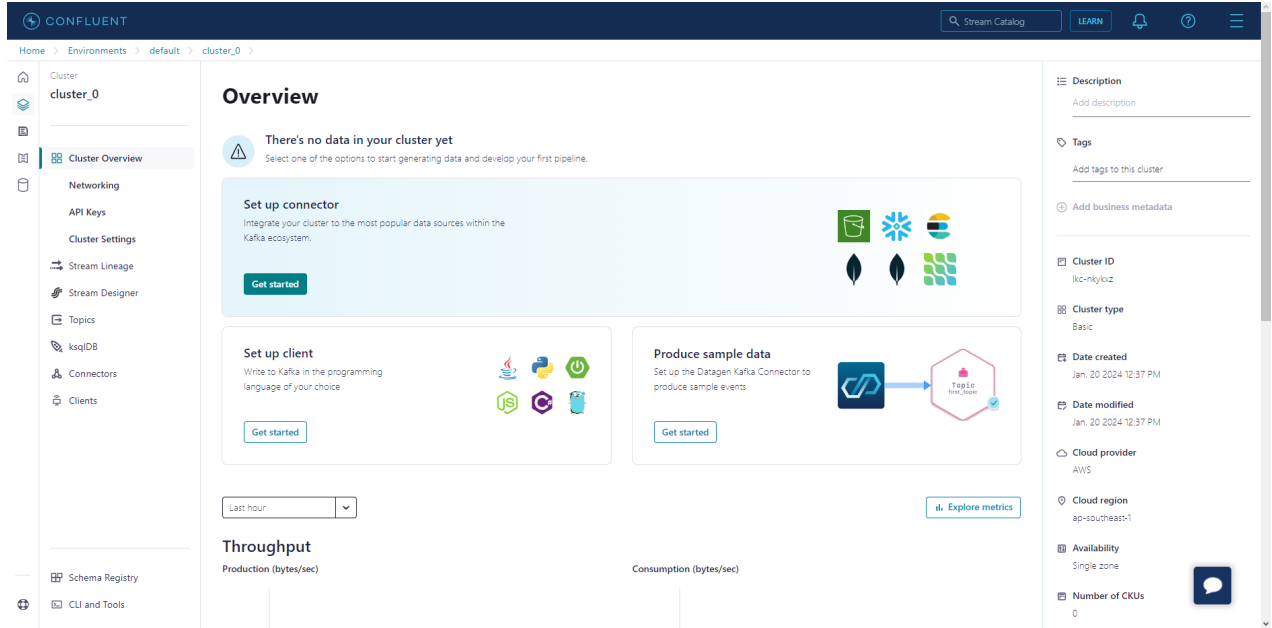
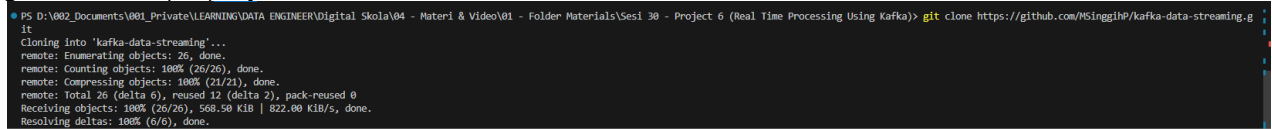


1. Introduction about Confluent

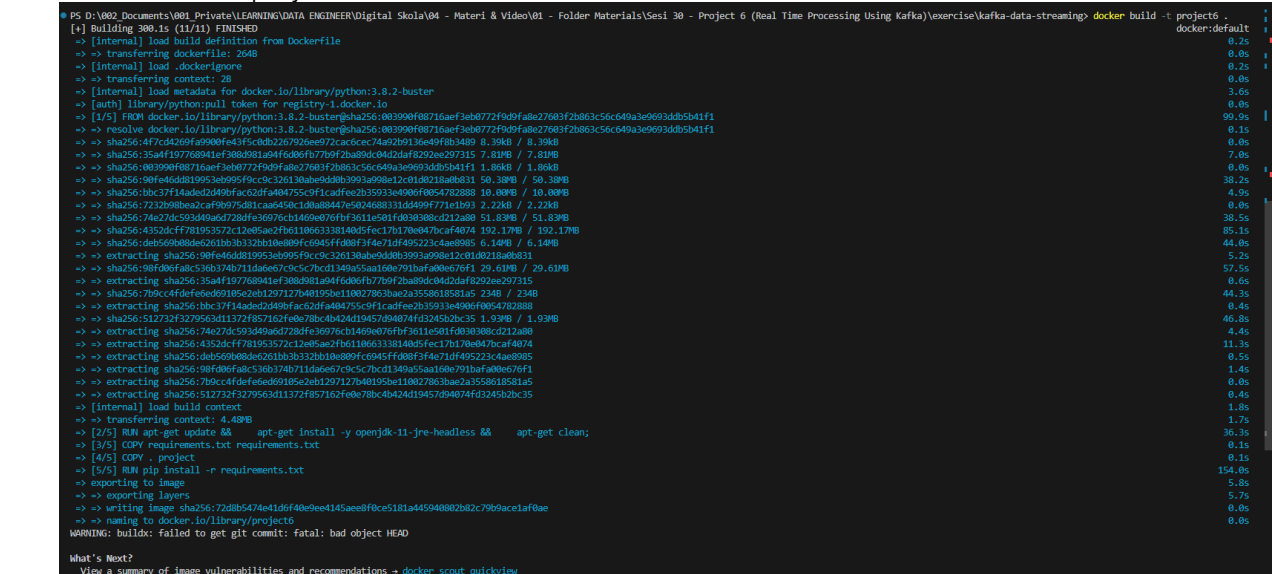


2. git clone a repo project

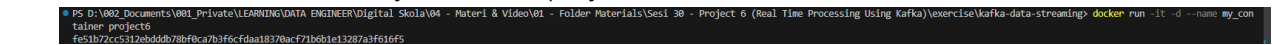


3. Docker build and run and check version

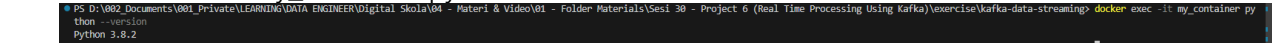
- docker build -t project6 .



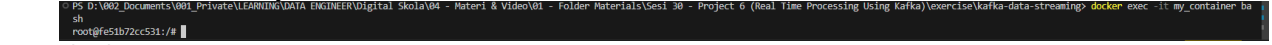
- docker run -it -d --name my_container project6



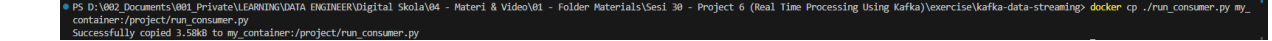
- docker exec -it my_container python --version



- docker exec -it my_container bash



- docker cp ./run_consumer.py my_container:/project/run_consumer.py



- run for check consumer topic data-row: ``sudo docker exec -it my_container python /project/run_consumer.py /project/credentials.ini ridwan-data-row``

```
0:00 PM,19.5,0.69,1.678,0.135,2,11,06/22/2014 8:00 PM,OhioStreetBea
Waiting...
Topic ridwan-data-raw, Event msg = Rainbow Beach,06/22/2014 08:00:00
PM,20.3,1.04,1.475,0.187,3,11,2,06/22/2014 8:00 PM,RainbowBeach2014
06222000
Waiting...
Topic ridwan-data-raw, Event msg = Calumet Beach,06/23/2014 05:00:00
AM,19.5,0.86,1.413,0.083,2,10,7,06/23/2014 5:00 AM,CalumetBeach2014
06230500
Waiting...
Topic ridwan-data-raw, Event msg = 63rd Street Beach,06/23/2014 05:0
0 AM,16.9,0.76,1.282,0.126,2,11,5,06/23/2014 5:00 AM,63rdStreetBea
ch201406230500
Waiting...
Topic ridwan-data-raw, Event msg = Montrose Beach,06/23/2014 05:00:00
AM,17.6,0.42,1.595,0.148,2,10,9,06/23/2014 5:00 AM,MontroseBeach20
1406230500
Waiting...
Topic ridwan-data-raw, Event msg = Osterman Beach,06/23/2014 05:00:00
AM,18.4,0.68,1.662,0.121,2,11,06/23/2014 5:00 AM,OstermanBeach20
1406230500
Waiting...
Topic ridwan-data-raw, Event msg = Ohio Street Beach,06/23/2014 05:0
0 AM,18.4,0.68,1.662,0.121,2,11,06/23/2014 5:00 AM,OhioStreetBeach
201406230500
Waiting...
Rainbow Beach,06/22/2014 07:00:00 PM,20.4,1.06,1.567,0.178,2,11,2,06
/22/2014 7:00 PM,RainbowBeach201406221000
Calumet Beach,06/22/2014 08:00:00 PM,20.5,0.94,1.368,0.099,3,10,8,06
/22/2014 8:00 PM,CalumetBeach201406222000
63rd Street Beach,06/22/2014 08:00:00 PM,20.1,0.78,1.707,0.171,6,11,
3,06/22/2014 8:00 PM,63rdStreetBeach201406222000
Montrose Beach,06/22/2014 08:00:00 PM,17.7,0.35,1.586,0.175,2,10,9,0
6/22/2014 8:00 PM,MontroseBeach201406222000
Osterman Beach,06/22/2014 08:00:00 PM,17.4,1.08,1.225,0.147,2,11,5,0
6/22/2014 8:00 PM,OstermanBeach201406222000
Ohio Street Beach,06/22/2014 08:00:00 PM,19.5,0.69,1.678,0.135,2,11,
06/22/2014 8:00 PM,OhioStreetBeach201406222000
Rainbow Beach,06/22/2014 08:00:00 PM,20.3,1.04,1.475,0.187,3,11,2,06
/22/2014 8:00 PM,RainbowBeach201406222000
Calumet Beach,06/23/2014 05:00:00 AM,19.5,0.86,1.413,0.083,2,10,7,06
/23/2014 5:00 AM,CalumetBeach201406230500
63rd Street Beach,06/23/2014 05:00:00 AM,16.9,0.76,1.282,0.126,2,11,5,
06/23/2014 5:00 AM,63rdStreetBeach201406230500
Montrose Beach,06/23/2014 05:00:00 AM,17.6,0.42,1.595,0.148,2,10,9,0
6/23/2014 5:00 AM,MontroseBeach201406230500
Osterman Beach,06/23/2014 05:00:00 AM,18.4,0.68,1.662,0.121,2,11,5,0
6/23/2014 5:00 AM,OstermanBeach201406230500
Ohio Street Beach,06/23/2014 05:00:00 AM,18.4,0.68,1.662,0.121,2,11,
06/23/2014 5:00 AM,OhioStreetBeach201406230500
Waiting...
PS D:\002_Documents\001_Private\LEARNING\DATA ENGINEER\Digital Sko
la\04 - Part1 & Video\01 - Folder Materials\Set1 30 - Project 6 (
Real Time Processing Using Kafka)\exercise\kafka-data-streaming\do
cker exec -it my_container python /project/run_consumer.py /proje
ct/credentials.ini ridwan-data-raw
Waiting...
Topic ridwan-data-raw, Event msg = Osterman Beach,06/22/2014 08:00
:00 PM,17.4,1.08,1.225,0.147,2,11,5,06/22/2014 8:00 PM,OstermanBea
ch201406222000
Waiting...
Waiting...
Waiting...
Waiting...
Waiting...
```

7. run ETL script using spark streaming

- source: topic raw-data
- destination: topic clean-data
- run: `sudo docker exec -it my_container spark-submit --packages org.apache.spark:spark-sql-kafka-0-10_2.12:3.0.3 /project/hands-on/clean_data.py`

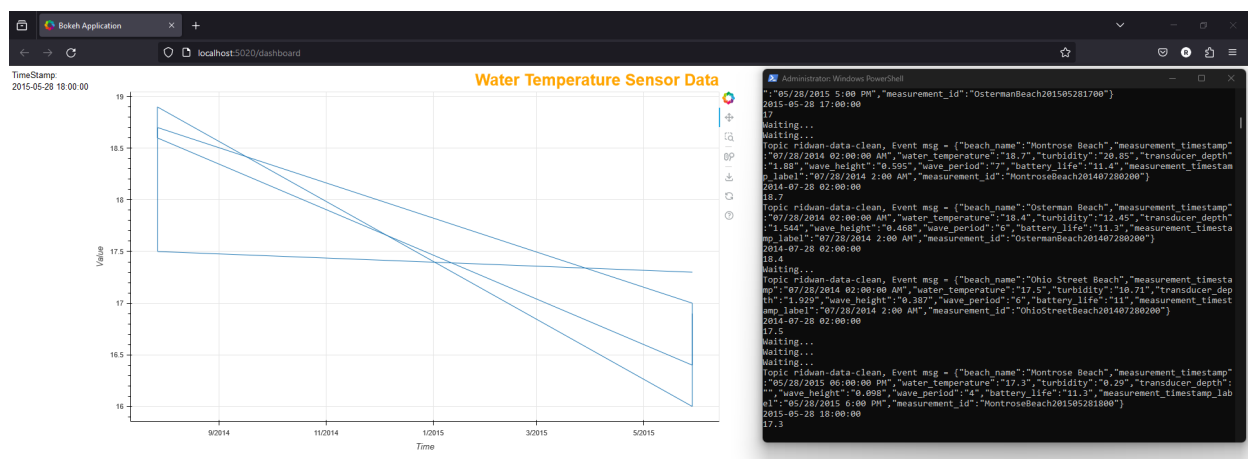
```
ach,05/24/2015 01:00:00 PM,11.6,0.48,0.104,3,
11,2,05/24/2015 1:00 PM,OstermanBeach201505241
300
Topic ridwan-data-raw, Event msg = Calumet Bea
ch,05/24/2015 09:00:00 PM,14.8,1.7,0.138,3,11
.4,05/24/2015 9:00 PM,CalumetBeach201505242100
Waiting...
Topic ridwan-data-raw, Event msg = Calumet Bea
ch,05/28/2015 01:00:00 PM,17.3,0.72,0.121,2,1
1.3,05/28/2015 1:00 PM,CalumetBeach20150528130
0
Waiting...
Topic ridwan-data-raw, Event msg = 63rd Street
Beach,07/21/2014 05:00:00 AM,21.0,0.48,1.837,0.
102,9,11.4,07/21/2014 5:00 AM,63rdStreetBeach2
01407210500
Waiting...
Topic ridwan-data-raw, Event msg = Montrose Be
ach,07/21/2014 05:00:00 AM,16.6,1.6,1.689,0.11
8,10,11.8,07/21/2014 5:00 AM,MontroseBeach2014
07210500
Waiting...
h201505241300
Montrose Beach,05/24/2015 01:00:00 PM,12.0,15.0,
.104,3,11.5,05/24/2015 1:00 PM,MontroseBeach2015
05241300
Osterman Beach,05/24/2015 01:00:00 PM,11.6,0.48,
0.104,3,11.2,05/24/2015 1:00 PM,OstermanBeach20
1505241300
Calumet Beach,05/24/2015 09:00:00 PM,14.8,1.7,0.
.138,3,11.4,05/24/2015 9:00 PM,CalumetBeach20150
524100
Calumet Beach,05/28/2015 01:00:00 PM,17.3,0.72,,
0.121,2,11.3,05/28/2015 1:00 PM,CalumetBeach2015
05281300
Calumet Beach,07/21/2014 05:00:00 AM,21.7,0.71,1.
.462,0.055,10,11.6,07/21/2014 5:00 AM,CalumetBea
ch201407210500
63rd Street Beach,07/21/2014 05:00:00 AM,21.0,0.48
.1.837,0.102,9,11.4,07/21/2014 5:00 AM,63rdStree
tBeach201407210500
Montrose Beach,07/21/2014 05:00:00 AM,16.6,1.6,1.
.689,0.118,10,11.8,07/21/2014 5:00 AM,MontroseBe
ach201407210500
Topic ridwan-data-raw, Event msg = Osterman Beach,05/
24/2015 11:00:00 AM,10.9,0.52,0.128,3,11.2,05/24/201
5 11:00 AM,OstermanBeach201505241100
Waiting...
Topic ridwan-data-raw, Event msg = Montrose Beach,05/
24/2015 12:00:00 PM,11.6,0.18,,0.097,4,11.5,05/24/201
5 12:00 PM,MontroseBeach201505241200
Waiting...
Waiting...
Waiting...
Waiting...
Waiting...
Waiting...
Waiting...
Waiting...
Waiting...
Topic ridwan-data-raw, Event msg = Calumet Beach,07/2
1/2014 05:00:00 AM,21.7,0.71,1.462,0.055,10,11.6,07/2
1/2014 5:00 AM,CalumetBeach201407210500
Waiting...
DD[1111] at start at NativeMethodAccessorImpl.java
:0) (first 15 tasks are for partitions Vector(0))
24/01/20 17:02:51 INFO TaskSchedulerImpl: Adding t
ask set 277.0 with 1 tasks
24/01/20 17:02:51 INFO TaskSchedulerImpl: Starting ta
sk 0.0 in stage 277.0 (TID 281, (ds1b72cc53), excu
tor driver, partition 0, PROCESS_LOCAL, 8572 byte
s)
24/01/20 17:02:51 INFO Executor: Running task 0.0
in stage 277.0 (TID 281)
24/01/20 17:02:51 INFO KafkaConsumer: [Consumer cl
ientId=consumer-spark-kafka-source-780259a8-a0e6-4
811-8416-e26c5e138d4b-903663869-executor-1, group
Id=spark-kafka-source-780259a8-a0e6-4811-8416-e26c
5e138d4b-903663869-executor] Seeking to offset 21
86 for partition ridwan-data-raw-0
24/01/20 17:02:51 INFO SubscriptionState: [Consum
er clientId=consumer-spark-kafka-source-780259a8-a0
e6-4811-8416-e26c5e138d4b-903663869-executor-1, g
roupId=spark-kafka-source-780259a8-a0e6-4811-8416-
e26c5e138d4b-903663869-executor] Seeking to EARLI
EST offset of partition ridwan-data-raw-0
```

- run for check consumer topic data-clean: `sudo docker exec -it my_container python /project/run_consumer.py /project/credentials.ini ridwan-data-clean`

```
Administrator: Windows PowerShell
Topic ridwan-data-clean, Event msg = {"beach_name":"Montrose Beach","measurement_timestamp":"07/22/2014 11:00:00 AM","wa
ter_temperature":"14.7","turbidity":"0.54","transducer_depth":"1.661","wave_height":"0.095","wave_period":"3","battery_lif
e":"11.7","measurement_timestamp_label":"07/22/2014 11:00 AM","measurement_id":"MontroseBeach201407221100"}
Waiting...
Topic ridwan-data-clean, Event msg = {"beach_name":"Osterman Beach","measurement_timestamp":"07/22/2014 11:00:00 AM","wa
ter_temperature":"15","turbidity":"0.74","transducer_depth":"1.344","wave_height":"0.138","wave_period":"3","battery_lif
e":"11.5","measurement_timestamp_label":"07/22/2014 11:00 AM","measurement_id":"OstermanBeach201407221100"}
Topic ridwan-data-clean, Event msg = {"beach_name":"Ohio Street Beach","measurement_timestamp":"07/22/2014 11:00:00 AM",
"water_temperature":"22","turbidity":"1.1","transducer_depth":"1.743","wave_height":"0.151","wave_period":"4","battery_l
ife":"11.3","measurement_timestamp_label":"07/22/2014 11:00 AM","measurement_id":"OhioStreetBeach201407221100"}
Waiting...
Topic ridwan-data-clean, Event msg = {"beach_name":"Rainbow Beach","measurement_timestamp":"07/22/2014 11:00:00 AM","wa
ter_temperature":"17","turbidity":"0.94","transducer_depth":"1.531","wave_height":"0.096","wave_period":"5","battery_lif
e":"11.2","measurement_timestamp_label":"07/22/2014 11:00 AM","measurement_id":"RainbowBeach201407221100"}
Topic ridwan-data-clean, Event msg = {"beach_name":"Calumet Beach","measurement_timestamp":"07/22/2014 12:00:00 PM","wa
ter_temperature":"19.5","turbidity":"1.16","transducer_depth":"1.386","wave_height":"0.088","wave_period":"2","battery_l
ife":"11.5","measurement_timestamp_label":"07/22/2014 12:00 PM","measurement_id":"CalumetBeach201407221200"}
Waiting...
Topic ridwan-data-clean, Event msg = {"beach_name":"63rd Street Beach","measurement_timestamp":"07/22/2014 12:00:00 PM",
"water_temperature":"16.9","turbidity":"0.69","transducer_depth":"1.757","wave_height":"0.08","wave_period":"5","battery
_life":"11.3","measurement_timestamp_label":"07/22/2014 12:00 PM","measurement_id":"63rdStreetBeach201407221200"}
Waiting...
Topic ridwan-data-clean, Event msg = {"beach_name":"Montrose Beach","measurement_timestamp":"07/22/2014 12:00:00 PM","wa
ter_temperature":"15.6","turbidity":"0.65","transducer_depth":"1.601","wave_height":"0.091","wave_period":"3","battery_l
ife":"11.7","measurement_timestamp_label":"07/22/2014 12:00 PM","measurement_id":"MontroseBeach201407221200"}
Waiting...
Topic ridwan-data-clean, Event msg = {"beach_name":"Osterman Beach","measurement_timestamp":"07/22/2014 12:00:00 PM","wa
ter_temperature":"14.8","turbidity":"0.74","transducer_depth":"1.295","wave_height":"0.133","wave_period":"3","battery_l
ife":"11.5","measurement_timestamp_label":"07/22/2014 12:00 PM","measurement_id":"OstermanBeach201407221200"}
Waiting...
```

8. run script for consume data from topic clean-data

- connect to realtime dashboard
- run without docker: `python -m bokeh serve --port 5020 --show hands-on/dashboard.py`



- don't forget to pip install -r requirements.txt in your local device