Building streaming data pipelines with Google Cloud Dataflow and Confluent Cloud

Elena Cuevas,
Senior Partner Solutions Engineer @ Confluent

August 2021
01. Data in Motion
   The new Data in Motion Paradigm

02. Product Overview
   Introduction to Confluent

03. Demo Use Case and Architecture
   How Confluent and Google partner to fulfill the Data in Motion paradigm

04. Demo
   Building streaming data pipelines with Google Cloud Dataflow and Confluent Cloud

05. Q&A
Data in Motion
Enterprises require total connectivity and instant reaction, 24x7, anywhere, in real-time. But they can’t get there with traditional, historical databases filled with data at rest. They need a complete streaming platform, dually capable of setting data in motion and analyzing that data in real-time, and which is globally interconnecting the clouds and on-premises data centers.

Today, the digital realm is as important as the physical world in how business is transacted.
A New Paradigm is Required for Data in Motion:
Continuously processing evolving streams of data in real-time

Real-time Event Streams and Analysis

- A Sale
- A shipment
- A Trade
- A Customer Experience

Rich front-end customer experiences
Real-time backend operations
At Confluent, streaming is in our DNA.

We help the world’s largest organizations make it part of theirs.

https://www.confluent.io/customers/
Cloud-native, Complete, Everywhere
with Kafka at its core

- Fully Managed ‘NoOps’ on AWS, Azure, GCP
- Security & Data Governance
- Infinite Storage
- ksqlDB & Stream Processing, Analytics
- Connectors
- APIs, UIs, CLIs
What makes Confluent unique?

**Cloud-Native**
Available as a fully managed service that is a serverless, infinitely scalable, elastic, secure, and globally interconnected. Our self-managed service inherits all the work born in the cloud.

**Complete**
ksqlDB, Connect, Schema Registry, and more
Capable of end-to-end applications
Kafka from the people who made it

**Everywhere**
Global availability on AWS, Azure, and GCP
Bridge on-prem to cloud with cluster linking
Extend streaming apps across clouds
Confluent: Everywhere

Fully-Managed Service

**Confluent Cloud**
Apache Kafka Re-engineered for the Cloud

Available on the leading public clouds

Self-Managed Software

**Confluent Platform**
The Enterprise Distribution of Apache Kafka

Deploy on any platform, on-prem or cloud

Both: Subscription products where price scales with usage
Large Ecosystem for Event Streaming
Easily connect to 100+ data systems
Large Partner Network
Consulting Partners, Cloud Partners, OEM Partners, Tech Partners

https://www.confluent.io/partners/
Demo Architecture
Demo

Confluent Cloud

Real-time user application sends entries

Schema Registry

Streaming Analytics & Processing (ksqlDB)

Cloud Connectors

Infinite Storage

Promotions

BigQuery

Process Winners

Dataflow

Winners

BigQuery

Cloud-native Services

Copyright 2021, Confluent, Inc. All rights reserved. This document may not be reproduced in any manner without the express written permission of Confluent, Inc.
Demo

Real-time user application sends entries to Confluent Cloud

- topic: entries
- topic: application
- topic: ...

Confluent Cloud processes the data with the following steps:

1. **PCollection: Entries**
   - Window (60s)
   - CoGroupByKey
   - ParDo: toKV

2. **PCollection: Winners**
   - WithKeys
   - PCollection: Entries
   - Window (60s)
   - ParDo: toTableRow

3. Promotions sent to BigQuery

Cloud-native Services connect to Confluent Cloud with Azure, AWS, and Salesforce.
Q&A
Stay in touch!

Confluent Blog
cnfl.io/developer

Community Slack
cnfl.io/slack

Global Training
cnfl.io/kafka-training
Demo Architecture

Confluent Cloud

Customer services & apps on-premises (US)

Customer services & apps self-managed in AWS

Infinite Storage

Streaming Analytics & Processing (ksqlDB)

Schema Registry

Cluster Linking

Connectors

Cloud Connectors

Customer services & apps on-premises (EU)

Real-time analytics triggered a push alert "Fraudulent transaction!" to the B2C customer

Cloud-native Services

Microsoft Azure

AWS

Cloud-native Services

mongoDB

Atlas
Thank you!

@elena_cxc

linkedin.com/in/ecuevasc