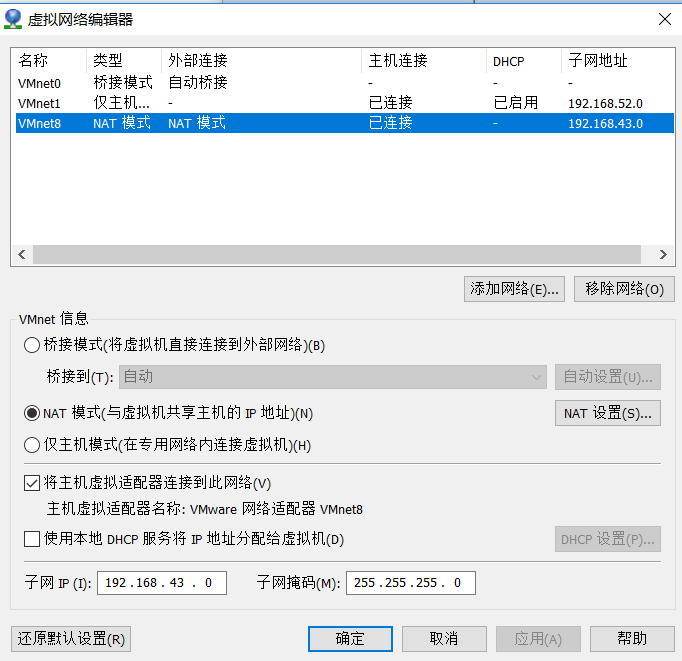
操作手册

1. 虚拟机配置简介

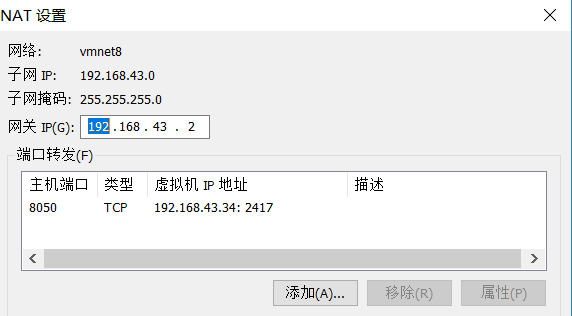
虚拟机版本VMwareWorkstation12系统CentOS7，拥有两个普通用户cjl和ELK，所有用户登陆密码均为cjl970627。在起始命令行页面使用root用户登录后，执行startx开启图形界面。虚拟机firewalld已停用，改换成iptables，且开机自启动，防火墙已经开放环境所需要端口。selinux已关闭。java版本1.8.0\_152。/etc/hosts已配置为master。但是SSH无法配置，尚未解决。深度学习平台搭建虚拟机里调用的是CPU资源。

1. 配置网络

虚拟机网络为NET模式，虚拟机IP为192.168.43.34，需要将主机的VMnet8虚拟网卡和虚拟机保持在同一43网段，取消使用本地DHCP服务分配IP：



网关设置为192.168.43.2，将在虚拟机中搭建的网站端口映射到物理机的8050端口，即可使用其他虚拟机访问。建议在其他搭建的虚拟机上攻击网站，后续的拦截工作会将攻击IP封锁。



1. 搭建网站

**cd /root/Desktop/apache-tomcat-7.0.82/bin**执行**./startup.sh**

在浏览器里输入192.168.43.34:7070/shop即可访问。

1. 反向代理

**cd /root/nginx/sbin**，执行**./nginx**开启反向代理，由于整合在一台机，所以仅端口号发生变化，在浏览器里输入192.168.43.34:2417/shop即可访问网站。网站日志生成在/home/ELK/ELK/nginxlog/predictdata/access.log。由于后续每日零时会定时将access.log改名并上传到HDFS，所以，定时在每次传输完毕后重新启动nginx，重新产生access.log。

1. 搭建HDFS

**cd /home/ELK/ELK/hadoop/sbin**，执行**./start-dfs.sh**启动HDFS。由于不知什么原因导致SSH免密始终无法配上，所以只能在启动HDFS时不断输入密码cjl970627。启动完成后，使用jps查看相关进程，有NameNode和DataNode说明启动成功。

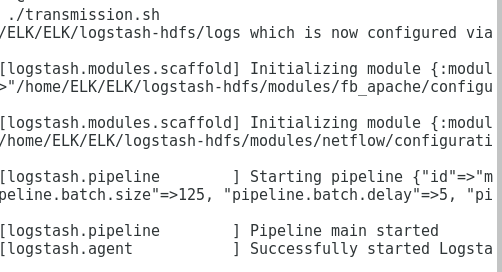
Sbin目录下有脚本transmission.sh，执行即可打通与HDFS的传输通道，并且触发执行cut\_nginx\_log.sh，将/home/ELK/ELK/nginxlog/ predictdata的access.log转化成成一份以昨日日期命名的日志，并上传HDFS。

1. 生成数据的分类情况、处理

训练数据存储在/home/ELK/ELK/nginxlog/traindata下，5种类型（正常general、慢链接slowlinks、DDOSddos、撞库hitlibrary、爆破blasting）的日志各一份，还有一份是在执行打标签脚本./label.sh后将5种类型日志整合到一起作为训练数据的train\_data.txt。

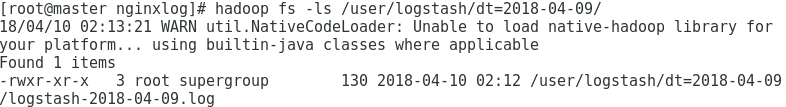
预测数据存储在/home/ELK/ELK/nginxlog/ predictdata下，有一份我们之前用来预测的数据predict\_data.txt，共5000条。当前访问（攻击）网站的数据会存储在access.log。

**cd /home/ELK/ELK/hadoop/sbin**下先执行（在生产实际中会每日零时定期执行）**./transmission.sh**。打开和HDFS的传输通道。



待屏幕显示Pipeline main started

**新终端执行/home/ELK/ELK/hadoop/sbin/cut\_nginx\_log.sh**定期切割日志脚本，会生成一份以昨日日期命名的日志。（目的是在生产实践中在每天零时将access日志切割，这一天的日志将用于预测，进入HDFS保存3天）该脚本执行后，会触发Logstash传输文件到HDFS。



查看一下传输成功后，**新终端执行/home/ELK/ELK/hadoop/sbin/stop.sh**关闭logstash（生产环境中会定时在前两个脚本执行完10分钟后执行）。

1. 搭建ELK用于日志分析

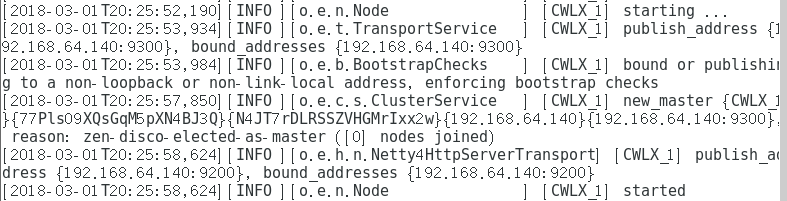
在**新终端root用户**下输入指令**sysctl -w vm.max\_map\_count=262144**

检查：sysctl -a|grep vm.max\_map\_count

会显示：vm.max\_map\_count = 262144

输入指令**su - ELK**转为ELK用户，**cd /home/ELK/ELK/elasticsearch/bin**执行

**./elasticsearch &**来启动Elasticsearch，稍等1分钟看到started即加载完毕。终端不要关闭。



若要检索训练日志，**再开一个终端，使用root用户**

**cd /home/ELK/ELK/logstash-traindata/bin**，

执行**./logstash -f traindata.conf**，用于传输训练数据到Elasticsearch检索。

若要检索预测日志，再开一个终端，使用root用户

**cd /home/ELK/ELK/logstash-predictdata/bin**

执行**./logstash -f predictdata.conf**，用于传输预测数据到Elasticsearch检索。

**再开一个终端，cd /home/ELK/ELK/kibana/bin**执行**./kibana**，

打开浏览器进入192.168.43.34:5601进入Kibana，可视化训练、预测数据。终端不要关闭。

在Kibana页面左侧的**Management，点击Index Patterns，create index**。**新建训练数据索引名train\*，预测数据索引名prediction\*。**训练数据索引train可以在Visualize里多维度可视化，预测数据索引prediction可在Discover里看到数据流量。

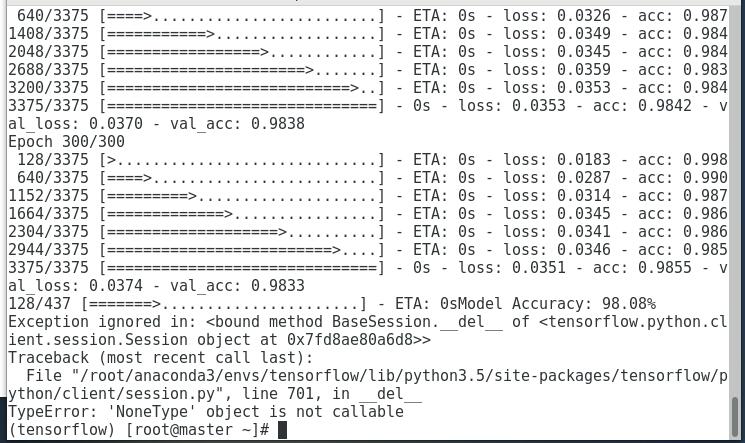
1. 训练神经网络

在**新终端执行指令source activate tensorflow**，进入到tensorflow环境中。

**cd /home/ELK/ELK/nginxlog/traindata**

执行**python3 LSTM.py**，训练神经网络模型。

有时候可能会出现一个NoneType报错，如下，重新执行**python3 LSTM.py**即可正常运行。



运行完后会在/home/ELK/ELK/nginxlog/traindata下生成三份文件securitai-lstm-model.h5,securitai-lstm-model.json,securitai-weights.h5，还有一个路径logs，用于Tensorboard作图。

1. 神经网络可视化

**在刚才的训练模型终端下执行**

**tensorboard --logdir = ‘/home/ELK/ELK/nginxlog/traindata/logs’**

跳出：TensorBoard 0.4.0rc2 at http://localhost.localdomain:6006 (Press CTRL+C to quit)

在浏览器中输192.168.43.34:6006端口即可进入Tensorboard窗口。

1. 离线预测

**在tensorflow环境中cd /home/ELK/ELK/nginxlog/predictdata**，

执行**python3 predict.py**进行预测。

当前预测代码中的输入数据是当日产生的access.log，若要改变输入数据要在代码的首尾两处（有注释）将access.log换成目标数据即可，注意代码行缩进，目标数据要存在于当前predictdata/路径下。predictdata/路径下还有一份之前准备好的预测数据predict\_data.txt，共5000条。

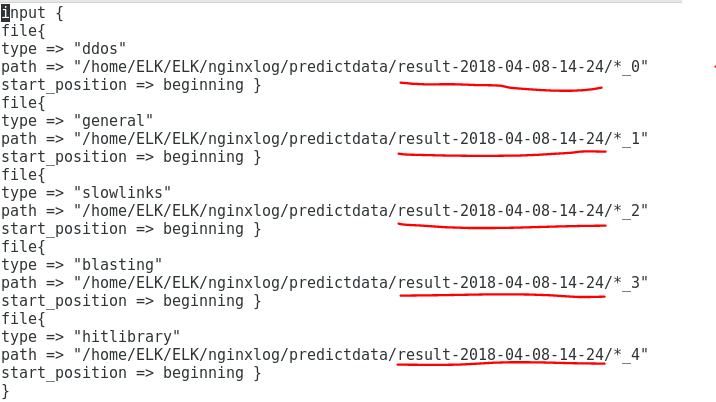
预测过程中会出现matplotlib的图像，要**关闭图像后代码才可继续运行**。运行结束后会生成一个result-timestamp命名的目录，目录下有预测的识别结果日志。

1. 实时预测

**执行/home/ELK/ELK/nginxlog/predictdata/Heartbeat.sh**，该脚本每分钟执行一次预测，预测对象为每分钟access.log的增量数据，该脚本中**嵌套执行attackresult.sh脚本和classification.sh脚本，**最终将预测结果中的被识别出的恶意攻击ip和结果日志分类整合到指定路径下。由Logstash收集通过Kibana展示。

1. 离线预测结果可视化

在**新终端下cd /home/ELK/ELK/logstash-result/bin**，在bin下**vim result.conf**



更改需要可视化的结果文件的时间。然后在bin下执行

**./logstash -f result.conf**启动Logstash。由于Logstash启动晚于识别结果生成，先需要返回**cd /home/ELK/ELK/nginxlog/predictdata/result-预测时间/**，将预测文件移出当前路径，然后再重新放进当前路径。

**mv 2018-04\* ../** #将文件移出当前路径到上一级目录

然后到上一级目录下执行

**mv 2018-04\* predictdata/** #文件重新移入到结果路径

Logstash将识别结果数据传输到Kibana后，在Kibana页面左侧的**Management，点击Index Patterns，create index，新建result\*索引**。在左侧的Visualize下可以多维度可视化预测结果（根据需要可视化展示的字段可以选择饼图、表格、Top10柱状图、云图等）。

1. 实时预测结果可视化

在**新终端下cd /home/ELK/ELK/logstash-result/bin**，

然后在bin下执行**./logstash -f result.conf**启动Logstash。

此时Logstash会将分类汇总得到的不同类型的预测结果日志文件从末尾开始读取数据。数据以不同的type汇总到一个索引之中。

Logstash将识别结果数据传输到Kibana后，在Kibana页面左侧的**Management，点击Index Patterns，create index，新建result\*索引**。在左侧的Visualize下可以多维度可视化预测结果（根据需要可视化展示的字段可以选择饼图、表格、Top10柱状图、云图等）。

1. 离线拦截

执行**ipset create blacklist hash:ip maxelem 100000**

#创建一个名称为 blacklist 的IP集合，最大容量为10000

执行**incrond start**  #开始执行触发。

**cd /ELK/ELK**，执行**./intercept.sh**开启拦截。拦截效果是封杀IP，并且会发送邮件给管理员（当前邮箱设置为我本人的邮箱，若要更改则提前

**vim /root/Desktop/warning.sh**，在最后一个收件人邮箱参数处更改为目的收件人邮箱即可）

1. 实时拦截

执行**ipset create blacklist hash:ip maxelem 100000**

#创建一个名称为 blacklist 的IP集合，最大容量为10000

执行**incrond start** #开始执行触发。

**Heartbeat.sh脚本中嵌套执行的脚本attackresult.sh会嵌套执行intercept.sh（条件是再次生成了恶意攻击IP文件attackip.log）**，即每分钟内但凡有攻击日志，就会执行拦截。

设置触发条件如下，执行**incrontab -e**

IN\_CLOSE\_WRITE表示只有当blacklist.log文件被写入后关闭时才会触发warning.sh脚本发送邮件。达到实时发送邮件的效果。



1. 监控

在新终端执行**systemctl start httpd && systemctl start zabbix-server**

再执行**systemctl start zabbix-agent**,开启监控平台。

在浏览器输入192.168.43.34/zabbix即可进入监控平台。

**点击“检测中”选项，点击“聚合图形”点击已经创建好的shop**，即可进行监控服务器性能。。