Treehouse
SOFTWARE

White Paper

Hot Topic or Hot Potato?

Enterprises need a highly fault-tolerant way for heterogeneous applications and databases to act in a coordinated fashion to accomplish business goals

Introduction

Enterprises are constantly demanding more from their information systems—and creating more data—than ever before. To stay agile and responsive to new or imposed business requirements, emerging technologies such as cloud computing, analytics, social media, and mobility, IT organizations must find a way to "unlock" mainframe data so that it can integrate with open systems (RDBMSs like Oracle, SQL Server and DB2 on other platforms).

Most of the leading-edge integration platforms being offered today fail to cope adequately with the real world of the mainframe: performance-sensitive workloads, cost containment, 24x7 operation with carefully-scheduled maintenance, and above all the proprietary nature of legacy mainframe databases. Direct access to these data sources is not for the faint of heart.

This white paper delves into the issues of why a well-architected, comprehensive, robust and scalable replication solution is the key to enabling legacy databases to exchange data reliably and effectively throughout the enterprise. This Makes it possible for IT and business users to access corporate data regardless of where it resides. Furthermore, effective enterprise replication can be combined with other techniques to overcome technology constraints and maximize IT effectiveness.

Hot Topic or Hot Potato?

Large enterprises have IT infrastructures consisting of multiple applications and data systems spread over a variety of different platforms. This heterogeneous environment has evolved over the years and is often not interoperable. Driven by this paradigm and by further globalization of the worldwide markets, companies face rapidly-changing business requirements which lead to

Editor's Note

At Treehouse Software, we cite "data replication" as one of our key competencies. But the term—like so many others in the IT arena, including some that are mentioned in this white paper—suffers from lack of clear definition. In the early days of our flagship ADABAS-to-RDBMS solution, tRelational/DPS, we coined the terms "materialization" and "propagation" to refer to what are often dubbed "ETL" (Extract-Transform-Load) and "CDC" (Change[d] Data Capture). We have also characterized our multiple offerings in this space as "data migration" solutions, though we prefer "data transfer", since many of our customer implementations involve ongoing coexistence of source and target systems rather than one-time migrations. At times we have used "data synchronization", and even made that the basis for naming our DPSync product.

Searching the Internet for "data replication" not only yields results for Treehouse and our competitors, it may also return information on disk-mirroring hardware and the like. Today we offer customers robust and comprehensive solutions that move data between a multitude of heterogeneous sources and targets, unidirectionally or bidirectionally, in bulk or in real time, within and across mainframes, open systems and Windows platforms throughout the enterprise. Until somebody comes along with a better idea, and for the sake of clarity within this writing, we'll continue to refer to these as "enterprise data replication" solutions.

demands for new solutions in data integration, supporting functionalities such as high availability, seamlessness across platforms and the requirement for up-to-date information. Data exchange in a heterogeneous IT infrastructure means harmonization of different data formats and data models for data exchange solutions. Very often, this data exchange is a complex and tedious task that represents a major cost factor. Data exchange is also time-sensitive and critical, hence reliability and auditability of all data movements is important.

But this exchange must be done in a controlled manner, as the new generation of data exchange is more complex than ever. Current methods for data extraction and transfer between databases do not meet time and cost requirements owing to myriad issues, including:

- Increasing data volumes;
- Limited batch windows;
- Mainframe solutions (from IBM and other vendors) are expensive and complicated;
- Performance and CPU burden on mainframe system.

Vendors come and go with spectacular-sounding ideas to solve this problem, but these ideas do not always scale, or cannot always cope with the real world that enterprise IT departments inhabit. From our experience and unique perspective, mainframe cycles are still an expensive resource to be managed carefully. Mission-critical databases are supporting mission-critical operational applications, and can't be burdened with complex, performance-killing ad-hoc queries and "data exploration". Platforms and applications require periodic maintenance, necessitating outages. What customers need is a highly fault-tolerant way for heterogeneous applications and databases to act in a coordinated fashion to accomplish business goals.

These days, "data virtualization" is a hot topic. The objective of data virtualization is to abstract data sources across the enterprise to present the client (data consumer) with access to logical, composite "business entities" whose technical implementation—location, structure, access method, etc.—are deliberately obscured. This is a noble and lofty goal, but it is really just an extension of the concept of "data federation"—which has been with us for at least two decades. Federated databases are those where middleware enables clients to access heterogeneous data sources transparently, e.g., to issue a SQL query which JOINs mainframe IMS/DB segments with Oracle tables on Linux. This often requires a metadata/management layer to manage and collate field/column names, proprietary datatypes, etc.

Improvements in processor speed, cost/performance ratios and software sophistication have made data virtualization more capable than data federation used to be. Nevertheless, knitting together disparate data sources and attempting to harmonize them into seamlessness is subject to the age-old principle that the system is only as robust as its weakest component. Thus, any suboptimal aspect of the interface to a given data source tends to render the whole system brittle. When it works, it can work very well; when it breaks, it's a nightmare.

Pitfalls of Integrating Legacy Mainframe Databases

Let's take the example of a popular nonrelational legacy database, Software AG's ADABAS. Thousands of mainframe sites worldwide run high-performance, bet-your-business applications using ADABAS as the back end. ADABAS, which has been around for over 40 years, manifests an interesting feature: the result set returned to a query may contain data that has been updated but not yet committed—indeed, updates that may never be committed if they are subsequently backed out. Therefore, a client issuing a query that does not take this possibility into account may receive "invalid" results—results that, when issued only a short time later (after the backout), would be different.

Modern relational databases offer configurable "transaction isolation" to help assure accurate-to-the-client and consistent query results, but ADABAS has no concept of transaction isolation. This is the way the database has always worked.

A robust data replication system, such as is offered by Treehouse Software's tcVISION, avoids such issues by synchronizing each data source with a readily-available, consistent target database, first by doing a bulk load from source(s) to target and then by replicating only changes—only committed changes—from source(s) to target. So there can never be ambiguity as to whether a query against the target database involves uncommitted data.

By definition, data replication necessitates redundancy and increased storage costs. A couple of decades ago, storage was limited and expensive; today—with everyone talking about "Big Data"—storage is abundant and cheap.

Data replication solutions over the years have been criticized for being inconsistently implemented and difficult to manage. This was mainly a function of the tools used—and of the tools not used. We have long held that "our real-world customers are our biggest competitors", in that we often find that customers have done in-house development of their own data replication systems. Certainly, there is nothing magical about such systems; our customers employ skilled developers who are fully capable of creating them. And often the initial implementation works well, and becomes entrenched such that other solutions are resisted even when the in-house system must be extended beyond its original capabilities and starts to break down.

These days, choosing to implement a commercial data replication software system is a "no-brainer", as long as it is one that comprehensively covers—efficiently, natively and scalably—the full gamut of data sources and targets that the organization might reasonably contemplate using. There is enormous benefit to leveraging a single solution to consistently design, implement and manage replication scenarios and processes. Data replication becomes simply another service offered by the professional IT center.

Getting the Best of Both Worlds in the Real World

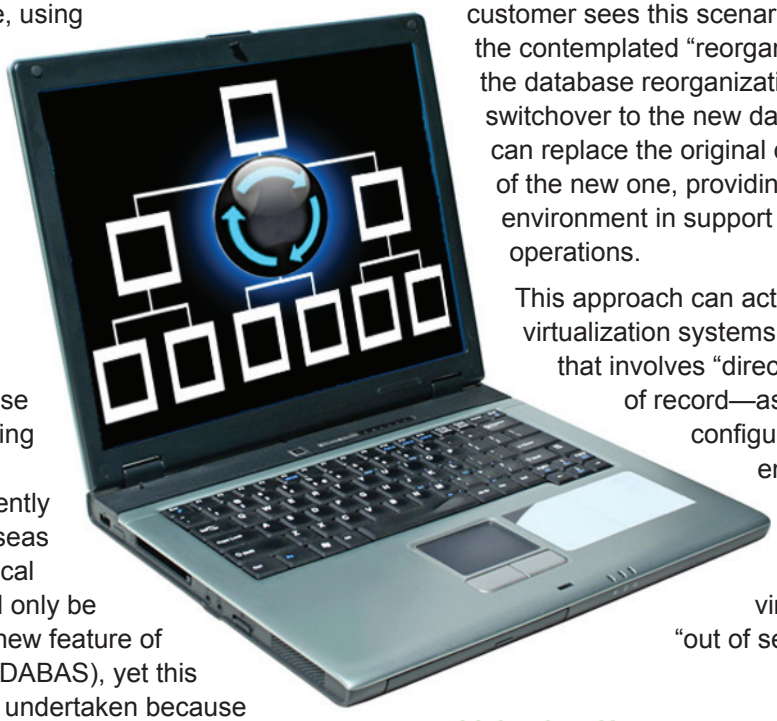
Savvy IT organizations also recognize that data replication and data virtualization do not represent an either/or choice. Hybrid implementations can be effective: virtualize directly to the system of record where that can be done readily and manageably (and with acceptable performance impact on the source system), and virtualize to a replica where direct access is problematic or the workload is better handled on the replica.

Note that well-architected replication solutions not only incur minimal overhead on the source system, they also enable the "heavy lifting" of data transformation and application to target to be placed on the platform where capacity is most available and manageability is maximized. For example, tcVISION can be configured to replicate from a variety of mainframe data sources with only the change-capture process being executed on the mainframe and all other processing done in a "lower-TCO" environment.

We were recently working with a customer in the real world that is phasing out legacy mainframe systems. At the same time, however, the customer wants to move to a Service-Oriented Architecture (SOA). The two initiatives were in conflict, because exposing mainframe-based processes as services would have increased mainframe capacity needs, necessitating an unacceptably-costly upgrade. However, by replicating the mainframe data to a Microsoft SQL Server environment, services could be exposed and the processing contained to the Windows environment—where, ultimately, they would need to reside anyway, after the decommissioning of the mainframe. (This customer scenario actually added another layer of complexity, since some of the services being contemplated were read/write in nature, meaning that updates would also have to flow from SQL Server back to the mainframe in real time. Fortunately, such a bidirectional synchronization scenario is fully and easily supported by tcVISION.)

Creating New Opportunities by Leveraging Homogeneous Replication

The foregoing discussion has, to a degree, implicitly focused on "heterogeneous replication", synchronizing data from one data source to an unlike target, e.g., from CA-IDMS to DB2 LUW. However, many large enterprises have been doing "homogeneous replication" for years in synchronizing data to a remote "hot" or "cold" disaster recovery/high-availability site, using products like E-Net Corp.'s Remote Recovery Data Facility (RRDF) and Enterprise Data Replicator (EDR) for mainframe data sources.



Today's sophisticated data replication products provide both heterogeneous and homogeneous replication capabilities, expanding the use cases and ultimately increasing the value of the customer's investment in them. We recently visited with a customer overseas that faced a conundrum: critical business requirements could only be satisfied by implementing a new feature of the database (in this case, ADABAS), yet this implementation could not be undertaken because it would necessitate a complete reorganization of the entire database—estimated to take weeks in duration, during which the database could not be used! Obviously, such an outage would not be tolerable. Yet, using a solution like tcVISION, database files can be migrated to a replica—a new ADABAS database, having the new feature implemented—in phases, and synchronized using replication techniques, with the process continuing for as long as necessary until all data are migrated and only a brief interruption to switch systems (a "zero downtime" cutover) can be executed.

Many data centers are facing the same demands for high availability. Using homogeneous replication involving a separate machine partition, or even another machine, enables extended maintenance windows to be scheduled and fault tolerance to be provided. In fact, our customer sees this scenario as a further benefit to the contemplated "reorganization replication": after the database reorganization is completed and the switchover to the new database is complete, they can replace the original database with a replica of the new one, providing a parallel failover environment in support of establishing 7x24 operations.

This approach can actually lend robustness to virtualization systems, in that a virtualization that involves "direct interface" to the system of record—as long as it can be configured to switch to the "mirror" environment dynamically and quickly—alleviates the possibility that the service using the virtualized data may find it "out of service".

Living in a Heterogeneous Real World

IT organizations must continue to embrace the mainframe and legacy databases for the indefinite future, and must harmonize these in an effective technology architecture. A comprehensive and flexible replication solution, such as Treehouse Software's tcVISION, that addresses the real-world aspects of enterprise computing should be the centerpiece to such an architecture.

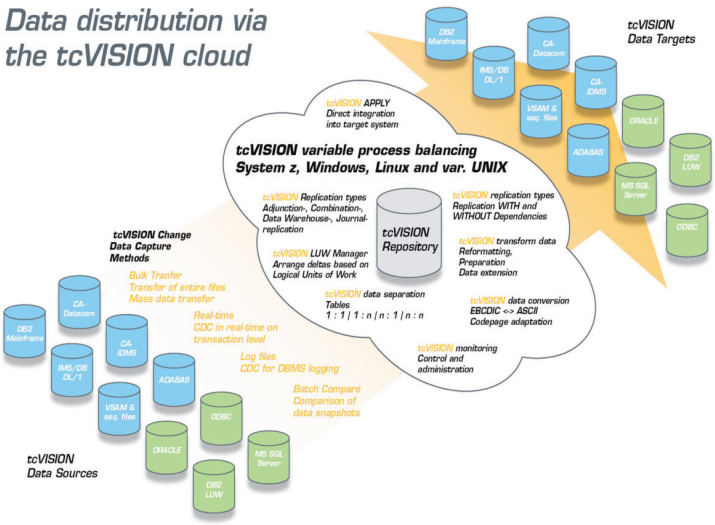
tcVISION – Intelligent, Guaranteed, Efficient and Transparent Data Exchange

tcVISION is an innovative, state-of-the-art solution to migrate, replicate, distribute and synchronize data across the enterprise, featuring a variety of mechanisms to capture and apply the data to the target systems—with efficiency, transparency and auditability.

The tcVISION philosophy: Move data – as much as necessary, as little as possible

tcVISION's replication methods focus only on changed data, reducing the data volume actually transferred. With tcVISION, changes at the data source level (Insert, Update, Delete) are automatically detected and transformed for data exchange according to specified requirements. Data exchange in a heterogeneous environment is efficient and simplified with tcVISION, whether the purpose is synchronization, replication, distribution or migration. tcVISION covers all aspects of data exchange and supports bi-directional replication of data in real time, or on a scheduled or event-driven basis.

Data distribution via the tcVISION cloud



tcVISION supports data exchange between mainframe-based systems like IMS/DB and DB2, between mainframe and open-systems servers like DB2 and Oracle, as well as within an open-systems environment (e.g., DB2 LUW to SQL Server or Oracle). tcVISION's unmatched array of Change Data Capture methods includes "Loopback Suppression" for bi-directional updates, ensuring data integrity and replication efficiency between the data sources and targets: changes received from a source system and applied to the target are not unintentionally propagated back.

tcVISION Advantages and Benefits

- Substantial improvements and faster implementation for all data integration activities.
- High efficiency for the usage of data across system borders because only changed data is transferred.
- Highest possible degree of up-to-date information used in decision making processes through real-time synchronization.
- Quicker response and maximum agility for new business requirements.
- Open design for faster implementation of new technologies and systems.
- Significantly shorter project cycles for data migration or ETL projects as no programming is required.
- Project autonomy for ETL processes because mainframe data can be integrated with no programming and no mainframe knowledge.
- Reduced mainframe workload due to workload-optimized processing.

For more information on tcVISION, visit our website at www.treehouse.com/tcVISION or contact a Treehouse Software account representative for a demo.

Treehouse SOFTWARE

About Treehouse Software, Inc.

Since 1982, Treehouse Software has been serving enterprises worldwide with industry-leading software products and outstanding technical support. Today, Treehouse is a global leader in providing data migration, replication and integration solutions for the most complex and demanding heterogeneous environments, as well as feature-rich, accelerated-ROI offerings for information delivery, business intelligence and analytics and application modernization. Treehouse Software customers are able to:

- REPLICATE Data Anywhere
- INTEGRATE Data Everywhere
- MODERNIZE Data from Legacy Applications
- ANALYZE Data for Business Advantage

With unmatched comprehensiveness of tools and depth of experience, Treehouse Software applies proven approaches to help customers and partners mitigate risk and profit sooner from modernization benefits.

For more information, contact us:

Email: sales@treehouse.com

Phone: 1.724.759.7070

Treehouse SOFTWARE

2605 Nicholson Road, Suite 1230
Sewickley, PA 15143