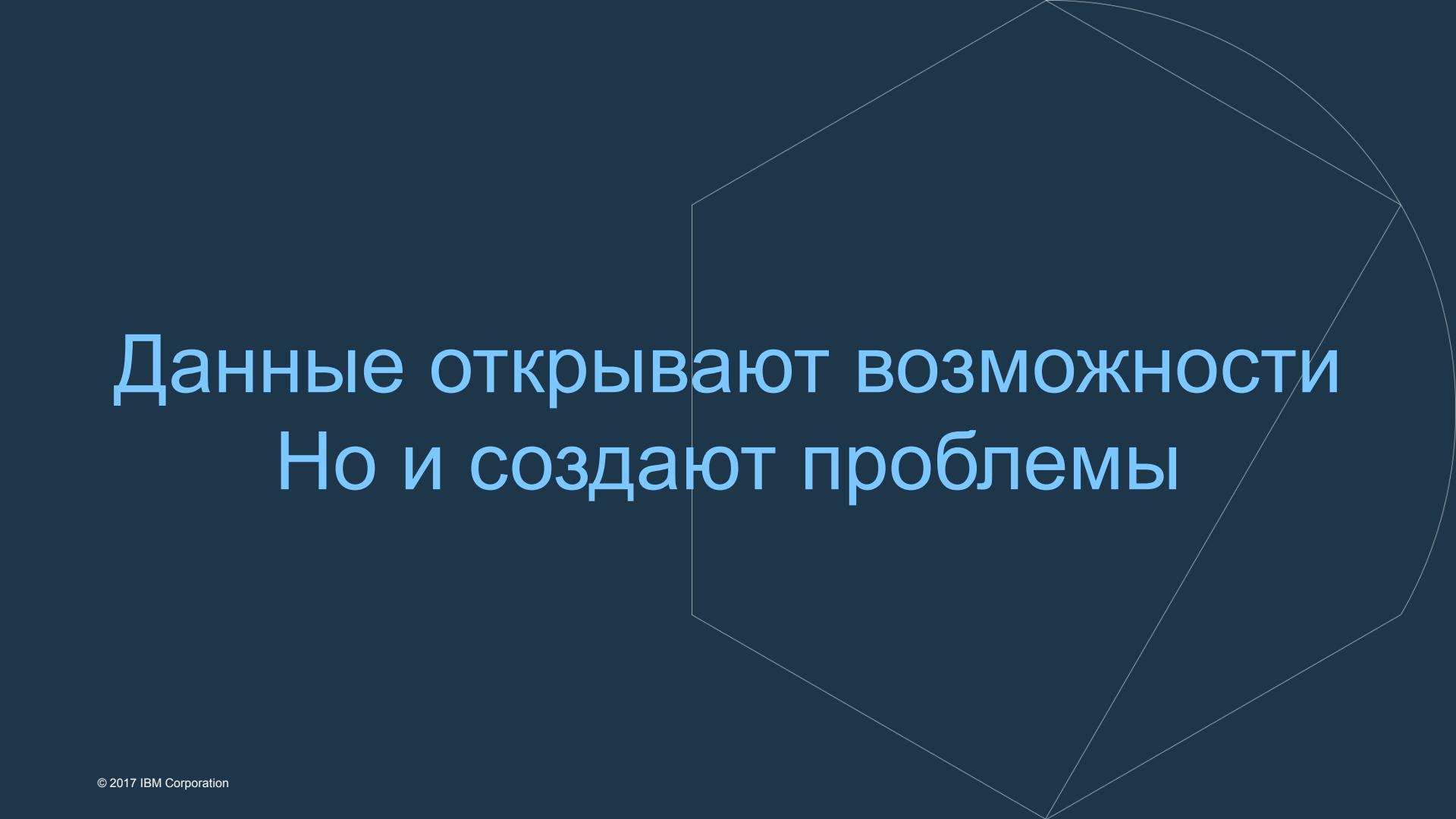


Объектное хранение

Новый стандарт для видеоконтента и Big Data

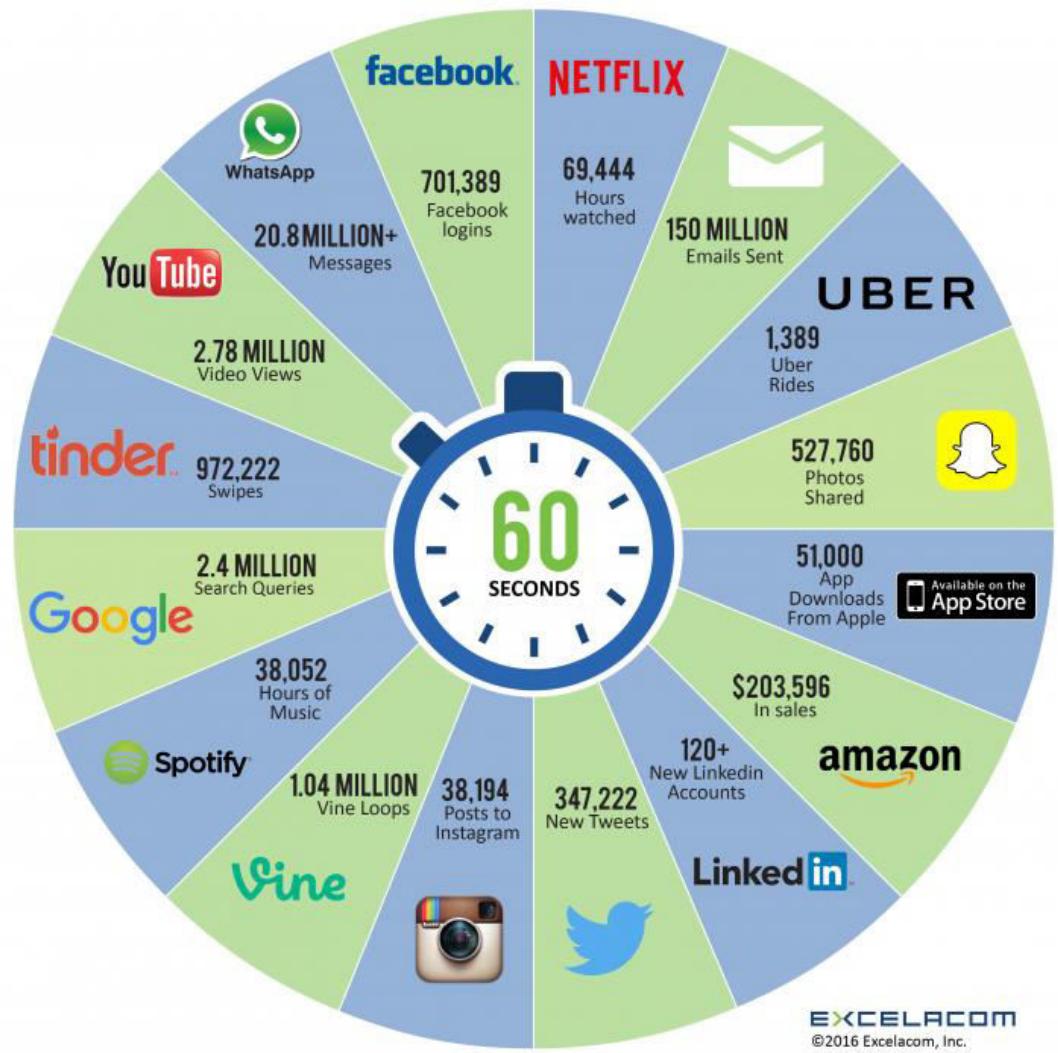
Сергей Сергиенко

Технический эксперт по системам хранения данных IBM



Данные открывают возможности
Но и создают проблемы

Что происходит в
интернете всего лишь
за 60 секунд?



Развитие облачных и мобильных технологий, Интернета вещей, социальных сетей и аналитики ведет к стремительному росту объемов данных

Постоянное появление новых, большей частью неструктурированных данных приводит к возникновению и расширению экосистемы данных в каждой организации

75 млрд
устройств, подключенных
к Интернету, к 2020 г.²

90 %
всех данных было создано
в течение последних 2 лет¹

80%
неструктурированных
данных³

Прогнозируемый объем данных, эксабайты

Источники

1. Science Daily, "Большие данные - во благо или во вред: 90 % всех данных в мире было создано за последние два года", 2013 г.
2. Business Insider, Morgan Stanley: 75 млрд устройств будет подключено к Интернету вещей к 2020 г., 2013 г.
3. "Откройте цифровую вселенную возможностей: разнообразные данные и все возрастающая ценность Интернета вещей", Исследование и анализ IDC, проведенные по заказу EMC Digital Universe, апрель 2014 г.



Неструктурированные данные

Имеют ряд отличий от привычных корпоративных данных

Структурированные данные

- Важно расположение и «структура» данных – их форма (базы данных, списки, таблицы)
- Оптимизированы по объёму
- Легко сортируются, обрабатываются и редактируются
- Требуют высокую скорость доступа

Неструктурированные данные

- Важно содержание данных (люди на фото, пейзаж на видео, информация генов, показатели датчиков)
- Неоптимизированные данные, имеющие много «лишней» информации
- Не требуется высокая скорость доступа, но требуется высокая надежность и продолжительность хранения («хранить всё и, по возможности, вечно»)

Файловые хранилища

не отвечают потребностям «больших данных»

Свойства файловых хранилищ

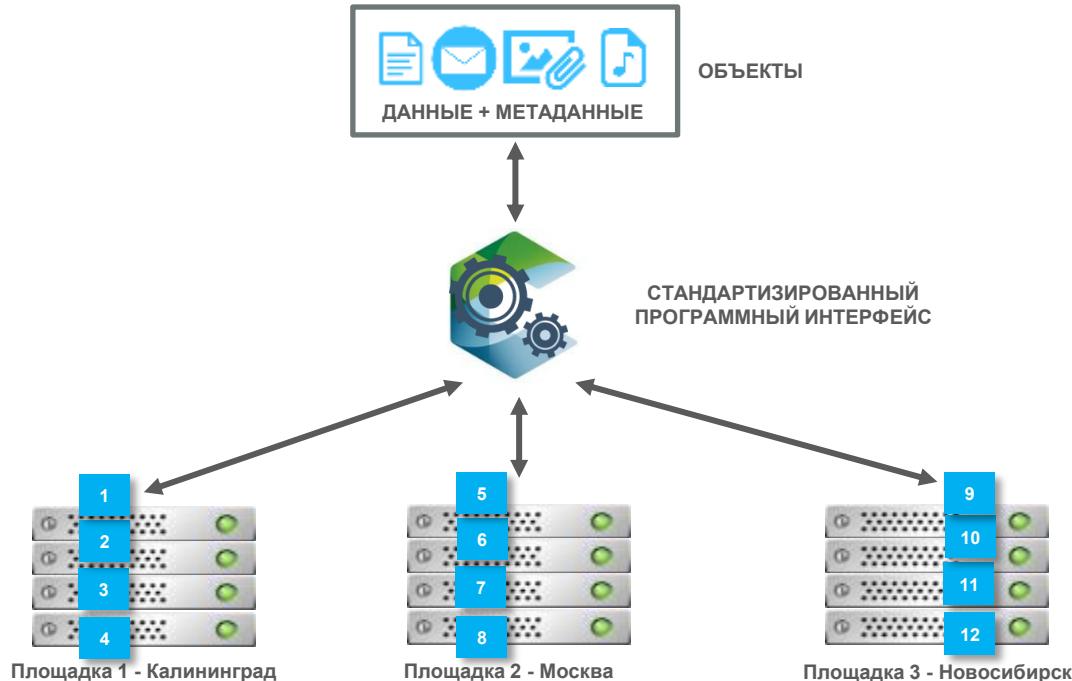
- Структурированное иерархическое дерево папок и файлов
- Атрибуты файлов: имя, дата создания, владелец, права пользователей
- Ограничение на максимальный размер файловой системы и размера файла, «локальность» файловой системы

Несоответствие потребностям

- У неструктурированных данных нет потребности в иерархии, т.к. первично их содержание
- Для описания неструктурированных данных этих атрибутов недостаточно
- Отдельные объекты неструктурированных данных могут занимать сотни гигабайт, а в совокупности – петабайты ёмкости

Требуется новая архитектура хранения данных

Концепция объектной архитектуры хранения данных

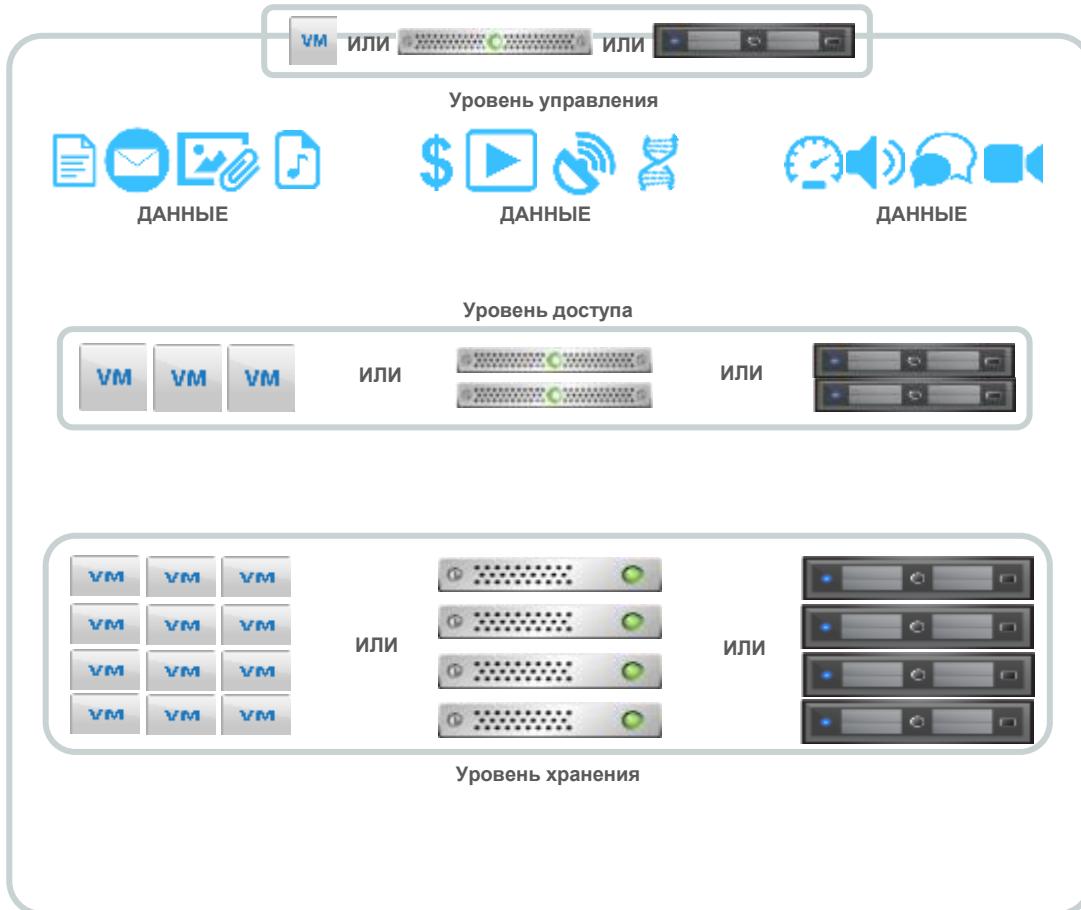


Устройство

КОНТЕНТ +
ОПИСАНИЕ, ТЕГИ

- AMAZON S3
- OPEN STACK SWIFT
- SIMPLE OBJECT
OVER HTTP

ВИРТУАЛЬНЫЕ ИЛИ
ФИЗИЧЕСКИЕ
ДИСКОВЫЕ РЕСУРСЫ

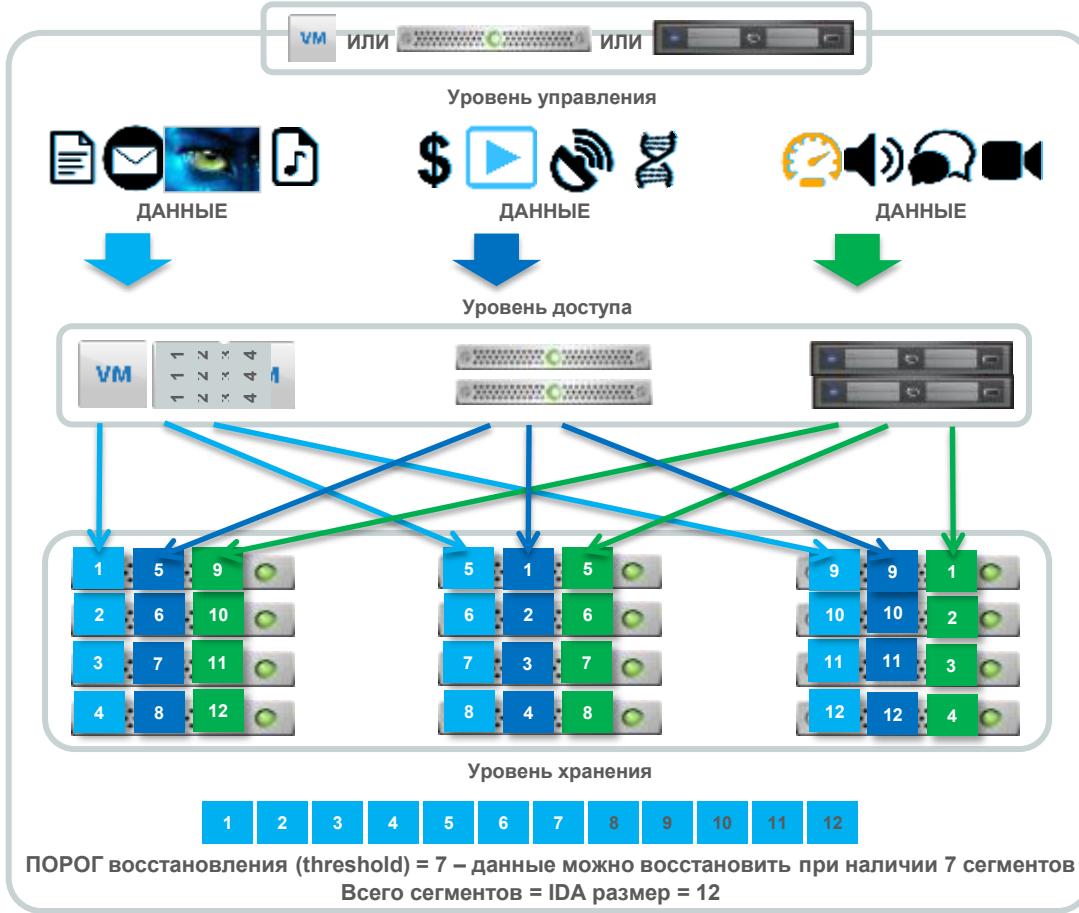


Уровень управления
Модуль управления для
конфигурирования и
администрирования

Уровень доступа
предоставляет доступ
для приложений к
системе хранения и
данным

Уровень хранения
непосредственное
хранение данных

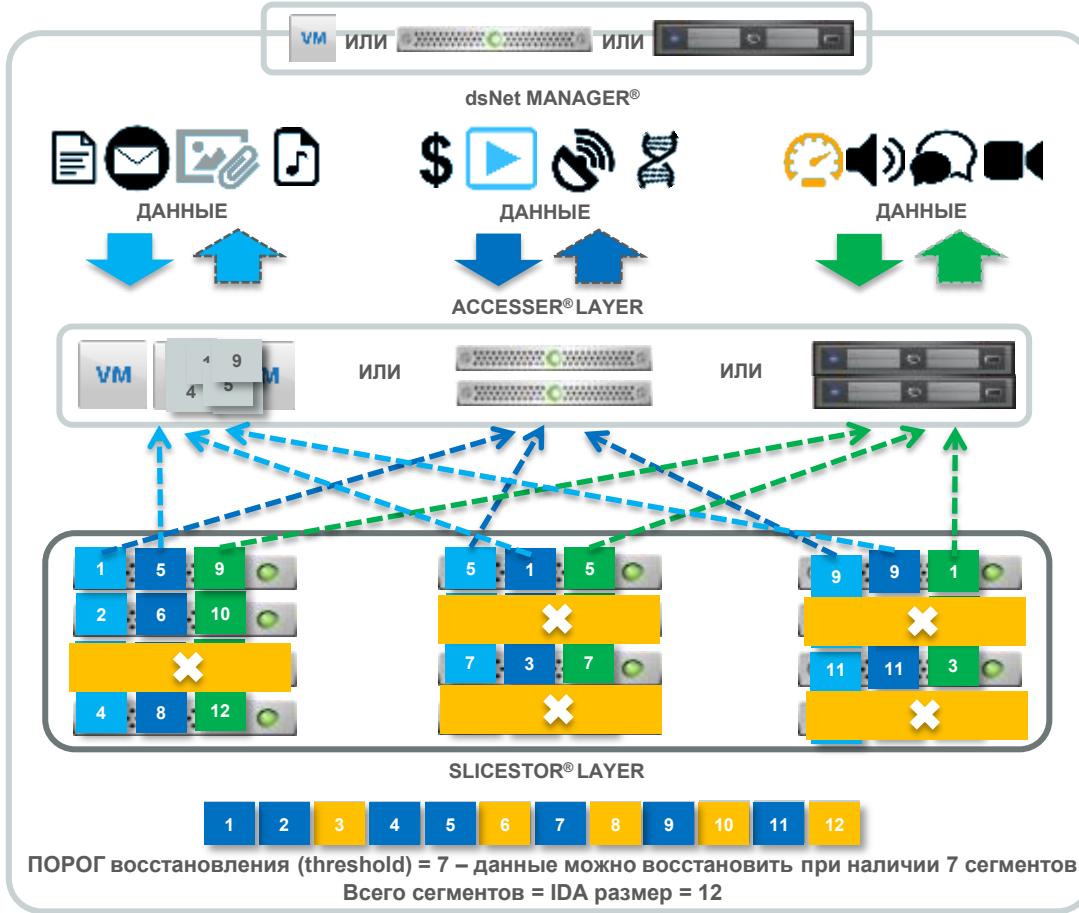
Распределение и запись данных



Данные шифруются и делятся на сегменты

Сегменты распределяются между разными дисками, узлами хранения, а также географически между разными сайтами

Чтение и восстановление данных

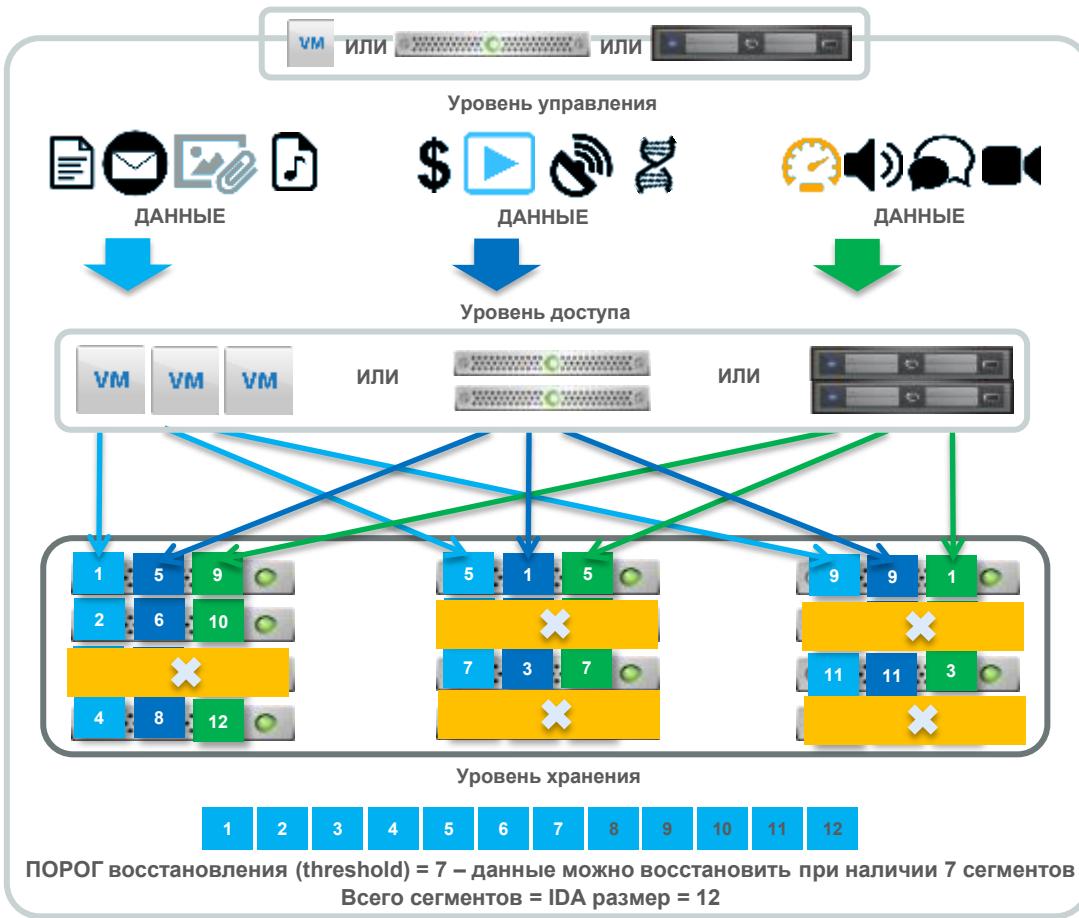


Порог восстановления устанавливается администратором в зависимости от требований к надежности системы

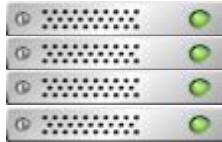
(теоретически можно установить надежность с «бесконечным» количеством «девяток»)

Данные восстанавливаются и расшифровываются независимо от того, какие узлы или сайты недоступны

Узел управления

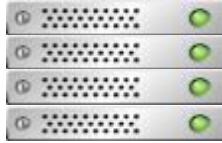


Технология объектного хранения для больших данных



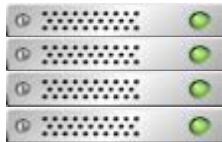
Расширяемость

Объем, скорость получения и уровень защиты контента напрямую зависит от количества узлов хранения



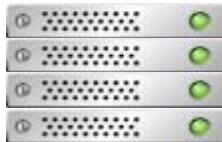
Доступность

Контент всегда доступен пользователям по стандартизованным протоколам



Содержание

Подробное описание объекта позволяет быстрее находить необходимый контент



Защищенность

Распределение ресурсов хранения между площадками и использование дисперсионного алгоритма защищает от потери данных и несанкционированного доступа

Объектное хранение как новый стандарт для видеоконтента и Big Data

Сергей Сергиенко

Технический эксперт по системам хранения данных IBM

Sergey.Sergienko@ru.ibm.com