

ТЕХНОЛОГИЯ AZURE BLOB STORAGE ДЛЯ СОЗДАНИЯ ХРАНИЛИЩ ДАННЫХ НАУЧНЫХ ИССЛЕДОВАНИЙ

Глоба Л.С., Омельченко А.И.

Институт телекоммуникационных систем КПИ им. Игоря Сикорского

E-mail: aomelchenko@its.kpi.ua

AZURE BLOB STORAGE TECHNOLOGY FOR CREATING THE STORAGE OF RESEARCH DATA

The work is devoted to improving the structure and speed of access to scientific research data, the reliability of their storage, by creating a data storage based on Azure BLOB Storage technology from Microsoft. A detailed overview of the technology for develop the "cloud" storage from Microsoft. The main technical features of BLOB storage technology are considered.

В настоящее время ряд научных организаций получают уникальные данные, для которых необходимо гарантировать достоверность, сохранность, а также предоставлять к ним доступ в соответствии с утвержденным регламентом.

Исследования проводятся в разнообразных отраслях науки, таких как: зоология, геология, океанология и пр. Данные исследований описывают состояние научных исследований, имеют разнообразный формат и представление: результаты измерений, фото материал, видео данные, аудио файлы.

В связи с уникальностью данных возникает необходимость обеспечить их структурированность, надежность хранения и быстрый доступ к ним широкому кругу ученых всем заинтересованным научным организациям [1].

Для хранения данных научных исследований возможно использовать разные схемы хранения: наращивать ресурсы собственного Дата Центра, использовать ресурсы «облачных» хранилищ данных, а также использовать ресурсы удаленного Дата Центра. В статье проведен анализ возможных характеристик доступа к информации для «облачных» хранилищ данных ведущих компаний, их предоставляющих.

Критерием выбора способа хранения являются требования к условиям хранения информации по научным–исследованиям: скорость записи данных, скорость считывания данных, скорость удаления данных, время отклика при обращении к информации, что особенно критично для «облачных» сервисов.

С целью определения возможности использования «облачных» ресурсов были проанализированы результаты тестирования ресурсов хранения, предоставляемых компаниями Microsoft, Amazon, Google [2].

Первым тестом выполнялась проверка на скорость записи данных. «Облачное» хранилище Microsoft показало лучшие результаты в тестах на

скорость записи данных (любого типа) в 14 из 23 отдельных тестов. На втором и третьем месте с большим отставанием: Amazon и Google.

Результаты тестирования при чтении данных (как в целом, так и с файлами размера > 1МБ) из «облачного» хранилища так же показывают преимущество по скорости для платформы Microsoft, а с достаточно большим отставанием на втором месте идет Amazon, за которым следуют другие вендоры с примерно одинаковыми результатами. Результаты тестирования чтения файлов размером более одного мегабайта приблизительно сравниваются, с небольшим отставанием от Microsoft следуют Amazon и Google.

При проведении тестирования по скорости удаления файлов также имеется значительное преимущество для «облачного» хранилища от Microsoft. На втором месте с большим отставанием Amazon, на третьем - Google.

Тестирование по времени отклика при обращении к информации, хранящейся в «облачном» хранилище, а также реакции сервисов было измерено из расчета времени ответа на запросы операций чтения/записи/удаления за период в 30 дней. Сервис Microsoft продемонстрировал лучшие результаты по среднему значению (0.48 секунды). С небольшим отставанием в результатах на втором месте идет Amazon, далее с большим отставанием располагаются другие облачные вендоры.

Из проведенных тестов видно, что Microsoft занимает лидирующие позиции в большинстве проведенных тестов, поэтому в дальнейшем рассматривается технология организации «облачного» хранилища от компании Microsoft.

Рассмотрим службу хранения, на базе облачных технологий от Microsoft Azure — хранилище BLOB-объектов Azure.

Хранилище BLOB-объектов Azure — это служба хранения большого количества неструктурированных данных объектов в облаке, таких как текстовые или двоичные данные (например, документы, файлы мультимедиа или установщики программных компонент), к которым можно получить доступ практически из любой точки мира по протоколу HTTP или HTTPS.

Хранилища BLOB-объектов или хранилища объектов можно использовать для предоставления данных в открытом доступе или для конфиденциального хранения данных, генерируемых прикладным программным обеспечением.

Наиболее частые способы использования хранилища BLOB-объектов:

- Обслуживание изображений или документов непосредственно в браузере;
- Хранение файлов для распределенного доступа;
- Поточковая передача видео и аудио;
- Хранение резервных копий и восстановление данных, аварийное восстановление и архивация;
- Хранение данных для анализа локальной службой или службой, размещенной в Azure.

Исходя из способов использования хранилища BLOB-объектов для хранения данных, функциональных возможностей, можно сделать вывод о соответствии данной технологии требованиям хранения данных научных исследований, выдвигаемых при создании центров хранения данных научных исследований:

Данные научных исследований необходимо разграничить по определенным правилам доступа (направлениям исследований), например, биологические данные исследований необходимо предоставить, только для специалистов биологов. Технология «облачного» хранилища Azure позволяет реализовать это требование с помощью средств разграничения доступа.

По умолчанию данные больших двоичных объектов в учетной записи хранения доступны только владельцу учетной записи хранения, также для проверки подлинности запросов к хранилищу BLOB-объектов используют ключ доступа к учетной записи. Существует два варианта предоставления другим пользователям доступа к данным больших двоичных объектов:

- **Анонимный доступ:** предоставить общий анонимный доступ к контейнеру или его большим двоичным объектам.

- **Подписанные URL-адреса:** предоставить клиентам подписанный URL-адрес (SAS), который обеспечивает делегированный доступ к ресурсу в учетной записи хранения пользователя. Для этого доступа существует возможность указать разрешения и интервал времени доступа.

Служба позволяет настраивать и давать доступ или его ограничивать к любому объекту в хранилище.

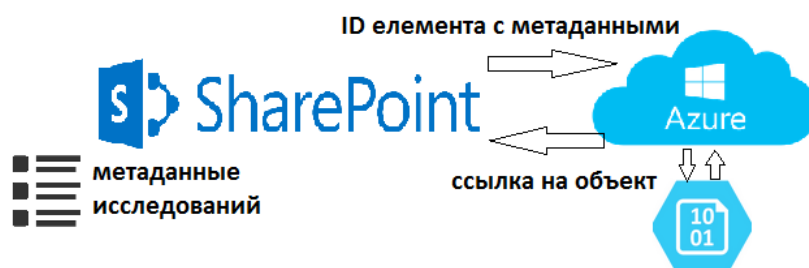


Рис 1. Архитектура взаимодействия SharePoint с Azure Blob Storage.

Управление метаданными научных исследований реализовано в рамках портала НАНЦ на базе SharePoint 2013[3]. Элементы хранятся в списках, каждый элемент имеет уникальный идентификатор. Такое построение хранилища, позволит изолировать данные исследований от их метаданных (Рис. 1). В такой реализации в составе метаданных будут храниться ссылки на данные исследований, представленные в виде объектов НЦАД, что позволит реализовать простое управление объектами, их копирование и восстановление при необходимости [4].

Исходя из рассмотренных функциональных компонент и программных средств можно сделать вывод, что служба хранения Blob объектов позволяет решить поставленные задачи хранения неструктурированного значительного

объема данных и обеспечить безопасность хранения, которая необходима для уникальных исследований, проводимых, например, НАНЦ в Антарктиде.

Исходя из вышесказанного делаем выводы:

- Служба хранилища Azure, основанная на способах использования хранилища BLOB-объектов для хранения данных и их функциональных возможностях, является универсальным и удобным способом хранения данных научных исследований;

- Задачу конфиденциальности, быстрого доступа и разграничения прав доступа на данные исследования между учеными и научными организациями, решает интеграция порталного решения и хранилища Azure с помощью стандартных средств SharePoint. Рассмотрены алгоритмы загрузки в «облачное» хранилище и получения из него больших двоичных объектов, которые учитывают конфиденциальность данных проводимых исследований [5];

- На основании теоретических исследований и представленных алгоритмов передачи и хранения данных реализуется прикладное программное обеспечение для автоматической загрузки и отображения результатов научных исследований.

Література

1. Глоба Л.С., Мороз И.В., Новогрудская Р.Л., Мочалкина К.С., Кузин И.О. Создание единого информационного пространства данных антарктических исследований, Украинский Антарктический Журнал, № 10-11, 2011, с. 343-351.
2. Глоба Л. С. Оптимізація функціонування порталу НАНЦ при роботі з гетерогенними даними / Л. С. Глоба, В.М. Наконечний, Р.Л. Новогрудська, О.О. Привар // VII Міжнародна Антарктична Конференція «Антарктичні дослідження: нові горизонти та пріоритети», Київ, 2015. – С.101.
3. Cloud storage: How do Amazon, Google and Microsoft stack up? Режим доступу: <http://www.techrepublic.com/article/cloud-storage-how-do-amazon-google-and-microsoft-stack-up/> Дата доступу: 30 серпня 2016.
4. Олексенко А.О., Глоба Л.С., Новогрудська Р.Л., Розробка процедури публікації метаданих в українському антарктичному центрі, Сьома міжнародна науково-технічна конференція "Проблеми телекомунікацій": Збірник тез. К.: НТУУ «КПІ», 2013, – С. 165-167.
5. Upload and Download files from Azure Storage [Електронний ресурс] – Електр. текстові дані. – Режим доступу: <https://blogs.msdn.microsoft.com/webapps/2014/09/24/upload-and-download-files-from-azure-storage/> Дата доступу: 30 серпня 2016.