



**Big Data –
creating business
value by leveraging
data to drive product
innovation, reduce
time-to-discovery,
and replace physical
experimentation with
simulation.**

What is Big Data

Big data is a term applied to a new generation of software, applications, and system and storage architecture, all designed to derive business value from unstructured data. Advanced tools, software, and systems are required to capture, store, manage, and analyze the data sets, all in a timeframe that preserves the intrinsic value of the data.

The big data concept has existed for over ten years, but was originally the exclusive domain of advanced university research and national laboratories. The term big data is now applied more broadly to cover commercial environments, thanks to massive technology advances, including: faster and cheaper processing power, clustered computing, lower cost SATA storage, and network performance improvements that enable companies to perform computing tasks that previously required highly sophisticated and costly computing systems.

Creating Value with Big Data

Four distinct applications segments comprise the big data market, each with varying levels of need for performance and scalability. The four big data segments are: 1) Design (engineering collaboration), 2) Discover (core simulation – supplanting physical experimentation), 3) Decide (analytics), and 4) Deposit (Web 2.0 and data warehousing). Panasas ActiveStor scale-out NAS systems, together with PanFS, the Panasas high performance parallel file system, are designed to specifically address the design and discover segments.

DESIGN

Value is created in the big data design segment by using data to drive product innovation, improve time-to-design, create transparency in the process flow, and by massively reducing the cost of physical design (or by replacing it all-together). For example, 3D modeling software allows aeronautic engineers and automotive designers to experiment with esoteric design without the costly need for physical models or wind tunnels. Designers, utilizing software tools that model computational fluid dynamics (CFD), can model air and fluid flow, and add thermal and acoustic characteristics to a model to get an accurate view of a physical design. Once they have their models, engineers can perform “what if” scenarios that allow their designs to be thoroughly, yet cheaply and easily vetted.

Many other benefits are enjoyed in the big data design segment in areas like engineering collaboration – whereby groups of engineers have access to a consistent data set in order to reduce the possibility for error. Similarly, manufacturing efficiencies can be gained by combining data from field failures to improve product quality.

While the variety of design applications is vast, all share a common need for a single global namespace to aggregate massive amounts of unstructured data, seamless storage scalability and the highest parallel performance possible to achieve rapid time-to-results. Panasas ActiveStor, with PanFS, the Panasas high performance parallel file system, is a proven leader in the design segment where performance and scale make the difference.

DISCOVER

Organizations in the big data discover segment create value by replacing costly physical scientific experimentation in the field with innovative computer simulation.

The traditional means of scientific discovery has depended on physical experimentation to prove a hypothesis. For example, within the biosciences sector, wet laboratory experimentation has been the norm. However, physical experimentation is time consuming, expensive, and requires highly skilled researchers often to perform repetitive tasks. Just as traditional physical

experimentation has a transformative effect, new computer methods have allowed researchers to combine differing data sets to unleash new discovery faster. This can be done by improving data transparency, simulating physical experiments, fusing disparate data types such as seismic and telemetric models, or other advanced mathematical modeling in the biosciences, energy, finance, government, and academic research markets.

Researchers in the Discover segment are both big data creators as well as big data users. As such, this segment has a high need for extreme, seamless scale as the end point is unknown. A single namespace that grows to meet the exponential growth in unstructured data is essential for Discover applications. The input data sets can grow to the tens or hundreds of terabytes which would choke a traditional NAS system. The Panasas parallel file system allows all nodes on a compute cluster to have parallel and direct access to the storage pool, enabling simulations to scale with the workload. Panasas ActiveStor is a proven leader in the design segment where performance and scale matters.

Big Data Application Segments

