

Apache YARN ve MapReduce ile Word Count Uygulaması

Apache YARN ve MapReduce ile Word Count Uygulaması

Görev 1: **turkish_cities_wc.txt** dosyasındaki şehir isimlerinin kaç kez tekrarlandığını MapReduce frameworkü ile çözünüz.

Not: Bunun için **mapreduce.wordcount.jar** uygulama dosyasını kullanabilirsiniz.

Adım 1: Veri setini sistem üzerinden indiriniz.

Adım 2: HDFS üzerinde **turkish_cities_input** adında bir klasör oluşturunuz.

Adım 3: İndirdiğiniz **turkish_cities_wc.txt** verisini HDFS üzerinde bu klasöre yükleyiniz.

Not: mapreduce.wordcount.jar dosyası da aynı çalışma klasöründe lokalde bulunmalıdır.

Apache YARN ve MapReduce ile Word Count Uygulaması

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

Adım 5: Resource Manager ara yüzünden görev durumlarını gözlemleyiniz.

Not: Adım 4 için önceki görevde oluşturulan **input** klasörü de komutta belirtilmelidir.

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

miuul

miuul.com