**Create ec2 instance (save the key generated after creation in safe and memorable location):**

* Server type = Ubuntu server 18 LTS - 64-bit (x86) (t2.Large 8GB of RAM)
* AMI ID (for reference) = ubuntu/images/hvm-ssd/ubuntu-bionic-18.04-amd64-server-20200112 (ami-0fc20dd1da406780b)

**Access instance via PuTTy**

* <https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/putty.html>

**Install wget:**

* sudo apt install wget

**Install jq:**

* sudo apt install jq

**Install Java 8 and Java 11 JDKs and JREs**:

* sudo apt update
* sudo apt install --reinstall default-jre default-jdk default-jre-headless default-jdk-headless openjdk-11-jdk openjdk-11-jdk-headless openjdk-11-jre openjdk-11-jre-headless openjdk-8-jdk openjdk-8-jdk-headless openjdk-8-jre openjdk-8-jre-headless

*remove the “—reinstall” option if this is the first installation attempt for Java*

**Check Java installation:**

* java –version

should say currently uses Java 11

**Switch to Java 8 as default:**

* sudo update-alternatives --config java (pick the number for Java 8 and press ENTER)

**Move to root (in case, not there already:**

* cd

**Download Confluent Platform Enterprise 5.0 or above (choose option #1 OR #2 but not both):**

1. sudo wget <https://packages.confluent.io/archive/5.4/confluent-5.4.0-2.12.tar.gz>
2. Visit <https://www.confluent.io/previous-versions/?_ga=2.51348321.970298139.1584974979-1665271302.1581446223&_gac=1.175301910.1584726100.EAIaIQobChMIu5_g2syp6AIVzv7jBx01YQ7sEAAYAiAAEgIBQ_D_BwE>

Then copy and paste the link to the download file of your choice into a text editor (you can grab the link by right clicking on the desired button and choosing “copy link address”). Remove all the characters after the “tar.gz”. Append newly modified link to the end of a “sudo wget” command in the instance terminal. Execute command.

**Unpack Confluent Platform Tarbel files:**

* tar –xf confluent-5.4.0-2.12.tar.gz
* wait for above command to finish executing (no progress bar will be shown) It should take about 10-15 seconds.

*If getting any errors concerning invalid options, type the command manually and retry. Copying and pasting the code from the Word document doesn’t always translate.*

**Delete Tarbel:**

* rm confluent-5.4.0-2.12.tar.gz
* if prompted with question about removing file: type “y” and press ENTER

**Create Confluent environment variable:**

* export CONFLUENT\_HOME=~/confluent-5.4.0

**Add Confluent bin directory to path:**

* export PATH=$PATH:$CONFLUENT\_HOME/bin

*Both the path and environment variable are temporary and thus must be recreated after closing shell*

**Install Confluent CLI:**

* curl -L --http1.1 https://cnfl.io/cli | sh -s -- -b ~/confluent-5.4.0/bin

**Test boot Confluent Platform:**

* confluent local start

**OR** $CONFLUENT\_HOME/bin confluent local start (do not remove dollar sign)

*“confluent local stop” will stop all Kafka processes.*

*“confluent local destroy” should be executed before restarting or else you might get error concerning corrupt indexes and the kafka broker will not start.*

***Stop Confluent Platform***

* confluent local stop

**OR** $CONFLUENT\_HOME/bin confluent local start (do not remove dollar sign)

**Move to predesignated kafka-connect directory:**

* cd ~/confluent-5.4.0/share/java (or something similar)

**Should be directory with the following:**

*acl kafka kafka-connect-storage-common*

*confluent-common kafka-connect-activemq kafka-rest*

*confluent-control-center kafka-connect-elasticsearch kafka-serde-tools*

*confluent-hub-client kafka-connect-ibmmq ksql*

*confluent-kafka-mqtt kafka-connect-jdbc monitoring-interceptors*

*confluent-metadata-service kafka-connect-jms rest-utils*

*confluent-rebalancer kafka-connect-replicator schema-registry*

*confluent-security kafka-connect-s3*

**Download Debezium plugin:**

* sudo wget <https://repo1.maven.org/maven2/io/debezium/debezium-connector-mysql/0.7.2/debezium-connector-mysql-0.7.2-plugin.tar.gz>
* sudo wget https://repo1.maven.org/maven2/io/debezium/debezium-connector-mysql/1.1.0.Final/debezium-connector-mysql-1.1.0.Final-plugin.tar.gz

**Unpack Debezium plugin:**

* tar –xf debezium-connector-mysql-0.7.2-plugin.tar.gz

**Delete Tarbel file:**

* rm debezium-connector-mysql-0.7.2-plugin.tar.gz

**Start Kafka services:**

* confluent local destroy
* confluent local start

**Link Debezium connector:**

* curl -k -X POST -H "Accept:application/json" -H "Content-Type:application/json" localhost:8083/connectors/ -d '{ "name": "inventory-connector", "config": { "connector.class":"io.debezium.connector.mysql.MySqlConnector", "database.hostname": "hostname", "database.port":"3306", "database.user": "user", "database.password": "password","database.server.id": "1", "database.server.name": "dbz", "table.whitelist":"inventory.usa\_covid19\_test", "database.history.kafka.bootstrap.servers": "localhost:9092","database.history.kafka.topic": "schema-registry.inventory" , "transforms": "route", "transforms.route.type":"org.apache.kafka.connect.transforms.RegexRouter", "transforms.route.regex": "([^.]+)\\.([^.]+)\\.([^.]+)", "transforms.route.replacement": "$3"}}'

// whitelist will watch the specified database OR table

// database.history.kafka.topic specifies that topic the CDC will write to, I assume.

**List topics (should see topics created from the link above including names of database and table):**

* kafka-topics --list --zookeeper localhost:2181 (from root directory)

**Should see database created topics at the bottom of console output like so:**

dbz

dbz.inventory

dbz.inventory.test

dbz.inventory.test2

**Create consumer to view all messages:**

kafka-avro-console-consumer \

--bootstrap-server localhost:9092 \

--property schema.registry.url=http://localhost:8081 \

--topic dib\_posts --from-beginning --max-messages=1 | jq '.'

**Create Elastic Search Connector:**

* ~/confluent-5.4.0/etc/kafka-connect-elasticsearch/
* vim quickstart-elasticsearch.properties
* **Update .properties file with the following text:**

name=elastic-sink

connector.class=io.confluent.connect.elasticsearch.ElasticsearchSinkConnector

tasks.max=1

topics=dip\_posts

connection.url=connection

transforms=unwrap,key

transforms.unwrap.type=io.debezium.transforms.ExtractNewRecordState

#transforms.unwrap.type=io.debezium.transforms.UnwrapFromEnvelope

transforms.unwrap.drop.tombstones=false

transforms.key.type=org.apache.kafka.connect.transforms.ExtractField$Key

transforms.key.field=id

key.ignore=false

type.name=post

behavior.on.null.values=delete

* confluent local load elasticsearch-sink

**View Elastic Search records (limited to 5):**

* curl 'https://connection /index/\_search?pretty'

**OR**

* curl -XGET 'https://connection/index'

**Show active connectors:**

* confluent local status connectors|grep -v Writing| jq '.[]'| xargs -I{connector} confluent local status {connector}| grep -v Writing| jq -c -M '[.name,.connector.state,.tasks[].state]|join(":|:")'| column -s : -t| sed 's/\"//g'| sort

View Data in Kibana:

* Click link below “Use Elasticsearch data” that read “Connect to your Elasticsearch index”
* Visualize data

**Delete index:**

* curl -XDELETE 'https://connection/index/’

**Helpful links:**

* Kafka tutorials: <https://developer.confluent.io/>
* Unwrap tutorial (Docker) - <https://github.com/debezium/debezium-examples/tree/master/unwrap-smt>
* Debezium installation tutorial (didn’t work for me, but might still be helpful): <https://rmoff.net/2018/03/24/streaming-data-from-mysql-into-kafka-with-kafka-connect-and-debezium/>
* Debezium plugin repo: <https://repo1.maven.org/maven2/io/debezium/debezium-connector-mysql/>