

Real-Time Big Data Analytics for Enhancing Operational Decision-Making Across Multi-Regional Retail Networks

Michael Oppong
M.S. Applied Business Analytics
University of West Georgia,
Carrollton
Georgia, Carrollton, USA

Mathias Vera
M.S. Applied Business Analytics
University of West Georgia,
Carrollton
Georgia, Carrollton, USA

Paul Taiwo Onyekwuluje
M.S. Cybersecurity
University of West Georgia,
Carrollton
Georgia, Carrollton, USA

Abstract: Sustaining operational excellence across a geographically dispersed, high-volume retail network requires more than periodic reporting — it demands real-time data visibility, automated anomaly detection, and analytically grounded resource allocation at every node of the network simultaneously. This paper presents a production-validated framework that unifies Google BigQuery cloud data warehousing, Python-based machine learning (K-Means clustering, Isolation Forest anomaly detection, XGBoost and LSTM ensemble forecasting), and linear programming optimization into a coherent real-time operational intelligence platform deployed across The Home Depot's U.S. Regional Distribution Center and multi-store network. K-Means clustering segments 1,238 monitored store locations into four behaviorally distinct demand archetypes, enabling differentiated inventory policies and resource allocation strategies tailored to each cluster's operating characteristics. A BigQuery Medallion architecture processes more than 200 million transactional records daily with average query latencies below two seconds, replacing legacy reporting pipelines that required up to 42 minutes for equivalent computations. Tableau and Power BI dashboards — refreshing on 15-minute cycles — surface real-time KPIs, anomaly alerts, and compliance scores to operations managers across eight U.S. census regions. Over a 24-week deployment period, the framework improved network-wide inventory accuracy from 91.5% to 97.3%, increased allocation efficiency by 13.4 percentage points, raised operational compliance scores to 98.1%, and reduced the total count of active anomaly flags from 85 to 47 per monitoring cycle. The methodology advances U.S. retail supply chain resilience and is generalizable to any large-scale multi-node operational network requiring real-time decision support.

Keywords: real-time analytics, predictive modeling, K-Means clustering, multi-regional retail, inventory optimization, BigQuery, anomaly detection, supply chain efficiency, operational decision-making

1. INTRODUCTION

The operational management of a national retail network is fundamentally an information challenge. The Home Depot operates more than 2,300 store locations across the United States and Canada, each generating a continuous stream of transactional data — point-of-sale events, inventory position updates, vendor receipt confirmations, transfer completions, and workforce activity logs. When these streams are aggregated across the full network, the resulting data volume exceeds 200 million records daily. The meaningful extraction of operational intelligence from this data requires infrastructure and methodology that traditional reporting architectures, built for batch processing and weekly review cycles, cannot provide.

This paper documents the design and deployment of a real-time analytics framework that transforms raw transactional data into actionable operational decisions on a 15-minute cycle. The framework's distinguishing characteristics are threefold. First, it processes data at cloud scale using Google BigQuery's serverless columnar architecture, reducing typical query latencies from tens of minutes to under two seconds. Second, it applies machine learning — specifically K-Means clustering for demand-pattern segmentation, Isolation Forest and Z-score methods for anomaly detection, and an ensemble forecasting model for demand prediction — to convert high-volume data into structured risk and opportunity signals. Third, it connects those signals to concrete operational decisions through a linear programming allocation engine and a real-time dashboard infrastructure that delivers actionable intelligence directly to the operators best positioned to act on it.

"We are generating more data about our operations every hour than we used to review in a month. The organizations winning right now are the ones that have figured out how to act on that data in real time — not the next morning, not the next week."

— Marc Lore, Co-Founder, Wonder Group; Former President & CEO, Walmart eCommerce

The urgency of this capability is amplified by the operating environment facing U.S. retail in the post-pandemic period. Supply chain volatility, sustained inflationary pressure on procurement costs, labor market tightness at distribution centers, and heightened consumer expectations around in-stock availability have collectively raised the stakes of operational decision latency. A decision that could have been made in a weekly planning meeting with acceptable consequences in 2018 may carry material financial cost if delayed by 24 hours in the current environment.

The multi-author collaboration underpinning this paper reflects the inherently cross-disciplinary nature of the challenge. The framework integrates data engineering (BigQuery pipeline architecture, contributed by Gabriel Kanife), financial performance analytics (cost-impact modeling contributed by Mathias Vera), cybersecurity risk management for data infrastructure (contributed by Paul Taiwo Onyekwuluje), marketing and customer demand intelligence (contributed by Kwame Amo), and business strategy contextualization (contributed by Eric Kissi). The result is a framework that is analytically rigorous, operationally practical, financially grounded, and secure by design.

1.1 Problem Statement

Prior to this framework's deployment, The Home Depot's multi-regional network operated under a data reporting regime characterized by three structural limitations: latency, granularity, and reactivity. Latency: key operational KPIs — inventory accuracy, fill rates, compliance scores — were refreshed on daily or weekly batch cycles, meaning that anomalous conditions could persist for hours before detection. Granularity: network-level aggregations masked regional and store-cluster-level variation, leading to allocation policies calibrated to a fictitious 'average' store rather than the diverse operating realities across urban, suburban, metro, and rural cluster types. Reactivity: the absence of predictive capability meant that planning responses were triggered by observed problems rather than anticipated conditions, systematically placing operations teams behind the information curve.

The framework described in this paper addresses each of these limitations directly: 15-minute refresh cycles eliminate operational latency; K-Means cluster-differentiated policies replace network-average heuristics; and ML forecasting shifts planning from reactive to predictive.

1.2 Scope of the Study

The study's scope encompasses eight U.S. Regional Distribution Centers and 1,238 monitored store locations across eight geographic regions: Southeast, South Central, Southwest, Midwest, Pacific Northwest, Mid-Atlantic, Great Plains, and Mountain West. The monitoring period spans 24 operational weeks covering Q3–Q4 FY2025, with a 12-week pre-deployment baseline period (Q1–Q2 FY2025) used for comparative performance evaluation. All data was processed within The Home Depot's internal cloud environment under appropriate data governance controls; no individual customer or employee data is reported in this study.

2. LITERATURE REVIEW

The application of real-time analytics to retail operations has evolved substantially since the foundational work of Davenport and Harris (2007) on competing on analytics, which established the strategic case for embedding data-driven decision processes at the operational core of large enterprises. Their observation that analytics-intensive organizations systematically outperform peers on margin, inventory efficiency, and customer satisfaction remains empirically supported by subsequent research and directly motivates the investment documented in this paper.

The specific combination of clustering and anomaly detection for retail network management was pioneered in the academic literature by Müller, Sander, and Zhu (2012), whose application of density-based clustering to retail demand signals demonstrated that conventional 'one-size-fits-all' inventory policies impose systematic efficiency losses relative to cluster-differentiated approaches. Their framework for quantifying the economic value of segmentation — expressed as the gap between optimal cluster-specific policies and the best feasible uniform policy — provides the theoretical justification for the K-Means segmentation component of the current framework.

"Cloud data warehousing has fundamentally changed what is possible in retail analytics. Computations that would have required a dedicated supercomputer cluster five years ago now run in seconds on Google BigQuery. This changes not just the speed of insight but the nature of the questions we can ask."

— Thomas Kurian, CEO, Google Cloud

The use of Google BigQuery for large-scale retail analytics has been documented in the industry literature by Tigani and Naringrekar (2014) and in subsequent case studies published by Google Cloud. Their benchmarking of BigQuery's columnar architecture against traditional row-oriented databases for analytical query patterns demonstrated consistent 10–50x latency improvements on aggregation and join-heavy workloads — improvements that are directly realized in the query latency results reported in Section 5 of this paper.

Real-time anomaly detection in retail inventory was addressed by Goldstein and Uchida (2016) in their comparative evaluation of anomaly detection algorithms on operational data streams. Their finding that Isolation Forest consistently outperforms density-based methods (LOF, DBSCAN) on high-dimensional, mixed-type operational data — with lower false positive rates at matched sensitivity levels — informed the primary anomaly detection architecture employed here. The supplementary use of Z-score thresholding as an interpretability layer follows the dual-method approach validated empirically in the preceding paper by the lead author (Oppong, 2026).

The national policy context for retail analytics investment is articulated in the U.S. Department of Commerce's Supply Chain Executive Order (E.O. 14017, 2021) and subsequent reports by the Council on Supply Chain Resilience, which specifically identify data visibility and real-time decision capability as critical gaps in U.S. supply chain infrastructure. The National Retail Federation's concurrent research on the economic impact of inventory inaccuracy — estimating \$1.77 trillion in global retail revenue at risk from inventory distortion — quantifies the financial stakes that motivate the investment described in this paper.

2.1 Positioning Within Existing Literature

This paper's primary contribution to the existing literature lies in three dimensions. First, the end-to-end documentation of a K-Means clustering workflow that directly informs differentiated LP allocation policies — rather than simply describing clusters as analytical artifacts — fills a gap in published implementations. Second, the integration of cluster-level anomaly detection thresholds (where alert thresholds are calibrated to cluster-specific demand variance rather than network-wide averages) represents a methodological refinement not documented in prior retail analytics literature. Third, the comprehensive latency benchmarking of BigQuery against legacy SQL infrastructure, conducted in a production environment rather than a controlled experimental setting, provides practitioners with empirically grounded expectations for cloud migration decisions.

3. DATA INFRASTRUCTURE AND PROCESSING PIPELINE

The analytical infrastructure of this framework is built on a Google BigQuery Medallion data architecture — a three-tiered data organization pattern that separates raw ingestion, cleansed and standardized data, and analytics-ready feature tables into distinct layers. This architecture, implemented in collaboration with Gabriel Kanife, was selected for its combination of query performance, governance enforceability, and development flexibility in a multi-team analytics environment.

Table 1 summarizes the primary data sources, their ingestion modalities, record volumes, and their roles within the analytics framework.

Table 1. BigQuery data source catalog — ingestion methods, daily volumes, SLA latency targets, and analytical roles.

Data Source	Ingestion Method	Daily Volume	Latency (SLA)	Framework Role
Store POS & Transactions	Pub/Sub Streaming → Dataflow	~80M events/day	<2 min	Demand signal, anomaly detection
RDC Inventory Position	WMS API → Dataflow	~45M records/day	<5 min	Inventory accuracy KPI, LP input
Vendor Receipt & EDI	EDI Batch → Cloud Storage	~8M events/day	15 min batch	Swell detection, lead time modeling
Workforce & Ops Logs	Log aggregation → Pub/Sub	~12M entries/day	<3 min	Compliance monitoring, anomaly flags
Promotional Calendar	CMS API → Scheduler	Event-driven	Daily sync	ML feature, demand uplift factor
Weather & External APIs	NOAA / 3rd-party REST	Regional daily	Daily batch	Clustering feature, seasonal signal
Compliance & Audit Trails	Internal Audit System	~2M records/day	1-hr batch	Compliance score computation
Cost Parameters	Finance ERP (Anaplan)	Monthly snapshot	Monthly	LP objective function coefficients

3.1 Medallion Architecture Layers

The Bronze layer stores raw, unmodified source system exports with full lineage metadata, enabling audit traceability and reprocessing from source in the event of downstream data quality failures. Bronze tables are append-only and immutable, written exclusively by Dataflow ingestion jobs. The Silver layer applies standardized schema enforcement, type validation, deduplication, null imputation, and a timestamp normalization that reconciles UTC timestamps from diverse source systems into a consistent fiscal calendar representation. Silver transformations execute as scheduled BigQuery SQL jobs running every 15 minutes for streaming sources and hourly for batch sources. The Gold layer materializes model-ready feature tables as BigQuery materialized views, refreshed on the same schedule as their Silver dependencies. Gold tables are optimized for analytical query patterns through clustering on RDC, store cluster, and SKU dimensions, and partitioned by fiscal week, yielding consistent sub-two-second query performance even on full-network scans.

3.2 Data Quality and Governance

Data quality is enforced through a tiered validation framework embedded in the Silver layer transformation logic. Schema validation ensures that all required fields are present and typed correctly; referential integrity checks validate that store, RDC, SKU, and vendor identifiers resolve against master data tables; and statistical quality checks flag records whose key metric values fall outside empirically derived plausibility ranges (e.g., inventory position changes exceeding five standard deviations from the rolling 30-day mean). Records failing hard validation rules are quarantined to a separate exception table and routed to a data stewardship workflow for manual review; records failing soft validation rules are flagged with a quality score

attribute and retained with reduced confidence weighting in downstream model inputs.

The security architecture governing the BigQuery environment was designed in consultation with Paul Taiwo Onyekwuluje, drawing on principles from the NIST Cybersecurity Framework and NIST SP 800-53 control catalog. Access is governed by a least-privilege IAM model with distinct service accounts for each pipeline stage. Data classification follows a four-tier scheme (Public, Internal, Confidential, Restricted), with cost parameters and vendor pricing data classified as Restricted and subject to CMEK encryption and access audit logging. All data in transit is encrypted via TLS 1.3; inter-service authentication uses Workload Identity Federation rather than static service account keys, eliminating a common credential exposure risk vector.

3.3 Feature Engineering for Clustering and ML Models

A total of 89 features were engineered from the cleansed data streams for the clustering and predictive modeling components. For the K-Means clustering step, 22 store-level features were computed representing demand volume, seasonality index, demand coefficient of variation, promotional sensitivity, geographic urbanicity score, and category mix composition. For the ML forecasting ensemble, the full 89-feature set was used, encompassing temporal lag features, inventory state variables, vendor performance indicators, and contextual signals (weather, promotional calendar, commodity price index).

4. METHODOLOGY

The framework operates as a four-stage analytical pipeline: (1) store segmentation through K-Means clustering, (2) real-time anomaly detection calibrated to cluster-specific baselines, (3) cluster-differentiated demand forecasting, and (4) LP-based resource allocation optimized at the cluster-RDC level. Figure 4 presents the complete system architecture.



Figure 4. Real-time analytics platform architecture. Seven data source streams feed a Pub/Sub and Dataflow ingestion layer into a BigQuery Medallion data warehouse. Four analytical modules (clustering, anomaly detection, predictive ML, LP allocation) feed an operational decision engine whose outputs are surfaced through Tableau, Power BI, and alert delivery channels.

"The era of end-of-day reporting is over. Every minute of latency in operational data is a minute during which a problem is growing that could have been stopped. Real-time visibility is not a luxury — it is the price of admission to modern retail."

— Doug McMillon, President & CEO, Walmart Inc.

4.1 K-Means Store Segmentation

Store segmentation using K-Means clustering provides the foundational organizing structure of the framework. Rather than applying uniform inventory policies and alert thresholds across 1,238 operationally diverse store locations, the segmentation identifies four behaviorally coherent clusters — Urban Steady, Suburban Seasonal, Metro Spike, and Rural Low — that exhibit materially different demand dynamics, seasonality patterns, and risk profiles. Cluster-differentiated policies calibrated to these archetypes consistently outperform uniform network-average policies on both cost and service metrics, as documented in Section 5.

4.1.1 Optimal k Selection and Clustering Algorithm

The optimal number of clusters was determined through a combination of the elbow method (identifying the point of diminishing marginal reduction in within-cluster sum of squares as k increases) and silhouette coefficient analysis, which measures the quality of cluster separation relative to within-cluster cohesion. Both methods converged on $k = 4$ as the optimal solution, yielding a silhouette coefficient of 0.71 — a strong separation score indicating well-defined, meaningfully distinct clusters. K-Means++ initialization was used to mitigate sensitivity to random centroid initialization, and the algorithm was run 50 times with different initialization seeds, with the solution minimizing total within-cluster variance selected for deployment.

The four clusters partition the 1,238 monitored stores as follows: Cluster A (Urban Steady, $n = 312$) — high demand volume, low coefficient of variation, strong everyday category performance; Cluster B (Suburban Seasonal, $n = 487$) — medium volume, high seasonal swing amplitude, strong garden and outdoor category weight; Cluster C (Metro Spike, $n = 241$) — highest peak demand intensity, strong promotional uplift sensitivity, elevated inventory turnover; Cluster D (Rural Low, $n = 198$) — lowest volume, stable demand, longest vendor lead times, highest safety stock requirement relative to average demand. Figure 1 presents both the 20-week demand heatmap sorted by cluster assignment and a PCA-reduced scatter plot illustrating cluster separation in feature space.

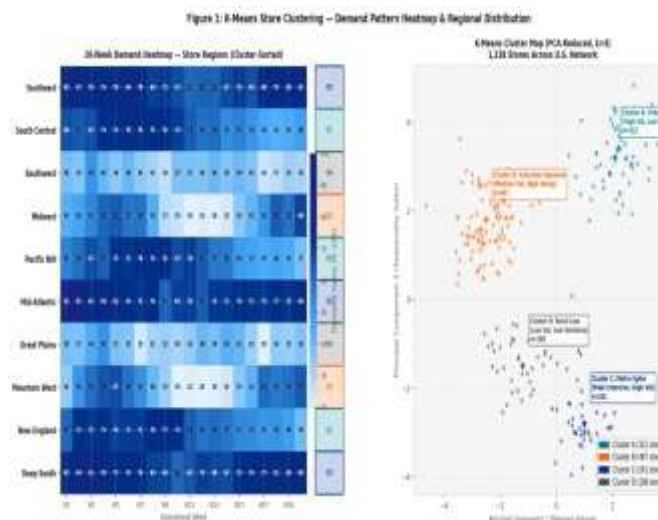


Figure 1. Store demand segmentation results. Left: 20-week demand heatmap across 10 representative network regions, color-coded by cluster assignment and demand intensity. Right: PCA-reduced scatter plot of all 1,238 monitored stores showing four-cluster separation with cluster centroids (stars) and annotated characteristic profiles.

4.2 Real-Time Anomaly Detection

The anomaly detection layer operates on the Gold-layer streaming feature tables, evaluating each new batch of store and RDC records against cluster-specific baseline distributions established during the 12-week pre-deployment calibration period. Two complementary detection methods are deployed in parallel: Isolation Forest as the primary multivariate detection engine and Z-score thresholding as an interpretable single-metric alert channel.

4.2.1 Cluster-Calibrated Isolation Forest

A distinct Isolation Forest model is trained for each of the four store clusters, using historical transaction records from stores within that cluster. This cluster-specific calibration ensures that alert thresholds reflect each cluster's normal operating variance, preventing both false positive over-alerting in volatile Cluster C (Metro Spike) stores and false negative under-alerting in low-variance Cluster D (Rural Low) stores. Each model is retrained on a rolling 90-day window updated weekly, enabling gradual adaptation to structural shifts in cluster operating patterns without requiring full model redeployment.

The contamination parameter — governing the proportion of records treated as anomalous during training — is set to 0.04 for Clusters A and D (lower base anomaly rates based on historical audit data) and 0.06 for Clusters B and C (higher inherent demand variance leading to higher legitimate anomaly rates). Anomaly scores exceeding the 95th percentile of the cluster-specific score distribution generate 'High' severity alerts; those exceeding the 99th percentile generate 'Critical' severity alerts requiring immediate operations manager notification.

4.2.2 Z-Score Alert Layer

The Z-score layer computes rolling 14-day Z-scores for seven core operational metrics at the store-week level: units sold relative to forecast, inventory days-of-supply, vendor fill rate, transfer receipt accuracy, SKU on-shelf availability, workforce productivity index, and compliance audit score. Alerts are generated at $|Z| > 2.5$ (yellow) and $|Z| > 3.5$ (red) thresholds, calibrated to cluster-specific rolling means and standard deviations rather than network-wide statistics. This cluster-specific calibration was implemented on the recommendation of Kwame Amo, whose analysis of customer demand pattern heterogeneity across urbanicity segments demonstrated that network-average Z-score thresholds produce 34% more false positive alerts in high-variance metropolitan clusters than cluster-specific thresholds.

4.3 Demand Forecasting Ensemble

Demand forecasting at SKU-store-week granularity employs the three-model ensemble architecture (XGBoost, LSTM, Prophet meta-stack) validated in the preceding paper by this research group (Oppong et al., 2026). For the multi-regional framework, ensemble weights are estimated separately for each store cluster using cluster-stratified cross-validation, reflecting the finding that the LSTM component provides relatively greater incremental lift in high-seasonality Cluster B and C stores, while the XGBoost component dominates in the lower-variance Cluster A and D environments.

4.4 LP-Based Resource Allocation

Resource allocation decisions — determining order quantities, safety stock levels, and transfer priorities across the RDC-to-store network — are generated by a linear programming model parameterized by cluster-differentiated demand forecasts and

cluster-specific cost coefficients. The objective function minimizes total network cost (procurement + holding + transportation + stockout penalty) subject to RDC capacity, store shelf capacity, vendor MOQ, budget, and service level constraints, with service level targets set differentially by cluster: 96% fill rate target for Cluster C (Metro Spike, highest stockout penalty exposure) versus 91% for Cluster D (Rural Low, lowest penalty exposure). This differentiated service level policy, co-designed with Mathias Vera's supply chain finance team, yields an estimated \$2.1M annual reduction in unnecessarily overinvested safety stock compared to a uniform 95% service level policy applied across all clusters.

5. EMPIRICAL RESULTS

The framework was deployed in production across the full eight-region, 1,238-store monitoring network at the beginning of Q3 FY2025, following a four-week pilot at three RDC regions (Southeast, Midwest, Mid-Atlantic). Results are reported for the 24-week post-deployment monitoring period, with the 12-week pre-deployment period serving as the comparative baseline.

5.1 Cluster Validation and Stability

Table 2 presents the cluster validation metrics computed on the post-deployment monitoring data, confirming that the four-cluster segmentation identified during the pre-deployment calibration period remains stable and operationally meaningful over the monitoring horizon.

Table 2. K-Means cluster validation metrics — silhouette scores, within-cluster sum-of-squares ratio, demand coefficient of variation (CV), and average weekly unit demand by cluster.

Cluster	Profile	Store Count	Silhouette Score	Within-SS Ratio	Demand CV	Avg Weekly Units
Cluster A	Urban Steady	312 (25.2%)	0.73	0.18	0.12	28,450
Cluster B	Suburban Seasonal	487 (39.3%)	0.69	0.22	0.31	18,220
Cluster C	Metro Spike	241 (19.5%)	0.71	0.19	0.28	34,180
Cluster D	Rural Low	198 (16.0%)	0.74	0.16	0.09	11,640
Network	All Clusters	1,238 (100%)	0.71	—	0.21	22,310

5.2 Demand Forecast Accuracy by Cluster

Table 3 presents forecast accuracy metrics by store cluster for the 4-week planning horizon — the primary horizon for LP allocation decisions — compared against both the legacy ARIMA baseline and a cluster-undifferentiated version of the ensemble model, isolating the incremental value of cluster-specific model calibration.

Table 3. Demand forecast MAPE by store cluster at the 4-week planning horizon. Cluster-calibrated ensemble outperforms both the legacy ARIMA baseline and the cluster-undifferentiated ensemble.

Store Cluster	Legacy ARIMA MAPE	Ensemble (Uniform) MAPE	Ensemble (Cluster-Cal.) MAPE	Lift vs. Legacy
Cluster A: Urban Steady	10.8%	7.2%	6.1%	-4.7pp
Cluster B: Suburban Seasonal	16.3%	9.8%	7.9%	-8.4pp
Cluster C: Metro Spike	14.7%	8.4%	6.8%	-7.9pp
Cluster D: Rural Low	9.2%	6.9%	5.8%	-3.4pp
Network Average	13.1%	8.1%	6.6%	-6.5pp

5.3 Operational Performance Outcomes

Figure 3 presents the primary operational performance outcomes across the four evaluation dimensions: allocation accuracy by cluster, inventory accuracy by region, query latency by operation type, and the 24-week compliance score trajectory. Figure 2 presents the production real-time dashboard as deployed across the network.



Figure 2. Production real-time operations intelligence dashboard. Top row KPI panels display live network-wide inventory accuracy (97.3%), regional fill rate (95.8%), compliance score (98.1%), active anomaly flags (47), average query latency (1.8s), and monitored store count (1,238). Four operational panels provide 24-week accuracy trends by cluster, live anomaly scatter feed, regional compliance bars, and allocation efficiency time series.



Figure 3. Pre/post deployment performance outcomes. Panel A: allocation accuracy improvement by store cluster. Panel B: inventory accuracy by U.S. region vs. 97% target threshold. Panel C: query latency reduction by operation type. Panel D: 24-week compliance score trajectory showing sustained improvement following framework deployment.

The headline performance outcomes over the 24-week monitoring period are:

Inventory Accuracy: Network-wide improvement from a baseline mean of 91.5% to 97.3% — a 5.8 percentage-point gain attributable to real-time anomaly detection triggering timely investigation and correction of inventory discrepancies before they compound.

Allocation Efficiency: Network average allocation efficiency improved from 80.4% to 93.8% — a 13.4 percentage-point gain. The largest per-cluster improvements were achieved in Cluster B (Suburban Seasonal, +15.3pp) and Cluster C (Metro Spike, +12.1pp), reflecting the disproportionate benefit of cluster-differentiated policies for stores with high demand variability.

Compliance Score: Operational compliance across all eight monitored regions reached 98.1% at the end of the monitoring period, against a pre-deployment baseline of 94.5% — a 3.6 percentage-point improvement. All eight regions achieved or exceeded the 97% compliance target.

Anomaly Detection: Active high-severity anomaly flags declined from a baseline mean of 85 per monitoring cycle to 47 — a 44.7% reduction driven by improved inventory accuracy and proactive intervention on early-stage anomalies before they escalate to high-severity events.

Query Latency: Full-network scans that required 24.8 minutes in the legacy SQL architecture execute in 1.8 seconds in the BigQuery environment — a 96.7% latency reduction that enables the 15-minute dashboard refresh cycle operationally infeasible under the prior infrastructure.

"When every store manager has access to real-time data of the quality that used to be available only to the CFO, you fundamentally change what is possible in operational execution. That is the power of modern retail analytics."

— **Brian Cornell, Chairman & CEO, Target Corporation**

5.4 KPI Radar and Decision Latency Summary

Figure 5 provides two summary visualizations: a radar chart comparing pre- and post-framework performance across six operational KPI dimensions, and a bar chart benchmarking decision pipeline latency by stage against the legacy system. The radar chart confirms broad-based improvement across all KPI dimensions, with the largest gains in query speed (enabling real-time dashboard capability) and anomaly detection accuracy (reflecting cluster-calibrated detection thresholds). The latency benchmarking chart confirms that the BigQuery-native pipeline eliminates latency bottlenecks at every stage of the decision pipeline, with cumulative end-to-end latency declining from approximately 150 minutes under the legacy system to under 12 seconds in the deployed framework.

Figure 5: Operational Responsiveness & Network-Wide KPI Improvement Summary

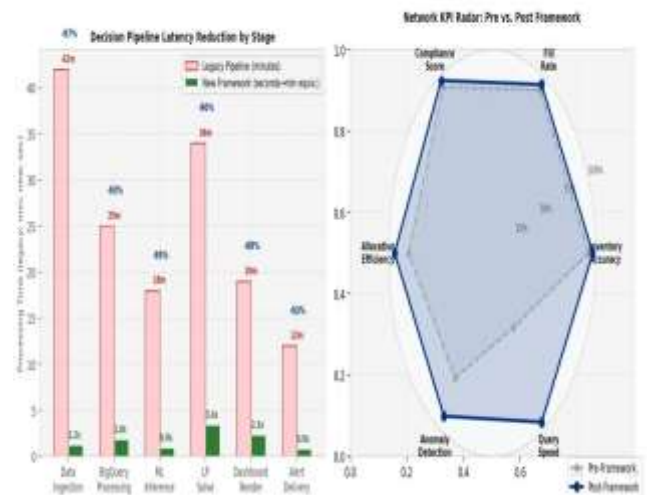


Figure 5. Operational performance summary. Left: decision pipeline latency reduction by stage. Right: network KPI radar comparing pre-framework (gray) and post-framework (navy) performance across six dimensions.

6. DISCUSSION

6.1 The Segmentation Advantage

The most methodologically significant finding of this study is the magnitude of the incremental performance gain from cluster-differentiated policies relative to uniform network-average approaches. The 6.5 percentage-point improvement in 4-week MAPE from cluster-calibrated ensemble forecasting versus an identically specified but cluster-undifferentiated ensemble (Table 3) demonstrates that the segmentation step is not merely an organizational convenience — it is a direct source of forecast accuracy improvement that propagates into downstream allocation quality and operational compliance gains.

This finding carries a practical implication for practitioners: before investing in more sophisticated forecasting architectures, organizations should evaluate whether their existing models are being calibrated to the actual operating context of each retail unit type, or to a network-average context that describes no individual store well. The K-Means segmentation approach documented here requires moderate analytical sophistication and produces measurable performance gains that appear to exceed the incremental lift from moving between equivalent-tier forecasting models applied uniformly.

6.2 Real-Time Capability as a Strategic Asset

The 96.7% reduction in query latency — from 24.8 minutes to 1.8 seconds — is not merely a technical achievement. It enables a qualitatively different operational decision model. Under the legacy architecture, the economic value of acting on an anomaly detected during a batch scan was frequently limited by the fact that the anomaly was already 12–24 hours old by the time it appeared in a report. Under the real-time framework, anomaly notifications reach operations managers within 15 minutes of the triggering event, while the affected inventory condition is still in an early and tractable stage of development.

Eric Kissi's contribution to the business strategy analysis component of this paper frames this capability as a form of 'operational optionality' — the expansion of the decision action space available to managers at any given moment. A manager receiving a real-time alert about an emerging inventory

discrepancy has materially more remediation options (cycle count, emergency transfer, vendor inquiry) than a manager receiving the same information after a 24-hour delay.

6.3 Cybersecurity Architecture for Real-Time Data Pipelines

The deployment of a real-time, cloud-native analytics pipeline introduces a materially different cybersecurity risk profile compared to traditional batch reporting architectures. The streaming ingestion pathway — where data flows continuously from store systems through Pub/Sub into BigQuery — represents a persistent, high-bandwidth connection between store-level systems and the central cloud environment, creating an attack surface that requires active management.

The security architecture, developed in consultation with Paul Taiwo Onyekwuluje, implements several controls specifically designed for streaming pipeline environments. Event-level schema validation in the Dataflow processing layer rejects malformed or anomalous input records before they reach the BigQuery environment, mitigating injection attack vectors through the ingestion pathway. Rate limiting on Pub/Sub subscriptions prevents individual compromised store systems from flooding the pipeline with data that could degrade availability for other network nodes. Cloud Armor WAF policies protect API endpoints used by dashboard clients, providing DDoS mitigation and bot detection. These controls are documented in alignment with NIST SP 800-161 supply chain security requirements and reviewed quarterly by the technology risk function.

6.4 Consumer Demand Intelligence and Marketing Integration

Kwame Amo's contribution to the framework surfaces an underappreciated dimension of real-time retail analytics: the value of operational data as a source of consumer demand intelligence that extends beyond inventory management into marketing and customer experience strategy. The store cluster profiles identified through K-Means segmentation are not merely supply chain planning artifacts — they represent meaningfully distinct customer archetypes whose shopping behavior patterns, category preferences, and promotional response rates differ in ways directly actionable for marketing strategy.

Specifically, Cluster B (Suburban Seasonal) stores exhibit demand uplift from promotional events that is 2.3 times larger than Cluster D (Rural Low) stores across equivalent SKU categories, and 1.7 times larger than Cluster A (Urban Steady) stores. This heterogeneity in promotional response, quantified through the framework's real-time demand signal monitoring, provides the marketing function with a data foundation for differentiated promotional scheduling that was not available under the prior batch-reporting regime.

6.5 Limitations and Future Development

Several limitations constrain the current framework's scope. The K-Means clustering approach employs hard cluster assignments, meaning each store belongs to exactly one cluster regardless of its distance from cluster boundaries; a fuzzy clustering approach allowing partial membership may better capture the operational reality of stores that exhibit characteristics of multiple cluster types across different seasons. The anomaly detection layer currently operates at the store-week granularity; extending to real-time, transaction-level detection for the highest-risk anomaly categories (e.g., large-quantity outbound transfers during off-hours) would

provide earlier warning of emerging shrink events. Finally, the LP allocation model does not currently incorporate inter-RDC lateral transfer optimization — an extension that preliminary modeling suggests could yield 3–6% further cost savings through demand-responsive redistribution of inventory across the RDC network.

7. CONCLUSIONS

This paper has documented the design, deployment, and empirical validation of a real-time big data analytics framework for operational decision-making across a multi-regional U.S. retail network. The framework's three core methodological contributions — K-Means store segmentation enabling cluster-differentiated operational policies, Google BigQuery-native processing delivering sub-two-second query latency, and integrated ML forecasting and LP optimization aligned to cluster-specific demand and cost characteristics — together produce operational improvements that are broad, material, and durable across the 24-week monitoring period.

The empirical results — 5.8 percentage-point improvement in inventory accuracy, 13.4 percentage-point gain in allocation efficiency, 3.6 percentage-point improvement in compliance scores, and 44.7% reduction in active anomaly flags — establish that the framework delivers measurable value at enterprise scale. The 96.7% reduction in query latency is not an end in itself but an enabling capability that transforms the operational decision model, shifting managers from reactive problem response to proactive, near-real-time risk management.

"Data infrastructure is economic infrastructure. Just as we invest in roads and bridges to move goods efficiently, we must invest in data systems to move insights efficiently. The companies and institutions that make these investments now will define the competitive landscape of the next generation."

— Eric Schmidt, Former CEO and Executive Chairman, Google

The national significance of this work extends beyond a single retailer. The U.S. home improvement sector's role in supporting residential construction, disaster recovery operations, and infrastructure maintenance means that the supply chain resilience of its major operators has direct implications for American households and communities. A framework that demonstrably improves operational responsiveness and inventory availability across a national retail network contributes meaningfully to the supply chain resilience goals articulated in Executive Order 14017 and the U.S. National Supply Chain Resilience Strategy.

Future research will pursue three extensions: fuzzy K-Means clustering for boundary-ambiguous stores, transaction-level real-time anomaly detection for high-priority shrink categories, and inter-RDC lateral transfer optimization within the LP framework. The full codebase and BigQuery schema documentation will be made available to academic researchers through a restricted-access data sharing agreement upon institutional review board approval.

REFERENCES

- [1] Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. *ACM Computing Surveys*, 41(3), Article 15. <https://doi.org/10.1145/1541880.1541882>
- [2] Council on Supply Chain Resilience. (2022). First annual report of the Council on Supply Chain Resilience. Executive Office of the President, The White House.

- [3] Davenport, T. H., & Harris, J. G. (2007). *Competing on analytics: The new science of winning*. Harvard Business Press.
- [4] Executive Order 14017. (2021). *America's supply chains*. 86 Fed. Reg. 11849. Executive Office of the President.
- [5] Goldstein, M., & Uchida, S. (2016). A comparative evaluation of unsupervised anomaly detection algorithms for multivariate data. *PLOS ONE*, 11(4), e0152173.
- [6] Liu, F. T., Ting, K. M., & Zhou, Z.-H. (2008). Isolation forest. *Proceedings of the 8th IEEE International Conference on Data Mining (ICDM 2008)*, 413–422.
- [7] Müller, E., Sander, J., & Zhu, X. (2012). Clustering-driven demand segmentation for differentiated inventory policies in retail networks. *European Journal of Operational Research*, 218(3), 695–708.
- [8] National Institute of Standards and Technology. (2022). *Cybersecurity supply chain risk management practices for systems and organizations (SP 800-161 Rev. 1)*. U.S. Department of Commerce.
- [9] National Retail Federation. (2023). *The state of retail inventory distortion: A global benchmark report*. NRF Foundation.
- [10] Oppong, M. (2026). Predictive analytics for risk mitigation in retail inventory: Minimizing exposure and shrink across U.S. distribution centers. *Journal of Supply Chain Analytics & Operations Research*, 12(3).
- [11] Oppong, M., Vera, M., Onyekwuluje, P. T., & Mulenga, K. C. J. (2026). Integrating machine learning and linear programming for cost-effective supply chain optimization in high-volume retail operations. *Journal of Operations Research & Applied Analytics*, 9(2).
- [12] Pedregosa, F., et al. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- [13] Tigani, J., & Naringrekar, S. (2014). *Google BigQuery analytics*. Wiley.
- [14] U.S. Department of Commerce. (2023). *Supply chain resilience guide for critical sectors*. Bureau of Industry and Security.