Class project for Introduction to NoSQL Databases: 1980 US census population characteristic research

Tomoaki Tanaka

7/9/2016

MongoDB sharding

Set up MongoDB sharded cluster on AWS EC2

- Create three EC 2 instances
- Set up Config server, Query Router, 3 shards on EC2 instances

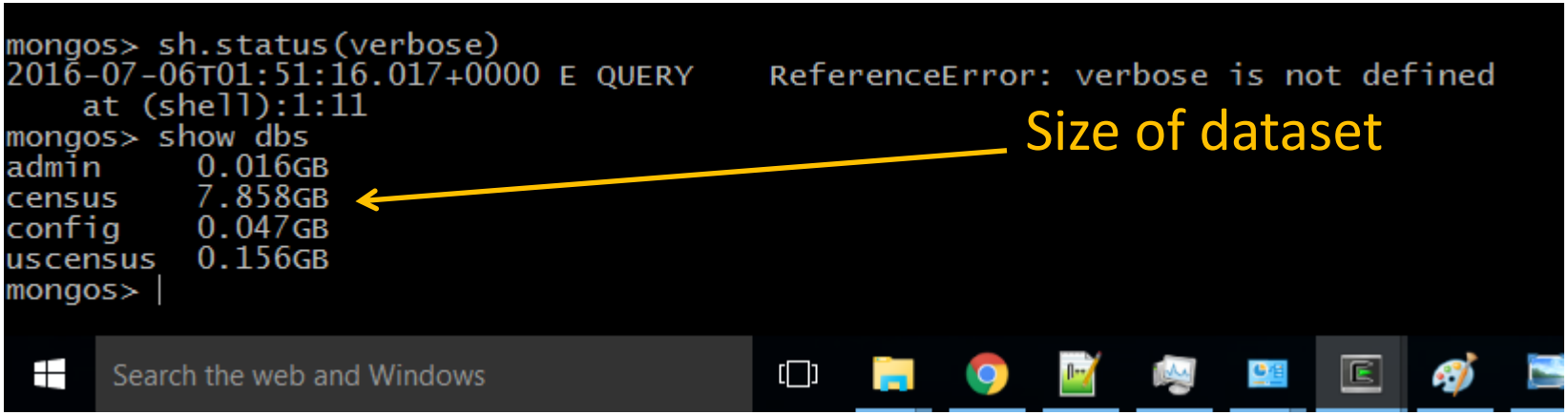
Use AWS Dataset: 1980 US Census

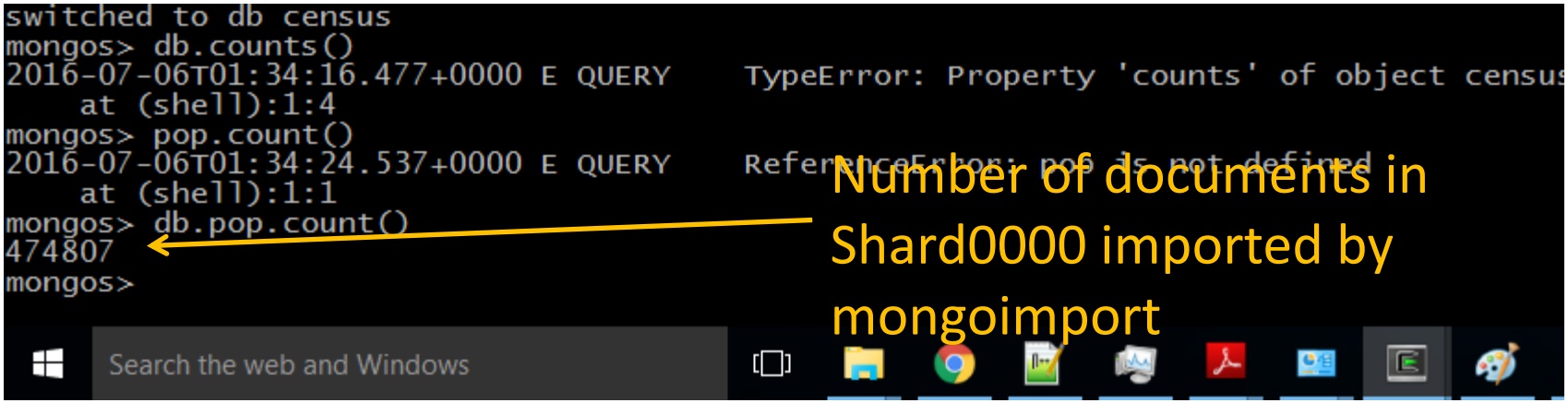
https://aws.amazon.com/datasets/1980-us-census/?tag=datasets%23keywords%23economics

Sharding status:

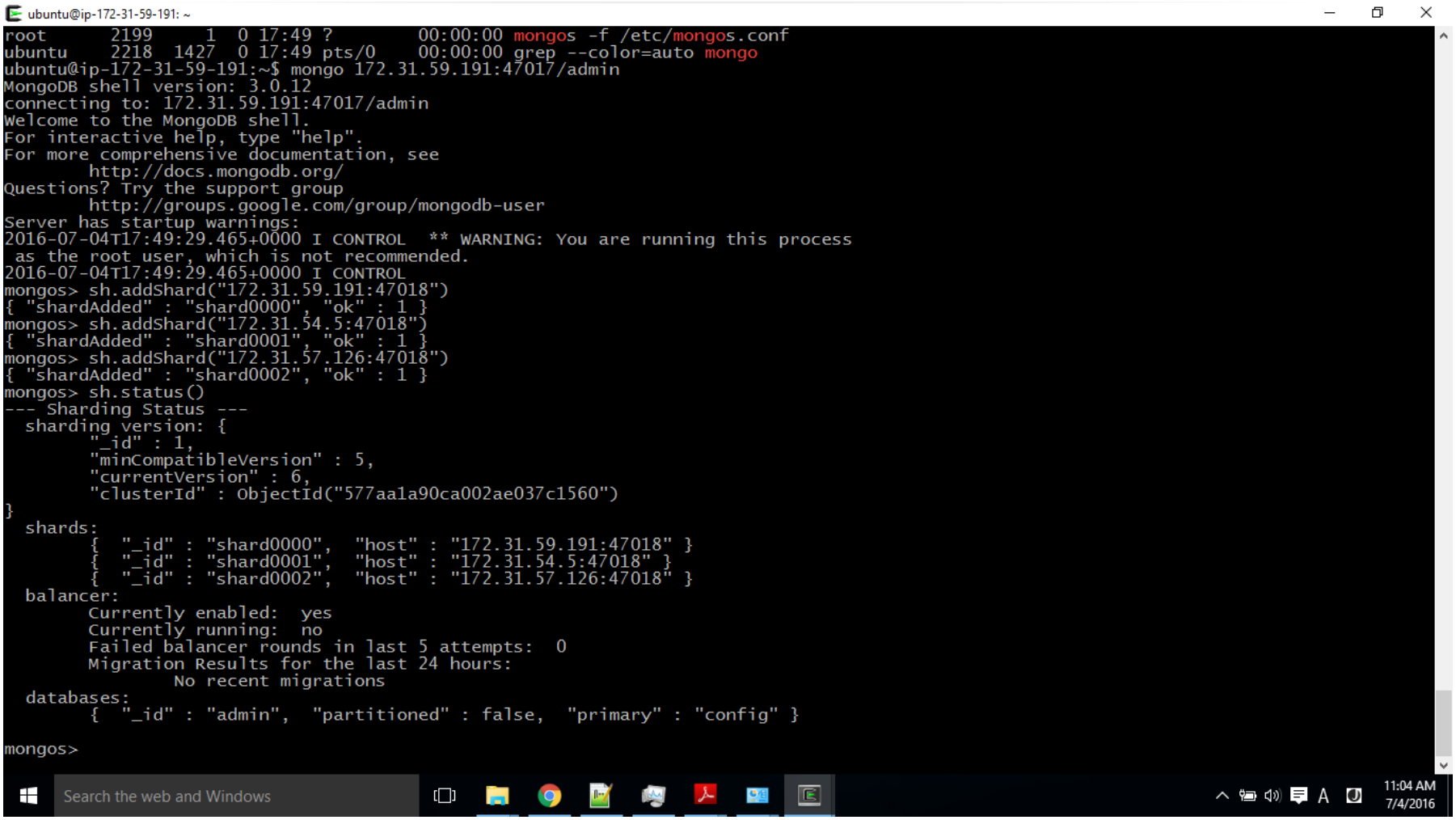
- No. of documents: 793,340
- No. of chunks: 58

Imported dataset

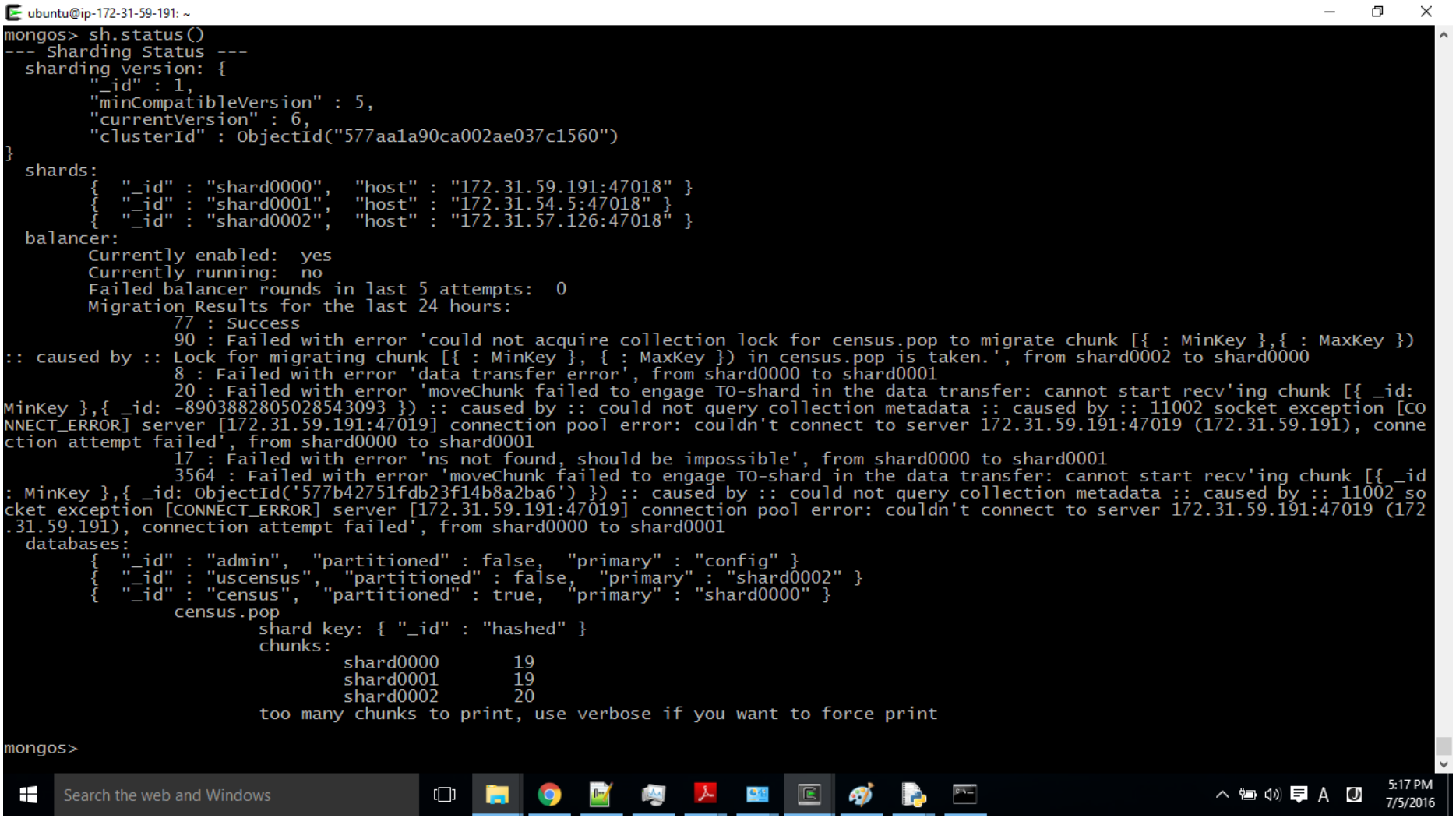




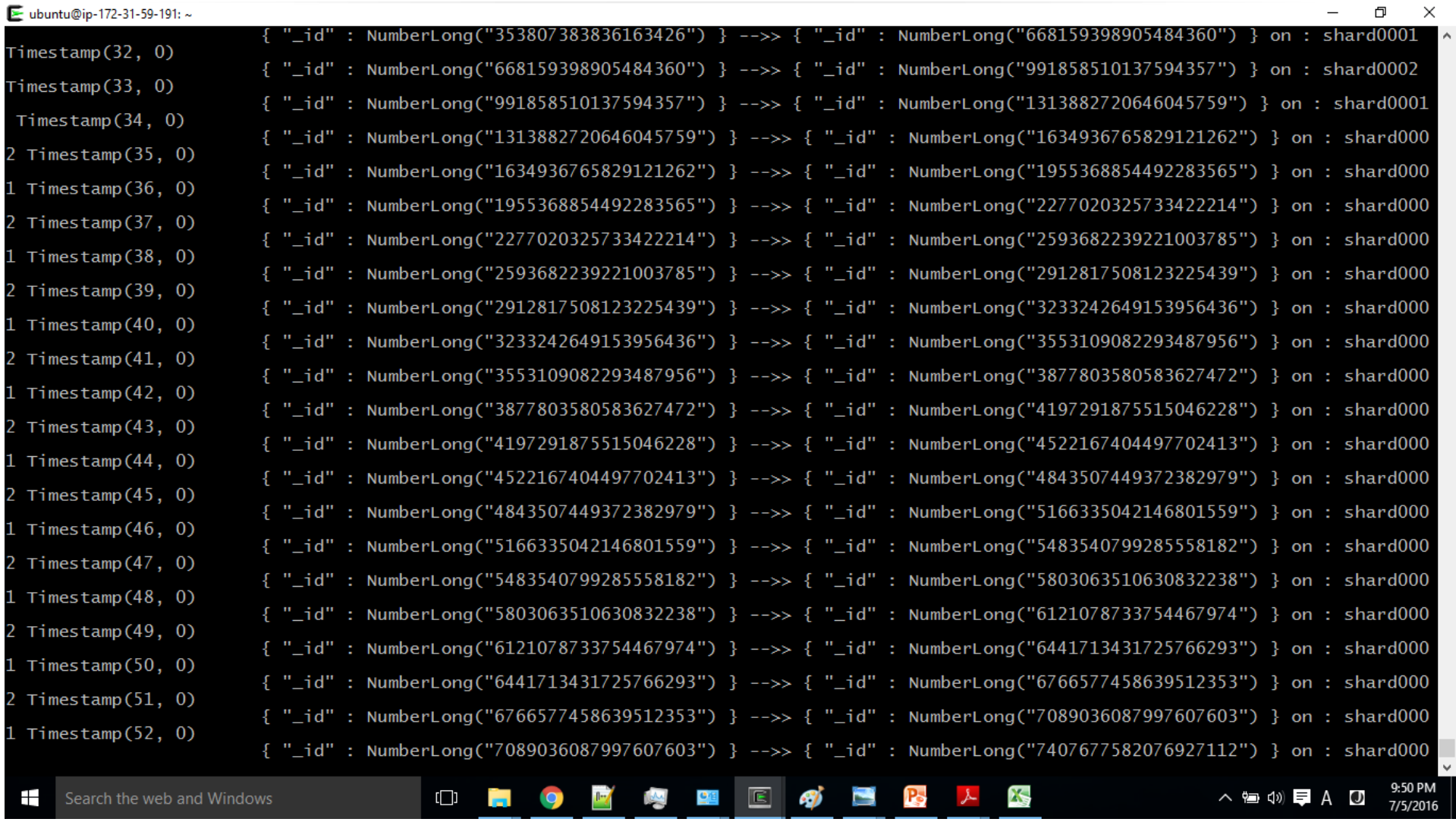
sh.status() before sharding



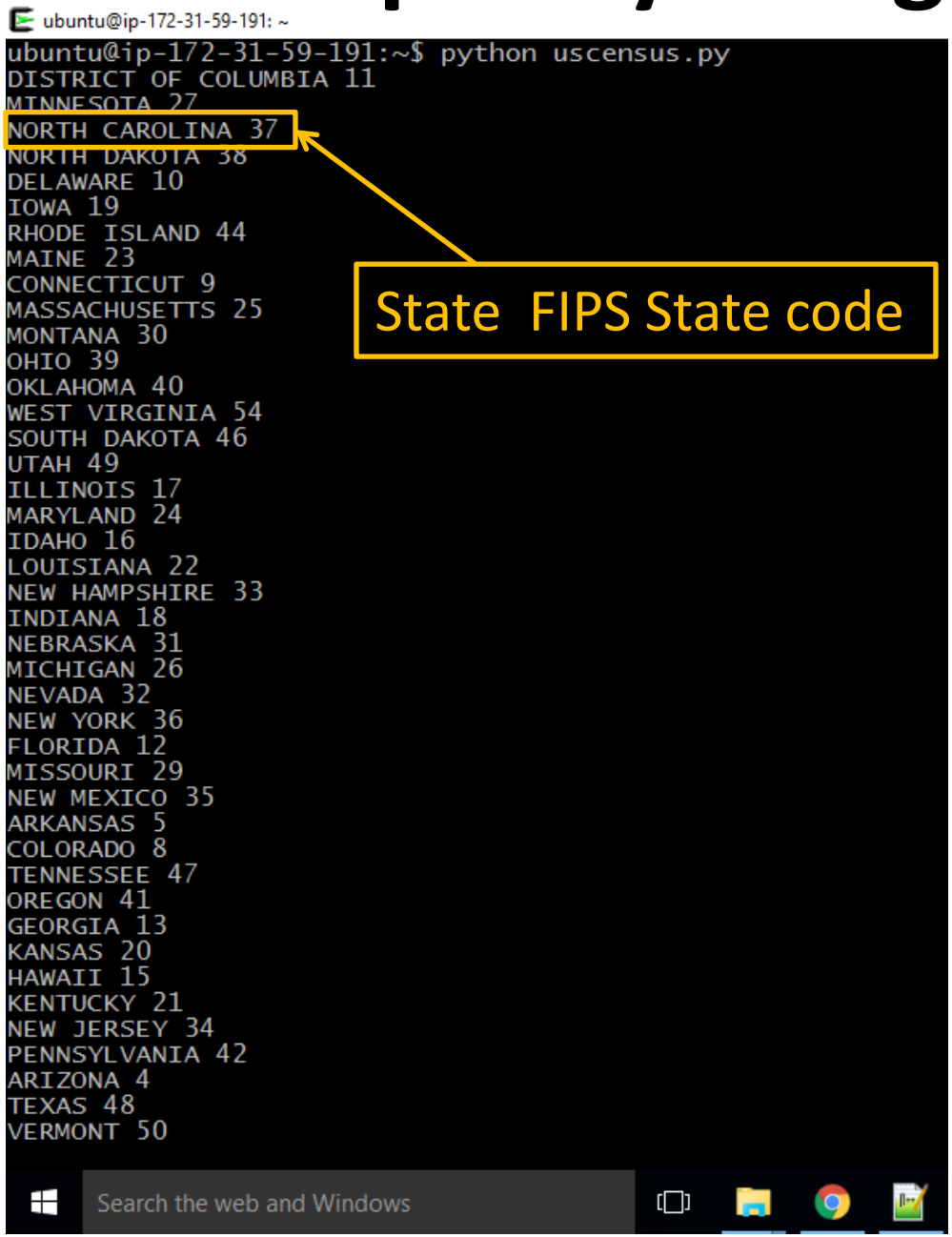
sh.status( ) after sharding



sh.status(true)



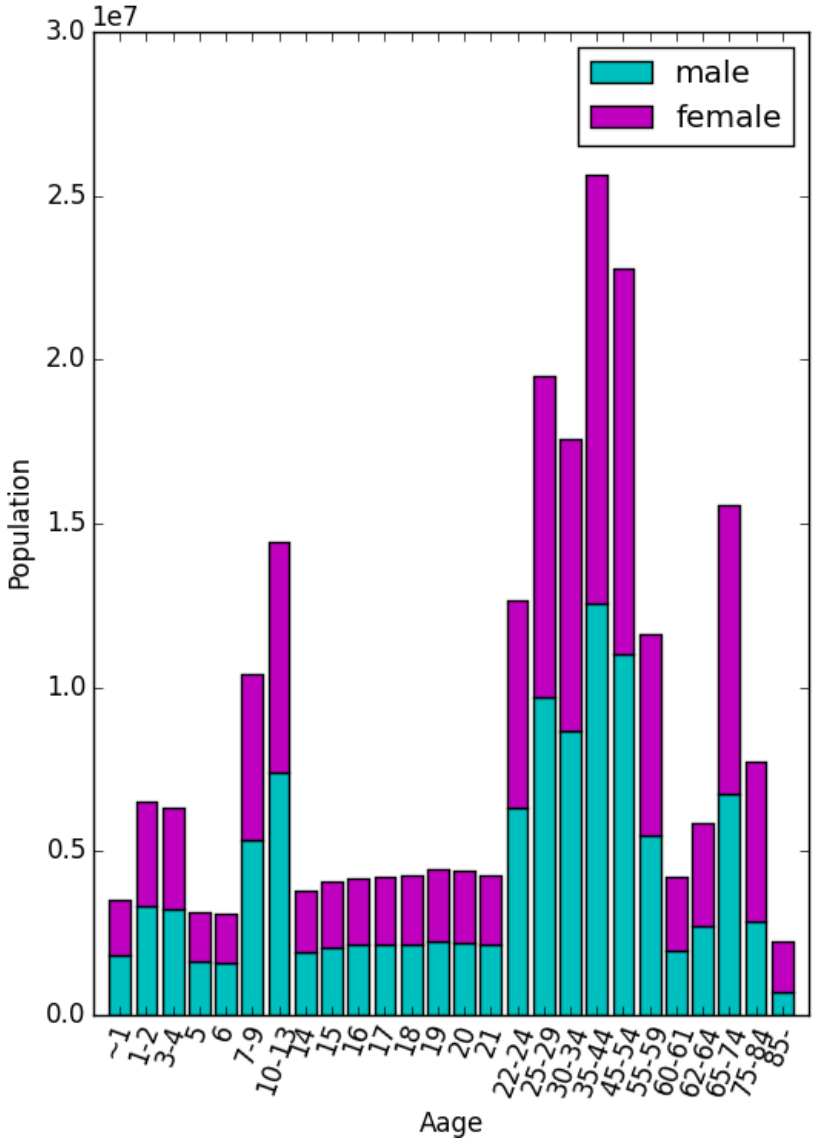
Data output by using Pymongo



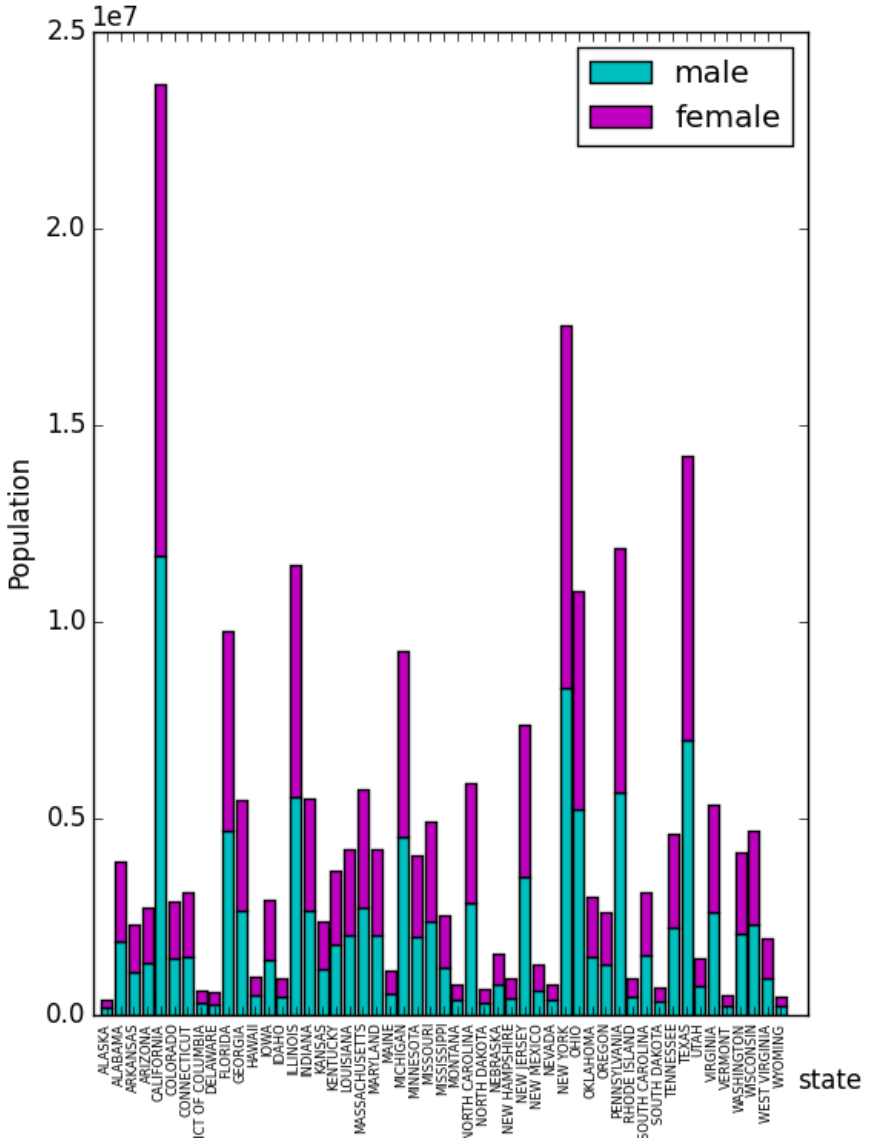
uscensus.py retrieved the population by each age rage, each state and sex.

Result: 1980 US population

Total population by age



Total population by state



Result: 1980 US population

Population table for each state

State	population
IDAHO	943935
LOUISIANA	4205900
MINNESOTA	4075970
NEW HAMPSHIRE	920610
NORTH CAROLINA	5881766
UTAH	1461037
ILLINOIS	11426518
INDIANA	5490224
NEBRASKA	1569825
RHODE ISLAND	947154
MICHIGAN	9262078
NEVADA	800493
NEW YORK	17558072
DISTRICT OF COLUMBIA	638333
FLORIDA	9746324
MISSOURI	4916686
NEW MEXICO	1302894
ARKANSAS	2286435
COLORADO	2889964
MAINE	1124660
TENNESSEE	4591120
WEST VIRGINIA	1949644
OREGON	2633105
GEORGIA	5463105
KANSAS	2363679

State	population
DELAWARE	594338
HAWAII	964691
KENTUCKY	3660777
MASSACHUSETTS	5737037
NEW JERSEY	7364823
PENNSYLVANIA	11863895
ARIZONA	2718215
CONNECTICUT	3107576
TEXAS	14229191
NORTH DAKOTA	652717
MONTANA	786690
VERMONT	511456
CALIFORNIA	23667902
OHIO	10797630
MARYLAND	4216975
ALASKA	401851
VIRGINIA	5346818
WYOMING	469557
IOWA	2913808
WASHINGTON	4132156
ALABAMA	3893888
MISSISSIPPI	2520638
OKLAHOMA	3025290
SOUTH DAKOTA	690768
SOUTH CAROLINA	3121820
WISCONSIN	4705767