APACHE SPARK SQL

* Penjelasan

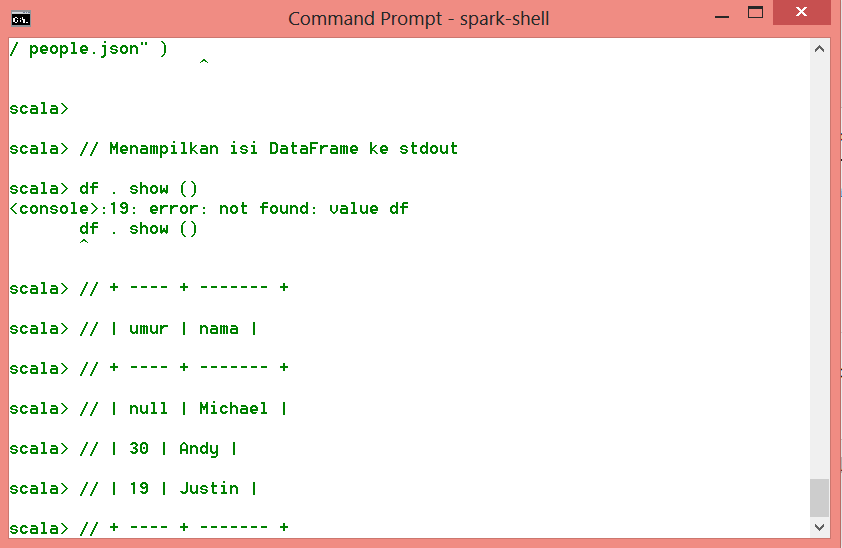
apache Spark SQL adalah modul Spark untukpemrosesan data terstruktur yang disediakan oleh Spark SQL untuk memberikan Spark dengan lebih banyak informasi tentang struktur data dan perhitungan yang sedang dilakukan

* Gambar di bawah ini masih error karena belum di definisikan oleh spark

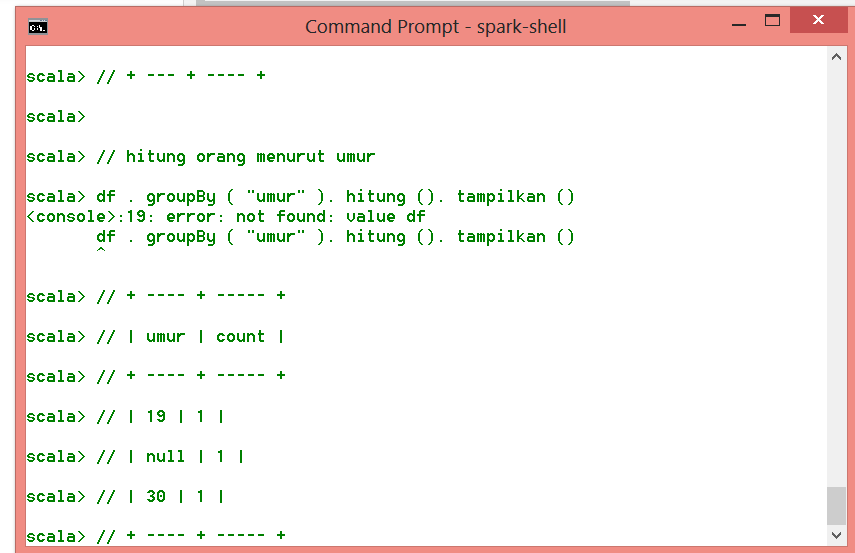


* Membuat Dataframe

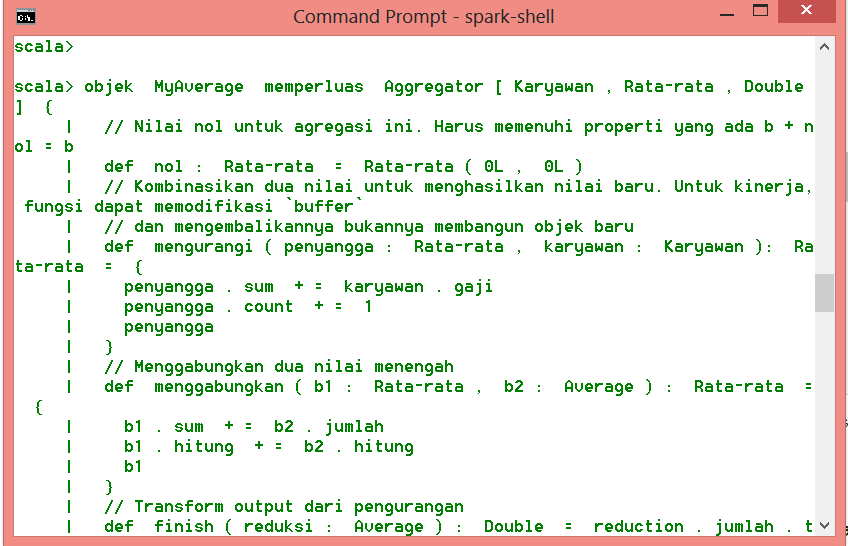
Dengan RDD spark session maka dapat membuat dataframe

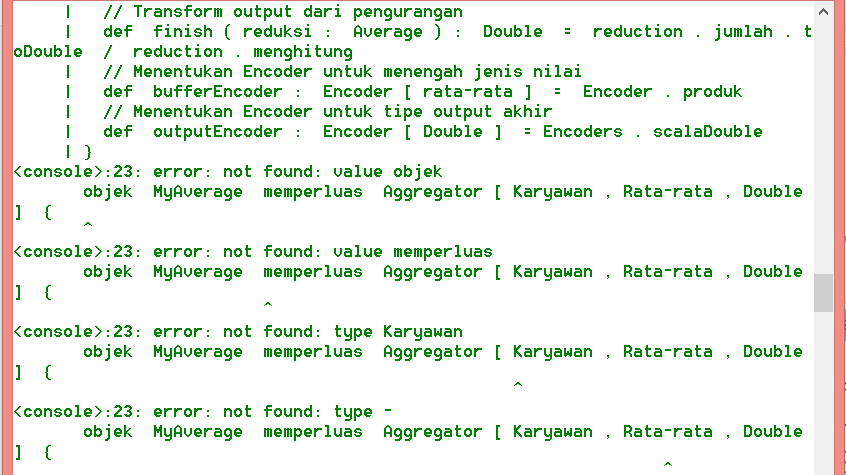


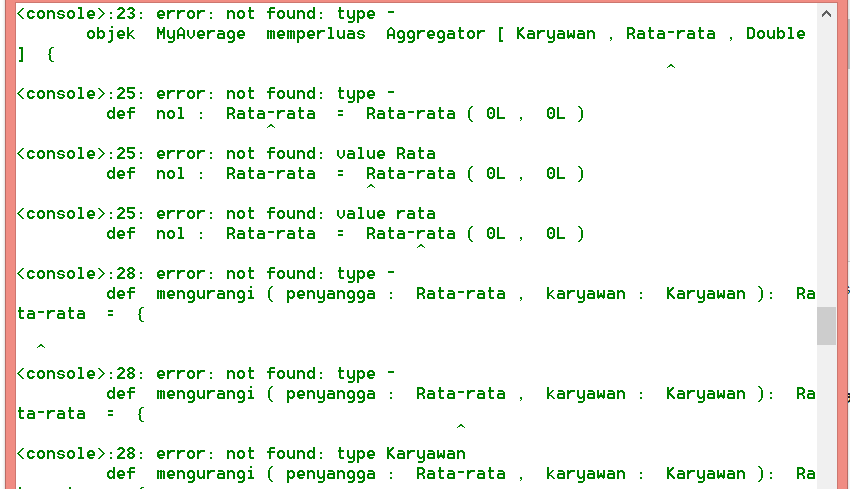
* Mnghitung umur dengan dataframe

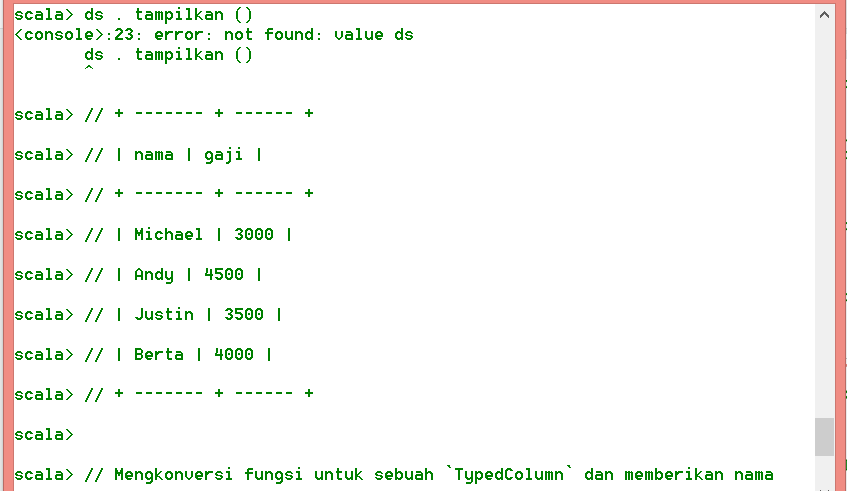


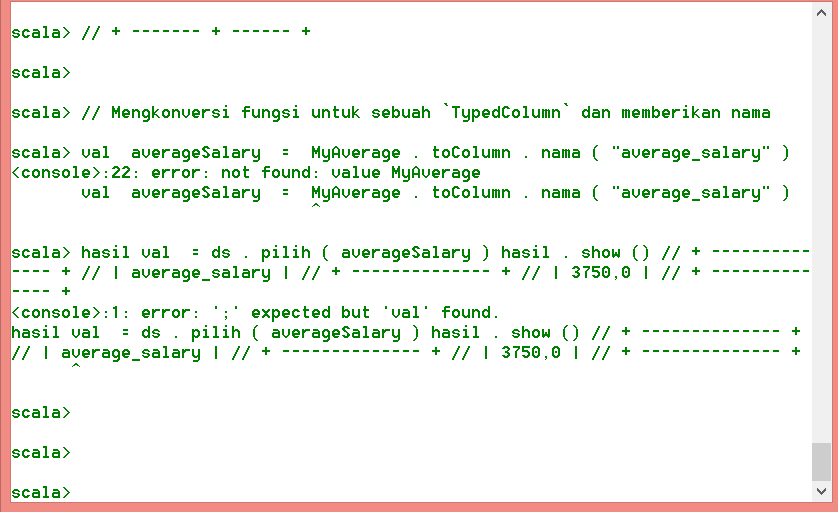
* Data frame juga dapat di gunakan untuk memperluar aggregator (karyawan, rata, dll)
* Dan juga bisa untuk menghitung gaji karyawan











Datataset juga mirip dengan RDD, namun, dataset menggunakan serialisasi Java atau Kryo yang menggunakan [Encoder](https://spark.apache.org/docs/latest/api/scala/index.html#org.apache.spark.sql.Encoder) khusus untuk membuat serialize objek untuk pemrosesan atautran smisi melalui jaringan. Sementara kedua encoders dan serialisasi standard bertanggungjawab untuk mengubah sebuah objek menjadi byte.