Glossary terms from week 3

Accuracy: An element of quality testing used to confirm that data conforms to the actual entity being measured or described

Business rule: A statement that creates a restriction on specific parts of a database

Completeness: An element of quality testing used to confirm that data contains all desired components or measures

Conformity: An element of quality testing used to confirm that data fits the required destination format

Consistency: An element of quality testing used to confirm that data is compatible and in agreement across all systems

Data dictionary: A collection of information that describes the content, format, and structure of data objects within a database, as well as their relationships

Data lineage: The process of identifying the origin of data, where it has moved throughout the system, and how it has transformed over time

Data mapping: The process of matching fields from one data source to another

Integrity: An element of quality testing used to confirm that data is accurate, complete, consistent, and trustworthy throughout its life cycle

Quality testing: The process of checking data for defects in order to prevent system failures; it involves the seven validation elements of completeness, consistency, conformity, accuracy, redundancy, integrity, and timeliness

Redundancy: An element of quality testing used to confirm that no more data than necessary is moved, transformed, or stored

Schema validation: A process to ensure that the source system data schema matches the target database data schema

Timeliness: An element of quality testing used to confirm that data is current

Terms and definitions from previous weeks

A

Application programming interface (API): A set of functions and procedures that integrate computer programs, forming a connection that enables them to communicate

Applications software developer: A person who designs computer or mobile applications, generally for consumers

Attribute: In a dimensional model, a characteristic or quality used to describe a dimension

B

Business intelligence (BI): Automating processes and information channels in order to transform relevant data into actionable insights that are easily available to decision-makers

Business intelligence governance: A process for defining and implementing business intelligence systems and frameworks within an organization

Business intelligence monitoring: Building and using hardware and software tools to easily and rapidly analyze data and enable stakeholders to make impactful business decisions

Business intelligence stages: The sequence of stages that determine both BI business value and organizational data maturity, which are capture, analyze, and monitor

Business intelligence strategy: The management of the people, processes, and tools used in the business intelligence process

C

Columnar database: A database organized by columns instead of rows

Combined systems: Database systems that store and analyze data in the same place

Compiled programming language: A programming language that compiles coded instructions that are executed directly by the target machine

Contention: When two or more components attempt to use a single resource in a conflicting way

D

Data analysts: People who collect, transform, and organize data

Data availability: The degree or extent to which timely and relevant information is readily accessible and able to be put to use

Data governance professionals: People who are responsible for the formal management of an organization's data assets

Data integrity: The accuracy, completeness, consistency, and trustworthiness of data throughout its life cycle

Data lake: A database system that stores large amounts of raw data in its original format until it's needed

Data mart: A subject-oriented database that can be a subset of a larger data warehouse

Data maturity: The extent to which an organization is able to effectively use its data in order to extract actionable insights

Data model: A tool for organizing data elements and how they relate to one another

Data partitioning: The process of dividing a database into distinct, logical parts in order to improve query processing and increase manageability

Data pipeline: A series of processes that transports data from different sources to their final destination for storage and analysis

Data visibility: The degree or extent to which information can be identified, monitored, and integrated from disparate internal and external sources

Data warehouse: A specific type of database that consolidates data from multiple source systems for data consistency, accuracy, and efficient access

Data warehousing specialists: People who develop processes and procedures to effectively store and organize data

Database migration: Moving data from one source platform to another target database

Database performance: A measure of the workload that can be processed by a database, as well as associated costs

Deliverable: Any product, service, or result that must be achieved in order to complete a project

Developer: A person who uses programming languages to create, execute, test, and troubleshoot software applications

Dimension (data modeling): A piece of information that provides more detail and context regarding a fact

Dimension table: The table where the attributes of the dimensions of a fact are stored

Design pattern: A solution that uses relevant measures and facts to create a model in support of business needs

Dimensional model: A type of relational model that has been optimized to quickly retrieve data from a data warehouse

Distributed database: A collection of data systems distributed across multiple physical locations

E

ELT (extract, load, and transform): A type of data pipeline that enables data to be gathered from data lakes, loaded into a unified destination system, and transformed into a useful format

ETL (extract, transform, and load): A type of data pipeline that enables data to be gathered from source systems, converted into a useful format, and brought into a data warehouse or other unified destination system

Experiential learning: Understanding through doing

F

Fact: In a dimensional model, a measurement or metric

Fact table: A table that contains measurements or metrics related to a particular event

Foreign key: A field within a database table that is a primary key in another table (Refer to primary key)

Fragmented data: Data that is broken up into many pieces that are not stored together, often as a result of using the data frequently or creating, deleting, or modifying files

Functional programming language: A programming language modeled around functions

G

Google DataFlow: A serverless data-processing service that reads data from the source, transforms it, and writes it in the destination location

I

Index: An organizational tag used to quickly locate data within a database system

Information technology professionals: People who test, install, repair, upgrade, and maintain hardware and software solutions

Interpreted programming language: A programming language that uses an interpreter, typically another program, to read and execute coded instructions

Iteration: Repeating a procedure over and over again in order to keep getting closer to the desired result

K

Key performance indicator (KPI): A quantifiable value, closely linked to business strategy, which is used to track progress toward a goal

L

Logical data modeling: Representing different tables in the physical data model

M

Metric: A single, quantifiable data point that is used to evaluate performance

O

Object-oriented programming language: A programming language modeled around data objects

OLAP (Online Analytical Processing) system: A tool that has been optimized for analysis in addition to processing and can analyze data from multiple databases

OLTP (Online Transaction Processing) database: A type of database that has been optimized for data processing instead of analysis

Optimization: Maximizing the speed and efficiency with which data is retrieved in order to ensure high levels of database performance

P

Portfolio: A collection of materials that can be shared with potential employers

Primary key: An identifier in a database that references a column or a group of columns in which each row uniquely identifies each record in the table (Refer to foreign key)

Project manager: A person who handles a project's day-to-day steps, scope, schedule, budget, and resources

Project sponsor: A person who has overall accountability for a project and establishes the criteria for its success

Python: A general purpose programming language

Q

Query plan: A description of the steps a database system takes in order to execute a query

R

Resources: The hardware and software tools available for use in a database system

Response time: The time it takes for a database to complete a user request

Row-based database: A database that is organized by rows

S

Separated storage and computing systems: Databases where data is stored remotely, and relevant data is stored locally for analysis

Single-homed database: Database where all of the data is stored in the same physical location

Snowflake schema: An extension of a star schema with additional dimensions and, often, subdimensions

Star schema: A schema consisting of one fact table that references any number of dimension tables

Strategy: A plan for achieving a goal or arriving at a desired future state

Subject-oriented: Associated with specific areas or departments of a business

Systems analyst: A person who identifies ways to design, implement, and advance information systems in order to ensure that they help make it possible to achieve business goals

Systems software developer: A person who develops applications and programs for the backend processing systems used in organizations

T

Tactic: A method used to enable an accomplishment

Target table: The predetermined location where pipeline data is sent in order to be acted on

Throughput: The overall capability of the database's hardware and software to process requests

Transferable skill: A capability or proficiency that can be applied from one job to another

V

Vanity metric: Data points that are intended to impress others, but are not indicative of actual performance and, therefore, cannot reveal any meaningful business insights

W

Workload: The combination of transactions, queries, data warehousing analysis, and system commands being processed by the database system at any given time

Mark as completed