ESP-JETA
ESP Journal of Engineering & Technology Advancements
ISSN: 2583-2646/ Volume 6 Issue 3 September 2026
Paper Id: JETA-V6I3P114/ Page No: 117-122

Original Article

Unified Data Modeling in GCP and AWS Environments for Retail Decision Systems

Shires Gorgilli
Southern University A&M College, Baton Rouge, LA

Received Date: 20 August 2026 Revised Date: 25 August 2026 Accepted Date: 01 September 2026

Abstract: *Many retail organizations are finding that the Google Cloud Platform (GCP) and Amazon Web Services (AWS) "hybrid cloud" is now enabling them to scale and perform well on their decision systems. With their unique architecture, different storage models, and processing capabilities, however, they add fragmentation, duplicate and inefficiencies which could impact analytics-driven decision making. It outlines a unified data-modeling framework, backed by creating schema conventions, constraint and governability of data across clouds and making data interoperable when conducting retail analytics across clouds. It eliminates data redundancy, latency, supports compliance and provides time to insight and efficient operations by leveraging the necessary canonical schemas, federated query engines, standard formats and metadata-driven governance. This paper brought up light on the theoretical views of Unified modeling approach and suggested an improved modeling approach to implement the same in the retail world, which could help the retailers to have unified and parallel access to the similar compliant Analytics environment with GCP and AWS environments; thereby, supporting them in taking better decision making with adequate understanding.*

Keywords: *Unified Data Modeling, Retail Analytics, GCP, AWS, Hybrid Cloud*

I. INTRODUCTION

In the present day, retail operates in a highly complicated and information-rich context, and has essentially been transformed into a data-driven and real-time analytical decision-making firm to thrive in today's business landscape. As the adoption of hybrid clouds continues to rise, retail companies leverage hybrid cloud solutions running on Google Cloud Platform (GCP) and Amazon Web Services (AWS) to maximize the value of their data, performance and scalability. The challenge to deal with these heterogeneous tools at the cloud-native level, however, is enormous for consistency, availability and interoperability of data. However, these challenges can be addressed if a single type of data is available that can easily be combined with other data in the GCP and AWS environments, and provides a consistent schema design and governance, enabling a fast and reliable processing and analysis of retail data [1][2][3]. Because of the differences in architectural paradigms of various clouds, it poses problems in a hybrid cloud. Whether it's data storage format, query layers, orchestration capabilities, as well as other features, GCP services—such as Big Query, Cloud Storage, and Data Flow are, to varying extents, different than AWS services such as Redshift, S3, and Glue. Without a shared model approach, retail organizations have to provide multiple versions of their data model, different models for their metrics, latency between platforms and more. All these inefficiencies degrade the current performance of key retail tasks like customer demand forecasting and the implementation of promotional campaigns or cutting the supply chain, where timely and suitable information is a major problem [4][5][6]. There is an opportunity to define in a single Data Model the common "data semantics" and "data structure" for the individual technologies like GCP and AWS. Data can be mapped to standard models that can use AI to implement Cross Platform Data Warehouses, Federated queries as well as AI-driven Analytics pipelines without the need to re-model data for each data platform. A business strategy can help save a business money on operating expenses, save on the integration process, reduce delivery time to insight for the retail decision makers [7][8][9]. Another added advantage apart from aligning information technology is integrated modeling which can be used to improve compliance and data governance. Retailers will be at greater ease with controlling sensitive customer information to which GDPR/CCPA and beyond will apply. Access to different cloud platforms of the same data-type might be easier; it should help ease access to data that have the same data classification as well as tracking and access control and make the traction easier towards audit and lower compliance-related risks on both cloud platforms should help complement that as I would add too. Retailers can integrate metadata-based governance layers, and develop a unique modeling to achieve centralized control and access to foundational data to allow their decision systems to operate at an optimal level while adhering to the legal and ethical requirements [10][11][12]. This paper provides in-depth discussion on unified data modelling that can be used for implementation of retail decision systems on the GCP and AWS platform. It starts with in-depth knowledge of multi-cloud retailing systems' technical and functional challenges, followed by the retailing task's context to which the need to have a unified model is introduced. It further delves into all the key principle, tools and techniques of developing shared data models with it and it discusses the implementation framework and experiences from the field. The research offered will provide a blueprints for retail businesses to streamline end-to-end analytics, add more accuracy to their decisions and unlock the best of hybrid cloud [13][14][15].

The next part investigates the issues and problems associated with creating a retailing business model comprising a mixture of GCP as well as AWS; the problems and technicalities that have been motivating applications of retail toward a unified modeling view to set up scalable, robust and accurate decisions systems in retail [16][17][18].

II. CHALLENGES OF MULTI-CLOUD RETAIL DATA ECOSYSTEMS

Externally, they do give the retailers the chance to become redundant, specialise and be flexible but they also create siloes of operations and can create the hybrid structure of GCP and AWS, harnessing the best fit of both. The cloud platforms have also been separated from the model, which leads to different ways they view the data, store data, and offer a query optimizer; there are a number of redundancies, unbalanced measurements and unseen views of the data. For example, while Big Query is columnar, best optimized for server less queries and Redshift is clustered, best optimized for SQL queries. This is because the optimisation and index can vary if the same data is used to perform different queries. This is one reason why it is challenging to conceptualize a consistent reporting/predictive analytics solution for this variation: some tweaks to one system may not mesh with the other [19][20][21]. Another large danger is a data duplication. Retail data is from the point-of-sale, e-commerce, supply chain and third party data pipelines, and is processed in parallel on both GCP and AWS. In contrast to the default team model, teams can create their own model; some of the calculations behind KPIs and metrics that will be reported, such as the customer lifetime value or inventory turnover, can also differ per team, again allowing for a considerable amount of storage space. These kinds of discrepancies can work against stakeholders' trust in the analytic product, and it might be necessary to have enough stakeholders involved in the decision-making process as resolution time may necessitate further disagreement and/or attempts should be made to resolve any discrepancies between the output of the analytics [22][23].

Slowness and inadequate query performance is one of the reasons for this growth in complexity. These retail decision systems close offline and can be used in real-time, which is used to offer dynamic pricing, offers and stock. Both joins and aggregations can be quite compute-intensive and slow when both types of platform are used as data sets to be joined or aggregated, where large data (e.g. multi-year transaction logs, product catalogues) is being conjoined for ad hoc purposes on the platforms' architectures or from one to more platforms. This agility can be easily simulated with an incompatible file format such as Avro to Parquet (defaults) and schema-inconsistencies can occur, as well as inhomogeneous-partitioning issues. Also, hybrid cloud deployments experience governance deficiencies [24][25]. The sellers will be responsible in handling the information belonging to their customers, intensively monitoring customers' information (such as encrypting, safeguarding access to their information, complying with new regulations concerning the privacy of customers' information, etc.). Sellers need to implement stringent measures to manage customer data—including access management, compliance to the new privacy regulations and encryption methods—and be able to manage it responsibly. These differences in the nature of the metadata systems, role based access systems, and tools to trace lineage systems in GCP and AWS can be their own issue in governance to solve.

Figure 1: Challenges Of Multi-Cloud Retail Data Ecosystems: A Visual Breakdown of Key Obstacles Faced by Retailers, Including Rising Costs, Data Silos, Security and Compliance Complexities, And A Growing Skills Gap.

No model compliance audit can be as time consuming and is affected more by a data stewardship – this creates a higher regulatory risk and reputation risk [26][27]. This way, both the technical and operational hurdles emphasize the need to have a single data modelling framework that would allow to remove platform-specific quirks, have one source of truth and allow faster and more efficient access to data for decision systems as shown in Figure 1. Based on this knowledge, the next section discusses the concepts and methods that are applicable to unified data modeling, and demonstrates the ability of the proposed approach to partially overcome the problem of data fragmentation, as well as enhance the scalability and performance of the retail analytics [28][29][30][31].

III. PRINCIPLES AND TECHNIQUES OF UNIFIED DATA MODELING

Multi-cloud retail ecosystems aren't a problem for a single domain, discipline or approach to data modelling – it's a multi-cloud challenge. Unified data modeling starts with abstracting key business elements like customer profiles, transactions, inventory hierarchy, as well as their supplier relationships into a common semantic system that can be applied consistently to AWS and GCP environments, as shown in Figure 2 using common semantic data modeling. Now, the retailers can develop schemas and transformation rules for each of their individual platforms, instead of having to develop one optimized model for each, and their analytics teams can now base their efforts on one "version of the truth," with the data potentially stored in multiple places [1][3][7]. The central issue to this approach is schema standardization. Common data types, name definitions and group definition of entities and relationships can be easily preserved in unified modeling systems for seamlessly integrating between services (Big Query, Redshift etc.) with minimal refactoring effort. More importantly, the standardization can be enhanced by using several software models, where the most appropriate ones are selected in order to abstract their information from the data structure towards objects relevant to the business. These layers enable data engineers to design scalable ingestion pipelines, and business users to design easy-to-understand data models for reporting and seeking advanced data analytics, without worrying about the wrong interpretation of data (metrics) [8][10][12].

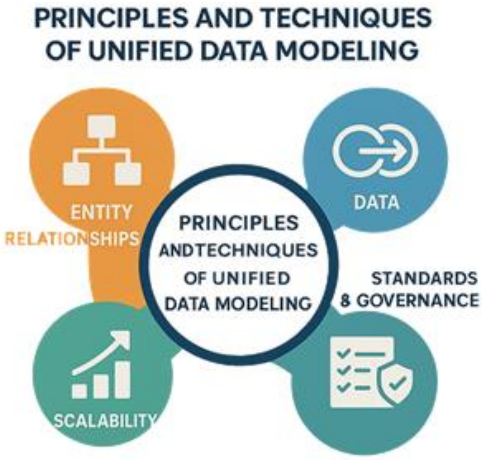

Figure 2: Principles And Techniques of Unified Data Modeling – Introduces Important Concepts in UDM Like the Entity Relationships, Data Integration, Standards & Governance, And Scalability, To Build a Consistent and Efficient Data Model.

One of the other is the cross platform one: format/federated query is synchronized. This wisely would reduce friction between the different clouds that any organization might be using in cross cloud processing if closed encoding formats were avoided and formats like Parquet and ORC, accepted by wide swaths of the user base, were used. An alternative way to achieve the same is to replicate datasets, which uses a query engine tool or virtualization layer, and is more storage and performance efficient. There are other tools like a query engine, virtualization layer, and so on that replicate data sets to query Big Query / Redshift with a single query. Federated capabilities such as these can be a huge benefit in retail decision applications where certain subsets of e-commerce transactions (deep server-side), or serious data lakes (historical inventory or demand forecasting) are being utilized.

A higher level of harmonized models integrity in the frame of metadata-Gov. Higher level of consistency of the harmonized models within the framework of metadata-Gov. These are not limited to GCP and are an added layer of governance specified in the Catalos, schema versioning, data lineage and access control. These catalogues work along with the role-based access controls and automated audit trails, which guarantee adherence to privacy regulations and, at the same time, keep analytics highly agile. When the metadata rules are added, integrated into the system is the automatic dissemination of access restrictions, quality control to prevent sensitive information from customers, and information relating to transactions [18, 20] are included. Finally, automation and reusability of frameworks simplify the deployment and maintenance of the models. By deploying using infrastructure-as-code and declarative modeling languages, the same scripts can be used to deploy the same schema, transformations, and governance policies to both clouds. This will lead to lower operational costs, increased reliability and produce uniformity in the quickly changing retail data space.

Let's now focus on a few principles we followed to align data modeling between the two cloud providers, GCP and AWS, and the problems of data analysis in the retail sector.

Table 1: Unified Data Modeling Techniques for GCP-AWS Retail Environments

Technique	Purpose in Retail Context	Problem Addressed
Canonical Schema Design	Standardizes entities like customers, products, and transactions.	Eliminates metric discrepancies across Big Query and Redshift.
Layered Semantic Modeling	Separates technical data structures from business-facing abstractions.	Reduces misinterpretation by non-technical stakeholders.
Parquet/ORC Data Formats	Provides cross-platform compatibility for bulk analytics workloads.	Mitigates file format incompatibilities and query inefficiency.
Federated Query Engines	Enables single-query analytics across GCP and AWS datasets.	Avoids duplication and minimizes latency in decision systems.
Metadata-Driven Governance	Centralizes schema, lineage, and access controls.	Simplifies compliance (GDPR, CCPA) and audit processes.

The next level of interaction is modelling how they can be applied using integration workflows and architectures as outlined in these modelling principles. The latter will provide details on the possible implications for a retail organisation, the data modelling will be normalised via native and third party technologies, and the flexibility in adopting the different types of decision systems [22][23][24] will be emphasized.

IV. IMPLEMENTATION FRAMEWORKS AND CASE INSIGHTS

To apply the rules of unified data modeling to the retail operations, retailers need to establish a schema that can be applied to synchronize ingest, transform, and query between GCP and AWS. Typically, these rely on both platform native services and abstraction services, so while each platform is still capable of operating independently, they all operate harmoniously together as one mapping analysis ecosystem. One such architecture that leverages this involves standardising the ingestion side with data fed into the "staging spaces" on GCP as well as on AWS from Ecommerce, Pops, supply chain and customer engagement systems. The shared transformation logic that data engineers use to process their data is typically standardized (and sometimes even coded into tools such as Apache Beam in Dataflow and Glue ETL jobs) so that all their data always adheres to canonical schema standards. These scripts can be changed with versions, automatically and automatically generated from changes in Data to Derived Data, and be readily adapted to future changes in Data sources and/or regulatory requirements [5][8][15]. Unified analytics is determined by query federation and virtualization add-ons. Business analysts can query both Big Query and Redshift (and third-party) virtualized data without copying the data to both data stores. Such federated strategies minimize latency in important retail decision-making processes, like matching in real-time inventory presence with predicted demand to optimize promotions in real-time [12][17].

Finally, the use of machine learning in decisions systems is adding to when there is an integrated knowledge model. They can deploy using these predictive models and services in a uniform manner, like forecast demand, dynamic pricing, or churn prediction, and they don't have to repeat the pipelines on clouds, such as Vertex AI (GCP) or Sage Maker (AWS). But the consistency would be having a central feature store, without retraining or having schema problems with other teams that use engineered features on other clouds [19][22]. Studies of retail organisations have proven that organisational gains are actually obtained through unified modelling, including cost reduction of managing both ETLs and rapid implementation of new features in analytics and a robust guarantee of the outcomes of decision making. As an example, one multinational retailer who simplified both its GCP and AWS data environments achieved a 40 present savings in their time-to-insight over promotional analytics and 25 present savings on infrastructure costs due to minimal data set and pipeline over-replication costs [20][23].

Below table provides details of the specific strengths and integration points of GCP and AWS services in an integrated modelling approach and how these provide complementary capabilities that can be chained together for integrated retail analytics architecture.

Table 2: GCP vs. AWS Integration Points in Unified Retail Data Modeling

Aspect	GCP Strength	AWS Strength	Integration Strategy
Data Warehousing	Big Query: server less, highly scalable query engine.	Redshift: optimized for clustered, large-scale analytics.	Use federated queries to join across both for hybrid workloads.
Data Transformation	Dataflow (Apache Beam) for streaming and batch ETL.	Glue for automated cataloguing and ETL orchestration.	Standardize transformation logic with shared templates across both.
Machine Learning	Vertex AI for rapid model deployment.	Sage Maker for scalable, production-grade ML training.	Share unified feature stores to maintain consistency across clouds.
Storage	Cloud Storage for raw and curated datasets.	S3 a cost-efficient, durable object storage.	Leverage common formats (Parquet/ORC) for seamless data interchange.
Governance & Metadata	Data Catalo for schema and lineage tracking.	Glue Data Catalo for governance integration.	Federate catalogues or use third-party metadata hubs for a centralized view.

Establishing a framework and patterns of integration, the last part wraps up the hybrid retail information systems' technical fragmentation and how unified modeling can help speed up and standardize decisions at scale. The last section will outline the long-term prospects and ways of upgrading the practices by automating them and making them AI-powered to achieve improvements

V. CONCLUSION

Combining these technologies from the GCP and AWS space unfolds a paradigm shift in the architecture of decision systems across the retail industry, ensuring that organizations have the ability to run advanced analytics with high speed, accuracy and reliability, without sacrificing governance. Technical and governance conformity: Schema conformity, more flexible multi-layered semantic models and federated query engines, and less complex governance structures will enable retailers to overcome the inherent operation silo and inefficiency in hybrid cloud. This will not only enable everyone to be more aligned and leave no room for duplicate efforts and conflicting KPIs, it will also create a single source of analytics, machine learning and a single set of accurate information on-the-fly for everyone to decide. The efficiencies are just the beginning of the benefits, though. By doing this, Unified modeling makes the most of what the time requires, allowing retailers to quickly gain insight and adjust their pricing strategies, offer timely promotions and optimize the stock. Those savings in costs have the added benefit of eliminating pipeline redundancies and storage, along with centralization. And that the centralized control makes keeping up with the ever-changing data privacy regulations (GDPR, CCPA, etc.) continuity a breeze. These are all contributing to increased consumption of decision systems, greater trust in decisions generated by the analytics and increased synergy between organizations and technical parties. Future automation will take place in conjunction with further improvements in the unified modeling strategies, which will also be supported by AI. Consequently, companies can replicate entire data models, governance policies and processing flows in minutes in GCP, or even AWS, and the artificial intelligence-powered optimizers will automatically ensure workload is balance across the clouds based on performance and price. Predictive models for governance will be added to metadata catalogues helping to automatically flag up compliance violations, eliminating from the development cycles the time spent on governance and allowing development cycles to run much faster whilst at the same time flagging up any schema drift violations. For data-driven markets, unified data modeling has become a requirement for competitiveness among the retail organizations trying to remain competitive. Implementing the systems and workflows detailed in this guide will give the BI teams and analytics peace of mind that the way they decide will be consistent, flexible and scalable between the cloud platforms in which they reside. The ability to provide the seamless integration of GCP and AWS data environments will prove critical to help retailers not only stay on the cutting edge of change in the industry, but be the driving force behind change.

VI. REFERENCES

[1] George, J. (2022). Optimizing hybrid and multi-cloud architectures for real-time data streaming and analytics: Strategies for scalability and integration. World Journal of Advanced Engineering Technology and Sciences, 7(1), 10-30574.

[2] Pamisetty, A. (2022). A Comparative Study of AWS, Azure, and GCP for Scalable Big Data Solutions in Wholesale Product Distribution. Azure, and GCP for Scalable Big Data Solutions in Wholesale Product Distribution (December 28, 2022).

[3] Okra, B. P., Aduloju, T. D., Ajani, O. O., Ounce, O., & Amah, L. (2021). A cross-platform data mart synchronization model for high availability in dual-cloud architectures. Journal of Advanced Education and Sciences, 1(1), 70-77.

[4] Globe, L., & Kartashov, A. (2024). OPTIMIZING DISTRIBUTED DATA STORAGE IN MULTI-CLOUD ENVIRONMENTS:

ALGORITHMIC APPROACH. Information and Telecommunication Sciences, (2), 4-12.
[5] Shoo, S. K. (2023). Open-source ETL Framework using Big Data tools Orchestration on AWS Cloud Platform.
[6] Mandala, N. R. (2022). Data Integration in Heterogeneous Systems. ESP Journal of Engineering & Technology Advancements, 2(4).
[7] Zhang, Q., Li, S., Li, Z., Xing, Y., Yang, Z., & Dai, Y. (2015). CHARM: A cost-efficient multi-cloud data hosting scheme with high availability. IEEE Transactions on Cloud computing, 3(3), 372-386.
[8] Pendyala, S. K. Cloud-Driven Data Engineering: Multi-Layered Architecture for Semantic Interoperability in Healthcare.
[9] Shade, F., & Schubert, P. (2020). A semantic data lake for harmonizing data from cross-platform digital workspaces using ontology-based data access.
[10] [10]. Dawson, C. (2023). Cloud Data Governance: Strategies for Managing Big Data in Distributed Environments.
[11] Ferula, S. (2023). The "Delta" Case: New AWS Data Platform Implementation (Doctoral dissertation, Polytechnic di Torino).
[12] Taylor, M., & Scott, A. (2021). Federated query processing in big data integration approaches and prospects. European Journal of Business and Innovation Research, 9(2), 79-87.
[13] Gleam, L. C., Kari, M. R., Zimmermann, L., Kohlbacher, O., Stinkhorn, H., Decker, S., & Bayan, O. (2020). Enabling ad-hoc reuse of private data repositories through schema extraction. Journal of biomedical semantics, 11(1), 6.
[14] Angamuthu, M. (2025). Optimizing Multi-Cloud Business Intelligence: A Framework for Balancing Cost, Performance, and Security. Journal of Computer Science and Technology Studies, 7(4), 427-437.
[15] Brora, P. (2024). Comparative Review: Top Cloud Service Providers ETL Tools-AWS vs. Azure vs. GCP. International Journal of Computer Engineering and Technology (IJCET) Volume, 15, 203-208.
[16] Ruíz-Ceniceros, J. A., Aguilar-Calderon, J. A., Tripp-Barb, C., & Zaldívar-Colado, A. (2023). Dynamic Canonical Data Model: An Architecture Proposal for the External and Data Loose Coupling for the Integration of Software Units. Applied Sciences, 13(19), 11040.
[17] Qi, Z., Ding, J., Javari, R. H., Denture, Y., Ming, X., & Tiara, P. (2023). Fed Sarah: A novel low-latency federated learning algorithm for consumer-centric personalized recommendation systems. IEEE Transactions on Consumer Electronics, 70(1), 2675-2686.
[18] Gunawardena, R. S. (2022). Dynamic Access Control Techniques and Their Role in Preserving Data Confidentiality in Multi-Cloud Retail Solutions. Journal of Computational Intelligence for Hybrid Cloud and Edge Computing Networks, 6(12), 12-22.
[19] Kodakandla, P. Cross-Cloud Data Engineering: Unified Big Data Workflows with AWS and GCP.
[20] Aquarelle, J. I., Asoka, A., Ojukwu, P. U., & Olajuwon, O. J. (2024). Data management solutions for real-time analytics in retail cloud environments. Engineering Science & Technology Journal, 5(11), 3180-3192.
[21] Kumar, R., Li, A., & Wang, W. (2018). Learning and optimizing through dynamic pricing. Journal of Revenue and Pricing Management, 17(2), 63-77.
[22] Kiss, E. (2025). Predictive Analytics for Cloud Resource Planning and Cost Forecasting.
[23] Roman, D., Nikola, N., Soil, A., Elvesæter, B., Song, H., Prod an, R., ... & Ceccarelli, R. (2021, September). Big data pipelines on the computing continuum: Ecosystem and use cases overview. In 2021 IEEE Symposium on Computers and Communications (ISCC) (pp. 1-4). IEEE.
[24] Adeyinka, A. (2023). Automated compliance management in hybrid cloud architectures: A policy-as-code approach.
[25] Wallace, F., MacKay, L., Anderson, C., & Martin, S. (2025). AI for Cross-Cloud Interoperability and Configuration Management.
[26] Arawak, R. (2024). The Role of Infrastructure as Code in Disaster Recovery and Business Continuity. Journal Homepage: http://www. Imran. us, 14(11).
[27] Peace, P., & Agora, H. (2024). Assessing the Role of Metadata in Data Governance Policies.
[28] Siring, A. (2018). Retail Fraud Detection via Log Analysis and Stream Processing. Computer Fraud & Security Bulletin, 2018, 21-32.
[29] Radhakrishnan, P., & Heath, R. (2021). Next-gen cloud synergy: Transforming finance, health, and retail through unified digital frameworks. International Journal of Business Management and Economic Review, 4(1), 1-10.
[30] Brawley, C. (2023). Understanding the Efficacy of Cloud Native Sensitive Data Protection Capabilities (Doctoral dissertation, Dublin, National College of Ireland).