

1. Advantages of Cloud Data Storage

Storing extremely large volumes of information on a local area network (LAN) is expensive. High capacity electronic data storage devices like file servers, Storage Area Networks (SAN) and Network Attached Storage (NAS) provide high performance, high availability data storage accessible via industry standard interfaces. However, electronic data storage devices have many drawbacks, including that they are costly to purchase, have limited lifetimes, require backup and recovery systems, have a physical presence requiring specific environmental conditions, require personnel to manage and consume considerable amounts of energy for both power and cooling.

Cloud data storage providers, such as AmazonS3, provide cheap, virtually unlimited electronic data storage in remotely hosted facilities. Information stored with these providers is accessible via the internet or Wide Area Network (WAN). Economies of scale enable providers to supply data storage cheaper than the equivalent electronic data storage devices.

Cloud data storage has many advantages. It's cheap, doesn't require installation, doesn't need replacing, has backup and recovery systems, has no physical presence, requires no environmental conditions, requires no personnel and doesn't require energy for power or cooling. Cloud data storage however has several major drawbacks, including performance, availability, incompatible interfaces and lack of standards.

2. Disadvantages of Cloud Data Storage

Performance of cloud data storage is limited by bandwidth. Internet and WAN speeds are typically 10 to 100 times slower than LAN speeds. For example, accessing a typical file on a LAN takes 1 second, accessing the same file in cloud data storage may take 10 to 100 seconds. While consumers are used to slow internet downloads, they aren't accustomed to waiting long periods of time for a document or spreadsheet to load.

Availability of cloud data storage is a serious issue. Cloud data storage relies on network connectivity between the LAN and the cloud data storage provider. Network connectivity can be affected by any number of issues including global networks disruptions, solar flares, severed underground cables and satellite damage. Cloud data storage has many more points of failure and is not resilient to network outages. Network outages mean the cloud data storage is completely unavailable.

Cloud data storage providers use proprietary networking protocols often not compatible with normal file serving on the LAN. Accessing cloud data storage often involves ad hoc programs to be created to bridge the difference in protocols.

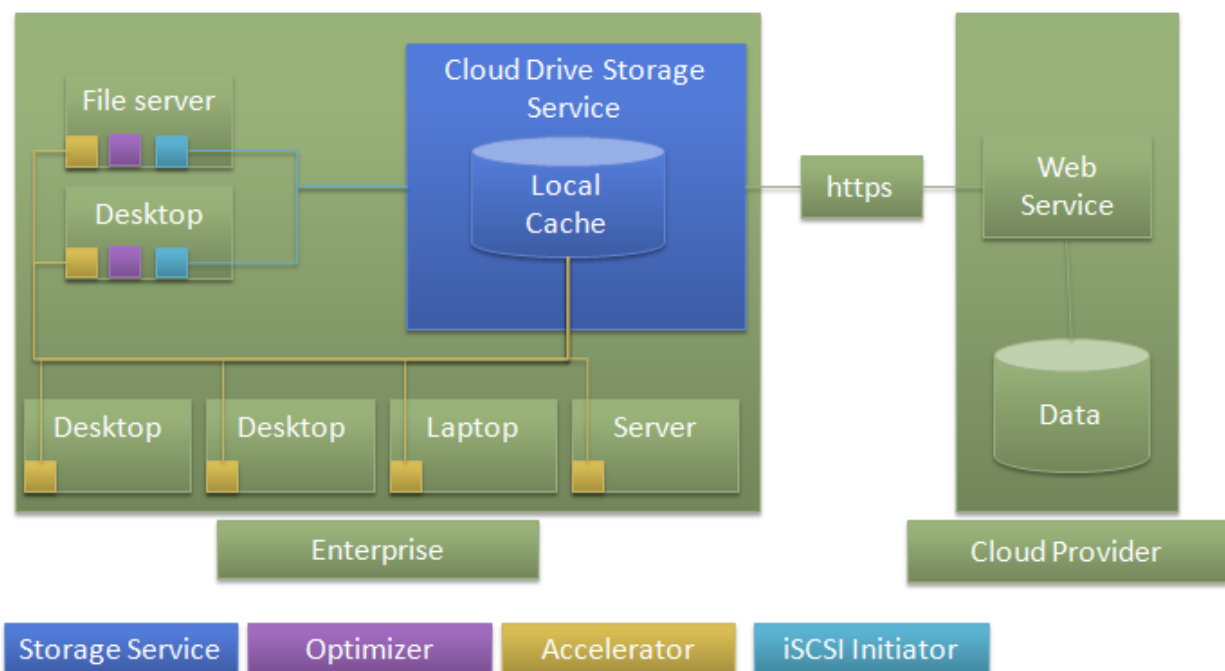
The cloud data storage industry doesn't have a common set of standard protocols. This means that different interfaces need to be created to access different cloud data storage providers. Swapping or choosing between providers is complicated as their protocols are incompatible.

The cloud drive data storage is small enough to be used on laptops while having enterprise class features that enable it to be scaled out to the largest organization.

3. Cloud Drive Architecture

Cloud Drive is a gateway to cloud storage. Cloud Drive supports many cloud data storage providers including Microsoft Azure, Amazon S3, Amazon EC2, Rackspace, EMC Atmos, Nirvanix, GoGrid, vcloud, Zetta, Scalify, Dunkel, Mezeo, Box.net, Webdav and FTP. Cloud Drive hides the complexity of the underlying protocols allowing you to deploy cloud storage as simply as deploying storage via an IP SAN.

Cloud Drive is like an IP SAN that never runs out of space. As usage increases, Cloud Drive starts “offloading” data to the cloud data provider. Cloud Drive caches and optimizes traffic to/from cloud storage dramatically increasing performance and availability while also reducing network traffic.



Computers on the LAN access data via the block based iSCSI protocol. The storage service communicates via an internet connection with the cloud data storage provider. When the iSCSI initiator saves data to the data storage server, it initially stores the data in the local cache. Each data unit is uniquely located within the local cache and is flagged as either “online” in the local cache or “offline” in the cloud data storage provider. All data units in the local cache are checked periodically for usage. Least recently used (or “dormant”) data units are uploaded to the cloud data storage provider, flagged as “offline” and deleted from the local cache.

4. Cloud Drive Storage Service

The cloud drive storage service is simple to install and configure. It can be installed on a range of hardware, from laptop for personal use, a server in the office, or a cluster of high end 64 bit servers for the enterprise. Once the service is installed and configured, many clients can connect to it using the iSCSI protocol.

The storage service reduces the data storage requirements while maintaining performance by moving the least recently used data to the cloud data storage provider as well as one or more of the data storage accelerators. Cloud Drive accelerates performance by assuming that actual writes to data can happen anytime before a subsequent read to the same data. Cloud Drive accelerates performance by scheduling this “delayed” write data to periods of low activity and by not downloading data from the cloud data storage provider when the “delayed” write data has wholly overwritten data stored in the cloud. Cloud Drive further accelerates performance by assuming that delete operations can happen anytime after the data is downloaded.

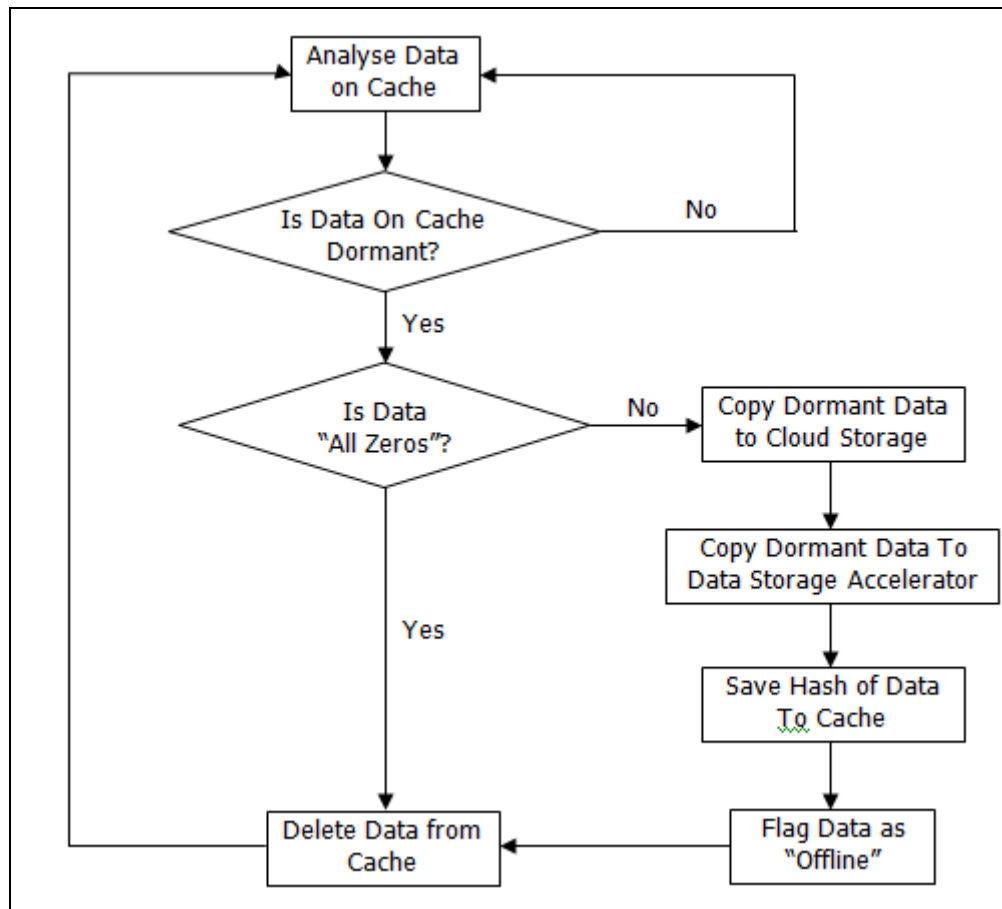


Figure 1 – Upload data to Cloud Storage

5. Cloud Drive Optimizer

An optional component, cloud drive optimizer, improves performance, reduces bandwidth and reduces your data storage requirements. The optimizer should be installed on all iSCSI clients using the cloud drive storage service.

The data storage optimizer has access to the virtual hard drive to optimize the data stored in the local cache. The optimizer periodically reads virtual hard drive or virtual file share metadata including directories, filenames, permissions and attributes in order to maintain that data in the local cache. In this way, the data storage optimizer also accelerates performance of the data storage server by preventing data other than file data from being identified as “dormant”. The data storage optimizer also reduces storage requirements of the data storage server by periodically overwriting “all zeros” to unused parts of the virtual hard drive. The data storage optimizer is also adapted to periodically run disk checking utilities against the virtual hard drive to prevent important internal file systems data structures from being marked as dormant.

6. Cloud Drive Network Accelerator

An optional component, cloud drive network accelerator, improves the performance and availability of the Storage Service. This component can be installed on all computers in the home, office or enterprise.

The network accelerators allow the office to reclaim all those “small spaces” of data storage already available on the 10’s, 100’s or 1000’s of computers within the enterprise. A typical office with 100 computers having on average 100 GB of space available could potentially reclaim $100 \times 100 \text{ GB} = 10 \text{ TB}$ of data storage space by reclaiming and consolidating this unused space. Network accelerators boost performance and improve resilience to slowness or unavailability of the cloud data storage providers by redundantly storing data uploaded to the cloud data storage provider on the local network in the already existing “unused spaces”

Network accelerators work like a massive cache within the enterprise. In the above example, the Storage Services local cache is complimented by a 10 TB onsite cache running throughout the enterprise

7. Cloud Drive Solution

Cloud Drive increases the apparent availability of the cloud data storage provider. If the local cache satisfies 99% of requests for data without requiring the cloud data storage provider, the apparent availability of the cloud data storage provider is increased 100 fold and 99% of data accesses occur at local network speeds rather than the network connection speeds to the cloud data storage provider. Cloud Drive also manages the data formatting and communication with the cloud data storage provider while allowing seamless access to data using standard protocols such as iSCSI and NFS. Further, Cloud Drive allows concurrent processing of read and writes requests to different data as well as synchronized and serialized access to the same data.

Cloud Drive virtualizes data storage by allowing a limited amount of physical data storage to appear many times larger than it actually is. Cloud Drive allows fast, expensive physical data storage to be supplemented by cheaper, slower remote data storage without incurring substantial performance degradation. Cloud Drive also reduces the physical data storage requirements to a small fraction of the total storage requirements, while the rest of the data can be “offloaded” into slower, cheaper online cloud data storage providers.