**CAPAIAN PEMBELAJARAN II**

**BIG DATA**

Nama : Elisabet Maria Pereira da Costa Leite

Nim : 04317038

**FAKULTAS ILMU KOMPUTER**

**UNIVERSITAS NAROTAMA**

**SURABAYA**

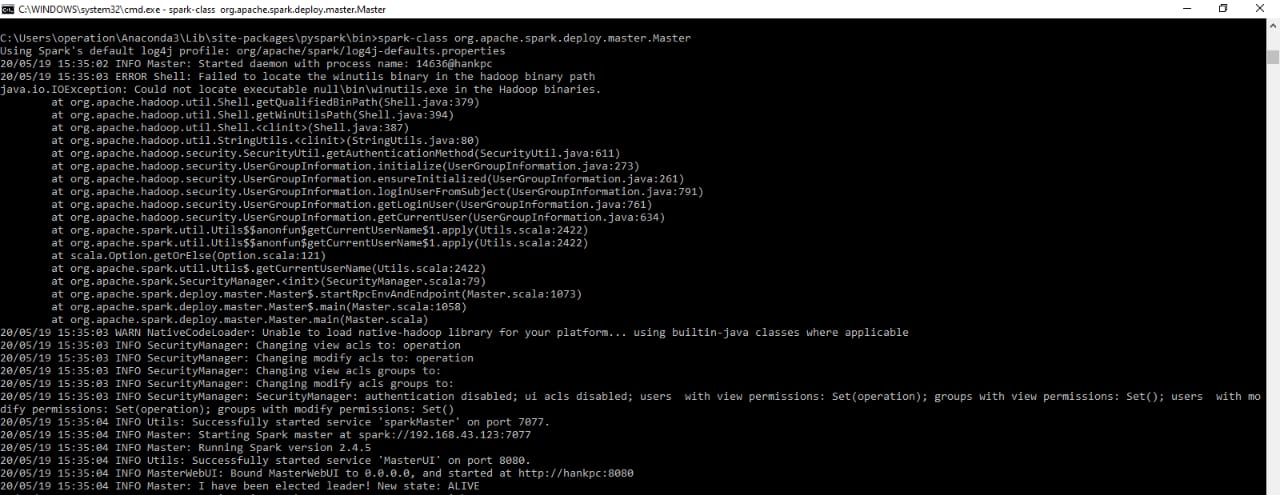
**2020**

**Komputasi paralel menggunakan Pyspark melalui 2 laptop**

1. Langkah pertama yang harus dilakukan adalah membuat master klaster pada laptop , sebelum membuat master klaster pastikan dua laptop tersebut sudah terhubung pada jaringan yang sama dan menempatkan dataset yang mau digunakan ke direktori ata Path yang sama seperti : :  
   \Users\lella leite\Desktop\ Movie Recommender with PySpark\movies.csv\
2. Membuat master klaster pada laptop A.

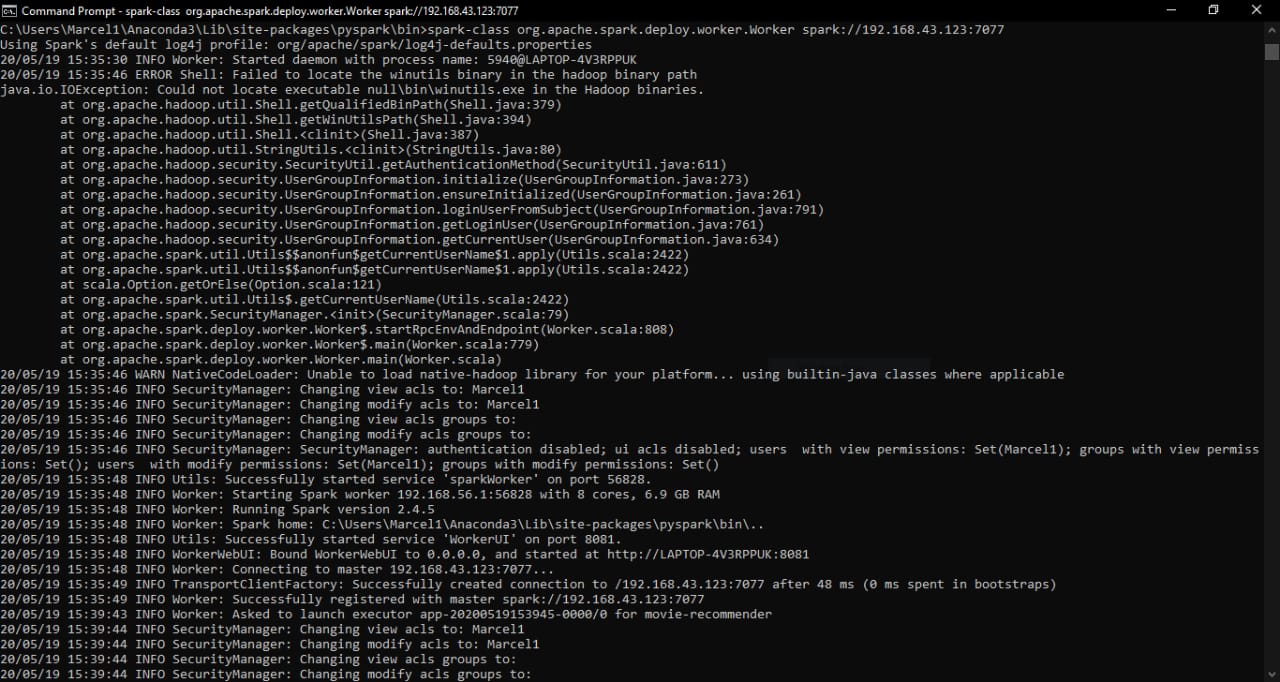
* Membuka terminal, kemudian masuk ke direkori :

cd \Users\user\anaconda3\Lib\site-packages\pyspark\bin. Kemudian setting master klaster dengan perintah sebagai berikut **spark-class org.apache.spark.deploy.master.Master.** Setelah mengetikan perintah tersebut maka kita akan mendapatkan IP address dan port dari master klaster yang akan digunakan untuk menambahkan komputer slave.

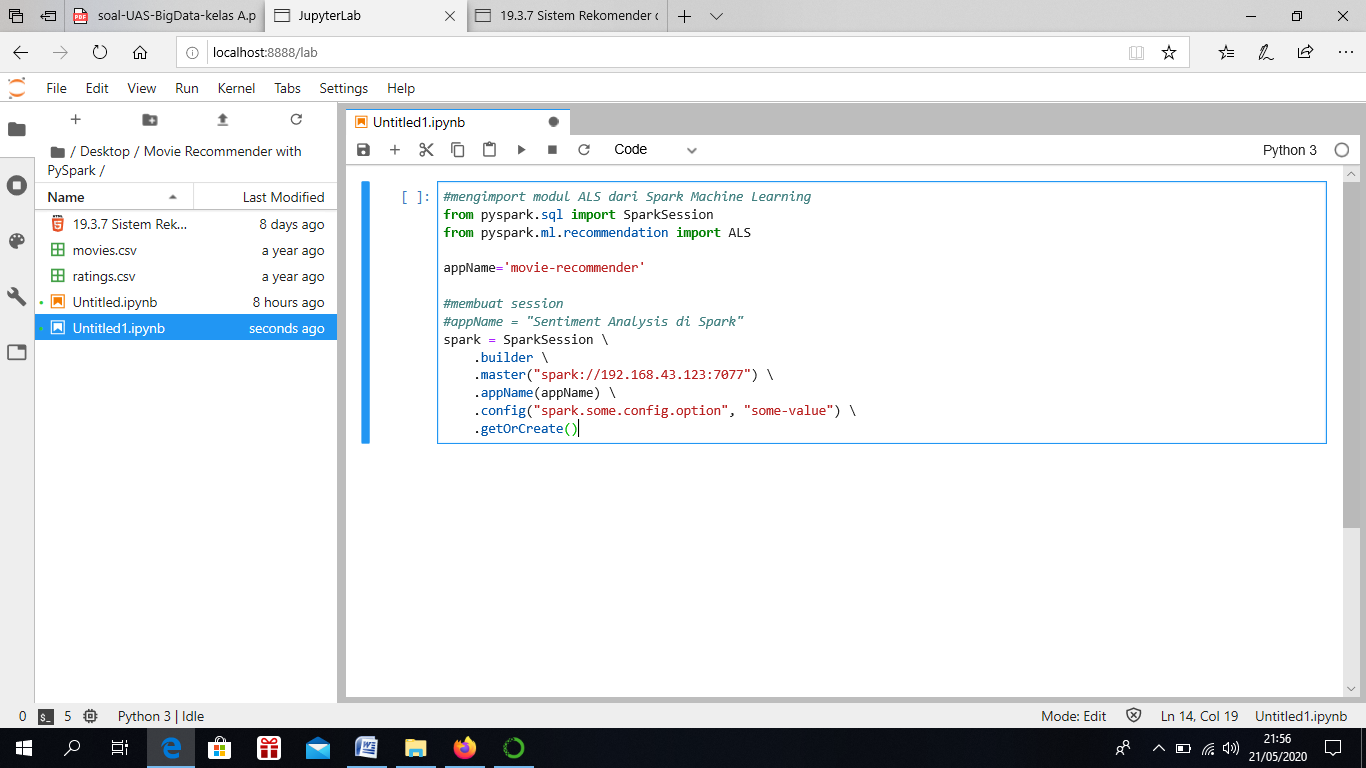


1. Selanjutnya adalah menambahkan slave pada laptop B

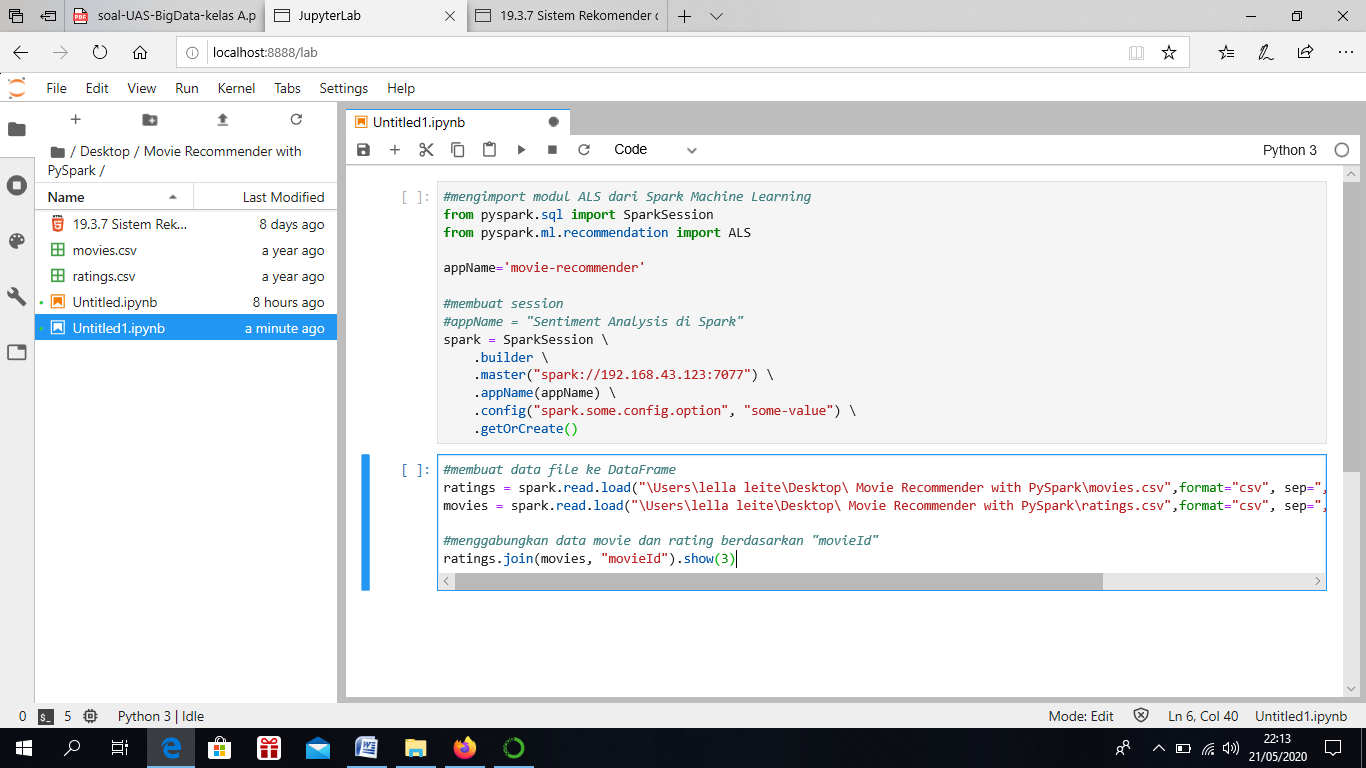
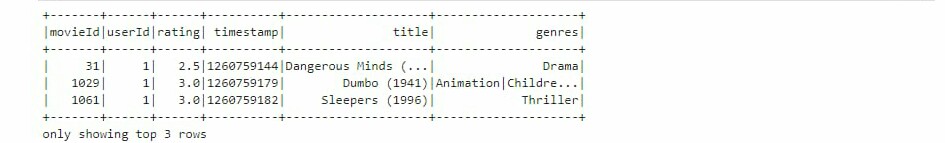
* Membuka terminal kemudian mengetikan perintah cd \Users\<nama\_user>\anaconda3\Lib\site-packages\pyspark\bin untuk masu ke direktori.
* Setelah berhasil masuk ke direktori selanjutnya adalah melakukan setting slave pomputer dengan perintah **spark-class org.apache.spark.deploy.worker.Worker spark://192.168.43.123:7077**



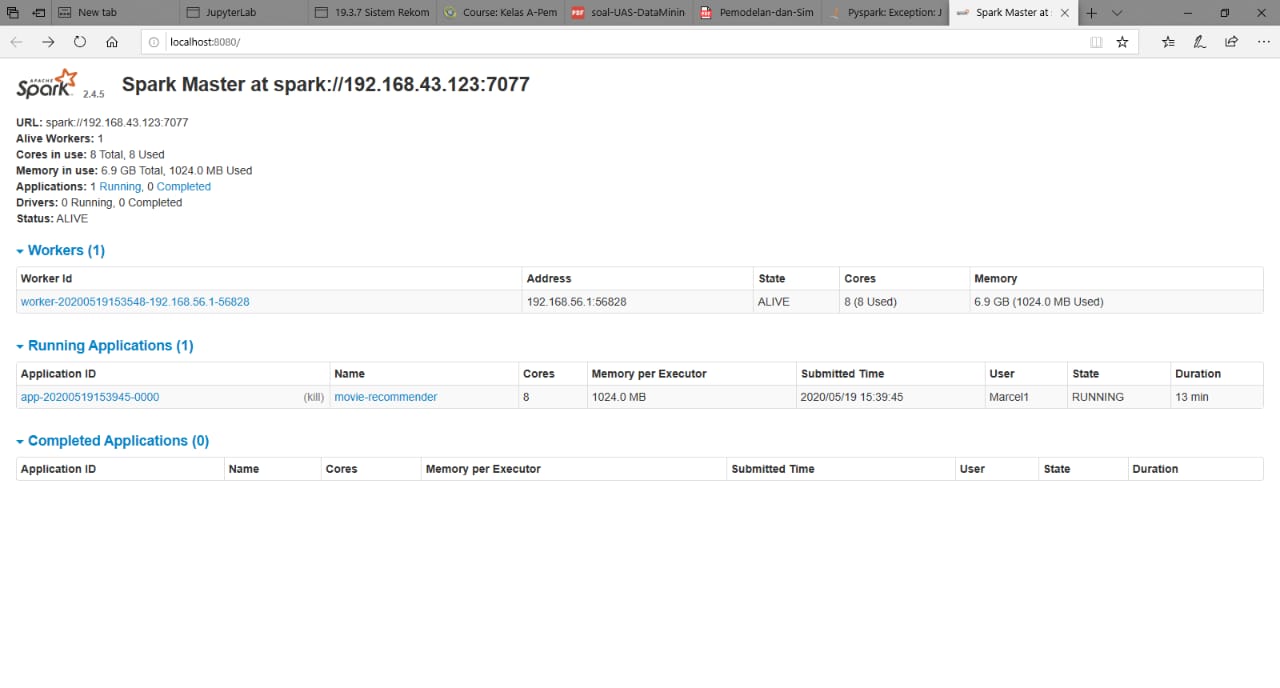
1. Selanjutnya adalah mengecek sistem monitoring pada master klaster spark. Untuk melihatnya dapat di akses pada <http://localhost:8080/>. Setelah berhasil menambahkan worker pada master klaster langkah berikutnya adalah menjalankan pyspark pada Jupiter lab.
2. Langkah pertama yang akan dilakukan pada master klaster atau pada laptop A adalah mengimport modul dan membuat session yang akan digunakan. Perlu diperhatikan untuk masukan IP dan Port pada master sesuai dengan yang sudah didaptakan.



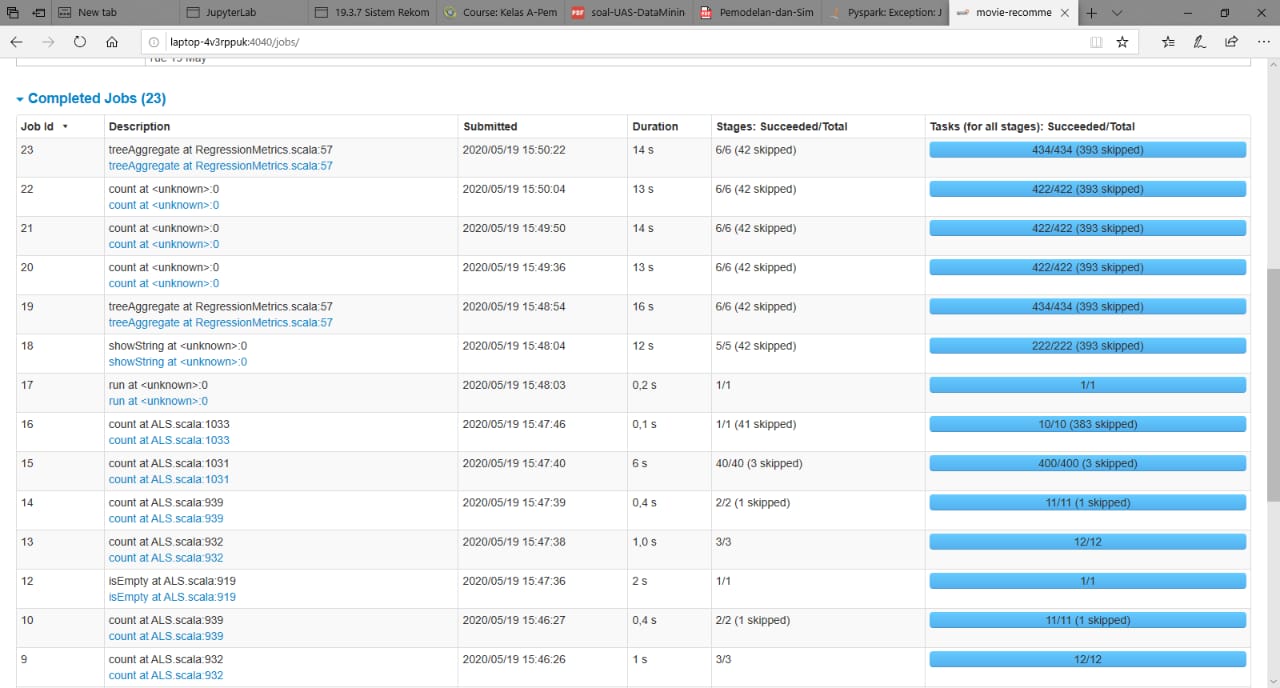
1. Selanjutnya adalah memuat file atau dataset yang akan digunakan.



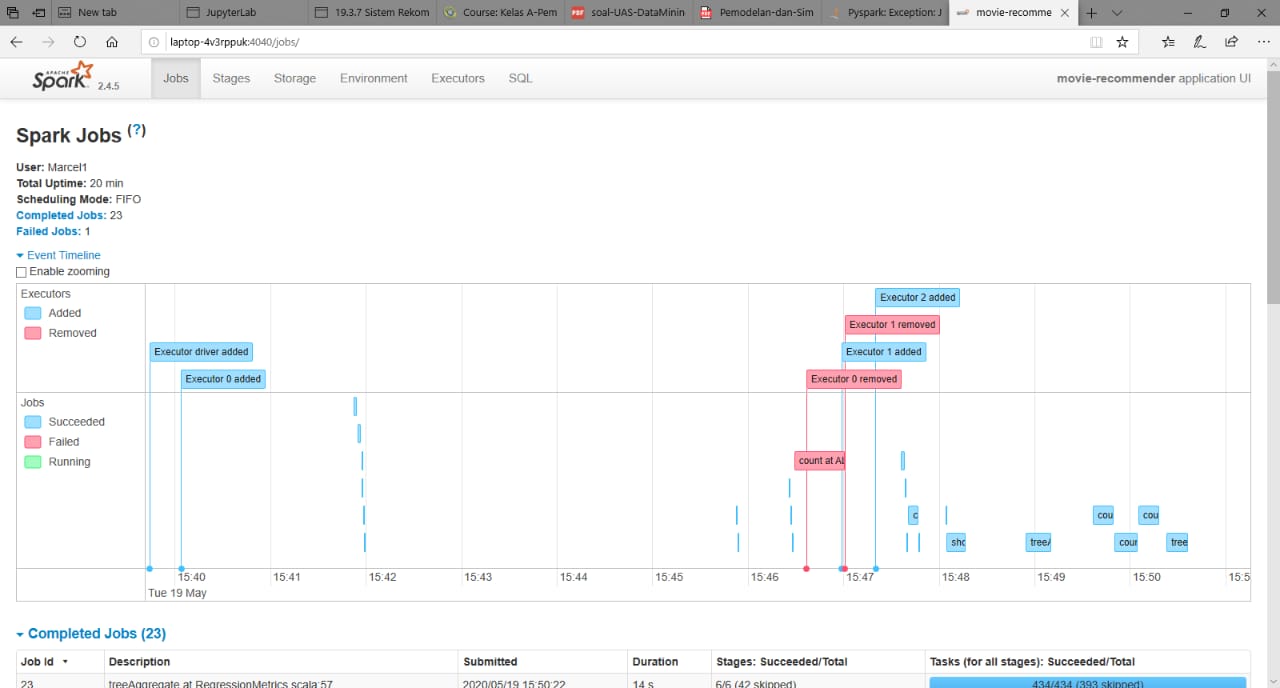
Setelah itu jalankan semua cells, selanjutnya adalah periksa kembali sistem monitoring pada spark. Bisa dilihat pada gambar dibawah telah tercatat 1 aplikasi yang sudah berhasil di submit dengan nama movie-recommender.



Kemudian periksa kode pysparknya apakah tereksekusi dengan baik pada progress job App, untuk mengecek bisa klik pada bagian nama App.



Bisa dilihat juga bar progress timeline, disini kita bisa melihat lamanya waktu eksekusi program *movie-recommender.*

**

Terakhir kita evaluasi seberapa akurat sistem recomender kita.

