

# Azure Data Factory: Data Flows – first blood

Kamil Nowiński

# About me

Kamil Nowinski



**Microsoft**  
**CERTIFIED**  
Solutions Associate  
SQL Server 2012



Microsoft Data Platform **MVP**

Speaker, blogger, data enthusiast

Data Engineer at ASOS ([www.asos.com](http://www.asos.com))

13+ yrs experience as DEV/DBA

Member of the Data Community PL

Project member of „SCD Merge Wizard”

Founder of blog SQLPlayer ([www.SQLplayer.net](http://www.SQLplayer.net))

SQL Server Certificates:

MCITP, MCP, MCTS, MCSA, MCSE Data Platform,

MCSE Data Management & Analytics

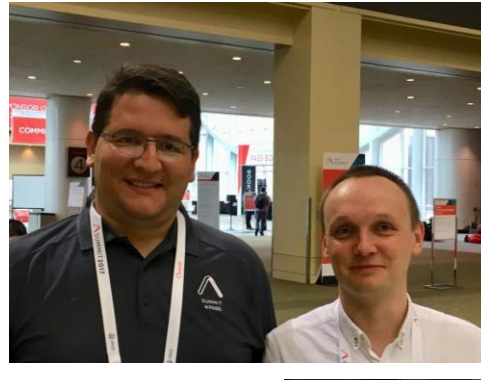
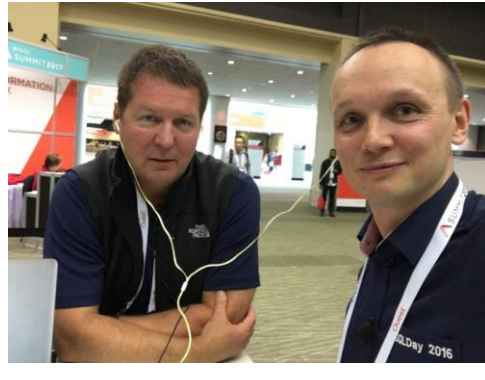
Moreover: Bicycle, Running, Digital photography

@NowinskiK, @SQLPlayer

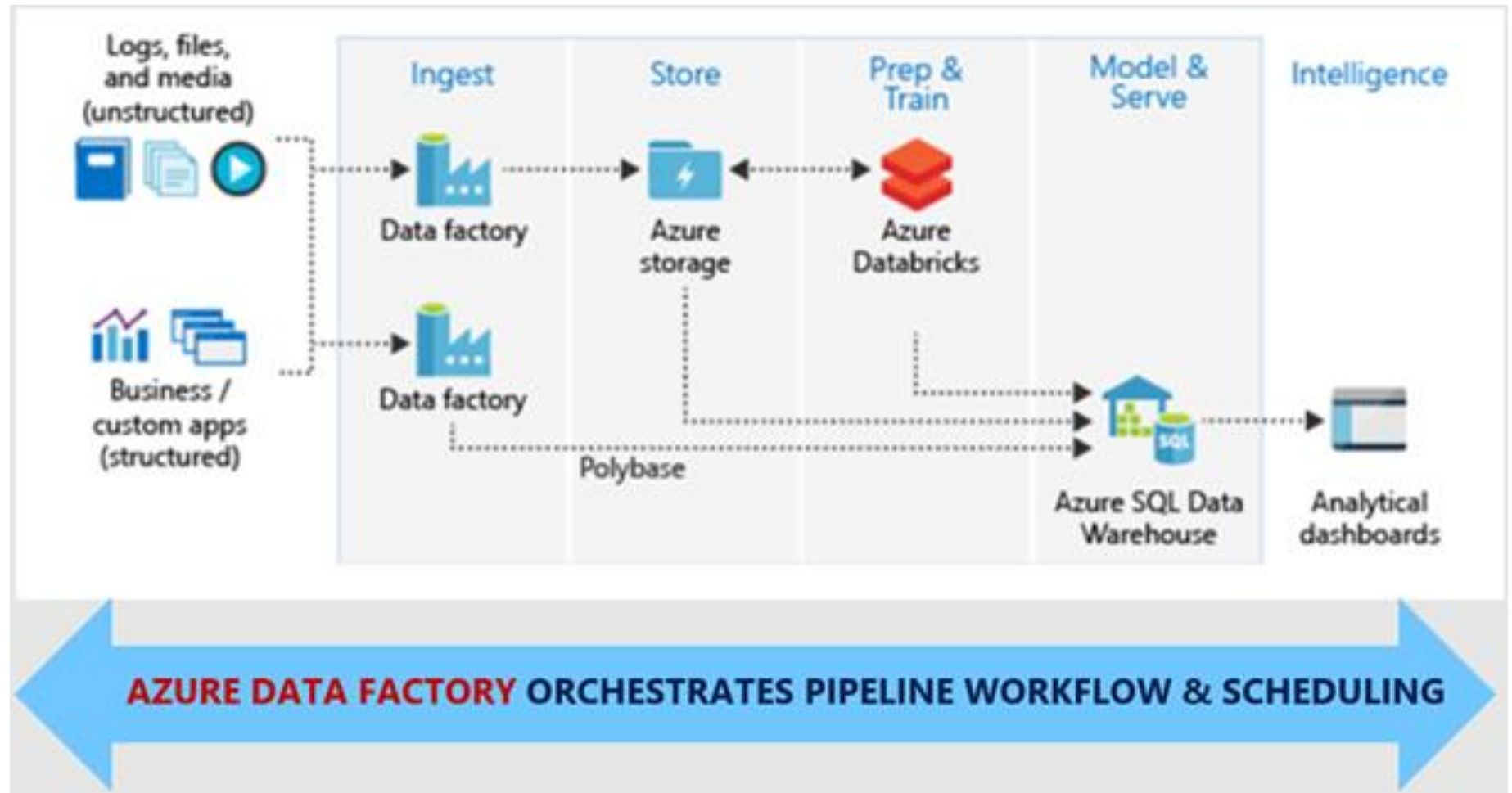




# PODCAST – Interviews with...



# What the Azure Data Factory is?



# How to create Azure Data Factory with Data Flow?

Home > Data factories > New data factory

Data factories

ASOS.com Ltd

+ Add

≡ Edit columns

⋮ More

Filter by name...

NAME

↑↓

bigfactory555

BigFactoryCDM

BigFactoryDF

BigFactoryPP

BigPlayer

New data factory

\* Name ⓘ

MyDataFactory666 ✓

\* Subscription

Visual Studio Enterprise ▼

\* Resource Group ⓘ

☐ Create new

☒ Use existing

rg-datafactory ▼

Version ⓘ

V2 with data flow (preview) ▼

\* Location ⓘ

Southeast Asia ▼

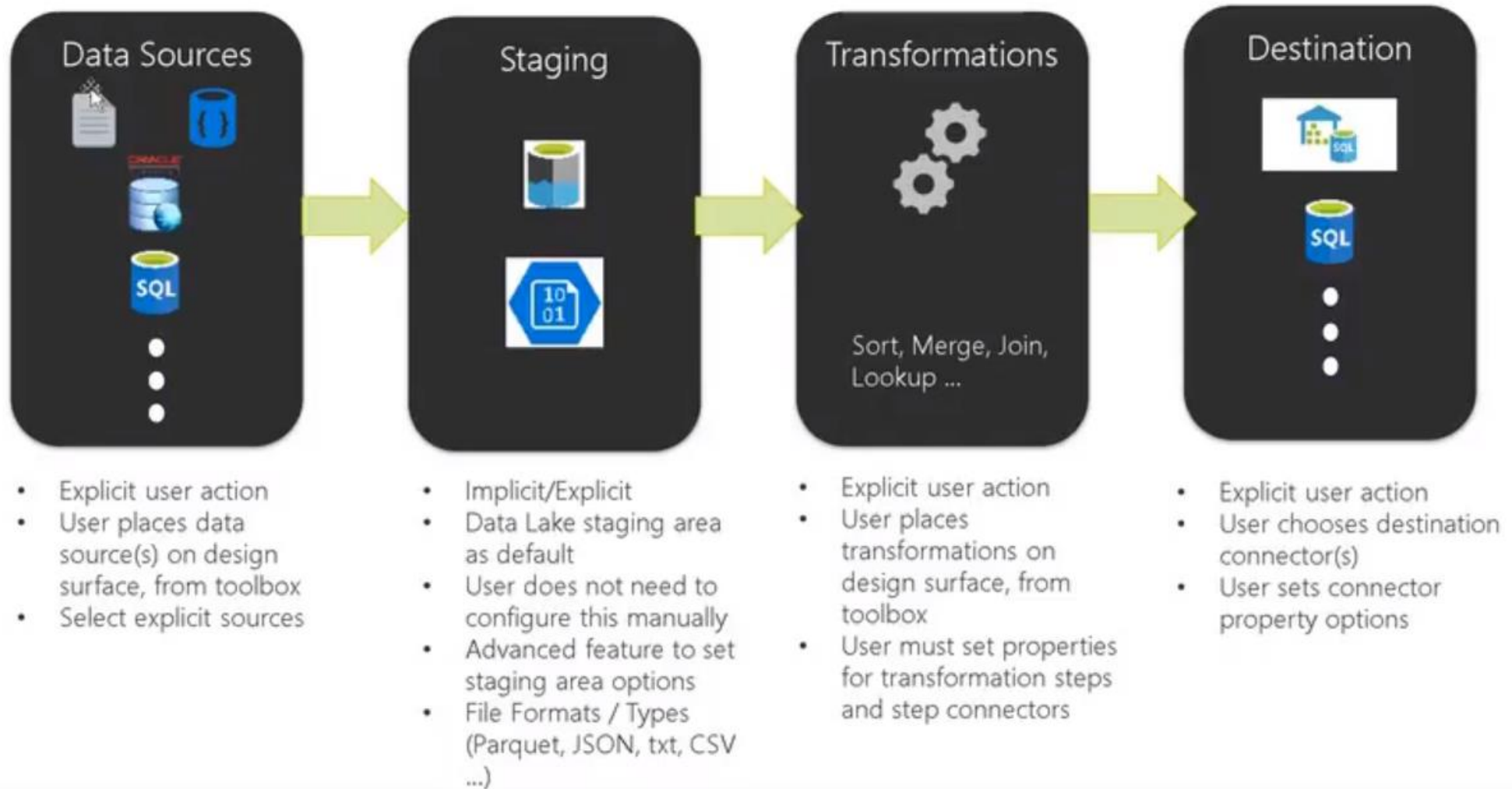
Include data flow sample

☐

ADF Data Flow is currently in private preview. The normal Azure SLAs do not apply to use of this preview feature and all support must route through this email address:  
adfdatabflowext@microsoft.com



# What the hell Data Flows are?



# Authoring of Azure Data Factory (v2)

Microsoft Azure

Search resources

BigPlayer Data Factory Publish All Validate All Refresh Discard All ARM Template

Factory Resources Filter Resources

- Pipelines 2
- Datasets 9
  - Badges
  - BadgesBlob
  - BadgesBlobWithHeader
  - BadgesStatsByName
  - BadgesStatsByNameBlob
  - Crimes\_BlobCsv
  - Src\_Users
  - Users\_BlobCsv
  - UsersTest
- Data Flows 3
  - StackOverflow
    - badgesGroupByName
    - badgesGroupByName2
    - users

users

Debug Validate

sourceUsers Import data from Users\_BlobCsv

Select1 Renaming sourceUsers to Select1 with columns 'DisplayName', 'DownVotes', 'LastAccessDate', 'Location', 'Reputation'

FilterByReputation Filtering rows using expressions on columns 'Reputation'

GroupByLocation Aggregating data by 'Location' producing columns 'SumOfReputation', 'SumOfViews', 'Count'

SortByLocation Sorting rows on columns 'Location'

Wrong Conditionally distributing the data in 2 groups, based on columns 'Location', 'Location', 'Location', 'Location', 'Location'

AllRight Conditionally distributing the data in 2 groups, based on columns 'Location', 'Location', 'Location', 'Location', 'Location'






General External dependencies

Name \* users




Description

# Data Flow: Components = Actions \*





## Multiple inputs/outputs

-  New Branch
-  Join
-  Conditional Split
-  Union
-  Lookup

## Schema modifier

-  Derived Column
-  Aggregate
-  Surrogate Key


## Row modifier

-  Exists
-  Select
-  Filter
-  Sort

## Custom

-  Extend

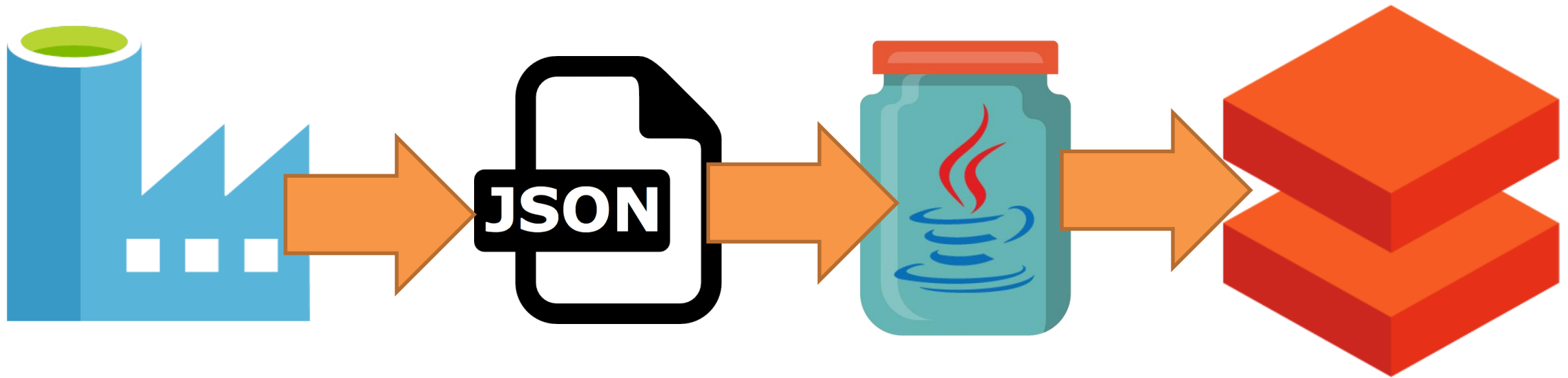
## Destination

-  Sink

\* With some small exceptions



# What is going on behind the scenes?



JAR

Azure  
Databricks

DEMO TIME





# Thank you



[kamil@nowinski.net](mailto:kamil@nowinski.net)



@NowinskiK

@SQLPlayer



SQLPlayer.net

## Kamil Nowinski

Microsoft Data Platform MVP

MCSE Data Platform & MCSE Data Management and Analytics