

# DELL EMC ISILON BIG DATA STORAGE AND ANALYTICS SOLUTIONS

## Efficient, Flexible In-Place Hadoop Analytics

### ESSENTIALS

- Industry's first and only scale-out NAS solution with native Hadoop support
- Reduce costs and accelerate results with in-place data analytics
- Increase efficiency with > 80% storage utilization and data deduplication
- Support multiple Hadoop distributions simultaneously
- Gain added operational flexibility with multi-protocol support
- Fast, easy scalability to over 20 PB in a single Isilon cluster

### THE BIG DATA OPPORTUNITY

The rapid growth of unstructured data represents a significant challenge for many enterprises across a wide range of industries today. In an environment of constrained IT budgets, providing sufficient storage capacity and managing it effectively can be daunting. Many organizations are, however, realizing that Big Data is a valuable asset that can be leveraged to uncover new opportunities to accelerate their business and gain a competitive advantage. Central to realizing this opportunity is Hadoop, an innovative big data analytics engine designed specifically to analyze large-scale data sets of unstructured file data.

### CHALLENGES WITH TRADITIONAL HADOOP DEPLOYMENTS

Traditional deployments of Hadoop have largely been implemented on a dedicated infrastructure, not integrated with or connected to any other applications. In addition to requiring a separate capital investment and added management resources, this approach poses a number of other inefficiencies and risks:

- Inefficient direct attached storage (DAS) with poor utilization compounded by mirroring typically up to three times or more. In addition to unnecessary capital costs, this infrastructure adds to data center costs and is management-intensive.
- Manual ingest of large-scale data sets into a separate Hadoop environment is time and resource consuming. This can lead to significant delays in the time to benefit or insight from the Hadoop analytics effort.
- Data and analytics results cannot be accessed or shared with other enterprise applications due to the lack of industry-standard protocol support. Organizations are also limited to running only one specific version of a Hadoop distribution.
- Single-point-of-failure risk with the Hadoop NameNode in traditional deployments.
- Lack of enterprise-level data protection options including snapshot backup and recovery and data replication capabilities for disaster recovery purposes.

## THE DELL EMC ISILON APPROACH

DELL EMC® Isilon® Big Data storage and analytics solutions combine a powerful yet simple, highly efficient, and massively scalable storage platform with integrated support for Hadoop analytics. DELL EMC Isilon is the first and only scale-out NAS platform to incorporate native support for the Hadoop Distributed File System (HDFS) layer. With your unstructured data on Isilon, these capabilities allow you to quickly implement an in-place data analytics approach and avoid unnecessary capital expenditures, increased operational costs, and time-consuming replication of your Big Data to a separate infrastructure. Simply connect your analytics compute resources to your Isilon storage system and you're ready to begin your analytics projects immediately.

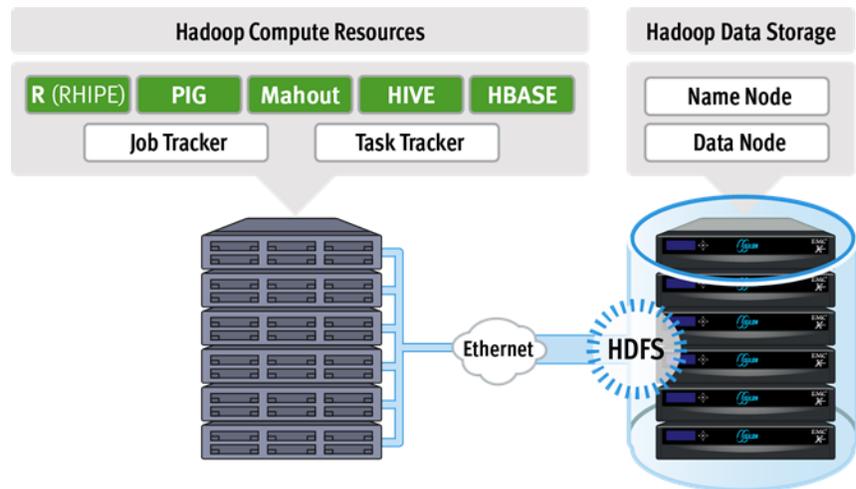


Figure: In-Place Analytics with DELL EMC Isilon Scale-Out NAS

## REDUCE COSTS

With DELL EMC Isilon scale-out network-attached storage (NAS), you can streamline your storage infrastructure by consolidating large-scale file and unstructured data assets, eliminating silos of storage. Isilon solutions also allow you to achieve a storage utilization rate of over 80 percent. And with Isilon SmartDedupe™ data deduplication, you can further reduce your Big Data storage requirements by up to 35 percent by eliminating redundant data. This unmatched efficiency means that with Isilon, you need less storage capacity and physical space to house the same amount of data than with other alternatives—reducing both initial capital outlay and ongoing costs. And because an Isilon solution is simple to manage, it requires fewer IT resources for storage administration than traditional storage systems. This further reduces overall operating costs.

By leveraging Isilon's native HDFS support and in-place Big Data analytics capabilities, you can also avoid significant capital expenditures required for a separate, dedicated Hadoop infrastructure along with the associated operational costs.

## ACCELERATE RESULTS

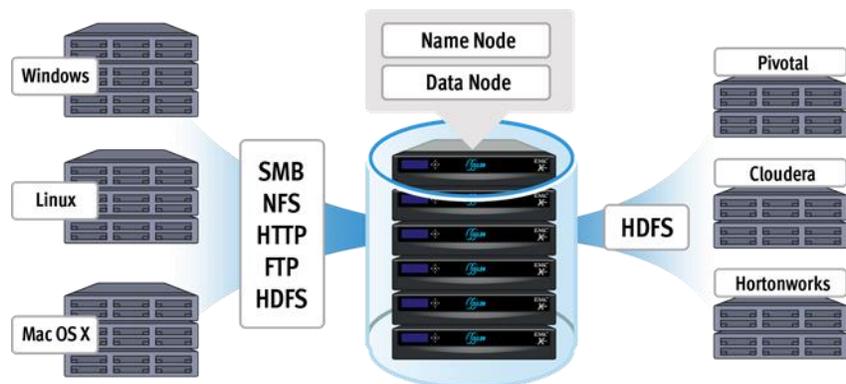
Isilon's in-place data analytics approach allows you to eliminate the time and resources required to replicate your Big Data into a separate infrastructure. For example, it can take over 24 hours to copy 100 TB of data over a 10Gb line. Instead, with Isilon, you can initiate data analytics projects immediately and get results in a matter of minutes. The ability to carry out in-place analytics without the lengthy data ingest phase enables support for analytics at the speed of business.

## INCREASE FLEXIBILITY

By treating HDFS as an over the wire protocol, you can quickly deploy a comprehensive big data analytics solution that can combine multiple industry-standard Hadoop distributions with Isilon scale-out NAS storage to provide a highly flexible Big Data storage and analytics ecosystem.

Isilon storage and analytics solutions support multiple instances of Apache Hadoop distributions from different vendors simultaneously including Pivotal HD, Cloudera CHD and Hortonworks Data Platform. This allows you to leverage the specific tools you need for each of your unstructured data analytics projects. You're also not locked-in to using any particular vendor's Hadoop distribution or version. These Hadoop distributions can also be combined with Isilon scale-out NAS as the basis for implementing a Hadoop-as-a-Service offering.

In addition to native HDFS 1.0 and HDFS 2.0 support, Isilon solutions include integrated support for a wide range of industry-standard protocols, including NFS, SMB, HTTP, FTP, and REST-based Object access for your cloud initiatives. As a result, with Isilon, you can provide a highly flexible storage infrastructure that allows you to simplify workflows, accelerate business analytics projects, support cloud infrastructure initiatives, and get more value from your enterprise applications and data.



**Figure: Simultaneous Support for Multiple Hadoop Distributions and NAS Protocols**

## GAIN MASSIVE SCALABILITY

With DELL EMC Isilon scale-out storage solutions, you can have massive room for growth for your unstructured data assets and related analytics projects. Isilon solutions scale easily from 18 terabytes (TB) to over 20 petabytes (PB) of capacity in a single Isilon cluster. The DELL EMC Isilon OneFS® operating system allows a storage system to grow symmetrically or independently as more space or processing power is required. This provides a true grow-as-you-go approach and the ability to scale out as business needs dictate. With Isilon, you can scale capacity and performance. The DELL EMC Isilon AutoBalance™ function allows you to quickly add nodes to your Isilon cluster as needed without downtime, manual data migration, or application logic reconfiguration, thereby saving IT management resources and avoiding operational interruptions.

## PROTECT AND SECURE YOUR BIG DATA ASSETS

DELL EMC Isilon storage solutions provide unsurpassed levels of data protection and availability for unstructured data. With Isilon OneFS, our solutions also eliminate the “single-point-of-failure” risk associated with traditional Hadoop deployments. We do this by enabling all nodes in a DELL EMC Isilon storage cluster to become, in effect, NameNodes. This greatly improves the resiliency of your Hadoop environment.

The DELL EMC solution for Hadoop also offers end-to-end data protection options for your unstructured data. For fast and efficient data backup and recovery, you can schedule snapshots as frequently as needed to meet your specific recovery-point objectives. For reliable disaster recovery protection, our storage solutions provide fast data replication, along with push-button failover and failback simplicity, to further increase the availability of your data assets.

## TAKE THE NEXT STEP

Contact your DELL EMC sales representative or authorized reseller to learn more about how the DELL EMC Isilon archive solutions can benefit your organization.

Also see our solutions in the DELL EMC Store at <https://store.emc.com/isilon>.

