HDFS INTEGRATION WITH APACHE ATLAS

Step 1: Copy the HDFS Bridge to Hadoop Lib Directory

```
cp /home/u1/hdfs_hook/atlas-hdfs-bridge-2.1.0.7.1.8.0-801.jar /opt/hadoop-3.3.2/lib/

chown hadoopZetta:hadoopZetta /opt/hadoop-3.3.2/lib/atlas-hdfs-bridge-2.1.0.7.1.8.0-801.jar

chmod 0755 /opt/hadoop-3.3.2/lib/atlas-hdfs-bridge-2.1.0.7.1.8.0-801.jar
```

Step 2: Copy the HDFS Bridge to Atlas Lib Directory

```
cp /home/u1/hdfs_hook/atlas-hdfs-bridge-2.1.0.7.1.8.0-801.jar /opt/apache-atlas-2.4.0/server/webapp/atlas/WEB-INF/lib/

chown hadoopZetta:hadoopZetta /opt/apache-atlas-2.4.0/server/webapp/atlas/WEB-INF/lib/atlas-hdfs-bridge-2.1.0.7.1.8.0-801.jar

chmod 0755 /opt/apache-atlas-2.4.0/server/webapp/atlas/WEB-INF/lib/atlas-hdfs-bridge-2.1.0.7.1.8.0-801.jar
```

Step 3: Copy the HDFS Extractors to Atlas Directory

```
cp -r /home/u1/hdfs_hook/extractors /opt/apache-atlas-2.4.0/

chown -R hadoopZetta:hadoopZetta /opt/apache-atlas-2.4.0/extractors

chmod -R 0755 /opt/apache-atlas-2.4.0/extractors
```

Step 4: Copy the HDFS Model to Atlas Models Directory

```
mkdir -p /opt/apache-atlas-2.4.0/models/1000-Hadoop

cp /home/u1/hdfs_hook/1120-hdfs_model_v2.json /opt/apache-atlas-2.4.0/models/1000-Hadoop/

chown hadoopZetta:hadoopZetta /opt/apache-atlas-2.4.0/models/1000-Hadoop/1120-hdfs_model_v2.json
```

```
chmod 0755 /opt/apache-atlas-2.4.0/models/1000-Hadoop/1120-
hdfs_model_v2.json
```

Step 5: Ensure Curl is Installed

```
apt-get update

apt-get install curl -y
```

Step 6: Register the HDFS Model in Atlas

```
curl -u admin:admin -X POST -H "Content-Type: application/json" -d @/opt/apache-
atlas-2.4.0/models/1000-Hadoop/1120-hdfs_model_v2.json
http://192.168.22.10:21000/api/atlas/v2/types/typedefs
```

Step 7: Set Atlas Environment Variables in .bashrc for hadoopZetta

```
echo -e "\n# Atlas Environment Variables" >> /home/u1/.bashrc

echo "export ATLAS_CONF_DIR=/opt/apache-atlas-2.4.0/conf" >> /home/u1/.bashrc

echo "export ATLAS_LOG_DIR=/opt/apache-atlas-2.4.0/logs" >> /home/u1/.bashrc

echo "export BASEDIR=/opt/apache-atlas-2.4.0/extractors" >> /home/u1/.bashrc

chown hadoopZetta:hadoopZetta /home/u1/.bashrc

chmod 0644 /home/u1/.bashrc
```

Step 8: Add HDFS Wrapper Function to .bashrc for hadoopZetta

```
cat << 'EOF' >> /home/u1/.bashrc
function hdfs() {
 echo "Function hdfs called with args: $@"

if [ "$1" = "dfs" ]; then

 echo "Running wrapper with: ${@:2}"

 /opt/hadoop-3.3.2/bin/hdfs-wrapper.sh "${@:2}"

else echo "Running original hdfs with: $@"

/opt/hadoop-3.3.2/bin/hdfs "$@"
```

```
fi }
export -f hdfs EOF
```

Step 9: Copy the HDFS Wrapper Script to Hadoop Bin Directory

```
cp /home/u1/hdfs_hook/hdfs-wrapper.sh /opt/hadoop-3.3.2/bin/

chown hadoopZetta:hadoopZetta /opt/hadoop-3.3.2/bin/hdfs-wrapper.sh

chmod 0755 /opt/hadoop-3.3.2/bin/hdfs-wrapper.sh
```

Step 10: Verify

```
su - hadoopZetta -c "source /home/u1/.bashrc && hdfs dfs -ls /"  # Test the wrapper
as hadoopZetta

cat /opt/apache-atlas-2.4.0/logs/application.log  # Check Atlas logs for errors
```