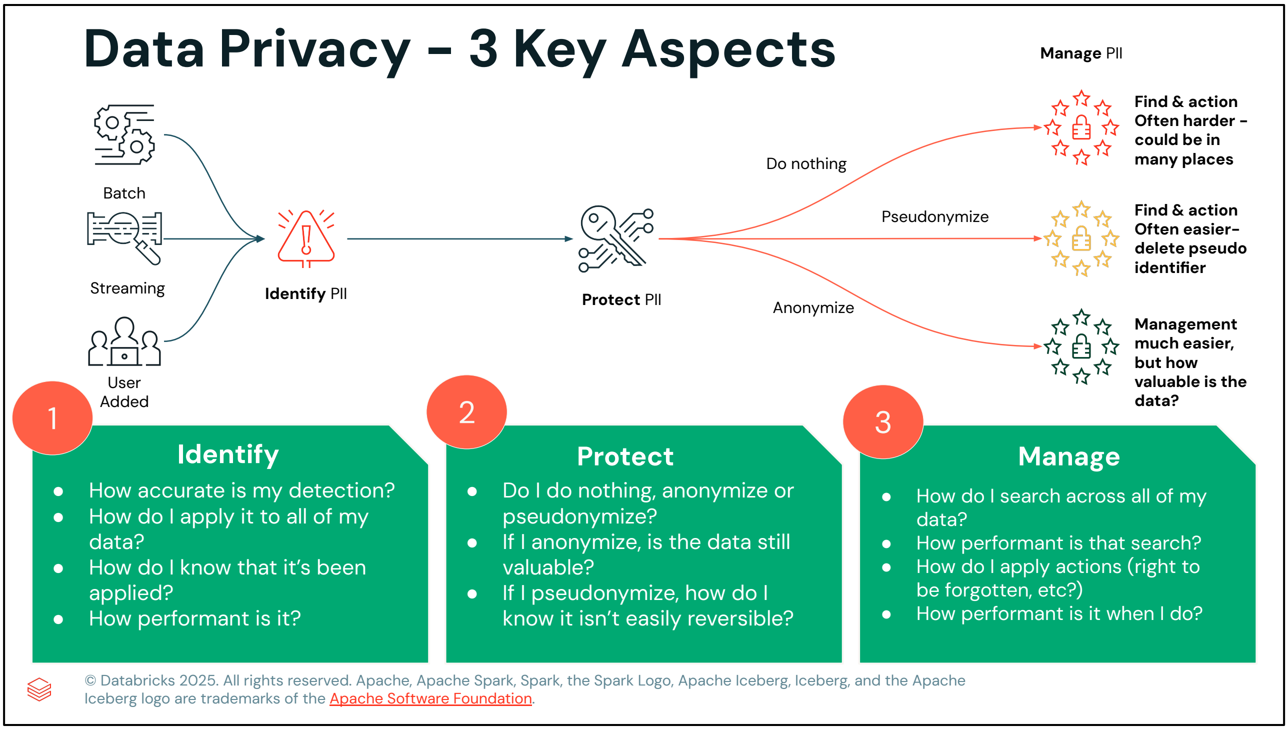
databricks

Storing Data Securely

LECTURE

Data Privacy



Data Privacy – 3 Key Aspects

Manage PII
Batch
Streaming
User Added
Identify PII
Protect PII
Do nothing
Pseudonymize
Anonymize
Find & action Often harder - could be in many places
Find & action Often easier– delete pseudo identifier
Management much easier, but how valuable is the data?

1

Identify

- How accurate is my detection?
- How do I apply it to all of my data?
- How do I know that it's been applied?
- How performant is it?

2

Protect

- Do I do nothing, anonymize or pseudonymize?
- If I anonymize, is the data still valuable?
- If I pseudonymize, how do I know it isn't easily reversible?

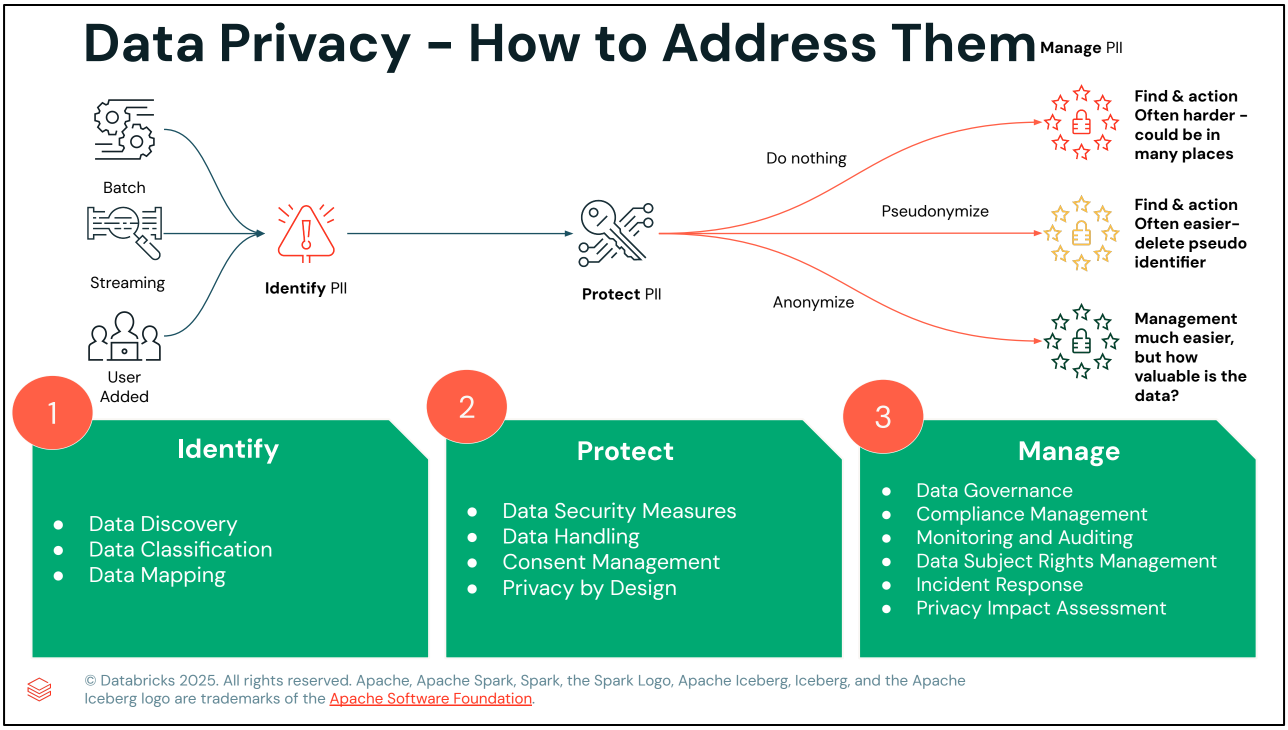
3

Manage

- How do I search across all of my data?
- How performant is that search?
- How do I apply actions (right to be forgotten, etc?)
- How performant is it when I do?

When we're working with Data Privacy, there are three key aspects to consider in your organization in how to manage your data independently and where it is coming from:

1. First, we must identify what is our current state, and we must ask ourselves questions like:
    a. What data, where is its source located, and where will it be consumed?
    b. How is the data handled?
    c. How accurate is my detection of what data must be protected?
    d. How do I know some kind of security has been applied?
2. Second, we need to assess and evaluate our options:
    a. Should I do nothing? Is it worth to secure this data?
    b. Should I anonymize or pseudonymize my data?
    c. How can I get my data back after it is secured?
3. Third, it is about managing your data:
    a. How do I search all my data?
    b. How do I apply actions to my data where users have rights such as “to be forgotten,” “to be rectified,” or “to restrict processing?”



Data Privacy – How to Address Them

Manage PII

Batch

Streaming

User Added

Identify PII

Protect PII

Do nothing

Pseudonymize

Anonymize

Find & action Often harder - could be in many places

Find & action Often easier- delete pseudo identifier

Management much easier, but how valuable is the data?

1

Identify

- Data Discovery
- Data Classification
- Data Mapping

2

Protect

- Data Security Measures
- Data Handling
- Consent Management
- Privacy by Design

3

Manage

- Data Governance
- Compliance Management
- Monitoring and Auditing
- Data Subject Rights Management
- Incident Response
- Privacy Impact Assessment

The "identify" aspect of data privacy suggests:

- Data Discovery: Organizations need to conduct comprehensive data inventories to understand what personal and sensitive information they hold. This involves identifying, categorizing, and labeling sensitive data to understand its location and usage within the organization.
- Data Classification: Once data is discovered, it should be classified based on its sensitivity, such as PII, financial data, and health records. This classification helps in applying appropriate privacy measures and compliance protocols.
- Data Mapping: Creating a clear picture of how data flows through the organization, including where it's stored, who has access to it, and how it's used.

The "protect" aspect suggests:

- Data Security Measures: Implementing technical safeguards like encryption, access controls, and firewalls to prevent unauthorized access to sensitive data.
- Data Handling: Minimization and Limiting data collection to only what is necessary for specific business purposes, aligning with regulations like GDPR.
- [Out of Databricks Scope]Consent Management: Obtaining explicit and informed consent from individuals before collecting and processing their personal data.
- [Out of Databricks Scope]Privacy by Design: Incorporating privacy considerations into the development of new products, services, and processes

- from the outset.

The "manage" aspect suggests:

- Data Governance: Establishing policies, procedures, and best practices for handling personal data throughout its lifecycle.
- Compliance Management: Ensuring adherence to relevant data protection regulations such as GDPR, CCPA, and HIPAA.
- Ongoing Monitoring and Auditing: Regularly assessing and updating privacy practices to address evolving threats and regulatory requirements.
- [Out of Databricks Scope]Data Subject Rights Management: Implementing processes to handle data subject access requests (DSARs) and other privacy rights efficiently.
- [Out of Databricks Scope]Incident Response: Developing and maintaining plans for responding to data breaches and other privacy incidents.