

Glossary terms from module 4

Terms and definitions from Course 1 Module 4

Apache Hadoop: A framework that distributes the processing of data across clusters of computers

Business data request: A business question that can be answered with data

Check-in: Upload code to a main repository so others can access it and review it

Collaboration: Members of the data analytics team visit the central system to review data requests, including queries and data delivery methods

Data access management: The process of implementing features including password protection, user permissions, and encryption to protect data and related processes

Data documentation: A written guide to the data contained in a dataset, how it was collected, and how it's organized

Dataproc: A fully managed service that maximizes open-source data tools for batch processing, querying, streaming, and machine learning

Documentation: A process for preserving all of the details and conversations about each request in one place

Historical records: Past data or information that's been stored and can be accessed or retrieved when needed

Identity Access Management (IAM): The process of assigning certain individual roles or groups with access to particular resources

Parent-child relationship: A feature of a ticketing system that enables users to divide a single ticket into subtickets, which can be worked on simultaneously