Maximize the ROI of Your Mainframe Data

How to Choose the Right Tool to Integrate Your Mainframe into Your Business
**Table of Contents**

Executive Summary ................................................. 2

Look for a Consistent Approach to Data Access ................. 4
- Standard Connections to Each Subsystem ................... 4
- Data Delivery Consistency ..................................... 4
- Source Data Availability ....................................... 4
- Data Store and Transport Independence .................... 5
- Case Study: Con-way Freight .................................. 5

The Right Data at the Right Time ................................. 6
- Batch Processing ................................................. 6
- Change Data Capture and Real-Time Latency Processing ... 6
- Case Study: A Midwestern Insurance Company .............. 7

Performance and Restart/Recovery ............................... 7
- Multithreading .................................................. 7
- Offload Processing ............................................. 8
- Offload Persistent Data Store .................................. 8
- Integrated Restart/Recovery ................................... 8
- Case Study: West Bend Mutual Insurance ................... 9

Conclusion .................................................................. 10
Executive Summary

You've been hearing for years that the mainframe is a dinosaur headed for extinction. Yet your own experience suggests otherwise. Your mainframe is extremely reliable, it processes large numbers of transactions, and its availability statistics are impressive. No wonder analysts estimate that mainframes still manage 70 to 80 percent of the world’s most critical data. Far from being obsolete technology, your mainframe may remain a critical part of your IT infrastructure for years to come. And that means it needs to exchange data smoothly with all your other business systems.

Traditionally, mainframe integration has been a painful process: writing code to unload legacy data sources, more code to convert and transfer the data, and still more code to load the data into your target system. Developing and maintaining this code is costly, and it distracts the IT organization from other important tasks—which may leave you wondering whether your mainframe is delivering any return on investment at all.

As businesses develop ever more ways to leverage data on their mainframes, system programmers and business architects are under increasing pressure to find an affordable way to integrate mainframes with other systems. Custom coding has become too expensive to be feasible, yet choosing the wrong product from the variety of available mainframe integration solutions on the market can force you to use scarce resources to customize the product to your environment. In some cases, this may cost as much as, or even more than, an in-house coded solution.

Integrating mainframe data into your business decision-making applications isn’t optional, but it doesn’t have to be painful. To make the process as smooth and simple as possible, you need to choose the right integration method and product. This white paper examines the critical issues to consider, starting with a checklist to guide your search for a mainframe integration toolset:

- Does it provide standard native connections to each data source?
- Is the data delivery consistent across each source?
- Does the toolset provide access to all the major sources on the mainframe?
- Does the tool support most of the typical mainframe data stores (DB2, VSAM, IMS, IDMS, etc)?
- Does the tool require proprietary communication protocols or does it use standard TCP/IP?
- Does the toolset furnish the most critical latency options?
- What are the performance characteristics of the toolset?
- Does the toolset supply an end-to-end restart/recovery facility?
We also discuss basic needs such as a consistent approach to data access for each mainframe source, the ability to access and propagate that data in any time frame or latency, and the ability to enhance performance through offload processing and multithreading. We conclude that a complete end-to-end restart/recovery solution is the key to true mainframe data integration.

In addition, you’ll find case studies detailing how several companies solved their mainframe integration challenges:

- An insurance company that deployed a product to deliver a single unified platform for all its business integration needs

- A freight company that replaced an outdated tool with one that could provide real-time replication from multiple mainframe databases to an operational data store (ODS) of its choosing

- Another insurance company that found a product that readily implemented its requirement for an ODS with real-time integration of mainframe data
Look for a Consistent Approach to Data Access

Standard Connections to Each Subsystem

A good mainframe integration tool distinguishes itself from the rest in the way it accesses most of the data sources it claims to support. It must have native access to data sources and/or standard access to data source change capture exits.

For example, if you’re looking for a product that accesses IBM’s Information Management System (IMS), the product itself should be able to talk DLI or Data Language/Interface to the database and provide a connection via any of the native connection methods: DLI, Batch Message Processing (BMP), Database Management Batch (DBB), or Open Database Access (ODBA). The product should not be dependent on a COBOL or PL1 program under the covers to connect to IMS and unload the data to a flat file (for subsequent reading from the integration tool). If you need to read changes from, say, DB2, the tool itself should be able to read from the standard DB2 change capture exit, the IFI306 interface.

A mainframe integration tool that lacks this feature will require additional coding and support.

Data Delivery Consistency

All too often, business units buy a “single integrated mainframe integration solution” that turns out to be bundled pieces of different software from different vendors. IT departments end up struggling to stitch the technology together with code and customization for each data source involved, and sometimes each database or table within each data source. This requires an IT team with special expertise in each piece of software as well as in the data source itself. You can avoid this pitfall by choosing a solution that uses a common software and common interface for all supported data source types. This commonality ensures that your integration solution can accommodate not just your current data sources but also those you may use in the future.

In addition, a mainframe integration solution must deliver data to the integration platform or integration tool in a consistent way. In many cases, the toolset delivers different source data in different formats, requiring the IT organization to do significant amounts of coding and customization to standardize the formats—in effect, writing code to integrate the integration tool. This is hardly an acceptable solution. Instead, choose a data integration tool that delivers all the data to your target system in a single format or view, as if all the data sources came from the same data source type. Your solution should also provide bidirectional functionality, enabling it to write back to its mainframe sources.

Source Data Availability

No matter how consistent data delivery is, it’s useless if it only provides that consistency for a handful of data source types. Your mainframe integration solution should be able to support multiple mainframe source data stores, both the ones you have and those you may acquire in the future, whether by organic growth or through a merger. Otherwise, you risk finding yourself back at square one: coding source access support for new sources, buying and customizing a different product to handle them, or simply excluding that source from your integration project.

A quality mainframe integration tool addresses most, if not all, of the major mainframe data stores, including new architectures such as z/Linux, and provides consistent access to all of them. This is both a good measure of the tool and sound preparation for the future.
Data Store and Transport Independence

Pass up any integration tools that require you to land data in a specific type of data store or transport it using a proprietary transport mechanism or queue, even if you’re already licensed for that type of data store or transport method. These lock you into decisions that may prove extremely expensive if you decide to make a change further down the development road. Instead, look for data store and data transport independence. That includes multiple methods of transporting the data, the ability to change target data stores, optional stores for landing intermediate data, and the option of delivering the data all the way to any target of your choosing with no intermediate landing. This is, after all, what data integration concerns: using sources, transport, and targets chosen by the organization, not the technology vendor.

Data store and transport independence has other benefits, as well. By enabling the IT organization to use a single vendor and tool for the integration effort, it makes testing easier and simplifies ongoing maintenance of the integration application.

Case Study: Con-way Freight

Con-way Freight, one of the largest shippers in the United States, was using an outdated tool for data propagation. It only provided access to one or two mainframe sources; more importantly, it did not furnish target data store independence. The company needed to pull data from DB2 and IMS on the mainframe, as well as Oracle on the midrange side, to load a historical ODS on a Netezza data warehouse appliance. This would subsequently feed its ETL and business intelligence processes.

The source data consisted of more than 1,000 tables containing data on more than 60,000 shipments each day; each shipment record might change as much as 30 times daily. Because the outdated tool provided no centralized metadata, maintenance and administration were cumbersome. “It could take four to five days just to set up one table, and then we would have to redo the entire process again when we moved it to production,” says Roopali Doshi, technical lead for Con-way’s Enterprise Business Intelligence Group.

To modernize its mainframe integration, Con-way chose Informatica® PowerExchange® for its ability to connect to multiple databases and integrate tightly with Informatica PowerCenter®, the company’s ETL tool of choice. “We can recover and retrieve changed data for the last couple of days if necessary,” Doshi says. “PowerExchange has the ability to switch between batch, real time and change data capture easily, without the need for the manual changing of items on the mainframe side.”
The Right Data at the Right Time

Batch Processing

One of the most basic requirements of a mainframe integration tool is the ability to process data in a batch mode. Batch processing serves a number of functions for an IT team, including delivering all of the data from a source to a target on demand and materializing the target data store to prepare it to receive changes in a change data capture (CDC) or real-time environment. It also provides ad hoc access to mainframe data and facilitates testing or analyzing the data. The ability to set up sources for batch access can also be considered the federation of a database. An outstanding mainframe integration tool combines this useful and necessary function with CDC capability.

Change Data Capture and Real-Time Latency Processing

Change Data Capture or CDC has become a fundamental requirement for mainframe data integration over the last decade. Real-time latency processing, the ability to deliver mainframe transactions immediately to the business decision-making data stores, is one of the most desirable solutions for IT departments because of its positive effects on performance. It’s no longer necessary to move all the data in a source to the target for each business decision because the data is continually moved and synchronized as changes occur within the source.

Various mainframe integration tools claim some sort of CDC facility, but they don’t necessarily take data usability and reusability into account. Can the changes be propagated more than once? Can they be propagated to multiple targets at the same time? The ability to “capture once and consume many”—that is, the ability to deliver captured changes to multiple targets many times—is a critical requirement in choosing a mainframe integration solution. The solution must also avoid performing destructive reads of the changed data; as the change is read and propagated, it should not be removed from the persistent change data store. This impacts how the change capture tool handles restart and recovery functions.

The ability to maintain source data referential integrity to the target is also an important consideration. Many integration projects require the use of source key and -nonkey fields to populate target keys. This requirement can pose significant problems if the tool being used can capture changes but lacks a mechanism to maintain the referential integrity all the way through propagation to the target data store.

CDC takes advantage of the strengths of each platform your business uses. Although business decision-making software often resides on midrange systems, a mainframe typically handles transaction and high-volume batch processing because of its superior processing power. This factor can now be an advantage, not a dilemma. Most IT organizations are not necessarily looking to someday move data off the mainframe; they want an integration tool that lets them exploit the mainframe. A tool that provides CDC and the ability to deliver those changes in real time for all major adapters provides a competitive advantage.
Case Study: A Midwestern Insurance Company

This large insurance company needed to consolidate its mainframe data from IMS and DB2 (as well as SQL Server) to an operational data store (ODS) in near real time to create a holistic view of all claims or policy data. The ODS, which would reside in SQL Server, needed to receive CDC from the source mainframe systems in less than 15 minutes.

The company determined that it needed a tool that could handle both batch and change data capture, supply a consistent approach to data access, and allow access to all sources through a single product. For the IMS sources, it also needed the flexibility to capture changes from IMS logs.

“We needed a single-solution vendor to handle these multiple sources and Informatica was really the only tool on the market (without OEMing multiple tools) that could do it all through a single product,” says the insurer’s director of data services, adding that the company was pleased with PowerExchange’s ability to capture changes from either the control region directly (synchronous capture) or IMS logs.

Today, using PowerExchange, the company processes tens of thousands of transactions, totaling hundreds of gigabytes of information, each business week. “If Informatica’s product had not been able to solve our requirements, the project would not have been successful,” the director says. “Informatica provided us with the ability to consolidate mainframe and data store information in a near real-time environment, sourced from a single vendor.

Performance and Restart/Recovery

Performance and restart/recovery are critical to modernizing a business through mainframe integration. Today’s business world requires IT systems to be flexible and adaptable to demand, able to handle both organic business growth and the possibility of mergers and acquisitions. The IT organization must be able to tune the performance of a mainframe integration tool easily to incorporate new systems, data sources, and volumes of data. The integration tool must also have a complete restart/recovery solution. Anything less will make an integration tool merely a short-term solution for a growing company.

Multithreading

Access to mainframe data is only part of the equation for a true mainframe integration tool. A tool must be able to scale to handle large amounts of data without jeopardizing performance. Multithreading, the distinct component of a mainframe integration tool that processes up to several hundred pipelines of data from the same source simultaneously, thus becomes a vital requirement because it allows the tool to process more data in the same amount of time. Even IT organizations that currently have low volumes of data are wise to adopt this solution as part of a plan for future growth.

True multithreading is an optional process that is controlled not by archaic parameter files or the compiling of mainframe programs, but through the integration tool’s application interface. Users must have control of multithreading through the tool’s GUI so they can tune each individual application or session quickly and easily by changing requirements or data volumes.

“Our company faces challenges in finding and maintaining skill sets to support the mainframe. Critical parts of our business are now and will continue to utilize the mainframe in the future. As long as we access data on the mainframe, PowerExchange will be utilized.”
Offload Processing

For obvious reasons, processing cost per machine is a big concern for IT organizations planning mainframe integration initiatives. IT organization costs can be incurred in many different ways. Some companies pay for MIPS (million instructions per second) on the mainframe internally or through leasing time on the mainframe. Some actually incur costs on the midrange systems. Cost control is critical.

A mainframe integration tool with offload processing capability can address these issues by allowing the IT organization to control where the costs of processing occur. Offload processing tells the mainframe integration tool where to perform most of the heavy work. For example, a company that needs to protect the mainframe source system from as much impact as possible can use offload processing to move raw mainframe data to the target system, where it would be converted, formatted, and filtered for propagating to the target. On the other hand, a company that wants most of the work done on the source side can assign the software to perform the work there before passing it on for propagation.

In either case, an integration tool should allow you to choose where the bulk of your data processing will affect the CPU, both for individual extractions and when processing and propagating captured changes.

Offload Persistent Data Store

For IT organizations strapped for mainframe processing power or the costs associated with it, a best-in-class mainframe integration tool will also include the ability to relocate the persistent data store it uses to hold change capture data. This ability further improves performance by shifting the processing load rather than placing it on the same system that holds the source data.

Integrated Restart/Recovery

A complete end-to-end restart/recovery solution may be the most important component of a mainframe integration tool. Many available tools may have a restart/recovery solution for individual parts of the process: one for the mainframe piece, another for the transport tool, and a third for the midrange propagation component. However, this also creates many potential points of failure. And in the event of failure, none of these individual solutions will save you from having to completely rematerialize the target system.

When starting over is not an option—especially in cases with complex or high-demand latency requirements—only a single end-to-end restart/recovery solution can recover if the data source, capture mechanism, transport, or propagation engine fails. The integration tool should handle any restart or recovery from a single location, such as the process control or the monitoring interface, rather than requiring separate pieces in each location. The solution should also maintain recovery information in a persistent data store that can handle commits to the target and the integration tool at the same time (usually a relational target).

Moreover, the mainframe integration tool should also be configurable to restart itself if the failure meets certain criteria, and this configuration should be possible through the tool’s GUI for every individual session or extraction.

A truly comprehensive and robust mainframe integration tool does not require you to start over and reintegrate after a failure.
Case Study: West Bend Mutual Insurance

West Bend Mutual Insurance Company was growing, and with that growth came a massive increase in customer data from multiple sources, much of it in mainframe-based VSAM and DB2 systems as well as SQL Server. To improve its strategic decision making, the company needed a single consolidated view of customer information. To achieve this goal, the company decided to create a master data management (MDM) system that would capture changes from all source systems and deliver the data in real time to an MDM portal.

West Bend adopted Informatica PowerExchange to capture data and deliver it in real time to the data integration solution, Informatica PowerCenter. This solution provides complete integration across systems, with common data delivery and complete restart/recovery protection.

“Having a single platform that allows us to access and capture changed data from multiple platforms has been a huge win for us,” says Spencer Tabbert, IT manager of Data Integration Services for West Bend Mutual Insurance. “It has speeded up our development time, saves us on support time, and it has been extremely important for us because we were able to have consistent integration processes across the board for VSAM, DB2, and SQL Server. That was just not something we could have done if we had to interface with one custom solution that handled VSAM, another for DB2, and yet another for SQL Server. I am sure we would have ended up with all very different integration processes to move that data into our data store.”

PowerExchange even helped West Bend conquer its most vexing integration challenge, its Policy Administration System, a proprietary file system of VSAM files that are accessed only via special I/O modules containing compression algorithms. PowerExchange is able to call user exits where appropriate, at the access level, the record level, and even the field level. “This was another big win for us, how quickly we utilized PowerExchange to access our custom VSAM files through our proprietary I/O modules,” Tabbert says. “We had access to all of our files in less than a week.”

West Bend prioritized restart and recovery functionality, according to Tabbert: “The capability for us to bring down our RT [real-time change capture] processes, knowing that we can start them back up at any point in time and pick up where we left off, is integral to ensuring that we have the appropriate integrity around data that is also housed in our customer data store. Unless there is that type of solution in place, it would not even be an option for us to run this system.

“The difficult part of the mainframe [integration] has always been from the perspective of the integration [with other systems] and understanding [of the mainframe itself],” Tabbert adds, concluding, “By providing tools like PowerExchange coupled with PowerCenter, Informatica makes that integration much easier.”

“Having a single platform that allows us to access and capture changed data from multiple platforms has been a huge win for us. It has speeded up our development time, saves us on support time, and it has been extremely important for us because we were able to have consistent integration processes across the board for VSAM, DB2, and SQL Server.”
Conclusion

For most IT organizations, the mainframe is here to stay as part of the data infrastructure. Whether you call it mainframe integration, modernization, federation, or migration, the question isn’t whether to do it, but how—and how to find a best-of-breed tool to do it well. Given the plethora of mainframe integration solutions available, IT decision makers need to know not just what criteria separate the wheat from the chaff but also how to design a plan and select a tool that suits both today’s needs and tomorrow’s potential requirements.

In this paper, we set out the criteria on which IT decision makers should base this decision:

- a consistent approach to data access
- independence of data store and transport
- access to all major data sources
- standard native access to each data source for batch or change data capture

We also examined the failings of subpar tools in the face of changing business needs and latency requirements—for example, the inability to access data stores in batch or CDC modes, or limits to the number and type of data capture sources or new architectures a tool can accommodate. Finally, we discussed the importance of selecting a mainframe integration tool that meets all these requirements while also offering the ability to increase integration performance and provide true end-to-end restart/recovery capabilities.

As you evaluate mainframe integration tools, don’t just take a vendor’s word that its solution will meet these critical criteria. Follow the Missouri state motto: “Show me.” Require the vendor to install its solution on your systems and demonstrate it using your data and your infrastructure. How many experts did the vendor call in to make the solution work? How long did it take to install and set up? When problems arose, did the documentation provide all the answers, or did the experts have to call the vendor’s development team—and if so, would you have the same level of access to that development team? Were problems solved with simple parameter changes, or did the experts make behind-the-scenes code changes they couldn’t or wouldn’t share with you? Above all, were the results satisfactory?
Every company in the three case studies achieved results beyond their expectations by choosing the Informatica platform as their data integration tool. As these case studies show, PowerExchange actually exceeds the criteria we’ve set out as basic to a mainframe integration tool. PowerExchange is also tightly integrated with PowerCenter, Informatica’s enterprise data integration solution, to provide end-to-end integration of a mainframe with other business systems—thus reducing the total cost of ownership of both the mainframe and the solution.

With native access to all major mainframe source data stores, including VSAM, DB2, IMS, ADABAS, IDMS, and DATACOM, as well as custom proprietary data stores, PowerExchange supplies a consistent approach to mainframe data access. It also standardizes data delivery through its integration with PowerCenter, delivering data in a consistent relational format regardless of source without requiring delivery into any particular data store or using any special transport or queue. PowerExchange’s large selection of source adapters raises it to the top tier of comprehensive mainframe integration tools.

PowerExchange also accesses and delivers source data to the target data store or system in batch, CDC, and real-time modes, further enhancing an IT organization’s ability to integrate new sources of data from organic growth or mergers and acquisitions. PowerExchange and PowerCenter provide still more flexibility by supporting processing on z/Linux.

Most notably, the complete integration of the PowerExchange/PowerCenter solution gives it best-in-class performance and restart/recovery facilities: multithread processing controlled through the PowerCenter GUI, offload processing allowing the IT team to control whether data is processed on the target or source system, and persistent data store offloading that gives the IT organization complete flexibility in choosing where to hold changed data when creating a CDC or real-time environment for mainframe integration. If a failure occurs at any point in the process, the Informatica solution also integrates and stores all recovery information, ensuring that the appropriate response is easy to track, understand, and restart from the propagation side—without missing or duplicating data.

A true modern mainframe integration solution more than meets all of these criteria. The failure to meet any of them can severely restrict an IT organization—and doesn’t that make the point of data integration moot?