

## TEZ Installation on YARN in Single Node Hadoop Cluster

### Step 1: Download TEZ tar file from tez.apache.org

```
cd ~
```

```
wget https://dlcdn.apache.org/tez/0.7.1/apache-tez-0.7.1-bin.tar.gz --no-check-certificate
```

### Step 2: Untar and configure libraries in hadoop

```
tar xvzf apache-tez-0.7.1-bin.tar.gz
```

```
sudo mv apache-tez-0.7.1-bin /usr/local/tez
```

```
cd /usr/local/tez
```

```
hadoop fs -mkdir -p /apps/tez
```

```
hadoop fs -put tez-0.7.1.tar.gz /apps/tez/
```

### Step 3: Create tez-site.xml file and update the configured path

```
mkdir /usr/local/tez/conf
```

```
vi /usr/local/tez/conf/tez-site.xml
```

```
<configuration>  
  <property>  
    <name>tez.lib.uris</name>  
    <value>${fs.default.name}/apps/tez/tez-0.7.1.tar.gz</value>  
  </property>  
</configuration>
```

## Step 4: Disable virtual memory checks in yarn-site.xml

**vi /usr/local/hadoop/etc/hadoop/yarn-site.xml**

```
<property>
  <name>yarn.nodemanager.vmem-check-enabled</name>
  <value>>false</value>
  <description>Whether virtual memory limits will be enforced for
containers</description>
</property>
```

Restart the yarn services:-

```
stop-yarn.sh
start-yarn.sh
```

## Step 4: Add environment variables into .bashrc file

**vi ~/.bashrc**

```
export TEZ_HOME=/usr/local/tez
export TEZ_CONF_DIR=${TEZ_HOME}/conf
export HADOOP_CLASSPATH="${HADOOP_CLASSPATH}:${TEZ_CONF_DIR}:$(find
${TEZ_HOME} -name "*.jar" | paste -sd ":")"
```

**source ~/.bashrc**

## Step 5: Installation get completed. Get into hive and test

hive

```
hive> set hive.execution.engine;
```

```
hive> select state,city, profession, count(*),max(amount),sum(amount) from  
txnrecords t inner join tblcustomer c on t.custno = c.custno group by  
state,city,profession order by city;
```

```
hive> select a.custno,a.salesamount,b.profession from (select  
custno,sum(amount) salesamount from txnrecords where amount > 10 group by  
custno) a inner join (select custno,profession from tblcustomer where age > 50)  
b on a.custno = b.custno;
```

```
hive> set hive.execution.engine=tez;
```

```
hive> select state,city, profession, count(*),max(amount),sum(amount) from  
txnrecords t inner join tblcustomer c on t.custno = c.custno group by  
state,city,profession order by city;
```

```
hive> select a.custno,a.salesamount,b.profession from (select  
custno,sum(amount) salesamount from txnrecords where amount > 10 group by  
custno) a inner join (select custno,profession from tblcustomer where age > 50)  
b on a.custno = b.custno
```

**!!! THANK YOU !!!**