

# Accelerate Data Preparation for Snowflake

## Background

On-premise Data Lakes and Data Warehouses are expensive and difficult to manage. Modern Cloud platforms like Snowflake's Zero Management Cloud Data Warehouse provide scalable, low cost computing, enabling organizations to store larger volumes of data in more cost effective ways. As data volumes and the number of data sources increases, data preparation challenges grow and those with the technical capabilities to use traditional data management tools become overburdened and unable to respond to the businesses asks in a timely manner. This presents an opportunity for organizations to outcompete with successful curation of differentiated, insight rich data. Trifacta provides an [industry-leading](#), machine learning-driven cloud data preparation solution natively integrated with Snowflake to deliver clean, trusted, and well-prepared data to support your analytics and AI initiatives. Trifacta and Snowflake combine to create a modern solution for preparing large volumes of data in an agile, powerful and cost effective environment to accelerate data driven decision making. Data teams more than ever need a modern platform to be able to profile, structure, clean and enrich their data for valuable onboarding, analytics and reporting, and Machine Learning and AI initiatives.

## Challenges

While a growing number of companies are migrating their data to Snowflake to improve data onboarding processes, modernize reporting and analytics, and accelerate machine learning and AI, making data fit for purpose is no easy feat due to the varying formats, shapes and sizes of data businesses deal with. The existing data management solutions such as ETL tools are not equipped to adequately clean and prepare data in Snowflake because of the following limitations:

**Lack of self-service:** The existing tools were primarily designed for IT/technical users as opposed to business users who understand data best and rely on data to gain business insights and make everyday decisions.

**Rigid architectural design:** Most legacy tools were designed to process structured data with predefined schema, they are unable to refine and prepare raw data in less structured formats that modern cloud platforms like Snowflake can handle, thus limiting the analytics use cases companies can solve for.

“

Trifacta has empowered our business users to accelerate data preparation and provide a 3x improvement in turnaround-time. Snowflake's unlimited scale, on-demand execution clusters, and support of a rich set of data formats have allowed Malwarebytes to centralize all data from transactions including billions of telemetry logs. With the integration of Trifacta and Snowflake, we can accelerate the process of preparing data in Snowflake and empower our business to do more with self-service analytics.”

Manjunath Vasishta  
Director of Data Science and Engineering,  
Malwarebytes

## Supported Services



Amazon  
S3



Amazon  
EMR



Amazon  
Glue



Amazon  
Sagemaker



Data Robot



Looker

“Organizations maximize the value of their data’s insights when working with high-quality, well-prepared data, and Trifacta is a strong example of a solution that accelerates the cleaning and preparation of data for analytics and reporting within Snowflake. Partnering with Trifacta allows our shared customers to utilize the scalability and concurrency benefits of Snowflake’s unique architecture, while gaining additional efficiencies and agility when preparing structured, semi-structured, and unstructured data for downstream Analytics.”

Brian Stevens  
CTO, Google Cloud

### About Snowflake

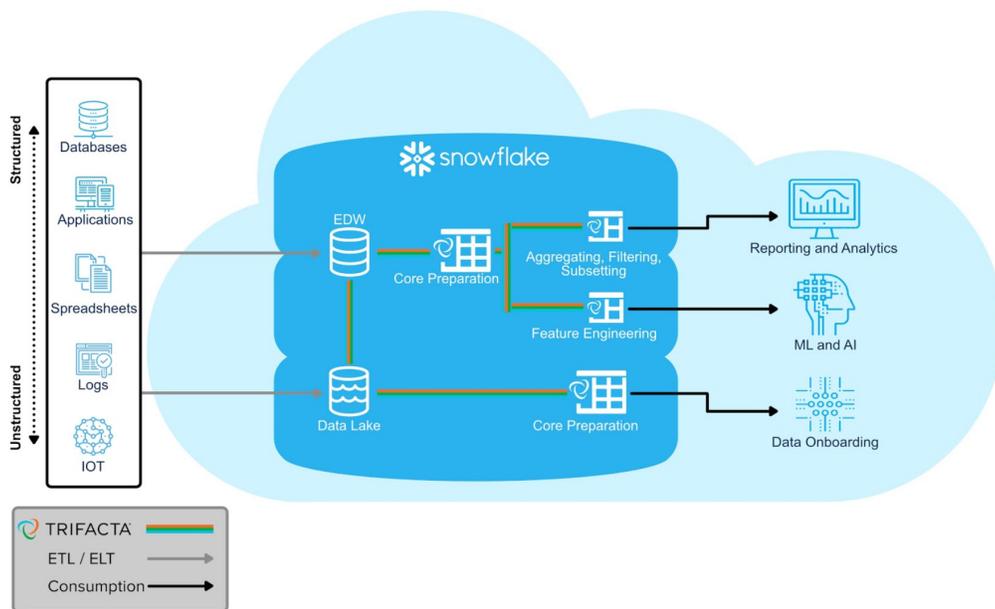
Snowflake is the only data warehouse built for the cloud, enabling the data-driven enterprise with instant elasticity, secure data sharing and per-second pricing, across multiple clouds. Snowflake combines the power of data warehousing, the flexibility of big data platforms and the elasticity of the cloud at a fraction of the cost of traditional solutions. Snowflake: Your data, no limits. Find out more at [snowflake.com](https://www.snowflake.com)

**Governance and Collaboration:** Most business facing tools are desktop applications with no centralized governance and no collaboration, creating issues with lineage, transitioning work, security, and auditing.

## Why Trifacta for Snowflake

In today’s world, organizations compete on differentiated data. To foster efficient, secure and scalable data operations, organizations must empower data professionals to instantly connect to the data they need, to profile, structure, blend and clean that data to make it fit for purpose, to automate that work for repeatable value, and to collaborate in a centrally governed platform. The scale, flexibility and low cost of Snowflake’s Zero Management Cloud Data Warehouse combined with Trifacta’s easy to use, powerful data preparation platform enable businesses to build a modern data lake or data warehouse to empower all of their data professionals with the tools needed to make data driven business decisions in a fraction of the time. Spend 80% less time wrangling your data and 80% more time driving valuable business decisions with Trifacta and Snowflake.

### Reference Architecture



### Onboard Diverse Data Faster

Onboarding data to data products and services in Snowflake requires agility in taking disparate sources of data in varying formats and schemas, and then blending, cleaning, and restructuring that data for analysis. Companies who monetize data and sell analytics have to deal with new and unfamiliar data every time new clients provide data. These organizations need to quickly understand the unfamiliar data, and then prepare that data for their

### About Trifacta

Organizations that embrace data-driven decision making compete on differentiated data. Trifacta empowers data professionals of all levels of technical expertise to connect and wrangle data into its final state for reporting, analytics and machine learning, all in a tightly governed, cloud native environment. Trifacta blends visual guidance and machine learning to create an intuitive user experience built to accelerate time to value and automate repeatable workflows. Learn more at [trifacta.com](http://trifacta.com).

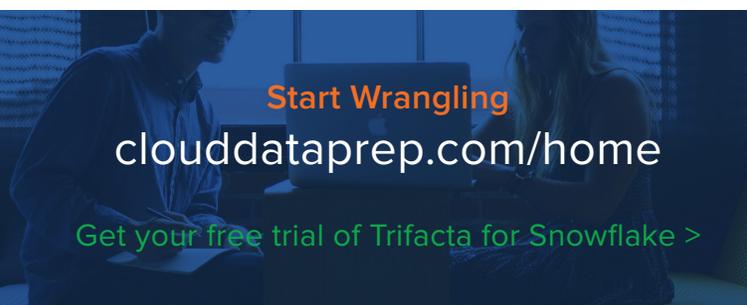
downstream models, analytics, and other data products and services. Trifacta connects to supported data lakes like Amazon S3, and Trifacta is uniquely designed to handle structured, semi-structured and unstructured data with ease, making it easy to take diverse and unfamiliar data, and profile, structure, clean, blend, and validate to the format and quality that the downstream use requires, before publishing to Snowflake for data onboarding use cases.

### Modernize Reporting and Analytics

Trifacta provides business teams with the automation capabilities of technical platforms with the ease of use of Excel. The platform is centrally governed and access and permissions are easily configurable. Data never leaves the source systems, integrating with the platforms native security. Businesses are able to be more agile, ensure faster time to insight with higher quality data, and collaborate with other teams. The combined ease of use, agility, governance and collaboration creates a platform that enables accelerated turn around time whenever new data or new business requirements come in.

### Accelerate ML/AI initiatives

Machine Learning and AI initiatives rely on clean high quality data. So much so that data scientists spend over 80% of their time on cleaning data, engineering features, and refining their data before training and deploying models. Trifacta provides an environment for data scientists, data engineers, and data analysts to collaborate and build repeatable processes to ensure faster time to value. Join disparate sources of data, standardize columns and normalize, scale, and engineer features with immediate previews at each step of the way. Trifacta's visual guidance and built in machine learning create an interface focused on ease of use, instant validation, and powerful transformations, ensuring users of all skill levels can perform the transformations they need without the need for code or debugging, improving time to deploy models in supported data science platforms like Amazon Sagemaker and DataRobot.



Start Wrangling  
[clouddataprep.com/home](http://clouddataprep.com/home)

Get your free trial of Trifacta for Snowflake >



Trifacta is the industry pioneer and established leader of the global market for data preparation technology. The company draws on decades of academic research in machine learning and data visualisation to make the process of preparing data faster and more intuitive. More than 100,000 data wranglers in 10,000 companies worldwide use Trifacta solutions across cloud, hybrid and on-premises environments to support a variety of analytic and operational use cases. Leading organizations such as Deutsche Boerse, Google, Kaiser Permanente, New York Life and PepsiCo count on Trifacta to accelerate time-to-insight and discover opportunities that drive success. Learn more at [trifacta.com](http://trifacta.com).

For Additional Questions, Contact Trifacta  
[www.trifacta.com](http://www.trifacta.com) | [team@trifacta.com](mailto:team@trifacta.com)

Experience the Power of Data Wrangling Today  
[www.trifacta.com/start-wrangling](http://www.trifacta.com/start-wrangling)