

WHITE PAPER

Transforming Data Practices into Tangible Business Outcomes

witboost



Chapters

Introduction **1**

Technology Lock-In Challenges **2**

Data Best Practices **3**
Build with Data Interoperability in Mind 4
Apply Platform Engineering Principles 4
Always Develop Evolutionary Data Architectures 4
Ensure Data Documentation Is Part of Business Value 5
Shift Data Governance Left 5
Automate the Data Lifecycle 6
Prioritize Data Security 6
Infuse Product Thinking into Data Engineering 6

Defining the Needed Platform Features **7**
Customization and Composition Capabilities 7
Data Shaping with Templates, Technology Adapters, and Policies 8
- Templates 8
- Policies 8

Build vs. Buy **9**
Time 9
Cost 9

How Witboost Meets Data Platform Needs for Business Value **10**
Accelerated Integration Using Witboost Open-Source Starter Kits 10

1. Introduction

In an era where data is the lifeblood of enterprise innovation and decision-making, the landscape of data engineering and discovery is rapidly evolving. Today's enterprises are navigating a complex web of data-related challenges, from generating insights to ensuring efficient data management. The surge in volume, variety, velocity, and complexity of data and the unstoppable growth of unstructured data require the means for converting it into business value.

While self-service data strategies are a growing approach to data management, 49% of recent respondents to a **Forrester survey** say it's challenging to maintain a user-friendly environment and boost data trust and quality. Navigating these complexities requires investment in data management solutions that transform data practices into tangible business value.

A data management platform that is truly technology-agnostic can create a customizable, scalable, and agile data tech stack. Such a future-proof platform model would avoid obsolescence through the ability to reshape itself and respond to emerging business needs and constraints. The key is to find a solution, enabling:

- Data Project Streamlining
- Business-Driven Data Discovery
- Computational Policies for non-bypassable Data Governance
- Broad automation capabilities
- Integration of various platforms
- Customized technology within any organization based on their rules and policies
- A self-service approach that ensures data trust and quality

Having this flexibility enables new technology integration and interoperability; the development of new policies; and the introduction of constraints, processes, or practices. These advantages pave the way for significant cost reductions around subscription models, migration strategies, maintenance, and management.

The purpose of this white paper is to lay out some of the challenges, best practices, and major decisions surrounding data management.

2. Technology Lock-In Challenges

There are countless tools available in the current data landscape that promise to unlock the potential of enterprise data. Most of these technologies often lead to fragmented data practices, hindering the ability to derive meaningful insights. Employing any data tech stack, technology, or tool that is not technology-agnostic presents countless challenges, which all begin with technology lock-in.

Most data platform vendor solutions today lack openness and technology-agnostic interoperability; instead, they force organizations to fall victim to vendor lock-in. However, the path of building a platform also has inherent lock-in challenges.

The DIY route is often chosen based on the assumption of greater control—particularly when it comes to cost. This 'sunken cost fallacy' is soon revealed as data growth and needs must give way to greater openness and interoperability, which requires ongoing investment to create even a limited illusion of interoperability. This not only limits the organization's agility in responding to new data trends and technologies but also constrains openness and interoperability. Such limitations can hinder an organization's capacity to handle increasing data volumes or to leverage emerging data technologies and tools effectively.

Vendor lock-in can and will lead to escalating costs, starting with subscription licensing and continuing with the need to develop workarounds, which often achieve a less-than-ideal level of interoperability.

In an environment where data technologies evolve rapidly, avoiding technology lock-in is critical for sustaining a competitive advantage and ensuring the long-term effectiveness and flexibility of data management strategies.

By implementing a data platform model that is technology-agnostic, businesses have the flexibility to use any data technologies for storage, mining, analytics, and visualization. It is also imperative that the data platform be capable of implementing processes, policies, and rules for governance through automation. Since these are only several of the attributes of the ideal platform, it's best to start with a more comprehensive (but by no means exhaustive) list of needed platform features.

Both options can lock an organization into a legacy technology stack that makes it impossible to integrate new tools and approaches. This becomes the root of many other issues with data management, including:

- Compatibility issues.
- Lack of openness and interoperability.
- Increased IT expenditures.

Any of these can stop an organization from creating effective data practices. By implementing a technology-agnostic data platform, organizations achieve a solution capable of:

- Providing seamless integration with various technologies and tools
- Impose internal data practices
- Achieving a high level of integration to transform disparate data fragments into valuable business insights
- Driving strategic decisions and fostering a data-driven culture
- Delivering agile alignment and elevating data practices for tangible business value with major cost savings in:
  - Subscriptions
  - Migrations
  - Upkeep, management, and maintenance

This white paper explores the necessary elements of this transformative platform approach and how a modular design and technology-agnostic framework empowers enterprises. The goal in a business landscape where data is the currency of competitive innovation and customer experience (CX) is to enable every organization to:

- Craft their own data standards.
- Embed data governance best practices at every step.
- Transform data practices into tangible business value.

Moving beyond the old technology lock-in paradigm used by most platforms today to a vendor-agnostic approach empowers organizations to deliver integrated, flexible, and efficient data management. To grasp why this shift is essential for harnessing the true value of data, organizations must clearly understand the benefits of proper data practices.

3.
Data Best Practices

The relentless and growing flow of data can be like a raging river in every organization, mandating the means to direct and harness it effectively via well-structured data practices. A recent study shows that **76% of respondents** feel their data management systems are inadequate for business needs. Most are using around 19 different tools across their data and infrastructure environments.

By understanding the advantages of implementing proper data practices, organizations can see how they are the foundation for transforming their tech stack and data engagement.

BUILD WITH DATA INTEROPERABILITY IN MIND

The cornerstone of effective data management is interoperability, which organizations can only achieve through a technology-agnostic platform that uses automation, policies, and sophisticated metadata/data practices. By designing systems that facilitate easier data discovery and consumption across various technologies and products, enterprises can unlock greater levels of efficiency and insight.

This approach significantly reduces the time, costs, and silos that impede data integration to:

- Minimize the need for complex data translations or transformations
- Enable seamless interaction across different platforms and technologies
- Democratize data by making it more easily accessible for improved decision-making and efficiency

APPLY PLATFORM ENGINEERING PRINCIPLES

Data engineers face a great deal of cognitive load due to the need for broad tool and practice mastery, testing, troubleshooting, and data operations. These challenges lower productivity and make it difficult to reach the primary goal of converting data into business value.

Platform engineering principles, which incorporate self-service capabilities into the design process, are the best way to overcome this challenge and:

- Enhance productivity
- Foster a sense of ownership among software engineering teams
- Streamline workflows
- Improve the quality of the code and systems

These principles create an environment where the focus shifts from just “keeping the lights on” to innovating and pushing the boundaries of what data can achieve.

ALWAYS DEVELOP EVOLUTIONARY DATA ARCHITECTURES

Flexibility, scalability, and resilience are key attributes of any data architecture within a constantly evolving technology landscape.

A critical aspect here is the prioritization of modularity and component decoupling, allowing data engineers to update or replace parts of the data architecture without affecting it in its entirety. This approach not only future-proofs the data architecture but also simplifies maintenance and upgrades.

ENSURE DATA DOCUMENTATION IS PART OF BUSINESS VALUE

Data documentation is a vital component of data governance that should be integrated early in the data project lifecycle. It must be accessible to make sure users understand and trust the data. The documentation process involves clearly defining data sources, structures, and uses, making it easier for all stakeholders—from data scientists to business analysts—to leverage data effectively.

SHIFT DATA GOVERNANCE TO THE LEFT

The best method for integrating data governance early into the data engineering lifecycle is via a governance shift left approach. This enables organizations to achieve more efficient, accurate, and trusted data management.

The approach eliminates the historical separation of data and metadata management by harmonizing them within the data lifecycle, yielding significant benefits such as:

- Early quality control
- Elimination of process hand-offs
- Automatic data-metadata synchronization
- Automatic alignment of the data catalog with governance policies, which are **automatically enforced through code** to ensure reliability
- Elimination of data hand-offs and wasted time by empowering the data engineering team to create accessible, usable, and secure data for faster time to market

Governance shift left leads to a more agile and cost-effective data governance framework that positions organizations to manage data focused on strategic business outcomes. The enterprise can thus always ensure security, compliance, and quality are integral to the data engineering process, enabling data engineers to:

- Create context-aware governance policies that are documented, accessible, and self-explanatory
- Strengthen their data infrastructure
- Build trust among data users
- Expedite their transition to a data-driven future redefined by data governance
- Leverage metadata to create intelligent automation and enhance data consumption
- Ensure data and metadata are always in sync

Implementing policies as code enables the creation of metadata quality standards that must be adhered to and enforced during the CI/CD process. The result is complete and adaptive data documentation where end-to-end governance ensures accuracy, consistency, and compliance.

AUTOMATE THE DATA LIFECYCLE

Automation stands as a pivotal aspect of modern data practices and when building/maintaining any data platform. It brings comprehensive improvements in deployment, testing, accuracy, security, and compliance. By automating every aspect of the data lifecycle, organizations can enhance efficiency and reduce the likelihood of human error and unnecessarily high costs.

The ideal approach is to start with a machine-readable, logical model for generating automation protocols; these should be independent of specific technologies to show representations of components, data, and metadata relationships.

Integration tests during the delivery process are also vital to prevent breaking changes while also ensuring data consistency. Data contracts, meanwhile, help to fully automate data testing and maintain data integrity during deployment and operation. This includes structured and automated vulnerability assessments and modeling for breaking changes. Both are crucial for identifying and mitigating potential security threats early in the data lifecycle.

PRIORITIZE DATA SECURITY

Global data breaches led to over **6.4 million data records being exposed** in Q1 2023 alone, while **74% of data breaches** could be attributed to human factors, such as errors or misuse. This shows how imperative it is for every organization to make data security a priority for data management.

The best approach here involves the implementation of fine-grained authentication and authorization mechanisms to ensure only authorized personnel have access to sensitive data.

Organizations should also emphasize privacy-preserving data mining (PPDM) techniques, which enable meaningful data extraction from large data sets while preserving the privacy of individual data entries.

This holistic focus on leveraging automation for proper data governance and privacy ensures data security and compliance are a foundational aspect of data practices.

INFUSE PRODUCT THINKING INTO DATA ENGINEERING

When organizations view data assets and pipelines as **buildable data projects** that solve specific problems, they can focus on creating business value.

Embracing all of the above data practices offers a roadmap for effective data management that drives the extraction of tangible business-driven value. Organizations can effectively leverage these practices to drive operational innovation, efficiency, and sustainability. However, effectively using them requires looking at the pros and cons surrounding the fundamental decision of buying versus building a platform.

4.

Defining the Needed Platform Features

Becoming a data-driven enterprise is a complex undertaking with many hurdles. By implementing a data platform model that is technology-agnostic, businesses have the flexibility to use any data technologies for storage, mining, analytics, and visualization. It is also imperative that the data platform be capable of implementing processes, policies, and rules for governance through automation.

In an ideal world, data engineers and users within an enterprise would choose a data platform that brings the following attributes (by no means an exhaustive list) into their data tech stack:

- True tech-agnostic agility
- Scalability
- Governance
- Automation
- Interoperability

These capabilities deliver a future-proof platform that would enable all the best practices discussed in this white paper and many more. However, there are a few additional critical features, which we discuss in greater detail below.

CUSTOMIZATION AND COMPOSITION CAPABILITIES

The often-discussed technology lock-in challenge for most data platforms and technologies within a data stack prohibits scalability, true interoperability/integration, and broad customization.

To remove these limitations, the platform should be designed in an intuitive way that prioritizes customization and composability to meet diverse data platform needs. The architecture must be technology-agnostic to allow for tailored solutions that empower an enterprise to adapt the platform to their unique data landscapes.

DATA SHAPING WITH TEMPLATES, TECHNOLOGY ADAPTERS, AND POLICIES

There are inherent limitations that come with the modern data tech stack, catalogs, and the ability to constantly update data and metadata for growing productization uses. This creates a major time and cost drain for enterprises. It also increases the risk of data errors, misuse, and security vulnerabilities.

The ideal platform should thus be capable of leveraging templates, technology adapters, and policies for effective data shaping to counter these and other limitations.

Templates

The platform should give data engineers and practitioners the ability to create fully automated, reusable, tested, and validated templates. Templates provide a structured pathway or “golden path” for users, setting clear expectations and establishing best practices for technology use.

Easy implementation will significantly cut time-to-market and costs while increasing standardization through best practices that maximize quality, security, and compliance.

Policies

Policies create quality gates on how any specific technology can and should be used. They also allow defining rules of interaction between platform users, such as the definition of and compliance with data contracts. Policies additionally create pattern standardization across the tech stack.

Data engineers and practitioners need the ability to create computational governance policies for deployment and runtime. They should also be capable of collecting results in a unified location for better auditing and monitoring to manage the end-to-end governance process of all data components.

5.

Build vs. Buy

The decision to build or buy a data platform is pivotal for any organization aiming to optimize their data operations. This choice involves evaluating not just the initial investment but also long-term implications in terms of scalability, adaptability, and total cost of ownership (TCO).

	BUY witboost	BUILD	Benefits of Witboost
TIME TO MARKET	3 MONTHS	2 YEARS	ROLL-OUT GUARANTEED SUCCESS
LONG LASTING COMMITMENT	NO	YES	PROTECTION FROM CHANGE OF CONTROL, CHANGE OF PRIORITIES, STAKEHOLDER SUPPORT
CAPEX COST	200K $	> 2M $	90% LOWER IMPLEMENTATION COST
TCO	300K $	1.5M $ / YEAR	80% LOWER TCO
REGULAR UPDATES	EACH MONTH	MANAGE THE ROADMAP	1-2 FTE SAVING
TECHNICAL SUPPORT	INCLUDED	DEDICATED TEAM	2-3 FTE SAVING
OUT OF THE BOX INTEGRATIONS	AN INCREASING ECOSYSTEM OF INTEGRATION	IN HOUSE DEVELOPMENT	80% LOWER COST OF TECHNOLOGY INTEGRATION
ADOPTION	EASY ONBOARDING THANKS TO UI AND GUARDRAILS	DEVELOPER "LEVEL" TOOLS WILL BE A BARRIER FOR ADOPTION	IMPROVE AND BOOST THE ADOPTION OF THE PLATFORM
PLATFORM TEAM REPUTATION	VERY FAST DELIVERY VALUE AND A HIGH DEGREE OF MATURITY	INCREMENTAL RELEASES, CREATING REWORK ON DOMAINS AND BECOMING A BOTTLENECK	PLATFORM TEAM REPUTATION IS A CRUCIAL ELEMENT FOR PLATFORM ADOPTION
DESIGN CHALLENGES	RELY ON SOLID DESIGN FOUNDATION, FLEXIBLE ENOUGH TO ADAPT TO EVERY TECHNOLOGICAL AND PRACTICE CONTEXT	RISK OF BEING TIED TO LEGACY DECISIONS AND TECHNOLOGIES THAT BECOME OUTDATED	REDUCE RISK OF FAILURE OR EVOLVABILITY READBLOCKS

Time

Building a custom data solution offers the advantage of a platform tailored to specific organizational needs; however, it comes with significant commitments. On average, developing a custom data platform can take upwards of three years—a substantial time investment compared to the average three months it takes to integrate a purchased solution.

Cost

Financially, the capital expenditure for building a custom solution can be substantial, often exceeding $2 million, compared to around $200,000 for procuring an off-the-shelf solution. Additionally, the TCO for a DIY solution averages around $1.5 million per year, significantly higher than the $300,000 associated with purchased systems.

These statistics underscore the advantages of purchasing over building, particularly in terms of quicker deployment, lower initial costs, and reduced TCO. However, when it comes to flexibility, agility, scalability, and avoidance of vendor lock-in, both options are equally suspect.

6.

How Witboost Meets Data Platform Needs for Business Value

The long list of platform features needed by the modern enterprise is beyond the scope of this white paper. However, Witboost encompasses and exceeds all attributes covered above, enabling teams to:

- Quickly gather meaningful and contextual insights from complex and large analytics architectures.
- Focus on what's important and avoid useless or misleading metrics.
- Extend control capabilities that govern data and deployment to harness data's full potential.

As a data experience platform, Witboost embodies an API-first approach to ensure flexible, secure data project streamlining. Its design gives data engineers and practitioners the control and freedom to:

- Streamline their data projects
- Build automated data provisioning processes driven by data governance guardrails
- Easily access a real and high-fidelity view of the entire lifecycle of all data projects
- Deliver unified, self-service experiences across the entire lifecycle of a data project
- Offer data practitioners an interoperable ecosystem of data tools

ACCELERATED INTEGRATION USING WITBOOST OPEN-SOURCE STARTER KITS

Witboost speeds up integration with its open-source Starter Kits that enable data stakeholders to:

- Access **Provisioners** to automate the lifecycle of a resource tied to a technology and implement use cases (e.g., ACL management, application deployment) while adhering to organizational standards
- Create **Templates** to provide a structured pathway for users, set clear expectations, and establish best practices
- Simplify the deployment of technology-specific components
- Facilitate rapid and efficient integration into existing systems
- Establish transparent and automatically enforced contracts between the platform and users

- Create guardrails for business-driven data initiatives that are high-quality and secure
- Leverage tools to monitor and audit data pipelines, access, and usage
- Protect the organization from data breaches and compliance violations
- Implement data mesh, data lake, data warehouse (DWH), and ML platforms with high-quality standards and velocity
- Avoid technology or vendor lock-in via a technology-agnostic design and principles
- Create a fully extensible and customizable data architecture via plugins and microservices on both the back and front ends

The demands associated with proper data management will continue to change and grow more complex in unpredictable ways. To meet such challenges, organizations must implement data practices that act as the foundation on which data management solutions are built. At the same time, the resulting platform embodying those data practices must be capable of endless agility, flexibility, and future-proof scalability to meet those emerging challenges.

While the Witboost Starter Kit will expand to include additional components, thefirst release is now available with Provisioners and Templates.
To learn more about the Witboost Starter Kit, click here.

To see how Witboost can transform your data and data practices into tangible business value through a measurable ROI, try out Witboost's ROI estimator here.

Sources

1. https://ecm.capitalone.com/WCM/software/self-service-data-strategies---capital-one-software.pdf (page 6)
2. Enterprise Strategy Group eBook, Being a Data-First Leader Continues to Matter, August 2023 (Slide 12)
3. https://www.verizon.com/business/resources/reports/dbir/ (opening page)
4. https://www.statista.com/statistics/1307426/number-of-data-breaches-worldwide/
5. https://www.gartner.com/en/information-technology/glossary/dataops

About Witboost



Witboost is a pioneering platform **streamlining complex data projects across various platforms**, enabling seamless **data production and consumption**. This unified approach empowers you to **fully utilize your data without platform-specific hurdles**, fostering smoother collaboration across teams.

It seamlessly blends business-relevant information, **data governance processes**, and **IT delivery**, ensuring technically sound data projects aligned with strategic objectives. **Witboost facilitates data-driven decision-making** while maintaining data security, ethics, and regulatory compliance.

Moreover, **Witboost maximizes data potential through automation, freeing resources for strategic initiatives.** Apply your data for growth, innovation, and competitive advantage.

Find out more at: www.witboost.com

witboost