

zadara

CLOUD TECHNICAL OVERVIEW

Zadara delivers a complete range of fully managed cloud services, including compute, networking, and storage solutions, tailored to support the needs of today's enterprises and service providers.

While this document focuses on Zadara's edge architecture, it is important to highlight the core components that enable multi-edge cloud management within Zadara's suite of services.

Index

1. Edge Architecture

2. Compute Services

- Virtual Machines (VMs)
- Autoscaling Groups
- · Load Balancers
- Kubernetes Integration
- GPU Cloud

3. Storage Services

- · Block Storage
- · File Storage
- · Object Storage

4. Networking Services

- Virtual Private Cloud (VPC)
- NAT Gateways (NGW)
- DNS Service
- Security Groups

5. Architecture Abstraction

- Local Provisioning Portal
- Command Center

6. zStorage

 zStorage resource allocation in a multi-tenant environment

7. zCompute

- · zCompute System Features
- Elastic Compute / Instance Types
- Virtual Private Clouds (VPCs)
- Load Balancers
- Auto-Scaling Groups
- Kubernetes and Docker Support
- zCompute EBS

8. zStorage System Features

- Scalability
- Low-Cost, Low-Latency Transport
- Storage Array Controllers
- · Object Storage
- Security
- · Standards Compliance
- · Fully Managed, Upgrades Included
- Replicate to Any Location
- Multi-Zone High Availability for VPSA Storage Array and Object Storage
- Cloud Hydration Service
- · Backup to Object Storage Service
- Microsoft VSS
- Docker Containers
- Volume Migration
- Remote Clone

9. zNetworking System Features

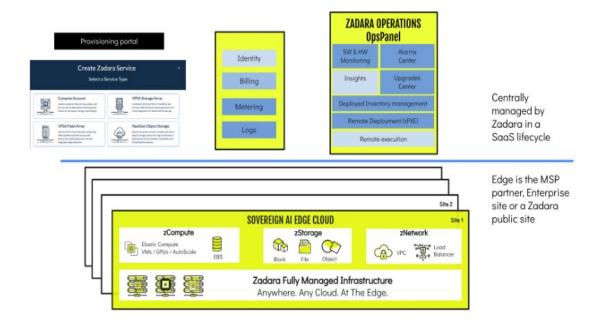
- Virtual Networking
- VLANs and Virtual Networks
- · Security Groups
- Load Balancing

10. Zadara's zCloud - Private AI/LLM with One-Touch Deployment

- · Key Features
- Use Cases

11. Federated Edge Program

Zadara offers a comprehensive suite of fully managed cloud services, encompassing compute, networking, and storage solutions designed to meet the demands of modern enterprises and service providers.

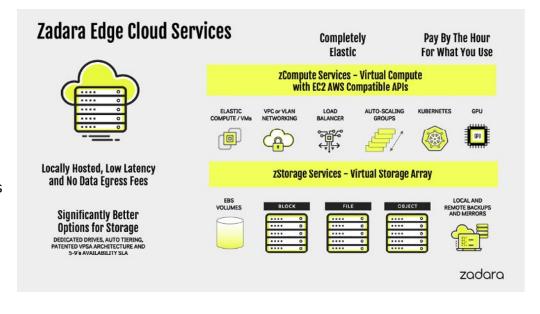


This document describes Zadara's edge architecture, but it is worth mentioning the central components that allow multi-edge cloud management as part of the Zadara services:

- 1. The Zadara public provisioning portal, allows you to provision compute accounts, Storage Arrays, and Object Storage services within locations that have been authorized for this use.
- 2. The Zadara Operations panel allows the Zadara Operations and Support teams to operate the cloud as a managed service.

Edge Architecture

Zadara's edge architecture is engineered for high availability, scalability, and flexibility, ensuring optimal performance across various workloads.



Compute Services

Zadara's compute services, known as zCompute, offer a robust platform for developing, deploying, and managing applications. Key features include:

- Virtual Machines (VMs): Provision and manage x86-based VMs with intelligent scheduling and load mitigation to ensure optimal performance.
- Autoscaling Groups: Automatically adjust the number of VM instances in response to workload demands.
- Load Balancers: zCompute utilizes load balancers to manage incoming application traffic by distributing it across multiple Virtual Machine (VM) instances. Zadara supports both Application Load Balancers (ALBs) and Network Load Balancers (NLBs).
- Kubernetes Integration: Support for container orchestration, facilitating the deployment and management of containerized applications.
- GPU Cloud: Access to GPU resources for compute-intensive tasks such as AI and machine learning workloads.
- Flexible API's: Either us extended Zadara API or use EC2 compatibility APIs

These services are designed to provide secure, flexible, and scalable compute resources, whether deployed in colocation facilities, customer (private) data centers, or public clouds.

Storage Services

Zadara delivers enterprise-grade storage solutions supporting various data types and access protocols:

- Block Storage: Offers comprehensive elastice block storage capabilities, reducing latency and increasing throughput. This service is based on Zadara's Virtual Private Storage Array (VPSA) technology and is fully compatible with AWS EBS API for seamless application integration. Block services may be also consumed as a dedicated network accessible VPSA allowing sharing block storage between different compute platforms.
- File Storage: Provides scalable file shares via NFS or SMB protocols, allowing file access to all associated VM instances. This service is based on the Zadara VPSA NAS offering.
- Object Storage: Supports both private and public object storage services, accessible via AWS S3 API or Swift interface. Private Object Storage is directly connected to your Virtual Private Cloud (VPC), while public Object Storage is accessible via a public IP.

These storage services are designed to be scalable, high-performing, and compatible with a wide range of applications and workloads.

Networking Services

Zadara's networking services are designed to provide secure and efficient connectivity within the cloud environment:

- Virtual Private Cloud (VPC): Enables the creation of isolated network environments within the Zadara cloud.
- NAT Gateways (NGW): Allow instances without public IP addresses to access the internet securely, providing an additional layer of protection for workloads.



- DNS Service: Offers an easy-to-manage DNS solution for private networks, seamlessly integrated with cloud resources.
- Security Groups: Act as virtual firewalls, controlling inbound and outbound traffic to instances based on defined rules.

These networking services ensure secure, efficient, and flexible connectivity for applications and services within the 7adara cloud.

Account and Identity Services

Zadara's compute services are designed with multiple tenancy in mind

- Tenant accounts: Allow MSP to add tenants to the system
- Tenant project: Allow Tenant admin to further partition accounts into separate sub-tenants regular users (Members) log in the context of a project.
- Project switcher: Allow members to have permissions on multiple projects and easily switch between them
- Policies: Fine grained policies that define what a user can and can't do in the system. Different policies are available for Zadara API vs AWS compatibility APIs

Architecture Abstraction

The minimal set of services provided on the edge are the storage services known in this document as zStorage. These services are topped with compute and network services referenced in this document as zCompute.

The above dictated that the core of Zadara's cloud architecture are at least two storage nodes (SN). The Zadara's cloud management is running on the storage nodes and is facilitated through two primary software components:

- Local Provisioning Portal: Enables users to manage and allocate cloud resources efficiently. This portal may be exposed to the tenants and allows them to provision the main 4 services:
 - zCompute account
 - Virtual Private Storage Array (VPSA)
 - Private Object Storage (NGOS)
 - · Object Storage account
- Command Center: Provides centralized management and monitoring capabilities for the entire cloud infrastructure. This portal is administrator only.

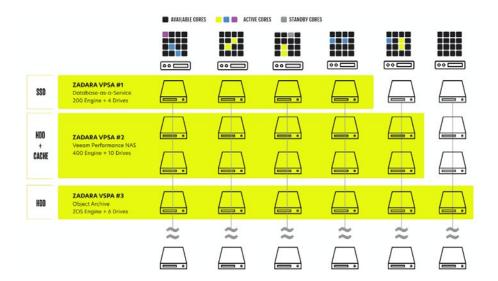
zStorage

zStorage may grow in node count where the system aggregates all the resources of these nodes into a larger pool of storage devices and computation resources that provide the storage services basically it's a multi-tenant software-defined storage.



zStorage resource allocation in a multi-tenant environment.

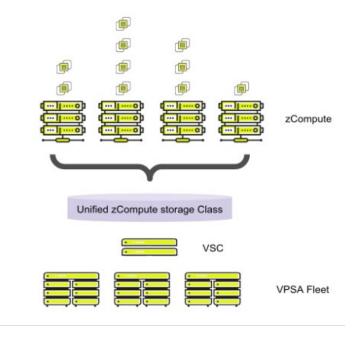
The structure of both object storage and VPSAs is the same. Each is composed of virtual controllers running isolated on dedicated cores on the storage nodes, hardware based network isolation with SR-IOV passthrough to the controller, and a bunch of dedicated storage drives. Effectively, tenants do not share resources as can be seen in the diagram below.



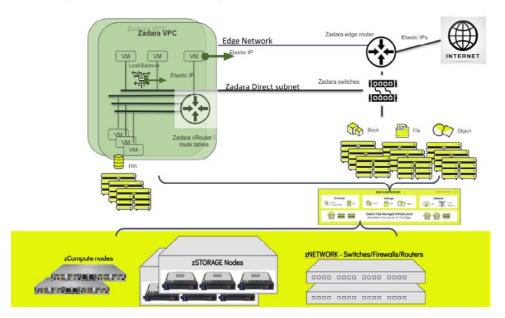
zCompute

zCompute services require an additional set of nodes called compute nodes (CN). Same with the storage nodes, these nodes aggregate all the compute resources from all the participating compute nodes into a multi-tenant compute service. The elastic volumes for the virtual machines in the system are provided by zStorage in an abstracted manner to the user.

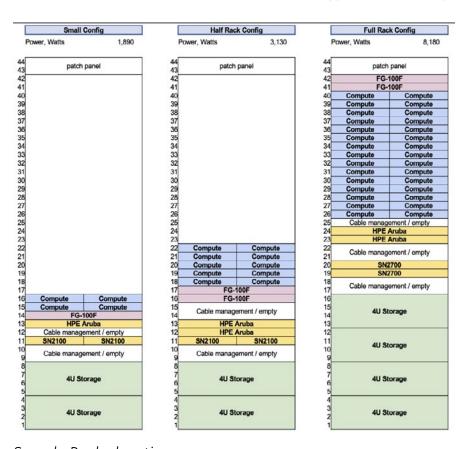
At a high level, zCompute and zStorage are stacked together to provide all the cloud services. Both zCompute and zStorage can be scaled individually to support specific user requirements.



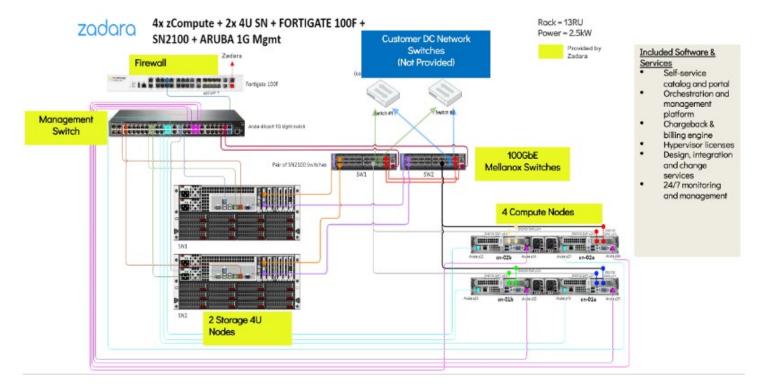
A typical edge location diagram can be seen below. Edge locations that are connected to enterprise facilities or work in a hybrid cloud mode may differ.



Here are a few examples of the physical view of the system in your data center. Starting with a minimal installation and expanding to a full rack. Many edge locations span multiple racks, multiple compute nodes models are available to offer a choice of CPU types and GPU options.



Sample Rack elevations



Zadara Cloud - Base Infrastructure configuration

zCompute System Features

Zadara's zCompute platform offers a range of features tailored to meet diverse computational needs, including various instance types and robust support for containerized environments like Kubernetes and Docker.

Elastic Compute / Instance Types

zCompute provides a selection of predefined and customizable instance types to accommodate workloads of varying sizes and performance requirements. These range from standard instances suitable for small applications (2vCPU, 4GB RAM) to premium instances (96vCPU, 800GB RAM) designed for performance-intensive and memory-demanding tasks. Users can select from predefined Virtual Machine (VM) images (Windows or Open Source O/S's) or customize their own as needed. zCompute also supports GPU instances that are directly connected with PCI-passthrough to Nvidia L4 or L40s GPUs. The environment also supports directly connected NVMe resources. All of the features are customizable and are architected to meet specific customer requirements.

Zadara's zCompute platform offers comprehensive networking and scaling features to ensure secure, efficient, and scalable cloud environments. Here's an overview of their Virtual Private Clouds (VPCs), Load Balancers, and Auto-Scaling Groups:

Virtual Private Clouds (VPCs)

Zadara enables the creation of isolated network environments called Virtual Private Clouds. Within a VPC, you have full control over your networking setup, including IP address ranges, subnets, route tables, network gateways, and security groups. This isolation ensures that your resources operate securely and efficiently within a defined virtual network. The Tenant administrator can further divide the environment into specific Projects, which allows further isolation between VPCs. This function allows a single tenant the ability to manage production environments and other essential workloads like development to co-exist without interference and to be easily managed.

Load Balancers

To distribute incoming application traffic across multiple instances, Zadara offers both Application Load Balancers (ALBs) and Network Load Balancers (NLBs):

- Application Load Balancer (ALB): Operates at the application layer (Layer 7) and is ideal for HTTP and HTTPS traffic. It provides advanced routing features based on content, making it suitable for web applications.
- Network Load Balancer (NLB): Functions at the transport layer (Layer 4) and is designed for high-performance, low-latency traffic handling. It is suitable for TCP traffic and can manage millions of requests per second.

These load balancers automatically distribute incoming traffic across multiple targets, such as instances or IP addresses, enhancing fault tolerance and ensuring high availability.

Auto-Scaling Groups

Zadara's Auto-Scaling Groups provide an automated mechanism to manage the number of VM instances in response to application load or failure conditions. Key features include:

- Dynamic Scaling: Automatically adjusts the number of instances based on predefined criteria, such as CPU utilization or network traffic, ensuring optimal performance during varying load conditions.
- Health Monitoring: Continuously monitors instance health, replacing any that fail to maintain application availability.
- Cost Efficiency: By scaling in during low-demand periods, it helps optimize resource utilization and reduce costs.

Auto-Scaling Groups can be associated with load balancer target groups to ensure that traffic is efficiently distributed among the available instances.

Kubernetes and Docker Support

Zadara offers comprehensive support for container orchestration platforms, particularly Kubernetes and Docker:

- Kubernetes Integration: Zadara's cloud is AWS-compatible, making it suitable for running EKS-D (Amazon's Elastic Kubernetes Service - Distro) workflows outside of AWS. This compatibility allows users to utilize AWS-oriented utilities such as the AWS Cloud Controller Manager (CCM), the AWS Elastic Block Store (EBS) Container Storage Interface (CSI) driver, and the AWS Load Balancer Controller. Additionally, Kubernetes-native tools like the Kubernetes Cluster Autoscaler, which leverages AWS Auto-Scaling Groups API, are supported.
- Docker Integration: Zadara integrates Docker container technology into its virtual controllers directly on VPSA with CSI, enabling users to run processing tasks directly within the storage environment. This setup provides direct, low-latency access to SSDs and HDDs, resulting in high-throughput input/output operations per second (IOPS) for Docker containers running within the storage.

zCompute EBS

The Zadara Storage platform Virtual Private Storage Array (VPSA) provides the underlying storage for zCompute EBS Volumes

- Storage Classes using Virtual Storage Controller (VSC)
- Support for scale out and differing performance needs
- SSD
- SSD Tiered
- HDD Only

^{*}Graphic on next page





zStorage System Features

Zadara transforms storage-related costs from a variable mix of equipment and management expenses to a predictable, on-demand, pay-per-use, elastic service that greatly simplifies planning, streamlines budgeting, and improves return on investment (ROI).

Scalability

Zadara is delivered as a collection of storage nodes: standard servers running Linux, with storage options that include hard drives, SSDs, and NVMe flash. Connectivity is built with Mellanox 100Gb Ethernet switches and NICs. Based on standard building blocks (storage nodes), Zadara allows you to start with as few as two storage nodes and scale to as many as 100 storage nodes. This software-defined storage scales in every dimension: 1) CPU resources for the storage controllers; 2) drive resources for raw storage; and 3) transport for interconnectivity. All storage services and tiers can be delivered from the same cluster of storage nodes. A single VPSA can support up to 200TiB's of capacity with snapshots, mirroring, and backup to Object Storage functions built-in. Additionally, a VPSA can be scaled to 500TiB for large data repositories in which capacity is the primary function with data protection offloaded to other applications.

Low-Cost, Low-Latency Transport

Using redundant 100Gb interconnects and the iSCSI extensions for RDMA (iSER), we built a low-cost, low-latency network fabric with sufficient bandwidth for the most demanding workloads. All Zadara storage solutions use iSER as the protocol for the interconnect and, optionally, for host connectivity. Zadara supports Fibre Channel, iSCSI, iSER, NFS, CIFS, S3, and Swift host connectivity.

Storage Array Controllers

Zadara Storage Nodes are leveraging KVM to run and manage virtual controllers (Zadara Engines), which are virtual machines running on the storage nodes. Each controller is assigned dedicated CPU cores on a storage node plus dedicated disks across several nodes, so there is no over-subscription of these resources. The combination of these dedicated processor cores, memory, and disks comprises a Virtual Private Storage Array. The workloads in each VPSA are completely isolated from each other and each can deliver both file and block storage based on hard disk drives or solid state drives (SAS or NVMe). Zadara storage clouds can run thousands of virtual arrays simultaneously.

Object Storage

In addition to storage arrays for block and file storage, Zadara offers the industry's only fully managed private enterprise cloud object storage solution. Zadara object storage utilizes dedicated storage resources, is compatible with public object storage, and supports the same interface (Amazon S3 RESTful API, Swift API), while allowing customers to keep the data on-premises or in the cloud. This provides both higher levels of security, but also provides consistent levels of performance. Our next gen (NGOS) has been re-written with performance in mind, enabling Strong Consistency throughout the data put and get process. The storage and proxy controllers have larger processor and RAM resources along with larger cache volumes to further enhance performance over the previous generation. The Object Storage can scale from 300TB to 10's of PB's in a single system.

Security

Security is integral to Zadara, starting with workload isolation. Drives are mapped to a single array at a time, and if an array ever vacates a drive, it is scrubbed before being made available again. Administrator accounts are never shared between arrays, and the drives in each array can be encrypted with their own, customer-managed key. The password to the key is never stored durably and is never shared outside the array. Arrays support IPSec for encryption of data in transit between the array and clients, and array-array replication traffic is always encrypted. Clients are mapped to individual arrays, and a client can access only the LUNs or shares in arrays to which it has been granted access.

Standards Compliance

Zadara conducts ongoing security testing of its clouds and maintains security certifications such as ISO 27001, SOC 2 Type 2, and HIPAA. GDPR compliance is a shared responsibility. We offer a wide set of controls to help you maintain GDPR compliance. For more details, please visit the Zadara website. There you will find information about the security measures we have in place.

Fully Managed, Upgrades Included

Zadara is a fully managed storage service with 24/7 monitoring and support. And it doesn't end there. Zadara automatically upgrades hardware without any application impact. For Zadara running in the public cloud, this happens seamlessly and without the need for any work on your part. When a Zadara storage node running on your premises has reached its end of life, we ship you a new node. The data from the EOL node is copied — online and without a performance impact — to the new storage node. The drives in the EOL node are software shredded, and then the node is returned to Zadara. All of these tasks are managed and monitored by the Zadara NOC, so the only thing you need to do is swap the old gear for the new gear in the rack.

Replicate to Any Location

Zadara Enterprise Storage-as-a-Service provides simple, secure, and powerful on-demand data mobility services between virtual private arrays, regardless of location. Use Zadara Remote Mirror for disaster recovery and business continuity applications. Deploy Zadara Remote Clone for automated migrations, offline processing tasks, and data distribution. Either way, you get one-to-one or one-to-many replication service across racks, data centers, clouds, and continents, without extra license or service level expenses.

Multi-Zone High Availability for VPSA Storage Array and Object Storage

In certain regions, and in all on-premises deployments, Zadara supports Multi-Zone High Availability. When a Zadara Storage Cloud is deployed in a Multi-Zone High Availability model, the deployment is split into two Protection Zones. Each Protection Zone is deployed in a different location or Availability Zone located in the same metropolitan area — generally within 2ms of each other (max. 5ms and 20ms for Object Storage). VPSA Storage Array resources are split across two Protection

Zones, ensuring data availability is maintained, even in the case of an Availability Zone outage. For VPSA Object Storage, the system will create a complete Data Policy according to the one selected during the creation, ensuring the solution can sustain a complete region loss.

Cloud Hydration Service

Zadara's cloud hydration enables you to adopt cloud computing using a cost-effective and practical way of moving corporate data into the cloud. Cloud Hydration allows migration to the cloud of both online production environments and data that does not need to be continuously online. Cloud Hydration can be targeted to any storage medium from any vendor or cloud storage provider.

Backup to Object Storage Service

Using a Zadara storage array, you can create an automatic, snapshot-based, continuous, incremental backup to low-cost, object storage. This eliminates the need for host-based backup software and provides a simple, easy-to-use, high-performance, block-based backup. And data backed up by Zadara can be restored back to the original volume or anywhere else you choose.

Microsoft VSS

Microsoft Volume Shadow Copy Service (VSS) enables online, point-in-time snapshots of Microsoft SQL Server, Exchange, SharePoint, and other Windows-based enterprise applications in Zadara storage arrays. You can replicate these snapshots to any of Zadara's dozens of public storage clouds around the world, including AWS, Azure, Google Cloud Platform, and others. Zadara's Snapshots mechanism is very efficient, in terms of capacity and performance. Block volumes and file shares are supported. Snapshot policies define the Snapshots lifecycle via the enforcement of creation and deletion policies. Snapshot Policies are "global" entities, and you can apply instances of the policies to one or more volumes. Snapshots can be taken as frequently as every minute. There is no limit to the number of snapshots you may keep per volume.

Docker Containers

Zadara storage arrays incorporate Docker container technology into the Zadara Engines, a pair of virtual controllers. The Zadara Engines have direct, low-latency access to your SSDs and HDDs, and therefore provide high throughput IOPS to the Docker container running within. Like the Zadara Engines, the Container has dedicated CPU and memory, which can be increased and decreased on the fly, non-disruptively. Uniquely, Zadara offers Docker high availability, thanks to Zadara's dualengine architecture with auto-failover.

Volume Migration

Volume migration allows you to transparently move your Zadara storage array data and snapshots between different pools without affecting servers and workstations mounted to the volume or share. The most common case is to move volumes from a high-performance storage pool to a lower-tier pool when performance requirements are no longer imposed on the data set. Other uses, such as moving between pools with different durability levels, can increase storage efficiency for lower cost per GB or increase performance using the same storage media.

Remote Clone

Remote Clone makes a given snapshot of a source volume instantly available (before data is copied) as a volume on another VPSA, in the same cloud or in a different cloud over any distance. Unlike Mirroring that might take a long time to replicate the data (depending on capacity and the link bandwidth), the cloned new volume is available immediately.



zNetworking System Features

Virtual Networking

Zadara's virtual networking layer follow the AWS VPC concepts. Itenables the creation of isolated network environments, allowing users to define internal private networks, public networks that allow the choice to assign fixed elastic IPs or usage of NAT gateway for the virtual machines (VMs). This setup ensures that resources operate securely within defined VPC. VPC are isolated from each other to the level that IP address ranges may overlap between VPCs.

VLANs and Virtual Networks

In addition to the VPC concepts. The platform supports the allocation of Virtual Local Area Networks (VLANs) and virtual networks to cloud tenants. VLANs are used to segment network traffic, enhancing security and performance, while virtual networks define a set of available IP addresses within a specific network segment. This combination allows for tailored network configurations to meet diverse customer demands.

Security Groups

Security groups act as virtual firewalls, controlling inbound and outbound traffic to instances based on defined rules. This feature provides an additional layer of security, ensuring that only authorized traffic can access specific resources within the cloud environment.

Load Balancing

Zadara's scale-out architecture for cloud-native applications includes load balancer services, enabling customers to provision managed load balancers for fault tolerance and application scalability. This ensures efficient distribution of incoming traffic across multiple instances, enhancing application performance and reliability.

Zadara's zCloud - Private AI/LLM with One-Touch Deployment and RAG for Real-Time Insights

Zadara's zCloud platform is the perfect solution to support private AI and Large Language Model (LLM) deployments on-premises, allowing enterprises to securely integrate corporate data while ensuring low-latency real-time inference. With built-in support for Terraform-based one-touch deployment utilizing Kubernetes and Docker, organizations can rapidly provision AI infrastructure without complex configurations.

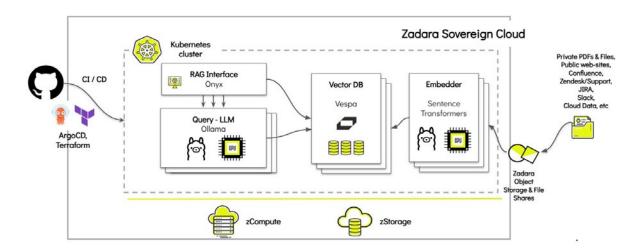
Key Features:

- One-Touch AI Deployment with Terraform: Automate the entire AI/LLM infrastructure setup—including compute, storage, and networking-using Terraform for fast, repeatable deployments.
- Seamless Integration with Vector Databases: Pinecone, Weaviate, or ChromaDB support ensures high-performance storage and retrieval of embeddings, enabling efficient similarity search and context-aware responses.
- Retrieval-Augmented Generation (RAG) for Enhanced Insights: AI models can dynamically retrieve relevant corporate data from vector databases, ensuring accurate, real-time responses to user queries. This reduces hallucinations and improves model reliability.
- On-Prem Deployment for Data Security & Sovereignty: Keep sensitive corporate data within private infrastructure, fully compliant with regulatory requirements while leveraging Al-powered analytics.



- Real-Time Inference for Low-Latency AI Applications: AI/LLM workloads are co-located with corporate data for minimal latency, allowing instant access to business-critical insights.
- Scalable, Fully Managed Infrastructure: zCloud's AI platform scales compute, storage, and network resources on demand, ensuring cost-efficient scaling based on workload demands.

Use Cases:



- Enterprise Chatbots & Assistants Private AI models trained on internal documents, HR policies, and knowledge bases.
- Fraud Detection & Compliance Real-time analysis of transactional data for anomaly detection.
- Predictive Analytics Al-driven insights for demand forecasting and business intelligence.
- Medical & Legal AI Secure, compliance-driven models trained on sensitive case data.

With Zadara's private AI cloud, enterprises can deploy, scale, and optimize AI-powered solutions faster than ever—all while maintaining full control over corporate data privacy.

Federated Edge Program

Zadara's Federated Edge program introduces a fully managed, distributed cloud architecture designed to enable hosting providers and Managed Service Providers (MSPs) to offer on-demand edge computing services without significant upfront investments. This program allows MSPs to:

- Create New Revenue Streams: Offer global IT services to customers without the need to build new data centers or deploy in new regions.
- Deliver Global Products and Services: Leverage a network of hosting service provider locations to host location-sensitive workloads, optimizing for latency, price, performance, and security/compliance.
- Compete with Public Cloud Providers: Provide features comparable to major public clouds, with the advantage of localized services through Federated Edge Zones in cities worldwide.
- Address Data Sovereignty Concerns: Offer customers control over the physical location of their data, easing concerns around data sovereignty and compliance.

This program empowers MSPs to expand their service offerings and compete effectively in the evolving cloud services market.

In summary, Zadara's cloud architecture and offerings provide a comprehensive, flexible, and scalable platform designed to meet the diverse needs of modern enterprises and service providers. Their fully managed services ensure high availability, security, and performance across compute, storage, and networking resources.

Zadara Alliance Partners

Our technology partner ecosystems aim to achieve a win-win situation for all parties involved. Some key objectives are:

Revenue Acceleration:

 Tailored industry solutions drive adoption and infrastructure consumption across compute, storage, and networking layers.

Vertical Expansion Strategy:

 GTM alignment on real-world use cases (e.g., FinTech, Media, Healthcare) accelerates customer onboarding and repeatable growth.

Innovation with Purpose:

· Partners co-develop next-gen capabilities addressing vertical pain points—transforming complexity into competitive advantage.

Customer-Centric Solutions:

· End-to-end offerings deliver seamless integration and better outcomes—boosting satisfaction and retention.

Market Penetration & Expansion:

 Validated use cases become case studies for entry into adjacent industries, shortening sales cycles and increasing TAM.

Operational Efficiency:

Shared frameworks and automation enable partners to scale while optimizing cost structures and delivery models.

Partner Ecosystem Stack

