

# YARN Çözüm

İlk olarak servisleri çalıştıralım.

```
start-all.sh
```

Adım 1: Veri setini indiriniz.

```
wget https://github.com/erkansirin78/datasets/raw/master/turkish_cities_wc.zip
```

```
unzip turkish_cities_wc.zip
```

Adım 2: HDFS üzerinde turkish\_cities\_input adında bir klasör oluşturunuz.

```
hdfs dfs -mkdir /user/train/datasets/turkish_cities_input
```

Adım 3: İndirdiğiniz turkish\_cities\_wc.txt verisini HDFS üzerinde bu klasöre yükleyiniz.

```
hdfs dfs -put turkish_cities_wc.txt /user/train/datasets/turkish_cities_input
```

Adım 4: YARN kullanarak mapreduce.wordcount.jar programını çalıştırınız.

```
yarn jar mapreduce.wordcount.jar com.veribilimiokulu.mapreduce.MR2WordCount /user/train/datasets/turkish_cities_input /user/train/data
```

Adım 5: Resource Manager ara yüzünden görev durumlarını gözlemleyiniz.

Adım 6: HDFS üzerinde oluşturduğunuz output klasörün içerisindeki dosyayı head komutu ile okuyarak, şehir isimlerinin kaç kez tekrarlandığını gözlemleyiniz.

```
hdfs dfs -cat /user/train/datasets/turkish_cities_output/part-r-00000
```