



Scalable Cloud Data Warehousing Models For Global Logistics Tracking Systems

Vinod Battapothu

Independent Researcher, India

vinod.battapothu.researcher@gmail.com

ABSTRACT: Logistics tracking systems involving complex global distribution networks emit massive amounts of data that are typically stored in on-premises data warehouses followed by an export to a data mart for analytical consumption. Analytical operations are also often Monte Carlo simulations, sensitive to how closely the empirical data distribution matches the real-world distribution. Newer capabilities of cloud-based data warehouses can provide better scalability, elastic execution environments, and routing or geospatial capabilities with significantly reduced operational overhead.

Three data warehousing aspects common in traditional designs pursued by cloud-based data analytics of captured logistical shipment data over global tracking routes are analysed with respect to a data warehouse model supporting both real-time or near-real-time workloads: (a) multi-tenancy of the data warehouse schema, (b) data modeling features supporting these analytics, especially the temporal dimension, and (c) additional techniques supporting a global-scale query workload of such a model, especially the ability to elastically scale compute resources for query execution. Results suggest that given the massive scale of global tracking workloads, a design employing a multi-tenant schema configuration is reasonable if not essential to maintain a reasonable cost for operating the data warehouse.

KEYWORDS: Logistics Data Analytics, Cloud Data Warehousing, Global Shipment Tracking, Real-Time Analytics, Multi-Tenant Data Warehouses, Temporal Data Modeling, Monte Carlo Simulation, Geospatial Analytics, Scalable Query Processing, Elastic Cloud Infrastructure.

I. INTRODUCTION

Recent research points to an ever-increasing interest in the logistics and supply chain domain, where the enhancement of the governance, risk and compliance aspects of commercial planning and execution is now strategically relevant. The successful global coordination of increasingly complex logistics networks introduces a range of risk and quality analysis issues that support decision making through the alignment of traditional analytical tools with real time operational data and their visualization through more appropriate information dashboards. In this context, the formation of distributed yet interrelated data it is crucial for the effectiveness of predictive analysis methods.

Cloud public data services allow system owners to establish global distributed control dashboards without high implementation costs and reduced governance constraints on the opening of sensitive company data. The fulfillment of these requirements through the construction of separate cloud data warehouses for each company within the logistics network requires the architectural development of a global data warehouse capable of supporting the information needs of all network companies. For cloud data warehouses, the analysis of the architectural components and solution paradigms must be complemented by the big data characteristics of the geographical distribution of information, in the dimension of scalability, in the relation of data locality and processing, in the real-time nature of data traceability, quality checks and controls on compliance with international rules.

1.1. Overview of the Study's Purpose and Objectives

We present an objective, scholarly, evidence-based synthesis of scalable cloud data warehousing models for global logistics tracking systems. By the nature of their operations and geography, logistics networks generate vast amounts of data on deliveries. Emerging opportunities for cloud data analytics now allow the construction of a global data warehouse to support tracking, analytics, and reporting for the entire network. Such a business model is favorable for cloud data warehouses since the workload is predictable and can

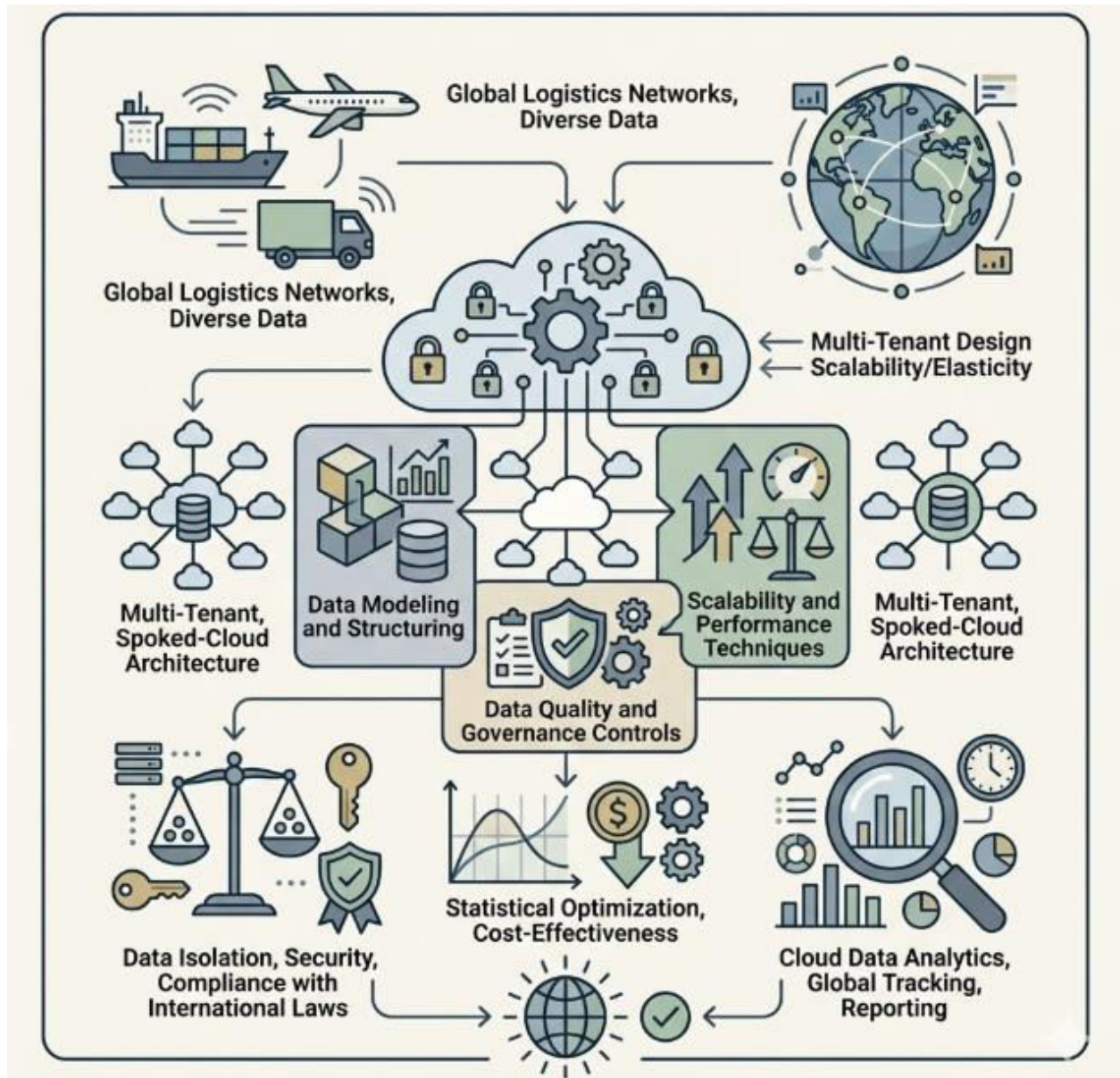


Fig 1: A Multi-Tenant, Spoked-Cloud Architecture for Global Logistics Tracking: Balancing Scalability, Elasticity, and Governance

However, a global tracking warehouse that serves multiple customers also introduces complex challenges. Scalability and elasticity can be supported natively by the cloud infrastructure, yet raising clouds is often accompanied by a low signal-to-noise ratio. A multi-tenant design in a spoked-cloud architecture can bring down operational overhead significantly, yet it raises issues around data isolation, security, and compliance with the laws of international logistics. Moreover, globally distributed and dynamic workloads must also be statistically optimized to utilize the system resources in a cost-effective manner.

The development of a global tracking warehouse is broken into three main components: data modeling and structuring, scalability and performance techniques, and data quality and governance controls. These components are examined through the questions listed above, which evaluate the paradigms and techniques for scalable cloud data warehousing in the context of a logistics tracking system.



II. BACKGROUND AND CONTEXT

The data warehousing landscape for logistics tracking systems is currently experiencing considerable activity, driven in part by the requirements of the global COVID-19 pandemic response and the recovery process. Consumer demand for real-time information about goods in transit, especially orders of increasing complexity, is reflected in a wide range of new academic and commercial contributions. Relatively recent literature on data warehouses for logistics is extensive and varied. The specific interest here is in scalable, cost-efficient constructions for global shipping tracking systems. As with other workloads that are suitable for cloud-based architecting, solutions should exploit the inherent multi-tenancy of the public cloud. Their distributed nature makes such tracking systems prime candidates for follow-the-sun data processing, but this demands special techniques to maintain low latency in a global operating environment. Adequate data quality is also non-negotiable. Regardless of the predictive value of deterministic modeling, end-user expectations demand that records deviate as little as possible from the realities of the present moment.

The capabilities and functions of public cloud vendors operating in the data analytics space have expanded rapidly. Proprietary cloud data warehouses developed by major commercial players are gaining traction in large enterprises and more cautious smaller organizations. Requests posed to these services by different tenants can be processed in either the same or separate sets of control logic. While the former approach offers smaller per-query processing overheads and faster response times, it complicates the spatial data distribution from a data locality perspective. Multiple tenants sharing a schema in a cloud data warehouse introduce a new range of data protection and integrity considerations.

These must encompass database system-level isolation, additional security controls, and potentially different retention periods for different clients, such as airlines and passenger accounting authorities. Decision support functionality for logistics tracking is supported in part by an appropriate data modeling approach. The design of the primary facts and dimensions should reflect the lifetime of shipments, maintain suitable links with external temporal and spatial data, and support both streaming and batch updating of the warehouse.

Metric	Value
Temporal freshness	0.86
Completeness ratio	0.92
Data match ratio	0.9
Anomaly detection ratio	0.07
Attribute-on-time ratio	0.88

2.1. Key Considerations in Data Warehouse Design for Logistics

Current cloud innovations facilitate the realization of scalable data management architectures for global distribution systems that enable comprehensive tracking of shipments across their entire journeys. Cloud data warehousing reduces the total cost of ownership compared to on-premises solutions, provided that organizations are willing to adopt a multi-tenant approach. While such models deliver significant operational advantages, they may nevertheless entail some disbenefits by exposing organizations to performance issues arising from the resource demand of other tenants and potentially challenging compliance with regulatory requirements, such as data jurisdiction rules. The specific dataset and usage characteristics of international logistics tracking systems suggest a dedicated schema model as preferable. Such an approach incurs a higher operational cost but can offer improved isolation, higher likely security, and easier management of compliance implications.

As analytics become increasingly cloud-based, the distribution of data in space, time, and frequency of use will necessitate the integration of data at a global level, with reconciling temporal and spatial dimensions now moving to the forefront with respect to data modeling. The near real-time ingestion of global tracking data through a cloud data warehouse further amplifies the challenge. Decision makers need to understand the required data freshness and analytics speed to determine whether a near real-time stream-processing or a batch-processing paradigm is appropriate for a given processing requirement. Nevertheless, for many applications, the benefits of advanced analytics using actual latest-node dwell times will outweigh the lower costs of historical-batch processing.

Equation A. Temporal freshness (timeliness) metric

Step 1: Define event timestamps

- Expected update time: t_{exp}
- Actual update time: t_{act}



Step 2: Define delay

$$\Delta t = t_{\text{act}} - t_{\text{exp}}$$

If updates arrive early, you can clamp:

$$\Delta t^+ = \max(0, \Delta t)$$

Step 3: Convert delay into a 0–1 score

Choose an SLA threshold T (e.g., “acceptable delay”).

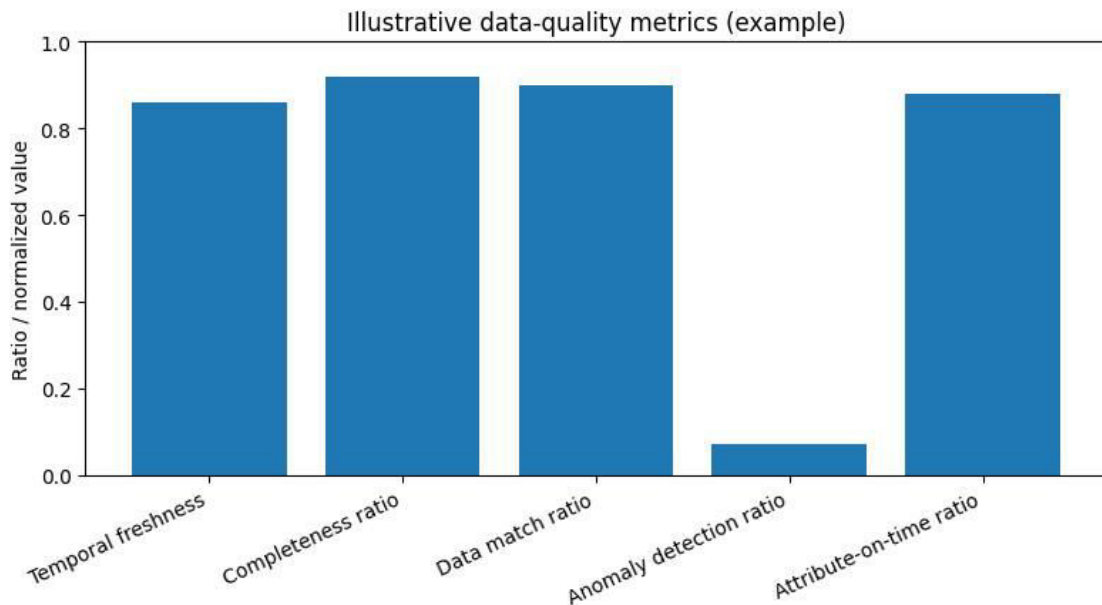
$$F = 1 - \frac{\Delta t^+}{T}$$

Step 4: Clamp to valid range

$$F_{\text{clamped}} = \min(1, \max(0, F))$$

So the **temporal freshness score** is:

$$F_{\text{clamped}} = \min \left(1, \max \left(0, 1 - \frac{\max(0, t_{\text{act}} - t_{\text{exp}})}{T} \right) \right)$$



III. DATA WAREHOUSE ARCHITECTURES FOR LOGISTICS

Many transaction processing and data analytics workloads have effectively utilized off-the-shelf cloud services in a pay-per-use manner. However, global logistics tracking systems, often involving geographically distributed heterogeneous data sources and heavy resource consumption, have not yet converged. Distributed and partitioned data processing and storage capabilities have evolved to support such workloads, breaking the earlier one-size-fits-all approach of on-demand offerings. Traditional cloud data warehouses have achieved the objectives of scalability, elasticity, and operational simplicity, superseding commercial or in-house constructed systems.

Within the context of global logistics tracking requirements, the essential differentiation between traditional on-premises data warehouse architectures and cloud-based alternatives is on operational overhead. Traditional on-premises architectures require a continuous ITOps effort, accounting for a substantial fraction of the infrastructure TCO, while cloud DWaaS offerings inherently require minimal operational effort and associated cost. The main remaining aspect to decide is whether a multi-tenant or dedicated-schema DWaaS offering is more suitable for global tracking workloads. Global cross-organizational data sharing generally has an impact on data sense-making, isolation, and security aspects, as well as on regulatory compliance, especially with respect to personally identifiable information.



3.1. Traditional vs. Modern Cloud Data Warehouses

The comparison begins with traditional data warehouses, designed for a fixed load that is expected to change only slowly as the business grows. As long as the load remains within deployment limits, maintenance is relatively straightforward and predictable.

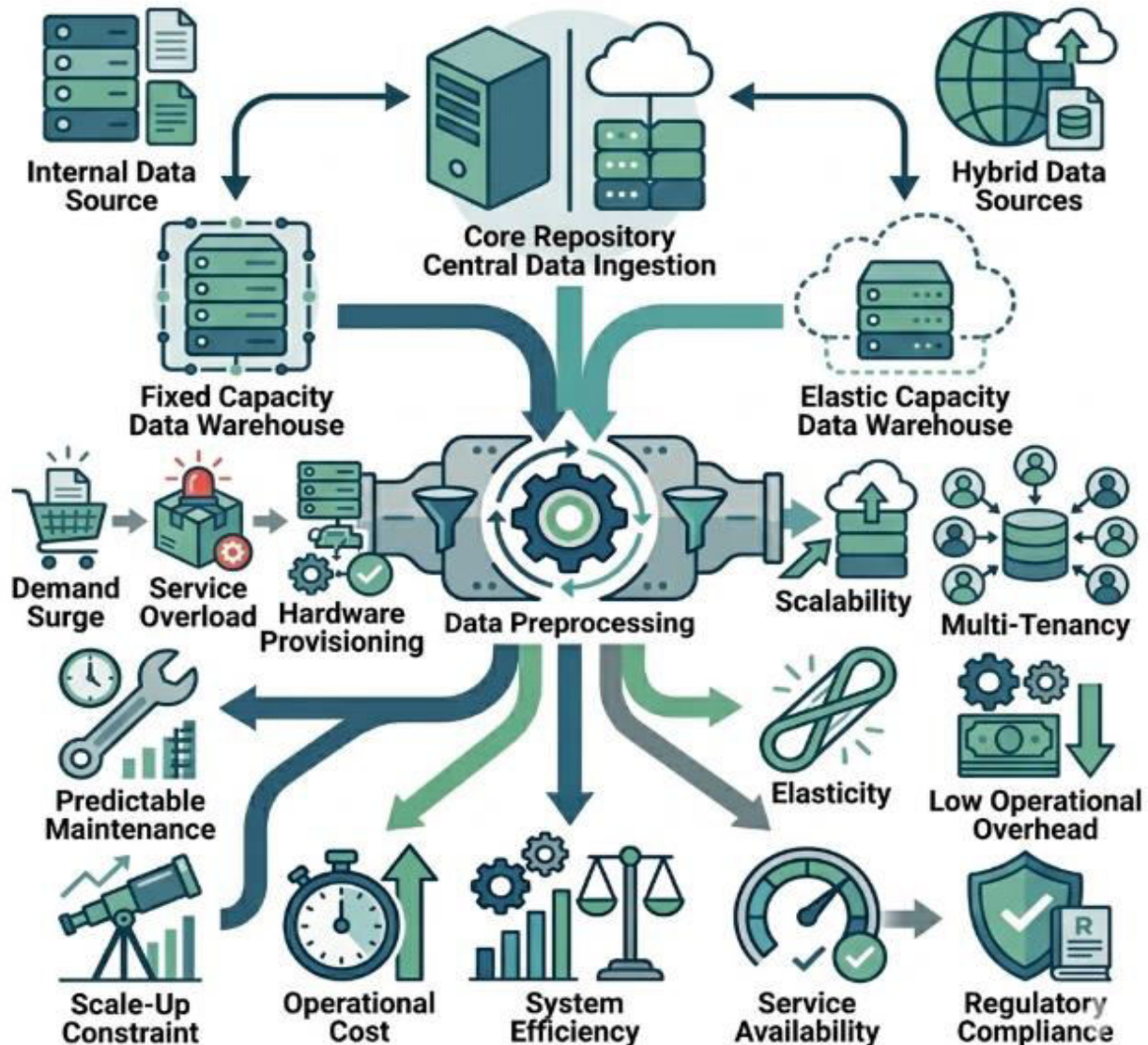


Fig 2: Comparative Analysis of Fixed vs. Elastic Data Warehouse Architectures: Operational Implications of Cloud-Native Multi-Tenancy on Public Sector Scaling

However, when demand peaks or exceeds capacity, either the online services become delayed or unusable, or impending demand must be forecast, hardware provisioning arranged, and added capacity tested before becoming available. These issues are not new, and many development teams address them during a project. What is different now is the existence of cloud data warehouses, which—given access to external systems—can offer at least a workable solution.

Unlike a traditional data warehouse, a cloud data warehouse can scale up to meet demand. The obvious solution is to create multiple copies, each serving a subset of users, or one copy with distinct roles. This approach is certainly possible and appears—at least anecdotally—common, but is it really a cloud data warehouse? Is it not rather an on-premises data warehouse that happens to be running on an external service provider? The question is moot. Much consulting documentation continues to refer to on-premises solutions as “cloud ready,” as long as external services can



be incorporated. Nevertheless, the discussion is helpful, as it identifies multi-tenancy as an essential cloud feature that enables other features such as elasticity and low operational overhead.

3.2. Multi-tenant Versus Dedicated Schemas

The fundamental schemes used in most cloud DWaaS implementations are multi-tenant in nature, where multiple customers share the same physical resources and the same underlying database engines. For cloud data warehousing services supporting globally distributed logistics processes, a dedicated schema per principle can be considered, even if such a design does not conform to the standard. Each principal and its data can be better isolated to meet data sovereignty, security, and privacy protection requirements. Also, regulatory or organizational policies may require the isolation of sensitive data from other customers even when making use of a cloud service supporting a multi-tenant architecture.

Logistics principles supporting large-scale cloud data warehouse solutions for global tracking purposes may benefit from a dedicated schema per principle. However, such a design imposes additional operational overhead in terms of provisioning, scaling, monitoring, and maintaining the dedicated underlying database engines, and it may obscure the cloud service from being perceived as a DWaaS. Fig. 8 illustrates a workload such a distribution orchestrates, showing the processing of shipments moving from Singapore to New Zealand across different logistics principles in the network. Three shipping events may be completed within hours of each other, but the warehouse design enables the event with larger space requirements to be processed first, such that the warehouse is not running at full utilization yet and the two smaller events can be processed next at lower data freshness latencies.

Equation B. Completeness ratio

Step 1: Define required fields count

- Total required attributes: A_{req}

Step 2: Count present attributes

- Present (non-null, valid) attributes: A_{present}

Step 3: Ratio

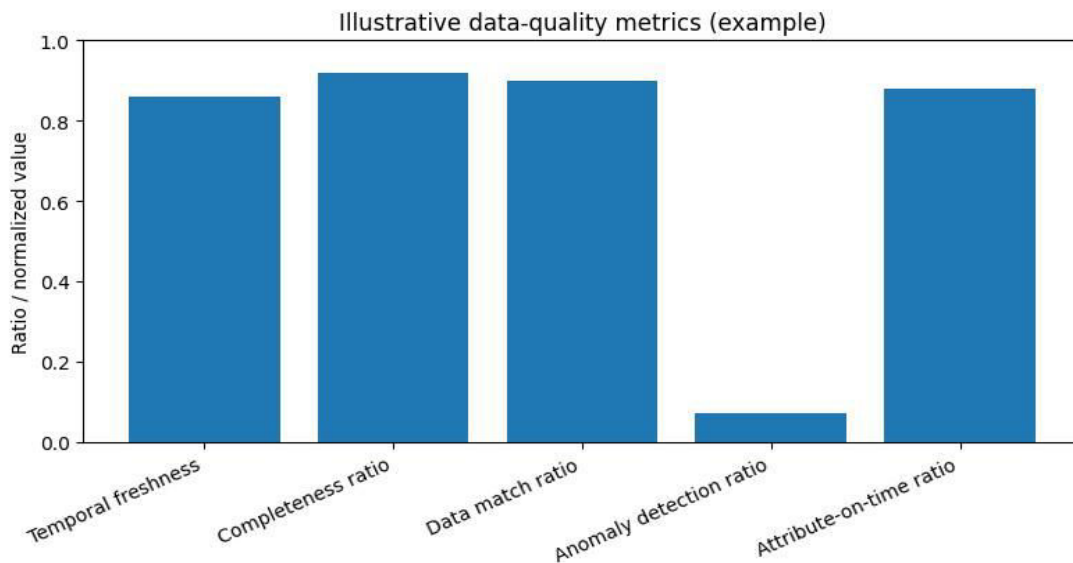
$$C = \frac{A_{\text{present}}}{A_{\text{req}}}$$

Clamp to [0,1] if needed.

IV. DATA MODELING FOR GLOBAL TRACKING

The data modeled in a cloud warehouse geared towards global logistics tracking is an event sequence associated with a parcel's shipment. Each shipment typically obeys the same flow from origin to destination, albeit with a number of routing options and an associated flow sequence that is only occasionally subject to change. The relevant movement is that of the shipment itself, but events process within the parcel's flow—pickups, transfers, and deliveries—are analyzed in conjunction with the shipment.

To allow easy analysis with other applications, a primary business key with a defined scope must be established within the staging area. Both the temporal and spatial dimensions of an event should be described. The temporal dimension tracks the time of the event and ideally stores a date and timestamp in UTC format. In practice, though, the time zone(s) of carriers and the timestamps applied are frequently inconsistent. The spatial dimension can consist of a combination of attributes covering the identity of the main carrier and the place of occurrence as captured in the original travel chain. Routing analysis is typically directed at dwell times, so a conformed departure dimension, spanning worldwide locations along with their GPS coordinates, is required. From this, a carefully defined set of slowly changing dimensions can be built: departure-stop dimensions that represent the current flow from origin to destination, the flow seat of the carrier responsible for the shipment at the moment of processing, and the current flow router.



Equation C. Data match ratio

Step 1: Define matching predicate

Let each record i have:

- observed event type e_i
- expected event type from partner mapping $\hat{e}_i$

Define indicator:

$$m_i = \begin{cases} 1, & e_i = \hat{e}_i \\ 0, & \text{otherwise} \end{cases}$$

Step 2: Average over N events

$$M = \frac{1}{N} \sum_{i=1}^N m_i$$

4.1. Temporal and Spatial Dimensions

Data movement through a distributed logistics network can be viewed as a sequence of shipment events appointed to carriers over time. For each event, active logistics partners, the current load, systemic routing, and other relevant business data are represented by non-temporal dimensions. Besides a timestamp, tracking events can be accompanied by additional information about monitored installations, checked inspections, applied transports, and passed transshipments. In this extended fact constellation, lanes for the complete shipment remain available, and a dedicated logical grouping of the main partner dimensions supports a descriptive routing analysis of shipments. Other dimensions describe the duration of transports and the dwell time of loads at hubs and terminals. Arrival delays can be investigated relative to planned transports and dwell times.

Logistics organizations adopt cloud data warehousing more in the direction of analytics as a service, scaling up or down according to the business cycle. Batch-oriented architectures remain acceptable for global workload characteristics. Incoming events tend to exhibit sudden spikes rather than smooth streaming patterns. Top latency requirements concern the freshness of tracking information rather than the speed of updates. Global tracking systems thus commonly prefer a near real-time availability of fresh data. Such batch-mode storage and processing also align well with distribution and load-balancing issues of geographically distributed cloud resources.

Tenants	Multi-tenant cost/tenant (example)	Dedicated cost/tenant (example)
7	305.7142857142857	490.0
8	282.5	490.0
9	264.44444444444446	490.0
10	250.0	490.0
11	238.1818181818182	490.0



Tenants	Multi-tenant cost/tenant (example)	Dedicated cost/tenant (example)
12	228.33333333333334	490.0

4.2. Real-time vs. Batch Ingestion Paradigms

Concerning the ingestion paradigm supporting the data warehouse for a real-time global tracking workload, the need for sufficient freshness and the resulting latency target favor a real-time approach. However, a real-time ingestion mode does not introduce an additional dimension for query processing; ingestion can follow streamed micro-batch processing and batch processing paradigms, thus either conforming to or supporting the temporal scale of the analytical processing. A real-time ingestion mode ideally supports supply chain events as they occur at source nodes. It can, however, also focus on a subset of events where real-time query executions are expected to help decision-making, such as those likely affecting the customer experience.

Real-time ingestion facilitates limited timeliness of change-data capture for the route and dwelling-time dimensions, but it may also incorporate delta updates for all temporally partitioned tables. Sensible timeliness for events at volume nodes justifies processing-fresh data, but the ally-dimensioned volume distribution along the settled supply chain still enables traditional batch-analytical or hybrid-analytical modes, with processing-lengthy freshness constraints. Combined real-time alongside batch update supports sufficiently refreshed data handling and implies the real-time Layer-1 throughput, also qualified to diminish batch latency at key sections.

V. SCALABILITY AND PERFORMANCE TECHNIQUES

When catering to a global workload, distributed storage and compute mechanisms improve scalability and performance. In cloud-based data warehouses such as Amazon Redshift, Google BigQuery, and Snowflake, the user can choose how to partition the data in the underlying storage. Data can be split across physical storage nodes by specifying a partitioning column; however, in some cases, it may not be obvious how to generate a splitting column using the standard partitioning operation. Among the data found in shipment events, a natural choice of partitioning is event time. Data can also be distributed across multiple clusters for improving scalability and elasticity, which has been made possible by routing data to different clusters in a multi-cluster setup in Snowflake or via Federated Queries in BigQuery utilizing BigQuery Data Transfer Services and scheduled queries. During off-peak hours, the amount of data being routed is usually low, and therefore, additional clusters can be granted to meet such high workloads. The auto-scaling capability can be further tuned by utilizing utilization metrics such as concurrency utilization, CPU utilization, storage utilization, and queues to optimize scaling decisions.

As a result of the highly variable and geographically distributed nature of global logistics workloads, it is important to eliminate cold-start latency as much as possible. The most popular queries of a data warehouse can be cached and resources can be provisioned to colocate data and processing as close as possible to the region from where a workload originates. After analyzing a historical set of workload execution traces, the information can be formulated in a mapping of regions to data questions. During peak demand for certain types of questions in certain regions, the processing capabilities can be proactively stepped to meet the forecasted demands. Such workload-aware resource allocation is also similar to the previously discussed concept of geo-patterns.

Equation D. Anomaly detection ratio

Step 1: Define anomaly flag

$$a_i = \begin{cases} 1, & \text{event } i \text{ is flagged anomalous} \\ 0, & \text{otherwise} \end{cases}$$

Step 2: Ratio

$$A = \frac{1}{N} \sum_{i=1}^N a_i$$

Note: **Lower is not always “better”**—a very low A can mean either “clean operations” or “weak detection.”

5.1. Distributed Storage and Compute

Distributed storage and compute are critical for globally distributed cloud data warehouses. Storage servers are usually organized in cloud storage services as large object stores but can also take the form of an object store with key value tables as in Google BigTable or columnar NoSQL stores. They employ redundant data replication across multiple regions to ensure high availability. Computing resources involve the services provided in Infrastructure as a Service

cloud environments: virtual machines from colocated pools that can be dynamically allocated or de-allocated on demand via automation tools. Latency affecting workloads are thus always served by resources deployed or shipped closer to the user or the data center.

Multi-tenancy or shared schema models usually require high availability and low latency on the storage level. Consequently, queries executing on several data shards tend to see their costs grow exponentially with the number of shards. Optimization techniques trading performance for cost include caching results from commonly executed queries in a scalable low-latency storage (in-Memory Databases, NoSQL key-value stores) that store small portions of the global dataset to avoid shuttling data all over the system and enable real-time response for popular requests. In environments where query execution volume is high enough for the problem of efficiently distributing the data traffic load among the database response resources to be noticeable, a workload-aware resource auto-scaling strategy can be employed to ensure continuous optimal resource allocation.

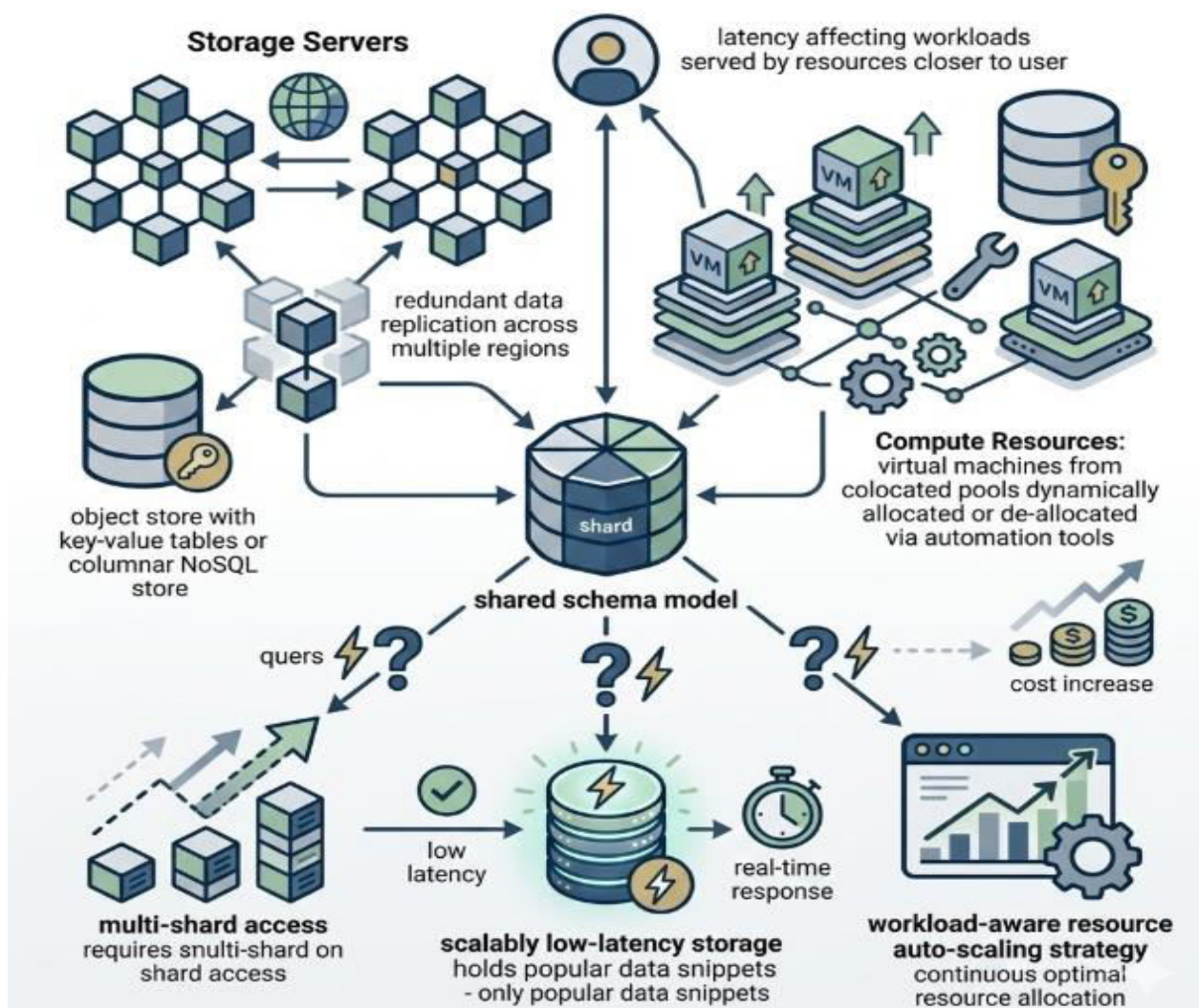


Fig 3: Dynamic Scalability in Global Cloud Data Warehousing: Optimizing Multi-Tenant Latency through Workload-Aware Auto-Scaling and Distributed Caching

5.2. Query Optimization for Geographically Distributed Workloads

Query optimization for geographically distributed workloads can address the impacts of data locality, caching, and workload-aware resource management on response time. For analytical processing, reducing data-transit time between warehouse and processing nodes is a primary optimization strategy. For situations in which global data availability is important, a query latency of minutes is acceptable, allowing either data-caching solutions or hot-region processing at the edge of the logistics network.



An obvious solution for reducing response time for geographically distributed workloads is to have the compute nodes located wherever the data reside. This usually requires a full data copy, which incurs high costs. Because this data redundancy often changes, a more cost-efficient solution is to keep only a hot copy of the frequently used data and the cold copy that incurs low maintenance cost. Such a partitioning strategy can reduce response time and data-transit volume for both cache solutions and data-transit operations. Another optimization strategy is workload-aware resource management, which steers or replicates a specific data copy closer to the event based on the workload distribution pattern. Although such a control strategy may not yield the best performance in isolation, it is suitable for query processing with user-perceived performance levels.

VI. DATA QUALITY, GOVERNANCE, AND COMPLIANCE

Data quality is a key aspect of data warehousing that requires attention throughout the entire data lifecycle. In the context of the proposed models to support global logistics tracking systems, the definition of quality criteria and the governance procedures that help to maintain the data quality at an acceptable level are essential. For example, data processing in accordance with GDPR requires including consent management capabilities in the systems using the data warehouse, and such capabilities can be supported by having the consent information stored, monitored, and analyzed in the data warehouse. Although these two aspects relate to data quality in different ways, there is an intersection for large distributed systems—failure detection—and it is critical to be able to track the cause of failures, and furthermore, that information is leveraged to prevent future recurrences.

The set of data quality criteria used to define quality procedures starts with defining and regularly assessing data quality metrics appropriate for the usage of the data, such as accuracy, completeness, consistency, conformity, currency, timeliness, detectability, and plausibility. Once the metrics are defined, appropriate procedures for the logistics domain can include checks on the detectability to ascertain that on-time events for shipments are indeed the latest events, alerts for plausible anomalies or deviations detected using historical data, and checks on the consistency of the routing data used in processing the events, for example, inter-country checks on France routing data, where no routing using Italy and Spain should be proposed on any valid shipment. Engaging the appropriate business users in the quality review process ensures that the data used for key analysis are maintained at a suitable quality level, helping to prioritize investments in quality controls when required.

Equation E. Attribute-on-time ratio

Step 1: Define on-time predicate for attribute j of event i

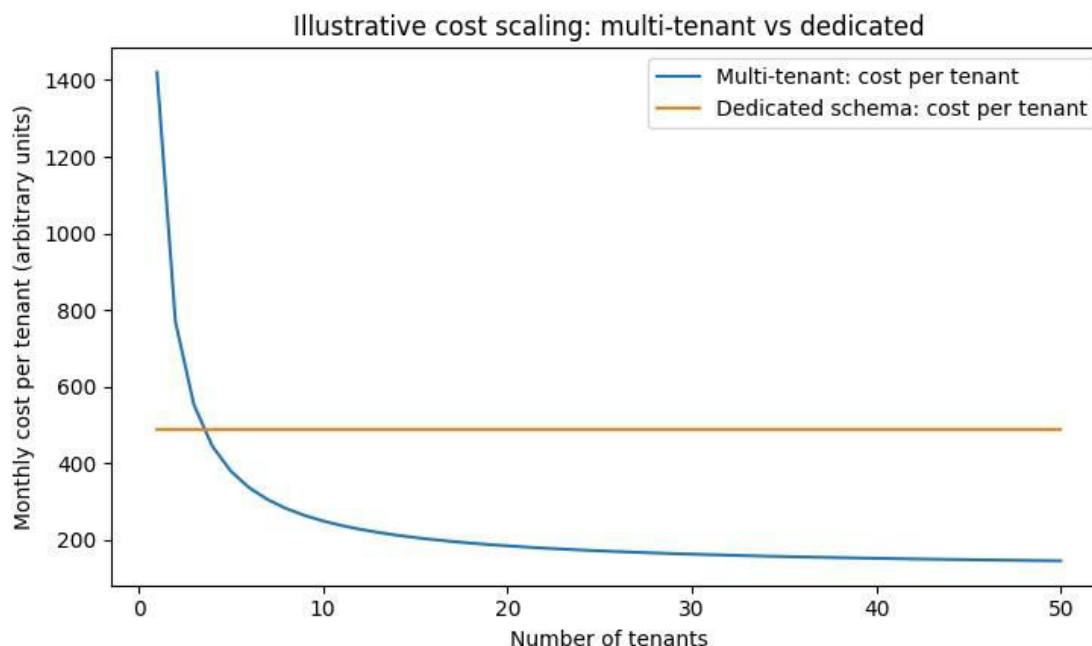
- attribute observed time: $t_{i,j}$
- deadline: $d_{i,j}$

$$o_{i,j} = \begin{cases} 1, & t_{i,j} \leq d_{i,j} \\ 0, & \text{otherwise} \end{cases}$$

Step 2: Ratio over all event-attribute pairs

If there are N events and K tracked attributes per event:

$$O = \frac{1}{NK} \sum_{i=1}^N \sum_{j=1}^K o_{i,j}$$



6.1. Metadata Management and Lineage

Today's global logistics networks continuously generate vast amounts of tracking data. A high number of stakeholders at different levels need to analyse this information at various granularities to understand when shipments will arrive and how these times can be improved. Traceability in this context is essential, because the transport of goods is often governed by national and international regulations. Special requirements for visibility and reporting tools add another layer of complexity. Logistics service providers as well as customs, tax, and other governmental authorities require accurate and trustworthy information. Metadata management, lineage tracking, and data quality are widely recognised elements of commercial systems.

Maintaining accurate metadata is fundamental for analytics. Metadata repositories are typically supported by dedicated database systems and include descriptions of the structure and content of data in data warehouses, logs of user actions, search indexes, the lineage of each analysis stream, conformance check and validation rules, and semantically enriched data. Metadata also provides underlying mechanisms to support the definition, implementation, and evolution of analytical tools and ensure compliance with regulations. Impact analysis is made possible by capturing the complete data lineage of all delivered information products. Nevertheless, logistics data quality is an especially challenging aspect. The processes that create and modify logistics data involve an extended number of stakeholders, some of which are not motivated by the final audience nor subject to scrutiny by internal quality controls. Data transformation often entails using information created by partners, external authors, or automatic systems; thus, concealment or premeditation of errors is highly probable.

Clusters	Utilization rho (example)	Mean latency ms (example)
7	1.1428571428571428	148849.999999999988
8	1.0	148849.999999999988
9	0.8888888888888888	12349.999999999995
10	0.8	6350.000000000001
11	0.7272727272727273	4350.0
12	0.6666666666666666	3349.999999999995

6.2. Data Quality Metrics and Validation

Data quality is vital for reliable analytical results, enabling accurate reporting, decision-making, and visualization. Adequate data quality measures promote informed confidence in data and insights derived from it, fostering rapid and appropriate business actions. The absence of data quality mechanisms increases the likelihood of poor-quality data entering the cloud environment and supports the need for automated detection and validation within cloud data warehouses.



Scrutiny of the SCF process illustrated its reliance on multiple global logistics partners for shipment data, raising data quality concerns. Though LSPs aim to fulfil SCF's Official unparalleled responsibility, system inefficiencies have affected global business. The data warehouse should address these issues, and the historical DDL should regulate incoming event data. The following key metrics help assess replacement and modification data quality:

- Temporal freshness: Delay in actual replacement/modification compared with expected time, crucial for timely action/decision. Fresh data can still be valid for reporting even after delivery.
- Completeness ratio: All replacement details must be present in the actual data; missing fulfillment details indicate reasons for Visual Inspection Assign Operation.
- Data match ratio: Actual event/event_type should correspond with the source Logistics Partner. Although unexpected entries may occur, high percentages are expected.
- Anomaly detection ratio: Flagging events using std., mean, forecasting models, or clustering techniques ensures monitoring and management of abnormal and unusual operations.
- Attribute-on-time ratio: Capturing of all possible attributes within expected time (when available for each operation type).
- Overall data quality score: Index value derived from the above factors helps indicate when a data quality control mechanism should trigger corrective processes.

VII. CONCLUSION

The objective of the study was to assess and characterize cloud data warehousing architectures and solutions for a global data handling and analysis system in logistics and distribution. An enterprise-oriented analytical architecture was examined as a case study, and the facilities and capabilities offered by modern cloud platforms were discussed. Without advocating a specific technology or product and while bearing Parikh's work in mind, distinct data financing and warehousing architectures were rationalized for different use cases and loads. Potential design dimensions and techniques for a global cloud storage and data warehouse system were proposed.

As large, complex, and widely distributed solutions, global logistics tracking systems pose several unique challenges. These include the volume and variety of data generated and acquired within a scaling network; the need for reliable analysis and prediction to enable corrective action in real time, anywhere in the world; and the imperative to be prepared for regional storage and data handling legislation without burdening internal structures. Cloud support greatly enhances the possibilities for data warehousing and handling. However, the services that cloud environments offer can differ dramatically, depending on the region, country, and even the cloud provider.

Enterprise Architectural Weighting

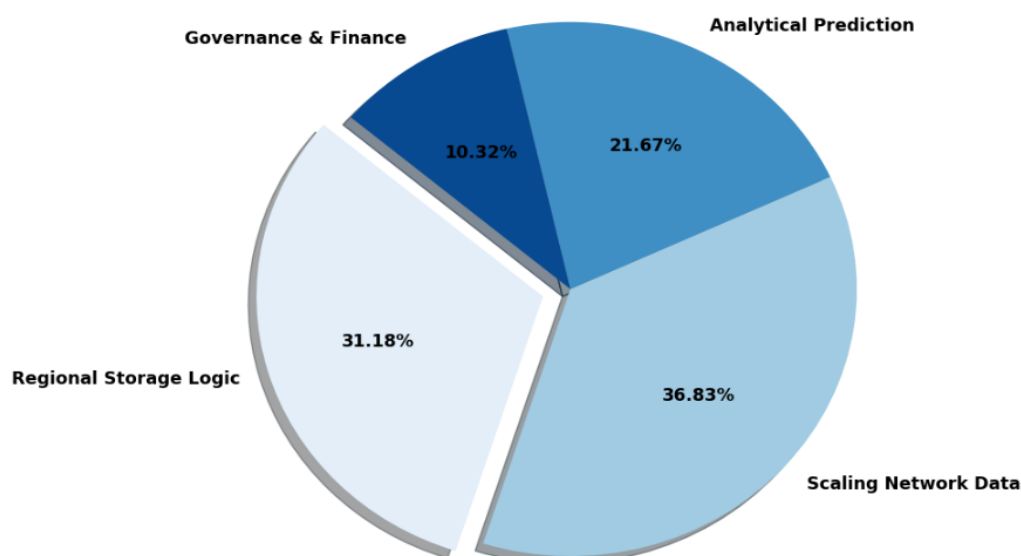


Fig 4: Enterprise Architectural Weighting



7.1. Final Thoughts and Future Directions

On-premise data warehouses are being adopted with an ever-decreasing operational overhead, and cloud-native data solutions have a promising future in global logistics. Cache-optimizing and query locality-aware data distribution approaches are particularly relevant for geographically distributed environments, where supporting a dedicated cloud tenant merely for LHS, with Data Warehouse as-a-service functionality geared toward system integrators and customized services for end-users, DQ-DQ monitoring, prevention, detection, data cleaning, and quality control measures in close alignment with worldwide trade requirements addressees; however, these need to be systematically defined and implemented until reliable and established data quality becomes a competitive differentiator. Countermeasures for ensuring that valid data is shared with AI and ML applications running in other cloud tenants will also determine the long-term adoption and overall success of such systems.

Inclusion of additional sources (e.g., ship owners/operators, owners of container/intermodal cargoes) can also advance forecasts for remaining times (or other KPIs) of the (sub-)legs to improve door-to-door delivery distances and/or windows for last-leg deliveries. Owners/operators of intermodal transport carriers (combined inland shipping and road transport) should likewise be integrated. Moreover, as the previous discussion revealed, fast-shifting level data (dimension tables, mostly) have a tendency toward real-time update forms, warranting adoption of real-time ingestion. Compressing these mass data streams into streaming small data packages (incl. mini-batches) is wanted, with streaming processing being a suitable processing mode for the streams. Processed, outputted small data packages at the same-to-similar granulator level as the input data streams can actually support near-real-time updating of a cloud data engine and joining a near-real-time data warehouse analytics solution.

REFERENCES

- [1] Sharma, A., & Rani, R. (2022). Machine learning approaches for healthcare fraud detection. *Journal of Healthcare Engineering*, 2022, 1–13.
- [2] Gadi, A. L. The Role of Digital Twins in Automotive R&D for Rapid Prototyping and System Integration.
- [3] Zhang, Y., Zhao, L., & Chen, X. (2022). Intelligent analytics for healthcare financial fraud detection. *IEEE Access*, 10, 90124–90136.
- [4] Kothapalli Sondinti, L. R., & Syed, S. (2022). The Impact of Instant Credit Card Issuance and Personalized Financial Solutions on Enhancing Customer Experience in the Digital Banking Era. *Universal Journal of Finance and Economics*, 1(1), 1223. Retrieved from <https://www.scipublications.com/journal/index.php/ujfe/article/view/1223>
- [5] Arasu, A., & Kaushik, R. (2014). Data cleansing: A context dependent approach. *Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data*, 135–146.
- [6] Singireddy, J. (2022). Leveraging Artificial Intelligence and Machine Learning for Enhancing Automated Financial Advisory Systems: A Study on AI-Driven Personalized Financial Planning and Credit Monitoring. *Mathematical Statistician and Engineering Applications*, 71 (4), 16711–16728.
- [7] Armbrust, M., Das, T., Davidson, A., Ghodsi, A., Or, A., Rosen, J., Stoica, I., Wendell, P., Xin, R., & Zaharia, M. (2021). Delta Lake: High-performance ACID table storage over cloud object stores. *Proceedings of the VLDB Endowment*, 13(12), 3411–3424.
- [8] Garapati, R. S. (2022). AI-Augmented Virtual Health Assistant: A Web-Based Solution for Personalized Medication Management and Patient Engagement. Available at SSRN 5639650.
- [9] Alsharo, M., Alnsour, Y., & Alabdallah, M. (2020). How habit affects continuous use: Evidence from Jordan's national health information system. *Informatics for Health and Social Care*, 45(1), 43–56. <https://doi.org/10.1080/17538157.2018.1540423>
- [10] Chava, K., Chakilam, C., & Recharla, M. (2021). Machine Learning Models for Early Disease Detection: A Big Data Approach to Personalized Healthcare. *International Journal of Engineering and Computer Science*, 10(12), 25709–25730. <https://doi.org/10.18535/ijecs.v10i12.4678>
- [11] Babcock, J., Chaudhuri, S., & Das, G. (2004). Dynamic sample selection for approximate query processing. *Proceedings of the 2004 ACM SIGMOD International Conference on Management of Data*, 539–550.
- [12] Sriram, H. K. (2022). Advancements in Credit Score Analytics using Deep Learning and Predictive Modeling Techniques. Available at SSRN 5255128.
- [13] Bifet, A., & Gavaldà, R. (2007). Learning from time-changing data with adaptive windowing. *Proceedings of the 2007 SIAM International Conference on Data Mining*, 443–448.
- [14] Muthusamy, S., Kannan, S., Lee, M., Sanjairaj, V., Lu, W. F., Fuh, J. Y., ... & Cao, T. (2021). Cover Image, Volume 118, Number 8, August 2021. *Biotechnology and Bioengineering*, 118(8), i-i.



- [15] Kalisetty, S., & Ganti, V. K. A. T. (2019). Transforming the Retail Landscape: Srinivas's Vision for Integrating Advanced Technologies in Supply Chain Efficiency and Customer Experience. *Online Journal of Materials Science*, 1, 1254.
- [16] Gottimukkala, V. R. R. (2021). Digital Signal Processing Challenges in Financial Messaging Systems: Case Studies in High-Volume SWIFT Flows.
- [17] Chen, M., Mao, S., & Liu, Y. (2014). Big data: A survey. *Mobile Networks and Applications*, 19(2), 171–209.
- [18] Dwaraka Nath Kummari. (2022). Fiscal Policy Simulation Using AI And Big Data: Improving Government Financial Planning. *Kurdish Studies*, 10(2), 934–945. <https://doi.org/10.53555/ks.v10i2.3855>
- [19] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794.
- [20] Davuluri, P. N. Event-Driven Compliance Systems: Modernizing Financial Crime Detection Without Machine Intelligence.
- [21] Das, T., Zhu, A., Li, S., Narayanamurthy, S., & Bhat, P. (2013). Distributed and fault-tolerant streaming computation in Spark. *Proceedings of the ACM Symposium on Cloud Computing*, 1–12.
- [22] Siva Hemanth Kolla. (2022). Knowledge Retrieval Systems for Enterprise Service Environments. *International Journal of Intelligent Systems and Applications in Engineering*, 10(3s), 495–506. Retrieved from <https://ijisae.org/index.php/IJISAE/article/view/8037>
- [23] Dean, J., & Ghemawat, S. (2008). MapReduce: Simplified data processing on large clusters. *Communications of the ACM*, 51(1), 107–113.
- [24] Uday Surendra Yandamuri. (2022). Cloud-Based Data Integration Architectures for Scalable Enterprise Analytics. *International Journal of Intelligent Systems and Applications in Engineering*, 10(3s), 472–483. Retrieved from <https://ijisae.org/index.php/IJISAE/article/view/8005>
- [25] DeCandia, G., Hastorun, D., Jampani, M., Kakulapati, G., Lakshman, A., Pilchin, A., Sivasubramanian, S., Voss, P., & Vogels, W. (2007). Dynamo: Amazon's highly available key-value store. *Proceedings of the 21st ACM Symposium on Operating Systems Principles*, 205–220.
- [26] Sriram, H. K., ADUSUPALLI, B., & Malempati, M. (2021). Revolutionizing Risk Assessment and Financial Ecosystems with Smart Automation, Secure Digital Solutions, and Advanced Analytical Frameworks.
- [27] Davuluri, P. N. (2020). Improving Data Quality and Lineage in Regulated Financial Data Platforms. *Finance and Economics*, 1(1), 1-14.
- [28] Adusupalli, B., Pandiri, L., & Singireddy, S. (2019). DevOps Enablement in Legacy Insurance Infrastructure for Agile Policy and Claims Deployment. *risk*, 7(12).
- [29] Elmagarmid, A. K., Ipeirotis, P. G., & Verykios, V. S. (2007). Duplicate record detection: A survey. *IEEE Transactions on Knowledge and Data Engineering*, 19(1), 1–16.
- [30] Dwaraka Nath Kummari. (2022). Machine Learning Approaches to Real-Time Quality Control in Automotive Assembly Lines. *Mathematical Statistician and Engineering Applications*, 71(4), 16801–16820. Retrieved from <https://philstat.org/index.php/MSEA/article/view/2972>
- [31] Fader, P. S., Hardie, B. G. S., & Lee, K. L. (2005). "Counting your customers" the easy way: An alternative to the Pareto/NBD model. *Marketing Science*, 24(2), 275–284.
- [32] Inala, R. (2022). Engineering Data Products for Investment Analytics: The Role of Product Master Data and Scalable Big Data Solutions. *International Journal of Scientific Research and Modern Technology*, 155-171.
- [33] Davuluri, P. N. (2020). Improving Data Quality and Lineage in Regulated Financial Data Platforms. *Finance and Economics*, 1(1), 1-14.
- [34] Meda, R. Enabling Sustainable Manufacturing Through AI-Optimized Supply Chains.
- [35] Ghemawat, S., Gobioff, H., & Leung, S. T. (2003). The Google file system. *Proceedings of the 19th ACM Symposium on Operating Systems Principles*, 29–43.
- [36] Kannan, S. (2021). Advanced Computational Technologies in Vehicle Production, Digital Connectivity, and Sustainable Transportation: Innovations in Intelligent Systems, Eco-Friendly Manufacturing, and Financial Optimization. *Universal Journal of Finance and Economics*.
- [37] Yandamuri, U. S. (2021). A Comparative Study of Traditional Reporting Systems versus Real-Time Analytics Dashboards in Enterprise Operations. *Universal Journal of Business and Management*, 1(1), 1–13. Retrieved from <https://www.scipublications.com/journal/index.php/ujbm/article/view/1357>
- [38] Gottimukkala, V. R. R. (2022). Licensing Innovation in the Financial Messaging Ecosystem: Business Models and Global Compliance Impact. *International Journal of Scientific Research and Modern Technology*, 1(12), 177-186.
- [39] Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer.
- [40] Vadisetty, R., Polamarasetti, A., Guntupalli, R., Raghunath, V., Jyothi, V. K., & Kudithipudi, K. (2022). AI-Driven Cybersecurity: Enhancing Cloud Security with Machine Learning and AI Agents. Sateesh kumar and



Raghunath, Vedapada and Jyothi, Vinaya Kumar and Kudithipudi, Karthik, AI-Driven Cybersecurity: Enhancing Cloud Security with Machine Learning and AI Agents (February 07, 2022).

[41] Kalisetty, S., Vankayalapati, R. K., Reddy, L., Sondinti, K., & Valiki, S. (2022). AI-Native Cloud Platforms: Redefining Scalability and Flexibility in Artificial Intelligence Workflows. *Linguistic and Philosophical Investigations*, 21(1), 1-15.

[42] Garapati, R. S. (2022). Web-Centric Cloud Framework for Real-Time Monitoring and Risk Prediction in Clinical Trials Using Machine Learning. *Current Research in Public Health*, 2, 1346.

[43] Hu, Y., Koren, Y., & Volinsky, C. (2008). Collaborative filtering for implicit feedback datasets. *Proceedings of the 2008 IEEE International Conference on Data Mining*, 263–272.

[44] Amistapuram, K. (2022). Fraud Detection and Risk Modeling in Insurance: Early Adoption of Machine Learning in Claims Processing. Available at SSRN 5741982.

[45] Davuluri, P. S. L. N. (2021). Event-Driven Compliance Systems: Modernizing Financial Crime Detection Without Machine Intelligence. *Journal of International Crisis and Risk Communication Research*, 339–354. <https://doi.org/10.63278/jicr.vi.3636>

[46] Meda, R. (2022). Integrating Edge AI in Smart Factories: A Case Study from the Paint Manufacturing Industry. *International Journal of Science and Research (IJSR)*, 1473-1489.

[47] Jagadish, H. V., Gehrke, J., Labrinidis, A., Papakonstantinou, Y., Patel, J. M., Ramakrishnan, R., & Shahabi, C. (2014). Big data and its technical challenges. *Communications of the ACM*, 57(7), 86–94.

[48] Segireddy, A. R. (2020). Cloud Migration Strategies for High-Volume Financial Messaging Systems.

[49] Khatri, V., & Brown, C. V. (2010). Designing data governance. *Communications of the ACM*, 53(1), 148–152.

[50] Amistapuram, K. (2021). Digital Transformation in Insurance: Migrating Enterprise Policy Systems to .NET Core. *Universal Journal of Computer Sciences and Communications*, 1(1), 1–17.

[51] Kumar, P., Singh, R., & Verma, A. (2022). Explainable machine learning models for fraud detection in healthcare insurance. *Knowledge-Based Systems*, 247, 108763.

[52] Nagabhyru, K. C. (2022). Bridging Traditional ETL Pipelines with AI Enhanced Data Workflows: Foundations of Intelligent Automation in Data Engineering. Available at SSRN 5505199.

[53] Lahiri, M., & Venkatasubramanian, S. (2013). Robust record linkage. *Proceedings of the 2013 ACM SIGMOD International Conference on Management of Data*, 101–112.

[54] Challa, K. (2021). Cloud Native Architecture for Scalable Fintech Applications with Real Time Payments. *International Journal Of Engineering And Computer Science*, 10(12).

[55] Leskovec, J., Rajaraman, A., & Ullman, J. D. (2014). *Mining of massive datasets* (2nd ed.). Cambridge University Press.

[56] Adusupalli, B., Singireddy, S., Sriram, H. K., Kaulwar, P. K., & Malempati, M. (2021). Revolutionizing Risk Assessment and Financial Ecosystems with Smart Automation, Secure Digital Solutions, and Advanced Analytical Frameworks. *Universal Journal of Finance and Economics*, 1(1), 101-122.

[57] Linden, G., Smith, B., & York, J. (2003). Amazon.com recommendations: Item-to-item collaborative filtering. *IEEE Internet Computing*, 7(1), 76–80.

[58] Meda, R. (2021). Digital Infrastructure for Predictive Inventory Management in Retail Using Machine Learning. *International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE)*, DOI, 10.

[59] Lin, J., Kolcz, A., & Szymanski, B. K. (2012). Large-scale machine learning at Twitter. *Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data*, 793–804.

[60] Sheelam, G. K. Power-Efficient Semiconductors for AI at the Edge: Enabling Scalable Intelligence in Wireless Systems. *International Journal of Innovative Research in Electrical, Elec-tronics, Instrumentation and Control Engineering (IJIREEICE)*, DOI, 10.

[61] Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). Big data: The next frontier for innovation, competition, and productivity. McKinsey Global Institute.

[62] Vadisetty, R., Polamarasetti, A., Guntupalli, R., Rongali, S. K., Raghunath, V., Jyothi, V. K., & Kudithipudi, K. (2021). Legal and Ethical Considerations for Hosting GenAI on the Cloud. *International Journal of AI, BigData, Computational and Management Studies*, 2(2), 28-34.

[63] Davenport, T., & Kalakota, R. (2019). The potential for artificial intelligence in healthcare. *Future Healthcare Journal*, 6(2), 94–98.

[64] Ramesh Inala. (2022). Cross-Domain MDM Integration Using AI-Driven Data Governance: A Case Study In Financial Technology Architecture. *Migration Letters*, 19(2), 280–304. Retrieved from <https://migrationletters.com/index.php/ml/article/view/11982>

[65] Marz, N., & Warren, J. (2015). *Big Data: Principles and Best Practices of Scalable Realtime Data Systems*. Manning Publications.



- [66] Aitha, A. R. (2021). Optimizing Data Warehousing for Large Scale Policy Management Using Advanced ETL Frameworks.
- [67] Zaharia, M., Chowdhury, M., Franklin, M. J., Shenker, S., & Stoica, I. (2010). Spark: Cluster computing with working sets. *Proceedings of the 2nd USENIX Conference on Hot Topics in Cloud Computing*, 1–7.
- [68] Challa, K. (2022). The Future of Cashless Economies Through Big Data Analytics in Payment Systems. *International Journal of Scientific Research and Modern Technology*, 60-70.
- [69] Zaharia, M., Das, T., Li, H., Shenker, S., & Stoica, I. (2012). Discretized streams: Fault-tolerant streaming computation at scale. *Proceedings of the 24th ACM Symposium on Operating Systems Principles*, 423–438.
- [70] Segireddy, A. R. (2021). Containerization and Microservices in Payment Systems: A Study of Kubernetes and Docker in Financial Applications. *Universal Journal of Business and Management*, 1(1), 1–17.
- [71] Zhai, C., & Massung, S. (2016). Text data management and analysis: A practical introduction to information retrieval and text mining. ACM & Morgan Claypool.
- [72] Davuluri, P. N. (2020). Event-Driven Architectures for Real-Time Regulatory Monitoring in Global Banking.
- [73] Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, 5, 135–146.
- [74] Annareddy, V. N., Preethish Nandan, B., Kommaragiri, V. B., Gadi, A. L., & Kalisetty, S. (2022). Emerging Technologies in Smart Computing, Sustainable Energy, and Next-Generation Mobility: Enhancing Digital Infrastructure, Secure Networks, and Intelligent Manufacturing
- [75] Goutham Kumar Sheelam. (2022). Reconfigurable Semiconductor Architectures For AI-Enhanced Wireless Communication Networks. *Kurdish Studies*, 10(2), 1027–1040. <https://doi.org/10.53555/ks.v10i2.3867>
- [76] Batarseh, F. A., & Yang, R. (2019). Federal data science: Transforming government and society. Academic Press.
- [77] Kolla, S. K. (2021). Architectural Frameworks for Large-Scale Electronic Health Record Data Platforms. *Current Research in Public Health*, 1(1), 1–19. Retrieved from <https://www.scipublications.com/journal/index.php/crph/article/view/1372>
- [78] Bhasin, H., & Bhatia, P. (2020). Clickstream data mining for web analytics and customer behavior modeling: A review. *ACM Computing Surveys*, 53(6), 1–34.
- [79] Maguluri, K. K., Pandugula, C., Kalisetty, S., & Mallesham, G. (2022). Advancing Pain Medicine with AI and Neural Networks: Predictive Analytics and Personalized Treatment Plans for Chronic and Acute Pain Managements. *Journal of Artificial Intelligence and Big Data*, 2(1), 112-126.
- [80] Goutham Kumar Sheelam, "Semiconductor Innovation for Edge AI: Enabling Ultra-Low Latency in Next-Gen Wireless Networks," *International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE)*, DOI: 10.17148/IJARCCE.2022.111258
- [81] Abedjan, Z., Golab, L., & Naumann, F. (2016). Profiling relational data: A survey. *The VLDB Journal*, 24(4), 557–581.
- [82] Yandamuri, U. S. (2022). Big Data Pipelines for Cross-Domain Decision Support: A Cloud-Centric Approach. *International Journal of Scientific Research and Modern Technology*, 1(12), 227–237. <https://doi.org/10.38124/ijsrmt.v1i12.1111>
- [83] Dwaraka Nath Kummari. (2022). AI-Driven Audit Frameworks For Enhancing Compliance In Modern Manufacturing Systems. *Migration Letters*, 19(S8), 2150–2177. Retrieved from <https://migrationletters.com/index.php/ml/article/view/11912>
- [84] Patel, S., Shah, M., & Gupta, R. (2022). Healthcare billing anomaly detection using machine learning. *Journal of Biomedical Informatics*, 132, 104128.
- [85] Baesens, B., Van Vlasselaer, V., & Verbeke, W. (2021). Fraud analytics using descriptive, predictive, and social network techniques: A guide to data science for fraud detection (2nd ed.). Wiley.
- [86] Avinash Reddy Aitha. (2022). Deep Neural Networks for Property Risk Prediction Leveraging Aerial and Satellite Imaging. *International Journal of Communication Networks and Information Security (IJCNIS)*, 14(3), 1308–1318. Retrieved from <https://www.ijcnis.org/index.php/ijcnis/article/view/8609>
- [87] Barocas, S., Hardt, M., & Narayanan, A. (2019). Fairness and machine learning. fairmlbook.org (Book manuscript).
- [88] Paleti, S. (2022). Financial Innovation through AI and Data Engineering: Rethinking Risk and Compliance in the Banking Industry. Available at SSRN 5250726.
- [89] Gottimukkala, V. R. R. (2020). Energy-Efficient Design Patterns for Large-Scale Banking Applications Deployed on AWS Cloud. *power*, 9(12).
- [90] Ahmad, M. A., Eckert, C., & Teredesai, A. (2018). Interpretable machine learning in healthcare. *Proceedings of the ACM Conference on Health, Informatics, and Data Science*, 1–10.
- [91] Aitha, A. R. (2022). Cloud Native ETL Pipelines for Real Time Claims Processing in Large Scale Insurers. Available at SSRN 5532601.



- [92] Inala, R. Advancing Group Insurance Solutions Through Ai-Enhanced Technology Architectures And Big Data Insights.
- [93] Avinash Reddy Segireddy. (2022). Terraform and Ansible in Building Resilient Cloud-Native Payment Architectures. *International Journal of Intelligent Systems and Applications in Engineering*, 10(3s), 444–455. Retrieved from <https://www.ijssae.org/index.php/IJISAE/article/view/7905>
- [94] Paleti, S., Singireddy, J., Dodda, A., Burugulla, J. K. R., & Challa, K. (2021). Innovative financial technologies: Strengthening compliance, secure transactions, and intelligent advisory systems through ai-driven automation and scalable data architectures. *Secure Transactions, and Intelligent Advisory Systems Through AI-Driven Automation and Scalable Data Architectures* (December 27, 2021).
- [95] Armbrust et al. (2010) – A view of cloud computing, *Communications of the ACM*
- [96] Sheelam, G. K., & Nandan, B. P. (2022). Integrating AI And Data Engineering For Intelligent Semiconductor Chip Design And Optimization. *Migration Letters*, 19, 2178-2207.
- [97] Kimball, R., & Ross, M. (2013). *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling*. John Wiley & Sons
- [98] Dodda, A., Lakkarasu, P., Singireddy, J., Challa, K., & Pamisetty, V. (2022). Optimizing Digital Finance and Regulatory Systems Through Intelligent Automation. *Secure Data Architectures, and Advanced Analytical Technologies*.
- [99] Chaudhuri, S., & Dayal, U. (1997). An Overview of Data Warehousing and OLAP Technology. *ACM SIGMOD Record*, 26(1), 65–74.
- [100] Pandiri, L., Singireddy, S., & Adusupalli, B. (2020). Digital Transformation of Underwriting Processes through Automation and Data Integration. *Global Research Development (GRD) ISSN*, 2455-5703.
- [101] Stonebraker, M., Abadi, D. J., DeWitt, D. J., Madden, S., Paulson, E., Pavlo, A., & Rasin, A. (2010). MapReduce and Parallel DBMSs: Friends or Foes? *Communications of the ACM*, 53(1), 64–71.
- [102] Keerthi Amistapuram, "Energy-Efficient System Design for High-Volume Insurance Applications in Cloud-Native Environments," *International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering (IJIREEICE)*, DOI 10.17148/IJIREEICE.2020.81209
- [103] Grolinger, K., Higashino, W. A., Tiwari, A., & Capretz, M. A. M. (2013). Data Management in Cloud Environments: NoSQL and NewSQL Data Stores. *Journal of Cloud Computing*, 2(22), 1–24.
- [104] Kolla, S. H. (2021). Rule-Based Automation for IT Service Management Workflows. *Online Journal of Engineering Sciences*, 1(1), 1–14. Retrieved from <https://www.scipublications.com/journal/index.php/ojes/article/view/1360>
- [105] Vadisetty, R., Polamarasetti, A., Guntupalli, R., Raghunath, V., Jyothi, V. K., & Kudithipudi, K. (2021). Privacy-Preserving Gen AI in Multi-Tenant Cloud Environments. Sateesh kumar and Raghunath, Vedapraada and Jyothi, Vinaya Kumar and Kudithipudi, Karthik, *Privacy-Preserving Gen AI in Multi-Tenant Cloud Environments* (January 20, 2021).
- [106] Nandan, B. P. (2022). AI-Powered Fault Detection In Semiconductor Fabrication: A Data-Centric Perspective.