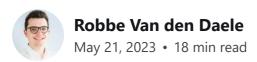


18 min read

Microsoft Sentinel

Demystifying Data Collection Rules and Transformations



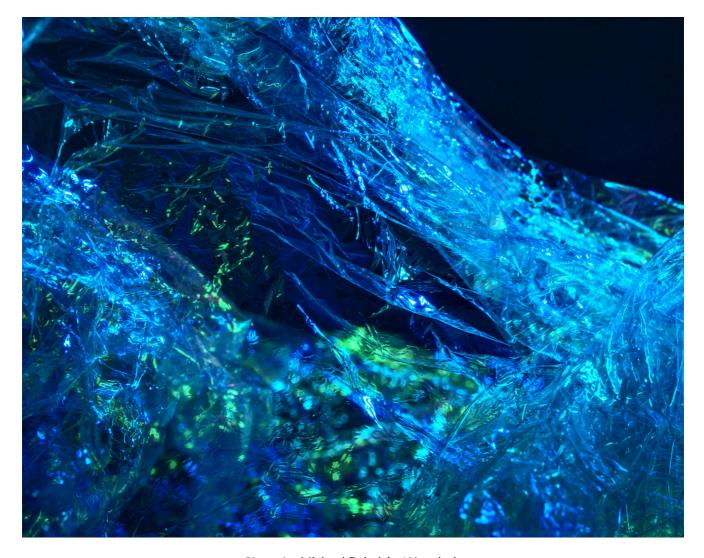


Photo by Michael Dziedzic / Unsplash

Introduction

Data Collection Rules theoretically



What are DCRs

DCR use cases regarding Microsoft Sentinel

Structures of Data Collection Rules

Data ingestion flow in Microsoft Sentinel

Custom data ingestion and transformations

How transformations work

Using multiple destinations

Costs for transformations

Creating transformations

Creating transformations for AMA sources

Creating transformations for standard tables

Why DCRs are so complicated

Logstash and DCRs

Golden tip

Introduction

Data Collection Rules in Azure can be very confusing once you really start using all of their potential. During my experience using them in complex setups, I identified a couple of pitfalls I would like to tackle for you. These pitfalls include the complex structures and types of DCRs, which type and structure of DCR you need for which use case, what the limitations and possibilities are of the different DCRs, and how you need to create the different DCRs.

In this post, you will learn what DCRs are, what structures and flavors there are, and how you can overcome the pitfalls mentioned earlier. All of this will be focussed on creating data collection rules for Microsoft Sentinel use cases.

Data Collection Rules theoretically

To clarify how data collection rules work, we will first have to go over some basics.

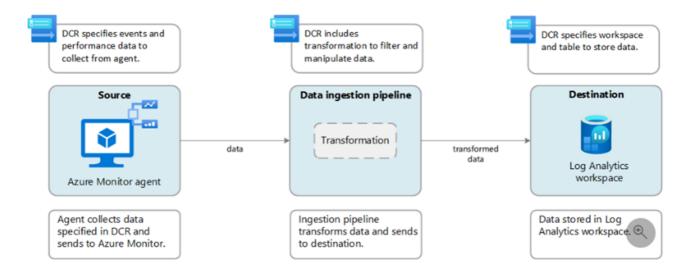
What are DCRs

Data Collection Rules are Azure resources that define the data collection process in Azure Monitor. It defines the details of a particular data collection scenario including which data should be collected, how to transform the data, and where to send that data. It is the new way of ingesting data into multiple types of destinations in Azure. This reflects in the fact that for example all Microsoft Sentinel data connectors that use the MMA agent or Log Analytics workspace ID and Key, are legacy in the Azure portal. When you try to create a data connector using the new way, you will find yourself creating Data Collection Rule in most cases.

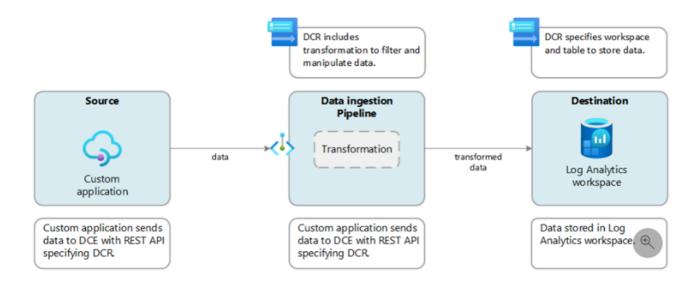
DCR use cases regarding Microsoft Sentinel

For Microsoft Sentinel, there are a couple of primary use cases for which you can use Data Collection Rules:

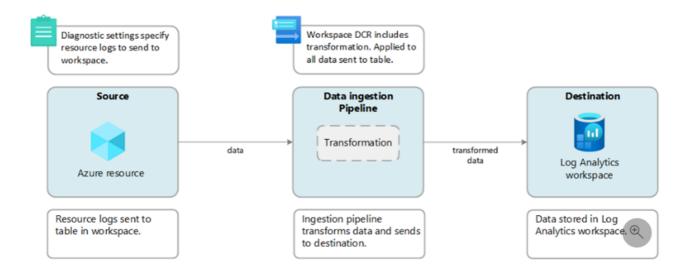
Azure Monitor Agent – You can create DCR rules with an
association to an Azure Monitor Agent, to receive the data from an
AMA agent and send it to a Log Analytics Workspace. This can be
done for Windows Events, Linux Syslog events, or third-party
syslog forwarding via a Syslog server



 Custom Logs – You can send data to the REST API that is connected to a Data Collection Endpoint which is linked to a Data Collection Rule.



 Workspace transformations – When you want to use transformations for a legacy workload that does not support DCR or Azure service connectors, you can associate a DCR to the workspace and use it for supported tables.



Regardless of the source that is used for the DCRs, they are **very powerful for filtering and normalization** of data. Since DCRs are available, you have a lot of power to choose

- Which data can be forwarded to Sentinel.
- How data must be normalized to fit in a certain table
- Under which complex conditions an event must be dropped (using KQL conditions instead of filtering at the source, which is generally very limited)

Structures of Data Collection Rules

There are two main structures for creating data collection rules, which are the 'Custom Logs' structure and the 'Azure Monitor Agent' structure. I found in practice that there is no hard limit between these structures, which means you can combine these structures in one ARM template (not recommended though). After you read the below, keep in mind that in essence, the big difference between the two structures is that Custom Logs uses streamDeclartions and need a Data Collection Endpoint and that Azure Monitoring Agent uses DataSources

instead of streamDeclarations and does not need a Data Collection Endpoint.

Custom logs

A Custom Logs DCR contains the following sections:

• StreamDeclarations – This section contains the different types of data that will be sent via the HTTP endpoint. Each stream is an object with a Key and a value. The Key is the stream name, which needs to start with 'Custom-'. The value is a list of top-level properties that are contained in the JSON data that will be sent.

 Destinations – The destinations to where the data needs to be sent (yes, multiple are possible)

- DataFlows This section ties the other sections together. It defines the following properties for each stream declared in streamDeclarations
- Streams Takes the streamsDeclarations
- Destination Takes the destinations
- TransformKql kql used to transform data from the streamDeclarations to the destinations
- OutputStream which table in the workspace the data will be sent to. The value needs to be 'Microsoft-[tableName]' when data is being ingested in standard Log Analytics tables, or 'Custom-[tableName]' when data is ingested in a custom-created table

Azure Monitor Agent

A DCR for Azure Monitor Agent contains the following sections

- DataSources Contains the unique source of monitoring data in its own format. Each data source has its own data source type, where each type defines its own unique set of properties that must be specified. The currently available data source types are:
- Extension VM extensions used by Log Analytics solutions and Azure services
- PerformanceCounters For both Windows and Linux
- Syslog Syslog events on Linux
- WindowsEventLogs Windows Events

More info about the data source types and their properties can be found in the sample ARM template at https://learn.microsoft.com/en-us/azure/azure-monitor/agents/data-collection-rule-sample-agent

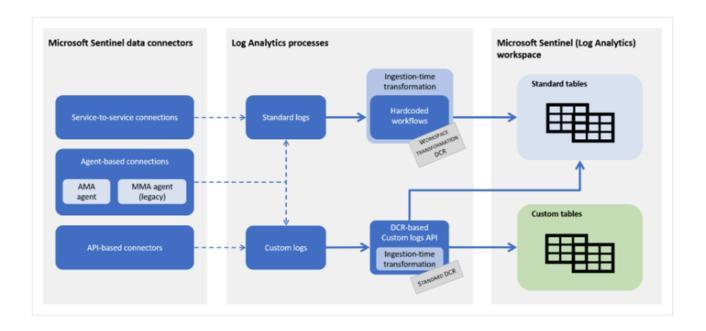
Destination – Destinations where data should be sent to

 DataFlows – Indicate which streams should be sent to which destinations (can have all of the properties like the DataFlows part described in the Custom Logs format)

• Streams - This unique handle describes a set of data sources that will be transformed and schematized as one type. Each data source requires one or more streams, and one stream can be used by multiple data sources. All data sources in a stream share a common schema. Use multiple streams, for example, when you want to send a particular data source to multiple tables in the same Log Analytics workspace. This is defined in both the DataSources and the DataFlows

Data ingestion flow in Microsoft Sentinel

Data ingestion flow in Microsoft Sentinel can be a little overwhelming when you are first learning about it. In the Microsoft documentation you will find a schema that can help you with understanding the different ingestion flows:



Below, we will go over the different flows based on the data connectors that are found at the left of the schema.

Service-to-service connections

When you would like to create transformations for service-to-service connections, your only option is to create a **workspace transformation DCR**. More info about how to create workspace transformation DCRs and what they are can be found later in the blog post.

Agent-based connections

Agent-based connections can use **both a workspace transformation DCR and a standard DCR**. When you use a workspace transformation, you are creating a DCR formed in the 'Azure Monitor Agent' format, without a DataSource configured in the DCR. By doing this, the DCR

transforms data for all data being ingestion in a specific table, without caring what the data source is (as long as it is not another DCR).

When you create a standard DCR, you can create transformation for specific data sources. The structure that is being used here is also the 'Azure Monitor Agent' structure of the DCR.

API-based connectors

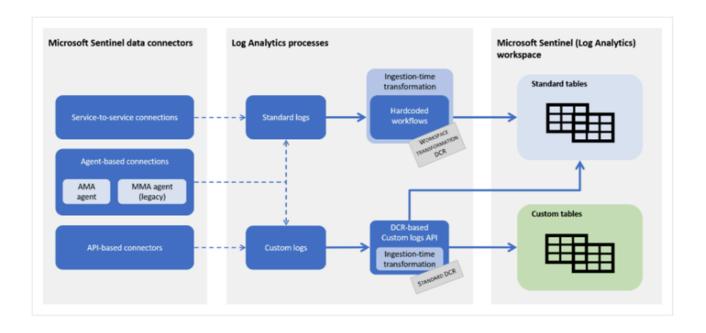
API-based connectors use **Standard DCRs** to send and transform data in tables. These DCRs are always created in the 'Custom logs' DCR format, which also means that there is a **Data Collection Endpoint needed for these DCRs**.

Custom data ingestion and transformations

Going forward, we will be focusing a little bit more on the custom data ingestion part of the Data Collection Rules.

How transformations work

We already mentioned that you can create transformation in DCRs to manipulate your data and that this must be done either by workspace transformation DCRs or by standard DCRs. Below you will find more info about both of them. Keep in mind the below schema while learning about them:



Workspace transformation DCR

When you create a workspace transformation DCR, you will find that the DCR is created in the structure of – what Microsoft calls – an **Azure Monitor Agent structure**. This DCR will have the DataSources,
Destinations, DataFlows, and Stream properties, but the DataSources field will remain empty.

It is important to note that a workspace can only have one workspace transformation DCR, but that DCR can contain separate transformations for each input stream. Below you will find a workspace transformation DCR created for the AADNonInteractiveUserSignInLogs table and the SecurityEvent table:

```
"kind": "WorkspaceTransforms",
"properties": {
   "dataSources": {},
    "destinations": {
        "logAnalytics": [
                "workspaceResourceId": "[parameters('workspaces_rvd_weu_sentinelbase_la_externalid')]",
                "name": "042c76ff99d44b20838e866acc1c8e1f"
    "dataFlows": [
            "streams": [
                "Microsoft-Table-AADNonInteractiveUserSignInLogs"
            "destinations": [
                "042c76ff99d44b20838e866acc1c8e1f"
            "transformKql": "source"
        },
            "streams": [
               "Microsoft-Table-SecurityEvent"
            "destinations": [
                "042c76ff99d44b20838e866acc1c8e1f"
            "transformKql": "source"
```

This type of DCR does not care what the source of the data is, as long as the data is not coming from another DCR. The supported tables that can be used for workspace transformation DCRs are currently limited and can be found here. **Workspace transformation DCR can only be used for transformations in standard tables**.

Standard DCR

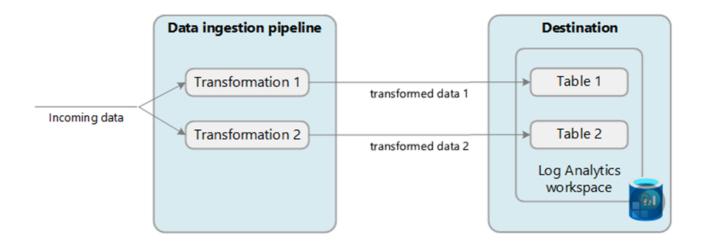
A standard DCR can be created in **both the Custom Logs and Azure Monitor Agent structure**. The important part here is that each connector or log source can have its own dedicated DCR, though multiple connectors or sources can share a common standard DCR as well. These DCRs are currently only supported for AMA agents and workflows using the Log Ingestion API but can **send data to both custom tables and a couple of standard tables**. The supported standard tables are currently

limited and can be found here, at the time of writing the supported standard tables are:

- CommonSecurityLog
- SecurityEvents
- Syslog
- WindowsEvents

Using multiple destinations

With transformation, you can send data to multiple destinations in a Log Analytics Workspace, where you can provide a KQL transformation for each destination. Currently, the tables in the DCR must reside in the same Log Analytics Workspace.



An example of a use case for sending data to multiple destinations is when you want to send debug and verbose logs to a custom table configured as basic logs, and other more import severities to an analytics table for interactive querying.

Costs for transformations

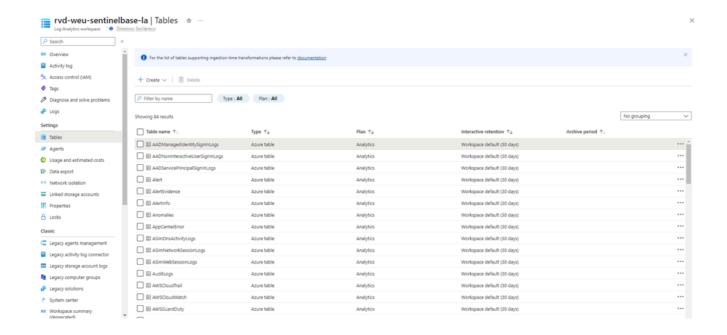
While transformations themselves don't incur direct costs, some scenarios can result in additional charges. These are described and explained in the Microsoft documentation: https://learn.microsoft.com/en-us/azure/azure-monitor/essentials/data-collection-transformations#cost-for-transformations

Creating transformations

You can create transformations and DCRs either by using the Azure portal or by deploying ARM templates via REST API or your favorite client libraries. In this section, we will be focussing on creating DCRs via the Azure portal, since I found this the easiest way to create a DCR (due to the complexity of the ARM structure of a DCR). Once we created them via the portal, we can then export the ARM templates for later use in DevOps scenarios.

Creating a workspace transformation DCR

If you still remember, there can only be one workspace transformation DCR. This DCR can be used for transformations from service-to-service connectors or the AMA agent, to standard and supported Log Analytics tables. To create a workspace transformation, go to your **Log Analytics workspace > Tables**. Here you will find all the tables present in your workspace.

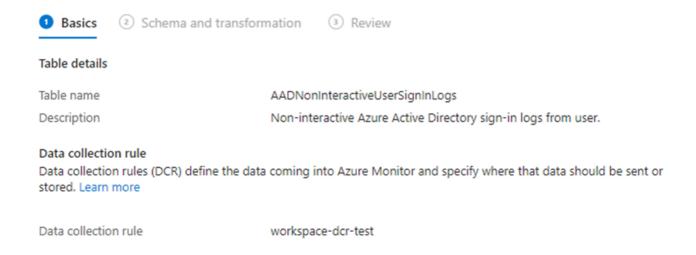


When you click on the three dots on the right side of the table, you will be able to create a transformation

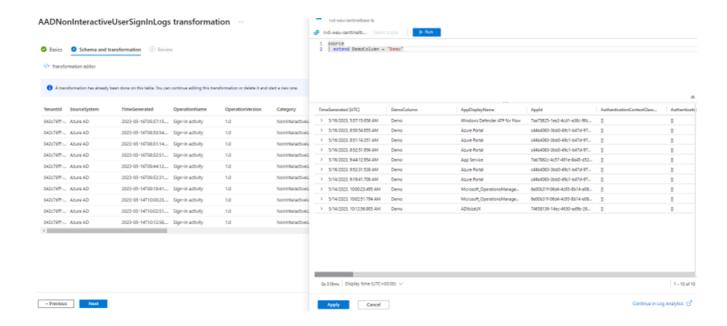


On the first page, you will be able to create a DCR. Note that if you already have a workspace transformation DCR, you will have to use that one (since only one workspace transformation DCR can exist)

AADNonInteractiveUserSignInLogs transformation

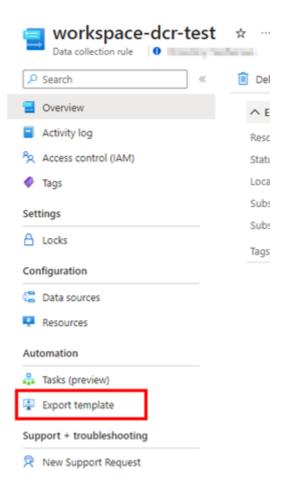


In the schema and transformation tab, you will be presented with the data that is currently present in the table. This is very handy to create your transformation query since you can interactively test the query with the current data in your table. Once you have developed the query, you can save the transformation and create the DCR.



When you go to the Data Collection Rule in the Azure Portal, you can view and export the ARM template. Here you will find your DCR with the

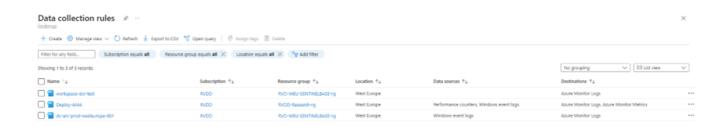
transformations for every table:



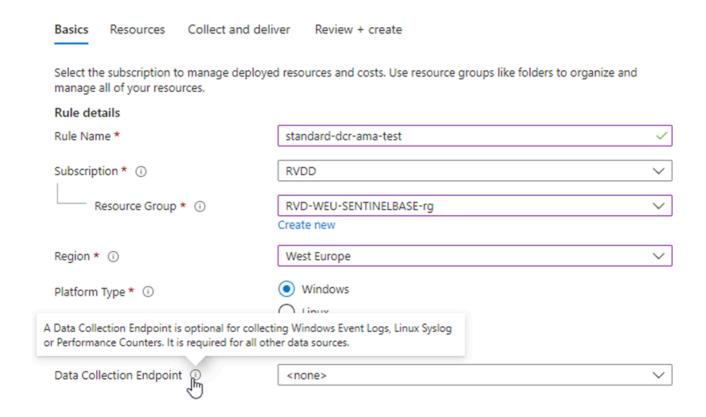
```
"dataFlows": [
        "streams": [
           "Microsoft-Table-SecurityEvent"
        "destinations": [
           "042c76ff99d44b20838e866acc1c8e1f"
        "transformKql": "source"
   },
        "streams": [
           "Microsoft-Table-AuditLogs"
        "destinations": [
           "042c76ff99d44b20838e866acc1c8e1f"
        "transformKql": "source\n| extend Id = \"1\"\n"
   },
       "streams": [
           "Microsoft-Table-AADNonInteractiveUserSignInLogs"
        "destinations": [
            "042c76ff99d44b20838e866acc1c8e1f"
        "transformKql": "source\n| extend DemoColumn_CF = \"Demo\"\n"
```

Creating a standard DCR for AMA option 1

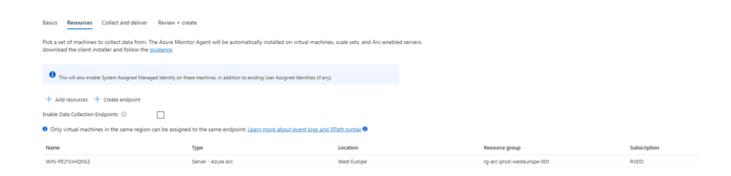
There are two experiences for creating standard DCRs for AMA sources (regarding Sentinel). The first one is by creating the DCR on the DCR page via the create button.



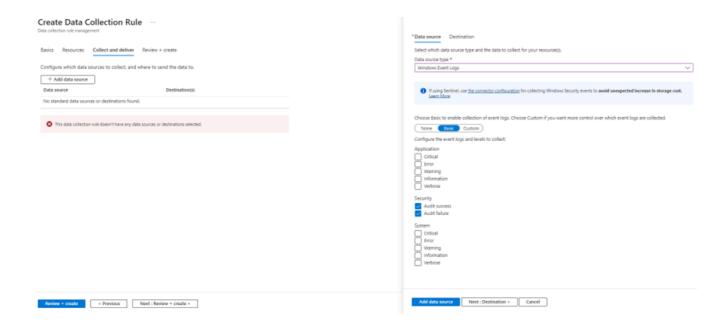
Here you start configuring the name, region, and resource group of the DCR. Depending on the OS of the device, you choose Linux or Windows. A Data Collection Endpoint is not needed for Windows Logs, Syslog logs, or Performance counters:



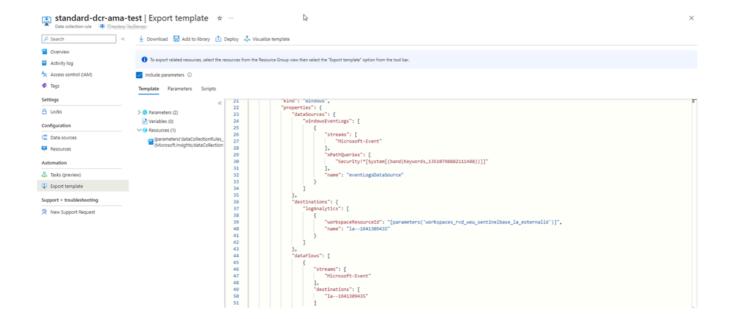
On the next page, you choose the resource from where you want to collect logs



In the Collect and Deliver tab, you can configure the types of logs you want to collect and where the logs need to be sent. Depending on whether you choose Linux or Windows as OS, this page will look different.



When you finish these configurations, you can create the DCR and check the ARM template in the Azure portal.

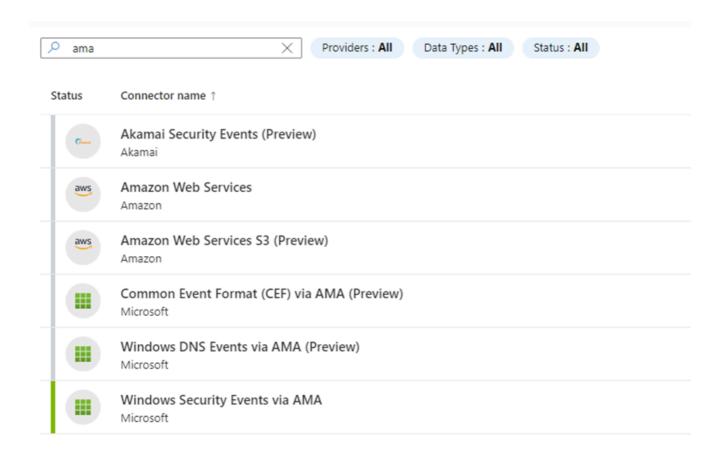


You might be wondering why we did not have the option to create a transformation in the portal while we were creating the DCR. And in fact, I do not know either. I suspect that this has something to do with the fact that the logs of an Azure Monitoring Agent are being sent to standard tables by default. However, this does not mean that transformations are

not supported for standard DCRs with AMA sources. Later in the post, we will be going over how we can create transformation anyway.

Creating a standard DCR for AMA option 2

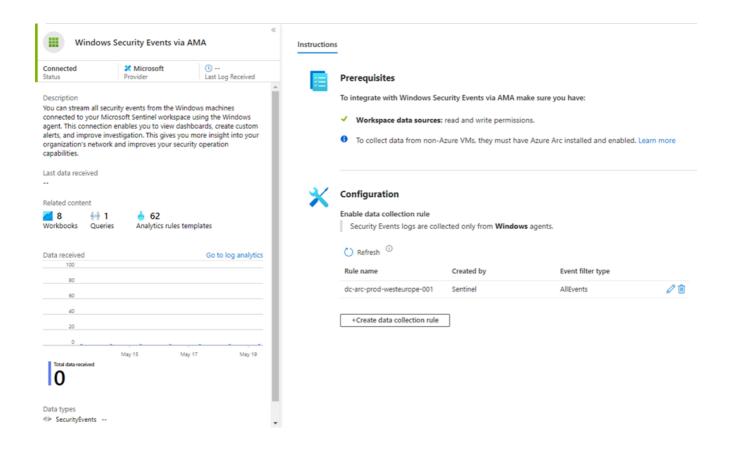
The second option for creating AMA-related DCRs is by going to the Microsoft Sentinel Data Connectors blade. When we search for AMA, we get three options:



Depending on what type of data you want to ingest, you select one of the three data connectors. The differences between the DCRs created by the different data connectors are very little since they are all **Standard DCRs** in the Azure Monitoring Agent format. What is different is the user experience in the portal, since you will have other configuration options for Windows Logs than you have for CEF forwarding for example.

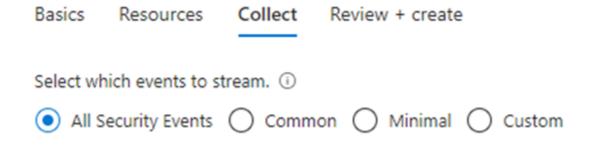
Windows Security Logs:

When you open the Windows Security Events via the AMA page, you will be able to create a DCR:



Here you will be able to create a DCR for Windows Servers just like you can in option 1, but you will find that you have other options in the 'Collect' tab to configure which events you want to ingest:

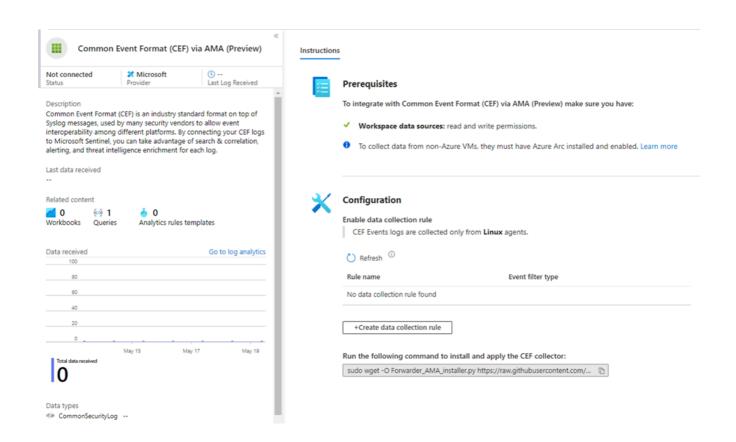




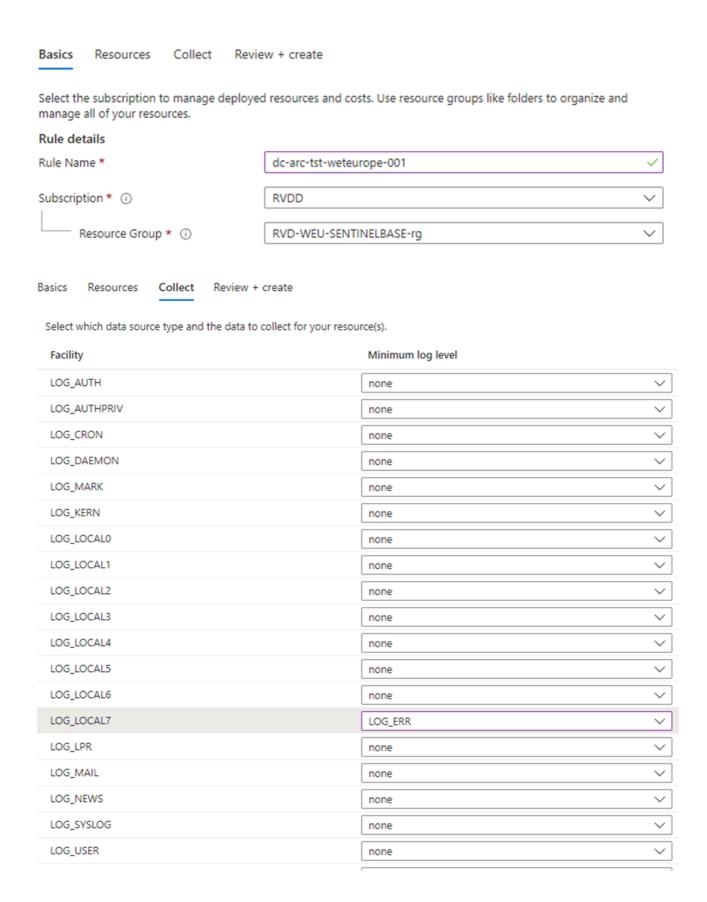
The reason behind this is that we are configuring a DCR for a specific use case, which is ingesting Security related events rather than all events from a specific severity or class. This does not mean the DCR will be in other formats, the portal just tries to help you create a DCR for your use case.

Common Event Format via AMA:

Another example is the CEF forwarding via AMA. In this data connector, you will also be able to create your DCR.



Here you will be able to create your DCR for CEF forwarding, where you will find that the Collect tab has other options again in comparison to creating Windows Security Events DCRs:



Again, this does not mean the DCR will be in other formats, the portal just tries to help you create a DCR for your use case.

Comparing the Windows and CEF DCRs

Now that we created DCRs for the different data connectors, we will compare the templates that were created by them.

When we check the DCR for **Windows Security Events**, we see that the data connector created a stream to ingest the data in the SecurityEvents table. This makes it easy for us since we do not have to search ourselves which table is best suiting for Windows Security Logs. We also see that the data connector created an XPath query based on the level of event we choose to ingest in the 'Collect' tab, which makes sure we do not have to decide which events we want to ingest:

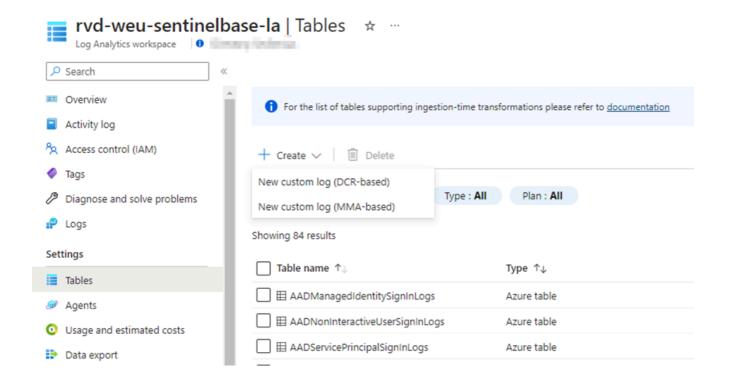
When we check the CEF DCR template, we see that the data connector now created a stream to send events to the CommonSecurityLog table, and changed the data sources to Syslog with the syslog facilities that we choose in the 'Collect' tab

```
"dataSources": {
    "syslog": [
            "streams": [
                 "Microsoft-CommonSecurityLog"
            ],
             "facilityNames": [
                "local7"
            "logLevels": [
                 "Error",
                 "Critical",
                 "Alert",
                 "Emergency"
             "name": "sysLogsDataSource-1688419672"
"destinations": {
    "logAnalytics": [
            "workspaceResourceId": "[parameters('workspaces rvd weu sentinelbase la externalid')]",
             "name": "DataCollectionEvent"
"dataFlows": [
         streams": [
             'Microsoft-CommonSecurityLog'
         destinations": [
             "DataCollectionEvent"
```

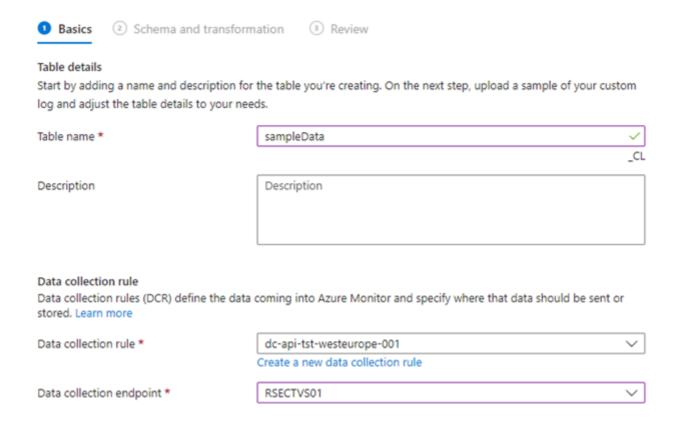
When comparing bot templates, we can conclude that the **settings and configurations** of the DCRs are **different** and specifically created for a different use case, but that **both DCRs** are still **standard DCRs in the Azure Monitoring Agent format**.

Creating standard DCR for Log Ingestion API sources

When you want to ingest data via the Log Ingestion API, you need to create a DCR with a Data Collection Endpoint. To start creating such a DCR, you will need to go to the tables of your log analytics workspace, and create a custom log DCR:



In the Basic tab, you need to choose a table name where your events will flow to. If you pay attention, you will see that this is a custom table that will be created. This is perfect if you want to ingest data in a custom table, but as mentioned earlier you can also send data via a standard DCR to a couple of build-in tables. In this view, this is not something you can choose to do. However, there is a **workaround that we will cover later in this post**. For now, choose a table name, DCR name, and a Data Collection Endpoint that you will use:



In the Schema and Transformation tab, you will see that you need to upload a sample file of your data. This is because the portal needs to know the format of the data so it can be included in the template. If you are awake, you should know by now that this type of DCR will be a **standard DCR in the Custom Logs format**, since this format requires you to define your columns:

```
"streamDeclarations": {
                                "Custom-testtable_CL": {
                                    "columns": [
                                           "name": "TimeGenerated",
                                           "type": "datetime"
                                        {
                                           "name": "Key1",
                                           "type": "string"
                                        {
                                           "name": "Key2",
                                           "type": "int"
                                           "name": "Key3",
                                           "type": "boolean"
                                           "name": "Key4",
                                           "type": "string"
```

Once the file is uploaded, you can create the transformation in the GUI just like we did for the workspace transformation DCR:



If we check the DCR template again, we now see that the template is in the Custom Logs format, and includes the components we created via the portal:

Creating transformations for AMA sources

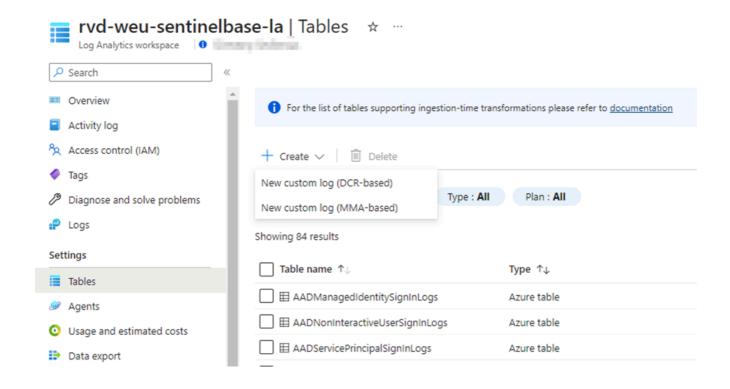
If you remember the section where we created DCRs for the AMA connectors, we were not able to create a transformation KQL in the portal. This does not mean transformations are not supported for these DCRs. You can just add a transformation KQL and an outputStream to the ARM template:

The only thing you need to make sure of is that the columns you use in the tranformation KQL are in fact columns that are known for the DCR source table. When this is not the case, the deployment of the ARM template will fail. The table where the columns need to come from is the table that is defined in the streams part.

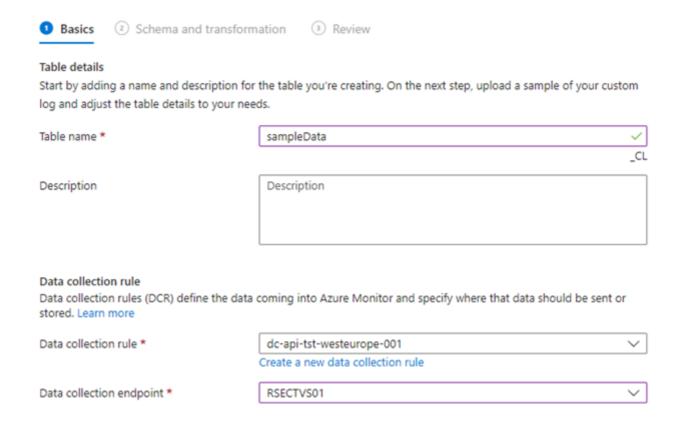
Creating transformations for standard tables

If you remember the section where we created a DCR for the Log Ingestion API, we mentioned that it was not possible in the GUI to choose a standard table, even though sending logs via a standard DCR support a couple of standard tables. To work around this issue, we can easily follow the below steps.

Create a DCR for Log Ingestion API as mentioned earlier:



Choose a random table name, we will delete the table later:



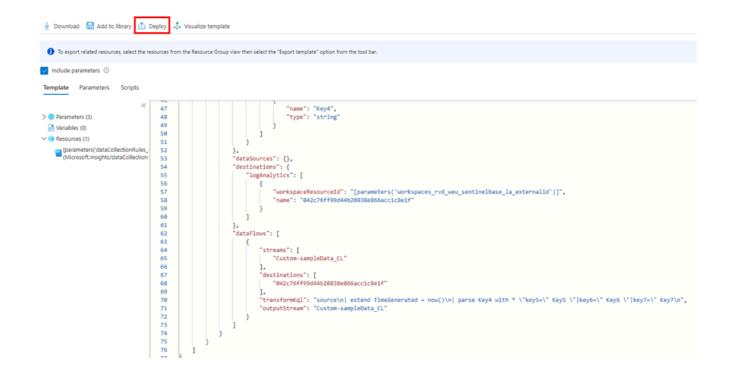
In the Schema and transformation part, you can upload your sample data again and create your transformation.



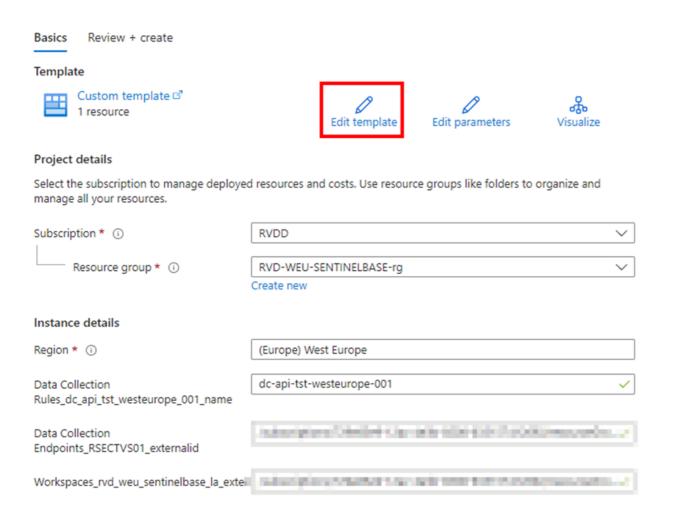
Once you have created your DCR and checked the template, you will see that the custom table is filled in as outputStream:



To change this to a supported standard table, click on Deploy:



Now edit the template:



And change the outputStream to a standard Microsoft table (remember that standard tables need to have Microsoft- in front of them).

Once you have saved the template, you can deploy the template again to the same existing DCR with the newly created table. Remember that you will need to make sure that the **columns that are created or changed after the tranformationKql need to represent the columns of the standard table you chose in the outputStream**. After you have deployed the ARM template, you can delete the custom table we created for developing the transformation query.

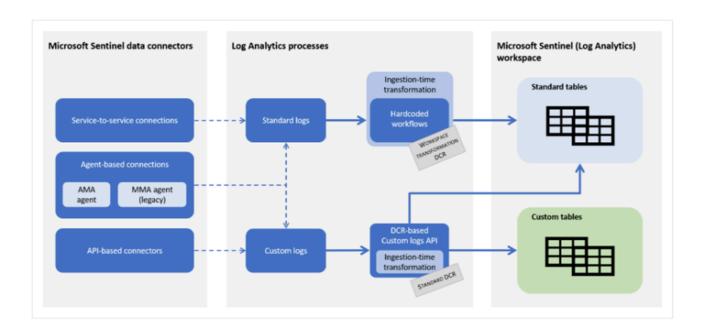
Why DCRs are so complicated

There are in my opinion a couple of problems with DCRs (at the time of writing this post) that make it so hard for people to understand how they work. First of all, the Microsoft documentation related to DCRs and transformation is all over the place. During the creation of this blog post, I had at a certain point 15 tabs open in my browser where I constantly had to switch between Microsoft documentation pages to gather all the info I needed.

Secondly, the two structures described by Microsoft of DCRs (being Custom Logs or Azure Monitoring Agent) are poorly documented and bring a lot of confusion. The only difference in essence is that Custom Logs

use streamDeclarations and needs a DCE, and that AMA uses data sources and doesn't need a DCE. What is also not clear in the documentation is that you can perfectly combine these two structures in one template (even though I wouldn't recommend it) and that outputStreams and transformationKql are perfectly useable in the AMA format (while it is only present in the Custom Logs format of the documentation).

The last problem I would like to describe is related to the below schema:



Don't get me wrong, the schema is good and quickly shows what DCRs are capable of. But one of the problems is that you can find on the schema that standard DCRs can send data to standard and custom tables. However, in some DCR creation flows in the portal you only have the option to choose a custom table (referring to the creating standard DCR for Log Ingestion API part), which brings again a lot of confusion. Another example is that you can create transformations for DCRs that use the Custom Logs format while this is not possible for DCRs that use the AMA format, even though you can perfectly add a transformationKql to an AMA formatted DCR and redeploy the template.

Even though DCRs are in my opinion complicated, I think they are very powerful and are a must-learn and understand for anyone who is regularly deploying advanced data connectors in Microsoft Sentinel.

Logstash and DCRs

In this blog post, we focussed on how DCRs are created, how they work, and what the pitfalls are. Even though we will not be talking about Logstash in this post, I would like to mention that Logstash is a very powerful alternative to transformations if you want to do data normalization and filtering before it is being sent to Sentinel. It even doesn't have to be an alternative to transformations, since you can now combine Logstash and transformations by using the new DCR-supported output plugin of Logstash. I highly recommend checking Logstash and the DCR support out, since it can be a very powerful combination if you want to create custom on-premise data connectors and still want to ingest them in standard log analytics tables. Logstash uses the Log Ingestion API, which means you can use standard DCRs in the custom logs format to manipulate your data.

If you want to learn more about using Logstash and DCRs, I recommend reading the post created by Koos Goossens! Once you understand this setup, make sure to harden your Logstash connections to Microsoft Sentinel.

Golden tip

I recently stumbled on a very cool workbook created by the community, which **helps you with creating and visualizing DCRs in your workspace**. An explanation of the workbook and how it works can be found here: https://techcommunity.microsoft.com/t5/microsoft-sentinel-

blog/create-edit-and-monitor-data-collection-rules-with-the-data/ba-p/3810987.

One of the most powerful tools in this workbook is the feature where you can create DCRs for Windows servers based on NSA and MITRE categories:



Which is in my opinion far more powerful than ingesting Event IDs via the 'all, common, minimal' filter. Or having to plot Event IDs to the MITRE framework manually. Like I did in my previous post.







Transition from

Detecting non-

MDE Device

Microsoft Sentinel to Defender XDR -Practical challenges

Introduction
Microsoft
announced on the...

Jul 4, 2025 12 min read

privileged Windows Hello abuse

Introduction I recently followed a live session of Dirk...

Apr 26, 16 min 2025 read

Discovery -Improving the monitored network page

Introduction This blogpost is probably the first ...

Mar 19, 6 min 2025 read

Hybrid Brothers © 2025

Sign up Privacy policy

Powered by Ghost