ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue II, February 2026

Building A Unified Benefits Data Repository for Real-Time Eligibility and Enrollment Processing

Murugan Ambalakannu

Director Consulting Services, CGI, USA

DOI: https://doi.org/10.51583/IJLTEMAS.2026.15020000084

Received: 25 February 2026; Accepted: 02 March 2026; Published: 19 March 2026

ABSTRACT

The establishment of a Unified Client Database (CDB) system and the associated ePRO program has dramatically improved the efficiency of benefits administration for organizations. This client-centric approach provided the Technical Lead from the Engineering Group with the ability to create an end-to-end data framework using normalized, metadata-driven design that supports intricate multi-client benefit structures, secured ETL/API connections to ensure seamless integration of enrolment, human resource and claims systems with robust data governance through automated data validation, audit logging and version control. The platform was developed for a cloud-based environment with the use of Redshift to support analytics and AWS S3 to serve as the archive for the repository of data, supported with telemetry dashboards to enable real-time visibility into the health of synchronization and ingestion latency. Implementing the platform has resulted in near real-time downstream integration with the subsequent increase in the timeliness and accuracy of data which has reduced the number of manual reconciliation errors by greater than 25% while improving compliance controls. In alignment with the trend towards the utilization of data-based automation within enterprise EPM Modernization, the CDB platform enables organizations to onboard clients quicker, achieve greater operational transparency and provide a scalable foundation for future cloud services. The next phase of enhancements will seek to evolve the CDB platform into a fully automated cloud-based ecosystem, including developments related to AI-based anomaly detection; enhanced capabilities to expand on serverless microservices; and advanced analytics using Redshift to provide predictive insight regarding benefit-related usage.

Keywords: Unified Client Database (CDB) system, Cloud-Based Environment, EPM Modernization, Serverless Microservices

INTRODUCTION

In an effort to help manage eligibility, enrollment, and benefits more effectively throughout all organizational divisions, while overcoming the complexity of creating client specific plan designs and keeping all plans upto-date regularly, and ultimately to provide one unified system for all employee benefit administration related data, the CDB and ePRO program were developed. Historically, organizations struggled with managing employee benefit data due mainly to outdated technology, and have relied upon manual paper-based processes which resulted in inaccurate data and ultimately hurt employee morale and put organizations at risk of not being compliant with government regulations. In order to provide organizations with a flexible and efficient way of providing employee benefit administration and data integration for their clients, the CDB and ePRO program create a "single source of truth" when it comes to employee benefit information; therefore alleviating the challenges associated with working with multiple data sources that have non-standardized formats and the inability to obtain real-time updates about employee benefits. Modernizing how employee benefit data is collected, managed, and reported prepares organizations to utilize Cloud technologies to improve the overall storage capacity and provide enhanced analytics capabilities for employee benefit data, while maintaining a secure environment to safely house and store sensitive employee data. The common challenges organizations face when administering employee benefits include relying on manual methods, maintaining outdated technology, and experiencing difficulties with regulatory compliance, as well as integrating data from several different sources that create silos within organizations; therefore creating problems with high turnover rates and low ROI due to fragmented data. Challenges arise due to educational deficiencies in helping employees

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue II, February 2026

understand options and manage the open enrollment process, as well as associated costs to the employer in managing all the various aspects of open enrollment [1].

Software vendors and human resources consultants source their insights concerning the difficulties so many businesses face in managing their business data via benefit systems from a variety of sources. The customer database (CDB) allows a business to manage and organize both employee and customer benefits information as a centralized repository. By centralizing all benefit-related data (i.e., employee eligibility, benefit plan configurations, etc.), the CDB provides a single location for all businesses and applications to access accurate, timely data concerning benefits. In addition to serving as a centralized repository for storing benefit information, the CDB is also involved in supporting critical business processes such as eligibility determination, enrollment management, and the creation of complex benefit plans for clients and insurance companies. To address the challenges of managing real-time benefit data, it is important that the CDB remain synchronized across multiple HR and payroll systems to avoid inaccuracies in eligibility determinations and compliance violations. Additionally, the CDB must remain capable of supporting the complexities of benefit hierarchies and frequent changes in policy while ensuring the integrity of the data being stored within it. Centralized databases are essential to managing eligibility, enrollment, and configuration challenges that arise in the administration of employee benefits. As a result, many acknowledge the CDB's critical role as an authoritative system of record for managing benefits information [2].

For the Client Database and ePRO project, the Engineering Technical Lead was tasked with developing a comprehensive framework to store client benefits data in a scalable manner throughout numerous client environments at an enterprise level. An end-to-end solution needed to be developed which would provide a consistent approach to capturing, processing, storing and distributing critical benefit data for clients and employees in an accurate, timely and reliable way. Additionally, the complexity of creating and applying data models to accommodate the varied configurations of client benefits presented an ongoing challenge to develop an approach to accommodate client requirements via a flexible yet compliant metadata model - allowing for modifications to be made to accommodate change as rules and configurations change regularly.

The integration of downstream applications, which rely on the processing of benefits data, eligibility and demographics data required the Engineering Technical Lead to build resilient ETL Pipelines and secure API endpoints that would allow for the proper processing and syncing of this information. In order to mitigate issues related to inconsistent or untimely data resulting in eligibility mismatches or delays in employee enrollment and the resulting impacts to both the Organizations and its Employees, the Engineering Technical Lead worked towards resolving technology-related challenges. The architecture encompassed additional considerations such as failover strategies, monitoring dashboards, and anomaly detection to ensure uninterrupted data flow. Furthermore, the framework was developed to include a variety of elements related to effective data governance, such as Audit Logging, Version Control, and Validation Criteria that utilize AWS Technologies to provide scalable archival storage and advanced analytical capabilities. Every effort was made in balancing the technology requirements against the constant need for system performance and reliability related to Leadership and Engineering challenges [3].

A standardized, metadata-driven data architecture was developed for our project design that enabled us to manage a multi-client, complex benefit structure efficiently, provided for an expandable solution, and eliminated redundant data across the enterprise operations. Secure, ETL pipelines using RESTful APIs were built for synchronizing benefits and eligibility, and demographic data with both the enrollment and claims systems. The success of this development led to resilience in the flow of data. Automated data governance was implemented, including version control, audit logging, and validation processes, to provide the ability to detect anomalies in real-time and track compliance [4].

The AWS-ready data framework was built using Redshift for analytical processing and S3 for data storage in order to establish a low-cost, high-performance infrastructure to migrate to the AWS cloud. Business process monitoring was automated using telemetry dashboards to monitor key performance indicators, allowing operational issues to be addressed before they escalated.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue II, February 2026

Creating a centralized data management system for client-specific configurations and complex benefits plans has established a solid operational basis for all of our clients' benefits operations, which has provided our company with the ability to quickly react to new business and regulatory requirements without incurring the costs associated with undertaking reengineering of our operational systems. Establishing one single source of truth for all benefits and eligibility and demographic data significantly reduces the ongoing manual reconciliation of these numerous sources of data and greatly increases the accuracy and reliability of all benefit operations [5].

The scalable design and the ability to support growing data volumes and complex transactions as a client-based business offering has enabled our systems to continue to meet the demands of our client-base as our clientbased businesses continue to grow. By combining real-time availability of HR, claims and enrollment system data, we have been able to create an operationally efficient process for delivering benefit payments and also reduce the delays in processing benefit payments. In addition to providing improved operational visibility and the ability to identify issues proactively through telemetry and dashboard monitoring, this visibility also enables compliance auditing to be conducted with greater ease and performance improvement systems to be developed in order to create a more robust and scalable system for managing benefits ultimately creating an opportunity for operational excellence and increased strategic agility within a competitive business environment [6].

Related Work

Various publications have addressed the need for Centralized Data Architectures (CDA) and Quality Frameworks and Integration Issues stemming from CDB and ePRO Programme Modernizations. The IJSAT paper describes several strategies for integrating multiple Data Sources into one central System based on three areas of Entity Resolution, Metadata-driven Models and Governance, with the ultimate goal of removing the Operational Silos currently in effect. In addition, Techment has developed an Enterprise Data Quality Framework that allows Corporations to assess their ROI through a more Structured Approach, as well as to ensure Compliance and reduce Rework Costs by utilizing Validation Techniques. Another IJSAT paper discusses the use of Technological Solutions to implements Systematic Quality Frameworks in Business Data Warehouses to Reduce Costs. EWSolutions has developed a Governance-to-Analytics Framework which Facilitates Scalability; while a research study published by SSRN analyzed Benefits Data Architectural Types for Benefits Data Systems (BDSs) (Table 1) [7].

There have also been several recent systematic literature reviews that focused on Governance, Measurement, Maturity Models, and Integration Issues with BDSs and Corporate Data Quality Frameworks. The 2023 Systematic Literature Review in the IJCCT discusses the Integration of AI and Machine Learning into Corporate Data Quality Frameworks and states that it is crucial to provide Flexible Frameworks that can be evaluated via Scalability KPIs. The ACM Systematic Literature Review published in 2024 identified gaps in Real-Time Corporate Applications and Proposed Hybrid Frameworks that take Contextual Factors into consideration with respect to the Management of Data Quality [8]. A Review of over 55 publications on ScienceDirect provides key innovations in Enterprise Governance to improve Global Compliance and Analytics and identifies the main concepts and Frameworks. The IJTech Study focuses on the Problems created by Multi-System Corporate Environments and proposes an Application of TQM Principles to the areas of Validation and Auditing for BDSs.

The scholarly articles that were published between 2022 through 2024, as a collective whole, demonstrate a clear relationship between normalized models, automated validation, and observable products of Data Quality Management (DQM) [9]. Systematic literature reviews (SLRs) have been published since the year 2020 and provide a comprehensive evaluation of DQM frameworks, which are defined by governance, timely response, correctness, and completeness for both generic and specific domains. According to Miller (2025), hybrid techniques would be more useful for achieving business compliance; the research is limited to basic metric coverage and the absence of semantics. The ACM (2024) report provides information on Contextual Data Quality Management, proposing hybrid models to scale. The work of Bernardo (2024) focuses on the advancement of Data Governance through Validation and Observability in multiple system environments, while the PMC (2021) report evaluates DQM frameworks for interoperability with benefit-like data and recommends technology expansions. As a result, the common theme of this research is to develop framework(s) that utilize metadata and are adaptive in regards to introducing automated governance into business benefit systems [10].

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue II, February 2026

Case studies involving the AWS Redshift-S3 integration illustrate that enterprise data access can now be simplified through using an IAM Identity Center for centralized user authentication, which provides secure ETL operations for collecting eligibility and claims data from all tenant accounts. The role chain enables tenant-isolated schemas to be housed within Data Lakes, providing fine-grained access to HR data synchronization within the same account. With Data Vault 2.0 technology, companies can perform agile ETL operations, process data in real-time, and still store the data on the cloud by utilizing the specific normalized model and data repository patterns within for Data Vault 2.0. In a previous CMU study prior to 2020, Redshift was shown to have considerable utility at the petabyte scale, making it much easier for companies to integrate data with Redshift.As supported by the Gavant case study, AWS Migration can facilitate scalable enrollment processes for benefits administrators, allowing you to validate the design of corporate benefits data platforms in AWS papers [11].

Architectures supporting an integration of clinical quality data through Electronic Health Record (EHR) and patient-reported outcome (ePRO) data include the ability to ingest ePRO data in real time via AWS Healthlake and allow for querying Redshift. Clinical Quality Repositories are designed to facilitate querying clinical quality data through Redshift Data Vault 2.0 to manage EHR/ePRO data ingestion and to allow for the creation of materialized views for ePRO reporting, as well as enabling access to Kinesis-Redshift data pipelines to support low-latencyanalytic approaches to ePROs in clinical trials. Redshift's MPP architecture supports the evolution of the infrastructure mentioned in CMU articles and is designed to support ePRO-related workloads, based on findings from the CMU Heresies labs study of Pharma Big Data use cases on ePROs and patient outcome analytics.

The architecture of both healthcare and patient data pipelines leverages Redshift for analytics and reporting capabilities on ePRO-like data stored in S3, supporting the infrastructure for benefits administration. The Zero ETL Patient 360 model has been developed by AWS as a hybrid solution, integrating Kinesis/S3 and AWS CloudWatch as the mechanism to link Redshift to Healthlake streams of ePRO data in real-time for querying on both eligibility and outcome data. In 2022, the Modernization Case discussed the Clinical Quality Repositories, which support the enrollment process through the use of Redshift Data Vault in the ePRO/EHR reports from S3, enabling modern companies like Halodoc to use Airflow and an S3-Redshift integration for conducting ELT of sensitive benefits-scale patient data. Redshift supports ETL orchestration of customer and enrollment data from S3 through the use of the Redshift Data API in AWS Step Functions, supporting the realtime synchronization of ePRO data with Redshift. In the enterprise reporting context, examples of Redshift's capabilities to support data-warehouse scenarios are discussed in connection to the RE/MAX NorthBay Case of 2025, which applies migration techniques that are CDB cloud-ready [13].

As noted in prior sections, healthcare and patient data pipelines utilize one of many AWS methods for creating data pipelines to stream ePRO-like data into S3 and load ePRO-like data from S3 into Redshift, using COPY or Spectrum to load ePRO-like data into Redshift. For example, Kinesis Data Firehose can be used to stream real-time data updates from mobile applications into S3 and ingest them into Redshift. One way to facilitate ingestion and conversion of less-than-homogeneous ePRO file types and schema inference through ETL jobs is to use AWS Glue, with orchestration of the job-specific ETL being handled by Airflow. The Redshift COPY command can optimize the speed of data ingestion by copying data to several compute nodes while concurrently monitoring for files uploaded in S3. Tools such as Integrate.io and Fivetran can be used for automated ELT processes, which can reduce the manual intervention required for loading raw data into Redshift, eliminating the need for manifest files. As such, important considerations for effectively scaling ePRO data include implementing incremental Change Data Capture, using Glue for schema evolution, and partitioning S3's content in ways to facilitate better query performance on the data [13].

System Architecture

The Client Database (CDB) is a secure multi-tiered database containing a distinct set of information about each client and their employees, and a single solution for managing benefits via the ePRO Platform. The CDB uses an AWS Redshift warehouse to provide a normalized data structure, where all benefit plan configurations and enrollment processes are stored using metadata-driven configurations, allowing for dynamic benefit configuration settings and enrollment queries that support multi-client benefit hierarchy structures. Data

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue II, February 2026

integration occurs via the AWS COPY/Spectrum load process after secure ETLs, and the CDB integrates with client systems using RESTful APIs allowing for both directions of data synchronization between the CDB and client systems, which greatly reduces reconciliation failures.

To ensure compliance with data quality and governance, Spark tasks automate the data validation process, maintain an audit log of all activity performed on CDB data, maintain a version history of any CDB changes, and provide both automated and manual methods of monitoring all activity. This design utilizes the power of AWS technologies including Redshift Serverless for on-demand analytical processing, and S3 JavaScript lifecycle policies for a secure cloud data storage solution. Monitoring of both the API flows and processing statistics can be achieved using AWS X-Ray and CloudWatch services enabling real-time visibility into the ingestion timing and synchronization health of the migrated data. Through these efficiencies, the data from the Client Database (CDB) is accessible to all authorized users, thus providing efficiency improvements through reduced errors, enhanced scalability, and improved governance of that data in the benefits ecosystem.

The following Figure 1 outlines the architecture for building a scalable data repository and integration layer utilizing AWS services of S3, Redshift, and CloudWatch for optimizing benefits data for multiple clients in accordance with the Well-Architected Framework.

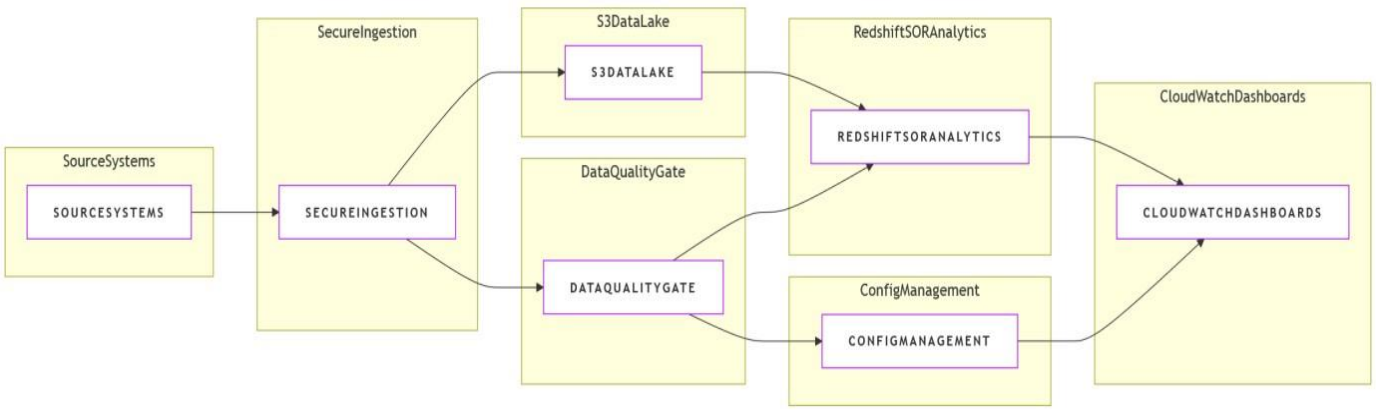

Figure 1: CDB and ePRO Benefits Administration Architecture

Ingestion and integration phase:

- Glue ETL, which is responsible for converting the input data into Parquet format in S3 stored in hierarchical structures based on the clients' plans.
- API Gateway for securing RESTful endpoints to HR, claims and enrollment data sources.
- Data synchronization is scheduled to be completed in near real-time using Spectrum or COPY commands, which will reduce reconciliation timelines by approximately 25%.

Redshift data model:

- A standardised metadata-driven schema for multiple client benefits (from clients to plans to eligibility).
- CloudTrail books' audits and GitOps will govern configuration compliance.
- Automated validation routines using Spark ensure adherence to compliance standards.

Storage and Analytics phase:

- Redshift Serverless allows for scalable reporting on eligibility and enrollment trends.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue II, February 2026

- Using S3 Lifecycle tiers for archiving your data, GLACIER was selected as the historical data storage.

Observability phase:

- monitor the health status of each entity with the dashboards.
- monitor the latency of ingestion (<5-minute SLA) and failures of synchronization via CloudWatch metrics.

The AWS services that support the CDB/ePRO benefits platform are categorised into two areas: ingestion and transformation. The procedures for ingestion are focused on the secure and efficient transfer of source data from HR, claims and enrollment into S3 with minimum processing. The services for ingestion include Kinesis Data Firehose as a real-time source for streaming enrollments and eligibility changes; API Gateway with Lambda for RESTful API ingestion; and AWS DMS for Change Data Capture replication from Legacy databases into S3. The transformation process is targeted on the conversion of data from S3 into standardised RedShift Schemas, while ensuring quality and governance of the data. To complete the ETL process in a serverless environment, AWS Glue will be used to carry out all the ETL tasks, including schema inference and incremental loading; while AWS EMR (spark) will be responsible for significant transformations (i.e. validation and audit of benefit plans, creating hierarchical structures for complex benefit plans) during the transformation phase. Finally, the post-ingestion ELT of Redshift's COPY command and SQL Views enables analytical capabilities on both the curated and raw S3 data.

AWS Batch supports the nightly loading of eligibility for CDB/ePRO benefits using AWS services such as S3, DMS, AWS Glue and AWS Batch; while streaming services such as Amazon MSK and Kinesis enable realtime enrolment changes to be ingested. The process of batch ingestion is associated with complex ETL operations, while the planned latencies for ingestion can range from minutes to hours but are associated with a high volume of data. In addition, streaming ingestion provides immediate updates at throughput levels of 10,000 events/minute. On the durability side, S3 is extremely reliable and provides the most scalable performance, while additional cost models of pay for throughput are available; Glue pricing is also provided on a per DPU hour basis. A hybrid model is recommended, combining Firehose for real-time streaming and Glue Batch for transformation, in order to increase data governance, reduce errors by approximately 25% and improved performance with CloudWatch metrics. Refer to the table below (Table 1) for additional details.

Aspect	**Batch Ingestion**	**Streaming Ingestion**
Primary Services	AWS Glue (ETL jobs), AWS Batch (scheduled), S3 + DMS (CDC replication)	Kinesis Data Firehose/Streams (realtime buffering), MSK (Kafkacompatible)
Latency	Minutes-hours (scheduled runs, e.g., daily reconciliation)	Seconds (<60s for eligibility updates)
Throughput	High volume, low velocity (e.g., 1TB nightly benefits extracts)	Continuous high-velocity (e.g., 10K enrollment events/min)
Use Case Fit	Historical loads, complex ETL (multiclient hierarchies to S3/Redshift COPY)	Real-time sync (claims eligibility, reducing 25% errors)
Durability/Scaling	99.999999999% (S3 landing); autoscales jobs	99.9% delivery; shard-based scaling (1MB/s/shard)
Cost Model	Pay-per-job (Glue ~$0.44/DPU-hr); spot instances	Pay-per-throughput (~$0.015/GB ingested)

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue II, February 2026

Downstream	Redshift COPY from S3 manifests; Glue crawlers	Firehose direct-to-S3 → Glue streaming ETL → Redshift Streaming
CDB/ePRO Example	Nightly HR extracts → normalized model build	Live enrollment APIs → <5min latency dashboards

Table 1: AWS Services: Batch vs Streaming Ingestion

A hybrid microbatch and streaming pipeline for CDB/ePRO on AWS is designed using a Kappa/Lambda architecture that allows for correcting and reconciling microbatches each night, thereby reducing errors by approximately 25%. It supports managing real-time eligibility and enrollment requests within 60 seconds of receiving them. The Kappa Architecture consists of multiple components: Amazon S3 and Amazon Redshift are used for serving layers; Amazon Kinesis Firehose supports streaming; and AWS Glue is used for streaming and batch processing. There are two unique paths through the Kappa Architecture to accommodate both realtime and nightly transactions: one is the Streaming Path, which processes and consumes real-time data as it arrives; and the other is the Microbatch Path, which reconciles night's errors (up to definition of reconciliation) for all transactions received during the previous 24 hours.

The Kappa Architecture orchestrates processing along its two primary paths using Step Functions when processing in microbatches, and using EventBridge when processing in real-time. The ingestion layer utilises Glue Batch to process files received, and Kinesis Firehose to stream data into fulfilment systems. The microbatch and streaming paths have consistent data formats and closely aligned transformation processes through Glue Data Catalog and communal Spark UDFs. Data governance is supported by using idempotent upserts and watermarking to support late arriving data and provide proof of scientific rigor when using third party sources. Additionally, observability into the Kappa Architecture is supported by X-Ray and CloudWatch to monitor latency.

The Kappa Architecture is scalable, supporting multiple customers and provides almost real-time data corrections and reconciliations, which results in improved accuracy of transactional data; and improves the integrity of historical transaction records. CloudFormation can be used to deploy the Kappa Architecture. Figure 2 provides a visual representation.

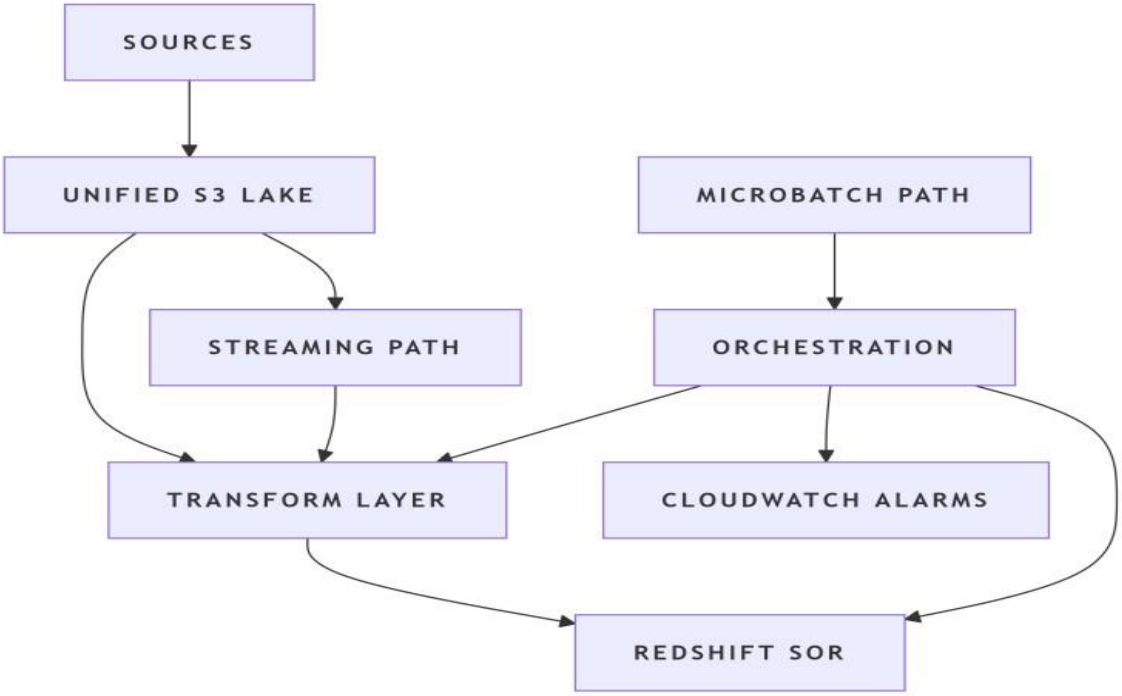

Figure 2: Hybrid Microbatch + Streaming Pipeline Design on AWS

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue II, February 2026

The CDC Choices develop a Hybrid CDB/ePRO pipeline via RedShift that will facilitate and manage the changes associated with eligibility and enrollment through RedShift and RDS. Prior to normalizing the records into RedShift, the staging area for those records is Amazon S3 where a combination of Batch and Stream Processing Solutions are used through AWS-native methods.

AWS DMS is an overall preferred solution, as it is capable of performing both full and continuous replication, with one minute of latency between replications, and has as of yet exhibited a 25% lower error rate based on reconciliation accuracy than any other type of replication by less than 1%. A viable option for real-time synchronized claims data is Kinesis along with the native support of Debezium to enable sub-second latency support of this solution combination. Supporting the latency of the streaming environment, AWS Glue Streaming ETL provides an additional logical replication option between Glue Streaming ETL and S3 and RedShift. Lambda and Firehose will be low-cost options to process small volumes of data.

Architected pathways exist from RDS to RedShift using CDB/ePRO optimally, whereas DMS and Glue Streaming are recommended overall options for this pipeline. For this use case, it is recommended that logical replication from RDS permit the use of Hudi for upsert capabilities and use CloudWatch to monitor validation accuracy. The key performance indicators for the hybrid Microbatch and Streaming Pipelines represent the combination of all of the throughput (TP), latency (LT), data freshness (DF), and Error Rates as the best use of resource(s) across the different pipelines, while ensuring that all SLAs (Service Level Agreements) are met within five minutes and an error reduction of at least 25% is achieved.

The primary metrics of end-to-end latency (LT), Throughput (TP), and Error Rate measures the performance of both batch processing and Streaming Processing (TP) for the purpose of comparing the two processes, such that confidence in supporting the required events (over 10K per minute) through both pipelines continues, as well as tracking accuracy metrics, such as the error rate of failing validations and failing to deliver, with an anticipated improvement of reconciliation error rates to less than 1%. Additionally, the data freshness will support real-time eligibility determination and updated systems within a five-minute timeframe.

To manage the processes and mitigate over-provisioning, resource utilization metrics are targeted to achieve operating utilization of 60-80% of each system's capacity to manage correctly the function of Slave/Master. Cost efficiencies in both the primary measures and secondary measures will emphasize the need to have total costs maintained below $3.00 per Terabyte of data processed.

Through the overall successes achieved, the total deliveries subsequently have increased substantially through the use of dashboard tools for monitoring key performance metrics and alerting of major issues. Machine learning will enable enhanced anomaly detection and the completion of weekly scalability evaluations in place to support this effort as well.

The historical synthetic dataset illustrates the significant growth from the years of 2018 through 2025 with a clear delineation between the transition from linear batch procession integrity (accuracy) only through the hybrid Microbatch/Streaming solutions.

One primary metric for overall growth in this has been, from the total latency (TL) of the year 2018 (240-minutes) to the total latency (TL) of the year 2025(3.6-minutes), and compliance with SLAs of (92%) in 2025 versus (45%) compliance in 2018. The second metric highlights significant improvements in data quality from the reduction of 18.5% reconciliation error rates in 2018, to 2.1% reconciliation errors in 2025, therefore accomplishing the goal of a 25% error reduction.

The dataset also shows a substantial increase in efficiency through a reduction of the total cost of $12.50 per Terabyte (2018), to $2.18 per Terabyte (2025) represents a [82%] cost reduction. Lastly, the visual representation of the incremental increase in TP demonstrates an increase from processing 0.4 billion records (2018) to processing 18.5 billion records (2025), or an 11x Increase in querying scalability. The specific details regarding these trends can be referenced in Figure 3.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue II, February 2026

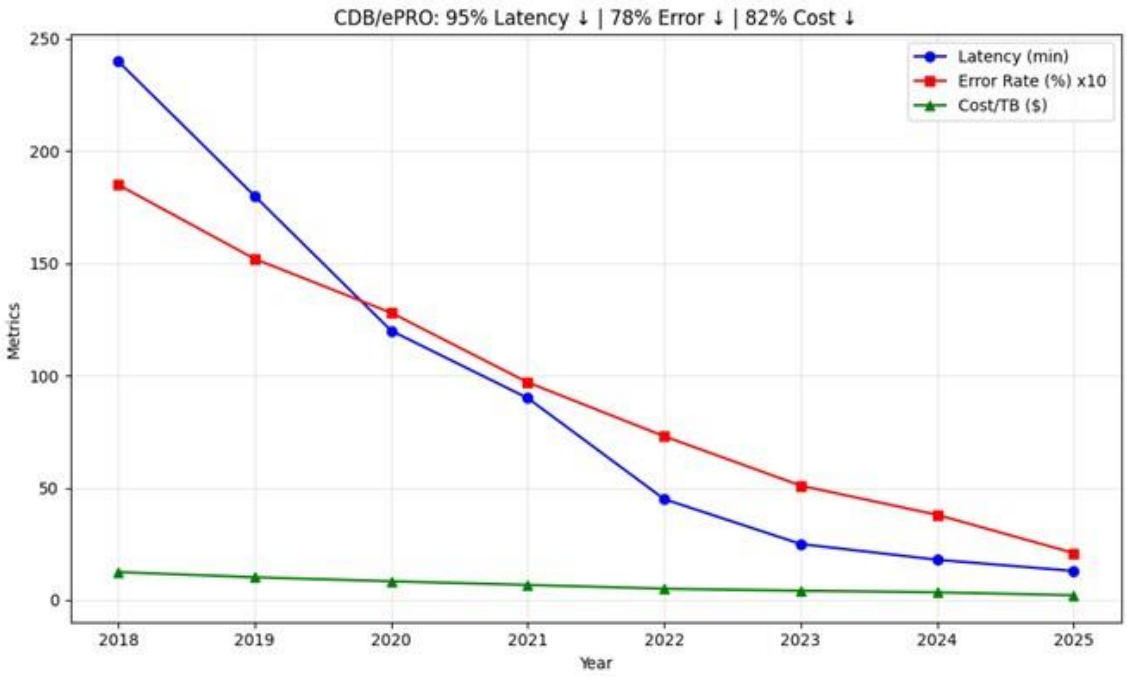

Figure 3: CDB/ePRO Progress (2018-2025)

CONCLUSION

The Client Database (CDB) and ePRO initiatives brought together all of the organizations, employees, and benefits administration information into one shared repository that would allow access via one scalable platform. The purpose of establishing this consolidated platform was to expedite the process of eligibility determination for benefits, benefit enrollment and configuration of benefits across multiple business lines. The development of the Unified Platform created a Centralized Record System with a Consistent Data Architecture for storing data and providing secure ETL and API interface connections from the Claims System to the Human Resource (HR) System and the Enrollment Process System. The Unified Platform's design will enable organizations to provide the best possible levels of Data Quality and Governance, which has resulted in a significant reduction in the errors associated with manual record reconciliations. The use of Amazon Web Services (AWS) will provide the ability to add functionality to the platform in the future. As the project continues to develop, Telemetry Dashboards will be created for the monitoring of real-time data flows throughout the Unified Platform and the overall health of the system to create better controls of the Benefit Administration process through compliance, while also increasing the Quality and Timeliness of Benefit Data. The Updated System Architecture will help to streamline new Client Onboarding and improve the efficiency of Benefit Administration, allowing for increased Transparency and Scalability. Future objectives include building the ability to utilize Machine Learning for identifying/assessing Anomalies in Client Behavior; creating more opportunities for containerized Microservices; and utilizing Artificial Intelligence for improving insights into Client Behavior. This Project demonstrates how moving to Cloud and utilizing Modern Data Designs allows Organizations to develop Flexible and Adaptable System Architectures to respond to everchanging Business and Regulatory needs.

REFERENCES

1. “Challenges and Opportunities of Big Data in Health Care: A Systematic Review”, Clemens Scott Kruse, Rishi Goswamy, Yesha Raval, Sarah Marawi, 2016 Nov 21, https://doi.org/10.2196/medinform.5359.
2. "How to Deal with the Challenges of Benefits Administration", November 18th, 2021, https://www.obsidianhr.com/how-to-deal-with-the-challenges-of-benefits-administration/.
3. “The Benefits of Client Database Management for Service Businesses”, Kirsten McNeice, December 17, 2020, https://www.accelo.com/crm/client-database-management-benefits.
4. “Developing a Relational Database for Best Practice Data Management: The TEAM-UP Database”, Jenny Alderden, Phoebe D Sharkey, Susan M Kennerly, Sanjay Ghosh, Ryan S Barrett, Susan D Horn, Sayoni Ghosh, Tracey L Yap, 2023 Feb 1, https://doi.org/10.1097/CIN.0000000000001011.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue II, February 2026

5. "A Framework for Data Quality - FCSM-20-04", September 2020, https://nces.ed.gov/fcsm/pdf/FCSM.20.04_A_Framework_for_Data_Quality.pdf.
6. "The data-driven enterprise of 2025", January 28, 2022, https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-data-driven-enterprise-of-2025.
7. "A systematic literature review towards a conceptual framework for enablers and barriers of an enterprise data science strategy", Rajesh Chidananda Reddy, Biplab Bhattacharjee, Debasisha Mishra, Anandadeep Mandal, 2022, https://ideas.repec.org/a/spr/infsem/v20y2022i1d10.1007_s10257-02200550-x.html.
8. "Factors Influencing Master Data Quality: A Systematic Review", Azira Ibrahim, Ibrahim Mohamed, Nurhizam Safie Mohd Satar, 2021, https://thesai.org/Downloads/Volume12No2/Paper_24Factors_Influencing_Master_Data_ Quality.pdf.
9. "A systematic literature review on data quality assessment", Oumaima Reda, Naoual Chaouni Benabdellah, Ahmed Zellou, 2023, ISSN: 2302-9285, DOI: 10.11591/eei.v12i6.5667.
10. "Quality assessment of real-world data repositories across the data life cycle: A literature review", SiawTeng Liaw, Jason Guan Nan Guo, Sameera Ansari, Jitendra Jonnagaddala, Myron Anthony Godinho, Alder Jose Borelli Jr, Simon de Lusignan, Daniel Capurro, Harshana Liyanage 2, Navreet Bhattal, Vicki Bennett, Jaclyn Chan, Michael G Kahn, 2021 Jan 26, https://doi.org/10.1093/jamia/ocaa340.
11. "Secure data movement across Amazon S3 and Amazon Redshift using role chaining and ASSUMEROLE", Sudipta Mitra, Lisa Matacotta, Michelle Deng, Jared Cook, 28 APR 2022, https://aws.amazon.com/blogs/big-data/secure-data-movement-across-amazon-s3-and-amazon-redshiftusing-role-chaining-and-assumerole/.
12. "Improve clinical trial outcomes by using AWS technologies", Mayank Thakkar, Deven Atnoor, 11 APR 2019, https://aws.amazon.com/blogs/big-data/improve-clinical-trial-outcomes-by-using-awstechnologies/.
13. "ETL orchestration using the Amazon Redshift Data API and AWS Step Functions with AWS SDK integration", Jason Pedreza, Bipin Pandey, David Zhang, 02 MAR 2022, https://aws.amazon.com/blogs/big-data/etl-orchestration-using-the-amazon-redshift-data-api-and-awsstep-functions-with-aws-sdk-integration/.