Introduction

Today, enterprises produce enormous amounts of data in their daily operations. This includes transactional data generated from online and offline orders, machine data collected from equipment and sensors, and even social data created from likes and comments on social media platforms.

When all your data is organized and integrated, powerful analytics tools can generate business insights that translate into competitive advantages for your business. But without consolidated data, your data analytics will be inaccurate and incomplete.

Data integration solutions can automate data pipelines and generate a visual dataflow for priceless insights into your organization. By uncovering strategic strengths and weaknesses, you can empower improvements in your business practices and processes. The result is more informed decisions, greater employee productivity, and better data management.

However, if your data integration solution is stagnant, your company will miss new growth opportunities and lose potential revenue streams to your competitors. That's why in this guide, we'll review and compare some integration platforms for both on-premises and cloud deployment.

Data Integration Market Trends

Data Integration is the process of combining data from multiple sources into a unified view for downstream applications or analytics. It allows data teams to centralize data for other uses to maximize query performance (from the consolidated data store) and minimizing the impact on back-end data sources.

Traditionally, data integration has involved three key processing components: extracting data from the sources, transforming into the unified view, and loading the unified data into the destination. Together, these processes are referred to as ETL. As enterprises began to generate more data and expect more from it, data integration tools became more granular and user friendly.

Data integration is often used for analytics purposes, sending the resulting data to analytical databases, data warehouses, or data marts, and organizing the data in specific analytic formats (star-schemas, multi-dimensional, etc.). But new purposes for data integration include enriching data within an application such as CRM and bringing together all the data need for a new application such as customer or supply chain 360.

The rise of cloud computing has birthed cloud integration technologies. Cloud integration connects different applications, systems, IT environments, and repositories to enable the real-time exchange of data and processes. Cloud integration can consist of cloud-to-cloud integration, cloud-to-on-premises integration, or a combination of both. As Software-as-a-Service (SaaS) solutions become widely adopted, many enterprises now use a Hybrid Integration Platform (HIP) to blend their hybrid mix of on-premises applications and cloud-based SaaS.

ETL vs ELT

ETL stands for extract, transform, and load, one type of process used for data integration. Each step is interrelated and come together into a complete process to deliver a unified set of data. Extraction pulls the data from the source, Transformation shapes the data into the final form needed, and Loading puts the data into the destination data store.

ELT, or extract, load, and transform, is a new data integration model that has emerged with data lakes and cloud analytics. Data is extracted from the source and loaded into a destination still in its original or raw form. The raw data is transformed within the destination to a second form that is then ready for analytics.

An ETL process could be a single synchronous process, or steps can be separated and run individually. In the latter, there will often be an intermediate data store used to manage the in-flight data. However, ELT processes can be sophisticated, especially when dealing with data lakes, as they are often dealing with complex data formats, multiple data sources, and a great deal of preparation. Cloud analytics ELT from SaaS applications tend to be simpler since it only takes data from 1 or 2 SaaS sources.

ETL is the more traditional method of doing data integration. It has well-known best practices and a good number of tools to support it. However, developing ETL processes can be slow, cumbersome, and costly. ELT is more modern and embraces the agile methodologies of today. Teams can iterate quickly on projects, and analysts can often create their ELT pipelines without coding a single line.

Datameer Spectrum (ETL++)

Datameer Spectrum is a comprehensive ETL++ data integration platform with wizard-driven simplicity that enables faster and easier self-service data integration without writing one line of code. Once set up, Spectrum's robust operationalization and governance features enable reliable, automated, and secure data pipelines to ensure a consistent data flow.

Spectrum supports analyst self-service data preparation and data engineering use cases, enabling a single hub for all data preparation across an enterprise. With its roots in Datameer X, Datameer's premier on-premises solution, Spectrum provides you with the best ETL data integration functionality that are proven at large enterprises. What Spectrum adds is ease of use, full data preparation, elasticity, and manageability in a single package, allowing you to make your data ready for analytics 10 to 20 times faster at a fraction of the cost.

Spectrum provides a hybrid ETL and ELT platform for flexibility to support both data integration forms your organization needs on the same scalable platform. Spectrum is cloud-native on all three major cloud platforms (AWS, Azure, GCP) and carries with it the elasticity and cost economics you would expect from the cloud. Spectrum also bridges any data sources you have regardless of type, format, and location (cloud or on-premises).

Spectrum vs. Traditional Data Integration Vendors

Traditional data integration platforms typically handle data in the same straightforward way. The data is structured, which takes less space. All the data is loaded on traditional, on-premises servers within the enterprise data center. Currently, Talend and Informatica are some of the popular traditional data integration vendors:

Talend promises a comprehensive integration platform that covers a full range of integration scenarios, including real-time data pipelines, data quality and governance, and application and API integration. The Talend platform has a variety of different platform services and user interfaces/tools, with each service designed for a specific function, such as the platform for on-premises data integration.

Informatica's product family supports two main forms of data integration: ETL and point-to-point synchronization. The main data integration products work with the ETL data flow style model. While Informatica's platform is not lacking in capabilities, its many enterprise features are only available as add-ons or separate products.

Spectrum vs Talend

Integrated vs. Dis-integrated

- The Talend platform consists of a complex set of tools and services designed for specific purposes, but the services are not well integrated.
- Spectrum offers a single integrated platform and toolset that provides a unified data integration hub for any of your use cases.

User Experience

- Working with Talend requires using multiple tools, each with specific functionality to design a piece of your data pipeline or manage the pipeline.
- Spectrum provides a single, seamless user experience that supports all aspects of data pipeline creation and management, including discovery and exploration, design, testing, and deployment – without any coding.

Data Source Connectors

- Talend claims to have over 1,000 connectors AND components, which is misleading as each connector has anywhere from 10 to 15 components for specific tasks.
- Spectrum has an extensive array of over 80 connectors designed to work with different sources – databases, data warehouses, files, SaaS applications, and cloud services – in various formats – structured, semi-structured, and unstructured.

Hybrid-Cloud Support

- Talend's platform is unsuited for many hybrid environments due to the mismatches in functionality between Talend's on-premises and cloud offerings, particularly in the connectors supported.
- Spectrum's platform has the same functionality regardless of whether running it on-premises or in the cloud. It also lets you seamlessly burst or migrate pipeline workloads into the cloud.

First-Class Data Preparation

- With Talend, data preparation is an after-thought. Talend data preparation is designed to help cleanse and transform data and has a limited set of functions.
- Spectrum, data preparation is a central part of the platform and a critical piece of the user experience. The easy, spreadsheet-style user interface allows an analyst to create data pipelines rapidly regardless of technical skills.

BI Tool Integration

- With Talend, your data destination is always a database or data warehouse. This creates an extra hop for BI tools to use the data, making end-to-end operationalization difficult.
- Spectrum supports the ability to directly send data to BI platform servers in their native format, including Tableau, PowerBI, Qlik, Looker, and ThoughtSpot.

Pricing

- Talend has a per-user pricing model, which is somewhat old-fashioned. It is costly per user and requires you to purchase add-ons for additional items such as working with big data, data governance, and using their APIs.
- Spectrum offers a simple, cost-effective pricing model based on the number of users and compute resources required for data pipelines.

Comparison Table

DATAMEER SPECTRUM	TALEND
Single integrated platform that forms a unified data integration hub.	Disconnected tools and services designed for specific purposes.
A single, seamless user experience that supports all aspects of data pipeline creation and management without any coding.	A disjointed, frustrating user experience that requires users to jump around between multiple tools - each with limited functionality.
An extensive array of over 80 connectors designed to work with different sources regardless of where you are processing data.	An inconsistent set of connectors depending on which platform you are using and has an extremely limited set in the cloud.
Same functionality in all environments - can seamlessly bridge on-premises sources into the cloud and lets you burst or migrate pipeline workloads into the cloud.	Difficult to run in a hybrid environment - often impractical due to the mismatches in functionality between on-premises and cloud offerings.
Directly send data to all the BI platforms in their native format, simplifying end-to-end processes, lowering your costs, speeding data delivery.	No integration with BI tools, forcing an extra hop for data delivery to BI users and slow down the creation of operationalized BI data pipelines.
Simple pricing model that can start small, scale effectively with an organization, and help onboard the broader analytics and data community.	Complex, expensive pricing model that disincentivizes organizations to add their broader analytics and data community.

Spectrum vs Informatica

Integration Approaches

- Both Informatica PowerCenter (the legacy data integration product) and Cloud Data Integration offers a very traditional ETL approach.
- Spectrum supports a more general approach to data integration, freeing designers to piece together dataflows that best support their needs. This includes both ETL and ELT (extract, load, and transform) approaches, or just general-purpose orchestrated data pipelines.

User Experience & Ease of Use

- Informatica's user interface is workflow/dataflow oriented, where the user adds individual tasks to extract, transform, and load the data.
- Spectrum's wizard-led extraction and loading, easy spreadsheet-style interface, over 300 powerful single-click functions make it faster and easier to design ETL flows.

Data Preparation & Transformation

- Informatica is designed to support standard data transformation needs specific to data integration. It contains 110+ basic transformation functions and another 30 higher-level tasks.
- Spectrum has integrated data preparation capabilities and a library of over 300 functions to enable complete data preparation and transformation capability within data pipelines.

Asset Management

- Informatica and Spectrum provide similar folder and component asset organization and management of data integration components.
- Spectrum takes this model a step further. It automates and persists managed datasets within the flow and provides data cataloging and search features.

Job Execution

- Informatica only has its proprietary Blaze engine available for on-premises deployments.
- Spectrum is specifically designed to use scalable Spark clusters as the job execution engine.

Security & Governance

- Informatica requires add-ons for advanced security and a separate additional product for data governance.
- Spectrum offers integrated security and governance while providing enterprise-class security and governance capabilities.

Pricing & Packaging

- Informatica's pricing has limited transparency. Their on-premises pricing is expensive, and their Data Integration Essentials leaves out many of the enterprise features.
- Spectrum provides integrated enterprise features and more, all integrated into a single platform with transparent, consumable cloud-based pricing.

Comparison Table

DATAMEER SPECTRUM	INFORMATICA (POWERCENTER AND CLOUD DATA INTEGRATION)
Supports multiple forms of integration including ETL, data integration, data preparation, and data engineering.	Only supports ETL and point-to-point (no transformations) integration.
Supports ETL, ELT, data preparation, data engineering and data science pipeline jobs.	Supports only ETL jobs.
Has an easy-to-use spreadsheet-style, and wizard-driven user experience that is code-free; Includes a large library of over 300 functions and tasks	Offers a more complicated user experience that is dataflow oriented and code-free; Includes a limited library over 140 functions and tasks
Supports both asset tagging and annotations, and full search to discover assets	Only supports asset tagging and has no asset search
Integrated enterprise security.	Extra cost for enterprise security.
Integrated data governance.	Separate product with additional cost for data governance.
Uses fully elastic Spark cluster to execute jobs (included).	Extra cost for elastic cluster, uses proprietary engine for jobs (included).
Offers a simple pricing model, at competitive cloud-based pricing in a single package.	Offers a complex pricing model that can be expensive, with many add-ons costs or additional products for enterprise features.

Spectrum vs. Cloud Data Integration Vendors

The emergence of cloud data integration tools means organizations can now connect different applications, systems, repositories, and IT environments for real-time data processes. Businesses can also integrate their cloud applications with traditional on-premises systems. Currently, Fivertran and Matillion are some of the popular cloud data integration vendors:

Fivetran is a cloud-based ELT data integration platform that offers a simple, reliable way to replicate and synchronize data into your cloud data warehouse (CDW). It is a basic, reliable service that lets you set up “connections” between your data sources – primarily SaaS applications, cloud services, and cloud databases – and your cloud data warehouse.

Matillion is one of the younger, cloud-based ETL solutions on the market. It consists of three components: the underlying platform, a graphical data orchestration tool, and a management tool.

Matillion does not have a storage and execution engine, and all data processed in a data flow is stored in its intermediate form in your cloud data warehouse tables.

Spectrum vs Fivetran

Data Integration Capabilities

- Fivetran supports one simple data integration pattern – ELT.
- Spectrum supports a wide range of data integration patterns: ETL, ELT, data preparation, and data science pipelines.

Operationalization

- Fivetran supports a simple operationalization model – each connection has scheduled synchronizations.
- Spectrum has a scalable, flexible job execution system for operationalizing jobs that can manage large volumes of data, run reliably, is extremely easy to operate, and connects to enterprise management tools.

Security and Governance

- Fivetran offers a few key security capabilities, including encryption, column hashing and blocking, role-based access control, and SSO/SAML.
- Spectrum provides enterprise-class security and governance features, such as Encryption, data masking and anonymization, Kerberos integration, fine-grained access controls, integration with enterprise security (LDAP/AD, SSO/SAML), and more enable complete data security and privacy.

Connectivity and Hybrid Cloud

- Fivetran supports an extensive suite of intelligent connections to data sources, especially SaaS applications, cloud services, and cloud databases. It lacks connectivity and bridging to on-premises and hybrid data sources, particularly enterprise data warehouses.
- Spectrum supports a wide range of data sources, both cloud-based and enterprise, including enterprise data warehouses and data lakes – Teradata, Netezza, Hadoop, Hive. It also can use secure protocols, enterprise security controls, and encryption to access and bridge on-premises sources to the cloud securely.

Transformation and Data Preparation

- Fivetran's transformation capabilities require writing sophisticated SQL code or using dbt packages.
- Spectrum includes a powerful yet easy-to-use data preparation capability that allows analysts and data scientists to shape data to their needs without any coding.

Hidden Costs

- The "T" in Fivetran's ELT is performed in your cloud data warehouse, such as Snowflake, and creates hidden CDW costs.
- Using Spectrum's ETL model, all transformations, joins, and aggregations to make the data analytics-ready are performed in-transit within Spectrum, using its compute infrastructure, which is included in the transparent pricing model.

Comparison Table

DATAMEER SPECTRUM	FIVETRAN
Provides a fully-featured data integration and pipeline platform to bridge all your data sources and create analytics-ready data in minutes.	Fivetran is a simple, reliable service for replication, synchronizing, and transforming data into a cloud data warehouse.
Supports many data integration capabilities, including ETL, ELT, data preparation, data engineering, and data pipelines for data science.	Supports a simple ELT model for integration for individual data sources that cannot combine multiple sources into a common analytics view.
Offers easy to manage, scalable operationalization of data integration jobs to ensure reliable data delivery for analytics.	Offers a simple operationalization model for data replication and synchronization connections.
Supports a deep suite of security and governance features, evolved through work with large enterprises in highly regulated industries.	Provides cloud-based security and managed security operations for the running service.
Effortlessly bridges your on-premises and hybrid data sources to the cloud with secure, scalable access and protocols.	Lacks connectivity to and the ability to integrate on-premises data sources into the cloud.
Includes a rich array of over 300 point-and-click transformation functions, all usable without writing any code.	Supports SQL coding for transformations or the use of single-source dbt packages.
Performs all data transformations and integrations in-flight, using its own compute engine, and lands only analytics-ready data, eliminating hidden CDW costs or credit burning.	Forces you to store duplicate copies of data and perform all transformation and integration in the target CDW ringing up extra CDW costs or credit burns.

Spectrum vs Matillion

Faster Data Pipeline Definition

- Matillion's data orchestration tool creates overly complex data flows with many components because users must string together many components in even a simple data flow, among other inefficient processes.
- Spectrum's spreadsheet-style UI is the antithesis of complex data flow-style ones. The UI makes it extremely easy to put together all the transformation operations needed, and the interactive nature of it allows a user to see the impact of each operation.

Greater Sophistication, Same Simplicity

- Matillion offers a limited set of 75 components, of which only 25 are dedicated to data transformation.

- Spectrum supports a deep library of close to 300 functions, each applicable graphically without coding, to tame even the most complex data for more sophisticated data engineering tasks.

First-Class Data Preparation

- Matillion's data preparation is performed with the limited set of 25 transformation components in their library. There are no capabilities for basic areas such as data cleansing and de-duplication.
- With Spectrum, data preparation is a central part of the platform and a critical piece of the user experience. The easy, spreadsheet-style user interface allows any analyst to create data pipelines rapidly regardless of technical skills.

Hybrid-Cloud Support

- Matillion offers connectors that can work with more traditional data sources such as Oracle, Teradata, Netezza, and others typically on-premises. But don't count on Matillion to create pipelines to integrate data from your on-premises sources.
- Spectrum uses secure protocols, enterprise security controls, and encryption can access and bridge on-premises sources to the cloud securely and have robust data retention policies. Spectrum also lets you seamlessly burst or migrate pipeline workloads into the cloud.

Robust Security and Governance

- Matillion offers only minimal security capabilities with your basic user and role-based controls, LDAP integration, and Single Sign-On (SSO). It does not provide encryption or obfuscation and has no data governance features.
- Spectrum provides enterprise-class security and governance features, including encryption, data masking and anonymization, Kerberos integration, fine-grained access controls, integration with enterprise security (LDAP/AD, SSO/SAML), and more enable complete data security and privacy.

Single Hub Supporting Many Use Cases

- Matillion is only suitable for data integration of cloud and SaaS data sources into a cloud data warehouse. Matillion does not offer sophisticated functions for real data engineering, algorithmic or encoding functions for data science, and robust security to bridge a hybrid cloud.
- Spectrum offers a single integrated platform that offers all the capabilities for many different use cases, providing a versatile, unified data integration hub. Data scientists and engineers alike can make use of the spreadsheet-style UI and wizard-driven operations.

BI Tool Integration

- With Matillion, your data destination is always a cloud data warehouse or data lake (Databricks). Because Matillion relies on the CDW for its storage and processing, it creates an extra hop for BI tools to use the data, making end-to-end operationalization with BI processes difficult.
- Spectrum supports the ability to directly send data to BI platform servers in their native format, including Tableau, PowerBI, Qlik, Looker, and ThoughtSpot. For operationalized BI use cases, this simplifies the end-to-end process, lowers your costs, and speeds up data delivery.

Greater Scalability and Performance

- Matillion itself is a single cloud compute server instance only routing jobs. There is no intelligence as to how to break down jobs optimally for execution, which can severely limit Matillion's scalability and performance and can force jobs to take a back seat if other database workloads take priority.
- Spectrum runs its own elastic Spark-based compute cluster under the covers to give jobs the scale and performance they need automatically. The patented Smart ExecutionTM intelligently breaks down jobs into smaller components and executes them in a parallelized, optimal way.

No Hidden Costs

- Matillion's reliance on your cloud data warehouse creates hidden CDW costs above and beyond the Matillion costs. Data transformation often requires many-way joins and unions, and aggregations which are extremely "expensive" compute operations in a CDW.
- Using Spectrum's ETL model all transformations, joins, and aggregations to make the data analytics-ready are performed in-transit within the Spectrum Spark-based elastic compute cluster, which is included in the transparent pricing model.

Comparison Table

DATAMEER SPECTRUM	MATILLION
Interactive, spreadsheet-style UI makes Spectrum fast and easy to apply operations on the data and create data pipelines.	Offers a complex data flow UI for data pipeline orchestration, making it difficult and complicated to create data pipelines.
Offers a deep library of 300+ functions within the same easy-to-use UI, allowing data engineers to quickly craft more sophisticated data pipelines.	A small set of components (75) with limited transformation capabilities makes it impossible to create sophisticated data engineering pipelines.
The easy, spreadsheet-style user interface allows any analyst to create data pipelines rapidly regardless of technical skills.	Because of its limited set of components and transformation capabilities, and complex UI, Matillion is not suitable for data preparation.
Uses secure protocols, enterprise security controls, and encryption to bridge on-premises sources within your hybrid cloud environment.	Offers no enterprise-grade security capabilities making it risky to use in bridging on-premises data in a hybrid cloud.
Offers encryption, data masking and anonymization, Kerberos integration, fine-grained access controls, integration with enterprise security (LDAP/AD, SSO/SAML).	Offers only minimal security capabilities and lacks enterprise security capabilities like encryption and obfuscation, no data governance features.
Suitable for different use cases with a versatile, unified data integration hub for data integration, self-service data preparation, data engineering, data science, and hybrid-cloud.	Ideal use is for data integration of cloud and SaaS data sources into cloud data warehouse, does not support data engineering, data science, and hybrid cloud use cases.

Supports the ability to directly send data to all the popular BI platforms in their native format, simplifying end-to-end processes.	Lack of integration with BI tools makes challenging to create operationalized BI data pipelines.
Has an automated, scalable Spark-based elastic compute cluster and patented Smart ExecutionTM optimizer to maximize scalability and performance.	Relies on a CDW for processing and does not have a robust optimizer, giving it performance and scalability issues.
Supplies its own Spark-based elastic compute clusters included in the price to eliminate hidden CDW costs.	Relies on a CDW for its processing generating extra hidden costs on top of the Matillion costs.

Conclusion

To ensure your organization is maximizing the value of business data, you need the right data integration solution. You need an ETL solution that offers an easy-to-use, extremely agile dataflow definition user experience, integrates packaging on an elastic cloud-native infrastructure, and has transparent pricing you can expect from the cloud data integration.

That's why Datameer Spectrum is the ideal solution for your enterprise.

Datameer Spectrum provides all the critical enterprise-grade ETL data integration and pipeline capabilities as fully featured data integration platforms, such as Talend and Informatica. However, Spectrum offers a comprehensive set of capabilities but provides it in a well-integrated platform with a seamless, easy user experience.

Spectrum offers something for everyone – data engineers, analysts, and data scientists – and provides a shared hub for this entire community to collaborate with a pricing model that fosters more extensive team usage.

To experience Spectrum firsthand, you can request a personalized demo with one of our experts. Or you can visit our Datameer Spectrum microsite to learn more.