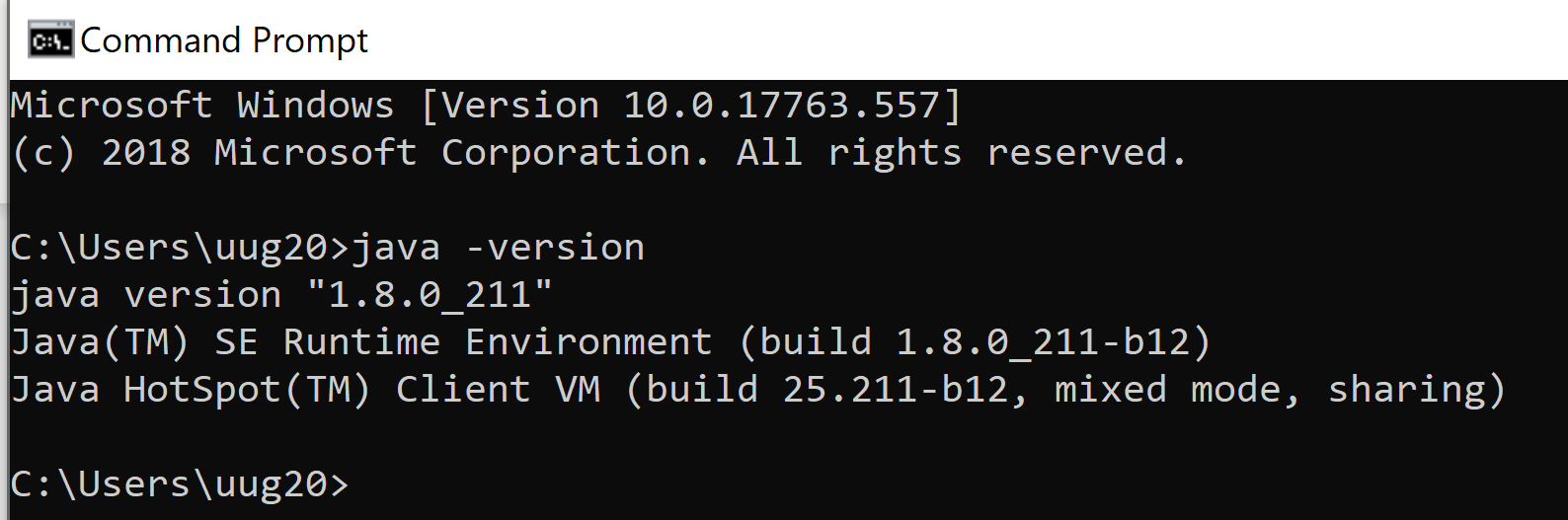
# Spark & Intellij Kurulumu

## Windows

### Adım

PySpark’ı çalıştırabilmemiz için bilgisayarlarımızda Java 7 ve Python 2.6 ya da daha ileri sürümlerinin yüklü olması gerekmektedir. İlk adım olarak Java’nın yüklü olup olmadığı kontrol edilmelidir. Bunun için;

Command Prompt’a girilir ve “java -version” komutu çalıştırılır. Eğer yüklü ise aşağıdaki çıktı görülür.



Eğer “‘java’ is not recognized as an internal or external command, operable program or batch file.” yazıyor ise Java’nın yüklenmesi gerekmektedir. Yüklü değil ise aşağıdaki adımlar izlenir. Yüklü ise 2. Adım’a geçiş yapabilirsiniz.

Java Yüklemesi

<https://www.java.com/en/download/> sitesinden Java indirilir.

Yüklemesi gerçekleştirilir.

Yükleme gerçekleştirildikten sonra tekrardan command prompt’a girilerek “java -version” komutu çalıştırılır ve başarılı olarak yüklendiği test edilir.

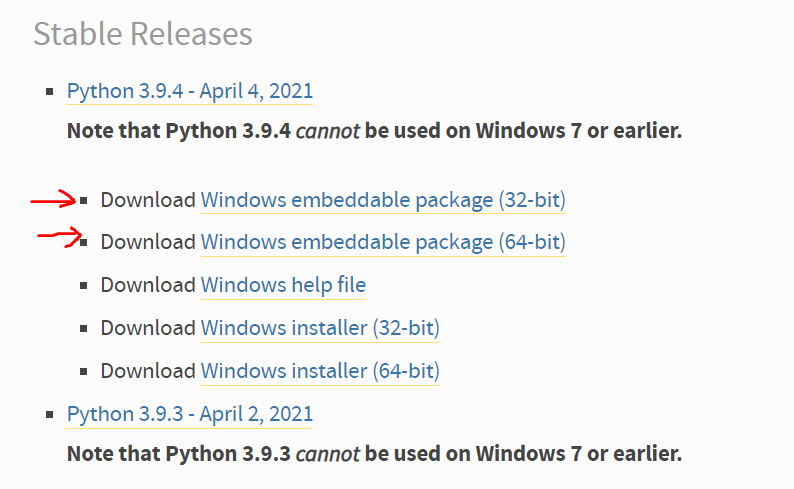
### Adım

Bu adımda Python yüklemesini sağlayacağız. İlk olarak Java’nın yüklü olup olmadığını kontrol ettiğimiz gibi Pyhthon’a da bakmamız gerekiyor. Tekrardan Command Prompt’a girilerek “python --version” komutu çalıştırılır. Yüklü olması durumunda Python’ın versiyon bilgileri ekrana gelecektir.

Eğer “ ‘python’ is not recognized as an internal or external command, operable program or batch file.” ya da benzeri bir mesaj alınıyor ise Python’ın yüklenmesi gerekmektedir. Yüklü değil ise aşağıdaki adımlar izlenir. Yüklü ise 3. Adım’a geçiş yapabilirsiniz.

Python Yüklemesi

<https://www.python.org/downloads/windows/> sitesine girilir.



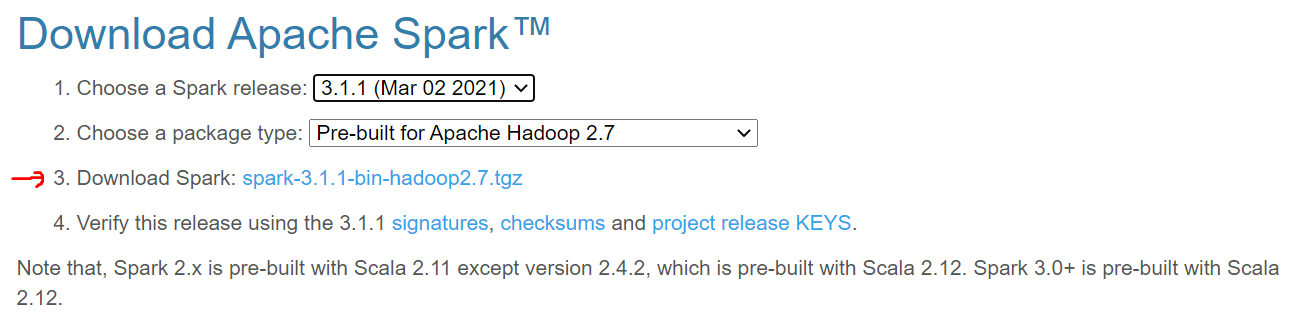
3.9.4 versiyonunun ilgili Windows paketi indirilir (Windows 32 bit ya da Windows 64 bit).

Yükleme yapılırken dikkat edilmesi gereken bir nokta vardır. Customize Python bölümünde “Add python.exe to Path” seçilmelidir. Aksi takdirde PySpark’ın bazı özellikleri kullanılamayabilir.

Yükleme gerçekleştirildikten sonra tekrardan command prompt’a girilerek “python --version” komutu çalıştırılır ve başarılı olarak yüklendiği test edilir.

### ADIM

Bu adımda Spark’ın Windows makineye yükleme adımları gerçekleştirilir. <http://spark.apache.org/downloads.html> adresine gidilerek ekran görüntüsündeki versiyonları ile Spark indirilir.



İndirildikten sonra dosya zip’ten çıkartılır. Daha önceki yaptığımız gibi burada bir yükleme söz konusu değildir.

Zip’ten çıkartmış olduğumuz dosyalar “spark-3.1.1-bin-hadoop2.7” adlı bir klasöre çıkartılacaktır. Bu dosyamızı C ya da D sürücüsündeki bir path’e kopyalanır. Örnek olarak “C:\spark\spark-3.1.1-bin-hadoop2.7” diyelim.

Test etmek için Command Prompt’tan ilgili klasöre gidilir. CMD’de bir klasöre gitme komutu cd’dir. “cd C:\spark\spark-3.1.1-bin-hadoop2.7\bin” dersek ilgili path’e gitmiş oluruz. Path’te olduğumuzdan emin olduktan sonra “pyspark” komutu çalıştırılır. Bu komut PySpark Shell’e giriş yapmamızı sağlayacaktır.

PySpark Shell’de kullanılabilecek komutlar vardır. Örneğin “sc.version” yazıp Enter’a basarsak bize Spark’ın versiyonunu söyleyecektir. Çıkış yapmak için “exit()” yazılır.

Spark’ın kendisi Windows’ta çalışmak üzere dizayn edilmemiştir. Stabil çalışabilmek için winutils.exe adında bir program kullanır. Fakat default’unda winutils.exe’nin path’i atanmaz. Bu yüzden “ ERROR Shell: Failed to locate the winutils binary in the hadoop binary path java.io.IOException: Could not locate executable null\bin\winutils.exe in the Hadoop binaries.” hatası alınabilir.

Bu hata PySpark Shell’in çalışmasına engel değildir. Fakat bazı fonksiyonları çalışmayacaktır (örneğin spark-submit).

### ADIM

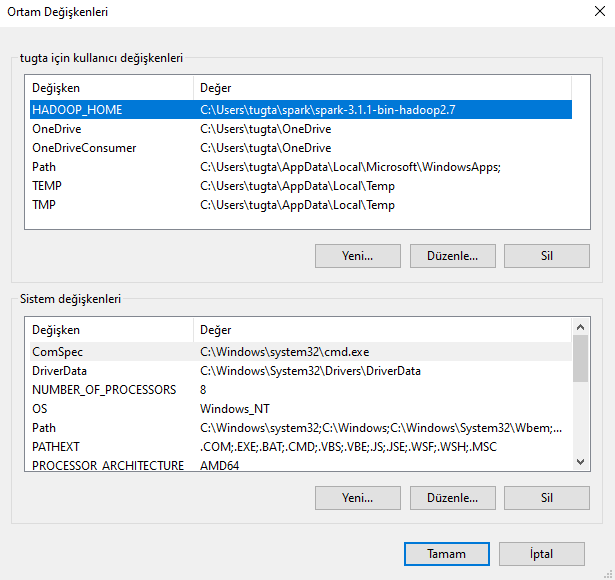
Winutils Yükleme

Winutils.exe’yi yükleme adımları aşağıdaki gibidir.

“<http://github.com/steveloughran/winutils>” adresinden Hadoop versiyonuna uyumlu olan Winutils indirilir. Biz bu durumda 2.7.1’i indireceğiz. Bu yüzden “<https://github.com/steveloughran/winutils/tree/master/hadoop-2.7.1/bin>” altında bulunan winutils.exe indirilir. İndirilen dosya “C:\spark\spark-3.1.1-bin-hadoop2.7\bin” altına taşınır.

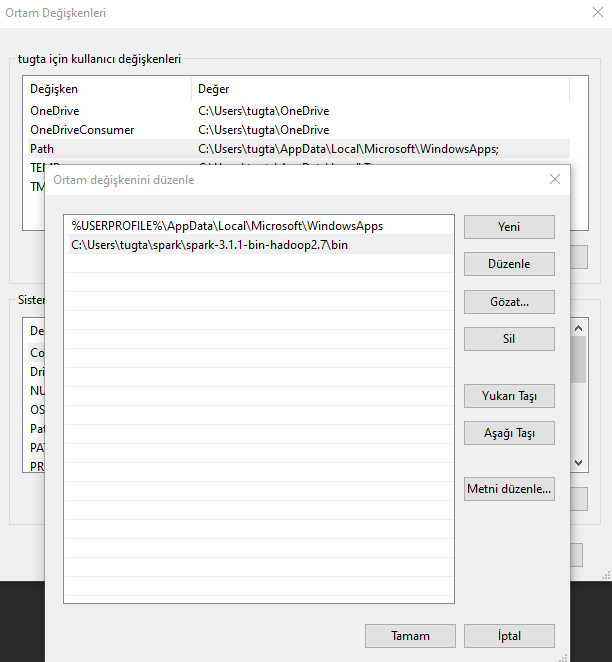
Bu aşamada Hadoop’un kullanabilmesi için bir system environment tanımlamamız gerekmektedir.

Bu pathi user environment kısmında SPARK\_HOME ve HADOOP\_HOME şeklinde eklememiz gerekiyor.



User Enviroment içinde bulunan Path değişkenini de

C:\spark\spark-3.1.1-bin-hadoop2.7\bin pathini ekliyoruz.



Artık “pyspark” komutunu çalıştırdığımız zaman winutils ile alakalı hataların geçmiş olduğunu göreceksiniz Ayrıca pyspark enviroment path’e eklendiği için komut satırına doğrudan pyspark yazarak da spark konsolu başlatabiliriz .

## Mesajları Temizleme

PySpark Shell’e girildiği zaman ekrana çokça mesaj basılacaktır. Bu mesajların tipleri INFO, ERROR, WARN’dur. Gösterilen mesajları azaltmak için gerekli işlemler aşağıdaki gibidir.

SPARK\_HOME\conf altında bulunan ”log4j.properties.template” dosyası ”log4j.properties” adıyla kopyalanır.

Notepad ile yeni oluşturduğumuz dosya düzenlenir.

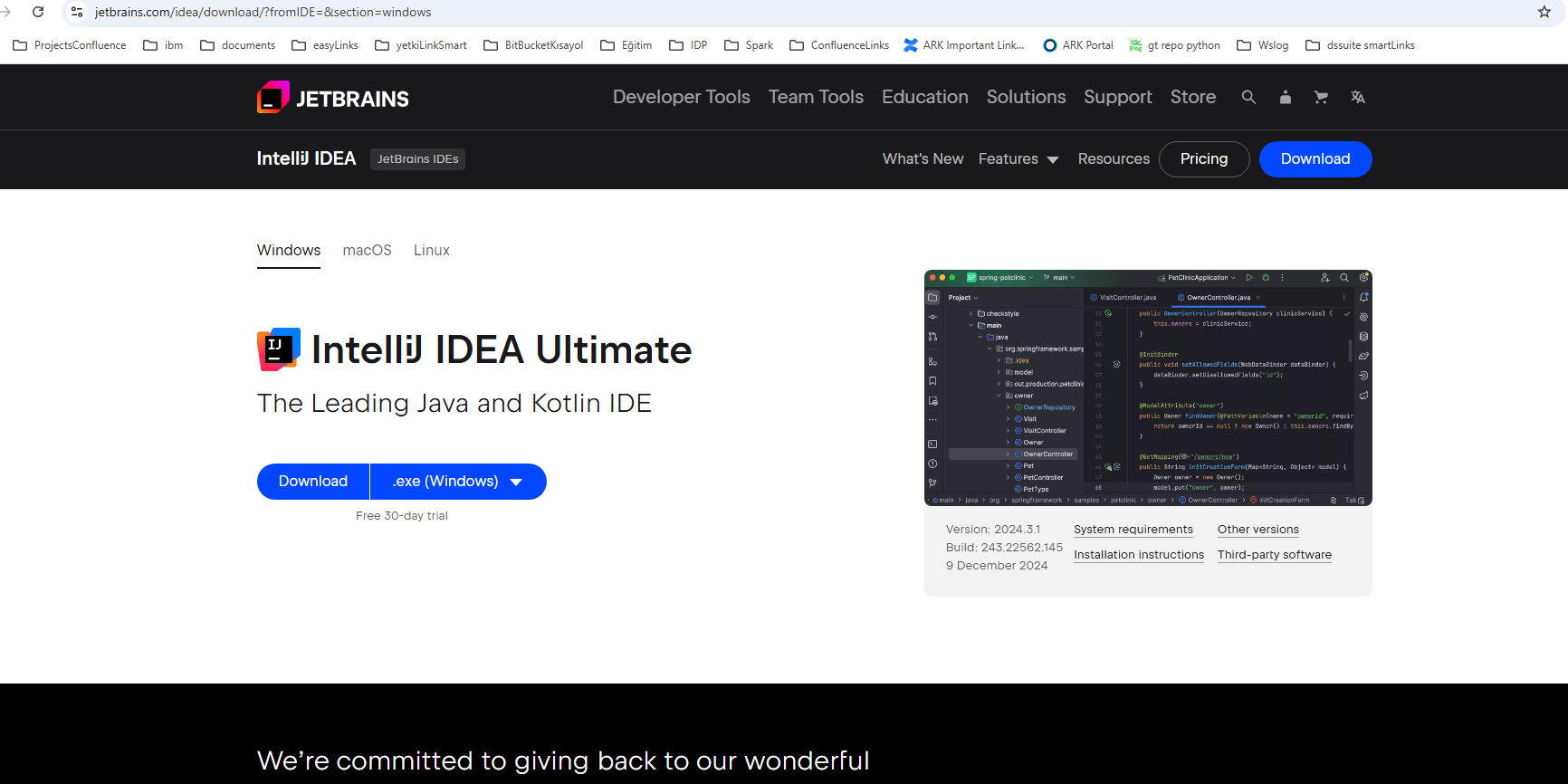
log4j.rootCategory satırındaki değer yerine “WARN” yazılır ve kaydedilir. Bu şekilde shell’i tekrar açtığımız zaman çıkan INFO mesajları temizlenmiş olur.

## Intellij Kurulumu

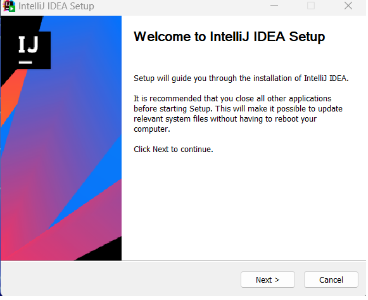
Trial versiyon yerine öğrencilere sunulan abonelikten yararlanabilirsiniz.

Ihttps://www.jetbrains.com/idea/download/?fromIDE=&section=Windows

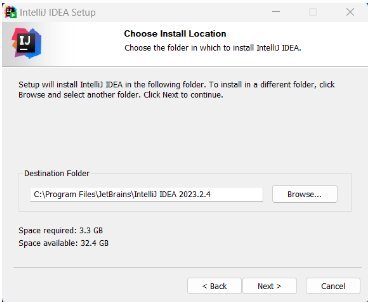
Adresine gidiyoruz.



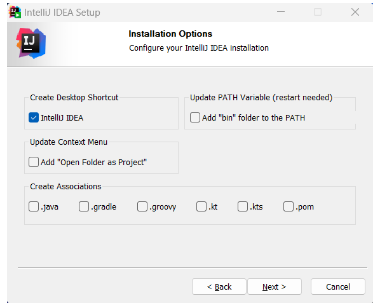
.exe dosyasını açın ve aşağıdaki ekranı göreceksiniz, ardından İleri'ye tıklayın:



Kurulum yapılacak konumu seçin, konumu Gözat'a tıklayarak değiştirebilirsiniz . Aksi takdide aynı kalır. Bundan sonra İleri'ye tıklayın.



Masaüstü kısayolu oluşturma gibi uygun seçenekleri işaretleyin. Daha sonra İleri'ye tıklayın.



İleri'ye tıklayın, ardından IntelliJ kurulumu başlayacaktır. Aşağıdaki ekran görünecektir:

