Data Management Glossary

A

Access path: The route through a system by which data is found, accessed and retrieved

Agile methodology: An approach to software development which takes incremental, iterative steps that are reviewed and checked at each stage

Analytics: The systematic study of data to discover and interpret meaningful patterns

Anonymised data: Data from which personal information has been removed or encrypted to ensure that an individual person cannot be identified

Application integration: A continual process of moving data or functions from one software program to another

B

Big data: An extremely large data set that can be analysed by computer to discover patterns and trends

Business intelligence: The process of analysing data to provide actionable information that will inform business decision-making

C

Change data capture: A function in a database that tracks changes in the source data and applies them to the target data

Cloud integration: The process of uploading data to the cloud, where it is subsequently processed and consolidated into a cloud-based database

D

Database: A collection of information which is kept in a storage system in an organised, structured manner

Dataset: A collection of individual but related items of data that can be accessed and processed as individual items or collectively
Data assurance: The process of identifying and amending errors and inconsistencies in a set of data

Data cleansing: The process of identifying and amending or removing inaccurate data from a dataset

Data cluster: A sub-group of data which shares similar characteristics and is significantly different to other clusters in a database, usually defined by the statistical technique of cluster analysis

Data compliance: The process of ensuring that a dataset confirms to the rules specified by national or international laws, or the standards set by trade bodies

Data consolidation: The extraction and integration of data from multiple sources into a single database

Data discovery: The process, which can include data mining or data modelling, of identifying all the data required for a project or task

Data governance: The management of an organisation’s data, including its quality, structure, accessibility and security

Data integration: The continual movement of data from different sources into one or more alternative data stores

Data lake: An unstructured store of raw data held in a flat architecture

Data lineage: The path taken by an item of data from its creation, including any changes to its storage and any processing undertaken

Data management: The administration of the process by which data is created, stored, protected and processed

Data mapping: The process of modelling or illustrating how data will move from a source data store to a target data store

Data masking: The process in which a database is copied and sensitive data is obfuscated, to provide a neutral environment for testing or training

Data migration: The process of moving data from one data store to another

Data mining: The process of examining large data stores to identify patterns or extract usable data

Data modelling: The process of recording the structure of a data stores and the relationships between the data within it

Data profiling: The process of assessing a set of data for completeness and accuracy, often involving the creation of a set of statistics about the data
Data quality: An assessment of how fit-for-purpose a set of data is, based on variables such as completeness and accuracy.

Data science: An interdisciplinary field concerning the methodology of extracting, processing, storing and analysing data.

Data warehouse: See ‘enterprise data warehouse’.

Deduplication: The process of identifying and removing duplicate items of data, sometimes replacing the duplicates with a reference that points to the remaining item of data.

Electronic data interchange (EDI): the process of exchanging standard business documents automatically from one computer to another, often between different organisations.

Enterprise data warehouse: A centralised repository for an organisation’s data, usually consisting of all the organisation’s data or a significant part of it.

Extract, transform and load (ETL): A process that combines three functions to move data from its source to a new storage system: data extraction, processing and integration into a target data store.

Extraction: The process of retrieving data from a source for processing, migration or integration.

File format: A standard, specific way in which an item of data is held in a computer file.

Granularity: The level of detail in a set of data.

Hadoop: A framework of open source software programs used for processing large datasets.

Impact analysis: A risk assessment focusing on the likely impact of any changes to a set of data, the database in which it’s stored or a target system.

Integration testing: The process in which individual units of a software program are combined and tested as a group.

Legacy system: An old, often obsolete computer system or database that may still be in use.
M

Master data management: The process and governance of connecting all of an organisation’s important data into one master file, creating a common point of reference

Metadata: A set of data that summarises information about other data in a database, such as the data’s date of creation, any standards used or its size

N

Network integration: The process of connecting different networks so that data can flow between systems

O

Open database connectivity (ODBC): An interface created by the SQL Access Group for accessing data using a standard set of commands

P

Predictive analytics: The analysis of patterns formed by existing data to predict future trends

R

Real-time analytics: The analysis of data to identify meaningful patterns as soon as it enters a system

Relational database: A database which stores data in tables, ensuring that it is relatively easy to access the information within it

Repository: A database containing related data or metadata from other databases

S

Scalability: The capability of a system or process to handle increasing amounts of data

Schema: An outline or plan of a database’s structure

Source system: The database from which a specific piece of data originated

Staging database: A temporary storage area in which data is processed during an extract, transform and load procedure

Standards: The rules by which a set of data is processed or stored, often defined by an industry body
Structured query language (SQL): A standard computer programming language used for accessing and processing the data in a database

System testing: The process of testing an entire software program to confirm that the entire system is working correctly

Target system: The database to which an item of data is to be moved for storage

Transformation: The process of converting data from one format to another, usually to match the format required by a target system

Unit testing: The process of testing the smallest components of a software program to ensure that each component is working correctly

Virtualisation: The process of replicating and integrating data from disparate sources, without transferring the original data itself

Visualisation: The representation of a set of data in a visual format to aid understanding, such as a flow chart or picture