Data Quality

TASK 7 REPORT

CONTENTS

DWH TESTING AND BUGS CREATION

BUG 1

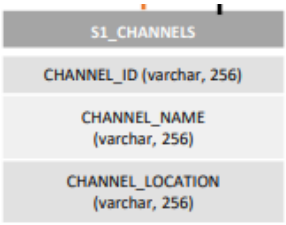
Summary Invalid column name "CHANNELLLOCATION" in the LND_LAYER.S1_CHANNELS

Test Rail Test Case: LINK

Priority: Major

Severity: Trivial

Description: Invalid column name "CHANNELLLOCATION" in the LND_LAYER.S1_CHANNELS table that does not match the provided column name "CHANNEL_LOCATION" according to DWH_schema.pdf document.



Screenshot the rows with invalid column name:

	COLUMN_NAME	DATA_TYPE
1	CHANNEL_ID	VARCHAR2(256 BYTE)
2	CHANNEL_NAME	VARCHAR2(256 BYTE)
3	CHANNELLOCATION	VARCHAR2(256 BYTE)

Steps to reproduce:

1. Login to database as LND_LAYER user. S1_CHANNELS table exists.
2. Execute following query:

```
SELECT    all_tab_cols.COLUMN_NAME , all_tab_cols.DATA_TYPE , all_tab_cols.DATA_LENGTH , all_tab_cols.NULLABLE

FROM all_tab_cols

WHERE all_tab_cols.TABLE_NAME='S1_CHANNELS'

AND all_tab_cols.owner = 'LND_LAYER';
```

Expected Result: CHANNEL_ID, CHANNEL_NAME, CHANNEL_LOCATION

Actual Result: Script from step 2 return following rows

	COLUMN_NAME	DATA_TYPE
1	CHANNEL_ID	VARCHAR2(256 BYTE)
2	CHANNEL_NAME	VARCHAR2(256 BYTE)
3	CHANNELLOCATION	VARCHAR2(256 BYTE)

So it is invalid column name "CHANNELLLOCATION" in the LND_LAYER.S1_CHANNELS.
RCA: ALTER TABLE LND_LAYER.S1_CHANNELS RENAME COLUMN CHANNELLOCATION to CHANNEL_LOCATION

BUG 2

Summary Incorrect Data lenght in LND_LAYER.S1_PRODUCTS
Test Rail Test Case: LINK

Priority: Major

Severity: Trivial

Description: LND_LAYER.S1_PRODUCTS table contain incorrect data length, which is incorrect according to the DWH_SCHEMA.pdf and DWH_schema_description.docx documents

S1_PRODUCTS
PRODUCT_ID (varchar, 256)
COST (varchar, 256)
PRODUCT_NAME (varchar, 256)

Screenshot the rows with invalid data length:

	COLUMN_NAME	DATA_TYPE	DATA_LENGTH	NULLABLE
1	PRODUCT_ID	VARCHAR2	240	Y
2	COST	VARCHAR2	240	Y
3	PRODUCT_NAME	VARCHAR2	240	Y

Steps to reproduce:

1. Login to database as LND_LAYER user. S1_PRODUCTS table exists.
2. Execute following query:

```
SELECT  all_tab_cols.COLUMN_NAME , all_tab_cols.DATA_TYPE , all_tab_cols.DATA_LENGTH , all_tab_cols.NULLABLE

FROM all_tab_cols

WHERE all_tab_cols.TABLE_NAME='S1_PRODUCTS'

AND all_tab_cols.owner = 'LND_LAYER';
```

Expected Result: data_length = 256

Actual Result: Script from step 2 return following rows

	COLUMN_NAME	DATA_TYPE	DATA_LENGTH	NULLABLE
1	PRODUCT_ID	VARCHAR2	240	Y
2	COST	VARCHAR2	240	Y
3	PRODUCT_NAME	VARCHAR2	240	Y

RCA:

```
ALTER TABLE LND_LAYER.S1_PRODUCTS
        MODIFY PRODUCT_ID VARCHAR2(256);

ALTER TABLE LND_LAYER.S1_PRODUCTS
MODIFY COST VARCHAR2(256);

ALTER TABLE LND_LAYER.S1_PRODUCTS
            MODIFY PRODUCT_NAME VARCHAR2(256);
```

BUG 3

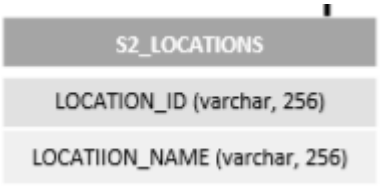
Summary Incorrect Data lenght in LND_LAYER.S2_LOCATIONS

Testrail Test Case: LINK

Priority: Major

Severity: Trivial

Description: Table contain incorrect data length, which is incorrect according to the DWH_SCHEMA.pdf and DWH_schema_description.docx documents



Screenshot the rows with invalid data length:

	COLUMN_NAME	DATA_TYPE	DATA_LENGTH	NULLABLE
1	LOCATION_ID	VARCHAR2	230	Y
2	LOCATION_NAME	VARCHAR2	230	Y

Steps to reproduce:

Login to database as LND_LAYER user. S1_CHANNELS table exists.
Execute following query:

```
SELECT    all_tab_cols.COLUMN_NAME , all_tab_cols.DATA_TYPE , all_tab_cols.DATA_LENGTH , all_tab_cols.NULLABLE

FROM all_tab_cols

WHERE all_tab_cols.TABLE_NAME='S2_LOCATIONS'

AND all_tab_cols.owner = 'LND_LAYER';
```

Expected Result: data_length = 256

Actual Result: Script from step 2 return following rows

	COLUMN_NAME	DATA_TYPE	DATA_LENGTH	NULLABLE
1	LOCATION_ID	VARCHAR2	230	Y
2	LOCATION_NAME	VARCHAR2	230	Y

RCA:

```
ALTER TABLE LND_LAYER.S2_LOCATIONS
        MODIFY LOCATION_ID VARCHAR2(256);

ALTER TABLE LND_LAYER.S2_LOCATIONS
              MODIFY LOCATION_NAME VARCHAR2(256);
```

BUG 4

Summary 1 Column Affected. Incorrect Column Name in DWH_LAYER.DWH_CLIENTS
Testrail Test Case: LINK

Priority: Major

Severity: Trivial

Description: Table containcolumn with incorrect anme “FIRST_PURRCHASE_DATE”, which is incorrect according to the DWH_SCHEMA.pdf and DWH_schema_description.docx documents

DWH_CLIENTS
CLIENT_ID (int) PK
CLIENT_SRC_ID (varchar, 10)
FIRST_NAME (varchar, 100)
MIDDLE_NAME (varchar, 100)
LAST_NAME (varchar, 100)
EMAIL (varchar, 50)
PHONE_NUMBER (varchar, 100)
VALID_FROM (DATE)
VALID_TO (DATE)
IS_VALID (varchar, 1)
FIRST_PURCHASE_DATE (DATE)

Screenshot the rows with invalid name:

10	IS_VALID	CHAR
11	FIRST_PURRCHASE_DATE	DATE

Steps to reproduce:

Login to database as dwh_LAYER user.DWH_CLIENTS table exists.
Execute following query:

```
SELECT

      all_tab_cols.column_name,   all_tab_cols.data_type, all_tab_cols.data_length,  all_tab_cols.data_precision,
all_tab_cols.nullable

FROM    all_tab_cols

WHERE   all_tab_cols.table_name = 'DWH_CLIENTS'

AND     all_tab_cols.owner = 'DWH_LAYER';
```

Expected Result: 'FIRST_PURCHASE_DATE’

Actual Result: Script from step 2 return following rows

10	IS_VALID	CHAR
11	FIRST_PURRCHASE_DATE	DATE

RCA:
ALTER TABLE DWH_LAYER.DWH_CLIENTS RENAME COLUMN FIRST_PURRCHASE_DATE to FIRST_PURCHASE_DATE

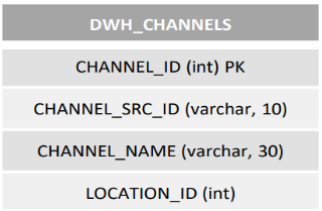
BUG 5

Summary : 2 Columns Affected. Incorrect Data Type and Constraint Type in DWH_LAYER.DWH_CHANNELS
Testrail Test Case: LINK
Priority: Major

Severity: Trivial

Description:

Column CHANNEL_NAME has data length 50 instead of 30.
Column LOCATION_ID has not “NOT NULL” constraint, takes Nullable as Y and is also a Foreign Key to the DWH_LOCATIONS table, which is incorrect according to the DWH_SCHEMA.pdf and DWH_schema_description.docx documents:



Steps to reproduce:

1. Login to database as dwh_LAYER user.DWH_CHANNELS table exists.
2. Execute following query:

```
SELECT  all_tab_cols.column_name,  all_tab_cols.data_type,  all_tab_cols.data_length,
all_tab_cols.data_precision,  all_tab_cols.nullable

FROM  all_tab_cols

WHERE  all_tab_cols.table_name = 'DWH_CHANNELS'

  AND all_tab_cols.owner = 'DWH_LAYER';
```

Expected Result:

column_name	data_type	data_length	nullable
CHANNEL_ID	Int		N
CHANNEL_SRC_ID	VARCHAR	10	Y
CHANNEL_NAME	VARCHAR	30	Y
LOCATION_ID	Int		N

Actual Result:

	COLUMN_NAME	DATA_TYPE	DATA_LENGTH	DATA_PRECISION	NULLABLE
1	CHANNEL_ID	NUMBER	22	(null)	N
2	CHANNEL_SRC_ID	VARCHAR2	10	(null)	Y
3	CHANNEL_NAME	VARCHAR2	50	(null)	Y
4	LOCATION_ID	NUMBER	22	(null)	Y

RCA: ALTER TABLE DWH_LAYER.DWH_CHANNELS MODIFY CHANNEL_NAME VARCHAR2(30);

ALTER TABLE DWH_LAYER. DWH_CHANNELS MODIFY (LOCATION_ID NOT NULL);

BUG 6

Summary : 2 Columns Affected. Incorrect Column Name and Precision in DM_LAYER.DM_MAIN_DASHBOARD
Testrail Test Case: LINK
Priority: Major

Severity: Trivial

Description: Column TOTAL_COST has data precision 16 instead of 18.
Column CLIENT_IAST_NAME _ID is incorrect according to the DWH_SCHEMA.pdf and DWH_schema_description.docx documents:

DM_MAIN_DASHBOARD
ID (int), PK
CLIENT_FIRST_NAME (varchar, 100)
CLIENT_LAST_NAME (varchar, 100)
CLIENT_MIDDLE_NAME (varchar, 70)
EMAIL (varchar, 50)
PHONE_NUMBER (varchar, 100)
CHANNEL_NAME (varchar, 30)
LOCATION_NAME (varchar, 100)
PRODUCT_NAME (varchar, 150)
TOTAL_COST (number, 18,4)

Steps to reproduce:

1. Connect to DM_LAYER
2. EXECUTE CURRENT S

```
SELECT    all_tab_cols.COLUMN_NAME   , all_tab_cols.DATA_TYPE  , all_tab_cols.DATA_LENGTH ,
all_tab_cols.DATA_PRECISION , all_tab_cols.NULLABLE

FROM all_tab_cols

WHERE all_tab_cols.TABLE_NAME='DM_MAIN_DASHBOARD'  AND all_tab_cols.owner = 'DM_LAYER';
```

Expected Result:

COLUMN_NAME	DATA_TYPE	DATA_LENGTH	DATA_PI
ID	NUMBER	22	(null)
CLIENT_FIRST_NAME	VARCHAR	100	(null)
CLIENT_LAST_NAME	VARCHAR	100	(null)
CLIENT_MIDDLE_NAME	VARCHAR	70	(null)
EMAIL	VARCHAR	50	(null)
PHONE_NUMBER	VARCHAR	100	(null)
CHANNEL_NAME	VARCHAR	30	(null)
LOCATION_NAME	VARCHAR	150	(null)
TOTAL_COST	NUMBER	18	4

Actual Result:

	COLUMN_NAME	DATA_TYPE	DATA_LENGTH	DATA_PRECISION
1	ID	NUMBER	22	(null)
2	CLIENT_FIRST_NAME	VARCHAR2	100	(null)
3	CLIENT_IAST_NAME	VARCHAR2	100	(null)
4	CLIENT_MIDDLE_NAME	VARCHAR2	70	(null)
5	EMAIL	VARCHAR2	50	(null)
6	PHONE_NUMBER	VARCHAR2	100	(null)
7	CHANNEL_NAME	VARCHAR2	30	(null)
8	LOCATION_NAME	VARCHAR2	100	(null)
9	PRODUCT_NAME	VARCHAR2	150	(null)
10	TOTAL_COST	NUMBER	22	16

RCA:

ALTER TABLE DM_LAYER.DM_MAIN_DASHBOARD MODIFY TOTAL_COST NUMBER(18,4)

ALTER TABLE DM_LAYER.DM_MAIN_DASHBOARD RENAME COLUMN CLIENT_IAST_NAME to CLIENT_LAST_NAME

BUG 7

Summary :11 Duplicated Rows in LND_LAYER.S1_CLIENTS
Testrail Test Case: LINK
Priority: Major

Severity: Trivial

Description: 11 extra rows in LND_LAYER.S1_CLIENTS is incorrect according to the DWH_SCHEMA.pdf and DWH_schema_description.docx documents.

Steps to reproduce:

3. Connect to LND_LAYER
4. EXECUTE CURRENT SCRIPT
   SELECT CLIENT_ID FROM S2_CLIENTS GROUP BY CLIENT_ID HAVING COUNT(CLIENT_ID)>1;

Expected Result: Query returns 0 rows.

Actual Result: 11 rows returned after executing the script

	CLIENT_ID
1	244
2	241
3	247
4	240
5	250
6	242
7	243
8	245
9	246
10	248
11	249

BUG 8

Summary : 5 Duplicated Rows in LND_LAYER.S2_LOCATIONS
Testrail Test Case: LINK
Priority: Major

Severity: Trivial

Description: 5 extra rows in LND_LAYER.S2_LOCATIONS table, it is incorrect according to the DWH_SCHEMA.pdf and DWH_schema_description.docx documents:

> In source and landing layers all columns should have datatype varchar (256), no constraints are presented. Data is loaded from source to loading 1 to 1, w/o any transformations, validations and so on.

Steps to reproduce:

5. Connect to LND_LAYER
6. EXECUTE CURRENT SCRIPT

```
SELECT LOCATION_ID FROM S2_LOCATIONS GROUP BY LOCATION_ID HAVING COUNT(LOCATION_ID)>1;
```

Expected Result: 0 rows

Actual Result:

	LOCATION_ID
1	33
2	35
3	34
4	36
5	37

BUG 9

Summary : 176 Unloaded Rows in LND_LAYER.S1_SALES
Testrail Test Case: LINK
Priority: Major

Severity: Trivial

Description: 176 rows in CLIENT_DB.S1_SALES table aren't loaded, it is incorrect according to the DWH_SCHEMA.pdf and DWH_schema_description.docx documents:

> In source and landing layers all columns should have datatype varchar (256), no constraints are presented. Data is loaded from source to loading 1 to 1, w/o any transformations, validations and so on.

Steps to reproduce:

7. Connect to LND_LAYER
8. EXECUTE CURRENT SCRIPT:

```
SELECT COUNT(*)

FROM LND_LAYER.S1_SALES

UNION

SELECT COUNT(*)

FROM CLIENT_DB.S1_SALES;
```

Expected Result: equal count

Actual Result:

	COUNT(*)
1	22224
2	22400

BUG 10

Summary : 5 Unloaded Rows in LND_LAYER.S2_CLIENT_SALES
Testrail Test Case: LINK
Priority: Major

Severity: Trivial

Description: 5 Unloaded Rows in LND_LAYER.S2_CLIENT_SALES table, it is incorrect according to the DWH_SCHEMA.pdf and DWH_schema_description.docx documents:

> In source and landing layers all columns should have datatype varchar (256), no constraints are presented. Data is loaded from source to loading 1 to 1, w/o any transformations, validations and so on.

Steps to reproduce:

9. Connect to LND_LAYER
10. EXECUTE CURRENT SCRIPT:

```
SELECT COUNT(*)

FROM LND_LAYER.S2_CLIENT_SALES

UNION

SELECT COUNT(*)

FROM CLIENT_DB_2.S2_CLIENT_SALES;
```

Expected Result: equal count

Actual Result:

	COUNT(*)
1	1291
2	1296

BUG 11

Summary : Incorrect Constraint. 3 Columns Affected in DWH_LAYER.DWH_SALES Table.
Testrail Test Case: LINK
Priority: Major
Severity: Trivial

Description: CLIENT_ID, CHANNEL_ID, PRODUCT_ID must have a “NOT NULL” constraint, it is incorrect according to the DWH_SCHEMA.pdf and DWH_schema_description.docx documents:

All columns in DWH are nullable, for the exception of PK/FK columns.

Steps to reproduce:

11. Logged in as user DWH_LAYER
12. EXECUTE CURRENT SCRIPT:
    SELECT all_tab_cols.COLUMN_NAME, all_tab_cols.DATA_TYPE, all_tab_cols.DATA_LENGTH,
    all_tab_cols.DATA_PRECISION, all_tab_cols.NULLABLE
    FROM all_tab_cols
    WHERE all_tab_cols.TABLE_NAME='DWH_SALS' AND all_tab_cols.owner = 'DWH_LAYER';

Expected Result: equal count

Actual Result:

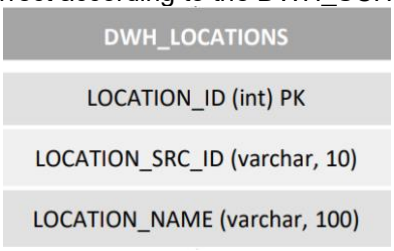
	COUNT(*)
1	1291
2	1296

BUG 12

Summary : Incorrect Data Type in DWH_LAYER.DWH_LOCATIONS. 1 Column Affected

Testrail Test Case: LINK
Priority: Major
Severity: Trivial

Description: In table DWH_LAYER.DWH_LOCATIONS Column LOCATION_NAME has data type VARCHAR with 256 bytes it's incorrect according to the DWH_SCHEMA.pdf document:



Steps to reproduce:

13. Logged in as user DWH_LAYER
14. EXECUTE CURRENT SCRIPT:
    SELECT all_tab_cols.column_name, all_tab_cols.data_type, all_tab_cols.data_length, all_tab_cols.data_precision, all_tab_cols.nullable
    FROM all_tab_cols
    WHERE all_tab_cols.table_name = 'DWH_LOCATIONS'

Expected Result:

column_name	data_type	data_length	nullable
LOCATION_ID	Int		N
LOCATION_SRC_ID	VARCHAR	10	Y
LOCATION_NAME	VARCHAR	100	Y

Actual Result:

	COLUMN_NAME	DATA_TYPE	DATA_LENGTH	DATA_PRECISION
1	LOCATION_ID	NUMBER	22	(null)
2	LOCATION_SRC_ID	VARCHAR2	10	(null)
3	LOCATION_NAME	VARCHAR2	256	(null)

RCA:

ALTER TABLE DWH_LAYER. DWH_LOCATIONS

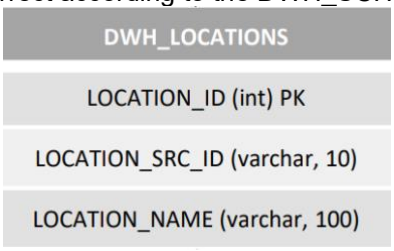
MODIFY LOCATION_NAME VARCHAR2(100);

BUG 13

Summary : Incorrect Data Type in DWH_LAYER.DWH_LOCATIONS. 1 Column Affected

Testrail Test Case: LINK
Priority: Major
Severity: Trivial

Description: In table DWH_LAYER.DWH_LOCATIONS Column LOCATION_NAME has data type VARCHAR with 256 bytes it's incorrect according to the DWH_SCHEMA.pdf document:



Steps to reproduce:

15. Logged in as user DWH_LAYER
16. EXECUTE CURRENT SCRIPT:
    SELECT all_tab_cols.column_name, all_tab_cols.data_type, all_tab_cols.data_length, all_tab_cols.data_precision, all_tab_cols.nullable
    FROM all_tab_cols
    WHERE all_tab_cols.table_name = 'DWH_LOCATIONS'

Expected Result:

column_name	data_type	data_length	nullable
LOCATION_ID	Int		N
LOCATION_SRC_ID	VARCHAR	10	Y
LOCATION_NAME	VARCHAR	100	Y

Actual Result:

	COLUMN_NAME	DATA_TYPE	DATA_LENGTH	DATA_PRECISION
1	LOCATION_ID	NUMBER	22	(null)
2	LOCATION_SRC_ID	VARCHAR2	10	(null)
3	LOCATION_NAME	VARCHAR2	256	(null)

RCA:

ALTER TABLE DWH_LAYER. DWH_LOCATIONS

MODIFY LOCATION_NAME VARCHAR2(100);