

How to Build a Modern Data Architecture with Legacy Data



Introduction

Business competitiveness depends on an organization's ability to leverage data. Despite technological advances, enterprises still have trouble even accessing their data, especially legacy mainframe data. As a result, most companies have a fragmented data architecture that doesn't support their strategic goals.

Meanwhile, competitive pressures are building. Today's most successful companies operate beyond function-specific analytics and even interdepartmental analytics, enabling enterprise-wide analytics that incorporate data from a combination of internal and external sources. They're also using machine learning to answer questions they haven't been able to answer previously.

Unlike newer companies, which were born in the cloud, wellestablished companies haven't had the benefit of using all their data from day one. To compete effectively, they must integrate data from many disparate sources, although mainframe data is often missing because it's too difficult to access. With Precisely data integration software, any business can create a modern data architecture that includes any data source regardless of the data's type, format, origin, or location in a manner that's fast, easy, cost-effective, secure, and future-proof.



Different organizations are at different stages of data-driven maturity, but they all tend to face the following common challenges:

- **Data access**. Companies tend to lack the expertise they need to liberate data from all their systems, including their mainframes. In addition, individuals, departments, and business units hesitate to share their data with others for fear of losing control of it.
- Environmental complexity. Enterprise systems have expanded out from mainframes and enterprise data warehouses to mobile, cloud, and the IoT.
- Performance. Ever-growing data volumes overwhelm systems, which slows down processing and results in missed SLAs.
- Data quality. Poor data quality negatively impacts the accuracy of data analytics. It also affects the reliability of machine learning.
- Multiple data types and formats. More data types and formats make data integration more complex than ever.
- Data delays. Businesses lack timely access to data because it's trapped in systems.

- Data strategy. Many organizations lack a holistic data strategy that contemplates all their data. This issue is so critical that some organizations have appointed Chief Data Officers.
- Inadequate resources. Tight IT budgets and the current shortage of data-related talent slow progress.
- **Competitive pressures.** Data-savvy competitors are stealing customers and changing the competitive landscape.

Why Most "Modern Data Architectures" are Incomplete

2.5 quintillion bytes of data are created every day¹ and 90% of the world's data has been created in the past two years.² To stay relevant, companies need fast insights from data at scale.

However, modern data architecture don't just include modern data sources such as mobile, web, IoT and social data.

Decades worth of critical business data has accumulated in traditional data stores, and this valuable insight cannot be ignored.

¹ Domo Data	Never	Sleeps	6.0
² IBM Cloud			

Mainframe-Specific Challenges

Many different issues make it difficult to work with mainframe data, such as:

- **Data structure.** The hierarchical nature of mainframe data differs from relational data. In addition, the data structures can be so complicated (e.g., nested, repeating elements) that data processing speeds suffer. Further, COBOL Redefines enable data to be interpreted in different ways by different applications, so logic must be written to make the data usable with non-mainframe data.
- Metadata. Hundreds of copybooks may map to a single file. Usually a data tag specifies which copybook to use so mainframe applications can understand what to do with the data and how that maps to the copybook.
- **Data mapping.** Unlike an RDBMS, which requires data to be entered into a table or column, nothing enforces a set data structure on the mainframe. While a copybook describes how the data should appear, it can be changed in a manner that no longer matches the copybook.
- Different storage formats. As data storage costs have fallen over time, data compression has become less critical so numeric values stored one way on a mainframe may be stored differently in the cloud. Therefore, it's important to understand how to make mainframe data types legible to the outside world.

The complexity of modern data architecture requires flexible tools that will work with all data sources and across multiple environments. Precisely's data integration software makes connecting to any data sources, from traditional mainframes to the latest cloud platforms, simple.

How to Create a Modern Data Architecture

Step 1: Assess

The most effective way to start creating a sustainable modern data architecture is to align to your company's current and future goals. How well is your data architecture suited to meet those business goals? What does the business want to achieve that it's unable to or could do more effectively if it had access to richer datasets? Do you have an architecture that will also support the future needs of the company?

The next question is what data exists – and where? For example, is it in legacy systems in the data center, does it come from realtime data streams, cloud-based applications and repositories? Businesses often struggle to understand the complete scope of their data inventory. If some data assets are invisible, data analytics and machine learning accuracy may be compromised.

Another question is what data do you lack that is necessary to answer a query or meet a business goal? In other words, identifying the missing data you need may prove to be as important as understanding what data exists. For example, if a hotel wants to predict room night sales during a given week, it has to consider more than historical transaction data and seasonality during the same week in previous years. It also has to consider other external impacts such as weather and nearby events.

Still another consideration is latency – how often do you need fresh data from the mainframe to satisfy the needs of your use case? For example, if the hotel referenced above wants historical mainframe data for analytics, then a daily refresh might be sufficient. However, if the same hotel needs to update its room inventory in a cloud-based reservation system, transactions recorded in the mainframe would have to be fed to the cloud in real-time.

In addition, organizations need to contemplate regulatory compliance. While it's important to understand what data is available, it's also important to consider applicable laws or regulations your business must comply with.

Step 2: Plan Your Approach

While it's an admirable goal to identify an organization's wealth of data assets and make them available so insights can advance business objectives, it's wise to have a plan that ensures good intentions aren't derailed by unforeseen circumstances.

Take time to understand the following:

- What challenges will you face getting the data you need when you need it? Think about barriers to accessing the data at all, such as whether you're able to identify all data sources and destination targets regardless of the platforms and technologies including private, public, multi, or hybrid cloud(s). Also consider, whether you're able to access that data in the time frame required, especially if you need it in real-time. If it's a technical obstacle, look for an easyto-use data integration tool that's flexible, agile, and platformagnostic. If it's a people obstacle, you may have to navigate cultural and political issues. Quite often, a lack of data access is due to a combination of both technology and people challenges.
- Which stakeholders should be involved? Solid problem-solving means getting the right stakeholders involved from across the organization so you're in a position to enable a modern data

architecture that truly meets the needs of the business. Both business and IT stakeholders need to be involved.

- What sort of governance is required? Data governance can be tied to regulatory compliance or can be a mechanism for effectively managing your data as a valuable and strategic business asset. Either way, tracking where your data came from, how it's being used, and how it changed along the way, is a very good idea. Large enterprises tend to have governance committees, smaller companies may lack an equivalent. Regardless of an organization's size, governance policies should be memorialized in employeerelated documentation and supported automatically by your data integration solution with features such as end-to-end data lineage.
- What level(s) of security are required? Different types of data require different types of security. Providing secure access to all enterprise data and metadata lineage across platforms becomes critical for the modern data platform. That should also be addressed as part of the plan and also easily implementable in the data integration tool. Your data integration software should be flexible enough to connect to any relevant data source while adapting to an organization's unique governance and security policies

Step 3: Implement Your Plan

Once you have a plan in place, it's time to inventory the data in a level of detail that may not have been done previously. Then comes the traditionally hard part which is getting the data, organizing it, getting it into a usable form, and making it available to reporting systems, data analytics systems, or machine learning systems.

Regardless of how much you have tested, some bumps in the road are inevitable once you are in production. Some to be aware of include:

- Missed SLAs due to slower than expected performance
- Lost data from system outages or connectivity disruptions especially when moving data to/from the cloud
- Spikes in system resource utilization
- The need to manually tune your processes

Using data integration software that has a small footprint, with optimized performance across any platform and guaranteed data delivery can simplify the process greatly.

Step 4: Sustain

Once your modern data architecture is in place, it needs to be sustained versus maintained. Maintenance aims to preserve the status quo, which is not only inadvisable but virtually impossible given how quickly technology and business models change. Even if it were possible to freeze a data environment, it would only be capable of bridging the time of the freeze and the past. Also, maintenance tends to be done on a periodic basis, which may not be often enough to keep an organization competitive.

	Maintain	Sustain	
Purpose	Preserve status quo	Succeed despite change	
Characteristics	Periodic	Continuous	
Insight	Hindsight	Foresight	
Ecosystem	Locked into existing technology	Technology agnostic	

Sustainability is success that adapts to change. It's a continuous process that anticipates technological change rather than a period event. A solid data integration platform is technology agnostic so it can support whatever data sources, use cases and processing platforms exist today and in the future. Similarly, if your company faces a merger or acquisition, you'll be in a position to access and blend the data from both entities without the usual hassle, time, and expense. Rather than coding, tuning and re-writing integration jobs for each new environment, you can design once and deploy anywhere.

Enable Accurate Analytics and Machine Learning

Truly insightful analytics and machine learning require two things: access to data and good quality data. Precisely enables both.

Customers adopting or expanding their analytics are able to gain insights that were impossible to get before by combining modern and legacy data. Examples include customer journeys, supply chain optimization, fraud detection, and other cross-departmental insights that can increase revenue and reduce operating costs. With Precisely, you can access, profile and combine any data you need whether it's internal or external, static or streaming. You can also assure data quality, including golden records, which enables more accurate insights.

If your company uses machine learning, Precisely helps ensure that your training data is reliable. User-friendly data profiling tools help you understand whether the body of data you have is adequate to solve a target problem, and data quality software helps ensure your data is valid, accurate and fit for purpose.other external impacts such as weather and nearby events.

Drive More Value from Talent

IT budgets are always limited, and most of them are flat. Given that fact, IT departments and data teams need to minimize the need for rare and expensive skills and make better use of existing talent. Instead of spending valuable time and budget building and tuning their own data integration pipelines, organizations rely on Precisely's technology and expertise. That way, they can spend more time doing higher-value tasks such as mining data and training machine learning models.

Precisely's drag-and-drop simplicity enables:

- Easy onboarding
- Greater operational continuity when a member of the data team leaves
- Freedom from the anxiety, time, and expense that typically accompanies working with unfamiliar and complex data types, such as those found on the mainframe
- Increased efficiency
- Faster time to value

Future-Proof Your Data Strategy

Adaptability and simplification are critical in a successful modern data architecture. Look for tools that transcend operating systems and execution platforms, so you avoid lock-in and are always ready to take advantage of next-generation technologies. Precisely's application data integration and data quality products are designed with Intelligent Execution to simplify data management and future proof applications, enabling you to:

- Graphically design your jobs once and deploy them anywhere cloud, single-server, distributed platforms, or hybrid environments – with no changes or tuning required
- Easily move applications from standalone server environments to cloud and distributed platforms, including Spark and Hadoop
- Future-proof applications for emerging compute frameworks

Precisely's solutions enable your organization to successfully meet the challenges of today, while being prepared for the future.

Learn how <u>Precisely</u> can support your enterprise.

Precisely is a global leader in data integrity, providing accuracy and consistency in data for 12,000 customers in more than 100 countries, including 90 percent of the Fortune 100. Precisely enables companies to integrate, verify, locate, and enrich their data to power better business decisions. To learn more, visit www.precisely.com.

www.precisely.com