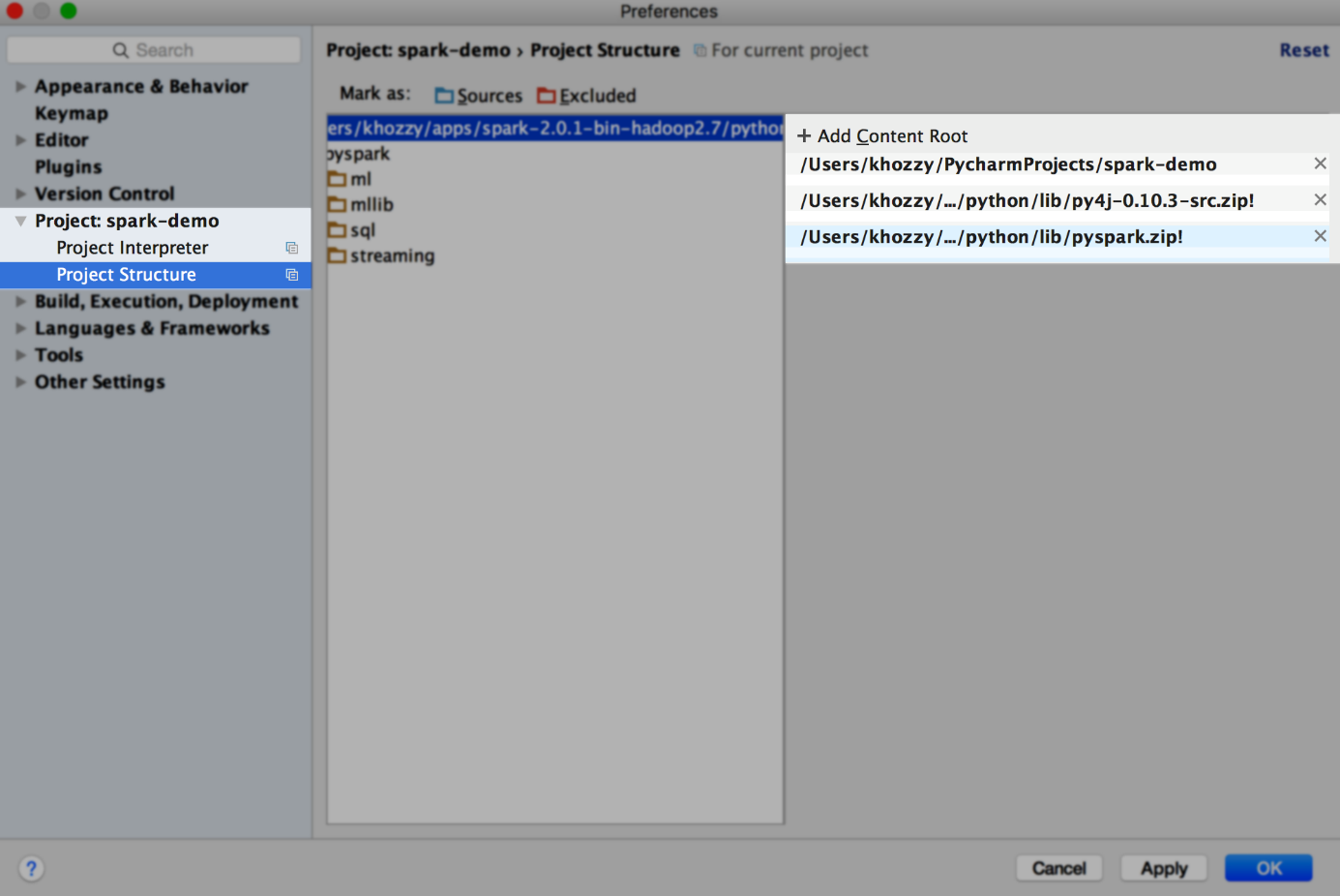
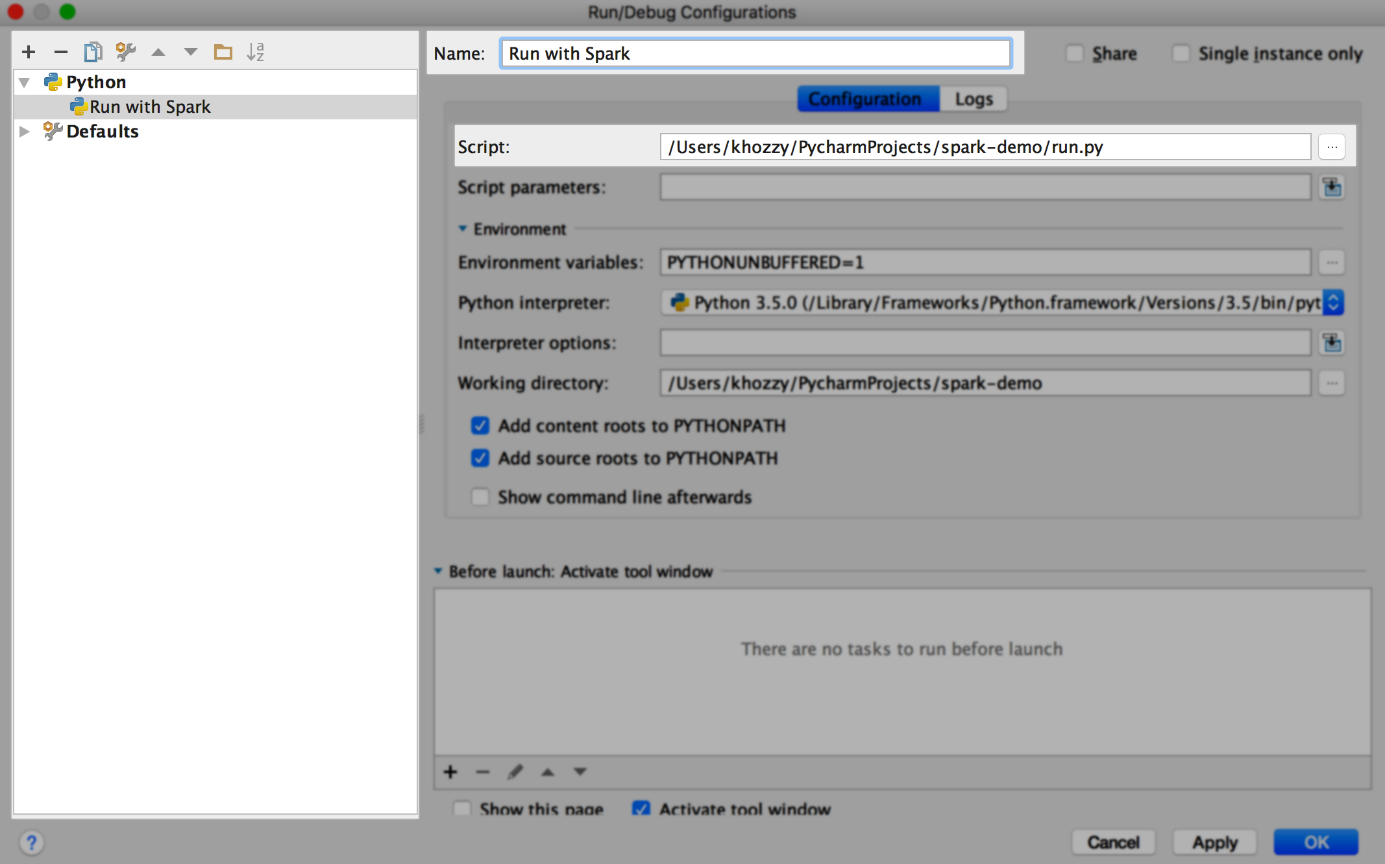
Apache Spark Coding With Python

Pycharm kullanarak bir apache spark projesi oluşturuyor isek öncelikle spark’ı pycharm’a entegre etmemiz gerekiyor.Onu da şu şekilde yapıyoruz.

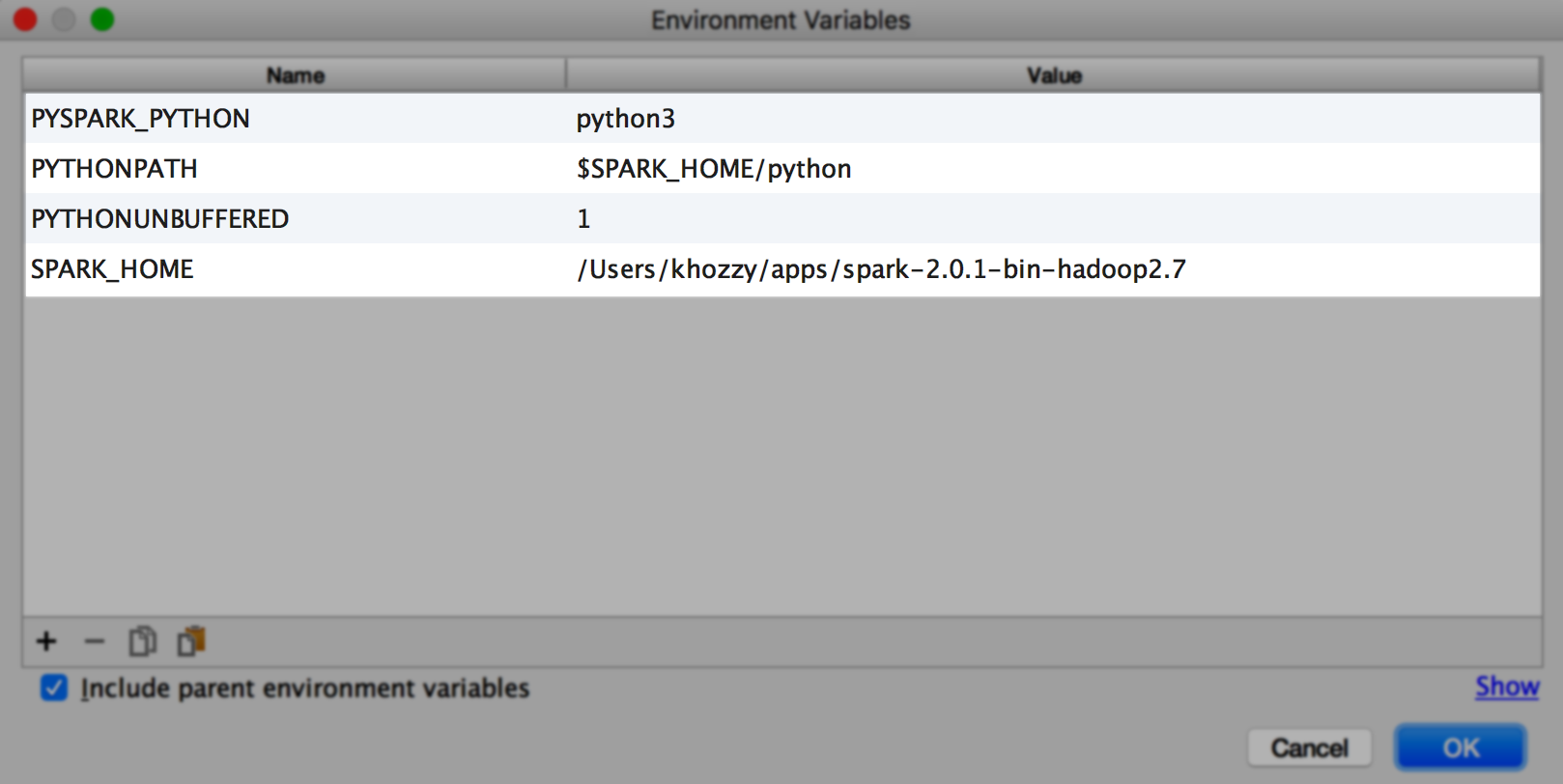
* **Add required libraries** : PyCharm → Preferences ... → Project spark-demo → Project Structure → Add Content Root. Select all ZIP files from $SPARK\_HOME/python/lib. Apply changes.



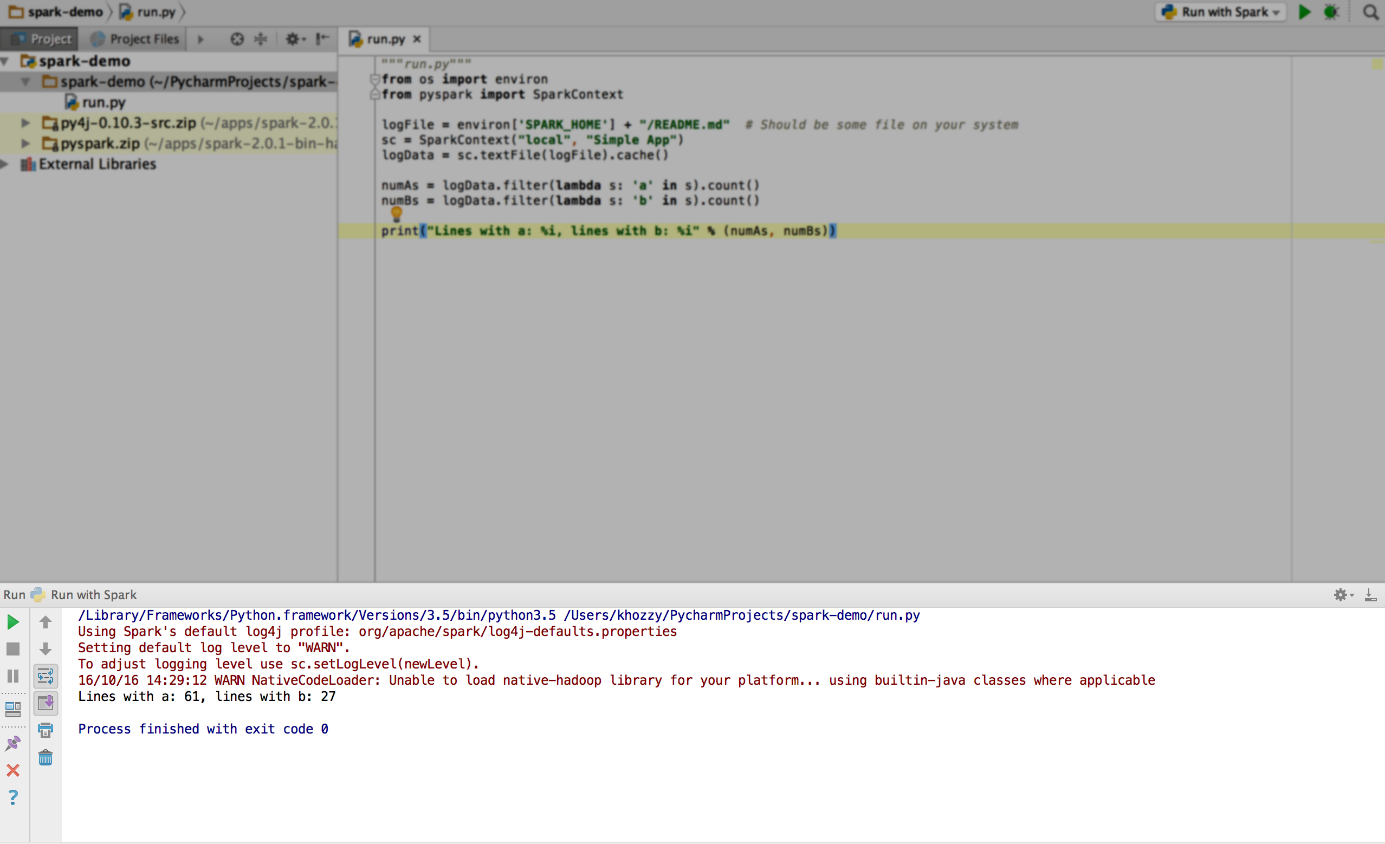
* **Create a new run configuration** : Go into Run → Edit Configurations → + → Python. Name it "Run with Spark" and select the previously created file as the script to be executed.



* **Add environment variables**: Inside created configuration add the corresponding environment variables. Save all changes.



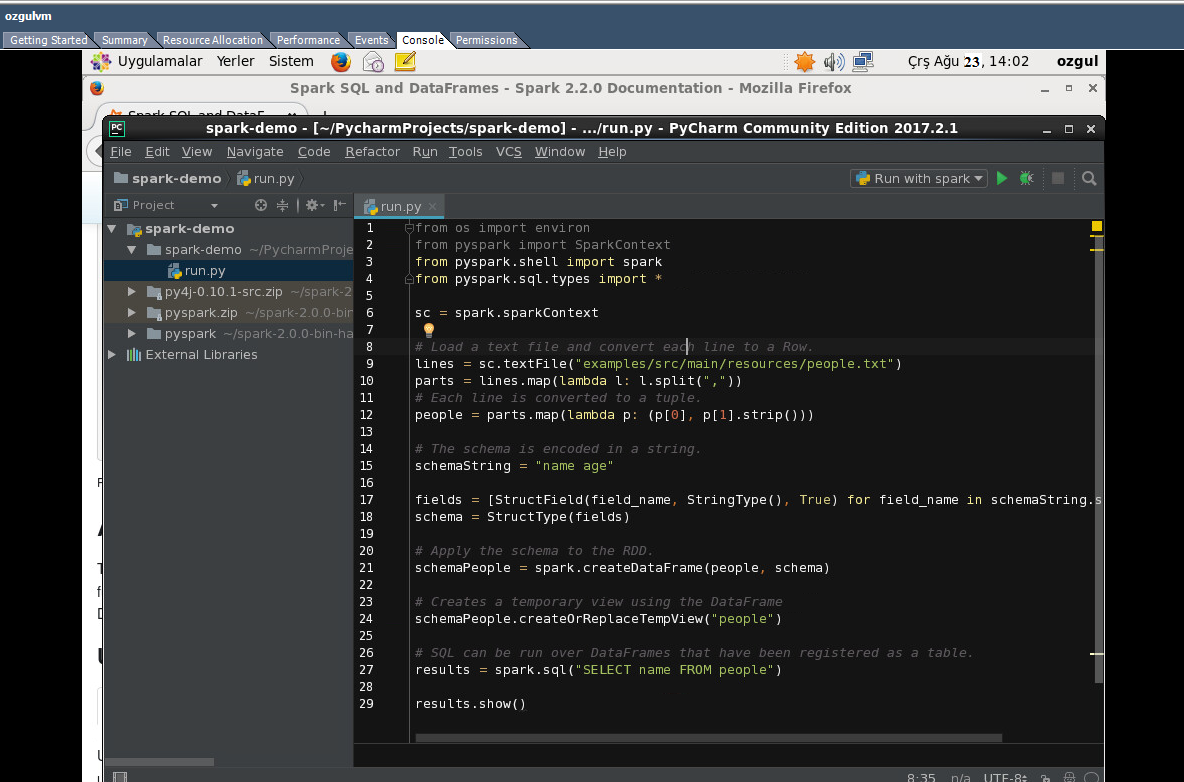
* **Run the script:** Run → Run 'Run with Spark'. You should see that the script is executed properly within Spark context.



Bu şekilde entegre ettikten sonra projemizin kodlarını yazmaya başlayabiliriz.

**Programmatically Specifying the Schema**

Bu kodda yapılandan bahsetmek gerekirse : Öncelikle spark’a bağlanılıyor ve içinde hazır olarak gelen people.txt dosyasına ulaşmak hedefleniyor.Oraya ulaştıktan sonra hedefimiz people dosyasının içindeki insanların isimleri şema halinde göstermek.Gerekli sql komutları yardımıyla da kodumuzu yazıyoruz.



Program çıktısı olarak karşımıza bu gelmektedir.

*# +-------+*

*# | name|*

*# +-------+*

*# |Michael|*

*# | Andy|*

*# | Justin|*

*# +-------+*