pg14安装citus

测试lz4

**目录**

[一、rejoin方式恢复 1](#_Toc1675008602)

[3.1 查看主库状态和进程信息 1](#_Toc506445291)

[3.2 将新主archive拷贝到老主库 1](#_Toc1348312776)

[3.3 旧主执行rejoin预操作 1](#_Toc845383088)

[3.4 旧主执行rejoin正式操作 1](#_Toc601751464)

[3.5 旧主库启动ltclusterd进程 1](#_Toc1136361725)

[二、主备切换 2](#_Toc1259439304)

[3.6 旧主执行预切换操作 2](#_Toc1797557496)

[3.7 旧主执行切换操作 2](#_Toc748889276)

[3.8 验证集群状态 4](#_Toc180406665)

[三、 更换vip操作 4](#_Toc1995392738)

[3.9 停掉主库和备库的keepalived 4](#_Toc1461116014)

[3.10 编辑keepalived.conf文件（主备都执行） 4](#_Toc501343853)

[3.11 root重新启动keepalived进程 4](#_Toc1507790190)

[四、 clone方式恢复主库(主库异常宕机场景) 4](#_Toc1122688730)

[3.12 旧主库执行clone欲检测 4](#_Toc1238162568)

[3.13 旧主执行clone操作 5](#_Toc681740946)

[3.14 旧主库启动 5](#_Toc1194822677)

[五、 高可用主库维护重启操作 5](#_Toc265149242)

[3.15 备库停掉keepalived 5](#_Toc334742769)

[3.16 主库重启，需要在lightdb用户下执行 5](#_Toc1762047090)

[3.17 查看集群状态，确认primary的Paused?状态为yes 6](#_Toc925949500)

[3.18 主库停止数据库实例 6](#_Toc1766740338)

[3.19 修改数据库参数 6](#_Toc348473697)

[3.20 启动主库 6](#_Toc609520110)

[3.21 恢复主库ltclusterd 6](#_Toc707492580)

[3.22 查看集群状态，确认primary的Paused?状态为no 6](#_Toc210838621)

修订记录

|  |  |  |  |
| --- | --- | --- | --- |
| 日期 | 修订版本 | 描述 | 作者 |
| 2022/05/07 | V1.0 | LightDB高可用主库宕机恢复流程 | 姚崇 |

审阅

|  |  |
| --- | --- |
| 姓名 | 职位 |
|  |  |
|  |  |

分发

|  |  |
| --- | --- |
| 姓名 | 单位 |
|  |  |
|  |  |

# 一、安装PG14

## 下载源码包编译

下载源码包，解压

cd /home/lightdb/yc/postgresql-14.4/

./configure --prefix=/home/lightdb/pgsqlcn --enable-nls='gbk' --with-openssl --with-libxml --with-zlib --with-uuid=ossp --without-ldap --with-segsize=32 --with-lz4

#然后执行

make -sj && make install

## 指定环境变量

export PG\_PORT=10002

export PGUSER=lightdb

export PGHOME=/home/lightdb/pgsqlcn

export PGDATA=/data1/cn14

export PATH=${PGHOME}/bin:${PGHOME}/tools/bin:${PGHOME}/tools/sbin:${PATH}

export LD\_LIBRARY\_PATH=${PGHOME}/lib:${PGHOME}/lib/ltext:${PGHOME}/tools/lib64:${LD\_LIBRARY\_PATH}

export PGPORT=10002

export PGUSER=lightdb

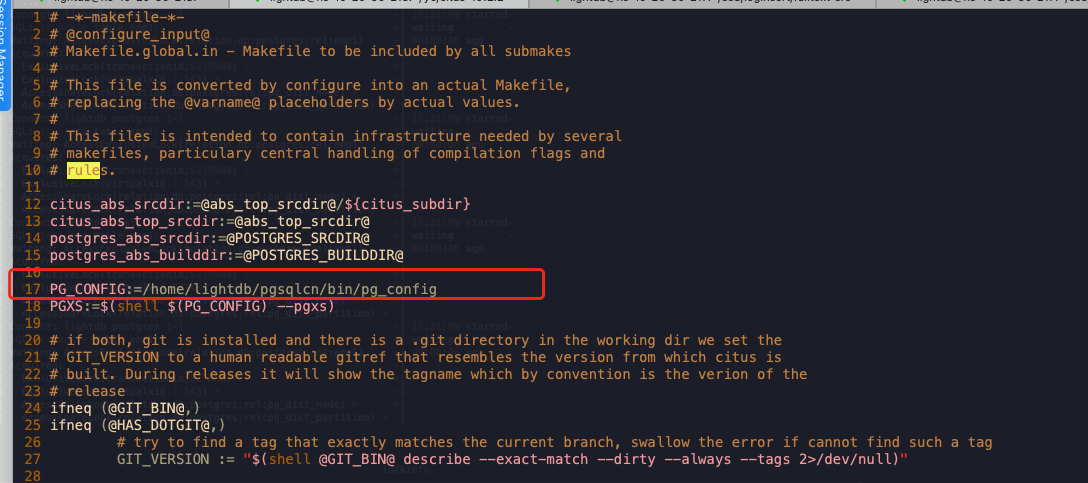
export PGHOST=10.20.137.201

## 编译安装citus（需要依赖curl-devel）

cd citus-10.2.2/

./configure --without-lz4 --without-zstd

修改Makefile.global，将PG\_CONFIG修改为pg14的pg\_config目录



make && make install

## initdb创建数据库

# Source global definitions

if [ -f /etc/bashrc ]; then

. /etc/bashrc

fi

export PG\_PORT=1111

export PGUSER=lightdb

export PGHOME=/home/lightdb/pgsqlcn

export PGDATA=/data1/14dn1

export PATH=${PGHOME}/bin:${PGHOME}/tools/bin:${PGHOME}/tools/sbin:${PATH}

export LD\_LIBRARY\_PATH=${PGHOME}/lib:${PGHOME}/lib/ltext:${PGHOME}/tools/lib64:${LD\_LIBRARY\_PATH}

export PGPORT=1111

export PGUSER=lightdb

export PGHOST=10.20.137.201

指定如上环境变量后source ~/.bashrc14dn1 直接initdb创建数据库，其余2个DN以此类推

initdb

## 配置postgresql.conf和pg\_hba.conf

listen\_addresses = '\*'

effective\_cache\_size=272GB

wal\_buffers=128MB

max\_wal\_size=174036MB

min\_wal\_size=58012MB

random\_page\_cost=1.0

max\_parallel\_maintenance\_workers=12

default\_statistics\_target=256

min\_parallel\_index\_scan\_size=128MB

transform\_null\_equals=on

max\_worker\_processes=96

max\_parallel\_workers=96

logging\_collector=on

enable\_partitionwise\_aggregate=on

shared\_buffers=48GB

bgwriter\_lru\_maxpages=4572

max\_slot\_wal\_keep\_size=174036MB

parallel\_setup\_cost=10000

lt\_stat\_statements.track\_planning=on

track\_io\_timing=on

log\_min\_messages=info

temp\_buffers=64MB

commit\_siblings=10

canopy.shard\_count=48

min\_parallel\_table\_scan\_size=2GB

wal\_sync\_method=fdatasync

fsync=off

full\_page\_writes=off

maintenance\_work\_mem = 2048MB

work\_mem=2048MB

commit\_delay=10

commit\_siblings=10

synchronous\_commit=off

shared\_preload\_libraries='citus'

gin\_pending\_list\_limit=2048MB

default\_toast\_compression=lz4

## 启动数据库并创建集群

#所有节点执行create extension citus;

create extension citus;

#下面cn执行, 构建1个CN 3个DN节点

select master\_add\_node('127.0.0.1',1111);

select master\_add\_node('127.0.0.1',2222);

select master\_add\_node('127.0.0.1',3333);

# 创建表并指定lz4压缩

CREATE TABLE lem\_db\_log (

id serial NOT NULL ,

db\_log\_time timestamp(6) not null,

db\_log\_message text COLLATE pg\_catalog.default,

log\_tsv tsvector NOT NULL DEFAULT ''

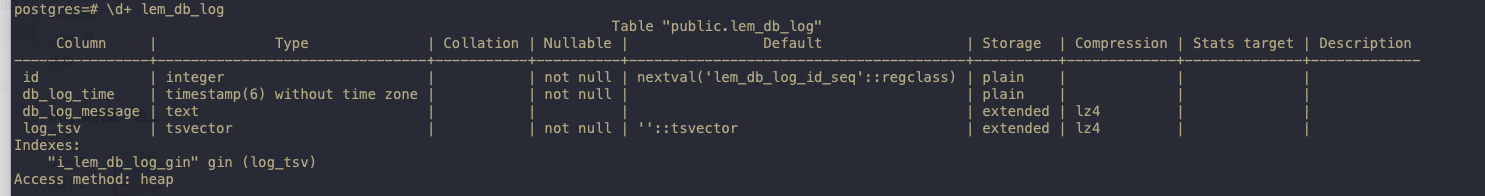
);

create index i\_lem\_db\_log\_gin on lem\_db\_log using gin(log\_tsv);

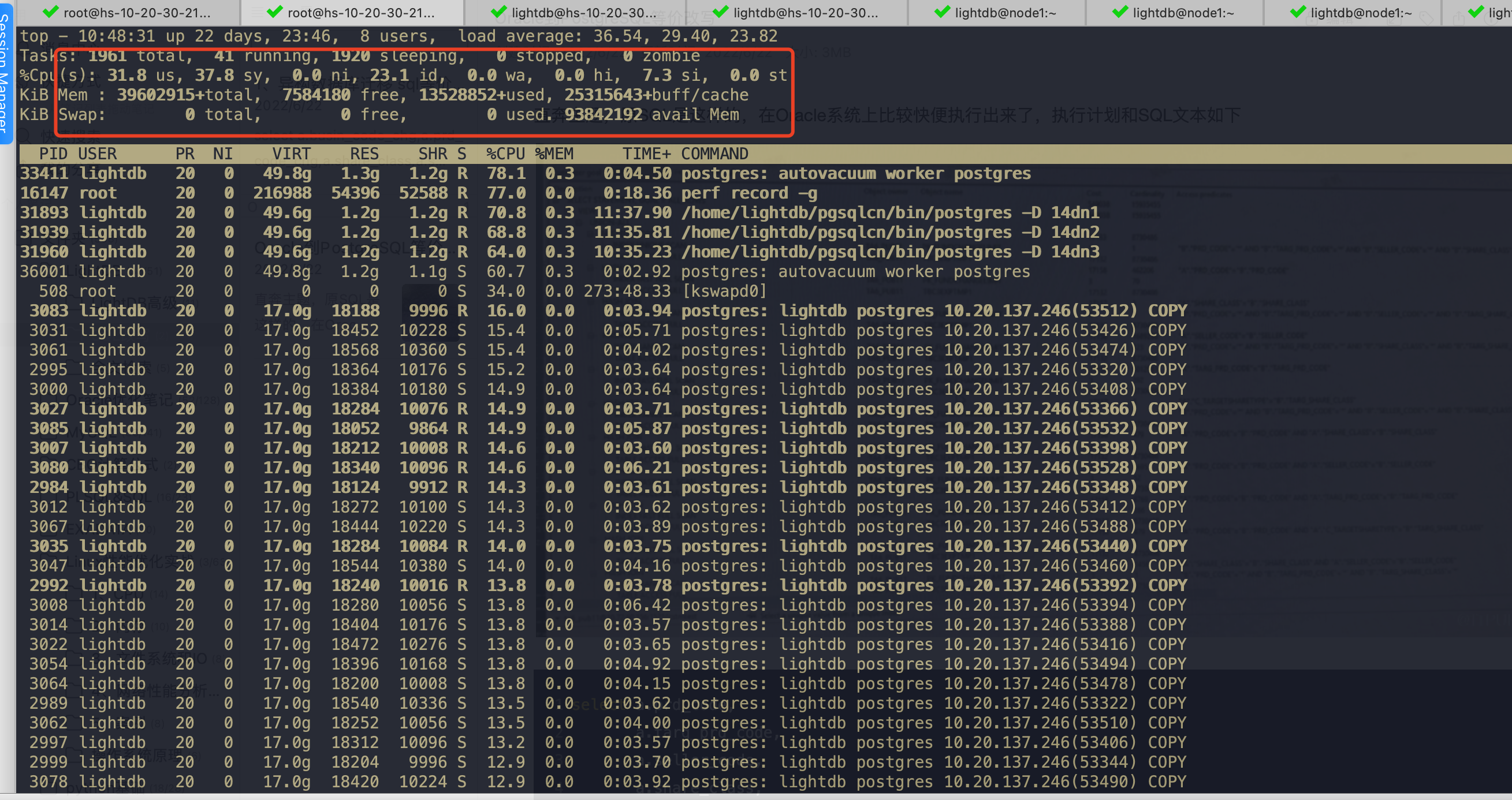
SELECT create\_distributed\_table('lem\_db\_log', 'id');

ALTER TABLE lem\_db\_log ALTER COLUMN db\_log\_message SET COMPRESSION lz4;

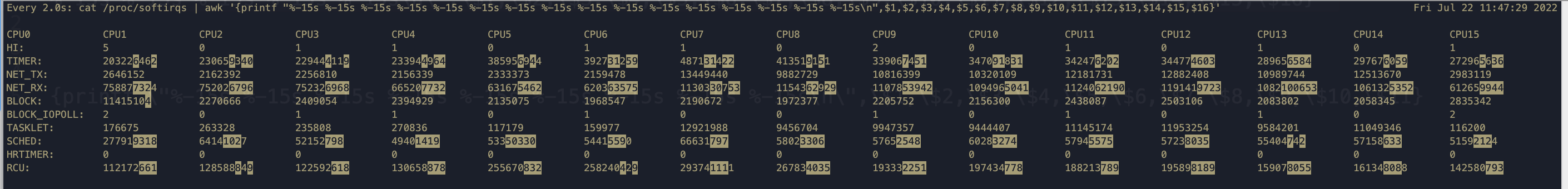
ALTER TABLE lem\_db\_log ALTER COLUMN log\_tsv SET COMPRESSION lz4;



# 测试copy入库



watch -d "cat /proc/softirqs | awk '{printf \"%-15s %-15s %-15s %-15s %-15s %-15s %-15s %-15s %-15s %-15s %-15s %-15s %-15s %-15s %-15s %-15s\n\",\$1,\$2,\$3,\$4,\$5,\$6,\$7,\$8,\$9,\$10,\$11,\$12,\$13,\$14,\$15,\$16}'"



## 回环地址流量很高，没个CPU软中断都很高

# sar -n DEV 2 222

01:28:00 PM IFACE rxpck/s txpck/s rxkB/s txkB/s rxcmp/s txcmp/s rxmcst/s

01:28:02 PM p7p1 102.00 110.00 10.91 32.25 0.00 0.00 0.00

01:28:02 PM p7p2 0.00 0.00 0.00 0.00 0.00 0.00 0.00

01:28:02 PM p3p1 59461.00 4123.50 84956.00 272.45 0.00 0.00 0.00

01:28:02 PM p3p2 0.00 0.00 0.00 0.00 0.00 0.00 0.00

01:28:02 PM lo 74326.00 74326.00 75346.15 75346.15 0.00 0.00 0.00

01:28:02 PM p5p1 0.00 0.00 0.00 0.00 0.00 0.00 0.00

01:28:02 PM p5p2 0.00 0.00 0.00 0.00 0.00 0.00 0.00

01:28:02 PM em1 0.00 0.00 0.00 0.00 0.00 0.00 0.00

01:28:02 PM em2 0.00 0.00 0.00 0.00 0.00 0.00 0.00

01:28:02 PM em4 0.00 0.00 0.00 0.00 0.00 0.00 0.00

01:28:02 PM em3 0.00 0.00 0.00 0.00 0.00 0.00 0.00

01:28:02 PM IFACE rxpck/s txpck/s rxkB/s txkB/s rxcmp/s txcmp/s rxmcst/s

01:28:04 PM p7p1 1728.00 1723.50 206.59 209.77 0.00 0.00 0.00

01:28:04 PM p7p2 0.00 0.00 0.00 0.00 0.00 0.00 0.00

01:28:04 PM p3p1 114671.00 8771.00 163869.19 574.07 0.00 0.00 0.00

01:28:04 PM p3p2 0.00 0.00 0.00 0.00 0.00 0.00 0.00

01:28:04 PM lo 93412.00 93412.00 165984.80 165984.80 0.00 0.00 0.00

01:28:04 PM p5p1 0.00 0.00 0.00 0.00 0.00 0.00 0.00

01:28:04 PM p5p2 0.00 0.00 0.00 0.00 0.00 0.00 0.00

01:28:04 PM em1 0.00 0.00 0.00 0.00 0.00 0.00 0.00

01:28:04 PM em2 0.00 0.00 0.00 0.00 0.00 0.00 0.00

01:28:04 PM em4 0.00 0.00 0.00 0.00 0.00 0.00 0.00

01:28:04 PM em3 0.00 0.00 0.00 0.00 0.00 0.00 0.00

01:28:04 PM IFACE rxpck/s txpck/s rxkB/s txkB/s rxcmp/s txcmp/s rxmcst/s

01:28:06 PM p7p1 1477.50 1470.50 211.65 222.49 0.00 0.00 0.00

01:28:06 PM p7p2 0.00 0.00 0.00 0.00 0.00 0.00 0.00

01:28:06 PM p3p1 114876.50 9188.50 164097.17 599.04 0.00 0.00 0.00

01:28:06 PM p3p2 0.00 0.00 0.00 0.00 0.00 0.00 0.00

01:28:06 PM lo 92468.50 92468.50 176905.52 176905.52 0.00 0.00 0.00

01:28:06 PM p5p1 0.00 0.00 0.00 0.00 0.00 0.00 0.00

01:28:06 PM p5p2 0.00 0.00 0.00 0.00 0.00 0.00 0.00

01:28:06 PM em1 0.00 0.00 0.00 0.00 0.00 0.00 0.00

01:28:06 PM em2 0.00 0.00 0.00 0.00 0.00 0.00 0.00

01:28:06 PM em4 0.00 0.00 0.00 0.00 0.00 0.00 0.00

01:28:06 PM em3 0.00 0.00 0.00 0.00 0.00 0.00 0.00

###如下每个CPU都很高, 96逻辑CPU中间部分略，%soft都维持在20%左右

01:34:45 PM CPU %usr %nice %sys %iowait %irq %soft %steal %guest %gnice %idle

01:34:46 PM all 16.63 0.00 44.29 0.00 0.00 18.15 0.00 0.00 0.00 20.93

01:34:46 PM 0 12.37 0.00 45.36 0.00 0.00 24.74 0.00 0.00 0.00 17.53

01:34:46 PM 1 19.00 0.00 44.00 0.00 0.00 17.00 0.00 0.00 0.00 20.00

01:34:46 PM 2 18.18 0.00 40.40 0.00 0.00 23.23 0.00 0.00 0.00 18.18

01:34:46 PM 3 21.21 0.00 44.44 0.00 0.00 19.19 0.00 0.00 0.00 15.15

01:34:46 PM 4 14.58 0.00 43.75 0.00 0.00 19.79 0.00 0.00 0.00 21.88

01:34:46 PM 5 17.53 0.00 47.42 0.00 0.00 16.49 0.00 0.00 0.00 18.56

01:34:46 PM 6 13.40 0.00 45.36 0.00 0.00 21.65 0.00 0.00 0.00 19.59

... ...

01:34:46 PM 88 14.29 0.00 39.80 0.00 0.00 19.39 0.00 0.00 0.00 26.53

01:34:46 PM 89 18.95 0.00 45.26 0.00 0.00 13.68 0.00 0.00 0.00 22.11

01:34:46 PM 90 20.62 0.00 41.24 0.00 0.00 15.46 0.00 0.00 0.00 22.68

01:34:46 PM 91 14.29 0.00 44.90 0.00 0.00 15.31 0.00 0.00 0.00 25.51

01:34:46 PM 92 15.79 0.00 41.05 0.00 0.00 14.74 0.00 0.00 0.00 28.42

01:34:46 PM 93 24.24 0.00 31.31 0.00 0.00 9.09 0.00 0.00 0.00 35.35

01:34:46 PM 94 12.24 0.00 51.02 0.00 0.00 13.27 0.00 0.00 0.00 23.47

01:34:46 PM 95 10.20 0.00 52.04 0.00 0.00 17.35 0.00 0.00 0.00 20.41

## 怀疑citus设置问题

更换节点

select master\_remove\_node('127.0.0.1',1111);

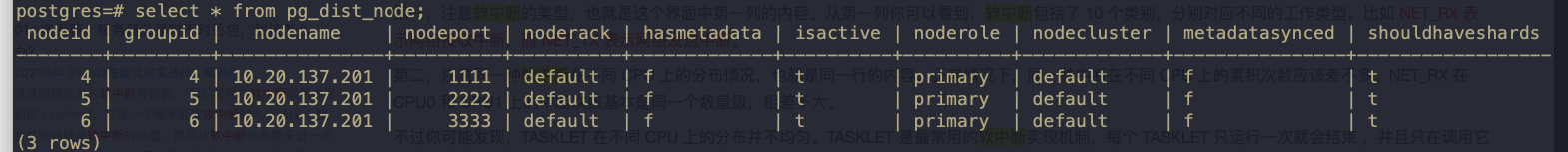
select master\_remove\_node('127.0.0.1',2222);

select master\_remove\_node('127.0.0.1',3333);

select master\_add\_node('10.20.137.201',1111);

select master\_add\_node('10.20.137.201',2222);

select master\_add\_node('10.20.137.201',3333);



CREATE TABLE lem\_db\_log (

id serial NOT NULL ,

db\_log\_time timestamp(6) not null,

db\_log\_message text COLLATE pg\_catalog.default,

log\_tsv tsvector NOT NULL DEFAULT ''

);

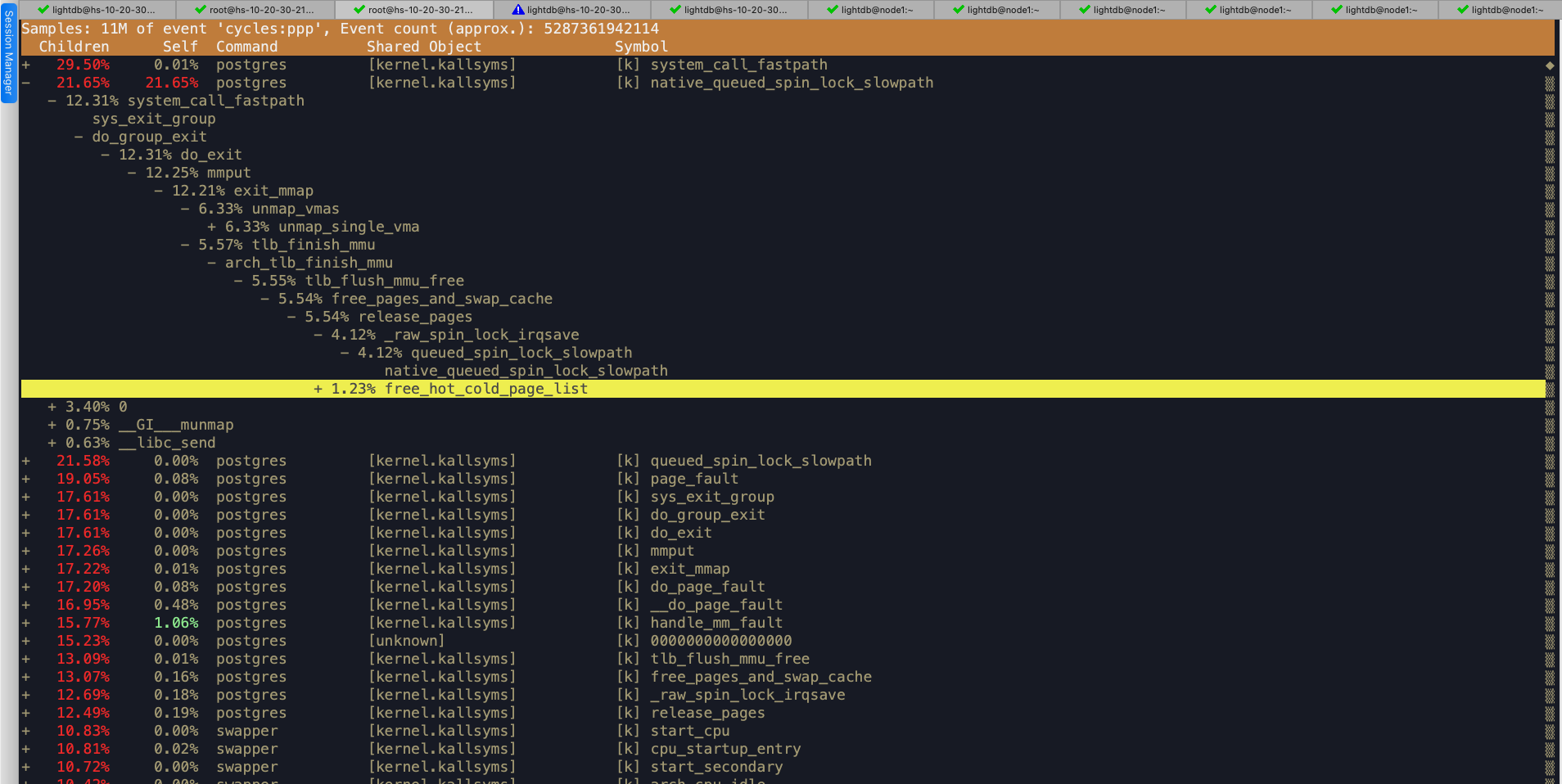
create index i\_lem\_db\_log\_gin on lem\_db\_log using gin(log\_tsv);

SELECT create\_distributed\_table('lem\_db\_log', 'id');

ALTER TABLE lem\_db\_log ALTER COLUMN db\_log\_message SET COMPRESSION lz4;

ALTER TABLE lem\_db\_log ALTER COLUMN log\_tsv SET COMPRESSION lz4;

## perf record -g



## 关闭ssl

alter system set ssl=off;

alter system set ssl\_ciphers = 'HIGH:MEDIUM:+3DES:!aNULL';

alter system set citus.node\_conninfo = 'sslmode=prefer';

##Postgresql.conf

citus.enable\_repartition\_joins=true

citus.node\_connection\_timeout = 6000000

citus.max\_intermediate\_result\_size = -1

citus.node\_conninfo='sslmode=prefer'

citus.max\_cached\_conns\_per\_worker=2

citus.executor\_slow\_start\_interval=5ms

ssl=off

ssl\_ciphers = 'HIGH:MEDIUM:+3DES:!aNULL'

也不行

## 原因就是COPY不能提高高并发