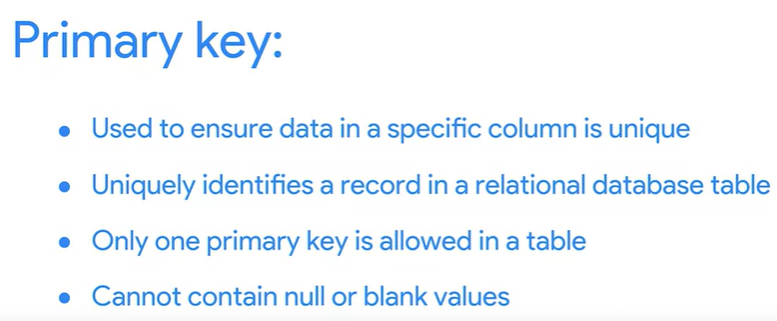
**Relational database** is a database that contains a series of related tables that can be connected via their relationships

**Primary key** is an identifier that references to a column in which each value is unique



**Foreign key** is a field within a table that is a primary key in another table

A white background with blue text

Description automatically generated

A table only contains a primary key but can contain lots of foreign keys

A primary key may also be constructed using multiple columns of a table. This type of primary key is called a **composite key**

**Metadata is used in database management to help data analysts interpret the contents of the data within the database**

**3 common types of metadata**

**Descriptive**

**Structural**

**Administrative**

**The benefits of metadata**

**Reliability**

Data analysts use reliable and high-quality data to identify the root causes of any problems that might occur during analysis and to improve their results. If the data being used to solve a problem or to make a data-driven decision is unreliable, there’s a good chance the results will be unreliable as well.

Metadata helps data analysts confirm their data is reliable by making sure it is:

* Accurate
* Precise
* Relevant
* Timely

It does this by helping analysts ensure that they’re working with the right data and that the data is described correctly. For example, a data analyst completing a project with data from 2022 can use metadata to easily determine if they should use data from a particular file.

**Consistency**

Data analysts thrive on consistency and aim for uniformity in their data and databases,  and metadata helps make this possible. For example, to use survey data from two different sources, data analysts use metadata to make sure the same collection methods were applied in the survey so that both datasets can be compared reliably.

When a database is consistent, it’s easier to discover relationships between the data inside the database and data that exists elsewhere. When data is uniform, it is:

* Organized: Data analysts can easily find tables and files, monitor the creation and alteration of assets, and store metadata.
* Classified: Data analysts can categorize data when it follows a consistent format, which is beneficial in cleaning and processing data.
* Stored: Consistent and uniform data can be efficiently stored in various data repositories. This streamlines storage management tasks such as managing a database.
* Accessed: Users, applications, and systems can efficiently locate and use data.

Together, these benefits empower data analysts to effectively analyze and interpret their data.

**Explore public datasets**

**Open data** helps create a lot of **public datasets** that you can access to make data-driven decisions. Here are some resources you can use to start searching for public datasets on your own:

* The [Google Cloud Public Datasets](https://cloud.google.com/public-datasets) allow data analysts access to high-demand public datasets, and make it easy to uncover insights in the cloud.
* The [Dataset Search](https://datasetsearch.research.google.com/) can help you find available datasets online with keyword searches.
* [Kaggle](https://www.kaggle.com/datasets?utm_medium=paid&utm_source=google.com+search&utm_campaign=datasets&gclid=CjwKCAiAt9z-BRBCEiwA_bWv-L6PpACh6RzmrJjQjmNGCCE7kky1FCtc6Jf1qld-4NwDMYL0WsUyxBoCdwAQAvD_BwE) has an Open Data search function that can help you find datasets to practice with.
* Finally, [BigQuery](https://cloud.google.com/bigquery/public-data) hosts 150+ public datasets you can access and use.

**Public health datasets**

1. [Global Health Observatory data](https://www.who.int/data/collections): You can search for datasets from this page or explore featured data collections from the World Health Organization.
2. [The Cancer Imaging Archive (TCIA) dataset](https://cloud.google.com/healthcare/docs/resources/public-datasets/tcia): Just like the earlier dataset, this data is hosted by the Google Cloud Public Datasets and can be uploaded to BigQuery.
3. [1000 Genomes](https://cloud.google.com/life-sciences/docs/resources/public-datasets/1000-genomes): This is another dataset from the Google Cloud Public resources that can be uploaded to BigQuery.

**Public climate datasets**

1. [National Climatic Data Center](https://www.ncei.noaa.gov/products): The NCDC Quick Links page has a selection of datasets you can explore.
2. [NOAA Public Dataset Gallery](https://www.climate.gov/maps-data/datasets): The NOAA Public Dataset Gallery contains a searchable collection of public datasets.

**Public social-political datasets**

1. [UNICEF State of the World’s Children](https://data.unicef.org/resources/dataset/sowc-2019-statistical-tables/): This dataset from UNICEF includes a collection of tables that can be downloaded.
2. [CPS Labor Force Statistics](https://www.bls.gov/cps/tables.htm): This page contains links to several available datasets that you can explore.
3. [The Stanford Open Policing Project](https://openpolicing.stanford.edu/): This dataset can be downloaded as a .csv file for your own use.

**Glossary terms from module 3**

**Terms and definitions for Course 3, Module 3**

**Administrative metadata:** Metadata that indicates the technical source of a digital asset

**CSV (comma-separated values) file:** A delimited text file that uses a comma to separate values

**Data governance:** A process for ensuring the formal management of a company’s data assets

**Descriptive metadata:** Metadata that describes a piece of data and can be used to identify it at a later point in time

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

**FROM:** The section of a query that indicates where the selected data comes from

**Geolocation:** The geographical location of a person or device by means of digital information

**Metadata:** Data about data

**Metadata repository:** A database created to store metadata

**Naming conventions:** Consistent guidelines that describe the content, creation date, and version of a file in its name

**Normalized database:** A database in which only related data is stored in each table

**Notebook:** An interactive, editable programming environment for creating data reports and showcasing data skills

**Primary key:** An identifier in a database that references a column in which each value is unique (Refer to foreign key)

**Redundancy:** When the same piece of data is stored in two or more places

**Schema:** A way of describing how something, such as data, is organized

**SELECT:** The section of a query that indicates the subset of a dataset

**Structural metadata:** Metadata that indicates how a piece of data is organized and whether it is part of one or more than one data collection

**WHERE:** The section of a querythat specifies criteria that the requested data must meet

**World Health Organization:** An organization whose primary role is to direct and coordinate international health within the United Nations system