**Pyspark基本语法（同盾DC）—展现sql内容**

————20190812：韵方

这章我们来说说使用pyspark如何从sql拿到数据并展示出来，其默认拿到的数据结构又是怎么样的。

直接上代码吧，随便从一个数据库中拿到一列名为task\_id的数据，并将它用变量df命名记录下来，随后我们直接将df打印出来：



来看看结果：



于是我们就知道了，这样直接通过连接得到的sql的数据结构是一个名为DataFrame（这里要明确一下，这样获取到的DataFrame是一个只读的数据结构，具体原因目前我还没有研究，以后明白了在spark的原理中与大家进行解释）的东西，那么，它又是怎么构成的呢。我们都知道数据库的数据肯定是以行为单位的，所以我们将其保存成一个数组，再将数组打开，一行一行的输出：



最后我们得到了这么一个东西（每行）：



于是我们可以知道，在DataFrame中每行都是一个Row类，这就是通过pyspark以这种方式获取到数据库值时拿到的数据结构。那么我们如何展示它呢，这里我再来写两种展示的方法（第一种在上面已经写过了就是把它保存成list逐行输出那个）：

通过collect（）方法输出：



输出结果如下：



这样内容是全部输出来了，但是都挤在了一行，很难看，所以我们通过show（）方法输出：



后面20参数的意思是只输出20行，具体结果如下：



是不是好看很多呢。