Booz Allen Hamilton

SQL Suggested Practices

**Gene Belford**

**July 2018**

Table of Contents

[1. Suggestions 5](#_Toc519494775)

[1.1. Incorporate 5](#_Toc519494776)

[1.2. Avoid 5](#_Toc519494777)

[2. Naming 5](#_Toc519494778)

[2.1. Reserve Words 5](#_Toc519494779)

[2.2. Tables and Views 5](#_Toc519494780)

[2.3. Column/Field Names 6](#_Toc519494781)

[2.4. Functions and Procedures 6](#_Toc519494782)

[2.5. Variables 6](#_Toc519494783)

[2.6. Constants Variables 6](#_Toc519494784)

[2.7. Packages 6](#_Toc519494785)

[2.8. A 6](#_Toc519494786)

[3. Comments 7](#_Toc519494787)

[3.1. Copyright 7](#_Toc519494788)

[3.2. Boilerplate 7](#_Toc519494789)

[3.3. Changes 7](#_Toc519494790)

[4. SQL Practices (Suggested vs. Best) 7](#_Toc519494791)

[4.1. Suggestions 7](#_Toc519494792)

[4.1.1. Templates 8](#_Toc519494793)

[5. Data Query Language 8](#_Toc519494794)

[5.1. Grant and Revoke 8](#_Toc519494795)

[5.2. Rollback and Commit 8](#_Toc519494796)

[5.2.2. Whitespaces and Newlines 8](#_Toc519494797)

[5.2.3. Upper and Lower Case 8](#_Toc519494798)

[5.2.4. Aliases 8](#_Toc519494799)

[5.2.5. Comments 9](#_Toc519494800)

[5.2.6. Avoid 9](#_Toc519494801)

[5.2.7. Organization 9](#_Toc519494802)

[5.3. Best 9](#_Toc519494803)

[6. Team 9](#_Toc519494804)

[7. Data Management 9](#_Toc519494805)

[7.1. Authoritative Data Source (ADS) 9](#_Toc519494806)

[7.1.1. Authoritative Data Source Library 10](#_Toc519494807)

[7.2. Data Governance 10](#_Toc519494808)

[7.3. Dimensions 10](#_Toc519494809)

[7.3.1. Completeness 10](#_Toc519494810)

[7.3.2. Uniqueness 10](#_Toc519494811)

[7.3.3. Validity 10](#_Toc519494812)

[7.3.4. Accuracy 10](#_Toc519494813)

[7.3.5. Consistency 10](#_Toc519494814)

[7.3.6. Timeliness 10](#_Toc519494815)

[8. Data Migration 10](#_Toc519494816)

[9. Oracle 11](#_Toc519494817)

[9.1.1. Oracle Database Express Edition 11g Release 2 11](#_Toc519494818)

[9.1.2. SQL Developer Data Modeler 18.1 11](#_Toc519494819)

[9.2. Working Schema Creation Notes 11](#_Toc519494820)

[9.2.1. PDS 11](#_Toc519494821)

[9.2.2. PDS\_AKO 12](#_Toc519494822)

[9.2.3. PDS\_DEV 13](#_Toc519494823)

[10. Business Data Matrix 15](#_Toc519494824)

[10.1. Bus Matrix 15](#_Toc519494825)

[10.2. Roles 16](#_Toc519494826)

[10.3. Metadata 16](#_Toc519494827)

[10.4. Logical Data Map 16](#_Toc519494828)

[10.5. Data Profile 17](#_Toc519494829)

[10.6. Dimensional Model Dev 18](#_Toc519494830)

[10.7. Source System Tracking Report 18](#_Toc519494831)

[10.8. Fixed Length Flat File Layout 19](#_Toc519494832)

[11. Enumeration Reference 19](#_Toc519494833)

[11.1. Theory 19](#_Toc519494834)

[11.2. Instructions 19](#_Toc519494835)

[11.3. Error Messages 24](#_Toc519494836)

[12. Navy VPN Access 25](#_Toc519494837)

[13. References 25](#_Toc519494838)

[14. Footnotes 26](#_Toc519494839)

Figures

[Figure 1 -Bus Matrix 15](#_Toc519494960)

[Figure 2 - Roles 16](#_Toc519494961)

[Figure 3 - Logical Data Map 17](#_Toc519494962)

[Figure 4 - Data Profile 17](#_Toc519494963)

[Figure 5 - Dimensional Model Dev 18](#_Toc519494964)

[Figure 8 - Source System Tracking Report 18](#_Toc519494965)

[Figure 7 - Fixed Length Flat File Layout 19](#_Toc519494966)

[Figure 8 - 20](#_Toc519494967)

[Figure 9 - 20](#_Toc519494968)

[Figure 10 - 21](#_Toc519494969)

[Figure 11 – 21](#_Toc519494970)

[Figure 12 – 21](#_Toc519494971)

[Figure 13 – 22](#_Toc519494972)

[Figure 14 – 22](#_Toc519494973)

[Figure 15 – 22](#_Toc519494974)

[Figure 16 – 22](#_Toc519494975)

[Figure 17 – 23](#_Toc519494976)

[Figure 18 – 23](#_Toc519494977)

[Figure 19 – 23](#_Toc519494978)

[Figure 20 – 24](#_Toc519494979)

# Suggestions

Conventions

Below are some suggested naming conventions for the SQL programming language. They should be followed while developing software in SQL for good maintenance and readability of code. IF you do not a rule, create your own. What is important is that you develop a consistent style. And yes, the idea for these rules was borrowed from Java to develop a measure of consistency.

Camel case: It consists of compound words or phrases such that each word or abbreviation begins with a capital letter or first word with a lowercase letter, rest all with capital.

## Incorporate

* + Use white space and indentation to make code easier to read.
  + Use only standard SQL functions instead of vendor specific functions for reason of portability.
  + Include comments in the code, (not just SQL,) often, you will be doing yourself a favor down the road.

## Avoid

* + Plurals
  + Quoted identifiers

# Naming

## Reserve Words

* + Reserve words should be uppercase.

## Tables and Views

* + Tables names should be nouns, in mixed case with the first letter of each internal word capitalized. Views name should also be capitalized just like table names. When possible, the name should be singular.
  + Use whole words and must avoid acronyms and abbreviations.

## Column/Field Names

* + Column/field names should be descriptive nouns, in camel case. Avoid the generic names like ‘id’ for ‘employeeId’.

## Functions and Procedures

* + Functions and procedures should start with a verb, in mixed case with the first letter lower and with the first letter of each internal word capitalized.
  + Prefix and/or suffix

## Variables

* + Should not start with underscore (‘\_’) or dollar sign (‘$’) characters.
  + Should be mnemonic i.e., designed to indicate to the casual observer the intent of its use.
  + One-character variable names should not be used.
  + Prefixes

## Constants Variables

* + Should be all uppercase with words separated by underscores (“\_”).
  + Prefixes

## Packages

* + The prefix of a unique package name is always written in all-lowercase ASCII letters.
  + Subsequent components of the package name vary according to an organization’s own internal naming conventions.

## A

* + B

# Comments

## Copyright

## Boilerplate

## Changes

# SQL Practices (Suggested vs. Best)

Let me state right up front that anything I propose is a “suggestion” because I am willing to accept that I what I think is “best” will be unacceptable to others. So, I welcome your comments and suggestions.

## Suggestions

If you have been doing SQL development for anytime at all you know that it is repetitious and tedious.

### Templates

#### Data Manipulation Language

# Data Query Language

## Grant and Revoke

## Rollback and Commit

#### Data Define Language

#### Data Control Language

### Whitespaces and Newlines

### Upper and Lower Case

### Aliases

#### Table

#### Column

### Comments

#### Blocks

#### Single Line

### Avoid

#### Be Specific

SELECT \* FROM <table\_name>;

### Organization

## Best

# Team

# Data Management

## Authoritative Data Source (ADS)

DoD 8320.2: Authoritative Source: A source of data or information that is recognized by members of a Community of Interest (COI) to be valid or trusted because it is considered highly reliable or accurate or is from an official publication or reference (e.g., the United States (U.S.) Postal Service is the official source of U.S. mailing ZIP codes). (DoD 8320.2)

### Authoritative Data Source Library

ADS is a web accessed (http://ads.msrr.dmso.mil) set of libraries (DMSO, Army, Navy, Air Force, MEL) of metadata on M&S data and knowledge source. It provides general description, quality, and access information for each source. [DMSO 1039, AF 690, Army 755, Navy 1164]

### ADS Build Checklist

1. Function: Raw\_to\_GUID
2. Table: EnumTypeRef
3. Table: EnumValueRef
4. Table: EnumValueAssocRef

## Data Governance

Data Quality Scorecard

## Dimensions

### Completeness

* The proportion of stored data against the potential of "100% complete".

### Uniqueness

* No thing will be recorded more than once based upon how that thing is identified.

### Validity

* Data are valid if it conforms to the syntax (format, type, range) of its definition.

### Accuracy

* The degree to which data correctly describes the "real world" object or event being described.

### Consistency

* The absence of difference, when comparing two or more representations of a thing against a definition.

### Timeliness

* The degree to which data represent reality from the required point in time.

# Data Migration

* Is there a list of business processes that need to be migrated?
* Is there a list of business interfaces that will be needed?
* Is there a list of reference tables?
* So, what happens to data not migrated get SPS is shut down? (Archiving status?)
* Is any collecting statistics on the number of new and current contracting data? (How is ‘current contracting data’ defined?) At least for on the LD.)
* Are the ePS Governance Board and Data Management Working Group (DMWG) meeting?
* data migration business rules
* Is there an approved list of file type(s) and contract type(s) to be migrated?
* Important contract terms (metadata) must be tracked and reviewed. (Has this been defined?)
* A validation standard for contract documents must be defined and executed.
* Can the data owner(s) produce a list of contracts to be migrated to ePS? Can this list be used to determine what supplemental data needs to be migrated? Do bidder(s) and/or contractor(s) without new or current contracts get migrated to ePS?
* Procurement Data Standard (PDS)
* Purchase Request Data Standard (PRDS)

# Oracle

With any Oracle download there is a “OTN License Agreement” that you have to accept before you can do the actual download. (It is usually found in the top third of the page.)

### Oracle Database Express Edition 11g Release 2

<http://www.oracle.com/technetwork/database/database-technologies/express-edition/downloads/index.html>

### SQL Developer Data Modeler 18.1

<http://www.oracle.com/technetwork/developer-tools/sql-developer/overview/index.html>

sqldeveloper-18.1.0.095.1630-x64.zip

## Working Schema Creation Notes

### PDS

SQL> connect / as sysdba

Connected.

SQL> create user pds identified by pds;

SQL> grant create session to pds;

SQL> drop user pds;

select tablespace\_name from dba\_tablespaces;

create tablespace pds\_tabspace

datafile 'pds\_tabspace.dat'

size 10M autoextend on;

create temporary tablespace pds\_tabspace\_temp

tempfile 'pds\_tabspace\_temp.dat'

size 5M autoextend on;

create user pds

identified by pds

default tablespace pds\_tabspace

temporary tablespace pds\_tabspace\_temp;

grant create session to pds;

grant create table to pds;

grant unlimited tablespace to pds;

GRANT CREATE ANY PROCEDURE TO pds ;

GRANT CREATE VIEW TO pds ;

GRANT CREATE SYNONYM TO pds ;

GRANT CREATE SEQUENCE TO pds ;

GRANT CREATE ANY TRIGGER TO pds ;

### PDS\_AKO

create tablespace pds\_ako\_tabspace

datafile 'pds\_ako\_tabspace.dat'

size 10M autoextend on ;

create temporary tablespace pds\_ako\_tabspace\_temp

tempfile 'pds\_ako\_tabspace\_temp.dat'

size 5M autoextend on ;

create user pds\_ako

identified by pds\_ako

default tablespace pds\_ako\_tabspace

temporary tablespace pds\_ako\_tabspace\_temp ;

grant create session to pds\_ako ;

grant create table to pds\_ako ;

grant unlimited tablespace to pds\_ako ;

GRANT CREATE ANY PROCEDURE TO pds\_ako ;

GRANT CREATE VIEW TO pds\_ako ;

GRANT CREATE SYNONYM TO pds\_ako ;

GRANT CREATE SEQUENCE TO pds\_ako ;

GRANT CREATE ANY TRIGGER TO pds\_ako ;

Connection Name: pds\_ako

pds\_ako

pds\_ako

Type: basic Role: default

Port: 1521

Service Name: XE

### PDS\_DEV

SQL> connect / as sysdba

create tablespace pds\_dev\_tabspace

datafile 'pds\_dev\_tabspace.dat'

size 10M autoextend on ;

create temporary tablespace pds\_dev\_tabspace\_temp

tempfile 'pds\_dev\_tabspace\_temp.dat'

size 5M autoextend on ;

create user pds\_dev

identified by pds\_dev

default tablespace pds\_dev\_tabspace

temporary tablespace pds\_dev\_tabspace\_temp ;

grant create session to pds\_dev ;

grant create table to pds\_dev ;

grant unlimited tablespace to pds\_dev ;

GRANT CREATE ANY PROCEDURE TO pds\_dev ;

GRANT CREATE VIEW TO pds\_dev ;

GRANT CREATE SYNONYM TO pds\_dev ;

GRANT CREATE SEQUENCE TO pds\_dev ;

GRANT CREATE ANY TRIGGER TO pds\_dev ;

Connection Name: pds\_dev

pds\_dev

pds\_dev

Type: basic Role: default

Port: 1521

Service Name: XE

# Business Data Matrix

The workbook contains a collection of eight (8) spreadsheets that are useful in designing and developing data warehouses. They were developed and/or taken from Ralph Kimball’s series of data warehousing books and others. Generally, they can be used with any type of data quality issue with a little imagination.

## Bus Matrix

“The Bus Matrix is a key design tool representing the organization’s core business processes and associated dimensionality. It’s the architectural blueprint providing the top-down strategic perspective to ensure data in the DW/BI environment can be integrated across the enterprise, while agile bottom-up delivery occurs by focusing on a single business process at a time.” (1)



Figure -Bus Matrix

## Roles

“A single physical dimension can be referenced multiple times in a fact table, with each reference linking to a logically distinct role for the dimension. For instance, a fact table can have several dates, each of which is represented by a foreign key to the date dimension. It is essential that each foreign key refers to a separate view of the date dimension so that the references are independent. These separate dimension views (with unique attribute column names) are called roles.” (1)

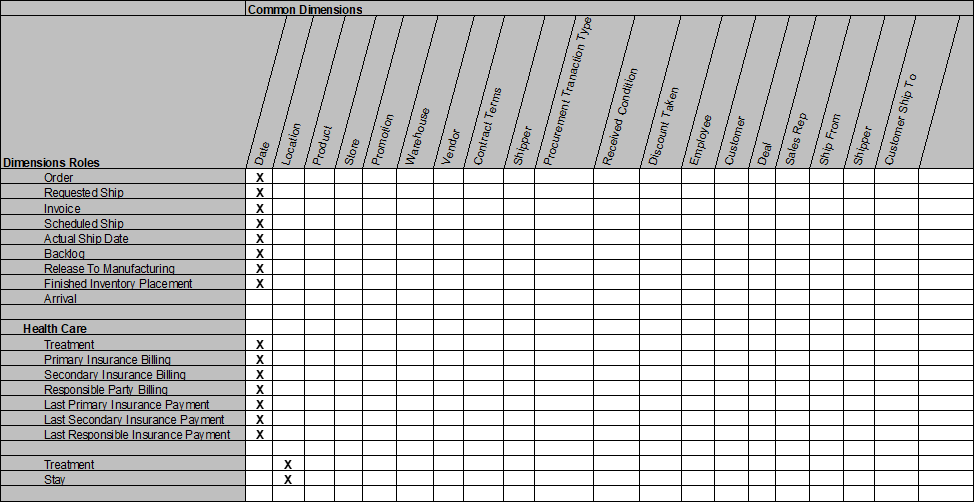


Figure - Roles

## Metadata

Metadata is all the information that defines and describes the structures, operations, and contents of the DW/BI system. Technical metadata defines the objects and processes which comprise the DW/BI system. Business metadata describes the data warehouse contents in user terms, including what data is available, where did it come from, what does it mean, and how does it relate to other data. Finally, process metadata describes the warehouse’s operational results.” (2)

## Logical Data Map

“The logical data map is a document that shows how source fields map to target fields within an ETL process.

The logical data map (LDM) is a high-level specification for the ETL team or developer and thus should exist before ETL development starts. It can be produced once the data warehouse schema is close to finalization and once source systems have been identified to feed the data warehouse. In addition to providing a business specification for the ETL development, it can also be used by QA to help validate the ETL.” (3)

“It is a document which shows the relationship between the primary source and the final point and it is usually presented in a form of tables or a spreadsheet format. The logical data map (called sometimes the lineage report) includes:” (5)

* Components of the target: target table name, column name, table type (fact, dimension or subdimension), data type (ex: a number) and SCD type
* Components of the source: the source database, table name, column name, data type
* Transformation



Figure - Logical Data Map

## Data Profile

“Data profiling is the systematic up-front analysis of the content of a data source, all the way from counting the bytes and checking cardinalities up to the most thoughtful diagnosis of whether the data can meet the high-level goals of the data warehouse.

Data profiling practitioners divide this analysis into a series of tests, starting with individual fields and ending with whole suites of tables comprising extended databases. Individual fields are checked to see that their contents agree with their basic data definitions and domain declarations. It is especially valuable to see how many rows have null values, or have contents that violate the domain definition.” (4)



Figure - Data Profile

The scripts below are written for Oracle DBMS. Similar functionality is available from most database providers.

The script provides most of the data profiling data that might be needed, including comments on the column.

SELECT atc.owner,

atc.table\_name,

atc.column\_name,

atc.column\_id AS "Ordinal",

atc.data\_type,

atc.data\_length,

atc.data\_precision,

atc.num\_nulls,

acc.comments,

atc.data\_default,

atc.low\_value,

atc.high\_value

-- , atc.\*

FROM all\_tab\_cols atc

LEFT OUTER JOIN all\_col\_comments acc ON atc.table\_name = acc.table\_name

AND atc.column\_name = acc.column\_name

WHERE atc.owner = 'PDS'

-- AND atc.table\_name = ''

ORDER BY atc.owner,

atc.table\_name,

atc.column\_id;

This script shown below with provide a list of “distinct” column values, and the number of occurrences of the value.

SELECT DISTINCT enum\_type,

COUNT(rec\_id) AS "Rec Cnt"

FROM enum\_value\_ref

GROUP BY enum\_type

ORDER BY enum\_type;

## Dimensional Model Dev



Figure - Dimensional Model Dev

## Source System Tracking Report

Contains useful information regarding the sources data for the warehouse. Can be used as input to calculating processing volumes and estimating warehouse growth.



Figure - Source System Tracking Report

## Fixed Length Flat File Layout

A description of the positional layout of data in a flat text file. (6)



Figure - Fixed Length Flat File Layout

# Enumeration Reference

## Theory

Metadata

## Instructions

1. Duplicate the **Enum\_Ref\_Create** workbook and rename as appropriate.
2. Rename the **<source>** tab to reflect the data source.

This is an optional step, load the source data into the <source> tab.

1. Open the **<Blank>** tab and replace **<appl\_name>** value in in B2 with the name of the data source. (Note: this eliminates the need to update the cell every
2. Insert a new tab after the **<blank>** tab.

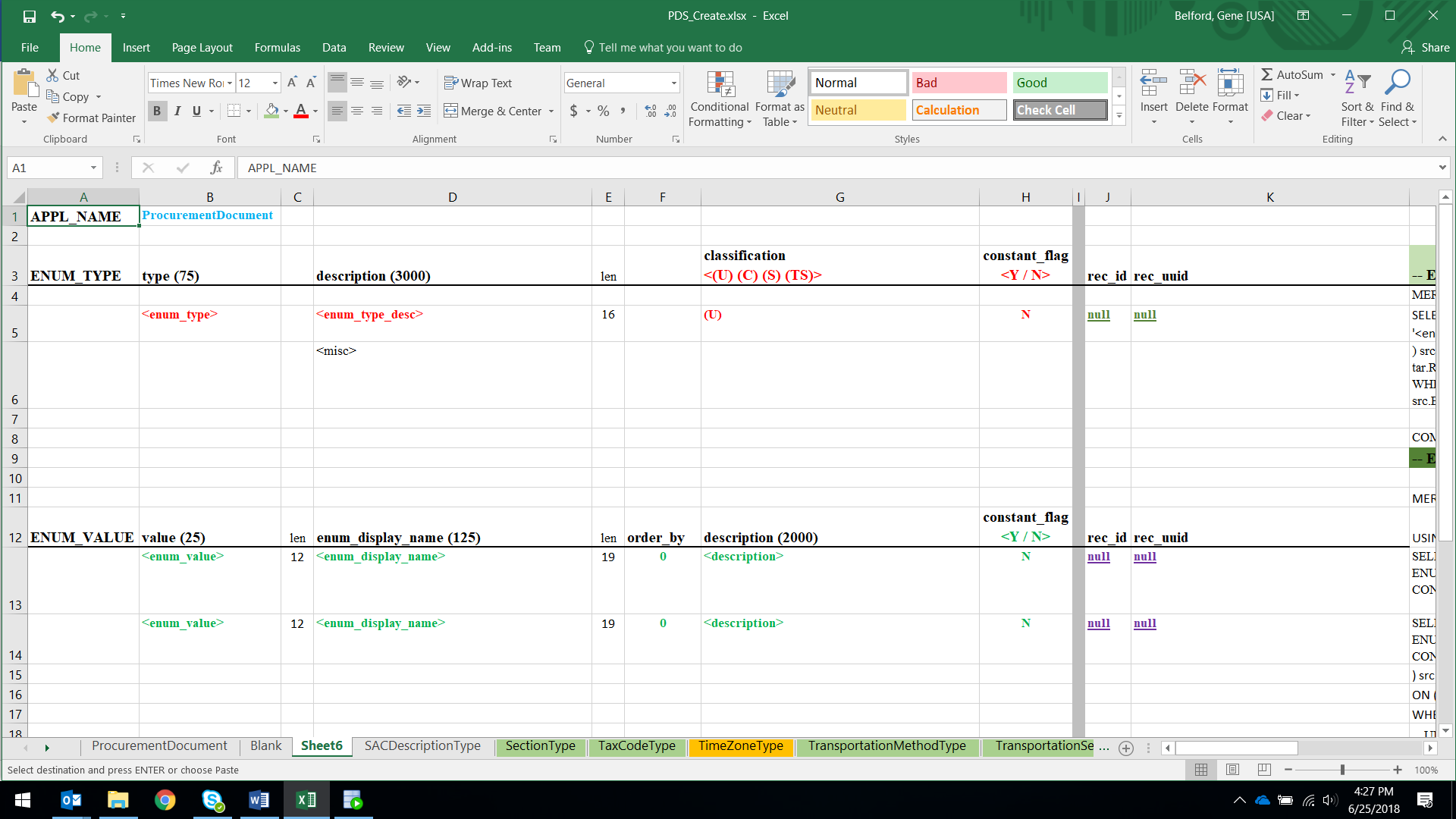


Figure -

1. Select and copy the whole “Blank” tab and paste it in cell A1 in the **<Ref\_1>** tab.
2. From the **<source>** tab identify the name for the new reference. Rename the **<Sheet#>** tab with that name.
3. ENUM\_TYPE



Figure -

Enter that name into the cell B5 **<enum\_type>**.

Update cell C5 **<enum\_type\_desc>** with a description for the reference. If there is a description in the source use that. If not either duplicate the name, or provide your own description.

Select a classification level in cell F5 **<(U) (C) (S) (TS)>**. This is optional and will default to (U) if left blank.

Set the constant flag in cell G5 **<Y / N>.** This is optional and will default to (N) if left blank.

Highlight and select the ENUM\_TYPE merge script, cells K4:K8.

Paste in IDE script window. Replace the double quotes (“) with a blank.

Execute. (A new entry should be made in the ENUM\_TYPE\_REF table.

Retrieve the new <REC\_ID> and <REC\_UUID> for the new row created.

Replace the respective values in cells I5 and J5.

{ Save as necessary }

1. ENUM\_VALUE



Figure -

A



Figure –



Figure –

A

{ Save as necessary }



Figure –

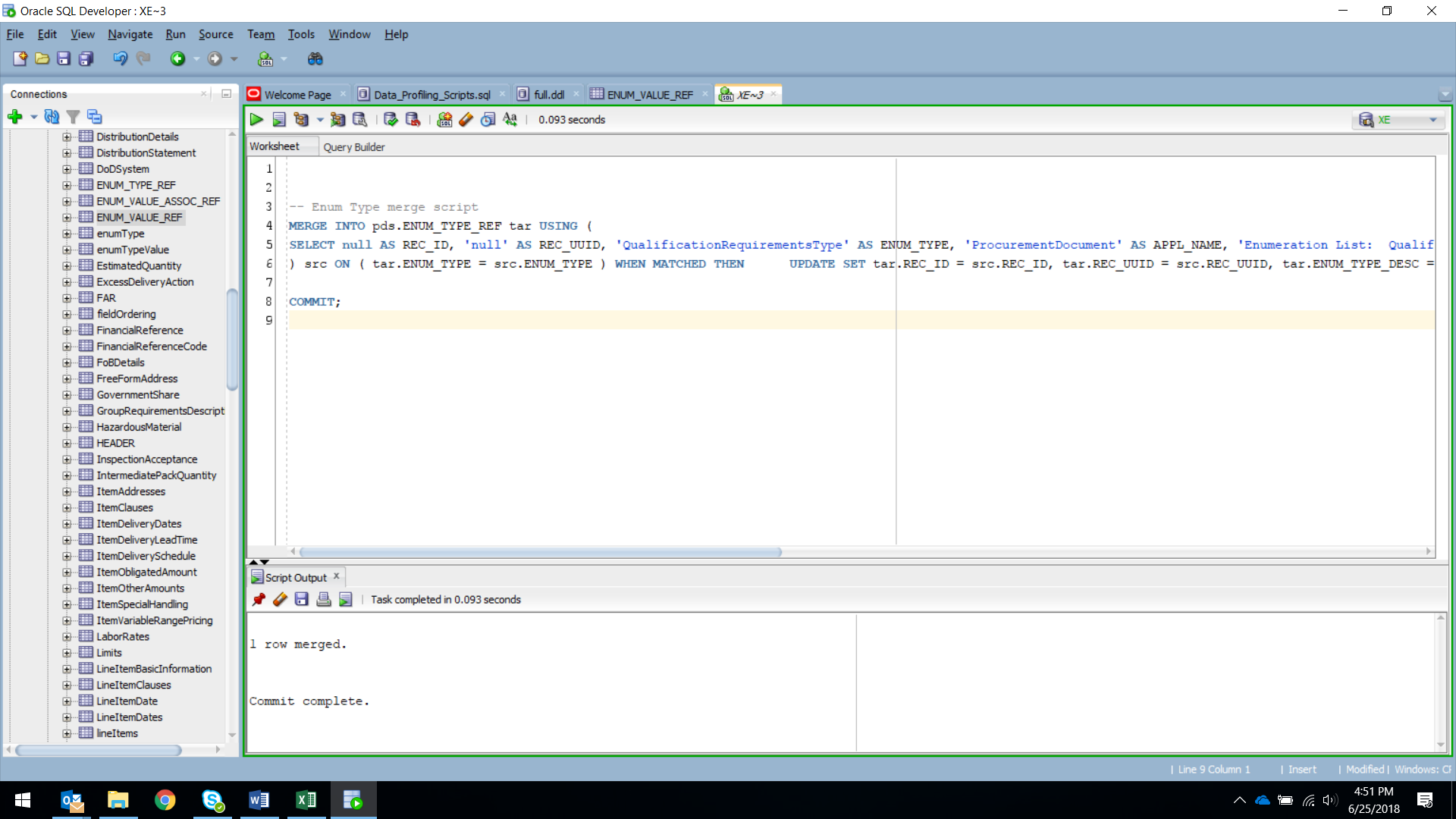


Figure –

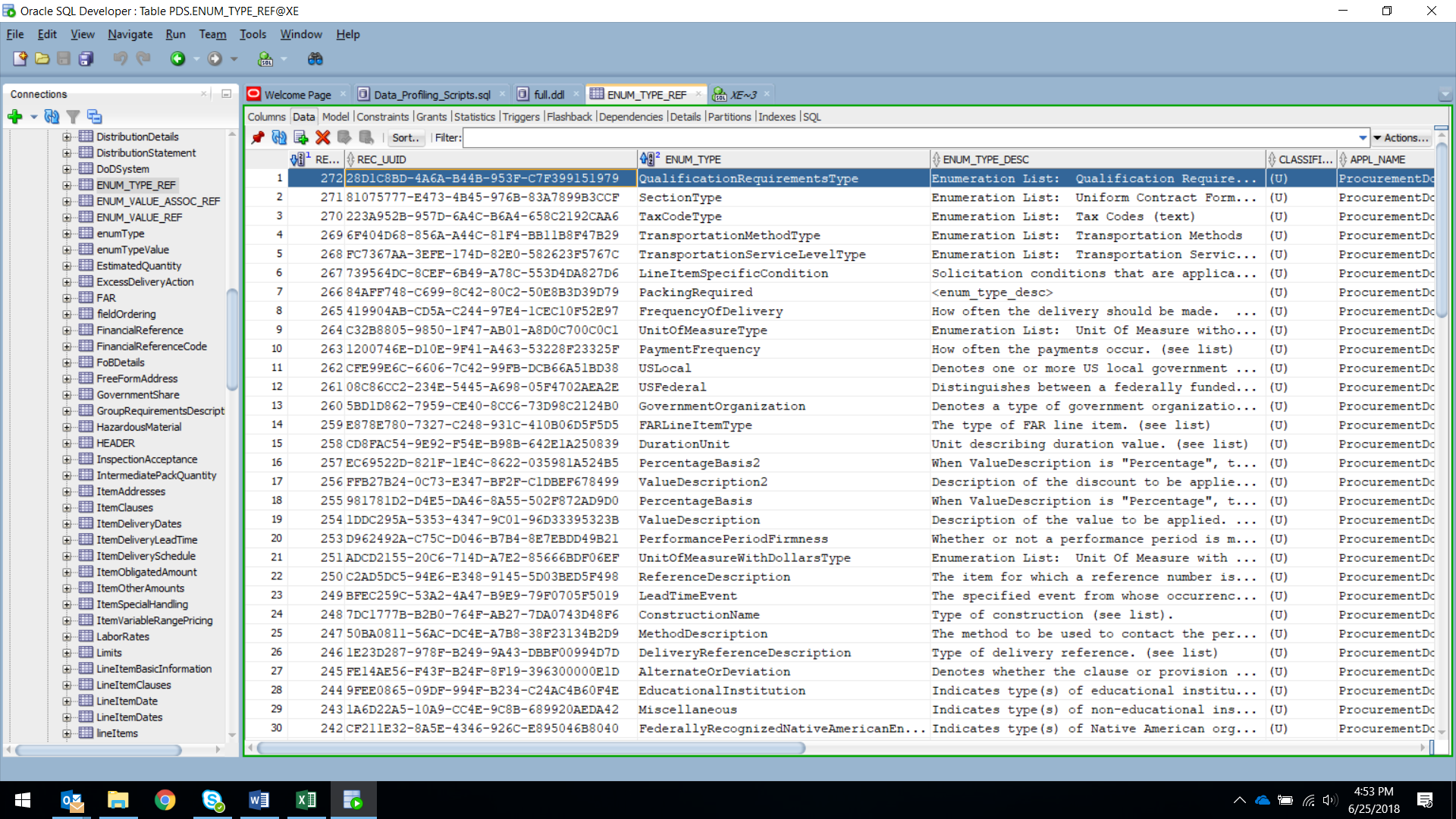


Figure –



Figure –



Figure –

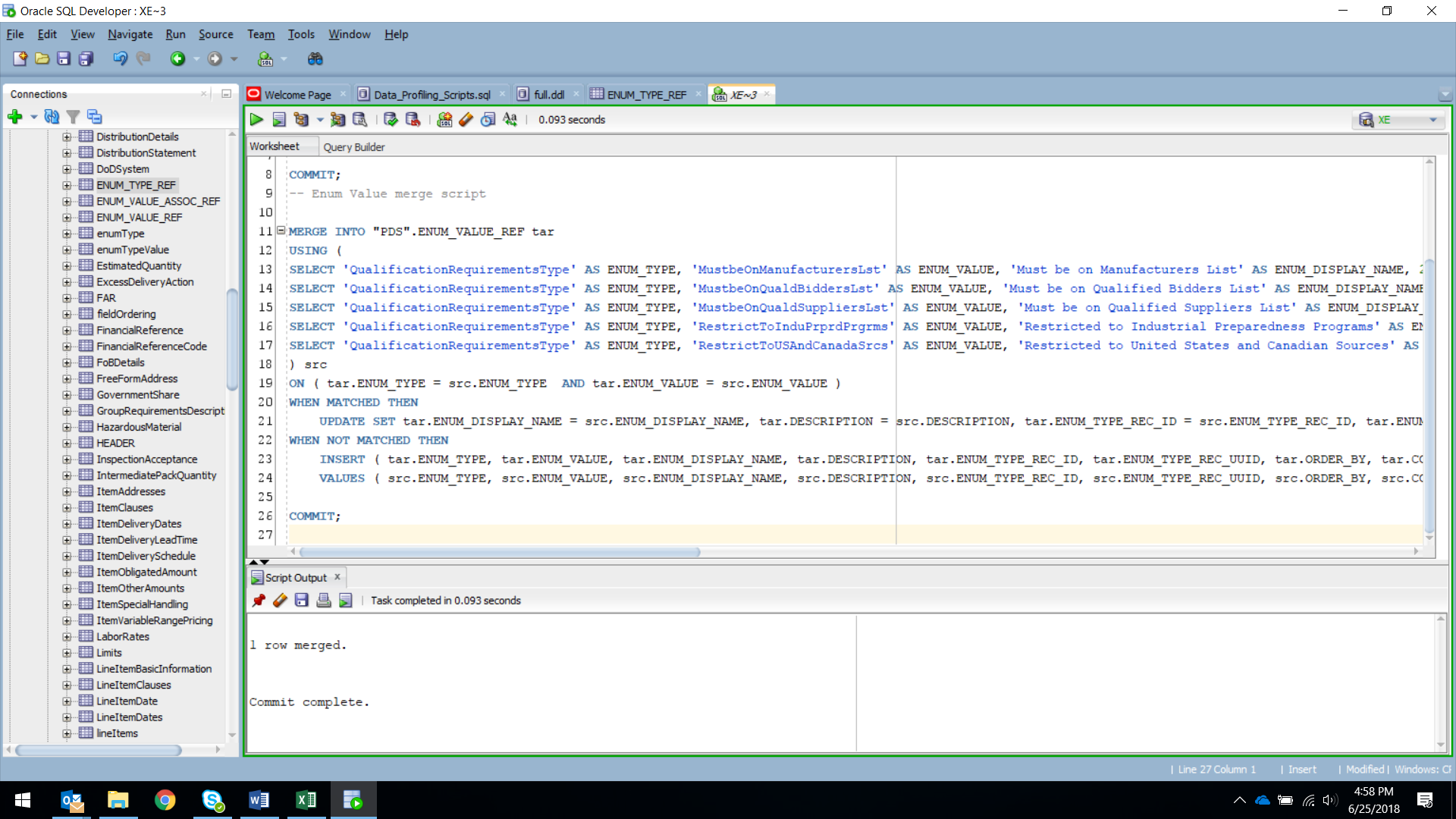


Figure –

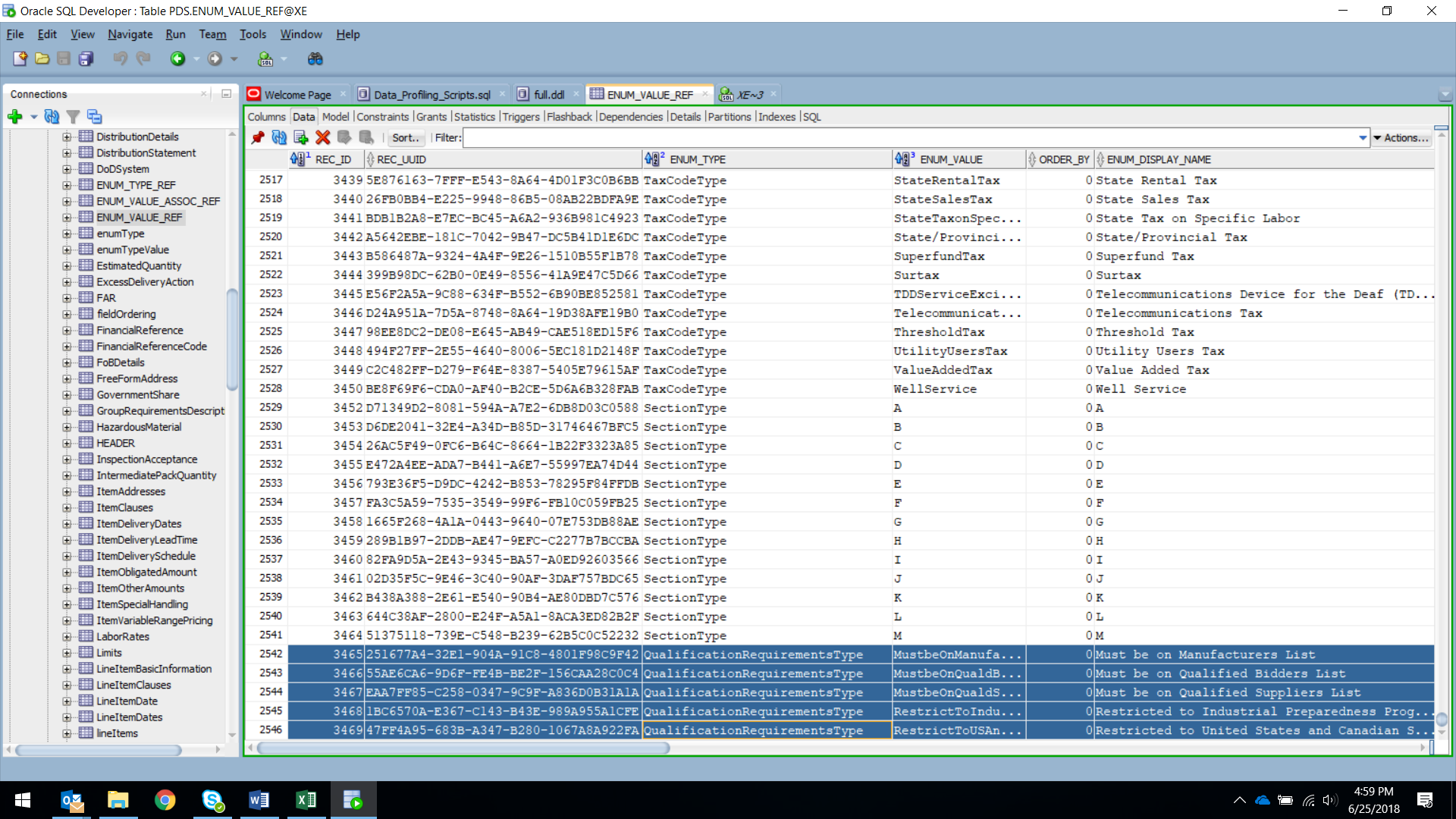


Figure –

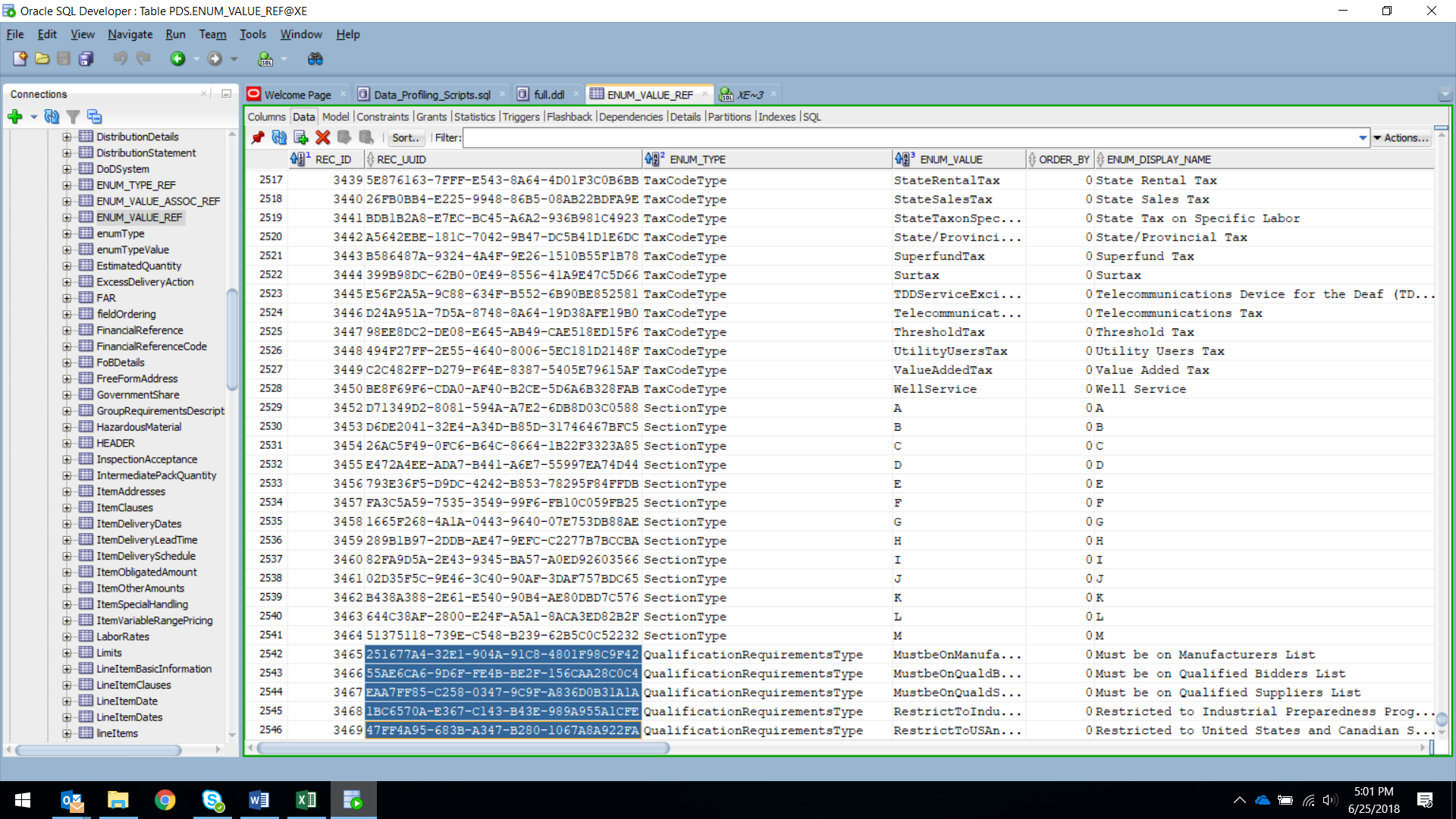


Figure –

## Error Messages

Authoritative Data Source

# e-mail

* @?/rdbms/admin/utlmail.sql
* @?/rdbms/admin/prvtmail.plb
* @?/rdbms/admin/utlsmtp.sql
* @?/rdbms/admin/utltcp.sql
* GRANT EXECUTE ON utl\_mail TO prds\_dev;
* GRANT EXECUTE ON dbms\_network\_acl\_admin TO prds\_dev;
* GRANT SELECT ON dba\_network\_acl\_privileges TO prds\_dev;
* GRANT SELECT ON dba\_network\_acls TO prds\_dev;

# Navy VPN Access

<https://guestmanager.bah.com/sponsorportal/PortalSetup.action?portal=0aa72342-6ee5-11e5-a1bf-e4c722f2fa9a>

To establish VPN Connectivity:

1. In the Notification Area of the Windows taskbar, right-click the RAS Tools icon
2. Under Tools, click Enable Hotspot Mode
3. Select a wi-fi network from the Wireless Network Connections dialog box and click Connect
4. Open your web browser to confirm you have Internet access
5. Select your DoD EMAIL CAC certificate and click OK
6. Type your CAC PIN when prompted and click OK
7. If the Setup Control-Warning box appears, click Always
8. When the ‘You may begin working’ screen appears, your SSL-VPN connection is established

To terminate VPN Connectivity:

To disconnect from a wi-fi hotspot, right-click the wireless network connection icon in the Notification Area of the Windows taskbar, select the wi-fi network, and click Disconnect.

# References

<https://spssite.caci.com>

<https://www.acq.osd.mil/dpap/pdi/docs/p2p%20training%20presentations/NavyEPS_Placemat_final_05.24.17_v9.pdf>

<https://www.dau.mil/library/defense-atl/blog/How-the--Electronic--Procurement-System--Ate-the-Acquisition--Elephant>

<http://www.i2insights.com/library/defense_budget-documents/fy2017-defense_budget/1319N/1319N-0605013N-R.pdf>

<http://www.navair.navy.mil/nawctsd/Resources/Library/Acqguide/sps.htm>

<http://www.secnav.navy.mil/rda/OneSource/Pages/eBusiness/DON%20eBusiness%20Solutions/SPS.aspx>

<http://www.caci.com/business/sps_prog.shtml>

<https://www.omicsonline.org/open-access/why-data-profiling-and-how-2277-1891.1000118.pdf>

<http://www.etltools.org/extraction.html>

# Footnotes

(1) The Data Warehouse Toolkit, 3rd Edition (Wiley, 2013)

**(2)** The Data Warehouse Lifecycle Toolkit, 2nd Edition (Wiley, 2008)

**(3)** Andrew Smith, <http://www.andrewleesmith.co.uk/>

**(4)** Kimball Group, <http://www.kimballgroup.com/wp-content/uploads/2012/05/DT59SurprisingValue.pdf>

**(5)** ETL Tools - ETL and Data Warehousing Portal, <http://www.etltools.org/extraction.html>

**(6)** The Data Warehouse ETL Toolkit, (Wiley, 2004)