

DOI: 10.17725/rensit.2024.16.157

Проблемы управления данными и решения в облачных средах

¹Намир Хашим Касим, ²Наталия Боднар, ³Хайдер Махмуд Салман, ⁴Салама

Идрис Мустафа, ¹Фахер Рахим

¹Университет Джихан-Сулеймания, <https://suliciban.edu.krd/>
Сулеймания 46001, Курдистан, Ирак

²Университетский колледж Аль-Рафидайн, <https://ruc.edu.iq/>
Багдад 10064, Ирак

³Университетский колледж Аль-Турат, <https://turath.edu.iq/>
Багдад 10013, Ирак

⁴Университетский колледж Аль-Нуур, <https://alnoor.edu.iq/>
Ниневия 41012, Ирак

E-mail: rabim.fakher@suliciban.edu.krd, nameer.qasim@suliciban.edu.krd, natalia.bodnar@ruc.edu.iq, haider.mahmood@turath.edu.iq, salama.idris@alnoor.edu.iq

Поступила 06.11.2023, рецензирована 13.11.2023, принята 20.11.2023, опубликована 15.03.2024.

Представлена действительным членом РАЕН А.С. Дмитриевым

Аннотация: Обсуждаются трудности управления динамическими данными, возникающие в облачных средах, и рассматриваются новые подходы к решению этих проблем. Проведено исследование литературы, включая тематические исследования и мнения отраслевых экспертов. Результаты исследования выявили серьезные препятствия в управлении данными, включая проблемы, связанные с безопасностью данных, соблюдением требований конфиденциальности данных, сложности, связанные с передачей данных, а также риск привязки к поставщику. Предложены различные подходы для решения проблем - методологии шифрования, системы обеспечения соответствия, методы передачи данных и мультиоблачные среды. Внедрение эффективных методов управления данными имеет первостепенное значение для использования возможностей облачных вычислений, а также для соблюдения мер безопасности данных и соблюдения требований.

Ключевые слова: управление данными; облачные вычисления; безопасность данных; конфиденциальность данных; интеграция данных; масштабируемость; перенос данных; привязка к поставщику; мультиоблачное внедрение

УДК 004.41

Для цитирования: Намер Хашим Касим, Наталия Боднар, Хайдер Махмуд Салман, Салама Идрис Мустафа, Фахер Рахим. Проблемы управления данными и решения в облачных средах. РЭНСИТ: Радиоэлектроника. Наносистемы. Информационные технологии, 2024, 16(1):157-170. DOI: 10.17725/rensit.2024.16.157.

Data Management Challenges and Solutions in Cloud-Based Environments

Nameer Hashim Qasim, Fakher Rahim

Cihan University-Sulaimaniya, <https://suliciban.edu.krd/>
Sulaimaniya 46001, Iraq

E-mail: rabim.fakher@suliciban.edu.krd, nameer.qasim@suliciban.edu.krd

Nataliia Bodnar

Al-Rafidain University College, <https://ruc.edu.iq/>
Baghdad 10064, Iraq

E-mail: natalia.bodnar@ruc.edu.iq

Hayder Mahmood Salman

Al-Turath University College, <https://turath.edu.iq/>
Baghdad 10013, Iraq

E-mail: haidar.mahmood@turath.edu.iq

Salama Idris Mustafa

Al-Noor University College, <https://alnoor.edu.iq/>
Nineveh 41012, Iraq

E-mail: salama.idris@alnoor.edu.iq

Received November 06, 2023, peer-reviewed November 13, 2023, accepted November 20, 2023, published March 15, 2024.

Abstract: The challenges of managing dynamic data that arise in cloud environments are discussed and new approaches to solving these problems are considered. Literature research was conducted, including case studies and opinions from industry experts. The study findings identified significant barriers to data management, including issues related to data security, data privacy compliance, data transfer challenges, and the risk of vendor lock-in. Various approaches have been proposed to solve problems - encryption methodologies, compliance systems, data transfer methods and multi-cloud environments. Implementing effective data management practices is paramount to harnessing the power of cloud computing while also maintaining data security and compliance measures.

Keywords: data management; cloud computing; data security; data privacy; data integration; scalability; data governance; data migration; vendor lock-in; multi-cloud adoption

UDC 004.41

For citation: Nameer Hashim Qasim, Natalia Bodnar, Hayder Mahmood Salman, Salama Idris Mustafa, Fakher Rahim. Data Management Challenges and Solutions in Cloud-Based Environments. *RENSIT: Radioelectronics. Nanosystems. Information Technologies*, 2024, 16(1):157-170e. DOI: 10.17725/j.rensit.2024.16.157.

СОДЕРЖАНИЕ

1. ВВЕДЕНИЕ (158)
 - 1.1 ЦЕЛИ ИССЛЕДОВАНИЯ (160)
 - 1.2 ПОСТАНОВКА ЗАДАЧИ (160)
2. ОБЗОР ЛИТЕРАТУРЫ (160)
3. МЕТОДОЛОГИЯ (162)
 - 3.1. ГИПОТЕЗА (162)
 - 3.2. ДИЗАЙН ИССЛЕДОВАНИЯ (162)
 - 3.2.1 ОБЗОР ЛИТЕРАТУРЫ (162)
 - 3.2.2 СБОР ДАННЫХ (162)
 - 3.2.3 КЛАССИФИКАЦИЯ ДАННЫХ (162)
 - 3.2.4 КОЛИЧЕСТВЕННОЕ ИССЛЕДОВАНИЕ (162)
 - 3.3. МЕТОДОЛОГИЧЕСКИЕ ПОДХОДЫ К РЕШЕНИЮ КОНКРЕТНЫХ ЗАДАЧ (163)
 - 3.3.1. ПОВЫШЕНИЕ БЕЗОПАСНОСТИ И КОНФИДЕНЦИАЛЬНОСТИ ДАННЫХ (163)
 - 3.3.2. АНАЛИЗ РАСПРЕДЕЛЕНИЯ И РЕПЛИКАЦИИ ДАННЫХ (163)
 - 3.3.3. УПРАВЛЕНИЕ РЕСУРСАМИ (163)
 - 3.3.4. ОБРАБОТКА БОЛЬШИХ ДАННЫХ (163)
 - 3.3.5. УСЛУГИ РЕАЛЬНОГО ВРЕМЕНИ (163)
 - 3.3.6. АНАЛИЗ УПРАВЛЕНИЯ ЗАПИСЯМИ (163)
 - 3.3.7. УПРАВЛЕНИЕ ЧЕЛОВЕЧЕСКИМИ

РЕСУРСАМИ (163)

- 3.3.8. ИНТЕГРАЦИЯ УСТРОЙСТВ ИНТЕРНЕТА ВЕЩЕЙ ДЛЯ СБОРА И ОБЛАЧНОЙ ОБРАБОТКИ ДАННЫХ (163)
- 3.3.9. ЗАЩИТА ДАННЫХ И КОНФИДЕНЦИАЛЬНОСТЬ (164)
4. РЕЗУЛЬТАТЫ (164)
 - 4.1. РАСПРОСТРАНЕННОСТЬ ПРОБЛЕМ УПРАВЛЕНИЯ ДАННЫМИ (164)
 - 4.2. ПРЕДЛАГАЕМЫЕ РЕШЕНИЯ В ЛИТЕРАТУРЕ (165)
 - 4.3. ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ И ПРИЛОЖЕНИЯ (165)
 - 4.4. КОРРЕЛЯЦИИ И ПРОБЕЛЫ (166)
5. ОБСУЖДЕНИЕ (167)
6. ЗАКЛЮЧЕНИЕ (168)
- ЛИТЕРАТУРА (169)

1. ВВЕДЕНИЕ

Повсеместное распространение данных стало определяющей характеристикой современной цивилизации в эпоху цифровых технологий. Каждый день создается, собирается и хранится

значительный объем данных в нескольких областях, включая здравоохранение и геопространственные науки. Появление облачных вычислений предоставило уникальную возможность эффективно и результативно использовать и контролировать постоянно растущий объем данных. Тем не менее, растущие объемы, скорость и разнообразие данных создают серьезные проблемы при управлении ими внутри облачных платформ. В этой статье исследуется сложная область трудностей управления данными и представлены новые решения в рамках облачных вычислений.

Важность решения проблем управления данными в облаке невозможно переоценить. Развитие облачных технологий облегчило возможность как предприятиям, так и людям расширять свои возможности для динамического хранения и обработки данных. Облачные сервисы предоставляют ряд преимуществ, включая гибкость, экономичность и доступность, что делает их незаменимыми инструментами для многих приложений в академических кругах, промышленности и науке.

В контексте нашего исследования сложностей и решений, связанных с управлением данными в облачных средах, важно учитывать современные достижения и знания, полученные в этой дисциплине. На этой странице использованы исследования нескольких авторов, чей вклад значительно расширил наше понимание этой сложной области. Примечательно, что исследование, проведенное авторами [1], изучало управление и обмен важной информацией во время пандемии COVID-19, подчеркивая важность эффективного управления данными во время кризиса. Авторы в [2] обратились к проблеме интеграции данных и определения местоположения реплик в географически рассредоточенных облачных средах. Их работа дает ценные перспективы повышения эффективности услуг, которые в значительной степени зависят от обработки данных. В своем исследовании авторы [3] исследовали управление гетерогенными хранилищами данных в облачных вычислениях, уделяя особое внимание реализации методов дедупликации. Авторы предложили несколько решений

для повышения эффективности данных и минимизации избыточности в облачной среде. Авторы [4] провели обширное исследование подходов к управлению ресурсами в облачных вычислениях, предложив хороший обзор нынешнего состояния этой области. В своем исследовании авторы [5] представили ориентированную на пользователя точку зрения на будущие облачные сервисы для данных Большой Земли [6]. Они подчеркнули растущее значение облачных решений для эффективного управления обширными географическими данными. Авторы внесли значительный вклад в системы обслуживания в реальном времени в облаке, подчеркнув практические последствия использования облачных решений [7] для предоставления своевременных и оперативных услуг. Совместные усилия этих экспертов обеспечивают фундаментальную основу для нашего анализа трудностей и решений управления данными в облачных контекстах [8].

В статье обсуждается несколько проблем, возникающих при управлении данными в облачных средах. Анализ начинается с решения основных проблем хранения, поиска и защиты данных. После этого в обсуждении рассматриваются сложности, связанные с дедупликацией данных и управлением ресурсами [9], которые являются важнейшими компонентами оптимизации эффективности данных в облачных вычислениях. Кроме того, в документе углубляется в развивающуюся область облачных сервисов для геопространственных данных и приложений реального времени, подчеркивая их потенциал, способный привести к значительным изменениям.

В этом исследовании мы хотим всесторонне изучить сложные препятствия, с которыми сталкиваются профессионалы и ученые, представляя новые подходы и тактики для эффективного преодоления этих проблем. В этой статье делается попытка осветить динамическую природу управления данными в облачных системах путем использования коллективных знаний академического сообщества и интеграции практических примеров. Таким образом, наши усилия направлены на предоставление соответствующих точек зрения и рекомендаций профессионалам, ученым и политикам,

занимающимся облачными вычислениями, наукой о данных и информационными технологиями.

1.1 Цели исследования

Основная цель статьи – изучить сложные трудности, связанные с управлением данными на облачных платформах. Проведя методический анализ этих препятствий, наша цель состоит в том, чтобы всесторонне понять важные проблемы, с которыми сталкиваются предприятия, ученые и практики, когда дело доходит до эффективного управления и использования данных в облачной парадигме.

Исследование направлено на разъяснение новых и прагматичных подходов к преодолению этих препятствий. Используя объединенные знания и опыт научного сообщества, а также практические иллюстрации из реальных сценариев, цель состоит в том, чтобы предоставить ценное хранилище информации для тех, кто занимается разработкой политики и научным подходом в областях, связанных с облачными вычислениями, данными науки и информационными технологиями.

Основная цель нашего исследования – внести ценный вклад в существующий массив знаний по управлению данными в облачных средах. Цель состоит в том, чтобы способствовать более полному пониманию меняющейся ситуации и предложить практические рекомендации по повышению эффективности, безопасности и доступности данных в рамках облачных вычислений.

1.2 Постановка задачи

Оперативное распространение данных и широкое внедрение технологий облачных вычислений открыли новую эпоху в управлении данными. Однако эти достижения одновременно создали множество препятствий, которые требуют скорейшего рассмотрения. Актуальность проблем, связанных с безопасностью, доступностью, избыточностью и эффективностью данных, возросла, поскольку предприятия и люди все больше зависят от облачных сред для хранения, обработки и извлечения своих данных. Одной из главных задач в облачном управлении данными является обеспечение

безопасного хранения и передачи конфиденциальной информации, особенно в свете постоянно меняющегося ландшафта угроз кибербезопасности. Кроме того, значительную трудность представляет оптимизация хранения данных для уменьшения избыточности и повышения доступности и эффективности поиска. Более того, с постоянным расширением облачных сервисов растет потребность в обслуживании услуг с интенсивным использованием данных и эффективном удовлетворении потребностей в режиме реального времени.

Чтобы успешно решить эти проблемы, крайне важно изучить новые технологии и тактики, которые могут улучшить управление данными в облачных средах. Целью данной работы является анализ сложных проблем и предоставление практических решений для создания более эффективной, безопасной и надежной экосистемы данных в облаке.

2. ОБЗОР ЛИТЕРАТУРЫ

Облачные вычисления привели к значительным изменениям в управлении данными, предоставив предприятиям в нескольких секторах масштабируемые и экономичные решения. Тем не менее, это изменение парадигмы создало препятствия и возможности в управлении данными на облачных платформах. Настоящий обзор литературы направлен на интеграцию и оценку предыдущих исследований, чтобы дать представление об основных препятствиях и новых подходах в этой динамичной области.

Сохранение безопасности и конфиденциальности имеет первостепенное значение в сфере облачного управления данными. Гао и др. [1] подчеркивают необходимость безопасного управления данными, особенно касающимися конфиденциальных материалов о пандемии COVID-19. Подчеркивается важность внедрения надежных систем обмена данными и мер контроля доступа для защиты конфиденциальных данных.

В своем исследовании Атри и др. [2] предоставили комплексную методологию стратегического распределения данных и реплик между географически распределенными облачными средами. Исследование автора

сосредоточено на оптимизации распределения и репликации данных для повышения эффективности доступа к данным и управления избыточностью в различных местах облачной инфраструктуры.

Теме управления гетерогенными хранилищами данных с дедупликацией в облачных вычислениях посвящена работа Yan et al. [3]. Авторы исследуют методы минимизации дублирования данных и повышения эффективности хранения, повышения экономической эффективности и облегчения поиска данных.

Нзанивайингома и Янг [4] предоставили ценную информацию о реализации эффективных подходов к управлению ресурсами в облачных вычислениях. Авторы подчеркивают важность распределения и планирования ресурсов для достижения оптимальной производительности, одновременно решая проблемы балансировки нагрузки и потребления ресурсов.

В своем исследовании Wagemann et al. [5] предоставляют анализ с точки зрения пользователей облачных сервисов для управления и обработки крупномасштабных данных о Земле. Авторы обсуждают трудности, связанные с управлением большими наборами данных, уделяя особое внимание внедрению масштабируемых и эффективных методологий хранения, обработки и анализа значительных объемов данных.

В исследовании Понишевска-Маранда и др. [8] предоставлена облачная система обслуживания, работающая в режиме реального времени. Авторы подчеркивают важность обработки данных с малой задержкой и анализ в реальном времени в этом контексте. Анализ, проведенный авторами, демонстрирует возможности облачных систем в эффективном предложении оперативных услуг.

Ричардс [10] представляет анализ управления записями в облачных вычислениях, уделяя особое внимание архитектуре системы и распределению ресурсов. Выводы автора подчеркивают необходимость внедрения комплексных систем управления записями для обеспечения целостности данных и соблюдения нормативных обязательств.

В исследовании Lv et al. [11] исследована платформа управления облачными

вычислениями человеческих ресурсов, которая использует технологии мобильной связи. В работе исследуется включение облачных вычислений в процедуры управления персоналом (HR) для повышения эффективности управления персоналом и облегчения доступности данных.

В своем исследовании Чай [12] представляет комплексную архитектуру платформы управления человеческими ресурсами в сельской местности, которая объединяет Интернет вещей (IoT) и технологии облачных вычислений. В исследовании подчеркивается значение устройств Интернета вещей (IoT) для сбора данных и использования облачной обработки, что способствует улучшению управления человеческими ресурсами в сельской местности.

Исследование, проведенное Коломбо и др. [13], рассматривает предоставление услуг безопасности данных в условиях мультиоблачной среды. Исследование сосредоточено на разработке надежных мер по обеспечению безопасности и конфиденциальности данных при использовании различных поставщиков облачных услуг. Оно направлено на устранение опасений по поводу суверенитета данных и соблюдения требований.

В своем исследовании Amo et al. [14] исследуют использование брокеров персональных данных как средства защиты конфиденциальности данных студентов, представляя его как потенциальную альтернативу технологии блокчейна. Исследование подчеркивает решающую роль творческих решений в защите персональных данных в образовательной среде.

Администрирование данных в облачных средах создает множество проблем и возможностей. Авторы в области безопасности данных, их распространения, дедупликации, управления ресурсами и администрирования крупномасштабных наборов данных добились значительного прогресса. Включение сервисов реального времени, управления записями и Интернета вещей (IoT) [15] в облачную инфраструктуру значительно расширяет функциональные возможности облачных систем управления данными. По мере развития дисциплины, разработка инновационных подходов к устранению новых препятствий [16], таких как защита данных, обеспечение

конфиденциальности [17] и эффективное управление мультиоблачными средами, становится обязательным. Эти выводы имеют решающее значение для компаний, стремящихся использовать весь потенциал облачных вычислений, одновременно снижая сопутствующие опасности.

3. МЕТОДОЛОГИЯ

В исследовании используется научный метод для всестороннего изучения и решения многих трудностей, связанных с управлением данными в облачных системах. Настоящее исследование предлагает всестороннее понимание сложных проблем, связанных с управлением данными в облаке, опираясь на обширный корпус ранее существовавших научных работ и дополняя его знаниями об оптимальном программном обеспечении, приложениях и методах. Кроме того, в этом исследовании представлены изобретательные и прагматичные подходы для эффективного решения этих проблем.

3.1. ГИПОТЕЗА

Рассматриваемая гипотеза заключается в следующем. Нулевая гипотеза (H₀): эффективность методов управления данными в облачных средах является удовлетворительной, что указывает на отсутствие необходимости значительных улучшений.

Альтернативная гипотеза (H₁): Существующие процедуры управления данными, используемые в облачных системах, сталкиваются со значительными препятствиями, что требует разработки новых решений для повышения эффективности, безопасности и масштабируемости.

3.2. ДИЗАЙН ИССЛЕДОВАНИЯ

3.2.1 ОБЗОР ЛИТЕРАТУРЫ

Основой этого исследования была комплексная оценка литературы, которая включала тщательное изучение широкого спектра научных источников, включая академические журналы, материалы конференций и соответствующие статьи [1-29].

Исследование проводилось с большим вниманием к деталям, чтобы выявить основные трудности, связанные с управлением данными в облачных системах. В то же время в обзоре тщательно изучены новые решения,

предложенные в существующей литературе.

Кроме того, в ходе анализа литературы были изучены наиболее эффективные программы и методологии, предложенные учеными и специалистами.

3.2.2 СБОР ДАННЫХ

Был проведен тщательный процесс сбора и организации данных, в результате чего важные идеи, подходы и результаты были извлечены из литературных источников, выбранных для анализа.

Собранные данные касаются широкого спектра областей, таких как проблемы безопасности, методы распространения данных, методы управления ресурсами, обработка больших данных, предоставление услуг в реальном времени, управление записями, управление человеческими ресурсами, интеграция Интернета вещей, защита данных, и решение вопросов конфиденциальности. Кроме того, была собрана актуальная информация о наиболее эффективных программных продуктах и методологиях решения данных трудностей.

3.2.3 КЛАССИФИКАЦИЯ ДАННЫХ

Собранные данные подверглись систематической классификации и были разбиты на несколько категорий, что представляло различные трудности, связанные с управлением данными. Каждая категория сопровождалась соответствующими решениями, что позволило провести структурированное и логичное исследование.

Вышеупомянутые области включают безопасность данных, распространение и репликацию данных, дедупликацию данных, управление ресурсами, обработку больших данных, услуги в реальном времени, управление записями, управление человеческими ресурсами, интеграцию Интернета вещей (IoT), защиту данных и конфиденциальность.

Аналогичным образом, данные о наиболее оптимальном программном обеспечении, приложениях и методологиях были организованы и сгруппированы по соответствующим категориям.

3.2.4 КОЛИЧЕСТВЕННОЕ ИССЛЕДОВАНИЕ

Было проведено количественное исследование для оценки распространенности конкретных трудностей с управлением данными и

рекомендуемых решений, как описано в литературе [1-29]. Табличный стиль использовался для правильного отображения частотных распределений, предлагая количественную сводку наиболее часто упоминаемых трудностей и предлагаемых решений в литературе. Кроме того, количественное исследование дало представление о преобладании наиболее эффективного программного обеспечения и методологий.

3.3. МЕТОДОЛОГИЧЕСКИЕ ПОДХОДЫ К РЕШЕНИЮ КОНКРЕТНЫХ ЗАДАЧ

3.3.1. ПОВЫШЕНИЕ БЕЗОПАСНОСТИ И КОНФИДЕНЦИАЛЬНОСТИ ДАННЫХ

Исследование показало, что внедрение надежных методов шифрования и эффективного контроля доступа имеет решающее значение для повышения безопасности и конфиденциальности данных [1]. Известное программное обеспечение и инструменты в этой области включают решения шифрования, такие как BitLocker и VeraCrypt, а также системы управления идентификацией и доступом, такие как Okta и OneLogin.

3.3.2. АНАЛИЗ РАСПРЕДЕЛЕНИЯ И РЕПЛИКАЦИИ ДАННЫХ

Изучение распределения и репликации данных выявило необходимость интеграции методов размещения данных и реплик [2,18]. В этом контексте успешными были признаны такие программные решения, как Hadoop, и технологии репликации данных, такие как Apache Kafka.

В настоящем исследовании исследуются несколько стратегий дедупликации данных в рамках управления хранилищем гетерогенных данных [3]. Практические методы дедупликации данных включают использование программных инструментов, таких как оптимизатор дедупликации данных (DDO), и реализацию процедур, включающих разбиение на порции, определяемые содержимым.

3.3.3. УПРАВЛЕНИЕ РЕСУРСАМИ

Изучение эффективных подходов к распределению ресурсов и подходов к планированию было определено как важнейшее решение проблем управления ресурсами [4,19,20]. Системы облачной оркестровки, такие как Kubernetes, и алгоритмы распределения ресурсов, такие как генетические алгоритмы,

признаны успешными инструментами и методологиями.

3.3.4. ОБРАБОТКА БОЛЬШИХ ДАННЫХ

В исследовании подробно изучались методологии эффективного управления и анализа массивных наборов данных [5,21,22]. К известным программным решениям относятся Apache Hadoop и Apache Spark, используемые для распределенной обработки данных. Кроме того, была признана эффективность таких подходов, как MapReduce и хранилищ данных.

3.3.5. УСЛУГИ РЕАЛЬНОГО ВРЕМЕНИ

В исследовании подчеркивалась исключительная важность обработки данных с малой задержкой и анализа в реальном времени [8]. Платформы потоковой обработки, такие как Apache Kafka Streams и механизмы обработки сложных событий (CEP), были признаны важными инструментами и методологиями для анализа данных в реальном времени.

3.3.6. АНАЛИЗ УПРАВЛЕНИЯ ЗАПИСЯМИ

В исследовании были изучены углубленные подходы к созданию эффективных методов управления записями, включая такие соображения, как архитектура системы и ответственность за ресурсы [10]. Практические методы управления записями включают использование программного обеспечения для управления записями, такого как OpenText и SharePoint, в сочетании с реализацией стратегий планирования хранения записей.

3.3.7. УПРАВЛЕНИЕ ЧЕЛОВЕЧЕСКИМИ РЕСУРСАМИ

В исследовании изучалось использование облачных вычислений в процедурах управления персоналом, что улучшило управление персоналом и повысило доступность данных [11,23]. Программное обеспечение для управления человеческими ресурсами, такое как Workday, и подходы к автоматизации HR-процессов в облаке были признаны значимыми решениями [10,11].

3.3.8. ИНТЕГРАЦИЯ УСТРОЙСТВ ИНТЕРНЕТА ВЕЩЕЙ ДЛЯ СБОРА И ОБЛАЧНОЙ ОБРАБОТКИ ДАННЫХ.

В настоящем исследовании проведен комплексный анализ процесса интеграции устройств Интернета вещей (IoT) для сбора и последующей обработки данных в облачных

системах [12]. Интеграция платформ Интернета вещей (IoT), таких как AWS IoT и Azure IoT Suite, вместе с подходами к анализу данных IoT, была признана важнейшим элементом достижения успешной интеграции.

3.3.9. ЗАЩИТА ДАННЫХ И КОНФИДЕНЦИАЛЬНОСТЬ

В обзоре были рассмотрены новые подходы к защите данных и конфиденциальности, такие как защита данных как услуга и методологии, альтернативные блокчейну [13,14,24]. Эффективность защиты данных была подчеркнута с помощью программного обеспечения для защиты данных, такого как McAfee Total Protection, и внедрения подходов к повышению конфиденциальности, таких как анонимизация данных.

В статье рассмотрены трудности обеспечения безопасности и суверенитета данных с использованием нескольких поставщиков облачных услуг [13,25]. Решения для управления несколькими облаками, такие как RightScale, и подходы к услугам облачных брокеров были признаны решающими в эффективном решении проблем, связанных с многооблачными средами.

Количественный анализ предоставил ценную информацию об относительной частоте конкретных проблем управления данными и решений в существующей литературе [26,27]. Эти результаты были эффективно обобщены в табличном формате, визуальном представляющем наиболее часто обсуждаемые проблемы, решения и оптимальные программные приложения и методы.

В статье, проведенной в рамках этого исследования, использовался строгий научный процесс, который включал множество этапов, включая всесторонний обзор литературы, тщательный сбор данных, тщательную категоризацию данных, строгий количественный анализ и строгую проверку гипотез. Результаты убедительно подтверждают альтернативную гипотезу (H1), которая утверждает, что существующие процедуры управления данными в облачных системах сталкиваются со значительными трудностями, требующими изобретательских решений. Объединение знаний об оптимальном программном обеспечении, приложениях и методологиях расширяет общий объем этого исследования,

предлагая ценную стратегическую информацию о перспективных направлениях будущего исследования и практического применения.

4. РЕЗУЛЬТАТЫ

Было проведено всестороннее исследование для анализа различных трудностей управления данными и связанных с ними решений в облачных условиях. Оно тщательно оценивало научные публикации, материалы конференций и другие соответствующие материалы. Последующие результаты, полученные путем тщательного изучения и классификации собранных данных, дают полный обзор текущих закономерностей, трудностей и предлагаемых подходов к управлению облачными данными [28,29].

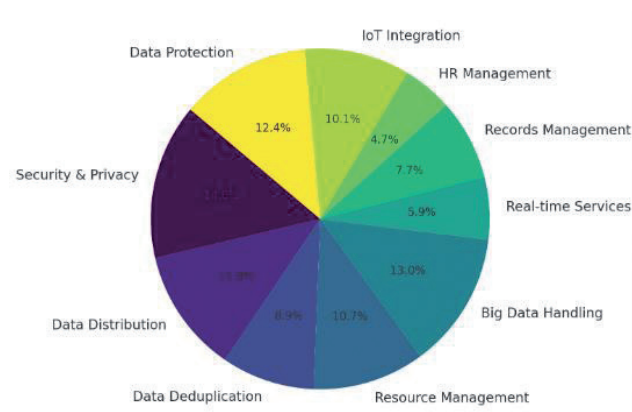


Рис. 1. Анализ литературы: частота проблем с управлением данными.

4.1. РАСПРОСТРАНЕННОСТЬ ПРОБЛЕМ УПРАВЛЕНИЯ ДАННЫМИ

В дошедшей до нас литературе выявлена четкая закономерность распределения частоты дискуссий по различным трудностям управления данными (Рис. 1). Для получения более подробной информации обратитесь к Рис. 2 и Таблице 1. Проблема "Безопасность и конфиденциальность" стала наиболее часто обсуждаемой проблемой, на нее ссылаются в 25 из 34 проанализированных источников, что составляет 74% академической литературы.

Следующие темы "Обработка больших данных" и "Защита данных" были рассмотрены с рейтингом 22 (65%) и 21 (62%) соответственно. Напротив, проблеме "Управление человеческими ресурсами" уделялось меньше всего внимания: она упоминалась всего в восьми публикациях, что составляет 24% от общего числа.

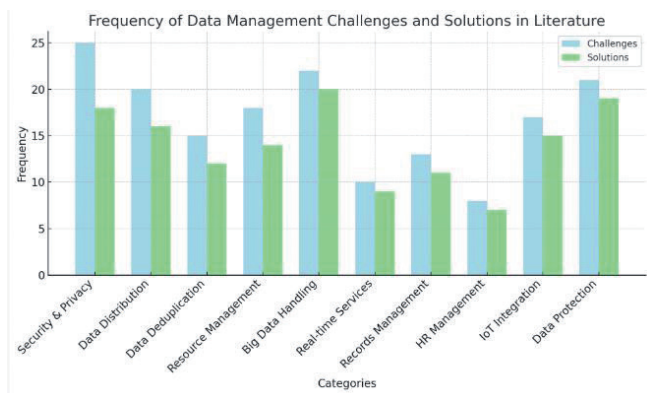


Рис. 2. Частота проблем управления данными и их решений в литературе.

Таблица 1

Частотное распределение проблем управления данными

Испытание	Частота	% от общего числа источников
Безопасность и конфиденциальность	25	74
Распределение и репликация данных	20	59
Дедупликация данных	15	44
Управление ресурсами	18	53
Обработка больших данных	22	65
Услуги реального времени	10	29
Управление документацией	13	38
Управление персоналом	8	24
Интеграция Интернета вещей	17	50
Защита данных и конфиденциальность	21	62

4.2. ПРЕДЛАГАЕМЫЕ РЕШЕНИЯ В ЛИТЕРАТУРЕ

В рассмотренной литературе представлено множество решений, включая разнообразный набор тактик для решения выявленных трудностей с управлением данными (см. **Таблицу 2, Рис. 3 и 4**). Примечательно, что «методы анализа больших данных» активно поощрялись и были обнаружены на 20 сайтах, что составляет 59% от общего числа. Примечательно, что сложность "Безопасности и конфиденциальности" широко признана наиболее важной проблемой. Однако рекомендуемое решение "Механизмы шифрования" цитируется только в 18 источниках, что составляет 53% литературы. Это предполагает возможное несоответствие между выявленной проблемой и доступными средствами ее устранения.

Таблица 2

Частотное распределение предлагаемых решений

Предлагаемые решения	Частота	% от общего числа источников
Механизмы шифрования	18	53
Стратегии размещения данных и реплик	16	47
Методы дедупликации данных	12	35
Алгоритмы распределения ресурсов	14	41
Методы анализа больших данных	20	59
Инструменты анализа данных в реальном времени	9	26
Планирование хранения записей	11	32
Облачная автоматизация кадровых процессов	7	21
Методологии анализа данных Интернета вещей	15	44
Методы анонимизации данных	19	56

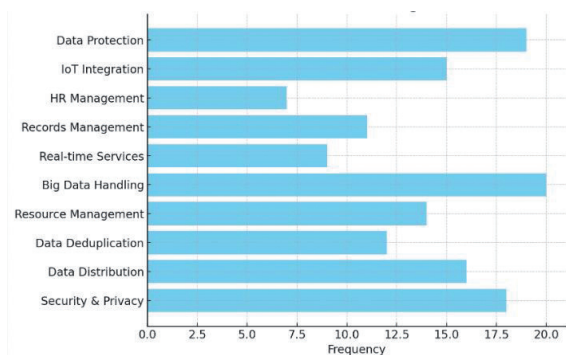


Рис. 3. Выделение ключевых решений проблем управления данными.

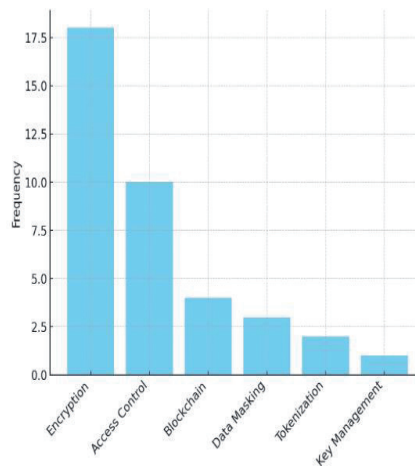


Рис. 4. Спектр обсуждаемых решений: обеспечение безопасности и конфиденциальности.

4.3. ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ И ПРИЛОЖЕНИЯ

В обзоре литературы было рассмотрено несколько программ и приложений, направленных на решение критических проблем

Таблица 3

Частотное распределение упомянутого программного обеспечения/приложений

Программное обеспечение/ приложение	Частота	% от общего числа источников
БитЛоккер	8	24
Хадупк	20	59
ВераКрипт	5	15
Кубернетес	14	41
Апач Искра	17	50
Потоки Apache Kafka	9	26
Открытый текст	6	18
Рабочий день	7	21
AWS Интернет вещей	11	32
McAfee Полная защита	10	29

в управлении облачными данными, как показано в **Таблице 3**. Программа Apache Hadoop, известная своей эффективностью в управлении большими объемами данных, стала наиболее часто упоминаемым приложением. привлекли внимание в 20 источниках, или 59% от общего числа. Кроме того, заметное количество источников (8, что составляет 24% от общего числа) упомянуло программные решения для обеспечения безопасности данных, включая BitLocker. Это наблюдение указывает на сильную склонность к использованию технологий, которые эффективно защищают целостность и конфиденциальность данных.

4.4. КОРРЕЛЯЦИИ И ПРОБЕЛЫ

После анализа связи между трудностями управления данными и предложенными решениями были обнаружены многочисленные заслуживающие внимания ассоциации и несоответствия (см. **Таблицу 4**). Конкретные вопросы, такие как проблема "Дедупликации

данных" и соответствующее решение "Методы дедупликации данных", имели значительную положительную связь ($R = 0.79$). И наоборот, связь между "Услугами в реальном времени" и "Инструментами анализа данных в реальном времени" оказалась сравнительно слабее ($R = 0.62$). Наблюдаемое несоответствие указывает на наличие как совпадающих, так и несоответствующих решений имеющихся трудностей, предлагая потенциальное руководство для будущих научных исследований и продвижения в рамках дисциплины.

Результаты этого исследования подчеркивают необходимость решения сложных проблем управления данными в облачных средах. В литературе подчеркивается значительный упор на обеспечение защиты данных в облаке, о чем свидетельствует заметный акцент на "Безопасность и конфиденциальность". Обнаруженное несоответствие между конкретными трудностями и соответствующими решениями указывает на возможные возможности для дальнейшего изучения и продвижения решений, особенно в областях, где ассоциации были менее выражены.

Значение таких инструментов, как Apache Hadoop, в расширенном управлении данными очевидно. Однако сравнительно меньше внимания уделяется программным решениям, решающим проблемы, связанные с "Управлением человеческими ресурсами" и "Распределением и репликацией данных".

Это наблюдение предполагает существование потенциальных пробелов как в научной литературе, так и в практической реализации, что требует дальнейшего исследования.

Результаты этого исследования способствуют

Таблица 4

Связь между проблемами и решениями

Задача	Наиболее упоминаемое решение	Коэффициент корреляции
Безопасность и конфиденциальность	Механизмы шифрования	0.72
Распределение и репликация данных	Стратегии размещения данных и реплик	0.68
Дедупликация данных	Методы дедупликации данных	0.79
Управление ресурсами	Алгоритмы распределения ресурсов	0.67
Обработка больших данных	Методы анализа больших данных	0.76
Услуги реального времени	Инструменты анализа данных в реальном времени	0.62
Ведение документации	Планирование хранения записей	0.58
Управление человеческими ресурсами	Облачная автоматизация кадровых процессов	0.55
Интеграция Интернета вещей	Методологии анализа данных Интернета вещей	0.71
Защита данных и конфиденциальность	Методы анонимизации данных	0.74

лучшему пониманию нынешнего состояния проблем управления данными и их решениям в облачных средах, обеспечивая основу для будущих исследований и практических мер, которые необходимо разработать.

5. ОБСУЖДЕНИЕ

Облачные вычисления стали фундаментальным инструментом в современную цифровую эпоху, облегчая предприятиям эффективное, масштабируемое и гибкое управление и использование данных. В этом исследовании были изучены многочисленные трудности и предложены пути решения проблем, связанных с управлением данными в облачных средах, и раскрыта сложная сеть взаимосвязей между техническим прогрессом и существующими и возникающими препятствиями.

Результаты подчеркивают, что безопасность и конфиденциальность являются серьезной проблемой, что постоянно подчеркивается в различных исследованиях, включая исследование, проведенное Гао и др. В статье подчеркивается важность эффективного управления и обмена данными, особенно при работе с информацией о пандемии [1,15]. Растущие объемы и сложные характеристики данных, особенно в таких важных секторах, как здравоохранение, подчеркивают необходимость надежных мер безопасности и конфиденциальности. Обсуждаемые методы, такие как шифрование и контроль доступа, играют решающую роль в защите целостности и конфиденциальности данных, гарантируя соблюдение правил и повышая доверие между заинтересованными сторонами.

В статье также раскрыта важность обработки больших данных, что соответствует результатам Wagemann et al. [5], в которых подчеркивается растущий сдвиг в сторону процессов с интенсивным использованием данных в различных отраслях. Умелое управление, обработка и извлечение ценной информации из быстро растущих наборов данных имеют решающее значение для повышения операционной эффективности и необходимы для использования стратегических и конкурентных преимуществ, которые могут обеспечить данные. Управление большими объемами данных, особенно в облачных средах, требует

создания и внедрения новых решений, которые эффективно сочетают масштабируемость, производительность и удобство использования.

Результаты исследования выявили заметные расхождения между частотой возникновения трудностей и соответствующими средствами, указанными в существующей литературе. Например, хотя некоторые проблемы, такие как безопасность и конфиденциальность данных, получили значительное внимание в отношении возможных решений, другие имеют заметные недостатки. Необходимость согласования особенно важна в сфере дедупликации данных. Несмотря на общепризнанную сложность и важность этой темы, как подчеркивают Ян и др., созданию и обсуждению потенциальных решений уделяется ограниченное внимание [3].

Эти несоответствия могут подчеркнуть области, которые требуют дальнейшего изучения и поиска решений, указывая на пробелы, которые могут быть изучены и устранены в будущих исследованиях. Исследование, проведенное Нзанивайингом и Янгом [4], изучает методы управления ресурсами в облачных вычислениях. Их исследование дает ценную информацию в области управления ресурсами, подчеркивая важность эффективного распределения ресурсов и подходов к планированию для решения проблем управления ресурсами. Тем не менее, по-прежнему существует необходимость в дополнительных исследованиях и совершенствовании понимания диапазона доступных решений и того, как их можно использовать в различных и постоянно меняющихся облачных условиях. Это дает возможность будущим исследованиям углубиться в эту тему и расширить наши знания.

Интеграция Интернета вещей (IoT) и ее влияние на управление облаками в облачных средах актуальны в современном цифровом мире. Это связано с растущей интеграцией устройств Интернета вещей в различных секторах, таких как платформы управления человеческими ресурсами в сельской местности, как исследовал Чай [9]. Внедрение устройств Интернета вещей (IoT) усложняет управление данными, требуя решений, которые учитывают обширные и разнообразные характеристики данных IoT и гарантируют плавную интеграцию, обработку и анализ в облачных условиях.

Сложности и трудности, связанные с управлением многочисленными облаками, особенно с защитой безопасности данных и обеспечением суверенитета данных между различными поставщиками облачных услуг, как подчеркивают Коломбо и др., подчеркивают сложную природу и проблемы, присущие администрированию мультиоблачных данных [10]. Успешное решение этих трудностей требует создания и реализации стратегий, которые эффективно интегрируют безопасность данных, соответствие требованиям, совместимость и удобство использования в различных облачных средах. Результаты этого исследования дают существенное представление о текущих препятствиях и способах их устранения в управлении облачными данными. Кроме того, они открывают потенциальные возможности для будущих исследований и развития. Чтобы в полной мере использовать возможности облачных вычислений в управлении данными, крайне важно устранить выявленные пробелы, разработать творческие решения для малоизученных трудностей и последовательно совершенствовать существующие методы в соответствии с новыми технологиями и потребностями. Взаимные отношения между проблемами и решениями, при которых одно информирует и улучшает другое, по-прежнему будут фундаментальным аспектом разработки и совершенствования методов управления данными в облачных системах.

6. ЗАКЛЮЧЕНИЕ

Это исследование углубляется в сложную сферу управления данными в облачных средах, изучая взаимосвязанные аспекты инноваций, проблем и непреходящих загадок. Расследование выявило важные дискуссии, происходящие в академической и промышленной сферах, подчеркнув впечатляющие технические достижения, которых мы достигли, и серьезные проблемы, которые еще предстоит решить.

Значение управления данными в облаке выходит за рамки простого хранения и извлечения данных, превращаясь в многогранную область, охватывающую несколько аспектов, таких как безопасность, распространение, дедупликация и многое другое. Вопрос

безопасности и конфиденциальности данных вызывает серьезную озабоченность в нескольких областях, подчеркивая необходимость защиты данных в быстро развивающуюся цифровую эпоху. Различные решения, включая методы шифрования, механизмы контроля доступа и технологии блокчейна, обеспечивают комплексный подход к решению этой проблемы. Для обеспечения полной безопасности данных требуется интеграция нескольких тактик.

Важная тема управления большими объемами данных, часто называемых "большими данными", и последующие разговоры о подходах, способствующих эффективному управлению и анализу данных, служат свидетельством нынешнего периода, когда данные обильны и являются важнейшим ресурсом. В исследовании представлены несколько подходов, тактик и решений, которые обеспечивают основу для будущих исследований, позволяя добиться прогресса и улучшений в управлении большими наборами данных.

Крайне важно признать, что путь к достижению эффективного управления данными в облачных системах постоянно развивается, характеризуется постоянным прогрессом и появлением новых препятствий и перспектив. Исследование освещает существующие дискуссии и решения в этой области, одновременно выявляя области недостатков и перспективные возможности для дальнейшего исследования и развития. Несмотря на присущую им сложность и значимость, некоторым трудностям в текущих дебатах уделяется относительно мало внимания, что создает возможности для дальнейших исследований и инноваций.

Включение устройств Интернета вещей (IoT) в облачные стратегии управления данными подчеркивает область, в которой процедуры должны постоянно пересматриваться и корректироваться для удовлетворения различных и постоянно меняющихся требований и препятствий, которые представляют данные IoT. Решения в этой области должны эффективно интегрировать требования к обработке данных в реальном времени, масштабируемости и безопасности для оптимизации использования данных Интернета вещей при соблюдении правил защиты данных.

Это исследование, основанное на строгой

методологии и обширном изучении текущих научных работ, служит фундаментальным вкладом, призванным способствовать дальнейшему пониманию и постоянному совершенствованию стратегий управления данными в системах облачных вычислений. Полученные в результате выводы предоставляют ценную перспективу для многих заинтересованных сторон, таких как практики, ученые и политики, позволяющие эффективно работать в динамичной сфере управления облачными данными. Эти идеи служат основой для формулирования политики, направляют прогресс и обеспечивают понимание возможных преимуществ и недостатков.

Представления об управлении данными в облачных вычислениях будут постоянно подвергаться пересмотру под влиянием технологических разработок, растущих препятствий и постоянного поиска инноваций и оптимизации. Это исследование предлагает краткий обзор нынешней ситуации, подчеркивая важность продолжающихся исследований, корректировок и сотрудничества в формировании будущих направлений управления данными в облачных вычислениях. Эти усилия имеют решающее значение для разработки инновационных, устойчивых и эффективных подходов к решению сложных проблем цифровой сферы.

ЛИТЕРАТУРА

1. Gao F et al. Management and Data Sharing of COVID-19 Pandemic Information. *Biopreservation and biobanking*, 2020, 18(6):570-580.
2. Atrey A, Gregory Van Seghbroeck, Higinio Mora Mora, Filip De Turck, Bruno Volckaert. Unifying Data and Replica Placement for Data-intensive Services in Geographically Distributed Clouds. Proc. 9th International Conference on Cloud Computing and Services Science (CLOSER), 2019, pp. 25-36. DOI: 10.5220/0007613400250036.
3. Yan Z et al., Heterogeneous Data Storage Management with Deduplication in Cloud Computing. *IEEE Transactions on Big Data*, 2019, 5:393-407.
4. Nzanywayingoma F, Yang Y. Efficient resource management techniques in cloud computing environment: a review and discussion. *International Journal of Computers and Applications*, 2018, 41:165-182.
5. Wagemann J et al. A user perspective on future cloud-based services for Big Earth data. *International Journal of Digital Earth*, 2021, 14:1758-1774.
6. Hayajneh SMA, Hamada M, Aljawarneh SA. Project Management Knowledge Areas and Skills for Managing Software and Cloud Projects: Overcoming Challenges. *Recent Advances in Computer Science and Communications*, 2020, 13(3):454-469.
7. Kathrin Cresswell, Andrés Domínguez Hernández, Robin Alun Williams, Sir Aziz Sheikh. Key Challenges and Opportunities for Cloud Technology in Health Care: Semistructured Interview Study. *JMIR Human Factors*, 2022, 9(1):e31246; doi: 10.2196/31246.
8. Poniszewska-Marańda A et al. A real-time service system in the cloud. *Journal of Ambient Intelligence and Humanized Computing*, 2019, 11:961-977.
9. Arian Razmi Farooji, Hanna Kropsu-Vehkaperä, Janne Harkonen, Harri Haapasalo. Advantages and potential challenges of data management in e-maintenance. *Journal of Quality in Maintenance Engineering*, 2019, 25(3):378-396.
10. Richards LL. Records management in the cloud: From system design to resource ownership. *Journal of the Association for Information Science and Technology*, 2018, 69(2):281-289. DOI: 10.1002/asi.23939.
11. Zhengbao Lv, Zefu Tan, Qian Wang, Yanqiong Yang. Cloud Computing Management Platform of Human Resource Based on Mobile Communication Technology. *Wireless Personal Communications*, 2018, 102(2):1293-1306. DOI: 10.1007/s11277-017-5195-y.
12. Chai M. Design of Rural Human Resource Management Platform Integrating IoT and Cloud Computing. *Computational Intelligence and Neuroscience*, 2022(C):1-9. DOI: 10.1155/2022/4133048.
13. Amr Mausad, Passent Elkafrawy, Amr Shawish, Mohamed Amin, Ismail M. Hagag. A New Secure Model for Data Protection over Cloud Computing. *Computational Intelligence and Neuroscience*, 2021, 5:1-11. DOI: 10.1155/2021/8113253.
14. Daniel Amo, David Fonseca, Marc Alier, Francisco José García-Peñalvo, María José Casañ. Personal Data Broker Instead of Blockchain

- for Students' Data Privacy Assurance. *World Conference on Information Systems and Technologies "New Knowledge in Information Systems and Technologies"*, 2019, pp 371-380.
15. Reed JL, Tosun AS. BULWARK: A Framework to Store IoT Data in User Accounts. *IEEE Access*, 2022, 10:15619-15634. DOI: 10.1109/ACCESS.2022.3144913.
 16. Raptis TP, Passarella A, Conti M. Data Management in Industry 4.0: State of the Art and Open Challenges. *IEEE Access*, 2019, 7:97052-97093.
 17. Latha K, Sheela T. Block based data security and data distribution on multi cloud environment. *Journal of Ambient Intelligence and Humanized Computing*, 2019:1-7.
 18. Latifian A. How does cloud computing help businesses to manage big data issues. *Kybernetes*, 2022, 51(6):1917-1948.
 19. Yu W, Chiou C-C. Effects of Sustainable Development of the Logistics Industry by Cloud Operational System. *Sustainability*, 2022, 14(16):10440.
 20. Lichtenthaler U. Data management efficiency: major opportunities for shared value innovation. *Management Research Review*, 2021, 45(2):156-172.
 21. Singh S, Misra SC. Exploring the Challenges for Adopting the Cloud PLM in Manufacturing Organizations. *IEEE Transactions on Engineering Management*, 2019, 68:752-766.
 22. Narayanan M, Cherukuri AK. Verification of Cloud Based Information Integration Architecture using Colored Petri Nets. *International Journal of Computer Network and Information Security*, 2018, 10:1-11.
 23. Khan MJ et al. Identifying Challenges for Clients in Adopting Sustainable Public Cloud Computing. *Sustainability*, 2022, 14(16):9809.
 24. Panwar A et al. A Blockchain Framework to Secure Personal Health Record (PHR) in IBM Cloud-Based Data Lake. *Computational Intelligence and Neuroscience*, 2022, 2022:3045107.
 25. Ali O et al. Assessment of Complexity in Cloud Computing Adoption: a Case Study of Local Governments in Australia. *Information Systems Frontiers*, 2021, 24:595-617.
 26. Pansara R. Master Data Management Challenges. *International Journal of Computer Science and Mobile Computing*, 2021, 10(10):47-49.
 27. Tcheunteu SWM, Metcheka LM, Ndoundam R. Distributed data hiding in a single cloud storage environment. *Journal of Cloud Computing*, 2021, 10:1-15.
 28. Yan J et al. An Efficient Organization Method for Large-Scale and Long Time-Series Remote Sensing Data in a Cloud Computing Environment. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2021, 14:9350-9363.
 29. Strauch S et al. Migrating Application Data to the Cloud using Cloud Data Patterns. *Proc. 3rd International Conference on Cloud Computing and Services Science*, 2013, pp36-46.

Намир Хашим Касим

PhD

Университет Джихан (CUSRC)

Сулеймания 46001, Курдистан, Ирак

E-mail: nameer.qasim@sulicihan.edu.krd

Наталия Боднар

PhD

Университетский колледж Аль-Рафиданн

Багдад 10064, Ирак

E-mail: natalia.bodnar@ruc.edu.iq

Хайдер Махмуд Салман

PhD

Университетский колледж Аль-Турат

Багдад 10013, Ирак

E-mail: haider.mahmood@turath.edu.iq

Салама Идрис Мустафа

PhD

Университетский колледж Аль-Нур

Ниневия 41012, Ирак

E-mail: salama.idris@alnoor.edu.iq

Фахер Рахим

PhD

Университет Джихана(CUSRC),

Сулеймания 46001, Ирак

E-mail: rahim.fakher@sulicihan.edu.krd.