

<https://doi.org/10.5281/zenodo.7882256>

УДК 004

ИНФОРМАЦИОННАЯ СИСТЕМА ДЛЯ ДОСТУПА К СЕТЕВЫМ ФАЙЛОВЫМ ХРАНИЛИЩАМ РАЗЛИЧНОГО ТИПА

А.В. Алексеева,

студент 4 курса, напр. «Программная инженерия»

Б.С. Юдинцев,

к.т.н., ст.преп. кафедры ВМиК,

УУНиТ,

г. Уфа

Аннотация: В статье проводится сравнительный анализ базовых сетевых протоколов для доступа к файлам в корпоративных хранилищах данных. В статье также рассмотрены существующие IT-решения в области распределённого (облачного) хранения данных, показаны особенности интеграции этих информационных систем с сетевыми хранилищами различного типа. На основе сравнительного анализа и выявленных недостатков существующих решений рассматривается архитектура информационной системы, позволяющая взаимодействовать с файловыми хранилищами различного типа (FTP, SMB/CIFS, S3). Представлен подход, в рамках которого предлагается использование единого программного интерфейса (Application Programming Interface, API), реализующего модель REST (Representational State Transfer) для управления файлами данных в хранилищах, использующих различные протоколы сетевого доступа. Данный подход также предполагает интеграцию NoSQL базы данных (БД) MongoDB для хранения метаданных файлов и быстрого поиска необходимых данных в хранилищах. Показана актуальность реализации представленной информационной подсистемы для предприятий малого и среднего бизнеса.

Ключевые слова: сетевые хранилища, ftp, smb, s3, rest api, mongo

INFORMATION SYSTEM FOR ACCESSING NETWORK ATTACHED STORAGES OF DIFFERENT TYPES

A.V. Alekseeva,

4th year student, direction "Software engineering"

B.S. Yuditsev,

Candidate of Technical Sciences, Senior Lecturer Department of VMIK,
UUST,
Ufa

Annotation: The article provides a comparative analysis of the basic network protocols for accessing files in corporate network attached storages (NAS). The article also discusses existing IT solutions in the field of distributed (cloud) data storage systems, shows the features of the integration of these information systems with NAS of various types. Based on a comparative analysis and identified shortcomings of existing solutions, an information system architecture for interaction with file storages of various types (FTP, SMB/CIFS, S3) is considered. An approach is presented that proposes the use of a Application Programming Interface (API), which implements the Representational State Transfer (REST) model for managing data files in storages using various network access protocols. This approach also involves the integration of a NoSQL database (MongoDB) to store file metadata and use fast search operations for the necessary data in storages. The relevance of the implementation of the presented information subsystem for small and medium-sized businesses is shown.

Keywords: network attached storages, NAS, ftp, smb, s3, rest api, mongo

В настоящее время применение корпоративных сетевых хранилищ для хранения файлов данных, используемых во внутренних бизнес-процессах, фактически является стандартом для обеспечения безопасной и надёжной информационной инфраструктуры предприятия. В качестве хранимых данных (в зависимости от вида деятельности предприятия) могут быть, например, файлы проектной и рабочей документации (для проектных организаций), офисная

документация, файлы пользовательских заявлений и документов (для госорганизаций, занимающихся предоставлением электронных государственных услуг) и т.д. Таким образом, специализированные системы хранения предоставляют ряд критически важных функций: резервное копирование и восстановление данных, возможность гибкой настройки прав доступа к хранимым данным, а также различные средства интеграции в информационную сеть предприятия (например, с помощью протокола доступа к каталогам LDAP).

Особенности интеграции сетевых хранилищ в ИТ-инфраструктуру предприятия в первую очередь определяются сетевым протоколом, выбранным для доступа к хранимым данным. Рассмотрим далее наиболее популярные решения в рамках корпоративного сегмента.

FTP-хранилище – система хранения на основе File Transfer Protocol (протокол передачи файлов) является наиболее простым решением для создания системы хранения файлов, благодаря большому количеству свободных реализаций данной системы, а также упрощённой процедуре развёртывания, для которой достаточно обычной серверной машины с установленным дистрибутивом Linux (например, ProFTPD [1]). При этом система предоставляет все необходимые операции для управления файлами и каталогами (сохранение, удаление, изменение и т.д.), а также уровень авторизации пользователей [2]. Среди основных преимуществ данного типа хранилищ можно выделить:

1. Большое количество свободных реализаций клиентских приложений и библиотек для работы с FTP на всех популярных языках программирования, что упрощает доступ к данным через пользовательские устройства (мобильные и стационарные), а также упрощает интеграцию сервисных приложений, в том числе собственной (внутрикорпоративной) разработки.

2. Возможность интеграции системы авторизации сервера с корпоративной базой данных пользователей (например, SQL или LDAP) [1].

3. Возможность ограничения прав пользователей на выполнение конкретных типов операций, формирование «чёрных списков» [1].

Однако существует и ряд недостатков:

1. Протокол FTP имеет ряд известных уязвимостей (например, FTP-Bounce [3]), что не позволяет обеспечить безопасный доступ к данным в хранилище из внешней сети.

2. Проблема уязвимости протокола может быть решена с помощью использования защищённой реализации протокола – SFTP [1]. В таком случае, программные решения и библиотеки для стандартного FTP будут неприменимы для взаимодействия с хранилищем, что усложняет интеграцию корпоративных сервисных приложений и клиентских устройств.

3. В отличие от HTTP не содержит метаданных (MIME-тип) о передаваемых данных [2], что может приводить к проблемам с декодированием данных при чтении.

4. Протокол не предоставляет инструментов для централизованного управления несколькими хранилищами, что усложняет горизонтальное масштабирования системы хранения данных.

5. Нет встроенных механизмов поиска необходимых файлов.

Ввиду указанных недостатков FTP-хранилища крайне редко используются как основные системы хранения в современных корпоративных сетях. Более популярным решением для задач хранения файлов данных предприятия являются сетевые хранилища, на основе протокола Server Message Block (SMB/CIFS, далее SMB) и NFS.

В качестве основных преимуществ таких систем хранения можно выделить:

1. Встроенные механизмы интеграции хранилища в корпоративную сеть предприятия и на клиентских стационарных компьютерах (особенно, для машин на основе ОС Windows).

2. Гибкая настройка политик безопасности и прав доступа пользователей и групп к каталогам, а также инструменты централизованного управления политиками при наличии контроллера домена [4].

3. Наличие свободных реализаций протокола и приложений для работы с ним (Samba, FreeIPA) [4, 5], что позволяет построить надёжную и современную IT-инфраструктуру для организаций малого и среднего бизнеса.

Тем не менее, SMB хранилища также имеют и ряд недостатков:

1. Ранние версии протокола (SMB1 и SMB2) не имеют механизмов шифрования, уязвимы к перехвату трафика (атаки типа Adversary-in-The-Middle) и, соответственно, не могут быть использованы для организации безопасного удалённого доступа к хранилищу. Новая версия протокола (SMB3) решает основную часть проблем с безопасностью за счёт внедрения механизмов шифрования и новых методов авторизации пользователей, но усложняет интеграцию корпоративных сервисных приложений и клиентских устройств, а также требует обновления контроллеров домена и клиентских систем до актуальных версий для полноценной интеграции [6, 7].

2. Замедление скорости передачи данных при работе с большим количеством малоразмерных файлов [8].

3. Медленное выполнение поисковых запросов при большом количестве файлов и каталогов в хранилище. Как и в случае с передачей данных подобные запросы инициируют многократные вызовы протокола, что замедляет работу системы [8].

4. Несмотря на то, что SMB хранилища легко интегрируются в ОС клиентских компьютеров (интеграция сводится к простому «монтированию» каталогов или целых хранилищ в виде отдельных сетевых дисков в клиентской системе), при наличии нескольких хранилищ в сети поисковые запросы по файлам необходимо выполнять на каждом хранилище отдельно.

Ряд недостатков, присущих системам хранения на основе протокола SMB, были устранены в рамках разработанного компанией Amazon сервиса хранения данных S3 (Simple Storage Service). Основная идея данного сервиса заключается в хранении большого объема данных в исходном формате без формирования классической иерархической структуры каталогов. Каждый файл хранится в виде объекта с дополнительными атрибутами (метаданными). Объектам присваивается идентификационный номер, который используется для доступа к соответствующим данным [9]. Как известно, доступ к Amazon S3 осуществляется на основе подписочной модели монетизации [10], тем не менее, существуют свободно распространяемые решения, предоставляющие все основные

возможности S3 и позволяющие развернуть подобные системы хранения на базе собственной IT-инфраструктуры. Наиболее популярным среди таких решений является MiniO [11].

Среди основных конкурентных преимуществ S3-хранилищ можно отметить:

1. S3-хранилища относятся к классу облачных систем хранения (Cloud Storage), что предоставляет широкие возможности для масштабирования системы хранения и балансировки нагрузки [10, 11].

2. Наличие встроенных средств репликации данных и мониторинга [10, 11].

3. Наличие встроенных средств аналитики хранимых данных (для определённых типов контента) [10, 11].

4. S3 стандарт предоставляет классический REST API для управления данными в хранилище (используются стандартные методы протокола HTTP: GET, POST, PUT, DELETE), что значительно упрощает интеграцию с другими корпоративными сервисами [12].

5. Использование стандартного протокола HTTPS позволяет реализовать безопасный удалённый доступ к данным при использовании современных методов авторизации [13].

Также можно отметить и некоторые недостатки:

1. Развёртывание и интеграция собственного S3-хранилища требует серьёзное расширение IT-инфраструктуры предприятия. Необходима установка новых серверов с дисковыми массивами (JBOD/JBOF), при этом для подобного решения минимальные аппаратные требования определяют наличие 8-ми дисков серверного типа, серверные процессоры серии Xeon Gold, 128 GB RAM и ширину сетевого канала от 25 Гбит/с [14]. Также необходимы соответствующие компетенции со стороны обслуживающего персонала. Подобные вложения не всегда целесообразны задачам, которые решают системы хранения данных в рамках предприятий малого и среднего бизнеса.

2. Также известные решения на основе S3 модели хранения данных не предоставляют встроенных инструментов миграции данных со старых хранилищ (SMB или FTP).

3. В качестве технического недостатка S3-хранилищ можно отметить ограниченность инструментов поиска данных без

использования идентификатора. В данных системах возможно выполнение гибки поисковых запросов на основе SQL, но такие запросы могут выполняться только для данных в формате JSON и CSV [15].

Исходя из проведённого анализа основных протоколов для управления файлами данных и сетевых хранилищ на их основе, можно выделить следующие ключевые проблемы:

1. Отсутствие встроенных механизмов для выполнения гибкого поиска (по регулярным выражениям или полнотекстового) файлов данных по наименованию или атрибутам (метаданным).

2. Отсутствие единого интерфейса (API) для взаимодействия с хранилищами разного типа или с несколькими однотипными хранилищами (такими как FTP, SMB) и с возможностью безопасного удалённого доступа клиентов с различных клиентских устройств, в том числе мобильных.

3. Необходимость реализации собственных инструментов миграции данных между хранилищами разного типа при обновлении ИТ-инфраструктуры предприятия (например, при переходе с SMB на S3).

Решением указанных проблем может быть платформа от Microsoft – Azure Storage Service, которая предоставляет инструменты для интеграции и совместного использования хранилищ различного типа (рис. 1): FileStorage (SMB) и BlobStorage (объектное хранилище, аналог S3 от Amazon) [16].

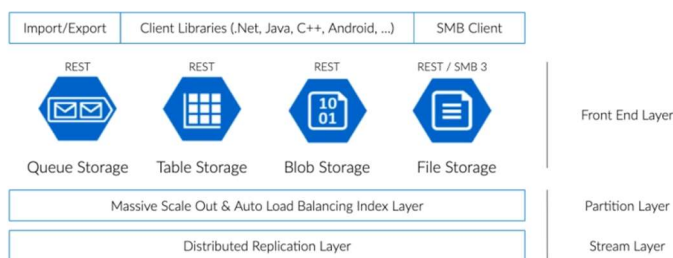


Рисунок 1 – Архитектура платформы Azure Storage Service [16]

Но здесь стоит отметить, что доступ к данной платформе осуществляется на основе подписочной модели монетизации, также

стоит учитывать, что инфраструктура данного сервиса будет находиться не в локальной сети предприятия, а на внешних ресурсах. Хранение данных на внешнем контуре может не соответствовать внутренней корпоративной политике ряда организаций, а, если организация хранит личные данные пользователей, то данное решение будет противоречить законодательству Российской Федерации [17].

Альтернативным решением, в таком случае, может быть разработка вспомогательной информационной системы (сервиса), позволяющей централизованно взаимодействовать с различными типами сетевых хранилищ, при этом компоненты такой системы не должны предъявлять высокие требования к IT-инфраструктуре для развёртывания на локальных ресурсах организации. Архитектура предлагаемого решения представлена на рисунке 2:

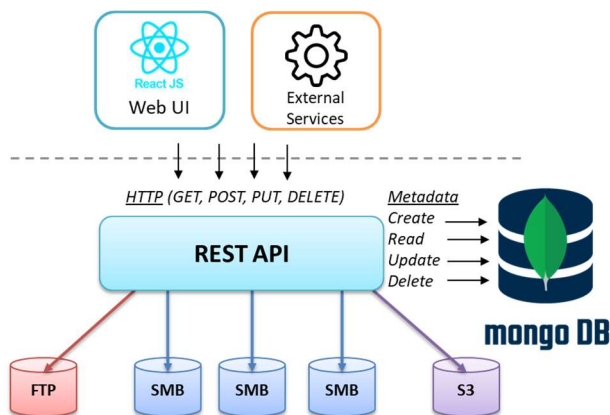


Рисунок 2 – Архитектура информационной системы для доступа к сетевым файловым хранилищам различного типа

Рассмотрим далее основные компоненты предлагаемого решения и их взаимодействие.

Центральным компонентом информационной системы является классический REST API, который предоставляет необходимые HTTP-методы для управления данными в хранилищах. Также данный компонент выполняет функции маршрутизации запросов к соответствующим хранилищам, используя их основные протоколы (FTP, SMB и т.д.). Запрос к API может иметь тип

application/json или multipart/form-data (для больших файлов), при этом непосредственно в хранилища передаются только данные файла (payload), а метаданные (metadata), находящиеся в других полях передаваемого json сохраняются в БД Mongo. Метаданные могут содержать точный адрес сохраняемого файла (с указанием типа и адреса хранилища), формат, теги, дополнительное описание. Таким образом, при поисковых запросах к REST API (GET) можно использовать встроенные возможности Mongo для поиска файлов по тегам, регулярным выражениям (в имени файла), а также использовать продвинутый метод полнотекстового поиска на основе Apache Lucene [18].

В качестве внешних компонентов, взаимодействующих с REST API, могут выступать:

1. Web UI (графический веб-интерфейс) – представляет собой React приложение, с помощью которого можно получить удобный доступ к данным в хранилищах на любых клиентских устройствах, поддерживающих современные браузеры.

2. External Services – любые внутренние корпоративные сервисы предприятия, которым также необходим доступ к данным (анализаторы данных, планировщики задач, системы рассылки и т.п.).

Исходя из сказанного выше, можно сделать вывод, что современные IT-решения предлагают достаточно широкий выбор инструментов и сервисов для хранения данных. Тем не менее, возникает ряд сложностей при работе с большим объёмом данных, расположенных на нескольких сетевых хранилищах, а также и при необходимости интеграции этих хранилищ в автоматизированные бизнес-процессы предприятия. Особенно данная проблема актуальна для предприятий малого и среднего бизнеса, которые (ввиду финансовых или внутрикорпоративных ограничений) не могут перейти на качественно новые решения, такие как Azure Storage или Amazon S3. Таким образом, разработка и внедрение в IT-инфраструктуру таких организаций вспомогательной информационной системы, которая решает проблемы централизованного взаимодействия с хранилищами данных и устраняет ряд недостатков используемых протоколов, является актуальной задачей.

Список литературы

[1] The ProFTPD Project. Project Documentation [Электронный ресурс] – URL: <http://www.proftpd.org/docs/> (дата обращения: 06.04.2023).

[2] RFC959: FTP. Declarative Specifications [Электронный ресурс] – URL: https://www.w3.org/Protocols/rfc959/5_Declarative.html. (дата обращения: 06.04.2023).

[3] Linux.org. NMAP FTP Bounce Attack [Электронный ресурс] – URL: <https://www.linux.org/threads/nmap-ftp-bounce-attack.4493/> (дата обращения: 06.04.2023).

[4] Энциклопедия «Касперского». Что такое контроллер домена (domain controller) [Электронный ресурс] – URL: <https://encyclopedia.kaspersky.ru/glossary/domain-controller/> (дата обращения: 06.04.2023).

[5] Samba – opening windows to a wider world [Электронный ресурс] – URL: <https://www.samba.org/> (дата обращения: 06.04.2023).

[6] Microsoft Learn. Server Message Block Overview [Электронный ресурс] – URL: [https://learn.microsoft.com/ru-ru/previous-versions/windows/it-pro/windows-server-2012-R2-and-2012/hh831795\(v=ws.11\)](https://learn.microsoft.com/ru-ru/previous-versions/windows/it-pro/windows-server-2012-R2-and-2012/hh831795(v=ws.11)) (дата обращения: 06.04.2023).

[7] Microsoft Learn. Защита трафика SMB от перехвата [Электронный ресурс] – URL: <https://learn.microsoft.com/ru-ru/windows-server/storage/file-server/smb-interception-defense?source=recommendations&tabs=group-policy> (дата обращения: 06.04.2023).

[8] Microsoft Learn. Slow SMB files transfer speed [Электронный ресурс] – URL: <https://learn.microsoft.com/en-us/windows-server/storage/file-server/troubleshoot/slow-file-transfer?source=recommendations> (дата обращения: 06.04.2023).

[9] Amazon Simple Storage Service. What is Amazon S3 Amazon Simple Storage Service [Электронный ресурс] – URL: <https://docs.aws.amazon.com/AmazonS3/latest/userguide/Welcome.html> (дата обращения: 06.04.2023).

[10] Amazon Web Services. Цены на облачное хранилище Amazon S3 Simple Storage Service [Электронный ресурс] – URL: <https://aws.amazon.com/ru/s3/pricing/> (дата обращения: 06.04.2023).

[11] MinIO. Enterprise Grade, High Performance Object Storage [Электронный ресурс] – URL: <https://min.io/product/overview>. (дата обращения: 06.04.2023).

[12] Yandex Cloud Object Storage. Как пользоваться S3 API [Электронный ресурс] – URL: <https://cloud.yandex.ru/docs/storage/s3/> (дата обращения: 06.04.2023).

[13] MinIO Object Storage for Linux. Configure MinIO for Authentication using OpenID [Электронный ресурс] – URL: <https://min.io/docs/minio/linux/operations/external-iam/configure-openid-external-identity-management.html>. (дата обращения: 06.04.2023).

[14] MinIO. Recommended Hardware & Configuration [Электронный ресурс] – URL: <https://min.io/product/reference-hardware> (дата обращения: 06.04.2023).

[15] Amazon Simple Storage Service. Filtering and retrieving data using Amazon S3 Select [Электронный ресурс] – URL: <https://docs.aws.amazon.com/AmazonS3/latest/userguide/selecting-content-from-objects.html> – (дата обращения: 06.04.2023).

[16] Glenn's SQL Server Performance. Azure Storage Primer [Электронный ресурс] – URL: <https://glennsqlperformance.com/2020/01/30/azure-storage-primer/> (дата обращения: 06.04.2023).

[17] КонсультантПлюс. Статья 1 [Электронный ресурс] – URL: http://www.consultant.ru/document/cons_doc_LAW_165838/3d0cac60971a511280cbba229d9b6329c07731f7/ (дата обращения: 06.04.2023).

[18] MongoDB. Full-Text Search: What Is It And How It Works | MongoDB [Электронный ресурс] – URL: <https://www.mongodb.com/basics/full-text-search> – (дата обращения: 06.04.2023).

Bibliography (Transliterated)

[1] The ProFTPD Project. Project Documentation [Electronic resource] – URL: <http://www.proftpd.org/docs/> (date of access: 04/06/2023).

[2] RFC959: FTP. Declarative Specifications [Electronic resource] – URL: https://www.w3.org/Protocols/rfc959/5_Declarative.html. (date of access: 04/06/2023).

[3] Linux.org. NMAP FTP Bounce Attack [Electronic resource] – URL: <https://www.linux.org/threads/nmap-ftp-bounce-attack.4493/> (accessed 04/06/2023).

[4] Kaspersky Encyclopedia. What is a domain controller [Electronic resource] – URL: <https://encyclopedia.kaspersky.ru/glossary/domain-controller/> (date of access: 04/06/2023).

[5] Samba – opening windows to a wider world [Electronic resource] – URL: <https://www.samba.org/> (date of access: 04/06/2023).

[6] Microsoft Learn. Server Message Block Overview [Electronic resource] – URL: [https://learn.microsoft.com/ru-ru/previous-versions/windows/it-pro/windows-server-2012-R2-and-2012/hh831795\(v=ws.11\)](https://learn.microsoft.com/ru-ru/previous-versions/windows/it-pro/windows-server-2012-R2-and-2012/hh831795(v=ws.11)) (date of access: 04/06/2023).

[7] Microsoft Learn. Protecting SMB traffic from interception [Electronic resource] – URL: <https://learn.microsoft.com/ru-ru/windows-server/storage/file-server/smb-interception-defense?source=recommendations&tabs=group-policy> (date of access: 04/06/2023).

[8] Microsoft Learn. Slow SMB files transfer speed [Electronic resource] – URL: <https://learn.microsoft.com/en-us/windows-server/storage/file-server/troubleshoot/slow-file-transfer?source=recommendations> (date of access: 04/06/2023).

[9] Amazon Simple Storage Service. What is Amazon S3 Amazon Simple Storage Service [Electronic resource] – URL: <https://docs.aws.amazon.com/AmazonS3/latest/userguide/Welcome.html> (accessed 04/06/2023).

[10] Amazon Web Services. Prices for cloud storage Amazon S3 Simple Storage Service [Electronic resource] – URL: <https://aws.amazon.com/en/s3/pricing/> (date of access: 04/06/2023).

[11] MinIO. Enterprise Grade, High Performance Object Storage [Electronic resource] – URL: <https://min.io/product/overview>. (date of access: 04/06/2023).

[12] Yandex Cloud Object Storage. How to use the S3 API [Electronic resource] – URL: <https://cloud.yandex.ru/docs/storage/s3/> (date of access: 04/06/2023).

[13] MinIO Object Storage for Linux. Configure MinIO for Authentication using OpenID [Electronic resource] – URL:

<https://min.io/docs/minio/linux/operations/external-iam/configure-openid-external-identity-management.html>. (date of access: 04/06/2023).

[14] MinIO. Recommended Hardware & Configuration [Electronic resource] – URL: <https://min.io/product/reference-hardware> (accessed 04/06/2023).

[15] Amazon Simple Storage Service. Filtering and retrieving data using Amazon S3 Select [Electronic resource] – URL: <https://docs.aws.amazon.com/AmazonS3/latest/userguide/selecting-content-from-objects.html> – (Accessed: 04/06/2023).

[16] Glenn's SQL Server Performance. Azure Storage Primer [Electronic resource] – URL: <https://glennsqlperformance.com/2020/01/30/azure-storage-primer/> (accessed 04/06/2023).

[17] ConsultantPlus. Article 1 [Electronic resource] – URL: http://www.consultant.ru/document/cons_doc_LAW_165838/3d0cac60971a511280cbba229d9b6329c07731f7/ (date of access: 04/06/2023).

[18] MongoDB. Full-Text Search: What Is It And How It Works | MongoDB [Electronic resource] – URL: <https://www.mongodb.com/basics/full-text-search> – (date of access: 04/06/2023).

© А.В. Алексеева, Б.С. Юдинцев, 2023

Поступила в редакцию 14.03.2023

Принята к публикации 06.04.2023

Для цитирования:

Алексеева А.В., Юдинцев Б.С. Информационная система для доступа к сетевым файловым хранилищам различного типа // Инновационные научные исследования. 2023. № 4-1(28). С. 95-107. URL: <https://ip-journal.ru/>