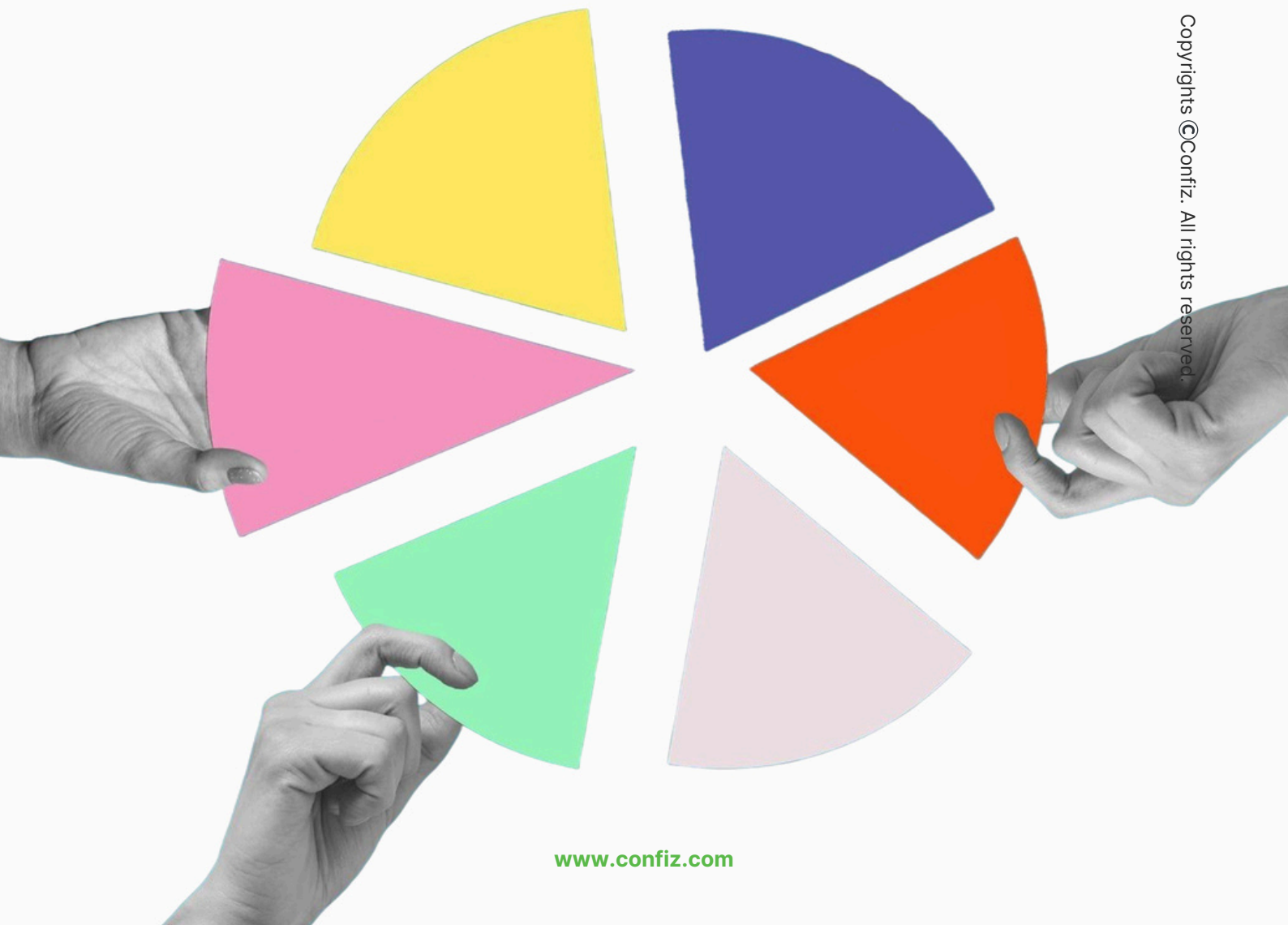


Data Sharing Consortium

Unlocking mutual value while remaining compliant, scalable, and future-ready.



About the Authors



Faizan Tariq

faizan.tariq@confiz.com

With over 12 years of experience in designing and implementing full-stack scalable solutions, with expertise in cloud modernization majorly transforming legacy applications into next-gen cloud infrastructure. Specialization in integrating Blockchain, Generative AI, Machine Learning and Agentic AI into future-ready solutions. Leading technology strategies, architecting enterprise-grade solutions, and leading engineering teams to deliver innovative, impactful projects that push the boundaries of cloud and AI technologies.



Abdul Samad

abdul.samad02@confiz.com

With over 10 years of experience delivering high-impact solutions for Fortune 500 companies across retail, supply chain, and emerging AI domains. His work includes modernizing legacy systems, building resilient cloud-native solutions, and driving measurable value through cost savings, scalability, and enhanced customer experience. His expertise spans data engineering, backend, frontend, DevOps, and AI/ML, making him a catalyst for enterprise-wide technical excellence.

Table of Contents

1. Executive Summary
2. Market Dynamics: Challenges, Opportunities, and Trends
3. Why a Data Consortium is Relevant Now?
4. Data Sharing Consortium model
5. Clean Room Visualization
6. Vector Embeddings Visualization
7. Outcomes and Success
8. Conclusion
9. About Confiz
10. Where to Find Us?
11. References

Executive Summary

This paper presents a case study of a strategic data consortium between two leading US-based brands: **FreshHarvest**, a retail grocery chain, operating over 800 grocery stores nationwide, serving millions of customers weekly and **Pebkac**, an e-commerce snacking brand and one of FreshHarvest's vendors, with 150 physical locations.

Both serving overlapping demographics, partnered to deepen customer insights beyond their verticals. Instead of sharing raw customer data, they can use privacy-preserving techniques such as vector embedding-based ML enrichment via Delta Sharing and secure clean rooms for non-PII and aggregated data. This approach enables richer, actionable insights for improved personalization, marketing optimization, and demand forecasting.

Market Dynamics: Challenges, Opportunities, and Trends.

Repetitive Challenges

Isolated first-party data, limiting strategic insights.

Blind spots in cross-retail behavior.

Under-optimized marketing due to siloed visibility.

Teams spend up to 40% of their time just finding and aligning data across partners and platforms using online tools.

Potential Opportunity

Privacy preserving collaboration to uncover cross-retail behavioral patterns, lifecycle signals and segment enrichment.

Unlock new strategic insights.

Industry Landscape

Decline of third-party cookie-based insights.

Shift towards first-party data enrichment via secure collaboration.

Greater expectations for relevant, cross-category customer experiences.

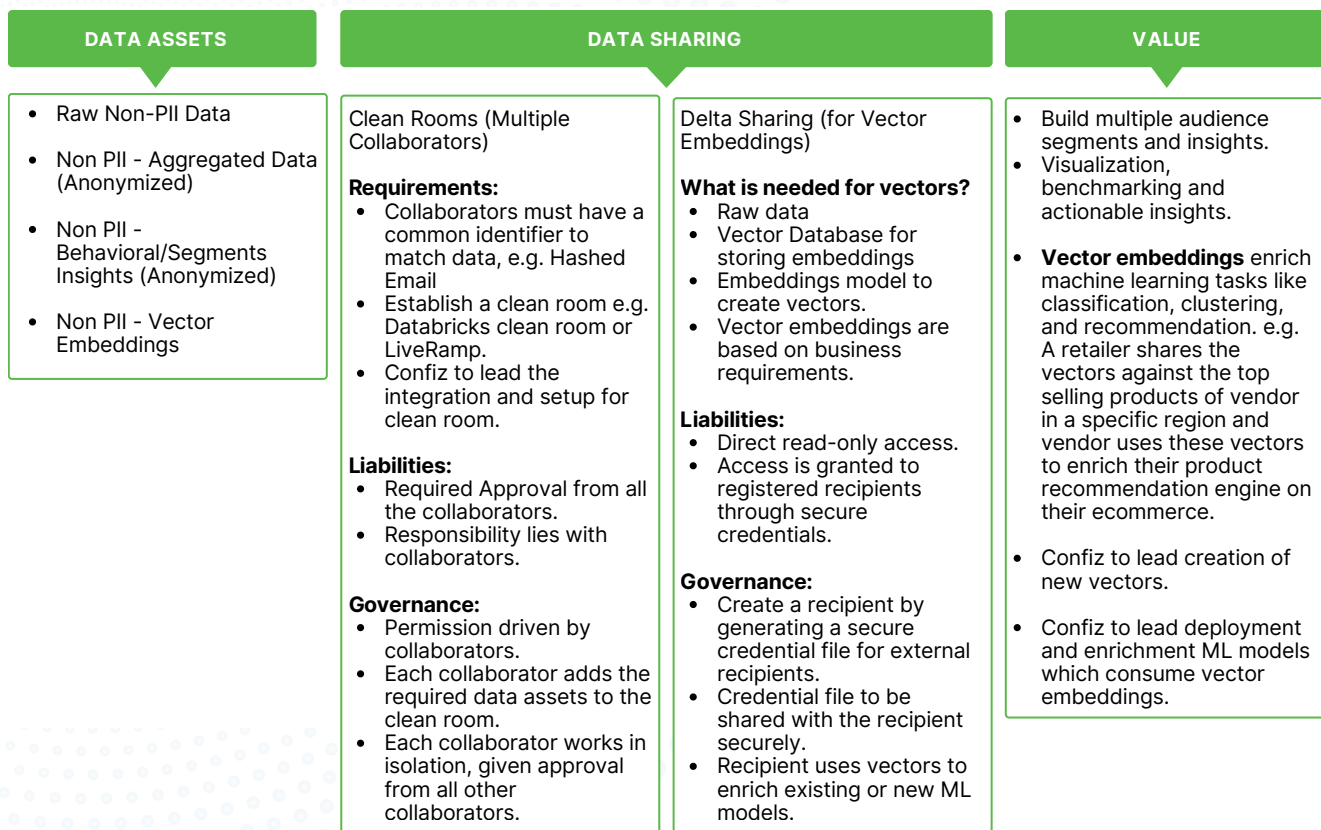
Why a Data Consortium is Relevant Now?

- Siloed data restricts the ability to unlock mutual value.
- Rising costs for customer and data acquisition.
- Data privacy regulations (GDPR, CCPA) forcing compliant approaches.
- Maturity of technologies like Delta Sharing, vector embeddings, and clean rooms.

Data Sharing Consortium Model

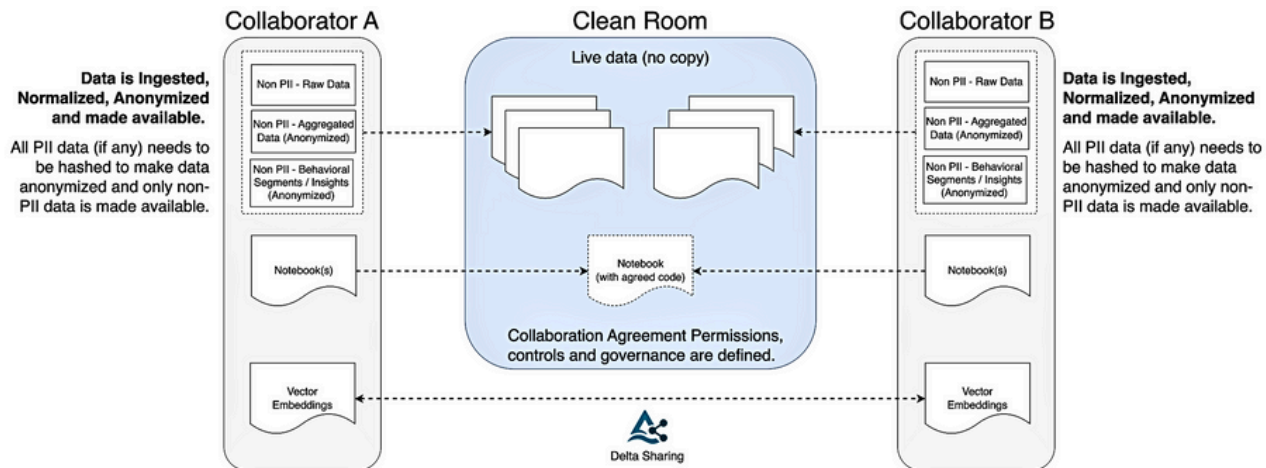
A **data consortium** is a governed partnership where two or more parties share datasets or derived outputs to generate collective value while maintaining compliance, privacy, and data sovereignty.

- Data sharing consortium model has following components:
 - Data Assets
 - Embeddings Model (To generate vector embeddings)
 - Vector Embeddings
 - Delta Sharing
 - Clean Rooms
- For Data Sharing we have 2 verticals:
 - Vector Embeddings are exchanged securely using Delta Sharing,
 - While use of Clean Rooms is limited to the handling of raw and aggregated data processed under strict controls.
- Each collaborator enhances its internal ML systems with partner-generated vectors, e.g. cross-domain purchase intent or product similarity.



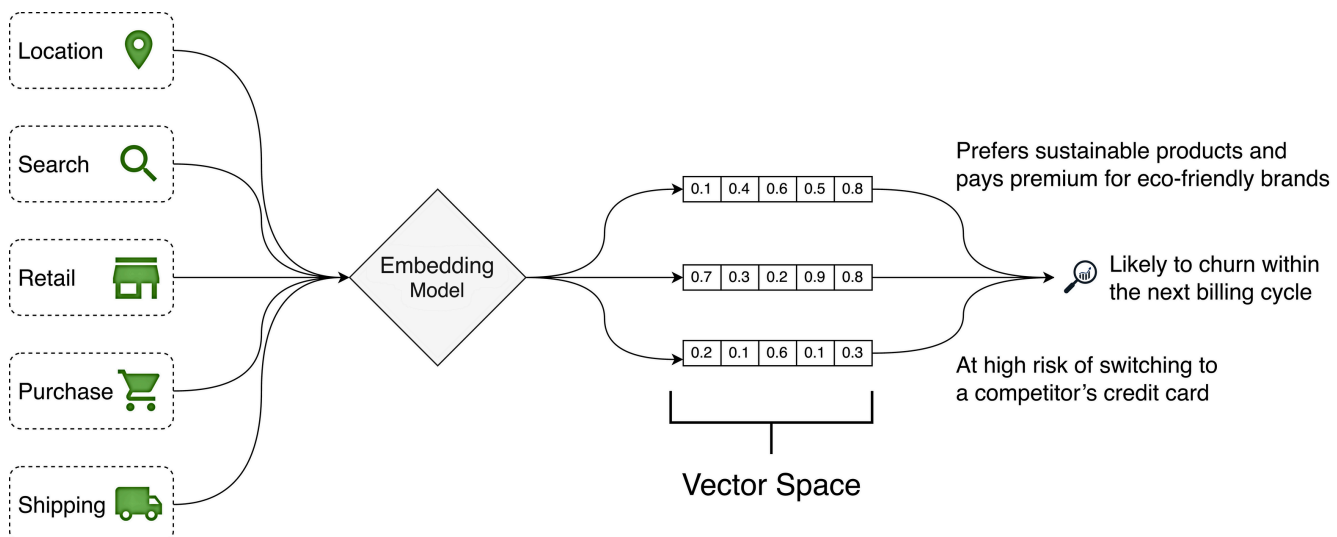
Clean Room Visualization

The recommended visualization of data consortium model is as follows, focusing on Clean Rooms for sharing Non-PII raw and aggregated data and Delta Sharing for sharing vector embeddings for ML enrichment.



Vector Embeddings Visualization

The visualization of how vector embeddings are being built from raw data and being consumed as signals in a vector space for ML enrichment is as follows:



Outcomes

Build Embeddings

Business Fit	Strong alignment with tailored business goals
Cost	Higher upfront and ongoing costs
Time to Market	Slower due to development and validation
Internal Capabilities	Requires strong in-house data/ML expertise
Customization & Control	Full customization and control over embedding logic
Security and Compliance	Aligned with internal compliance and security standards
Integration Needs	Requires custom integration pipelines
Scalability & Flexibility	Highly flexible and scalable if well-architected
Long-Term Strategic Value	High value as a core asset and competitive edge

Success

- Pebkac shares the aggregated data against the top trending newly introduced products within a clean room and FreshHarvest uses it to drive product innovation or category creation on their side.
- FreshHarvest shares the aggregated data against the top trending snacking products within a clean room and Pebkac uses it to drive product innovation or category creation on their side.
- FreshHarvest shares product level vector embeddings (e.g. cosine similarity across catalogs) against the top selling in-store products of Pebkac in a specific region and Pebkac uses these vectors to enrich their product recommendation engine on their e-commerce.
- Pebkac shares customer intent vectors and FreshHarvest consumes it to model high-propensity buyers for special events and holiday bundles.

CONCLUSION

Conclusion

This data consortium focuses how modern collaboration via vector embeddings with secure data exchange platforms like Delta Sharing enables two retailers to share intelligence, not data. By avoiding direct PII data exchange and leveraging controlled model enrichment techniques, FreshHarvest and Pebkac unlock mutual value while remaining compliant, scalable, and future-ready.

NCLUSIO

Ready to Explore What's Possible?

Book a complimentary 30-minute consultation with a Confiz Azure Architect. Let's discuss how scalable AI infrastructure can help address your business challenges and support your growth.

[Book a Consultation](#)

Copyrights © Confiz. All rights reserved.

About Confiz

A global technology solutions and consulting company that empowers forward-thinking enterprises, including Fortune 100 companies, to strengthen and transform their digital core. Our expertise spans Data & Enterprise AI, Microsoft Business Applications, Cloud Services, Mobility, Security, and Bespoke Software Development.

With 19+ years of experience and a global workforce, Confiz is recognized for its commitment to innovation, industry expertise, and a global delivery mindset. Operating across North America, EMEA, ANZ, APAC, and now LATAM, and backed by 8 Development Centers worldwide, we drive business growth through technology, helping our clients build a better future today and tomorrow.

Where to Find Us?

Our Worldwide Offices



www.confiz.com



marketing@confiz.com



+1 425 365 0857

References

Clean Room Architecture

- <https://learn.microsoft.com/en-us/azure/databricks/clean-rooms/>

Delta Sharing

- <https://learn.microsoft.com/en-us/azure/databricks/delta-sharing/>
- <https://delta.io/sharing/>