WHITE PAPER
Sponsored content

Big Data Is Mission-Critical. Is Yours Protected?

New storage solutions are tailored for today's infrastructure

In just a few years, big data has advanced from scattered experimental projects to achieve mission-critical status in digital enterprises, and its importance is increasing. According to IDC, by 2020, organizations able to analyze all relevant data and deliver actionable information will earn $430 billion more than their less analytically oriented peers.[1] Big-data analytics, once performed on an occasional basis, are now performed daily at many enterprises, including Amazon, Walmart, and UPS.

Big-data backup challenges

Big data's climb to the top rung of the information ladder means it must be treated with care. Backup, once an afterthought for big-data volumes, is now essential and must be completed promptly and reliably. IT leaders are taking heed. In a recent ESG survey on the top five 2018 data center modernization priorities, 31% of respondents indicated that improving data backup and recovery was a high priority. Increasingly, backup and recovery resources will be dedicated to big data. IDC predicts the big-data market will be worth $210 billion by 2020, up from 150.8 billion in 2017.[2]

To back up data effectively, many businesses will have to make significant operational changes. Legacy architectures that were sufficient for earlier-generation workloads must be revamped with modern architectures to protect environments running big data workloads such as Hadoop. For example:

- **Many organizations have implemented client-server backup architectures**, which work well for conventional workloads but were never intended to handle large quantities of big data or continuous analytical processing of unstructured data. This data, including its corresponding points in time, is every bit as critical as large quantities of transactional data contained in more traditional CRM or ERP systems.
- **Backing up multiple copies of the same data sets** squanders storage resources. An effective backup architecture for big-data environments should eliminate needlessly replicated data sets to conserve storage resources, enable faster processing of data, and reduce costs.
- **Data integrity is important for organizations** because they are making business decisions based on mission-critical data analytics. If data becomes corrupted because of human error, then inaccurate information will be replicated in typical Hadoop big-data environments.
- **With the need for daily analytical processes** in some organizations, service-level agreements (SLAs) demand that backup tasks be performed rapidly. In addition, regulatory compliance usually requires data to be retrieved promptly.
- **Outdated backup architectures** impede big-data traffic. For example, replicated blocks of data flowing from multiple data

1 DC FutureScape: Worldwide Big Data, Business Analytics, and Cognitive Software Predictions, International Data Corporation, 2016.

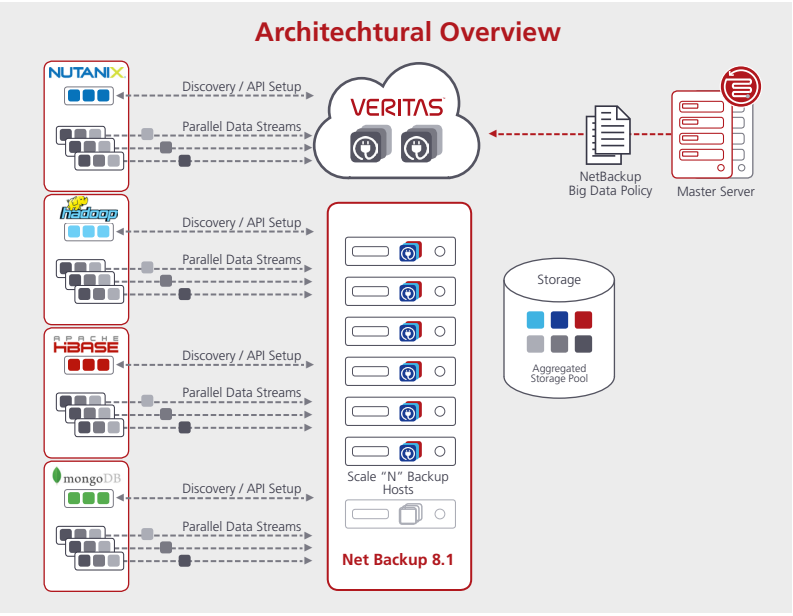
2 *Worldwide Semiannual Big Data and Analytics Spending Guide,* IDC, 2017.

VERITAS™
The truth in information.

IDG
STRATEGIC MARKETING SERVICES

nodes into the Hadoop NameNode create a bottleneck that slows big-data streams.

- **Backup for Hadoop environments** has led many organizations to create time-consuming, multistep workarounds that include the addition of secondary Hadoop clusters and extra storage. The complexity of these measures creates the risk of data loss at each step and the risk of missing the restore time objective (RTO).



The Veritas solution: NetBackup 8.1

Veritas NetBackup 8.1 is designed to meet the needs of big-data backup. It includes the

Veritas NetBackup Parallel Streaming Framework, which is designed for large, scale-out, backup, and multinode cluster workloads in Hadoop environments. By backing up the Hadoop Distributed File System (HDFS) natively, the agentless NetBackup 8.1 eliminates the need for complex workarounds. An agentless architecture is an important concept for modern environments. Because there is no agent footprint on the cluster nodes, you need not be concerned with managing each respective agent, saving time and money. And there is no need to worry about upgrading Hadoop.

Here's how it works: A downloadable plugin for NetBackup 8.1 scans the Hadoop cluster for metadata, then partitions and distributes the workload among backup hosts, pooling resources from NetBackup Media Servers or clients. Job resources are managed to maximize backup and recovery performance for parallel streaming as each Hadoop node sends data to the backup hosts. Big-data policies are fully integrated into the NetBackup user interface. By using plug-ins, you don't have to wait until the next Hadoop release.

The Veritas solution is unlike the Hadoop big-data architecture, which replicates three copies of data from each node through the NameNode and sends all data through backup hosts to storage, resulting in unnecessary storage usage and higher storage costs. In contrast, NetBackup 8.1 captures and stores only unique data, reducing storage space usage by two-thirds and saving 67% in secondary storage costs. In addition, backup is performed three times faster.

Although Hadoop performs data replication, it does not provide point-in-time data protection, which protects against human errors and data corruption. NetBackup 8.1 fills this gap, performing point-in-time backup to preserve critical data and meet SLA and compliance mandates.

Because the Veritas solution relies on an agentless NetBackup plugin (rather than an agent that sits on the Hadoop cluster), the analytics performance of the cluster is not affected by backup and recovery processes. In addition, the management headaches associated with agents, such as installing, managing, and upgrading, are eliminated.

Scalability is important for big-data environments, and the Veritas solution can be augmented to meet the demands of expanding Hadoop data volumes. As nodes are added to the Hadoop cluster, additional NetBackup nodes likewise can be added.

In addition, NetBackup 8.1 supports multiple environments in addition to Hadoop, such as Nutanix hyperconverged infrastructure, Apache HBase, and MongoDB. NetBackup 8.1 supports 40 cloud environments through integrated cloud connectors.

Conclusion

Big-data analytics have emerged from the test lab to take their place in the enterprise mainstream, establishing a track record of strategic value at many leading corporations. With business success or failure hanging in the balance, organizations must quickly analyze large quantities of data in real time.

The infrastructure supporting big-data backup and recovery must adapt to this mission-critical role. The Veritas NetBackup Parallel Streaming Framework overcomes the obstacles to big-data backup inherent in big-data environments. Veritas NetBackup 8.1 delivers faster performance while reducing risk and consuming fewer storage resources, opening new horizons for digital enterprises in their quest to gain a strategic edge from big-data analytics.

For more information, please click here