

Análisis de las temperaturas mundiales

Integrantes: Ariel Barraza
Valentina González
Maximiliano Vargas
Profesor: Aidan Hogan
Auxiliar: Alberto Moya
Ayudantes: Adriana Concha
Daniel D. Pinto
Gabriel Norambuena
Fecha de entrega: 21 de Junio de 2019

Índice de Contenidos

1. Objetivos	1
2. Datos	1
2.1. Origen	1
2.2. Países	1
2.3. Datos meteorológicos	1
2.4. Estaciones meteorológica	2
3. Métodos y herramientas	2
3.1. Filtro de datos	2
3.1.1. Mediciones de temperatura	3
4. Resultados	3
4.1. Temperatura promedio global por año desde 1900	3
4.2. Temperaturas máximas de Australia de los últimos 118 años	4
4.3. Promedio de temperaturas máximas, mínimas y promedio por año en US	5
5. Conclusiones y aprendizaje	5

Lista de Figuras

1	Archivo csv países	1
2	Archivo csv datos meteorológicos	2
3	Archivo csv estaciones meteorológicas	2
4	Gráfico que representa la temperatura promedio del mundo con respecto al tiempo .	3
5	Gráfico que representa la temperatura máxima de Australia con respecto al tiempo .	4
6	Gráfico que representa la variación de la temperatura máxima de Australia con respecto al tiempo	4
7	Gráfico que representa el promedio de temperaturas máximas, mínimas y temperatura promedio por año en EEUU.	5

Lista de Tablas

1	Extracto de mediciones filtradas	3
---	--	---

1. Objetivos

El principal objetivo del proyecto es realizar un análisis a un dataset que contiene la temperatura de la tierra en distintos lugares, con el fin de mostrar cuanto ha cambiado el clima con el paso del tiempo. Se quiere analizar en detalle el promedio de temperaturas por país.

Otro objetivo es utilizar y aprovechar las tecnologías que se enseñaron en el curso CC-5212 para realizar análisis de una gran magnitud de datos.

2. Datos

2.1. Origen

Los datos fueron recolectados por NOAA's Global Historical Climatology Network (GHCN). Estos datos fueron producidos gracias a resúmenes de clima de estaciones terrestres a través del mundo que han sido sometidos a un conjunto común de revisiones de control de calidad.

Los datos se obtuvieron a través de más de 20 fuentes, en donde se encuentran incluidos algunos datos de cada año desde 1763 hasta la fecha.

El volumen de los datos usados en este proyecto es el orden de 36 millones de filas (3 millones de filas por año, por 12 años).

2.2. Países

Se tiene un csv el cual contiene la siglas del país y el nombre del país. Los datos de una misma fila se encuentran separados por comas. En la siguiente figura se muestran algunos datos del archivo.

```
AC,Antigua and Barbuda
AE,United Arab Emirates
AF,Afghanistan
AG,Algeria
AJ,Azerbaijan
AL,Albania
AM,Armenia
AO,Angola
AQ,American Samoa [United States]
AR,Argentina
AS,Australia
AU,Austria
AY,Antarctica
BA,Bahrain
BB,Barbados
BC,Botswana
BD,Bermuda [United Kingdom]
```

Figura 1: Archivo csv países

2.3. Datos meteorológicos

Se tiene un csv el cual contiene la temperatura que registro una estación. El csv contiene los siguientes campos: (codigo-estacion, maxima-temperatura-del-dia, minima-temperatura-del-dia, temperatura-tiempo-observacion, precipitación, Nevada, profundidad de la nieve, otros-elementos).

En la siguiente figura se muestran algunos datos del archivo.

```

US1MOCW0004,20100101,PRCP,0,,N,
US1MOCW0004,20100101,SNOW,0,,N,
US1MODG0003,20100101,PRCP,0,,N,
US1MODG0003,20100101,SNOW,0,,N,
US1MODG0003,20100101,SNWD,0,,N,
US1MODG0003,20100101,WESD,0,,N,
US1MODG0003,20100101,WESF,0,,N,
US1MSHD0003,20100101,PRCP,41,,N,
US1LAEB0041,20100101,PRCP,18,,N,
US1MICX0001,20100101,PRCP,28,,N,
US1MICX0001,20100101,SNOW,46,,N,
US1MICX0001,20100101,SNWD,229,,N,
US1MICX0001,20100101,WESF,28,,N,
US1NCMK0016,20100101,PRCP,3,,N,
US1NCTR0002,20100101,PRCP,3,,N,
US1NHST0018,20100101,PRCP,15,,N,
US1NHST0018,20100101,SNOW,20,,N,

```

Figura 2: Archivo csv datos meteorológicos

2.4. Estaciones meteorológica

Se tiene un csv el cual contiene la información de la estación meteorológica en donde se hizo la medición. El csv contiene los siguientes campos: (codigo-estacion, latitud, longitud, elevacion, nombre ciudad estacion, codigo-pais, pais).

En la siguiente figura se muestran algunos datos del archivo.

```

ACW00011604,171.167,-617.833,10.1,ST JOHNS COOLIDGE FLD
ACW00011647,171.333,-617.833,19.2,ST JOHNS
AE000041196,253.330,555.170,34.0,SHARJAH INTER. AIRP GSN 41196
AEM00041194,252.550,553.640,10.4,DUBAI INTL 41194
AEM00041217,244.330,546.510,26.8,ABU DHABI INTL 41217
AEM00041218,242.620,556.090,264.9,AL AIN INTL 41218
AF000040930,353.170,690.170,3366.0,NORTH-SALANG GSN 40930
AFM00040938,342.100,622.280,977.2,HERAT 40938
AFM00040948,345.660,692.120,1791.3,KABUL INTL 40948
AFM00040990,315.000,658.500,1010.0,KANDAHAR AIRPORT 40990
AG000060390,367.167,32.500,24.0,ALGER-DAR EL BEIDA GSN 60390
AG000060590,305.667,28.667,397.0,EL-GOLEA GSN 60590
AG000060611,280.500,96.331,561.0,IN-AMENAS GSN 60611
AG000060680,228.000,54.331,1362.0,TAMANRASSET GSN 60680
AGE00135039,357.297,0.6500,50.0,ORAN-HOPITAL MILITAIRE

```

Figura 3: Archivo csv estaciones meteorológicas

3. Métodos y herramientas

Se realizará el análisis de los datos con la ayuda de Pig, la cual es una plataforma de alto nivel que permite crear programas MapReduce utilizados en Hadoop.

La metodología a seguir es realizar filter, join y group By para poder obtener la información que se necesita. Como se está trabajando con un gran volumen de datos, se debe filtrar primero las condiciones de los datos que se quieren analizar para luego realizar operaciones de map/reduce, lo que permite disminuir la cantidad de datos para realizar las otras operaciones.

3.1. Filtro de datos

3.1.1. Mediciones de temperatura

Tabla 1: Extracto de mediciones filtradas

USC00027281,20000101,TMAX,133	20000101	TMAX	133
USC00027281,20000101,TMIN,22	20000101	TMIN	22
USC00034988,20000101,TMAX,167	20000101	TMAX	167
USC00034988,20000101,TMIN,-17	20000101	TMIN	-17
USC00099466,20000101,TMAX,167	20000101	TMAX	167
USC00099466,20000101,TMIN,78	20000101	TMIN	78
USC00109601,20000101,TMAX,-22	20000101	TMAX	-22

4. Resultados

4.1. Temperatura promedio global por año desde 1900

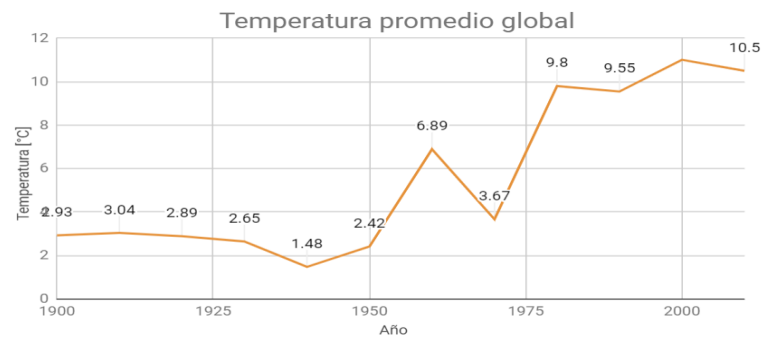


Figura 4: Gráfico que representa la temperatura promedio del mundo con respecto al tiempo

4.2. Temperaturas máximas de Australia de los últimos 118 años

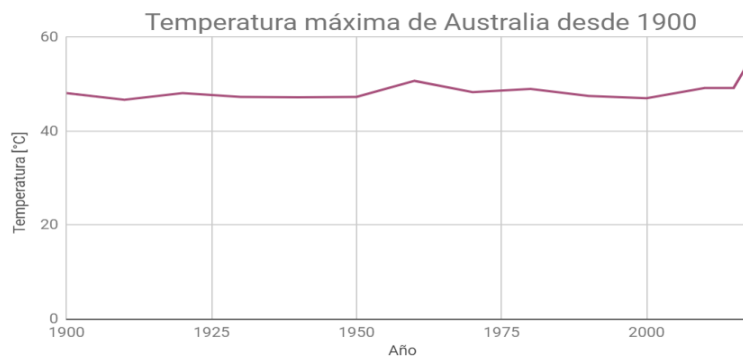


Figura 5: Gráfico que representa la temperatura máxima de Australia con respecto al tiempo

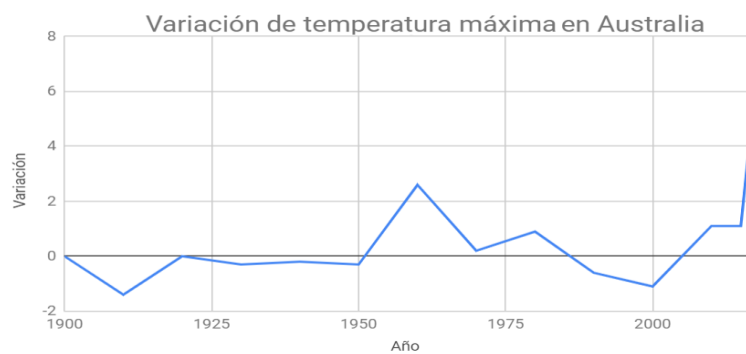


Figura 6: Gráfico que representa la variación de la temperatura máxima de Australia con respecto al tiempo

4.3. Promedio de temperaturas máximas, mínimas y promedio por año en US

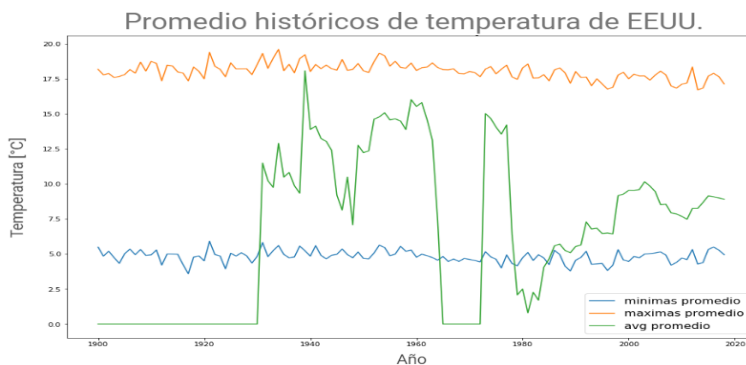


Figura 7: Gráfico que representa el promedio de temperaturas máximas, mínimas y temperatura promedio por año en EEUU.

5. Conclusiones y aprendizaje

Pig es una buena herramienta para realizar map/reduce y tiene el beneficio que se parece a la sintaxis de SQL, por lo que se percibe en forma familiar para el usuario.

Respecto de los datos, es posible concluir que existe una tendencia a un alza en la temperatura promedio de los países del mundo y en particular en Australia. Esto es apreciable en las Figuras adjuntas al reporte.

Se puede decir además que los datasets son confiables puesto que provienen de una organización gubernamental. Además los resultados obtenidos calzan con los resultados esperados (búsquedas en la web: papers sobre el tema entre otros), esto valida el dataset aún más.