What Is ADF?

- ADF is defined as a **data integration service**.
- The aim of ADF is to fetch data from one or more data sources and convert them into a format that we process.
- The data sources might contain noise that we need to filter out. ADF connectors enable us to pull the interesting data and remove the rest.
- ADF to ingest data and load the data from a variety of sources into Azure Data Lake Storage.
- It is the **cloud-based ETL** service that allows us to create data-driven pipelines for **orchestrating** data movement and transforming data at scale.

What Is a Data Integration Service?

- Data integration involves the collection of data from one or more sources.
- Then includes a process where the data may be transformed and cleansed or may be augmented with additional data and prepared.
- Finally, the combined data is stored in a data platform service that deals with the type of analytics that we want to perform.
- This process can be automated by ADF in an arrangement known as **Extract, Transform, and Load (ETL)**.

What Is ETL?

1) Extract

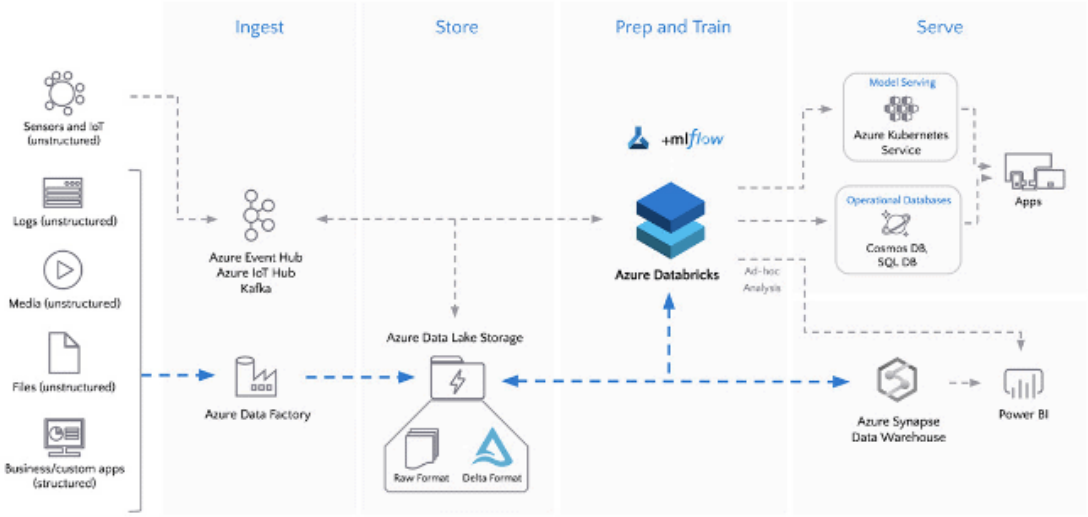
- In this extraction process, data engineers define the data and its source.
- **Data source**: Identify source details such as the subscription, resource group, and identity information such as secretor a key.
- **Data**: Define data by using a set of files, a database query, or an Azure Blob storage name for blob storage.

2) Transform

- Data transformation operations can include combining, splitting, adding, deriving, removing, or pivoting columns.
- Map fields between the data destination and the data source.

3) Load

- During a load, many Azure destinations can take data formatted as a file, JavaScript Object Notation (JSON), or blob.
- Test the ETL job in a test environment. Then shift the job to a production environment to load the production system.



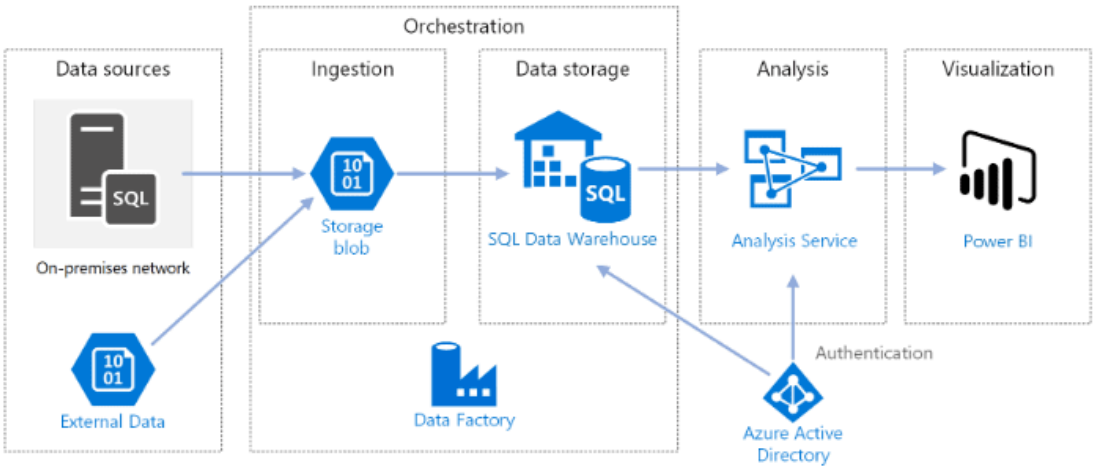
4) ETL tools

- Azure Data Factory provides approximately 100 enterprise connectors and robust resources for both code-based and code-free users to accomplish their data transformation and movement needs.

What Is Meant By Orchestration?

- Sometimes ADF will instruct another service to execute the actual work required on its behalf, such as a Databricks to perform a transformation query.

- ADF hardly orchestrates the execution of the query and then prepare the pipelines to move the data onto the destination or next step.



Copy Activity In ADF

- In ADF, we can use the Copy activity to copy data between data stores located on-premises and in the cloud.
- After we copy the data, we can use other activities to further transform and analyze it.
- We can also use the DF Copy activity to publish transformation and study results for business intelligence (BI) and application consumption.

1) Monitor Copy Activity

- Once we've created and published a pipeline in ADF, we can associate it with a trigger.
- We can monitor all of our pipelines runs natively in the ADF user experience.
- To monitor the Copy activity run, go to your DF **Author & Monitor** UI.
- On the **Monitor** tab page, we see a list of the pipeline runs, click the **pipeline name** link to access the list of activity runs in the pipeline run.

2) Delete Activity In ADF

- Back up your files before you are deleting them with the **Delete activity** in case you wish to restore them in the future.
- Make sure that Data Factory has to write permissions to delete files or folders or from the storage store.

How ADF work?

1) Connect and Collect

- Enterprises have data of various types such as structured, unstructured, and semi-structured.
- The first step collects all the data from a different source and then move the data to a centralized location for subsequent processing.
- We can use the Copy Activity in a data pipeline to move data from both cloud source and on-premises data stores to a centralized data store in the cloud.

2) Transform and Enrich

- After data is available in a centralized data store in the cloud, transform, or process the collected data by using ADF mapping data flows.
- ADF supports external activities for executing our transformations on compute services such as **Spark, HDInsight Hadoop, Machine Learning, Data Lake Analytics.**

3) CI/CD and Publish

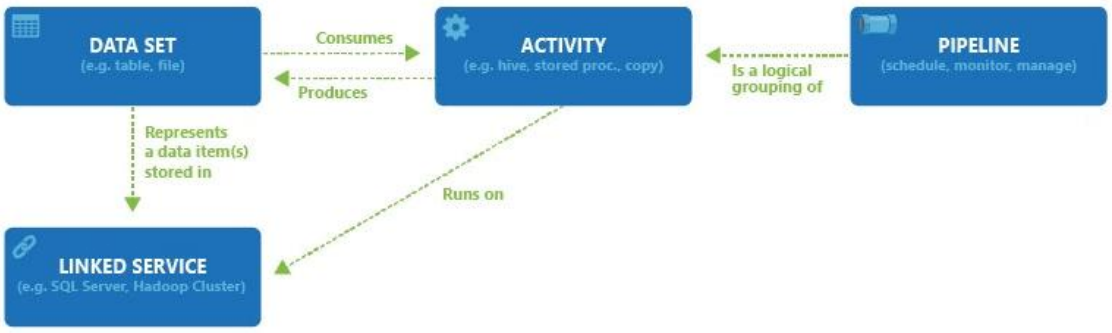
- ADF offers full support for CI/CD of our data pipelines using GitHub and Azure DevOps.
- After the raw data has been refined, ad the data into **Azure SQL Database, Azure Data Warehouse, Azure CosmosDB**

4) Monitor

- ADF has built-in support for pipeline monitoring via Azure Monitor, PowerShell, API, Azure Monitor logs, and health panels on the Azure portal.

5) Pipeline

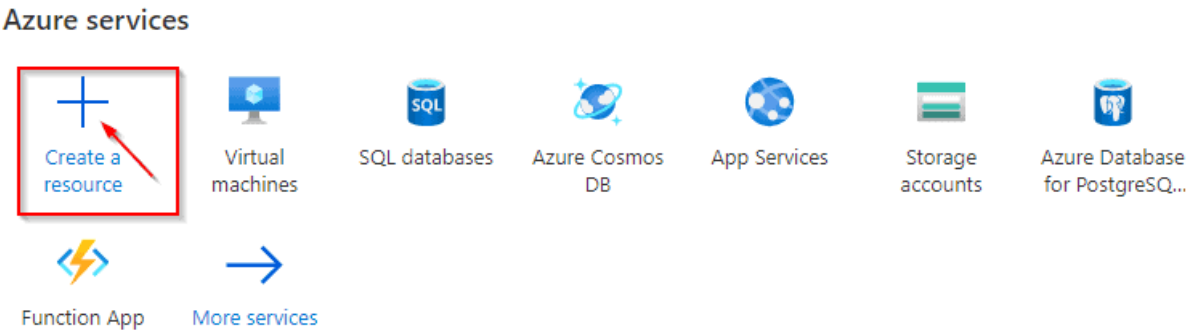
- A pipeline is a logical grouping of activities that execute a unit of work. Together, the activities in a pipeline execute a task.

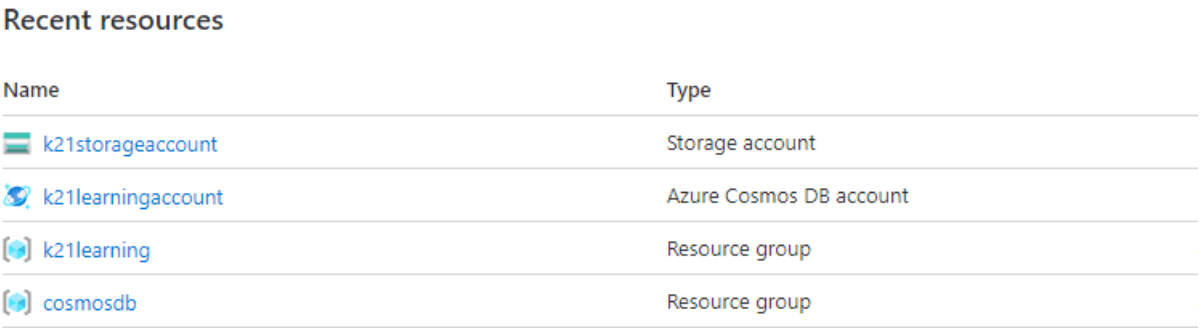


How To Create An ADF

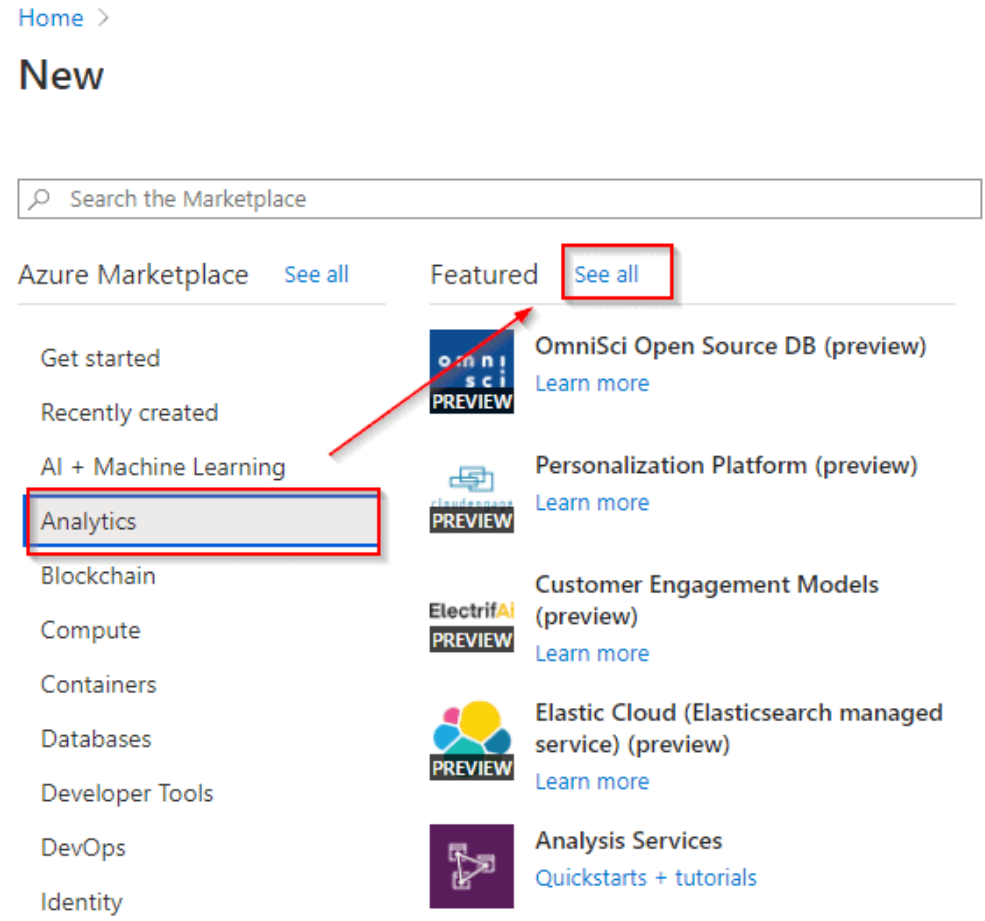
1) Go to the Azure portal.

2) From the portal menu, Click on Create a resource.

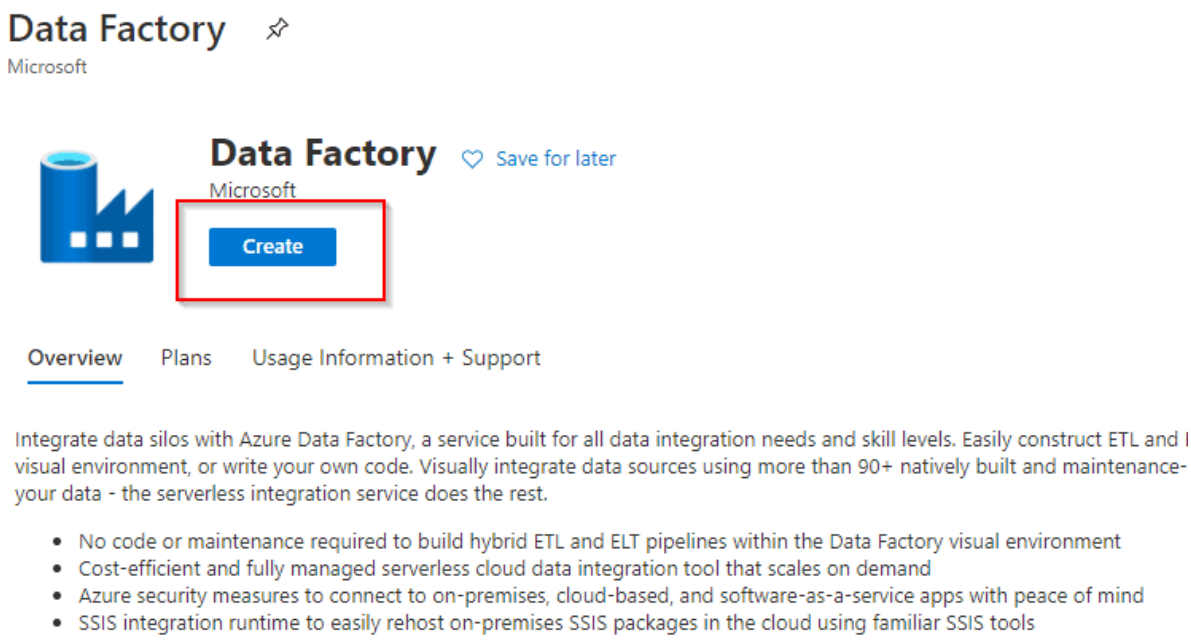




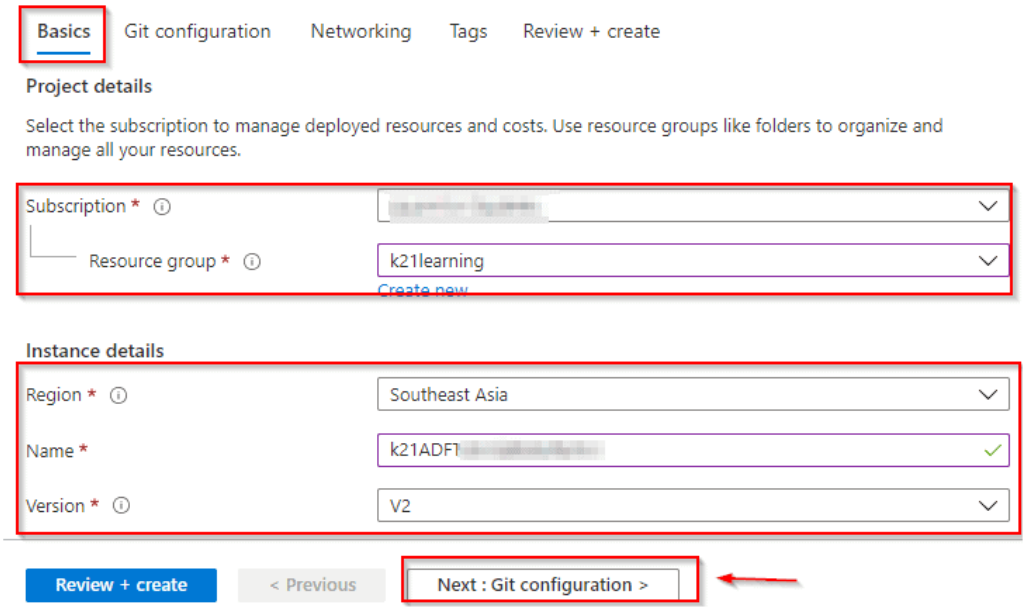
3) Select Analytics, and then select see all.



4) Select Data Factory, and then select Create

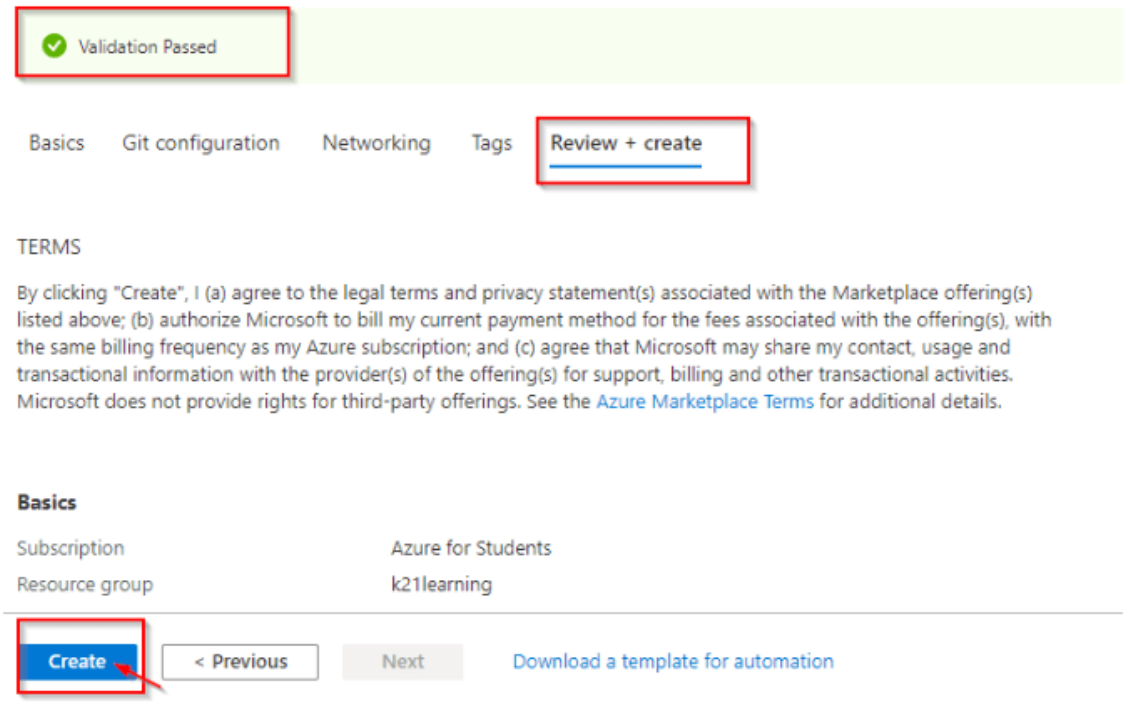


5) On the Basics Details page, Enter the following details. Then Select Git Configuration.



6) On the Git configuration page, Select the Check the box, and then Go To Networking.

7) On the Networking page, don't change the default settings and click on Tags, and the Select Create.



8) Select Go to resource, and then Select Author & Monitor to launch the Data Factory UI in a separate tab.