**Nama : Muhamad Fadil**

**NPM : 1706042812**

**Tugas : Tugas akhir Big Data – Pengembangan Arsitektur Pengolahan Big Data**

**Gempa Bumi Indonesia Tahun 1965 - 2016**

1. **Pendahuluan**

Data ini merupakan data yang menginformasikan mengenai data gempa bumi yang terjadi di dunia sejak tahun 1965 sampai 2016. Dari data tersebut diambil dan di sortir gempa bumi yang terjadi diindonesia. Dari data gempa bumi ini juga digunakan untuk mendapatkan titik gempa bumi terbesar, negara yang mendapat jumlah gempa terbanyak, jarang terjadi gempa bahkan negara yang tidak pernah terjadi gempa.

1. **Latar belakang**

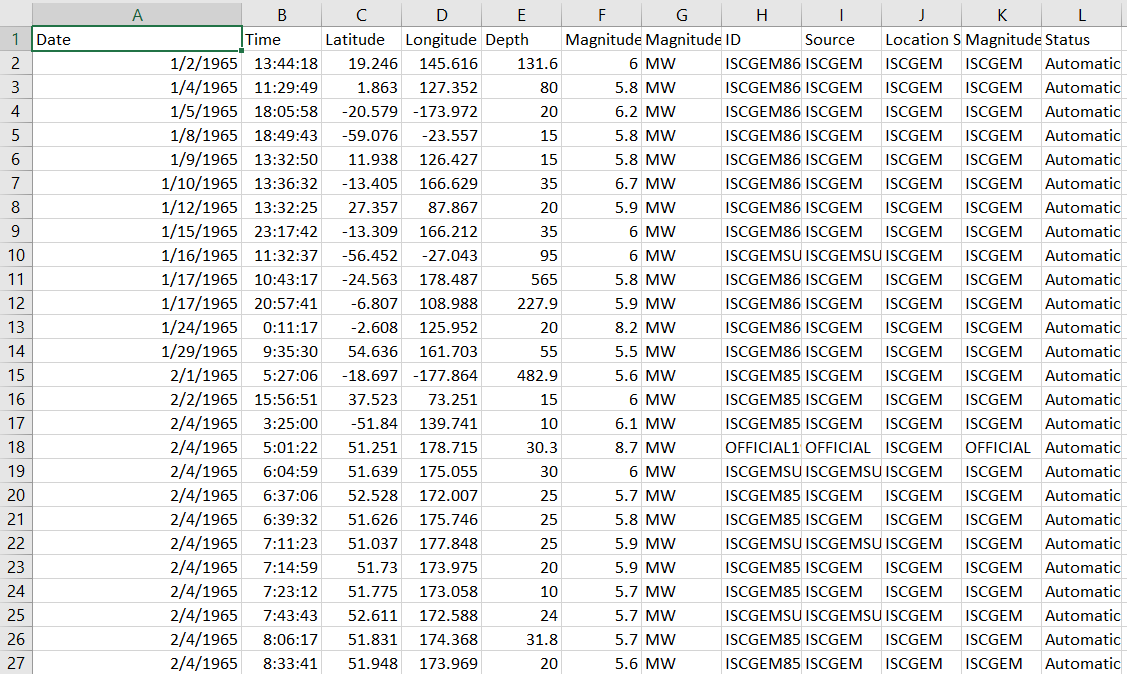
Diambilnya topik dengan data-data gempa bumi dari seluruh negara di dunia ini untuk mengetahui letak persisnya lokasi yang telah terjadi gempa. Selain itu, alasan lainnya adalah untuk mengetahui lokasi dimana gempa itu sering terjadi.

Hal ini penting untuk kegiatan kehidupan di bumi. Dari lokasi dan seringnya terjadi gempa bumi pada suatu tempat, suatu kehidupan pada lokasi tersebut akan menyesuaikan kehidupan dengan lingkungan alamnya yang sering terjadi gempa bumi. Suatu wilayah yang sering terjadi gempa bumi akan mempengaruhi kehidupan masyarakat itu sendiri seperti desain arsitektur bangunan yang tidak seperti daerah yang jarang terjadi gempa bumi. Biasanya daerah yang sering terjadi gempa bumi bangunannya akan didesain agar tidak mudah roboh ketika terjadi gempa bumi, seperti jepang dan lainnya. Selain itu, daerah yang sering terjadi gempa bumi akan membuat cara budaya dan lainnya berubah.

1. **Rancangan teknologi big data**

Data diolah dengan menggunakan Apache Spark yang diintergrasikan dengan jupyter yang dikoneksikan dengan AWS. Apache spark merupakan sebuah engine (perangkat lunak) analisis terpadu super cepat untuk memproses data dalam skala besar, meliputi Big Data dan mechine learning. Secara lebih detailnya, Apache spark dapat didefinisikan sebagai perangkat lunak untuk memproses data dalam skala besar secara in-memory, dilengkapi dengan API pengembangan yang elegan dan ekspresif guna memudahkan para pekerja data dalam mengeksekusi pekerjaan-pekerjaan yang membutuhkan perulangan akses yang cepat terhadap data yang diproses, pada proyek ini adalah data yang diproses dengan tambahan SQL, secara efisien.

Gambaran data yang diambil seperti pada gambar 1.

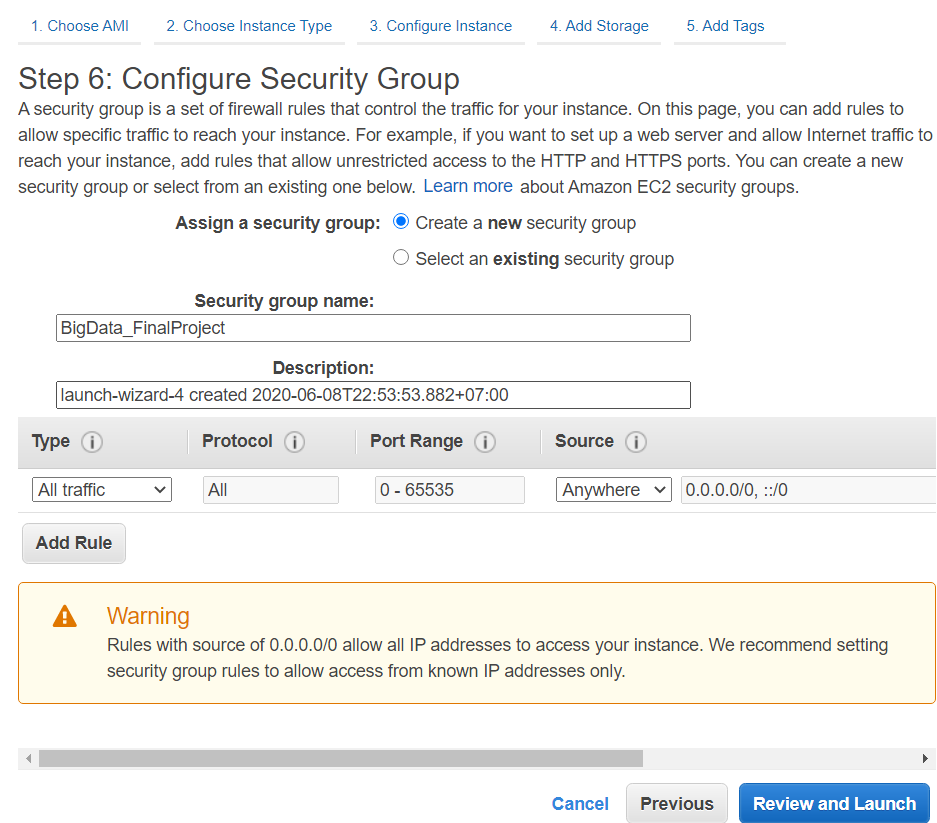


Gambar 1. Gambar data dari seluruh kejadian gempa bumi

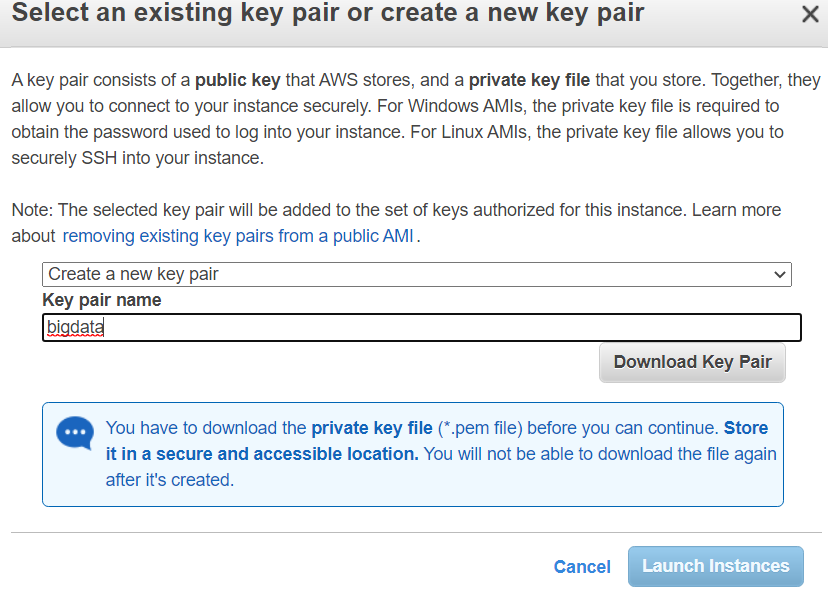
1. **Implementasi**

**Bagian 1. Membuat ec2 pada aws**

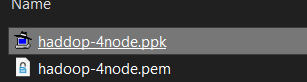
* Tahap1



* Tahap 2

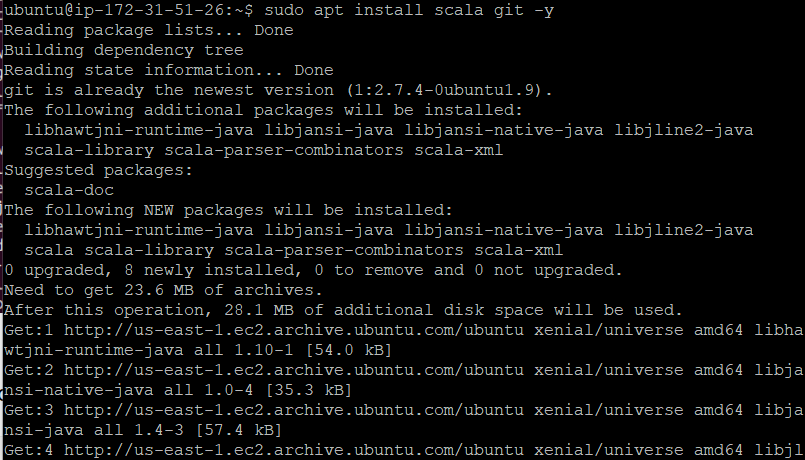


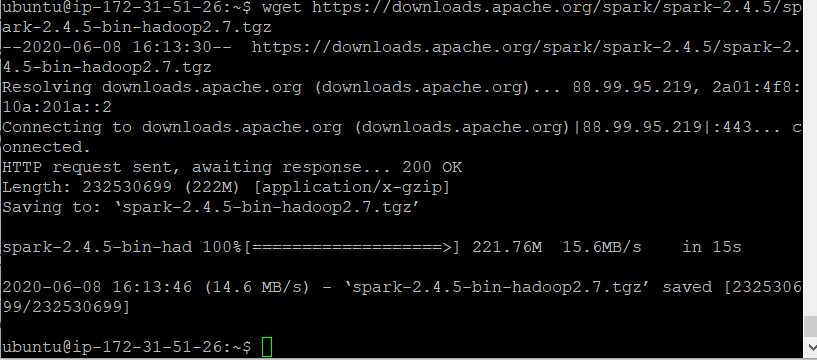
* Tahap 3









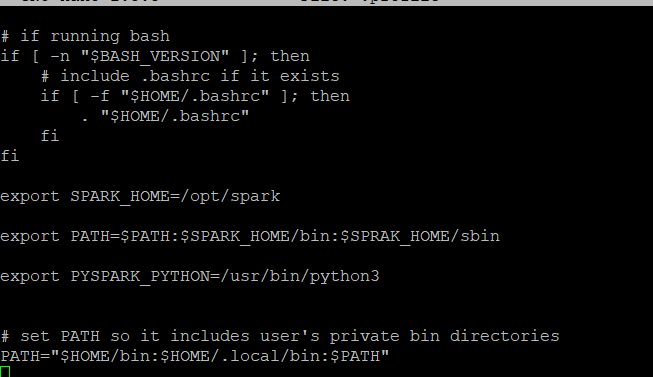






* Tahap 4 menambahkan perintah

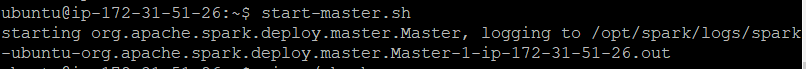




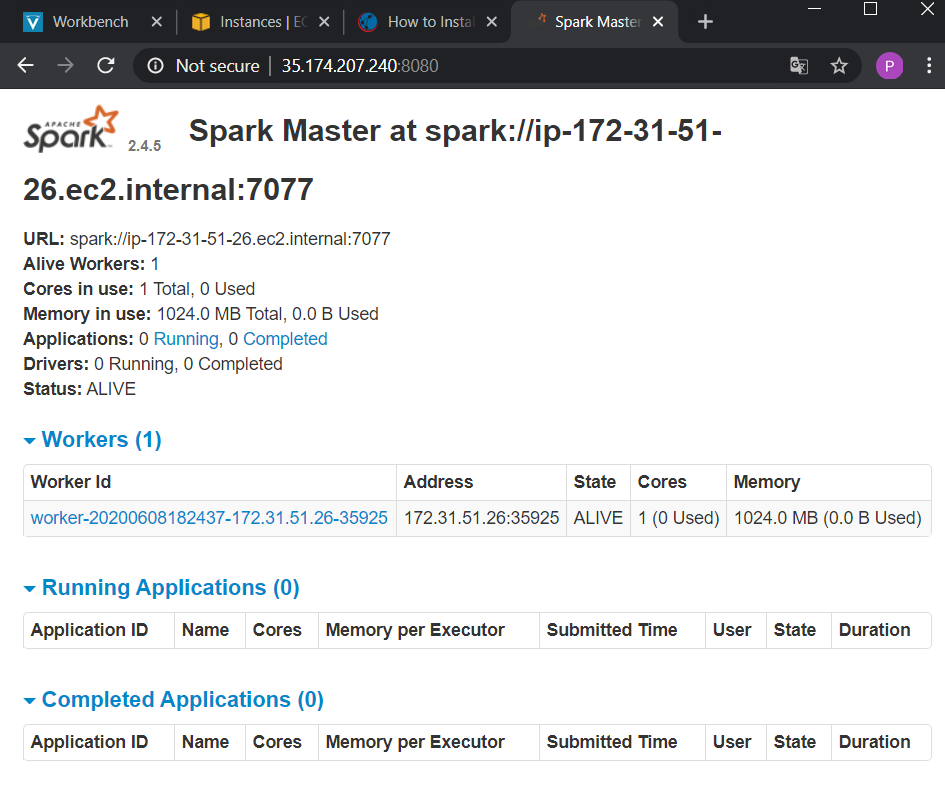


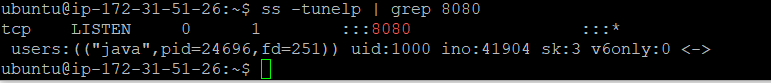


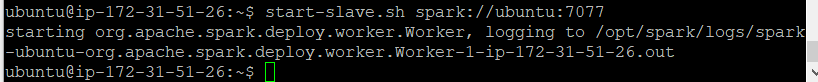




Hasil apache spark pada alamat 35.174.207.240:8080



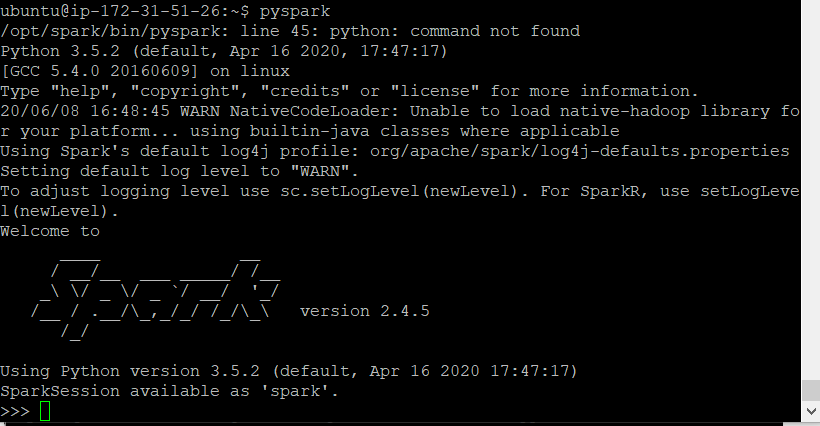




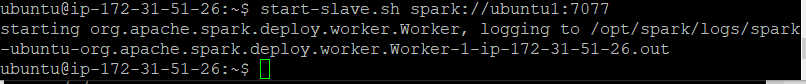
Pengecekan spark-shell



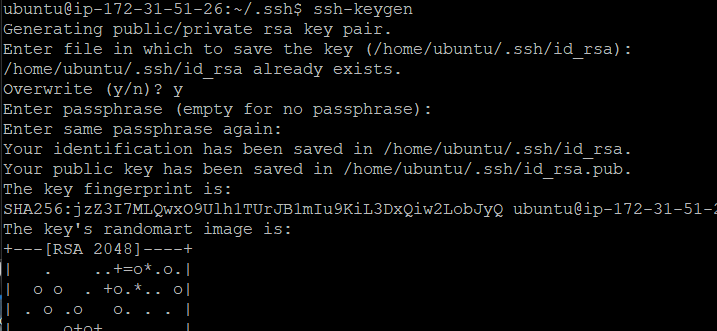
Pengecekan pyspark

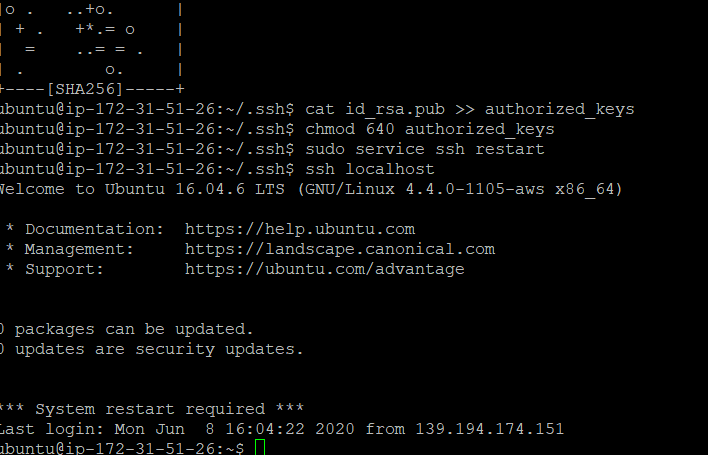


Membuat slave pada spark



Mengakifkan localhost





**Bagian 2. Menginstall jupyter**

1. Menginstall pip

Sudo apt install pip

1. Menginstall Jupiter

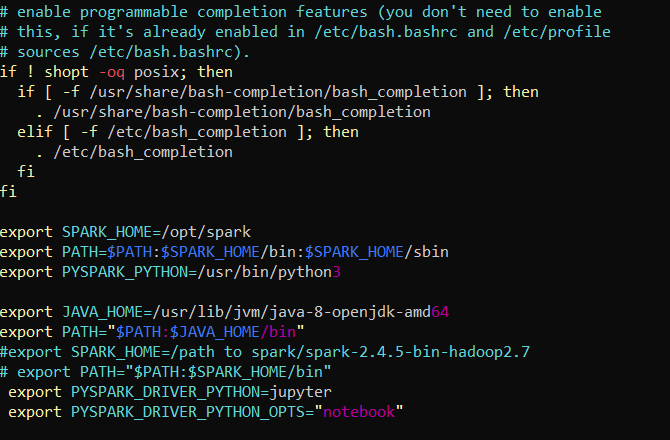
Pip3 install jupyter

1. Menginstall toree

Pip3 install toree

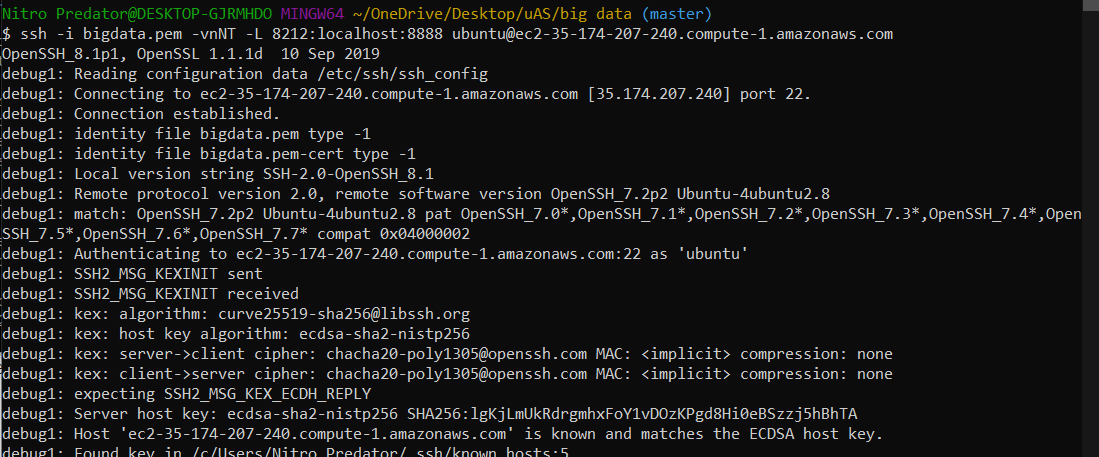
1. Menambahkan pada bashrc (vim ~/.bashrc)

 export JAVA\_HOME=/usr/lib/jvm/**installed java 8**  
 export PATH="$PATH:$JAVA\_HOME/bin"  
 export SPARK\_HOME=/**path to spark**/spark-2.3.1-bin-hadoop2.7  
 export PATH="$PATH:$SPARK\_HOME/bin"  
 export PYSPARK\_DRIVER\_PYTHON=jupyter  
 export PYSPARK\_DRIVER\_PYTHON\_OPTS="notebook"

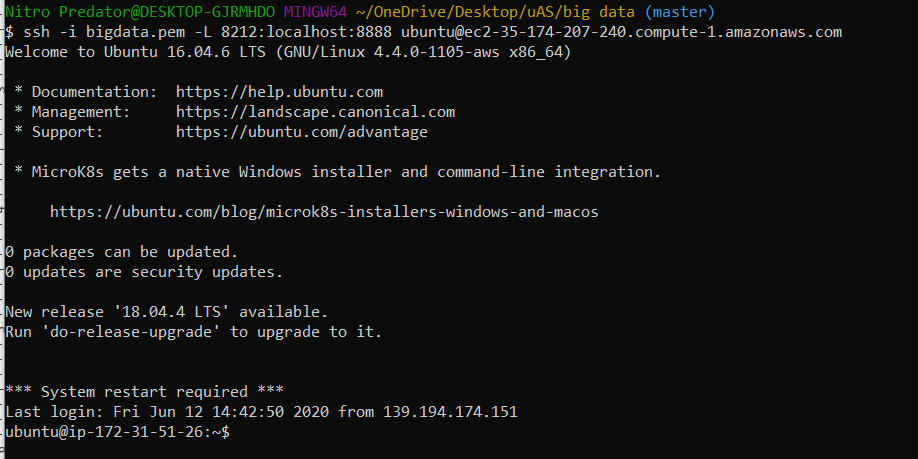


1. Mengkoneksikan ke jupyter dengan menggunakan Git Bash

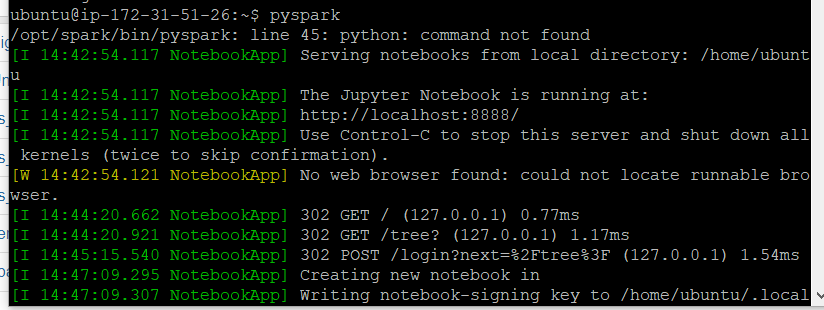
* Error. jika tidak dapat tersambung (open failed)



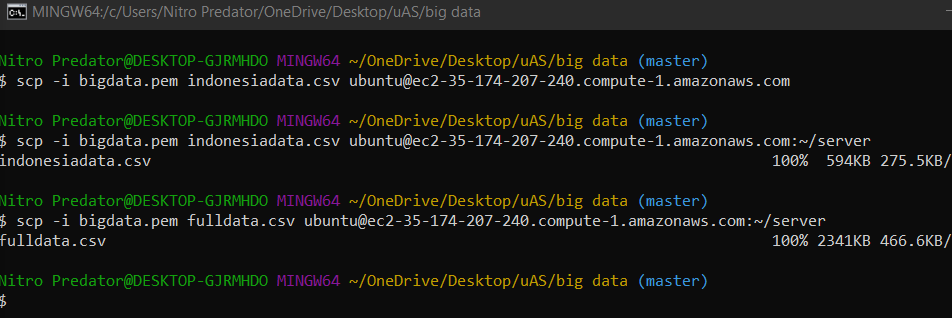
* Melakukan ssh agar dapat terkoneksi dan dibuka di laptop



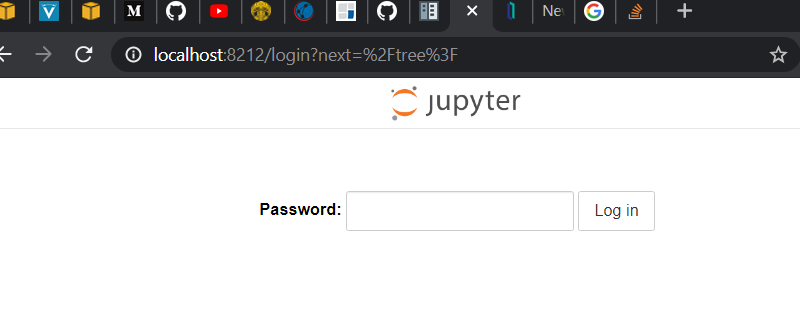
1. Melakukan pyspark pada putty



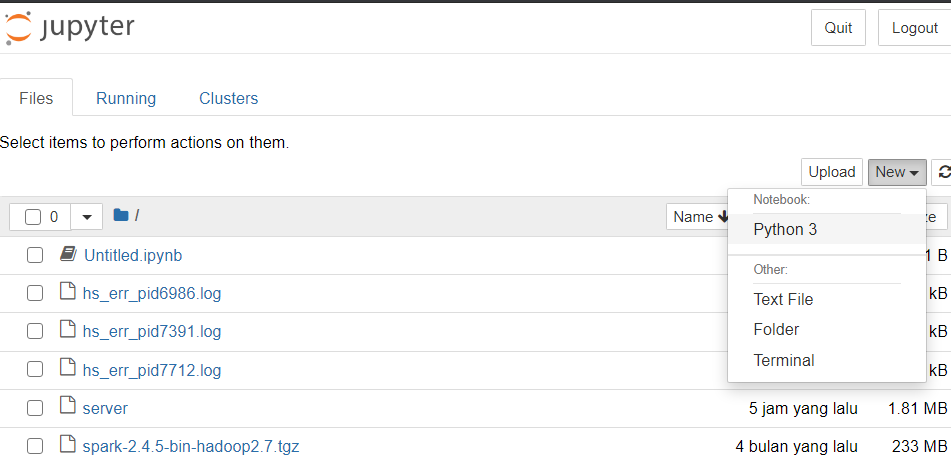
1. Mengupload data csv ke server



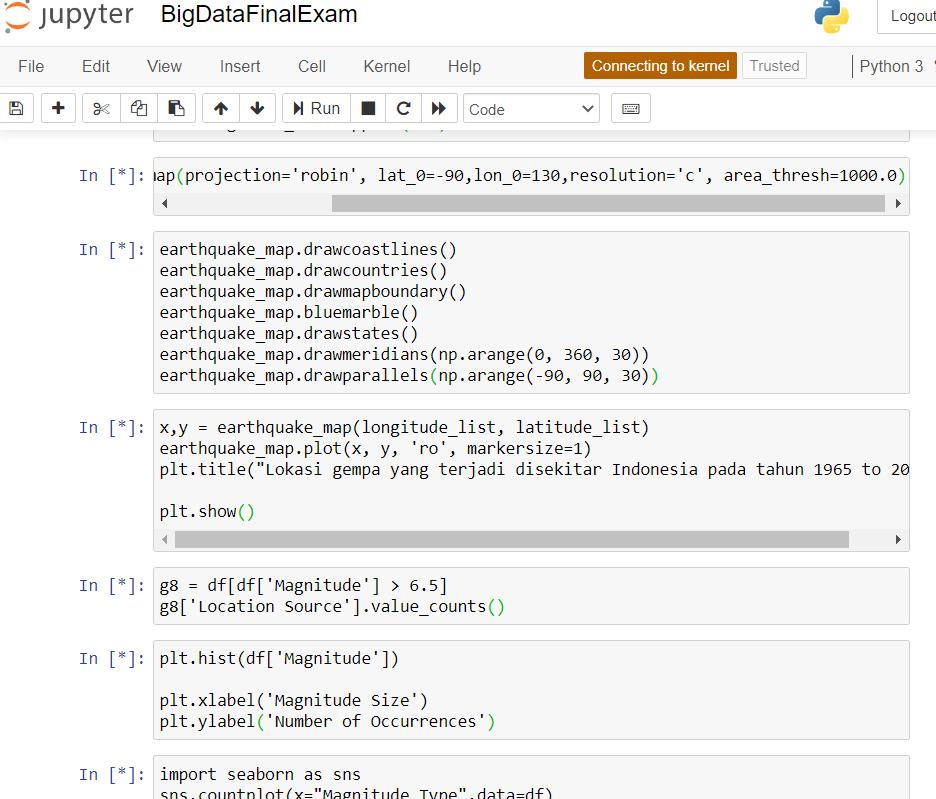
1. Akses alamat localhost:8212 pada laptop, dengan menggunakan password yang telah dibuat



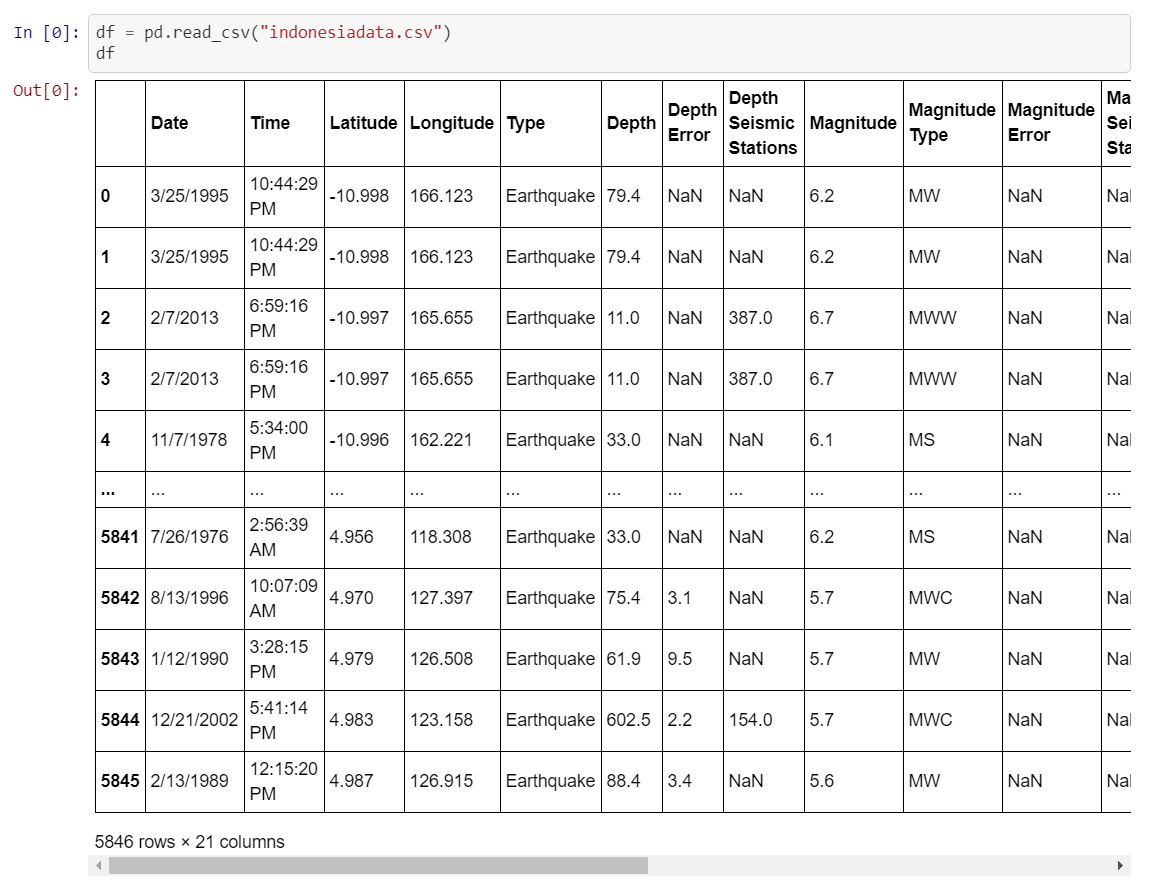
Membuat file baru



Contoh code program, untuk program yang utuh terlampir pada folder code



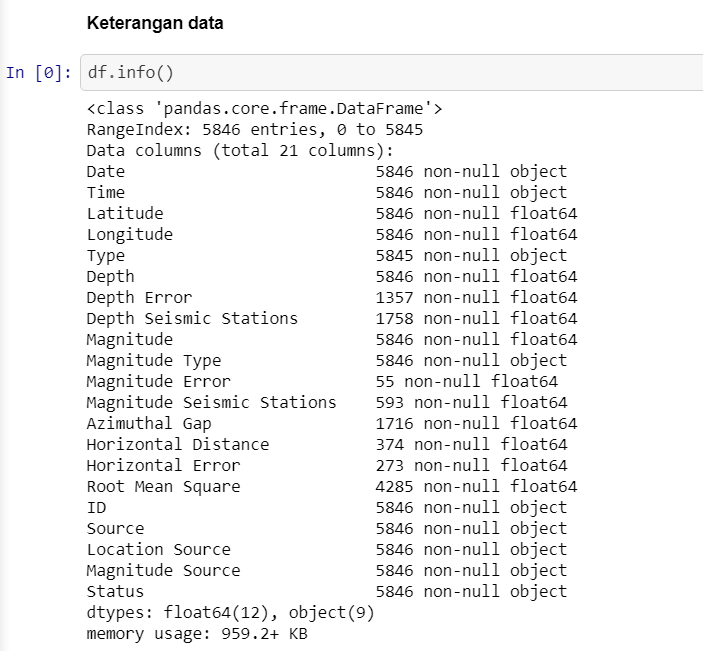
Hasil yang didaptkan



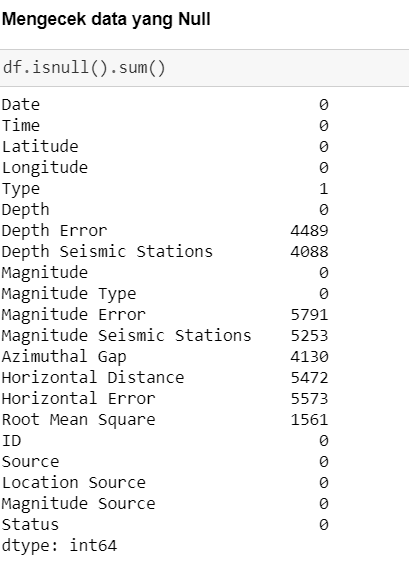
1. **Analisa**

Pada analisa ini saya juga menggunakan google colab untuk memastikan hasil yang benar dari code yang telah dibuat. Hasil didapatkan dari analisa ini, sebagai berikut:

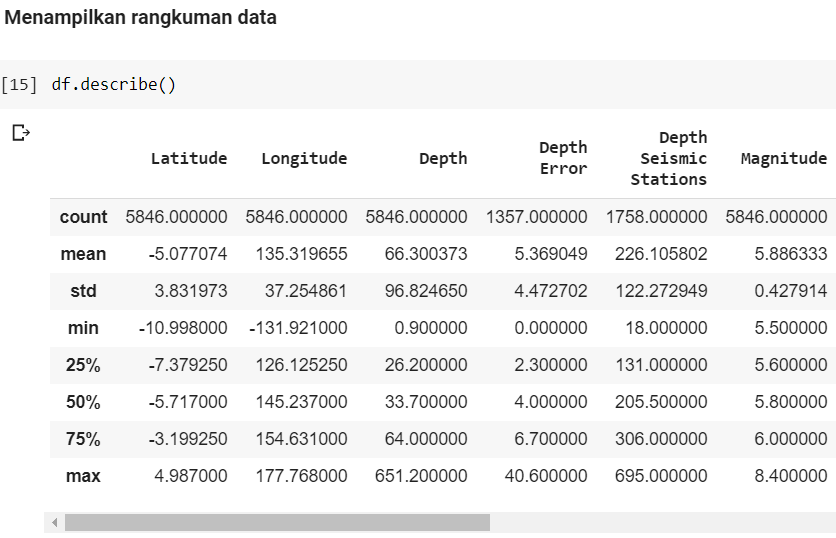
* Keterangan data yang didapat dari gempa yang terjadi di Indonesia

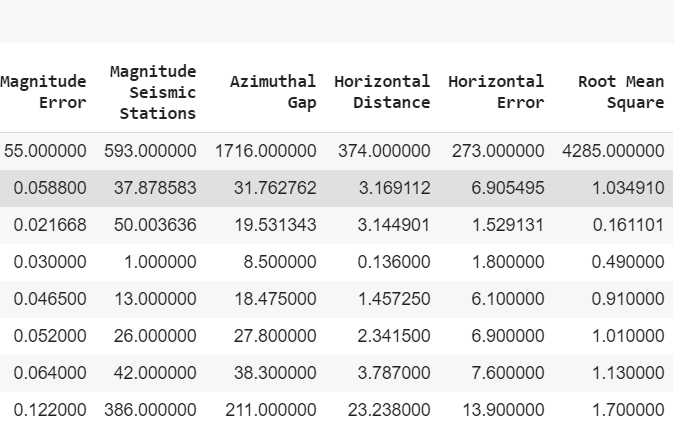


* Mengecek data yang nilainya null

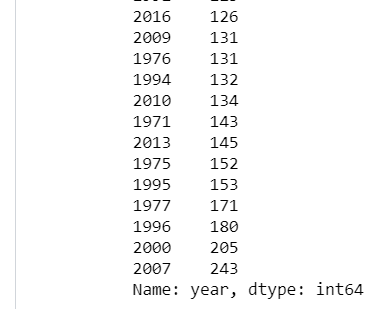
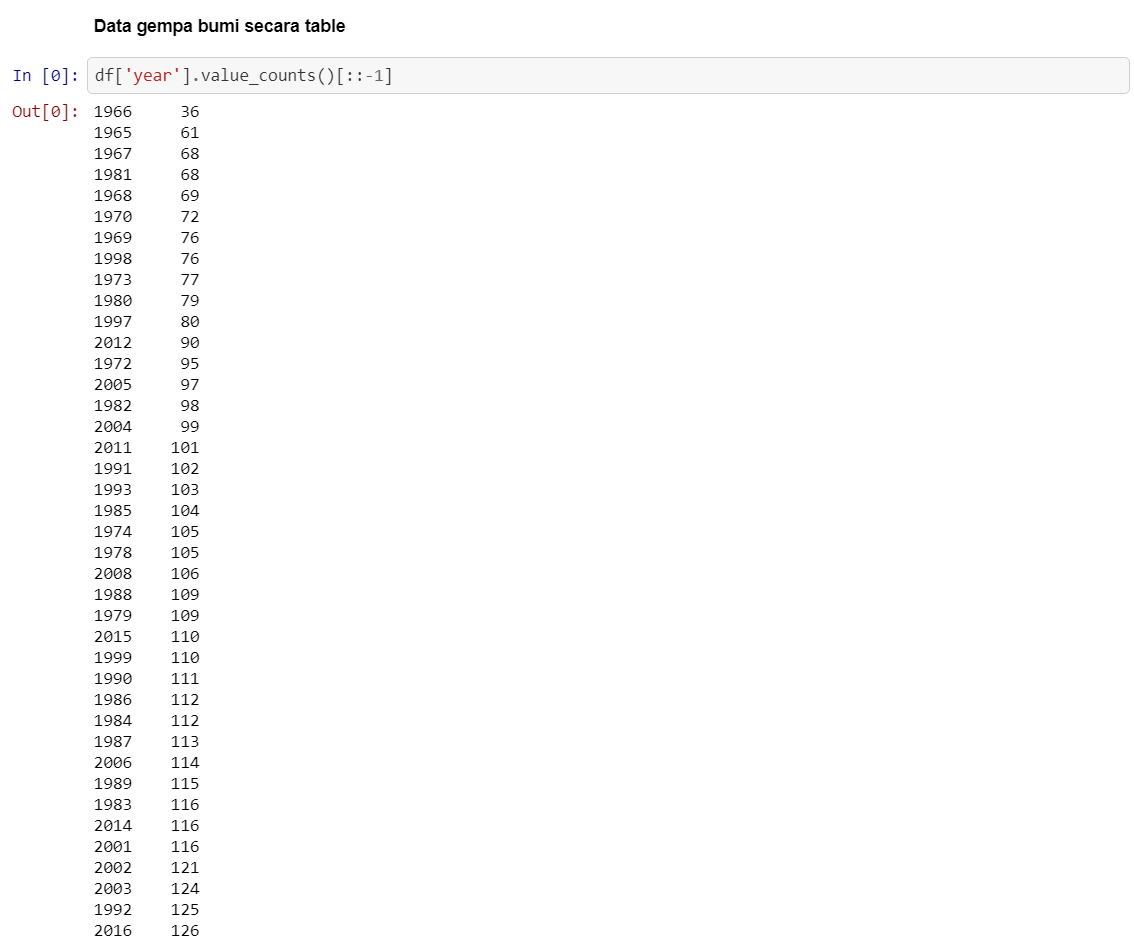


* Rangkuman data





* Banyaknya kasus gempa (dalam tahun)



Dari data-data yang diambil, dapat dilihat bahwa Indonesia menjadi salah satu negara yang sering terjadi gempa.

1. **Kesimpulan**

Gempa yang terjadi di Indonesia dari tahun 1965 sampai dengan 2016 tercatat terdapat 5847 kasus gempa bumi yang terdeteksi. Tidak hanya sering terjadinya gempa, skala magnitude yang terjadi juga besar. Gempa yang terjadi terbanyak pada tahun 2007 yaitu sebanyak 243 kali. Dari data-data ini, Indonesia menjadi suatu negara yang tiap tahunnya dapat terjadi gempa bumi.

Berganti topik, untuk aplikasi yang digunakan yaitu, digunakan jupyter yang diintegrasikan dengan spark yang terdapat pada AWS EC2. Data file .csv diupload ke server AWS EC2 agar dapat digunakan.

**Alamat yang dapat diakses untuk program python:**

1. Alamat untuk resource code: <https://github.com/muhfadil26/BigDataFinalExam>
2. Alamat Vidio: **https://youtu.be/TPhoP2\_DYqA**