

Accelerate Data Preparation on AWS

Background

Cloud based data analytics solutions have evolved over the past few years to make it much easier to establish data lakes and warehouses for business intelligence (BI) while leveraging artificial intelligence (AI) and machine learning. Cloud solution providers like Amazon provide a wide variety of services that cover data processing, warehousing, storage, real-time analytics, operational analytics, catalogs and the ability to easily incorporate 3rd party data through the use of the Amazon Data Exchange.

While these services have eliminated the long startup costs and time associated with similar on-premise solutions, they were designed to be used by data engineering experts, who are increasingly hard to find and expensive to hire. As a result, data analysts and data scientists are unable to service themselves when it is time to blend and prepare data and every analytics project still requires time-consuming back and forth requests between themselves and IT.

Trifacta was created to change this paradigm and unblock data analysts and scientists from the IT department. Trifacta leverages the amazing depth of capabilities available on the Amazon cloud while providing the ease of use of tools like Excel, with added visualizations and data quality information, paired with the automation capabilities of tools like Python. Trifacta automates many of the steps, like JSON file conversion and translation of discovery steps into operational data pipelines, that used to require manual coding.

With Trifacta, data professionals can truly do it yourself.

“

Trifacta on AWS provides a savings of \$124,800 and 3x the productivity to start. Marketing analysts and users can more easily munge data as needed and automate the productionalized transformations.”

John Gardner
Sr. Business Intelligence Analyst, Autodesk

Supported Services



Amazon S3



Amazon Redshift



Amazon EMR



Amazon Glue



AWS Data Exchange



Amazon Aurora



Amazon Athena



Amazon SageMaker



Amazon RDS



AWS Identity and Access Management



Amazon QuickSight

“Clean and annotated training data is the foundation of modern machine learning. It fuels state of the art algorithms in computer vision and natural language understanding; however, acquiring it takes time and resources. We are very excited to have Trifacta join the Machine Learning Competency Program to help our customers spend less time preparing their data and more time creating intelligence.”

Joseph Spisak
Global Lead for Machine Learning Partnerships,
Amazon Web Services, Inc.

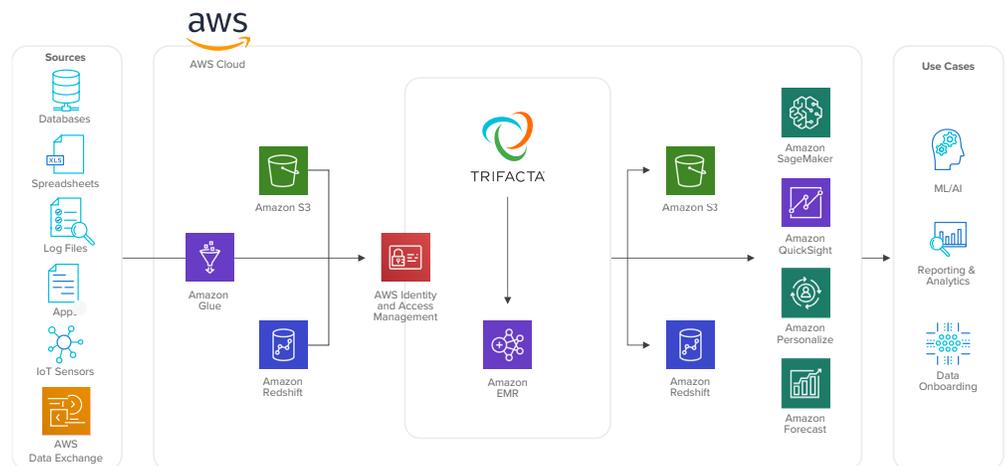
About AWS

Amazon Web Services (AWS) is the world’s most comprehensive and broadly adopted cloud platform, offering over 165 fully featured services from data centers globally. Millions of customers—including the fastest-growing startups, largest enterprises, and leading government agencies—trust AWS to power their infrastructure, become more agile, and lower costs. Learn more about all AWS services at aws.amazon.com

Why Trifacta for AWS

Trifacta, an [AWS ML Competency and Data & Analytics Competency solution](#), is a data preparation service on Amazon Web Services (AWS) that leverages the Amazon ecosystem services across Amazon S3, Amazon EMR, Amazon Redshift, Amazon SageMaker, Amazon IAM as well as AWS Data Exchange to enable analysts, data scientists, data engineers, and business users to prepare data of any form and size, quickly transform data from its raw format into a refined state for analytics and/or machine learning initiatives. Trifacta is natively integrated with AWS ecosystem, allowing organizations to easily scale computing capacity to meet changing requirements. Whether you are building a cloud data lake with Amazon S3, modernizing your legacy data warehouse to Amazon Redshift for better BI/Reporting, or launching ML/AI project in Amazon SageMaker, Trifacta automates your data preparation to allow faster time to insights and innovation with clean, connected, secure and timely data.

Reference Architecture



Trifacta for data preparation in AWS delivers:

Seamless Integration with AWS Ecosystem

Architected for the cloud and natively integrated across a range of Amazon services including Amazon S3, Amazon EMR, Amazon Redshift, Amazon SageMaker, AWS Data Exchange and Amazon IAM Role for ease of management, greater agility and enterprise-class security.

Easy access to 3rd Party Data on AWS Data Exchange

Good data analysis doesn’t happen on an island, which is why AWS Data Exchange aggregates a wide variety of 3rd party data for you. Trifacta helps analytics teams easily augment and enrich their analysis by more efficiently onboarding 3rd party data. Trifacta’s self-service data preparation platform helps you explore, transform and join together diverse data and accelerate the process of getting data ready to use`.

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.

Accelerated Data Preparation on AWS

Automate data preparation process with a visual, interactive and AI-powered platform to ensure clean, connected and trusted data is immediately available on AWS to support data services, modern BI/Reporting, and AI/ML initiatives.

Centralized Data Governance and Access Control

Centralizes data governance, security, lineage and access control to a single platform instead of disparate spreadsheets or desktops that are impossible to manage, reducing operational burden and cost.

Business Self-service, Intelligent Data Preparation

Empower business users who know the data best with simple, interactive, visual, and machine learning-powered platform to accelerate data preparation and increase productivity and time to insights.

Superior Data Services with AWS Data Lake

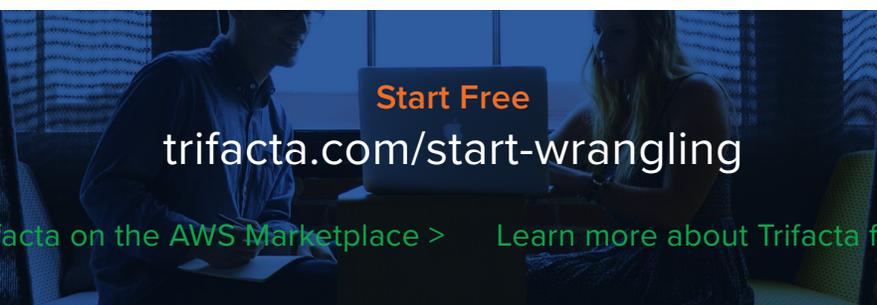
Trifacta quickly transforms and standardizes messy data from internal and external systems into clean and well-prepared data in AWS data lake such as Amazon S3, accelerating data lake adoption and enabling superior data services.

Accelerated Data Preparation for BI Reporting

Trifacta expedites data preparation on AWS with a simple, interactive and intelligent platform, ensuring clean, connected and timely data is immediately available on AWS, ready for all your BI reporting needs.

Automated Data Prep for AI/ML

Trifacta automates data preparation for data scientists and developers working on ML/AI projects on AWS by leveraging services such as Amazon SageMaker, minimizing the time spent on data wrangling while allowing data scientists and engineers to focus on building and training models, as well as interpreting the results.



Start Free

trifacta.com/start-wrangling

[Get Trifacta on the AWS Marketplace >](#) [Learn more about Trifacta for AWS >](#)



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.

For Additional Questions, Contact Trifacta
www.trifacta.com | team@trifacta.com

Experience the Power of Data Wrangling Today
www.trifacta.com/start-wrangling