

Trabajo CDR's de Milán

Para profundizar en el uso de MongoDB para el tratamiento de datos procedentes de los registros de datos o CDR (Call Detail Record) almacenados por las operadoras de telecomunicaciones, utilizaremos los datos de la ciudad de Milán tomados por Telecom Italia desde el 1 de noviembre de 2013 hasta el 1 de enero de 2014 disponibles en el siguiente enlace:

<https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/EGZHFV>

Cada alumno elegirá una semana completa de datos que incluya la totalidad de los datos tomados de lunes a domingo y realizará las siguientes acciones:

- 1.- Descargar los ficheros “.txt” correspondientes y crear un programa en Python para cargar los datos en la colección “Milan_CDR_c” de la base de datos “Milan_CDR_db”.
- 2.- Encuentra los países con los que se interactúa.
- 3.- Encuentra que país es con el que más se interactúa además de Italia
- 4.- ¿Qué celda comunica más con el extranjero?
- 5.- Encuentra la celda con más actividad de smsin, smsout callin, callout, internet y la total
- 6.- Crea una colección con una documento por celda en el que aparezcan los acumulados de los diferentes campos.
- 7.- Crea una colección con una documento por celda y hora en el que aparezcan los acumulados de los diferentes campos.
- 8.- Realiza un estudio de las celdas 4259 (Bocconi), 4456 (Navigli), 5060 (Duomo), 1419 (terreno agrícola), 2436 (área industrial), 4990 (aeropuerto de Linate), 945 (residencial aislado) y 5048 (residencial céntrico)