

СРАВНИТЕЛЬНЫЙ АНАЛИЗ АРХИТЕКТУР ОБЛАЧНЫХ СИСТЕМ, ОПЕРИРУЮЩИХ BIG DATA

Е.В. Чернышова¹, В.В. Кондусова¹, Г.А. Спесивцев¹

¹ФГБОУ ВО «Воронежский государственный лесотехнический университет имени Г.Ф. Морозова»

Аннотация. В статье приводится сравнительный анализ архитектур облачных систем, оперирующих Big Data. Архитектура облачных систем, работающих с Big Data, это комплексная инфраструктура, которая представляет собой систему хранения, обработки и работы с большими объемами данных их анализа в облачной среде. Анализ архитектур, оперирующих Big Data, позволяет выявить их недостатки и плюсы, а также особенности их работы.

Ключевые слова: распределенная архитектура, сервер-клиентская архитектура, Big Data, облачные системы.

COMPARATIVE ANALYSIS OF CLOUD SYSTEMS ARCHITECTURES OPERATING BIG DATA

E.V. Chernyshova¹, V.V. Kondusova¹, G.A. Spesivtsev¹

¹Voronezh State University of Forestry and Technologies named after G.F. Morozov

Abstract. The article provides a comparative analysis of the architecture of cloud systems operating big data. The architecture of cloud systems working with Big Data is a complex infrastructure, which is a system for storing, processing and working with large volumes of analysis data in a cloud environment. Analysis of the architectures that operate Big Data allows us to identify their disadvantages and advantages, as well as the features of their work.

Keywords: distributed architecture, server-client architecture, Big Data, cloud systems.

Распределенная архитектура считается одной из основных архитектурных моделей в работе с Big Data в облачной системе. Данная архитектура легко масштабируема и производительна это достигается за счет того, что в этой модели данные хранятся и обрабатываются на нескольких серверах или узлах данные

разбиваются на части что обеспечивает анализ больших объёмов данных и параллельную обработку. Примерами распределенных облачных систем для Big Data являются Apache Hadoop, Apache Spark и Google Cloud Dataflow. Данная системы предоставляют инструменты обработки анализа данных и хранения данных в облаке используя для этого распределенную архитектуру и с ее помощью достигают масштабируемости и производительности системы.

К основным характеристикам распределенной архитектуры Big Data, а облачной системе относят:

1. Горизонтальное масштабирование: это возможность увеличения производительности, а также обработки данных благодаря возможности добавления новых узлов.
2. Отказоустойчивость: если один из узлов выдает ошибку и перестает работать данные сохраняются автоматически, и обработка проходит без прерываний.
3. Параллельная обработка: одновременная обработка данных на нескольких узлах ускоряет процесс анализа данных.
4. Распределенное хранение: для обеспечения баланса доступности и нагрузки данные распределяются автоматически.
5. Управление данными: выполнения задач на различных узлах и распределение данных автоматически управляется системой.

Также рассмотрим недостатки распределенной архитектуры Big Data в облаке:

1. Сложность управления: управление данной архитектурой требует наличия специальных знаний, а также навыков что вызывает трудности для части организации.
2. Затраты на инфраструктуру: развертывание обслуживание и поддержка распределенной архитектуры требует значительных затрат на обслуживание серверов и инфраструктуру.
3. Безопасность: распределенная архитектура требует увеличения количества сервера, а это влияет на риск утечки данных или несанкционированного доступа к информации и это требует дополнительных способов сохранения безопасности.
4. Согласованность данных: между различными узлами распределённой архитектуры возникают проблемы согласованности данных что приводит к ошибкам в анализе.

Распределенная архитектура Big Data в облаке в целом является эффективным решением для работы с большими объемами данных, однако стоит отметить, что для успешной реализации стоит учитывать ее особенности и недостатки, чтобы избежать рисков и обеспечивать эффективный анализ данных.

Другой популярной архитектурой облачных систем, оперирующих Big Data, является сервер-клиентская архитектура. В данной архитектуре данные хранятся на сервере централизованно, а приложения клиентов получают доступ к данным через интерфейс или с помощью API. Примерами таких облачных систем являются Amazon Web Services (AWS) и Microsoft Azure.

Сейчас идет постоянный рост объемов данных и сильной необходимостью эффективного и быстрого их анализа и за этого клиента серверная архитектура все больше набирает популярность среди решений для предприятий. Данная модель является системой, где с клиентскими устройствами через облачную инфраструктуру взаимодействуют серверы, хранящие данные и обрабатывающие их, обеспечивая возможность доступа к данным и анализу большого объема информации в реальном времени.

Преимущества сервер-клиентской архитектуры Big Data в облаке:

1. Масштабируемость: По мере необходимости серверы в облаке могут быть масштабируемы горизонтально и вертикально, что обеспечит увеличение объема хранимых данных и производительность.
2. Отказоустойчивость: обеспечение с помощью облачных серверов высокой доступности данных и при отказе одного сервера перенаправление нагрузки на другие узлы, уменьшая риск потери информации.
3. Гибкость: для обеспечения гибкости в работе с информацией устройства клиентов имеют доступ к данным из любой точки мира через интернет.
4. Быстродействие: для ускорения и оптимизации бизнес-процессов обработка и анализ данных больших объемов информации происходят на облачных серверах.

Рассмотрим недостатки сервер-клиентской архитектуры Big Data в облаке:

1. Высокие затраты на обслуживание: для организации с ограниченным бюджетом это может быть серьезной проблемой, так как поддержание и обновление серверов требует значительных финансовых затрат, как на техническую поддержку, так и на инфраструктуру.
2. Одиночная точка отказа: при выходе из строя сервера, на котором хранятся данные, это приводит к недоступности информации для приложений.

клиентов, а отсутствие резервного копирования повлечёт за собой потерю ценной информации.

3. Низкая отказоустойчивость: если происходит сбой в сети или отказ сервера временно может быть потерян доступ к данным что приведет к задержке работы с информацией и данными.

4. Безопасность данных: сервер-клиентская архитектура подразумевает хранение всей информации на централизованных серверах что увеличивает риск утечки информации и несанкционированного доступа к ней. Из-за этого приходится принимать дополнительные меры для обеспечения безопасности данных на серверах.

В целом, сервер-клиентская архитектура, как и распределенная имеет свои плюсы и минусы, которые стоит учитывать при внедрении и разработки системы обработки и внедрения данных. Можно сказать, что сервер-клиентская архитектура — это гибкое и эффективное решение для работы с большими данными которая обеспечивает должную производительность. Из чего следует что организациям проще и быстрее адаптироваться к меняющемуся потребностям рынка что обеспечит эффективность бизнес-процессов.

Сравнивая две архитектурные модели, которые мы рассмотрели выше это распределенная архитектура и сервер-клиентской архитектуры, можно выделить следующие преимущества и недостатки каждой из них:

1. Распределенная архитектура:

Преимущества: высокая производительность, горизонтальное масштабирование, отказоустойчивость, пероральная обработка, распределённое хранение, управление данными.

Недостатки: сложность настройки и управления, требовательность к ресурсам, затраты на инфраструктуру, безопасность.

2. Сервер-клиентская архитектура:

Преимущества: простота использования, централизованное управление данными, гибкость и быстродействие.

Недостатки: ограничения по производительности и масштабируемости, высокие затраты на обслуживание, низкая отказоустойчивость, одиночная точка отказа.

В зависимости от конкретных задач организации или компании и потребностей можно выбрать оптимально подходящую архитектуру для работы с Big Data в облаке. При выборе подходящей системы нужно учитывать много факторов требования к производительности системы также к безопасности данных

масштабируемости системы и другие важные факторы. Таким образом, проведенный сравнительный анализ архитектур распределенной архитектуры и сервера-клиентской архитектуры позволяет выбрать самый подходящий вариант для конкретных потребностей и задач организации. Каждая из рассмотренных моделей имеет свои недостатки и преимущества, которые нужно учитывать при выборе системы для работы с Big Data в облаке.

Список литературы

1. Шипилова Е.А., Некрылов Е.Е., Курченкова Т.В. Анализ и моделирование траекторий поведения пользователей онлайн-сервисов с использованием платформы RETENTIONEERING // Моделирование систем и процессов. – 2022. – Т. 15, № 3. – С. 82-93.
2. Strength test of the industrial building's load-bearing structures / Sazonova S.A., Nikolenko S.D., Zyazina T.V., Chernyshova E.V., Kazbanova I.M. // Journal of Physics: Conference Series. III International Conference on Metrological Support of Innovative Technologies (ICMSIT-III-2022). Krasnoyarsk. – 2022. – С. 22016
3. Карпенко, А. Облачные архитектуры: разработка устойчивых и экономичных облачных приложений. (2023): 50-53.
4. Новикова Т. П. Управление данными: лабораторный практикум / Т. П. Новикова. – Воронеж, 2022. – 106 с.
5. Вайгенд, А. BIG DATA. Вся технология в одной книге". (2019): 14-23.
6. Су, К. Теоретический минимум по Big Data. Всё что нужно знать о больших данных. (2020): 33-43.
7. Разработка специального программного обеспечения стеганографического скрывает информации в аудиофайлах / Жуматий В.П., Денисенко Д.И., Чернышова Е.В. // Информатика: проблемы, методы, технологии. Материалы XX Международной научно-методической конференции ; под ред. А.А. Зацаринного, Д.Н. Борисова. – 2020. – С. 1022-1031.
8. Полуэктов А.В., Макаренко Ф.В., Ягодкин А.С. Использование сторонних библиотек при написании программ для обработки статистических данных // Моделирование систем и процессов. – 2022. – Т. 15, № 2. – С. 33-41.

References

1. Shipilova E.A., Nekrylov E.E., Kurchenkova T.V. Analysis and modeling of behavior trajectories of users of online services using the RETENTIONEERING platform // Modeling of systems and processes. – 2022. – T. 15, No. 3. – P. 82-93.
2. Strength test of the industrial building's load-bearing structures / Sazonova S.A., Nikolenko S.D., Zyazina T.V., Chernyshova E.V., Kazbanova I.M. // Journal of Physics: Conference Series. III International Conference on Metrological Support of Innovative Technologies (ICMSIT-III-2022). Krasnoyarsk. – 2022. – C. 22016.
3. Karpenko, A. Cloud Architectures: Developing Resilient and Cost-Effective Cloud Applications. (2023): 50-53.
4. Novikova T. P. Data management: laboratory workshop / T. P. Novikova. – Voronezh, 2022. – 106 p.
5. Weigend, A. BIG DATA. All technology in one book. (2019): 14-23.
6. Su, K. Theoretical minimum for Big Data. Everything you need to know about Big Data. (2020): 33-43.
7. Development of special software for steganographic hiding of information in audio files / Zhumatiy V.P., Denisenko D.I., Chernyshova E.V. // Computer science: problems, methods, technologies. Materials of the XX International Scientific and Methodological Conference. Edited by A.A. Zatsarinny, D.N. Borisova. – 2020. – P. 1022-1031.
8. Poluektov A.V., Makarenko F.V., Yagodkin A.S. The use of third-party libraries when writing programs for processing statistical data // Modeling of systems and processes. - 2022. – Vol. 15, No. 2. – pp. 33-41.