Flink & Kafka

An Apache Love Story
June 18, 2020

Introduction - Who Is The Data Jedi

- Completed PhD in Computer Science in 2015 from the University of North Texas
- Experienced Software Engineer, Data Architect, Data Scientist, and University Instructor
- Involved in Big Data and Spark since 2015
- Joined HWX/Cloudera as a SE with the US Federal team
- DS SME 2018
- Hobbies: Travel, Fitness, Texas, and Beer

Why Use Apache Flink??

siliconANGLE [the voice of enterprise and emerging tech]

CLOUD AI SECURITY INFRA BLOCKCHAIN POLICY BIG DATA APPS EMERGING TECH •••

theCUBE Coverage from SiliconANGLE's livestreaming video studio

UPDATED 17:00 EDT / APRIL 16 2018

BIG DATA

Apache Flink helps Netflix process 3 trillion events every day

BY MARK ALBERTSON

The processing demands for a video content service like Netflix Inc. are almost unimaginable. A consumer audience of over 109 million subscribers enjoys 125 million hours of TV and movie content via the online subscriber service every single day.

That places great demand on the company's data ingestion pipeline and stream processing engines, which must handle 3 trillion daily events involving 12 petabytes of data. One of the platforms used by Netflix is Apache Flink, an open-source tool for distributed stream and batch data processing.

i·t for all HOME POD

ARTICLE Technology › Data Analytics

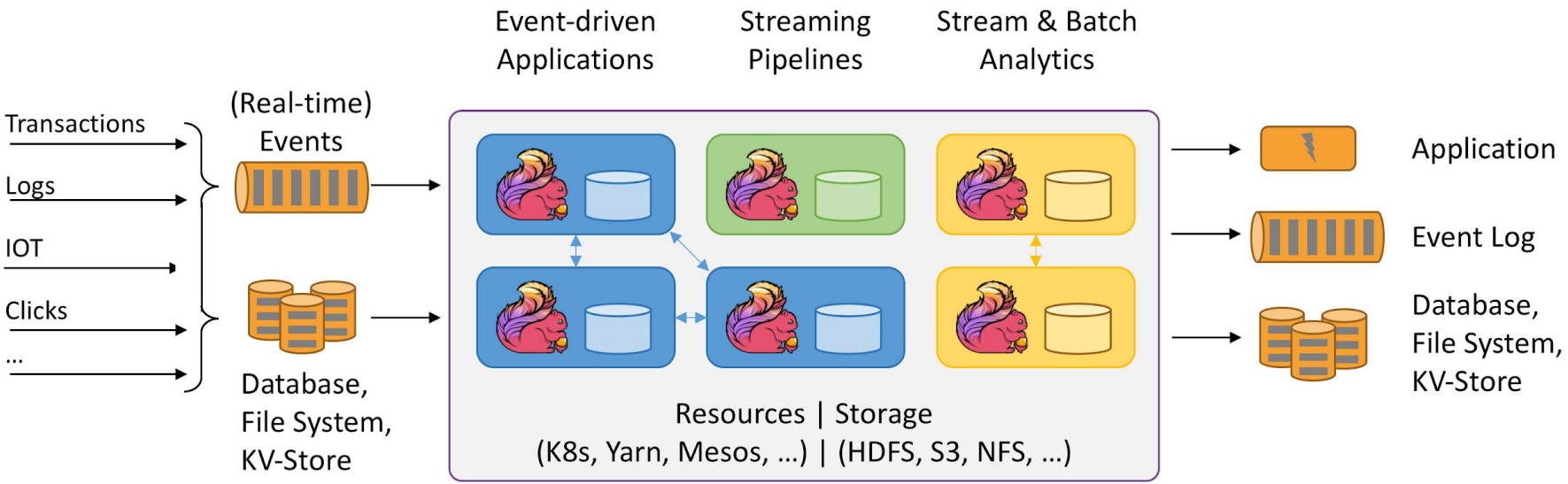
IoT Data Processing With Apache Flink: A Game Changer?

Organizations implementing an IoT strategy face the challenge of finding the right data processing architecture. Apache Flink is an excellent option for processing streaming IoT data.

IoT For All - May 15, 2019

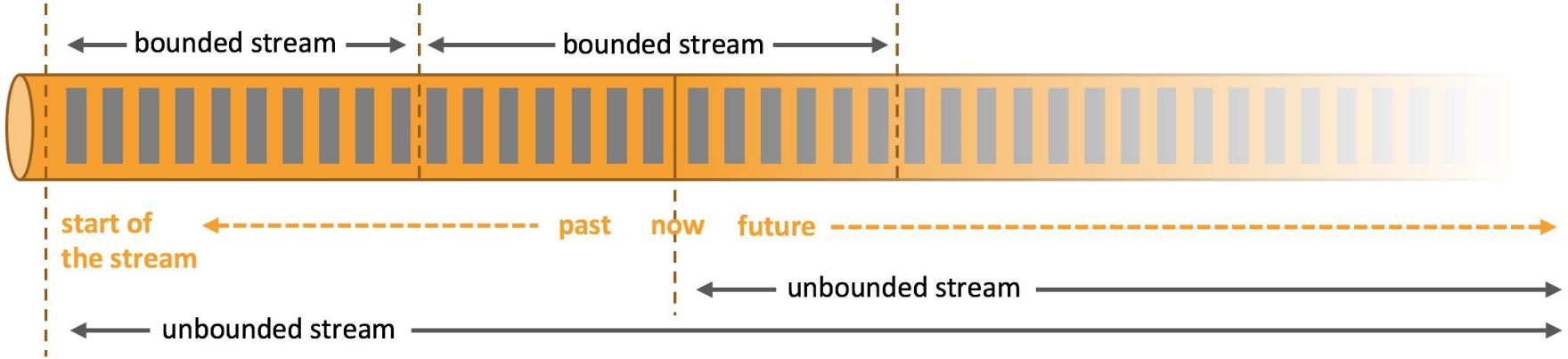
Illustration: © IoT For All

Apache Flink Is A Distributed Data Processing System



Flink Unifies Stream and Batch Processing

- Processes *unbounded* (stream) and *bounded* (batch) data
- Processes *recorded* (offline) and *live* (real-time) data
- Serves most streaming & batch use cases
  - Data Pipelines, Analytics, CEP, Event-driven Applications



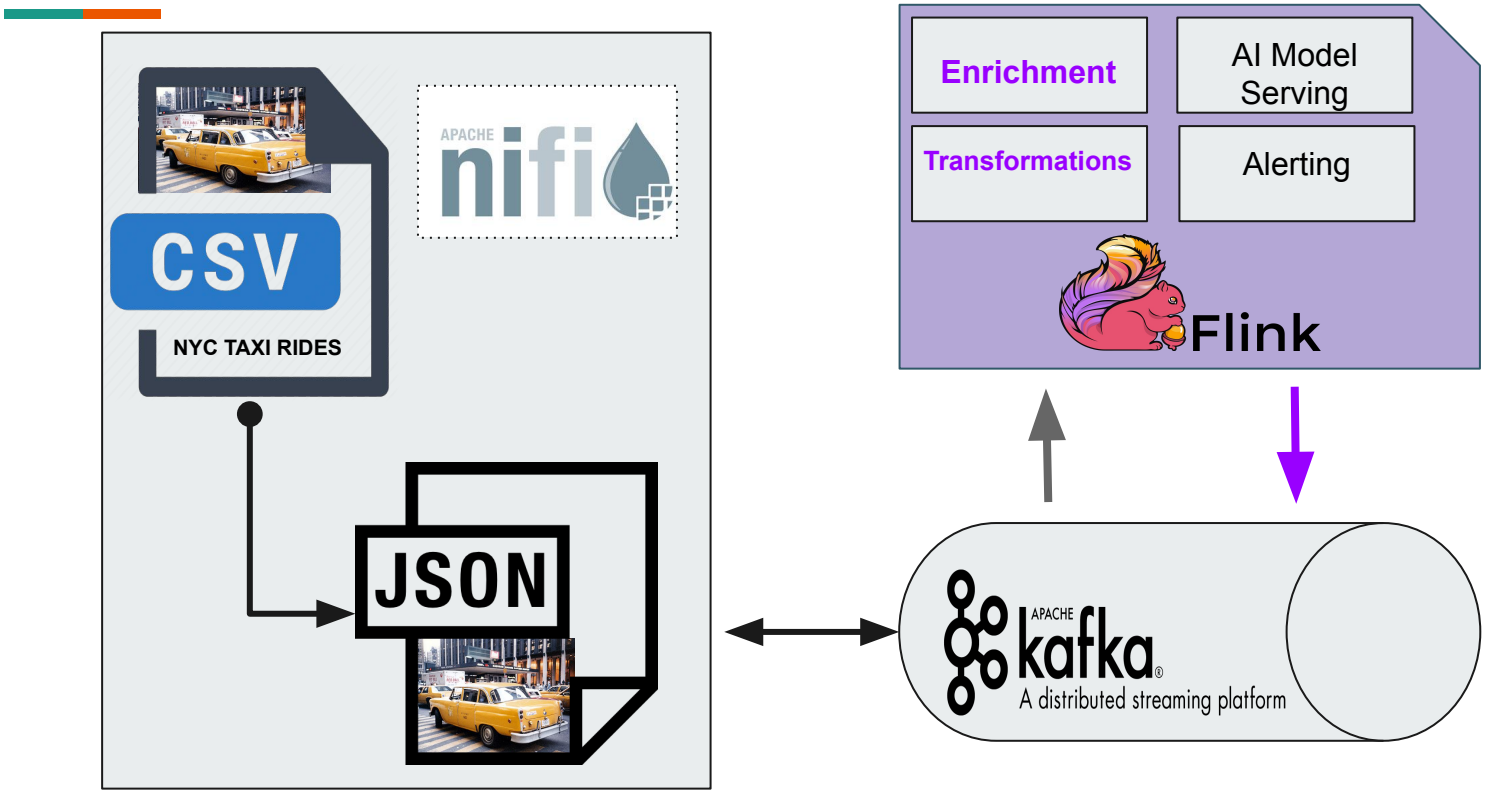
Event-Time and Processing-Time

This is called **event time**

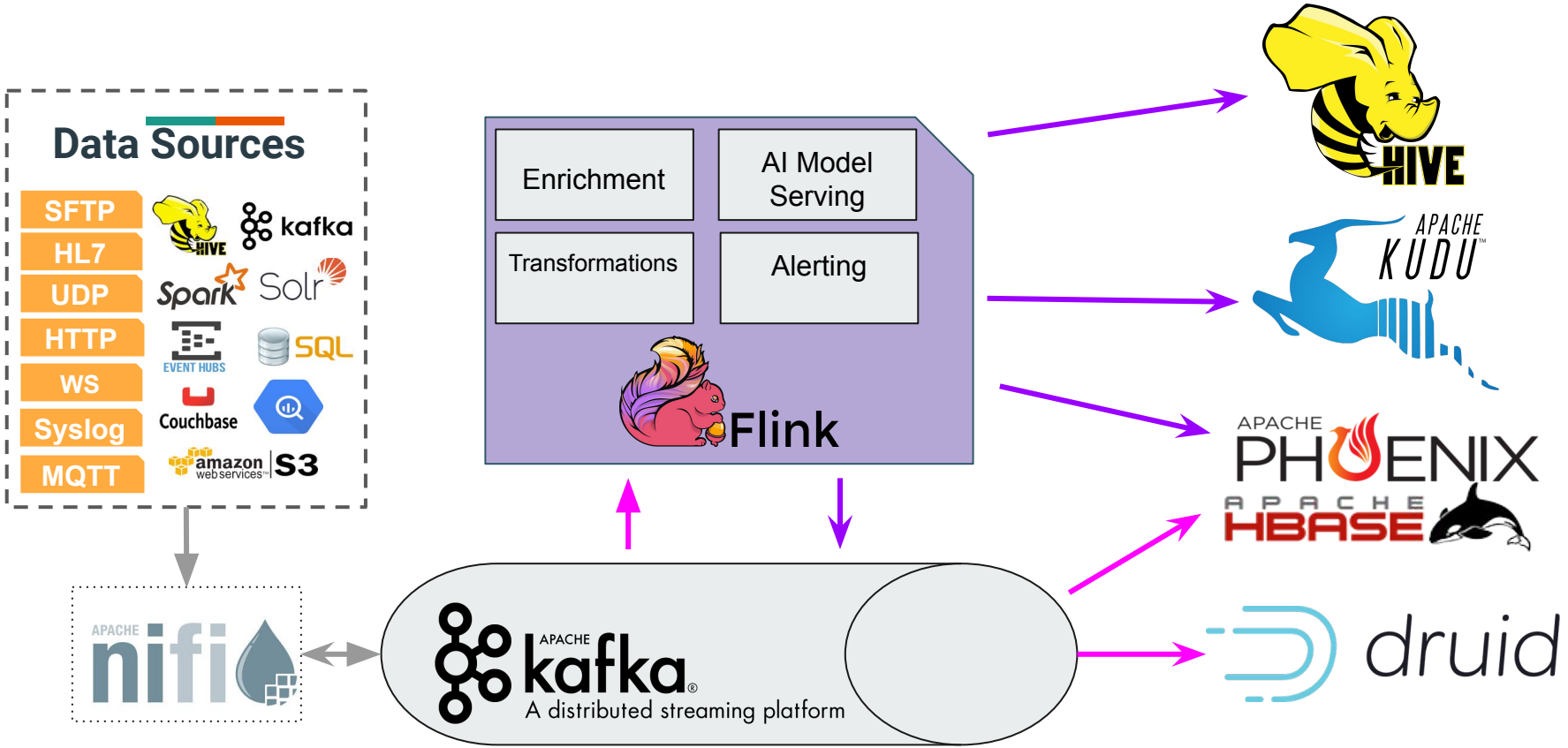
Episode IV	Episode V	Episode VI	Episode I	Episode II	Episode III	Episode VII
A New Hope	The Empire Strikes Back	Return of the Jedi	The Phantom Menace	Attack of the Clones	Revenge of the Sith	The Force Awakens
1977	1980	1983	1999	2002	2005	2015

This is called **processing time**

Today's Demo Reference Architecture



Apache Zoo Animal Reference Architecture



Additional Documentation

- Additional project documentation on Flink and Kafka from Ian's Github collection
- Netflix Real-time Stream Processing Platform
- Apache Flink helps Netflix process 3 trillion events every day
- IoT Data Processing With Apache Flink: A Game Changer?

THANK YOU!