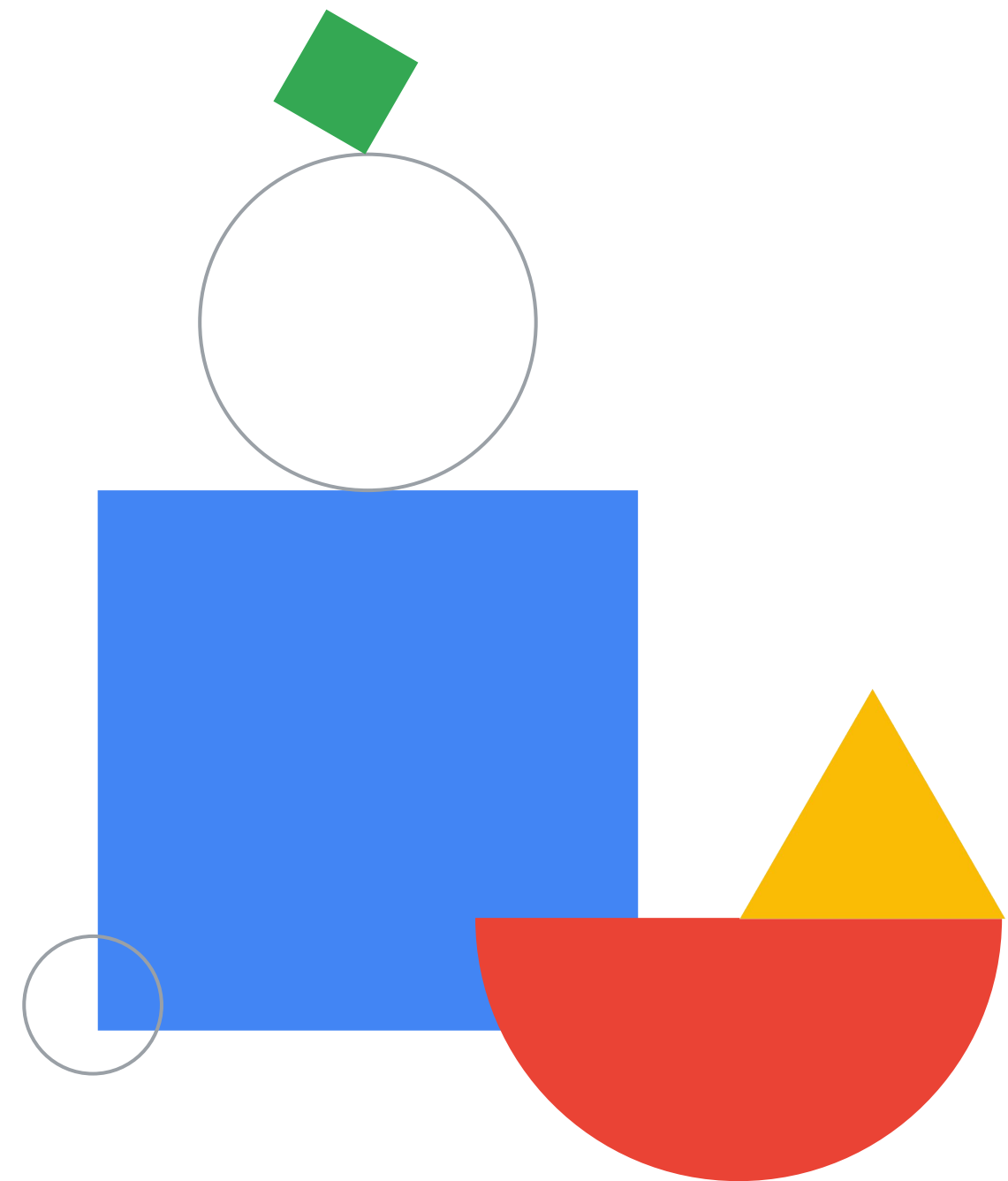Google Cloud

Building Batch Data Pipelines

Summary

Summary

- When to use ELT versus ETL

- Transforming data in BigQuery and ensuring data quality

- Using Dataflow to overcome BigQuery's limitations as an ETL solution

- Dataproc simplifies Hadoop workloads on Google Cloud

- Use Cloud Storage instead of HDFS with Dataproc

- Using Dataflow to process batch data

- 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

Google Cloud

Data Engineering on Google Cloud learning path

✓ Modernizing Data Lakes and Data Warehouses with Google Cloud

✓ Building Batch Data Pipelines on Google Cloud

3 Building Resilient Streaming Analytics Systems on Google Cloud

4 Smart Analytics, Machine Learning and AI on Google Cloud

Google Cloud