

Cloud Dataprep by Trifacta, the sole native data preparation service for the Google Cloud Platform

Background

Data lakes, data warehouses, and machine learning/artificial intelligence (ML/AI) applications have been historically expensive and difficult to manage in on-premise architectures. The rise of cost-effective, scalable data storage and elastic processing in the cloud has completely flipped the analytics paradigm. Modern cloud platforms with serverless automated data services now enable organizations to take a new, easier, and more efficient approach to analytics. Yet, with cloud adoption and mass digitalization, the volume and form of data are also drastically increasing. This is causing increased data preparation challenges, and those with the technical capabilities to use traditional data management tools are becoming overburdened and consequently unable to respond in a timely manner to business demands. At the same time, this presents an opportunity for organizations to outcompete through the successful curation of differentiated, insight-rich data. Cloud Dataprep by Trifacta on the Google Cloud Platform (GCP) offers a modern solution for preparing large volumes of data in an agile, powerful, and cost-effective environment to accelerate data-driven decision making.

Why Trifacta for Google

While Google, with GCP, entered the cloud platform market late, they did so in a very different way by proposing an end-to-end suite of applications. Especially within the analytics space, and given that the center of gravity of data is now in the cloud, Google decided to offer a modern, fully managed smart analytics suite that includes batch and real-time data ingestion, data storage and processing at scale, data preparation, reporting and dashboarding, and ML/AI applications.

There is one common challenge that every analytics initiative has to tackle. Clean, structured, and normalized data is needed to fuel trustworthy reports or accurate predictions. In other words, self-service data preparation has become critical to solving this now well-known hurdle, which easily consumes up to 80% of any data project.

“

Dataprep by Trifacta is incredibly user friendly and the machine learning suggestions help us reduce a big chunk of the labor-intensive process of data wrangling, so we can analyze and mine food purchasing data across 46 states, 521 distributors, and 1,000 producer groups representing over 100,000 farms and vendors to reimagine local and sustainable sourcing and regional supply chains.”

Linda Mallers
President and CEO, Farmlogix

Supported Services



Google Cloud Storage



Google BigQuery



Google Dataflow



Google Data Fusion



Google Cloud AIM



Google Cloud Console



Cloud AutoML

“Cloud Dataprep is an intelligent data service that allows users to visually explore, clean and interactively prepare their data. We selected Trifacta to help power this new service because it was incredibly advanced, super intuitive for people to use immediately, and had a cloud architecture that integrated naturally with Google Cloud Platform.”

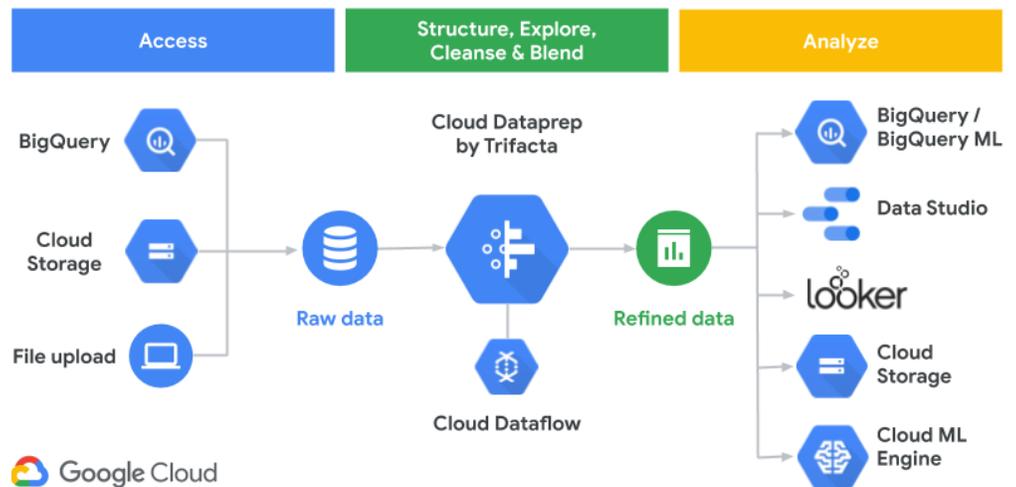
Brian Stevens
CTO, Google Cloud

About Google Cloud

Google Cloud Platform, offered by Google, is a suite of cloud computing services that run on the same infrastructure that Google uses internally for its end-user products, such as Google Search and YouTube. Alongside a set of management tools, it provides a series of modular cloud services including computing, data storage, data analytics, and machine learning. Learn more about Google Cloud Platform at cloud.google.com

Google and Trifacta have partnered to offer Cloud Dataprep on the GCP, the sole data preparation solution available on the GCP. Cloud Dataprep by Trifacta offers the unmatched Trifacta wrangling experience for GCP customers. Cloud Dataprep is available in the GCP console and adheres to the same Google Cloud Platform’s consumption, invoicing, and security principles to offer a seamless Google Cloud Platform experience.

Reference Architecture



How Does Cloud Dataprep Work?

Information Security and Privacy

Cloud Dataprep provides a web-interface for users to define the transformation logic (wrangling recipe) and schedule job execution. The Cloud Dataprep service translates the wrangling recipe into a Cloud Dataflow job which reads, transforms, and writes the data into or between BigQuery and Google Cloud Storage (GCS) within the user’s GCP project, while never leaving that project.

Cloud Dataprep anonymously stores wrangling recipes in the form of metadata within the GCP Cloud SQL. Therefore, the Cloud Dataprep service does not process or store any of the physical data in Trifacta’s GCP project. The physical data is only persisted within the GCP project, which can be configured for storage in any of GCP’s regions/zones.

The Cloud Dataprep service only exists within the GCP ecosystem and can only be launched from the GCP console after authorization through GCP IAM (OAuth).

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.

Business Intelligence and Analytics with the Google Cloud Platform

Business intelligence (BI) is becoming an ever-increasing critical factor to inform business strategy and measure operational efficiency within data-driven organizations. The business decisions must happen fast, formed on trustworthy information. This high demand for data insight induces increased pressure on data analysts and business analysts to produce accurate analysis in an extremely quick turnaround.

These data professionals want to break free from the time-consuming IT-dependent processes and schema-rigid data warehouses to build and deliver on agile BI analysis both for ad-hoc requests and recurring reporting. They want to be empowered with scalable self-service analytic solutions to tackle any business data demands and automate their delivery chain.

Google Cloud Platform offers just that with their Business Intelligence solution, a comprehensive suite of data analytics tools that provides flexibility, scalability, collaboration, and advanced analytics, built-in. With just an email account, anyone can benefit from this solution and be up and running in minutes leveraging the full suite including Cloud Dataprep by Trifacta to prepare data easily and at scale for analytics.

To learn more about the leveraging GCP for business intelligence and analytics, download the brief [“Getting Started with Self-Service Analytics on BigQuery with Cloud Dataprep.”](#)



Start Wrangling

clouddataprep.com/home

[Learn More about Google Cloud Dataprep >](#)



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