European Journal of Computer Science and Information Technology, 12 (6),15-23, 2024
Print ISSN: 2054-0957 (Print),
Online ISSN: 2054-0965 (Online)
Website: https://www.eajournals.org/
Publication of the European Centre for Research Training and Development -UK

Amplifying Big Data Utilization in Healthcare Analytics Through Cloud and Snowflake Migration

Jinesh Kumar Chinnathambi
Subject Matter Expert(SME)//4x AWS Certified Technologist,
Leading Health Insurance Company,
Richmond, Virginia, United States.

doi: https://doi.org/10.37745/ejcsit.2013/vol12n61523 Published September 16, 2024

Citation:Chinnathambi J.K. (2024) Amplifying Big Data Utilization in Healthcare Analytics Through Cloud and Snowflake Migration, *European Journal of Computer Science and Information Technology*, 12 (6),15-23

Abstract: *Amplifying the utilization of big data in healthcare analytics through cloud and Snowflake migration presents a significant opportunity to enhance data-driven insights and decision-making in the healthcare sector. This migration makes it easier to move large amounts of healthcare data to the cloud. Applications deployed in could are scalable for in-depth analysis in Health Care industry. The cloud is becoming more popular for storing data and running applications because it can easily grow with your needs, requires little to no management, improves security, and offers budget flexibility. The benefits of the cloud are obvious -- once you get there. Moving to the cloud requires planning, strategy, and the right tools for data migration. [1] By using Snowflake's advanced data warehousing tools, healthcare organizations can smoothly handle and analyze their complex and varied data. This helps them quickly uncover important insights and make better decisions. The shift to cloud technology and Snowflake has the potential to significantly enhance real-time analytics, personalized patient care, and evidence-based decision-making in healthcare. When healthcare organizations leverage big data in a cloud-based setting, they can discover valuable insights from their data, ultimately improving clinical outcomes, operational efficiency, and healthcare delivery. This study explores how the adoption of cloud and Snowflake in healthcare analytics can bring about transformative change and create new possibilities for leveraging data and generating insights in the healthcare sector*.

Keywords*:* Healthcare Analytics, Big Data, Cloud Migration, Snowflake, Data Insights, Decision-making, Scalability, Data Security, Real-time Analytics, Operational Efficiency, Cloud Infrastructure, Big Data Management, Data Warehousing, Cloud Technology, Snowflake Migration, AWS Services, Data Warehousing

INTRODUCTION

In the present era driven by data, Big Data is reshaping numerous industries, and healthcare stands out as one of the most profoundly affected. Healthcare professionals, researchers, and administrators grapple with vast amounts of data, ranging from electronic health records to genomic information. When leveraged through advanced analytics, this data holds the potential to notably enhance patient care, optimize operations, and propel advancements in medical research. However, the challenge lies in effectively managing and extracting value from this data, particularly within the constraints of traditional on-premise infrastructures.By embracing cloud infrastructure and Snowflake's innovative data management platform,

healthcare organizations can address the challenges of scalability, interoperability, and data security, while unlocking the full potential of big data analytics. This introduction sets the stage for exploring the benefits and implications of cloud and Snowflake migration in amplifying big data utilization, paving the way for enhanced healthcare outcomes, operational efficiencies, and evidence-based insights.The recent adoption of cloud technology and the use of cloud-native data platforms like Snowflake have brought about transformative solutions to this obstacle. The shift to a cloud-based infrastructure delivers scalability, adaptability, and cost-effectiveness, effectively meeting the demands of managing big data. Concurrently, Snowflake offers a managed data warehousing solution designed for the cloud, streamlining the storage and analysis of large datasets. With the integration of cloud technology and Snowflake, healthcare organizations are now well-equipped to enhance the utilization of big data in healthcare analytics, leading to improvements across various aspects of healthcare delivery. In the next 10 years, cloud computing for healthcare is expected to increase at a compound annual growth rate (CAGR) of 17.4% and reach a valuation of $280 million, highlighting that the pandemic has been a catalyst for the growth of healthcare cloud use. [2]

BACKGROUND

The advent of Big Data has brought about profound changes across diverse sectors, with healthcare standing out as a primary beneficiary. Within the healthcare industry, an extensive array of data is generated, encompassing patient records, clinical trial data, and intricate genomic sequences. Effectively harnessing this vast pool of data carries significant implications for patient outcomes, healthcare policies, and overall industry growth. However, efficiently managing and extracting valuable insights from such extensive data remains a complex challenge with traditional on-premises infrastructure due to its limitations in handling the volume, velocity, and variety inherent in healthcare data.

The emergence of cloud computing has revolutionized data management, offering organizations a powerful, scalable, and cost-efficient alternative to traditional setups. Its key advantages, such as on-demand scalability, pay-per-usage cost structure, and universal accessibility, are pivotal for effective big data management. Additionally, streamlined collaboration, enhanced security, and disaster recovery make cloud technology the preferential choice for the healthcare industry. To create a better patient experience, healthcare organizations seek solutions that help them gain actionable insights and uncover new patterns in disparate data sets. AWS Marketplace offers a curated catalog of healthcare analytics solutions that speed the discovery, visualization, and sharing of data and insights. These solutions empower organizations to unlock the value of their data, better the patient experience, and more accurately predict patient outcomes. [3]

While migration to the cloud addresses many challenges associated with big data, it does not entirely eliminate them. This is where Snowflake, a cloud-native data platform, plays a crucial role. Snowflake provides a comprehensive, zero-management solution tailored specifically for data warehousing in a cloud environment. It facilitates the storage of large volumes of structured and semi-structured data across multiple clouds, enabling seamless data management, rapid data-driven insights, secure data sharing, and real-time analytics. Leveraging the power of cloud technology and the innovation of Snowflake, healthcare providers are well-equipped to enhance their utilization of big data, driving innovation, improving patient care, and achieving exceptional operational efficiency.

SAMPLE ON-PREMISE HEALTHCARE DATA ANALYTICS APPLICATION AS-IS

In the traditional model, healthcare organizations relied on on-premise applications supported by robust databases such as SQL Server and Oracle. These resilient database systems efficiently stored and managed

a wide range of healthcare data, including patient records, lab results, billing information, and clinical study data. The applications built on these platforms played a crucial role for healthcare providers, allowing them to handle and retrieve critical data. However, as the volume of data continued to increase and the demand for advanced analytics grew, it became evident that these systems alone could not deliver the scalability, speed, and efficiency needed to support the complex analytics processes essential for managing healthcare data.

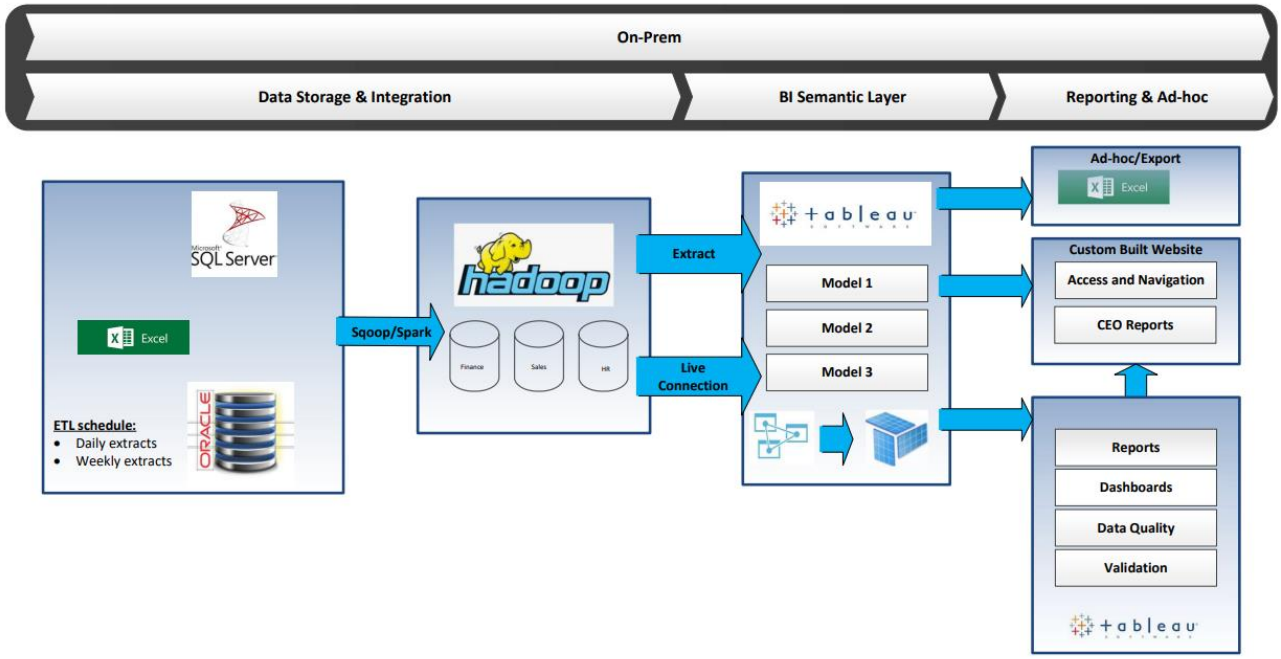

FIGURE 1: on-premises application

In response to these challenges, the data previously stored in SQL Server and Oracle was migrated to the Hadoop ecosystem, which offers greater capabilities for handling big data and complex analytics. With its distributed computing architecture and cost-effective storage features, Hadoop was employed to manage and process extensive data sets. It not only facilitated the management of large volumes of structured data from SQL Server and Oracle but also handled semi-structured and unstructured data, enriching the data landscape. Tools within the Hadoop ecosystem, including Hive and Pig, were utilized for data processing and analytics, delving deep into these expansive data repositories to extract valuable insights. The insights obtained were then visualized using Tableau, a prominent data visualization platform. This empowered healthcare providers with intuitive, interactive dashboards and reports, transforming raw, complex data into understandable, actionable information. The reports generated by Tableau guided healthcare professionals and stakeholders in making data-driven decisions to enhance patient care, streamline operations, and formulate healthcare strategies. Healthcare costs can quickly spin out of control. Misallocation of resources can quickly bring down quality of care. To keep efficiency and profitability moving in the right direction, you need to see all your key healthcare reporting metrics across hospitals, programs, and regions. You need to cut that data many different ways and share it with key employees in order to manage your business more effectively. [4]

HADOOP TO SNOWFLAKE MIGRATION BY LEVERAGING SNOWFLAKE UTILITIES

Assumptions: The approach outlined considers an on-premise Hadoop data warehouse and an AWS Snowflake environment. It assumes that the setup for Virtual Private Cloud (VPC), Virtual Private Server (VPS), AWS Direct Connect, Accounts, IAM User/Role/Policy, S3 buckets, Snowflake Databases, Schemas, credentials, roles, and warehouses is either already established or will be arranged before commencing the migration. It is noted that the processes for data modeling that can develop DDL's/data types compatible with Snowflake are already in place. The estimated maximum size of the data warehouse set to be migrated is 10TB.

Approach Description: The migration process is divided into two parts: History data migration and Incremental data migration.

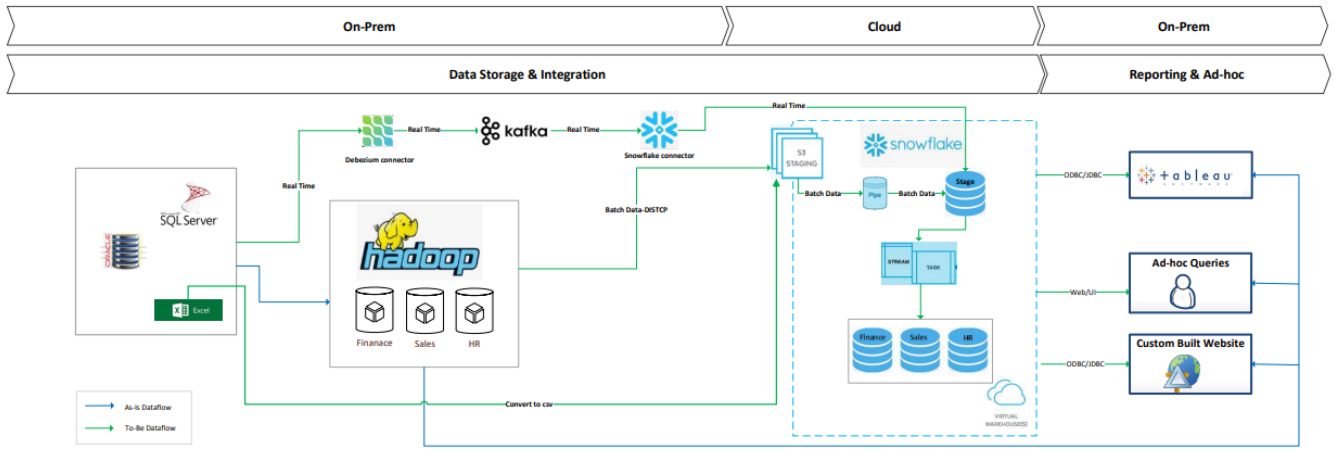

FIGURE 2: Snowflake migration by leveraging Snowflake utilities

During the History Data Migration phase, DISTCP is utilized to copy the parquet files from on-premise HDFS to AWS-S3. An external stage, pointing to S3 where parquet files are located, is created. Snowpipe is then provisioned to move the data from the external stage into the stage or reporting table(s) within Snowflake. Configuration is set up such that the S3 event notification of the external stage alerts Snowpipe when new data arrives. Streams/tasks are employed for any transformation logic orchestration.

In the Incremental Data Migration phase, Kafka is used to stream data from the source systems to AWS-S3 at predefined intervals, using JDBC or log-based Change Data Control (CDC). Change data capture (CDC) is a set of software design patterns. It allows users to detect and manage incremental changes at the data source. CDC technology lets users apply changes downstream, throughout the enterprise. CDC captures changes as they happen. [5] A Snowflake connector is set up to load data into the stage table(s). Following that, Snowflake streams/tasks are utilized to merge data from the stage to reporting schemas/tables. For Reporting, a copy of the existing Tableau reports is created and the connection string is adjusted to point to Snowflake.

HADOOP TO SNOWFLAKE MIGRATION BY LEVERAGING AWS SERVICES

Assumptions: The following strategy takes into account an existing on-premise Hadoop data warehouse and an AWS Snowflake environment. Snowflake delivers the Data Cloud — a global network where thousands of organizations mobilize data with near-unlimited scale, concurrency, and performance. Inside the Data Cloud, organizations unite their siloed data, easily discover and securely share governed data, and execute diverse analytic workloads. [6] It is assumed that essential components, such as Virtual Private Cloud (VPC), Virtual Private Server (VPS), AWS Direct Connect, Accounts, IAM User/Role/Policy, S3 buckets, external stages, Snowflake Databases, Schemas, credentials, roles, and warehouses, are already established or will be set up before commencing the migration. Additionally, it is assumed that there is an existing data modeling process for creating Data Definition Languages (DDL's)/datatypes compatible with Snowflake. The data warehouse set for migration is estimated to be approximately 10TB.

Description: This solution/framework has been successfully deployed in my current organization, enabling various teams to migrate their on-premise data warehouses to Snowflake (AWS). To enhance security, data protection software is used to tokenize sensitive information before cloud migration and to detokenize within the Snowflake semantic layer views.

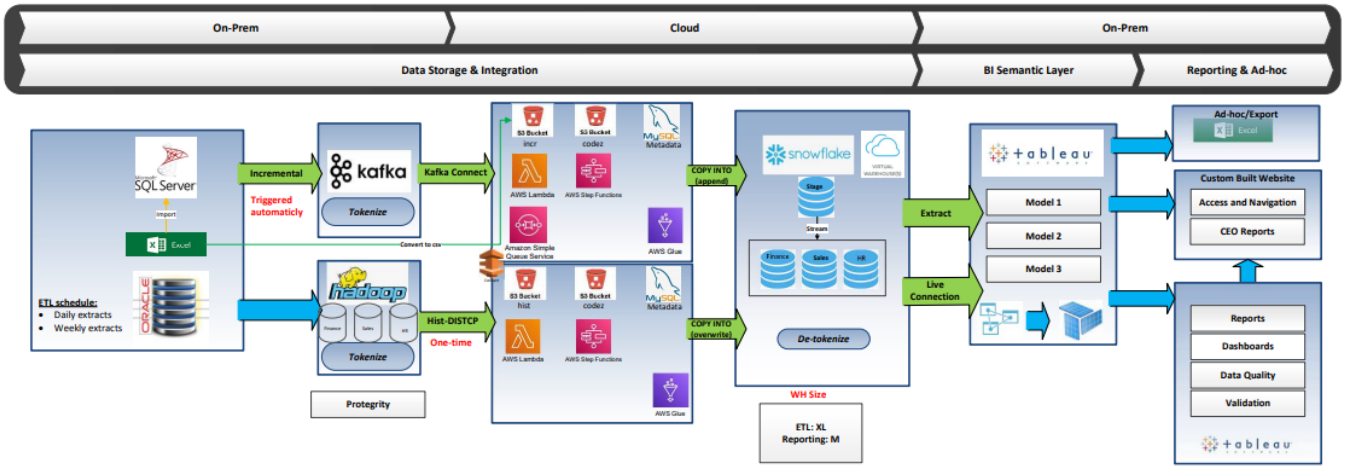

FIGURE 3: Snowflake migration by leveraging AWS services

For Historical Data Migration, DISTCP is used to copy parquet files from on-premise HDFS to AWS-S3. Amazon Simple Storage Service (Amazon S3) is an object storage service offering industry-leading scalability, data availability, security, and performance. [7] AWS services such as Lambda/Step function/Glue, alongside Python scripts, are orchestrated to export parquet files from the AWS-S3 landing zone to the S3 external stage, and then copy them into Snowflake reporting schemas. The S3-Codez bucket contains SnowSQL scripts, executed by AWS Lambda/Step function/Glue for any required transformations.

The Incremental Data Migration leverages Kafka, streaming data from source systems to AWS-S3 at regular intervals using JDBC or log-based Change Data Control (CDC). AWS Lambda is employed to

export parquet files from the AWS-S3 landing zone to the S3-External stage and copy the data into Snowflake stage schemas. Additionally, Snowflake streams merge the data from the stage to reporting schemas/tables. Similar to historical data migration, AWS Lambda/Step function/Glue with Python scripts may be used to execute the SnowSQL scripts located in the S3-Codez bucket for transformations from stage to reporting schemas/tables. AWS-RDS (MySQL) is utilized to store all the metadata required to drive the framework. Similar implementation capabilities are anticipated for future projects.

SNOWFLAKE MIGRATION BEST PRACTICES

1. Establish the Migration Approach: Identify the processes to be migrated as they currently exist, those that will need reengineering, and areas where performance can be improved.
2. Preparation of 'To-Be' Architecture: Create a diagram of the intended architecture and review it with a Snowflake Solution Partner or Expert.
3. Identify Tools and Strategy: Determine the set of tools needed for migration, those to be decommissioned post-migration, the development environments, and the deployment strategy.
4. Team Formation: Create a migration team and assign the roles and responsibilities.
5. Initiate a Proof of Concept (POC): Select a small set of tables/datasets for migration as part of a POC for post-migration validation, beginning small-scale instead of adopting a 'Big Bang' approach.
6. Develop a Test Plan: Plan the set of strategies to test the migrated data and system.
7. Set Up Environments: Configure Snowflake environments based on the determined usage and requirements.
8. Execute the Migration: Implement the actual process of migration as per the plan.
9. Establish a Data Pipeline: Construct a pipeline to move data from sources to the Snowflake data warehouse to keep the data up-to-date.
10. Implement the Test Plan: Execute the earlier formulated test plan to ensure the stable functioning of the system.
11. Parallel Operation: Run the legacy and Snowflake warehouses simultaneously, carrying out comparison testing for a specific time period to ensure data accuracy and system reliability.
12. Monitor Budget and Resource Utilization: Keep a track of the budget and billing aspects, adjusting resource allocation for optimization.
13. Final Cutover: Once validation is successful, switch operations to the Snowflake warehouse.
14. Decommission Legacy Systems: Post successful cutover and after ensuring data accuracy in Snowflake, retire the legacy data warehouse system.

FINANCIAL ADVANTAGES OF SNOWFLAKE MIGRATION IN HEALTHCARE DATA ANALYTICS

Moving to Snowflake can yield significant cost savings in the field of healthcare data analytics:

Infrastructure Efficiency: Snowflake's cloud-based nature eliminates the need to set up and maintain physical servers, resulting in substantial infrastructure cost savings, allowing for resources to be allocated to critical aspects of healthcare.
Usage-Based Payment Model: Snowflake's scalability and "pay-as-you-use" model adapt to the volume of data and compute resources needed, preventing organizations from overpaying for unused services or capacity.
Streamlined Operations: By simplifying complex data operations, Snowflake enables data scientists and analysts to focus on generating actionable insights, leading to savings in time and resources, translating into financial efficiencies.

Reduced Administrative Overhead: Snowflake's automated maintenance tasks lead to minimization of administrative costs typically associated with tasks such as indexing, tuning, and backup procedures.
Cost-Effective Large-Scale Analysis: Snowflake's separate compute and storage capabilities offer cost-effective solutions for large-scale healthcare data analysis.
Readmission Rate Reduction: Snowflake-powered predictive analytics can help healthcare providers predict potential health risks, contributing to the reduction of readmission rates, which are a significant cost driver for healthcare organizations.
Improved Patient Care: Advanced healthcare analytics facilitated by Snowflake can enhance patient care, potentially reducing complications and associated costs linked to extended treatments and additional procedures.

While there are initial migration costs, the substantial savings and efficiencies gained in the mid to long term can offset these expenses. The scale of cost savings will vary based on the size and complexity of the existing setup and databases.

IMPLEMENTATION IN OTHER INDUSTRIES

Beyond healthcare, Snowflake's advanced data solutions have been leveraged in various other industries, proving its versatility and robustness.

1. Financial Services: In the financial industry, Snowflake provides a unified and secure platform for all financial data, helping businesses generate real-time insights. It allows for a comprehensive view of a customer's data, improved cross-selling, and risk management. Organizations can efficiently analyze complex, diverse, and large scale data sets for improved decision making, compliance reporting, and risk analysis.
2. Retail: Retail businesses use Snowflake to better understand and connect with their customers. Snowflake can handle high volumes of data from various sources - like sales transactions, website visits, or social media interactions - and provide insights to personalize the customer experience, optimize supply chains, and predict future trends.
3. Media and Entertainment: Companies in the media and entertainment industry use Snowflake to analyze consumer behavior and preferences to tailor their content and marketing strategies. It helps them churn massive volumes of data from different platforms to derive meaningful insights.
4. Education: Educational institutions use Snowflake to analyze student data and track progress, helping them improve curricula and teaching methods. It also aids in research by providing a platform for manipulating and analyzing large data sets.
5. Manufacturing: In manufacturing, Snowflake is used to optimize supply chains, track and predict equipment maintenance, handle inventory, and streamline production plans. It provides a unified view of all the data points required for smooth operation and decision-making.
6. Public Sector: Government entities use Snowflake to manage and analyze large volumes of data for public services, reporting, and decision-making. It aids in areas like public health management, crime prediction and prevention, and resource allocation.

Beyond these, Snowflake is proving invaluable across numerous other industries, anywhere large volume data storage and robust data analytics are required. Its secure, scalable, and easy-to-use platform makes it an essential tool for any data-driven business. For organizations everywhere, data is the gateway to new

opportunities. It's limitless in use and the amount generated is increasing at an exponential rate, largely due to the growing number of connected devices and percentage of the world gaining internet access — both fixed and mobile. [8]

CONCLUSION

In summary, the strategic initiative to enhance big data utilization in healthcare analytics through migration to cloud environments and Snowflake heralds a new era of healthcare progress. This transformative transition empowers healthcare organizations to effectively manage, analyze, and derive meaningful insights from extensive and intricate data, thereby enhancing patient outcomes, improving operational efficiencies, and fostering innovation in healthcare. Tens of billions of dollars are flowing annually into health care innovation and research and development (R&D) efforts. Big areas of focus include virtual health technologies, artificial intelligence (AI) tools to support clinical and administrative activities and data analytics capabilities for clinical predictions and recommendations. [9]

The transition to a cloud-based model brings essential scalability, cost-effectiveness, and flexibility crucial for managing diverse and voluminous healthcare data. Snowflake further amplifies these advancements by facilitating seamless data management, real-time analytics, secure data sharing, and providing an integrated, unified data platform. Unified Data Analytics is a new category of solutions that unify data processing with AI technologies, making AI much more achievable for enterprise organizations and enabling them to accelerate their AI initiatives. [10] Advanced computational methods and data analytics can scrutinize and interpret intricate datasets, uncover patterns, and establish meaningful correlations. When combined with AI's adaptive and predictive capabilities, this has the potential to significantly revolutionize diagnoses, treatment plans, and healthcare operations. AI's proficiency in machine learning equips healthcare professionals with foresight, enabling the prediction of potential health risks and proactive interventions. [13]

The potential unleashed by integrating the inherent capabilities of healthcare big data with the resilient and adaptable infrastructure offered by cloud technology and Snowflake is considerable. This transformative landscape not only contributes to improved decision-making and advancements in patient care practices but also has a positive impact on return on investment (ROI). Hence, the strategic utilization of concurrent cloud migration and Snowflake for data analytics represents a game-changing approach in the healthcare industry, significantly amplifying the value of big data for these pivotal, life-enhancing services.

REFERENCES

[1] Tools for Data Migration https://www.snowflake.com/trending/tools-data-migration/
[2] Why Healthcare Is Moving Toward Cloud Computing https://www.forbes.com/councils/forbestechcouncil/2024/03/19/why-healthcare-is-moving-toward-cloud-computing/
[3] Healthcare analytics https://aws.amazon.com/marketplace/solutions/healthcare/analytics?aws-marketplace-cards.sort-by=item.additionalFields.headline&aws-marketplace-cards.sort-order=desc&awsf.aws-marketplace-healthcare-store-use-cases=*all
[4] Healthcare Analytics https://www.tableau.com/solutions/gallery/healthcare-analytics

[5] What Is Change Data Capture? https://www.informatica.com/resources/articles/what-is-change-data-capture.html
[6] Snowflake - Mobilize Your Data. Join the Data Cloud. https://aws.amazon.com/financial-services/partner-solutions/snowflake/
[7] What is Amazon S3? https://aws.amazon.com/s3/
[8] How do you become a data-driven organization? https://aws.amazon.com/executive-insights/content/how-do-you-become-a-data-driven-organization/
[9] Accelerating The Pace Of Innovation In Health Care https://www.forbes.com/sites/united-healthcare/2023/10/30/accelerating-the-pace-of-innovation-in-health-care/
[10] Unified Data Analytics https://www.databricks.com/glossary/what-is-unified-analytics
[11] Grammarly. (2024). Grammarly (Sep 4 version) [Large language model]. https://app.grammarly.com/
[12] Healthcare Cloud Computing Market https://www.persistencemarketresearch.com/market-research/healthcare-cloud-computing-market.asp
[13] Jinesh Kumar Chinnathambi. (2024). Harnessing Data Analytics and Artificial Intelligence in Healthcare: A New Dawn in Patient Care. *The International Journal of Science & Technoledge*, *12*(8). Retrieved from https://www.internationaljournalcorner.com/index.php/theijst/article/view/173817