Install CDH5 with Cloudera Manager

Document Author: arho.virkki@vtt.fi

This procedure is tested with Ubuntu 14.04. For details, see: http://www.cloudera.com/content/cloudera/en/downloads/cloudera_manager/cm-5-4-6.html

Pre-requisites: multiple, Internet-connected Linux machines, with SSH access, and significant free space in /var and /opt.

```
wget http://archive.cloudera.com/cm5/installer/latest/cloudera-manager-installer.bin
chmod u+x cloudera-manager-installer.bin
sudo ./cloudera-manager-installer.bin
```

For CDH cluster installation to Ubuntu 14.04 server, enable password-less sudo privileges

```
sudo visudo

# No password
%sudo ALL=NOPASSWD: ALL
```

CHD Installer is really pedantic about fully qualified domain names (FQDN). Check that the hosts themselves report their correct hostname with

```
hostname -f
```

And not e.g. 'localhost'. Static names can be configured under */etc/hosts* (for each machine inside the cluster), for example

```
# IP      FQDN     shortname
127.0.0.1       localhost
192.168.1.221   cdh cdh
```

Finally, open Cloudera manager at

```
http://192.168.1.221:7180/cmf/
```

If all went well, you can find the web services (most interestingly, Hue) at the following addresses:

- Hue: http://192.168.1.221:8888
- Application summary http://192.168.1.221:8088
- DFS NameNode http://192.168.1.221:50070

All services and their ports are listed in the Cloudera documentation: http://www.cloudera.com/content/cloudera/en/documentation/core/latest/topics/cm_ig_ports_cm.html

Configuration

Enable Beeline/Hue/Hive to write to their default temporal location in HDFS:

```
sudo -u hdfs hadoop fs -mkdir /user/anonymous
sudo -u hdfs hadoop fs -chown anonymous /user/anonymous
```