

Управление данными, 25

ФИО преподавателя: Смирнов Михаил Вячеславович

e-mail: smirnovmgupi@gmail.com

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.

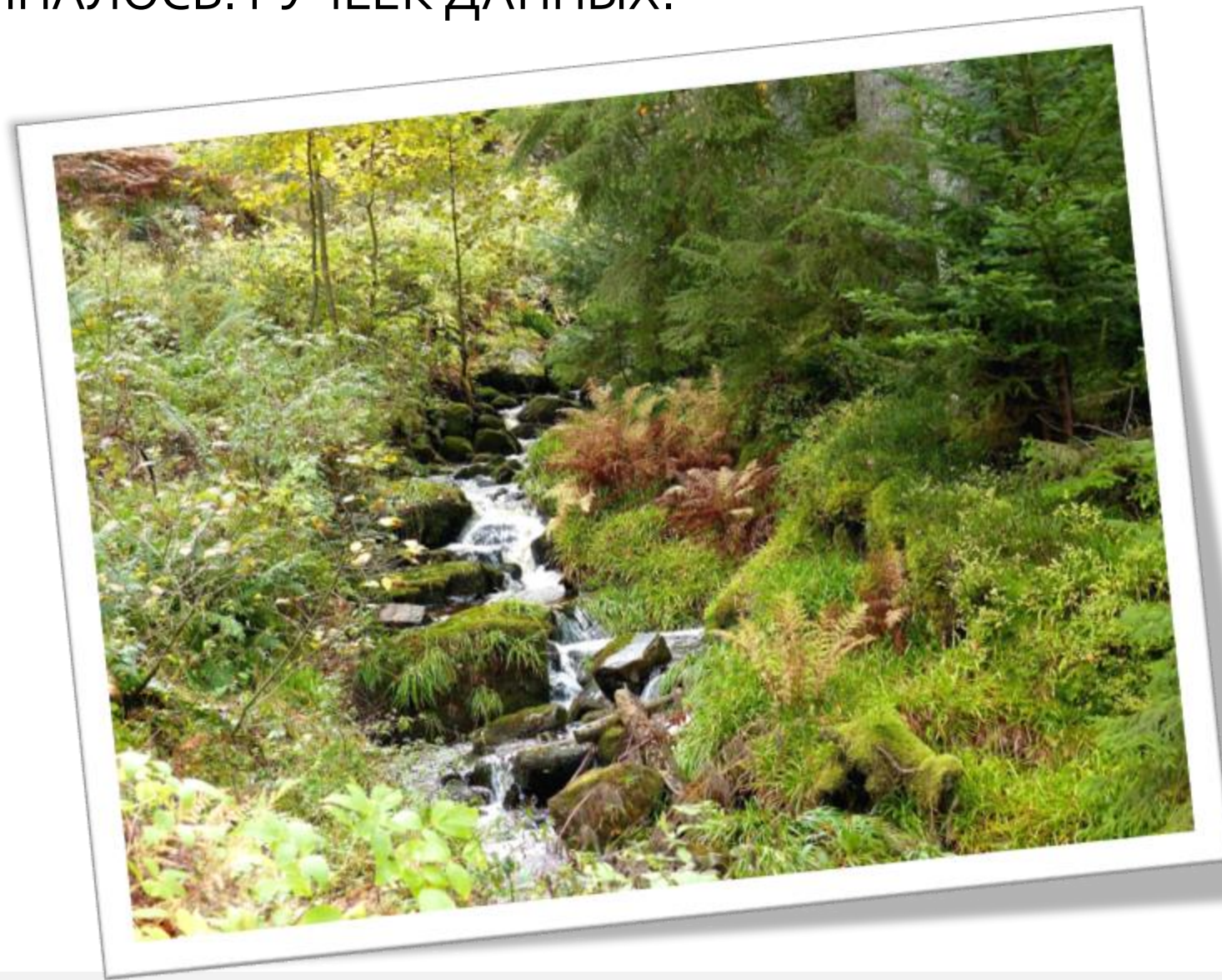




ДАННЫЕ В ЦИФРОВОМ МИРЕ

Эволюция данных и подходов к их хранению.

КАК ВСЕ НАЧИНАЛОСЬ. РУЧЕЕК ДАННЫХ.



ДАННЫЕ В ВИДЕ ФАЙЛОВ.

Public

1,2 ТБ из 1,8 ТБ свободно

iPhone (Михаил) Camera Roll Backup

Public (\WDMYCLOUD) (E:) > iPhone (Михаил) Camera Roll Backup > .*

Размер	Измен	Имя	Тип	Размер	Измен
<DIR>	07.12.2	<DIR>			
<DIR>	06.04.2	* 3332F8AF-B552-4114-BF55-3D99BC8BAE51	JPG	3 845 347	05.1
<DIR>	06.04.2	* 3A862322-2D8A-430A-BE10-F3DF6AFBCA56	jpg	4 028 563	29.0
<DIR>	09.04.2	* 8F4EECF0-1AE3-4A4A-BBF0-F0641F495878	JPG	2 841 724	22.1
<DIR>	26.05.2	* 8F4EECF0-1AE3-4A4A-BBF0-F0641F495878_1	JPG	2 841 724	05.1
<DIR>	25.03.2	* 9CC7561E-7563-4996-96DE-6809B400169C	jpg	2 658 941	24.0
1 546	12.05.2	* Clipboard01	jpg	230 316	30.0
		* F114DBFF-6FB2-4E00-B9B2-697E4045F35F	JPG	2 371 039	21.1
		* FullSizeRender	JPG	2 593 073	05.1
		* IMG_0011	JPG	1 934 320	19.1
		* IMG_0014	JPG	581 041	16.0
		* IMG_0040	JPG	101 882	05.1
		* IMG_0056	JPG	2 280 232	05.0
		* IMG_0058	MOV	34 230 667	05.0
		* IMG_0117	JPG	4 629 035	03.0
		* IMG_0118	JPG	4 459 237	03.0
		* IMG_0119	JPG	115 407	15.1
		* IMG_0125	JPG	1 724 866	15.1
		* IMG_0146	JPG	67 975	01.0
		* IMG_0147	JPG	61 621	01.0
		* IMG_0148	JPG	76 245	01.0
		* IMG_0149	JPG	85 681	01.0
		* IMG_0154	JPG	119 577	01.0
		* IMG_0156	JPG	110 837	01.0
		* IMG_0157	JPG	108 569	02.0
		* IMG_0164	JPG	1 966 123	17.1
		* IMG_0165	JPG	123 689	01.1
		* IMG_0166	JPG	119 084	02.1
		* IMG_0209	PNG	190 950	15.1

0 байт из 3 002,7 МБ в 0 из 584 файле(ах), 0 из 0 каталогах

F6 Переместить

F7 Новый каталог

F8 Удалить

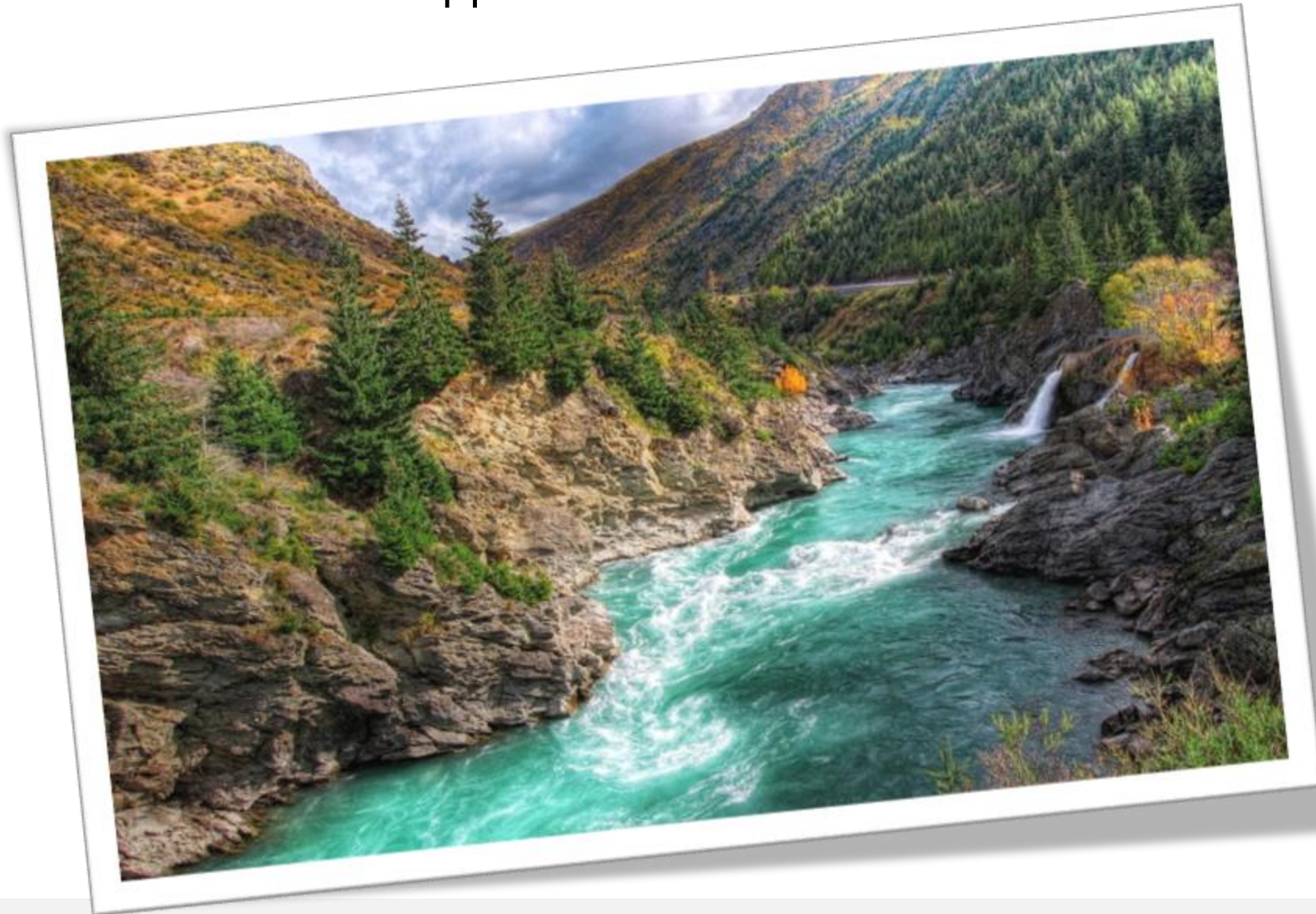
Alt+F4 Выйти

НУ, ЭТО ЖЕ НЕ ПРОБЛЕМА?



**И ТАК
СОЙДЕТ!**

ДАННЫХ ВСЕ БОЛЬШЕ. РЕКА ДАНЫХ.



РЕЛЯЦИОННЫЙ ПОРЯДОК В ДАННЫХ. РЫБАЛКА ЛИНКА.

Озеро Дейа

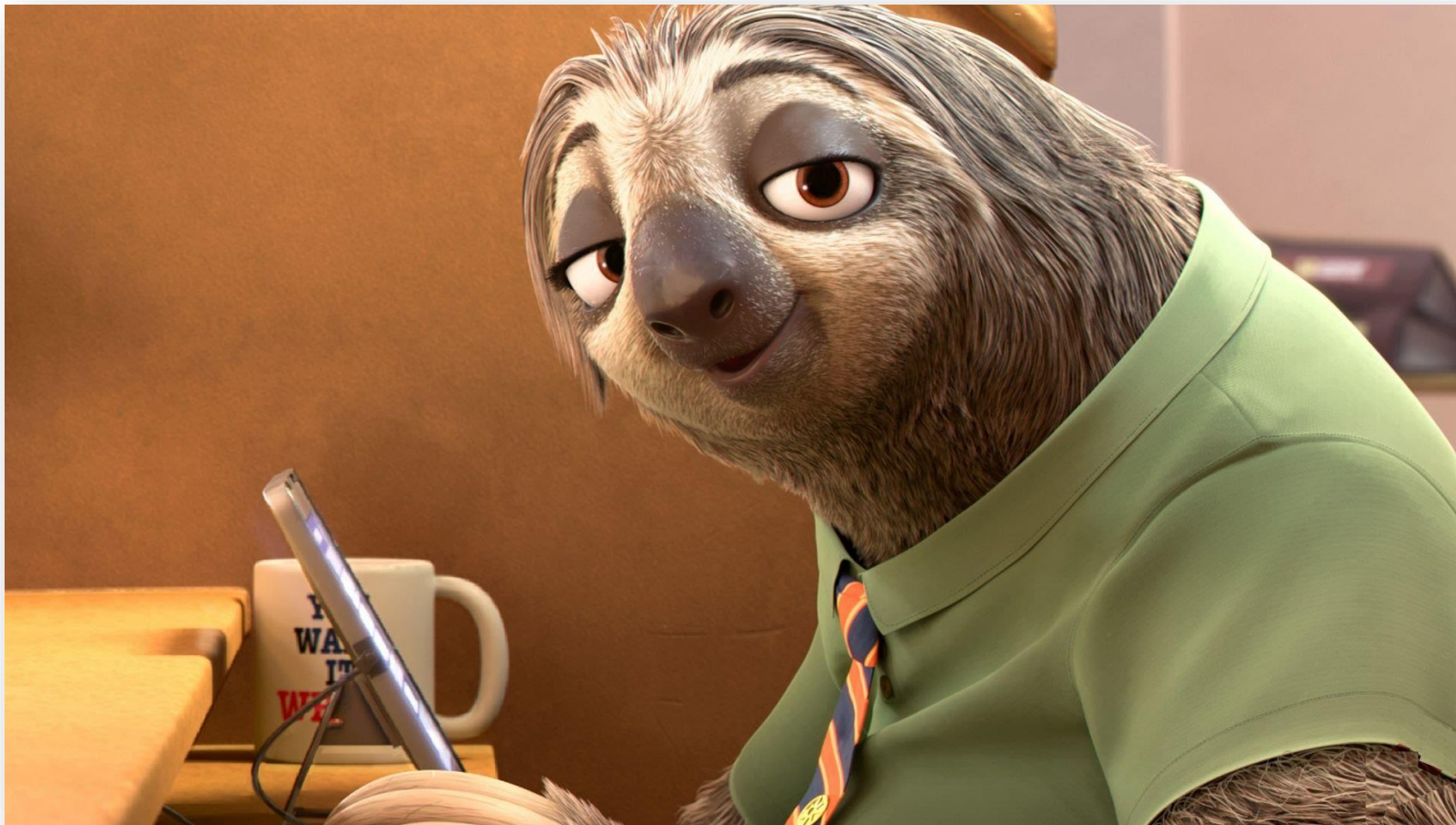
ДАТА	КОЛ-ВО	РЫБА
17.05.2021	412	Муранья
03.07.2021	84	Сладкая рыба-пастух
03.07.2021	9	Лорд Чапу-Чапу
03.07.2021	117	Веселый карп

Рыба Муранья

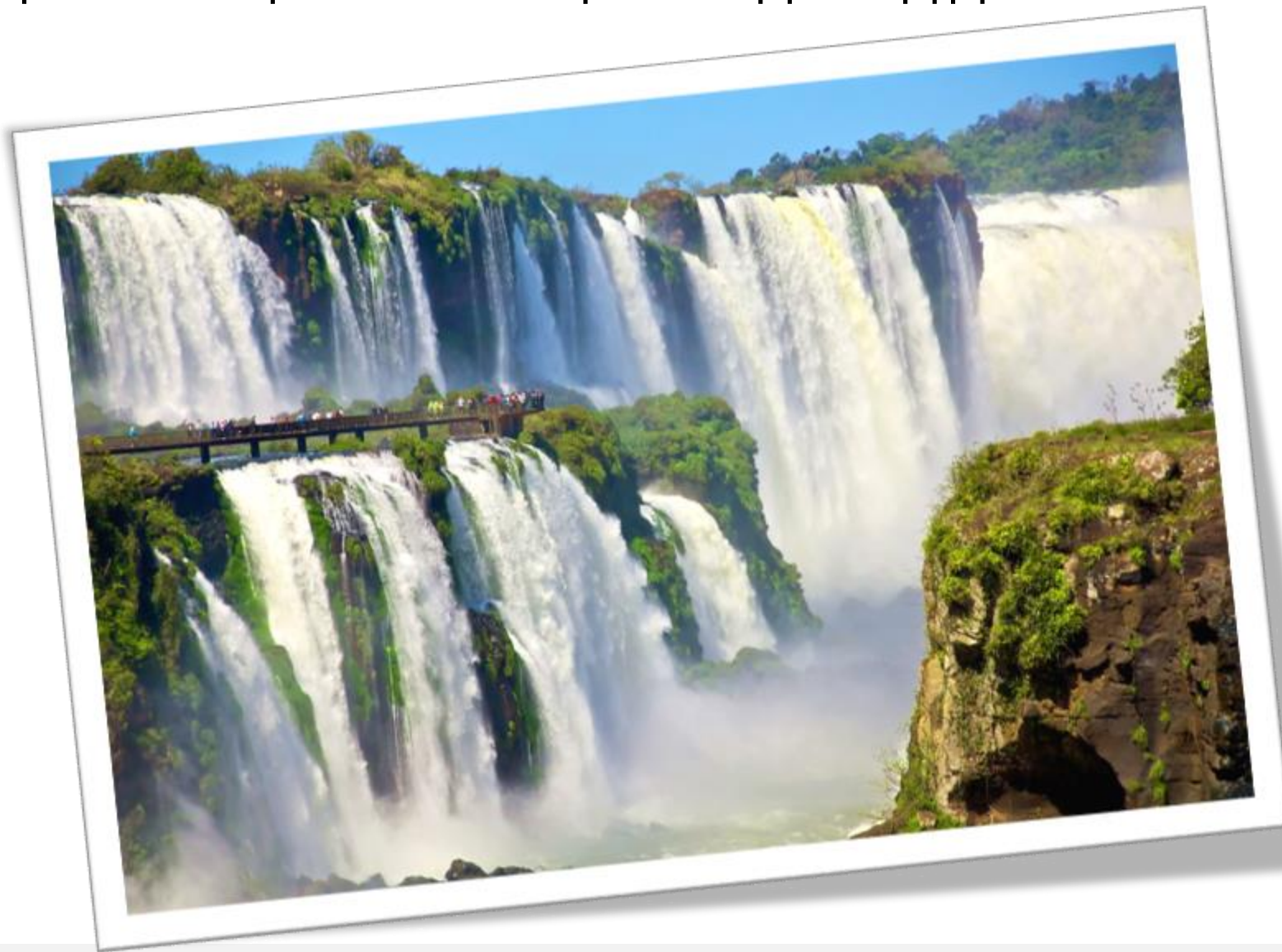
ДАТА	КОЛ-ВО	ВОДОЕМ
17.05.2021	412	Озеро Дейа
03.07.2021	84	Озеро Джаррах
03.07.2021	9	Озеро Кора
03.07.2021	117	Озеро Нирвата



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



Пользователь

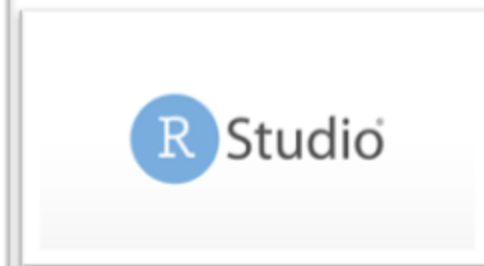
ПК
Браузер

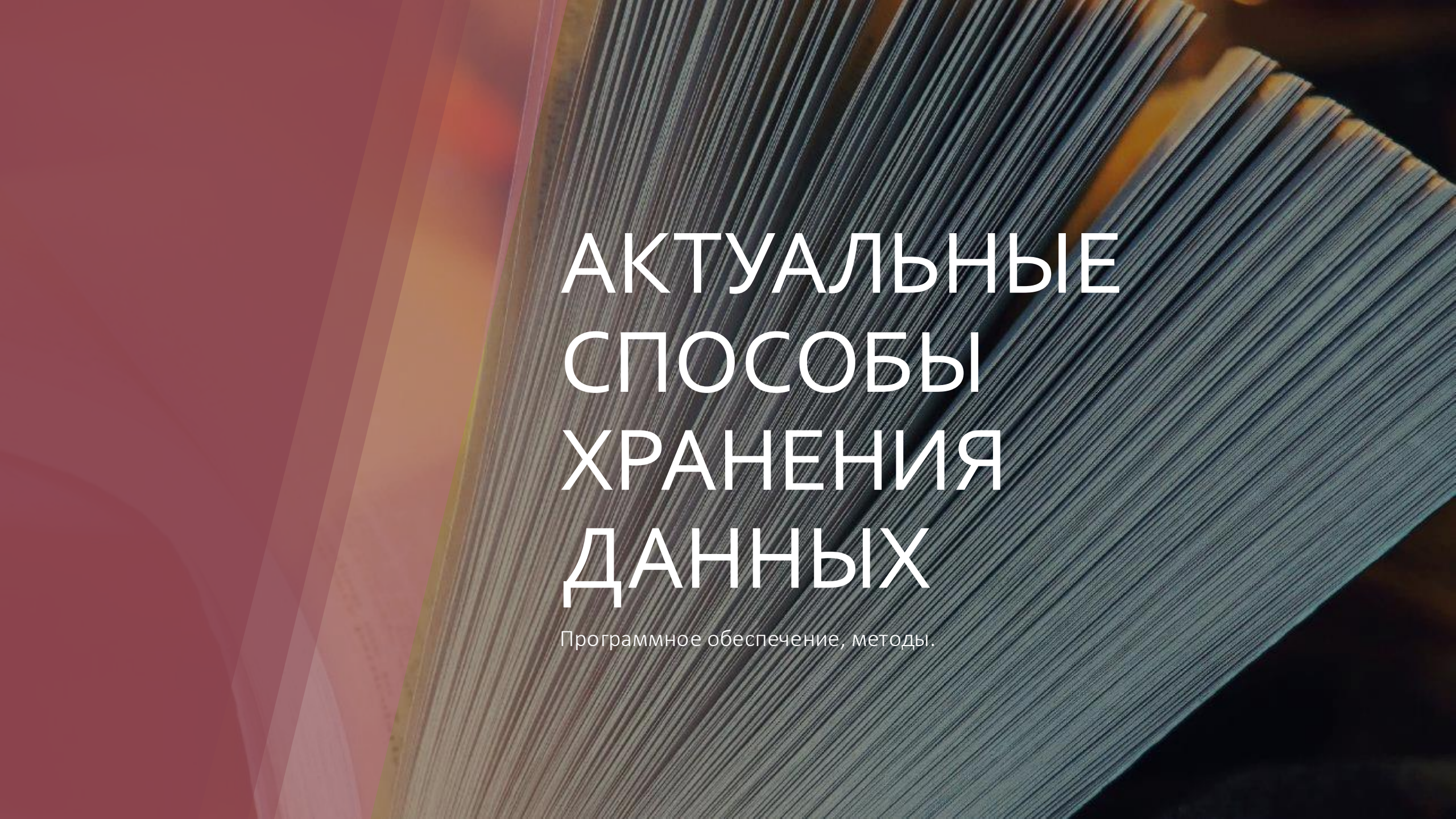


Пользователь



Смартфон
Приложение





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

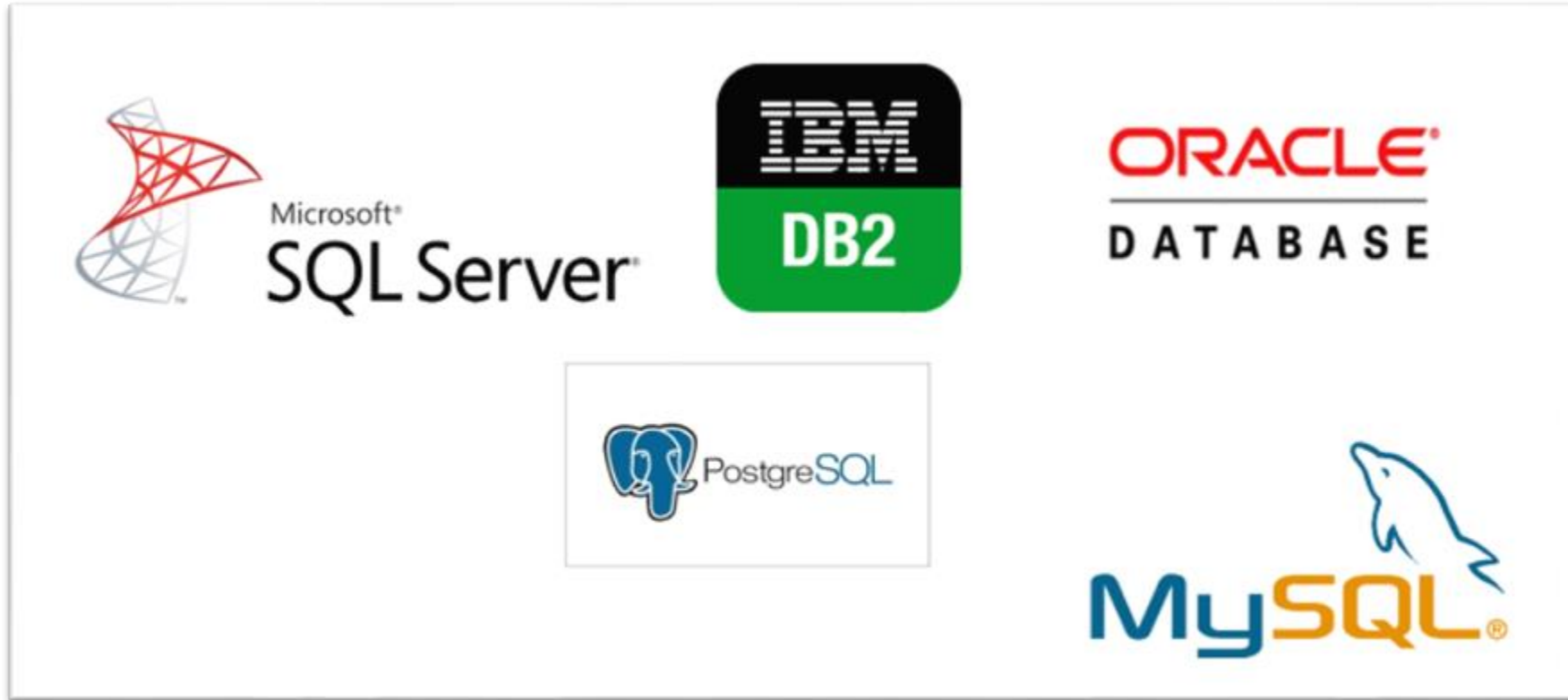
Реляционная модель хранения



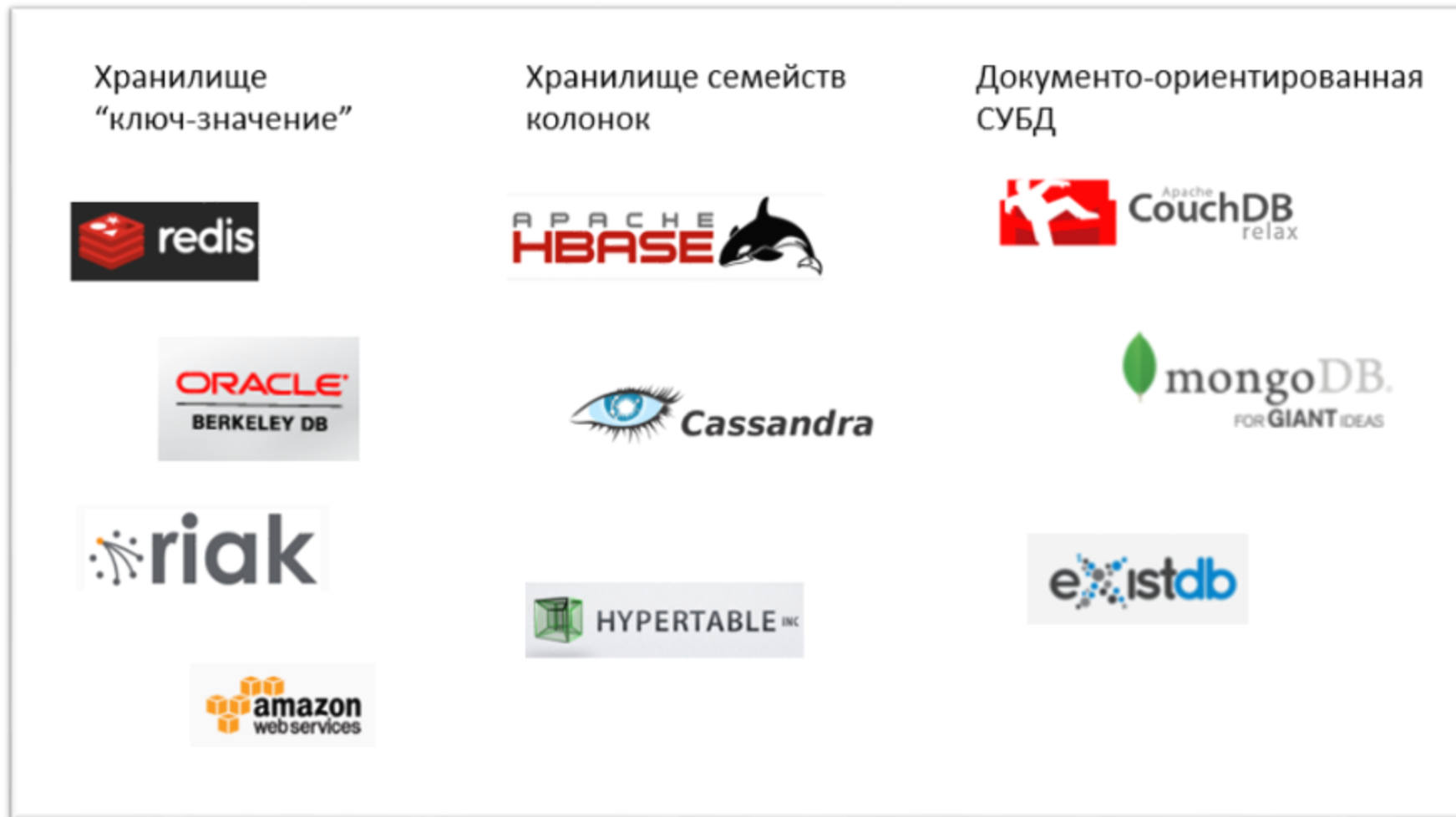
Постреляционная модель хранения

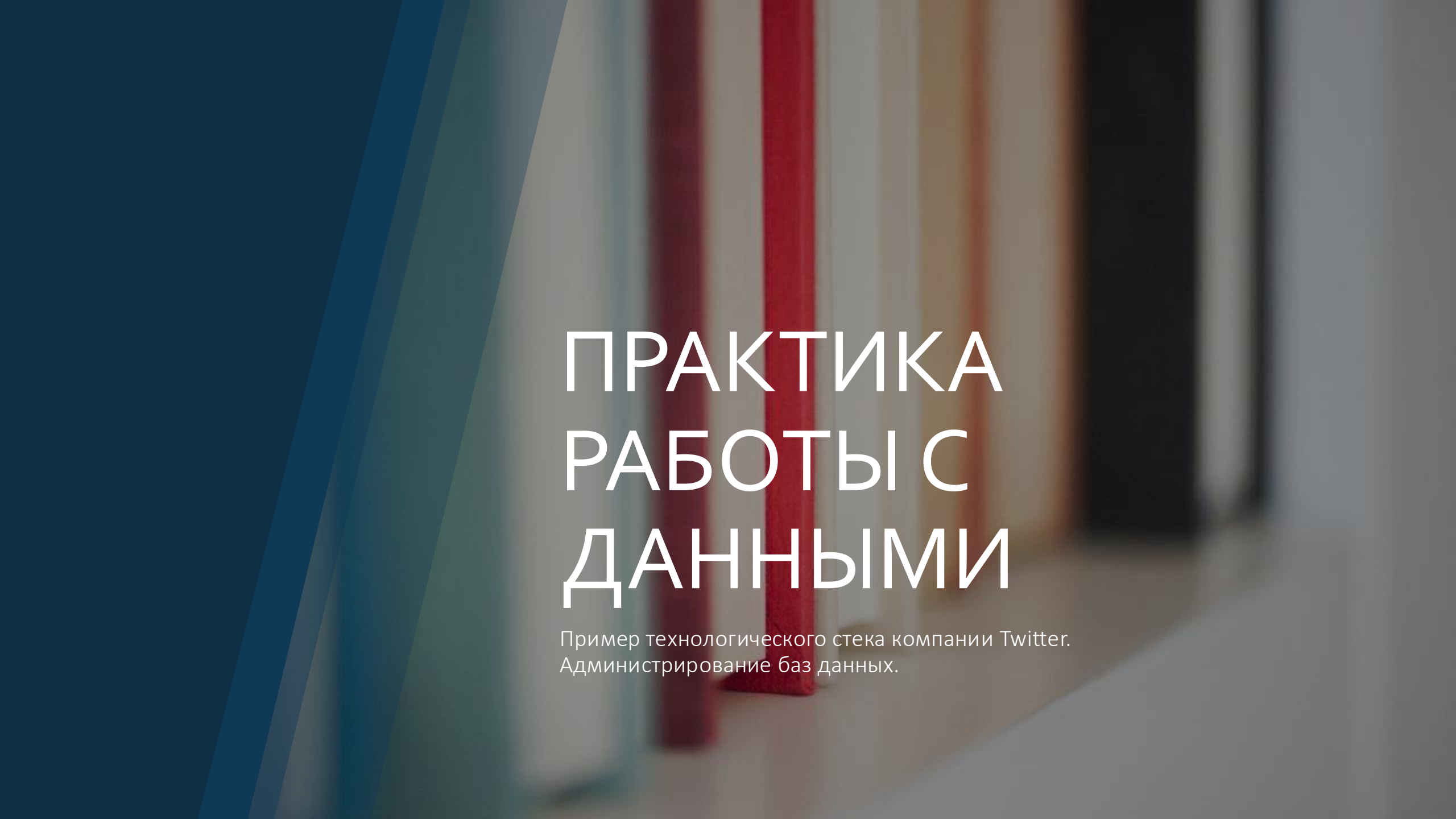


ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ РЕЛЯЦИОННОЙ ПАРАДИГМЫ



ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ ПОСТРЕЛЯЦИОННОЙ ПАРАДИГМЫ





ПРАКТИКА РАБОТЫ С ДАННЫМИ

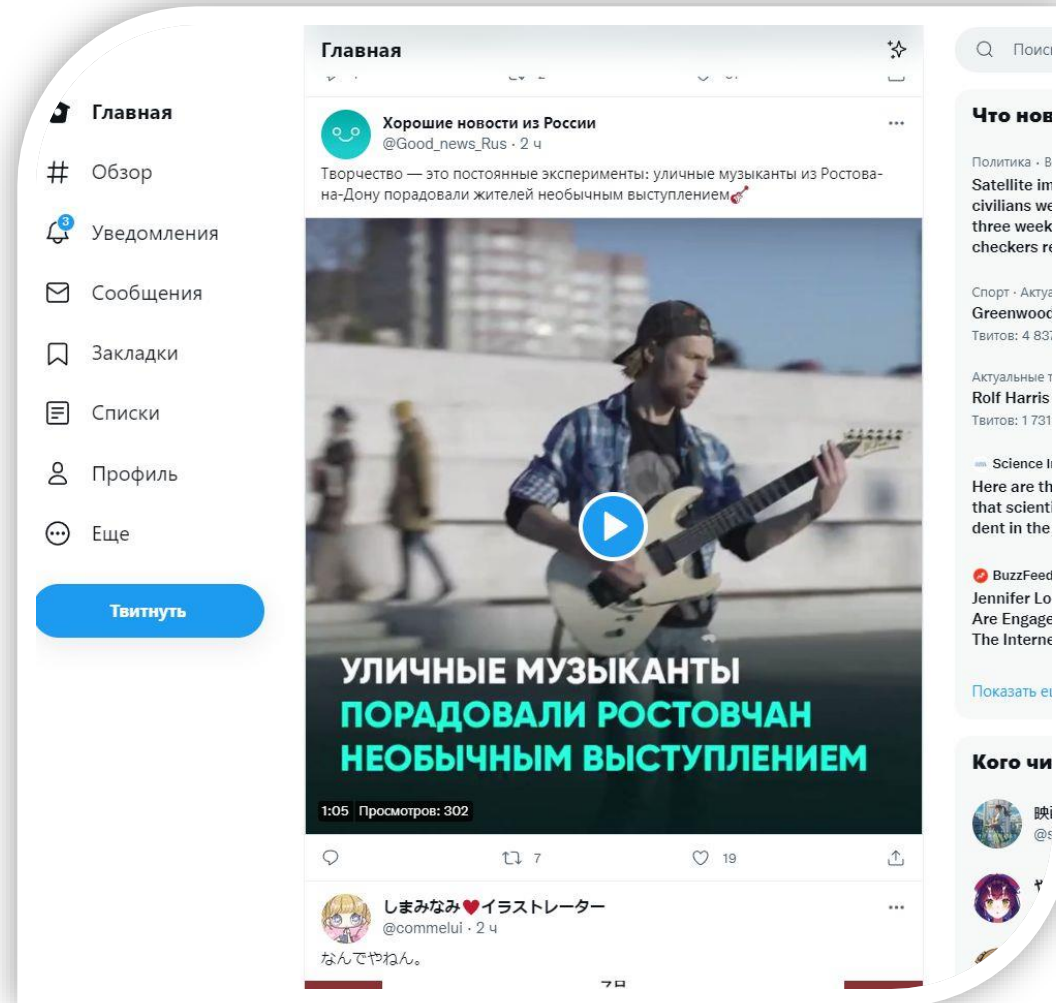
Пример технологического стека компании Twitter.
Администрирование баз данных.



ОПЫТ КОМПАНИИ TWITTER.

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

ВЕБ-ПРИЛОЖЕНИЕ (СОЦИАЛЬНАЯ СЕТЬ) TWITTER



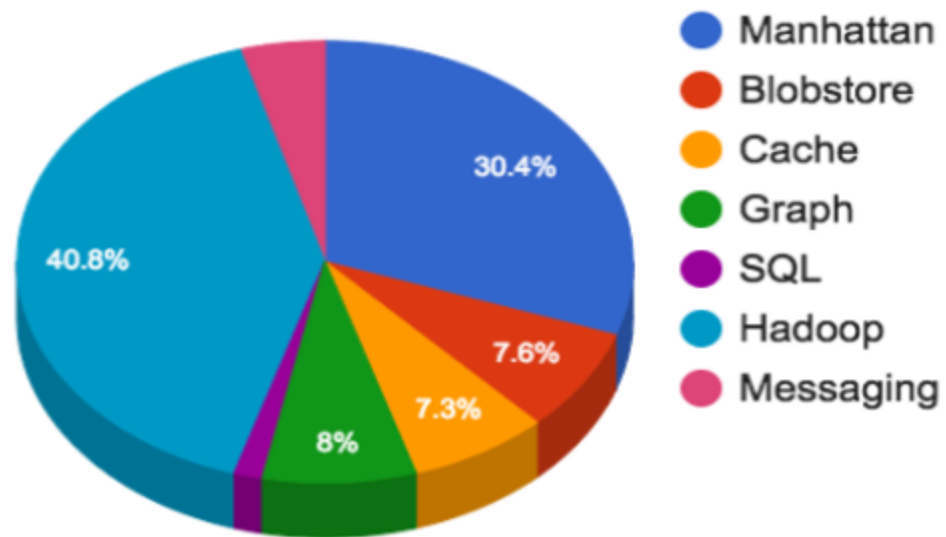
ПРАКТИЧЕСКИЕ ЗАДАЧИ, СТОЯВШИЕ ПЕРЕД СПЕЦИАЛИСТАМИ TWITTER



- 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 кэшировать 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

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



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.

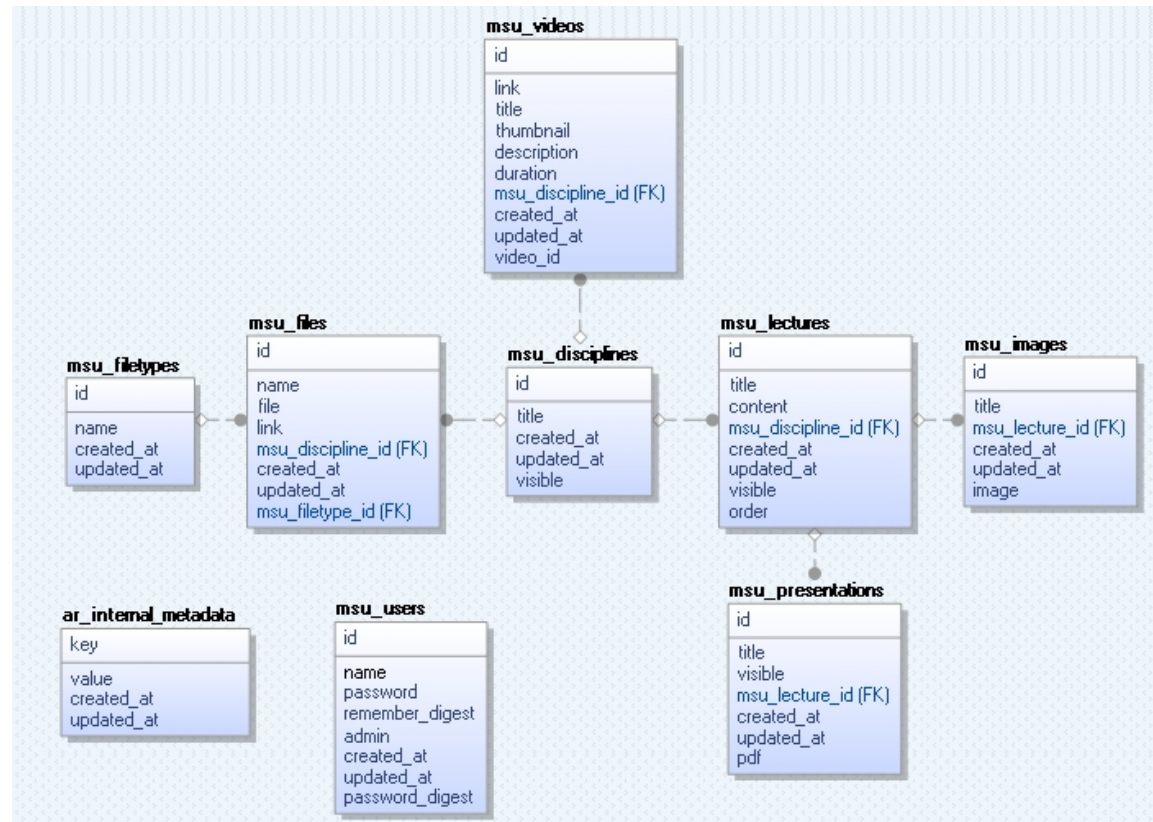
Оптимизация работы компонентов системы хранения данных.



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



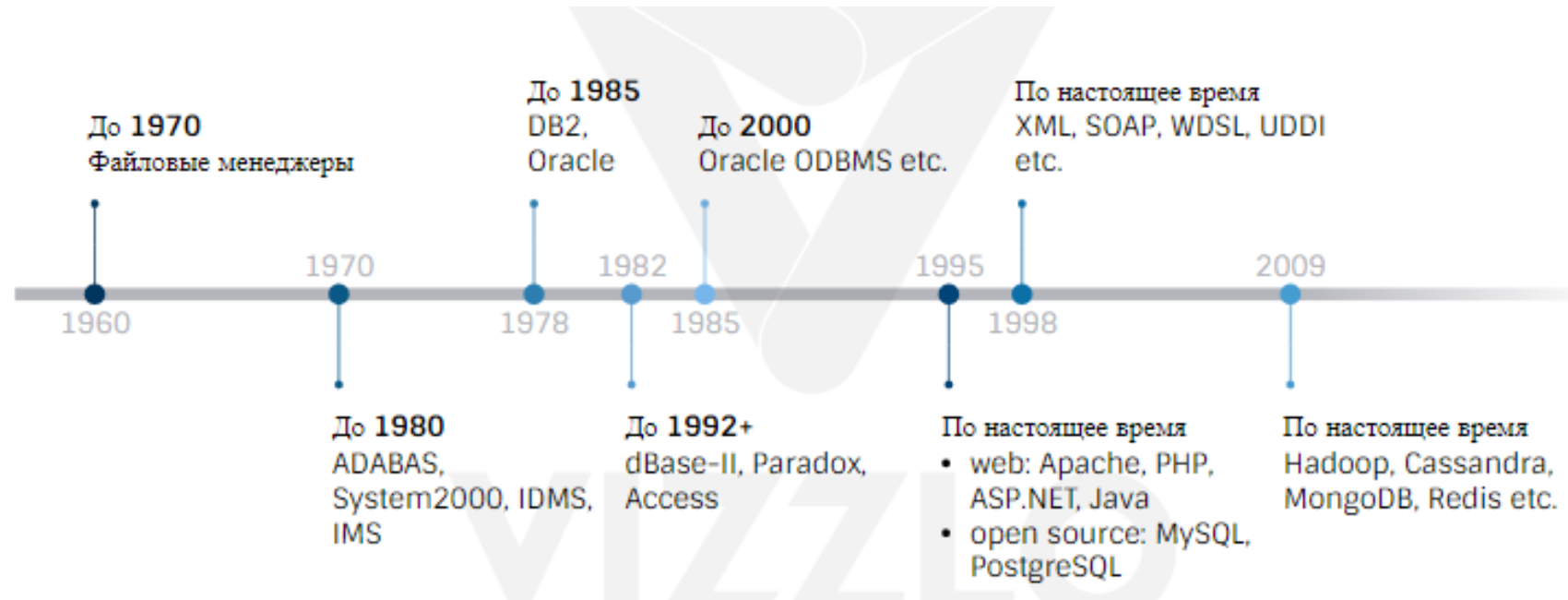
Типы и образцы баз данных

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!