

Streaming and Replicating Db2 for i Data Into “The Cloud”

John Westcott

Db2 for i Analytics Consultant

JWestcott@ibm.com

Db2 Web Query Team email: QU2@us.ibm.com

IBM Systems Expert Labs
Db2 Web Query for i

...extract data from files, possibly transform it and then load to some other destination perhaps to MS SQL Server, MySQL, Oracle, CSV, etc.

...automate the consolidation of data from multiple servers, including some data from non-IBM i sources such as Microsoft SQL Server, Geospatial Coordinates, External Weather API, etc.

...move data bi-directionally between IBM i partition?

...automate a repeatable process get data into (fill in the blank) cloud service?

...process a regularly received spreadsheet, CSV, Text, JSON, XML, etc. file and automatically do data clean up and then move the data into Db2 for i.

...create a Db2 for i data warehouse, data mart, ODS (operational data store) for improved reporting, business intelligence, insights, etc.

...do a bunch of pre-processing on IBM i data before it's fed it into an AI/ML service

...send data near real-time to another library, table, non Db2 for i database, a Kafka stream, Cloud database, Cloud Service, etc.

And a billion more need to....

Essential Need...

Our “mission statement”

An automated and repeatable process
optionally transforming or cleaning data
to motion data from somewhere to somewhere

Our expanded “mission statement”

An automated and repeatable process

- ✓ On an Interval
- ✓ Near Real-Time

optionally transforming or cleaning data

- ✓ Dates, “Smart Fields”, etc.
- ✓ Split or combine files
- ✓ Aggregate data
- ✓ Derived and calculated values

to motion data from somewhere to somewhere

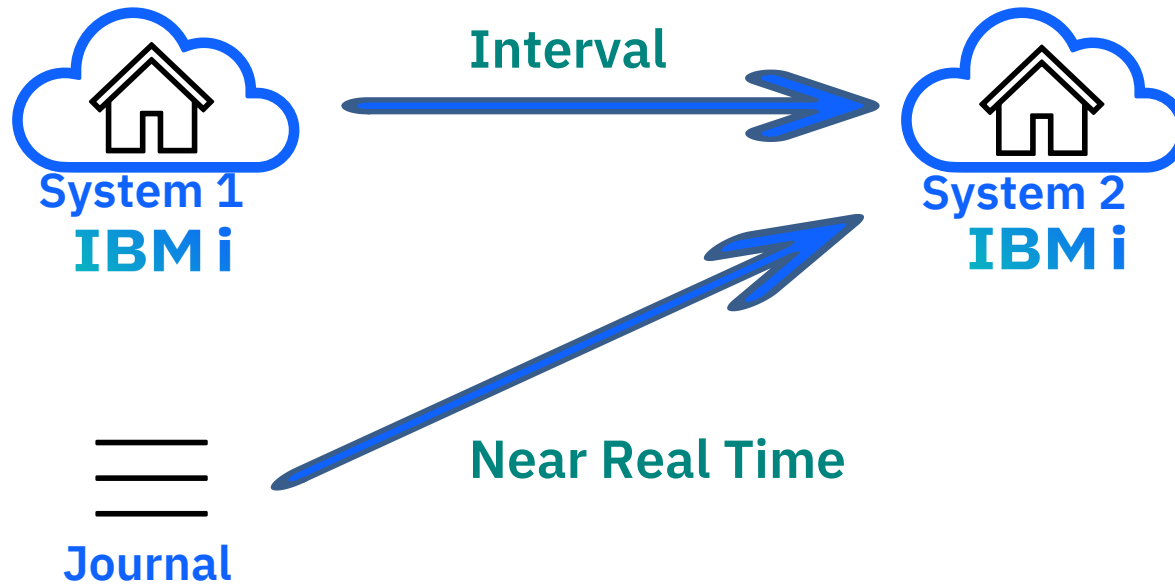
- ✓ From one or more sources to one or more targets
- ✓ IBM i and non-IBM i

Motioning Data – The Concepts And Enhancing And Untangling

Building and Participating in a Data Pipeline

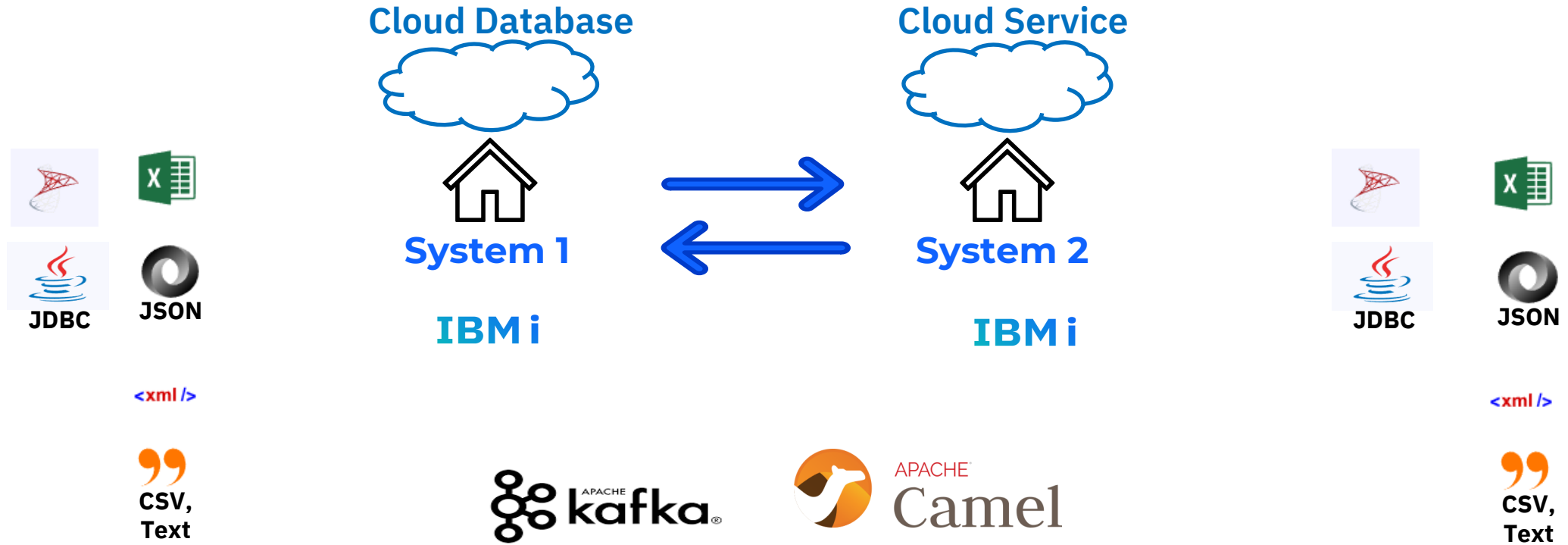
Motioning Data – Some Basic Cases

~~Pretty easy, no worries~~ *Yuck, reading journals and all the journaling stuff*



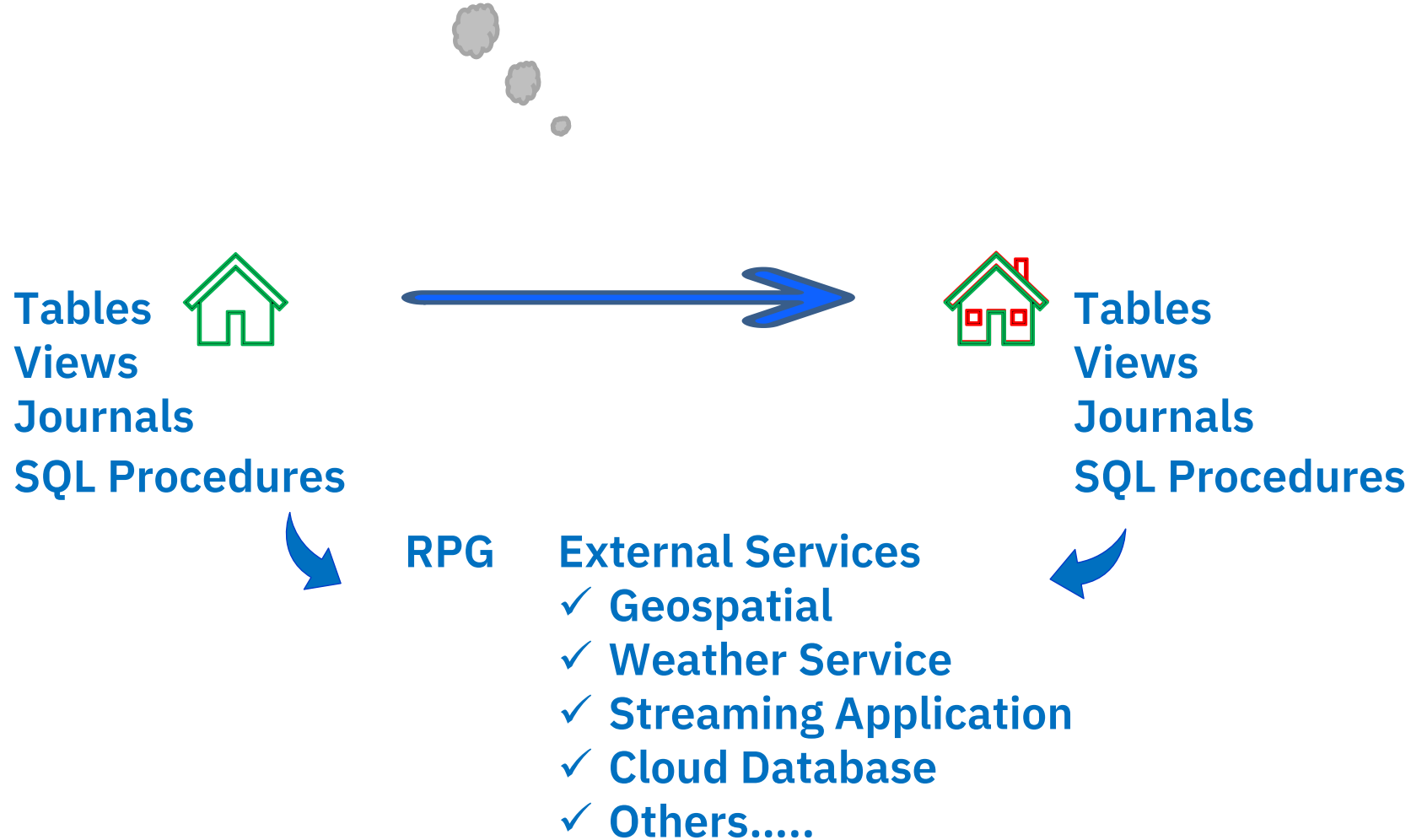
Motioning Data – More Complex Cases

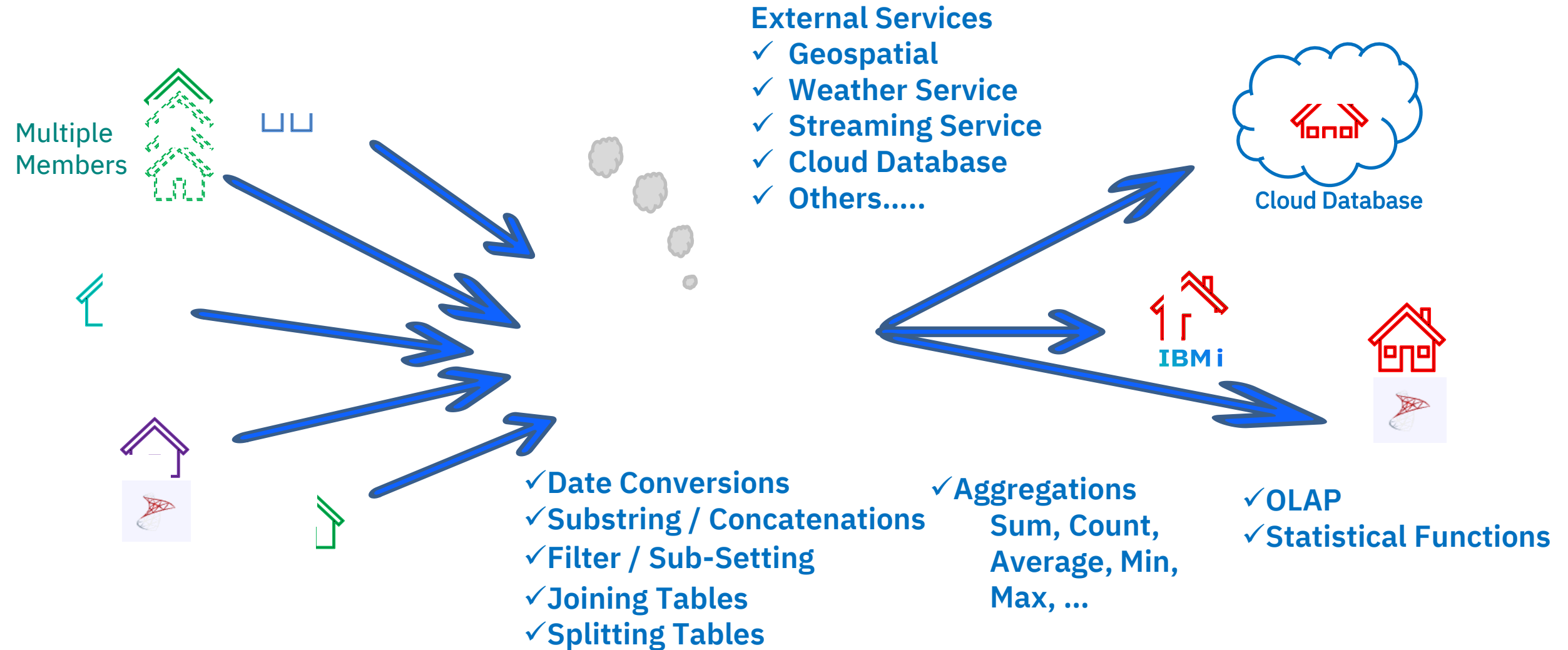
Non IBM i sources and targets ...and CLOUD sources and targets ...using Open Source technologies



Motioning and Enhancing Data

Clean, Enhance, and Transform using a “Factory”





How Do I Get Started?

Let's talk about one way...

DataMigrator - 5733WQX

A good place to start especially for IBM i environments

Extract, Transform, and Load Tool (ETL)

Low Cost

Integrated with and runs on IBM i (7.3+)

Understands Db2 for i

SQL based – set vs row at a time

Part of Db2 Web Query for i

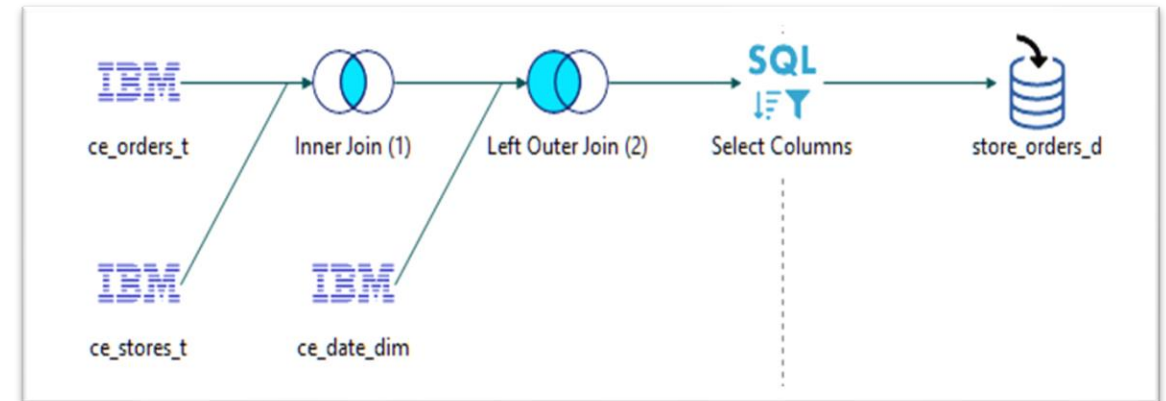
Benefits

Low Code

- GUI
- Drag and Drop
- PC Design Client

SQL Helpful; not necessary

Access to non IBM i databases

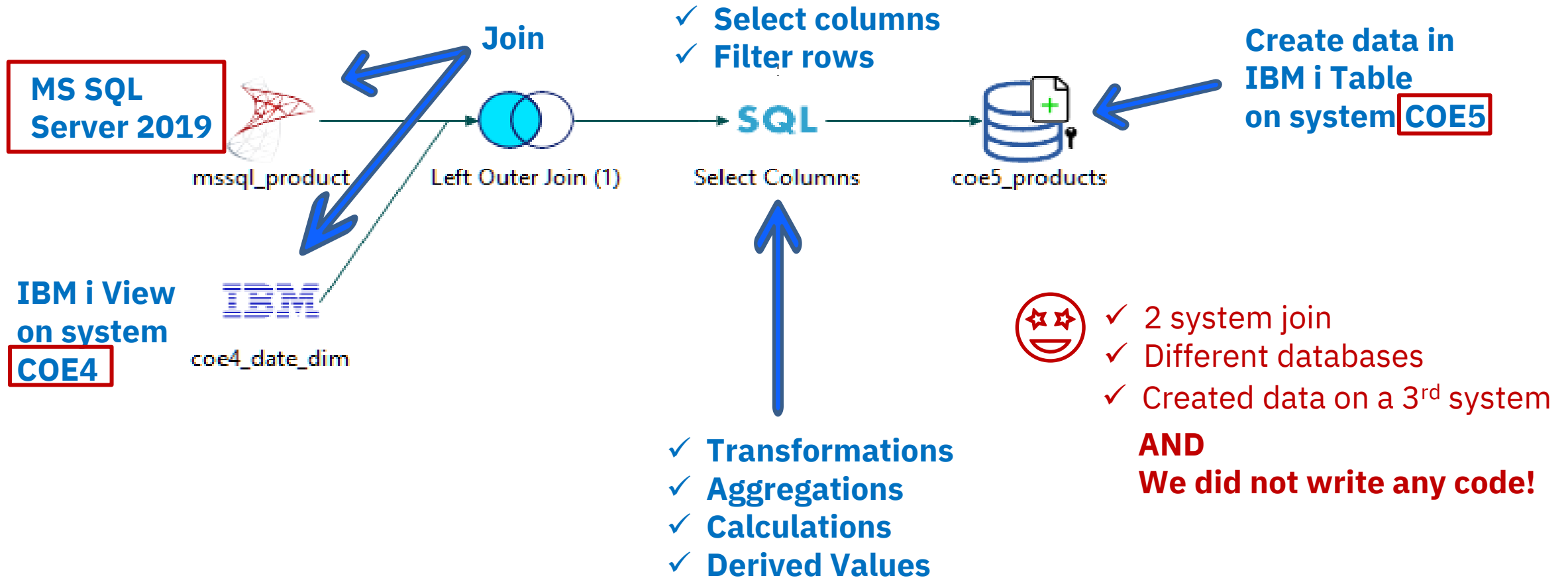


The screenshot displays the IBM DataMigrator interface. At the top, there is a ribbon menu with tabs for 'Home' and 'Flow'. The 'Flow' tab is active, showing various icons for editing and viewing. Below the ribbon is the 'Object Browser' pane, which lists various objects in a tree view. The main workspace shows a data flow diagram with a source object and a target object connected by a blue arrow. A callout box labeled 'New target' is positioned over the target object. Below the diagram, the text 'SQL' and 'Select Columns' is visible.

Name	Type	Description
CTCV71	Server	
DB2ICOE4	Server	
Application Directories		
foccache(Temporary)	Directory	Session Directory: Files Lost on Disco
baseapp	Directory	Default Directory: Files Always Availa
john_westcott	Directory	
sample_flow_with_s...	Flow	
coe4_date_dim	Warehouse	Date Dimension Table - IBM v1
coe3_products	DB2/DBZ vwarehouse	
mssql_product	Server JDBC/AzureDB	Products sold or used in the manufact
product_date_dim	Fixed Format Files	
sql_productcategory_t	MS SQL Server JDBC/AzureDB	High-level product categorization.

DataMigrator – Low Code To NO CODE

Let's look at the replay



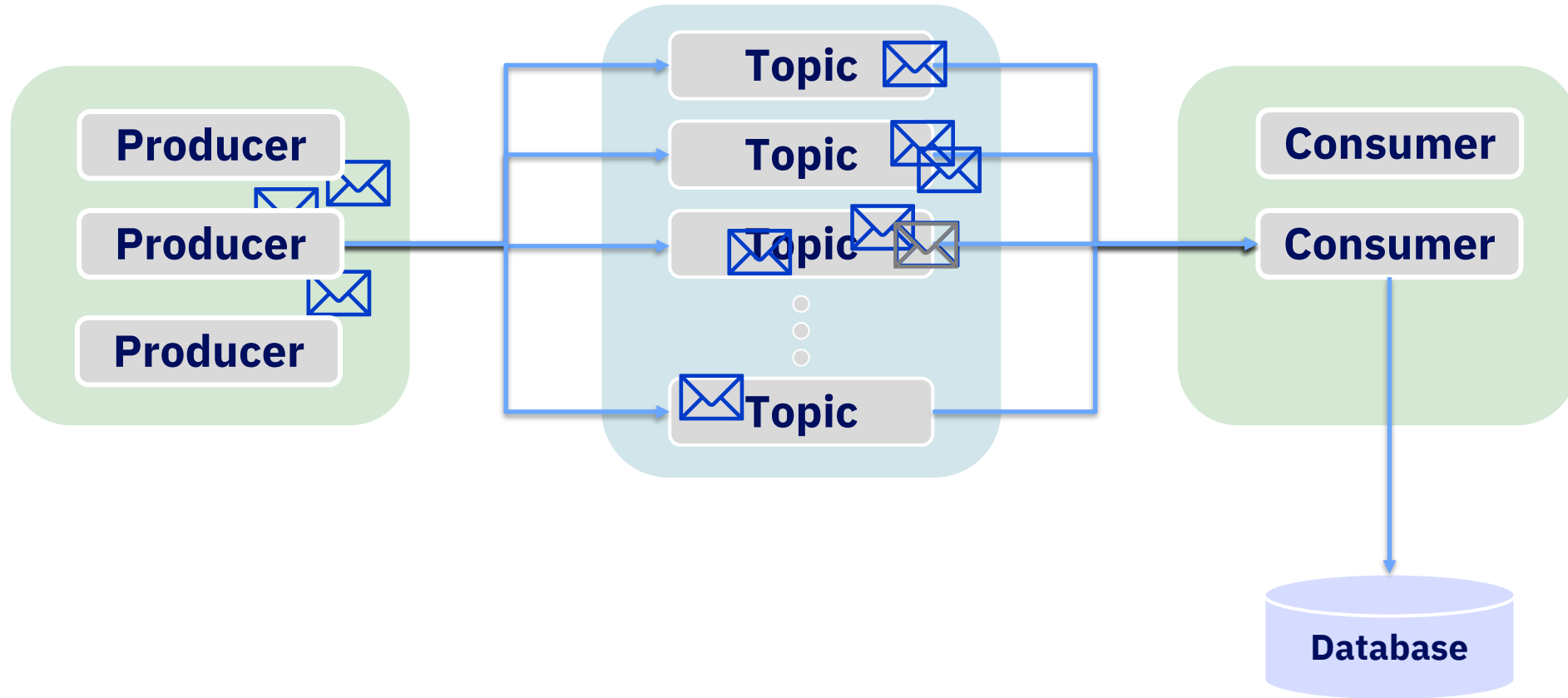
Streaming with Apache Kafka



Apache Kafka is a real-time, fault tolerant, highly scalable streaming messaging system

Simplified Kafka Workflow

Generalized example of how a Kafka flow could work



At the end of each Business Day
send IBM i Sales Order Data
to a Cloud Application (Watson Studio)
and to an Azure Data Lake
using Kafka

Kafka Workflow – Action Plan

Call to action

Who

From IBM i

To ~~Watson Studio and MS Azure~~ **Kafka**



With Kafka, we are not sending data directly to Watson and Azure. We are producing messages for some consumer to process.

What

Sales Order Information

- Sales by Store
- Returns by Store
- Sales for Event by Store

} Three (3) Kafka topics

When

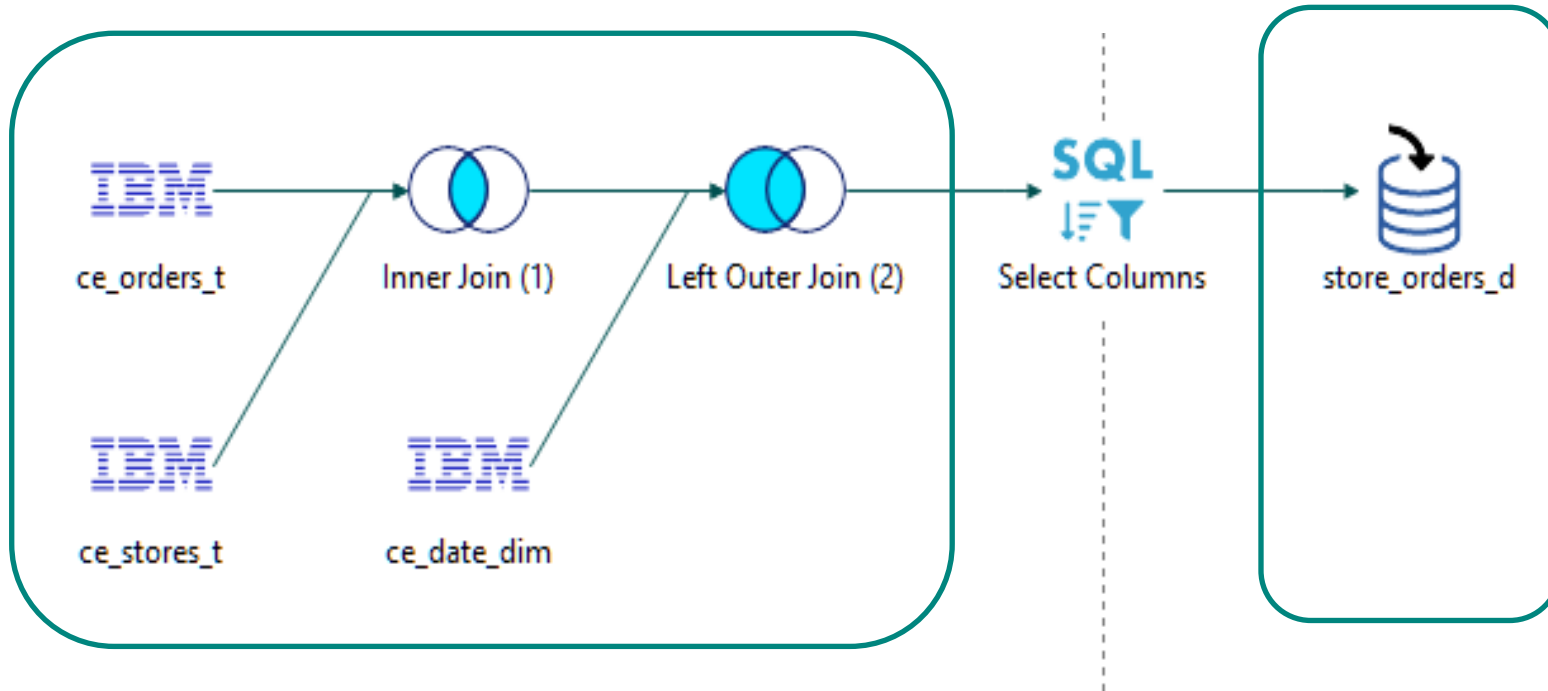
At End of Business Day

How

In JSON format

Kafka Workflow – Data Flow

Repeated use a temp (staging) table

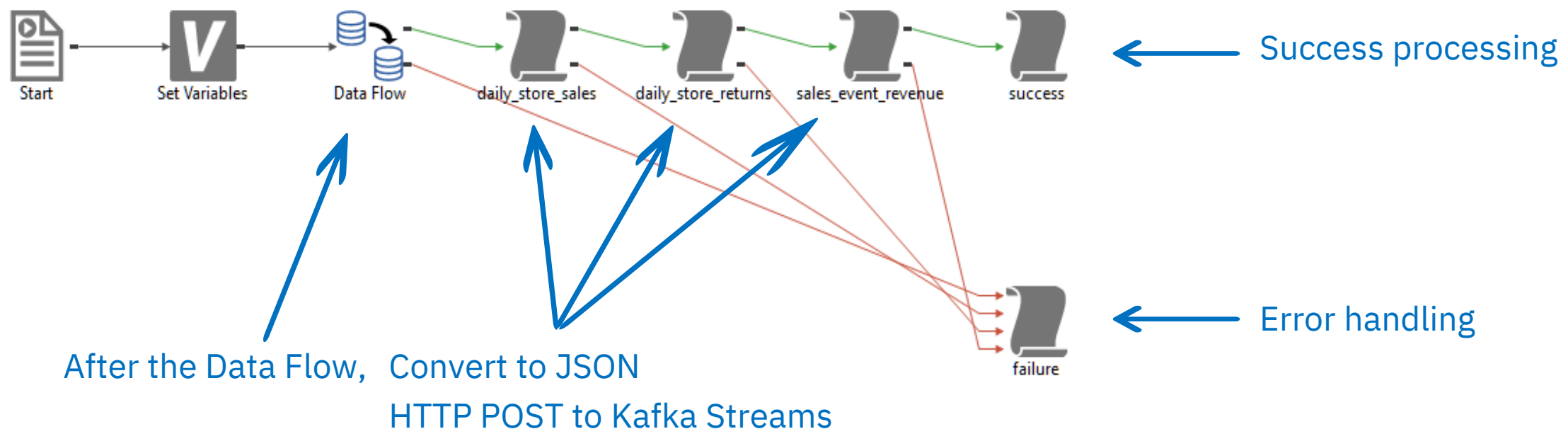


Staging Table will be processed three times

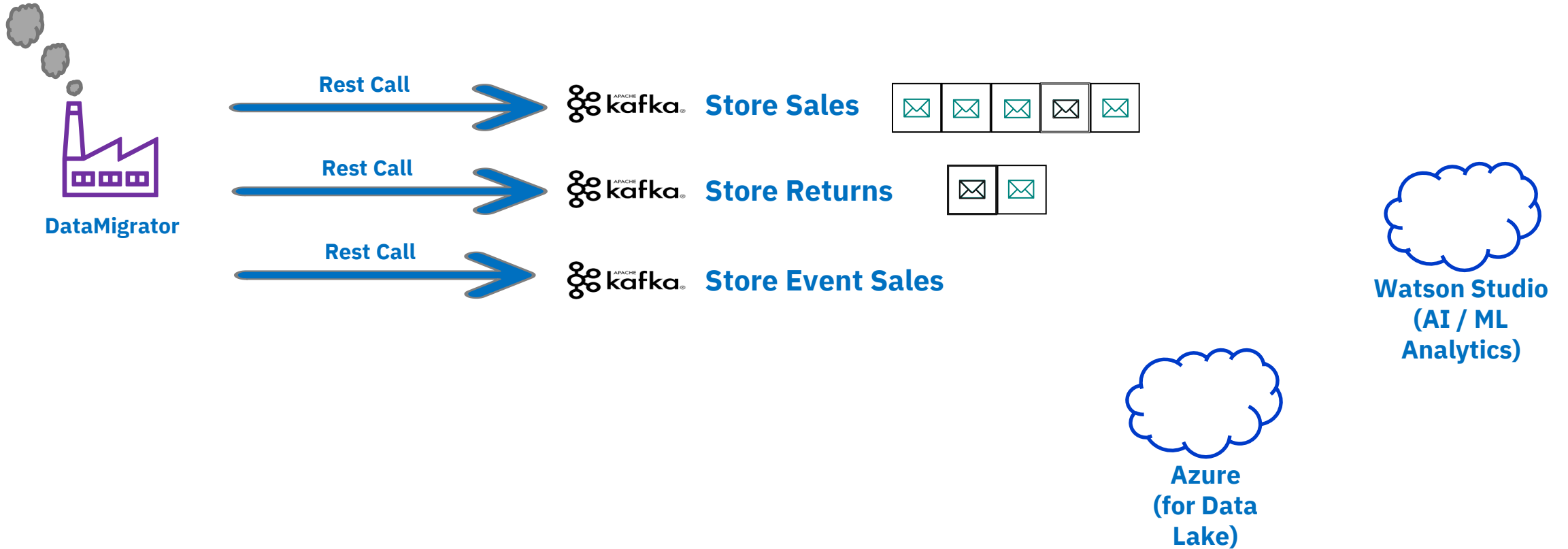
1. Sales by Store
2. Returns by Store
3. Sales for Event by Store

Kafka Workflow – Process Flow

For this example, the flow's where the magic happens



After the Data Flow, Convert to JSON
HTTP POST to Kafka Streams



Using Data Queues and Camel



Apache Camel is an Open-Source integration framework providing rules-based routing.

New call to action

Who

- **From** IBM i
- **To** Kafka **Via Camel**

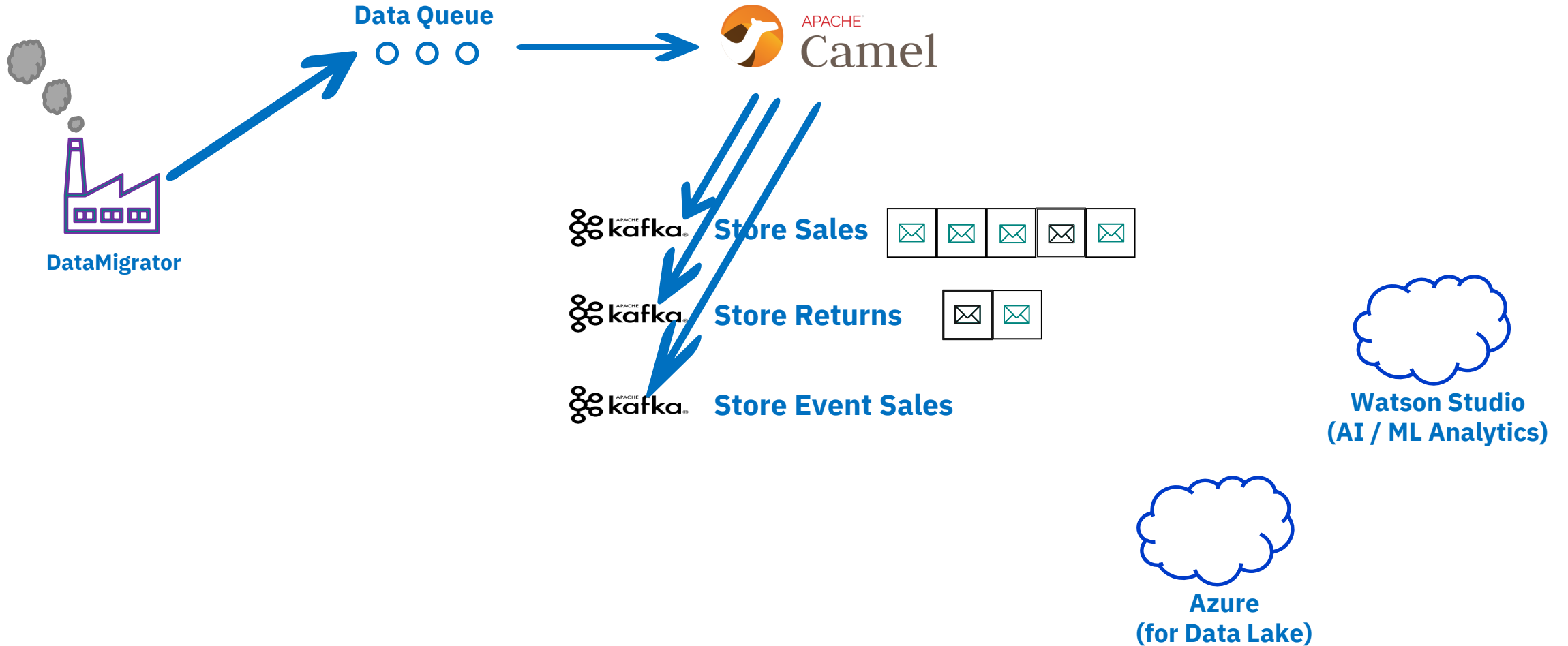
What: Sales Order Information

- Sales by Store
- Returns by Store
- Sales for Event by Store

When: At End of Business Day

How: In JSON format

Camel + Kafka Workflow – Flow



Streaming à la Debezium



Debezium provides services to capture changes in your databases and post these changes to *change event streams*. Applications simply read these streams to see the change events in the same order in which they occurred.

Translating

We read your journals and produce Kafka topics where applications can later consume the messages – maybe in the order they occurred.

The element Db-ezium

Does more than we'll talk about today...

Has no IBM i connector And that's the point... just because Debezium does not have an IBM i adapter DOES NOT mean we can't participate !!

Debezium feeds Kafka with specific information about journal activities

Must follow the Debezium rules for Kafka topics

Predefined JSON specifications for each transaction type: INSERT, UPDATE, DELETE

```
{  
  "schema": { ... },  
  "payload": {  
    "before": { ... },  
    "after": { ... },  
    "source": { ... }  
  },  
},  
"operation-type": "...",  
"cdc_transaction_id": "...",  
"cdc_transaction_ts": "...",  
"cdc_transaction_rrn": "..."
```

Pipelining Debezium

Yes, a data pipeline of journal data





Links

Db2 Web Query for i Website

ibm.biz/db2webqueryi

Db2 Web Query for i Wiki

ibm.biz/db2wqwiki

Db2 Web Query Getting Started Enablement

<https://ibm.biz/db2ilabservices>

Db2 Web Query Version 2.3.0

Videos: <http://ibm.biz/db2wq-230-videos>

EZ-Install

<http://Ibm.biz/db2wqezinstall-info>



Other Links

Compliance Automation Reporting Tool (CART)

<https://www.ibm.com/support/pages/ibm-i-security#cart>

Db2 Web Query, DataMigrator - John Westcott

<http://ibmidb2ianalytics.blogspot.com>

Db2 Web Query for i - Doug Mack

<http://db2webqueryi.blogspot.com>

Db2 for i Blog - Kent Milligan

<http://db2ibmi.blogspot.com/>



Questions?

Thank You

John Westcott

Db2 for i Analytics Consultant

JWestcott@ibm.com

Db2 Web Query Team: **QU2@us.ibm.com**

© Copyright IBM Corporation 2022. All rights reserved. The information contained in these materials is provided for informational purposes only and is provided AS IS without warranty of any kind, express or implied. Any statement of direction represents IBM's current intent, is subject to change or withdrawal, and represent only goals and objectives. IBM, the IBM, and the IBM i logo are trademarks or registered trademarks of International Business Machines Corporation, in the United States and/or other countries. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on [ibm.com/trademark](https://www.ibm.com/trademark).

IBM Systems Expert Labs
Db2 Web Query for i

This document was developed for IBM offerings in the United States as of the date of publication. IBM may not make these offerings available in other countries, and the information is subject to change without notice. Consult your local IBM business contact for information on the IBM offerings available in your area.

Information in this document concerning non-IBM products was obtained from the suppliers of these products or other public sources. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

IBM may have patents or pending patent applications covering subject matter in this document. The furnishing of this document does not give you any license to these patents. Send license inquiries, in writing, to IBM Director of Licensing, IBM Corporation, New Castle Drive, Armonk, NY 10504-1785 USA.

All statements regarding IBM future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only.

The information contained in this document has not been submitted to any formal IBM test and is provided "AS IS" with no warranties or guarantees either expressed or implied.

All examples cited or described in this document are presented as illustrations of the manner in which some IBM products can be used and the results that may be achieved. Actual environmental costs and performance characteristics will vary depending on individual client configurations and conditions.

IBM Global Financing offerings are provided through IBM Credit Corporation in the United States and other IBM subsidiaries and divisions worldwide to qualified commercial and government clients. Rates are based on a client's credit rating, financing terms, offering type, equipment type and options, and may vary by country. Other restrictions may apply. Rates and offerings are subject to change, extension or withdrawal without notice.

IBM is not responsible for printing errors in this document that result in pricing or information inaccuracies.

All prices shown are IBM's United States suggested list prices and are subject to change without notice; reseller prices may vary.

IBM hardware products are manufactured from new parts, or new and serviceable used parts. Regardless, our warranty terms apply.

Any performance data contained in this document was determined in a controlled environment. Actual results may vary significantly and are dependent on many factors including system hardware configuration and software design and configuration. Some measurements quoted in this document may have been made on development-level systems. There is no guarantee these measurements will be the same on generally-available systems. Some measurements quoted in this document may have been estimated through extrapolation. Users of this document should verify the applicable data for their specific environment.

IBM, the IBM logo, ibm.com AIX, AIX (logo), AIX 5L, AIX 6 (logo), AS/400, BladeCenter, Blue Gene, ClusterProven, Db2, ESCON, i5/OS, i5/OS (logo), IBM Business Partner (logo), IntelliStation, LoadLeveler, Lotus, Lotus Notes, Notes, Operating System/400, OS/400, PartnerLink, PartnerWorld, PowerPC, pSeries, Rational, RISC System/6000, RS/6000, THINK, Tivoli, Tivoli (logo), Tivoli Management Environment, WebSphere, xSeries, z/OS, zSeries, Active Memory, Balanced Warehouse, CacheFlow, Cool Blue, IBM Systems Director VMControl, pureScale, TurboCore, Chiphopper, Cloudscape, Db2 Universal Database, DS4000, DS6000, DS8000, EnergyScale, Enterprise Workload Manager, General Parallel File System, , GPFS, HACMP, HACMP/6000, HASM, IBM Systems Director Active Energy Manager, iSeries, Micro-Partitioning, POWER, PowerExecutive, PowerVM, PowerVM (logo), PowerHA, Power Architecture, Power Everywhere, Power Family, POWER Hypervisor, Power Systems, Power Systems (logo), Power Systems Software, Power Systems Software (logo), POWER2, POWER3, POWER4, POWER4+, POWER5, POWER5+, POWER6, POWER6+, POWER7, System i, System p, System p5, System Storage, System z, TME 10, Workload Partitions Manager and X-Architecture are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries, or both. If these and other IBM trademarked terms are marked on their first occurrence in this information with a trademark symbol (® or ™), these symbols indicate U.S. registered or common law trademarks owned by IBM at the time this information was published. Such trademarks may also be registered or common law trademarks in other countries.

A full list of U.S. trademarks owned by IBM may be found at: <http://www.ibm.com/legal/copytrade.shtml>.

Adobe, the Adobe logo, PostScript, and the PostScript logo are either registered trademarks or trademarks of Adobe Systems Incorporated in the United States, and/or other countries.

AltiVec is a trademark of Freescale Semiconductor, Inc.

AMD Opteron is a trademark of Advanced Micro Devices, Inc.

InfiniBand, InfiniBand Trade Association and the InfiniBand design marks are trademarks and/or service marks of the InfiniBand Trade Association.

Intel, Intel logo, Intel Inside, Intel Inside logo, Intel Centrino, Intel Centrino logo, Celeron, Intel Xeon, Intel SpeedStep, Itanium, and Pentium are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

IT Infrastructure Library is a registered trademark of the Central Computer and Telecommunications Agency which is now part of the Office of Government Commerce.

Java and all Java-based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates.

Linear Tape-Open, LTO, the LTO Logo, Ultrium, and the Ultrium logo are trademarks of HP, IBM Corp. and Quantum in the U.S. and other countries.

Linux is a registered trademark of Linus Torvalds in the United States, other countries or both.

Microsoft, Windows and the Windows logo are registered trademarks of Microsoft Corporation in the United States, other countries or both.

NetBench is a registered trademark of Ziff Davis Media in the United States, other countries or both.

SPECint, SPECfp, SPECjbb, SPECweb, SPECjAppServer, SPEC OMP, SPECviewperf, SPECcapc, SPECchpc, SPECjvm, SPECmail, SPECimap and SPECsfs are trademarks of the Standard Performance Evaluation Corp (SPEC).

The Power Architecture and Power.org wordmarks and the Power and Power.org logos and related marks are trademarks and service marks licensed by Power.org.

TPC-C and TPC-H are trademarks of the Transaction Performance Processing Council (TPPC).

UNIX is a registered trademark of The Open Group in the United States, other countries or both.

Other company, product and service names may be trademarks or service marks of others.