This Project aims to create the Maven Project by configuring the dependency in pom.xml file and applying data ingestion and data transformation

* PySpark
* Scala
* Postgress DataBase or MySQL db
* Any IDE such as IntelliJ

**Setup Process**

Step 1: Install Java JDK (Java development kit)- version 8

Step 2: Set JAVA\_HOME environment variable.

Step 3: Download Hadoop's winutils.exe.

Step 4: Install IntelliJ IDE

Step 5: Install Scala Plugin

Step 6: Install PostgreSQL Database on Local machine

Step 7: Install PGAdmin a web-based GUI tool

**Creating a SparkSession**

Created Spark Session

val spark = SparkSession

.builder()

.appName("HelloSpark")

.config("spark.master", "local")

.enableHiveSupport()

.getOrCreate()

println("Created Spark Session")

**Interacting with RDBMS Tables**

We can export the .csv files to RDBMS using “sqoop-export” or

Can directly create some records using SQL query language

Now use modular approach to access MySQL or PostgreSQL db.

Created maven project using mnv package.

Created seperate .scala files for relevant functions

def getPostgresServerDatabase(): String = {

logger.info("getPostgresServerDatabase() started")

val pgURL = "jdbc:postgresql://localhost:5432/futurex"

pgURL

}

def getPostgresCommonProps(): Properties = {

logger.info("getPostgresCommonProps() started")

val pgConnectionProperties = new Properties()

pgConnectionProperties.put("user", "postgres")

pgConnectionProperties.put("password", "\*\*\*\*\*\*\*")

pgConnectionProperties

}

**Configured Logging Support**

Logging is an essential component of any good software development process. Through logging, you can detect and fix issues, understand the behavior of users, and even obtain valuable business insights.

configured log4j logging support in our project.

Added all required dependencies to pom.xml file by using the URL :- https://mvnrepository.com/

Log4j is a simple and flexible logging framework. We used below given dependencies for log4j

<dependency>

<groupId>log4j</groupId>

<artifactId>log4j</artifactId>

<version>1.2.17</version>

</dependency>

**Key Features**

* Designed Effective and Robust Maven Project.
* Implemented Logger Support.
* Configuring Dependencies in pom.xml file.
* Project is in accordance with Error Handling, File auditing ( checking for data quality and consistency).
* Used Modular programming approach with multiple files in a Package.
* Used JSON File to hold the inputs (Parameterize Fields).