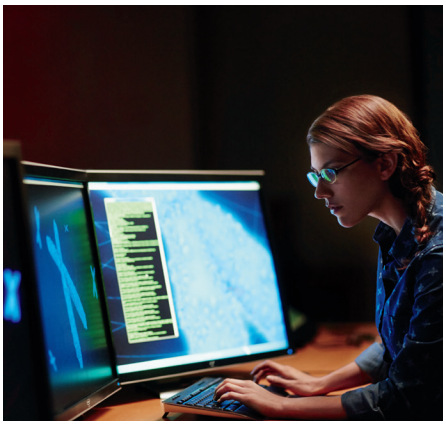
CASE STUDY

Intel® Network Platforms Group Big Data Analytics

Business Intelligence

Software-Defined Storage Enables Agility and Faster Analytics

Intel and Coho Data developed a data-centric infrastructure that combines software-defined storage and big data analytics.



UBS AG, a Fortune 100 financial institution, strengthens its competitive stance by leveraging business intelligence (BI) derived from increasingly sophisticated analytics. Its IT group sought to maintain the functional benefits of a mature enterprise storage system, while incorporating BI tools into the system to perform analytics tasks efficiently on the enterprise data in situ. Through convergence of the traditionally separate enterprise storage and BI environments, a new architecture established a stronger, more resilient foundation for enterprise-scale, big data operations and real-time analytics. Coho Data collaborated with Intel on this reference architecture and built a flexible, scalable solution using a software-defined storage model and a data-centric infrastructure.

Challenges

- **Analytics sprawl**. Uncoordinated, disconnected analytics environments—beyond IT control—can cause inefficiencies and threaten business continuity.
- **Excess expenditures**. Projections on storage needs for an analytics project can be difficult, leading to potential purchases of hardware well beyond required capacity, resulting in needless expenditures.
- **Inadequate management** and data services. Today's analytics environments have not kept pace with storage and virtualization technologies, leading to a lack of centralized management and inadequate data services.

Solution

- **Scale-out enterprise storage**. Coho Data storage services, Cloudera CHD5* data analytics (containarized with Docker*), and a foundation built on Intel's reference architecture establish a data-centric infrastructure optimized for business intelligence.
- **High-performance data-centric enterprise infrastructure**. Validated software ingredients and hardware components from third-party vendors round out the architecture to meet the needs of enterprises developing big data solutions.

Results

- **Improved analytics** and near real-time BI. The data-centric infrastructure enhances the capabilities to scale storage requirements fluidly and handle analytics operations more efficiently. The development of a single shared environment has opened opportunities for innovative data uses throughout the company and made it easier to comply with regulatory mandates that apply to the financial services industry.

"By 2016, IDC predicts that hyperscale data centers will house more than 50 percent of raw compute capacity and 70 percent of raw storage capacity worldwide, becoming the primary consumer/adopters of new compute and storage technologies."
— Storage Newsletter.com

Software-Defined Storage Adds Flexibility and Scalability to Big Data Environments

"In the third calendar quarter of 2014. . .up 39.4 percent year over year, the software-defined storage platforms market is benefiting from an increased desire to control costs by utilizing commodity hardware when building storage systems."

– Jingwen Li, Research Analyst, IDC

Maximizing Business Data Value

Business data—in growing volumes—has escalated the need to improve data availability and storage capacity for large-scale analytics. Organizations rethinking the infrastructures of IT services often favor solutions enabling improved BI—such as those provided by Apache Hadoop* and Spark*. Building data-centric infrastructures, however, must typically be accomplished without compromising existing traditional compute and storage resources.

Within this context, enterprise IT organizations, cloud hosting providers, and cloud service providers interested in BI solutions are evaluating approaches that take advantage of emerging technologies. These technologies include software-defined networking (SDN) and software-defined storage (SDS) for their flexibility, scalability, centralized management, and efficient use of resources.

The reference architecture developed is well suited to the requirements of data-centric operations, massive-volume data analytics, and BI applications. The data-centric storage platform integrates hosted BI capabilities within a scale-out enterprise storage system. Using Coho Data's DataStream* storage modules, the deployment solved a number of problems associated with enterprise BI deployments and used parallel computing processes for data-intensive analytical operations.

Technical Challenges in Building Enterprise Big Data Solutions

Many enterprises recognize the value in big data analytics but find their current infrastructure inadequate to working with big data workloads. Two key technical challenges stand in the way:

- **The mismatch between big data workloads and the traditional enterprise storage architecture**. Cloudera advises creating a data hub or big data lake—in silos away from other enterprise storage—for BI/analytics environments (see Figure 1). Duplication of data and inefficient data copying across environments raise costs and increase complexity.

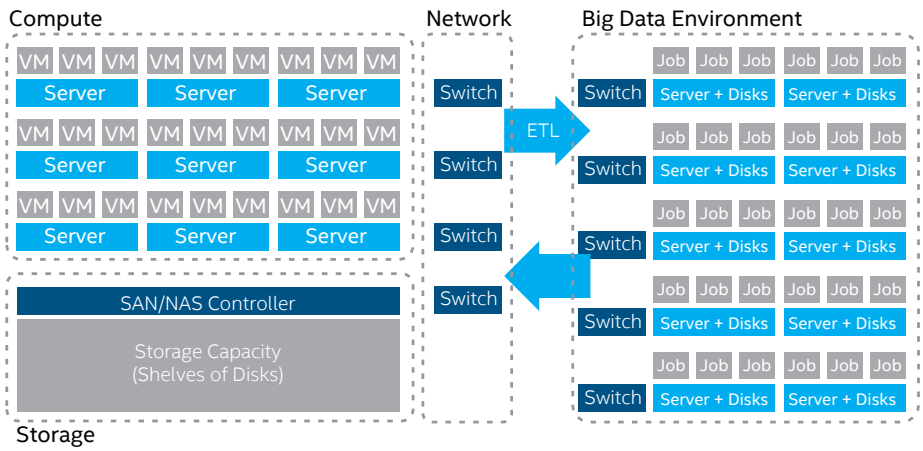


Figure 1. Mismatch between big data requirements and traditional enterprise storage creates inefficiency.

- **Meeting IT Data Management Responsibilities.** Key data management responsibilities for IT groups include ensuring availability, managing archiving, retaining data for compliance, and planning disaster recovery. All of these tasks become more challenging with the proliferation of bespoke Hadoop clusters and multiple data copies circulating to and from IT storage repositories.

Coho Data meets this challenge by unifying enterprise storage and hosting BI apps in an architecture that scales compute and connectivity in balanced proportion to storage capacity and performance.

Other factors that needed to be addressed during this collaboration:

Pooled and shared resources: Using commodity hardware, the solution needed to support resource sharing among constituents.

Hadoop optimization: To meet service-level agreements and multi-tenant requirements, Hadoop had to be configured and optimized for varying workloads and reliable performance.

Strong support for future big data implementations: The solution had to provide flexibility and a solid foundation for future big data applications.

Solution Components

The solution includes these components (see Figure 2):

Cloudera CDH5 (deployed using Docker container technology) provides an enterprise-capable implementation of Hadoop and big data analytics.

Nuage Networks Virtualized Services Platform delivers automated, policy-driven SDN functionality.

Intel® Network Interface Cards, Intel® Solid-State Drives, and Intel® processor-based server hardware establish high-speed, network communications, responsive storage access, and exceptional processor performance.

Arista Software-Driven Cloud Networking ensures rapid provisioning and control of cloud services, network-wide virtualization, and automated management of network applications.

Coho Data DataStream nodes provide an underlying high-performance scale-out storage layer and transparently integrate the other components in the architecture. The DataStream nodes also provide data-centric container hosting directly within the storage system and SDN integration for dynamic horizontal scalability.

Participation by Intel® Network Builders helped organize partners and coordinate the engagement.

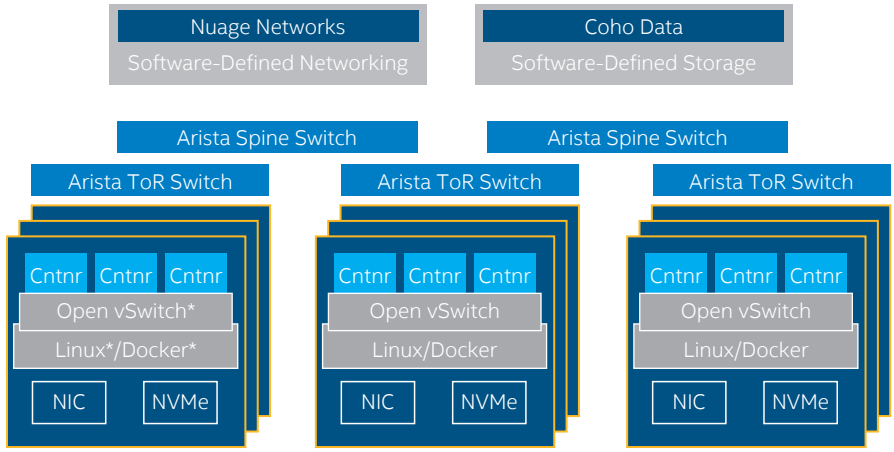


Figure 2. Solution components and shared storage resources.

What is a Data-Centric Infrastructure?

A data-centric infrastructure is characterized by low-cost tiered storage, fluid elasticity of the storage resources, easy access to rich data services, a complete database platform, and enterprise-caliber responsiveness. In contrast, an application-centric infrastructure is optimized for specific applications, often in a rigid, inflexible way.

Data-centric infrastructures are frequently built around an Apache Hadoop* framework. This form of infrastructure effectively supports complex analytical operations common to big data use cases. The infrastructure delivers rapid access to data resources, whether located on-premises or in the cloud, in a unified way, typically within a single address space.

Building a Private Cloud Infrastructure

"Enterprise IT organizations can learn from the web-scale firms as they build out their private cloud infrastructure. They need the performance of the latest flash technology, but more crucially, they need a flexible scale-out model that uses commodity hardware and simpler management to build out their storage infrastructure to accommodate many different application workloads. The Coho Data team has built a solution with high-performance scale-out capability combined with cost-efficient hardware that challenges the price-performance of both traditional storage and many of the flash arrays in the market. They've created a new operational model for scaling storage more efficiently in the enterprise."

— Ashish Nadkarni, Research Director - Storage, IDC

Test Results and Outcome

Relying on clusters that combined Intel® Solid-State Drives (Intel® SSDs) and Coho Data's software-defined storage (SDS) approach (with approximately 200 terabytes raw capacity per cluster), data access improvements were substantial. Other unexpected improvements were also observed: the SDS reference architecture proved extremely capable at supporting the addition of data analytics tools with ease and efficiency. This characteristic suggests that the architecture offers substantial business value for creating environments optimized for big data applications.

For more extensive information on the test results and the SDS reference architecture, refer to the whitepaper titled "Converging Enterprise Storage and Business Intelligence: A Reference Architecture for Data-Centric Infrastructure," available at: **https://networkbuilders.intel.com/resources**.

Confirming Flash-Based Storage Performance Benefits for Analytics

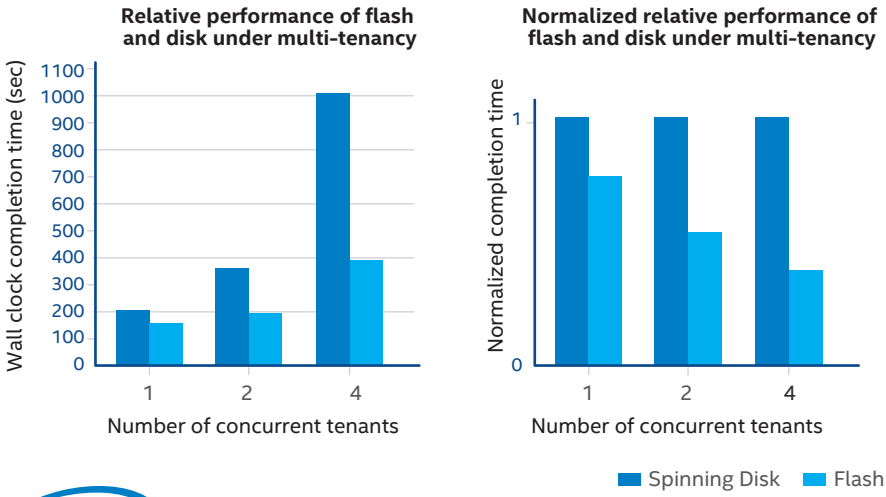
In testing performed in a comparable configuration—but not set up with the Coho Data storage components—Intel staff members tested standard spinning hard disk drives against Intel SSDs in a multi-tenant environment. The results, as shown in Figure 3, indicated that individual jobs took up to 2.5 times longer to complete on spinning drives than on the flash-based system. For routine computational tasks with less I/O involved, the dfference was still apparent, but less striking.

More Control with Software-Defined Storage

Centralized control is at the heart of SDS, providing a mechanism to support storage needs dynamically across a multi-tenant cloud environment, and, in this case, to accommodate the big data management requirements of a major financial institution.

Following the deployment, the SDS reference architecture was praised for improving the capabilities to take action on business data. Observers noted that the integration with the network file system, rapid provisioning, and containerized approach to analytics application use streamlined the process of performing tasks on primary data and lowered the barriers to building out big data clusters.

This collaborative engagement demonstrated that BI derived from big data can be successfully managed using SDN and SDS technology.



For more information on Intel® Network Builders, visit **networkbuilders.intel.com**.

Intel technologies' features and benefits depend on system configuration and may require enabled hardware, software or service activation. Performance varies depending on system configuration. Check with your system manufacturer or retailer or learn more at [intel.com].

INFORMATION IN THIS DOCUMENT IS PROVIDED IN CONNECTION WITH INTEL' PRODUCTS. NO LICENSE, EXPRESS OR IMPLIED, BY ESTOPPEL OR OTHERWISE, TO ANY INTELLECTUAL PROPERTY RIGHTS IS GRANTED BY THIS DOCUMENT. EXCEPT AS PROVIDED IN INTEL'S TERMS AND CONDITIONS OF SALE FOR SUCH PRODUCTS, INTEL ASSUMES NO LIABILITY WHATSOEVER, AND INTEL DISCLAIMS ANY EXPRESS OR IMPLIED WARRANTY, RELATING TO SALE AND/OR USE OF INTEL PRODUCTS INCLUDING LIABILITY OR WARRANTIES RELATING TO FITNESS FOR A PARTICULAR PURPOSE, MERCHANTABILITY, OR INFRINGEMENT OF ANY PATENT, COPYRIGHT OR OTHER INTELLECTUAL PROPERTY RIGHT. UNLESS OTHERWISE AGREED IN WRITING BY INTEL, THE INTEL PRODUCTS ARE NOT DESIGNED NOR INTENDED FOR ANY APPLICATION IN WHICH THE FAILURE OF THE INTEL PRODUCT COULD CREATE A SITUATION WHERE PERSONAL INJURY OR DEATH MAY OCCUR.

Intel may make changes to specifications and product descriptions at any time, without notice. Designers must not rely on the absence or characteristics of any features or instructions marked "reserved" or "undefined." Intel reserves these for future definition and shall have no responsibility whatsoever for conflicts or incompatibilities arising from future changes to them. The information here is subject to change without notice. Do not finalize a design with this information.

The products described in this document may contain design defects or errors known as errata which may cause the product to deviate from published specifications. Current characterized errata are available on request. Contact your local Intel sales office or your distributor to obtain the latest specifications and before placing your product order. Copies of documents which have an order number and are referenced in this document, or other Intel literature, may be obtained by calling 1-800-548-4725, or by visiting Intel's Web site at www.intel.com.

Copyright © 2015 Intel Corporation. All rights reserved. Intel, the Intel logo, Experience What's Inside, and the Intel Experience What's Inside logo are trademarks of Intel Corporation in the U.S. and/or other countries.

*Other names and brands may be claimed as the property of others.
Printed in USA 0515/JL/MESH/PDF ♻ Please Recycle 332532-001US