

# 120 petabytes: IBM building largest data storage array ever

August 29 2011, by Bob Yirka

---

(PhysOrg.com) -- As computers big and small grow ever faster, new and better ways to store more data must be developed as well to keep up with the demand. It wasn't all that long ago that a 1 gigabyte hard drive on a personal computer seemed the stuff of science fiction. On the higher end, ever faster supercomputers require not just more data storage but an ability to save and retrieve data faster as well, otherwise they'd spend more and more time dedicated to doing nothing but searching for that data. To address the situation, IBM is apparently hard at work assembling the largest data array ever, which will utilize the fastest data storage and retrieval system ever devised, allowing for the storage of 120 petabytes of data (120 million gigabytes) using its newly refined [GPFS](#) file system that is capable of indexing 10 billion files in just forty three minutes.

Housed in IBM's Almaden, California research facility, the new as yet unnamed data array is reportedly being underwritten by an unnamed client with a lot of money to spend. Based on the situation with the U.S. government's financial situation, it seems more likely the client is someone like Microsoft, Apple or Google, though for what purpose such a huge array would serve is a mystery. Traditionally, supercomputers and their huge data servers have been used to model weather, military or physics experiments. Now however, the focus might be on ways to crunch ever more data in a faster fashion, which sounds a lot like something Google might be interested in doing, especially when you consider that so much data is moving onto the cloud.

The array works by connecting 200,000 standard (likely 1 terabyte drives) together in a traditional warehouse, though more tightly compacted together than usual in special extra wide drawers. The drives will be kept cool by normal air-conditioning in conjunction with a water cooling system, something [IBM](#) has been using for years with its supercomputers; when finished the array will dwarf current systems, which are typically able to store just 15 petabytes of data. To make sure drive failures don't slow down computations on the supercomputer, data is stored in a redundant fashion over multiple drives, which are striped to allow for accessing large chunks of data simultaneously. When a drive dies, data is moved slowly to its replacement, spreading processing time over a relatively long period, allowing the processor to focus on its main tasks.

The new array, while seemingly for the moment to be almost overkill, is likely to be followed by other even larger systems as more and more data storage becomes necessary. The hope though, is that some new breakthrough will occur that will allow for huge increases in [data storage capacity](#) that won't mean simply adding more and more conventional [storage](#) medium to existing technology.

**More information:** via [Technology Review](#)

© 2011 PhysOrg.com

Citation: 120 petabytes: IBM building largest data storage array ever (2011, August 29) retrieved 19 April 2024 from <https://phys.org/news/2011-08-petabytes-ibm-largest-storage-array.html>

<p>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.</p>
--