Data Literacy for Data Science Lesson Glossary

Welcome! This alphabetized glossary contains many of the terms in this course. These terms are important for you to recognize when working in the industry, participating in user groups, and participating in other certificate programs.

Term	Definition	Video where the term is introduced
ACID-compliance	Ensuring data accuracy and consistency through Atomicity, Consistency, Isolation, and Durability (ACID) in database transactions.	Relational Database Management System
Cloud-based Integration Platform as a Service (iPaaS)	Cloud-hosted integration platforms that offer integration services through virtual private clouds or hybrid cloud models, providing scalability and flexibility.	Data Integration Platforms
Column-based Database	A type of NoSQL database that organizes data in cells grouped as columns, often used for systems requiring high write request volume and storage of time-series or IoT data.	NoSQL
Data at rest	Data that is stored and not actively in motion, typically residing in a database or storage system for various purposes, including backup.	Considerations for Choice of Data Repository
Data integration	A discipline involving practices, architectural techniques, and tools that enable organizations to ingest, transform, combine, and provision data across various data types, used for purposes such as data consistency, master data management, data sharing, and data migration.	Data Integration Platforms
Data Lake	A data repository for storing large volumes of structured, semi-structured, and unstructured data in its native format, facilitating agile data exploration and analysis.	Data Marts, Data Lakes, ETL, and Data Pipelines
Data mart	A subset of a data warehouse designed for specific business functions or user communities, providing isolated security and performance for focused analytics.	Data Marts, Data Lakes, ETL, and Data Pipelines
Data pipeline	A comprehensive data movement process that covers the entire journey of data from source systems to destination systems, which includes data integration as a key component.	Data Integration Platforms
Data repository	A general term referring to data that has been collected, organized, and isolated for business operations or data analysis. It can include databases, data warehouses, and big data stores.	Data Collection and Organization
Data warehouse	A central repository that consolidates data from various sources through the Extract, Transform, and Load (ETL) process, making it accessible for analytics and business intelligence.	Data Collection and Organization
Document-based Database	A type of NoSQL database that stores each record and its associated data within a single document, allowing flexible indexing, ad hoc queries, and analytics over collections of documents.	NoSQL
ETL process	The Extract, Transform, and Load process for data integration involves extracting data from various sources, transforming it into a usable format, and loading it into a repository.	Data Marts, Data Lakes, ETL, and Data Pipelines
Graph-based Database	A type of NoSQL database that uses a graphical model to represent and store data, ideal for visualizing, analyzing, and discovering connections between interconnected data points.	NoSQL
Key-value store	A type of NoSQL database where data is stored as key-value pairs, with the key serving as a unique identifier and the value containing data, which can be simple or complex.	NoSQL
Portability	The capability of data integration tools to be used in various environments, including single-cloud, multi-cloud, or hybrid-cloud scenarios, provides flexibility in deployment options.	Data Integration Platforms
Pre-built connectors	Cataloged connectors and adapters that simplify connecting and building integration flows with diverse data sources like databases, flat files, social media, APIs, CRM, and ERP applications.	Data Integration Platforms
Relational databases (RDBMSes)	Databases that organize data into a tabular format with rows and columns, following a well-defined structure and schema.	Data Collection and Organization
Scalability	The ability of a data repository to grow and expand its capacity to handle increasing data volumes and workload demands over time.	Considerations for Choice of Data Repository
Schema	The predefined structure that describes the organization and format of data within a database, indicating the types of data allowed and their relationships.	Considerations for Choice of Data Repository
Streaming data	Data that is continuously generated and transmitted in real-time requires specialized handling and processing to capture and analyze.	Considerations for Choice of Data Repository
Use cases for relational databases	Applications such as Online Transaction Processing (OLTP), Data Warehouses (OLAP), and IoT solutions where relational databases excel.	Relational Database Management System
Vendor lock-in	A situation where a user becomes dependent on a specific vendor's technologies and solutions, making it challenging to switch to other platforms.	Data Integration Platforms