DECOMMISSIONING MAINFRAMES... SIMPLIFIED

AN ARBUTUS SOFTWARE SOLUTION PAPER

CONTENTS

EXECUTIVE SUMMARY

This solution paper describes the challenges inherent with mainframe decommissioning projects, and the technologies and tactics that have been proven to make large projects such as these run more smoothly and cost effectively.

Decommissioning a mainframe computer is a major undertaking, there's no question about it. However, a variety of new tools exist that can make the process less difficult, less costly, and far faster to achieve than you might have thought possible. For example, query tools by Arbutus can access and read your mainframe source data directly, without requiring ETL programming or staging the data, regardless of the accuracy or reliability of your source documentation, or your access to expert resources.

...access and read mainframe source data directly, without requiring ETL programming or staging.

Arbutus technologies can assist you with quantifying and addressing data quality as a by-product of query, and help you discover or confirm your business rules through query, rather than documentation. You can also avoid unnecessary ETL with Arbutus, and easily archive any necessary mainframe data on inexpensive PC-based computers, before turning off the switch. Archiving the data on PCs makes possible ongoing query access to your mainframe legacy data, thereby helping you remain in compliance with regulatory requirements governing your historical legacy data.

Look for this symbol: » throughout for 10 easy-to-use Arbutus commands or functions that will help you perform sophisticated data analysis with just a few clicks.

WHAT ARE THE CHALLENGES?

Decommissioning a mainframe is like an extreme data migration project, because once you turn off the mainframe, there's no going back; you've lost access to its data forever. Decommissioning therefore shares many characteristics with data migration projects. Leaving aside the hardware issues inherent in such a project, some key challenges of decommissioning a mainframe include:

- Reading source legacy data stored in difficult to access or arcane formats
- Addressing missing or out-of-date source documentation, or having limited access to subject-matter experts
- Detecting and eliminating duplication with existing systems (especially as a result of M&A)
- Identifying data quality issues, including uncovering unknown or possibly suspect data
- Reconciling conversion results back to the source data
- Satisfying regulatory requirements to maintain historical data and keeping it accessible for a number of years, especially when data is omitted during system conversion

Many organizations attempt to utilize the perceived cost savings of an an in-house, proprietary technology solution to address the issues above. However, the hidden costs of protracted timelines, increased staff resources and less than acceptable data quality often make this a poor decision.

...query all defined mainframe data - with no ETL programming.

HOW CAN ARBUTUS HELP?

Arbutus technologies can not only solve all the issues listed above, they can also dramatically reduce the timelines and costs involved with such a project:

- **Arbutus Servers –** Native servers for zOS/MVS mainframes (zSeries) and AS/400 (iSeries) can directly access and define virtually all existing legacy data (including VSAM, QSAM, IMS, DB2, and ADABAS). Where possible, they can also make use of available data definitions (such as in COBOL, Easytrieve, PL/1, etc.) for automatic data definition generation. Arbutus Servers are integrated, so the full suite of historical mainframe legacy data can be easily extracted to an Arbutus Windows Server for future accessibility.

- **Arbutus Analyzer –** Lets you quickly and easily query all defined mainframe data – with no ETL programming – to quantify and identify data quality issues, validate documentation and uncover undocumented data characteristics.

- **Arbutus LegacyLink™** - Provides direct ODBC access to all defined or staged mainframe data directly from end-users' preferred applications, while respecting all security rights and issues. For example, you can easily read VSAM, IMS, and other legacy data with Excel or existing data profiling tools.

A TYPICAL DECOMMISSIONING PROJECT

The first step in any data migration project is to know your data. This often involves a long and costly process of reviewing the source documentation. Based on this review, the source data is then mapped to the target system.

During this process, assumptions are made about the quality of the documentation and of the data itself. These assumptions are then used to prepare a budget for the remaining process. Of course, this process can be severely compromised when the quality of the documentation is poor, or when you have limited access to subject matter experts.

Next, ETL processes are created based on the review and assumptions. This is typically the first time you get to see the data, so it is only at this point that you know whether your earlier assumptions bear any resemblance to reality. Unfortunately, countless "surprises" are typically discovered at the ETL phase, necessitating multiple iterations of the ETL processes, until the result is satisfactory.

...validate the parts of the documentation that are critical to the process.

HOW IS AN ARBUTUS PROJECT DIFFERENT?

The overall process governing a mainframe decommissioning project is similar with Arbutus, but the focus, especially early on, is different. Because Arbutus can query the source data so easily, source documentation is still reviewed, but at this point it is considered suspect, as we let the data "speak for itself".

Early query access helps you validate the parts of the documentation that are critical to the process. This involves identifying data that does not fit with the documented model and using these exceptions to either adjust the model or classify the data as an error.

Error identification is a critical part of early querying. By quantifying error rates and categories early on, a much more informed project schedule can be created. More importantly, by identifying specific data quality issues at the outset, you are able to create processes to address these issues.

THE 4 I's of DATA QUALITY ISSUES

At Arbutus, we group data quality issues into four broad categories: invalid, improper, incomplete or inconsistent. With Arbutus technology, you can test the full range of data quality problems in each of these areas.

- **Invalid data** - flat-out wrong data, this type of data is supposed to be packed but is actually text (or garbage). Often occurs as a result of a system error, but may also indicate an overloaded field or other undocumented characteristic.

  - **Verify** - *this command can be used to identify invalid fields, such as date fields that are correctly formatted but contain an error (2011/17/31), or an invalid leap year (2007/02/29).*

  - **Verify All** - *use this command to identify invalid data in all fields within a file or table, rather than verifying the validity of individual fields separately.*

...easily validate your organization's unique business rules.

- **- indicates a command or function in Arbutus**

- **Improper data** - technically correct data that violates a business rule, such as a blank field, or data that is improperly formatted (e.g., product numbers that are not entered in a standard format, like 99-9999999).

  In contrast to invalid data, business rules are situation-dependent, so it may be ineffective or impossible to run a single command that identifies all exceptions. Instead, Arbutus offers query capabilities that can easily validate your organization's unique business rules:

  - **Quick Filters** - *identify a characteristic in the data, then isolate anything that is like, or unlike, that item, for example: AMOUNT<0 or CITY="NEW YORK"*

  - **Format** - *identify improper data. For example, setting the filter FORMAT(SSN)<>"999-999-999" identifies all items that violate the business rule for valid formatting of social security numbers.*

  - **Classify** - *identify every unique instance present in the data. Ideal for easily validating any coded information, such as transaction types, location codes, product numbers or department numbers. Can be used to identify data quality issues and highlight incomplete documentation.*

...easily test any business rule for incomplete data.

- **Incomplete data** - data that is expected to form a numerical sequence, like invoice numbers, but does not. Although technically not the same thing, duplicated items are usually included in this category, such as missing or duplicated invoice numbers. Another common example in this category is key fields, which may not necessarily be sequential, but should rarely be duplicated.

  - **Gaps on Check_No** - *identify all missing check numbers*
  - **Duplicates on Customer_No** - *identify duplicated customer keys*

- **Inconsistent data** - covers all multi-element rules, whether at the record level (e.g., field 'total' should be field 'quantity' times field 'price', or as found in Arbutus: **quantity*price**), or at the file level (e.g., record type '2' should follow record type '1'). It can also span multiple files, such as missing foreign keys. Inconsistent data is a special type of improper data, because it violates a certain business rule.

  Arbutus's query capabilities allow you to easily test any business rule for inconsistent data. While the **quantity*price** example above is straightforward or common, Arbutus' Expression Builder offers over 90 data manipulation functions, and supports expressions of any complexity. Tests may also have multiple values based on defined criteria, so a field might have multiple different calculations, depending on the characteristics of the transaction.

  - **Join** - *perform an unmatched join to determine missing or unused keys between one related table and another. For example, every employee in the current payroll file should have a matching record in the employee master database table. An unmatched join would identify employees as paid that are not in the employee master table (i.e., are invalid or have not yet been entered in the employee database).*

DATA 'ERRORS' CAN YIELD SYSTEM INFORMATION

Not all "data quality issues" are, in fact, data quality issues. In some cases, you are actually discovering an undocumented characteristic, such as an undocumented transaction type or an overloaded field. In an overloaded field, the documentation might indicate that the field contains the date of the customer's PO, but in certain situations, it actually contains the customer's account number. A standard test for date validity with Arbutus detects these errors, helping you refine your understanding of the source data.

SOURCE DATA QUERY REDUCES ETL PROGRAMMING

One very important opportunity presented by source data query is to reduce your ETL programming.

Business processes change over time, resulting in changes or additions to existing systems and their documentation. In the case of legacy systems on mainframes, documentation may suggest various transaction types or formats are valid, when in fact some are accurate, some are partially accurate, and some haven't been used for decades.

Rather than creating a transform for each documented type of transaction or data, Arbutus allows an ETL programmer to easily identify whether that type of data even exists in the files to be migrated.

Why create a complex transform for a type of data that doesn't even exist?

Perhaps there are only seven instances in the file; it may be more expedient to export and then manually re-enter them. Arbutus offers specific capabilities to help identify the data that is actually present.

...easily identify whether a type of data even exists in the files to be migrated.

Earlier, the Classify command was used to identify unexpected transaction types and incomplete documentation, but it can also be used in the reverse manner.

- **Classify** - *identify transaction types missing from the data, and dramatically reduce your ETL.*

ETL MADE EASY

Rather than creating custom ETL processes, why not let Arbutus transform your data? You already have direct access to the source data, and Arbutus offers a full range of data transformation capabilities. This can be especially valuable in decommissioning projects, as the ETL processes will not be required once the conversion is complete.

With Arbutus, you can take any of the source data and re-format it to meet almost any need. If the target is a relational database system, Arbutus offers a variety of techniques to not only transform, but also to load that data.

RECONCILE THE MIGRATION

An important component of any migration project, including decommissioning, is to verify that the process has executed correctly. This is particularly important for decommissioning projects, as there is typically no going back.

Because Arbutus has the ability to access both the source and target systems, it is the logical choice to confirm the correctness of the result. In fact, that checking can be built right into the processes.

...you can even query the staged data as if it were still on the mainframe.

- **Verify Results** - *verify fields in total based on a key value, like customer totals, or at a detailed level, such as with individual entries.*

STAGE YOUR MAINFRAME DATA EASILY

Once your migration is complete, the decommissioned system must finally be turned off. There is always some concern at this point, as there is no going back. If some data was not transferred, but is needed in the future for any reason (perhaps regulatory or business) then it may be extremely difficult, or even impossible, to retrieve this data.

> *What you need is to archive your mainframe data, at a reasonable cost, without requiring the mainframe.*

Arbutus Servers are seamlessly integrated, so it is as easy as a few clicks to transfer an entire data set from your mainframe to the Arbutus Windows Server. The Arbutus Windows Server is fully compatible with all mainframe data and data types (e.g., EBCDIC, PACKED, ZONED), so no ETL programming is required to stage your data. This is true whether your data is in DB2, IMS, VSAM, or any other file formats found on your mainframe.

> *With Arbutus, you can archive the entire data file, so you can go back whenever you want to review any data.*

You can even query the staged data as if it were still on the mainframe. This presents a very powerful opportunity, as it may mean that you do not have to transfer as much historical information to the new system. Since historical queries can easily be fulfilled from the staged source data, you only need to transfer data that is required for the operation of the new system.

PROCEDURES AUTOMATE YOUR TESTS AND ETL

All the capabilities described in this paper can be executed on the fly, or saved as procedures within Arbutus, for repeated execution. Since Arbutus technology is based on an underlying command language, it is easy to place these commands in procedures. Also, given the 4GL nature of the language, the commands are typically very simple and English-like, making them easy to modify as your needs change.

EASE OF USE

Arbutus technology features an easy to use graphical user interface (GUI) that not only provides easy access to all the commands, functions, data types and other characteristics mentioned earlier, but it also presents data in a spreadsheet-like, table format, regardless of the source data.

Even complex VSAM, IMS and QSAM mainframe data can be viewed and queried queried as if it already existed in a spreadsheet. More importantly, this spreadsheet-like view of the data does not limit the number of allowable rows or columns, like typical spreadsheet software. Whether you have a million rows, a billion, or more, Arbutus technology is designed to handle the rigors of real, big data.

...complex mainframe data can be viewed and queried as if it already existed in a spreadsheet.

This tabular view of the data presents very powerful opportunities to understand your data because you can actually "see" the data and dynamically query it. This may be the first opportunity you have had to work with your data in such a flexible format.

STAGING COSTS OF A DECOMMISSIONED MAINFRAME

At today's prices, in 2011, you can pick up a 3-terabyte PC-compatible drive for $100. Before you store your mainframe data, you'll use the Arbutus Windows Server, which automatically compresses data 85% to 90%, achieving "ZIP-level" compression. That equates to storage costs of $100 for every 20 to 30 terabytes.

> *For the first time in history, the cost of archiving your entire mainframe data set on PC-compatible equipment is surprisingly cost-effective.*

Staged historical data can also be made available for use by business analysts or executives throughout the organization. Arbutus LegacyLink is an ODBC driver that provides read-only access to data residing on any Arbutus Server, allowing it to be exposed as an ODBC data source to targeted groups of end-users and their ODBC-compliant PC applications (e.g., MS Excel, MS Access or Crystal Reports) with a few quick clicks of a mouse.

What does this mean for your organization? Simply this:

...contact us today to set up a 30-minute web demo.

With Arbutus, end-users gain on-demand access to all necessary data - including decommissioned mainframe legacy data - within their familiar PC-based applications, whenever they need it.

THE LAST WORD

While the benefits of decommissioning the mainframe are clear for many organizations, the process and technology to achieve this objective are still laden with risk and hidden costs. Arbutus represents a simpler, low-risk, cost-effective approach that allows organizations to decommission their mainframe while avoiding the many hurdles that negatively impact these projects.

ARBUTUS

#270 - 6450 Roberts Street
Burnaby, British Columbia
Canada V5G 4E1

Toll-free: 877.333.6336
Direct: 604.456.6336
Fax: 604.437.7872
info@ArbutusSoftware.com
www.ArbutusSoftware.com

Based on 25 years of innovation excellence, Arbutus delivers the very best in purpose-built audit analytics technology to meet the exacting demands of today's business environment. Auditors, business analysts, and fraud investigators rely on Arbutus to enhance their testing, analysis, and compliance capabilities.

The data universe is wide and varied. One of our core strengths as a technology firm is the ability to easily work with all types of data, both legacy and non-legacy. Arbutus solutions allow auditors, IT, and business professionals to overcome many of their current constraints in areas such as data migration, data quality, fraud detection, and data analysis.

Arbutus Audit Analytics, our flagship product suite, is a proven solution used by auditors, business professionals, IT, and management all over the world. With outstanding customer service, strong product support, and flexible licensing, Arbutus Software offers the best value for advanced data conversion, migration, and analysis solutions.

Arbutus Software Inc. is a privately held company based in Greater Vancouver. For more information about our company or products, please contact us.