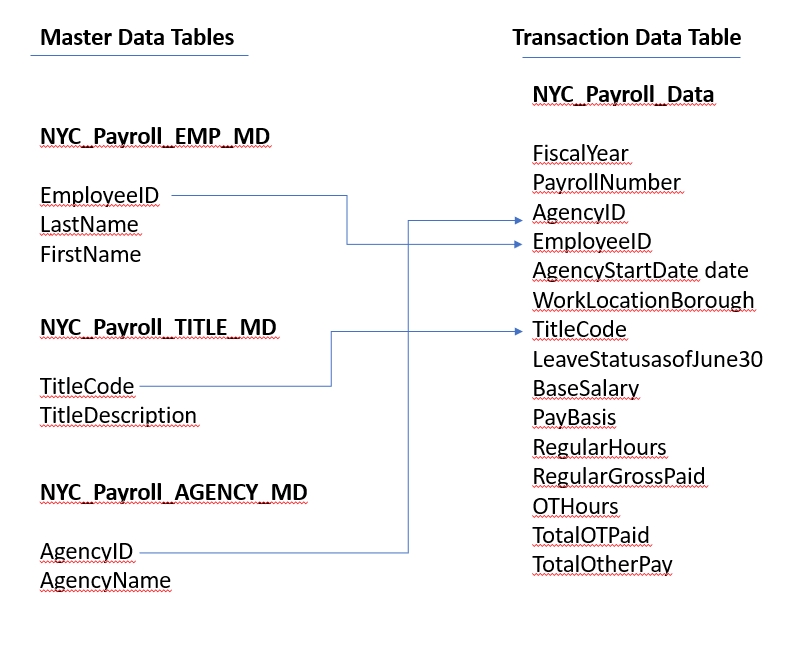
Project Introduction

The City of New York would like to develop a Data Analytics platform on Azure Synapse Analytics to accomplish two primary objectives:

1. Analyze how the City's financial resources are allocated and how much of the City's budget is being devoted to overtime.
2. Make the data available to the interested public to show how the City’s budget is being spent on salary and overtime pay for all municipal employees.

You have been hired as a Data Engineer to create high-quality data pipelines that are dynamic, can be automated, and monitored for efficient operation. The project team also includes the city’s quality assurance experts who will test the pipelines to find any errors and improve overall data quality.

The source data resides in Azure Data Lake and needs to be processed in a NYC data warehouse in Azure Synapse Analytics. The source datasets consist of CSV files with Employee master data and monthly payroll data entered by various City agencies.



For this project, you'll do your work in the Azure Portal, using several Azure resources including:

* Azure Data Lake Gen2 (Storage account with Hierarchical Namespaces checkbox checked when creating)
* Azure SQL DB
* Azure Data Factory
* Azure Synapse Analytics

Instructions for using a temporary Azure account to complete the project are on the next page.

You'll take screenshots as proof of work for this project, so remember to collect these screenshots throughout the project steps. A checklist is provided at the end of each step so you can double-check you've collected all of the deliverables.

You'll also need to create a Github repository for this project. At the end of the project, you will connect your Azure pipelines to Github and submit the URL or contents of the repository.

## Project Instructions

For this project, you'll do your work in the Azure Portal, using several Azure resources including:

* Azure Data Lake Gen2
* Azure SQL DB
* Azure Data Factory
* Azure Synapse Analytics

Instructions for using a temporary Azure account to complete the project are on the previous page.

You'll take screenshots as proof of work for this project, so remember to collect these screenshots throughout the project steps. A checklist is provided at the end of each step so you can double-check you've collected all of the screenshot deliverables.

You'll also need to create a Github repository for this project. At the end of the project, you will connect your Azure pipelines to Github and submit the URL or contents of the repository.

When you submit your project, it will be assessed against this [**project rubric**](https://learn.udacity.com/rubric/4857). Take a moment to review it now and periodically throughout the project to make sure you're meeting the requirements.

https://web.azuresynapse.net/en/authoring/explore/linked/storageaccounts/UdacityData1-udacitystorage2023%2Fadlsnycpayroll-louis-b?subFolderPath=&workspace=%2Fsubscriptions%2F7916f776-e9e8-4273-9d6a-27d794cf81bc%2FresourceGroups%2Flouis-rg%2Fproviders%2FMicrosoft.Synapse%2Fworkspaces%2Flouis-synapse-ws1

The access level is set to private because anonymous access is disabled on this storage account.

ADLS Gen2 operation failed for: Operation returned an invalid status code 'Conflict'. Account: 'test112233access'. ErrorCode: 'EndpointUnsupportedAccountFeatures'. Message: **'This endpoint does not support BlobStorageEvents or SoftDelete**. Please disable these account features if you would like to use this endpoint.'. RequestId: '25012059-801f-001e-24f5-310186000000'. TimeStamp: 'Mon, 18 Dec 2023 21:03:03 GMT'..Operation returned an invalid status code 'Conflict'

IF NOT EXISTS (SELECT \* FROM sys.external\_file\_formats WHERE name = 'CsvFormatWithHeader')

    CREATE EXTERNAL FILE FORMAT [CsvFormatWithHeader]

    WITH ( FORMAT\_TYPE = DELIMITEDTEXT ,

           FORMAT\_OPTIONS (

             FIELD\_TERMINATOR = ',',

             FIRST\_ROW = 2,

             STRING\_DELIMITER = '"',

             USE\_TYPE\_DEFAULT = FALSE

            ))

GO

--SELECT \* FROM sys.external\_file\_formats;

IF NOT EXISTS (SELECT \* FROM sys.external\_data\_sources WHERE name = 'container1\_test33221100\_dfs\_core\_windows\_net')

    CREATE EXTERNAL DATA SOURCE [container1\_test33221100\_dfs\_core\_windows\_net]

    WITH (

        LOCATION = 'abfss://container1@test33221100.dfs.core.windows.net'

    )

GO

CREATE EXTERNAL TABLE dbo.louis15 (

    [EmployeeID] bigint,

    [LastName] nvarchar(400),

    [FirstName] nvarchar(400)

    )

    WITH (

    LOCATION = 'EmpMaster.csv',

    DATA\_SOURCE = [container1\_test33221100\_dfs\_core\_windows\_net],

    FILE\_FORMAT = [CsvFormatWithHeader]

    )

GO

SELECT TOP 10 \* FROM dbo.louis15

GO

## Data Factory Network Settings

**Managed virtual network**

Choose whether you want the default AutoResolveIntegrationRuntime to be provisioned on demand inside an ADF-managed virtual network. If this setting is disabled, after the data factory is created, you can still choose whether to provision explicitly created Azure integration runtime inside an ADF-managed virtual network.

[Learn more](https://learn.microsoft.com/azure/data-factory/managed-virtual-network-private-endpoint?WT.mc_id=Portal-Microsoft_Azure_DataFactory)

Enable Managed Virtual Network on the default AutoResolveIntegrationRuntime

**Self-hosted integration runtime inbound connectivity to Azure Data Factory service**

Choose whether to connect your self-hosted integration runtime to Azure Data Factory via public endpoint or private endpoint. This applies to self-hosted integration runtime running either on premises or inside customer managed Azure virtual network

[Learn more](https://learn.microsoft.com/azure/data-factory/data-factory-private-link?WT.mc_id=Portal-Microsoft_Azure_DataFactory)

Connect via

* Public endpoint
* Private endpoint

: Please create a master key in the database or open the master key in the session before performing this

**Account selection method**

From Azure subscription

Enter manually

**Fully qualified domain name**

tcp:louis-test-synapse-ws1.sql.azuresynapse.net,1433

**Database name**

@{linkedService().DBName}

**Authentication type**

System Assigned

Managed identity name:

**louis-test-synapse-ws1**

Managed identity object ID:

**5d5c59dd-81be-4efb-93d8-a2ca483f2bfe**

Grant workspace service managed identity access to your Azure Synapse Analytics.