|  |
| --- |
|  |
| Fact Table Techniques |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| REVISION HISTORY | | | | | |
| Ver. | Description of Change | Author | Date | Approved | |
| Name | Effective Date |
| 1.0 | Initial status | Arina Marchenko | 27-NOV-2017 |  |  |

Contents

[1. Data 3](#_Toc499507517)

[2. Dimensional Tables 3](#_Toc499507518)

[3. Schema 4](#_Toc499507519)

[4. Data generation 5](#_Toc499507520)

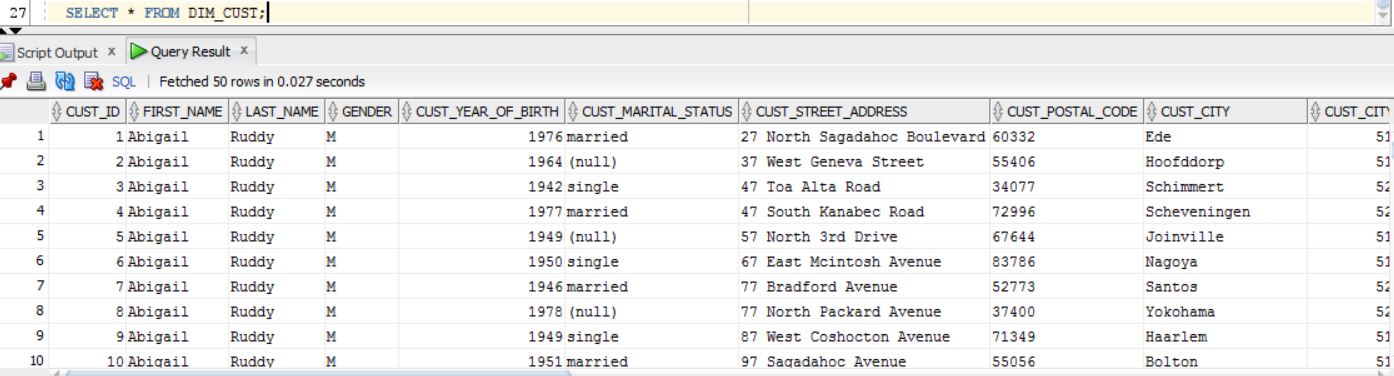
[5. Reports 6](#_Toc499507521)

# Data

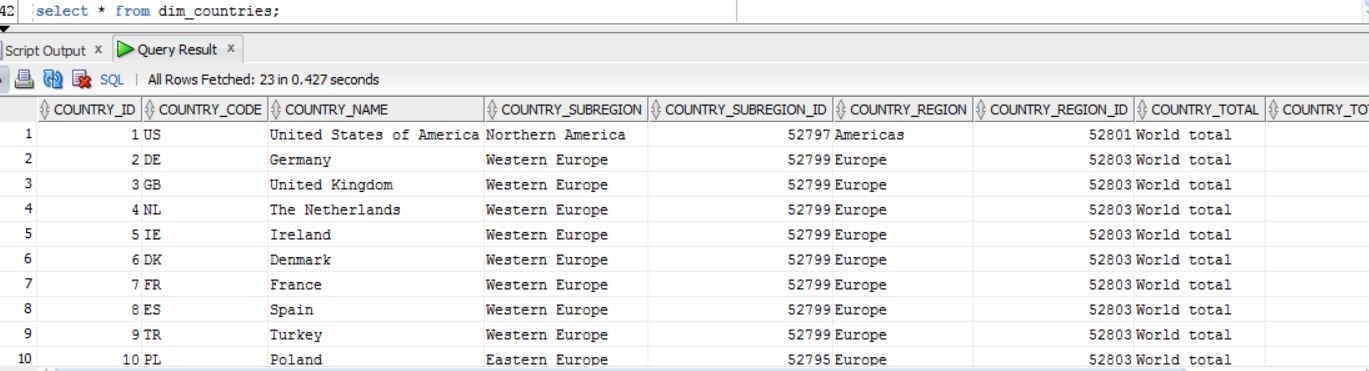
For dimensional tables and fact table I used information from sh.customers, sh.countries, sh.products, sh.channels and dim\_date, which I generate for the previous tasks.

# Dimensional Tables

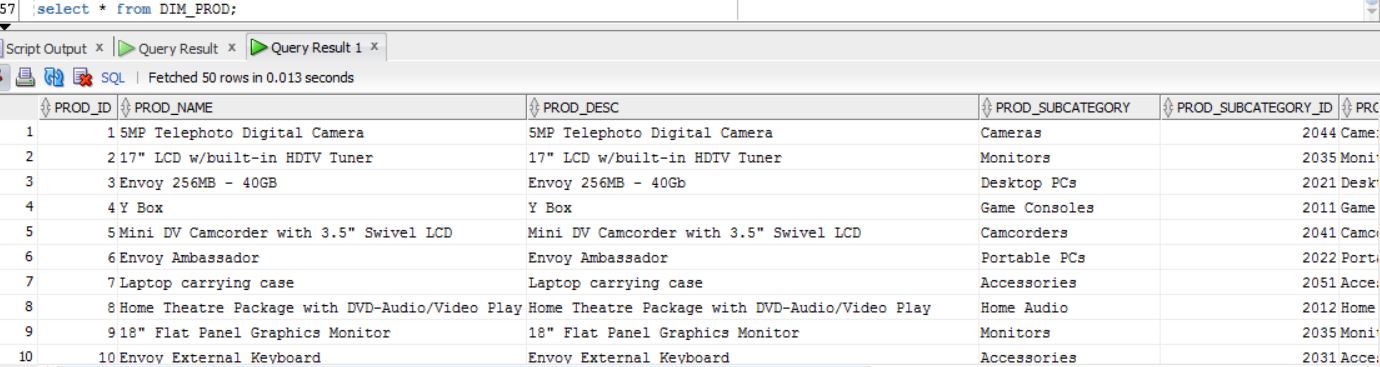
Dim\_cust was based on sh.customers and include information, such as CUST\_ID, FIRST\_NAME, LAST\_NAME, GENDER, CUST\_YEAR\_OF\_BIRTH, CUST\_MARITAL\_STATUS, CUST\_STREET\_ADDRESS, CUST\_POSTAL\_CODE, CUST\_CITY, CUST\_CITY\_ID, CUST\_STATE\_PROVINCE, CUST\_STATE\_PROVINCE\_ID, COUNTRY\_ID, CUST\_MAIN\_PHONE\_NUMBER, CUST\_INCOME\_LEVEL, CUST\_CREDIT\_LIMIT, CUST\_EMAIL, CUST\_TOTAL, CUST\_TOTAL\_ID, CUST\_SRC\_ID, CUST\_EFF\_FROM, CUST\_EFF\_TO, CUST\_VALID. There are 55550 rows in this table.



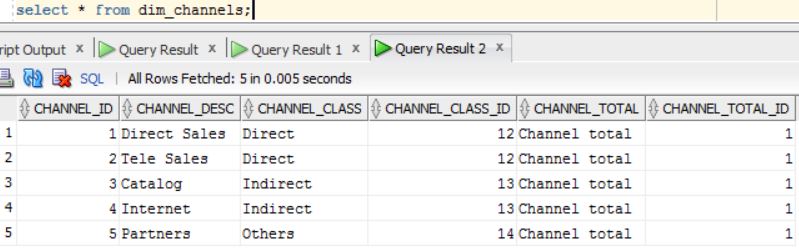
Dim\_countries was based on sh.countries and include information such as COUNTRY\_ID, COUNTRY\_CODE, COUNTRY\_NAME, COUNTRY\_SUBREGION, COUNTRY\_SUBREGION\_ID, COUNTRY\_REGION, COUNTRY\_REGION\_ID, COUNTRY\_TOTAL, COUNTRY\_TOTAL\_ID, COUNTRY\_NAME\_HIST. There are 23 rows here.



Dim\_prod was based on sh.products and include PROD\_ID, PROD\_NAME, PROD\_DESC, PROD\_SUBCATEGORY, PROD\_SUBCATEGORY\_ID, PROD\_SUBCATEGORY\_DESC, PROD\_CATEGORY, PROD\_CATEGORY\_ID, PROD\_CATEGORY\_DESC, PROD\_LIST\_PRICE. There are 72 rows here.



Dim\_channels id based on sh.channels and include CHANNEL\_ID, CHANNEL\_DESC, CHANNEL\_CLASS, CHANNEL\_CLASS\_ID, CHANNEL\_TOTAL, CHANNEL\_TOTAL\_ID. There are 5 rows here.



Dim\_date. We used to generate date dimension in previous task.

INSERT INTO Date\_Dimension

SELECT SYSDATE+rownum-365\*1018 as Date\_id,

to\_number(to\_char(SYSDATE+rownum-365\*1018-7\*1, 'D'))as Day\_per\_week,

to\_number(to\_char(extract(day from SYSDATE+rownum-365\*1018)))as Day\_per\_month,

to\_number(to\_char(SYSDATE+rownum-365\*1018, 'DDD')) AS Day\_per\_year,

to\_number(to\_char(SYSDATE+rownum-365\*1018, 'W')) AS Week\_per\_month,

to\_number(to\_char(SYSDATE+rownum-365\*1018, 'IW')) AS Week\_per\_year,

to\_number(to\_char(extract(month from SYSDATE+rownum-365\*1018))) as Month\_number,

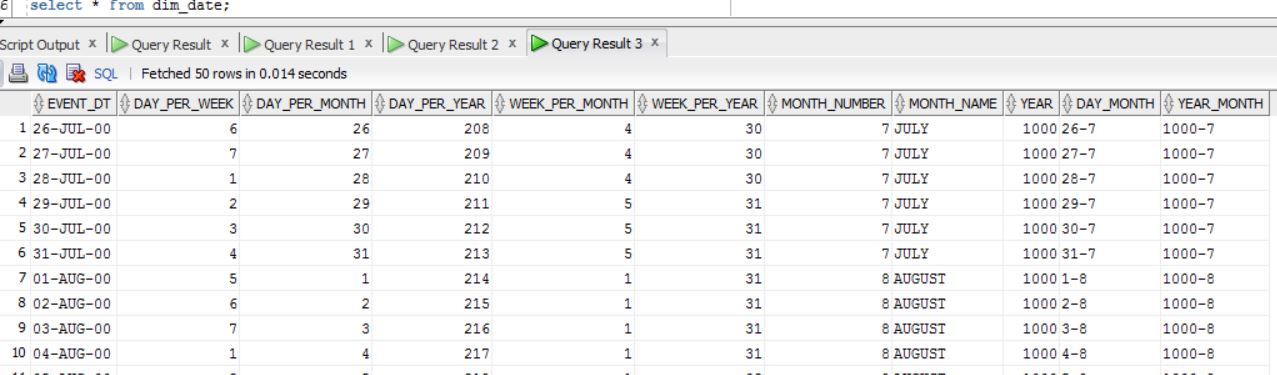
to\_char(SYSDATE+rownum-365\*1018, 'MONTH', 'NLS\_DATE\_LANGUAGE=English') AS Month\_name,

to\_number(to\_char(extract(year from SYSDATE+rownum-365\*1018))) as Year,

to\_char(extract(day from SYSDATE+rownum-365\*1018) || '-'|| to\_char(extract(month from (SYSDATE+rownum-365\*1018)))) as Day\_Month,

to\_char(extract(year from SYSDATE+rownum-365\*1018) || '-'|| to\_char(extract(month from (SYSDATE+rownum-365\*1018)))) as Year\_Month

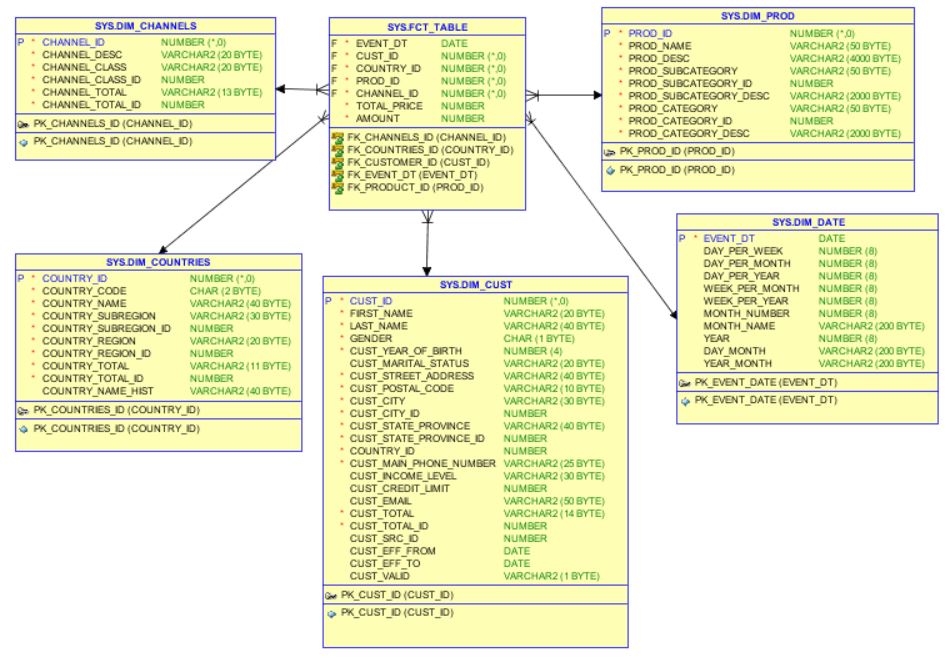
FROM dual CONNECT BY rownum <=365\*1018;



Based on everything we can make a conclusion that business of the table is tv sales. So it is daily sale per customer per channel per country per product. Fact table include four foreign keys and 2 measures:total price and amount.

# Schema

Here is the schema of our dimensionals and a fact table.



# Data generation

Create table fct\_table (

EVENT\_DT DATE not null,

CUST\_ID NUMBER,

COUNTRY\_ID NUMBER,

PROD\_ID NUMBER,

CHANNEL\_ID NUMBER,

TOTAL\_PRICE NUMBER,

AMOUNT NUMBER);

drop table fct\_table cascade constraints;

BEGIN

FOR i IN 1..500

LOOP

dbms\_random.seed

(

i \* 3

)

;

INSERT INTO fct\_table

( Event\_dt,

cust\_id,

country\_id,

prod\_id,

channel\_id,

total\_price,

amount

)

SELECT TRUNC ( SYSDATE - DBMS\_RANDOM.VALUE ( 1, 1700 ) ) AS EVENT\_DT ,

ROUND (DBMS\_RANDOM.VALUE (1,

(SELECT MAX(CUST\_ID)FROM DIM\_CUST

))) AS CUST\_ID ,

ROUND ( DBMS\_RANDOM.VALUE (1,

(SELECT MAX(COUNTRY\_ID)FROM DIM\_COUNTRIES

))) AS COUNTRY\_ID ,

ROUND ( DBMS\_RANDOM.VALUE (1,

( SELECT MAX(PROD\_ID) FROM DIM\_PROD

)) ) AS PROD\_ID ,

ROUND ( DBMS\_RANDOM.VALUE (1,

( SELECT MAX(CHANNEL\_ID) FROM DIM\_CHANNELS

)) ) AS CHANNEL\_ID ,

ROUND ( DBMS\_RANDOM.VALUE (10, 500) ) AS total\_price ,

ROUND ( DBMS\_RANDOM.VALUE (1, 20) ) AS AMOUNT

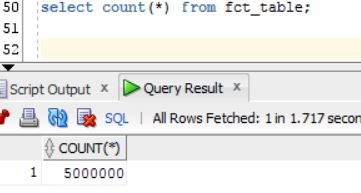
FROM dual

CONNECT BY level <= 10000;

END LOOP;

END;

/



# Reports

With group by cube we can find the information about product category per 2017 year and their total sales.

SELECT p.prod\_category,

fct.event\_dt,

TO\_CHAR(SUM(fct.total\_price),'9,999,999,999') AS Sales

FROM dim\_prod p,

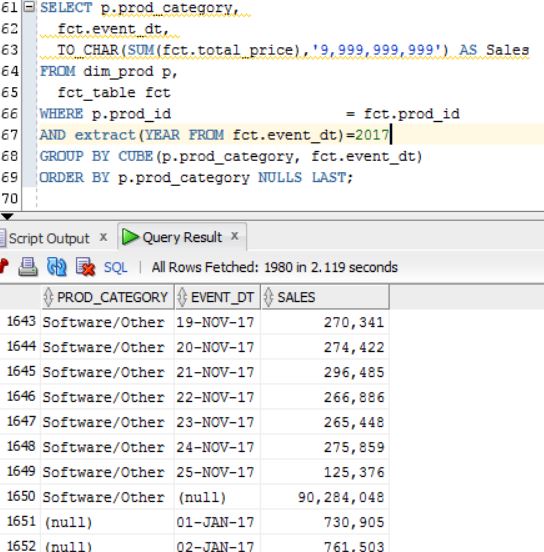
fct\_table fct

WHERE p.prod\_id = fct.prod\_id

AND extract(YEAR FROM fct.event\_dt)=2017

GROUP BY CUBE(p.prod\_category, fct.event\_dt)

ORDER BY p.prod\_category NULLS LAST;



With group by rollup we can find the information about categories per each channels and their date. Also here is total sales.

SELECT DECODE(GROUPING( s.channel\_desc),1,'All channels', s.channel\_desc) AS channel,

DECODE(GROUPING( p.prod\_category),1,'All categories', p.prod\_category) AS categories,

fct.event\_dt,

TO\_CHAR(SUM(fct.total\_price),'9,999,999,999') AS Sales

FROM dim\_channels s,

fct\_table fct,

dim\_prod p

WHERE s.channel\_id = fct.channel\_id

AND fct.prod\_id =p.prod\_id

GROUP BY ROLLUP(s.channel\_desc, p.prod\_category, fct.event\_dt)

ORDER BY s.channel\_desc,

p.prod\_category NULLS LAST;

