

Name: Abhishek Kanoujia

DATA ENGINEERING BATCH 1

DAY 23 ASSIGNMENT

Class hand written notes:-

08-08-2022
Tuesday

Azure databricks

Date: (29.01.2022)

What is Azure data factory?

cloud based data integration service that allow you create data-driven workflows in the cloud for orchestrating and automating data movement and data transformation.

ADF Azure use for:

- Supporting data migrations.
- Getting data from a client's server or online data to an Azure Data lake.
- Carrying out various data integration process.
- Integration data from different ERP system and loading it into Azure. (hosted on Azure)

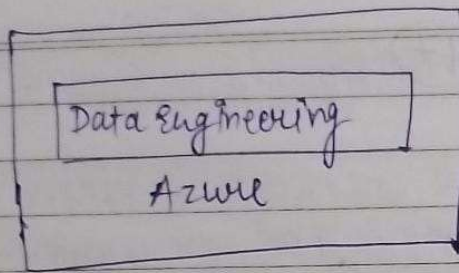
four thing:

sql
python
pyspark
sparksql

azure databricks
azure data lake/ adls gen 2
azure data factory
azure synapse

[azure devops.]

(cloudbased integration service)



(Shortcut)

(ADB, ADL, ADLS, ADFAS, ADF)

Azure Data-factory: Data driven pipeline

Move \rightarrow transform \rightarrow run data in the pipeline of ADF.

Step:1 connect & collect

Step:2 transform & enrich

Step:3 Publish

In microsoft Azure migrate btw. two clouds data stores and btw. on-premise data store and cloud data store

Copy Activity in Azure data Factory copies data from source data store to sink data store

(Azure Blob storage, Azure Cosmos DB)

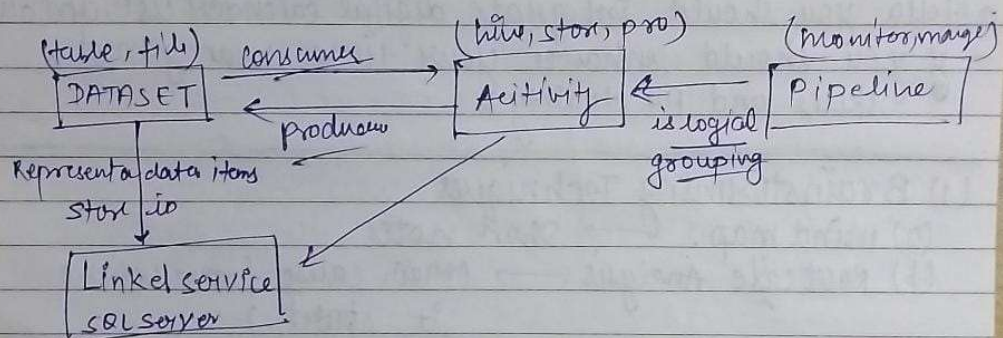
(ADF Pipeline) (ADF Key Component)

\Rightarrow Datasets represent data structure within the data stores.

⇒ A pipeline is group of activities.

⇒ Activities define the actions to perform on your data.

⇒ Linked service define the information needed for Azure Data Data factory to connect to external Resource.

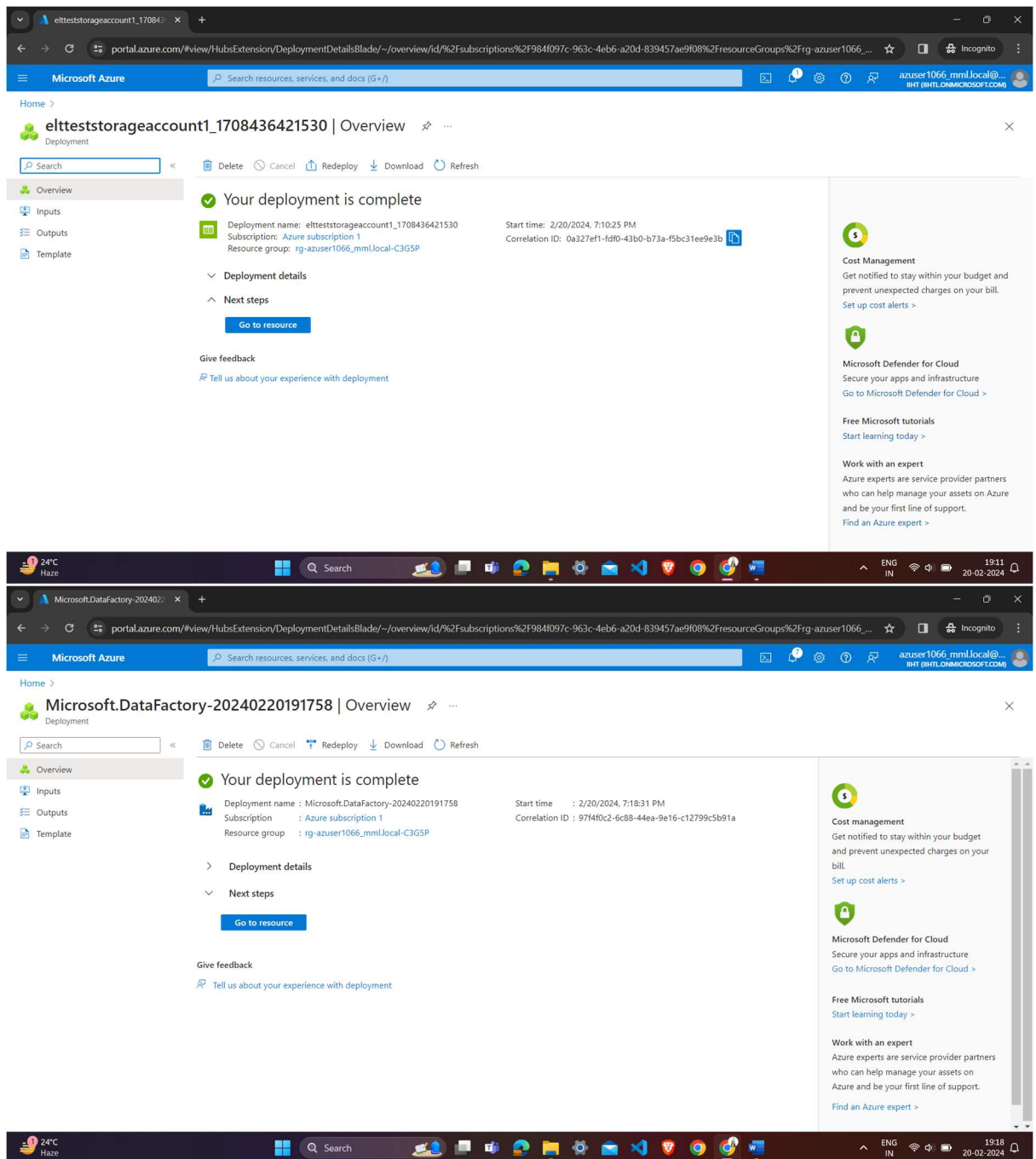


You can use one of the following tool or API to create data pipeline in Azure Data factory.

- Azure Portal
- Visual Studio.
- PowerShell
- Net API
- Rest API
- Azure Resource management plat

Data Factory blob to blob copy files hand on :-

1. Create two storage account
2. Create one eltteststroage1 and other is eltteststroage22
3. Create data factory account and do the configuration part screen short attached below.



blobtoblob1066 - Microsoft Azure

portal.azure.com/#@ihlt.onmicrosoft.com/resource/subscriptions/984f097c-963c-4eb6-a20d-839457ae9f08/resourceGroups/rg-azuser1066_mmlLocal-C3GSP/providers/Microsoft.Data...

Microsoft Azure Search resources, services, and docs (G+)

Home > Microsoft.DataFactory-20240220191758 | Overview >

blobtoblob1066 Data factory (V2)

Search Delete

Overview

- Activity log
- Access control (IAM)
- Tags
- Diagnose and solve problems

Settings

- Networking
- Managed identities
- Properties
- Locks

Getting started

- Quick start

Monitoring

- Alerts
- Metrics
- Diagnostic settings

Essentials

Resource group (move) : rg-azuser1066_mmlLocal-C3GSP

Status : Succeeded

Location : East US

Subscription (move) : Azure subscription 1

Subscription ID : 984f097c-963c-4eb6-a20d-839457ae9f08

Type : Data factory (V2)

Getting started : [Quick start](#)

Azure Data Factory Studio

[Launch studio](#)

Quick Starts Tutorials Template Gallery Training Modules

24°C Haze

blobtoblob1066 - Microsoft Azure

blobtoblob1066 - Azure Data Factory

Microsoft Azure Data Factory blobtoblob1066

Search factory and documentation

Microsoft recently announced the public preview of Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click [here](#) to get started with Fabric Data Factory!

Azure Data Factory allows you to configure a Git repository with either Azure DevOps or GitHub. Git is a version control system that allows for easier change tracking and collaboration. [Learn more](#)

[Set up code repository](#)

Data factory

blobtoblob1066

New

Ingest
Copy data at scale once or on a schedule.

Orchestrate
Code-free data pipelines.

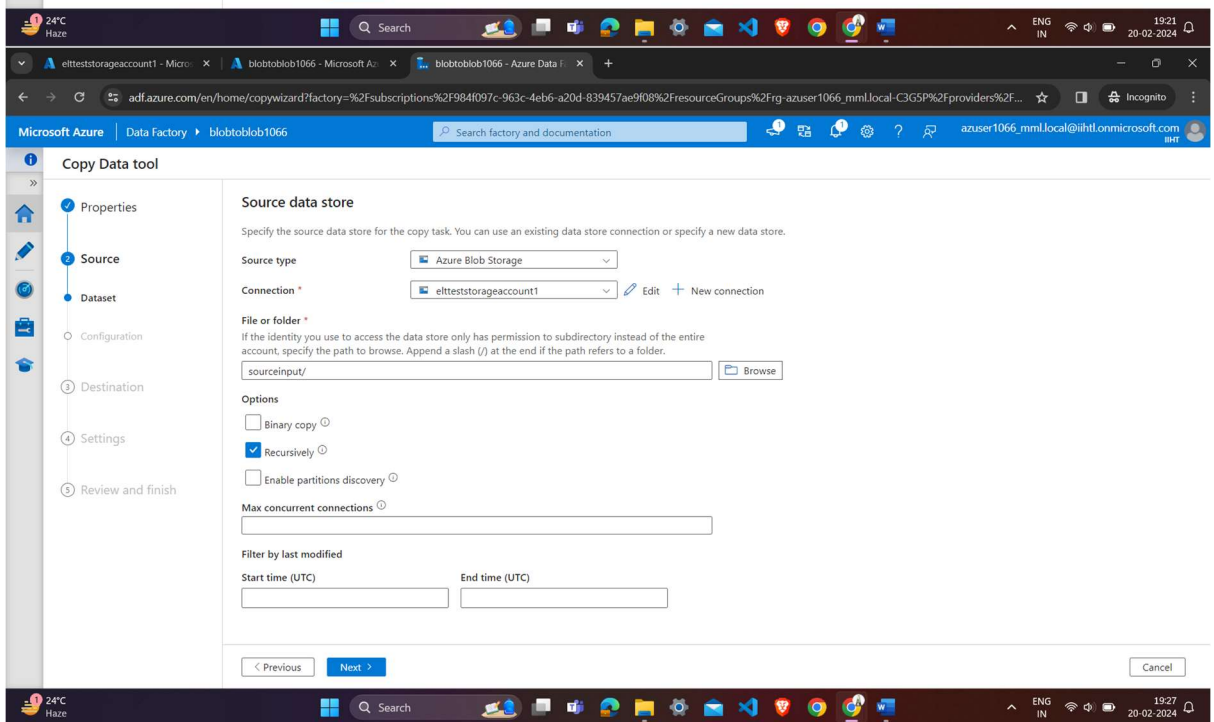
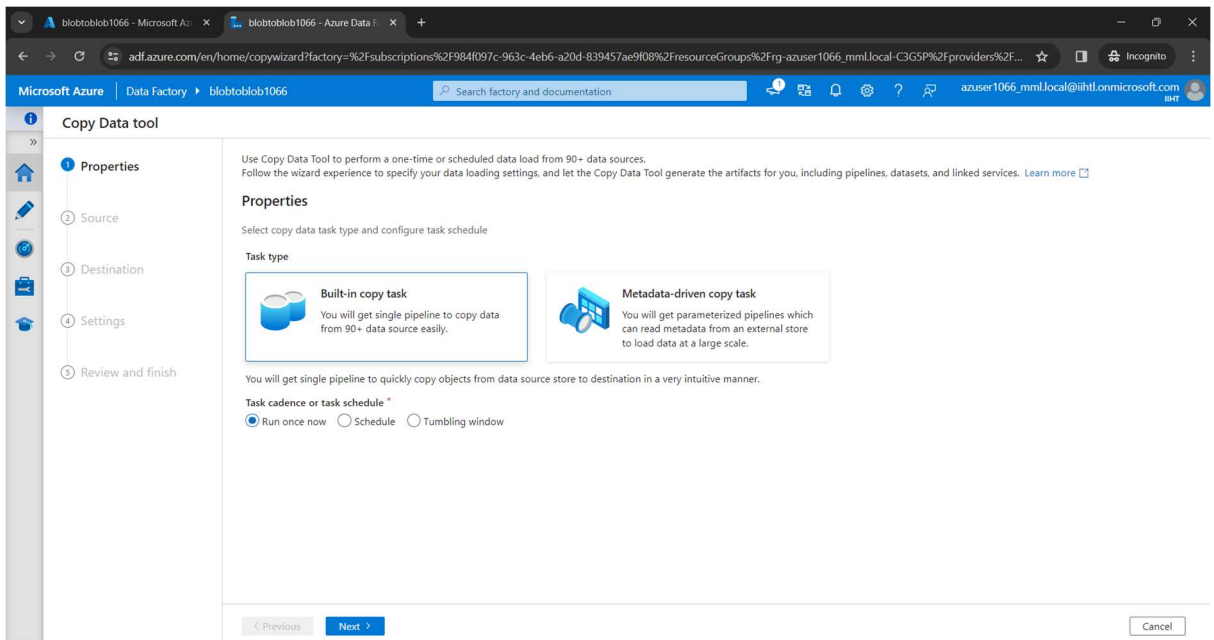
Transform data
Transform your data using data flows.

Configure SSIS
Manage & run your SSIS packages in the cloud.

Recent resources

No items to show

24°C Haze



Microsoft Azure | Data Factory | blobblob1066

Copy Data tool

Properties
Source
Destination
Dataset
Configuration
Settings
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 Blob Storage

Connection: elteststorageaccount1 Edit + New connection

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.

sourceinput/ Browse

File name: Bibadata.csv

Copy behavior: Select...

Max concurrent connections:

Block size (MB):

Metadata: + New

< Previous Next > Cancel

Microsoft Azure | Data Factory | blobblob1066

Microsoft recently announced the public preview of Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click [here](#) to get started with Fabric Data Factory!

All pipeline runs > CopyPipeline_gmz - Activity runs

Rerun Cancel Refresh Update pipeline List Gantt

Error details

Copy_gmz

Error code: UserErrorSourceBlobNotExist

Failure type: User configuration issue

Details: ErrorCode=UserErrorSourceBlobNotExist, Type=Microsoft.DataTransfer.Common.Shared.HybridDeliveryException, Message=The required Blob is missing. Please contact Blob Storage team if you need further support. ContainerName: sourceinput, path: sourceinput/_Source=Microsoft.DataTransfer.ClientLibrary, Type=Microsoft.WindowsAzure.Storage.StorageException, Message=The remote server returned an error:

Copy data

Copy_gmz

Monitor in Azure Metrics Export to CSV

Activity	Start	Duration	Integration runtime	User properties	Destination
Copy_gmz	2/20/2024, 7:36:04 PM	14s	AutoResolveIntegrator		sourceinput/