

# EMPOWERING BUSINESS TO PREPARE DATA OF ALL SHAPES AND SIZES

Trifacta & Cloudera

**Industry**

Data Preparation & Discovery

**Website**

www.trifacta.com

**Company Overview**

Trifacta is the global leader in data wrangling software. Leveraging decades of innovative research in human-computer interaction, scalable data management and machine learning to make the process of preparing data faster and more intuitive. Thousands of customers are unlocking the potential of their data with Trifacta.

**Product Overview**

- \_ Access and transform data in Cloudera
- \_ Removes data prep dependency on IT
- \_ Integrates with Cloudera security, governance and lineage solutions

**Benefits**

For the Business

- \_ Discover and transform data for downstream analytics and ML without the bottleneck of IT
- \_ Easily enrich data sets to improve the results of BI reporting or advanced analytics
- \_ Machine-learning based recommendations guide the user to the final result
- \_ Enable more use cases on your Cloudera Enterprise environment

For IT

- \_ Enable the business to leverage Cloudera Enterprise big data environment
- \_ Provide self-service data access in a secure and governed fashion
- \_ Better meet SLAs for data access and reporting

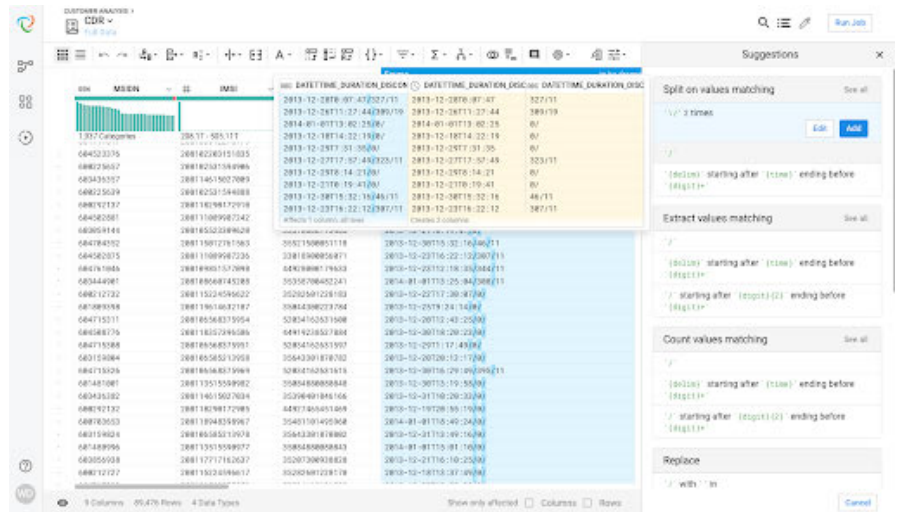
**Overview**

An enterprise data warehouse enables organizations to land data of all shapes and sizes for enterprises to explore and leverage for analytics and ML driven applications. However, organizations are struggling with the wide volume and variety of data formats to quickly gain value out of their data. In fact, it has been widely publicized that up to 80% of the overall analysis process is typically spent cleaning or preparing data. Enterprises are increasingly looking to empower their business groups to explore and prepare data in their enterprise data hubs in a governed and secure environment to quickly gain value from their data assets

**Data Innovation**

The joint solution from Trifacta and Cloudera enables business and IT departments to partner in driving their organization's efforts to innovate with data. With the solution, organizations can empower their data analysts to perform exploratory data preparation, leveraging the processing power of Apache Spark, on all of the data that is stored in an enterprise data hub without heavy reliance on specialized data engineering & data science teams.

Trifacta is designed to help business analysts prepare data without having to manually write code. The solution enables users to visualize the content of data stored in Cloudera and interact with that content to define transformation rules that output the data in the desired form for analysis. All of this is done in a secure and governed fashion, with Trifacta integrations with Apache Kerberos, Apache Sentry and Cloudera Navigator.



## Cloudera Data Warehouse Solution Highlights

- \_ End-to-End platform covering the complete data lifecycle
- \_ Enterprise-grade features for scalability, security and governance
- \_ Intuitive business ready tools with powerful functions
- \_ Agile, schema-less data and pipeline modeling
- \_ Interactive interface with visual data profiling for iterative preparation
- \_ Visual data exploration over any attribute, value or metric

## Data Wrangling to Accelerate Cloudera Usage and Business Value

Trifacta enables end-user preparation on data of all shapes and sizes in Cloudera for analysis by accelerating the following aspects of the data wrangling process:



### Discovery

Discover what is in your data and how it might be useful for different analytic explorations to quickly identify the value of a data set through Trifacta's interactive profiling capabilities.

### Structuring

Trifacta's Predictive Transformation provides intelligent transform suggestions driven through the user's interactions to structure their data for downstream reporting and analytics.

### Cleansing

Trifacta automatically detects data types and missing or inconsistent values throughout the transformation process providing users with immediate cues to a dataset's fit and validity.

### Enrichment

Trifacta's data enrichment features make standardizing data, joining datasets and aggregating data outputs to the right level, faster and more accurate - making it easy to enrich analytics with various types of data from different departments and organizations.

### Validate

Trifacta provides statistics of each data preparation job to ensure results are complete. Validated data sets are critical for ensuring the accuracy of reporting and analytics projects.

### Publish

Once data preparation is complete, Trifacta allows users to publish back to Cloudera Enterprise in a variety of file formats to fit the requirements for the analytics project. Jobs can be scheduled for projects that require consistent updates.

### About Trifacta

Trifacta's mission is to create radical productivity for people who analyze data. We're deeply focused on solving for the biggest bottleneck in the data lifecycle, data wrangling, by making it more intuitive and efficient for anyone who works with data.

### About Cloudera

At Cloudera, we believe that data can make what is impossible today, possible tomorrow. We empower people to transform complex data into clear and actionable insights. We are the leading platform provider for machine learning, analytics and data management built for the cloud. The world's largest enterprises trust Cloudera to help solve their most challenging business problems.

[Learn more at cloudera.com](https://www.cloudera.com)