COP5725 - Database Management Systems

Project Deliverable 3: Weather Recording System

Group 19: He,Jiahui,
Shi,Qinxuan,
Wang,Shihuan,
Zhang,Guanglong

Contents

1 Quality of the transformation of the ER diagram into a collection of relation schemas. 3
1.1 Improved ER Diagram 3
1.2 Relational Schema 3

2 Quality of the transformation of the collection of relation schemas into a collection of SQL table schemas. 6

1 Quality of the transformation of the ER diagram into a collection of relation schemas.

1.1 Improved ER Diagram

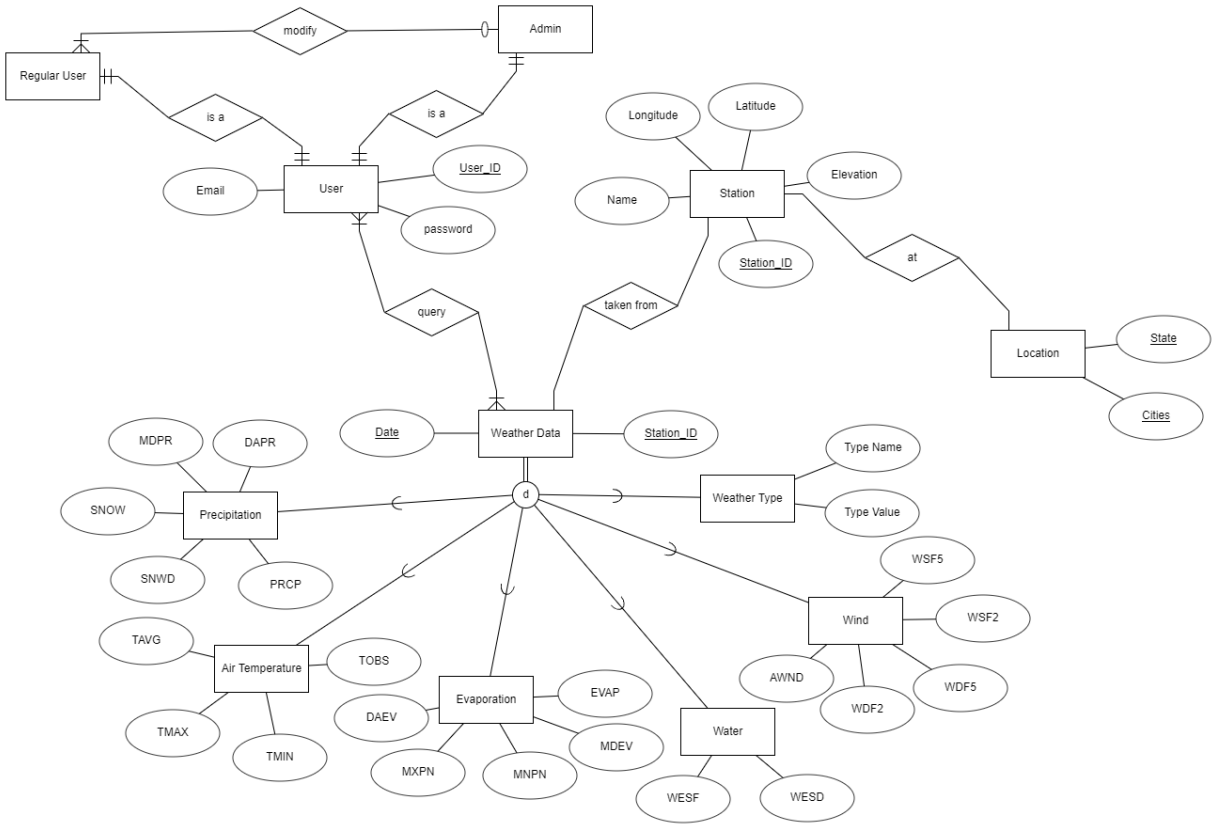


Figure 1: ER Diagram

1.2 Relational Schema

User(User_ID: Integer, Email: String, password: String)

We know from the ER diagram that the 'User' entity acts as a superclass, and it has three attributes. The first attribute is the unique attribute 'User_ID', the domain of this attribute is Integer. The second attribute is 'password', the domain of this attribute is String. The third attribute is 'Email', the domain of this attribute is String.

Regular_User(User_ID: Integer)

Admin(User_ID: Integer)

The subclasses of the superclass 'User' have two entities, they are disjoint from each other. One is the 'Regular_User' entity, with the unique attribute 'User_ID', the domain of this attribute is String. The other is the 'Admin' entity, with the unique attribute 'User_ID', the domain of this attribute is String.

Station(Station_ID: String, Name: String, Latitude: Float, Longitude: Float, Elevation: Float)

All the weather data is taken from the physical stations. The relation of 'Station' has attributes of (Station_ID: String, Name: String, Latitude: Float, Longitude: Float, Elevation: Float). The primary key is Station_ID. The data type of Station_ID is strings that contain characters and numbers with lengths no longer than 10.

Name_ contains characters only. Latitudes and Longitudes are numbers with a maximum length of 13 and the maximum number of decimal is 10. The positive and negative values of these two values indicate their directions(North, South, East, and West). Elevation's values are ranged from 0 to 999 with one decimal place.

Location(State: String, City: String)

When users query a set of weather data of a certain location, they input City and State. The data type of State and City is String with a length no longer than 20.

WeatherData(Station_ID: String, Data_Time: Date)

As the ER diagram shows, 'WeatherData' is the superclass of all subclasses(e.g. AirTemperature, Water). Its two attributes are also primary keys(Station_ID: String, Data_Time: Date). The data type of Station_ID is a unique string that contains characters and numbers with lengths no longer than 10. Date contains data type of Date.

Precipitation(Station_ID: String, Data_Time: Date, DAPR: Float, MDPR: Float, SNOW: Float, SNWD: Float, PRCP: Float)

According to the ER diagram, 'Precipitation' is a subclass of its super class 'WeatherData'. Just like the 'WeatherData', 'Precipitation' also has two attributes, 'Station_ID' and 'Data_Time', which form a pair to become the primary key of this relationship. Learnt from the data file and its explanation document, 'DAPR' is short for Number of days included in the multiday precipitation total, 'MDPR' is short for Multiday precipitation total, 'PRCP' means precipitation, 'SNWD' stands for snow depth and 'SNOW' refers to snowfall. All these attributes are measured in inches and the data type is float.

AirTemperature(Station_ID: String, Data_Time: Date, TAVG: Integer, TMAX: Integer, TMIN: Integer, TOBS: Integer)

Also as a subclass of 'WeatherData', 'AirTemperature' has a pair of primary keys (Station_ID, Data_Time) and its own attributes. 'TAVG' means the average temperature, 'TMAX' means the maximum temperature, 'TMIN' refers to the minimum temperature and 'TOBS' represents temperature at the time of observation. We choose Fahrenheit as the unit of measurement, and the data types of these attributes are all integers.

Evaporation(Station_ID: String, Data_Time: Date, DAEV: Integer, MXPN: Float, MNPN: Float, MDEV: Float, EVAP: Float)

As an important member of weather information, 'Evaporation' is also a subclass of 'Weather', like other subclasses, it uses the pair (Station_ID, Data_Time) as its primary keys. In addition, 'DAEV' refers to the number of days included in the multiday evaporation total, 'MXPN' is short for daily maximum temperature of water in an evaporation pan, 'MNPN' is daily minimum temperature of water in an evaporation pan, 'MDEV' means multiday evaporation total and 'EVAP' means evaporation of water from evaporation pan. The data type of 'DAEV' should be integer and the data types of the other attributes are floats.

Water(Station_ID: String, Data_Time: Date, WESF: Float, WESD: Float)

We can learn from the ER Diagram that the relation 'Water' is a subclass of the super class 'WeatherData'. Since Generalizations are not represented by an own relation, the key of 'WeatherData' (Station_ID: String, Data_Time: Date) is also be used as the key of 'Water'. Besides, 'Water' has two other attributes. 'WESF' means the water equivalent of snowfall, which domain is Integer, and 'WESD' means the water equivalent of snow on the ground, which domain is also Integer.

Wind(Station_ID: String, Data_Time: Date, WSF5: Float, WSF2: Float, WDF5: Integer, WDF2: Integer, AWND: Float)

The relation of 'Wind' is also the subclass of 'WeatherData'. So, the key is (Station_ID: String, Data_Time: Date). There are several other attributes. 'WSF5' means fatest 5-second wind speed, the domain of this attribute is Float. 'WSF2' means fatest 2-minute wind speed, the domain of this attribute is Float. 'WDF5' means the direction of fatest 5-second wind speed, the domain of this attribute is Float. 'WDF2' means the

direction of fatest 2-minute wind speed, the domain of this attribute is Float. 'AWMD' means average wind speed, the domain of this attribute is Float.

WeatherType(Station_ID: String, Data_Time: Date, TypeName: Integer, TypeValue: Boolean)

The relation of 'WeatherType' is also the subclass of 'WeatherData'. So, the key is (Station_ID: String, Data_Time: Date). There are two other attributes. 'TypeName' is the type of the Weather, since the set of data uses number like '01', '02' to represent the weather Type, we decide to define the domain as Integer. 'TypeValue' is exactly a boolean value, it means whether or not the weather belongs to the type.

2 Quality of the transformation of the collection of relation schemas into a collection of SQL table schemas.

According to the relational schema, the 'User' entity has three attributes. The user registers for an account by using Email and Password. We define them as VARCHAR2(20). Since 'User_ID' is a number string automatically generated by the system, we define this as Integer.

According to the relational schema, 'Regular_User' has one attribute. We define them as Integer which is the same as 'User' table.

According to the relational schema, 'Admin' has one attribute. We define them as Integer which is the same as 'User' table.

```
create TABLE "User"(
    User_ID Integer not null,
    Email VARCHAR2(20),
    password VARCHAR2(20),
    primary key (User_ID)
    );
create TABLE Regular_User(
    User_ID Integer not null,
    primary key (User_ID)
    );
create TABLE Admin(
    User_ID Integer not null,
    primary key (User_ID)
    );
```

Figure 2:

	COLUMN_NAME	DATA_TYPE	NULLABLE	DATA_DEFAULT	COLUMN_ID	COMMENTS
1	USER_ID	NUMBER(38,0)	No	(null)	1	(null)

Figure 3:

	COLUMN_NAME	DATA_TYPE	NULLABLE	DATA_DEFAULT	COLUMN_ID	COMMENTS
1	USER_ID	NUMBER(38,0)	No	(null)	1	(null)

Figure 4:

	COLUMN_NAME	DATA_TYPE	NULLABLE	DATA_DEFAULT	COLUMN_ID	COMMENTS
1	USER_ID	NUMBER(38,0)	No	(null)	1	(null)
2	EMAIL	VARCHAR2(20 BYTE)	Yes	(null)	2	(null)
3	PASSWORD	VARCHAR2(20 BYTE)	Yes	(null)	3	(null)

Figure 5:

All the weather data is taken from the physical stations. The relation of 'Station' has attributes of (Station_ID: String, Name: String, Latitude: Float, Longitude: Float, Elevation: Float). The primary key is Station_ID. The data type of Station_ID is strings that contain characters and numbers with lengths no longer than 10. Name_ contains characters only. Latitudes and Longitudes are numbers with a maximum length of 13 and the maximum number of decimal is 10. The positive and negative values of these two values indicate their directions(North, South, East, and West). Elevation's values are ranged from 0 to 999 with one decimal place.

When users query a set of weather data of a certain location, they input City and State. The data type of State and City is String with a length no longer than 20.

As the ER diagram shows, 'WeatherData' is the superclass of all subclasses(e.g. AirTemperature, Water). Its two attributes are also primary keys(Station_ID: String, Data_Time: Date). The data type of Station_ID is a unique string that contains characters and numbers with lengths no longer than 10. Date contains data type of Date.

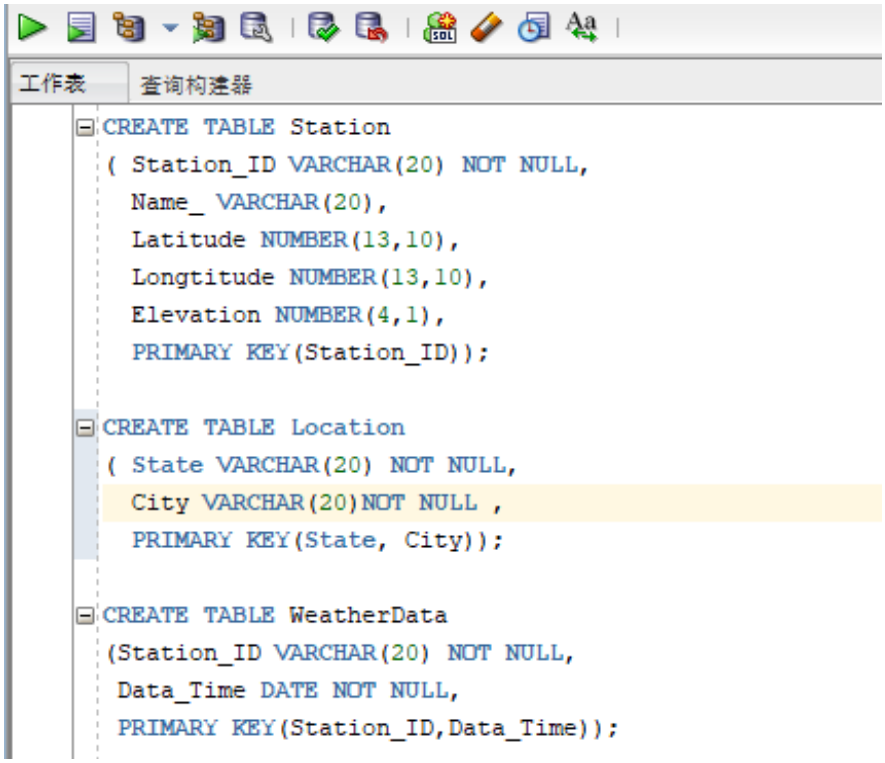


Figure 6:

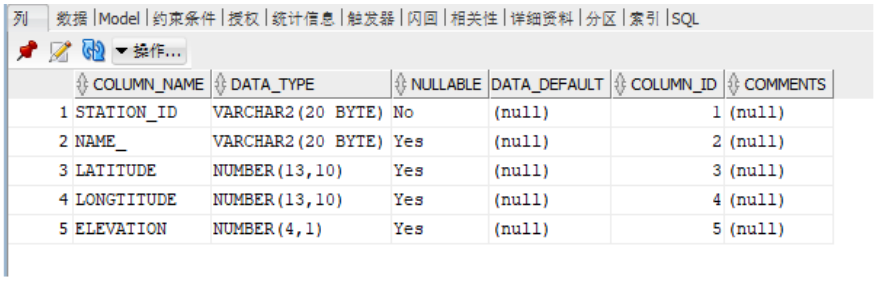


Figure 7:

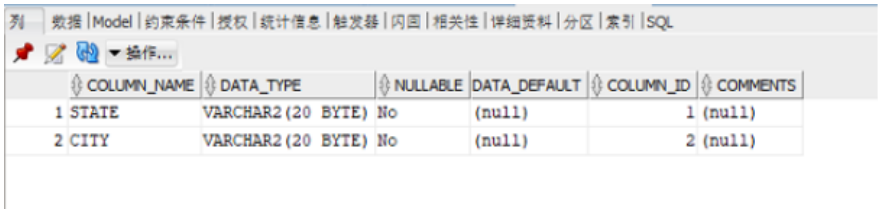


Figure 8:

Figure 9:

According to the relation schema and the dataset, except 'Station_ID' and 'Data_Time', which should be the same as in the 'WeatherData' table, the other attributes in the table 'Precipitation' are reserved for two decimal places, and the integer part is set to three digits. Combined with the data type of the oracle database, numeric(5, 2) is selected as the data type.

In this table, attributes 'Station_ID' and 'Data_Time' should be the same as in the 'WeatherData' table since 'AirTemperature' is a subclass of 'WeatherData'. Besides, the data type of all other attributes should be integer as we do have this data type in oracle database.

Like other subclasses, attributes 'Station_ID' and 'Data_Time' should be the same as in the 'WeatherData'. In addition, we use integer for 'DAEV' since the count of days must be an integer. The other attributes in the table 'Evaporation' are reserved for two decimal places, and the integer part is set to three digits. Taking the data type of the oracle database into consideration, numeric(5, 2) is selected as the data type.

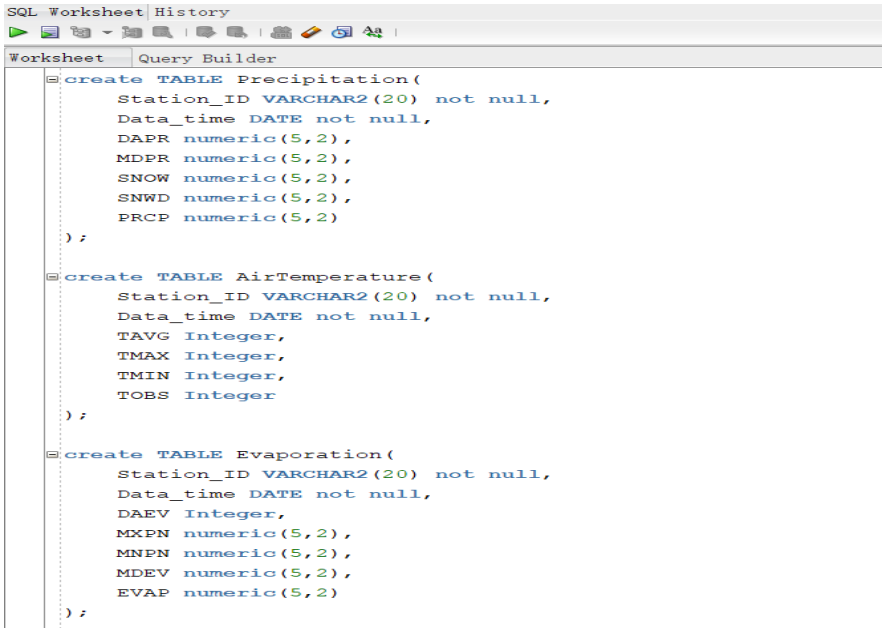
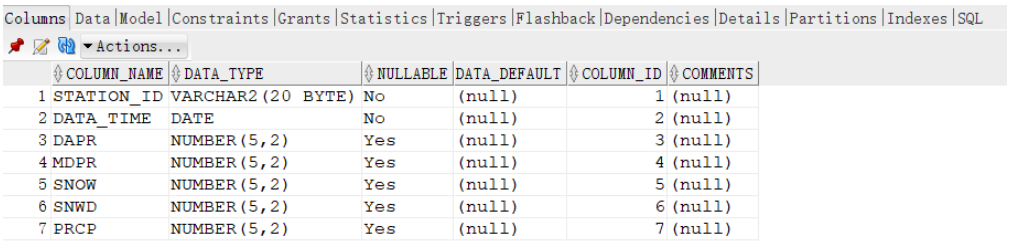


Figure 10:



	COLUMN_NAME	DATA_TYPE	NULLABLE	DATA_DEFAULT	COLUMN_ID	COMMENTS
1	STATION_ID	VARCHAR2(20 BYTE)	No	(null)	1	(null)
2	DATA_TIME	DATE	No	(null)	2	(null)
3	DAPR	NUMBER(5,2)	Yes	(null)	3	(null)
4	MDPR	NUMBER(5,2)	Yes	(null)	4	(null)
5	SNOW	NUMBER(5,2)	Yes	(null)	5	(null)
6	SNWD	NUMBER(5,2)	Yes	(null)	6	(null)
7	PRCP	NUMBER(5,2)	Yes	(null)	7	(null)

Figure 11:

Columns | Data | Model | Constraints | Grants | Statistics | Triggers | Flashback | Dependencies | Details | Partitions | Indexes | SQL

Actions...

	COLUMN_NAME	DATA_TYPE	NULLABLE	DATA_DEFAULT	COLUMN_ID	COMMENTS
1	STATION_ID	VARCHAR2(20 BYTE)	No	(null)	1	(null)
2	DATA_TIME	DATE	No	(null)	2	(null)
3	TAVG	NUMBER(38,0)	Yes	(null)	3	(null)
4	TMAX	NUMBER(38,0)	Yes	(null)	4	(null)
5	TMIN	NUMBER(38,0)	Yes	(null)	5	(null)
6	TOBS	NUMBER(38,0)	Yes	(null)	6	(null)

Figure 12:

Columns | Data | Model | Constraints | Grants | Statistics | Triggers | Flashback | Dependencies | Details | Partitions | Indexes | SQL

Actions...

	COLUMN_NAME	DATA_TYPE	NULLABLE	DATA_DEFAULT	COLUMN_ID	COMMENTS
1	STATION_ID	VARCHAR2(20 BYTE)	No	(null)	1	(null)
2	DATA_TIME	DATE	No	(null)	2	(null)
3	DAEV	NUMBER(38,0)	Yes	(null)	3	(null)
4	MXPN	NUMBER(5,2)	Yes	(null)	4	(null)
5	MNPN	NUMBER(5,2)	Yes	(null)	5	(null)
6	MDEV	NUMBER(5,2)	Yes	(null)	6	(null)
7	EVAP	NUMBER(5,2)	Yes	(null)	7	(null)

Figure 13:

According to the relational schema, the primary key is (Station_ID, Data_Time), which is the same as Weather Data table, so I do not describe this again. Since both of the attributes 'WESF' and 'WESD' are required two decimals and the integer part will not exceed three digits, we define them as numeric(5, 2).

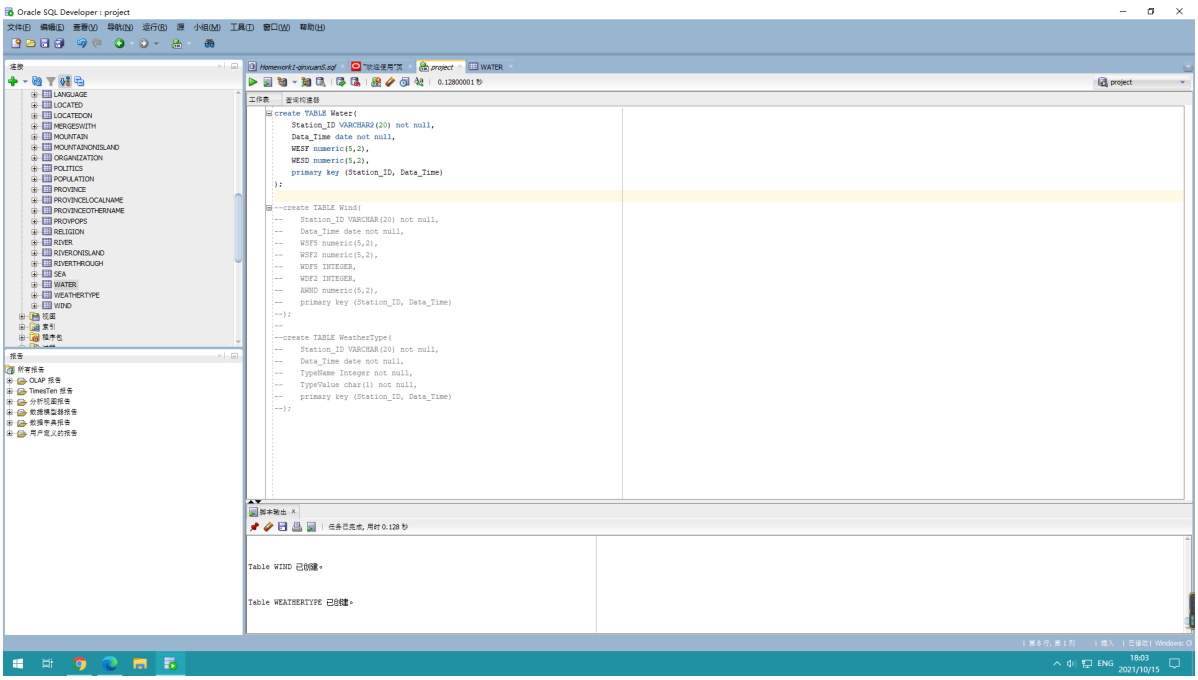


Figure 14: Water-SQL

Here is the snapshot of the Table of Water.

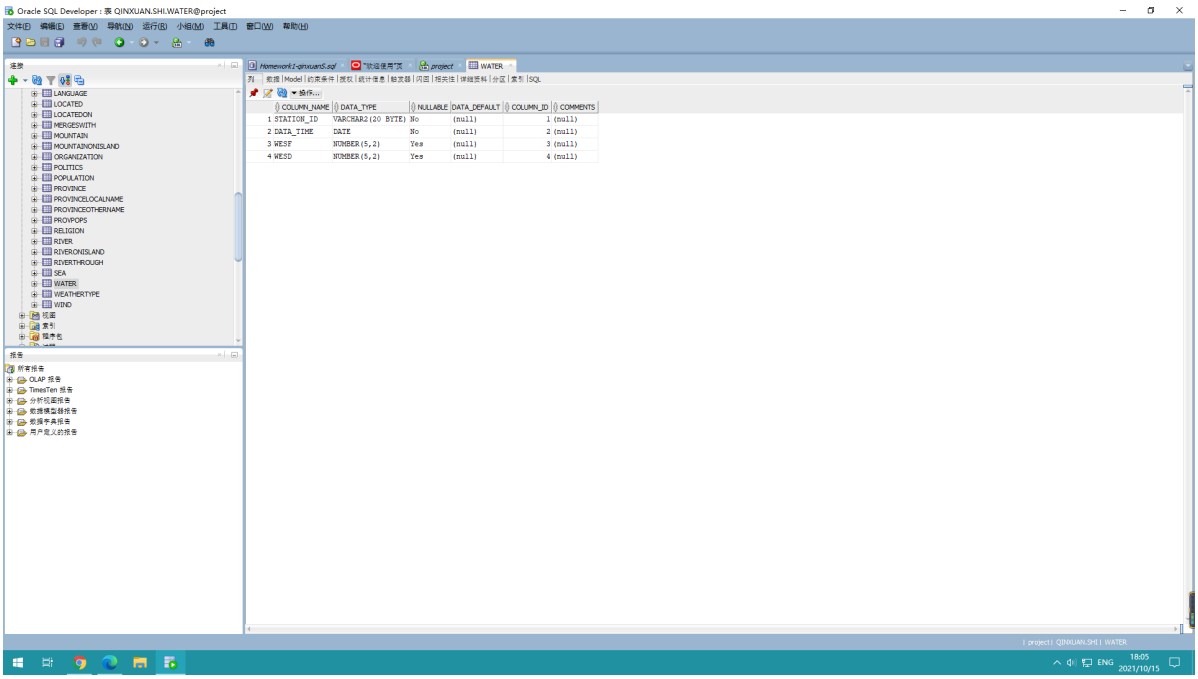


Figure 15: Water-Table

According to the relational schema, the primary key is also (Station_ID, Data_Time), which is the same as Weather Data table. Since the attributes 'WSF5', 'WSF2' and 'AWND' are required two decimals and the integer part will not exceed three digits, we define them as numeric(5, 2). Besides, the attributes 'WDF5' and 'WDF2' are using Integer to represent wind speed, so we define them to be Integer.

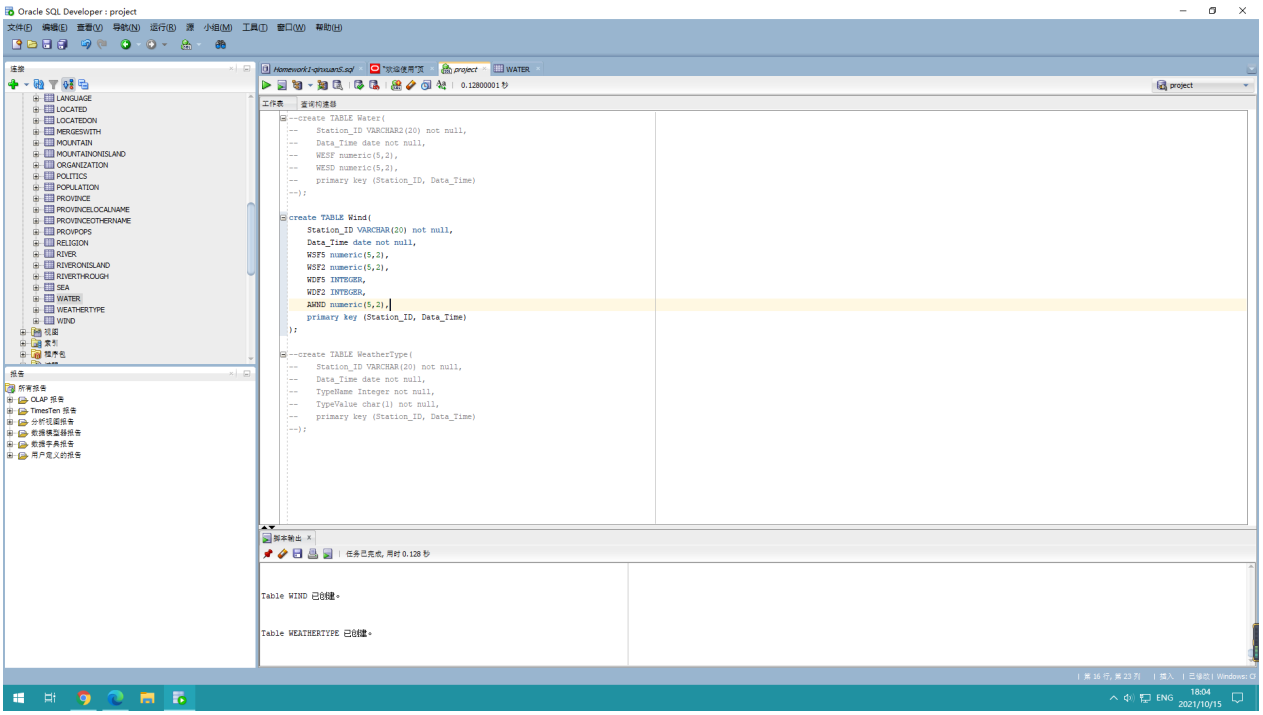


Figure 16: Wind-SQL

Here is the snapshot of the Table of Wind.

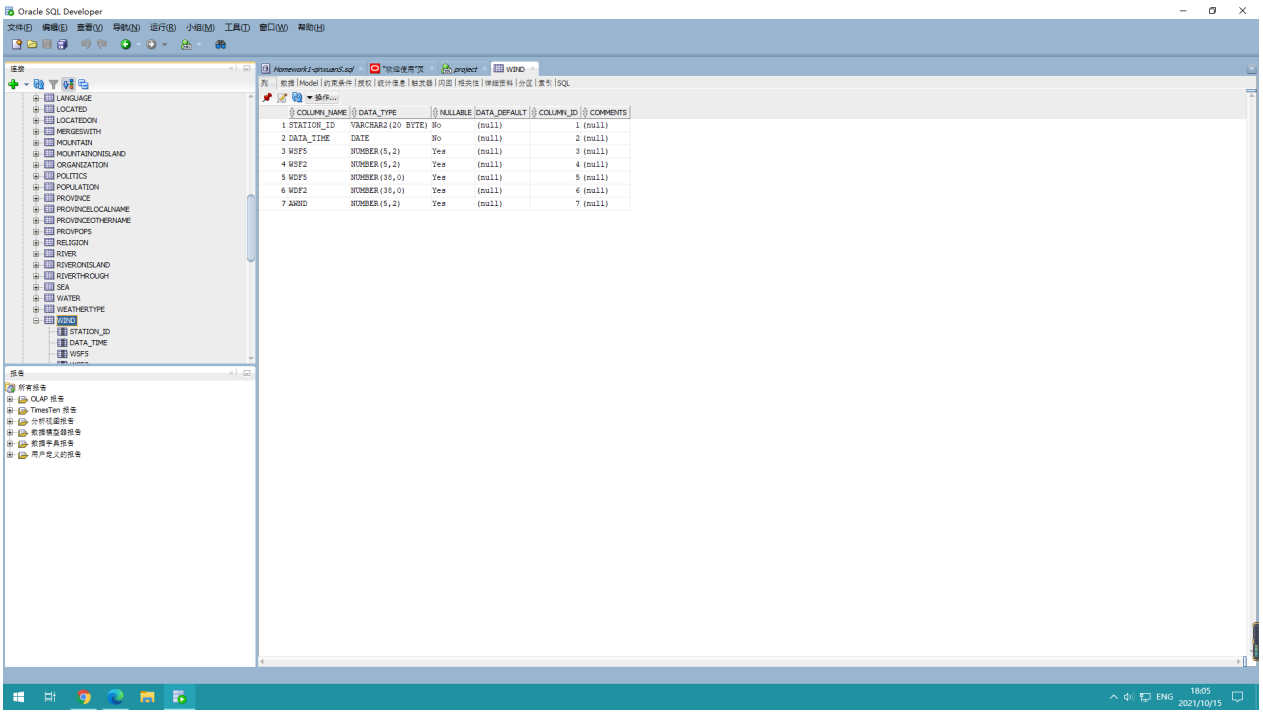


Figure 17: Wind-Table

According to the relational schema, the primary key is (Station_ID, Data_Time), which is the same as Weather Data table. Since data sets use format like '01', '02' to represent different weather Type, we define 'TypeName' to be Integer. The 'TypeValue' is exactly a boolean value, which only contains two possible results 'true' or 'false'. However, the Oracle database does not support boolean type, we choose char(1) to represent this boolean value. '0' means false and '1' means true.

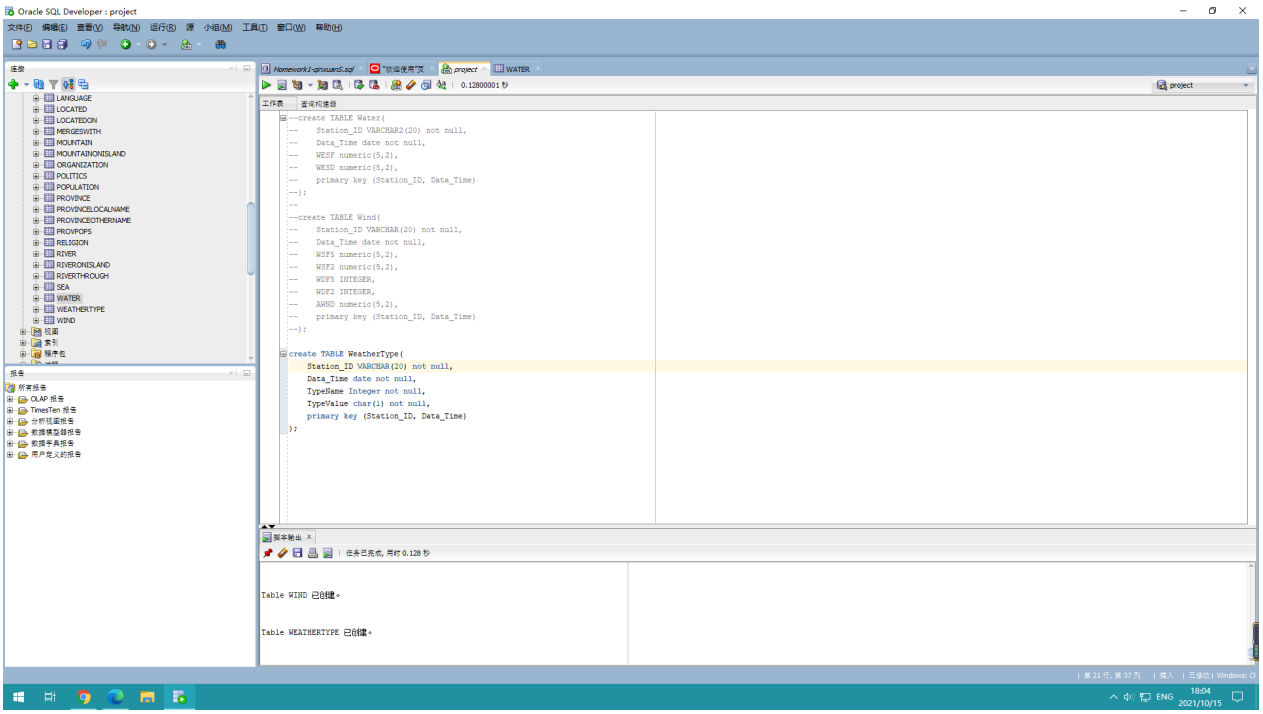


Figure 18: WeatherType-SQL

Here is the snapshot of the Table of WeatherType.

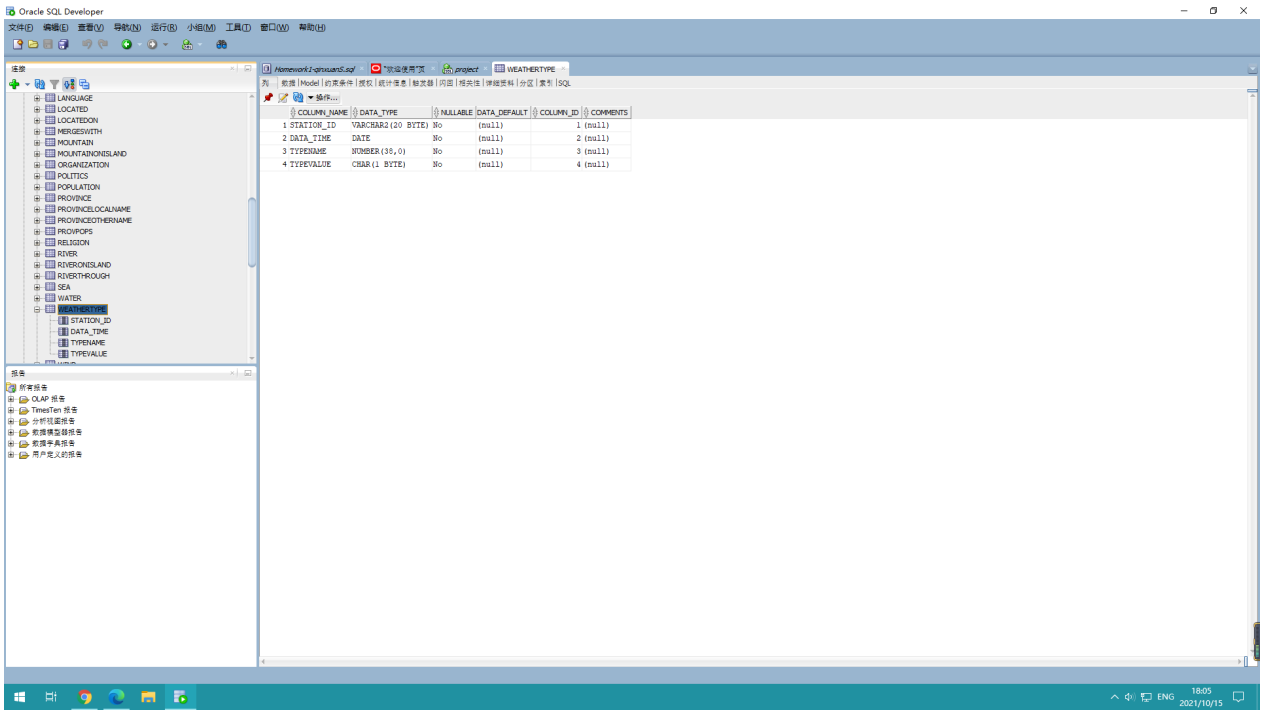


Figure 19: WeatherType-Table