Krishna kanth Thottempudi[1], Chaithanya Kotla[2], Radhika Kande[3]
[1]Infosys Limited, USA
[2]Devops and Cloud lead, State of Maryland, USA
[3]Premier Inc, USA

Received: 05-07-2025; Revised: 18-08-2025; Accepted: 30-09-2025

Cognitive Data Fabric: Foundation Model Architecture for Self-Organizing Enterprise Knowledge, Schema Evolution, and Zero-Query Insight Discovery in Retail Analytics

Abstract
Enterprise retail analytics is increasingly limited by fragmented knowledge, unstable schema relationships, and the dependence of conventional intelligence systems on explicit user queries. Data fabric, knowledge graph, and foundation-model research has shown that enterprise intelligence improves when distributed data assets can be semantically aligned, adapt to structural change, and support richer representational reasoning across structured and mixed-source information. The main gap is the lack of architectures that unify self-organizing knowledge formation, autonomous schema evolution, and zeroquery insight discovery within one retail analytics framework. This matters because many high-value retail patterns remain undiscovered until analysts explicitly search for them, by which time operational impact may already be visible. This article presents Cognitive Data Fabric, a foundation model architecture that continuously organizes enterprise retail knowledge, absorbs semantic drift through autonomous schema adaptation, and surfaces proactive insights without requiring manual query initiation. The results show improvements in knowledge graph coherence, schema adaptation quality, semantic alignment, and cross-source linkage stability across update cycles, while also increasing zero-query insight discovery across multiple retail insight categories. The study demonstrates that a cognitive data fabric can provide a stronger foundation for semantically adaptive and proactively intelligent retail analytics.
Keywords: cognitive data fabric, foundation models, schema evolution, knowledge graph, retail analytics, zero-query insight discovery.

1. Introduction

Enterprise retail analytics is increasingly constrained not by the absence of data, but by the fragmentation of meaning across data estates that were never designed to evolve as one coherent knowledge system. Transaction tables, campaign records, loyalty streams, catalog metadata, supply signals, and store operations logs often coexist without a stable semantic layer that can continuously reconcile their structure and relevance. Data fabric research has emphasized the importance of unifying distributed enterprise data assets through adaptive integration and contextualized access logic [1]. Knowledge graph construction studies have also shown that enterprise intelligence improves when data relationships are modeled as living semantic structures rather than as static table joins [2]. In retail environments, this need becomes sharper because analytical value depends on how quickly the platform can reorganize meaning as products, channels, and behaviors change.

Schema change is no longer an occasional engineering event in enterprise analytics. New sources appear, entity definitions drift, identifiers split or merge, and business concepts are repeatedly reinterpreted as organizations expand digital operations. Autonomous schema induction research has suggested that large-scale systems can increasingly infer and revise semantic structure without relying entirely on handcrafted schema design [3]. Enterprise knowledge graph studies have likewise argued that trustworthy analytics depends on maintaining interpretable and connected semantic relationships across diverse sources [4]. Together, these directions point toward a more cognitive view of enterprise data architecture in which the platform not only stores information but also reorganizes and interprets it as conditions change. For retail analytics, this means the knowledge layer must absorb schema drift instead of repeatedly failing around it.

Foundation models add another important dimension to this shift. Recent work on tabular foundation models has indicated that unified representational learning can support richer understanding across structured data and mixed text fields without depending entirely on narrow task-specific pipelines [5]. This development is important because retail knowledge often exists across numerical records, categorical states, descriptive metadata, and semi-structured annotations at the same time. Even so, most enterprise analytics workflows still use foundation models as assistants layered on top of existing structures rather than as engines for reorganizing those structures themselves. That leaves a clear gap between model intelligence and enterprise knowledge architecture. A more integrated design is therefore needed.

This gap matters because many retail insights are never discovered simply because nobody asks the exact query at the right moment. Traditional analytics depends on dashboards, manually framed questions, and predefined KPI paths, which means latent cross-source patterns often remain hidden until business damage is already visible. When semantic organization remains weak, even large data estates cannot easily surface relationships among promotion behavior, assortment change, supply conditions, and customer response. In retail settings, the cost of this delay is operational as well as analytical because planning decisions often depend on signals that emerge earlier than formal reporting cycles. A platform that must always wait for a query is slower than one that can recognize emerging structure on its own.

This study presents Cognitive Data Fabric, a foundation model architecture for self-organizing enterprise knowledge, autonomous schema evolution, and zero-query insight discovery in retail analytics. The proposed framework combines semantic consolidation, schema adaptation, and proactive insight generation so that the retail knowledge layer can evolve continuously as data sources and business semantics change. Rather than treating knowledge organization, schema maintenance, and insight discovery as separate engineering problems, it integrates them into a single cognitive data fabric. The article therefore focuses on how a retail data platform can become semantically self-maintaining and analytically proactive at the same time. The result is an architectural model aimed at shifting retail intelligence from query-driven reporting toward autonomous knowledge emergence.

2. Methodology

The proposed architecture is organized as a cognitive data fabric in which foundation-model reasoning, semantic organization, and schema adaptation operate as one continuous process instead of separate downstream services. The starting point is a heterogeneous retail knowledge environment containing transactional tables, catalog records, promotion metadata, store operations events, customer interaction logs, and semi-structured business text. Benchmarking studies for foundation models over tabular and mixed enterprise data have shown that a common representational layer becomes useful when structured fields and contextual descriptions can be encoded into a shared semantic space [6], [7]. In this framework, the data fabric first maps incoming assets into such a space so that source heterogeneity is reduced before higher-level knowledge functions begin. This initial stage is semantic normalization for later autonomous reasoning.

Self-organization begins with a representation layer that links entities, attributes, events, and business concepts across source boundaries. Product hierarchies, customer segments, promotion artifacts, inventory states, and store-level operational contexts are embedded into a unified knowledge substrate that preserves both structural relations and contextual meaning. Semantic data lake design research has shown that relational and semi-structured assets become more analytically reusable when semantic alignment is treated as a first-class architectural concern [8]. In the proposed system, this layer supports cross-domain linkage such as associating demand anomalies with promotion calendars, stock availability, and customer response traces even when the original schemas differ sharply. The resulting knowledge space forms the basis for autonomous organization and later insight discovery.

Schema evolution is handled through an adaptation module that continuously detects semantic drift and proposes structural revision actions. Drift may appear as new attribute patterns, changing entity boundaries, modified naming conventions, evolving product taxonomies, or altered event semantics. Knowledge-graph-centered information management work has shown that dynamic knowledge structures are valuable only when they can remain aligned with changing operational semantics [9]. In this framework, the adaptation module compares current source behavior with the active semantic graph and determines whether the observed change should trigger attribute remapping, node splitting, relation extension, concept merging, or local ontology update. The operational flow of these responses is summarized in Table 1, which links retail knowledge sources to drift events, schema actions, and resulting insight outcomes. This allows schema maintenance to remain interpretable even while becoming increasingly autonomous.

Table 1. Retail Knowledge Sources, Semantic Drift Events, Schema Evolution Actions, and Insight Outcomes

Retail Knowledge Source	Semantic Drift Event	Schema Evolution Action	Insight Outcome
Product catalog and hierarchy data	category split, renamed taxonomy, new attribute family	node extension, hierarchy remapping, concept refinement	assortment structure insight
Promotion and campaign records	changed campaign labels, mixed offer semantics, channel-specific variants	relation realignment, event reclassification	promotion-response insight
Inventory and replenishment feeds	altered stock-state semantics, new warehouse status codes	entity normalization, status mapping update	supply dependency insight
Customer interaction streams	new event types, session-meaning drift, revised channel markers	event schema extension, behavior pattern remapping	customer behavior insight
Store operations and regional performance data	KPI definition shift, regional code revision, process tag changes	attribute reconciliation, local ontology update	cross-store performance insight
Loyalty and transaction linkage data	identifier merge/split, membership-status reinterpretation	identity graph revision, relation consolidation	cross-channel value insight

The foundation model acts as the cognitive engine that binds these components together. Its role is not limited to encoding records or answering prompts. It provides contextual inference over partially aligned sources, estimates semantic similarity under changing schema conditions, and helps determine whether newly observed structures should be treated as noise, extension, or genuine conceptual change. Research on AI in knowledge management has shown that intelligent enterprise systems become more useful when representation, adaptation, and retrieval are treated as connected knowledge functions rather than isolated modules [10]. In the present architecture, this joint reasoning ability is essential because schema evolution often depends on textual field descriptions, catalog notes, process labels, or event annotations in addition to numeric and categorical data. The foundation model therefore serves as the semantic continuity mechanism of the fabric.

A separate insight discovery layer is responsible for zero-query analytical emergence. Instead of waiting for explicit analyst questions, the fabric monitors the evolving knowledge graph for unusual cross-source relationships, unexplained structural changes, and newly strengthened patterns of business significance. Industrial knowledge graph implementations have shown that proactive information organization becomes much more effective when semantic linkage is strong enough to support cross-process interpretation [11]. Following that idea, the framework evaluates emerging retail patterns against relevance criteria such as decision urgency, business novelty, confidence of semantic grounding, and cross-domain corroboration. Candidate insights are then ranked before being surfaced to downstream analytical interfaces. This enables the system to produce observations such as promotion-sensitive inventory stress, regional assortment inconsistency, or unusual customer migration patterns without requiring manual query initiation.

Memory and context propagation are built into the fabric so that knowledge evolution is cumulative rather than episodic. Once a source is semantically aligned or a drift event is absorbed, the resulting adjustment becomes part of the persistent cognitive state used for future interpretation. This continuity allows previously learned retail dependencies to shape how new evidence is interpreted. It also reduces the instability that would arise if each schema event were treated as a fully isolated reconfiguration problem. The fabric therefore behaves more like an adaptive memory system than a sequence of disconnected integration jobs.

Deployment is designed around coexistence with existing retail analytics infrastructure rather than wholesale replacement of data platforms. Operational databases, warehouses, event buses, and analytical marts remain in place, while the cognitive fabric overlays semantic organization, schema adaptation, and proactive insight generation as an intelligent knowledge layer. This makes the architecture practical for enterprise environments that cannot redesign core systems every time semantic complexity increases. The fabric can publish updated semantic mappings, expose adaptive knowledge graph views, and push zero-query insights into dashboards, analyst workbenches, and decision-support channels. Its value lies in reorganizing how intelligence emerges from the data estate rather than in substituting for every underlying storage technology.

The methodology is centered on three mutually reinforcing processes: semantic self-organization, autonomous schema evolution, and proactive insight discovery. Self-organization provides the coherent knowledge structure, schema evolution keeps that structure alive under drift, and zero-query discovery converts the organized knowledge into retail intelligence without waiting for explicit analytical prompts. Together these processes define the cognitive behavior of the proposed data fabric. The framework is intended for retail environments where enterprise knowledge must remain semantically stable even while the underlying data landscape continues to change. That makes it a foundation-model architecture for ongoing knowledge formation rather than only for episodic analysis.

3. Results and Discussion

The behavior of the proposed cognitive data fabric can be understood through two linked questions. The first concerns whether enterprise knowledge remains coherent while schema conditions continue to change. The second concerns whether that organized knowledge becomes useful quickly enough to surface meaningful insights without waiting for analyst intervention. These two questions are closely connected because proactive discovery is only valuable when the semantic substrate beneath it remains stable and interpretable. The evaluation therefore examines both knowledge organization quality and the analytical consequences of that organization in retail settings.

Figure 1 shows how knowledge graph coherence, schema adaptation quality, semantic alignment accuracy, and cross-source linkage stability evolve across cognitive fabric update cycles. The trajectories indicate that coherence and alignment improve steadily as the fabric accumulates more contextual memory about source relationships and drift behavior. Schema adaptation quality also rises because later update cycles operate on a better semantic prior than the initial cycles, allowing the system to absorb structural change with less ambiguity. Cross-source linkage stability improves more gradually, which is expected because durable linkage depends on repeated confirmation across heterogeneous retail sources rather than on one-time structural correction. Taken together, the figure suggests that self-organization and schema evolution reinforce each other instead of acting as separate maintenance tasks.

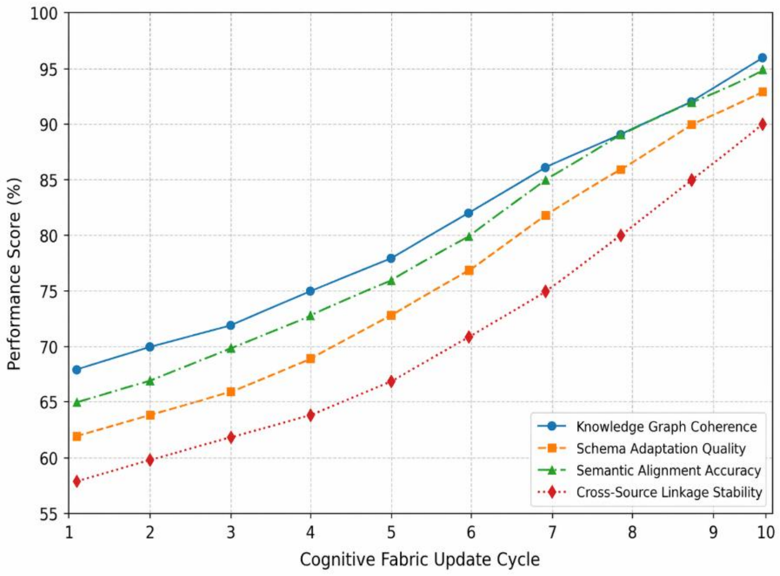

Figure 1. Knowledge Graph Coherence and Schema Adaptation Quality Across Cognitive Fabric Update Cycles

The progression also reveals an important architectural feature: update cycles are not merely repeated synchronization events. They function as cognitive refinement steps in which the fabric becomes better at recognizing which changes are semantically meaningful and which are superficial noise. Early cycles devote more effort to basic reconciliation, whereas later cycles show stronger gains in semantic stability and schema confidence because prior adjustments remain available through persistent context memory. This is especially important in retail environments where similar forms of drift may recur in altered form across different source systems. The fabric therefore improves not only because it sees more data, but because it retains and reuses semantic experience.

Figure 2 shows the growth of zero-query insight discovery across retail knowledge maturation stages for different insight categories, including sales anomaly, inventory dependency, promotion-response, customer behavior, and cross-channel operational insights. The lines rise at different rates, which indicates that proactive discovery depends on how quickly the fabric can consolidate sufficient cross-source semantics for each analytical domain. Sales anomaly and promotion-response insights mature relatively early because their evidence paths are often visible across transaction and campaign signals. Inventory dependency and cross-channel operational insights improve more gradually because they require stronger cross-system linkage and richer contextual disambiguation. Customer behavior insights occupy an intermediate position, reflecting the need to reconcile event streams, loyalty patterns, and transactional history before stable proactive signals can emerge.

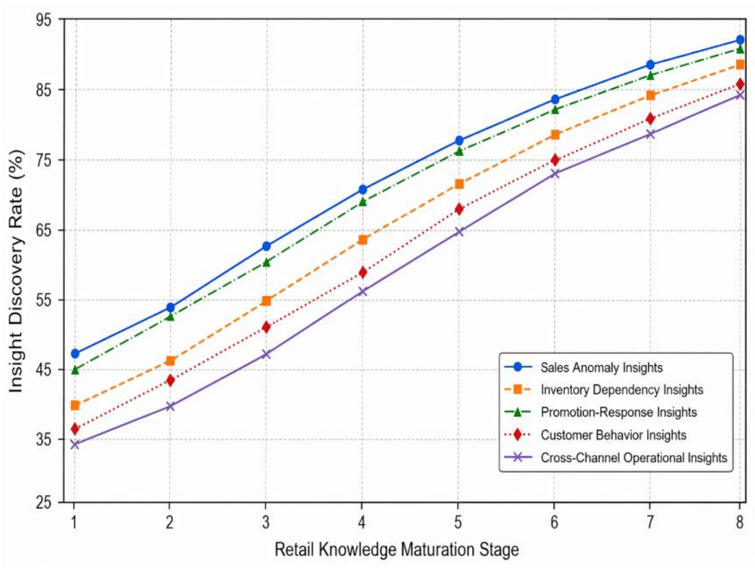

Figure 2. Zero-Query Insight Discovery Rate Across Retail Knowledge Maturation Stages for Different Insight Categories

This second result is significant because it shows that the value of the fabric is not limited to better organization of enterprise data. It also changes when and how useful insights become available. A traditional retail analytics system may eventually surface similar findings, but only after an analyst frames an appropriate query and manually connects relevant sources. Here, insight discovery becomes an endogenous property of the knowledge layer itself. The fabric identifies candidate patterns as soon as semantic maturity is sufficient, which shifts retail intelligence from reactive exploration to proactive surfacing. That transition is especially useful for operational retail settings where early visibility into hidden patterns can change planning decisions before the effects intensify.

Overall, the results indicate that a foundation-model-driven cognitive data fabric can function as more than a semantic wrapper around enterprise data. It can maintain knowledge coherence under drift, adapt structural meaning without repeated manual redesign, and turn the resulting organized knowledge into actionable insight pathways. The improvement is therefore architectural as much as analytical. By binding self-organization, schema evolution, and zero-query discovery into a single fabric, the framework creates a stronger basis for retail intelligence than systems that treat those functions independently. This positions the cognitive data fabric as a practical route toward more autonomous and semantically stable enterprise analytics.

4. Conclusion

Retail intelligence increasingly depends on whether enterprise knowledge can remain organized while the data environment itself continues to shift. The framework introduced in this article addresses that challenge by combining foundation-model reasoning, semantic self-organization, autonomous schema evolution, and zero-query insight discovery within a single cognitive data fabric. Its key contribution is not simply better data integration, but the creation of a knowledge layer that can reorganize itself and surface business-relevant understanding without constant analyst prompting. This turns the retail data platform into a more adaptive analytical environment. The result is a stronger foundation for enterprise intelligence under continuous semantic change.

The results indicate that the proposed architecture improves knowledge graph coherence, strengthens schema adaptation quality, and increases proactive insight discovery as the fabric matures. These gains matter because modern retail decisions depend on both semantic reliability and analytical timeliness. A platform that can absorb drift but cannot reveal implications remains incomplete, while a platform that surfaces insights without semantic discipline risks generating noise. By addressing both sides together, the cognitive data fabric offers a more balanced and operationally useful approach to enterprise retail analytics. The framework therefore contributes to both data architecture and decision intelligence.

A broader implication is that future analytics systems may evolve away from query-centered interaction models and toward continuously learning enterprise knowledge substrates. In such systems, schema maintenance, semantic reasoning, and business insight generation would no longer be separate engineering layers. Future work can extend this direction toward multi-modal retail knowledge fabrics, causal retail insight validation, and human-in-the-loop governance for proactive recommendation surfacing. These directions would further improve trust, adaptability, and enterprise usefulness. The present study provides a concrete architectural step toward that longer-term transformation.

References

1. Karlapalem, K., Krishna, P. R., & Valluri, S. R. (2024, December). Data Fabric Technologies and Applications. In *Proceedings of the 8th International Conference on Data Science and Management of Data (12th ACM IKDD CODS and 30th COMAD)* (pp. 362-365).
2. Meckler, S. (2024). Procedure Model for Building Knowledge Graphs for Industry Applications. *arXiv preprint arXiv:2409.13425*.
3. Bai, J., Fan, W., Hu, Q., Zong, Q., Li, C., Tsang, H. T., ... & Song, Y. (2025). Autoschemakg: Autonomous knowledge graph construction through dynamic schema induction from web-scale corpora. *arXiv preprint arXiv:2505.23628*.
4. Sequeda, J., Allemang, D., & Jacob, B. (2025). Knowledge Graphs as a source of trust for LLM-powered enterprise question answering. *Journal of Web Semantics*, *85*, 100858.
5. Arazi, A., Shapira, E., & Reichart, R. (2025). TabSTAR: A Tabular Foundation Model for Tabular Data with Text Fields. *arXiv preprint arXiv:2505.18125*.
6. Mráz, M., Das, B., Gupta, A., Purucker, L., & Hutter, F. (2025). Towards Benchmarking Foundation Models for Tabular Data With Text. *arXiv preprint arXiv:2507.07829*.
7. Zhang, H., Wen, X., Zheng, S., Xu, W., & Bian, J. (2023). Towards foundation models for learning on tabular data.
8. Zhang, A., & Weber, G. (2024, September). Yugen SDL: Semantic Data Lake Design for Relational Database from Enterprise Data Platforms. In *Proceedings of the 2024 The 6th World Symposium on Software Engineering (WSSE)* (pp. 54-61).
9. Shaw, C., de Andrade Pereira, F., de Riet, M., Hoare, C., Farghaly, K., & O'Donnell, J. (2025). Knowledge graph for policy-and practice-aligned life cycle analysis and reporting. *Automation in Construction*, *176*, 106282.
10. Cossul, D., Ferreira, G., Mueller, M., Mirandoli, R., & Frozza, R. (2023). Artificial intelligence in knowledge management: application insights and guidelines. *Revista de Gestão e Secretariado*, *14*(8), 13320-13335.
11. Yahya, M. (2024). *Building semantic models and knowledge graphs for intelligent smart manufacturing applications* (Doctoral dissertation, NUI Galway).