



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

Voice SIP Cluster Service metrics and alerts

5/6/2026

---

## Contents

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

Find the metrics Voice SIP Cluster Service exposes and the alerts defined for Voice SIP Cluster Service.

| Service                   | CRD or annotations?               | Port  | Endpoint/Selector     | Metrics update interval |
|---------------------------|-----------------------------------|-------|-----------------------|-------------------------|
| Voice SIP Cluster Service | Supports both CRD and annotations | 11300 | http://:11300/metrics | 30 seconds              |

See details about:

- Voice SIP Cluster Service metrics
- Voice SIP Cluster Service alerts

## Metrics

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

| Metric and description   | Metric details  | Indicator of |
|--|---|--------------|
| <b>http_client_request_duration_seconds</b><br>HTTP client time from request to response, measured in seconds. | <b>Unit:</b> seconds<br><b>Type:</b> histogram<br><b>Label:</b> le, target_service_name<br><b>Sample value:</b> | Latency      |
| <b>http_client_response_count</b><br>Number of received HTTP client responses.                                 | <b>Unit:</b> N/A<br><b>Type:</b> counter<br><b>Label:</b> target_service_name<br><b>Sample value:</b>           | Traffic      |
| <b>kafka_producer_queue_depth</b><br>Number of Kafka producer pending events.                                  | <b>Unit:</b> N/A<br><b>Type:</b> gauge<br><b>Label:</b> kafka_location<br><b>Sample value:</b>                  | Traffic      |
| <b>kafka_producer_queue_age_seconds</b><br>Age of the oldest producer pending event, measured in seconds.      | <b>Unit:</b> seconds<br><b>Type:</b> gauge<br><b>Label:</b> kafka_location<br><b>Sample value:</b>              | Traffic      |

| Metric and description   | Metric details  | Indicator of |
|--|---|--------------|
| <b>kafka_producer_error_total</b><br>Number of Kafka producer errors.  | <b>Unit:</b> N/A<br><b>Type:</b> counter<br><b>Label:</b> kafka_location<br><b>Sample value:</b>            | Errors       |
| <b>log_output_bytes_total</b><br>Total amount of log output in bytes.  | <b>Unit:</b> bytes<br><b>Type:</b> counter<br><b>Label:</b> level, format, module<br><b>Sample value:</b>   | Traffic      |
| <b>sipnode_requests_total</b><br>Number of processed requests.   | <b>Unit:</b> N/A<br><b>Type:</b> counter<br><b>Label:</b> tenant, request<br><b>Sample value:</b>           | Traffic      |
| <b>sipnode_pending_requests_current</b><br>Number of pending requests.   | <b>Unit:</b> N/A<br><b>Type:</b> gauge<br><b>Label:</b> tenant, request<br><b>Sample value:</b>             | Traffic      |
| <b>sipnode_requests_queue_size</b><br>Number of postponed requests.  | <b>Unit:</b> N/A<br><b>Type:</b> gauge<br><b>Label:</b><br><b>Sample value:</b>                             | Saturation   |
| <b>sipnode_sips_request_duration_seconds</b><br>Duration of the request processed by SIP Cluster Service, measured in seconds. | <b>Unit:</b> seconds<br><b>Type:</b> histogram<br><b>Label:</b> le, tenant, request<br><b>Sample value:</b> | Traffic      |
| <b>sipnode_events_total</b><br>Call events streamed to Redis.  | <b>Unit:</b> N/A<br><b>Type:</b> counter<br><b>Label:</b> tenant, event<br><b>Sample value:</b>             | Traffic      |
| <b>sipnode_ha_writes_total</b><br>Number of HA writes to Redis.  | <b>Unit:</b> N/A<br><b>Type:</b> counter<br><b>Label:</b><br><b>Sample value:</b>                           | Traffic      |
| <b>sipnode_ha_reads_total</b><br>Number of HA reads from Redis.  | <b>Unit:</b> N/A<br><b>Type:</b> counter<br><b>Label:</b><br><b>Sample value:</b>                           | Traffic      |
| <b>sipnode_monitoring_events_total</b><br>Number of monitoring events submitted to Kafka.                                      | <b>Unit:</b> N/A<br><b>Type:</b> counter<br><b>Label:</b> tenant<br><b>Sample value:</b>                    | Traffic      |
| <b>sipnode_redis_restored_calls_total</b>  | <b>Unit:</b> N/A<br><b>Type:</b> counter<br><b>Label:</b><br><b>Sample value:</b>                           | Traffic      |

| Metric and description   | Metric details  | Indicator of |
|--|---|--------------|
| Total number of restored calls from Redis cache.   | <b>Type:</b> counter<br><b>Label:</b><br><b>Sample value:</b>   |              |
| <b>sipnode_sips_restarts_total</b><br>Total number of SIP Server restarts.   | <b>Unit:</b> N/A<br><b>Type:</b> counter<br><b>Label:</b><br><b>Sample value:</b>   | Errors       |
| <b>sipnode_sips_disconnects_total</b><br>Total number of SIP Cluster Service disconnections from SIP Server.   | <b>Unit:</b> N/A<br><b>Type:</b> counter<br><b>Label:</b><br><b>Sample value:</b>   | Errors       |
| <b>sipnode_redis_state</b><br>Current Redis connection state.  | <b>Unit:</b> N/A<br><b>Type:</b> gauge<br><b>Label:</b> redis_cluster_name<br><b>Sample value:</b>  | Errors       |
| <b>sipnode_ors_tlib_latency_msec</b><br>T-Library latency from Orchestration Service to SIP Cluster, measured in milliseconds.   | <b>Unit:</b> milliseconds<br><b>Type:</b> histogram<br><b>Label:</b> le, ors<br><b>Sample value:</b>  | Latency      |
| <b>sipnode_ors_health_check</b><br>SIP Cluster Service to Orchestration Service health check.  | <b>Unit:</b> N/A<br><b>Type:</b> gauge<br><b>Label:</b><br><b>Sample value:</b>   | Traffic      |
| <b>service_version_info</b><br>Displays the version of Voice SIP Cluster 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. | <b>Unit:</b> N/A<br><b>Type:</b> gauge<br><b>Label:</b> version<br><b>Sample value:</b><br>service_version_info{version="100.0.1000006"}<br>1 |              |
| <b>sipnode_treatment_not_applied</b><br>Number of unsuccessful treatments.   | <b>Unit:</b> N/A<br><b>Type:</b> counter<br><b>Label:</b> tenant<br><b>Sample value:</b>  | Errors       |
| <b>sipnode_default_routing_total</b><br>Total number of default routed calls.  | <b>Unit:</b> N/A<br><b>Type:</b> counter<br><b>Label:</b> tenant<br><b>Sample value:</b>  | Traffic      |
| <b>sipnode_envoy_proxy_status</b><br>Status of the Envoy proxy:<br>-1 - error<br>0 - disconnected  | <b>Unit:</b> N/A<br><b>Type:</b> gauge<br><b>Label:</b><br><b>Sample value:</b> 1   | Health       |

| Metric and description  | Metric details   | Indicator of |
|---|--|--------------|
| 1 - connected   |  |              |
| <b>sipnode_config_node_status</b><br>Status of the config node connection:<br>0 - disconnected<br>1 - connected                             | <b>Unit:</b> N/A<br><br><b>Type:</b> gauge<br><b>Label:</b><br><b>Sample value:</b> 1        | Health       |
| <b>sipnode_health_level</b><br>Health level of the SIP node (SIP Cluster Service):<br>-1 - fail<br>0 - starting<br>1 - degraded<br>2 - pass | <b>Unit:</b> N/A<br><br><b>Type:</b> gauge<br><b>Label:</b><br><b>Sample value:</b> 2        | Traffic      |
| <b>sipnode_call_state_health_check</b><br>SIP Cluster Service to Call State Service health check.   | <b>Unit:</b> N/A<br><br><b>Type:</b> gauge<br><b>Label:</b> memberId<br><b>Sample value:</b> | Health       |
| <b>sips_hastate</b><br>Current HA state of SIP Server:<br>0 - Unknown<br>1 - backup<br>2 - primary  | <b>Unit:</b> N/A<br><br><b>Type:</b> gauge<br><b>Label:</b><br><b>Sample value:</b> 2        |              |
| <b>sips_calls</b><br>Current number of calls.   | <b>Unit:</b> N/A<br><br><b>Type:</b> gauge<br><b>Label:</b><br><b>Sample value:</b>          | Traffic      |
| <b>sips_call_rate</b><br>Call rate.   | <b>Unit:</b> N/A<br><br><b>Type:</b> gauge<br><b>Label:</b><br><b>Sample value:</b>          | Traffic      |
| <b>sips_cpu_usage_sips</b><br>SIP Server CPU usage.   | <b>Unit:</b> N/A<br><br><b>Type:</b> gauge<br><b>Label:</b><br><b>Sample value:</b>          | Saturation   |
| <b>sips_cpu_usage_main</b><br>SIP Server main thread CPU usage.   | <b>Unit:</b> N/A<br><br><b>Type:</b> gauge<br><b>Label:</b><br><b>Sample value:</b>          | Saturation   |
| <b>sips_cpu_usage_cm</b><br>CPU usage of the call manager thread.   | <b>Unit:</b> N/A<br><br><b>Type:</b> gauge<br><b>Label:</b>                                  | Saturation   |

| Metric and description  | Metric details  | Indicator of |
|---|---|--------------|
|   | <b>Sample value:</b>  |              |
| <b>sips_calls_created</b><br>Total number of created calls.                         | <b>Unit:</b> N/A<br><b>Type:</b> gauge<br><b>Label:</b><br><b>Sample value:</b> | Traffic      |
| <b>sips_abandoned_calls</b><br>Total number of abandoned calls.                     | <b>Unit:</b> N/A<br><b>Type:</b> gauge<br><b>Label:</b><br><b>Sample value:</b> | Errors       |
| <b>sips_rejected_calls</b><br>Total number of rejected calls.                       | <b>Unit:</b> N/A<br><b>Type:</b> gauge<br><b>Label:</b><br><b>Sample value:</b> | Errors       |
| <b>sips_dialogs_created</b><br>Total number of created SIP dialogs.                 | <b>Unit:</b> N/A<br><b>Type:</b> gauge<br><b>Label:</b><br><b>Sample value:</b> | Traffic      |
| <b>sips_call_recording_failed</b><br>Number of failed call recording sessions.      | <b>Unit:</b> N/A<br><b>Type:</b> gauge<br><b>Label:</b><br><b>Sample value:</b> | Errors       |
| <b>sips_urs_response_1_to_5_sec</b><br>Number of URS responses from 1 to 5 seconds. | <b>Unit:</b> N/A<br><b>Type:</b> gauge<br><b>Label:</b><br><b>Sample value:</b> | Latency      |
| <b>sips_urs_response_more_5_sec</b><br>Number of URS responses more than 5 seconds. | <b>Unit:</b> N/A<br><b>Type:</b> gauge<br><b>Label:</b><br><b>Sample value:</b> | Latency      |
| <b>sips_user_data_updates</b><br>Number of UserData updates.                        | <b>Unit:</b> N/A<br><b>Type:</b> gauge<br><b>Label:</b><br><b>Sample value:</b> | Traffic      |
| <b>sips_routing_timeouts</b><br>Number of routing timeouts.                         | <b>Unit:</b> N/A<br><b>Type:</b> gauge<br><b>Label:</b><br><b>Sample value:</b> | Errors       |
| <b>sips_trequest_rate</b><br>T-Requests rate.                                       | <b>Unit:</b> N/A<br><b>Type:</b> gauge<br><b>Label:</b>                         | Traffic      |

| Metric and description  | Metric details  | Indicator of |
|---|---|--------------|
|   | <b>Sample value:</b>  |              |
| <b>sips_treatment_rate</b><br>TApplyTreatment requests rate.                                      | <b>Unit:</b> N/A<br><b>Type:</b> gauge<br><b>Label:</b><br><b>Sample value:</b>                     | Traffic      |
| <b>sips_userdata_rate</b><br>UserData change rate.  | <b>Unit:</b> N/A<br><b>Type:</b> gauge<br><b>Label:</b><br><b>Sample value:</b>                     | Traffic      |
| <b>sips_sips_memory_usage</b><br>Memory usage of the SIP Server process.                          | <b>Unit:</b> N/A<br><b>Type:</b> gauge<br><b>Label:</b><br><b>Sample value:</b>                     | Saturation   |
| <b>sips_stat_fetch_total</b><br>Number of successful SIP Server statistic fetches.                | <b>Unit:</b> N/A<br><b>Type:</b> counter<br><b>Label:</b><br><b>Sample value:</b>                   | Other        |
| <b>sips_sip_response_time_ms</b><br>SIP Server metric of response time, measured in milliseconds. | <b>Unit:</b> milliseconds<br><b>Type:</b> histogram<br><b>Label:</b> le<br><b>Sample value:</b>     | Latency      |
| <b>sips_trunk_in_service</b><br>Trunk devices that are in service.                                | <b>Unit:</b> N/A<br><b>Type:</b> gauge<br><b>Label:</b> device_name, tenant<br><b>Sample value:</b> | Traffic      |
| <b>sips_trunk_ncallscreated</b><br>Number of created calls per trunk.                             | <b>Unit:</b> N/A<br><b>Type:</b> gauge<br><b>Label:</b> device_name, tenant<br><b>Sample value:</b> | Traffic      |
| <b>sips_trunk_noos_detected</b><br>Number of trunks that are out of service.                      | <b>Unit:</b> N/A<br><b>Type:</b> gauge<br><b>Label:</b> device_name, tenant<br><b>Sample value:</b> | Errors       |
| <b>sips_trunk_n4xx_received</b><br>Number of received 4xx messages.                               | <b>Unit:</b> N/A<br><b>Type:</b> gauge<br><b>Label:</b> device_name, tenant<br><b>Sample value:</b> | Errors       |
| <b>sips_trunk_n5xx_received</b><br>Number of received 5xx messages.                               | <b>Unit:</b> N/A<br><b>Type:</b> gauge<br><b>Label:</b> device_name, tenant                         | Errors       |

| Metric and description  | Metric details  | Indicator of |
|---|---|--------------|
|   | <b>Sample value:</b>  |              |
| <b>sips_trunk_n6xx_received</b><br>Number of received 6xx messages.                           | <b>Unit:</b> N/A<br><b>Type:</b> gauge<br><b>Label:</b> device_name, tenant<br><b>Sample value:</b> | Errors       |
| <b>sips_softswitch_in_service</b><br>Softswitch devices that are in service.                  | <b>Unit:</b> N/A<br><b>Type:</b> gauge<br><b>Label:</b> device_name, tenant<br><b>Sample value:</b> | Traffic      |
| <b>sips_softswitch_ncallscreated</b><br>Number of created calls per softswitch device.        | <b>Unit:</b> N/A<br><b>Type:</b> gauge<br><b>Label:</b> device_name, tenant<br><b>Sample value:</b> | Traffic      |
| <b>sips_softswitch_noos_detected</b><br>Number of softswitch devices that are out of service. | <b>Unit:</b> N/A<br><b>Type:</b> gauge<br><b>Label:</b> device_name, tenant<br><b>Sample value:</b> | Errors       |
| <b>sips_softswitch_n4xx_received</b><br>Number of received 4xx messages.                      | <b>Unit:</b> N/A<br><b>Type:</b> gauge<br><b>Label:</b> device_name, tenant<br><b>Sample value:</b> | Errors       |
| <b>sips_softswitch_n5xx_received</b><br>Number of received 5xx messages.                      | <b>Unit:</b> N/A<br><b>Type:</b> gauge<br><b>Label:</b> device_name, tenant<br><b>Sample value:</b> | Errors       |
| <b>sips_softswitch_n6xx_received</b><br>Number of received 6xx messages.                      | <b>Unit:</b> N/A<br><b>Type:</b> gauge<br><b>Label:</b> device_name, tenant<br><b>Sample value:</b> | Errors       |
| <b>sips_msml_in_service</b><br>MSML devices that are in service.                              | <b>Unit:</b> N/A<br><b>Type:</b> gauge<br><b>Label:</b> device_name<br><b>Sample value:</b>         | Traffic      |
| <b>sips_msml_ncallscreated</b><br>Number of created calls per MSML device.                    | <b>Unit:</b> N/A<br><b>Type:</b> gauge<br><b>Label:</b> device_name<br><b>Sample value:</b>         | Traffic      |
| <b>sips_msml_noos_detected</b><br>Number of MSML devices that are out of service.             | <b>Unit:</b> N/A<br><b>Type:</b> gauge<br><b>Label:</b> device_name                                 | Errors       |

| Metric and description   | Metric details  | Indicator of |
|--|---|--------------|
|  | <b>Sample value:</b>  |              |
| <b>sips_msml_n4xx_received</b><br>Number of received 4xx messages.                           | <b>Unit:</b> N/A<br><b>Type:</b> gauge<br><b>Label:</b> device_name<br><b>Sample value:</b> | Errors       |
| <b>sips_msml_n5xx_received</b><br>Number of received 5xx messages.                           | <b>Unit:</b> N/A<br><b>Type:</b> gauge<br><b>Label:</b> device_name<br><b>Sample value:</b> | Errors       |
| <b>sips_msml_n6xx_received</b><br>Number of received 6xx messages.                           | <b>Unit:</b> N/A<br><b>Type:</b> gauge<br><b>Label:</b> device_name<br><b>Sample value:</b> | Errors       |
| <b>sips_dp_state</b><br>Dial Plan Service state:<br>0 - Out-Of-Service<br>1 - In-Service     | <b>Unit:</b> N/A<br><b>Type:</b> gauge<br><b>Label:</b><br><b>Sample value:</b> 1           | Traffic      |
| <b>sips_dp_queue_size</b><br>Size of the request queue to Dial Plan Service.                 | <b>Unit:</b> N/A<br><b>Type:</b> gauge<br><b>Label:</b><br><b>Sample value:</b>             | Traffic      |
| <b>sips_dp_avg_queue_time</b><br>Average queue time (msec) of requests to Dial Plan Service. | <b>Unit:</b> milliseconds<br><b>Type:</b> gauge<br><b>Label:</b><br><b>Sample value:</b>    | Latency      |
| <b>sips_dp_connections</b><br>Number of connections to Dial Plan Service per URL.            | <b>Unit:</b> N/A<br><b>Type:</b> gauge<br><b>Label:</b><br><b>Sample value:</b>             | Traffic      |
| <b>sips_dp_active_connections</b><br>Number of active connections to Dial Plan Service.      | <b>Unit:</b> N/A<br><b>Type:</b> gauge<br><b>Label:</b><br><b>Sample value:</b>             | Traffic      |
| <b>sips_dp_req_rate</b><br>Request rate to Dial plan Service.                                | <b>Unit:</b> N/A<br><b>Type:</b> gauge<br><b>Label:</b><br><b>Sample value:</b>             | Traffic      |
| <b>sips_dp_400_errors</b><br>Dial Plan Service 400 type of errors.                           | <b>Unit:</b> N/A<br><b>Type:</b> gauge  | Errors       |

| Metric and description  | Metric details  | Indicator of |
|---|---|--------------|
|   | <b>Label:</b><br><b>Sample value:</b>   |              |
| <b>sips_dp_404_errors</b><br>Dial Plan Service 404 type of errors.                                  | <b>Unit:</b> N/A<br><b>Type:</b> gauge<br><b>Label:</b><br><b>Sample value:</b>   | Errors       |
| <b>sips_dp_4xx_errors</b><br>Dial Plan Service 4xx type of errors.                                  | <b>Unit:</b> N/A<br><b>Type:</b> gauge<br><b>Label:</b><br><b>Sample value:</b>   | Errors       |
| <b>sips_dp_500_errors</b><br>Dial Plan Service 500 type of errors.                                  | <b>Unit:</b> N/A<br><b>Type:</b> gauge<br><b>Label:</b><br><b>Sample value:</b>   | Errors       |
| <b>sips_dp_501_errors</b><br>Dial Plan Service 501 type of errors.                                  | <b>Unit:</b> N/A<br><b>Type:</b> gauge<br><b>Label:</b><br><b>Sample value:</b>   | Errors       |
| <b>sips_dp_5xx_errors</b><br>Dial Plan Service 5xx type of errors.                                  | <b>Unit:</b> N/A<br><b>Type:</b> gauge<br><b>Label:</b><br><b>Sample value:</b>   | Errors       |
| <b>sips_dp_timeouts</b><br>Dial Plan Service timeouts.  | <b>Unit:</b> N/A<br><b>Type:</b> gauge<br><b>Label:</b><br><b>Sample value:</b>   | Errors       |
| <b>sips_dp_average_response_latency</b><br>Dial Plan Service average response latency.              | <b>Unit:</b><br><b>Type:</b> gauge<br><b>Label:</b><br><b>Sample value:</b>       | Latency      |
| <b>sips_sipproxy_in_service</b><br>SIP Proxy Service state:<br>0 - Out-Of-Service<br>1 - In-Service | <b>Unit:</b> N/A<br><b>Type:</b> gauge<br><b>Label:</b><br><b>Sample value:</b> 1 | Traffic      |
| <b>trunk_config_synced_count</b><br>Number of trunks synchronized with SIP Server.                  | <b>Unit:</b> N/A<br><b>Type:</b> gauge<br><b>Label:</b><br><b>Sample value:</b>   |              |
| <b>trunk_config_cached_count</b>  | <b>Unit:</b> N/A  |              |

| Metric and description  | Metric details  | Indicator of |
|---|---|--------------|
| Number of trunks obtained from the config node.   | <b>Type:</b> gauge<br><b>Label:</b><br><b>Sample value:</b>                       |              |
| <b>trunk_config_cfg_node_error_count</b><br>Number of failed attempts to read from the config node. | <b>Unit:</b> N/A<br><b>Type:</b> counter<br><b>Label:</b><br><b>Sample value:</b> |              |
| <b>trunk_config_tlib_connection</b><br>Number of trunks with the T-Library connection.              | <b>Unit:</b> N/A<br><b>Type:</b> gauge<br><b>Label:</b><br><b>Sample value:</b>   |              |

## Alerts

The following alerts are defined for Voice SIP Cluster Service.

| Alert                         | Severity | Description  | Based on                         | Threshold   |
|-------------------------------|----------|--|----------------------------------|---|
| Too many Kafka pending events | Critical | <p>Too many Kafka producer pending events for pod {{ \$labels.pod }}.</p> <p>Actions:</p> <ul style="list-style-type: none"> <li>Ensure there are no issues with Kafka, {{ \$labels.pod }} pod's CPU, and network.</li> </ul>                                    | kafka_producer_queue_depth       | Too many Kafka producer pending events for service {{ \$labels.service,container }} (more than 100 in 5 minutes). |
| Dial Plan node is overloaded  | Critical | <p>Dial Plan node is overloaded as the response latency increases.</p> <p>Actions:</p> <ul style="list-style-type: none"> <li>Check that the inbound call rate to SIP Server is not too high.</li> <li>Check the Dial Plan node CPU and memory usage.</li> </ul> | sips_dp_average_response_latency | Dial Plan node is overloaded as the response latency increases (more than 1000).                                  |

| Alert                    | Severity | Description   | Based on  | Threshold   |
|--------------------------|----------|---|---|---|
|                          |          | <ul style="list-style-type: none"> <li>Check the network connection between SIP Server and Dial Plan nodes.</li> </ul>  |   |   |
| Dial Plan Queue Increase | Critical | <p>Because Dial Plan requests are huge in size or there is a connection issue with the Dial Plan node, the processing queue size increases in size.</p> <p>Actions:</p> <ul style="list-style-type: none"> <li>Check SIP Server inbound call rate.</li> <li>Check the connection between SIP Server and the Dial Plan node.</li> </ul>            | sips_dp_queue_size  | The processing queue size is greater than 10 requests for 1 minute. |
| SIP Proxy overloaded     | Critical | <p>SIP Proxy is overloaded.</p> <p>Actions:</p> <ul style="list-style-type: none"> <li>Check SIP Proxy nodes for CPU and memory usage.</li> <li>If SIP Proxy nodes have acceptable CPU and memory usage, then check for errors or a "hang-up" state which could delay SIP Proxy in forwarding.</li> <li>Check the SBC side for network</li> </ul> | sips_sip_response_time_millis_sum,<br>sips_sip_response_time_millis_count | Response time is greater than 20 milliseconds for 1 minute          |

| Alert                     | Severity | Description   | Based on              | Threshold   |
|---------------------------|----------|---|-----------------------|---|
|                           |          | delays.   |                       |   |
| SIP Node HealthCheck Fail | Critical | <p>SIP Node health level fails for pod {{ \$labels.pod }}.</p> <p>Actions:</p> <ul style="list-style-type: none"> <li>• Check for failure of dependent services (Redis/Kafka/SIP Proxy/GVP/Dial Plan).</li> <li>• Check for Envoy proxy failure, then restart the pod.</li> </ul>   | sipnode_health_level  | SIP Node health level fails for pod {{ \$labels.pod }} for 5 minutes. |
| Kafka not available       | Critical | <p>Kafka is not available for pod {{ \$labels.pod }}.</p> <p>Actions:</p> <ul style="list-style-type: none"> <li>• If the alarm is triggered for multiple services, ensure there are no issues with Kafka. Restart Kafka.</li> <li>• If the alarm is triggered only for pod {{ \$labels.pod }}, check if there is an issue with the pod.</li> </ul> | kafka_producer_state  | Kafka is not available for pod {{ \$labels.pod }} for 5 minutes.      |
| Pod Status Error          | Warning  | <p>Actions:</p> <ul style="list-style-type: none"> <li>• Restart the pod. Check if there are any issues with the pod after restart.</li> </ul>  | kube_pod_status_phase | Pod {{ \$labels.pod }} is in Failed, Unknown, or Pending state.       |
| Pod Status NotReady       | Warning  | Pod {{ \$labels.pod }} is in NotReady   | kube_pod_status_ready | Pod {{ \$labels.pod }} is in NotReady                                 |

| Alert                           | Severity | Description   | Based on   | Threshold   |
|---------------------------------|----------|---|--|---|
|                                 |          | <p>state.</p> <p>Actions:</p> <ul style="list-style-type: none"> <li>Restart the pod. Check if there are any issues with the pod after restart.</li> </ul>  |  | state for 5 minutes.  |
| Container Restarted Repeatedly  | Critical | <p>Container {{ \$labels.container }} was repeatedly restarted.</p> <p>Actions:</p> <ul style="list-style-type: none"> <li>Check if the new version of the image was deployed.</li> <li>Check for issues with the Kubernetes cluster.</li> </ul>  | kube_pod_container_status_restarts_total   | Container {{ \$labels.container }} was restarted 5 or more times total within 15 minutes.   |
| Ready Pods below 60%            | Critical | <p>The number of statefulset {{ \$labels.statefulset }} pods in the Ready state has dropped below 60%.</p> <p>Actions:</p> <ul style="list-style-type: none"> <li>Check if the new version of the image was deployed.</li> <li>Check for issues with the Kubernetes cluster.</li> </ul> | kube_statefulset_status_replicas_ready, kube_statefulset_status_replicas_current | For the last 5 minutes, fewer than 60% of the currently available statefulset {{ \$labels.statefulset }} pods have been in the Ready state. |
| Pods scaled up greater than 80% | Critical | <p>The current number of replicas is more than 80% of the maximum number of replicas.</p> <p>Actions:</p>   | kube_hpa_status_current_replicas, kube_hpa_spec_max_replicas                     | For 5 consecutive minutes, the number of replicas is more than 80% of the maximum number of replicas.                                       |

| Alert                       | Severity | Description  | Based on  | Threshold  |
|-----------------------------|----------|--|---|--|
|                             |          | <ul style="list-style-type: none"> <li>Check if max replicas must be modified based on load.</li> </ul>  |   |  |
| Pods less than Min Replicas | Critical | <p>The current number of replicas is less than the minimum replicas that should be available. This might be because Kubernetes cannot deploy a new pod or pods are failing to be active/ready.</p> <p>Actions:</p> <ul style="list-style-type: none"> <li>If all services have the same issue, then check Kubernetes nodes and Consul health.</li> <li>If the issue is only with the SIP Cluster service, then check pod logs or the deployment manifest/helm errors.</li> </ul> | <p>kube_hpa_status_current_replicas<br/> kubernetes_hpa_spec_min_replicas</p> | <p>For 5 consecutive minutes, the number of replicas is less than the minimum replicas that should be available.</p> |
| Pod CPU greater than 80%    | Critical | <p>Critical CPU load for pod {{ \$labels.pod }}.</p> <p>Actions:</p> <ul style="list-style-type: none"> <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</li> </ul>   | <p>container_cpu_usage_seconds_total<br/> container_spec_cpu_requests</p>     | <p>Container {{ \$labels.container }} CPU usage exceeded 80% for 5 minutes.</p>                                      |

| Alert                       | Severity | Description   | Based on   | Threshold  |
|-----------------------------|----------|---|--|--|
|                             |          | <p>load.</p> <ul style="list-style-type: none"> <li>Collect the service logs for pod {{ \$labels.pod }}; raise an investigation ticket.</li> </ul>  |  |  |
| Pod CPU greater than 65%    | Warning  | <p>High CPU load for pod {{ \$labels.pod }}.</p> <p>Actions:</p> <ul style="list-style-type: none"> <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 for pod {{ \$labels.pod }}; raise an investigation ticket.</li> </ul> | <p>container_cpu_usage_seconds_total</p> <p>container_spec_cpu_period</p>                          | <p>Container {{ \$labels.container }} CPU usage exceeded 65% for 5 minutes.</p>    |
| Pod memory greater than 80% | Critical | <p>Critical memory usage for pod {{ \$labels.pod }}.</p> <p>Actions:</p> <ul style="list-style-type: none"> <li>Check whether the horizontal pod autoscaler has triggered and if the maximum number of pods has been reached.</li> <li>Check Grafana</li> </ul>   | <p>container_memory_working_set_bytes</p> <p>kube_pod_container_resource_requests_memory_bytes</p> | <p>Container {{ \$labels.container }} memory usage exceeded 80% for 5 minutes.</p> |

| Alert                       | Severity | Description   | Based on   | Threshold   |
|-----------------------------|----------|---|--|---|
|                             |          | <p>for abnormal load.</p> <ul style="list-style-type: none"> <li>Restart the service for pod {{ \$labels.pod }}.</li> </ul>   |  |   |
| Pod memory greater than 65% | Warning  | <p>High memory usage for pod {{ \$labels.pod }}.</p> <p>Actions:</p> <ul style="list-style-type: none"> <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 for pod {{ \$labels.pod }}; raise an investigation ticket.</li> </ul> | <p>container_memory_working_set_bytes, kube_pod_container_resource_requests_memory_bytes</p> | <p>Container {{ \$labels.container }} memory usage exceeded 65% for 5 minutes.</p>  |
| Redis not available         | Critical | <p>Redis is not available for pod {{ \$labels.pod }}.</p> <p>Actions:</p> <ul style="list-style-type: none"> <li>If the alarm is triggered for multiple services, ensure there are no issues with Redis. Restart Redis.</li> <li>If the alarm is triggered only for pod {{ \$labels.pod }},</li> </ul>  | <p>redis_state</p>   | <p>Redis is not available for pod {{ \$labels.pod }} for 5 consecutive minutes.</p> |

| Alert   | Severity | Description   | Based on                       | Threshold  |
|---|----------|---|--------------------------------|--|
|   |          | check if there is an issue with the pod.  |                                |  |
| Too many Kafka producer errors                                | Critical | <p>Kafka responds with errors at pod {{ \$labels.pod }}.</p> <p>Actions:</p> <ul style="list-style-type: none"> <li>For pod {{ \$labels.pod }}, ensure there are no issues with Kafka.</li> </ul>   | kafka_producer_error_total     | More than 100 errors for 5 consecutive minutes.  |
| SIP Server main thread consuming more than 65% CPU for 5 mins | Warning  | <p>Main thread consumes too much CPU.</p> <p>Actions:</p> <ul style="list-style-type: none"> <li>Collect SIP Server Main thread logs; that is, log files without index in the file name (appname_date.log files). Raise an investigation ticket.</li> </ul>   | sips_cpu_usage_main            | Main thread consumes too much CPU (more than 65% for 5 consecutive minutes).   |
| Calls activity drop   | Warning  | <p>A noticeable reduction in the number of active calls on a specific SIP Server and no new calls are arriving for processing.</p> <p>Actions:</p> <ul style="list-style-type: none"> <li>If a problematic SIP Server is primary, do a switchover, and then restart the former primary server.</li> <li>If a problematic SIP Server is</li> </ul> | sips_calls, sips_calls_created | The absolute value of active calls on a specific SIP Server dropped by more than 30 calls in 2 minutes and no new calls are arriving at the SIP Server for processing. |

| Alert                 | Severity | Description   | Based on                   | Threshold  |
|-----------------------|----------|---|----------------------------|--|
|                       |          | <p>backup, restart the backup server. Collect SIP Server Main thread logs; that is, log files without index in the file name (appname_date.log files). Raise an investigation ticket.</p>   |                            |  |
| Dial Plan Node Down   | Critical | <p>No Dial Plan nodes are reachable from SIP Server and all connections to Dial Plan nodes are down.</p> <p>Actions:</p> <ul style="list-style-type: none"> <li>• Check the network connection between SIP Server and the Dial Plan node host.</li> <li>• Check the Dial Plan node CPU and memory usage.</li> </ul> | sips_dp_active_connections | All connections to Dial Plan nodes are down  |
| Dialplan Node problem | Warning  | <p>Dial Plan node rejects requests with an error or it doesn't respond to requests and requests are timed out.</p> <p>Actions:</p> <ul style="list-style-type: none"> <li>• Check the network connection between SIP Server and the Dial Plan host.</li> <li>• Check that Dial Plan nodes are running.</li> </ul>   | sips_dp_timeouts           | During 1 minute, the Dial Plan node rejects more than 5 requests with an error or more than 5 requests time out because the Dial Plan node fails to respond. |

| Alert                          | Severity | Description   | Based on              | Threshold   |
|--------------------------------|----------|---|-----------------------|---|
| Routing timeout counter growth | Warning  | <p>The trigger detects that routing timeouts are increasing.</p> <p>Actions:</p> <ul style="list-style-type: none"> <li>Check the URS_RESPONSE_MORE5SEC stat value. If it's increasing, then investigate why URS doesn't respond to SIP Server in time.</li> <li>Check SIPS-to-URS network connectivity.</li> </ul>   | sips_routing_timeouts | The absolute value of NROUTINGTIMEOUTS on a specific SIP Server increased by more than 20 in 2 minutes. |
| SIP trunk is out of service    | Critical | <p>SIP trunk is out of service.</p> <p>Actions:</p> <ul style="list-style-type: none"> <li>For Primary and Secondary trunks: <ul style="list-style-type: none"> <li>Troubleshoot SIP Server-to-SBC network connection. Collect network stats and escalate to the Network team to resolve network issues, if necessary.</li> </ul> </li> <li>Troubleshoot the SBC. For Inter-SIP Server trunks: troubleshoot the SIP Server-to-</li> </ul> | sips_trunk_in_service | SIP trunk is out of service for more than 1 minute.   |

| Alert                            | Severity | Description   | Based on                   | Threshold   |
|----------------------------------|----------|---|----------------------------|---|
|                                  |          | SIP Server network connection. Collect network stats and escalate to the Network team to resolve network issues, if necessary.  |                            |   |
| Media service is out of service  | Critical | <p>Media service is out of service.</p> <p>Actions:</p> <ul style="list-style-type: none"> <li>• Troubleshoot the SIP Server-to-Resource Manager (RM) network connection. Collect network stats and escalate to the Network team to resolve network issues, if necessary.</li> <li>• Troubleshoot RM, consider RM restart.</li> <li>• After 5 minutes, redirect traffic to another site.</li> </ul> | sips_msml_in_service       | Media service is out of service for more than 1 minute. |
| SIP softswitch is out of service | Critical | <p>Actions:</p> <ul style="list-style-type: none"> <li>• Troubleshoot the SIP Server-to-SBC network connection. Collect network stats and escalate to the Network team to resolve network issues,</li> </ul>  | sips_softswitch_in_service | SIP softswitch is out of service.                       |

---

| Alert                       | Severity | Description  | Based on                 | Threshold                    |
|-----------------------------|----------|--|--------------------------|------------------------------|
|                             |          | <p>if necessary.</p> <ul style="list-style-type: none"> <li>• Troubleshoot the SBC.</li> </ul>   |                          |                              |
| SIP Proxy is out of service | Critical | <p>Actions:</p> <ul style="list-style-type: none"> <li>• Troubleshoot the SIP Server-to-SIP Proxy nodes network connections. Collect network stats and escalate to the Network team to resolve network issues, if necessary.</li> <li>• Troubleshoot SIP Proxy nodes.</li> </ul> | sips_sipproxy_in_service | SIP Proxy is out of service. |