ZERO ETL

HOW TO MINIMIZE THE ETL RISK OF MAINFRAME LEGACY DATA

AN ARBUTUS SOFTWARE WHITEPAPER

Organizations with mainframes continue to face the escalating costs of dealing with legacy data and systems, which present unique challenges and risks. This paper outlines newer technologies that minimize the challenges, risks and most importantly, the costs associated with Mainframe legacy data.

New agile data access and query technology allows you to work with your mainframe and other legacy data with essentially no ETL. The ability to have immediate SQL and query access to your VSAM, IMS and other legacy data opens up a number of powerful opportunities. These include:

Legacy data... a 'perfect storm' of risk

- Giving your existing profiling and other tools direct access to your legacy data
- Quantifying data quality before planning starts
- Identifying data quality issues for correction
- Staging your data easily, to minimize mainframe impact
- Confirming source documentation
- Querying source legacy data to better understand it, and improve mappings
- Validating requirements by prototyping a proposed data mart with live data

The technology allows you to query and profile your source legacy data during the planning stages. The result is fewer surprises and a dramatic lowering of risk.

INTRODUCTION

Most of us labor under the certainty of belief that data fits into two broad categories, Legacy and Relational. Relational data is stored in SQL databases and is generally easy to access. Legacy data, on the other hand, inevitably involves a significant ETL investment before you can work with the data. In the legacy category we often include mainframe sources like VSAM, IMS and flat files, but it is often a catch phrase for any "older" data.

Aside from being perceived as difficult to access, the other typical characteristic of legacy data is that it has less built-in quality controls. Of course, if the very data that is most problematic is also the most difficult to access, then you have a "perfect storm" of risk.

The first category of risk can cripple a project, depending on the severity of the data quality issue. The latter category, at a minimum, can extend the timelines as you strive to address this unexpected data.

The bottom line is that for many project managers, legacy data is synonymous with risk, complexity and cost.

THE SOLUTION

What if the commonly held belief about the necessity of up-front ETL was wrong? What if legacy data was as easy to access as data in Oracle, or other relational databases? It is easy to see that this could completely change the structure of projects involving legacy data.

The biggest change could be to implement a comprehensive data profiling approach; an approach that includes all the mainframe, legacy and other data sources. Minimize your reliance on assumptions or suspect documentation. Instead, generate as much metadata as possible directly from your source data.

This paper puts forward the concept of "data centric" planning. Project Managers can use data profiling technologies to read the source data from the most complex file structures encountered in mainframe and other legacy systems. While there are certainly files that can't be read, these are becoming a rare exception.

INSTANT WAREHOUSE: A ZERO ETL TECHNOLOGY

Instant Warehouse is technology that bridges the gap between legacy data and SQL standards. It allows you to include all your data sources, including legacy data, during the planning process. Legacy data is perceived as being complex, or not standards based. While true, it is only its physical characteristic. The data itself is no more complex than any other, and Instant Warehouse bridges the standards gap. All of your legacy data can be presented in a tabular format, for use with our own tools or through standard SQL queries.

At a minimum, Instant Warehouse can extend your profiling capabilities well beyond what you have come to expect, but that is only the start. A more complete list of benefits includes:

- Avoid costly ETL programming.
- Integrate data from disparate platforms.
- Implement business rules quickly.
- Discover and quantify data quality and other issues.
- Stage your data easily. Instant Warehouse can stage any mainframe data on Windows with the click of a mouse
- Utilize other profiling technologies and tools, without the need to ETL
- Expose all of the source data (whether legacy or otherwise) as a SQL database.

With this new approach, organizations access their source data directly, hours or even minutes after the request is received. They can you apply a wide range of data quality and profiling technologies, and can have much better information with which to plan projects.

Increase "buy in" to the project

Since the source data can be made available as a virtual database, the project team can even let the end users themselves have read-only access the source data. End users can read the data with their tools of choice, be they Excel, Access, Crystal Reports, or other tools. Not only will this increase "buy in" to the project, but with more end user eyes on the data you are much more likely to discover the rules and anomalies in the data, before you build a data warehouse.

In addition to merely exposing the source data, the project team has the ability to control the semantic presentation of the data and also to transform the data dynamically. This allows them to prototype a proposed data mart before it is built. The users can experience, or even define, the solution during the planning stages, and do this with the full set of real data. This capability provides a powerful tool to validate and refine the specification before work begins.

CONCLUSION

Legacy data is traditionally considered one of the highest risk aspects of a DW project. This need no longer be the case. The result can be simpler projects, improving your chances of success and improving timeliness. The result is to simultaneously reduce both risks and costs. These benefits accrue while adding more flexibility to the process.

THE AUTHOR

Grant Brodie, CA, is President of Arbutus Software Inc. He has over 30 years experience in mainframe data access, query and data analysis. Grant's career before Arbutus was spent in computer audit, first with the Deloitte organization, and then founding ACL Software, a leading audit software firm.

#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.