GDPR Compliance in the Mastadon Data Pipeline

Ensuring General Data Protection Regulation (GDPR) compliance is of utmost importance in handling personal data within your data pipeline. Here are key steps taken to respect GDPR regulations:

1. Data Anonymization:

Personal information is a critical aspect of GDPR compliance. To adhere to GDPR requirements, any personal data that is not relevant to the analysis is either deleted or hashed before processing. This ensures that no sensitive personal information is exposed during the data pipeline.

2. Data Minimization:

Only the necessary data required for analysis is retained. Any extraneous or irrelevant information is discarded to minimize data exposure and potential privacy risks.

3. Security Measures:

Both HDFS and HBase storage systems are secured to prevent unauthorized access. Proper access controls, encryption, and authentication mechanisms are implemented to protect data at rest. This safeguards data throughout the pipeline.

4. Data Lake Management:

To mitigate data retention risks, a strict policy is in place to delete data from the data lake once it has been processed. This ensures that personal data is not stored indefinitely, and retention is strictly aligned with the purposes of data processing as specified in the GDPR.

5. Data Protection Impact Assessment (DPIA):

A DPIA is conducted to identify and mitigate potential privacy risks in the data pipeline. This assessment helps in making informed decisions about data processing practices and safeguards.

6. Consent Management:

If applicable, data processing activities are conducted only after obtaining explicit consent from data subjects, as required by GDPR. Consent records are maintained and managed accordingly.

7. Data Subject Rights:

Mechanisms are in place to accommodate data subject rights as defined by GDPR. This includes the ability to access, correct, or delete their personal data.

8. Documentation and Compliance Records:

Detailed documentation of data processing activities, safeguards, and compliance measures is maintained. This documentation ensures transparency and accountability in the event of regulatory inquiries.

9. Regular Auditing and Compliance Checks:

Regular audits are conducted to verify that GDPR compliance measures are consistently followed throughout the data pipeline. This includes ensuring that the pipeline conforms to any changes or updates in GDPR regulations.

10. Data Breach Response:

A well-defined protocol is established to promptly respond to any data breaches in accordance with GDPR requirements. Data subjects and supervisory authorities are notified as required.

By taking these steps, the data pipeline demonstrates a commitment to respecting GDPR regulations, safeguarding personal data, and upholding the privacy rights of individuals as they pertain to the collected data.