Data collection best practices

Article • 11/27/2024

This section reviews best practices for collecting data using Microsoft Sentinel data connectors. For more information, see Connect data sources, Microsoft Sentinel data connectors reference, and the Microsoft Sentinel solutions catalog.

Prioritize your data connectors

Learn how to prioritize your data connectors as part of the Microsoft Sentinel deployment process.

Filter your logs before ingestion

You might want to filter the logs collected, or even log content, before the data is ingested into Microsoft Sentinel. For example, you might want to filter out logs that are irrelevant or unimportant to security operations, or you might want to remove unwanted details from log messages. Filtering message content might also be helpful when trying to drive down costs when working with Syslog, CEF, or Windows-based logs that have many irrelevant details.

Filter your logs using one of the following methods:

- **The Azure Monitor Agent**. Supported on both Windows and Linux to ingest Windows security events. Filter the logs collected by configuring the agent to collect only specified events.

- **Logstash**. Supports filtering message content, including making changes to the log messages. For more information, see Connect with Logstash.

> ⓘ **Important**
>
> Using Logstash to filter your message content will cause your logs to be ingested as custom logs, causing any free-tier logs to become paid-tier logs.
>
> Custom logs also need to be worked into analytics rules, threat hunting, and workbooks, as they aren't automatically added. Custom logs are also not currently supported for Machine Learning capabilities.

Alternative data ingestion requirements

Standard configuration for data collection might not work well for your organization, due to various challenges. The following tables describe common challenges or requirements, and possible solutions and considerations.

> ⓘ Note
>
> Many solutions listed in the following sections require a custom data connector. For more information, see Resources for creating Microsoft Sentinel custom connectors.

On-premises Windows log collection

Expand table

Challenge / Requirement	Possible solutions	Considerations
Requires log filtering	Use Logstash Use Azure Functions Use LogicApps Use custom code (.NET, Python)	While filtering can lead to cost savings, and ingests only the required data, some Microsoft Sentinel features aren't supported, such as UEBA, entity pages, machine learning, and fusion. When configuring log filtering, make updates in resources such as threat hunting queries and analytics rules.
Agent cannot be installed	Use Windows Event Forwarding, supported with the Azure Monitor Agent	Using Windows Event forwarding lowers load-balancing events per second from the Windows Event Collector, from 10,000 events to 500-1000 events.
Servers do not connect to the internet	Use the Log Analytics gateway	Configuring a proxy to your agent requires extra firewall rules to allow the Gateway to work.

Challenge / Requirement	Possible solutions	Considerations
Requires tagging and enrichment at ingestion	Use Logstash to inject a ResourceID Use an ARM template to inject the ResourceID into on-premises machines Ingest the resource ID into separate workspaces	Log Analytics doesn't support role-based access control (RBAC) for custom tables. Microsoft Sentinel doesn’t support row-level RBAC. **Tip**: You might want to adopt cross workspace design and functionality for Microsoft Sentinel.
Requires splitting operation and security logs	Use the Microsoft Monitor Agent or Azure Monitor Agent multi-home functionality	Multi-home functionality requires more deployment overhead for the agent.
Requires custom logs	Collect files from specific folder paths Use API ingestion Use PowerShell Use Logstash	You might have issues filtering your logs. Custom methods aren't supported. Custom connectors might require developer skills.

On-premises Linux log collection

Expand table

Challenge / Requirement	Possible solutions	Considerations
Requires log filtering	Use Syslog-NG Use Rsyslog Use FluentD configuration for the agent Use the Azure Monitor Agent/Microsoft Monitoring Agent Use Logstash	Some Linux distributions might not be supported by the agent. Using Syslog or FluentD requires developer knowledge. For more information, see Connect to Windows servers to collect security events and Resources for creating Microsoft Sentinel custom connectors.

Challenge / Requirement	Possible solutions	Considerations
Agent cannot be installed	Use a Syslog forwarder, such as (syslog-ng or rsyslog.	
Servers do not connect to the internet	Use the Log Analytics gateway	Configuring a proxy to your agent requires extra firewall rules to allow the Gateway to work.
Requires tagging and enrichment at ingestion	Use Logstash for enrichment, or custom methods, such as API or Event Hubs.	You might have extra effort required for filtering.
Requires splitting operation and security logs	Use the Azure Monitor Agent with the multi-homing configuration.	
Requires custom logs	Create a custom collector using the Microsoft Monitoring (Log Analytics) agent.	

Endpoint solutions

If you need to collect logs from Endpoint solutions, such as EDR, other security events, Sysmon, and so on, use one of the following methods:

- **Microsoft Defender XDR connector** to collect logs from Microsoft Defender for Endpoint. This option incurs extra costs for the data ingestion.
- **Windows Event Forwarding**.

> ⓘ Note
>
> Load balancing cuts down on the events per second that can be processed to the workspace.

Office data

If you need to collect Microsoft Office data, outside of the standard connector data, use one of the following solutions:

Expand table

Challenge / Requirement	Possible solutions	Considerations
Collect raw data from Teams, message trace, phishing data, and so on	Use the built-in Office 365 connector functionality, and then create a custom connector for other raw data.	Mapping events to the corresponding recordID might be challenging.
Requires RBAC for splitting countries/regions, departments, and so on	Customize your data collection by adding tags to data and creating dedicated workspaces for each separation needed.	Custom data collection has extra ingestion costs.
Requires multiple tenants in a single workspace	Customize your data collection using Azure LightHouse and a unified incident view.	Custom data collection has extra ingestion costs. For more information, see Extend Microsoft Sentinel across workspaces and tenants.

Cloud platform data

⛶ Expand table

Challenge / Requirement	Possible solutions	Considerations
Filter logs from other platforms	Use Logstash Use the Azure Monitor Agent / Microsoft Monitoring (Log Analytics) agent	Custom collection has extra ingestion costs. You might have a challenge of collecting all Windows events vs only security events.
Agent cannot be used	Use Windows Event Forwarding	You might need to load balance efforts across your resources.
Servers are in air-gapped network	Use the Log Analytics gateway	Configuring a proxy to your agent requires firewall rules to allow the Gateway to work.
RBAC, tagging, and enrichment at ingestion	Create custom collection via Logstash or the Log Analytics API.	RBAC isn't supported for custom tables Row-level RBAC isn't supported for any tables.

Related content

For more information, see:

- Predeployment activities and prerequisites for deploying Microsoft Sentinel
- Best practices for Microsoft Sentinel
- Connect data sources

Feedback

Was this page helpful? 👍 Yes 👎 No

Provide product feedback | Get help at Microsoft Q&A