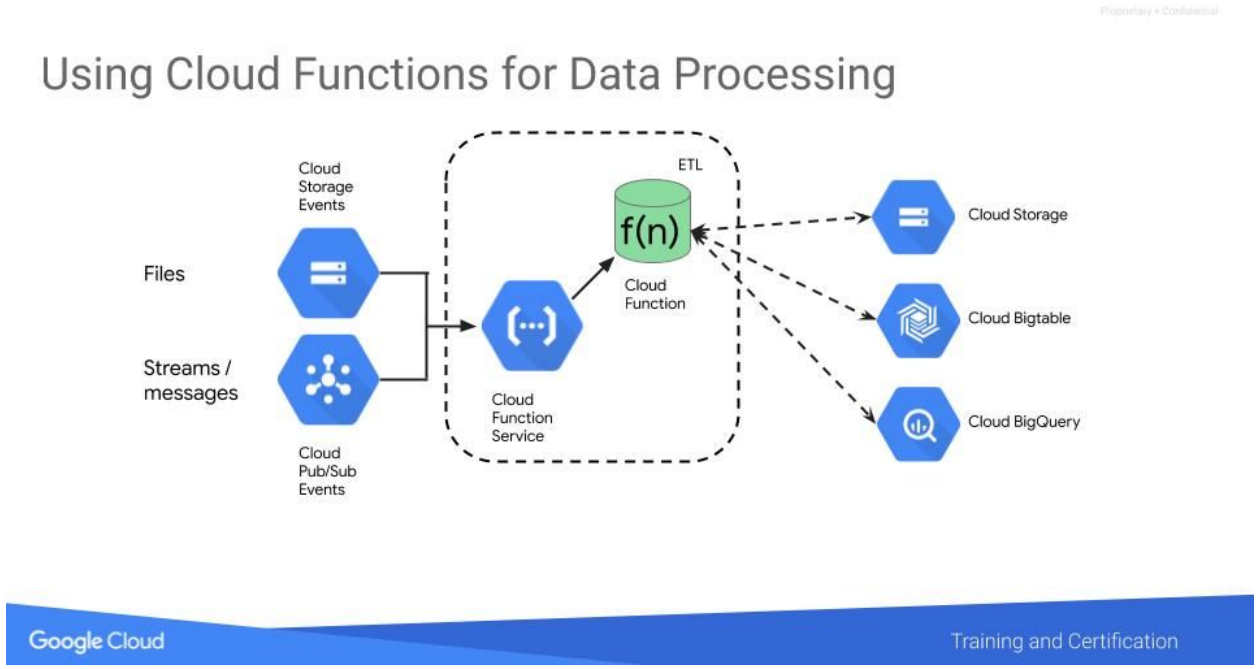
Cloud Functions and Data



A Cloud Function is a serverless, stateless, execution environment for application code. You deploy your code to the Cloud Functions service and set it up to be triggered by a class of events. Mobile application developers use the HTTP (web) event. Data Engineers mainly use events that are associated with Cloud Storage or Cloud Pub/Sub but there are many other triggers available.

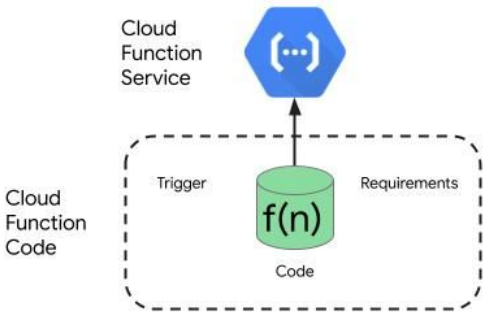
When the event occurs, it triggers the Cloud Function to run. Each time an event occurs and the function is run, it is a fresh instance without history. For example, if you wanted to create a Cloud Function that counts the number of times it is called, it would have to store that counter information externally, such as in Cloud Storage. When you deploy a Cloud Function, you can specify requirements so that common libraries are loaded into the environment. Because Cloud Functions are lightweight and stateless, you can construct microservices applications that are highly scalable.

In Data Engineering, Cloud Functions are often used at data ingress, when data is uploaded to a Cloud Storage bucket or when data arrives as a Cloud Pub/Sub message. The Cloud Function often is used to perform ETL -- Extract, Transform, and Load. In the illustration, the Cloud Function uses APIs to work with common data storage components. For example, it might extract metadata from image files uploaded to Cloud Storage and save the metadata in BigQuery for analysis.

It is possible to assemble a microservices-based workflow using Cloud Functions. You can trigger periodic events using Cloud Scheduler. However, for data processing there are tools such as Cloud Dataproc Workflow Templates and Cloud Composer that are designed to manage workflows without having to code the service yourself.

Cloud Functions has Stackdriver integration so you can monitor your application.

Deploying code to Cloud Functions



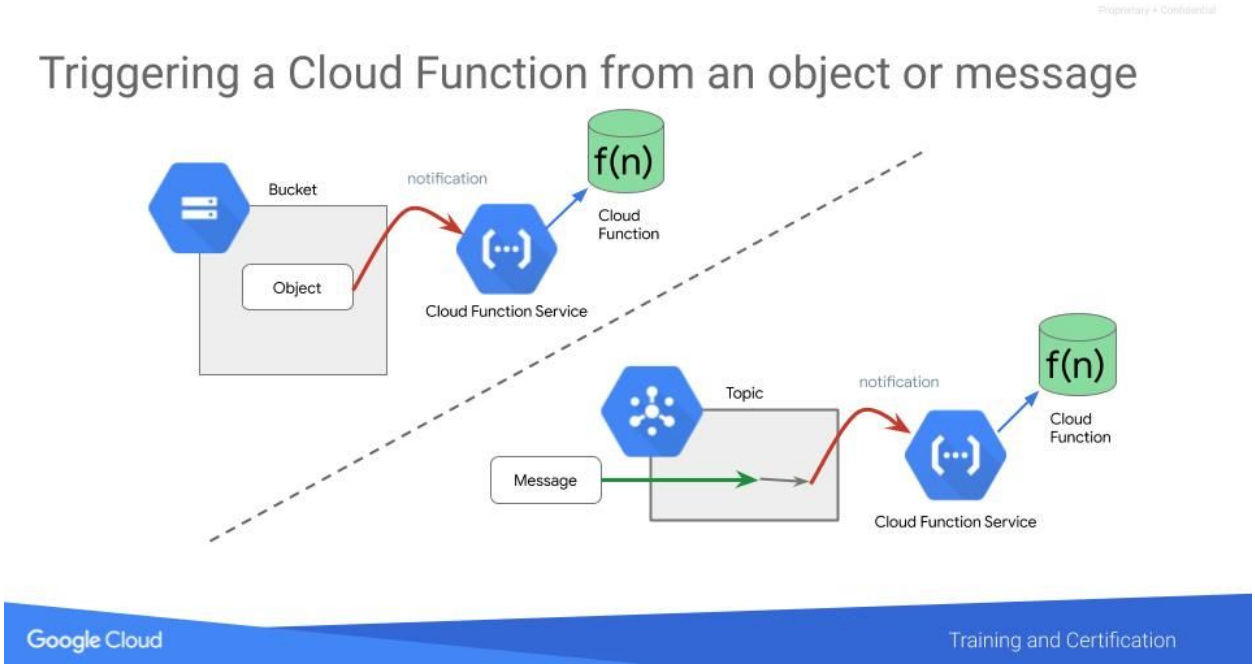
A Cloud Function is written in Python, Node.js or Go.

There are specific requirements for each language.

For example, in Python, the file main.py contains the definitions for one or more Cloud Functions.
A file called requirements.txt is used by pip, the Python package manager, to incorporate dependencies into the runtime environment.
Some dependendent software is not available through pip. You can package these and supply them to Cloud Functions as well.

The Cloud Function code can be deployed to the service through Console, the gcloud command line, or from your local computer.
At that time you specify the trigger that will cause the Cloud Function to run, such as the trigger bucket for Cloud Storage or the trigger topic for Cloud Pub/Sub.

https://cloud.google.com/functions/docs/writing/#functions-writing-file-structuring-python



The bucket must be in the same project as the Cloud Function.

- Authentication
- Send watch request
    - Sync notification event
- add, update, remove object
- Notification
- Waits for acknowledgement

If the app is unreachable for 20 seconds, the notification is retired.
If the app is reachable, but does not acknowledge, then exponential backoff 30 seconds after fail up to max 90 minutes for up to 7 days.

A user-defined HTTP callback (a webhook).

Node.js, Python, Go

Triggers:
HTTP functions
Background functions -- Cloud Pub/Sub or Cloud Storage event