

# Monitor Azure Data Factory pipelines

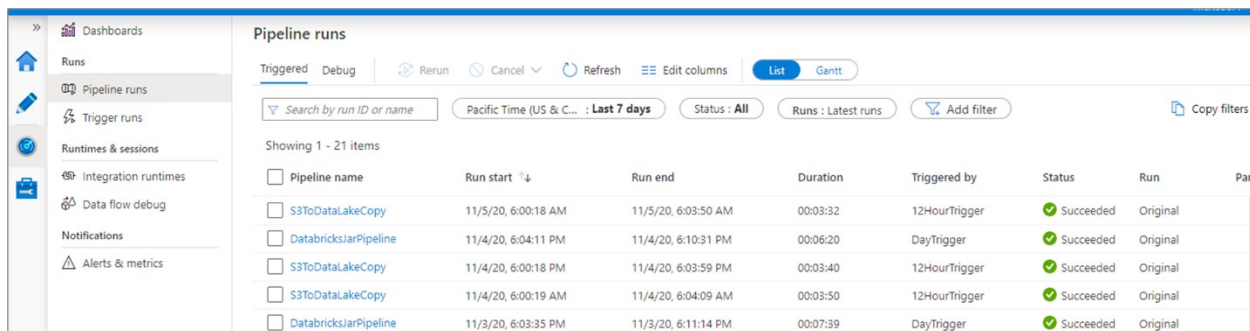
## Note:

You are not required to complete the processes, tasks, activities, or steps presented in this example. The various samples provided are for illustrative purposes only and it's likely that if you try this out you will encounter issues in your system.

Once you've created and published a pipeline in Azure Data Factory, you can associate it with a trigger or manually kick off an on-demand run. You can monitor all of your pipeline runs natively in the Azure Data Factory user experience. To open the monitoring experience, select the **Monitor & Manage** tile in the data factory blade of the Azure portal. If you're already in the Azure Data Factory UX, click on the **Monitor** icon on the left sidebar.

## Monitor pipeline runs

The default monitoring view is a list of triggered pipeline runs in the selected time period. You can change the time range and filter by status, pipeline name, or annotation. Hover over the specific pipeline run to get run-specific actions such as rerun and the consumption report.



Pipeline name	Run start	Run end	Duration	Triggered by	Status	Run
S3ToDataLakeCopy	11/5/20, 6:00:18 AM	11/5/20, 6:03:50 AM	00:03:32	12HourTrigger	Succeeded	Original
DatabricksJarPipeline	11/4/20, 6:04:11 PM	11/4/20, 6:10:31 PM	00:06:20	DayTrigger	Succeeded	Original
S3ToDataLakeCopy	11/4/20, 6:00:18 PM	11/4/20, 6:03:59 PM	00:03:40	12HourTrigger	Succeeded	Original
S3ToDataLakeCopy	11/4/20, 6:00:19 AM	11/4/20, 6:04:09 AM	00:03:50	12HourTrigger	Succeeded	Original
DatabricksJarPipeline	11/3/20, 6:03:35 PM	11/3/20, 6:11:14 PM	00:07:39	DayTrigger	Succeeded	Original

List view for monitoring pipeline runs

The pipeline run grid contains the following columns:

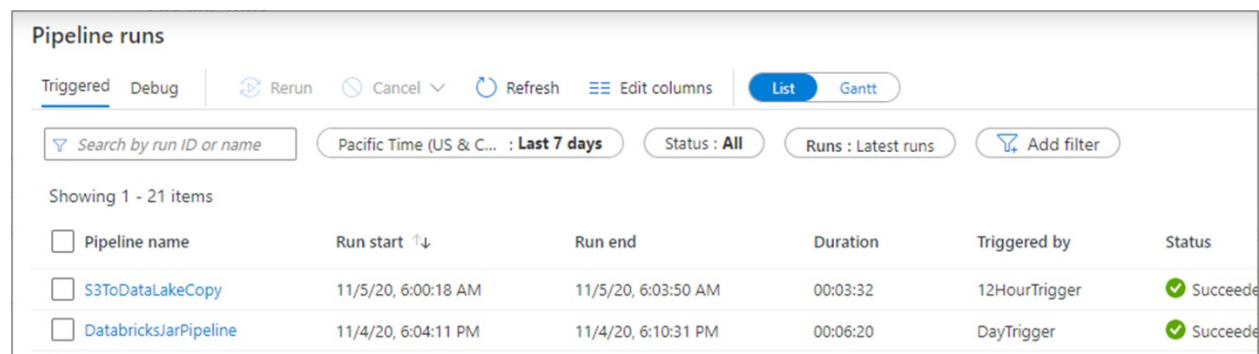
### Column name Description

Pipeline Name	Name of the pipeline
Run Start	Start date and time for the pipeline run (MM/DD/YYYY, HH:MM:SS AM/PM)
Run End	End date and time for the pipeline run (MM/DD/YYYY, HH:MM:SS AM/PM)
Duration	Run duration (HH:MM:SS)
Triggered By	The name of the trigger that started the pipeline
Status	Failed, Succeeded, In Progress, Canceled, or Queued
Annotations	Filterable tags associated with a pipeline
Parameters	Parameters for the pipeline run (name/value pairs)
Error	If the pipeline failed, the run error

## Column name Description

Run ID ID of the pipeline run

You need to manually select the **Refresh** button to refresh the list of pipeline and activity runs. Autorefresh is currently not supported.

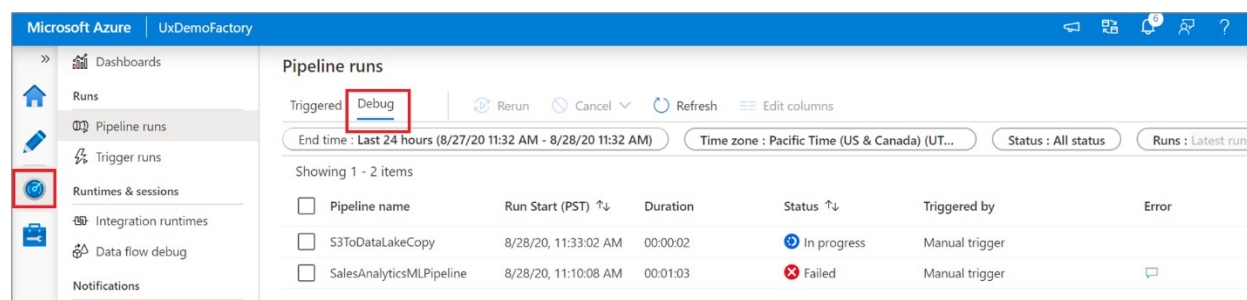


The screenshot shows the 'Pipeline runs' interface. At the top, there are tabs for 'Triggered' and 'Debug'. Below the tabs are buttons for 'Rerun', 'Cancel', 'Refresh', and 'Edit columns'. There are also buttons for 'List' and 'Gantt'. A search bar is present with the text 'Search by run ID or name'. Below the search bar, there are filters for 'Pacific Time (US & Canada) : Last 7 days', 'Status : All', and 'Runs : Latest runs'. The table shows 1 - 21 items. The columns are: Pipeline name, Run start, Run end, Duration, Triggered by, and Status. The table lists two runs: 'S3ToDataLakeCopy' and 'DatabricksJarPipeline'.

Pipeline name	Run start	Run end	Duration	Triggered by	Status
S3ToDataLakeCopy	11/5/20, 6:00:18 AM	11/5/20, 6:03:50 AM	00:03:32	12HourTrigger	Succeeded
DatabricksJarPipeline	11/4/20, 6:04:11 PM	11/4/20, 6:10:31 PM	00:06:20	DayTrigger	Succeeded

Refresh button

To view the results of a debug run, select the **Debug** tab.



The screenshot shows the 'Pipeline runs' interface with the 'Debug' tab selected. The left sidebar shows a navigation menu with 'Runs' selected. The main area shows the 'Debug' tab with buttons for 'Rerun', 'Cancel', 'Refresh', and 'Edit columns'. There are filters for 'End time : Last 24 hours (8/27/20 11:32 AM - 8/28/20 11:32 AM)', 'Time zone : Pacific Time (US & Canada) (UT...)', 'Status : All status', and 'Runs : Latest runs'. The table shows 1 - 2 items. The columns are: Pipeline name, Run Start (PST), Duration, Status, Triggered by, and Error. The table lists two runs: 'S3ToDataLakeCopy' and 'SalesAnalyticsMLPipeline'.

Pipeline name	Run Start (PST)	Duration	Status	Triggered by	Error
S3ToDataLakeCopy	8/28/20, 11:33:02 AM	00:00:02	In progress	Manual trigger	
SalesAnalyticsMLPipeline	8/28/20, 11:10:08 AM	00:01:03	Failed	Manual trigger	

Select the View active debug runs icon

## Monitor activity runs

To get a detailed view of the individual activity runs of a specific pipeline run, click on the pipeline name.

### Pipeline runs

Triggered
Debug
Rerun
Cancel
Refresh
Edit col

Pacific Time (US & C... : Last 7 days

Showing 1 - 21 items

<input type="checkbox"/> Pipeline name	Run start ↑↓	Run end
<input type="checkbox"/> S3ToDataLakeCopy	11/5/20, 6:00:18 AM	11/5/20, 6:03
<input type="checkbox"/> DatabricksJarPipeline	11/4/20, 6:04:11 PM	11/4/20, 6:10
<input type="checkbox"/> S3ToDataLakeCopy	11/4/20, 6:00:18 PM	11/4/20, 6:03
<input type="checkbox"/> S3ToDataLakeCopy	11/4/20, 6:00:19 AM	11/4/20, 6:04

View activity runs

The list view shows activity runs that correspond to each pipeline run. Hover over the specific activity run to get run-specific information such as the JSON input, JSON output, and detailed activity-specific monitoring experiences. Activities such as data flow, copy, and Azure Databricks have a dedicated view that can be seen by clicking on the eyeglasses icon.

Dashboards

Runs

Pipeline runs

Trigger runs

Runtimes & sessions

Integration runtimes

Data flow debug

Notifications

Alerts & metrics

All pipeline runs > SalesAnalyticsMLPipeline - Activity runs

SalesAnalyticsMLPipeline

List Gantt

Rerun Rerun from activity Rerun from failed activity Refresh

Copy data Location\_HTTP

Copy data Customer\_Salesforce

Copy data Products\_SAP

Wrangling Data Flow (Preview) SalesDataPrep

Data flow SalesAnalytics

Activity runs

Pipeline run ID a600eabe-19fb-4d0b-bd8d-d20b21223923

All status

Showing 1 - 5 of 5 items

Activity name	Activity type	Run start ↑↓	Duration	Status
Location_HTTP	Copy	11/5/20, 12:12:44 PM	00:00:15	✓ Succeeded
Clickstream_S3	Copy	11/5/20, 12:12:44 PM	00:00:27	✗ Failed
Customer_Salesforce	Copy	11/5/20, 12:12:44 PM	00:00:10	✓ Succeeded
POS_SQL	Copy	11/5/20, 12:12:44 PM	00:00:36	✗ Failed
Products_SAP	Copy	11/5/20, 12:12:44 PM	00:00:08	✓ Succeeded

View the activity runs in the monitoring tool

## Column name

## Description

Activity Name	Name of the activity inside the pipeline
Activity Type	Type of the activity, such as Copy, ExecuteDataFlow, or AzureMLExecutePipeline
Actions	Icons that allow you to see JSON input information, JSON output information, or detailed activity specific monitoring experiences
Run Start	Start date and time for the activity run (MM/DD/YYYY, HH:MM:SS AM/PM)
Duration	Run duration (HH:MM:SS)
Status	Failed, Succeeded, In Progress, or Canceled
Integration Runtime	Which Integration Runtime the activity was run on
User Properties	User-defined properties of the activity

**Column name****Description**

Error If the activity failed, the run error

Run ID ID of the activity run

If an activity failed, you can see the detailed error message by clicking on the icon in the error column.

The screenshot shows the 'Activity runs' section of the Azure Data Factory portal. On the left, there's a list of activities: 'Copy data' (Customer\_Salesforce) and 'Copy data' (Products\_SAP). Below this is a table of activity runs. The first row is selected, showing 'Pipeline run ID a600eabe-19fb'. The table has columns for 'All status', 'Showing 1 - 5 of 5 items', 'Integration runtime', and 'DefaultIntegrationRuntime (East US)'. An 'Error details' modal is open, displaying the following information:

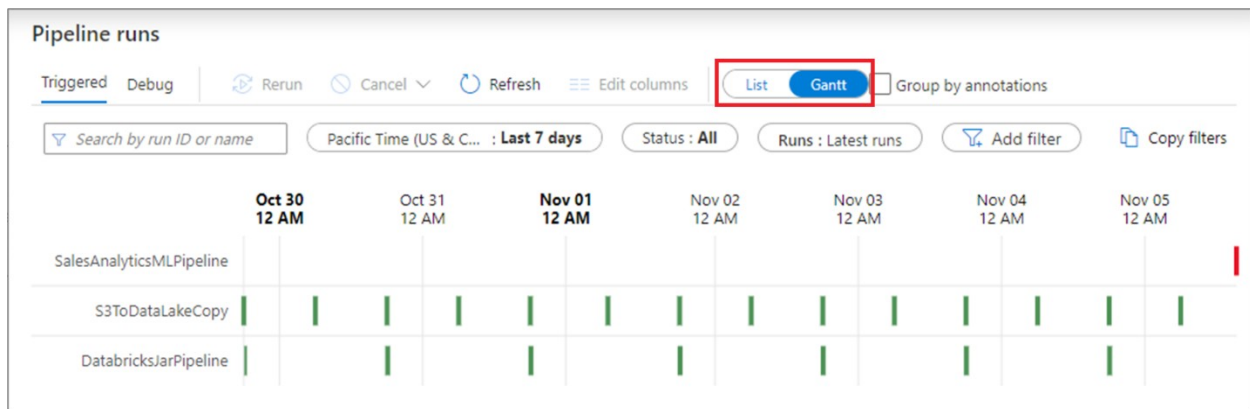
- Error code:** 2200 [Troubleshooting guide](#)
- Failure type:** User configuration issue
- Details:** 'Type=Amazon.S3.AmazonS3Exception,Message=The specified bucket does not exist,Source=AWSSDK.Core,'Type=Amazon.Runtime.Internal.HttpResponseException,Message=The remote server returned an error: (404) Not Found.,Source=AWSSDK.Core,'Type=System.Net.WebException,Message=The remote server returned an error: (404) Not Found.,Source=System,'

At the bottom of the modal, there's a feedback section: 'How helpful or unhelpful was this error message?' with five stars. The bottom of the screenshot shows the 'DefaultIntegrationRuntime (East US)' and a red box highlighting a speech bubble icon next to the run ID 'ce92dc2d-cdeb-477e'.

Viewing activity run errors

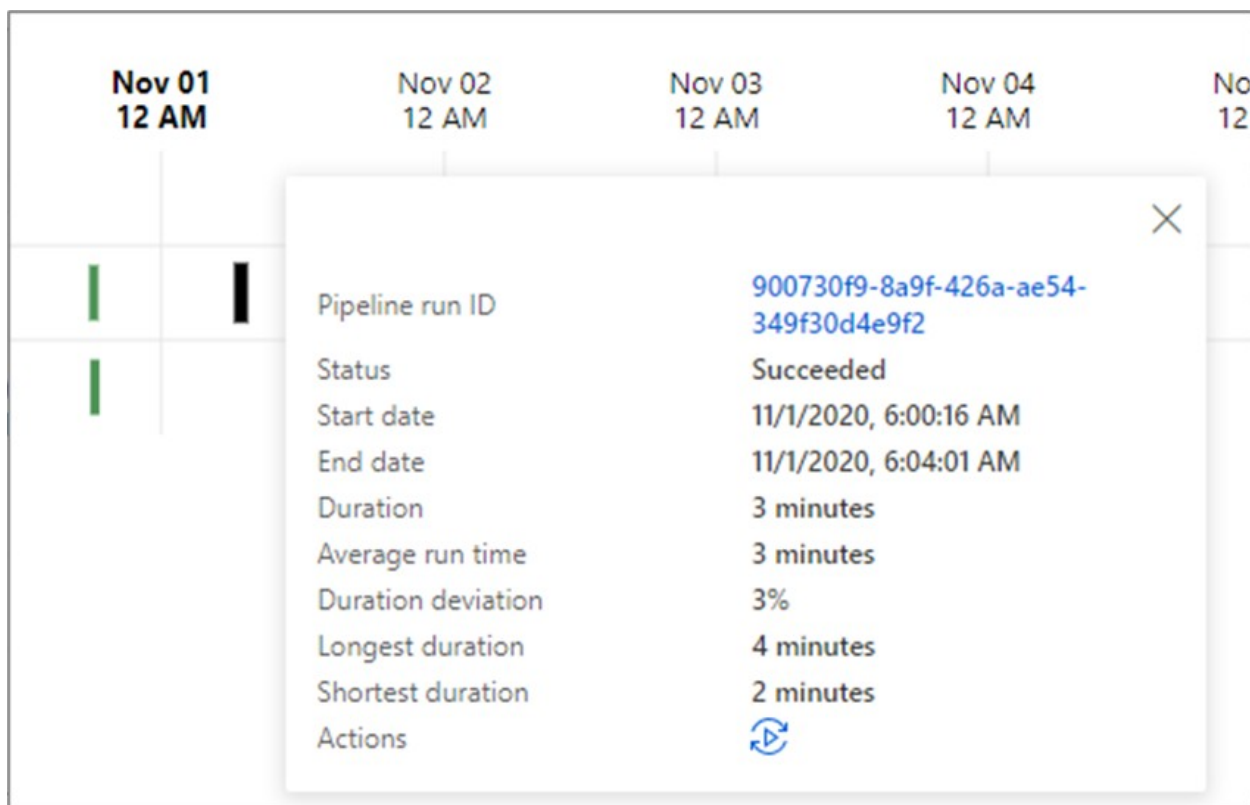
## Gantt view

A Gantt chart is a view that allows you to see the run history over a time range. By switching to a Gantt view, you will see all pipeline runs grouped by name displayed as bars relative to how long the run took. You can also group by annotations/tags that you've create on your pipeline. The Gantt view is also available at the activity run level.



Example of a Gantt chart

The length of the bar informs the duration of the pipeline. You can also select the bar to see more details.

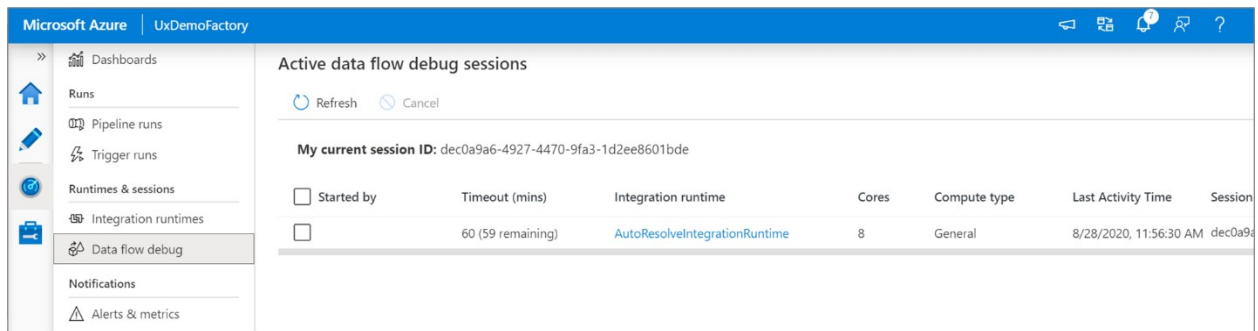


Gantt chart duration

## Monitor data flow debug sessions

Mapping data flows allow you to build code-free data transformation logic that runs at scale. When building your logic, you can turn on a debug session to interactively work with your data using a live Spark cluster.

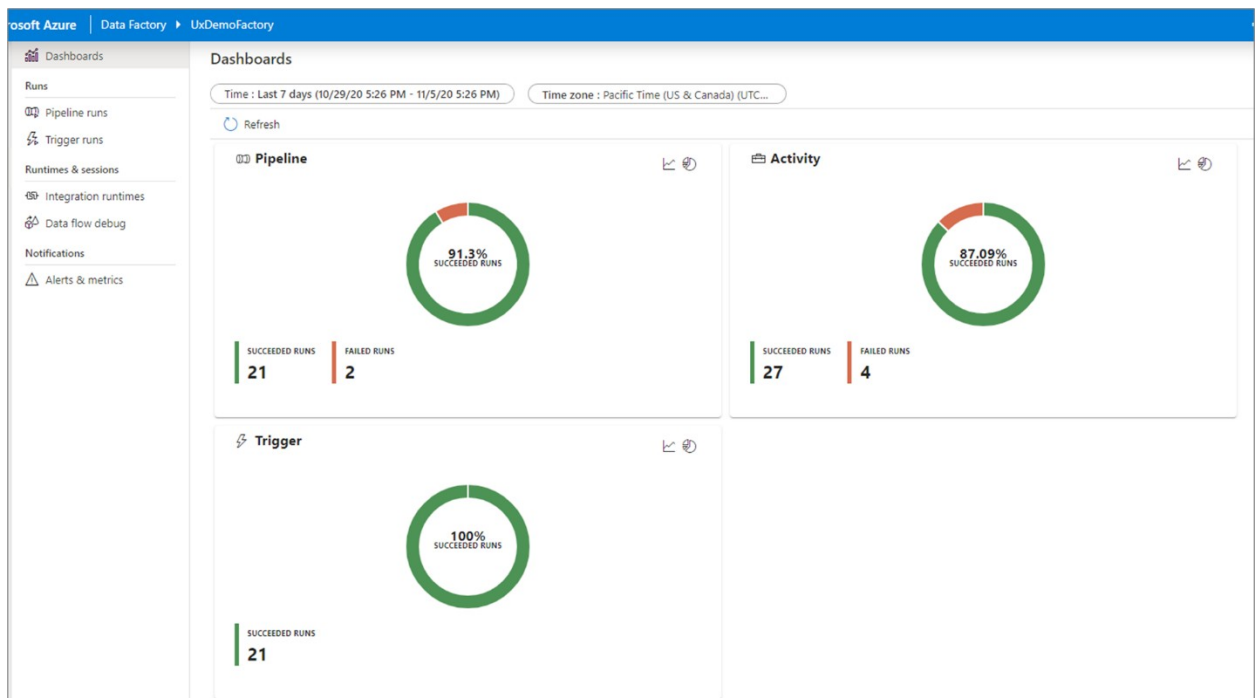
You can monitor active data flow debug sessions across a factory in the **Monitor** experience.



View data flow debug session

## Monitoring dashboards

To view a high-level visual display of your pipeline, activity, and trigger runs, go to the **Dashboards** pane. Here you can see the number of successes and failures over a specified time period.



reading the monitor dashboard

## Monitor using Azure Monitor

Azure Monitor provides base-level infrastructure metrics and logs for most Azure services. Azure diagnostic logs are emitted by a resource and provide rich,



frequent data about the operation of that resource. Azure Data Factory (ADF) can write diagnostic logs in Azure Monitor.

Data Factory stores pipeline-run data for only 45 days. Use Azure Monitor if you want to keep that data for a longer time. With Monitor, you can route diagnostic logs for analysis to multiple different targets.

- **Storage Account:** Save your diagnostic logs to a storage account for auditing or manual inspection. You can use the diagnostic settings to specify the retention time in days.
- **Event Hub:** Stream the logs to Azure Event Hubs. The logs become input to a partner service/custom analytics solution like Power BI.
- **Log Analytics:** Analyze the logs with Log Analytics. The Data Factory integration with Azure Monitor is useful in the following scenarios:

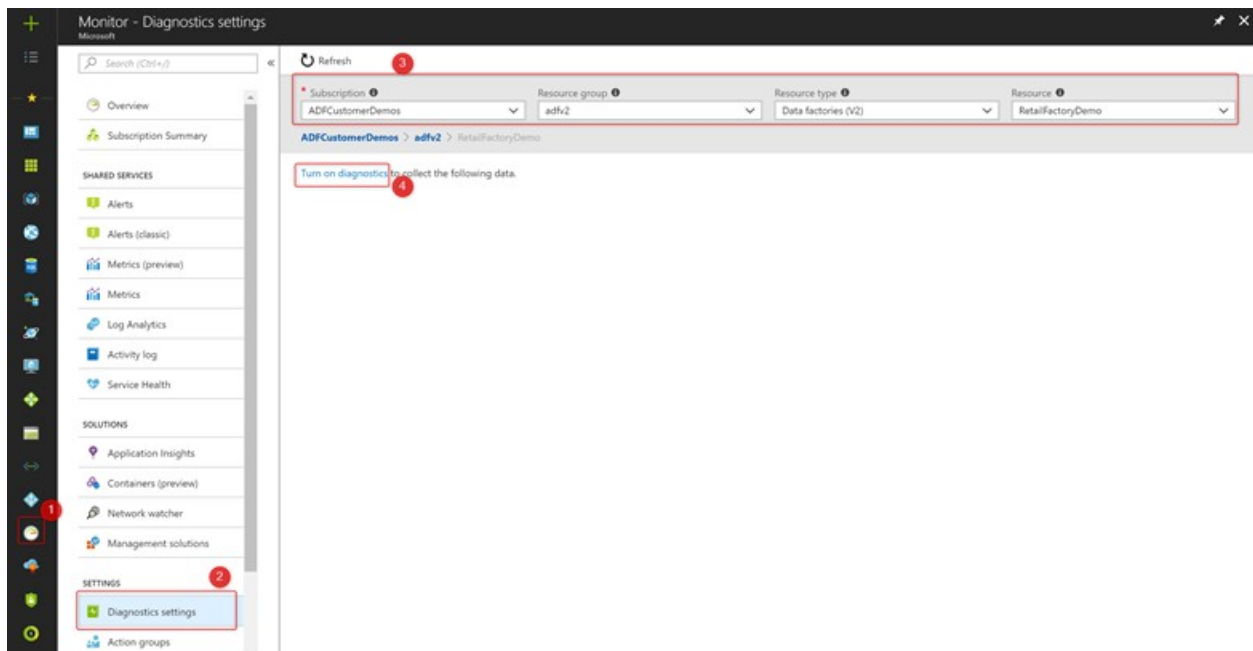
You can also use a storage account or event-hub namespace that isn't in the subscription of the resource that emits logs. The user who configures the setting must have appropriate Azure role-based access control (Azure RBAC) access to both subscriptions.

## Configure diagnostic settings and workspace

Create or add diagnostic settings for your data factory.

1. In the portal, go to Monitor. Select **Settings > Diagnostic settings**.
2. Select the data factory for which you want to set a diagnostic setting.
3. If no settings exist on the selected data factory, you're prompted to create a setting. Select **Turn on diagnostics**.





Create a diagnostic setting if no settings exist

**If there are existing settings on the data factory, you see a list of settings already configured on the data factory. Select **Add diagnostic setting**.**

Subscription \* ⓘ Resource group ⓘ Resource type ⓘ Resource ⓘ

>  >

Diagnostic settings are used to configure streaming export of platform logs and metrics for a resource to the destination of your choice. You may create up to five different diagnostic settings to send different logs and metrics to independent destinations. [Learn more about diagnostics settings](#)

Diagnostics settings

Name	Storage account	Event hub	Log Analytics workspace	Edit setting
ADFv2Diagnostics	-	-	loganalyticsworkspace	<a href="#">Edit setting</a>

[+ Add diagnostic setting](#)

Click 'Add Diagnostic setting' above to configure the collection of the following data:

- ActivityRuns
- PipelineRuns
- TriggerRuns
- SSISPackageEventMessages
- SSISPackageExecutableStatistics
- SSISPackageEventMessageContext
- SSISPackageExecutionComponentPhases
- SSISPackageExecutionDataStatistics
- SSISIntegrationRuntimeLogs
- AllMetrics

Add a diagnostic setting if settings exist

1.

## Diagnostics settings

 Save  Discard  Delete  Provide feedback

A diagnostic setting specifies a list of categories of platform logs and/or metrics that you want to collect from a resource, and one or more destinations that you would stream them to. Normal usage charges for the destination will occur. [Learn more about the different log categories and contents of those logs](#)

Diagnostics settings name

Diagnostics settings name \*  

### Category details

#### log

☒ ActivityRuns

☒ PipelineRuns

☒ TriggerRuns

☒ SSISPackageEventMessages

☒ SSISPackageExecutableStatistics

☒ SSISPackageEventMessageContext

☒ SSISPackageExecutionComponentPhases

☒ SSISPackageExecutionDataStatistics

☒ SSISIntegrationRuntimeLogs

#### metric

☒ AllMetrics


### Destination details

☒ Send to Log Analytics

#### Subscription



#### Log Analytics workspace



#### Destination table ⓘ

☐ Azure diagnostics ☒ Resource specific

☐ Archive to a storage account

☐ Stream to an event hub

Name your settings and select a log-analytics workspace