Data factory

Managed service in the cloud for scale-out serverless data integration, data transformation and orchestration

Create setup

vinoworld-dev-rg
Resource group

How do I troubleshoot issues with this resource group?   What are the best practices for managing this

+ Create   ⚙ Manage view ∨   🗑 Delete resource group   ↻ Refresh   ↓ Export to CSV   ⋈ Open query

Search

- Overview
- Activity log
- Access control (IAM)
- Tags
- Resource visualizer
- Events
- Settings
  - Deployments
  - Security

∨ Essentials

Resources   Recommendations

Filter for any field...   Type equals **all** ✕   Location equals **all** ✕   ＋ Add filter

Showing 1 to 1 of 1 records.   ☐ Show hidden types ⓘ

☐ Name ↑↓	Type ↑↓
☐ ▦ adfsvinoworld	Storage account

---

vinoworld-dev-adf01
Data factory (V2)

How do I troubleshoot pipeline failures in this Data factory (V2)?   List triggers associated with this Data factory (V2).   +1

Search

- Overview
- Activity log
- Access control (IAM)
- Tags
- Diagnose and solve problems
- Resource visualizer
- Settings
- Getting started
- Monitoring
- Automation
- Help

🗑 Delete

∧ Essentials

Resource group (move)	: vinoworld-dev-rg	Type : Data factory (V2)
Status	: Succeeded	Getting started : Quick start
Location	: Germany West Central	
Subscription (move)	: Azure subscription 1	
Subscription ID	: b435e57d-913c-4896-baf9-12e1971e7539	

Azure Data Factory Studio

Launch studio

SERVERS

- vinoworld-dev-sql1.database.wind...
  - Tables
    - dbo.test
      - Columns
        - id (int, null)
        - description (varchar(10), null)
      - Keys
      - Constraints
      - Triggers
      - Indexes
      - Statistics
    - Dropped Ledger Tables
  - Views
  - Synonyms
  - Programmability

▷ Run ☐ Cancel ⚡ Disconnect ⟳ Change | Database: vinoworld-dev-sqldb ⌄ | ⛁ Estimated

```sql
1   create table dbo.test(id  int, description varchar(10))
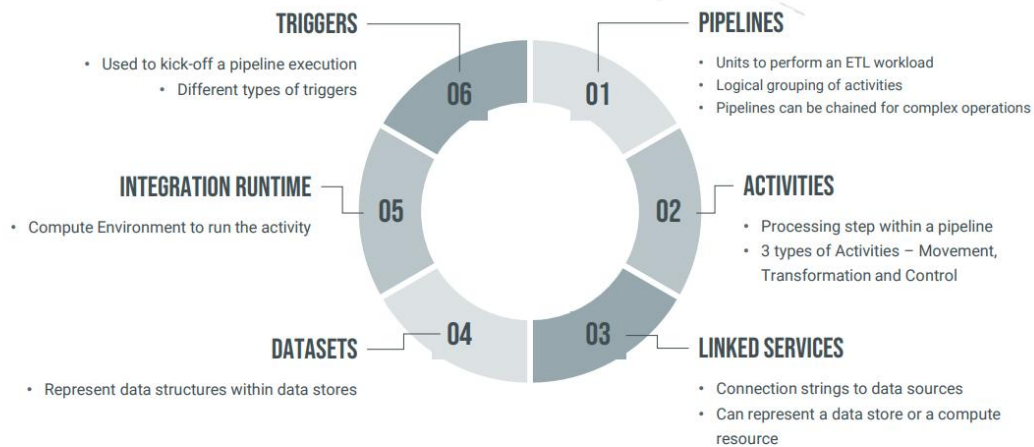```

**Messages**

1:08:55 PM    Started executing query at Line 1
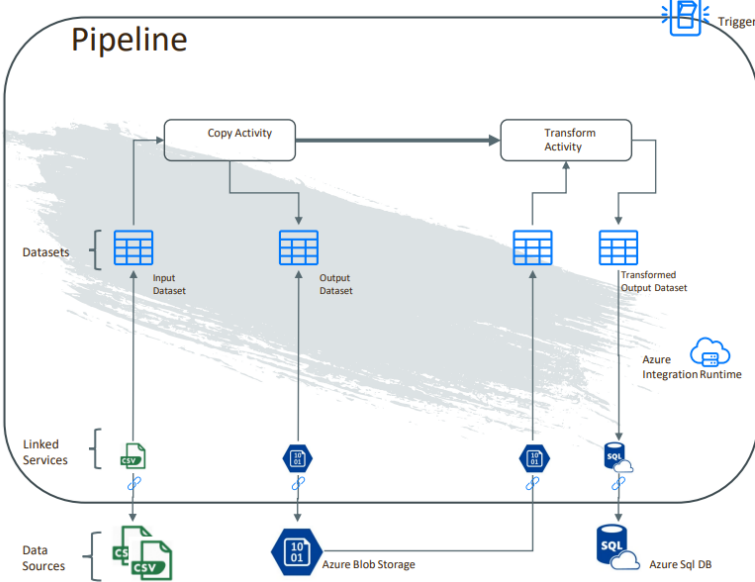              Commands completed successfully.
              Total execution time: 00:00:00.166

*What are the main components of Azure Data Factory?*



**TRIGGERS**
- Used to kick-off a pipeline execution
- Different types of triggers

**06**

**PIPELINES**
- Units to perform an ETL workload
- Logical grouping of activities
- Pipelines can be chained for complex operations

**01**

**INTEGRATION RUNTIME**
- Compute Environment to run the activity

**05**

**ACTIVITIES**
- Processing step within a pipeline
- 3 types of Activities – Movement, Transformation and Control

**02**

**DATASETS**
- Represent data structures within data stores

**04**

**03**

**LINKED SERVICES**
- Connection strings to data sources
- Can represent a data store or a compute resource

# SO HOW ARE THESE COMPONENTS RELATED IN AZURE DATA FACTORY?

*Let's look under the hood*

**Pipeline**

Copy Activity → Transform Activity

Trigger

Datasets
- Input Dataset
- Output Dataset
- Transformed Output Dataset

Azure Integration Runtime

Linked Services

Data Sources
- Azure Blob Storage
- Azure Sql DB

---

**Pipelines**
*Group of Activities*

**Triggers**
*Switch to schedule & execute your pipeline*

**Activities**
*Operations on your dataset*

**Datasets**
*Representation of your data source*

**Integration Runtime**
*Compute Infrastructure for ADF*

**Linked Services**
*connection strings to your data sources*

# Building the Data Pipeline - Creating the Containers



## Source-

## New connection

# File or folder

## Source data store

Specify the source data store for the copy task. You can use an existing data store connection or specify a new data store.

Source type        ☐ Azure Data Lake Storage Gen2 ▾

Connection *       ☐ ls_adfsvinoworld_dev ▾    ✎ Edit    + New connection

landing/sampledata.zip      🗁 Browse

**Options**

☑ Binary copy ⓘ

**Compression type**

ZipDeflate (.zip) ▾

**Compression level**

▾

☐ Preserve zip file name as folder ⓘ

☑ Recursively ⓘ

☐ Delete files after completion ⓘ

   ‹ Previous    Next ›

# Destination data store

## Copy Data tool

- ✓ Properties
- ✓ Source
- ③ Destination
- ● Dataset
- ○ Configuration
- ④ Settings
- ⑤ Review and finish

### Destination data store

Specify the destination data store for the copy task. You can use an existing data store connection or specify a new data store.

Destination type     [ 🔷 Azure Data Lake Storage Gen2   ⌄ ]

Connection *     [ 🔷 ls_adfsvinoworld_dev   ⌄ ]    ✏ Edit   + New connection

**Folder path**

If the identity you use to access the data store only has permission to subdirectory instead of the entire account, specify the path to browse.

[ raw ]    📁 Browse

**File name**

[ ]

**Compression type**

[ None                      ⌄ ]

**Copy behavior** ⓘ

[ Select...                     ⌄ ]

**Max concurrent connections** ⓘ

[ ]

[ < Previous ]   [ Next > ]

# Settings

## Copy Data tool

- ✓ Properties
- ✓ Source
- ✓ Destination
- ④ Settings
- ⑤ Review and finish

### Settings

Enter name and description for the copy data task, more options for data movement

Task name *     [ pl_CopyAndUnzipfile ]

Task description     [ ]

Enable logging ⓘ     ☐

Enable staging ⓘ     ☐

> Advanced

[ < Previous ]   [ Next > ]                       [ Cancel ]

# Review

## Copy Data tool

- ✓ Properties
- ✓ Source
- ✓ Destination
- ✓ Settings
- ⑤ **Review and finish**
  - ● Review
  - ○ Deployment

### Summary

You are running pipeline to copy data from Azure Data Lake Storage Gen2 to Azure Data Lake Storage Gen2.

Azure Data Lake Storage Gen2 ⟶ Azure Data Lake Storage Gen2

✏ Edit

#### Properties

| | |
|---|---|
| Task name | pl_CopyAndUnzipfile |
| Task description | |

✏ Edit

#### Source

| | |
|---|---|
| Connection name | ls_adfsvinoworld_dev |
| Dataset name | SourceDataset_be7 |
| File name | sampledata.zip |

✏ Edit

#### Destination

[< Previous]  [Next >]

---

## Copy Data tool

- ✓ Properties
- ✓ Source
- ✓ Destination
- ✓ Settings
- ⑤ **Review and finish**
  - ● Review
  - ○ Deployment

### Summary

You are running pipeline to copy data from Azure Data Lake Storage Gen2 to Azure Data Lake Storage Gen2.

| | |
|---|---|
| Connection name | ls_adfsvinoworld_dev |
| Dataset name | SourceDataset_be7 |
| File name | sampledata.zip |

✏ Edit

#### Destination

| | |
|---|---|
| Connection name | ls_adfsvinoworld_dev |
| Dataset name | DestinationDataset_be7 |

✏ Edit

#### Copy settings

| | |
|---|---|
| Timeout | 0.12:00:00 |
| Retry | 0 |
| Retry interval (sec) | 30 |
| Secure output | false |
| Secure input | false |

[< Previous]  [Next >]

## Click on monitor

**Copy Data tool**

Azure Data Lake Storage Gen2 → Azure Data Lake Storage Gen2

- ✓ Properties
- ✓ Source
- ✓ Destination
- ✓ Settings
- **5 Review and finish**
  - Review
  - Deployment

## Deployment complete

| Deployment step | Status |
|---|---|
| Validating copy runtime environment | ✓ Succeeded |
| ❯ Creating datasets | ✓ Succeeded |
| ❯ Creating pipelines | ✓ Succeeded |
| ❯ Running pipelines | ✓ Succeeded |

Datasets and pipelines have been created. You can now monitor and edit the copy pipelines or click finish to close Copy Data Tool.

[ Finish ]  [ Edit pipeline ]  [ Monitor ]

---

**Dashboards**

Runs
- Pipeline runs
- Trigger runs
- Change Data Capture (previ...

Runtimes & sessions
- Integration runtimes
- Data flow debug

Notifications
- Alerts & metrics

### Pipeline runs

**Triggered**  Debug  |  ⟳ Rerun  ⊘ Cancel options ∨  ⟳ Refresh  ≡≡

▽ Filter by run ID or name   Chennai, Kolkata, Mu... : **Last 24 hours**

Status : **All**   Runs : **Latest runs**   Triggered by : **All**   ▽ Add filter

Showing 1 - 1 items

| Pipeline name ↑↓ | Run start ↑↓ | Run end ↑↓ |
|---|---|---|
| pl_CopyAndUnzipfile | 9/4/2025, 2:18:54 PM | 9/4/2025, 2:... |

**Pipeline run consumption**

**Name**
pl_CopyAndUnzipfile

**Status**
✓ Succeeded

**Run ID**
f7baa01f-b782-453f-ab54-171868261533

| | Quantity | Unit |
|---|---|---|
| **Pipeline orchestration** | | |
| Activity runs | 1 | Activity runs |
| **Pipeline execution** | | |
| **Azure integration runtime** | | |
| Data movement activities | 0.0667 | DIU-hour |

Learn more ↗
Pricing calculator ↗

[ Close ]

## Verify details on storage

# raw ...
Container

- Overview
- Diagnose and solve problems
- Access Control (IAM)
- > Settings

+ Add Directory   ↑ Upload   ↻ Refresh   | 🗑 Delete   📋 Copy   📋 Paste   ▭ Rename   🖉 Acquire lease   🖉 Break lease   ▦ Edit columns

raw  >  sampledata  >  Arancione

**Authentication method:** Access key (Switch to Microsoft Entra user account)

| Search blobs by prefix (case-sensitive) | | | Only show active objects ⌄ |

Showing all 8 items

| | Name | Last modified | Access tier | Blob type | Size | Lease state | |
|---|---|---|---|---|---|---|---|
| ☐ | 📁 [..] | | | | | | ⋯ |
| ☐ | 📄 Arancione_2022_01.csv | 9/4/2025, 2:19:13 PM | Hot (Inferred) | Block blob | 10.88 KiB | Available | ⋯ |
| ☐ | 📄 Arancione_2022_02.csv | 9/4/2025, 2:19:14 PM | Hot (Inferred) | Block blob | 10.97 KiB | Available | ⋯ |
| ☐ | 📄 Arancione_2022_03.csv | 9/4/2025, 2:19:14 PM | Hot (Inferred) | Block blob | 10.81 KiB | Available | ⋯ |
| ☐ | 📄 Arancione_2022_04.csv | 9/4/2025, 2:19:14 PM | Hot (Inferred) | Block blob | 10.98 KiB | Available | ⋯ |
| ☐ | 📄 Arancione_2022_05.csv | 9/4/2025, 2:19:14 PM | Hot (Inferred) | Block blob | 11.08 KiB | Available | ⋯ |
| ☐ | 📄 Arancione_2022_06.csv | 9/4/2025, 2:19:14 PM | Hot (Inferred) | Block blob | 10.95 KiB | Available | ⋯ |
| ☐ | 📄 Arancione_2022_07.csv | 9/4/2025, 2:19:14 PM | Hot (Inferred) | Block blob | 10.89 KiB | Available | ⋯ |
| ☐ | 📄 Arancione_2022_08.csv | 9/4/2025, 2:19:14 PM | Hot (Inferred) | Block blob | 10.85 KiB | Available | ⋯ |

Add or remove favorites by pressing Ctrl+Shift+F

# Importing Semi-Structured Data - Building the Pipeline

## Source

### Copy Data tool



### Click on preview data to see details

# Copy Data tool

- ✓ **Properties**
- ② **Source**
  - • Dataset
  - • Configuration
- ③ Destination
- ④ Settings
- ⑤ Review and finish

## File format settings

**File format**

| DelimitedText ▾ |  | Detect text format |  | 👓 Preview data |

**Column delimiter**

| Comma (,) ▾ |

☐ Edit

**Row delimiter**

| Default (\r,\n, or \r\n) ▾ |

☐ Edit

☑ First row as header ⓘ

> **Advanced**

**Compression type**

| No compression ▾ |

**Additional columns** ⓘ

＋ New

[ < Previous ]  [ Next > ]

---

# Destination

## Copy Data tool

- ✓ **Properties**
- ✓ **Source**
- ③ **Destination**
  - • Dataset
  - ○ Configuration
- ④ Settings
- ⑤ Review and finish

## Destination data store

Specify the destination data store for the copy task. You can use an existing data store connection or specify a new data store.

**Destination type**

| 🗄 Azure Data Lake Storage Gen2 ▾ |

**Connection** *

| 🗄 ls_adfsvinoworld_dev ▾ |  ✏ Edit  ＋ New connection

**File name**

| |

**Copy behavior** ⓘ

| Select... ▾ |

**Max concurrent connections** ⓘ

| |

**Block size (MB)** ⓘ

| |

**Metadata** ⓘ

＋ New

[ < Previous ]  [ Next > ]

# Settings

## Copy Data tool

- ✓ Properties
- ✓ Source
- ✓ Destination
- ④ Settings
- ⑤ Review and finish

### Settings

Enter name and description for the copy data task, more options for data movement

| | |
|---|---|
| Task name * | CopyPipeline_0a9 |
| Task description | |
| Data consistency verification ⓘ | ☐ |
| Fault tolerance ⓘ | ▾ |
| Enable logging ⓘ | ☐ |
| Enable staging ⓘ | ☐ |
| ⟩ Advanced | |

⟨ Previous    Next ⟩

---

# Settings

## Copy Data tool

- ✓ Properties
- ✓ Source
- ✓ Destination
- ✓ Settings
- ⑤ Review and finish
  - ● Review
  - ○ Deployment

### Summary

You are running pipeline to copy data from Azure Data Lake Storage Gen2 to Azure Data Lake Storage Gen2.

Azure Data Lake Storage Gen2 ⟶ Azure Data Lake Storage Gen2

✎ Edit

#### Properties

| | |
|---|---|
| Task name | CopyPipeline_0a9 |
| Task description | |

✎ Edit

#### Source

| | |
|---|---|
| Connection name | ls_adfsvinoworld_dev |
| Dataset name | SourceDataset_0a9 |
| Column delimiter | , |
| Escape character | \ |
| Quote char | " |
| First row as header | true |
| Folder path | sampledata/Arancione |

✎ Edit

#### Destination

| | |
|---|---|
| Connection name | ls_adfsvinoworld_dev |
| Dataset name | DestinationDataset_0a9 |
| Column delimiter | , |
| Escape character | \ |

⟨ Previous    Next ⟩

**Copy Data tool**

| | |
|---|---|
| ✓ Properties | **Summary** |
| ✓ Source | You are running pipeline to copy data from Azure Data Lake Storage Gen2 to Azure Data Lake Storage Gen2. |
| ✓ Destination | Column delimiter , |
| ✓ Settings | Escape character \ |
| ⑤ Review and finish | Quote char " |
| • Review | First row as header true |
| ○ Deployment | Folder path sampledata/Arancione |

*Edit*

**Destination**

| | |
|---|---|
| Connection name | ls_adfsvinoworld_dev |
| Dataset name | DestinationDataset_0a9 |
| Column delimiter | , |
| Escape character | \ |
| Quote char | " |
| First row as header | true |
| Folder path | salesdata |

*Edit*

**Copy settings**

| | |
|---|---|
| Timeout | 0.12:00:00 |
| Retry | 0 |
| Retry interval (sec) | 30 |
| Secure output | false |
| Secure input | false |

# Click on monitor

**Copy Data tool**



Azure Data Lake Storage Gen2 → Azure Data Lake Storage Gen2

## Deployment complete

| Deployment step | Status |
|---|---|
| Validating copy runtime environment | ✓ Succeeded |
| › Creating datasets | ✓ Succeeded |
| › Creating pipelines | ✓ Succeeded |
| › Running pipelines | ✓ Succeeded |

Datasets and pipelines have been created. You can now monitor and edit the copy pipelines or click finish to close Copy Data Tool.

[Finish] [Edit pipeline] [Monitor]

**Pipeline runs**

Triggered | Debug | Rerun | Cancel options | Refresh | Edit columns

Filter by run ID or name | Chennai, Kolkata, Mu... : **Last 24 hours** | Pipeline name : C

Runs : Latest runs | Triggered by : **All** | Add filter | ✕

Showing 1 - 1 items

| Pipeline name ↑↓ | Run start ↑↓ | Run end ↑↓ |
|---|---|---|
| CopyPipeline_0a9 | 9/4/2025, 2:52:35 PM | 9/4/2025, 2:52:58 PM |

**Pipeline run consumption**

**Name**
CopyPipeline_0a9

**Status**
✓ Succeeded

**Run ID**
5de22be5-d531-402b-973e-7e7bc60fbc66

| | Quantity | Unit |
|---|---|---|
| **Pipeline orchestration** | | |
| Activity runs | 1 | Activity runs |
| **Pipeline execution** | | |
| **Azure integration runtime** | | |
| Data movement activities | 0.0667 | DIU-hour |

Learn more ☐
Pricing calculator ☐

Close

# Verify data

Home > Storage center | Storage accounts (Blobs) > adfsvinoworld | Containers >

**cleansed**
Container

Search

- Overview
- Diagnose and solve problems
- Access Control (IAM)
- Settings

+ Add Directory | ↑ Upload | ↻ Refresh | 🗑 Delete | Copy | Paste | Rename | Acquire lease | Break lease | Edit columns

cleansed > salesdata

**Authentication method:** Access key (Switch to Microsoft Entra user account)

Search blobs by prefix (case-sensitive) | Only show active objects

Showing all 8 items

| | Name | Last modified | Access tier | Blob type | Size | Lease state |
|---|---|---|---|---|---|---|
| | 📁 [..] | | | | | |
| | Arancione_2022_01.csv | 9/4/2025, 2:52:56 PM | Hot (Inferred) | Block blob | 10.88 KiB | Available |
| | Arancione_2022_02.csv | 9/4/2025, 2:52:56 PM | Hot (Inferred) | Block blob | 10.97 KiB | Available |
| | Arancione_2022_03.csv | 9/4/2025, 2:52:56 PM | Hot (Inferred) | Block blob | 10.81 KiB | Available |
| | Arancione_2022_04.csv | 9/4/2025, 2:52:56 PM | Hot (Inferred) | Block blob | 10.98 KiB | Available |
| | Arancione_2022_05.csv | 9/4/2025, 2:52:56 PM | Hot (Inferred) | Block blob | 11.08 KiB | Available |
| | Arancione_2022_06.csv | 9/4/2025, 2:52:56 PM | Hot (Inferred) | Block blob | 10.95 KiB | Available |
| | Arancione_2022_07.csv | 9/4/2025, 2:52:56 PM | Hot (Inferred) | Block blob | 10.89 KiB | Available |
| | Arancione_2022_08.csv | 9/4/2025, 2:52:56 PM | Hot (Inferred) | Block blob | 10.85 KiB | Available |

Add or remove favorites by pressing Ctrl+Shift+F

**Naming conventions**

# NAMING CONVENTIONS

*Naming conventions for resources and components within Azure and Azure Data Factory*

| RESOURCE GROUP | STORAGE ACCOUNT | ADLS GEN2 | DATA FACTORY |
|---|---|---|---|
| **{ABCDEF}-{ENV}-RG** | **ST{ABCDEF}{ENV}001** | **ADLS{ABCDEF}{ENV}** | **{ABCDEF}-{ENV}-ADF** |
| e.g.: *vinoworld-dev-rg* | e.g.: *adlsvinoworlddev001* | e.g.: *adlsvinoworlddev* | e.g.: *vinoworld-dev-adf* |
| Suffix "rg" for resource group | Prefix "st" for storage | Prefix "adls" for Azure Data Lake Storage | Suffix "adf" for Azure Data Factory |
| Suffix "env" identifies environment | Specify "env" to identify environment | Specify "env" to identify environment | Specify "env" to identify environment |

*Naming the components of your project using best practices and doing it consistently makes the solution more manageable and easy to maintain.*

# PipelineActivities and Parameters


# Activity Dependencies – Examples

## Create pipeline and drag drop validation select  dataset (cleansed sales)




## Click on set variable and run pipeline

**Set variable execute when validation succeed or failed**

# Skip activity

# Expression & variabls

**Keyvault setup**

# Create a key vault  ···

Basics     Access configuration     Networking     Tags     **Review + create**

◉ View Automation Template

### Basics

| | |
|---|---|
| Subscription | Azure subscription 1 |
| Resource group | vinoworld-dev-rg |
| Key vault name | vinoworld-dev-kvault |
| Region | Germany West Central |
| Pricing tier | Standard |
| Soft-delete | Enabled |
| Purge protection during retention period | Disabled |
| Days to retain deleted vaults | 90 days |

### Access configuration

[ Previous ]  [ Next ]  [ Create ]

---

Home > vinoworld-dev-kvault

🔒 **vinoworld-dev-kvault** | Secrets  ☆  ···
Key vault

✅ Creating the secret 'adfsvinoworld-dev'.
The secret 'adfsvinoworld-dev' has been successfully created.  ✕

🔍 Search        « 

🕐 Overview
📋 Activity log
🔑 Access control (IAM)
🏷 Tags
🔧 Diagnose and solve problems
☰ Access policies

+ Generate/Import   ↻ Refresh   ⬆ Restore Backup   🔗 Manage deleted secrets   </> View sample code

ℹ The secret 'adfsvinoworld-dev' has been successfully created.

| Name | Type | Status | Expiration date |
|---|---|---|---|
| adfsvinoworld-dev | | ✓ Enabled | |

## Azure Key Vault - Create Linked Service

## Linked services

Linked service defines the connection information to a data store or compute. Learn more ⧉

+ New

Filter by name                   Annotations : Any

Showing 1 - 1 of 1 items

| Name ↑↓ | Type ↑↓ |
|---------|---------|
| ls_adfsvinoworld_dev | Azure Data Lake Storage Gen2 |

---

🔑 Azure Key Vault

Azure key vault selection method ⓘ
⦿ From Azure subscription    ⃝ Enter manually

Azure subscription ⓘ
Azure subscription 1 (b435e57d-913c-4896-baf9-12e1971e7539)          ⌄

Azure key vault name *
vinoworld-dev-kvault                                              ⌄    ↻
Edit key vault

Authentication method
System-assigned managed identity                                     ⌄

Managed identity name: **vinoworld-dev-adf01**
Managed identity object ID: **a94b1e30-b4e5-402b-8dba-9ba724c394b9**
Grant Data Factory service managed identity access to your Azure Key Vault. Learn more ⧉

Test connection
⦿ To linked service    ⃝ To secret

Annotations
+ New

> Parameters

✓ Connection successful

| Create | Back |    🔌 Test connection    Cancel