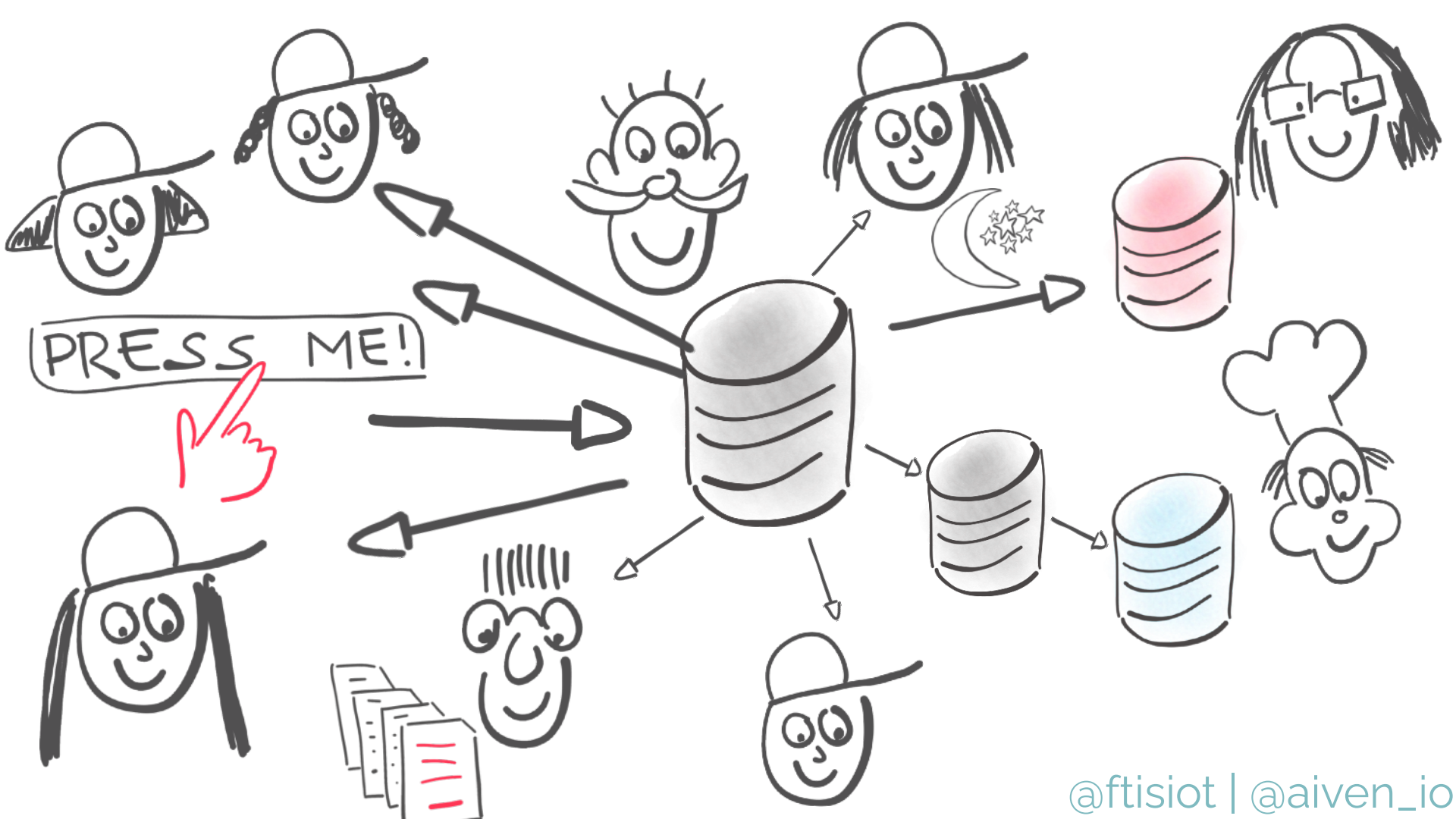
aiven

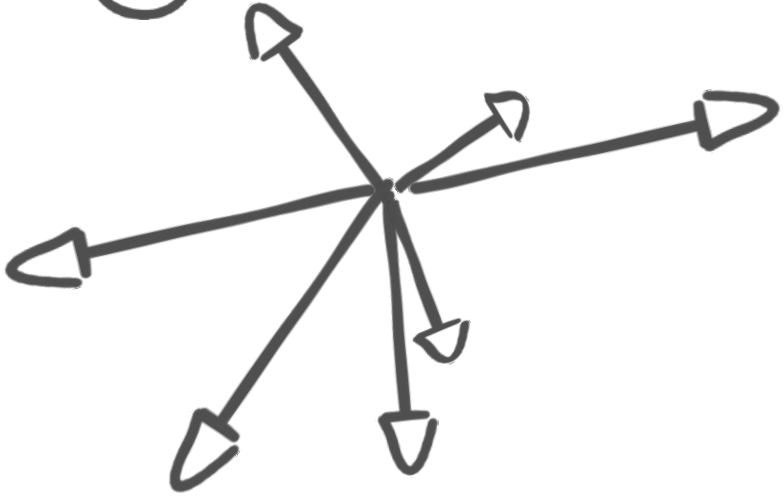
I Don't Want to Miss a Thing

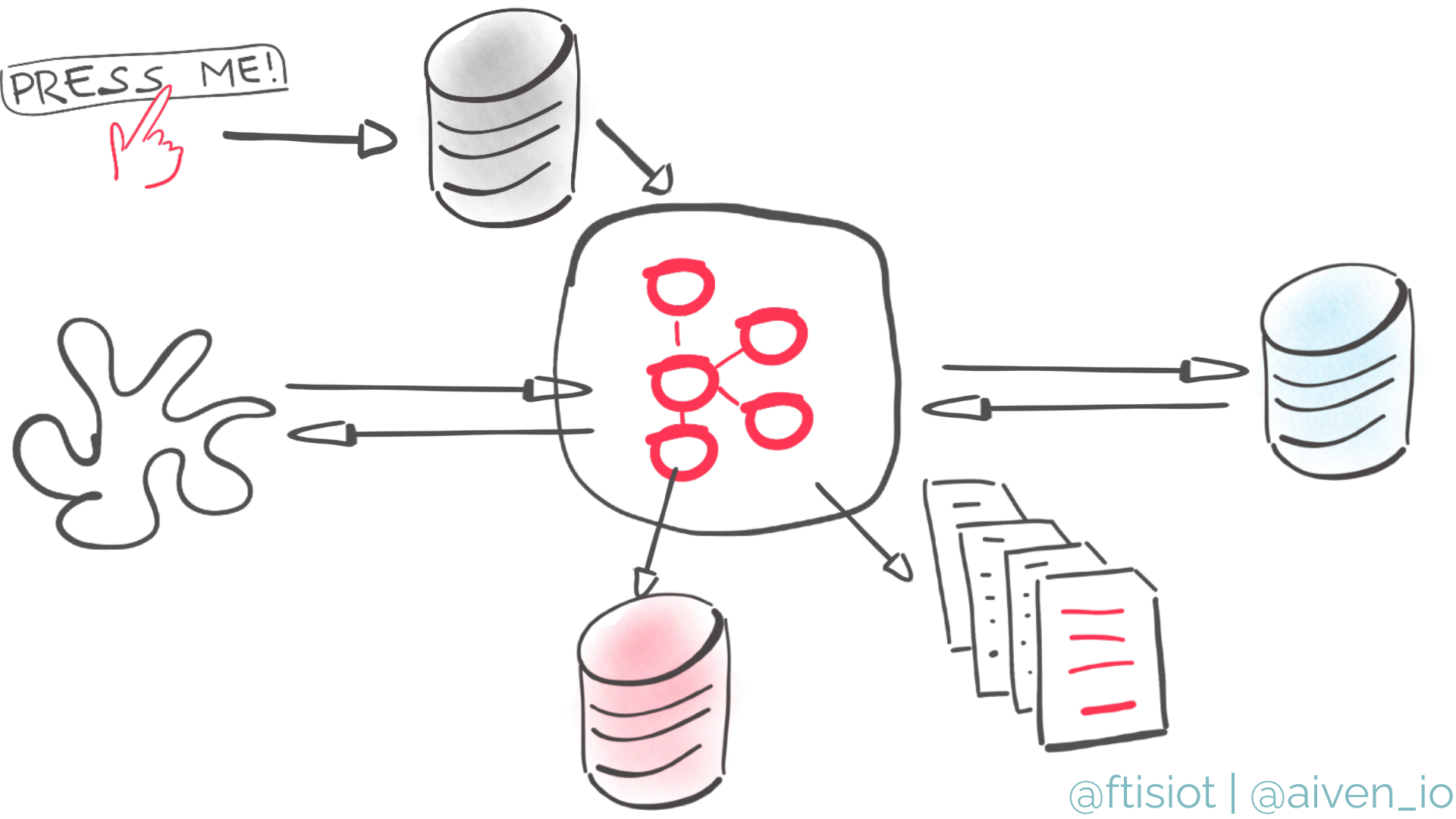
Track Database Changes with Apache Kafka

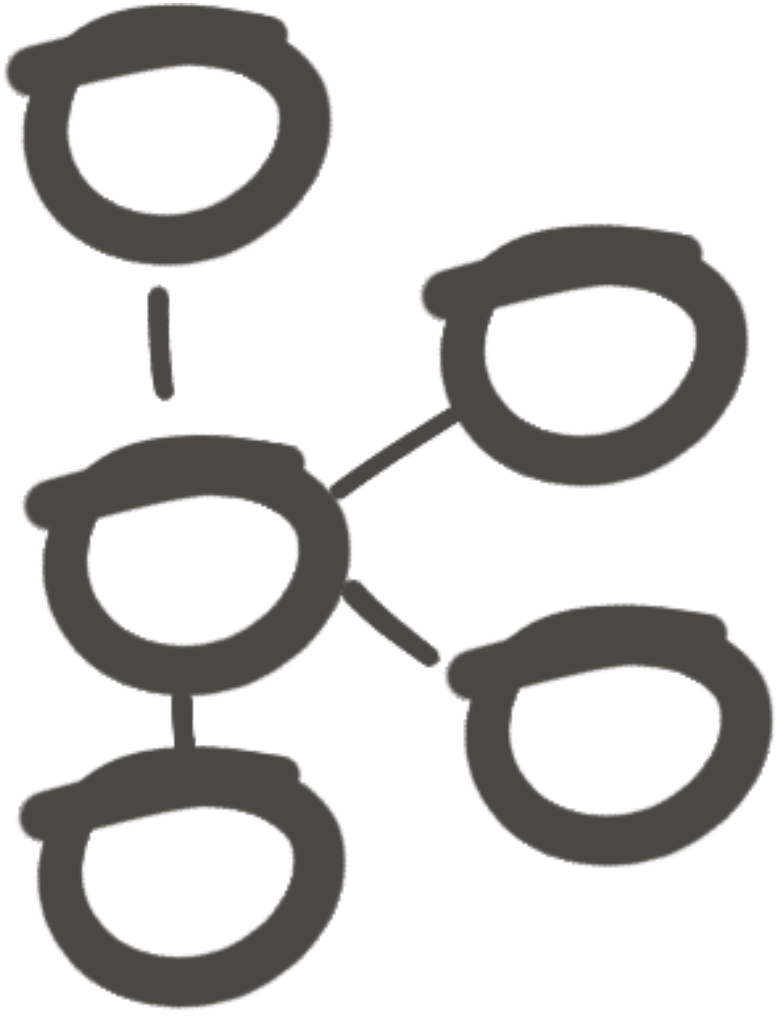
Francesco Tisiot - Developer Advocate

@ftisiot


PRESS ME!




PRESS ME!



What is Apache Kafka?

Topic A

Producer

4 3 2 1 0

Consumer

Topic B

Producer

3 2 1 0

Consumer

Consumer

Events in a Database?

Table

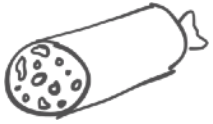
vs

Log

The Fridge Dilemma

Table

5



3

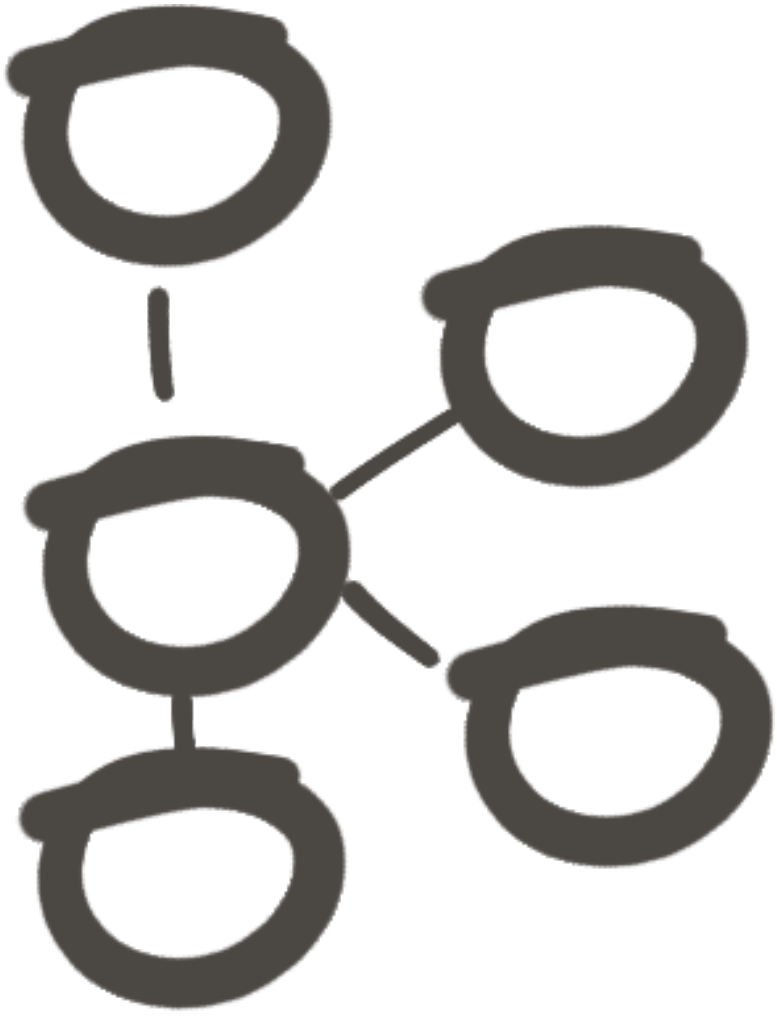
6

Log

+7 -2

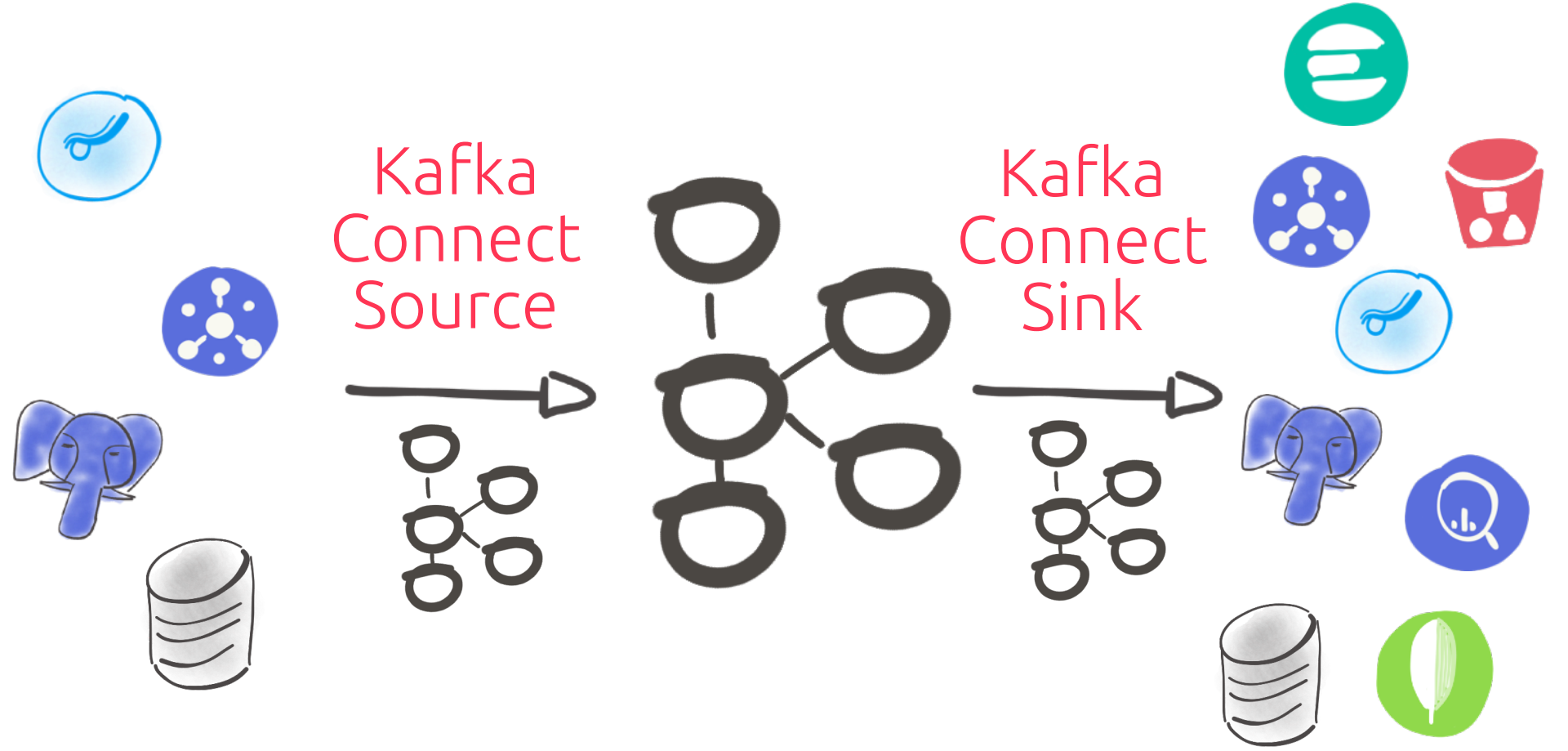
+3

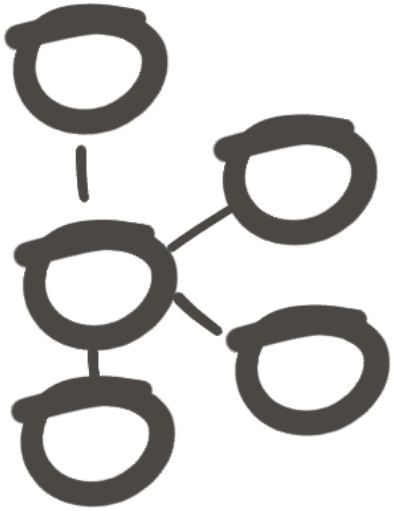
+6 -5 +5



Integrating Apache Kafka

@ftisiot | @aiven_io

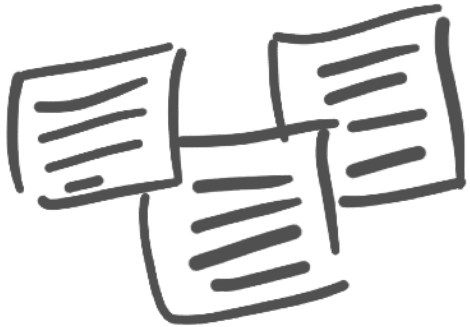

Kafka
Connect
Source
Kafka
Connect
Sink



Changes in Kafka?

Query Based Approach

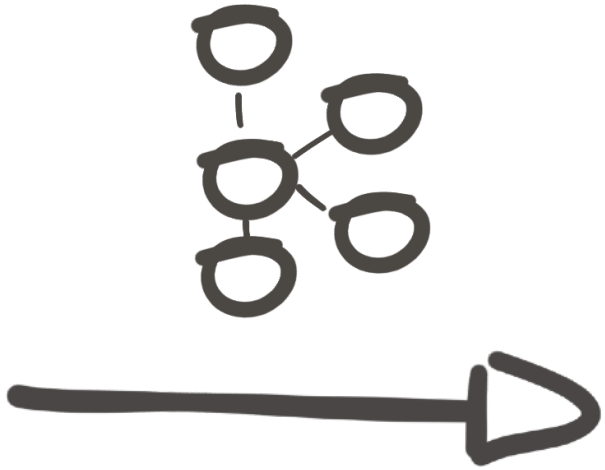
List of Tables



Polling Interval

Query Mode

Pasta Name
Fusilli
Pennette
Farfalle
Spaghetti
Lasagne



Fusilli, Pennette, Farfalle

Fusilli, Pennette, Farfalle

Fusilli, Pennette, Farfalle

Fusilli, Pennette, Farfalle, **Spaghetti**

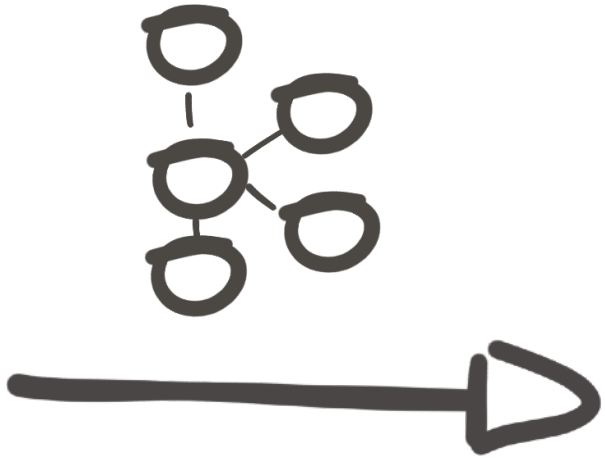
Fusilli, Pennette, Farfalle, Spaghetti

Fusilli, Pennette, Farfalle, Spaghetti

Fusilli, Pennette, Farfalle, Spaghetti, **Lasagne**

Bulk Mode

Id	Pasta Name
1	Fusilli
2	**Penne**
3	Farfalle
4	**Spaghetti**
5	**Lasagne**



1 - Fusilli

2 - Pennette

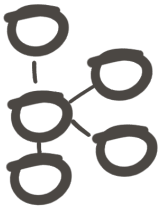
3 - Farfalle

4 - Spaghetti

5 - Lasagne

Incremental Mode

Pasta Name	Created date	Update date
Fusilli	2022-03-01 10:00:00	
Penne	2022-03-01 10:00:00	**2022-03-01 10:03:00**
Farfalle	2022-03-01 10:00:00	
Spaghetti	**2022-03-01 10:01:00**	
~~Lasagne~~	~~2022-03-01 10:02:00~~	



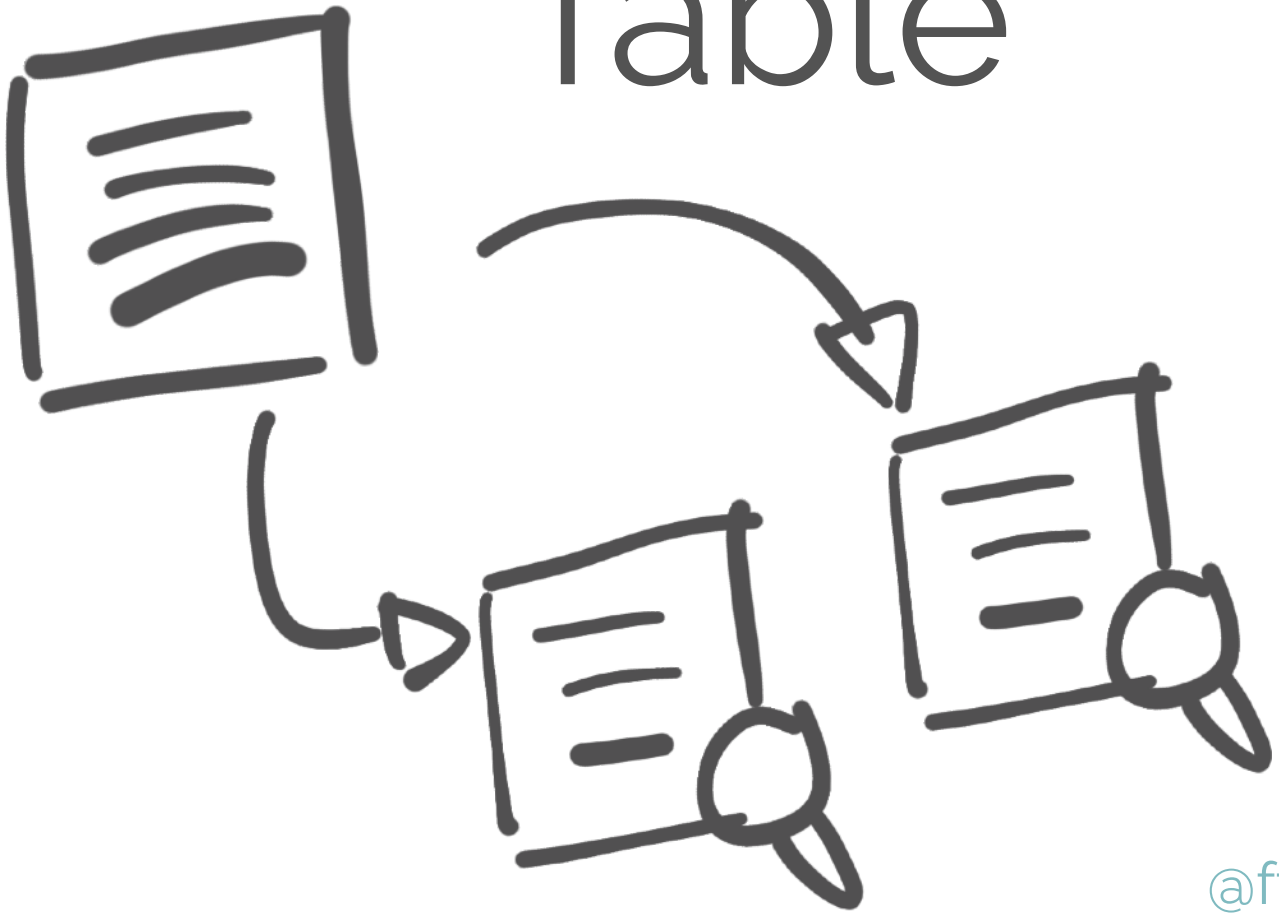
Fusilli Pennette Farfalle Spaghetti Lasagne Penne

Timestamp Mode

```
{
 "name": "pg-bulk-source",
 "connector.class": "io.aiven.connect.jdbc.JdbcSourceConnector",
 "connection.url": "jdbc:postgresql://<HOST>:<PORT>/<DB>?<SSL>",
 "connection.user": "avnadmin",
 "connection.password": "verysecurepassword123",
 "table.whitelist": "pasta",
 "poll.interval.ms": "10000",
 "topic.prefix": "pg_source_",
 "mode": "bulk",
 "tasks.max":"1"
}
```

Max 1 per Table!

“Table”



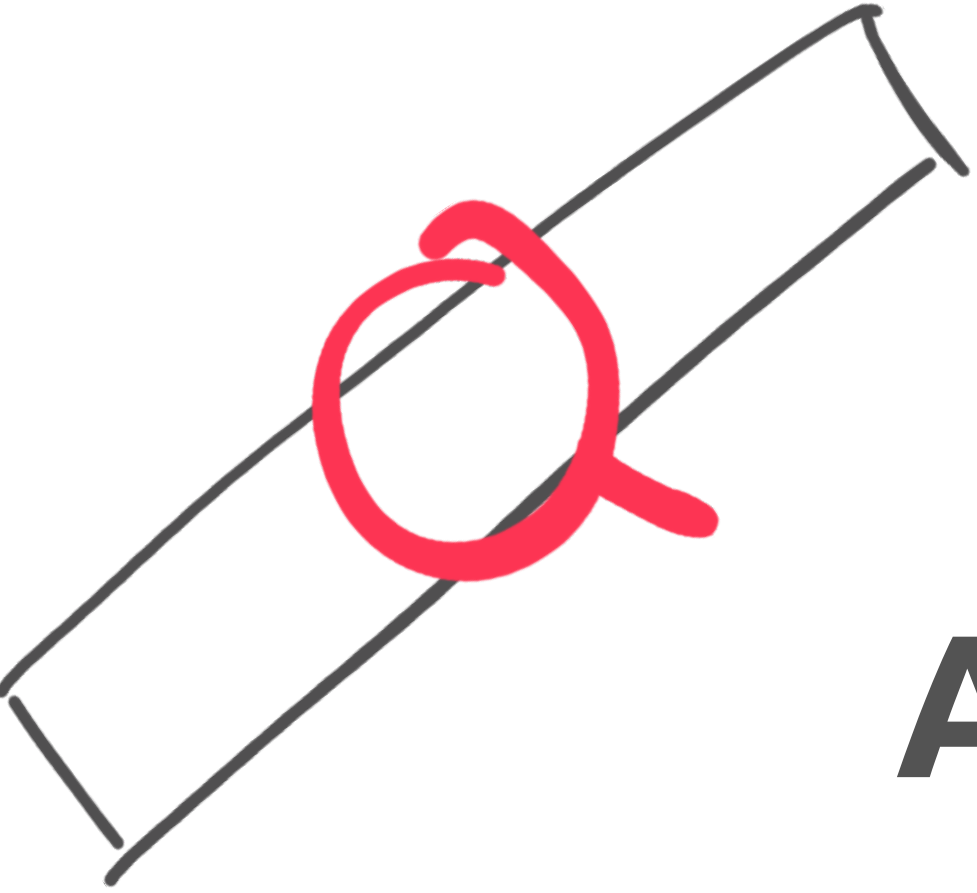
Limits

Updates/Deletions

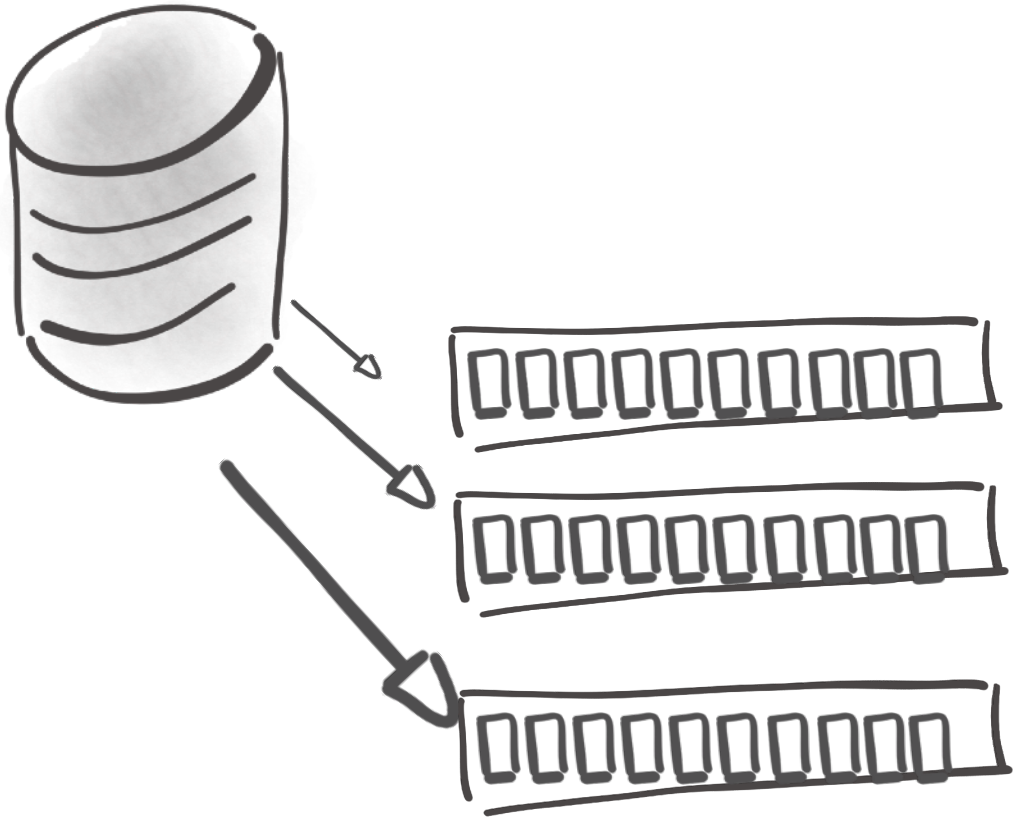
Load on the DB

Polling Time

Out of Order Events



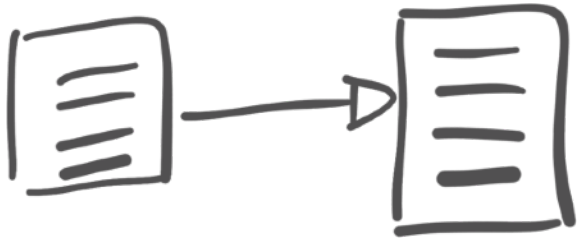
Log Based Approach



Write Ahead Log

Debezium

Replication Slot



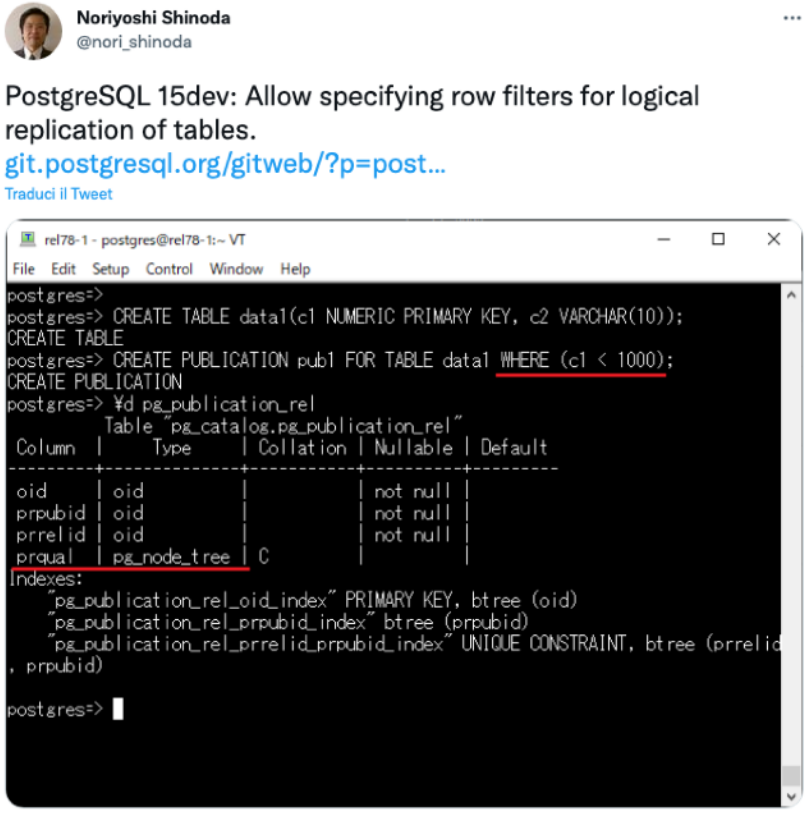
<COL 1, COL 2>

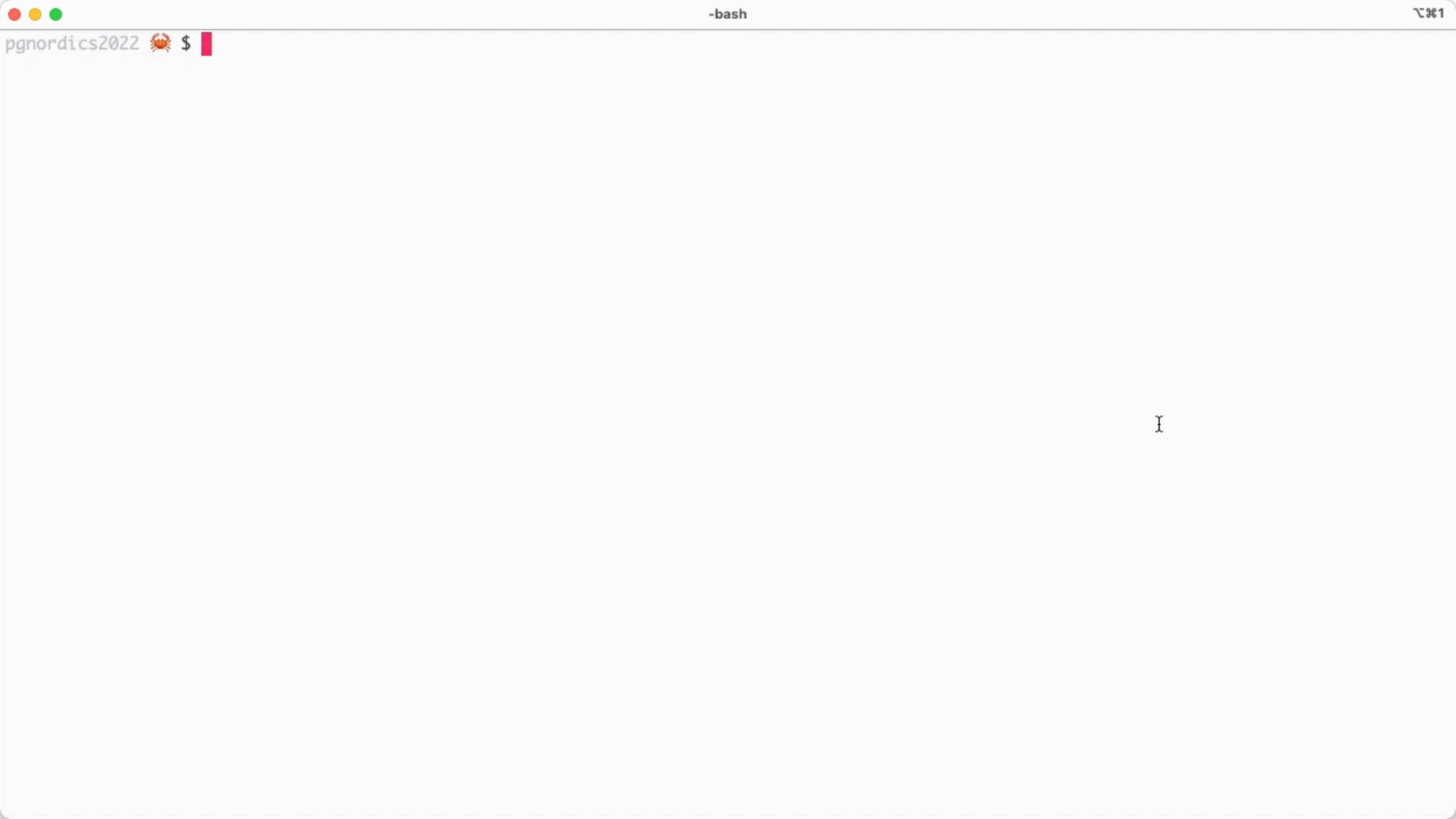
Replica Identity

Publication

```
{
 "name": "cdc-source-pg-json",
 "connector.class":
"io.debezium.connector.postgresql.PostgresConnector",
 "database.hostname": "<HOST>",
 "database.port": "<PORT>",
 "database.user": "avnadmin",
 "database.password": "verysecurepassword123",
 "database.dbname": "<DB>",
 "database.sslmode": "require",
 "plugin.name": "wal2json",
 "slot.name": "<SLOT_NAME>",
 "publication.name": "<PUBLICATION_NAME>",
 "table.include.list": "public.pasta",
 "database.server.name": "<TOPIC_PREFIX>",
 "tasks.max": 1
}
```

Publication




-bash
⌥⌘1
pgnordics2022 🦀 $

Debezium Gotchas

Plugin Types: decoderbufs, wal2json, pgoutput

Log Sequence Numbers (LSNs) Updates

Data Types support

https://debezium.io/documentation/reference/stable/connectors/postgresql.html

Benefits Log Based CDC

Real Time

All Events

Enhanced Metadata

Watch Out for Failovers

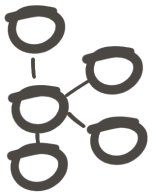
Replication slots are gone!

Accept the data loss

Restart the connector from scratch

Handle failover

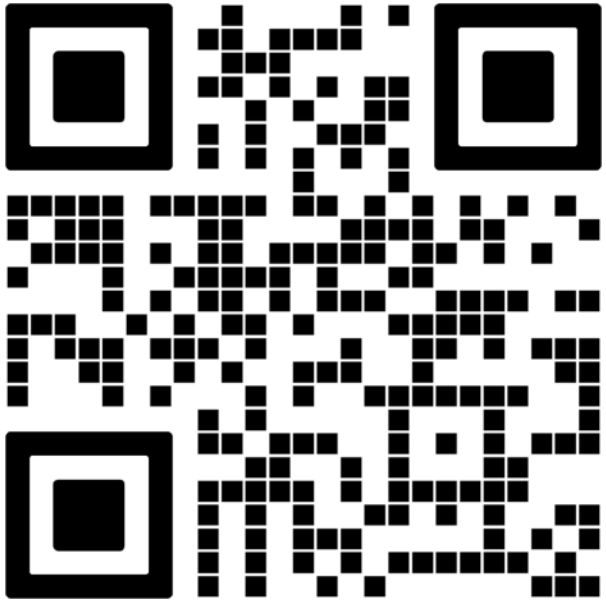
	INSERT	UPDATE	DELETE	NOTES
Query Bulk	✓*	✓*	✓*	
Query Timestamp	✓*	✓*	✗	
Log - Debezium	✓	✓	✓	



Apache Kafka docs

Debezium

JDBC Source Connector in Action

Debezium Connector in Action

https://aiven.io