GVP Deployment Guide

Deploying GVP Multi-Site Environments

4/30/2025

Deploying GVP Multi-Site Environments

This topic describes deployment solutions and configurations for Genesys Voice Platform (GVP) in multi-site or large scale environments. It contains the following sections:

Overview

To ensure your Genesys Voice Platform multi-site environment is secure and functioning efficiently, consider the following key factors:

- A solution that requires the use of a virtual IP address (VIP) across WAN is likely unacceptable, since the use of virtual IPs is better suited to LAN environments.
- In a WAN environment, sites are interconnected by individual links between each site. These individual links can fail and create islands of sites even though locally the site could still be operational.
- Policy enforcement must be consistent across all sites.
- Reporting functions must be consistent across all sites and must be customized to provide individual site reports and multi-site or overall environment reports.
- Resource sharing must be enabled between sites to mitigate failures or if spill-over traffic occurs.
- SIP Server instances within the same site can use all of GVP's available resources within the site.
- A Resource Manager HA configuration is required for multi-site, when integrated with SIP-S.
- In an environment with multi-site deployments, each Resource Manager pair requires a matching pair of Report Servers.

Consider the key factors described in this section when you are planning a multi-site deployment.

Scalability

A single GVP site typically includes a Resource Manager instance (or a High Availability [HA] pair), a Reporting Server instance (or an HA pair), and a pool of Media Control Platform instances. Within a single site, scalability is typically limited by the number of call attempts per second (CAPS) that the

Reporting Server can support.

However, scalability can occur across multiple physical sites, enabling policy enforcement, resource sharing, and reporting across multiple sites. In addition, historical and real-time reports can be filtered to generate site reports (by using site identification) or system-wide reports (no site identification).

Multi-Tenancy

In hosted environments where multiple tenants are deployed, GVP can be deployed across multiple sites and media services for different tenants can be serviced by any one of the sites. There are two requirements for this type of deployment:

- Enforcement of tenant policies must be applied consistently across all the sites. Usage limits for a tenant must be applied globally across all the sites at all times. For example, if a tenant has a usage limit of 100, the maximum number of concurrent calls that can be serviced across all the sites at all times is 100.
- Collection of operational data, such as peak and summary usage, must be aggregated correctly across all the sites and can be filtered on a single-site basis. This applies to both historical and real-time reporting.

Disaster Recovery

Implementation of a disaster recovery (DR) plan is critical in multi-site deployments. It means that one site in the multi-site deployment is designated as the DR site and is enabled and ready to be fully functional in the event that any given site is out-of-service, even if it is out for an extended period of time. In addition, operational data must be replicated to the disaster recovery site.

A DR site deployment enables:

- Servicing of incoming requests at full capacity.
- Access to and reporting of data from the failed site.

For more information about DR sites, see Deploying GVP Multi-Site Environments.

Sites and Segments

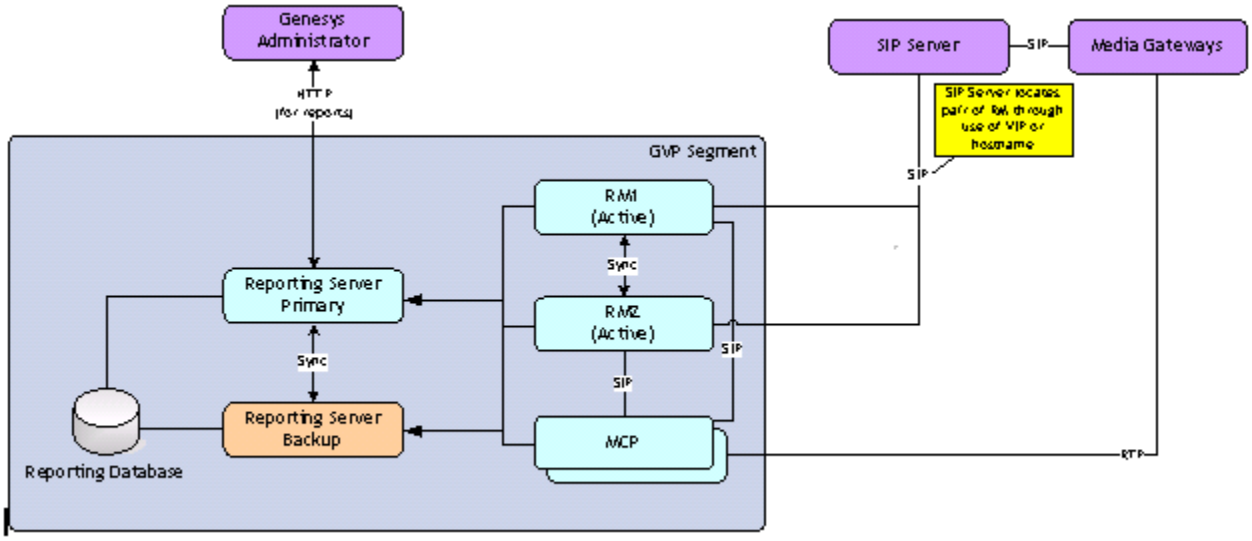
The Reporting Server writes reporting data to the database, which provides single-site and overall system reports, and can prove to be a bottleneck, based on known performance metrics. Therefore, the core design of the solution that is proposed in this section is scaled upward by the deployment of multiple sets of the Reporting Server and multiple sets of the database storage units.

GVP Segment Defined

A GVP segment can be defined as a logical grouping of core components, such as Resource Manager, Reporting Server, and logical resource groups (LRG) that include Media Control Platforms, Call Control Platforms, or CTI Connectors. The figure below depicts the high-level relationship between the core

GVP components.

The components in a segment must be deployed locally in the same site. However, one or more segments can be deployed within a site. Genesys Administrator provides a mechanism to identify and display the segments and sites. In addition, when resource groups are created, the user can choose the segment to which the resource group will belong, rather than the Resource Manager to which it will belong.

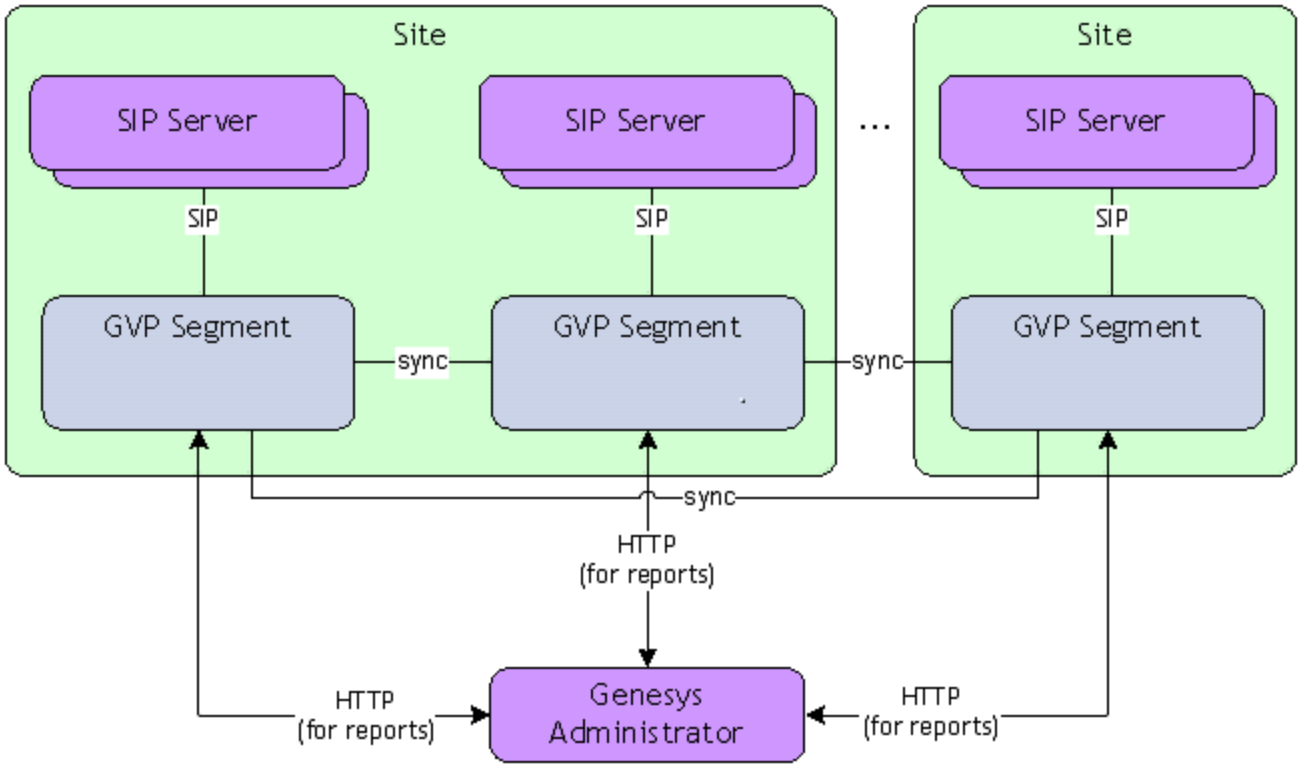


Relationship Between GVP Core Components

Within the segment, one or more SIP Servers can connect to a pair of Resource Manager instances to access media services or through SIP communication, IVR Server. In addition, Genesys Administrator can retrieve historical and real-time reports through the Reporting Server.

The Resource Manager and the Media Control Platform contribute to the events that are logged to the reporting database by the Reporting Server. The Resource Manager and Media Control Platform generate operational data, such as summary and peak information and send it to the Reporting Server.

To make a single site scalable, multiple segments can be deployed within the site. To make multiple sites scalable, segments can be deployed across multiple sites. In this solution, all the segments are considered to be working together as a single large deployment, regardless of the boundaries between them. The illustration below shows an example of how the multiple segments can apply to both a single site or multiple sites.



Segmentation of Single or Multiple Sites

In this solution, three elements enable the GVP segments to work together as a large deployment.

1. Certain components synchronize with all other segments. This is a higher-level synchronization than the synchronization that occurs locally, such as the HA synchronization between two active Resource Manager instances or two Reporting Server instances. In other words, local HA synchronization is designed to ensure continuity of operations for the same component, while synchronization across sites is for elements or counters that are globally shared. In this solution, only the Resource Manager (for policy enforcement) and the Reporting Server (for historical and real-time reporting) require synchronization across sites.
2. The Resource Manager monitors the local segment to determine if the media servers are able to handle all incoming requests or if their capacity is exhausted. If this happens and resource sharing is enabled, the Resource Manager can forward the SIP requests to another segment.
3. Genesys Administrator is the GUI from which aggregated real-time and historical reports from all the Reporting Server instances across all segments can be extracted. All segments in the deployment have a site identifier, which Genesys Administrator uses to aggregate specific reports. The site identifier enables users to generate reports on a per-site basis. In addition, multiple segments can have the same site identifier.

Synchronization Across Sites

This section provides information about how site synchronization, policy enforcement, and resource

sharing occurs across GVP multi-site environments. It also describes segment and network recovery after a failure.

Site Policy Enforcement

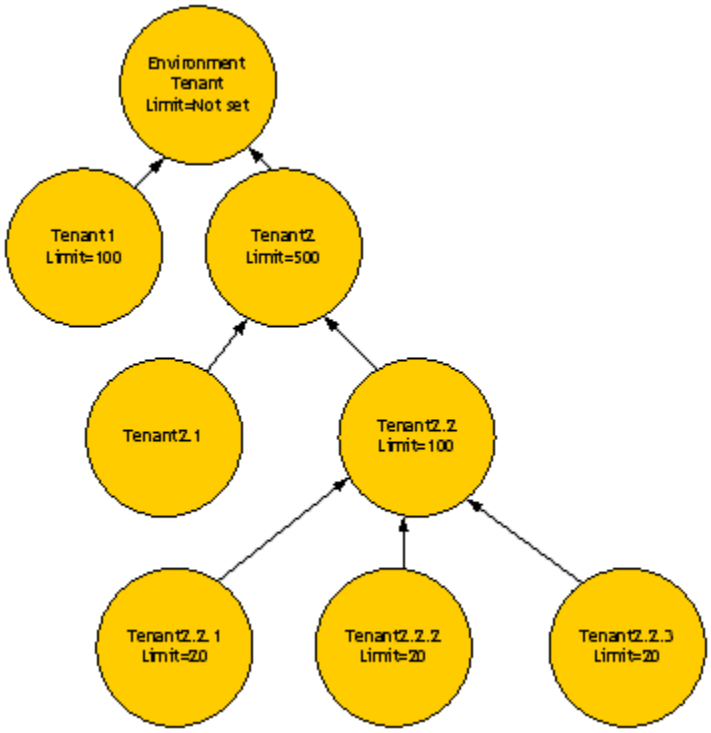
In a multi-site environment, policies are categorized as static or dynamic to help the user understand which policies are dynamically enforced. The Resource Manager enforces mostly static policies, which means it reads the value from Configuration Server to determine how the policy is enforced.

The Resource Manager tracks the current usage limits in memory for dynamic policies. To accurately enforce the policy usage limit for each tenant and IVR Profile and ensure all calls are accounted for, the usage limit is routed through the Resource Manager.

Usage Limit Counters

After the call is established, the Resource Manager stays in the SIP messaging path, so that it can track the time of call termination. The Resource Manager keeps a counter for each tenant and IVR Profile, based on the current usage.

In distributed environments, where the same counter is tracked system-wide, each usage limit counter is subdivided into smaller units and shared with other segments so that each segment can locally track each counter. This process ensures the accuracy of each local counter. For example, in a deployment that has 3 segments and a tenant with a usage limit of 100, the counter for the segments can be subdivided into 40, 40, and 20. INSERT_TEXT The figure below provides an example of a tenant hierarchy and its usage limits.



Tenant Hierarchy and Usage Limits

Role of the Coordinator

The Resource Manager maintains a synchronized connection with other Resource Manager peer instances across multiple segments and uses an election algorithm to select a coordinator. The coordinator's role is to assign usage limit values to each segment for dynamic policies, by dividing each usage limit counter for each tenant. The values are assigned, based on a weight parameter, which defaults to 100 if it is not set.

Election Algorithm

To simplify the election algorithm, each segment is assigned a segment identifier or number. When the Resource Managers (including the active pairs) connect to each other through the synchronization port, they assign a coordinator to divide the counters, based on the segment identifier (together with the local Resource Manager identifier). The sequence of events occurs as follows:

1. A Resource Manager instance (RM) broadcasts an election message to all other connections, which includes the segment identifier.
2. If RM does not hear from other connections within a certain amount of time (for example, 5 seconds), then RM declares victory and broadcasts itself as the coordinator.
3. If RM hears from another Resource Manager instance with a lower identifier, RM waits for the timeout to expire and then listens for the broadcast result. If RM does not hear the result within the timeout, then RM sends the election message again (see Step 1).
4. At the end of the election algorithm, the coordinator broadcasts another message with the division of the counters as the result. By default, the counters are divided evenly among the segments. See the example in the table below.
5. After the counters are divided, the Resource Manager pairs for each segment enforce the usage limits locally.

Table: Division of Counters Among Segments

Tenant	Segment 1 (coordinator)	Segment 2	Segment 3
Environment	Not set	Not set	Not set
T1	34	33	33
T2	167	167	166
T2.1	Not set	Not set	Not set
T2.2	34	33	33
T2.2.1	7	6	6
T2.2.2	7	6	6
T2.2.3	7	6	6

Customizing Division of Counters

The division of counters can be customized by adding a weight value to each segment. Each segment has a weight of 100 by default. The coordinator reads the weight values and uses them to divide the counters. The coordinator adds up all the weight values for all segments, giving counters to each segment in proportion to its own weight. For example, if segments 1, 2, and 3 have weights 300, 100,

50, respectively, segment 1 is assigned 300/450 (or 2/3) of the counters. the table below contains the segments from the table above with adjusted weights factored into the equation.

Table: Customized Division of Counters With Weights

Tenant	Segment 1 coordinator (weight 300)	Segment 2 (weight 100)	Segment 3 (weight 50)
Environment	Not set	Not set	Not set
T1	67	22	11
T2	334	111	55
T2.1	Not set	Not set	Not set
T2.2	67	22	11
T2.2.1	14	4	2
T2.2.2	14	4	2
T2.2.3	14	4	2

Segment Failure

If a Resource Manager instance detects a peer segment failure (both Resource Manager instances within the segment fail), the surviving segments re-issue the election process to elect a new coordinator. The new coordinator then distributes the counters, recalculated with the remaining weights, among the remaining segments after the election algorithm is complete. In the table below, see how the counters are divided when segment 3 fails and segment 1 is elected as the coordinator.

Table: Division of Counters With Weights After Re-Election

Tenant	Segment 1 coordinator (weight = 325)	Segment 2 (weight =125)	Segment 3 (down)
Environment	Not set	Not set	--
T1	75	25	--
T2	375	125	--
T2.1	Not set	Not set	--
T2.2	75	25	--
T2.2.1	15	5	--
T2.2.2	15	5	--
T2.2.3	15	5	--

Segment Recovery

When a Resource Manager instance recovers and re-joins the system, no action is taken if the instance belongs to an active segment. When a Resource Manager instance from a failed segment recovers and re-joins the system, the existing segments re-issue the election process to elect a new coordinator. The new coordinator distributes the counters among the remaining segments, recalculating them with the remaining weights.

New Segments Joining

If a new segment joins the system, the counters are further sub-divided among the segments. It is possible for some segments to have more existing calls than the new usage limit. The Resource Manager does not attempt to drop calls because of the new (and lower) limit, but allows over-usage temporarily.

> **Tip**
> When the Resource Manager allows over-usage temporarily, any new incoming calls are rejected until a sufficient number of existing calls drop, and bring the current usage lower than the new usage limit.

Network Disconnections

> **Tip**
> When the Resource Manager allows over-usage temporarily, any new incoming calls are rejected until a sufficient number of existing calls drop, and bring the current usage lower than the new usage limit.

A WAN link failure can create islands of segments that can operate independently. A disconnection from the WAN link is treated as a disconnected segment by the surviving segments. The separate islands independently issue the election process to find a new coordinator. When this happens, the system has two full sets of usage limits because the islands do not see each other. The table below provides results when Site A and B are disconnected from each other.

Table 44: Island Segments After a WAN Link Failure

	Site A (Island 1)		Site B (Island 2)
Tenant	Segment 1 coordinator (weight = 300)	Segment 2 (weight = 100)	Segment 3 (weight = 50)
Environment	Not set	Not set	Not set
T1	75	25	100
T2	375	125	500
T2.1	Not set	Not set	Not set
T2.2	75	25	100
T2.2.1	15	5	20
T2.2.2	15	5	20
T2.2.3	15	5	20

Network Recovery

When the network is recovered, the segments in the system are re-connected. The segments issue another election process to find a new coordinator. Similar to segment recovery, some segments might end up with more existing calls than the new usage limit. The Resource Manager does not attempt to drop calls because of the new (and lower) limit, but allows over-usage temporarily.INSERT_TEXT

> **Tip**
>
> When the Resource Manager allows over-usage temporarily, any new incoming calls are rejected until a sufficient number of existing calls drop, and bring the current usage lower than the new usage limit.

GVP Multi-Site Reporting

In GVP multi-site environments, Reporting Server instances in each GVP segment collect data. Genesys Administrator queries the Reporting Server instances to generate historical and real-time reports on a per-site or system-wide basis. Genesys Administrator can access site information from Configuration Server by checking the gvp.site folder in the Annex section of the `Provisioning > Environment > Applications` folder. It reads the site information at the user session logon and retains it throughout the session.

The following reports can be aggregated to provide system-wide data:

- IVR Profile Call Arrival
- Component Call Arrival
- Tenant Call Arrival
- VAR IVR Action Usage
- ASR/TTS Usage
- All VAR Summary reports

In a multi-site environment, GA/GAX Plug-in can generate aggregated Call Arrivals, VAR IVR Action, and VAR Last IVR Action reports across multiple sites. The aggregation is based on the same time period and same query parameters. For example, if the report requested is a daily Call Arrivals report on Feb. 2 using IVR Profile A, then, all call arrivals data on the same date (Feb. 2) using the same IVR profile (IVR Profile A) from multiple RS are summed up and generated into a total call arrivals report.

Call Peaks data are reported individually by RM, so GA/GAX plug-in will only generate Call Peaks reports one site at a time.

In a multi-site environment, one pair of RS is needed for every pair of RM. RMs use the same set of

tenants and IVR Profiles. RS has three call peaks reports: queried by components (RM, MCP, and so on), queried by tenant, queried by IVR Profile. RS can only distinguish call peaks from which RM when the call peaks report is queried by RM component.

The call peaks reports queried by IVR Profile and by tenant will ignore which RM the call peaks data come from.

Genesys Administrator Reporting Interface

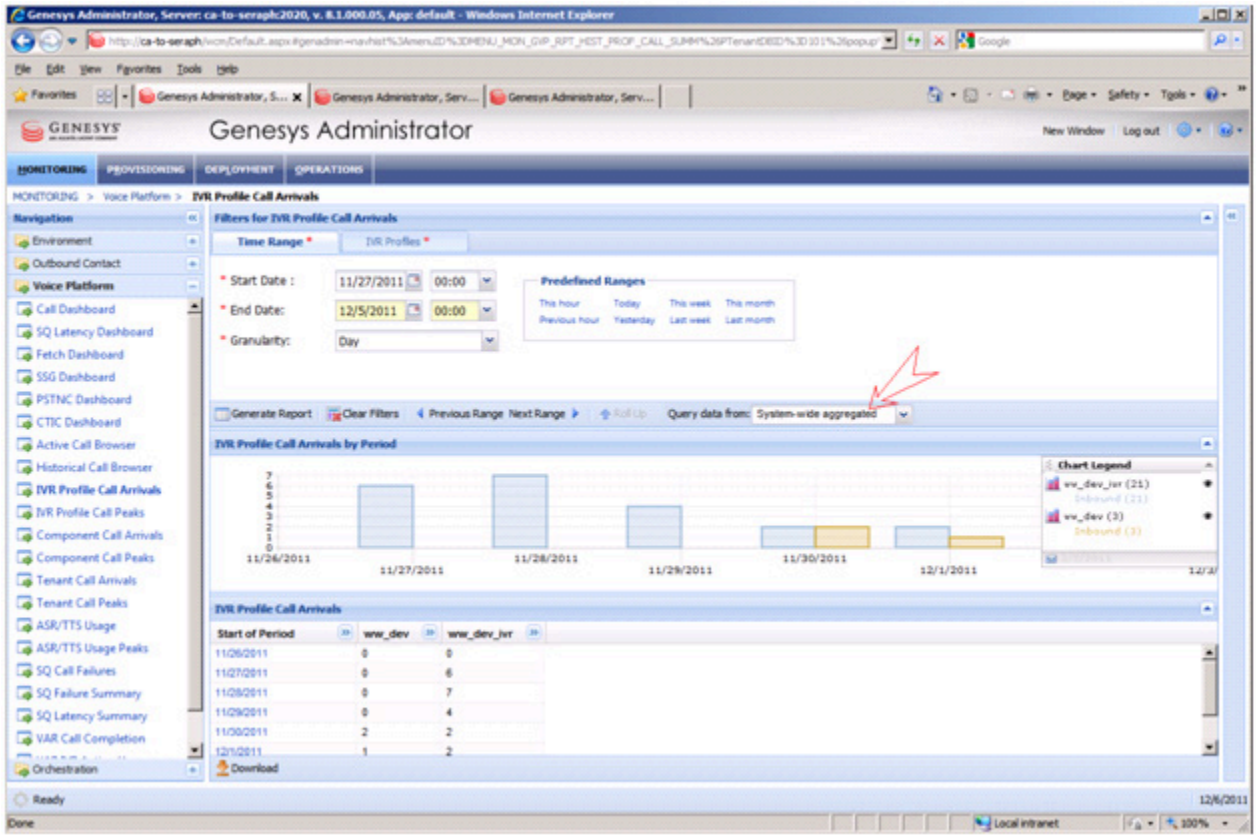
This section describes the additional menu selections that are displayed on Genesys Administrator's Monitoring tab when GVP is deployed in a multi-site configuration.

Operational Reports

IVR Profile Call Arrivals, Tenant Call Arrivals, and ASR/TTS Usage Reports

On the **Generate Report** bar of these report pages of the GUI, you can choose a site from a drop-down list. This drop-down list is displayed when there are multiple sites configured in Configuration Server, whether a Reporting Server is present or not in Genesys Administrator's **Application Connections** settings.

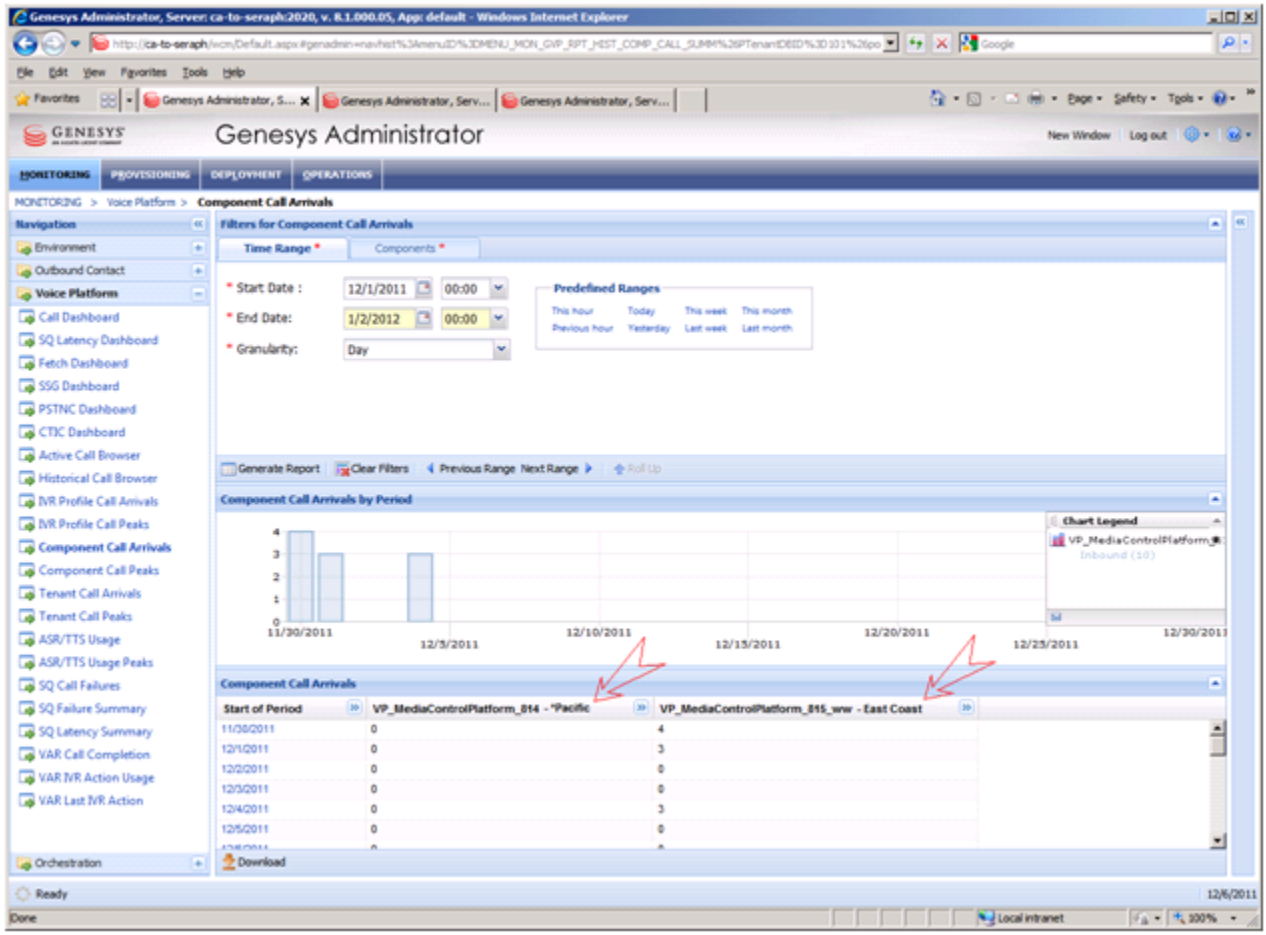
The default selection is `All-Sites`. When this option is selected, the arrivals data must be summarized across all sites on the selected IVR Profiles.



Query Date From Field Genesys Administrator

Component Call Arrivals and ASR/TTS Usage Reports

On the Component Call Arrivals section of these report pages in the GUI, the site name is appended to the component name to indicate the site to which the component belongs. See the figure below.



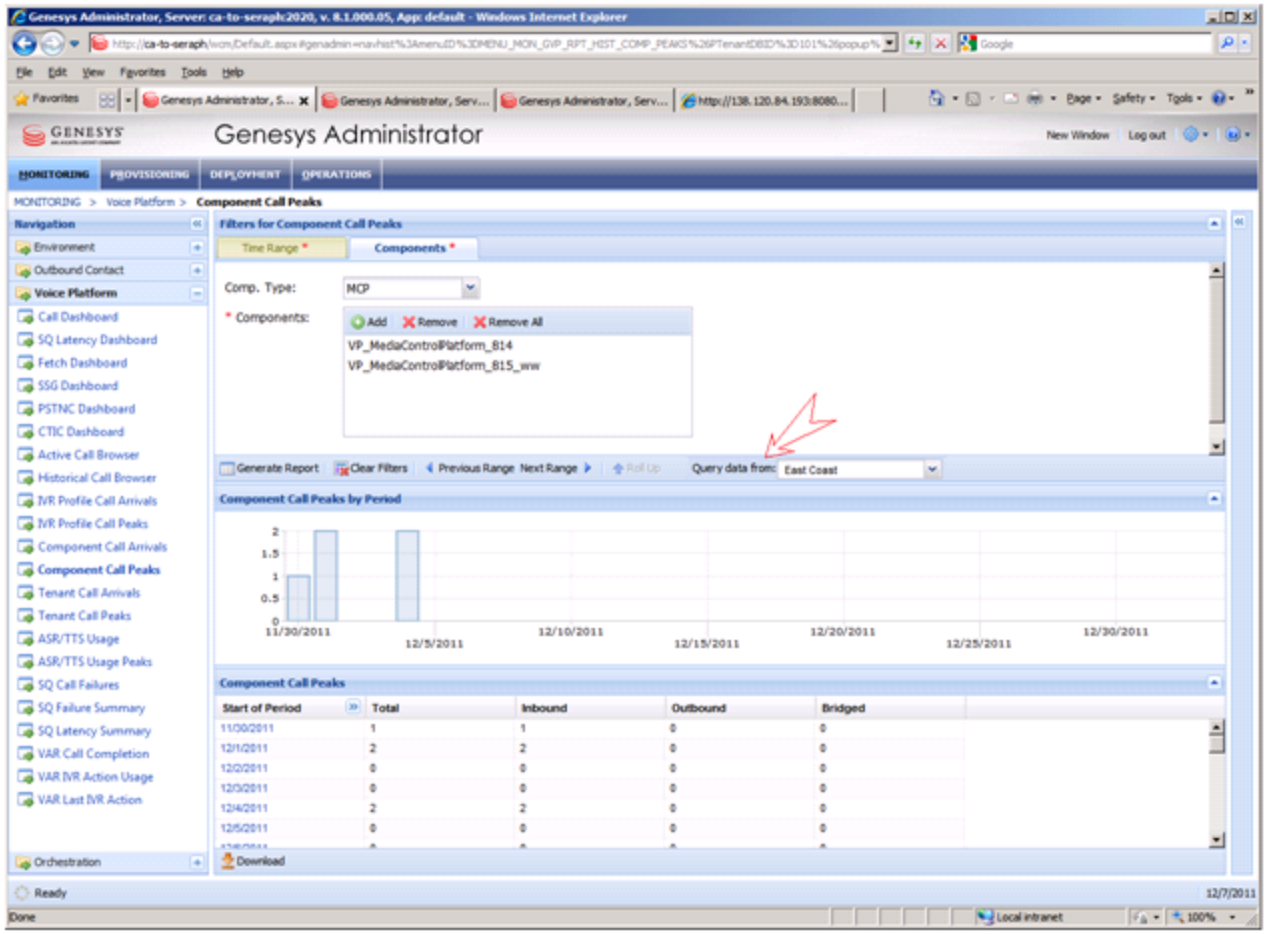
Component Call Arrival Genesys Administrator

IVR Profile Call Peaks, Component Call Peaks, Tenant Call Peaks, and ASR/TTS Usage Peaks Reports

On the **Generate Report** bar of this report GUI, you can choose a site from a drop-down list. The drop-down list does not include the `All-Sites` option.

The default selection is the first site that appears at the top of the drop-down list. See the figure below.

> **Tip**
>
> Currently, Peaks reports display only one selection.



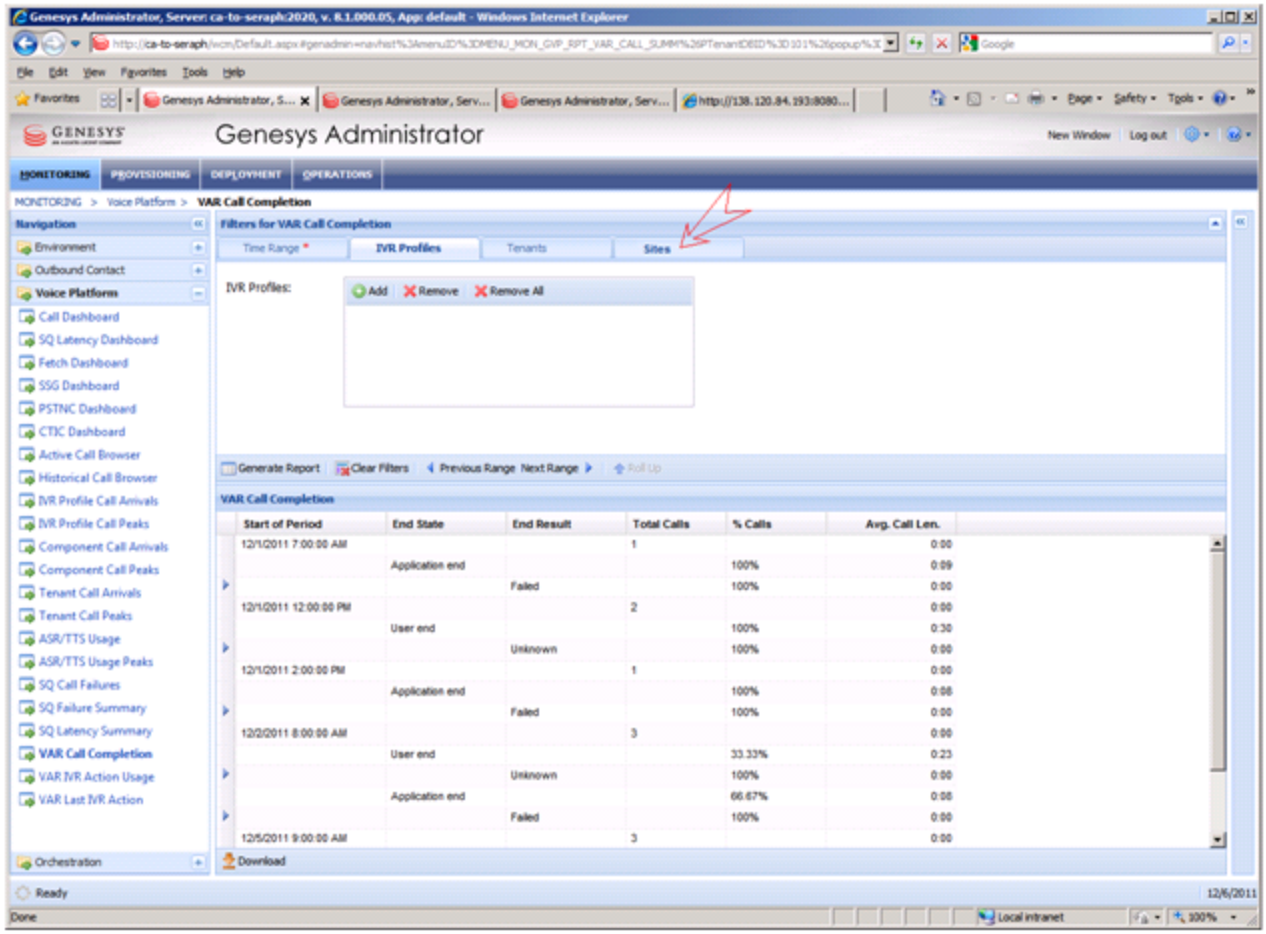
Component Call Peaks Genesys Administrator

VAR Reports

VAR Call Completion Report

On the **Filters for VAR Call Completion** section of this report page of the GUI, you can choose from a selection of sites from a drop-down list. See the figure below.

The default selection is `All-Sites`. When this option is selected, data is merged from the values that are returned from multiple queries.

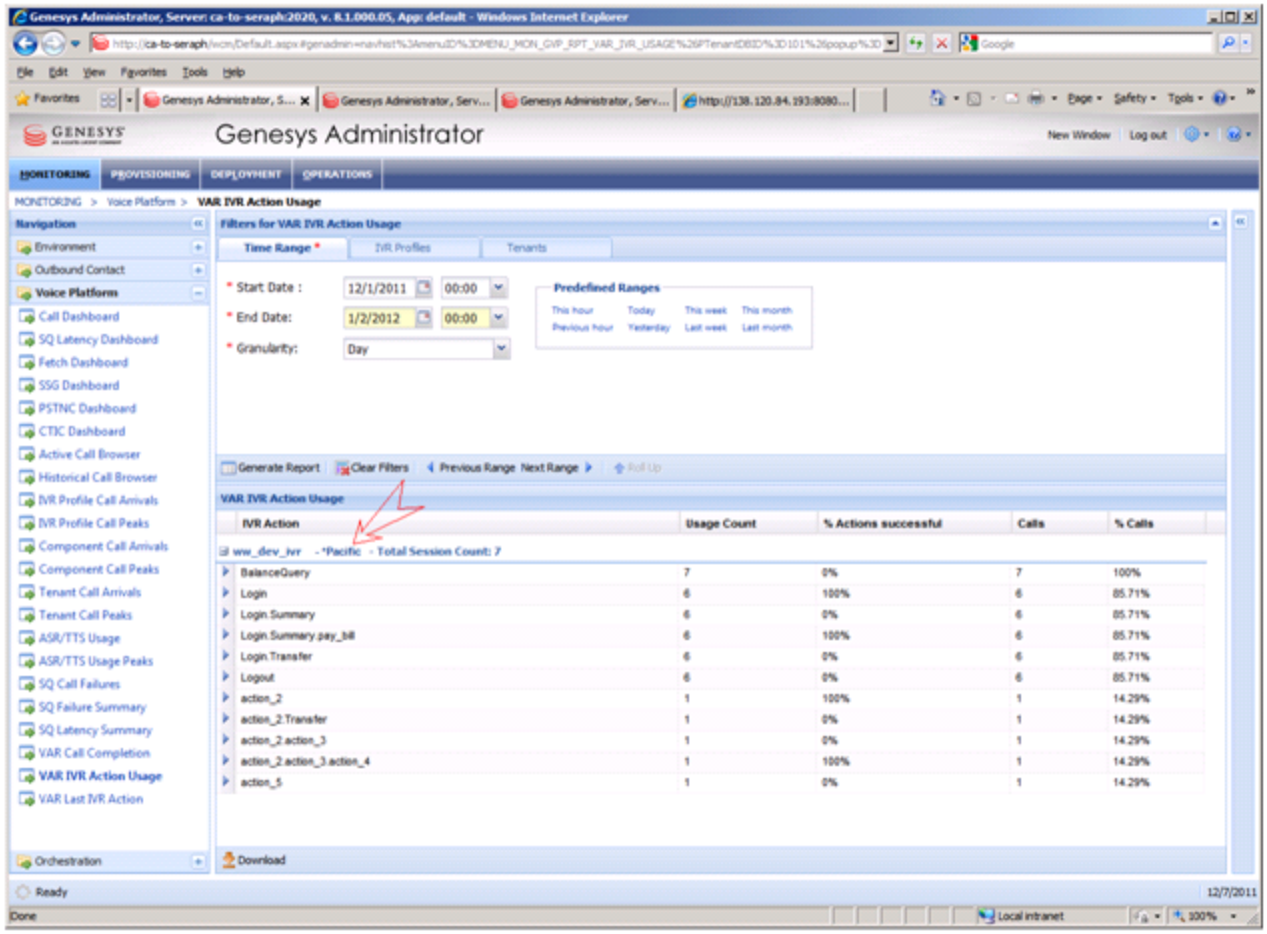


VAR Call Completion Genesys Administrator

VAR IVR Action Usage Report

On the **VAR IVR Action Usage** section of this report page in the GUI, you can choose from a selection of sites from a drop-down list. See the figure below.

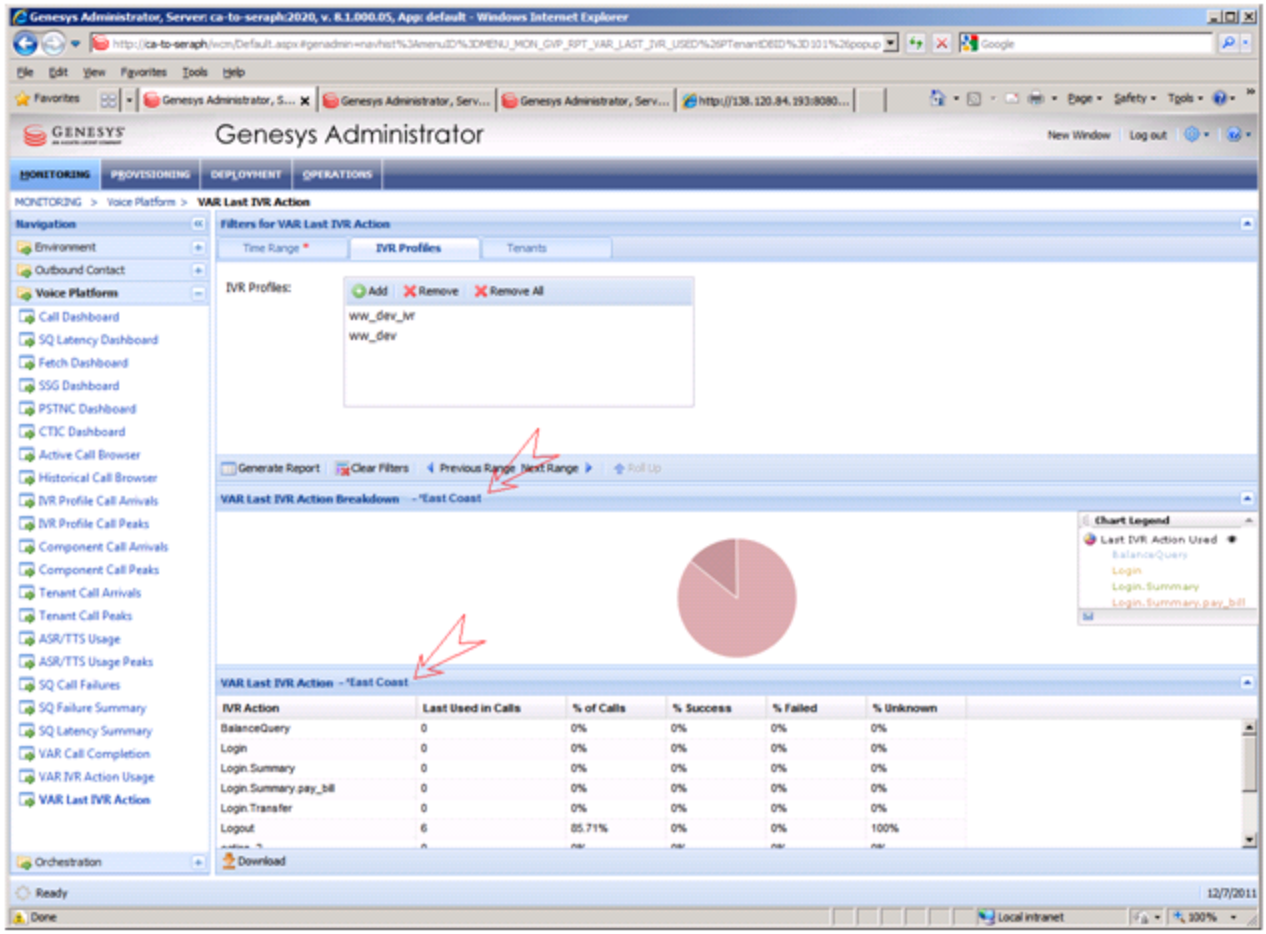
The default selection is `All-Sites`.



VAR IVR Action Usage Genesys Administrator

On the **VAR Last IVR Action Breakdown** and **VAR Last IVR Action** sections of this report page in the GUI, you can choose from a selection of sites from a drop-down list. See the figure below.

The default selection is `All-Sites`.



VAR Last IVR Action Genesys Administrator

Deploying Disaster Recovery Sites

A Disaster Recovery (DR) deployment is basically organized as two regular segments, with the capacity to replicate reporting data between them to ensure operational reports are available when a segment fails. **Structure of a Disaster Recovery Segment** (below) depicts the structure of a DR segment.

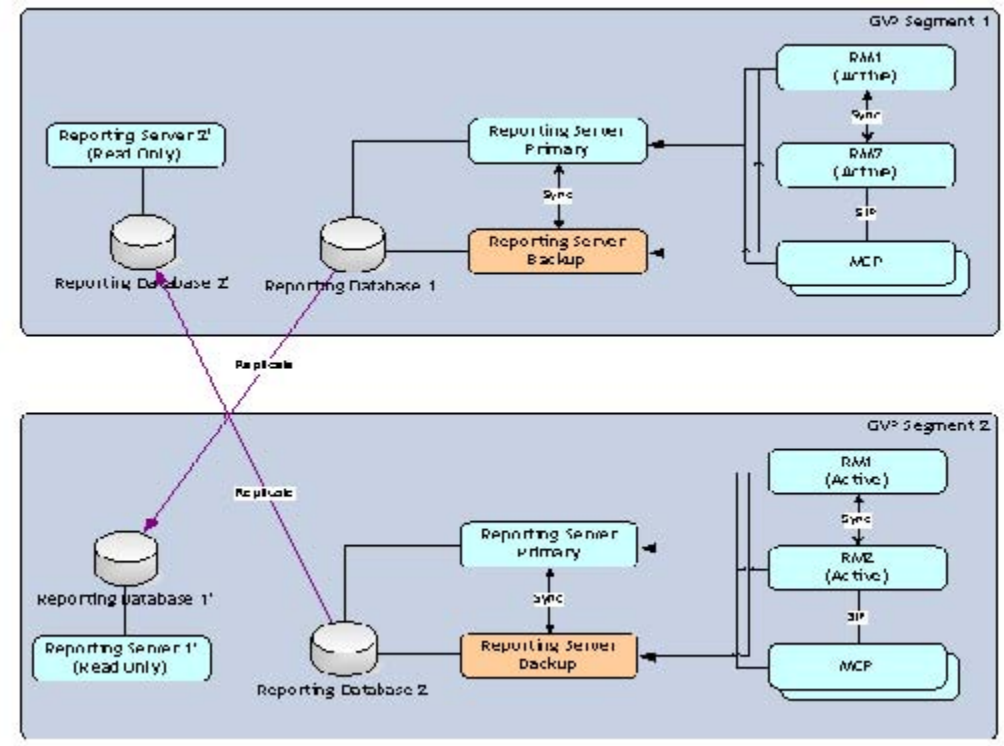
The Disaster Recovery deployment consists of a Primary (or hot) site and a DR (or cold) site. The Reporting Server instance at the Primary site is configured to accept incoming calls. Its `rs.histonly.enabled` configuration option is set to `false`.

- The Reporting Server instance at the DR site is configured so that it does not accept incoming calls until another DR site is no longer reachable. Its `rs.histonly.enabled` configuration option is set to `true`.

To deploy the DR site, create a replicated reporting database and a read-only Reporting Server instance so that the data is replicated during runtime. This ensures that the replicated instance has the most recent data available at all times.

Only the Primary sites are included in the **Site** drop-down list and **All Sites** queries in Genesys Administrator.

> **Tip**
>
> In this release, the Primary (hot site) and DR (cold site) configuration is supported by the Reporting Server only (not Resource Manager).



Structure of a Disaster Recovery Segment

Role of the Read-Only Reporting Server in Disaster Recovery

The role of the read-only Reporting Server is to provide operational reports only. These servers stay active despite the fact that the segment might be a cold DR site, so that operational reports are available in both sites at all times. In this way, Genesys Administrator can automatically select one of the working Reporting Server instances to retrieve operational reports.

Add the DR Reporting Server Connection

Go to the primary Reporting Server's connections tab and add a connection to the DR Reporting Server.

Reporting Data Queries

In DR deployments, Genesys Administrator queries data from the Reporting Server in the following order:

For the historical data:

- HA Primary Reporting Server
- HA Backup Reporting Server (that synchronizes with the HA Primary Reporting Server)
- Read-only DR Reporting Server (that synchronizes with the HA primary Reporting Server)

For real-time data:

- HA Primary Reporting Server
- HA Backup Reporting Server (that synchronizes with the HA Primary Reporting Server)

For aggregate data:

- The queries are repeated *n* times for *n* Primary sites, where *n* is the number of Primary sites. For example, if there are 10 Primary sites, the query is repeated 10 times.

Disaster Recovery Modes of Operation

A DR site or segment can be running in hot or cold mode. A hot DR segment acts as a normal site and participates in the election process to accept incoming calls. A cold DR segment does not process incoming calls and starts as passive. The Resource Manager instances in a cold DR segment still connect to the other segments and participate in the election algorithm. As long as all of the hot DR sites are running and active, the coordinator does not assign usage limits to the cold DR site. If one of the hot DR sites goes down, the cold DR site becomes active and the coordinator assigns usage limits to the (now active) cold DR site.

After a cold DR site becomes active and begins accepting calls, the site does not become passive again until all hot DR sites are back online and active. The coordinator can assign a zero usage limit to a cold DR site, once it determines all hot DR sites are back online. The cold DR site can then become passive again.