

# Azure Data Factory

Cloud-based data workflows and orchestration

Presented by Brian Nolan and Sreenivas Nalla



# Overview

- What is an Azure Data Factory?
- Use Cases
- Case Study - Data Migration
- Advantages
- Demonstration
- References
- Questions
- Contact Information



# What is an Azure Data Factory?

- **Microsoft Azure cloud-based service**
- **Tool to create, orchestrate, and manage data movement and enrichment**
- **Pull, extract, transform, analyze, and push data**
  - Extract-transform-load (ETL)
  - Extract-load-transform (ELT)
- **Integrates on-premise and cloud environments**
  - Databases (SQL and NoSQL)
  - Files
  - APIs
  - Services/Apps
- **Integrates with other Azure Services**
  - Function Apps
  - Logic Apps
  - Machine Learning
  - Stream Analytics





## Use Cases

- Data migrations
- Sync between data sources
- Data analysis
- Data transformation

# Case Study

## Data Migration



- **General Workflow**
  - Orchestrated via Azure Data Factory
    - Extract source content and push to staging database
    - Execute data transformation logic
    - Push transformed data to destination system
  - All version controlled via git
  - Repeatable
- **Source** - Oracle | **Destination** - Salesforce
  - 20 GB data
  - 551 tables / 5453 fields
- **Source** - Salesforce | **Destination** - SQL Server
  - 30 GB data
  - 26 tables / 1504 fields



# Advantages

- **Connectors to 100+ services, databases, and APIs already exist**
  - Oracle, Amazon S3, Salesforce, Twilio
  - Create your own
- **Work with data located almost anywhere**
  - On-prem
  - Cloud
- **Pay for what you use**
  - Building data flows are free
  - Pay for execution
- **Version control the factory process**
  - Built for git
  - Easily generate templates or clone



# Advantages

- **Monitor “factory” execution process**
  - Execution generated notifications
  - Logs
- **Security and privacy focused**
  - Credentials encrypted
  - No data caching
- **Multiple development options**
  - Visual Studio
  - Data Factory website
  - PowerShell
  - Auto-generate via .NET library

# Demo

- Data Migration to Salesforce
- Integration with Logic App
- Automation for Salesforce Backup
- Automated Notifications





What is an Azure Data Factory Pipeline?

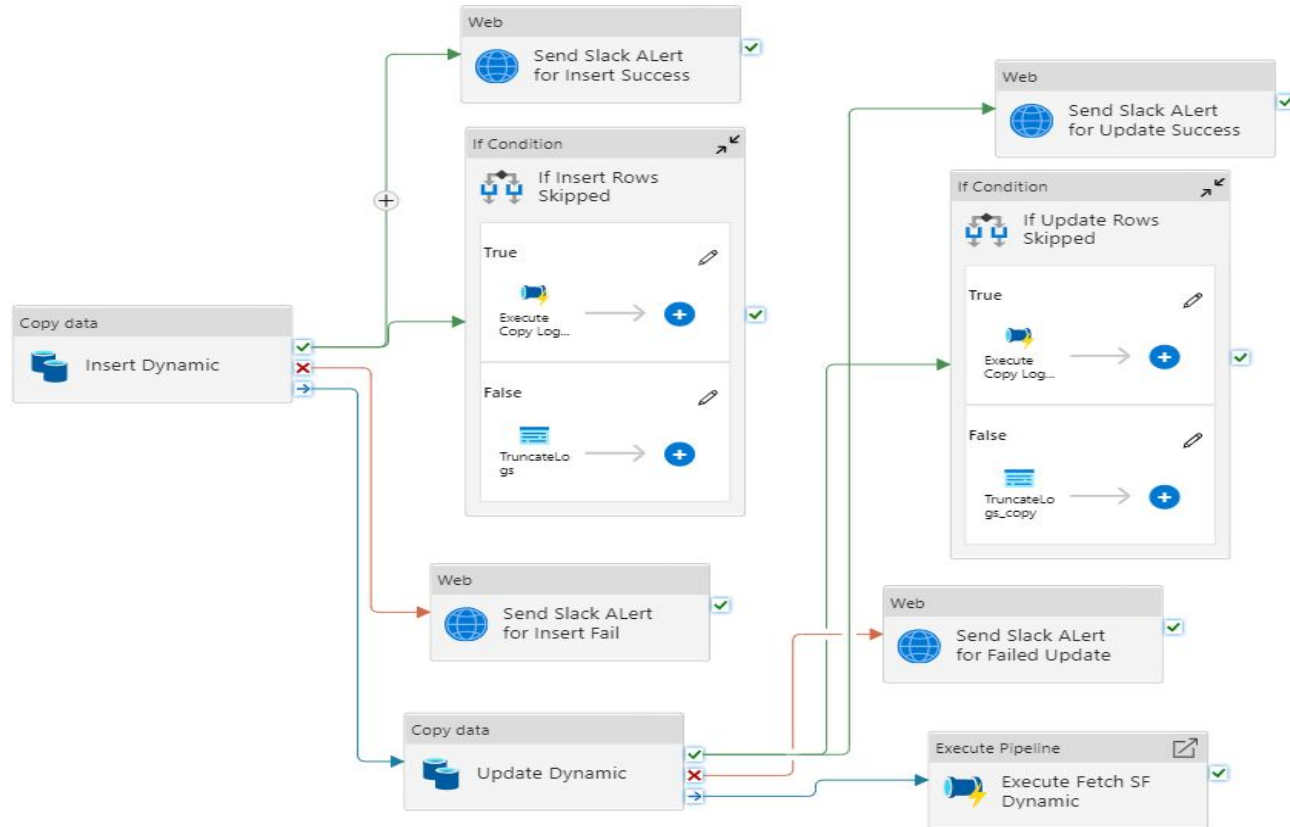
# A set of data-driven workflows

*With a data factory, you can create pipelines that ingest data from disparate data stores and use them to manage activities as a set.*

***Together, the activities in a pipeline perform a task.***



# Data Migration to Salesforce



# Integration with Logic App

The screenshot displays the Microsoft Azure Logic App interface. At the top, the navigation bar shows "Microsoft Azure" and a search bar. Below it, the breadcrumb path is "Home > Logic apps > SLACK-ADNotification >". The main content area shows a list of six successful data sync actions, each represented by a "1\_OracletoAzure\_SYNC" connector icon. The actions are as follows:

- 1\_OracletoAzure\_SYNC: SUCCESS: Rows AMTRUST\_USER\_Y\_GROUP\_RULE Read: 28 Rows Copied: 28 Rows Skipped: 0 @Sreeni 🌟
- 1\_OracletoAzure\_SYNC: SUCCESS: Rows [redacted]\_FRC\$ Read: 0 Rows Copied: 0 Rows Skipped: 0 @Sreeni 🌟
- 1\_OracletoAzure\_SYNC: SUCCESS: Rows [redacted]\_LINE\_ITEM\_NON\_US\_TAX\_CODE Read: 0 Rows Copied: 0 Rows Skipped: 0 @Sreeni 🌟
- 1\_OracletoAzure\_SYNC: SUCCESS: Rows [redacted]\_USER\_E\_FIR\$ Read: 0 Rows Copied: 0 Rows Skipped: 0 @Sreeni 🌟
- 1\_OracletoAzure\_SYNC: SUCCESS: Rows [redacted]\_DTYM Read: 103 Rows Copied: 103 Rows Skipped: 0 @Sreeni 🌟
- 1\_OracletoAzure\_SYNC: SUCCESS: Rows [redacted]\_J\_DMS\_FOLDER Read: 0 Rows Copied: 0 Rows Skipped: 0 @Sreeni 🌟
- 1\_OracletoAzure\_SYNC: SUCCESS: Rows [redacted]\_CONT\_ADDRESS Read: 138052 Rows Copied: 138052 Rows Skipped: 0 @Sreeni 🌟

At the bottom of the interface, there are two empty action slots, each with an "Add an action" button.

# Automation for Salesforce Backup

The screenshot displays a workflow automation interface. On the left, a 'Lookup' activity labeled 'LIST OF Objects' is connected to a 'ForEach' loop. The 'ForEach' loop is titled 'BACKUP Objects' and contains an 'Activities' section with a 'Copy Data from...' activity. Below the workflow canvas are tabs for 'Parameters', 'Variables', 'Settings', and 'Output', with a '+ New' button under 'Parameters'. On the right, a 'new trigger' configuration panel is shown. It includes the following fields and options:

- Name \***: Daily
- Description**: (empty text area)
- Type \***: Schedule
- Start date \***: 1/25/2023, 6:32:28 PM
- Time zone \***: Coordinated Universal Time (UTC)
- Recurrence \***: Every 1 Day(s)
- Advanced recurrence options**:
  - Execute at these times**: Hours 22, Minutes (empty)
  - Schedule execution times**: 22:32
  - Specify an end date
  - End On \***: 3/31/2023, 6:32:28 PM
- Annotations**: + New
- Start trigger**: (empty)

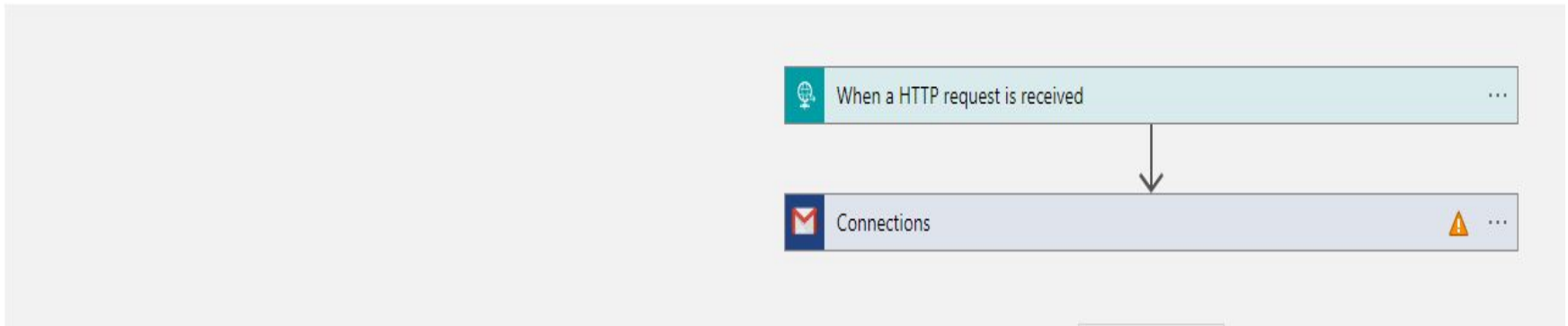
At the bottom of the configuration panel are 'OK' and 'Cancel' buttons.



# Automated Notifications

Logic Apps Designer ...

Save Discard Run Trigger Designer Code view Parameters Templates Connectors Help Info



**Coming up  
Next Time**



## **Diving deeper into Azure Data Factories**

- Azure Data Factory development environment
- Creating pipelines to move data

# References

- **Introduction to Azure Data Factory**  
<https://learn.microsoft.com/en-us/azure/data-factory/introduction>
- **Azure Data Factory Connectors**  
<https://learn.microsoft.com/en-us/azure/data-factory/connector-overview>
- **Source Control Azure Data Factory**  
<https://learn.microsoft.com/en-us/azure/data-factory/source-control>
- **Azure Portal**  
<https://portal.azure.com/#home>

**Questions?**



**DOWNLOAD THE PRESENTATION**







# Contact Information

[www.gobrio.com](http://www.gobrio.com)

**Brian Nolan**

[bnolan@gobrio.com](mailto:bnolan@gobrio.com)

**Sreenivas Nalla**

[snalla@gobrio.com](mailto:snalla@gobrio.com)

