Micro Focus
ArcSight Platform

Technical Requirements for the ArcSight Platform

Monday, July 12, 2021

This Technical Requirements document describes the requirements and guidelines for the ArcSight Platform 21.1. The platform enables you to deploy a combination of security, user, and entity solutions into a single cluster within the Container Deployment Foundation (CDF) environment. The core services for this CDF environment, including the Dashboard and user management, are provided by a common layer called Fusion.

Recommended Platforms

Micro Focus recommends the tested platforms listed in this document.

Product	This Release	Upgrade from Version
ArcSight Command Center for Enterprise Security Manager	7.5.0	7.4.0
ArcSight Intelligence	6.3.0	6.2.0
ArcSight Fusion	1.3.0	1.2.0
ArcSight Layered Analytics	1.2.0	1.1.0
ArcSight Management Center	3.0	2.9x
ArcSight Platform	21.1.0	20.11.0 or 20.11.1
ArcSight Recon	1.2.0	1.1.1
ArcSight SOAR	3.1.0	3.0.0 or 3.0.1
Transformation Hub	3.5.0	3.4 or 3.4.1

> ⚠ On AWS, upgrade from Transformation Hub 3.4.0 to Transformation Hub 3.5.0 is not supported.

> ⚠ Customers running on platforms not provided in this document or with untested configurations will be supported until the point Micro Focus determines the root cause is the untested platform or configuration. According to the standard defect-handling policies, Micro Focus will prioritize and fix issues we can reproduce on the tested platforms.

Additional Documentation

The ArcSight Platform documentation library includes the following resources.

- Administrator's Guide for ArcSight Platform, which contains installation, user, and deployment guidance for the ArcSight software products and components that you deploy in the containerized platform.
- User’s Guide for Fusion in the ArcSight Platform, which is embedded in the product to provide both context-sensitive Help and conceptual information.
- ArcSight Platform Release Notes, which provides an overview of the products deployed in this suite and their latest features or updates.
- Product Support Lifecycle Policy, which provides information on product support policies.

Software Requirements

This section lists the software needed to install and run the ArcSight Platform.

Category	Operating System
Certified OS (minimal installation)	**For CDF:** • Red Hat Enterprise Linux 8.2 (x86, x64) • CentOS 8.2 (x86, x64)
	For the database: • Red Hat Enterprise Linux 7.9 (x86, x64) • CentOS 7.9 (x86, x64)
Supported OS (minimal installation)	**For CDF:** • Red Hat Enterprise Linux 8.2 (x86, x64) • Red Hat Enterprise Linux 8.1 (x86, x64) • Red Hat Enterprise Linux 7.9 (x86, x64) • Red Hat Enterprise Linux 7.8 (x86, x64) • Red Hat Enterprise Linux 7.7 (x86, x64) • CentOS 8.2 (x86, x64) • CentOS 8.1 (x86, x64) • CentOS 7.9 (x86, x64) • CentOS 7.8 (x86, x64) • CentOS 7.7 (x86, x64)
File systems	One of the following: • EXT3 • EXT4 (recommended) • Logical Volume Manager (LVM) • XFS
Data Collection	SmartConnector 7.14 or later
Browser	• Google Chrome • Mozilla Firefox Browsers should not use a proxy to access Container Deployment Foundation (CDF) applications because this might result in inaccessible web pages.

Supported Data Types and SmartConnectors/FlexConnector Types

This section describes the data types and SmartConnectors/FlexConnector types Intelligence supports.

Data Types	Supported Smart Connectors
Access	SmartConnector for Microsoft Windows Event Log – Native Application and System Event Support SmartConnector for Microsoft Windows Event Log – Unified Application and System Event Support
Active Directory	SmartConnector for Microsoft Windows Event Log – Native Application and System Event Support
VPN	SmartConnector for Microsoft Network Policy Server File SmartConnector for Pulse Secure Pulse Connect Secure Syslog SmartConnector for Citrix NetScaler Syslog SmartConnector for Nortel Contivity Switch Syslog
Web Proxy	SmartConnector for Microsoft Forefront Threat Management Gateway File SmartConnector for Squid Web Proxy Server File SmartConnector for Blue Coat Proxy SG Multiple Server File
Repository	FlexConnector Type - ArcSight FlexConnector Regex File

Additional Considerations

Consider the following:

- A fuller set of SmartConnectors is supported for those sources that provide data of relevance to the Intelligence analytics models. Micro Focus might need to examine sample logs to optimize analysis of data from this broader set of sources.
- For supported data types, Intelligence provides support for new devices that provide data of relevance to the Intelligence analytics models. For more information, see *Adding Support for New Devices* in the Administrators Guide for ArcSight Platform.
- Intelligence supports the SmartConnectors listed. However, additional capabilities you might deploy, such as Recon, might support a wider set of SmartConnectors/FlexConnector types.

- Micro Focus advises against configuring event aggregation for data to be processed by ArcSight Intelligence. If you wish to use ArcSight Intelligence with aggregated events, contact Micro Focus Customer Support.

Hardware Requirements and Tuning Guidelines

The section lists the guidelines for a deployment with all of the following software installed.

- Command Center for ESM
- Intelligence
- Recon
- SOAR
- Transformation Hub

Command Center for ESM Hardware Requirements and Tuning Guidelines

These guidelines apply to the requirements for deploying Command Center for ESM to a single node. You might have additional components deployed to that node, such as ESM, which have additional requirements.

The hardware requirements are based on dedicated resources allocations. In virtual environments, where there is a risk of over-subscription of the physical hardware, ensure the Fusion system meets these hardware requirements to avoid installation and functionality issues.

If you install Command Center for ESM on the same node as ESM server, you should keep some unused resource capacity on the node. For more information, see the Administrator's Guide for ArcSight Platform.

- "System Sizing" on the next page
- "Disk Space" on the next page

System Sizing

This section provides guidance for node requirements.

Category	Requirement
Worker nodes	1
vCores (per node)	8
RAM (per node)	32 GB

Disk Space

This section lists the minimum disk space needed to run Command Center on ESM. In some environments, you might deploy ESM Command Center with other capabilities, which would have additional disk space requirements.

Partition	Disk Space
/opt	200 GB
swap	16 GB
/home	50 GB

Intelligence Hardware Requirements and Tuning Guidelines

This section describes the requirements and guidelines for a deployment with Intelligence, Transformation Hub, Fusion, Recon, and Database installed.

Intelligence Workload

This section describes the total workload for Intelligence, which depends on the following factors.

The number of events collected by the SmartConnectors from the data sources and sent to the different storage components, that is, Elasticsearch, Transformation Hub, and the database.

The number of events and the number of entities processed by the Intelligence Analytics component to produce the Intelligence Analytics results that are sent to the different storage components, that is, Elasticsearch and the database.

- "Database Cluster" below
- "Hardware Requirements" below
- "Hardware Specification Metrics" on the next page

Database Cluster

Your deployment can have a non-collocated database cluster. In a non-collocated database cluster, the database is not deployed on the worker nodes in the CDF cluster. Instead, the database is deployed on dedicated nodes that make up the database cluster, and this cluster is not a part of the CDF cluster.

Hardware Requirements

The hardware requirements for Intelligence comprise the following:

- Processing requirements based on the Events per second (EPS) and the number of entities.
- Storage requirements based on the EPS, the number of entities, and the number of days' events.

Hardware Specification Metrics

The hardware specifications provided were determined for the following metrics.

Hardware	Metric
EPS	5000, 25000
Entities	15000
Master High Availability	Yes
Additional Storage Factor (additional storage required to avoid sizing errors)	20%
Replication Factor (replicas of data required to ensure data resiliency)	1
Intelligence Analytics Run Frequency	Once a Day on 1 Day's Events
Storage for Elasticsearch and the Database	30 Days
Storage for the System Sizing	30 Days

Use the given information to determine the processing and storage requirements for different values of the metrics.

Intelligence System Sizing

This section lists the system sizing used to determine the processing and storage requirements for the specified metrics.

- "System Sizing" below
- "Database Sizing" on the next page

System Sizing

Type	EPS	Number of Nodes	CPU per Node (core)	CPU per Node (threads)	RAM per Node (GB)	Database storage per Node (GB)	Elasticsearch Storage per Node	Transformation Hub Storage per Node	System Storage per Node	Total Storage per Node (GB)
Master	5000	3	8	16	64	-	-	-	-	500
	25000	3	8	16	64	-	-	-	-	500

Type	EPS	Number of Nodes	CPU per Node (core)	CPU per Node (threads)	RAM per Node (GB)	Database storage per Node (GB)	Elasticsearch Storage per Node	Transformation Hub Storage per Node	System Storage per Node	Total Storage per Node (GB)
Worker	5000	3	12	24	64	"Database Sizing" below	3459.233	488.162	800	4747.395
	25000	6	24	48	128	"Database Sizing" below	8648.083	1220.405	800	10668.49

Database Sizing

EPS	Number of nodes	CPU per Node (core)	CPU per Node (threads)	RAM per Node (GB)	Database Storage per Node	System Storage per Node	Total Storage per Node
5000	3	12	24	64	2613.18	500	3113.18
25000	3	24	48	128	13065.9	500	13565.9

Intelligence Processing Requirements

This section lists the processing requirements. You will need to tune Intelligence Analytics based on the events per second (EPS) in your environment.

- "5000 EPS" below
- "25000 EPS" on the next page

5000 EPS

The following table provides the Intelligence processing requirements for the specified metrics.

Component	Number of Instances	CPU per Instance	RAM per Instance (GB)	Total CPU for Component	Total RAM for Component (GB)
Intelligence UI	1	1	0.2	1	0.2
Intelligence API	1	1	1	1	1
H2	1	1	1	1	1

Component	Number of Instances	CPU per Instance	RAM per Instance (GB)	Total CPU for Component	Total RAM for Component (GB)
Intelligence Exports	1	1	1	1	1
HDFS NameNode	1	1	0.5	1	0.5
HDFS DataNode	3	1	0.5	3	1.5
Logstash	6	2	2	12	12
Intelligence Analytics Driver	1	1	5	1	5
Intelligence Analytics Executor	21	1	5	21	105
Elasticsearch Master	1	2	2	2	2
Elasticsearch data	3	10	10	30	30

Intelligence Analytics Tuning Parameters

Parameters	Values
Parallelism	32
Number of Executors	21
Number of Cores per Executor	2
Memory per Executor (GB)	6
Driver Memory (GB)	5
esBatchEntries	0
esBatchBytes (MB)	5

Increase the number of Logstash instances if the Kafka partitions are increased and there is sufficient CPU and RAM.
Increase the number of Executors if there is sufficient CPU and RAM.

25000 EPS

The following table provides the Intelligence processing requirements for the specified metrics.

Component	Number of Instances	CPU per Instance	RAM per Instance (GB)	Total CPU for Component	Total RAM for Component (GB)
Intelligence UI	1	1	0.2	1	0.2
Intelligence API	1	1	1	1	1
H2	1	1	1	1	1
Intelligence Exports	1	1	1	1	1
HDFS NameNode	1	1	0.5	1	0.5
HDFS DataNode	6	1	0.5	6	3
Logstash	12	2	2	24	24
Intelligence Analytics Driver	1	1	5	1	5
Intelligence Analytics Executor	48	1	7	48	336
Elasticsearch Master	1	2	2	2	2
Elasticsearch data	6	16	16	96	96

Intelligence Analytics Tuning Parameters

Parameters	Values
Parallelism	40
Number of Executors	48
Number of Cores per Executor	2
Memory per Executor (GB)	7
Driver Memory (GB)	5
esBatchEntries	0
esBatchBytes (MB)	15

> Increase the number of Logstash instances if the Kafka partitions are increased and there is sufficient CPU and RAM.
> Increase the number of Executors if there is sufficient CPU and RAM.

Elasticsearch and Database Storage Requirements

This section lists the storage requirements for Elasticsearch and the database, which is incremental. It encompasses the storage capacity for both the raw events and the Intelligence Analytics data.

Component	EPS	Number of Instances	Disk Size per Instance per Day (GB)	Total Disk Size per Day (GB)
Elasticsearch	5000			
		3	3459.233	10377.699
	25000			
		6	8648.083	51888.498
Database	5000			
		3	2613.18	7839.54
	25000			
		3	13065.9	39197.7

Transformation Hub Storage Requirements

The section describes the storage for Transformation Hub, which is non-incremental and is a buffer for storing only the raw events.

The following are applicable for storing events in Transformation Hub:

- Events are stored only for the Kafka retention period. Default is 2 days.
- Events beyond the maximum Kafka partition size are removed. Default is 60 GB.
- The storage capacity is independent of the number of entities.

The maximum storage for Transformation Hub is determined by the following formula:

```
Maximum storage = Number of Kafka Partitions * Maximum Partition Size * Number
of Kafka Instances
```

The default compression used is GZIP (recommended).

Recon Hardware Requirements and Tuning Guidelines

This section describes the requirements and guidelines for Recon. These hardware requirements for Recon are based on dedicated resource allocations. In virtual environments, where there is a risk of over subscription of the physical hardware, ensure that the Recon system meets these hardware requirements to avoid installation and functionality issues.

The total workload for Recon depends on your data received through SmartConnectors or ArcSight Enterprise Security Manager (ESM) and on the number of events captured by those data sources each day. For example, each day, your environment might have thousands of events.

At the same time, someone might be updating details about the events or new information can be coming in about the entities associated with the events. Recon must be able to process all of these types of transactions. Thus, this document lists the requirements for small, medium, and large workloads.

Micro Focus based these recommendations on the maximum workload achievable while still maintaining stability of the system resources in our labs. It is possible you might need to further adjust the tuning values for satisfactory performance in your environment.

> The system sizing was tested in an ArcSight Recon environment without SSL communication.

- "Recon Small Workload System Sizing" below
- "Recon Medium Workload System Sizing" on page 15
- "Recon Large Workload System Sizing" on page 17
- "Recon Extra Large Workload System Sizing" on page 20

Recon Small Workload System Sizing

This section provides environment requirements for a small workload environment when deploying ArcSight Recon. It provides guidance for hardware requirements and tuning the performance of the workload. You might compare this information with the guidance for medium workloads.

- "Small Workload Distribution" on the next page
- "Small Workload System Sizing" on the next page
- "Small Workload Database Resource Pools Tuning" on the next page
- "Small Workload Transformation Hub Tuning" on page 15

Small Workload Distribution

The following table provides an example of how event ingestion activities might occur in a small workload.

Application	Category	Expected Workload
Microsoft Windows	Events per second	375
Fortinet Fortigate	Events per second	375
Infoblox NIOS	Events per second	375
Blue Coat, Check Point, Cisco	Events per second	375
ArcSight Recon	Events per second Searches (concurrent)	1500 3

Small Workload System Sizing

The following table provides a small workload example.

Category	Requirement
Single node (master and worker)	1
CPU cores (per node)	8
RAM (per node)	32
Disks (per node)	1
Storage per day (1x)	15 GB
Total disk space (1.5 billion events)	500 GB

Small Workload Database Resource Pools Tuning

The following table provides a small workload example.

Category	Property	Value
Database	tm_concurrency tm_memory	5 6,000

Resource pools	ingest_pool_memory_size ingest_pool_planned_concurrency	30% 12
Schedule	plannedconcurrency tm_memory_usage maxconcurrency	5 10,000 7

Small Workload Transformation Hub Tuning

The following table provides a small workload example.

Property	Quantity
# of Kafka broker nodes in the Kafka cluster	1
# of ZooKeeper nodes in the ZooKeeper cluster	1
# of Partitions assigned to the Kafka Topics*	12
# of Replicas assigned to each Kafka Topic	1
# of Message replicas for the __consumer_offsets Topic	1
Schema Registry nodes in the cluster	1
Kafka nodes required to run Schema Registry	1
# of CEF-to-Avro Stream Processor instances to start**	0/2

*Kafka topics - th-arcsight-avro, mf-event-avro-enriched, and, if connectors are configured to send to Transformation Hub in CEF format, th-cef

**If connectors are configured to send Avro format to Transformation Hub, you can set the quantity to 0 because there is no need to convert CEF to Avro.

Recon Medium Workload System Sizing

This section provides environment requirements for a medium workload environment when deploying ArcSight Recon. It provides guidance for hardware requirements and tuning the performance of the workload. You might compare this information with the guidance for small workloads.

- "Medium Workload Distribution" on the next page
- "Medium Workload System Sizing" on the next page
- "Medium Workload Database Resource Pools Tuning" on the next page
- "Medium Workload Transformation Hub Tuning" on page 17

Medium Workload Distribution

The following table provides an example of how event ingestion activities might occur in a medium workload.

Application	Category	Expected Workload
Microsoft Windows	Events per second	6000
Fortinet Fortigate	Events per second	7600
Infoblox NIOS	Events per second	4000
Blue Coat, Check Point, Cisco	Events per second	1900
ArcSight Recon	Events per second Searches (concurrent)	19500 3

Medium Workload System Sizing

The following table provides a medium workload example.

Category	Requirement
Single node (master and worker)	1 (G10 -L7700)
CPU cores (per node)	24
RAM (per node)	192
Disks (per node)	4 (7200 rpm)
Storage per day (1x)	0.9 TB
Total disk space (1.5 billion events)	10.8 TB

Medium Workload Database Resource Pools Tuning

The following table provides a medium workload example.

Category	Property	Value
Database	active_partitions tm_concurrency tm_memory	8 5 6,000

Resource pools	ingest_pool_memory_size ingest_pool_planned_concurrency	30% 12
Schedule	plannedconcurrency tm_memory_usage maxconcurrency	5 10,000 7

Medium Workload Transformation Hub Tuning

The following table provides a medium workload example.

Property	Quantity
# of Kafka broker nodes in the Kafka cluster	1
# of ZooKeeper nodes in the ZooKeeper cluster	1
# of Partitions assigned to the Kafka Topics*	12
# of Replicas assigned to each Kafka Topic	1
# of Message replicas for the __consumer_offsets Topic	1
Schema Registry nodes in the cluster	1
Kafka nodes required to run Schema Registry	1
# of CEF-to-Avro Stream Processor instances to start**	0/2

*Kafka topics - th-arcsight-avro, mf-event-avro-enriched, and, if connectors are configured to send to Transformation Hub in CEF format, th-cef

**If connectors are configured to send Avro format to Transformation Hub, you can set the quantity to 0 because there is no need to convert CEF to Avro.

Recon Large Workload System Sizing

This section provides environment requirements for a large workload environment when deploying ArcSight Recon. It provides guidance for hardware requirements and tuning the performance of the workload. You might compare this information with the guidance for medium workloads.

- "Large Workload Distribution" on the next page
- "Large Workload System Sizing" on the next page
- "Large Workload Database Resource Pools Tuning" on page 19
- "Large Workload Transformation Hub Tuning" on page 19

Large Workload Distribution

The following table provides an example of how event ingestion activities might occur in a large workload.

Application	Category	Expected Workload
Microsoft Windows	Events per second	40000
Fortinet Fortigate	Events per second	40000
Infoblox NIOS	Events per second	30000
Blue Coat, Check Point, Cisco	Events per second	10000
ArcSight Recon	Events per second Searches (concurrent)	120000 5

Large Workload System Sizing

The following table provides a large workload example.

CDF Infrastructure and Transformation Hub, Fusion, and Recon Capabilities

Category	Requirement
# of Nodes	3 (1 node with master, worker, and NFS; 2 nodes with worker only) (G10 - L7700)
CPU cores (per node)	24
RAM (per node)	192
Disks (per node)	4 (7500 rpm)
Storage per day (1x)	0.2 TB
Total disk space (250 billion events)	5 TB

Database

Category	Requirement
# of Database nodes	6 (G10 -L7700)
CPU cores (per node)	24
RAM (per node)	192
Disks (per node)	4 (7500 rpm)

Storage per day (1x)	4.2 TB
Total disk space (250 billion events)	20 TB
Redundant copies of data	1

Large Workload Database Resource Pools Tuning

The following table provides a large workload example.

Category	Property	Value
Database	active_partitions	8
	tm_concurrency	5
	tm_memory	10,000
Resource pools	ingest_pool_memory_size	30%
	ingest_pool_planned_concurrency	12
Scheduler	plannedconcurrency	5
	tm_memory_usage	10,000
	maxconcurrency	7

Large Workload Transformation Hub Tuning

The following table provides a large workload example.

Property	Quantity
# of Kafka broker nodes in the Kafka cluster	3
# of ZooKeeper nodes in the ZooKeeper cluster	3
# of Partitions assigned to the Kafka Topics*	72
# of Replicas assigned to each Kafka Topic	2
# of Message replicas for the __consumer_offsets Topic	3
Schema Registry nodes in the cluster	3
Kafka nodes required to run Schema Registry	3
# of CEF-to-Avro Stream Processor instances to start**	0/24
Kafka Override Parameters	**Quantity**

arcsight.eventbroker.kafka.KAFKA_NUM_IO_THREADS	256
arcsight.eventbroker.kafka.KAFKA_NUM_NETWORK_THREADS	52
arcsight.eventbroker.kafka.KAFKA_NUM_REPLICA_FETCHERS	145

*Kafka topics - th-arcsight-avro, mf-event-avro-enriched, and, if connectors are configured to send to Transformation Hub in CEF format, th-cef

**If connectors are configured to send Avro format to Transformation Hub, you can set the quantity to 0 because there is no need to convert CEF to Avro.

Recon Extra Large Workload System Sizing

This section provides environment requirements for an extra large workload environment when deploying ArcSight Recon. It provides guidance for hardware requirements and tuning the performance of the workload.

Also, this section provides recommendations for a 500K environment; however, these extrapolated requirements are based on our testing of EPS in the range 200K through 300K.

You might compare this information with the guidance for large workloads.

- "Extra Large Workload Distribution" below
- "Extra Large Workload System Sizing" on the next page
- "Extra Large Workload Database Resource Pools Tuning" on the next page
- "Extra Large Workload Transformation Hub Tuning" on page 22

Extra Large Workload Distribution

The following table provides an example of how event ingestion activities might occur in a extra large workload.

Application	Category	200K EPS	300K EPS	500K EPS
			Expected Workload	
Microsoft Windows	Events per second	60000	90000	150000
Fortinet Fortigate	Events per second	60000	90000	150000
Infoblox NIOS	Events per second	60000	90000	150000
Blue Coat, Check Point, Cisco	Events per second	20000	30000	50000
ArcSight Recon	Events per second	200000	300000	500000
	Searches (concurrent)	5	5	5

Extra Large Workload System Sizing

The following table provides a extra large workload example.

Transformation Hub/Recon

Category	200K EPS	300K EPS	500K EPS
	Expected Workload		
# of Worker nodes	3 (G10 -L7700)	4 (G10 -L7700)	5 (G10 -L7700)
CPU cores (per node)	24	24	24
RAM (per node)	192	192	192
Disks (per node)	4 (7200 rpm)	4 (7200 rpm)	4 (7200 rpm)
Storage per day (1x)	1 TB	1.5 TB	2.5 TB
Total disk space (250 billion events)	10 TB (transient)	10 TB (transient)	10 TB (transient)

Database

Category	200K EPS	300K EPS	500K EPS
	Requirement		
# of database nodes	9 (G10 -L7700)	15 (G10 -L7700)	25 (G10 -L7700)
CPU cores (per node)	24	24	24
RAM (per node)	192	192	192
Disks (per node)	4 (7200 rpm)	4 (7200 rpm)	4 (7200 rpm)
Storage per day (1x)	8 TB	12 TB	20 TB
Total disk space (500 billion events)	26 TB	26 TB	26 TB
Fault tolerance level*	0	1	1

* For more information, see the *High Availability* description in the Administrator's Guide for ArcSight Platform.

Extra Large Workload Database Resource Pools Tuning

The following table provides a extra large workload example.

Category	Property	200K EPS	300K EPS	500K EPS
		Value		

Database	active_partitions	8	8	8
	tm_concurrency	5	5	5
	tm_memory	10,000	10,000	10,000
Resource pools	ingest_pool_memory_size	30%	30%	30%
	ingest_pool_planned_concurrency	12	12	12
Scheduler	plannedconcurrency	5	5	5
	tm_memory_usage	10,000	16,000	24,000
	maxconcurrency	7	7	7
	max_parallelism	6	6	6

Extra Large Workload Transformation Hub Tuning

The following table provides a extra large workload example.

Property	**200K EPS**	**300K EPS**	**500K EPS**
	Quantity		
# of Kafka broker nodes in the Kafka cluster	3	4	5
# of ZooKeeper nodes in the ZooKeeper cluster	3	3	3
# of Partitions assigned to the Kafka Topics*	108	162	270
# of Replicas assigned to each Kafka Topic	2	2	2
# of Message replicas for the __consumer_offsets Topic	3	3	3
Schema Registry nodes in the cluster	3	3	3
Kafka nodes required to run Schema Registry	3	3	3
# of CEF-to-Avro Stream Processor instances to start**	0/36	0/48	0/80
Kafka Override Parameters	**Quantity**	**Quantity**	**Quantity**
arcsight.eventbroker.kafka.KAFKA_NUM_IO_THREADS	256	256	256
arcsight.eventbroker.kafka.KAFKA_NUM_NETWORK_THREADS	52	52	52
arcsight.eventbroker.kafka.KAFKA_NUM_REPLICA_FETCHERS	145	145	145

*Kafka topics - th-arcsight-avro, mf-event-avro-enriched, and, if connectors are configured to send to Transformation Hub in CEF format, th-cef

**If connectors are configured to send Avro format to Transformation Hub, you can set the quantity to 0 because there is no need to convert CEF to Avro.

SOAR Hardware Requirements and Tuning Guidelines

This section describes the SOAR workload. The total workload for SOAR depends on the number of correlation events ingested and cases processed daily.

For example, each day, your environment might have hundreds of correlation alerts sent to SOAR and playbooks are executed for them. At the same time, your analysts might be working on manual investigations, taking reports, etc. SOAR must be able to process all of these types of transactions. Thus, this document lists the requirements for small and medium workloads.

Micro Focus based these recommendations on the maximum workload achievable while still maintaining stability of the system resources in our labs. It is possible that you might need to further adjust the tuning values for satisfactory performance in your environment.

- "Small Workload System Sizing" below
- "Medium Workload System Sizing" below

Small Workload System Sizing

The following table provides guidance for a small workload environment (up to 250 incidents per day).

Category	Requirement
Single node (master and worker)	1
CPU cores (per node)	4
RAM (per node)	16
Disks (per node)	1
Total disk space (two years)	200 GB

Medium Workload System Sizing

The following table provides guidance for a medium workload environment (up to 1000 incidents per day).

Category	Requirement
Single node (master and worker)	1
CPU cores (per node)	6

RAM (per node)	24
Disks (per node)	1
Total disk space (two years)	750 GB

Network File System Options

This section describes the available network file system (NFS) options.

- Required File Systems
- NFS Minimum Directory Sizes

Required File Systems

The following table lists the minimum required file systems.

Category	Minimum Requirement
NFS Types	• Amazon EFS • HPE 3PAR File Persona • Linux-based NFS • NetApp
NFS Server Versions	• NFSv4 • NFSv3

NFS Minimum Directory Sizes

The following table lists the minimum required size for each of the NFS installation directories.

Directory	Minimum Size	Description
{NFS_ROOT_DIRECTORY}/itom-vol	130 GB	This is the CDF NFS root folder, which contains the CDF database and files. The disk usage will grow gradually.
{NFS_ROOT_DIRECTORY}/db-single-vol	Start with 10 GB	This volume is only available when you did not choose PostgreSQL High Availability (HA) for CDF database setting. It is for CDF database. During the install you will not choose the Postgres database HA option.

{NFS_ROOT_DIRECTORY}/db-backup-vol	Start with 10 GB	This volume is used for backup and restore of the CDF Postgres database. Its sizing is dependent on the implementation’s processing requirements and data volumes.
{NFS_ROOT_DIRECTORY}/itom-logging-vol	Start with 40 GB	This volume stores the log output files of CDF components. The required size depends on how long the log will be kept.
{NFS_ROOT_DIRECTORY}/arcsight-volume	10 GB	This volume stores the component installation packages.

Firewall Ports

This section lists the ArcSight Platform capabilities firewall ports. These ports need to be available when you deploy the associated capability.

- ArcMC
- CDF Vault
- CDF Management Portal
- Database
- Intelligence
- Kubernetes
- NFS
- SmartConnector
- SOAR
- Transformation Hub

ArcMC

Ports	Direction	Description
32080, 9000	Inbound	Used for Transformation Hub and ArcMC communication

CDF Vault

Ports (TCP)	Node	Description
8200	Master	Used by the `itom-vault` service which provides a secured configuration store All cluster nodes should be able to access this port for the client connection.
8201	Master	Used by the `itom-vault` service which provides a secured configuration store Web clients must be able to access this port for peer member connections.

CDF Management Portal

Ports (TCP)	Node	Description
3000	Master	Used only for accessing the CDF Management portal during CDF installation from a web browser Web clients must be able to access this port during the installation of CDF. After installation, web clients use port 5443 to access the CDF Management portal.
5443	Master	Used for accessing the CDF Management portal post CDF deployment from a web browser Web clients must be able to access this port for administration and management of CDF.
5444	Master	Used for accessing the CDF Management portal post CDF deployment from a web browser, when using two-way (mutual) SSL authentication Web clients must be able to access this port for administration and management of CDF, when using two-way (mutual) SSL authentication.

Database

The database requires several ports to be open on the local network. It is not recommended to place a firewall between nodes (all nodes should be behind a firewall), but if you must use a firewall between nodes, ensure the following ports are available:

Ports	Description
TCP 22	Required for the Administration Tools and Management Console Cluster installation wizard
TCP 5433	Used by database clients, such as vsql, ODBC, JDBC, and so on
TCP 5434	Used for Intra-cluster and inter-cluster communication
UDP 5433	Used for database spread monitoring
TCP 5438	Used as Management Console-to-node and node-to-node (agent) communication port
TCP 5450	Used to connect to Management Console from a web browser and allows communication from nodes to the Management Console application/web server
TCP 4803	Used for client connections

UDP 4803	Used for daemon to daemon connections
UDP 4804	Used for daemon to daemon connections
UDP 6543	Used to monitor daemon connections

Intelligence

In addition to the ports used by CDF, Transformation Hub, and the database, Intelligence uses the following ports when firewall is enabled. Ensure that the following ports are available:

Ports	Direction	Node	Description
TCP 30820	Inbound	Worker (HDFS Namenode)	Used for the database to connect to HDFS during Analytics processing
TCP 30070	Inbound	Worker (HDFS Namenode)	Used for Hadoop Monitoring Dashboard (Optional)
TCP 30010	Inbound	Worker (HDFS Datanodes)	Used for communication between the HDFS NameNode and the HDFS DataNodes
TCP 30210	Inbound	Worker (HDFS Datanodes)	Used by the database to establish secure communication with HDFS during Analytics processing

Kubernetes

Ports (TCP)	Node	Description
2380	Master	Used by the `etcd` component which provides a distributed configuration database All the master nodes should be able to access this port for the `etcd` cluster communication.
4001	Master	Used by the `etcd` component which provides a distributed configuration database All cluster nodes should be able to access this port for the client connection.
5000	Master	Used by `kube-registry` component which handles the management of container image delivery All cluster nodes should be able to access this port to communicate with the local container registry.

7443	Master	*(Conditional)* Used by the Kubernetes API server when you perform one of the following methods of installation: • Use the provided scripts • Install manually and on the same node as ESM All cluster nodes should be able to access this port for internal communication.
8443	Master	*(Conditional)* Used by the Kubernetes API server when you manually install and the installation is not on the same node as ESM. All cluster nodes should be able to access this port for internal communication.
8472	All nodes	*Uses UDP protocol* Used by the Flannel service component which manages the internal cluster networking All cluster nodes should be able to access this port for internal communication.
10250	All nodes	Used by the Kubelet service which functions as a local node agent that watches pod specifications through the Kubernetes API server All cluster nodes should be able to access this port for internal communications and worker node Kubelet API for exec and logs.
10251	All nodes	Used by `Kube-scheduler` component that watches for any new pod with no assigned node and assigns a node to the pod All cluster nodes should be able to access this port for internal communication.
10252	All nodes	Used by `kube-controller-manager` component that runs controller processes which regulate the state of the cluster All the cluster nodes should be able to access this port for internal communication.
10256	All nodes	Used by the `Kube-proxy` component, which is a network proxy that runs on each node, for exposing the services on each node All the cluster nodes should be able to access this port for internal communication.

NFS

Ports (TCP)	Node	Description
111	NFS server	Used by `portmapper` service All the cluster nodes should be able to access this port.
2049	NFS server	Used by `nfsd` daemon All the cluster nodes should be able to access this port. This port must be open even during a single- node deployment.
20048	NFS server	Used by `mountd` daemon All the cluster nodes should be able to access this port.

SmartConnector

Ports	Direction	Description
• 1515 (Raw TCP) • 1999 (TLS)	Inbound	Used by SmartConnector to receive events
• 9092 (Non-SSL) • 9093 (SSL)	Outbound	Used by SmartConnector to send data to Transformation Hub

SOAR

The SOAR cluster listens on the following NodePorts on all Kubernetes Master and Worker Nodes, but Micro Focus suggests you only use the ports on the master virtual IP.

Port	Description
32200	Data from ESM
32201	Data from Qradar
32202	Data from McAffee

Transformation Hub

Ports (TCP)	Direction	Description
2181	Inbound	Used by ZooKeeper as an inbound port
9092	Inbound	Used by Kafka during non-SSL communication
9093	Inbound	Used by Kafka when TLS is enabled
32080	Outbound	Used by Transformation Hub to send data to ArcMC
32181	Outbound	Used by ZooKeeper as an outbound port
443	Inbound	Used by ArcMC
9000	Inbound	Used by ArcMC
9999, 10000	Inbound	Used by the Transformation Hub Kafka Manager to monitor Kafka
39001, 39050	Outbound	Used by ArcMC to communicate with Connectors in Transformation Hub

Examples of Deployment Scenarios

You can deploy the ArcSight Platform capabilities in a variety of ways. The most basic deployment option is an all-in-one system that contains a limited number of capabilities on a single node. The single-node deployment is suitable for small workloads or to use as a proof-of-concept environment. For large workloads, you will need a multi-node environment, possibly with multiple masters. There are many scenarios and considerations involved in creating your environtment. Please see "Reviewing the Considerations and Best Practices" in the *Administrator's Guide to ArcSight Platform*.

This section provides some examples on how you could deploy one or more capabilities. Use these examples as a general guidance for planning your environment.

- "Multiple Master and Worker Nodes for High Availability" on the next page
- "Single Master, Multiple Workers, and a High-availability Database" on page 37
- "Everything on a Single Node" on page 40

Multiple Master and Worker Nodes for High Availability

In this scenario, which **deploys Intelligence with high availability**, you have three master nodes connected to three worker nodes and a database cluster. Each node runs on a separate, dedicated, connected host. All nodes have the same operating system, such as CentOS 7.8. Each Worker Node processes events, with failover to another Worker Node if a Worker fails. All of these environments require an external server to support NFS.

- **Diagram of this Scenario**
- **Characteristics of this Scenario**
- **Guidance for Node Configuration**

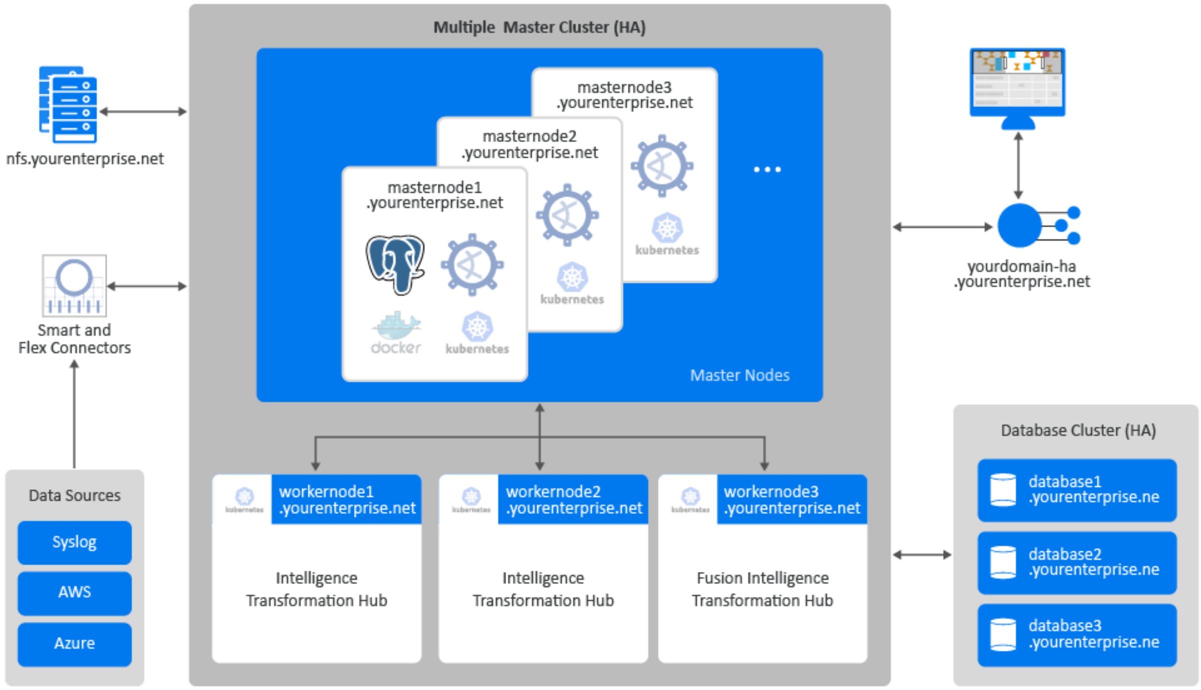
You can run this configuration in development and testing. It is the recommended configuration for highly available environments.

> If this scenario resembles your intended deployment, you might want to use the `example-install-config-intelligence-high_availability.yaml` config file with the ArcSight Platform Installer. See "Configuring the Deployed Capabilities" in the *Administrator's Guide for ArcSight Platform*.

The worker nodes process events, with failover to another worker node in the event of a worker failure. There are no single points of failure. You need a minimum of nine physical or VM environments: three dedicated master nodes, three or more dedicated worker nodes, and a database cluster. You also need a customer-provisioned, highly available NFS server (external NFS).

Diagram of this Scenario

***Figure 1.** Example deployment of Intelligence in a high-availability cluster*


Multiple Master Cluster (HA)
nfs.yourenterprise.net
masternode3 .yourenterprise.net
masternode2 .yourenterprise.net
masternode1 .yourenterprise.net
kubernetes
docker
...
Master Nodes
yourdomain-ha .yourenterprise.net
Smart and Flex Connectors
Data Sources
Syslog
AWS
Azure
workernode1 .yourenterprise.net
workernode2 .yourenterprise.net
workernode3 .yourenterprise.net
Intelligence Transformation Hub
Intelligence Transformation Hub
Fusion Intelligence Transformation Hub
Database Cluster (HA)
database1 .yourenterprise.ne
database2 .yourenterprise.ne
database3 .yourenterprise.ne

Characteristics of this Scenario

This scenario has the following characteristics:

- The Kubernetes cluster has three master nodes and three worker nodes, so that it can tolerate a failure of a single master and still maintain master node quorum.
- A FQDN hostname for a virtual IP is used so that clients accessing master nodes have a single reliable hostname to connect to that will shift to whatever is the current primary master node. For example, `yourdomain-ha.yourenterprise.net`.
- Transformation Hub's Kafka and ZooKeeper are deployed to all worker nodes with data replication enabled (1 original, 1 copy) so that they can tolerate a failure of a single node and still remain operational.
- Intelligence services, as well as Transformation Hub's platform and processing services, are allocated across all worker nodes so that, if one of the nodes fails, Kubernetes can move all of the components to the other node and still remain operational.
- Fusion is allocated to a single worker node.

- For the NFS configuration, use an NFS server that has high availability capabilities so that it is not a single point of failure.
- The database cluster has three nodes with data replication enabled (1 original and 1 copy) so that it can tolerate a failure of a single node and remain operational.

Guidance for Node Configuration

The following table provides guidance for deploying the capabilities across multiple nodes to support a large workload.

Node Name	Description	RAM	CPU Cores	Disk Space	Ports
Master Nodes 1-3 `masternodeNN.yourenterprise.net`	CDF Management Portal	256 GB	32	5 TB	"CDF Vault" on page 27 CDF Management Portal "Kubernetes" on page 29 "NFS" on page 31
Database Nodes 1-3 `databaseNN.yourenterprise.net`	Database	192 GB	24	28 TB	"Database" on page 28
Worker 1 `workernode1.yourenterprise.net`	Intelligence Transformation Hub	256 GB	32	5 TB	"Kubernetes" on page 29 "Transformation Hub" on page 32
Worker 2 `workernode1.yourenterprise.net`	Intelligence Transformation Hub	256 GB	32	5 TB	"Kubernetes" on page 29 "Transformation Hub" on page 32
Worker 3 `workernode1.yourenterprise.net`	Fusion Intelligence Transformation Hub	256 GB	32	5 TB	ArcMC Intelligence "Kubernetes" on page 29 "Transformation Hub" on page 32

Single Master, Multiple Workers, and a High-availability Database

In this scenario, which **deploys Intelligence with high availability on the ArcSight Database**, you have a single master node connected to three worker nodes and a database cluster. This scenario supports an environment with modest EPS and minimal number of nodes. However, it allows for futher scaling with multiple worker nodes. Each worker node runs on a separate, dedicated, connected host. All nodes have the same operating system, such as CentOS 7.8.

- **Diagram of this Scenario**
- **Characteristics of this Scenario**
- **Guidance for Node Configuration**

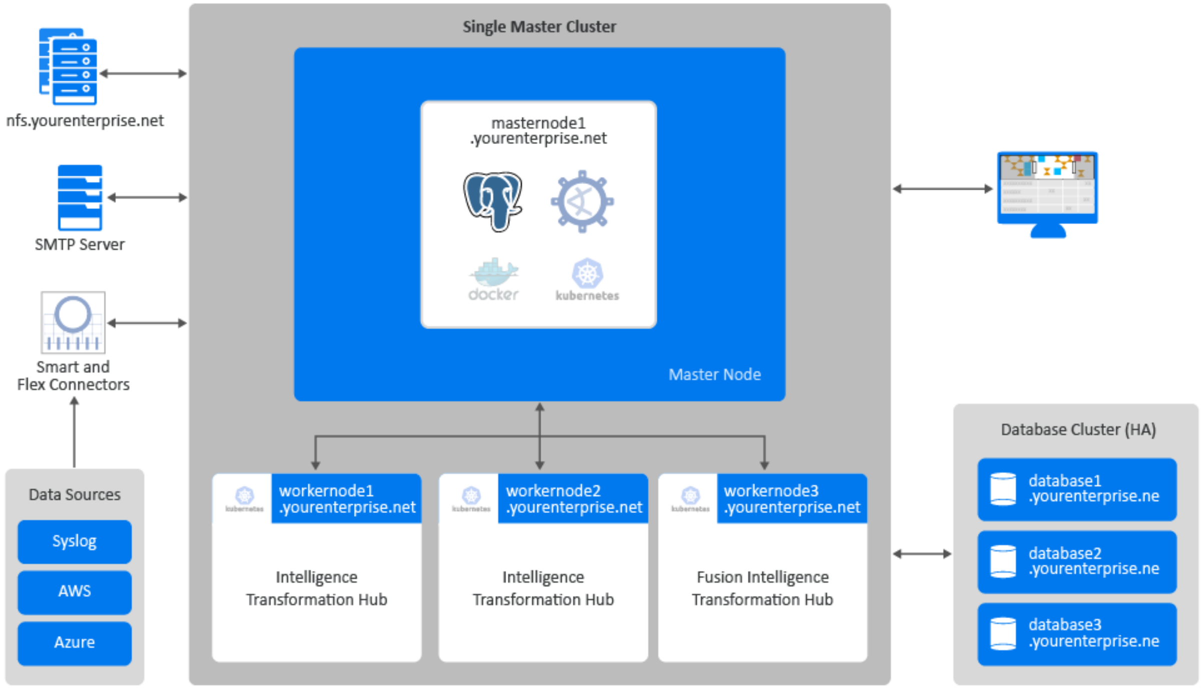
You can run this configuration in development and testing. This the recommended configuration for having a highly available database.

> If this scenario resembles your intended deployment, you might want to use the `example-install-config-intelligence-scale_db.yaml` config file with the ArcSight Platform Installer. See "Configuring the Deployed Capabilities" in the *Administrator's Guide for ArcSight Platform*.

You need a minimum of nine physical or VM environments: three dedicated master nodes, three or more dedicated worker nodes, and a database cluster. You also need a customer-provisioned, highly-available NFS server (External NFS) and an SMTP server.

Diagram of this Scenario

***Figure 1.** Example deployment of Intelligence and Recon*


Single Master Cluster
masternode1 .yourenterprise.net
docker
kubernetes
Master Node
nfs.yourenterprise.net
SMTP Server
Smart and Flex Connectors
Data Sources
Syslog
AWS
Azure
workernode1 .yourenterprise.net
Intelligence Transformation Hub
workernode2 .yourenterprise.net
Intelligence Transformation Hub
workernode3 .yourenterprise.net
Fusion Intelligence Transformation Hub
Database Cluster (HA)
database1 .yourenterprise.ne
database2 .yourenterprise.ne
database3 .yourenterprise.ne

Charcteristics of this Scenario

This scenario has the following characteristics:

- The Kubernetes cluster overall is not highly available since it is deployed with only one master node.
- A FQDN hostname for a virtual IP is used so that clients accessing master nodes have a single reliable hostname to connect to that will shift to whatever is the current primary master node. For example, `yourdomain-ha.yourenterprise.net`.
- Transformation Hub's Kafka and ZooKeeper are deployed to all worker nodes with data replication enabled (1 original, 1 copy) so that they can tolerate a failure of a single node and still remain operational.
- Transformation Hub's ZooKeeper is deployed to all worker nodes with data replication across the nodes so that it can tolerate a failure of a single node and still remain operational.
- Intelligence services, Fusion, and Transformation Hub's platform and processing services are allocated across all worker nodes so that, if one of the nodes fails, Kubernetes can move all of the components to the other node and still remain operational.

- The database cluster has three nodes with data replication enabled (1 original and 1 copy) so that it can tolerate a failure of a single node and remain operational.
- For the NFS configuration, use an NFS server that has high availability capabilities so that it is not a single point of failure.

Guidance for Node Configuration

The following table provides guidance for deploying the Intelligence across multiple nodes to support a medium workload.

Node Name	Description	RAM	CPU Cores	Disk Space	Ports
Master Node `masternode1.yourenterprise.net`	CDF Management Portal (Optional) Fusion	256 GB	32	5 TB	CDF Management Portal "Kubernetes" on page 29 "NFS" on page 31
Database Nodes 1-3 `databaseNN.yourenterprise.net`	Database	192 GB	24	28 TB	"Database" on page 28

Node Name	Description	RAM	CPU Cores	Disk Space	Ports
Worker 1 `workernode1.yourenterprise.net`	Intelligence Fusion Transformation Hub	256 GB	32	5 TB	ArcMC "Intelligence" on page 29 "Kubernetes" on page 29 "Transformation Hub" on page 32
Worker 2 `workernode2.yourenterprise.net`	Intelligence Fusion Transformation Hub	256 GB	32	5 TB	ArcMC "Intelligence" on page 29 "Kubernetes" on page 29 "Transformation Hub" on page 32
Worker 3 `workernode3.yourenterprise.net`	Fusion Intelligence Transformation Hub	256 GB	32	5 TB	ArcMC "Intelligence" on page 29 "Kubernetes" on page 29 "Transformation Hub" on page 32

Everything on a Single Node

In this scenario, which **deploys ESM Command Center on a single node**, you have the master and worker node co-located. You can include ArcSight SOAR as an optional capability on the same node.

- Diagram of this Scenario
- Scenario Characteristics
- Guidance for Node Configuration

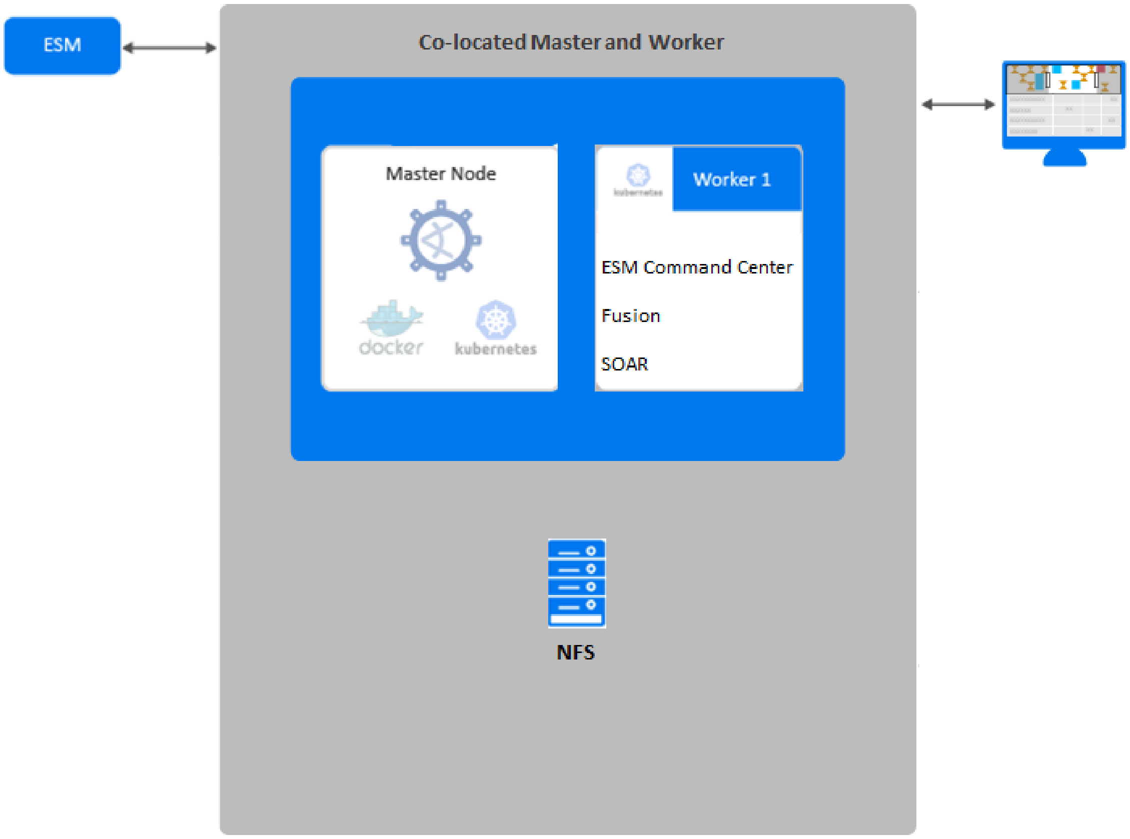
You can run this configuration in development and testing environments.

> ✓ If this scenario resembles your intended deployment, you might want to use the `example-install-config-esm_cmd_center-single-node.yaml` config file with the ArcSight Platform Installer. See "Configuring the Deployed Capabilities" in the *Administrator's Guide for ArcSight Platform*. The configuration in the example file describes a single-node deployment, but you can add more worker nodes to the file.

You need a minimum of one physical or VM environment to support master, worker, and NFS server on a single node. If you intend to install ESM Manager on the same machine, install ESM Manager first. ESM Manager uses port 8443, so master-api-ssl-port is set to a different port to avoid a conflict.

Diagram of this Scenario

Figure 2. *Example deployment of ESM Command Center on a single node*



Characteristics of this Scenario

This scenario has the following characteristics:

- The Kubernetes cluster has a single node to which you deploy ESM Command Center, Fusion, and (optionally) SOAR.

  > Having a single master node creates a single point of failure. As a result, if you intend to add worker nodes, this configuration is not recommended for high availability (HA) environments.

- FIPS 140 mode is enabled.
- For the NFS configuration, an NFS server that has high availability capabilities so that it is not a single point of failure.

Guidance for Node Configuration

The following table provides guidance for deploying ESM Command Center and associatd capabilities on a single node to support a small workload.

Node Name	Description	RAM	CPU Cores	Disk Space	Ports
Master Node `yourdomain-node.yourenterprise.net`	CDF Management Portal	256 GB	32	5 TB	"CDF Vault" on page 27 CDF Management Portal "Kubernetes" on page 29 "NFS" on page 31
Worker 1	ESM Command Center Fusion SOAR (optional)	32 GB	8	300 GB	ArcMC "Kubernetes" on page 29 "SOAR" on page 31

Send Documentation Feedback

If you have comments about this document, you can contact the documentation team by email. If an email client is configured on this computer, click the link above and an email window opens with the following information in the subject line:

Feedback on ArcSight Platform Technical Requirements (ArcSight Platform 21.1)

Just add your feedback to the email and click send.

If no email client is available, copy the information above to a new message in a web mail client, and send your feedback to Documentation-Feedback@microfocus.com.

We appreciate your feedback!

Legal Notices

Copyright Notice