Get started with data modeling, schemas, and databases

Choose the right database

How data moves

Data-processing with Dataflow

Organize data in BigQuery

Review: Data models and pipelines

Video: Wrap-up
1 min

Reading: Glossary terms from week 1
10 min

Quiz: Weekly challenge 1
12 questions

[Optional] Review Google Data Analytics Certificate content

Glossary terms from week 1

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

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

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

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

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

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)

Functional programming language: A programming language modeled around functions

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

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

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

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

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)

Python: A general purpose programming language

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

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

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

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

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

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

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

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

E

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

I

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

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

M

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

P

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

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

S

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

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

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

Mark as completed