



ELTE | IK
INFORMATIKAI KAR

BigData architektúrák és elemző módszerek

1. gyakorlat

Bemutatkozás

- Varga Dániel, adjunktus (Információs Rendszerek Tanszék)
- E-mail: vargadaniel@inf.elte.hu
- Weboldal: vargadaniel.web.elte.hu
- Szoba: Déli épület, 2.506
- Használt rendszerek:
 - canvas.elte.hu - ZH-k
 - MS Teams – üzenetek, hirdetések

Tematika

- Architektúra ismeretek
 - Hadoop
 - Spark
 - MapReduce
- Elemző módszerek
 - Python (pandas, sklearn, numpy)
 - Vizualizáció (matplotlib)
 - SparkML, SKLearn

Követelmények

- Három részből fog állni:
 - Hadoop: beadandó
 - Spark: ZH
 - Python: ZH
- A három jegy átlaga megadja a gyakorlati jegyet

Hadoop

- Nyílt forráskódú keretrendszer
- Lehetőséget ad nagy adathalmazok feldolgozására számítógép klaszterek használatával
- Fő részei
 - Hadoop Distributed File System (HDFS) elosztott fájlrendszer
 - MapReduce programozási modell
 - YARN erőforrás menedzser



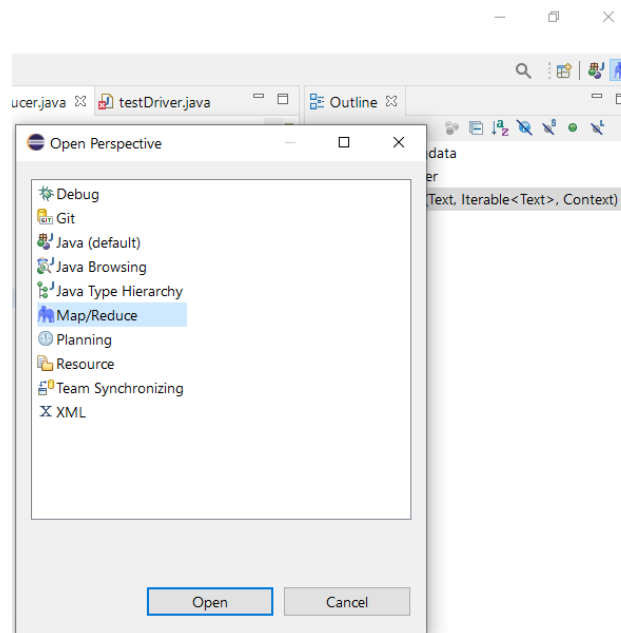
Mai gyakorlat célja: hadoop környezet kialakítása

- Töltsd le a [BigData.7z](#) fájlt és csomagold ki a C:\BigData mappába
- Add hozzá a felhasználó PATH környezeti változójához a C:\BigData\hadoop-3.3.6\bin útvonalat.
- BigData.7z letöltési linkek:
 - <http://oktnb16.inf.elte.hu/ggombos/bigdata/BigData.7z>
 - <https://drive.google.com/file/d/1aytbxZ18CNtQ1JepEb1zM7LpmVRpLYmq/view?usp=sharing>



Fejlesztői környezet: Eclipse

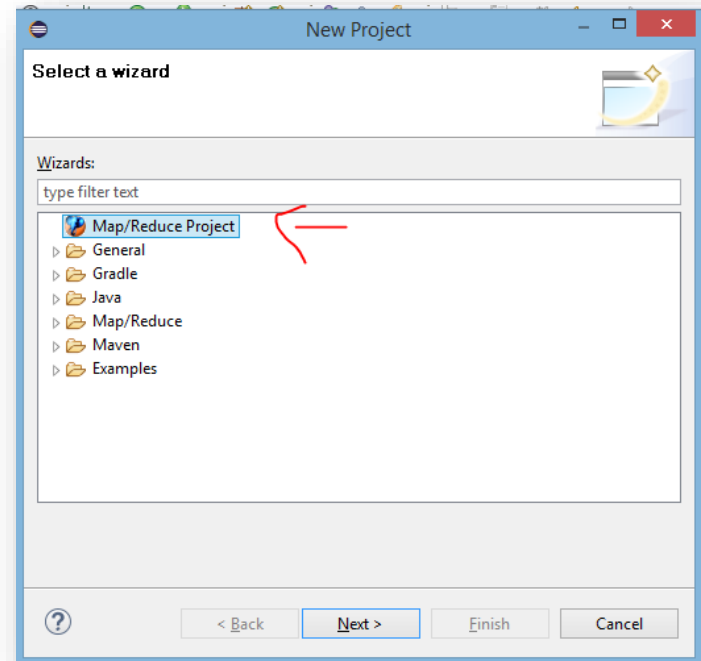
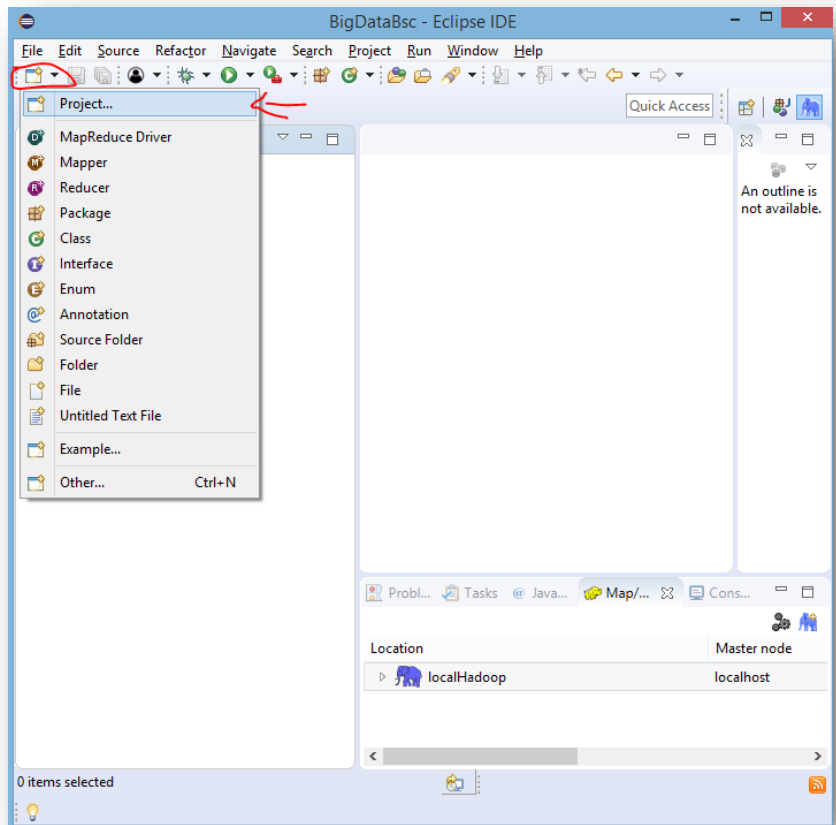
- Indítsuk el az Eclipset (a C:\BigData mappából) és állítsuk be a MapReduce nézetet.
- Logoláshoz a következő fájlt kell bemásolni a projekt src mappájába:
<https://vargadaniel.web.elte.hu/download/log4j.properties>



Nézet beállítása

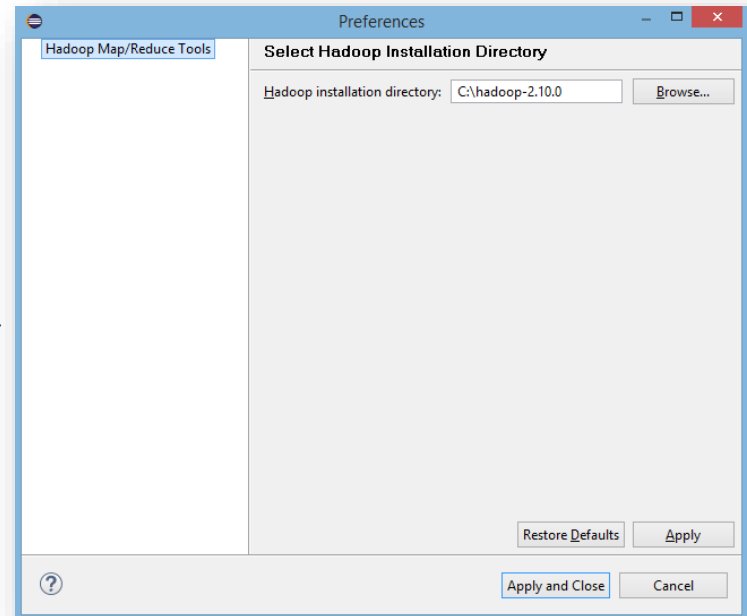
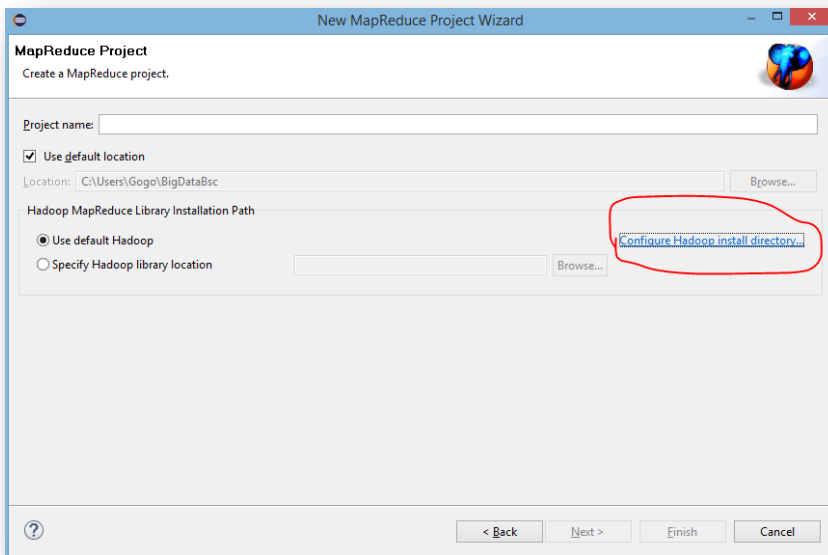
MapReduce projekt létrehozása Eclipseben

- 1. lépés: projekt létrehozása



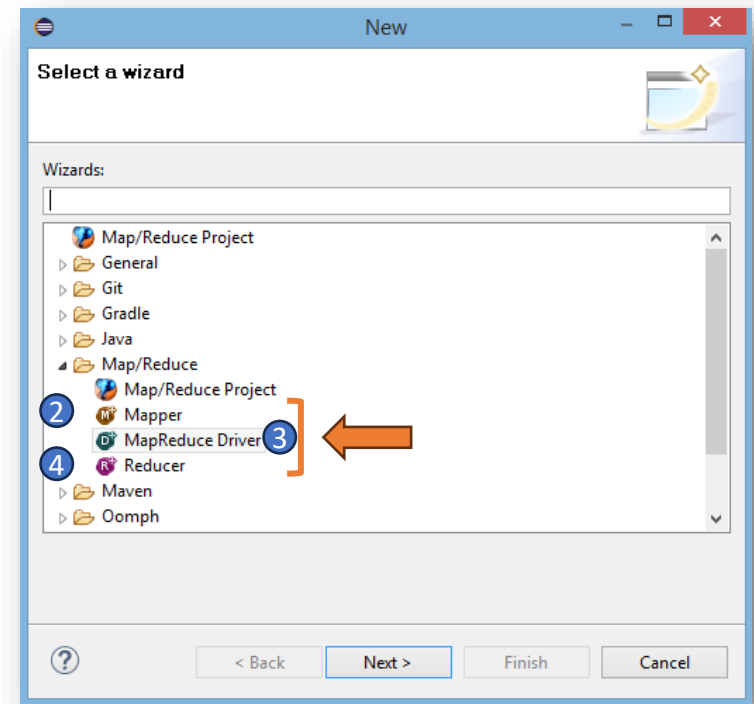
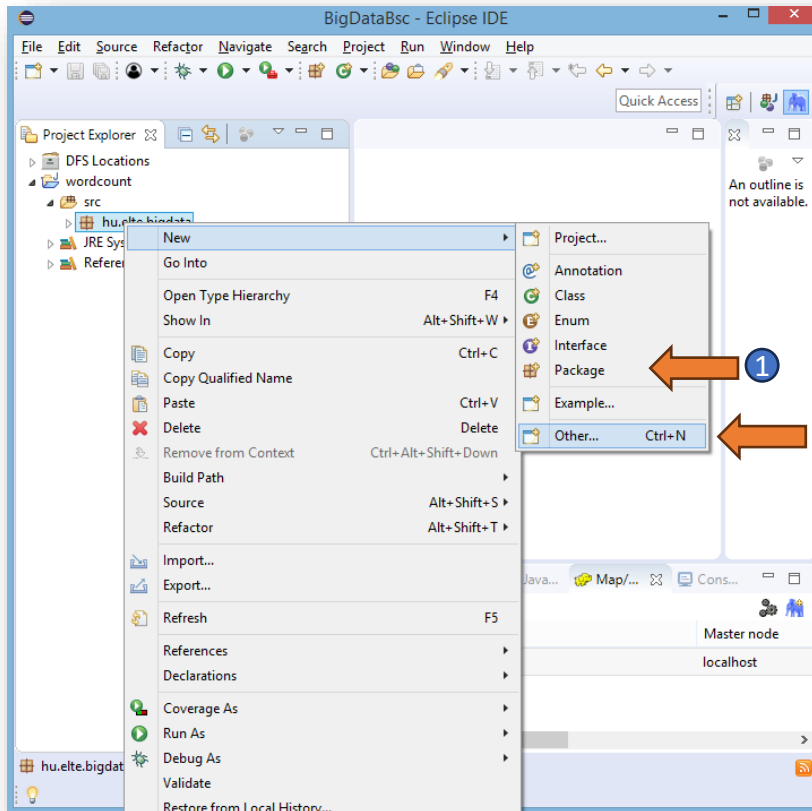
MapReduce projekt létrehozása Eclipseben

- 2. lépés: Hadoop elérési útvonal beállítása



MapReduce projekt létrehozása Eclipseben

- 3. lépés: Java package létrehozása
- 4. lépés: Mapper, Reducer és Driver létrehozása



Feladat: WordCount

