

Incremental Updates vs Full Reload **Change Data Capture**

Minutes vs Hours

CONNX Solutions

September 2007

TABLE OF CONTENTS

Executive Summary.....	3
Accessing Data from a Variety of Data Sources and Platforms.....	4
Approaches to Moving Changed Data.....	4
Change Data Capture.....	7
How CONNX DataSync Moves Changed Data.....	8
Benefits of The CONNX DataSync Solution.....	13
Closing Thoughts and Contact Information.....	16

EXECUTIVE SUMMARY

Companies usually move changed data from source to target by performing extracts and loads. This approach often involves custom programming, monitoring and moving an entire data volume. The CONNX® DataSync tool solves these issues with its point-and-click interface and pre-programmed selections. Tasks can be automated during successful and problematic synchronizations using CONNX DataSync's built-in error monitoring.

Companies often need to migrate their data. CONNX DataSync creates a target schema that matches the current schema and transforms the data types automatically to ANSI/ISO compliant versions, at the same time it transfers the source data to the target. No written code is needed to perform any of the data transformations or data conversions.

In summary, CONNX DataSync moves the company's enterprise data from virtually any source location to virtually any target data source and performs incremental updates from source to target at a fraction of the usual full update time.

ACCESSING DATA FROM A VARIETY OF DATA SOURCES AND PLATFORMS

In the modern-day business environment, large organizations can have a variety of data repositories that operate either in association with off-the-shelf software packages or as custom solutions to business needs. Different departments in an organization might run SAP® on an Oracle® database, PeopleSoft® on Oracle, or CINCOM SYSTEMS® MRP program on IBM® VSAM. Some companies run tools that require permanent data stores from multiple vendors.

Complicating this further, data produced by one department, such as manufacturing, may be needed by another department, such as finance. The finance department's analysis tools or skill sets may require a different database platform than the manufacturing department uses to store its data.

In addition to data access issues, companies often face the daunting task of migrating a database system to a different platform. Total corporate demand for data may exceed the bandwidth of the server where the data natively resides.

APPROACHES TO MOVING CHANGED DATA

There are four typical approaches to moving changed data between systems:

- Performing extracts and loads
- Log shipping
- Timestamps
- Trigger based methods

Performing Extracts and Loads

Figure 1 shows performing extracts and loads. This approach requires writing an extract script and a load script that moves data from the production server to form a copy on the target reporting server.

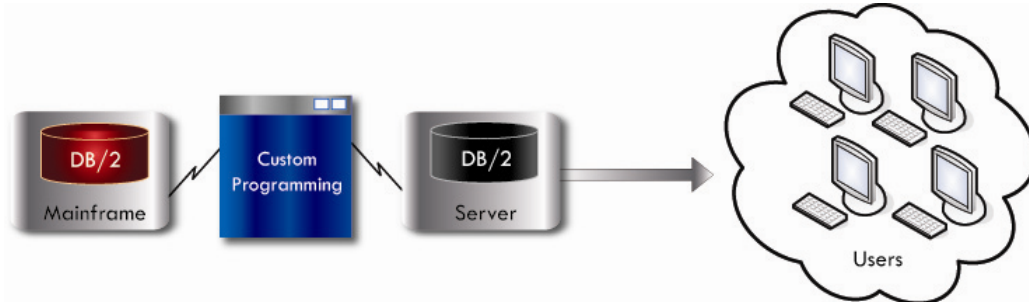


Figure 1: Performing Extracts and Loads

There are several problems with this approach.

- Some custom programming is required to accomplish the extract
- Moving the entire volume of data can be expensive and time consuming
- The batch process must be monitored for conditions like running out of temporary disk space or losing connections.

This approach is inefficient because it takes time and resources to create each extract request.

Log Shipping

This approach sends copies of the source database transaction log to the target database. This log contains a list of actions performed against the database. These transactions are then replayed against the target database. Log shipping can be a cost-effective solution since it only moves changed data and has very little source system impact.

There are several problems with this approach.

- Both the source and target databases must use the same database systems.
- Both the source and target databases must use the same type of transaction log.

Timestamps

A timestamp is a field that marks a record with the date and time of the last action against the record. This approach can be used to move changed data to a new target server.

There are several problems with this approach.

- This method doesn't work unless the system was designed to access and use the timestamp fields.
- All the source tables must include time stamps that are modified each time the table changes.

Trigger Based Methods

When an insert, update, or delete occurs, a trigger fires to forward the action to another data store.

There are several problems with this approach.

- This method requires custom programming.
- This method slows down the source database due to trigger activity.

CHANGE DATA CAPTURE

Change Data Capture can address the task of data migration and the issues that extracts and loads pose. This tool captures the source data incremental changes as opposed to full updates, and delivers them to the target data source (a data warehouse or data mart) which can be used for analysis and production tasks. With change data capture, a snapshot of the source data versus the entire database is delivered to the target, being a data warehouse or data mart.

There are two common methods of change data capture:

- **Batch-Oriented Change Data Capture (Pull Change Data Capture)**

Batch-Oriented Change Data Capture takes a snapshot of the source database and duplicates the data into the target database. This is usually done in a batch or on-demand environment.

- **Live/Real-Time Change Data Capture (Push Change Data Capture)**

Live Change Data Capture, captures changes as they happen in the source database and performs the exact same changes on the target database. This is usually done in a live data or transaction-oriented environment.

HOW CONNX DATASYNC MOVES CHANGED DATA

CONNX DataSync is a batch-oriented (or pull) change data capture tool. It captures incremental changes and delivers them at user-defined intervals to a data warehouse or data mart in a fraction of the time required by full updates. It allows users to move a complete snapshot of a pre-defined list of tables into the data warehouse. DataSync also incrementally updates the current snapshot of data through a simple-to-use scheduling interface. Figure 3 illustrates the CONNX DataSync process.

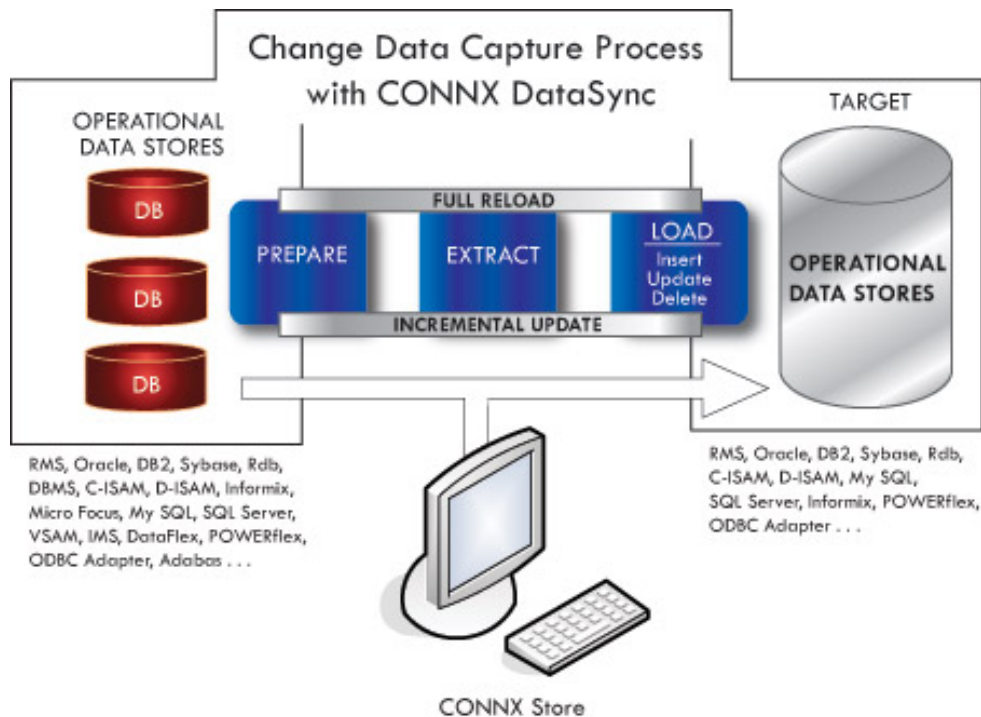


Figure 3: CONNX DataSync Process

The primary difference between the CONNX DataSync tool and general Change Data Capture is the tool's unique technique of hash keys to detect which records have been added, deleted, or changed on a scheduled basis. CONNX DataSync uses that information to insert, update or delete the corresponding data in the data warehouse snapshot tables. Incremental updates are delivered from source to target considerably faster than full updates because the hash keys are stored along with the primary key of each record in a separate CONNX private data store.

How CONNX DataSync Resolves the Problems Extracts and Loads Pose

The CONNX DataSync tool resolves the extract and load script programming, monitoring, cost and time issues with point-and-click selections. Built-in error monitoring makes it possible to perform different tasks automatically depending on whether synchronization was or was not completed successfully. Since all data transformations (such as converting from big-endian to little-endian format or from hardware specific time formats to ANSI/ISO SQL formats) occur automatically, less manual intervention means reduced time and expense.

With the CONNX DataSync tool, primary keys or all keys associated with a table can be transferred to a reporting server. The same keys can be used on both the target and on the source; you can customize the keys to perform different tasks on the target and on the source.

The CONNX DataSync tool moves only changed data. In a table with millions of rows of thousands of characters each, it is much faster to move a few hundred or a few thousand changes than it is to transfer the entire table. The task groupings can be easily separated into different synchronizations to automatically provide each department with the information they want when they want it.

How CONNX DataSync Resolves Data Migration Issues

The CONNX DataSync tool also helps users solve data migration issues.

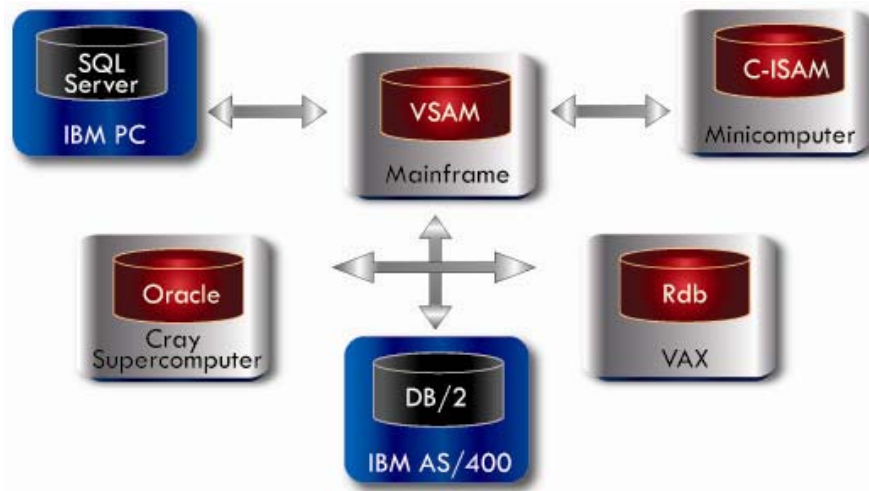


Figure 4: Data Migration with CONNX DataSync

Migrating a database system between different operating systems and databases usually requires a considerable amount of time. With the CONNX DataSync Tool, the data and schema can be easily transferred between systems. The only tasks that need to be performed are selecting the source and target databases and synchronization frequency. Translations used to move the schema and data to the new target are performed automatically.

IT staff often spend hours analyzing and mapping out the required data transformations for a typical data migration. The CONNX DataSync tool can minimize manual analysis and mapping by creating a target schema to match the source schema, and automatically transforming the source data types to ANSI/ISO-compliant versions, while the data migrates to the target. It performs all of the data transformations and data conversions without adding a single line of code.

CONNX can be used to develop a standards based GUI front end for a company's old data stores using front ends like Java and VB.NET. After CONNX DataSync is used to create a target database, any standards based tools used to access the source database can be used to access the target database. The standards based tools remain completely unchanged. The only task to be completed is to change the name of the data source.

The CONNX DataSync Tool Performs Several Unique Product Functions

- **Extracts Data On Demand**

Extracts data as a complete snapshot of a pre-defined list of tables from the source and moves it to the target data warehouse or data mart.

- **Full Reloads On Demand**

Creates a copy of the source data warehouse and moves this copy into the target data warehouse.

- **Incremental Updates On Demand**

Employs unique technique of hash keys to quickly detect when records are updated, deleted, or inserted. By storing the hash keys along with the primary key of each record in a separate CONNX private data store, the CONNX DataSync tool performs incremental updates from source to target at a fraction of the usual full update time.

Initial full update from source to target may take two to three times longer than the equivalent transfer using FTP (for example: 3.5 hours) because the full update also:

- ✓ Inserts data into SQL Server
- ✓ Creates CRCs
- ✓ Stores information on CONNX Store
- ✓ Transforms the source schema data types from to the target schema data types.

Other costs are also associated with the initial full update, such as location on a Wide Area Network.

The incremental update usually runs four to ten times faster than the full update, a significant decrease in the synchronization time.

- **Scheduled Incremental Update**

On a scheduled basis, detects which records have been added, deleted, or changed, and inserts, updates, or deletes the corresponding data in the data warehouse snapshot tables. Each file in the synchronization process may be scheduled independently, or as a group.

- **Scheduled Full Reloads**

On a scheduled basis, purges all prior snapshot data in the data warehouse. A completely new snapshot of the data is placed in the data warehouse. Each file in the synchronization process may be scheduled independently, or as a group.

- **Creates Data Mart Model Schema**

Automatically performs a complete analysis of all source database schemas and applicable transformations for the target data mart schema.

- **Creates Data Mart Model Transformations**

Automatically creates data type transformations of all data elements from source schema to target schema.

- **Works with Applications**

Once the target source is populated, CONNX, the engine powering DataSync, can be used with other tools for ad hoc reporting, application development, and Web development through OLE DB, ODBC, or JDBC drivers. CONNX provides the flexibility of querying either against the data warehouse and data marts, or against the live operational data sources.

- **Creates Multiple Data Mart Instances**

Additional copies of a source database can be created for each database system in the enterprise that has been logged as a suitable target database in the CONNX Data Dictionary (CDD). The CONNX DataSync Administrator tool allows you to override the default target database for group synchronization and specify any target database known to the system with two mouse clicks.

- **Error Detection and Correction**

CONNX DataSync allows you to automatically detect error conditions, and re-sync until a successful sync has been performed.

- **Task Triggering**

Additional tasks can be triggered based on synchronization status. A triggered task can be as simple as sending an e-mail to notify successful sync completion, or as complex as calling an ETL (Extract, Transform, Load) tool to perform additional transformations against the synchronized data.

BENEFITS OF THE CONNX DATASYNC SOLUTION

- **Scalable**

The CONNX DataSync tool offers multithreaded power to leverage the force of inexpensive modern hardware. Multiple CPU 64-bit machines with thousands of MIPS of throughput can be purchased for a few thousand dollars. Since the CONNX DataSync tool is specifically designed for environments running on multiple machines, this can drastically reduce the load against production systems. If demand warrants, multiple synchronization servers can be added to meet increasing demand. As demand for data expands, you can continue to add inexpensive hardware to meet your needs.

- **Secure**

In addition to the current database security, CONNX can easily add additional security protection. CONNX not only features table- and column-level security used to restrict access to specific types of information, but also includes row level security. For example, an employee will be allowed access to his/her own personnel record, but not to the records of other employees.

- **Easy to Use**

In many cases, the CONNX DataSync tool is up and running a production schedule within an hour of installation. Importing a source database and a target database into the CONNX repository only requires the user to provide the connection information and to select the tables to transfer. A mirror image can then be made of the entire database system or any desired subset. Synchronizations can be run at any given time or on a regular, automated schedule.

- **Minimizes Project Risk**

- ✓ Eliminates manual/one-time programming efforts.
- ✓ Provides option for automatic synchronization between originating and target data sources.
- ✓ Provides consistent methodology for determining and synchronizing adds, changes and deletes.

- **Reduces Computer Resources Required**

- ✓ Streamlines effort required to populate target data sources
- ✓ Time required for updates is reduced significantly

BENEFITS OF THE CONNX DATASYNC SOLUTION

- **Saves Time and Money**

- ✓ Eliminates the need to create and maintain “homegrown” solutions.
- ✓ Eliminates the need for extracts.
- ✓ Lowers the total cost of ownership.
- ✓ Achieves time savings with incremental updates versus full reloads.

Actual Customer Results Using the CONNX Data Synchronization Tool				
		Full Reload	Incremental	CONNX Data Synchronization Server Specifications
DB2 to Oracle				
	20 million rows (333 tables)	3.5 hours	12 minutes	Quad Processor 700 MHz, 1 GB RAM, Windows NT 4.0
RMS to SQL Server				
	142 million rows	8 hours	2 hours, 40 minutes	Single processor 1.1 GHz AMD, 1 GB RAM, Windows 2000
	100,000 rows (1 wide table over slow WAN)	20 minutes	1 minute	Dual processor 400 MHz, 2 GB RAM, Windows 2000
VSAM to SQL Server				
	4.2 million rows (20 tables over slow WAN)	4 hours	48 minutes	Dual processor 600 MHz, 512 MB RAM, Windows Server 2003

Figure 5: Customer Benchmark Results using the CONNX DataSync Tool

In Figure 5 Change Data Capture Operations within legacy databases are contrasted with full database restores to target relational databases. The results show that incremental updates were between three and twenty times faster than full reloads for these customers.

CLOSING THOUGHTS

Whether a company is facing a data migration project or is looking to resolve the issues extracts and loads can pose, CONNX DataSync is the solution. CONNX DataSync moves the company's enterprise data from virtually any source location to virtually any target data source and performs incremental updates from source to target at a fraction of the usual full update time.

CONTACT INFORMATION

For more information regarding the CONNX DataSync tool, please call, e-mail or visit the CONNX Solutions Web site:

CONNX Solutions.
425-519-6600
sales@connx.com
www.connx.com

2039 152nd Ave. NE
Redmond, WA 98052

CONNX Solutions and CONNX are registered trademarks of CONNX Solutions. All other trademarks, registered trademarks, product names, and company names mentioned herein are acknowledged as the property of their respective owners.