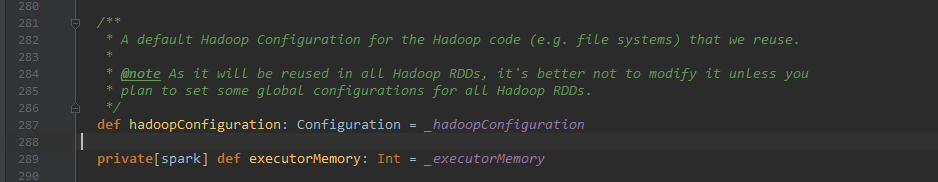
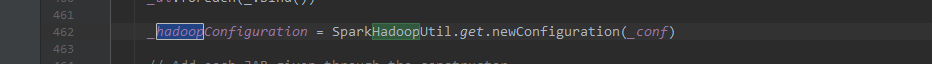
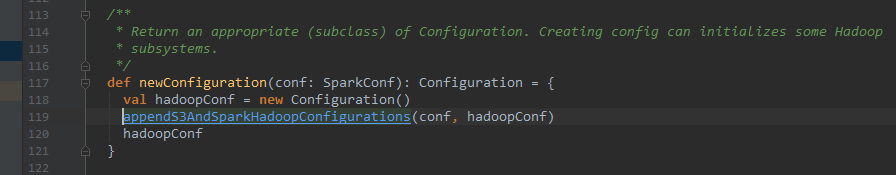
Hadoop相关配置

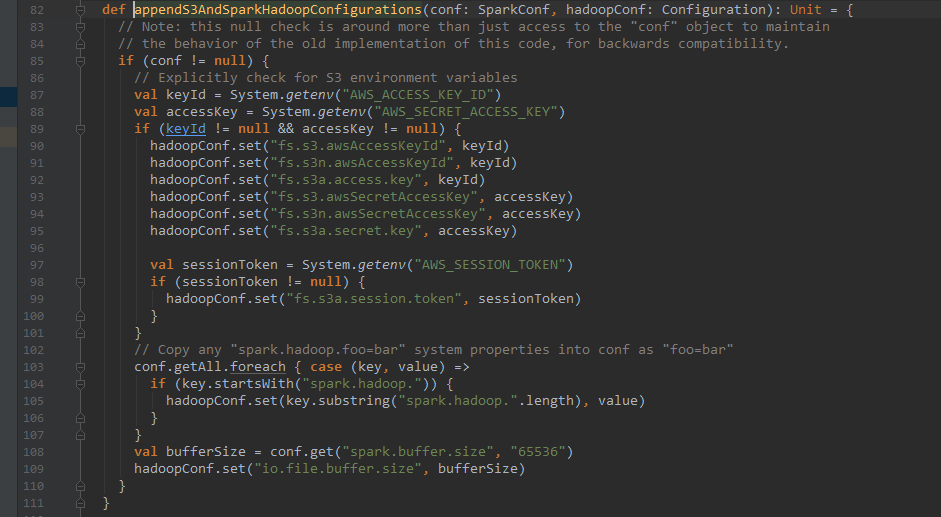
默认情况下，Spark使用HDFS作为分布式文件系统，所以需要Hadoop的配置信息。相关代码如下：



或者：







从以上图片代码可知获取的信息包括：

1. 将Amazon S3文件系统的awsAccessKeyId和awsSecretAccessKey加载到Configuration；
2. 将spark.hadoop.开头的属性复制Hadoop的Configuration；
3. 将spark.buffer.size设置为io.file.buffer.size