



# Universidad Técnica Federico Santa María

INGENIERÍA CIVIL EN INFORMÁTICA

## APACHE BEAM Y APACHE AIRFLOW

*Base de Datos a Gran Escala*

Autores:

Benjamín Nicolás Daza Jiménez

Daniel Rodrigo Maturana Cristino

Sebastián Andrés Von Kunowsky Lepe

## Apache Beam

La implementación de Apache Beam no presentó grandes desafíos, ya que la experiencia adquirida en la tarea anterior facilitó su desarrollo. El único problema fue lograr que funcionara tanto en la nube como de forma local, lo que requirió entender las diferencias en el manejo de los paths entre ambos entornos.

## Apache Airflow

La implementación de Apache Airflow sí presentó diversos desafíos que dificultaron el desarrollo de la tarea. Entre ellos estuvieron la configuración en contenedores e integrarlo con Visual Studio Code, abarcando el archivo `devcontainer.json`, el archivo `docker-compose.yml` y los scripts de dependencias (`*.sh`). Además, surgió el problema de lograr que Airflow reconociera correctamente la carpeta `dags`. Finalmente, otro desafío fue hacer que Airflow pudiera enviar el mensaje a Kafka, ya que el contenedor no reconocía el servicio; esto se solucionó añadiendo una instrucción para levantar el servicio de Kafka junto con la app.