

УДК 004.4

**МЕХАНИЗМ ЗАГРУЗКИ И ОТОБРАЖЕНИЯ ДОКУМЕНТОВ
АНТАРКТИЧЕСКИХ ИССЛЕДОВАНИЙ НА БАЗЕ BLOB STORAGE**А. И. Омельченко¹, Л. С. Глоба¹, И. В. Мороз²

¹Національний технічний університет України «Київський політехнічний інститут», Інститут телекомунікаційних систем, пр. Індустріальний, 2, Київ, Україна 03056,
e-mail: a_omelchenko@outlook.com, lgloba@its.kpi.ua

²Державна установа Національний антарктичний науковий центр МОН України;
e-mail: morozko21_63@mail.ru

Реферат. Работа посвящена повышению структурированности, надежности хранения, повышению скорости доступа к данным антарктических исследований Национального центра антарктических данных (НЦАД) за счет создания хранилища данных на базе технологии Azure BLOB Storage от компании Microsoft. Проведен анализ показателей, которые определяют эффективность доступа к «облачным ресурсам», определены основные критерии. Проведен детальный обзор технологии организации «облачного» хранилища от компании Microsoft. Рассмотрены основные технические особенности технологии хранения BLOB-объектов. Предложены алгоритмы загрузки данных Государственного учреждения Национальный антарктический научный центр (ГУ НАНЦ) в хранилище Azure, а также алгоритм их отображения

Механізм завантаження та відображення документів антарктичних досліджень на базі BLOB STORAGE

А. І. Омельченко, Л. С. Глоба, І. В. Мороз

Реферат. Работа посвящена підвищенню структурованості, надійності зберігання, підвищенню швидкості доступу до даних антарктичних досліджень Національного центру антарктичних даних (НЦАД) за рахунок створення сховища даних на базі технології Azure BLOB Storage від компанії Microsoft. Проведено аналіз показників, які визначають ефективність доступу до «хмарних ресурсів», визначено основні критерії. Проведено детальний огляд технології організації «хмарного» сховища від компанії Microsoft. Розглянуто основні технічні особливості технології зберігання BLOB-об'єктів. Запропоновано алгоритми завантаження даних Державної установи Національний антарктичний науковий центр (ДУ НАНЦ) до сховища Azure та алгоритм їх відображення.

The process of loading and displaying documents of Antarctic research based on BLOB STORAGE

A.I. Omelchenko, L.S. Globa, I. V. Moroz

Abstract. The work is dedicated to increasing the structuring, storage reliability and fast access to data of Antarctic Research of the National Antarctic Data Centre based on Azure BLOB Storage technology from Microsoft. The analysis of the indicators to determine the access efficiency to “cloud resources,” to find the main criteria. A detailed over-

view of the main features of “cloud” storage technologies from Microsoft is presented. The basic definition of storage technology BLOB-objects is determined. Data loading algorithms for State Institutions National Antarctic Scientific Center in Azure storage and their submission are proposed.

Key words: Antarctic bibliography, Antarctic research, Antarctica, Ukraine, Ukrainian Antarctic Research Expedition, Vernadsky Station, BLOB-storage, Azure, cloud, algorithms, storage.

1. Вступ

В соответствии с Резолюцией 4.1 XXII АТСМ 1998 года ученые Государственного учреждения Национальный антарктический научный центр (ГУ НАНЦ), как и ученые двадцати стран мира, проводят исследования в Антарктиде, 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. За период пребывания исследователей в Антарктиде формируется большое количество уникального, неструктурированного, и, в связи с условиями среды, не воспроизводимого материала, хранящегося в Национальном центре антарктических данных (НЦАД). [1] Данные исследований описывают текущее состояние и условия в Антарктиде, имеют разнообразный формат и представление:

- результаты измерений;
- фотоматериал;
- видеоданные;
- аудиофайлы.

В связи с уникальностью данных возникает необходимость обеспечить их структурированность, надежность хранения и быстрый доступ к ним широкому кругу ученых, научным учреждениям в Украине, а также заинтересованным зарубежным научным организациям [2].

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

2. Анализ показателей производительности при доступе к «облачным» ресурсам

Критерием выбора способа хранения являются требования к условиям хранения информации по антарктическим 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 [3].

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 находились Amazon и Google.

Результаты тестирования при чтении данных (как в целом, так и с файлами размера > 1МБ) из «облачного» хранилища также показывают преимущество по скорости для платформы Microsoft, 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 равны, с небольшим отставанием за 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.

3. 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;
- потоковая 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 и функциональных возможностей можно сделать вывод о соответствии данной технологии требованиям хранения данных антарктических исследований, выдвигаемых при создании ИЦАД.

3.1. Основные понятия службы BLOB-объектов

Объекты BLOB представляются в двоичном формате, который является универсальным. Объект данных исследований является унифицированным представлением данных, которые имеют различные форматы и структуру представления.

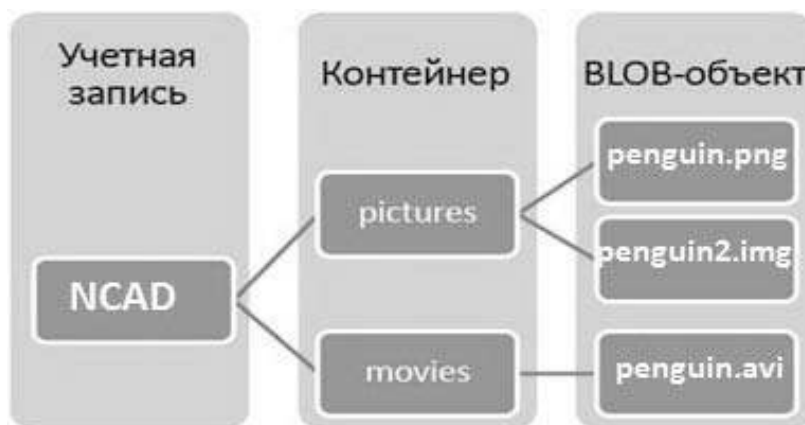


Рис 1. Компоненты службы BLOB-объектов

• **Учетная запись хранения.** Весь доступ к хранилищу Azure осуществляется с помощью учетной записи хранения. Это может быть учетная запись хранения общего назначения или учетная запись хранилища BLOB-объектов, специально предназначенная для хранения больших двоичных объектов и других объектов.

• **BLOB-объект.** Файл любого типа и размера. Служба хранилища Azure предлагает три типа больших двоичных объектов: блочные, страничные и добавочные.

— *Блочные BLOB-объекты* используют для хранения текстовых и двоичных файлов, таких как документы и файлы мультимедиа. Добавочные большие двоичные объекты схожи с блочными BLOB-объектами, так как состоят из блоков, однако они оптимизированы для операций добавления и поэтому полезны для сценариев ведения журнала. Один блочный BLOB-объект или добавочный большой двоичный объект может содержать до 50 000 блоков размером до 4 МБ каждый с общим размером немного более 195 ГБ (4 МБ X 50 000).

— *Страничные BLOB-объекты* могут иметь размер до 1 ТБ. Они более эффективны для частых операций чтения и записи. Виртуальные машины Azure используют страничные BLOB-объекты в качестве дисков данных и дисков операционной системы.

— *Добавочные BLOB-объекты* оптимизированы для операций добавления.

• **Контейнер.** В контейнере группируется набор BLOB-объектов. Все BLOB-объекты должны содержаться в контейнере. Учетная запись может содержать неограниченное число контейнеров. Контейнер может хранить неограниченное количество больших двоичных объектов. Все знаки в имени контейнера должны быть строчными.

Служба хранилища Azure по умолчанию защищает данные, ограничивая доступ к учетной записи пользователя, который владеет ключами доступа к учетной записи. При необходимости предоставить доступ к данным больших двоичных объектов в своей учетной записи хранения важно сделать это без ущерба для безопасности ключей доступа к учетной записи. Кроме того, существует возможность зашифровать данные больших двоичных объектов, чтобы обеспечить их безопасную отправку по сети в службу хранилища Azure.

3.2 Разграничение доступа для «облачного» хранилища Azure

Данные антарктических 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, а 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 к любому объекту в хранилище или его ограничивать.

Azure поддерживает шифрование больших двоичных объектов данных на стороне клиента и сервера.

— **клиентская библиотека** службы хранилища для .NET поддерживает шифрование данных в клиентском прикладном программном обеспечении перед их отправкой в службу хранилища Azure и расшифровку данных во время скачивания. Библиотека также поддерживает интеграцию с хранилищем ключей Azure для управления ключами учетной записи хранения.

— **на стороне сервера:** с помощью этой функции служба хранилища Azure автоматически шифрует данные перед сохранением в хранилище и расшифровывает их перед извлечением. Шифрование, расшифровка и управление ключами полностью прозрачны для пользователей.

Управление метаданными антарктических исследований реализовано в рамках портала ГУ НАНЦ на базе SharePoint 2013 [7]. Элементы хранятся в списках, каждый элемент имеет уникальный идентификатор. Такое построение хранилища позволит изолировать данные исследований от их метаданных (рис. 2). 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. [8]



Рис 2. Архитектура взаимодействия SharePoint с Azure Blob Storage

4. Алгоритмы доступа

4.1. Алгоритм отправки в «облачное» хранилище больших двоичных объектов [8]

Последовательность действий для отправки больших двоичных объектов представлена ниже на рис. 3:

1. Для отправки файла большого двоичного объекта в контейнер необходимо получить ссылку на контейнер.
2. Полученную в пункте 1 ссылку необходимо использовать для получения ссылки на большой двоичный объект.
3. Получив ссылку на большой двоичный объект, необходимо отправить в него любой поток данных.
4. Отправка потока данных создает большой двоичный объект, если его не существует, или заменяет его, если он существует.

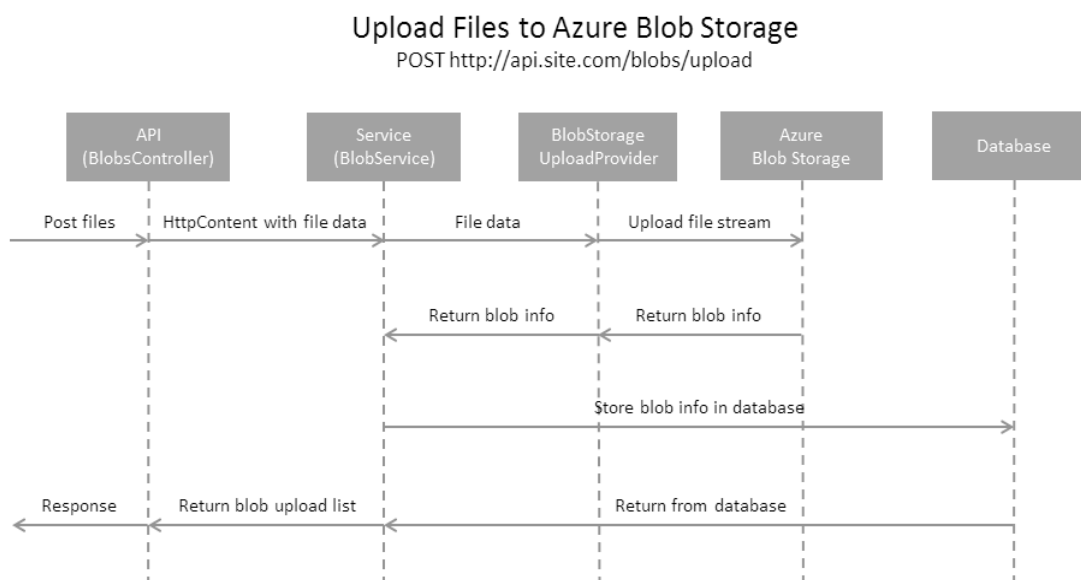


Рис. 3. Отправка BLOB-объекта в контейнер.

4.2. Алгоритм получения пользователем из «облачного» хранилища больших двоичных объектов [9].

Последовательность действий для скачивания больших двоичных объектов представлена ниже на рис 4.:

1. Для загрузки BLOB-объектов сначала нужно получить ссылку на BLOB-объект.
2. Далее необходимо вызвать метод для загрузки потока данных.
3. Метод для загрузки потока данных используется для переноса содержимого большого двоичного объекта в объект потока, который затем можно сохранить в локальном файле или отобразить на веб странице.

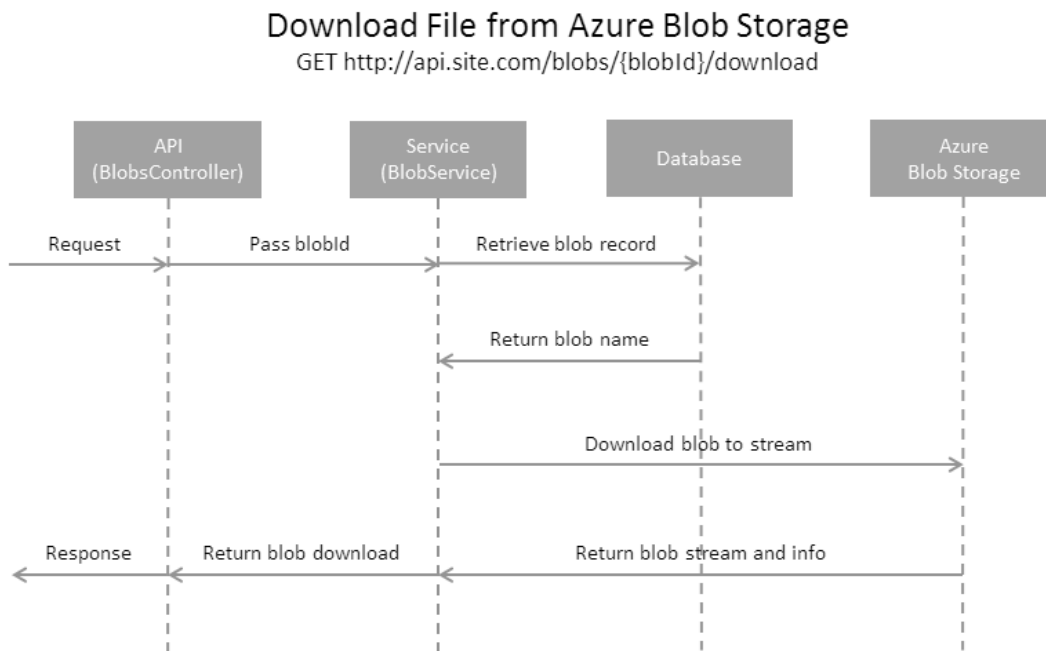


Рис. 4. Загрузка BLOB-объекта с контейнера.

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 ГУ НАНЦ в Антарктиде.

Выводы

1. Служба хранилища Azure, основанная на способах использования хранилища BLOB-объектов для хранения данных и их функциональных возможностях, является универсальным и удобным способом хранения данных исследований в Антарктиде.
2. Задачу конфиденциальности, быстрого доступа и разграничения прав доступа на данные исследования между учеными и научными учреждениями Украины, а также зарубежными научными организациями, решает интеграция портала ГУ НАНЦ и хранилища Azure с помощью стандартных средств SharePoint.
3. Рассмотрены алгоритмы загрузки в «облачное» хранилище и получения из него больших двоичных объектов, которые учитывают конфиденциальность данных исследований в Антарктиде.
4. На основании теоретических исследований и представленных алгоритмов передачи и хранения данных реализуется прикладное программное обеспечение для автоматической загрузки и отображения результатов исследований в Антарктиде.

Список літератури

1. **Глоба Л. С.**, Мороз І. В., Новогрудская Р.Л., Мочалкина К. С., Кузін І. О. Создание единого информационного пространства данных антарктических исследований, *Український Антарктичний 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. **Глоба Л. С.** Модель Интернет-портала Национального Центра Антарктических Данных / Л. С. Глоба, И. А. Кузін, К. С. Мочалкина, Р. Л. Новогрудская // Прикладные интеллектуальные системы, основанные на семантических сетях: Тр. междунар. Конференции OSTIS-2012, 16-18 февр. 2012 г., Минск, Белоруссия. – Минск, 2012. – С. 501–506.
5. **Новогрудська Р. Л.** Інформаційно-комунікаційне середовище співробітника НАНЦ / Р. Л. Новогрудська, Я. І. Федяй // X Міжнародна Науково-технічна Конференція «ПРОБЛЕМИ ТЕЛЕКОМУНІКАЦІЙ 2016», 9.04-22.04.2016 р. Київ, Україна. – Київ, 2016. – С. 375–377.
6. **Microsoft** about Share Point 2013 [Електронний ресурс] – Електр. текстові дані. – Режим доступу: <http://technet.microsoft.com/en-us/library/cc303422.aspx>. Дата доступу: 25 березня 2016.
7. **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.
8. **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.
9. **Upload** and Download files from Azure Storage [Електронний ресурс] – Електр. текстові дані. – Режим доступу: <http://www.intstrings.com/ramivemula/articles/file-upload-and-download-to-azure-blob-storage/> Дата доступу: 30 червня 2016.
10. **NASA** portal. Global change master directory – Режим доступу: <http://gcmd.nasa.gov/> – Дата звернення: 07.12.2016.