European
Power Platform
Conference

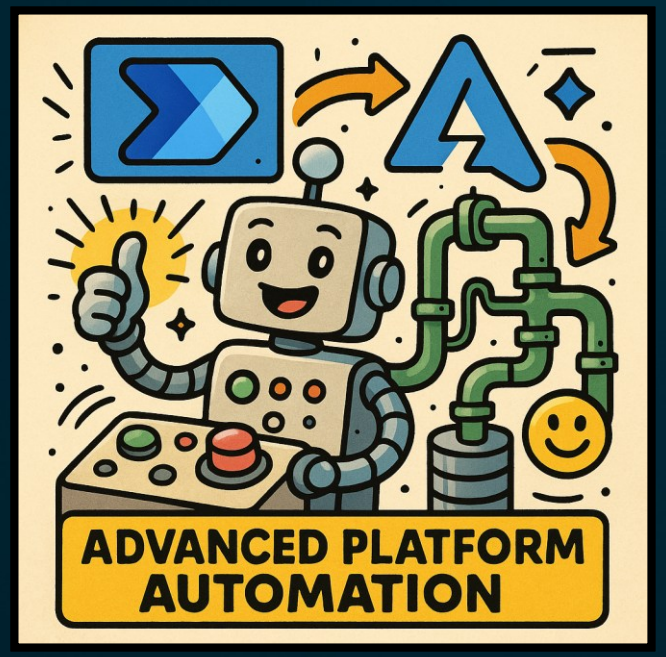
PowerUp!

WELCOME
TO

VIENNA 2025

JOIN THE
CONVERSATION

#EPPC25

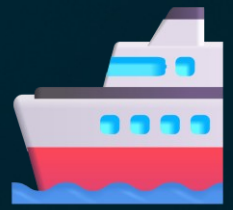


Lab 4 – Azure Data Factory

Advanced Integrations Workshop @ EPPC

What ever you dream of – in Functions you can build it!

Azure Data Factory



Code-Free ETL as a service

Ingest

- Multi-cloud and on-premise hybrid copy data
- 100+ native connectors
- Serverless and auto-scale
- Use wizard for quick copy jobs

Control Flow

- Design code-free data pipelines
- Generate pipelines via SDK
- Utilize workflow constructs: loops, branches, conditional execution, variables, parameters, …

Data Flow

- Code-free data transformations that execute in Spark
- Scale-out with Azure Integration Runtimes
- Generate data flows via SDK
- Designers for data engineers and data analysts

Schedule

- Build and maintain operational schedules for your data pipelines
- Wall clock, event-based, tumbling windows, chained

Monitor

- View active executions and pipeline history
- Detail activity and data flow executions
- Establish alerts and notifications

WHAT IS AZURE DATA FACTORY

READ THE DOCS

A CLOUD-BASED DATA INTEGRATION SERVICE THAT ORCHESTRATES DATA MOVEMENT & TRANSFORMATION BETWEEN DIVERSE DATA SOURCES ⟷ CLOUD COMPUTE RESOURCES AT SCALE

DATA INTEGRATION CHALLENGE

WHAT PROBLEM DOES IT SOLVE?

BIG DATA — DIVERSE DATA SOURCES, DURATIONS — STORED DATA, CLEANED & ENRICHED — ACTIONABLE INSIGHTS

DATABASES, FILES, DATA STREAMS, DEVICES → DATA INGEST → DATA TRANSFORMATION → DASHBOARDS, REPORTS

DATA OPS TEAM — DATA MOVEMENT — DATA ANALYSTS

HELP! We need to collect all this data!

WE NEED DATA WITH RELEVANT CONTEXT TO DO OUR ANALYSIS

HOW CAN WE ORCHESTRATE (DATA MOVEMENT) AND OPERATIONALIZE (WORKFLOW)

AZURE DATA FACTORY

- Raw, unorganized data stored in relational, non-relational, and other storage systems
- We need actionable business insights but raw data lacks the correlating context (by itself) for analysis
- Big data needs a scalable service than can ORCHESTRATE the data movement and OPERATIONALIZE the data processing workflows!

THE SCENARIO YOU RECOGNIZE

GAME DEV COMPANY

- HAVE: Petabytes of game log data in the cloud
- WANT: Actionable insight into my customers
  - gaming preferences
  - player demographics
  - user behaviors
- USES
  - develop new features
  - improve player experience
  - upsell, cross-sell and drive business growth.

DATA ANALYST

WHAT I NEED..

- reference data from our on-prem services
  - customer info
  - marketing campaign
  - game metadata
- * combined with gameplay data logs from cloud
  - game play behaviors
  - success, failure rates etc.

WHAT I MUST DO

- collect data from sources
- process data to transform it
- publish data into warehouse for analysts.

AUTOMATE THIS WORKFLOW (TRIGGER IT BY AN EVENT OR MANUALLY — MONITOR & MANAGE DAILY SCHEDULES)

HOW DOES ADF WORK?

CAN YOU TELL ME WHAT ADF DOES AT A HIGH LEVEL?

LET'S BREAK IT DOWN

- AZURE DATA FACTORY IS A SET OF INTERCONNECTED SYSTEMS...
- THAT GIVE YOU AN END-TO-END PLATFORM...
- FOR YOUR DATA ENGINEERING!

DESIGN WIZARD OR CODE + COMPUTE

CONNECT & COLLECT

CONNECTORS

1. Build an information production system to connect all data sources and adapt ingest to their diverse intervals and speeds!
2. Collect all data in a centralized location to facilitate processing (e.g. transformation) next

COPY ACTIVITY

TRANSFORM & ENRICH

3. Transform the collected data using actions to aggregate, filter, clean etc.

Use code-free UI based mapping data transformation graph creation! OR use compute to transform data by hand.

MAPPING DATA FLOW

CI/CD & PUBLISH

4. Manage data pipeline ops using Azure DevOps and GitHub! Incrementally develop and deploy your E-T-L process then publish processed data into Azure Data Warehouse, Azure SQL Database, Azure Cosmos DB – or your BI analytics engine!

PIPELINE OPS

MONITOR & ALERT

5. Once your data pipeline is created and deployed, monitor it for the success/failure rates!

Use Azure Monitor, API, PowerShell, health panels on Azure Portal!

AZURE MONITOR

LET'S TALK DATA INTEGRATION PATTERNS

The most common pattern is ETL

- EXTRACT = connect to sources & copy data to central store
- TRANSFORM = process the data for analysis & data transformation at scale
- LOAD = move the data to data warehouse or analytics engines for business insights

Another pattern is ELT where raw (native) data is itself loaded (stored) before the transformation phase....

AZURE DATA FACTORY HAS CODE-FREE ETL AS A SERVICE!

You always have the option to do hand-coded transformation using Azure compute... but mapping data flows provide a UI-based wizard to simplify your pipeline setup..

5 ASPECTS IT COVERS!

1. INGEST DATA
   - * Copy data on-prem or in cloud!
   - * 90+ native connectors
   - * Serverless, auto-scaling!
2. CONTROL FLOW
   - * Design (via UI) or Create (via SDK) your data pipelines
   - * Utilize workflow constructs (loops, parameters, variables...) for efficiency
3. DATA FLOW
   - * Code-free data transforms (execute automagically in Spark)
   - * Scale out with Azure Integration Runtimes, generate data flows with SDK (code) or Designer (UI)
4. SCHEDULE OPS
   - * Build/Maintain ops schedules for your data pipelines
   - * Options include: Wall clock, Event based, Tumbling windows, chaining
5. MONITOR OPS
   - * View active executions and pipeline history – details at activity/pipeline
   - * Establish alerts for key events or progress notifications

7 THINGS TO KNOW ABOUT AZURE DATA FACTORY

1. ENTERPRISE READY — Data Integration At cloud scale!
2. ENTERPRISE DATA READY — 90+ connectors! It just works.
3. CODE-FREE TRANSFORMATION — UI driven mapping data flows
4. RUN CODE ON ANY AZURE COMPUTE — For hands-on data transformations
5. MANY SSIS PACKAGES RUN ON AZURE (Reboot on-prem in ADF in 3 steps)
6. ADF CAN MAKE DATA OPS SEAMLESS — source control, automated deploy, simple templates
7. SECURE DATA INTEGRATION — Managed virtual networks protect against data exfiltration, simplify your networking!

6 INTEGRATION RUNTIME (IR)

INTEGRATED RUNTIME (IR) is COMPUTE INFRASTRUCTURE USED BY ADF TO PROVIDE FULLY MANAGED

- DATA FLOWS (transformation)
- DATA MOVEMENT (movement)
- ACTIVITY DISPATCH (route to service)
- SSIS PACKAGE EXECUTION (in managed compute)

3 IR TYPES

IR Types | Capability | N/W

- AZURE-SSIS → SSIS package execution (public or private) | N/A
- SELF HOSTED → Activity Dispatch + Data Movement | public or private
- AZURE → Activity Dispatch + Data Movement + Data Flow | public or private

POTENTIAL TO HAVE HYBRID DATA INTEGRATION SOLUTIONS

- SELF-HOSTED
- AZURE-SSIS
- AZURE

PICK WHAT WORKS FOR YOUR DATA NEEDS ... AND YOUR NETWORKING REQUIREMENTS!

5 ADDITIONAL CONCEPTS

1 CONTROL FLOW

REFERS TO THE ORCHESTRATION OF PIPELINE ACTIVITIES INCLUDING:

- CHAINING ACTIVITIES
- BRANCHING ACTIVITIES
- CUSTOM STATE PASSING
- DEFINING PARAMETERS
- PASSING ARGUMENTS (for pipeline run)
- LOOPING CONTAINERS

Think of ADF as the conductor for an orchestra of activities! KNOWS THE BIG PICTURE

2 PIPELINE RUN

AN INSTANCE OF PIPELINE EXECUTION

(arguments) A1 A2 A3 (parameters)

INSTANTIATE PIPELINE BY PASSING ARGUMENTS (values) TO PARAMETERS (placeholders) DEFINED IN IT.

PASS ARGUMENTS manually OR BY DEFINING triggers

3 TRIGGERS

TRIGGER EVENT — DIFFERENT TRIGGER TYPES FOR VARIOUS EVENTS!

UNIT OF PROCESSING THAT DETERMINES WHEN TO KICK OFF PIPELINE RUNS.

4 PARAMETERS

"READ-ONLY" KEY/VALUE PAIRS POPULATED FROM RUN CONTEXT AT EXECUTION

BOTH 'DATASET' AND 'LINKED SERVICE' ARE STRONGLY-TYPED, REUSABLE, REFERENCEABLE PARAMETER ENTITIES!!

ACTIVITY REFERENCES DATASET TO CONSUME DATA PROPERTIES, AND 'LINKED SERVICE' FOR THE CONNECTION INFO (to data store for copy, compute resource for flow)

5 VARIABLES

ARE USED INSIDE PIPELINES TO STORE TEMPORARY VALUES, STATE

USED WITH PARAMETERS TO PASS VALUES BETWEEN ACTIVITIES, DATA FLOWS, PIPELINES.

2 ACTIVITIES

REPRESENT A SINGLE PROCESSING STEP IN THE PIPELINE e.g.

- COPY ACTIVITY — copy data from one store (source) to other (destination)
- HIVE ACTIVITY — run a Hive query on an HDInsight cluster

THREE TYPES OF ACTIVITY SUPPORTED BY ADF

- Data Movement
- Data Transformation
- Control

1 PIPELINES

A1 A2 A3 — PIPELINE STAGE

- A PIPELINE IS A LOGICAL GROUPING OF ACTIVITIES THAT PERFORM ONE UNIT OF WORK
- AN ADF INSTANCE CAN HAVE ONE OR MORE PIPELINES!

Example: Pipeline that ingests data from Azure Blob (A1), runs Hive query on HDInsight to partition it (A2) and moves result to next store (A3)

Benefit: Manage correlated activities as a single entity!

Execution: Activities may be chained (run sequentially) or be independent (run in parallel) within pipeline!

3 DATASETS

- TWO USES
  - REPRESENT DATA SOURCE
  - REPRESENT COMPUTE RESOURCE
- REPRESENT CONNECTION INFORMATION TO LINK ADF TO EXTERNAL SERVICES
- WORKS CLOSELY WITH DATASETS. THIS SPECIFIES HOW TO CONNECT & DATASET DEFINES WHAT DATA LOOKS LIKE...

data sources for INGEST

compute resources for TRANSFORM

DATA STORE — DATA SET (view)

- REPRESENTS DATA STRUCTURE GIVING SELECT VIEW INTO DATA STORE
- POINTS TO THE SPECIFIC DATA SUBSET TO USE IN ACTIVITY...
- .. FOR INPUT & OUTPUT

5 MAPPING DATA FLOWS

ADF EXECUTES LOGIC ON SPARK CLUSTER FOR YOU (automagically!) & NO NEED TO MANAGE OR MAINTAIN YOUR OWN CLUSTER!

- CREATE & MANAGE DATA TRANSFORMATION GRAPHS THAT WORK ON ANY SIZE DATA
- BUILD UP REUSABLE LIBRARY OF DATA TRANSFORMATION ROUTINES!
- EXECUTE TRANSFORM ROUTINES IN PIPELINES (AS ACTIVITY) TO SCALE!

HEY PRESTO!

4 LINKED SERVICES

HERE ARE THE KEY TERMS AND CONCEPTS YOU NEED TO KNOW TO USE ADF!

AND A FEW KEY TERMS

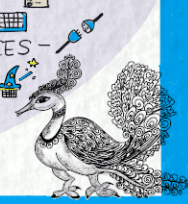
- CONTROL FLOW
- PIPELINE RUN
- TRIGGERS
- PARAMETERS
- VARIABLES

ADF TOOLKIT

LET'S TALK ABOUT COMPONENTS OF AZURE DATA FACTORY

6 MAIN CONCEPTS

- PIPELINES
- ACTIVITIES
- DATASETS
- LINKED SERVICES
- DATA FLOWS
- INTEGRATION RUNTIMES



ADF Concepts

- Pipelines
- Activities
- Datasets
- Linked services
- Data Flows
- Integration Runtimes

Azure Data Factory: Use Cases

- Automate data migration from on-premises to cloud.
- Create ETL workflows for data transformation.
- Integrate data from diverse sources for analytics.

Demo: Azure Data Factory

- SaaS
- 100s customers each with individual database
- Each customer has users (userid, name, email)

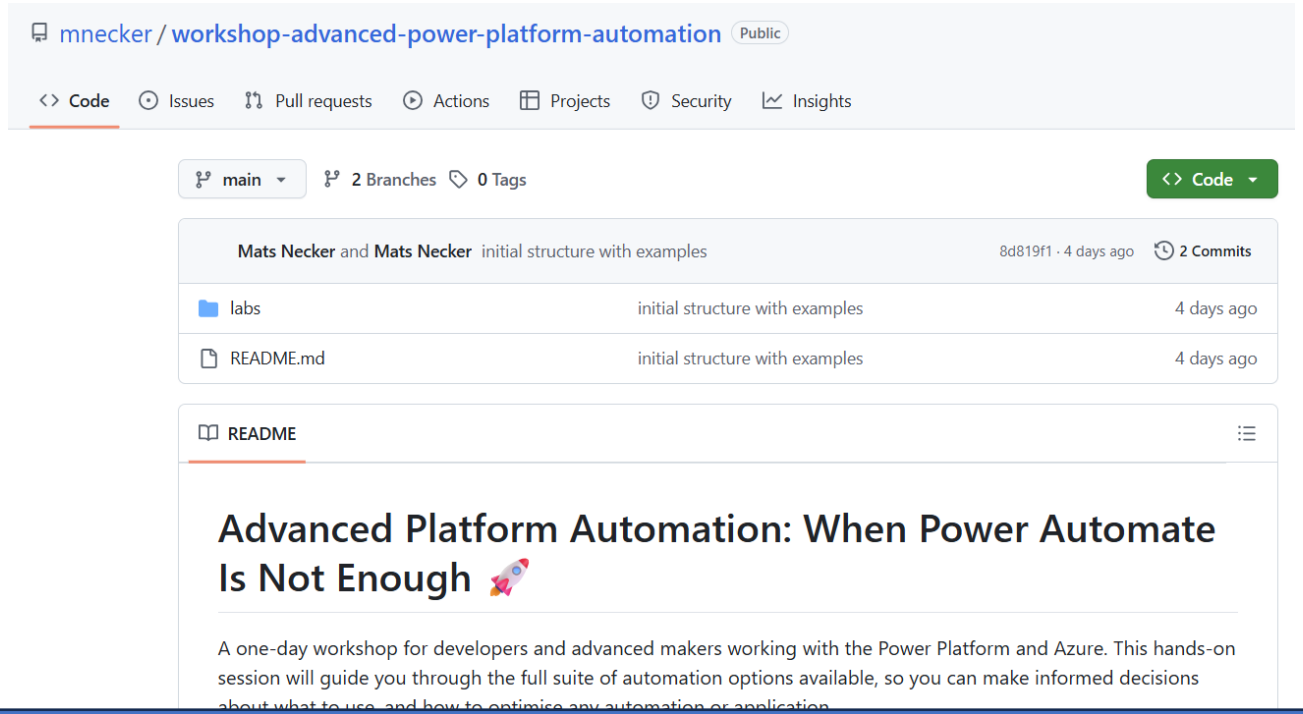
- Business case: implement discovery service and single sign on
- Question: how many users registered with multiple customers do we have

Creating the demo

- Create a demo is an automation in itself
  - Use AI to help
    - You should be comfortable with reading code + errors and have some basic understanding of what's going on.
    - Rise of expert beginner
  - Beware of hallucinations
    - Prompt "how to execute sql command for azure sql database using az cli"

Let‘s build it!

The Lab Material



https://github.com/mnecker/workshop-advanced-power-platform-automation

CONTACT US

George Doubinski

Mats Necker