Azure Data Factory:
Data Flows – first blood

Kamil Nowiński

About me

Kamil Nowinski

Microsoft Certified
Solutions Associate
SQL Server 2012

MCSE Data Platform

MVP Microsoft® Most Valuable Professional

Microsoft Data Platform **MVP**

Speaker, blogger, data enthusiast

Data Engineer at ASOS (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)

SQL Server Certificates:

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

MCSE Data Management & Analytics

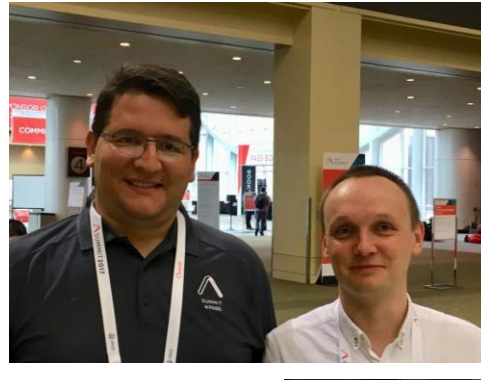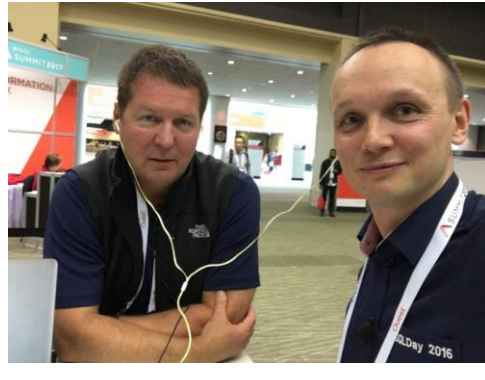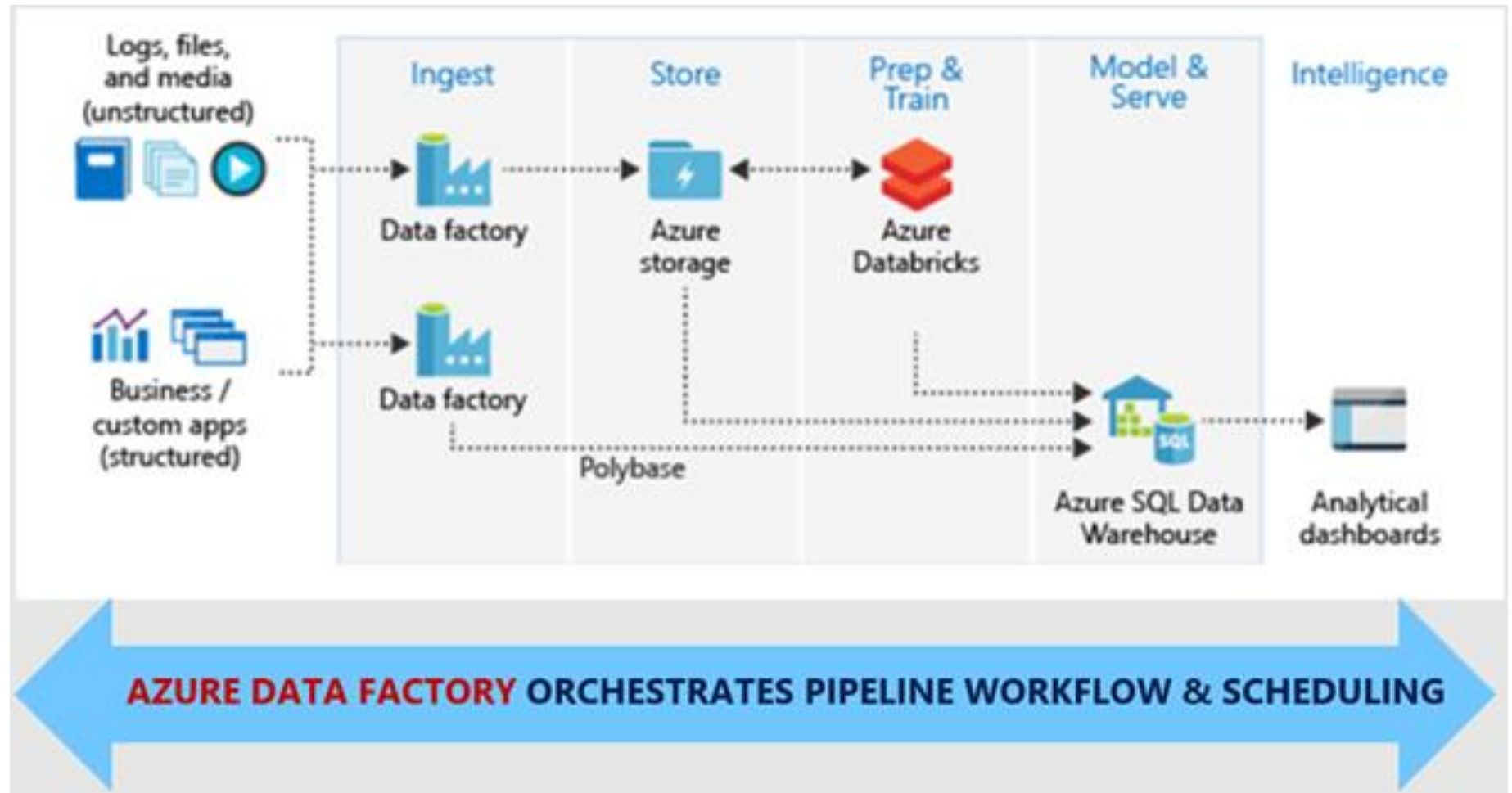Moreover: Bicycle, Running, Digital photography

@NowinskiK, @SQLPlayer

redgate FRIENDS

SQL Player
Play with data & have fun!

PODCAST – Interviews with…

What the Azure Data Factory is?

How to create Azure Data Factory with Data Flow?

What the hell Data Flows are?

Data Sources
- Explicit user action
- User places data source(s) on design surface, from toolbox
- Select explicit sources

Staging
- Implicit/Explicit
- Data Lake staging area as default
- User does not need to configure this manually
- Advanced feature to set staging area options
- File Formats / Types (Parquet, JSON, txt, CSV ...)

Transformations

Sort, Merge, Join, Lookup ...
- Explicit user action
- User places transformations on design surface, from toolbox
- User must set properties for transformation steps and step connectors

Destination
- Explicit user action
- User chooses destination connector(s)
- User sets connector property options

SQL Player
Play with data & have fun!

Authoring of Azure Data Factory (v2)

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
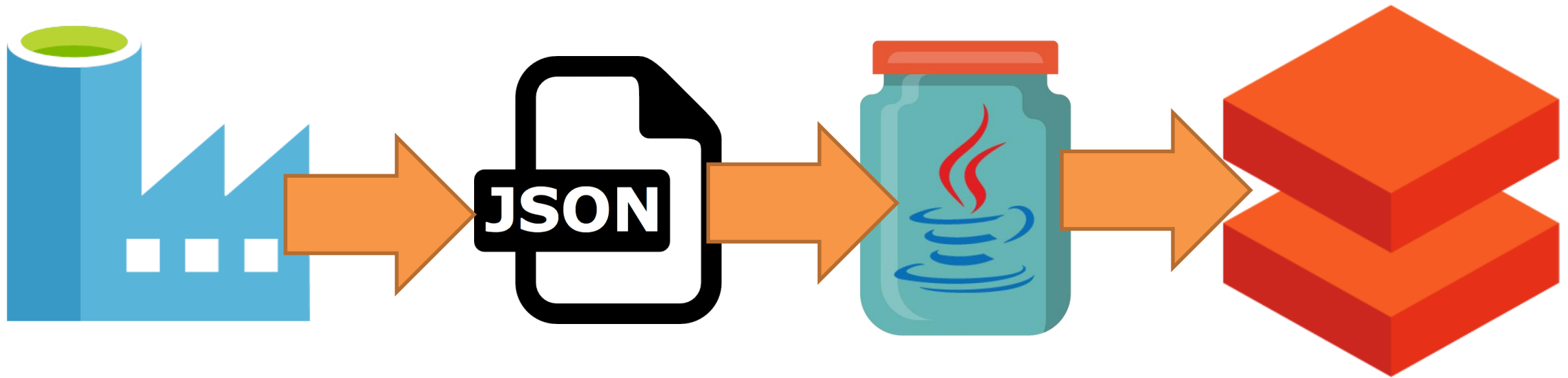
- Sink

* With some small exceptions

SQL Player
Play with data & have fun!

What is going on behind the scenes?

JAR

Azure
Databricks

SQL Player
Play with data & have fun!

DEMO TIME

SQL Player
Play with data & have fun!

Thank you

✉ kamil@nowinski.net

🐦 @NowinskiK   @SQLPlayer

🔗 SQLPlayer.net

Kamil Nowinski

Microsoft Data Platform MVP
MCSE Data Platform & MCSE Data Management and Analytics

SQL Player
Play with data & have fun!