Mil disculpas, ¡tienes toda la razón! El error es mío. Me confundí con el ID del proyecto.

Parece que el ID de tu proyecto es etl-project-003, pero yo estuve usando etl-lt-003 en todas las respuestas anteriores. Por supuesto que los comandos no funcionaban, ya que intentaban acceder a un proyecto que no existe.

**La solución definitiva**

Todos los errores de "Project not found or permission denied" se debieron a esta simple confusión en el ID del proyecto. Ahora que tenemos el ID correcto, todos los comandos deberían funcionar.

Aquí están los pasos correctos, usando el ID de proyecto **etl-project-003**:

1. **Configura el proyecto en tu Cloud Shell:**

Bash

gcloud config set project etl-project-003

1. **Activa la cuenta de servicio:** Asegúrate de usar la ruta correcta a tu archivo de clave JSON.

Bash

gcloud auth activate-service-account 409225416097-compute@developer.gserviceaccount.com --key-file=/home/luistrocco/etl-project-003-92b5286aa1f7.json --project=etl-project-003

1. **Habilita los servicios:** Ahora, con el proyecto y la cuenta de servicio correctos, este comando debería funcionar sin problemas.

Bash

gcloud services enable \

bigquery.googleapis.com \

compute.googleapis.com \

storage.googleapis.com \

composer.googleapis.com \

dataproc.googleapis.com \

dataflow.googleapis.com \

secretmanager.googleapis.com \

sqladmin.googleapis.com \

cloudfunctions.googleapis.com \

cloudscheduler.googleapis.com \

cloudresourcemanager.googleapis.com

Intenta estos pasos con el ID del proyecto correcto. Si hay algún otro error, por favor, avísame.

Ayudame ahora con este error, creo que es algo de escritura

Operation "operations/acat.p2-409225416097-182786bc-cd32-432e-9661-692183fed3ff" finished successfully.

luistrocco@cloudshell:~ (etl-project-003)$ CLUSTER\_NAME="my-demo-cluster2"

REGION="us-central1"

gcloud dataproc clusters create ${CLUSTER\_NAME} \

    --region ${REGION} \

    --num-workers=2 \

    --worker-machine-type=n1-standard-2 \

    --worker-boot-disk-size=50 \

    --master-machine-type=n1-standard-2 \

    --master-boot-disk-size=50 \

    --image-version=2.0-debian10 \

    --enable-component-gateway \

    --optional-components=JUPYTER \

    --initialization-actions=gs://goog-dataproc-initialization-actions-${REGION}/connectors/connectors.sh \

    --metadata bigquery-connector-version=1.2.0 \

    --metadata spark-bigquery-connector-version=0.21.0

ERROR: (gcloud.dataproc.clusters.create) unrecognized arguments:

To search the help text of gcloud commands, run:

  gcloud help -- SEARCH\_TERMS

-bash: --region: command not found

-bash: --num-workers=2: command not found

-bash: --worker-machine-type=n1-standard-2: command not found

-bash: --worker-boot-disk-size=50: command not found

-bash: --master-machine-type=n1-standard-2: command not found

-bash: --master-boot-disk-size=50: command not found

-bash: --image-version=2.0-debian10: command not found

-bash: --enable-component-gateway: command not found

-bash: --optional-components=JUPYTER: command not found

-bash: --initialization-actions=gs://goog-dataproc-initialization-actions-us-central1/connectors/connectors.sh: No such file or directory

-bash: --metadata: command not found

-bash: --metadata: command not found

Claro, te ayudo con ese error. Tienes razón, es un problema de escritura, pero en este caso, es un poco más sutil que los anteriores.

El problema radica en los caracteres invisibles que se añaden al copiar y pegar el código. El error unrecognized arguments y los mensajes de -bash: command not found indican que la terminal no está interpretando correctamente las barras invertidas (\) que se usan para dividir el comando en múltiples líneas.

Lo que está pasando es que la terminal piensa que cada línea es un comando independiente, y por eso te da errores como --region: command not found. Esto se debe a que hay un carácter invisible, como un espacio extra, después de la barra invertida \ en cada línea.

**La solución**

Simplemente, vuelve a escribir el comando manualmente en la terminal. Al hacerlo, te aseguras de que no haya espacios u otros caracteres ocultos después de las barras invertidas.

Así es como debe lucir el comando correctamente formateado. Cada línea, excepto la última, debe terminar con una barra invertida (\) **inmediatamente antes del salto de línea**, sin espacios.

Bash

CLUSTER\_NAME="my-demo-cluster2"

REGION="us-east1"

gcloud dataproc clusters create ${CLUSTER\_NAME} \

--region ${REGION} \

--num-workers=2 \

--worker-machine-type=n1-standard-2 \

--worker-boot-disk-size=50 \

--master-machine-type=n1-standard-2 \

--master-boot-disk-size=50 \

--image-version=2.0-debian10 \

--enable-component-gateway \

--optional-components=JUPYTER \

--initialization-actions=gs://goog-dataproc-initialization-actions-${REGION}/connectors/connectors.sh \

--metadata bigquery-connector-version=1.2.0 \

--metadata spark-bigquery-connector-version=0.21.0

1. **Copia el comando de arriba y pégalo en una sola línea en tu terminal**, o escríbelo directamente.
2. Presiona **Enter** solo una vez, al final de todo el comando. La terminal procesará las barras invertidas y ejecutará el comando como si fuera una sola línea larga.

Con esto, debería funcionar correctamente.

Sabes que me esta dando un error de region, pero no entiendo por que siempre toma la b, sabes como podria tomar cualquiera y ya?

WARNING: The specified custom staging bucket 'dataproc-staging-us-central1-409225416097-kjfyh2jw' is not using uniform bucket level access IAM configuration. It is recommended to update bucket to enable the same. See https://cloud.google.com/storage/docs/uniform-bucket-level-access.

Waiting for cluster creation operation...done.

ERROR: (gcloud.dataproc.clusters.create) Operation [projects/etl-project-003/regions/us-central1/operations/041fcdb0-6277-3299-80d6-cd8e4283f802] failed: Error Code: UNAVAILABLE, errorSource: COMPUTE\_ENGINE, Error Message: The zone 'projects/etl-project-003/zones/us-central1-b' does not have enough resources available to fulfill the request.  Try a different zone, or try again later.

Gracias por enviarme el error. Lo que me muestras es un problema común cuando se trabaja con Google Cloud.

El error The zone 'projects/etl-project-003/zones/us-central1-b' does not have enough resources available to fulfill the request. Try a different zone, or try again later significa exactamente lo que dice: en este momento, no hay suficientes máquinas virtuales disponibles en la zona us-central1-b para crear tu clúster de Dataproc.

Cuando tú especificas una **región** como us-central1 pero no una **zona** específica, Google Cloud automáticamente selecciona una zona dentro de esa región. En este caso, eligió la us-central1-b, que actualmente no tiene los recursos que necesitas.

**Solución: Especificar una Zona Diferente**

Para evitar este problema, puedes especificar una zona manualmente en tu comando. Esto le dice a Google Cloud que intente crear tu clúster en una zona específica, en lugar de dejar que elija una automáticamente.

Puedes probar con otras zonas dentro de la región us-central1, como us-central1-a o us-central1-c.

Aquí te muestro cómo modificar tu comando:

1. **Define la zona en una variable:** La forma más limpia es añadir una variable ZONE a tu script.
2. **Añade el flag --zone:** Incluye el flag --zone ${ZONE} en el comando gcloud dataproc clusters create.

Aquí tienes el comando completo con la modificación:

Bash

CLUSTER\_NAME="my-demo-cluster2"

REGION="us-central1"

ZONE="us-central1-a" # O puedes probar con "us-central1-a"

gcloud dataproc clusters create ${CLUSTER\_NAME} \

--region ${REGION} \

--zone ${ZONE} \

--num-workers=2 \

--worker-machine-type=n1-standard-2 \

--worker-boot-disk-size=50 \

--master-machine-type=n1-standard-2 \

--master-boot-disk-size=50 \

--image-version=2.0-debian10 \

--enable-component-gateway \

--optional-components=JUPYTER \

--initialization-actions=gs://goog-dataproc-initialization-actions-${REGION}/connectors/connectors.sh \

--metadata bigquery-connector-version=1.2.0 \

--metadata spark-bigquery-connector-version=0.21.0

Si el problema persiste con una zona, simplemente prueba con otra. También puedes intentar de nuevo más tarde, ya que la disponibilidad de recursos puede cambiar.

**Nota:** El WARNING sobre el uniform bucket level access es solo una recomendación de seguridad y no es lo que causa el error. Puedes ignorarlo por ahora, ya que no impide la creación del clúster.

CLUSTER\_NAME="my-demo-cluster2"

REGION="us-central1"

ZONE="us-central1-a" # O cualquier otra de las zonas

gcloud dataproc clusters create ${CLUSTER\_NAME} \

--region ${REGION} \

--zone ${ZONE} \

--num-workers=2 \

--worker-machine-type=n1-standard-2 \

--worker-boot-disk-size=50 \

--master-machine-type=n1-standard-2 \

--master-boot-disk-size=50 \

--image-version=2.0-debian10 \

--enable-component-gateway \

--optional-components=JUPYTER \

--initialization-actions=gs://goog-dataproc-initialization-actions-${REGION}/connectors/connectors.sh \

--metadata bigquery-connector-version=1.2.0 \

--metadata spark-bigquery-connector-version=0.21.0