

MEMSQL-KAFKA-VFLOW KURULUM VE AYARLARI

1) Memsql (Singlestore) kurulumu:

Singlestore üyeliği: <https://www.singlestore.com/free-software/>

<https://docs.singlestore.com/v7.3/guides/deploy-memsql/self-managed/linux/step-2/>

<https://docs.singlestore.com/v7.3/guides/deploy-memsql/self-managed/linux/step-3/> (Cluster in a box seçeneğiyle)

Mysql client indirilmeli

Python bağlantısı için connector: <https://docs.singlestore.com/v7.3/guides/use-memsql/connecting-to-memsql/python/>

Kurulumu test etmek için bağlantı denenebilir:

(password yerine belirlenen şifre yazılmalı)

```
mysql --host=127.0.0.1 -P 3306 --user=root --password=testpw --loose-default-auth=mysql_native_password
```

Bağlandıktan sonra "vflow" isimli bir veri tabanı oluşturulabilir. **Farklı bir isim kullanılacaksa** python scriptinde ilgili düzenleme yapılmalı.

Veri tabanı oluşturulduktan sonra "db_format" isimli dosyada belirtilen formatta bir tablo oluşturulmalı.

2) Kafka kurulumu: kafka-kurulum.pdf

Kafka ile dışarıdan veri alabilmek için server.properties dosyasında ilgili değişikliklerin yapılması gerekiyor.

```
nano /usr/local/kafka/config/server.properties
```

***ipadresiniz kısmını düzenleyin:**

```
port = 9092
```

```
advertised.host.name = 127.0.0.1
```

```
advertised.listeners=PLAINTEXT://ipadresiniz:9092
```

```
listeners = PLAINTEXT://0.0.0.0:9092
```

***Yüksek bir veri akışı beklendiğinde kafka logları disk limitini aşabilir.**

***Bu durumun önlenmesi için Log tutma süresinin ihtiyaca göre düzenlenmesi gerekiyor.**

log.retention.hours= logların kaç saat tutulacağı. Çok yer kaplaması istenmiyorsa 1 yapılabilir.

log.retention.check.interval.ms= log tutma limitinin hangi aralıklarla kontrol edileceği.

Log tutma süresi 1 saat olarak ayarlandıysa saatte 2 defa kontrol edilmesi için check.interval.ms=300000 verilebilir

3) vflow kurulumu: <https://github.com/VerizonDigital/vflow>

vflow kurulduktan sonra aşağıdaki komut ile topicleri listeleyin.

```
bin/kafka-topics.sh --list --zookeeper localhost:2181
```

Topic ismi olarak "vflow.ipfix" gözüküyorsa ayrıca oluşturulması gerekiyor.

Topic oluşturma:

```
cd /usr/local/kafka
```

```
bin/kafka-topics.sh --create --zookeeper localhost:2181 --replication-factor 1 --partitions 1 --topic vflow.ipfix
```

4) Kurulumlar tamamlandığında aşağıdaki komut ile kafkada verilerin gözükmesi gerekiyor.

```
/usr/local/kafka/bin/kafka-console-consumer.sh --bootstrap-server localhost:9092 --topic vflow.ipfix
```

5) read.py içerisinde veri tabanı bilgileri düzenlenmeli. (şifre, db adı gibi bilgiler)

read.py içerisinde düzenleme yapıldıktan sonra aşağıdaki bağlantılarda verilen dosyalar indirilmeli ve read.py scriptinin bulunduğu dizine atılmalı

```
wget https://git.io/GeoLite2-ASN.mmdb
```

```
wget https://git.io/GeoLite2-City.mmdb
```

Sonrasında verileri veri tabanına gerçek zamanlı aktaracak scripti çalıştırabiliriz.

*start-vflow.sh içerisinde verilen read.py scriptinin dosya dizini düzenlenebilir.

```
./start-vflow.sh
```

6) Veri tabanındaki eski kayıtların silinmesi isteniyorsa ilgili cron job tanımlanabilir.

clear_old_data.sh dosyasında ilgili kayıtların silinmesi için gereken memsql komutu bulunmaktadır.

Db bağlantısı için gerekli **password**, **database** ve silinmesi istenilen **saat aralığı** düzenlenebilir.

Cron job eklemek için ilgili komut:

```
crontab -e
```

İçerik (clear_old_data.sh dosyasının dizini düzenlenebilir):

```
0 * * * * /root/clear_old_data.sh > /root/cronoutput.log 2>&1
```