CCDC Inject

INJECT NAME	Install Mahout Machine Learning Environment
INJECT ID	SOFT10T

INJECT DESCRIPTION:
Using the following instructions, install the Mahout Machine Learning environment on the Debian platform.

Step 1: Install Java

Mahout requires Java to run. Ensure you have Java installed by running:

```
java -version
```
If Java is not installed, install it using:

- **Ubuntu/Debian**: `sudo apt update && sudo apt install openjdk-11-jdk`

Step 2: Install Apache Hadoop (Optional)

Mahout is optimized for working with Hadoop. If you plan to use Mahout with Hadoop, install it by following:

```
sudo apt install hadoop
```
Ensure Hadoop is correctly set up by running:

```
hadoop version
```

Step 3: Download Apache Mahout

Download the latest Mahout release from the official website:

```
wget https://downloads.apache.org/mahout/0.14.0/apache-mahout-
distribution-0.14.0.tar.gz
```
Extract the archive:

```
tar -xvzf apache-mahout-distribution-0.14.0.tar.gz
cd apache-mahout-distribution-0.14.0
```

Step 4: Set Environment Variables

Add Mahout to your system's environment variables. Open the `.bashrc` or `.zshrc` file:

```
nano ~/.bashrc
```
Add the following lines:

```
export MAHOUT_HOME=~/apache-mahout-distribution-0.14.0
export PATH=$MAHOUT_HOME/bin:$PATH
```
Apply the changes:

```
source ~/.bashrc
```

Step 5: Verify Installation

To check if Mahout is installed correctly, run:

```
mahout
```
You should see the Mahout help menu displaying available commands.

Step 6: Running a Sample Job

To test Mahout, try running a k-means clustering example:

```
mahout org.apache.mahout.clustering.kmeans.KMeansDriver
```
If Hadoop is installed and configured, you can run Mahout jobs in a distributed manner:

```
mahout kmeans -i input-data -c clusters -o output -dm
org.apache.mahout.common.distance.EuclideanDistanceMeasure -x 10 -k 5
```

Conclusion

You have successfully installed Apache Mahout. You can now use it for machine learning tasks such as clustering, classification, and recommendation systems.

INJECT DELIVERABLE

Respond a business memo that documents, with screenshots, all of the verification steps.

- Verify Java
- Verify Hadoop
- Verify Mahout
- Verify test job, k-means clustering