

CLOUDERA

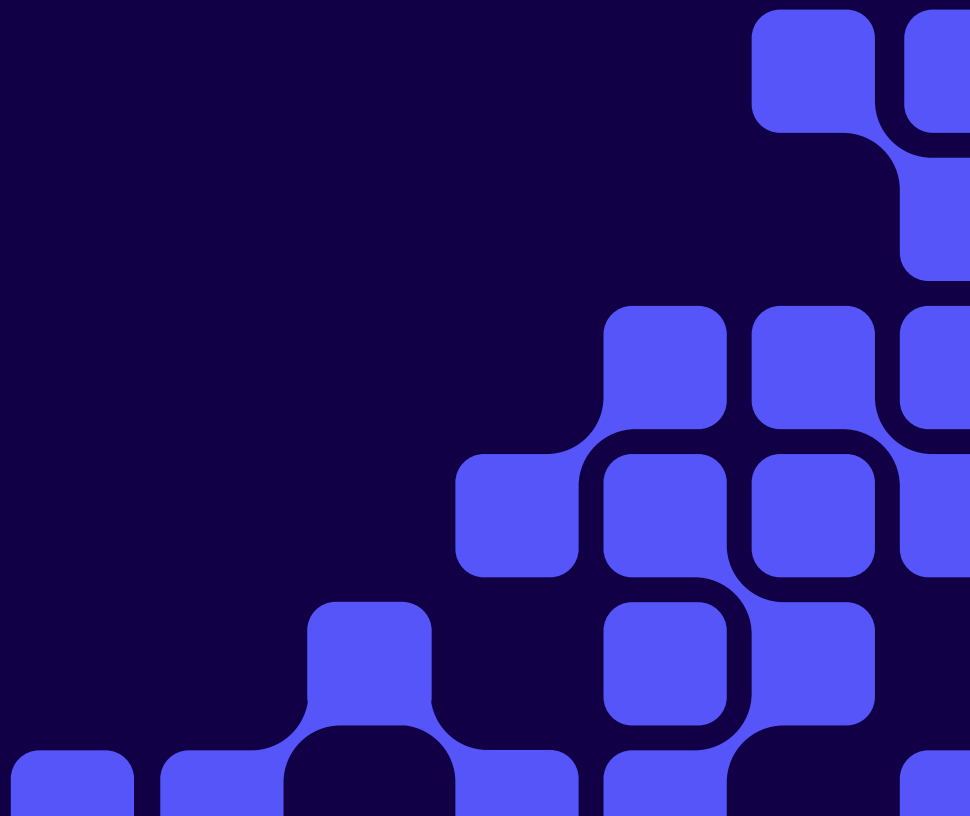
NUTANIX

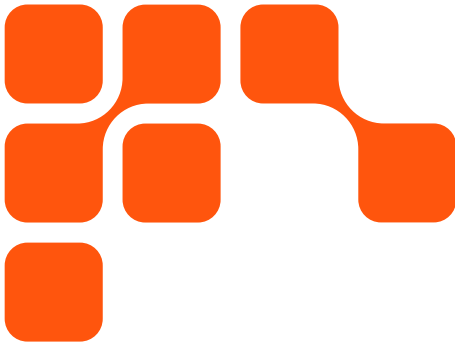
JOINT TECHNICAL WHITEPAPER

Cloudera on Premises on Nutanix

End-to-End Deployment, Storage
Integration & Data Services Validation

By Yash Gulati, Kuldeep Sahu, and Vishnu Prasad G J





Executive Summary

This white paper documents the joint field validation conducted by Cloudera and Nutanix teams to confirm that Cloudera Base on premises and its Data Services — Cloudera Data Engineering, Cloudera AI, and Cloudera Data Warehouse — run successfully on Nutanix AHV hyperconverged infrastructure, using Nutanix Objects as the S3-compatible object store and Nutanix Files as the NFS storage layer for AI workloads. Five integration scenarios were tested across a production-representative environment running Cloudera Runtime 7.3.2 on Nutanix AOS 7.5.1 and AHV 11.0.1.

Key Validation Highlights

- Cloudera Base (HDFS + Hive + Spark): Full CRUD via S3A on Nutanix Objects — validated (Non-RAZ)
- Cloudera Data Engineering: Log offload to Nutanix Objects
- Cloudera AI: Workbench provisioned with Nutanix Files NFS
- Cloudera Data Warehouse: Hive VW created; Parquet table location confirmed in Nutanix Objects
- RAZ + Nutanix Objects: Architectural incompatibility documented — STS AssumeRole HTTP 501

1. Business Context & Joint Solution

As data volumes grow and AI workloads move from public cloud to hybrid and on-premises environments, enterprises face a critical infrastructure decision: how to run a modern, cloud-native data platform without the egress costs, compliance risks, and latency of public cloud — while still supporting the analytics, AI, and data engineering use cases that drive business value.

The answer for a growing number of enterprises is Hyperconverged Infrastructure (HCI) paired with a proven data platform. Nutanix AHV and Cloudera Base on premises address this problem directly

Why Nutanix + Cloudera?

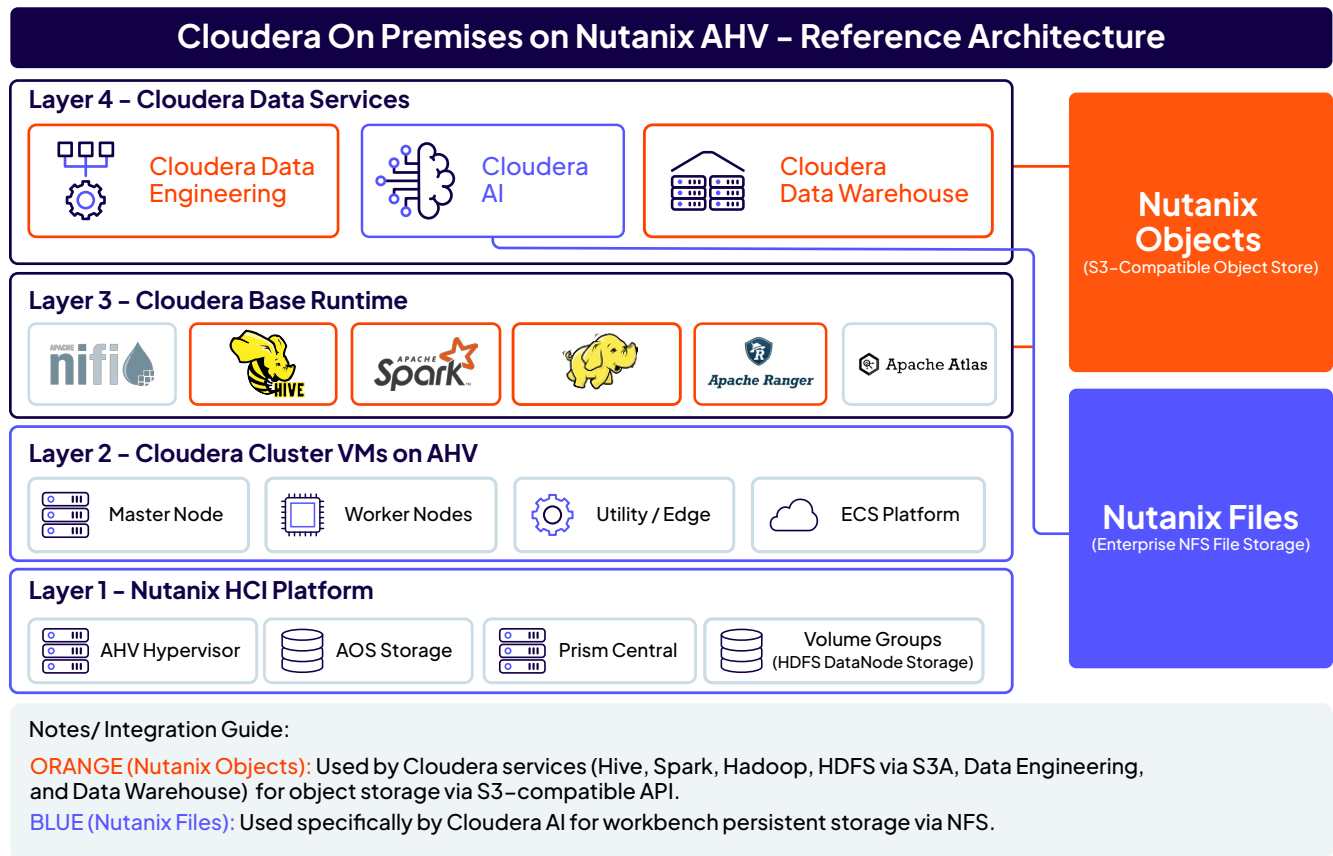
- **Simplified Infrastructure:**
Nutanix AHV converges compute, storage, and networking — managed through a single pane (Prism Central). No separate SAN, NAS, or hypervisor licensing required.
- **S3-Compatible Storage On-Premises:**
Nutanix Objects delivers an S3-compatible API endpoint natively on the cluster, allowing Cloudera to use its existing S3A connector — the same connector used in the cloud — without any cloud dependency.
- **NFS for AI Workloads:**
Nutanix Files delivers enterprise NFS shares for Cloudera AI Workbench file storage — removing the need for a dedicated NAS appliance.
- **Data Sovereignty & Security:**
All data stays on-premises on Nutanix infrastructure. Kerberos, AutoTLS, Ranger, and Knox provide a defense-in-depth security posture.
- **Validated Blueprint:**
This white paper provides a tested, field-validated reference architecture — reducing deployment risk and shortening time-to-production.

Validated Technology Stack

Component	Version	Role
Nutanix AHV / AOS	11.0.1 / 7.5.1	Hypervisor + software-defined storage (HCI platform)
Nutanix Prism Central	pc.7.5.1	Unified management for VMs, Objects, and Files
Nutanix Objects	5.3.0.1	S3-compatible object storage
Nutanix Files	5.3	Enterprise NFS file storage
RHEL (RedHat Enterprise Linux)	9.6	Operating system for all Cloudera cluster VMs
Cloudera Manager	7.13.2	Centralised Cloudera cluster management & configuration
Cloudera Runtime (Unified)	7.3.2	Core runtime — Hive, Spark, HDFS, Ranger, Knox, Atlas
Cloudera Data Services (ECS)	1.5.5 SP2	Kubernetes platform for Cloudera Data Engineering, Cloudera AI, Cloudera Datawarehouse

2. Reference Architecture

The validated architecture places Cloudera cluster VMs on Nutanix AHV, with all data services connecting to Nutanix Objects via the S3A protocol over an internal FQDN endpoint. Nutanix Files provides the NFS backend for AI workloads. Prism Central serves as the unified management plane for all Nutanix infrastructure resources.



Key Architectural Decisions

- Volume Groups for HDFS DataNodes:**
 Nutanix Volume Groups (VGs) are used for HDFS DataNode storage — they persist independently of VM lifecycle, protecting HDFS data if a VM is accidentally deleted. Standard vDisks are not recommended for DataNode storage.
- Custom FQDN for Nutanix Objects:**
 A DNS-resolvable FQDN (cloudera-ntnx.ntnxlab.local) registered via Prism Central is used for all S3A connections. The Objects SSL certificate is distributed to all Cloudera nodes and the CM truststore.
- HDFS RF2 + AOS RF2:**
 We have used HDFS RF2 + Nutanix AOS RF2 as part of this validation. For production deployments, customers can choose between two options based on their resiliency and storage requirements:
 - Option A: AOS RF1 + HDFS RF3 – 3 total copies, optimized for storage efficiency.
 - Option B: AOS RF2 + HDFS RF2 – 4 total copies, providing balanced layered resiliency.
- Active Directory Integration:**
 AD is used as both the DNS provider (A-Records for Nutanix Objects/Files FQDNs) and identity provider for Kerberos, LDAP, and Knox SSO across Cloudera Data Platform.
- Security Stack:**
 AutoTLS encrypts all inter-service traffic. Ranger enforces fine-grained data access policies. Knox provides SSO gateway access. Atlas manages metadata and lineage.

3. Storage Integration Validation

Service / Component	Validation Finding	Result
Cloudera Base — HDFS/Hive (Non-RAZ)	Full CRUD via S3A on Nutanix Objects. External table mapped to s3a://cloudera-ntnx. SELECT, INSERT, and cross-engine read validated.	PASS
Cloudera Base — Spark (Non-RAZ)	Spark Shell read Hive metastore catalog over Nutanix S3A. Table creation (Parquet) and INSERT/SELECT confirmed end-to-end.	PASS
Cloudera Base — RAZ + Nutanix Objects	AWS STS AssumeRole (required by Knox IDBroker) returns HTTP 501 Not Implemented from Nutanix Objects. Dynamic credential exchange not supported.	INCOMPATIBLE
Cloudera Data Engineering	Ozone log offload configured via Cloudera API to Nutanix Objects bucket (cloudera-ntnx). HTTP 200 OK confirmed. Sample Spark/Airflow jobs executed successfully with logs persisted to Nutanix Objects.	PASS
Cloudera AI — NFS Workbench	NFS share (ntnx-file-share) created on Nutanix Files, mounted on Cloudera Manager node, and attached as external NFS storage for Cloudera AI Workbench provisioning. Workbench status: Ready.	PASS
Cloudera Data Warehouse	Hive Virtual Warehouse created. Parquet table (customer_details) created, data inserted, queried. Table location confirmed in Nutanix Objects at s3a://cloudera-ntnx/warehouse/	PASS

3.1 Cloudera Base — HDFS, Hive & Spark (Non-RAZ)

Configuration Applied

Nutanix Objects was configured as the S3A storage backend via Cloudera Manager Safety Valve (cluster-wide core-site.xml). Hive warehouse directories were remapped to s3a://cloudera-ntnx.

core-site.xml — Key S3A Properties (Cloudera Manager Safety Valve)

```
fs.s3a.access.key = [REDACTED]
fs.s3a.secret.key = [REDACTED]
fs.s3a.endpoint = http://10.54.59.142 <!-- Nutanix Objects IP -->
fs.s3a.path.style.access = true
fs.s3a.signing.algorithm = AWSS3V4SignerType
fs.s3a.connection.ssl.enabled = false
fs.s3a.change.detection.mode = none
fs.s3a.endpoint.region = us-east-1
```

Hive Warehouse Directories (Hive Service — Safety Valve)

```
hive.metastore.warehouse.dir = s3a://cloudera-ntnx/warehouse/
  tablespace/managed/hive

hive.metastore.warehouse.external.dir = s3a://cloudera-ntnx/
  warehouse/tablespace/external/hive
```

Functional Testing Performed

- Data staging (S3A):**
 CSV data staged to bucket path s3a://cloudera-ntnx/warehouse/.../employee/ using HDFS S3A client. Upload and listing verified successfully.
- External table creation (Hive):**
 Created an external table employee_ext pointing to the Nutanix Objects S3 path via Beeline CLI. Table registered in Hive Metastore.
- Read validation — SELECT:**
 Verified the data correctly loading from data stored in Nutanix Objects.
- Write validation — INSERT (Tez):**
 Validated the data insertion via Hive Tez engine. Data was successfully inserted in the table.
- Spark interoperability:**
 Spark Shell launched with Hive metastore catalog. Existing Hive table read over S3A, new Parquet table created and queried — both against Nutanix Objects. Cross-engine consistency confirmed.

3.2 Ranger Authorization Service (RAZ) — Architectural Finding

RAZ was tested to determine whether fine-grained S3 path-level access control could be applied to Nutanix Objects. The testing revealed a fundamental architectural incompatibility.

Configuration Attempted Page

- **Knox IDBroker Setup:**
Long-lived Nutanix Objects static keys injected into Knox IDBroker via Save Alias. Custom cab-providers topology configured to point IDBroker at the Nutanix Objects FQDN.
- **Ranger RAZ Parameters:**
ranger.raz.bootstrap.servicetypes=s3 configured. IDBroker CAB address and Knox token URL set to cluster FQDNs.
- **User Mapping Options Tested:**
AWS Role ARN format (arn:aws:iam:::role/S3RazRole) — failed: Nutanix Objects cannot process AWS Role ARNs. AWS User ARN format — failed: Knox IDBroker requires STS token exchange, which Nutanix Objects does not support.

Root Cause

Cloudera RAZ requires Knox IDBroker to call the AWS STS AssumeRole API to exchange principal identities for short-lived session tokens. Nutanix Objects supports the core S3 data-plane APIs (GET, PUT, DELETE, LIST) and static key authentication, but does not implement the AWS STS control-plane API. When IDBroker calls the STS endpoint, it receives HTTP 501 Not Implemented — the dynamic credential exchange cannot complete.

Incompatible: RAZ + Nutanix Objects — AWS STS AssumeRole is not implemented in Nutanix Objects (HTTP 501). This is an architectural gap, not a configuration error.

Note: - RAZ support is a Cloudera recommendation and Nutanix Objects roadmap item. As per Cloudera's discussion with Nutanix Engineering Team, the feature is under consideration and planned to be covered in future Nutanix Objects release.

3.3 Cloudera Data Engineering — Nutanix Objects Log Offload

Configuration Applied

Cloudera Data Engineering log offload was configured via the Cloudera Management Console REST API, directing Ozone log storage to the Nutanix Objects bucket and FQDN endpoint. The key API parameters used:

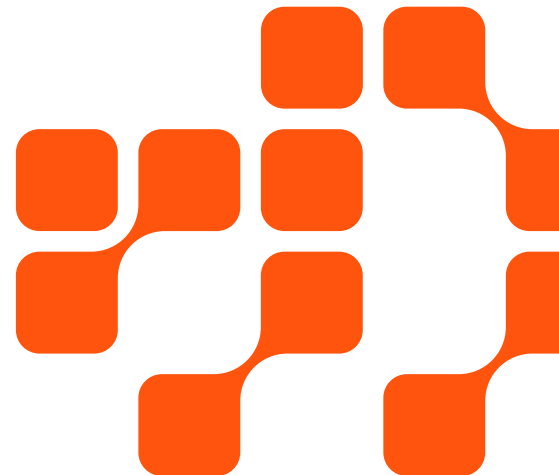
setEnvironmentSetting API — Key Parameters

```
"logType": "OZONE"  
"ozoneLogsBucket": "cloudera-ntnx" "ozoneLogsPath": "/logs"  
"ozoneS3RestUrl": "https://cloudera-ntnx.ntnlab.local"
```

API Response: HTTP/1.1 200 OK → Configuration accepted

Functional Testing Performed

- **Spark jobs:**
Some example Spark jobs were loaded and executed successfully in the Cloudera Data Engineering Virtual Cluster.
- **Airflow job:**
An example Airflow job was loaded and execution completed successfully.
- **Log verification in Nutanix Objects:**
Cloudera Data Engineering log directory structure confirmed in Nutanix Objects browser: logs/dex/cluster-*/apps/spark-*/driver/executor_1/ — stderr log file (stderr_*.gz)
- **Kubernetes secrets validation:**
Cloudera Data Engineering ozone-secrets (Kubernetes Opaque type) inspected to confirm S3 credentials correctly mounted in the dex-app namespace.



3.4 Cloudera AI — Nutanix Files NFS Workbench

Configuration Applied

An NFS share was created on Nutanix Files (ntnx-nfs) via Prism Central. The share was mounted on the Cloudera Manager node, the workbench directory was created with the required ownership, and the Cloudera AI Workbench was provisioned using the Nutanix NFS path.

Nutanix Files NFS Share — Configuration

Share Name: ntnx-file-share
Share Path: /ntnx-file-share
Protocol: NFS (NFSv3 + NFSv4)
Client Access: Read-Write
Max Size: 1000 GiB | Compression: Enabled
Host FQDN: ntnx-nfs.ntnslab.local

Cloudera AI Workbench Provisioning — Key Settings

NFS Server Type: External
NFS Path: nfs://ntnx-nfs.ntnslab.local:/ntnx-file-share/workbench1
NFS Protocol: 4.1
Namespace: cai-ntnx | Environment: ntnslab

Functional Testing Performed

- **DNS & NFS export:**
nslookup confirmed FQDN resolution.
showmount -e listed /ntnx-file-share —
NFS export confirmed active.
- **Mount & permissions:**
NFS share mounted at /mnt/cai. Workbench directory created with chown 8536:8536 and chmod 2775 (required by Cloudera AI Workbench service account).
- **MLXControlPlane provisioning:**
All provisioning events completed: CSI drivers installed, storage provisioned, storage mounted, access point created, AI application installed — no errors.
- **Workbench status:**
Cloudera AI Workbench provisioning was successful (version 2.0.51).
- **Directory verification:**
NFS workbench directory confirmed populated post-provisioning; required workbench directories created with correct ownership.

3.5 Cloudera Data Warehouse — Hive Virtual Warehouse

Configuration & Testing

A Hive Virtual Warehouse (test_hive_vw) was deployed on the ECS platform. SQL workloads were executed via the Cloudera Data Explorer against the Hive VW, with table storage backed by Nutanix Objects.

Cloudera Datawarehouse Hive VW — Storage Location Confirmed (DESCRIBE FORMATTED output)

Table: customer_details (MANAGED_TABLE, STORED AS PARQUET)
Database: test_hive_vw | Owner: ygulati
Location: s3a://cloudera-ntnx/warehouse/tablespace/managed/hive/test_hive_vw.db/customer_details
delta_0000001_0000001_0000/
000000_0 (1.05 KiB — confirmed in Nutanix Objects browser)

Functional Testing Performed

- **Database creation:**
Created a database test_hive_vw and confirmed in Hive metastore.
- **Table creation (Parquet):**
Created a parquet table 'customer_details' inside the database and confirmed storage location as s3a bucket.
- **Data insertion and Read:**
Inserted data into the table and confirmed the insertion and read was successful from Nutanix objects bucket.
- **Table location verification:**
Confirmed the table location as s3a://cloudera-ntnx/warehouse/.../customer_details by using the DESCRIBE FORMATTED command
- **HDFS CLI validation:**
Confirmed the delta directory readable via the S3A connector from the Cloudera master node by using the hdfs dfs -ls s3a://cloudera-ntnx/warehouse/.../customer_details command.

4. Conclusion, Benefits & Next Steps

Four of five tested integration scenarios passed end-to-end field validation. The fifth (RAZ) produced a clearly documented architectural finding — providing precise guidance for the Nutanix Objects roadmap and Cloudera field teams. This white paper forms the technical foundation for formal Cloudera–Nutanix certification.

Key Benefits

- **Drop-In S3 Compatibility:**
Nutanix Objects works natively with the S3A connector built into Cloudera Runtime — no code changes, no custom plugins. Configure endpoint and keys in `core-site.xml` and it works.
- **Unified On-Premises Object Storage:**
A single Nutanix Objects bucket serves Hive, Spark, Cloudera Data Engineering log offload, and Cloudera Data Warehouse — simplifying storage operations and capacity planning without cloud dependency or egress cost.
- **NFS for AI Without a NAS Appliance:**
Nutanix Files replaces a dedicated NAS for Cloudera AI Workbench — enterprise NFS at lower cost and with Prism Central unified management.
- **Simplified HCI Operations:**
AHV eliminates separate hypervisor, SAN, and NAS licensing. VMs, storage, Objects, and Files all managed through a single Prism Central console.

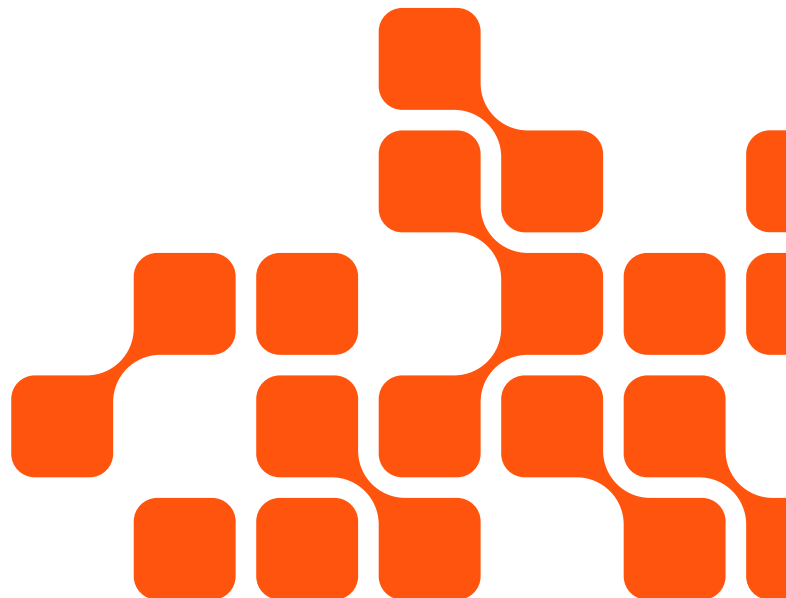
Production Recommendations

- **HDFS Replication and Nutanix AOS Resiliency:**
HDFS follows the standard Hadoop architecture, where the default replication factor is 3 to ensure data durability and availability. However, because Nutanix AOS also provides data protection by replicating storage blocks at the infrastructure layer (typically RF2), the HDFS replication factor can be reduced to 2. This balances redundancy across both the application and storage layers while lowering the overall storage footprint and infrastructure cost.

For deployments where minimizing storage consumption is the primary objective, customers may choose Nutanix AOS RF1. In this configuration, it is recommended to retain the standard HDFS replication factor of 3 to maintain data protection at the application layer. This results in an effective overall redundancy of 3 copies (compared to 4 copies with HDFS RF2 + AOS RF2), offering additional storage savings while accepting a lower level of storage-layer fault tolerance.

- **Register a custom FQDN for Nutanix Objects:**
Required for SSL-enabled S3A and Cloudera Data Engineering integration. Configure via Prism Central → Manage FQDN & SSL Certificates.
- **Use Volume Groups for HDFS DataNode storage:**
VGs persist independently of VM lifecycle — critical for HDFS data durability on Nutanix. Standard vDisks are not recommended.
- **Distribute the Nutanix Objects SSL certificate:**
Inject into the Java cacerts store on all cluster nodes and Cloudera Manager truststore to enable HTTPS S3A connections.

Note: Cloudera does not require WORM or versioning on its object storage buckets – enabling WORM would in fact conflict with Cloudera’s normal write operations, so both features would remain disabled on any bucket used by Cloudera services.



References

Cloudera on Nutanix AHV Deployment Blog:

<https://community.cloudera.com/t5/Developer-Blogs/Running-Cloudera-on-premises-on-Nutanix-AHV-An-End-to-End/ba-p/413520>

Cloudera On-Premises Deployment Guide (ECS):

https://github.com/cloudera/ClouderaSetup/blob/main/OnPremise/ECS/Cloudera_On-Premise_Deployment_Guide-ECS.pdf

Cloudera + Nutanix Best Practices Guide:

https://github.com/cloudera/ClouderaSetup/blob/main/OnPremise/ECS/Nutanix_Cloudera_Best_Practices.pdf

Cloudera RAZ Documentation (Cloudera Data Platform 7.3.2):

<https://docs.cloudera.com/cdp-private-cloud-base/7.3.2/security-ranger-aws-access-control-with-raz/>

Nutanix Objects Documentation (S3-Compatible Object Storage)

https://portal.nutanix.com/page/documents/details?targetId=Nutanix-Objects-v5_3:Nutanix-Objects-v5_3

Nutanix Files Documentation (Enterprise NFS Files Storage)

https://portal.nutanix.com/page/documents/details?targetId=Nutanix-Files-v5_3:Nutanix-Files-v5_3

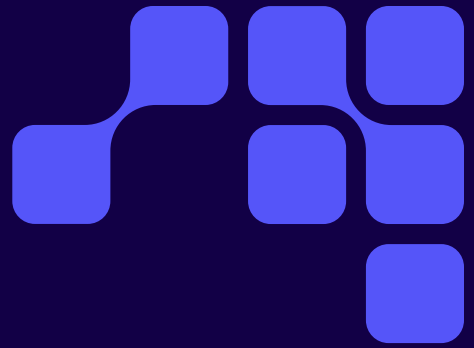
Prism Central Documentation

https://portal.nutanix.com/page/documents/details?targetId=Prism-Central-Guide-vpc_7_5:mul-intro-doc-organization-pc-c.html

About Cloudera

Cloudera is the only hybrid data and AI platform company that large organizations trust to bring AI to their data anywhere it lives. Unlike other providers, Cloudera delivers a consistent cloud experience that converges public clouds, on-prem data centers, and the edge, leveraging a proven open-source foundation. As the pioneer in big data, Cloudera empowers businesses to apply AI and assert control over 100% of their data, in all forms, improving security, governance, and real-time and predictive insights. The world's largest brands across all industries rely on Cloudera to transform decision-making and ultimately boost bottom lines, safeguard against threats, and save lives.

To learn more, visit [Cloudera.com](https://cloudera.com) and follow us on [LinkedIn](#) and [X](#).



CLOUDERA

Cloudera, Inc. | 6220 America Center Dr, 5th Fl, San Jose, CA 95002 USA | cloudera.com

© 2026 Cloudera, Inc. All rights reserved. Cloudera and associated marks are trademarks or registered trademarks of Cloudera, Inc. All other company and product names may be trademarks of their respective owners. 0000-001 Aug19, 2026