

UNIVERSITY OF WATERLOO

Faculty of Mathematics

**RESOLVING SCATTERED INFORMATION FROM DIFFERENT
SYSTEMS WITHIN THE WORKPLACE**

CI Global Asset Management

Advanced Analytics Department

Toronto, Ontario

Prepared by

Calista Kurniawan

ID 21000788

1B Honours Mathematics

July 20, 2023

July 20, 2023

Dear PD Evaluators,

For the past few weeks, I have been working on writing my first technical report titled “Resolving Scattered Information from Different Systems Within the Workplace”. This report is written to complete a Co-operative Education Program course, which is in progressing currently after my 1B Honours Mathematics term. As stated on the cover page, this report has been written and prepared by me and resources used within the report have been credited at the end of this report.

Currently, I am working as a Data Scientist at CI Global Asset Management, under the Advanced Analytics department located in Toronto, Ontario. My current role is to read and understand the different pipelines and projects used within the team and document the functionalities and data within Confluence. This requires the ability to trace code in languages such as Python, SQL, or Lambda, as well as the ability to organize and document the information about the different functions properly so that clients using the product can implement it in their systems and occupations. The purpose of this technical report is to give solutions regarding the issue of scattered information within multiple workplaces including CI Global Asset Management.

I would like to acknowledge the assistance received by people within the PD11 team and the classmates who assisted in the peer-editing stage of writing. I have checked my report for spelling and grammar mistakes to the best of my abilities. This report was written entirely by me and has not received any previous academic credit at this or any other institution.

Sincerely,

Calista Kurniawan

21000788

Calista Kurniawan

Executive Summary

CI Global Asset Management, one of Canada's largest investment companies, focuses on investment and wealth services to meet the needs of all clients and investors. CI focuses on multiple divisions such as mutual funds, exchange-traded funds, and private pools (private funds), and assists portfolio managers with making decisions based on the market. As a member of the Advanced Analytics team, it is a priority to create and manage services available for portfolio managers to use and aid them to make proper decisions for their clients. These services require large amounts of data with information regarding clients, and funds. Thus, the purpose of this report is to improve the data within the company to enhance the company's efficiency, accuracy, and customer service.

Scattered information is a problem that can be found in various companies, as it may seem difficult to change processes of data storage and combine multiple into one source of truth. This report displays the simplicity and importance of resolving this issue of scattered data.

To summarize, this report's goal is to understand what scattered information means and looks like, while giving solutions that are realistic and manageable to implement within the company. Though every company may have different priorities and budgeting, this report is available to give recommendations, though a company must overview its current layout and carefully discuss with the different systems' teams to choose the proper solution tailored for the company. For CI Global Asset Management, AWS' S3 data lake may be the best solution due to the company's experienced use of Amazon Web Services and familiarity with S3 in certain teams.

Table of Contents

EXECUTIVE SUMMARY	ii
LIST OF FIGURES	iv
1.0 INTRODUCTION	1
2.0 LITERATURE REVIEW	2
3.0 ANALYSIS	3
3.1 Centralized Data Storage	3
3.1.1 Amazon Web Services S3	5
3.1.2 Microsoft SharePoint	5
3.1.3 Snowflake	5
3.1.4 OneTrust	6
3.2 Knowledge Management Systems	6
3.2.1 Document Management Systems	7
3.2.2 Business Intelligence Systems	8
3.2.3 Customer Relationship Management Systems	8
4.0 DISCUSSION	9
5.0 CONCLUSION	11
6.0 RECOMMENDATIONS	12
REFERENCES	13
ACKNOWLEDGMENTS	15

List of Figures

Figure 1 – Example of a Centralized Data Storage	3
Figure 2 – Knowledge-based Systems Architecture	7

1.0 Introduction

In today's society, organizations often find themselves struggling against the challenges of scattered information across multiple platforms and systems. As businesses continue to grow and adapt to evolving technology, the issue of fragmented data arises, impacting decision-making and efficiency.

For context, CI Global Asset Management is a global asset and wealth management advisory service, aiding clients with mutual funds and investments. As a member of the advanced analytics team, responsible for managing and optimizing datasets and AI training models, it is crucial to have accurate datasets as well as consistency throughout databases within the company. Due to the company utilizing different systems, an issue arises where the systems do not connect to the same data cloud, resulting in scattered information. To overcome this hurdle, centralizing the data storage through the merging of platforms through the different systems and knowledge management systems are two promising approaches to resolve this data issue. By integrating software solutions and understanding how scattered information lies within a business, organizations can achieve enhanced data visibility, efficient workflows, and improved productivity. This technical report aims to explore and understand the benefits of resolving scattered databases by merging platforms from different systems and offering software recommendations and implementations to improve the business environment.

2.0 Literature Review

Scattered data refers to information that is disorganized, fragmented, or spread across multiple sources, systems, or locations within a business (Saunders, 2016). Scattered information can have various negative impacts on a company, including reduced efficiency and productivity, inaccurate decision-making, data security risks, and the creation of data silos. Not keeping track of the information also results in old and irrelevant data being preserved due to the lack of clean-up or maintenance, as well as inefficient business processes which may be costly for a company (Showell, 2022). Companies that have not updated their technical data storages tend to have scattered information, due to the problem of having multiple databases that other teams may not have access to. This results in the creation of data silos, a barrier where the “collection of data held by one group is not easily or fully accessible by other groups in the same organization” (“Data silos, why they’re a problem, & how to fix it”, 2023).

Currently, CI Global Asset Management has different systems connected to individual databases, causing issues when selecting a database to be the company’s main source of information. The company has the front, middle, and back offices controlling their own data sets, while users grab from whichever system they need. The issue with this layout is that because there is no main database that everyone shares, the same documented information may be inputted differently, resulting in inaccurate records. In summary, scattered data can have significant detrimental effects on businesses, impacting security, collaboration, and overall operational efficiency. By addressing the challenges associated with scattered information and implementing solutions for centralizing and organizing data, businesses can reduce these negative impacts and start unlocking the full potential of their data.

3.0 Analysis

3.1 Centralized Data Storage

Firstly, a solution to resolving the issue of scattered data is by updating the current data format and implementing centralized data storage. A centralized data storage is a single location where the company stores and maintains data, thereby having a single source of data truth where all systems can refer to the same location when information is needed.

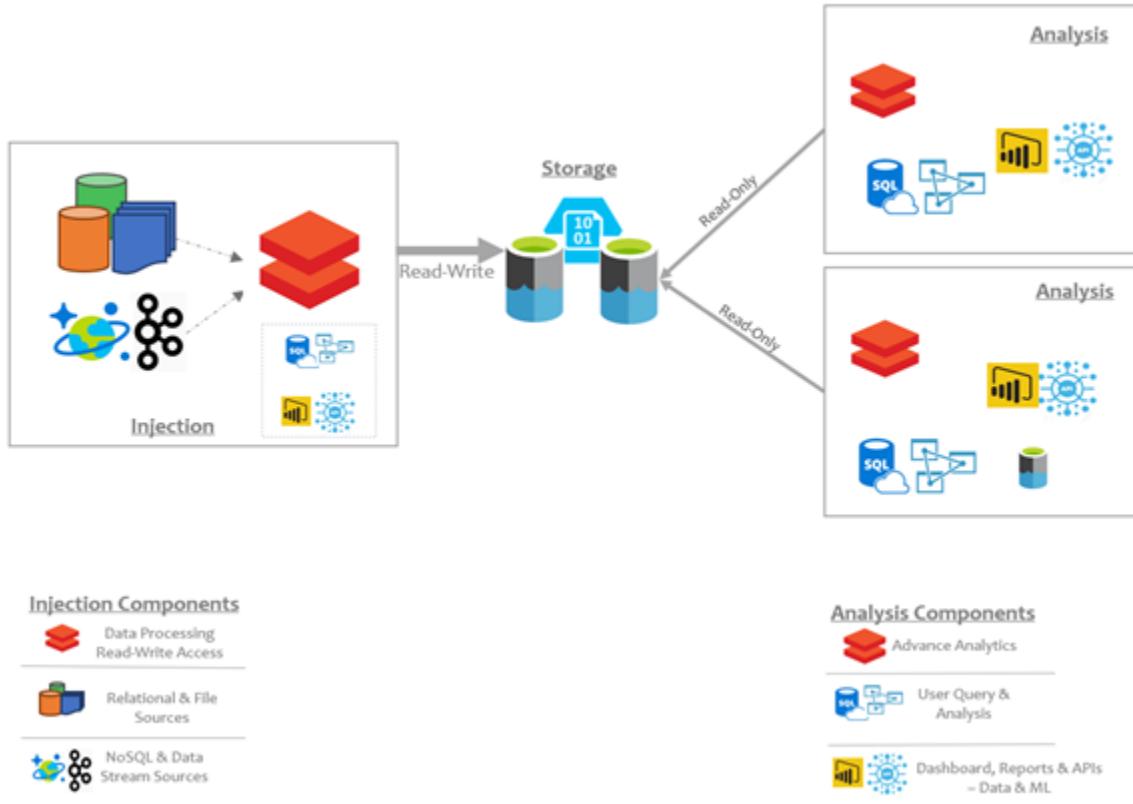


Figure 1

Example of a Centralized Data Storage

Note. This table was created to describe the concept of centralized data storage by Rengachari, P. (2022, November 7). Centralized data storage system: How your business can benefit? - ilink digital. <https://www.ilink-digital.com/insights/blog/centralized-data-storage-system-how-your-business-can-benefit/>

Figure 1 displays an example of how centralized data storage can be organized and used within a company. It displays how different data containers and sets may come from different systems but are all connected to one storage, the single source of truth. “The company needs to implement centralized data storage, as it allows us to have a single source of truth where every team in the company can access the same information, without having to worry about whether or not it is accurate to the other teams’ information” (L. Day, personal communication, July 4, 2023). Lisa states how with the use of centralized data storage, the company’s internal teams can have better communication and understanding of each other’s data, and how they all correlate to each other without having to review whether different databases align with each other. “81 percent of businesses had reported that custom apps reduced inefficient tasks, with a further 74 percent stating they increase team productivity” (Saunders, 2016). Saunders’ statistics display how impactful centralized data storages have the ability to be within a company, and how it can benefit efficiency, thus leading to better outcomes being produced. Lisa also mentions how “implementing a centralized data storage will allow the company to focus more on the analysis part, as we can start using data analysis platforms such as DataRobot or Dataiku and connect it to the centralized data storage” (L. Day, personal communication, July 4, 2023). Data analysis platforms will be more relevant and useful after implementing centralized data storage, as it will increase productivity and more accurate data reports. The next few sections give examples of services and software that can aid in transitioning to centralized data storage.

3.1.1 Amazon Web Services S3

AWS uses Amazon S3 as its primary storage platform, allowing clients to build a data lake (“Central storage: Amazon s3 as the data lake storage platform”, 2023). S3 is useful for companies that are already using Amazon’s other services such as SageMaker Studio, Athena, and Lambda. When already integrated into these services, S3 can be easily linked for smooth data access and decreases the number of steps to connect databases with the services. Thus, AWS may be the best choice for a company that is already familiar with its’ other services and wants to link its data to those said services in order to create a smoother transfer of data. The catch with Amazon Web Services

3.1.2 Microsoft SharePoint

Similarly, an example of a platform with the ability to manage and organize a company’s information is Microsoft’s SharePoint (“What is SharePoint?”, 2023). SharePoint is a cloud-based service allowing employees to create and share documents with each other in the cloud (“What is SharePoint?”, 2023). Microsoft’s platform is optimal for companies who have fewer databases, but a large number of documents that may need to constantly be passed around to different teams. SharePoint also starts with a price of 5 dollars per month, with other plans available to suit the company’s needs (“What is SharePoint?”, 2023).

3.1.3 Snowflake

Snowflake is an alternate cloud data warehouse that has the ability to store and analyze a company’s data records in one place (“How does Snowflake work?”, 2022). The platform is easily scalable so that all companies regardless of size can use their cloud data warehouse. Similar to AWS, Snowflake offers cloud services that assist in the security, availability, and

development of the data within the cloud. Since Snowflake is known within the data industry, many data analysis companies are already able to sync with Snowflake, thus creating a smoother linking process between platforms that can benefit the company's efficiency and accuracy.

3.1.4 OneTrust

Finally, an option is to use OneTrust, which is a platform that manages and protects a company's data, creating a strong security and private cloud for your data to be stored within. News Bites states that "centralizing on OneTrust gives organizations a single source of data truth that can intelligently create workflows and automate business decisions that achieve compliance and build trust" ("Data Processing Consent Sharing Systems and Related Methods", 2023). This means OneTrust prioritizes creating centralized data storage as the single source of truth and thrives to secure the main data storage for privacy and data governance. Overall, whether the priority is linking services, easy document sharing, or security, some various platforms and services can tend to a company's needs and create centralized data storage to increase efficiency, accuracy, and data protection.

3.2 Knowledge Management Systems

Secondly, Knowledge Management Systems are another way to resolve the scattered information problem within the company. Knowledge Management Systems are platforms that are designed to capture, organize, store, retrieve, and share knowledge within an organization, based on simple questions asked to the system's query.

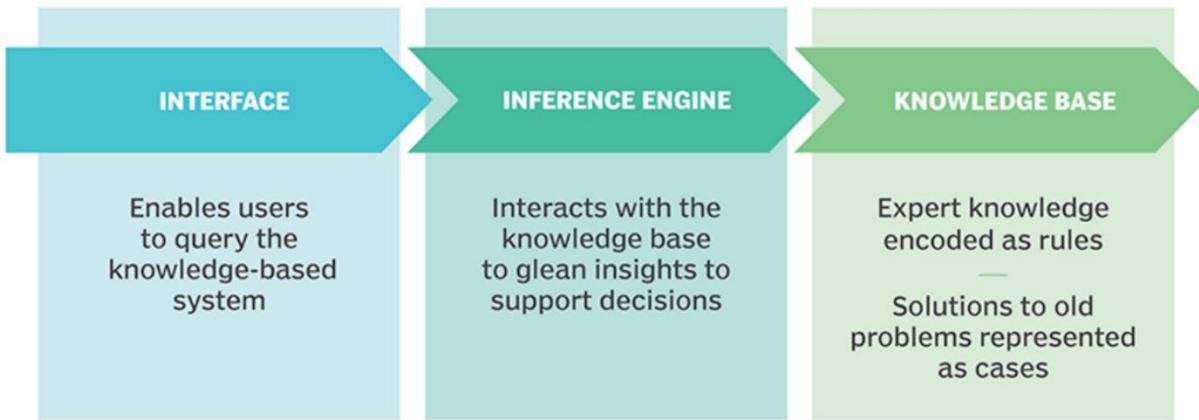


Figure 2

Knowledge-based Systems Architecture

Note. This visual was created to describe the architecture of Knowledge Management Systems by Amsler, S. (2023). What is knowledge management? -- Definition from whatis. Com. Content Management. <https://www.techtarget.com/searchcontentmanagement/definition/knowledge-management-KM>

The primary goal of a knowledge management system is to facilitate collaboration and knowledge-sharing among individuals and teams. There are different types of Knowledge Management Systems such as Document Management Systems, Content Management Systems, Enterprise Search Engines, and more. The purpose of having various types is so that every company can achieve its goals and needs with Knowledge Management Systems.

3.2.1 Document Management Systems

Similar to Microsoft SharePoint, Document Management Systems help organize, store, and retrieve various types of documents, such as text files, spreadsheets, presentations, and PDFs. The system's control of the cloud makes it easier to locate and access scattered data. DMS

systems are more applicable for companies that have large amounts of files with client information that need to be kept confidential and protected, while still needing to pass these files along to other internal teams.

3.2.2 Business Intelligence Systems

For a company heavily working with data, BI systems have the ability to collect, analyze and visualize data from various sources to provide relevant insights. The system can help consolidate scattered data from multiple systems or databases, perform analysis, and create reports that facilitate decision-making. Business Intelligence Systems are important due to the fact that it allows organizations to ask questions in simple sentences, with the return result of data that can be analyzed for the business' benefit ("What is business intelligence and how does it work?", 2023).

3.2.3 Customer Relationship Management Systems

Relevant to CI Global Asset Management, CRM systems centralize customer-related data. This includes information such as contact information, history, and interactions. By implementing a Customer Relationship Management System, it will enable better customer relationship management, personalized marketing, and improved customer service. Since CI Global Asset Management deals with clients and information about their funds, it is important to have an organized and accurate dataset, holding client information, thus having a Customer Relationship Management System is needed and a priority.

4.0 Discussion

After researching, the results display how crucial it is to fix scattered information issues, because of how detrimental the consequences affect a company. Based on the analysis, there are multiple applications available to help centralize data storage which helps create consistency and prevent repetitive entries within the database. These applications also allow workers to shift their focus on analyzing the data, instead of wasting precious time cleaning and organizing dirty data sets. By prioritizing and having the budget for choosing an application suitable for the company's database, the software enhances the company's efficiency, collaboration, and accuracy, resulting in clients trusting the business, and staying afloat with society's evolution with data. The research also shows how it does not cost a fortune to purchase services that manage centralized data storage for the company. Centralized data storage will help modernize and update the company's tactics while benefiting the business, through syncing the systems together while efficiently organizing the data, thus creating a better environment for the workers, resulting in happier clients. In addition, implementing a Customer Relationship Management System with centralized data storage will enable even better customer service, improving the company and its reputation.

For the company, CI Global Asset Management's main priorities are the clients and optimizing service usage. With this knowledge, the best platform would be AWS S3 data storage, with a CRM system supporting it. The reason for this recommendation is because of the company's experience and previous exposure to Amazon Web Services such as Athena and SageMaker Studio, that can easily be linked to AWS S3 data storage. The CRM system can be used to optimize the use of client information and benefit from better customer services and client advice. On the contrary, OneTrust, Microsoft SharePoint, and the DMS system may be the

least applicable for CI compared to the other platforms. OneTrust and SharePoint may not be the most suited choices due to their focus on documents and files, where the company does not seem to display issues with document sharing, and thus may not be the most suitable for CI.

5.0 Conclusion

To summarize, companies need to prioritize resolving issues related to scattered information and data. By addressing the challenges posed by scattered data, such as fragmented systems, outdated data entries, and inefficient use of workers' time, companies can obtain numerous benefits. For instance, they can enhance the decision-making process, sync databases, improve customer experiences, and allow more collaboration between teams in the company. Moreover, a centralized and integrated approach to data management allows companies to identify patterns and encourage analytics to thrive for better results. In conclusion, by being aware and acting towards scattered information and data issues, companies can position themselves for sustained growth and maximize their potential within a technology-advancing world.

6.0 Recommendations

The recommendations listed in this section are more focused and tailored towards CI Global Asset Management and its priorities, although it applies to other companies who are dealing with a similar situation. A recommendation for a service to use as centralized data storage is to use AWS' S3 data lake. This is because the company already uses S3 for some data storage in certain teams, and the company also uses services from AWS such as SageMaker Studio and Athena, thus making a smoother experience when transferring data and using the services. Outsourcing is a good way to make sure centralizing is done properly by a service that is professionally experienced in the field and can optimize for the company's needs. It is also crucial to budget properly for these issues because it is the source of numerous issues that may arise when neglected.

References

- Amsler, S. (2023). What is knowledge management? -- Definition from whatis. Com. Content Management.<https://www.techtarget.com/searchcontentmanagement/definition/knowledge-management-KM>
- Central storage: Amazon s3 as the data lake storage platform—Storage best practices for data and analytics applications. (2023). Amazon Web Services.
<https://docs.aws.amazon.com/whitepapers/latest/building-data-lakes/amazon-s3-data-lake-storage-platform.html>
- Data silos, why they're a problem, & how to fix it. (2023). Talend - A Leader in Data Integration
& Data Integrity. <https://www.talend.com/resources/what-are-data-silos/>
- How does Snowflake work? A simple explanation of the popular data warehouse. (2022, September 27). mparticle. <https://www.mparticle.com/blog/how-does-snowflake-work/>
- OneTrust LLC issued a patent titled "Data Processing Consent Sharing Systems and Related Methods". (2023, Jun 28). News Bites - Private Companies
<http://search.proquest.com.proxy.lib.uwaterloo.ca/wire-feeds/onetrust-llc-issued-patent-titled-data-processing/docview/2830025519/se-2>
- Rengachari, P. (2022, November 7). Centralized data storage system: How your business can benefit? - ilink digital. <https://www.ilink-digital.com/insights/blog/centralized-data-storage-system-how-your-business-can-benefit/>

Saunders, K. (2016). Is scattered information destroying your business? Fresh Business Thinking. <https://www.freshbusinessthinking.com/purpose/is-scattered-information-destroying-your-business/42340.article>

Showell. (2022, March). The dangers of scattered content and how to solve it.

<https://www.showell.com/resources/the-dangers-of-scattered-content-and-how-to-solve-it>

What is business intelligence and how does it work? | ibm. (2023). IBM.

<https://www.ibm.com/topics/business-intelligence>

What is SharePoint? - Microsoft support. (2023). Microsoft. <https://support.microsoft.com/en-us/office/what-is-sharepoint-97b915e6-651b-43b2-827d-fb25777f446f>

Acknowledgments

I would like to thank my supervisor in CI Global Asset Management, Lisa, for allowing me to ask her questions about the company's current data storage status and advising solutions that are relevant and crucial to resolving the issue. Due to the information given by Lisa, I was able to proceed to create a technical report that is tailored and useful for CI Global Asset Management and its needs. I would also like to thank all those who are involved with the PD11 course, for providing insightful and applicable information that is heavily used in this technical report.