

Manual

环境配置

需要安装以下环境：

- Docker
- Java 8
- Maven
- [mongosh](#)
- [MongoDB Database Tools](#)
- [Mongo Compass](#)

数据准备：将 db-generation 解压缩放置到 \${PROJECT}/（指代项目目录）下，重命名为 data 目录

部署方法

脚本介绍

- /data-generate 目录
 - 该脚本会使用 /data 目录下的数据，并将生成的新数据一并放置在 /data 下
- /script 目录
 - setup.sh、setup.bat：初始化并启动容器，包括数据加载等，用于第一次启动
 - shutdown.sh、shutdown.bat：停止容器的运行，并删除容器
 - start.sh、start.bat：开始容器的运行（第一次启动请使用 setup.sh）
 - stop.sh、stop.bat：停止容器的运行

部署流程

数据生成

使用脚本：/data-generate/genTable_mongoDB10G.py

```
# 进入到相应目录，然后运行 genTable_mongoDB10G.py
python3 genTable_mongoDB10G.py
```

启动服务

首次启动：使用脚本 /script/setup.sh

```
# 进入到 /script 目录，然后运行 setup.sh
./setup.sh
```

setup.sh 脚本说明

```
#!/bin/bash
```

```
# hdfs
docker-compose -f ../hdfs/docker-compose.yml up -d
sleep 10
# 创建目录, 将 /data/articles/ 上传至 hdfs
docker exec -it namenode hadoop fs -mkdir /articles/
docker exec -it namenode hadoop fs -put /data/articles/ /articles/

# redis
docker pull redis:7.0.0
docker run -itd --name redis -p 6379:6379 redis
docker network connect hdfs_hadoopnet redis --ip 172.21.0.6

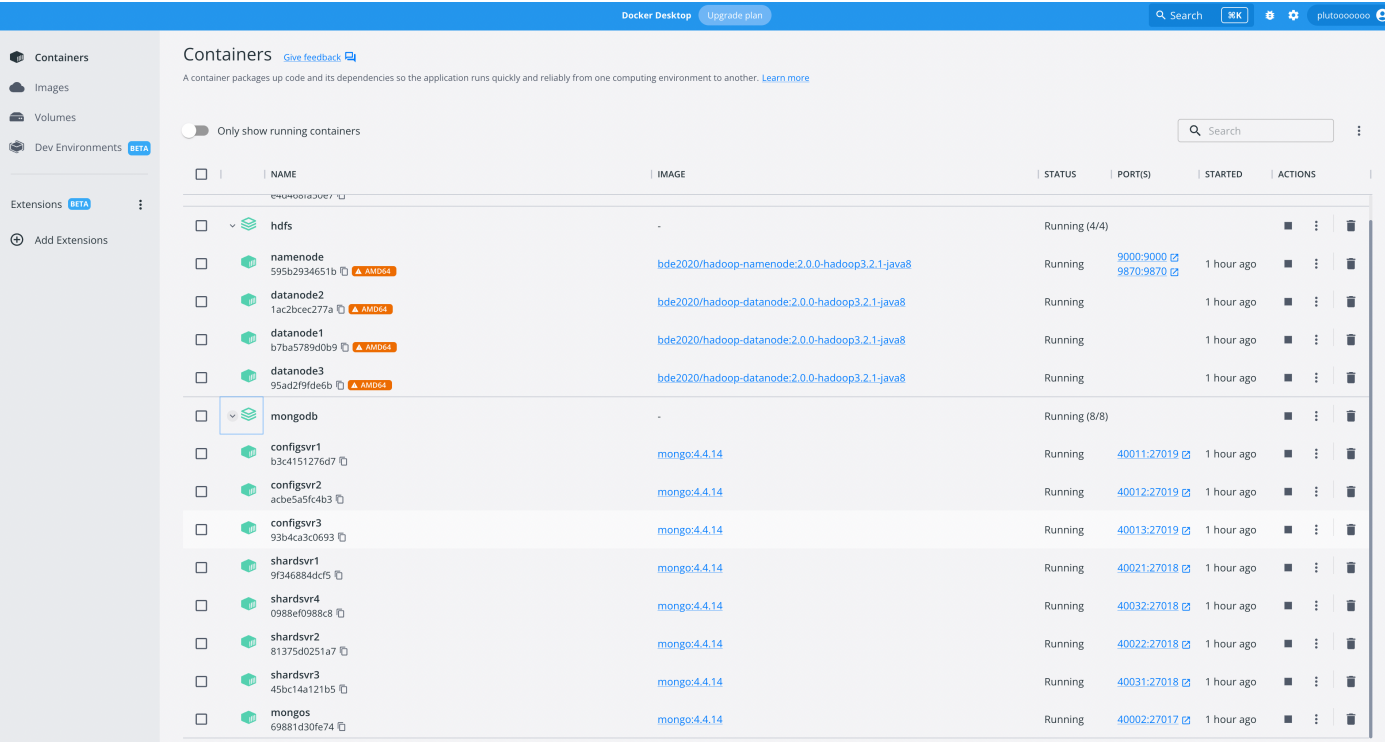
# mongodb
# config servers
docker-compose -f ../mongodb/docker-compose-configsvr.yml up -d
mongosh --host localhost --port 40011 ../mongodb/setup-configsvr.js
# shard servers
docker-compose -f ../mongodb/docker-compose-shardsvr.yml up -d
mongosh --host localhost --port 40021 ../mongodb/setup-shardsvr-rs1.js
mongosh --host localhost --port 40031 ../mongodb/setup-shardsvr-rs2.js
# mongos
docker-compose -f ../mongodb/docker-compose-mongos.yml up -d
sleep 5
mongosh --host localhost --port 40002 ../mongodb/setup-mongos.js
docker network connect hdfs_hadoopnet mongos --ip 172.21.0.7

# load data
# 使用 mongoimport 直接导入生成的数据
mongoimport --host localhost --port 40002 -d ddbms -c user --file ../data/user.dat
mongoimport --host localhost --port 40002 -d ddbms -c article --file
../data/article.dat
mongoimport --host localhost --port 40002 -d ddbms -c read --file ../data/read.dat
# 使用 js 脚本生成 be-read 和 rank 表
mongosh --host localhost --port 40002 ../mongodb/generate-beread.js
mongosh --host localhost --port 40002 ../mongodb/generate-rank.js
```

环境查看说明

Mongo

首先在 Docker Desktop 中 进入 mongos 容器：



在 Terminal 中 运行如下命令可以查看导入的数据

```
# 进入 mongod
mongo

# 查看 db
show dbs

# 使用 ddbms 并查询 user 表
use ddbms
db.user.find()
```

Containers

Images

Volumes

Dev Environments

Extensions

Add Extensions

mongo mongo:4.4.14

Running

LogsInspectTerminalStats

MongoDB shell version v4.4.14
connecting to mongo:4.4.14/127.0.0.1:27017/?compressor=disabled&sslServiceName=mongo
Implicit session: session { "id" : UUID("e677e08b-46b2-468d-8551-791639ba565") }
MongoDB server version: 4.4.14
Welcome to the MongoDB shell.
For interactive help, type "help".
For more comprehensive documentation, see
https://docs.mongodb.com/
Questions? Try the MongoDB Developer Community Forums
https://community.mongodb.com/

The server generated these startup warnings when booting:
2022-11-27T14:13:45.984+08:00: Access control is not enabled for the database. Read and write access to data and configuration is unrestricted

mongo> show dbs
admin 0.000GB
config 0.001GB
dbsize 0.001GB
mongo> use dbsize
switched to db dbsize
mongo> db.user.find()
{ "_id" : ObjectId("6383756a393677ab59706568"), "timestamp" : "1506328859011", "id" : "u11", "uid" : "11", "name" : "user11", "gender" : "male", "email" : "email11", "phone" : "phone11", "dept" : "dept3", "grade" : "grade1", "language" : "e
n", "region" : "Hong Kong", "role" : "role1", "preferTags" : "tags10", "obtainedCredits" : 26 }
{ "_id" : ObjectId("6383756a393677ab59706568"), "timestamp" : "1506328859012", "id" : "u12", "uid" : "12", "name" : "user12", "gender" : "male", "email" : "email12", "phone" : "phone12", "dept" : "dept13", "grade" : "grade2", "language" : "e
n", "region" : "Hong Kong", "role" : "role0", "preferTags" : "tags33", "obtainedCredits" : 5 }
{ "_id" : ObjectId("6383756a393677ab5970656a"), "timestamp" : "1506328859014", "id" : "u14", "uid" : "14", "name" : "user14", "gender" : "male", "email" : "email14", "phone" : "phone14", "dept" : "dept12", "grade" : "grade3", "language" : "z
h", "region" : "Hong Kong", "role" : "role1", "preferTags" : "tags40", "obtainedCredits" : 66 }
{ "_id" : ObjectId("6383756a393677ab5970656b"), "timestamp" : "1506328859013", "id" : "u13", "uid" : "13", "name" : "user13", "gender" : "male", "email" : "email13", "phone" : "phone13", "dept" : "dept4", "grade" : "grade1", "language" : "z
h", "region" : "Hong Kong", "role" : "role1", "preferTags" : "tags91", "obtainedCredits" : 21 }
{ "_id" : ObjectId("6383756a393677ab5970656d"), "timestamp" : "1506328859016", "id" : "u16", "uid" : "16", "name" : "user16", "gender" : "female", "email" : "email16", "phone" : "phone16", "dept" : "dept4", "grade" : "grade1", "language" : "z
h", "region" : "Hong Kong", "role" : "role2", "preferTags" : "tags35", "obtainedCredits" : 85 }
{ "_id" : ObjectId("6383756a393677ab5970656e"), "timestamp" : "1506328859018", "id" : "u18", "uid" : "18", "name" : "user18", "gender" : "male", "email" : "email18", "phone" : "phone18", "dept" : "dept15", "grade" : "grade1", "language" : "z
h", "region" : "Hong Kong", "role" : "role0", "preferTags" : "tags49", "obtainedCredits" : 16 }
{ "_id" : ObjectId("6383756a393677ab5970656f"), "timestamp" : "1506328859019", "id" : "u19", "uid" : "19", "name" : "user19", "gender" : "male", "email" : "email19", "phone" : "phone19", "dept" : "dept8", "grade" : "grade1", "language" : "z
h", "region" : "Hong Kong", "role" : "role1", "preferTags" : "tags31", "obtainedCredits" : 10 }
{ "_id" : ObjectId("6383756a393677ab59706570"), "timestamp" : "1506328859020", "id" : "u20", "uid" : "20", "name" : "user20", "gender" : "male", "email" : "email20", "phone" : "phone20", "dept" : "dept17", "grade" : "grade2", "language" : "z
h", "region" : "Hong Kong", "role" : "role2", "preferTags" : "tags31", "obtainedCredits" : 69 }
{ "_id" : ObjectId("6383756a393677ab59706573"), "timestamp" : "1506328859023", "id" : "u23", "uid" : "23", "name" : "user23", "gender" : "female", "email" : "email23", "phone" : "phone23", "dept" : "dept19", "grade" : "grade2", "language" : "z
h", "region" : "Hong Kong", "role" : "role1", "preferTags" : "tags18", "obtainedCredits" : 8 }
{ "_id" : ObjectId("6383756a393677ab59706574"), "timestamp" : "1506328859024", "id" : "u24", "uid" : "24", "name" : "user24", "gender" : "male", "email" : "email24", "phone" : "phone24", "dept" : "dept7", "grade" : "grade4", "language" : "z
h", "region" : "Hong Kong", "role" : "role1", "preferTags" : "tags20", "obtainedCredits" : 13 }
{ "_id" : ObjectId("6383756a393677ab59706575"), "timestamp" : "1506328859025", "id" : "u5", "uid" : "5", "name" : "user5", "gender" : "male", "email" : "email5", "phone" : "phone5", "dept" : "dept10", "grade" : "grade2", "language" : "zh", "region" : "Hong Kong", "role" : "role1", "preferTags" : "tags6", "obtainedCredits" : 90 }
{ "_id" : ObjectId("6383756a393677ab59706576"), "timestamp" : "1506328859017", "id" : "u17", "uid" : "17", "name" : "user17", "gender" : "male", "email" : "email17", "phone" : "phone17", "dept" : "dept15", "grade" : "grade1", "language" : "z
h", "region" : "Hong Kong", "role" : "role2", "preferTags" : "tags31", "obtainedCredits" : 82 }
{ "_id" : ObjectId("6383756a393677ab59706577"), "timestamp" : "1506328859004", "id" : "u4", "uid" : "4", "name" : "user4", "gender" : "male", "email" : "email4", "phone" : "phone4", "dept" : "dept10", "grade" : "grade1", "language" : "zh", "region" : "Hong Kong", "role" : "role0", "preferTags" : "tags18", "obtainedCredits" : 11 }
{ "_id" : ObjectId("6383756a393677ab59706578"), "timestamp" : "1506328859003", "id" : "u3", "uid" : "3", "name" : "user3", "gender" : "male", "email" : "email3", "phone" : "phone3", "dept" : "dept2", "grade" : "grade2", "language" : "zh", "region" : "Hong Kong", "role" : "role2", "preferTags" : "tags34", "obtainedCredits" : 60 }
{ "_id" : ObjectId("6383756a393677ab59706579"), "timestamp" : "1506328859002", "id" : "u2", "uid" : "2", "name" : "user2", "gender" : "male", "email" : "email2", "phone" : "phone2", "dept" : "dept13", "grade" : "grade3", "language" : "zh", "region" : "Hong Kong", "role" : "role0", "preferTags" : "tags17", "obtainedCredits" : 77 }
{ "_id" : ObjectId("6383756a393677ab5970657f"), "timestamp" : "1506328859028", "id" : "u28", "uid" : "28", "name" : "user28", "gender" : "male", "email" : "email28", "phone" : "phone28", "dept" : "dept3", "grade" : "grade1", "language" : "e
n", "region" : "Hong Kong", "role" : "role0", "preferTags" : "tags20", "obtainedCredits" : 43 }
{ "_id" : ObjectId("6383756a393677ab59706584"), "timestamp" : "1506328859035", "id" : "u35", "uid" : "35", "name" : "user35", "gender" : "female", "email" : "email35", "phone" : "phone35", "dept" : "dept16", "grade" : "grade1", "language" : "z
h", "region" : "Hong Kong", "role" : "role0", "preferTags" : "tags38", "obtainedCredits" : 88 }
{ "_id" : ObjectId("6383756a393677ab59706585"), "timestamp" : "1506328859027", "id" : "u13", "uid" : "13", "name" : "user33", "gender" : "male", "email" : "email33", "phone" : "phone33", "dept" : "dept14", "grade" : "grade1", "language" : "e
n", "region" : "Hong Kong", "role" : "role0", "preferTags" : "tags17", "obtainedCredits" : 23 }
{ "_id" : ObjectId("6383756a393677ab59706586"), "timestamp" : "1506328859034", "id" : "u34", "uid" : "34", "name" : "user34", "gender" : "male", "email" : "email34", "phone" : "phone34", "dept" : "dept11", "grade" : "grade2", "language" : "z
h", "region" : "Hong Kong", "role" : "role1", "preferTags" : "tags19", "obtainedCredits" : 61 }
{ "_id" : ObjectId("6383756a393677ab59706587"), "timestamp" : "1506328859036", "id" : "u36", "uid" : "36", "name" : "user36", "gender" : "female", "email" : "email36", "phone" : "phone36", "dept" : "dept0", "grade" : "grade2", "language" : "z
h", "region" : "Hong Kong", "role" : "role0", "preferTags" : "tags28", "obtainedCredits" : 24 }
Type "help" for more
mongo>

查看数据分片

```
mongo> sh.status()
--- Sharding Status ---
  sharding version: {
    "_id" : 1,
    "minCompatibleVersion" : 5,
    "currentVersion" : 6,
    "clusterId" : ObjectId("6383754ed5180dab2a303515")
  }
  shards:
    { "_id" : "shardsvr_rsl", "host" : "shardsvr_rsl/shardsvr1:27018,shardsvr2:27018", "state" : 1, "tags" : [ { "DBMS1", "DBMS1-2" } ] }
    { "_id" : "shardsvr_rs2", "host" : "shardsvr_rs2/shardsvr3:27018,shardsvr4:27018", "state" : 1, "tags" : [ "DBMS2", "DBMS1-2" ] }
  active mongoses:
    "4.4.14" : 1
  autosplit:
    Currently enabled: yes
  balancer:
    Currently enabled: yes
    Currently running: no
    Failed balancer rounds in last 5 attempts: 0
    Migration Results for the last 24 hours:
      No recent migrations
  databases:
    { "_id" : "config", "primary" : "config", "partitioned" : true }
    { "_id" : "ddbms", "primary" : "shardsvr_rsl", "partitioned" : true, "version" : { "uuid" : UUID("51ad99c4-def4-41d0-a1ld-8e6be6ed9e76"), "lastMod" : 1 } }
      ddbms.article
        shard keys: { "category" : 1, "aid" : 1 }
        unique: false
        balancing: true
        chunks:
          shardsvr_rsl 3
          shardsvr_rs2 2
        { "category" : { "$minKey" : 1 }, "aid" : { "$minKey" : 1 } }--> { "category" : "science", "aid" : { "$minKey" : 1 } } on : shardsvr_rsl Timestamp(1, 0)
        { "category" : "science", "aid" : { "$minKey" : 1 } }--> { "category" : "science", "aid" : { "$maxKey" : 1 } } on : shardsvr_rsl Timestamp(1, 1)
        { "category" : "science", "aid" : { "$maxKey" : 1 } }--> { "category" : "technology", "aid" : { "$minKey" : 1 } } on : shardsvr_rs2 Timestamp(1, 2)
        { "category" : "technology", "aid" : { "$minKey" : 1 } }--> { "category" : "technology", "aid" : { "$maxKey" : 1 } } on : shardsvr_rs2 Timestamp(1, 3)
        { "category" : "technology", "aid" : { "$maxKey" : 1 } }--> { "category" : "science", "aid" : { "$maxKey" : 1 } } on : shardsvr_rsl Timestamp(1, 4)
        tag: DBMS1-2 { "category" : "science", "aid" : { "$minKey" : 1 } }--> { "category" : "science", "aid" : { "$maxKey" : 1 } }
        tag: DBMS2 { "category" : "technology", "aid" : { "$minKey" : 1 } }--> { "category" : "technology", "aid" : { "$maxKey" : 1 } }
      ddbms.beread
        shard keys: { "category" : 1, "_id" : 1 }
        unique: false
        balancing: true
        chunks:
          shardsvr_rsl 3
          shardsvr_rs2 2
        { "category" : { "$minKey" : 1 }, "_id" : { "$minKey" : 1 } }--> { "category" : "science", "_id" : { "$minKey" : 1 } } on : shardsvr_rsl Timestamp(1, 0)
        { "category" : "science", "_id" : { "$minKey" : 1 } }--> { "category" : "science", "_id" : { "$maxKey" : 1 } } on : shardsvr_rsl Timestamp(1, 1)
        { "category" : "science", "_id" : { "$maxKey" : 1 } }--> { "category" : "technology", "_id" : { "$minKey" : 1 } } on : shardsvr_rs2 Timestamp(1, 2)
        { "category" : "technology", "_id" : { "$minKey" : 1 } }--> { "category" : "technology", "_id" : { "$maxKey" : 1 } } on : shardsvr_rs2 Timestamp(1, 3)
        { "category" : "technology", "_id" : { "$maxKey" : 1 } }--> { "category" : "science", "_id" : { "$maxKey" : 1 } } on : shardsvr_rsl Timestamp(1, 4)
        tag: DBMS1-2 { "category" : "science", "_id" : { "$minKey" : 1 } }--> { "category" : "science", "_id" : { "$maxKey" : 1 } }
        tag: DBMS2 { "category" : "technology", "_id" : { "$minKey" : 1 } }--> { "category" : "technology", "_id" : { "$maxKey" : 1 } }
      ddbms.rank
        shard keys: { "temporal_granularity" : 1, "_id" : 1 }
        unique: false
        balancing: true
        chunks:
          shardsvr_rsl 3
          shardsvr_rs2 4
        { "temporal_granularity" : { "$minKey" : 1 }, "_id" : { "$minKey" : 1 } }--> { "temporal_granularity" : "daily", "_id" : { "$minKey" : 1 } } on : shardsvr_rsl Timestamp(1, 0)
        { "temporal_granularity" : "daily", "_id" : { "$minKey" : 1 } }--> { "temporal_granularity" : "daily", "_id" : { "$maxKey" : 1 } } on : shardsvr_rsl Timestamp(1, 1)
        { "temporal_granularity" : "daily", "_id" : { "$maxKey" : 1 } }--> { "temporal_granularity" : "monthly", "_id" : { "$minKey" : 1 } } on : shardsvr_rs2 Timestamp(1, 2)
        { "temporal_granularity" : "monthly", "_id" : { "$minKey" : 1 } }--> { "temporal_granularity" : "monthly", "_id" : { "$maxKey" : 1 } } on : shardsvr_rs2 Timestamp(1, 3)
```

Hadoop

访问 localhost:9097 可以看到数据成功导入

Browse Directory

/articles/articlesGo!

Show25entriesSearch:

<input type="checkbox"/>	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	
<input type="checkbox"/>	-rw-f--f--	root	supergroup	12 KB	Nov 27 21:18	3	128 MB	.DS_Store	
<input type="checkbox"/>	drwxr-xr-x	root	supergroup	0 B	Nov 27 21:18	0	0 B	article0	
<input type="checkbox"/>	drwxr-xr-x	root	supergroup	0 B	Nov 27 21:18	0	0 B	article1	
<input type="checkbox"/>	drwxr-xr-x	root	supergroup	0 B	Nov 27 21:18	0	0 B	article10	
<input type="checkbox"/>	drwxr-xr-x	root	supergroup	0 B	Nov 27 21:18	0	0 B	article11	
<input type="checkbox"/>	drwxr-xr-x	root	supergroup	0 B	Nov 27 21:18	0	0 B	article12	
<input type="checkbox"/>	drwxr-xr-x	root	supergroup	0 B	Nov 27 21:18	0	0 B	article13	
<input type="checkbox"/>	drwxr-xr-x	root	supergroup	0 B	Nov 27 21:18	0	0 B	article14	
<input type="checkbox"/>	drwxr-xr-x	root	supergroup	0 B	Nov 27 21:18	0	0 B	article15	
<input type="checkbox"/>	drwxr-xr-x	root	supergroup	0 B	Nov 27 21:18	0	0 B	article16	
<input type="checkbox"/>	drwxr-xr-x	root	supergroup	0 B	Nov 27 21:18	0	0 B	article17	
<input type="checkbox"/>	drwxr-xr-x	root	supergroup	0 B	Nov 27 21:18	0	0 B	article18	
<input type="checkbox"/>	drwxr-xr-x	root	supergroup	0 B	Nov 27 21:18	0	0 B	article19	
<input type="checkbox"/>	drwxr-xr-x	root	supergroup	0 B	Nov 27 21:18	0	0 B	article2	
<input type="checkbox"/>	drwxr-xr-x	root	supergroup	0 B	Nov 27 21:18	0	0 B	article20	