



WHITEPAPER

# Improve PostgreSQL Performance with Memblaze<sup>®</sup> PBlaze SSD

## Executive Summary

For most companies, cutting down the IT costs and improving the infrastructure's efficiency are the first areas in a Chief Information Officer sight. In RDBMS area, PostgreSQL is a very popular open source database. More small medium business company had migrated their data into the PostgreSQL. In general, the OLTP type database need the low latency storage system indeed, that means to apply the high speed network device and expensive storage system. As SSDs become more common, you'll also hear more about Non-Volatile Memory Express, a.k.a. NVMe. The PBlaze SSD is a kind of NVMe device. Let take a look what's happened when the PostgreSQL run on the Memblaze PBlaze SSD.

In the whitepaper, it compares the performance of an Online Transaction Processing(OLTP) workload executed on PostgreSQL database with PBlaze SSD device and SAS hard disks RAID, and unveils the PBlaze SSD performance and QoS.

## PostgreSQL Introduction

PostgreSQL implements the majority of the SQL:2011 standard, is ACID-compliant and transactional avoiding locking issues using multiversion concurrency control (MVCC), provides immunity to dirty reads and full serializability; handles complex SQL queries using many indexing methods that are not available in other databases.

A page write which is in process during an operating system crash might be only partially completed, leading to an on-disk page that contains a mix of old and new data. The row-level change data normally stored in WAL will not be enough to completely restore such a page during post-crash recovery. Storing the full page image guarantees that the page can be correctly restored.

If turn on the `full_page_writes` parameter, the PostgreSQL server writes the entire content of each disk page to Write-Ahead-Log(WAL) during the first modification of that page after a checkpoint. But it will increase the amount of data that must be written to WAL, thus lead to storage layer get heavy IO operations.

## About the Test

Based on TPCC and modeled after an industry benchmarking association scenario, BenchmarkSQL is an open source and easy to use Java Database Connectivity (JDBC) benchmark application closely resembling the TPC-C standard for OLTP. The test was driven by BenchmarkSQL with a setting of 1000 warehouse and models a set of five transactions driven by a group of simulated operators. The transactions modeled are: New-order, Payment, Order Status, Delivery, Stock Level.

The overall test environment is as follows:

### Server

- DellPowerEdge R730xd
- 2\* Intel XeonE5-2620(6 cores) v3 CPU
- 8\* 8GB DRAM

### Storage

- 1 x Memblaze 1.6T PBlaze4
- 6 x 600GB 15K SAS Disks, construct a RAID0 virtual disk by LSI RAID

### Benchmark Tool

- Benchmarksql 4.1.1, TPC-C test tool for PostgreSQL
- pgcluu 2.4, PostgreSQL performances monitoring and auditing tool

### Software

- PostgreSQL 9.2.14
- CentOS 7.1

### Test Pre-condition

Before testing, following operations needs to be considered:

1. 4KB alignment partition in the PBlaze4 and be sure the partition offset from 1MB.
2. Enable irqbalance service and set the cpu frequency for performance level.
3. Tuning the OS kernel parameter for PostgreSQL application. Avoiding the OS swap function, and tuning the shared memory parameter/file system dirty page parameter. Like the bellows:
 

```
vm.swappiness=0
kernel.shmmax=21474836480
net.core.rmem_max=4194304
net.core.wmem_max=4194304
kernel.sem=50100 64128000 50100 1280
vm.dirty_background_bytes=33554432
```
4. Using XFS, the default block size of PostgreSQL is 8KiB, but in the current linux environment, the operation of “mount xfs” will get” Function not implemented” error with 8KiB blocksize, so setting 4KiB. More allocation groups imply that more parallelism can be achieved when allocating blocks and inodes. The size of nvme device partition is 1601GB, and we will set the agsize=2g. Set inode64/nobarrier/nolargeio/allocsize=16M as the mount option.
5. For PostgreSQL database cluster, set 3/4 total memory as the database effective memory, and set shared\_buffer equal to 6GB, set max connection equal to 600.
6. Configure 1000 warehouse in the Benchmarksql tool and set 128 terminals.
7. Preconditioning PBlaze4 and be sure the file system available capacity less than 50%.

### Test Results

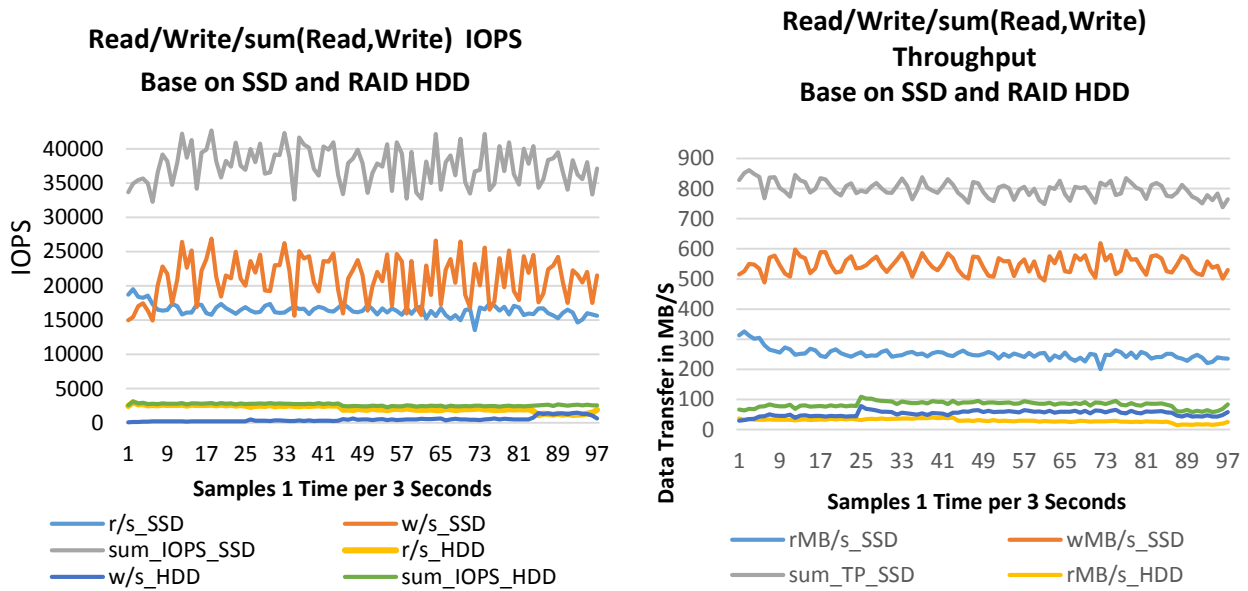


Figure 1, the IOPS and Throughput of Benchmark Test based on Different Storage Media

Figure 1 clearly demonstrates the performance difference of IOPS and throughput for the PBlaze SSD and RAID HDD. Comparing the test results, the sum IOPS of PBlaze SSD is about 10 times the sum IOPS of RAID HDD, and the sum throughput of PBlaze SSD is about 8 times the sum throughput of RAID HDD.

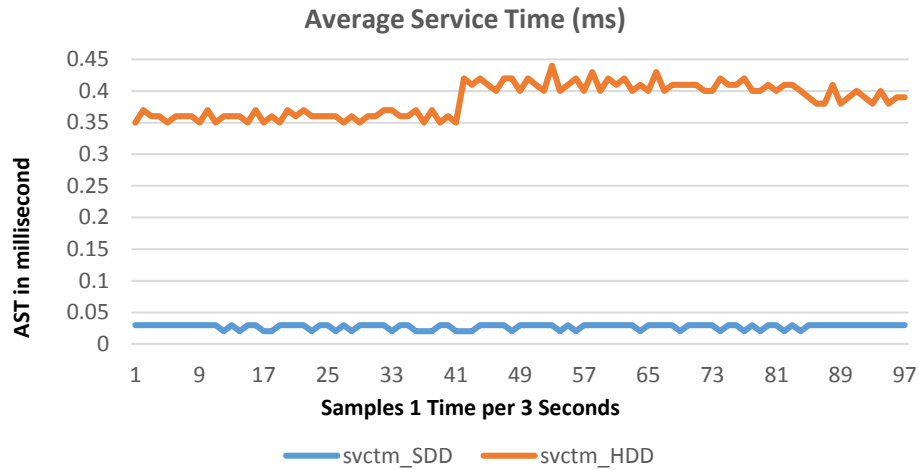


Figure 2, the Average Service Times of Benchmark Test based on Different Physical Device

As showed in Figure2, PBlaze SSD device ensure consistent lower latency.

From the above, Memblaze PBlaze SSD displays its powerful performance for the PostgreSQL database benchmark test.

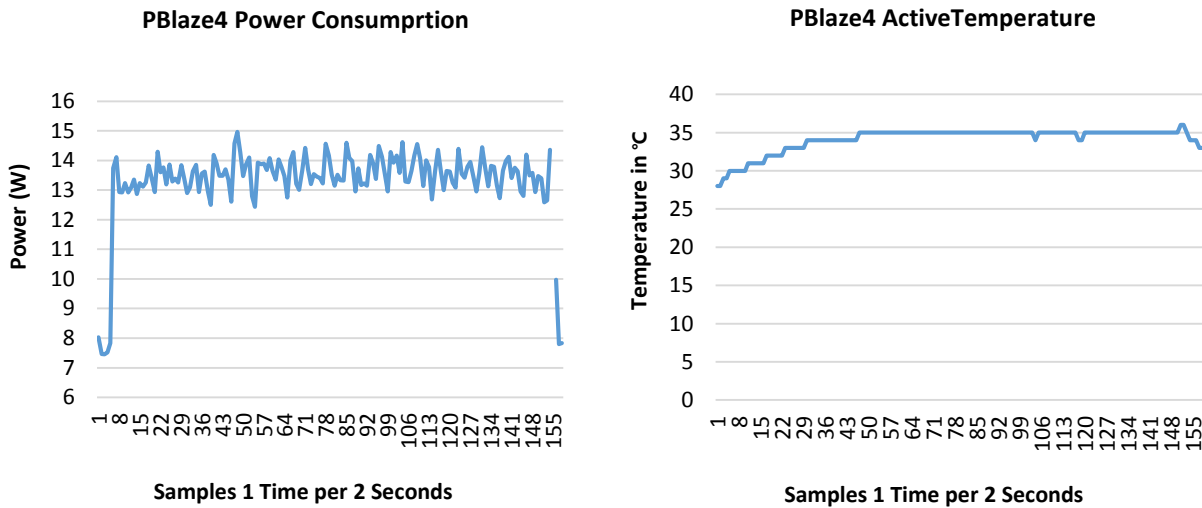


Figure 3, PBlaze SSD Power Consumption and Temperature during the benchmark test

As illustrated in Figure 3, both power consumption and temperature metrics demonstrate that PBlaze SSD ensures stable working state delivery throughout the testing process.

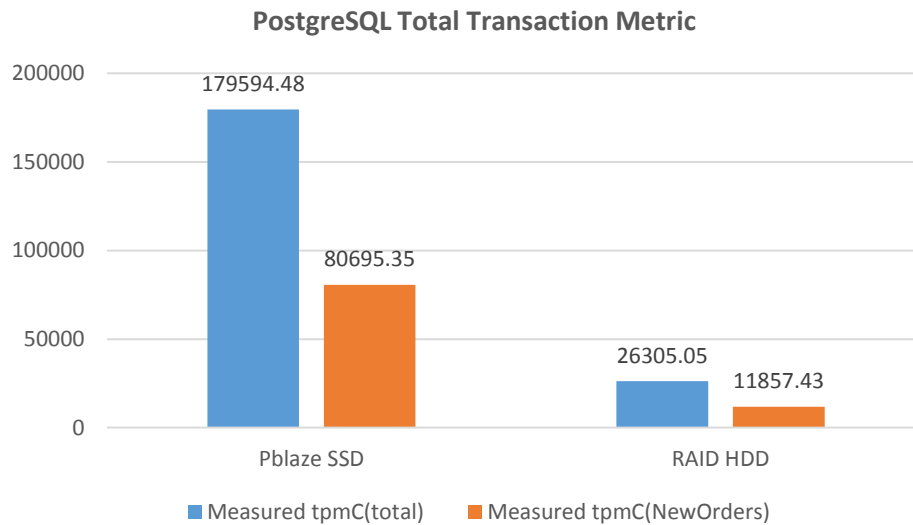


Figure 4, PostgreSQL Total and New-Orders Transaction Metric

As demonstrate in Figure 4, the total transactions per minute (tpmC) includes all five transactions models. The transaction metric based on PBlaze SSD is much higher than the test results based on RAID HDD.

## Conclusions

This test clearly demonstrates the amazing RDBMS benchmark TPMC value based on Memblaze PBlaze SSD. The power and temperature of PBlaze SSD is relatively steady. For the simple RDBMS environment, the Memblaze PBlaze SSD is an idea choice for replacing the special storage system to increase performance and reduce TCO.

### DISCLAIMER

Information in this document is provided in connection with Memblaze products. Memblaze provides this document “as is”, without warranty of any kind, neither expressed nor implied, including, but not limited to, the particular purpose. Memblaze may make improvements and/or changes in this document or in the product described in this document at any time without notice. The products described in this document may contain design defects or errors known as anomalies or errata which may cause the products functions to deviate from published specifications.

### COPYRIGHT

©2016 Memblaze Corp. All rights reserved. No part of this document may be reproduced, transmitted, transcribed, stored in a retrieval system, or translated into any language in any form or by any means without the written permission of Memblaze Corp.

### TRADEMARKS

Memblaze, PBlaze are trademarks of Memblaze Corporation. Other names mentioned in this document are trademarks/registered trademarks of their respective owners.

### USING THIS DOCUMENT

Though Memblaze has reviewed this document and very effort has been made to ensure that this document is current and accurate, more information may have become available subsequent to the production of this guide. In that event, please contact your local Memblaze sales office or your distributor for latest specifications before placing your product order.