CISCO
SECURE

Cisco Secure Network Analytics Data Store Design Guide

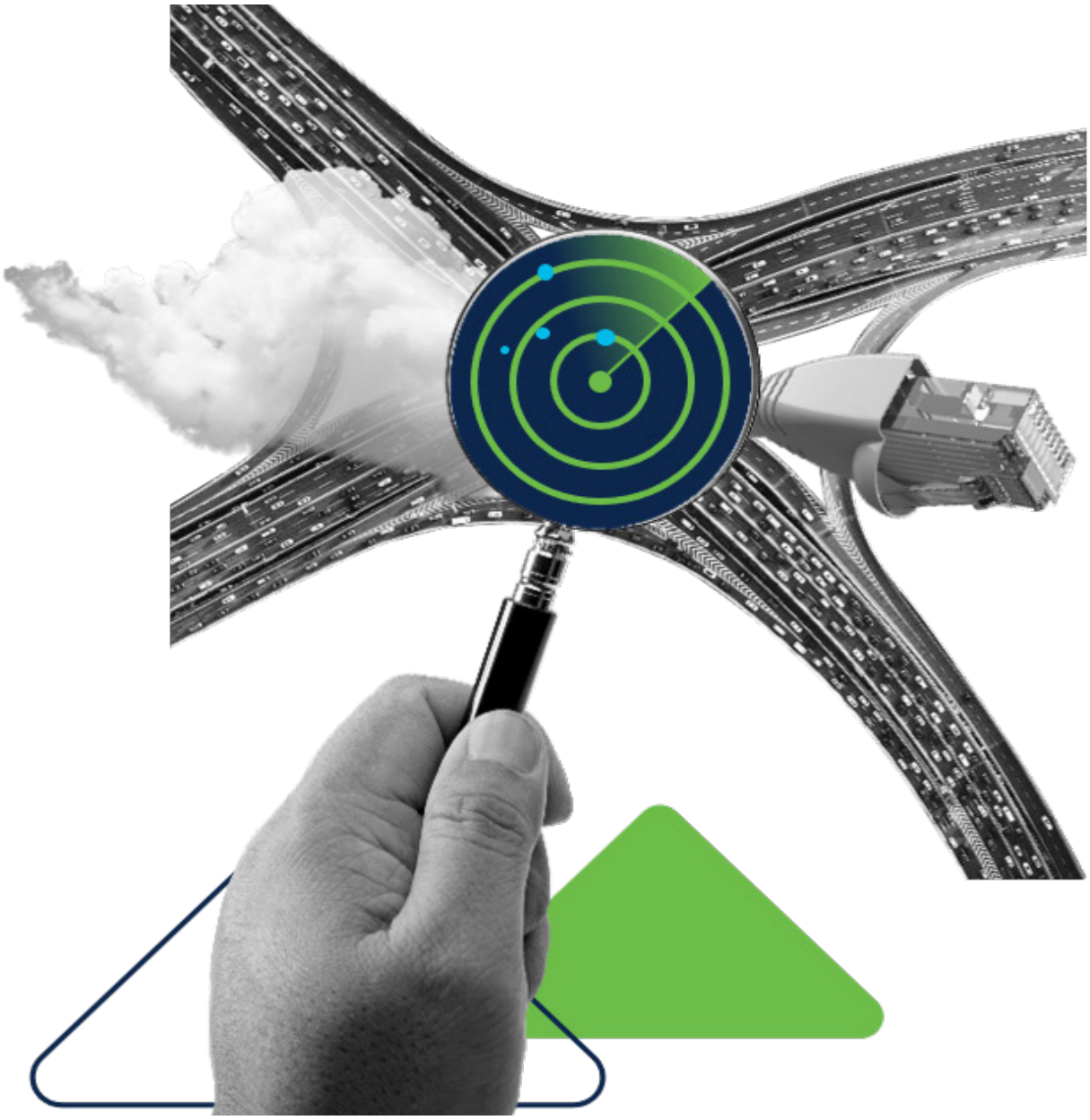


Table of Contents

Audience

The audience for this document includes sales engineers, field consultants, professional services specialists, IT and security managers, partner engineers, and customers who want to take advantage of an infrastructure built to deliver efficiency and enable innovation.

Purpose of this Document

This document provides the fundamental design principles needed to understand how the Secure Network Analytics (formerly Stealthwatch) Data Store scales telemetry consumption, provides data resiliency, and increase search performance for the most demanding enterprise class environments.

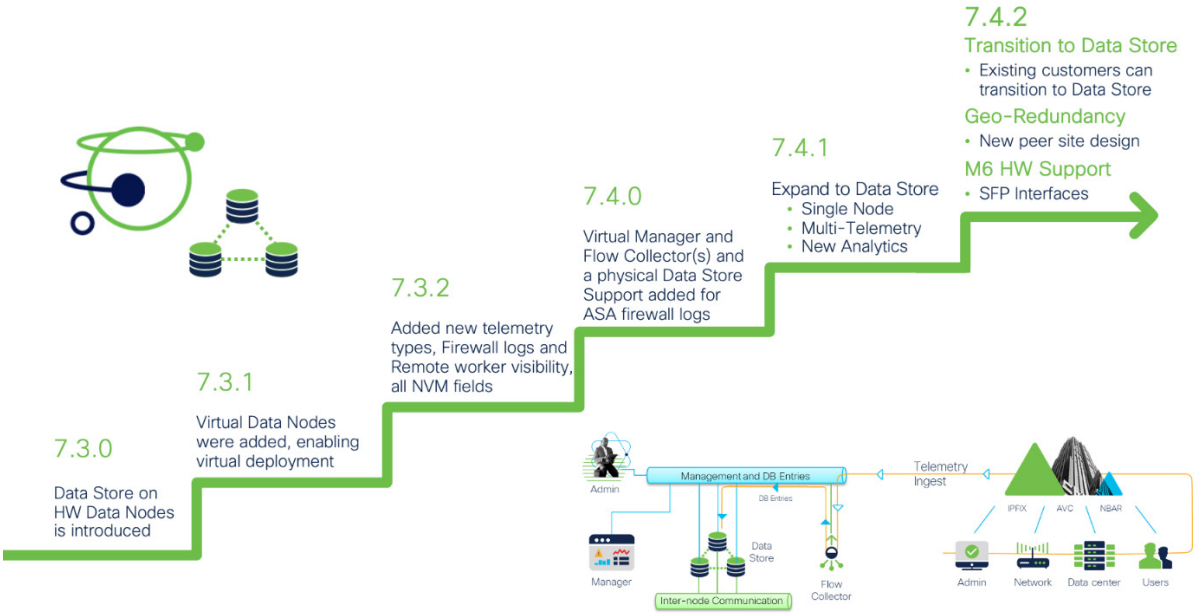
Executive Summary

The Cisco Secure Network Analytics Data Store solution provides a central repository to store network telemetry collected by Flow Collectors. The Data Store is comprised of a cluster of data nodes, each containing a portion of the network's flow data, and a backup copy of a separate data node's data. Because all the flow data is in one centralized database, as opposed to spread across multiple Flow Collectors, the Secure Network Analytics Manager can retrieve query results from a single Data Store faster than querying all Flow Collectors individually. The Data Store database cluster gives customers improved fault tolerance, improved query response, faster graph and chart population when using Cisco Secure Network Analytics. The Data Store architecture allows for independent scaling of flow collection and flow storage functions to better fit customer needs and budgets.

Technology Overview

Cisco Secure Network Analytics provides enterprise-wide network visibility and applies advanced security analytics to detect and respond to threats in real time. Using a combination of behavioral modeling, machine learning, and global threat intelligence, Secure Network Analytics can quickly, and with high confidence, detect threats such as command-and-control (C&C) attacks, ransomware, distributed-denial-of-service (DDoS) attacks, illicit crypto mining, unknown malware, and insider threats. With a single, agentless solution, Cisco Secure Network Analytics delivers comprehensive threat monitoring across the entire network, for both clear and encrypted communications.

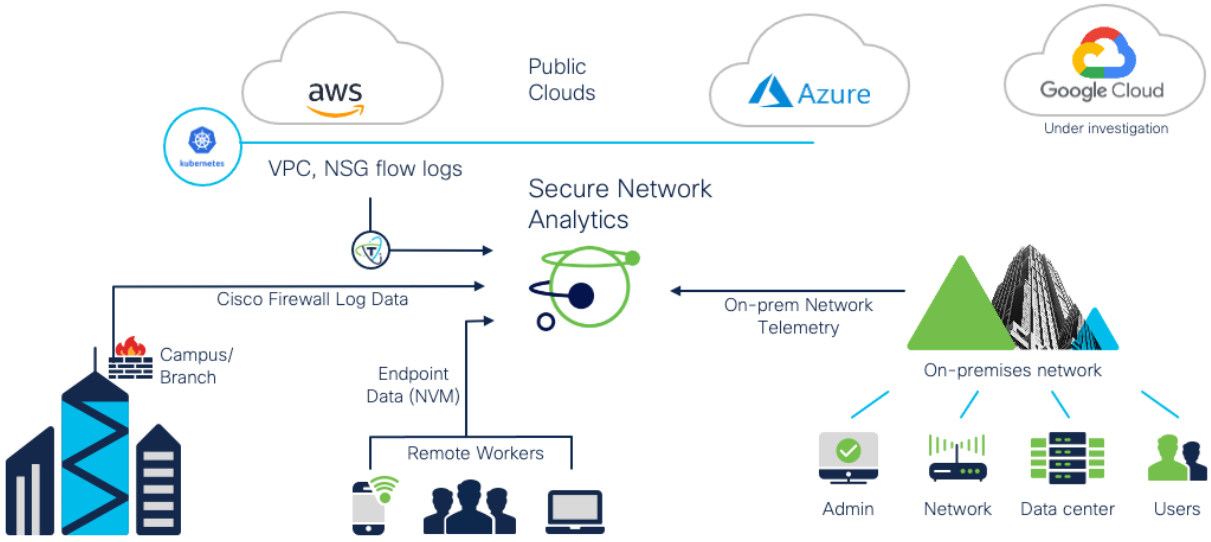
The Secure Network Analytics Data Store was first introduced in release 7.3.0 as a hardware appliance offering. In the following 7.3.1 release the Secure Network Analytics Data Store was expended to also include virtual appliances on VMWare's ESXi hosts and KVM hosts. In the 7.3.2 release, new telemetry types (Firewall logs from FTD and NVM flow) were able to be ingested. In 7.4.0 release, the Secure Network Analytics Data Store was expanded to allow a virtual Manager and Flow Collector to support a physical data store. In release 7.4.1, the Secure Network Analytics Data Store architecture was further expanded to allow the deployment of single node data store (help lower the entry bar) and allowing the intermix of physical and virtual appliances to build a Secure Network Analytics Data Store deployment. Release 7.4.1 also introduced multi-telemetry support on the data store Flow Collector and Converged Analytics. Now with release 7.4.2, it is easy to transition to a data store architecture and take advantage of more retention, faster query responses, better scalability, and data redundancy (with the new feature called peer site). Also, with this release performance of the manager, flow collector, flows sensor and data store has been improved with the support of the new UCSC M6 (hardware) appliances.



A Secure Network Analytics Data Store deployment consists of a Manager, a Data Store with 1, 3 or more data nodes and minimally 1 Flow Collector. The Manager, Data Store and Flow Collector appliances can be either physical or virtual appliances and they can be mixed. The Data Store's data nodes must be the same, either the same physical hardware or virtual with the exact same resources.

Multi-telemetry Ingest and Visibility

Secure Network Analytics can ingest (via a single Flow Collector) telemetry from multiple sources such as On-premises networks, remote workers (Cisco Secure Client/AnyConnect Network Visibility Module flow), Cisco firewall log data, and even public clouds (AWS and Azure) using Cisco Telemetry Broker.



The VPC and NSG flow logs from AWS and Azure provide insight into the activities of hosts residing within cloud environments. Meta data from Flow Logs centers around the network activity, like IPFIX/NetFlow and the Cisco Telemetry Broker pulls the flow logs from AWS S3 buckets or Azure Blob Storage via secure HTTPS connections and transforms the telemetry to IPFIX. The IPFIX is then forwarded to the Flow Collector for analysis.

Cisco FTD and ASA firewalls send firewall log data (100K eps for +30 days) to the same Flow Collector. The FTD/ASA Syslogs are viewed in the Manager using a unified event viewer. The viewer allows you to customize timeframes (so you can go back and forth in time), filter exclusively on security events and use per column filters to quickly isolate data of interest.

The Cisco Secure Client (formerly AnyConnect Secure Mobility Client) stores all network traffic telemetry records, even when users are not using a VPN. When the user connects to a VPN all stored flow data is sent to the same Flow Collector for processing. Detections are carried out on the On-network NVM flows (behavioral, custom security events and converged analytics). For Off-network flows (not connected via VPN) updates will by-pass the Flow Collector flow caches allowing for historical Network Visibility Module (NVM) flow data to be stored and viewed properly in report builder. No detections are currently applied to Off-network traffic. A Data Store deployment is required to support off-network telemetry and to retain all NVM telemetry records.

Analytics

Cisco Secure Network Analytics uses a combination of behavioral modeling and multilayered machine learning for its analytical engines as follows:

- Behavioral modeling – It applies close to 100 different security events or heuristics that look at various types of traffic behavior, such as scanning, beaconing host, brute force login, suspect data hoarding, suspect data loss, etc. These security events feed into high-level logical alarm categories.
- Multilayered machine learning – Applies machine learning, both supervised and unsupervised, to discover advanced threats and malicious communications. Integrates with a cloud-based multistage machine learning analytics pipeline which correlates threat behaviors seen in the enterprise with those seen globally.

Behavioral detections mentioned come from a deployed Secure Network Analytics Flow Collector (which collects the flow data processes it and analyzes it), Multilayered machine learning comes from Secure Cloud Analytics monitoring public clouds and hybrid networks.

Converged Analytics is where both Behavioral (on premises) and Multilayer machine learning (cloud) detections from both are converging into one detection engine, which will be our primary detection engine going forward.

Cisco Secure Network Analytics offers early access to advanced event capabilities and UI workflows with Converged Analytics, which provides new and effective alerts that require less manual configuration. Converged Analytics assigns appropriate roles to devices and uses this information, along with data collected using additional detection capabilities, to provide optimized alerts.

When you enable Converged Analytics, related features are switched on within your deployment. These additional capabilities function in parallel with your existing detections and interface. Continue to monitor your alarms, security events, and Manager as you do today, while also taking advantage of our new experimental detections and interface capabilities.

When you open a Converged Analytics alert in the Manager, you can view the supporting observations that led the system to generate the alert. From these observations, you can also view additional context about the entities involved, including the traffic that they transmitted and, if available, external threat intelligence.

Converged Analytics consumes additional system resources. Review your resource consumption prior to enabling this optional feature. Analytics feature sets and their general availability are limited to specific deployment types and may change with each future release.

For clarity, the Behavioral detections that are part of SNA will still be generated by the non-data store flow collector(s), irrespective if Converged Analytics is enabled.

Note the following are only applicable when enabling Converged Analytics:

- Converged Analytics supports a single Cisco Secure Network Analytics Data Store domain. Deployments may also contain 1 or more non-Data Store domains. For example, if you deploy one Data Store and use

separate domains for the Flow Collectors then Converged Analytics cannot be used for this design.

- You cannot enable Converged Analytics until you have deployed a minimum of 1 node and added a minimum of 1 Data Store Flow Collector Flow (NetFlow) on your network.
- If your system is consuming sFlow, the accuracy of your detections and the performance of your system may be impacted.
- Only the Admin user can enable and disable Converged Analytics.
- Converged Analytics in release 7.4.2 supports fail over deployments. Thus, if Managers are configured in high availability mode, then Converged Analytics can be enabled on both the primary and secondary Managers. Note the following:
    - Alerts and observations are not synchronized to the secondary manager.
    - Jobs do not run on the secondary manager when the primary is active.
    - When the secondary is promoted, then jobs start to run, but alerts and observations will be different.
    - When the primary is promoted back, alerts and observations will not match to the secondary.
    - Configuration of alerts is not synchronized across Manager.

Performance Enhancements

Cisco Secure Data Store provides world class Network Detection and Response (NDR) performance from a small on-premises footprint, which uses distributed engines combined with the centralized cloud assisted analytics. This is illustrated in the table below, which shows the performance enhancements when running reports with Converged Analytics disabled.

For example, we can see that a single Node Data Store best performance at 500K FPS is around 7 minutes to return results on average, and up to 30 minutes worst case. But if we grow the Data Store into a three-node cluster then performance at 700K FPS is remarkably better with typical reports running in 2 minutes and worst case 20 minutes. Also, if we compare the single Node Data Store at 1M FPS with the three Node Data Store at 2M FPS, then we see a vast query improvement of 3 nodes over a single node. In both instances we are seeing typical reports running in 1/3 of the time.

On-premises with Converged Analytics Disabled	Flows Per Second	Unique Host Count	Report Times (Avg/Max)
Previous Distributed Arch FC5210	300K	33M	<60 / hours
Single Node Data Store – DN6300 (M6)	500K	33M	<7min / <30min
Single Node Data Store – DN6300 (M6)	1M	33M	<15min / <65min
Three Node Data Store – DN6300 (M6)	700K	33M	<2min / <20min
Three Node Data Store – DN6300 (M6)	2M	33M	<5min / <45min
Three Node Data Store – DN6300 (M6)	3M	33M	<10min / <100min

When Converged Analytics is enabled (which provides the added benefit of converged cloud and on-premises analytics and detection) then the performance of the Data Store must be within the supported unique host count and flows per second range for your deployment based on guidance below:

On-premises with Converged Analytics Enabled	Flows Per	Unique Host Count

	Second	
Single Node Data Store (with single FC) – DS6200 (M5)	250K	1.3M
Single Node Data Store (with single FC) – DN6300 (M6)	600K	1.3M
Three Node Data Store (with four FCs) – DS6200 (M5)	600K	1.3M
Three Node Data Store (with four FCs) – DN6300 (M6)	600K	1.3M
Three Node Data Store (with four FCs)– DN6300 (M6)	850K	0.7M

Manager

The Secure Network Analytics Manager (SMC) aggregates, organizes, and presents analysis from Flow Collectors, the Identity Services Engine, and other sources. It uses graphical representations of network traffic, identity information, customized summary reports, and integrated security and network intelligence to deliver comprehensive security analytics.

The Manager also serves as the Central Manager for all Secure Network Analytics appliances. The Manager is the centralized user interface in which all configuration changes, management and investigative workflows take place.

PID	Description	Form Factor
ST-SMC2300-K9	Secure Network Analytics Management Console 2300	1RU
L-ST-SMC-VE-K9	Secure Network Analytics Management Console VE	N/A

Data Store

The Secure Network Analytics Data Store provides a central repository to store a network's telemetry, collected by Flow Collectors. The Data Store is comprised of one, three or more data nodes, with each data node containing a portion of ingested data, and a backup of a separate data node's data. Instead of data being spread across multiple flow collectors and databases, all network telemetry is consolidated in the centralized database. This centralized architecture is optimized to provided fast query results, rapid graph and chart population and long-term data storage. Additionally, the Data Store improves fault tolerance, offering the highest level of data redundancy and enterprise class deployment resiliency.

This new database architecture is a major improvement to the existing distributed model. The following sections will explore the technical details of this evolutionary architecture to explain how it operates, and the various deployment designs so that you can determine the one that works best for your specific environment.

For customers desiring a small footprint, the Secure Network Analytics single node Data Store can be deployed as either virtual or physical appliances and supports up to four flow collectors. It can easily expand to a full 3 node cluster, from 1 node to 3 nodes and then N+1 for horizontal scaling, using identical data node appliances.

A virtual single data store node scales to 225,000 Flows Per Second and a physical single data node scales up to 500,000 Flows Per Second. Note, a single data store deployment does not support data resiliency.

PID	Description	Form Factor
ST-DS6200-K9	Secure Network Analytics Data Store 6200	6RU
ST-DS6200-D1-K9	Secure Network Analytics Data Store Spare Node	2RU
ST-DN6300-K9	Secure Network Analytics Data Node	2RU
L-ST-DS-VE-K9	Secure Network Analytics Data Node VE	N/A

Flow Collector

The Flow Collector leverages enterprise telemetry such as NetFlow, IPFIX (Internet Protocol Flow Information Export), and other types of flow data from existing infrastructure such as routers, switches, firewalls, endpoints, and other network infrastructure devices. The Flow Collector can receive enhanced flow from ETA enabled exporters as well as collect telemetry from proxy data sources. This enhanced flow and proxy data can be analyzed by the cloud-based, multilayered machine learning engine to deliver zero-day threat detection.

PID	Description	Form Factor
ST-FC4300-K9	Secure Network Analytics Flow Collector 4300	1RU
L-ST-FC-VE-K9	Secure Network Analytics Flow Collector VE	N/A

Cisco Telemetry Broker

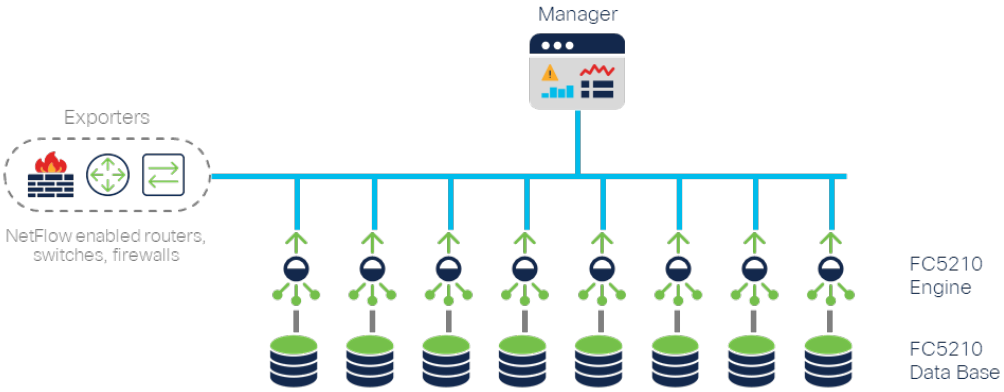
Cisco Telemetry Broker vastly simplifies the consumption of telemetry data from business-critical tools. It can broker hybrid cloud data, filter unneeded data, and transform data into a more usable format. We seek to democratize telemetry for all.

PID	Description	Form Factor
TB-SEC-SUB	Cisco Telemetry Broker Parent Subscription license	N/A
TB-ESS-100GB	Cisco Telemetry Broker 100GB/day license	N/A
ST-TB2300-K9	Cisco Telemetry Broker 2300 appliance	1 RU

Architecture Overview

Traditional Distributed Architecture

In a traditional Secure Network Analytics deployment, one or more Flow Collectors ingests network telemetry, deduplicates flow data, performs analysis, and reports directly to the Manager. To resolve user-submitted queries, including graphs and charts, the Manager queries all the managed Flow Collectors. Each flow collector returns matching results to the Manager. The Manager collates the information from the different result sets, then generates a graph or chart displaying the results. In this deployment, each Flow Collector stores flow data on a local database. See the following diagram for an example.



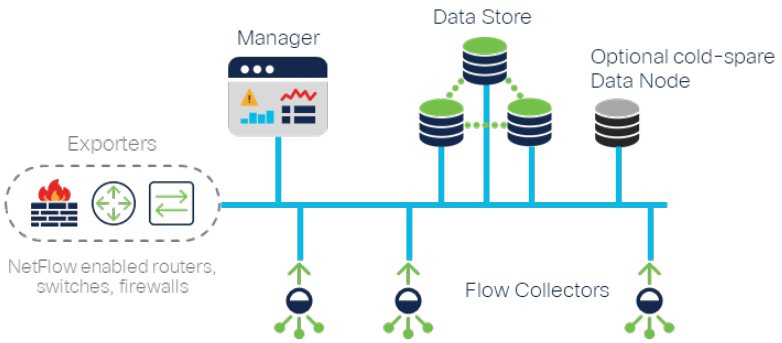
This Distributed Architecture poses design challenges when deployments scale across many Flow Collectors or when flow rates are multiple millions of flows per second (FPS).

Data Store Architecture

In a Secure Network Analytics deployment, the Data Store sits logically between your Manager and Flow Collectors. A Secure Network Analytics Data Store is comprised of one, three, or more data nodes. A physical Data Store deployment can support up to 12 DS6200s or 36 DN6300s. When deploying multiple data nodes to meet flow ingest or retention requirements, the resources of each data node appliance combine to from a single Data Store.

You can expand the Data Store up to a maximum of 36 data nodes. Secure Network Analytics supports deploying the Data Store with 1,3,4,5, N+1 data nodes, where each data node must be of the same type and have identical resources allocated. Deploying a Data Store with 3, 6, 9, N+3 provides optimal performance outcomes. Support for Data Stores deployed with DS6200 (M5) data nodes being expanded using DS6300 (M6) data nodes is planned for a future release.

In a Data Store architecture one or more Flow Collectors ingests, deduplicates, and performs analysis on network flow, firewall log, or client endpoint data. The processed telemetry is then sent directly to the Data Store for storage. The Data Store collates the information from the different data sets, storing it within the Data Store database, and distributing it equally to all data nodes within the Data Store.

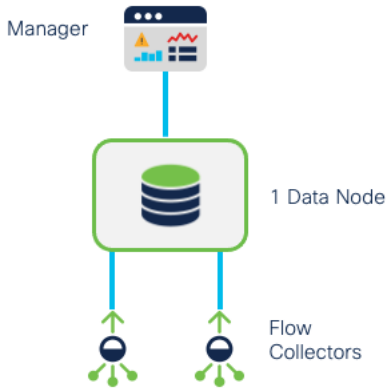


This centralized design allows the Manager to send all queries to a single destination instead of querying each Flow Collector individually. In addition to overall greater retention times, the Data Store architecture enables faster overall query performance, which scale linearly as more nodes are added to the Data Store. The Data Store supports an optional cold spare data node, which can be added to the Data Store's database in the event of a failure with one of the managed data nodes.

Single Node Data Store

The central Data Store architecture (starting in release 7.4.1) allows you to install a single virtual or physical data store node with a physical or virtual Manager and flow collector(s). The advantage of this architecture is that it reduces the number of appliances required to be deployed making the installation, easier and quicker. This single node Data Store supports all the beneficial outcomes, like Converged Analytics, multi telemetry ingestion (firewall logs from FTD/ASA, NVM data and NetFlow), as the multi-node Data Store.

The deployment can be just a single data node as shown below. The primary reasons to go beyond a single node would be for increased retention, better query responses or to provide data resiliency.



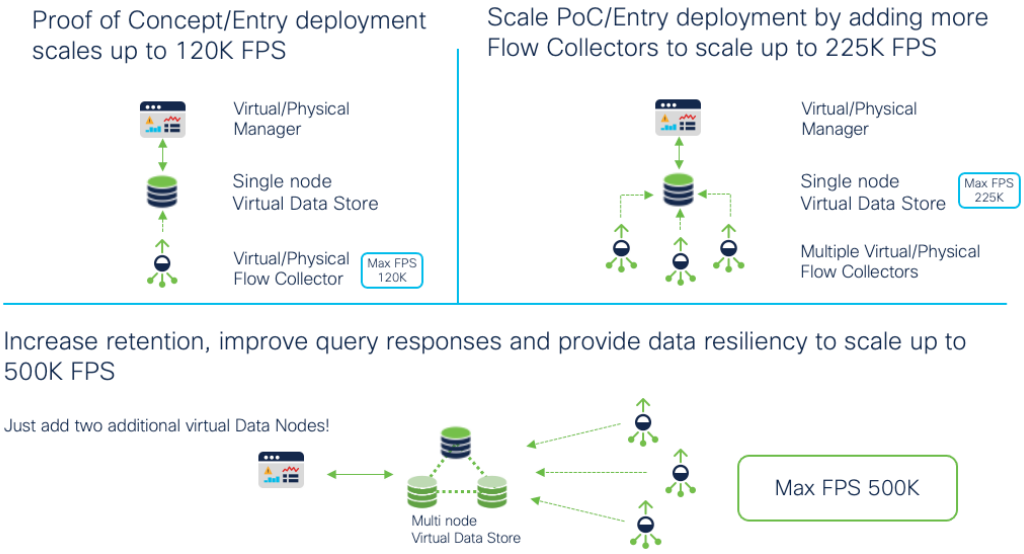
The virtual Data Store helps to lower the bar of entry by reducing costs and still provides great performance (225,000 FPS). It is not just for new customers, it can be used for existing deployments where the customer wants to take advantage of Converged Analytics, Multi Telemetry ingest, query time performance improvements, and scale improvements.

The virtual Data Store node can be used for proof of concepts, to build lab environments, new deployments and expand existing deployments. All that is required is to install a virtual Manager, virtual data node and a virtual Flow Collector (maximum of 4 Flow Collectors supported for a single virtual or physical Data Store) with the virtual machine specifications listed below, which will scale up to 225,000 FPS.

Virtual Resourcing recommended for 30K FPS and 50K internal hosts.

Virtual	vCPU	RAM	Storage
Manager	6	40 GB	200 GB
Flow Collector	6	32 GB	200 GB
data node	6	32 GB	2.5 TGB
Total	18	104 GB	2.9 TB

In the adoption examples below the virtual single node Data Store can start as an entry deployment of up to 120,000 FPS and then scale up to 225,000 FPS by adding more Flow Collectors. If more scale (up to 500,000 FPS), retention, data resiliency or faster query responses are required, then two additional data nodes can be added to create a multi-node Data Store.



Of course, a single physical Data Store node can be used for even greater performance scaling up to 600,000 FPS and supporting 1.3M hosts when using the DN6300 data node.

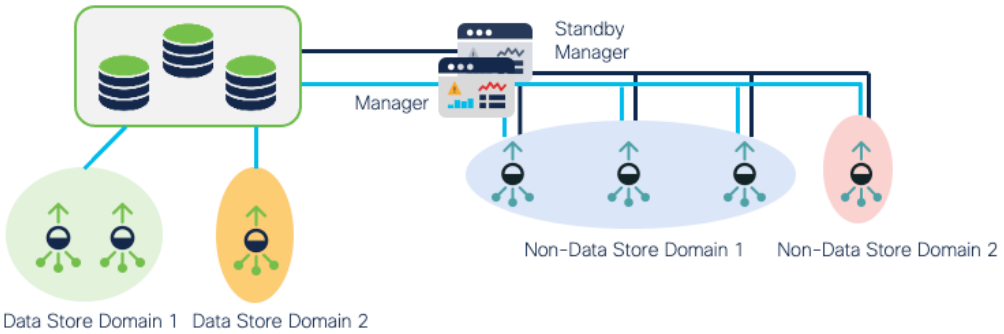
Secure Network Domains

A network domain is a grouping of hosts and other devices that you want to monitor and manage. Flow Collectors exist within domains, and you can have multiple domains within one Secure Network Analytics system. Domains are completely independent of other domains, and every domain contains the Host Group tree. Separate Network Domains can be used to separate the ingest of telemetry for customers. So, customer A ingested telemetry is processed by Flow Collector A and customer B telemetry by Flow Collector B (also helps when customer A and B use overlapping IP addresses as the domains keep the telemetry separate).

Data store and non-data store domains can be created (called hybrid configuration) in a deployment.

- Data Store Domain: The Flow Collector sends its telemetry to the Data Store data nodes for storage.
- Non-Data Store Domain: The Flow Collector stores its telemetry locally on the Flow Collector or on the Flow Collector database (5000 Series only).
- Hybrid Configuration: In Secure Network Analytics with a hybrid configuration, you can configure a Data Store domain and Non-Data Store domain. When you configure your Flow Collectors, you can choose which domain they will use, which determines where they send data.

Note: From release 7.4.1 onwards you can only have one Data Store (can be 1/3 Nodes or more, which can be physical or virtual) per deployment. Also, you can only have one ISE integration per domain and within the domain you can have multiple ISE instances.



Note: Converged analytics does not support multiple data store domains

Data Resiliency

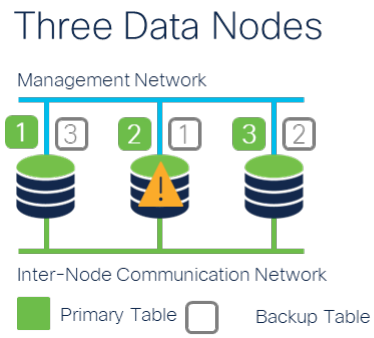
One of the improvements coming with the Secure Network Analytics Data Store and the new centralized architecture is the implementation of K-safety functionality.

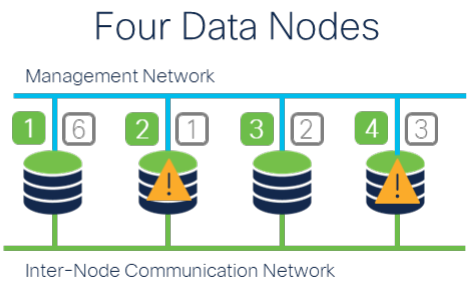
K-safety enables the fault tolerance in the Secure Network Analytics Data Store database cluster. The value K represents the number of times the data in the database cluster is replicated. These replicas allow other nodes to take over query processing for any failed nodes, which results in no data loss upon an unexpected node failure, allowing processing and ingestion to continue during a failure event.

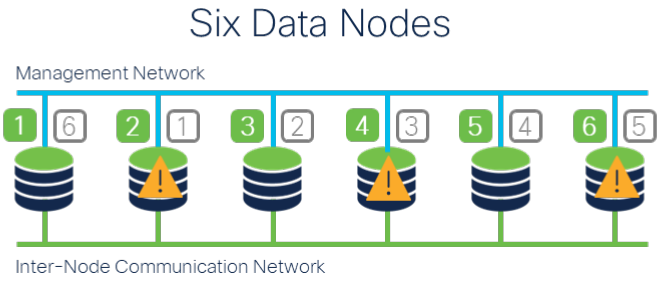
Each single node will write a copy of its own tables onto its associated buddy node to ensure resiliency across the database.

A Secure Network Analytics Data Store deployment with a single DS6200 cluster (i.e., three data nodes) or three individual data nodes DN6300/VE's will have a K-Safety of 1. This configuration can lose any 1 data node out of the 3 without interruption to query actions or the storage of data. This is because each data node's tables are written to a single peer to create a mirror copy data node's primary table. When a Secure Network Analytics Data Store is expanded by adding more nodes the K-Safety feature will allow for up to 50% of nodes failing while maintaining complete functionality. Note, while up to 50% of the nodes failing is a theoretical maximum, we support losing up to 40% of the nodes failing, where a node's table is not completely down.

See examples of each design and the associated data node redundancy below. The green square represents the primary database table, while the grey square represents the replicated copy in case of data node failure.







A Secure Network Analytics Data Store design can also support the use of a cold-spare data node. If this is required the Data Store will be setup as normal, and afterwards another data node can be manually added to central manager but will not be added to the Data Store's database until a data node has failed. Once added the Data Store will replicate the data from the existing data nodes restoring the databases to normal operation. This cold node support was an initial feature of the Data Store design before N+1 data node support was available. Now that Data Stores can have 3+ data nodes, we highly recommend adding a spare node as an active node to get increased performance and retention times in nominal all up scenarios.

Network Connectivity

Hardware Data Store

For hardware deployments, there are two interface configurations for handling appliance Management Communications and Flow Ingestion. The configuration will be made during the initial setup on the Manager, Flow Collector, and each data node.

Physical Network Connectivity (for SMC2210, FC4210 and DS6200):

- Secure Network Analytics Data Store appliances support either a Base-T (copper) or SFP (fiber) configuration for Management Communication and Flow Ingestion.
- The Manager and Flow Collector can be configured with a 10-Gbps Ethernet SFP+ (default) or 1-Gb/10-Gb BASE-T Ethernet LAN port (RJ45 connector).
- Each data node can be configured with a 10-Gbps Ethernet SFP+ (default) or 10-Gb BASE-T Ethernet LAN port (RJ45 connector).

Physical Network Connectivity (for SMC2300, FC4300 and DN6300):

- Secure Network Analytics Data Store appliances support either a SFP 1Gbps (Fiber or Base-T) or 10Gbps SFP+ (fiber/Base-CU – DAC direct attached cable) configuration for Management Communication and Flow Ingestion.
- The Manager and Flow Collector can be configured with a 1Gbps SFP (Fiber/Base-T) or 10Gbps SFP+ (Fibre/Base-CU - DAC direct attached cable) for management.
- Each data node can be configured with a 1Gbps (Fibre/Base-T) or 10Gbps SFP+ (Fiber/Base-CU - DAC direct attached cable) for management and 10Gbps SFP+ (Fiber/Base-CU - DAC direct attached cable) for inter-data node communications.
- For M6 appliances only SFP-based NICs are used for both management and production connections. There is a single 100Mbps/1Gbps Base-T (RJ45) CIMC connection on each appliance.

While each appliance can be configured independently, it is recommended to configure all appliances using the same media. It is also recommended to deploy the appliances using 10Gbps connections.

In deployment scenarios where Flow Collectors are not co-located with the Data Store, it will be necessary to ensure adequate bandwidth exists between them to support Data Store updates from the Flow Collector (note Flow Collector encrypts and compresses the data sent by 90% before sending it to the Data Store). Optimum design is to allow for a minimum of 4Gbps of bi-directional burst traffic between these appliances. If the bandwidth is lower than 4Gbps then the ingest performance of the data store will be affected (it will be slower) but the Data Store will still function.

Virtual Data Store

It is recommended that all network adapter types are set to VMXNET 3 to obtain optimal performance. If Secure Network Analytic virtual appliances are distributed across multiple ESXi/KVM hosts it is recommended that vNIC uplinks are 10Gbps, redundant, and NIC teaming is implemented.

data node devices will need two vNICs one for management communication and the other for internode communication.

Flow Ingestion

Getting flow out of a network and into a Flow Collector engine is an early step in establishing a successful Secure Network Analytics deployment. Typically, for most Secure Network Analytic deployments, NetFlow will be enabled at the access, distribution, and core to give administrators the entire conversation both north-south and east-west.

Each network is different and as such special consideration is needed when designing how to get NetFlow into the Flow Collector.

There are several types of telemetry exporters that are supported by Secure Network Analytics. Supported telemetry protocols include:

- NetFlow version 5 – fixed set of IPv4 network telemetry fields
- NetFlow version 9 –template-based format of customer IPv4/IPv6 network telemetry fields
- IPFIX – Optimal telemetry format for NetOps events and SecOps detections
- Sampled NetFlow – a sampling of traffic is sent in either NetFlow v5 or v9 format. Sampled NetFlow is adequate for network trending and capacity planning but is insufficient for security outcomes.
- sFlow – Supported on some 3rd party network devices, useful for NetOps, but not for SecOps
- Syslog – Standard message for sending and receiving connection, eventing, or notification messages. Syslog can provide addition context to network data such as adding as URL data.
- NSEL – NetFlow Security Event Logging is provided by the Cisco ASA and FTD, the protocol includes fields such as pre and post network address translations.

There are additional data sources supported by Secure Network Analytics:

- Remote Worker Network Visibility Module (NVM) flow data from Cisco Secure Client endpoints – This enables Secure Network Analytics to store and view the NVM fields. Also allows existing policy violation rules to trigger from NVM flows and carry out NetFlow detections based on the NVM traffic. Custom Security Events based on the endpoint connections can be created.
- Firewall logs – Ingests and store firewall event logs for Security Analytics and Logging (SAL On Prem). Provides increased storage at a larger retention period for Cisco ASA (via syslog to Manager) and FTD firewalls. Security events exported to the Manager.
- AWS VPC flow logs – Cisco Telemetry Broker enables you to configure VPC flow log inputs to consume AWS VPC flow logs from an S3 bucket., transform them into IPFIX, and send the IPFIX to the SNA flow collector (or other destinations). This gives you visibility and detection capabilities for cloud workloads using the network telemetry from the cloud.
- Azure NSG2 flow logs – Cisco Telemetry Broker enables you to configure NSG Flow Log inputs to consume Azure NSG Flow Logs from an Azure Storage Account, transform them into IPFIX, and send the IPFIX to the SNA flow collector (or other destinations). This gives you visibility and detection capabilities for cloud workloads using the network telemetry from the cloud.
- Proxy Logs – Gathers user information from your network proxy servers. Provides URLs and application names of the traffic inside a network going through the proxy server (Blue Coat, McAfee, Cisco WAS and Squid supported)

Typically, NetFlow is collected from the WAN edge, campus core and access layers from a wide range of exporters: switches, routers, firewalls, wireless LAN controllers and flow sensors.

Most exporters only support two destinations for exporting telemetry and if there are multiple tools with collectors needing this data there is a potential for a constraint on where telemetry can be sent. This can be solved by deploying Cisco Telemetry Broker which can ingest numerous forms of telemetry and applies rule to send to multiple destinations at line rate.

Traffic Profiles

Telemetry profiles should be taken into consideration when setting the retention period for the Data Store. Factors to consider include the flow per second rate, types of traffic profiles, and requirements for how long it is necessary to store the data.

Standard Enterprise traffic profile:

This is the typical environment that closely resembles a campus architecture with standard NetFlow enabled exporters. In these environments the Flow Collectors typically consolidate flows from five or more exporters as the network communication traverses the campus network. This traffic profile is most common and suitable for most designs.

Physical Appliance Sizing Table – Individual M5 Nodes (data node Capacity 15 TB) (3 Nodes = DS6200)

Flow Rate (FPS)	# FC 4210	# data nodes for 90 days storage	# data nodes for 180 days storage	# data nodes for 360 days storage
250K	1	3	6	12
500K	1	6	12	21
1M	2	12	21	
2M	4	21		
3M	6	33		

Physical Appliance Sizing Table – Individual M6 Nodes (data node Capacity 23 TB) (1 Node = DN6300)

Flow Rate (FPS)	# FC 4300	# data nodes for 90 days storage	# data nodes for 180 days storage	# data nodes for 360 days storage
250K	1	3	4	7
500K	1	4	7	13
1M	2	7	14	26
2M	3	14	27	
3M	5	21		

Service Provider traffic profile:

Service Providers have unique environments consisting of high numbers of unique hosts, very little deduplication of flows and large amounts of sampled flow data. Due to this profile more appliances are required for flow collection and data storage to meet the flow rate and retention times. The typical guidance would be to plan for twice the number of flow collectors as recommended for an Enterprise traffic profile. Due to optimizations and storage characteristics of large data sets, we have not seen a need for increased data nodes for similar data retention times.

Sizing Virtual Data Store Node Deployments

While physical appliances have fixed resources and are easily scaled and sized using the previous tables, virtual Data Store deployments may require additional calculations to support specific retention times.

As of version 7.4.2 expanding the storage size of virtual data nodes is not supported. However, users may add new data node of equivalent resourcing (CPU, Memory, Storage) to grow the Data Store if additional retention is required.

Data Store Retention Dependencies:

Flow data retention, the number of days telemetry can be stored, within the data store is primarily dependent on the following:

- Telemetry ingest rate (ex: NetFlow Flows Per Second, Firewall Events Per Second, NVM Flows per Second)
- Number of days to retain the flow data (i.e, meta data from network, firewall, or NVM endpoint)
- Number of days to retain the flow interface stats data (i.e. exporter interface data used to track a conversation through a network)

Flow data retention can be lower or higher if one of the above dependency's changes. For example, if the flow interface data retention is adjusted from 7 (default) to 30 days and the same traffic arrival rate then the flow data retention will be lower as the flow interface data will consume more of the data store storage, leaving less room for flow/firewall/NVM data.

Let's have a look at a typical enterprise scenario, where physical appliances will be deployed:

- DN6300 data node Capacity (M6) = 23 TB
- Number of days to retain the flow data = 30 days
- Traffic Arrival Rates
    - NetFlow Flows Per Second = 1M FPS
    - Firewall Events Per Second = 5000 EPS
    - NVM Flows Per Second = 1M FPS
- Number of days to retain the flow interface stats data (example 7 days is default) = 7 days

To support these requirements, we would need approximately 94.63 TB of storage, which would require a data store with 5 data node appliances. The NetFlow flow data would consume 48 TB, Firewall data would consume 2.43 TB, NVM flows would consume 33 TB and flow interface statistics would consume 11.2 TB.

If we were to use the DS6200 data node (capacity 15TB) then we would need 7 data nodes, due to their lower storage capacity.

In some environments there is a need to store flow interface data longer, let's see the impact to storage requirements when making this change:

- DN6300 data node Capacity (M6) = 23 TB
- Number of days to retain the flow data = 30 days
- Traffic Arrival Rates
    - NetFlow Flows Per Second = 1M FPS
    - Firewall Events Per Second = 5000 EPS
    - NVM Flows Per Second = 1M FPS
- Number of days to retain the flow interface stats data (example 7 days is default) = 30 days

By increasing the flow interface data to a month, we would need approximately 131.43 TB of storage, which would be 6 data node appliances. In this case we would expect the flow data to consume 48 TB, as calculated above, however the flow interface stats now consume 48 TB, which as you may have expected is nearly double the storage than the first scenario.

To help users estimate the size of their Data Store deployment, we have provided a simple calculator based upon actual data from both extensive internal testing as well as real world customer experience. Please click on the icon to open the calculator. Once open you will see the fields you can modify as outlined cells.

The data node Capacity column is preset with the values for the physical appliances; however, if you are deploying virtual machines, you can modify this number to reflect the virtual storage you plan to allocate per data nodes, and the calculator will work out the number of virtual data nodes required based on the other parameters.

Click below to open the embedded Data-Store-retention-calculator:

Data Store Sizing Calculation:

If you are interested in how we built the calculator we based, it on the following:

Formula for Calculating NetFlow Flows Stats and Flow Interface Stats

[(Avg FPS / 1000) x 1.6GB x Retention days] / DN# = Data Storage required per data node

1. Divide the FPS by 1000
2. Multiply this number by 1.6 GB. Note: A day's worth of storage at an ingress rate of 1,000 FPS is 1.6GB
3. Multiply this number by the required retention time in number of days.
4. Divide this number by 3, which is the minimum number of virtual data nodes required for resiliency (supports minimum of 1 or 3 virtual data nodes and additional data nodes can be added N+1)
5. As we showed in the above scenarios, this formula must be run for both flow data and flow interface stats then combined to obtain the storage size and data node quantity.

Formula for Calculating Firewall Events Per Second *(100 EPS = 1.62GB per day)*

(Avg Number of EPS x 1.62GB) x Retention days = Total GB required

Formula for Calculating NVM Flow Stats *(1000 FPS NVM = 1.1GB per day)*

[(Avg NVM FPS/ 1000) x 1.1GB x Retention days] = Total GB required

Let's walk through the details in another example for clarity, but keep in mind the calculator does all this for you, so

please take advantage of it. Let's take an example, where the environment being monitored has a mix of telemetry from the network, firewalls, and endpoints (remote workers) all being ingested by a flow collector and Data Store. Expanding the data collection from just the network to firewalls and endpoints (remote workers) further extends visibility and enhances context. With firewall telemetry extends the visibility to the network perimeter via the firewall log collection and NVM telemetry extends visibility to remote workers (users).

For this example, the environment produces 50,000 daily average Flows Per Second (FPS) from the network, 5000 Events Per Second (EPS) from the firewalls and 50,000 daily average NVM Flows Per Second (FPS) from the endpoints. The customer needs storage for 90 days and they want to deploy a data store with 3 virtual data nodes. We will calculate the flow stats and flow interface stats (for the network flows) then firewall events per second and finally the NVM flow stats, then we will combine all four results:

Flow stats: [(50,000/1000) x 1.6 x **90**]

1. Daily average FPS = 50,000
2. 50,000 average FPS / 1,000 = 50
3. 50 x 1.6 GB = 80 GB for one day's worth of flow storage
4. 80 GB x 90 days = 7200 GB (7.2 TB) of total storage for 90 days

Flow interface stats: [(50,000/1000) x 1.6 x **7**]

1. Daily average FPS = 50,000
2. 50,000 average FPS / 1,000 = 50
3. 50 x 1.6 GB = 80 GB for one day's worth of flow storage
4. 80 GB x 7 days = 560 GB (0.56 TB) of total storage for 7 days

Firewall Events per second (in TB): [5000 x 1.62 x **90**]

1. Daily average EPS = 5,000
2. 5,000 average EPS / 100 = 50
3. 50 average EPS x 1.62 GB = 81 GB for one day's worth of flow storage
4. 81 GB x 90 days = 7290 GB (7.29 TB) of total storage for 90 days

NVM Flow stats (in TB): [(50,000/1000) x 1.1 x **90**]

5. Daily average FPS = 50,000
6. 50,000 average FPS / 1,000 = 50
7. 50 x 1.11 GB = 55GB for one day's worth of flow storage
8. 55 GB x 90 days = 4950GB (4.95 TB) of total storage for 90 days

7.2 TB of flow stats data + 0.56 TB flow interface stats data + 7.29 TB Firewall stats + 4.95TB of NVM flow stats = 20 TB of storage space required.

20 TB of total storage / 3 data nodes = ~6.6 TB per data node.

Requirements and Considerations

Licensing Requirements

Your Secure Network Analytics deployment requires a Flow Rate (FPS) Smart License; the Data Store itself does not require an additional license.

Basic Hardware Specifications

Appliance	Platform	Gen	Height	SW Version
SMC2300	UCSC-C225	M6SX	1 RU	7.4.2 and above
DN6300	UCSC-C245	M6SX	2 RU	7.4.2 and above
FC4300	UCSC-C225	M6SX	1 RU	7.4.2 and above
SMC2210	UCSC-C220	M5SX	1 RU	7.3.0 and above
DS6200	UCSC-C240 (3x)	M5SX	6 RU (3x 2 RU)	7.3.0 and above
FC4210	UCSC-C220	M5SX	1 RU	7.3.0 and above

When configuring a Secure Network Analytics Data Store deployment all Secure Network Analytics appliances must be on the same version.

Basic Virtual Appliance Specifications

Manager VE

Users	Memory Reserved	CPU Reserved	Storage	SW Version
Up to 9	40 GB	6	200 GB	7.3.1 and above
10 or more	64 GB	12	480 GB	7.3.1 and above

Flow Collector VE With Data Store

Flows Per Second	Memory Reserved	CPU Reserved	Storage	SW Version
Up to 10,000	24 GB	2	200 GB	7.3.1 and above
Up to 30,000	32 GB	6	200 GB	7.3.1 and above
Up to 60,000	64 GB	8	200 GB	7.3.1 and above
Up to 120,000	128 GB	12	200 GB	7.3.1 and above

Data Store VE 1 x data nodes

Flows Per Second	Memory Reserved	CPU Reserved	Storage	SW Version
Up to 30,000	32 GB	6	2.25 TB	7.4.1 and above
Up to 60,000	32 GB	6	4.5 TB	7.4.1 and above
Up to 120,000	32 GB	12	9 TB	7.4.1 and above
Up to 225,000	64 GB	18	18 TB	7.4.1 and above

Data Store VE 3 x data nodes

Flows Per Second	Memory Reserved	CPU Reserved	Storage	SW Version
Up to 30,000	32 GB per data node	6 per data node	1.5TB per data node 4.5 TB total Data Store	7.3.1 and above
Up to 60,000	32 GB per data node	6 per data node	3 TB per data node 9 TB Data Store	7.3.1 and above
Up to 120,000	32 GB per data node	12 per data node	6 TB per data node 18 TB Data Store	7.3.1 and above
Up to 250,000	64 GB per data node	18 per data node	10 TB per data node 30 TB Data Store	7.3.1 and above
Up to 500,000	64 GB per data node	18 per data node	15 TB per data node 45 TB Data Store	7.3.1 and above

Hardware Networking Requirements

Take consideration of the following requirements for management and inter-node communications when designing Data Store deployment using hardware appliances:

Management communication requirements:

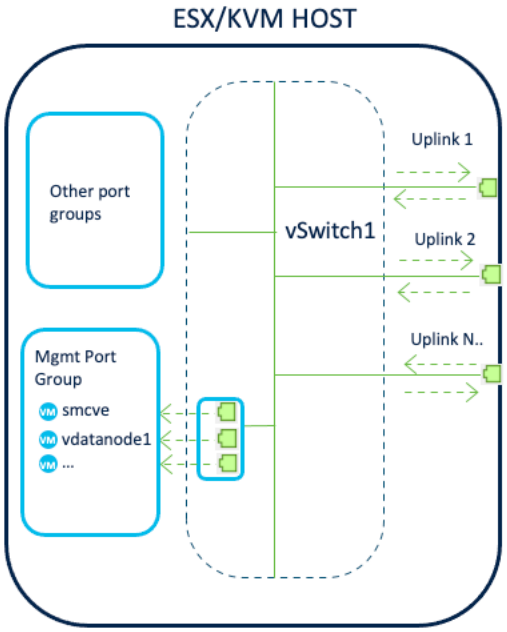
1. 10 Gbps throughput is required for the Data Store's management interface, which can be BASE-T, SFP (Fibre/Base-T M6) or SFP+ (Fibre/Base-CU - DAC direct attached cable M6)
2. Clock skew must be kept to 1 second or lower between and among your data nodes.
3. When using 10G throughput it is recommended that there is 4 Gbps of bandwidth available between each Flow Collector and the Data Store's data nodes. The optimum design is to allow for a minimum of 4Gbps of bi-directional burst traffic between these appliances. If the bandwidth is lower than 4Gbps then the ingest performance of the Data Store will be affected (it will be slower) but the Data Store will still function.

Inter-node communication requirements:

1. 10 Gbps full duplex switched layer-2 connection exists between the data node and supports 6.4Gbps of throughput or greater.
2. Round-trip time (RTT) latency between data nodes must not exceed 200 microseconds.
3. Clock skew must not exceed 1 second between data nodes.

Virtual Networking Requirements

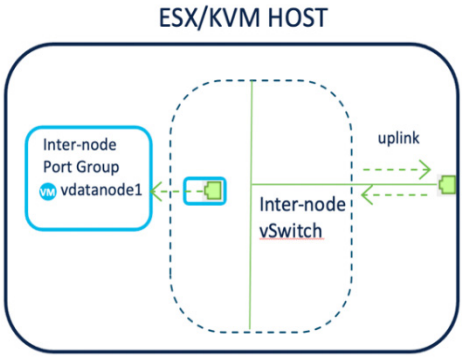
Take consideration of the following requirements for management and inter-node communications when designing Data Store deployment using virtual appliances:



Management communication requirements:

In a design where Secure Network Analytic devices are deployed across ESX/KVM hosts it is required that the physical adapter uplink that is used for management communications is 10G.

In this example there is a defined port group named Management Port Group. The vNIC eth0 interfaces for each Secure Network Analytic Device are assigned to the Management Port Group. It is required that the interfaces in the management port group and its associated vSwitch has at least one 10G uplink.
It is a best practice to ensure that more than one 10G uplink is available for redundancy and NIC teaming is implemented.



Inter-node communication requirements :

For a virtual deployments with data nodes spread across multiple ESX/KVM hosts it is required that the physical adapter uplink for each inter-node eth1 vNIC is 10Gbps.
It is a best practice to ensure that more than one 10G uplink is available for redundancy and NIC teaming is implemented.
Ensure the inter-node eth1 vNIC is assigned a distributed or standard port group tagged with a VLAN that has been configured on the networking device uplink. This ensures each internode interface on each data node can communicate with the corresponding inter-node interfaces on the other data nodes within their private network.

Caveats as of version 7.4.2

Take into consideration the following limitations and features that are not supported for the 7.4.2 release of the Secure Network Analytics Data Store:

- Java Client Support / Desktop Client.
- Host Classifier Application is not supported.
- The data forwarder or data exporter on the flow collectors.
- Flow Adapter service.
- Cannot mix hardware and virtual Data Store nodes, must be homogenous.

Design and Deployment Considerations

Below is a list of things to know when designing and deploying a Data Store:

- A Secure Network Analytics Manager supports one Data Store. Multiple domains are supported, and a Manager can support multiple non-Data Store and Data Store domains.
- A Data Store cannot typically be stretched across two data centers due to inter data node latency requirements and that they need to be on the same layer 3 network.
- Data nodes must be all virtual or all physical, cannot mix types in a single Data Store
- If expanding from a single node to a multiple node Data Store, a minimum of 3 data nodes is required. The Data Store does not support a 2 data node configuration.
- Separate Flow Collectors are required for non-Data Store and Data Store.
- Thus, if you have an existing deployment and want to add Data Store you will need a second Flow Collector.
- Note, there is a way to copy the existing domain configs to the new Data Store domain, but manual sync will be required for now.
- You can transition your existing Flow Collectors to use the Data Store database without loss of pre-transition data or visibility. Once you have completed the initial transition process, you can preserve your pre-existing data until you no longer need it. By completing the transition process, your Flow Collector will solely become a Data Store Flow Collector. All the pre-existing Non-Data Store data that the Flow Collector is storing will be deleted and resources will be recovered, thereby improving the performance of your Flow Collector.
- If adding a Flow Collector 5K to a Data Store, the FC5200/FC5210 Engine and Database appliances are both required for the initial installation.

Feature and Functional Considerations:

- As of SNA 7.5.0, detections are carried out on Firewall logs flows (Behavioural and Custom Security Events). With the Data Store.
- Firewall logs can be seen using the Secure Analytics and Logging application.
- Firewall logging supports log ingest from up to 5 Flow Collectors.
- Neither network detections nor Customer Security Events CSEs are applied to firewall logs.
- Remote Worker enhancements with release 7.4.2
    - Detections are now available for remote workers using Full VPN:
        - All 98 behavioural analytics events on the Flow Collector run against on-network remote worker telemetry.
        - Custom Security Events can be applied to on-network remote worker telemetry.
        - Converged Analytics network detections 50+ are applied to on-network Remote Worker telemetry.
    - Flow search does not show NVM specific fields.
    - Use the Endpoint Traffic (NVM) report in the Report Builder application to view Endpoint data.
    - Detections are carried out on NVM flows (Behavioural, Custom Security Events and Converged Analytics.
    - No detections are applied to off-network traffic telemetry, but this historical NVM flow data can be viewed in the NVM endpoint traffic reports in Report Builder.
- DTLS is not supported for Syslog or endpoint telemetry (NVM) ingest.
- For ISE integrations, you can configure one integration per domain but within that domain you can have multiple ISE instances.

Designs and Recommendations

Hardware or Virtual

The first decision to make in any Data Store design is whether to deploy the solution physically or virtually. There are considerations to be made based on the unique requirement of a given environment. For example, some environments are focus on reducing hardware footprints and would prefer a virtual solution, while others may require extremely high flow ingest rates or long-term storage which tend to require a hardware deployment.

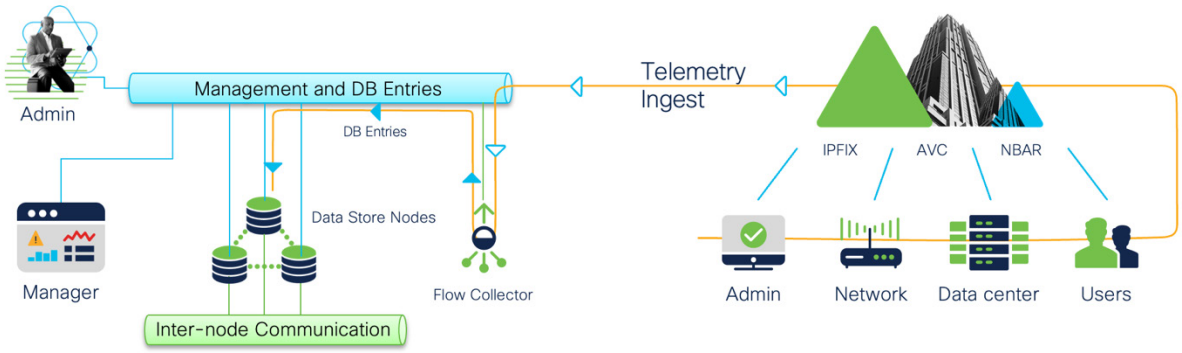
The largest factor for many customers will be the scale of the deployment.

If the design calls for ingest rates above 500K FPS, then the solution should be designed for hardware. For networks that produce less than 500K FPS it is likely the virtual solution will make the most sense for users.

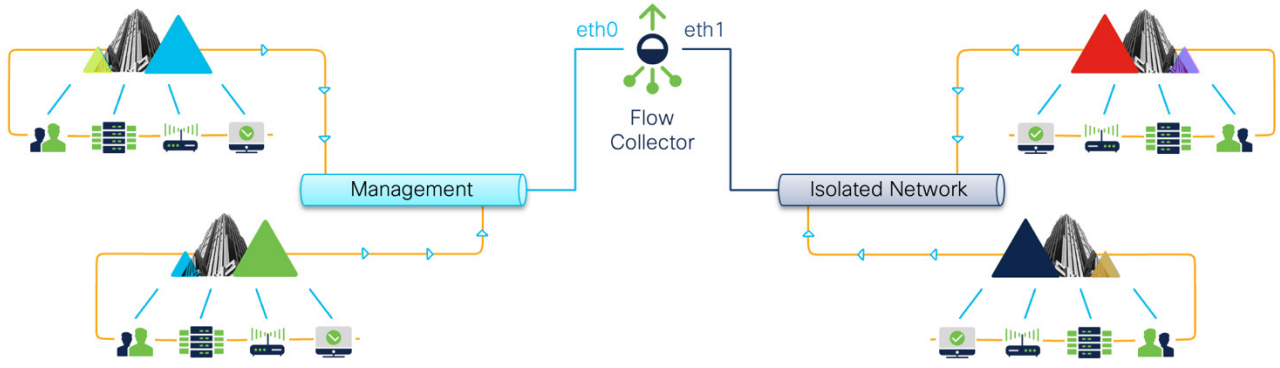
Placing the Secure Network Analytics Data Store

As a repository for flow data collected by the Flow Collector, and as the centralized repository against which a Secure Network Analytics Manager runs queries, install the data nodes at a location in the network that is accessible by the Flow Collector and the Secure Network Analytics Manager.

It is common in Enterprise networks to deploy Secure Network Analytics on a management network. In this design exporters will need network connectivity to the management interface on the Flow Collector. Telemetry from these network exporters will then be de-duplicated and processed by the Flow Collector's engine and stored as database entries in the Data Store.



There is an additional design consideration to allow the Flow Collector to collect telemetry from standalone or isolated networks by using a second collection interface on the Flow Collector. This can be beneficial in a design where a test lab or a DMZ is segmented away from the production network.



This design makes it possible to physically segment all management communications from other isolated areas of the network.

Standard Design - Hardware

Let's start with a simple design where a standard three node Data Store is deployed with one Manager, Flow Collector, and single Nexus switch. After which, we'll consider other configurations and design choices which provide increased redundancy and scale.

This design will include the following Secure Network Analytics appliances:

- SMC2210/SMC2300
- FC4210/FC4300
- DS6200 (3 data nodes)/ 3 x DN6300

Management Communication

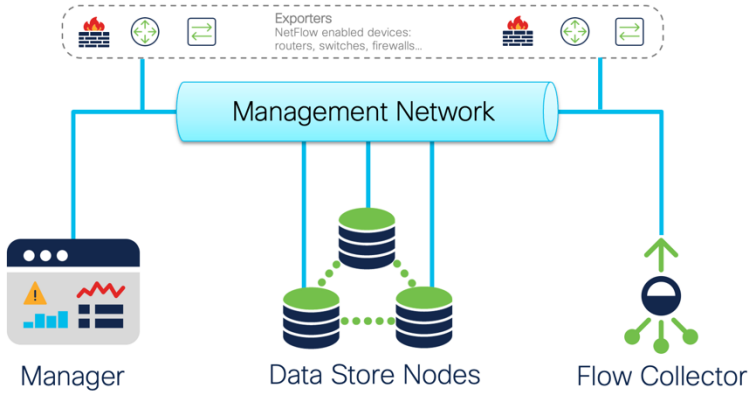
For best performance the Secure Network Analytics Manager should be co-located with the data nodes to ensure the optimal performance of queries. Each Secure Network Analytics appliance must have a routable IP address assigned to the eth0 management port.

Secure Network Analytics appliances all have a logical interface eth0 that is used for encrypted management and data communications between all appliances. This is also how configuration changes are pushed by the Manager to other appliances as well as queries made to the Data Store by the Manager.

When deploying physical Data Store appliances, the logical eth0 management port can be configured to use a SFP (BASE-T copper/Fiber) 1G/10G port or SFP+ (Base-CU/DAC/TwinAX) 10G port. While each appliance is configured independently, it is recommended to use the same media type and 10Gbps connectivity between all Secure Network Analytics appliances.

The Secure Network Analytics Manager establishes management of all the other Secure Network Analytics appliances in the ecosystem using Central Manager.

It is required that all Secure Network Analytics appliances including each data node in the Data Store deployment be on the same network (i.e., Layer-3 subnet).



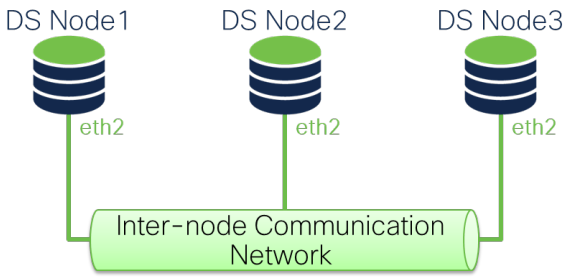
It is possible to have Secure Network Analytics appliances on different networks if they can reach the management (i.e., eth0) interface of the other appliances.

Ensure all ports are open that are necessary for inter-device communications. See the Communication Ports section at the end of the document for more information.

The Secure Network Analytics appliances must be managed by the same Manager. Ensure each appliance uses the same domain name server and NTP (Network Time Protocol) source.

Inter-node Communication

The data nodes are required to be co-located with each other to minimize latency for critical inter-node communication. Each Data Store Node will communicate on a *private non-routable* LAN that will be used for the inter-node communication. Traffic sent between data nodes is not encrypted and so it is required that the data nodes are installed so that they can reach each other via the private LAN IP address (do not install a date node in a different LAN or different Data Center. If you do then the installation will fail).



Cisco recommends the use of the Nexus 7000 and Nexus 9000 series switches for the inter-node communication switch.

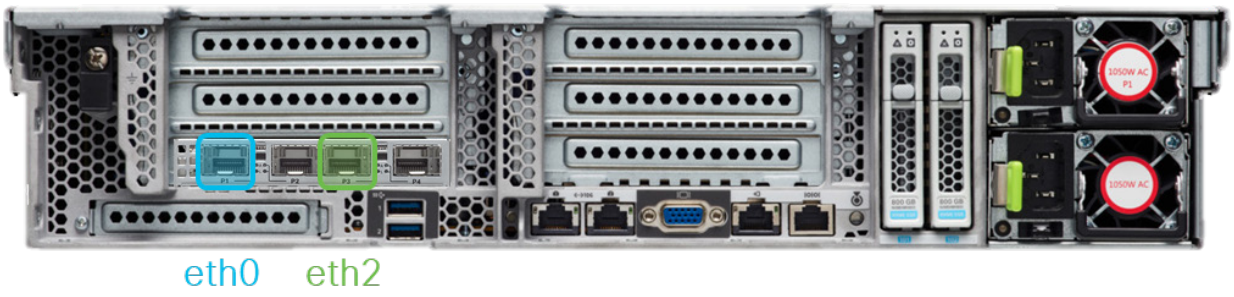
This design features a single link using the local eth2 interface on each data node to communicate with each data nodes in the deployment.

Each data node will use this link for writing data and other critical database communications.

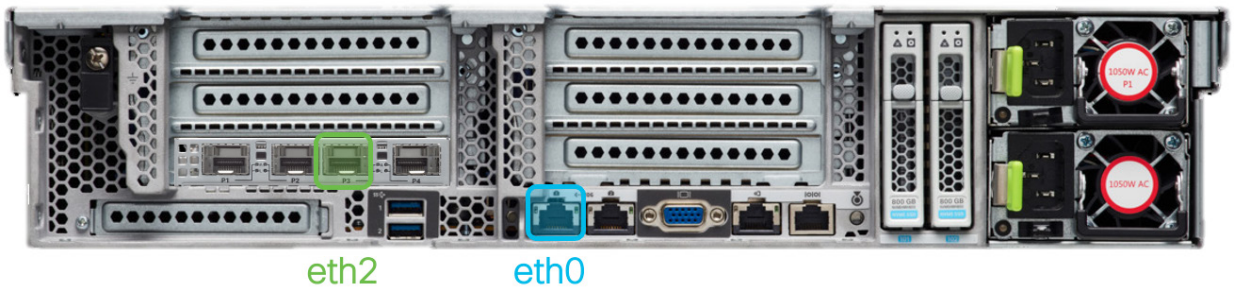
DS6200

There are two management configuration options with the DS6200 design, the default which uses the SFP port only, and an alternative which uses the two onboard RJ-45 ports and the first two SFP ports.

SPF mode is the default mode where the 10-Gbps Ethernet SFP+ ports are used for all network communication. In this configuration the left most SFP port is the eth0 management port and the third SFP+ port from the left is used for inter-node communication to a single inter-node communication switch as indicated below:



Mixed mode is an optional mode where a 1-Gb/10-Gb BASE-T Ethernet LAN port (RJ45 connector) for management is used and the third SFP+ port from the left is the 10-Gbps Ethernet SFP+ used for inter-node communication to a single inter-node communication switch as illustrated:



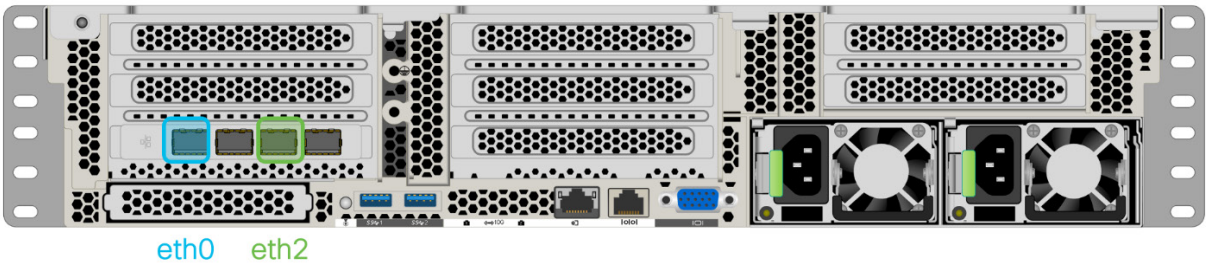
For additional details on appliance specifications refer the Data Store spec sheet:

https://www.cisco.com/c/en/us/support/security/stealthwatch/products-technical-reference-list.html

DN6300

There is one management configuration option with DN6300 design, and the default uses the SFP port.

By default, the 10-Gbps Ethernet SFP+ ports are used for all network communication. In this configuration the left most SFP port is the eth0 management port and the third SFP+ port from the left is used for inter-node communication to a single inter-node communication switch as indicated below:



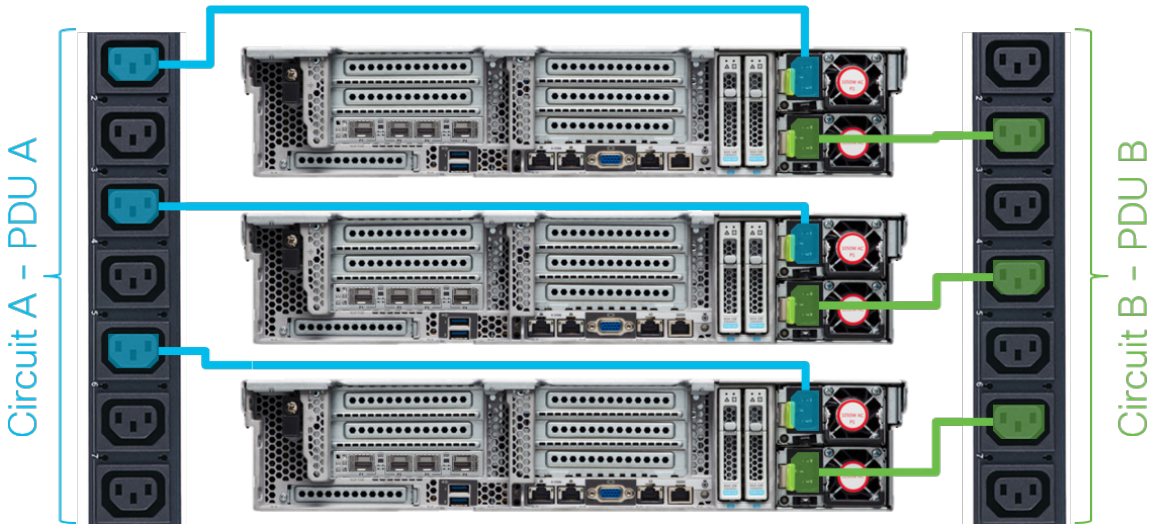
For additional details on appliance specifications refer the Data Store spec sheet:

https://www.cisco.com/c/en/us/support/security/stealthwatch/products-technical-reference-list.html

Power Redundancy

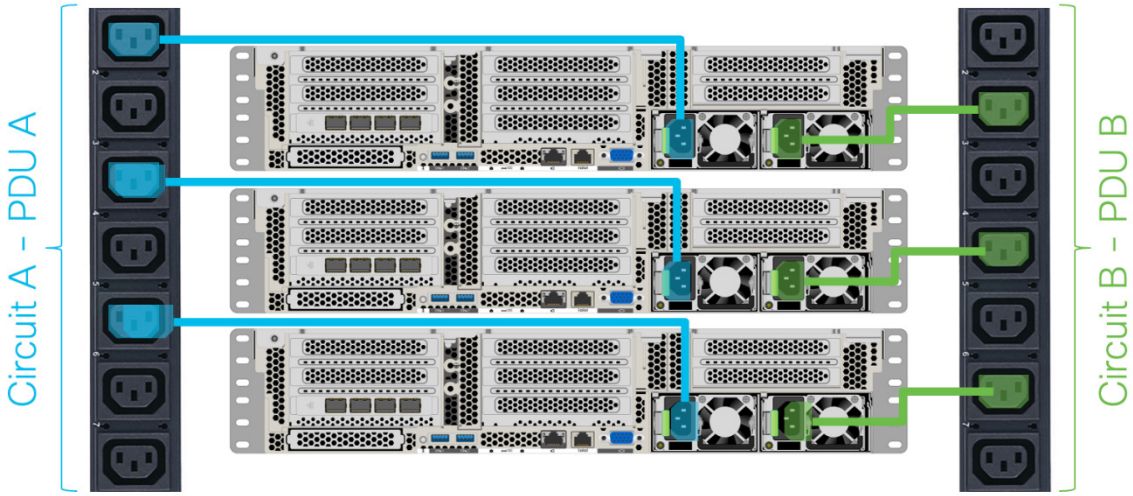
Each Data Store Node will be shipped with redundant PSUs labelled in the diagram below. This will ensure you can maintain power redundancy in case of an outage.

DS6200



PID	Description	Power
ST-DS6200	Secure Network Analytics Data Store 6200	Redundant [770W or 1050 W] AC 50/60 Auto Ranging (100V to 240V)

DN6300

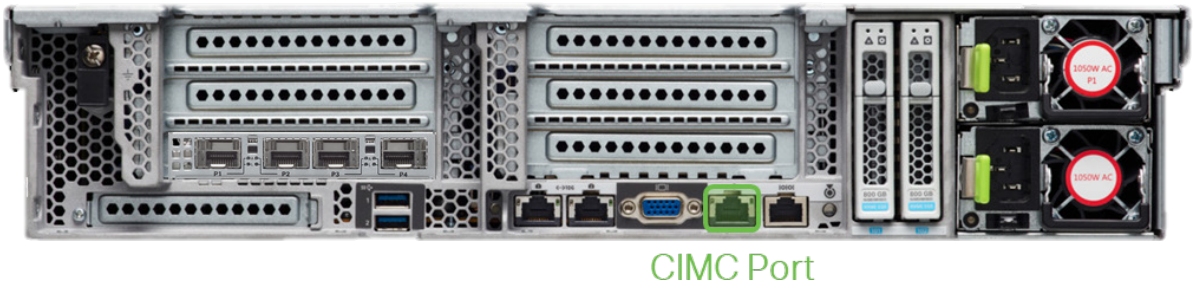


PID	Description	Power
ST-DN6300	Secure Network Analytics Data Store 6300	Redundant [1050 W] AC 50/60 Auto Ranging (100V to 240V)

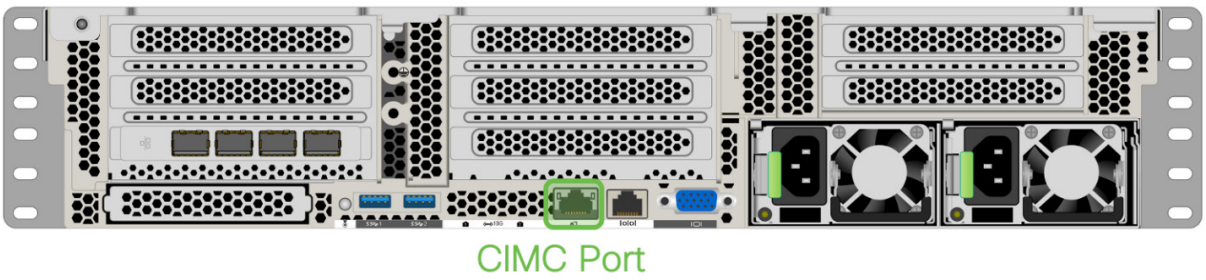
Out of Band Management

Out of band management can be configured using the Cisco Integrated Management Controller (CIMC) port. This port is labeled with a symbol on the back of the appliance as indicated in the picture below.

DS6200



DN6300



For further details on configuring CIMC for remote access please follow the guidance found in the Cisco UCS Installation Guide:

DS6200: https://www.cisco.com/c/en/us/td/docs/unified_computing/ucs/c/hw/C240M5/install/C240M5.html

DN6300: https://www.cisco.com/c/en/us/td/docs/unified_computing/ucs/c/hw/C245m6/install/c245m6.html

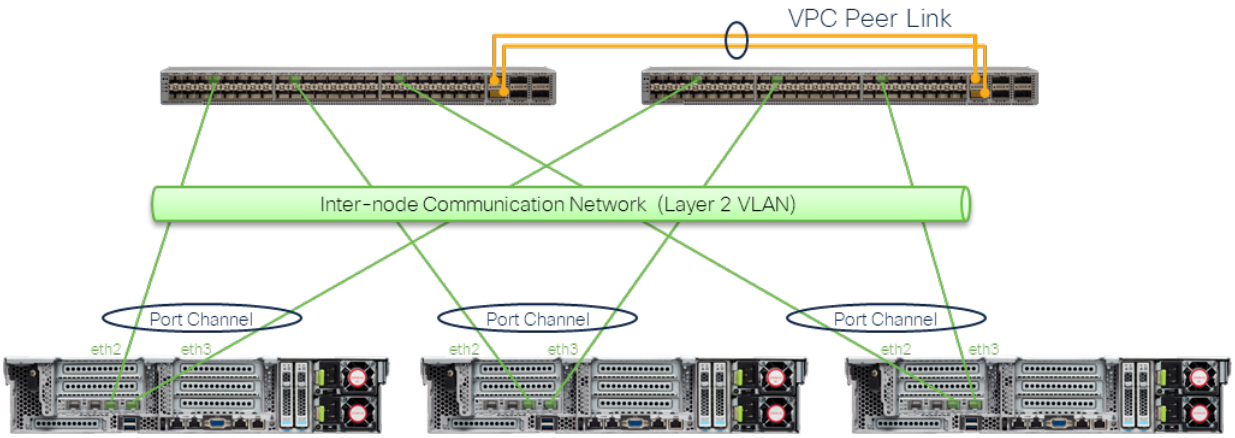
Adding Redundancy and Scale to the Hardware Design

Redundancy

Dual-switch architecture - In this redundant switch configuration, the inter-node communications have redundancy via a dual-link port channel (802.3ad LACP is used for the dual port channel configuration) using both the eth2 and eth3 ports on each data node. If using Cisco Nexus switches, we highly recommend Virtual Port Channeling:

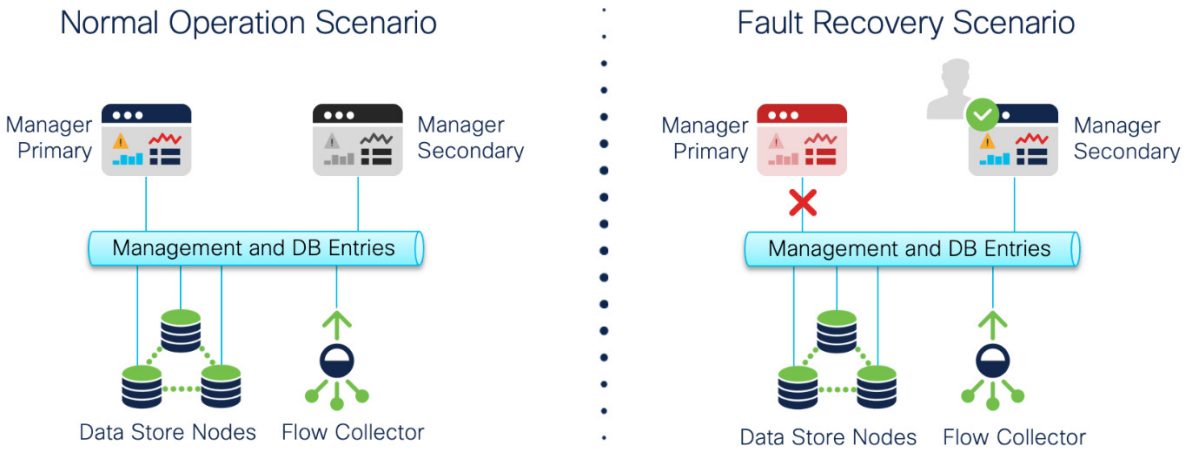
https://www.cisco.com/c/dam/en/us/td/docs/switches/datacenter/sw/design/vpc_design/vpc_best_practices_design_guide.pdf

The appliance configuration is completed in the Initial Setup Tool on each data node in the Data Store.



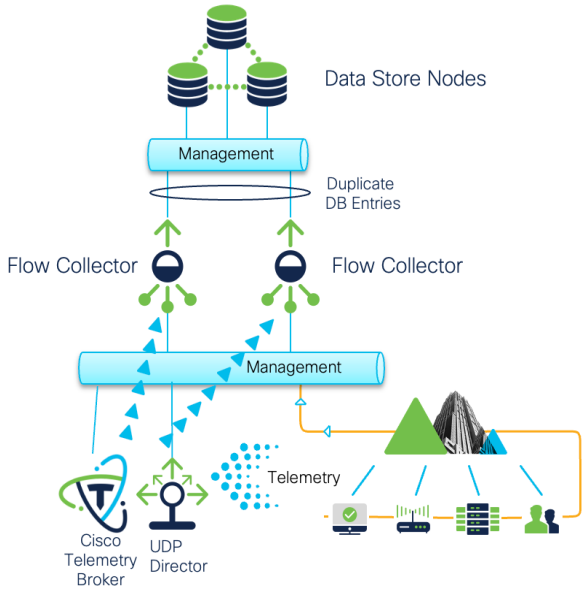
Redundant Manager (SMC) 2210/2300 – Adding a redundant Manager is the most common form for redundancy in Secure Network Analytics. The architecture includes a total of two Managers with one serving as the primary and the other as the secondary. The secondary operates in the pair as read only.

Configuration changes made to the primary are automatically synched to the secondary. If the primary Manager goes down the secondary must be manually promoted to primary before it is fully operational.



Redundant FCNF 4210/4300 – To prevent losing the ability to process flow data in the event a Flow Collector goes down, you can add multiple FCNF 4210's/4300's for redundancy. The flow data would be sent to both Flow Collectors and both Flow Collectors would be managed by the same Manager. A flow that is sent to both collectors will count against the flow rate license twice. Keep this in mind when building a design.

Additionally, a UDP Director or Cisco Telemetry Broker would be recommended for the redundant flow collector option to prevent having to configure multiple destinations from the exporters.



High-Availability UDP Director 2210 – NetFlow is sent to a virtual IP address shared between two UDP Directors in the high-availability design for the UDP Director. One UDP Director will be the primary online device and the other UDP director will be the secondary offline device. If the primary node should fail, the secondary node takes over automatically and becomes the primary.

Two physical 1Gbps copper cross connect cables are required between the appliances. The device must be located within a range of each no more than 300 feet.

High-Availability Cisco Telemetry Broker 2300 – NetFlow is sent to a virtual IP address shared between two (or more) Telemetry Brokers in the high-availability design. One Telemetry Broker will be the Active online device (meaning it passes the telemetry) and the other Telemetry Broker(s) will be designated Passive (meaning it is not passing telemetry). If an active broker node stops passing telemetry or otherwise loses connectivity with Telemetry Broker, one of the Passive broker nodes is promoted to Active broker node and starts passing telemetry.

You can assign either a virtual IPv4 or virtual IPv6 address, or both, to a cluster. Telemetry Broker uses this virtual IP address to communicate with the cluster and promote Passive broker nodes to Active broker nodes when an Active broker node loses connectivity with Telemetry Broker.

You must configure the IPv4 or IPv6 VIP IP addresses to be in the same subnet as the primary IP addresses in the cluster, since the VIP must be in the same subnet. This ensures proper routing via the preconfigured Gateway and fast failover. If the VIP addresses are not in the same subnet as the primary IP addresses of the Telemetry Network interfaces, or if the Telemetry Network interfaces within a Cluster are configured with different subnets, then it is very likely that high availability will not work.

Retention and Scalability

Expanding the size of the Data Store – Expanding the size of the physical Data Store is accomplished by adding additional DS6200/DN6300 data nodes. The Secure Network Analytics Data Store supports a maximum of 12 DS6200s, which equates to 36 data nodes in total. Adding additional data nodes to the Data Store requires the data nodes to be registered with the Central Manager and additional database setup steps to bring them online.

The virtual Data Store can start with a single data node and be expanded from a single data node to three data nodes and then you can add N+1 data nodes there after (Ex; 1 to 3 then 3 to 4 or 3 to 6 data nodes) for improving the scale, performance, and storage.

Note: The Data Store does not support a 2 data node configuration.

For detailed steps on how to expand the Data Store find the documentation linked below:

https://www.cisco.com/c/en/us/support/security/stealthwatch/products-installation-guides-list.html

Scaling Flow Collection – Bottlenecks can occur when flow rates exceed a collector's maximum flows per second thresholds causing the Flow Collector to drop telemetry packets. To prevent loss of telemetry data a UDP Director or Cisco Telemetry Broker and additional Flow Collectors can be deployed.
If expanding an existing deployment, the recommended approach is to plan a maintenance window where the existing Secure Network Analytics Flow Collector can be taken offline. During this window the UDP Director/Cisco Telemetry Broker is configured to use the IP address of the existing single Flow Collector, preventing unnecessary changes to production switches and routers. The original Flow Collector and new Flow Collectors can be uniquely IP addressed and then flow forwarding rules can be defined within UDP Director/Cisco Telemetry Broker to distribute the flows across multiple collectors.

Standard Virtual Design

Single Host ESX/KVM Design – Virtual Single Node Deployment

Overview

For a virtual single Data Store node design use a single ESX/KVM host to deploy the Manager (SMCVE), Flow Collector (FCNFVE) and virtual data node (VDATANODE) on, as there is no data resiliency with a single data node. Of course, you could deploy the Manager, Flow Collector, and data node on separate hosts in the same data center, but this would not be an optimal design for performance when data is being queried.

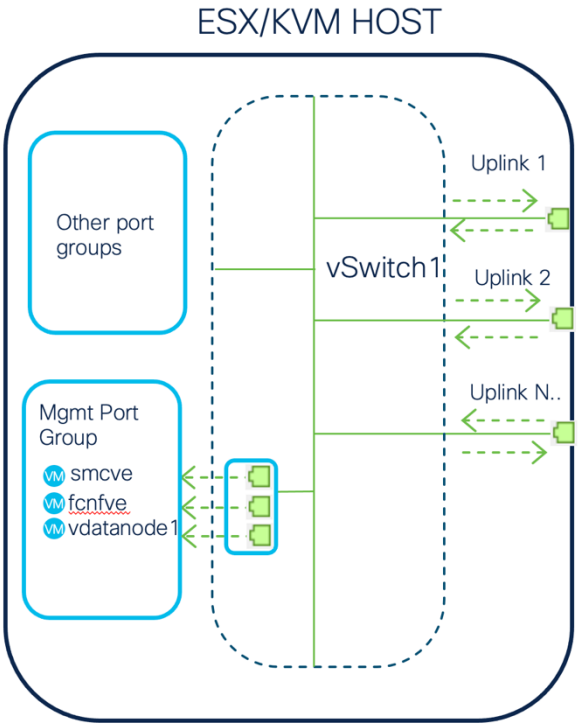
Below is a recommended design where each virtual appliance is deployed on the same ESX/KVM hosts. This design will scale to 225K FPS with minimum virtual machine resources required.

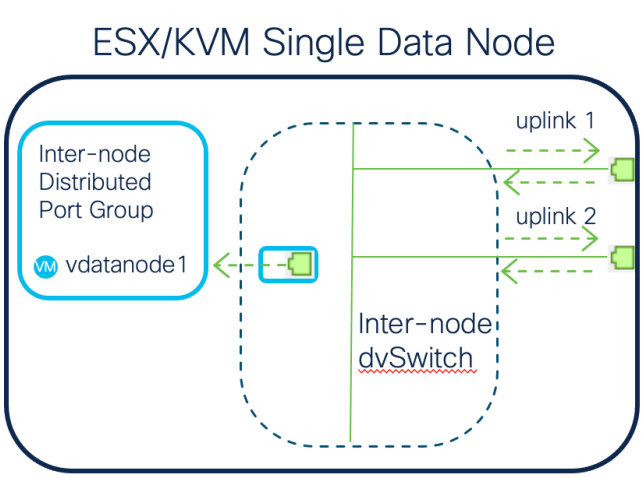
This is the easiest design to deploy and configure, but with all Secure Network Analytic devices stored on one ESX/KVM host, you will not have data resiliency if the ESX/KVM host fails.

Management Communication

Ensure that the mgmt eth0 vNIC on each Secure Network Device is minimally on the same vSwitch. It is recommended to segment these vNICs into a port group with a dedicated vSwitch.

This will ensure other VM's and their activity such as snapshots or intensive network activity does not impact the Secure Network Analytic devices.





Inter-node Interface Configuration

When you deploy the single data node, you will still be required to enable at a minimum two network interfaces. One for management (eth0) and one for inter-node (eth1) communication with other data nodes. If you do not do this, then the installation will fail. Of course, initially there will only be one data node, but this allows you to expand from 1 data node to 3+ at a later date. The inter-node eth1 vNIC on each data node is in a port group tagged with the same VLAN across each ESX/KVM host. It is recommended this port group is created on a separate vSwitch that is used only by the inter-node vNICs.

Multiple ESX/KVM Host Design – Virtual 3 Node Deployment

Overview

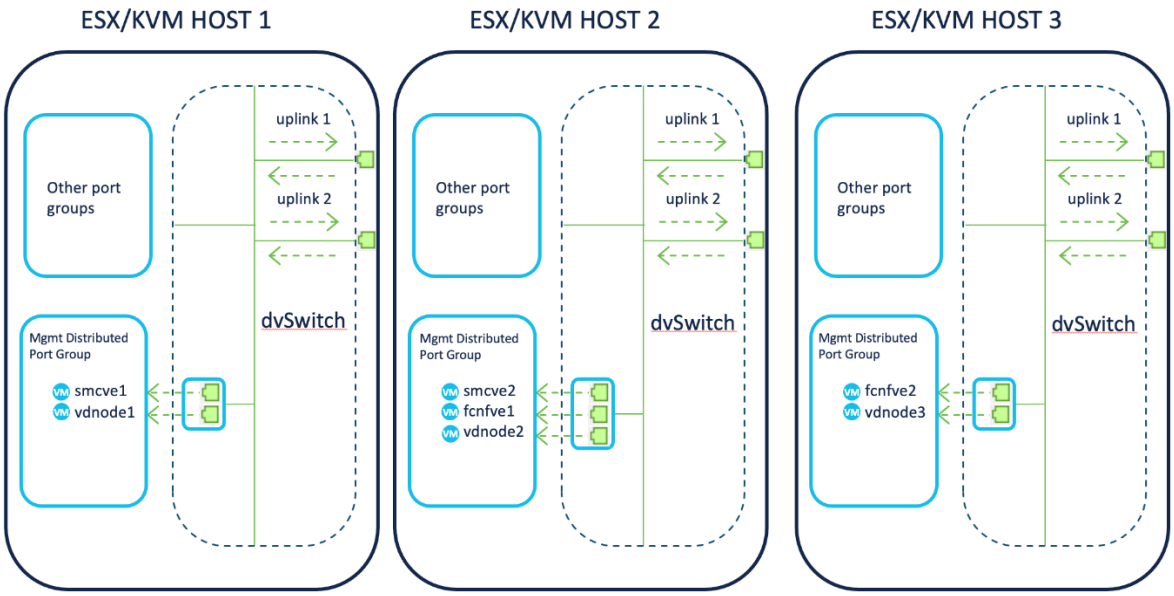
For virtual Data Store designs using multiple ESX/KVM hosts, it is recommended to deploy each virtual data node on a separate ESX/KVM host to ensure data resiliency across the Data Store. It also can help spread required virtual resources such as compute and memory across multiple hosts. Below is a recommended design where each virtual data node is distributed across three different ESX/KVM hosts.

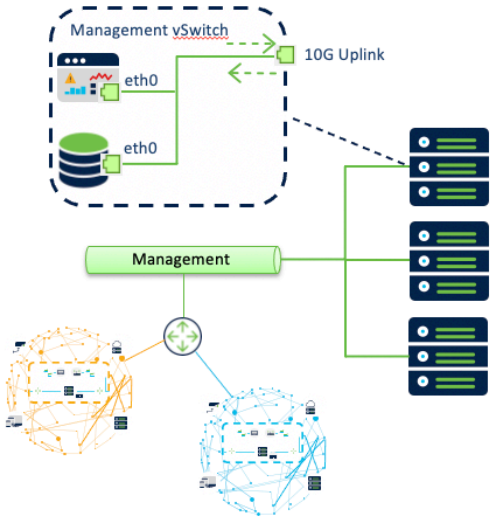
In this design it is recommended to deploy a primary and secondary Manger and two Flow Collectors. Having a primary and secondary Manager ensures that if an ESX/KVM host was to fail there would always be one Manager online. Having two flow collectors is important if a customer wants to store flow redundantly or needs multiple collectors to scale their flows per second ingest.

This design is the most resilient and provides for higher flow rates but requires the most resources. Take this into consideration for your design.

Management Communication

Ensure that the mgmt eth0 vNIC on each Secure Network Device is in a port group tagged with the same VLAN across each ESX/KVM host. It is recommended to use a distributed vSwitch and a distributed port group if possible.

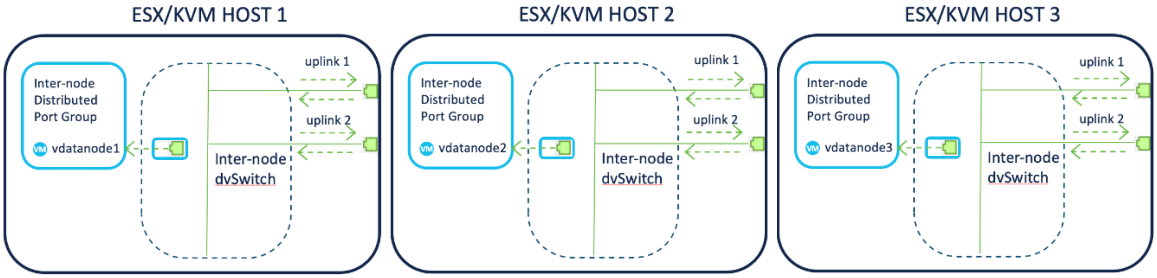




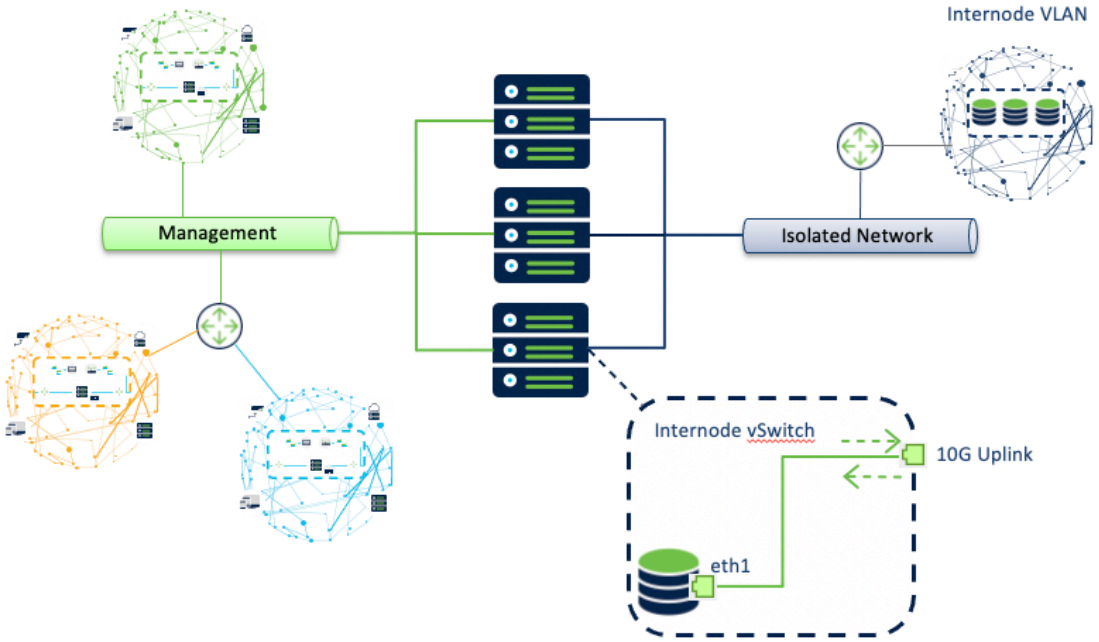
Ensure that the network device each uplink is connected to and its associated switchport interface has the correct VLAN configured to ensure all devices can communicate over their management interface.

Inter-node Interface Configuration

Ensure that the inter-node eth1 vNIC on each data node is in a port group tagged with the same VLAN across each ESX/KVM host. It is recommended this port group is created on a separate vSwitch that is used only by the inter-node vNICs.



Additionally, it is recommended to have redundant 10G uplinks. Ensure that the network device each uplink(s) is physically connected to and its associated switchport has the correct VLAN configured. This will ensure internode vNICs can communicate over their network.

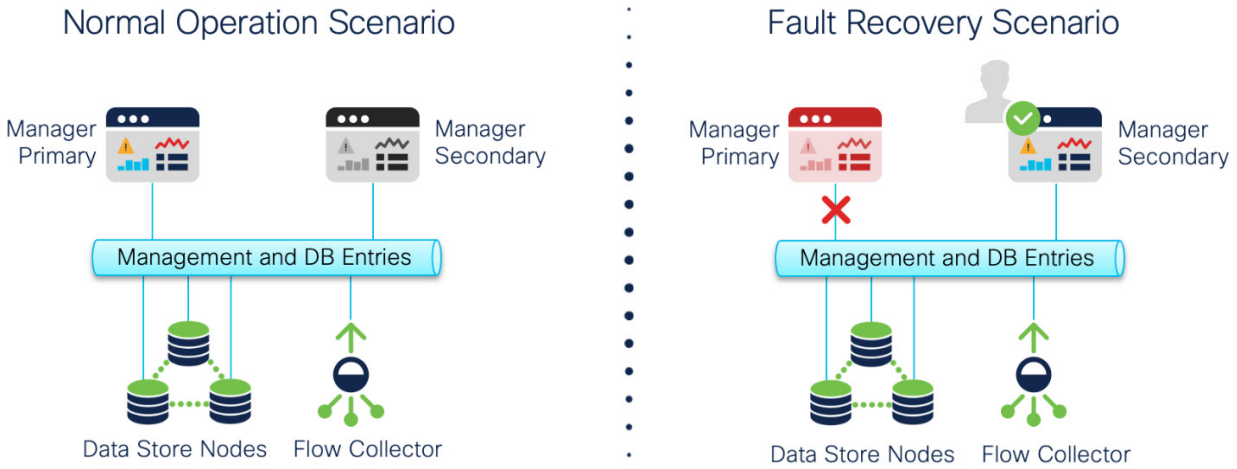


Adding Redundancy and Scale to Virtual Designs

Redundancy

Redundant Manager VE – Adding a redundant Manager is the most common form for redundancy in Secure Network Analytics. The architecture includes a total of two Managers with one serving as the primary and the other as the secondary. The secondary operates in the pair as read only.

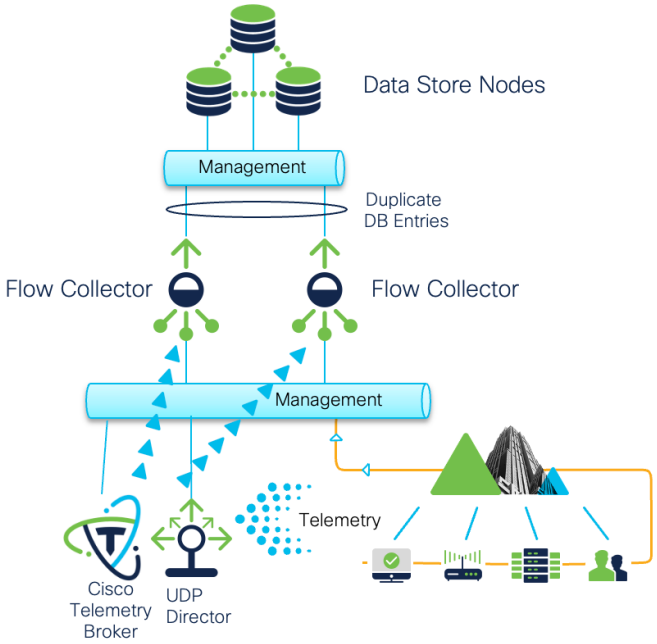
Configuration changes made to the primary are automatically synched to the secondary. If the primary Manager goes down the secondary must be manually promoted to primary before it is fully operational.



In virtual deployments deployed across multiple ESX/KVM hosts it is recommended to deploy a Manager on two different hosts. This ensures in the event of a failure to one of the ESX/KVM hosts an SMC will be available.

Redundant Flow Collector VE – To prevent losing the ability to process flow data in the event a Flow Collector goes down, you can add multiple Flow Collector VE's for redundancy. The flow data would be sent to both Flow Collectors and both Flow Collectors would be managed by the same Manager. A flow that is sent to both collectors will count against the flow rate license twice. Keep this in mind when building a design.

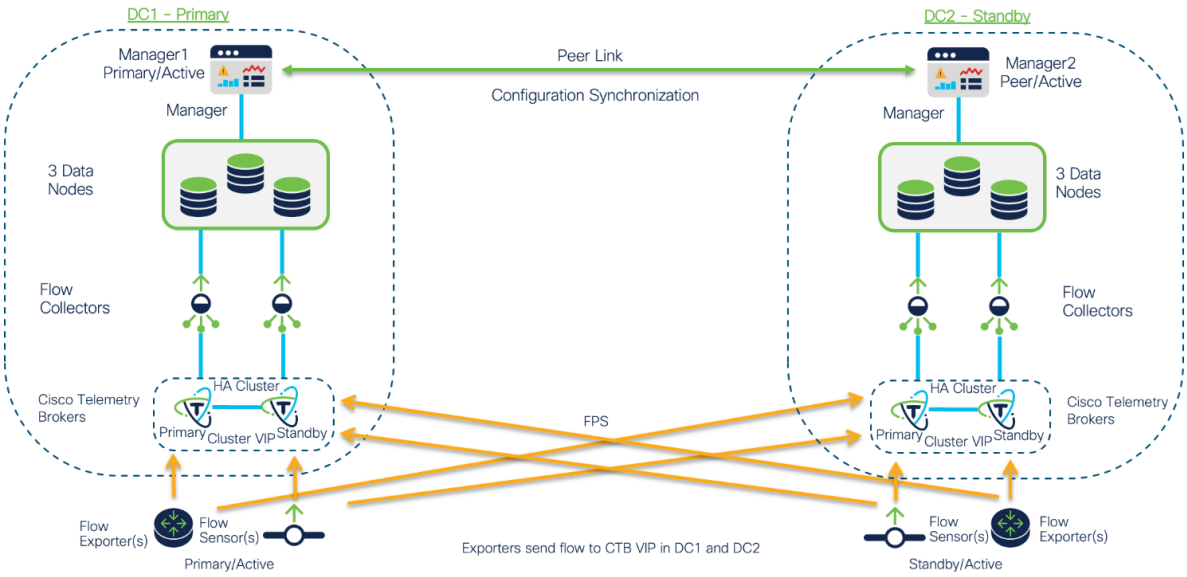
Additionally, a UDP Director or Cisco Telemetry Broker would be recommended for the redundant flow collector option to prevent having to configure multiple destinations from the exporters.



Redundancy with Dual Data Centers (Peer Sites) – To prevent losing the ability to process flow data in the event an entire data center goes down, you can add one Manager, data node(s), Flow collector(s) and Cisco Telemetry Brokers per Data Center. Both managers will be active (users can us both managers and run reports etc) and configuration changes will be synchronized between the Managers. The Managers will only see their own Data Store domain and not the Data Store domain of the other Manager.

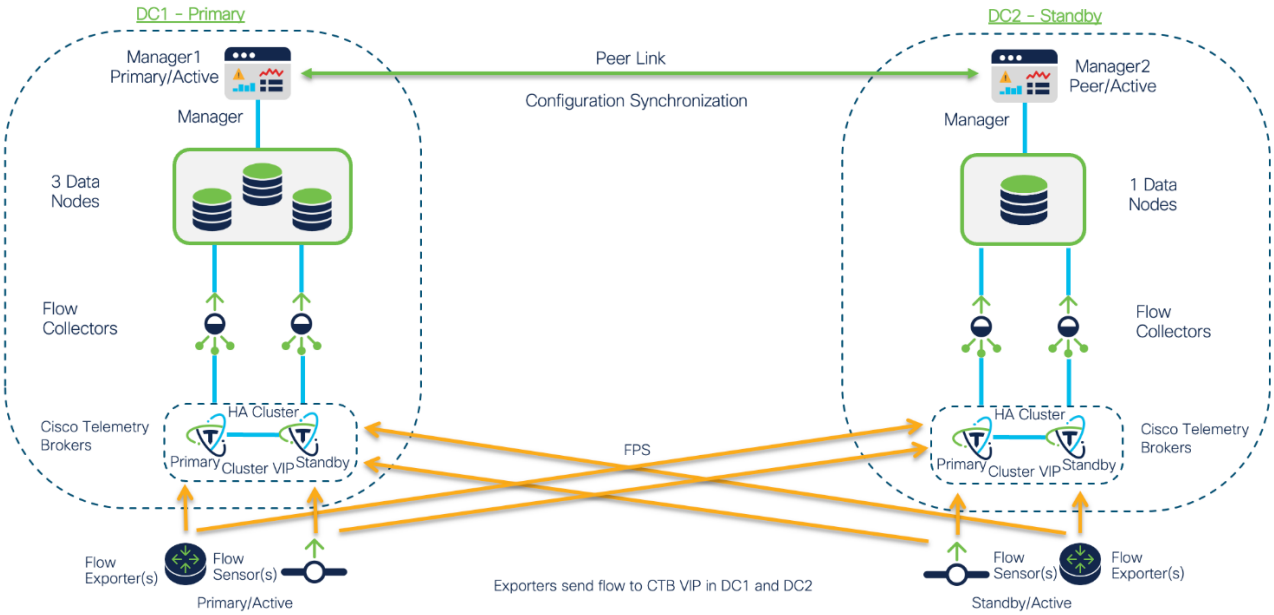
Option 1 – You can deploy a Manager, 3 data node Data Store, 2 Flow Collectors and two Cisco Telemetry Brokers in the primary and standby data centers. Configuration changes between the Managers will be synchronized using a peer link between the Managers (they will need to be able to communicate with each other between the data centers). Flow data will be sent to the virtual IP of the Cisco Telemetry Brokers (which would be deployed in high availability mode) and the active Telemetry Broker will send the flow data to both Flow collectors. The Flow collectors will then send the data to the Data Store data nodes. If one Telemetry Broker appliance fails, then the other will automatically take control. If one of the Flow collectors fail, then the other Flow Collector will still be online and be able to process the flows. If one of the Data Store data nodes fail, then the others will take over with no data loss. If the Manager fails or if the entire data center becomes unavailable (may be due to a disaster or maintenance) then the Standby data will become the primary.

Option 1 – Dual DC (3 dnode) with Primary/Active and Peer/Active Managers



Option 2 – This option is identical to option 1 with the exception that you deploy only one Data Store data node (physical or virtual) in the Standby Data Center. This offers those customers a choice to save money by deploying a single virtual or hardware data node in the Standby data center. Majority of our customers fit within the limitations for the single data node deployment when it comes to 225,000 FPS for Single data node (virtual) or 500,000 FPS for Single data node (physical).

Option 2 - Dual DC (3 dnode & 1 dnode) with Primary/Active and Peer/Active Managers



The advantages of option 2 is a reduction in the Capex and highly optimized footprint for the Standby DC. While supporting Geo-redundancy between DCs.

In summary redundancy can be provided at the data node level (with 3 plus data nodes provides data resiliency, if one node should fail) at the rack level (3 plus nodes deployed in different racks on different power supplies in the event of power failure at the rack level) and at the site level (using Geo Redundancy with Peer Sites feature).

Retention and Scalability

Scaling Flow Collection – Bottlenecks can occur when flow rates exceed a collector's maximum flows per second thresholds causing the Flow Collector to drop telemetry packets. To prevent loss of telemetry UDPD/Cisco Telemetry Brokers VE's and additional Flow Collector VE's can be deployed. The recommended approach is to plan a maintenance window where the Secure Network Analytics Flow Collector can be taken offline. During this window the UDP Director/Cisco Telemetry Broker is configured to use the IP address of the existing single Flow Collector, preventing unnecessary changes to production switches and routers. The original Flow Collector and new Flow Collectors can be uniquely IP addressed and then flow forwarding rules can be defined to distribute the flows per second load across multiple collectors.

Transitioning to a Data Store Architecture

With release 7.4.2 you can transition Flow Collectors to a Data Store architecture and take advantage of improved query and reporting performance, while demanding geographic redundancy. Additional benefits of the Data Store Architecture is increased telemetry retention (90 days to 1 year and beyond) and the capability for the Flow Collector to ingest telemetry from multiple sources (NetFlow, NVM flow and Firewall logs) all on one appliance. The Data Store also allows Converged Analytics to be enabled, which connects the detections to security frameworks like MITRE to get enough technical context to describe and understand how an adversary is acting.

You can transition existing Flow Collector 4K and 5K models and re-use existing Managers and Flow Sensors. All that needs to be done is to upgrade the appliances to release 7.4.2 prior to the transition. The transition can be easily done in several simple steps as shown below.

Step 1 – Upgrade all existing appliances and Flow Collectors 4K and 5K to 7.4.2

Step 2 – Add a Data Store Domain and deploy a Data Store data node in it.

Step 3 – **Start** the transition by configuring the Flow Collector to send all new telemetry to the Data Store data node.

Note: Manager will query the Data Store for new telemetry and the data base on the Flow Collector for older telemetry. Older telemetry stays on the Flow Collector.

- Step 4 – Once the data on the Flow Collector data based has aged out (typical after 30 days) then **finish** the transition on the Flow Collector. Once completed the Flow Collector will continue to send all new telemetry to the Data Store and the data base on the flow collector will be removed. This frees up resources on the Flow Collector and allows it to run up to three times faster than before.

The following shows in more detail "The How" for the transition process.

Transition Setup

- From the Manager web UI
  - Step 1: Create a Data Store domain
  - Step 2: Setup sync between non-Data Store domain to Data Store domain
  - Step 3: Sync the domains

Initiate Transition

- From the Manager CLI (SystemConfig as root)
  - Step 4 – Add the data node(s) to Central Manager
  - Step 5 – Enable SSH on the Data Store
  - Step 6 – Initialize the Data Store
  - Step 7 – Pick the flow collector and domain for transitioning
  - Step 8 – Acknowledge the flow collector transition

Monitor Transition

- From the Manager web UI
  - Central Manager>Inventory tab will show a transition flag (Data Store Transition) next to the flow collector
  - Central Manager>Data Store tab will show "Oldest Record (days ago)" for NetFlow, NVM and Firewall logs.
  - Once there is 30 days for each then the transition can be completed

Complete Transition

- From the Manager CLI (SystemConfig as root)
  - Step 9 – Select Data Store then Complete Transition and then the flow collector to transition
  - Step 10 – Acknowledge to complete the transition (note all old data on the flow collector will be deleted)

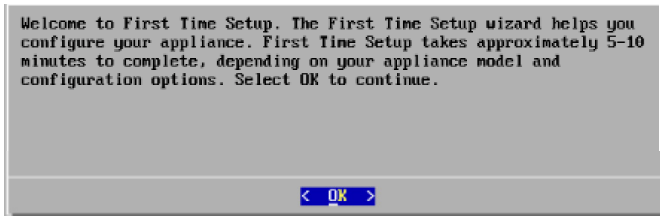
Software Installation and Configuration

Initial Setup Tool

The Initial Setup Tool will be used to configure the interfaces used and the management interface IP address settings. It is required that you use the same subnet (all appliances on the same Layer-3 network/subnet) when assigning management IP addresses for each appliance. The configuration order is the Manager configured first, followed by the data node(s) and then the data store Flow Collector.

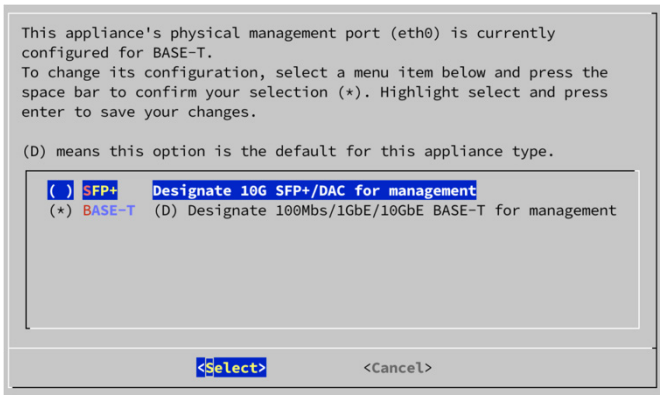
Note: If you place one data node in a different data center and on a different Layer-3 subnet for the management network, then the initialization of the data store will fail. The management interfaces for the data nodes MUST be on the same layer 3 network/subnet.

After configuring the management interface IP address settings, the appliance will reboot to continue the IST process.



For the Manager (2210 M5 hardware only), data node (M5 hardware only) and Flow Collector (4210 M5 hardware only) appliances there is an additional step to choose (after the Welcome message) the appliances physical management port (eth0).

For SMC2210/FC4210 it can be 10G SFP+ or BASE-T 100Mbps/1GbE/10GbE. Note BASE-T is the default for the SMC2210 and FC4210. Whereas SFP+ is the default for the DS6200.

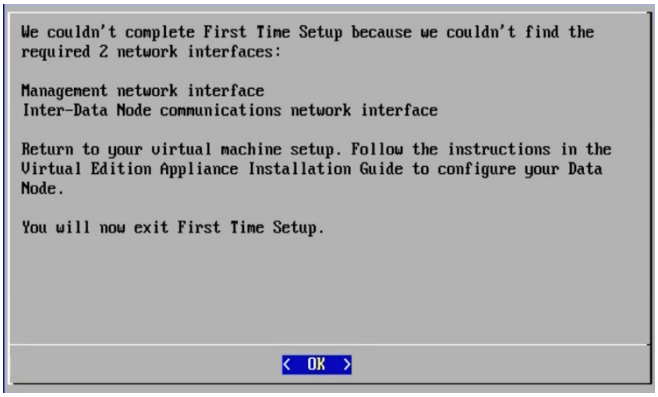


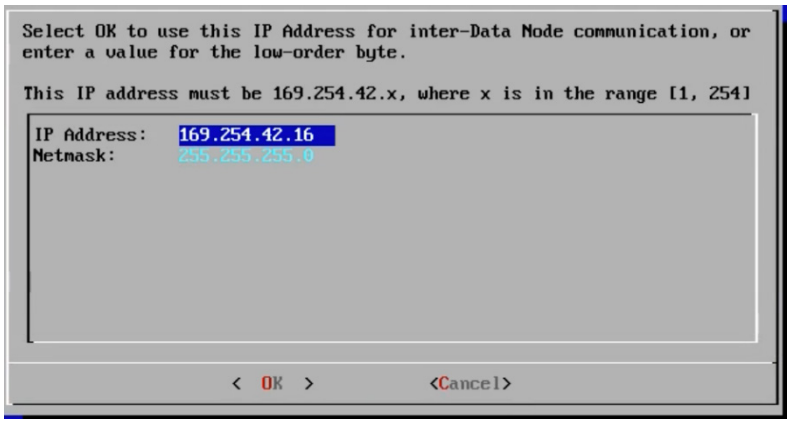
Note: Screen shot above is from Manager/Flow Collector not Data Node

For SMC2300 and FC4300 (M6 hardware appliances) there is no additional step as these appliances only support SFP+ ports. So, no configuration is required by the user.

For the data node appliances, the IST will have an additional option to configure the interface configuration and IP address settings for the inter-node communications. There are two interface configuration options for hardware data nodes. The first being a single link and the second being a port channel.

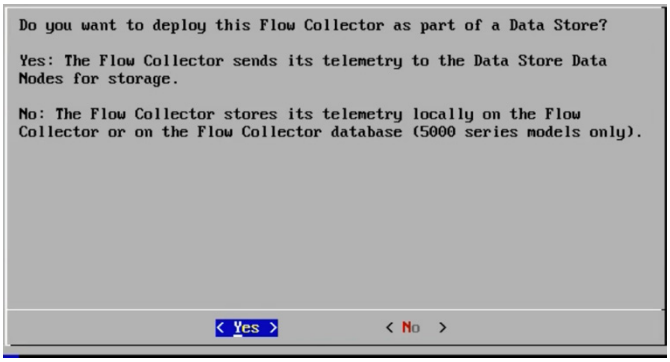
Note: For virtual data node deployments make sure to have configured at least two network interfaces. One for management and one for inter node communication. Otherwise, you will see the message below.





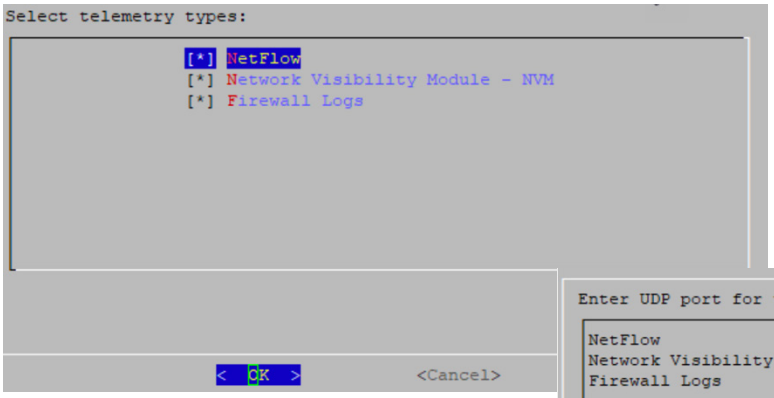
If one Management and at least one inter-data node interface is found, then you can configure the inter-data node interface with an IP address. Recommend accepting the default IP address automatically chosen for you. You can change the last octet of the IP address if required but it must reside within the 169.254.42.0/24 subnet.

Flow Collector appliances there is an additional step to choose whether the device will be part of a Data Store deployment.



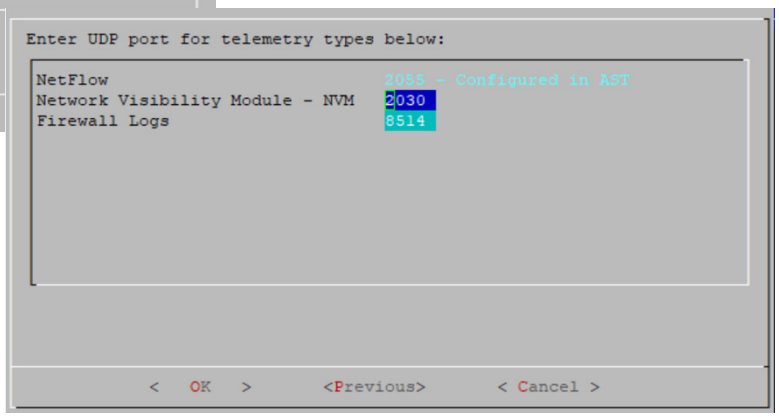
After you choose to configure your Flow Collector for use with Data Store, you cannot change this configuration. Select Yes only if you plan to deploy a Data Store to your network. Once the configuration has been completed you must RFD the appliance to change the Data Store selection.

For a Flow Collector appliance that is configured to work with a data store then you will be asked to select which telemetry types (and which UDP ports to use) to ingest on the Flow Collector. You can select all telemetry types or select one or more types. The types that can be selected are NetFlow, Network Visibility Module – NVM and Firewall Logs.



Note: You can update the telemetry settings after the IST configuration using the Flow Collector Advance settings.

Note: If the root or sysadmin passwords are forgotten, the Recovery option within SystemConfig can be used to reset the passwords.



Appliance Setup Tool

After the Initial Setup Tool, the next step is to complete the Appliance Setup Tool (AST). The AST is accessed via a web browser by navigating to the management IP address of each appliance. The appliances must be configured in priority order starting with the Primary Manager. Following it, the AST must be completed on each data node then the Flow Collector(s).

The Appliance Setup Tool is responsible for configuring the following areas:

1. Verifying the password settings and changing the admin web interface password.
2. Validating the management interface IP address settings.
3. Defining the device's host name, Network Domain (Ex vsphere.local). For the Manager define the Manager domain (name of the data store or non-data store domain the manager is in), Manager Domain type (Non Data Store or Data store).
4. Configuring the domain name server(s) the device will use.
5. Providing the NTP source(s) the device will use.
6. Register the device with Central Manager (register the Manager, data node(s) & flow collector(s)), this will be the primary Manager (SMC2210). When registering a Flow Collector select the domain (chose name configured in step 3) and Flow Collection Port.

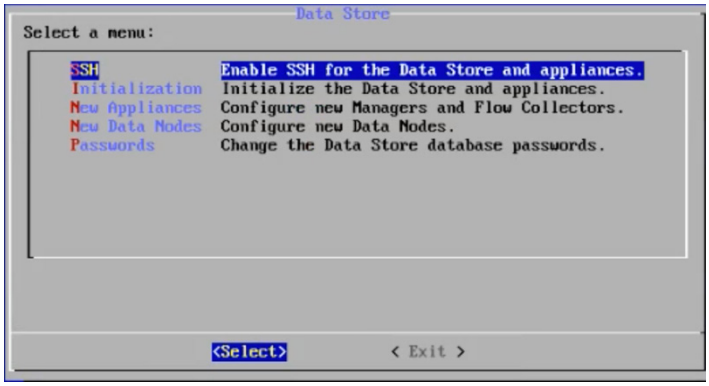
After completing the AST each device will reboot.

SystemConfig

If the root or sysadmin passwords are forgotten, the Recovery option within SystemConfig can be used to reset the passwords.

Manager Data Store Configuration

Following the IST and AST there is a final necessary step to configure the communication between the Manager and the data nodes. The configuration is completed on the Manager from the SystemConfig GUI and consists of completing the Datastore Secure Connectivity setup tool. You will need to enable SSH and then select "Initialization" from the Data Store option in the SystemConfig GUI. Do this after all Managers, Flow Collectors and data nodes have been added to the Central Manager Inventory.



The outcome of this step is a secure, SSL certificate based, communication channel between the Manager and the data nodes is established and database user accounts will be configured.

For more information on the IST, AST, and Manager Data Store Configuration:

https://www.cisco.com/c/en/us/support/security/stealthwatch/products-installation-guides-list.html

Summary

Effective network security analytics is not a function of applying just one technique. To stay ahead of evolving threats, a network visibility and analytics solution needs to be able to use a combination of methods. This begins by collecting and storing telemetry data at scale for comprehensive visibility on which analytical techniques such as behavioral modeling and machine learning can be applied.

Today with the new Secure Network Analytics Data Store customers can now meet the needs of the world's most demanding and dynamic enterprise-class and service provider networks. The Data Store architecture allows for ingesting up to 3M FPS and telemetry from over 30M endpoints which can be redundantly stored for more than half a year.

This new architecture transforms the capabilities of existing customers and offers a great solution to new customers who demand the fastest query response times and require data stored at scale for extended periods of time.

Learn more

Product and solution pages:

Cisco XDR

Cisco Secure Network Analytics

Cisco Secure Cloud Analytics

Encrypted Traffic Analytics

Secure Analytics Training Center

Reference

Communication Ports

Cisco highly recommends placing the data nodes within your firewall. Below is a list of the ports you will need to have open to ensure connectivity between the Secure Network Analytics Appliances.

From (Client)	To (Server)	Port	Protocol Purpose
Manager	Flow Collectors and data nodes	22/TCP	SSH, required to initialize Data Store database
Manager	Flow Collectors and data nodes	443/TCP	HTTPS, required for secure communications between appliances
Manager	data nodes	5444/TCP	Vertica Management Console secure communications
Flow Collectors	Manager	443/TCP	HTTPS, required for secure communications between appliances
NetFlow Exporters	Flow Collectors – NetFlow	2055/UDP	NetFlow ingestion
sFlow Exporters	Flow Collectors – sFlow	6343/UDP	sFlow ingestion
data nodes	All other data nodes	22/TCP	SSH, required to initialize Data Store database and for database administration tasks
data nodes	Manager	443/TCP	HTTPS, required for secure communications between appliances
data nodes	All other data nodes	4803/TCP **4803/UDP**	Inter-data node messaging service
data nodes	All other data nodes	4804/UDP	Inter-data node messaging service
data nodes	All other data nodes	5433/UDP	Vertica Messaging service monitoring
data nodes	All other data nodes	6543/UDP	Inter-data node messaging service
All Appliances data nodes	data nodes All other data nodes	5433/TCP6543/UDP	Vertica client connections Inter-data node messaging service
NTP Service All Appliances	Manager, Flow Collectors, and data nodes data nodes	123/UDP5433/TCP	NTP, required for time synchronization Vertica client connections
Administrator workstations NTP Service	Manager, FC and data nodes Manager, Flow Collectors, and data nodes	443/TCP123/UDP	WebUI Management NTP, required for time synchronization
Administrator workstations	Manager, FC and data nodes Manager, FC and data nodes	22/TCP443/TCP	Management/Troubleshooting WebUI Management

Design Checklist

Virtual or Hardware?

Which interface will be used for management on the Manager?

SMC2210
10-Gbps Ethernet SFP+ (default)
1-Gb/10-Gb BASE-T Ethernet LAN port (RJ45 connector)

SMC2300

1/10-Gbps Ethernet SFP+ (preferred 10-Gbps)

Which interface will be used for management on each Flow Collector?

FC4210
10-Gbps Ethernet SFP+ (default)
1-Gb/10-Gb BASE-T Ethernet LAN port (RJ45 connector)

FC4300
1/10-Gbps Ethernet SFP+ (preferred 10-Gbps)

Which interface will be used for management on each data node?

DS6200
10-Gbps Ethernet SFP+ (default)
1-Gb/10-Gb BASE-T Ethernet LAN port (RJ45 connector)

DN6300

1/10-Gbps Ethernet SFP+ (preferred 10-Gbps)

Management IP Addressing:
Manager: (Ex: 192.168.1.100/24)
DS Node1: (Ex: 192.168.1.111/24)
DS Node2: (Ex: 192.168.1.112/24)
DS Node3: (Ex: 192.168.1.113/24)
Flow Collector: (Ex: 192.168.1.200/24)
Default Gateway: (Ex: 192.168.1.1/24)

Inter-node communication IP Addressing :
(MUST be within the 169.254.42.0/24 subnet)
DS Node1: (Ex: 169.254.42.11/24)
DS Node2: (Ex: 169.254.42.12/24)
DS Node3: (Ex: 169.254.42.13/24)

Passwords:
(CLI) root:
(CLI) sysadmin:
(web) admin:
(CLI db initialization) dbadmin:
(CLI db initialization) readonlyuser:

Domain Name Servers:
DNS Server1:
DNS Server2:

Network Time Protocol Servers:
NTP Server1:
NTP Server2:
NTP Server3:

Host Name and Domains:
Host Name:

Network Domain: (Ex: vsphere.local)
Manager Domain: (Ex: SNA domain name for manager)
Manager Domain Type: (Ex: Data Store or Non-Data Store)

Certificate Configurations:

If CA specific certs are required for device identity certificates, they need to be generated for each device and added via Central Manager.