

Iniciación a CDC con Debezium.

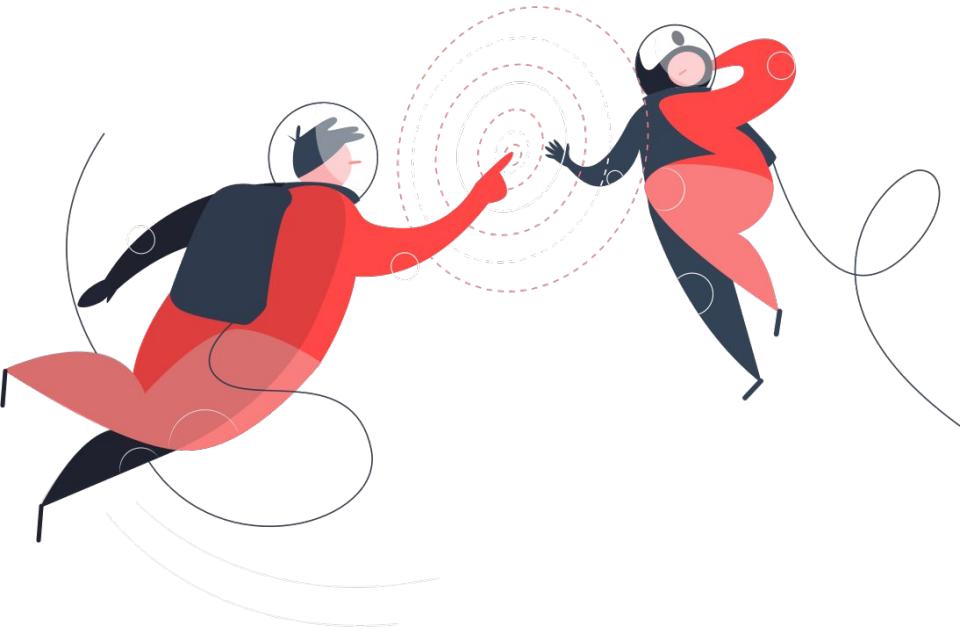
Abriendo tu CORE al tiempo real

¿Quiénes somos?



Jesús Pau de la Cruz,
Software Architect.

jpau@paradigmadigital.com
<https://www.linkedin.com/in/jesuspaulacruz/>



José Alberto Ruiz Casarrubios,
Software Architect.

jaruiz@paradigmadigital.com
<https://www.linkedin.com/in/jaruizcasarrubios/>

Qué vamos a ver.

1. Contexto y motivación.
2. ¿Qué es Change Data Capture?.
3. Debezium.
4. Live Demo
5. Conclusiones

01.

• • •

Iniciación a CDC con Debezium: Abriendo tu core al tiempo real

Contexto y motivación

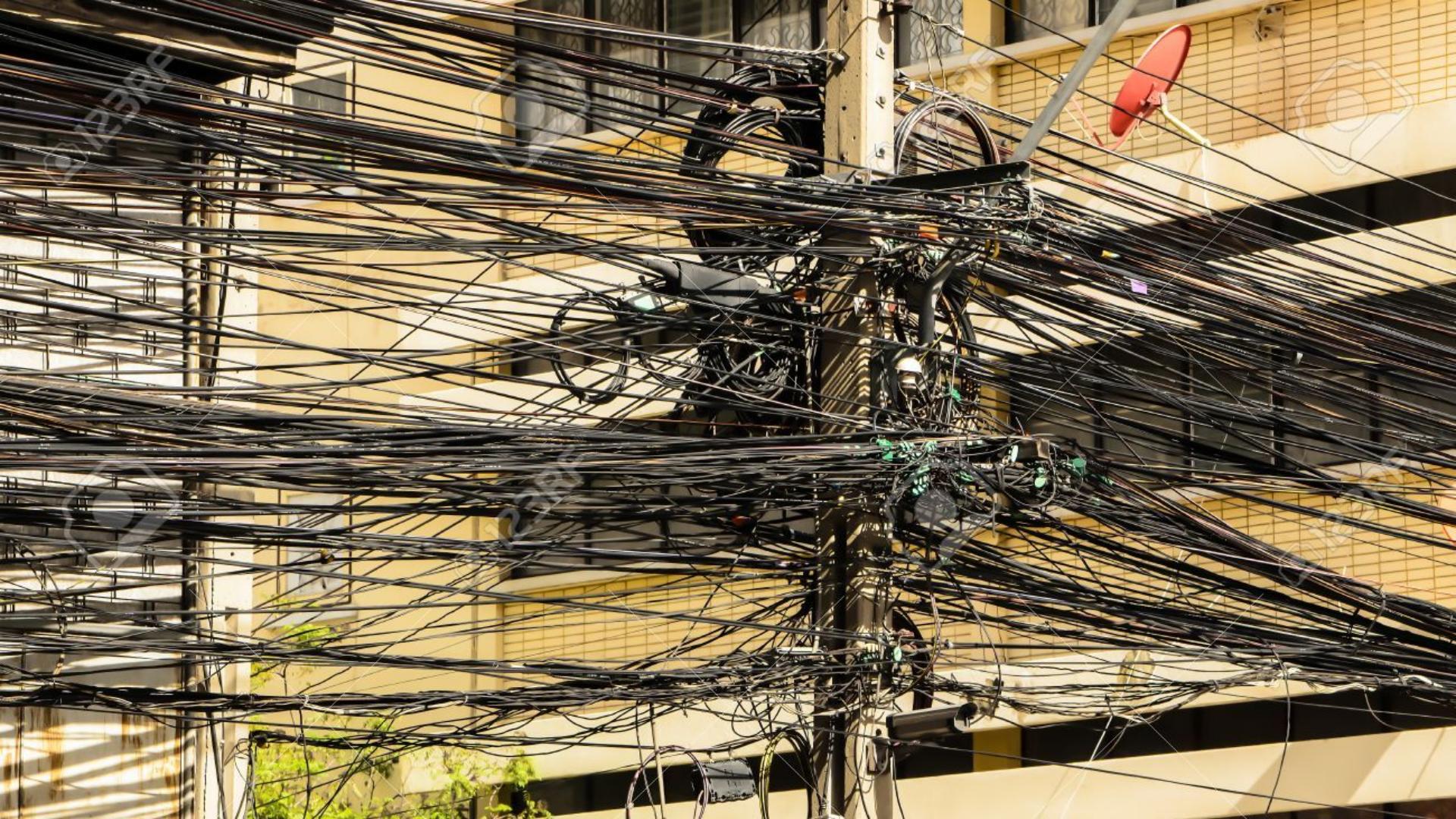


Ya manejamos mucha
información pero ahora

...

**necesitamos
hacerlo en
TIEMPO REAL**







Múltiples orígenes de datos



Falta de escalabilidad



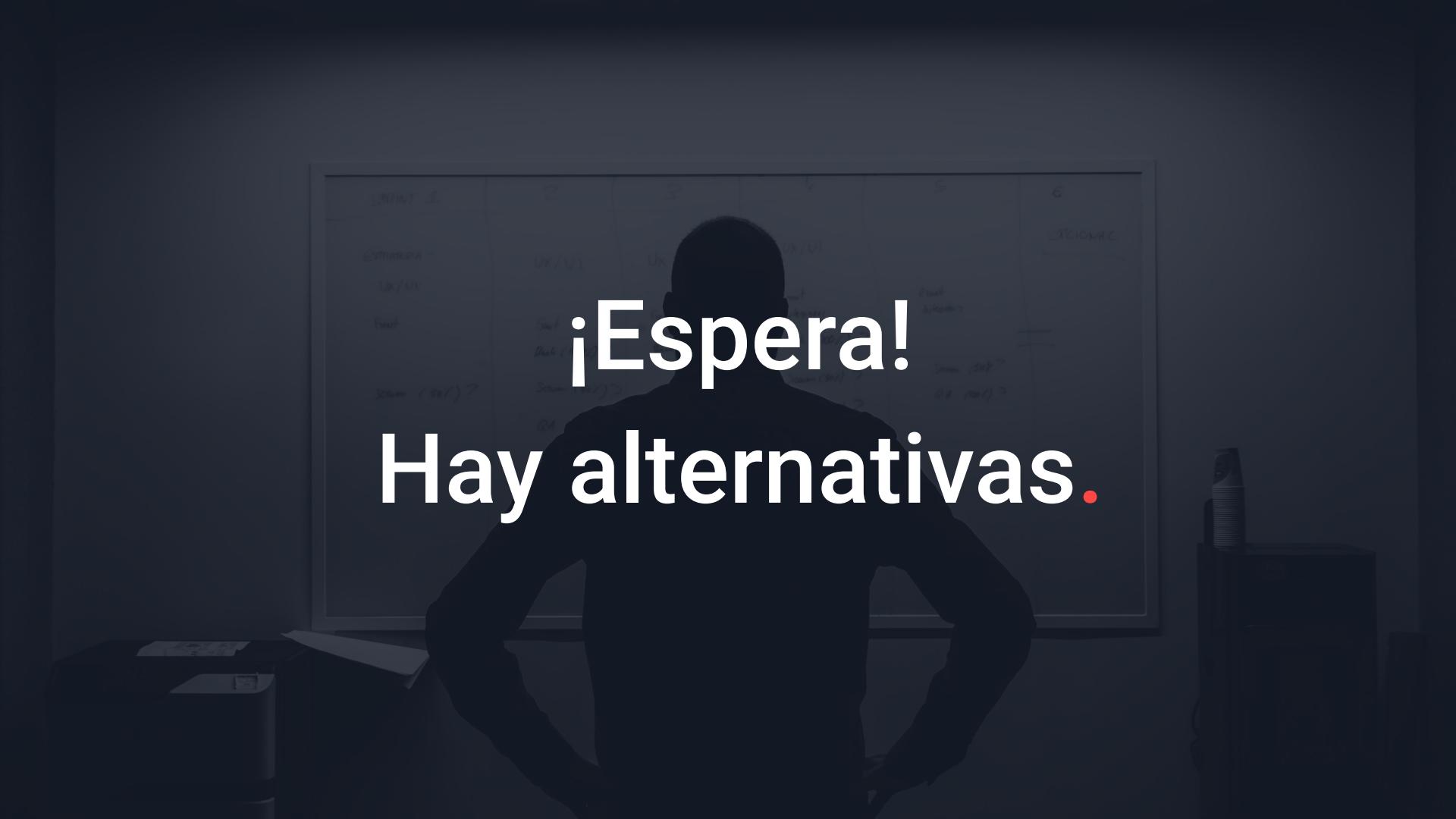
Evolución complicada



Conciliación y replicación pesada

**Es muy
complicado.**



A dark, semi-transparent background image shows a person from the waist up, sitting at a desk and facing the camera. They appear to be in an office or study environment. In the background, there's a whiteboard with various handwritten notes and diagrams, including words like 'STRATEGY', 'UX/UI', 'Front', 'Back', 'Scrum (scrum) ?', 'QA', 'Event', 'Sprint', 'Sprint (sprint) ?', 'QA (qa) ?', and 'Gd (gd) ?'.

¡Espera!
Hay alternativas.

02.

• • •

Iniciación a CDC con Debezium: Abriendo tu core al tiempo real

¿Qué es Change Data Capture? (CDC)

“

Change Data Capture: En las bases de datos, la captura de datos modificados es un conjunto de patrones de diseño de software que se utilizan para determinar y rastrear los datos que han cambiado para que se puedan tomar medidas utilizando los datos modificados.



“

Change Data Capture: En las bases de datos, la captura de datos modificados es un conjunto de patrones de diseño de software que se utilizan para determinar y rastrear los datos que han cambiado para que **se puedan tomar medidas utilizando los datos modificados. EN TIEMPO REAL**



¿Qué aplicaciones tiene CDC?



Replicación en tiempo real.

Disponer de copias sincronizadas en tiempo real para :

- analítica,
- backup o DR
- Data Lake
- Data Warehouse



Modernización de aplicaciones

Dentro de la estrategia de modernización se puede aplicar en:

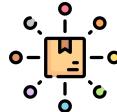
- migración a Cloud
- ruptura de monolito
- propagación de datos sin acoplamiento en microservicios
- Actualización de caches



Mejorar la escalabilidad.

Al replicar en tiempo real posibilita:

- escalabilidad horizontal de los motores de datos
- actualizar modelos de lectura (CQRS)
- eliminar “ventanas batch”, ETLs o procesos pesados



Event Streaming y Sourcing

Mediante CDC y el bus de eventos generamos un streaming de eventos de las diferentes fuentes replicadas que podemos manejar en tiempo real. Además, tenemos todo el historial de cambios hasta llegar al estado actual (event sourcing)



Procesar diferentes orígenes

CDC se puede aplicar a varios orígenes a la vez, permitiendo:

- vistas enriquecidas con datos de varias fuentes, actualizadas en tiempo real
- procesos de analíticas avanzada
- simplificación de ETLs y procesos de conciliación

¿Qué debemos tener en cuenta?



Datos en crudo

Los cambios que se propagan son los realizados en base de datos, a nivel de fila, con lo que no se dispone de un contexto de Negocio asociado en el que se producen.



Nuevas piezas que mantener

Aunque son sistemas no intrusivos con las aplicaciones existentes, estamos incrementando el número de componentes globales y sus dependencias. Debemos establecer mecanismos de monitorización y trazabilidad



Diseño de replicación.

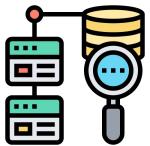
Para que sea eficiente se debe diseñar muy bien el proceso de replicación, incluyendo qué datos se quiere replicar y a qué destino(s) se van a enviar



Cambios de modelo.

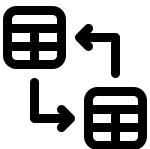
El proceso de replicación está totalmente asociado a la estructura del sistema origen. Cualquier cambio de modelo se debe gestionar para mantener el proceso de replicación funcionando de forma correcta

Tipos de CDC



Query-Based

Se basan en la existencia de columnas en las tablas que indican cuándo ha sido la última vez que ha cambiado un dato. Es intrusivo, afecta a la base de datos, depende de las aplicaciones para actualizar el timestamp y los DELETES son problemáticos



Trigger-Based

Se basa en la actualización de una tablas “shadow” en las que se recogen los cambios efectuados en las tablas principales. No depende de las aplicaciones pero afecta al rendimiento de la base de datos



Log-Based

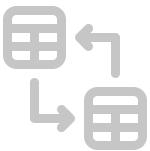
Se basan en el acceso y lectura de los logs transaccionales generados por las propias bases de datos. La mayoría de las bases de datos actuales disponen de esta característica

Tipos de CDC



Query-Based

Se basan en la existencia de columnas en las tablas que indican cuándo ha sido la última vez que ha cambiado un dato. Es intrusivo, afecta a la base de datos, depende de las aplicaciones para actualizar el timestamp y los DELETES son problemáticos



Trigger-Based

Se basa en la actualización de una tablas “shadow” en las que se recogen los cambios efectuados en las tablas principales. No depende de las aplicaciones pero afecta al rendimiento de la base de datos



Log-Based

Se basan en el acceso y lectura de los logs transaccionales generados por las propias bases de datos. La mayoría de las bases de datos actuales disponen de esta característica



¿Por qué CDC & log-based?



Mínimamente intrusivo.

Son elementos que se instalan de forma independiente, no necesitando (prácticamente) modificar el sistema actual para que puedan funcionar.



No penaliza el sistema.

Se basan en la lectura de los ficheros de logs, evitando la necesidad de polling o peticiones constantes al motor de base de datos



Volumen de datos contenido

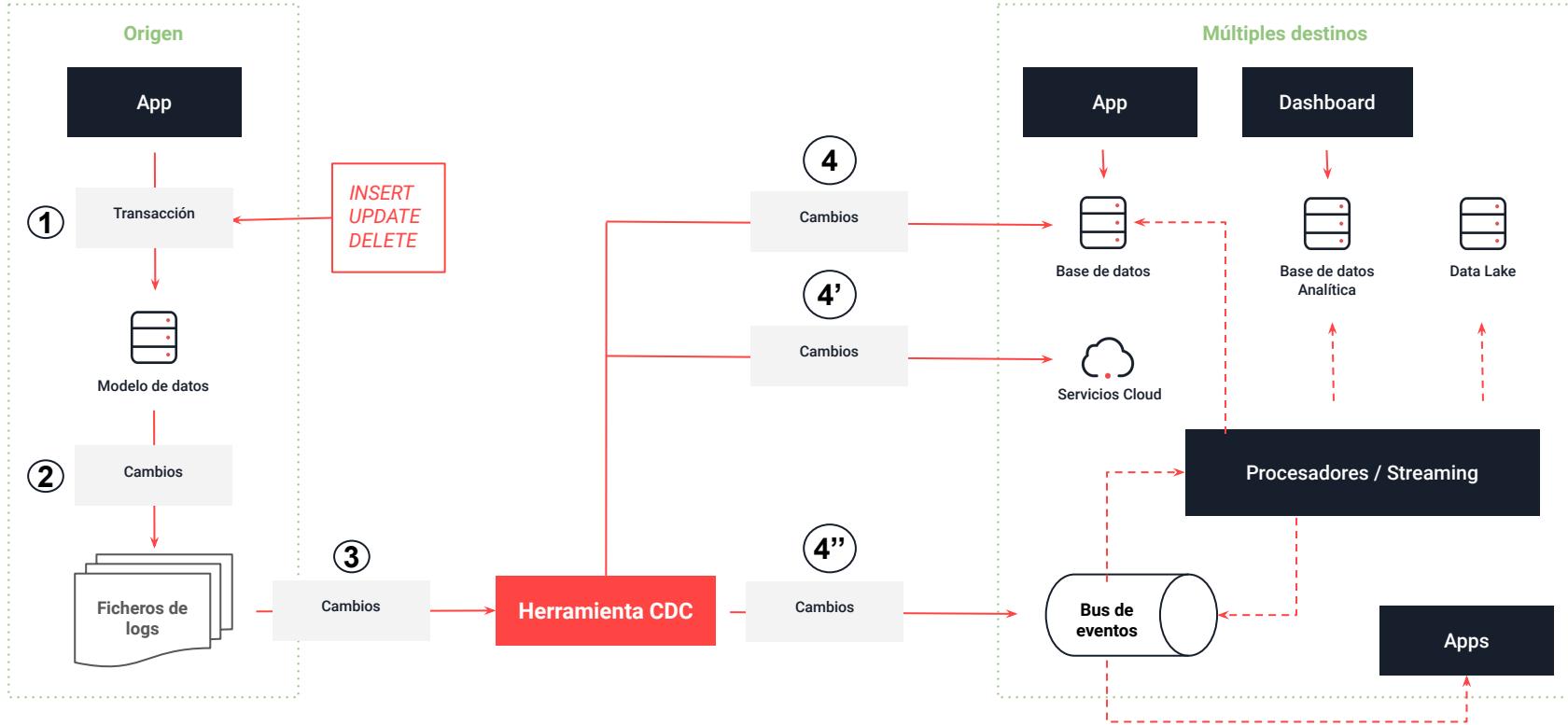
Al manejar solo incrementos se consigue que los datos que viajan entre sistemas tengan un volumen menor



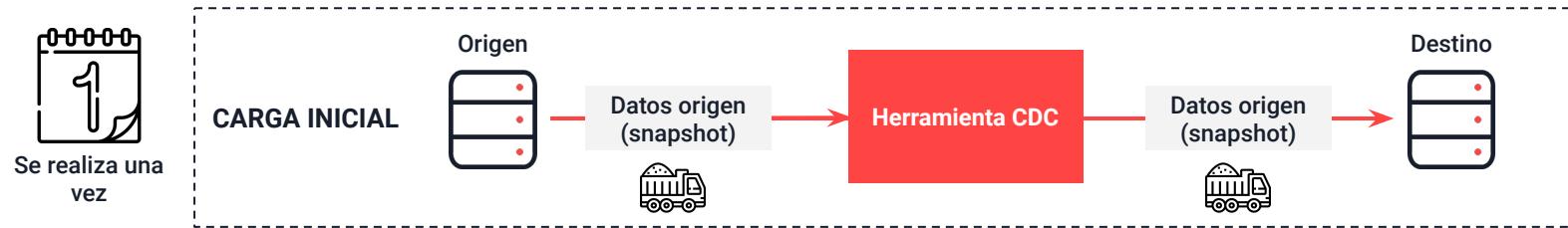
Tiempo real

Los cambios son propagados en el momento en el que se producen, ayudando a la toma de decisiones

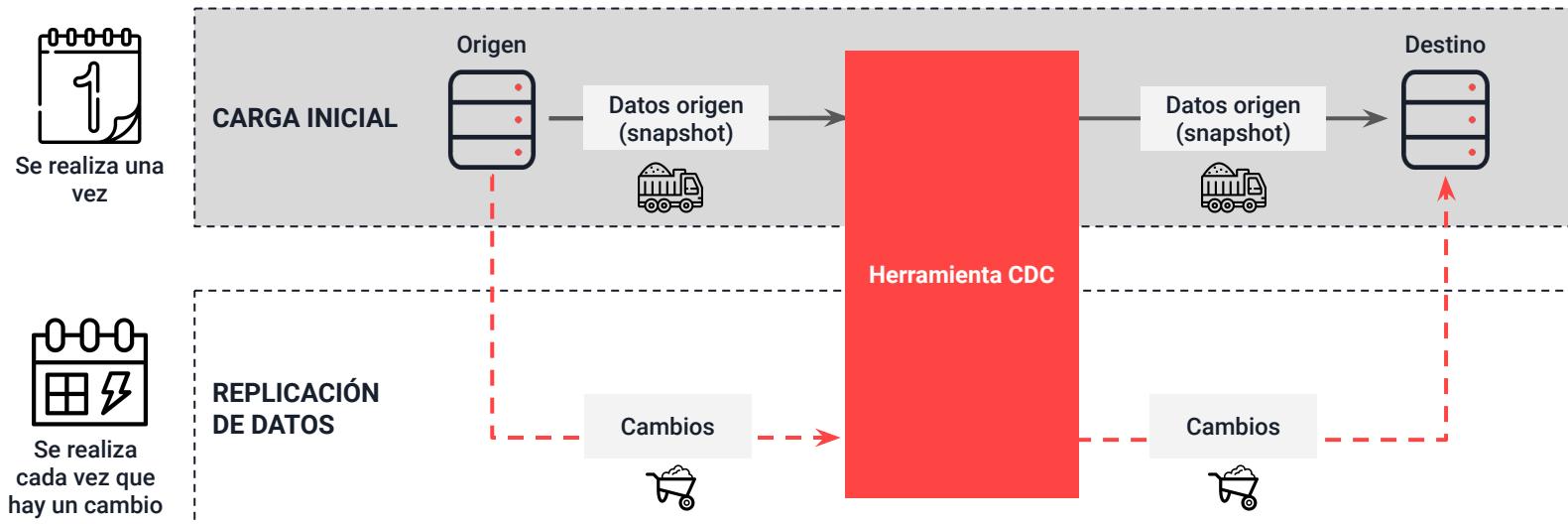
¿Cómo funciona?



Carga inicial y replicación



Carga inicial y replicación



03.

• • •

Iniciación a CDC con Debezium: Abriendo tu core al tiempo real

Debezium

¿Qué es Debezium?

Debezium es una plataforma distribuida, Open Source, para realizar Change Data Capture.



Sponsored by
Red Hat

Esponsorizado por Red Hat

Debezium es un proyecto Open Source esponsorizado por Red Hat e incluido dentro de sus soluciones para integración de datos



Kafka & Kafka Connect

Debezium se apoya totalmente en Apache Kafka y Apache Kafka Connect, proporcionando conectores capaces de monitorizar los cambios en bases de datos y replicarlos sobre Apache Kafka



CDC Log based

La mayoría de conectores de Debezium acceden a los ficheros de logs de las bases de datos para detectar los cambios producidos. No añade overhead al motor de base de datos

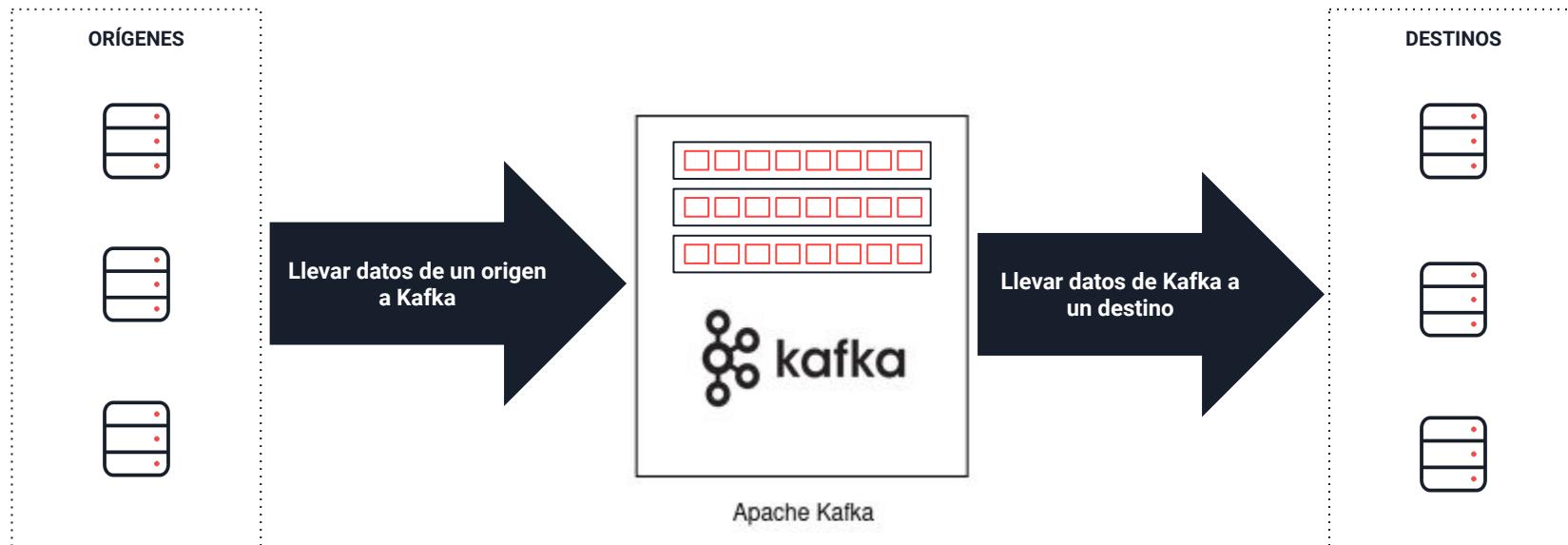


Múltiples bases de Datos

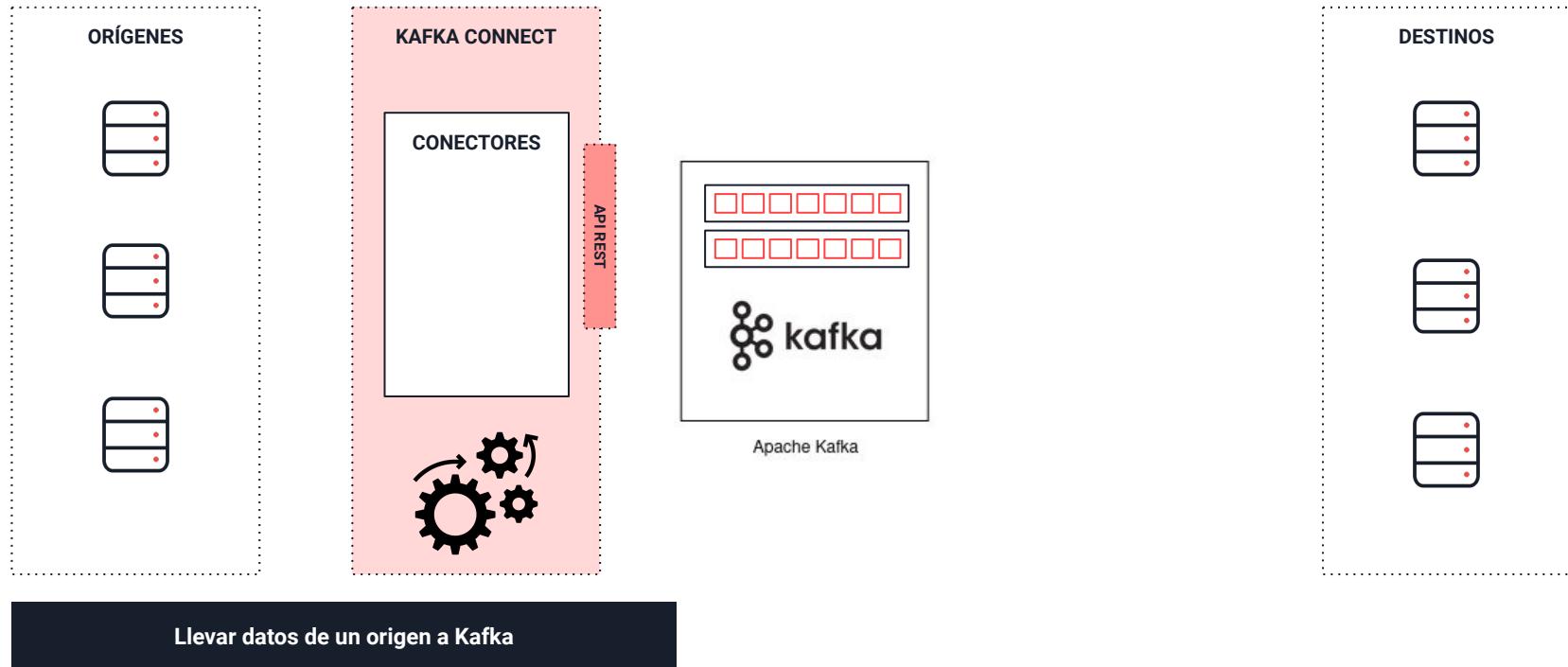
Actualmente, Debezium proporciona conectores para los principales motores de bases de datos. Añaden nuevos conectores cada cierto tiempo.

Kafka Connect

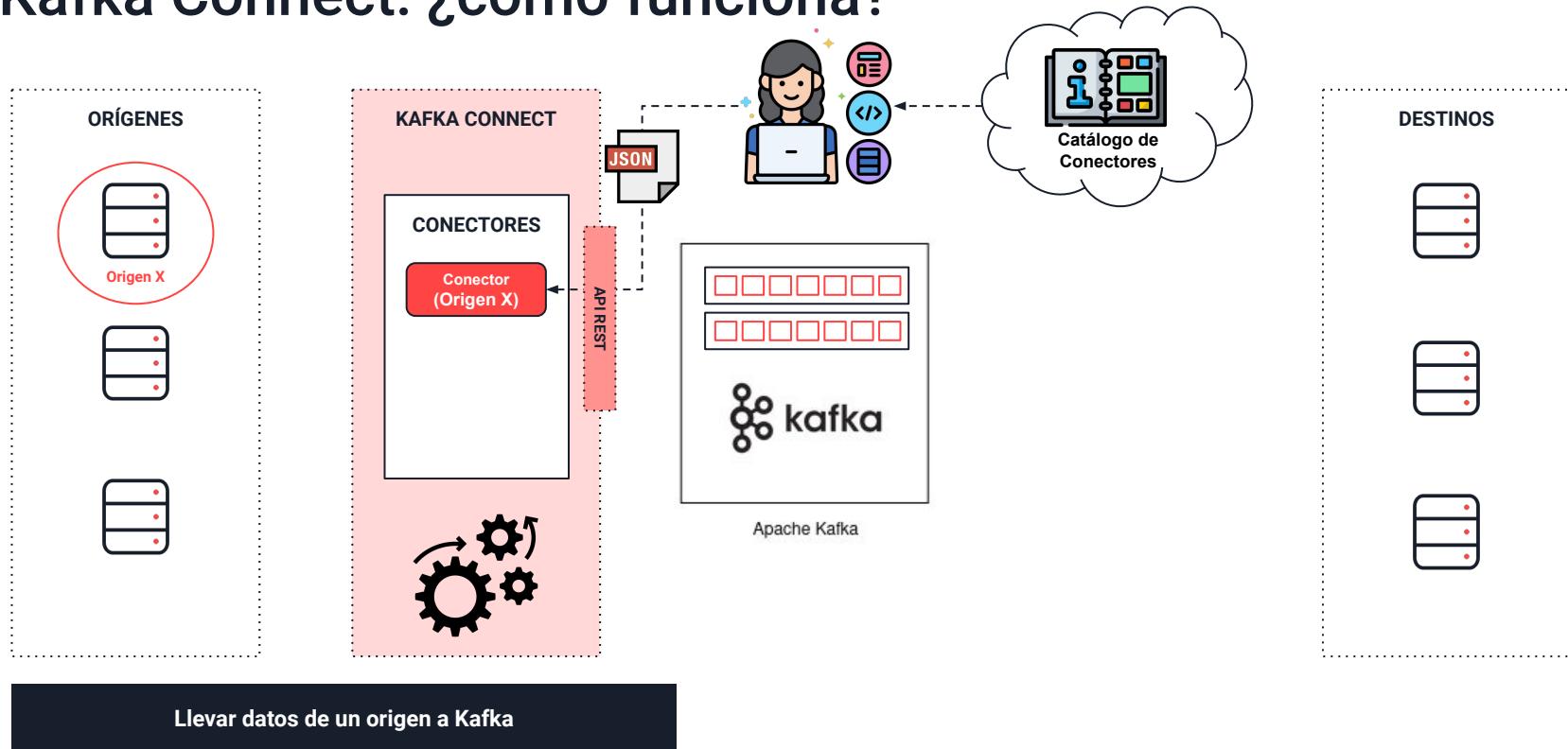
Kafka Connect es parte de Kafka. Se usa para llevar datos desde un origen a Kafka (“data source”) como desde Kafka a un destino (“data sink”) utilizando una serie de componentes pre-construidos, llamados conectores



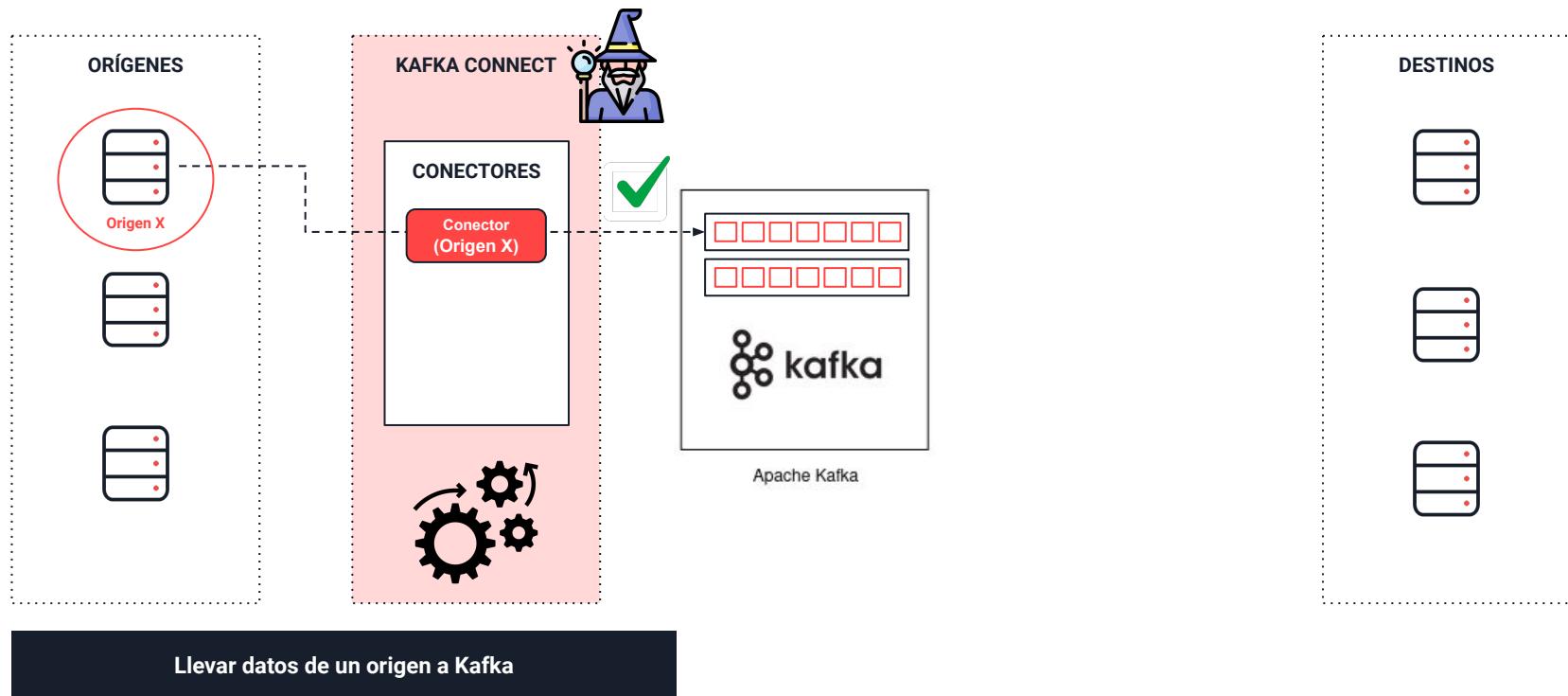
Kafka Connect: ¿cómo funciona?



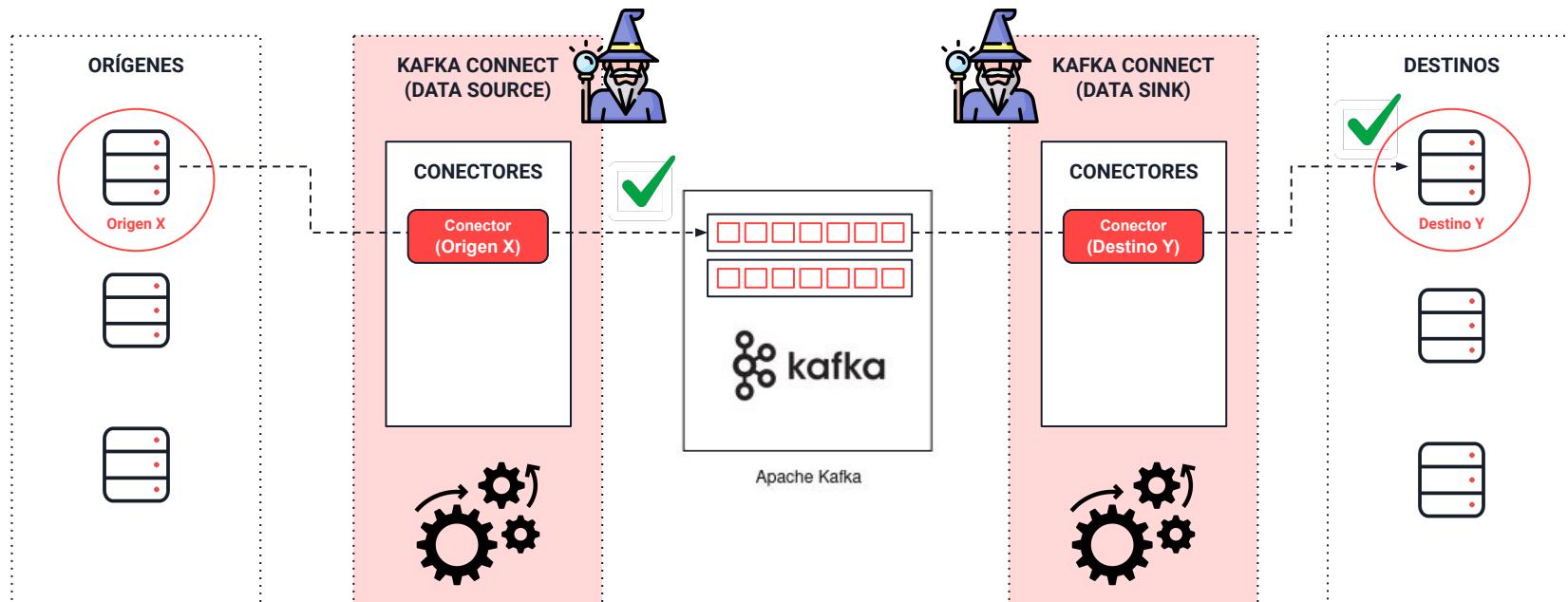
Kafka Connect: ¿cómo funciona?



Kafka Connect: ¿cómo funciona?



Kafka Connect: ¿cómo funciona?

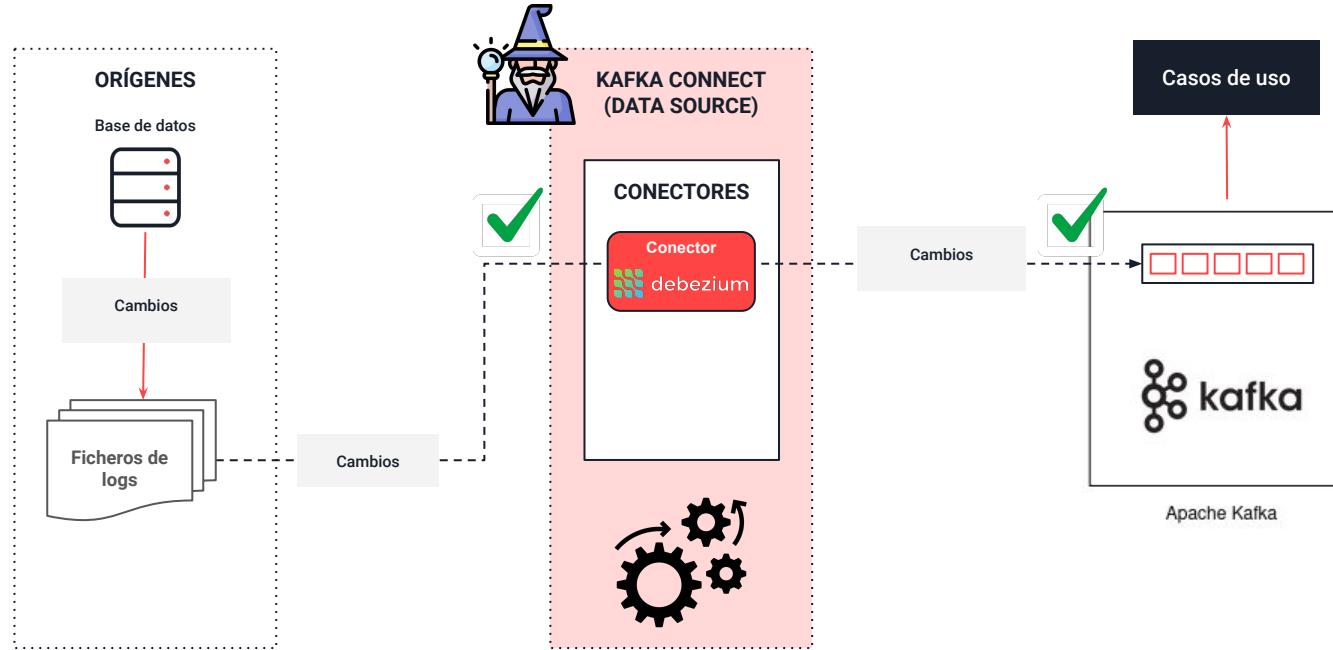


Llevar datos de un origen a Kafka

Llevar datos de Kafka a un destino

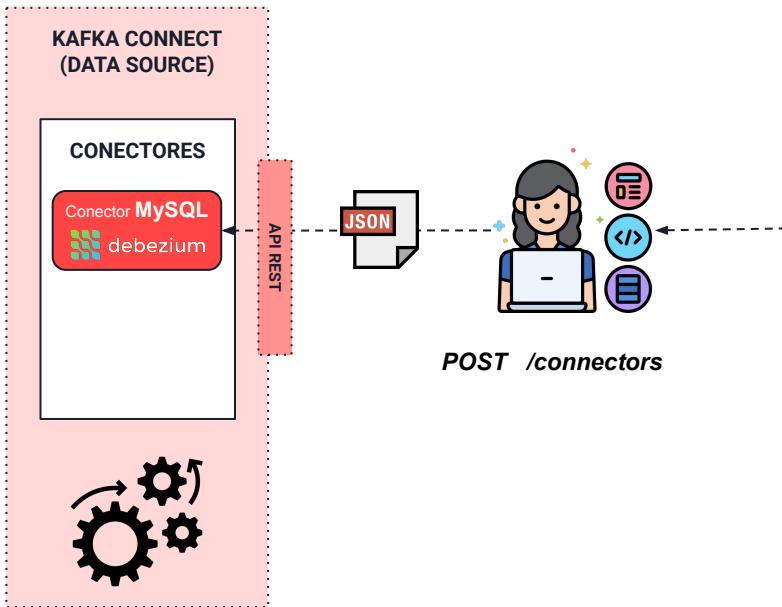
¿Cómo funciona Debezium?

Debezium se basa en **conectores para Kafka Connect** que implementan la lógica de obtención de cambios en las distintas bases de datos que soporta y llevan esa información a Kafka



¿Cómo funciona Debezium? Catálogo de conectores

En la documentación de Debezium puedes encontrar la lista de conectores disponibles en cada versión y todos los detalles de uso



Catálogo de Conectores de Debezium

Connectors

MySQL

Debezum connector for MySQL

A new capturing implementation for the Debezum MySQL connector has been created as of Debezum 1.5, based on the common connector framework used by all the other Kafka Connect connectors of Debezum.

MySQL has a binary log (binlog). The Debezum MySQL connector As MySQL is typically set up to process events from the binlog.

Connector properties

The Debezum MySQL connector has numerous configuration properties that you can use to achieve the right connector behavior for your application. Many properties have default values. Information about the properties is organized as follows:

- Required connector configuration properties
- Advanced connector configuration properties
- MySQL database driver configuration properties that control how Debezum processes events that it reads from the database history topic.
- Passthrough database driver properties
- Passthrough database driver properties that control the behavior of the database driver.

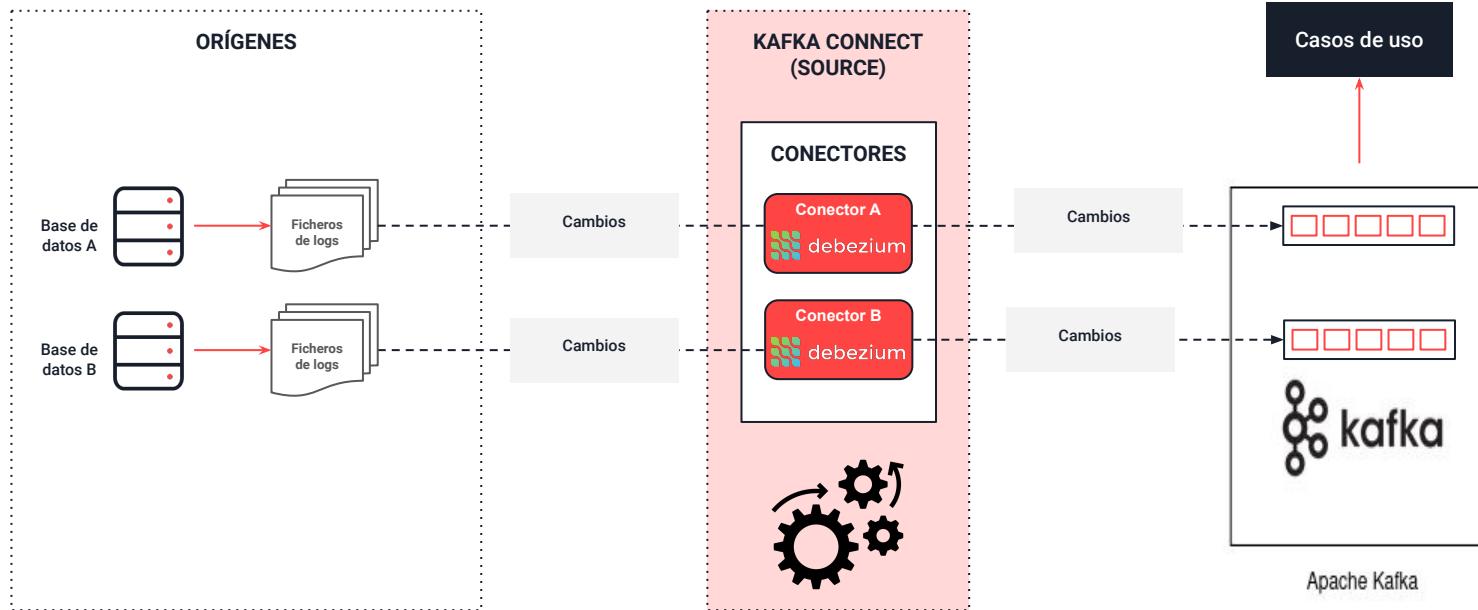
The following configuration properties are required unless a default value is available.

Related Debezum MySQL connector configuration properties

Property	Default	Description
name	No default	Unique name for the connector. Attempting to register again with the same name fails. This property is required by all Kafka Connect connectors.
connector.class	No default	The fully qualified name of the Java class for the connector. Always specify <code>io.debezium.connector.mysql.MySqlConnector</code> for the MySQL connector.
tasks.max	1	The maximum number of tasks that should be created for this connector. The MySQL connector always uses a single task and therefore does not use the value, so the default is always acceptable.
database.hostname	No default	IP address or host name of the MySQL database server.
database.port	3306	Integer port number of the MySQL database server.
database.user	No default	Name of the MySQL user to use when connecting to the MySQL database server.
database.password	No default	Password to use when connecting to the MySQL database server.
database.server.name	No default	Logical name that identifies and provides a namespace for the particular MySQL database server/master in which Debezum is capturing changes. The logical name should be unique across all other connectors, since it is used as a prefix for all Kafka topic names generated by this connector. Only alphanumeric characters, hyphens, dots and underscores must be used in the database server logical name.

¿Y si tengo varias fuentes que replicar?

Se pueden registrar múltiples conectores dependiendo del número de fuentes que se quieran replicar



¿Cómo es un conector de Debezium?



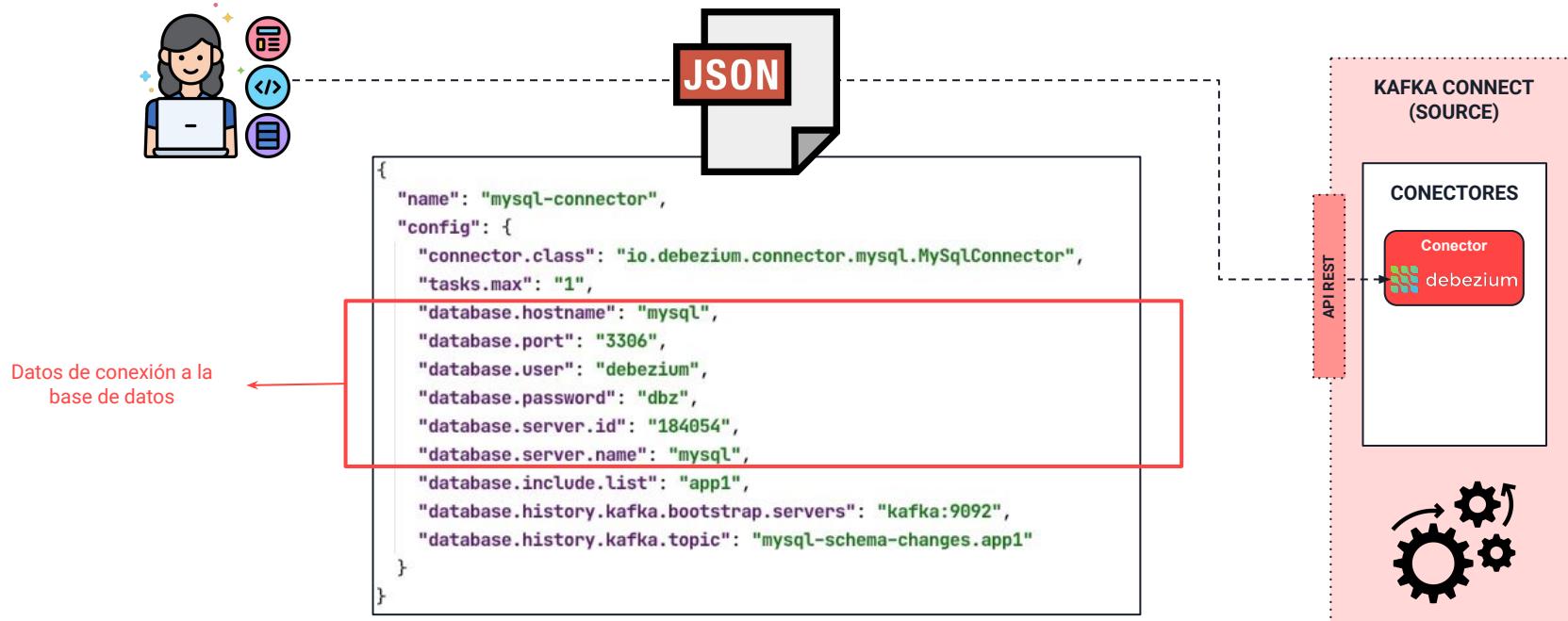
¿Cómo es un conector de Debezium?



¿Cómo es un conector de Debezium?



¿Cómo es un conector de Debezium?



¿Cómo es un conector de Debezium?



¿Cómo son los mensajes que publica Debezium?

Debezium nos permite capturar los cambios que se producen en una base de datos. ¿Qué nos interesa saber cuando hablamos de un cambio?

CLAVE = ¿QUÉ REGISTRO HA CAMBIADO?	
ESQUEMA (DEL PAYLOAD DE LA CLAVE)	
PAYLOAD (CLAVE)	<p>¿Cuál es el identificador del registro?</p> <pre>CLAVE</pre>
CONTENIDO = ¿QUÉ DATOS HAN CAMBIADO?	
ESQUEMA (DEL PAYLOAD DEL CONTENIDO)	
PAYLOAD (CONTENIDO)	<p>CONTENIDO = ¿QUÉ DATOS HAN CAMBIADO?</p> <ul style="list-style-type: none">- ¿qué valor tenía antes?- ¿qué valor tiene después?- ¿dónde se ha hecho la operación?- ¿qué operación se ha hecho? <pre>BEFORE AFTER SOURCE DATOS OPERACIÓN</pre>

Otros puntos de interés



Transformaciones

Debezium proporciona “out-of-the-box” una serie de transformaciones con las que podemos hacer filtrado de mensajes o enviar los cambios a diferentes topics



Debezium Server

Todavía en desarrollo. Consiste en una aplicación que nos permitirá procesar cambios en fuentes de datos y replicarlos en destinos “no Kafka” como AWS Kinesis, Google Cloud Pub/Sub, Apache Pulsar or Redis (Stream)



Debezium UI

Con la versión 1.7 se ha incluido la primera release de la interfaz gráfica de Debezium. Actualmente solo permite trabajar con los conectores de MySQL, PostgreSQL, MongoDB pero incluirán el resto en próximas versiones

¿Puedo hacer CDC con otros conectores “no Debezium”?



Home > Connect External Systems to Confluent Cloud >

Microsoft SQL Server Source (JDBC) Connector for Confluent Cloud

Note
If you are installing the connector locally for Confluent Platform, see JDBC Connector (Source and Sink) for Confluent Platform.

Home > Connect External Systems to Confluent Cloud >

MySQL Source (JDBC) Connector for Confluent Cloud

Note
If you are installing the connector locally for Confluent Platform, see JDBC Connector (Source and Sink) for Confluent Platform.

Home > Connect External Systems to Confluent Cloud >

PostgreSQL Source (JDBC) Connector for Confluent Cloud

Note
If you are installing the connector locally for Confluent Platform, see JDBC Connector (Source and Sink) for Confluent Platform.

Home > Connect External Systems to Confluent Cloud >

Oracle Database Source Connector for Confluent Cloud

Note
If you are installing the connector locally for Confluent Platform, see JDBC Connector (Source and Sink) for Confluent Platform.

¿Qué tengo que tener en cuenta al elegir un conector?

Hay que tener en cuenta que muchos conectores, incluido alguno de Debezium, no son “log based” por lo que afectan al rendimiento del motor de base de datos. Estudia los requerimientos y el volumen de operaciones y datos a replicar



Home > Connect External Systems to Confluent Cloud

MySQL source (JDBC) Connector for Confluent Cloud

Note

If you are installing the connector locally for Confluent Platform, see JDBC Connector (Source and Sink) for Confluent Platform.

Debezium Starter Kit



jaruizes Added transformations 8e09158 11 days ago 4 commits

connectors	Added transformations	11 days ago
doc/img	Added transformations	11 days ago
infra	Added transformations	11 days ago
.gitignore	Initial import	19 days ago
README.md	Added transformations	11 days ago
docker-compose.yaml	Added transformations	11 days ago

README.md

Debezium Starter Kit

En este repositorio se puede encontrar una PoC lista para poder empezar a trabajar con Debezium. La versión seleccionada es la 1.7.

Features

Dentro de las features que podrás encontrar en la PoC se encuentran:

- Entorno local usando Docker Compose
- Replicación desde los siguientes orígenes de datos: MySQL y Postgresql
- Kafka UI para que puedas ver los mensajes de replicación que llegan a Kafka
- Debezium UI, liberada en la versión 1.7 de Debezium, para gestionar los conectores de forma gráfica
- Adminer, para poder realizar cambios en base de datos sin necesitar ningún cliente adicional

<https://github.com/paradigmadigital/debezium>

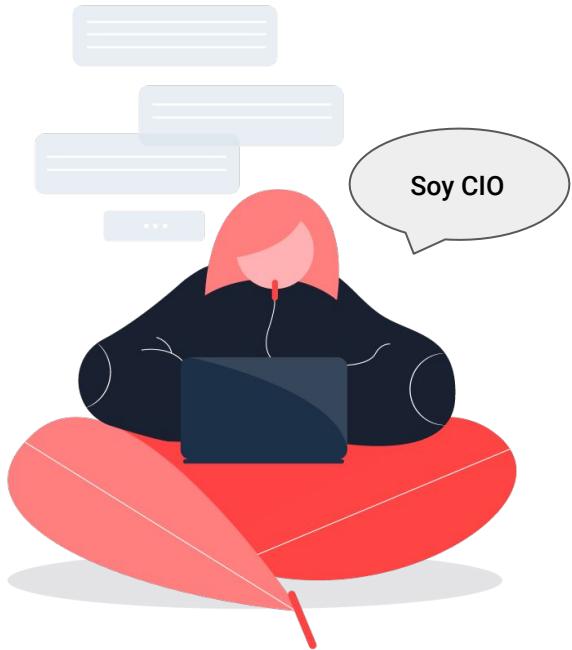
04.

•••

Iniciación a CDC con Debezium: Abriendo tu core al tiempo real



Introducción a CDC con Debezium (Live Demo)



04.01

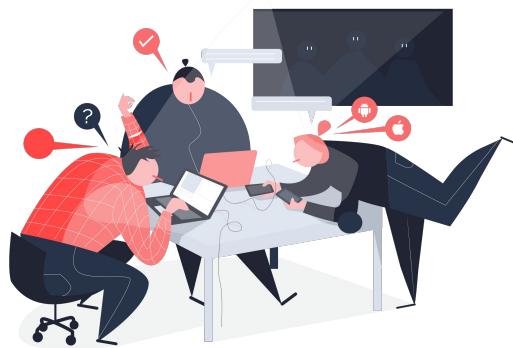
•••

Iniciación a CDC con Debezium: Abriendo tu core al tiempo real

Live Demo

Contexto y motivación.

Ecosistema de aplicaciones.



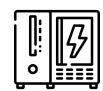
Facturación



Captación de clientes



Pedidos/Altas



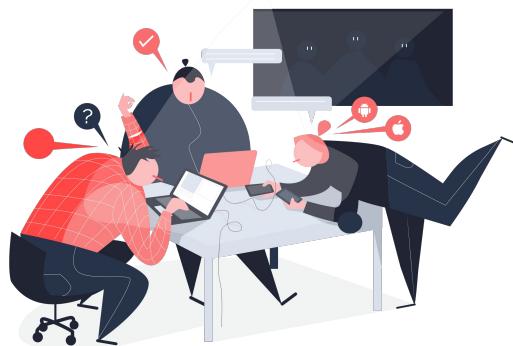
Infraestructura



Atención al cliente

...

Ecosistema de aplicaciones.



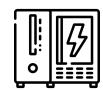
Facturación



Captación de clientes



Pedidos/Altas



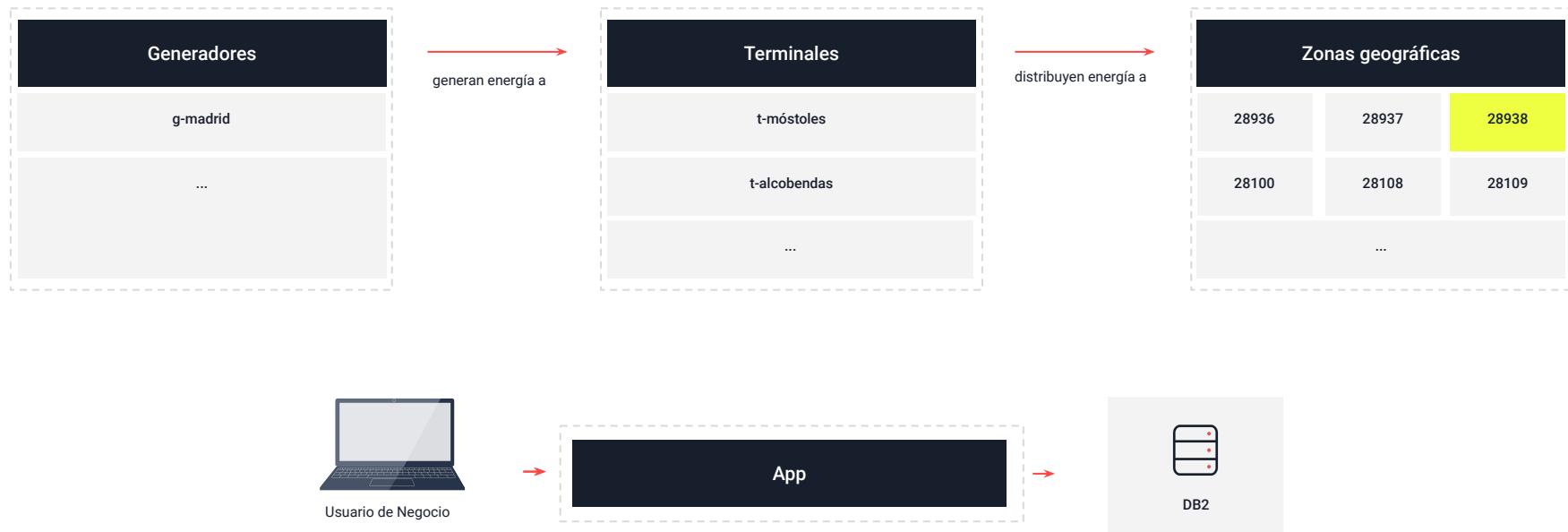
Infraestructura



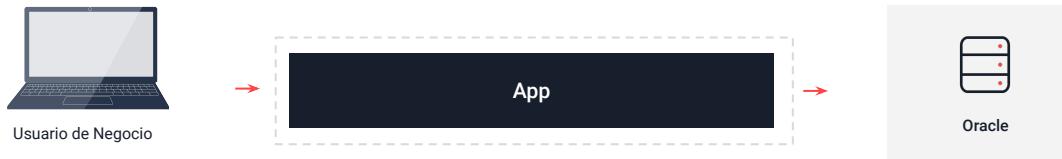
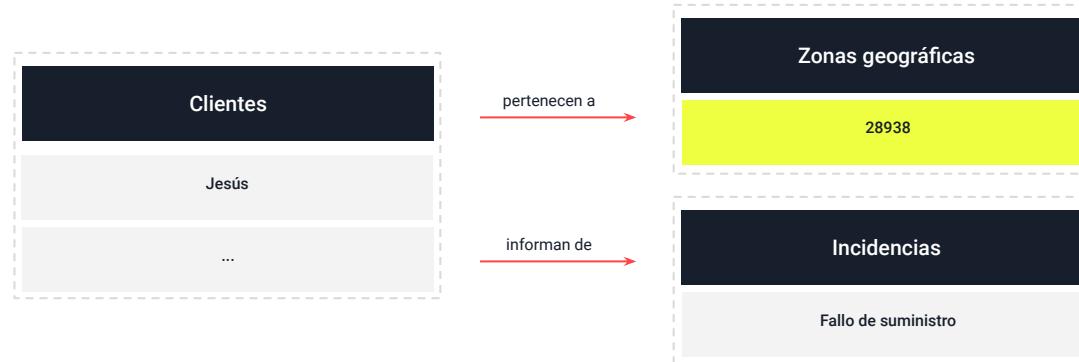
Atención al cliente

...

Estado actual de la gestión de la infraestructura.



Estado actual de la gestión de att. al cliente.



Necesidades de negocio

Visión global en **tiempo real** del estado de la infraestructura

Notificación en **tiempo real** de cambios en las incidencias de clientes

Mejora de rendimiento del sistema de gestión de incidencias

Visión en **tiempo real** del estado de las incidencias notificadas por los clientes

Sistema de alertas en **tiempo real** para informar de problemas en la infraestructura

Funcionalidades adicionales para el sistema de gestión de incidencias

...



...

Escenario técnico complicado



Tecnología obsoleta



Evolución compleja



Sistemas no escalables



Infraestructura



Atención al cliente

Silos de información



Procesos pesados





KEEP
CALM
AND
REAL
TIME

04.02

••• Iniciación a CDC con Debezium: Abriendo tu core al tiempo real ————— Live Demo

Estrategia.

No queremos un Big Bang



Priorización de necesidades de negocio

Necesidad de negocio	Módulo/Aplicación	
Visión global en tiempo real del estado de la infraestructura.	Infraestructura	
Sistema de alertas en tiempo real para informar de problemas en la infraestructura.	Infraestructura	
Visión en tiempo real del estado de las incidencias notificadas por los clientes.	Atención al cliente	
Sistema de alertas en tiempo real para notificar cambios en las incidencias de clientes.	Atención al cliente	
Mejora de rendimiento del sistema de gestión de incidencias.	Atención al cliente	
Implementar funcionalidades adicionales para el sistema de gestión de incidencias.	Atención al cliente	
...		

Priorización de necesidades de negocio

Necesidad de negocio	Módulo/Aplicación
Visión global en tiempo real del estado de la infraestructura.	Infraestructura
Sistema de alertas en tiempo real para informar de problemas en la infraestructura.	Infraestructura
Visión en tiempo real del estado de las incidencias notificadas por los clientes.	Atención al cliente
Sistema de alertas en tiempo real para notificar cambios en las incidencias de clientes.	Atención al cliente
Mejora de rendimiento del sistema de gestión de incidencias.	Atención al cliente
Implementar funcionalidades adicionales para el sistema de gestión de incidencias.	Atención al cliente
...	

Roadmap

		Necesidad de negocio	Módulo/Aplicación		
Iteración I	Visión global en tiempo real del estado de la infraestructura.	Infraestructura			
Iteración II					
Iteración III	Visión en tiempo real del estado de las incidencias notificadas por los clientes.	Atención al cliente			
Iteración IV	Mejora de rendimiento del sistema de gestión de incidencias.	Atención al cliente			
		...			
				Solución incremental - Integración iterativa	

04.03

•••

Iniciación a CDC con Debezium: Abriendo tu core al tiempo real

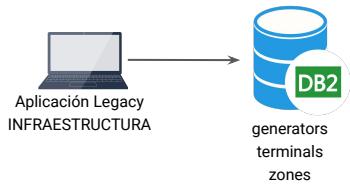
Live Demo

Implementando soluciones en tiempo real.

Roadmap (Iteración I)

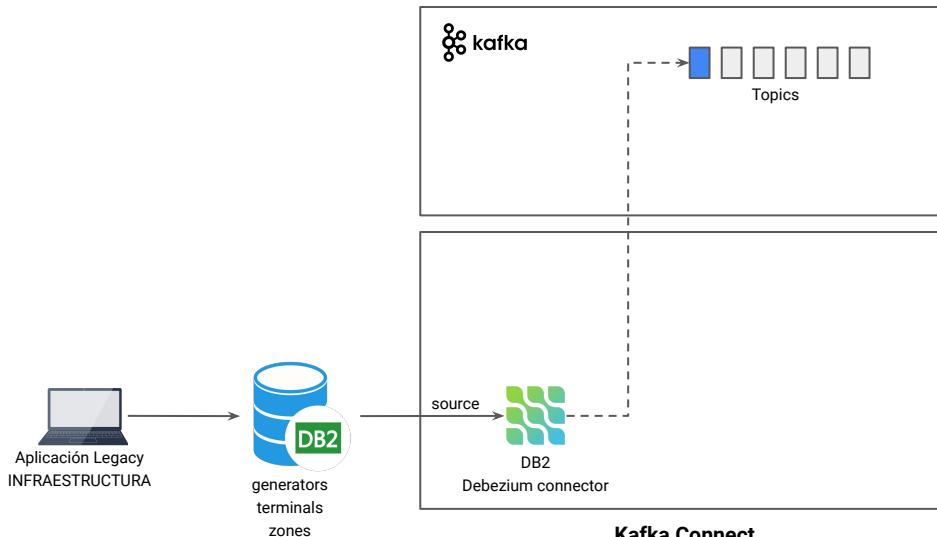
Necesidad de negocio	Módulo/Aplicación	
Visión global en tiempo real del estado de la infraestructura.	Infraestructura	
Sistema de alertas en tiempo real para informar de problemas en la infraestructura.	Infraestructura	
Visión en tiempo real del estado de las incidencias notificadas por los clientes.	Atención al cliente	
Sistema de alertas en tiempo real para notificar cambios en las incidencias de clientes.	Atención al cliente	
Mejora de rendimiento del sistema de gestión de incidencias.	Atención al cliente	
Implementar funcionalidades adicionales para el sistema de gestión de incidencias.	Atención al cliente	
...		

Construyendo solución (Iteración I)



Legacy

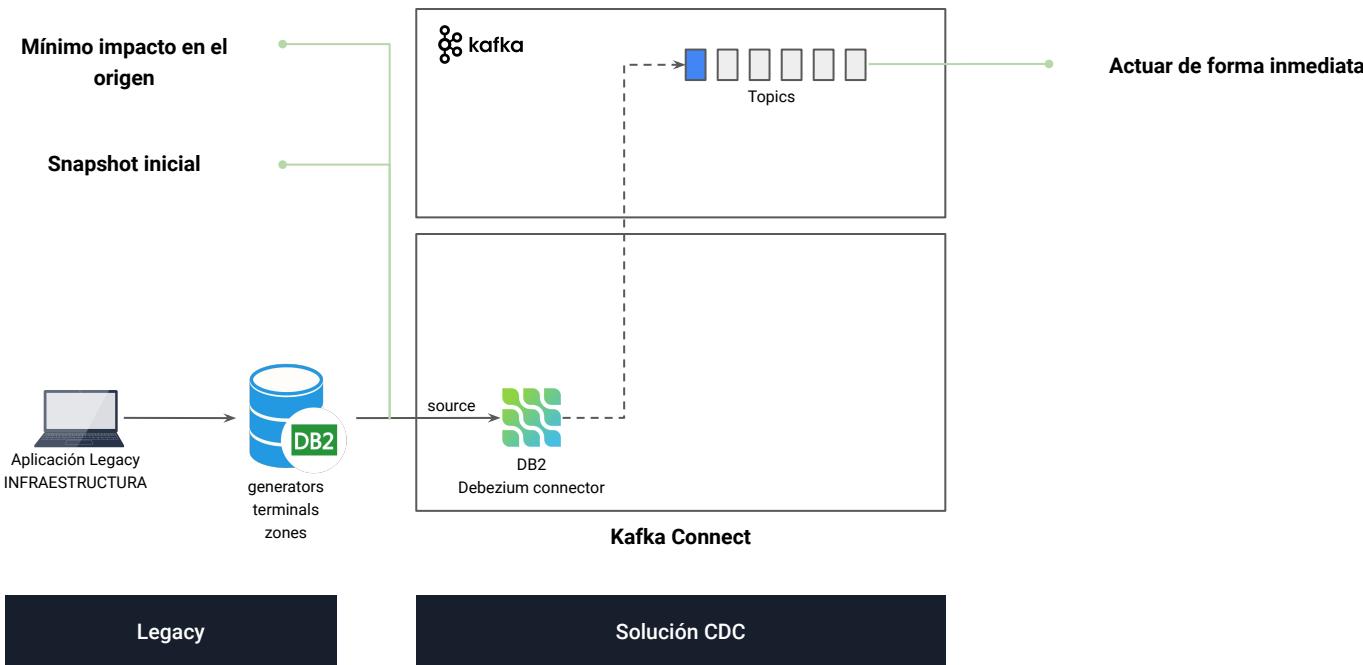
Construyendo solución (Iteración I)



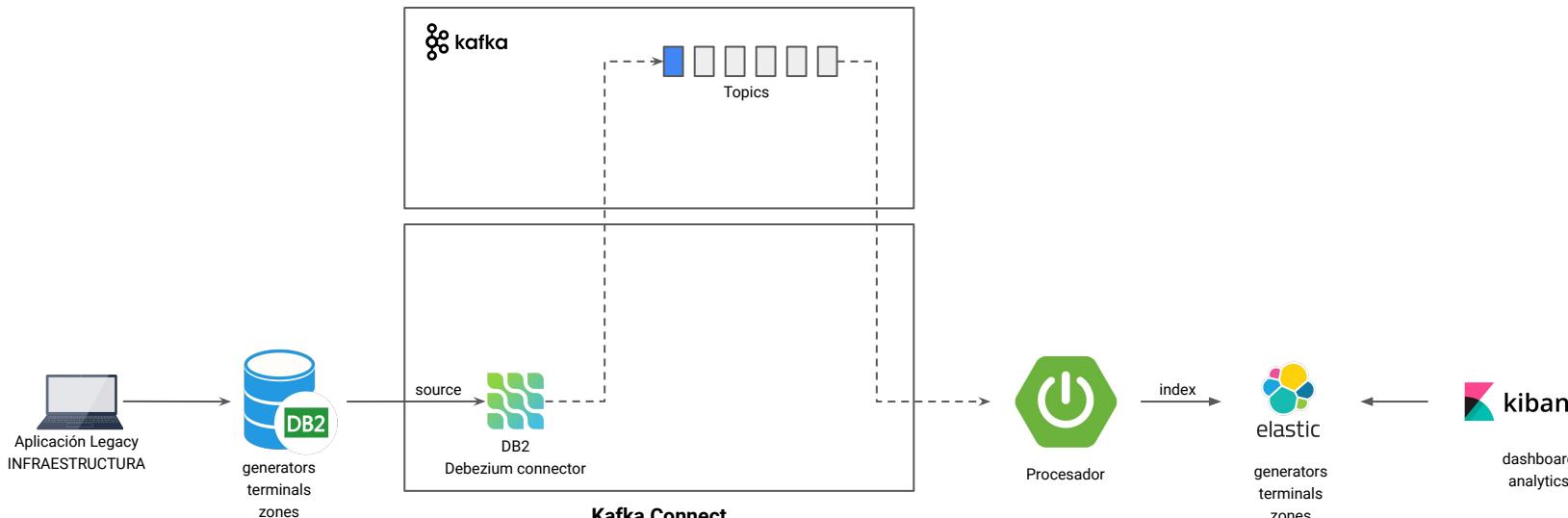
Legacy

Solución CDC

Construyendo solución (Iteración I)



Construyendo solución (Iteración I)



Legacy

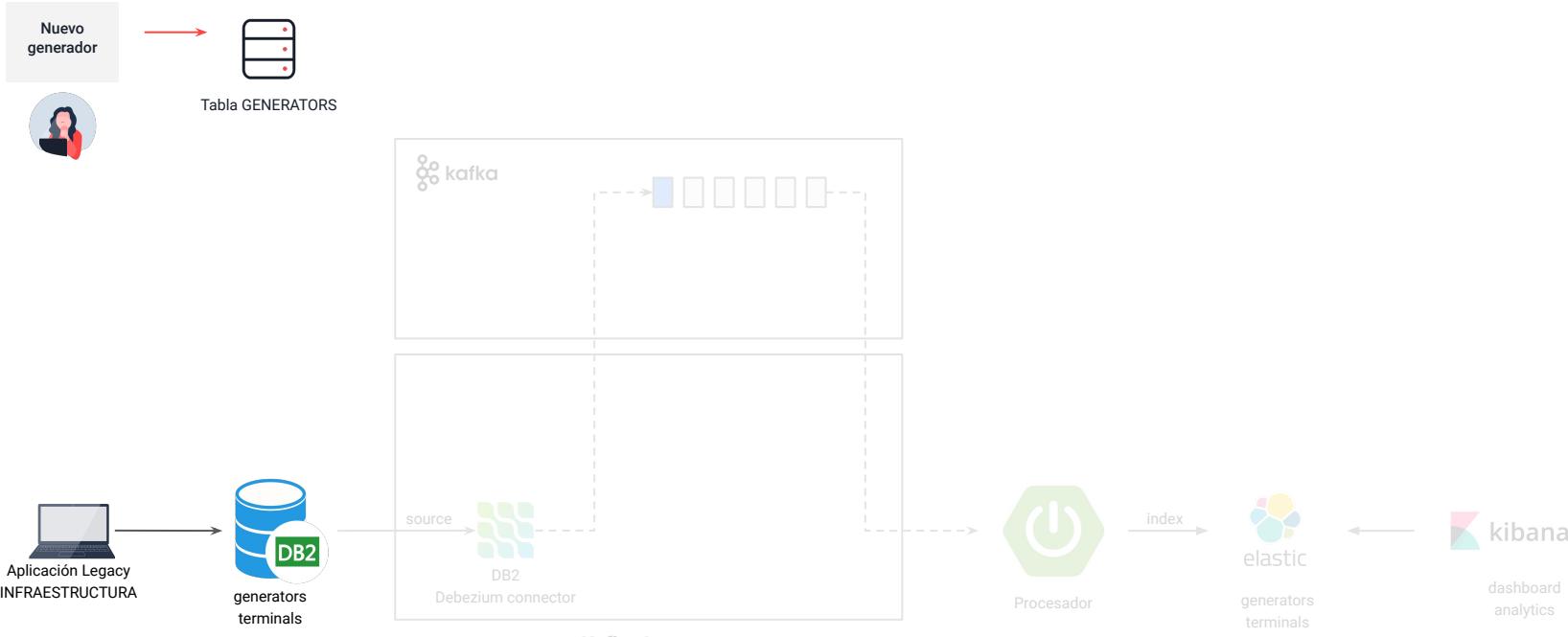
Solución CDC

Caso de uso (Visión estratégica de la infraestructura)

Live Demo (Iteración I).



¿Qué ocurre? (Iteración I)



Legacy

Solución CDC

Caso de uso (Visión estratégica de la infraestructura)

¿Qué ocurre? (Iteración I)

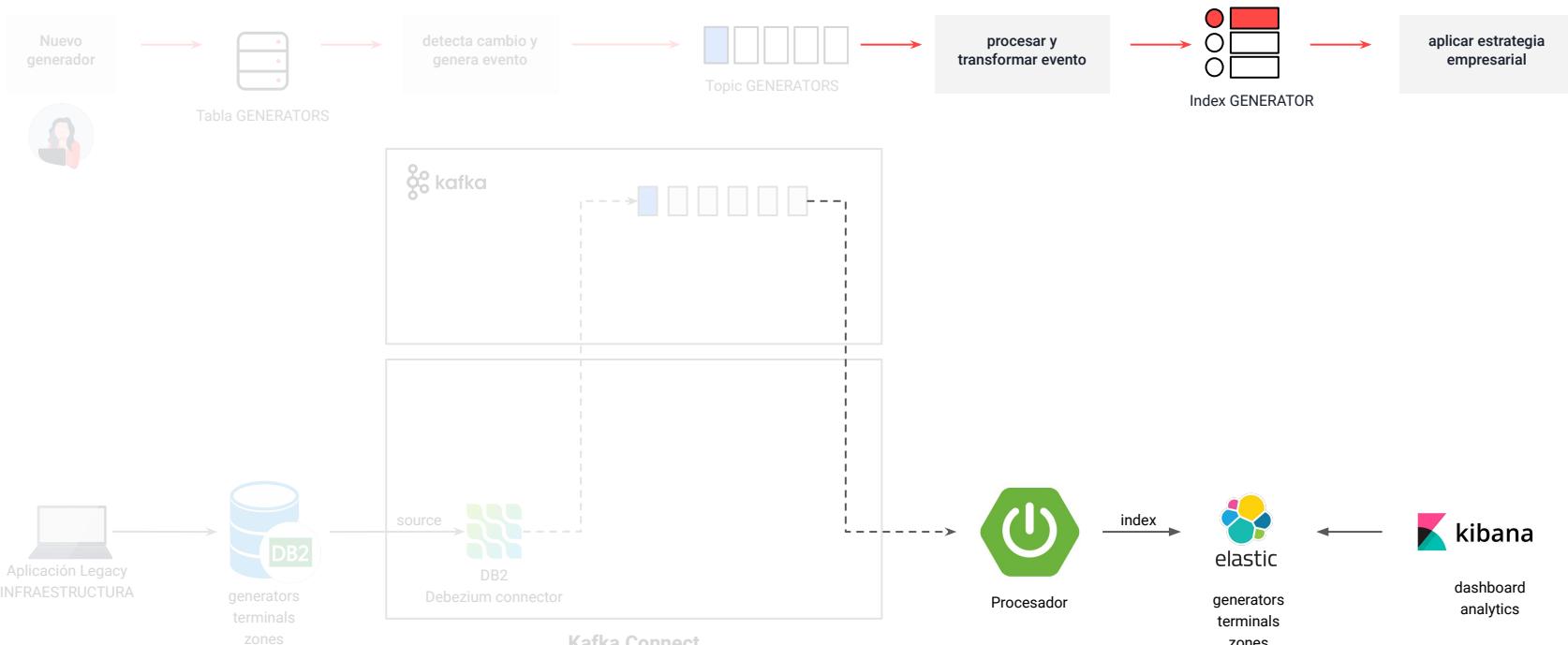


Legacy

Solución CDC

Caso de uso (Visión estratégica de la infraestructura)

¿Qué ocurre? (Iteración I)

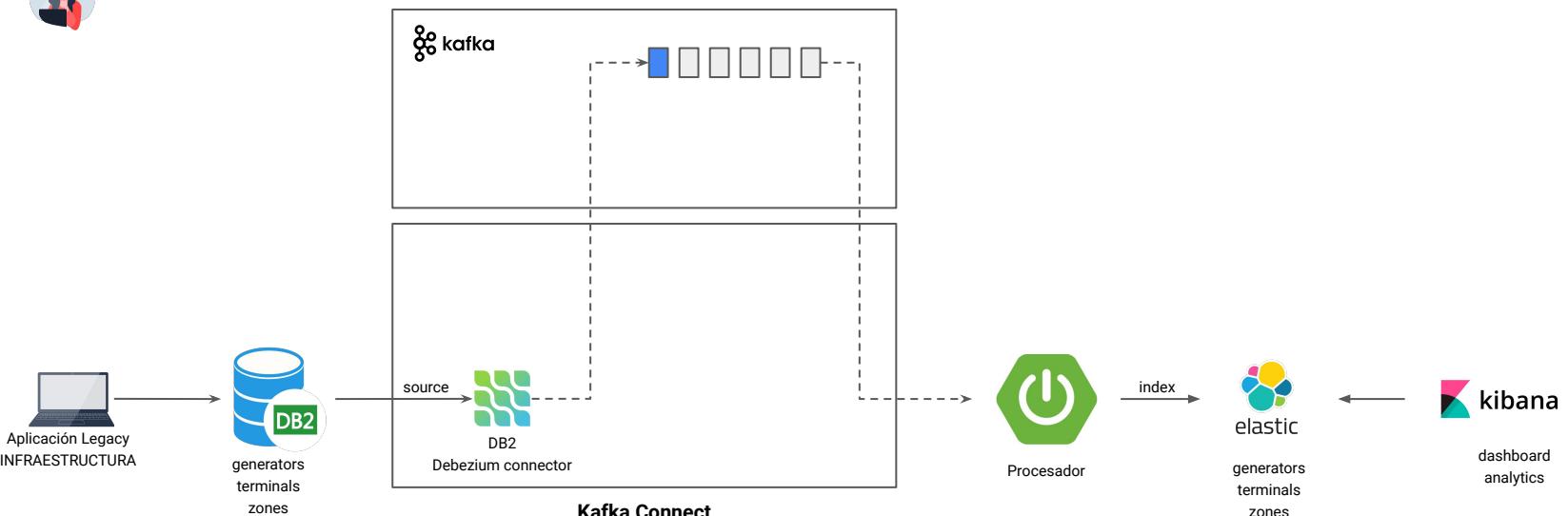
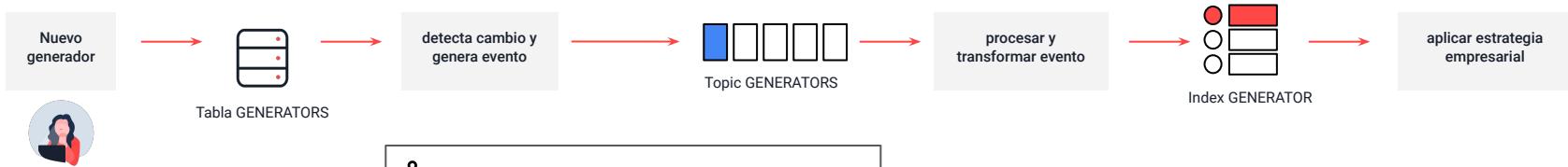


Legacy

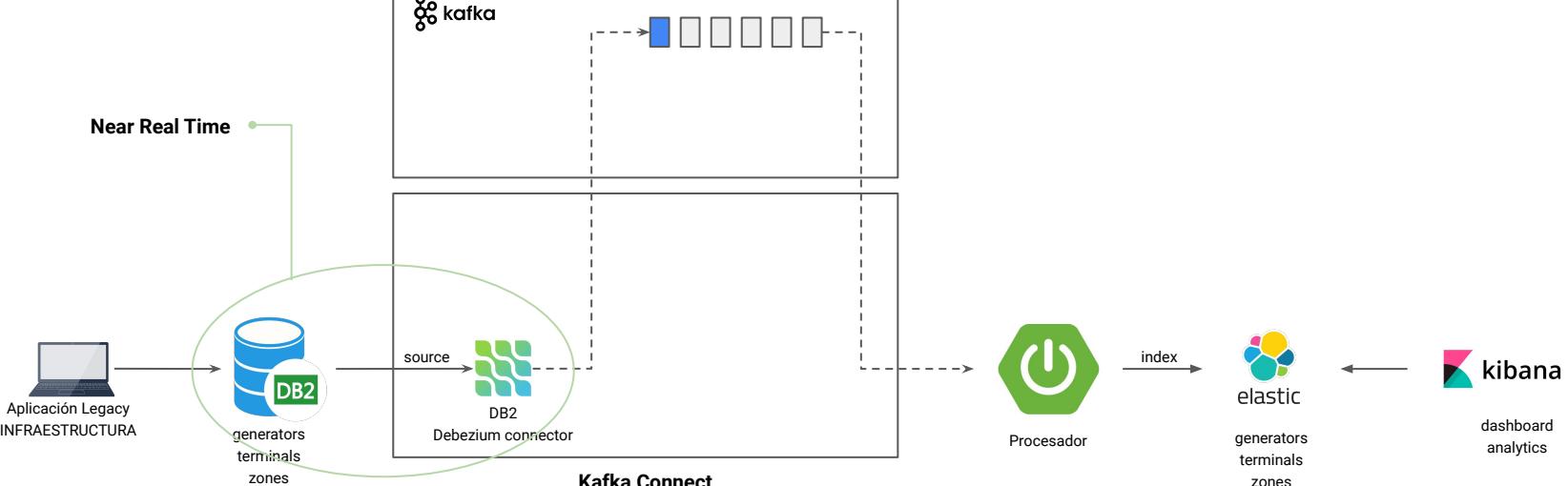
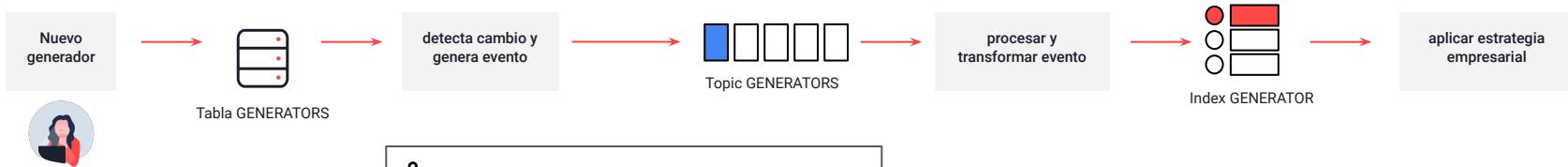
Solución CDC

Caso de uso (Visión estratégica de la infraestructura)

¿Qué ocurre? (Iteración I)



¿Qué ocurre? (Iteración I)



Legacy

Solución CDC

Caso de uso (Visión estratégica de la infraestructura)

Hitos (Iteración I).

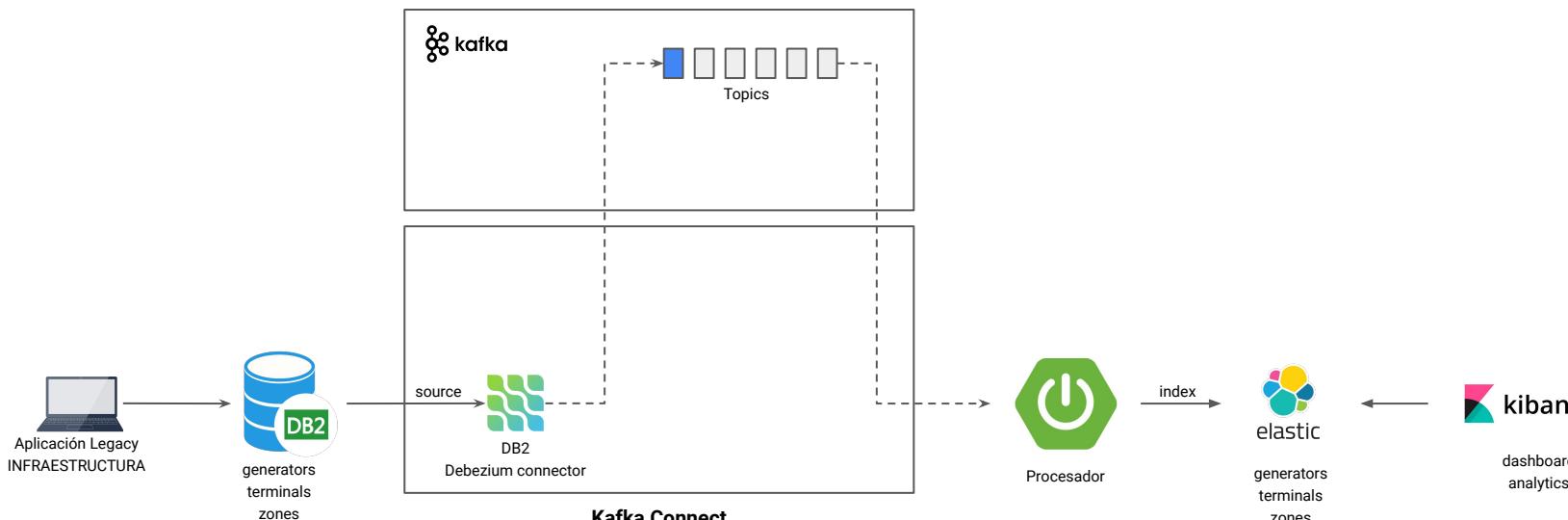
Cumplir las demandas de negocio

Solución en tiempo real

Roadmap (Iteración II)

Necesidad de negocio	Módulo/Aplicación
Visión global en tiempo real del estado de la infraestructura.	Infraestructura
Sistema de alertas en tiempo real para informar de problemas en la infraestructura.	Infraestructura
Visión en tiempo real del estado de las incidencias notificadas por los clientes.	Atención al cliente
Sistema de alertas en tiempo real para notificar cambios en las incidencias de clientes.	Atención al cliente
Mejora de rendimiento del sistema de gestión de incidencias.	Atención al cliente
Implementar funcionalidades adicionales para el sistema de gestión de incidencias.	Atención al cliente
...	

Solución (Iteración I)

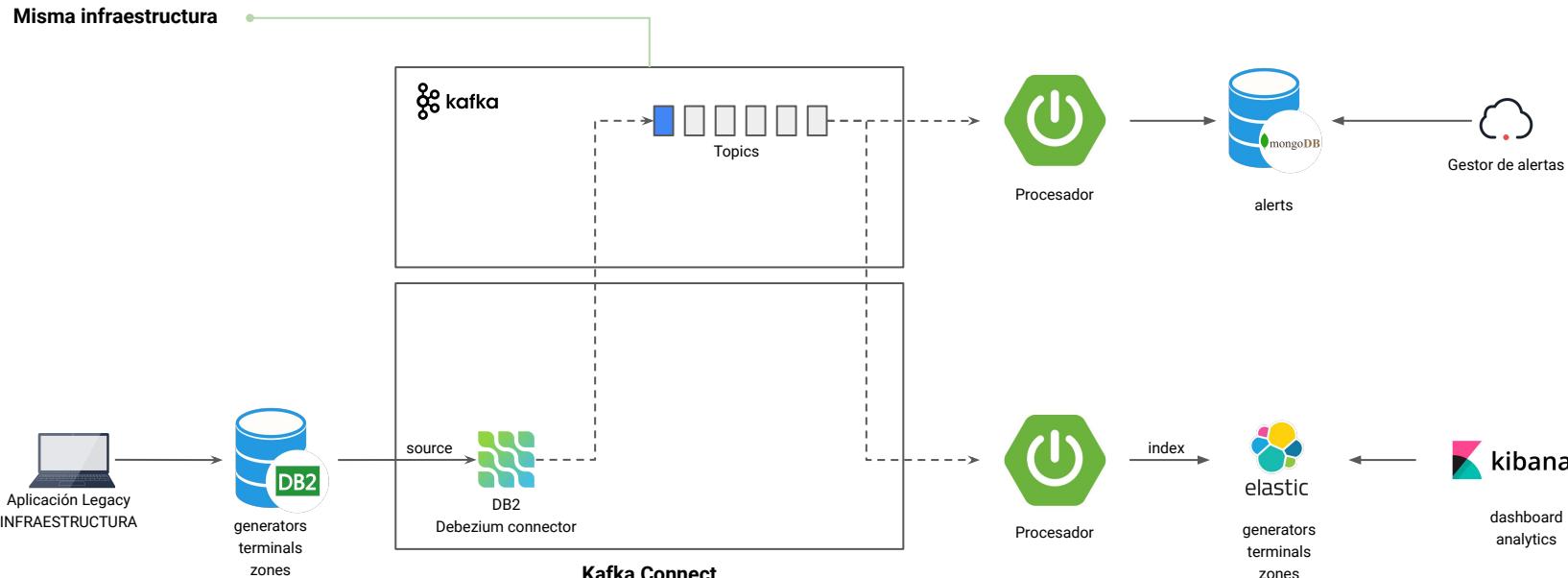


Legacy

Solución CDC

Caso de uso (Visión estratégica de la infraestructura)

Construyendo solución (Iteración II)



Legacy

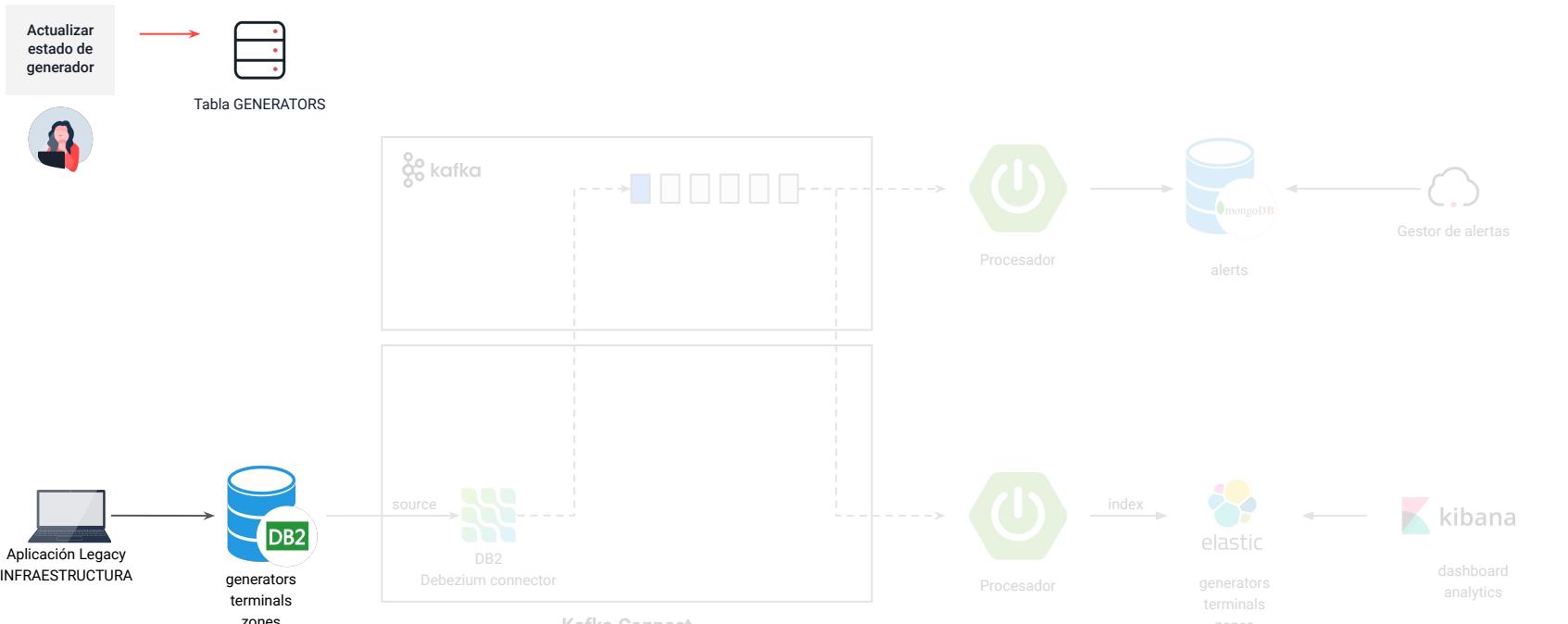
Solución CDC

Caso de uso (Visión estratégica de la infraestructura/Sistema de alertas)

Live Demo (Iteración II).



¿Qué ocurre? (Iteración II)

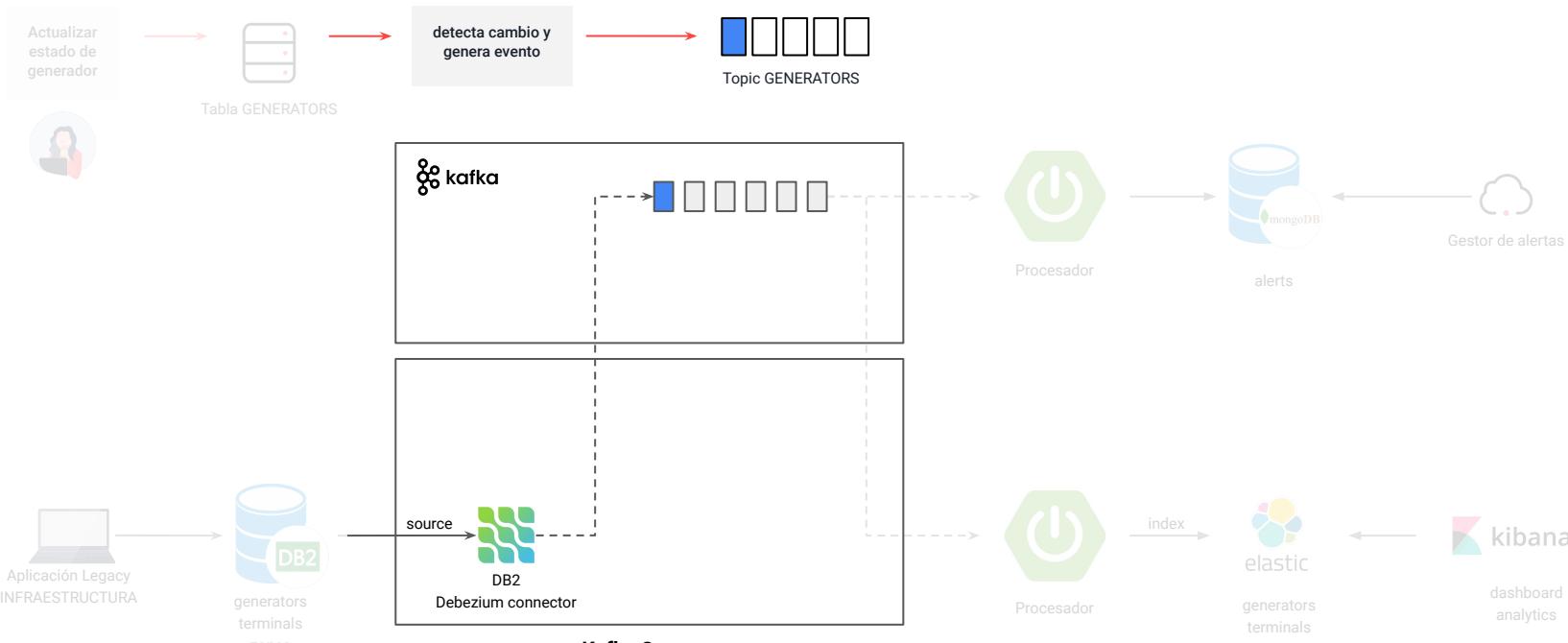


Legacy

Solución CDC

Caso de uso (Visión estratégica de la infraestructura/Sistema de alertas)

¿Qué ocurre? (Iteración II)

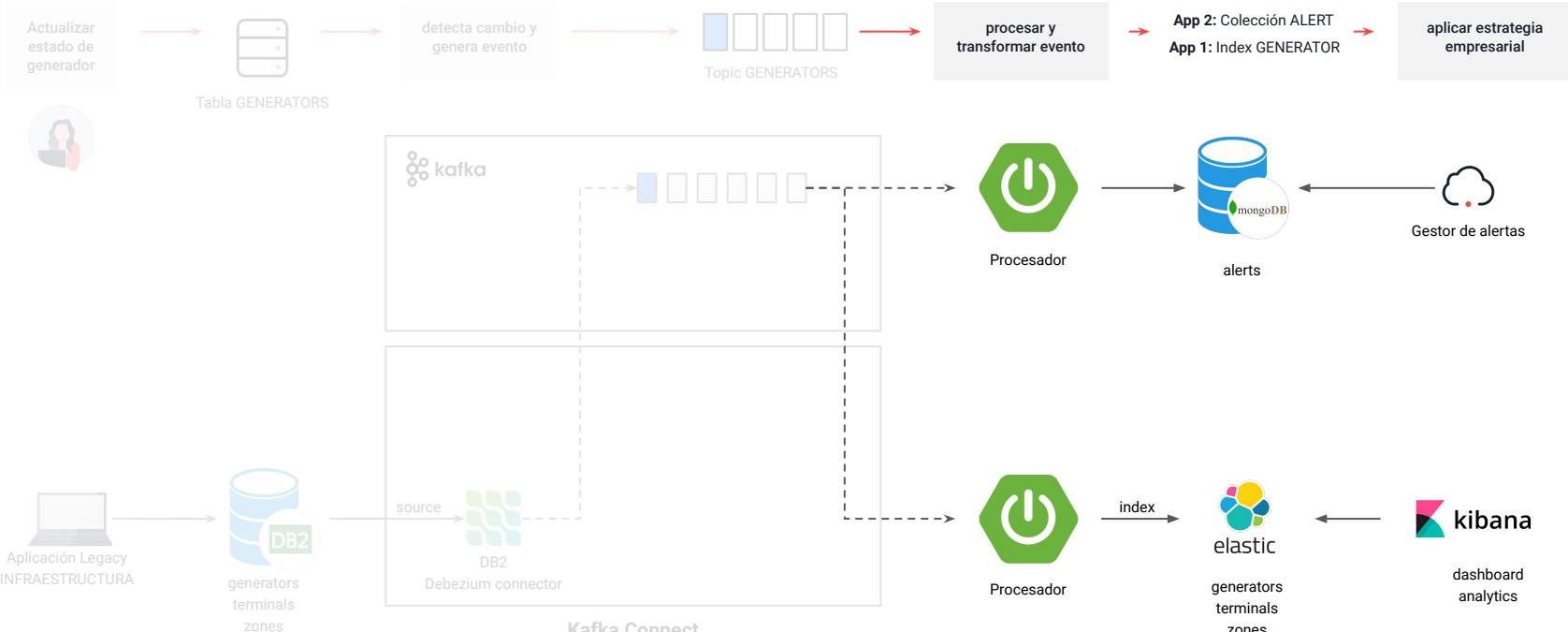


Legacy

Solución CDC

Caso de uso (Visión estratégica de la infraestructura/Sistema de alertas)

¿Qué ocurre? (Iteración II)

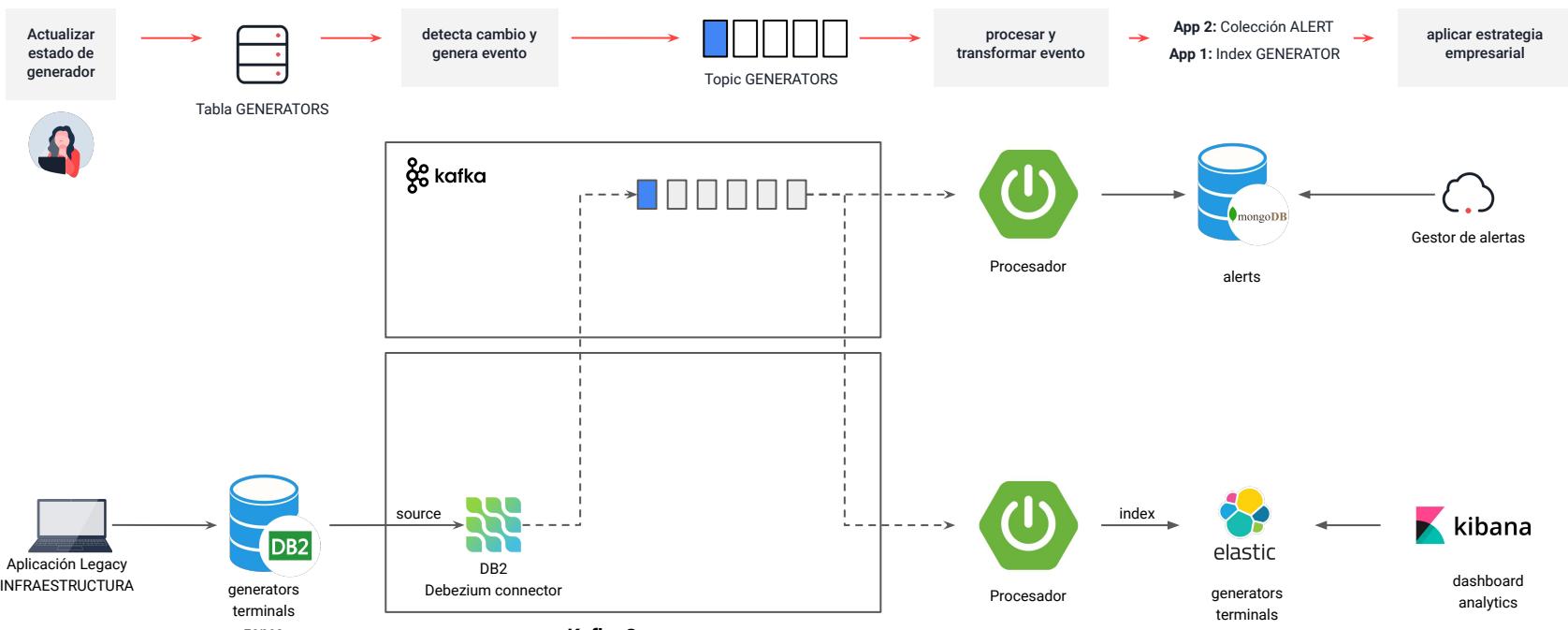


Legacy

Solución CDC

Caso de uso (Visión estratégica de la infraestructura/Sistema de alertas)

¿Qué ocurre? (Iteración II)



Legacy

Solución CDC

Caso de uso (Visión estratégica de la infraestructura/Sistema de alertas)

Hitos.

~~Cumplir las demandas de negocio~~

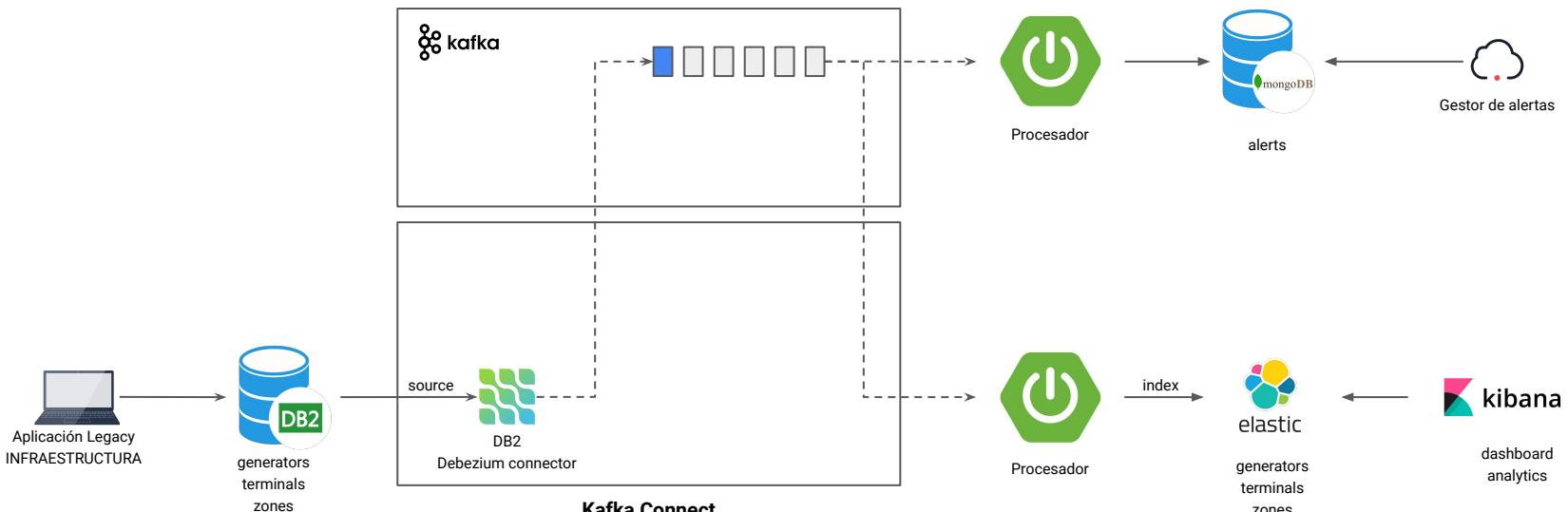
~~Solución en tiempo real~~

Añadir nuevo caso de uso a la solución

Roadmap (Iteración III)

Necesidad de negocio	Módulo/Aplicación	
Visión global en tiempo real del estado de la infraestructura.	Infraestructura	
Sistema de alertas en tiempo real para informar de problemas en la infraestructura.	Infraestructura	
Visión en tiempo real del estado de las incidencias notificadas por los clientes.	Atención al cliente	
Sistema de alertas en tiempo real para notificar cambios en las incidencias de clientes.	Atención al cliente	
Mejora de rendimiento del sistema de gestión de incidencias.	Atención al cliente	
Implementar funcionalidades adicionales para el sistema de gestión de incidencias.	Atención al cliente	
...		

Solución (Iteración II)

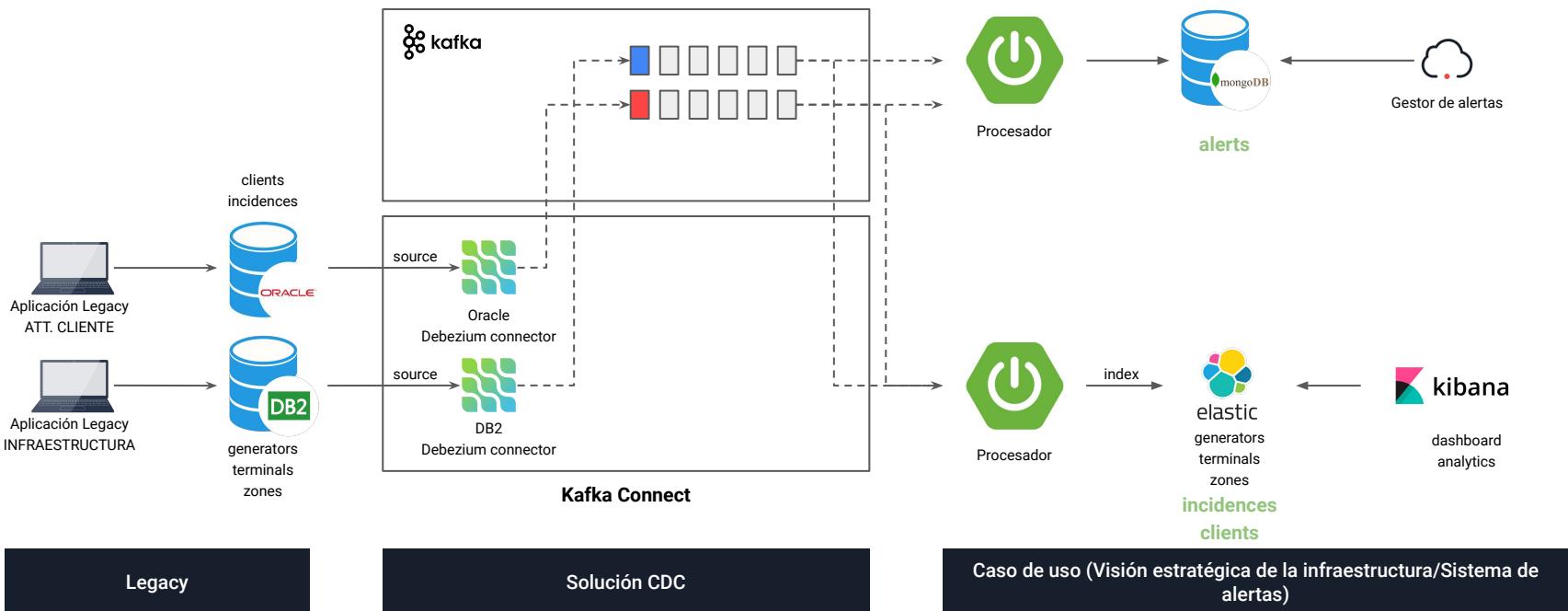


Legacy

Solución CDC

Caso de uso (Visión estratégica de la infraestructura/Sistema de alertas)

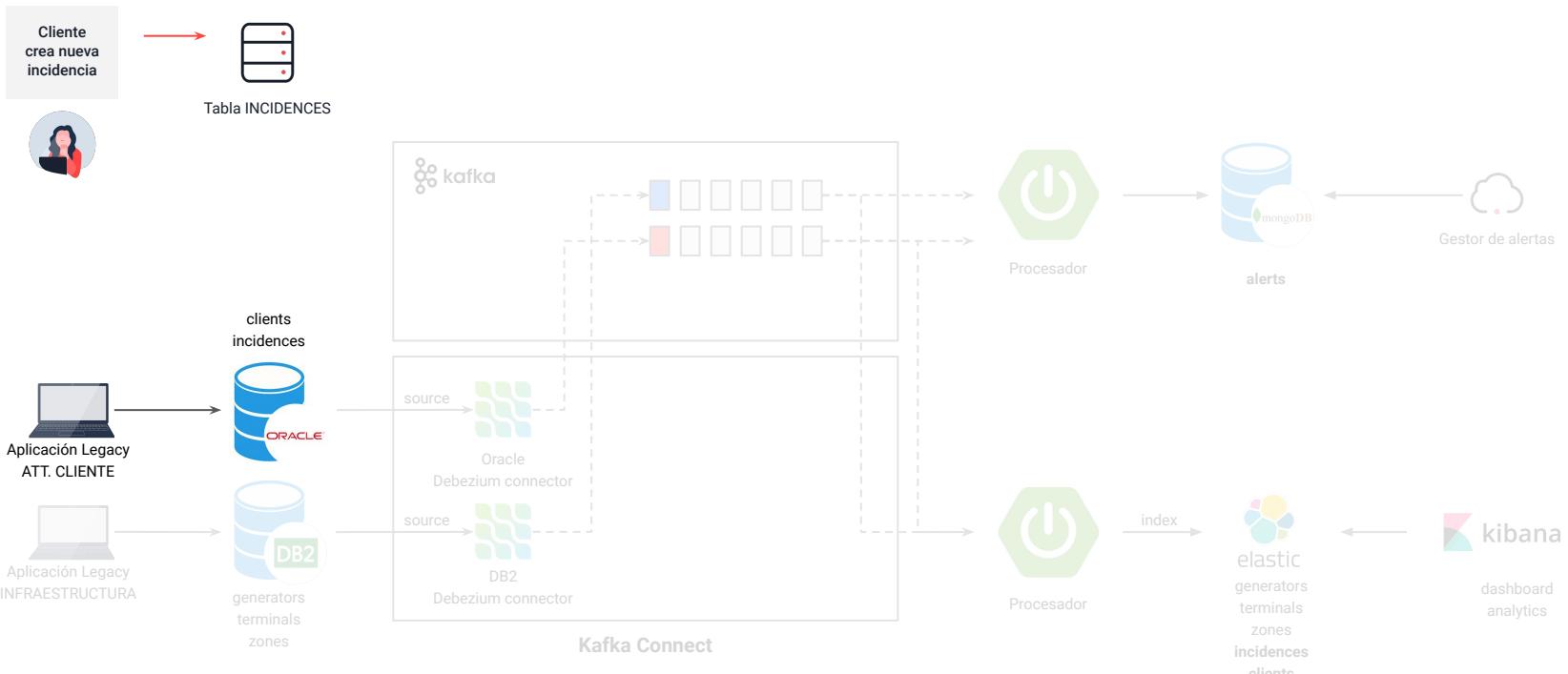
Construyendo solución (Iteración III)



Infraestructura: Solución Real Time - Live Demo.



¿Qué ocurre? (Iteración III)

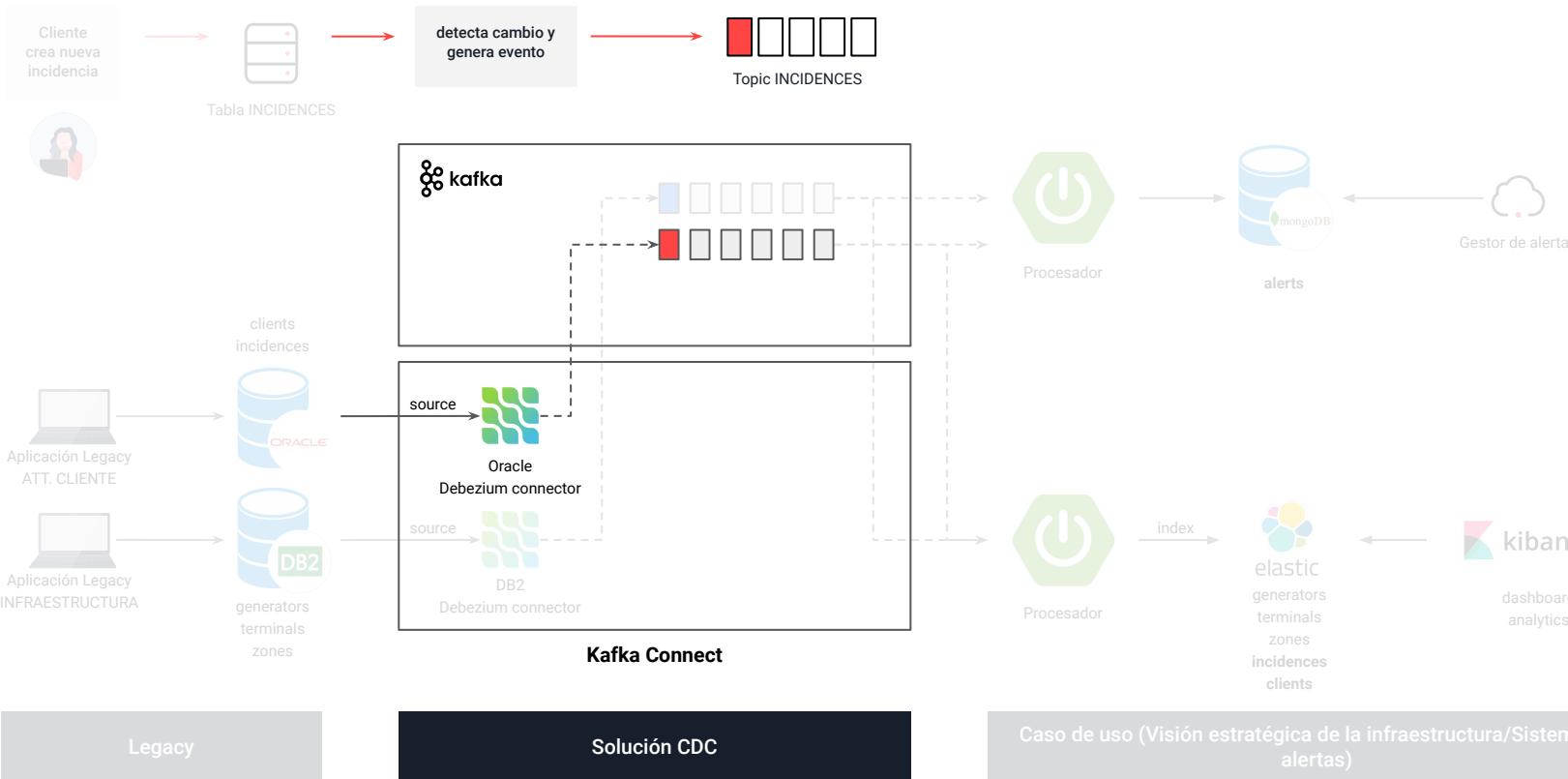


Legacy

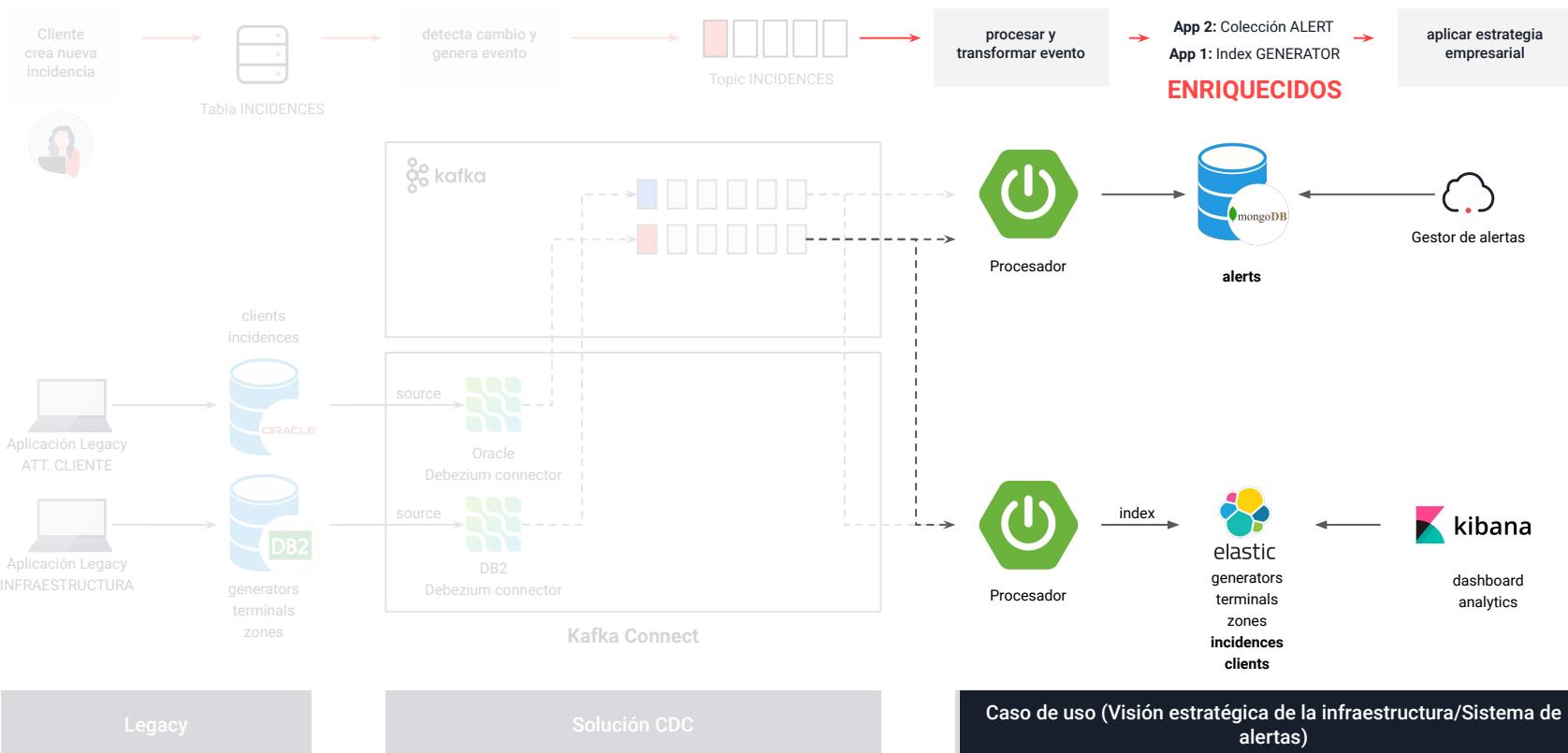
Solución CDC

Caso de uso (Visión estratégica de la infraestructura/Sistema de alertas)

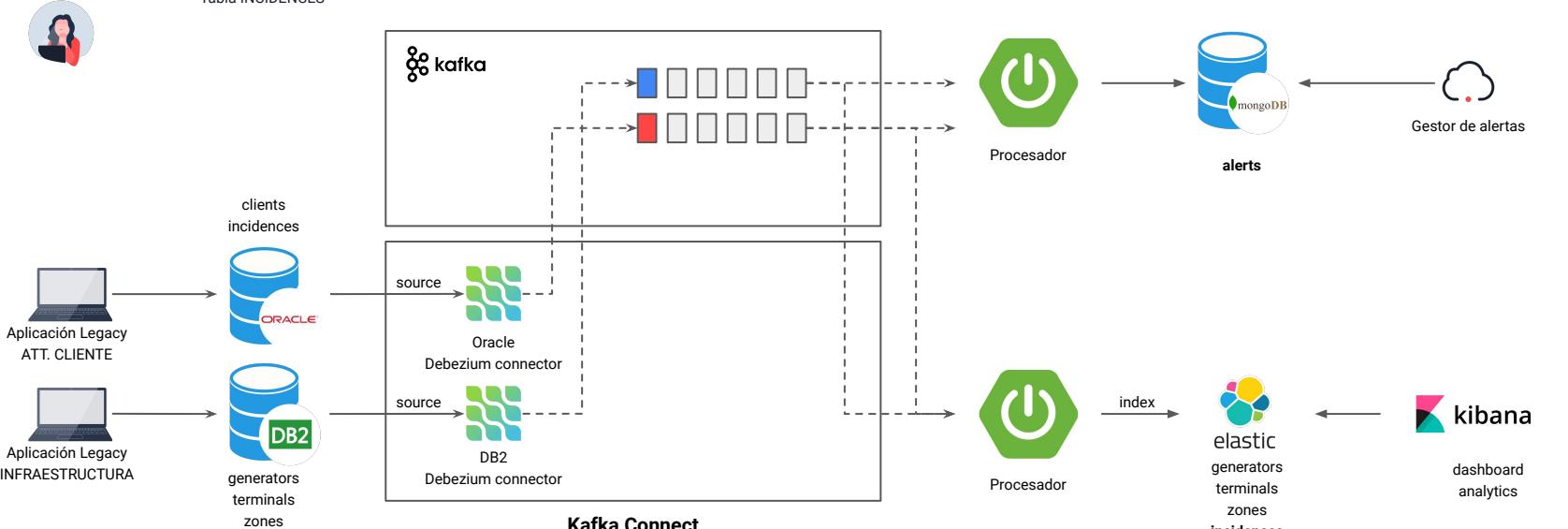
¿Qué ocurre? (Iteración III)



¿Qué ocurre? (Iteración III)



¿Qué ocurre? (Iteración III)



Legacy

Solución CDC

Caso de uso (Visión estratégica de la infraestructura/Sistema de alertas)

Hitos.

~~Cumplir las demandas de negocio~~

~~Solución en tiempo real~~

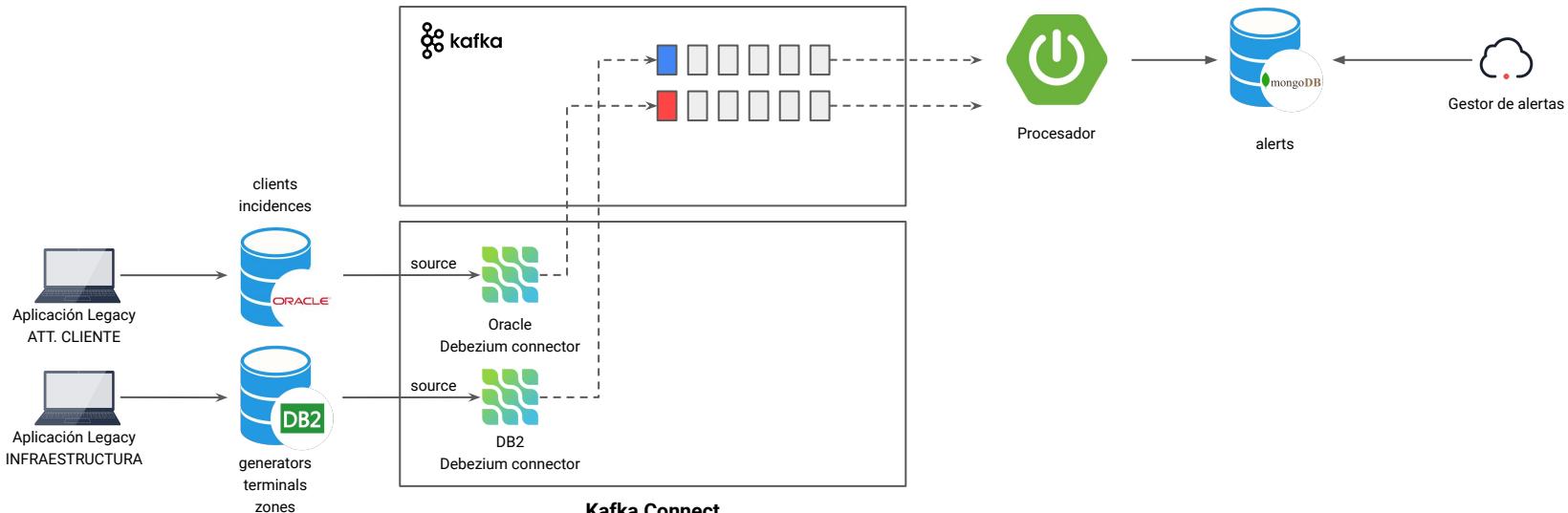
~~Añadir nuevo caso de uso a la solución~~

Integración de información entre módulos

Roadmap (Iteración IV)

Necesidad de negocio	Módulo/Aplicación
Visión global en tiempo real del estado de la infraestructura.	Infraestructura
Sistema de alertas en tiempo real para informar de problemas en la infraestructura.	Infraestructura
Visión en tiempo real del estado de las incidencias notificadas por los clientes.	Atención al cliente
Sistema de alertas en tiempo real para notificar cambios en las incidencias de clientes.	Atención al cliente
Mejora de rendimiento del sistema de gestión de incidencias.	Atención al cliente
Implementar funcionalidades adicionales para el sistema de gestión de incidencias.	Atención al cliente
...	

Solución (Iteración III)

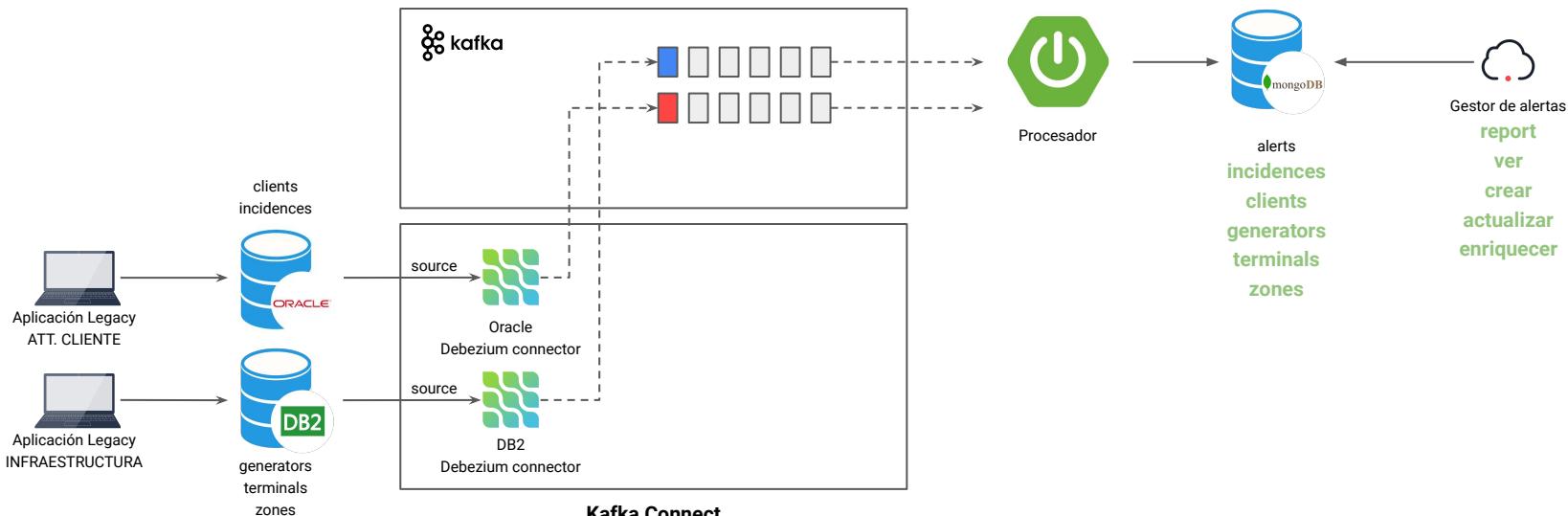


Legacy

Solución CDC

Caso de uso (Sistema de alertas)

Construyendo solución (Iteración IV) (I)

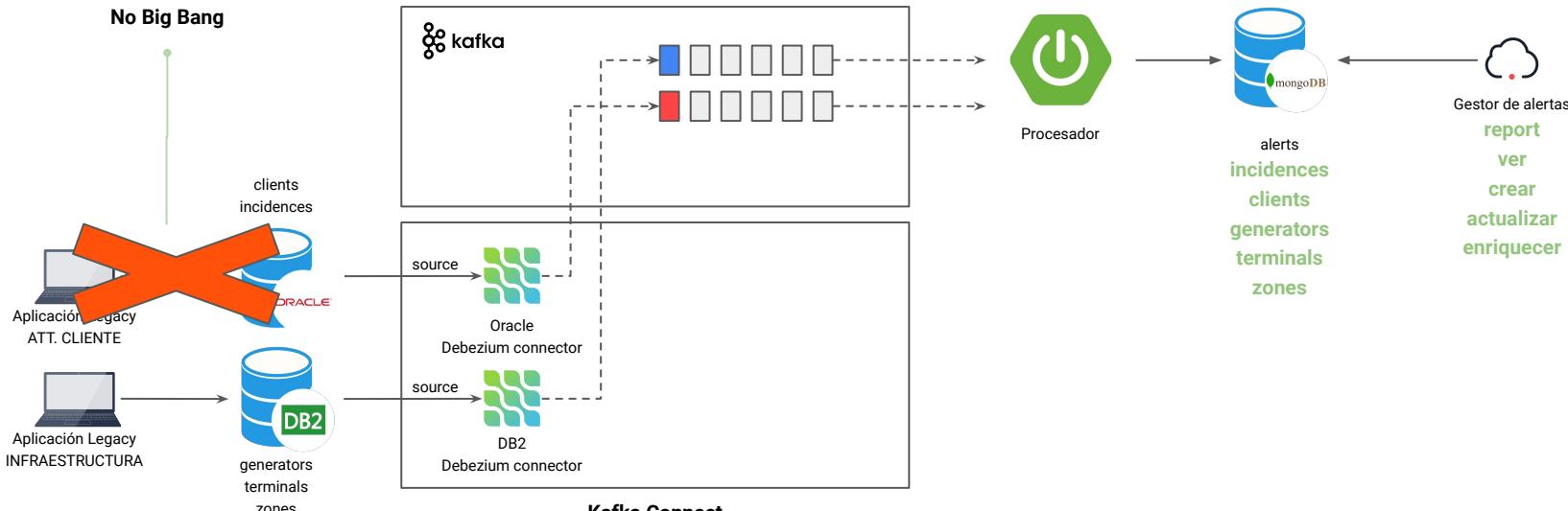


Legacy

Solución CDC

Caso de uso (Sistema de alertas + Modernización)

Construyendo solución (Iteración IV) (I)

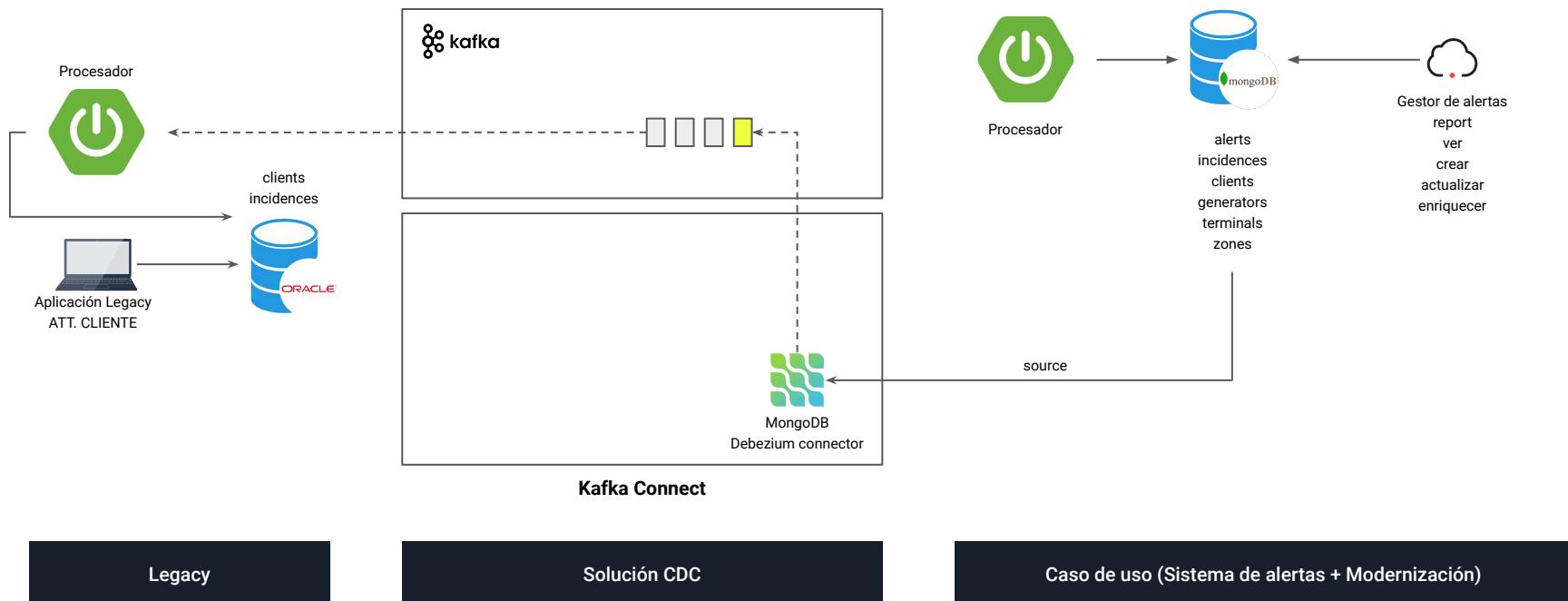


Legacy

Solución CDC

Caso de uso (Sistema de alertas + Modernización)

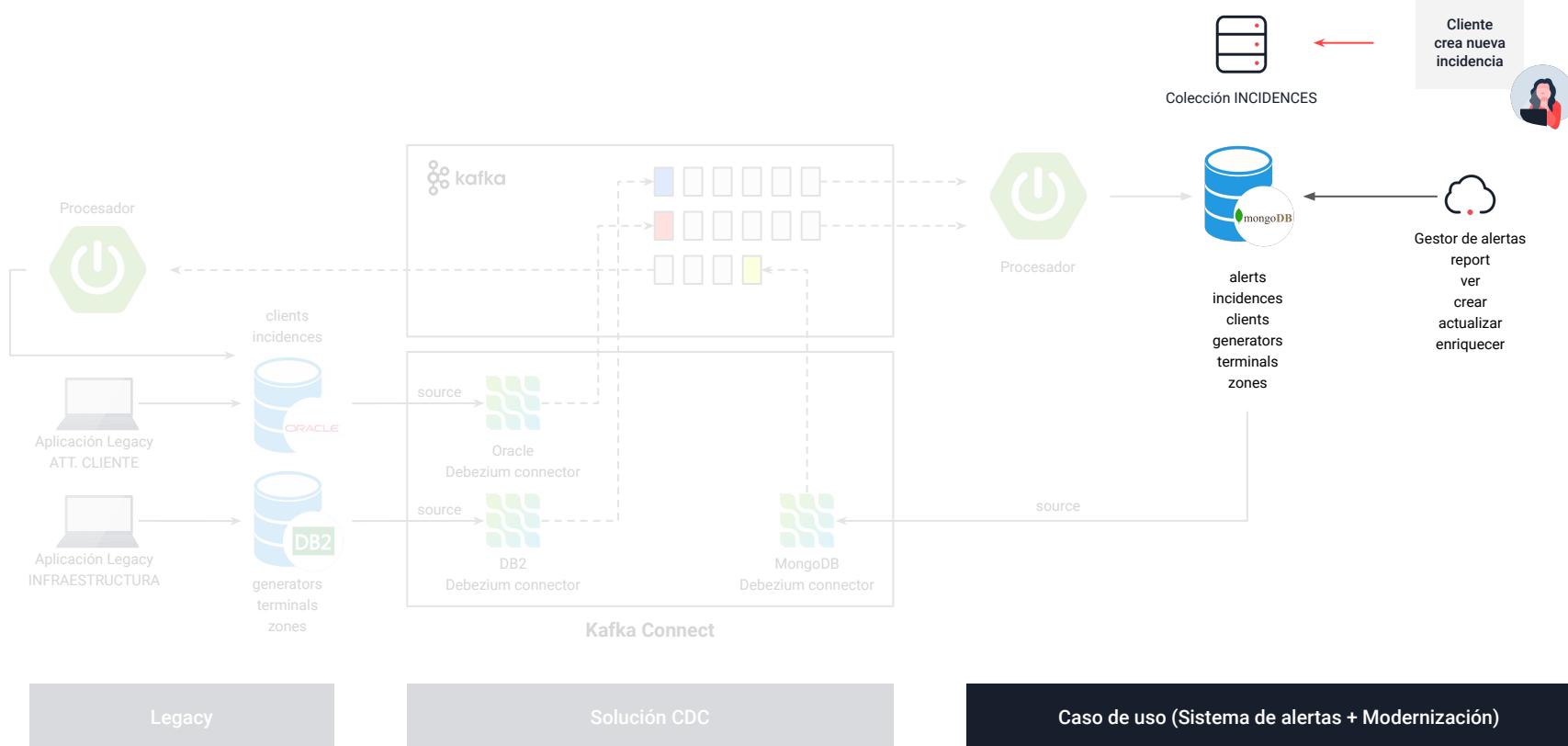
Construyendo solución (Iteración IV) (II)



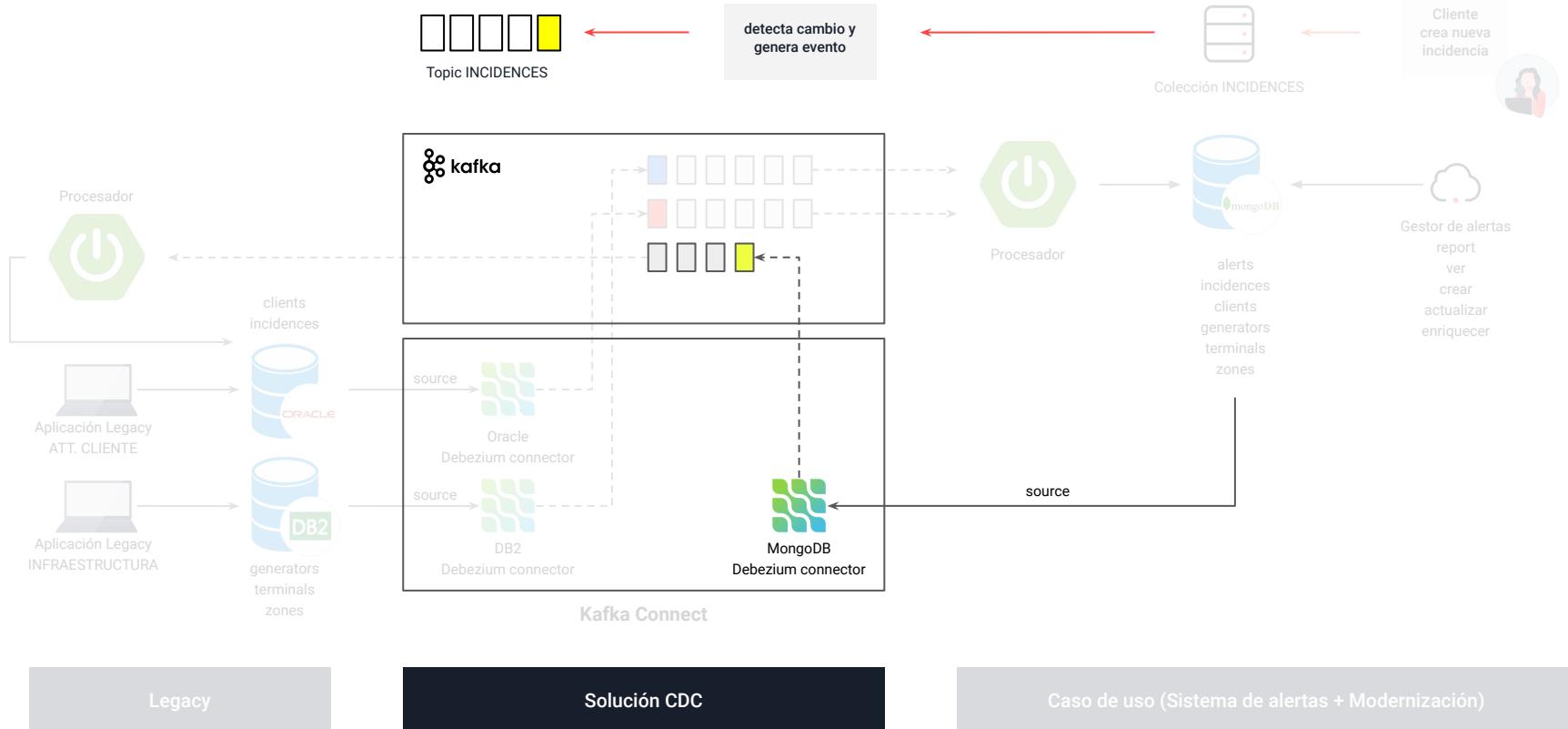
Infraestructura: Solución Real Time - Live Demo.



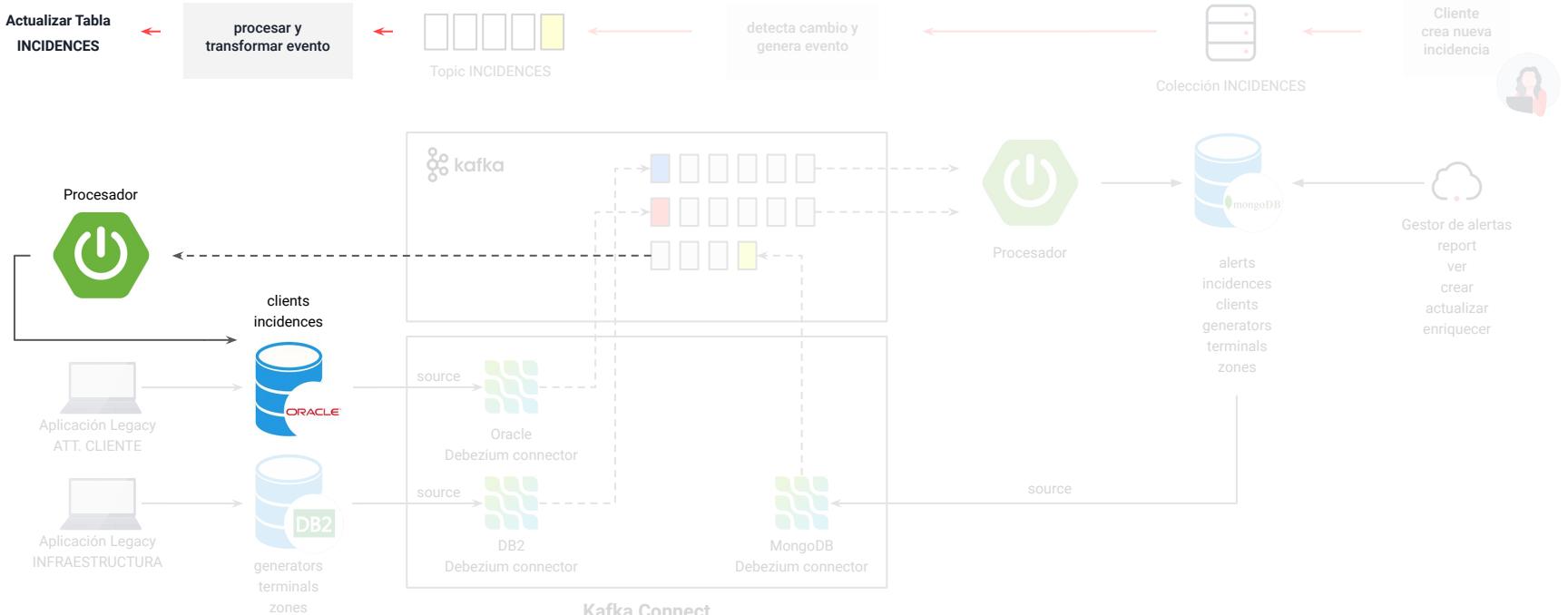
¿Qué ocurre? (Iteración IV)



¿Qué ocurre? (Iteración IV)



¿Qué ocurre? (Iteración IV)

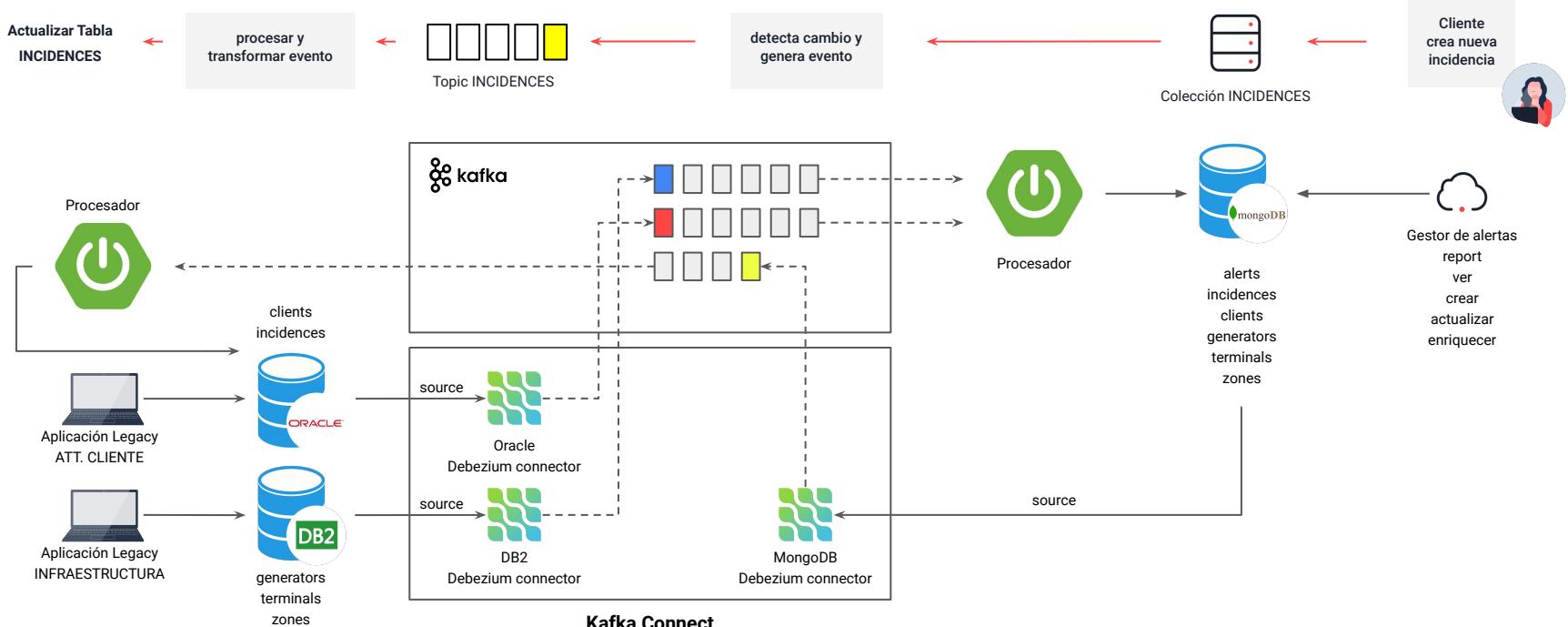


Legacy

Solución CDC

Caso de uso (Sistema de alertas + Modernización)

¿Qué ocurre? (Iteración IV)



Legacy

Solución CDC

Caso de uso (Sistema de alertas + Modernización)

Hitos.

~~Cumplir las demandas de negocio~~

~~Solución en tiempo real~~

~~Añadir nuevo caso de uso a la solución~~

~~Integración de información entre módulos~~

Transformación digital - El camino a la modernización

04.03

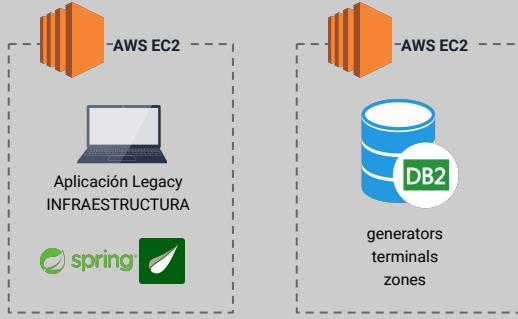
• • •

Iniciación a CDC con Debezium: Abriendo tu core al tiempo real

Live Demo

Infraestructura Demo.

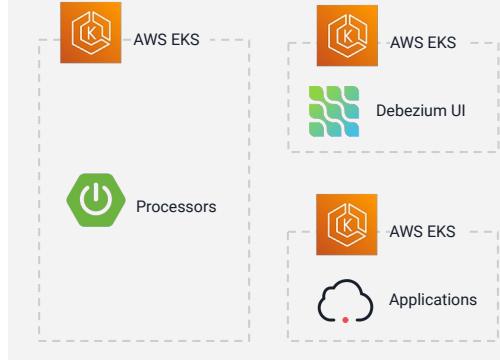
RUNTIME - LEGACY



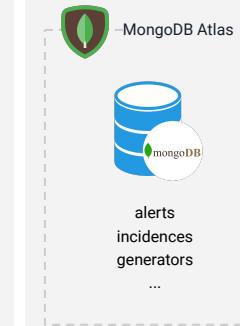
EVENT PROCESSING & STREAMING LAYER



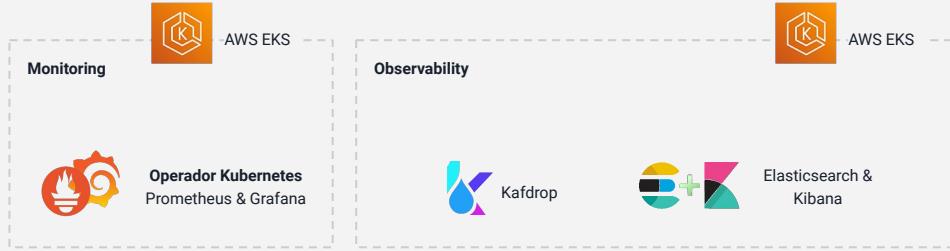
OPERATIONAL PLATFORM



DATA



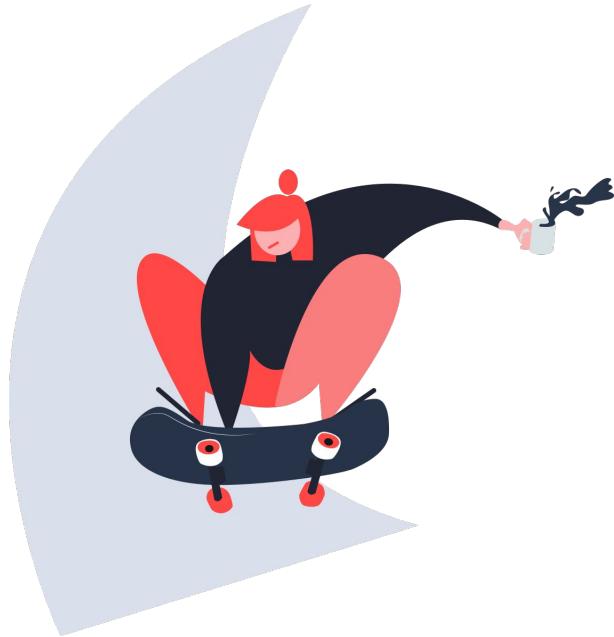
SUPPORTING



DEVELOPMENT



¿Escenario técnico complicado?

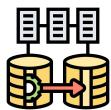


05.

• • • Iniciación a CDC con Debezium: Abriendo tu core al tiempo real

Conclusions.

¿Qué hemos aprendido hoy?



CDC

Es una técnica que nos permite replicar cambios desde una o varias fuentes de datos a uno o varios destinos en **tiempo real**. Nos permite hacer soluciones abiertas, incrementales y facilita la convivencia entre sistemas



Log-Based

Siempre que puedas elige métodos de CDC **basados en ficheros de logs** ya que el impacto sobre el motor de datos es mínimo y no necesitarás realizar muchas acciones adicionales para ponerlo en marcha



CDC en procesos de Modernización

CDC te puede ayudar en los procesos de transformación digital al **permitir realizar soluciones abiertas, poco intrusivas, incrementales y facilitar mucho la convivencia** a nivel de datos entre sistemas legacy y modernos, aplicando replicación bidireccional si es necesario



Debezium

Es una **herramienta para hacer CDC, soportada por Red Hat, que se basa en el uso de conectores de Kafka Connect**. Nos permite, principalmente, llevar cambios de fuentes de datos a Kafka para que sean procesados para satisfacer casos de uso

¿Qué implica?



Infraestructura

Si queremos aplicar CDC tenemos que tener en cuenta que necesitamos infraestructura (y sus entornos) que tendrá que ser gestionada y mantenida. Por ejemplo, en el caso de Debezium, necesitamos un PaaS y Kafka



Conocimiento del modelo

Tenemos que conocer muy bien el sistema origen y el modelo de datos para poder replicar aquellas tablas que realmente aportan valor en el caso de uso que queramos implementar.



Trabajar con datos en crudo

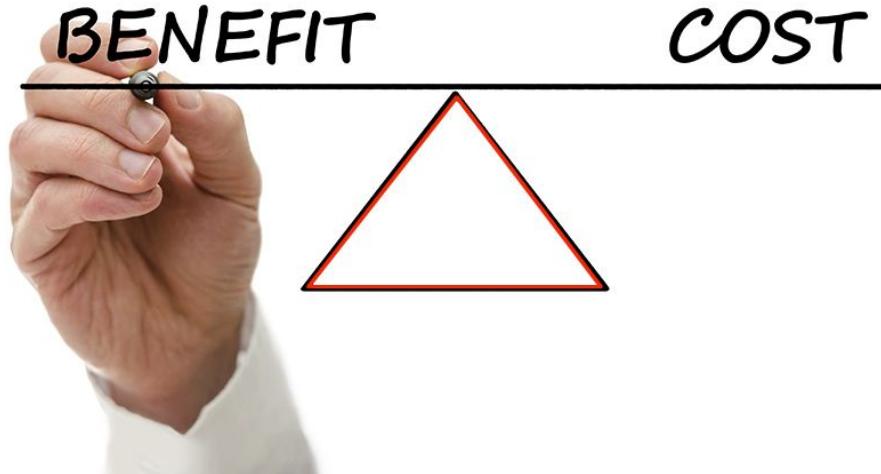
Los cambios que se propagan son los realizados en base de datos, a nivel de fila, con lo que no se dispone de un contexto de Negocio asociado en el que se producen.



Establecer un gobierno

El proceso de replicación está totalmente asociado a la estructura del sistema origen. Cualquier cambio de modelo se debe gestionar para mantener el proceso de replicación funcionando de forma correcta

¿La solución para todos los males?



Lecturas complementarias



Un vistazo a Debezium: una herramienta completa de CDC



Primeros pasos con Debezium

Lecturas relacionadas



[¿Qué es Oracle GoldenGate?](#)



[Replicación unidireccional con Oracle GoldenGate](#)

¡Muchas gracias!.



Jesús Pau de la Cruz,
Software Architect.

jpau@paradigmadigital.com
<https://www.linkedin.com/in/jesuspaudelacruz/>



José Alberto Ruiz Casarrubios,
Software Architect.

jaruiz@paradigmadigital.com
<https://www.linkedin.com/in/jaruizcasarrubios/>