

Modelagem e Extração

Por Gustavo Simões



work flow



Dataprep

Para reorganizar, analisar os dados de forma automática e inserindo os dados no BigQuery.



BigQuery

Modelagem de tabelas temporárias, fixas e consultas.



Colab

Pipeline, consulta de queries e automatização de extração.



Bucket GCP

Armazenamento em cloud para backup de arquivos e armazenamento de saídas.



Passos para a extração



1. Dataprep

Atualize o GCP com os novos arquivos gerados que o pipeline irá modelar e inserir as tabelas no BigQuery;

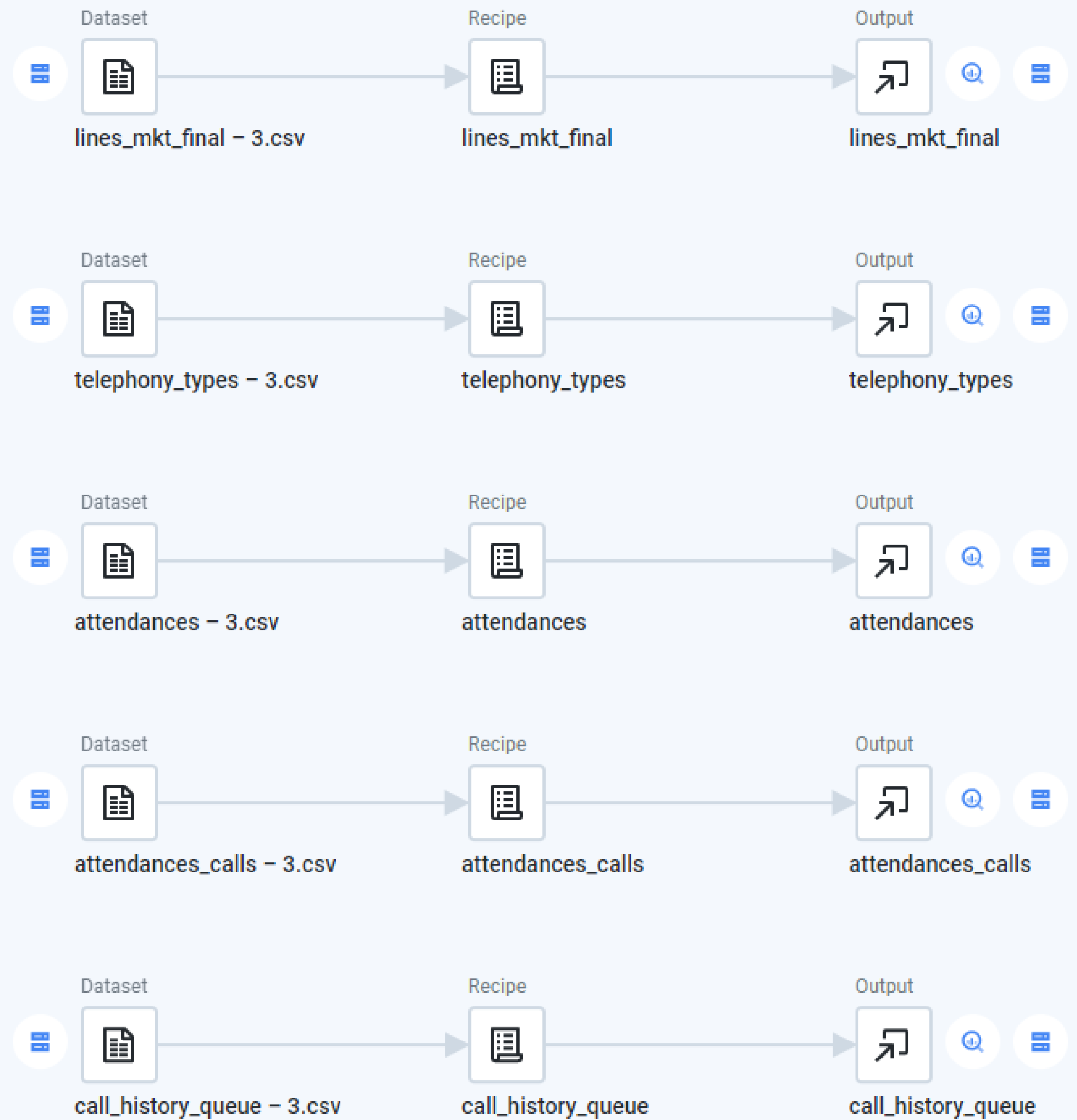
2. BigQuery

O BigQuery já está configurado para atualizar as consultas conforme o parametro necessário.

3. Colab or Python

Anexe a chave (.JSON) no local indicado e execute o código. Ele ira mostrar; as queries feitas, visualização de dados, extração da consulta em formato xlsx e backup no GPC tornando a consulta disponível para download.

Pipelines



lines_mkt_final

pipeline

- 1 Convert row 1 to header
 - 2 Delete "row.names"
 - 3 Lock id type to Integer
 - 4 Lock line_id type to Integer
 - 5 Lock midia type to String
 - 6 Lock campanha type to String
 - 7 Lock fonte type to String
 - 8 Lock pagina type to String
 - 9 Lock dominio type to String
 - 10 Lock destino type to String
 - 11 Lock operacao type to String
 - 12 Lock growth type to String
 - 13 Set fonte to `IFMISSING($col, NULL())`
 - 14 Set pagina to `IFMISSING($col, NULL())`
 - 15 Set dominio to `IFMISSING($col, NULL())`
 - 16 Set destino to `IFMISSING($col, NULL())`
 - 17 Split created_at at 1 positions
 - 18 Split created_at2 at 1 positions
 - 19 Delete created_at3
 - 20 Rename created_at4 to 'created_hour'
 - 21 Rename created_at1 to 'created_at'
-

telephony_types

pipeline

- 1 Convert row 1 to header
- 2 Create id1 from examples of id
- 3 Delete id
- 4 Rename id1 to 'id'
- 5 Delete "row.names"
- 6 Set pattern to `IFMISSING($col, NULL())`
- 7 Lock types type to `String`
- 8 Set types to `IFMISSING($col, NULL())`
- 9 Set created_at to `IFMISSING($col, NULL())`
- 10 Lock pattern type to `String`
- 11 Lock description type to `String`
- 12 Lock id type to `Integer`
- 13 Split created_at at 1 positions
- 14 Split created_at2 at 1 positions
- 15 Split created_at4 at 1 positions
- 16 Delete created_at3
- 17 Delete created_at5
- 18 Rename created_at2 to 'created_hour'
- 19 Rename created_at1 to 'created_at'

attendances

pipeline

- 1 Convert row 1 to header
- 2 Delete "row.names"
- 3 Set id to `NUMFORMAT($col, '#.##', '', '.')`
- 4 Set team_id to `NUMFORMAT($col, '#.##', '', '.')`
- 5 Set customer_id to `NUMFORMAT($col, '#.##', '', '.')`
- 6 Set status_id to `NUMFORMAT($col, '#.##', '', '.')`
- 7 Set type_id to `NUMFORMAT($col, '#.##', '', '.')`
- 8 Lock monthly_value type to Decimal
- 9 Create type_atte from examples of type_id
- 10 Set created_at to `IFMISSING($col, NULL())`
- 11 Lock type_atte type to String
- 12 Lock type_id type to Integer
- 13 Lock status_id type to Integer
- 14 Lock protocol type to String
- 15 Lock customer_id type to Integer
- 16 Lock team_id type to Integer
- 17 Lock id type to Integer
- 18 Split created_at at 1 positions
- 19 Split created_at2 at 1 positions
- 20 Split created_at4 at 1 positions
- 21 Delete created_at3
- 22 Rename created_at1 to 'created_at'
- 23 Rename created_at2 to 'created_hour'
- 24 Delete created_at5

attendances_calls

pipeline



```
1 Convert row 1 to header
2 Delete "row.names"
3 Set id to NUMFORMAT($col, '#.##', ',', '.')
4 Set attendance_id to NUMFORMAT($col, '#.##', ',', '.')
5 Set queue_number to NUMFORMAT($col, '#.##', ',', '.')
6 Set line_id to NUMFORMAT($col, '#.##', ',', '.')
7 Set main_connection to NUMFORMAT($col, '#.##', ',', '.')
8 Create main_connection1 from examples of
  main_connection
9 Rename main_connection1 to 'main_connection_status'
10 Set token to IFMISSING($col, NULL())
11 Set queue_number to IFMISSING($col, NULL())
12 Lock main_connection_status type to String
13 Lock main_connection type to Integer
14 Lock line_id type to Integer
15 Lock queue_number type to Integer
16 Lock phone type to String
17 Lock token type to String
18 Lock attendance_id type to Integer
19 Lock id type to Integer
20 Split created_at at 1 positions
21 Split created_at2 at 1 positions
22 Split created_at4 at 1 positions
23 Delete created_at3
24 Delete created_at5
25 Rename created_at2 to 'created_hour'
26 Rename created_at1 to 'created_at'
```

call_history_queue

pipeline



- 1 Convert row 1 to header
- 2 Delete "row.names"
- 3 Set id to `NUMFORMAT($col, '#.##', '', '.')`
- 4 Set team_id to `NUMFORMAT($col, '#.##', '', '.')`
- 5 Set user_id to `NUMFORMAT($col, '#.##', '', '.')`
- 6 Set queue to `NUMFORMAT($col, '#.##', '', '.')`
- 7 Set lines_id to `NUMFORMAT($col, '#.##', '', '.')`
- 8 Set ddd to `NUMFORMAT($col, '#.##', '', '.')`
- 9 Lock phone type to String
- 10 Lock token type to String
- 11 Lock lines_id type to Integer
- 12 Lock queue type to Integer
- 13 Lock user_id type to Integer
- 14 Set queue_log_verb_types_id to `NUMFORMAT($col, '#.##', '', '.')`
- 15 Lock queue_log_verb_types_id type to Integer
- 16 Set wait to `NUMFORMAT($col, '#.##', '', '.')`
- 17 Set duration to `NUMFORMAT($col, '#.##', '', '.')`
- 18 Set duration_pos_attendance to `NUMFORMAT($col, '#.##', '', '.')`
- 19 Lock wait type to Integer
- 20 Lock duration type to Integer
- 21 Lock duration_pos_attendance type to Integer
- 22 Set queue_log_modality_types_id to `NUMFORMAT($col, '#.##', '', '.')`
- 23 Lock queue_log_modality_types_id type to Integer
- 24 Lock locality type to String
- 25 Lock uf type to String
- 26 Set locality to `IFMISSING($col, NULL())`
- 27 Split created_at at 1 positions
- 28 Split created_at2 at 1 positions
- 29 Split created_at4 at 1 positions
- 30 Delete created_at5
- 31 Delete created_at3
- 32 Rename created_at1 to 'created_at'
- 33 Rename created_at2 to 'created_hour'
- 34 Split created_hour at 1 positions

Estrutura BigQuery

athenas-364914			
Conexões externas			
Consultas salvas (6)			
Consultas do projeto			
LIB_TOTAL			
LIGACOES_NAO_RECPT			
LIGACOES_RECEPT			
PORCENTAGEM_ATENDENT...			
QT_ATENDIMENTOS			
TIKET_MÉDIO			
Athenas2			
Ligacoes_dist			
attendances			
attendances_calls			
call_history_queue			
lines_mkt_final			
telephony_types			

<input type="checkbox"/>	<u>status_id</u>	INTEGER	NULLABLE
<input type="checkbox"/>	<u>type_id</u>	INTEGER	NULLABLE
<input type="checkbox"/>	▼ <u>type_atte</u>	RECORD	NULLABLE
<input type="checkbox"/>	<u>id</u>	INTEGER	NULLABLE
<input type="checkbox"/>	<u>team_id</u>	INTEGER	NULLABLE
<input type="checkbox"/>	<u>customer_id</u>	INTEGER	NULLABLE
<input type="checkbox"/>	<u>protocol</u>	STRING	NULLABLE
<input type="checkbox"/>	<u>status_id</u>	INTEGER	NULLABLE
<input type="checkbox"/>	<u>type_id</u>	INTEGER	NULLABLE
<input type="checkbox"/>	<u>type_atte</u>	STRING	NULLABLE
<input type="checkbox"/>	<u>monthly_value</u>	FLOAT	NULLABLE
<input type="checkbox"/>	<u>created_at</u>	DATETIME	NULLABLE
<input type="checkbox"/>	<u>created_hour</u>	TIME	NULLABLE
<input type="checkbox"/>	<u>monthly_value</u>	FLOAT	NULLABLE
<input type="checkbox"/>	<u>created_at</u>	DATETIME	NULLABLE
<input type="checkbox"/>	<u>attendance_id</u>	INTEGER	NULLABLE

Obrigado!

