

Storage options for transformed data

So far, you've learned that after locating and processing data, you need to store it. Luckily, there are several storage options to choose from, including storing data locally, or in the cloud. In this reading, you'll learn more about where transformed data is stored, and explore examples of tools that manage this task.

The data storage phase

Storing data is the action of saving data so a computer or device can access it. Every time you save a photo, write a note on your phone, or enter data into a spreadsheet, you're relying on data storage. In the world of cloud data analytics, storage becomes even more important because of the large volumes of data involved. Storage is crucial, because without it, there wouldn't be a space to keep the refined information for analysis or future use.

Types of data storage

As a reminder, the storing phase of the data journey comes after the collecting and processing phases. There are several ways to store data in cloud computing, including options for data that hasn't yet been processed. Organizations may choose to store their pre-processed data if they don't have a particular use-case, or if they want to start new analysis from scratch. Review the table to learn more about the types of data storage available and their uses:

Type of data storage	Uses
Database	Organized repositories of structured or unstructured data
Data warehouses	Specifically designed database for answering business questions
Data lakes	Storage that houses raw data
Cloud storage	Online storage for all kinds of data, typically serving as a part of a data lake

There are plenty of storage options for organizations to consider, including hybrid options like a

data lakehouse. These hybrid storage options have characteristics of more than one storage solution, giving organizations more flexibility for their data.

Databases

Databases are like digital filing cabinets; they're structured to hold data in an organized way, making it faster to find the exact piece of information you're looking for. Databases are typically classified into two main categories according to their function: transactional and analytical.

Transactional databases support applications like smartphone applications, websites, or payment systems. Transactional databases are usually always on, and support the reading and writing of small amounts of data at high speeds.

Analytical databases are designed to store and process huge amounts of data in seconds or minutes. They help businesses answer questions about vast amounts of data to make business decisions. Analytical databases are sometimes called online analytical processing (OLAP) databases. Common types of databases include relational databases and NoSQL databases. Relational databases contain a series of tables that can be connected to form relationships, and usually supports SQL. NoSQL Databases store massive amounts of structured and unstructured data that can be retrieved at high speeds.

Data warehouses

A data warehouse is a specially designed database that consolidates data from multiple source systems for data consistency, accuracy, and efficient access. Services like BigQuery are modern data warehouses where you can store data and run powerful queries.

Data lakes

A data lake is a storage system that stores large amounts of raw data in its original format until it's needed. Because the data is stored in its raw form, data lakes can cost less than data warehouses, and can offer more flexibility in how the data is used. Data lakes offer a variety of features like scalability, security, and performance. They can be used to store and process large amounts of data from a variety of sources, including structured, semi-structured, and unstructured data. Google Cloud Storage, Amazon S3, and Apache Hadoop distributed file system (HDFS) are common examples of data lake storage.

Cloud storage

Cloud storage solutions enable users to save data online so it can be accessed from any device, anytime, anywhere. Cloud storage also offers users flexibility to access their data. Common examples of cloud storage include Google Drive, DropBox, iCloud Drive, Google Cloud Storage, and Amazon S3.

Cloud data storage is a solution that enables organizations to help keep, access, and maintain digital data on off-site, cloud-based servers. Data warehouses can also be deployed in the cloud, because they're hosted on remote servers by a cloud service provider.

Key takeaways

Data storage is crucial in cloud data analytics, serving as a space for processed data to live and be used for analysis. From data warehouses like BigQuery, to data lakes like Google Cloud Storage, there are a variety of tailored options for different organizational needs. Cloud solutions like Google Drive or DropBox help empower users to access their data anytime, from any device, emphasizing the growing importance of flexibility in data storage and retrieval.

Resources for more information

Use the following resource to learn more about storing data with BigQuery:

- This resource can help you learn the basics of using Google's BigQuery for storing and querying data: [What is BigQuery?](#)