**Big Data Technology  
Final Project**

**Weather Data Streaming, Analysis and visualization**

**Course: Big Data Technology (CS-523)**

**Professor: Mrudula Mukadam**

**Group Members**

* 1. **Thanh Do Nguyen**
  2. **Van Ngoc Trau Nguyen**
  3. **Mihret Ghebremichael Kidane**
  4. **Melaku Gossaye Denbel**

1. **Setup environment:**

* **Kafka:**
* Download Kafka: <https://archive.apache.org/dist/kafka/2.4.1/kafka_2.12-2.4.1.tgz>
* Extract and change directory to the extracted file:

cd /home/cloudera/Downloads/kafka\_2.12-2.4.1

* **Check service status**

service --status-all

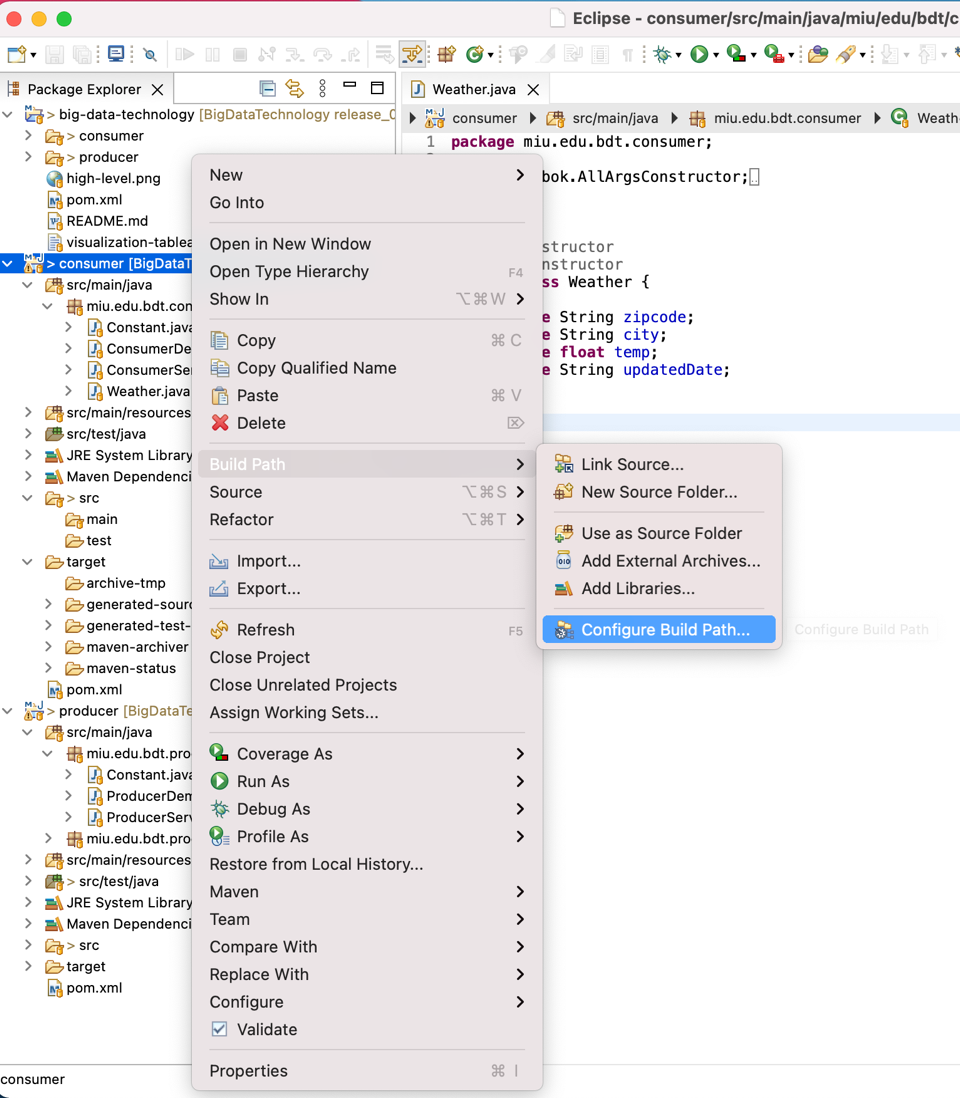
* **Start Zookeeper if it’s not running:**

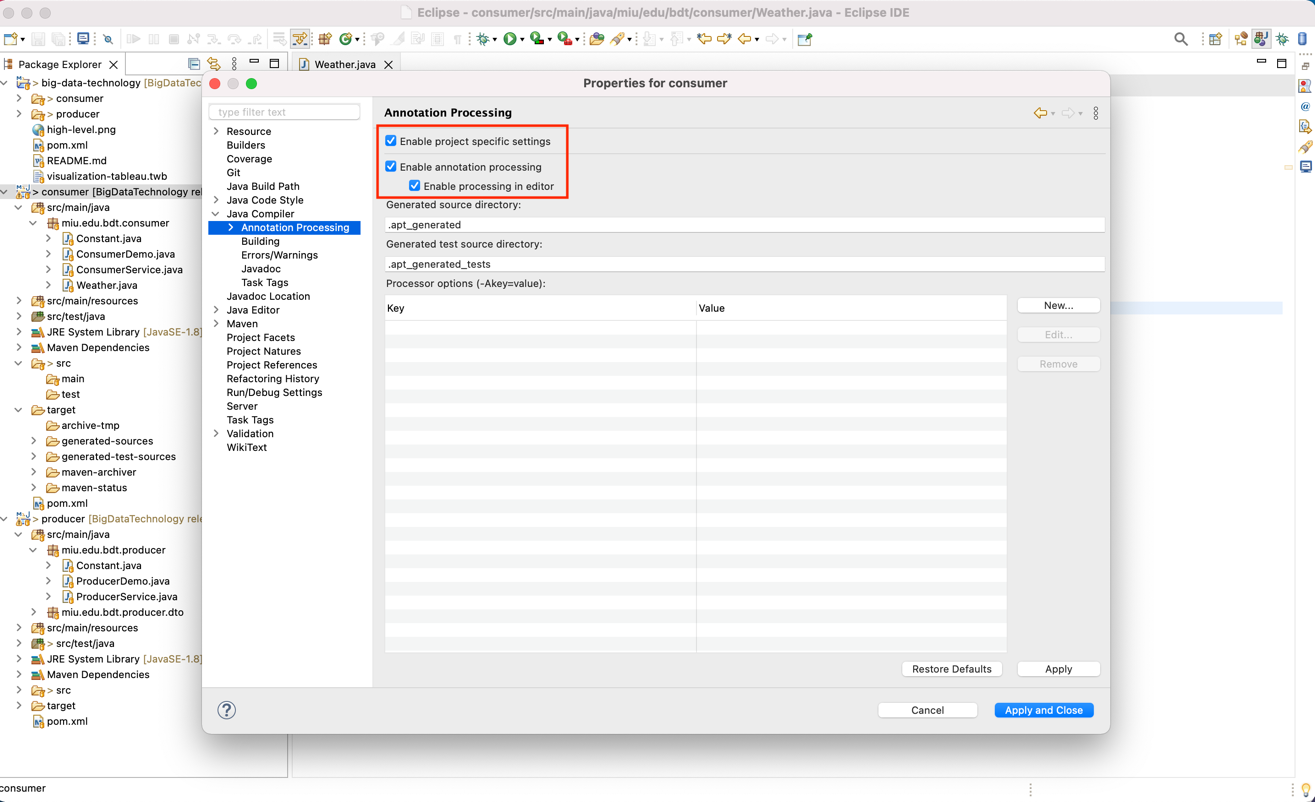
bin/zookeeper-server-start.sh config/zookeeper.properties

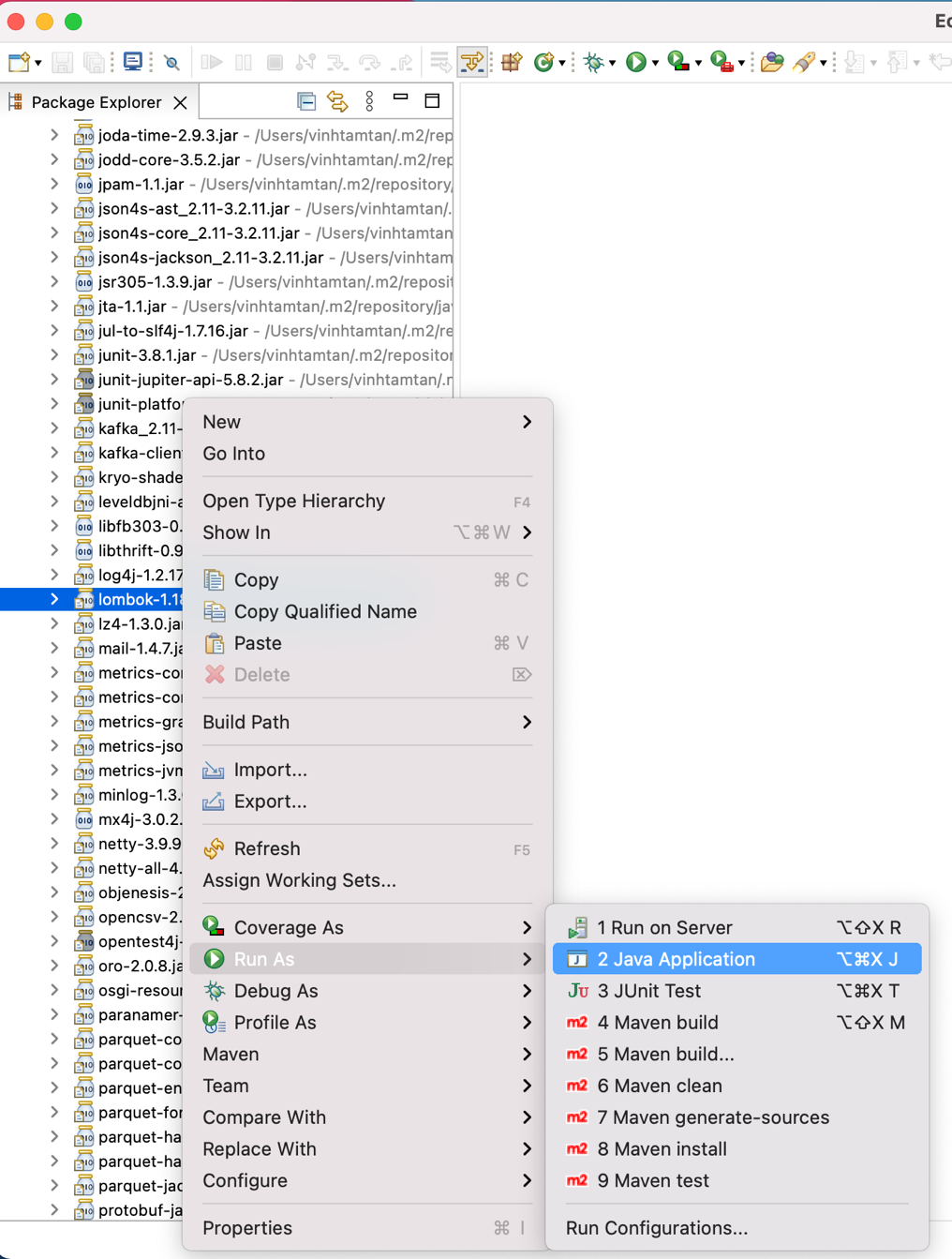
* **Start Kafka service:**

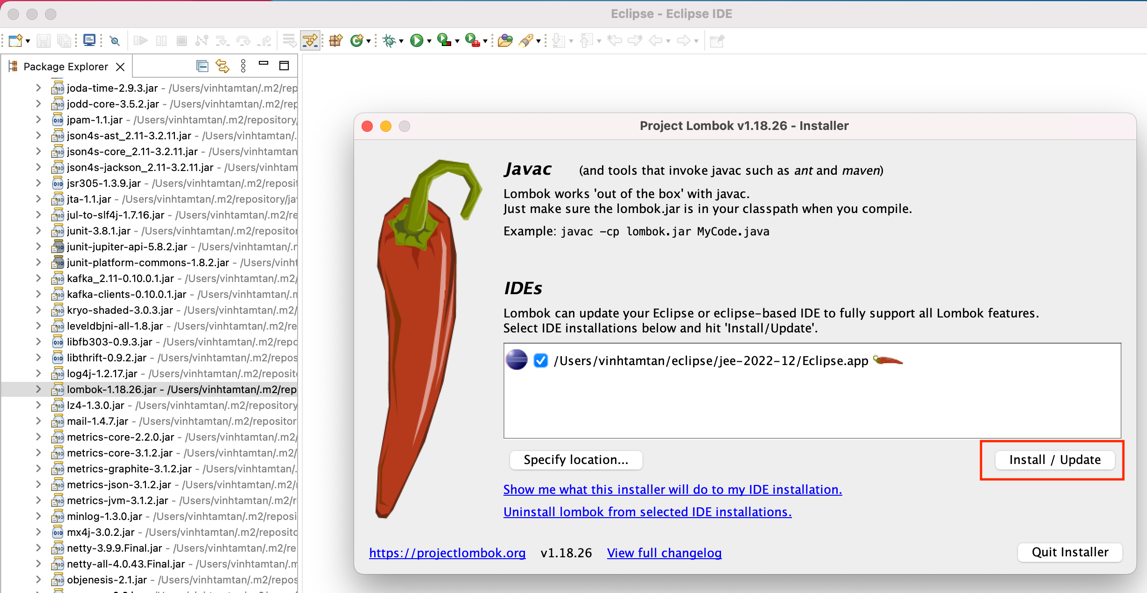
bin/kafka-server-start.sh -daemon config/server.properties

1. **Some IDEs have problems with “lombok-1.18.26.jar”:** make sure you setup this library correctly



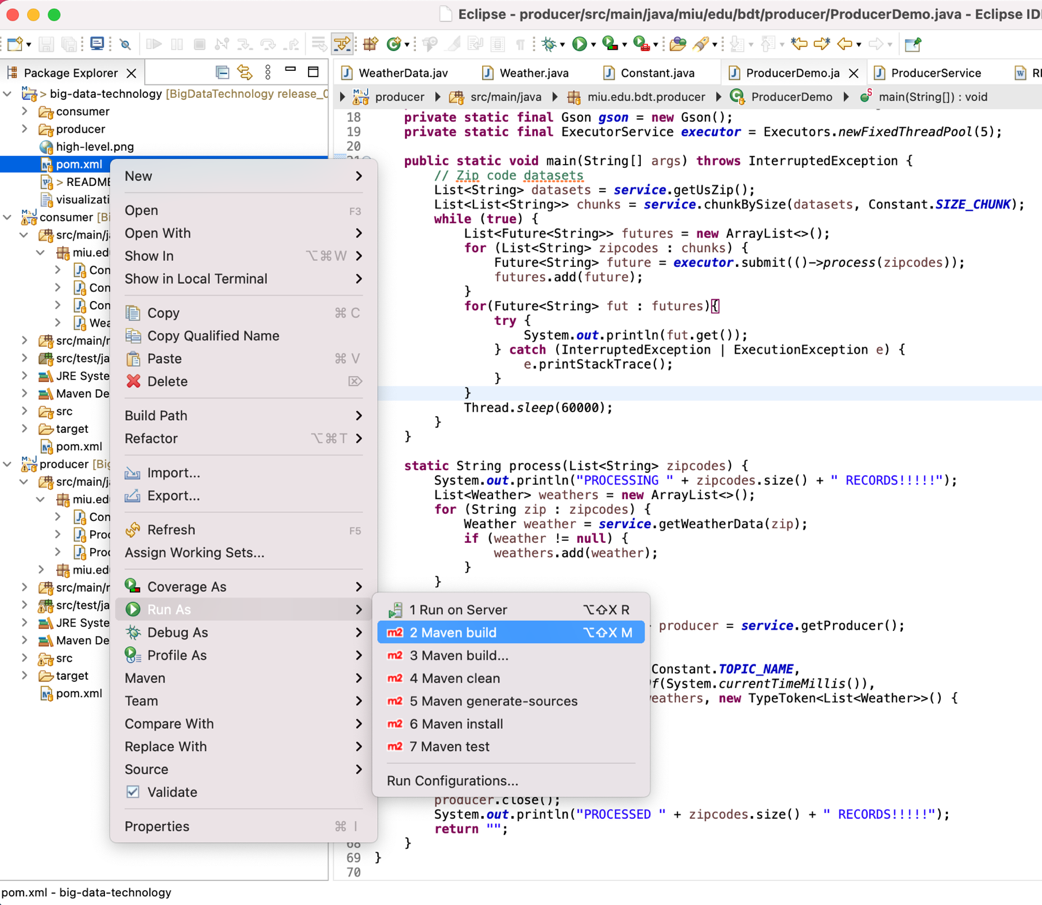






1. **Build projects:** use command-line OR run directly on Eclipse

* cd ../project
* mvn clean package

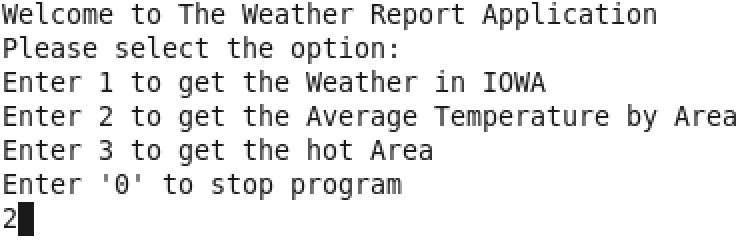


1. **Run producer and consumer programs with jar files:** always run jar file with dependencies

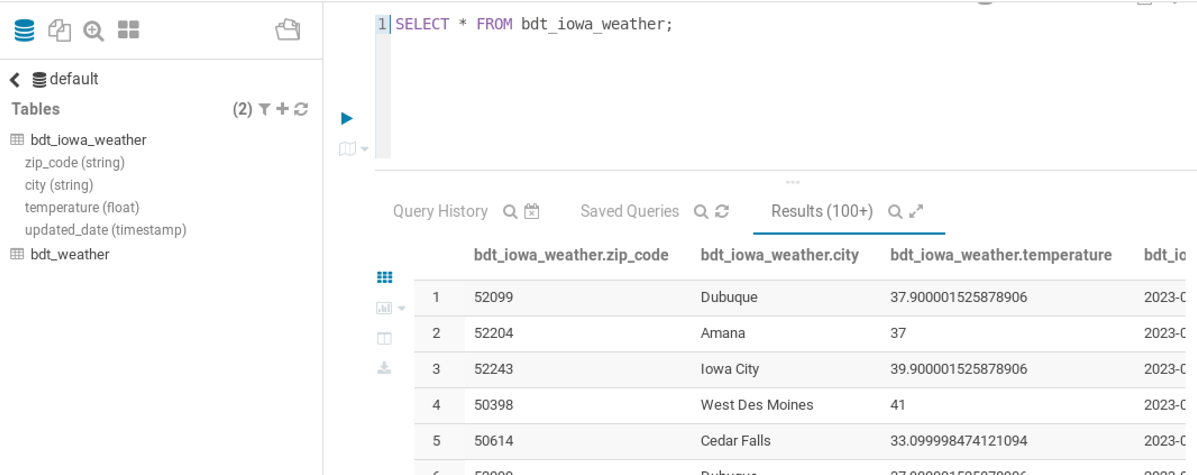
* java -jar 'consumer/target/consumer-1.0-SNAPSHOT-jar-with-dependencies.jar'
* java -jar 'producer/target/producer-1.0-SNAPSHOT-jar-with-dependencies.jar'

1. **Run visualizer with jar file:** run jar file (without dependencies)

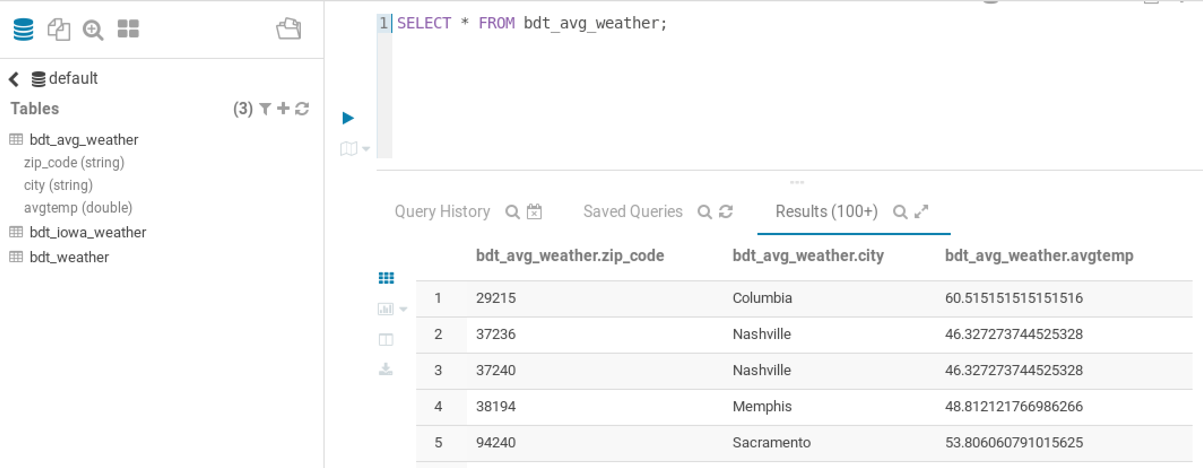
* java -jar 'visualizer/visualizer-1.0-SNAPSHOT.jar'
* Select option to view: each option will execute different SQL statements.



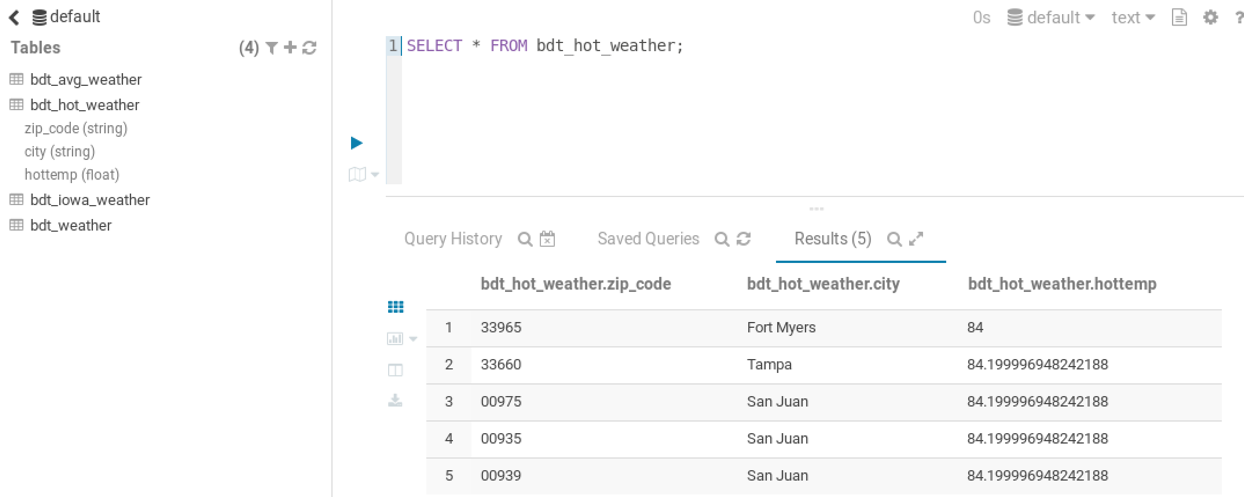
Option 1: get the Weather in IOWA only



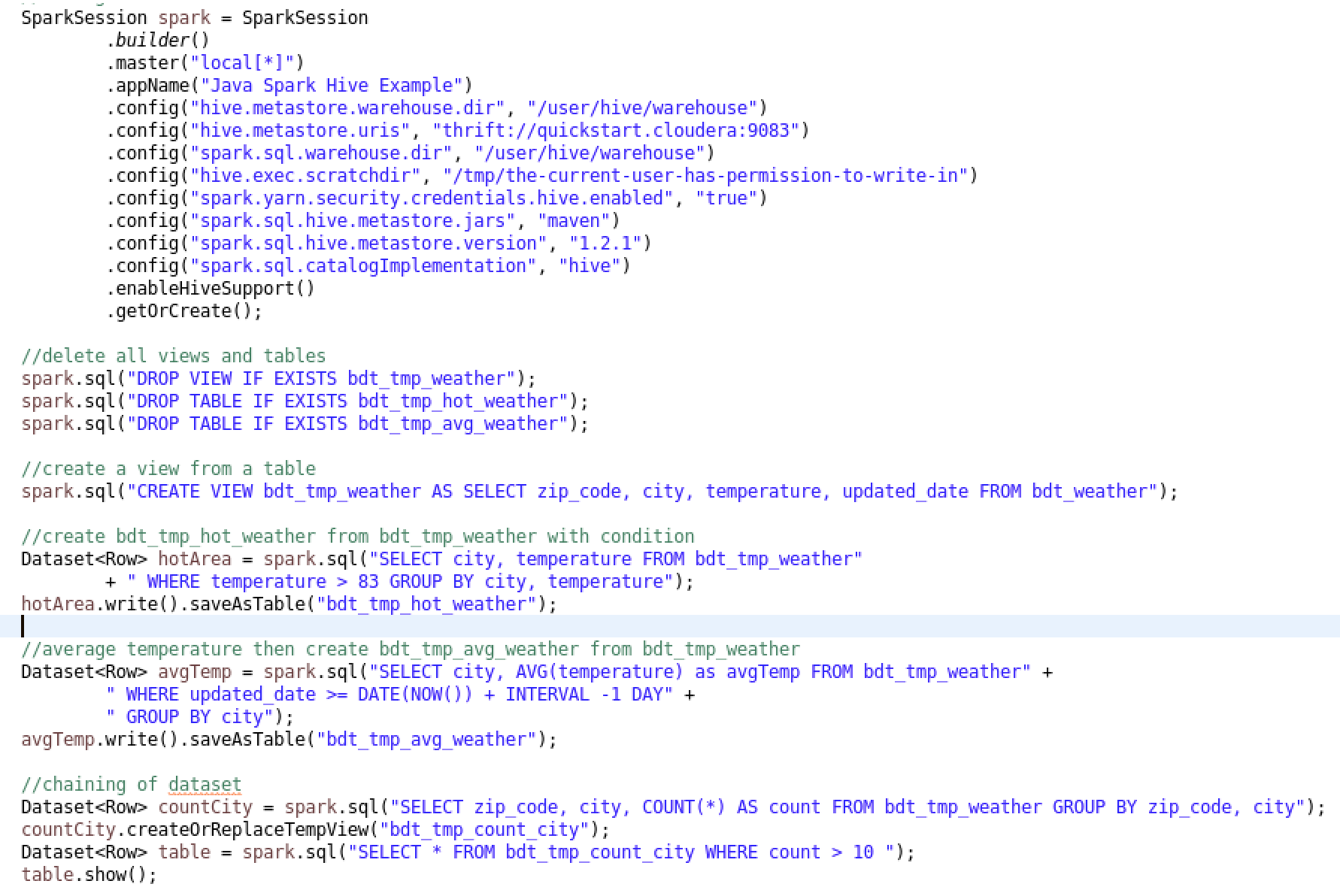
Option 2: get the last 7 days average temperature grouping by zipcode

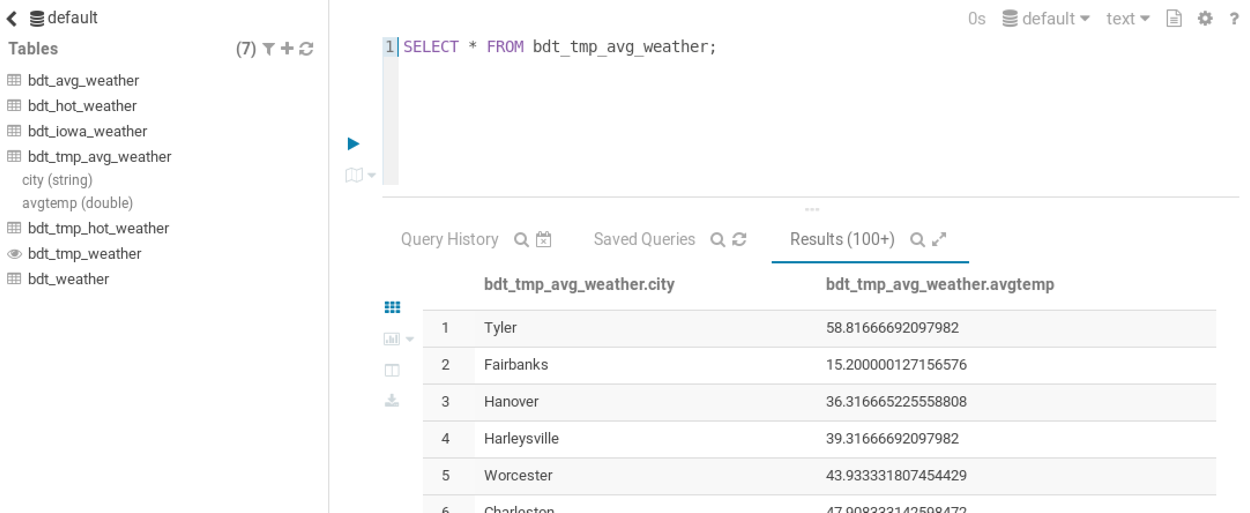


Option 3: get the area with temperature greater than 83 F degrees



Some more advanced configuration and more Spark SQL integration with HIVE

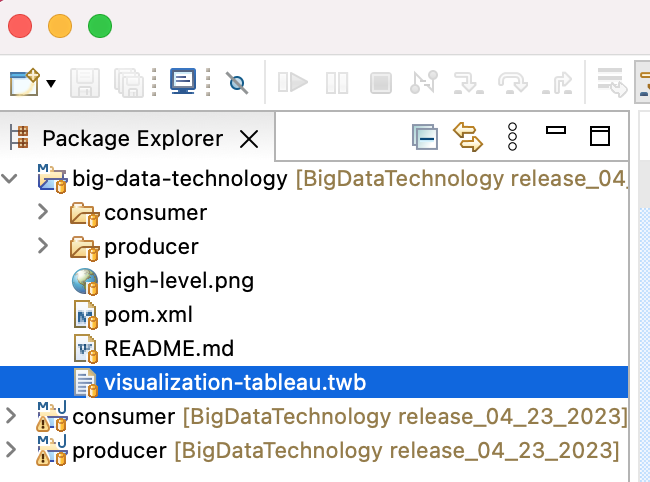




Make sure you include these codes in pom.xml. It’s very important. I took me 2 hours to find the solution by including these codes into the project.



1. **Visualization with Tableu:** please open “visualization-tableau.twb” from resource folder to see the result.



Make sure you installed “Tableau-desktop” and “ClouderaHiveODBC” before opening the file.

+ <https://www.cloudera.com/downloads/connectors/hive/odbc/2-6-1.html>

+ <https://www.tableau.com/products/desktop/download>

1. **Analyze projects:** please refer the “FinalProject\_presentation\_v2.pptx”