

## GUIONES DE PRÁCTICAS – MÓDULO 3

### Práctica 3

#### Contexto

En esta práctica vamos a implementar una aplicación para explorar datos meteorológicos. Los sensores meteorológicos recopilan cada cierto tiempo datos de muchas localizaciones a lo largo de todo el mundo, obteniendo enormes volúmenes de datos semi-estructurados y orientados a registro, lo cual lo convierte en un buen candidato para análisis de datos.

Ejemplos de estos datos son los obtenidos por el *National Climatic Data Center* (NCDC). Los datos recopilados se almacenan en ficheros ASCII, en los que cada línea es un registro. El formato admite gran número de elementos meteorológicos, muchos de los cuales son opcionales o con longitud variable. En la página <ftp://ftp.ncdc.noaa.gov/pub/data/noaa/> se encuentran los datos meteorológicos para cada año.

Vamos a considerar un registro cualquiera (Ejemplo 1) y explicar algunos de los campos.

```
002900290709999991905010106004+64333+023450FM-  
12+000599999V0202301N008219999999N0000001N9-  
01391+999999102641ADDGF1029919999999999999999
```

*Ejemplo 1. Registro con los valores de los campos para una medida del sensor*

El campo **029070** es el identificador de la estación. El siguiente resaltado, **19050101**, representa la fecha en la que se recogió la observación (año/mes/día). El tercer dato resaltado es la temperatura del aire en grados Celsius multiplicada por 10, en el ejemplo **-0139** representa **-13.9** grados Celsius. El dato siguiente en negrita e itálica, **1**, indica un código de calidad de lectura.

## Ejercicio

Vamos a considerar que en el fichero *sample\_ncdc.text* hay una muestra de datos con el formato anteriormente explicado.

Considerando solo los datos que tienen **1** como código de calidad de lectura:

1. Calcular la máxima temperatura registrada para cada año.
2. Calcular la temperatura media para cada año.
3. Calcular el día, mes y año en el que se alcanza la temperatura máxima.

## Entrega

Se debe entregar un fichero pdf en el que se especifique el enunciado de cada ejercicio, la solución en spark (con algún comentario explicativo) y captura de pantalla con la ejecución y los resultados obtenidos en la Shell.