

Teniendo en cuenta los resultados del ejercicio anterior. Y que sigo operando en el IDE de Eclipse.

Para crear una nueva cola que disponga del 60 % de los recursos en Hadoop, tengo que modificar la configuración de YARN y establecer una nueva cola en el archivo `capacity-scheduler.xml`. Se han seguido los siguientes pasos:

1. **Configurar YARN para soportar colas:** Modifica el archivo `capacity-scheduler.xml` para definir una nueva cola con el 60 % de los recursos.
2. **Modificar el archivo de configuración `capacity-scheduler.xml`:** para agregar la definición de la nueva cola.
3. **Ejecutar el trabajo MapReduce del ejercicio anterior en la nueva cola.**

Paso 1: Configurar YARN para soportar colas

Primero, el archivo `capacity-scheduler.xml` está configurado correctamente en el clúster Hadoop.

Paso 2: Modificar el archivo `capacity-scheduler.xml`

Voy a añadir la definición de la nueva cola en el archivo `capacity-scheduler.xml` para asignarle el 60 % de los recursos.

```
<configuration>

<property>

<name>yarn.scheduler.capacity.root.queues</name>

<value>default,newqueue</value>

</property>

<property>

<name>yarn.scheduler.capacity.root.default.capacity</name>

<value>40</value>

</property>

<property>

<name>yarn.scheduler.capacity.root.newqueue.capacity</name>
```

```

<value>60</value>

</property>

<property>

<name>yarn.scheduler.capacity.root.newqueue.maximum-capacity</name>

<value>100</value>

</property>

<property>

<name>yarn.scheduler.capacity.root.newqueue.state</name>

<value>RUNNING</value> </property> <property>
<name>yarn.scheduler.capacity.root.newqueue.acl_submit_applications</name> <v
alue>*</value>

</property>

<property>

<name>yarn.scheduler.capacity.root.newqueue.acl_administer_queue</name>

<value>*</value>

</property>

</configuration>

```

Este archivo me ha definido dos colas: `default` , que tiene el 40 % de los recursos, y `newqueue` , que tiene el 60 % de los recursos.

Paso 3: Reiniciar

Después de realizar estos cambios, es necesario reiniciar el ResourceManager de YARN para que los cambios tengan efecto.

```
$ yarn resourcemanager --refreshQueues
```

Paso 4: Ejecutar el trabajo antiguo MapReduce en la nueva cola

```

import org.apache.hadoop.conf.Configuration;
import org.apache.hadoop.fs.Path;

```

```

import org.apache.hadoop.io.DoubleWritable;

import org.apache.hadoop.io.Text;

import org.apache.hadoop.mapreduce.Job;

import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;

import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;

public class SalesJob {

    public static void main(String[] args) throws Exception {

        Configuration conf = new Configuration();

        conf.set("mapreduce.job.queue.name", "newqueue"); Job job = Job.getInstance(conf, "Sales Analysis"); job.setJarByClass(SalesJob.class);
        job.setMapperClass(SalesMapper.class);
        job.setReducerClass(SalesReducer.class); job.setOutputKeyClass(Text.class);
        job.setOutputValueClass(DoubleWritable.class);

        FileInputFormat.addInputPath(job, new Path(args[0]));
        FileOutputFormat.setOutputPath(job, new Path(args[1]));

        System.exit(job.waitForCompletion(true) ? 0 : 1);

    }

}

```

Este código especifica que la cola `newqueue` en la configuración del trabajo MapReduce, cuando se inicie la ejecución del job se van a utilizar los recursos de la cola `newqueue`, que tiene el 60 % de los recursos disponibles.