Microsoft

# StorSimple Solution for File Share Deployments

September 2013

# Table of contents

# Copyright information

# Introduction

The growth of data storage continues to be more prominent with each passing year, and with that the growth in unstructured data, compared to structured data, continues to grow as well. In 2005, it was projected that approximately 36% of stored data was unstructured, contrasting with the projection for 2014 were approximately 77% of data will be unstructured.  Common examples of unstructured data range from Microsoft Office files and Adobe PDF documents to media content for video, graphic and audio files. Unstructured content is harder to organize and prioritize without understanding the content.  For example, a user has created two word documents, one has a schedule for an event that will be edited and updated many times over the next few weeks by various users and the second document has an event summary which will remain static and unchanged. Both are identified as the same file type and created by the same user however one document is a high priority document since it will need to be accessed frequently while the other document can most likely be archived.  However, since both files have all the same attributes there would be no way to distinguish this without understanding the content of the file, therefore, likely both files will be stored on expensive, high performance storage.  On a larger scale, file share deployments, with many files, would have the same difficulty without an intelligent way to manage that placement of data within the storage infrastructure.

The StorSimple solution is a cloud integrated storage (CiS) solution that addresses the complexities of unstructured data in file share deployments. The StorSimple solution uses Windows Azure cloud storage as an extension of the on premise solution and automatically tiers data across on premise storage and cloud storage. This paper will discuss the StorSimple solution for file share deployments and the attributes of the cloud integrated solution that simplify the complexity associated with large quantities of unstructured data.

## Target Audience

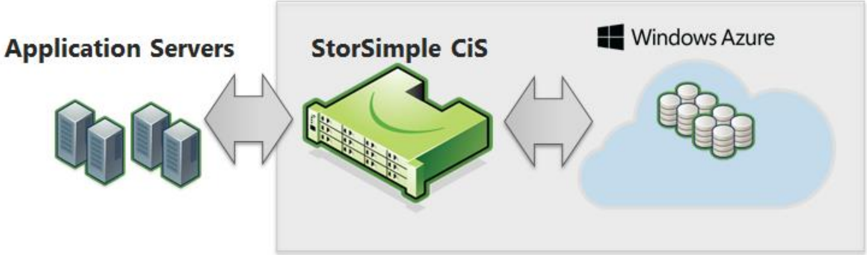
Those reading this document may have varying degrees of familiarity with the StorSimple solution, therefore, the objective of this paper is multifaceted. For those less familiar with the StorSimple solution, the intention of this paper is to provide an overview of the StorSimple solution and how it is differentiated from other storage solutions, both traditional and cloud integrated. For those familiar with the StorSimple solution, and have or will purchase the solution, the intention of this paper is to provide a set of best practices, functional and operation, for maximum optimization.

In general, this paper discusses topics with a specific focus on file share deployments. Some information may generally apply to StorSimple solutions and is included to provide the most complete set of information.

# StorSimple Solution Overview

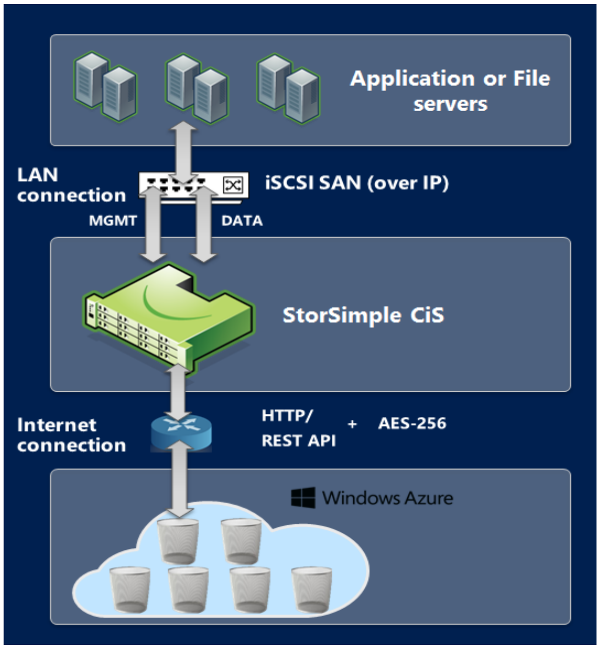
The StorSimple solution is an on premise SAN storage device that resides in the local customer data center. The software provided with the StorSimple solution allows the on premise storage device to connect with the Windows Azure cloud, creating another tier of storage as an extension of the on premise storage solution. The StorSimple solution does not require any additional integration with the application programming interfaces (APIs) to connect with the Windows Azure cloud. The on premise storage and cloud storage resources are seen as a single container of storage.

**Figure 1) Storage infrastructure with StorSimple and Windows Azure**



The StorSimple solution uses iSCSI to communicate with the application or host. Application servers can be physical or virtualized (Hyper-V or VMware). On the other end, the StorSimple solution uploads data to the cloud over a secure HTTP internet connection.  Data is encrypted using AES-256-CBC prior to storing it on the cloud storage service.  The below figure shows the infrastructure layout.

**Figure 2) Infrastructure overview with StorSimple**

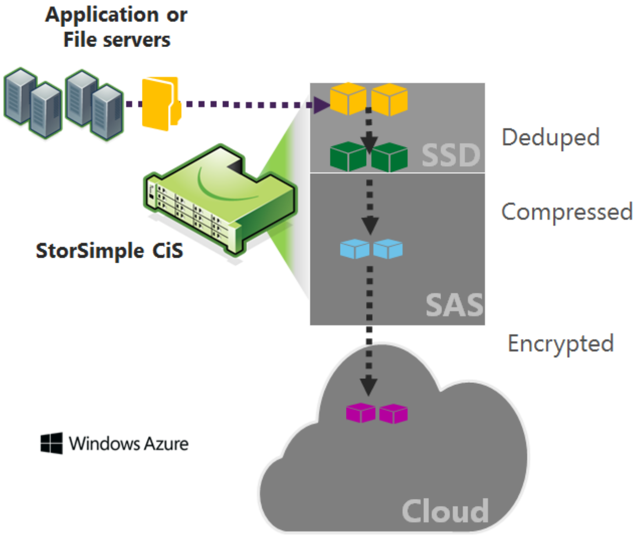


## Automatic tiering

The StorSimple solution is a SAN solution that processes data in blocks. When data is sent from a client or host application it is broken into chunks of data, blocks, to be stored in the SSD tier of the StorSimple solution.  The SSD tier is generally small in capacity and when it reaches a threshold for fullness, the oldest data in the SSD tier will be deduplicated, compressed and then pushed, or "spilled", to the HDD tier. The HDD tier continues to fill until it approaches the threshold for fullness, the oldest blocks will be encrypted and pushed to Azure cloud storage.

For a read request, if the data blocks are in SSD they are read directly from the SSD layer. However, if the blocks, corresponding to the requested data, reside on HDD or cloud storage then they are pulled back into the SSD tier. The data is decompressed and unencrypted prior to being pulled back into the SSD tier, however it remains deduplicated. When it comes to read blocks that were pulled back from lower tiers, the SSD tier acts as a read cache with the corresponding performance of that tier.

**Figure 3) StorSimple data flow model, including deduplication, compression and encryption.**



For more information on StorSimple deduplication refer to online documentation for Deduplication Reduction of Data.

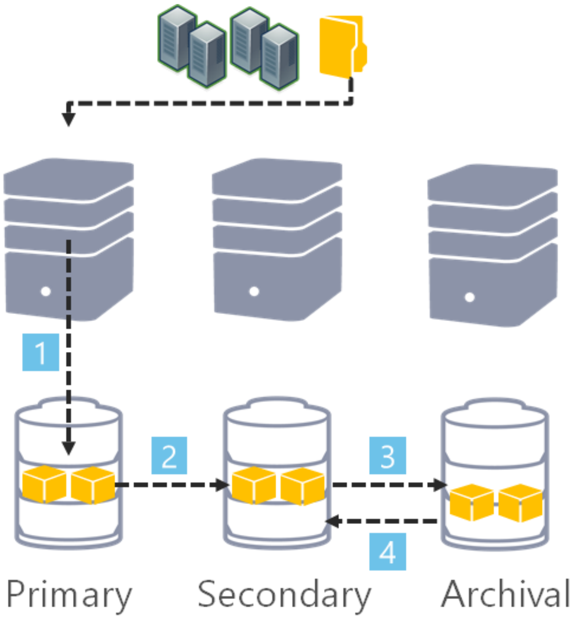
# Benefits of StorSimple solution for file share deployments

## Simplified management of file share data

StorSimple is a block-storage solution, however it is especially beneficial for file share environments where there are large sets of data that have no distinguishing features to identify the priority associated with the file or the components of the file. The automated life cycle of data, as discussed in the previous section, allows data to continually move within the three tiers of storage, SSD, HDD, and cloud storage without policy or administrative management over the data. Data that is high priority data has the required read and write performance consistent with SSD storage while other data is pushed to the HDD and cloud storage tiers. This is especially significant for files where only a portion of the file would need to be retained in SSD storage. For example, a file has multiple components: metadata and content objects. However, metadata maybe the only portion of the file that needs to be accessed. Therefore, the blocks corresponding to the metadata for the file can be retained in the SSD tier while the other components of the file can be pushed out to the HDD and cloud storage tiers. The automated and granular management of block data done by the StorSimple solution is a simplified approach compared to traditional storage approaches to data management.

The following diagram shows the flow of data (associated with a file) for traditional storage environments, having on premise primary, secondary and archival storage. Moving data to storage resources that provide adequate performance on the lowest cost storage media is a manual, policy based, activity by an administrator. The diagram shows the number of policies required for moving data between tiers, and this would be minimally acceptable. For file share data, the large quantities of data may require a very granular policy process that addresses files on an individual bases to be effective however this can become unmanageable.

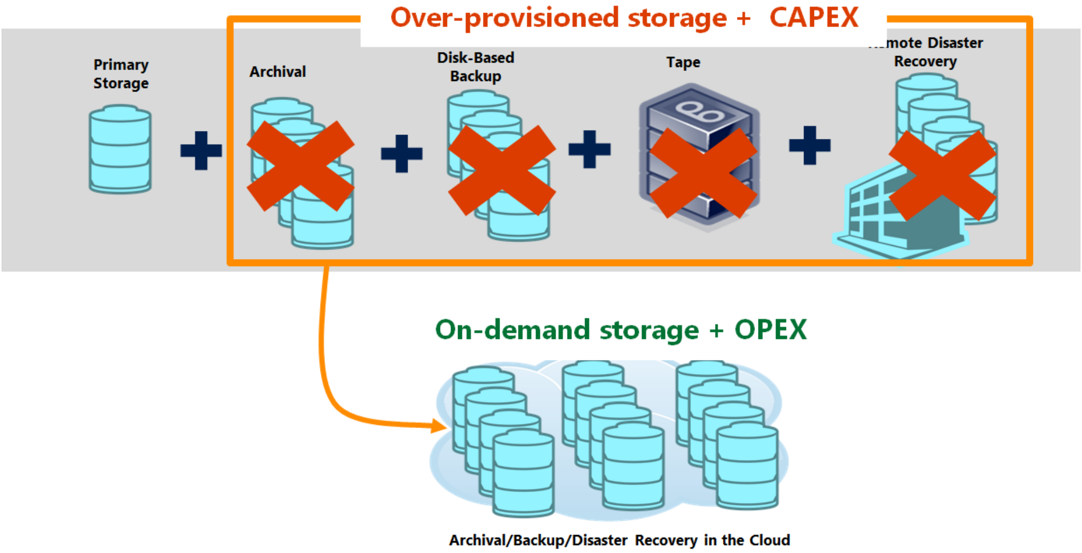
**Figure 4) Tiering storage across traditional storage infrastructures**



## Minimize over-provisioning of storage for large file share environments

With traditional storage environments, storage administrators have to estimate expected growth when they purchase their storage infrastructure. Running out of storage is not an option which means that storage administrators regularly buy more storage than they actually need. The excess storage typically sits idle for some period of time, consuming data center resources. This is especially problematic in file share environments where the amount of data is growing at unexpected rates. Rather than underestimate the growth rate of file share data, storage administrators will increase the over-provisioning percentage. The StorSimple solution addresses this from another perspective, where only a subset of data is stored on premise and all other data is scaled-over to cloud storage. The cloud storage is provided as the demand for storage arises. The StorSimple solution is completely integrated with Windows Azure during the setup of the on premise solution without any addition API integration. Storage administrators do not incur any additional work when using the cloud, all data is moved to the cloud based on the frequency that the data is accessed therefore over time all aged data that is no longer needed will be tiered to the cloud and the on premise StorSimple solution is maintain with a fixed capacity of data (referred to as the working set or "hot" data). This allows for storage in the cloud to be paid for when it is actually used and storage becomes an operation expenditure (OPEX) rather than capital expenditure (CAPEX). The below figure shows the amount of infrastructure that can be consolidated into cloud resources and transitioned to an operation expenditure.

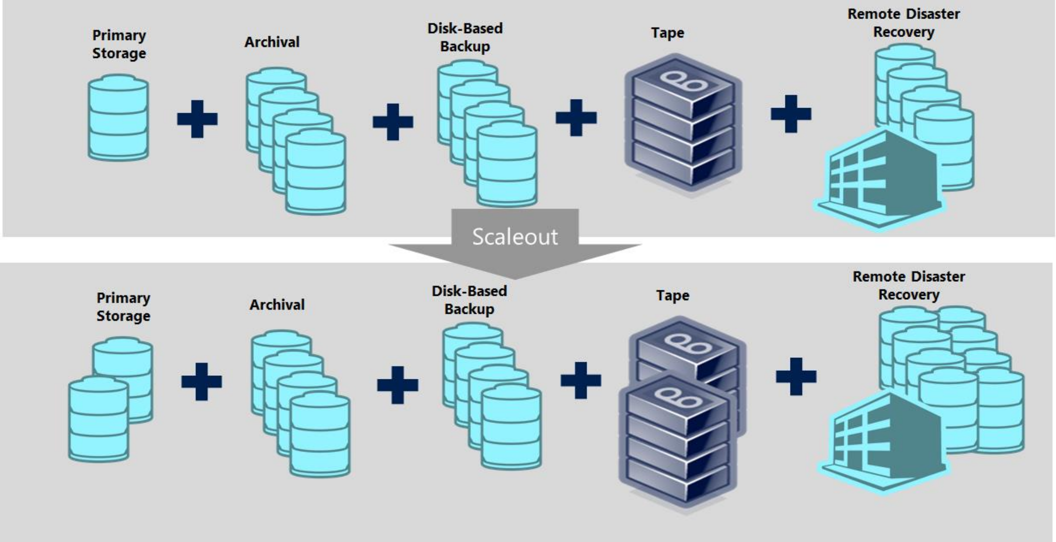
**Figure 5) the transition of storage infrastructure to the cloud for on-demand storage**


Over-provisioned storage + CAPEX
Primary Storage
Archival
Disk-Based Backup
Tape
...mote Disaster Recovery
On-demand storage + OPEX
Archival/Backup/Disaster Recovery in the Cloud

## Consolidated primary, secondary, backup and archival into a single solution for file share environments

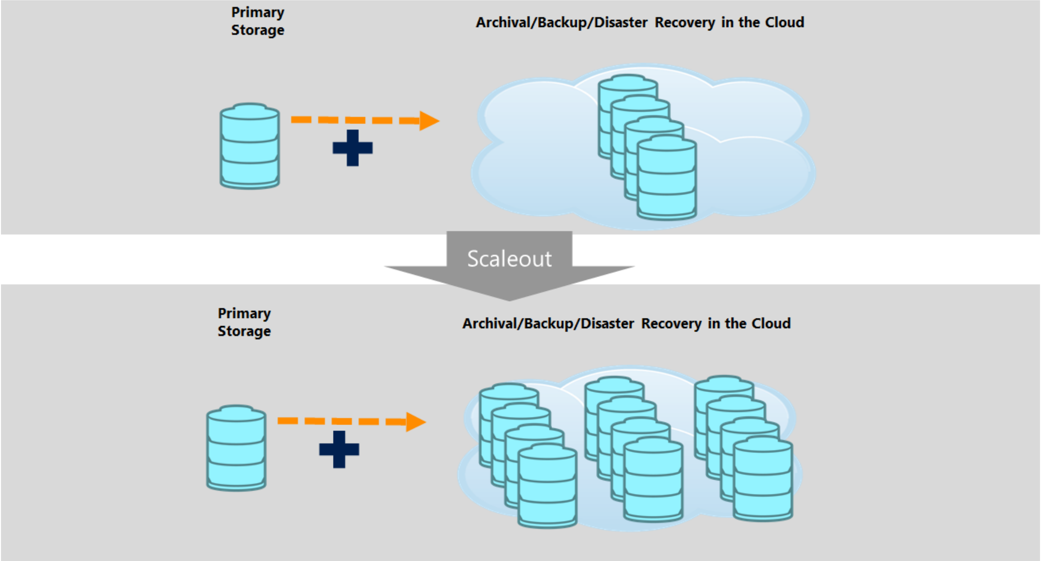
The growing amount of unstructured data in file share environments means that there is also a growing set up backup data corresponding to the primary set of data. For each file that is stored, there is at least one backup copy, or snapshot, or redundant copy of the data stored.  In more conservative storage deployments, where data availability is inherent to the business there will be many copies of the data distributed across multiple data centers. What this means is a sprawling set of storage infrastructure to support the availability of data files when data is lost or disaster strikes. StorSimple approaches this data sprawl, common with traditional storage approaches, differently by using the cloud as the storage resource for backup and recovery needs. A subset of snapshots are stored local for quick restores however the majority of backup data resides in the cloud, consolidated into a single solution for backup, recovery, archival and disaster recovery. This reduces the amount on premise storage to only primary storage. The below diagram shows a typical traditional storage infrastructure for proper data protection.

Figure 6) Traditional storage data center sprawl and growth across all data protection solutions



The following diagram show how the StorSimple solution eliminates data center sprawl and consolidates storage growth related to data protection solutions into the cloud while still meeting a range of backup, recovery, archival and disaster recovery needs.

Figure 7) StorSimple solution for consolidated data protection, reducing on premise data center sprawl


Primary Storage
Archival/Backup/Disaster Recovery in the Cloud
Scaleout
Primary Storage
Archival/Backup/Disaster Recovery in the Cloud

# Deployment and data migration

When deploying the StorSimple solution, data may need to be migrated from another storage system. This can be done using a variety of methods however this section will focus on disk mirroring and file copy utilities. Disk mirroring uses the synchronization of data between two volumes using RAID on the host side server. File copy utilities copies data using a file copy utility, such as robocopy, from the existing storage system to the StorSimple solution.

When initially deploying and creating volumes on the StorSimple solution, it is important to create volumes that have a capacity equal to or greater than the existing volume (on the storage system the data will be migrated from).

## Migrating from a NAS versus SAN

When migrating data from a network attached storage (NAS) device, the data is copied from the existing storage device to the volume mounted on the windows server and then routing the user or application to the StorSimple solution. This method of migrating data requires a file copy utility to migrate the data and planning in regards to a specific cutover period when the client or application is moved from the existing NAS storage to the Windows server, serving the data from the volumes on the StorSimple solution.

When migrating data from a storage area network (SAN) device, the SAN is served by a host, often a Windows server, on which the SAN volume is mounted and can be mirrored with the volume on the StorSimple solution. The disk mirroring option is available on the server hosting the SAN volumes. The file copy utilities can also be used to migrate data from the SAN volumes mounted on one Windows server to another Windows server mounting the volumes on the StorSimple solution.

## Disk mirroring

Disk mirroring is built-in to Windows server and provides disk mirroring (RAID1) for the volume. Disk mirroring allows existing data to be synchronized between the two volumes and all subsequent writes to be performed on both volumes. The mirror can be broken when the StorSimple volume has a consistent view with the original volume. This process allows the administrator to break the mirror and use only the volume on the StorSimple solution. The following conditions must be true to use disk mirroring for migration:

- Existing storage volume is mounted on the Windows server
- The disk is converted from a basic disk to a dynamic disk (prior to setting up the mirror)
- The disk/volume must be formatted as a NTFS volume

When the resynchronization (resync) process completes, the disks representing both volumes, should contain synchronized data. At this point the mirror, of the original storage volume, should be removed. It is recommended to take a snapshot of the volume prior to removing the original volume or breaking the mirror. Take the old SAN volume offline and break the mirror. Verify that any operations continue as expected. Taking the snapshot provides a backup for before and after the migration.

Mirroring volumes is only recommended for the purpose of migrating data from another volume. Once data is migrated the mirror should be broken. Disk mirroring is not intended for long term usage.

There are positive and negative trade-offs with the disk mirroring method. On the positive side, the disks remain online while the mirroring is in progress and no changes are required during or after the mirroring process to the volume, minimizing impact to the end-user.  The mirroring process is very simple to setup, execute and complete. On the negative side, disk mirroring has restriction regarding NAS devices, mirrors cannot be created on a NAS device or any share that is mounted using a drive letter. The mirroring process mirrors all blocks on the existing disk including empty disks, and a basic disk must be converted into a dynamic disk.

## File copy utilities

File copy utilities can be used to copy data from an existing storage system to the StorSimple solution. The volumes of each storage system would need the corresponding volumes mounted on a Windows server. Windows Server has a built-in file copy utility called robocopy that is recommended when migrating data to the volumes on the StorSimple solution. The migration process consists of multiple phases: initial data copy, incremental sync of data, final data copy & cutover (no more writing to source volume), mapping of shares and applications to the StorSimple solution.

It is recommended to consider the following variables prior to migrating the data

- Make sure the final cutover time is determined and aligns with a maintenance window or a period of low activity. During the cutover, existing data cannot be modified and all file locks need to be released
- Know how much data is going to be migrated and the expected time the migration will take with the robocopy utility
- Backup the configuration information for accessing existing storage, such as the Share names and Share permissions

For file shares, migrating data from existing storage solution to the StorSimple solution will automatically tier the data so that it resides on the storage tier that is appropriate for the service level regardless of the unstructured nature. It is recommended to copy the oldest data first to the StorSimple solution volume. This will help to properly tier the data, moving the oldest data to the SAS and cloud tiers as newer data is migrated later in the process. The most active data will ultimately be tiered on the SSD layer.

During the migration process, it is recommended to take a snapshot on the original NAS or SAN volume, mount the snapshot and robocopy from the mounted snapshot as a point-in-time source. This will prevent issues open files during the migration process. This is an alternative to using the retry and wait option as the files may be open during the copy operation which will result in an error. The copy operation will continue by skipping the open file or the file which encounters an error.  Any data in an open file will be copied during the next incremental phase of transfers.  It is also recommended to log the operation for better performance. Robocopy should be run in administrator mode to ensure it has enough privileges to copy data and all the attributes.

During the incremental sync phase, the incremental sync should be run on a regular basis to reduce the amount of data that will need to be copied during the final data copy and cutover portion of the migration.

During the final copy and cutover, access to the source volume should be restricted to prevent any changes to the data while the final data copy is completed. This should be done during a planned downtime period, disable user access to the source volume and close open handles to the files. Once the final copy and cutover is complete user access can resume on the destination volume residing on the StorSimple solution.

There are considerations for using a file copy utility instead of other migration methods.

A file copy utility is beneficial for the following reasons:

- Enables data copy from a NAS device
- Allows the consolidation of data from multiple location (can copy data from multiple volumes on to a single volume on the StorSimple solution)
- Allows copying old data first to optimize the tiering of data during migration
- DFS-R can be used for continuous synchronization of the data

A file copy utility has limitations for the following reasons:

- Requires extensive planning for the entire migration process
- Data is copied over the network if using a different server for running the file copy application
- Open files cannot be copied using robocopy, robocopy will not copy the file until it has been closed pro-longing the incremental copy portion of the migration process.
- Shares may need to be mapped again

# Primary storage for file shares

## Choosing a StorSimple solution

StorSimple offers various models, allowing the user to choose the solution that align with their business needs. All the StorSimple storage models are a SAN storage solution and support file share deployments. Each StorSimple solution runs the same software and supports all features. The key difference is around storage capacity in the three tiers (SSD, HDD, and cloud). The following table outlines the specification for the StorSimple product line.

**Table 1) StorSimple solution models and specifications**

| | 5020 | 7020 | 5220 | 7220 |
|---|---|---|---|---|
| **Raw Local Capacity** | 2TB | 4TB | 10TB | 20Tb |
| **Local SSD Capacity** | 400GB | 600GB | 1.2TB | 2TB |
| **Usable Local Capacity** | 4-10TB | 8-20TB | 20-50TB | 40-100TB |
| **Maximum Capacity (TB)** | 100TB | 200TB | 300TB | 500TB |

Note: Capacities in table 1 are calculate where TB=10^12.

When choosing which StorSimple solution there are several things to consider: the total capacity of the existing data set (on the current storage system) and the performance required to accommodate the number of expected users. The usable local capacity depends on the amount of data that can be deduplicated. For file shares, users can generally expect a deduplication rate of 2 times. For example, a

StorSimple 5020 solution deployed in a file share environment would have a local raw capacity of 2TB yielding a usable capacity (the amount of data that can actually be stored) of 4TB.

When calculating the required capacity, this is the working set of data. Any growth or data that is to be archived or used for backup purposes is not included in this calculation. That data would be scaled over to the cloud as the data set grows. The cloud storage is available on-demand as the data set grows therefore it is not necessary to factor (or guess) at expected growth for the file share deployment.

When calculating the performance requirement, his is a function of the number of users to be supported and the current service levels (MB/s) being supported on the current system. This should be the baseline for deciding which StorSimple solution. A local Technical Service Professional (TSP) can assist with proper sizing and guidance when selecting a StorSimple solution.

The StorSimple solution approach differs somewhat from how traditional storage solutions are sized. Traditional storage solutions factor in a number of variables related to data protection therefore, depending on the level of protection a business requires, the product or solution they choose would vary. With the StorSimple solution all the models deliver recovery point objectives (RPO) and recovery time objectives (RTOs) that meet the needs of most businesses. The StorSimple data protection solutions leverages the cloud, providing flexibility and long term retention. This is discussed in the following data protection sections of this document.

# Scalability

One of the things that continues to be problematic for file share deployments is the growth of primary storage. How to plan and scale the storage infrastructure is an ongoing question for IT managers. For example, if there was 25TB of data last year and 60TB of data this year and a projected 100TB (plus) for next year, how do I properly scale out while maintain budgets and costs? Additionally, planning for the space and resources to accommodate this amount of storage, and keep all the data optimized so that it can be accessed when needed is a daunting task in traditional scale out architectures where all storage is on premises. StorSimple approaches scalability of primary storage differently and is based on the concept that the amount of true primary storage remains relatively fixed.  For example, an employee has 5 projects that they are working on, two of the projects are new and fairly dynamic with continually changing content, while the other three projects are in the later stages and most content is in a complete state and the files are reviewed occasionally. The general, idea is that an employee would only ever have two hot projects at any one time so all other projects (and related content) can be tiered to the cloud, available when and if needed.

So what this means is that the on premises StorSimple solution will always hold the active data however as the data set grows and capacity needs to scale, data is pushed to the cloud and the cloud storage is where the solution scales. All cloud data is continuously available to the end user so it is not like archiving to an unusable third-party remote site. The cloud is just an extension of the on premise storage solution. If data that has been moved to the cloud is accessed, it can be pulled back to the local on premise SSD tier.

Each StorSimple solution has a maximum capacity, shown in table 1. The maximum total capacity for a solution should not be exceeded to maintain expected performance.

# Primary storage deployment considerations

There are a couple of things that can be done to optimize the StorSimple solution during the initial deployment of the system. When setting up the StorSimple volumes, there is an option to select a volume type that is optimized for file share deployments. When creating a new volume, for a file share, select the File Server Volume type. When a volume type is selected StorSimple automatically adjusts the internal priority of the volume data to optimize performance and utilization of local storage across all volumes and all data. Regardless of the volume type, data tiering decisions are still made based on the oldest data being pushed to the lower tier of storage first.

On the Windows Server, StorSimple volumes should be accessible using the Microsoft iSCSI initiator, by supplying one of the interface IP addresses as the target portal or discovery IP address. All StorSimple volumes are auto discovered once they are connected through iSCSI to the StorSimple solution.  Once connected with iSCSI, open the Disk Management snap-in to initialize your volumes. Next, configure them to use the NTFS file system with a 64KB allocation unit size (AUS), supply a drive letter, and perform a quick format.  For more information on installing and Configuring Microsoft iSCSI Initiator please click here.

Following best practices when migrating data is also a factor in optimizing the system during the initial deployment. Follow the best practices stated in the deployment and data migration section of this document. The main concept for migration is copy all of the older file share data first and then the recent data, this will help in immediately tiering the oldest, less used data to the cloud.

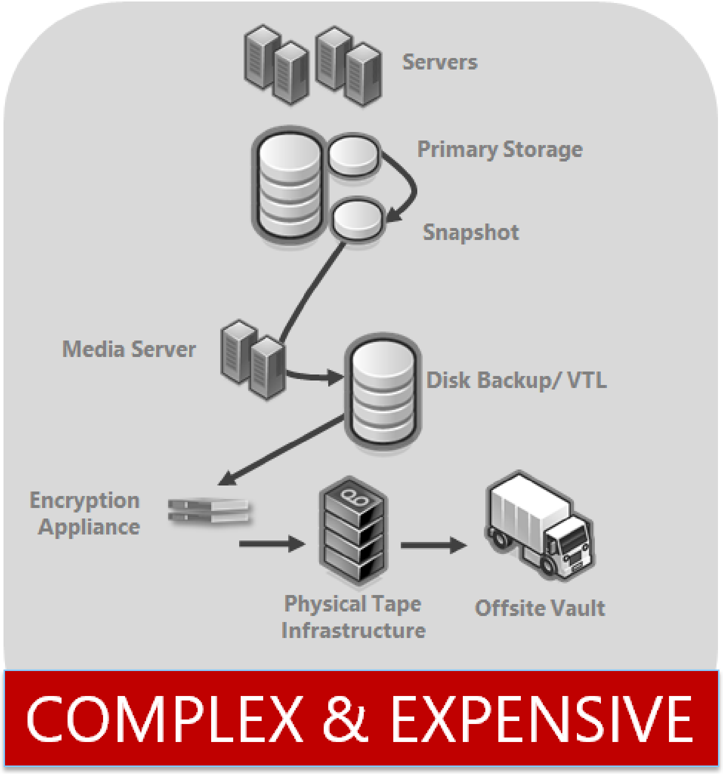
Migration should be done during periods where there is minimal activity on the systems. The time it takes to migrate data from the older storage system to the StorSimple solution is dependent on the amount of data and any shared resources. In general terms, during a period of migration, a user can expect to see write throughput between 20 to 100 MB/s depending on how full the solution is. For a solution that has data on the SSD tier only, expected performance would be around 100 MB/s (up to 200 MB/s for deployments with MPIO). For a solution that has SSD and HDD tiers that are relatively full and pushing data to the cloud, expected performance would be up to the 20 MB/s range, depending on Internet bandwidth available. Deployments configured with MPIO can expect to see higher write throughput. For more information about MPIO refer to the online knowledge base article, MPIO Setup Windows Server 2008R2.

# Data protection for file shares

Data protection is an important piece of any IT infrastructure. Ensuring business continuity in the case of simple data loss such as an accidentally deleted file or complete recovery in the face of disaster is essential. Data protection solutions, while necessary can be extremely costly, not just from equipment but also for the time and expertise required to plan and manage backup solutions that meet the RTO and RPO of the business.

Traditional storage approaches employ a strategy that requires multiple levels of data protection, additional licensing and sometimes a completely separate solution and infrastructure from the primary storage infrastructure. What this means is data centers are sprawling with multiple solutions, each having specific requirements and management needs. A typical data center might look something like this when data protection solutions are incorporated (for simple backup and disaster recovery).

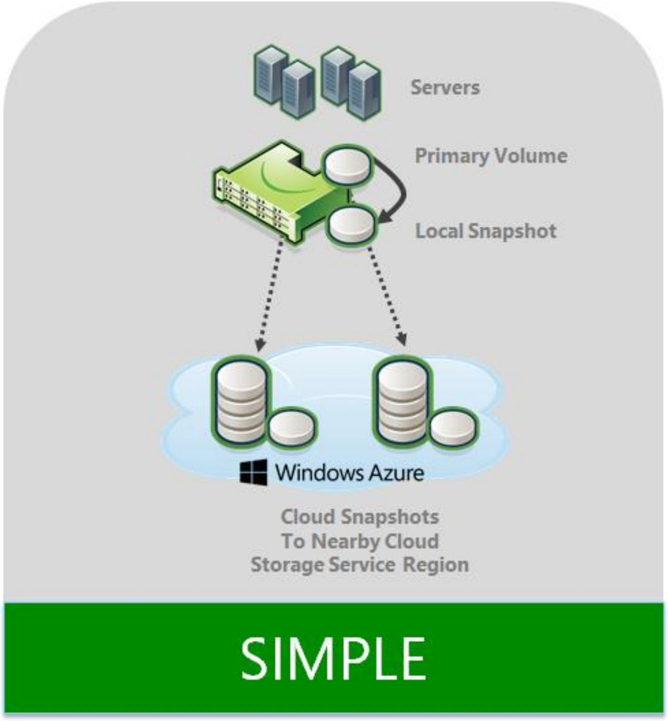
Figure 8) Traditional storage data protection solutions



When dealing with file share deployments, this problem grows exponentially with the amount of primary data being stored.  For each backup and recovery point, the amount of storage needed and the additional management to maintain that solution grows. Eventually, the traditional storage approach for data protection in file share environments is fragmented at best with multiple solution accomplishing different tasks.

The StorSimple solution takes a different approach to data protection by leveraging the cloud as a single solution for the majority of a businesses and backup and recovery needs. The consolidated solution provides some amount of backup and recovery on the local appliance with local snapshots. The rest is consolidated in the cloud: archival of data for long term retention (replacing tape), cloud snapshots and thin restores replacing disk backup, VTL and disaster recovery systems. All data is automatically encrypted when it is transferred to the cloud without any additional integration or solution. The above data center is consolidated into a single solution that is completely integrated without addition management. The figure below shows the consolidated solution for data protection with the StorSimple solution.

**Figure 9) StorSimple data protection solution**



For more information on the StorSimple Data Protection solution refer to the Data Protection with StorSimple technical document.

# Backup policy guidance for file shares

The file set size will affect the file share backup time and bandwidth utilization. It is very important that organizations pay attention to the file set size and adjust the bandwidth accordingly.

Backup methods and policies are highly influenced by RPO & RTO. The recovery point objective (RPO) and the recovery time objective (RTO) are two very specific parameters that are closely associated with recovery. This is often associated with your maximum allowable or maximum tolerable outage.

The RTO is really used to dictate your use of replication or backup to tape or disk. If the RTO is zero then an enterprise may decide to have a completely redundant infrastructure with replicated data offsite. If the RTO is 48 hours or 72 hours then maybe tape backup is OK for that specific application. Again, that dictates the kind of data protection solution you want in place.

So both, RTO and RPO, really influence the kind of redundancy or backup infrastructure an enterprise should put together. StorSimple solution with cloud integrated storage allows enterprises to meet a range of RPO and RTOs given the needs of the business through on-site local snapshots and cloud snapshots that provide high performance restores or flexible disaster recovery solutions.
Backup policies for Local Snapshots and Cloud Snapshots are created separately for each volume group (a group of related volumes with the same backup schedule). The backup policies define the following variables:

- **Backup Type** – the type of backup to perform (snapshot, Cloud Snapshot, Cloud Clone)
- **Schedule** – at what interval and what frequency the backup is executed

- **Retention** – how much backup copies to retain
- **State** – whether the policy is enabled or disabled

When creating backup policies, it is recommended, at a minimum, that a local and cloud backup policy be created for each volume group, one for each backup type offered by StorSimple.

Each volume can have up to 256 backup copies, across all policies that reference that volume in the associated volume groups. Alternatively, a policy can be set to retain "all" backup copies, but the policy will become disabled once the limit on any volume in the associated volume group is reached.

In order to create the most appropriate backup policies for your volumes, it is important to understand the retention requirements for each type of backup. As an example, when transitioning from a daily snapshot to a weekly Cloud Snapshot, it is recommended that at least two weeks' worth of daily snapshots be retained.

Minimally, it is recommended to take a cloud snapshot daily and retain it for one day for all volumes, regardless of the data protection or disaster retention policy. A cloud snapshot can help to maintain consistent performance of the StorSimple solution, especially during times when the StorSimple solution is hit with an influx of data. When a cloud snapshot is taken, all the blocks that makeup the cloud snapshots are essentially copied to cloud (however they also stay in the SSD or HDD tier, if that is where the block currently resides). Subsequently, if a large number of writes come into the StorSimple solution and data begins to be pushed from the SSD tier to the HDD tier to the cloud tier there will not be a bottleneck pushing data to the cloud since a large percentage of the blocks already exist in cloud storage, only changed blocks since the cloud snapshot was taken would need to be pushed to the cloud.

For more extensive data protection and disaster recovery policies, the following outlines an example setup for various backup needs.

**Example Policy**

- **Short-term backup**
  - A local snapshot policy
  - Scheduled to run every 30 minutes
  - Retain for 8 hours
  - Creates a total of 16 local snapshots
- **Medium-term backup**
  - A Cloud Snapshot policy
  - Scheduled to daily
  - Retain for 30 days
  - Creates a total of 30 cloud snapshots
- **Long-term backup**
  - Scheduled to run monthly
  - Retention of thirty-six months
  - Creates a total of 36 cloud snapshots copies (1 per month for 36 months)

# Best Practices

It is highly recommended that all the best practices be closely adhered to. These best practices will ensure realization of full potential of the StorSimple solution.

Prior to the deployment of the StorSimple solution, it is recommended to review the following documentation and complete any pre-work in preparation for the installation and deployment:

- Review and complete the Pre-Installation Checklist. This will help you identify information that will be needed during the initial configuration of your StorSimple solution.
- Review the Hardware Replacement Guide and follow the steps in the Rack Mount Guide to safely mount the appliance in the rack.
- Review the Monitoring Indicators Guide and familiarize yourself with the LEDs and other components that notify you if there are any hardware issues.
- Review the online documentation for powering and cabling instructions of StorSimple

Refer to the StorSimple Deployment Setup Guide for step-by-step instruction for setting up the StorSimple solution.

## Configuration Best Practices

### Optimal availability

Ensure that all high availability components are in good health at all times.

**Recommendation:** In the case of a failure that may take a redundant component out of service, replace the failed component as soon as possible to return the system to a highly available configuration (in all areas).

### Configuring Volumes

When using iSCSI volumes, special care needs to be taken to ensure that a volume is not concurrently accessed from multiple host servers. Accessing a volume from multiple host servers, except in case of a cluster of servers setup in a failover configuration, can corrupt it and render it completely unusable. StorSimple addresses this problem by the use of access control records (ACRs). In order to ensure that only the intended server have access to an appropriate set of volumes, the following best practice recommendations should be followed when configuring ACRs.

**Recommendations:**

1. When assigning more than one ACR to a volume, care should be taken that the combinations of the ACRs do not expose the volume in a way such that it can be concurrently accessed by more than one non-clustered host. The StorSimple appliance is designed to display a pop-up warning message if multiple ACRs expose the volume to more than one host. However the StorSimple solution doesn’t actively block I/Os that originate from multiple hosts.

## Selecting Optimal Volume Size

StorSimple supports a maximum volume size of 100TB. Every StorSimple volume, regardless of its size, is thinly provisioned implying that a volume can be configured to be of any size but only space corresponding to the actual amount of data that is written is consumed on the appliance (local data) and the cloud (cloud data). For example: a 10TB volume with 500GB of data written to it will only consume 500GB of space.

While the space is consumed on an as-needed basis, the actual configured size of the volume does result in some overheads with regards to the amount of time it takes to create or restore from cloud-based backups (Cloud Snapshots). This is because during a backup or restore, the entire volume space needs to be scanned to determine which blocks contain data and therefore the larger the configured size of the volume the more the overhead.

**When choosing volume size, the following best practices are recommended:**

- On Windows versions prior to Windows2012, there is no support for space reclamation commands on thin-provisioned storage, deleted files will continue to occupy block storage space. You should create volumes that are sized to your current needs plus certain margin. You can always expand the volumes later on. If you size the volumes too large, you cannot shrink them later on.
- It is recommended that volumes be no larger than 64TB for optimal RTO on large capacity volumes.
    - For a maximum RTO of 3-4 hours (for thin restore), we recommend 16TB volumes.
    - During disaster recovery, multiple volumes can be restored in parallel (up to 16). RTO is approx. 10-15min/TB of allocated capacity.
- Volumes cannot span Cloud storage accounts. It is important to size total volumes in a Cloud storage account to no more than 50% of the storage account's maximum capacity. For example, in Azure, a storage account has a limit of 200TB. As a best practice, no more than 100TB of the cloud storage, for that account, should be provisioned in volumes.

## Configure optimal NTFS allocation unit size

Once a volume has been created on the StorSimple solution most Windows applications require a file system to be created on it before it can be used. When formatting a volume, Windows provides the option to choose an AUS. The AUS represents the smallest amount of disk space that is allocated to store data. By default Windows uses a 4KB AUS for volumes. While this is reasonable for smaller files, Microsoft recommends selecting a 64KB AUS as a best practice to achieve the maximum deduplication results. Given the growth in data and file sizes, the 64KB AUS recommendation stands for file shares as well. It is recommended to format all StorSimple volumes with a 64KB AUS.

## Configuring Index Search and Virus Scan Applications

When applications such as index search or virus scan are used to scan data that is stored on a StorSimple volume, special care needs to be taken to ensure that the data that is marked 'cold' and tiered out to the cloud is not accessed and inadvertently made 'hot'.

**Recommendations:**

1. Configure the index search application (example: file indexer, SharePoint search crawler, etc.) or the virus scan application to always perform incremental operations. This implies that only the new data that is most likely still on the local tiers is operated on, and the cold data is not accessed by way of a full scan

2. Disable any automatically configured full scan operations

3. Ensure that the correct search filters and settings configured so that only the intended types of files get scanned. For example, image files (JPEG, GIF, TIFF) and engineering drawings should not be scanned during the incremental nor full index rebuild

# Operational Best Practices

The following best practices are for the operation of StorSimple storage systems.

## Network Connectivity Requirements

StorSimple offers a cloud-integrated storage solution that requires an active and working connection to the internet at all times. This connection is used for the various operations such as off-ramping non-working set 'cold' data to the cloud, taking Cloud Snapshots, synchronizing the time, etc. For the solution to perform optimally it is recommended that the following networking best practices be adhered to.

**Recommendations:**

1. The minimum internet network bandwidth for the StorSimple solution should be at least 20 Mbps at all times. This 20 Mbps bandwidth should be dedicated and not shared with any other applications. Assuming a data change rate of 1%, for the data size the bandwidth requirements can be the following.

   10TB – 20Mbps

   50TB – 50Mbps

   100TB – 100Mbps

2. Configure StorSimple QoS (Quality of Service) templates to enable variable throttling of the network throughput by the appliance at different times of the day. QoS templates can be used very effectively in conjunction with backup schedules to leverage additional network bandwidth for cloud operations during off-peak hours

3. The actual bandwidth required for a particular deployment should be calculated based on the size of the deployment and the required RTO. Please refer to the StorSimple Backup and Restore Best Practices technical report for further information on this topic

4. Ensure network connectivity to the internet is available at all times. This recommendation also applies to use cases where a very small amount exists on the StorSimple solution and the data has not tiered out to the cloud. While the StorSimple solution can easily buffer temporary glitches in the network connectivity, prolonged outages will result in the storage becoming unavailable and iSCSI error messages being returned to the application. The StorSimple solution has a built-in alert mechanism that will send an alert message, if configured and displayed, an alert on the Web GUI

5. If the network infrastructure supports jumbo frames, they can be configured on the data ports that transmit the iSCSI traffic between the host servers and the StorSimple solution. The StorSimple ports are configured to automatically accept jumbo frames if the network and the iSCSI initiators are setup to support them. Jumbo frames should not be configured on the network port that is connected to the internet.

6. The iSCSI SAN traffic should be carried on a dedicated network from the corporate LAN, having both MPIO and jumbo frames enabled, if possible. However, there may be issues enabling jumbo frames on a corporate LAN and internet enabled network. Cloud access should be on a corporate LAN or other internet enabled network. Different interfaces should be used for iSCSI and cloud access.

# Data Security

## Encryption and Cloud Storage

StorSimple storage systems provide the highest security standards to secure the data. To keep the data even more secure, enterprises can deploy virtual private clouds between their data center and the cloud storage service provider, which will isolate and dedicate a part of the cloud storage provider's infrastructure solely for them. This prevents users from accessing cloud storage resources from outside the corporate network.

StorSimple storage systems implement industry proven security practices to ensure security of the data in motion and data at rest.

**Data at rest** is encrypted using AES-256-CBC prior to storing it on the cloud storage service. AES-256-CBC is the strongest commercially available encryption mechanism. The key used for the encryption is generated by the customer and there is no need to share the key with the cloud provider. The data will be unusable in the event that the cloud storage service provider is asked to turn over your data, or if they happen to lose a device or media that data is stored on.

**Data in-transit** between the on-site StorSimple solution and cloud storage service, data is encrypted using SSL. Storage data in transit between servers and the StorSimple solution can be isolated to a non-routable, unreachable, or physically separate iSCSI network, and volume access can be controlled by IQN, and CHAP authentication, including mutual authentication. StorSimple storage systems implement multiple security measures to secure the data.

Client security certificates are used to secure RPC communication between the StorSimple solution and host integration tools. Challenge-Handshake Authentication Protocol (CHAP) is used for mutual authentication with servers before establishing a communication link with them. Access Control Records are used to verify the initiator or server before allowing access to a volume. StorSimple storage systems verify the name of the access control record, CHAP user associated with it and the iSCSI Qualified Name (IQN) of the initiator that is trying to access the volume.

StorSimple Data Protection Console (DPC) is a Microsoft Management Console (MMC) snap-in that can be installed on the Windows host server. DPC provides multiple backup options for the StorSimple storage systems. DPC allows creating and exporting client certificates which are required for securing the RPC communications between the StorSimple storage system and the Windows host. Exported client certificates have the following attributes:

# User Access and Identity Management

A common concern for those considering cloud storage services is how to restrict access for the terminated employees or users who should not have access to the data on the StorSimple storage systems or on the cloud. StorSimple storage systems address the user and data security concerns in multiple ways for the enterprises.

StorSimple Management Console is designed to ensure with the principle of "simplicity first" to minimize operational burden and costs. The StorSimple Management Console is integrated directly into the solution itself, and does not require additional hardware or software.

All the StorSimple storage systems use role-based permissions to control user access to the management console web site. Administrators can configure role-based user access to the StorSimple storage system console. Based on the role, the logged in user will have access to specific areas on the management web site. StorSimple storage systems support both local users and Active Directory users to authenticate. Organizations have more control on managing user accounts from a centralized location using active directory. StorSimple uses Light Weight Directory Access Protocol (LDAP) for reading and editing directories over an IP network. While creating a user choose the 'Authentication' as 'LDAP' creates that user in active directory. If the user is specified as local, then provide a password for the specified user is required. Active Directory can use their AD credentials to login to StorSimple appliance. Both Windows 2008 and 2003 domain controllers are supported by StorSimple storage systems.

Additionally, Multi-Factor Authentication (MFA) may be used however it is not directly supported by the StorSimple solution. Addition setup is required to use MFA. MFA is typically implemented using secure tokens, which must be presented when logging into or accessing various parts of the management console for the cloud storage service. This, coupled with granular roles-based access control (RBAC) to determine what privilege level users have to the cloud storage management dashboard minimizes the threat posed by a malicious user. Threats from unauthorized or terminated employees can be prevented with proactive employment practices that revoke access tokens, making it virtually impossible for them to access or manipulate your data or services.

# Conclusion

The StorSimple solution is an innovative way to simplify the large quantities of unstructured data associated with file share deployments. The key characteristic of file share data that makes it most appropriate for StorSimple deployment is that data in file shares is active for a short period of its existence (hot data) and infrequently or rarely accessed (cold data) for the majority of its life cycle. As a result, bulk of the data in file shares is rarely used. By automatically tiering cold data to cloud storage, enterprises can eliminate storage infrastructure sprawl, reduce data management complexity due to large amounts of cold data and dramatically reduce their overall storage costs.