



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.

Voice Microservices Private Edition Guide

Call State Service metrics and alerts

12/18/2025

Contents

- [1 Metrics](#)
- [2 Alerts](#)

Find the metrics Call State Service exposes and the alerts defined for Call State Service.

Service	CRD or annotations?	Port	Endpoint/Selector	Metrics update interval
Call State Service	Supports both CRD and annotations	11900	http://11900/metrics	30 seconds

See details about:

- Call State Service metrics
- Call State Service alerts

Metrics

Voice Call State Service exposes Genesys-defined, Call State Service-specific metrics as well as some standard Kafka metrics. You can query Prometheus directly to see all the metrics that the Call State Service exposes. The following metrics are likely to be particularly useful. Genesys does not commit to maintain other currently available Call State Service metrics not documented on this page.

Metric and description	Metric details	Indicator of
callthread_call_threads Number of monitored call threads.	Unit: N/A Type: counter Label: Sample value:	Saturation
callthread_envoy_proxy_status Status of the envoy proxy: -1 - error 0 - disconnected 1 - connected	Unit: N/A Type: gauge Label: Sample value:	
callthread_health_level Health level of the agent node: -1 - error 0 - fail 1 - degraded 2 - pass	Unit: N/A Type: gauge Label: Sample value:	
callthread_healthcheck_generic_exception Generic error during health check.	Unit: N/A Type: gauge	

Metric and description	Metric details	Indicator of
	Label: Sample value:	
callthread_redis_state Current Redis connection state: -1 – error 0 – disconnected 1 – connected 2 – ready	Unit: N/A Type: gauge Label: Sample value:	Errors
http_client_request_duration_seconds HTTP client time from request to response, in seconds.	Unit: seconds Type: histogram Label: target_service_name Sample value:	
http_client_response_count The number of HTTP client responses received.	Unit: N/A Type: counter Label: target_service_name, tenant, status Sample value:	
kafka_consumer_rcv_messages_total Number of messages received from Kafka.	Unit: N/A Type: counter Label: topic, tenant, kafka_location Sample value:	Traffic
kafka_consumer_error_total Number of Kafka consumer errors.	Unit: N/A Type: counter Label: topic, kafka_location Sample value:	Errors
kafka_consumer_latency Consumer latency is the time difference between when the message is produced and when the message is consumed. That is, the time when the consumer received the message minus the time when the producer produced the message.	Unit: Type: histogram Label: topic, tenant, kafka_location Sample value:	Latency
kafka_consumer_rebalance_total Number of Kafka consumer re-balance events.	Unit: N/A Type: counter Label: topic, kafka_location Sample value:	
kafka_consumer_state Current state of Kafka consumer.	Unit: N/A Type: gauge Label: topic, kafka_location Sample value:	
kafka_producer_messages_total Number of messages received from	Unit: N/A Type: counter	Traffic

Metric and description	Metric details	Indicator of
Kafka.	Label: topic, tenant, kafka_location Sample value:	
kafka_producer_queue_depth Number of Kafka producer pending events.	Unit: N/A Type: gauge Label: kafka_location Sample value:	Saturation
kafka_producer_queue_age_seconds Age of the oldest producer pending event, in seconds.	Unit: seconds Type: gauge Label: kafka_location Sample value:	
kafka_producer_error_total Number of Kafka producer errors.	Unit: N/A Type: counter Label: kafka_location Sample value:	Errors
kafka_producer_state Current state of the Kafka producer.	Unit: N/A Type: gauge Label: kafka_location Sample value:	
log_output_bytes_total Total amount of log output, in bytes.	Unit: bytes Type: counter Label: level, format, module Sample value:	

Alerts

The following alerts are defined for Call State Service.

Alert	Severity	Description	Based on	Threshold
Kafka events latency is too high	Critical	<p>Actions:</p> <ul style="list-style-type: none"> If the alarm is triggered for multiple topics, ensure there are no issues with Kafka (CPU, memory, or network overload). If the alarm is triggered only for topic {{ 	kafka_consumer_latency_bucket	Latency for more than 5% of messages is more than 0.5 seconds for topic {{ \$labels.topic }}.

Alert	Severity	Description	Based on	Threshold
		<code>\$labels.topic</code> }}, check if there is an issue with the service related to the topic (CPU, memory, or network overload).		
Too many Kafka consumer failed health checks	Warning	Actions: <ul style="list-style-type: none"> If the alarm is triggered for multiple services, make sure there are no issues with Kafka, and then restart Kafka. If the alarm is triggered only for {{ <code>\$labels.container</code> }}, check if there is an issue with the service. 	<code>kafka_consumer_error_total</code>	Health check failed more than 10 times in 5 minutes for Kafka consumer for topic {{ <code>\$labels.topic</code> }}.
Too many Kafka consumer request timeouts	Warning	Actions: <ul style="list-style-type: none"> If the alarm is triggered for multiple services, make sure there are no issues with Kafka, and then restart Kafka. If the alarm is triggered only for {{ <code>\$labels.container</code> }}, check if there is an issue with the service. 	<code>kafka_consumer_error_total</code>	More than 10 request timeouts appeared in 5 minutes for Kafka consumer for topic {{ <code>\$labels.topic</code> }}.
Too many Kafka consumer crashes	Critical	Actions:	<code>kafka_consumer_error_total</code>	More than 3 Kafka consumer crashes in 5 minutes for

Alert	Severity	Description	Based on	Threshold
		<ul style="list-style-type: none"> If the alarm is triggered for multiple services, make sure there are no issues with Kafka, and then restart Kafka. If the alarm is triggered only for {{ \$labels.container }}, check if there is an issue with the service. 		topic {{ \$labels.topic }}.
Pod status Failed	Warning	Actions: <ul style="list-style-type: none"> Restart the pod. Check if there are any issues with the pod after restart. 	kube_pod_status_phase	Pod {{ \$labels.pod }} is in Failed state.
Pod status Unknown	Warning	Actions: <ul style="list-style-type: none"> Restart the pod. Check if there are any issues with pod after restart. 	kube_pod_status_phase	Pod {{ \$labels.pod }} is in Unknown state for 5 minutes.
Pod status Pending	Warning	Actions: <ul style="list-style-type: none"> Restart the pod. Check if there are any issues with the pod after restart. 	kube_pod_status_phase	Pod {{ \$labels.pod }} is in Pending state for 5 minutes.
Pod status NotReady	Critical	Actions: <ul style="list-style-type: none"> Restart the pod. Check if there are any issues with the pod after restart. 	kube_pod_status_ready	Pod {{ \$labels.pod }} is in NotReady status for 5 minutes.

Alert	Severity	Description	Based on	Threshold
Container restarted repeatedly	Critical	<p>Actions:</p> <ul style="list-style-type: none"> Check if the new version of the image was deployed. Check for issues with the Kubernetes cluster. 	kube_pod_container_status_restarts_total	Container {{ \$labels.container }} was restarted 5 or more times total within 15 minutes.
Max replicas is not sufficient for 5 mins	Critical	The desired number of replicas is higher than the current available replicas for the past 5 minutes.	kube_statefulset_replicas, kube_statefulset_status_replicas	The desired number of replicas is higher than the current available replicas for the past 5 minutes.
Kafka not available	Critical	<p>Actions:</p> <ul style="list-style-type: none"> If the alarm is triggered for multiple services, make sure there are no issues with Kafka, and then restart Kafka. If the alarm is triggered only for pod {{ \$labels.pod }}, check if there is an issue with the pod. 	kafka_producer_status, kafka_consumer_status	Kafka is not available for pod {{ \$labels.pod }} for 5 consecutive minutes.
Redis not available	Critical	<p>Actions:</p> <ul style="list-style-type: none"> If the alarm is triggered for multiple services, make sure there are no issues with Redis, and then restart Redis. If the alarm is triggered only for pod {{ \$labels.pod }}, check if there is an issue with the pod. 	callthread_redis_status	Redis is not available for pod {{ \$labels.pod }} for 5 consecutive minutes.

Alert	Severity	Description	Based on	Threshold
		<code>\$labels.pod }</code> }, check if there is an issue with the pod.		
Pod CPU greater than 65%	Warning	High CPU load for pod <code>{{ \$labels.pod }}</code> .	<code>container_cpu_usage_seconds_total</code> <code>container_spec_cpu_period</code>	Container <code>{{ \$labels.container }}</code> CPU usage exceeded 65% for 5 minutes.
Pod CPU greater than 80%	Critical	Critical CPU load for pod <code>{{ \$labels.pod }}</code> .	<code>container_cpu_usage_seconds_total</code> <code>container_spec_cpu_period</code>	Container <code>{{ \$labels.container }}</code> CPU usage exceeded 80% for 5 minutes.
Pod memory greater than 65%	Warning	High memory usage for pod <code>{{ \$labels.pod }}</code> .	<code>container_memory_working_set_bytes</code> <code>kube_pod_container_resource_requests_memory_bytes</code>	Container <code>{{ \$labels.container }}</code> memory usage exceeded 65% for 5 minutes.
Pod memory greater than 80%	Critical	Critical memory usage for pod <code>{{ \$labels.pod }}</code> .	<code>container_memory_working_set_bytes</code> <code>kube_pod_container_resource_requests_memory_bytes</code>	Container <code>{{ \$labels.container }}</code> memory usage exceeded 80% for 5 minutes.
Too many Kafka pending events	Critical	Actions: <ul style="list-style-type: none"> Ensure there are no issues with Kafka or <code>{{ \$labels.container }}</code> service's CPU and network. 	<code>kafka_producer_queue_depth</code>	Too many Kafka producer pending events for service <code>{{ \$labels.container }}</code> (more than 100 in 5 minutes).