

# Управление данными, 22

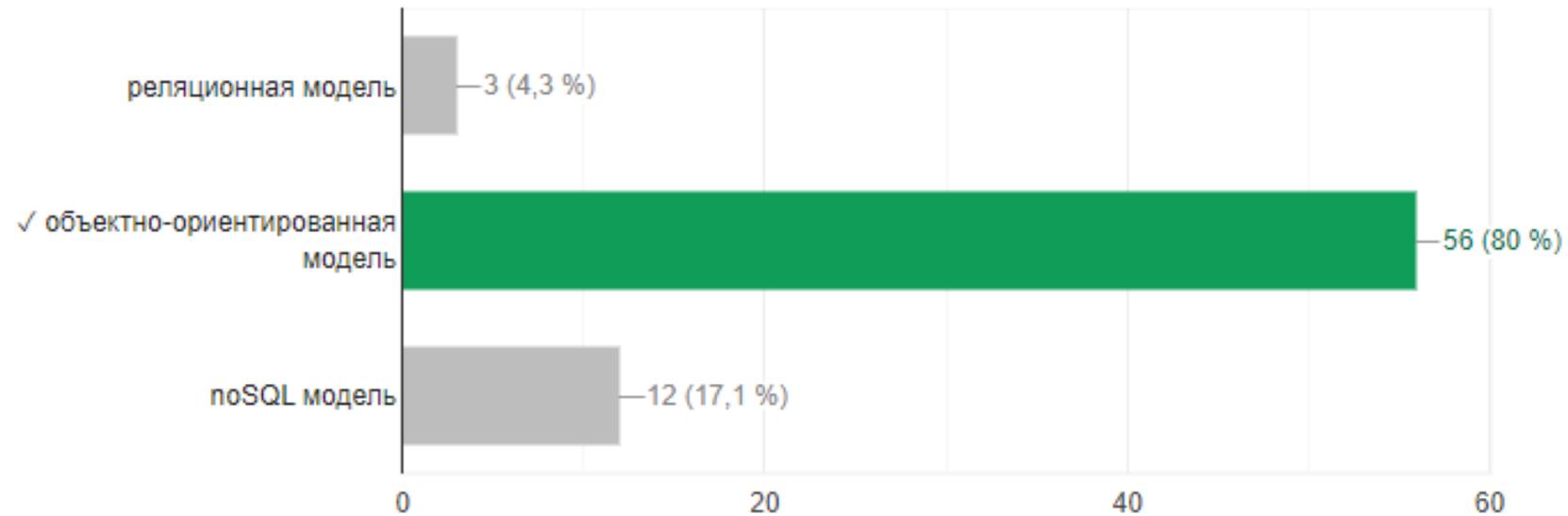
ФИО преподавателя: Смирнов Михаил Вячеславович

e-mail: [smirnovmgupi@gmail.com](mailto:smirnovmgupi@gmail.com)

Какая модель баз данных не будет рассмотрена в курсе Управление данными?

 Копировать

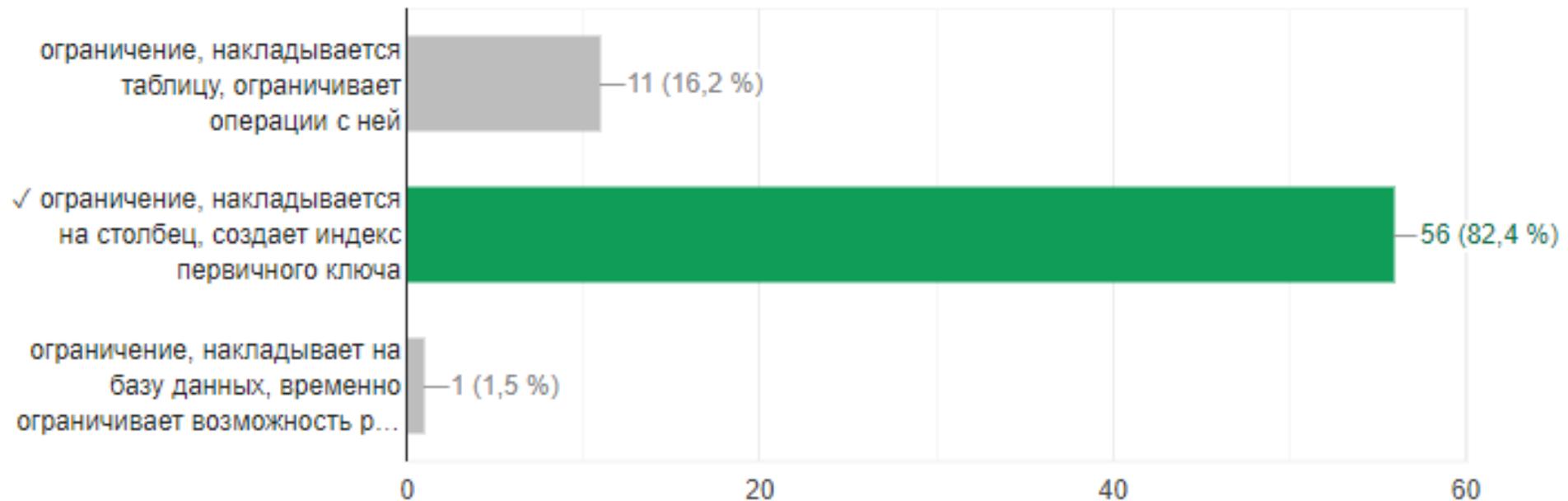
Верных ответов: 55 из 70



## Ограничение CONSTRAINT, это

 Копировать

Верных ответов: 56 из 68



# ОПРЕДЕЛИМ ПРЕДМЕТ ДИСКУССИИ



Данные.

Форма представления информации в виде фактов.

Факты можно изучать, классифицировать, по ним можно делать выводы и получать новые знания.

Благодаря данным с вашего Apple Watch я могу узнать практически все о вашем образе жизни. И даже спрогнозировать, что вы будете завтра... или через неделю.



Базы данных.

Программное обеспечение, фреймворки, технологии, главной задачей которых является обеспечение возможности безопасного хранения данных.

Их ОЧЕНЬ много.

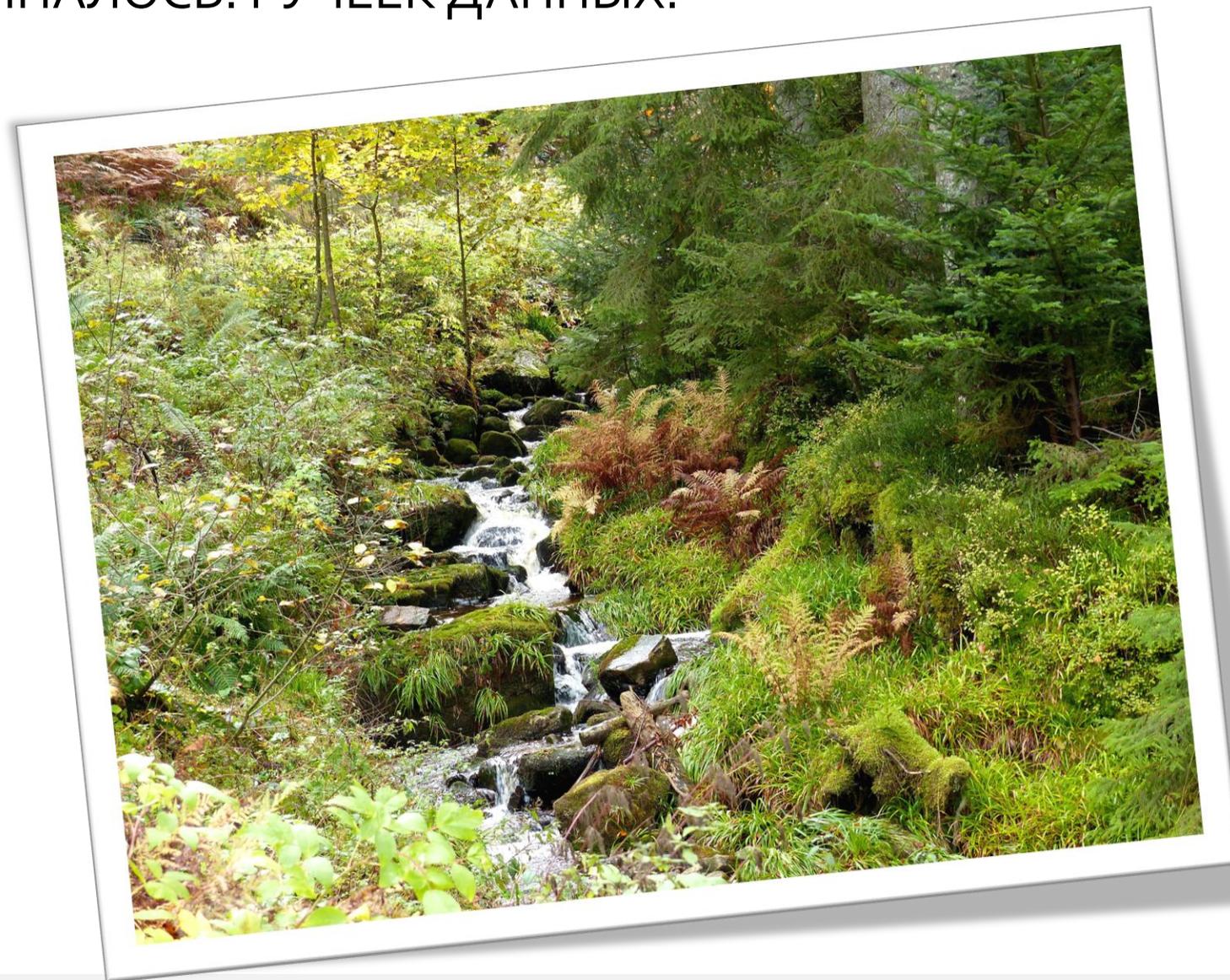




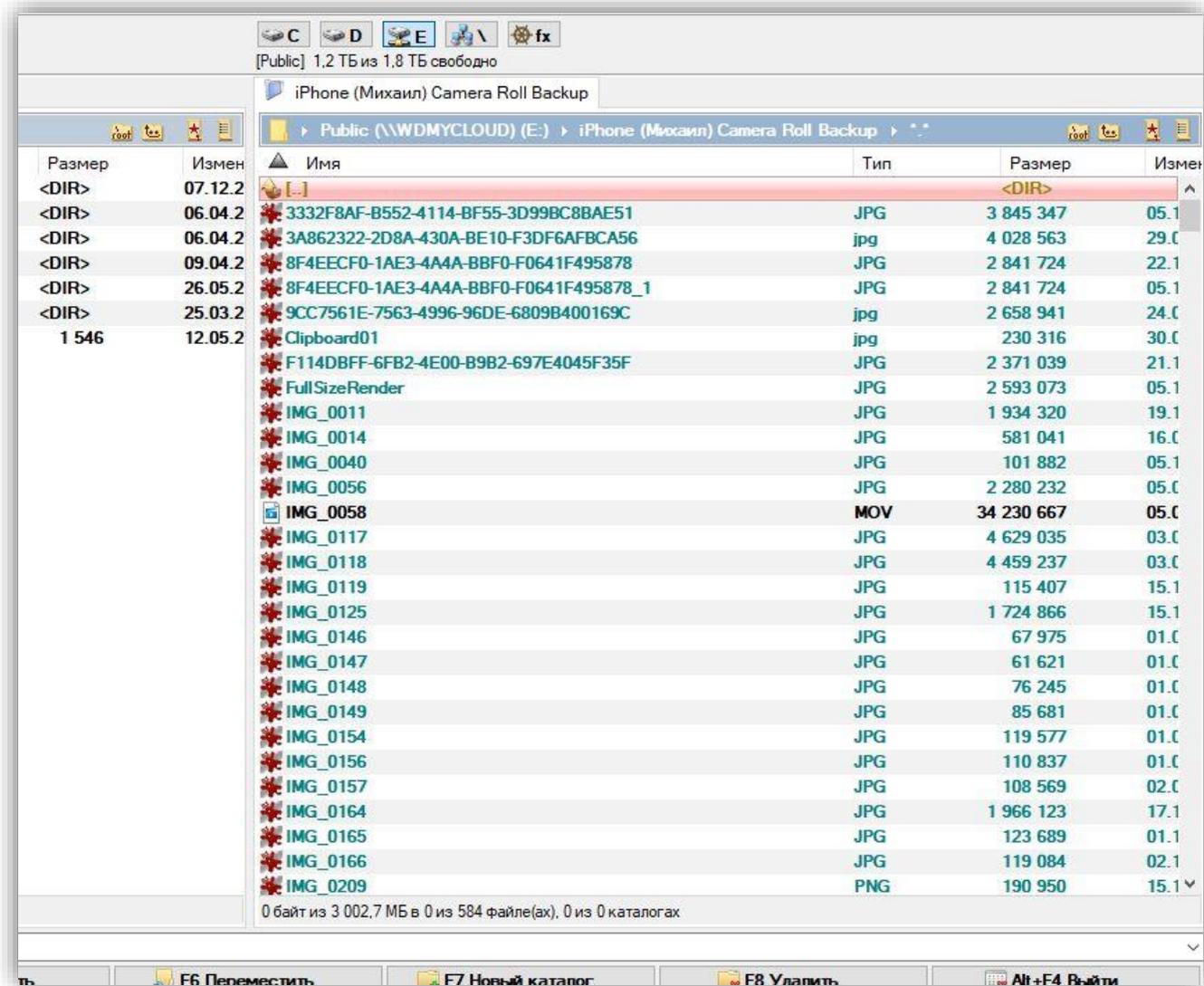
# ДААННЫЕ В ЦИФРОВОМ МИРЕ

Эволюция данных и подходов к их хранению.

КАК ВСЕ НАЧИНАЛОСЬ. РУЧЕЕК ДАННЫХ.



# ДААННЫЕ В ВИДЕ ФАЙЛОВ.



НУ, ЭТО ЖЕ НЕ ПРОБЛЕМА?



**И ТАК  
СОЙДЕТ!**

ДАННЫХ ВСЕ БОЛЬШЕ. РЕКА ДАННЫХ.



# РЕЛЯЦИОННЫЙ ПОРЯДОК В ДАННЫХ. РЫБАЛКА ЛИНКА.

Озеро Дейа

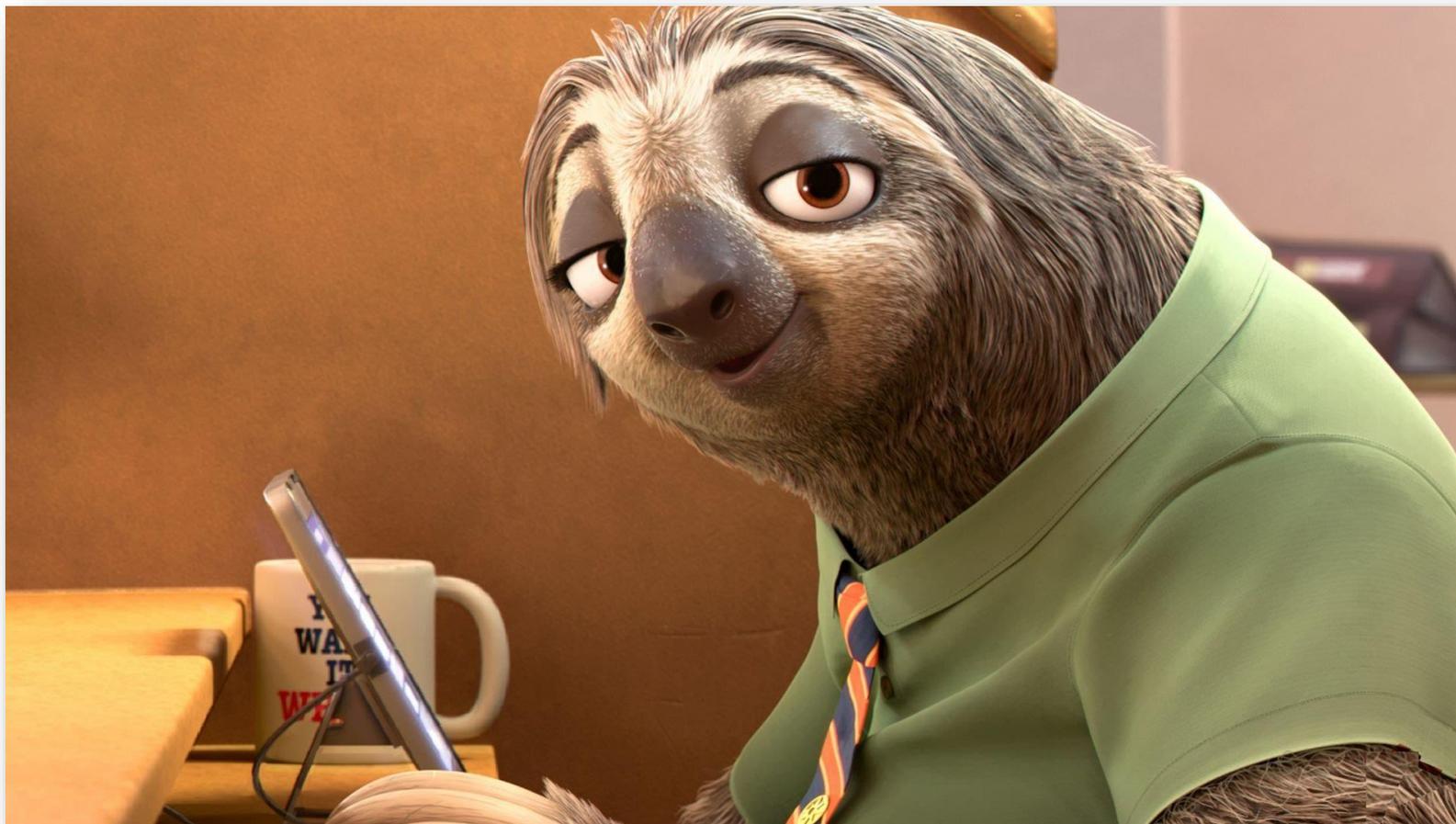
ДАТА	КОЛ-ВО	РЫБА
17.05.2021	412	Муранья
03.07.2021	84	Сладкая рыба-пастух
03.07.2021	9	Лорд Чапу-Чапу
03.07.2021	117	Веселый карп

Рыба Муранья

ДАТА	КОЛ-ВО	ВОДОЕМ
17.05.2021	412	Озеро Дейа
03.07.2021	84	Озеро Джаррах
03.07.2021	9	Озеро Кора
03.07.2021	117	Озеро Нирвата



НАДЕЖНО, НО... НЕБЫСТРО.



# ЭПОХА СОЦСЕТЕЙ И ЦИФРОВИЗАЦИИ. ВОДОПАД ДАННЫХ.



# ТОЧКИ ПРИЛОЖЕНИЯ ДАННЫХ В СОВРЕМЕННОМ МИРЕ

Государство

Мобильные приложения

Компьютерные игры

Социальные сети

Интернет магазины

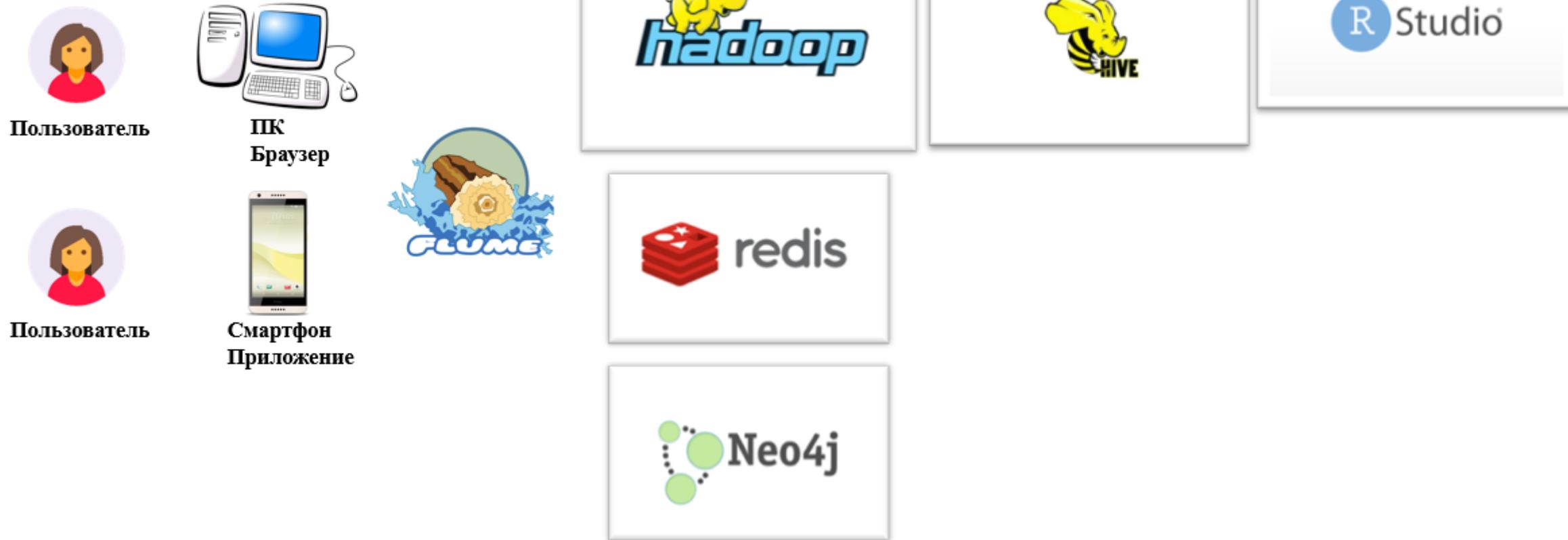
Интернет вещей

Десктопные приложения

Корпорации

Банки и финучреждения

# ПРОСТОЙ СЦЕНАРИЙ "ПРИКЛЮЧЕНИЯ ДАННЫХ" В СОЦИАЛОЧКАХ



# АКТУАЛЬНЫЕ СПОСОБЫ ХРАНЕНИЯ ДАННЫХ

Программное обеспечение, методы.

# РЕЛЯЦИОННЫЕ И ПОСТРЕЛЯЦИОННЫЕ БАЗЫ ДАННЫХ



*Реляционные базы данных и хранилища данных.*

Все упорядочено по строгим правилам.

Дружелюбный, похожий на естественный язык программирования (SQL).

Гарантирует правильность и целостность данных. Всегда. Не даст совершить ошибку.

Долгая подготовка, дорогие изменения.



*Постреляционные базы данных и хранилища данных.*

Упорядочим потом. Как-нибудь. Но это неточно.

Много скобочек. Разных. Очень много! (JSON).

Гарантирует правильность данных. Когда-нибудь. Но это неточно.

Короткая подготовка. Моментальные изменения.



# РЕЛЯЦИОННЫЕ И ПОСТРЕЛЯЦИОННЫЕ БАЗЫ ДАННЫХ

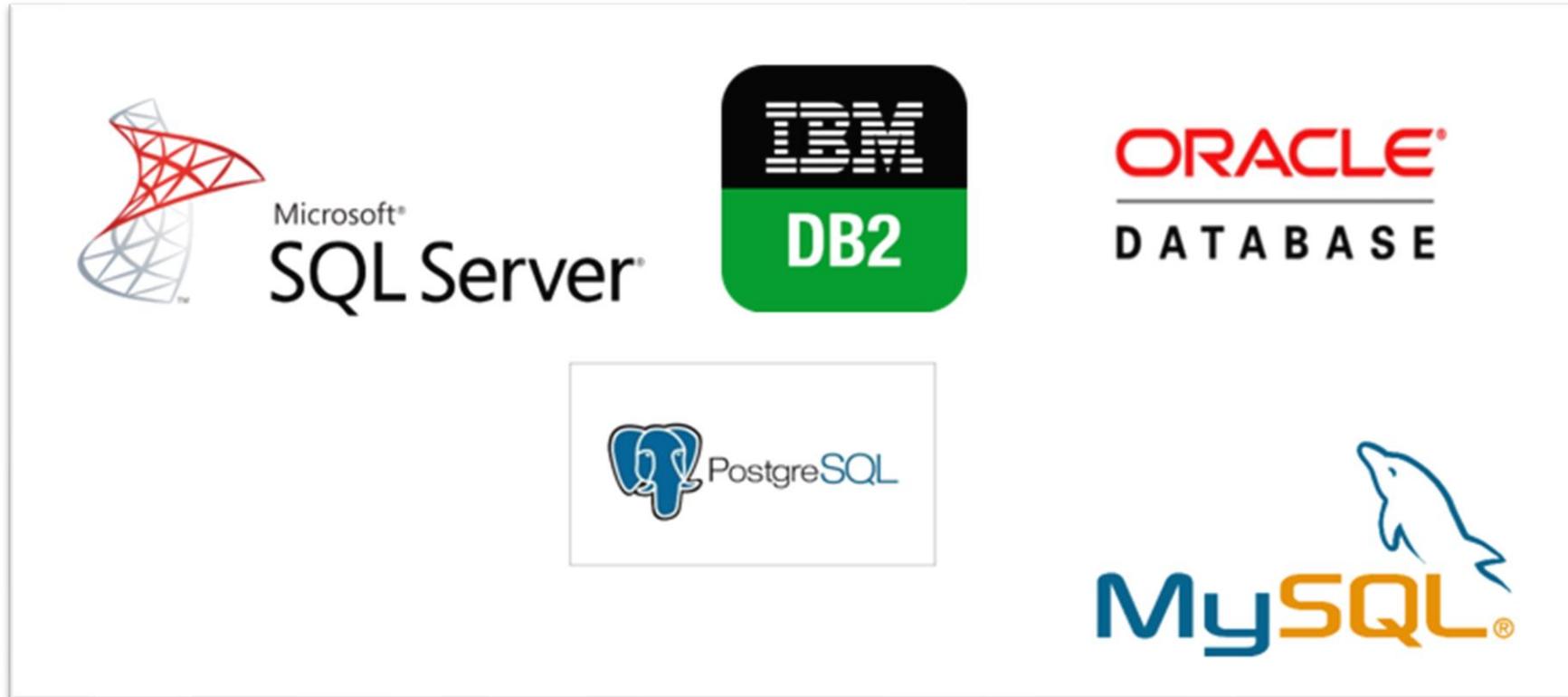
Реляционная модель хранения



Постреляционная модель хранения



# ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ РЕЛЯЦИОННОЙ ПАРАДИГМЫ



# ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ ПОСТРЕЛЯЦИОННОЙ ПАРАДИГМЫ

Хранилище  
"ключ-значение"



Хранилище семейств  
колонок



Документо-ориентированная  
СУБД



# ПРАКТИКА РАБОТЫ С ДАННЫМИ

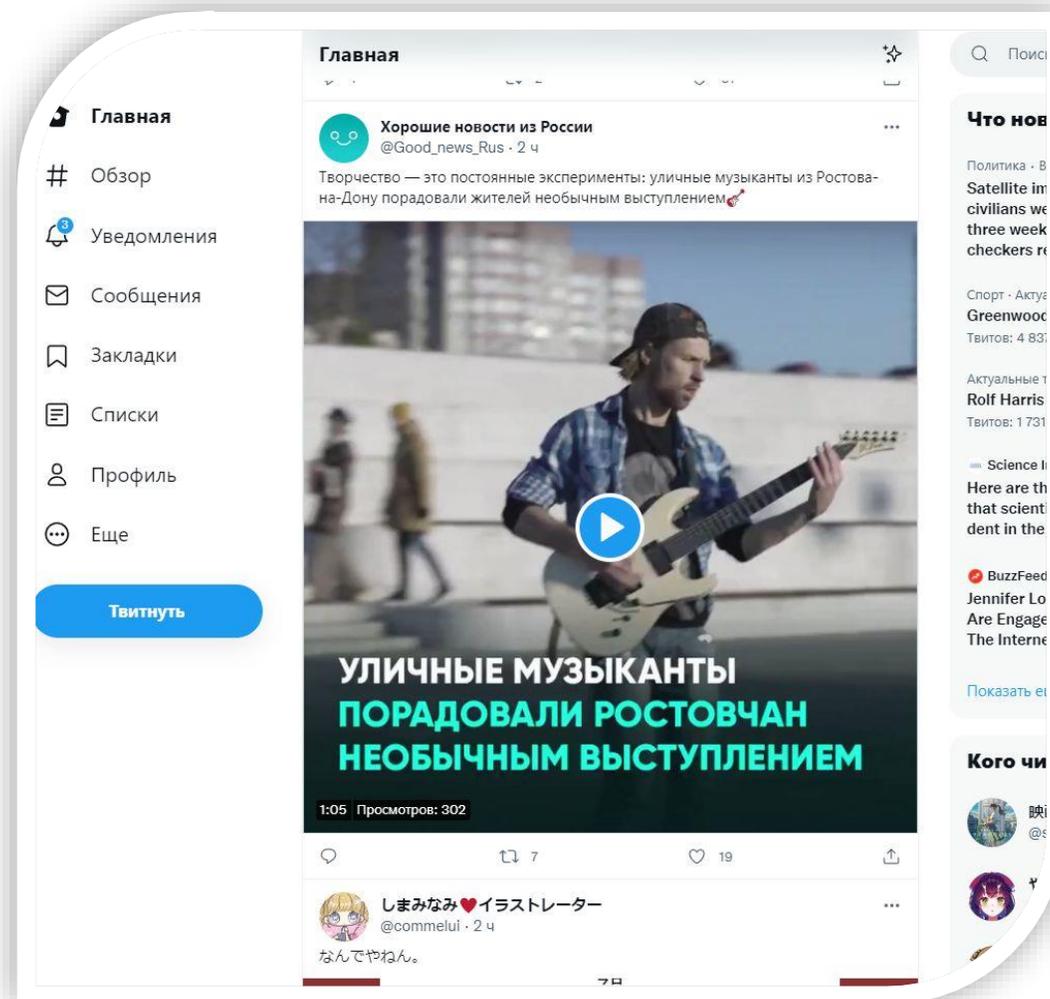
Пример технологического стека компании Twitter.  
Администрирование баз данных.



## ОПЫТ КОМПАНИИ TWITTER.

Десятилетний опыт построения действительно огромной и комплексной базы данных для удовлетворения нужд огромной аудитории сервиса.

# ВЕБ-ПРИЛОЖЕНИЕ (СОЦИАЛЬНАЯ СЕТЬ) TWITTER



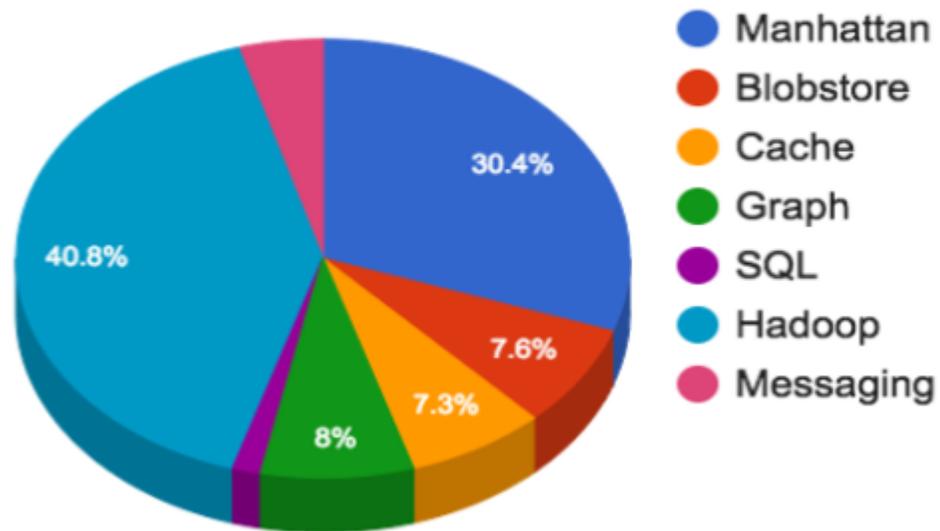
# ПРАКТИЧЕСКИЕ ЗАДАЧИ, СТОЯВШИЕ ПЕРЕД СПЕЦИАЛИСТАМИ TWITTER



- За годы работы суммарный объем данных, накопленный на серверах (без мультимедиа) – около 500 петабайт (квадриллион байт, google it (с)). Где и как хранить то, что уже есть... а главное – где и как хранить то, новое, что приходит бешеными темпами?
- Обработка оперативной информации – твиты, личные сообщения, должна происходить моментально. Учитывая их количество, это десятки миллионов QPS по всему миру (в СЕКУНДУ, КАРЛ!!)
- Управление системой рекомендаций (что почитать, на кого подписаться...). Социальный граф огромного размера!
- Мультимедиа пользователей (сотни миллиардов картинок, видеороликов) – куда это девать?
- Чем кешировать метаинформацию (действия пользователей, таймлайны) для моментального доступа клиентских приложений к ней?
- Где хранить финансовую, рекламную и другую информацию строгой отчетности? Потеряется твит или личное сообщение – пользователь будет недоволен, но это не критично? Потеряется оплаченная реклама – это огромный удар по репутации компании!

[https://blog.twitter.com/engineering/en\\_us/topics/infrastructure/2017/the-infrastructure-behind-twitter-scale](https://blog.twitter.com/engineering/en_us/topics/infrastructure/2017/the-infrastructure-behind-twitter-scale)

# СТРУКТУРА ТЕХНОЛОГИЧЕСКОГО СТЕКА TWITTER ДЛЯ ХРАНЕНИЯ ДАННЫХ (ПОСТОЯННО ЭВОЛЮЦИОНИРУЕТ)



1. Кластеры Hadoop для вычислений и HDFS.
2. Кластеры Manhattan для всех хранилищ key-value с малой задержкой.
3. Хранилища Graph для шардированных кластеров MySQL.
4. Кластеры Blobstore для всех крупных объектов (видео, изображения, бинарные файлы...).
5. Кластеры кэширования.
6. Кластеры обмена сообщениями.
7. Реляционные хранилища (MySQL, PostgreSQL и Vertica).

[https://blog.twitter.com/engineering/en\\_us/topics/infrastructure/2017/the-infrastructure-behind-twitter-scale](https://blog.twitter.com/engineering/en_us/topics/infrastructure/2017/the-infrastructure-behind-twitter-scale)



## А ЧТО С ЭКСПЛУАТАЦИЕЙ ГОТОВЫХ БАЗ ДАННЫХ?

Задачи администраторов программного обеспечения баз данных.

# ЗАДАЧИ АДМИНИСТРАТОРОВ БАЗ ДАННЫХ



*Чтобы все работало.*

Постоянная проверка данных.

Мониторинг правильности работы интерфейсов и программного обеспечения.

Автоматизация задач всем доступными способами, чтобы можно было попить кофейку и посмотреть сериальчик, пока хорошо отлаженная система работает (прохладная история про мастера PERL скриптов из комьюнити).



*Чтобы все работало максимально хорошо.*

Эволюция программного обеспечения.

Эволюция программного кода.

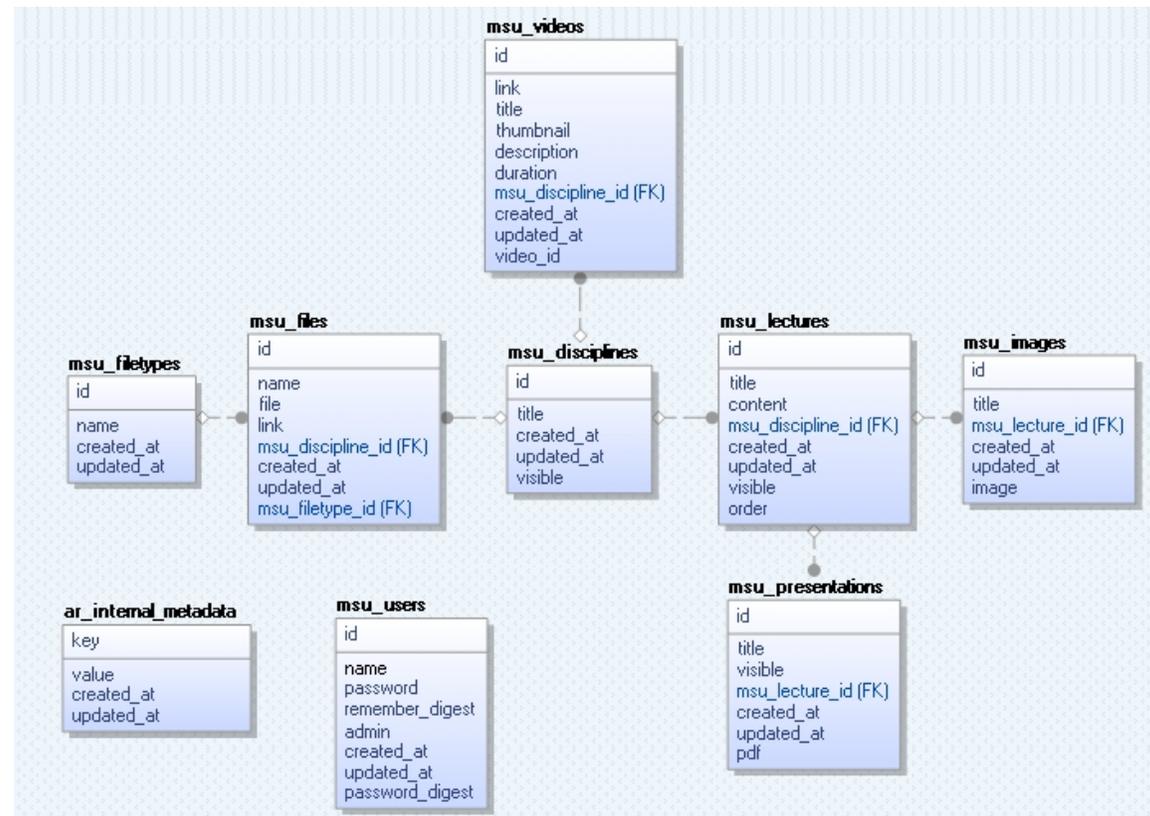
Оптимизация работы компонентов системы хранения данных.



*Чтобы никто это не сломал.*

Мероприятия, связанные с постоянной защитой данных пользователей (зачастую не приводящие к положительному результату, но все-таки необходимые).

# Простые структуры хранения данных для приложений



# Типы и образцы баз данных

Приложение	Пример пользователя	Количество пользователей	Примерный размер
Контакты менеджера по продажам	Менеджер по продажам	1	2000 строк
Прием пациентов в больнице	Больница	15-50	100000 строк
CRM	Отдел продаж, работы с клиентами, etc	500	10 млн. строк
ERP	Вся организация	5000	10 млн. + строк
E-commerce	Интернет пользователи	Возможно свыше 1 млн.	1 млрд + строк
Data mining	Бизнес-аналитики	25	100000 - 1 млн. строк

# Функции приложений баз данных

## Основные функции приложения баз данных

Обработка форм

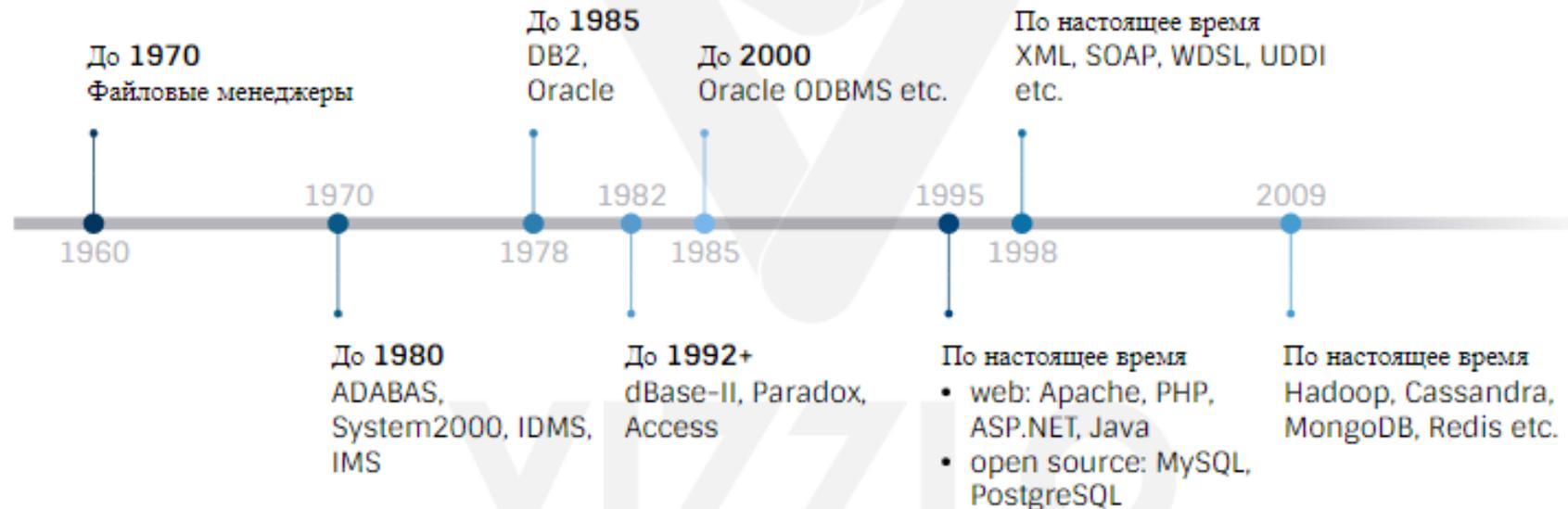
Обработка запросов пользователей

Создание и обработка отчетов

Реализация логики приложения

Управление приложением

# Развитие технологий хранения данных



# Чтение на дом

- Русский Кренке, стр. 24-76.
- Системы баз данных (полный курс), стр. 31-51.
- Английский Кренке, стр. 3-34.

Спасибо за внимание!