

## PROGRAMA PASANTÍAS GLOBANT - TALENTO DIGITAL PARA CHILE 2023

### BIENVENIDA AL CODE CHALLENGE

Te invitamos a participar de un Code Challenge, para Globant es una etapa una etapa esencial para evaluar tus habilidades técnicas y tu capacidad para enfrentar desafíos en el mundo digital.

Las entregas serán evaluadas por un equipo de profesionales expertos del área de datos de Globant, conjuntamente con los antecedentes entregados en la etapa anterior para decidir quienes pasan a última etapa del proceso de selección de pasantes que corresponde a una entrevista.

#### Plazos de ejecución

**Inicio:** 11.10.2023

**Cierre:** 13.10.2023 a las 11:00 horas.

#### Instrucciones generales para el Code Challenge:

**1. Acceso a la Plataforma:** Tendrás acceso a un enlace para enviar tus respuestas, el cual permite grabar sólo una entrega por postulante. Asegúrate de tener todos los elementos listos para el envío antes de iniciar la carga de archivos y de entender su funcionamiento antes de la fecha programada.

**2. Desafío Técnico:** El desafío consta de instrucciones técnicas que evaluarán tus habilidades en programación, resolución de problemas y comprensión de conceptos relevantes. Asegúrate de seguir detalladamente cada indicación estipulada en el documento, puedes releer las veces que sea necesario.

**3. Tiempo Límite:** El Code Challenge debe completarse en el tiempo asignado (Tope viernes 13 de octubre a las 11:00 horas). Asegúrate de comenzar a tiempo para poder abordar todos los requerimientos.

**4. Preparación:** No dudes en revisar tus conocimientos técnicos y repasar conceptos relevantes antes del desafío. Puedes utilizar recursos en línea y libros de referencia si es necesario.

**5. Ética y Originalidad:** Por favor, completa el Code Challenge de manera ética y original.

**6. Evaluación y Resultados:** Una vez que completes el desafío, el equipo de selección de Globant, revisará tus respuestas. Te informaremos sobre los resultados y las próximas etapas del proceso de selección en un plazo razonable.

Estamos muy felices por tu participación en este proceso y esperamos ver tus habilidades técnicas en acción.

**¡Te deseamos el mayor de los éxitos!**

## CODE CHALLENGE, INSTRUCCIONES:

Desarrollar un pequeño proceso de ETL que recoja datos desde una fuente, los transforme y luego los cargue en BigQuery. Durante este proceso, es necesario aplicar metodologías Agile y utilizar herramientas de control de versiones.

- Utilizar Python o Java para escribir un script que simule la extracción de datos desde una fuente. Puede ser una API ficticia o un archivo CSV.
- Limpiar y transformar los datos: eliminar registros duplicados, completar valores faltantes, etc.
- Usar SQL para segmentar y agrupar la información.
- En esta fase, es recomendable usar SQL básico: consultas, agrupamientos, DDL, DML y filtros.
- Utilizar la SDK de GCP para cargar los datos transformados en BigQuery.
- El candidato o candidata debe garantizar que la estructura de la tabla en BigQuery sea adecuada para los datos.
- Desarrollar un proceso simple que garantice que la ingestión de datos se realice de manera periódica (ejemplo: cada 24 horas). Esto se puede hacer a través de herramientas como Apache Airflow o mediante cron jobs.
- Usar Git para control de versiones.
- Debe haber evidencia de ramificación (branching) y fusión (merging) apropiados en el repositorio.
- Incluir un `README.md` con instrucciones sobre cómo ejecutar el proceso y entender el flujo de trabajo. Documentar los supuestos utilizados para el desarrollo del etl con respecto a los datos.
- Organizar el desarrollo del proyecto en sprints. Documentar brevemente las tareas realizadas en cada sprint y los obstáculos encontrados.
- Utilizar herramientas como Trello o Jira para gestionar y organizar las tareas.

**Bonus:** si el candidato o candidata tiene experiencia con Scala o Javascript, puede incluir un pequeño snippet o componente que demuestre su capacidad para entender y trabajar con estos lenguajes, aunque no es el enfoque principal de este desafío.

## ENTREGABLES

- Repositorio Git con el código fuente, instrucciones y documentación.
- URL de la tabla en BigQuery con los datos cargados.
- Evidencia de la orquestación de las ingestas (por ejemplo, capturas de pantalla de Airflow o logs del cron job).
- Tablero Agile con la planificación y gestión del proyecto.

### Ejemplos de archivo CSV:

```
transaction_id,product_name,quantity_sold,date_sold,price,store_location,user_agent
1001,Smartphone,2,2023-10-01,300,New York,"Mozilla/5.0 (Android)" 1002,Laptop,1,2023-10-02,800,Los Angeles,"Mozilla/5.0 (Windows)" 1002,Laptop,,2023-10-02,,,,"Mozilla/5.0 (Windows)" # Registro duplicado con valores faltantes
1003,Headphones,5,2023-10-03,50,Chicago,"Mozilla/5.0 (iOS)" 1004,Smartwatch,3,2023-10-04,150,San Francisco,"Mozilla/5.0 (Android)" 1005,Tablet,2,2023-10-05,400,Boston,"Mozilla/5.0 (macOS)" 1006,Camera,1,2023-10-06,500,Miami,"Mozilla/5.0 (Windows)" 1007,Charger,10,2023-10-07,15,Seattle,"Mozilla/5.0 (Linux)" 1008,Mouse,4,2023-10-08,25,Dallas,"Mozilla/5.0 (Android)" 1009,Keyboard,3,2023-10-09,45,Houston,""
1010,Monitor,2,2023-10-10,200,Philadelphia,"Mozilla/5.0 (macOS)"
```