



Intro to

# Neo4j Change Data Capture for Event-Driven Applications

28 June, 2025 | Nokia Bangalore

**Abhishek Das**

Co-Founder & CTO,  
SourcingXPress



# What we are **building**?

- A platform to bridge the gap between opportunities and talents.
- A new way of looking at Talent match and sourcing strategies tapping into community of people that aspires to build on cutting edge tech.

# Neo4j Database

built for Operational And Analytical Workloads

## Intelligent Operational Systems

Graph Transactions,  
Storage & Querying



## Better Predictions for Analytics

Graph Analytics, ML,  
& Data Science

5 months into **production**

**40K+** Active Jobseekers

**300+** Active Recruiters

Time to hire reduced by **~70%**



# Neo4j change data capture

Building event driven applications with Aura  
Enterprise and Neo4j Enterprise

# Agenda

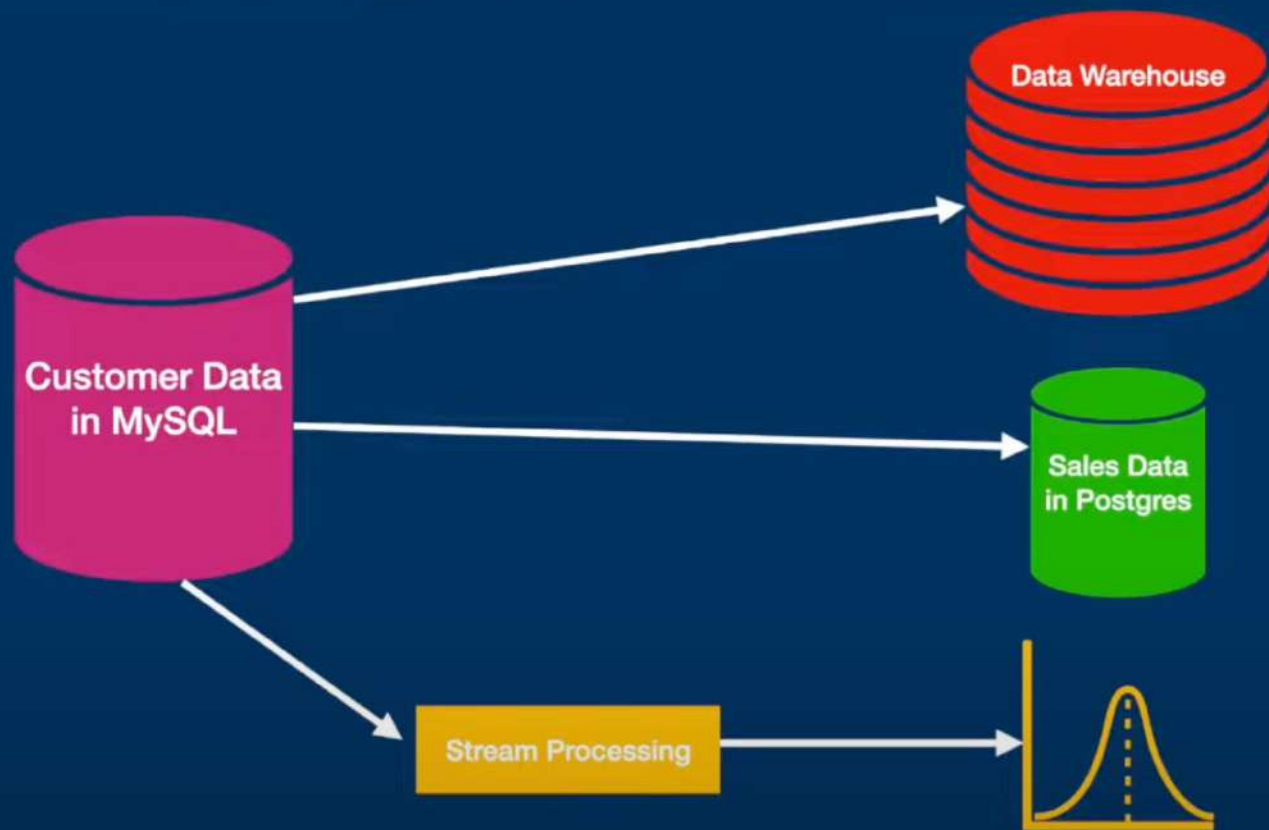
1. What is **Change Data Capture**
2. Why do you **need** it
3. How does it **work**
4. Lets see it in **action**

Questions in the Q&A

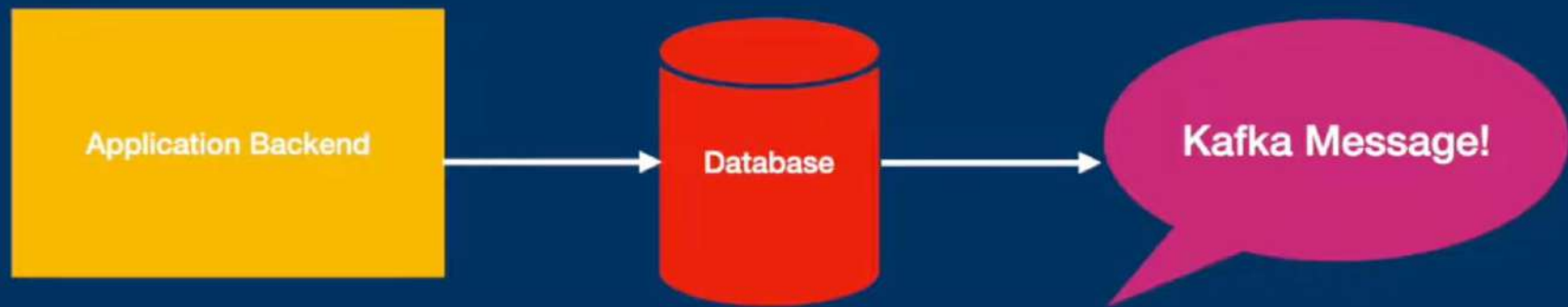


# CHANGE DATA CAPTURE (CDC)

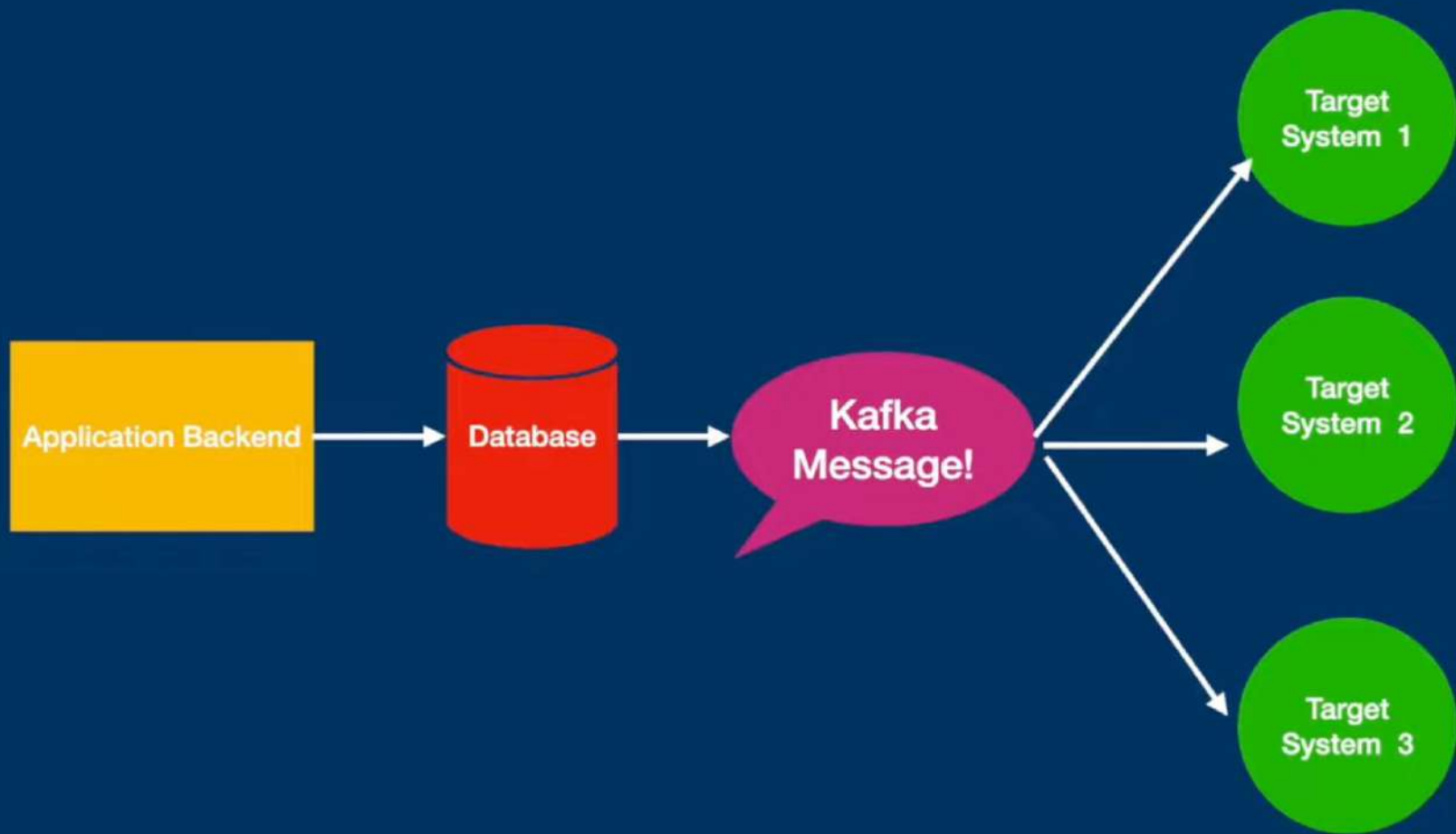
**PROCESS** of recognizing when data has changed in a **SOURCE SYSTEM** so a **DOWNSTREAM SYSTEM** can take an **action** based on that change





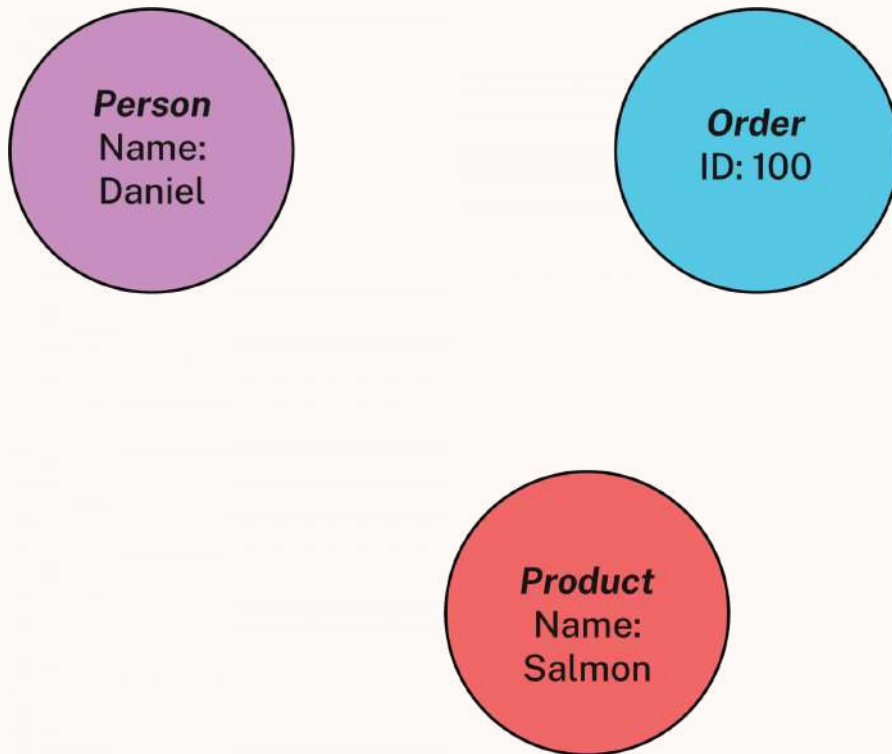


**Every Data Mutation (INSERT, UPDATE, DELETE)**



# What Happens Every Time Data Changes?

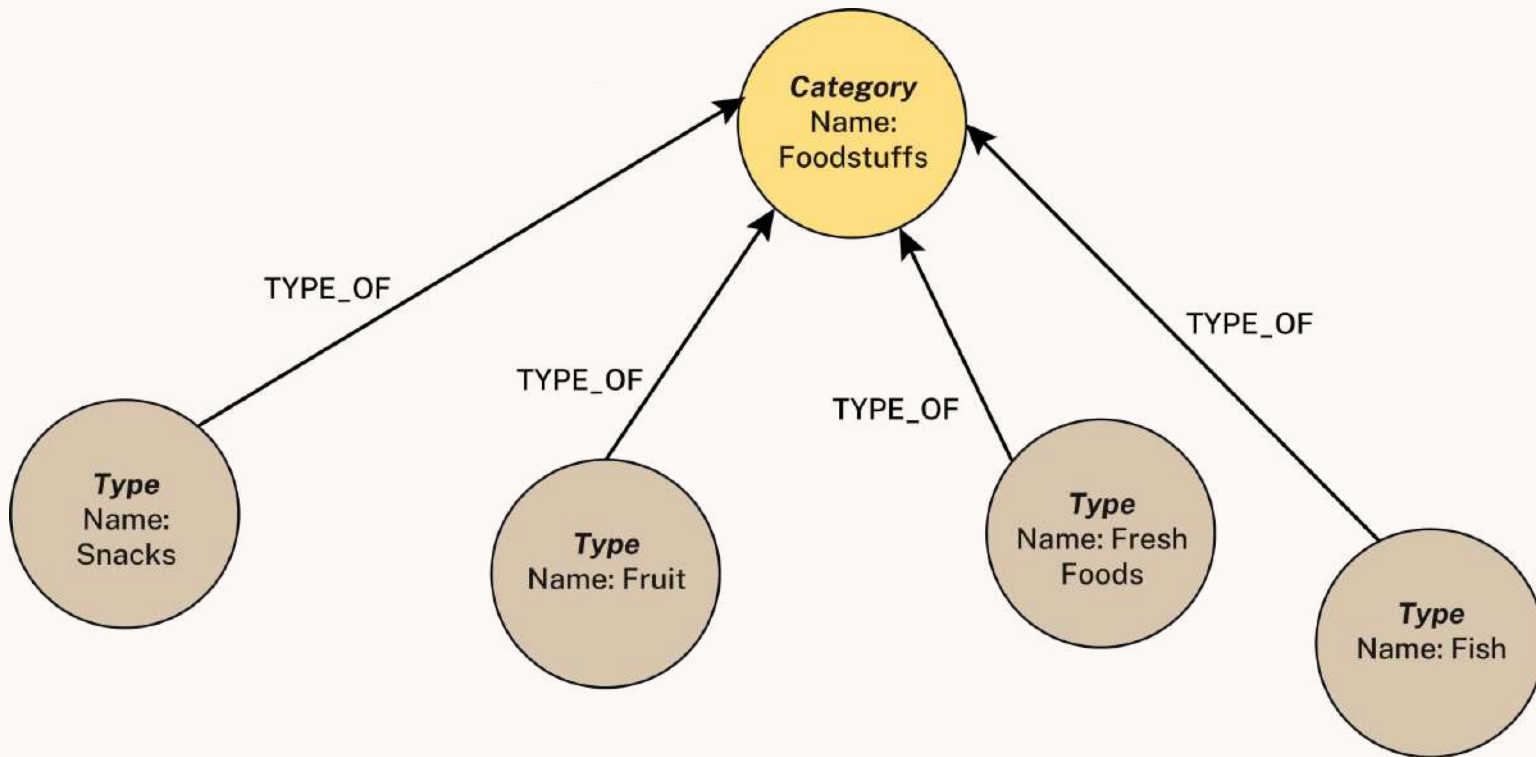
# Nodes



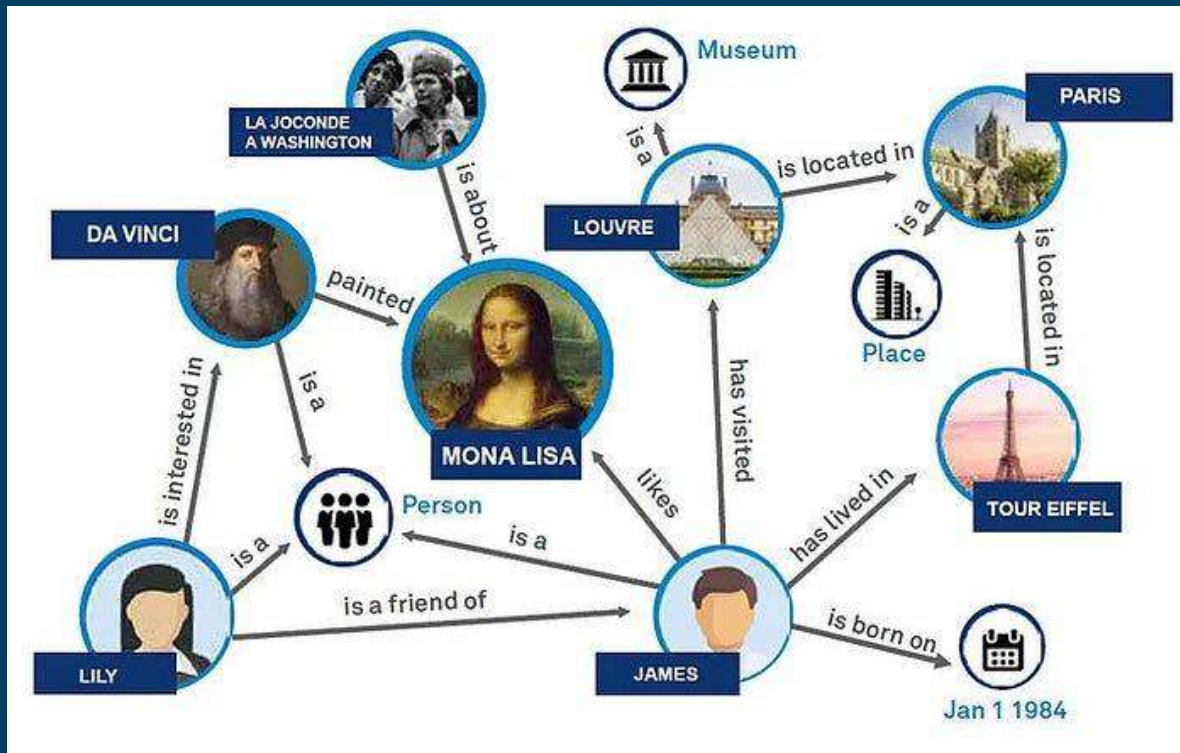
# Relationships



# Organizing Principle(s)



# Semantic Relationships



- Source system pushes change to Kafka
- Target system listens to topic, and consumes messages
- Target system applies changes

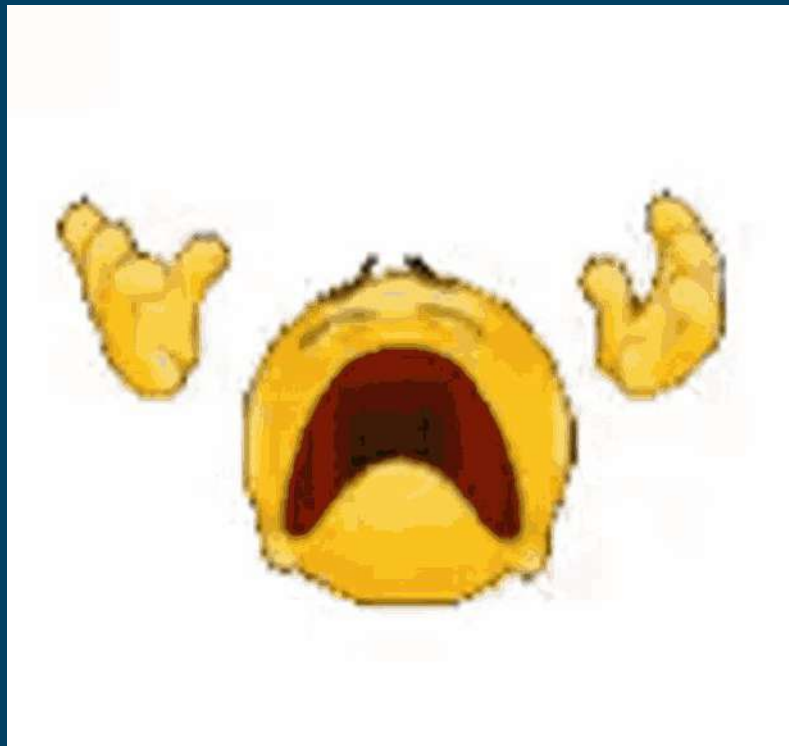


# What Happens Every Time Data Changes?

- Replicate data in other databases (Data Warehouses or Data Lakes)
- Stream processing based on data changes
- Invalidate or update cache
- Asynchronous jobs based on data changes

**DONE : )**

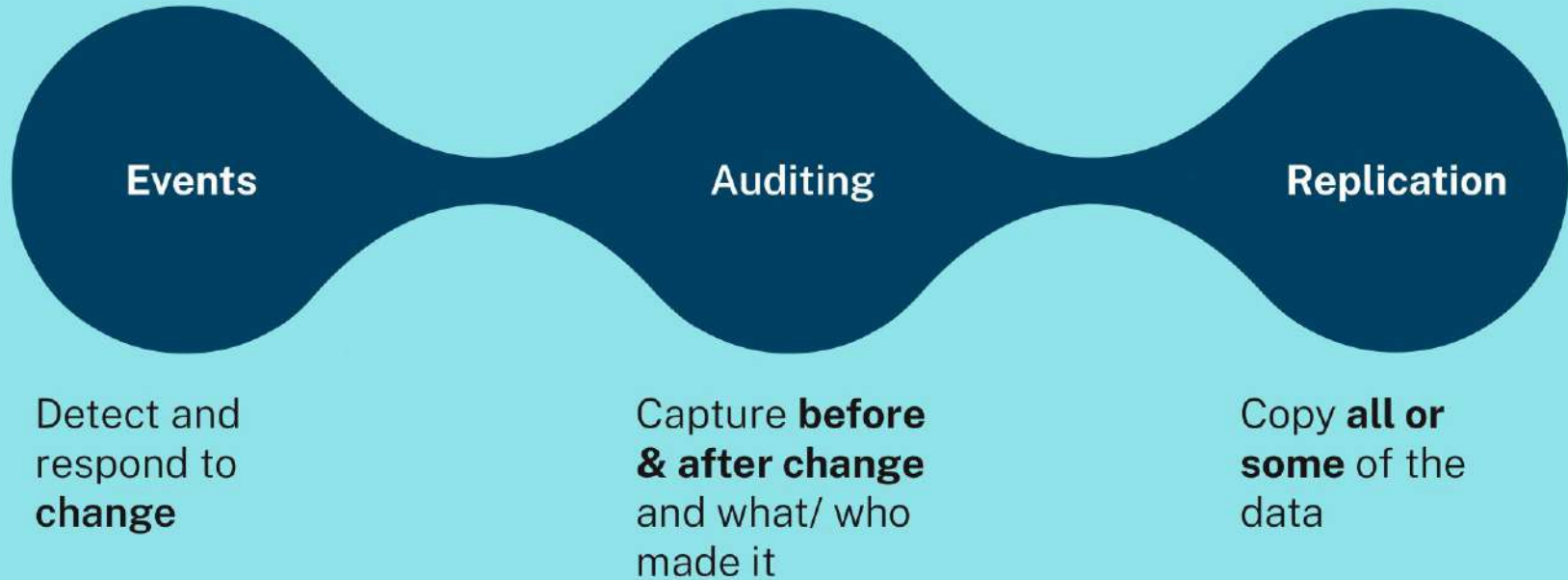
# Before CDC



# After CDC



# What is change data capture & how is it used



# Use cases for CDC in our Early Access Preview

## Crime Prevention

Updating entity resolution models, staying up to date on persons of interest



neo4j



## Connecting data!

Keep disparate data sources up to date.

## Application Audits

Track what applications are composed of, security and compliance reviews in enterprise applications.



## Identity Access Management

Notify access systems when a user's role / access permissions change

# How changes are captured

No plugins or components, no schema changes, no overhead. Get transactional integrity, low latency, scalable & Cypher-free queries for **changes in your graph**.

Step 1  
**Graph Changes**

Step 2  
**Updates Tx Log**

Results  
**cdc.query**

Transaction log based CDC is a state of art implementation **available in many Enterprise RDBMS** –like Oracle and SQL Server.



# SET OPTION txLogEnrichment "mode"



OFF

The **default** no additional info in the logs

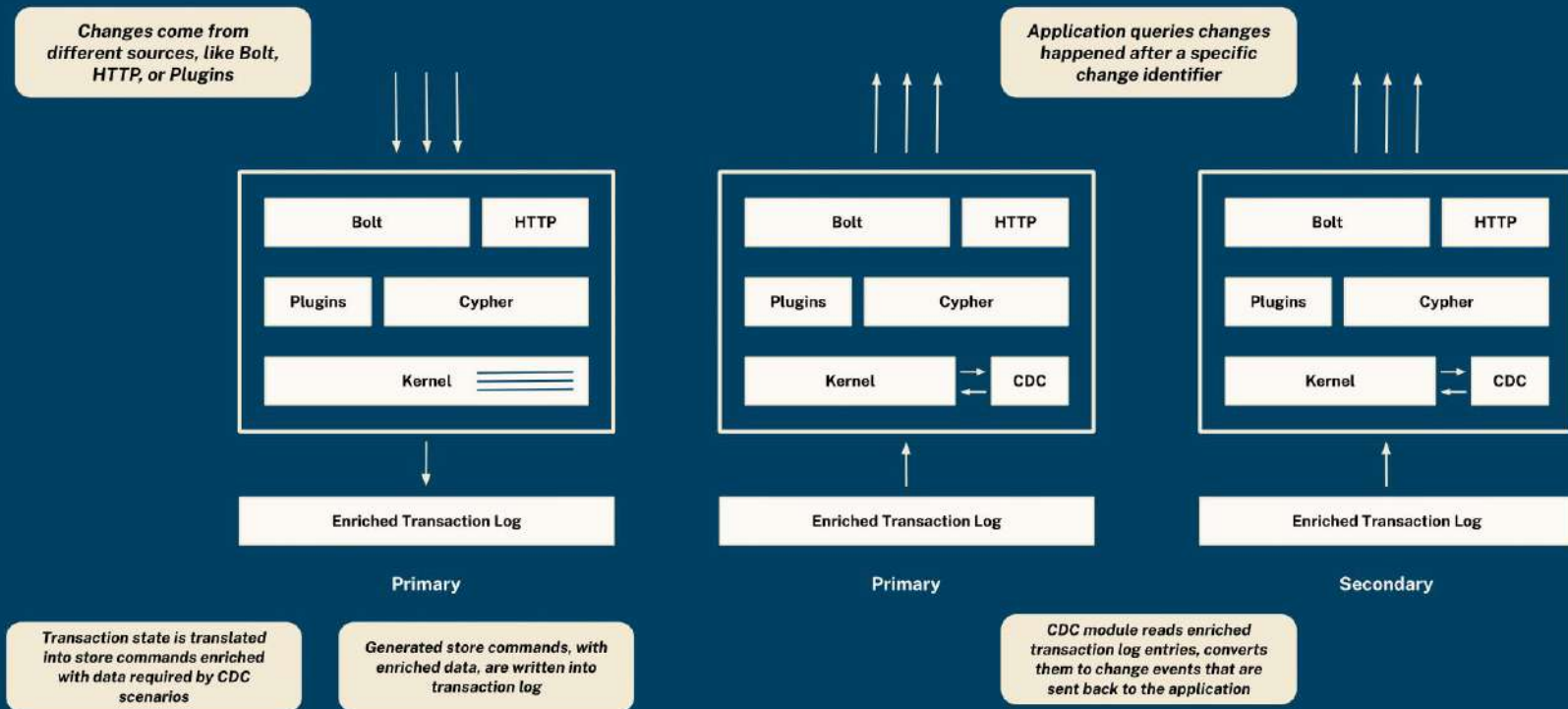
DIFF

Records the **before & after state of what changed** plus metadata - ideal for most event based use cases

FULL

The **before & after state, regardless of what changed** plus metadata - ideal for auditing.

# Change Data Capture under the hood



\*CDC supports single instances as well

# CDC applications and transaction logs



Application calls CDC procedures, and tracks the transactions (Tx) it has read e.g.  $Tx_1$  with a cursor



As time passes, more transactions are written to the logs. The size of the logs can be set in days or GBs.



✓ If the CDC application goes down, the cursor must continue from the last transaction it read.

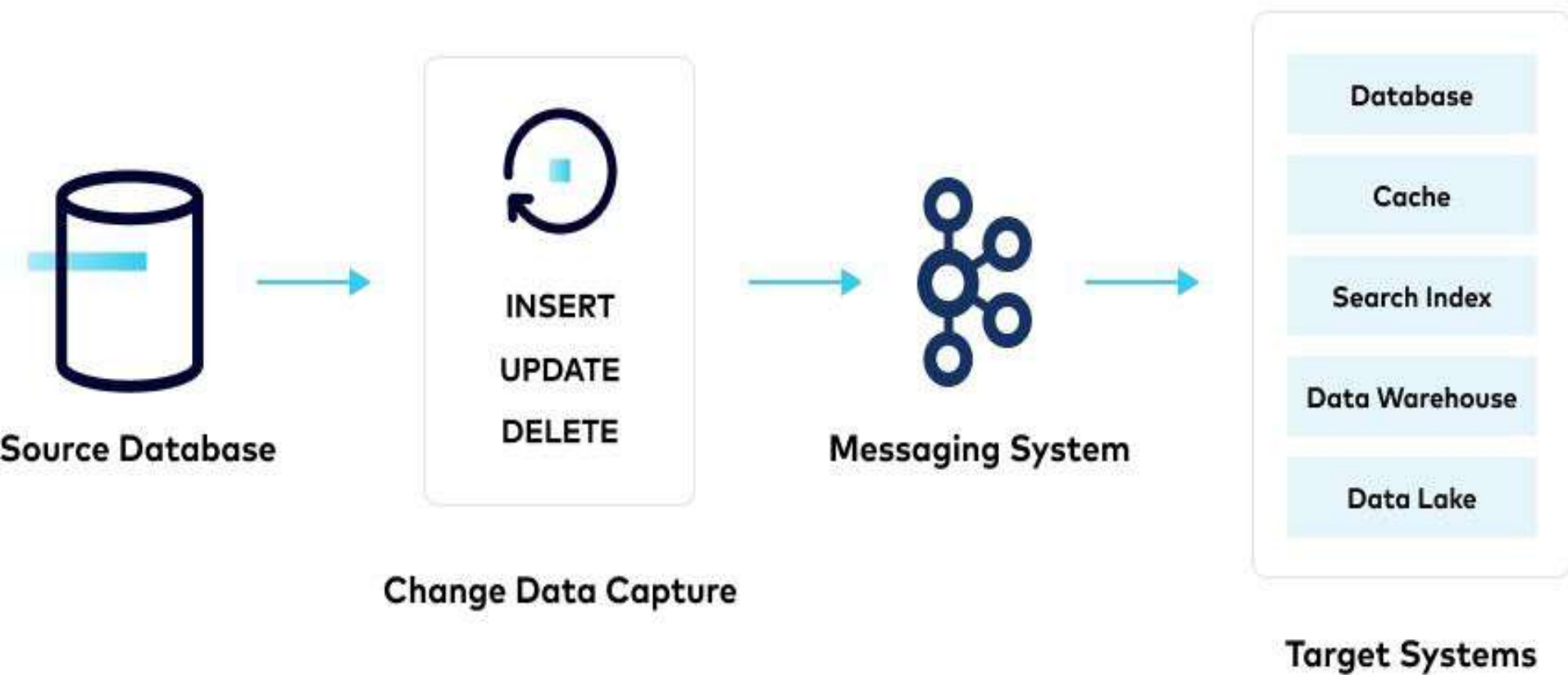


✗ Events are lost IF downtime exceeds the retention policy and logs rollover,  $T_2$  will no longer exist.

## Recommendations

Example the Client has an RTO of 12 hrs, using DIFF

- Use historical trend data to identify max Tx log size for **busiest 12 hr** period e.g. 100GB
- For DIFF plan **50% extra data** e.g. 150GB.
- Available disk space for logs must be greater than 150GB.
- Configure **time based** log retention min 12 hrs **and** a **max usage**  $\geq 150GB$ .
- Set up appropriate monitoring /alerting.



[GET STARTED](#)[DOCS](#)[POWERED BY](#)[COMMUNITY](#)[APACHE](#)[DOWNLOAD KAFKA](#)

### HIGH THROUGHPUT

Deliver messages at network limited throughput using a cluster of machines with latencies as low as 2ms.



### SCALABLE

Scale production clusters up to a thousand brokers, trillions of messages per day, petabytes of data, hundreds of thousands of partitions. Elastically expand and contract storage and processing.



### PERMANENT STORAGE

Store streams of data safely in a distributed, durable, fault-tolerant cluster.



### HIGH AVAILABILITY

Stretch clusters efficiently over availability zones or connect separate clusters across geographic regions.

## ECOSYSTEM



### BUILT-IN STREAM PROCESSING

Process streams of events with joins, aggregations, filters, transformations, and more, using event-time and exactly-once



### CONNECT TO ALMOST ANYTHING

Kafka's out-of-the-box Connect interface integrates with hundreds of event sources and event sinks including Postgres,



### CLIENT LIBRARIES

Read, write, and process streams of events in a vast array of programming languages.



### LARGE ECOSYSTEM OPEN SOURCE TOOLS

Large ecosystem of open source tools: Leverage a vast array of community-driven tooling.

[Product ▾](#)[Solutions ▾](#)[Docs](#)[Customers](#)[Pricing](#)[Resources ▾](#)[Sign In](#)[Sign Up →](#)[NEW FEATURE RELEASE: ASSETS ▸](#)

# It's time to update your orchestrator.

Write Python. Wrap it in a @flow. Prefect brings dynamic scaling, built-in observability, and flexible workflow design to your data pipelines.

[GET STARTED ▸](#)[PREFECT vs. AIRFLOW ▸](#)



WELCOME TO TEMPORAL

# What if your code was crash-proof?

Failures happen. Temporal makes them irrelevant.  
Build applications that never lose state, even when everything else fails.

[Get Started for Free](#)[Get Started with OSS](#)

```
@workflow.defn
class SleepForDaysWorkflow:
    # Send an email every 30 days, for the year
    @workflow.run
    async def run(self) -> None:
        for i in range(12):
            # Activities have built-in support for timeouts and retries!
            await workflow.execute_activity(
                send_email,
                start_to_close_timeout=timedelta(seconds=10),
            )

            # Sleep for 30 days (yes, really)!
            await workflow.sleep(timedelta(days=30))
```

# The **fastest** analytical database for business intelligence

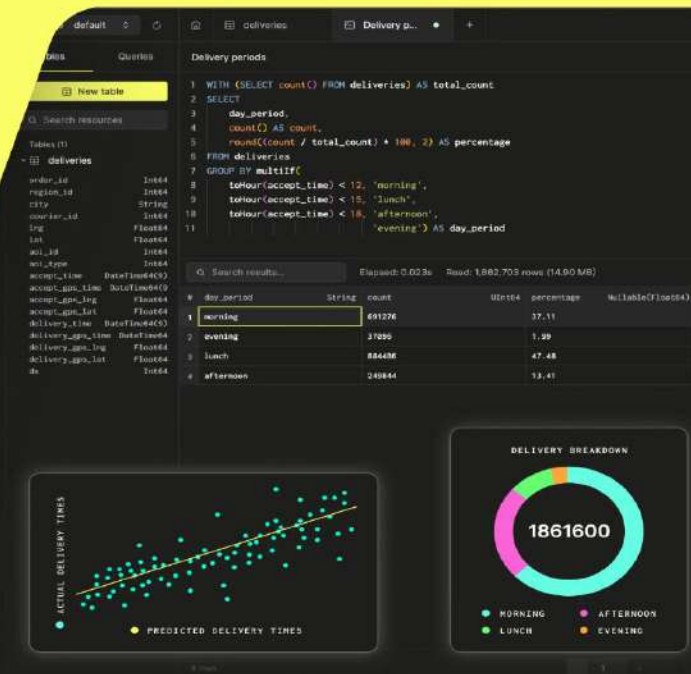
Unlock faster queries without skyrocketing costs.

Start free cloud trial

Download open-source



Open House videos: ClickHouse deep dives, customer stories and user interviews. [Watch now!](#)



CLICKHOUSE IS TRUSTED BY





Apache Superset™ is an open-source modern data exploration and visualization platform.

 Star 66,817
  Watch 1,531
  Fork 15,233

[Get Started](#)



# Querying CDC: CALL

**earliest** obtains the change identifier for the earliest available change\*,  
**current** returns the one for last committed transaction.

`cdc.earliest()`

`cdc.current()`

`cdc.query`  
`("change-id")`

**query** gives you the changes that have happened in the database after the change-id.

\*Based on log retention settings.

# Example

```
CALL cdc.current()
```

```
id "A3V16ZaLLUmnipHLFkWrLA0AAAAAAAAABQAAAAAAAAAA"
```

```
Rows: 1
```

```
CALL cdc.query("A3V16ZaLLUmnipHLFkWrLA0AAAAAAAAABQAAAAAAAAAA")
```

```
// OR you can pass in selectors
```

```
CALL cdc.query("A3V16ZaLLUmnipHLFkWrLA0AAAAAAAAABQAAAAAAAAAA", selectors)
```

# ChangeEvent {

```
"id": "AyajseMlIk_bttd6Mak0eL4AAAAAAAAAAIwAAAAAAAAAA",  
"txId": 35,  
"seq": 0,  
...  
}
```

```
"metadata": {
```

```
  "txStartTime": DateTime(2023, 10, 11, 10, 57, 46, 93500000, tzinfo=<UTC>),
```

```
  "txCommitTime": DateTime(2023, 10, 11, 10, 57, 46, 9780000, tzinfo=<UTC>),
```

```
  "txMetadata": {
```

```
    "app": "your-apps-name",
```

```
    "type": "user-direct"
```

```
  },
```

```
  "authenticatedUser": "neo4j",
```

```
  "executingUser": "neo4j",
```

```
  "connectionType": "bolt",
```

```
  "connectionClient": "192.168.0.5:61120",
```

```
  "connectionServer": "192.168.0.1:7687",
```

```
  "serverId": "e199d2ca",
```

```
  "captureMode": "DIFF"
```

```
}
```

```
"event": {
  "elementId": "4:26a3b1e3-2522-4fdb-b6d0-fa31a93478be:62",
  "eventType": "n",    // Change was for a node, r for relationships
  "operation": "c",    // Create operation, "u" for update, "d" for delete
  "keys": {},
  "labels": [ "Artist" ],
  "state": {
    "after": {
      "labels": [ "Artist", ],
      "properties": {
        "name": "Justin Signed"
      }
    },
    "before": null // This is create, so only see the after state
  }
}
```



# Selectors enable filtering so apps don't have to

```
{  
  "select": "n",                      // "r" for relationships "e" for entities  
  "elementId": "4:b7e35973-0aff-42fa-873b-5de31868cb4a:1",  
  "key": {  
    "property": "value",  
    "otherProperty": "value"  
  },  
  "labels": ["Artist"],  
  "operation": "c",                   // "u" for update "d" for delete  
  "changesTo": ["name", "lastName"] // If both name and lastName are updated  
}
```

Note: Key matching is only possible when specific constraints are defined on the node

Coming soon to select customers

Build loosely coupled event streaming

- **No schema changes**
- **No need to do soft deletes**
- **Simpler Cypher**

Just specify patterns for selectors



## Neo4j Connector for Confluent & Neo4j Connector for Apache Kafka

with Change Data Capture





# Neo4j AuraDB: Fully Managed Graph Database

Mirror your data design like sketching on a whiteboard and adapt effortlessly to changing business needs. Store data and relationships natively with a flexible schema.

[Start Free](#)[View Resources](#)[How AuraDB Works](#)[Capabilities](#)[Use Cases](#)[Customers](#)[Plans](#)[Partners](#)[Resources](#)

## Uncover Hidden Patterns

Reveal insights with Cypher queries and graph algorithms for your connected data.



## Build Applications With Ease

Simplify development using a native graph model, flexible schema, and expressive Cypher.



## Always-On, Zero Admin

99.95% uptime SLA, automated upgrades, patches, and maintenance.



## Enterprise-Grade Security

Protect your data and ensure compliance with robust security features.



WHY CONFLUENT

PRODUCTS

PRICING

USE CASES

RESOURCES

[Start For Free](#)



# The Data Streaming Platform

**Stream, connect, process, and govern** your data with an all-in-one, real-time platform from the pioneer in data streaming. Build faster, scale smarter, and turn data chaos into instantly accessible and usable data products with the market leading Data Streaming Platform.

[Start for Free](#) →

[Explore the Platform](#) →

Or get started through a cloud marketplace for \$1,000 in free credit.



[Amazon Web Services](#)



[Google Cloud](#)



[Microsoft Azure](#)



Founded by the original co-creators of Apache Kafka®

50K+

Clusters operated in Confluent Cloud

3 Trillion

Messages written per day

>1 Exabyte

Data processed per year



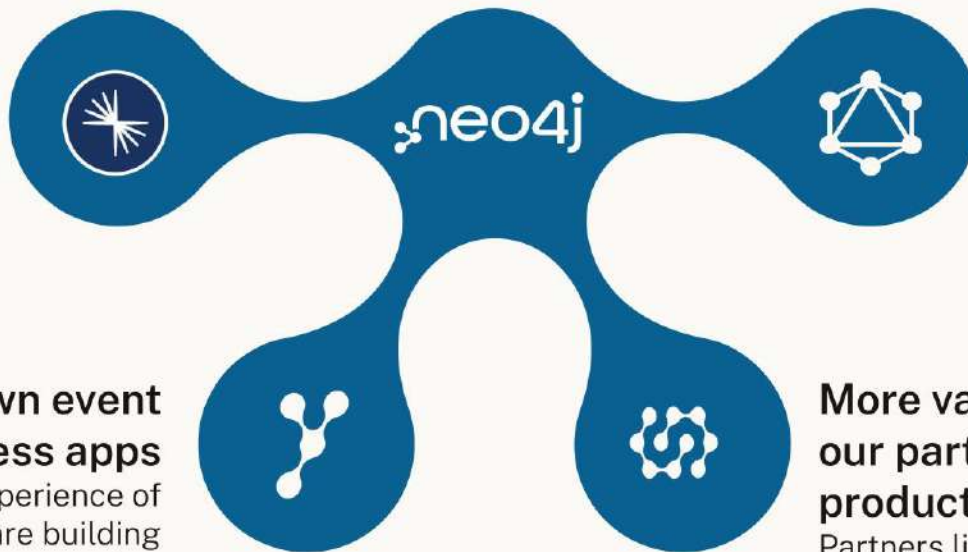
Hi there! How can I help you today?



# Graph at the heart of an ecosystem

## Neo4j Connector for Confluent

Build loosely coupled  
streaming pipelines.



## GraphQL

Build applications with  
subscriptions updated in  
real time.

## Create your own event driven business apps

Improve the user experience of  
the applications you are building  
in your organisation.

## More value from our partner's products

Partners like GraphAware  
are offering great new  
capabilities.

# Getting started with CDC

## Links embedded

Section 1 of 3

### Feedback on Neo4j CDC

We announced the availability of a Beta of Neo4j's Change Data Capture (CDC) at NODES 2023.

Over the course of the Beta, you will be able to

- Use Cypher to control change data capture (full control)
- Create applications that can poll for changes in real-time
- Use in Aura or Neo4j Clusters

If you participated in the EAP we are keen to capture your feedback.

Disclaimer: Being in Beta means that you have the opportunity to provide feedback on a pre-release version of the product.

neo4j Docs

Introduction

Getting Started

Procedures

Selectors

Example usage of Change Data Capture

Change Data Capture

Intro

### Deployment Center

Choose deployment options, drivers and tools.

Graph Database

Graph Data Science

Drivers

Libraries and Connectors

Tools


AuraDB Fully Managed Cloud Service

AuraDB solutions offer flexible plans for your fully managed Neo4j graph database service, wherever you need it.

Aura solutions offer:

Neo4j Developer Blog — @neo4j


### Creating a Custom Connector in Confluent Cloud to Sink Data to Aura for Real-Time Analysis (Part 2)

 Stu Moore, Product Manager  
Aug 06 · 10 mins read

On the 26th of July, we announced support for Confluent's Custom Connectors as a way to connect Confluent Cloud to AuraDB, our fully managed database-as-a-service. In this week's blog, I am going to show you how easy it is to set up a sink connector to create or update data in a database in Aura when messages are written to a topic.

Neo4j Developer Blog — @neo4j

### Using Neo4j AuraDB as a Source and Sink for AWS MSK

 Stu Moore, Product Manager  
Mar 21 · 12 mins read

Neo4j now supports AWS Managed Streaming for Kafka (MSK) through MSK Connect, configured as a source and sink with AuraDB Free.

Back in January, I took on the Product Manager role for Neo4j Connectors—the products which enable customers to stream data into and out of Neo4j using products like **Apache Kafka** / **Apache Spark**.

# Q&A

# Neo4j Monthly Contest



## Welcome to the Neo4j Community Contest!

### Another winner every month!

We're giving away a DJI Mini SE Drone or a prize of equal value every month. The first name drawn at random will snag the prize. Just fill out the form below to enter. We'll pick the winner on the first Wednesday of each month for those who entered the previous month.

Anyone can participate. The lucky winner will get an email from us. Once the winner responds with the Affirmation of Eligibility and Liability/Publicity release filled out, we'll ship the prize right to them. Enter the contest today!

#### Form Title

First Name

bob

Last Name



[dev.neo4j.com/nmc-apac](https://dev.neo4j.com/nmc-apac)

# NODES 25

Back for its seventh year! The online conference for **developers and data pros** ready to learn the **latest graph best practices**.

## CONFERENCE DATE

November 6, 2025

## CALL FOR PAPERS

Submit by June 15

## EVENT FORMAT

Live sessions from community and Neo4j experts - **24 hours** of technical talks **across all timezones**

## THEMES

### Applications

Libraries, Frameworks, and Platforms

### AI Engineering

GenAI, Knowledge Graphs, and RAG

### Data Intelligence

ML, Graph Data Science, and Models

### Graphs

Visualization, Data Integrations, and Tips & Tricks

### Architecture

Frameworks, Data Platforms, Clouds and Beyond



[neo4j.com/nodes-2025](https://neo4j.com/nodes-2025)

Slides→ Connect

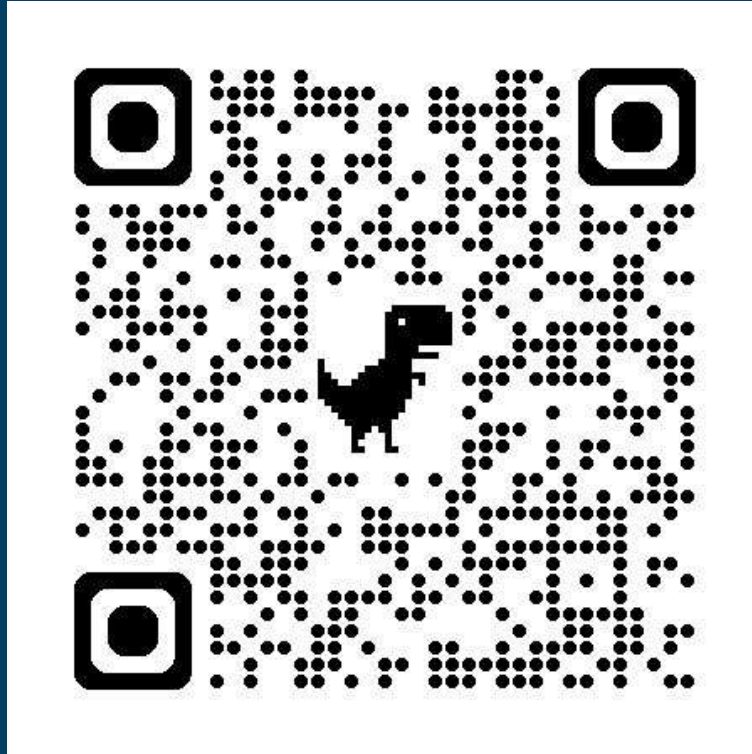
# Slides will be added to this repo within a week





Slides→ Connect

# GraphDB Bangalore Meetups

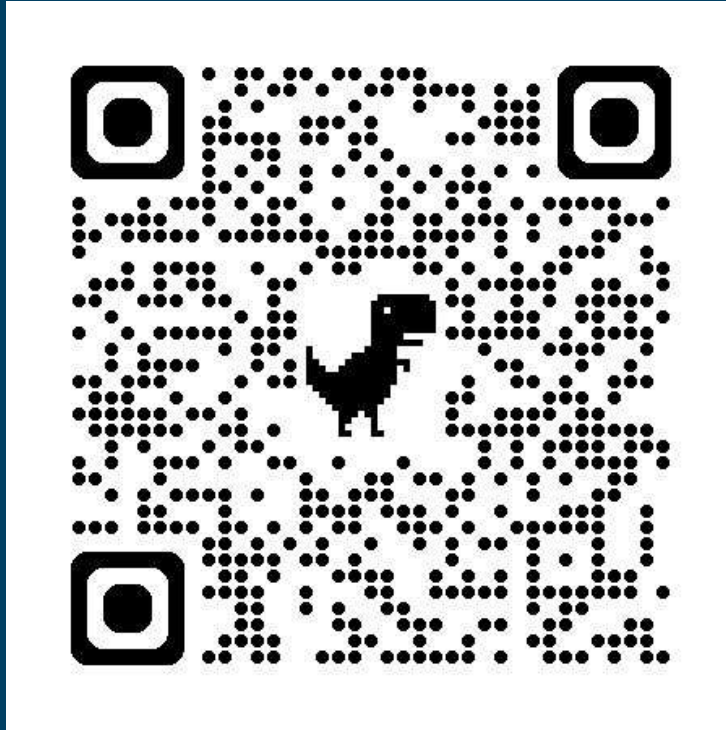


# Spread the word about **Graphs!**

Help your colleagues learn  
more about the technology  
powering our society.

Slides→ Connect

# Connect on LinkedIn



# Thank You

Get early access on

[www.sourcingxpress.com](http://www.sourcingxpress.com)

**Abhishek Das**

abhishek.das@sourcingxpress.com

**SOURCINGXPRESS**

Connect on LinkedIn

