

# Cloud Sync White Paper

---

Based on DSM 6.1

# Table of Contents

Introduction	3
Product Features	4
Synchronization	5
Architecture	
File System Monitoring (Local Change Notification)	
Event/List Monitoring (Remote Change Notification)	
Consistency Check	
Data Encryption Algorithm	8
Usage Scenarios	9
Efficiently Access Public Cloud Data from Local NAS	
Collaboration Across Multiple Storage Spaces	
Perform Offsite Backup to Multiple Public Clouds	
Synology NAS as a Centralized Data Storage for Multiple Clouds	
Performance Benchmark	11
Testing Bed	
Conclusion	13

# Introduction

Nowadays, we are being offered a wide variety of storage options, such as direct attached storage, networked storage, and Internet-based public storage services. Since different types of storage services are designed for different purposes, we often use several of them to meet different needs. Cloud Sync is designed to facilitate efficient and real-time data exchange, and it bridges private network storage servers with public storage services through a single framework designed for data synchronization. With the support for various protocols (e.g., WebDAV) and storage systems (e.g., OpenStack Swift), Cloud Sync can effectively manage potential private and on-premise storage.

Cloud Sync offers advanced features such as path mapping and encryption. It is equipped with advanced control over sync directions and traffic controls, and is highly flexible and versatile in its task types. This paper outlines the technical designs of Cloud Sync and offers the details regarding its performance.

# Product Features

**Real-time synchronization:** Synology Cloud Sync automatizes the synchronization of data between a Synology NAS and cloud storage in real-time, instantly delivering local updates to the remote storage, while pulling down the remote changes as frequently as every ten seconds by default.

**One-way or two-way settings:** Sync direction can be customized by session (a subtask created within a cloud storage connection) to meet different usage scenarios.

**Customizable cloud storage polling period:** The polling interval can be configured from ten seconds to one day, enabling users to determine their own polling period according to data update frequency and system resource consumption.

**Multiple subtasks:** With each cloud connection, multiple pairs of folders can be synchronized; therefore, local and remote data do not need to share identical directory structure.

**One-to-one and one-to-many topology:** Cloud Sync allows one local folder to be synchronized to more than one cloud destination, making multiple offsite backups possible.

**Data archive option:** With one-way synchronization, an option is available to prevent file deletions on the destination. This feature facilitates what is broadly defined as incremental backup, in which only additions and modifications are updated to the cloud server, and no deletion is automatically initiated by Cloud Sync.

**Data encryption and compression:** Synchronized data can be encrypted on client-side via Cloud Sync before the data is uploaded, preventing unauthorized access of the data on the remote server. In addition, data compression can reduce outbound traffic and storage consumption.

The next section offers more details regarding the encryption design for Cloud Sync.

# Synchronization

Cloud Sync serves as a client device for cloud storage servers. Though Cloud Sync offers a shared framework and feature set, certain synchronization functions are limited on the server side. Cloud Sync delivers a unified experience when users connect Synology NAS to various cloud storage services coming with different designs and capabilities. This chapter draws on some of our achievements.

## Architecture

Cloud Sync consists of five major components:

- **Unified Sync Framework:** a carefully-designed framework adaptable to various storage interfaces and file systems.
- **Event/List Monitor:** monitors file changes on cloud storage.
- **File System Monitor:** monitors file changes on Synology NAS.
- **Cloud Sync Database:** retains local records of the synchronized files and metadata.
- **Web-based User Interface:** replaces complicated command lines with intuitive graphic user interface.

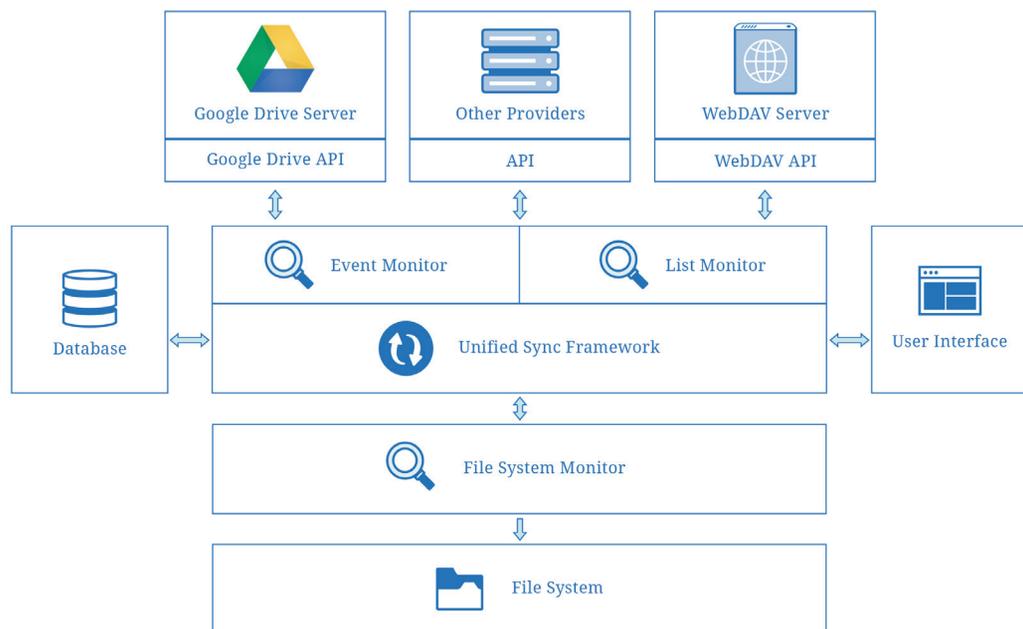


Figure 1: Cloud Sync architecture

Cloud Sync can synchronize data in real-time or according to the set schedule. Once a file is modified, Cloud Sync sends the file updated from the local destination to the remote destination and downloads the file updated from the remote destination to the local destination. Cloud Sync implements local file system and cloud change notifications to instantly deliver and download the updated file.

## File System Monitoring (Local Change Notification)

Cloud Sync leverages DSM's advanced Inotify API to monitor the files modified on your local Synology NAS, in which the changes made are always instantly updated to the remote cloud storage.

## Event/List Monitoring (Remote Change Notification)

Cloud Sync keeps track of the remote changes occurring on cloud storage through polling. The polling mechanism varies according to different storage services and protocols; however, it can, in general, be categorized into event-based providers and list-based providers.

### ◦ Event-based Providers

Change providers offer APIs that allow third-parties to fetch the delta between each polling event. According to the configured polling periods, Cloud Sync periodically demands the deltas and processes inbound synchronization accordingly.

The supported providers in this category include the following (last updated in April 2016):

- Amazon Drive
- Baidu Cloud
- Box
- Dropbox (including Dropbox for Business)
- Google Drive (including Dropbox for Business)<sup>1</sup>
- Microsoft OneDrive (including Office 365 and OneDrive for Business)<sup>2</sup>

### ◦ List-based Providers

For providers or protocols unable to offer delta information, Cloud Sync leverages LIST function to compare the directory structures in the local and remote sync folders. Cloud Sync uses the computing powers of NAS server to generate the difference occurred during each polling interval. When system activity is required during the scanning of local directory trees, the connections with list providers can prevent your Synology NAS from entering hibernation mode, in case the local file system monitored by Cloud Sync cannot be fully cached.

The providers supported in this category include the following (last updated in April 2016):

- Alibaba Cloud OSS
- Amazon S3
- B2 Backblaze
- Google Cloud Storage
- hicloud S3

1. Google Drive features an ID-based file system, which allows an account to contain multiple files sharing the same filename. To solve the problems regarding filename conflicts on NAS, Cloud Sync keeps a record of the ID of every synchronized file in the database, and appends a serial number at the end of each filename with the same name.
2. Tasks created through OneDrive's old API do not contain event-based provider capability.

- hubiC
- IBM SoftLayer
- Megafon MegaDisk
- OpenStack Swift-compatible storage
- Rackspace
- S3-compatible storage
- SFR NAS Backup
- WebDAV
- Yandex Disk

## Consistency Check

Cloud Sync keeps a local database to record the synchronization status of each file. The attributes recorded in the database are very useful when local and remote folders need to be compared for consistency check. The database is also very effective for verifying the changed events (e.g., modified files) and reducing API usage and unnecessary downloads/uploads.

Cloud Sync enables users to configure whether they want to enable advanced consistency check in the sync process of each session, during which an additional attribute - file hash - will be compared for consistency verification. A cloud storage system needs to provide hash information in order to execute advanced consistency check. The chart below presents hash availability (please refer to the [Help](#) article for more information on file hash).

Action	Advanced consistency check enabled				Advanced consistency check disabled			
	type	size	mtime	hash	type	size	mtime	hash
Compare the cloud and local file in the event of a relink	v	v		v	v	v		
Compare the cloud and local file attribute before downloads	v	v	v	v	v	v	v	
Determine whether a filename needs to be renamed due to filename conflicts (compare a downloaded file's attribute with a local file sharing the same filename)	v	v		v	v	v		

# Data Encryption Algorithm

## ◦ Overview

Storing data on remote and public locations may increase the risk for hackers or unauthorized parties to access your data. Cloud Sync lowers these threats by protecting your data backups with passwords.

In addition to HTTPS transmission encryption, a method is adopted to prevent data from being intercepted by hackers during data transfer. Cloud Sync also offers data encryption feature to ensure the data stored on a cloud is securely protected and cannot be accessed by other parties including the public cloud service itself.

With Cloud Sync, data can be decrypted even when the source NAS is out of service. The following offers the details regarding the encryption flow and decryption options.

## ◦ Encryption

When data encryption is enabled, each file from the sync task will be encrypted by AES through a randomly generated 256-bit key. The AES session key is subsequently encrypted using a 2048-bit RSA key and a user-defined primary key (password), with these two keys protecting the encrypted key.

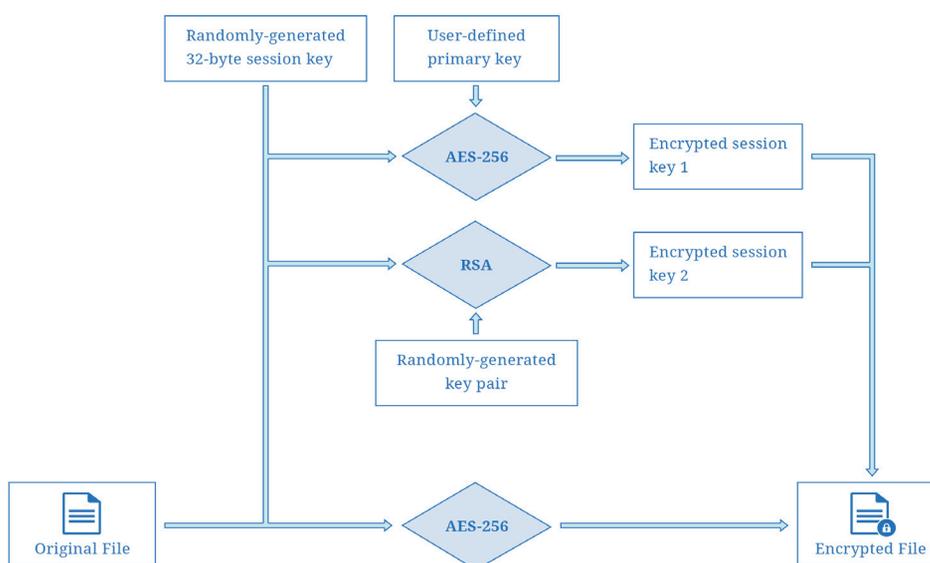


Figure 2: Encryption flow diagram

## ◦ Decryption

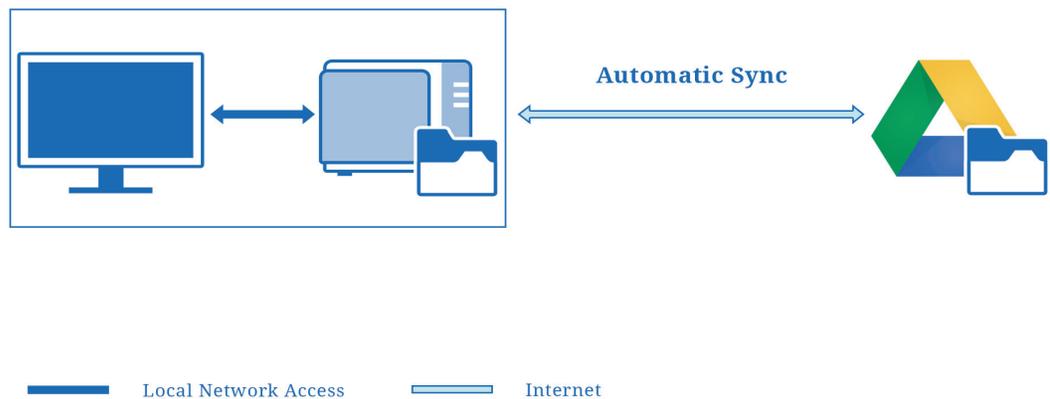
An encrypted file is automatically decrypted when files are downloaded from Cloud Sync; that is, when two NAS servers are linked to the same public cloud, files can be exchanged without sacrificing the confidentiality guaranteed by data encryption. The files encrypted and uploaded by a server can be read by another NAS sharing the same password.

If your NAS is stolen or unavailable and you need the data stored on a cloud storage, a decryption tool enables you to manually decrypt data on Windows or Ubuntu. With this decryption tool, users can decrypt a file or folder using either the primary key or the RSA private key generated during task creation.

# Usage Scenarios

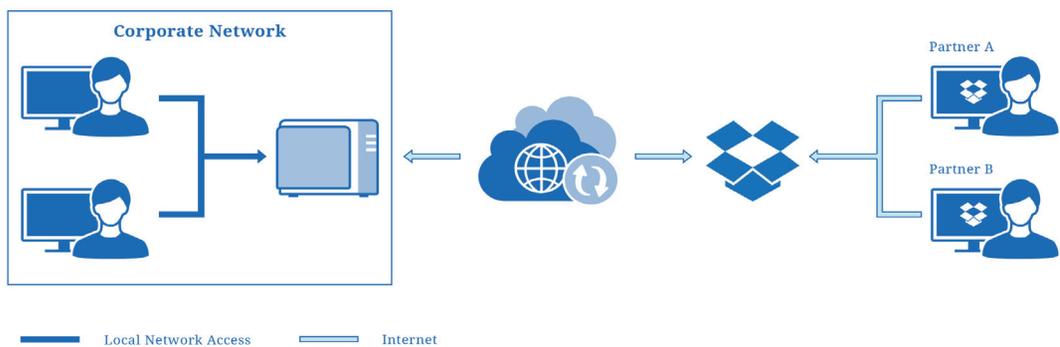
## Efficiently Access Public Cloud Data from Local NAS

Cloud Sync caches public cloud data on the NAS server and allows fast local access to data, thereby reducing the latency often encountered when accessing data on a public cloud. The files modified on your local Synology NAS will be instantly synced to your public clouds, saving you the hassle of needing to manually update the modified files to the public clouds.



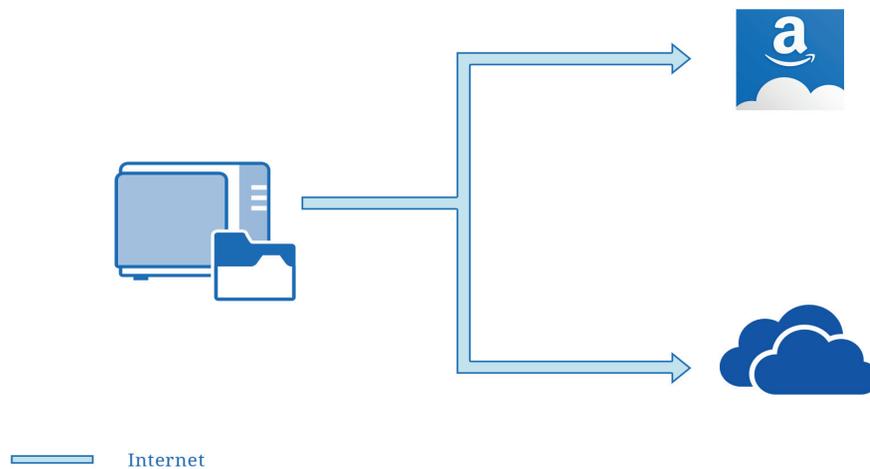
## Collaboration Across Multiple Storage Spaces

Cloud Sync facilitates collaboration with external parties on a public cloud when the external parties are not permitted to access the NAS storage server. In such cases, both the local employees and external correspondents can quickly and easily access data via LAN and can enable public cloud and Cloud Sync to automatically execute data exchange.



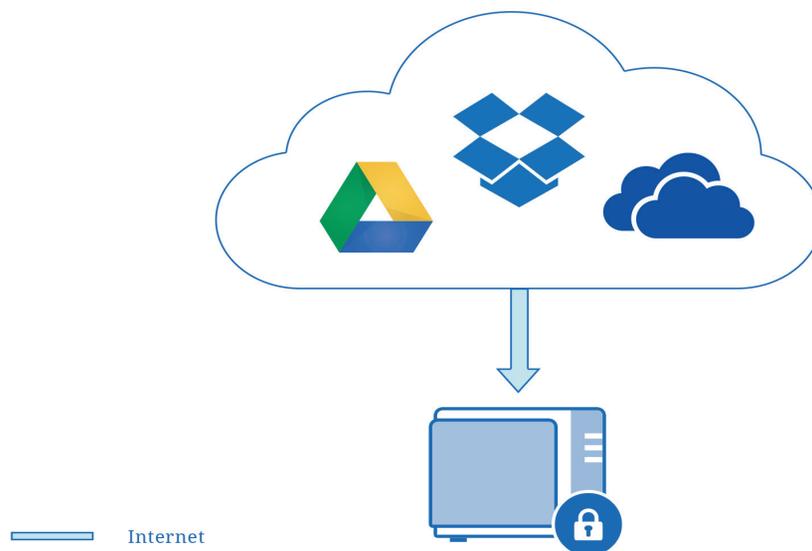
## Perform Offsite Backup to Multiple Public Clouds

Instead of setting up and maintaining a data center in a remote location, users can easily leverage public storage space as a cost-effective offsite backup destination by enabling the one-way synchronization (upload-only).



## Synology NAS as a Centralized Data Storage for Multiple Clouds

Backing up your public storage such as Google Drive and Dropbox to your private Synology NAS is always recommended. This can be easily configured with Cloud Sync's one-way synchronization (download-only). Another convenient and welcomed feature of Cloud Sync is that your online Google Doc can be saved to and backed up in Microsoft Office or JPEG formats.



# Performance Benchmark

## Testing Bed

In the performance test, we set up a 1Gb Ethernet environment and installed Cloud Sync on five different NAS servers. The specifications of these five models are listed below:

- RS3614xs+: Ext4 on RAID 5 with twelve 1TB hard drives
- DS3615xs: Ext4 on RAID 5 with twelve 1TB hard drives
- DS716+: Ext4 on RAID 1 with two 1TB hard drives
- DS416: Ext4 on RAID 5 with four 1TB hard drives
- DS216j: Ext4 on RAID 1 with two 1TB hard drives

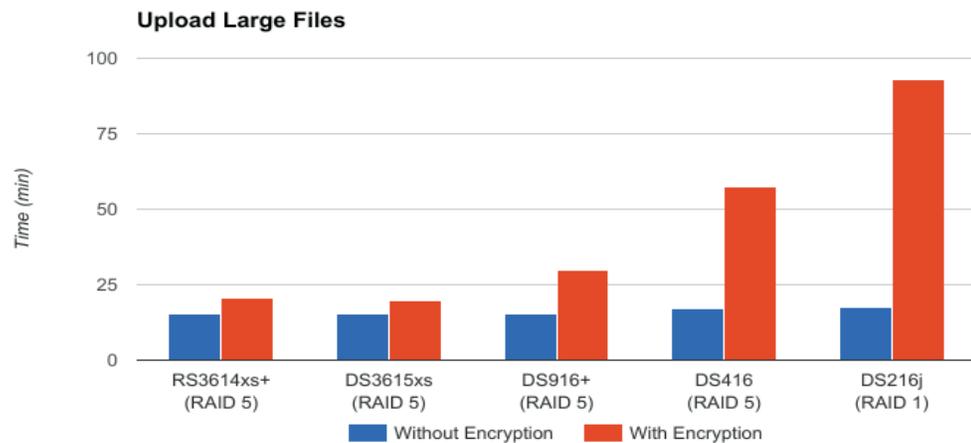
Each of the aforementioned NAS model was installed with DSM v8451 and Cloud Sync v0858, and Cloud Sync was connected to a WebDAV server with the following specifications:

- WebDAV Server: Windows IIS
- Hard disk: Intel 535 120G
- Memory: 16 GB

### Large Files

The chart below shows the evaluation results of the large file cases:

- Number of files: 20
- Size of each file: 5 GB
- Concurrent transfers: 3



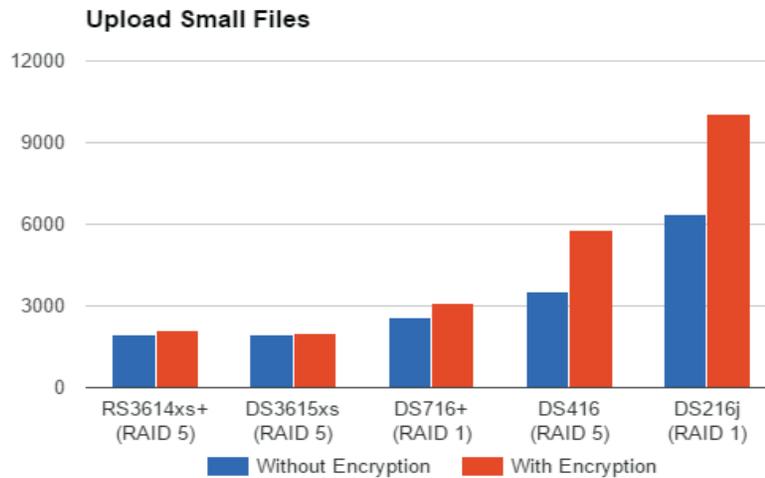
The blue bar shows the results of the cases without data encryption, in which transferring a large file without encryption required about 13.3 minutes. A comparison with the 1Gb Ethernet theoretical values showed that the network I/O was the bottleneck.

The red bar shows the results of the cases with the encryption function turned on, in which the computation power was required for data encryption. As a result, the network I/O was the bottleneck for models associated with strong computation power. However, for models with weaker CPU (e.g., DS216j and DS416), the bottleneck was the CPU.

### ◦ Small Files

The chart below shows the results of the small file cases:

- Number of files: 100,000
- Size of each file: 1 MB
- Concurrent transfers: 10

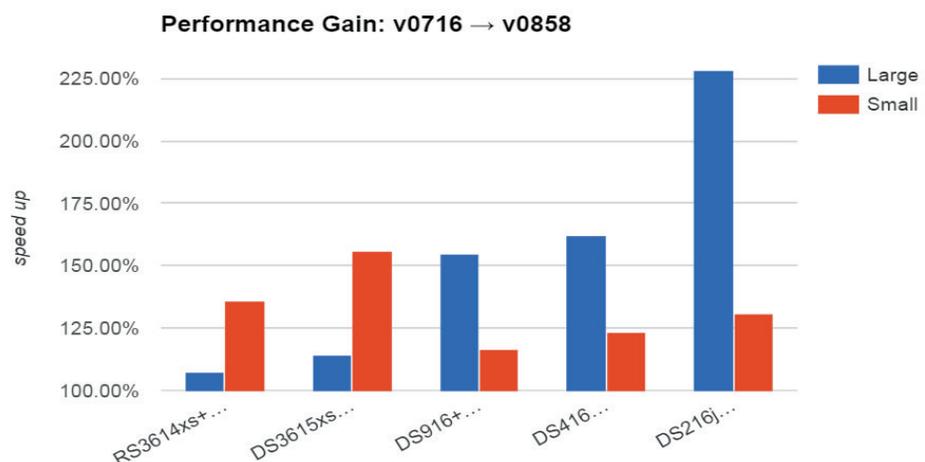


For models with powerful computation power (i.e., RS3614xs+ and DS3615xs), the Disk I/O was the bottleneck, leading to similar results for both the non-encryption and encryption tests.

Contrarily, for models with weaker computation power, relatively more time was required for data encryption. Therefore, models with weaker computation power consumed more time in encrypting files.

### ◦ Performance Enhancement

For Cloud Sync v0858, the I/O operations were further optimized and were equipped with a more responsive synchronization experience. The chart below shows the performance enhancement for plaintexts between v0716 and v0858, with the same dataset shown below. In short, all the tests experienced a 10% ~ 50% increase in their performance, and the speed for low-end models (e.g., DS216j) doubled.



## Conclusion

With Cloud Sync, you can seamlessly synchronize files between a Synology NAS and multiple cloud storages. The client-side encryption equipped by Cloud Sync surpasses the traditional file synchronization in availability and confidentiality, and also offers an excellent solution to overcome safety concerns associated with public cloud storage services. Moreover, with its advanced configurable options, Cloud Sync meets the demands of almost all types of usage scenarios.