



EKONOMICKÁ UNIVERZITA  
Fakulta hospodárskej informatiky  
Katedra aplikovanej informatiky



## INŠTALÁCIA HADOOP 3.3.3 NA LINUX UBUNTU 20.04

**Meno študentov:** Martin Jankech, Patrik Hajdučík

**Stupeň/ročník:** druhý/2.ročník

**Školský rok:** 2022/2023

**Meno prednášajúceho:** Ing. Jaroslav Kultán PhD.

**Meno cvičiaceho:** SCHMIDT, Peter, Ing. Mgr., PhD.

## Obsah

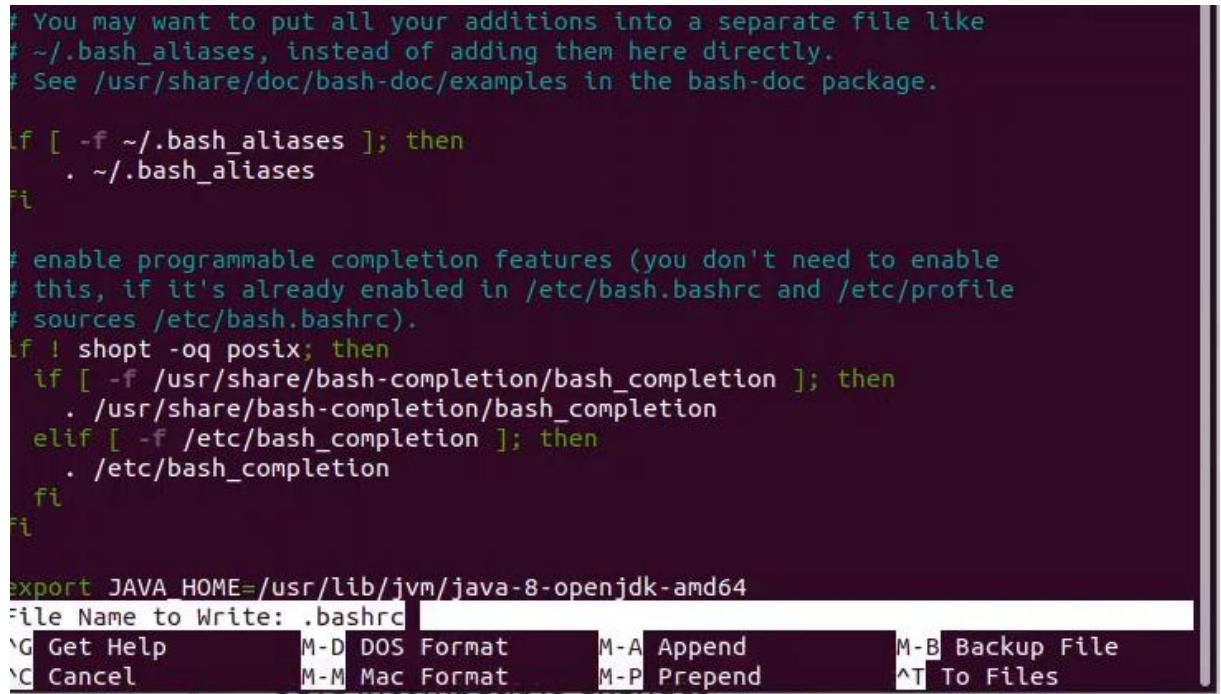
Úvod.....	3
Postup inštalácie.....	4
Záver.....	9
Bibliografia.....	10

## Úvod

V tomto zadání si ukážeme a popíšeme postup akým sa inštaluje Hadoop 3.3.3 na operačný systém Linux a jeho distribúciu Ubuntu 20.04.

## Postup inštalácie

1. Prvým krokom je samozrejme inštalácia Linuxu pokiaľ ho nemáme. My sme si ho spustili na virtuálnom stroji pomocou nástroja Oracle VM Virtualbox. Verzia na ktorej tu budeme ukazovať je ubuntu 20.04.
2. Pokiaľ nemáme nainštalovanú javu tak prvým krokom je jej inštalácia. Príkaz na nainštalovanie javy 1.8 je `- sudo apt install openjdk-8-jdk -y`
3. To či bola Java nainštalovaná správne si môžeme skontrolovať príkazom `java -version`
4. ďalej potrebujeme nastaviť systémové premennej
5. pomocou príkazu `which java` zistíme relatívnu cestu k jave
6. my však potrebujeme plnú cestu a tak použijeme príkaz `readlink -f /usr/bin/javac`
7. pomocou príkazu `sudo nano .bashrc` otvoríme súbor kde nastavíme systémovú premennú
8. na koniec pridáme `export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64` premennú `JAVA_HOME` podobne ako na obrázku. Súbor zavrieme `ctrl+x` a dáme `y` aby sa nám uložil.



```
# You may want to put all your additions into a separate file like
# ~/.bash_aliases, instead of adding them here directly.
# See /usr/share/doc/bash-doc/examples in the bash-doc package.

if [ -f ~/.bash_aliases ]; then
    . ~/.bash_aliases
fi

# enable programmable completion features (you don't need to enable
# this, if it's already enabled in /etc/bash.bashrc and /etc/profile
# sources /etc/bash.bashrc).
if ! shopt -oq posix; then
    if [ -f /usr/share/bash-completion/bash_completion ]; then
        . /usr/share/bash-completion/bash_completion
    elif [ -f /etc/bash_completion ]; then
        . /etc/bash_completion
    fi
fi

export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64
```

9. `File Name to Write: .bashrc`
10. Ďalej nainštalujeme openssh ktorý bude potrebný neskorej pre správne fungovanie hadoopu  
`sudo apt install openssh-server openssh-client -y`

11. Pomocou príkazu povolíme aby používateľ hadoopu nemusel stále zadávať ssh heslo  
`ssh-keygen -t rsa -P "" -f ~/.ssh/id_rsa`
12. Pomocou príkazu `cat` uložíme verejný kľúč ako autorizované kľúče v adresári ssh:  
`cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys`
13. Nastavíme povolenia pre svojho používateľa pomocou príkazu `chmod`:  
`chmod 0600 ~/.ssh/authorized_keys`

14. skontrolujeme nastavenia ssh na localhoste  
`ssh localhost`

15. ďalšími krokmi je stiahnutie a inštalácia hadoopu

wget <mirror link> pre verziu 3.3.3 treba nahradiť <mirror link>  
<https://dlcdn.apache.org/hadoop/common/hadoop-3.3.3/hadoop-3.3.3.tar.gz>

16. rozbalíme súbor

```
tar xzf hadoop-3.3.0.tar.gz
```

17. ďalej budeme modifikovať nasledujúce súbory

```
bashrc
hadoop-env.sh
core-site.xml
hdfs-site.xml
mapred-site.xml
yarn-site.xml
```

18. ##.bashrc

```
sudo nano .bashrc
```

Pridajte toto na koniec súboru a nezabudnite ho uložiť

```
#Hadoop Related Options
```

```
export HADOOP_HOME=/home/meno_pouzivatela!!!/hadoop-3.3.0
```

```
export HADOOP_INSTALL=$HADOOP_HOME
```

```
export HADOOP_MAPRED_HOME=$HADOOP_HOME
```

```
export HADOOP_COMMON_HOME=$HADOOP_HOME
```

```
export HADOOP_HDFS_HOME=$HADOOP_HOME
```

```
export YARN_HOME=$HADOOP_HOME
```

```
export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_HOME/lib/native
```

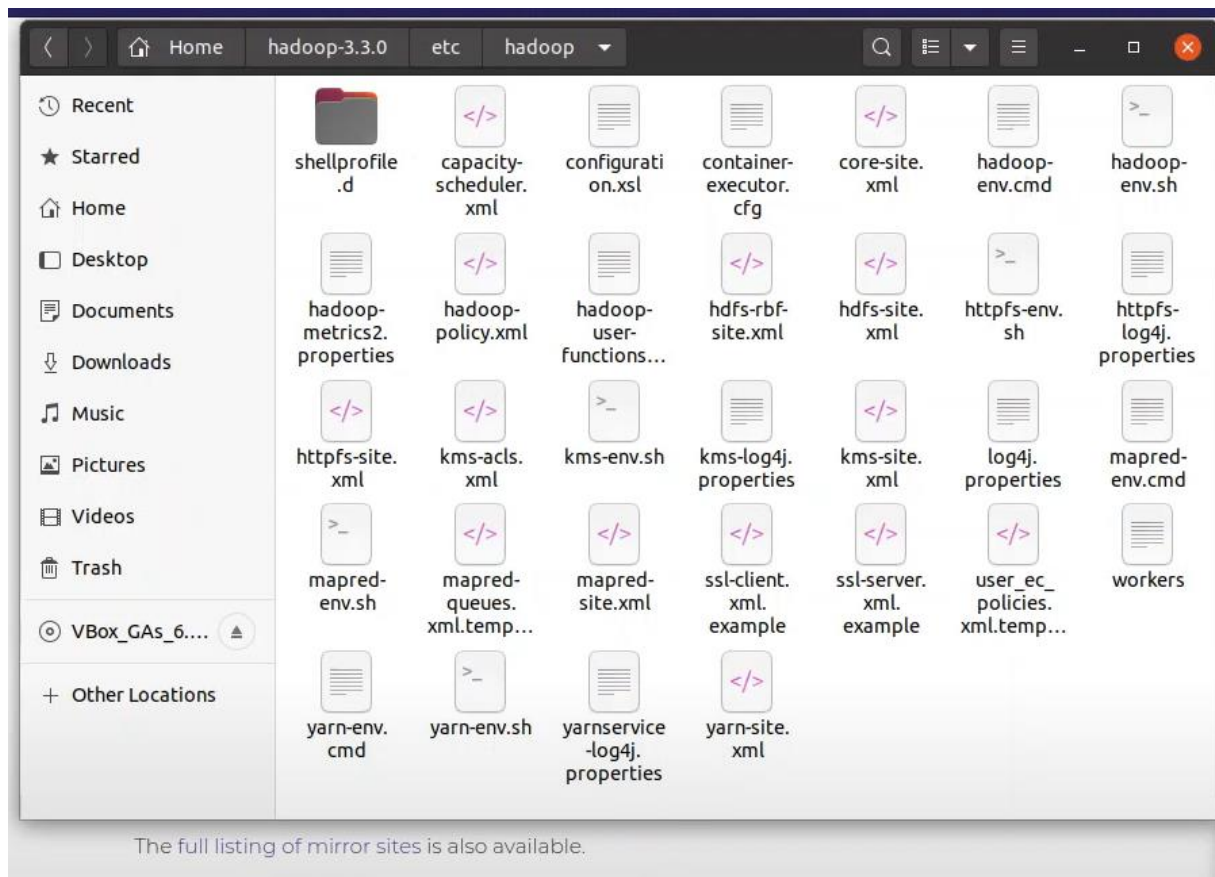
```
export PATH=$PATH:$HADOOP_HOME/sbin:$HADOOP_HOME/bin
```

```
export HADOOP_OPTS="-Djava.library.path=$HADOOP_HOME/lib/native"
```

19. Spustíte príkaz na termináli, aby boli viditeľné nové premenné prostredia.

```
source ~/.bashrc
```

20. ostatné súbory nájdeme v sobore etc a ďalej hadopp



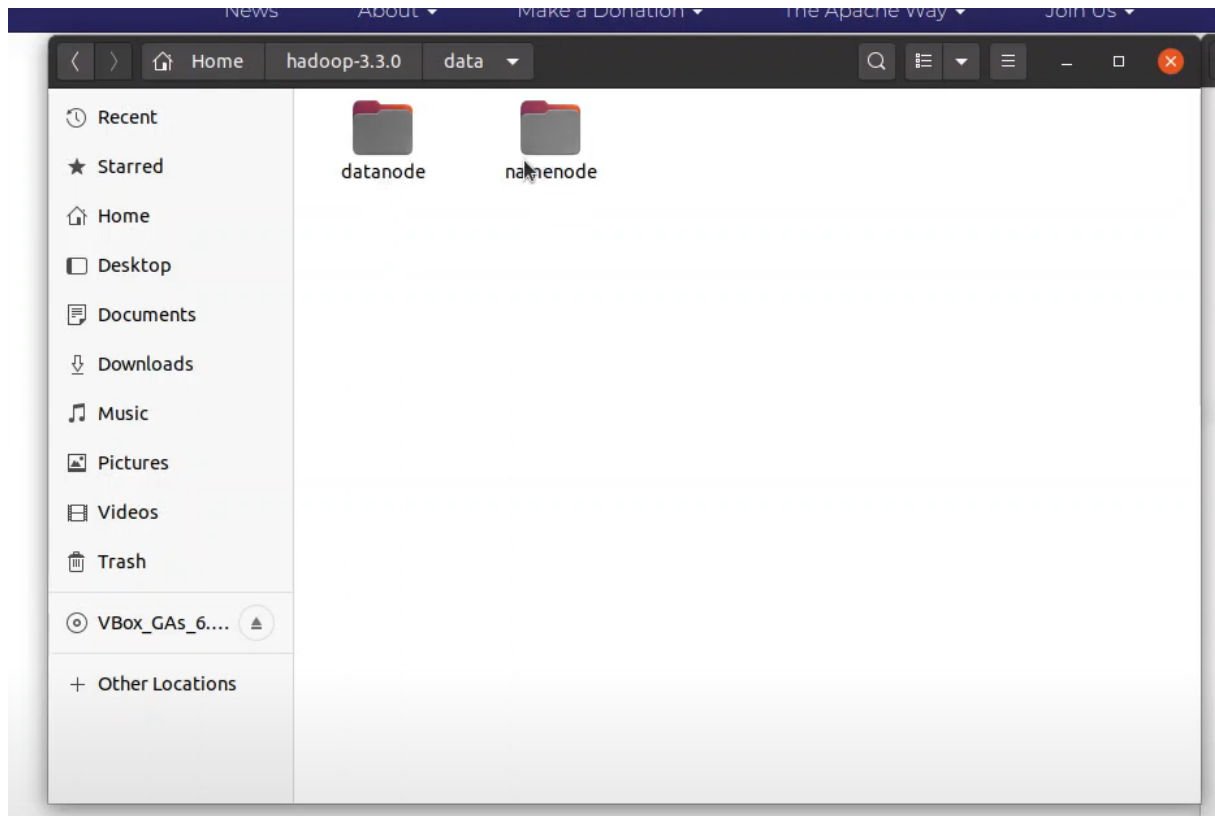
21. ##hadoop-env.sh

22. Odkomentovať JAVA\_HOME odstrániť # a modifikovať na  
 JAVA\_HOME=/usr/lib/jvm/java-8-openjdk-amd64

23. ##core-site.xml

```
<configuration>
  <property>
    <name>fs.default.name</name>
    <value>hdfs://localhost:9000</value>
  </property>
</configuration>
```

24. Vytvorte adresár pre datanodes a namenodes a pridajte jeho umiestnenie do súboru hdfs-site.xml



## 25. ##hdfs-site.xml

amit zmenš za meno aktuálneho Linux používateľa

```
<property>
  <name>dfs.replication</name>
  <value>1</value>
</property>
<property>
  <name>dfs.namenode.name.dir</name>
  <value>/home/amit/hadoop-3.3.0/data/namenode</value>
</property>
<property>
  <name>dfs.datanode.data.dir</name>
  <value>/home/amit/hadoop-3.3.0/data/datanode</value>
</property>
```

## 26. ##mapred-site.xml

```
<property>
  <name>mapreduce.framework.name</name>
  <value>yarn</value>
</property>
```

## 27. ##yarn-site.xml

```
<property>
  <name>yarn.nodemanager.aux-services</name>
  <value>mapreduce_shuffle</value>
</property>
<property>
  <name>yarn.nodemanager.aux-services.mapreduce.shuffle.class</name>
  <value>org.apache.hadoop.mapred.ShuffleHandler</value>
</property>
<property>
  <name>yarn.resourcemanager.hostname</name>
  <value>127.0.0.1</value>
</property>
<property>
  <name>yarn.acl.enable</name>
  <value>0</value>
</property>
<property>
  <name>yarn.nodemanager.env-whitelist</name>

  <value>JAVA_HOME,HADOOP_COMMON_HOME,HADOOP_HDFS_HOME,HADOOP_
CONF_DIR,CLASSPATH_PERPEND_DISTCACHE,HADOOP_YARN_HOME,HADOOP_M
APRED_HOME</value>
</property>
```

## 28. source ~/.bashrc

29 .Naformátujeme HDFS NameNode  
*hdfs namenode -format*

30 .Zapneme Hadoop Cluster  
*./start-dfs.sh*  
*./start-yarn.sh*  
*jps*

## 31. prístup k Hadoop UI z prehliadača

Hadoop NameNode UI: <http://localhost:9870>  
YARN Resource Manager: <http://localhost:8088>



## Záver

V tejto práci sme si popísali postup pomocou ktorého nainštalujeme Hadoop 3.3.3 na Linux. Postup bol otestovaný a dvoch počítačoch preto veríme že by mal byť funkčný.

## Bibliografia

<https://www.youtube.com/watch?v=8HLxeMMGGcQ&t=765s>