**INSTITUTO TECNOLÓGICO DE TIJUANA**

**SUBDIRECCIÓN ACADÉMICA**

**DEPARTAMENTO DE SISTEMAS Y COMPUTACIÓN**

**SEMESTRE AGOSTO-DICIEMBRE 2018**

**CARRERA:**

**Ing. Informática**

**MATERIA Y SERIE**

**Datos Masivos BDD-1704 IF9A**

**TÍTULO**

**Ejercicio 5**

**UNIDAD A EVALUAR**

**Unidad 1**

**NOMBRE Y # DE CONTROL DEL ALUMNO:**

**Beltrán Bustos Claudio Israel 14212317**

**GRUPO**

**BDD-1704 IF9A**

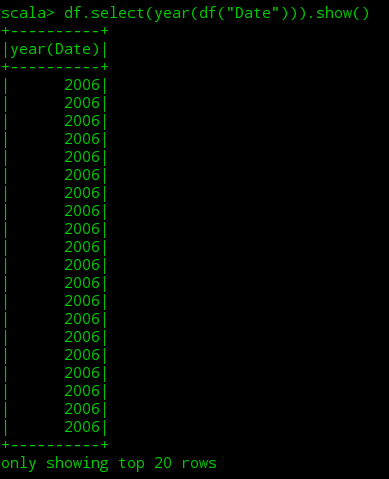
**NOMBRE DEL MAESTRO**

**José Christian Romero**

A continuación, se mostrarán screenshots junto con la descripción de 5 funciones date time para dataframes usando spark apache (Scala):

**1.- year()**

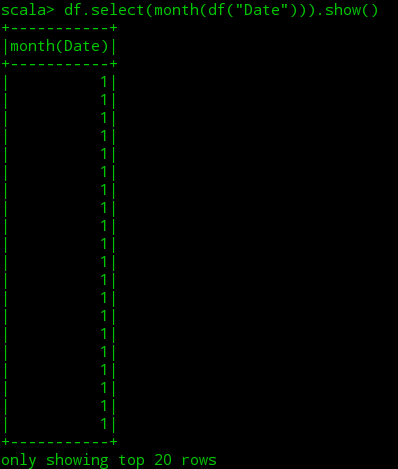
|  |
| --- |
| df.select(year(df("Date"))).show() |



Ésta función nos permite obtener el año existente en un registro de dataframe. En éste caso, se obtuvo el año de la columna ‘Date’ ya que es esa columna la que contiene un formato de tipo date time. Esto es útil para futuras operaciones ya que nos permite obtener la fecha histórica de un registro con date time. Éstas funciones de tiempo nos permiten realizar comparaciones en donde podemos guardarlas en una vista para luego acceder a ellas cuando lo deseemos.

**2.- month()**

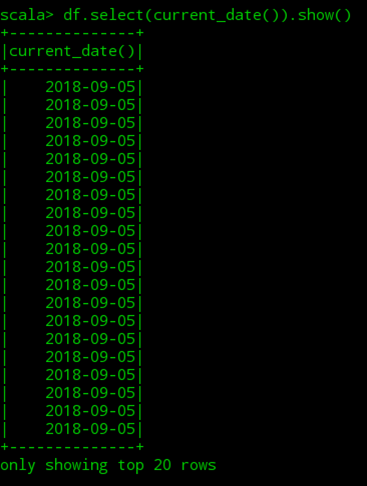
|  |
| --- |
| **df.select(month(df("Date"))).show()** |



Al igual que la función year(), la función month() nos permite extraer el mes de la columna con formato date time, en nuestro caso es la columna ‘Date’. Estas funciones permiten la extracción del tiempo en términos de año, mes y día. Esto lo podemos utilizar para futuras comparaciones de tiempo con otra tabla de la misma situación, por decir, comparando dos tablas: una tabla histórica y otra tabla con datos actuales, comparamos el tiempo de las dos y podemos realizar consultas generadas en vistas relacionadas con el tiempo de ambos.

**3.- current\_date()**

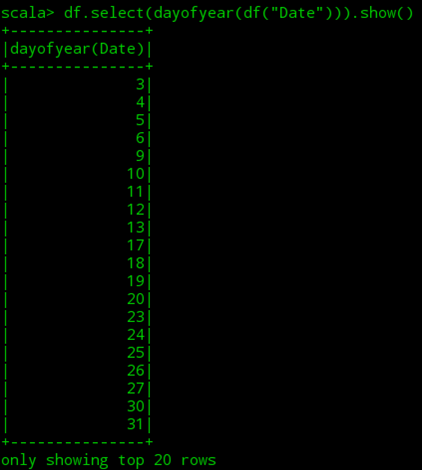
|  |
| --- |
| **df.select(current\_date()).show()** |



La función currentDate() nos regresa la fecha actual de acuerdo a la fecha de nuestra computadora. Se crea una columna temporal en donde despliega la fecha en formato año-mes-día. Como se ve en la imagen, la consulta de la fecha se hizo el día 05 del mes 09 del 2018. Esto es importante ya que al momento de hacer reportes uno puede guardar la fecha actual de cuando abrió o realizó operaciones en el dataframe.

**4.- dayofyear()**

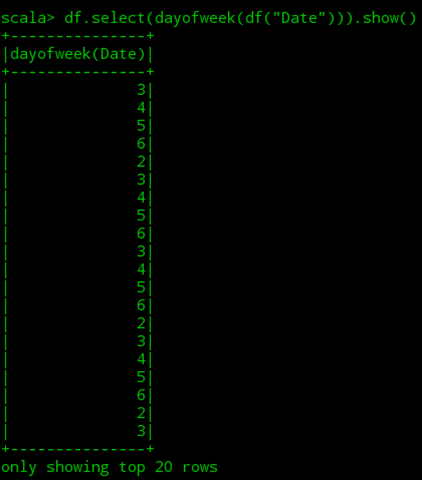
|  |
| --- |
| **df.select(dayofyear(df("Year"))).show()** |



La función dayofyear() nos regresa el día del año que contienen los registros de la columna con formato date time, en nuestro caso es ‘Date’. Al igual que las demás funciones, esto nos permite realizar operaciones de tiempo o comparación con otra tabla para analizar el tiempo y relación de los datos.

**5.- dayofweek()**

|  |
| --- |
| **df.select(dayofweek(df("Date"))).show()** |



Ésta función nos regresa el día de la semana de los registros de la columna que contenga el formato date time, en éste caso es ‘Date’. Al igual que las demás funciones, esto nos permite realizar operaciones de tiempo o comparación con otra tabla para analizar el tiempo y relación de los datos.