

АНАЛИЗ ИНФОРМАЦИОННЫХ РЕШЕНИЙ ДЛЯ ГЛОБАЛЬНО РАСПРЕДЕЛЕННОЙ ОБРАБОТКИ КОРПОРАТИВНЫХ ДАННЫХ В СИСТЕМАХ РЕАЛЬНОГО ВРЕМЕНИ С ПРИМЕНЕНИЕМ ОБЛАЧНЫХ ТЕХНОЛОГИЙ

ANALYSIS OF INFORMATION SOLUTIONS FOR GLOBALLY DISTRIBUTED PROCESSING OF CORPORATE DATA IN REAL-TIME SYSTEMS USING CLOUD TECHNOLOGIES

D. Titov
O. Romashkova
S. Chiskidov

Summary. The article provides a comprehensive analysis of information solutions used for globally distributed processing of corporate data in real-time systems using cloud technologies. In the conditions of accelerating digital transformation and globalisation of business, companies are faced with the need to efficiently process large volumes of data coming from various geographically distributed sources. Modern solutions, including distributed databases, streaming data processing technologies, and cloud services from leading providers are considered. Special attention is paid to microservice architecture and containerisation as tools for developing flexible and scalable applications. Practical cases of successful implementation of these technologies in large corporations are presented, the results obtained and benefits for business are analysed. On the basis of the study, recommendations are formulated on the effective implementation and use of the technologies under consideration to improve the productivity and competitiveness of organisations in today's digital economy.

Keywords: globally distributed data processing, enterprise data, real-time systems, cloud technologies, distributed databases, streaming data processing, containerisation, microservice architecture, scalability.

Титов Дмитрий Андреевич

Аспирант, ГАОУ ВО города Москвы «Московский
городской педагогический университет (МГПУ)»
pointtitov@yandex.ru

Ромашкова Оксана Николаевна

Доктор технических наук, профессор, профессор,
Российская академия народного хозяйства
и государственной службы при Президенте РФ
(РАНХиГС)», г. Москва, Россия
ox-rom@yandex.ru

Чискидов Сергей Васильевич

Кандидат технических наук, доцент, профессор
ФГБВОУ ВО «Академия гражданской защиты МЧС России»
chis69@mail.ru

Аннотация. В статье проводится всесторонний анализ информационных решений, применяемых для глобально распределённой обработки корпоративных данных в системах реального времени с использованием облачных технологий. В условиях ускоряющейся цифровой трансформации и глобализации бизнеса, компании сталкиваются с необходимостью эффективной обработки больших объёмов данных, поступающих из различных географически распределённых источников. Рассмотрены современные решения, включая распределённые базы данных, технологии потоковой обработки данных, а также облачные сервисы ведущих провайдеров. Особое внимание уделено микросервисной архитектуре и контейнеризации как инструментам для разработки гибких и масштабируемых приложений. Приведены практические кейсы успешного внедрения этих технологий в крупных корпорациях, проанализированы полученные результаты и выгоды для бизнеса. На основе проведенного исследования сформулированы рекомендации по эффективному внедрению и использованию рассматриваемых технологий для повышения производительности и конкурентоспособности организаций в условиях современной цифровой экономики.

Ключевые слова: глобально распределённая обработка данных, корпоративные данные, системы реального времени, облачные технологии, распределённые базы данных, потоковая обработка данных, контейнеризация, микросервисная архитектура, масштабируемость.

Введение

В современную эпоху цифровой трансформации и глобализации организации сталкиваются с постоянно растущими объемами данных, которые требуют эффективной и своевременной обработки. Глобально распределенная обработка корпоративных данных становится критически важной для обеспечения конкурентоспособности и оперативности бизнес-процессов.

Это обусловлено необходимостью быстрого принятия решений на основе актуальной информации, доступной в режиме реального времени независимо от географического расположения подразделений компании.

Системы реального времени играют ключевую роль в этом процессе, обеспечивая непрерывный поток данных и возможность мгновенной реакции на изменения внешней и внутренней среды. Они позволяют организа-

циям мониторить операции в режиме онлайн, оптимизировать логистические цепочки, управлять рисками и улучшать качество обслуживания клиентов. В условиях высокой динамики рынка задержки в обработке данных могут приводить к упущенным возможностям и снижению эффективности бизнеса.

Облачные технологии выступают в качестве фундаментального инструмента для реализации глобально распределенных систем обработки данных. Они предоставляют масштабируемые и гибкие ресурсы, позволяющие обрабатывать большие объемы данных с высокой скоростью и надежностью. Кроме того, ИТ-решения, разработанные с применением облачных технологий, снижают затраты на инфраструктуру, обеспечивают доступность сервисов из любой точки мира, а также упрощают интеграцию различных информационных систем. Это особенно важно для корпоративных структур с разветвленной сетью филиалов и подразделений [1, 2].

Цель данной статьи — провести детальный анализ информационных решений, используемых для глобально распределенной обработки корпоративных данных в системах реального времени с применением облачных технологий. Для достижения поставленной цели необходимо решить следующие задачи:

1. Проанализировать современные распределенные базы данных и технологии обработки данных, такие как NoSQL и NewSQL системы, а также платформы потоковой обработки данных, включая Apache Kafka и Apache Flink. Анализ их характеристик позволит определить наиболее подходящие решения для различных корпоративных задач.
2. Исследовать модели предоставления облачных услуг (IaaS, PaaS, SaaS) и их применение в контексте корпоративных информационных систем.
3. Рассмотреть платформы ведущих облачных провайдеров, таких как AWS, Microsoft Azure и Google Cloud, и оценить их возможности для поддержки систем реального времени.
4. Проанализировать практические кейсы успешной реализации глобально распределенных систем обработки данных в крупных корпорациях. Это позволит выявить эффективные стратегии внедрения и использования рассматриваемых технологий, а также оценить полученные результаты и выгоды для бизнеса.

Проведение данного анализа имеет важное значение для понимания текущих тенденций и перспектив развития информационных технологий в корпоративной сфере. Результаты исследования будут полезны для ИТ-менеджеров, системных архитекторов и специалистов по данным, занимающихся разработкой и внедрением высокопроизводительных и масштабируемых систем обработки данных. Кроме того, статья способствует фор-

мированию рекомендаций по оптимизации существующих решений и повышению эффективности бизнес-процессов за счет использования передовых технологий.

Информационные технологии для глобально распределенной обработки данных в системах реального времени

Современные корпоративные структуры все чаще сталкиваются с необходимостью обработки огромных объемов данных, поступающих из различных географически распределенных источников. Для эффективного управления этими данными в режиме реального времени используются специализированные информационные технологии, которые обеспечивают высокую производительность, масштабируемость и надежность систем. В данном разделе рассматриваются ключевые технологии и решения, способствующие реализации глобально распределенной обработки данных в корпоративной среде с использованием облачных технологий.

Одной из фундаментальных основ для обработки больших объемов данных являются распределенные базы данных. Традиционные реляционные базы данных (SQL) не всегда справляются с требованиями современных приложений, особенно когда речь идет о горизонтальном масштабировании и обработке неструктурированных данных [3, 4]. В связи с этим широкое распространение получили NoSQL-системы, такие как Cassandra, MongoDB и HBase. Они обеспечивают высокую доступность и позволяют эффективно работать с распределенными данными, поддерживая модель eventual consistency, что приемлемо для многих приложений.

В дополнение к NoSQL-системам, развивается концепция NewSQL — класс реляционных систем управления базами данных, которые сочетают в себе преимущества реляционных моделей с возможностями горизонтального масштабирования NoSQL [5]. Примеры таких систем включают CockroachDB и Google Spanner. Они обеспечивают сильную согласованность данных и поддерживают транзакции ACID, что важно для критически важных корпоративных приложений.

Для обработки потоковых данных в реальном времени используются технологии потоковой обработки, такие как Apache Kafka и Apache Flink. Apache Kafka является распределенной платформой для передачи и хранения потоков данных, позволяя обрабатывать миллионы событий в секунду. Она используется для организации очередей сообщений и обеспечения высокой производительности и надежности передачи данных между сервисами. Apache Flink, в свою очередь, предоставляет возможности для распределенной обработки потоков данных и выполнения сложных аналитических

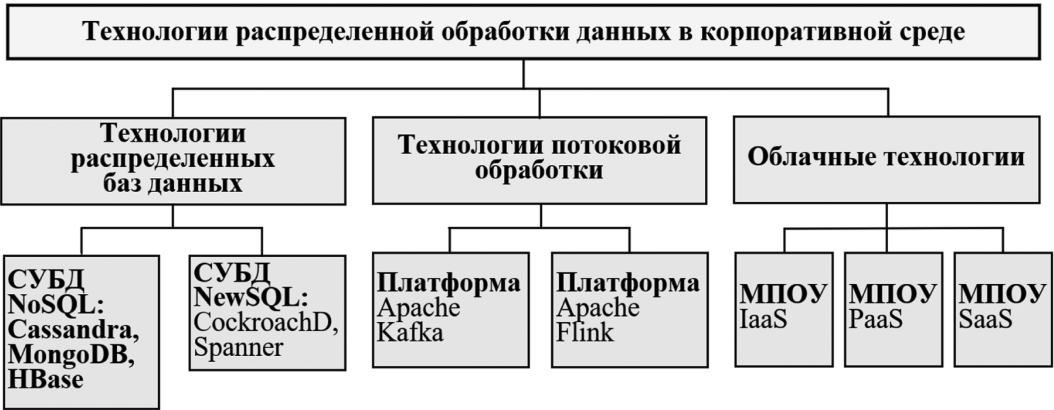


Рис. 1. Схема классификации технологий распределенной обработки данных в корпоративной среде

задач в режиме реального времени. Его архитектура позволяет обрабатывать данные с низкими задержками и высокой пропускной способностью [6, 7].

Облачные технологии играют ключевую роль в обеспечении инфраструктуры для глобально распределенной обработки данных. Модели предоставления облачных услуг (МПОУ) — IaaS (Infrastructure as a Service), PaaS (Platform as a Service) и SaaS (Software as a Service) — позволяют организациям выбирать наиболее подходящий уровень абстракции и управляемости. IaaS предоставляет виртуализированные вычислительные ресурсы, PaaS предлагает платформы для разработки и развертывания приложений, а SaaS обеспечивает доступ к готовым программным продуктам через интернет [8].

Возможная схема классификации технологий распределенной обработки данных в корпоративной среде изображена на рисунке 1.

Крупнейшие зарубежные облачные провайдеры, такие как Amazon Web Services (AWS), Microsoft Azure и Google Cloud Platform (GCP), предлагают широкий спектр сервисов для поддержки систем реального времени.

Например, AWS предоставляет сервисы Amazon Kinesis для сбора и обработки потоковых данных, Amazon DynamoDB как высокопроизводительную NoSQL базу данных, и AWS Lambda для безсерверных вычислений. Microsoft Azure предлагает аналогичные сервисы, включая Azure Event Hubs, Azure Cosmos DB и Azure Functions.

Таблица 1.

Возможности популярных сервисов и решений для поддержки систем реального времени от отечественных облачных провайдеров

№ п/п	Облачный провайдер	Название сервиса	Назначение сервиса	Зарубежный аналог
1	Яндекс.Облако	Яндекс.Cloud Functions	Серверлесс-архитектура, которая позволяет запускать код в ответ на события	Amazon CloudFront
2	Яндекс.Облако	Яндекс.Stream	Сервис для обработки потоковых данных	Amazon Kinesis
3	Mail.Ru Cloud Solutions	Mail.Ru Cloud	Платформа для развертывания приложений и сервисов, обрабатывающие данные в реальном времени	Amazon Kinesis
4	Ростелеком	Restream IIoT	Платформа для для сбора, первичной обработки и доставки данных с датчиков в ИС компании-клиента	Siemens MindSphere
5	Selectel	Облачные серверы и Kubernetes	Развертывание контейнеризированных приложений	Red Hat OpenShift, VMware Tanzu, SUSE Rancher
6	Cloud4Y	—	Платформы для развертывания приложений и сервисов, обрабатывающие данные в реальном времени	Microsoft Azure, Amazon Web Services, Google Cloud Platform
7	VK	VK Cloud		
8	СБЕР	SberCloud –		
9	МТС	Cloud MTS		
10	ВымпелКом	Beeline Cloud		

GCP предоставляет Google BigQuery для анализа больших данных и Google Pub/Sub для асинхронной передачи сообщений.

Перечень популярных сервисов и решения для поддержки систем реального времени, предлагаемых отечественными облачными провайдерами, приведены в таблице 1.

Микросервисная архитектура становится все более популярной в разработке корпоративных приложений, позволяя разделять сложные системы на небольшие, независимо разворачиваемые компоненты.

Контейнеризация с использованием Docker и оркестрация контейнеров с помощью Kubernetes упрощают развертывание и управление микросервисами в облачной среде [6, 9].

Это обеспечивает гибкость и ускоряет процесс разработки и обновления приложений, что особенно важно для систем, работающих в реальном времени.

Практические кейсы применения этих технологий демонстрируют их эффективность. Например, глобальные компании, такие как Netflix и Uber, активно используют микросервисную архитектуру и облачные технологии для обеспечения надежной и масштабируемой обработки данных в реальном времени. Netflix использует AWS для развертывания своей инфраструктуры, включая Amazon Kinesis и Amazon DynamoDB для обработки и хранения данных о пользовательской активности, что позволяет предоставлять персонализированные рекомендации и улучшать качество сервиса [10]. Uber применяет Apache Kafka и Apache Flink для обработки данных о поездках и оптимизации маршрутов в режиме реального времени, обеспечивая высокую производительность и низкие задержки.

Анализ полученных результатов и выгод от внедрения этих технологий показывает значительное улучшение показателей производительности и эффективности бизнес-процессов. Компании отмечают снижение затрат на инфраструктуру, повышение скорости обработки данных и улучшение качества сервисов для конечных пользователей. Использование облачных технологий позволяет организациям быстро масштабировать ресурсы в соответствии с текущими потребностями, избегая избыточных расходов и сложностей управления физической инфраструктурой [11].

В целом, интеграция распределенных баз данных, технологий потоковой обработки данных и облачных решений создает прочную основу для глобально распределенной обработки корпоративных данных в системах реального времени [12]. Это открывает новые

возможности для анализа больших данных, принятия обоснованных решений и повышения конкурентоспособности на рынке. Однако успешное внедрение этих технологий требует тщательного планирования, учета специфики бизнес-процессов и обеспечения безопасности и надежности систем.

Заключение

Проведенный анализ информационных решений для глобально распределенной обработки корпоративных данных в системах реального времени с использованием облачных технологий подтвердил их критическую важность для современного бизнеса. Рост объемов данных и необходимость их оперативной обработки в глобальном масштабе требуют от компаний внедрения передовых технологических решений, способных обеспечить высокую производительность, масштабируемость и надежность.

Выводы по результатам анализа свидетельствуют о том, что использование распределенных баз данных, таких как NoSQL и NewSQL системы, позволяет эффективно управлять большими объемами структурированных и неструктурированных данных. Технологии потоковой обработки данных, представленные платформами Apache Kafka и Apache Flink, обеспечивают возможность обработки событий и данных в режиме реального времени с минимальными задержками, что является критически важным для принятия оперативных бизнес-решений.

Облачные технологии предоставляют необходимую инфраструктуру и сервисы для реализации таких систем. Модели предоставления облачных услуг (IaaS, PaaS, SaaS) и решения ведущих провайдеров (AWS, Azure, Google Cloud) позволяют компаниям быстро масштабировать ресурсы, снижать затраты и фокусироваться на разработке и совершенствовании бизнес-приложений. Микросервисная архитектура и контейнеризация с использованием Docker и Kubernetes упрощают процесс разработки, развертывания и управления приложениями, повышая их гибкость и устойчивость к нагрузкам.

Рекомендации для внедрения и использования рассматриваемых технологий включают следующие аспекты. Во-первых, компаниям следует провести тщательный анализ своих бизнес-процессов и требований к обработке данных, чтобы выбрать наиболее подходящие технологии и инструменты. Во-вторых, важно обеспечить профессиональное обучение персонала и привлечение экспертов для успешного внедрения новых систем. В-третьих, особое внимание необходимо уделить вопросам безопасности и защиты данных, включая разработку стратегий резервного копирования, восстановления после сбоев и соблюдения нормативных требований.

Практические кейсы успешной реализации рассмотренных технологий демонстрируют значительные преимущества, такие как повышение скорости обработки данных, улучшение качества сервисов и увеличение конкурентоспособности. Поэтому компаниям, стремящимся к лидерству в своих отраслях, рекомендуется активно внедрять и адаптировать эти технологии в соответствии с собственными потребностями и стратегическими целями.

В заключении можно сказать, что будущее отрасли отражает уверенность в том, что дальнейшее развитие информационных технологий будет еще более тесно связано с облачными решениями и системами реального времени. Ожидается, что интеграция с технологиями искусственного интеллекта и машинного обучения позволит создать еще более интеллектуальные и адап-

тивные системы обработки данных. Развитие сетей 5G и концепции Edge Computing расширит возможности обработки данных на периферии сети, снижая задержки и повышая эффективность.

В условиях стремительного развития технологий и роста требований к обработке данных компании должны быть готовы к постоянным изменениям и инновациям. Инвестиции в современные информационные технологии и облачные решения станут ключевым фактором успеха и устойчивого развития в цифровой экономике. Внедрение глобально распределенных систем обработки данных в реальном времени не только удовлетворяет текущие потребности бизнеса, но и закладывает фундамент для будущего роста и адаптации к новым вызовам рынка.

ЛИТЕРАТУРА

1. Затирко О.К., Пынько Л.Е. Облачные технологии и их применение в сфере экономики и информационных технологий: достоинства и недостатки // В сборнике: Актуальные проблемы развития ЕАЭС в условиях современных глобальных изменений. Материалы первой Всероссийской (национальной) научно-практической конференции. Под редакцией И.В. Цвигун, В.С. Колодина. 2019. С. 249–253.
2. Громей Д.Д., Лебеденко Е.В. Математическое обеспечение поддержки процесса управления схемой реляционной базы данных в задачах горизонтального масштабирования // Моделирование, оптимизация и информационные технологии. 2019. Т. 7. № 2 (25). С. 65–79.
3. Кондрусева С.А. Технологии сбора и обработки больших массивов информации // В сборнике: Прикладная математика и информатика: современные исследования в области естественных и технических наук. Материалы VI Международной научно-практической конференции (школы-семинара) молодых ученых. 2020. С. 574–578.
4. Самарев Р.С. Обзор состояния области потоковой обработки данных. Труды Института системного программирования РАН. 2017. Т. 29. № 1. С. 231–260.
5. Латыпова Э.М., Кононов Н.А. К вопросу моделей облачных услуг SaaS, PaaS и IaaS: особенности и возможности // В сборнике: Наука, общество, образование в условиях цифровизации и глобальных изменений. Материалы III Международной научно-практической конференции. 2022. С. 7–9.
6. Манойло В.Е. Разработка инфраструктуры приложений на основе Docker и Kubernetes // Научно-технические инновации и веб-технологии. 2023. № 1. С. 17–20.
7. Горелов Г.В., Ромашкова О.Н., Чан Т.А. Качество управления речевым трафиком в телекоммуникационных сетях /Москва, Радио и связь, 2001. С.132
8. Ромашкова О.Н., Ломовцев Р.С., Пономарева Л.А. Компьютерная поддержка принятия управленческих решений для образовательной системы регионального уровня //Вестник Рязанского государственного радиотехнического университета. 2019. № 67. С. 50–58.
9. Наумов М.А., Ромашкова О.Н., Ермакова Т.Н., Чискидов С.В. Модель управления корпоративными вычислительными ресурсами образовательного комплекса// Современная наука: актуальные проблемы теории и практики. Серия: Естественные и технические науки. 2021. № 4-2. С. 40–47.
10. Заболотникова В.С., Ромашкова О.Н. Анализ методов кластеризации для эффективного управления процессами в налоговой службе // Фундаментальные исследования. 2017. № 9-2. С. 303–307.
11. Рябовичева О.В., Ромашкова О.Н., Ермакова Т.Н., Чискидов С.В. Процесс обработки и передачи виртуальных данных в вычислительных комплексах и компьютерных сетях вуза //Современная наука: актуальные проблемы теории и практики. Серия: Естественные и технические науки. 2022. № 7–2. С. 85–92.
12. Петрова А.М., Ромашкова О.Н., Ермакова Т.Н., Чискидов С.В. Модели процессов функционирования информационной системы мониторинга климата и окружающей среды в арктическом регионе // Современная наука: актуальные проблемы теории и практики. Серия: Естественные и технические науки. 2022. № 4-2. С. 104–110.

© Титов Дмитрий Андреевич (pointtitov@yandex.ru); Ромашкова Оксана Николаевна (ox-rom@yandex.ru);

Чискидов Сергей Васильевич (chis69@mail.ru)

Журнал «Современная наука: актуальные проблемы теории и практики»