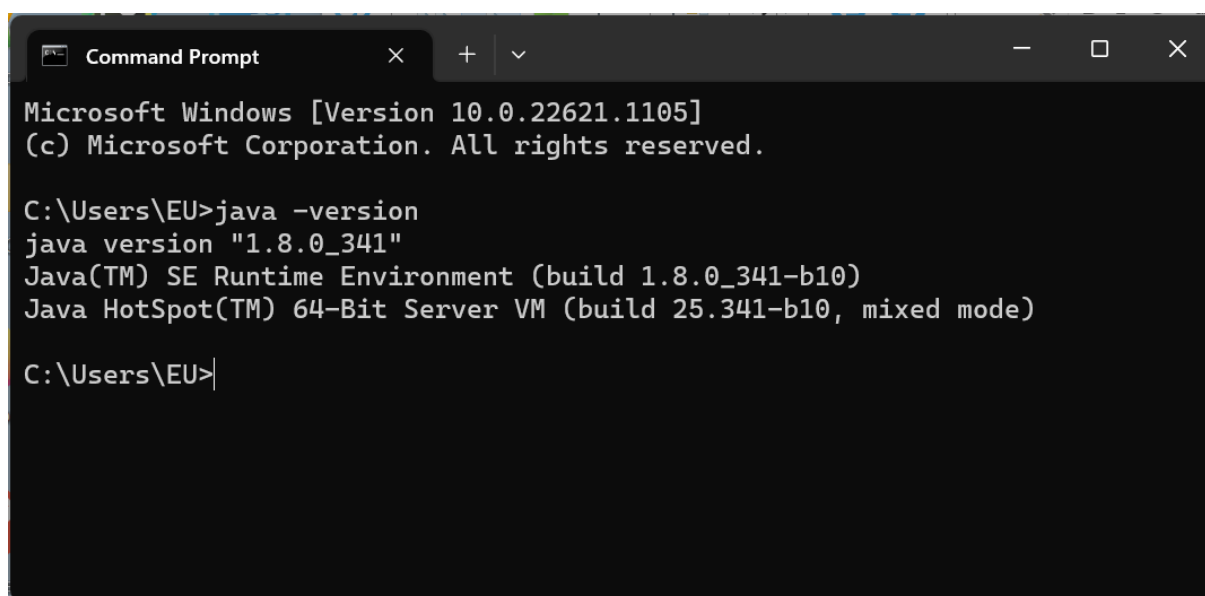


Inštalácia Hadoop 3.3.0 na Windows 10

1. Kontrola nainštalovanej verzie JAVA. Hadoop nepodporuje všetky verzie JAVA preto je dôležité mať nainštalovanú kompatibilnú verziu. Štandardne Hadoop 3.x požaduje verziu Java 8, Hadoop 2.x zase Java 7. Vyššie verzie buď vôbec nepodporujú alebo sú zdrojom viacerých problémov
2. Otestovaná je verzia fungujúca s Hadoop 3.3.0 je napr. Java 1.8.0_341,
3. V command prompte príkazom **java -version** zistím nainštalovanú verziu



```
Microsoft Windows [Version 10.0.22621.1105]
(c) Microsoft Corporation. All rights reserved.

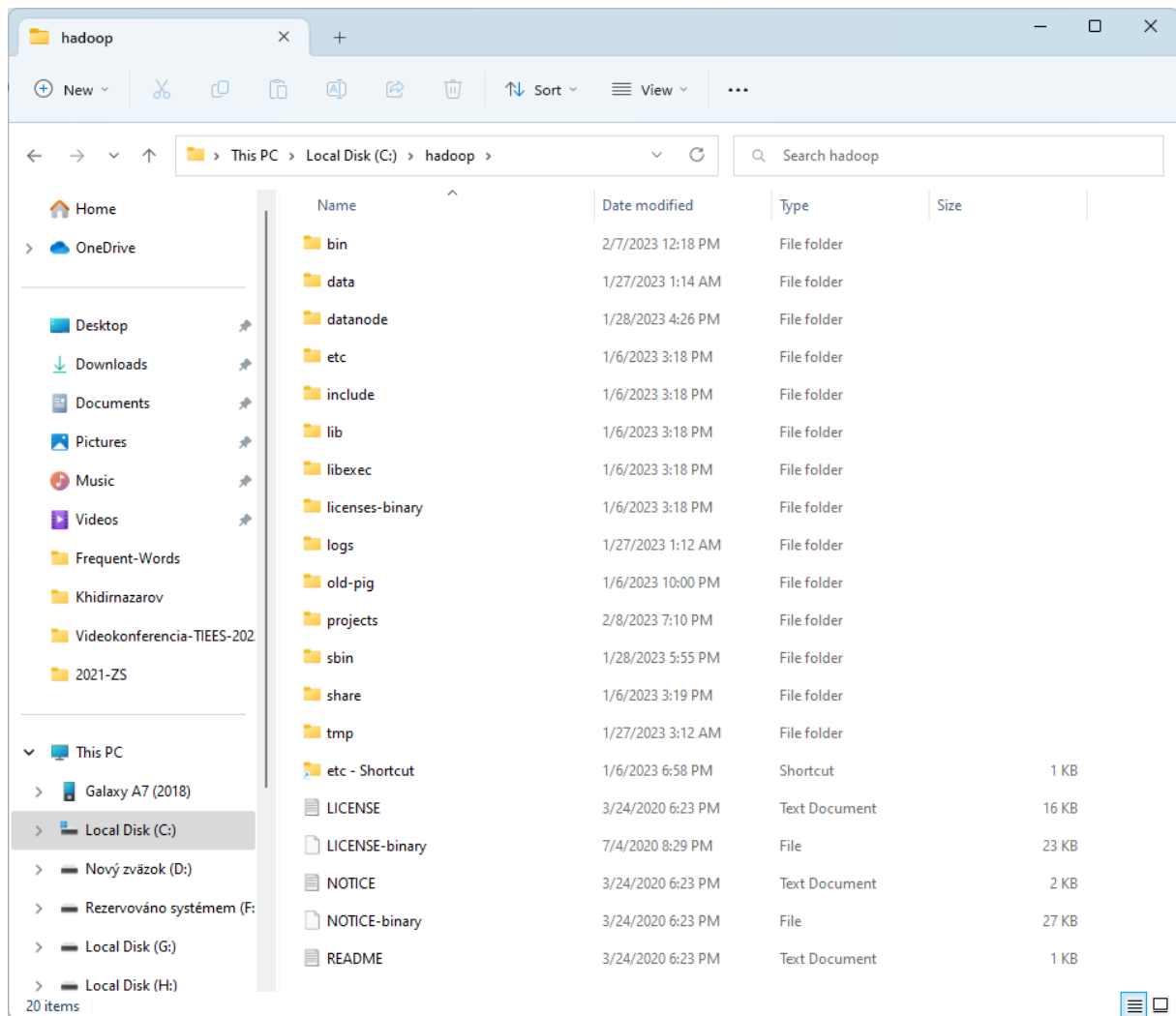
C:\Users\EU>java -version
java version "1.8.0_341"
Java(TM) SE Runtime Environment (build 1.8.0_341-b10)
Java HotSpot(TM) 64-Bit Server VM (build 25.341-b10, mixed mode)

C:\Users\EU>
```

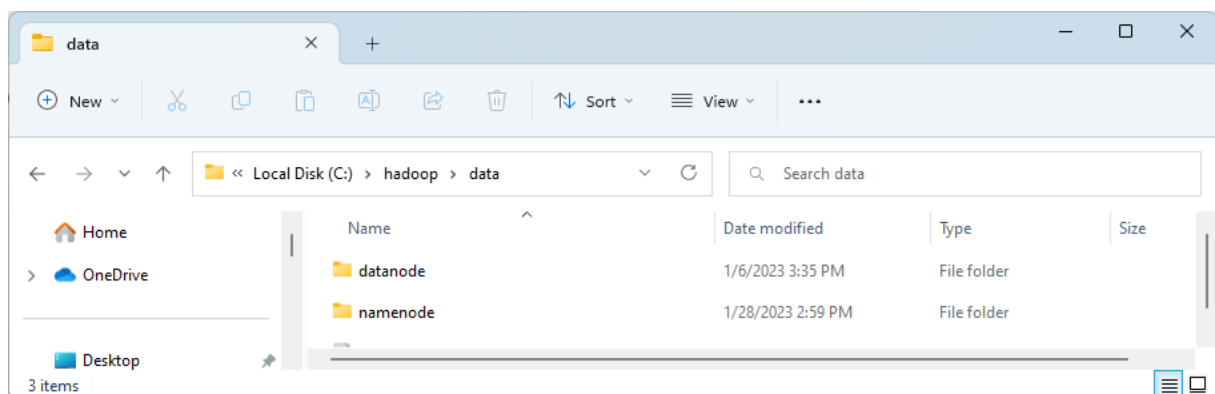
4. Ak Java nie je nainštalovaná alebo je to nevhodná verzia treba ju vymazať a nainštalovať odporúčanú verziu.
5. jdk-8u341-windows-x64.exe môžete stiahnuť
6. Stiahnutie Hadoop 3.3.0
<http://apache.mirror.amaze.com.au/hadoop/common/hadoop-3.3.0/hadoop-3.3.0.tar.gz>

Môžete použiť aj inú verziu, ale treba počítať s tým že bude treba riešiť problémy

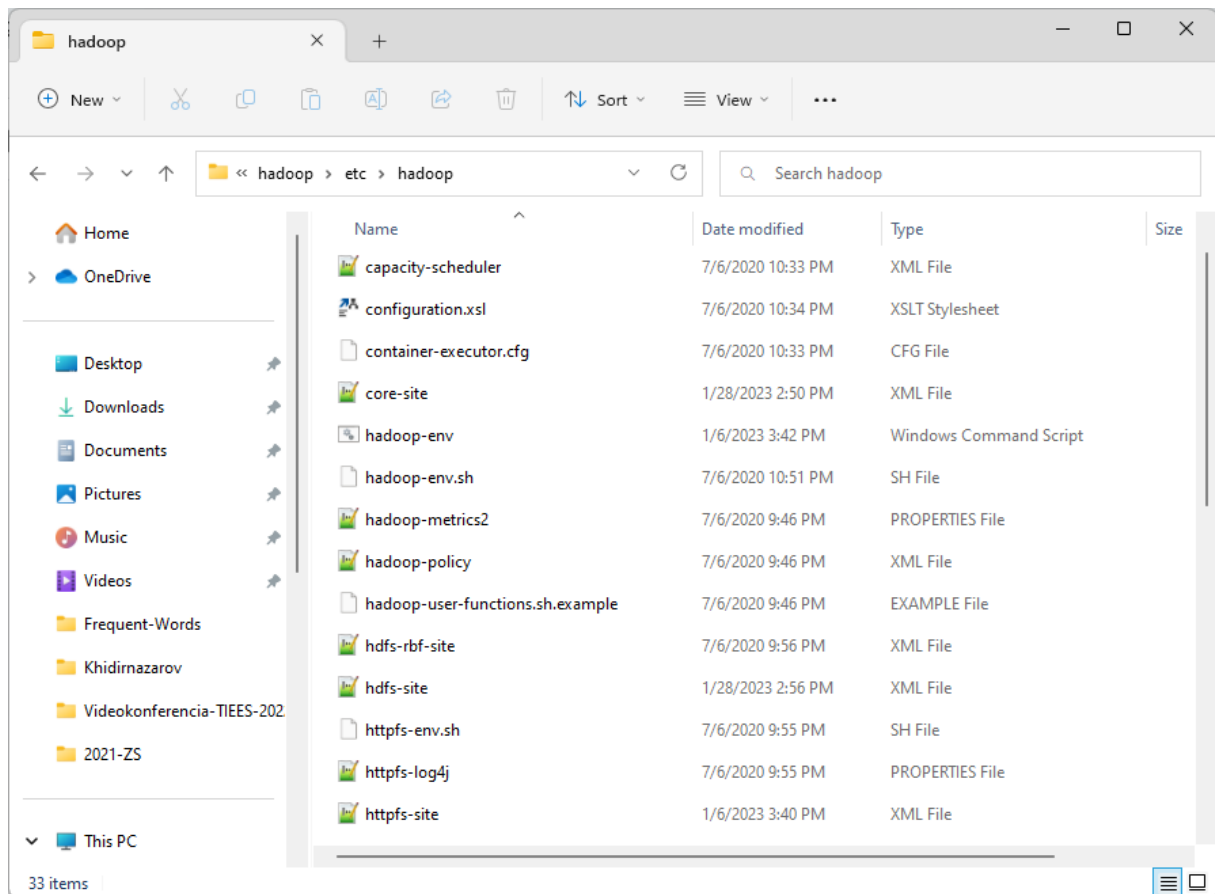
7. Rozbalený adresár nahráme na cieľové miesto. V princípe môže byť v inom adresári alebo na inom mieste, ale najmenej problémov sa vyskytne keď je v roote na bootovacom disku v našom prípade disk C. Názov adresára Hadoop 3.3.0 sme premenovali na Hadoop. *Pozn. Je to dobré urobiť ešte pred tým než nastavíme cesty k súborom.*



8. V adresári data vytvoríme 2 adresáre **datanode** a **namenode**



9. Stiahneme si konfiguračný súbor pre Hadoop 3.3.0 - 3.3.0 Configuration Files Apache Hadoop.zip, ktorý obsahuje jeden adresár **bin**. Obsah tohto adresára nakopírujeme do pôvodného adresára bin, a pôvodné súbory prepíšeme.
10. V ďalšom kroku si otvoríme adresár **etc** a následne **hadoop**



11. V tomto adresári budeme konfigurovať 4 súbory xml a 1 cmd

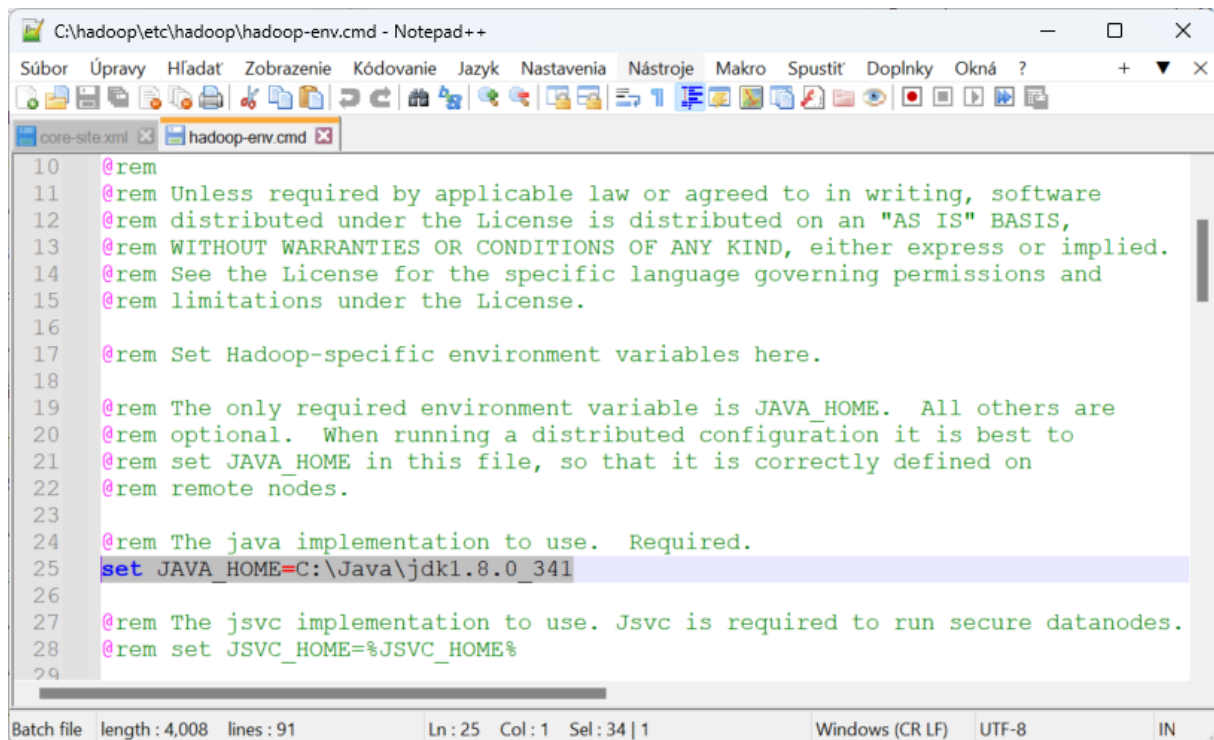
- core-site.xml
- hadoop-env.cmd
- hdfs-site.xml
- mapred-site.xml
- yarn-site.xml

12. core-site.xml

doplníme konfiguráciu na

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

13. Nastavíme hadoop-env.cmd set JAVA_HOME=C:\Java\jdk1.8.0_341



```
10 @rem
11 @rem Unless required by applicable law or agreed to in writing, software
12 @rem distributed under the License is distributed on an "AS IS" BASIS,
13 @rem WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
14 @rem See the License for the specific language governing permissions and
15 @rem limitations under the License.
16
17 @rem Set Hadoop-specific environment variables here.
18
19 @rem The only required environment variable is JAVA_HOME. All others are
20 @rem optional. When running a distributed configuration it is best to
21 @rem set JAVA_HOME in this file, so that it is correctly defined on
22 @rem remote nodes.
23
24 @rem The java implementation to use. Required.
25 set JAVA_HOME=C:\Java\jdk1.8.0_341
26
27 @rem The jsvc implementation to use. Jsvc is required to run secure datanodes.
28 @rem set JSVC_HOME=%JSVC_HOME%
29
```

14. Nasleduje hdfs-site.xml, kde nastavíme konfiguráciu na

```
<configuration>
  <property>
    <name>dfs.replication</name>
    <value>1</value>
  </property>
  <property>
    <name>dfs.namenode.name.dir</name>
    <value>file:///C:/hadoop/data/namenode</value>
  </property>
  <property>
    <name>dfs.datanode.data.dir</name>
    <value>datanode</value>
  </property>
</configuration>
```

15. Mapred-site.xml

```
<configuration>
  <property>
    <name>mapreduce.framework.name</name>
```

```
<value>yarn</value>
</property>
</configuration>
```

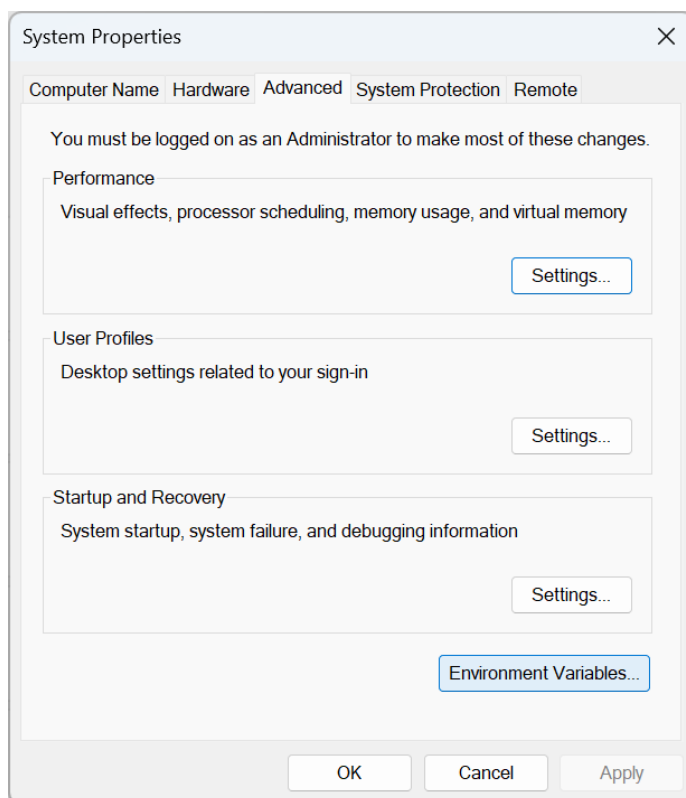
16. Yarn-site.xml upravíme na

```
<configuration>
  <property>
    <name>yarn.nodemanager.aux-services</name>
    <value>mapreduce_shuffle</value>
  </property>
  <property>
    <name>yarn.nodemanager.auxservices.mapreduce.shuffle.class</name>
    <value>org.apache.hadoop.mapred.ShuffleHandler</value>
  </property>
</configuration>
```

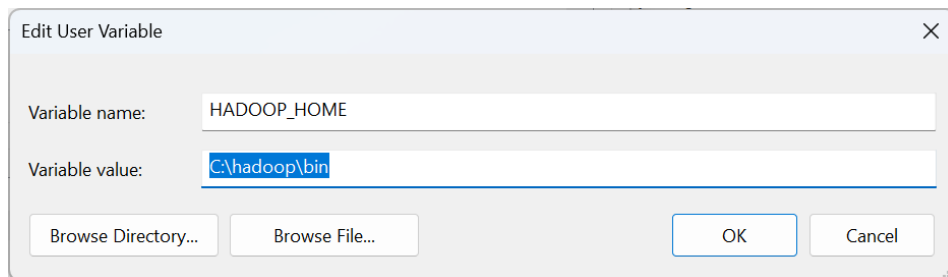
Nezabudnite všetky konfiguračné zmeny uložiť.

17. Nasleduje nastavenie prostredia

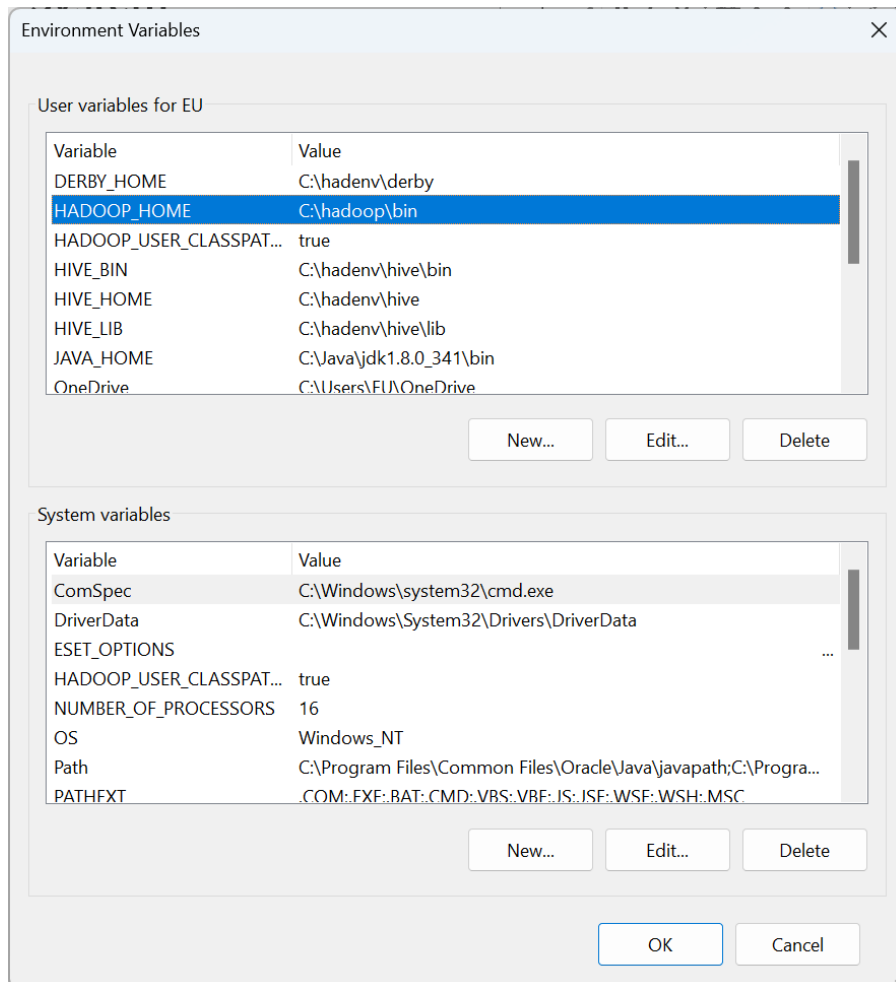
- Control panel / System / Advanced system settings
- V okne System properties vyberieme Enviroment variables



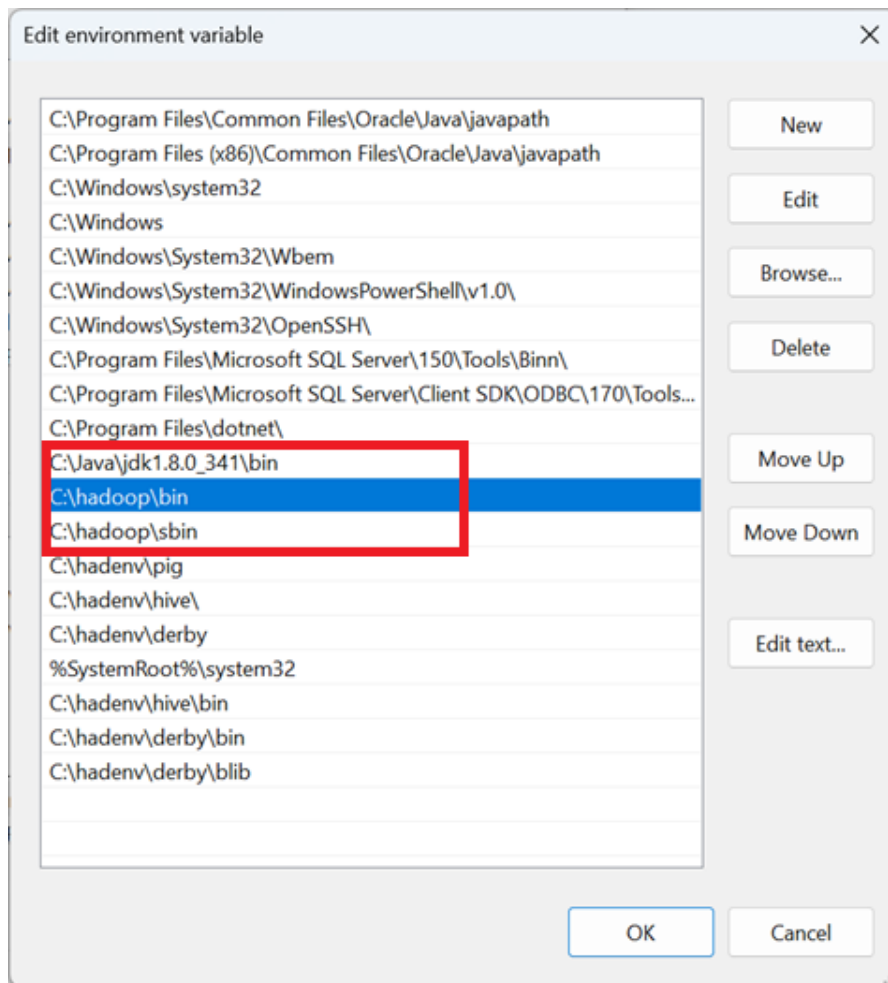
- Skontrolujeme či máme dobre nastavené používateľské premenné. Pri prvej inštalácii nám bude chýbať HADOOP_HOME, cez tlačidlo NEW si otvoríme editačné okno a vyplníme názov premennej HODOOP_HOME a hodnotu premennej C:\hadoop\bin, čiže adresár kde sú spúšťače súbory



Po potvrdení sa údaje zapíšu do tabuľky



- Rovnako postupujeme aj pri zadávaní Systémových premenných. Tu si musíme nastaviť správnu cestu preto klikneme na **path** a následne na tlačidlo **Edit..**
- Ak v rámečku vyznačené cesty v zozname nemáme, klikneme na tlačidlo **New**. Ak cesty v zozname sú, len sú nesprávne môžeme použiť tlačidlo **Edit**.



- f. Dôležité je mať zadané tieto položky
- i. C:\Java\jdk1.8.0_341\bin
 - ii. C:\hadoop\bin
 - iii. C:\hadoop\sbin

Zmeny potvrdíme tlačidlom OK, a okno System Properties zavrieme tlačidlom OK.

18. Naformátujem namenode príkazom `hdfs namenode -format`

```

Command Prompt
Microsoft Windows [Version 10.0.22621.1105]
(c) Microsoft Corporation. All rights reserved.

C:\Users\EU>hdfs namenode -format
  
```

Keď systém akceptuje príkaz tak sa nás spýta, či chceme namenode preformátovať

```
Command Prompt - hdfs nan x + v
op.num.users = 10
2023-02-15 18:12:33,062 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.t
op.windows.minutes = 1,5,25
2023-02-15 18:12:33,066 INFO namenode.FSNamesystem: Retry cache on namenode
is enabled
2023-02-15 18:12:33,066 INFO namenode.FSNamesystem: Retry cache will use 0.
03 of total heap and retry cache entry expiry time is 600000 millis
2023-02-15 18:12:33,068 INFO util.GSet: Computing capacity for map NameNode
RetryCache
2023-02-15 18:12:33,068 INFO util.GSet: VM type = 64-bit
2023-02-15 18:12:33,068 INFO util.GSet: 0.029999999329447746% max memory 88
9 MB = 273.1 KB
2023-02-15 18:12:33,068 INFO util.GSet: capacity = 2^15 = 32768 entrie
s
Re-format filesystem in Storage Directory root= C:\hadoop\data\namenode; lo
cation= null ? (Y or N) |
```

Po zadání Y sa namenode preformátoval

```
Command Prompt x + v
2023-02-15 18:15:51,408 INFO common.Storage: Storage directory C:\hadoop\da
ta\namenode has been successfully formatted.
2023-02-15 18:15:51,440 INFO namenode.FSImageFormatProtobuf: Saving image f
ile C:\hadoop\data\namenode\current\fsimage.ckpt_00000000000000000000 using
no compression
2023-02-15 18:15:51,531 INFO namenode.FSImageFormatProtobuf: Image file C:\
hadoop\data\namenode\current\fsimage.ckpt_00000000000000000000 of size 397 b
ytes saved in 0 seconds .
2023-02-15 18:15:51,539 INFO namenode.NNStorageRetentionManager: Going to r
etain 1 images with txid >= 0
2023-02-15 18:15:51,543 INFO namenode.FSImage: FSImageSaver clean checkpoi
nt: txid=0 when meet shutdown.
2023-02-15 18:15:51,543 INFO namenode.NameNode: SHUTDOWN_MSG:
/*****
SHUTDOWN_MSG: Shutting down NameNode at N-S-20306305234/192.168.56.1
*****/
C:\Users\EU>|
```

Príkazom C:\Users\EU>start-dfs.cmd spúšťame namenode a datanode a príkazom C:\Users\EU> start-yarn.cmd spúšťame yarn.

Na obr. vidíme spustené 4 okná s názvom Apache Hadoop Distribution.

Príkazom JPS môžeme otestovať či sa vykonávajú všetky služby


```
Feb 15, 2023 6:23:46 PM com.sun.jersey.server.impl.application.WebApplicationImpl _initiate
INFO: Initiating Jersey application, version 'Jersey: 1.19 02/11/2015 03:25 AM'
Feb 15, 2023 6:23:46 PM com.sun.jersey.guice.spi.container.GuiceComponentProviderFactory getComponentProvider
INFO: Binding org.apache.hadoop.yarn.server.nodemanager.webapp.JAXBContextResolver to GuiceManagedComponentProvider with
the scope "Singleton"
Feb 15, 2023 6:23:46 PM com.sun.jersey.guice.spi.container.GuiceComponentProviderFactory getComponentProvider
INFO: Binding org.apache.hadoop.yarn.webapp.GenericExceptionHandler to GuiceManagedComponentProvider with the scope "Sin
gleton"
Feb 15, 2023 6:23:46 PM com.sun.jersey.guice.spi.container.GuiceComponentProviderFactory getComponentProvider
INFO: Binding org.apache.hadoop.yarn.server.nodemanager.webapp.NMWebServices to GuiceManagedComponentProvider with the s
cope "Singleton"
2023-02-15 18:23:46,499 INFO handler.Context
ppData/Local/Temp/jetty-0_0_0-8042-_-any-2
arn/hadoop-yarn-common-3.3.0.jar!/webapps/no
2023-02-15 18:23:46,511 INFO server.Abstract
2}
2023-02-15 18:23:46,511 INFO server.Server:
2023-02-15 18:23:46,511 INFO webapp.WebApps:
2023-02-15 18:23:46,511 INFO nodemanager.No
2023-02-15 18:23:46,513 INFO util.JvmPauseMo
2023-02-15 18:23:46,519 INFO client.DefaultM
2023-02-15 18:23:46,581 INFO nodemanager.No
2023-02-15 18:23:46,593 INFO nodemanager.No
2023-02-15 18:23:46,786 INFO security.NMCont
th id 215912575
2023-02-15 18:23:46,787 INFO security.NMTok
-1271597086
2023-02-15 18:23:46,787 INFO nodemanager.No
with total resource of <memory:8192, vCores

C:\Users\EU>stop-all.cmd
This script is Deprecated. Instead use stop-dfs.cmd and stop-yarn.cmd
SUCCESS: Sent termination signal to the process with PID 17512.
INFO: No tasks running with the specified criteria.
stopping yarn daemons
SUCCESS: Sent termination signal to the process with PID 10792.
SUCCESS: Sent termination signal to the process with PID 8424.
INFO: No tasks running with the specified criteria.

C:\Users\EU>start-dfs.cmd
C:\Users\EU>start-yarn.cmd
starting yarn daemons
C:\Users\EU>
```

```
C:\Users\EU>jps
1376 Jps
15680 ResourceManager
7892 DataNode
14424 NameNode
6392 NodeManager

C:\Users\EU>
```

Vykonávanie programov môžeme overiť aj spustením localhost:8088

All Applications														
ID	User	Name	Application Type	Application Tags	Queue	Application Priority	StartTime	LaunchTime	FinishTime	State	FinalStatus	Running Containers	Allocated CPU V-Cores	Allocated Memory MB
No data available in table														

A localhost:9870

Namenode information x +

localhost:9870/dfshealth.html#tab-overview

Hadoop Overview Datanodes Datanode Volume Failures Snapshot Startup Progress Utilities

Overview 'localhost:9000' (✓active)

Started:	Wed Feb 15 19:06:29 +0100 2023
Version:	3.3.0, raa96f1871bfd858f9bac59cf2a81ec470da649af
Compiled:	Mon Jul 06 20:44:00 +0200 2020 by brahma from branch-3.3.0
Cluster ID:	CID-57311021-587a-4657-803a-0fc07e488df6
Block Pool ID:	BP-1524194147-192.168.56.1-1676482860455

Summary

Security is off.
Safemode is off.
1 files and directories, 0 blocks (0 replicated blocks, 0 erasure coded block groups) = 1 total filesystem object(s).
Heap Memory used 118.37 MB of 467 MB Heap Memory. Max Heap Memory is 889 MB.
Non Heap Memory used 52.3 MB of 53.31 MB Committed Non Heap Memory. Max Non Heap Memory is <unbounded>.

Configured Capacity:	0 B
Configured Remote Capacity:	0 B
DFS Used:	0 B (100%)
Non DFS Used:	0 B
DFS Remaining:	0 B (0%)
Block Pool Used:	0 B (100%)
DataNodes usages% (Min/Median/Max/stdDev):	0.00% / 0.00% / 0.00% / 0.00%
Live Nodes	0 (Decommissioned: 0, In Maintenance: 0)
Dead Nodes	0 (Decommissioned: 0, In Maintenance: 0)
Decommissioning Nodes	0
Entering Maintenance Nodes	0
Total Datanode Volume Failures	0 (0 B)

Zastaviť vykonávanie služieb Hadoopu môžeme pomocou príkazu **stop**

```
Microsoft Windows [Version 10.0.22621.1105]
(c) Microsoft Corporation. All rights reserved.

C:\Users\EU>stop-yarn.cmd
stopping yarn daemons
SUCCESS: Sent termination signal to the process with PID 11016.
SUCCESS: Sent termination signal to the process with PID 1048.

INFO: No tasks running with the specified criteria.

C:\Users\EU>stop-dfs.cmd
```