

```

1  from pymongo import MongoClient
2  import bson
3  import requests
4
5  URL = "https://blockchain.info"
6  LATEST BLOCK = '/latestblock'
7  RAW BLOCK = '/rawblock/'
8
9  client = MongoClient()
10
11  database = client['bitcoin-cluster']
12
13  Entity = database['Entity']
14  Blocks = database['Block']
15  Transaction = database['Transaction']
16  AddressEntity = database['AddressEntity']
17  control = database['control']
18  AddressChange = database['AddressChange']
19  count = 0
20
21
22  def updateLastBlock(lastBlock):
23      try:
24          control.update one({"name": "lastblock"}, {"$set": {"value": lastBlock}})
25          Blocks.insert(lastBlock)
26      except Exception as e:
27          print("Fail on update the last block ", e)
28          raise e
29
30
31  def getTheLastBlock():
32      lastBlockControl = control.find one({
33          "name": "lastblock"
34      })
35
36      if lastBlockControl is not None:
37          lastBlock = lastBlockControl['value']
38      else:
39          control.insert({"name": "lastblock"})
40          lastBlock = requests.get(URL + LATEST BLOCK).json()
41          lastBlock = requests.get(URL + RAW BLOCK + lastBlock['hash']).json()
42          updateLastBlock(lastBlock)
43
44      return lastBlock
45
46
47  def populateTransactionsDatabaseWhenNecessary(lastBlock, MAX POPULATION=1e6):
48      print("doing count transactions")
49      population = Transaction.count documents({})
50      print(population, " transactions!")
51      while True:
52          if population >= MAX POPULATION:
53              return
54          try:
55              print("Calling the last block...")
56              actualBlock = requests.get(URL + RAW BLOCK +
57                  lastBlock['prev block']).json()
58              Transaction.insert many(actualBlock['tx'])
59              population += len(actualBlock['tx'])
60              updateLastBlock(actualBlock)
61          except Exception as e:
62              print("fail on insert transaction ", e)
63              raise e
64
65          lastBlock = actualBlock
66          print("Actual Population: ", population)
67
68  def executeH1Clustering():
69      print("Proccess transactions with h1: ")

```

```

70     for transaction in Transaction.find():
71
72         count transactions()
73
74         addresses = get all address in transaction(transaction)
75
76         entityToMerge = get all entity and remove address already in db(addresses)
77
78         newEntityId = get new entity id(entityToMerge)
79
80         update database addressEntity(addresses, entityToMerge, newEntityId)
81
82 def executeH2Clustering():
83     change wallets = {}
84     for transaction in Transaction.find():
85         count transactions()
86
87         addresses = get all address in transaction(transaction)
88
89         if len(transaction['out']) > 1:
90             first time = 0
91             first address = None
92
93             for output in transaction['out']:
94                 if output.get('addr') is not None and not (output['addr'] in
95                     addresses or output['addr'] in change wallets):
96                     first address = output['addr']
97                     first time += 1
98
99             if first time == 1:
100                 add change wallet(addresses, first address)
101                 change wallets[first address] = True
102
103 def add change wallet(addresses, first address):
104     if len(addresses):
105         entity = AddressEntity.find one({"address": addresses[0]})
106         if entity is not None:
107             first time on db = AddressEntity.find one({"address": first address})
108             if first time on db is None or not len(first time on db):
109                 AddressEntity.insert one({"address": first address, "entity":
110                     entity['entity']})
111
112             if AddressChange.find one({"address": first address}) is None:
113                 AddressChange.insert one({"address": first address, "entity":
114                     entity['entity']})
115
116 def update database addressEntity(addresses, entityToMerge, newEntityId):
117     # if merge, update new entityId
118     if len(entityToMerge) > 1:
119         AddressEntity.update many({"entity": {"$in": entityToMerge}}, {"$set":
120             {"entity": newEntityId}})
121     # insert all addresses without entity
122     newAddressesEntity = [{"address": address, 'entity': newEntityId} for address
123         in addresses]
124     if len(newAddressesEntity):
125         AddressEntity.insert many(newAddressesEntity)
126
127 def count transactions():
128     global count
129     if not count % 10000:
130         print(count)
131         count += 1
132
133 def get all address in transaction(transaction):
134     addresses = []
135     for inputAddress in transaction['inputs']:

```

```
135         address = getAddress(inputAddress)
136         if address is not None:
137             addresses.append(address)
138         addresses = list(set(addresses))
139         return addresses
140
141
142     def get_all_entity_and_remove_address_already_in_db(addresses):
143         addressesEntity = AddressEntity.find({"address": {"$in": addresses}})
144         entityToMerge = []
145         for addressEntity in addressesEntity:
146             addresses.remove(addressEntity['address'])
147             if addressEntity['entity'] not in entityToMerge:
148                 entityToMerge.append(addressEntity['entity'])
149         return entityToMerge
150
151
152     def get_new_entity_id(entityToMerge):
153         if len(entityToMerge) == 1:
154             newEntityId = entityToMerge[0]
155         else:
156             newEntityId = bson.objectid.ObjectId()
157         return newEntityId
158
159
160     def getAddress(inputAddress):
161         address = None
162         if inputAddress.get('prev') is not None:
163             address = inputAddress['prev'].get('addr')
164         if address is None and inputAddress.get('prev out') is not None:
165             address = inputAddress['prev out'].get('addr')
166         return address
167
168
169     def main():
170         global count
171         populateTransactionsDatabaseWhenNecessary(getTheLastBlock())
172         count = 0
173         executeH1Clustering()
174         count = 0
175         executeH2Clustering()
176
177     main()
```