Load Balanced Flow Collection

highlight

Ensuring scalable centralised flow data collection

Highlight's flow collection agent is designed to be deployed by managed service providers on a customer by customer basis. This involves installation inside a customer's network with the option for multiple collectors per customer. For larger partners, this is not a scalable solution and requires a bespoke installation every time. A solution is to deploy flow collectors in the core of their customer's networks to standardise the installation but this brings additional challenges:

- Capacity management of each collector
- Configuration management to move from one collector to another
- No resiliency with extended outage if hardware fails

Highlight have successfully set up and tested a load balanced solution to address the challenges of multiple flow collectors.

Why load balance?

Scalability

Each flow collector has a recommended capacity for flow collection, depending on the number of devices connected and the amount of data collected. Once the thresholds are reached additional flow collectors need to be manually provisioned and all new devices configured to point to an additional collector. **This is not a scalable solution.**

Resilience

In the event of a flow collector failing, all flow data is lost until either the flow collector is repaired or a new flow collector is provisioned. **This may result in extended outages and loss of flow data.**

Management

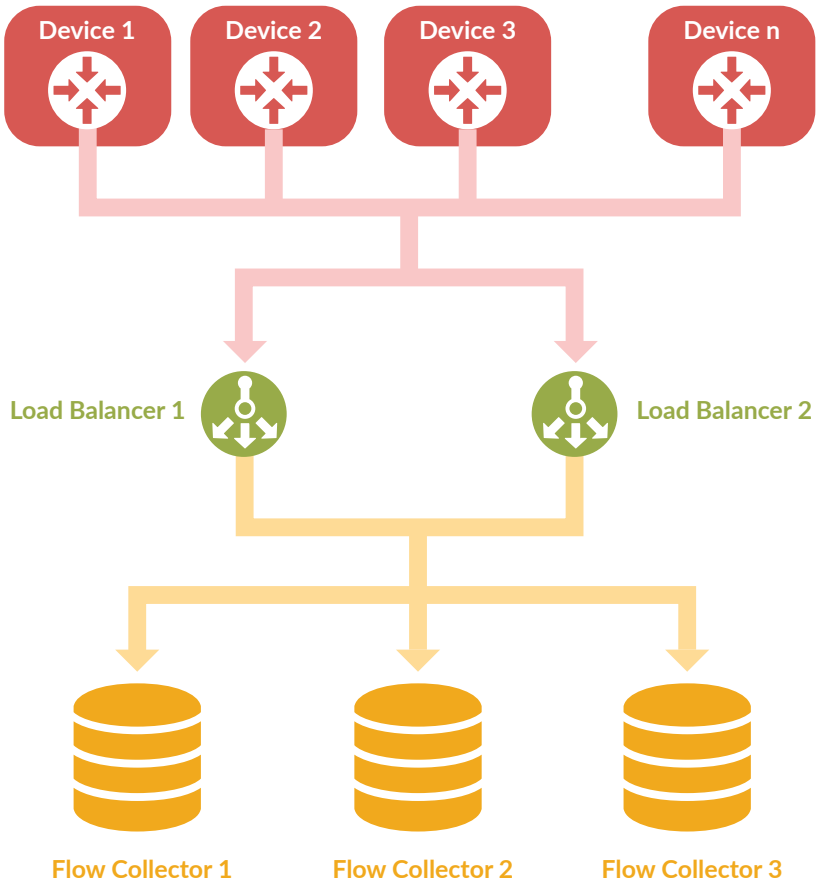
The addition and removal of flow collectors is a manual process, involving not only configuring the collector, but reconfiguring all the flow devices. **This is time-consuming and inefficient.**

The optimal solution

Highlight set and achieved the following goals:

- All flow data from one device must always be sent to the same flow collector to ensure reliable and consistent data, this is known as a sticky connection
- Flow is evenly balanced between the collectors when multiple flow collectors are connected to two load balancers
- All flow data is kept and the sticky connection is retained when a load balancer fails or a flow collector is rebooted
- Flow is redistributed evenly between the remaining collectors and a new sticky connection is established when one of the collectors is unresponsive for 10 minutes
- Any flow collector is treated as new when it comes back online after 10 minutes

Recommended architecture and configuration



A pair of industry standard load balancers were used to manage the flow traffic, configured as an active-passive clustered pair. A single virtual IP address was assigned to the load balancer cluster with all flow traffic sent to that IP address.

Each load balancer was configured with:

- **Full proxy** - traffic is sent to each flow collector by altering the MAC address on each flow packet
- **Least cons** - all incoming flow traffic is balanced between the available flow collectors, sending new flows to the collector with the least number of connections, known as least cons
- **Health check or monitoring** - ICMP pings are sent to determine the status of each flow collector
- **Persistence or affinity**- a sticky connection is retained to the same flow collector unless the device stops sending flow data for six hours
- **High availability (HA)** - all flows redistributed to other collectors if a flow collector is unavailable for more than 10 minutes

The standard, recommended flow collector configuration was used and no changes were made to the configuration of Highlight during the setup and testing of a load balanced solution.

The collection of flow data in Highlight can be scalable, resilient and easily managed using centralised industry standard load balancers, as shown with our recommended architecture above.