**如何自定义HDInsight Storm 日志文件大小**

HDInsight Storm集群采用Log4j对topology日志进行收集并存储在Azure Blob Storage，默认情况下，只会保存一定时间内的日志，history的日志会被冲刷。但是用户可以通过修改相关参数对日志文件大小和保存时间进行自定义修改。以下为相关步骤：

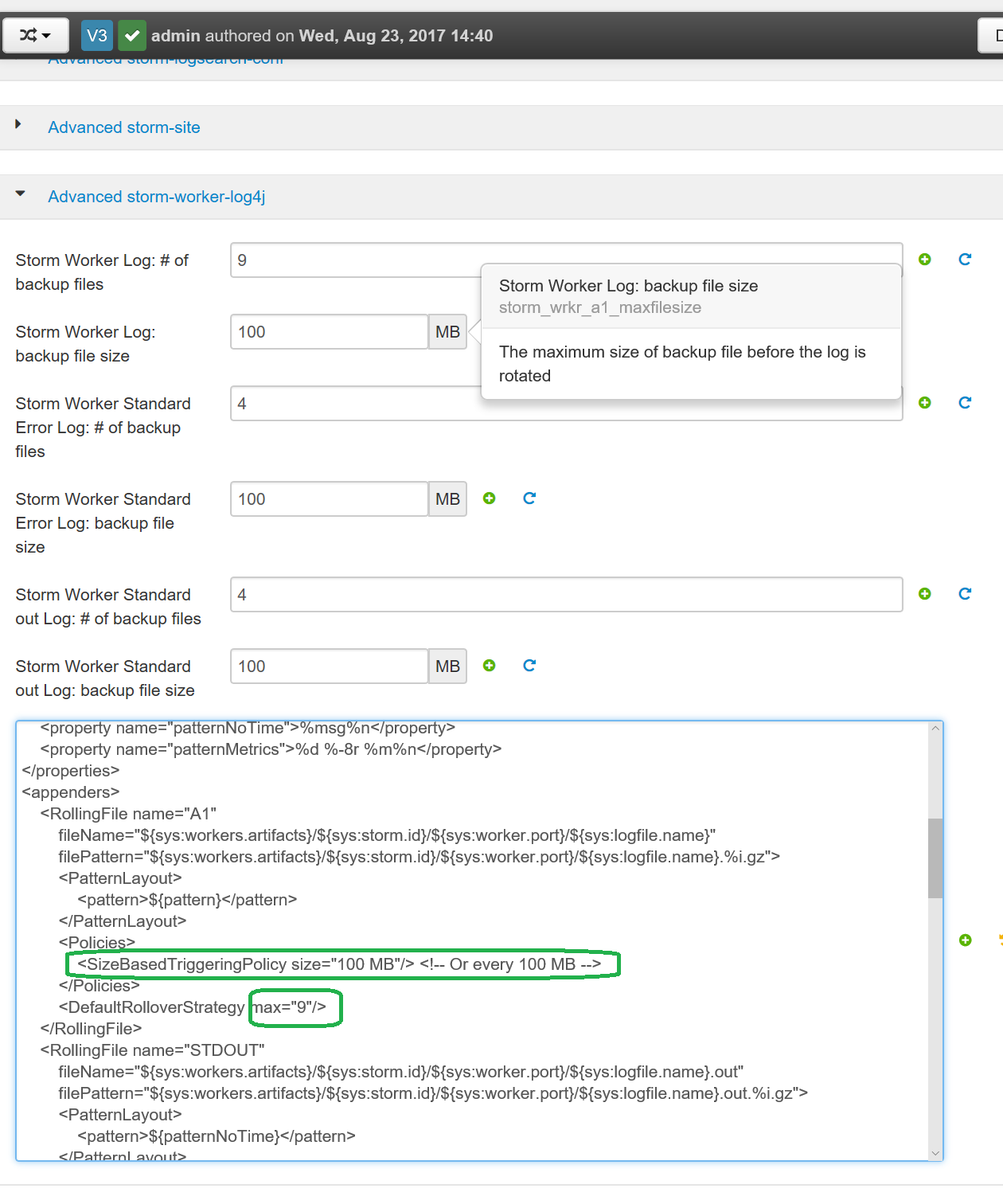
**先决条件：**

HDInsight Storm 集群

登录集群的ssh工具，例如：Putty或者Xshell.

**步骤：**

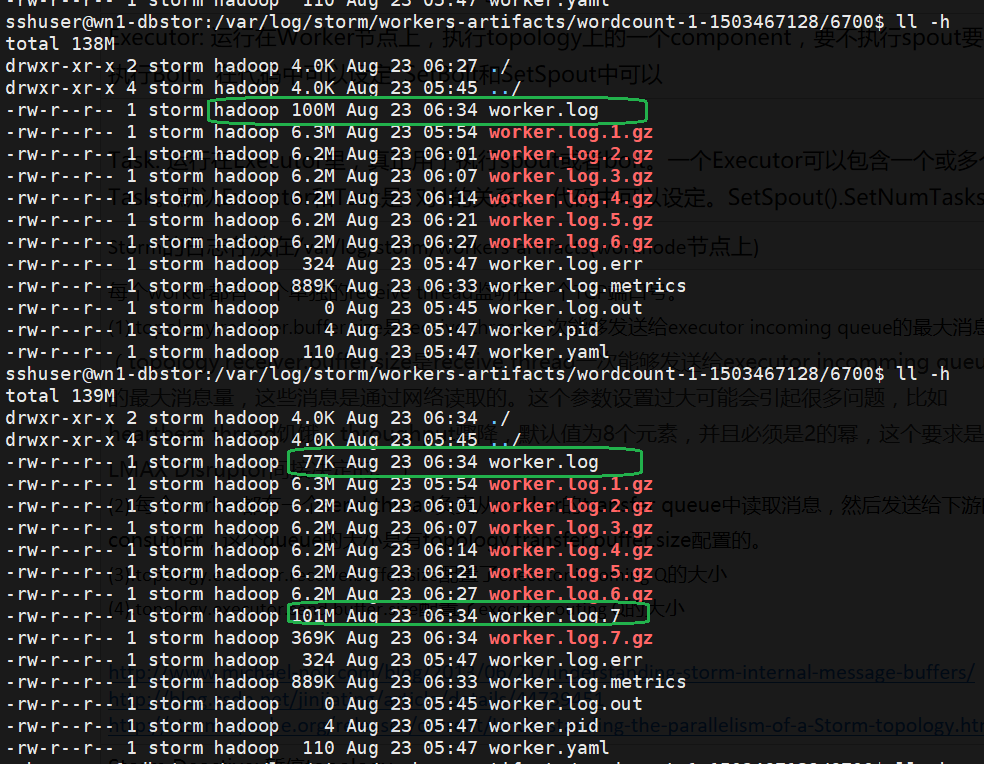
1. 创建完集群后，登录ambari界面，可以查看到集群Log4j保存日志的默认参数。



参数SizeBasedTriggeringPolicy size=”100 MB”作用是如果日志文件大小达到100MB，那么此日志文件会被进行压缩，生成后缀为.gz文件。

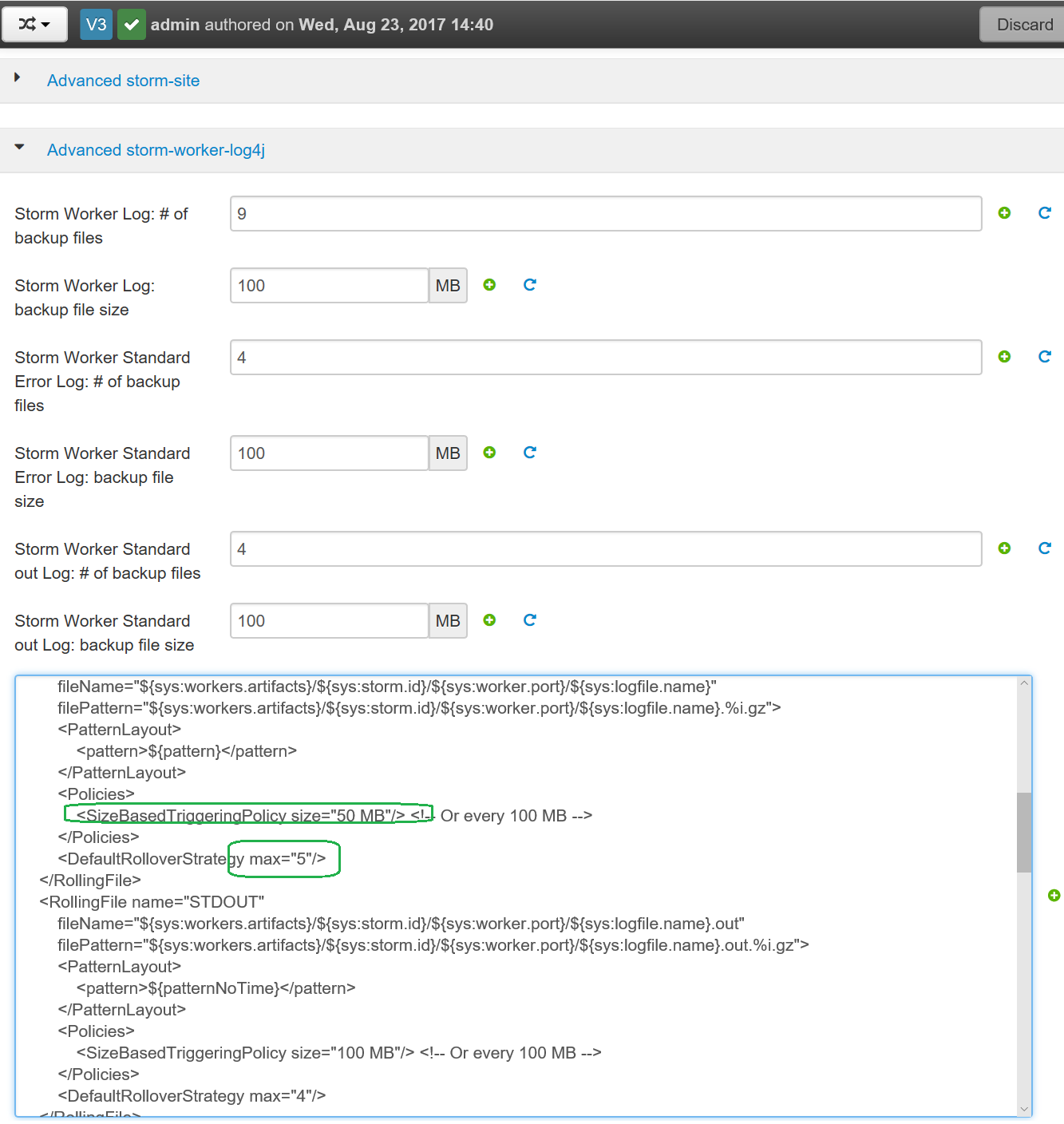
参数DefaultRolloverStrategy max="9" 作用是默认情况下，此类gz文件保留9份。注：最老版本的gz文件会被冲刷。

1. 向storm集群提交topology，在worker节点路径/var/log/storm/workers-artifacts下查看日志文件。

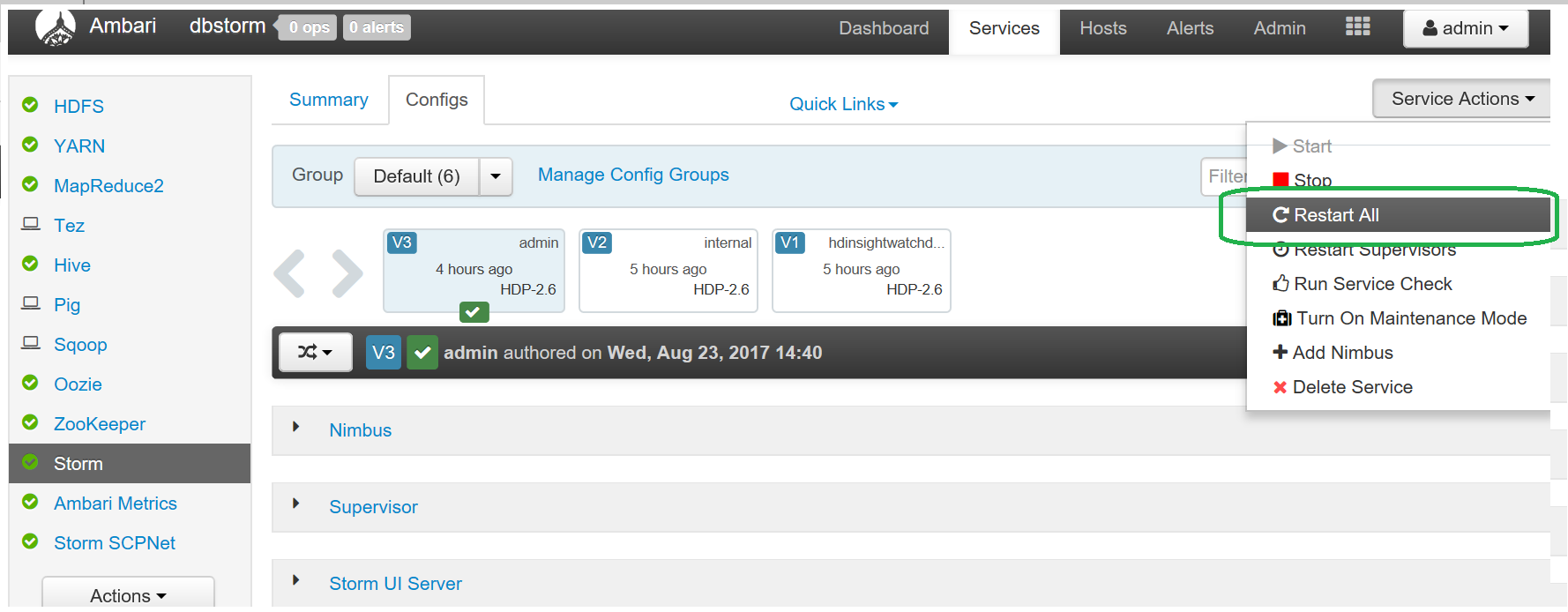


从此图中我们可以看到worker.log文件size达到100M时，马上被重命名为worker.log.7，并且随即进行了压缩。注：gz文件会保留9份。

1. 如果客户想要对此日志文件保留策略进行自定义的话，可以通过修改第一步中说明的两个参数。

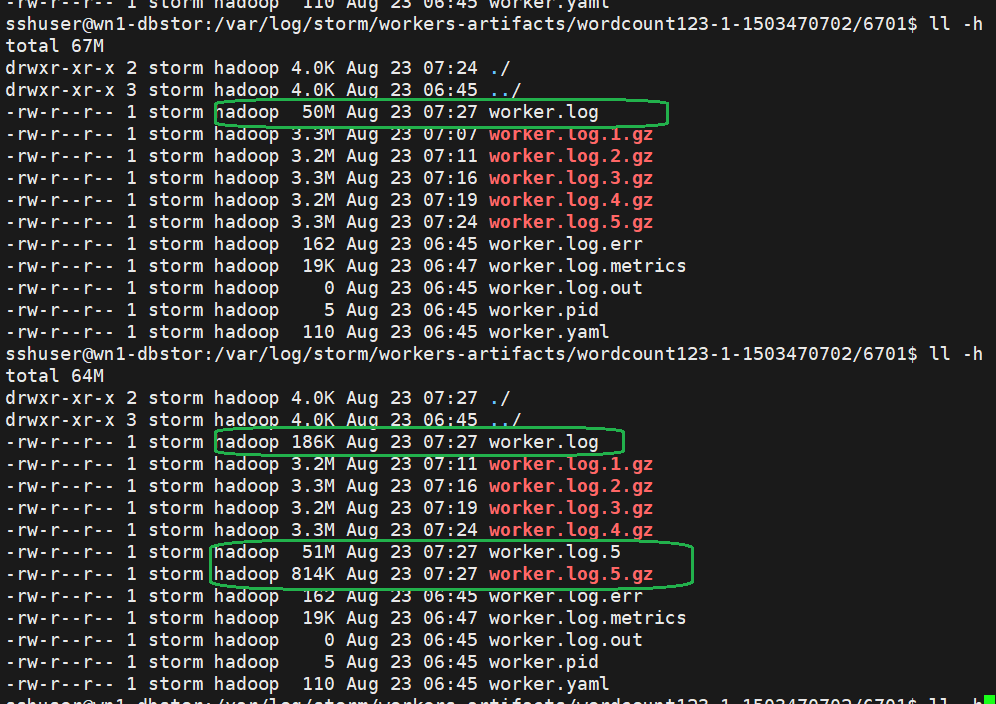


点击右上方Save按钮，然后选择一个合适的时间点restart storm相关服务。操作如下：



通过以上步骤，将此修改deploy到了集群的每个节点，包括headnode和worknode。可以在/usr/hdp/XXXX/storm/log4j2/worker.xml查看到更新。

1. 观察新topology的日志文件变化。文件增长到50M的时候，随即被压缩。并且，以gz结尾的文件值保存了5个。



总结：通过以上步骤可以对HDInsight Storm 日志文件进行自定义，以上演示是将单个日志文件保存大小从100M改为50M，且总数从10个减少为5个。当然，完全可以增大单个日志文件大小和文件总数，只要增大第一步中提到了2个参数值的大小即可。