

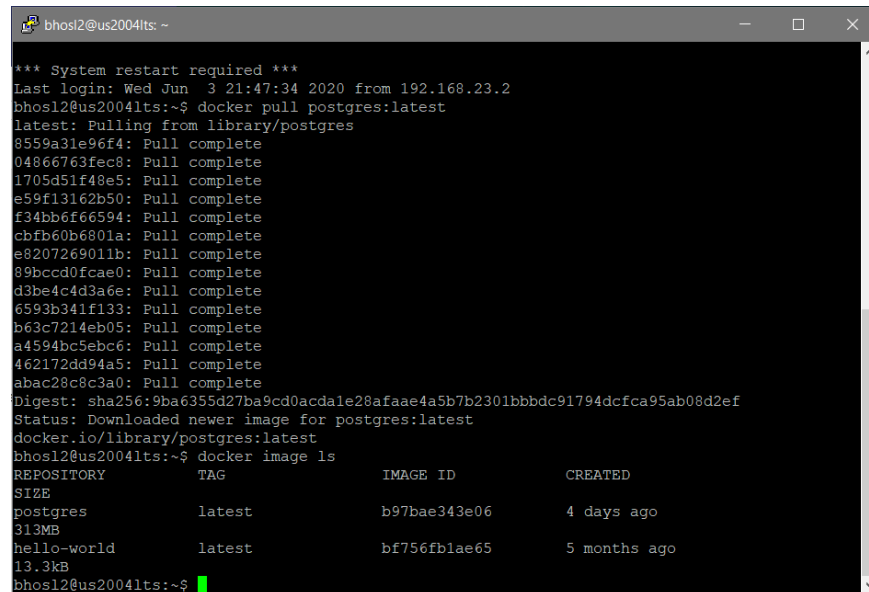
Containerization and Big Data: Exercise 2

Brandon Hosley

June 14, 2020

2 Initializing PostgreSQL Container

2.1 Pulling PostgreSQL



```
bhos12@us2004lts: ~  
*** System restart required ***  
Last login: Wed Jun  3 21:47:34 2020 from 192.168.23.2  
bhos12@us2004lts:~$ docker pull postgres:latest  
latest: Pulling from library/postgres  
8559a31e96f4: Pull complete  
04866763fec8: Pull complete  
1705d51f48e5: Pull complete  
e59f13162b50: Pull complete  
f34bb6f66594: Pull complete  
cbfb60b6801a: Pull complete  
e8207269011b: Pull complete  
89bccd0fcae0: Pull complete  
d3be4c4d3a6e: Pull complete  
6593b341f133: Pull complete  
b63c7214eb05: Pull complete  
a4594bc5ebc6: Pull complete  
462172dd94a5: Pull complete  
abac28c8c3a0: Pull complete  
Digest: sha256:9ba6355d27ba9cd0acd9e28afaae4a5b7b2301bbbd9c91794dcfca95ab08d2ef  
Status: Downloaded newer image for postgres:latest  
docker.io/library/postgres:latest  
bhos12@us2004lts:~$ docker image ls  
REPOSITORY          TAG             IMAGE ID        CREATED  
SIZE  
postgres            latest         b97bae343e06    4 days ago  
313MB  
hello-world         latest         bf756fblae65    5 months ago  
13.3kB  
bhos12@us2004lts:~$
```

2.2 Running PostgreSQL

```
bhos12@us2004lts: ~
running bootstrap script ... ok
performing post-bootstrap initialization ... ok
syncing data to disk ... ok

Success. You can now start the database server using:

    pg_ctl -D /var/lib/postgresql/data -l logfile start

initdb: warning: enabling "trust" authentication for local connections
You can change this by editing pg_hba.conf or using the option -A, or
--auth-local and --auth-host, the next time you run initdb.
waiting for server to start....2020-06-14 13:48:16.361 UTC [46] LOG:  starting PostgreSQL
12.3 (Debian 12.3-1.pgdg100+1) on x86_64-pc-linux-gnu, compiled by gcc (Debian 8.3.0-6) 8.3.
64-bit
2020-06-14 13:48:16.363 UTC [46] LOG:  listening on Unix socket "/var/run/postgresql/.s.PG
L.5432"
2020-06-14 13:48:16.383 UTC [47] LOG:  database system was shut down at 2020-06-14 13:48:1
UTC
2020-06-14 13:48:16.388 UTC [46] LOG:  database system is ready to accept connections
done
server started

/usr/local/bin/docker-entrypoint.sh: ignoring /docker-entrypoint-initdb.d/*

2020-06-14 13:48:16.451 UTC [46] LOG:  received fast shutdown request
waiting for server to shut down....2020-06-14 13:48:16.453 UTC [46] LOG:  aborting any act
e transactions
2020-06-14 13:48:16.454 UTC [46] LOG:  background worker "logical replication launcher" (P
53) exited with exit code 1
2020-06-14 13:48:16.454 UTC [48] LOG:  shutting down
2020-06-14 13:48:16.472 UTC [46] LOG:  database system is shut down
done
server stopped

PostgreSQL init process complete; ready for start up.

2020-06-14 13:48:16.563 UTC [1] LOG:  starting PostgreSQL 12.3 (Debian 12.3-1.pgdg100+1) o
x86_64-pc-linux-gnu, compiled by gcc (Debian 8.3.0-6) 8.3.0, 64-bit
2020-06-14 13:48:16.565 UTC [1] LOG:  listening on IPv4 address "0.0.0.0", port 5432
2020-06-14 13:48:16.567 UTC [1] LOG:  could not create IPv6 socket for address "::: Addre
family not supported by protocol
2020-06-14 13:48:16.572 UTC [1] LOG:  listening on Unix socket "/var/run/postgresql/.s.PGS
.5432"
2020-06-14 13:48:16.589 UTC [55] LOG:  database system was shut down at 2020-06-14 13:48:1
UTC
2020-06-14 13:48:16.594 UTC [1] LOG:  database system is ready to accept connections
bhos12@us2004lts:~$
```

2.3 Titanic Dataset

Creating DB and Table for Titanic Dataset in PostgreSQL DBMS.

```
bhos12@us2004lts: ~  
mydb=# \dt  
List of relations  
Schema | Name | Type | Owner  
-----+-----+-----+-----  
public | passengers | table | postgres  
(1 row)  
  
mydb=# \copy passengers (survived, pclass, name, sex, age, siblings_spouses,  
\copy: parse error at end of line  
mydb=# parents_children, fare) from '/path/to/titanic.csv' CSV HEADER;  
ERROR: syntax error at or near "parents_children"  
LINE 1: parents_children, fare) from '/path/to/titanic.csv' CSV HEAD...  
^  
mydb=# \copy passengers (survived, pclass, name, sex, age, siblings_spouses, parents_child  
ren, fare) from '/path/to/titanic.csv' CSV HEADER;  
/path/to/titanic.csv: No such file or directory  
mydb=# \copy passengers (survived, pclass, name, sex, age, siblings_spouses, parents_child  
ren, fare) from './titanic.csv' CSV HEADER;  
COPY 887  
mydb=#
```

2.4 Analytical Query

Running an analytic query “average age of passengers who survived the Titanic disaster by ticket class” and results.

```
bhos12@us2004lts: ~  
bhos12@us2004lts:~$ docker exec -it mypostgreslname bash  
root@790d0f25a321:/# psql -U postgres mydb  
psql (12.3 (Debian 12.3-1.pgdg100+1))  
Type "help" for help.  
  
mydb=# SELECT avg(age), pclass FROM passengers WHERE survived = 1 GROUP BY pclass;  
avg | pclass  
-----+-----  
35.96264705882353 | 1  
21.41109243697479 | 3  
26.17045977011494 | 2  
(3 rows)  
  
mydb=#
```