CloudSync

John E. Anderson

July 15, 2019

# Introduction

**CloudSync** is a command line Python script which data between repositories. A *repository* is a place where files and folders are stored. This can be a local hard drive on a system, a folder on a remote file server, or a cloud-based server, such as Amazon S3. Data in a repository is organized into files and folders. Attributes of a file include:

* Name
* Date and time of last modification
* Size (in bytes)
* Data

The content of files is not interpreted and is treated as a series of binary bytes. Files may have 0 length. CloudSync synchronizes data between two repositories. One repository is called the *source* and the other the *target*. There are several types of synchronization between the source and target that are supported:

**Replicate** An exact copy of all files and folders are made. Any file or folder that does not exist on the target that exists on the source is copied to the target. Any file or folder that exist on the target but not the source is deleted from the target. If the file or folder exists on both the source and target, its contents are copied if the files are different (different modification time and/or size). If the files are the same (same timestamp and size), nothing is done.

**Update** Any file that exists on the source but not the target is copied. Any file on the target that is older than the same file on the source is overwritten with the source file data. Any file on the target that does not exist on the source is left alone.

**Synchronize** Contents of the source and target are compared. Any file that exists on one but not the other is copied. Any file of the same name that is newer replaces the older file, on either the source or target.

**Replicate** and **Update** can modify data on the target, with the source requiring read access. Synchronize may modify data on either the source or target, requiring both to have read/write access.

As an option, when a file in a repository is to be overwritten, and the repository supports it, the older version of the file may be preserved. This can be done on Amazon S3, for example. Also, older versions of a file may be purged in a repository, separate from any synchronize operation or as an option.