



Data Engineering on
Google Cloud Platform

Building Batch Data Pipelines Course Summary

Hello and welcome to our course on data engineering.
My name is ...

Course Summary

- When to use ELT vs ETL
- Transforming data in BigQuery and ensuring data quality
- Using Dataflow to overcome BigQuery's limitations as an ETL solution

Course Summary

- Cloud Dataproc simplifies Hadoop workloads on GCP
- Use Cloud Storage instead of HDFS with Cloud Dataproc

Course Summary

- Cloud Data Fusion allows you to visually design, build, and run data processing pipelines
- Cloud Composer, Cloud Functions and Cloud Scheduler are the glue for your data pipelines

Course Summary

- Using Cloud Dataflow to process batch data