

## ***Tarea Programada #2***

- La tarea se debe entregar en el Github según el profesor haya asignado los grupos.
- Los documentos deben ser entregados en el Dropbox según el profesor haya asignado los archivos.
- Toda tarea debe ser defendida ante el profesor, de tal manera todos los estudiantes deben poder explicar la solución satisfactoriamente.
- ¡Buena Suerte!

## Estadísticas del Tiempo

La tarea consiste en 3 partes básicas.

1. Un Web Crawler que inserte datos en un almacenamiento que pueda ser leído por Hadoop.
2. Jobs de Hadoop que tomen la información del Web Crawler, la agregue y además, inserte en una Base de Datos de MariaDB/MySQL
3. Una aplicación Web que lea de la base de datos de MariaDB y retorne los resultados.

## Web Crawler

El Web Crawler debe de ingresar a la página <http://en.tutitiempo.net/climate> y bajar las estadísticas del tiempo a un archivo en hadoop..

Los datos a guardar deben ser iguales a los de la página, incluyendo el año y la siguiente tabla.

- |        |   |
|--------|---|
| 1. T   | Average annual temperature              |
| 2. TM  | Annual average maximum temperature      |
| 3. Tm  | Average annual minimum temperature      |
| 4. PP  | Rain or snow precipitation total annual |
| 5. V   | Annual average wind speed               |
| 6. RA  | Number of days with rain                |
| 7. SN  | Number of days with snow                |
| 8. TS  | Number of days with storm               |
| 9. FG  | Number of foggy days                    |
| 10. TN | Number of days with tornado             |
| 11. GR | Number of days with hail                |

Tomen en cuenta que el proceso de procesar e insertar estos datos puede tomar tiempo así que no lo dejen para el último minuto. Esto debe ser la primer parte de la tarea.

El formato específico del archivo en hadoop lo pueden diseñar ustedes.

Pueden omitir cualquier información que falta o que no se ajuste a los formatos válidos.

## Hadoop

Una vez que tengan los datos de las páginas web, debemos calcular la siguiente información:

1. Para cada una de las 11 variables necesitaremos
  - a. Los 10 países con los máximos promedios generales
  - b. Los 10 países con los mínimos promedios generales
  - c. Para cada país el año en que cada uno de las variables fue la máxima
  - d. Para cada país el año en que cada uno de las variables fue la mínima
  - e. El promedio de temperatura para cada continente, en grupos de 10 años
  - f. Por País la estación que tiene los valores máximos

- g. Por País la estación que tiene los valores mínimos
- h. Por Continente los países con los valores máximos
- i. Por Continente los países con los valores mínimos

Esta información debe de insertarse en una base de datos de MariaDB, el diseño de la base de datos lo deben de hacer ustedes.

### **Interface**

Se debe crear una página web, en donde se pueda visualizar la información que es calculada por el Hadoop. Note que puede poder buscar de alguna manera en la página web, por país, continente o en general.