

15721- DATABASE SYSTEMS

Wire Protocol
Memcached

Siddharth Santurkar, Nitin Chandra Badam, Di Xiao

Overview

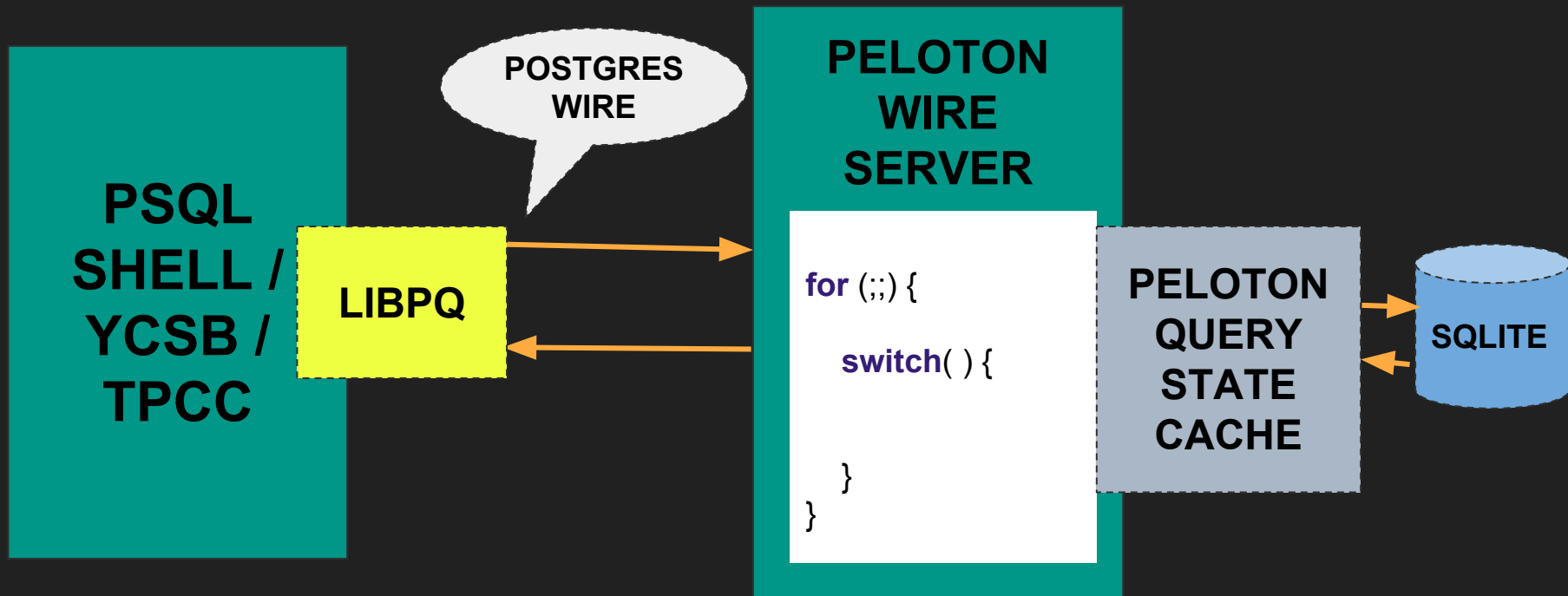
- Implement the Wire Protocol for YCSB + TPC-C support (100%)
 - Backed with sqlite for correctness testing
 - Integrated with the new Peloton Query cache
- Support Memcached API with Peloton (75%)
 - Tested with benchmarks: YCSB, memslap

The Wire Protocol

In collaboration with

Jianhong Li, Runshen Zhu and Ryan Xian

Wire Protocol for Peloton



SIMPLE QUERY PROTOCOL

- From PSQL
 - StartupMessage (F)
 - Query (Q)
- At Server
 - AuthenticationOk (R)
 - ErrorResponse (E)
 - ReadyForQuery (Z)
 - EmptyQueryResponse (I)
 - RowDescription (T)
 - DataRow (D)
 - CommandComplete (C)

DEMO



Filter: tcp.port == 57721 Expression... Clear Apply Save

No.	Time	TCP Delta	Source	Destination	Protocol	Length	Info
19	0.000000	0.000000	127.0.0.1	127.0.0.1	TCP	74	58551 > 57721 [SYN] Seq=0 Win=43690 Len=0 MSS=65495 SACK_PERM=1 TSval=26325984 TSecr=26325984
20	0.000057	0.000057	127.0.0.1	127.0.0.1	TCP	74	57721 > 58551 [SYN, ACK] Seq=0 Ack=1 Win=43690 Len=0 MSS=65495 SACK_PERM=1 TSval=26325984 TSecr=26325984
21	0.000048	0.000048	127.0.0.1	127.0.0.1	TCP	66	58551 > 57721 [ACK] Seq=1 Ack=1 Win=43776 Len=0 TSval=26325965 TSecr=26325965
22	0.000089	0.000089	127.0.0.1	127.0.0.1	TCP	152	58551 > 57721 [PSH, ACK] Seq=1 Ack=1 Win=43776 Len=86 TSval=26325965 TSecr=26325965
23	0.000028	0.000028	127.0.0.1	127.0.0.1	TCP	66	57721 > 58551 [ACK] Seq=1 Ack=87 Win=43776 Len=0 TSval=26325965 TSecr=26325965
24	0.075409	0.075409	127.0.0.1	127.0.0.1	TCP	385	57721 > 58551 [PSH, ACK] Seq=1 Ack=87 Win=43776 Len=319 TSval=26325984 TSecr=26325984
25	0.000018	0.000018	127.0.0.1	127.0.0.1	TCP	66	58551 > 57721 [ACK] Seq=87 Ack=320 Win=44800 Len=0 TSval=26325984 TSecr=26325984

Ethernet II, Src: 00:00:00_00:00:00 (00:00:00:00:00:00), Dst: 00:00:00_00:00:00 (00:00:00:00:00:00)
Internet Protocol Version 4, Src: 127.0.0.1 (127.0.0.1), Dst: 127.0.0.1 (127.0.0.1)
Transmission Control Protocol, Src Port: 58551 (58551), Dst Port: 57721 (57721), Seq: 1, Ack: 1, Len: 86

Data (86 bytes)
Data: 000000560003000075736572007369646468617274680064...

[Length: 86]		
000	00 00 00 00 00 00 00 00 00 00 00 00 08 00 45 00	E.
010	00 8a e6 57 40 00 40 06 56 14 7f 00 00 01 7f 00	...W@.@. V.....
020	00 01 e4 b7 e1 79 ce 84 55 99 59 c0 03 2c 80 18	y.. U.Y.,...
030	01 56 fe 7e 00 00 01 01 08 0a 01 91 b3 cd 01 91	.V.~....
040	b3 cd 00 00 00 56 00 03 00 00 75 73 65 72 00 73	...V.. .user.s
050	69 64 64 68 61 72 74 68 00 64 61 74 61 62 61 73	iddharth .databas
060	65 00 73 69 64 64 68 61 72 74 68 00 61 70 70 6c	e.siddha rth.appl
070	69 63 61 74 69 6f 6e 5f 6e 61 6d 65 00 70 73 71	ication name.psq
080	6c 00 63 6c 69 65 6e 74 5f 65 6e 63 6f 64 69 6e	l.client encodin
090	67 00 55 54 46 38 00 00	g.UTF8..



Filter: tcp.port == 57721 Expression... Clear Apply Save

No.	Time	TCP Delta	Source	Destination	Protocol	Length	Info
19	0.000000	0.000000	127.0.0.1	127.0.0.1	TCP	74	58551 > 57721 [SYN] Seq=0 Win=43690 Len=0 MSS=65495 SACK_PERM=1 TSval=2
20	0.000057	0.000057	127.0.0.1	127.0.0.1	TCP	74	57721 > 58551 [SYN, ACK] Seq=0 Ack=1 Win=43690 Len=0 MSS=65495 SACK_PER
21	0.000048	0.000048	127.0.0.1	127.0.0.1	TCP	66	58551 > 57721 [ACK] Seq=1 Ack=1 Win=43776 Len=0 TSval=26325965 TSecr=26
22	0.000089	0.000089	127.0.0.1	127.0.0.1	TCP	152	58551 > 57721 [PSH, ACK] Seq=1 Ack=1 Win=43776 Len=86 TSval=26325965 TS
23	0.000028	0.000028	127.0.0.1	127.0.0.1	TCP	66	57721 > 58551 [ACK] Seq=1 Ack=87 Win=43776 Len=0 TSval=26325965 TSecr=2
24	0.075409	0.075409	127.0.0.1	127.0.0.1	TCP	385	57721 > 58551 [PSH, ACK] Seq=1 Ack=87 Win=43776 Len=319 TSval=26325984
25	0.000018	0.000018	127.0.0.1	127.0.0.1	TCP	66	58551 > 57721 [ACK] Seq=87 Ack=320 Win=44800 Len=0 TSval=26325984 TSecr

Internet Protocol Version 4, Src: 127.0.0.1 (127.0.0.1), Dst: 127.0.0.1 (127.0.0.1)
Transmission Control Protocol, Src Port: 57721 (57721), Dst Port: 58551 (58551), Seq: 1, Ack: 87, Len: 319
Data (319 bytes)

0040	b3 cd 52 00 00 00 08 00 00 00 00 53 00 00 00 18	..R.... ..S....
0050	54 69 6d 65 5a 6f 6e 65 00 55 53 2f 45 61 73 74	Timezone .US/East
0060	65 72 6e 00 53 00 00 00 1a 61 70 70 6c 69 63 61	ern.S... applica
0070	74 69 6f 6e 5f 6e 61 6d 65 00 70 73 71 6c 00 53	tion nam e.psqli.S
0080	00 00 00 19 63 6c 69 65 6e 74 5f 65 6e 63 6f 64	clie nt_encod
0090	69 6e 67 00 55 54 46 38 00 53 00 00 00 1c 73 65	ing.UTF8 .S....se
00a0	72 76 65 72 5f 76 65 72 73 69 6f 6e 00 39 2e 35	rver ver sion.9.5
00b0	64 65 76 65 6c 00 53 00 00 00 23 73 65 73 73 69	devel.S. ...#sessi
00c0	6f 6e 5f 61 75 74 68 6f 72 69 7a 61 74 69 6f 6e	on_autho rization
00d0	00 70 6f 73 74 67 72 65 73 00 53 00 00 00 17 44	.postgre s.S....D
00e0	61 74 65 53 74 79 6c 65 00 49 53 4f 2c 20 4d 44	ateStyle .ISO, MD
00f0	59 00 53 00 00 00 19 73 65 72 76 65 72 5f 65 6e	Y.S....s erver en
0100	63 6f 64 69 6e 67 00 55 54 46 38 00 53 00 00 00	coding.U TF8.S...
0110	19 69 6e 74 65 67 65 72 5f 64 61 74 65 74 69 6d	.integer _datetim
0120	65 73 00 6f 6e 00 53 00 00 00 1b 49 6e 74 65 72	es.on.S. ...Inter
0130	76 61 6c 53 74 79 6c 65 00 70 6f 73 74 67 72 65	valStyle .postgre
0140	73 00 53 00 00 00 14 69 73 5f 73 75 70 65 72 75	s.S....i s superu
0150	73 65 72 00 6f 6e 00 53 00 00 00 23 73 74 61 6e	ser.on.S. ...#stan
0160	64 61 72 64 5f 63 6f 6e 66 6f 72 6d 69 6e 67 5f	dard con forming
0170	73 74 72 69 6e 67 73 00 6f 6e 00 5a 00 00 00 05	strings. on.Z....
0180	49	I

Click to change configuration profile



Filter: tcp.port == 57721 Expression... Clear Apply Save

No.	Time	TCP Delta	Source	Destination	Protocol	Length	Info
24	0.0754090	0.075409000	127.0.0.1	127.0.0.1	TCP	385	57721 > 58551 [PSH, ACK] Seq=1 Ack=87 Win=43776 Len=319 TSval=26325984
25	0.0000180	0.000018000	127.0.0.1	127.0.0.1	TCP	66	58551 > 57721 [ACK] Seq=87 Ack=320 Win=44800 Len=0 TSval=26325984 TSec
506	344.17783344	344.1778300	127.0.0.1	127.0.0.1	TCP	119	58551 > 57721 [PSH, ACK] Seq=87 Ack=320 Win=44800 Len=53 TSval=2641202
507	0.0000410	0.000041000	127.0.0.1	127.0.0.1	TCP	66	57721 > 58551 [ACK] Seq=320 Ack=140 Win=43776 Len=0 TSval=26412028 TSe
508	0.2116950	0.211695000	127.0.0.1	127.0.0.1	TCP	126	57721 > 58551 [PSH, ACK] Seq=320 Ack=140 Win=43776 Len=60 TSval=264120
509	0.0000170	0.000017000	127.0.0.1	127.0.0.1	TCP	66	58551 > 57721 [ACK] Seq=140 Ack=380 Win=44800 Len=0 TSval=26412081 TSe
510	5.2697760	5.269776000	127.0.0.1	127.0.0.1	TCP	98	58551 > 57721 [PSH, ACK] Seq=140 Ack=380 Win=44800 Len=32 TSval=264133
511	0.0000300	0.000030000	127.0.0.1	127.0.0.1	TCP	66	57721 > 58551 [ACK] Seq=380 Ack=172 Win=43776 Len=0 TSval=26413398 TSe
512	0.1451170	0.145117000	127.0.0.1	127.0.0.1	TCP	105	57721 > 58551 [PSH, ACK] Seq=380 Ack=172 Win=43776 Len=39 TSval=264134
513	0.0000130	0.000013000	127.0.0.1	127.0.0.1	TCP	66	58551 > 57721 [ACK] Seq=172 Ack=419 Win=44800 Len=0 TSval=26413435 TSe
522	5.1210130	5.121013000	127.0.0.1	127.0.0.1	TCP	88	58551 > 57721 [PSH, ACK] Seq=172 Ack=419 Win=44800 Len=22 TSval=264147
523	0.0000410	0.000041000	127.0.0.1	127.0.0.1	TCP	66	57721 > 58551 [ACK] Seq=419 Ack=194 Win=43776 Len=0 TSval=26414715 TSe
524	0.0004290	0.000429000	127.0.0.1	127.0.0.1	TCP	163	57721 > 58551 [PSH, ACK] Seq=419 Ack=194 Win=43776 Len=97 TSval=264147
525	0.0000110	0.000011000	127.0.0.1	127.0.0.1	TCP	66	58551 > 57721 [ACK] Seq=194 Ack=516 Win=44800 Len=0 TSval=26414715 TSe

```
0000 00 00 00 00 00 00 00 00 00 00 00 00 08 00 45 00 .....E.
0010 00 4a e6 5d 40 00 40 06 56 4e 7f 00 00 01 7f 00 .J.]@.@. VN.....
0020 00 01 e4 b7 e1 79 ce 84 56 44 59 c0 04 ce 80 18 ....y.. VDY.....
0030 01 5e fe 3e 00 00 01 01 08 0a 01 93 0e 7b 01 93 .^.>....{...
0040 09 7b 51 00 00 00 15 53 45 4c 45 43 54 20 2a 20 .{Q....S ELECT *
0050 46 52 4f 4d 20 41 3b 00 FROM A;.
```


EXTENDED QUERY PROTOCOL

- YCSB and TPCC use JDBC to run benchmarks
- JDBC requires the extended query protocol support
- Main packets supported
 - **Parse(P)**: Query/Prepared Statement string + Prepared Statement handle (cache the handle)
 - **Bind(B)**: (Prep Statement only) - Handle + Bind parameters
 - **Describe(D)**: Indicator of statement type
 - **Execute(E)**: Execute the complete statement on Sqlite
 - **Sync(S)**: End of txn indicator
- Txn awareness through the cache
- Coarse-grained locking, because, Sqlite

EXAMPLES - PARSE

13:43:28,388 [../../src/backend/wire/protocol.cpp:318:exec_parse_message] INFO - PARSE message

13:43:28,388 [../../src/backend/wire/protocol.cpp:324:exec_parse_message] INFO - Prep stmt: S_2

13:43:28,388 [../../src/backend/wire/protocol.cpp:327:exec_parse_message] INFO - Parse Query: INSERT INTO
"USERTABLE" VALUES (\$1, \$2, \$3, \$4, \$5, \$6, \$7, \$8, \$9, \$10, \$11)

13:43:28,388 [../../src/backend/wire/protocol.cpp:345:exec_parse_message] INFO - NumParams: 11

EXAMPLES - BIND

13:43:28,389 [./../src/backend/wire/protocol.cpp:384:exec_bind_message] INFO - BIND message

13:43:28,389 [./../src/backend/wire/protocol.cpp:386:exec_bind_message] INFO - Portal name:

13:43:28,389 [./../src/backend/wire/protocol.cpp:388:exec_bind_message] INFO - Prep stmt name: S_2

13:43:28,389 [./../src/backend/wire/protocol.cpp:460:exec_bind_message] INFO - TEXT mode

13:43:28,389 [./../src/backend/wire/protocol.cpp:464:exec_bind_message] INFO - Bind param (size: 1) : 0

13:43:28,389 [./../src/backend/wire/protocol.cpp:460:exec_bind_message] INFO - TEXT mode

13:43:28,389 [./../src/backend/wire/protocol.cpp:464:exec_bind_message] INFO - Bind param (size: 100) : RMVg9zN9>[\$]:
-:Slr'NS=%([[-==KrxKB)|k~yN-;GRYT ^kc%[gV5T&`Ao(1/k\$,A0\+=1BQgBLy{QGQTZS3RF,;_PYL?,/V5\$

EXAMPLES - EXECUTE

13:43:28,389 [../../src/backend/wire/protocol.cpp:560:exec_execute_message] INFO - EXECUTE message

13:43:28,389 [../../src/backend/wire/protocol.cpp:583:exec_execute_message] INFO - Executing query: INSERT INTO "USERTABLE" VALUES (\$1, \$2, \$3, \$4, \$5, \$6, \$7, \$8, \$9, \$10, \$11)

13:43:28,389 [../../src/backend/wire/sqlite.h:179:ExecPrepStmt] INFO - Executing statement.....

13:43:28,389 [../../src/backend/wire/sqlite.h:182:ExecPrepStmt] INFO - ret: 101

13:43:28,389 [../../src/backend/wire/sqlite.h:184:ExecPrepStmt] INFO - column count: 0

13:43:28,389 [../../src/backend/wire/protocol.cpp:236:complete_command] INFO - complete command tag: INSERT 0 1

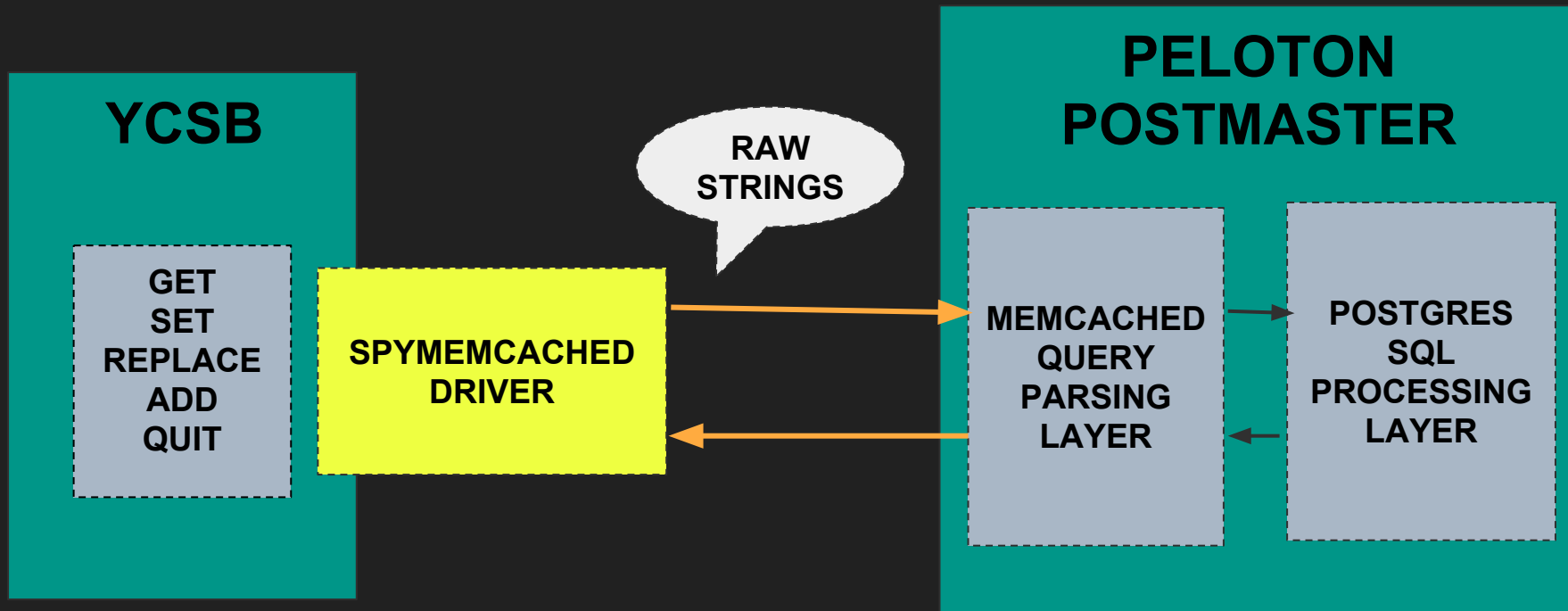
```
siddharth@siddharth-Q502LAB: ~/Documents/CMU/15-721/oltpbench 161x41
14:47:59,297 (DBWorkload.java:721) INFO - Rate limited reqs/s: Results(nanoSeconds=100999429270, measuredRequests=1341) = 13.27730275004949 requests/sec
14:47:59,382 (POSTGRESCollector.java:39) DEBUG - Error while collecting DB parameters: No results were returned by the query.
14:47:59,383 (DBWorkload.java:586) INFO - Output into file: outputfile.88.res
14:47:59,384 (DBWorkload.java:590) INFO - Output Raw data into file: outputfile.88.raw
Our result uploadercom.oltpbenchmark.util.ResultUploader@2928854b
14:47:59,384 (DBWorkload.java:595) INFO - Output summary data into file: outputfile.88.summary
14:47:59,385 (DBWorkload.java:601) INFO - Output db config into file: outputfile.84.db
14:47:59,386 (DBWorkload.java:607) INFO - Output benchmark config into file: outputfile.84.ben
14:47:59,421 (DBWorkload.java:616) INFO - Grouped into Buckets of 5 seconds
14:47:59,435 (DBWorkload.java:644) INFO - *****
14:47:59,435 (DBWorkload.java:645) INFO - Completed Transactions:
ReadRecord/02 [ 289] *****
InsertRecord/03 [ 272] *****
ScanRecord/04 [ 250] *****
UpdateRecord/05 [ 268] *****
DeleteRecord/06 [ 276] *****
ReadModifyWriteRe... [ 0]

14:47:59,436 (DBWorkload.java:646) INFO - Aborted Transactions:
ReadRecord/02 [ 0]
InsertRecord/03 [ 0]
ScanRecord/04 [ 0]
UpdateRecord/05 [ 0]
DeleteRecord/06 [ 0]
ReadModifyWriteRe... [ 0]

14:47:59,436 (DBWorkload.java:647) INFO - Rejected Transactions:
ReadRecord/02 [ 0]
InsertRecord/03 [ 0]
ScanRecord/04 [ 0]
UpdateRecord/05 [ 0]
DeleteRecord/06 [ 0]
ReadModifyWriteRe... [ 0]

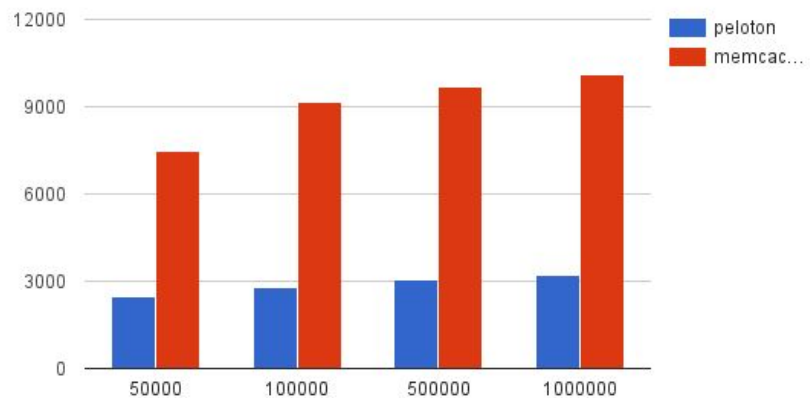
14:47:59,437 (DBWorkload.java:648) INFO - Unexpected Errors:
ReadRecord/02 [ 0]
InsertRecord/03 [ 0]
ScanRecord/04 [ 0]
UpdateRecord/05 [ 0]
DeleteRecord/06 [ 0]
ReadModifyWriteRe... [ 0]
siddharth@siddharth-Q502LAB:~/Documents/CMU/15-721/oltpbench$
```

Implementing Memcached Support



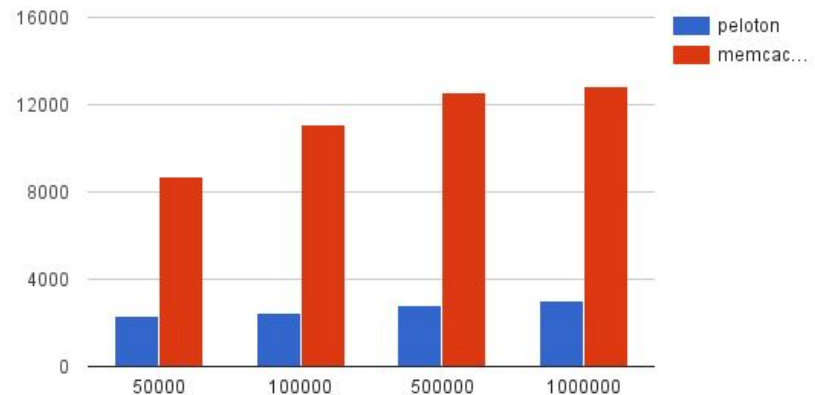
DEMO

peloton and memcached



90:10

peloton and memcached



50:50

Issues

- Performance profiling of wire protocol
 - Potential Bottlenecks due to `std::vector` and `std::string`
- Stack smashing
- Prepared statements!
- SQL Constraints support (INSERT ON CONFLICT UPDATE)
- Performance Tuning
 - Upon profiling we observed that string operations, new and index scan for searching the table
are the major bottlenecks

Future work

- Replacing Berkeley sockets with Libevent for wire protocol
- Integration with Peloton query parser and planner
- Integrate Memcached in new frontend
- Use hand coded execution plan for Memcached operations or compiled query