

# TPC-H performance measure

Keisuke Suzuki

2012 年 12 月 19 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 1 by index scan on l\_shipdate

look-ahead を切った状態での IO 性能が乱れていたのを、その原因を探る。

### 2.1 random read microbenchmark

測定時の条件

- look-ahead (read-ahead): 0
- iosize: 8KB
- raw device access

#### 2.1.1 With O\_DIRECT flag

まず OS のバッファリングなどを切った状態でのアクセス時の性能を示す。

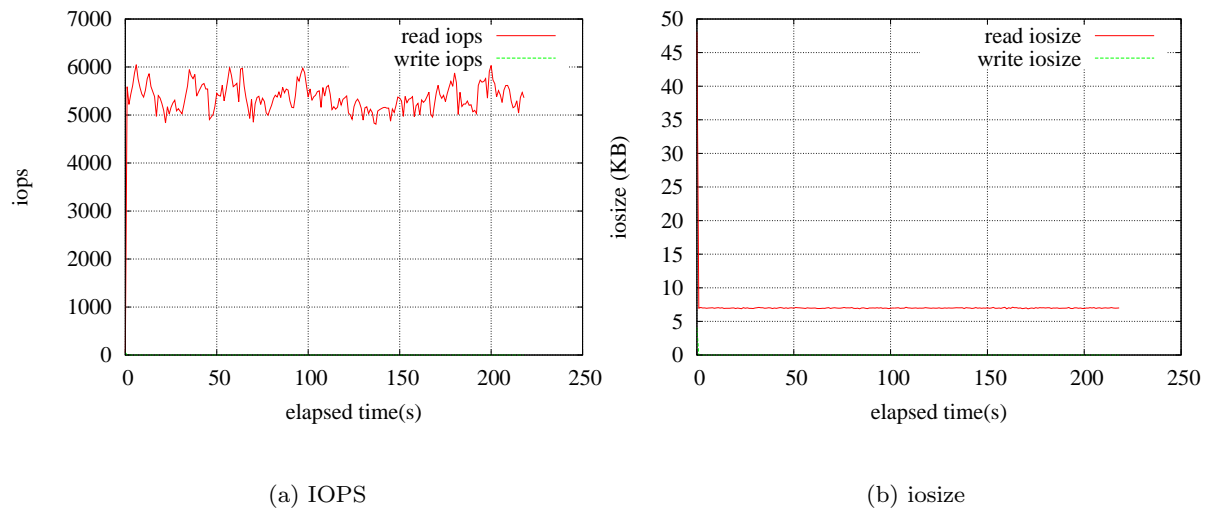


図 1 IO spec (iosize = 8KB)

表 1 IO spec average (iosize = 8KB)  
(IOPS = (io issued by benchmark / elapsed time))

IOPS	MBPS
4578	37.5

benchmark 実行時の IO の発行状況を blktrace で監視すると以下の通りであった。

```
9,0    1      1      0.0000000000 38969  Q   R 1894937090 + 16 [randomread]
9,0    1      2      0.000011450 38969  U   N [randomread] 0
9,0    1      3      0.000331059 38969  Q   R 1645272306 + 13 [randomread]
9,0    1      4      0.000338372 38969  Q   R 1645272319 + 3 [randomread]
9,0    1      5      0.000339942 38969  X   R 1645272319 / 1645272320 [randomread]
9,0    1      6      0.000347454 38969  U   N [randomread] 0
9,0    1      7      0.000611996 38969  Q   R 2143216519 + 16 [randomread]
9,0    1      8      0.000615720 38969  U   N [randomread] 0
```

...

CPU1 (md0):

Reads Queued:	1,172K,	7,882MiB	Writes Queued:	0,	0KiB
Read Dispatches:	0,	0KiB	Write Dispatches:	0,	0KiB
Reads Requeued:	0		Writes Requeued:	0	
Reads Completed:	0,	0KiB	Writes Completed:	0,	0KiB
Read Merges:	0,	0KiB	Write Merges:	0,	0KiB
Read depth:	0		Write depth:	0	
IO unplugs:	1,000,000		Timer unplugs:	0	

Throughput (R/W): 0KiB/s / 0KiB/s

Events (md0): 2,273,525 entries

### 2.1.2 Without O\_DIRECT flag

O\_DIRECT flag を使用しない状態でのアクセス時の性能を示す。

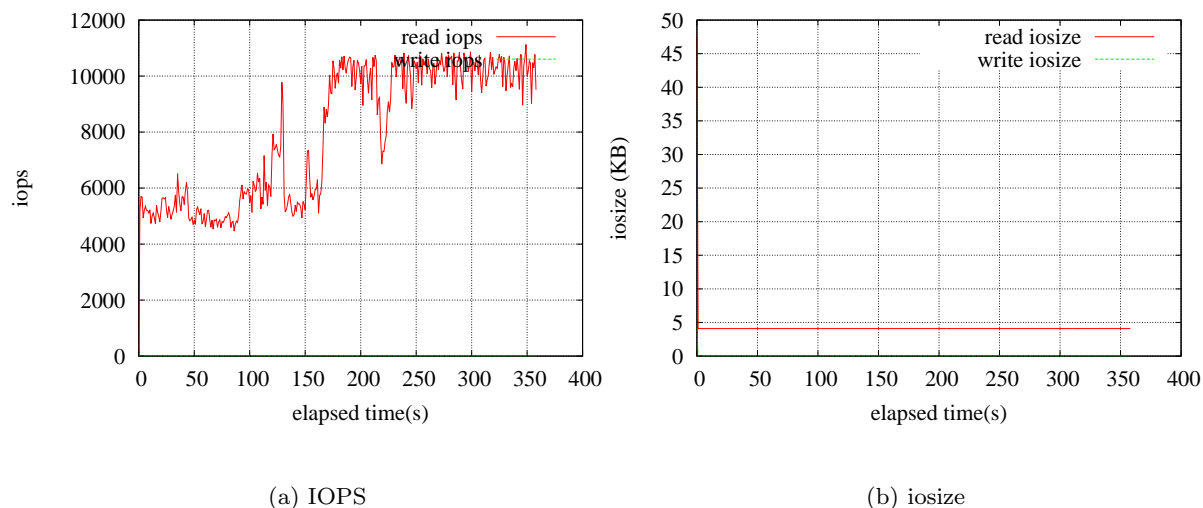


図 2 IO spec (iosize = 8KB)

表 2 IO spec average (iosize = 8KB)

(IOPS = (io issued by benchmark / elapsed time))

IOPS	MBPS
2787	22.8

図 2(b) を見ると、クエリ実行時と同様の性能の乱れが見られる。この乱れは、iodrive を OS の buffering を有効にして使用したときに生じる、デバイスの特性に由来するものではないかと考えられる。

また、iosize に着目すると、本来 8KB で IO を発行しているはずであるが、実際の IO は半分の 4KB で出ている。iostat で計測した IOPS と、benchmark で計算した IOPS も数値としてはあっていない。

そこで blktrace で benchmark 実行時の IO の発行状況を監視すると以下の通りであった。

```

9,0    1      1      0.000000000 39618  Q   R 1894937088 + 8 [randomread]
9,0    1      2      0.000009686 39618  U   N [randomread] 0
9,0    1      3      0.000274736 39618  Q   R 1894937096 + 8 [randomread]
9,0    1      4      0.000278264 39618  U   N [randomread] 0
9,0    1      5      0.000503299 39618  Q   R 1894937104 + 8 [randomread]
9,0    1      6      0.000509047 39618  U   N [randomread] 0
9,0    1      7      0.000794094 39618  Q   R 1645272304 + 8 [randomread]
9,0    1      8      0.000800924 39618  U   N [randomread] 0
9,0    1      9      0.001145873 39618  Q   R 1645272312 + 8 [randomread]
9,0    1     10      0.001152079 39618  U   N [randomread] 0

```

...

CPU1 (md0):

```

Reads Queued:      2,860K,    11,442MiB Writes Queued:      0,          OKiB
Read Dispatches:    0,          OKiB Write Dispatches:    0,          OKiB
Reads Requeued:     0              Writes Requeued:      0
Reads Completed:    0,          OKiB Writes Completed:    0,          OKiB
Read Merges:        0,          OKiB Write Merges:        0,          OKiB
Read depth:         0              Write depth:         0
IO unplugs:         2,860,599      Timer unplugs:      0

```

Throughput (R/W): 0KiB/s / 0KiB/s

Events (md0): 5,721,230 entries

この結果をみると、OS から発行される IO は 4KB のサイズになっている。これは OS がページサイズ単位 (4KB) に IO を分割して発行しているのではないかと考えられる。実際、他のサイズで IO を発行した場合も、OS からは 4KB の IO として発行されていることが確認できた。以下は、16KB のサイズでアクセスした場合の結果である。

```

9,0    1      1      0.000000000 39810  Q   R 1894937088 + 8 [randomread]
9,0    1      2      0.000008815 39810  U   N [randomread] 0
9,0    1      3      0.000352988 39810  Q   R 1894937096 + 8 [randomread]
9,0    1      4      0.000359066 39810  U   N [randomread] 0
9,0    1      5      0.000656633 39810  Q   R 1894937104 + 8 [randomread]
9,0    1      6      0.000662659 39810  U   N [randomread] 0

```

...

CPU1 (md0):

```

Reads Queued:      4,831K,    19,327MiB Writes Queued:      0,          OKiB
Read Dispatches:    0,          OKiB Write Dispatches:    0,          OKiB

```

Reads Requeued:	0		Writes Requeued:	0	
Reads Completed:	0,	OKiB	Writes Completed:	0,	OKiB
Read Merges:	0,	OKiB	Write Merges:	0,	OKiB
Read depth:	0		Write depth:	0	
IO unplugs:	4,831,926		Timer unplugs:	0	

Throughput (R/W): 0KiB/s / 0KiB/s

Events (md0): 9,663,852 entries