

(<https://blog.appliedai.com/etl-tools/>)

# UNDERSTANDING ETL TOOLS

*By Francesca Ling ([https://Blog.appliedai.com/Author/Francesca\\_ling/](https://Blog.appliedai.com/Author/Francesca_ling/)) | April 30, 2018 (<https://Blog.appliedai.com/Etl-Tools/>)*

ETL, or Extract, Transform, Load is the process of integrating data from multiple applications (systems), converting them to a single format or structure and then loading the data into the target, often a data warehouse. This process is essential for data analysis, business intelligence, and other related tasks – particularly in businesses with a wide range of data sources and formats to consider.

Selecting the right tool to do so is integral to ensuring the success of not only the specific action, but also for the overall goals and efforts of the business. To learn more about ETL and get a better understanding of how businesses use it, visit our ETL blog post (<http://blog.appliedai.com/etl>).

## The ETL Ecosystem

In the past it was common for businesses to have several ETL tools that operated independently of one another. Today, it is becoming increasingly common to have **a single ETL tool as part of a greater data integration effort**. To do so allows these processes to be seen as contributing to the overall profitability of the organization – instead of as a siloed data project.

*Source: SAS*

To help businesses who are looking towards a greater data integration strategy, it can be helpful to isolate single components to better understand their part in the whole – as is the case with ETL tools.

## Choosing an ETL Solution

The world of ETL tools has greatly evolved over the years to include a much wider range of capabilities and setups. Many come in cloud based versions, giving a greater degree of scalability, availability, and security; with lower infrastructure costs.

There are **a few criteria that can help you when evaluating potential ETL tools**, it is important to decide which of these will be most essential for your business needs. However, generally speaking some of the most important factors to consider include:

**TASKS:** Depending on the needs of your business, the importance of certain functionalities over others will vary. Day-to-day tasks such as data conversion, joining records, filtering, grouping, and combining data should be included with any tool. Some come with the capacity for more advanced tasks such as web methods, rebuilding indexes, handling arrays, and processing unstructured data.

**CONNECTIONS:** Any ETL solution must be able to connect to Excel, SharePoint, FIX, Salesforce, Hadoop, FTP, and others. Without this functionality, **the processing power of the tool is irrelevant as it will not be usable**. However, keep in mind that all tools can connect to a database/RDBMS, but only some have native client drivers that enable greater performance when compared with ODBC.

**WORKFLOW:** Being able to create effective workflows to organize and connect all of these tasks is key. Some of the most important workflows to establish include: constraint (criteria), branching, grouping, and looping (repeating).

**EXECUTING:** Being able to understand how an ETL package runs is essential – this includes how long it takes, when it started (and ended), who began the progress, if it was successful or not, and in the case of failure, what the error message received was. Execution also includes the capacity to run at predetermined times, restart in the case of failure, and limit the duration of the execution.

**PERFORMANCE:** This again is where the needs of your business will greatly impact your decision. For those who need greater capacity, many ETL tools include features such as bulk loading or the ability to cache the lookup table, to name a few.

**MANAGEMENT:** This can mean anything from being able to configure packages to run at the same time, to setting alert frequency, and creating different users and setting their permissions.

The value of each of these criteria against each other will vary depending on the size of your business, the goals you have for your data, and other similar factors.

## Leading ETL Vendors

Major tech companies have developed tools with incredible functionality to suit the needs of a wide range of organizations. However, a number of growing tech companies are starting to offer even more features and capabilities for specialized needs and requirements.

Some major vendors to consider when exploring ETL tools include:

Show

10

⬆

⬇

⬆

⬇

⬆

entries

Search:

Show

10

⬆

⬇

⬆

⬇

⬆

entries

Search:

Name	Year Founded	Status	Additional Features
Informatica	1993	Public	-Range of prebuilt transformations -Embeddable engine for real-time and batch data execution
Stitch	2016	Private	-Was created from RJMetrics
IBM: Infosphere Information Server	2008	Public	-Netezza integration for faster loading

Oracle Data Integrator (ODI)	2006	Public	-Separation of declarative rules from implementation details -ELT architecture can use RDBMS engine
ETLeap	2013	Private	-Data wrangling enables working off sample data alone
SAP Business Objects Data Services (BODS)	2007	Public	-Structured and unstructured data integration -Web-based DI administrator for repository management
CloverETL	2002	Private	-Open source based on Java -Has its own transformation language for complex validation rules
Microsoft SQL Server Integration Services (SSIS)	2005	Public	-Transformation is processed in the memory, making the integration process in SQL server much faster
SAS Data Management	2006	Public	-Access to Hadoop via Impala or Pivotal HAWQ -Role-based GUI with drag-and-drop functionality
Matillion	2011	Private	-Tools built specifically for Redshift, BigQuery, Snowflake

Showing 1 to 10 of 10 entries

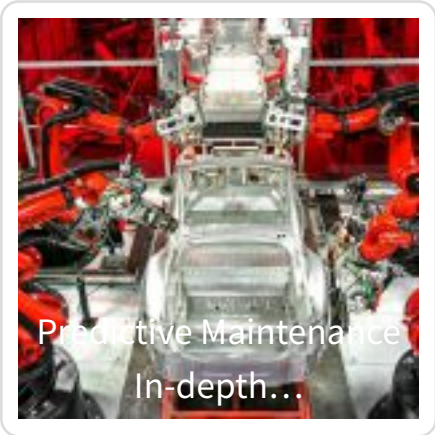
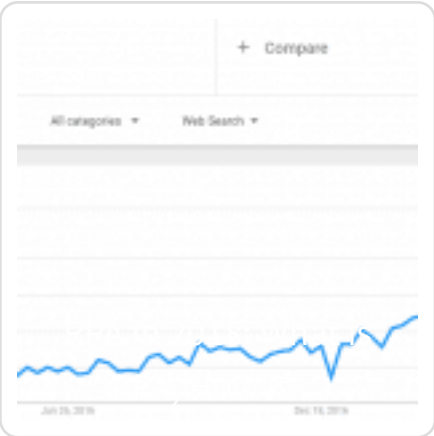
◀ Previous Next ▶

Showing 1 to 10 of 11 entries

◀ Previous [Next](#) ▶

Interested in finding out more modern ways to manage your data? Our blog (<https://blog.appliedai.com/>) covers a wide range of related topics to help you discover what new technologies could help you the most. Additionally, our list of vendors (<https://appliedai.com/vendor-selection>) is full of information about tools, solutions, and the companies that bring them to you.

Related Posts:



"Digital is the wires, but digital intelligence, or artificial intelligence as some people call it, is about much more than that. This next decade is about how you combine those and become a cognitive business. It's the dawn of a new era."

VIRGINIA M. (GINNY) ROMETTY, CHAIRMAN, PRESIDENT AND CHIEF EXECUTIVE OFFICER, IBM

(<https://blog.appliedai.com/ibm/>)

ETL (<https://Blog.appliedai.com/Category/Data/Etl/>)  
Data (<https://Blog.appliedai.com/Tag/Data/>), Data Warehousing  
(<https://Blog.appliedai.com/Tag/Data-Warehousing/>), Etl (<https://Blog.appliedai.com/Tag/Etl/>)

## EXPLORE APPLIED AI

USE CASE GUIDES

(<https://appliedai.com/use-case-guide>)

AI STARTUPS & SOLUTION PROVIDERS

(<https://appliedai.com/vendors>)

## ABOUT

BLOG

(<https://blog.appliedai.com>)

ABOUT US

(<https://appliedai.com/about-us>)

CONTACT US

(<https://appliedai.com/contact-us>)

## PLAN YOUR AI JOURNEY

VENDOR SELECTION

(<https://appliedai.com/vendor-selection>)

## OTHER

PRIVACY POLICY (<https://appliedai.com/privacy-policy>)



(<https://www.facebook.com/appliedAI>)



(<https://twitter.com/appliedaicom>)



(<https://www.linkedin.com/company/11071351>)

applied  (<https://appliedai.com>)

ALL RIGHTS RESERVED © 2017