Considerations for Mainframe Application Modernisation

Data Migration and Data Integration Strategies
Mainframe Application Modernisation: Data Migration and Integration Considerations

Executive Summary | 1
Business Challenge | 1
Requirements for Data Migration and Integration | 2
DMExpress Mainframe Application Modernisation Solution | 2
- EBCDIC to ASCII Conversion with Support for Various Mainframe Sources, Formats, and Data Types | 3
- Robust Data Transformation Capabilities | 4
- Open Systems Equivalents of the Mainframe Data Integration Processes | 6
- Deep Support for Rehosting of Mainframe Applications | 7
- Bi-Directional Support | 8
- Support for Changed Data Capture Processing | 8
- Embedded Data Protection Capabilities | 8

Conclusion | 10
Executive Summary

Mainframe application modernisation initiatives, regardless of the implementation approach, all require careful attention to the translation and transformation of the data in order to maximise success. The complex and specific nature of mainframe data sources and formats pose challenging obstacles in these projects. According to one analyst, fully 84% of the data migration portions of application modernisation initiatives fail.

In addition, the inability to perform the equivalent data integration and transformation processes in the open systems environment with the required levels of performance and scalability often impedes the success of mainframe application modernisation projects.

This paper examines the issues surrounding the data migration and data integration aspects of mainframe application modernisation initiatives, and provides specific functional recommendations to consider in order to address these issues.

Business Challenge

Mainframe-based production applications are both widespread and mission critical. Various analysts report that at least 70% of the world’s transactional production applications are running on mainframe platforms, and that a typical enterprise organisation has millions of lines of legacy code in production.

But various business and technology drivers are motivating companies to modernise or re-platform their mainframe applications to run on open systems such as Windows and UNIX. These drivers include lower hardware and software costs resulting from a reduction in MIPS, wider availability of skill sets, greater extensibility of applications, and others.

Mainframe application modernisation initiatives are often complex and include many phases, each of which has unique organisational and technical challenges. Even migrating the data between the platforms, one of the earliest — and seemingly simplest — phases, is rife with hidden complexity. This phase is a critical prerequisite to all subsequent phases, and yet it is a frequent point of failure. One analyst survey found that the success rate (percentage delivered on time and on budget) for the data migration portion of [application modernisation] projects is just 16%.

These metrics should not be surprising, since most mainframe applications have been accumulating data for years or decades, resulting in volumes of legacy data that can exceed hundreds of terabytes. In addition, mainframe data sources and formats are often difficult to examine, manipulate, and convert.

Another critical component of mainframe modernisation initiatives is to be able to perform corresponding data integration and transformation tasks, such as sorting, merging, copying, joining, in the open systems environment with high levels of performance, scalability and reliability, and without requiring significant or complex development efforts.

KEY TAKEAWAYS

1. Multiple business and technology drivers are motivating companies to modernise or re-platform their mainframe applications to run on open systems.

2. Data translation and transformation processing is a critical and challenging component.

3. Syncsort DMExpress provides deep functional capabilities to support the specific and complex requirements of mainframe application modernisation projects.

4. DMExpress delivers a fast, resource-efficient solution for mainframe data migration and data integration on open systems.

5. Syncsort’s 40 years of mainframe software leadership gives DMExpress a depth and breadth of functionality that is unmatched by any other data integration software.

“For mainframe application modernisation initiatives, migrating data between the platforms is rife with hidden complexity.”

Organisations must plan for and properly address these data considerations in order to successfully accomplish their mainframe application modernisation initiatives. Various implementation approaches exist for modernising mainframe-based applications, including:

- Rearchitecting mainframe applications for the open systems environment
- Replacing mainframe applications with open systems-based packaged applications
- Rehosting existing mainframe code on open systems

For all of these approaches, the requirements around accurately and efficiently handling the data are critical for success.

**Requirements for Data Migration and Integration**

A critical — and challenging — part of a mainframe modernisation initiative is converting the often massive amounts of mainframe EBCDIC data into ASCII format while preserving packed-decimal and binary numeric values. Since the volume of data to be migrated can be huge, the data integration software must be able to process large data sets without failing. Ideally, the software should be able to process the data both quickly (short elapsed time) and efficiently (minimum use of hardware resources).

The mainframe data must not only be translated from the mainframe format to ASCII, it must subsequently be transformed into the appropriate format so that it can be loaded into the open systems-based application. Transformation of the records is not a trivial process, since mainframe production processing often uses complex, hierarchical record structures to pack significant amounts of information into a single data set. Several levels of information are typically bundled into a single data set through the use of multiple record types, multiple composite field groupings, and arrays — sometimes varying in repetition from record to record. This hierarchical data organisation often needs to be broken up and normalised for relational storage.

The data integration software used to process these complex records must not only be able to understand the complicated record compositions, but must also be able to leverage the existing metadata that describes these complex structures.

Mainframe applications store these data structures in a variety of highly specialised formats, creating special challenges for data integration software that must recognise and process the data. Without built-in support for these mainframe sources, record formats, and data types, the projects require time consuming and error-prone manual interaction and custom coding. The data integration software used for the initiative should have deep support for COBOL copybook metadata, mainframe data sets such as VSAM, and the various mainframe, UNIX and Windows data types (packed-decimal, big and little Endian binary numbers, IBM floating-point, IEEE floating-point, and so on).

**DMExpress Mainframe Application Modernisation Solution**

Syncsort, a recognised leader in mainframe and open systems-based data integration software, provides deep functional capabilities and extremely efficient processing to support the specific and complex requirements associated with mainframe application modernisation projects.

At the core of the mainframe application modernisation solution is Syncsort’s DMExpress extreme performance data integration software. Through the use of proprietary sorting algorithms, I/O optimisation, parallel processing, and dynamic environmental monitoring techniques developed and refined over 40 years with over 12,000 customers including more than 90% of the Fortune 100, DMExpress delivers highly efficient data integration solutions with significantly reduced CPU time, elapsed time, and disk I/O activity for data processing applications. DMExpress leverages the benefits of high efficiency and low resource utilisation — resulting from the development of the fastest data processing utilities on the mainframe — with the latest architectural innovations of open systems platforms.

“Since the volume of data to be migrated can be huge, your data integration software must be able to quickly and efficiently process large data sets without failing.”
In practical terms, this equates to the shortest elapsed time, requiring the least amount of hardware resources, to process massive amounts of data.

On November 14, 2008, DMExpress set a world record for data integration performance by extracting, transforming, cleansing, and loading 5.4 TB of data in 57 minutes on commodity hardware. There is no faster or more resource efficient solution than DMExpress — neither packaged software, utility, nor custom coding — for migrating mainframe data for use with open systems-based applications and for performing data integration processing on open systems.

**EBCDIC to ASCII Conversion with Support for Various Mainframe Sources, Formats, and Data Types**

DMExpress provides deep mainframe support developed through more than 40 years of mainframe data software leadership. Syncsort’s mainframe-based software products have 50% market share, and Syncsort’s mainframe data sort application is the most frequently installed third party software application on IBM mainframes. As a result of this extensive mainframe experience, the depth and breadth of DMExpress’ mainframe data conversion functionality is unmatched by any other data integration software.

DMExpress provides built-in support for:
- Mainframe sources including fixed (blocked and unblocked), variable (blocked and unblocked), VSAM, and other sequential formats.
- Direct connection to the mainframe from open systems servers.
- Processing and converting EBCDIC and ASCII character data.
- All mainframe and open systems numeric data types, including binary integers, displayable (leading or trailing sign), zoned decimal, packed decimal, IBM floating point, IEEE floating point, etc. As a result, the need for any custom coding to support data types is eliminated. DMExpress avoids all of the data corruption problems associated with using FTP or other simple EBCDIC-ASCII converters on data that contains packed or binary values.
- Robust, extensive COBOL copybook support. While some products include high level copybook support, once the record layouts utilise non-trivial copybook formats, custom coding is required. In contrast, DMExpress provides unmatched support for complex formats, including "OCCURS", "OCCURS DEPENDING ON", "REDEFINES," and so on.
- Normalisation techniques while writing converted data to database tables, sequential files, indexed files, XML files, standard input, named pipes, or an invoking program.
- Heterogeneously combining data from the mainframe with data read from database tables, sequential files, indexed files, XML files, standard input, named pipes, or an invoking program.
- Record formats of fixed length, variable length, delimited text, variable length, Fortran unformatted, Micro Focus fixed and variable, Micro Focus line sequential, Micro Focus indexed, and ACUCOBOL-GT indexed records.
- Standard (ASCII, EBCDIC, folded ASCII, folded EBCDIC, multinational, UNICODE), locale-defined, and user-defined (one-for-one, one-for-two, two-for-one, and two-for-two replacements) collating sequences.

In addition, DMExpress provides built-in support for various rehosting system equivalents of mainframe data formats, to support rehosting initiatives.

**Robust Data Transformation Capabilities**

As part of the mainframe application modernisation process, it is usually necessary to perform transformations on the translated (ASCII) data before it can be loaded into the open systems based application. For example, DMExpress is often used to normalise legacy hierarchical records prior to loading into relational database management systems. DMExpress provides built-in support for a wide range of transformations, including capabilities to aggregate, join, sort, merge, copy and partition records to the various targets, and to filter, reformat, and transform records prior to output.

“Mainframe applications store data structures in a variety of highly specialised formats, creating special challenges for data integration software that must recognise and process the data.”
DMExpress provides data type and format conversions, comparison operators, logical operators, and an extensive library of transformation functions. DMExpress also provides a robust graphical custom expression development environment for defining and standardising higher level data transformations.

DMExpress’ built-in support for mainframe data, coupled with extensive transformation capabilities and a custom development environment, speeds and simplifies transformation processing of the converted data, while eliminating errors and enabling standardisation and consistency.

CUSTOMER EXAMPLE
A major US telecommunications company was migrating its mainframe-based new user account activation system to a UNIX-based HP platform. The system is a mission-critical production application that is required to activate all new accounts. With hundreds of thousands of new subscribers to their network every day, eliminating any system downtime associated with the migration was essential for this initiative.

The company initially planned to migrate all of the data from the mainframe system to the UNIX system using custom C++ code. The custom coded application would connect to the mainframe and the OLTP database, create approximately 1 terabyte of flat files, then translate, transform and load the data into the new system. The company mapped out the approach in detail, flowcharting all the processing, and found that they would need to code hundreds of steps of various transformation processes. Next, they wrote some of the code, processed a subset of the data, and calculated that processing the full data set through the entire conversion process would take a total of 8 days.

The projected total processing time forced the company to explore other approaches to handle the migration. Using DMExpress, the company was able to replace custom coding techniques with a graphical, point-and-click development environment, drastically simplifying the development of the application. More importantly for the company, the overall processing time for converting all of the data was compressed to 21 hours, reduced from 192 hours. The company closed on Saturday night, and when it opened for business again the following Monday morning, their UNIX-based account activation system was fully populated with all of the legacy data, and it was operating — in production — to activate new user accounts.

“"The depth and breadth of DMExpress’ mainframe data conversion functionality is unmatched by any other data integration software.""
Utilising data integration software that is specifically optimised for handling mainframe migration projects eliminates the risks associated with using utilities or custom coding techniques.

For data migration projects, Gartner recommends that organisations should “[a]void custom-coded or desktop-tool-based approaches — such approaches can create quality challenges and are limited in their ability to be reused and extended over time. In addition, they do not provide any capabilities for discovery and capture of metadata... Readily available packaged data integration tools and data quality tools can provide greater productivity, higher quality results, and more opportunities for reuse than custom-coded approaches.”

Custom Coding

Custom coded solutions are reported to be the most common method for converting data as part of mainframe migration initiatives, often employing COBOL on the mainframe. Custom coded scripts and programs are initially attractive, due to low initial costs and the availability of developer skills. However, these solutions quickly become saddled with problems, since they are error-prone, brittle, and difficult and expensive to debug and maintain. COBOL solutions on the mainframe for data migration can be particularly costly in terms of elapsed time, data latency, and MIPS. Executing data conversion programs can cost expensive CPU cycles, and the expansion of the numeric formats can substantially increase the size of data resulting in lengthened data transport durations.

Mainframe Utilities

Most mainframe-based file transfer programs have an option to automatically convert a file from EBCDIC to ASCII. These utilities can be used to translate or unpack binary data to ASCII. However, users should be aware of the following limitations.

- Mainframe processing is expensive compared with processing on open systems.
- Mainframe utilities often have limits on the number of job steps.
- Mainframe utilities expand the data.
- The “unpacking” process can cause the record layouts in the programs to no longer match the translated data.

Finally, FTP is only useful if the data that is being moved is all character. Using these commands to translate binary data compromises the integrity of the data.

Unix Utilities

The UNIX based conversion utilities, such as dd, have similar limitations, which make them unsuitable for most mainframe modernisation initiatives. For example, dd cannot convert packed decimals, has no support for COBOL copybook, and has limitations in speed and scale. The limitation that presents the greatest risk, however, is that the presence of a single typo in the syntax can corrupt some or all of the data on disk.

Alternatives to Data Integration Software: Custom Coding and Utilities

For data migration projects, Gartner recommends that organisations should “[a]void custom-coded or desktop-tool-based approaches — such approaches can create quality challenges and are limited in their ability to be reused and extended over time. In addition, they do not provide any capabilities for discovery and capture of metadata... Readily available packaged data integration tools and data quality tools can provide greater productivity, higher quality results, and more opportunities for reuse than custom-coded approaches.”

Utilising data integration software that is specifically optimised for handling mainframe migration projects eliminates the risks associated with using utilities or custom coding techniques. DMExpress provides the ability to handle massive data volumes, deep support for mainframe data, and support for the variety of data migration scenarios that are likely to be encountered in mainframe modernisation projects including rehosting and long running projects where bi-directional support and data protection is required.

Typically, the same transformations performed with the mainframe application must be performed with the new application running in the open systems environment. A critical component of a successful implementation is the ability to run the same data transformation functions (sorts, joins, etc.) with the highest levels of performance and scalability, and with a minimum amount of development resources. The broad feature set and extreme performance of DMExpress enables organisations to confidently reproduce the equivalent data integration processes that were executing in the mainframe environment with speed and scale.

**CUSTOMER EXAMPLE**

A large international telecommunications company was consolidating multiple mainframe-based PeopleSoft financial applications, and migrating them to run in a Linux environment. Their data integration requirements included merging the data from multiple source systems, replicating the functionality from the mainframe applications in the new Linux environment, improving the overall performance of the applications, and minimising the impact on end users associated with the migration. The company had expected that the Linux system sort would match the feature set and at least match the performance they were accustomed to on the mainframe, but quickly realised that the Linux system sort was too slow for sorting the data to meet their needs. They also determined that Linux system sort had feature limitations that were impacting their applications, including the inability to filter, or to total amounts when summarising the data.

The company deployed DMExpress into the Linux environment for data integration. DMExpress capabilities that were utilised included:

- Sort
- Aggregation with Total

Additionally, DMExpress enables the company to view the binary data from within the application directly, allowing them to map the data types and view the actual number represented in the bytes. This helps verify the data mappings and helps troubleshoot data problems without having to write code to temporarily convert data. This sampling also allows the company to view the EBCDIC data in ways that UNIX utilities do not.

With DMExpress running in the Linux environment, the company is able to replicate the functionality of the mainframe, at lower cost, and with better performance. The Linux based processing of the data integration tasks — which includes sorting, summarising, and aggregating data, as well as splitting files into large numbers of target files — is **8 times faster than on the mainframe.**
Deep Support for Rehosting of Mainframe Applications

Many mainframe modernisation projects involve rehosting the mainframe application to run in an open systems environment. Rehosting allows companies to recompile their existing source code, usually with little or no modifications, so that the applications can run on open systems-based operating systems, saving time and development resources. Rehosting enables organisations to quickly migrate applications off mainframes, even for applications that will undergo a full rearchitecting effort at some future date.

DMExpress provides deep support for rehosting environments. This support enables organisations to use DMExpress’ graphical development environment to quickly develop new open systems-based job steps to sort, merge, join, filter, summarise, aggregate, and reformat their data, and to easily integrate these steps directly into the new JCL.

In addition, depending on the data types chosen, rehosting mainframe applications may require data translation processing. DMExpress provides built-in support for the various rehosting vendors’ equivalents of mainframe data formats, to enable DMExpress to read and write these data file formats directly.

CUSTOMER EXAMPLE

One of the world’s largest food and beverage companies is migrating thousands of mainframe jobs to Windows, to run under the Micro Focus Enterprise Server. There are literally thousands of jobs involving sorting and COBOL programs that are required to run accurately and efficiently in the new environment in order to meet the SLAs on the Windows platform. Typical sort jobs involve hundreds of GB of data per job. The customer is using DMExpress tightly integrated within their Micro Focus rehosting environment to successfully run all of the data transformation processing in the Windows environment, decreasing processing times as much as 93% (80% on average) over mainframe processing, and 82% over the default capabilities in the Micro Focus environment. The Micro Focus implementation team had specifically recommended that the customer consider DMExpress due to DMExpress’ tight integration with the Micro Focus Enterprise Server environment and its extreme performance characteristics.

“A critical component of a successful implementation is the ability to run the same data transformation functions with the highest levels of performance and scalability, and with a minimum amount of development resources.”
**Bi-Directional Support**

In many projects, bi-directional support (EBCDIC to ASCII and ASCII to EBCDIC) is required. This may be due to a long running migration in which the application is running on both platforms for a lengthy period of time, requiring the application data to be synchronised until the mainframe application is decommissioned. Another scenario requiring bi-directional support arises when the output from the modernised mainframe application must be read back into other mainframe applications. In these cases, the new application’s output data (in ASCII format) must be transformed to EBCDIC before it can be read into the mainframe applications. DMExpress provides built-in support for converting data from EBCDIC to ASCII, and for converting from ASCII back to EBCDIC.

**Support for Changed Data Capture Processing**

Changed Data Capture (CDC) processing is a data integration technique in which an older file and a new file are compared in order to identify changes that have occurred. Only data that has changed is updated in the master file. This type of processing is often useful for data migration projects, since the bulk of the data can be migrated in advance, and just the updates can be processed on an ongoing basis until the legacy application is decommissioned. DMExpress is often used to perform changed data capture processing in addition to the initial batch processing of the full data sets for data migration projects. The CDC process to update a table with millions of records of data, for example, can take just two or three minutes to complete with DMExpress, making ongoing, rapid updates to the open systems based application practical.

**Embedded Data Protection Capabilities**

Long running modernisation initiatives can take weeks or months to complete, and may require processes to back up the data periodically and to restore the data as needed. Embedded data protection capabilities for the target data files and any intermediate files minimise risk of data loss during these modernisation initiatives.

To protect data on modernised systems, Syncsort provides block-based data protection capabilities, which are extremely efficient at capturing data and storing it on disk with a minimal disk footprint. They also provide immediate access to the data via direct server mapping to snapshot images. Even multi-terabyte data sets can be returned to production within minutes. Sophisticated integration with server virtualisation platforms provides additional fast recovery models, including the ability to restore a failed physical system as a virtual machine within minutes.

“A large food and beverage company uses DMExpress within their Micro Focus rehosting environment to successfully run data transformation processing in Windows, decreasing processing times as much as 93% over mainframe processing.”
“DMExpress provides built-in support for converting data from EBCDIC to ASCII, and for converting from ASCII back to EBCDIC.”

CUSTOMER EXAMPLE
A manufacturing company had been running approximately thirty production applications on the mainframe, consuming about 800 MIPS every year. In order to reduce costs, this company took on an initiative to migrate all of their applications off the mainframe onto IBM AIX workstations, recompiling their COBOL applications to run as IBM COBOL for AIX applications. They plan to sunset all of their mainframe processing over time.

They are using DMExpress to convert the flat files and VSAM files from their production mainframe applications into ASCII, and then transform the data into the appropriate formats to load into the AIX-based applications.

In addition, the company is converting their mainframe JCLs to run in a rehosting environment in AIX. They are using DMExpress to perform data transformation processing in the AIX environment in conjunction with the new applications on a production basis, replicating the data transformation functionality from the mainframe environment, and achieving greater performance than was experienced on the mainframe.

They also continue to receive mainframe files from their customers, and they use DMExpress to convert these files to ASCII format, and transform and load this data into their new systems. Finally, they use DMExpress to translate and transform the output files back to EBCDIC to send back to their customers in the required mainframe formats.

CUSTOMER EXAMPLE
In order to reduce mainframe MIPS and associated costs, a major international investment bank is rehosting their legacy mainframe-based equity trade settlement application on Linux servers. Due to merger and acquisition activities, they plan to consolidate multiple trade settlement applications and data into one application, which will grow their data sizes in the application by 400%.

With their previous data transformation capabilities, they projected they would be unable to process the combined data each night before the next day’s trading began. To alleviate this data performance bottleneck, they implemented DMExpress within the rehosting environment, successfully processing all of the data 700% faster each night.

As a result, the bank has reduced costs by running on Linux servers instead of on the mainframe, and has eliminated the risk of missing SLAs to the business, even as data volumes grow.
**Conclusion**

Multiple business and technology drivers are motivating companies to modernise or re-platform their mainframe applications to run on open systems such as Windows and UNIX. These mainframe application modernisation initiatives are often complex and include many phases. Effectively handling the data translation and transformation processing comprises a critical and challenging set of requirements. DMExpress, based on an extensive history of mainframe data integration experience and expertise, is the most functionally capable and performance orientated solution for the various requirements associated with mainframe application modernisation initiatives. To learn more or to schedule a consultation, call +44 (0)1732 849000 or visit www.syncsort.com.

“To alleviate data performance bottlenecks, a major bank with extreme data growth implemented DMExpress within their rehosting environment, successfully processing the data 700% faster each night.”