Hadoop `sysctl.conf` parameters.

Performace tuning Hadoop at kernel level. sysctl is an interface that allows you to make changes to a running Linux kernel. With /etc/sysctl.conf you can configure various Linux networking and system settings such as:

- Limit network-transmitted configuration for IPv4
- Limit network-transmitted configuration for IPv6
- Turn on execshield protection
- Prevent against the common `syn flood attack`
- Turn on source IP address verification
- Prevents a cracker from using a spoofing attack against the IP address of the server.
- Logs several types of suspicious packets, such as spoofed packets, source-routed packets, and redirects.

File System > 1. `fs.file-max` Increase size of file handles and inode cache

```
[ahmed@server ~]# echo 'fs.file-max = 943718' >> /etc/sysctl.conf
```

Swappiness : Do less swapping

1. `vm.dirty_ratio` setting virtual memory ratio.
2. `vm.swappiness` How often swap should be used. 0 is least, 60 default.

```
[ahmed@server ~]# echo 'vm.dirty_ratio=10' >> /etc/sysctl.conf
[ahmed@server ~]# echo 'vm.swappiness=0' >> /etc/sysctl.conf
```

Connection Settings

1. `net.core.netdev_max_backlog` Increase number of incoming connections backlog.
2. `net.core.somaxconn` Increase number of incoming connections.

```
[ahmed@server ~]# echo 'net.core.netdev_max_backlog = 4000' >> /etc/sysctl.conf
[ahmed@server ~]# echo 'net.core.somaxconn = 4000' >> /etc/sysctl.conf
```

TCP settings

1. `net.ipv4.tcp_sack` Disable select acknowledgments
2. `net.ipv4.tcp_dsack` Allows TCP to send “duplicate” SACKs.
3. `net.ipv4.tcp_keepalive_time` How often TCP sends out keepalive messages when keepalive is enabled. Default: 2hours.
4. `net.ipv4.tcp_keepalive_probes` How many keepalive probes TCP sends out, until it decides that the connection is broken. Default value: 9.
5. `net.ipv4.tcp_keepalive_intvl` How frequently the probes are send out. Multiplied by `tcp_keepalive_probes` it is time to kill not responding connection, after probes started. Default value: 75sec i.e. connection will be aborted after ~11 minutes of retries.

6. `net.ipv4.tcp_fin_timeout` Time to hold socket in state FIN-WAIT-2, if it was closed by our side. Peer can be broken and never close its side, or even died unexpectedly. Default value is 60sec. Usual value used in 2.2 was 180 seconds, you may restore it, but remember that if your machine is even underloaded WEB server, you risk to overflow memory with kilotons of dead sockets, FIN-WAIT-2 sockets are less dangerous than FIN-WAIT-1, because they eat maximum 1.5K of memory, but they tend to live longer. `Cf.tcp_max_orphans`.
7. `net.ipv4.tcp_rmem` The three values setting the minimum, initial, and maximum size of the Memory Receive Buffer per connection. They define the actual memory usage, not just TCP window size.
8. `net.ipv4.tcp_wmem` The same as tcp_rmem, but just for Memory Send Buffer per connection.
9. `net.ipv4.tcp_retries2` This value influences the timeout of an alive TCP connection, when RTO retransmissions remain unacknowledged. Given a value of N, a hypothetical TCP connection following exponential backoff with an initial RTO of `TCP_RTO_MIN` would retransmit N times before killing the connection at the (N+1)th RTO. The default value of 15 yields a hypothetical timeout of 924.6 seconds and is a lower bound for the effective timeout. TCP will effectively time out at the first RTO which exceeds the hypothetical timeout. RFC 1122 recommends at least 100 seconds for the timeout, which corresponds to a value of at least 8.
10. `net.ipv4.tcp_synack_retries` Number of times SYNACKs for a passive TCP connection attempt will be retransmitted. Should not be higher than 255. Default value is 5, which corresponds to ~180seconds.

```
[ahmed@server ~]# echo 'net.ipv4.tcp_sack = 0' >> /etc/sysctl.conf
[ahmed@server ~]# echo 'net.ipv4.tcp_dsack = 0' >> /etc/sysctl.conf
[ahmed@server ~]# echo 'net.ipv4.tcp_keepalive_time = 600' >> /etc/sysctl.conf
[ahmed@server ~]# echo 'net.ipv4.tcp_keepalive_probes = 5' >> /etc/sysctl.conf
[ahmed@server ~]# echo 'net.ipv4.tcp_keepalive_intvl = 15' >> /etc/sysctl.conf
[ahmed@server ~]# echo 'net.ipv4.tcp_fin_timeout = 30' >> /etc/sysctl.conf
[ahmed@server ~]# echo 'net.ipv4.tcp_rmem = 32768 436600 4194304' >> /etc/sysctl.conf
[ahmed@server ~]# echo 'net.ipv4.tcp_wmem = 32768 436600 4194304' >> /etc/sysctl.conf
[ahmed@server ~]# echo 'net.ipv4.tcp_retries2 = 10' >> /etc/sysctl.conf
[ahmed@server ~]# echo 'net.ipv4.tcp_synack_retries = 3' >> /etc/sysctl.conf
```

Disable IPv6 Defaults. We dont use these anyway.

```
[ahmed@server ~]# echo 'net.ipv6.conf.all.disable_ipv6 = 1' >> /etc/sysctl.conf
[ahmed@server ~]# echo 'net.ipv6.conf.default.disable_ipv6 = 1' >> /etc/sysctl.conf
[ahmed@server ~]# echo 'net.ipv6.conf.lo.disable_ipv6 = 1' >> /etc/sysctl.conf
```

Execute below command to make it permanent.

```
[ahmed@server ~]# sysctl -p
```

Next update limits.

```
[ahmed@server ~]# echo '* - nofile 65536' >>/etc/security/limits.conf
[ahmed@server ~]# echo '* - nproc 65536' >>/etc/security/limits.conf
```

More Details on IPv4

```
http://www.cyberciti.biz/files/linux-kernel/Documentation/networking/ip-sysctl.txt
http://www.cyberciti.biz/files/linux-kernel/Documentation/sysctl/
```