

In [104]:

```
import cx_Oracle
```

In [105]:

```
con = cx_Oracle.connect('hr/password@localhost:1521/xe')
cur = con.cursor()
```

In [106]:

```
create_table = """CREATE TABLE HRD_NET(
    NO NUMBER(38),
    COURSE_NAME VARCHAR2(500),
    ACADEMY VARCHAR2(500),
    FULL_ADDRESS VARCHAR2(500),
    SIDO VARCHAR2(500),
    SIGUNGU VARCHAR2(500),
    TEL VARCHAR2(500),
    START_DATE VARCHAR2(500),
    END_DATE VARCHAR2(500),
    EMPLOY_SCORE NUMBER(30),
    EMPLOY_WAGE NUMBER(30),
    TRAIN_PRICE NUMBER(30),
    RECRUIT_LIMIT NUMBER(30),
    WEEKEN_ABLE VARCHAR2(500),
    TRAIN_PERIOD_DAY NUMBER(30),
    TRAIN_PERIOD_TIME NUMBER(30),
    COUR_SATISFACTION NUMBER(30),
    OCCUPATION VARCHAR2(500),
    TRAIN_LEVEL NUMBER(10),
    NCS VARCHAR2(500),
    RELEATED_CERTIFICATE VARCHAR2(500)
) """
cur.execute(create_table)
```

In [107]:

```
sql_insert = """INSERT INTO HRD_NET (NO,COURSE_NAME,ACADEMY,FULL_ADDRESS, SIDO,
    SIGUNGU,TEL,START_DATE,END_DATE,EMPLOY_SCORE,EMPLOY_WAGE,TRAIN_PRICE,
    RECRUIT_LIMIT, WEEKEN_ABLE, TRAIN_PERIOD_DAY, TRAIN_PERIOD_TIME,
    COUR_SATISFACTION, OCCUPATION, TRAIN_LEVEL, NCS, RELEATED_CERTIFICATE)
    """
```

In [108]:

```
#no 데이터 정제
no_cleansing = "SELECT NVL(TO_NUMBER(NO),NULL) AS NO"
```

In [109]:

```
#coursename 정제
courseName_cleansing = """NVL(REGEXP_REPLACE(LTRIM(RTRIM(COURSENAME)),' ','_'),NULL) AS
COURSENAME"""
```

In [110]:

```
#ACADEMY 정제
academy_cleansing = """NVL(REPLACE(LTRIM(RTRIM(ACADEMY)),' ','_'),NULL) AS ACADEMY"""
```

In [111]:

```
#ADDRESSTEL -> FULL_ADDRESS
full_address_cleansing= """NVL(RTRIM(SUBSTR(REPLACE(REPLACE(ADDRESSTEL, ' ')
    ,'( ',''),1,REGEXP_INSTR(REPLACE(REPLACE(ADDRESSTEL, ' ')','(',')',' '))
),NULL)
    AS FULLADDRESS
```

In [112]:

```
#ADDRESSTEL -> SIDO
sido_cleansing= """
                NVL(RTRIM(SUBSTR(REPLACE(ADDRESSTEL, ' ( ', '1', '3')), NULL) AS SIDO
                """
```

In [113]:

```
#ADDRESSTEL -> SIGUNGU
sigungu_cleansing= """
                NVL(RTRIM(SUBSTR(SUBSTR(REPLACE(REPLACE(ADDRESSTEL, ' ) '), ' ( ', ''),
                REGEXP_INSTR(REPLACE(REPLACE(ADDRESSTEL, ' ) '), ' ( ', ''), ' ')+1,
                REGEXP_INSTR(REPLACE(REPLACE(ADDRESSTEL, ' ) '), ' ( ', ''), ' ☎ ') -3), 1,
                REGEXP_INSTR(SUBSTR(REPLACE(REPLACE(ADDRESSTEL, ' ) '), ' ( ', ''),
                REGEXP_INSTR(REPLACE(REPLACE(ADDRESSTEL, ' ) '), ' ( ', ''), ' ')+1,
                REGEXP_INSTR(REPLACE(REPLACE(ADDRESSTEL, ' ) '), ' ( ', ''), ' ☎ ') -3), ' '))), NULL)
                AS SIGUNGU
                """
```

In [114]:

```
#ADDRESSTEL -> TEL
tel_cleansing= """
                NVL(SUBSTR(REPLACE(REPLACE(ADDRESSTEL, ' ) '), ' ( ', ''),
                REGEXP_INSTR(REPLACE(REPLACE(ADDRESSTEL, ' ) '), ' ( ', ''), ' ☎ ') +3), NULL)
                AS TEL
                """
```

In [115]:

```
#TERM -> START_DATE
start_date_cleansing= """
                SUBSTR(TERM, 1, REGEXP_INSTR(TERM, ' ~ ') -1) AS STARTDATE
                """
```

In [116]:

```
#TERM -> END_DATE
end_date_cleansing= """
                TO_DATE(SUBSTR(TERM, REGEXP_INSTR(TERM, ' ~ ') +3), 'YYYY-MM-DD') AS ENDDATE
                """
```

In [117]:

```
#EMPLOYRATE -> EMPLOY_SCORE
employ_score_cleansing = """TO_NUMBER(REGEXP_SUBSTR(EMPLOYRATE, '[0-9]+'))
                """
```

In [118]:

```
#EMPLOYWAGE -> EMPLOY_WAGE
employ_wage_cleansing = """REGEXP_SUBSTR(EMPLOYWAGE, '[0-9]+')
                """
```

In [119]:

```
#TRAINPRICE -> TRAIN_PRICE
train_price_cleansing = """CASE WHEN TRAINPRICE = '훈련비 : -' THEN NULL
                ELSE TO_NUMBER(REGEXP_SUBSTR(REPLACE(TRAINPRICE, '훈련비 : '), '[0-9]+'))
                END AS TRAINPRICE
                """
```

In [120]:

```
#RECRUITLIMIT -> RECRUIT_LIMIT
recruit_limit_cleansing = "" CASE WHEN RECRUITLIMIT = '모집정원 : ' THEN NULL
                             ELSE REGEXP_SUBSTR(RECRUITLIMIT, '[0-9]+')
                             END AS RECRUITLIMIT
""
```

In [121]:

```
#WEEKENDABLE -> WEEKEND_ABLE
weeken_able_cleansing = "" CASE WHEN WEEKENDABLE LIKE '주말여부 : -' THEN NULL
                             WHEN WEEKENDABLE LIKE '%주말+주중' THEN 'all'
                             WHEN WEEKENDABLE LIKE '%주중' THEN 'wkdays'
                             WHEN WEEKENDABLE LIKE '%주말' THEN 'wkend'
                             END AS WEEKENDABLE
""
```

In [122]:

```
#TRAINPERIOD -> TRAIN_PERIOD_DAY
train_period_day_cleansing = ""CASE WHEN TRAINPERIOD = '훈련시간 : ' THEN NULL
                             ELSE REGEXP_SUBSTR(TRAINPERIOD, '[0-9]+')
                             END AS TRAINPERIOD_DAY
""
```

In [123]:

```
#TRAINPERIOD -> TRAIN_PERIOD_TIME
train_period_time_cleansing = ""TO_NUMBER(REPLACE(REGEXP_SUBSTR(TRAINPERIOD, '총 [0-9]+'), '총 '))
                             AS TRAINPERIOD_TIME
""
```

In [124]:

```
# CORSATISFACTIION -> COUR_SATISFACTION
cour_satisfaction_cleansing = ""CASE WHEN CORSATISFACTION = '과정 만족도 : -' THEN NULL
                             ELSE TO_NUMBER(REGEXP_SUBSTR(SUBSTR(CORSATISFACTION,13), '[0-9]+'))
                             END AS CORSATISFACTION
""
```

In [125]:

```
# OCCUPATION -> OCCUPATION
occupation_cleansing = ""CASE WHEN OCCUPATION IS NOT NULL
                             THEN SUBSTR(REPLACE(OCCUPATION, '직종 (직무수준) : ', ''),1,
                             REGEXP_INSTR(REPLACE(OCCUPATION, '직종 (직무수준) : ', ''), ' ') -1)
                             ELSE NULL
                             END AS OCCUPATION
""
```

In [126]:

```
# OCCUPATION -> TRAIN_LEVEL
train_level_cleansing = ""TO_NUMBER(REPLACE(REGEXP_SUBSTR(OCCUPATION, '[0-9]+수준'), '수준', ''))
                             AS TRAINLEVEL
""
```

In [127]:

```
# NCS -> NCS
ncs_cleansing = ""CASE WHEN NCS = 'NCS적용여부 : 적용' THEN 'YES'
                     WHEN NCS = 'NCS적용여부 : 미적용' THEN 'NO'
                     ELSE NULL
                     END AS NCS
""
```

In [128]:

```
# RELEATEDCERTIFICATE -> RELEATED_CERTIFICATE
releated_certificate_cleansing = """CASE WHEN RELEATEDCERTIFICATE IS NOT NULL
                                THEN REPLACE(LTRIM(REPLACE(RELEATEDCERTIFICATE, '관련
자격증 :','')), ' ','_')
                                ELSE NULL
                                END AS RELEATEDCERTIFICATE
                                """
```

In [129]:

```
t = sql_insert + ' ' + no_cleansing + ',' + courseName_cleansing + ',' + academy_cleansing + ',' + f
ull_address_cleansing + ',' + sido_cleansing + ',' + sigungu_cleansing + ',' + tel_cleansing + '
','+start_date_cleansing + ',' + end_date_cleansing + ',' + employ_score_cleansing + ',' + emp
loy_wage_cleansing + ',' + train_price_cleansing + ',' + recruit_limit_cleansing + ',' + w
eeken_able_cleansing + ',' + train_period_day_cleansing + ',' + train_period_time_cleansing +
',' + cour_satisfaction_cleansing + ',' + occupation_cleansing + ',' +
train_level_cleansing + ',' + ncs_cleansing + ',' + releated_certificate_cleansing + ' FROM R
AW_HRD'
cur.execute(t)
```

In [130]:

```
con.commit()
```