Executive summary

Run compute-intensive Apache Hadoop big data workloads faster with Dell EMC PowerEdge R640 servers

A Hadoop cluster of PowerEdge R640 servers powered by 2nd Generation Intel Xeon Scalable processors completed three compute-heavy big data workloads in less time than previous-generation Dell EMC PowerEdge R630 servers by processing more data per second

Your organization likely generates large volumes of data from numerous sources continuously. This data can range from how long users are on a web page to the length of routine sales team video calls. Extracting insight from this disparate information often requires running complex, compute-intensive workloads quickly on multiple data sets.

Aging servers typically cannot deliver the speed that newer servers can offer for these compute-intensive workloads. Current-generation servers can deliver a performance improvement that helps your organization now and allows you to continue accumulating and using data effectively. Faster servers can process and analyze data more quickly, so marketing teams, for example, can more quickly determine whom to target for their next email campaign.

Analyze more data per second

Up to 112% greater throughput while analyzing words in a large document

Identify topics in documents more quickly

Up to 52% less waiting for document analysis

In our data center, an Apache Spark™ cluster of current-generation Dell EMC™ PowerEdge™ R640 servers featuring 2nd Generation Intel® Xeon® Scalable processors outperformed a cluster of previous-generation Dell EMC PowerEdge R630 servers in three compute-intensive, big data workloads from Intel HiBench 7.1. The workloads identified topics in a large document, classified information to make a prediction, and counted words in a data set. Moving these workloads to new PowerEdge R640 servers and getting better performance can help your organization meet today's demands and offer the computing power necessary to face the challenges of tomorrow.

In addition to newer, faster Intel processors to run queries and algorithms, the PowerEdge R640 servers have more drive bays than their predecessors. More drive bays allow you to add more storage to each server and store more data, which could help prevent server bottlenecks and promote speedy access to databases.

Complete compute-heavy Hadoop big data workloads in less time

The cluster of three current-generation Dell EMC PowerEdge R640 servers powered by 2nd Generation Intel Xeon Scalable processors ran three Hadoop® workloads from Intel HiBench 7.1: Latent Dirichlet allocation (LDA), RandomForest (RF), and WordCount. For the LDA workload, which analyzes words in a document and then provides a summary, the PowerEdge R640 solution achieved throughput of more than 4 MB per second—more than double the throughput of the previous-generation solution. The current-generation solution needed just over 17 minutes to process 4.5 GB of data for the LDA test. Compared to the previous-generation solution, which needed 36 minutes, the solution completed the workload in less than half the time. Processing more data per second could allow your marketing team, for example, to organize customer reviews on a product or service more quickly.

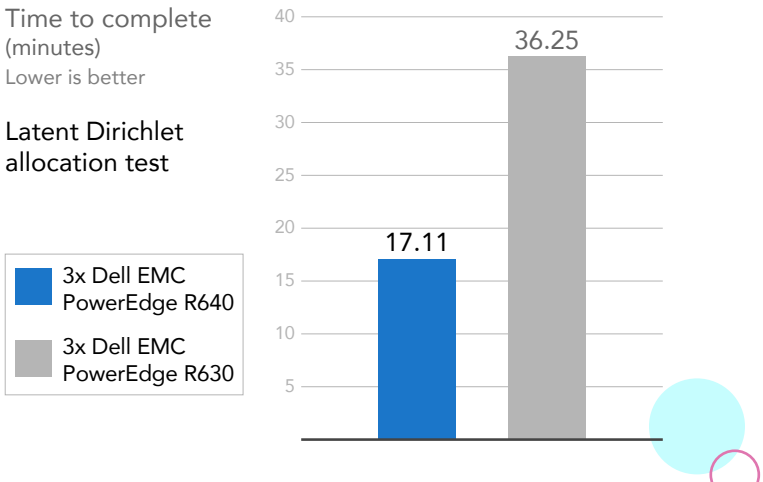
The PowerEdge R640 solution also completed the two other Hadoop big data workloads in less time while processing more data per second. Not only could you deliver analysis to decision makers more quickly with the PowerEdge R640, but you could use the extra time, for example, to re-run the LDA workload to ensure accuracy.

Conclusion

Moving compute-intensive, Hadoop big data workloads to current-generation Dell EMC PowerEdge R640 servers powered by 2nd Generation Intel Xeon Scalable processors could allow your organization to better meet the data analysis challenges of today. Faster analysis of large data sets means getting insight into your organization, products, and services sooner, which could help your organization grow and beat its competition.

Identify topics in documents more quickly

Up to 52% less waiting for document analysis



The winning solution at a glance

Dell EMC PowerEdge R640 server

- Dense 1U, two-socket server
- 24 DDR4 DIMM slots
- Up to 76.8 TB of storage

2nd Generation Intel Xeon Scalable processor platform

- Offers multiple levels of performance to match your workloads, including Bronze, Silver, Gold, and Platinum
- Supports Intel Optane™ DC persistent memory,[1] a new memory and storage technology for workload acceleration

Read the report at http://facts.pt/jqj8hg1 ▶

1 Intel, "2nd Gen Intel Xeon Scalable Processors Brief," accessed November 7, 2019, https://www.intel.com/content/www/us/en/products/docs/processors/xeon/2nd-gen-xeon-scalable-processors-brief.html.

Principled Technologies®

Facts matter.®