

DataSet
california Quarterly Hospital Sales 2020 - XLSV

Data cleaning

.csv generated

Databricks

- group by ()
- order by ()
- count
- Average
- sort ()
- aggregate ()

Data input →

Mapper

Shuffling
&
Sorting

Reducer

Analysis

Output