



**High Performance Connectors
for
Microsoft SQL Server
Integration Services
September 2005**

Evolutionary Technologies International, Inc.
816 Congress Avenue
Suite 1450
Austin, Texas 78701
www.eti.com

Microsoft®
GOLD CERTIFIED

Partner



Contents

Executive Summary 1
Background 2
 The traditional trade-off: speed of implementation vs. speed of execution 3
ETI Solution 4
 Requirements for the ETI High Performance Connectors 6
ETI Integration with SQL Server 2005 7
Case study: 8
Transferring data from Teradata using Integration Services 8
 Test Results 9
 Test data 9
 Configuration 9
 Results 10
 Conclusion 11

Executive Summary

Today, enterprises face large volumes of data that are not accessible for performing cost-effective, near real-time analysis. For certain types of source systems, e.g. legacy, mainframes, etc., real-time access is thwarted by proprietary access modes. Traditional integration solutions build silos for reporting and analysis and require a disparate set of tools to create the solution. This silo solution tends to be expensive and hard to maintain and doesn't offer a consistent view of the business.

With its release of SQL Server 2005, Microsoft provides best-of-breed engine-based data integration, analysis and reporting in keeping with its commitment to provide customers with high-value, low-cost software solutions.

Founded in 1991, Evolutionary Technologies International, Inc. (ETI) provides access to proprietary source systems with significant TCO advantages. With the release of SQL Server 2005, ETI integrates its products with the Microsoft® Business Intelligence platform.

ETI Solution® Version 5 is a collaborative development environment that helps large enterprises automatically build and maintain native data interfaces that would otherwise have to be written by hand. A key differentiator between ETI and other connectors or integration development platforms is that the executables are source code and therefore, no proprietary software from ETI has to run in the execution environment. Consequently, ETI can produce connectors for streaming source data directly into the Microsoft Business Intelligence Platform without requiring the data to be staged in intermediate tables.

Background

One of the hardest problems facing most strategic IT projects is a cultural divide caused by the different skill sets of the technical people involved. Typically, the individuals trained in the new technology do not have knowledge of the source data technology. However, any programmer who wanted to write an interface would have to know the following:

- The access languages used by both the source and target systems
- The schema definition of both the source and target databases
- How to interface to the communication software used between the host computers.

Flat files are often used to bridge this skills gap. In that case, the technical staff assigned to the source systems creates a flat file dump (or dumps) of the desired data, and the developers code against the file layouts provided with the flat file. Unfortunately, this strategy requires two extra I/O steps, which degrades performance.

Gateway products reduce the time required for developers to design and implement applications that take data from multiple source systems. The applications are implemented in such a way that they can interpret any request (usually a SQL statement) and use that information to translate between the source data format and the format expected by the requesting application. However, because gateways also use interpretive code to read and translate each record, performance is often unacceptable when large volumes of data are involved.

The traditional trade-off: speed of implementation vs. speed of execution

ETL tools provide connectivity using standardized access methods for importing source metadata at design time. Common ways of doing this include ODBC, OLEDB, etc. These tools provide a “codeless,” GUI-driven, Integrated Development Environment (IDE) for higher productivity. On the other hand, several source systems provide high performance, bulk data exchange tools that typically have weak design time support.

Microsoft offers customers the best of both worlds – rapid development through the SQL Server 2005 Business Intelligence Development Studio (BIDS), *and* a runtime providing SQL Server Integration Services (SSIS) users and SQL Server Analysis Services (SSAS) users with rapid access to the data.

In the joint ETI-Microsoft solution, BIDS can be used to configure SSIS components to call an ETI™ High Performance Connector (HPC) with the SQL statement that it would have otherwise passed to the OLEDB driver. During the design and development process, the ETI HPC simply invokes the OLEDB driver so the developer can:

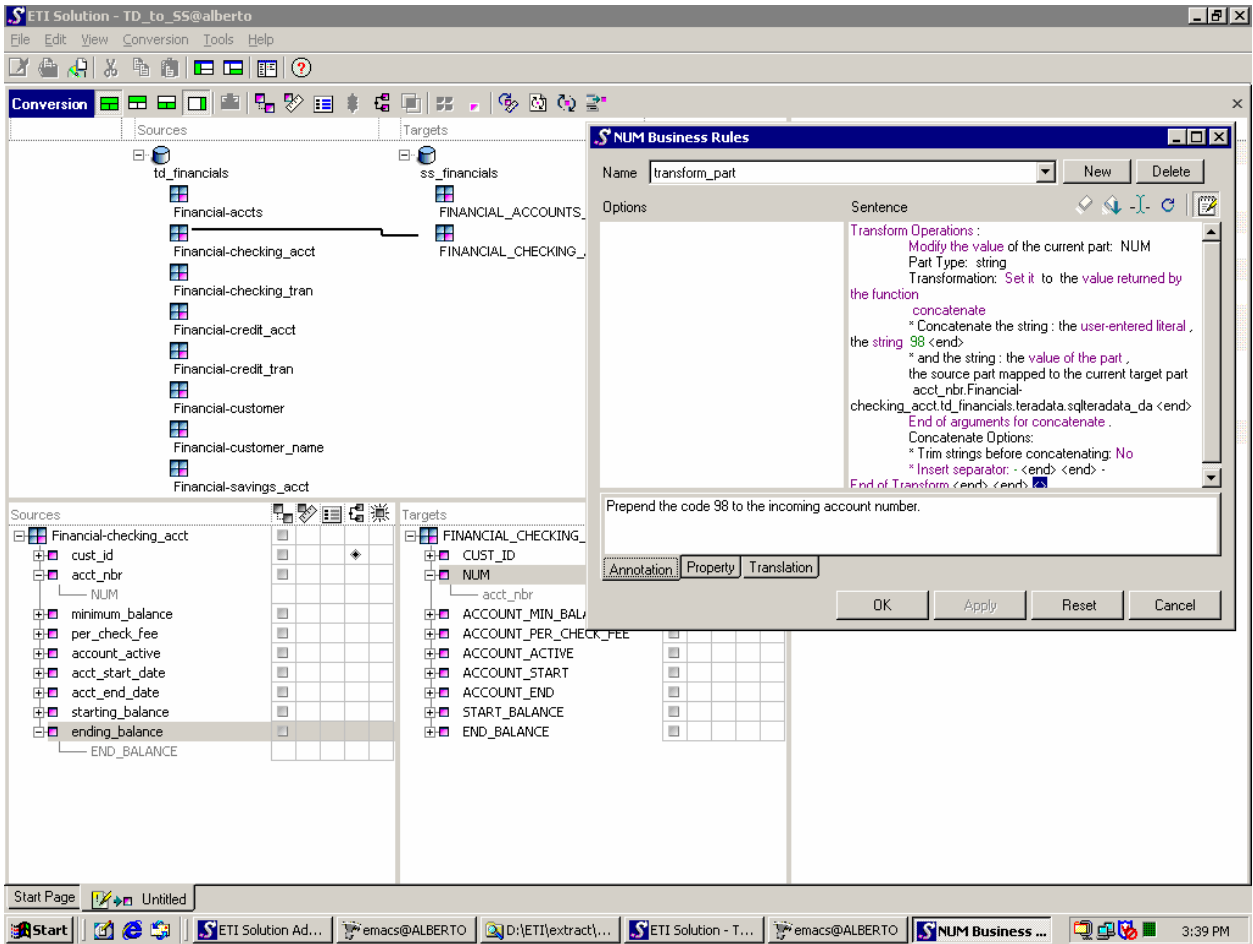
- Dynamically access the source schema information
- Specify the desired sequence of tasks
- Apply any desired transformations, and
- Test the resulting specification.

During the runtime of the BI solution, instead of invoking the OLEDB driver the ETI HPC invokes the appropriate utility to efficiently access data from the source system. However, rather than writing the records to a file (the default process for most fast export utilities), the ETI HPC intercepts the records, performs the necessary data type conversions, and streams them directly into the Microsoft SSIS Data Flow Engine. There, it is available for further data cleansing or transformation, or streaming directly into destination systems including SQL Server Analysis Services. The runtime performance using HPC is comparable to the performance of the source system’s native fast export utilities.

ETI Solution

ETI Solution® Version 5 is a collaborative development environment that helps large enterprises automate and maintain native data interfaces that would otherwise have to be written by hand. ETI Solution helps eliminate the gap between the data that is actually stored in production databases and the implicit knowledge contained in the minds of data owners and business process owners. By providing these (usually) non-technical users with a smart client (for the data owner) and/or reports (for the business process owner), ETI Solution not only speeds implementation, but captures this information in the metadata used to generate the code. Built on the Microsoft® .NET technology, ETI Solution can be used globally from a single installation, thereby providing the best of both worlds – centralized control *and* visibility with distributed development. This architecture also allows ETI and its partners to offer hosted solutions to mid-sized businesses.

Figure 1 illustrates the interface using ETI Solution to create a custom connector between Teradata and SQL Server (or vice versa). Customers who want a custom connector between SSIS and other production systems will use the ETI Integration Design Studio to provide the specifications for generating the native adapter.



*Figure 1
Using ETI Solution to Create a Custom Connector
To SQL Server 2005*

Requirements for the ETI High Performance Connectors

While ETI's smart clients are relatively easy to learn and use, the goal for the HPC was to provide a single, generic connector for each data source that:

- Could be installed and treated as a single Data Flow component in the SSIS environment
- Would provide BI designers the flexibility of using the vendor's OLEDB driver to interactively build and test their solution
- Would provide significantly faster performance at runtime.

As Figure 2 illustrates, the ETI HPC toggles between interfacing to a vendor's OLEDB driver for the data source and a vendor-supplied fast export utility.

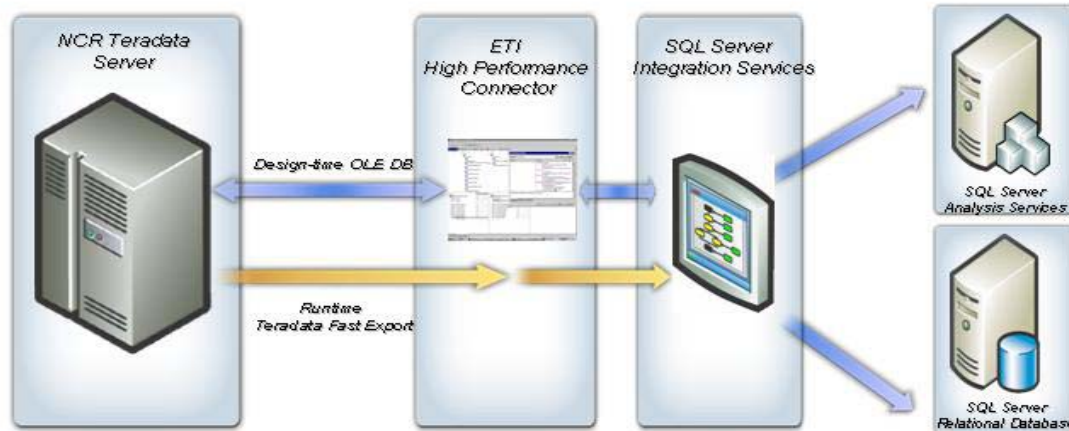


Figure 2
ETI HPC at Design and Runtime

Because the ETI HPC simply streams the data to SSIS rather than writing it to an intermediate file, the performance at runtime is comparable to that of the native fast export utility which writes to a file.

ETI Integration with SQL Server 2005

In SSIS, the ETI HPC is accessed through a source connector written in managed .Net code and available for both 32 and 64 bit environments. This connector is available in the BIDS environment as just another Data Flow Source Adapter and provides a simple drag and drop interface for creating data cleansing and transformation solutions.

Figure 3 below illustrates this user experience.

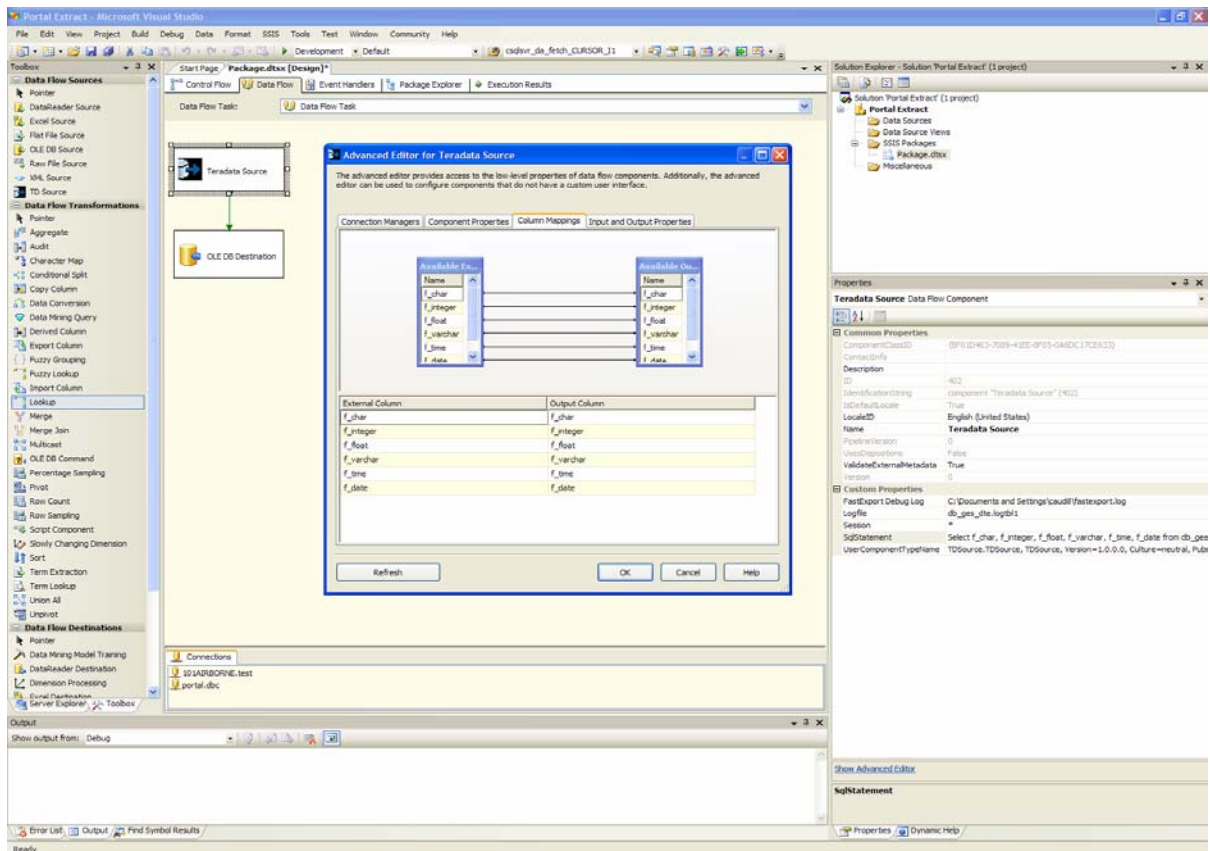


Figure 3

Case study:

Transferring data from NCR Teradata using Integration Services

One of the beta customers of SQL Server 2005 Integration Services, a large telecom company, wanted to use the analytical capabilities of SQL Server 2005 against a large Teradata source system. However, the performance of the OLEDB driver provided by Teradata was not adequate for performing timely analysis.

To address this problem, ETI created a Teradata HPC. As illustrated in Figure 4 below, the HPC:

- Invokes the Teradata FastExport utility with the appropriate Teradata FastExport script
- Intercepts each record and performs any necessary data type transformations, and
- Streams the data directly into SSIS.

In this way, the HPC eliminates the disk I/O normally incurred by the Teradata FastExport utility.

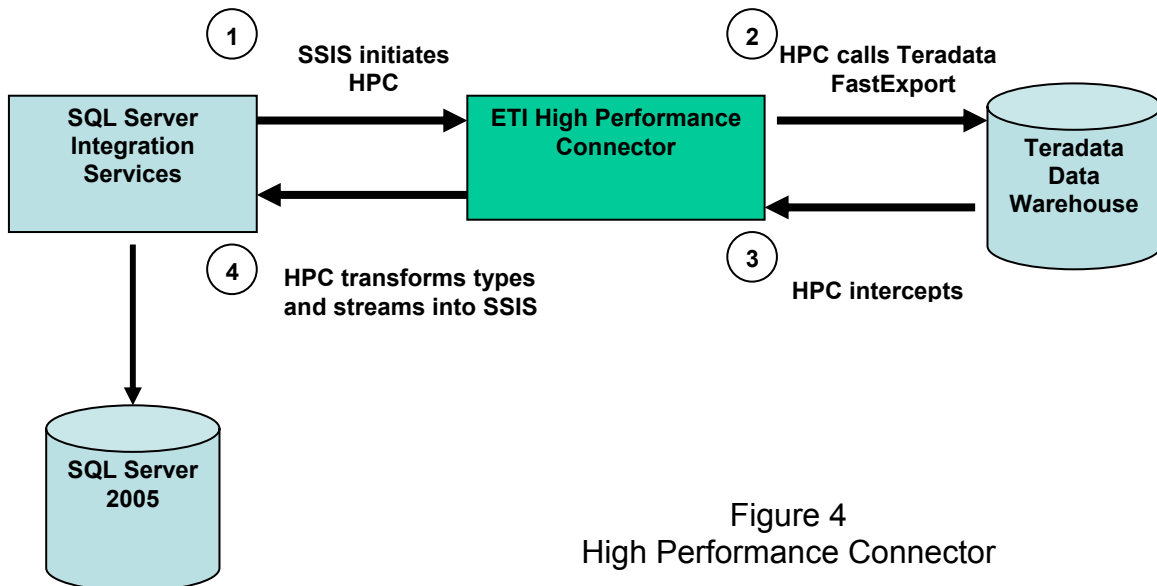


Figure 4
High Performance Connector

Test Results

The beta team performed the following test to compare the relative performance of the ETI HPC to both the OLEDB driver provided by Teradata and the FastExport utility provided by Teradata.

Test data

A table with 11.7 million rows was transferred across a Wide Area Network (WAN) to load the SQL Server Database using SQL Server Integration Services.

Configuration

Hardware

HP DL585 AMD x64 Server (4 CPUS with 32GB RAM)

½ TB on Hitachi SAN connected to server using two Emulex Fiber cards

Software

Windows 2003 x64

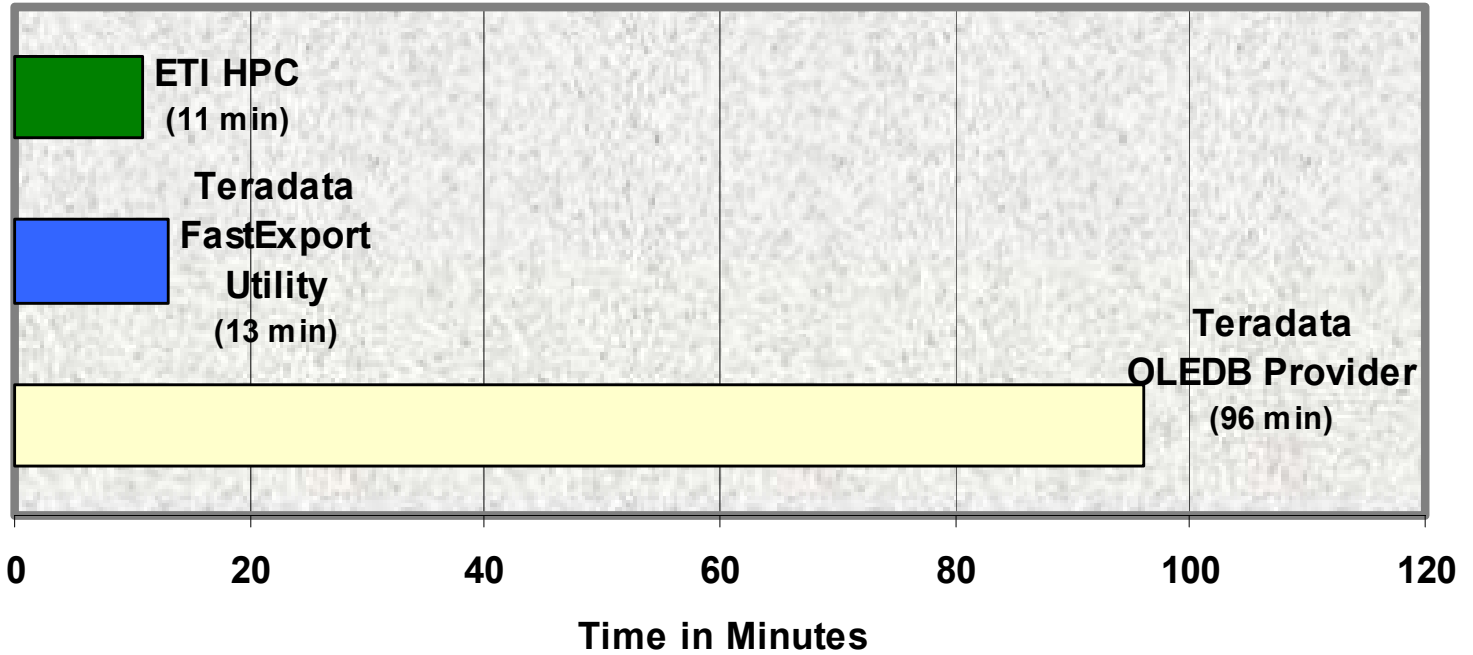
SQL Server 2005, April x64 Customer Technology Preview Build

Network

10 Mbps across WAN (measured using FTP)

Results

*Transfer Time of 11.7 Million Row Table
from Teradata in Location A to SQL Server in Location B*



*Figure 5
Throughput Results*

Not only is the standalone Fast Export Utility 10% slower than the ETI HPC, the added system administration and disk space required to bypass the HPC would make a standalone solution more costly to maintain over time.

	<i>ETI HPC Using SQL Server 2005 Integration Services</i>	<i>Teradata's Fast Export Utility</i>	<i>Teradata OLEDB Provider Using Integration Services</i>
<i>Transfer of 11.7 million row table from Teradata in Location A to SQL Server 2005 server in Location B</i>	<i>1 million rows per minute</i> Approximately 11 minutes to transfer table through Integration Services from Teradata database to SQL Server 2005 database.	<i>Approximately 900,000 rows per minute</i> Between 11 to 12 minutes to transfer table to a file between servers using Fast Export and about 1 minute to load the data from the file into a table in SQL Server using Integration Services.	<i>120,000 rows per minute</i> 96 minutes to transfer table from Teradata to SQL Server using SQL Server Integration Services.

***A SQL Server to SQL Server transfer is about 8 million rows per minute**

Conclusion

ETI and Microsoft share a common vision of what they want for the enterprise customer—scalability, usability, and a strong price/performance value proposition. "ETI Solution Version 5 is an efficient and fast way to obtain data out of a non-Microsoft based environment and into SQL Server 2005," said Donald Farmer, group program manager of SQL Server Integration Services (SSIS) at Microsoft Corp. "ETI Solution also removes the requirement to land the data prior to loading it into SSIS. This reduces the overall cost of storage, and also enables Teradata users to leverage Microsoft's cutting-edge data integration seamlessly with their data." SQL Server 2005 provides the developer with an exceptional set of tools for speeding the design and implementation of world-class BI applications. In tandem, ETI can ensure that those applications have efficient access to the data they need, regardless of where it resides in the enterprise.