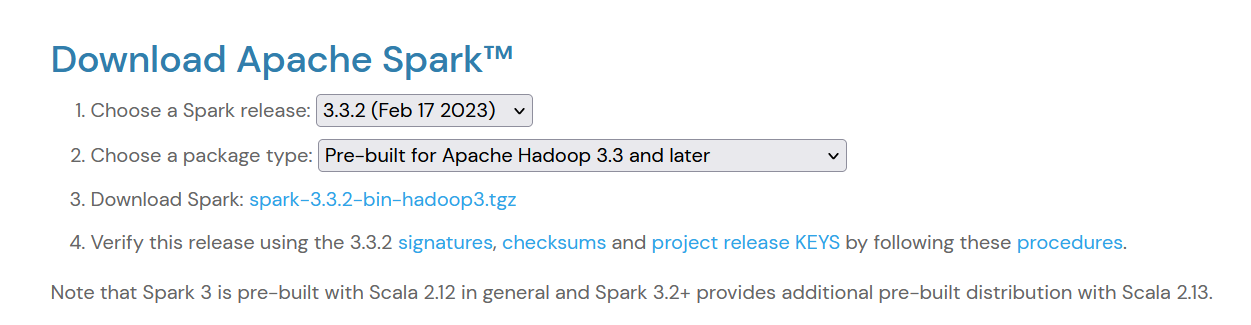
# Spark inštalácia

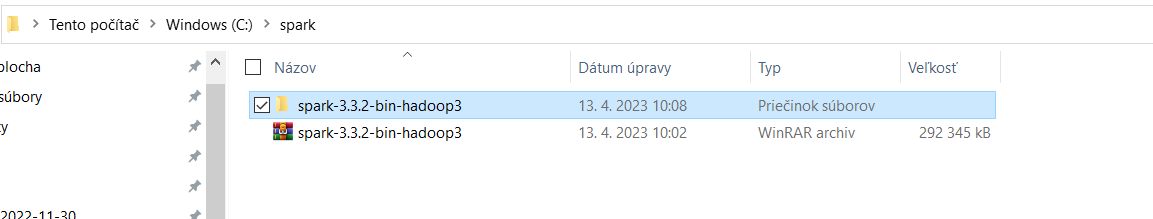
Treba mat nainštalovanú javu a python a hadoop

Stiahneme Apache Spark s oficiálnej stránky : <https://spark.apache.org/downloads.html>

My sme použili verziu pre hadoop 3.3.0



Na C disku vytvoríme priečinok spark a nakopíruje do neho stiahnutý Spark a extrahujeme ho



Ďalej je potrebné nastaviť užívateľské a systémové premenne prostredia

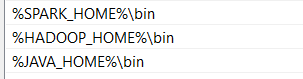
Používateľské premenne prostredia takto



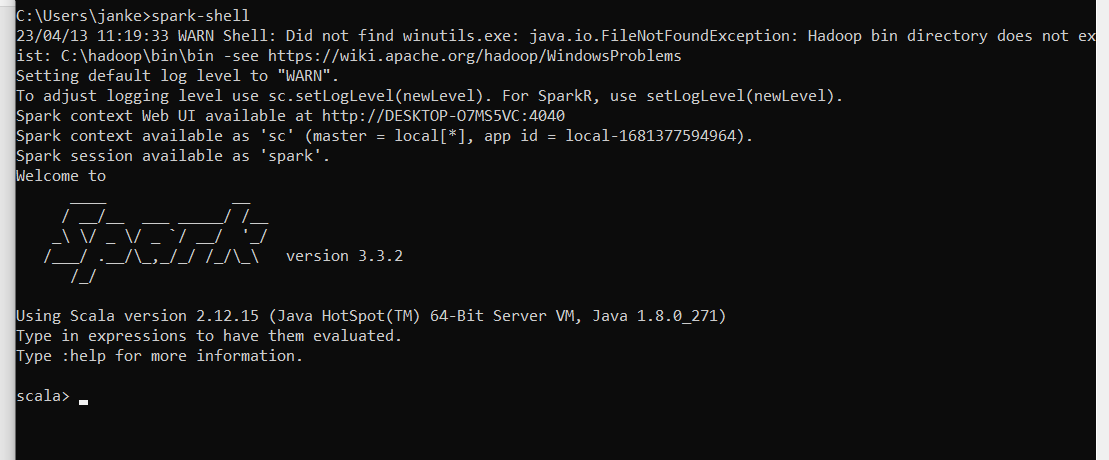




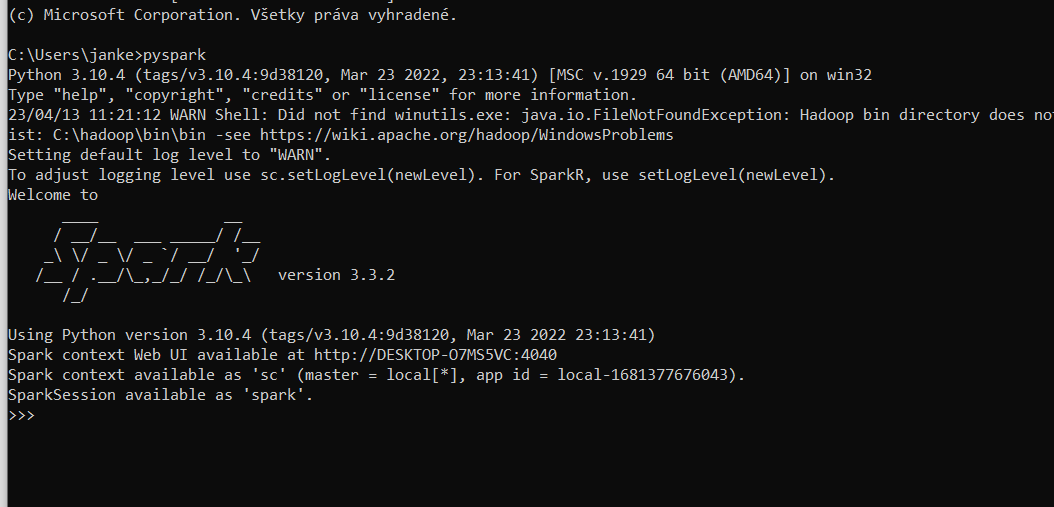
A do path v systémových premenných prostredia pridáme



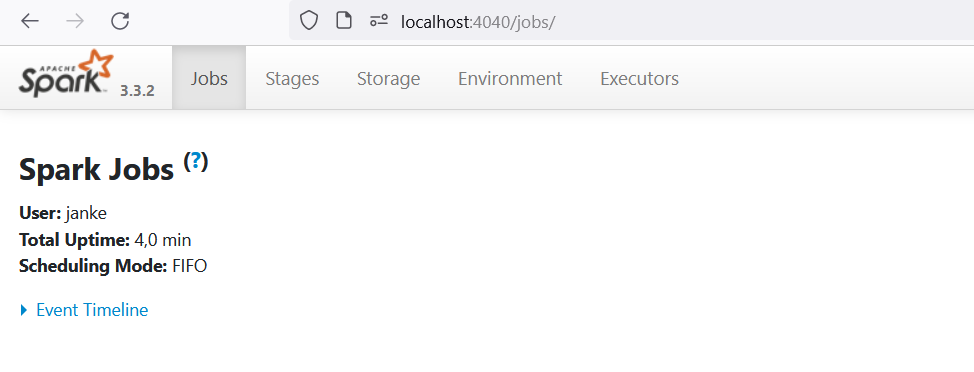
Do cmd dáme príkaz spark-shell a mal by sa nám spustiť spark



Príkazom pyspark pustime Pyspark



Taktiež môžeme navštíviť adresu <http://localhost:4040/jobs/> kde môžeme pristúpiť k webovému UI sparku



# Analáza log filu pomocou sparku