OpenTSx

Architecture Overview

Architecture overview...

Architecture:
Identify Patterns & Buildingblocks

3 typical flow patterns:

(1) Linear flow ...

(2) Bi-directional flow ...

(3) Complex process flows ...

3 typical flow patterns:

(1) Linear flow …

(2) Bi-directional flow …

(3) Complex process flows …

Easy:

1. reusable
2. scalable
3. ready to use
4. ready to improve

⇒ **Target for simplification …**

A Standardized Processing Procedure for Episodes
used for social media analysis on Hadoop:

- The predecessor of **OpenTSx** is **Hadoop.TS**

  (https://www.researchgate.net/publication/269687614_Hadoop_TS_Large-Scale_Time-Series_Processing)

- Hadoop.TS used a variety of Hadoop ecosystem projects
  (Sqoop, Flume, Hive, Spark, Yarn, HDFS, Solr, HBase, Impala, Mahout, Giraph)

- Managing the data flow at scale was **possible**, but complex.

OVERVIEW - DATA FLOW
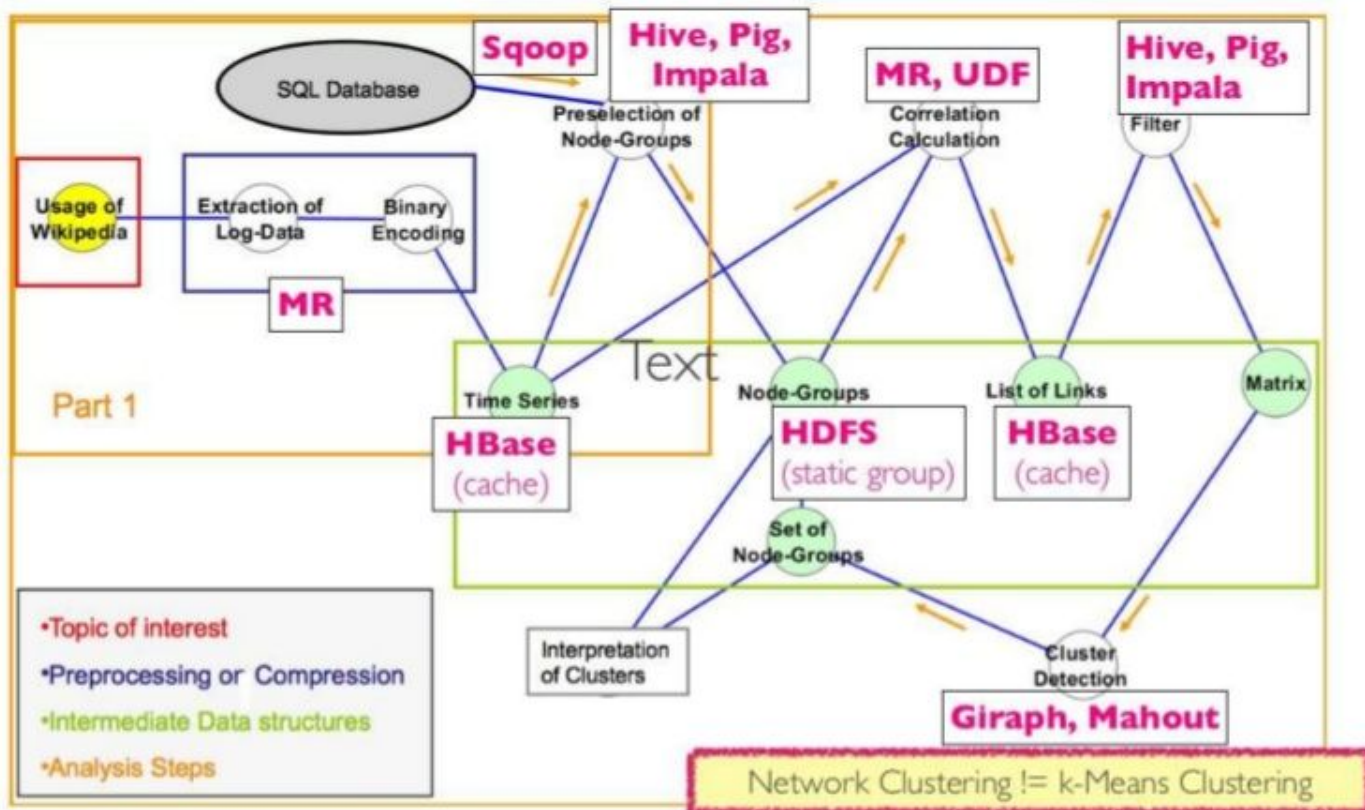
This example illustrate the variety
of interconnected components from our
previous implementation using the Hadoop ecosystem.

The resulting complexity of a solution can become a blocker!

Architects have to simplify ...

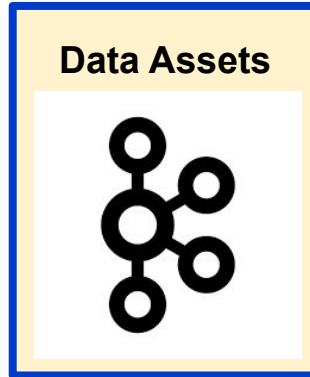... and the Kafka Ecosystem helps <u>you</u> on this journey!

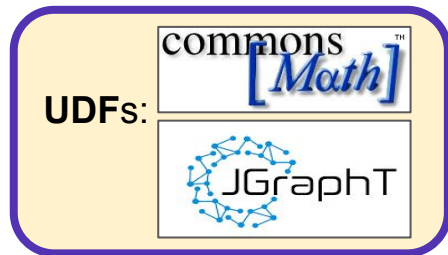# KEEP CALM

AND

Simplify

Building Blocks:

Data flows are no
longer transient.
The event log acts as
single source of truth.

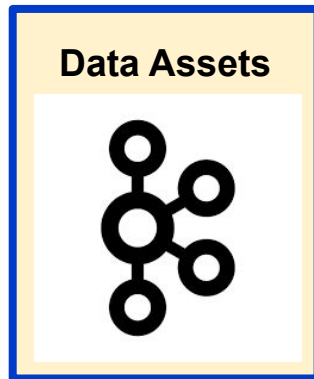**Data Assets**

*Paradigm Shift in
Data Management*

Building Blocks:

Domain specific logic is
implemented in small
reusable components:

UDFs:

commons
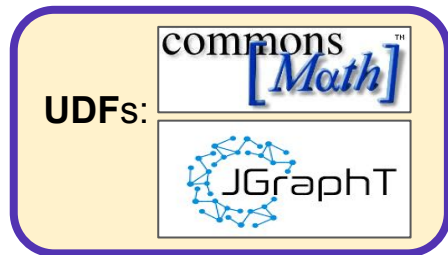[*Math*]™

JGraphT

Domain Driven Design

Data flows are no
longer transient.
The event log acts as
single source of truth.

Data Assets

*Paradigm Shift in
Data Management*

Building Blocks:

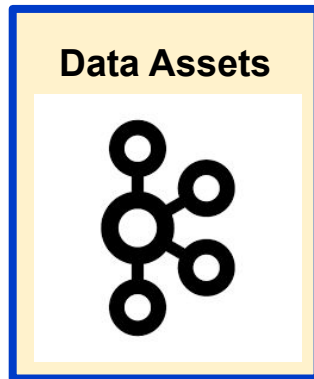Domain specific logic is implemented in small eusable components:

UDFs: commons [Math]

JGraphT

Domain Driven Design

Source Connectors Integrate input side …

Kafka Connect
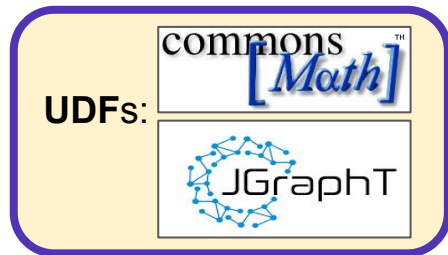
APACHE PLC4X

Legacy and Future Systems

Data flows are no longer transient.
The event log acts as single source of truth.

Data Assets

Paradigm Shift in Data Management

Building Blocks:

Domain specific logic is
implemented in small
reusable components:

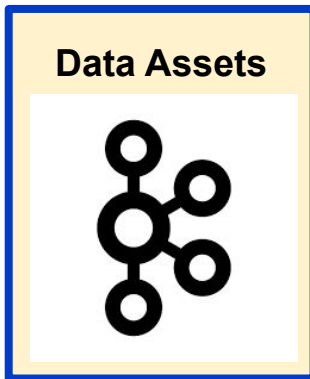**UDFs:** commons [Math] / JGraphT

Domain Driven Design

Source Connectors
integrate sources …

Kafka Connect — APACHE PLC4X

Legacy and Future Systems

Data flows are no
longer transient.
The event log acts as
single source of truth.

Data Assets

*Paradigm Shift in
Data Management*

Sink Connectors
integrate external targets …

Kafka Connect

elasticsearch

SCYLLA

neo4j

Special Purpose Systems

Confluent Platform for Time Series Processing
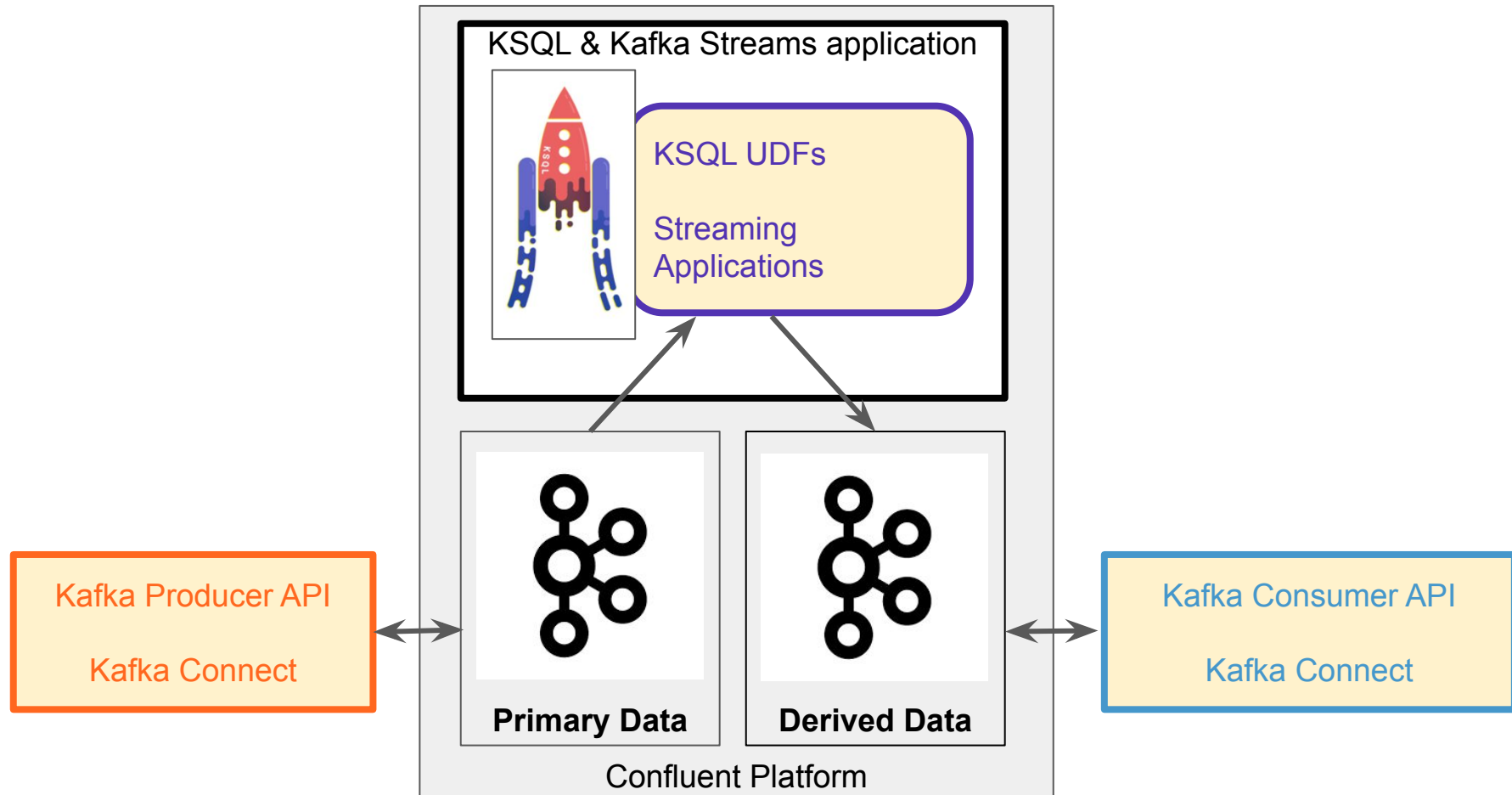
confluent

THANK YOU !!!

mirko@confluent.io

🐦 @semanpix