## Instituto Tecnológico de Costa Rica, Escuela de Computación, Bases de Datos II Prof. Rodrigo Núñez Núñez, Proyecto II, Big Data, parte III, 10%

## **ETLing**

Utilizando los data sets *detallados* (no los sumarizados), de listings y calendar de AirBnb, en <a href="http://insideairbnb.com/get-the-data.html">http://insideairbnb.com/get-the-data.html</a> , los mismos que usó para la parte II, proceda a realizar las siguientes actividades:

- 1. Diseñe e implemente un ETL usando Microsoft Integration Services que realice las siguientes operaciones:
- Diseñe las tablas del datamart, usando copo de nieve, use un solo datamart que va existir en MemSQL
- Realice el ETL entre diversas fuentes y MemSQL, las fuentes son:
  - Los datos sumarizados hechos en la parte 2 del proyecto desde Hive/Metastore
  - Datos del estado del clima para los años contenidos en el csv del booking de Airbnb para las ciudades ya seleccionadas, procure agregar todo lo posible, temperatura, precipitación, nieve, todas las variables posibles por día
  - Las configuraciones habitacionales: número de habitaciones, baños, parqueos, cocina, sala, terrazas, etc.
  - Migre todos los bookings de cada listing

## **Otros aspectos**

- El trabajo es individual
- Cualquier mínima sospecha de copia anulará el trabajo de los involucrados.
- La revisión será con cita en el lab H del Tec
- Fecha de revisión martes 6 de octubre, 7pm a 9pm