

Fundamentals of GCP

Google Cloud Platform Fundamentals: Big Data and Machine Learning

Version #1.1



Agenda



Agenda

CPUs on demand + Lab

A global filesystem + Lab

Google Cloud provides an earth-scale computer

Networking



FASTER (US, JP, TW) 2016

Data storage



IC (JP, HK, SG) 2013



Compute power

- Network
- Network sea cable investments
- Edge points of presence >100
- Edge node locations >1000

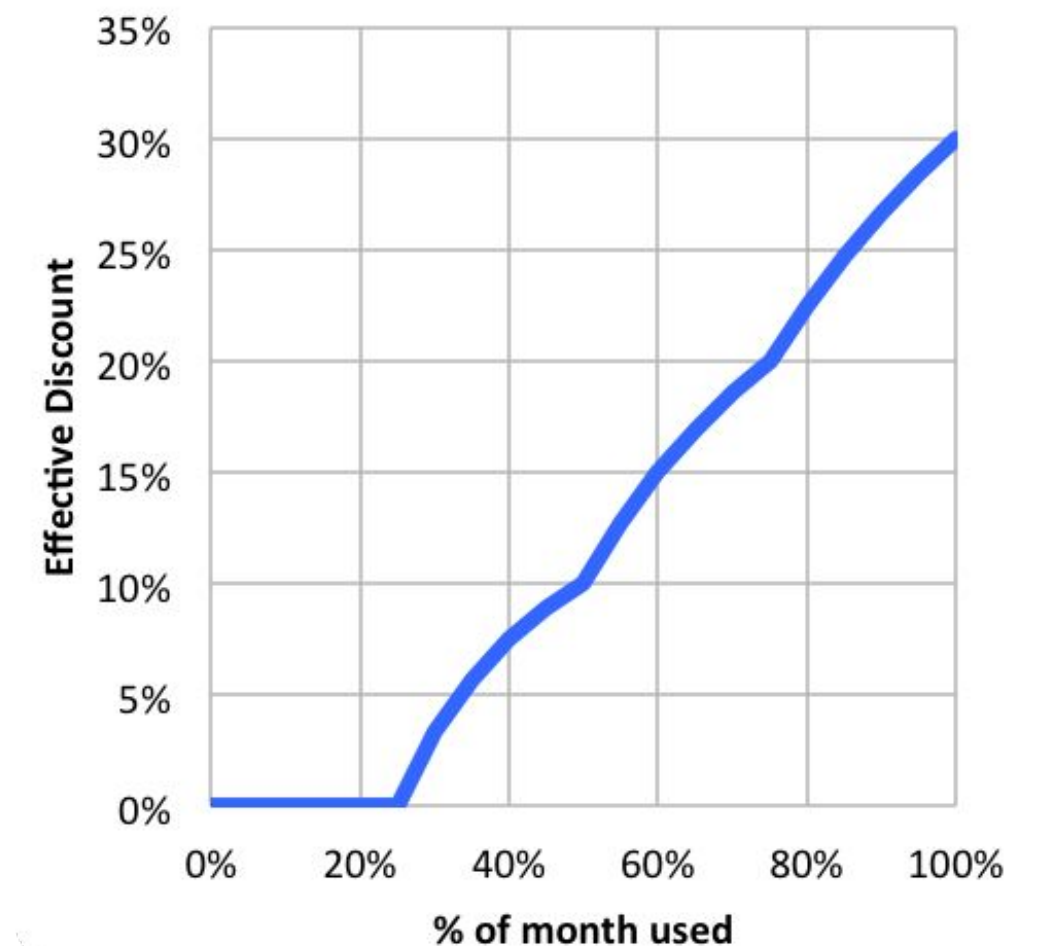
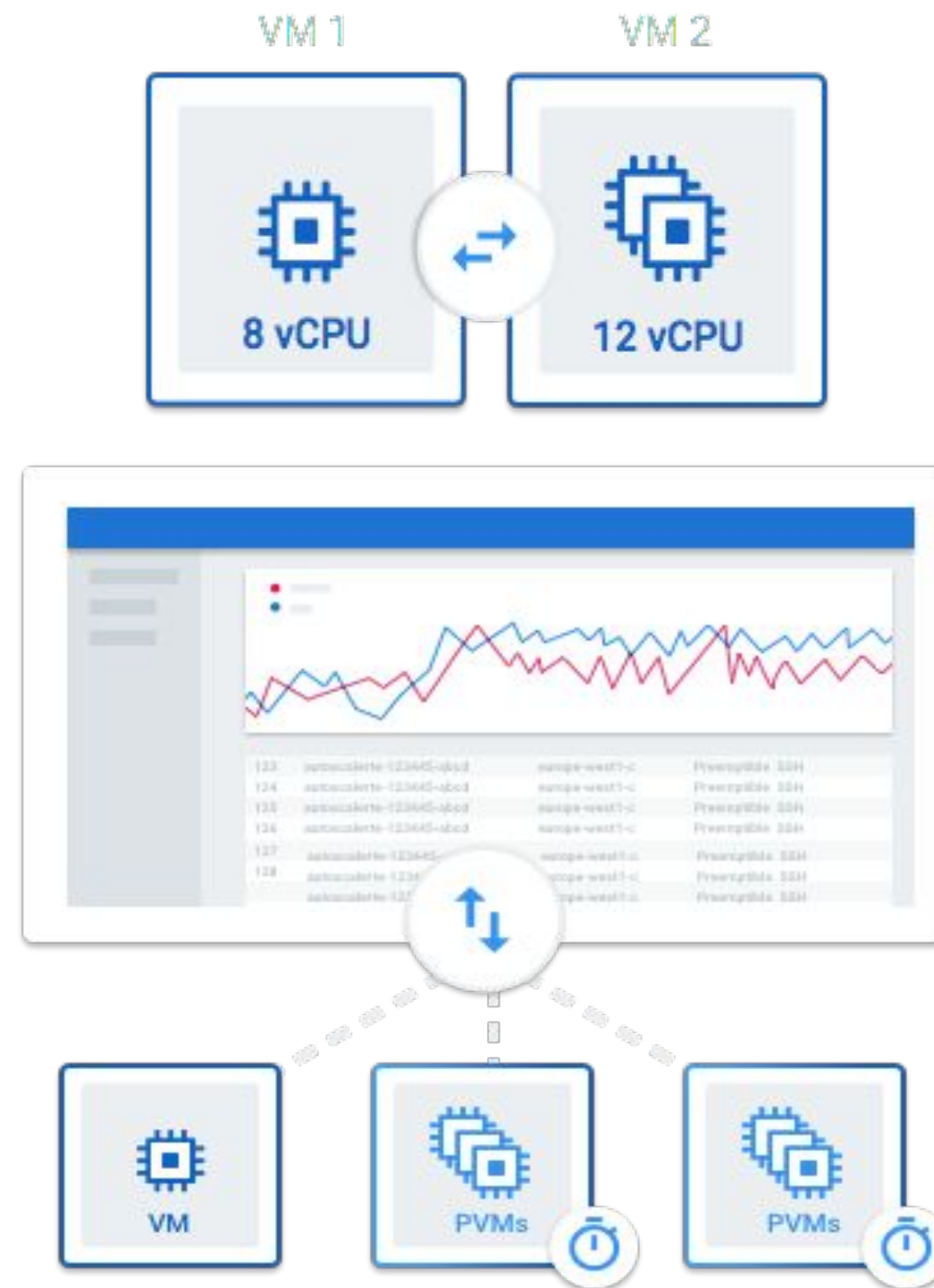
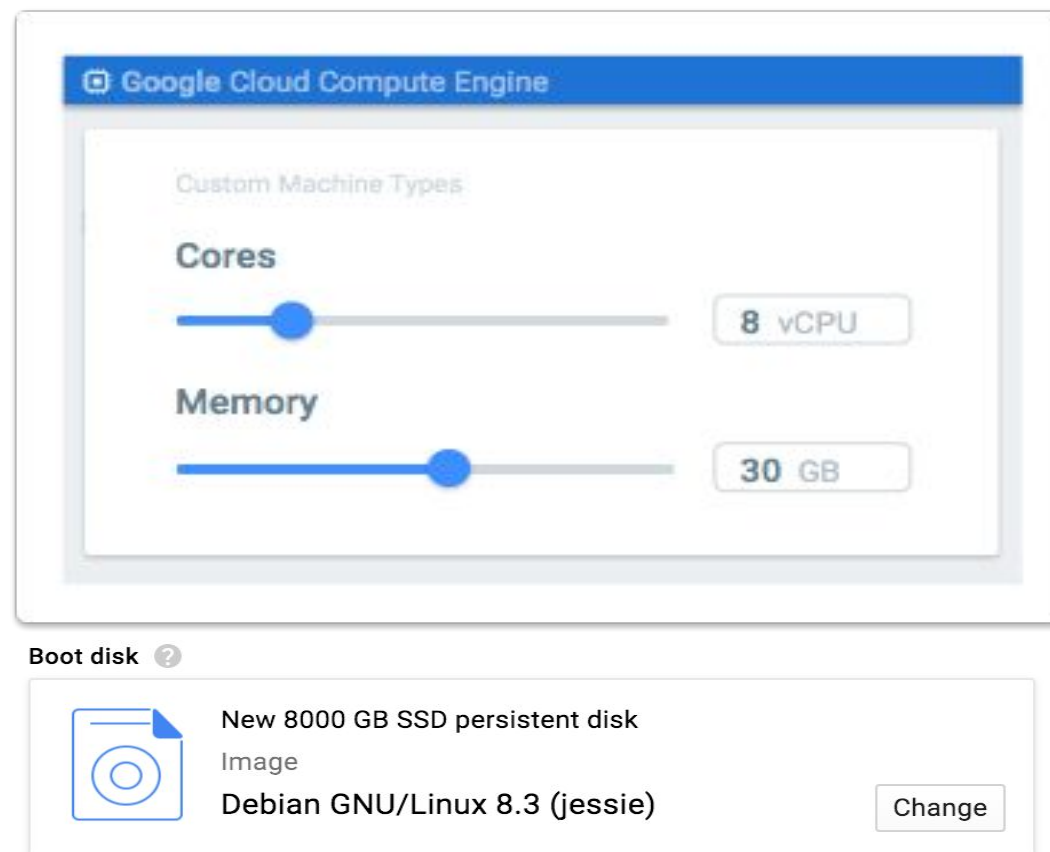
Indigo (SG, ID, AU) 2019

Monet (US, BR) 2017

Junior (Rio, Santos) 2017

Tannat (BR, UY, AR) 2017

Custom/changeable machine types, preemptible machines, and automatic discounts lead to simplicity and agility



Lab: Create a Compute Engine instance

Lab 1: Create a Compute Engine Instance

In this lab you will :

1. Create a Compute Engine instance
2. SSH into the instance
3. Install the software package git (for source code version control)

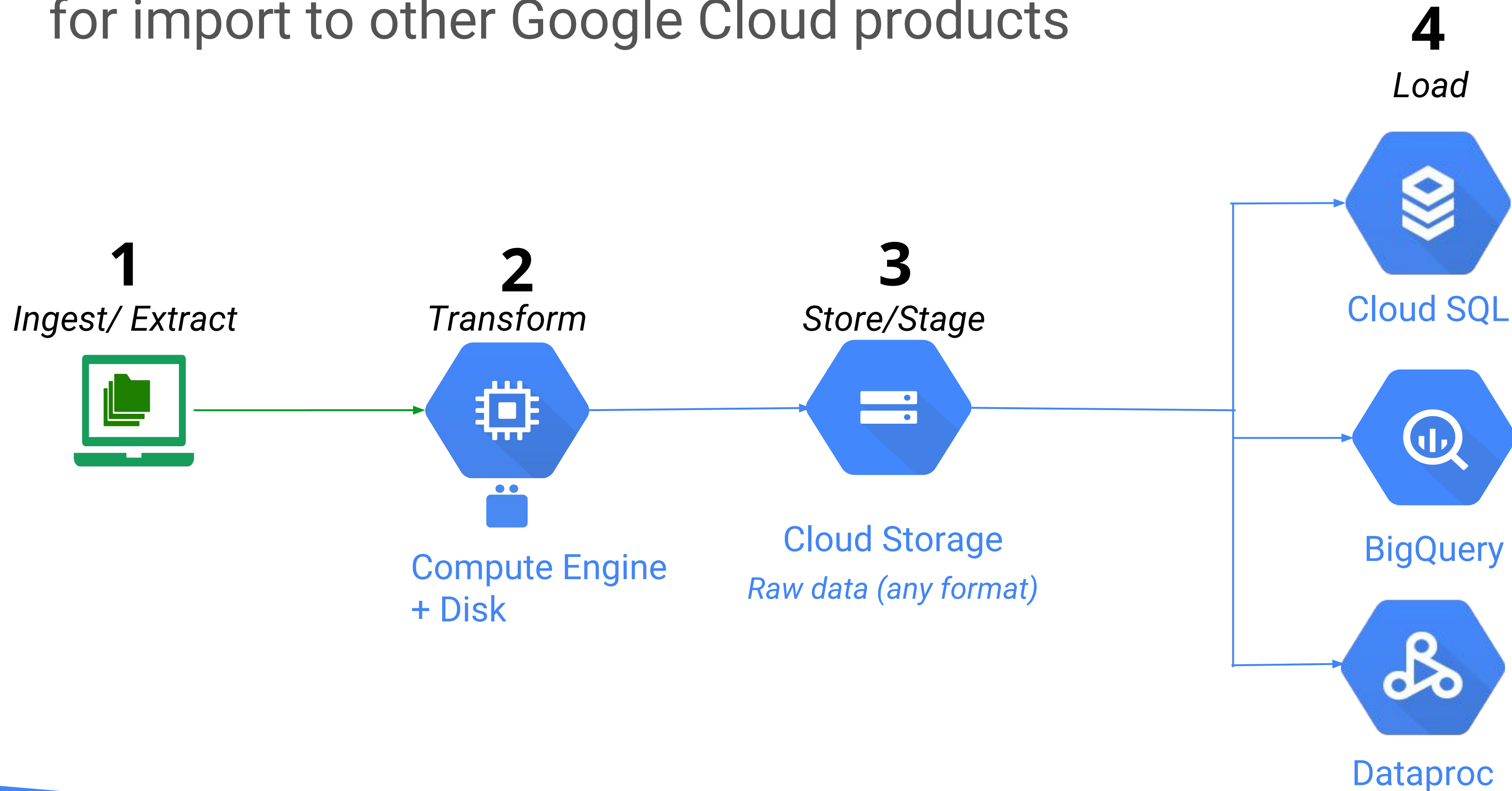


Agenda

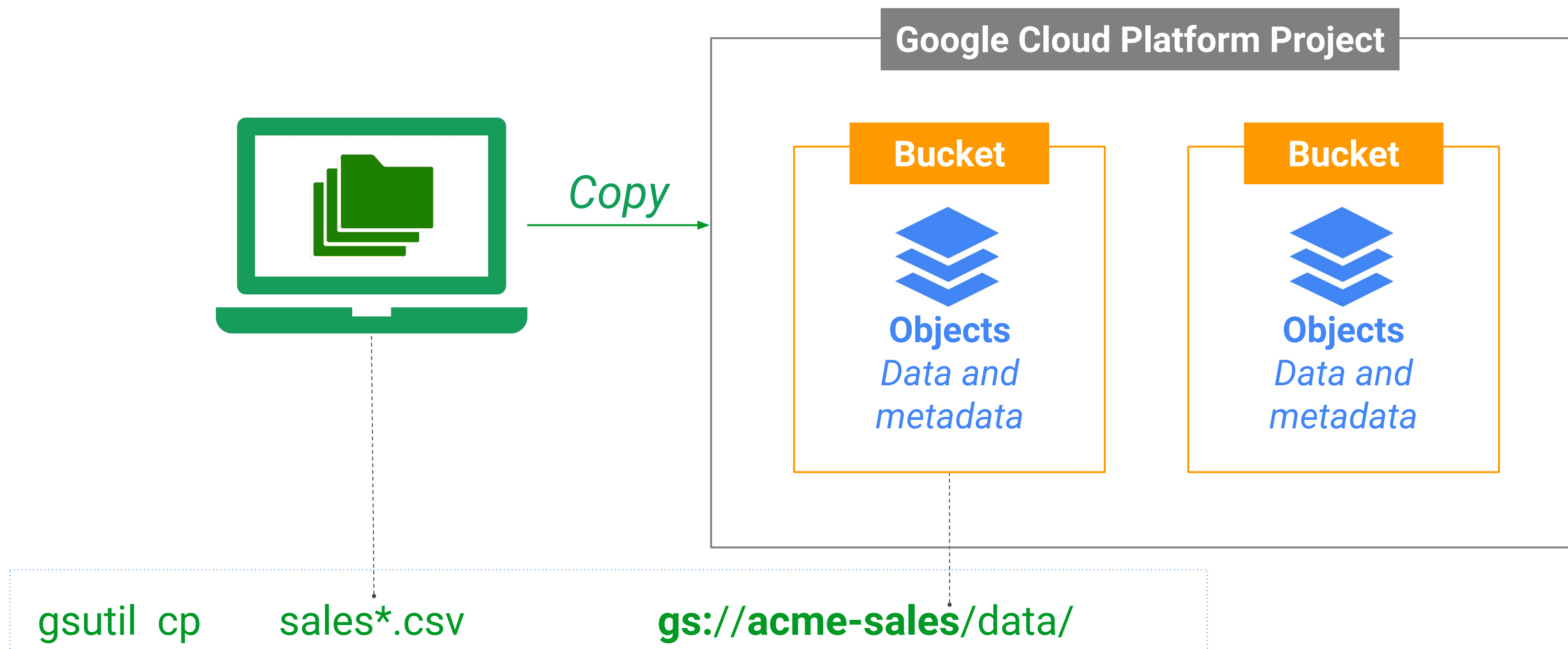
CPUs on demand + Lab

A global filesystem + Lab

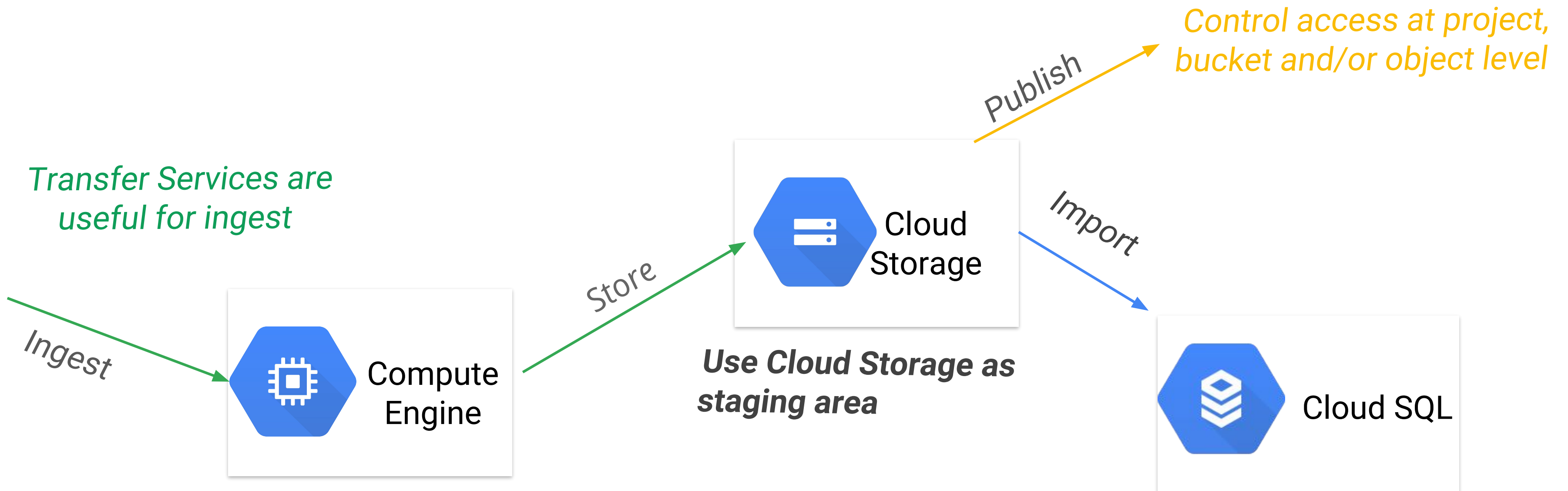
Use Cloud Storage for persistent storage and as staging ground for import to other Google Cloud products



Create a bucket and copy the data over using the Cloud SDK;
blobs are referenced through a `gs://.../` URL



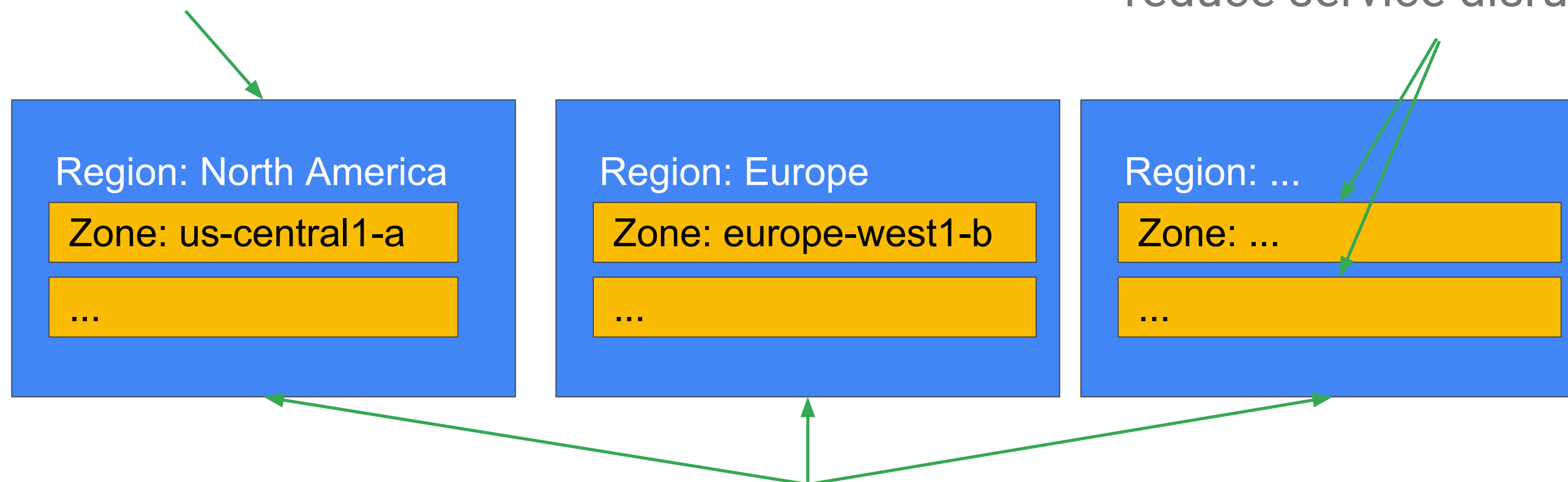
Cloud Storage gives you durability, reliability, and global reach



Control latency and availability with zones and regions

Choose the closest zone/region so as to reduce latency.

Distribute your apps and data across zones to reduce service disruptions.



Distribute your apps and data across regions for global availability.

Lab: Interact with Cloud Storage

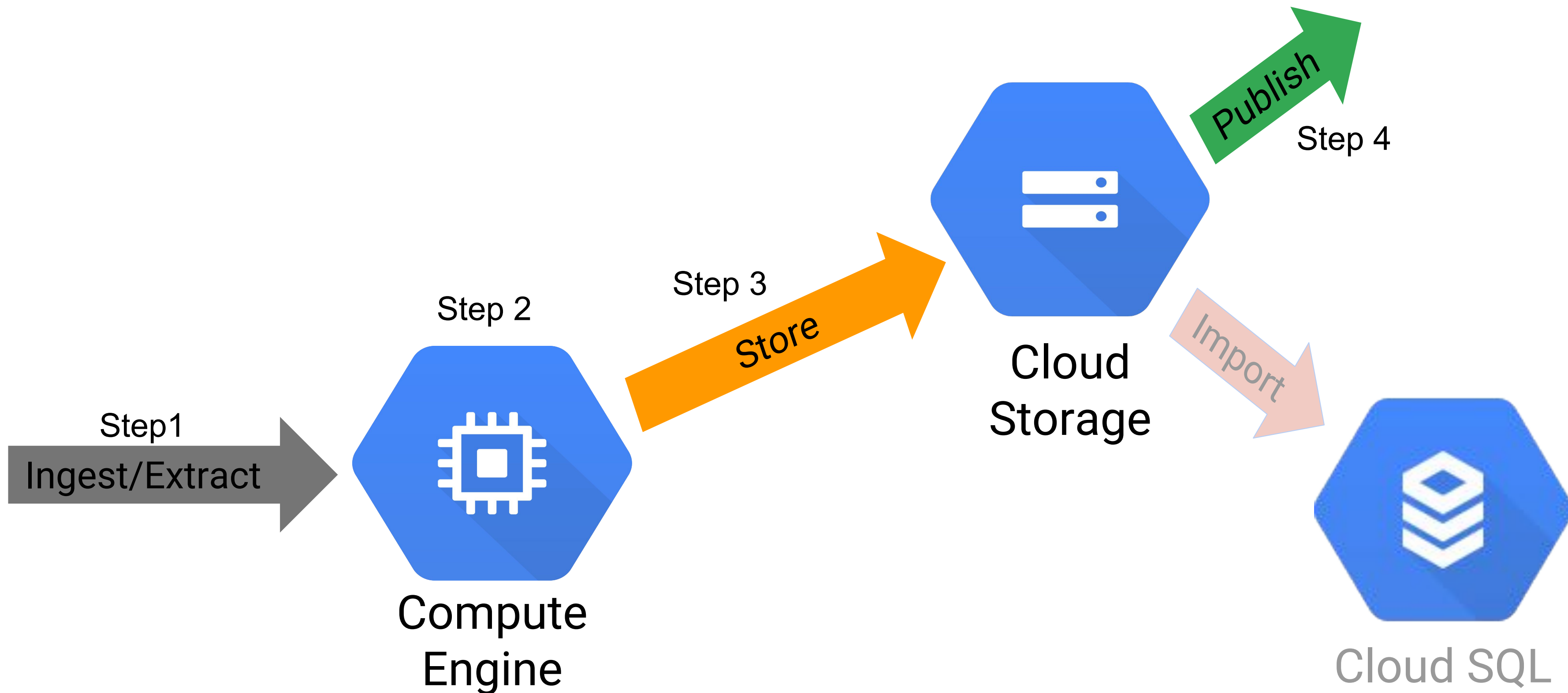
Lab 2: Interact with Cloud Storage

In this lab, you carry out the steps of an ingest-transform-and-publish data pipeline manually:

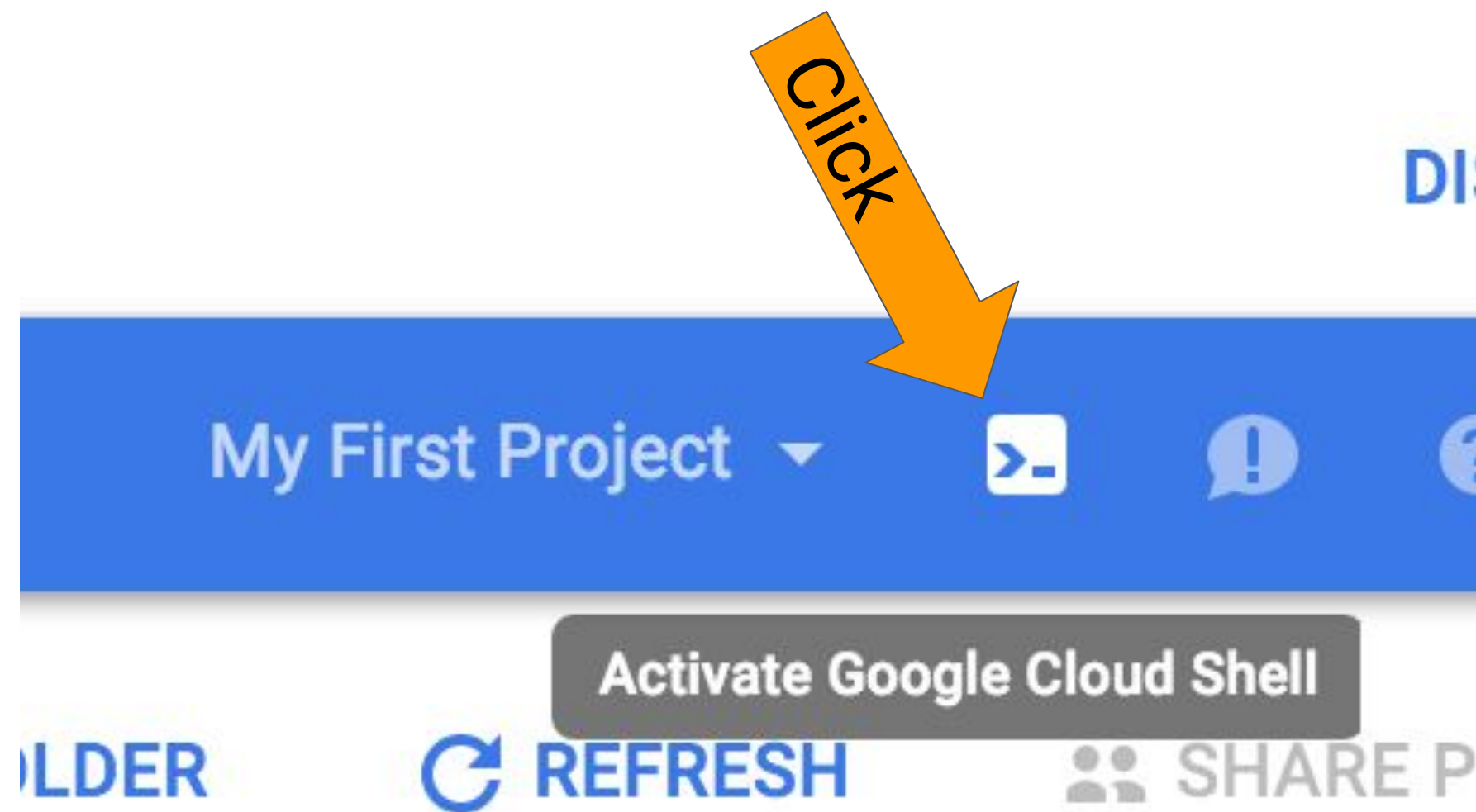
1. Ingest data into a Compute Engine instance
2. Transform data on the Compute Engine instance
3. Store the transformed data on Cloud Storage
4. Publish Cloud Storage data to the web



Ingest-Transform-Publish using core infrastructure



Cloud Shell gives you an easy command-line



Cloud Shell comes pre-installed with the tools, libraries, and so on you need to interact with Google Cloud Platform

Module Review

Module review (1 of 2)

Compute nodes on GCP are:
(select the correct option)

- ☐ Allocated on demand, and you pay for the time that they are up.
- ☐ Expensive to create and teardown
- ☐ Pre-installed with all the software packages you might ever need.
- ☐ One of ~50 choices in terms of CPU and memory

Module review answers (1 of 2)

Compute nodes on GCP are:
(select the correct option)

- ✓ Allocated on demand, and you pay for the time that they are up.
- ☐ Expensive to create and teardown
- ☐ Pre-installed with all the software packages you might ever need.
- ☐ One of ~50 choices in terms of CPU and memory

Module review (2 of 2)

Google Cloud Storage is a good option for storing data that:
(select all of the correct options)

- ☐ Is ingested in real-time from sensors and other devices
- ☐ Will be frequently read/written from a compute node
- ☐ May be required to be read at some later time
- ☐ May be imported into a cluster for analysis

Module review (2 of 2)

Google Cloud Storage is a good option for storing data that:
(select all of the correct options)

- ☐ Is ingested in real-time from sensors and other devices
- ☐ Will be frequently read/written from a compute node
- ✓ May be required to be read at some later time
- ✓ May be imported into a cluster for analysis

Resources

Compute Engine

<https://cloud.google.com/compute/>

Storage

<https://cloud.google.com/storage/>

Pricing

<https://cloud.google.com/pricing/>

Cloud Launcher

<https://cloud.google.com/launcher/>

Pricing Philosophy

<https://cloud.google.com/pricing/philosophy/>

cloud.google.com