HITACHI

White Paper

Sovereign AI Starts With Resilient Data.

Hitachi Vantara

Building a trusted, governed data foundation for the sovereign AI data center.



HITACHI

White Paper

Table of Contents

03 Executive Summary

04 Sovereignty Is More Than Data Residency

- 04 Jurisdictional and Regulatory Control
- 04 Operational Control
- 04 Cryptographic Control
- 04 Technology Control
- 04 Data Lifecycle Control
- 05 AI Control
- 05 Resilience Control

05 Architecture of a Sovereign AI Data Center

- 07 The Data Plane
- 07 The AI Infrastructure Plane
- 07 The Management and Governance Plane

07 A Sovereign Data Lakehouse

- 09 One Authoritative Copy
- 09 Analytics Without the Copies
- 09 From SQL Analytics to AI-Ready Knowledge

09 Sovereign Use of Pretrained LLMs

- 09 Pattern 1: Deploy the Model Inside the Sovereign Environment
- 09 Pattern 2: Connect to an External Model Through a Controlled Gateway
- 10 Pattern 3: Ground models through RAG
- 10 Fine-Tune Selectively
- 10 Govern the Model Lifecycle

10 Governance as a Reusable Service

11 From Sovereign Data to Autonomous Operations

11 Conclusion: Reliable. Resilient. Ready.

Executive Summary

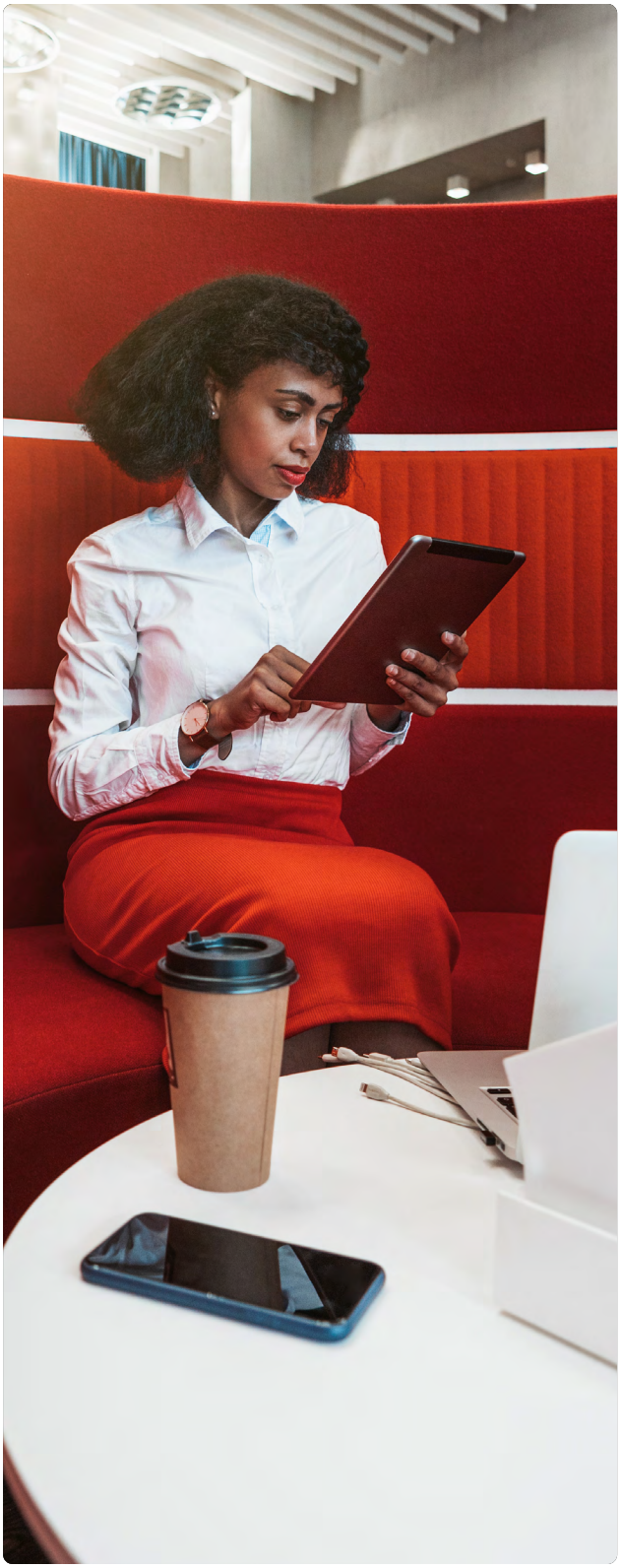
As AI systems gain the ability to process, generate and act on an organization's most sensitive data, control over that data has become a matter of national security for governments and a core requirement for digital enterprises.

Data residency no longer suffices. It describes where information sits, not who can reach it or whose laws govern the providers that handle it. It does not show who holds the encryption keys or whether the organization can keep operating and restore trusted data on its own.

A sovereign AI data center gives an organization or government sustained and provable control over its trusted data, metadata, models, infrastructure, operations and AI outcomes. It delivers cloud-like agility without surrendering authority of strategic digital assets or depending on any single provider or operating model.

Resilience is what makes sovereignty hold, and fragmentation is what breaks it. Data sits in silos, protection and governance tools do not share context and AI projects copy data faster than policy can follow. Integrated data resilience keeps data available, protected, and governed as one system. Sovereignty then becomes an outcome of resilient control, not a statement about location.

Hitachi Vantara closes that gap with an integrated platform built around enterprise data, so organizations can bring AI to their data without giving up control. The same foundation extends to Physical AI and increasingly autonomous operations.



Sovereignty Is More Than Data Residency

Data residency describes where data is stored; data localization requires certain data to stay within a prescribed boundary. Sovereignty goes further, applying legal, operational, technical and organizational control across the data lifecycle, whether the boundary is a country, region, regulated industry or enterprise-controlled environment, with carefully governed connections to public cloud services when appropriate. Unlike residency, which focuses on location, sovereignty focuses on maintaining control as conditions change.

Servers inside a country do not make a data center sovereign. What matters is who operates it; whose laws apply to the provider; who holds the encryption keys; where backups, metadata and telemetry go; and whether the organization can operate independently during a disruption. If policy breaks during failover, integrity cannot be established after an attack, or recovered data is used outside its approved boundary, sovereignty is compromised even if the infrastructure never leaves the country.

True data sovereignty requires sustained, provable control across seven dimensions. Each has a simple litmus test:

- Do you know which laws govern your providers?
- Can you govern operators and keep running if they fail?
- Do you alone hold the encryption keys?
- Can you change technology without losing your data?
- Do your data policies carry through to AI outputs?
- Can you decide what data AI sees and where it runs?
- Can you restore trusted data on your own?

If any answer is no, residency has not produced sovereignty. The sections that follow address each dimension in turn.

Jurisdictional and Regulatory Control

Do you know which laws govern your providers?

Data, metadata, models, logs, backups and derived information must follow the laws and policies that apply to each organization and workload. For a multinational organization, these can differ across Europe, Japan, the United States, the Middle East, India and beyond. A sovereign architecture identifies the jurisdiction of every data asset and enforces where it can be stored, processed, replicated, backed up or accessed.

Regulatory control means proving compliance, not asserting it, with independent evidence such as FIPS 140-3 validated cryptography, ISO/IEC 27001 certification and alignment with ISO/IEC 42001 for AI management. Depending on sector and jurisdiction, the EU's General Data Protection Regulation (GDPR), Network and Information System Directive 2 (NIS2), Digital Operational Resilience Act (DORA) and the Artificial Intelligence Act (AI Act) add obligations for resilience testing, incident reporting, third-party risk and AI governance.

Operational Control

Can you govern operators and keep running if they fail?

The organization must control who administers the platform and from where, how privileged access is approved, and how operator actions are recorded. Highly sensitive workloads may also require locally cleared personnel, separation of duties, dual authorization for critical actions and the ability to run without remote dependence on a foreign operator.

Cryptographic Control

Do you alone hold the encryption keys?

The organization should retain authority over encryption keys, secrets, certificates and rotation; encrypted data is not sovereign if another party can reach its keys, at rest or in motion. Bring-your-own-key and hold-your-own-key models help, but sovereignty also requires policy for backup keys, recovery procedures, hardware security modules and privileged access across the full lifecycle.

Technology Control

Can you change technology without losing your data?

Sovereignty requires portability and real exit options. Amazon S3®-compatible access, Apache Iceberg® tables, standard SQL, open APIs and portable model formats keep data accessible. Thus, the organization controls its own integration and is not permanently dependent on one infrastructure provider, cloud or operating model.

Data Lifecycle Control

Do your data policies carry through to AI outputs?

Sovereignty starts when data is created and continues through classification, processing, sharing, AI use, retention and secure deletion. Policy must follow data into curated datasets, embeddings, prompts, model outputs, reports and long-term AI memory, because derived information can be as sensitive as its source.

AI Control

Can you decide what data AI sees and where it runs?

The organization needs to know where models run, who selected them, which version is live and what data they can access. It must understand how prompts and outputs are retained, which tools agents can invoke and what happens when a model or agent acts outside policy.

Control extends to the intelligence itself: the origin of model weights, training-data lineage, known biases and limitations, behavior validated against intended use, and auditability for every deployed model.

It also reaches into economics. Governing GPU and CPU utilization, token consumption, model and API calls, storage and data movement through budgets, quotas, right-sizing and workload-aware model selection aligns AI operating cost with business value.

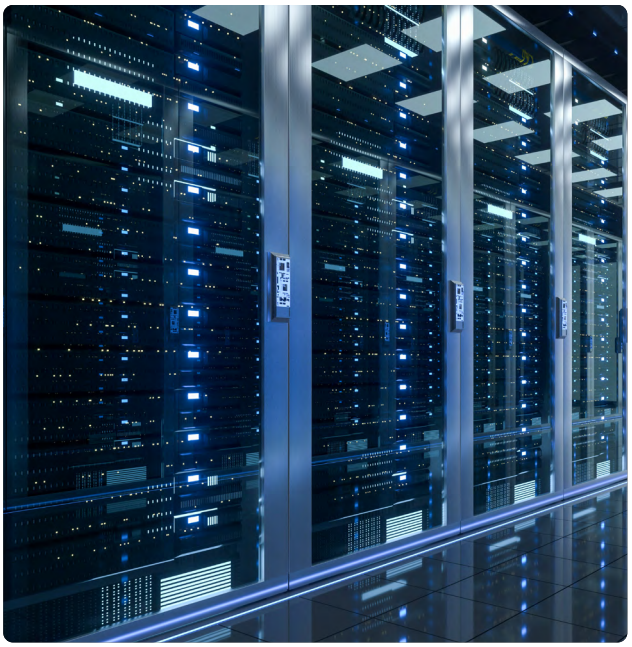
This is the difference between owning infrastructure and controlling intelligence.

Resilience Control

Can you restore trusted data on your own?

Resilience control determines whether every other dimension holds as conditions change. The organization must be able to withstand disruption, verify data integrity, identify a validated clean recovery point, recover critical services in business-dependency order and confirm that restored data still carries its governance, access, retention and residency policies. It must accomplish this without depending on an outside provider, even if a vendor's support is withdrawn. Restoring the latest copy is not enough if that copy cannot be trusted.

Because every other dimension depends on it, resilience is the foundation of sovereignty. Integrated data resilience makes that foundation practical, keeping data trusted, governed, protected, verifiable, recoverable and usable through disruption. It shapes the architecture that follows.



Architecture of a Sovereign AI Data Center

The seven dimensions define what must be controlled; architecture determines whether that control holds in practice. A sovereign AI data center should be built on a few key principles:

- **Start with data and govern the full chain.** Bring compute and intelligence to governed information where it already lives, and govern it from source records to embeddings, prompts, outputs and agent actions.
- **Separate storage from compute and stay open,** so processing engines and providers can change without relocating authoritative data or losing exit options.
- **Apply policy continuously,** as data moves and conditions change.
- **Design for disruption and prove resilience.** Plan for degraded or disconnected operation, and produce evidence of integrity, clean recovery and policy compliance.
- **Operate efficiently and sustainably.** High-density infrastructure, data reduction, intelligent power management and workload-aware utilization reduce footprint, power, cooling and cost.

SOVEREIGN AI DATA CENTER • CAPABILITY FRAMEWORK

Trusted Data. Complete Control. Intelligent Outcomes.

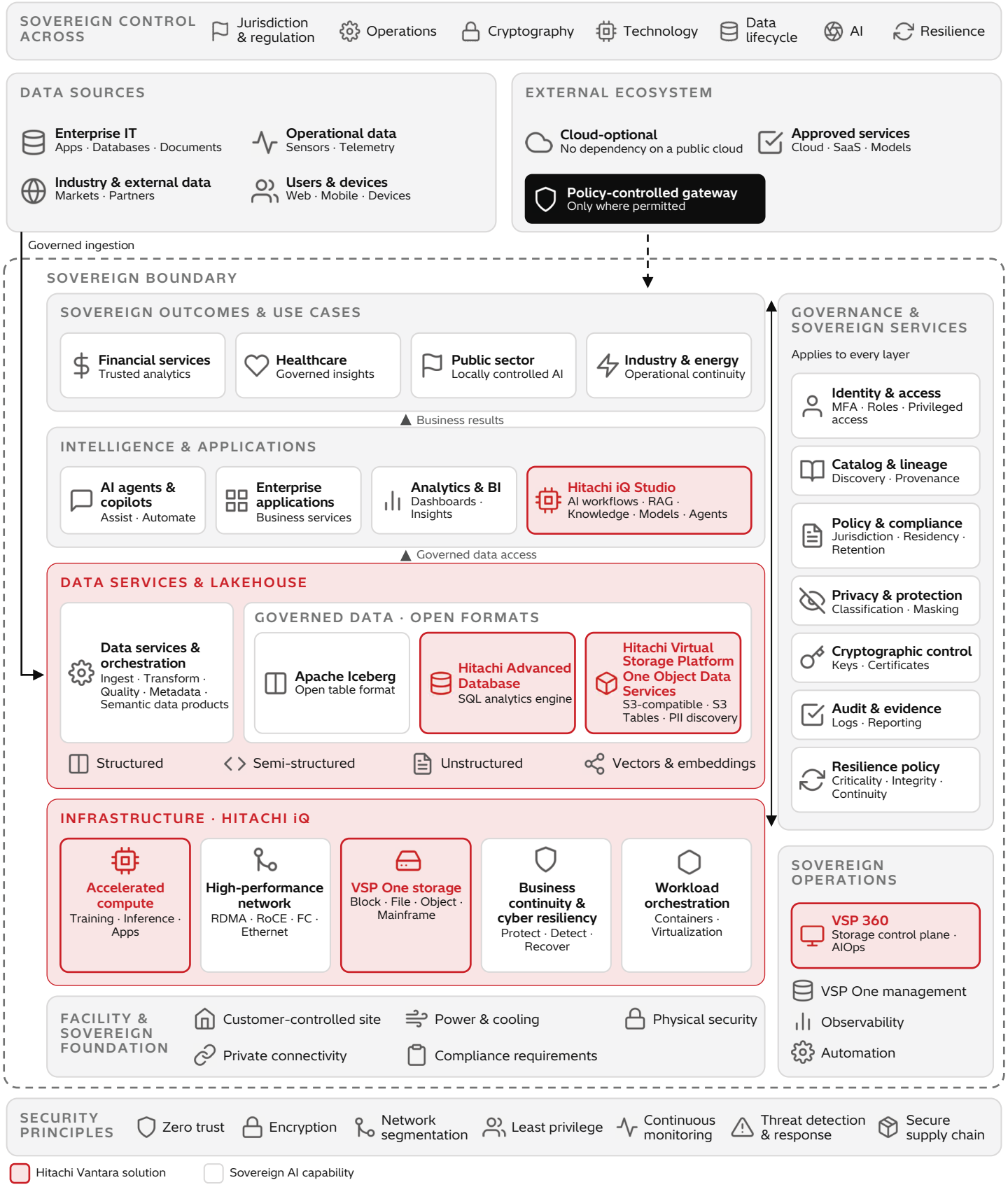


Figure 1. Sovereign AI data center capability framework

Figure 1 summarizes the capabilities a sovereign AI data center should include and how Hitachi Vantara applies these principles. Enterprise, operational, industry and user data enters through governed interfaces, and external services connect only through a secure gateway. Security and governance are built into each of the three planes that follow.

The Data Plane

Most organizations are challenged by silos: separate data infrastructure in different locations, each with its own tools, policies and administrators, and each a point where sovereignty can break. A sovereign data plane brings that infrastructure under one platform and a consistent governance model. Hitachi Virtual Storage Platform One (VSP One) does this across block, file, object, mainframe and software-defined data services, so infrastructure behavior can align with workload requirements.

In practice, transactional and low-latency inference data can use block services; documents and other unstructured content can use file services; and data lakes, archives, AI repositories and lakehouse tables can use object services. Policy determines not only where data is stored but also where it may be replicated, moved, processed or recovered, preserving jurisdictional, security, performance and operational requirements.

Not every dataset should move. Sensitive records can stay within a site, country, hospital, bank or government agency, while authorized metadata and selected data products are discoverable across the sovereign environment.



The AI Infrastructure Plane

Accelerated compute is only part of sovereign AI. Workloads also need a reliable data path that keeps them supplied with governed information inside the required boundary.

Training, fine-tuning, retrieval-augmented generation (RAG), inference, vector retrieval, model serving and agentic workflows each place different demands on that path. Hitachi iQ addresses them with integrated, production-ready infrastructure that brings accelerated computing, storage, networking and operations together in validated configurations, while keeping data and processing within the sovereign boundary.

The Management and Governance Plane

Sovereignty and resilience requirements hold only when they become policy that operations can apply, monitor and prove. Integrated data resilience provides the operating model for that, and data resilience zones turn it into policy.

Data resilience zones are the policy architecture for sustained control. A zone is a policy construct, not a physical availability zone or region. Starting from the minimum viable company, the critical services and data the organization needs to keep operating, each zone translates business criticality, jurisdiction, data classification and approved AI use into a defined policy profile for a class of services, workloads and data. That profile spans placement, access, encryption and key control, retention, immutable protection, integrity validation, recovery order, permitted processing, AI-use boundaries, operational mode, evidence and deletion.

Responsibilities stay distinct. Hitachi Virtual Storage Platform 360 (VSP 360) data management software provides the operational intelligence, visibility and policy-driven operations that help apply and monitor zone policy across the data foundation. VSP One implements the required placement, protection, availability and recovery behavior, and Hitachi iQ Studio governs models, agents and AI workflows. Together, they allow sovereignty and resilience requirements to persist as data moves, fails over, is recovered or is used by approved AI workflows.

A Sovereign Data Lakehouse

Many organizations hold multiple copies of the same data across object stores, warehouses, analytics platforms and AI environments. Every copy adds cost, governance complexity, security exposure and doubt about which version is authoritative. A sovereign data lakehouse is a practical entry point for reducing those copies while preserving open, governed access for analytics and AI.

SOVEREIGN DATA LAKEHOUSE · REFERENCE ARCHITECTURE

One copy of the data in open Iceberg tables. Engines can be replaced, and the organization keeps the keys and control.

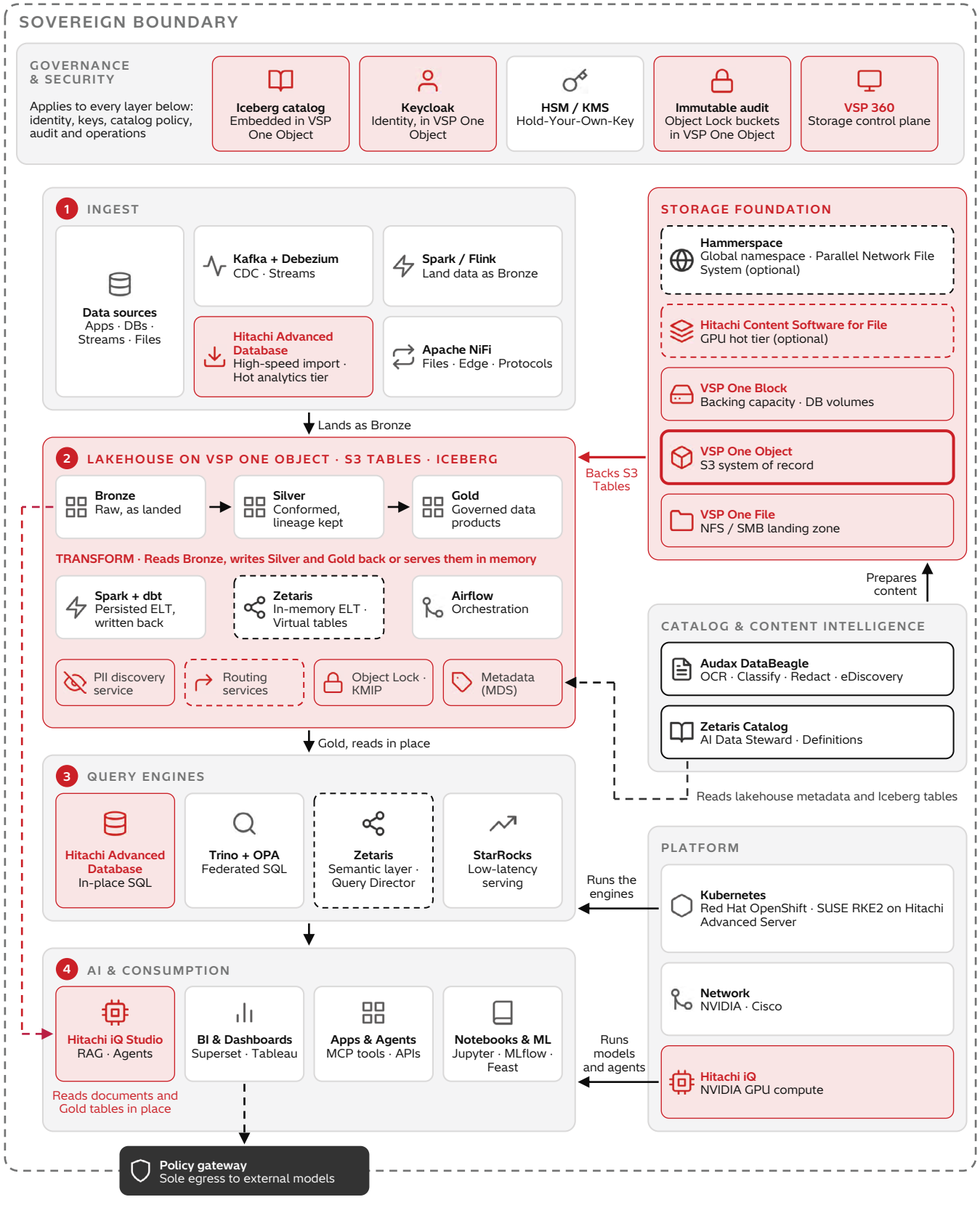


Figure 2. Sovereign data lakehouse reference architecture.

Figure 2 shows a representative implementation: one copy of data on VSP One Object, stored as open Apache Iceberg tables, so the engines above it can change while the organization and its chosen operator keep the keys and the control plane.

One Authoritative Copy

The lakehouse starts with one authoritative copy that every approved engine can use. VSP One Object provides that durable, scalable layer, with S3-compatible access, native S3 Tables and Apache Iceberg integration. This approach allows structured, semi-structured and unstructured data to stay in object storage while retaining table semantics such as schemas, snapshots and controlled evolution. Storage is separated from compute, and organizations choose the processing engine without relocating the data.

Data engineering and streaming tools then organize this information into a medallion architecture, with VSP One Object remaining the authoritative store:

- **Bronze data** preserves raw source information with acquisition time, source identity, classification and retention requirements.
- **Silver data** is standardized and validated, while preserving lineage back to source data.
- **Gold data** provides business-ready and AI-ready data products with agreed definitions, quality measures, access policies, lineage and conditions of use.

Analytics Without the Copies

Analytics should run where the data lives. Instead of copying data into a separate warehouse for every workload, Hitachi Advanced Database (HADB) provides SQL access to governed data on the VSP One platform for applications, reporting, notebooks and AI workflows.

Together, VSP One Object and HADB create a unified logical data plane across structured, semi-structured and unstructured data. Frameworks for large-scale transformation, streaming or machine learning remain optional, modular components.

The real advantage is reduced movement of sensitive data. The organization retains one governed source, one lineage model and one set of sovereignty controls while supporting different analytical and AI use cases. Assigning each data product to a data resilience zone carries its sovereignty, AI-use, protection and recovery profile from bronze to gold.

From SQL Analytics to AI-Ready Knowledge

A lakehouse becomes far more valuable when AI can safely consume it. Hitachi iQ Studio connects lakehouse tables with documents, files and other permitted sources, so models and agents retrieve approved information instead of relying on what they learned in pretraining.

HADB can provide precise SQL access to governed facts, while semantic retrieval can provide context from documents and unstructured information. An agent can combine both, provided that policy permits the operation.

This creates governed enterprise RAG. The model remains a reasoning layer rather than the system of record and receives approved content together with structured facts, provenance, user permissions, classification, geography, purpose and usage constraints.

Sovereign Use of Pretrained LLMs

With governed data in place, the question becomes how to bring models to it. Most organizations do not need to train a foundation model; pretrained large language models (LLMs) are a faster, more economical start. What matters is how they are introduced, governed and connected to enterprise data.

Organizations typically combine three deployment patterns, chosen by policies based on data classification, geography, purpose, operational dependency, model capability and evidence requirements.

Sovereignty does not require the largest model for every workload; larger models and GPU resources belong where they deliver measurable value.

Pattern 1: Deploy the Model Inside the Sovereign Environment

Where licensing permits, a pretrained open or commercial model can run on Hitachi iQ infrastructure inside the organization-controlled data center. Prompts, context, outputs, embeddings and logs stay within the boundary, and the enterprise controls version, runtime, access, updates and retirement.

Pattern 2: Connect to an External Model Through a Controlled Gateway

For selected workloads, organizations may use a managed model service through a policy-controlled gateway. Before information leaves the sovereign environment, the gateway can classify the request, mask prohibited fields, enforce geographic restrictions, limit retention and record the transaction, or block highly sensitive workloads entirely. A hybrid strategy should reflect that not all data carries the same sovereignty requirement.

Pattern 3: Ground Models Through RAG

Hitachi iQ Studio can connect the model to governed knowledge bases, lakehouse information, documents, databases and operational sources. The model receives only the context permitted for the user, agent, purpose, geography and workload.

This keeps proprietary data out of model weights, lets sources update without retraining, enforces permissions at retrieval time and ties answers to enterprise evidence. It also lets each jurisdiction use its own knowledge collections and keeps sensitive information out of prompts sent to external services.

Fine-Tune Selectively

Fine-tuning can be appropriate when an organization needs consistent domain terminology, format, classification behavior or task performance.

The training data should first be approved, classified, de-identified where necessary and connected to a documented legal or business purpose. The resulting adapter or fine-tuned model must be treated as a governed information asset because model weights can encode sensitive knowledge.

For many enterprise use cases, RAG, prompt control, tools and agent workflows should be considered before extensive fine-tuning. They provide greater transparency and allow data to remain separately governed.

Govern the Model Lifecycle

Every production model or agent should have an accountable owner, approved purpose, deployment location, version, permitted data classes and knowledge sources, tool and action permissions, evaluation criteria, risk classification, monitoring requirements, human-approval points, change history and retirement policy. Operational records should preserve prompts, retrieved context, sources, tools invoked, outputs, actions, evaluations and approvals as policy requires.

Hitachi iQ Studio provides orchestration model management, agent construction, guardrails, evaluation, monitoring and controlled deployment. The objective is to move from model experimentation to an operating discipline that can withstand regulatory, security and business scrutiny.



Governance as a Reusable Service

Every control described so far must adapt to where data and workloads live. Legal requirements, contractual obligations and risk tolerance vary by geography: A European service may require EU processing locations and AI risk controls. A Japanese service may mandate local privacy and transfer rules. And a public-sector environment may determine citizen data localization, cleared administrators and disconnected operation.

A sovereign platform packages these controls as reusable blueprints. Each workload is assigned to a data resilience zone along with its geography, owner, data classification, purpose, recovery objective and model-access policy. Each inherits the matching controls, including those for metadata, embeddings, prompts and agent memory, which can expose sensitive information even when source data stays local.

Industry blueprints build on that foundation by defining zone profiles. In financial services, zones for payment, customer and trading data carry transaction controls, model risk requirements, operational resilience, immutable retention and human-in-the-loop approval for consequential decisions. In healthcare, zones for clinical systems carry protected health information controls, consent, de-identification and recovery in clinical-service priority. Government and critical-infrastructure zones carry localization, sovereign key management and safety-critical continuity. Geography-specific overlays then apply local rules, such as the EU's GDPR, DORA or AI Act, where relevant.

Data resilience zones carry each blueprint's policy profile; VSP 360 applies and monitors it at the infrastructure level, VSP One implements the required data behavior and Hitachi iQ Studio governs the AI lifecycle. With approved blueprints in place, teams start each AI project from a sovereign foundation and focus on the business outcome.

From Sovereign Data to Autonomous Operations.

AI no longer just generates text and code. It already reasons, acts and controls physical systems, from substations to rail networks. That raises the stakes for the data behind it. A model drafting a report is a governance question, but a system adjusting a grid is a matter of safety, continuity and public trust.

The progression is already underway: sovereign data enables sovereign AI, which enables Physical AI and increasingly autonomous operations. Each step builds on the one before, so the data foundation determines how far and fast an organization can go.

Organizations should therefore build once rather than rebuild controls for every new AI requirement. A common sovereign data foundation keeps critical enterprise and operational data trusted, resilient, governed and under organizational control. It supports today's AI and the Physical AI and autonomous operations already being deployed.

That foundation turns control into competitive advantage. Those that build it now can put proprietary data to work, adopt new AI faster and adapt to shifts in technology, regulation and geopolitics without rearchitecting. In practice, grid operators can balance renewable supply and demand in real time. Manufacturers can run increasingly autonomous operations that correct faults before production stops. Transportation networks can adjust rail and traffic flows to demand. Each depends on trusted, governed operational data under the organization's control.

Delivering these outcomes takes expertise across data, AI and physical operations. Hitachi brings all three as one company, with more than 110 years in operational technology and decades in enterprise IT and digital engineering across energy, transportation, manufacturing and other mission-critical industries. Trusted by 86% of the Fortune Global 100, Hitachi understands both how to keep critical data under control and how to put it to work across the critical systems society depends on.

Conclusion: Reliable. Resilient. Ready.

Sovereign AI keeps organizations open to innovation by making the data foundation beneath it reliable, resilient and ready:

- **Reliable:** Critical data remains available and performant.
- **Resilient:** Critical data remains protected, recoverable, governed, portable, observable and under organizational control.
- **Ready:** Enterprise data is classified, contextualized, governed, accessible, verifiable and usable by approved analytics, models and agents.

The right architecture combines cloud-like agility, open data access, accelerated AI infrastructure, automated operations and policy-driven governance while preserving control over data, models, infrastructure and outcomes. Hitachi Vantara delivers it as a unified platform with integrated data resilience at its core.

The path forward starts with the seven questions. Hitachi Vantara works with organizations to assess control across each dimension, identify the services and data that make up their minimum viable company, build it into policy-driven governance and establish a sovereign data foundation and lakehouse before scaling AI. Each step produces evidence of control that leaders can show boards, auditors and regulators.

The result is a trusted digital foundation where governments, industries and enterprises can turn proprietary data into intelligence without surrendering ownership, independence or accountability.

Reliable. Resilient. Ready. Sovereign data. Sovereign AI. One trusted foundation.

Learn more about Hitachi Vantara's unified data lakehouse for the sovereign AI data center.

Learn More

About Hitachi Vantara

Hitachi Vantara, a Hitachi, Ltd. subsidiary, is the data foundation for innovation. We build resilient data storage and infrastructure the world's innovators rely on.

Hitachi Vantara

Corporate Headquarters
2535 Augustine Drive
Santa Clara, CA 95054 USA
hitachivantara.com | community.hitachivantara.com

Contact Information
USA: 1-800-446-0744
Global: 1-858-547-4526
hitachivantara.com/contact

© Hitachi Vantara LLC 2026. All Rights Reserved. All other trademarks, service marks and company names are properties of their respective owners.

Figure 2 trademarks: Iceberg = Apache Iceberg®, Kafka = Apache Kafka®, S3 = Amazon S3®, Keycloak®, Debezium™, Spark = Apache Spark®, Flink = Apache Flink®, Apache NiFi™, Zetaris®, Airflow = Apache Airflow®, Trinio™, OPA™, StarRocks™, Audax DataBeagle = Audax Labs DataBeagle™, Kubernetes®, Red Hat®, OpenShift®, SUSE®, NVIDIA®, Cisco®. Note that Hitachi Content Software for File is Powered by WEKA.

HV-BTD-WP-Sovereign-AI-Starts-with-Resilient-Data-14Oct26-A