FORMADOR: ALONSO MELGAREJO

alonsoraulmgs@gmail.com



laboratorio 18

INSTALACIÓN DE LIBRERÍAS

**LABORATORIO 18 – INSTALACIÓN DE LIBRERÍAS**

1. En Internet, buscamos un foro en donde nos den la librería para leer archivos XML, entramos al siguiente enlace:

|  |
| --- |
| <https://community.cloudera.com/t5/Support-Questions/Error-processing-XML-data-in-Spark-Unable-to-process-XML/td-p/187436> |

Y encontramos los datos de la librería que soluciona el problema:

Texto

Descripción generada automáticamente

1. Según el foro, la librería que debemos instalar es la siguiente:

|  |  |
| --- | --- |
| **providerId** | com.databricks |
| **artifactId** | spark-xml\_**2.10** |
| **version** | 0.4.1 |

Vemos que en el artifactId se incluye un número (spark-xml\_**2.10**), este indica la versión de SCALA con el que la librería es compatible (Scala 2.10)

1. SPARK funciona sobre el lenguaje SCALA, cuando instalamos una librería debemos asegurarnos que la librería sea compatible con la versión de SCALA que tenemos en nuestro clúster. Para averiguar qué versión de SCALA tenemos, sobre DATABRICKS seleccionamos la opción “Compute”

Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

1. Damos clic sobre el nombre de nuestro clúster

Interfaz de usuario gráfica

Descripción generada automáticamente

1. En la pestaña “Configuration” podemos encontrar que en nuestro clúster tenemos la versión “**2.12**” de SCALA

Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

No podemos instalar la librería “spark-xml\_**2.10**” en nuestro clúster ya que no es compatible con la versión de SCALA de nuestro clúster (**2.12**)

1. Buscaremos si el providerId (com.databricks) ha publicado una librería “spark-xml\_**2.12**”, en el navegador buscamos:

|  |
| --- |
| com.databricks:spark-xml\_**2.12** |

1. En el primer resultado encontremos la web de MAVEN (<https://mvnrepository.com>), el cual es el repositorio más grande de librerías, en este repositorio las empresas publican librerías gratuitas para que los desarrolladores puedan descargarlas. Vemos que se nos muestra el enlace MAVEN de la librería “spark-xml\_**2.12**”, damos clic sobre él.

**ENLACE**

|  |
| --- |
| <https://mvnrepository.com/artifact/com.databricks/spark-xml_2.12/0.5.0> |

Texto

Descripción generada automáticamente

1. Vemos que la librería “spark-xml\_**2.12**” para su versión “0.5.0” fue publicada el 30 de diciembre del 2018, tal vez haya una versión más reciente, damos clic sobre el nombre de la librería.

Captura de pantalla de un celular

Descripción generada automáticamente

1. Encontramos que hay una versión más reciente, damos clic sobre la versión más reciente para ver el detalle.

Captura de pantalla de un celular

Descripción generada automáticamente

1. Podemos encontrar los detalles de instalación de la librería:

Texto

Descripción generada automáticamente

|  |  |
| --- | --- |
| **providerId** | com.databricks |
| **artifactId** | spark-xml\_**2.12** |
| **version** | 0.16.0 |

La cadena de instalación de la librería compatible con nuestro clúster es:

|  |
| --- |
| com.databricks:spark-xml\_2.12:0.16.0 |

**IMPORTANTE: Siempre debemos de adaptar la librería a la versión exacta de SCALA que tengamos en el clúster, sino la librería no funcionará.**

1. Para instalar la librería, sobre DATABRICKS seleccionamos la opción “Compute”

Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

1. Damos clic sobre el nombre de nuestro clúster

Interfaz de usuario gráfica

Descripción generada automáticamente

1. En la pestaña “Libraries” damos clic sobre “Install new”

Interfaz de usuario gráfica, Aplicación, Sitio web

Descripción generada automáticamente

1. Seleccionamos la opción “Maven” y en el campo “Coodinates” ingresamos la cadena de instalación. Damos clic en “Install”.

Captura de pantalla de un celular

Descripción generada automáticamente

1. La librería se descargará desde el repositorio MAVEN y se instalará en el clúster

Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente