



Complete Hadoop workloads faster

13.9% less time to complete a TeraSort disk-intensive workload*

Increase throughput

16.3% more GB/s*

Boost throughput for big data workloads with Dell PowerEdge R750 servers featuring Dell PowerEdge RAID Controllers (PERC 11)

Compared to a legacy solution, this new Dell solution increased Hadoop performance in our testing

For organizations relying on Apache™ Hadoop® for big data storage and analysis, it's important to choose a computing solution that can deliver fast analytics for very large quantities of data. The more up-to-date the data and analysis, the more likely that businesses can effectively use that data to make an informed decision.

At Principled Technologies, we used the TeraSort benchmark to measure the Apache Hadoop performance of two server solutions in virtualized environments:

- A Dell™ PowerEdge™ R750 server with a Dell PERC 11 RAID controller
- A previous-generation Dell PowerEdge R740 server with a PERC 10 RAID controller

We found that in both metrics we tested—time to complete a workload and throughput while completing that workload—the PowerEdge R750 server featuring the Dell PERC 11 RAID controller offered stronger performance, finishing the workload faster with higher throughput.

**Dell PowerEdge R750 server with Dell PERC 11 RAID controller vs. Dell PowerEdge R740 server with PERC 10 RAID controller*

How we tested

We deployed two Hadoop environments, one with the Dell PowerEdge R750 server with Dell PERC 11 RAID controller and one with the PowerEdge R740 server with PERC 10 RAID controller. Both servers had SAS SSDs, which we chose to stress the RAID controller, and both environments used Linux VMs.

To test the solutions' capabilities around big data workloads, we used the TeraSort benchmark, which measures 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 big data workloads faster

Whether you're looking to big data for insights about everyday operations or answers about a new offering, the faster a solution can complete a big data workload, the more quickly your organization can use that data effectively. In our testing, the Dell PowerEdge R750 server with Dell PERC 11 RAID controller completed the TeraSort workload in 13.9 percent less time than the previous-generation PowerEdge R740 with PERC 10 RAID controller, showing that for this type of big data workload, newer is faster.

Time to complete TeraSort workload

Min:sec | Lower is better

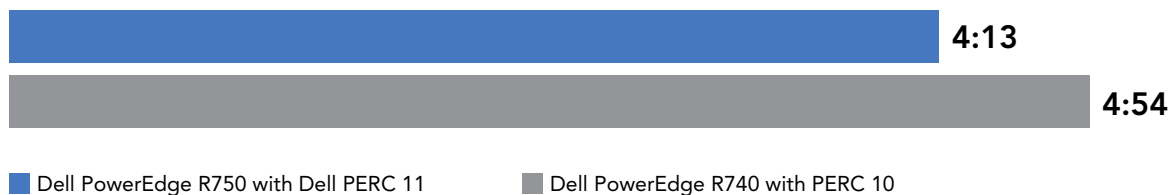


Figure 1: Time required for both solutions to complete a TeraSort workload on Apache Hadoop. Lower is better.
Source: Principled Technologies.

Increase throughput for big data workloads

In addition to measuring the time it took both solutions to complete a TeraSort workload, we measured the throughput, or amount of data per second, that both servers processed. The Dell PowerEdge R750 server featuring the Dell PERC 11 RAID controller sustained 16.3 percent higher throughput than the previous-generation solution.

Throughput on a TeraSort workload

GB/s | Higher is better

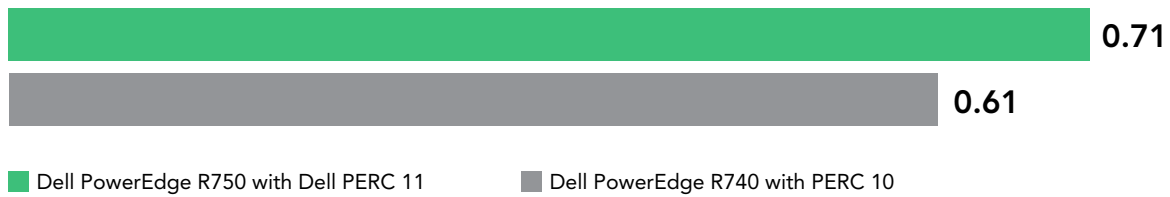


Figure 2: Throughput for both solutions during a TeraSort workload. Higher is better. Source: Principled Technologies.

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





Conclusion

For the most up-to-date decision-making based on big data, you need a solution that can analyze that data quickly. In our testing with the TeraSort benchmark on two Apache Hadoop environments, a Dell PowerEdge R750 server with Dell PERC 11 RAID controller completed a big data workload faster and with higher throughput than a previous-generation PowerEdge R740 server with PERC 10 RAID controller.

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 November 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/mJepPM7>



Facts matter.®

Principled Technologies is a registered trademark of Principled Technologies, Inc. All other product names are the trademarks of their respective owners. For additional information, review the science behind this report.