Two-Node CDC Setup: PostgreSQL + Kafka (KRaft) + Debezium

This document contains complete step-by-step instructions to deploy a two-node Change Data Capture (CDC) architecture using Docker on Rocky Linux.

Architecture Overview

Node 1: PostgreSQL (CDC Source)
Node 2: Kafka (KRaft mode) + Kafka Connect + Debezium
Denodo connects externally to Kafka on Node 2.

Prerequisites

- Rocky Linux 8 or 9
- Docker and Docker Compose installed on both nodes
- Network connectivity between Node 1 and Node 2
- Firewall ports open: 5432, 9092, 8083

Node 1 – PostgreSQL CDC Setup

1. Copy postgres-cdc-node1.zip to Node 1
2. Unzip the file
3. Run: docker compose up -d
4. Verify container cdc_postgres is running
5. Connect using: docker exec -it cdc_postgres psql -U dbuser -d mydb
6. Verify customers table exists

Node 2 – Kafka + Debezium Setup

1. Copy kafka-debezium-node2.zip to Node 2
2. Unzip the file
3. Edit docker-compose.yml and set NODE2_IP
4. Edit connect/register-postgres.json and set NODE1_IP
5. Run: docker compose up -d
6. Verify containers cdc_kafka and cdc_connect are running
7. Register Debezium connector via curl on port 8083

End-to-End Validation

Insert data into PostgreSQL on Node 1. Consume Kafka topic on Node 2 using kafka-console-consumer. Verify Debezium JSON change events are received.

Denodo Connection Details

Kafka Bootstrap: NODE2_IP:9092
Topic: pgcdc.public.customers
Message Format: Debezium JSON
Mode: Real-time CDC