

# Tarea 2

## Bases de Datos a gran escala

mail: claudio.torresf@usm.cl

### 1. Objetivo

El objetivo de esta tarea es analizar y comparar productos de diversos proveedores de nubes públicas con el fin de elegir la mejor tecnología acorde a las necesidades del cliente.

### 2. Descripción

“Data Solutions” es una empresa dedicada a la generación de reportes de Recursos Humanos, debido a su éxito se encuentra en rápido crecimiento. Dada la ambiciosa meta de expansión de la compañía, la gerencia te ha seleccionado para liderar la investigación y recomendación del proveedor de servicios en la nube más adecuado para soportar esta nueva etapa.

Tu misión consiste en investigar y comparar exhaustivamente **dos proveedores líderes** de tu elección (por ejemplo, Amazon Web Services, Google Cloud Platform, Microsoft Azure, entre otros). El objetivo final es presentar a la gerencia un análisis detallado que les permita tomar una decisión informada sobre qué plataforma en la nube se alinea mejor con las necesidades tecnológicas y estratégicas de “Data Solutions”.

Una vez seleccionados tus proveedores, deberás identificar y analizar los productos específicos que ofrecen para satisfacer los siguientes requerimientos clave, cruciales para la infraestructura de la expansión:

- **Plataforma de Machine Learning con soporte para Jupyter Notebooks:** Investiga qué servicios ofrecen tus proveedores para el desarrollo y experimentación de modelos de aprendizaje automático, asegurándote de que sean compatibles con el entorno de desarrollo interactivo Jupyter Notebooks. Esto es fundamental para la futura implementación de soluciones inteligentes en “Data Solutions”.
- **Cómputo:** Explora las soluciones de cómputo que cada proveedor pone a disposición, prestando especial atención a:
  - **Máquinas Virtuales (VMs):** ¿Qué tipos de instancias ofrecen y cuáles son sus características principales en términos de escalabilidad y rendimiento para las crecientes demandas de “Data Solutions”?
  - **Servicio de Contenedores como Servicio (CaaS):** ¿Cuál es su oferta para la gestión y orquestación de contenedores, considerando la agilidad y eficiencia que requiere la expansión?
  - **Procesamiento de Datos Batch y Stream Serverless con compatibilidad con Apache Beam:** ¿Qué servicios serverless proporcionan para el procesamiento de grandes volúmenes de datos en lote y en tiempo real, y cómo se integran con Apache Beam para el análisis de la información generada por la expansión?

- **Almacenamiento:** Analiza las opciones de almacenamiento que cada proveedor ofrece para los siguientes tipos de datos, considerando la escalabilidad y seguridad necesarias para el crecimiento de “Data Solutions”:
  - **Almacenamiento de Objetos (Object Storage):** ¿Cuáles son sus características de escalabilidad, durabilidad y acceso para el almacenamiento de datos no estructurados?
  - **Data Warehouse:** ¿Qué soluciones de almacenamiento analítico ofrecen para el análisis de grandes conjuntos de datos y la generación de insights estratégicos para la expansión?
  - **Sistema de Gestión de Bases de Datos Relacionales (RDBMS) compatible con MySQL:** ¿Cuál es su oferta de bases de datos relacionales gestionadas compatibles con MySQL para las aplicaciones transaccionales de “Data Solutions”?

### 3. Actividades

Para cada uno de los requerimientos mencionados, deberás (100 puntos):

1. **Identificar los servicios específicos** que ofrece cada proveedor. Debes seleccionar 1 servicio por proveedor cloud en cada categoría (20 puntos):
  - Plataforma de Machine Learning
  - VMs
  - CaaS
  - Procesamiento de Datos Batch y Stream Serverless
  - Almacenamiento de Objetos
  - Data Warehouse
  - RDBMS
2. **Describir brevemente las características principales** de cada servicio seleccionado en el punto 1, enfocándote en cómo podrían beneficiar a “Data Solutions”. Mencionar al menos dos características. (30 puntos)
3. **Comparar las similitudes y diferencias** entre las ofertas de ambos proveedores, destacando posibles ventajas y desventajas para las necesidades particulares de la empresa. **Un punto crucial a considerar en esta comparación es el modelo de precios de cada servicio, incluyendo posibles costos ocultos y opciones de escalabilidad de costos.** Mencionar al menos una similitud, una diferencia y los precios. (40 puntos)
4. **Conclusiones.** (10 puntos)

Finalmente, deberás elaborar un informe conciso y bien estructurado que presente tus hallazgos y conclusiones sobre la comparación de los proveedores y sus respectivos productos, incluyendo una recomendación justificada para la gerencia de “Data Solutions”. ¡El futuro de la expansión de la compañía está en tus manos!

## 4. Consideraciones

- La tarea puede ser desarrollada en grupo de máximo 3 estudiantes.
- El informe debe poseer el nombre de todos los integrantes del equipo de trabajo. Este debe ser en formato PDF.
- Se debe enviar un informe por equipo en la plataforma AULA.
- Está permitido utilizar Inteligencias Artificiales para el desarrollo de la tarea, pero debe agregar el prompt utilizado en el informe.

**Fecha entrega: 11 Mayo 2025**