



One liner

Transformation
In GCP happens
here

Our Interest

Dataflow Basics

What is it

In ETL, the data is extracted, transformed and loaded into the Datawarehouse/ Datalake.

GCP provides a service, Dataflow, which does the transformation part.

Unified Processing

Dataflow works with both

- batch data- data that's already been collected and is now being retrieved, and
- streaming data- data being collected in real or near real-time.

Serverless

Dataflow is Serverless.

No VMs for you to configure or scale.

Enables Horizontal Auto Scaling.

Fully Managed

Dataflow is Fully Managed.

Requiring zero effort in terms of hardware, setup, configuration, and maintenance.

Code Engine

Apache Beam is underlying code engine for Dataflow.

Supports SQL, Java & Python.