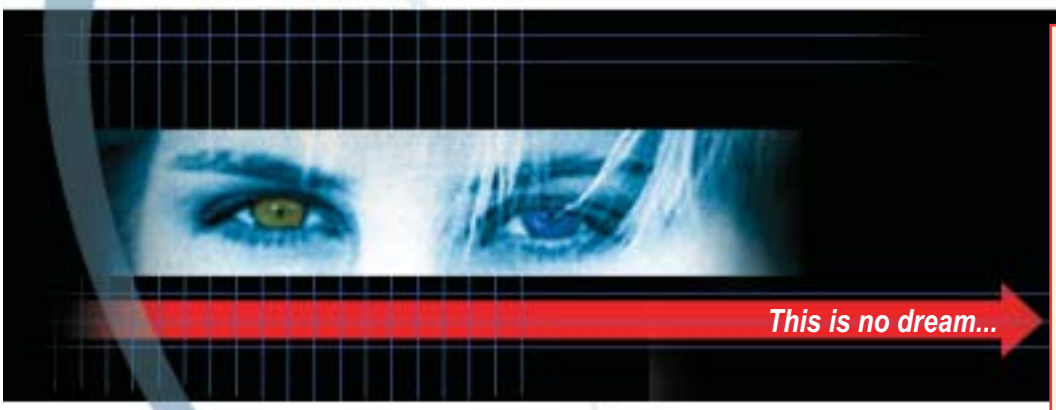
TREETIPS

Issue #32 • August, 2001

A Publication of Treehouse Software • www.treehouse.com

R.A.M. (Rapid ADABAS-to-RDBMS Movement)



State Government
ADABAS-to-NCR Teradata

Defense Contractor
ADABAS-to-ORACLE

Heavy Industry
ADABAS-to-DB2 Mainframe

Natural Resources
ADABAS-to-ORACLE

Educational Institution
ADABAS-to-Sybase & DB2

When faced with the reality of choosing a solution for getting your ADABAS source data to a data warehouse, datamart, or distributed application, the quest can seem quite daunting. Questions immediately start coming to mind, such as:

How much of the project should I expect a product to handle?

Can I talk to a vendor's customers?

What about the data modeling part?

The floodgates open, and you begin to realize the scope of the educational process you will need to undergo.

As a helping hand, we've produced this special **Buyers' Guide** issue of TREETIPS, where we provide a list of questions you should ask when looking for an ADABAS-to-RDBMS solution (page 3). We've also included a thread of postings that recently appeared on the SAG-L on-line discussion forum that deals with this very issue (page 4).

In upcoming issues of TREETIPS, there will be new installments of our ***Real World Series*** customer success stories.

Treehouse Software, Inc.

Inside This Issue

Editor's Sproutings by Joseph Brady

PRODUCT UPDATE: Preview of N2O V4.0.1

We've been working on several new enhancements for the upcoming release of **N2O**, including:

- N2OVRFY validates all local Archive Environment definitions (similar to the way it validates all Local Environment definitions).
- Batch Source Compare can compare two NATURAL objects in remote environments.
- N2OPURGE is able to delete/purge a range of objects in a library.
- **N2O** Event Purge utility purges old **N2O** utility records (Cancels, Checkouts, Rejects, and Transfers).
- N2OUE12N contains new fields. This has been changed to secure all options of the **N2O** Source Compare utility.
- N2OUE22N Batch Autocompile Exit is a new user-exit that can be used to override the JCL Library/Member and Popup window displayed for Batch Job Submissions.
- Added support for NATURAL types: Macro, Report, ExpertModel, Recording, Dialog, Class and Processor.
- An Interface program (N2OAPI2N) that retrieves either of the following without executing **N2O**:
  - Event details
  - Objects in the Event

TREETIPS

Editing, Writing, Design, and Layout
Joseph Brady

Contributors
Larry Jones, Paul Linn, Heather Snyder, Dan Vimont

Production and Distribution
Terri Hammerschmitt and Carolyn Henning

Back issues available upon request. Documentation for all products is available in hard copy or on CD-ROM.

Hard Copy Circulation: 8,000

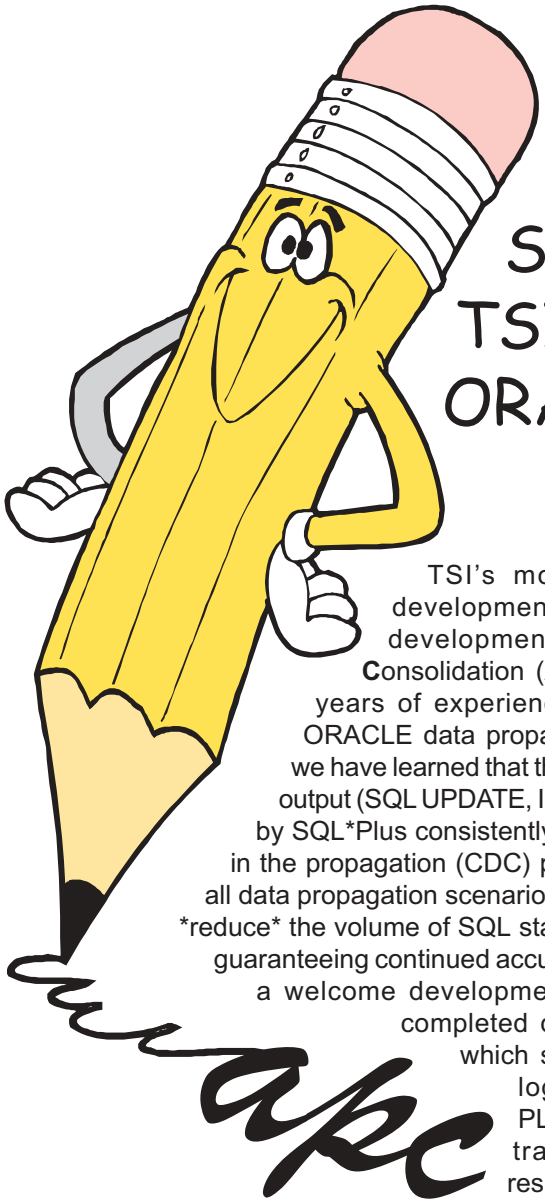
TSI Traveling Tales

Over the past few months, TSI representatives have visited sites in Canada, Germany, Florida, Colorado, New York, North Carolina, and Washington State.

Wayne Lashley, TSI Technical Representative attended the New England Regional SAGGROUP meeting in Rhode Island.

Hans-Peter Will, TSI European Technical Representative, attended the SAGD "Experience 2001" in Munich.

Hans-Peter Will also gave product demonstrations and participated in the Oracle Migration Tour in Germany along with TSI Sales Manager, **Mitch Doricich**.



TSI's most recent ORACLE-focused development work has been centered on development of the new **A**DABAS **P**LOG **C**onsolidation (APC) utility for **DPS**. From our years of experience in serving the ADABAS-to-ORACLE data propagation needs of our customers, we have learned that the processing of **DPS** propagation output (SQL UPDATE, INSERT, and DELETE statements) by SQL*Plus consistently represents the worst bottleneck in the propagation (CDC) process at almost all sites and in all data propagation scenarios. Therefore, anything that could *reduce* the volume of SQL statements (while at the same time guaranteeing continued accuracy of data replication) would be a welcome development. To this end, TSI has just completed development of the APC utility, which serves the purpose of removing logically-extraneous portions of PLOG transactions before the **DPS** transformation process begins, resulting in a reduction in the number of outputted SQL statements that will ultimately need to be processed by SQL*Plus.

Look for APC to be included in the next release of **DPS**.

BE AN INFORMED BUYER...

What Should I Ask When Looking at ADABAS-to-RDBMS ETL/CDC Tools?

TSI's **tRelational** and **DPS** development team were recently asked to come up with a list of questions that a site should ask any vendor offering a solution for ADABAS data migration, propagation, warehousing, etc. The following chart is what they came up with.

I. General Questions	**tRe/DPS?**
Is the product actually in production anywhere for both ETL and CDC processing?	✔
If so, what are the target RDBMS systems? ORACLE? DB2? SQL Server? Sybase?	✔
What do the references have to say about performance, features, ease of installation, ease of use, etc.?	✔
If the product depends on utilities or products provided by other vendors, what contingency plans are in place to provide for the eventuality that support/development of these other utilities or products is discontinued?	✔
II. Modeling and Mapping questions	
Does the system provide data modeling tools? Does the data modeling tool provide for importation of already-existing RDBMS schema? Does the data modeling tool provide for generation of DDL statements derived from a user-designed or user-imported model?	✔
How does the mapping facility provide for the following: field-to-column mapping, field-substring-to-column mapping, constant-to-column mapping, ISN-to-column mapping, multiple-concatenated-fields-to-column mapping, and various options for concatenation?	✔
For MUs and PEs, what are the options for normalization and denormalization? Can multiple occurrences of a repeating field be concatenated into a single column? Can they be mapped to separate columns in a table? Can they be mapped to multiple rows of a table? Can they be spun off into “child” tables?	✔
For null ADABAS fields, does the mapping and transformation process provide for the options of loading/updating the RDBMS with null values? (Some systems may only provide for loading/updating these with spaces/zeros, which obfuscates the actual null value, making it impossible to distinguish null values from actual spaces/zeros values.)	✔
How does the product provide direct support for value-based mapping (for example, to deal with ADABAS “record typing”)?	✔
Does the product provide direct support of primary key and foreign key constraints on target RDBMS tables?	✔
What facility is available for the mapping of a single ADABAS file to multiple RDBMS tables? Are foreign key constraints within such a set of tables supported?	✔
What facilities are available to inform a user when the “image” of an ADABAS file’s structure becomes “out of sync” with either the NATURAL DDM, the PREDICT file, and/or the ADABAS FDT for the file?	✔
Is complete, automatic generation of a complex, multi-table RDBMS schema from an ADABAS file available?	✔
Does the auto-generation process include any form of sampling actual production data to optimize the generated schema, such as determining the number of occurrences of PEs and MUs; the uniqueness of specific ADABAS descriptors that are being considered for RDBMS primary or foreign keys; or the maximum physical lengths and average lengths of fields being considered for VARCHAR RDBMS columns?	✔
What are the options for automatic primary key and foreign key constraint generation?	✔
Can modifications be made to any autogenerated schema, or must the generated schema be used?	✔
Does the modeling and mapping facility provide direct support for all of the major RDBMS systems on the market? ORACLE? DB2? SQL Server? Sybase?	✔
What modeling-and-mapping-related ADABAS file analysis facilities are available with the product? Specifically, how does it provide for analysis of (1) maximum actual physical occurrences of repeating fields, (2) maximum actual physical length and average physical length of alphanumeric fields, and (3) percentage-uniqueness of descriptors and superdescriptors?	✔

(continued on page 7)

Talk About ETL Tools on the SAG-L by Joseph Brady

The following is taken from a recent thread on the SAG-L. We have removed the names of those who posted the messages, since the SAG-L is a discussion forum that is not supposed to be used for marketing to its members.

Question

We are presently looking into acquiring an ETL tool (Extract, Transform, and Load) to assist in extracting data, primarily from our operational, mainframe, ADABAS databases to the data marts and other open systems applications we maintain on ORACLE on NT. Also with an eye to possibly setting up a Data Warehouse in the future.

We are looking at:

IBI Copy Manager

Informatica

Ardent Data Stage

SAS Data Warehouse Manager

SCH's dBridge

CA's InfoRefiner

Has anyone any experience with any of these tools? Has anyone conducted a survey and/or benchmark of ETL tools? Any recommendations? Any other names recommended but missing from above list?

Thanks in advance

Answers

AA

You listed some fine ETL tools, but are any of them able to read ADABAS? Reading flat files produced by NATURAL programs or PLOG extracts doesn't count. You may consider any of those tools for ORACLE-to-ORACLE or SQL Server-to-ORACLE transformations. We use Cognos' Decision Stream for doing RDBMS-to-RDBMS ETL work (which wasn't on your list but I think is comparable to the others).

However, when ADABAS C is your source database, you need to evaluate a whole different set of products. Otherwise, you will run into a brick wall when you try to scale your efforts. You will find past discussions on the pros/cons of these products, and a search of SAG-L archives would be the best way for you to delve into this information, especially if you search on the following product names:

tRelational and **DPS** - Treehouse Software

NatQuery - NatWorks

We have and use **tRelational/DPS**, and you can find out how we use it by looking at the most recent TREETIPS magazine. I will also entertain offline discussions about our use of this product to support our ORACLE-based data warehousing efforts.

BB

I would say that most of your ETL tools will be ODBC compliant, so you would pair up the ETL tool with a ODBC/SQL middleware product like Shadow from NEON Systems, Attunity, or others that take SQL and convert the queries into ADABAS direct calls.

The issue that you will encounter for large files will be the need to maintain some kind of descriptor (indexed) update-flag/timestamp field on your ADABAS file to read only updated records. This implies some application change to maintain the update or timestamp field or implement ADABAS triggers to maintain the updated records.

Also, how often would you need to do the extraction? A nightly read of the entire file may not be an issue. The benefit of the ETL tool in this case is giving the non-mainframe data warehouse person access to the mainframe data in a GUI environment.

AA

Many of these ETL tools handle ODBC calls, but that's so they can handle all the other RDBMS's that they haven't developed the ability to utilize native call interfaces with. The OBDC interfaces are only used if they have to be.

I mean, I guess that "if" all files get timestamped with audit data every time an update occurs, and "if" the DBA makes all such timestamps a descriptor field, and "if" you don't need to handle deletes or manage the kinds of changes that occur as found in the PLOG, and "if" your data mappings are straightforward (file A becomes table A and no MU/PEs are involved and no transformation is occurring), and "if" you aren't pumping serious amounts of data through, I guess you could do that. That's a big "if" that I do not believe is realistic.

Plus when you add the cost of buying an ETL tool AND a product like Attunity Connect or NEON Shadow Direct or Entire Access and ADABAS SQL Server, it would be hard to cost justify why you wouldn't just buy a tool that would allow you to do significant data mappings properly with transformation rules for complex file structures utilizing MU/PE structures for massive amounts of data and using proper change data capture techniques.

It's all about using the right tool for the job.

BB

I'd agree with most of what you stated, you do have to use the right tool for the job and I've seen it to be unrealistic in a lot of cases as well. I was just responding to your question "but are any of them able to read ADABAS?" Personally, I would lean away from data propagation all together & access the data directly from a client/mid-tier application to ADABAS. I understand there's a good case for data warehouses, but a lot of shops assume they have to move data off the mainframe to get to it from their desktop.

And to say "The OBDC interfaces are only used if they have to be." has me confused. Is not ODBC one of the most standardized APIs out there? Why would you want to lock yourself into a proprietary interface/vendor?

AA

It sounds like we're talking about two different things.

(continued on page 5)

ETL Tools on the SAG-L *(continued from page 4)*

Certainly if accessing data on the mainframe suffices as the sole requirement, tools like NEON Shadow Direct, Attunity Connect, Entire Access & ADABAS SQL Server, etc. are the right tools for the job. OBDC is a standard call interface, and I am not against its use to invoke the appropriate driver to access the appropriate database management system. I don't believe one should have to propagate data from OS/390-bases ADABAS C to an RDBMS just for the sole purpose of having it be accessible by an application running on UNIX or NT, either. At our shop, if we just want access to ADABAS data from our client-server or web applications, we use MQ Series to conduct an asynchronous dialogue between the environments. Some places have used products such as Sagavista or Tibco to build and manage their application interfaces and workflow. The kind of work that these tools are meant for, despite there being such a wide range of functionality, is called "Enterprise Application Integration".

EAI is good for accessing live up-to-date data in its source location, but data warehousing needs are just about never satisfied by EAI because of the fact that data warehouse schemas are often different from their transactional source systems, and it's good to isolate DW queries so they do not compete with your transactional system's resources. Also, because you may have so many disparate data sources, if you query the target systems you must pass data back-and-forth and are limited performance-wise to bandwidth availability, plus it is impossible to tune a query to perform well if the data is coming from multiple sources.

David Linthicum, the former CTO of SAGA Software and well-known EAI expert, writes in his book, "Enterprise Application Integration", that "Unlike EAI, which can support real-time data movement, data warehousing provides adequate business information but without up-to-the-minute access of information. In many cases, the data is weeks, even months old, and the data mart or data warehouse is updated through antiquated batch, extract-aggregate-and-load processes." He goes on to predict that as EAI matures, it will become feasible to have data warehousing solutions with real-time up-to-the-minute data. Until then (it's me speaking again), we must balance our need to have voluminous data propagated with appropriate transformations against our need to get this data propagated as quickly as possible. EAI is not mature enough yet to handle a real-time data warehouse, and if and when it does get there, I suspect it will be on the high-end from companies such as Tibco, webMethods, Iona, etc., which already have the capability to do simple transformations and to manage workflow.

Now, this is important because it sets the context for what I meant about not wanting to use ODBC if a native call interface exists: as performance matters greatly in this balancing needs between real-time vs. volume of data, the better the performance is of the retrieval of data for processing, the better things will be for either moving closer to real-time or for propagating more data (or for increasing each one slightly - your choice). That is why if you wish to buy a tool for data propagation, you should select one that is built with the understanding of how to most efficiently extract the kind of data you need for handling "change data capture" and data transformation. This is why I said you would only want to use ODBC for this process if you had no better solution, and direct call interfaces are a little better than ODBC. Even that's not perfect, as direct call interfaces still compete for resources on your source data system. As all the data you'd ever need is captured on ADASAV tapes and PLOGs, the best performance is gained by using tools that can extract data from these sources.

In using **tRelational/DPS**, our data warehouse in ORACLE on UNIX contains several large files from ADABAS, and it is always current up to the end of the last business day (current business day will be available tomorrow). This is by choice, though, as we could have it be as of the previous PLOG flip about 30 minutes after the flip occurs (conservative estimate). This is much better than the weeks or months that **David Linthicum** ascribes to the age of DW data built by batch load processes, mainly because of the efficiencies built in to the "right tool for the job".

BB

I think we are on the same wavelength, but coming from different perspectives.

You wrote:

>He goes on to predict that as EAI matures, it will become feasible to have data warehousing solutions with real-time up-to-the-minute data.

Yes, I've read his book & actually spoken with him before. I agree EAI has a long way to mature, but that doesn't stop ETL vendors from trying to jump in and play in the space. They are doing it, currently, most cases in a "polling" mode to pull data off the mainframe at X sec intervals to achieve near real-time updates, which is a poor implementation, but works. Or when available using database triggers. Why do you think vendors are now talking about change data capture coupled with an ETL tool? Once "real-time push technology" for going from mainframe to EAI tools comes of age, then we are talking something significant. Maybe ADABAS 7 triggers will prove to be viable.

You wrote:

>I suspect it will be on the high-end from companies such as Tibco, webMethods, Iona, etc., which already have the capability to do simple transformations and to manage workflow.

They may be high-end for workflow management & GUI transformations, but do they have mainframe expertise? They would have to find partners.

(continued on page 6)

ETL Tools on the SAG-L *(continued from page 5)*

You wrote:

>not wanting to use ODBC if a native call interface exists: as performance matters greatly in this balancing needs between real-time

Agreed, with 2 exceptions: 1. Don't assume an ODBC or JDBC driver is going to be a poor performer, not all drivers are created equal, it depends on the underling middleware architecture. 2. A lot of shops are looking for vendor independence, they don't want to be locked into a specific, proprietary solution. The EAI or ETL tool has to offer standardized APIs like (ODBC,JDBC,SAP/BAPI,IDOC) to do database or application integration. Example: I've integrated a 3270 CICS/NATURAL application to SAP using BAPI using EAI vendor A. (I have actually done this with a 3 second response time to the CICS end-user.) Now vendor A gets bought out by big-bad vendor C, and I don't like that combination of vendor letters. I now want a better solution because, I now need to integrate to ORACLE Financials as well. What's the effort of replacing the EAI solution with vendor B? Most of the work was already done in the interface, adapter, or end-point & they are standard APIs...lucky me.

You wrote:

>In using **tRelational/DPS**, our data warehouse in ORACLE on UNIX contains several large files from ADABAS, and it is always current up to the end of the last business day (current business day will be available tomorrow). This is by choice, though, as we could have it be as of the previous PLOG flip about 30 minutes after the flip occurs (conservative estimate). This is much better than the weeks or months that **David Linthicum** ascribes to the age of DW data built by batch load processes, mainly because of the efficiencies built in to the "right tool for the job".

Perfect product for migrating ADABAS data nightly, but it's database specific and not real time.

Another Answer (and final words in this thread) posted by Treehouse Software:

This is an important and interesting thread, because the discussion has highlighted the differences between middleware and ETL/CDC tools, and between one-size-fits-all products and those that are designed to solve a specific need.

tRelational and **DPS** are indeed database-specific in terms of migration source—ADABAS (it should be pointed out that the products *natively* support ORACLE, DB2, Microsoft SQL Server and Sybase as targets—no add-on "adapters" or "selectable units" required). This is by design, since Treehouse Software has many years of expertise and experience in developing products complementary to ADABAS and other Software AG products. **DPS** is indeed not real-time. This is also by design. It is not a middleware solution, and does not pretend to be. It is a data propagation system (hence the name!) for ETL and CDC, ADABAS to RDBMS, not necessarily for nightly use but on any interval that the customer chooses—the smallest unit of time being the time that it takes to switch a PLOG and process its contents. The fact that it is not real-time, that it does not require connection to, nor impose a load upon, the source ADABAS database is one of the product's strengths—but only one, as there are many others.

TSI has spent millions of dollars and years of effort (since 1994) in developing **tRelational** and **DPS**. When the requirement is to migrate, especially with complex mapping relationships, ADABAS data to relational databases to support system conversion, operational data stores, data marts and data warehouses, we believe ours is the best solution on the market. With over 20 customers who have purchased the products, and at least that many actively conducting trials and pilot projects, we believe this claim is supported. Oracle Corporation has endorsed these products as the recommended solution for migrating ADABAS data to ORACLE.

The migration of ADABAS data to an RDBMS can easily be accomplished with TSI's **tRelational** and **DPS** product set. ●

It was the best of data, it was the worst of data

When suffering the slings and arrows of viewing the data in your Protection Log, you can be overwhelmed by its sometimes voluminous and complex form.

TSI's **AUDITRE** can be used to help sort out an otherwise intimidating mountain of information by efficiently processing the Protection Log data.

Taking user defined direction through **AUDITRE** parameter cards and input from the Protection Log, **AUDITRE** is able to determine the number and kinds of transactions (adds, updates, or deletes to one or more files) existing on the ADABAS Protection Log(s) IN ONE PASS and produce multiple reports or output the results to a sequential dataset(s) for later processing.

So sit back, relax, and take in some light PLOG reading with the help of **AUDITRE**.

What Should I Ask When Looking at ADABAS-to-RDBMS ETL/CDC Tools?

(continued from page 3)

	tRe/DPS?
III. ETL questions	
Does the product directly support the high-speed loader utilities of all of the major RDBMS products on the market?	✔
If the product requires direct database access for extraction of data from ADABAS, what have existing users found with respect to performance impact against the production system?	✔
If the product provides for multiple adhoc data-extraction queries to be run by end-users, what facilities are available to limit the threat that a single user or group of users will generate and submit queries that put an unacceptable burden on ADABAS?	✔
Does the ETL process entail the FTPing of files from one platform to another? If so, are these files fixed-format, delimited, or does the user have the option to choose the file format, or to choose other options for file compression to limit the time it takes to accomplish FTP transfers and to minimize file storage problems as much as possible? Additionally, how does the product deal with operating-system-specific file-size constraints?	✔
Does the product allow for the distinguishing between a physically "null" ADABAS field and one that contains either a blank or a zero?	✔
IV. CDC questions	
Does the product directly support the SQL-processing utilities of all of the major RDBMS products on the market?	✔
How does the product assure that no "UPDATE", "DELETE", or "INSERT" transactions against target RDBMS tables will be rejected due to violation of primary key or foreign key constraints? Explicitly, how does the presence of a primary key or foreign key constraint on a target RDBMS table affect the product's transformation of an ADABAS transaction into RDBMS transaction(s)?	✔
What is the methodology for ensuring that the CDC solution keeps the data on target RDBMS tables "in sync" with the data on the source ADABAS files?	✔
Does the CDC process entail the FTPing of files from one platform to another? If so, are these files fixed-format, delimited, or does the user have the option to choose the file format, or to choose other options for file compression to limit the time it takes to accomplish FTP transfers and to minimize file storage problems as much as possible?	✔
What steps must a user take during the ETL process to assure the feasibility of a subsequent PLOG-based CDC process? (e.g., should the file(s) be placed in an "exclusive use" state? If not, how is the transactional integrity of a subsequent PLOG-based CDC process ensured?)	✔
Does the CDC solution allow for the distinguishing between a physically "null" ADABAS field and one that contains either a blank or a zero?	✔
Does the system provide any type of "PLOG consolidation" to reduce the CDC transactional volume imposed on the target RDBMS? (Note that usage of ADACDC is not acceptable, as its consolidation methodology leads to inevitable violations of RDBMS transactional integrity with respect to primary key and foreign key constraints.)	✔

Demo TSI Products On-line

To set up a live, on-line demonstration of any TSI product, simply fill out the short form on the Treehouse Software Web site at **www.treehouse.com/webexform.html**. All you need is an Internet connection and a current Web browser (Netscape, Internet Explorer, etc.) to see TSI products in action right on your PC screen.

Become a Beta Test Site for TSI Products

We are looking for **current North American customers** to become Beta test sites for new releases of TSI products. If you are interested in becoming a Beta test site, please fill out the short form at **www.treehouse.com/betasites.html**, and a TSI representative will contact you.

TREETIPS

Treehouse Software products include:

Relational Products:

tRelational - ADABAS data analysis, relational modeling, and mapping tool; **DPS** parameter generator

tRelationalPC - Windows-based graphical interface to make the tasks of modeling and mapping even simpler.

Treehouse Remote Access (TRA) - Middleware that allows **tRelationalPC** to communicate with **tRelational** on the mainframe.

DPS - ADABAS to RDBMS data materialization, replication, and propagation software

UNIX Products:

SEDIT - XEDIT and ISPF/PDF compatible editor for UNIX and Windows

S/REXX - REXX-compatible language for UNIX and Windows

S/REXX Debugger - Optional graphical debugger for **S/REXX** programs

Software AG Related Products:

ADAREORG - File reorganization tool for ADABAS

ADASTRIP - Data extraction utility for ADABAS

* **AUDITRE** - Generalized ADABAS auditing facility

* **AUTOLOADER** - ADABAS file automatic unload/reload/dump utility

* **CHART for NATURAL** - NATURAL application analysis and documentation tool

DBAUDIT - Data integrity verification utility for ADABAS

* **N2O** - NATURAL application change management system

* **N2O/3GL** - 3GL support within **N2O** for PANVALET, LIBRARIAN, ENDEVOR, and PDSs

PEEK - ADABAS file browsing utility

* **PROFILER for NATURAL** - NATURAL quality assurance and testing tool

QDUMP - Incremental backup utility for ADABAS

RACE - NATURAL performance enhancer and "**R**edundant **A**DABAS **C**all **E**liminator"

* **SECURITRE** - ADABAS and NATURAL security interface to RACF, ACF2, and TOP SECRET

* **TRIM** - ADABAS and NATURAL performance monitor

* Indicates TSI Products that are marketed for TSI by international affiliates

Phone: (412) 741•1677 Fax: (412) 741•7245 E-mail: tsi@treehouse.com Web: http://www.treehouse.com

TREEHOUSE SOFTWARE, INC.
409 Broad Street, Suite 140
Sewickley, PA 15143 USA

PUT MAILING INFORMATION HERE FOR MAILING LIST QUANTITY ONLY