

Versioning and holds for accurate data

So far, you've learned that versioning and holds are ways to organize your data and files so that everyone on your team has access to the most accurate and recent versions. Versioning is the process of creating a unique way of referring to data. And a hold is a policy placed on a dataset that prevents its deletion, or prevents deletion capabilities for certain accounts. In this reading, you'll learn more about holds and versioning, and review examples of how they work in data analysis.

Version control

While version control is usually associated with managing changes to software code, versioning can also be useful in the data world. There are three types of version control: local, centralized, and distributed.

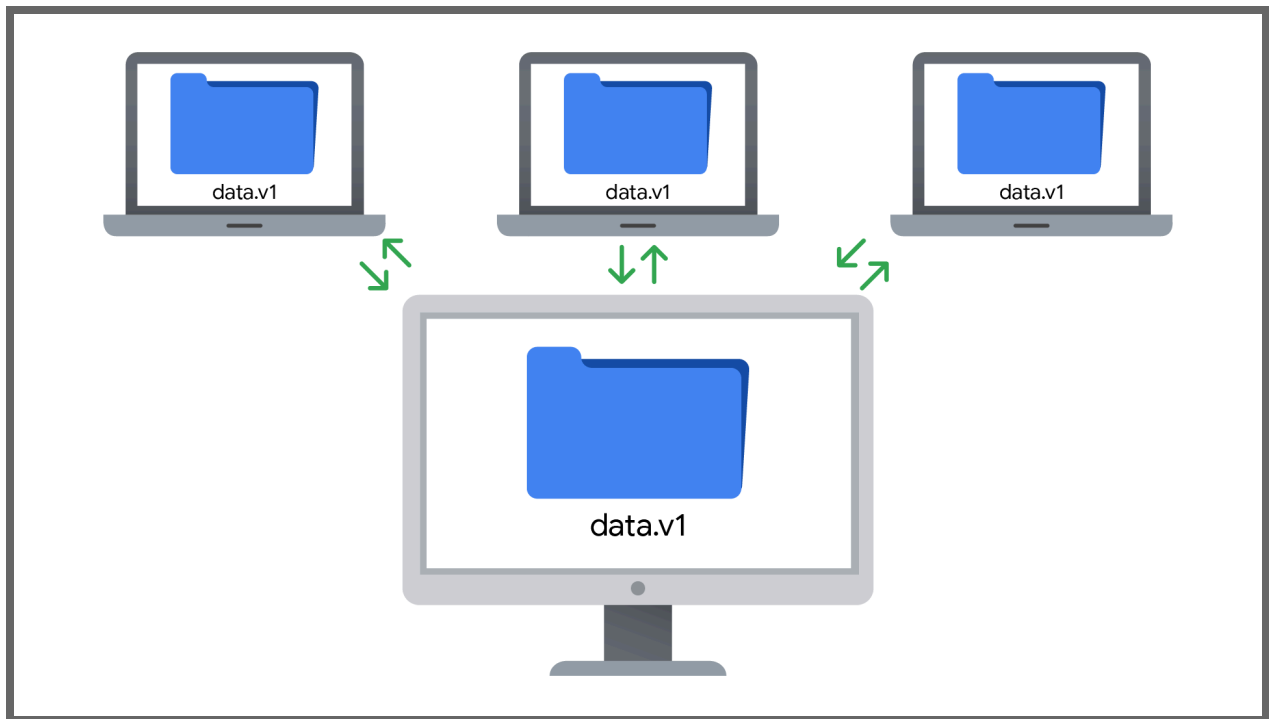
Localized control

When working as a data analyst, you may save data files, Excel sheets, or other reports on your computer. When updating files, it's important to use versioning to ensure that you're naming files so that it's clear which file is the most recent version. You can also use naming conventions provided by your organization to ensure there is a consistent naming method.



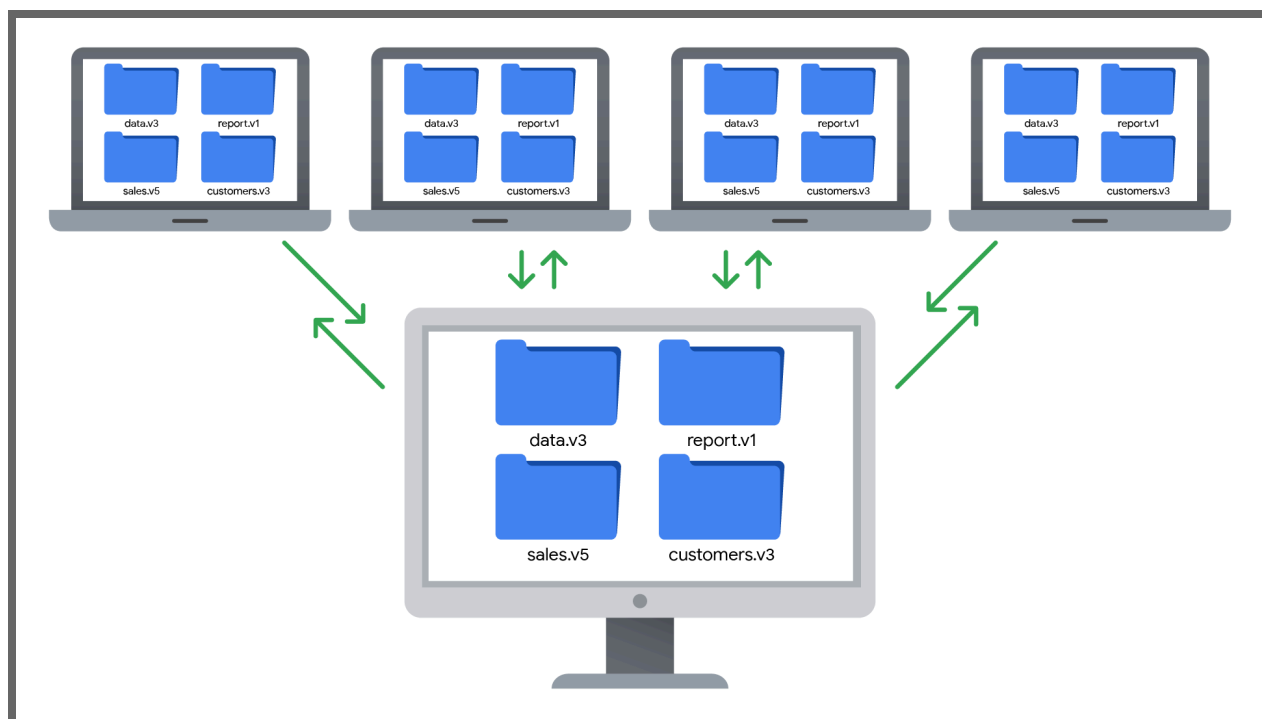
Centralized control

A centralized version control system means that all of your files, including your coworkers', are checked into a single database. This means anyone can access these files and can update them. To ensure that everyone on your team can find the most recent version of each file, use a consistent naming convention.



Distributed control

In a distributed control system, each team members' individual computer mirrors the central database. This means everyone has access to the same files in a centralized system, but this access appears on each individual computer. As you work on your individual computer, your changes are live in the central database.



Holds

A hold can be placed on a file to ensure that it cannot be deleted by someone without permissions to delete. In the Google ecosystem, holds override retention rules. Multiple holds can be placed on the same file. This means that you can create different holds for different staff members. Holds can also be specific to individual files. This means you may have a hold associated with your account, or a hold on one file but not another.

Key takeaways

Organizing data and files can be a huge challenge when there are multiple versions and multiple people working on the same project. Using version control or holds can provide added safeguards for your work. Version controls and holds can also ensure that everyone is working with the most up-to-date data and information.

Resources for more information

Learn more about how version control works in the data world with these links:

- Longwood Research Data Management provides information about different types of version control software:
<https://datamanagement.hms.harvard.edu/collect-analyze/version-control>
- NYU Libraries provides guidelines for version control in research data management:
https://guides.nyu.edu/data_management/version-control