# Deliverable2_Corrected_Data_Flow_Diagram

Tuesday, February 3, 2026 8:32 AM

Deliverable 2: Corrected Data Flow Diagram Notes

1. Remove excessive data collection at the mobile application using data minimization
The mobile application should only collect data that is strictly necessary for loan assessment and service delivery. Collecting excessive or unrelated personal information increases privacy risk and regulatory exposure. Applying data minimization ensures compliance with data protection principles and reduces the impact of potential data breaches.

2. Add explicit consent verification between the API Gateway and the Raw Data Database
Before any customer data is stored in the Raw Data Database, the API Gateway must verify that explicit user consent has been obtained. This verification step ensures that only lawfully collected data enters persistent storage and enforces compliance with consent requirements under data protection regulations.

3. Apply data classification, encryption, and retention policies in the Raw Data Database
All stored data should be classified based on sensitivity (e.g., personal, financial, or sensitive data). Appropriate encryption mechanisms must be applied both at rest and in transit, and retention policies should define how long data is stored and when it must be securely deleted to reduce long-term privacy and security risks.

4. Normalize and validate data in the preprocessing service
The preprocessing service should standardize data formats, handle missing or inconsistent values, and validate data against defined schemas. This step improves data quality, ensures consistency across datasets, and prevents corrupted or invalid data from being used in machine learning models.

5. Enable logging and explainability at the decision service
The decision service should maintain detailed logs of model inputs, outputs, and decision paths. Explainability mechanisms must be included to justify automated loan decisions, supporting transparency, auditability, and the ability to respond to customer or regulatory inquiries.

6. Mask or anonymize data before analytics and third-party sharing
Before data is used for analytics or shared with third parties, personally identifiable information must be masked or anonymized. This reduces the risk of re-identification, protects customer privacy, and ensures that secondary data usage remains compliant with legal and ethical requirements.