**SEO GOOGLE – DOU**

**SCHEMA\_NAME** : V2DIGT

**SOURCE\_SUBJECT** : SEO\_GOOGLE

**MANUAL\_FLAG** : N 🡺 load incremental data: update and insert target table

**TABLE\_LIST** : 🡪 KEY : FCT\_GOOGLE\_SRCH\_KEYWRD\_URL

🡪 URL : FCT\_GOOGLE\_SRCH\_URL

**seo\_master.sh:**

* Get parameters
* Get script folder
* Generate NFS to store flat files
* Keep latest records only (until 7 days) and delete the older records
* Definition of log function
* Definition of send email function
* Get key from file
* Decrypt JDBC URL file
* Get multiple schemas if exists
* Get table list
* Loads the data from cos bucket
* Call seo\_extractor.sh with all the necessary parameters
* Check logs and send email

**seo\_extractor.sh:**

* Definition of log function
* Get parameters
* Get script folder
* Get key from file
* Decrypt JDBC URL file
* Add import log into ts\_file
* Execute START\_LOAD script to get start timestamp
* Definition of start\_load()
  + Connects to DB
  + Sets last\_etl\_dt as the newest source create timestamp or [current date – 3] (if null)
  + Sets start\_and\_last\_dt
* Check whether manual refresh is set
* Connect to db2 ACS DB
* Truncate the below tables
  + V2DIGT0.L\_ FCT\_GOOGLE\_SRCH\_KEYWRD\_URL \_W3
  + V2DIGT0.L\_FCT\_GOOGLE\_SRCH\_KEYWRD\_URL\_ DOMAIN
  + V2DIGT0.L\_ FCT\_GOOGLE\_SRCH\_URL \_W3
  + V2DIGT0.L\_FCT\_GOOGLE\_SRCH\_URL\_ DOMAIN
* Move python script and needed config file to the temp folder
* Gather the data from the IBM domain using Python scripts
  + **pull\_data\_url.py:**
    - * Get arguments
      * Initiate the sample\_tools and give the required arguments : command line arguments, name of the API, version of the API, description of the application, filename of the application, list of additional command line flags and the OAuth scope used (PS: These are in the order of the parameters mentioned in the python script)
      * Create a dataframe with required columns.
      * Copy this empty DataFrame to a csv file in the location which was sent as arguments earlier.
      * Now, request data using the search Analytics query to get the response.
      * Store the response data in the data frame and append it into the same csv file created earlier
      * This csv file is returned as the output and loaded into the truncated table.
    - **pull\_data\_key.py:**
      * Follow the same steps as above, the only difference is the columns of the dataframe and key used to retrieve the data.
* Load the data retrieved using python scripts into the tables which were truncated earlier.
* Start with the ETL process
* **DB2\_FCT\_GOOGLE\_SRCH\_KEYWRD\_URL.sql**
* **Extract the rows from distinct PAGE\_COMPL\_URL**

DECLARE GLOBAL TEMPORARY TABLE SESSION.MPWEXT\_L\_V\_DIM\_URL\_HDP\_TEMP\_SEO\_GOOGLE\_SRCH\_KEYWRD\_URL\_1\_TEMP AS

(

SELECT DISTINCT(C),DIM\_PAGE\_URL\_ID,URN\_REG\_DIM\_URL,UPDT\_TS FROM V2DIGT2.V\_DIM\_URL

) WITH DATA ORGANIZE BY ROW NOT LOGGED;

* **Extract the latest updated rows alone from the above table**

DECLARE GLOBAL TEMPORARY TABLE SESSION.MPWEXT\_L\_V\_DIM\_URL\_HDP\_TEMP\_SEO\_GOOGLE\_SRCH\_KEYWRD\_URL\_1 AS

(

select \* from (select f.PAGE\_COMPL\_URL,f.DIM\_PAGE\_URL\_ID,f.URN\_REG\_DIM\_URL,row\_number() OVER (PARTITION BY PAGE\_COMPL\_URL ORDER BY UPDT\_TS DESC) as rn from SESSION.MPWEXT\_L\_V\_DIM\_URL\_HDP\_TEMP\_SEO\_GOOGLE\_SRCH\_KEYWRD\_URL\_1\_TEMP f) L where L.rn=1

) WITH DATA ORGANIZE BY ROW NOT LOGGED;

* **Remove the substrings ['%/index.html','/?lnk','//'] from PAGE\_COMPL\_URL if present from previous table**

DECLARE GLOBAL TEMPORARY TABLE SESSION.MPWEXT\_L\_V\_DIM\_URL\_HDP\_TEMP\_SEO\_GOOGLE\_SRCH\_KEYWRD\_URL\_2\_TEMP AS

(

select

(CASE

WHEN A.PAGE\_COMPL\_URL LIKE '%/index.html' THEN SUBSTR(A.PAGE\_COMPL\_URL, 1,(LENGTH(A.PAGE\_COMPL\_URL) - 11))

WHEN A.PAGE\_COMPL\_URL LIKE '%/?lnk%' THEN SUBSTR(A.PAGE\_COMPL\_URL, 1,instr(SUBSTR(A.PAGE\_COMPL\_URL, 9),'/?lnk')-1)

WHEN A.PAGE\_COMPL\_URL LIKE '%//' THEN SUBSTR(A.PAGE\_COMPL\_URL, 1,(LENGTH(A.PAGE\_COMPL\_URL)-2))

ELSE A.PAGE\_COMPL\_URL

END) as PAGE\_COMPL\_URL,A.DIM\_PAGE\_URL\_ID,A.URN\_REG\_DIM\_URL from SESSION.MPWEXT\_L\_V\_DIM\_URL\_HDP\_TEMP\_SEO\_GOOGLE\_SRCH\_KEYWRD\_URL\_1 A

) WITH DATA ORGANIZE BY ROW NOT LOGGED;

* **Keep the records with largest DIM\_PAGE\_URL\_ID from previous table**

DECLARE GLOBAL TEMPORARY TABLE SESSION.MPWEXT\_L\_V\_DIM\_URL\_HDP\_TEMP\_SEO\_GOOGLE\_SRCH\_KEYWRD\_URL\_2 AS

(

select \* from (select f.PAGE\_COMPL\_URL,f.DIM\_PAGE\_URL\_ID,f.URN\_REG\_DIM\_URL,row\_number() OVER (PARTITION BY PAGE\_COMPL\_URL ORDER BY DIM\_PAGE\_URL\_ID DESC) as rn from SESSION.MPWEXT\_L\_V\_DIM\_URL\_HDP\_TEMP\_SEO\_GOOGLE\_SRCH\_KEYWRD\_URL\_2\_TEMP f) L where L.rn=1

) WITH DATA ORGANIZE BY ROW NOT LOGGED;

* **Remove ‘/’ from the end of the PAGE\_COMPL\_URL**

DECLARE GLOBAL TEMPORARY TABLE SESSION.MPWEXT\_L\_V\_DIM\_URL\_HDP\_TEMP\_SEO\_GOOGLE\_SRCH\_KEYWRD\_URL\_3\_TEMP AS

(

select

(CASE

WHEN A.PAGE\_COMPL\_URL LIKE '%/' THEN SUBSTR(A.PAGE\_COMPL\_URL, 1,(LENGTH(A.PAGE\_COMPL\_URL)-1))

ELSE A.PAGE\_COMPL\_URL

END) as PAGE\_COMPL\_URL,A.DIM\_PAGE\_URL\_ID,A.URN\_REG\_DIM\_URL from SESSION.MPWEXT\_L\_V\_DIM\_URL\_HDP\_TEMP\_SEO\_GOOGLE\_SRCH\_KEYWRD\_URL\_2 A

) WITH DATA ORGANIZE BY ROW NOT LOGGED;

* **Keep the records with largest DIM\_PAGE\_URL\_ID from previous table**

DECLARE GLOBAL TEMPORARY TABLE SESSION.MPWEXT\_L\_V\_DIM\_URL\_HDP\_TEMP\_SEO\_GOOGLE\_SRCH\_KEYWRD\_URL\_3 AS

(

select \* from (select f.PAGE\_COMPL\_URL,f.DIM\_PAGE\_URL\_ID,f.URN\_REG\_DIM\_URL,row\_number() OVER (PARTITION BY PAGE\_COMPL\_URL ORDER BY DIM\_PAGE\_URL\_ID DESC) as rn from SESSION.MPWEXT\_L\_V\_DIM\_URL\_HDP\_TEMP\_SEO\_GOOGLE\_SRCH\_KEYWRD\_URL\_3\_TEMP f) L where L.rn=1

) WITH DATA ORGANIZE BY ROW NOT LOGGED;

* **Remove substrings ['http://','https://'] from PAGE\_URL if present from L\_FCT\_GOOGLE\_SRCH\_KEYWRD\_URL\_W3 and store in temp table**

DECLARE GLOBAL TEMPORARY TABLE SESSION.MPWEXT\_L\_FCT\_GOOGLE\_SRCH\_KEYWRD\_URL\_W3 AS

(

select

KEYWRD,

PAGE\_URL,

(CASE

WHEN PAGE\_URL LIKE 'http://%' THEN SUBSTR(PAGE\_URL, 8)

WHEN PAGE\_URL LIKE 'https://%' THEN SUBSTR(PAGE\_URL, 9)

ELSE PAGE\_URL

END) as PAGE\_URL\_TEMP,

SRC\_FILE\_CREATE\_DT,CLICKS,IMPR,CTR,POSITN,COUNTRY from V2DIGT0.L\_FCT\_GOOGLE\_SRCH\_KEYWRD\_URL\_W3

) WITH DATA ORGANIZE BY ROW NOT LOGGED;

* **Remove substrings ['http://','https://'] from PAGE\_URL if present from V2DIGT0.L\_FCT\_GOOGLE\_SRCH\_KEYWRD\_URL\_DOMAIN and store in temp table**

DECLARE GLOBAL TEMPORARY TABLE SESSION.MPWEXT\_L\_FCT\_GOOGLE\_SRCH\_KEYWRD\_URL\_DOMAIN AS

(

select KEYWRD,

PAGE\_URL,

(CASE

WHEN PAGE\_URL LIKE 'http://%' THEN SUBSTR(PAGE\_URL, 8)

WHEN PAGE\_URL LIKE 'https://%' THEN SUBSTR(PAGE\_URL, 9)

ELSE PAGE\_URL

END) as PAGE\_URL\_TEMP,

SRC\_FILE\_CREATE\_DT,CLICKS,IMPR,CTR,POSITN,COUNTRY from V2DIGT0.L\_FCT\_GOOGLE\_SRCH\_KEYWRD\_URL\_DOMAIN

) WITH DATA ORGANIZE BY ROW NOT LOGGED;

* **Insert the rows to SCHEMA1 table,from MPWEXT\_L\_FCT\_GOOGLE\_SRCH\_KEYWRD\_URL\_W3 include latest updated DIM\_PAGE\_URL\_ID of the corresponding PAGE\_URL**

insert into V2DIGT1.FCT\_GOOGLE\_SRCH\_KEYWRD\_URL(

URN\_REG\_DIM\_URL

,SRCH\_KEYWRD

,PAGE\_URL

,DIM\_PAGE\_URL\_ID

,SRC\_CREATE\_TS

,CLICK\_CNT

,IMPR\_CNT

,CLICK\_THRU\_RATE

,RANK\_POSITN

,CTRY\_ALPHA3\_COMPLNC\_CD

,SRC\_CREATE\_DT

,DATA\_SRC\_CD

,CREATE\_TS

,UPDT\_TS

) select

B.URN\_REG\_DIM\_URL,

A.KEYWRD,

A.PAGE\_URL,

B.DIM\_PAGE\_URL\_ID,

cast(A.SRC\_FILE\_CREATE\_DT as timestamp) as SRC\_CREATE\_TS,

A.CLICKS,

A.IMPR,

A.CTR,

A.POSITN,

A.COUNTRY,

A.SRC\_FILE\_CREATE\_DT,

'GSC' AS DATA\_SRC\_CD,

CURRENT\_TIMESTAMP AS CREATE\_TS,

CURRENT\_TIMESTAMP AS UPDT\_TS

from SESSION.MPWEXT\_L\_FCT\_GOOGLE\_SRCH\_KEYWRD\_URL\_W3 A

LEFT JOIN SESSION.MPWEXT\_L\_V\_DIM\_URL\_HDP\_TEMP\_SEO\_GOOGLE\_SRCH\_KEYWRD\_URL\_1 B ON lower(A.PAGE\_URL\_TEMP)=B.PAGE\_COMPL\_URL;

* **Insert the rows to SCHEMA1 table,from MPWEXT\_L\_FCT\_GOOGLE\_SRCH\_KEYWRD\_URL\_DOMAIN where, page url is NULL and include latest updated DIM\_PAGE\_URL\_ID.**

insert into V2DIGT1.FCT\_GOOGLE\_SRCH\_KEYWRD\_URL(

URN\_REG\_DIM\_URL

,SRCH\_KEYWRD

,PAGE\_URL

,DIM\_PAGE\_URL\_ID

,SRC\_CREATE\_TS

,CLICK\_CNT

,IMPR\_CNT

,CLICK\_THRU\_RATE

,RANK\_POSITN

,CTRY\_ALPHA3\_COMPLNC\_CD

,SRC\_CREATE\_DT

,DATA\_SRC\_CD

,CREATE\_TS

,UPDT\_TS

) select

C.URN\_REG\_DIM\_URL,

A.KEYWRD,

A.PAGE\_URL,

C.DIM\_PAGE\_URL\_ID,

cast(A.SRC\_FILE\_CREATE\_DT as timestamp) as SRC\_CREATE\_TS,

A.CLICKS,

A.IMPR,

A.CTR,

A.POSITN,

A.COUNTRY,

A.SRC\_FILE\_CREATE\_DT,

'GSC' AS DATA\_SRC\_CD,

CURRENT\_TIMESTAMP AS CREATE\_TS,

CURRENT\_TIMESTAMP AS UPDT\_TS

from SESSION.MPWEXT\_L\_FCT\_GOOGLE\_SRCH\_KEYWRD\_URL\_DOMAIN A

LEFT JOIN SESSION.MPWEXT\_L\_FCT\_GOOGLE\_SRCH\_KEYWRD\_URL\_W3 B ON A.PAGE\_URL=B.PAGE\_URL

LEFT JOIN SESSION.MPWEXT\_L\_V\_DIM\_URL\_HDP\_TEMP\_SEO\_GOOGLE\_SRCH\_KEYWRD\_URL\_1 C ON lower(A.PAGE\_URL\_TEMP)=C.PAGE\_COMPL\_URL

where B.PAGE\_URL IS NULL

;

* **LEVEL 1 UPDATE**

UPDATE V2DIGT1.FCT\_GOOGLE\_SRCH\_KEYWRD\_URL A SET A.URN\_REG\_DIM\_URL=B.URN\_REG\_DIM\_URL,A.DIM\_PAGE\_URL\_ID=B.DIM\_PAGE\_URL\_ID FROM SESSION.MPWEXT\_L\_V\_DIM\_URL\_HDP\_TEMP\_SEO\_GOOGLE\_SRCH\_KEYWRD\_URL\_1 B WHERE

(CASE WHEN A.PAGE\_URL LIKE 'http://%/' THEN lower(SUBSTR(SUBSTR(A.PAGE\_URL, 8), 1,LENGTH(SUBSTR(A.PAGE\_URL, 8))-1))

WHEN A.PAGE\_URL LIKE 'https://%/' THEN lower(SUBSTR(SUBSTR(A.PAGE\_URL, 9), 1,LENGTH(SUBSTR(A.PAGE\_URL, 8))-1))

WHEN A.PAGE\_URL LIKE 'http://%/index.html' THEN lower(SUBSTR(SUBSTR(A.PAGE\_URL, 8), 1,LENGTH(SUBSTR(A.PAGE\_URL, 8)) - 11))

WHEN A.PAGE\_URL LIKE 'https://%/index.html' THEN lower(SUBSTR(SUBSTR(A.PAGE\_URL, 9), 1,(LENGTH(SUBSTR(A.PAGE\_URL, 9)) - 11)))

WHEN A.PAGE\_URL LIKE 'http://%/?lnk%' THEN lower(SUBSTR(SUBSTR(A.PAGE\_URL, 8), 1,instr(SUBSTR(A.PAGE\_URL, 8),'/?lnk')-1))

WHEN A.PAGE\_URL LIKE 'https://%/?lnk%' THEN lower(SUBSTR(SUBSTR(A.PAGE\_URL, 9), 1,instr(SUBSTR(A.PAGE\_URL, 9),'/?lnk')-1))

ELSE lower(A.PAGE\_URL)

END) = B.PAGE\_COMPL\_URL and A.DIM\_PAGE\_URL\_ID is null;

UPDATE V2DIGT1.FCT\_GOOGLE\_SRCH\_KEYWRD\_URL A SET A.URN\_REG\_DIM\_URL=B.URN\_REG\_DIM\_URL,A.DIM\_PAGE\_URL\_ID=B.DIM\_PAGE\_URL\_ID FROM SESSION.MPWEXT\_L\_V\_DIM\_URL\_HDP\_TEMP\_SEO\_GOOGLE\_SRCH\_KEYWRD\_URL\_1 B

WHERE

(CASE WHEN A.PAGE\_URL LIKE 'http://%' THEN lower(SUBSTR(SUBSTR(A.PAGE\_URL, 8), 1,instr(SUBSTR(A.PAGE\_URL, 8),'/',-1)-1))

WHEN A.PAGE\_URL LIKE 'https://%' THEN lower(SUBSTR(SUBSTR(A.PAGE\_URL, 9), 1,instr(SUBSTR(A.PAGE\_URL, 9),'/',-1)-1))

ELSE lower(A.PAGE\_URL)

END)= B.PAGE\_COMPL\_URL AND A.DIM\_PAGE\_URL\_ID is null;

* **LEVEL 2 UPDATE**

UPDATE V2DIGT1.FCT\_GOOGLE\_SRCH\_KEYWRD\_URL A SET A.URN\_REG\_DIM\_URL=B.URN\_REG\_DIM\_URL,A.DIM\_PAGE\_URL\_ID=B.DIM\_PAGE\_URL\_ID FROM SESSION.MPWEXT\_L\_V\_DIM\_URL\_HDP\_TEMP\_SEO\_GOOGLE\_SRCH\_KEYWRD\_URL\_2 B WHERE

(CASE WHEN A.PAGE\_URL LIKE 'http://%/' THEN lower(SUBSTR(SUBSTR(A.PAGE\_URL, 8), 1,LENGTH(SUBSTR(A.PAGE\_URL, 8))-1))

WHEN A.PAGE\_URL LIKE 'https://%/' THEN lower(SUBSTR(SUBSTR(A.PAGE\_URL, 9), 1,LENGTH(SUBSTR(A.PAGE\_URL, 8))-1))

WHEN A.PAGE\_URL LIKE 'http://%/index.html' THEN lower(SUBSTR(SUBSTR(A.PAGE\_URL, 8), 1,LENGTH(SUBSTR(A.PAGE\_URL, 8)) - 11))

WHEN A.PAGE\_URL LIKE 'https://%/index.html' THEN lower(SUBSTR(SUBSTR(A.PAGE\_URL, 9), 1,(LENGTH(SUBSTR(A.PAGE\_URL, 9)) - 11)))

WHEN A.PAGE\_URL LIKE 'http://%/?lnk%' THEN lower(SUBSTR(SUBSTR(A.PAGE\_URL, 8), 1,instr(SUBSTR(A.PAGE\_URL, 8),'/?lnk')-1))

WHEN A.PAGE\_URL LIKE 'https://%/?lnk%' THEN lower(SUBSTR(SUBSTR(A.PAGE\_URL, 9), 1,instr(SUBSTR(A.PAGE\_URL, 9),'/?lnk')-1))

ELSE lower(A.PAGE\_URL)

END) = B.PAGE\_COMPL\_URL and A.DIM\_PAGE\_URL\_ID is null;

UPDATE V2DIGT1.FCT\_GOOGLE\_SRCH\_KEYWRD\_URL A SET A.URN\_REG\_DIM\_URL=B.URN\_REG\_DIM\_URL,A.DIM\_PAGE\_URL\_ID=B.DIM\_PAGE\_URL\_ID FROM SESSION.MPWEXT\_L\_V\_DIM\_URL\_HDP\_TEMP\_SEO\_GOOGLE\_SRCH\_KEYWRD\_URL\_2 B

WHERE

(CASE WHEN A.PAGE\_URL LIKE 'http://%' THEN lower(SUBSTR(SUBSTR(A.PAGE\_URL, 8), 1,instr(SUBSTR(A.PAGE\_URL, 8),'/',-1)-1))

WHEN A.PAGE\_URL LIKE 'https://%' THEN lower(SUBSTR(SUBSTR(A.PAGE\_URL, 9), 1,instr(SUBSTR(A.PAGE\_URL, 9),'/',-1)-1))

ELSE lower(A.PAGE\_URL)

END)= B.PAGE\_COMPL\_URL AND A.DIM\_PAGE\_URL\_ID is null;

* **LEVEL 3 UPDATE**

UPDATE V2DIGT1.FCT\_GOOGLE\_SRCH\_KEYWRD\_URL A SET A.URN\_REG\_DIM\_URL=B.URN\_REG\_DIM\_URL,A.DIM\_PAGE\_URL\_ID=B.DIM\_PAGE\_URL\_ID FROM SESSION.MPWEXT\_L\_V\_DIM\_URL\_HDP\_TEMP\_SEO\_GOOGLE\_SRCH\_KEYWRD\_URL\_3 B WHERE

(CASE WHEN A.PAGE\_URL LIKE 'http://%/' THEN lower(SUBSTR(SUBSTR(A.PAGE\_URL, 8), 1,LENGTH(SUBSTR(A.PAGE\_URL, 8))-1))

WHEN A.PAGE\_URL LIKE 'https://%/' THEN lower(SUBSTR(SUBSTR(A.PAGE\_URL, 9), 1,LENGTH(SUBSTR(A.PAGE\_URL, 8))-1))

WHEN A.PAGE\_URL LIKE 'http://%/index.html' THEN lower(SUBSTR(SUBSTR(A.PAGE\_URL, 8), 1,LENGTH(SUBSTR(A.PAGE\_URL, 8)) - 11))

WHEN A.PAGE\_URL LIKE 'https://%/index.html' THEN lower(SUBSTR(SUBSTR(A.PAGE\_URL, 9), 1,(LENGTH(SUBSTR(A.PAGE\_URL, 9)) - 11)))

WHEN A.PAGE\_URL LIKE 'http://%/?lnk%' THEN lower(SUBSTR(SUBSTR(A.PAGE\_URL, 8), 1,instr(SUBSTR(A.PAGE\_URL, 8),'/?lnk')-1))

WHEN A.PAGE\_URL LIKE 'https://%/?lnk%' THEN lower(SUBSTR(SUBSTR(A.PAGE\_URL, 9), 1,instr(SUBSTR(A.PAGE\_URL, 9),'/?lnk')-1))

ELSE lower(A.PAGE\_URL)

END) = B.PAGE\_COMPL\_URL and A.DIM\_PAGE\_URL\_ID is null;

UPDATE V2DIGT1.FCT\_GOOGLE\_SRCH\_KEYWRD\_URL A SET A.URN\_REG\_DIM\_URL=B.URN\_REG\_DIM\_URL,A.DIM\_PAGE\_URL\_ID=B.DIM\_PAGE\_URL\_ID FROM SESSION.MPWEXT\_L\_V\_DIM\_URL\_HDP\_TEMP\_SEO\_GOOGLE\_SRCH\_KEYWRD\_URL\_3 B

WHERE

(CASE WHEN A.PAGE\_URL LIKE 'http://%' THEN lower(SUBSTR(SUBSTR(A.PAGE\_URL, 8), 1,instr(SUBSTR(A.PAGE\_URL, 8),'/',-1)-1))

WHEN A.PAGE\_URL LIKE 'https://%' THEN lower(SUBSTR(SUBSTR(A.PAGE\_URL, 9), 1,instr(SUBSTR(A.PAGE\_URL, 9),'/',-1)-1))

ELSE lower(A.PAGE\_URL)

END)= B.PAGE\_COMPL\_URL AND A.DIM\_PAGE\_URL\_ID is null;

* **INSERT INTO V2DIGT2 FROM V2DIGT1**

INSERT INTO V2DIGT2.FCT\_GOOGLE\_SRCH\_KEYWRD\_URL SELECT

URN\_REG\_DIM\_URL

,SRCH\_KEYWRD

,PAGE\_URL

,DIM\_PAGE\_URL\_ID

,SRC\_CREATE\_TS

,CLICK\_CNT

,IMPR\_CNT

,cast(CLICK\_THRU\_RATE as NUMERIC(5,4)) AS CLICK\_THRU\_RATE

,RANK\_POSITN

,CTRY\_ALPHA3\_COMPLNC\_CD

,SRC\_CREATE\_DT

,DATA\_SRC\_CD

,CREATE\_TS

,UPDT\_TS

FROM V2DIGT1.FCT\_GOOGLE\_SRCH\_KEYWRD\_URL;

* **DB2\_FCT\_GOOGLE\_SRCH\_URL.sql**
* **Extract the rows from distinct PAGE\_COMPL\_URL**

DECLARE GLOBAL TEMPORARY TABLE SESSION.MPWEXT\_L\_V\_DIM\_URL\_HDP\_TEMP\_SEO\_GOOGLE\_URL\_1\_TEMP AS

(

SELECT DISTINCT(PAGE\_COMPL\_URL),DIM\_PAGE\_URL\_ID,URN\_REG\_DIM\_URL,UPDT\_TS FROM V2DIGT2.V\_DIM\_URL

) WITH DATA ORGANIZE BY ROW NOT LOGGED;

* **Extract the latest updated rows alone from the above table**

DECLARE GLOBAL TEMPORARY TABLE SESSION.MPWEXT\_L\_V\_DIM\_URL\_HDP\_TEMP\_SEO\_GOOGLE\_URL\_1 AS

(

select \* from (select f.PAGE\_COMPL\_URL,f.DIM\_PAGE\_URL\_ID,f.URN\_REG\_DIM\_URL,row\_number() OVER (PARTITION BY PAGE\_COMPL\_URL ORDER BY UPDT\_TS DESC) as rn from SESSION.MPWEXT\_L\_V\_DIM\_URL\_HDP\_TEMP\_SEO\_GOOGLE\_URL\_1\_TEMP f) L where L.rn=1

) WITH DATA ORGANIZE BY ROW NOT LOGGED;

* **Remove the substrings ['%/index.html','/?lnk','//'] from PAGE\_COMPL\_URL if present from previous table**

DECLARE GLOBAL TEMPORARY TABLE SESSION.MPWEXT\_L\_V\_DIM\_URL\_HDP\_TEMP\_SEO\_GOOGLE\_URL\_2\_TEMP as

(

select

(CASE

WHEN A.PAGE\_COMPL\_URL LIKE '%/index.html' THEN SUBSTR(A.PAGE\_COMPL\_URL, 1,(LENGTH(A.PAGE\_COMPL\_URL) - 11))

WHEN A.PAGE\_COMPL\_URL LIKE '%/?lnk%' THEN SUBSTR(A.PAGE\_COMPL\_URL, 1,instr(SUBSTR(A.PAGE\_COMPL\_URL, 9),'/?lnk')-1)

WHEN A.PAGE\_COMPL\_URL LIKE '%//' THEN SUBSTR(A.PAGE\_COMPL\_URL, 1,(LENGTH(A.PAGE\_COMPL\_URL)-2))

ELSE A.PAGE\_COMPL\_URL

END) as PAGE\_COMPL\_URL,A.DIM\_PAGE\_URL\_ID,A.URN\_REG\_DIM\_URL from SESSION.MPWEXT\_L\_V\_DIM\_URL\_HDP\_TEMP\_SEO\_GOOGLE\_URL\_1 A

) WITH DATA ORGANIZE BY ROW NOT LOGGED;

* **Keep the records with largest DIM\_PAGE\_URL\_ID from previous table**

DECLARE GLOBAL TEMPORARY TABLE SESSION.MPWEXT\_L\_V\_DIM\_URL\_HDP\_TEMP\_SEO\_GOOGLE\_URL\_2 as

(

select \* from (select f.PAGE\_COMPL\_URL,f.DIM\_PAGE\_URL\_ID,f.URN\_REG\_DIM\_URL,row\_number() OVER (PARTITION BY PAGE\_COMPL\_URL ORDER BY DIM\_PAGE\_URL\_ID DESC) as rn from SESSION.MPWEXT\_L\_V\_DIM\_URL\_HDP\_TEMP\_SEO\_GOOGLE\_URL\_2\_TEMP f) L where L.rn=1

) WITH DATA ORGANIZE BY ROW NOT LOGGED;

* **Remove ‘/’ from the end of the PAGE\_COMPL\_URL**

DECLARE GLOBAL TEMPORARY TABLE SESSION.MPWEXT\_L\_V\_DIM\_URL\_HDP\_TEMP\_SEO\_GOOGLE\_URL\_3\_TEMP as

(

select

(CASE

WHEN A.PAGE\_COMPL\_URL LIKE '%/' THEN SUBSTR(A.PAGE\_COMPL\_URL, 1,(LENGTH(A.PAGE\_COMPL\_URL)-1))

ELSE A.PAGE\_COMPL\_URL

END) as PAGE\_COMPL\_URL,A.DIM\_PAGE\_URL\_ID,A.URN\_REG\_DIM\_URL from SESSION.MPWEXT\_L\_V\_DIM\_URL\_HDP\_TEMP\_SEO\_GOOGLE\_URL\_2 A

) WITH DATA ORGANIZE BY ROW NOT LOGGED;

* **Keep the records with largest DIM\_PAGE\_URL\_ID from previous table**

DECLARE GLOBAL TEMPORARY TABLE SESSION.MPWEXT\_L\_V\_DIM\_URL\_HDP\_TEMP\_SEO\_GOOGLE\_URL\_3 as

(

select \* from (select f.PAGE\_COMPL\_URL,f.DIM\_PAGE\_URL\_ID,f.URN\_REG\_DIM\_URL,row\_number() OVER (PARTITION BY PAGE\_COMPL\_URL ORDER BY DIM\_PAGE\_URL\_ID DESC) as rn from SESSION.MPWEXT\_L\_V\_DIM\_URL\_HDP\_TEMP\_SEO\_GOOGLE\_URL\_3\_TEMP f) L where L.rn=1

) WITH DATA ORGANIZE BY ROW NOT LOGGED;

* **Remove substrings ['http://','https://'] from PAGE\_URL if present from L\_FCT\_GOOGLE\_SRCH\_URL\_W3 and store in temp table**

DECLARE GLOBAL TEMPORARY TABLE SESSION.MPWEXT\_L\_FCT\_GOOGLE\_SRCH\_URL\_W3 as

(

select PAGE\_URL,

(CASE

WHEN PAGE\_URL LIKE 'http://%' THEN SUBSTR(PAGE\_URL, 8)

WHEN PAGE\_URL LIKE 'https://%' THEN SUBSTR(PAGE\_URL, 9)

ELSE PAGE\_URL

END) as PAGE\_URL\_TEMP,

SRC\_FILE\_CREATE\_DT,CLICKS,IMPR,CTR,POSITN from V2DIGT0.L\_FCT\_GOOGLE\_SRCH\_URL\_W3

) WITH DATA ORGANIZE BY ROW NOT LOGGED;

* **Remove substrings ['http://','https://'] from PAGE\_URL if present from V2DIGT0.L\_FCT\_GOOGLE\_SRCH\_URL\_DOMAIN and store in temp table**

DECLARE GLOBAL TEMPORARY TABLE SESSION.MPWEXT\_L\_FCT\_GOOGLE\_SRCH\_URL\_DOMAIN as

(

select PAGE\_URL,

(CASE

WHEN PAGE\_URL LIKE 'http://%' THEN SUBSTR(PAGE\_URL, 8)

WHEN PAGE\_URL LIKE 'https://%' THEN SUBSTR(PAGE\_URL, 9)

ELSE PAGE\_URL

END) as PAGE\_URL\_TEMP,

SRC\_FILE\_CREATE\_DT,CLICKS,IMPR,CTR,POSITN from V2DIGT0.L\_FCT\_GOOGLE\_SRCH\_URL\_DOMAIN

) WITH DATA ORGANIZE BY ROW NOT LOGGED;

* **Insert the rows to SCHEMA1 table,from MPWEXT\_L\_FCT\_GOOGLE\_SRCH\_URL\_W3 include latest updated DIM\_PAGE\_URL\_ID of the corresponding PAGE\_URL**

insert into V2DIGT1.FCT\_GOOGLE\_SRCH\_URL(

URN\_REG\_DIM\_URL

,PAGE\_URL

,DIM\_PAGE\_URL\_ID

,SRC\_CREATE\_TS

,CLICK\_CNT

,IMPR\_CNT

,CLICK\_THRU\_RATE

,AVG\_RANK\_POSITN

,SRC\_FILE\_CREATE\_DT

,DATA\_SRC\_CD

,CREATE\_TS

,UPDT\_TS

) select

B.URN\_REG\_DIM\_URL,

A.PAGE\_URL,

B.DIM\_PAGE\_URL\_ID,

cast(A.SRC\_FILE\_CREATE\_DT as timestamp) as SRC\_CREATE\_TS,

A.CLICKS,

A.IMPR,

A.CTR,

A.POSITN,

A.SRC\_FILE\_CREATE\_DT,

'GSC' AS DATA\_SRC\_CD,

CURRENT\_TIMESTAMP AS CREATE\_TS,

CURRENT\_TIMESTAMP AS UPDT\_TS

from SESSION.MPWEXT\_L\_FCT\_GOOGLE\_SRCH\_URL\_W3 A

LEFT JOIN SESSION.MPWEXT\_L\_V\_DIM\_URL\_HDP\_TEMP\_SEO\_GOOGLE\_URL\_1 B ON lower(A.PAGE\_URL\_TEMP)=B.PAGE\_COMPL\_URL

;

* **Insert the rows to SCHEMA1 table,from MPWEXT\_L\_FCT\_GOOGLE\_SRCH\_URL\_DOMAIN where, page url is NULL and include latest updated DIM\_PAGE\_URL\_ID.**

insert into V2DIGT1.FCT\_GOOGLE\_SRCH\_URL(

URN\_REG\_DIM\_URL

,PAGE\_URL

,DIM\_PAGE\_URL\_ID

,SRC\_CREATE\_TS

,CLICK\_CNT

,IMPR\_CNT

,CLICK\_THRU\_RATE

,AVG\_RANK\_POSITN

,SRC\_FILE\_CREATE\_DT

,DATA\_SRC\_CD

,CREATE\_TS

,UPDT\_TS

) select

C.URN\_REG\_DIM\_URL,

A.PAGE\_URL,

C.DIM\_PAGE\_URL\_ID,

cast(A.SRC\_FILE\_CREATE\_DT as timestamp) as SRC\_CREATE\_TS,

A.CLICKS,

A.IMPR,

A.CTR,

A.POSITN,

A.SRC\_FILE\_CREATE\_DT,

'GSC' AS DATA\_SRC\_CD,

CURRENT\_TIMESTAMP AS CREATE\_TS,

CURRENT\_TIMESTAMP AS UPDT\_TS

from SESSION.MPWEXT\_L\_FCT\_GOOGLE\_SRCH\_URL\_DOMAIN A

LEFT JOIN SESSION.MPWEXT\_L\_FCT\_GOOGLE\_SRCH\_URL\_W3 B ON A.PAGE\_URL=B.PAGE\_URL

LEFT JOIN SESSION.MPWEXT\_L\_V\_DIM\_URL\_HDP\_TEMP\_SEO\_GOOGLE\_URL\_1 C ON lower(A.PAGE\_URL\_TEMP)=C.PAGE\_COMPL\_URL

where B.PAGE\_URL IS NULL

;

* **LEVEL 1 UPDATE**

UPDATE V2DIGT1.FCT\_GOOGLE\_SRCH\_URL A SET A.URN\_REG\_DIM\_URL=B.URN\_REG\_DIM\_URL,A.DIM\_PAGE\_URL\_ID=B.DIM\_PAGE\_URL\_ID FROM SESSION.MPWEXT\_L\_V\_DIM\_URL\_HDP\_TEMP\_SEO\_GOOGLE\_URL\_1 B WHERE

(CASE WHEN A.PAGE\_URL LIKE 'http://%/' THEN lower(SUBSTR(SUBSTR(A.PAGE\_URL, 8), 1,LENGTH(SUBSTR(A.PAGE\_URL, 8))-1))

WHEN A.PAGE\_URL LIKE 'https://%/' THEN lower(SUBSTR(SUBSTR(A.PAGE\_URL, 9), 1,LENGTH(SUBSTR(A.PAGE\_URL, 8))-1))

WHEN A.PAGE\_URL LIKE 'http://%/index.html' THEN lower(SUBSTR(SUBSTR(A.PAGE\_URL, 8), 1,LENGTH(SUBSTR(A.PAGE\_URL, 8)) - 11))

WHEN A.PAGE\_URL LIKE 'https://%/index.html' THEN lower(SUBSTR(SUBSTR(A.PAGE\_URL, 9), 1,(LENGTH(SUBSTR(A.PAGE\_URL, 9)) - 11)))

WHEN A.PAGE\_URL LIKE 'http://%/?lnk%' THEN lower(SUBSTR(SUBSTR(A.PAGE\_URL, 8), 1,instr(SUBSTR(A.PAGE\_URL, 8),'/?lnk')-1))

WHEN A.PAGE\_URL LIKE 'https://%/?lnk%' THEN lower(SUBSTR(SUBSTR(A.PAGE\_URL, 9), 1,instr(SUBSTR(A.PAGE\_URL, 9),'/?lnk')-1))

ELSE lower(A.PAGE\_URL)

END) = B.PAGE\_COMPL\_URL and A.DIM\_PAGE\_URL\_ID is null;

UPDATE V2DIGT1.FCT\_GOOGLE\_SRCH\_URL A SET A.URN\_REG\_DIM\_URL=B.URN\_REG\_DIM\_URL,A.DIM\_PAGE\_URL\_ID=B.DIM\_PAGE\_URL\_ID FROM SESSION.MPWEXT\_L\_V\_DIM\_URL\_HDP\_TEMP\_SEO\_GOOGLE\_URL\_1 B

WHERE

(CASE WHEN A.PAGE\_URL LIKE 'http://%' THEN lower(SUBSTR(SUBSTR(A.PAGE\_URL, 8), 1,instr(SUBSTR(A.PAGE\_URL, 8),'/',-1)-1))

WHEN A.PAGE\_URL LIKE 'https://%' THEN lower(SUBSTR(SUBSTR(A.PAGE\_URL, 9), 1,instr(SUBSTR(A.PAGE\_URL, 9),'/',-1)-1))

ELSE lower(A.PAGE\_URL)

END)= B.PAGE\_COMPL\_URL AND A.DIM\_PAGE\_URL\_ID is null;

* **LEVEL 2 UPDATE**

UPDATE V2DIGT1.FCT\_GOOGLE\_SRCH\_URL A SET A.URN\_REG\_DIM\_URL=B.URN\_REG\_DIM\_URL,A.DIM\_PAGE\_URL\_ID=B.DIM\_PAGE\_URL\_ID FROM SESSION.MPWEXT\_L\_V\_DIM\_URL\_HDP\_TEMP\_SEO\_GOOGLE\_URL\_2 B WHERE

(CASE WHEN A.PAGE\_URL LIKE 'http://%/' THEN lower(SUBSTR(SUBSTR(A.PAGE\_URL, 8), 1,LENGTH(SUBSTR(A.PAGE\_URL, 8))-1))

WHEN A.PAGE\_URL LIKE 'https://%/' THEN lower(SUBSTR(SUBSTR(A.PAGE\_URL, 9), 1,LENGTH(SUBSTR(A.PAGE\_URL, 8))-1))

WHEN A.PAGE\_URL LIKE 'http://%/index.html' THEN lower(SUBSTR(SUBSTR(A.PAGE\_URL, 8), 1,LENGTH(SUBSTR(A.PAGE\_URL, 8)) - 11))

WHEN A.PAGE\_URL LIKE 'https://%/index.html' THEN lower(SUBSTR(SUBSTR(A.PAGE\_URL, 9), 1,(LENGTH(SUBSTR(A.PAGE\_URL, 9)) - 11)))

WHEN A.PAGE\_URL LIKE 'http://%/?lnk%' THEN lower(SUBSTR(SUBSTR(A.PAGE\_URL, 8), 1,instr(SUBSTR(A.PAGE\_URL, 8),'/?lnk')-1))

WHEN A.PAGE\_URL LIKE 'https://%/?lnk%' THEN lower(SUBSTR(SUBSTR(A.PAGE\_URL, 9), 1,instr(SUBSTR(A.PAGE\_URL, 9),'/?lnk')-1))

ELSE lower(A.PAGE\_URL)

END) = B.PAGE\_COMPL\_URL and A.DIM\_PAGE\_URL\_ID is null;

UPDATE V2DIGT1.FCT\_GOOGLE\_SRCH\_URL A SET A.URN\_REG\_DIM\_URL=B.URN\_REG\_DIM\_URL,A.DIM\_PAGE\_URL\_ID=B.DIM\_PAGE\_URL\_ID FROM SESSION.MPWEXT\_L\_V\_DIM\_URL\_HDP\_TEMP\_SEO\_GOOGLE\_URL\_2 B

WHERE

(CASE WHEN A.PAGE\_URL LIKE 'http://%' THEN lower(SUBSTR(SUBSTR(A.PAGE\_URL, 8), 1,instr(SUBSTR(A.PAGE\_URL, 8),'/',-1)-1))

WHEN A.PAGE\_URL LIKE 'https://%' THEN lower(SUBSTR(SUBSTR(A.PAGE\_URL, 9), 1,instr(SUBSTR(A.PAGE\_URL, 9),'/',-1)-1))

ELSE lower(A.PAGE\_URL)

END)= B.PAGE\_COMPL\_URL AND A.DIM\_PAGE\_URL\_ID is null;

* **LEVEL 3 UPDATE**

UPDATE V2DIGT1.FCT\_GOOGLE\_SRCH\_URL A SET A.URN\_REG\_DIM\_URL=B.URN\_REG\_DIM\_URL,A.DIM\_PAGE\_URL\_ID=B.DIM\_PAGE\_URL\_ID FROM SESSION.MPWEXT\_L\_V\_DIM\_URL\_HDP\_TEMP\_SEO\_GOOGLE\_URL\_3 B WHERE

(CASE WHEN A.PAGE\_URL LIKE 'http://%/' THEN lower(SUBSTR(SUBSTR(A.PAGE\_URL, 8), 1,LENGTH(SUBSTR(A.PAGE\_URL, 8))-1))

WHEN A.PAGE\_URL LIKE 'https://%/' THEN lower(SUBSTR(SUBSTR(A.PAGE\_URL, 9), 1,LENGTH(SUBSTR(A.PAGE\_URL, 8))-1))

WHEN A.PAGE\_URL LIKE 'http://%/index.html' THEN lower(SUBSTR(SUBSTR(A.PAGE\_URL, 8), 1,LENGTH(SUBSTR(A.PAGE\_URL, 8)) - 11))

WHEN A.PAGE\_URL LIKE 'https://%/index.html' THEN lower(SUBSTR(SUBSTR(A.PAGE\_URL, 9), 1,(LENGTH(SUBSTR(A.PAGE\_URL, 9)) - 11)))

WHEN A.PAGE\_URL LIKE 'http://%/?lnk%' THEN lower(SUBSTR(SUBSTR(A.PAGE\_URL, 8), 1,instr(SUBSTR(A.PAGE\_URL, 8),'/?lnk')-1))

WHEN A.PAGE\_URL LIKE 'https://%/?lnk%' THEN lower(SUBSTR(SUBSTR(A.PAGE\_URL, 9), 1,instr(SUBSTR(A.PAGE\_URL, 9),'/?lnk')-1))

ELSE lower(A.PAGE\_URL)

END) = B.PAGE\_COMPL\_URL and A.DIM\_PAGE\_URL\_ID is null;

UPDATE V2DIGT1.FCT\_GOOGLE\_SRCH\_URL A SET A.URN\_REG\_DIM\_URL=B.URN\_REG\_DIM\_URL,A.DIM\_PAGE\_URL\_ID=B.DIM\_PAGE\_URL\_ID FROM SESSION.MPWEXT\_L\_V\_DIM\_URL\_HDP\_TEMP\_SEO\_GOOGLE\_URL\_3 B

WHERE

(CASE WHEN A.PAGE\_URL LIKE 'http://%' THEN lower(SUBSTR(SUBSTR(A.PAGE\_URL, 8), 1,instr(SUBSTR(A.PAGE\_URL, 8),'/',-1)-1))

WHEN A.PAGE\_URL LIKE 'https://%' THEN lower(SUBSTR(SUBSTR(A.PAGE\_URL, 9), 1,instr(SUBSTR(A.PAGE\_URL, 9),'/',-1)-1))

ELSE lower(A.PAGE\_URL)

END)= B.PAGE\_COMPL\_URL AND A.DIM\_PAGE\_URL\_ID is null;

* **INSERT INTO V2DIGT2 FROM V2DIGT1**

INSERT INTO V2DIGT2.FCT\_GOOGLE\_SRCH\_URL SELECT

URN\_REG\_DIM\_URL

,PAGE\_URL

,DIM\_PAGE\_URL\_ID

,SRC\_CREATE\_TS

,CLICK\_CNT

,IMPR\_CNT

,cast(CLICK\_THRU\_RATE as NUMERIC(5,4)) AS CLICK\_THRU\_RATE

,AVG\_RANK\_POSITN

,SRC\_FILE\_CREATE\_DT

,DATA\_SRC\_CD

,CREATE\_TS

,UPDT\_TS

FROM V2DIGT1.FCT\_GOOGLE\_SRCH\_URL;

* Check whether the process is successful and close the connection
* Check logs and send email