## **Dataprep**

There were some questions related to Dataprep, for example, if you want a quite easy implementation to deal with outliers what’s the best tool for that? transform recipes, and finally, how to schedule a Dataprep implementation, do you need Cloud Scheduler for that or you can do it directly from the Dataprep UI?

* Good understanding of the Apache Hadoop ecosystem
* IAM integration
* Benefits of preemptible nodes
* Best practices for migrating Hadoop clusters to Dataproc, such as always separating data from storage by using GCS
* Best practices for optimizing performance
* Connectors
* Apache Spark

Cloud Dataprep is an intelligent data service for visually exploring, cleaning, and preparing structured and unstructured data for analysis, reporting, and machine learning.  
Because Cloud Dataprep is **serverless and works at any scale, there is no infrastructure to deploy or manage.** Your next ideal data transformation is suggested and predicted with each UI input, so you don’t have to write code.  
With the automatic schema, datatype, possible joins, and anomaly detection, you can skip time-consuming data profiling and focus on data analysis.

D**ataprep provides the ability to detect, clean and transform data through a Graphical Interface without any programming knowledge.**  
Refer GCP documentation – Dataprep:- <https://cloud.google.com/dataprep/>

Cloud Dataprep by Trifacta is an intelligent data service for visually exploring, cleaning, and preparing structured and unstructured data for analysis. Cloud Dataprep is serverless and works at any scale. There is no infrastructure to deploy or manage. Easy data preparation with clicks and no code.  
Cloud Dataprep automatically detects schemas, datatypes, possible joins, and anomalies such as missing values, outliers, and duplicates so you get to skip the time-consuming work of profiling your data and go right to the data analysis.  
Cloud Dataprep automatically identifies data anomalies and helps you to take corrective action fast. Get data transformation suggestions based on your usage pattern. Standardize, structure, and join datasets easily with a guided approach.