



链滴

ClickHouse 初体验

作者: [flowaters](#)

原文链接: <https://ld246.com/article/1511510837158>

来源网站: [链滴](#)

许可协议: [署名-相同方式共享 4.0 国际 \(CC BY-SA 4.0\)](#)

追求更好的体验，本文采用docker方式。

环境部署

详细可参考[server官方docker镜像](#)和[client官方docker镜像](#)

启动Server

```
$ docker run -d --name some-clickhouse-server --ulimit nofile=262144:262144 yandex/clickhouse-server
```

启动Client

```
$ docker run -it --rm --link some-clickhouse-server:clickhouse-server yandex/clickhouse-client --host clickhouse-server
```

附一：启动自定义Server

如果需要自定义Server的启动参数，则：

```
$ docker run -d --name some-clickhouse-server --ulimit nofile=262144:262144 -v /path/to/your/config.xml:/etc/clickhouse-server/config.xml yandex/clickhouse-server
```

附二：其它安装包

参考[官方文档](#)

- Ubuntu: https://clickhouse.yandex/docs/en/getting_started/index.html
- From Source: https://clickhouse.yandex/docs/en/getting_started/index.html
- CentOS rpm: <https://github.com/Altinity/clickhouse-rpm-install>
- Gentoo : <https://github.com/kmeaw/clickhouse-overlay>

基本SQL

select 1

```
;) select 1
```

```
SELECT 1
```

```
┌───┐  
│ 1 │  
└───┘
```

```
1 rows in set. Elapsed: 0.004 sec.
```

select now

```
:) select now()
```

```
SELECT now()
```

```
now()
2017-11-24 02:34:19 |
```

1 rows in set. Elapsed: 0.003 sec.

AMPLab测试

如[官方文档](#),

下载数据

1. 申请S3账号：会有两笔预授权的1\$账单，据说3-5天后会返还。
2. 下载数据

```
s3cmd sync s3://big-data-benchmark/pavlo/text-deflate/tiny/ /User/abeffect/tmp/tiny/
```

建表

```
CREATE TABLE rankings_tiny
(
  pageURL String,
  pageRank UInt32,
  avgDuration UInt32
) ENGINE = Log;
```

```
CREATE TABLE uservisits_tiny
(
  sourceIP String,
  destinationURL String,
  visitDate Date,
  adRevenue Float32,
  userAgent String,
  cCode FixedString(3),
  lCode FixedString(6),
  searchWord String,
  duration UInt32
) ENGINE = MergeTree(visitDate, visitDate, 8192);
```

导入数据

```
#!/bin/sh
```

```
for i in tiny/rankings/*.deflate; do
  echo $i;
```

```

printf "\x1f\x8b\x08\x00\x00\x00\x00" | cat - $i | gzip -dc > ${i}.tmp
cat ${i}.tmp | wc -l
cat ${i}.tmp | clickhouse-client --host server-ip --query="INSERT INTO rankings_tiny FORM
T CSV";
done

```

```

for i in tiny/uservisits/*.deflate; do
  echo $i;
  printf "\x1f\x8b\x08\x00\x00\x00\x00" | cat - $i | gzip -dc > ${i}.tmp
  cat ${i}.tmp | clickhouse-client --host server-ip --query="INSERT INTO uservisits_tiny FOR
AT CSV";
done

```

查询

```

SELECT
  pageURL,
  pageRank
FROM rankings_tiny
WHERE pageRank > 300

```

pageURL	pageRank
nbizrgdziebsaecsecujfcqtnpcnxxwiopmddorcxnlijdzgoi	665
nbizrgdziebsaecsecujfcqtnpcnxxwiopmddorcxnlijdzgoi	665
dfqlfhuuwxxhhjvdtwljahxbwhqkutigdfcuiogfooprtmmomeyysquqlzvzmrxhktrbumglgwixsovxo isibirdlgtc	483
pucpwgjykbcelcevdzmzuaohojzjxgmpgojfehhabasvzbzpczqcubfbfekv	397
ffygkvsklpmup	332

5 rows in set. Elapsed: 0.007 sec. Processed 1.32 thousand rows, 90.28 KB (189.33 thousand r ws/s., 12.95 MB/s.)

```

SELECT
  substring(sourceIP, 1, 8),
  sum(adRevenue)
FROM uservisits_tiny
GROUP BY substring(sourceIP, 1, 8)
LIMIT 5

```

substring(sourceIP, 1, 8)	sum(adRevenue)
227.120.	0.7206825017929077
33.157.1	0.9653052091598511
134.195.	0.08809900283813477
172.239.	0.2393447756767273
19.26.11	0.13361740112304688

5 rows in set. Elapsed: 0.102 sec. Processed 10.00 thousand rows, 262.95 KB (98.18 thousand rows/s., 2.58 MB/s.)

```
SELECT
  sourceIP,
  sum(adRevenue) AS totalRevenue,
  avg(pageRank) AS pageRank
FROM rankings_tiny
ALL INNER JOIN
(
  SELECT
    sourceIP,
    destinationURL AS pageURL,
    adRevenue
  FROM uservisits_tiny
  WHERE (visitDate > '1980-01-01') AND (visitDate < '1980-04-01')
) USING (pageURL)
GROUP BY sourceIP
ORDER BY totalRevenue DESC
LIMIT 1
```

sourceIP	totalRevenue	pageRank
242.170.245.71	1.132317066192627	26

1 rows in set. Elapsed: 0.009 sec. Processed 1.38 thousand rows, 95.39 KB (150.26 thousand rows/s., 10.41 MB/s.)

进一步阅读

1. [DB-Engines Ranking](#): 目前上升非常快
2. [Berkeley BigData Benchmark](#): 相关OLAP产品的评测
3. [clickhouse rpm install](#): 安装rpm格式的client和server, 文中测试部分的client即通过此方法安装
4. [clickhouse jdbc](#): clickhouse的jdbc驱动
5. [tabix.ui](#): clickhouse gui界面, 无需安装, 直接进入
6. [ClickHouse之简单性能测试](#): 美国民用航班的数据, 从1987年到2017年, 有1.7亿条的测试case
7. [ClickHouse vs Amazon RedShift Benchmark](#)
8. 常见的benchmark:
 1. HiBench: Intel开发, 针对Hadoop
 2. Berkeley BigData Bench: AMPLab开发, 针对Spark
 3. Hadoop GridMix: Hadoop自带
 4. BigBench: Teradata、多伦多大学、InfoSizing、Oracle开发
 5. BigDataBenchmark: 中科院研发
 6. TPC-DS: 针对SQL on Hadoop

7. 其它Malstone、Cloud Harmony、YCSB、SWIM、LinkBench、DFSIO、Hive performance Benchmark(Pavlo)等等