

# Glossary terms from module 3

---

## Terms and definitions from Course 1 Module 3

**Archival policy:** An outline of where and how data is stored once an analysis project is complete

**Audit:** A formal examination of how users are accessing data in order to ensure safe and appropriate access, while identifying and solving any data concerns

**Automation:** The use of software, scripting, or machine learning to perform data analysis processes without human work

**Data analyst:** A professional who performs data analysis workflows, including importing, manipulating, calculating, and reporting on business data

**Data architect:** A professional who collaborates with data analysts, data scientists, and data engineers to design the infrastructure of a database

**Data engineer:** A professional who transforms data into a useful format for analysis, and gives it a reliable infrastructure

**Data lifecycle:** The sequence of stages that data experiences, which include plan, capture, manage, analyze, archive, and destroy

**Data management:** The process of establishing and communicating a clear plan for collecting, storing, and using data

**Data pipeline:** A series of processes that transport data from different sources to a destination for storage and analysis

**Data privacy:** Preserving a data subject's information any time a data transaction occurs

**Data retention:** The collection, storage, and management of data

**Data scientist:** A professional who works primarily in data research, identifying business questions, collecting data from multiple sources, organizing it, and finding answers

**Data-collection policy:** An outline that creates rules around how data is collected, and what resources may be used in the process

**Deletion policy:** An outline of when and how data is permanently destroyed

**Encryption:** The process of encoding information

**Hold:** A policy placed on a dataset that prevents its deletion, or prevents deletion capabilities for certain accounts

**Personally identifiable information (PII):** Data that is unique to an individual, and therefore can be used to distinguish one person from another

**Protected health information (PHI):** Health data that can identify an individual, like information about patient demographics, mental or physical health diagnoses or treatments, and payment records related to health care

**Retention policy:** An outline of how data is archived and deleted

**Security key:** An authentication method that uses a physical digital signature, or key, to verify a user's identity before accessing specific resources

**Versioning:** The process of creating a unique way of referring to data