

spark 与 hbase 的快捷交互

作者: [niko](#)

原文链接: <https://ld246.com/article/1505228368456>

来源网站: [链滴](#)

许可协议: [署名-相同方式共享 4.0 国际 \(CC BY-SA 4.0\)](#)

##以往在spark中读写hbase，往往需要进行相当多的操作，这里提供一个快捷的操作方式。

测试环境为本地模式，依赖为spark1.6.x, hbase0.98, 以及spark-hbase-connector。这里需要说的是，这种读写的方式没有提供原生api中的各种过滤接口，因此，适合进行归档的应用场景。如果要进行复杂过滤条件下的读操作，最好使用原生api。

项目的git地址 <https://github.com/nerdammer/spark-hbase-connector?spm=5176.doc28131.2.cycZVO>

##maven依赖如下：

```
<dependency>
<groupId>it.nerdammer.bigdata</groupId>
<artifactId>spark-hbase-connector_2.10</artifactId>
<version>1.0.3</version>
</dependency>
```

##废话不多说，直接上代码：

```
import org.apache.spark.{SparkConf, SparkContext}
import it.nerdammer.spark.hbase._

object FastSparkWithHbase {
  def main(args: Array[String]): Unit = {
    val sparkConf = new SparkConf().setAppName("hbase").set("spark.hbase.host", "192.168.93.11").setMaster("local[*]")
    val sc = new SparkContext(sparkConf)

    val rdd = sc.parallelize(0 to 20)
      .map(i => (i.toString, i + 1, "see" + util.Random.nextInt(100)))

    //写数据,需要保证表结构已存在
    rdd.toHBaseTable("spark_t").toColumns("col1", "col2").inColumnFamily("colfamily1").save()

    //读数据,不支持复杂过滤器,可以考虑转换成rdd (不建议)
    val hBaseRDD: HBaseReaderBuilder[(String, Int, String)] = sc.hbaseTable[(String, Int, String)]("spark_t")
      .select("id+1", "stringValue").inColumnFamily("mycf")

    hBaseRDD.foreach(println(_))

    sc.stop()
  }
}
```