Data Stack Modernization: Balancing Costs and Innovation

By Rahul Bagga

Information has become the lifeblood of modern businesses. But simply collecting vast amounts of data isn't enough. Organizations must adopt an appropriate solution to harness their power and gain a competitive edge. A well-designed data stack – a collection of technologies that ingest, store, manage, analyze, and visualize data – is essential to unlocking data's potential.

<div style="text-align: center; color: red;">

Building and maintaining the right data stack for your organization can be a complex endeavor.

</div>

Businesses must navigate a delicate balancing act: optimizing costs while fostering the innovation needed to gain a competitive edge. Across the globe, organizations are racing to modernize their data stacks and unlock transformative insights that promise competitive superiority.

The Evolution from Legacy to Modern Data Stacks

Think of a data stack as a comprehensive toolbox that empowers you to manage the entire lifecycle of your data. It includes the technologies and tools used to:

- **Collect:** Capture information from various sources, like customer interactions, website visits, and social media.
- **Process:** Clean, organize, and transform raw data into a usable format.
- **Store:** Securely house your data in a central repository for easy access and analysis.

- **Analyze:** Extract valuable insights and uncover hidden patterns to inform strategic decision-making.

Traditional data stacks, often reliant on on-premise infrastructure and siloed systems, struggle with scalability and agility. This struggle hinders businesses in their quest to extract valuable insights from the ever-expanding volume and complexity of data.

Modern data stacks (MDS) address these challenges using the power of the cloud. Cloud-based platforms like Snowflake and Databricks offer a robust and flexible data management environment with key advantages:

- **Unmatched Scalability**: Effortlessly scale storage and processing power to meet fluctuating data demands.
- **Enhanced Agility**: Rapidly deploy new data tools that adapt to evolving business needs.
- **Reduced Costs**: Eliminate expensive hardware maintenance and free up IT staff for strategic initiatives.
- **Faster Decision Making:** Empowers faster, data-driven decision-making through real-time insights.
- **Streamline Operations:** Automation of data pipelines increases efficiency.
- **Unlock advanced analytics:** Machine learning and AI capabilities unlock advanced insights.

The shift to the cloud is underway and accelerating. By 2025, 80% of organizations will move to multi-cloud and hybrid cloud strategies, and over 95% of new digital workloads will be deployed on cloud-native platforms [1].

The rush to adopt trendy data stacks often overshadows the need to tailor solutions that meet specific data handling, processing, and analytical needs.

This rapid shift to cloud-based data stacks isn't without its challenges. Many organizations struggle to find the right balance between cost-effectiveness

and innovation - a key reason is a misalignment between data infrastructure capabilities and overall business goals. Simply migrating everything to the cloud might not be the most efficient solution, and neglecting the current data landscape can lead to wasted resources and missed opportunities.

The rapidly expanding universe of data tools and technologies complicates the decision-making process for organizations striving to modernize their data ecosystems, with over 1,416 companies offering a wide range of data storage, processing, and analytics solutions [2]. Selecting the right toolset becomes a matter of technical suitability as well as a significant financial decision that can profoundly impact an organization's operational efficiency

According to a Couchbase survey taken in 2022, a typical enterprise spends more than $33 million a year on cloud services. 35% of that, or $8.75 million, is wasted on inflexible pricing plans, management tools that don't give users the level of control they need, and data not being stored where it needs to be [3].

Assess Unique Organizational Needs

The easily underestimated step of assessing the organization's unique requirements underscores the reality that a modern data stack (MDS) is not a universal solution. It answers a unique problem and must be carefully developed to accommodate the organization's specific data handling, processing, and analytical needs.

The difference between Snowflake and Databricks[4] is a good example of a nuanced data stack choice. While both platforms offer robust capabilities, they are designed with different primary use cases, as shown in the table below.

Platform/Feature	Databricks	Snowflake
Storage	Better for raw, unstructured data.	Better for reliability and ease of use for structured data
Use Case	Better for ML, AI, Data Science and Data Engineering. Collaborative notebooks in Python/Scala/R a big plus	Easier for analysts in business intelligence and companies looking to migrate existing data warehouse system
Price	Cheaper at high compute volumes. Not as predictable on cost.	Efficient at scaling down unused resources. More consistent, predictable costs.
Scalability	Infinitely scalable. Effective at high volume workloads.	Separate storage and compute makes for seamless concurrent queries.
Security	GDPR-compliant, role-based access control, encrypted at rest and and in motion	GDPR-compliant, role-based access control, encrypted at rest and in motion

As we can infer from the table above:

- **Databricks** is best for organizations focused on machine learning, AI, and collaborative data science with unstructured data.
- **Snowflake** is ideal for enterprises prioritizing structured data warehousing, efficient scaling of resources, and consistent budgeting with strong business intelligence capabilities.

Both products have several similarities. Without fully understanding your unique requirements, it's easy to pick the wrong one, resulting in wasted time and money. It's vital to understand the specific attributes of the platform and

make sure they align with your organization's needs. The following are a few key questions to consider.

1) **Does my organization need to migrate ALL of our systems?**

   Champion the principle of leveraging the right tool for the right job: a comprehensive data strategy can provide value through a combination of legacy systems and cutting-edge solutions.

   Consider the case of a Fortune 500 company with a mature data ecosystem centered around Snowflake as a cornerstone of its modern data stack. Despite the potential benefits of integrating additional modern tools, the organization found that its existing expertise in traditional ETL processes adequately addressed its current needs without necessitating a full pivot to newer MDS components [5].

   The shift from a one-size-fits-all approach to a more tailored, problem-solving orientation empowers organizations to create a data strategy that is both forward-looking and deeply anchored in the practical realities of their operational context.

2) **Does Your Organization Really Need Real-time Data Analytics?**

   Another pivotal consideration is the adoption of real-time data analytics. Real-time data processing can offer significant advantages in scenarios such as fraud detection, personalized marketing campaigns, and high-frequency trading. However, it comes with increased costs and complexity.

   The high-stakes world of financial trading involves transactions where seconds can mean the difference between significant profit or substantial loss. Firms in this industry rely on the most current data to make informed decisions on buying or selling assets. In this scenario, real-time data analytics is not just beneficial; it's critical.

   In contrast, the needs of a retail chain analyzing its sales operations (SO) are less time-constrained. Decisions such as restocking, layout

adjustments, or promotional strategies do not require second-by-second data updates. Batch processing can provide sufficient insights for making informed decisions without the additional expense of real-time analytics.

Ably, a platform specializing in real-time data services highlights the financial implications of maintaining real-time data infrastructure. According to a survey cited by Ably, companies that build and manage their own real-time systems often face significantly higher costs—typically $100,000 to $200,000 more annually—compared to traditional batch processing systems. This marks a potential increase in costs of about 2 to 3 times for maintaining real-time infrastructure over batch processing [6].

Modern data warehousing companies implement billing models based on process frequency within the system. Inserting data every minute incurs significantly higher costs for the customer than consolidating these operations into a single batch insertion every hour.

3) **What tangible ROI will the transition deliver?**

According to a survey by **451 Research** [7], 68% of organizations struggle to integrate different data tools effectively, leading to data silos that inhibit cohesive data analysis. Data silos occur when the same information is stored and managed separately in different departments/units within a business. This creates a lack of communication and collaboration between the departments and leads to major inefficiencies and duplication of efforts.

The key to success lies in creating a strategic, comprehensive approach to data stack modernization. Organizations must go beyond mere tool selection to create an action plan that considers how these technologies will integrate to deliver business value aligned with long-term business objectives. By doing so, they can avoid the pitfalls of data silos.

4) What hidden costs might be encountered during the transition?

The journey to a modern data stack ecosystem is often accompanied by excitement over the potential insights and efficiencies that new technologies promise. However, do not overlook the hidden costs involved in upgrading, which can significantly impact the overall ROI:

- **Integration Fees and Training Expenses:** Combining various tools into a seamless and integrated data environment often incurs significant expenses. This includes expenses related to connecting different systems and the necessary time to train teams on navigating the intricate landscape.

- **Ongoing Maintenance:** Maintaining a data stack requires attention. Updates and adjustments are often needed to ensure alignment with business needs and keep pace with technological advancements.

- **Non-Optimized Data Pipelines:** Casually employing resource-intensive processes, such as SQL cross-joins, without carefully considering their impact can significantly drive up operational costs and create bottlenecks. This inflates expenses and undermines the data stack's fundamental objective of facilitating quick, data-driven decisions. On the other hand, an over-reliance on nested views for data access simplification can lead to performance degradation and increased costs in consumption-based models.

An organization's strategic action plan must include mitigating risks, including those associated with these and other hidden costs.

A Template for Data Stack Modernization

Use this baseline to build a roadmap to underpin your organization's modernization plans.

1. **Conduct a Comprehensive Data Strategy Review**. Thoroughly assess your current data capabilities, needs, and challenges. This review should identify gaps in your existing infrastructure and opportunities for growth. Begin with a detailed analysis of costs by product, job, dashboard, or process:

Category	Operation	Annual Cost Breakdown (USD)	Current Data Capabilities	Business Needs	Performance Gaps	Opportunities for Growth
Executive Dashboard	P&L Dashboard	Licensing $ Personnel $ Maintenance $	Compliant with Current Specs	Enhanced Processing for Increased Workload	Query Response > 5s (Target: <1s)	Upgrade Processing Hardware
Process	Current Data Process	Licensing $ Personnel $ Maintenance $	Requires Data Quality Improvement	Ingesting Diverse Data Sources	No Real-Time Analysis Capabilities	Implement Real-Time Data Processing
Product	Inventory Mgmt System	Licensing $ Personnel $ Maintenance $	High Efficiency	Scalability for Expanding Inventory	System Not Cloud-Based	Migrate to Cloud-Based Solutions
Job	Marketing Analytics	Licensing $ Personnel $ Maintenance $	Improvable Analytics Platform	Predictive Analysis for Consumer Behavior	Lacks Predictive Modeling Capabilities	Develop Predictive Modeling Features

A baseline template for determining data strategy

2. **Evaluate and Select Tools Based on Use-Case Analysis**. Choose tools that align with your specific business requirements. Consider the primary use cases for your data, whether it's analytics, reporting, AI, or machine learning, and select tools that best support these needs.

3. **Budget for the Total Cost of Ownership.** Factor both the upfront costs of software and hidden costs related to integration, training, maintenance, and future scalability. This comprehensive budget approach will help ensure financial preparedness.

4. **Implement a Robust Training Program**. Equip your team with the necessary skills to use new technologies effectively. An investment in training helps minimize the learning curve and maximize the utility of your data stack.

5. **Optimize and Monitor Data Processes.** Regularly review and refine your data processes for efficiency and effectiveness. Optimization will reduce costly inefficiencies and ensure your data stack remains aligned with business objectives.

6. **Plan for Scalability and Future Growth**. Ensure your data infrastructure is designed to grow with your business. Scalability is critical to accommodate increased data volumes and evolving analysis needs.

7. **Engage Expert Consultants When Necessary**. Consider partnering with data strategy and technology experts. Their experience and insight can help navigate challenges and accelerate success.

The Balanced Path to Modernization

It is important to recognize that data infrastructure efficiency hinges on two crucial factors: a solution meeting business requirements and cost requirements. You want to choose the most cost-effective option that delivers the necessary performance and functionality. The costs associated with your data stack demand meticulous oversight.

The rise in popularity of cloud technology and subsequent decrease in computing and storage costs has led to a vast array of cloud-based platforms hitting the market. Cloud-based platforms like Snowflake and Databricks promise unparalleled scalability, flexibility, and revolutionary data capabilities, offering a competitive advantage that's hard to resist. However, a company must avoid succumbing to the latest technology trends; instead, follow a disciplined, strategic approach that includes thorough planning, careful tool selection, and proactive management of hidden costs.

The path to a modern data stack is fraught with challenges, but with the right strategy and guidance, it's a journey that can propel your business to new heights of efficiency and growth.

About us: [mXa](#), on the 20+ year foundation of [Method360](#), was founded to intentionally serve fast-growth companies and the unique challenges they face. We understand that inorganic and organic growth provokes change, ambiguity, and uncertainty that can deeply burden the organizations involved. By seeking to understand the human element in M&A and fast growth environments, mXa embraces a unique, contrarian approach in advising clients that seeks to realize maximum value for them in alignment with business objectives.

Interested in learning more about our capabilities or discussing your M&A or Data story? We're here to help.

References:

1. Gartner- Gartner Says Cloud Will Be the Centerpiece of New Digital Experiences
2. Matt Turck Machine Learning AI and Data (MAD) 2023 Landscape
3. Couchbase Enterprises Over-spending by More Than $8.75 Million on Cloud Services
4. Databricks vs Snowflake: A Complete 2024 Comparison
5. HighTouch You don't need the Modern Data Stack to get sh*t done
6. Ably Building realtime infrastructure: Costs and challenges
7. 451 Research