

Introduction

Contemporary technology often brings to mind cutting-edge advancements and Moore's Law, in which computing power is constantly improving, and with it, our ability to achieve previously unimagined feats with our data and software.

But that futuristic vision overlooks the very real importance and role of established technology and data. A business needs to look no further than legacy data to see just how vital the old can be in our race for the new.

Legacy data is the data that resides in long-standing systems, which are often foundational to the running of a business. For instance, those legacy systems could be an IBM i, a mainframe, or even in today's environment, enterprise data warehouses and relational databases. The legacy data in these mission-critical systems can be game-changing for a variety of applications. This data is often curated, high quality, and relevant to important aspects of the business. It is just the sort of data needed by AI, analytics, and other modern applications. What is often misunderstood about this data is how it can, and should, be used in real time to power a business.

This white paper explores why legacy data is frequently difficult to access in real time and describes use cases that illustrate how impactful legacy data can be when put to real time use.

What Does Real Time Access Mean for Legacy Data?

To begin, let's clarify how we are thinking about real time when it comes to legacy data. Real time access is obviously needed when transactions are taking place in those systems – for example, banking transactions commonly take place on mainframes – and that data constantly changes. Those changes need to be accessed in real time by downstream systems that rely on that data. But even if the data on the legacy systems doesn't change as frequently, there are use cases that require real time access to this data. For example, a fraud detection application needs to marry real time data streams with data from legacy systems that provide the context needed to accurately evaluate the likelihood of fraud.



Why Is Legacy Data Hard to Access?

By definition, legacy data comes from systems that were designed and created years, or even decades, ago. Quite obviously, these systems don't always integrate seamlessly with today's newest technology. For instance, many mainframe applications were built in a different world. That application may have been running for a long time and it may be the backbone of a business, but it was built in an age when storage and memory were expensive.

The developers of legacy systems and applications applied many design tricks to save on space. As a result, much of the data coming from these systems is generated in obscure formats that are unfamiliar to people who are tasked with deriving insights from next-generation platforms. The structure of legacy data itself is often entirely different from the structure of contemporary data. It may be hierarchical, rather than relational or semi-relational. This makes it incredibly challenging for those who are not legacy data experts to translate the data into a workable format that will also integrate with more recently generated data.

Additionally, enterprises often have security concerns about providing access to legacy data. It's not just a matter of the legacy data not being encrypted in the right way (which it often isn't). There's also the issue that the people maintaining legacy systems can be very protective of this data, which represents another hurdle to accessing it in real time.

What Are Some Common Use Cases That Are Not Real Time?

Not every use case for legacy data requires the data to be available in real time. For instance, many companies use legacy data for archival purposes. Companies will move the vast amount of data they have stored in their legacy systems into a cheaper storage system, such as a data lake or the cloud in an effort to save money. Another use case aimed at cost reduction is system optimization or application modernization. In these cases, the legacy data is not only put onto a new platform, but some of the processing associated with using the data is transferred as well. This happens frequently with batch processes that run nightly, and companies often want to use a lower cost system to run these.



Why Has Real Time Access Become More Important and What Are the Use Cases?

But while not all use cases involve legacy data needing to be available in real time, there is an increasing number of use cases where such immediate access is essential. Many of these have arisen because of the need to provide a better experience to customers. Here are some use cases driving real time access.

Use Case 1: Customer Experience

An important goal for enterprises today is ensuring excellent customer experience. This involves meeting or exceeding customer expectations for service no matter how the customer reaches out. That definitely includes making sure that the information provided to the customer – whether through a mobile app, web site or even a customer service agent – is always up to date. Companies have to be able to meet this demand, even if much of the data in question comes from legacy systems.

Take as an example an insurance company that wants to provide real time changes in a pension fund to its customers in its mobile application. It needs to do this in part due to customer expectations and in part because its competitors can offer this feature. If the insurance company cannot link its legacy data to the application, this creates risk for the business when its clients can choose to go to a competitor that is offering something that it cannot.

Use Case 2: BI and Analytics

Data plays an ever-increasingly important role in the daily decision making of businesses. Legacy data needs to be included in the BI and analytics a company runs so that the business has complete and accurate information at all times. Companies need real time access to legacy data to ensure that all teams are operating off of, and making decisions based upon, the most recent and relevant data, informed by the largest possible context of historical data as well. This is particularly true in industries where trends, suppliers, customers, competitors and markets change quickly. If companies don't have real time data from legacy systems, they're missing crucial context and cannot make the most intelligent response.



Use Case 3: Transactions

Having legacy data in real time for transactionally based uses is also very important. For instance, a core banking application might be running on a mainframe, meaning every transaction, from an ATM withdrawal to a bank deposit, must be kept up-to-date in real time so that customers, and the bank itself, have high visibility to prevent money laundering and fraud.

On the fraud and money laundering front, with transactional data, companies are also using analytics and AI to identify trends. Streaming data or IoT data must be combined with the context that legacy data provides in order to be of real value. If you don't have access to your legacy data, you simply can't engage in this type of risk mitigation. The legacy data is providing context to the application so that a decision about whether fraud is occurring can be accurately made. Without that context, companies are rudderless.

Use Case 4: Large Datasets

Large datasets that are accessed by the public often rely on legacy systems that run on mainframes. Government data is an example. If a state government maintains its court records on a mainframe, those records will include vital information on the criminality of individuals, as well as who is scheduled to appear at what time and in what courtroom. This all has to be readily available and searchable by any citizen at any time.

For a state government, enabling public use of an application on a mainframe is expensive because the system charges based on how much CPU is being used. At the same time, maintenance of such applications is also expensive, and government agencies often lack the personnel to fix issues that come up.

To overcome these challenges, and provide better customer service, the government agency needs to be able to rehost the application that is running on a legacy system onto a modern platform that incorporates the legacy data with real time updates, as court clerks are entering changes throughout the day and citizens cannot wait on finding out about those changes due to batch processing. Real time legacy data availability is key here because even if the data has been moved from a legacy application, citizens will expect a top-level experience and want to be able to perform the same searches that they always have.



Conclusion

The most advanced technologies and powerful use cases that organizations are investing in depend on legacy data coming from systems that may be decades old. Legacy data offers core transactional insights and key context to make sense of data from next-generation platforms. Blending legacy data with streaming newer data in real time fuels a broad range of digital initiatives, from customer experience to analytics and AI.

About Syncsort

Syncsort is the global leader in Big Iron to Big Data software. We organize data everywhere to keep the world working – the same data that powers machine learning, AI and predictive analytics. We use our decades of experience so that more than 7,000 customers, including 84 of the Fortune 100, can quickly extract value from their critical data anytime, anywhere. Our products provide a simple way to optimize, assure, integrate, and advance data, helping to solve for the present and prepare for the future. Learn more at syncsort.com.

© 2019 Syncsort Incorporated. All rights reserved. All other company and product names used herein may be the trademarks of their respective companies.