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](#_Toc128517890)

[Postup inštalácie 4](#_Toc128517891)

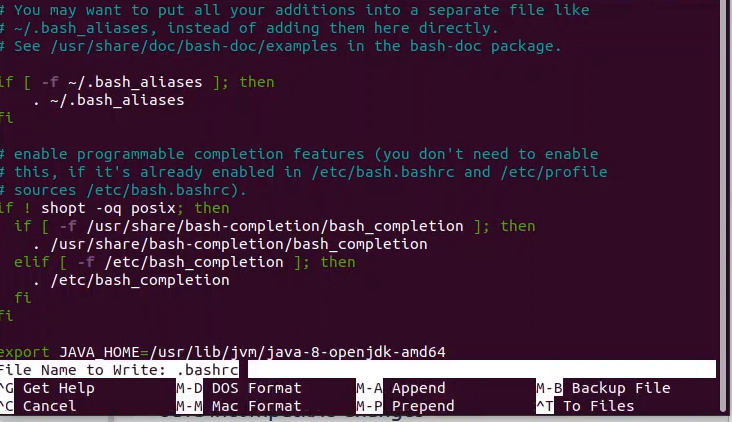
[Záver 9](#_Toc128517892)

[Bibliografia 10](#_Toc128517893)

# Úvod

V tomto zadaní 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* zistime 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.
9. **
10. Ďalej nainštalujeme openssh ktorý bude potrebný neskorej pre správne fungovanie hadoopu

*sudo apt install openssh-server openssh-client -y*

1. Pomocou príkazu povolíme aby používateľ hadoopu nemusel stále zadávať ssh heslo

*ssh-keygen -t rsa -P '' -f ~/.ssh/id\_rsa*

1. Pomocou príkazu cat uložíme verejný kľúč ako autorizované kľúče v adresári ssh:

*cat ~/.ssh/id\_rsa.pub >> ~/.ssh/authorized\_keys*

1. Nastavíme povolenia pre svojho používateľa pomocou príkazu chmod:

*chmod 0600 ~/.ssh/authorized\_keys*

1. skontrolujeme nastavenia ssh na localhoste

s *ssh localhost*

1. ďalšími krokmi je stiahnutie a inštalácia hadoopu

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

1. rozbalíme súbor

*tar xzf hadoop-3.3.0.tar.gz*

1. ďalej budeme modifikovať nasledujúce súbory

bashrc

*hadoop-env.sh*

*core-site.xml*

*hdfs-site.xml*

*mapred-site-xml*

*yarn-site.xml*

1. ##.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*

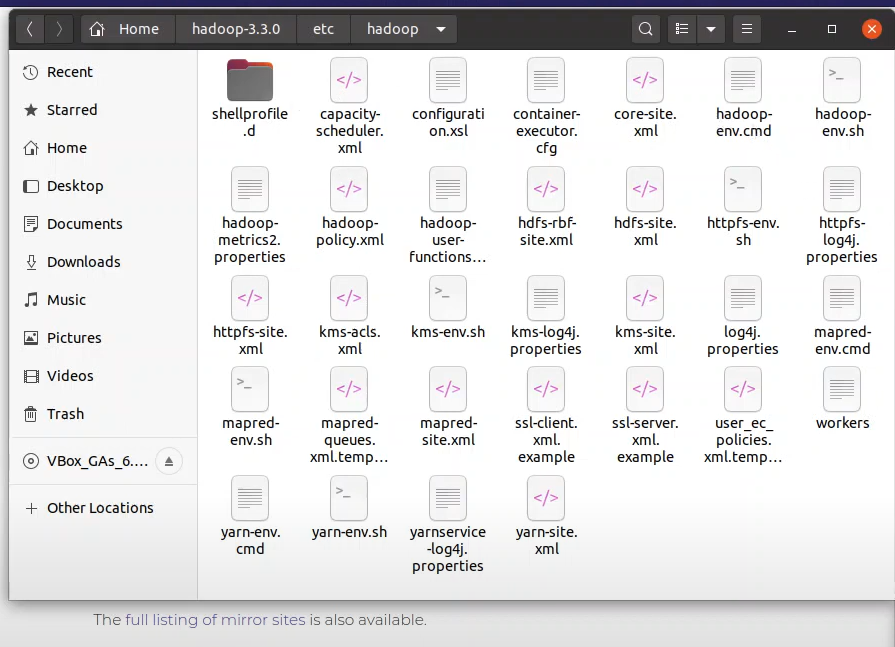
*export PATH=$PATH:$HADOOP\_HOME/sbin:$HADOOP\_HOME/bin*

*export HADOOP\_OPTS="-Djava.library.path=$HADOOP\_HOME/lib/native"*

1. Spustite príkaz na termináli, aby boli viditeľné nové premenné prostredia.

*source ~/.bashrc*

1. ostatné súbory nájdeme v sobore etc a ďalej hadopp



1. ##hadoop-env.sh
2. Odkomentovať JAVA\_HOME odstrániť # a modifikovať na

*JAVA\_HOME=/usr/lib/jvm/java-8-openjdk-amd64*

1. ##core-site.xml

*<configuration>*

*<property>*

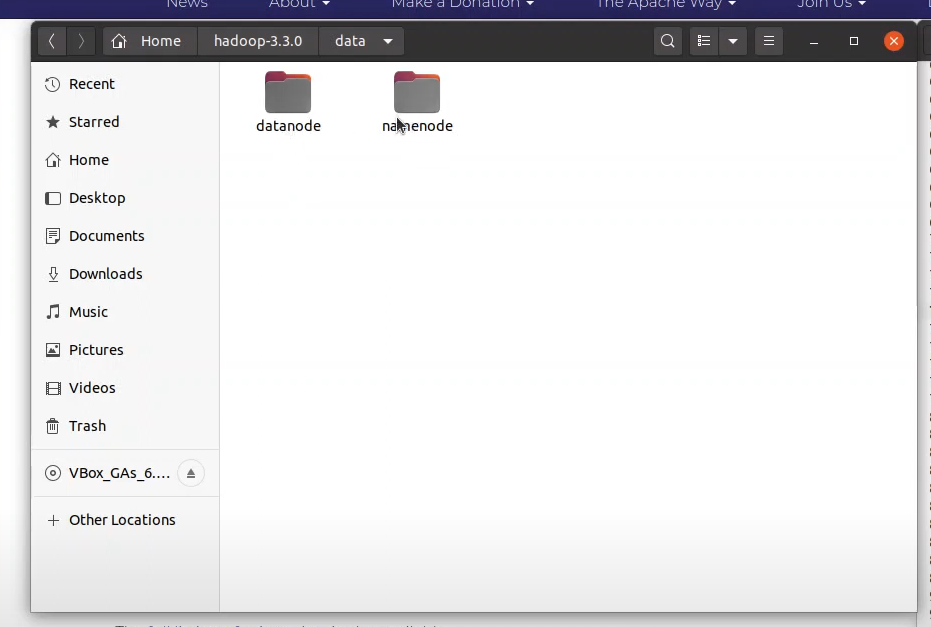
*<name>fs.default.name</name>*

*<value>hdfs://localhost:9000</value>*

*</property>*

*</configuration>*

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



1. ##hdfs-site.xml

amit zmeň 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>*

1. ##mapred-site-xml

*<property>*

*<name>mapreduce.framework.name</name>*

*<value>yarn</value>*

*</property>*

1. ##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\_MAPRED\_HOME</value>*

*</property>*

1. 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 praci sme si popísali postup pomocou ktorého nainštalujeme Hadoop 3.3.3 na Linux. Postup bol otestovaný a dvoch počítačoch preto veríme že by mal byt funkčný.

# Bibliografia

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