**EKONOMICKÁ UNIVERZITA V BRATISLAVE**

**Fakulta hospodárskej informatiky**



**Projekt z predmetu Big Data 01**

Bc. Richard Mišek

2023   Bc. Pavol Melko

**OBSAH**

**Analyzovať nástroje na tvorbu distribuovaného spracovania údajov a nástroja Hadoop**

* **Inštalácia Hadoop na systém Windows (Mišek Richard)**
* **Inštalácia Hadoop na systém Linux (Melko Pavol – ďalší súbor)**
* **Záver (Mišek Richard)**

# Inštalácia Hadoop na Windows

Inštalácia sa skladá z viacerých krokov :

1. V systéme je nutné mať nainštalovaný programovací jazyk JAVA verzie 8, pri používaní Hadoop verzie 3.x. Či je JAVA nainštalovaná zistíme po otvorení príkazového riadku (z angl. command line) a vpísaním príkazu java -version.

V prípade ak JAVA alebo jej požadovaná verzia nie je nainštalovaná, pristúpime ku kroku 2, inak prechádzame na krok 4

Obrázok, na ktorom je text

Automaticky generovaný popis

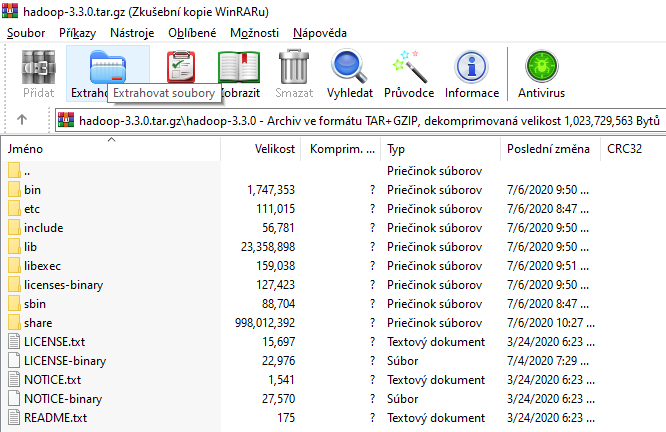
1. Na stiahnutie JAVA môžeme využiť tento odkaz, ktorý automaticky spustí jej sťahovanie. https://kai.ponuky.info/BD/jdk-8u341-windows-x64.exe
2. Následne po nainštalovaní, kde sa preklikáme poľami next, uložíme JAVU 2x do priečinku s názvom Java, na úložisku C, ktorý si vytvoríme, kde vytvoríme aj dva priecinky podľa počiatočných názvov čo nám ponúkla inštalácia jdk jre. Neodporúča sa ukladať do program files nakoľko medzi dvoma slovami je medzera a Hadoop inštalácia následne robí problémy.Obrázok, na ktorom je text

   Automaticky generovaný popis
3. Opäť overíme či bola JAVA správne nainštalovaná, zatvorením príkazového riadku a opätovným otvorením, kde vypíšeme java – version.

Obrázok, na ktorom je text

Automaticky generovaný popis

1. Po nainštalovaní JAVA prejdeme k samotnej inštalácii Hadoop 3.3.0, ktorý môžeme stiahnuť napríklad z adresy <https://kai.ponuky.info/BD/hadoop-3.3.0.tar.gz>
2. Rozbalený súbor extrahuje do priečinku, ktorý si vytvoríme na disku C. Priečinok nazveme hadoop a vložíme doň obsah z extrahovateľného priečinku hadoop-3.3.0.



1. V prípade ak sa export nepodarí, treba najskôr spustiť winrar ako správca a postup opakovať

Obrázok, na ktorom je text

Automaticky generovaný popis

1. V adresári hadoop vytvoríme priečinok data, do ktorého vytvoríme ďalšie dva priečinky datanode a namenode

Obrázok, na ktorom je text

Automaticky generovaný popis

1. Následne je nutné si stiahnuť konfiguračný súbor pre Hadoop 3.3.0, ktorý obsahuje adresár bin, ktorý nakopírujeme do pôvodného adresára bin a pôvodné súbory tak prepíšeme.

<https://kai.ponuky.info/BD/3.3.0%20Configuration%20Files%20Apache%20Hadoop.zip>

1. Otvoríme adresát etc a v ňom adresár hadoop.
2. Budeme tu konfigurovať 5 súborov, ktoré otvoríme napríklad pomocou notepad++ :

core-site.xml

hadoop-env.cmd

hdfs-site.xml

mapred-site.xml

yarn-site.xml

1. **core-site.xml**

doplníme konfiguráciu :

<configuration>

<property>

<name>fs.defaultFS</name>

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

</property>

</configuration>

1. **hadoop-env.cmd**

Nasledový riadok je 25.riadok v súbore. Nastavíme tak cestu k Java, ktorú si pozrite presne podľa nainštalovanej – môže sa líšiť. (V prípade ak sa Java ajtak nainštalovala do program files... tak ju presunieme opätovne do C: manuálne namiesto pôvodného priečinku)

set JAVA\_HOME=C:\Java\jdk1.8.0\_341

1. **hdfs-site.xml**

Nastavíme cesty k vytvoreným namenode a datanode

<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> file:///C:/hadoop/data/datanode</value>

</property>

</configuration>

1. **mapred-site.xml**

<configuration>

<property>

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

<value>yarn</value>

</property>

</configuration>

1. **yarn-site.xml**

<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>

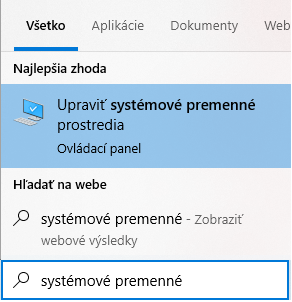
</property>

</configuration>

Všetky vykonané zmeny uložíme

1. Nastavenie prostredia

Do Windows prehliadača vpíšeme systémové premenné a stlačíme po otvorení premenné prostredia.

 Obrázok, na ktorom je text

Automaticky generovaný popis

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 spúšťacie súbory

Obrázok, na ktorom je text

Automaticky generovaný popis

V spodnom okne klikneme na path a nastavíme rovnako aj systémové premenné. Ľavým klikneme na Path, a následne na upraviť. Pridáme nižšie zobrazené tri cesty.

Obrázok, na ktorom je text

Automaticky generovaný popis

1. Naformátujeme namenode príkazom hdfs namenode -format
2. 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
3. Vykonávanie programov môžeme overiť aj spustením localhost:8088 A localhost:9870

Obrázok, na ktorom je text

Automaticky generovaný popis

Obrázok, na ktorom je text, snímka obrazovky, vnútri, niekoľko

Automaticky generovaný popis

**Záver**

Inštalácia bola uvádzaná na systéme windows. Je nutné sa nepomýliť v presných cestách nakoľko sa môžu líšiť. Podľa nasledovania tutoriálu by mal byť používateľ schopný previesť inštaláciu úspešne.