Google Search Appliance

Configuring Search Appliances for Load Balancing or Failover

Google Search Appliance software version 7.2 and later

Google

Google, Inc.
1600 Amphitheatre Parkway
Mountain View, CA 94043

Contents

Configuring Search Appliances for Load Balancing or Failover

This document contains the information that you need to set up two instances of the Google Search Appliance for load balancing or fail over.

About Load Balancing and Failover

You can configure two search appliances to provide load balancing or failover. To use a load-balancing configuration, you must purchase two licensed search appliances. You can use a hot backup search appliance in an active/active configuration provided you are able to fail over completely to a single search appliance if a single hardware failure occurs.

Load balancing distributes network traffic of different types to the appropriate applications. Load balancing can be used to distribute network traffic of a particular type to two or more instances of an application, dividing the work load between the instances. A load balancer is a software or hardware application that distributes network traffic. When you configure two Google Search Appliance systems for load balancing, search queries are distributed between the two systems.

Failover configurations typically involve two instances of an application or a particular type of hardware. The first instance, sometimes called the primary instance, responds to search queries. If the first instance fails, the second instance, sometimes called the secondary or standby instance, starts responding to search queries.

This document provides the information you need to create load balancing or failover configurations using two Google Search Appliance systems.

Google does not recommend specific load balancers to use with the search appliances. The configurations described in this document are expected to work with any equipment that complies with networking RFCs.

In load balancing configurations, Google recommends that you set up the load balancer to use *sticky sessions*, in which all search queries from a particular user are directed to the same search appliance. Sticky sessions ensure that each client receives consistent results. Some load balancers use sticky sessions as the default setting. See the load balancer's documentation for complete information.

Configuration Overview

You can set up the search appliances in the following configurations:

- A load balancing configuration in which there is a physical connection between the search appliances and the load balancer and each search appliance is on the same network or subnet as the load balancer.

- A load balancing configuration in which there is a logical connection to the load balance and each search appliance is potentially on different networks or subnets from the load balancer

- A failover configuration in which a switch fails over search queries from the search appliance that normally responds to search queries to a search appliance that does not normally respond to search queries and is used only for failover

You can use these configurations regardless of content location.

The following sections discuss the three configurations in detail and include diagrams each configuration. The diagrams use graphics of the Google Search Appliance model GB-1001, but any of the configurations can be used with other Google Search Appliance models.

Each configuration includes the following:

- Two Google Search Appliance systems.

- A border router, which is the source of queries. A border router is typically between your network and the Internet.

- A load balancer.

- The indexed content, which can be in any location the search appliance can crawl and index.

The computer in the illustrations represents an end point of the required network path to the content.

If you use connectors to traverse content for your search appliances, you must set up a separate connector manager for each Google Search Appliance in your load balancer configuration. This is because, by default, you cannot multiplex connections from a single connector manager to multiple appliances. A benefit of setting up separate connector manager instances for each search appliance is that you have no single point of failure for the connector at serve time.

Load Balancing with a Physical Connection

This configuration provides load balancing by distributing incoming search queries to the search appliances. The search appliances each respond to about half of incoming search queries.

The search appliances are set up with a physical connection between the two search appliances and the load balancer, or with the search appliances located on the same network or subnet as the load balancer.

The following diagram illustrates the components in this configuration. The legend that follows the diagram describes the components and lists the IP address of each component.
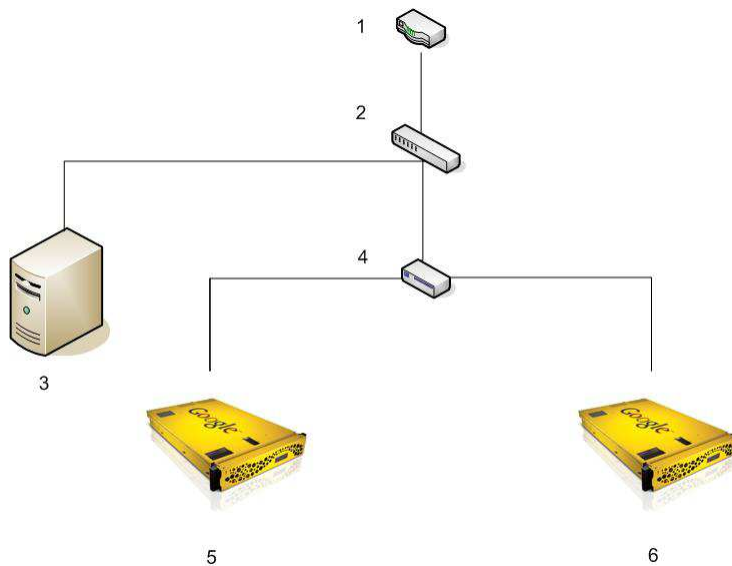
1. Border router at IP address 192.168.0.1

2. Switch at IP address 192.168.0.2

3. Content files on a host at IP address 192.168.0.10

4. Load balancer at two IP addresses: 192.168.0.3 externally and 192.168.1.1 internally

5. First Google Search Appliance at IP address 192.168.1.4.

6. Second Google Search Appliance at IP addresses 192.168.1.7.

Required Equipment

This configuration requires a load balancer and switch.

Benefits

This configuration is easy to set up.

If one search appliance goes off line, the load balancer directs all traffic to the remaining search appliance.

Considerations

During the crawl process, crawl traffic is directed through the load balancer, which creates the potential for network saturation. Therefore, it's important that you schedule the crawl on the two search appliances to take place at different times of day.

The load balancer is a single point of failure in this configuration. If the load balancer fails, you must have physical access to the load balancer and search appliances to restore search capabilities, because the load balancer must be fixed or the IP addresses of the search appliances must be changed.

Load Balancing with a Virtual Connection

This configuration provides load balancing by distributing incoming search queries to the search appliances. The search appliances each respond to about half of incoming search queries.

The search appliances are logically downstream from the load balancer, but potentially on different networks or subnets from the load balancer.

The following diagram illustrates the components of a virtual load balancing configuration. The legend that follows the diagram describes the components and lists the IP address of each component.
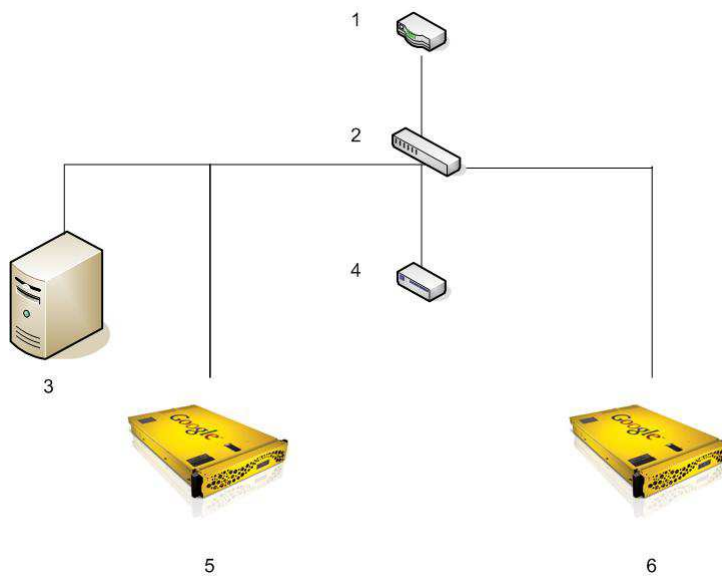
1. Border router at IP address 192.168.0.1

2. Switch at IP address 192.168.0.2

3. Content files on a host at IP address 192.168.0.10

4. Load balancer at IP address 192.168.0.3 with traffic proxy to .4 and .7

5. First Google Search Appliance at IP address 192.168.0.4. The search appliance can be located externally.

6. Second Google Search Appliance at IP address 192.168.0.7. The search appliance can be located externally.

Required Equipment

This configuration requires a load balancer and switch. The load balancer must be able to proxy traffic to external virtual IP addresses.

Benefits

In this configuration, there is no single point of failure.

Crawl traffic is not directed through the switch, so there is less risk of network saturation than in the first configuration.

If one search appliance goes offline, the load balancer directs all traffic to the remaining search appliance.

Considerations

This configuration requires a load balancer that supports load balancing or traffic proxying.

Access control lists (ACLs) for the router are more complex than in other configurations, because you must create rules for additional IP addresses.

Because of how search queries are handled, there is twice as much network traffic between the load balancer and the switch as in other configurations.

DNS Switchover

This configuration uses two search appliances configured as primary and second appliances. All queries are directed to the primary search appliance. If the primary appliance becomes unavailable, search fails over to secondary search appliance.

The following diagram illustrates the components of configuration supporting DNS switchover. The legend that follows the diagram describes the components and lists the IP address of each component.
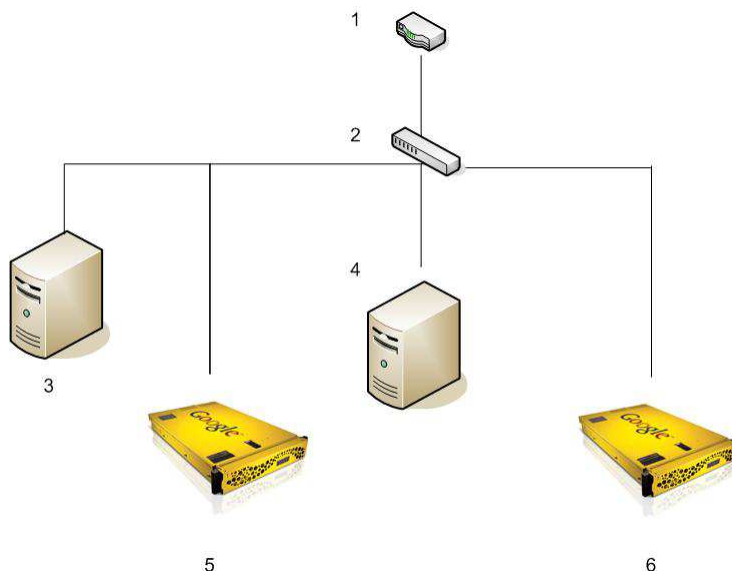
1. Border router at IP address 192.168.0.1

2. Switch at IP address 192.168.0.2

3. Content files on a host at IP address 192.168.0.10

4. DNS at IP address 192.168.0.3

5. First Google Search Appliance at IP address 192.168.0.4. The search appliance can be located externally.

6. Second Google Search Appliance at IP addresses 192.168.0.7. The search appliance can be located externally.

Required Equipment

This configuration requires a DNS server and switch.

Benefits

This configuration is easy to configure and has no special hardware requirements. There is no risk of network saturation during crawls The two search appliances can be located anywhere, physically or logically.

Considerations

This is not a true load balancing configuration. The configuration provides only failover.

Monitoring the Status of the Configuration

In all of these configurations, the load balancer needs to monitor the status of the search appliances. If a search appliance stops running, the load balancer stops sending requests to the search appliance.

To monitor the status of the search appliance, configure the load balancer to send periodic search requests to the search appliance. After each search request, close the connection to the search appliance. See the documentation for your load balancer for information on how to send HTTP requests from the load balancer to the search appliance.

Do not configure the load balancer to monitor the status of a search appliance by sending TCP packets to port 80 of the search appliance, making a connection, and sending a reset. Using TCP packets this way can cause a search appliance to become unresponsive.