/\* Explore GTIN\_13 tables - Randy Lisbona 10/29/2016 – 12/6/2016 \*/

/\* Summarize record counts in all tables \*/

select \* from information\_schema.tables where information\_schema.tables.table\_schema = 'gtin\_13';

flush tables;

/\* this is giving a slightly different record count for some tables, not sure why, use individual queries instead

select TABLE\_NAME, TABLE\_ROWS from information\_schema.tables where information\_schema.tables.table\_schema = 'gtin\_13';

/\* TABLE\_NAME TABLE\_ROWS \*/

/\* brand 4151 \*/

/\* brand\_group 3 \*/

/\* brand\_owner 32 \*/

/\* brand\_owner\_bsin 581 \*/

/\* brand\_type 2 \*/

/\* gs1\_gcp 1549550 \*/

/\* gs1\_gcp\_nb 264 \*/

/\* gs1\_gcp\_rc 16 \*/

/\* gs1\_gpc 38760 \*/

/\* gs1\_gpc\_hier 3298 \*/

/\* gs1\_prefix 1000 \*/

/\* gtin 844270 \*/

/\* label 2 \*/

/\* label\_gtin 3 \*/

/\* nutrition\_us 231 \*/

/\* pkg\_type 42 \*/

select \* from information\_schema.REFERENTIAL\_CONSTRAINTS where information\_schema.REFERENTIAL\_CONSTRAINTS.constraint\_schema = 'gtin\_13';

/\* no referential constraints found \*/

select \* from information\_schema.TABLE\_CONSTRAINTS where information\_schema.TABLE\_CONSTRAINTS.constraint\_schema = 'gtin\_13';

/\* All tables have Primary Keys \*/

select \* from information\_schema.KEY\_COLUMN\_USAGE where information\_schema.KEY\_COLUMN\_USAGE.constraint\_schema = 'gtin\_13';

/\* Tables and Keys

TABLE\_NAME Key1 Key2

brand BSIN

brand\_group BSIN

brand\_owner OWNER\_CD

brand\_owner\_bsin BSIN

brand\_type BRAND\_TYPE\_CD

gs1\_gcp GCP\_CD

gs1\_gcp\_nb prefix\_cd gcp\_length

gs1\_gcp\_rc RETURN\_CODE

gs1\_gpc GPC\_LANG GPC\_CD

gs1\_gpc\_hier GPC\_B\_CD

gs1\_prefix PREFIX\_CD

gtin GTIN\_CD

label LABEL\_ID

nutrition\_us GTIN\_CD

pkg\_type pkg\_type\_cd

\*/

/\* select record counts from all tables \*/

select 'brand' as 'TABLE\_NAME', count(\*) as 'TABLE\_ROWS' from gtin\_13.brand union

select 'brand\_owner' as 'TABLE\_NAME', count(\*) as 'TABLE\_ROWS' from gtin\_13.brand\_owner union

select 'brand\_owner\_bsin' as 'TABLE\_NAME', count(\*) as 'TABLE\_ROWS' from gtin\_13.brand\_owner\_bsin union

select 'brand\_type' as 'TABLE\_NAME', count(\*) as 'TABLE\_ROWS' from gtin\_13.brand\_type union

select 'gs1\_gcp' as 'TABLE\_NAME', count(\*) as 'TABLE\_ROWS' from gtin\_13.gs1\_gcp union

select 'gs1\_gcp\_nb' as 'TABLE\_NAME', count(\*) as 'TABLE\_ROWS' from gtin\_13.gs1\_gcp\_nb union

select 'gs1\_gcp\_rc' as 'TABLE\_NAME', count(\*) as 'TABLE\_ROWS' from gtin\_13.gs1\_gcp\_rc union

select 'gs1\_gpc' as 'TABLE\_NAME', count(\*) as 'TABLE\_ROWS' from gtin\_13.gs1\_gpc union

select 'gs1\_gpc\_hier' as 'TABLE\_NAME', count(\*) as 'TABLE\_ROWS' from gtin\_13.gs1\_gpc\_hier union

select 'gs1\_prefix' as 'TABLE\_NAME', count(\*) as 'TABLE\_ROWS' from gtin\_13.gs1\_prefix union

select 'gtin' as 'TABLE\_NAME', count(\*) as 'TABLE\_ROWS' from gtin\_13.gtin union

select 'nutrition\_us' as 'TABLE\_NAME', count(\*) as 'TABLE\_ROWS' from gtin\_13.nutrition\_us union

select 'pkg\_type' as 'TABLE\_NAME', count(\*) as 'TABLE\_ROWS' from gtin\_13.pkg\_type;

/\* Individual tables \*/

drop view if exists brand\_owner\_bsin\_subset;

create view brand\_owner\_bsin\_subset as

SELECT \* FROM gtin\_13.brand\_owner\_bsin

where OWNER\_CD between 27 and 30

order by OWNER\_CD, BSIN;

SELECT \* FROM gtin\_13.brand\_type;

SELECT \* FROM gtin\_13.pkg\_type;

select distinct(pkg\_type\_cd) from gtin; /\* Check how many package types there are: 23 rows returned \*/

select distinct(GCP\_CD) from gtin; /\* 52918 distinct rows returned for Company code \*/

/\* 15,502 distinct brand single identification number (BSIN) matched in gtin \*/

select count(\*) as Freq, A.GCP\_CD, A.BSIN, B.BRAND\_NM from gtin\_13.gtin A join gtin\_13.brand B on a.BSIN = B.BSIN GROUP BY A.GCP\_CD, A.BSIN, B.BRAND\_NM ;

select A.GTIN\_CD, A.GCP\_CD, A.BSIN, B.BRAND\_NM

from gtin\_13.gtin A left join gtin\_13.brand B on a.BSIN = B.BSIN

GROUP BY A.GCP\_CD, A.BSIN, B.BRAND\_NM

order by A.GTIN\_CD ; /\* 62,849 of 922,000 gtin records have BSIN codes \*/

select distinct(length(GTIN\_CD)) from gtin; /\* check this syntax \*/

select GTIN\_CD, length(GTIN\_CD) as GTIN\_LEN, GCP\_CD, length(GCP\_CD) as GCP\_LEN from gtin order by length(GTIN\_CD);

select A.GPC\_NM, B.GTIN\_CD, B.GPC\_S\_CD, B.GPC\_C\_CD, B.GPC\_C\_CD, B.PRODUCT\_LINE from

gtin\_13.gs1\_gpc A /\* 1,673,000 rows returned -Product names in various languages \*/

join gtin\_13.gtin B on B.GPC\_B\_CD = A.GPC\_CD ; /\*and A.GPC\_LANG = "EN";\*/

select M\_G, M\_OZ, count(\*) from gtin\_13.gtin

group by M\_G, M\_OZ;

drop view if exists gtin\_subset;

create view gtin\_subset as

SELECT A.\* FROM gtin\_13.gtin A join gtin\_13.nutrition\_us B on

A.GTIN\_CD = B.GTIN\_CD

where b.INGREDIENTS REGEXP '.WHEAT.'

limit 20;

drop view if exists gs1\_gcp\_subset;

create view gs1\_gcp\_subset as

SELECT A.\* FROM gtin\_13.gs1\_gcp A left join gtin\_subset B on

A.GCP\_CD = B.GCP\_CD or a.GCP\_CD = '73410'

order by GCP\_CD

limit 20;

/\* find records with 73410 anywhere in the GCP\_CD field \*/

SELECT A.\* FROM gtin\_13.gs1\_gcp A

where A.GCP\_CD REGEXP '.73410.';

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* The deeper we go (linking more tables) the fewer matching records we find \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\* Recommend using primarily the gtin table and brand table, gts\_gcp doesn't have much useful information \*\*\*\*\*\*\*\*\*\*/

select A.\* from

gtin\_13.gs1\_gcp A /\* 1,673,000 rows Global Company Prefixes \*/

join gtin\_13.gtin B on A.GCP\_CD = B.GCP\_CD /\* 918,000 rows matched 923,000 GTIN product item codes \*/

join gtin\_13.brand D on B.BSIN = D.BSIN /\* 527,000 rows matched 4,100 Brand names \*/

join gtin\_13.brand\_owner\_bsin E on D.BSIN = E.BSIN /\* 86,500 rows matched 581 Brand owner code of Brand \*/

join gtin\_13.brand\_owner F on E.OWNER\_CD = F.OWNER\_CD /\* 86,500 rows matched 32 Brand owner name of owner code \*/

join gtin\_13.nutrition\_us C on B.GTIN\_CD = C.GTIN\_CD /\* 52 rows matched 231 Nutrition information of item \*/

join gtin\_13.pkg\_type G on B.PKG\_TYPE\_CD = G.pkg\_type\_cd /\* 0 rows matched 42 Package description of item \*/

/\* extract example rows from each table \*/

SELECT \* FROM gtin\_13.gs1\_gcp\_nb

where prefix\_cd between '800' and '900';

SELECT \* FROM gtin\_13.gs1\_gcp\_rc;

SELECT \* FROM gtin\_13.gs1\_gpc

where GPC\_LANG = 'EN' ;

limit 15;

SELECT \* FROM gtin\_13.gs1\_gpc\_hier

limit 15;

SELECT \* FROM gtin\_13.gs1\_prefix

where COUNTRY\_ISO\_CD ='US'

limit 10;

SELECT A.\* FROM gtin\_13.nutrition\_us A join gtin\_subset B on

A.GTIN\_CD = B.GTIN\_CD

limit 200;

SELECT A.\* FROM gtin\_13.nutrition\_us A join gtin B on

A.GTIN\_CD = B.GTIN\_CD

limit 20;

SELECT \* FROM gtin\_13.label\_gtin;

SELECT \* FROM gtin\_13.label;

SELECT A.\* FROM gtin\_13.brand as A join brand\_owner\_bsin\_subset as B on

A.BSIN = B.BSIN

order by A.BSIN;

SELECT \* FROM gtin\_13.brand\_group;

SELECT \* FROM gtin\_13.brand\_owner;

select Count(\*) from gtin\_13.gtin where BSIN is null;

select count(\*), length(trim(leading '0' from GTIN\_CD)) as length from gtin\_13.gtin group by length(trim(leading '0' from GTIN\_CD));

/\* count number of rows by brand URL \*/

select count(\*) , BRAND\_LINK from gtin\_13.brand group BY BRAND\_LINK order by count(\*) DESC;

/\* count number of rows where the Global Location Number name is null or '', these are just showing the country that issued the code 512,813 rows \*/

select count(\*), ifnull(GLN\_CD,'') as GLN\_CD, GLN\_NM from gtin\_13.gs1\_gcp group by ifnull(GLN\_CD,''), GLN\_NM order by count(\*) desc;

/\* count the records that have a Gloal Location Number Code 505,740 \*/

select count(\*), GLN\_CD from gtin\_13.gs1\_gcp group by GLN\_CD having not isnull(GLN\_CD) order by count(\*) desc;

/\* Data for chart of Global Location Numbers\*/

select GCP\_CD, ifnull(GLN\_CD,'') as GLN\_CD, GLN\_NM from gtin\_13.gs1\_gcp

where is null(GLN\_NM)

order by ifnull(GLN\_CD,''), GCP\_CD ;

/\*Returns 1,095,279 rows with '' Global Location numbers out of 1,549,550\*/

select GCP\_CD, GLN\_CD, GLN\_NM from gtin\_13.gs1\_gcp

where GLN\_CD = ''

order by GLN\_CD ;

/\* Returns 540580 Global Location Numbers with company codes out of 1,549,550\*/

select GCP\_CD, GLN\_CD, GLN\_NM from gtin\_13.gs1\_gcp

where GLN\_CD <> '' and not (GLN\_NM like '%GS1%' or GLN\_NM like '%Unknown country%' or GLN\_NM = '' or GLN\_NM like 'Prefix never allocated%' or GLN\_NM like 'ReturnCode%');

order by GLN\_NM ;

/\* Returns 35,066 rows that have a Global Location Number but only the country that issued it, no company \*/

select GCP\_CD, GLN\_CD, GLN\_NM from gtin\_13.gs1\_gcp

where GLN\_CD <> '' and GLN\_NM like '%GS1%';

order by GLN\_CD ;

/\* Returns 2739 rows with an garbage GLN\_NM \*/

select GCP\_CD, GLN\_CD, GLN\_NM from gtin\_13.gs1\_gcp

where GLN\_CD <> '' and ( GLN\_NM like '%Unknown country%' or GLN\_NM = '' or GLN\_NM like 'Prefix never allocated%' or GLN\_NM like 'ReturnCode%')

order by GLN\_CD ;

/\* Overall summary table \*/

select Status, count(\*) from (

select GLN\_CD, GLN\_NM,

case

when GLN\_NM like '%GS1%' then 'GS1 Issue Country only'

when GLN\_CD = '' then 'Invalid or Missing'

when GLN\_NM like '%Unknown country%' or GLN\_NM = '' or GLN\_NM like 'Prefix never allocated%' or GLN\_NM like 'ReturnCode%' then 'Invalid or Missing'

else 'Global Location Number found'

end as Status

from gtin\_13.gs1\_gcp ) as derivedtable

group by Status

order by count(\*) desc;

/\* end Randy Lisbona exploratory queries \*/