



Microsoft

CLUDERA

Solution Brief

## Extend data insights across the enterprise with agility and flexibility

Cloudera Data Platform Private Cloud Base, validated on Microsoft Azure Stack HCI, brings a robust, scalable data platform to on-premises environments.

## Start with the right foundation for your data

It's not easy running an enterprise, and even more so if you're making IT decisions. The sheer amount of customer and operations data that enterprises must manage, store, and analyze can be intimidating. This big data comes in all shapes and forms—such as images, videos, telemetry, and Internet of Things (IoT) data. Often, these types of data are siloed in different databases and platforms, making it difficult for enterprises to analyze them comprehensively for insights. In addition, security requirements in some industries require some workloads and data to be siloed and managed on premises. Managing security in this environment can be problematic, with multiple security models applied across applications and clouds.

Cloudera Data Platform (CDP) Private Cloud Base can help enterprises deal with these challenges as a big-data platform that addresses both IT and business needs. CDP Private Cloud Base integrates big-data management and analytic experiences across the data lifecycle. It is simple to use, secure by design, and open and extensible; best of all, it can run on any infrastructure with any analytics and any data.

CDP Private Cloud Base is now available on Microsoft Azure Stack HCI, bringing together the best of both worlds for organizations that want more from their data but that are restricted to on-premises deployments due to regulations or security and privacy requirements. Azure Stack HCI is specifically engineered to help meet the needs of the modern datacenter by delivering a comprehensive Azure cloud experience on premises using an enterprise's own servers, and that is compatible with most preferred tools in IT environments.

## Modernize to a big-data platform within your own datacenter

Deliver end-to-end analytics and transactional and predictive workloads running on traditional clusters in your own datacenter. Power the data lifecycle from the edge to artificial intelligence (AI) with CDP Private Cloud Base, which helps provide:

- **Faster performance:** Deliver accelerated time to value by uncovering business insights with faster analytics and data management.
- **Improved cluster management:** Reduce platform operational costs and extend the life of hardware assets.
- **New powerful capabilities:** Create new avenues for generating value on data clusters with real-time data warehousing, stream processing, transactional operational database, graphics processing unit (GPU)-boosted data engineering and machine learning (ML), and much more.
- **A future-proof platform:** Manage data with object storage and vastly improved storage density and scalability on a foundation for cloud-native data services.

## CDP Private Cloud Base value

**2x**

**faster analytics  
to accelerate time  
to value**

Significant increases achieved from integrating the latest versions of Apache Impala, Apache Hive, and Apache Spark<sup>1</sup>

**20%**

**higher resource  
utilization  
to optimize costs**

Enhanced cluster management and resource scheduling to reduce overhead and boost utilization<sup>2</sup>

**50%**

**improved density  
for more  
cost-effective storage**

Increased storage density from HDFS erasure coding and augmented with Apache Ozone object storage<sup>2</sup>

**2x**

**better Infosec  
compliance  
reduces risk**

Dramatic CVE reduction and better encryption of user data through FIPS compliant modules<sup>2</sup>

**50%**

**stronger  
governance to ease  
regulatory compliance**

New SDX controls: deny by default, least privilege, policy tags, scalable audit, consistent enforcement<sup>2</sup>

CDP Private Cloud Base comprises a variety of components such as Apache Hadoop File System (HDFS), ApacheHive 3, Apache HBase, and Apache Impala, along with many other components for specialized workloads. All of these features in CDP work seamlessly in an Azure Stack HCI environment.

## Benefits of Azure Stack HCI

Azure Stack HCI allows enterprises to run scalable and secure virtualization on premises while taking advantage of connected and familiar Azure services for monitoring, scaling, backup, and more. Windows and Linux virtual machines (VMs) can run on premises on a platform that's managed with existing tools, processes, and skillsets, all delivered as an Azure subscription service.

Azure Stack HCI uses software-defined storage, networking technologies, and virtualization based on the Microsoft Hyper-V hypervisor. Azure Stack HCI is Azure Arc-enabled by design; therefore, one way to manage it is through the Azure portal.

By extending on-premises infrastructure to a hybrid environment, enterprises that deploy CDP Private Cloud Base on Azure Stack HCI can perform:

- Offsite backup and disaster recovery
- Centralized update management
- VM management from a single plane (the Azure portal)
- Cloud-based monitoring
- Security and advanced threat protection
- Centralized governance
- Cloud-based updates

## Streamline data pipelines at any scale with underlying security

CDP Private Cloud Base is an on-premises version of CDP that combines Cloudera Enterprise Data Hub and Hortonworks Data Platform Enterprise, in addition to new features and enhancements across the stack. This unified distribution is a scalable and customizable platform on which many types of workloads can be securely run.

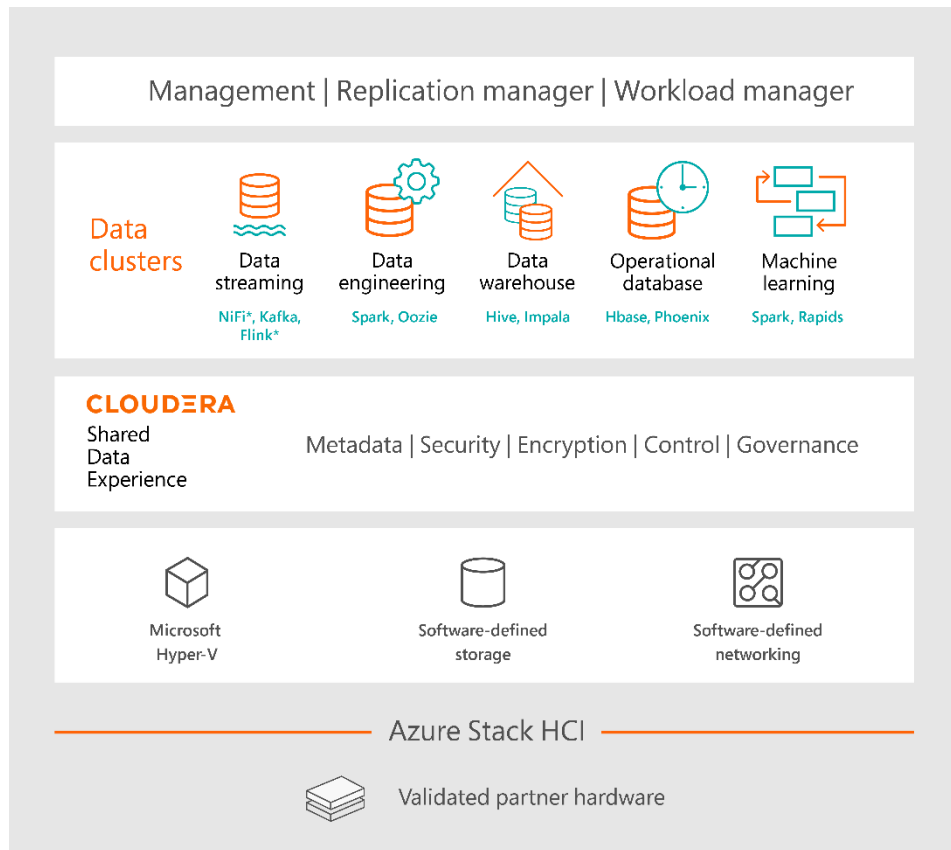


Figure 1. CDP on Azure Stack HCI

## Support security and governance for the entire data lifecycle

CDP Private Cloud Base helps ensure control and governance with built-in security for the entire data and analytics lifecycle. CDP Private Cloud Base features an integrated policy engine, Cloudera Shared Data Experience (SDX), to deploy security rules, authorizations, and policies at every step of the data lifecycle, while scaling to thousands of users without the burden of administration. Cloudera SDX manages user access, monitors activity, and audits usage of shared data across every workspace for consistent governance. To meet the compliance needs of highly regulated industries, CDP Private Cloud Base includes end-to-end governance of sensitive data, and it can extend to all clouds.

Enterprises gain additional security benefits when using CDP Private Cloud Base with Azure Stack HCI, which, supports new security capabilities on secured-core servers and secure network connectivity.



## Secured-core servers

Secured-core servers bridge hardware and firmware security for simplified, built-in security and preventative defenses. Combined with advanced Windows Server security features, Azure Stack HCI enhances the security features of CDP Private Cloud Base with:

- **Trusted Platform Module 2.0 (TPM 2.0):** TPM 2.0 secure crypto-processor chips provide a secure, hardware-based store for sensitive cryptographic keys and data, including system-integrity measurements.
- **Secure boot:** The secure boot security standard helps make sure that a device boots using only software that is trusted by the original equipment manufacturer (OEM).
- **Virtualization-based security (VBS):** VBS uses hardware-virtualization features to create and isolate a secure region of memory from the normal operating system, protecting an entire class of vulnerabilities against cryptocurrency mining attacks.
- **Hypervisor-protected code integrity (HVCI):** HVCI uses VBS to significantly strengthen code-integrity policy enforcement, including kernel-mode integrity.
- **Windows Defender System Guard Secure Launch:** Windows Defender System Guard Secure Launch is the first line of defense against exploits that try to take advantage of early-boot flaws or bugs.
- **Kernel Direct Memory Access (DMA) Protection:** The operating system and the system firmware work together to protect the system against malicious and unintended DMA attacks for all DMA-capable devices during the boot process.

## Secure connectivity updates

Azure Stack HCI also provides CDP Private Cloud Base with the following features for secure connectivity:

- **HTTPS and Transport Layer Security (TLS) 1.3 enabled by default:** HTTPS and TLS 1.3 protect the data of clients connecting to the server.
- **Encrypted DNS name-resolution requests with DNS-over-HTTPS:** The DNS client in Azure Stack HCI supports DNS-over-HTTPS (DoH), which encrypts DNS queries using the HTTPS protocol.
- **Server Message Block (SMB) AES-256 encryption:** Azure Stack HCI supports AES-256-GCM and AES-256-CCM cryptographic suites for SMB encryption. Windows automatically negotiates this more advanced cipher method when connecting to another computer that also supports it.
- **East-west SMB encryption controls for internal cluster communications:** When using Storage Spaces Direct, you can decide to encrypt or sign east-west communications within the cluster for higher security.
- **SMB Direct and remote direct memory access (RDMA) encryption:** Data is encrypted before placement, leading to far less performance degradation while adding AES-128 and AES-256 protected packet privacy.
- **SMB over QUIC:** Users and applications can securely and reliably access data from edge file servers. Mobile and telecommuter users no longer need a virtual private network (VPN) to access their file servers over SMB when using Windows.

## Use cases for CDP Private Cloud Base on Azure Stack HCI

CDP Private Cloud Base delivers powerful analytic, transactional, and machine learning workloads to deliver faster business insights on premises, with the control of the datacenter. CDP Private Cloud Base modernizes traditional monolithic cluster deployments into a powerful and efficient platform that enables many use cases, including:

- Integrated internal and external data as a foundation for data governance and models
- Infrastructure modernization for big data
- Fraud detection
- Improved customer recommendations/advice
- Predictive analytics
- Predictive insurability
- Accelerated product development/time to market
- Streamlined research and development
- Compliance with privacy and security regulations

### Example use case: Telecommunications

Telecommunications service providers are always looking for ways to increase their competitive advantage in the marketplace. With the race to expand its 5G user base, some telcos are analyzing their data for innovative 5G services to offer customers. To do so, they need a big-data system that can accommodate data not only from LTE/4G and 5G but also from wired services.

With CDP Private Cloud Base and Azure Stack HCI, telcos can build a network real-time analytic platform (NRAP), collecting large sums of data in real time and monitoring data by the minute. The NRAP can differentiate functions to lead digital transformation for 5G and provide distinguished services by organizing functions to a new efficiently structured architecture.

With this architecture, telcos could expand 5G services to millions more users. Additionally, by making use of data-driven decision making, it is possible to improve customer services overall by subdividing performance defects into customer, network, and data service units. These units can be addressed through AI-based tasks, enabling proactive quality responses to service problems.

# Accelerate data-driven decisions with CDP Private Cloud Base and Azure Stack HCI

CDP Private Cloud Base and Azure Stack HCI offer a collaborative solution that delivers high performance and fast insights from all enterprise data, on premises and with a single pane of glass. Organizations can innovate and improve services based on more efficient, data-driven decisions on a hybrid platform that evolves with their changing needs.

**Get started with a free trial of CDP Private Cloud Base at**

[www.cloudera.com/downloads/cdp-private-cloud-trial.html](http://www.cloudera.com/downloads/cdp-private-cloud-trial.html).

**Learn how Azure Stack HCI** can help you modernize infrastructure, consolidate virtualized workloads, and increase efficiency at [www.azure.com/hci](http://www.azure.com/hci).



<sup>1</sup>Tim Armstrong, David Rorke, Shant Hovsepian, and Justin Hayes. "New Multithreading Model for Apache Impala," Cloudera. October 2020. <https://blog.cloudera.com/new-multithreading-model-for-apache-impala/>

<sup>2</sup>Based on internal Cloudera product benchmarking.

© 2022 Microsoft Corporation. All rights reserved. This document is provided "as-is." Information and views expressed in this document, including URL and other Internet Web site references, may change without notice. You bear the risk of using it.

This document does not provide you with any legal rights to any intellectual property in any Microsoft product. You may copy and use this document for your internal, reference purposes.