Progress® DataDirect®

Overcoming 6 Challenges of Managing Information Products in a Modernized Hybrid Data Architecture

WHITEPAPER

Data Modernization: Embracing Distribution and Conceptual Information Products

Organizations are increasingly adopting a variety of distributed platforms for their applications and data, such as cloud service providers and Software-as-a-Service (SaaS) applications. The opportunity to migrate to the cloud has inspired technology architects to consider wholesale enterprise data modernization as a way of democratizing data use and increasing data literacy by making corporate data items across the data landscape more accessible. Yet these organizational attempts at data democratization remain constrained by several factors. These include ties to legacy on-premises applications, the need to accommodate hosted platforms and ongoing commitments to SaaS platforms and applications that hold enterprise data hostage.

As a result, the need to satisfy data-consumer expectations across a variety of data usage scenarios hinders the ability to achieve data modernization. For example, materializing consistent and up-to-date account information to present to a customer depends on accessing and governing account data. The challenge is in accessing the data in real time from various sources, including hybrid multicloud environments, on-premises systems, cloud-based data warehouses and hosted applications (e.g., data in an SaaS CRM system).

We have become acclimated to the oft-repeated assertion that "data is an asset," but perhaps it may be better to consider treating data (and, correspondingly, information) as a product that is manufactured from different source data sets according to design blueprints. Identifying the business use cases that direct the configuration of information products will drive the development of the design blueprint, as well as surface any potential issues. These issues can impede data accessibility, impact data quality and prevent compliance with externally imposed directives.

To adopt a "data product" approach, one must review the business drivers and data usage scenarios, and then consider the challenges of manufacturing those information products using data from a hybrid environment. Introducing data lifecycle patterns can help overcome those challenges and defuse the latent timebombs that eventually diminish data value.

Business Drivers Inspiring Data Product Blueprints

A data product approach presumes the objective of producing a usable data set composed of data pulled from different sources and across an increasingly distributed hybrid data landscape. In turn, information engineers must engage the information consumers, review the business scenarios and solicit their requirements to identify the information product configurations.

Some example business scenarios that motivate the manufacture of data products are:

- **Business initiatives**. Activities that require visibility into corporate information drawn from multiple business functions. Examples include mergers and acquisitions (M&A) that must accumulate customer information from a variety of business lines, corporate spinoffs that need to differentiate multiple products and product lines or partnership agreements that involve a limited set of services under prescribed conditions.

- **Omnichannel customer experience**. Over the past few years, agile businesses rapidly adapted their processes to accommodate different ways to meet customer needs across all lines of business via a variety of traditional and digital communication channels (e.g., online, mobile device, telephone, in person). In other words, the business must provide customers with a consistent and accurate view of all their interactions across all touchpoints in a way that is seamless and transparent to the customer.

- **Policy directives.** Organizations are becoming increasingly sensitive to the business value of compliance. Aside from the traditional viewpoint of legal observance and avoidance of penalties, companies are focusing attention on the GRC (Governance, Risk, Compliance) activities to not just minimally comply with laws but demonstrate their commitment to transparency and good corporate citizenship around governed operations. Activities such as integrated compliance, continuous auditability and mandatory reporting are driven by information artifacts produced with data drawn from across the data landscape.

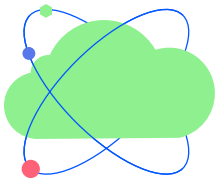
- **Process optimization.** Businesses looking to improve operational production and streamline automation of transaction processing need continuous operational reporting across all business functions to help identify process bottlenecks.

6 Challenges: Managing Information Products Across a Hybrid Environment

To support these drivers, an enterprise data strategy must accommodate three key objectives:

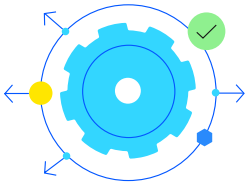
1. Seamless access to data
2. The ability to produce verifiably high-quality data products
3. Logical reference models for available data concepts such as customer, product, etc.

A corporate data platform modernization strategy must be robust enough to address all these drivers. However, the organic nature of incremental and siloed application, system and data migrations across a variety of cloud service providers and hosted/SaaS providers impacts the ability to deploy a coherent data modernization strategy that meets these goals. As a result, many organizations face these six challenges of managing information products across a complex data landscape:



Challenge #1: Data landscape complexity.

Modern data landscapes are increasingly complex. While some applications continue to operate on-premises, incremental migration of applications to a variety of cloud service providers and platforms leads to a dynamically evolving hybrid data environment. In addition, organizations that have subscribed to cloud-based SaaS platforms (e.g., for sales, marketing and customer relationship applications) have data managed by the SaaS providers. Ultimately, the increased distribution of data across a variety of file structures, platforms, systems and providers complicates the ability to access said data and limits the delivery of harmonized views.

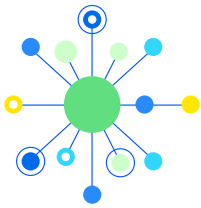


Challenge #2: Data quality requirements.

Data quality can be a challenge when there is only a single source of data and a limited number of data consumers. It becomes even more difficult when pulling data from multiple sources created under different circumstances and managed within distinct process flows. Streaming data from a distributed environment introduces new complications, such as synchronization and data currency. With the growing set of people accessing and using data, many have varying quality expectations that must be met.

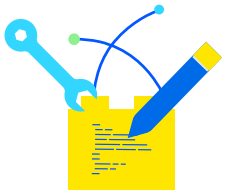
Challenge #3: Real-time integration.

Analysts and other data consumers rely on consolidated data to facilitate business processes such as enabling a positive omnichannel customer experience. To accommodate these processes, a modernized data environment must integrate and deliver unified views of information composed of data pulled from multiple sources in real time. However, a combination of data distribution and use of different tiers of cloud-based storage introduce increased latency into data requests, which impedes predictable real-time data delivery.



Challenge #4: Data fusion.

Many business initiatives, such as integrated recommendations or analysis for M&A, rely on data from different sources. And omnichannel customer processes require entity identification, identity resolution and a consolidated view of customer data. Conventional support for these processes can be tricky, requiring master data management techniques and tools. Intelligent data fusion is further complicated when requiring access to high-quality data from hosted environments.



Challenge #5: Unified data governance.

Enterprise-wide data literacy that satisfies data consumer quality and usability expectations implies the implementation of end-to-end process governance. While instituting monitoring and observability at all stages of data pipelines is difficult, facilitating governance within a hosted system requires alignment with the SaaS provider's best practices and coordination across the data pipelines.

Challenge #6: Compliance.

There is a growing inventory of laws, regulatory directives and industry standards mandated by external authorities requiring continuous compliance monitoring and auditability. In addition, companies adopting an ESG (environmental, social and corporate governance) strategy may have self-imposed compliance criteria that need to be monitored. These compliance directives must be incorporated into the data landscape, and in a modernized distributed environment, data policy compliance must be pervasive yet configurable.

Data Lifecycle Patterns Help Address These Challenges

Cloud modernization has transformed the digital landscape, making traditional data access patterns obsolete. Extracting data from operational systems and pumping that data through stages into a data warehouse may work in a traditional on-premises reporting and analytics environment, but organizations need to rethink how data operations in the modernized environment are architected and governed so that the "manufacturing" of data products is streamlined, and the end user's business drivers are supported.

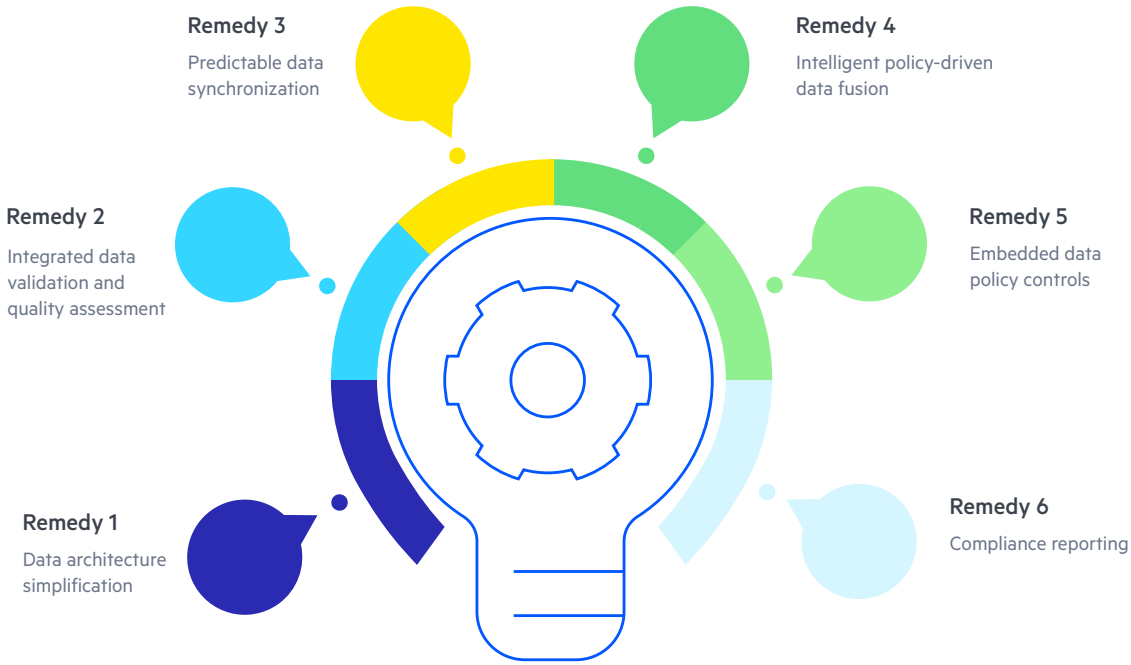
Three data lifecycle patterns can be incorporated into the data strategy as a way of overcoming the six challenges. These include:

- **Data lifecycle pattern #1: Managed connectivity.** By necessity, enabling access to all the data sources distributed across the modernized hybrid data environment implies connecting to a wide variety of systems, platforms, data organizations and file layouts. Data consumers should not be forced to learn the syntax and mechanics of every data source. The goal of this pattern is to integrate a data layer that is preconfigured with multiple connectors, providing access to a wide variety of data resources. A managed data connectivity product provides this data layer.

- **Data lifecycle pattern #2: Pipeline orchestration.** The rates of information consumption continue to increase, but expecting a centralized IT team to provide on-demand support for a growing number of downstream consumers creates an IT bottleneck and increases the friction in information production. This pattern empowers data engineers to work with the data consumers to configure their data pipelines and rely on enterprise services to manage, execute and orchestrate how information products are manufactured. Data pipeline orchestration tools scan the set of data pipelines, seek to optimize where redundant data accesses can be harmonized and synchronize data flows to support predictable data delivery.

- **Data lifecycle pattern #3: Virtualization.** While the data connectivity technology addresses data accessibility from the data sources distributed across the hybrid landscape, and the data pipeline orchestration supports the movement of data, the remaining objective is to simplify the reference models for commonly shared concepts. The goal of the virtualization pattern is to provide a logical semantic layer on top of accessed data. This facilitates the semantic layer in composing result sets that can be customized for each data user's role.

An organizational data strategy supporting modernization should explore how to incorporate these data lifecycle patterns within the hybrid data landscape, no matter what type of platforms, on-premises systems, cloud service provider storage paradigms or SaaS hosts it comprises. Applying these patterns for accessing source data, implementing quality and protection controls, monitoring for compliance with policy directives and delivering a unified view of commonly shared data entity objects can help overcome our six challenges.



- **Remedy #1: Data architecture simplification.** Complexity is rooted in the variety of systems, platforms and formats, and is compounded by distribution. By providing connectors to a wide variety of systems and platforms, a managed connectivity solution eliminates the data engineer's need to be familiar with the many potential interfaces for data access.

- **Remedy #2: Integrated data validation and quality assessment.** Managed connectivity solutions work together with data catalogs and data observability tools. They allow data engineers to partner with data consumers to solicit their data quality and usability requirements and define data validation policies that can be deployed within data pipelines. At the same time, the data observability tools monitor data flows and flag data values that vary from historical expectations.

- **Remedy #3: Predictable data synchronization.** Pipeline management and orchestration works with virtualization to identify sources of data delays, introduce caching and streamlined access and optimize to minimize data latency delays. This facilitates the predictable integration of data to meet real-time demands.

- **Remedy #4: Intelligent policy-driven data fusion.** A data layer powered by data virtualization allows for the specification of a semantic layer using reference data models for master data concepts such as *customer*. The semantic data layer removes the burden of harmonizing customer data plucked from a variety of sources. It also performs identity resolution and uses context-based rules to materialize a unified view of customer data based on the data consumer usage scenarios.

- **Remedy #5: Embedded data policy controls.** Translating data consumer expectations into defined sets of data rules is one source of data policies. Organizational access controls that limit data set access only to authenticated individuals are another source of policies. These policies need to be deployed no matter what data sets are accessed or how they are processed. Data pipeline orchestration technologies allow data engineers to integrate data policy controls across all data pipelines and provide the ability to alert data consumers when there are issues that need to be addressed.

- **Remedy #6: Compliance reporting.** Finally, probably the most critical source of data policies is the array of laws and regulations that continue to be passed and updated. The same techniques for integrating embedded data controls for data validation provided by data pipeline orchestration and data virtualization are used to facilitate continuous monitoring for compliance. The result is the ability to produce auditable reports demonstrating compliance with laws and regulations.

Considerations: What to Look for in Data Hubs for Information Products

We have examined the key characteristics of a data strategy that addresses the six challenges of managing the manufacture of information products that meet data user needs. A data hub provides seamless data access and facilitates the delivery of verifiably high-quality data products. Essentially a data layer that enables semantic reference models for master data concepts embedded within a virtualization engine helps you overcome critical barriers to success when modernizing to a hybrid data landscape.

When considering alternatives for a data hub, look for solutions that:

- Have a lightweight footprint that does not introduce additional computational burden to the data landscape
- Are engineered to leverage cluster computing for high availability and performance moderation through load balancing
- Can access many data sources through ODBC, JDBC and REST APIs
- Support data security through encryption at rest and in motion
- Simplify configuration and oversight using management APIs
- Provide managed data throttling to help with synchronization
- Enforce authentication of data consumer identities

The right data hub solution will empower organizations to access and share data in real time while providing access to a wide array of sources, regardless of location.

About the Author

David Loshin, president of data strategy consulting company Knowledge Integrity, Inc. (www.knowledge-integrity.com), is a recognized thought leader and expert consultant in the areas of data governance, quality, management, and analytics. David is a prolific author regarding information management best practices as the author of numerous books and papers, including Big Data Analytics: From Strategic Planning to Enterprise Integration with Tools, Techniques, NoSQL, and Graph, The Practitioner's Guide to Data Quality Improvement, and Master Data Management. David is a frequent invited speaker at conferences, web seminars, and sponsored websites and channels.

David is also a Senior Lecturer and Faculty Lead for Careers and External Relations, as well as the former director of the Master of Information Management program at University of Maryland's College of Information Studies. David can be reached through LinkedIn, or email via loshin@knowledge-integrity.com.

Disclaimer

Overcoming 6 Challenges of Managing Information Products in a Modernized Hybrid Data Architecture was prepared by David Loshin in their personal capacity. The opinions or representations expressed herein are the author's own and do not necessarily reflect the views of Progress Software Corporation, or any of its affiliates or subsidiaries. All liability with respect to actions taken or not taken based on the contents Overcoming 6 Challenges of Managing Information Products in a Modernized Hybrid Data Architecture are hereby expressly disclaimed. The content on this posting is provided "as is" with no representations made that the content is error-free.

Worldwide Headquarters

Progress Software Corporation
15 Wayside Rd, Suite 400, Burlington, MA 01803, USA
Tel: +1-800-477-6473

facebook.com/progresssw
twitter.com/progresssw
youtube.com/progresssw
linkedin.com/company/progress-software
progress_sw_

Progress