



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.

GVP Web Services API

Service Quality Latency Reports

5/10/2025

Contents

- 1 Service Quality Latency Reports
 - 1.1 Web Service Endpoint
 - 1.2 Input Parameters
 - 1.3 Output Format

Service Quality Latency Reports

The Service Quality Latency report describes a summary of the latency data observed from the MCP systems. Each report contains a [manifest section](#).

Web Service Endpoint

This report is available at the web service endpoint `/ems-rs/sqa/latency/details`.

Input Parameters

This report accepts the following Request URI parameters:

Parameter	Description
granularity	<p>This describes the unit of time for this report. It can take one of the following value:</p> <ul style="list-style-type: none">• HOUR• DAY• WEEK• MONTH <p>Each bin of the report will represent a set of latency results for a time period of size 'granularity' if aggregate-by parameter is set to 'time'.</p>
from	<p>This describes the start of the period the report would be for. If the "from" does not align with the granularity, the report will normalize the "from" time to the granularity boundary mark before the specified value. This parameter is required.</p>
to	<p>This describes the end of the period the report would be for. If the "to" does not align with the granularity, the report will normalize the "to" time to the granularity boundary mark after the specified value. This parameter is optional.</p>
comp-id	<p>This describes the target components. This may be specified zero or more times. If specified, the report will contain arrival data only for the specified set of comp-id's. If no comp-id is specified, the report will contain aggregated latency data for all the MCP components in the system.</p>
aggregate-by	<p>This can be specified at most once in the query. It</p>

Parameter	Description
	<p>may be set to one of the following values:</p> <ul style="list-style-type: none"> time component <p>This causes each bin of latency to be an aggregation of either 'time' or 'component'.</p>

Output Format

The output for this report consists of a list of <latencies> XML elements. Each <latencies> element contains a sequence of arrival data in the requested time range, for the specified set of components. Each <latencies> element contains a number of <bin> elements in a sequence, where each <bin> represents the latency data for a unit of time specified by granularity, and it is further broken down into the various latency category and latency types. In the report's manifest, in addition to the general manifest parameters as described in an earlier section, it also contains the latency thresholds configured for the latencies which has been measured. This report conforms to the RelaxNG schema SQUALatencyDetails.rng. [Download the GVP RNG Schemas](#) An example report body for this report is as follows:

```
<latencies comp-type="MCP" start="2010-02-04T16:00:00Z" end="2010-02-04T18:00:00Z"
granularity="HOUR">
  <components>
    <component href="/components/212" type="MCP" />
  </components>
  <bins count="2">
    <bin index="0">
      <category name="FETCHING" measured="210" avg="27.4476" min="0" max="3235" std-
dev="224.1228">
        <latency type="AUDIO_FETCH" measured="28" avg="18.3929" min="0" max="328" std-
dev="61.3154" estimate="118.9501" />
        <latency type="PAGE_FETCH" measured="182" avg="28.8407" min="0" max="3235" std-
dev="239.0884" estimate="420.9456" />
      </category>
      <category name="CALL_CONTROL" measured="92" avg="283.9565" min="15" max="20000" std-
dev="1526.3008">
        <latency type="CALL_ANSWER" measured="90" avg="67.7" min="15" max="3250" std-
dev="339.9106" estimate="625.1533" />
        <latency type="CALL_REJECT" measured="2" avg="10015.5" min="31" max="20000" std-
dev="14120.2153" estimate="33172.6531" />
      </category>
      <category name="RESPONSE" measured="2" avg="281" min="281" max="281" std-dev="0">
        <latency type="CUMULATIVE_RESPONSE" measured="1" avg="281" min="281" max="281" std-
dev="0" estimate="281" />
        <latency type="TRANSFER_RESPONSE" measured="1" avg="281" min="281" max="281" std-
dev="0" estimate="281" />
      </category>
      <category name="PROMPT" measured="272" avg="161.0735" min="0" max="3562" std-
dev="217.6774">
        <latency type="INBOUND_FIRST_PROMPT" measured="90" avg="208.3333" min="0" max="500"
std-dev="99.3131" estimate="371.2069" />
        <latency type="INITIAL_RESPONSE" measured="90" avg="276.0333" min="31" max="3562"
std-dev="364.4363" estimate="873.7089" />
        <latency type="INTERPROMPT" measured="91" avg="2.2308" min="0" max="16" std-
```

```
dev="5.4976" estimate="11.2469" />
  <latency type="OUTBOUND_FIRST_PROMPT" measured="1" avg="16" min="16" max="16" std-
dev="0" estimate="16" />
</category>
<category name="MRCP" measured="270" avg="28.5333" min="0" max="125" std-dev="22.674">
  <latency type="MRCP_SPEAK" measured="90" avg="7.8" min="0" max="16" std-
dev="7.8514" estimate="20.6763" />
  <latency type="MRCP_TTS_SESSION_ESTABLISH" measured="90" avg="70.8778" min="15"
max="125" std-dev="37.6494" estimate="132.6228" />
  <latency type="MRCP_TTS_SET_PARAMS" measured="90" avg="6.9222" min="0" max="16" std-
dev="7.7897" estimate="19.6973" />
</category>
<category name="EXECUTION" measured="74" avg="0" min="0" max="0" std-dev="0">
  <latency type="PAGE_COMPILE" measured="73" avg="0" min="0" max="0" std-dev="0"
estimate="0" />
  <latency type="JAVA_SCRIPT_EXECUTION" measured="1" avg="0" min="0" max="0" std-
dev="0" estimate="0" />
</category>
</bin>
<bin index="1" />
</bins>
</latencies>
```