# Pipeline Summary: Dynamic\_Copy\_to\_Datalake

This document summarizes the Azure Data Factory pipeline titled 'Dynamic\_Copy\_to\_Datalake'. The pipeline is designed to dynamically copy multiple datasets from web sources into Azure Data Lake Storage Gen2 (ADLS) in the Bronze layer for further processing.

## Overview

The pipeline enables scalable data ingestion by leveraging a dynamic and parameterized copy mechanism. It first reads metadata about the source files and then loops through each entry to copy the data into ADLS.

## Pipeline Components & Flow

1. 1. Lookup\_git (Lookup Activity)

- Reads metadata (like URLs, folder names) from a JSON file stored in ADLS.  
- Configured to fetch all rows (not just the first one).  
- Acts as the source of truth for what data needs to be copied.

1. 2. ForEach\_Git (ForEach Activity)

- Iterates over each metadata entry retrieved by the Lookup activity.  
- Executes the Copy activity for each file entry (sequentially).

1. 3. Copy\_to\_Bronze (Copy Activity)

- Source: Pulls data using HTTP GET from a web API that serves delimited text (CSV-like format).  
- Sink: Stores the output into ADLS Gen2 Bronze layer as '.txt' files.  
- Translator: Uses TabularTranslator to convert data types during the copy.  
- Uses parameterized datasets to dynamically adjust source and destination paths and file names.

## Key Features

- Fully dynamic pipeline with parameterized datasets.  
- Efficient metadata-driven design allows flexible scaling.  
- Simplifies ingestion of web-based data into a structured lakehouse Bronze layer.  
- Type conversion and format control built into the transformation pipeline.