EMERGERE TECHNOLOGIES
TECHNOLOGY ONWARDS AND UPWARDS

WHITEPAPER

THE JOURNEY FROM LEGACY DATA MODELS TO MODERN PROACTIVE AND PREDICTIVE DATA MODELS

Transforming Data Strategies for Agile, Insight-Driven Organizations

THE JOURNEY FROM LEGACY DATA MODELS TO MODERN PROACTIVE AND PREDICTIVE DATA MODELS

Transforming Data Strategies for Agile, Insight-Driven Organizations

CONTEXT

By 2030, the total amount of digital data created world-wide would quantify to 612 Zettabytes! Yes, data has emerged as the most valuable asset for any organization today. To take complete advantage of data as a resource for decision making and to future proof businesses in the current dynamic environment, data modernization is crucial. Digital native companies have an upper hand in encountering data modernization while for the traditional organizations, it becomes a key concern.

Smart industry-relevant solutions will not only smoothen the process of data modernization for digital immigrants/legacy companies but will also ensure optimal utilization of appropriate tools and resources. This white paper presents an approach to a seamless data model transformation, its challenges, steps and its real-time impact on the organization's performance.

> *"Administering Yesterday's solutions for the control and predictability of today's business processes don't seem to work well in this new era of cloud services"*

INTRODUCTION

With the advent of big data, organizations have started to realize the need for legacy data modernization. Day-to-day business processes generate mounds of data within which hides the knowledge that could influence how decisions are made, and forecasting is done. Crunching out the data for that knowledge is a skill rather than a task. Use of modern data architecture and data engineering tools will help businesses ace this skill and seamlessly scale their data volumes following their growth and progress.

Looking at the statistical data of research carried out in 2023 among 1420 IT decision makers at companies and organizations in eight different sectors (Financial services/insurance, Manufacturing & Logistics, Retail, Hospitality & Travel, Government/Public sector, Media & Entertainment, Energy sector, Healthcare, Pharma, Bio-tech/Life sciences), the major drivers of legacy data modernization were, improved security & efficiency, followed by increased scalability & flexibility, ensured cost reductions, and better integration with modern technologies[1].

"The greatest danger in times of turbulence is not the turbulence; it is to act with yesterday's logic."

– Peter Drucker, Father of modern management.

WHAT DRIVES DATA MODERNIZATION?

Legacy data models, developed before the advent of modern technologies, often support outdated applications and technologies, making them ill-suited for current business requirements. These systems are typically characterized by **rigid structures, siloed data, limited scalability, lack of automation, and outdated security protocols**, leading to increasing maintenance costs and minimal analytical capabilities. Businesses relying on such systems face challenges in data integration, cloud storage, reporting, and metadata management, leaving them vulnerable to cyberattacks and operational inefficiencies.

In contrast, **modern data architectures** can ingest data from thousands of sources simultaneously, validating, cleansing, standardizing, and enriching it to deliver insightful and actionable responses. They enable organizations to achieve immediate **cost efficiencies, improve forecasting accuracy, and enhance personalization through advanced analytics.** With these transformative capabilities, data modernization has become an essential priority for business leaders, offering a competitive edge in today's dynamic market landscape.

63% of organizations report that technical debt from outdated legacy systems moderately or severely impacts their operations. (June 2024, Censuswide research report) [2]

U.S. government agencies spend **$337 million annually** to maintain legacy systems, some of which are up to 51 years old and are no longer supported by vendors (2023 GAO report) [3]

95% of top global banking executives identify outdated legacy systems and core banking platforms as major barriers to optimizing data and achieving customer-centric growth (2022 World Retail Banking Report by Capgemini and Efma) [4]

Only **22% of healthcare** providers use modern operating systems due to high upgrade costs, compatibility issues, and lack of internal knowledge (2021 Kaspersky Healthcare Report) [5]

The Snowflake Revolution: Boosting Efficiency for the largest privately-owned Building Supply Material Distributor in the US

Predictive analytics demystified with **Emergere's** Data engineering solutions

How do I integrate data from disparate data sources/ERPs?

How do I get quick updates on inventory, supply chain logistics, client orders and sales?

How do I forecast demands on a regular basis to optimize inventory levels?

How do I streamline my operations to ensure timely responses to clients?

How do I enhance the security features to protect client and operational data?

CLIENT PROFILE

Our client is the **largest privately-owned, full-line distributor of specialty building materials in the United States.** Established two decades ago, the company operates across 47 active divisions and 450 locations nationwide with a turnover of $10 Billion. Its core business encompasses the procurement and distribution of building materials, the assembly of specialized materials, and providing comprehensive assistance with material installation.

The client envisioned a system that could predict demand spikes with precision and optimize their supply chain dynamically. This necessitated the integration of data from over 50 distinct sources and consolidation into a unified data warehouse, enabling **seamless enterprise-wide data analytics** on Quotes, Orders, Purchase orders, Stock receipts, Sales, Logistics, Inventory & General Ledger, fostering informed decision-making.

CHALLENGES

The client's legacy data platform faced several critical challenges that hindered the company's operational efficiency and innovation. **It could integrate only 25 data sources**, significantly **limiting its ability to scale** and accommodate the growing need for new integrations. The platform **lacked the agility to support advanced analytics initiatives**, creating roadblocks for data-driven decision-making.

Additionally, the complexity involved in developing and deploying new integrations further slowed down processes, while the **inability to handle semi-structured data** restricted the organization from leveraging diverse data formats essential for modern analytics and business operations.

PROPOSED SOLUTION

Considering the pain points noticed on assessing the client's legacy system, **Snowflake** was recommended as the best solution for the client. The platform offers the ability to deliver new integrations and insights to the business with both speed and scalability. It enables the **standardization of platforms** across the organization, reducing the need for extensive IT management and allowing IT teams to focus on more strategic, customer-facing initiatives.

Snowflake's **support for high-concurrency workloads** facilitates the parallel loading of thousands of tables from source systems, while its dynamic scalability ensures optimal performance during query execution. Furthermore, Snowflake's high-performing capabilities for **handling semi-structured data** allows for the seamless translation of JSON data into relational tables, further enhancing data integration and analytics capabilities.
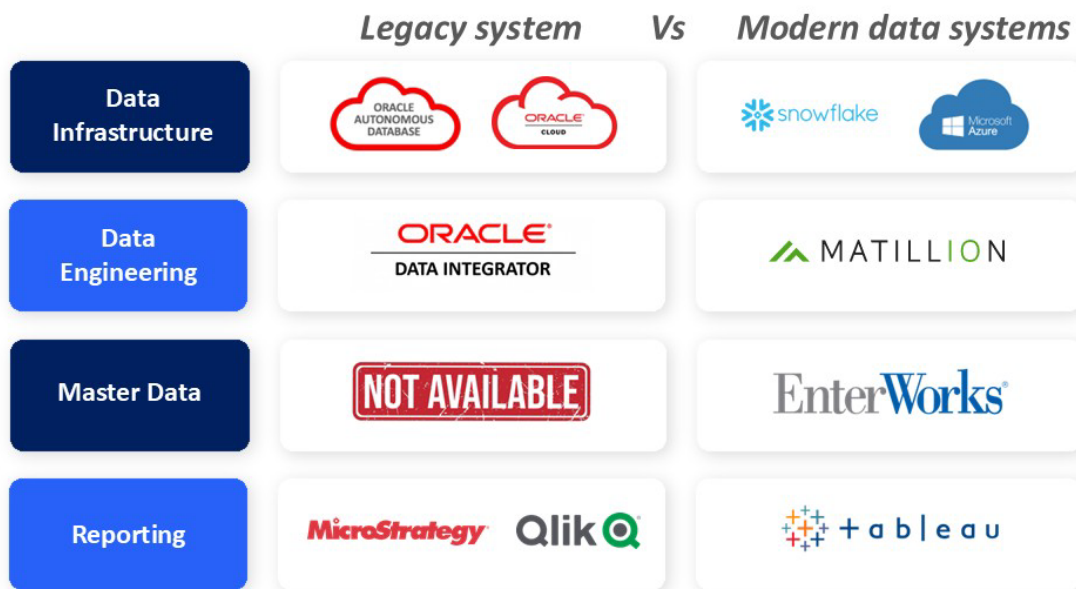
EDM Architecture
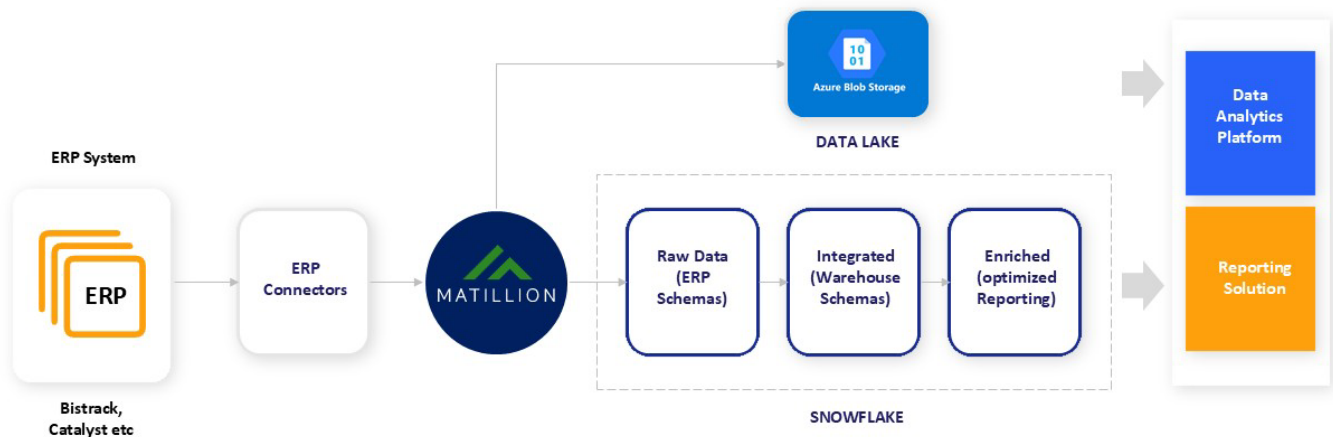
THE TRANSITION JOURNEY

Emergere Technologies played a pivotal role in the data modernization process by designing and implementing a comprehensive strategy to transform the client's data landscape by deriving advantages from Modern Data and Analytics Platforms like **Matillion, Snowflake, Tableau and Enterworks.**

EDM: Technology Stack

	Legacy system	Vs	Modern data systems
Data Infrastructure	ORACLE AUTONOMOUS DATABASE / ORACLE CLOUD		snowflake / Microsoft Azure
Data Engineering	ORACLE DATA INTEGRATOR		MATILLION
Master Data	NOT AVAILABLE		EnterWorks
Reporting	MicroStrategy / Qlik		tableau

DATA INGESTION

The transition journey began with a thorough assessment of the existing data sources, integrations, and data flows to uncover gaps, inefficiencies, and opportunities for improvement, revealing a complex landscape of data related to products, suppliers, customers, and locations scattered across diverse ERPs such as **Bistrack, Catalyst, Agility, Spruce, Eagle, and Ponderosa**, each with its own unique formats and structures.

Leveraging our expertise, we successfully integrated data from multiple source systems into a centralized repository, enabling robust and efficient daily ETL processes. With a refined data model and streamlined workflows, the pipeline seamlessly manages raw data ingestion, consolidation, and analytics, ensuring accurate and timely insights. The system is supported by a sophisticated infrastructure, incorporating hundreds of jobs and transformations to handle complex data workflows, while scheduled processes ensure smooth daily

DATA INTEGRATION

The data integration layer serves as the foundation for curated datasets, supporting all downstream systems. Central to the design and communication of the enterprise data warehouse (EDW) architecture is the **enterprise data warehouse bus matrix**, which acts as a pivotal tool for defining the EDW bus architecture.
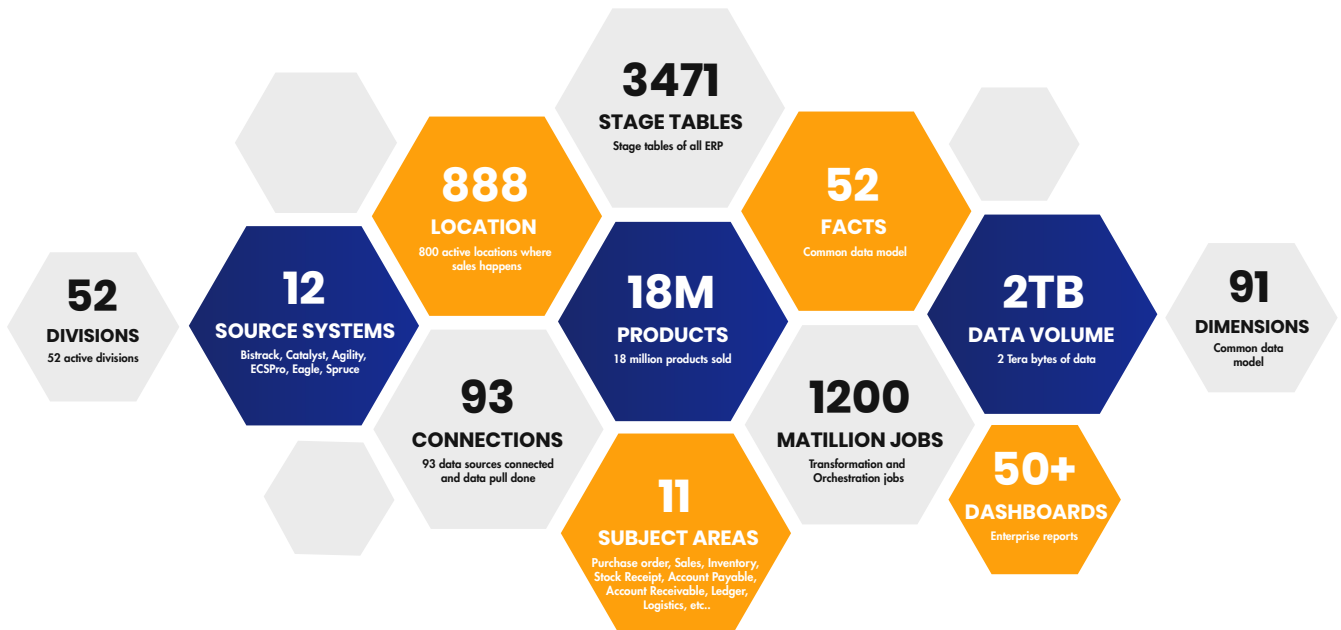
Reporting aggregation tables are constructed atop the data integration layer tables, with this layer playing a critical role in determining the business processes and the granularity of data. This phase also includes finalizing the facts and **dimensions** that define the architecture. Dimensions provide the contextual elements—such as the "who, what, where, when, why, and how"—that frame a business process event. These dimension tables hold the descriptive attributes leveraged by BI applications for filtering and grouping facts. In contrast, **facts** represent the measurable outcomes of business process events, which are predominantly numeric.

In our architecture, we primarily employ **Slowly Changing Dimension (SCD) Type 2** for historical tracking, ensuring comprehensive visibility into data changes over time. Both SCD Type 1 and Type 2 dimensions are constructed using a **template-driven approach**, promoting reusability, maintainability, and reduced development timelines. The **core transformation logic** is encapsulated within the data integration layer, providing a robust backbone for streamlined and efficient data processing.

DAILY LOAD STATISTICS

3471
STAGE TABLES
Stage tables of all ERP

888
LOCATION
800 active locations where sales happens

52
FACTS
Common data model

52
DIVISIONS
52 active divisions

12
SOURCE SYSTEMS
Bistrack, Catalyst, Agility, ECSPro, Eagle, Spruce

18M
PRODUCTS
18 million products sold

2TB
DATA VOLUME
2 Tera bytes of data

91
DIMENSIONS
Common data model

93
CONNECTIONS
93 data sources connected and data pull done

1200
MATILLION JOBS
Transformation and Orchestration jobs

50+
DASHBOARDS
Enterprise reports

11
SUBJECT AREAS
Purchase order, Sales, Inventory, Stock Receipt, Account Payable, Account Receivable, Ledger, Logistics, etc..

DATA MODELLING

In our data modeling efforts, we implemented a **star schema design** to structure data efficiently, with dimension tables supporting filtering and grouping and fact tables handling summarization. We streamlined ETL processes by adding control columns to enable efficient data tracking, updates, and soft deletes. UTC-based **datetime handling** was introduced to standardize global reporting.

Decimal handling was standardized to balance accuracy and storage, while fact aggregations were implemented to optimize reporting and reduce dependency on base tables. Derived measures and KPIs were integrated into the data warehouse schema during ETL, centralizing business logic and ensuring consistency. These steps enhanced scalability, performance, and alignment with business needs.

DATAMODEL RECOMMENDATIONS

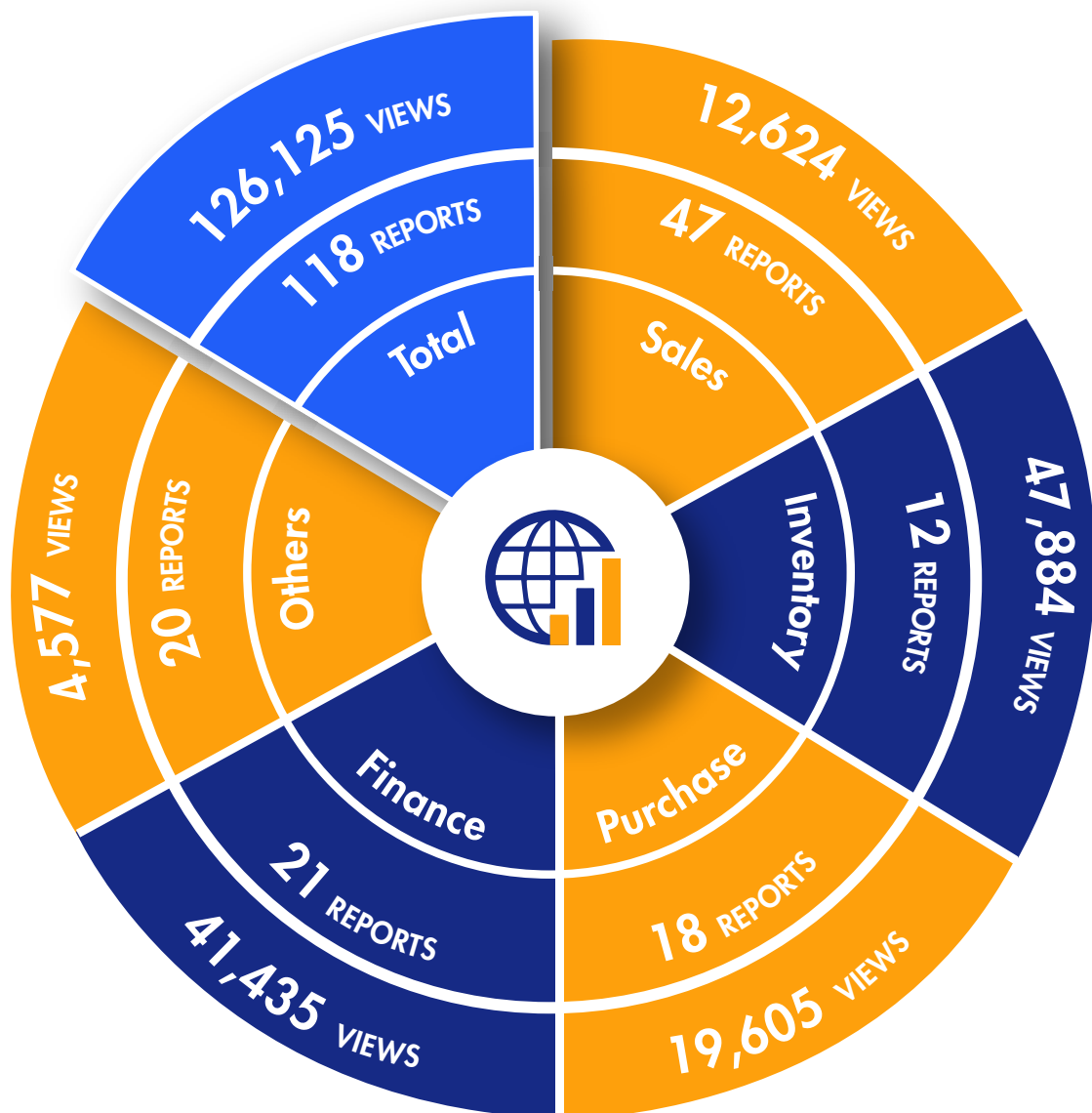| STAR SCHEMA	FACT AGGREGATIONS	KPIS
A well-designed model, then, is one that provides tables for filtering and grouping,and tables for summarizing. This design fits well with star schema principles: • Dimension tables support filtering and grouping • Fact tables support summarization	Define fact aggregations for reporting needs and also for performance optimization. Currently, there are no fact aggregations and all the data is being reported from the base tables.	Derived measures/Calculated KPIs, if possible, should be added as part of the DW schema. Avoid calculation logic in the reporting layer. Keep the business logic completely in ETL.

DATA VISUALIZATION

In our client's business landscape, data modernization dramatically enhanced data visualization by centralizing and standardizing data from multiple sources, such as inventory systems, logistics platforms, and sales records. Modernized systems enable real-time dashboards that visually represent critical metrics like stock levels, delivery timelines, and regional sales performance.

By replacing static, outdated reports with interactive visual tools, stakeholders can instantly monitor key performance indicators (KPIs), streamline inventory management, and optimize routes for timely deliveries. This level of enhanced visibility not only fosters proactive decision-making, but also reduces operational inefficiencies, and strengthens customer satisfaction.

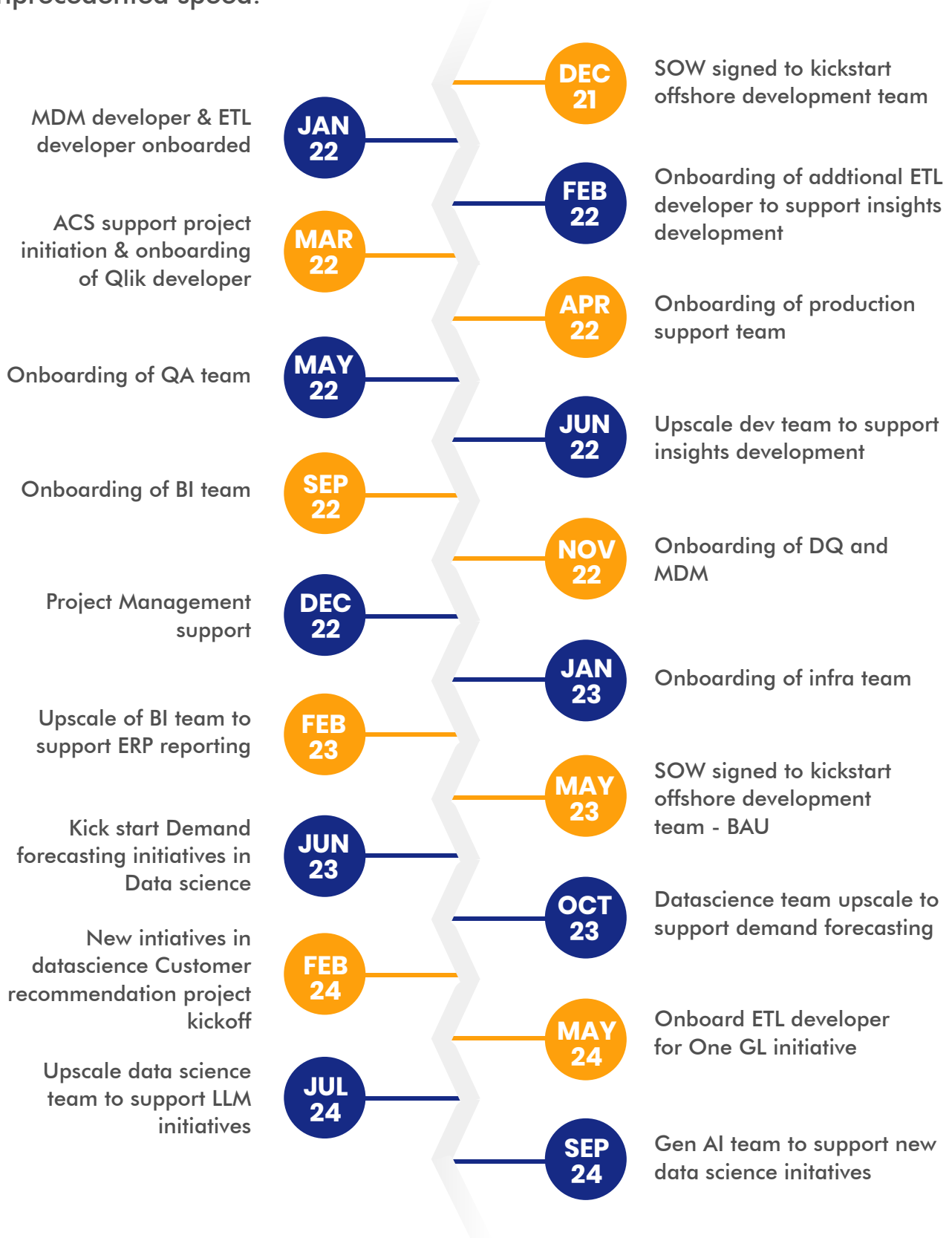## REPORTS AND USAGE OVER THE PAST 12 MONTHS

126,125 VIEWS — 118 REPORTS — Total

12,624 VIEWS — 47 REPORTS — Sales

47,884 VIEWS — 12 REPORTS — Inventory

4,577 VIEWS — 20 REPORTS — Others

41,435 VIEWS — 21 REPORTS — Finance

19,605 VIEWS — 18 REPORTS — Purchase

OUR JOURNEY WITH THE CLIENT: A TIMELINE OF SUCCESS

Emergere Technologies has strategically evolved to effectively support our client's data modernization initiatives. As their needs evolved, we expanded our capabilities, scaling infrastructure and talent to ensure efficiency, and sustained innovation, ultimately empowering the client to make data-driven decisions at unprecedented speed.

DEC 21 — SOW signed to kickstart offshore development team

JAN 22 — MDM developer & ETL developer onboarded

FEB 22 — Onboarding of addtional ETL developer to support insights development

MAR 22 — ACS support project initiation & onboarding of Qlik developer

APR 22 — Onboarding of production support team

MAY 22 — Onboarding of QA team

JUN 22 — Upscale dev team to support insights development

SEP 22 — Onboarding of BI team

NOV 22 — Onboarding of DQ and MDM

DEC 22 — Project Management support

JAN 23 — Onboarding of infra team

FEB 23 — Upscale of BI team to support ERP reporting

MAY 23 — SOW signed to kickstart offshore development team - BAU

JUN 23 — Kick start Demand forecasting initiatives in Data science

OCT 23 — Datascience team upscale to support demand forecasting

FEB 24 — New intiatives in datascience Customer recommendation project kickoff

MAY 24 — Onboard ETL developer for One GL initiative

JUL 24 — Upscale data science team to support LLM initiatives

SEP 24 — Gen AI team to support new data science initatives

CONCLUSION

Data keeps erupting like volcanoes across organizations from different platforms like computer terminals, smartphones, websites, social media networks, e-commerce sites and IoT devices. But what percent of this data is available in a format usable for analytics is still a challenge in many industries. This gap underscores an urgent need for a comprehensive, efficient, scalable, and cost effective data modernization strategy to empower organizations to harness the full value of their data.

In many industries, legacy systems are still deeply embedded in both public and private sectors, often acting as barriers to innovation. By transitioning to modern data models, these organizations can unlock immense potential. Modern data frameworks enable informed, data-driven decisions, provide the agility to scale operations seamlessly, and ensure future-readiness in a rapidly evolving technological landscape.

> *"Data modernization is about creating an organization that truly values data and uses it effectively at the entry-level. That's the real challenge—and the real opportunity—of data modernization in 2024 and beyond."*
>
> *– Ed Thompson, CTO and Co-founder of Matillion.[6]"*

These models are not just enablers but are essential for leveraging the transformative power of big data, advanced analytics, and emerging technologies like artificial intelligence and machine learning. **Embracing data modernization is no longer an option it's a strategic imperative for organizations seeking sustained growth and competitive advantage.**

However, this transformation journey is often hindered by critical challenges such as **resistance to change, data silos, cybersecurity concerns, and skill gaps.** With a proven track record in delivering innovative solutions, Emergere Technologies provides the necessary guidance and technical capabilities to break down silos, address security risks, and bridge the skill gap.

WHAT SETS US APART?

Many organizations have outgrown their legacy data system, and we understand that transitioning from a legacy system to a modern solution may seem daunting and disruptive. With our expertise, tools, and proven experience, we're here to help identify the challenges holding your business back and can guide you through retiring outdated systems and embarking on your data modernization journey prioritizing on all three essential elements of your organization-**people** (training, communication, and leadership), **process** (automation, reengineering, and agility) and **technology** (cloud, microservices, and data migration).

Architect your data infrastructure with us!

📞 **+1 469 401 7117**

✉️ **hello@emergertech.com**

🌐 **www.emergertech.com**

📍 **US (Corporate Office)**

8105 Rasor Blvd, Suite 60, Plano, Texas 75024

REFERENCES

01. https://www.statista.com/statistics/1420607/main-drivers-modernizing-legacy-applications-data/

02. https://censuswide.com/report/

03. https://www.gao.gov/assets/gao-23-106821-highlights.pdf

04. https://ibsintelligence.com/ibsi-news/world-retail-banking-report-2022-incumbent-banks-must-optimise-data-for-customer-centric-growth/

05. https://usa.kaspersky.com/about/press-releases/kaspersky-finds-73-of-healthcare-providers-use-medical-equipment-with-a-legacy-os/

06. https://www.matillion.com/blog/data-modernization-in-2024-what-you-need-to-know

07. https://www.matillion.com/blog/matillion-and-idg-survey-data-growth-is-real-and-3-other-key-findings

08. https://www.linkedin.com/pulse/legacy-systems-brief-study-how-outdated-impede-modernize-sharma-vkq3e/