This PDF is generated from authoritative online content, and is provided for convenience only. This PDF cannot be used for legal purposes. For authoritative understanding of what is and is not supported, always use the online content. To copy code samples, always use the online content.

Architecture

2/25/2026

Contents

Understand the architecture and components of Genesys Engage cloud private edition; the supported third-party back-end services; and how they all work together in both single- and multi-region deployments.

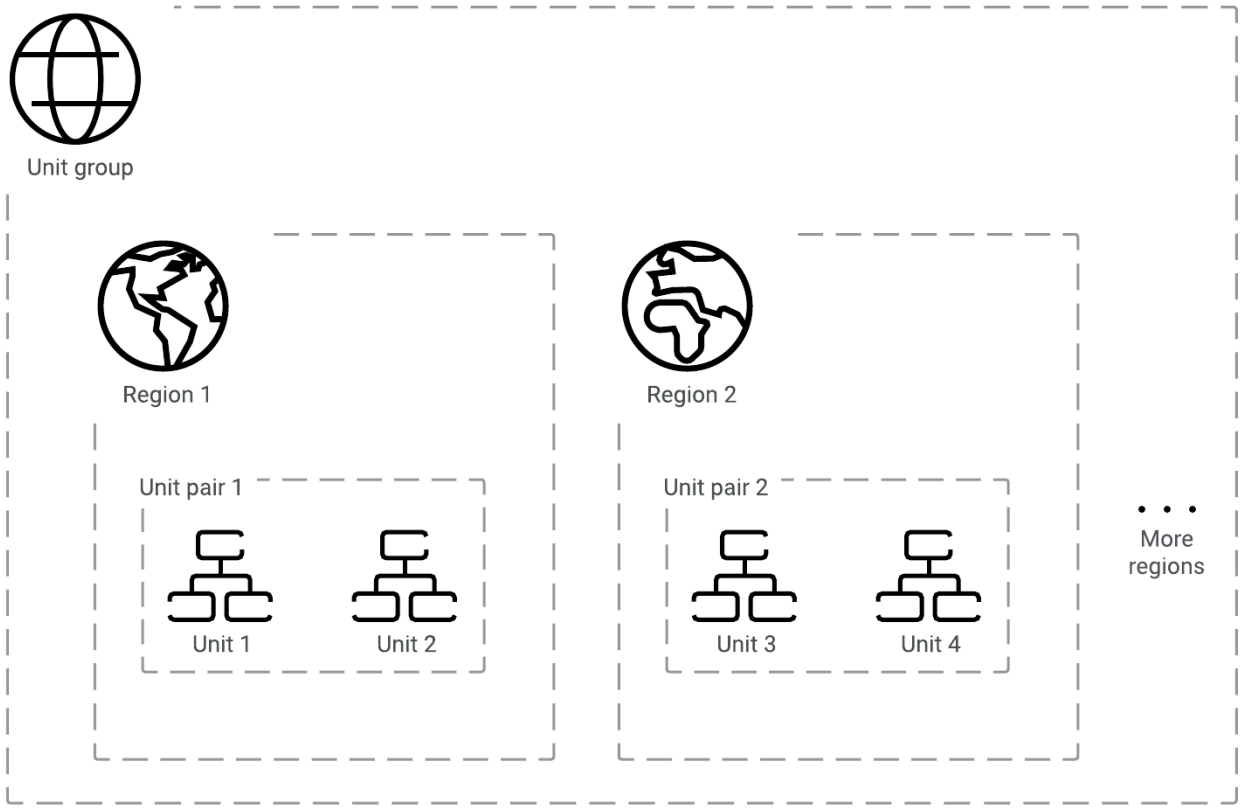
As mentioned in the About page, Genesys Engage cloud private edition gives you the flexibility to deploy your contact center on a public cloud or a private one—and even on bare metal servers that reside within your corporate data center.

Platform and network

Platform

The basic architecture for private edition involves three levels:

- A **unit** consists of all of the Genesys Engage and third-party services and resources required to create a single instance of Genesys Engage cloud private edition. This instance is hosted within a single region or data center.
- A **unit group** brings together a network of units to create a global platform for tenants that covers all geographical regions
- A **unit pair** consists of two units that are part of a unit group and that are both located within a specific geographical region



The following definitions describe important features of the private edition architecture:

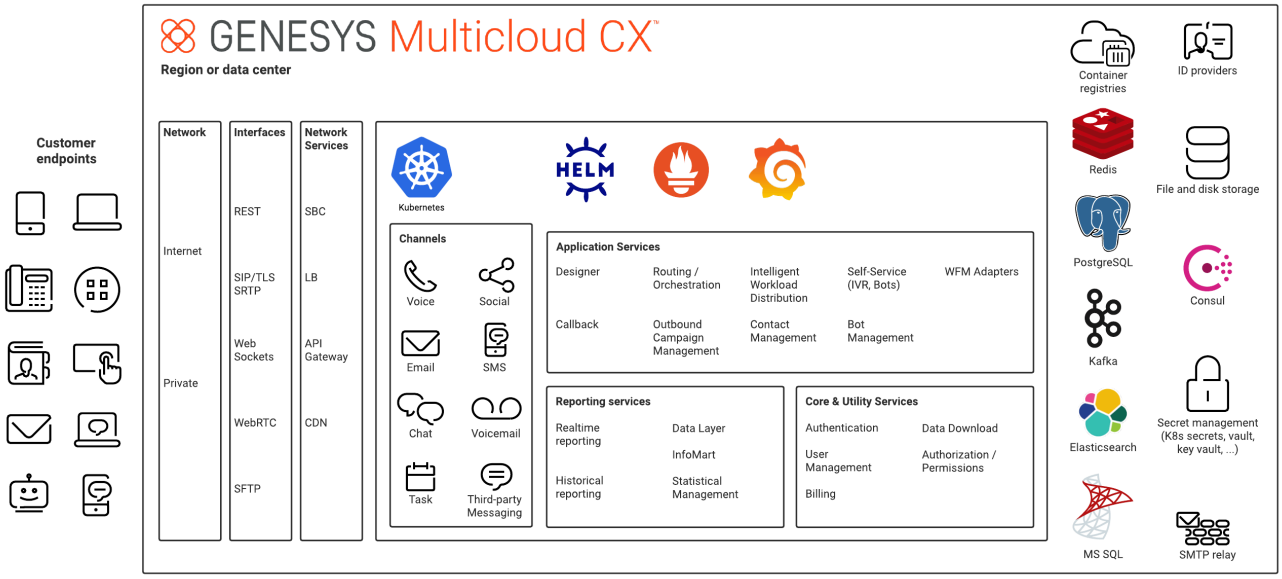
- **Region**—A set of isolated and physically separated Availability Zones deployed within a latency-defined perimeter and connected through a dedicated low-latency network within a specific geographical area.
- **Data center**—A building, a dedicated space within a building, or a group of buildings used to house computer systems and associated components, such as telecommunications and storage systems.
- **Availability Zone (AZ)**—A discrete location within a region that is designed to operate independently from the other Availability Zones in that region. Because of this separation, any given Availability Zone is unlikely to be affected by failures in other Availability Zones.
- **Tenant**—A business entity that has common goals and procedures, and occupies part or all of a contact center. Tenants that share a contact center could be different businesses, or different divisions within the same business.
- **Multi-tenancy**—The partitioning capacity for a platform to host and manage tenants. Each tenant is configured individually and separately.

Here is a more in-depth description of the characteristics of the three levels of the private edition architecture:

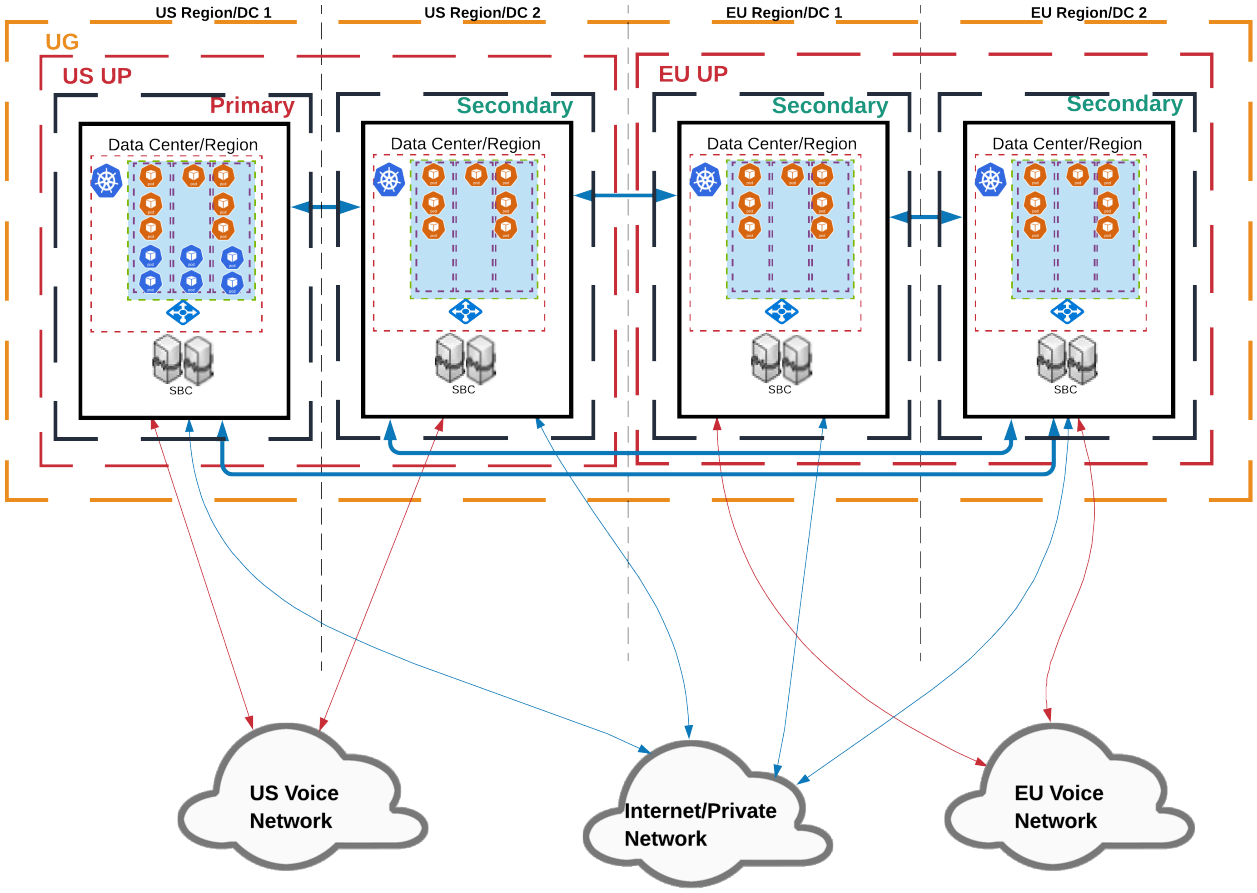
- **Units** can either be dedicated to a specific tenant or used for multiple tenants. The unit and its services and resources can be distributed across Availability Zones if the environment has them.

A unit is composed of the following:

- Network access services (load balancers, firewalls, SBC, and so on)
- A Kubernetes cluster with all of the private edition service pods
- Third-party services (Postgres, Redis, Consul, Kafka, and so on)



- **Unit pairs** provide the following capabilities. **Note:** Unit pairs are only supported by voice-related services at this time.
  - Redundancy within a geographical region. This geo-redundancy is built into the private edition services.
  - Tenants can be distributed across the two units to help reduce the blast area in case of a major failure
- **Unit groups** interconnect their constituent units by means of a network peering solution, and all inter-region traffic uses either your network connectivity or the network connectivity of your cloud provider. Each group contains a primary unit in each region. This primary regional unit hosts all of the private edition services, while the secondary regional unit hosts only a subset of private edition services. A unit group must contain at least one unit pair. If you add a new geographical region, then you must add a unit pair to the unit group in that geographical region.



Multiple regions and data centers

The platform supports deployment across multiple regions and data centers. This capability provides extra availability for the voice-related services, with a global view.

- **Call routing and processing**—The ability to distribute call processing across regions. Also, to centrally create and distribute Designer applications across regions.
- **Agent availability**—The ability to have a call processed by agents from any region
- **Data sovereignty**—The ability to contain the data (recordings, and so on) and processing of the call within the region in which the call originated
- **Reporting (Real-time and Historical)**—The ability to provide a global view across all regions
- **Tenant provisioning**—The ability to centrally provision the contact center across multiple regions
- **Callback**—The ability to use a central service to provide in-queue callback across regions

Subnets

Subnets are your responsibility: you must create a subnet for the Kubernetes cluster to accommodate the Genesys Engage services.

Network access

Content has been moved. Need a summary sentence and link to [[User:Jose.druker@genesys.com//PEGuide/NetworkOverview|]]

Supported services

Genesys Engage cloud private edition supports the services listed on the Genesys Engage services list.

Software requirements

Genesys Engage cloud private edition requires the software and versions listed on the software requirements page. Note that you are responsible for installing and deploying the appropriate third-party software in a way that best suits your requirements and the requirements of the Genesys Engage services.

Kubernetes clusters

All Genesys Engage services must run in Kubernetes. Required third-party services can be managed either outside Kubernetes or within Kubernetes. Kubernetes is responsible for managing the running of services, such as monitor restart, and so on.

Deployment

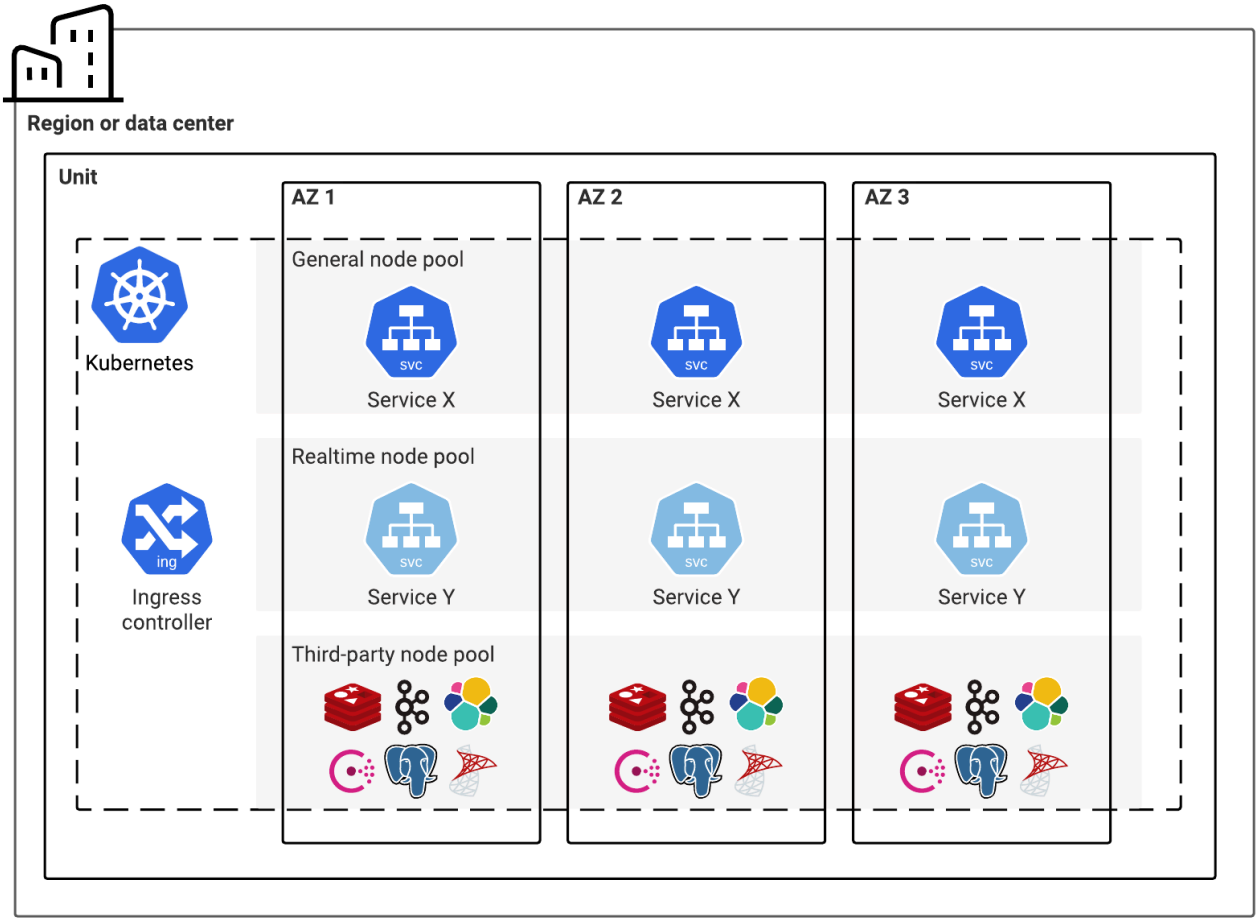
Genesys currently recommends that you use node pools to deploy Kubernetes for the Genesys Engage services that are hosted within each unit.

Richard, note that you can get rid of one level of bullets here.

- **Node pools**—Genesys recommends that you use the following node pools. Our Helm charts include overrides for the nodeSelector attribute. Use these overrides to assign a service to the appropriate node pool.
  - **General node pool**—This node pool is where most of the Genesys Engage services are deployed. ~~This type of pool uses general-purpose compute instances with Premium SSD, which provides~~ Use general-purpose compute instances with high-performance Solid State Drive (SSD) storage that provides a 4:1 ratio of memory GB to vCPU ("Premium SSD").
  - Real-time **node pool**—This node pool is for stateful voice services that require a drain time of 60

minutes or longer to maintain active voice sessions. ~~It uses~~ Use general-purpose nodes with Premium SSD.

- **Third-party service node pool (optional)**—~~This node pool is only needed~~ You need this node pool only if you are going to deploy data stores and other third-party services ~~in Kubernetes, —~~ such as Redis, Kafka, Postgres or Elasticsearch — within Kubernetes. These services generally need locally optimized storage~~. and will use the~~ Use storage-optimized nodes with directly attached NVMe and Premium SSD, which provide an 8:1 ratio of memory GB to vCPU.



Networking

For more information, see Network Settings.

Service priorities

For more information, see Service Priorities.

Autoscaling

Most services scale their Kubernetes pods by using the Horizontal Pod Autoscaler. However, this tool can only use CPU or memory metrics from the Kubernetes Metric Server in the HorizontalPodAutoscaler Object. Private edition also works with the Kubernetes cluster scaler.

Genesys uses the third-party KEDA open-source autoscaler for Genesys Engage services that require custom metrics from Prometheus. Use the included Helm override attributes to adjust the defaults for each service.

You must perform your own scaling operations on the Kubernetes control plane. The operational requirements of this scaling depend on the size of your contact center. For large installations, you might need to deploy multiple clusters and distribute the Genesys Engage services across them.

ConfigMaps

Private edition uses ConfigMaps to pass variables and data to the deployed services. This allows each service to be separate from its configuration data, which is a factor in making each service immutable. Genesys provides Helm override attributes that you use to set the configuration values for each service. For more information, see the appropriate service guide.

Operators

You can use OpenShift operators to deploy most third-party services into OpenShift. Note that Genesys does not provide operators to deploy Genesys Engage services.

Security

For more information, see Security Overview and Security Settings.

High-Availability

For more information, see High Availability and Disaster Recovery.

Data stores

Each service must have its own data store cluster or instances, which must not be shared in production environments unless they are under the same service group.

- All data stores must enable and deploy their high availability (HA) functionality
- All data stores must be distributed across Availability Zones, if they are available
- All data stores must support TLS connections and authentication, as appropriate

Here are the data stores used by each service:

Elasticsearch

Service	Type of Data	Shared across tenants	Cross region replication
Designer	Application Analytics data	Yes	No
IWD	Interaction and Queue Analytics data	Yes	No
TLM	Searchable telemetry data	Yes	No
UCS	Searchable contact and interaction history data	Yes	No
GWS WS	Searchable Statistics data	Yes	No
CXC	Campaign Analytics	Yes	No

Redis

Service	Type of Data	Shared across tenants	Cross region access
Pulse	runtime statistics	Yes	No
Tenant	stream of tenant data	Yes	Yes
CXC	runtime campaign and calling list status	Yes	No
Designer	config data	Yes	No
GES	runtime callback status and data	Yes	No
Nexus	runtime messaging session data	Yes	No
IWD	Historical reporting data	Yes	No
VMS (all of these services have separate keys (registrate, ORS, ORS stream, Callthread, Agent, Config, SIP, RQ))	runtime interaction, agent, registrations, config and routing request streams, scxml session data	Yes	Yes (not all)
GAuth	authentication session data	Yes	No
GWS	cached statistics, interaction and agent data.	Yes	No

Postgres and MS SQL

Service	Type of Data	Shared across tenants	Cross region replication
GCXI	metadata for reports	Yes	No
GVP RS - MS SQL	GVP reporting data	Yes	No
GVP CFG	config data	Yes	No
IXN	digital interaction and agent data	No	No
Pulse Permissions	config data	No	No
Tenant	config and campaign data	No	Yes
GES	config data	Yes	No
GIM	Historical reporting data	No	No
IWD	IWD config data	No	No
Nexus	config data	Yes	No
UCS	config data	Yes	No
UCS	contact, transcriptions, emails, interaction history	No	No
Gauth	config data	Yes	Yes
GWS	config data	Yes	Yes

File and disk storage

For more information, see Storage Requirements.

Email

The following private edition services send emails as part of their service:

- Voicemail
- GCXI
- Pulse

These services use standard mail agents on the operating system over SMTP via ports 25 and 587.

To use email with a service, you must set up the appropriate SMTP relay to relay messages from that service to your email system or email service. **Note:** This must be done from the Kubernetes clusters.

Content delivery networks (CDNs)

The WWE service that runs within private edition delivers static content. You can host this content from a CDN or from nginx running in the Kubernetes cluster.

Monitoring

Private edition provides appropriate interfaces for you to use your own monitoring tools. For the purposes of this software, monitoring encompasses:

- Metrics
- Logging
- Warnings
- Alerts

Monitoring (metrics)

Private edition provides a set of Prometheus-based metrics and defines an endpoint which the Prometheus platform can scrape. However, it does not provide a Grafana dashboard or Alert rule definitions.

Private edition uses OpenShift monitoring to verify all metrics provided by the pods.

Logging

Private edition provides the vast majority of its log data via stdout and stderr. In some exceptional cases, data is logged to disk.

Private edition uses OpenShift logging to verify all logs provided by the pods.

Integrations

Private edition support integrations with a wide variety of systems to provide an enriched customer experience, including in the following areas:

- Genesys Cloud for WEM, SMS, AI Solutions
- Bot platforms, such as Google Dialogflow and AWS Lex
- WFM platforms, such as Verint and Nice
- Email systems
- Identity providers
- Reporting platforms, including business intelligence tools

- Messaging and social platforms
- CRM and BPM systems
- Biometrics systems