DATA MODERNIZATION:

Modernize Your Legacy Systems without
Data Disruption

By David Loshin
**Recognized thought leader in all things data**
President of Knowledge Integrity, Inc.

Data environments have evolved organically, resulting in many different systems and ways to access and use data. Because these systems were designed originally to meet specific needs of various lines of business, it's typical for corporate data to exist in silos across the enterprise. As such, many organizations inherently rely on an array of legacy systems that are prime targets for modernization; however, they still need to access the existing environment until the future environment is production-ready.

That said, there is plenty of motivation for modernizing data environments, such as making emerging technologies (such as Hadoop or NoSQL tools like MongoDB) available, to provide scalable performance at significantly lower acquisition and operational costs. In addition, as analytics become entrenched within the organization, a strategy for embracing both conventional structured data, and a variety of sources of unstructured data for analysis, is necessary.

Legacy Systems Cannot Support the New Information-Powered Economy

Most legacy environments are not able to support the newer information-powered economy. Conventional databases may be sufficient for historical reporting and the typical "slice-and-dice" analyses driven by the data warehouse paradigm. However, predictive and prescriptive analytics rely on access to a wide variety of data sources and streams. To remain competitive, your business intelligence and analytics applications must access all of the data, quickly and easily.
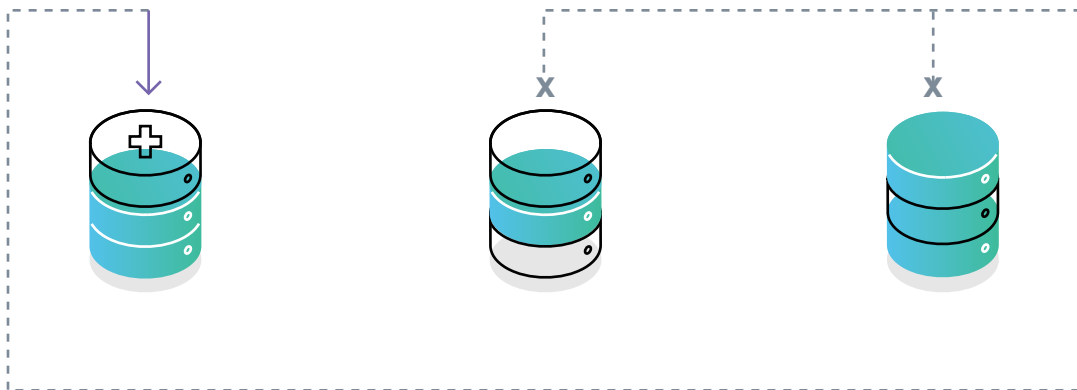
Data Modernization is the First Step

Because data modernization is the first step in an overall technology modernization strategy, maintaining stability for the application environment is critical. Transparent data access ensures this stability by leveraging the existing data landscape while enabling incremental technology adoption and transition. Addressing the challenges of data modernization requires a flexible environment that enables access regardless of where the data lives, and can adapt as business needs change.

Data connectivity solutions enable self-service access to a variety of data sources for analysis, while expediting application modernization. Transparent access to a broad range of data environments increases agility, lowers costs, reduces risks and speeds time-to-value for new technology adoption.

This paper examines market drivers and challenges associated with system and data modernization, and suggests best practices that balance new technology acquisition with bridging technologies for data connectivity that can extend the life of legacy systems as replacement applications are adopted. You'll find examples demonstrating the key benefits of high-performance, broad-based data connectivity, and learn what features to look for when evaluating data connectivity solutions.
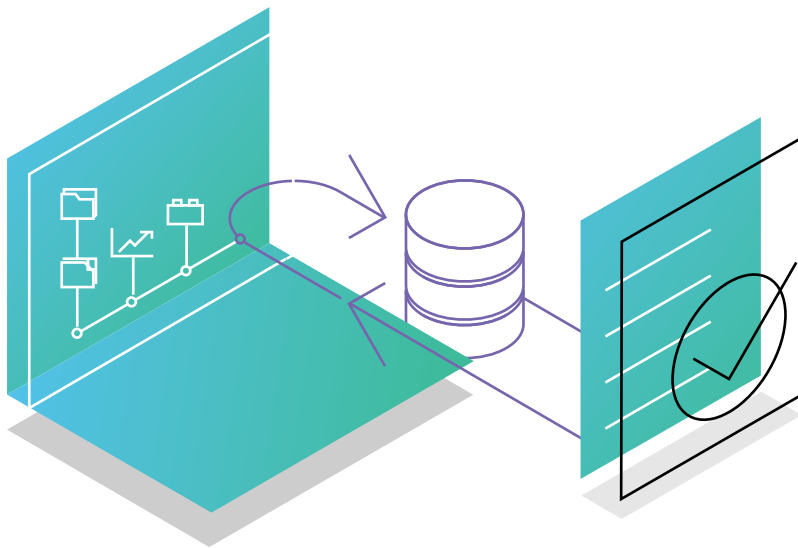
Market Drivers and Trends for Data Modernization

Adopting new technologies as they emerge is exciting but can be risky. Although new technologies are expected to provide costs savings and expanded capabilities, there can be unknown risks associated with transitioning from existing, trusted legacy solutions. Still, practical motivations for modernizing your data management environment are many:

- **Data utilization:** Data is no longer just input for transactions; it's a strategic asset. Data itself is no longer simply structured transaction records; it includes a wide variety of information artifacts from internal and external sources at volumes that legacy systems simply can't handle anymore.

- **Cross-platform integration:** Fully leveraging unstructured data, structured data in legacy systems and data managed off-premises in cloud/SaaS systems can be a challenge, requiring significant cost in resources, time and innovation to develop and deploy connectors to myriad data environments.

- **Evolving hybrid environment:** According to a recent survey report, respondents indicate using an increasing number of applications in off-premises environments, such as the public cloud or SaaS platforms[1]. The combination of internal systems, proprietary on-premises systems, SaaS and cloud-based applications such as Salesforce.com or Marketo, data warehouse appliances, and new platforms such as Hadoop have resulted in a rapidly changing enterprise system landscape.

- **Adoption of BI and analytics:** As the tools and techniques for reporting and analytics become more democratized, there is a growing demand for data availability, accessibility and utilization.

> Data is no longer just input for transactions; it's a strategic asset.

These issues highlight the need for a long-term strategy for legacy data migration and integration with modern data environments. However, most organizations wouldn't perform an immediate overhaul or "rip-and-replace" of the complete data management environment. It's more feasible to modernize selected applications and their related data systems incrementally.

Yet according to analyst firm Gartner, "most organizations are ill-prepared for the complexity of legacy data migration," because they are either unaware of the complexity or unable to address nontechnical issues such as coordinating people and processes[2]. This suggests the need for a more flexible future enterprise data access, integration and utilization environment that can adapt as business needs change. The centerpiece of this vision is flexible data connectivity that not only supports data availability no matter where the data lives, but also tackles the challenges associated with modernization.

1 Ring, Dr. Katy, "SaaS Vendors and IT Service Providers," 451 Research Game Changers, December 2015
2 Savage, Mel, "Legacy Data Migration Is a High-Risk Project– Be Prepared!"
December 5, 2014, Gartner Technical Professional Advice

Challenges for Modernization

The Hadoop ecosystem and the Hadoop Distributed File System (HDFS) in particular, presents a low-cost, yet easily scalable high-performance platform for managing the growing volumes of data used for transaction, operational and analytical processing. This environment can be deployed in-house or in a cloud environment, and easily extended on-demand by adding storage nodes. Either way, it lowers the costs of storage management, making it an obvious architecture to target as part of an application and data modernization strategy.

From a financial standpoint, a low-cost scalable infrastructure is enticing. Yet, there are challenges to consider prior to committing to any systemic application or data:

- **Understanding the current environment:** Most organizational data environments have evolved organically and have not followed any single set of architectural directives; understanding the various methods for capturing, storing and accessing data within the existing environment can therefore be challenging.

- **Siloed data organization:** The same or similar data concepts may be represented in different ways in different lines of business, but today's sophisticated analytics applications demand access to a horizontal view of data across the enterprise.

- **Reliance on legacy data and systems:** Many business processes use the data managed within legacy systems targeted for modernization, and it's difficult to figure out how to redirect those processes to the modernized systems.

> Today's sophisticated analytics applications demand access to a horizontal view of data across the enterprise.

- **Understanding the future environment:** Various emerging system architectures and data management paradigms (such as MongoDB, Hadoop and the conceptual data lake) all have unique characteristics that are relatively different from conventional mainframe data systems or relational database management systems. Yet a good understanding of how the future systems work is critical to any data migration.

- **The need for an unstructured data strategy:** Transaction processing, operational processing, and most business intelligence and reporting systems rely on structured data. However, one benefit of newer architectures is their ability to ingest and manage unstructured data, necessitating a strategy for effectively integrating and using unstructured data.

Clearly, managing these challenges requires preparation and makes a complete overhaul unwieldy and impractical.

Best Practices for Data Modernization

Addressing the challenges of data modernization is two-fold. First, you must address the technical engineering aspects associated with instantiating the new environment and migrating data sets from their existing sources. Second, you must ensure business continuity as the migration proceeds. Although the technical hurdles are challenging, there are more significant risks associated with a "rip-and-replace" approach to data modernization.

A more balanced approach to data modernization, coupled with good change management practices, a reasonable technology evolution roadmap and a program plan is your best bet for success. A roadmap will help you examine ways to balance the introduction of new technologies while continuing to enable access to existing systems, until migrating both functionality and users to a newer architecture is possible. The program plan provides details about the skills, processes and technologies required for a successful modernization. Together, the roadmap and program plan provide a comprehensive view of short-term and long-term milestones that supports the user community and sustains existing systems while enabling progress.
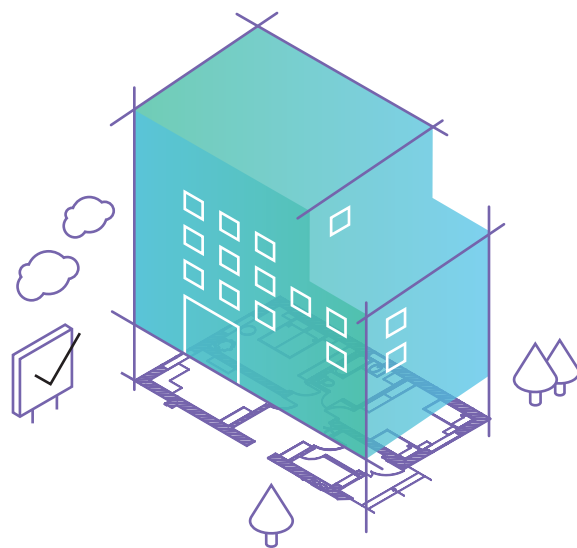
Program Plan

A program plan incorporates the following best practices to support users throughout what could be a relatively long transition period, during which both old and modern technologies must coexist.

- **Specify business needs for data:** How will business needs be influenced or impacted in a positive way by introducing newer technologies and processes?

- **Specify performance criteria:** How will you measure the degree of business value improvement as systems and users are moved to the new environment?

- **Identify the key stakeholders:** Who should be engaged in the decision making process until the systems selected for retirement have been decommissioned?

- **Document data consumption patterns:** Because data use will dictate the requirements for ensuring business continuity during any modernization plan, it is critical to understand and document data usage scenarios. This will help you to determine dependencies on existing legacy data assets and guide the timing of the migration tasks.

- **Catalog data assets:** Identify and prioritize the data assets that must be maintained for the sake of business continuity. What tools will enable continuous access to existing data sets as business applications are built using the new technology?

- **Modernization prioritization:** Which systems must be redesigned or moved first?

- **Provide self-service data access:** What techniques and tools will provide self-service access to existing data repositories? Exposing existing data assets to a broad developer and user community can inspire development of innovative applications that leverage modern technologies

- **Incorporate transition technology:** What bridging technologies (such as data virtualization, federation and data connectivity tools) will enable access to legacy data systems yet still maintain the expected level of service for application performance?

- **Education and training:** How will you train users to enable their transition from using the older systems to embracing the new technology?

Roadmap

The opportunity to incorporate new architectures into the organization may be driven by economics, but the success of the modernization program requires a modernization roadmap. That roadmap must go beyond basics of technology acquisition–it must incorporate these best practices for leveraging existing data assets.

A growing collection of use cases demonstrate the value of technologies that connect newer applications to existing data assets, and traditional applications to modern data architecture platforms. Using data connectivity technology to establish a bridge between emerging and legacy systems helps overcome the challenges of modernization.

Consider these two examples:

- **Transitioning to a modern NoSQL environment:** An organization decides to transition from SQL databases to MongoDB. This will enable them to track all customer interactions and transactions, and speed the capture, access and delivery of complete customer experience profiles to customer support representatives via dashboards. Existing reporting applications access data from the SQL database, but have difficulty executing the same queries against MongoDB. Introducing a data connectivity tool that exposes the data in MongoDB as normalized relational tables simplifies migrating applications and data to the new platform. At the same time, it provides self-service access to the new data environment, enabling existing users to report on and analyze data.

- **Employing a variety of Software as a Service (SaaS) data systems:** Another organization commits to using cloud-based SaaS applications for customer relationship management (CRM) and marketing applications. A data connectivity tool that provides high-speed access to all SaaS products used enables the organization to rapidly deploy internal data warehouses that connect securely to a variety of cloud-based systems, without significant additional investment in staff and resources.

> Using data connectivity technology to establish a bridge between emerging and legacy systems helps overcome the challenges of modernization.

These two examples highlight some key benefits of adopting high-performance, broad-based data connectivity tools as part of a data modernization program, including:

- **Self-service access:** Empowers existing users by continuing to provide real-time self-service access for reporting, business intelligence and analytics

- **Rapid access:** Streamlines layering the ODBC interface on top of optimized code for native access to target data environments

- **Faster time-to-value:** Enables instant connectivity to systems of record from anywhere

- **Increased agility:** Rapidly responds to new information requirements

- **Breadth of access:** Provides an omni-channel data experience, enabling connectivity to traditional database systems, emerging technologies such as the Hadoop ecosystem, NoSQL tools like MongoDB, unstructured data environments and cloud/SaaS-based products such as Salesforce.com and Marketo

- **Lowered costs and reduced risk:** Extends the lifetime of existing legacy systems
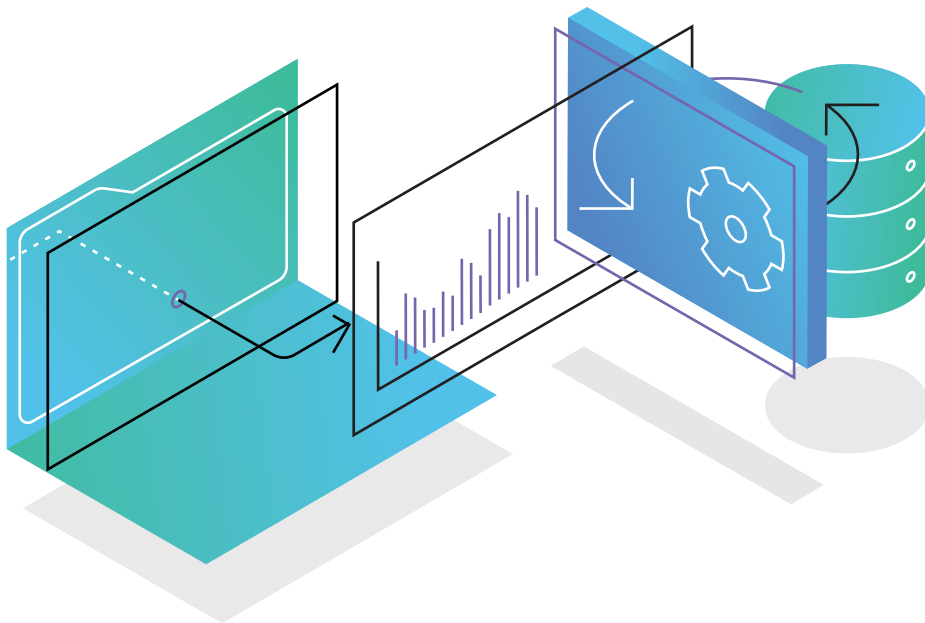
A Strategy for Data Connectivity: Key Solution Features

To provide a seamless transition, data and system modernization must leverage evolutionary changes and a strategy for connecting data from old and new sources. The future environment must provide seamless data availability, accessibility and utilization, no matter where the data lives or how it is managed. This means selecting a data connectivity solution with these features:

- **Hybrid:** Enables access to legacy environments, as well as the latest disruptive data sources such as HDFS or NoSQL data systems

- **Platform-independent:** Provides instant connectivity between applications and their data sources, whether they are deployed on premises or in the cloud

- **Comprehensive:** Supports a wide array of data environments

- **Industry standard:** Enables integration of legacy systems using industry standards for data access

- **Secure:** Provides secure and easy connectivity through any firewall

- **Real time:** Connects directly to data without making changes to legacy databases or moving/copying data

- **Production ready:** Incorporates into existing environment without complex integration

- **Vendor neutral:** Ensures broad compatibility and interoperability with tools and systems

- **High performance:** Provides fast data connectivity for data integration and ETL for analytics

This features checklist can help you select and evaluate candidate solutions for data connectivity. The best-of-breed solution exhibiting all of these features will provide stability and sustainability as your organization's modernization journey proceeds.
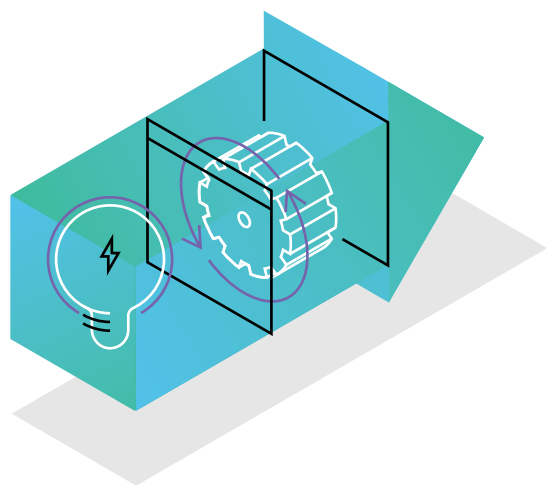
Key Takeaways

- Data environments have evolved organically to include many different systems, data silos, and methods for accessing and using data

- Organizations still rely on legacy systems, while acquiring emerging technologies such as Hadoop and NoSQL databases, and embracing BI and analytics to remain competitive

- Most legacy environments are not able to support the newer information-powered economy

- Data modernization is the first step in overall technology modernization

- Factors driving data modernization include the evolving hybrid environment, the democratization of BI/analytics, the potential of data for value growth and managing data across varied platforms

- A balanced approach to modernization maintains access to the existing environment until the future environment is production-ready

- The key to seamless modernization is a flexible data access environment that supports data availability, no matter where the data lives, and that can adapt as business needs change

- To succeed, any modernization program must incorporate best practices such as understanding the business needs for data, providing self-service access to data, and incorporating bridging technologies for data access

- The right data connectivity solution should be based on industry standards and support hybrid access across a wide range of legacy and emerging data sources

- Benefits of adopting broad-based connectivity tools as part of a data modernization program include self-service access, faster time-to-value, increased agility, lowered costs and reduced risk
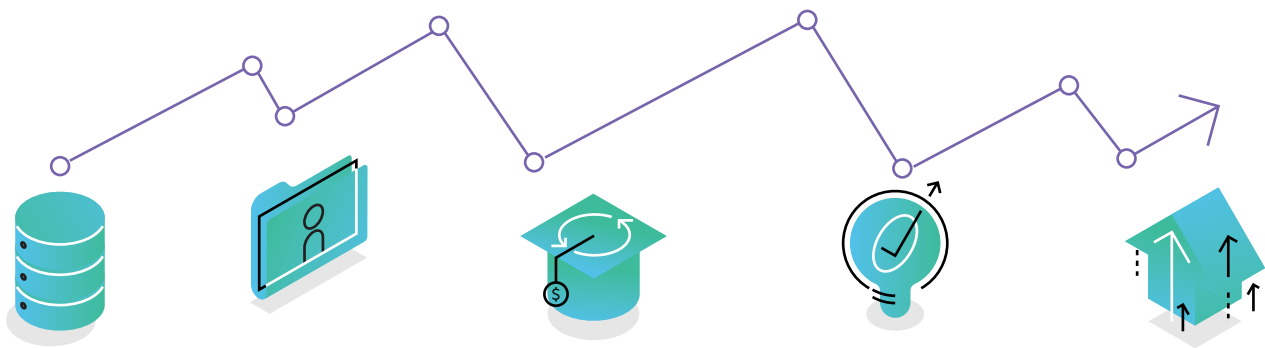
LEARN MORE

For more information on data connectivity solutions for your modernization program, please visit progress.com/datadirect-connectors

Author

David Loshin, president of Knowledge Integrity, Inc, is a recognized thought leader and expert consultant in the areas of analytics, big data, data governance, data quality, master data management and business intelligence. Along with consulting on numerous data management projects over the past 15 years, David is also a prolific author regarding business intelligence best practices, as the author of numerous books and papers on data management. David is a frequent invited speaker at conferences, web seminars and sponsored websites and TechTarget channels, and shares additional content at his notes and articles at http://dataqualitybook.com/.

About Progress

Progress (NASDAQ: PRGS) is a global leader in application development, empowering the digital transformation organizations need to create and sustain engaging user experiences in today's evolving marketplace. With offerings spanning web, mobile and data for on-premise and cloud environments, Progress powers startups and industry titans worldwide, promoting success one customer at a time. Learn about Progress at www.progress.com or 1-781-280-4000.

Worldwide Headquarters

Progress, 14 Oak Park, Bedford, MA 01730 USA Tel: +1 781 280-4000 Fax: +1 781 280-4095
On the Web at: www.progress.com
Find us on   (f) facebook.com/progresssw   (t) twitter.com/progresssw   (o) youtube.com/progresssw
For regional international office locations and contact information, please go to www.progress.com/worldwide

Progress®