PT

Speed completion of Hadoop workloads

27% less time to complete a TeraSort disk-intensive workload*

Boost throughput

36% more GB/s*

Finish Hadoop workloads faster and with higher throughput with Dell PowerEdge R750 servers featuring Dell PowerEdge RAID Controllers (PERC 11)

Compared to an HPE ProLiant DL380 Gen9 server with HPE Smart Array P440ar Controller

If your organization relies on big data analysis in Apache™ Hadoop® to inform everyday decision-making, your servers must be able to process large amounts of data quickly. The faster the data analysis, the more likely it is that your decision-makers are using more recent—and therefore more valuable—data.

Here at Principled Technologies, we used the TeraSort benchmark to measure the Apache Hadoop performance of two servers: a Dell™ PowerEdge™ R750 server with a Dell PERC 11 RAID controller, and an HPE ProLiant DL380 Gen9 server with HPE Smart Array P440ar Controller. The PowerEdge R750 server, featuring the Dell PERC 11 RAID controller, finished the TeraSort workload faster and offered stronger throughput, delivering overall better Hadoop performance.

Dell PowerEdge R750 server with Dell PERC 11 RAID controller vs. HPE ProLiant DL380 Gen9 server with HPE Smart Array P440ar Controller

Finish Hadoop workloads faster and with higher throughput with Dell PowerEdge R750 servers featuring Dell PowerEdge RAID Controllers (PERC 11)

January 2023 (Revised)

How we tested

We deployed two Hadoop environments, one on the Dell PowerEdge R750 server and one on the HPE ProLiant DL380 Gen9 server. Both servers used Linux VMs and had SAS SSDs, which we chose to stress the RAID controllers. The PowerEdge R750 server had a Dell PERC 11 RAID controller, while the HPE ProLiant DL380 Gen9 server had an HPE Smart Array P440ar Controller.

To test the solutions' capabilities around big data workloads, we used the TeraSort benchmark to test Apache Hadoop performance. Apache Hadoop is "a framework that allows for the distributed processing of large data sets" and is "designed to scale up from single servers to thousands of machines."[1] The TeraSort benchmark, part of the HiBench suite of benchmarks, sorts billions of records inside Apache Hadoop, making it a useful way to test Hadoop big data performance.

About the Dell PowerEdge R750 server

The Dell PowerEdge R750 is a full-featured, general-purpose 2U rack server featuring 3rd Gen Intel® Xeon® Scalable processors. According to Dell, the PowerEdge R750 is purpose-built to optimize application performance and acceleration with PCIe Gen 4.0 compatibility, eight channels of memory per CPU, and up to 24 NVMe® drives.[2] It also includes "I/O bandwidth and storage to address data requirements – ideal for: traditional corporate IT, database and analytics, virtual desktop infrastructure, AI/ML, and HPC."[3]

To learn more about the Dell PowerEdge R750, check out the spec sheet at https://i.dell.com/sites/csdocuments/Product_Docs/en/poweredge-R750-spec-sheet.pdf.

Finish Hadoop workloads faster and with higher throughput with Dell PowerEdge R750 servers featuring Dell PowerEdge RAID Controllers (PERC 11)

January 2023 (Revised) | 2

Complete your big data workloads faster

When you can complete your Hadoop workloads faster, you can put your big data to use more quickly and effectively, whether you're using it to inform your next big business strategy or the order of operations for today. In our tests, the Dell PowerEdge R750 server with Dell PERC 11 RAID controller finished a TeraSort workload in 27 percent less time than the HPE ProLiant DL380 Gen9 server with HPE Smart Array P440ar Controller.

Time to complete TeraSort workload
Min:sec | Lower is better

Dell PowerEdge R750 with Dell PERC 11	**4:13**
HPE ProLiant DL380 Gen9 with Smart Array P440ar	**5:47**

■ Dell PowerEdge R750 with Dell PERC 11   ■ HPE ProLiant DL380 Gen9 with Smart Array P440ar
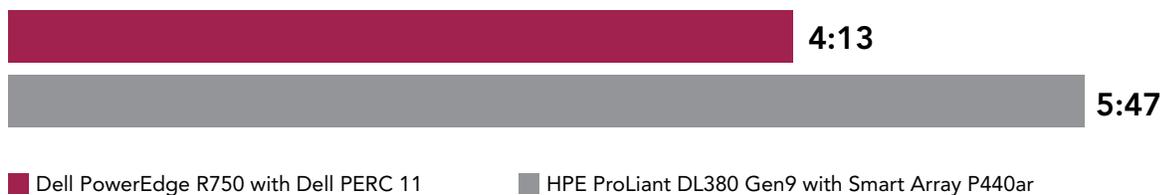
Figure 1: Time required for both solutions to complete a TeraSort workload on Apache Hadoop. Lower is better. Source: Principled Technologies.

Finish Hadoop workloads faster and with higher throughput with Dell PowerEdge R750 servers featuring Dell PowerEdge RAID Controllers (PERC 11)

January 2023 (Revised) | 3

Boost throughput on Apache Hadoop

When we measured the throughput of both solutions, the Dell PowerEdge R750 server featuring the Dell PERC 11 RAID controller delivered 36 percent more gigabytes per second (GB/s) than the HPE solution with the Smart Array P440ar Controller. Higher throughput means that your server solution can sustain a heavier workload in a given amount of time.

Throughput on a TeraSort workload
GB/s | Higher is better

Dell PowerEdge R750 with Dell PERC 11	0.71
HPE ProLiant DL380 Gen9 with Smart Array P440ar	0.52

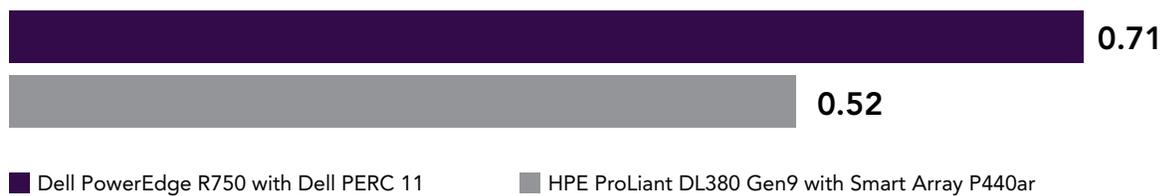■ Dell PowerEdge R750 with Dell PERC 11   ■ HPE ProLiant DL380 Gen9 with Smart Array P440ar

Figure 2: Throughput for both solutions during a TeraSort workload. Higher is better. Source: Principled Technologies.

About the Dell PERC 11 RAID controller

According to Dell, PERC 11 RAID controllers offer support for PCIe Gen 4, support for hot-swapping devices, non-volatile cache, secure enterprise key manager security, and more.[4]

Finish Hadoop workloads faster and with higher throughput with Dell PowerEdge R750 servers featuring Dell PowerEdge RAID Controllers (PERC 11)

January 2023 (Revised) | 4

Conclusion

In the area of big data analysis, speed is vital. We used the TeraSort benchmark to test two Apache Hadoop environments: a Dell PowerEdge R750 server with a Dell PERC 11 RAID controller, and an HPE ProLiant DL380 Gen9 server with an HPE Smart Array P440ar Controller. In both time to complete the workload and throughput, the Dell PERC 11 solution came out on top.

_____

1. "Apache Hadoop," accessed November 14, 2022, https://hadoop.apache.org/.
2. Dell, "Dell EMC PowerEdge R750 Spec Sheet," accessed November 14, 2022, https://i.dell.com/sites/csdocuments/Product_Docs/en/poweredge-R750-spec-sheet.pdf.
3. Dell, "Dell EMC PowerEdge R750 Spec Sheet."
4. Dell, "Dell PowerEdge RAID Controller 11 User's Guide PERC H755 adapter, H755 front SAS, H755N front NVMe, H755 MX adapter, H750 adapter SAS, H355 adapter SAS, H355 front SAS, H350 adapter SAS, H350 Mini Mo," accessed No-vember 14, 2022, https://www.dell.com/support/manuals/en-pa/perc-h755/perc11_ug/technical-specifications-of-perc-11-cards?guid=guid-aaaf8b59-903f-49c1-8832-f3997d125edf&lang=en-pa.

This project was commissioned by Dell Technologies.

Read the science behind this report at https://facts.pt/3m5epzN ▶

Principled Technologies®

Facts matter.®

Finish Hadoop workloads faster and with higher throughput with Dell PowerEdge R750 servers featuring Dell PowerEdge RAID Controllers (PERC 11)

January 2023 (Revised) | 5