PingCAP Education

数据库 OOM 问题诊断及处理之 TiDB Server

练习：概述

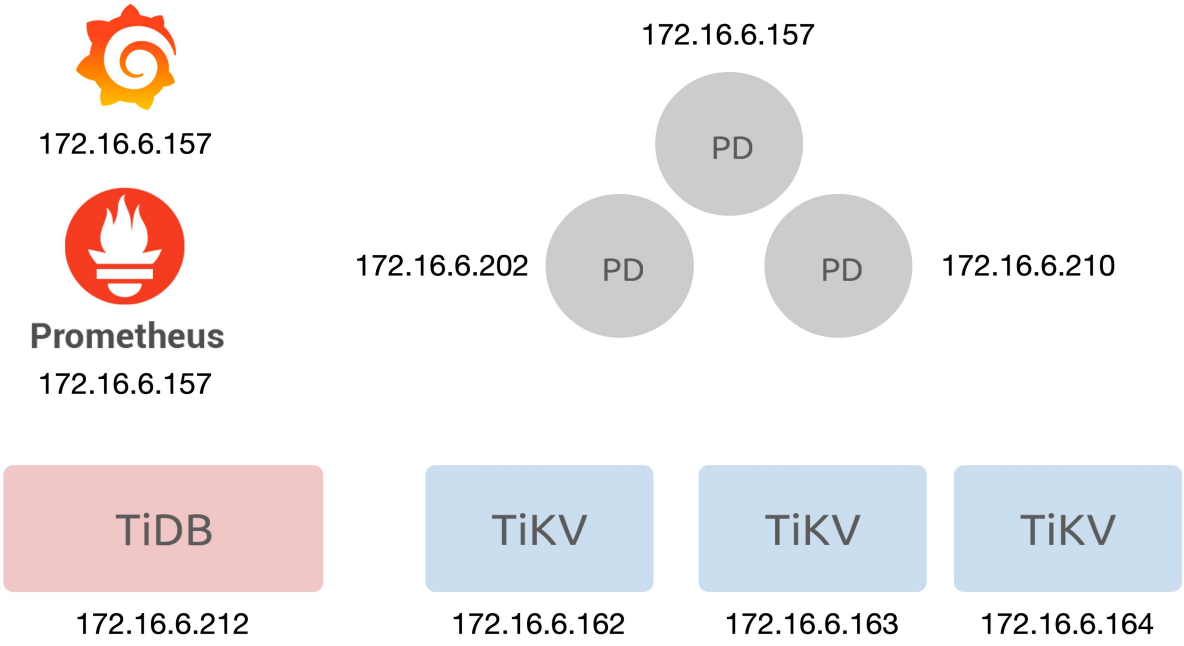
概述

在本课练习中，您将在 TiDB 数据库中运行耗内存较大的 SQL 语句，观察其对 TiDB Server 内存使用情况的影响，对其进行监控和限制。

实验环境要求

1. 注意，您的实验环境（包括 IP，端口号，用户名，密码和目录）可能和手册不同。
2. 可以连接到实验用虚拟机。

拓扑结构介绍



练习：消耗内存较大 SQL 语句对于 TiDB Server 的影响

1. 初始化环境

使用 TiUP bench 组件构造包括 customer、lineitem、nation 等在内的基础表结构以及数据：

```
# tiup bench tpch -H 172.16.6.212 -P 4000 -D tpch prepare

Starting component `bench`: /root/.tiup/components/bench/v1.6.1/tiup-bench tpch -H 172.16.6.212 -P 4000 -D tpch
prepare
creating nation
creating region
creating part
creating supplier
creating partsupp
creating customer
creating orders
creating lineitem
generating nation table
generate nation table done
generating region table
generate region table done
generating customers table
generate customers table done
generating suppliers table
generate suppliers table done
generating part/partsupplier tables
generate part/partsupplier tables done
generating orders/lineitem tables
generate orders/lineitem tables done
Finished

# mysql -h 172.16.6.212 -P4000 -uroot

mysql> show tables;
+------------------+
| Tables_in_tpch   |
+------------------+
| customer         |
| customer_test    |
| lineitem         |
| nation           |
| orders           |
| part             |
| partsupp         |
| region           |
| supplier         |
+------------------+
9 rows in set (0.00 sec)

mysql> select count(*) from customer;
+----------+
| count(*) |
+----------+
|   150000 |
+----------+
1 row in set (0.30 sec)
```

练习：消耗内存较大 SQL 语句对于 TiDB Server 的影响

```
# mysql -h 172.16.6.212 -P4000 -uroot

mysql> create table customer_test(
-> `C_CUSTKEY` bigint(20) PRIMARY KEY AUTO_INCREMENT,
-> `C_NAME` varchar(25) NOT NULL,
-> `C_ADDRESS` varchar(40) NOT NULL,
-> `C_NATIONKEY` bigint(20) NOT NULL,
-> `C_PHONE` char(15) NOT NULL,
-> `C_ACCTBAL` decimal(15,2) NOT NULL,
-> `C_MKTSEGMENT` char(10) NOT NULL,
-> `C_COMMENT` varchar(117) NOT NULL,
-> KEY `idx_cname` (`C_NAME`)
-> );
Query OK, 0 rows affected (0.11 sec)

# for i in `seq 60`; do mysql -uroot -P4000 -h172.16.6.212 -e "insert into
tpch.customer_test (C_NAME,C_ADDRESS,C_NATIONKEY,C_PHONE,C_ACCTBAL,C_MKTSEGMENT,C_COMMENT)
select C_NAME,C_ADDRESS,C_NATIONKEY,C_PHONE,C_ACCTBAL,C_MKTSEGMENT,C_COMMENT from
tpch.customer"; done;

mysql> select count(*) from customer_test;
+----------+
| count(*) |
+----------+
|  9000000 |
+----------+
1 row in set (0.30 sec)
```

练习：消耗内存较大 SQL 语句对于 TiDB Server 的影响

2. 执行测试 SQL 语句，观察内存消耗情况：

```
mysql> explain analyze select t.C_NAME,t.C_ADDRESS,s.C_CUSTKEY
from customer_test t,customer s where
t.C_NAME=s.C_NAME order by t.C_ADDRESS;
```

3. 反复执行测试用 SQL 语句，进行监控

反复执行测试用 SQL 语句：

```
# for i in `seq 10`; do mysql -uroot -P4000 -h172.16.6.212 -e "explain
analyze select t.C_NAME,t.C_ADDRESS,s.C_CUSTKEY from tpch.customer_test
t,tpch.customer s where t.C_NAME=s.C_NAME order by t.C_ADDRESS"; done;
```

通过监控 (http://172.16.6.157:3000)：Grafana监控 --> TiDB --> Server --> Memory Usage 监控内存使用呈上升趋势，如下图：

练习：消耗内存较大 SQL 语句对于 TiDB Server 的影响

找到内存占用较大的 SQL 语句：

<1> TiDB Dashboard SQL 语句分析

选择目标时间，以及数据库名称，可以看到 SQL list，其中可以通过『平均内存』看到集群中，执行过的 SQL 消耗的平均内存：

<2> TiDB Dashboard 慢查询

选择目标时间，以及数据库名称，可以看到 SQL list，按照『总执行时间』排序：

点击排序结果的第一条 SQL，并查看详情，可以看到该 SQL 使用的最大内存：

练习：消耗内存较大 SQL 语句对于 TiDB Server 的影响

4. 通过参数限制单条 SQL 语句对于内存的使用

修改参数 tidb_expensive_query_time_threshold，使测试 SQL 可以被记录到 tidb.log 中，如下：

```
mysql> show variables like '%tidb_expensive_query_time_threshold%';
+---------------------------------------+-------+
| Variable_name                         | Value |
+---------------------------------------+-------+
| tidb_expensive_query_time_threshold | 60    |
+---------------------------------------+-------+
1 row in set (0.02 sec)

mysql> set tidb_expensive_query_time_threshold=10;
Query OK, 0 rows affected (0.00 sec)

mysql> show variables like '%tidb_expensive_query_time_threshold%';
+---------------------------------------+-------+
| Variable_name                         | Value |
+---------------------------------------+-------+
| tidb_expensive_query_time_threshold | 10    |
+---------------------------------------+-------+
1 row in set (0.02 sec)
```

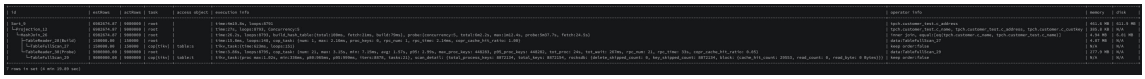
修改参数 tidb_mem_quota_query，限制测试 SQL 的内存使用为 107 M，如下：

```
mysql> show variables like '%tidb_mem_quota_query%';
+--------------------------+---------------+
| Variable_name            | Value         |
+--------------------------+---------------+
| tidb_mem_quota_query | 1073741824 |
+--------------------------+---------------+
1 row in set (0.01 sec)

mysql> set tidb_mem_quota_query=107374100;
Query OK, 0 rows affected (0.01 sec)

mysql> show variables like '%tidb_mem_quota_query%';
+--------------------------+---------------+
| Variable_name            | Value         |
+--------------------------+---------------+
| tidb_mem_quota_query | 107374100 |
+--------------------------+---------------+
1 row in set (0.01 sec)
```

练习：消耗内存较大 SQL 语句对于 TiDB Server 的影响

再次执行测试 SQL 语句，观察内存和磁盘的消耗情况：

```
mysql> explain analyze select t.C_NAME,t.C_ADDRESS,s.C_CUSTKEY from
customer_test t,customer s where
t.C_NAME=s.C_NAME order by t.C_ADDRESS;
```



```
-+----------+----------+
| id

 | memory   | disk     |
+-------------------------------
-------------------------------
-+----------+----------+
| Sort_9

 | 461.6 MB | 611.9 MB |
| └─Projection_12

 | 385.8 KB | N/A      |
|   └─HashJoin_26
37.7s, fetch:24.5s}
 | 4.94 MB  | 6.01 MB  |
|     ├─TableReader_28(Build)

 | 4.87 MB  | N/A      |
|     | └─TableFullScan_27

 | N/A      | N/A      |
|     └─TableReader_30(Probe)
, tot_proc: 24s, tot_wait: 267ms
 | 277.9 MB | N/A      |
|       └─TableFullScan_29
 8872154, rocksdb: {delete_skipp
 | N/A      | N/A      |
+-------------------------------
-------------------------------
-+----------+----------+
7 rows in set (4 min 19.89 sec)
```

练习：消耗内存较大 SQL 语句对于 TiDB Server 的影响

观察 TiDB Server 日志

可以通过过滤 tidb.log 中的 expensivequery 关键字，来发现内存占用高的 SQL，如下：

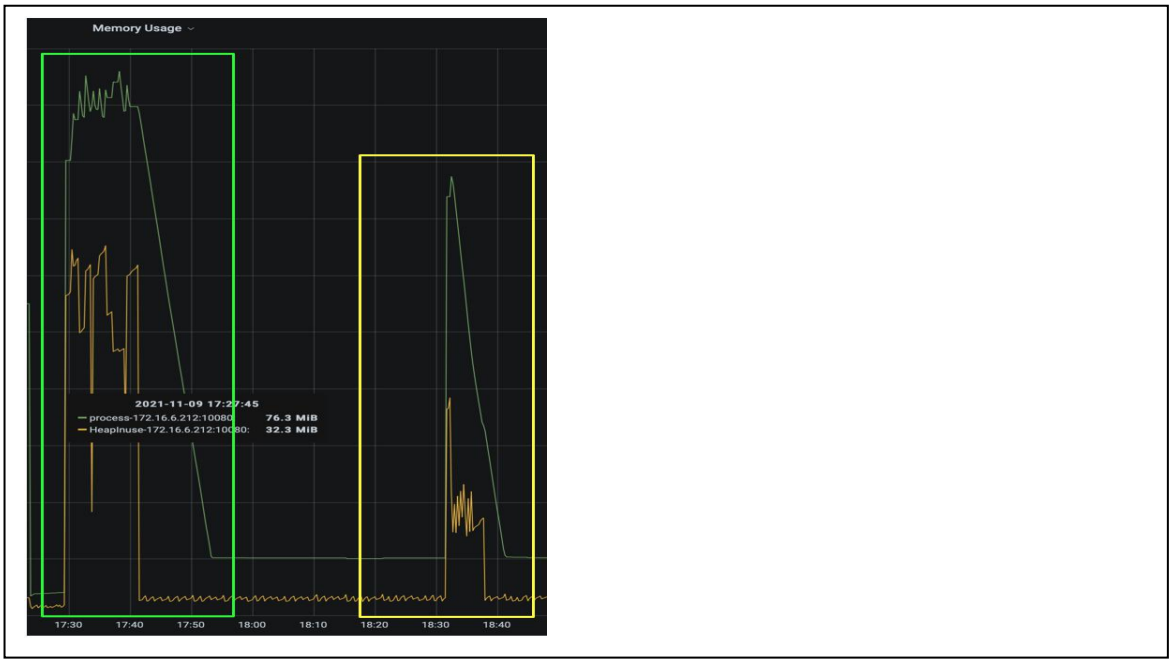
```
# more tidb.log | grep expensivequery
...
[2021/11/09 18:31:48.627 +08:00] [WARN] [expensivequery.go:178] [expensive_query] [cost_time=10.04978579s]
[cop_time=0.861190838s] [process_time=21.543s] [wait_time=0.265s] [request_count=17] [total_keys=6724098]
[process_keys=6724083] [num_cop_tasks=17] [process_avg_time=1.267235294s] [process_p90_time=2.276s]
[process_max_time=2.458s] [process_max_addr=172.16.6.163:20160] [wait_avg_time=0.015588235s]
[wait_p90_time=0.094s] [wait_max_time=0.095s] [wait_max_addr=172.16.6.163:20160]
[stats=customer:428984657937432579,customer_test:428985314136293379] [conn_id=123] [user=root]
[database=tpch] [table_ids="[93,100]"] [txn_start_ts=428986570778869761] [mem_max="541933395 Bytes (516.8 MB)"]
[sql="explain analyze select t.C_NAME,t.C_ADDRESS,s.C_CUSTKEY from customer_test t,customer s where
t.C_NAME=s.C_NAME order by t.C_ADDRESS"]
```

可以通过过滤 tidb.log 中的 memory exceeds quota, spill to disk now 关键字，来发现 SQL 执行过程中的一部分中间结果使用了磁盘作为临时存储，如下：

```
# more tidb.log | grep 'memory exceeds quota, spill to disk now'
...
[2021/11/09 18:31:45.665 +08:00] [INFO] [row_container.go:505] ["memory exceeds quota, spill to disk now."]
[consumed=515442826] [quota=107374100]
[2021/11/09 18:32:23.723 +08:00] [INFO] [row_container.go:311] ["memory exceeds quota, spill to disk now."]
[consumed=125727438] [quota=107374100]
[2021/11/09 18:32:25.512 +08:00] [INFO] [row_container.go:505] ["memory exceeds quota, spill to disk now."]
[consumed=107439794] [quota=107374100]
```

练习：消耗内存较大 SQL 语句对于 TiDB Server 的影响

最后，我们观察下 Grafana 监控 --> TiDB --> Server --> Memory Usage 内存使用情况，其中绿框部分是未 spill 到磁盘时，内存的使用情况。反之，黄框是中间结果 spill 到磁盘后的，内存使用情况：



清理实验数据

```
# tiup bench tpch -H 172.16.6.212 -P 4000 -D tpch clean
Starting component `bench`: /root/.tiup/components/bench/v1.6.1/tiup-bench tpch -H 172.16.6.212 -P 4000 -D tpch clean
DROP TABLE IF EXISTS lineitem
DROP TABLE IF EXISTS partsupp
DROP TABLE IF EXISTS supplier
DROP TABLE IF EXISTS part
DROP TABLE IF EXISTS orders
DROP TABLE IF EXISTS customer
DROP TABLE IF EXISTS region
DROP TABLE IF EXISTS nation
Finished
```