SOLUTION OVERVIEW

CAEVES Context LAKE™

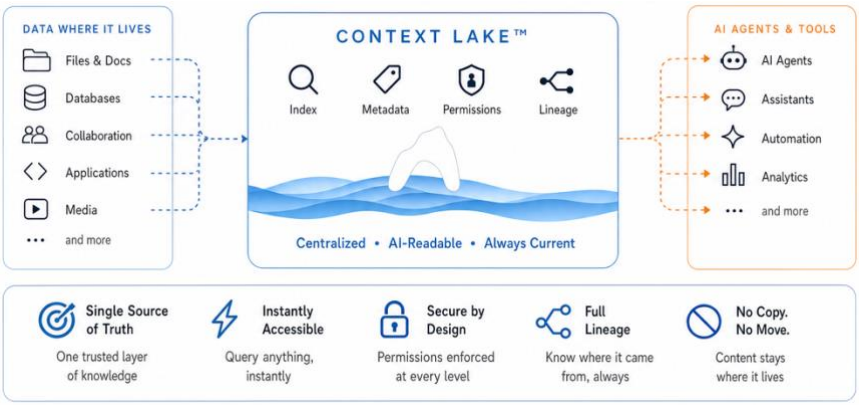
Connect, index and activate any data, for any AI assistant, in any cloud.

Enterprise knowledge is often scattered across decades of files, data repositories and storage platforms, making it difficult for people to find and for organizations to make available to search and AI.

CAEVES Context LAKE™ brings that information into reach, creating trusted context that AI can access without requiring organizations to migrate their data, replace existing infrastructure or compromise security and governance.

CAEVES Context LAKE™ is the secure context layer between the storage you already own and the AI assistants your teams already use. It connects to file shares, object stores and Microsoft 365 libraries across any hyperscaler, private datacenter or sovereign environment, indexes that content in place and read-only, and serves it to any AI assistant through the open Model Context Protocol (MCP). Nothing is migrated, nobody gains new access, and every answer points back to its source.

What is CAEVES Context LAKE™?
A centralized, AI-accessible memory layer that continuously organizes an organization's unstructured and structured data into instantly accessible knowledge: a single source of truth that AI agents can query, free of the limits of basic search or static context windows.



Unlike a data lake, nothing is copied into it. Content stays where it lives; the context lake holds the index, metadata, permissions and lineage.

SOLUTION HIGHLIGHTS

0	< 1 hr	> 70%	2–4%
BYTES MIGRATED	**TO A SEARCHABLE INDEX**	**LOWER AI COST**	**INDEX OVERHEAD**
Indexed where it lives, no copy of record	From deployment to first governed results	Compared with building custom RAG infrastructure	Versus 2–10x for typical vector stores

THE CHALLENGE

Your organization knows a lot. Your AI knows almost none of it.

01	02	03
People cannot find it	**AI cannot reach it**	**Sharing feels risky**
Decades of reports, contracts, drawings and project records sit in shares and archives nobody can search well.	Assistants only know what users paste in. Private documents stay invisible, so answers are generic or wrong.	Sending sensitive content to an external AI service is a non-starter in regulated and security-conscious industries.

CAEVES Context LAKE™ resolves all three at once: dark data becomes searchable for people and reachable for AI, while the data and its permissions stay inside your own boundary.

HOW IT WORKS

Five steps from dark data to governed, cited answers

1	**CONNECT**	Point Context LAKE™ at object stores, file shares and SharePoint libraries with read-only credentials. It never writes to, moves or renames source data.
2	**EXTRACT**	Self-hosted engines read PDFs, Office files and spreadsheets, apply OCR to scans, optionally use GPU vision models, and transcribe audio and video, including on-screen text, with timecodes.
3	**ENRICH**	Smart Actions redact identifiers and custom patterns, build a knowledge graph of people, organizations, locations and agreements, and feed Discover, a live overview of the whole estate.
4	**GOVERN**	NTFS and NFSv4 ACLs and POSIX ownership are captured at crawl time, mapped to directory identities and enforced with row-level security. Fail-closed by default.
5	**SERVE**	Hybrid lexical and semantic search in the console, CLI and REST API, and grounded AI through the built-in MCP server. Assistants sign in as the individual user and see only what that user may open.

ANY DATA · ANY AI

Connect what you already run. Serve every assistant you use.

PUBLIC CLOUD

Native and protocol-based connectors

Microsoft Azure: Blob Storage and ADLS Gen2 (native); Azure Files (SMB/NFS).

AWS: Amazon S3 (native); EFS (NFS); S3 File Gateway (SMB/NFS or its bucket).

Google Cloud: Cloud Storage (S3-compatible API); Filestore (NFS).

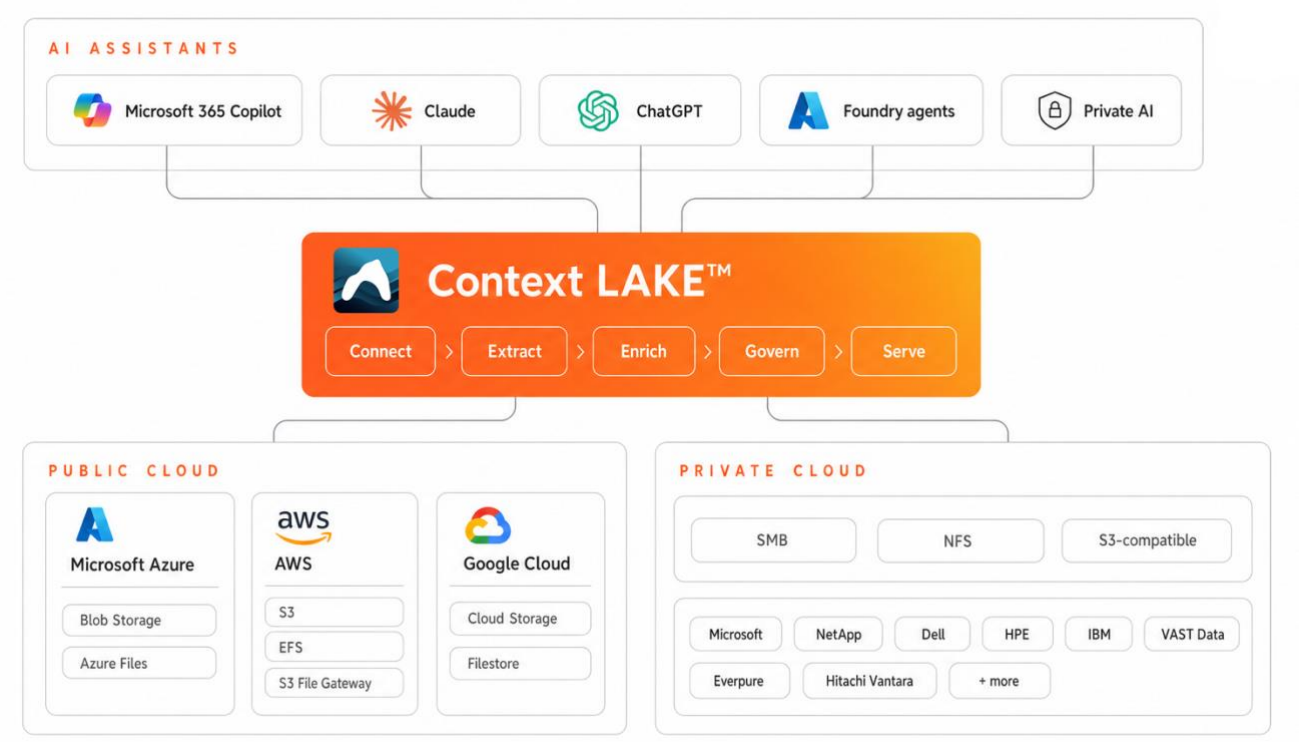
Microsoft 365: SharePoint Online and OneDrive through Microsoft Graph.

PRIVATE CLOUD AND ON PREMISES

Any SMB, NFS or S3-compatible storage

Works with storage from **Microsoft** (Windows Server, Azure Local), **NetApp, Dell, HPE, IBM, VAST Data, Everpure** (formerly Pure Storage), **Hitachi Vantara**, and S3-compatible stores such as MinIO and Wasabi.

Protocol-based, so no vendor-specific agent and no change to existing storage.



- **Any assistant, one governed index.** Microsoft 365 Copilot and Copilot Studio, Microsoft Foundry agents, Claude, ChatGPT, Gemini, Cursor and self-hosted models connect through MCP with OAuth 2.1 sign-in federated to Entra ID, Okta, Google or Keycloak.
- **Grounded and cited.** Answers reference the exact source document, so users can open the original behind any response.
- **Current, not a snapshot.** The index follows changes in the source, and switching or combining assistants needs no re-indexing.

SECURITY AND GOVERNANCE

Permissions travel with the data

1	2	3	4	5	6
Source ACLs	**Identity mapping**	**Redaction**	**Row-level security**	**User-scoped MCP**	**Cited, audited**
NTFS, NFSv4, POSIX captured at crawl	Entra ID, Okta, AD and others	Identifiers and custom patterns masked	Enforced per query, fail-closed	Assistant acts as the individual user	Every query and file open recorded

- **Read-only and in place.** CAEVES Context LAKE™ never writes to sources, and CAEVES never accesses customer data. Each source can be withheld from AI while remaining searchable by people.
- **Hardened by design.** Memory-safe services and connectors, TLS in transit, secrets sealed with authenticated encryption, signed container images, and audit logging of any send to an external provider.

SECURE PRIVATE AI

Keep both data and inference inside your boundary

MICROSOFT AZURE

Microsoft Foundry agents

Foundry agents call Context LAKE™ as a governed MCP tool, including over private, internal-only endpoints.

SOVEREIGN PRIVATE CLOUD

Foundry Local on Azure Local

Large models on customer hardware, including disconnected sites, grounded through Context LAKE™.

SECURE SITE, TEAM OR LAB

Self-hosted models, e.g. NVIDIA DGX Spark *NEW*

Run local AI models, fully air-gapped, answering through embedded Context LAKE™ platform.

Air-gapped ready. Extraction, OCR and transcription run on self-hosted engines, self-hosted models answer over MCP, and licenses validate offline, with zero egress. One solution to integrate in public cloud, hybrid, private cloud, sovereign and air-gapped environments.

USE CASES

Six ways organizations put Context LAKE™ to work

KNOWLEDGE, OPERATIONS, IT

Company file assistants

Assistants answer from policies, contracts and reports as the individual user, citing the original file.

Outcome: Answers that carry their own evidence, and AI that passes security review.

LEGAL, PROCUREMENT, FINANCE

Contract and counterparty intelligence

The knowledge graph maps counterparties, co-appearances and document subjects across agreements.

Outcome: Renewal and obligation dates surfaced before they lapse.

GOVERNANCE, SECURITY, STORAGE

Data estate governance

Discover shows composition by type, age, owner and tier, and flags world-readable, stale, unowned and duplicated files.

Outcome: A defensible baseline, with exposure reduced before AI rollout.

DEFENSE, PUBLIC SECTOR, FS, HEALTHCARE

Air-gapped and regulated environments

Everything runs on premises with self-hosted models, offline licensing and signed images.

Outcome: AI retrieval inside the boundary, with zero egress.

MEDIA, TRAINING, INVESTIGATIONS

Searchable media archives

Speech and on-screen text are transcribed and timecoded, so hits open at the exact moment.

Outcome: Hours of recordings searched in seconds.

SERVICE PROVIDERS, PLATFORMS

Multi-tenant service providers

Row-level tenant isolation, with a dedicated MCP endpoint, issuer and keys per tenant.

Outcome: Per-customer AI without per-customer infrastructure.

DEPLOYMENT AND PACKAGING

Deploy in your environment, license the capacity you connect

DEPLOYMENT

Runs wherever your data lives

Microsoft Azure: Cloud-native application package deployed from Microsoft Marketplace through an Azure Solution Template into your own Microsoft Azure subscription.

Any cloud or datacenter: Containers on Kubernetes (Helm) or Docker in AWS, Google Cloud, private and sovereign environments.

Evaluation / POC: Single-VM docker image, upgradeable to full version, without any rebuild or reindex.

Air-gapped and **multi-tenant** options available.

COMMERCIAL MODEL

Simple, capacity-based subscription

Annual or multi-year subscription on committed capacity, with no per-operation metering.

Graduated tiers: Starter, Business, Enterprise and Enterprise+, based on source capacity connected.

Procurement: Annual or multi-year subscription with offline license-file activation for on-premises, Azure, AWS and Google Cloud, or a Microsoft Marketplace private offer. AWS and Google Cloud marketplaces planned.

FREQUENTLY ASKED QUESTIONS

Do our files move or get copied?	No. Sources are accessed and indexed read-only, and originals stay exactly where they are. Context LAKE™ stores the index, metadata and permissions.
Can an assistant see everything indexed?	No. Per-file trimming is enforced in the database; assistants inherit the signed-in user's access, and sources can be withheld from AI.
Does any data leave our environment?	Processing runs on self-hosted engines. Any optional external send is audit-logged, and air-gapped mode has no egress.

GETTING STARTED

1. **Start a 30-day trial.** Share the sources and capacity you want to connect; we size a trial license to match.
2. **Deploy in your environment.** Your Azure subscription, other cloud tenant, datacenter or sovereign site.
3. **Connect, index and verify.** Register sources read-only, run the first crawl, search and open the original behind any hit.
4. **Connect an AI assistant.** Connect Microsoft Copilot, Copilot Studio, Claude, ChatGPT or any MCP client at your instance.

See Context LAKE™ on your own data.

Request a demo or a 30-day trial at caeves.com.

CAEVES and CAEVES Context LAKE™ are trademarks of CAEVES Technology, Inc. Other names are trademarks of their respective owners. Third-party storage is supported through standard protocols, not certification. Preview capabilities are labeled. Figures are indicative.