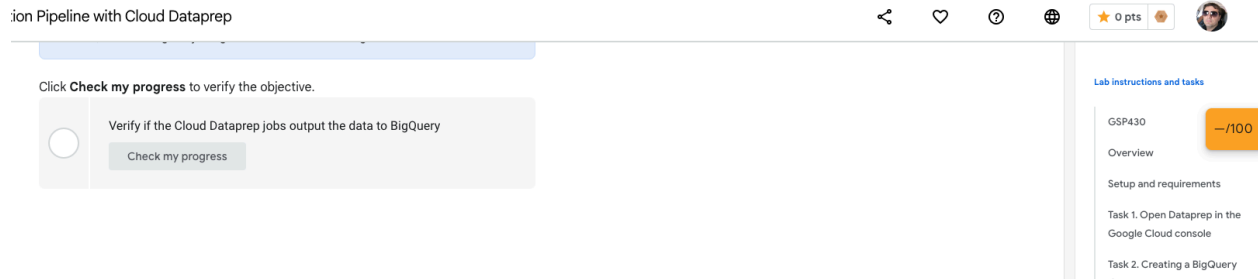


Google Skills Boost - Examen Final

LAB:

Realizar el siguiente LAB, al finalizar pegar un print screen donde se ve su perfil y el progreso final verificado



Título: Creating a Data Transformation Pipeline with Cloud Dataprep

Schedule: 1 hour 15 minutes

Cost: 5 Credits

Link:

https://www.cloudskillsboost.google/focuses/4415?catalog_rank=%7B%22rank%22%3A1%2C%22num_filters%22%3A0%2C%22has_search%22%3Atrue%7D&parent=catalog&search_id=32278924

Contestar las siguientes preguntas:

1. ¿Para que se utiliza data prep?
2. ¿Qué cosas se pueden realizar con DataPrep?
3. ¿Por qué otra/s herramientas lo podrías reemplazar? Por qué?
4. ¿Cuáles son los casos de uso comunes de Data Prep de GCP?
5. ¿Cómo se cargan los datos en Data Prep de GCP?
6. ¿Qué tipos de datos se pueden preparar en Data Prep de GCP?
7. ¿Qué pasos se pueden seguir para limpiar y transformar datos en Data Prep de GCP?
8. ¿Cómo se pueden automatizar tareas de preparación de datos en Data Prep de GCP?
9. ¿Qué tipos de visualizaciones se pueden crear en Data Prep de GCP?
10. ¿Cómo se puede garantizar la calidad de los datos en Data Prep de GCP?

Arquitectura:

El gerente de Analitca te pide realizar una arquitectura hecha en GCP que contemple el uso de esta herramienta ya que le parece muy fácil de usar y una interfaz visual que ayuda a sus desarrolladores ya que no necesitan conocer ningún lenguaje de desarrollo.

Esta arquitectura debería contemplar las siguiente etapas:

Ingesta: datos parquet almacenados en un bucket de S3 y datos de una aplicación que guarda sus datos en Cloud SQL.

Procesamiento: filtrar, limpiar y procesar datos provenientes de estas fuentes

Almacenar: almacenar los datos procesados en BigQuery

BI: herramientas para visualizar la información almacenada en el Data Warehouse

ML: Herramienta para construir un modelo de regresión lineal con la información almacenada en el Data Warehouse