Professional Certificate in Data Governance

Data Governance Tools and Technologies

Active Data Governance refers to the process of managing and overseeing data assets within an organization, ensuring that data is accurate, complete, and compliant with regulatory requirements. Related terms include Data Quality and Data Security. Active Data Governance involves the implementation of policies, procedures, and controls to manage data across its entire lifecycle, from creation to disposal. This includes data profiling, data validation, and data certification to ensure that data is reliable and trustworthy.

Data Architecture is the overall structure and organization of an organization's data assets, including the relationships between different data entities and the systems that manage and store data. Related terms include Data Modeling and Data Warehousing. Data Architecture involves the design and implementation of a framework that integrates data from various sources and provides a unified view of an organization's data assets. This includes the development of metadata management systems, data governance policies, and data quality procedures.

Data Asset is a collection of data that has value to an organization, such as customer information, financial transactions, or intellectual property. Related terms include Data Resource and Data Inventory. Data Assets are managed and protected through access controls, data encryption, and backup and recovery procedures to ensure their availability and integrity.

Data Certification is the process of verifying that data meets certain standards or requirements, such as accuracy, completeness, or consistency. Related terms include Data Validation and Data Verification. Data Certification involves the use of algorithms and statistical models to analyze data and identify errors or inconsistencies.

Data Classification is the process of categorizing data into different levels or categories based on its sensitivity, value, or risk. Related terms include Data Labeling and Data Tagging. Data Classification involves the use of metadata to assign attributes or properties to data, such as confidentiality, integrity, or availability.

Data Discovery is the process of identifying and locating data within an organization, including unstructured data such as emails, documents, or images. Related terms include Data Mapping and Data Inventory. Data Discovery involves the use of tools and techniques such as search engines, data mining, or text analytics to identify and categorize data.

Data Encryption is the process of converting plaintext data into ciphertext to protect it from unauthorized access or theft. Related terms include Data Protection and Data Security. Data Encryption involves the use of algorithms and keys to scramble data and make it unreadable to unauthorized users.

Data Governance Framework is a structured approach to managing and overseeing data assets within an organization, including policies, procedures, and standards. Related terms include Data Management and Data Quality. A Data Governance Framework involves the establishment of roles and responsibilities for data management, as well as the development of metrics and key performance indicators to measure data quality and compliance.

Data Integration is the process of combining data from multiple sources into a unified view, such as customer data or product data. Related terms include Data Aggregation and Data Consolidation. Data Integration involves the use of tools and techniques such as ETL (Extract, Transform, Load), data warehousing, or data virtualization to integrate data from different systems or applications.

Data Lineage is the process of tracking the origin, movement, and transformation of data throughout its lifecycle. Related terms include Data Provenance and Data Pedigree. Data Lineage involves the use of metadata to record the history of data, including changes, updates, or deletions.

Data Loss Prevention is the process of preventing unauthorized access or theft of sensitive data, such as personal identifiable information or financial data. Data Loss Prevention involves the use of tools and techniques such as data encryption, access controls, or intrusion detection to prevent data breaches or leaks.

Data Management is the process of planning, organizing, and controlling data assets within an organization, including data governance, data quality, and data security. Related terms include Data Administration and Data Stewardship. Data Management involves the establishment of policies, procedures, and standards for data management, as well as the development of metrics and key performance indicators to measure data quality and compliance.

Data Masking is the process of hiding or obscuring sensitive data, such as personal identifiable information or financial data, to protect it from unauthorized access or theft. Data Masking involves the use of algorithms and techniques such as data encryption or tokenization to disguise data and make it unreadable to unauthorized users.

Data Mining is the process of discovering patterns, trends, or relationships in large datasets, such as customer data or transactional data. Related terms include Data Analytics and Predictive Analytics. Data Mining involves the use of statistical models and machine learning algorithms to analyze data and identify insights or opportunities.

Data Modeling is the process of creating a conceptual, logical, or physical representation of data, such as entity-relationship diagrams or data flow diagrams. Related terms include Data Architecture and Data Design. Data Modeling involves the use of tools and techniques such as data modeling software or data modeling languages to create a blueprint for data management.

Data Profiling is the process of analyzing and understanding the characteristics of data, such as data quality, data distribution, or data relationships. Related terms include Data Analysis and Data Visualization. Data Profiling involves the use of statistical models and data visualization techniques to identify trends, patterns, or anomalies in data.

Data Quality is the process of ensuring that data is accurate, complete, and consistent, and meets the requirements of an organization. Related terms include Data Governance and Data Management. Data Quality involves the use of metrics and key performance indicators to measure data quality, as well as the development of policies and procedures to improve data quality.

Data Security is the process of protecting data from unauthorized access, theft, or damage, and ensuring that data is available and accessible to authorized users. Related terms include Data Protection and Data Encryption. Data Security involves the use of tools and techniques such as firewalls, intrusion detection, or access controls to prevent data breaches or leaks.

Data Stewardship is the process of managing and overseeing data assets within an organization, including data governance, data quality, and data security. Related terms include Data Management and Data Administration. Data Stewardship involves the establishment of policies, procedures, and standards for data management, as well as the development of metrics and key performance indicators to measure data quality and compliance.

Data Validation is the process of verifying that data meets certain standards or requirements, such as accuracy, completeness, or consistency. Related terms include Data Verification and Data Certification. Data Validation involves the use of algorithms and statistical models to analyze data and identify errors or inconsistencies.

Data Virtualization is the process of providing a unified view of data from multiple sources, such as customer data or product data, without physically integrating the data. Related terms include Data Integration and Data Aggregation. Data Virtualization involves the use of tools and techniques such as data virtualization software or data federation to create a virtual layer of data.

Data Warehousing is the process of storing and managing large amounts of data in a centralized repository, such as a data warehouse or data mart. Related terms include Data Integration and Data Analytics. Data Warehousing involves the use of tools and techniques such as ETL (Extract, Transform, Load), data modeling, or data mining to manage and analyze data.

ETL (Extract, Transform, Load) is a process of extracting data from multiple sources, transforming the data into a consistent format, and loading the data into a target system, such as a data warehouse or data mart. Related terms include Data Integration and Data Warehousing. ETL involves the use of tools and techniques such as ETL software or data integration platforms to manage and transform data.

Information Lifecycle Management is the process of managing the lifecycle of information, from creation to disposal, including data governance, data quality, and data security. Related terms include Data Management and Data Stewardship. Information Lifecycle Management involves the establishment of policies, procedures, and standards for information management, as well as the development of metrics and key performance indicators to measure information quality and compliance.

Master Data Management is the process of managing and overseeing master data, such as customer data or product data, to ensure that it is accurate, complete, and consistent across an organization. Related terms include Data Governance and Data Quality. Master Data Management involves the use of tools and techniques such as master data management software or data governance platforms to manage and synchronize master data.

Metadata is data that describes or explains other data, such as data definitions, data relationships, or data lineage. Metadata involves the use of metadata management systems or metadata repositories to store and manage metadata.

Predictive Analytics is the process of using statistical models and machine learning algorithms to analyze data and predict future trends or outcomes, such as customer behavior or market trends. Related terms include Data Mining and Data Analytics. Predictive Analytics involves the use of tools and techniques such as predictive analytics software or machine learning platforms to analyze data and make predictions.

Reference Data is a type of master data that provides a standardized set of values or codes for use across an organization, such as country codes or currency codes. Related terms include Master Data Management and Data Governance. Reference Data involves the use of reference data management systems or reference data repositories to store and manage reference data.

Risk Management is the process of identifying, assessing, and mitigating risks associated with data, such as data breaches or data losses. Related terms include Data Security and Data Protection. Risk Management involves the use of tools and techniques such as risk assessments, threat analysis, or vulnerability testing to identify and mitigate risks.

Taxonomy is a hierarchical structure or classification system used to organize and categorize data, such as product categories or customer segments. Taxonomy involves the use of taxonomy management systems or taxonomy repositories to store and manage taxonomies.

Text Analytics is the process of analyzing and extracting insights from unstructured data, such as text documents or social media posts. Text Analytics involves the use of tools and techniques such as text analytics software or natural language processing to analyze and extract insights from unstructured data.

Unstructured Data is a type of data that does not have a predefined format or structure, such as emails, documents, or images. Related terms include Structured Data and Semi-structured Data. Unstructured Data

involves the use of tools and techniques such as text analytics or data mining to analyze and extract insights from unstructured data.