

УДК: 004.6

В.М. Абас, канд. техн. наук (Университет Диалы, Ирак);
Н.А. Ахмед, (Университет Аль-Иракия – г. Багдад, Ирак);
С.Я. Егоров, проф. (ТГТУ, г. Тамбов, Россия)

ВНЕДРЕНИЕ МАСШТАБИРУЕМЫХ РЕШЕНИЙ ОБРАБОТКИ БОЛЬШИХ ДАННЫХ В РЕАЛЬНОМ ВРЕМЕНИ

Аннотация. В этой статье мы представляем ориентированную на поиск распределенную веб-сервисную систему для быстрого поиска документов, которая позволяет сервисам интеллектуального анализа данных генерировать сглаженные результаты практически в реальном времени на основе технологий коммерческих поисковых систем с быстрым поиском.

1. Введение. Решения для обработки больших данных (Big Data) в реальном времени, способные извлекать полезные закономерности и выводы, являются критически важными для принятия решений практически во всех сферах человеческой деятельности [1]. К ним относятся научные открытия, создание новых знаний, поддержка принятия решений, предотвращение преступлений, здравоохранение и национальная безопасность. Основная цель исследований, ориентированных на данные, – извлечение высококачественной информации из данных [2]. В этой статье мы обсуждаем реализацию масштабируемых решений для интеллектуального анализа данных при обработке больших данных в реальном времени в облачных экосистемах. В связи с постоянным ростом объема генерируемых и хранимых данных сегодня актуальной задачей является хранение этих данных, создание решений для интеллектуального анализа данных, способных генерировать новые алгоритмы, и реализация средств, способных работать с этими новыми решениями для хранения данных [3].

2. Основная часть. Работа с Big Data – это анализ больших объемов данных с помощью специальных технологий, которые позволяют обрабатывать и анализировать данные быстро и эффективно. Существует несколько методов обработки больших данных, таких как облачные вычисления, параллельная и распределенная обработка на сотнях или тысячах серверов в центрах обработки данных и потоковая передача бизнес-информации в режиме реального времени. Облачная инфраструктура для обработки больших данных в режиме реального времени содержит два ключевых компонента, которые вместе обеспечивают обработку и анализ потоковых данных в режиме реального времени. Первый компонент представляет новую архитектуру потоковой

обработки, называемую распределенными средствами потоковой обработки, и примеры ускорителей на основе сетки для обеспечения эффективного и масштабируемого анализа в режиме реального времени запросов данных типа map-reduce. Второй компонент представляет два средства массовой параллельной обработки, называемые GPU (графический процессор) и FPGA (программируемая пользователем вентильная матрица), для обеспечения обработки большого количества простых запросов на интеллектуальный анализ данных в режиме реального времени с очень низкой задержкой [4].

Облачная инфраструктура для обработки больших данных в режиме реального времени добавляет реактивность в процесс, который в противном случае является статическим анализом пакетных данных. Система должна учитывать растущий спрос на возможности обработки и анализа больших данных в реальном времени, включая быстрые результаты, такие как «что происходит прямо сейчас», и более ориентированные на тенденции «что будет дальше». Данные поступают как из традиционных каналов, так и из социальных сетей, будь то данные финансовых транзакций, записи данных вызовов в режиме, близком к реальному времени, или потоковые данные о кликах и присутствии. В условиях открывающейся бизнес-возможности обнадеживает то, что с точки зрения технологий такой анализ существим; инфраструктура для использования такого анализа существует и используется в отрасли; и, наконец, в реальности можно принимать обоснованные решения по событию, руководствуясь тем, что происходит прямо сейчас [5]. Текст представляет собой масштабируемую в реальном времени облачную инфраструктуру для потокового анализа больших данных с возможностью расширенной обработки данных и бизнес-аналитики с помощью облачных технологий.

Понимая ограничения и сложность тестирования приложений в реальном времени в облачных инфраструктурах в больших масштабах, мы разработали имитационную модель для понимания поведения представленного алгоритма на практике. В этом сценарии мы решили провести моделирование на распределенном отказоустойчивом движке обработки данных, который работает в производстве уже много лет и дает нам ценный опыт создания систем обработки данных в реальном времени. Мы стремились продемонстрировать, что предлагаемое решение проблем обработки больших данных в реальном времени может бесперебойно работать в облаке. Кроме того, мы смогли справиться с ситуацией, когда решение должно быть реализовано и интегрировано в платформу, где приложения ограничены работой в реальном времени.

При внедрении решений по обработке данных в рамках масштабируемой обработки больших данных для интеллектуального анализа данных следует учитывать несколько факторов. Это в основном проблемы приложений, обрабатывающих данные в оперативной памяти (*in-memory*), такие как высокая скорость передачи данных, которая может создать потенциальные ограничения для пропускной способности передачи данных, топология сетей поставщика облачных услуг, память, которая имеет первоочередное значение из-за ограничений, и необходимость сохранять состояние больших данных, например, когда данные должны оставаться неизменными в течение фиксированного срока.

Приложения по обработке больших данных должны быть эффективно интегрированы в традиционные ИТ-системы для повышения общего качества. Кроме того, использование общедоступных облачных сервисов может создать и другие трудности, например, связанные с обеспечением информационной безопасности log-файлов. Необходимо также учитывать различные методологии обеспечения соответствия и аудита. Распределение затрат, непосредственно связанных с обработкой данных в облачных сервисах, – еще один влиятельный фактор, который необходимо учитывать лицам, принимающим решения, для реализации оптимальной стратегии.

Во многих случаях существующие высокопроизводительные решения по интеллектуальному анализу данных были разработаны в основном для работы с пакетами больших данных с использованием обычных сетей и баз данных. Хотя мы не всегда можем полностью заменить пакетные решения обработки на потоковые или *in-memory* решения, совместное применение потоковой, пакетной или *in-memory* аналитики больших объемов данных может дать более значимые и полезные практические результаты в реальном времени. Сами по себе потоковые решения данных бесполезны, если они не включены в более широкий контекст, который может включать хранение данных для постоянных частей потока или даже пакетный анализ данных, который может учитывать периодичность, проявляемую в зрелых данных необработанных потоков данных.

3. Результат и анализ. В этом исследовании реализованы масштабируемые решения по интеллектуальному анализу данных для служб обработки больших данных в реальном времени в облачной экосистеме. В частности, мы предложили использовать распределенный брокер сообщений и крупномасштабную систему обработки для обработки больших данных в реальном времени на основе архитектуры Lambda. На основе интеллектуального анализа данных с большими

данными в реальном времени мы реализовали ряд ценных тематических исследований для оценки производительности среды обработки больших данных, которую мы построили, чтобы продемонстрировать осуществимость наших предлагаемых подходов. В этом исследовании мы предложили интегрировать распределенный брокер сообщений и крупномасштабную систему обработки, а также параллельно извлекать хранящиеся данные для выполнения набора задач обработки данных в другой топологии. Благодаря экспериментам, в ходе которых собирались миллионы log-файлов GPS, эти большие данные предоставили значительные результаты интеллектуального анализа данных в реальном времени для крупных прикладных сценариев, таких как службы на основе определения местоположения, транспортные системы или дорожная связь. В будущем мы будем и дальше использовать в наших алгоритмах передовые модели интеллектуального анализа данных, такие как обнаружение аномалий во множественных временных рядах и параллельный анализ затрат на объекты, прогнозирование скорости прибытия и онлайн-изменение затрат для динамического обновления элементов.

4. Заключение. В этой статье мы представили конкретную роль облачных вычислений как практической экосистемы в поддержке внедрения масштабируемых решений по интеллектуальному анализу данных для аналитики больших данных в реальном времени. Мы обсудили и объяснили сложные архитектуры, необходимые для внедрения таких сервисов в облаке для эффективного интеллектуального анализа больших данных, и динамические проблемы, которые были вызваны внедрением предлагаемых сервисов.

ЛИТЕРАТУРА

1. M.K. Gupta, P. Chandra. A comprehensive survey of data mining // International Journal of Information Technology. – 2020. – V. 12. – PP. 1243–1257. – URL: <https://link.springer.com/article/10.1007/s41870-020-00427-7>.
2. J. Chen, L. Ramanathan, M. Alazab. Holistic big data integrated artificial intelligent modeling to improve privacy and security in data management of smart cities // Microprocessors and Microsystems. – 2021. – V. 81. – Article 103722. – URL: <https://www.sciencedirect.com/science/article/pii/S014193312030867X>.
3. Y. Liu, W. Yu, W. Rahayu, T. Dillon. An evaluative study on IoT ecosystem for smart predictive maintenance (IoT-SPM) in manufacturing: Multiview requirements and data quality // IEEE Internet of Things Journal. – 2023. – V. 10. – No. 13. – PP. 60–84. – URL: <https://ieeexplore.ieee.org/abstract/document/10049619/>.

4. J. Wang, Y. Yang, T. Wang, R.S. Sherratt. Big data service architecture: a survey // Journal of Internet. – 2020. – V. 21. – №. 2. – PP. 393–405. – URL: <https://jit.ndhu.edu.tw/article/viewFile/2261/2274>.

5. E. Mehmood, T. Anees. Challenges and solutions for processing real-time big data stream: a systematic literature review // IEEE Access. – 2020. – V. 8. – PP. 123–143. – URL: <https://ieeexplore.ieee.org/iel7/6287639/6514899/09126812.pdf>.

УДК 681.178.9

Н.В. Гурько, студ.; П.А. Лазарева, доц.
(КНИТУ-КАИ им. А. Н. Туполева, г. Казань, Россия)

СИСТЕМА БЕСКОНТАКТНОГО УПРАВЛЕНИЯ ДИСПЛЕЕМ НА ОСНОВЕ КАМЕРЫ ГЛУБИННОГО ИЗОБРАЖЕНИЯ

В современном обществе всё большую популярность приобретают системы, использующие в основе принципа своей работы методы обработки изображений при помощи нейросетей. Частным случаем подобных систем считается определение положения рук на снимке, дающее возможность частично контролировать функционал любого устройства за счёт заранее указанных жестов. На основе подобного определения, при использовании камеры глубины можно реализовать механизм, схожий с привычным управлением любого сенсорного устройства, но без необходимости в касаниях.

Камеры глубины (depth cameras) – разновидность камер, позволяющих определить расстояние от устройства до наблюдаемых объектов сцены и сопоставить карту глубины с цветным RGB изображением (перенос точки из depth изображения на RGB). Реализация системы была осуществлена при помощи сенсора Kinect v1 и библиотеки OpenNI2, использовавшейся для получения и передачи данных с камеры, на языке программирования Python. Данная камера уже использовалась для создания бесконтактного управления консолями Xbox, однако механизм имел ряд недостатков, в том числе необходимость нахождения человека в полный рост на расстоянии не менее 3м от самой камеры и невозможность удобного использования её вне экосистемы самой игровой консоли.

Распознавание положения рук и отслеживание движений было реализовано с использованием библиотек OpenCV и MediaPipe, дающих возможность получить “скелет” руки через опорные точки, даже если часть точек будет закрыта или спрятана (рис. 1).