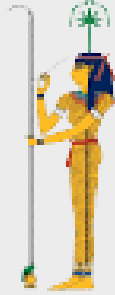


Big Data

Big Data Course

Mostafa Nabieh

وزارة الاتصالات
وتكنولوجيا المعلومات



MANNING



وزارة الاتصالات
وتكنولوجيا المعلومات
MINISTRY OF COMMUNICATIONS
AND INFORMATION TECHNOLOGY



UDACITY



PLURALSIGHT



YOUR SPACE TO LEARN
FUTURE SKILLS

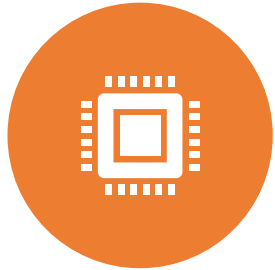
Mostafa Nabieh





Mostafa Nabieh

CONTENTS



WHAT IS DATA
ENGINEERING?



BIG DATA
ECOSYSTEM



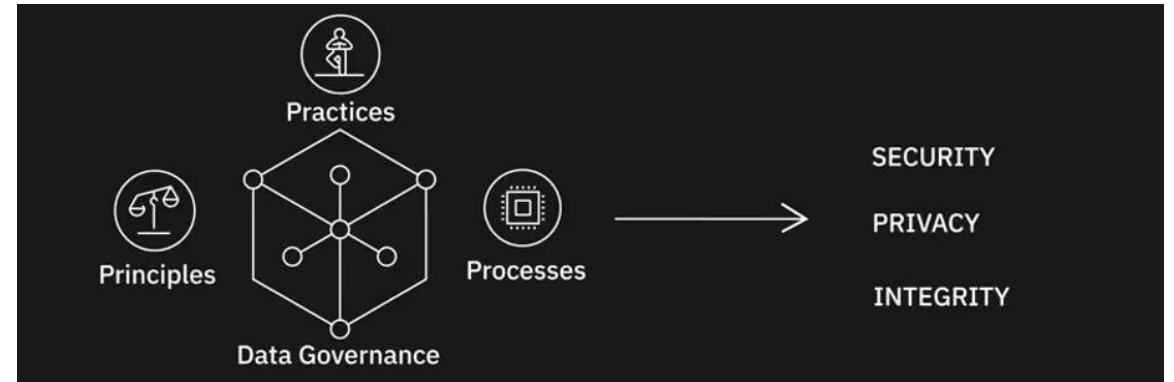
BIG DATA
LIFECYCLE



CAREER
OPPORTUNITIES

Governance and Compliance

- Data Governance is a collection of principles, practices, and processes to maintain the



Data that needs Governance

Personal: Personal Information (PI) and Sensitive Personal Information (SPI)

- Can be used to identify an individual
- Can be used to cause harm to an individual
- Can be traced back to an individual



Data that needs Governance

Industry-specific regulations:

- Health Insurance Portability and Accountability Act (**HIPAA**) for Healthcare
- Payment Card Industry Data Security Standard (**PCI DSS**) for Retail
- Sarbanes Oxley (**SOX**) for Finance



Compliance

Compliance covers the processes and procedures through which an organization adheres to regulations and conducts its operations in a legal and ethical manner.



Establish controls and checks in order to comply with regulations



Maintain a verifiable audit trail to establish adherence to regulations

Compliance

Compliance is an ongoing process requiring a blend of



PEOPLE



PROCESS



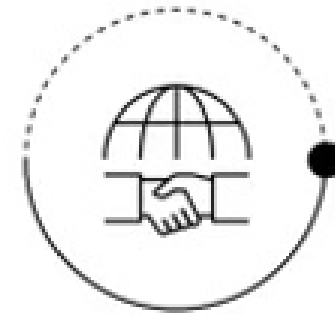
TECHNOLOGY

Data Lifecycle

Governance regulations require enterprises to **know their purpose and maintain transparency in their actions** at each step of the data lifecycle.

In the **Data Acquisition Stage**, you need to:

- Identify data that needs to be collected and the legal basis for procuring the data.
- Establish the intended use of data, published as a privacy policy.
- Identify the amount of data you need to meet your defined purposes.



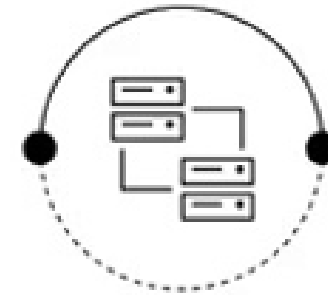
Acquisition

Data Lifecycle

Governance regulations require enterprises to **know their purpose and maintain transparency in their actions** at each step of the data lifecycle.

In the **Data Processing Stage**, you need to:

- Flesh out details of how exactly you are going to process personal data.
- Establish your legal basis for the processing of personal data.



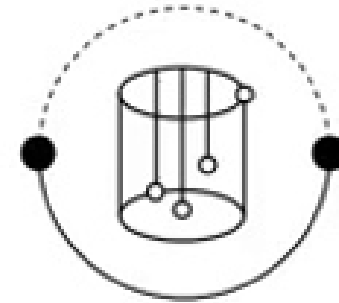
Processing

Data Lifecycle

Governance regulations require enterprises to **know their purpose and maintain transparency in their actions** at each step of the data lifecycle.

In the **Data Storage Stage**, you need to:

- Define where you will store the data.
- Establish specific measures you will take to prevent internal and external security breaches.



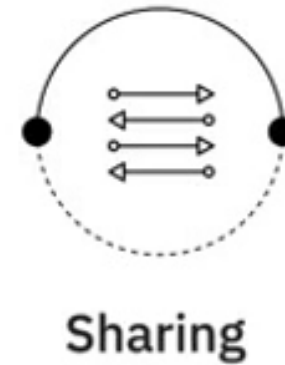
Storage

Data Lifecycle

Governance regulations require enterprises to **know their purpose and maintain transparency in their actions** at each step of the data lifecycle.

In the **Data Sharing Stage**, you need to:

- Identify third-party vendors in your supply chain that will have access to the collected data.
- Establish how you will hold third-party vendors contractually accountable to regulations.

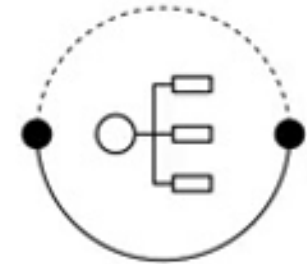


Data Lifecycle

Governance regulations require enterprises to **know their purpose and maintain transparency in their actions** at each step of the data lifecycle.

In the **Data Retention and Disposal Stages**, you need to:

- Define policies and processes you will follow for the retention and deletion of personal data after a designated time.
- Define how you will ensure deleted data is removed from all locations, including third-party systems.



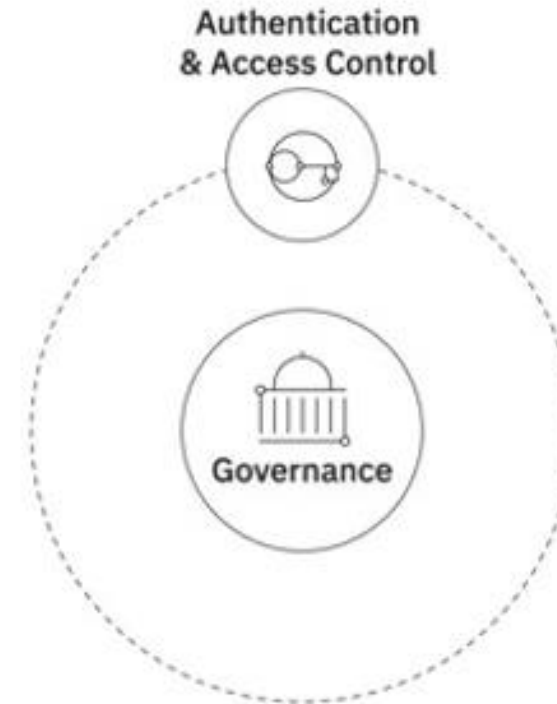
**Retention
and Disposal**

Technology as an Enabler

Today's tools and technologies provide several controls for ensuring organizations comply to governance regulations.

Authentication & Access Control

- Layered authentication processes
- Combination of passwords, tokens, and biometrics, to prevent unauthorized access
- Authentication systems verify that you are who you say you are
- Access Control systems ensure only authorized users have access to resources, both systems and data, based on their user group and role

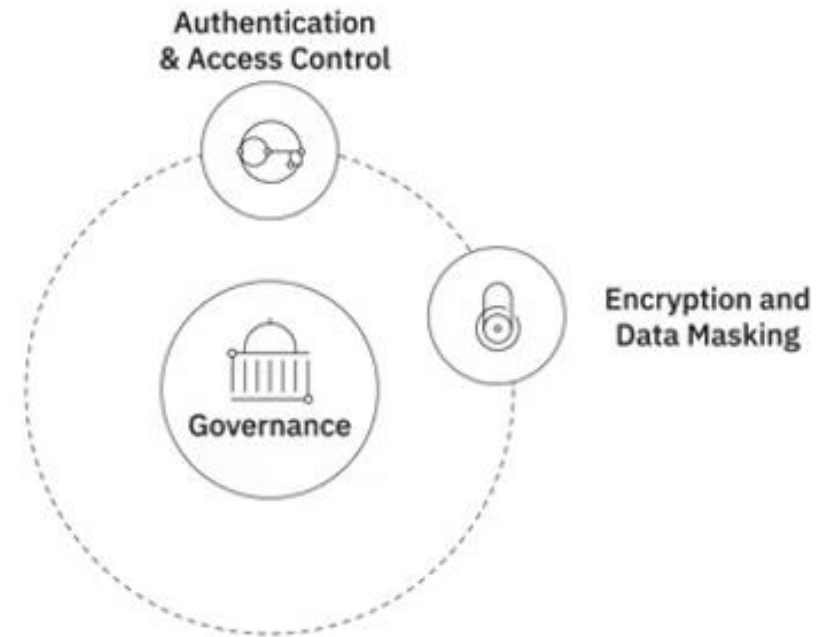


Technology as an Enabler

Today's tools and technologies provide several controls for ensuring organizations comply to governance regulations.

Encryption and Data Masking

- Encryption converts to an encoded format that can only be legible once it is decrypted via a secure key
- Encryption of data is available for:
 - Data at rest
 - Data in transit

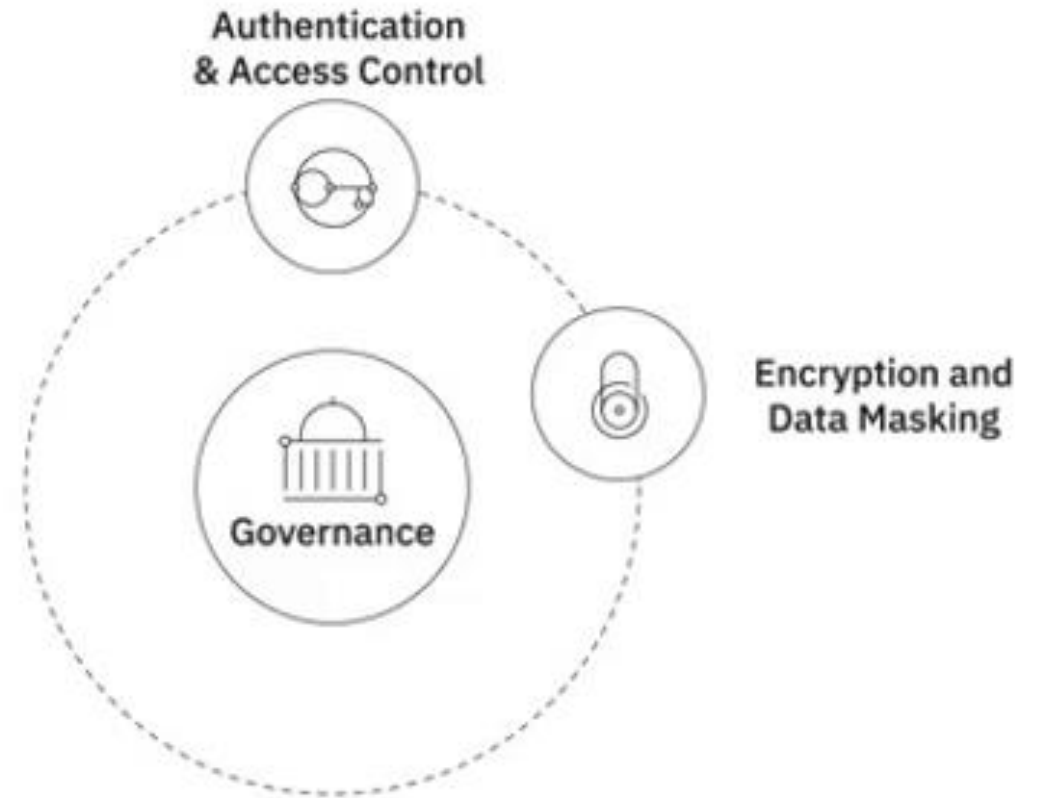


Technology as an Enabler

Today's tools and technologies provide several controls for ensuring organizations comply to governance regulations.

Encryption and Data Masking

- Data Masking provides anonymization of data for downstream processing and pseudonymization of data
- Anonymization abstracts the presentation layer without changing the data in the database itself
- Pseudonymization of data replaces personally identifiable information with artificial identifiers so that it cannot be traced back to an individual's identity

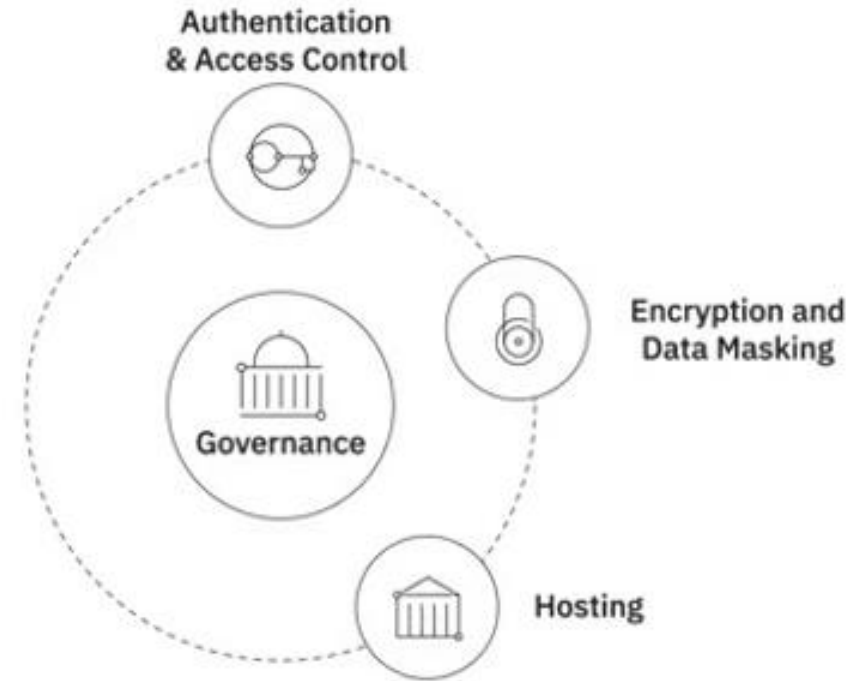


Technology as an Enabler

Today's tools and technologies provide several controls for ensuring organizations comply to governance regulations.

Hosting

- On-premise and cloud-based systems provide hosting options that comply with the requirements and restrictions for international data transfers

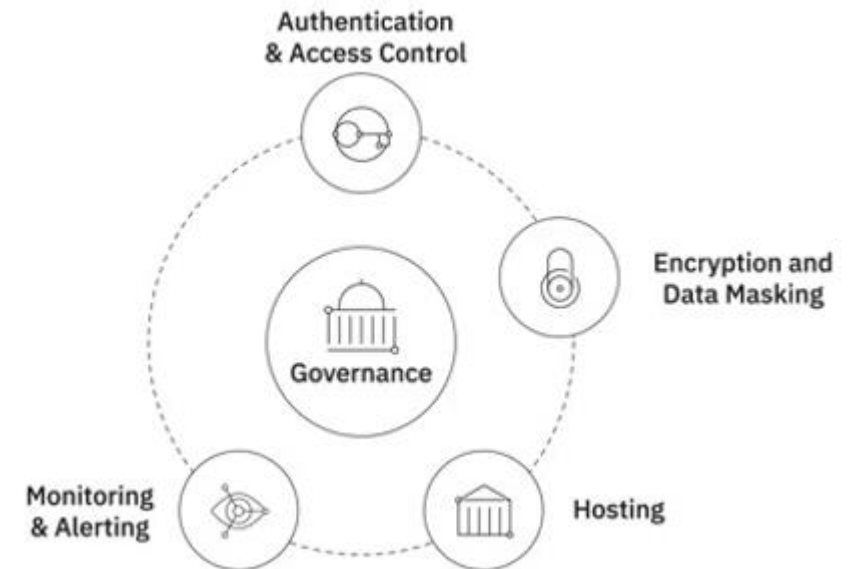


Technology as an Enabler

Today's tools and technologies provide several controls for ensuring organizations comply to governance regulations.

Monitoring & Alerting

- Security monitoring proactively monitors, tracks, and reacts to security violations across infrastructure, applications, and platforms
- Monitoring systems provide detailed audit reports that track access and other operations on the data
- Alerting functionalities flag security breaches so immediate remedial actions can be triggered
- Alerts are based on the severity and urgency level of a breach

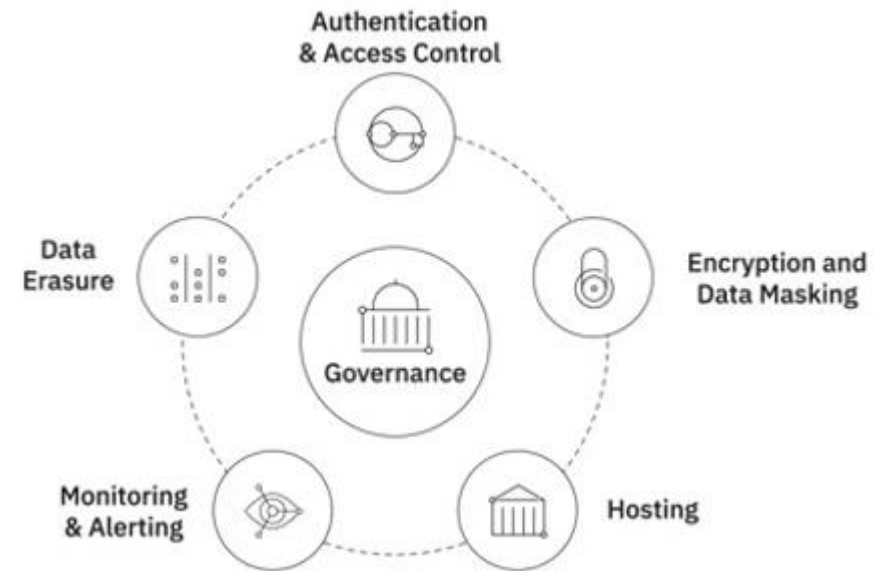


Technology as an Enabler

Today's tools and technologies provide several controls for ensuring organizations comply to governance regulations.

Data Erasure

- A software-based method of permanently clearing data from a system by overwriting.
- Data erasure prevents deleted data from being retrieved.



Thank
you