

Usar Branch “tarea”

BigData

Tarea

Unidad 2

Carga de datos “raw” en esquema publico

public

Aggregates

Collations

Domains

FTS Configurations

FTS Dictionaries

FTS Parsers

FTS Templates

Foreign Tables

Functions

Materialized Views

Operators

Procedures

Sequences

Tables (1)

batch_first_load

Columns (16)

person_name

user_name

email

personal_number

birth_date

address

phone

mac_address

ip_address

iban

accessed_at

session_duration

download_speed

upload_speed

consumed_traffic

unique_id

Query

Query History

1 SELECT

2 *

3 FROM

4 --golden_layer.pii_data

5 public.batch_first_load

6 --bronze_layer.batch_first_load

7 --WHERE

8 --unique_id = '830ff059-0526-4ec4-bbfd-850d6f318cff'

9 LIMIT 100

Scratch Pad

SELECT

person_name,

REGEXP_REPLACE(person_name, '^(ing\.)s*', ', i') AS

normalized_name,

unique_id

FROM silver_layer.batch_first_load

WHERE person_name ~* '^(ing?);'

Data Output

Messages

Notifications

Showing rows: 1 to 100

Page No: 1 of 1

	person_name character varying (60)	user_name character varying (60)	email character varying (100)	personal_number numeric	birth_date date	address character varying (250)
1	Ángela Isabel Cortez	ángelaisabelcortez	ángelaisabelcortez@hotmail.com	23340481821	2021-12-13	Callejón Guanajuato 675 Edif. 862 , Depto. 60
2	Abril Dulce María Cepeda Prado	abrildulcemariacepedaprado	abrildulcemariacepedaprado@gmail.com	54424328794	2017-03-23	Circunvalación Oaxaca 105 122, San Rubén I
3	Jerónimo Esperanza Casillas Villarreal	jerónimoesperanzacasillasvillarreal	jerónimoesperanzacasillasvillarreal@gmail.com	61143398545	1932-06-18	Eje vial Santa Lucía 866 Interior 357, San Cla
4	Gerónimo Guevara	gerónimoguevara	gerónimoguevara@gmail.com	38872377759	1949-01-15	Periférico Guevara 244 Interior 340, Vieja Ger
5	Irma Blanca Montes Monroy	irmablancamontesmonroy	irmablancamontesmonroy@yahoo.com	80074524176	1973-05-11	Eje vial de Anda 032 Edif. 824 , Depto. 028, V
6	Cristal Hernán Nájera Estrada	cristalhernánnájeraestrada	cristalhernánnájeraestrada@hotmail.com	34743610791	2004-10-14	Cerrada Morelos 246 004, San Cristal los alto
7	Reynaldo Hugo Jiménez Oquendo	reynaldohugojimínezquendo	reynaldohugojimínezquendo@hotmail.com	9374443233	1983-01-22	Cerrada Haití 207 Edif. 951 , Depto. 295, San
8	Citlali Saiz Solano	citlalisaisolano	citlalisaisolano@yahoo.com	38059902270	1942-02-02	Continuación Sur Casas 989 Edif. 895 , Depto
9	Guadalupe Paredes	guadalupeparedes	guadalupeparedes@gmail.com	85496915993	2004-09-09	Callejón México 564 Edif. 838 , Depto. 670, S
10	Alberto Covarrubias Chacón	albertocovarrubiaschacón	albertocovarrubiaschacón@gmail.com	25575392417	1986-12-13	Andador Brunei Darussalam 794 354, San Cc
11	Sr(a). Isaac Villalpando	sr(a).isaacvillalpando	sr(a).isaacvillalpando@yahoo.com	73446861012	2005-11-18	Peatonal Rodríguez 042 Interior 554, San Ad
12	Sr(a). Esperanza Puga	sr(a).esperanzapuga	sr(a).esperanzapuga@gmail.com	41850815089	1931-01-18	Pasaje Ortega 292 578, San Gustavo los alto

Creación de Esquemas bronce y normalización

- Se remueven duplicados y datos NULL
- Se remueven la combinación de correos y nombres de persona duplicados
- Títulos en nombre, “Ing.”, “Mtro”, “Lic.” etc.
- Acentos en correos

Query	Query History	Scratch Pad
1 2 3 4 5 6 7 8 9	<pre>SELECT * FROM --golden_layer.pii_data --public.batch_first_load bronze_layer.batch_first_load --WHERE --unique_id = '830ff059-0526-4ec4-bbfd-850d6f318cff' LIMIT 100</pre>	<pre>SELECT person_name, REGEXP_REPLACE(person_name, '(Ing\.)\s*', ', 'i) AS normalized_name, unique_id FROM silver_layer.batch_first_load WHERE person_name ~* '^(Ing°)';</pre>
Data Output	Messages	Notifications
Showing rows: 1 to 100 Page No: 1 of 1		
person_name character varying (60)	user_name character varying (60)	email character varying (100)
1 Irma Blanca Montes Monroy	irmablancamontesmonroy	irmablancamontesmonroy@yahoo.com
2 Citlali Saiz Solano	citlalisaisolano	citlalisaisolano@yahoo.com
3 Guadalupe Paredes	guadalupeparedes	guadalupeparedes@gmail.com
4 Aurelio Eloisa Loera	aurelioeloisaloera	aurelioeloisaloera@hotmail.com
5 Mariano Maldonado Malave	marianomaldonadomalave	marianomaldonadomalave@yahoo.com
6 Daniela Diana Escobedo Toledo	danieladianaescobedotoledo	danieladianaescobedotoledo@yahoo.com
7 Ángela Isabel Cortez	angelaisabelcortez	angelaisabelcortez@hotmail.com
8 Abril Dulce María Cepeda Prado	abrildulcemariacepedaprado	abrildulcemariacepedaprado@gmail.com
9 Jerónimo Esperanza Casillas Villarreal	jeronimoesperanzacasillasvillarreal	jeronimoesperanzacasillasvillarreal@gmail.com
10 Gerónimo Guevara	geronimoguevara	geronimoguevara@gmail.com
11 Cristal Hernán Nájera Estrada	cristalhernannajeraestrada	cristalhernannajeraestrada@hotmail.com
12 Reynaldo Hugo Jiménez Oquendo	reynaldohugojiminezquendo	reynaldohugojiminezquendo@hotmail.com
13 Alberto Covarrubias Chacón	albertocovarrubiaschacón	albertocovarrubiaschacon@gmail.com
14 Zoé Madrid	zoemadrid	zoemadrid@hotmail.com
15 Mauro Mojica	mauromojica	mauromojica@gmail.com
16 Pascual Jorge Almanza	pascualjorgealmanza	pascualjorgealmanza@gmail.com
17 Nelly Zamudio Valles	nellyzamudiovalles	nellyzamudiovalles@gmail.com
18 Luis Lira	luislira	luislira@hotmail.com
19		

Creación de capa Silver con nuevas tablas en esquema estrella

Processes x public.batch_first_L... x public.batch_first_L... x silver_layer.dim_pe... x bookstonedrive_db/pos

bookstonedrive_db/postgres@BookstoneDriven

Query Query History

```
1 SELECT
2 *
3 FROM
4 --golden_layer.pii_data
5 --public.batch_first_load
6 --bronze_layer.batch_first_load
7 silver_layer.fact_network_usage
8 --WHERE
9 --unique_id = '830ff059-0526-4ec4-bbfd-850d6f318cff'
```

Data Output Messages Notifications

Showing rows:

	unique_id character varying (100)	session_duration integer	download_speed integer	upload_speed integer	consumed_traffic integer
1	f2bd39fe-3cfb-4ea7-aaa9-117cdcd2c065	32192	511	474	1791882
2	53f6ac0e-38b0-4042-a1a2-86c8b170b8d...	33128	909	738	1492565
3	b47a8fcf-9510-4daa-a78d-17093759c6...	24921	325	481	1031638
4	d5b8fde4-9683-4a55-a6d4-a6707defb7...	27274	143	616	718639
5	d958fde0-b71e-4a1b-9d57-529d5cc343...	26280	93	207	151138
6	bd91cd07-0b72-44cd-b5cb-a3c9dc0f9a...	7471	206	431	343194
7	fb91debd-8f61-4c13-8639-e60a5be350...	25950	353	462	1433304
8	8a5cd440-15d8-40e7-9897-647fe2d3d2...	13305	287	216	1582589
9	5a883643-c859-46a8-8c8a-ac01fb5edd...	29141	980	588	708929
10	c134c1f9-ad29-4015-9650-e6dca4ff58ce	18972	297	486	1932306
11	d073b238-9308-4950-a78e-09a7c3416...	10157	502	525	407707

Query Query History

```
1 SELECT
2 *
3 FROM
4 --golden_layer.pii_data
5 --public.batch_first_load
6 --bronze_layer.batch_first_load
7 silver_layer.dim_date
8 --WHERE
9 --unique_id = '830ff059-0526-4ec4-bbfd-850d6f318cff'
```

Data Output Messages Notifications

unique_id character varying (100)	accessed_at time without time zone
1 f2bd39fe-3cfb-4ea7-aaa9-117cdcd2c065	15:50:29
2 53f6ac0e-38b0-4042-a1a2-86c8b170b8d	06:55:37
3 b47a8fcf-9510-4daa-a78d-17093759c6d7	01:43:05
4 d5b8fde4-9683-4a55-a6d4-a6707defb711	01:06:47
5 d958fde0-b71e-4a1b-9d57-529d5cc3438a	19:46:50
6 bd91cd07-0b72-44cd-b5cb-a3c9dc0f9a3f	14:19:15
7 fb91debd-8f61-4c13-8639-e60a5be3504e	01:38:07
8 8a5cd440-15d8-40e7-9897-647fe2d3d24f	23:55:46
9 5a883643-c859-46a8-8c8a-ac01fb5edd57	13:26:36
10 c134c1f9-ad29-4015-9650-e6dca4ff58ce	09:47:18
11 d073b238-9308-4950-a78e-09a7c34166b3	20:50:47
12 decef8a7-8a8d-4db3-8560-5e8fcb93d3af	01:30:43
13 c711d388-704e-4635-8831-46c98737ca4d	17:57:07

silver_layer

- > Aggregates
- > Collations
- > Domains
- > FTS Configurations
- > FTS Dictionaries
- > FTS Parsers
- > FTS Templates
- > Foreign Tables
- > Functions
- > Materialized Views
- > Operators
- > Procedures
- > 1.3 Sequences
- ✓ Tables (6)
 - > batch_first_load
 - > dim_address
 - > dim_date
 - > dim_finance
 - > dim_person
 - > fact_network_usage
- > Trigger Functions

Creación de esquema Golden y Tablas

golden_layer

Aggregates

Collations

Domains

FTS Configurations

FTS Dictionaries

FTS Parsers

FTS Templates

Foreign Tables

Functions

Materialized Views

Operators

Procedures

Sequences

Tables (5)

batch_first_load

non_pii_data

payment_data

pii_data

technical_data

Query

Query History

Scratch Pad

SELECT
*
FROM
--public_batch_first_load
--bronze_layer_batch_first_load
--silver_layer_fact_network_usage
golden_layer.non_pii_data
--WHERE
--unique_id = '830ff059-0526-4ec4-bbfd-850d6f318cff'
LIMIT 100

SELECT
person_name
REGEXP_REPLACE(person_name, '(ing\.)us', ',') AS
normalized_name
unique_id
FROM silver_layer.batch_first_load
WHERE person_name ~* '(ing*)';

SELECT
*
FROM
--golden_layer.pii_data
--public_batch_first_load
--bronze_layer_batch_first_load
silver_layer.fact_network_usage
--WHERE
--unique_id = '830ff059-0526-4ec4-bbfd-
850d6f318cff'
LIMIT 100

Data Output

Messages

Notifications

Showing rows: 1 to 100

Page No: 1 of 1

person_name	user_name	email	personal_number	birth_date	address	phone	mac_address	ip_address	iban	accessed_at	session_duration
MASKED	lic.s****	lic.s****	***MASKED***	***MASKED***	***MASKED***	***MASKED***	94.5a****	68.12****	GB83E****	20:16:26	
MASKED	aidag****	aidag****	***MASKED***	***MASKED***	***MASKED***	***MASKED***	c4.58****	59.39****	GB98J****	01:51:34	
MASKED	josef****	josef****	***MASKED***	***MASKED***	***MASKED***	***MASKED***	02:2d****	103.1****	GB10M****	17:45:59	
MASKED	mayte****	mayte****	***MASKED***	***MASKED***	***MASKED***	***MASKED***	06:d****	158.1****	GB27X****	11:57:33	
MASKED	karia****	karia****	***MASKED***	***MASKED***	***MASKED***	***MASKED***	ca.02****	115.4****	GB49X****	22:16:32	
MASKED	rubyc****	rubyc****	***MASKED***	***MASKED***	***MASKED***	***MASKED***	e8.3a****	36.94****	GB66Q****	06:42:17	
MASKED	espar****	espar****	***MASKED***	***MASKED***	***MASKED***	***MASKED***	20:d****	178.1****	GB85P****	01:59:27	

Query

Query History

Scratch Pad

SELECT
*
FROM
--public_batch_first_load
--bronze_layer_batch_first_load
--silver_layer_fact_network_usage
golden_layer.technical_data
--WHERE
--unique_id = '830ff059-0526-4ec4-bbfd-850d6f318cff'
LIMIT 100

SELECT
person_name
REGEXP_REPLACE(person_name, '(ing\.)us', ',') AS
normalized_name
unique_id
FROM silver_layer.batch_first_load
WHERE person_name ~* '(ing*)';

SELECT
*
FROM
--golden_layer.pii_data
--public_batch_first_load
--bronze_layer_batch_first_load
silver_layer.fact_network_usage
--WHERE
--unique_id = '830ff059-0526-4ec4-bbfd-
850d6f318cff'
LIMIT 100

Data Output

Messages

Notifications

Showing rows: 1 to 100

Page No: 1 of 1

	mac_address	ip_address	download_speed	upload_speed	min_session_duration	technical_issue
1	e0:9b:f3:b5:7b:06	128.140.196.219	511	474	536.5	false
2	fc:5b:80:14:3c:01	77.38.29.91	909	738	552.1	false
3	ee:1e:1b:10:71:37	81.188.174.113	93	207	438.0	false
4	e6:4b:b2:9c:3a:ae	26.47.173.119	353	462	432.5	false
5	fa:0d:0d:93:95:32	194.251.243.147	747	786	368.4	false
6	22:cd:2a:47:cd:ec	80.55.98.108	716	117	506.4	false
7	86:54:93:76:e2:5e	117.126.125.88	49	200	568.2	true
8	2a:ed:7c:63:a4:d1	216.250.77.180	64	686	492.3	false

Enrique Rodriguez Toscano