

RGPD avec Apache BEAM et BigQuery

Mehdi BEN HAJ ABBES

@mbha_phoenix

github.com/mbhaphoenix

DevFest Paris 2020

Règlement général sur la protection des données (RGPD)

- ☐ **Un règlement de l'Union européenne qui constitue le texte de référence en matière de protection des données à caractère personnel**
- ☐ **Les principales dispositions :**
 - **Privacy by design : sécuriser les DCP**
 - **Droit à l'effacement "à l'oubli"**
 - **Portabilité des données**
 - **Le consentement explicite**
 - **Notification des fuites**
 - **Nommer un DPO**
 - **...**

Crypto-shredding

- Wikipedia : the practice of 'deleting' data by deliberately deleting or overwriting the encryption keys. This requires that the data have been encrypted
- Thoughtworks : the practice of rendering sensitive data unreadable by deliberately overwriting or deleting encryption keys used to secure that data.

Clé de cryptage par id

id	name (DCP)	non DCP col
1	mehdi	toto
2	ben	bobo
3	haj	coco
4	abbes	hoho



id	keyset
1	CNXCzocLEm
2	QKWAowdHI
3	ZS5nb29nbG
4	NvbS9nb29

id	crypted DCP	non DCP col
1	AZvFDcZtK1	toto
2	YfDempUSO I	bobo
3	R0bKCvwC	coco
4	Acl+t3TIBP	hoho

BigQuery :

Les fonctions de chiffrement AEAD

En mode Batch

A base de Tink : lib open source de crypto par Google

Permet de :

- Créer des collections de clés (keyset) de chiffrement et de déchiffrement

`KEYS.NEW_KEYSET(key_type)`

- Utiliser ces clés pour chiffrer et déchiffrer les valeurs individuelles d'une table

`AEAD.ENCRYPT(keyset, plaintext, additional_data)`

`AEAD.DECRYPT_BYTES(keyset, ciphertext, additional_data)`

`AEAD.DECRYPT_STRING(keyset, ciphertext, additional_data)`

- Assurer la rotation des clés d'une keyset :

`KEYS.ROTATE_KEYSET`

BigQuery AEAD en batch

id	name (DCP)	non DCP col
1	mehdi	toto
2	ben	bobo
3	haj	coco
4	abbes	hoho

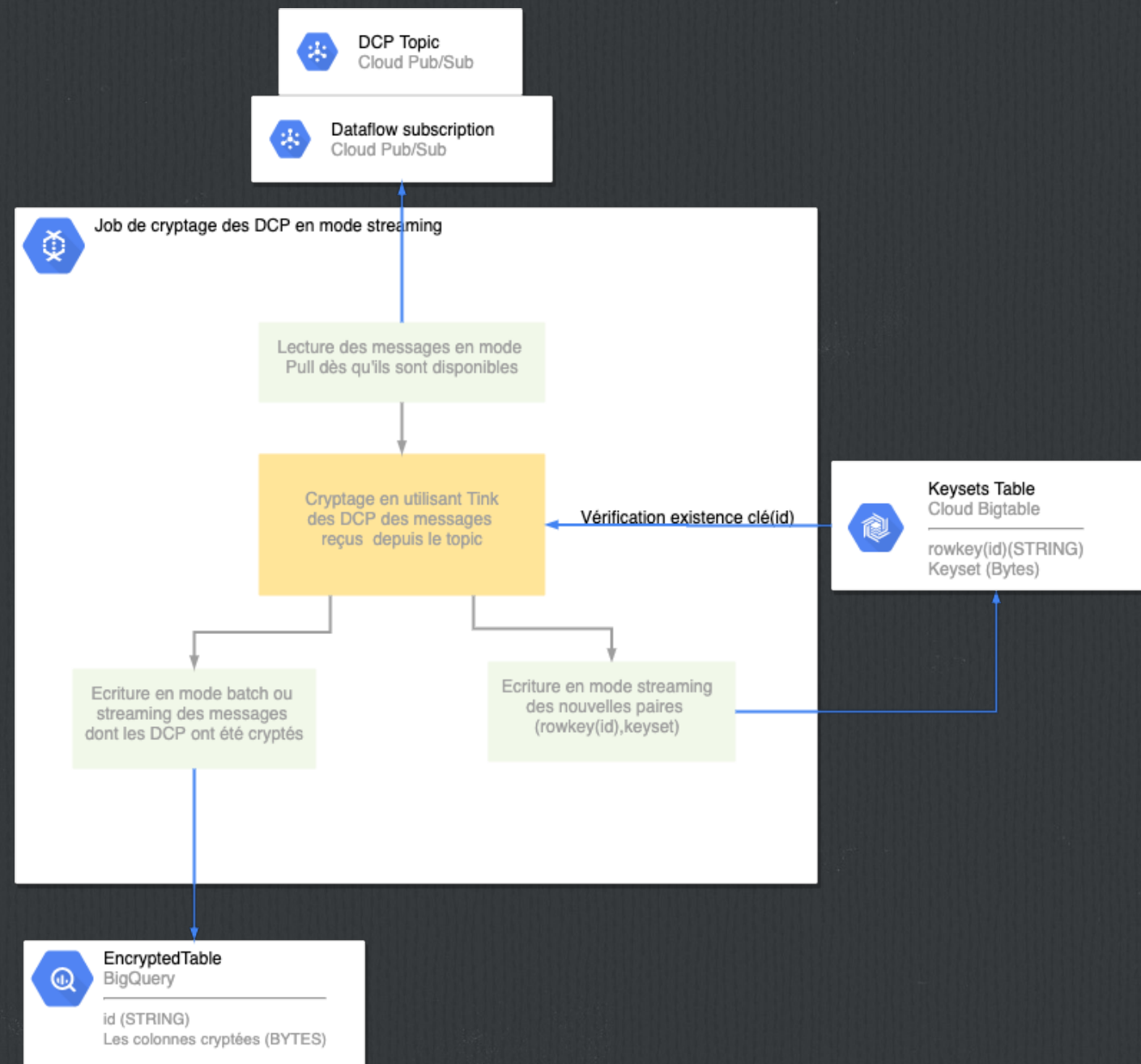
```
CREATE OR REPLACE TABLE
df_secure.keysets AS SELECT
id, KEYS.NEW_KEYSET('AEAD_AES_GCM_256') AS keyset
FROM
( SELECT DISTINCT(id)
  FROM df.plain)
```

id	keyset
1	CNXCzocLEm
2	QKWAowdHI
3	ZS5nb29nbG
4	NvbS9nb29

id	crypted DCP	non DCP col
1	AZvFDcZtK1	toto
2	YfDempUSO	bobo
3	R0bKCvwC	coco
4	Acl+t3TIBP	hoho

```
CREATE OR REPLACE TABLE
encrypted AS SELECT
plain.* EXCEPT(dcp),
AEAD.ENCRYPT((SELECT keyset FROM keysets
WHERE
keysets.id = plain.id),plain.dcp, plain.id ) AS cipher,
FROM plain
```


Tink en streaming



Ressources

- ❑ <https://github.com/mbhaphoenix/devfest-paris2020-rgpd-bq-beam>
- ❑ Antiséches RGD : <https://www.datagalaxy.com/blog/antiseches-rgpd/antiseche-rgpd-6-anonymisation-et-pseudonymisation/>
- ❑ BigQuery AEAD : https://cloud.google.com/bigquery/docs/reference/standard-sql/aead_encryption_functions
- ❑ Tink lib : <https://github.com/google/tink>

Merci