

Система управления неструктурированными данными в объектных хранилищах

Заглубоцкий А.В. (ИТМО)

Научный руководитель - кандидат технических наук, доцент Капитонов А.А.
(ИТМО)

Введение. К 2026 году объем неструктурированных данных в локальных, периферийных и публичных облачных хранилищах у крупных организаций может вырасти в 3 раза, по сравнению с 2023 годом [1]. К неструктурированным данным относятся текстовые документы, электронные письма, изображения, аудиофайлы, видеофайлы. Такие данные непредсказуемы, поэтому без контроля над их ростом организации будут вынуждены увеличивать бюджет на емкости хранилищ.

Основная часть. Одним из подходов хранения неструктурированных данных, который применяется для хранения медиафайлов, резервных копий, а также статических ресурсов, является объектное хранилище [2]. Доступ к объектам в объектном хранилище осуществляется через API, которое запрашивает метаданные объекта, чтобы получить его из любого места на любом устройстве [2]. Используя данный подход можно получить быстрый поиск и извлечение данных, а также возможность неограниченного масштабирования.

Реализация данного подхода может занять продолжительное время, поэтому организации чаще всего используют сторонние решения, которые могут развернуть у себя на серверах, либо арендовать у облачных провайдеров. При этом возникает проблема, что организации становятся зависимыми от провайдера, поэтому вынуждены мириться с недостатками предоставляемых услуг, так как процесс миграции на другого провайдера может быть сложным и долгим.

Основная идея данной работы заключается в реализации системы, которая стандартизирует взаимодействие с объектными хранилищами, позволяя легко интегрировать новых провайдеров. Помимо этого необходимо обеспечить восстановление данных между разными хранилищами, а также контролировать доступ к ним [3].

Выводы. По результатам работы предлагается система управления данными в объектных хранилищах.

Список использованных источников

1. Gartner, Hype Cycle for Storage and Data Protection Technologies, 2023 [Электронный ресурс]. - Режим доступа: <https://www.gartner.com/en/documents/4527399>.
2. IBM, What is object storage? [Электронный ресурс] - Режим доступа: <https://www.ibm.com/topics/object-storage>
3. Клеппман М. Высоконагруженные приложения. Программирование, масштабирование, поддержка. - СПб.: Питер, 2018 - 640 с.