关于udf：

<https://blog.csdn.net/sparkexpert/article/details/52832530>

一个变量的udf写法——比如某列全部加上“AAA”后缀：

val func:(String => String) = (a:String) => a.concat(“AAA”)

val udf1 = udf(func)

或者直接:

val udf1 = udf ( (a:String) => a.concat(“AAA”) )

然后dataframe中修改某列时，直接调用。

但是当udf要传入两个参数时，写法如下：

还是可以传上面的func，再传给udf，这里只提示udf的写法：

val udf2 = udf( (left:String, right:String) => left.concat(right) )

对于某个df中，假设已存在某列col1 ，StringType类型，可以如下：

df.withColumn(“col1 ”, udf2(col(“col1”), **lit**(“AAA”) ))

注意，这个lit（字面量）一定要加上，否则会提示报错：类型不匹配，需要Column，实际传的String.

另，如果希望填充某列：

可以使用如下语法，比如某列都赋值为“AAA”，如下：

df.withColumn(“col2 ”, **lit**(“AAA”) )

那col2这一列的值则全是“AAA“。

lit（s:Any）的作用，就是专门将s变成Column对象。

typedLit[T](literal:T)则是专门创建java中的类型

expr，提供表达式：如下：

变量名前面不需要美元符号。

