

# Presentación Ejecutiva: Proyecto Final Gran Escala

Luis Roman, Alfonso Kim, Fernanda Mora, Andrea García y Sergio Haro

5 de junio de 2015



# Arquitectura

El flujo de datos debe estar orientado a procesos y no tareas fragmentadas. El proceso de orquestación es vital para asegurar la resiliencia del flujo de datos.

En cuanto a la ingesta de datos preferimos Flume a Kite ya que Flume te permite procesar y filtrar directamente sin pasar por tantos procesos como en Kite (para leer de diferentes fuentes y esquemas, procesarlos y pasarlos a diferentes destinos etc..). Sqoop nos gustó por qué nos permite manejar los datos desde y hacia una base relacional. a pesar de ser menos potente que Flume ( por qué solo es para BD relacionales) nos gustó por su facilidad de uso (se conecta a Postgres y a otras bases de datos).

## Extras

Para mayor información dirigirse a la presentación extendida y a la documentación del trabajo [https://github.com/AndyGT/big-data/tree/master/grupos/LAFAS/Proy%20final\\_v4](https://github.com/AndyGT/big-data/tree/master/grupos/LAFAS/Proy%20final_v4)

“proyecto\_final.pdf”

“presentacion\_extendida.html”

