Home Work 5: Oracle SQL Class ETL

# USING SINGLE ROW FUNCTIONS IN AN ETL

-- Source table: tuition\_stg (see below)

-- Source table: states (see below)

-- Source table: school\_type (see below)

-- Target table: tuition\_ETL (created by you)

-- Do not be concerned with mistakes in data

-- Perform ETL from tuition\_stg to tuition\_etl

-- Perform individual transformation for each target column

-- Then place them all in one select statement

-- Using results from transformations

-- Create table tuition\_ETL

-- Determine data type by results of transformations

-- Determine size varchar by results of transformations

-- varchar column size should 10% larger than transformed data

-- While inserting into target

-- Remove any extra spaces while inserting into target

-- Convert all names to Initial Capital letters

-- Example Solution:

-- Source Row:

-- -1025, 19991018, LOMAX-HANNON JC, , 74, GREENVILLE , AL, 36037, , 5685267073, , , 33450, 1897808, 17068, 16124160, 16124160, 9461724, 7265232, , , , 9478792, 7282300, 16124160, 7282300, , 250000, , 4529459, 32959139, 35155632

-- Target Row:

-- 1, Is Grouped, 1025, 18-OCT-99, Lomax-Hannon Jc, PRIVATE, Greenville, AL, South, 36037, , (568) 526-7073 , 0, 1446, 1002

-- -----------------------------------------------------------------------------

-- ETL BUSINESS LOGIC FROM TUITION DATA

-- A. Group\_pk

-- New Field

-- B. School\_id

-- Source Field: groupid

-- Target Field No. 1:

-- if the string begins with '-'

-- then load 'Is Grouped' into grouped

-- else load 'Not Grouped' into grouped

-- Target Field No. 2: tuition\_etl.i\_group\_id

-- If dash, remove dash then insert this into school\_id

-- -----------------------------------------------------------------------------

-- C. Source Field: date\_joined 'YYYYMMDD'

-- Target Field: date\_joined

-- -----------------------------------------------------------------------------

-- D. Schnool Name

-- Source Field No. 1 : institution\_name is never null

-- Source Field No. 2: campus\_name sometimes null

-- Target Field: institution\_name

-- if campus\_name is null

-- then load institution\_name

-- else load campus\_name

-- -----------------------------------------------------------------------------

-- E. School\_type

-- Source Data: table: school\_type

-- Source Field: type\_inst

-- Target Field: institution\_name

-- -----------------------------------------------------------------------------

-- F. City : No transformation needed

-- Source Field : city

-- Target Field : city

-- -----------------------------------------------------------------------------

-- G. State

-- Source Field: state

-- Source Data: table: states

-- Target Field No. 1: state

-- Target Field No. 2: region use the states table to determine region

-- -----------------------------------------------------------------------------

-- H. Zip:

-- Source Field : zip

-- Target Field No. 1: zip5 First 5 values

-- Target Field No. 4: zip4 Last 4 values

-- -----------------------------------------------------------------------------

-- I. Phone:

-- Take least private phone number

-- Convert to format (555) 555-5555

-- Source Fields: work\_ph, cell\_ph, home\_ph

-- Target Field: phone

-- -----------------------------------------------------------------------------

-- J. avg\_local\_help\_per\_student:

-- Source Fields: Avergage (local06, state\_local\_grant\_contract) using fte\_count to divide

-- Target Field avg\_local\_help\_per\_student

-- -----------------------------------------------------------------------------

-- K. avg\_state\_help\_per\_student:

-- Source Fields: Avergage (state03, state06, state09, state\_local\_app) using fte\_count to divide

-- Target field: avg\_state\_help\_per\_student

-- -----------------------------------------------------------------------------

-- L. avg\_fed\_help\_per\_student

-- Source Fields: Avergage (federal03, federal07,federal07\_net\_pell,federal10,federal10\_net\_pell) using fte\_count to divide

-- Target Field avg\_fed\_help\_per\_student

-- =============================================================================

-- NOTES:

-- After reading the instructions above provided by the source data

-- you do some preliminary diffing and realize that the zip instructions

-- are incomplete.

--

-- You note that some numbers are not in 5 4 format

-- For example Sometines it is in 4 4 format

-- Just as you discover this your email lights up with a message from E.J. Pickering:

-- Zooie,

--

-- Sorry, forgot to tell you. Our system is very old and was not designed

-- for exporting. Because of this sometimes instead of 5 leading

-- digits you will only get 4. In thess cases please just add a zero

-- to make the 5th digit. (Our system sometimes drops the zero--not sure why).

--

-- Oh, also: Sometimes instead of 4 trailing digits you will only get 3. In

-- these cases please just add a zero to make the 4th digit.

-- (Same reason as above--again, sorry)

--

-- E.J. Pickering

-- SQL Engineer

-- ACME Anvils

-- You then scan notes from class and find the following:

--

-- data owner says expect 3 valid zip9 formats

-- 555554444

-- 55555-4444

-- 55555

--

-- however when you look at the actual data in tuition\_stg

-- you see there are actually 9 different types of zip code format

select trim(zip), length(trim(zip))

from tuition\_stg;

-- Hand sorted for readability

-- no spaces no dashes

-- 36037 5

-- 6902 4

-- 352940116 9

--

-- spaces

-- 67156 7456 10

-- 54716 745 9

-- 3031 4591 9

--

-- dashes

-- 60610-3431 10

-- 6021-6191 9

-- 31021-017 9

-- This homework is for you to practice and to learn.

-- You may choose any one of several options:

-- 1. Write the SQL that looks most like the class notes:

-- transform just 3 types

-- 555554444

-- 55555-4444

-- 55555

-- Then just move the others

-- with no transformation into the zip5 field or

-- enter 'Not 5 4 format' into zip5 or

--

-- 2. Transform all 9 types found

--



-- =============================================================================

-- TUITION TABLE

drop table tuition\_stg;

create table tuition\_stg (

groupid varchar2(500),

date\_joined varchar2(500),

institution\_name varchar2(500),

campus\_name varchar2(500),

type\_inst varchar2(500),

city varchar2(500),

state varchar2(500),

zip varchar2(500),

ziptype integer,

work\_ph varchar2(500),

cell\_ph varchar2(500),

home\_ph varchar2(500),

fte\_count varchar2(500),

net\_student\_tuition varchar2(500),

federal03 varchar2(500),

state03 varchar2(500),

state\_local\_app varchar2(500),

federal07 varchar2(500),

federal07\_net\_pell varchar2(500),

state06 varchar2(500),

local06 varchar2(500),

state\_local\_grant\_contract varchar2(500),

federal10 varchar2(500),

federal10\_net\_pell varchar2(500),

state09 varchar2(500),

fed\_state\_loc\_grants\_con varchar2(500),

private03 varchar2(500),

endowment03 varchar2(500),

priv\_invest\_endow varchar2(500),

auxother\_rev varchar2(500),

stable\_operating\_rev varchar2(500),

total03\_revenue varchar2(500)

);

Insert into TUITION\_STG values ('-1025','19991018',' LOMAX-HANNON JC',null,'74',' GREENVILLE ','AL','36037',null,'5685267073',null,null,'33450','1897808','17068','16124160','16124160','9461724','7265232',null,null,null,'9478792','7282300','16124160','7282300',null,'250000',null,'4529459','32959139','35155632');

Insert into TUITION\_STG values ('-100663','20041209',' Air Force Communty College ',' CCAF Montgomery','84','STAMFORD','CT','352940116',null,'3954871873','4924201253',null,'31400','2506282','2440214','17107620','17107620','8994152','6877843','1132421',null,'1132421','11434366','9318057','18240042',' ','472196',null,'472196','4477017','37005390','39985544');

Insert into TUITION\_STG values ('1408','19960704','University of Alabama at Birmingham','UA Birmingham','66','Birmingham','AL','6902',null,'6264285610','6135376382','8517627585','48310','5275735',null,'25429592','25429592','11363192','5508845','1916352','2400','1918752','11363192','5508845','27348344','7427597','490198',null,'490198','10061.29004','359450.29','1796802');

Insert into TUITION\_STG values ('1755','19960715',' University of Alabama at Birmingham ',' UA Birmingham','66','ATLANTA','GA',' 54716 745',null,null,null,'7006589290','35660','2538398','2480973','19295288','19295288','9114062','6419768','1367943',null,'1367943','11595035','8900741','20663232','10268684',null,null,'538832','4683969','39853910','43723596');

Insert into TUITION\_STG values ('-1733','20040131',' University of Alabama at Birmingham',' UA Birmingham','66','CHICAGO','IL',' 67156 7456',null,'4778392171',null,null,'42830','5461769','2831854',null,'20616852',null,'9852430','1264558',null,'1264558','16252634','12684284','21881410','13948842','819068',null,'819068','5135419','49351072','54899880');

Insert into TUITION\_STG values ('-1025','19971105',' ALABAMA STATE COLLEGE OF BARBER STYLING',null,'48',' FORSYTH ','GA',' 6021-6191',null,'8671874153',null,'8607476667','37670','2969939','2582813','19192740','19192740','10168088','7431012','1494454',null,'1494454','12750901','10013825',null,'11508279','673651','25000','673651','4196788','41255063','45543896');

Insert into TUITION\_STG values ('-1593','20090718',' NATL COLLEGE ED ',null,'84',' EVANSTON','IL','31021-017',null,'4119111565','6954325264',null,'44290','4382559','3108942','19744360','19744360','13211605','9654227','1225055',null,'1225055','16320547','12763169','20969416','13988224','1138610',null,'1138610','4905869','47146899','53132084');

Insert into TUITION\_STG values ('7789895','19981017',' SAINT JOHNS COLLEGE ',null,'74',' WINFIELD','KS','3031 4591',null,null,'8377247852',null,'47930','6769136','3162155','21909812','21909812','15313054','11759724','774214','3956','778170','18475208',null,'22687980','15700049','1449519',null,'1449519','6814877','55498820','61979400');

Insert into TUITION\_STG values ('-1595','20030125',' The University of Alabama','UA Tuscaloosa','48',' CHICAGO','IL',' 60610-3431',null,null,null,'6962526418','41040','2627916','3061494','19628344','19628344','13633430','10085782','1053452',null,null,'16694924','13147276','20681796','14200728','626850',null,'626850','5446271','46086363','51433912');

-- =============================================================================

-- STATES TABLE

drop table states;

create table states

( st\_id integer primary key,

state varchar2(20),

st char(2),

region varchar2(15));

insert into states values ('1', 'Alabama', 'AL', 'South');

insert into states values ('2', 'Alaska', 'AK', 'West');

insert into states values ('3', 'Arizona', 'AZ', 'West');

insert into states values ('4', 'Arkansas', 'AR', 'South');

insert into states values ('5', 'California', 'CA', 'West');

insert into states values ('6', 'Colorado', 'CO', 'West');

insert into states values ('7', 'Connecticut', 'CT', 'North East');

insert into states values ('8', 'Delaware', 'DE', 'South');

insert into states values ('9', 'Florida', 'FL', 'South');

insert into states values ('10', 'Georgia', 'GA', 'South');

insert into states values ('11', 'Hawaii', 'HI', 'West');

insert into states values ('12', 'Idaho', 'ID', 'West');

insert into states values ('13', 'Illinois', 'IL', 'Midwest');

insert into states values ('14', 'Indiana', 'IN', 'Midwest');

insert into states values ('15', 'Iowa', 'IA', 'Midwest');

insert into states values ('16', 'Kansas', 'KS', 'Midwest');

insert into states values ('17', 'Kentucky', 'KY', 'South');

insert into states values ('18', 'Louisiana', 'LA', 'South');

insert into states values ('19', 'Maine', 'ME', 'North East');

insert into states values ('20', 'Maryland', 'MD', 'South');

insert into states values ('21', 'Massachusetts', 'MA', 'North East');

insert into states values ('22', 'Michigan', 'MI', 'Midwest');

insert into states values ('23', 'Minnesota', 'MN', 'Midwest');

insert into states values ('24', 'Mississippi', 'MO', 'South');

insert into states values ('25', 'Missouri', 'MS', 'Midwest');

insert into states values ('26', 'Montana', 'MT', 'West');

insert into states values ('27', 'Nebraska', 'NE', 'Midwest');

insert into states values ('28', 'Nevada', 'NV', 'West');

insert into states values ('29', 'New Hampshire', 'NH', 'North East');

insert into states values ('30', 'New Jersey', 'NJ', 'North East');

insert into states values ('31', 'New Mexico', 'NM', 'West');

insert into states values ('32', 'New York', 'NY', 'North East');

insert into states values ('33', 'North Carolina', 'NC', 'South');

insert into states values ('34', 'North Dakota', 'ND', 'Midwest');

insert into states values ('35', 'Ohio', 'OH', 'Midwest');

insert into states values ('36', 'Oklahoma', 'OK', 'South');

insert into states values ('37', 'Oregon', 'OR', 'West');

insert into states values ('38', 'Pennsylvania', 'PA', 'North East');

insert into states values ('39', 'Rhode Island', 'RI', 'North East');

insert into states values ('40', 'South Carolina', 'SC', 'South');

insert into states values ('41', 'South Dakota', 'SD', 'Midwest');

insert into states values ('42', 'Tennessee', 'TN', 'South');

insert into states values ('43', 'Texas', 'TX', 'South');

insert into states values ('44', 'Utah', 'UT', 'West');

insert into states values ('45', 'Vermont', 'VA', 'North East');

insert into states values ('46', 'Virginia', 'VT', 'South');

insert into states values ('47', 'Washington', 'WA', 'West');

insert into states values ('48', 'West Virginia', 'WI', 'South');

insert into states values ('49', 'Wisconsin', 'WV', 'Midwest');

insert into states values ('50', 'Wyoming', 'WY', 'West');

commit;

-- =============================================================================

-- SCHOOL\_TYPE TABLE

drop table school\_type;

create table school\_type

( type\_id integer primary key,

type\_name varchar2(25));

insert into school\_type values (66, 'STATE UNIVERSITY');

insert into school\_type values (48, 'STATE COLLEGE');

insert into school\_type values (30, 'STATE COMMUNITY COLLEGE');

insert into school\_type values (38, 'LOCAL COMMUNITY COLLEGE');

insert into school\_type values (69, 'PRIVATE JR COLLEGE');

insert into school\_type values (74, 'PRIVATE');

insert into school\_type values (84, 'MILITARY');

commit;