

Deep Dive: Hybrid Cloud Data Integration

Three Products to Consider

KEVIN PETRIE

MAY 2021

This publication may not be reproduced or distributed without Eckerson Group's prior permission.

CUSTOM REPRINT PREPARED FOR TIMEXTENDER

TIMEX**TENDER**

About the Author



For 25 years **Kevin Petrie** has deciphered what technology means for practitioners as an industry analyst, writer, instructor, marketer, and services leader. Kevin launched, built, and led a profitable data services team for EMC Pivotal and ran field training at the data integration software provider Attunity (now part of Qlik). A frequent public speaker and the author of two books on data streaming, Kevin is also a data management instructor for eLearningCurve.

About Eckerson Group

Eckerson Group is a global research and consulting firm that helps organizations get more value from data. Our experts think critically, write clearly, and present persuasively about data analytics. They specialize in data strategy, data architecture, self-service analytics, master data management, data governance, and data science. Organizations rely on us to demystify data and analytics and develop business-driven strategies that harness the power of data. **Learn what Eckerson Group can do for you!**



About This Report

To conduct research for this report, Eckerson Group interviewed numerous industry experts and practitioners. The report was sponsored by Blitzz.io, TimeXtender, and Streamsets, who have exclusive permission to syndicate its content.

This is an excerpt from a larger report that profiles three cloud data integration products. For the full report, go to: <https://www.eckerson.com/articles/deep-dive-hybrid-cloud-data-integration>

Table of Contents

Executive Summary	4
Introduction	5
TimeXtender	10
Conclusion	17
About Eckerson Group	19
About the Reprint Sponsor	20

This is an excerpt from a larger report that profiles three cloud data integration products. For the full report, go to: <https://www.eckerson.com/articles/deep-dive-hybrid-cloud-data-integration>

Executive Summary

Data integration products have evolved to answer modern enterprise demands for automation, performance, and scale as they help businesses migrate to the cloud. Data engineers use these products to build, manage, and control data pipelines that ingest and transform data, from many sources to many targets. They support multiple use cases, ranging from batch loading to streaming, across on-premises and multi-cloud environments.

Modern data integration products help overburdened data teams at large and mid-sized enterprises increase their productivity and erase backlogs. They offer greater scalability, performance, and flexibility than prior generations of products. They offer a graphical user interface that reduces manual scripts, while still enabling data engineers to write custom code, if desired.

This report profiles three data integration products that offer distinct value. One goes broad, automating basic tasks across heterogeneous environments. Another goes deep and eliminates nearly all scripting in a given environment. A third optimizes performance and scale for a specific scenario, such as multi-target ingestion. Data leaders can use this report to evaluate these and other products' alignment with their data needs.

This is an excerpt from a larger report that profiles three cloud data integration products. For the full report, go to: <https://www.eckerson.com/articles/deep-dive-hybrid-cloud-data-integration>

Introduction

The human need to integrate data came long before computers, the telegraph, and the printing press. In 3800 BCE, scribes counted all the citizens of Babylonia on clay tiles to estimate their food requirements. Modern data integration, although more automated, remains equally challenging—and equally crucial to survival.

Data integration means combining two or more pieces of information. Today’s data engineers create and manage data pipelines that connect data from sources—such as files or databases—to targets such as cloud data platforms. They send data through those pipelines to feed operations and analytics. Finance, sales, and marketing applications all consume data, as do analytics projects that range from business intelligence (BI) to machine learning (ML) and other types of advanced analytics.

Data integration has four components. First is data ingestion, which extracts batches or increments of data (as well as their schema and metadata) from a source, then loads data to the target. Second, transformation combines, formats, structures, and cleanses data. It might also import or create data models. Third, data engineers manage their environments by designing, developing, testing, and deploying the data pipelines that ingest and transform data. They also monitor, tune, and reconfigure those pipelines. The final component, control, refers to tasks such as provisioning, version control, workflow orchestration, lineage, and documentation.

These capabilities support the so-called “extract, transform, and load” (ETL) process for batch data, as well as new variations such as ELT, ETLT, and streaming ETL.

Market Evolution

Enterprises use data integration products to support three types of business initiatives:

- > **Data modernization.** To reduce cost, improve efficiency, and enable digital transformation, enterprises are shifting to new cloud-based architectures.
- > **Data democratization.** Enterprises seek to empower managers throughout their organizations to make data-driven decisions and adopt self-service analytics.
- > **Advanced analytics.** Enterprises embrace use cases such as machine learning and artificial intelligence to create new insights—and even operationalize those insights in real time with automated action.

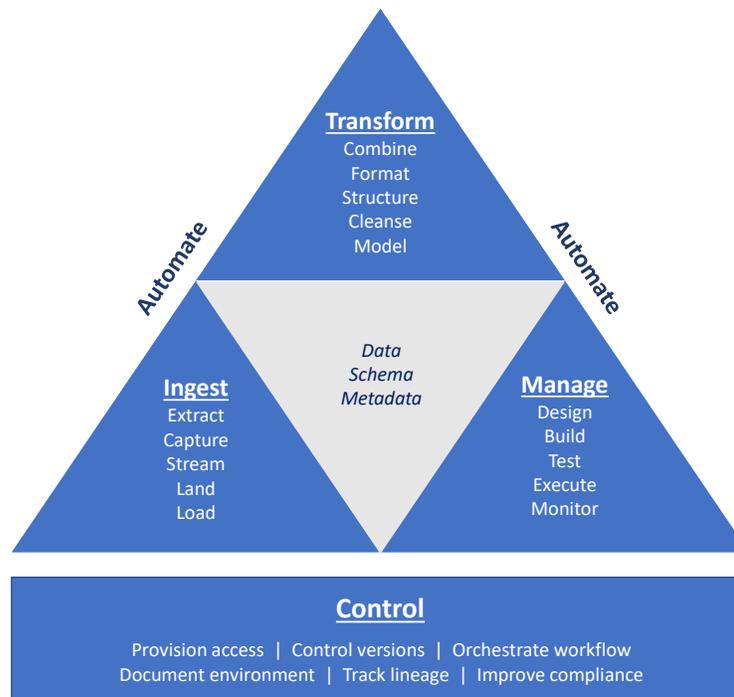
Data integration vendors have evolved to support these initiatives. Enterprises can evaluate their products by the following criteria:

- > **Platform support.** As environments grow increasingly heterogeneous, enterprises need data integration products that (1) connect many sources and targets or (2) provide specialized capabilities for selected platforms.
- > **Automation.** Many data engineers crave ways to reduce cumbersome, error-prone manual scripting processes with graphical, drag-and-drop interfaces. Automation also helps data scientists and analysts to perform basic data integration tasks.
- > **Streaming.** Enterprises need capabilities such as changed data capture to stream incremental updates from source to target rather than just loading batches. This supports high data volumes and real-time analytics.
- > **Cloud affinity.** As data teams shift operational and analytics workloads off-premises, they seek data integration products that align with the cloud. This includes cloud delivery models and usage-based pricing, as well as architectures that take advantage of elastic cloud compute resources.

Data Integration Framework

The following framework (Figure 1) helps visualize the components of data integration.

Figure 1. Data Integration Framework

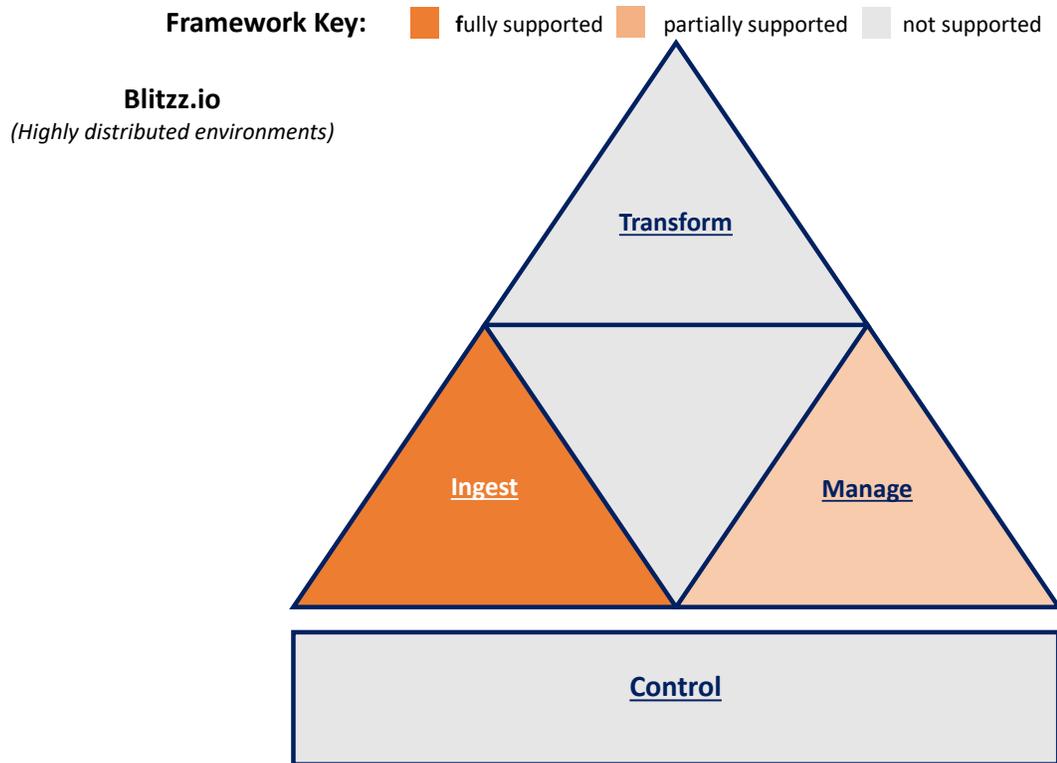


This framework shows how data integration products ingest and transform data, as well as schema and metadata, within a data pipeline. The sequence of these steps varies by enterprise requirements. You might transform data during the loading or streaming steps, or afterwards, depending on your workload requirements, latency needs, and so on. The “Manage” category shows the basic aspects of data pipeline management. “Control” functions help manage multiple data pipelines and stitch them into business and IT processes.

Three Products, Three Approaches

This report examines three products that take different approaches to data integration to help you understand the range of options and evaluate products that meet your needs. Together these three products cover the major categories of the market.

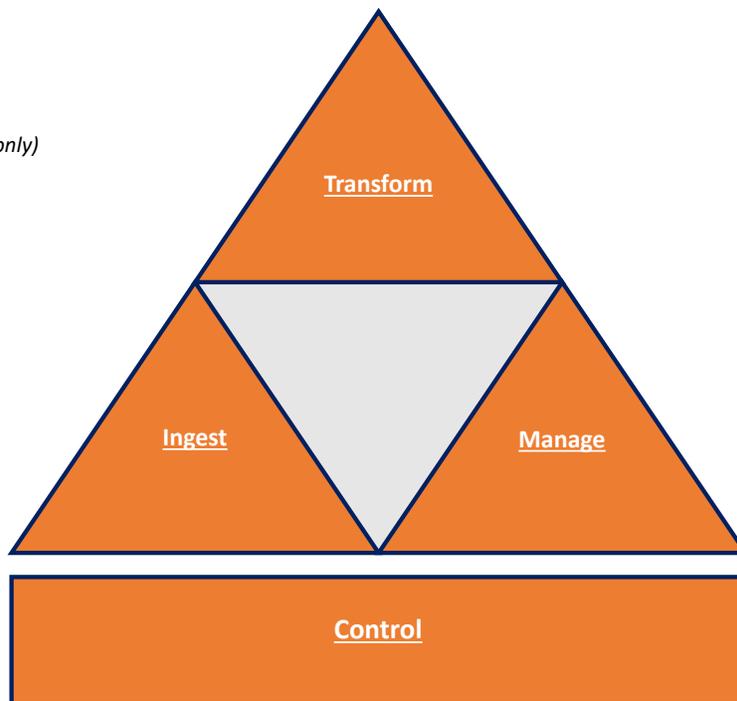
We will take a deep dive into products from Blitzz.io, StreamSets, and TimeXtender. Each targets a distinct customer type, solves a distinct problem, and therefore offers a unique set of capabilities. All three products address the core aspects of data ingestion, but only one offers full data integration functionality (albeit for a single type of environment). Whether narrowly or broadly focused, each vendor has specialties that differentiate it from the market landscape.



Blitzz.io Blitzz helps enterprises migrate on-premises data to distributed databases in hybrid, cloud, and multi-cloud environments. Data engineers use Replicant’s cloud-based architecture for multi-target data ingestion to support operational and analytical workloads. Although Blitzz supports fewer end points than older data integration vendors, it offers a scale and performance advantage for ingestion from on-premises sources such as Oracle and SQL Server, to distributed database targets such as MongoDB, CockroachDB, and Cassandra. Its Replicant product best serves data engineers and analysts that struggle to maintain accurate, real-time data across multiple copies, partitions, clusters, regions, and clouds.

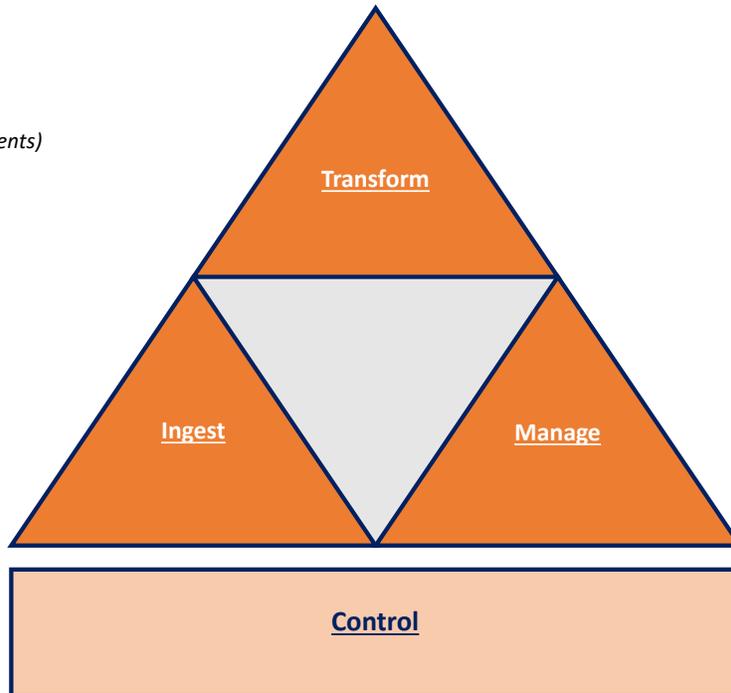
TimeXtender helps data engineers manage all aspects of data integration for on-premises, cloud, and hybrid environments. Although TimeXtender supports many data sources, it focuses on Microsoft targets. This product includes a graphical interface that overlays and automates the ingestion, transformation, and management capabilities of Microsoft tools such as SQL Server Integration Services and Azure Data Factory. TimeXtender caters to midsize enterprise users that want to simplify how they integrate structured data across SQL Server and Azure Synapse platforms.

TimeXtender
(Microsoft environments only)



StreamSets offers a comprehensive approach to integrating data with many data pipelines in heterogeneous environments. Enterprises use its product to configure and tune complex transformation tasks that support both standard BI and advanced analytics use cases. StreamSets focuses on continuously absorbing changes to source data, schema, and metadata so that enterprises can maintain resilient operations and analytics at scale. Its users are data engineers whose livelihoods depend on the sophistication, performance, and resiliency of fast-changing data pipelines.

Streamsets
(Heterogeneous environments)



The data integration landscape is crowded and diverse, but these three products cover the primary approaches. The following profiles explore the companies' backgrounds, target customers, product architectures, and key functionality. They should give you the right knowledge base to help evaluate other data integration products as well.

TimeXtender

Founded: 2006

Product: TimeXtender

Initial Product Launch: 2006

CEO: Heine Krog Iversen

Executive Summary

TimeXtender helps enterprises with 500 to 2,500 employees—and understaffed data teams within larger enterprises—automate how they integrate structured data in hybrid Microsoft environments. Data engineers can use a single interface to accelerate basic configurations, customize where needed, and orchestrate workflows. TimeXtender helps them overcome a history of technical debt by increasing the efficiency of data pipeline management and assisting compliance needs with lineage and documentation.

Company

Two decades ago, Heine Krog Iversen, a young entrepreneur and co-founder of an IT consulting firm, spotted an opportunity as he helped clients manage their script-intensive processes. That opportunity was automation.

Iversen and his wife Anne Krog Iversen launched TimeXtender in 2006 to help enterprises automate their bottlenecked BI processes. They started as a consultancy, but eventually developed a software product to address common customer needs for SQL Server data warehouses. They designed the product to improve efficiency by graphically configuring and automatically executing data integration tasks.

Iversen continues to evangelize the business benefits of automation. He positions TimeXtender as a holistic approach to data access, modeling, and compliance for Microsoft environments. Because it focuses on Microsoft, TimeXtender can go deeper to automate most or all aspects of data integration. TimeXtender solutions director Joseph Treadwell estimates that data engineers eliminate 95% of manual coding and scripting. This accelerates data pipeline design, workflow orchestration, and documentation, and simplifies compliance procedures—especially for environments that layer SQL data warehouse structures onto cloud object storage.

TimeXtender adapted its business model as it grew rapidly over the last decade. In 2015, the company formally exited the consulting business and decided to sell exclusively through value-added resellers. Today, TimeXtender has more than 3,300 customers and more than 200 partners supporting customers in 95 countries. Its product supports Microsoft targets and more than 200 data sources.

Customers

TimeXtender caters to enterprises with 500 to 2,500 employees—and understaffed data teams within larger enterprises—that need to streamline their data processes for Microsoft platforms. Typically, these data teams straddle hybrid environments. They migrate analytics and operational workloads to cloud targets such as Azure Data Warehouse, but maintain data pipelines back to on-premises sources. TimeXtender sells to all verticals, especially financial services, healthcare, government, retail, manufacturing, and professional services. Their buyers are data and IT leaders, influenced by business leaders, and their users are data engineers and architects.

TimeXtender's ideal customer has a history of technical debt. Data engineers may have hand-crafted brittle transformation scripts with little attention to compatibility or documentation. Those scripts broke when they encountered new data sources, pipelines, or versions of BI tools, forcing them to cancel upgrades or write new code from scratch. Other customers have connected BI tools directly to data sources, creating silos and bottlenecks.

Data teams implement TimeXtender to alleviate these problems in three ways. First, they start fresh by building a new Azure Data Warehouse (DW) or possibly an Azure Data Lake. They ingest and transform structured data in this new platform to address new use cases. Second, data teams modernize an existing data warehouse environment. They migrate the data, schema, and models from on-premises SQL Server or legacy systems into a consolidated Azure DW. Third, TimeXtender users land, format, and cleanse data in platforms such as Azure Databricks as preparation for machine learning or other advanced use cases.

The ideal TimeXtender user is a data engineer who handles projects like these for structured data. This user needs a flexible, automated product to speed up all aspects of data integration across Microsoft environments that contain data warehouse structures and data lake storage.

Product

TimeXtender provides a graphical interface that enables data engineers and architects to address all four components of our data integration framework: ingest, transform, manage, and control. Data engineers use the TimeXtender GUI to configure commands in SQL Server Integration Services (SSIS) or Azure Data Factory (ADF). TimeXtender then automatically generates the code they would otherwise need to hand-script line by line in ADF.

The TimeXtender product has an operational data exchange (ODX), modern data warehouse (MDW), and semantic layer. Each of these components aims to speed things up by making default configuration assumptions that users can approve or override. Together they enable users to design, create, execute, and monitor data pipelines that support ad hoc BI queries, reporting, and advanced analytics.

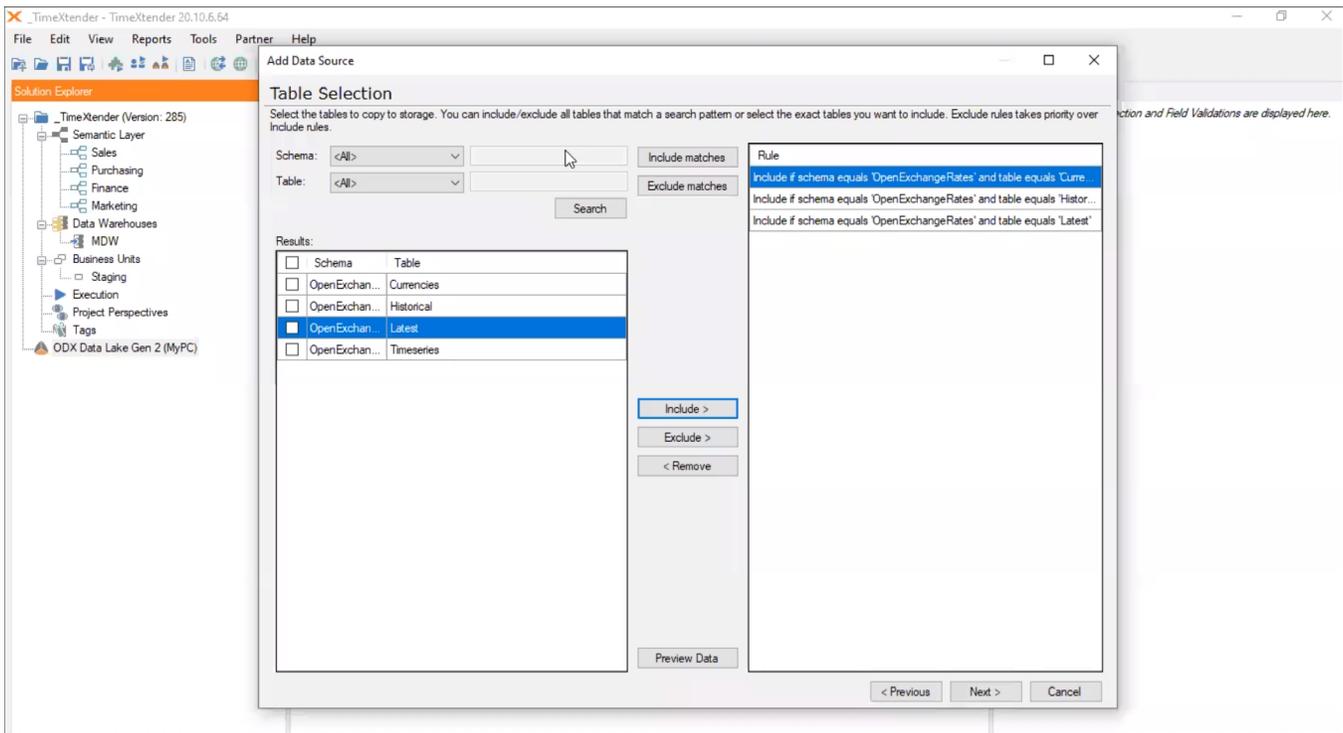
TimeXtender supports the ELT approach by default, with some customers opting for ELTLT instead. Users extract data from the source into the ODX, then load and transform it in the MDW before presenting it to the semantic layer for analytics. Although users configure and execute all major tasks through TimeXtender, they toggle over to Microsoft tools such as ADF to check execution status and results.

Operational Data Exchange

The ODX ingests data, schema, and metadata by landing it in SQL DB or more commonly Azure Data Lake Storage (ADLS). It discovers and lists the available data sources, ranging from databases to files to public data sets, for perusal and selection by the user. A popup window prompts users for credentials, then connects them to the selected sources and any necessary application programming interfaces (APIs). The interface preselects all data within each source for simplicity, but still enables the user to filter and de-select tables by schema and name.

Figure 2 illustrates the selection of tables for landing in the ODX.

Figure 2. Table Selection for Landing in the ODX



Next, users extract the data. TimeXtender configures the transfer and ongoing refreshes of each source and table with a few drag-and-drop steps, once again offering default assumptions that can be adjusted. TimeXtender users can check the status of data transfers in the Azure portal, then view their imported tables and incremental updates in ADLS. TimeXtender supports frequent updates to data, schema, and metadata. It does not provide real-time streaming because its users prefer to review and accept incoming data before sharing it with the business for analytics.

Data engineers gain two advantages by configuring data transfers in TimeXtender rather than ADF. First, they eliminate most or all manual scripting. Second, TimeXtender automatically absorbs ongoing source updates. It automatically adjusts to new or changed schema—fields, tables, and so on—during each scheduled refresh, so users can simply accept or decline the changes. ADF, in contrast, requires users to reconfigure the source manually for each change.

TimeXtender also simplifies data transfers between multiple environments to support the DevOps practice of continuous software integration and delivery. You can use a single interface to move data or metadata from development to test to production—for example, to assist frequent code iterations for machine-learning projects.

Modern Data Warehouse (MDW)

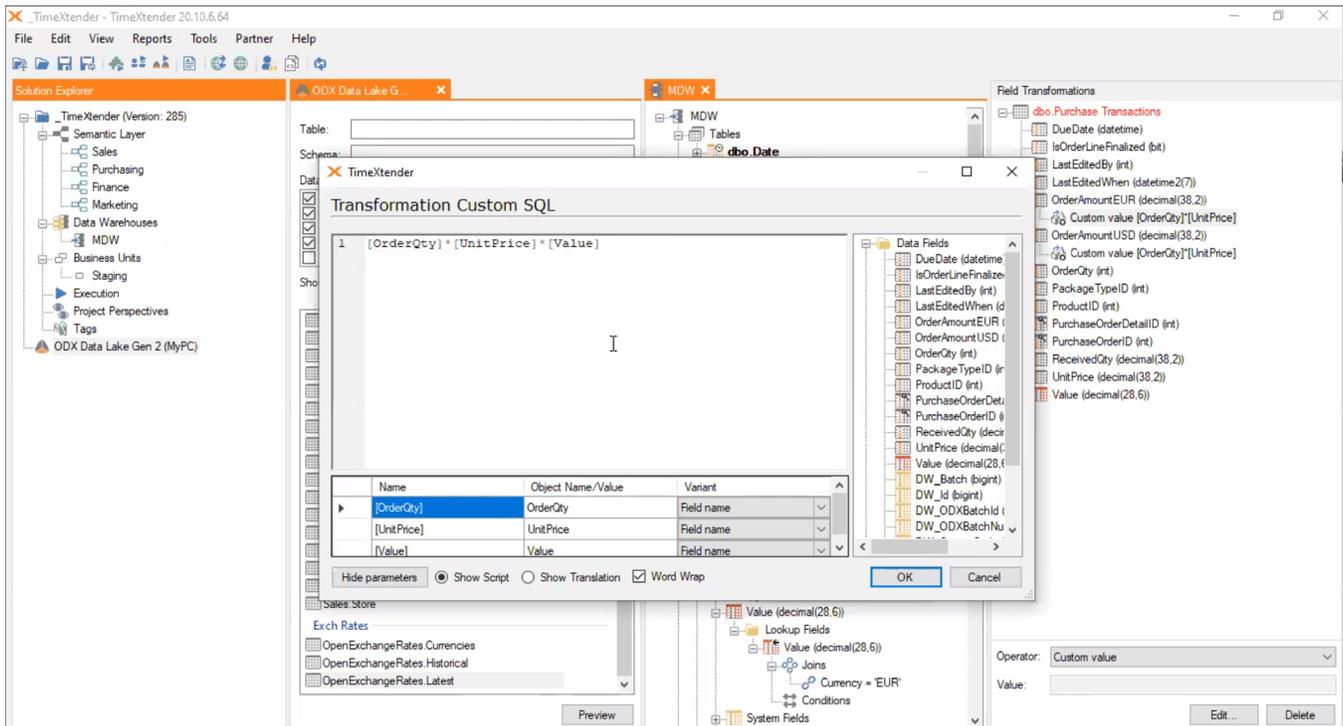
Most TimeXtender customers choose to transform data in the MDW because it offers the data warehouse constructs they need to prepare structured data for BI and advanced analytics use cases.

To start this process, users discover, select, and drag-and-drop the raw tables and fields from the ODX to the MDW. TimeXtender configures transformation tasks by preselecting options such as target data type and format, then letting you change those options if necessary. To combine data sets, you drag one table onto another, then follow the prompts in a popup window to configure your transformation logic. For example, the window will auto-populate script text for table joins based on a few clicks, then let you customize that script if necessary. You can also create or import a model that defines how tables, columns, and other elements relate to one another.

TimeXtender also supports more complex transformations, such as the matching and merging of multiple source tables into golden records. With a few clicks and keystrokes, you can create logic to look up, compare, and reconcile rows. Once the job executes, the resulting table fills gaps and removes duplicates in the source tables. You also can configure transformation logic to provide analysts with time-stamped versions of records. If a slowly changing dimension such as a customer's postal address changes, analysts will know when to include that customer in one regional study versus another.

Figure 3 illustrates the generation of transformation logic for user review, customization, and/or approval.

Figure 3. Generation of Transformation Logic



Semantic Layer

The semantic layer presents refined data—grouped into business-oriented semantic models—to familiar analytics tools such as Microsoft Power BI, Tableau, and Qlik. Data engineers can create these semantic models from tables, columns, or other entities in the MDW. This might be as simple as dragging a table from the MDW to the semantic layer, then defining its format and aggregation method in a popup window. As elsewhere, TimeXtender automates the basics while supporting optional customization. You can provision one semantic model to multiple BI tools in a single task, then let TimeXtender automatically integrate with the distinct syntax of each tool.

In the semantic layer and throughout the product, users maintain version control, including the ability to track different semantic models, annotate changes, and roll back to earlier versions. Users can also move semantic models between environments without manually rewriting code. When you drag a model from SQL data warehouse and drop it into SQL DB, TimeXtender automatically redeploys the underlying code base.

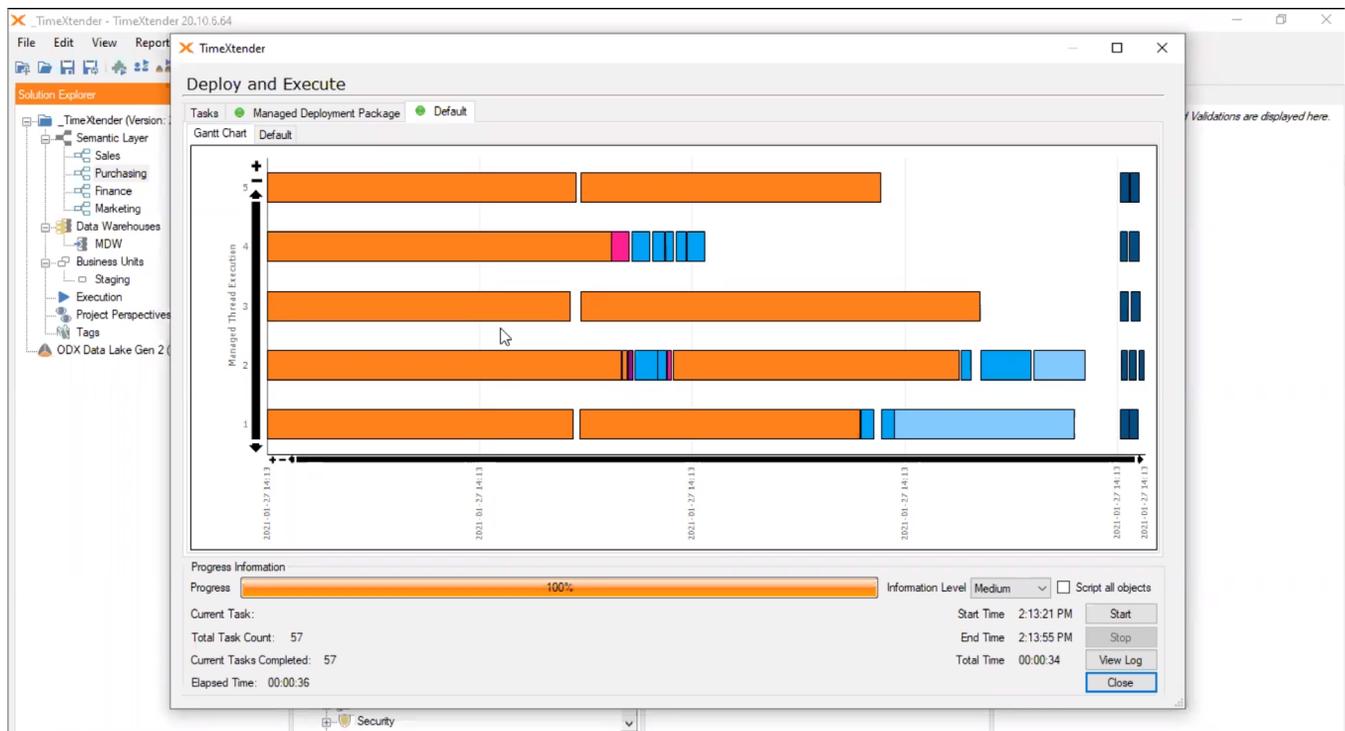
Deployment and Execution

Once users have configured their data pipeline from source, to ODX, to MDW, and the semantic layer, they click to deploy and execute. TimeXtender orchestrates the ingestion and transformation of all those

entities in multiple concurrent threads. It also assesses how these tasks and tables relate to one another, reordering tasks if necessary to improve efficiency. Users can view workloads for each thread, then scroll over a given thread to check the task status and execution time.

Figure 4 shows how TimeXtender orchestrates and visualizes task execution.

Figure 4. Task Execution View



Documentation and Compliance

TimeXtender provides additional visibility and documentation to support compliance and operational needs. It maintains a current data dictionary that describes data elements and metadata. Users can track the lineage of tables or columns from source to BI analyst, including all the ingestion and transformation tasks performed. With a few clicks, you can generate a full PDF document of your TimeXtender environment that provides business descriptions of tables, mappings, joins, custom transformation logic, and more.

Pricing

Product pricing is straightforward. TimeXtender charges a monthly subscription of \$3,000 per production instance of the data warehouse. This includes all features with unlimited data, users, and preproduction

environments. Customers also get discounts for prepaid, multi-year packages. They pay \$32,400 up front for one year, \$86,400 for three years, and \$126,000 for five years, for effective discounts of 10%, 20%, and 30%, respectively. Users can deploy TimeXtender on-premises or host it in the cloud.

Recommendation

TimeXtender helps data engineers and architects simplify how they manage and control data pipelines in hybrid Azure environments. It strikes a balance by fast-tracking basic configurations while allowing users to customize where needed. Its GUI, default assumptions, and orchestration features enable users to implement data flows faster and with less training than they would otherwise need to script the same jobs in SSIS or Azure Data Factory.

The ideal TimeXtender customer is a midsize enterprise with an understaffed data team that needs to:

- > Minimize traditional scripting processes to meet rising analytics demand.
- > Consolidate structured data from many operational sources onto Azure for analytics.
- > Manage artifacts and models across Microsoft platforms with a single interface.
- > Document the environment and track lineage to improve compliance.

Conclusion

Data integration tools have come a long way since Babylonian scribes counted citizens on clay tiles nearly 5,000 years ago, and even since enterprises started loading up data warehouses 30 years ago. Today they offer many ways to manage rivers of data, schema, and metadata. They ingest data both via real-time streaming and traditional batch loads, and connect new sources and targets that include IoT sensors, SaaS applications, and cloud databases. They transform data, manage data pipelines, and help control the complex environments in which they operate. Enterprises derive three primary benefits from data integration tools:

- 1. They simplify the data integration process.** Data engineers and analysts can configure, execute, monitor, and reuse standard pipelines with a graphical interface rather than scripted commands—but still customize where necessary. They consolidate data from many sources to one target, or broadcast one source to multiple targets. Some tools provide consolidated views, track lineage, and document environments to assist planning, monitoring, and compliance.
- 2. They improve performance and scale.** Data streaming enables real-time analytics and operations, and reduces the need to repeatedly load batches of unchanged data. Some tools spread processing across elastic compute resources to optimize latency and throughput. This supports high-performance ingestion to multiple targets, such as the partitions and replicas within a distributed database.
- 3. They help companies migrate to the cloud.** Data integration tools load batches of on-premises data to cloud targets, then stream continuous updates to synchronize datasets. This helps enterprises run analytics and operational workloads in hybrid, cloud, and multi-cloud environments.

These data integration products focus on distinct problems and offer distinct features. Some products go broad, automating basic tasks across heterogeneous environments. Other products go deep and eliminate nearly all scripting in a given environment. Still others optimize performance and scale for a specific scenario, such as multi-target ingestion to the distributed cloud database. Refer to the following buyer's guide questions as you compare other products to those profiled here.

Buyer's Guide Questions:

Does this product support your sources and targets?

Ideally your product should integrate with as many of your end points as possible. Carefully consider your likely future data sources and targets to reduce the need to buy another data integration tool next year.

Is the product architected to meet your performance requirements?

Evaluate how this product can ease or eliminate inevitable bottlenecks as data volumes rise. Take a hard look at your requirements for new targets such as distributed cloud databases. Certain products specialize in multi-target ingestion at low latency and with high throughput. These tools are worth consideration even if they do not support all the sources and targets in your environment.

What level of technical knowledge does it require?

Assess this product's fit with your team's skills. Some data engineers know how—and even prefer—to manage pipelines via command line scripting. Many others want to reduce scripting with automation wherever possible.

How does it support customization?

Many data teams need to create, manage, and reuse standard data pipeline components, but still inject custom scripts to support machine learning or other advanced use cases. Assess how this product integrates with notebooks, libraries, and other components of the AI/ML ecosystem to make this happen.

How is this product architected for the cloud?

Cloud delivery models (SaaS or managed service) and usage-based pricing simplify how you consume and pay for your data integration product. Each vendor should offer these options or have a committed path to get there. Also look for products that take advantage of elastic compute resources to achieve scale and performance.

About Eckerson Group



Wayne Eckerson, a globally known author, speaker, and consultant, formed **Eckerson Group** to help organizations get more value from their data. His goal was to provide organizations with expert guidance during every stage of their data and analytics journey.

Today, Eckerson Group helps organizations in three ways:

- > **Our thought leaders** publish practical, compelling content that keeps data analytics leaders abreast of the latest trends, techniques, and tools in the field.
- > **Our consultants** listen carefully, think deeply, and craft tailored solutions that translate business requirements into compelling strategies and solutions.
- > **Our advisors** provide one-on-one coaching and mentoring to data leaders and help software vendors develop go-to-market strategies.

Eckerson Group is a global research and consulting firm that focuses on data and analytics. Our experts specialize in data governance, self-service analytics, data architecture, data science, data management, and business intelligence.

Our clients say we are hard-working, insightful, and humble. It all stems from our love of data and desire to help organizations turn insights into action. We are a family of continuous learners, interpreting the world of data and analytics for you.

Get more value from your data. Put an expert on your side. [Learn what Eckerson Group can do for you!](#)



About the Reprint Sponsor

TimeXtender offers the fastest way to a modern data estate on Microsoft data platforms. TimeXtender provides

The logo for TimeXtender features the word "TIME" in a grey, sans-serif font, followed by a large, stylized orange "X" that overlaps the "T" of "TENDER", which is also in a grey, sans-serif font.

a no-code environment for automating data integration, extraction, cleansing and loading that enables rapid deployment and maintenance of a data estate for analytics and AI—saving our customers time and money. TimeXtender’s metadata-based approach also automates complete documentation of the data estate, including data lineage, securing private data and maintaining security settings.

TimeXtender’s mission is to empower customers with instant access to data that is ready for analysis, enabling them to make quality business decisions with data, mind and heart. We do this for one simple reason: because time matters. A Microsoft Gold Certified Partner, TimeXtender serves its 3,000+ customers, from mid-sized companies to Fortune 500, through its global network of partners. TimeXtender was founded in 2006 and is privately owned.