INF20010 / 60014 Assignment 2

Assignment Value: 15% of your final mark

The assignment is to be done groups of up to 2 people

Due Date/Time: 18th of May 2015 @08:30am

SUBMISSION REQUIREMENTS

All submissions must be made by a team using ESP via: https://esp.ict.swin.edu.au/.

Team Registration

- You must register your team in ESP BEFORE by Friday, 1st of May 2015
- **Students that are not in a registered team by this date will be required to do the assignment individually.**
- You cannot submit your assignment unless you join an ESP team.

Files required

A ZIP file containing:

- Tnnn_Ass2_Script.SQL : - where Tnnn is your ESP team number
- Tnnn_Ass2_Queries.SQL : - where Tnnn is your ESP team number

All files must have all team members student number and name in comments at the beginning of the file. Failure to include this will result in an immediate 50% penalty

GENERAL DESCRIPTION:

The aim of this assignment is to:

- Create a data warehouse.
- **E**xtracting data from source tables.
- **T**ransforming and cleaning the data, logging all changes.
- **L**oading the data into a data warehouse.

Finally you will be required to write queries based on data stored in the data warehouse.

Note:

1. All operations of ETL must be achieved by running SQL scripts. This scripts must work on any source data values. (If the convener decides to change the source data values, then your code must still work correctly. You cannot assume that a particular data value such as person Fred Smith or that Software Product 15 will always exist in the source data tables.)
2. **The script may only contain standard SQL, no PL/SQL may be used (i.e. no stored procedures, functions, or anonymous blocks).**

SOURCE DATA BACKGROUND:

HARRIS and RIGG were two medium sized organizations located in Melbourne and Brisbane. Both companies imported and sold similar products. Both companies sold their products via their on-line stores.

Customers for both companies came from all around Australia. Most customers remained loyal to the company that they first purchased items from. So generally HARRIS customers didn't buy from RIGG and RIGG customers didn't buy from HARRIS. However there are some exceptions.

HARRIS and RIGG have now merged and become the ACME organization. The product data have been merged and are now common to both branches of ACME. The customer and sales data have **not** been merged and are still stored separately within each branch. This means that there are two different customer tables and two different sales tables.

The convenor of INF20010/60014 has a copy of all source tables in his database account. Read-Only public access has been defined so you can refer to these tables. You cannot modify data in these tables. There is no need to copy the data in these tables to your account. Note: The data in the source tables may change!

SOURCE DATA TABLES:

These are the names of the existing source tables:

A2PRODUCT	A2SALEMELB	A2CUSTCATEGORY
A2CUSTBRIS	A2MANUFACTURER	DWTIME
A2CUSTMELB	A2SHIPPING	
A2SALEBRIS	A2PRODCATEGORY	

Source data from the source tables cannot be guaranteed to be clean.
Your task is to check data in the source tables.

- Some rows will be no errors
- Some rows may have errors that can be fixed
- Some rows may have errors that cannot be fixed and will be rejected
- Some rows will have multiple errors

Rows that have no errors or that can be fixed will be uploaded into Data Warehouse tables

ERROR EVENT TABLE

During the ETL process, you must keep track of which source data rows need to be fixed or rejected.

The schema for this table is:

```
A2ERROREVENT (
ERRORID              INTEGER,
SOURCE_ROWID         ROWID,
SOURCE_TABLE         VARCHAR2(30),
FILTERID             INTEGER,
DATETIME             DATE,
ACTION               VARCHAR2(6),
CONSTRAINT   ERROREVENTACTION
CHECK (ACTION IN ('SKIP','MODIFY'))
);
```

Do not alter this schema in anyway, doing so will result in a major loss of marks.

DATA WAREHOUSE TABLES:

You will create some tables in your account that simulate a Data Warehouse

These are the names of the data warehouse tables that you must create:

DWPROD *(DWPRODID, DWSOURCETABLE, DWSOURCEID, PRODNAME, PRODCATNAME, PRODMANUNAME, PRODSHIPNAME)*

DWCUST *(DWCUSTID, DWSOURCEIDBRIS, DWSOURCEIDMELB, FIRSTNAME, SURNAME, GENDER, PHONE, POSTCODE, CITY, STATE, CUSTCATNAME)*

DWSALE *(DWSALEID, DWCUSTID, DWPRODID, DWSOURCEIDBRIS, DWSOURCEIDMELB, QTY, SALE_DWDATEID, SHIP_DWDATEID, SALEPRICE)*

DWDATE has already been created. Do not recreate this table, but refer to it in your data warehouse.

PART 1. DATA WAREHOUSE SETUP (0 marks)

Almost all of the tasks below ask you to write SQL code. The code is to be placed in the Tnnn_Ass2_Script.SQL file. Each task must be separated by a line of code that contains a single / symbol.

Task 1.1

Write code to create the **A2ErrorEvent** table DDL and place the code into the Ass2_Script file. This table must **exactly** match the provided schema.

Add appropriate Drop Table and Create Sequence statements for this table in the appropriate section of the script file so that each time you run the script the code can be tested without any effects from previous attempts.

Task 1.2

Write the DDL code to create all required data warehouse tables according to the schema provided.

Add appropriate Drop Table and Create Sequence statements for this table in the appropriate section of the script file so that each time you run the script the code can be tested without any effects from previous attempts.

Task 1.3

Write the DDL to create the **GENDERSPELLING** table and the DML to insert the following data:

Invalid Value	New Value
MAIL	M
WOMAN	F
FEM	F
FEMALE	F
MALE	M
GENTLEMAN	M
MM	M
FF	F
FEMAIL	F

PART 2. PRODUCT TRANSFER (10 marks)

Task 2.1

Filter #1

a) Write code to insert a row into the **A2ERROREVENT** table for each row in the A2PRODUCT table which has a null product name value. The action must be set to 'SKIP'

b) **Testing**: You should test your code and ensure that **A2ERROREVENT** has been updated correctly.

Task 2.2

Filter #2

a) Write code to insert a row into the **Ass2ErrorEvent** table for each row in the A2Product table which has a null manufacturer code value. The action must be set to 'MODIFY'.

b) **Testing**: You should test your code and ensure that **A2ERROREVENT** has been updated correctly.

Task 2.3

Filter #3

a) Write code to insert a row into the **Ass2ErrorEvent** table for each row in the A2Product table which refers to a invalid row in the A2PRODCATEGORY, or is null. The action must be set to 'MODIFY'.

b) **Testing**: You should test your code and ensure that **A2ERROREVENT** has been updated correctly.

Task 2.4.1

Do **NOT** include this code in Ass2_Script File. This task is simply here to make the next task easier to understand.

Write a **select** statement that lists every row from the **A2Product** table that does not have the same ROWID as any of the source_rowid values in the ERROREVENT table.

Task 2.4.2.

Do **NOT** include this code in Ass2_Script File. This task is simply here to make the next task easier to understand.

Same as above but the select statement must join a2PRODUCT table to the A2MANUFACTURER, A2SHIPPING and a2PRODCATEGORY tables.

- The select statement must show the CategoryName of each product (from A2PRODCATEGORY)
- The select statement must show the manufacturer name of each product (from A2MANUFACTURER)
- The select statement must show the shipping DESCRIPTION of each product (from A2SHIPPING)

Task 2.4.3

a) Write code to insert rows into the **DWPROD** table.

Insert the rows from the A2PRODUCT table that do not have the same ROWID as any of the source_rowid values in the A2ERROREVENT table.

- All rows will require shipping, manufacturer and product category details

Testing: You should test your code and ensure that DWPROD table have been updated correctly.

Task 2.4.4

b) Write code to insert rows into the **DWPROD** table.

Insert the rows from theA2PRODUCT table that have the same ROWID as the source_rowid values in the ERROREVENT table and have a filterid value = 2.

- In each case, set the *PRODMANUNAME* to 'UNKNOWN'
- All rows will require shipping, and product category details.

Testing: You should test your code and ensure that the DWPROD tables have been updated correctly.

Task 2.4.5

c) Write code to insert rows into the **DWPROD** table.

Insert the rows from theA2PRODUCT table that have the same ROWID as the source_rowid values in the ERROREVENT table and have a filterid value = 3.

- In each case, set the *PRODCATNAME* to 'UNKNOWN'
- All rows will require shipping, and manufacturer name.

Testing: You should test your code and ensure that the DWPROD tables have been updated correctly.

Task 2.4.6

Testing: You should execute the **entire** Ass2_Script file to ensure that all tasks can run without error.

PART 3. CUSTOMER BRISBANE TRANSFER (15 marks)

This task deals with the A2CUSTBRIS table.

Task 3.1

Filter #4

a) Write code to insert a row into the **A2ErrorEvent** table for each row in the A2CUSTBRIS table where CustCatCode does not match a PK value in the A2CUSTCATEGORY source table. The action must be set to 'MODIFY'

b) **Testing**: You should test your code and ensure that A2ErrorEvent have been updated correctly.

Task 3.2

Filter #5

a) Write code to insert a row into the **A2ErrorEvent** table for each row in the A2CUSTBRIS table where PHONE contains a space or a hyphen. The action must be set to 'MODIFY'

b) **Testing**: You should test your code and ensure that A2ErrorEvent have been updated correctly.

Task 3.3

Filter #6

a) Write code to insert a row into the A2ErrorEvent table for each row in the A2CUSTBRIS table where PHONE Number with valid characters does not have length equal to 10. The action must be set to 'SKIP'

b) **Testing**: You should test your code and ensure that A2ErrorEvent have been updated correctly.

Task 3.4

Filter #7

a) Write code to insert a row into the A2ErrorEvent table for each row in the A2CUSTBRIS table where GENDER value is not M or F (not case sensitive). The action must be set to 'MODIFY'

b) **Testing**: You should test your code and ensure that A2ErrorEvent have been updated correctly.

Task 3.5.1

a) Write code to insert rows into the **DWCUST** table.

Insert the rows from the A2CUSTBRIS table that do not have the same ROWID as any of the source_rowid values in the A2ERROREVENT table.

- All rows will require Customer Category Name.
- All GENDER values should be in uppercase.

b) **Testing**: You should test your code and ensure that DWCUST table have been updated correctly.

Task 3.5.2

a) Write code to insert rows into the **DWCUST** table.

Insert the rows from theA2CUSTBRIS table that have the same ROWID as the source_rowid values in the ERROREVENT table and have a filterid value =4.

- In each case, set the *CUSTCATNAME* to 'UNKNOWN'
- All GENDER values should be in uppercase.

b) **Testing**: You should test your code and ensure that the DWCUST tables have been updated correctly.

Task 3.5.3

a) Write code to insert rows into the **DWCUST** table.

Insert the rows from theA2CUSTBRIS table that have the same ROWID as the source_rowid values in the ERROREVENT table and have a filterid value =5.

- In each case, remove all spaces or hyphens.
- All GENDER values should be in uppercase.

b) **Testing**: You should test your code and ensure that the DWCUST tables have been updated correctly.

Task 3.5.4

c) Write code to insert rows into the **DWCUST** table.

Insert the rows from theA2CUSTBRIS table that have the same ROWID as the source_rowid values in the ERROREVENT table and have a filterid value =7.

- In each case, replace the invalid GENDER value with the correct value according to the **GENDERSPELLING** table.
- Where the invalid GENDER value is not in the **GENDERSPELLING** table insert the value 'U' (i.e. unknown).
- All GENDER values should be in uppercase.

d) **Testing**: You should test your code and ensure that the DWCUST tables have been updated correctly.

Task 3.5.5

Testing: You should execute the **entire** Ass2_Script file to ensure that all tasks can run without error.

PART 4. CUSTOMER MELBOURNE TRANSFER (15 marks)

This task deals with the A2CUSTMELB table.

NOTE*: Normally the Melbourne Customer data would require the same filtering as the Brisbane Customer data, however in this assignment you may assume that all of the data values in A2CUSTMELB are clean.*

Task 4.1

a) Write code to **MERGE** rows into the **DWCUST** table.

Merge the rows from the A2CUSTMELB into DWCUST.

- Where the FIRSTNAME, SURNAME, & POSTCODE of the Melbourne Customer **matches** a record already in DWCUST table, update the DWCUST record to include the *DWSOURCEIDMELB*.
- Where the FIRSTNAME, SURNAME, & POSTCODE of the Melbourne Customer **do not match** a record already in DWCUST table, insert the row from A2CUSTMELB into DWCUST.
- All rows will require Customer Category Name.
- All GENDER values should be in uppercase.

b) **Testing**: You should test your code and ensure that DWCUST table have been updated correctly.

Task 4.2

Testing: You should execute the **entire** Ass2_Script file to ensure that all tasks can run without error.

PART 5. BRISBANE SALES TRANSFER (25 marks)

This task deals with the A2SALEBRIS table.

Task 5.1

Filter #8

a) Write code to insert a row into the A2ErrorEvent table for each row in the A2SALEBRIS table where PRODID does not match a DWPROD. *DWSOURCEID.* The action must be set to 'SKIP'

b) **Testing**: You should test your code and ensure that A2ErrorEvent have been updated correctly.

Task 5.2

Filter #9

a) Write code to insert a row into the A2ErrorEvent table for each row in the A2SALEBRIS table where CUSTID does not match a DWCUST. *DWSOURCEIDBRIS.* The action must be set to 'SKIP'

b) **Testing**: You should test your code and ensure that A2ErrorEvent have been updated correctly.

Task 5.3

Filter #10

a) Write code to insert a row into the A2ErrorEvent table for each row in the A2SALEBRIS table where SHIPDATE is earlier than SALEDATE. The action must be set to 'MODIFY'

b) **Testing**: You should test your code and ensure that A2ErrorEvent have been updated correctly.

Task 5.4

Filter #11

a) Write code to insert a row into the A2ErrorEvent table for each row in the A2SALEBRIS table where SALEPRICE is Null. The action must be set to 'MODIFY'

b) **Testing**: You should test your code and ensure that A2ErrorEvent have been updated correctly.

Task 5.5

a) Write the code that inserts records from A2SALEBRIS into DWSALE into the script file.

This code must ensure that:

- All rows not listed in the ERROREVENT table are uploaded to DWSALE
- Each new sale added to the DWSALE table must be given unique DWSALEID value. The DWSALEID value must be obtained from a the appropriate sequence as created in Part 1
- The DWSOURCEIDBRIS value must be set to the SALEID of the source table
- The DWCUSTID value must be set to the appropriate DWCUSTID of the DWCUST table
- The DWPRODID value must be set to the appropriate DWPRODID of the DWPROD table
- SALE_DWDATEID, SHIP_DWDATEID in DWSALE must be set to the appropriate DWDATE.DATEKEY

b) **Testing**: You should test your code and ensure that DWSALE have been updated correctly.

Task 5.6

a) Write the code that inserts records from A2SALEBRIS that have the same ROWID as the source_rowid values in the ERROREVENT table and have a filterid value =10 into DWSALE into the script file.

This code must ensure that:

- Each new sale added to the DWSALE table must be given unique DWSALEID value. The DWSALEID value must be obtained from a the appropriate sequence as created in Part 1
- The DWSOURCEIDBRIS value must be set to the SALEID of the source table
- The DWCUSTID value must be set to the appropriate DWCUSTID of the DWCUST table
- The DWPRODID value must be set to the appropriate DWPRODID of the DWPROD table
- **MODIFY the SHIPDATE so that it equals the SALEDATE plus 2 days**
- SALE_DWDATEID, SHIP_DWDATEID in DWSALE must be set to the appropriate DWDATE.DATEKEY

b) **Testing**: You should test your code and ensure that DWSALE have been updated correctly.

Task 5.7

a) Write the code that inserts records from A2SALEBRIS that have the same ROWID as the source_id values in the ERROREVENT table and have a filterid value =10 into DWSALE into the script file.

This code must ensure that:

- Each new sale added to the DWSALE table must be given unique DWSALEID value. The DWSALEID value must be obtained from a the appropriate sequence as created in Part 1
- The DWSOURCEIDBRIS value must be set to the SALEID of the source table
- The DWCUSTID value must be set to the appropriate DWCUSTID of the DWCUST table
- The DWPRODID value must be set to the appropriate DWPRODID of the DWPROD table
- **MODIFY the SALEPRICE so that it equals the maximum sale price for that product in A2SALEBRIS**
- SALE_DWDATEID, SHIP_DWDATEID in DWSALE must be set to the appropriate DWDATE.DATEKEY

b) **Testing**: You should test your code and ensure that DWSALE table have been updated correctly.

Task 5.8

Testing: You should execute the **entire** Ass2_Script file to ensure that all tasks can run without error.

PART 6. MELBOURNE SALES TRANSFER (5 marks)

This task deals with the A2SALEMELB table.

Task 6.1

Filter #12

a) Write code to insert a row into the A2ErrorEvent table for each row in the A2SALEMELB table where PRODID does not match a DWPROD. *DWSOURCEID.* The action must be set to 'SKIP'

b) **Testing**: You should test your code and ensure that A2ErrorEvent have been updated correctly.

Task 6.2

Filter #13

a) Write code to insert a row into the A2ErrorEvent table for each row in the A2SALEMELB table where CUSTID does not match a DWCUST. *DWSOURCEIDMELB.* The action must be set to 'SKIP'

b) **Testing**: You should test your code and ensure that A2ErrorEvent have been updated correctly.

Task 6.3

Filter #14

a) Write code to insert a row into the A2ErrorEvent table for each row in the A2SALEMELB table where SHIPDATE is earlier than SALEDATE. The action must be set to 'MODIFY'

b) **Testing**: You should test your code and ensure that A2ErrorEvent have been updated correctly.

Task 6.4

Filter #15

a) Write code to insert a row into the A2ErrorEvent table for each row in the A2SALEMELB table where SALEPRICE is Null. The action must be set to 'MODIFY'

b) **Testing**: You should test your code and ensure that A2ErrorEvent have been updated correctly.

Task 6.5

a) Write the code that inserts records from A2SALEMELB into DWSALE into the script file.

This code must ensure that:

- All rows not listed in the ERROREVENT table are uploaded to DWSALE
- Each new sale added to the DWSALE table must be given unique DWSALEID value. The DWSALEID value must be obtained from a the appropriate sequence as created in Part 1
- The DWSOURCEIDMELB value must be set to the SALEID of the source table
- The DWCUSTID value must be set to the appropriate DWCUSTID of the DWCUST table
- The DWPRODID value must be set to the appropriate DWPRODID of the DWPROD table
- SALE_DWDATEID, SHIP_DWDATEID in DWSALE must be set to the appropriate DWDATE.DATEKEY

b) **Testing**: You should test your code and ensure that DWSALE have been updated correctly.

Task 6.6

a) Write the code that inserts records from A2SALEMELB that have the same ROWID as the source_rowid values in the ERROREVENT table and have a filterid value =14 into DWSALE into the script file.

This code must ensure that:

- Each new sale added to the DWSALE table must be given unique DWSALEID value. The DWSALEID value must be obtained from a the appropriate sequence as created in Part 1
- The DWSOURCEIDMELB value must be set to the SALEID of the source table
- The DWCUSTID value must be set to the appropriate DWCUSTID of the DWCUST table
- The DWPRODID value must be set to the appropriate DWPRODID of the DWPROD table
- **MODIFY the SHIPDATE so that it equals the SALEDATE plus 2 days**
- SALE_DWDATEID, SHIP_DWDATEID in DWSALE must be set to the appropriate DWDATE.DATEKEY

b) **Testing**: You should test your code and ensure that DWSALE have been updated correctly.

Task 6.7

a) Write the code that inserts records from A2SALEMELB that have the same ROWID as the source_rowid values in the ERROREVENT table and have a filterid value =15 into DWSALE into the script file.

This code must ensure that:

- Each new sale added to the DWSALE table must be given unique DWSALEID value. The DWSALEID value must be obtained from a the appropriate sequence as created in Part 1
- The DWSOURCEIDMELB value must be set to the SALEID of the source table
- The DWCUSTID value must be set to the appropriate DWCUSTID of the DWCUST table
- The DWPRODID value must be set to the appropriate DWPRODID of the DWPROD table
- **MODIFY the SALEPRICE so that it equals the maximum sale price for that product in A2SALEMELB**
- SALE_DWDATEID, SHIP_DWDATEID in DWSALE must be set to the appropriate DWDATE.DATEKEY

b) **Testing**: You should test your code and ensure that DWSALE table have been updated correctly.

Task 6.8

Testing: You should execute the **entire** Ass2_Script file to ensure that all tasks can run without error.

PART 7. MELBOURNE SALES TRANSFER (ADVANCED) (20 marks)

Task 7.1

a) Until now no source data row has failed multiple filters, however some rows in A2SALEMELB do. No row that fails multiple filters is to be present in the DWSALE table. Modify your script so that this does not happen.

b) **Testing**: You should test your code and ensure that DWSALE table have been updated correctly.

PART 8. QUERIES (10 marks)

Write SQL queries based on your data warehouse to display the following information.
Note: The values used to create these examples will not match current values in the database tables.

1. List each weekday (e.g. Monday, Tuesday…) and the total sales (qty * saleprice) for that day. The list must be in descending total sequence. Note: 7 rows should be listed by this query
   E.g.

WEEKDAY	TOTAL SALES
FRIDAY	52400
MONDAY	47340
SUNDAY	46235 …

2. List each customer category and the total sales (qty * saleprice).
   The list must be in ascending total sequence.
   E.g.

CUSTCATNAME	TOTAL SALES
Exclusive	58200
Gold	61430
Silver	73439 …

3. List each product manufacturer and the total qty sold
   The list must be in descending total sequence.
   E.g.

PRODMANUNAME	TOTAL QTY SOLD
Lum-Tech	3210
Island Group	2956…

4. List the **top 10** customers based on total sales (qty * saleprice).
   The list must be in descending total sequence.
   E.g.

DWCUSTID	FISRTNAME	SURNAME	TOTAL SALES
1045	Fred	Smith	2400
7776	Sue	Davies	2260
56	Emma	Jones	1998 …

5. List the **bottom 10** products based on total qty sales
   The list must be in ascending total sequence.
   E.g.

DWPRODID	PRODNAME	TOTAL SALES
321	Product BD	65
95	Product DY	89
206	Product JJ	132 …

6. List the top city name based on total sales (qty * price) for each state
   The list must be in ascending state sequence.
   E.g.

STATE	CITY	TOTAL SALES
ACT	DOWNER	32010
NSW	TURRAMURRA	23400
QLD	HEATHWOOD	14200 …

REQUIREMENTS

-

Testing the script.

Before submitting the script file you must

- Copy the entire contents of the file and paste it into an SQL Developer worksheet.
- Press the Run script button 

The entire script should execute without error.

The script will drop tables and sequences, create tables and sequences, test ETL filters against source data, update the ErrorEvent table, upload data into the data warehouse tables and finally execute some SQL queries.