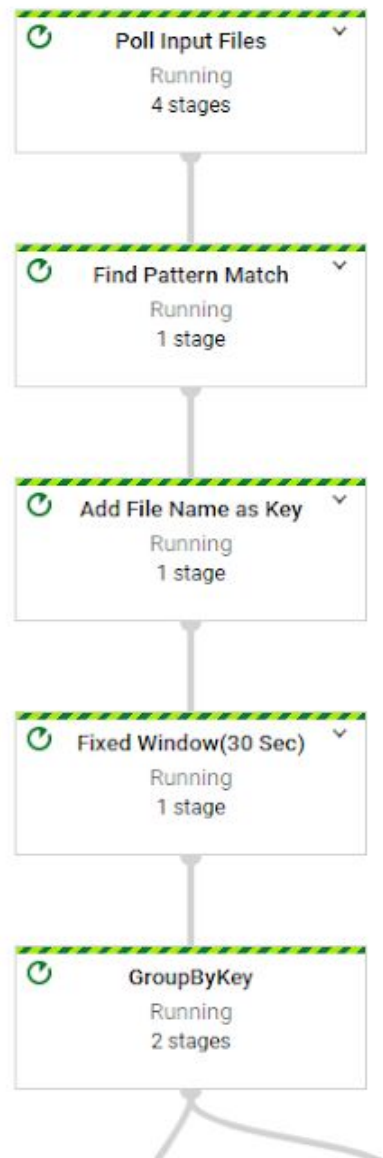
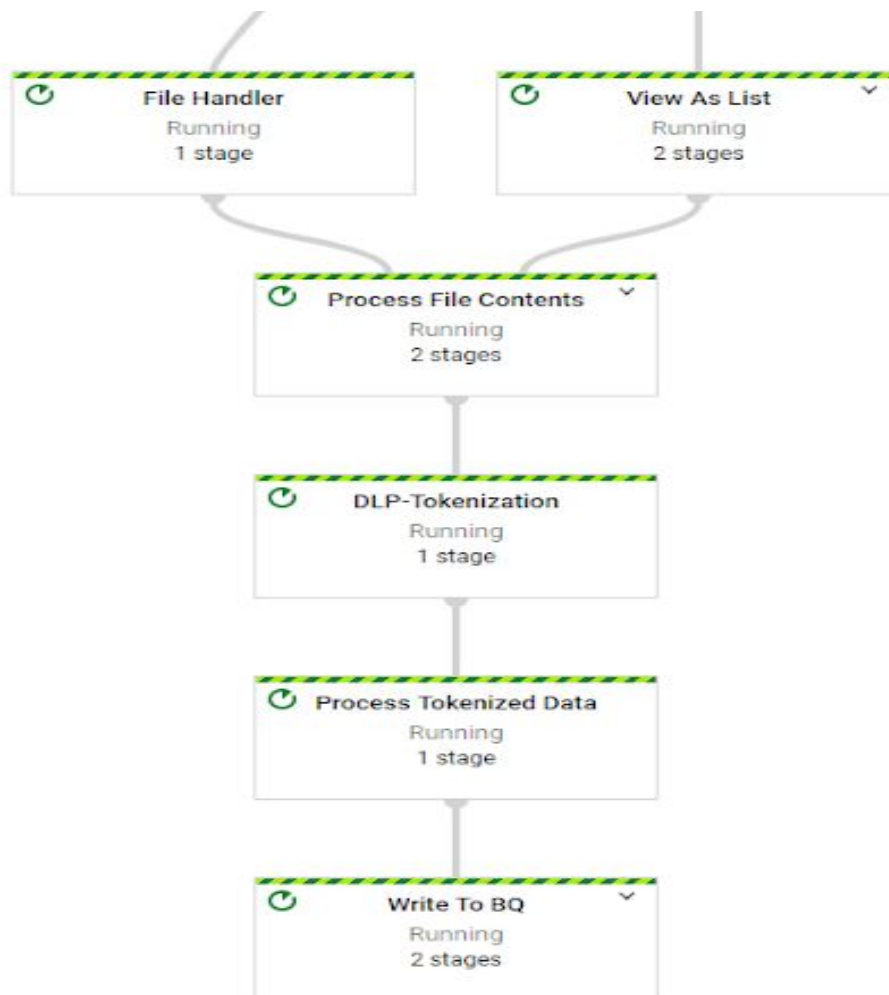


De-identification of PII (personally identifiable information) in large-scale datasets using Cloud DLP

Step 1: Data de-identification streaming pipeline





Queries:

Column details in sample data set

trial status: \$300.00 credit and 91 days remaining - with a full account, you'll get unlimited access to all of Google Cloud Platform.

Google Cloud Platform

My First Project

Search products and resources

Query

FEATURES & INFO

SHORTCUT

ry

ies

queries

is

+ ADD DATA

or your tables and datasets

gility-294813

Query editor

```
1 select table_name, column_name
2 from `deid_dataset.INFORMATION_SCHEMA.COLUMNS`
3 WHERE table_name="CCRecords_1564602825"
```

Run

Save query

Save view

Schedule query

More

Query results

SAVE RESULTS

EXPLORE DATA

Query complete (0.5 sec elapsed, 10 MB processed)

Job information

Results

JSON

Execution details

Row	table_name	column_name
1	CCRecords_1564602825	ID
2	CCRecords_1564602825	Card_Type_Code
3	CCRecords_1564602825	Card_Type_Full_Name
4	CCRecords_1564602825	Issuing_Bank
5	CCRecords_1564602825	Card_Number
6	CCRecords_1564602825	Card_Holders_Name
7	CCRecords_1564602825	Issue_Date
8	CCRecords_1564602825	Expiry_Date
9	CCRecords_1564602825	Billing_Date
10	CCRecords_1564602825	Card_PIN

arch and auto-complete powered by Data Catalog

Niemiec

Enable

Number of rows:

sylvan-agility-294813

Run

Save query

Save view

Schedule query

Query results

SAVE RESULTS

EXPLORE DATA

Query complete (1.3 sec elapsed, 0 B processed)

Job information

Results

JSON

Execution details

Row	number_of_rows
1	100000

Bucketing of the records: Job title

Free trial status: \$300.00 credit and 91 days remaining - with a full account, you'll get unlimited access to all of Google Cloud Platform.

Google Cloud Platform My First Project Search products and resources

BigQuery FEATURES & INFO SHORTCUT

history

queries

story

fers

uled queries

ventions

pine

nces + ADD DATA

arch for your tables and datasets

van-avility-294813

Query editor

```

1 select JobTitle, count(*) as number_of_records_found
2 from `deid_dataset.CCRecords_1564602825`
3 group by JobTitle

```

Run
Save query
Save view
Schedule query
More

Query results SAVE RESULTS EXPLORE DATA

Query complete (3.8 sec elapsed, 0 B processed)

Job information Results JSON Execution details

Row	JobTitle	number_of_records_found
1	Manager	19854
2	Executive	60109
3	Engineer	20037

Bucketing of age data:

FEATURES & INFO
 SHORTCUTS

ies

+ ADD DATA ▾

ur tables and datasets

-294813

Query editor

```

1 select Age, count(*) as number_of_records_found
2 from `deid_dataset.CCRRecords_1564602825`
3 group by Age
4 order by Age desc
          
```

Run ▾
 Save query
 Save view
 Schedule query ▾
 More ▾

Query results

SAVE RESULTS
 EXPLORE DATA ▾

Query complete (1.3 sec elapsed, 0 B processed)

Job information
Results
JSON
Execution details

Row	Age	number_of_records_found
1	60	2417
2	50	23057
3	40	23164
4	30	23239
5	20	28123

Masking transformation for the SSN:

console.cloud.google.com/bigquery?_ga=2.87114648.1065603283.1604600606-11819499.1604296305&_gac=1.152400331.1604600606

Free trial status: \$300.00 credit and 91 days remaining - with a full account, you'll get unlimited access to all of Google Cloud Platform.

Google Cloud Platform My First Project Search products and res

BigQuery FEATURES & INFO SHORTCUT

Query history

Saved queries

Job history

Transfers

Scheduled queries

Reservations

BI Engine

Resources + ADD DATA

Search for your tables and datasets

sylvan-agility-294813

Query editor

```
1 select SSN
2 from `deid_dataset.CCRecords_`
3 where regexp_contains(SSN, "@")
```

Run Save query Save view Schedule query More

Query results SAVE RESULTS EXPLORE DATA

Query complete (1.6 sec elapsed, 0 B processed)

Job information Results JSON Execution details

Row	SSN
1	***_**-2808
2	***_**-9932
3	***_**-7736
4	***_**-1044
5	***_**-6533
6	***_**-5613
7	***_**-4868
8	***_**-1515
9	***_**-2370
10	***_**-0190

deterministic encryption for the card_holders_name and card_number and card_pin entries:

console.cloud.google.com/bigquery?_ga=2.87114648.1065603283.1604600606-11819499.1604296305&_gac=1.152400331.1604666083.Cj0KCQIAhZT98RDmARIsAN2E-J1OG2HObmQQ1Jf0P6aTPpeWlKO

ee trial status: \$300.00 credit and 91 days remaining - with a full account, you'll get unlimited access to all of Google Cloud Platform.

Google Cloud PlatformMy First ProjectSearch products and resources

BigQueryFEATURES & INFOSHORTCUT

HistoryQueriesQuery editor

1 select card_holders_name, card_number, card_pin
2 from `deid_dataset.CCRecords_*`
3 limit 1
4

RunSave querySave viewSchedule queryMore

Query resultsSAVE RESULTSEXPLORE DATA

Query complete (0.8 sec elapsed, 0 B processed)
Job informationResultsJSONExecution details

Row	card_holders_name	card_number	card_pin
1	AcgXjWhsOs6hrjVsZVDGfKsq7legnMjq4qD0cv9N8+hTLw4=	AXRicZFCFNGUv4NF2gaLr+0GXwup3dtBSDmUlGrg/g==	VAGelBGb73HZwDxbu8Ia23yQlurHG5/j1eibTj5vwy0=

InfoType transformation was applied to the Additional Details column:

Row	additional_details
1	[IBAN_CODE] [EMAIL_ADDRESS] 491-756-3047 examiner.com Lemke[ONLINE_USER_ID]
2	[IBAN_CODE] [EMAIL_ADDRESS] 836-452-6042 github.com Hoppe[ONLINE_USER_ID]
3	[IBAN_CODE] [EMAIL_ADDRESS] 176-202-5004 qq.com Bahringer[ONLINE_USER_ID]
4	[IBAN_CODE] [EMAIL_ADDRESS] 779-719-4068 msn.com Mills[ONLINE_USER_ID]
5	[IBAN_CODE] [EMAIL_ADDRESS] 974-394-3017 businessweek.com Bogan[ONLINE_USER_ID]
6	[IBAN_CODE] [EMAIL_ADDRESS] 252-565-2609 toplist.cz Denesik[ONLINE_USER_ID]
7	[IBAN_CODE] [EMAIL_ADDRESS] 375-486-1852 huffingtonpost.com Paucek[ONLINE_USER_ID]
8	[IBAN_CODE] [EMAIL_ADDRESS] 914-567-0100 scientificamerican.com Ledner[ONLINE_USER_ID]
9	PK08 NOTU TJ05 M7HK DWOG RFUW [EMAIL_ADDRESS] 514-857-7874 timesonline.co.uk Lueilwitz[ONLINE_USER_ID]
10	[IBAN_CODE] [EMAIL_ADDRESS] 602-617-5896 scribd.com Harber[ONLINE_USER_ID]

Data Catalog

DismissEnable