

Data Cleanup Techniques Using Cloud Dataprep

Introduction

- Part of Google's BigData suite
- Google and Trifacta has partnered for last 2 years to launch Cloud Dataprep as a service
- Trifacta is an industry leader in data preparation
- Together with Google's serverless platform, and Trifacta's easy to use interface, the goal is to help customers to get more value out of their data

Big Challenges to turn Data into Insights

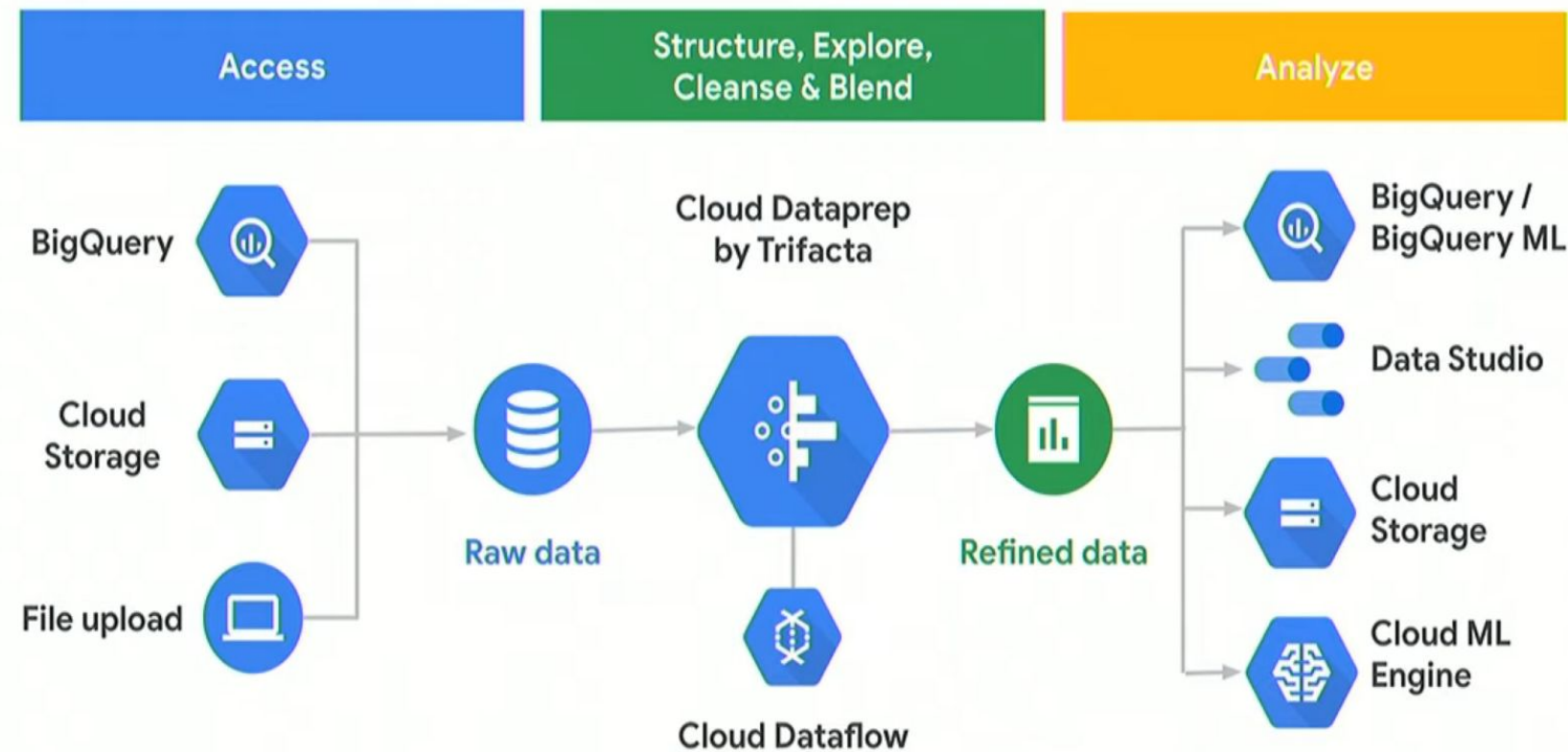
Cloud Dataprep Overview

Data preparation as a Service

Cloud Dataprep is an intelligent data service for **visually exploring, cleaning, and preparing/transforming structured and unstructured data for analysis**

With **zero infrastructure** to deploy or manage, data preparation job is more of clicks, than of code

Cloud Dataprep by Trifacta



Common use cases

Data Onboarding

To efficiently integrate external 3rd party data into standard formats or build analytical pipeline

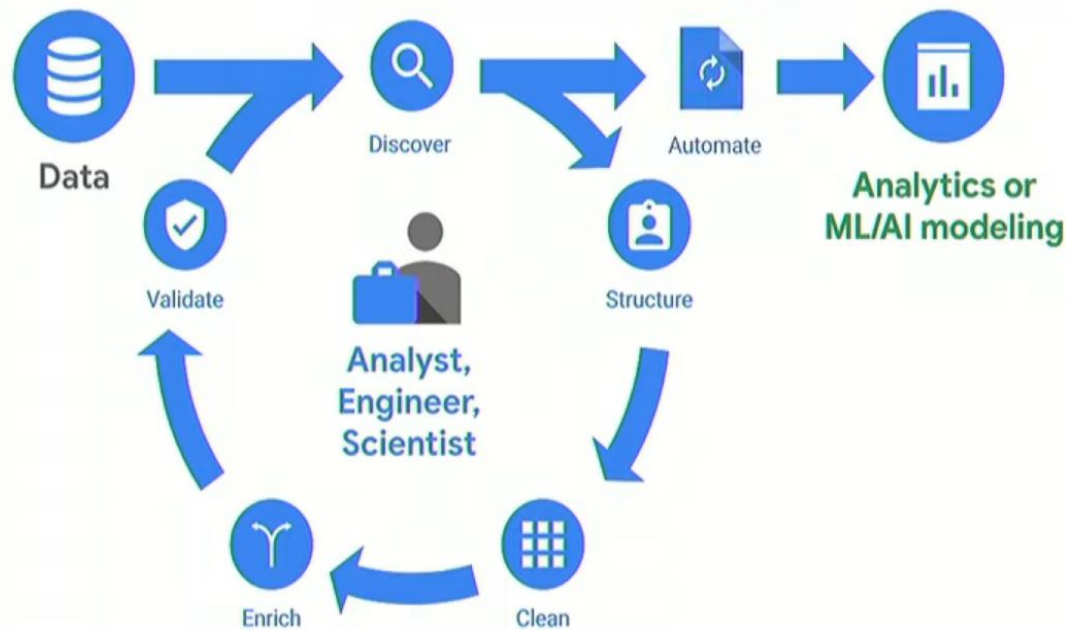
Data science and ML

By training over clean and well structured data

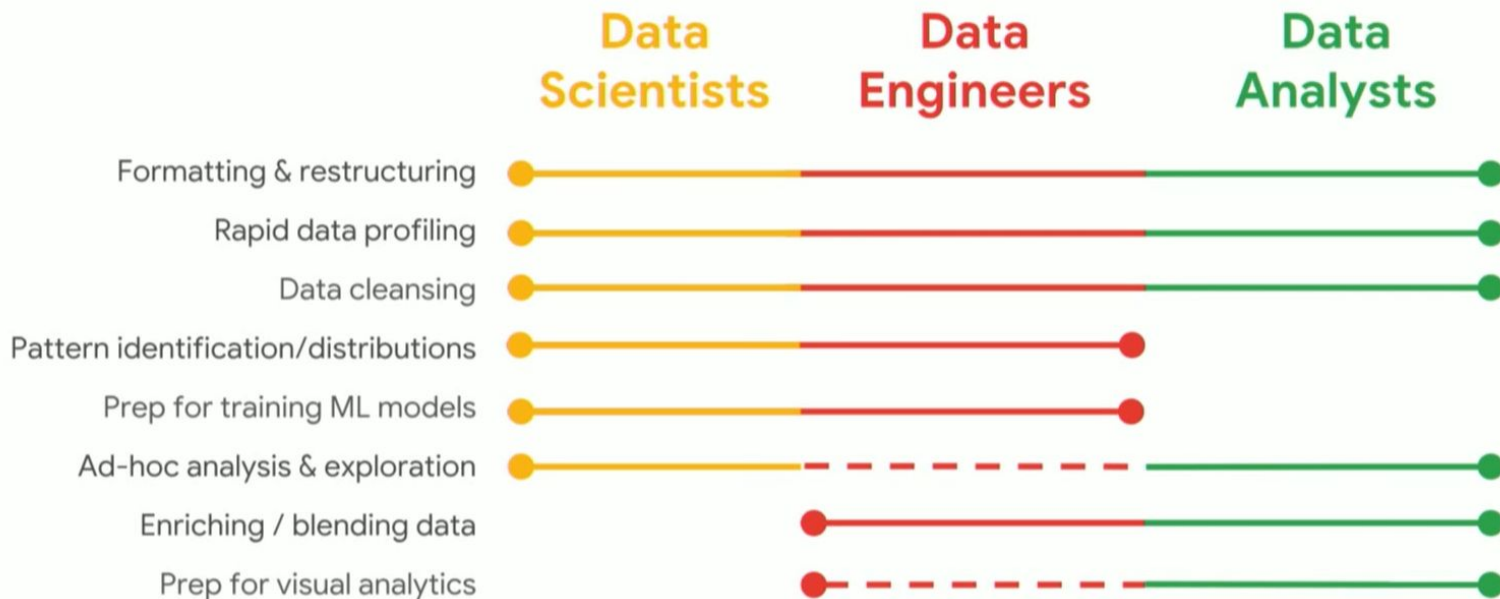
Marketing analytics

Improve marketing analytics by combining cross functional data sets

The data lifecycle with Cloud Dataprep



Who's putting Cloud Dataprep to use, and how?



Dataprep

DEMO

ML Specific Recipes

- One Hot Encoding
- Binning
- Integerize categorical variables
- Splitting the data into training and testing

Q & A

THANK YOU