

IBM's new architecture can double analytics processing speed

19 November 2010

At the Supercomputing 2010 conference, IBM today unveiled details of a new storage architecture design, created by IBM scientists, that will convert terabytes of pure information into actionable insights twice as fast as previously possible.

Ideally suited for cloud computing applications and data-intensive workloads such as digital media, [data mining](#) and financial analytics, this new architecture will shave hours off of complex computations without requiring heavy infrastructure investment. IBM won the [Storage Challenge](#) competition for presenting the most innovative and effective design in high performance computing with the best measurements of performance, scalability and storage subsystem utilization.

Running analytics applications on extremely large data sets is becoming increasingly important, but organizations can only continue to increase the size of their storage facilities so much. As businesses search for ways to harness their large stored data to achieve new levels of business insight, they need alternative solutions like cloud computing to keep up with growing data requirements as well as tackling workload flexibility through the rapid provisioning of system resources for different types of workloads.

"Businesses are literally running into walls, unable to keep up with the vast amounts of data generated on a daily basis," said Prasenjit Sarkar, Master Inventor, Storage Analytics and Resiliency, IBM Research - Almaden. "We constantly research and develop the industry's most advanced storage technologies to solve the world's biggest data problems. This new way of storage partitioning is another step forward on this path as it gives businesses faster time-to-insight without concern for traditional storage limitations."

Created at IBM Research - Almaden, the new General Parallel File System-Shared Nothing

Cluster (GPFS-SNC) architecture is designed to provide higher availability through advanced clustering technologies, dynamic file system management and advanced data replication techniques. By "sharing nothing," new levels of availability, performance and scaling are achievable. GPFS-SNC is a distributed computing architecture in which each node is self-sufficient; tasks are then divided up between these independent computers and no one waits on the other.

IBM's current GPFS technology offering is the core technology for IBM's High Performance Computing Systems, IBM's Information Archive, IBM Scale-Out NAS (SONAS), and the IBM Smart Business Compute Cloud. These research lab innovations enable future expansion of those offerings to further tackle tough big data problems.

For instance, large financial institutions run complex algorithms to analyze risk based on petabytes of data. With billions of files spread across multiple computing platforms and stored across the world, these mission-critical calculations require significant IT resource and cost because of their complexity. Using this GPFS-SNC design, running this complex analytics workload could become much more efficient, as the design provides a common file system and namespace across disparate computing platforms, streamlining the process and reducing disk space.

Source: IBM

APA citation: IBM's new architecture can double analytics processing speed (2010, November 19)
retrieved 27 January 2021 from <https://phys.org/news/2010-11-ibm-architecture-analytics.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.