Jessica Craig
IS 270
March 18, 2020

Dear UCLA Information Technology Planning Board,

The following policy brief is intended to introduce the emerging technology of Google Drive and evaluate how it might fit into our UCLA community members workstyle. When choosing to implement a new technology on a campus-wide scale, we need not only an understanding of the technology itself but an understanding of the community who will be utilizing it. The UCLA community thrives off of collective effort and the ability to communicate. The need to work together and join our individual intellect is one of the key aspects of academia and advancement, and it absolutely requires us to be able to share our work. This is one of the key reasons to address Google Drive as a potential tool for our community. Our campus exceeds beyond one location and the infrastructure of our technologies needs to accommodate that. Our community is comprised of people with very different schedules and locations, and during the limited hours they might actually all align, of course, there is not a sufficient amount of time to complete the work. The disconnects of time and place between collaborators may lead to hinder the research process. Cloud computing on a team-based collaboration platform, like Google Drive, may mitigate that issue. While there are several ways that a technology such as Google Drive may benefit our university, there are also potential challenges and issues that could arise with its implementation. This policy brief intends to cover the technology's definition, background, advantages, disadvantages, and future expectations in order to provide a complete evaluation of the information technology.

# Google Drive: Evaluating its Technical Capabilities for UCLA Workspaces

*Overview*

- Google Drive is a cloud-based storage system and has a complex technical stack. The infrastructure should be examined with depth during the consideration of its adoption.
- While Google Drive may have the potential to serve our community, cloud storage systems do pose higher risks in regard to data privacy.
- As a data storage service, Google Drive allows for a specific amount of information to be stored. The storage limit may not be sufficient for all of our community members.
- As an emerging technology, cloud storage systems pose a new territory for consumers who may not be familiar with how "the cloud" operates and maintains their data.
- Technology is constantly changing and being updated. How does Google Drive keep up with the market?

*Introduction*

Google Drive is a cloud-based file storage service and a file-sync solution, meaning that it can synchronize user data across several different devices. Google Drive allows users to save their data through the web interface which then connects their data to the off-site server through an Internet connection. They are able to connect via any device through the web, so there is no need to install specific software. This policy brief will introduce

ways that this technology can be implemented and address possible challenges of its use.

*Background*

Cloud storage is a developing technology of the 21st century and its adoption across platforms has been widely increasing. The introduction of cloud storage to the market was an advancement and alternative to the previous products and services for information storage. Previously, users relied exclusively on a variety of physical storage devices. The modern timeline of data storage starts in the 1970s with the magnetic tape floppy discs, cassette tapes and evolved to using CDs, USBs and other hard drives. But these devices exist in a tangible, portable, and modular form. Cloud services were up and running by the late 1990s, and as Marcus Oppitz and Peter Tomsu explain, "providing storage and computing capacity as a cloud service became one of the first applications of cloud service technology for consumers as well as for corporations. The idea was simple but convincing and based on the idea of resource sharing."[1] In 2005, one of the first widely successful cloud-based storage services was introduced, called Box.com and was followed by several other cloud computing technology such as Amazon Web Services and Rackspace Cloud in 2006, SpiderOak in 2007, DropBox and Microsoft OneDrive in 2008, WeTransfer in 2009, Google Storage in 2010, Apple's iCloud in 2011, and finally, Google Drive in 2012.[2] Since, Google Drive has grown to be one of the leading cloud-based storage solution for users across the world.

*Technical Infrastructure*

Google Drive relies on their cloud-based network. That is the essential function of the service, as it provides users with access to the cloud through their Google account. Users are given 15GB of free storage to start out with. This limit is applied not only to Google Drive, but Gmail and Google Photos as well. Google Drive is comprised of several applications; the core collaborative applications include Google Docs, Google Slides, Google Sheets, while other service applications include Gmail, Google Photos, Google Calendar, and Google Maps. These are all registered with one Google account. Documents created in any of these applications may be centrally stored in the Google Drive.

The technical infrastructure of Google Drive is complex, but the stack has three major storage components: node storage, cluster-level file systems, and structured storage.[3] Node storage is responsible for exporting and verifying disks on the network. It can be understood as the building block for higher-level storage and allows for resource sharing across users. The two cluster-level file systems that Google relies on are Google File Sharing (GFS) and Colossus. GFS was introduced in 2001 by Google and was intended to keep up with the increasing demands of their data processing. "[GFS is] a scalable distributed file system for large distributed data-intensive applications. It provides fault tolerance while running on inexpensive commodity hardware, and it delivers high aggregate performance to a large number of

[1] Marcus Oppitz and Peter Tomsu. *Inventing the Cloud Century.* (Cham: Springer International Publishing AG, 2018).
[2] Ibid.
[3] Andrew Fikes, "Storage Architecture and Challenges," July 29, 2010, https://cloud.google.com/files/storage_architecture_and_challenges.pdf.

clients."[4] Colossus is considered the next-generation cluster level file system.[5] For the structured storage, Google uses BigTable or Spanner. BigTable uses either GFS or Colossus for file storage and is a "distributed storage system for managing structured data that is designed to scale to a very large size: petabytes of data across thousands of commodity servers."[6] Spanner is explained as planet-scale structured storage, which would be the next generation of BigTable.[7] Lastly, for the data processing level in the stack, Google uses MapReduce. "MapReduce is a programming model and an associated implementation for processing and generating large data sets. Users specify a map function that processes a key/value pair to generate a set of intermediate key/value pairs, and a reduce function that merges all intermediate values associated with the same intermediate key."[8] To simplify this process, it can be thought of in terms of the hardware rather than software. As far as hardware goes, the central components include the user device (whether that be mobile smartphone, laptop, desktop, tablet, etc.) that connects to Google's control nodes through the Internet, which then accesses the data servers. These data servers are all connected to each other.

*Key Challenges and Issues*

1. Data Security

Data privacy is an increasingly urgent topic for tech consumers and should be a key area of examination for campus leaders when considering Google Drive. Since users are sharing their data with a distant and remote server, there are more possibilities for data breaches than there would be with a physical storage device. On a similar note, because Google Drive allows users to share among each other, there may be unintentional sharing with other users. Users need to understand how they can choose to share their data with each other, as Google Drive gives several options for it. The sharing options include sharing with the public (no access control), sharing by generating a secret URL to the file, or other additional options. For example, a UCLA user would have the option to share within the campus community either with or without the specific file URL. The creator may also choose to share with specific users by adding their email address. This is the most controlled option of file sharing, as it would require the collaborator to have permission to the file and they would need to sign into their account before accessing any of the data (assuming they have a Google account too). Of course, there is also the possibility of not sharing the file at all, which is the default setting whenever a file is created. While data sharing is a key component of Google Drive, it is important to take note of these examples. The multiple sharing options could lead to sharing by mistake. Our students, faculty, and staff should be made aware of possible data insecurities with Google Drive, as they relate to off-site servers and other users. Once they are aware, there are some

---

[4] Sanjay Ghemawat, Howard Gobioff, and Shun-Tak Leung, "The Google File System." *Proceedings of the 19th ACM Symposium on Operating Systems Principles*, (2003).

[5] Andrew Fikes, https://cloud.google.com/files/storage_architecture_and_challenges.pdf.

[6] Fay Chang, Jeffrey Dean, Sanjay Ghemawat, Wilson C. Hsieh, Deborah A. Wallach, Mike Burrows, Tushar Chandra, Andrew Fikes, and Robert E. Gruber, "Bigtable: A Distributed Storage System for Structured Data." *ACM Transactions on Computer Systems* 26, no. 2 (2008).

[7] Andrew Fikes, https://cloud.google.com/files/storage_architecture_and_challenges.pdf.

[8] Jeffrey Dean and Sanjay Ghemawat. "MapReduce: Simplified Data Processing on Large Clusters." *Communications of the ACM* 51, no. 1 (2008).

preventative measures that may be taken with instruction and awareness of possible consequences of data sharing within Google Drive.

2. Misunderstanding of the Product

Although cloud computing and cloud-based storage are widely used today, is it not uncommon to be confused by what "the cloud" really is. Users have accepted how it works for their daily life, yet comprehension of the underlying technology may be weak. This misunderstanding of cloud services poses a key issue, as it adds to the threat of data security. The way that Google Drive stores the data is invisible to the user. When users are giving their information away without knowing exactly how it is managed, they are trusting their personal information is in good hands. This could lead to users uploading sensitive or restricted material to their Google Drive. Since Google drive is a web-based platform, users must be aware of data leakage on the Internet. While the data might be understood to be locked into their Google account, it is still at risk as they are introducing it to the Web, which is always at risk. We need to ensure that our adoption and promotion of each of our campus technology platforms are accompanied by the knowledge of how the cloud technology works, so that our community can be informed users.

3. Storage Limits

Google Drive designates 15GB of storage to each new user. This is no-cost limit and can be distributed across the Google apps. A key issue when evaluating any storage system is the financial cost of storage. We cannot be sure that all students and employees would remain under the 15GB storage limit. The assumption that every UCLA is using the same amount of data storage would be a flaw in our service to them. The amount of storage each student needs for their academic purposes will likely depend on their course load, extracurricular activities, whether or not they are enrolled in online courses, and possibility their major of study. In the case that the user does need more data than Google allows for free, we should take into account our responsibility in promoting a service that eventually increases their academic fees. There may be the option to allot our UCLA users more storage space, with the financial cost on the institution, rather than the individuals. Either option should be considered

*Future Trends*

Cloud storage has been present and used for over a decade now, and it is not likely to go anywhere soon. As users adjust to the convenience of cloud storage, the products and services accommodating it will proliferate. There are already several competing businesses inventing new platforms for user friendly cloud storage. However, Google Drive undoubtedly stands out, as it is a product of one of the leading tech companies of our time. As Google introduces products to the market, consumers should be aware of the possible advantages and disadvantages of the business' domination of the industry. This is especially the case for communities choosing to adopt it on a wide scale, like UCLA. Google products are known for optimizing interoperability, user experience, and innovation. Alternatively, Google as a business will rely on its reputation and history of leadership. This could easily distract consumers, trusting that Google has the best products by default of their own prestige.

Bibliography

Chang, Fay, Jeffrey Dean, Sanjay Ghemawat, Wilson C. Hsieh, Deborah A. Wallach, Mike Burrows, Tushar Chandra, Andrew Fikes, and Robert E. Gruber. "Bigtable: A distributed Storage System for Structured Data." *ACM Transactions on Computer Systems* 26, no. 2 (2008): 1-26.

Dean, Jeffrey, and Sanjay Ghemawat. "MapReduce: Simplified Data Processing on Large Clusters." *Communications of the ACM* 51, no. 1 (2008): 107-113.

Fikes, Andrew. "Storage Architecture and Challenges," July 29, 2010, https://cloud.google.com/files/storage_architecture_and_challenges.pdf.

Ghemawat, Sanjay, Howard Gobioff, and Shun-Tak Leung. "The Google File System." *Proceedings of the 19th ACM Symposium on Operating Systems Principles*, (2003): 20-43.

Oppitz, Marcus, and Peter Tomsu. *Inventing the Cloud Century*. Cham: Springer International Publishing AG, 2018.