

Tarea 1

Otoño 2017

Usted ha sido contactado por la multinacional empresa de retail MARI (*Multinational Amazing Retail Inc.*). Sus ventas han sido tan buenas en la última década que han comenzado a invertir en I+D para no quedarse desactualizados respecto a los competidores más fuertes. Como se podrá imaginar, la experiencia de ellos está más ligado a lo que son las ventas al menor, pero el departamento de I+D insiste en traer conocimientos computacionales a la empresa, ya que la CEO visualiza que en un futuro próximo la única forma que les permita seguir siendo competitivos es a través del análisis real de las compras de sus clientes.

Actualmente, generan una cantidad de datos basado en boletas que con las tecnologías tradicionales cada vez tiene más complicaciones al momento de procesar. Dada esta situación, deciden contactarlo a usted para que les ayuden a dar con una solución. En una primera instancia para saber si pueden confiar en ustedes, les piden que hagan un Graph Schema, luego generar la estructura de datos necesaria para Java usando Apache Thrift, y finalmente inserten los datos en un cluster Hadoop.

Al ser Chile un buen lugar para probar nuevas tecnologías, la empresa les brindará acceso a una “cola” de Apache Kafka¹ desde donde obtendrán los datos en formato *JSON*, que corresponde a los datos que tendrá que insertar. A priori, al no saber que preguntas necesitará responder, su modelo de datos tendrá que ser capaz de almacenar “todos”² los campos que se incluyan.

En síntesis, las labores que tienen que realizar para ver si la empresa MARI sigue trabajando con ustedes son las siguientes:

- Definir un Graph Schema y escribir la estructura en Apache Thrift.
- Generar el *schema* para Java.
- Almacenar los datos en el HDFS de un cluster Hadoop, usando Pail y serializándolos con Apache Thrift.

¹ No necesita saber usar Apache Kafka, solo conectarse a este.

² Si cree que algo no debe ser considerado por su modelo puede hacerlo, pero tendrá que dar las razones de hacerlo.

La entrega debe contener los códigos en Java que se usen (80%) y un informe de no más de 4 páginas que detalle el trabajo realizado lo que corresponderá al 20% de la tarea.

Fecha de entrega códigos e informe: 16 de Abril del 2017 a las 23:59
No hay atrasos.