Change Data Capture with Debezium

☕ Java Vienna 🎡 | 03. Juni 2024

Hans-Peter Grahsl
@hpgrahsl

Developer 🥑 Advocate
Red Hat

Red Hat Developer

Red Hat Developer | DevNation

Hans-Peter Grahsl

"I'm passionate about event-driven architectures, distributed stream processing systems and data engineering."

Developer 🥑 Advocate @ Red Hat

based in Graz, Austria 🇦🇹

open-source enthusiast

Confluent Community Catalyst

MongoDB Champion since 2020

speaker at dev & tech conferences

@hpgrahsl@mastodon.social

Agenda

- What is Debezium and Change Data Capture (CDC)?
- Benefits of Log-based CDC
- Databases supported by Debezium
- Change Event Payload Structure
- Debezium deployment modes

- 🎬 **Debezium in Action** 🎬

Red Hat Developer | DevNation

Debezium Introduction

debezium.io

Red Hat Developer | DevNation

debezium

FAQ DOCUMENTATION RELEASES COMMUNITY BLOG

Debezium

Latest stable (2.6)

Development (2.7)

Stream changes from your database.

Debezium is an open source distributed platform for change data capture. Start it up, point it at your databases, and your apps can start responding to all of the inserts, updates, and deletes that other apps commit to your databases. Debezium is durable and fast, so your apps can respond quickly and never miss an event, even when things go wrong.

debezium / debezium

Type / to search

Code | Pull requests 64 | Actions | Projects | Security | Insights

debezium Public

Edit Pins | Watch 209 | Fork 2.4k | Starred 10k

debezium.io

Debezium in a Nutshell

- **fully open-source + very active community**

- **change data capture (CDC)** platform
  - based on **transaction logs**
  - **snapshotting**, **filtering, routing, flattening** etc.
  - web-based **UI**
- **large production deployments**

Currently supported Databases

- 10 databases
  - relational + non-relational
  - 8 production-ready
  - 2 incubating: Vitess + Informix
- + JDBC sink connector

Currently supported Databases

- 3 “categories” of connectors
  - core
  - community-led
  - external

- 🎊 DB support is steadily growing 🤩

Benefits of Log-based CDC

Red Hat Developer | DevNation

- low latency & little overhead
- no changes missed
- data model agnostic
- captures delete operations
- access to previous state
- get additional meta-data

Change Event Payload Structure

- Message **Key:** table's **primary key**
- Message **Value:**
  - **old & new data state**
  - **meta-data** on table, TX id, etc.
  - **operation type,** timestamp
- **Serialization** → JSON, Avro, ...
- **Cloud Events** spec also supported

```
{
  "before": {
    "id": 1004,
    "first_name": "Anne",
    "last_name": "Kretchmar",
    "email": "annek@noanswer.org"
  },
  "after": {
    "id": 1004,
    "first_name": "Anne Marie",
    "last_name": "Kretchmar",
    "email": "annek@noanswer.org"
  },
  ...
}
```

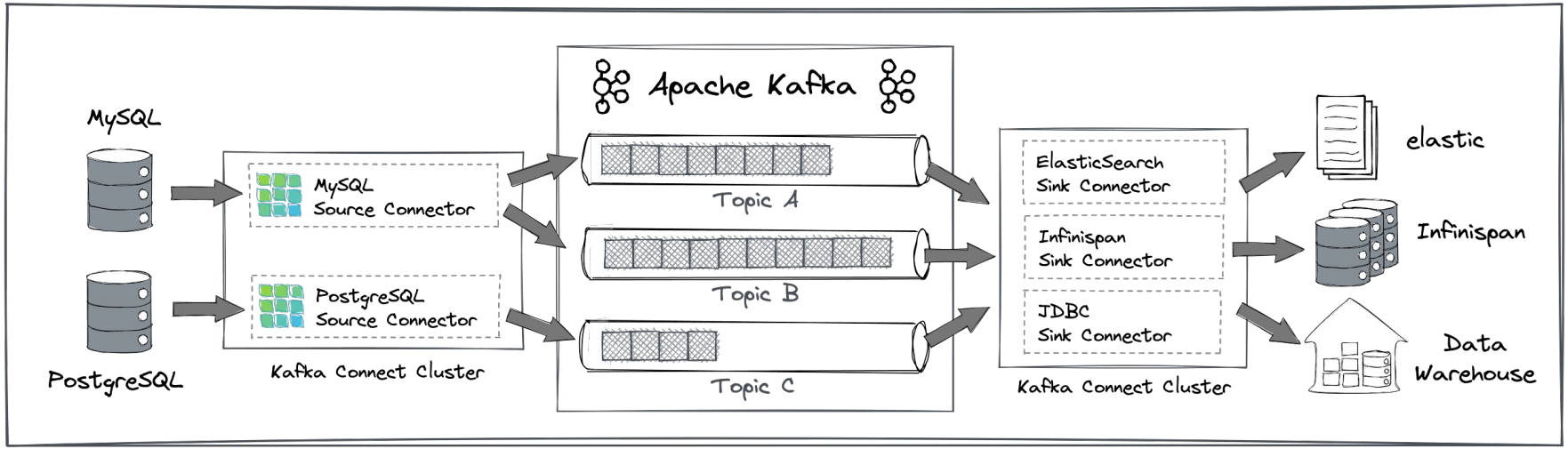
Change Event Payload Structure

- Message **Key:** table's **primary key**
- Message **Value:**
  - **old & new data state**
  - **meta-data** on table, TX id, etc.
  - **operation type,** timestamp
- **Serialization** → JSON, Avro, ...
- **Cloud Events** spec also supported

```
{
  ...
  "source": {
    "version": "2.0.1.Final",
    "connector": "postgresql",
    "name": "dbserver1",
    "ts_ms": 1686729166190,
    "snapshot": "false",
    "db": "postgres",
    "sequence": "[null,\"34488008\"]",
    "schema": "inventory",
    "table": "customers",
    "txId": 769,
    "lsn": 34488008,
    "xmin": null
  },
  "op": "u",
  "ts_ms": 1686729166517,
  "transaction": null
}
```

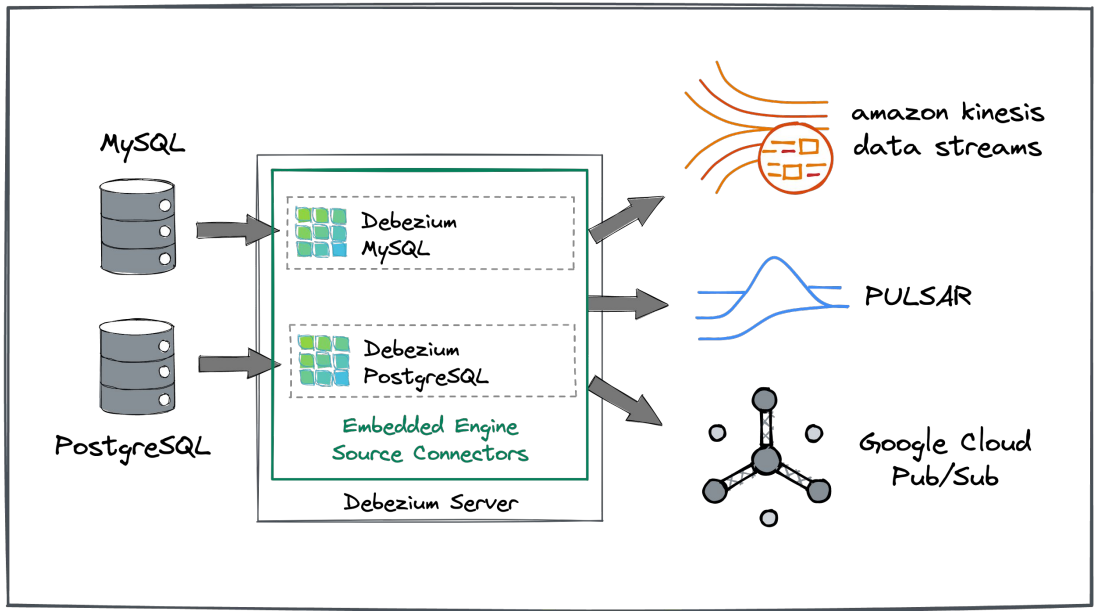
3 Deployment Modes for Debezium

Mode 1: Apache Kafka + Kafka Connect



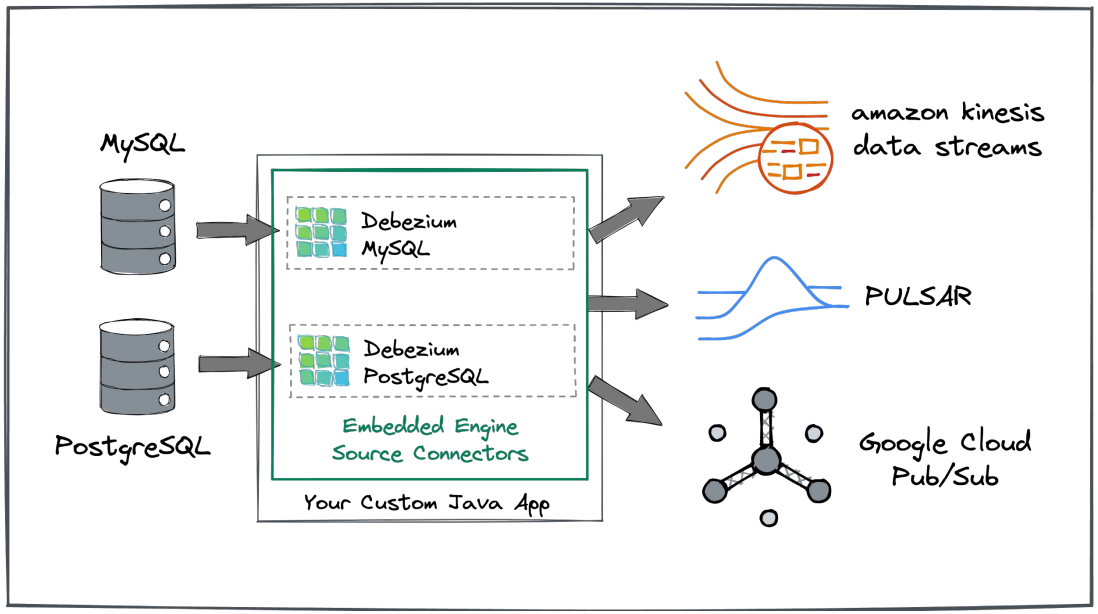
3 Deployment Modes for Debezium

Mode 2: Debezium Server



3 Deployment Modes for Debezium

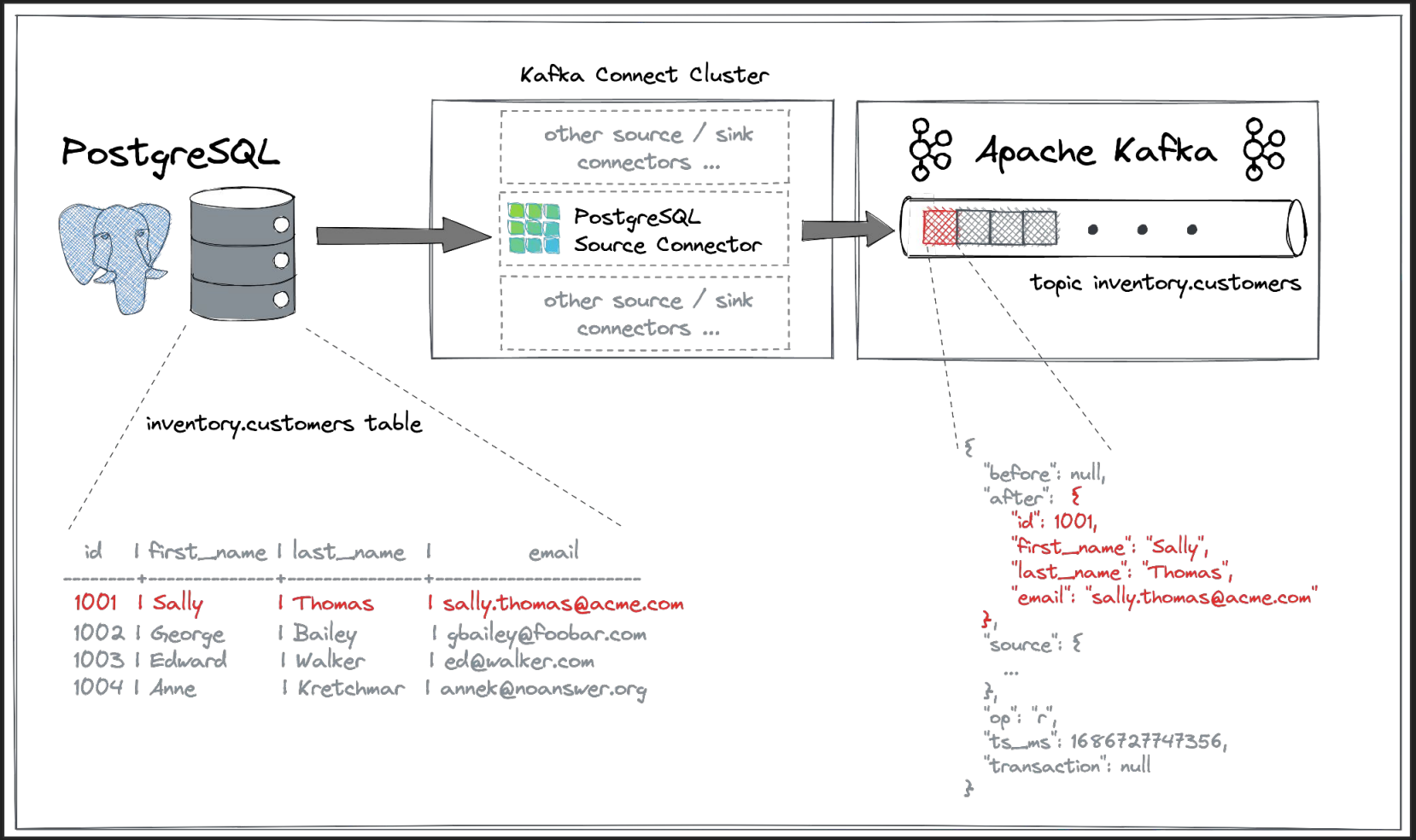
Mode 3: Debezium Embedded Engine

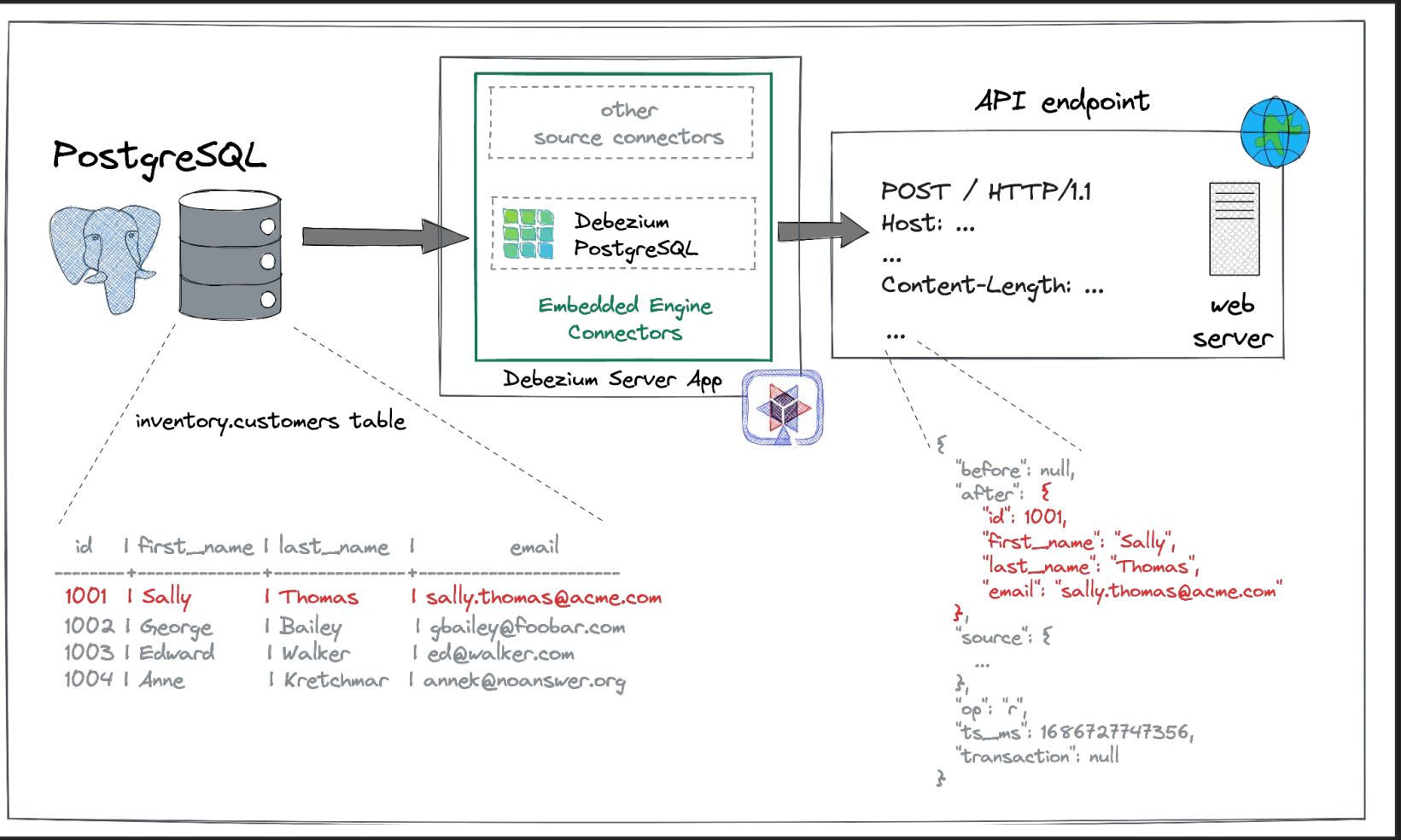


Red Hat Developer | DevNation

Debezium in Action

debezium.io


Kafka Connect Cluster
PostgreSQL
other source / sink connectors ...
PostgreSQL Source Connector
other source / sink connectors ...
Apache Kafka
topic inventory.customers
inventory.customers table
id | first_name | last_name | email
1001 | Sally | Thomas | sally.thomas@acme.com
1002 | George | Bailey | gbailey@foobar.com
1003 | Edward | Walker | ed@walker.com
1004 | Anne | Kretchmar | annek@noanswer.org
{
"before": null,
"after": {
"id": 1001,
"first_name": "Sally",
"last_name": "Thomas",
"email": "sally.thomas@acme.com"
},
"source": {
...
},
"op": "r",
"ts_ms": 1686727747356,
"transaction": null
}


PostgreSQL
other
source connectors
Debezium
PostgreSQL
Embedded Engine
Connectors
Debezium Server App
API endpoint
POST / HTTP/1.1
Host: ...
...
Content-Length: ...
...
web
server
inventory.customers table
id | first_name | last_name | email
1001 | Sally | Thomas | sally.thomas@acme.com
1002 | George | Bailey | gbailey@foobar.com
1003 | Edward | Walker | ed@walker.com
1004 | Anne | Kretchmar | annek@noanswer.org
{
"before": null,
"after": {
"id": 1001,
"first_name": "Sally",
"last_name": "Thomas",
"email": "sally.thomas@acme.com"
},
"source": {
...
},
"op": "r",
"ts_ms": 1686727747356,
"transaction": null
}

Try it → dn.dev/debezium-tutorial

Red Hat Developer | DevNation

Red Hat Developer

Debezium Tutorial

Books | Cheat Sheets | Upcoming Events | More Tutorials

1. Getting Started
   - Overview
     - Change Data Capture
     - Supported Databases
     - Deployment Modes
2. Elementary
   - Apache Kafka Connect
   - Debezium Server
   - Debezium UI

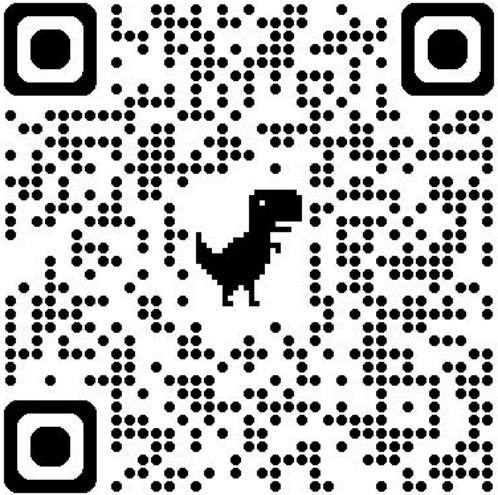
Debezium Tutorial / Welcome to the Debezium Tutorial

s Page

Welcome to the Debezium

Building Change Data Capture Pipeline

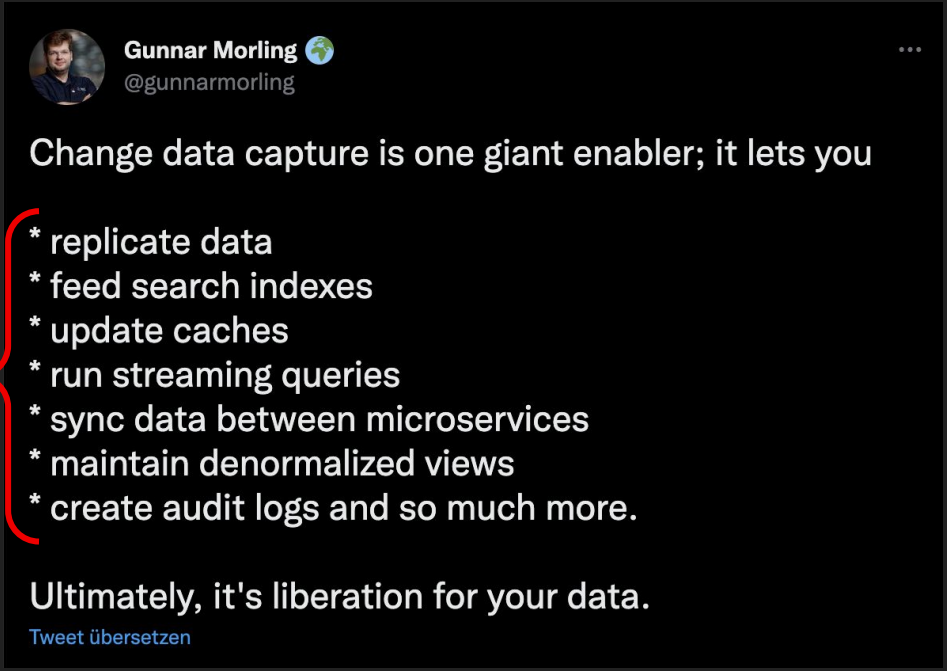
This tutorial helps you to get started with the open-sourc
will learn about selected features and different ways, how
lines using popular relational and non-relational database
stream captured change events to different messaging in
ited to Apache Kafka.



Debezium enables various Use Cases

Red Hat Developer | DevNation

🤩 All of these and many more can be powered by change data capture and Debezium! 🤩



Further CDC-related materials

- **What is change data capture (CDC)?**
  https://www.redhat.com/en/topics/integration/what-is-change-data-capture

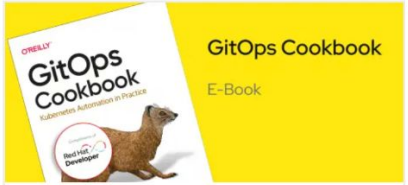
- **Using Change Data Capture for Stack Modernization**
  https://www.solutionpatterns.io/solution-pattern-modernization-cdc/solution-pattern-modernization-cdc/index.html

- **Use Case: Emerging Disease Detection**
  https://validatedpatterns.io/patterns/emerging-disease-detection/

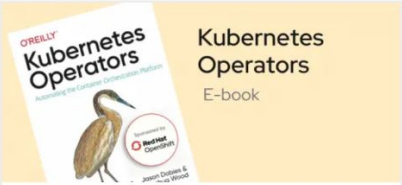
Developer Resources: Books, Tutorials,…

https://developers.redhat.com/eventtutorials

Featured Red Hat Developer e-books

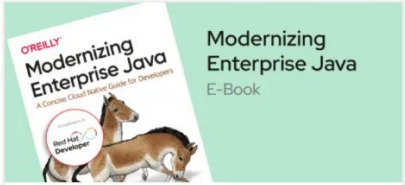


GitOps Cookbook: Kubernetes Automation in Practice



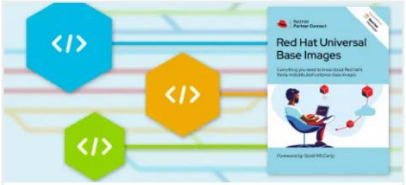
Kubernetes Operators e-book

Podman in Action



Modernizing Enterprise Java

OpenShift for Developers, Second Edition



Red Hat Universal Base Images (UBI)

Developer Portals: Prepare to Perform with Red Hat Developer

A developer's guide to setting supply chain security in DevSecOps

Red Hat Developer | DevNation

Thanks ! Questions ?

debezium.io