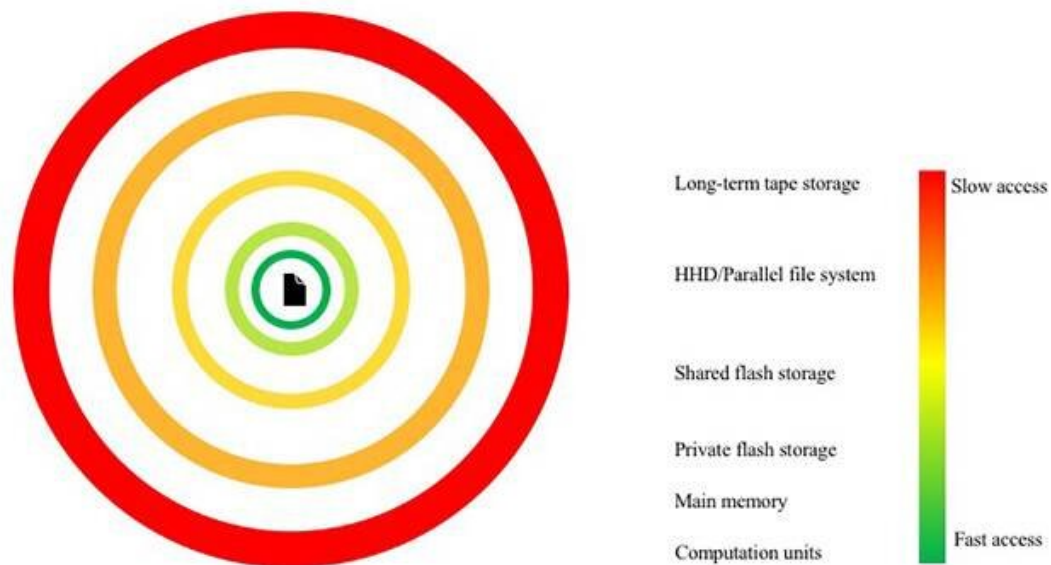


# A continuous data supply ensures data-intensive simulations can run at maximum speed

November 9 2020



Orchestrating the movement of data between more remote storage layers and main memory can speed up simulations by 2.5 times. Credit: KAUST

A pre-emptive memory management system developed by KAUST researchers can speed up data-intensive simulations by 2.5 times by eliminating delays due to slow data delivery. The development elegantly and transparently addresses one of the most stubborn bottlenecks in modern supercomputing—delivering data from memory fast enough to

keep up with computations.

"Reducing the movement of data while keeping it close to the computing hardware is one of the most daunting challenges facing computational scientists handling big data," explains Hatem Ltaief from the research team. "This is exacerbated by the widening gap between computational speed and memory transmission capacity, and the need to store high-volume data on remote storage media."

The key challenge in processing big data is the cost and scale of storing the data in memory. The faster the memory, the more expensive it is, and the faster the data need to be moved between computing elements. Because only relatively small capacities of the fastest memory are available on even the most powerful supercomputing platforms, system engineers add successively larger, slower and more remote layers of memory to hold the tera- and petabytes of data typical of [big data](#) sets.

"It is in this hostile landscape that our system comes into play by reducing the overhead of moving data in and out of remote storage hardware," says Ltaief.

Ltaief with colleagues David Keyes and Tariq Alturkestani developed their multilayer buffer system (MLBS) to work proactively to maintain the data as close as possible to the computing hardware by orchestrating data movement among memory layers.

"MLBS relies on a multilevel buffering technique that outsmarts the simulation by making it 'see' all the hundreds of petabytes of data as being in fast memory," says Alturkestani. "The buffering mechanism prevents the application from stalling when it would have needed to access data located on remote storage, allowing the application to proceed at full speed with asynchronous computing operations."

This synergism provided by MLBS achieved a speedup of 2.5 times for a three-dimensional seismic exploration simulation involving hundreds of petabytes of data movements using KAUST's Shaheen-2 supercomputer.

"This approach also reduces the energy required to move data to and from remote storage media, which can be hundreds of times higher than the energy to perform a single computation on local [memory](#)," says Ltaief. "Using MLBS, we can mitigate the energy overhead of data movement, which is one of the main goals of our center."

**More information:** Tariq Alturkestani et al. Maximizing I/O Bandwidth for Reverse Time Migration on Heterogeneous Large-Scale Systems, *Euro-Par 2020: Parallel Processing* (2020). [DOI: 10.1007/978-3-030-57675-2\\_17](https://doi.org/10.1007/978-3-030-57675-2_17)

Provided by King Abdullah University of Science and Technology

Citation: A continuous data supply ensures data-intensive simulations can run at maximum speed (2020, November 9) retrieved 2 May 2024 from <https://techxplore.com/news/2020-11-data-intensive-simulations-maximum.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.
---