CISCO

The bridge to possible

Extract Intelligence From Your Data—Wherever It Resides

With Cisco Data Intelligence Platform Powered by Cloudera Software

Your business depends on how much intelligence you extract from your data.

Big data was only the beginning, and organizations realize that all of their data has value—including online transaction processing systems, data warehouses, the Internet of Things, and data for machine-learning model training. Your data can be structured, unstructured, semi-structured, and object-based, with large and small file sizes.

A transition has occurred from on-premises big-data repositories to data residing almost anywhere. In today's hybrid-cloud world, data resides in many locations—on premises and/or in public clouds. Organizations analyze data on premises and/or in the cloud with traditional analytics and AI techniques to extract even more intelligence.

Cisco Data Intelligence Platform with Cloudera sets you up for a more flexible, fluid world where you can gather and process data on premises, move it to where it is needed, and process it locally or in a public cloud as your business needs dictate.

Benefits

- **Extract** more business value from your data— wherever it resides

- **Store** and operate on your data wherever your business needs dictate

- **Trust** Cisco and Cloudera for validated designs that help speed time to value while helping to reduce risk

ıı ıı ıı
CISCO
The bridge to possible

Help extract value from data

The Cisco Data Intelligence Platform with Cloudera is designed to help you get the most out of your data, wherever it resides, and wherever you want to extract knowledge from it. The platform supports a wide range of data, including:

- Operational databases
- Data warehouses
- Internet of Things
- Machine-learning data

It combines data storage with compute farms so you can analyze with standard compute engines and even AI frameworks that include GPU acceleration. It is designed to support and accelerate the following use cases:

- **Hybrid workloads:** Run your workload on premises and/or in the cloud with equal access to your data. Burst into the cloud during peak hours or during seasonal or urgent demands.
- **Hybrid pipelines:** Implement, orchestrate, and optimize data pipelines for easier management. Implement secure data exchange between your on-premises data center and your choice of public cloud.
- **Hybrid data integration:** Integrate data sources from multiple clouds. Simplify application development and ML model training that needs on-premises data sources or cloud-native data stores.
- **Hybrid DevOps:** Support agile development by developing software in the cloud, then running production software with sensitive data on premises.

- **Cloud-native applications:** Build applications that run in any cloud so that you can optimize cost, performance, and data residency.

Architecture for the future

The Cisco Data Intelligence Platform combines Cisco Unified Computing System™ (Cisco UCS®) servers as your on-premises cloud, using the capabilities of the Cloudera Data Platform (CDP) to integrate your data into a hybrid cloud data lake that is accessible from anywhere you wish to analyze it (Figure 1).

The platform brings together some of the largest open-source initiatives with Apache Ozone, Apache Hadoop, Kubernetes, and AI/ML platforms, all driven by the CDP Private Cloud Base for data storage and CDP Private Cloud Data Services for data analysis.

A storage tier with on-premises and cloud-based compute farms to enable you to gain value from data wherever it resides, and whether it is structured, unstructured, or object-based data. You have the freedom to move applications and data between your data center and multiple clouds and process it regardless of location.

CDP Private Cloud Base

The data portion of the solution provides the following components:

- The **Cloudera Unified Data Fabric** centrally orchestrates disparate data sources intelligently and securely between your data center and multiple clouds.

Apache Ozone innovations

The Cisco Data Intelligence Platform uses Apache Ozone for on-premises data storage. Ozone promises to break through limitations of Apache Hadoop, including:

- Scalability to exabyte storage capacity
- Support for billions of files, more storage per node, and larger drives
- Separation of control and data planes for higher performance.

What these improvements deliver to users include:

- Lower infrastructure costs
- Reduced software licensing costs
- Smaller data center footprint
- Support for more use cases

Storage-tier implementation

Cisco Validated Designs specify servers for the Apache Ozone storage cluster depending on your capacity and performance requirements:

Cisco UCS C240 M6 Rack Server

For environments needing high performance, the Cisco UCS C240 M6 can support up to 24 small-form-factor (SFF) Intel® SSDs including two Intel NVMe caching drives.

Cisco UCS S3260 Storage Server

When high capacity is required, the Cisco UCS S3260 is configured with up to 56 large-form-factor drives plus two NVMe caching drives. The server's unique architecture allows it to be configured with one or two 2-socket server nodes, enabling performance to be tuned to application needs.

- The **Cloudera Open Data Lakehouse** that brings together the benefits of a data lake and a data warehouse to enable multifunction analytics on both streaming and stored data in a cloud-native object store across hybrid and multicloud, all while helping reduce TCO. Strong data quality, reliability, and management (including security, governance, and lineage) are key qualities You can manage data directly within the native format appropriate for each cloud (for example, S3 when used in Amazon Web Services, or Ozone for an on-premises data center.)
- The **Cloudera Scalable Data Mesh** that helps organizations scale and optimize data alignment by enabling access to cross-functional teams all

under a single data infrastructure, treating data as a product owned by functional domains.

Cisco UCS servers form the on-premises storage tier presented by Apache Ozone (see sidebar). These servers support extremely fast data ingest and data engineering performed in the data lake. Cisco UCS gives you a highly scalable storage pool that can scale to exabyte size with automated deployment and single-pane management through the Cisco Intersight™ cloud-operations platform. This platform provides full lifecycle management of your infrastructure, including connection to the Cisco® Technical Assistance Center to proactively respond to hardware problems that may occur.
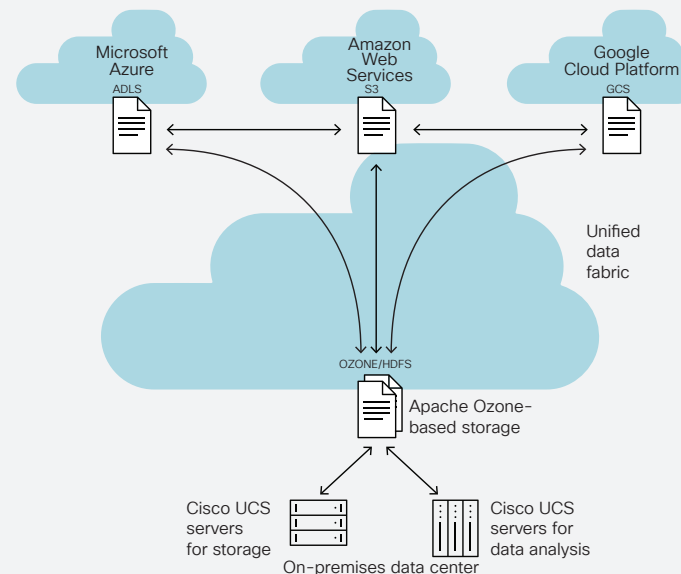

Figure 1. Cisco Data Intelligence Platform with Cloudera supports the Cloudera Open Data Lakehouse that supports native interfaces in the three major cloud providers

allialli
CISCO

The bridge to possible

Extract knowledge from your data

The compute farm is powered by the Cisco UCS X-Series Modular System with Intersight. The perfect complement to your public cloud deployments, this foundation for your private cloud is designed to think like software so you can think like tomorrow. The system hosts up to eight 2-socket servers and can be augmented with PCIe sleds to add GPU acceleration for your AI/ML workloads.
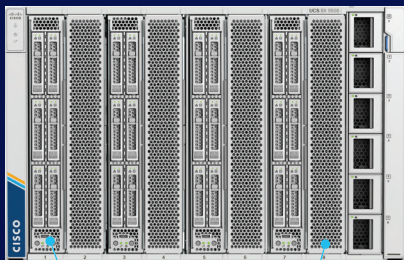
Cisco UCS X210c
M6 Compute Node

Cisco UCS X440p
PCIe Node

Cisco UCS X-Series Modular System

LE-84101-00   10/22

Cisco has created Cisco Validated Designs for all aspects of the solution. These specify the servers to use for storage depending on the size and performance requirements of your solution. The designs specify one of the three servers described in the sidebar on this page.

CDP Private Cloud Data Services

A data analytics compute farm is established with the Cisco UCS X-Series Modular System with Intersight (see sidebar). This flexible, adaptable, cloud-managed platform can support standard data analytics and also GPU-accelerated AI/ML frameworks through its PCIe node.

CDP Private Cloud Data Services establishes a platform for portable data analytics that enables you to move data and applications on premises or among your choice of clouds, helping you to quickly adapt to changing business conditions. The platform manages security, governance, metadata, replication, and automation across the data lifecycle. This enables you to run analytics on public clouds, on premises, and at the network edge.

Working together

The Cisco Data Intelligence Platform combines Cisco UCS servers, the Cisco Intersight cloud-operations platform, with Cloudera software to deliver a platform that gives you freedom to choose your cloud, your way:

- Cloud-scale architecture that can handle your needs regardless of size
- Rapid data ingest with Apache Ozone optimizations

- Independent scaling of storage and analysis capabilities with automated scaling and burst capacity to respond to workload changes
- Hybrid-cloud compute farm that is ready to support AI-based analysis and extraction of knowledge
- Lifecycle infrastructure management through Cisco Intersight
- Full-stack support from Cisco and Cloudera
- Scale storage/analysis capabilities independently; scales automatically in response to workload changes

Why Cisco and Cloudera

Cisco and Cloudera, working together, have resulted in multiple Cisco Validated Designs. These are tested and validated solutions that help reduce the complexity of deploying hybrid-cloud solutions. These designs help you size your solution to your needs, and help you deploy more rapidly, with less cost and risk.

This joint effort has resulted in a data platform that can help you extract more knowledge from all of your data, whether it is structured, unstructured, or object-based.

Learn more

Refer to the following Cisco Validated Design:

- Cisco Data Intelligence Platform on Cisco UCS C240 M5 with Cloudera Data Platform Running Apache Ozone