

«Сейфуллин оқулары – 18: « Жастар және ғылым – болашаққа көзқарас» халықаралық ғылыми -практикалық конференция материалдары = Материалы международной научно-практической конференции «Сейфуллинские чтения – 18: « Молодежь и наука – взгляд в будущее» - 2022.- Т.І, Ч.IV. - С.12-14

ИССЛЕДОВАНИЕ АРХИТЕКТУР ПОСТРОЕНИЯ И АЛГОРИТМОВ РАБОТЫ СИСТЕМ ХРАНЕНИЯ И АНАЛИЗА БОЛЬШИХ ДАННЫХ

*Буранбаев А.А., магистрант 2 курса
Казахский агротехнический университет им С. Сейфуллина, г.Нур-Султан*

Введение

Каждое устройство, имеющее доступ в интернет, является потенциальным источником информации в мире. Пользователи этих устройств генерируют разнообразные данные: личные, покупки, переходы по страницам сайта, также поступает информация о местоположении и передвижении, состояния климата, медицинских данных, предпочтений пользователей, проведения досуга. Эта информация обширна и создается с большой скоростью. В тоже время информация поступает с объектов инфраструктуры, разнообразных машин и устройств. Количество цифровой информации увеличивается быстро с увеличением устройств с доступом в интернет [1].

Со времени изобретения книгопечатания вся информация хранилась в основном на бумажных носителях. Во второй половине XX века начались появляться компьютеры, которые в настоящее время привели к новой парадигме работе с информацией. Мы живем в эпоху данных или информации. Для следующих целей или преимуществ данные являются ключевым фактором для всех организаций:

- Данные полезны при принятии решений.
- Знание ожиданий потребителей, чтобы помочь бизнесу стать лучше.
- Получать правильную информацию о рынке
- Оптимизация систем с помощью анализа данных.
- Данные эффективно увеличивают рыночную стоимость.
- Сбор данных предлагает информацию о предпочтениях и антипатиях [2].

Термин “большие данные” может означать всякий набор данных, который обрабатывать традиционными способами сложно или почти невозможно из-за их количества или сложности. Под традиционными способами можно представить РСУБД — реляционные системы управления базы данных, которые распространены повсеместно и считался универсальным инструментом, до появления “больших данных”.

Долгое время реляционные системы управления базами данных (РСУБД) считались универсальным решением, но экспоненциально растущие объемы, скорость и неоднородность данных показали непригодность РСУБД для использования в системах хранения и анализа больших данных. Наука о данных предполагает использование методов для анализа огромных объемов данных и извлечения содержащихся в них знаний.

Исследование архитектур построения и алгоритмов работы систем хранения

Феномен больших данных тесно связан с появлением науки о данных - дисциплины, которая сочетает в себе математику, программирование и научный инстинкт. Сейчас же Data science и большие данные разделились на разные дисциплины [3]. Получение выгоды от больших данных означает инвестирование в команды с этим набором навыков и окружение их организационной готовностью понимать и использовать данные для получения преимущества.

Современному человеку не спрятаться от больших данных. Отдельный человек физически не способен и никак не может успеть осмыслить процессы, которые происходят в информационном поле, в котором он находится. Современные системы позволяют довольно легко собирать некоторые данные в огромных масштабах и обрабатывать их. И этим во все возрастающей степени пользуются различные организации. Раньше это было невозможно. Но теперь есть Интернет [4].

Эти данные, как правило, накапливаются в мире, на предприятиях в результате работы приложений, аппаратуры и внешних источников. Невзирая на то, что появление больших данных сформировало новую дисциплину, но изначально они развивались в течение многих лет. Большие данные добавляют новые методы к традиционным аналитическим подходам, основанным на статистике, они усилены вычислительными ресурсами и подходами к выполнению аналитических алгоритмов. Анализ больших данных является слиянием таких дисциплин как: математика, статистика, компьютерная наука и специальные знания в конкретной предметной области. Выгоды и применимость при использовании больших данных, несомненно, очень широкие [5].

Для того, чтобы получить эти выгоды необходимо исследовать архитектуру построения и алгоритмы работы систем хранения и анализа больших данных, чтобы эффективно и быстро обрабатывать входящий поток данных. И на основе этого разрабатывать собственные решения для анализа и хранения больших данных.

Чтобы добиться этого необходимо изучить существующие системы для анализа больших данных, выявить их преимущества и недостатки, для формирования стандартов рынка.

Проанализировать процесс анализа и хранения больших данных существующих систем, чтобы найти оптимальный процесс с максимально возможным выходом. Проанализировать подходы к построению систем

хранения и анализа больших данных, чтобы выделить ключевые особенности в их построении[6].

Значимость исследования систем хранения и алгоритмов работы больших данных заключается в возможности практического применения апробированных технологий хранения и анализа больших данных. Выделив ключевые особенности и процесс построения, необходимо разработать архитектуру, которая будет отвечать всем современным стандартам рынка и будет проста в реализации. Полученная в результате разработки система должна предоставлять возможность работы с данными различных типов, а также обладать гибкостью и масштабируемостью.

Заключение

В последние годы бурное развитие Интернета, Интернет вещей, и облачных вычислений привели к взрывному росту объемов данных практически во всех отраслях и сферах бизнеса. Большие данные быстро превратились в горячую тему, которая привлекает большое внимание академических кругов, промышленности и правительств во всем мире. Необходимость быстро и надежно обрабатывать данных в постоянно растущем информационном мире, как никогда остра в наше время. И исследование в области больших данных может улучшить уже существующие дисциплины, став тем самым помощником для современного человека [7].

Список использованной литературы:

1. XiaolongJin, Benjamin W.Wah, Xueqi Cheng, Yuanzhuo Wang. Significance and Challenges of Big Data Research, Big Data Research: Volume 2, June 2015, pages 59-64
2. Dr. Kuldeep Singh Kaswan, Mr. Jagjit Singh Dhatteval. 2020. Big Data, An Introduction
3. Davy Cielen, Arno Meysman, Mohamed Ali. 2016. Introducing Data Science: Big Data, Machine Learning, and more, using Python tool. Manning Publications Co. – pp. 320
4. Жукова М. В., Деменок С.Л. Просто BigData. — СПб.: Страта, 2019. —178 с
5. Ritu Chauhan and Harleen Kaur. Big Data Applications for Data Analytics. D.P. Acharjya et al. (eds.), Computational Intelligence for Big Data Analysis, Adaptation, Learning, and Optimization, Springer International Publishing, Switzerland, 2015 - pp 166.
6. Основы Big Data: концепции, алгоритмы и технологии. Томас Эрл Издательство: "Баланс Бизнес Букс", 2018 — 320 с.
7. RadwaElshawi, SherifSakr, Domenico Talia, Paolo Trunfio. Big Data Systems Meet Machine Learning Challenges: Towards Big Data Science as a Service. Big Data Research: Volume 14, December 2018, pp 1-11