(Constructs K.T. Part 1)

(Debezium)

(Supported Databases)

* MySQL
* PostgreSQL
* OracleDB ⟹ SQL
* SQL Server
* Db2
* Vitess
* MongoDB ⟹ NoSQL.
* Cassandra3
* Cassandra4

[Kafka Connect]

Kafka Connect helps to quickly tranfer data between Kafka and other systems & vice-versa

Kafka connect operates as a separate servia beside the Kafka broker.

[Kafka Connect can pull & push data from Kafka].

Kafka connect runs with streaming as well as batch oriented fashion.

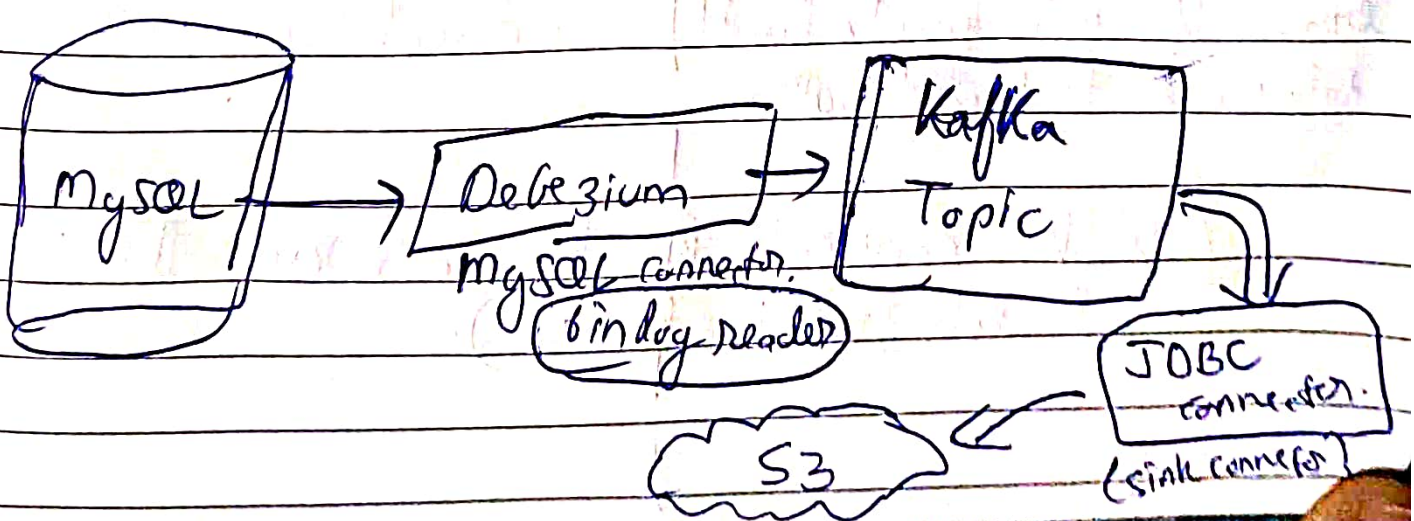
Type of Connectors.

* Source Connectors → { From databases which is the actual transactional databases to Kafka topics }.

* Sink Connectors → { takes data from Kafka topics and dumps into S3, delta lake or HDFS }.

Where does Debezium sits in?

Its a source connector which connects the transactional database to Kafka Topics.



Getting more insights b/w debezium and kafka connect.

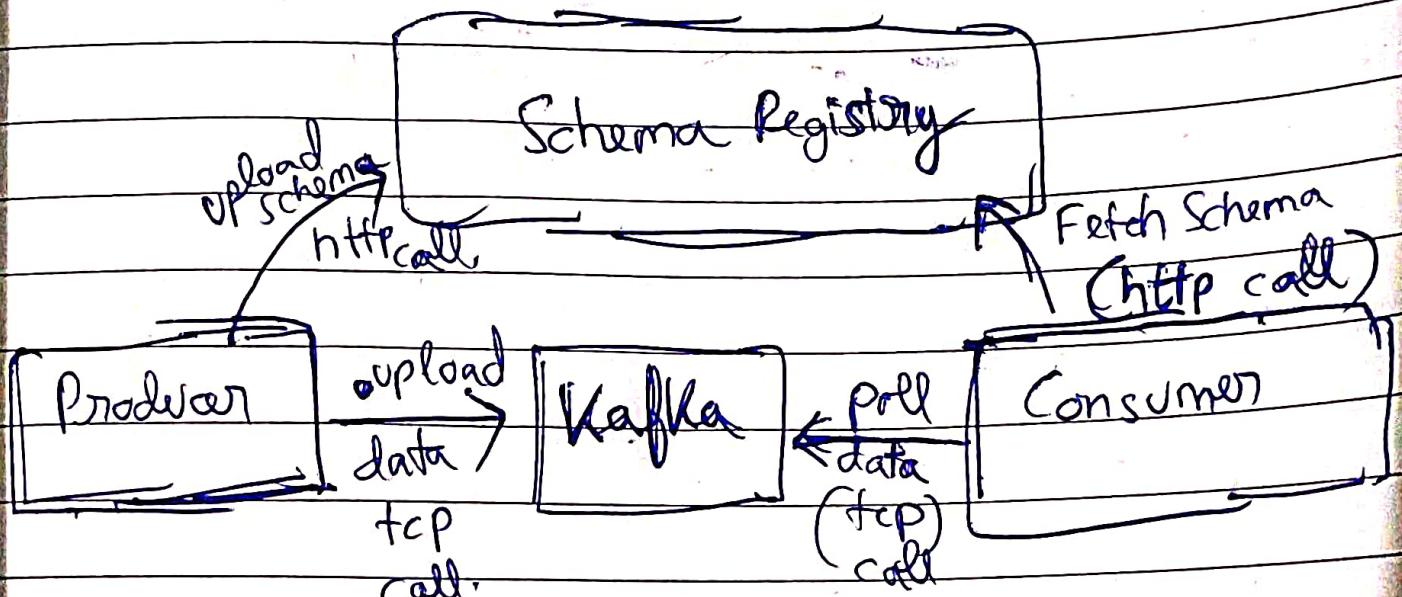
Kafka Connect → Its a connector runtime / framework which runs the process.
(Think of it like tomcat / docker).

Debezium → Contains the cdc logic and gets loaded by Kafka Connect.

Kafka Connect	Debezium Mysql
* Connector runtime / framework ⟶	Contains the cdc logic
* Provides functionalities like offset management, fault tolerance, task distribution.	* Implements Source Connector interfaces

(Kafka connect loads and runs the debezium classes).

~~(Continue VT Part 2)~~



Note, how uploading schema to the schema registry is http call.
Well, its because

i). It does not happen often, i.e. schema does not change much ~~during the event~~

ii). Consumer ~~caches~~ the schema to avoid network call for each batch from Kafka fetch.