Linux Data Warehouse Documentation 02

In Connections.py file, you'll see references to these drivers:

- mysqlsparkjar: This points to the JDBC driver for MySQL.

- oraclesparkjar: This points to the JDBC driver for Oracle

in the linux machine they are at /opt/spark/jars/

Update Spark Configuration (if needed):

- If you placed the JAR files outside the Spark `jars` directory, you might need to tell Spark where to find them. You can do this by:
    - **Spark Submit:** When submitting your Spark application, use the `--jars` option to include the paths to your JDBC driver JARs.
    - **`spark-defaults.conf`:** Add the following line to your `spark-defaults.conf` file (usually located at `/opt/spark/conf/spark-defaults.conf`):
    - `spark.driver.extraClassPath /path/to/mysql-connector-java-8.0.30.jar:/path/to/ojdbc8.jar`
    - `spark.executor.extraClassPath /path/to/mysql-connector-java-8.0.30.jar:/path/to/ojdbc8.jar`

        (Replace `/path/to/...` with the actual paths to your JAR files.)

Finding the Current Driver Locations

To find where the mysqlsparkjar and oraclesparkjar are currently located on your existing Linux server, you can use the following command in your terminal:

Bash

find / -name "mysql-connector-java*.jar" -o -name "ojdbc*.jar" 2>/dev/null

This command will search your entire file system (/) for files named mysql-connector-java*.jar or ojdbc*.jar and print their locations. The 2>/dev/null part suppresses any error messages that might occur during the search.