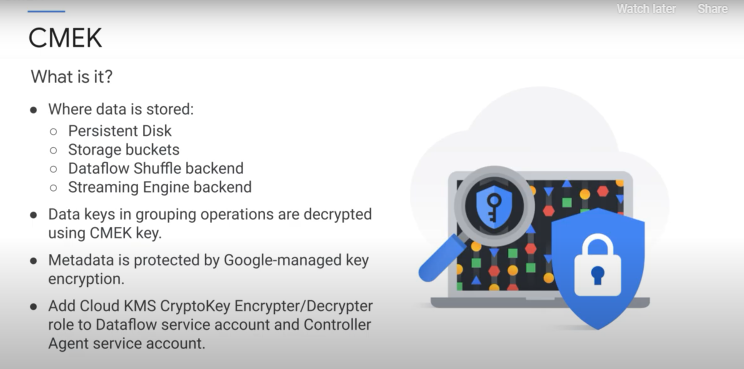
 CMEK stands for customer managed encryption key. During a Dataflow job's lifecycle, different storage locations are used to store data.

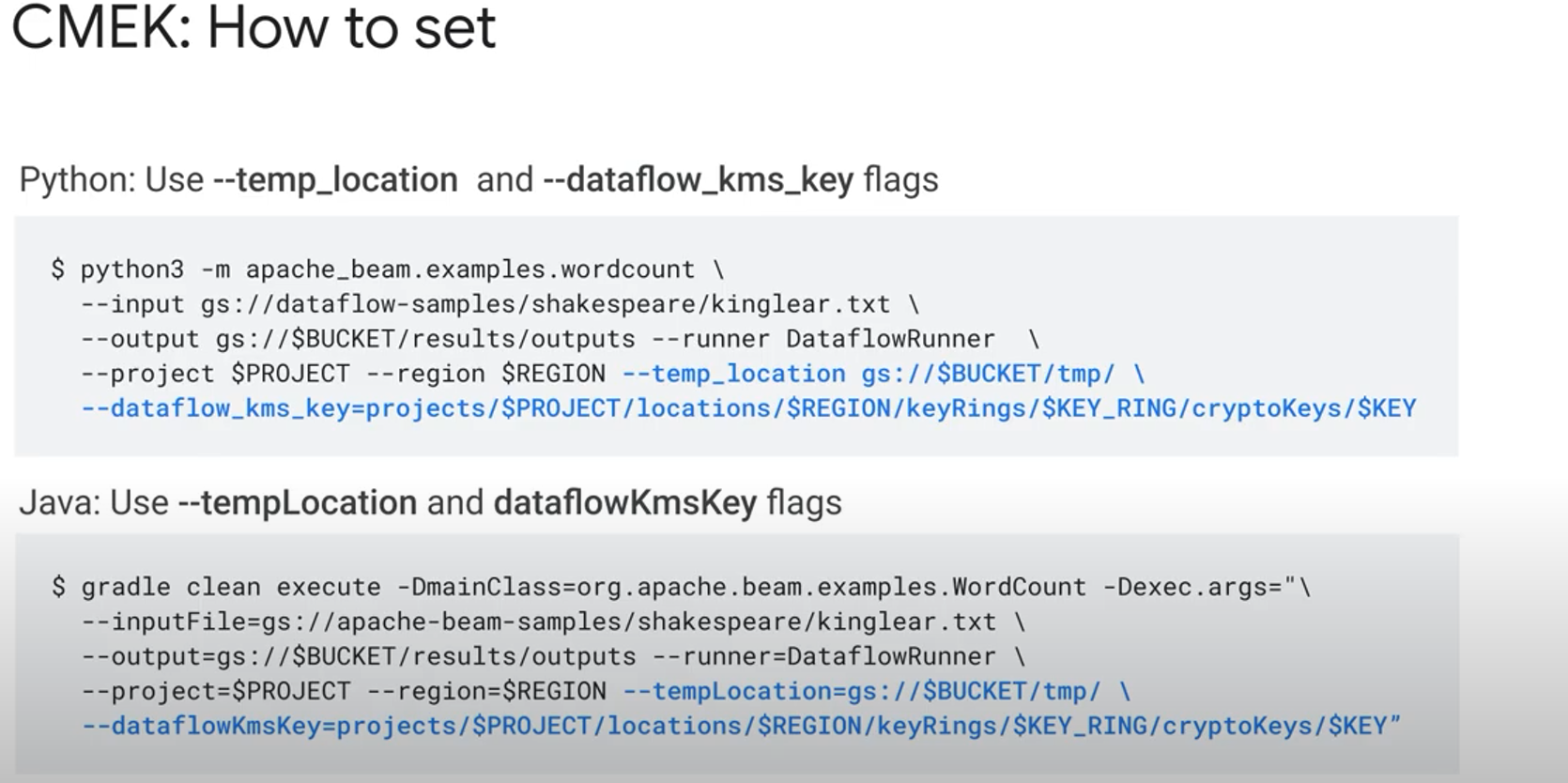
When a Dataflow job is created, a cloud storage bucket is used to store binary files containing pipeline code. A cloud storage bucket is also used to temporarily store export or import data.

While the job is running, persistent disks attached to Dataflow workers are used for persistent disk-based shuffle and streaming state storage. If a batch job is using Dataflow Shuffle, the backend stores the batch pipeline state during execution. If a job is using Dataflow Streaming Engine, the backend stores the streaming pipeline state during execution. By default, when data is stored in any of these locations, a Google-managed key is used to encrypt the data. CMEK allows you to encrypt data at rest using one of your symmetric keys stored in Google Cloud key management system.



 For an additional layer of security, you can hash or transform the key.

Job metadata includes the following: user-supplied data, such as job names, job parameter values, and pipeline graphs, and system generated data, such as job IDs and IP addresses of workers.



When you launch a job that uses CMEK, the region for your key and the regional input for your Dataflow job must be the same. Global or multiregional keys will not work. The bucket selected to temporarily store data must also be in the same region as the key. If you override the pipeline's worker region or zone to a different region than the region associated with your keys, regional keys will not work.