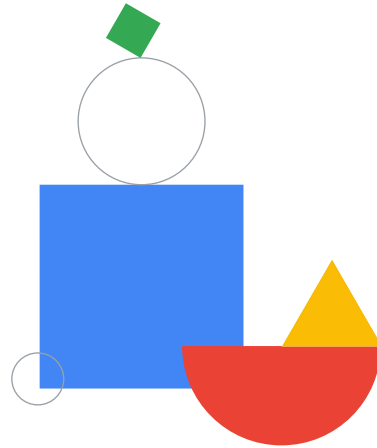


Building Batch Data Pipelines on Google Cloud



Welcome to Building Batch Data Pipelines on Google Cloud.

Facilities



Parking



Facilities

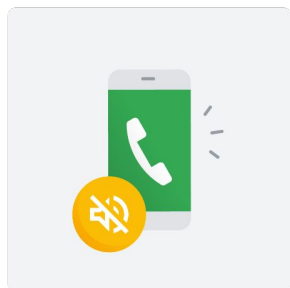


Food

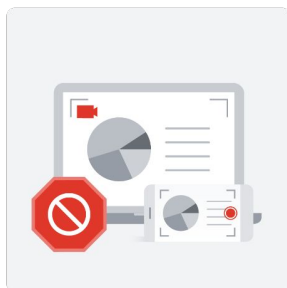
Facilities:

- Parking
- Facilities
- Food

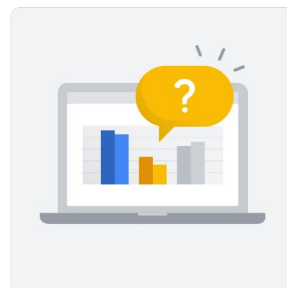
Etiquette



No calls



No recording



Ask questions

Course etiquette:

- Please silence your phone and take calls outside.
- Recording this class is prohibited.
- Ask questions interactively or via chat (online).

Data Engineering on Google Cloud learning path



Modernizing Data Lakes and Data Warehouses with Google Cloud



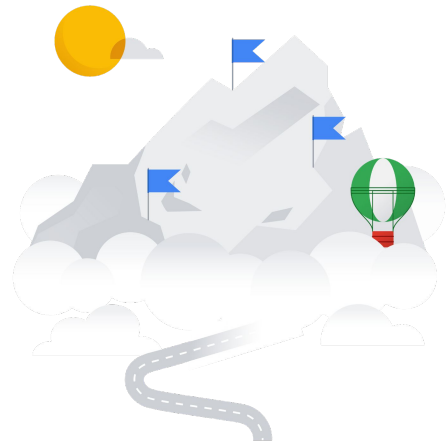
Building Batch Data Pipelines on Google Cloud




Building Resilient Streaming Analytics Systems on Google Cloud




Smart Analytics, Machine Learning and AI on Google Cloud



Building Batch Data Pipelines on Google Cloud is the second part of the Data Engineering on Google Cloud course.



Day 2 agenda



- 01 Introduction to Building Batch Data Pipelines
- 02 Executing Spark on Dataproc
- 03 Serverless Data Processing with Dataflow
- 04 Manage Data Pipelines with Cloud Data Fusion and Cloud Composer

In the first part of the course, we talked about data lakes and warehouses as storage options for your data.

Today, we'll concentrate on how to build batch data pipelines to get data into storage using extract load, extract transform load, and extract load transform routines.

We'll cover several technologies for data transformation on Google Cloud, including BigQuery, executing Spark on Dataproc, serverless data processing with Dataflow, and leveraging Google Cloud in pipelines with Cloud Data Fusion and Cloud Composer.