern of dots and lines runs horizontally across the middle of the image. The Russian text "Где мы были и где сейчас" is centered within this band.

Где мы были и где сейчас

История. Когда деревья были высокими

Выбор конца прошлого века

The Oracle logo, consisting of the word "ORACLE" in a white, sans-serif font with a registered trademark symbol, set against a solid red rectangular background.

ORACLE®

Нужные навыки для DBA

- установить Oracle Server из дистрибутива (Solaris)
- настроить Oracle Server
 - init.ora
 - listener.ora
- создать
 - табличные пространства
 - схемы
 - пользователей
- резервное копирование и восстановление
- мониторинг
- борьба 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
- знание предметной области и прикладной архитектуры
- хорошее знание инфраструктуры
- знание ОС
- загрузка и выгрузка данных - ETL
- миграция данных между разными СУБД

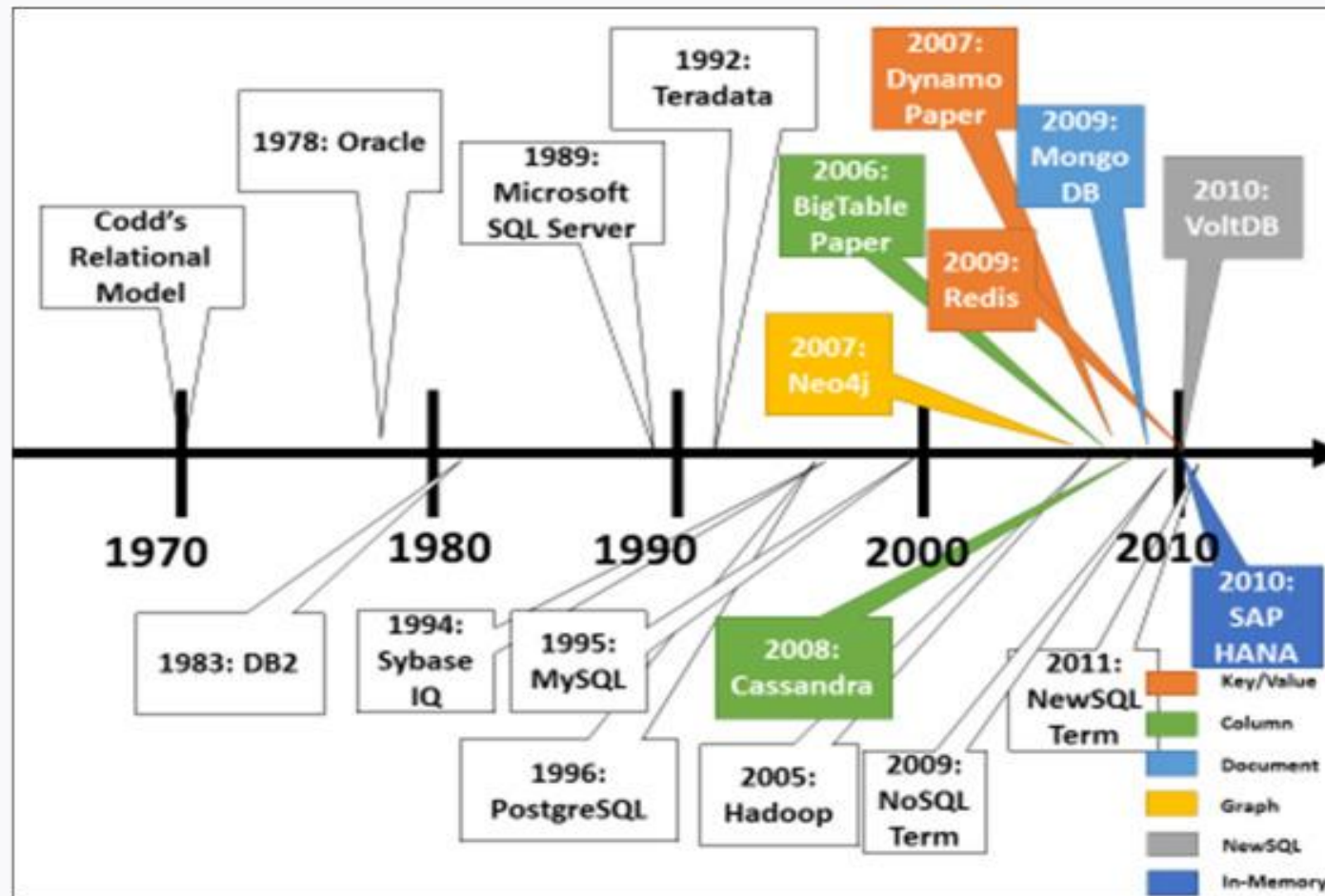
Раньше...



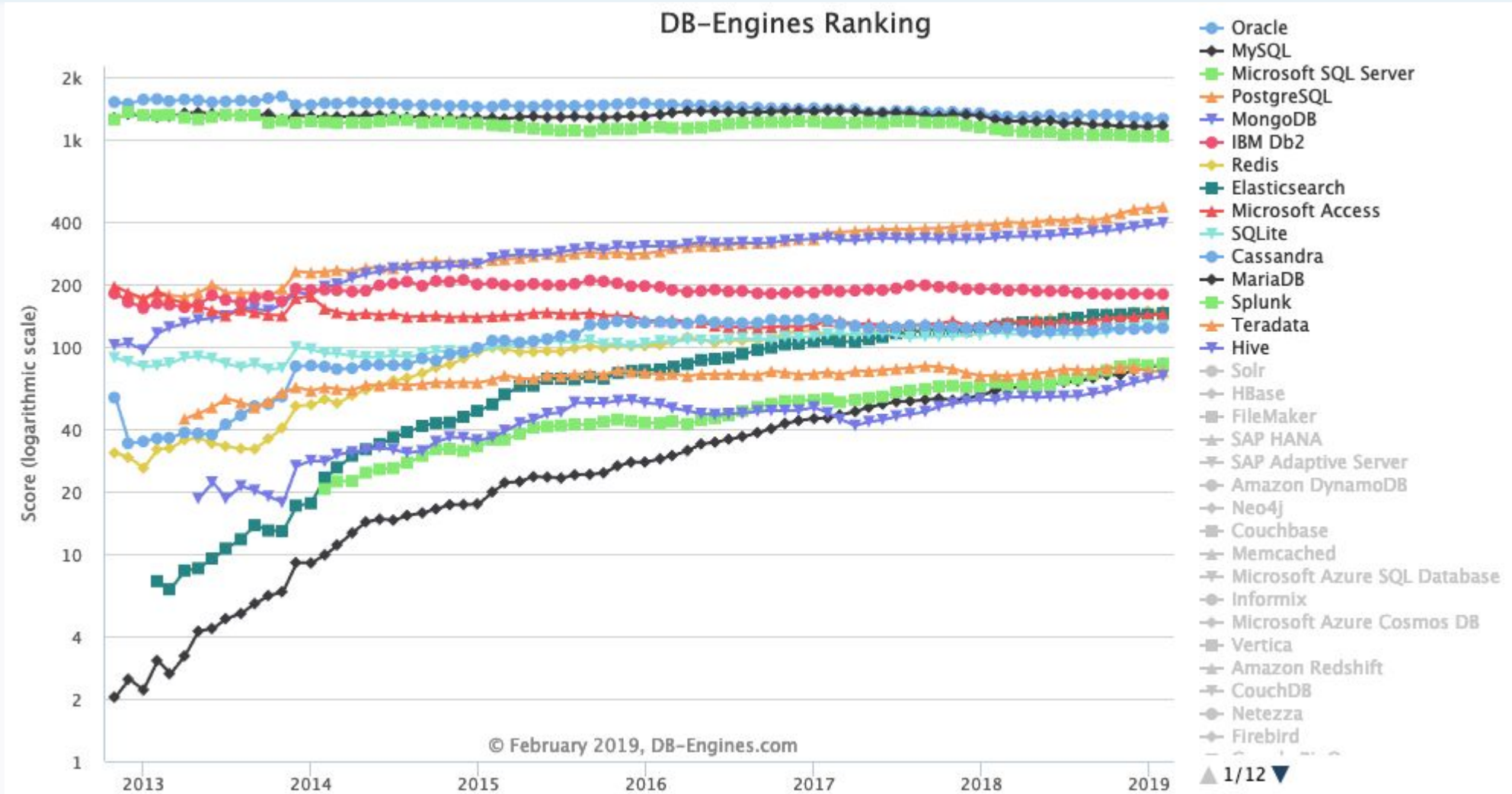
Сейчас...



История



Что выбрать?



Что же лучше использовать?

MySQL и MongoDB — когда и что лучше использовать

Тренды

[The Most Popular Databases – 2006/2021 - Update May 2021](#)

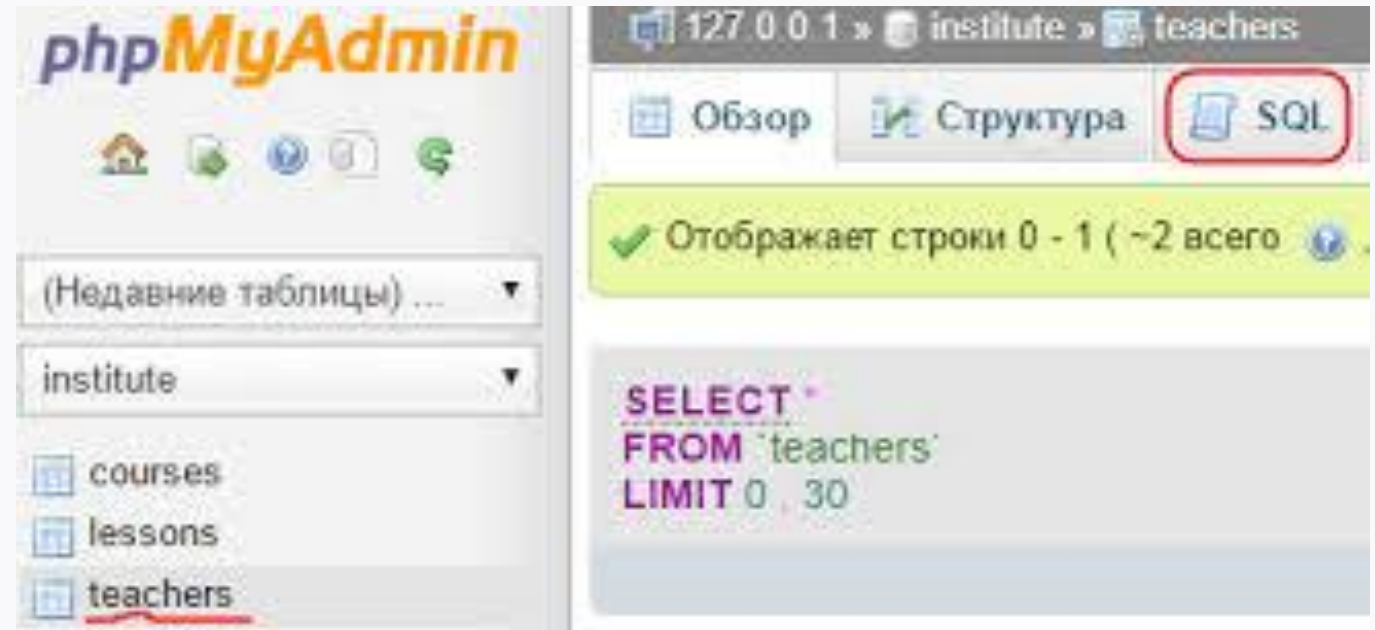
SQL

SQL

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

Например

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



NoSQL

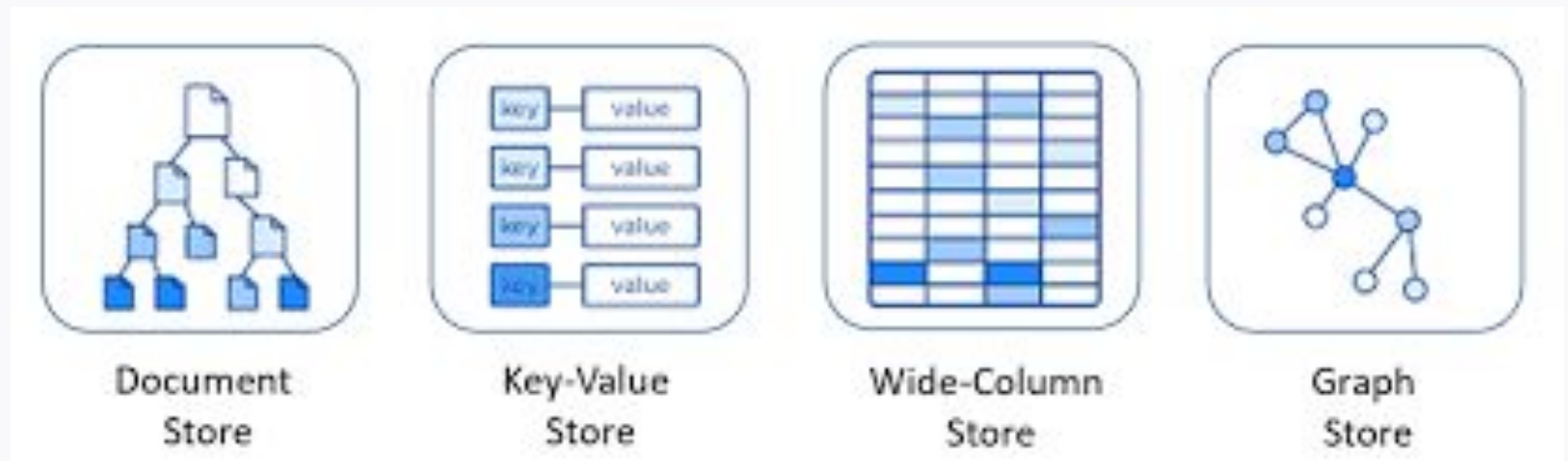
NoSQL

- 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



NewSQL

NewSQL

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

Например

- Amazon Aurora
- Google Spanner
- CockroachDB

Что дальше?

Gartner: The Future of Database Management Systems is Cloud!

What does this mean to the IT community?

- First, it confirms that more and more end-user organizations are deploying systems and applications **to the cloud**, including **replacing on-premises systems with SaaS**.

IT may or may not be driving these shifts.

- Second, organizations who want to take **advantage of new innovation in DBMS** are moving to the cloud. It is happening **only there, or at the least, in the cloud first**.

However, there is an increasing amount of innovation that will never get to on-premises, even when the vendor has on-premises products.

- Third, new pricing models, avoidance of capital expense in favor of operating expense, and leveraging a **pay-as-you-go** approach, seem to be driving the move to the cloud.

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

Где мы сейчас?

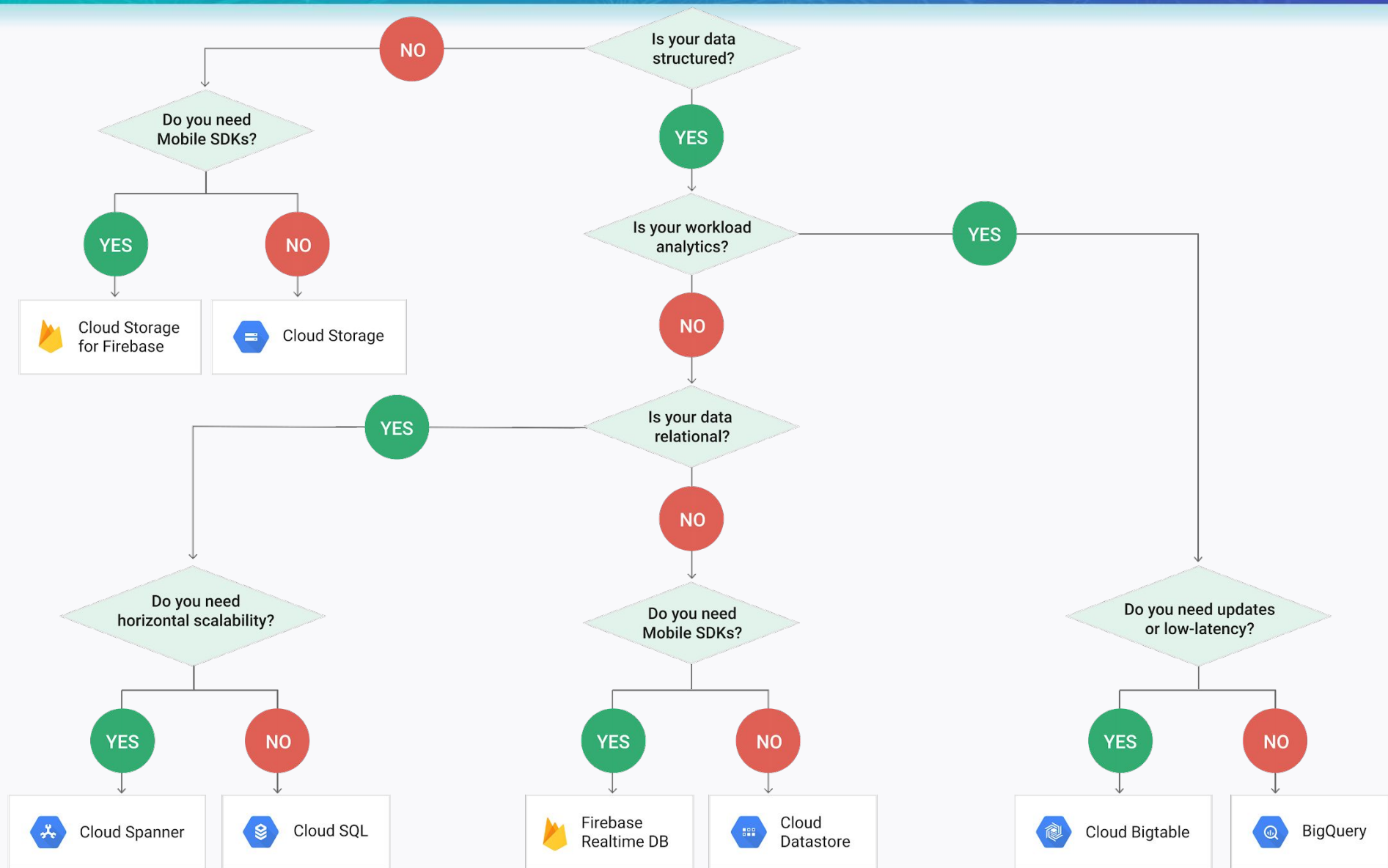
- в облаке или на пути к нему
- выбор:
- выбор технологий СУБД, даже в формате on premises - огромный
- к этому добавляется то что есть в облаке в виде managed services и SaaS

ограничения:

- навыки
- цена
- фичи в случае managed services и SaaS



на примере GCP



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 SQL на 5-м (PostgreSQL, MySQL и SQL Server)
- все остальное - NoSQL решения в условиях ограничений и соблазнов
- а хочется то Spanner и BigQuery!
- но как же дорого!
- а можете сделать так же как Spanner, но бесплатно?
- ну вы же профессионалы!



The background of the slide is a high-angle, blue-tinted aerial photograph of a dense urban skyline, likely New York City. Overlaid on this image is a semi-transparent blue band that contains a white network or mesh pattern of interconnected lines and dots. Centered within this band is the word "Вопросы?" in a large, white, sans-serif font.

Вопросы?

The background of the entire image is an aerial photograph of a city, likely New York City, showing a dense cluster of skyscrapers. A semi-transparent blue overlay covers the image, featuring a network of white lines and dots that suggest a digital or interconnected theme. The text is centered within this overlay.

Сложности выбора

Случай из практики

- построить CI/CD
- сделать review архитектуры
- и запустить в эксплуатацию
- микросервисы k8s
- Python/Django
- GCP
 - 2x региона - US и EU
 - Global Load balancer
 - Workloads
 - Google Kubernetes Engine
 - Data
 - Cloud Storage
 - Datastore
 - Cloud SQL (PostgreSQL)

Почему Cloud SQL?

как правило за этим вопросом следует неловкая пауза ;)

в данном случае потому что: **Django ORM**

поддерживает следующие СУБД

- PostgreSQL
- MariaDB
- MySQL
- Oracle
- SQLite

PostgreSQL выбрали из этого перечня скорее интуитивно. ну не Oracle же ;)

Чего не хватает?

- приложение сейчас развернуто в 2 регионах
- в планах - 3-ий (Азия)
- БД - в североамериканском регионе (Iowa)
- возможны задержки с доступом из Европы и Азии
- и перебои в обслуживании в случае простоя СУБД
- Django может работать с несколькими БД параллельно
- и делить их по чтению и записи
- а записи не так уж и много (>90% - чтение)
- read replica в Европе и Азии была бы компромиссным вариантом решения
- ну так а что тут сложного?

Что может Cloud SQL?

- HA
 - в рамках одного региона
 - посредством дисковой репликации
 - не используются механизмы PostgreSQL
 - возможно автоматическое и ручное управление - failover/failback
 - при переключении СУБД недоступна в течение нескольких минут
- RR - read replica
 - в рамках одного региона
 - hot standby
 - PostgreSQL streaming replication

Ограничения

- заказчик не хочет плодить сущности и использовать IaaS кроме как через GKE
- заказчик не хотел бы разворачивать self service PostgreSQL/MySQL
- ну и вообще Google Spanner вполне бы подошел
- если бы не его цена
- ну и еще с ним Django ORM работать не может
- а так штука то хорошая ;)
- а можете чтонибудь похожее сделать так чтобы и с Django ORM работало?

Что выбрали?

- CockroachDB - огонь! NewSQL
 - но стремно ;)
- MySQL Galera - тоже неплохо
 - много где используется
 - но MySQL
- **Patroni**
 - прекрасная интеграция с K8s
 - нет лишних сущностей
 - хорошо интегрируется с GCP LB

Выводы

- мир ИТ инфраструктуры существенно поменялся
- и это только начало
- инновации будут появляться в облаках
- и только потом силами стартапов переносится в on-premises
- SQL будет жить
- NoSQL набирает обороты
- но еще важнее уметь их правильно применять
- облака
- Kubernetes

The background of the slide is a high-angle, blue-tinted aerial photograph of a dense urban skyline, likely New York City. Overlaid on this image is a semi-transparent blue band that contains a white network or mesh pattern of interconnected lines and dots. Centered within this band is the word "Вопросы?" in a large, white, sans-serif font.

Вопросы?



МИНИТЕСТ

<https://forms.gle/WEQBLAhpeecjYnku5>

3-5 минут

The background of the slide is a high-angle, blue-tinted aerial photograph of a dense urban skyline, likely New York City. Overlaid on this image is a semi-transparent blue band that contains a white network or mesh pattern of interconnected lines and dots. Centered within this band is the main title text.

Программа курса

Программа курса









<https://otus.ru/lessons/nosql-bd>

The background of the slide is a high-angle, blue-tinted aerial photograph of a dense urban skyline, likely New York City. Overlaid on this image is a semi-transparent blue band that contains a white network diagram. This diagram consists of numerous small dots connected by thin white lines, creating a complex web-like structure that spans the width of the slide. Centered within this blue band is the text "Key Value" in a large, white, sans-serif font.

Key Value

Key Value

База данных «ключ-значение» top 8 key value in 2020

TOP NOSQL KEY VALUE DATABASES PAT INDEX™					SORT ▾	
	Amazon DynamoDB	7.6 8.5	🔥 84	≡+ Compare		Oracle NoSQL Database
	Aerospike	7.6 8.5	🔥 45	≡+ Compare		Oracle Berkeley DB
	Voldemort	7.6 8.8	🔥 45	≡+ Compare		InfinityDB
						Redis
						Riak KV













The background of the slide is a high-angle, blue-tinted aerial photograph of a dense urban skyline, likely New York City. Overlaid on this image is a semi-transparent blue band that contains a white network or mesh pattern of interconnected lines and dots. Centered within this band is the text "Document Store" in a large, white, sans-serif font.

Document Store

Document store

[Документоориентированная СУБД — Википедия](#)

<https://www.predictiveanalyticstoday.com/top-nosql-document-databases>

TOP NOSQL DOCUMENT DATABASES			PAT INDEX™	SORT ▼		
	MarkLogic	7.7 7.1 🔥 95	≡+ Compare		InterSystems Caché	7.6 7.9 🔥 95 ≡+ Compare
	MongoDB	7.6 6.2 🔥 83	≡+ Compare		OrientDB	7.6 7.3 🔥 67 ≡+ Compare
	Apache CouchDB	7.6 7.5 🔥 46	≡+ Compare		IBM Cloudant	7.6 8.2 🔥 45 ≡+ Compare
	BaseX	7.6 8.8 🔥 45	≡+ Compare		Couchbase Server	7.6 8.2 🔥 45 ≡+ Compare
	CrateDB	7.6 7.8 🔥 45	≡+ Compare		Azure Cosmos DB	7.6 8.3 🔥 45 ≡+ Compare
	eXist DB	7.6 8.2 🔥 45	≡+ Compare		IBM Informix	7.6 9.6 🔥 45 ≡+ Compare



Graph



Graph

Графовая база данных

<https://www.predictiveanalyticstoday.com/top-free-graph-databases>










The image features a high-angle, blue-tinted aerial photograph of a dense urban skyline, likely New York City. The image is split horizontally by a solid blue band. This band contains a faint, white, geometric network pattern of interconnected lines and dots. Centered within this band is the word "Column" in a large, white, sans-serif font.

Column

Column

Колоночные NoSQL

<https://www.predictiveanalyticstoday.com/top-wide-columnar-store-databases/>

TOP COLUMN-ORIENTED DATABASES			PAT INDEX™	SORT ▼	
	MariaDB	7.6 6.4	95	≡+ Compare	★
	ClickHouse	7.6 7.1	47	≡+ Compare	★
	Greenplum Database	7.6 8.7	47	≡+ Compare	★
	CrateDB	7.6 7.8	45	≡+ Compare	★
	Apache Hbase	7.6 8.4	45	≡+ Compare	★
	Apache Kudu	7.6 7.2	45	≡+ Compare	★
	Apache Parquet	7.6 8.2	45	≡+ Compare	★
	Hypertable	7.6 8.0	45	≡+ Compare	★
	MonetDB	7.6 7.9	45	≡+ Compare	★

В заключение

<https://www.predictiveanalyticstoday.com/?s=top%20nosq>

Варианты облаков

Варианты облаков

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

AWS <https://aws.amazon.com/ru/>

ЯндексОблако <https://cloud.yandex.ru/>

VKCloud <https://mcs.mail.ru/>

SberCloud (возможно заменим Azure) <https://sbercloud.ru/ru>

Selectel <https://selectel.ru/services/cloud/servers/>


Погнали в облака



Практика

Установим GCloud

<https://cloud.google.com/sdk/docs/downloads-apt-get>

The background of the image is an aerial photograph of a dense city skyline, likely New York City, with numerous skyscrapers. The image is overlaid with a semi-transparent blue gradient. A network of thin, light blue lines connects various points across the blue area, creating a web-like pattern. The word "BigQuery" is centered in the middle of the image in a white, sans-serif font.

BigQuery

BigQuery

200 млн записей

```
select count(*)  
from bigquery-public-data.chicago_taxi_trips.taxi_trips;
```

Сколько будет выполняться следующий запрос?

```
select payment_type, round(sum(tips)/sum(tips+fare)*100) tips_persent, count(*)  
from bigquery-public-data.chicago_taxi_trips.taxi_trips  
group by payment_type  
order by 3 desc;
```


The image features a high-angle, blue-tinted aerial photograph of a dense urban skyline, likely New York City, with numerous skyscrapers and buildings. A semi-transparent blue band with a white geometric network pattern of dots and lines runs horizontally across the center of the image. The Russian word "Порефлексируем" is written in white, bold, sans-serif font across this band.


Порефлексируем

Результаты опроса

https://docs.google.com/forms/d/1J1WmYFsSqaufHiVqF9_DK53E4zzgx68cuuMTz1MSxfo/edit#responses

Вопросы?

- **Что узнали нового?**


The image features a blue-tinted aerial view of a city skyline, likely New York City, with numerous skyscrapers. A semi-transparent network pattern of white lines and dots is overlaid on the image, particularly prominent in the center. The text 'ДЗ' is displayed in a large, white, sans-serif font in the center of the image.

ДЗ


ДЗ

нет его %)

но можно потренироваться в markdown и github

The background of the image is an aerial photograph of a dense city skyline, likely New York City, with numerous skyscrapers. The entire image is overlaid with a semi-transparent blue layer. A network of white lines and dots, resembling a molecular or digital structure, is visible across the blue overlay, particularly concentrated around the text area.

Заполните, пожалуйста,
опрос о занятии по ссылке в чате
<https://otus.ru/polls/52764/>



Спасибо за внимание!
Приходите на следующие вебинары

Аристов Евгений