

# **GENESYS**<sup>®</sup>

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

ORS metrics and alerts

8/20/2025

#### Contents

- 1 Metrics
- 2 Alerts

Find the metrics ORS exposes and the alerts defined for ORS.

Service	CRD or annotations?	Port	Endpoint/Selector	Metrics update interval
ORS	Supports both CRD and annotations	11200	http://:11200/metrics	30 seconds

See details about:

- ORS metrics
- ORS alerts

#### Metrics

You can query Prometheus directly to see all the metrics that the Voice Orchestration Service exposes. The following metrics are likely to be particularly useful. Genesys does not commit to maintain other currently available Orchestration Service metrics not documented on this page.

Metric and description	Metric details	Indicator of
<b>orsnode_callevents</b> Total number of received call events.	Unit: N/A Type: counter Label: Sample value:	Traffic
<b>orsnode_ha_writes</b> The number of HA writes to Redis.	Unit: N/A Type: counter Label: Sample value:	Traffic
<b>orsnode_ha_reads</b> The number of HA reads from Redis.	Unit: N/A Type: counter Label: Sample value:	Traffic
<b>orsnode_interactions</b> The number of active interactions.	Unit: N/A Type: gauge Label: Sample value:	Traffic
<b>orsnode_total_interactions</b> The total number of interactions that	Unit: N/A Type: counter	Traffic

Metric and description	Metric details	Indicator of
have been created.	Label: Sample value:	
orsnode_cleared_interactions	Unit: N/A	
The total number of call interactions that have been cleared.	Type: counter Label: Sample value:	
orsnode_strategies	Unit: N/A	
The number of strategies that are running.	Type: gauge Label: Sample value:	Traffic
orsnode_total_strategies	Unit: N/A	
The total number of strategies that have been created.	Type: counter Label: Sample value:	Traffic
orspode load errors	Unit: N/A	
The total number of strategy load errors.	Type: counter Label: Sample value:	Errors
orsnode_fetch_errors	Unit: N/A	
The total number of errors encountered when a strategy tried to fetch data from a Designer Application Server (DAS).	Type: counter Label: Sample value:	Errors
orsnode_config_errors	Unit: N/A	
The total number of strategy configuration errors.	Type: counter Label: Sample value:	Errors
orsnode_invoke_errors	Unit: N/A	
The total number of strategy invoke errors.	Type: counter Label: Sample value:	Errors
orsnode treatments	Unit: N/A	
The total number of strategy treatments.	Type: counter Label: Sample value:	Traffic
orsnode_failed_treatments	Unit: N/A	
The total number of failed strategy treatments.	Type: counter Label: Sample value:	Errors
orsnode_userdata_updates	Unit: N/A	
The total number of times that a strategy	Type: counter	Traffic

Metric and description	Metric details	Indicator of
updated user data.	Label: Sample value:	
<b>orsnode_scxml_transitions</b> The total number of SCXML transitions.	Unit: N/A Type: counter Label: Sample value:	Traffic
orsnode_scxml_events The total number of SCXML events.	Unit: N/A Type: counter Label: Sample value:	Traffic
<b>orsnode_scxml_error_events</b> The total number of SCXML error.* events.	Unit: N/A Type: counter Label: Sample value:	Errors
<b>orsnode_http_fetch_requests</b> The total number of HTTP fetch requests.	Unit: N/A Type: counter Label: Sample value:	Errors
<b>orsnode_http_fetch_duration</b> The HTTP fetch time, measured in milliseconds (ms).	Unit: milliseconds Type: histogram Label: Sample value:	Latency
<b>orsnode_http_fetch_errors</b> The total number of HTTP fetch errors.	Unit: N/A Type: counter Label: Sample value:	Errors
<b>orsnode_http_fetch_error_statu</b> Status of the HTTP fetch error.	Unit: S Type: histogram Label: Sample value:	Errors
<b>orsnode_urs_rlib_latency_msec</b> The Universal Routing Server (URS) rlib latency, measured in milliseconds (ms).	Unit: milliseconds Type: histogram Label: Sample value:	Latency
<b>orsnode_urs_rlib_errors</b> The total number of URS rlib errors.	Unit: N/A Type: counter Label: Sample value:	Errors
<b>orsnode_urs_rlib_requests</b> The total number of URS rlib requests.	Unit: N/A Type: counter	

Metric and description	Metric details	Indicator of
	Label: Sample value:	
<b>orsnode_urs_rlib_events</b> The total number of URS rlib events.	Unit: N/A Type: counter Label: Sample value:	
<b>orsnode_urs_rlib_timeouts</b> The total number of URS rlib timeouts.	Unit: N/A Type: counter Label: Sample value:	
<b>orsnode_redis_state</b> Current Redis connection state.	Unit: N/A Type: gauge Label: redis_cluster_name Sample value:	
<b>orsnode_redis_disconnect</b> The number of times that the ORS node disconnected from Redis.	Unit: N/A Type: counter Label: Sample value:	
<b>orsnode_sdr_messages_sent</b> The number of SDR messages that have been sent.	Unit: N/A Type: counter Label: Sample value:	
orsnode_rq_latency_msec Redis queue latency, measured in milliseconds (ms).	Unit: milliseconds Type: histogram Label: le, service Sample value:	Latency
orsnode_routing_latency_msec Routing latency, measured in milliseconds (ms).	Unit: milliseconds Type: histogram Label: Sample value:	Latency
orsnode_rstream_latency_msec Redis stream latency, measured in (ms).	Unit: milliseconds Type: histogram Label: le, node Sample value:	Latency
<b>orsnode_digital_latency_msec</b> Digital stream latency, measured in milliseconds (ms).	Unit: milliseconds Type: histogram Label: Sample value:	Latency
orsnode_sip_health_check ORS health check.	Unit: N/A Type: gauge	

Metric and description	Metric details	Indicator of
	Label: node Sample value:	
orsnode_ixn_health_check	Unit: N/A Type: gauge Label: Sample value:	
<b>orsnode_rq_state</b> Current Redis queue connection state.	Unit: N/A Type: gauge Label: Sample value:	
<b>orsnode_ixn_events</b> Total number of interaction stream events received.	Unit: N/A Type: counter Label: Sample value:	
<b>orsnode_rq_disconnect</b> Number of times the ORS node disconnected from the RQ Service.	Unit: N/A Type: counter Label: Sample value:	
<b>service_version_info</b> Displays the version of Voice Orchestration Service that is currently running. In the case of this metric, the labels provide the important information. The metric value is always 1 and does not provide any information.	Unit: N/A Type: gauge Label: version Sample value: service_version_info{version="100.0.10000 1	06"}
<b>orsnode_route_redirected</b> Total number of EventRouteUsed events without a ReferenceID.	Unit: N/A Type: counter Label: Sample value:	
<b>orsnode_balancer_stream_state</b> The state of the voice balancer stream.	Unit: N/A Type: gauge Label: balancer_stream_type Sample value:	
<b>orsnode_high_memory</b> Indicates when the ORS node is using a lot of memory.	Unit: N/A Type: gauge Label: Sample value:	
<b>orsnode_urs_rlib_state</b> Indicates a Tenant rlib request timeout.	Unit: N/A Type: gauge Label: Sample value:	
orsnode_stuck_interactions	Unit: N/A	

Metric and description	Metric details	Indicator of
The number of stuck interactions.	Type: gauge Label: Sample value:	
orsnode_urs_scxml_submit_requ	uests: N/A	
The total number of URS SCXMLSubmit requests.	Type: counter Label: Sample value:	
orsnode_urs_scxml_cancel_requ	ests N/A	
The total number of URS SCXMLQueueCancel requests.	Type: counter Label: Sample value:	
orsnode_urs_queue_submit_dor	e eitente	
Total number of URS queue.submit.done events.	Type: counter Label: Sample value:	
orsnode_health_level		
Summarized health level of the ORS	Unit: N/A	
-1 - fail	Type: gauge Label:	
0 - starting 1 - degraded 2 - pass	Sample value:	
orsnode_health_check_error	Unit: N/A	
Health check errors for the ORS node:	Type: gauge	Errors
1 – has error 0 – no error	Sample value:	
orsnode_running_applications	Unit: N/A	
The number of active sessions for each Designer application.	Type: gauge Label: Sample value:	
orsnode_failed_applications	Unit: N/A	
The number of failed sessions for each	Type: gauge Label:	
Designer application.	Sample value:	
orsnode_total_applications	Unit: N/A	
The total number of sessions created for each Designer application.	Type: gauge Label: Sample value:	
orsnode_failed_scripts	Unit: N/A	
The number of scripts that failed to load in the Tenant Service configuration	Type: gauge Label:	

Metric and description	Metric details	Indicator of		
management environment.	Sample value:			
orsnode_session_load_time_msec <sup>Unit:</sup> milliseconds				
The time it takes for the strategy to be compiled and go through its initial states.	Type: histogram Label: Sample value:			
orsnode service started	Unit: N/A			
Timestamp when the ORS node started.	Type: gauge Label: started Sample value:			
orsnode_total_terminal_request	<sup>ts</sup> Unit: N/A			
Total number of terminal requests (like Deliver, PlacelnQueue, StopProcessing for Digital and RequestClearCall, RequestRouteCall for Voice).	Type: counter Label: Sample value:			
orsnode_total_non_terminal_red	<b>dubrits</b> N/A			
Total number of non-terminal requests to the Interaction Server (for Digital) or SIP Server (for Voice).	Type: counter Label: Sample value:			
orsnode_sip_post_errors	Unit: N/A			
Total number of errors encountered in POST requests to the SIP node.	Type: counter Label: Sample value:	Errors		
orsnode pending tlib requests	Unit: N/A			
Total number of pending TLib requests.	Type: counter Label: Sample value:			
orsnode_sips_rest_connections	Unit: N/A			
The number of active REST connections with SIP Cluster Service.	Type: gauge Label: Sample value:			
orsnode_number_compiled_app	lications			
The number of compiled applications in the cache.	Type: counter Label: Sample value:			
orsnode_cached_applications_s	Unit:			
The sum of the sizes of the cached applications.	Type: gauge Label: Sample value:			
orsnode_tlib_latency_msec	Unit: milliseconds			
The TLib Rest API request latency,	Type: histogram	Latency		

Metric and description	Metric details	Indicator of
measured in (ms).	Label: le Sample value:	
orsnode_application_size	Unit:	
The compiled size of the Designer application.	Type: gauge Label: Sample value:	
orsnode_application_microstep	Count N/A	
The number of microsteps while executing the Designer application.	Type: gauge Label: Sample value:	
orsnode_application_run_time_	ngeit: milliseconds	
The length of time the Designer application was running, measured in milliseconds (ms).	Type: gauge Label: Sample value:	
orsnode_application_compiled_	date N/A	
The date on which the Designer application was compiled.	Type: gauge Label: Sample value:	
orsnode_application_last_invok	ed_date <sup>/A</sup>	
The date when the Designer application was last invoked.	Type: gauge Label: Sample value:	

### Alerts

The following alerts are defined for ORS.

Alert	Severity	Description	Based on	Threshold
Number of running strategies is too high	Warning	Too many active sessions. Actions: • Check whether the horizontal pod autoscaler has triggered and if the maximum number of pods has been reached.	orsnode_strategies	More than 400 strategies running in 5 consecutive seconds.

Alert	Severity	Description	Based on	Threshold
		<ul> <li>Check the number of voice, digital, and callback calls in the system.</li> </ul>		
Number of running strategies is critical	Critical	<ul> <li>Too many active sessions.</li> <li>Actions:</li> <li>Check whether the horizontal pod autoscaler has triggered and if the maximum number of pods has been reached.</li> <li>Check the number of voice, digital, and callback calls in the system.</li> </ul>	orsnode_strategies	More than 600 strategies running in 5 consecutive seconds.
Redis disconnected for 5 minutes	Warning	<ul> <li>Actions:</li> <li>If the alarm is triggered for multiple services, make sure there are no issues with Redis, and then restart Redis.</li> <li>If alarm is triggered only for pod {{ \$labels.pod }}, check if there is an issue with the pod.</li> </ul>	redis_state	Redis is not available for the pod {{ \$labels.pod }} for 5 minutes.
Redis disconnected for 10 minutes	Critical	Actions: • If the alarm is triggered for multiple services, make	redis_state	Redis is not available for the pod {{ \$labels.pod }} for 10 minutes.

Alert	Severity	Description	Based on	Threshold
		<ul> <li>sure there are no issues with Redis, and then restart Redis.</li> <li>If the alarm is triggered only for pod {{ \$labels.pod }}, check if there is an issue with the pod.</li> </ul>		
Pod status Failed	Warning	<ul> <li>Pod {{ \$labels.pod }} failed.</li> <li>Actions:</li> <li>One of the containers in the pod has entered a Failed state. Check the Kibana logs for the reason.</li> </ul>	kube_pod_status_ph	Pod {{ \$labels.pod }} is in Failed a <i>s</i> tate.
Pod in Unknown state	Warning	<ul> <li>Pod {{ \$labels.pod }} is in Unknown state.</li> <li>Actions:</li> <li>If the alarm is triggered for multiple services, make sure there are no issues with the Kubernetes cluster.</li> <li>If the alarm is triggered only for pod {{ \$labels.pod }}, check whether the image is correct and if the container is starting up.</li> </ul>	kube_pod_status_ph	Pod {{ \$labels.pod }} is in Unknown state for 5 affiinutes.
Pod in Pending state	Warning	Pod {{ \$labels.pod }} is in Pending state.	kube_pod_status_ph	Pod {{ \$labels.pod age} is in Pending state for 5

Alert	Severity	Description	Based on	Threshold
		<ul> <li>Actions:</li> <li>If the alarm is triggered for multiple services, make sure the Kubernetes nodes where the pod is running are alive in the cluster.</li> <li>If the alarm is triggered only for the pod {{ \$labels.pod }}, check the health of the pod.</li> </ul>		minutes.
Pod Not ready for 10 minutes	Critical	<ul> <li>Pod {{ \$labels.pod }} in NotReady state.</li> <li>Actions:</li> <li>If this alarm is triggered, check whether the CPU is available for the pods.</li> <li>Check whether the port of the pod is running and serving the request.</li> </ul>	kube_pod_status_rea	Pod {{ \$labels.pod }} in NotReady state for 10 Minutes.
Container restored repeatedly	Critical	Actions: • One of the containers in the pod has entered a failed state. Check the Kibana logs for the reason.	kube_pod_container_	Container {{ \$labels.container }} was restarted 5 ទុសសាទាទនំថាទន្ទ total within 15 minutes.
Pod memory greater than 65%	Warning	High memory usage for pod {{ \$labels.pod }}.	container_memory_v kube_pod_container_	Container {{ vorking set bytes Slabels container resource requests m - }} memory usage-

Alert	Severity	Description	Based on	Threshold
		<ul> <li>Actions:</li> <li>Check whether the horizontal pod autoscaler has triggered and if the maximum number of pods has been reached.</li> <li>Check Grafana for abnormal load.</li> <li>Collect the service logs; raise an investigation ticket.</li> </ul>		exceeded 65% for 5 minutes.
Pod memory greater than 80%	Critical	<ul> <li>Critical memory usage for pod {{ \$labels.pod }}.</li> <li>Actions:</li> <li>Check whether the horizontal pod autoscaler has triggered and if the maximum number of pods has been reached.</li> <li>Check Grafana for abnormal load.</li> <li>Restart the service.</li> <li>Collect the service logs; raise an investigation ticket.</li> </ul>	container_memory_v kube_pod_container	Container { { \$labels.container }} memory usage Wexcegated 80%For _r55MHrGeesequests_m
Pod CPU greater than 65%	Warning	High CPU load for pod {{ \$labels.pod }}. Actions:	container_cpu_usage container_spec_cpu_	Container { { slabels.container -}{CDU usage beriod usage exceeded 65% for 5 minutes.

Alert	Severity	Description	Based on	Threshold
		<ul> <li>Check whether the horizontal pod autoscaler has triggered and if the maximum number of pods has been reached.</li> <li>Check Grafana for abnormal load.</li> <li>Collect the service logs; raise an investigation ticket.</li> </ul>		
Pod CPU greater than 80%	Critical	<ul> <li>Critical CPU load for pod {{ \$labels.pod }}.</li> <li>Actions: <ul> <li>Check whether the horizontal pod autoscaler has triggered and if the maximum number of pods has been reached.</li> </ul> </li> <li>Check Grafana for abnormal load.</li> <li>Restart the service.</li> <li>Collect the service logs; raise an investigation ticket.</li> </ul>	container_cpu_usage container_spec_cpu_	Container { { slabels.container -se conds.torainer -se conds.torainer -se conds.torainer -secceded 80% for 5 minutes.