

Top Content Discovery Platform Makes a Solid Choice for Solidigm SSDs

Content-recommendation company Taboola prefers Solidigm™ solid state drives (SSDs) because of their reliability and price performance.



About Taboola

Taboola is the world's largest content-discovery platform on the open web, with a recommendation engine built on machine learning (ML) and massive datasets. It runs nine data centers with 10,000 servers worldwide, with tens of petabytes of storage on SSDs.

Reliability is Taboola's top prerequisite in SSD purchases. Price performance is the next most important factor. Taboola chose several years ago to standardize on Solidigm (formerly Intel®) SSDs based on their reliability and performance.

Taboola describes itself as the world's largest discovery platform, serving 360 billion tailored content recommendations to more than one billion people across the web each month. Powered by deep learning (DL) technology, Taboola uses unique data about people's interests and information consumption to recommend the right content to the right person at the right time. Taboola's content recommendations are found on high-profile sites like Bloomberg, NBC News, MSN, The Independent, and The Weather Channel, among many others.

Taboola's algorithm analyzes customer content and extracts a large number of "signals" used to match that content to the people who are most likely to engage with it. The DL technology begins recommending the content to those people, and then it refines its targeting based on people's actual reactions to the recommendations.

The use of artificial intelligence (AI) to predict audience interests is a centerpiece of the Taboola model, but that's not the most interesting technological challenge for Ariel Pisetzky, VP of IT at Taboola¹. For him, the greatest challenge is the mathematical and engineering problem of how best to scale Taboola's systems within their narrow budget constraints. "There are a lot of scale-out software solutions today in the open-source community that are meant to scale from a single server to thousands of nodes, so that engineering work has been done," says Pisetzky. "Now we need to make sure we can connect all of these pieces of the puzzle together in a coherent way, and in a way that allows a few people to manage a large install base."

Problem: Scaling storage

With nine data centers housing some 10,000 servers around the globe, Taboola has built its own private cloud, with each data center containing its own high-performance computing (HPC) infrastructure—essentially comprising small supercomputers. Each data center requires storage at a



massive scale to feed the machine learning (ML) algorithms at the core of its content-recommendation engine.

At present, Taboola processes around 100 TB of data per day, and it has tens of petabytes of storage capacity on SSDs distributed across its data centers.

Importantly, Taboola uses a hyperconverged infrastructure (HCI), meaning that its data centers are populated with multiple self-contained “building block” computing systems that each include their own tightly integrated computer, networking, and storage components. This architecture, which serves the needs of Taboola’s recommendation engine well, places a premium on the reliability of SSD storage.

If data center storage were handled separately and segregated from the compute element, the reliability of individual SSDs would be less important because failed units could be more easily identified, located, and swapped out for new ones. But because the SSDs are distributed among hundreds or thousands of hyperconverged systems, the processes for identifying and swapping out failed units are more labor-intensive, and maintenance costs rise quickly with small increases in failure rates.



Because a key objective of Taboola IT is to enable the smallest possible number of people to manage the largest installation base, the reliability of SSDs is a critical factor in the company’s purchase decisions. Performance and price enter the equation only after a high threshold of reliability is met.

Solution: Solidigm is the preferred SSD partner

In the past, Taboola purchased SSDs from multiple vendors. In Taboola’s experience, however, the reliability of the SSDs from other vendors did not meet their quality and reliability needs. Taboola decided to make a switch to purchasing Solidigm SSDs for the bulk of its data center storage needs. Solidigm SSDs provide the rock-solid reliability that Taboola requires and the high levels of performance it looks for, all at an attractive price.

Solidigm is a proven leader in SSD endurance. Solidigm’s first-generation quad-level cell (QLC) NAND drive for the data center—the Solidigm (formerly Intel) SSD D5-P4320—already offered up to 4x higher endurance than a competitor’s QLC NAND SSD. And the newer Solidigm (formerly Intel) SSD D5-P5316 provides industry-leading levels of endurance for QLC NAND SSDs at 0.41 drive writes per day (DWPD), and up to a 5x increase in random write endurance over previous-generation Solidigm QLC NAND SSDs.²

Most Taboola workloads revolve around feeding data to ML algorithms. Because this is a read-heavy type of workload, QLC SSDs are well suited to performing much of the storage work because they generally provide higher capacity at a lower cost per gigabyte (GB) than do triple-level cell (TLC) SSDs. For a few specialized types of workloads, such as large Microsoft SQL Server databases, Taboola finds it worthwhile to invest in TLC drives for even better endurance and reliability in a write-heavy environment.



Benefits

With a reliable storage infrastructure built on Solidigm SSDs, Taboola is able to scale its recommendation engine business even further with confidence that the storage will be there when it’s needed. The high reliability of these SSDs in Taboola’s decentralized, hyperconverged storage architecture keeps maintenance costs in check. This, in turn, helps the company cope with the issue of IT worker shortages.

“There is a certain ratio of workers to servers,” says Pisetsky, “and there aren’t enough engineers graduating every year, so we constantly need to be better with the resources we have.” One way Taboola avoids the need to scale its workforce in proportion to its infrastructure is by scaling its operations with larger and more reliable SSDs.

And Taboola knows that as it purchases SSDs over time, it will always be gaining performance advantages from Solidigm’s commitment to staying at the front edge of evolving SSD technologies.

Taboola is not a large company compared to the giant hyperscalers. Instead, with 10,000 servers, Taboola is at a scale that many medium-sized companies in the 1,000-server to 50,000-server range can relate to. The success of Taboola’s storage strategy of focusing on SSD reliability as a top priority may be a strategy that will also work well for many other companies.

Learn more

[Solidigm data center SSDs](#)

SOLIDIGM™

1 Interview with Ariel Pisetsky, VP of IT at Taboola, conducted May 26, 2022.

2 Intel. “QLC NAND Technology Is Ready for Mainstream Use in the Data Center.” 2021. <https://www.solidigm.com/content/dam/newco-aem-site/master/site/resources/qlc-nand-technology-ready-mainstream-use-data-center/qlc-nand-ready-for-data-center-white-paper.pdf>

3 [Repeat endnote 1] Interview with Ariel Pisetsky, VP of IT at Taboola, conducted May 26, 2022.

All information provided is subject to change at any time, without notice. Solidigm™ may make changes to manufacturing life cycle, specifications, and product descriptions at any time, without notice. The information herein is provided “as-is” and Solidigm does not make any representations or warranties whatsoever regarding accuracy of the information, nor on the product features, availability, functionality, or compatibility of the products listed. Please contact system vendor for more information on specific products or systems.

Refer to the spec sheet for formal definitions of product properties and features.

Solidigm technologies may require enabled hardware, software or service activation. No product or component can be absolutely secure. Your costs and results may vary. Performance varies by use, configuration and other factors. Other names and brands may be claimed as property of others. See our complete legal Notices and Disclaimers. Solidigm is committed to respecting human rights and avoiding complicity in human rights abuses. Solidigm products and software are intended only to be used in applications that do not cause or contribute to a violation of an internationally recognized human right.

Solidigm and the Solidigm logo are trademarks of Solidigm. Intel is a registered trademark of Intel Corporation. All other trademarks are the property of their respective owners.

© Solidigm 2022. All rights reserved.