UPSIDE
where DATA means BUSINESS

Data Wrangling Versus ETL: What's the Difference?

Three things that distinguish data prep from the traditional extract, transform, and load process.

By Wei Zheng

February 10, 2017

Over the past few years, data wrangling (also known as data preparation) has emerged as a fast-growing space within the analytics industry. Once an analysis bottleneck due to painful, time-consuming work preparing diverse data sources for reporting and analysis, data wrangling technologies have come a long way.

As head of products at Trifacta (a data wrangling software vendor), one of the questions I repeatedly get asked in meetings with clients, partners, and analysts is "What's the difference between data wrangling and ETL?" Given

analysts is, "What's the difference between data wrangling and ETL?" Given how the features of the two technology spaces overlap in functionality, it's a natural question to ask and one that the market needs to more clearly define.

To give you a clear understanding of the delineation between data wrangling and ETL, I'll describe the top three major differences between the two technologies.

1. The Users Are Different

The core idea of data wrangling technologies is that the people who know the data best should be exploring and preparing that data. This means business analysts, line-of-business users, and managers (among others) are the intended users of data wrangling tools. I can personally attest to the painstaking amount of design and engineering effort that has gone into developing a product that enables business people to intuitively do this work themselves.

In comparison, ETL technologies are focused on IT as the end users. IT employees receive requirements from their business counterparts and implement pipelines or workflows using ETL tools to deliver the desired data to the systems in the required formats.

Business users rarely see or leverage ETL technologies when working with data. Before data wrangling tools were available, these users' interactions with data would only occur in spreadsheets or business intelligence tools.

2. The Data Is Different

The rise of data wrangling software solutions came out of necessity. A growing variety of data sources can now be analyzed, but analysts didn't have the right tools to understand, clean, and organize this data in the appropriate format. Much of the data business analysts must deal with today comes in a growing variety of shapes and sizes that are either too big or too complex to work with in traditional self-service tools such as Excel. Data wrangling solutions are specifically designed and architected to handle diverse, complex data at any scale.

ETL is designed to handle data that is generally well-structured, often originating from a variety of operational systems or databases the organization wants to report against. Large-scale data or complex raw sources that require substantial extraction and derivation to structure are not one of ETL tools' strengths.

Additionally, a growing amount of analysis occurs in environments where the schema of data is not defined or known ahead of time. This means the analyst doing the wrangling is determining how the data can be leveraged for analysis as well as the schema required to perform that analysis.

3. The Use Cases Are Different

The use cases we see among users of data wrangling solutions tend to be more exploratory in nature and are often conducted by small teams or departments prior to being rolled out across the organization. Users of data wrangling technologies typically are trying to work with a new data source or new combination of data sources for an analytics initiative. We also see data wrangling solutions making existing analytics processes more efficient and

wrangling solutions making existing analytics processes more efficient and accurate as users can always have their eyes on their data as they prepare it.

ETL technologies originally gained popularity in the 1970s as tools primarily focused on extracting, transforming, and loading data into a centralized enterprise data warehouse for reporting and analysis via business intelligence applications. This continues to be the primary use case for ETL tools and one that they are extremely good at.

With some customers, we see data wrangling and ETL solutions deployed as complementary elements of an organization's data platform. IT leverages ETL tools to move and manage data so business users have access to explore and prepare the appropriate data with data wrangling solutions.

For Further Reading:

- Let the People Who Know the Data Best Do the Wrangling
- Case Study: NationBuilder Goes Data Wrangling
- The Logic of Disintermediation and the End of ETL as We've Known It

About the Author

As VP of products, Wei combines her passion for technology with experience in enterprise software to define and shape Trifacta's product offerings. Having founded several start-ups of her own, Wei believes strongly in innovative technology that solves real-world business problems. Most recently she led product management efforts at Informatica, where she helped launch several new solutions including its Hadoop and data virtualization products. You can contact the author at trifacta@teamlewis.com.
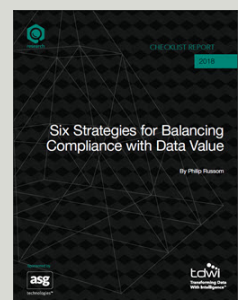
FEATURED RESOURCES
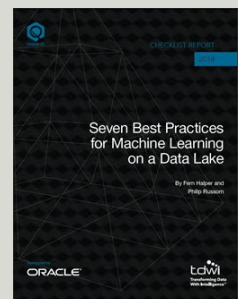
**TDWI Checklist
Seven Best Practices for
Machine Learning on a Data
Lake**

**TDWI Checklist
Six Strategies for Balancing
Compliance with Data Value**

Q&A with Jill Dyché

Find out what's keeping teams up at night and get great advice on how to face common problems when it comes to analytic and data programs. From head-scratchers about analytics and data management to organizational issues and culture, we are talking about it all with Q&A with Jill Dyche.

View Recent Article

Submit Your Questions to Jill

CONTRIBUTE TO UPSIDE

View Our Writer's Guide
Join Our Contributor Network

TDWI MEMBERSHIP
Accelerate Your Projects, and Your Career

TDWI Members have access to exclusive research reports, publications, communities and training.

Individual, Student, & Team memberships available.

Membership Information

FOLLOW TDWI

LINKS

Contact Us

SUBSCRIBE TO OUR NEWSLETTER

Transforming Data
With Intelligence™

1105 MEDIA
YOUR GROWTH. OUR BUSINESS.

LinkedIn
Twitter
Facebook
Google+
YouTube
Instagram

Become a TDWI
Training Partner
Become an Instructor or
Contributor
Marketing
Opportunities
Author Submissions
Reprints & Permissions
Privacy Policy
Vendor News
Newsletters
Press Releases
Sitemap
TDWI Europe

Become a Member

Subscribe

Privacy Policy
© 2018 TDWI
All Rights Reserved