Yadukishore Tatavarthi

Best practices of Managing GDPR with snowflake

Yadukishore Tatavarthi
Sr Resident Solutions Architect at Snowflake
Published Feb 25, 2022

+ Follow

This blog post is mainly aimed at Introducing GDPR Regulation, common 'Right to be forgotten options', common reference architecture patterns, and best practices of implementing the same using powerful and easy to use features in snowflake

What is the Challenge?

be stored in a highly secure area and organized in a way that gives access to personal data on request. Organizations typically carry petabytes of data distributed across multiple warehouses, data marts, clouds, technologies, backups, etc. Bringing all this data under GDPR Compliance is of high priority and also challenging. IT teams spend a lot of effort in identifying the personal data, reporting accurately, responding to audits, deleting the data based on the data subject request. This can take days/months and often leads to Poor GDPR compliance and massive fines. Snowflake Unique Architecture, inbuilt security, and several governance features helps organizations to easily comply with GDPR

What is GDPR?

The General Data Protection Regulation (GDPR) is a new EU Regulation governing the protection and processing of EU Personal Data. GDPR focuses on securing personnel data as well as how the data will be handled and how the entities processing the data will facilitate the exercise of data subject's rights.

Whom does GDPR Protect?

GDPR protects natural persons (I.e. Individuals, and not entities) within the EU. GDPR Covers any personal data or information that may be used to identify a person whether directly or indirectly. Pseudonymous data (tokenized or encrypted) is still considered personal data since pseudonymization can be reversed

It is to be noted that EU Citizens living outside of the EU (Ex: US) will also come under the GDPR Scope

GDPR key terms:

Processor: is the entity which processes the data on the instructions of the controller. Example: Snowflake

Controller: is the entity which determines the purposes and means of processing the personal data: Ex: Customer

Note: Organizations must consider GDPR as a part of the design process, not an afterthought. Snowflake provides several features to help organizations to be GDPR compliant, it is ultimately the organizations' responsibility for designing the architecture that is GDPR Compliant. Having said that there are certain responsibilities of the Processor which are listed in the Data processing agreement. Some of them include

1. Putting data processing addendums in place with customers and vendors

2. Should use the EU personal data only to provide service to them, not for any other use

3. Notifying Customers if there is a security incident that affects their EU Data. Based on my recollection SAAS provider (snowflake) will have 72 hours to inform the same and the customer will have another 72 hours to inform the end customers.

4. Being transparent on how Snowflake handle and process our customers EU personnel data on their behalf and keeping accurate records

5. Securing Customers EU Personnel data in our service which facilitates our customers' compliance with data Subject Requests

Controller – Snowflake customers are responsible for complying with GDPR independently from Snowflake

Snowflake provides the customer with several controls that may be used by the customer to retrieve, correct, delete or restrict personal data. This enables customers to comply with its GDPR Obligations including its obligations relating to responding to requests coming from data subjects or applicable data protection authorities

High level Features that are commonly Used: Time travel, Cloning, Data Classification, SWAP, Row Access Policies, Dynamic Masking Policies, Secure Views, Anonymization techniques, Enterprise SIEM Integration, Access History.

Note: Think of the above features as Lego blocks depending on organization needs, policies we can used one or several of them to build the solution you need.

Common Reference Architecture Patterns and Solutions: GDPR compliance can be extremely challenging if you don't have a well-thought-out database architecture, especially for handling the "right to erasure (right to be forgotten)" in GDPR Article 17. Once an individual's personally identifiable information (PII) is requested, organizations have 30 to 90 days in which to delete the individual's PII from their database.

Organizations have to determine on how data to be forgotten, flagged or anonymized or deleted and support for audits

1. **Foundational:** Tag all your data whether it is sensitive, Non-Sensitive, PII, GDPR Compliant, etc. Object tagging and data classification features of snowflake help you to

effective cataloging strategy help to easily identify the PII data when the situation arises

2. Isolate PII Data: Your data architecture/Model should be able to easily identify the PII data. There are multiple ways to do this. You can organize the PII data in a separate account, database, schema, or table level. The most used approach is doing this at table level.

a. For example, if the customer table has 300 Columns and 50 of them are PII data separate them into two different tables (Customer_PII), (Customer_NonPII) and link them by using an unintelligent key (ex: Hash key/pseudo-anonymized) so that you can easily drop the tables/delete the data related to PII data. This way you will preserve the Non-PII data and an efficient way to avoid costly updates to the table etc.

b. Note: Entire Data modeling of this is beyond the scope of this blog. This is to give an idea of the art of possible

c. Note: This approach is mostly applicable for green field implementations

3. Physical Deletion Of data: - TT

a. This is most common commonly used Technique. This can be achieved by setting the Time travel /Data retention time setting on the table level and simply delete the data

b. Snowflake has 3 types of tables. GDPR Provides organizations 30 days- 90 days to delete the data.

i. Permanent Tables: Set Data retention time to 23 Days (plus fail-safe of 7 days)

Note: Consider doing the batch deletions once a month with the help of some Audit tables which mark what data needs to be deleted

4. Physical Deletion of Data-2 (Clone/Swap):

a. Create a transient table (ex: cust_phi_clone) cloning production table containing PII data (ex: customer_phi) (you can choose to set zero-time travel)

b. Remove the data under investigation from both the databases

c. SWAP both the tables. Swapping essentially removes the time travel from the Production table

d. Restrict the access to cust_phi_clone to only Admins.

5. **Backing up of data bases**: Some customers wanted to keep the backup of the data for Auditing Purposes

a. As an example, keep the Time travel for Main customer data to be 32 days

b. Use a scheduled taks to create a Monthly backup:

```
CREATE OR REPLACE TRANSIENT DATABASE prd_ent_presentation_db_bu_2020_01 CLONE
```

c. Restrict the access to backup databases to only set of Administrators/stewards

```
DROP DATABASE prd_ent_presentation_db_bu_2020_01; --This data is instantly an
```

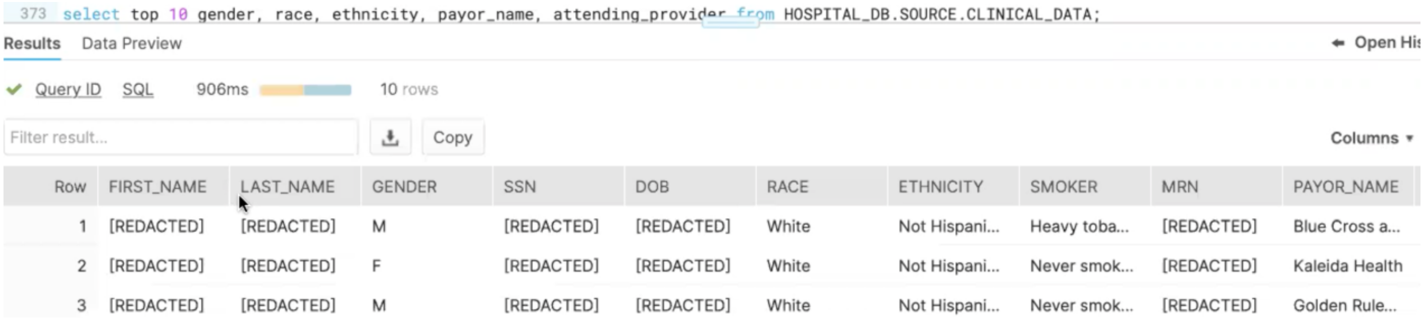
6. Logical Deletion Patterns:

a. HIPAA opt-out. We can use secure views and row access policies to Deal with the cases like patient wants to opt out to share his information, prescriber does not want to reveal the information, or a state law do not allow to share any information for patient's underage of 18 etc.

7. PII Anonymization Patterns:

a. Row Access Policies: We can restrict the access to certain rows based on the role user logs in. Marketing department cannot see sales data etc

b. Dynamic Data Masking Polices: Depending on the Role the user logs in we can mask/redact the data: Ex: HR can only see salary details



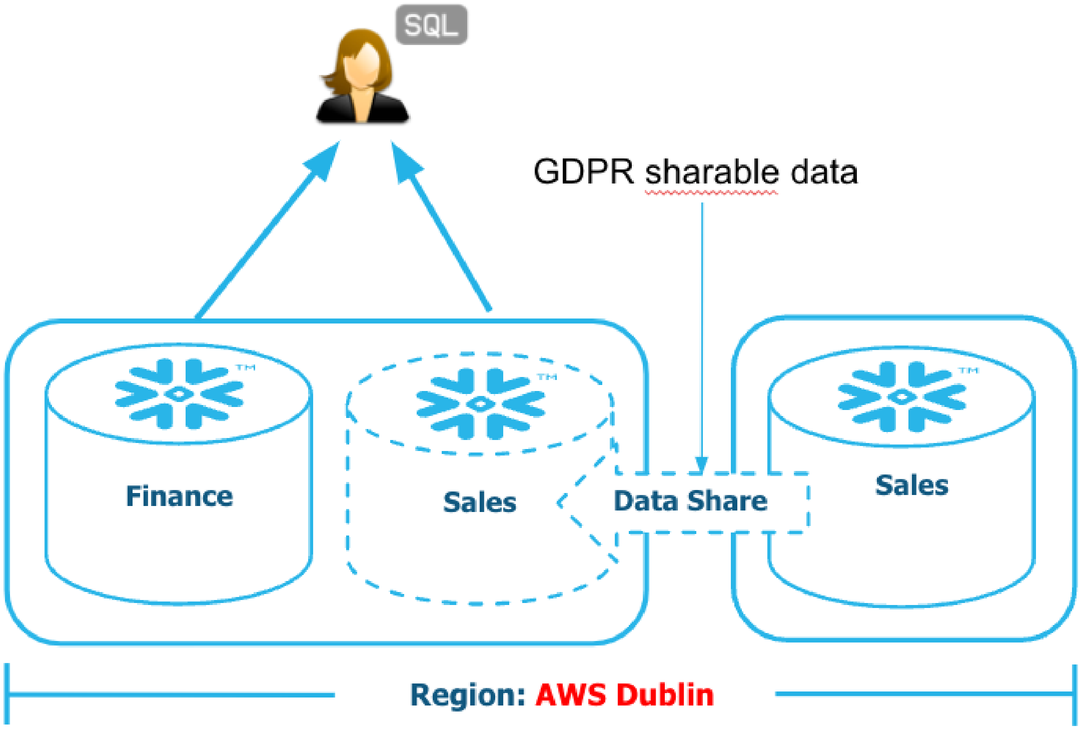
```
select top 10 gender, race, ethnicity, payor_name, attending_provider from HOSPITAL_DB.SOURCE.CLINICAL_DATA;
```

Row	FIRST_NAME	LAST_NAME	GENDER	SSN	DOB	RACE	ETHNICITY	SMOKER	MRN	PAYOR_NAME
1	[REDACTED]	[REDACTED]	M	[REDACTED]	[REDACTED]	White	Not Hispani...	Heavy toba...	[REDACTED]	Blue Cross a...
2	[REDACTED]	[REDACTED]	F	[REDACTED]	[REDACTED]	White	Not Hispani...	Never smok...	[REDACTED]	Kaleida Health
3	[REDACTED]	[REDACTED]	M	[REDACTED]	[REDACTED]	White	Not Hispani...	Never smok...	[REDACTED]	Golden Rule...

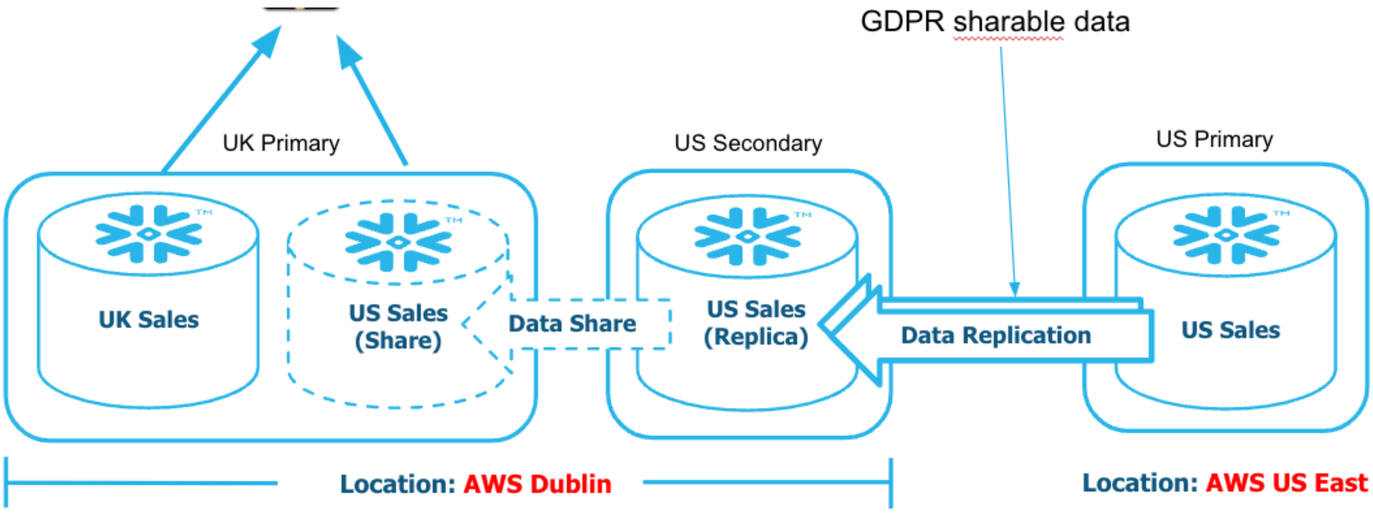
8. Secure Data Sharing: Helps to mitigate GDPR Risk

b. You can share the data in the same region or outside the region. We can also architect the solution to deal with the scenarios like “ This subject area data can not leave the country/region etc”

9. *Region Data Sharing:*



10. Cross Region Data sharing:



11. **Implement Audit Control Framework:** Maintain an audit control framework that tracks the data subject requests (who has requested, when, what data was deleted, rolled back, etc.). This can help for your audit requests and also helps to delete the PII data in case of accidental rollbacks etc. Snowflake's new feature Access History also helps you to easily determine who has accessed what data, when (high-level data lineage). This feature really helps organizations to respond to audit requests quickly, confidently

Final words: It is ultimately Organization's responsibility to ensure that their data is GDPR Compliant. Organizations must build their data architecture/designs by keeping this in mind. The above Point of view is the collation of thoughts based on my experience and I strongly suggest you seek advice from your legal team that this will suffice

Note: Opinions expressed are solely my own and do not express the views or opinions of my employer.

Yadukishore Tatavarthi

23 · 1 Comment

Like Comment Share

Pavan Allu 3mo

Great one Kishore !

Like Reply

To view or add a comment, sign in

More articles by this author

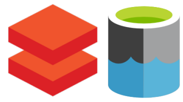
See all

How Lifesciences companies are…
Nov 13, 2020

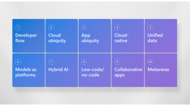
Best practices: Data migration from…
Oct 22, 2020

How to design a "Successful Multi Clo…
May 13, 2018

Others also viewed



Fast Incremental Data Loads into Delta Lake with ADLS Gen 2
Andrew M. Lucius · 1y



10 technologies coming together to help you build what’s next
Satya Nadella · 2w

Yadukishore Tatavarthi

Privacy Policy

Cookie Policy

Copyright Policy

Brand Policy

Guest Controls

Community Guidelines

Language