

# Researchers demonstrate breakthrough storage performance for big data applications

July 22 2011

---

Researchers from IBM today demonstrated the future of large-scale storage systems by successfully scanning 10 billion files on a single system in just 43 minutes, shattering the previous record of one billion files in three hours by a factor of 37.

Growing at unprecedented scales, this advance unifies data environments on a single platform, instead of being distributed across several systems that must be separately managed. It also dramatically reduces and simplifies data management tasks, allowing more information to be stored in the same technology, rather than continuing to buy more and more storage.

In 1998, IBM Researchers unveiled a highly scalable, clustered parallel file system called General Parallel File System (GPFS), which was furthered tuned to make this breakthrough possible. GPFS represents a major advance of scaling for storage performance and capacity, while keeping management costs flat. This innovation could help organizations cope with the exploding growth of data, transactions and digitally-aware sensors and other devices that comprise Smarter Planet systems. It is ideally suited for applications requiring high-speed access to large volumes of data such as data mining to determine customer buying behaviors across [massive data](#) sets, seismic data processing, risk management and financial analysis, weather modeling and scientific research.

## Driving New Levels of Storage Performance

Today's breakthrough was achieved using GPFS running on a cluster of 10 eight core systems and solid state storage, taking 43 minutes to perform this selection. The GPFS management rules engine provides the comprehensive capabilities to service any data management task.

GPFS's advanced algorithm makes possible the full use of all [processor cores](#) on all of these machines in all phases of the task (data read, sorting and rules evaluation).

GPFS exploits the solid state storage appliances with only 6.8 terabytes of capacity for excellent random performance and high data transfer rates for containing the metadata storage. The appliances sustainably perform hundreds of millions of data input-output operations, while GPFS continuously identifies, selects and sorts the right set of files among the 10 billion on the system.

"Today's demonstration of GPFS scalability will pave the way for new products that address the challenges of a rapidly growing, multi-zettabyte world," said Doug Balog, vice president, storage platforms, IBM. "This has the potential to enable much larger data environments to be unified on a single platform and dramatically reduce and simplify data management tasks such as data placement, aging, backup and migration of individual files."

The previous record was also set by IBM researchers at the Supercomputing 2007 conference in Reno, NV, where they demonstrated the ability to scan one billion files in three hours.

"Businesses in every industry are looking to the future of storage and data management as we face a problem springing from the very core of our success – managing the massive amounts of data we create on a daily

basis," said Bruce Hillsberg, director of [storage systems](#), IBM Research – Almaden. "From banking systems to MRIs and traffic sensors, our day-to-day lives are engulfed in data. But, it can only be useful if it is effectively stored, analyzed and applied, and businesses and governments have relied on smarter technology systems as the means to manage and leverage the constant influx of data and turn it into valuable insights."

IBM Research continues to develop innovative storage technologies to help clients not only manage data proliferation, but harness data to create new services. In the past year alone, IBM storage products included over five significant storage innovations invented by IBM Research including IBM Easy Tier, Storwize V7000, Scale-out Network Attached Storage (SONAS), IBM Information Archive and IBM Long Term File System (LTFS).

As the size of digital data increased 47 percent over last year, businesses are under tremendous pressure to quickly turn data into actionable insights, but grapple with how to manage and store it all. As new applications emerge in industries from financial services to healthcare, traditional data management systems will be unable to perform common but critical storage management tasks, leaving organizations exposed to critical data loss.

Anticipating these [storage](#) challenges decades ago, researchers from IBM Research – Almaden created GPFS to help businesses cope with the exploding growth of data, transactions and digitally-aware devices on a single system. Already deployed to perform tasks like backup, information lifecycle management, disaster recovery and content distribution, this technology's unique approach overcomes the challenge of managing unprecedented large file systems with the combination of multi-system parallelization and fast access to file system metadata stored on a [solid state storage](#) appliance.

**More information:** Additional details on the breakthrough can be found [here](#).

Provided by IBM

Citation: Researchers demonstrate breakthrough storage performance for big data applications (2011, July 22) retrieved 5 May 2024 from <https://phys.org/news/2011-07-breakthrough-storage-big-applications.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.