1 слайд

Всем добрый день, меня зовут Ащеулов Михаил, и в качестве диплома я выбрал тему «Разработка быстроразвёртываемого объектного хранилища»

С каждым годом, количество цифровых данных растёт. Люди делают всё больше фотографий, видео, документов, и т.д.

2 слайд

В 2020 году во всём мире было создано и скопировано 59 зетабайта или 59 триллионов гигабайтов данных. 44 из них являются неструктурированными.

Только представьте уже сейчас 75% всей цифровой информации – это неструктурированные данные.

Как и любая компьютерная парадигма, системы хранения должны эволюционировать, чтобы обрабатывать такое количество неструктурированных данных. Однако для начала я хотел бы объяснить, что такое неструктурированные данны е.

Данные, которые невозможно организовать для хранения в реляционной базе данных, обычно называют неструктурированными данными.

Текстовые документы, электронные письма и презентации являются текстовыми неструктурированными данными.

Примеры не текстовых неструктурированных данных включают видео, изображения и аудиофайлы.

3 слайд

У нас уже есть механизм хранения, который люди используют с появления современных компьютеров, файловая система. Но когда мы говорим об неструктурируемых данных, важно понимать, что система используемая для хранения данных должны очень хорошо масштабироваться. Однако масштабировать файловую систему это сложная задача. Вам требуется не только управлять ненужными метаданными и иерархией, которые навязывают файловые системы, но также необходимо учитывать такие аспекты обслуживания, как управление резервными копиями.

Недостаточно просто собрать неструктурированные данные. Вам также необходимо применить некоторый уровень организации, чтобы разобраться в данных. Такие методы, как анализ текста, автоматическая категоризация и автоматическая пометка, имеют решающее значение для получения бизнес-смысла для всех неструктурированных данных, которые вы собираете. Этого сложно добиться с файловыми системами, потому что они имеют фиксированные слои (или расположение).

Для решения этой проблемы появились объектные хранилища.

4 слайд

Так что же такое объектное хранилище ?

В отличии от файлов в файловой системе, объекты хранятся в плоской структуре. Есть просто пул объектов: никаких папок, каталогов или иерархий. Вы просто запрашиваете данный объект, предоставляя его идентификатор объекта. Объектны могут быть локальными или находиться на облачном сервере за тысячи километров, но, поскольку они находятся в плоском адресном пространстве, они извлекаются точно так же.

Очень важно, чтобы файлы были доступны через HTTP, чтобы обеспечить лёгкий доступ к файлу. Затем его можно подвергнуть анализу или другим методам. Объектное хранилище справляется с этим хорошо. Почти все объектные хранилища имеют REST API, которе помогает вам получить доступ к файлам через HTTP (S). API полезны не только при доступе к данным, они помогают вам авторизироваться, получать свойства файла и управлять разрешениями, в общем всё, что вам нужно будет делать вручную в файловой системе.

5 слайд

Важным аспектом является обработка метаданных. Объектное хранилище обеспечивает большую гибкость, поскольку метаданные объекта изначально не определены Метаданные не ограничиваются тем, что система хранения считает важным. Вы можете добавить любой тип или количество метаданных, например, вы можете назначить тип приложения, с которым связан объект; важность приложения; уровень защиты данных, который вы хотите назначить объекту. И так далее, возможности безграничны.

6 слайд

Подытожим. Объектные хранилища имеют следующие важные особенности:

* Данные хранятся как отдельные объекты
* Данные не помещаются в иерархию каталогов, а располагаются в плоском адресном пространстве
* Приложения определяют отдельные объекты данных по их уникальным адресам
* Каждый объект может содержать метаданные безграничные в плане кастомизации

7 слайд

Объектное хранилище обладает такими свойствами как:

**Масштабируемость**

Объектное хранилище может содержать практически любое количество данных без необходимости в разбиении набора данных на разделы.

**Эффективность**

Отсутствие иерархии означает отсутствие узких мест, возникающих вследствие использования сложных систем каталогов.

**Доступность**

Объектные системы хранения имеют механизмы для сохранения целостности данных, обеспечивают репликацию данных, последовательные обновления и отсутствие простоев.

8 слайд

На рынке есть несколько облачных объектных хранилищ от крупных IT компаний

…

Все они предоставляют услуги по использованию места на их серверах за определённую плату. Плата взимается за каждый запрос, а так же за само хранение данных. Однако не смотря на обещания крупных IT компаний в безопасности на 99,9999999%, по данным InfoWatch за прошлый год в сеть утекло более 14 миллиардов конфиденциальных записей.

И многие люди, боясь за очередную утечку хотят найти решение, которое позволит им развернуть объектное хранилище на базе собственных серверов.

9 слайд

В таком случае есть решение и называется оно MinIO.

MinIO - это стек серверного программного хранилища, совместимый с Amazon S3, он может обрабатывать неструктурированные данные, такие как фотографии, видео, файлы журналов, резервные копии и образы контейнеров, с максимальным поддерживаемым размером одного объекта 5 ТБ.

10 слайд

И мы плавно подходим к актуальности проблемы или почему я вообще выбрал такую тему для дипломной работы.

В современном мире каждый человек где-то хранит персональные данные, будь-то в облаке google или на флешке.

Однако довольно часто нужно передать какие-то данные своему другу, приятелю или коллеге.

И появляется два возможных решения этой задачи:

1. Самый безопасный, передать данные лично на флешке без посторонних глаз в тёмном переулке.
2. Самый простой, разместить файлы в облачном хранилище, предлагаемом IT гигантами

А что если я вам скажу, что есть третье решение ?

Локальное, быстроразвёртываемое объектное хранилище, которое запускается на вашем компьютере и обеспечивает удобство облачных хранилищ и безопасность флешки.

11 слайд

Чтобы было комфортно пользоваться системой, она должна быстро разворачиваться: не более 2-20 секунд в зависимости от конфигурации.

Чтобы конкурировать с MinIO объектное хранилище должно поддерживать работу с объектами размером более 5 терабайт

Чтобы любой человек мог воспользоваться системой, она должна быть максимально простой в запуске и не требовать 5 лет опыта работы с распределёнными системами хранения данных.

Однако чтобы удовлетворить потребности энтузиастов и продвинутых пользователей, объектное хранилище должно иметь все функции, присущие такому типу хранилищ.

12 слайд

УМЛ диаграмма

13 слайд

В дипломном проекте я планирую использовать такие технологии как …

Так же для тестирования будет использоваться Scala вкупе с фреймворком Gatling.