Научная статья
Original article
УДК 004

СОВЕРШЕНСТВОВАНИЕ УПРАВЛЕНИЯ ДАННЫМИ НА ПРЕДПРИЯТИИ. ВИРТУАЛИЗАЦИЯ ДАННЫХ ДЛЯ ПРЕДПРИЯТИЯ

IMPROVING DATA MANAGEMENT IN THE ENTERPRISE. DATA VIRTUALIZATION FOR THE ENTERPRISE

Авраменко Павел Витальевич, Магистрант ИТМО, Россия, г. Санкт-Петербург

Avramenko Pavel Vitalievich, ITMO Master's Student, Russia, St. Petersburg

АННОТАЦИЯ

Теперь данные в режиме реального времени требуются всем организациям для принятия мгновенных бизнес-решений и повышения ценности для клиентов. Но эти данные повсюду: они хранятся в облаке, на платформах социальных сетей, в операционных системах и на веб-сайтах. Не говоря уже о том, что дополнительные источники постоянно добавляются через новые инициативы, такие как аналитика больших данных, облачная и устаревшая модернизация приложений. Для устранения изолированности данных и ускорения доступа ко всей информации о предприятиях организации могут выбрать передовой метод интеграции данных, известный как виртуализация данных. **Цель:** Обосновать преимущество виртуализация данных для предприятия **Методы:** В ходе исследования применялся сравнительный анализ. **Результат.** В ходе исследования было проведено

сравнение виртуализации данных с традиционным метода интеграции данных − консолидацией данных. В результате были обоснованы преимущества применением на предприятии нового способа интеграции данных – виртуализации. **Выводы.** Виртуализация позволяет работать с данными, не перемещая их куда-либо и даже не зная их местоположения, что экономит ресурсы и снижает риск ошибок в отличии от дорогостоящего и подверженного ошибкам перемещения сложно-структурированных данных.

Этот процесс часто занимает много времени, является дорогостоящим и подвержен ошибкам. Не говоря уже о том, что источники данных отличаются друг от друга по структуре, и их число постоянно растет.

ANNOTATION

Now, real-time data is required by all organizations to make instant business decisions and increase value for customers. But this data is everywhere: it is stored in the cloud, on social media platforms, in operating systems and on websites. Not to mention that additional sources are constantly being added through new initiatives such as big data analytics, cloud and legacy application upgrades. To eliminate data isolation and speed up access to all enterprise information, organizations can choose an advanced data integration method known as data virtualization. Objective: To substantiate the advantage of data virtualization for the enterprise Methods: Comparative analysis was used in the course of the study. Result. In the course of the study, data virtualization was compared with the traditional method of data integration − data consolidation. As a result, the advantages of using a new way of data integration – virtualization - at the enterprise were justified. Conclusions. Virtualization allows you to work with data without moving them anywhere or even knowing their location, which saves resources and reduces the risk of errors, unlike expensive and error-prone movement of complex structured data.

This process often takes a long time, is expensive and error-prone. Not to mention the fact that data sources differ from each other in structure, and their number is constantly growing.

Ключевые слова: Управление данными, виртуализация, интеграция данных.

Keywords: Data management, virtualization, data integration.

Виртуализация данных − это подход к интеграции данных из различных источников в целостную логическую картину без физического перемещения. Проще говоря, данные остаются в исходных источниках, в то время как пользователи могут получить и проанализировать их практически через специальное промежуточное программное обеспечение.

Традиционный способ интеграции данных включает объединение разрозненных данных в едином хранилище - обычно хранилище данных - через процесс извлечения, преобразования, загрузки (ETL). Тем не менее, для того, чтобы информация потреблялась конечными пользователями, вы должны сначала извлечь ее из источников, затем преобразовать ее в формат и структуру (схему), которые соответствуют целевой системе, и, наконец, загрузить ее в эту систему. Если этап преобразования происходит после загрузки (например, когда данные собираются в информационное озеро или в информационное озеро), этот процесс называется ELT.

Во многих случаях компании выбирают двухуровневую архитектуру, в которой исходные данные сначала извлекаются и загружаются в информационное озеро, а затем подвергаются нескольким ЭПЗ для доступа к целевым хранилищам данных и/или базам данных. Затем различные потребительские приложения и инструменты, каждый со своим собственным семантическим слоем, могут использовать эти данные для BI-аналитики и других задач.

Будь то ETL или ELT, основная концепция остается той же: Массивные тома данных из многих разрозненных источников копируются в новую, консолидированную систему, переживая трансформации где-то по пути.

Такие структуры порождают ряд проблем. Данные постоянно тиражируются, и многочисленные процессы ETL протекают там. Это часто

приводит к несоответствиям и дублированию данных, поскольку пользователи требуют больше представлений данных.

Необходимо создать и обслуживать новое целевое хранилище, которое является дорогостоящим и требует больших ресурсов. Проблемы безопасности и управления существуют, так как иногда конфиденциальные данные не могут быть перемещены в облако или централизованное хранилище, оставаясь в его родном месте. Данные, как правило, представляются в плановом порядке, а не по запросу. Данные все еще могут быть разрозненными, если определенные группы пользователей не имеют к ним доступа.

Преимущества и ограничения виртуализации данных.

Виртуализация данных, основанная на совершенно ином подходе к интеграции данных, решает эти и другие проблемы и обеспечивает следующие преимущества.

Доступ в режиме реального времени. Вместо физического перемещения данных в новое место виртуализация данных позволяет пользователям получать доступ к исходным данным и манипулировать ими через виртуальный/логический уровень в режиме реального времени. ЭТЛ в большинстве случаев не требуется.

Низкая стоимость. Внедрение виртуализации данных требует меньше ресурсов и инвестиций по сравнению с созданием отдельного консолидированного хранилища.

Повышенная безопасность данных и управление ими. Информация не должна перемещаться куда-либо и уровни доступа могут управляться.

Маневренность. Все данные предприятия доступны через один виртуальный уровень для различных пользователей и различных вариантов использования.

Возможности самообслуживания для всех бизнес-пользователей. Они могут разрабатывать и проводить любые необходимые им отчеты и анализ, не беспокоясь о формате данных или о том, где он находится.

К минусам можно отнести следующее.

Одна точка отказа. Поскольку сервер виртуализации предоставляет единую точку доступа ко всем источникам данных, это также часто приводит к одной точке сбоя. Если сервер не работает, вы рискуете оставить все операционные системы без передачи данных.

Нет поддержки пакетных данных. В отличие от вышеупомянутого ETL, этот метод интеграции не поддерживает пакетное или массовое перемещение данных, которое может потребоваться в ряде случаев. Скажем, финансовая фирма имеет дело с большими объемами транзакционных данных, которые нужно обрабатывать раз в неделю.

В связи с этим возникает вопрос - когда стоит сделать выбор в пользу виртуализации данных?

В двух словах, виртуализация данных происходит через промежуточное программное обеспечение, которое является лишь унифицированным, виртуальным уровнем доступа к данным, построенным поверх многих источников данных. Этот слой представляет информацию, независимо от ее типа и модели, как единое виртуальное (или логическое) представление. Все это происходит по требованию и в режиме реального времени.

При подключении виртуализация данных загружает метаданные (детали исходных данных) и физические представления, если таковые имеются. Он сопоставляет метаданные и семантически схожие активы данных из различных автономных баз данных с общей виртуальной моделью данных или схемой уровня абстракции. Сопоставление определяет, как информация из каждой системы источников должна быть преобразована и переформатирована для нужд интеграции.

Большинство программного обеспечения виртуализации данных обеспечивает доступ для широкого круга деловых пользователей, инструментов и приложений, включая такие популярные решения, как Tableau, Cognos и Power BI.

Такая структура создает возможности самообслуживания для всех потребителей. Вместо того, чтобы делать запросы напрямую, они взаимодействуют с программным обеспечением, используемым в их повседневной работе, и это программное обеспечение, в свою очередь, взаимодействует с виртуальным уровнем, получая необходимые данные. Таким образом, потребителям не нужно беспокоиться о формате данных или их местоположении, так как эти сложности скрыты от них.

В течение многих лет компании объединяли данные из нескольких операционных систем, копируя их в хранилища данных, информационные магазины или информационные озера для дальнейшего анализа и принятия решений. Этот процесс часто занимает много времени, является дорогостоящим и подвержен ошибкам. Не говоря уже о том, что источники данных отличаются друг от друга по структуре, и их число постоянно растет. Вот почему многие организации выбирают программное обеспечение виртуализации данных, которое позволяет им просматривать, получать доступ и анализировать данные, не перемещая их куда-либо и даже не зная их местоположения.

Список литературы:

1. Business Analysis Body of Knowledge // URL: https://www.iiba.org/standards-and-resources/babok/ (Дата обращения: 12.03.2022)
2. Data Virtualization. An overview // URL: https://www.denodo.com/en/data-virtualization/overview (Дата обращения: 15.02.2022)
3. Easing Data Woes and Creating Tangible Business Value Through Data Virtualization in the Financial Services Industry // URL: https://www.datavirtualizationblog.com/easing-data-woes-and-creating-tangible-business-value-through-data-virtualization-in-the-financial-services-industry/ (Дата обращения: 25.02.2022)
4. Gartner // URL: https://www.gartner.com/reviews/market/analytics-business intelligenceplatforms (Дата обращения: 12.12.2021)

5. Iiba // URL: https://www.iiba.org/standards-and-resources/babok/babok guide-errata/ (Дата обращения: 22.02.2022)
6. Microsoft // URL: https://products.office.com/ru-ru/business/microsoft powerapps (Дата обращения: 02.03.2022).
7. Seagate // URL: https://www.seagate.com/www-content/our story/trends/files/idc-seagatedataage-whitepaper.pdf (Дата обращения:02.03.2022).

List of literature:

1. Business Analysis Body of Knowledge // URL: https://www.iiba.org/standards-and-resources/babok / (Accessed 12.03.2022)
2. Data Virtualization. An overview // URL: https://www.denodo.com/en/data-virtualization/overview (Accessed: 02/15/2022)
3. Easing Data Woes and Creating Tangible Business Value Through Data Virtualization in the Financial Services Industry // URL: https://www.datavirtualizationblog.com/easing-data-woes-and-creating-tangible-business-value-through-data-virtualization-in-the-financial-services-industry / (Date of request: 02/25/2022)
4. Gartner // URL: https://www.gartner.com/reviews/market/analytics-business intelligenceplatforms (Accessed date: 12.12.2021)
5. Iiba // URL: https://www.iiba.org/standards-and-resources/babok/babok guide-errata/ (Accessed: 02/22/2022)
6. Microsoft // URL: https://products.office.com/ru-ru/business/microsoft powerapps (Accessed: 02.03.2022).
7. Seagate // URL: https://www.seagate.com/www-content/our story/trends/files/idc-seagatedataage-whitepaper.pdf (Accessed date:02.03.2022).

© Авраменко П.В., 2022 Научно-образовательный журнал для студентов и преподавателей «StudNet» №6/2022.

Для цитирования: *Авраменко П.В.* СОВЕРШЕНСТВОВАНИЕ УПРАВЛЕНИЯ ДАННЫМИ НА ПРЕДПРИЯТИИ. ВИРТУАЛИЗАЦИЯ ДАННЫХ ДЛЯ ПРЕДПРИЯТИЯ// Научно- образовательный журнал для студентов и преподавателей «StudNet» №5/2022.