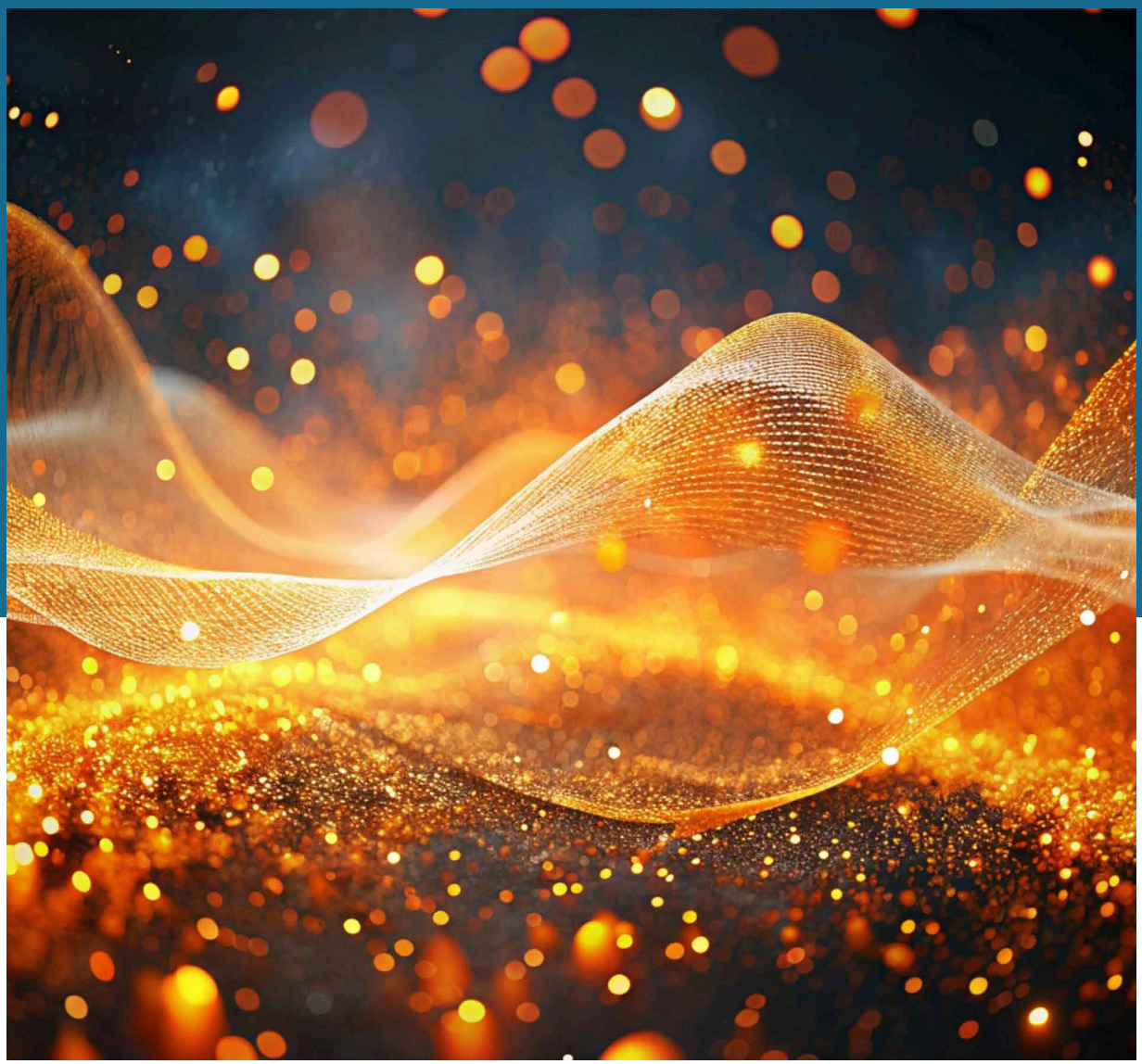
marlabs
driving digital agility

# Symphony of Innovation: The tech that personalized music for over 67 million users

## How Marlabs Helped a Leading Broadcasting Company Unlock Data Value with Databricks



www.marlabs.com

# CONTENTS



The data landscape is constantly transforming with ever-increasing volumes and complexities. In this dynamic environment, organizations must prioritize achieving a state of "data readiness." This involves not just collecting vast amounts of information, but ensuring its quality, accessibility, and usability. By becoming data-ready, organizations empower themselves to leverage customer data for more informed decision-making.

One key technology paving the way for data readiness is the data lake. Unlike traditional data warehouses, which are structured for specific purposes, data lakes act as central repositories for all an organization's raw data, regardless of format or source. This holistic approach empowers businesses to unlock new insights and make data-driven decisions that can transform their customer relationships.

As the largest audio entertainment platform, our client reaches over 100 million diverse listeners across North America. However, with siloed data across various legacy systems, their ability to innovate experiences and understand customers was constrained. By partnering with Marlabs on a modernization initiative cantered around implementing Databricks' Lakehouse capabilities on AWS, unified data analytics was achieved to better compete.

# Core Challenges: Fragmented Data, Delayed Insights

Over the last decade, our clients's data infrastructure could not keep pace with exponential growth in users, content types, and platforms. With crucial media insights trapped across various databases, third-party tools, and batch pipelines, the following were some of the primary challenges that emerged:

- Data operated and managed by different teams with no common goals.
- Batch processing caused unacceptable latency in accessing subscriber and engagement data.
- Inability to analyze streaming content quality, user interaction, and consumption in real-time.
- Separate systems hindered complete cross-channel customer profiling.
- Mergers & Acquisitions were taking too long to merge due to technology limitations.
- Unable to build a unified customer profile to come up with better loyalties.

Upwards of 85% of listener behaviour and digital feedback data was locked across:

- On-premise Teradata data warehouses with analytics latency nearing 4 weeks
- Multiple AWS stores like Redshift and SingleStore are managed for different use cases.
- Data control and access were scattered within separate systems without a single control pane.
- Master Data built on Informatica MDM was not servicing enterprise data reference.
- DynamoDB NoSQL tables for certain mobile app interactions
- Kafka streams for certain audio metadata analytics
- The use of cross-cloud technology platforms, including GCP, AWS, and Azure, posed a significant challenge.

This severe fragmentation manifested in quantifiable business impact including:

Churn prediction models only refreshed monthly enabling limited loss prevention.

- 4+ week delays in launching personalized packages to drive uptake.
- Missing holistic view of listeners leading to suboptimal content recommendations
- Ineffective targeting diminishing potential ad revenue.
- Increased cost of managing multiple platforms (GCP, AWS, and Azure)

With data access delays estimated to be causing ~$30M+ in yearly revenue leakage alone, it necessitated consolidating siloed information assets into a unified platform - enabling real-time gathering of insights from different systems and applying cutting-edge analytics at scale.

# The Databricks Lakehouse Solution

To meet the demands of 100M+ dynamic users, Marlabs designed a future-ready "lakehouse" solution on AWS consolidating all data into Delta Lake on Databricks leveraging cloud scale, open architecture, and breakthrough machine learning capabilities unified on one platform.

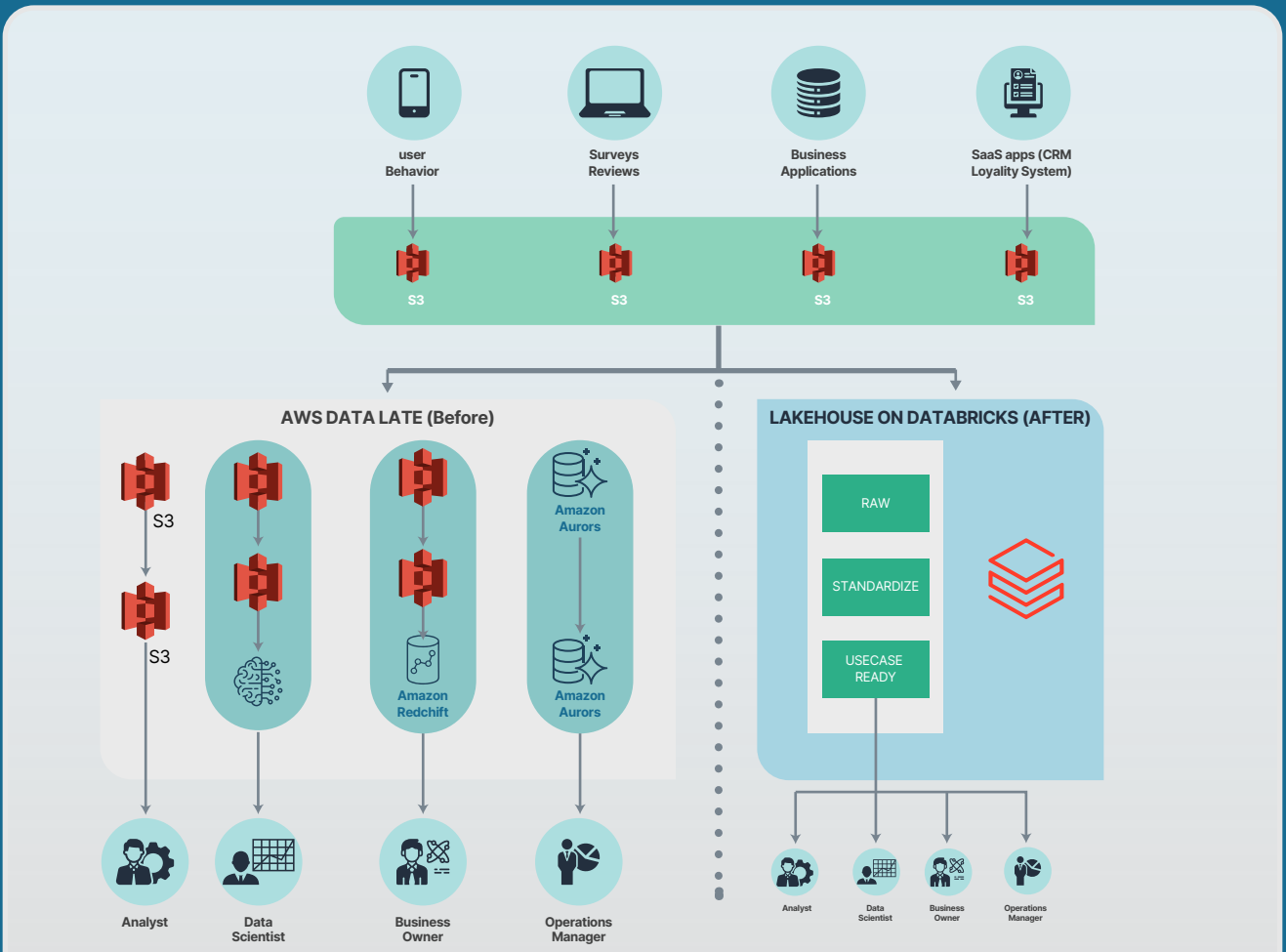
After evaluating alternatives, Databricks uniquely matched several key criteria:

- **Unified Platform -** To support various use cases and avoid the need for setup, management, and support, based on ingestion, processing, and consumption requirements of different LOBs and personas.
- **Hybrid Analytics Foundation -** As the pioneer behind groundbreaking open-source big data engines like Spark, Databricks provided a unified SQL, ML, and streaming analytics platform to ingest data at a massive scale from varied on-prem, cloud sources enabling deprecated system retirement.
- **Real-time Insights -** Spark-powered streaming and graph processing enabled new near real-time dashboards and algorithms providing granularity into listener behaviour changes facilitating real-time content adjustments - finding optimal points to introduce advertisements based on receptive data.
- **Impactful Intelligence -** Unified data and compute-enabled cross-data correlated insights on unified customer profiles, ML automation, and reproducible data science drove various monetization-oriented use cases before unfeasible - like geo-targeted ads, predictive churn models, and personalized engagements.
- **Future-proofing -** Open, cloud-based architecture with existing integrations mitigated vendor lock-in and provided extensibility to keep pace with evolving analytics needs - such as adding IoT metrics from new hardware-based products to the decision-making ecosystem.

The high-level depiction of changes in the architecture has led to enhanced data processing and reusability, regardless of the data consumers and the use case delivery approach. Every change or improvement in the factors relevant to decision-making is advanced to deliver broader audience adoption. The below diagram illustrates how the overall consumption of the data has changed the processing speed and steps, irrespective of the data producer's speed, format, technology, and complexity.

Data Lake is an excellent method for preparing data for targeted use cases and consumers. However, it requires the implementation of plugging components to ensure the availability and compatibility of the data. Despite addressing these challenges, Data Lake still encounters limitations, such as lengthier data processing cycles and data rewrites and overwrites. Furthermore, managing real-time data and batch workloads requires careful consideration of their velocity and technological advancements to meet the required quality parameters.

Databricks, on the other hand, manages streaming and batch data in a unified data processing approach which makes data to be ingested and processed only when needed, consume as much as needed but avoid managing different data copies and versions due to ACID features, and with the support to the extendable technology, features made it powerful to be analyzed directly by consuming the consumption layer data.




Key highlights on how modernizing the data lake house helped avoid creating multiple data copies

With Databricks Delta enabling more efficient and accessible analytics data online for the first time, the unified Customer360 foundation ingesting data streams across digital experiences enabled game-changing value across focus areas:

**10x**

Faster actionable analytics preventing subscriber churn.

**15%**

Lift in conversion leveraging engagement optimization.

**8%**

Increase in ad revenue from ultra-targeted campaigns.

**7%**

Reduction in OpEx savings from legacy system retirement.

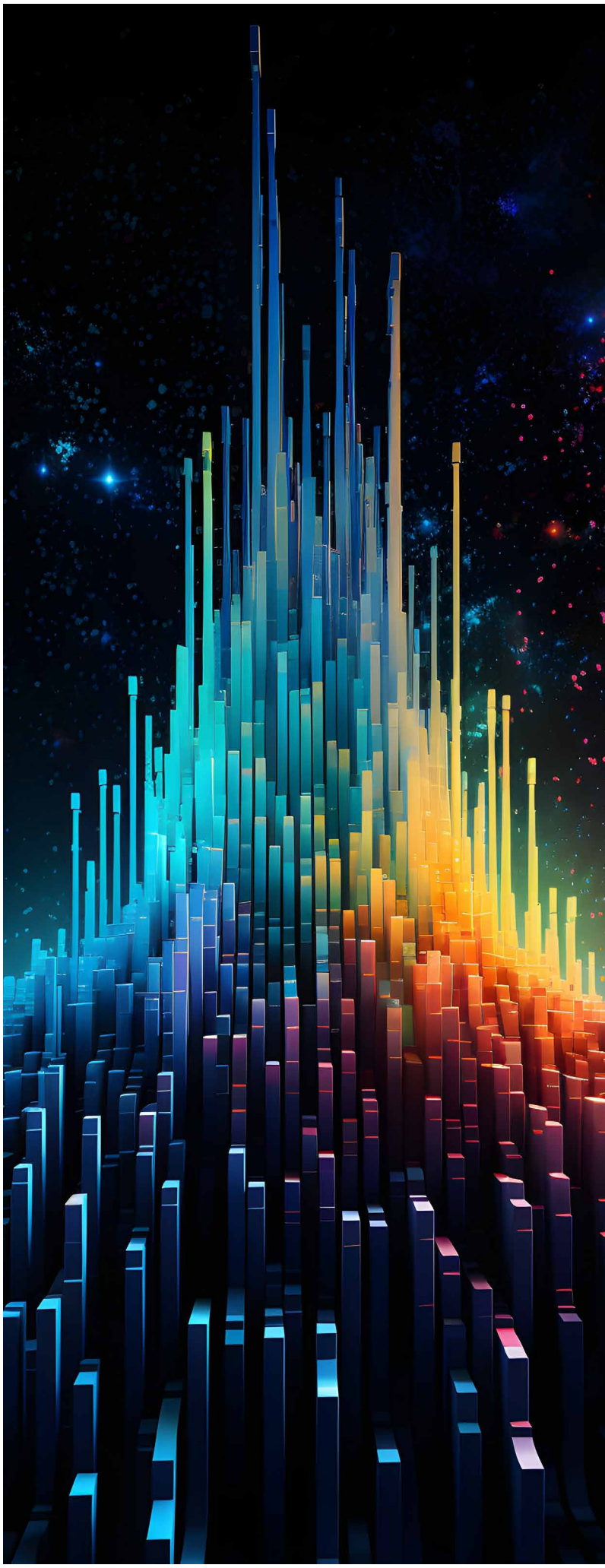
**18%**

Increase in revenue growth projection owing to the introduction of 4 new business processes.

# Architecture Powering Lakehouse on Databricks

The technical architecture implemented consists of key components powering a performant, reliable, and scalable modern data stack

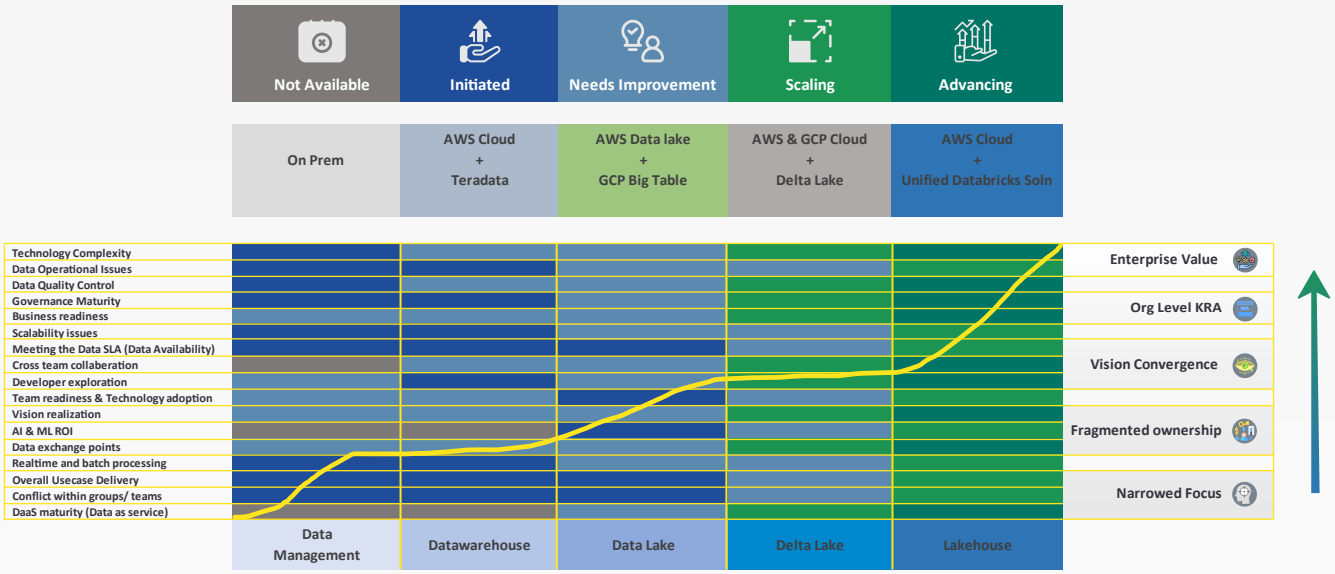
- AWS Infrastructure - Leveraging EC2, EMR, and S3 resources, the platform can expand storage and compute automatically to meet growing data demands.
- Kafka & Kinesis - Ingest real-time event streams from apps, subscribers, and devices using AWS streaming and queuing technologies.
- Delta Lake - Databricks Delta on S3 derived analytics datasets enriched using Spark and SQL analytics at scale.
- Governance – Access control, Lineage, and dictionary were built for enterprise-wide adoption which helped better manage the data entities and change management.
- Monte Carlo – Enterprise Data quality control room was designed to ensure the data movement within and across predefined metrics.
- ML Capabilities - MLflow, Hyperparameter Tuning manage the machine learning lifecycle - from experiment tracking to model deployment.
- Data Marketplace – Our client is exploring budding data subscriptions to external data providers without having to build integrations.



# Data platform modernization journey

The modernization journey was iterative with advancements and improving on the limitations identified with constant monitoring by Business and IT, the transformation maturity levels were carefully chartered with a focus on reusability and scaling without complete redesign as a strategy.

The maturity graph below shows the limitations that drove innovation through the years.



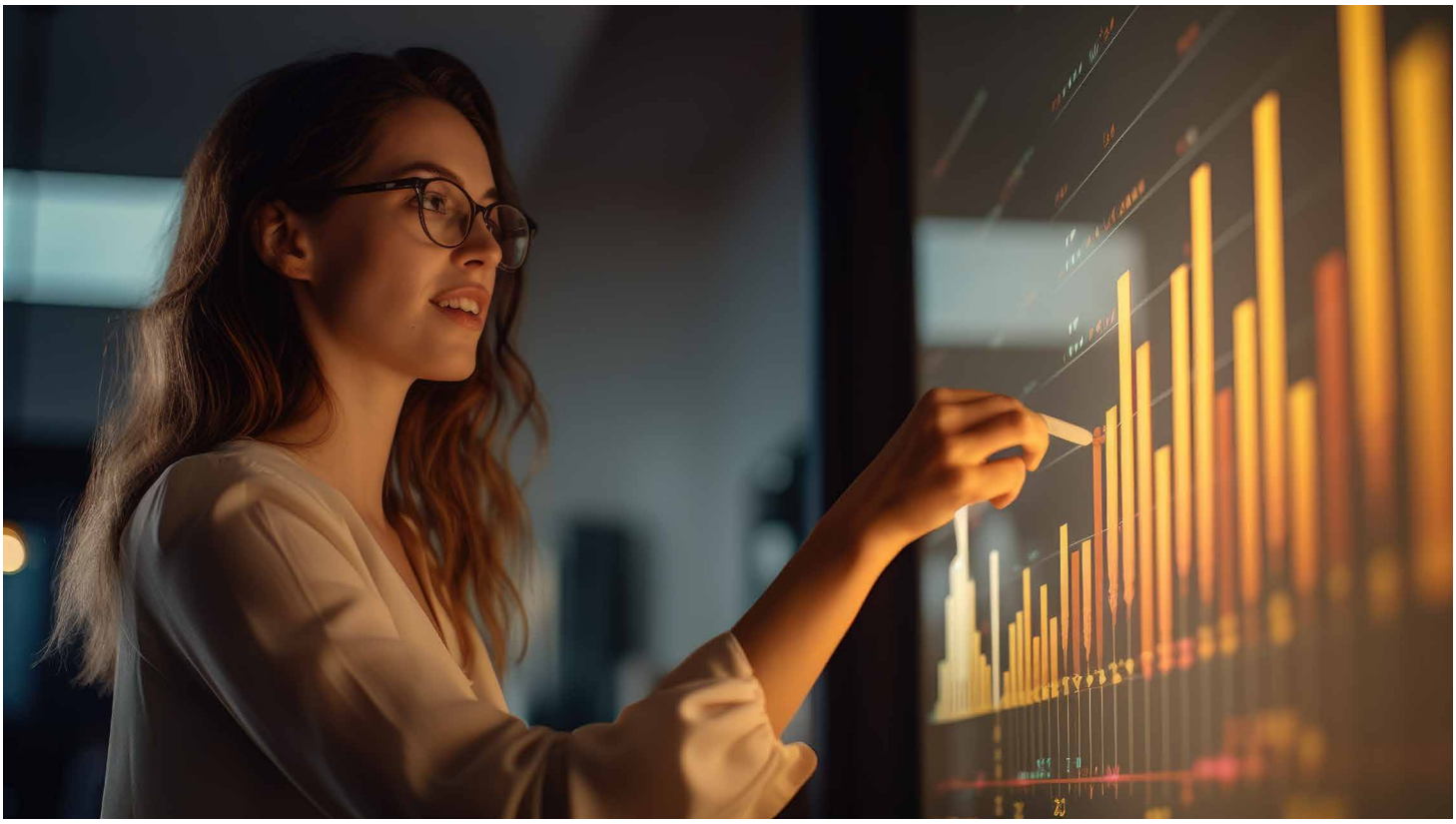

## Business Enhancement avenues were identified through the modernization as below:

While immense value has been realized since launching the Lakehouse initiative transforming our client's analytics maturity, ample potential exists to further entrench Databricks across the data-to-value lifecycle.

Marlabs has charted expansion opportunities around:

- Greater Data Consolidation - Ingesting remaining siloed social media, sales, and CRM system streams for complete unified profiling.
- Intelligence Automation - Optimizing recommendation and personalization through machine learning.
- Data Science Productivity - Providing curated features and improving sandbox tools boost model experimentation.
- Democratization - Real-time analytics exposures and dashboards integrated across portals enable self-service capabilities.
- Enterprise Data Governance - Implementing end-to-end lineage tracking and access controls as data assets scale.

As personalizing every listener's journey becomes critical for sustaining competitive differentiation and loyalty - while combating listener churn risks - enhancing Databricks' capabilities remains a pivotal growth driver Marlabs continues spearheading.



# The Path Forward

As our client expands products and doubles down on personalization, Databricks provides the launch point enabling deeper embedded intelligence - from individualized interactions to pioneering ahead of shifting consumer demand. Unified data and analytics establish information as an innovation accelerant for continued modernization.

With streaming content consumption displacing traditional media, seizing the advantage by leveraging data cross-enterprise is instrumental to unlocking the next horizon of possibilities. Marlabs sees abundant potential to further optimize intelligence configuring the platform as the customer-centric brain powering everything from real-time contextual recommendations to predicting emerging listener preferences through ML.

By tapping into data as the fuel energizing experiences across engagement channels, the vision is to equip our client with customer intelligence as a differentiating asset continually elevating competitive agility, relevance, and hyper-personalization. With a sturdy Lakehouse foundation now cemented, Marlabs is poised to chart the next leg of the data modernization journey hand-in-hand with our client.

**Mahesh Rayappa**
Senior Director - Data Engineering & Insights Practice, Marlabs

# About Marlabs

Marlabs designs and develops digital solutions that help our clients improve their digital outcomes. We deliver new business value through custom application development, advanced software engineering, digital-first strategy & advisory services, digital labs for rapid solution incubation and prototyping, and agile engineering to build and scale digital solutions. Our offerings help leading companies around the world make operations sleeker, keep customers closer, transform data into decisions, de-risk cyberspace, boost legacy system performance, and seize novel opportunities and new digital revenue streams.

Marlabs is headquartered in New Jersey, with offices in the US, Germany, Canada, Brazil and India. Its 2500+ global workforce includes highly experienced technology, platform, and industry specialists from the world's leading technical universities.

Marlabs Inc.(Global Headquarters) One Corporate Place South, 3rd Floor, Piscataway NJ - 08854-6116, Tel: +1 (732) 694 1000 Fax: +1 (732) 465 0100, Email: contact@marlabs.com.