Activate the Value of Big Data with SAS® and Intel: Storage Cost Reduction with Hadoop

The challenge of storing big data

The more data you capture for your business, the more potential value awaits you – if you can store and access the data effectively. To store, transform, analyze and understand all that information, you need a strong storage solution for fast, in-depth analytics.

Building on more than 20 years of engineering collaboration and problem-solving, SAS and Intel have the answer. SAS world-class analytic capabilities and big data technologies using the Intel® Xeon® processor E7 v2 family and Intel® SSDs provide integrated, scalable enterprise solutions that help you extract maximum value from your data while helping you manage data center space, cooling and power costs.

The benefits of Hadoop processing with SAS® and Intel

Many organizations are moving to Hadoop for data storage and processing – and for good reason. Distributed solutions using Hadoop can more efficiently store, process and manage large volumes of data while running on industry-standard servers. These solutions are easily scalable and highly available, with data stored redundantly across all cluster nodes.

By combining SAS software, Intel platforms and Hadoop, you can have a faster, more efficient and less costly data storage and processing infrastructure.

Storage cost reduction with Hadoop: A SAS® and Intel architecture for big data analytics

Storing your SAS data sets on SAN can be expensive to implement, operate and scale up, especially as your data needs increase. The SAS storage cost reduction architecture replaces SAN storage with Hadoop cluster storage. It then leverages this data by connecting a SAS grid to a Hadoop cluster, both of which are built on platforms using the Intel Xeon processor E7 v2 family. This architecture requires you to maintain only a limited amount of SAN to provide temporary space for data in the SAS grid. This architecture:

- Minimizes disruption to the SAS user community while delivering major cost reductions.
- Accommodates very wide data files, unlike traditional data management solutions.
- Offers highly cost-effective data resilience.
- Provides scalability to accommodate your data growth needs.
- Improves data governance, as all data resides in a common store.

With the SAS storage cost reduction architecture, you can store vast amounts of data on industry-standard servers while making that data accessible to a SAS grid for processing. This configuration offers the same SAS processing capabilities as those of a SAS grid and SAN solution, while providing substantial cost savings.

By combining SAS grid, Intel platforms and Hadoop, you achieve a highly scalable, highly available, lower cost storage infrastructure while maintaining your current analytical process. This architecture allows you to utilize your resources more efficiently, save money for your business and improve your data management strategy.

The power you need for strong analytics

SAS applications provide an integrated environment to solve a variety of complex business problems, so it's no surprise that they require strong computing performance and reliability. Fortunately, they get all that and more with platforms using the Intel Xeon processor E7 v2 family and Intel SSDs with NVMe.

The Intel Xeon processor E7-4890 v2 brings multiple advantages to analytics, data management and business intelligence. This processor can rapidly process large amounts of data in real time, keeping data in memory to overcome data-demanding workloads. Key gains over previous-generation technology include:

- **Higher core count.** Use fewer servers for the same work with 60 cores per server.

- **Increased memory capacity and I/O.** With a capacity of up to 6TB of memory, now solve very large problems through in-memory processing.

- **Increased computational capabilities.** Get both data fidelity and fast analysis.

- **Highly reliable.** Count on world-class uptime with a platform designed for five nines (99.999 percent) of reliability.

Intel Xeon processors E7 v2 work seamlessly with Intel SSDs with NVMe, which provide a fast PCI Express interface improving on previous-generation SATA drives. The Intel SSD DC P3700 Series with NVMe brings high storage performance directly to a server's processors. The Intel SSD DC S3700 Series also combines fast, consistent performance with high endurance and strong data protection.

Systems using the Intel Xeon processor E7-4890 v2 and Intel SSDs provide the flexibility, reliability, memory, storage speed and overall performance you need to enable big data analytics. By using the SAS storage cost reduction solution to replace SAS grid SAN storage with Hadoop, your big data system can yield valuable business insights and help improve your decision-making.
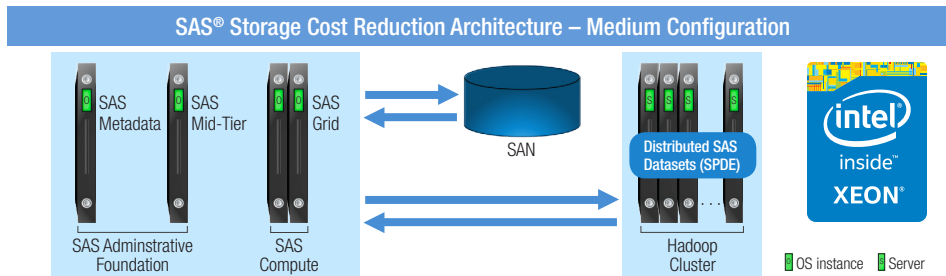
About SAS and Intel

SAS is the leader in business analytics software and services, and the largest independent vendor in the business intelligence market. Through innovative solutions, SAS helps customers at more than 70,000 sites improve performance and deliver value by making better decisions faster. Since 1976 SAS has been giving customers around the world THE POWER TO KNOW®.

Intel (NASDAQ: INTC) is a world leader in computing innovation. The company designs and builds the essential technologies that serve as the foundation for the world's computing devices. Additional information about Intel is available at intel.com.

Learn more

To learn more about how SAS and Intel can help you maximize the value of your data, contact your SAS or Intel representative or visit sas.com/intel.

SAS® Storage Cost Reduction Architecture – Medium Configuration

SAS Storage Cost Reduction Architecture			
SAS Administrative Foundation		**SAS Compute**	**Hadoop Cluster**
SAS Metadata	SAS Mid-Tier	SAS Grid	Cloudera or Hortonworks
RHEL 6.5	RHEL 6.5	RHEL 6.5	RHEL 6.5
1 OS instance (min) (Physical or Virtualized)	1 OS instance (min) (Physical or Virtualized)	2 OS instances (Physical or Virtualized)	4 servers (min)
2 sockets/server (min) Intel® Xeon® E5-2640 v3 (16 cores)	**2 sockets/server** (min) Intel Xeon E5-2640 v3 (16 cores)	**4 sockets/server** Intel® Xeon® E7-4890 v2 (60 cores)	Intel® Xeon® E5-2690 v3 (24 cores)
Starting at 128GB per OS instance	Starting at 128GB per OS instance	Starting at 256GB per OS instance	256GB/server (min)
8+ active cores per OS instance	8+ active cores per OS instance	Up to 60 active cores per OS instance	
Multiple OS instances may share a single physical machine to best leverage resources. Intel recommends 2 Boot Intel SSD DC S3700 Series - 200GB per OS instance running SAS.			
SAS Storage Cost Reduction Architecture – Storage and Network			**Hadoop Cluster**
Starts at 800GB Intel® SSD DC S3700 per OS instance	Starts at 800GB Intel SSD DC S3700 per OS instance	Starts at 4 Intel® P3700 1.6TB SSD (PCIe) per OS instance - reference SAS I/O and file system requirements	Starts at 4 Intel SSD DC S3700 (800GB each)
Total Intel SSD DC P3700 Series 1.6TB: 8+, Total Intel SSD DC S3700 Series - 800GB: 6+, Network recommendation: 10Gb Ethernet minimum.			
Software			
SAS Configuration	**Cloudera Hadoop**	**Hortonworks Hadoop**	**Third Party**
SAS Grid Manager, SAS Data Management, SAS Enterprise Model Management, SAS/ACCESS® Interface to Hadoop, SAS/ACCESS Interface to Impala, SAS Office Analytics	Impala, Core Components, Cloudera Manager	Ambari, HDFS, MapReduce, Nagios, Ganglia, Hive	MS Office 2013, PuTTy
The sample architecture referenced in the table above is for informational purposes only, is not a performance benchmark, and does not constitute a performance guarantee. SAS and Intel recommend that customers engage in a formal sizing exercise before acquiring hardware.			