

TPC-H performance measure

Keisuke Suzuki

2012 年 12 月 4 日

1 実験環境

- CPU : Xeon X7560 @ 2.27GHz x4
- Memory : 64GB
- DBMS : PostgreSQL 9.2
- RAID0 : iohdrive x8 (chunk size = 64KB)
- 各テーブルの primary key 上に B-tree index を構築
- Scale Factor = 100
- shared buffer = 8GB
- 各クエリの実行時の状況を iostat と mpstat で 1 秒おきに監視

2 Query : count lineitem

最初の数十秒間だけ少しスループットが大きくなる現象について。NUMA によるローカルメモリとリモートメモリのページキャッシュの書き込みのスループットの差が原因と考えられる。そこで、ローカルメモリが空の状態と半分埋まった状態からスタートし、スループットが大きくなっている時間に変化があるか確認する。

2.1 計測結果

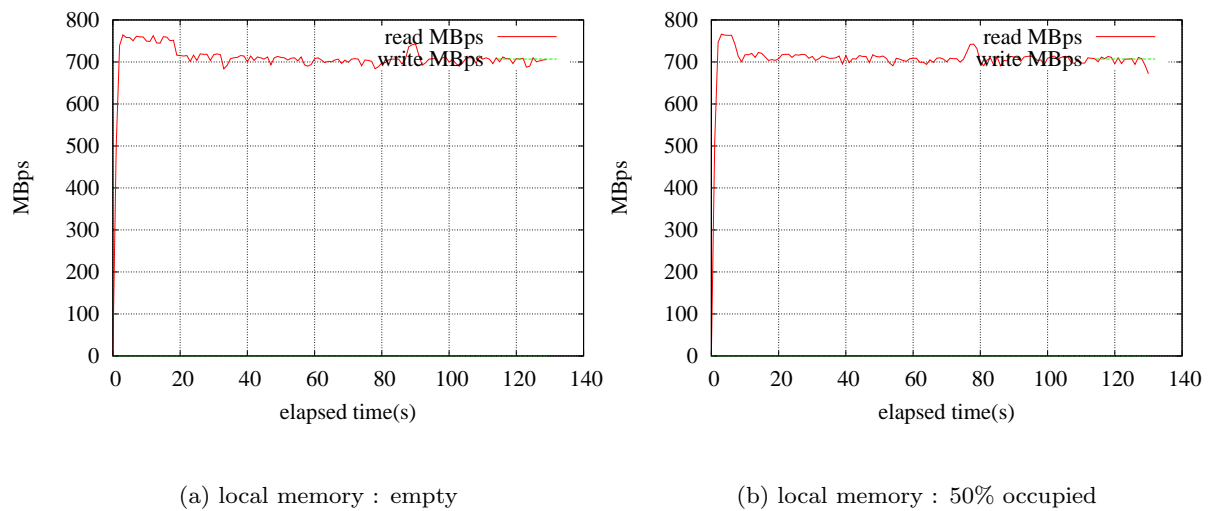


図 1 MBPS

図 1(a) と 1(b) でスループットが大きくなっている時間が大体半分程度になっている。numactl -hardware で一定時間おきにメモリの使用状況を追ってみても、始めにローカルメモリから使用されていることが確認できた。

2.2 Query 1 by index scan on l_shipdate

lineitem の l_shipdate attribute 上に B-tree index を構築し、lineitem 上のデータの選択率を絞って、index scan させる。クエリ及びクエリ実行プランは以下の通り。

Query 1

```
select
    l_returnflag, l_linestatus,
    sum(l_quantity) as sum_qty,
    sum(l_extendedprice) as sum_base_price,
    sum(l_extendedprice * (1 - l_discount)) as sum_disc_price,
    sum(l_extendedprice * (1 - l_discount) * (1 + l_tax)) as sum_charge,
    avg(l_quantity) as avg_qty,
    avg(l_extendedprice) as avg_price,
    avg(l_discount) as avg_disc,
    count(*) as count_order
from
    lineitem
where
    l_shipdate <= date '1992-02-28'
group by
    l_returnflag, l_linestatus
order by
    l_returnflag, l_linestatus;
```

Query execution plan

```
Sort  Sort Key: l_returnflag, l_linestatus
-> HashAggregate
    -> Index Scan using l_shipdate_idx on lineitem
        Index Cond: (l_shipdate <= '1992-02-28'::date)
```

2.3 baseline 計測

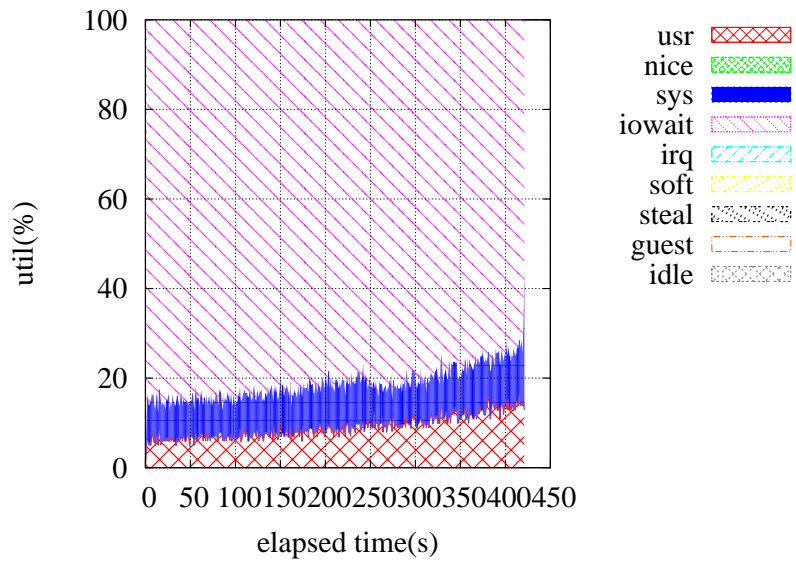
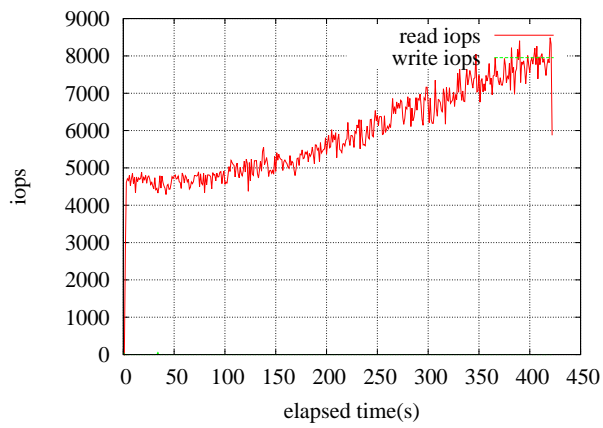
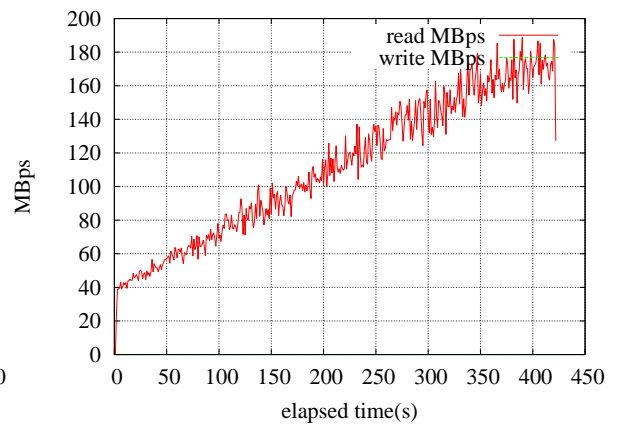


図 2 cpu utilization



(a) IOPS



(b) MBPS

図 3 IO spec

n

時間の経過と共に徐々にスループットが増加しているが、IOsize の増加から、look-ahead の効果ではないかと考えられる。

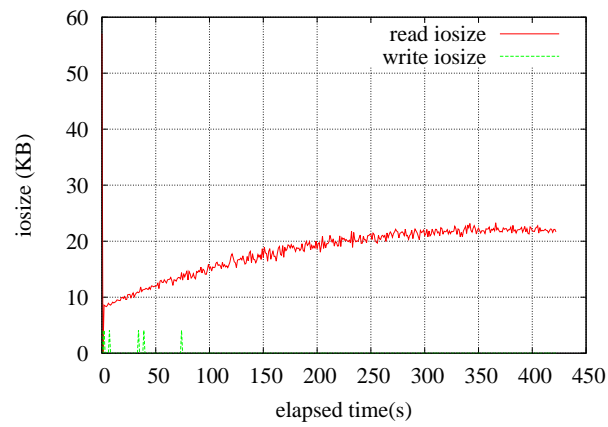


図 4 IO size

2.4 look-ahead を off にした場合

blockdev --setra 0 でデバイスに対する look-ahead を切って測定する。

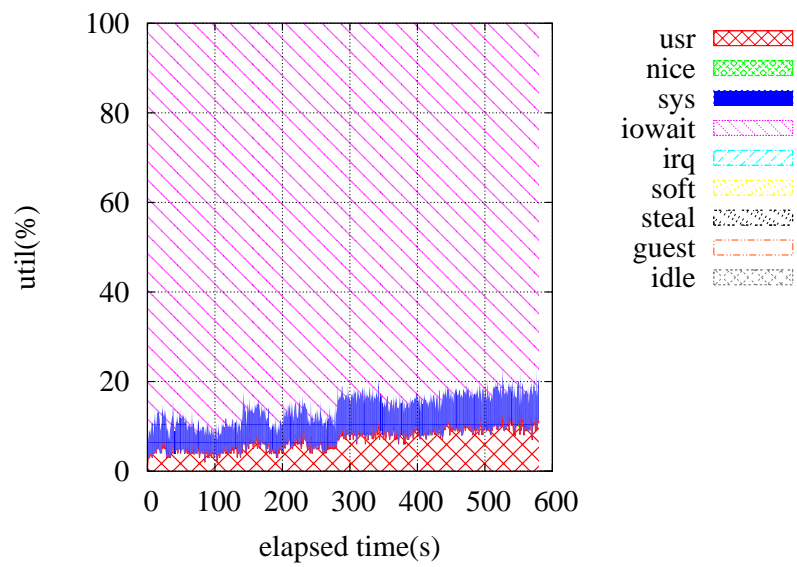
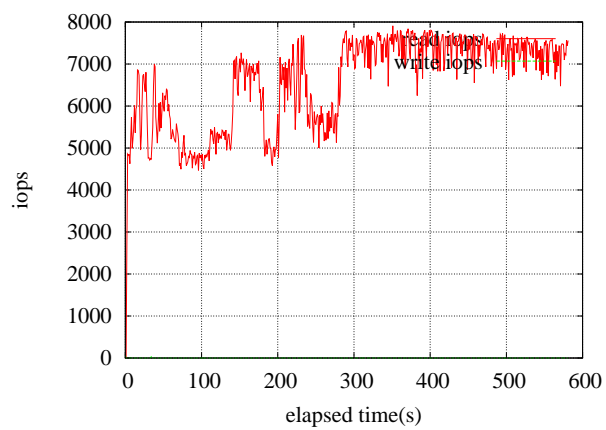
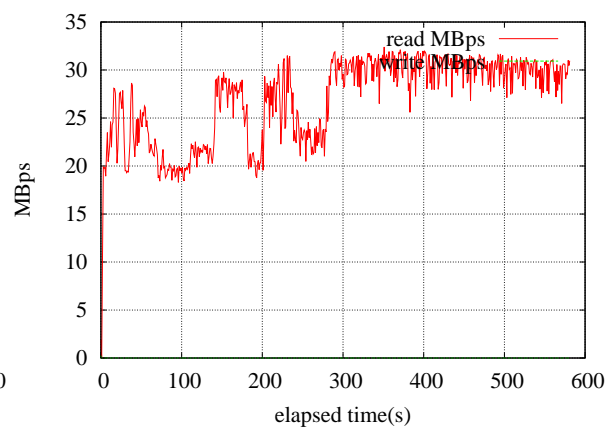


図 5 cpu utilization



(a) IOPS



(b) MBPS

图 6 IO spec

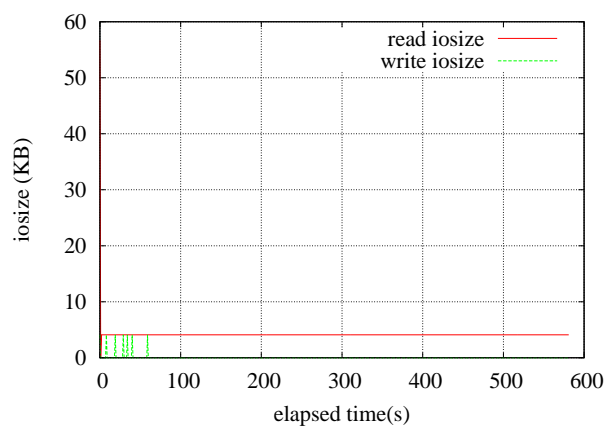


图 7 IO size