

Archive file

From Wikipedia, the free encyclopedia

An **archive file** is a file that is composed of one or more files along with metadata that can include source volume and medium information, file directory structure, error detection and recovery information, file comments, and usually employs some form of lossless compression. Archive files may also be encrypted in part or as a whole. Archive files are used to collect multiple data files together into a single file for easier portability and storage.

Computer archive files are created by file archiver software, optical disc authoring software, or disk image software that uses an archive format determined by that software. The file extension or file header of the archive file are indicators of the file format used.

Archive files are sometimes accompanied by separate parity archive (PAR) files that allow for additional error detection and recovery, particularly in recovery of missing package files in a multi-file archive.

Archives can have extensions like `.zip`, `.rar`, `.7z`, `.tar`, etc. (note: of these, `.tar` is the only one that does not include compression).

Archive files were originally used to store important files on alternate media (such as tapes) for preservation and recoverability. Archive files are now commonly used for transferring files and for software distribution and installation. A hard drive volume, RAID, and shadow copy/snapshot can also be considered archive files. The operating system would be the file archiver.

File archiver

From Wikipedia, the free encyclopedia

A **file archiver** is a computer program that combines a number of files together into one archive file, or a series of archive files, for easier transportation or storage. Many file archivers employ Archive formats that provide lossless data compression to reduce the size of the archive which is often useful for transferring a large number of individual files over a high latency network like the Internet.

The most basic archivers just take a list of files and concatenate their contents sequentially into the archive. In addition the archive must also contain some information about at least the names and lengths of the originals, so that proper reconstruction is possible. Most archivers also store metadata about a file that the operating system provides, such as timestamps, ownership and access control.

The process of making an archive file is called *archiving* or *packing*. Reconstructing the original files from the archive is termed *unarchiving*, *unpacking* or *extracting*.

Unix Archiver Tools

Unlike integrated archival and compression tools like PKZIP, Winzip, and WinRAR, the Unix tools **ar**, **tar**, **cpio** (for "archiver", "tape archiver" and "copy in/out" respectively) act as archivers but not compressors. Users of the Unix tools typically add compression by compressing the result of packing (and uncompressing before unpacking), most often using the **gzip** or **bzip2** programs. Modern tar programs can automatically invoke a (de)compression program, giving the appearance that tar itself handles compression and decompression.

This approach has two advantages:

- It follows the Unix toolbox concept that each program should accomplish a single, well-done task, as opposed to attempting to accomplish everything with one tool. As compression technology progresses, users may use different compression programs without having to modify or abandon their archiver.
- The archival program can potentially detect redundancy across two or more archived files, and need not store redundant data. An archiver compressing each archived file in isolation usually cannot exploit inter-file redundancies, because compression tends to hide redundancy.

In practice, compression with tar is usually done on the entire archive, and the second advantage listed above is not obtained. This type of compression - compressing the archive itself - has three disadvantages as compared with compressing within the archive:

- Extracting one file requires decompressing all the files that are before the file in the archive. This may take many minutes for a large archive.
- Modification is even more inconvenient than extraction - just changing a single character of one of the archived files will typically require that the entire archive be uncompressed, updated, and then recompressed.
- It's impossible or more difficult to take advantage of inter-file redundancy.

Archive format

From Wikipedia, the free encyclopedia

Jump to: navigation, search

An **archive format** is the file format of an archive file. The archive format is determined by the file archiver. Some archive formats are well-defined by their authors and have become conventions supported by multiple vendors and/or open-source communities.

Archive formats support features such as file concatenation, data compression, encryption, file spanning, parity/Cyclic redundancy check, checksum, self-extraction, self-installation, volume and directory structure information, package notes/description, and other meta-data.

Types of Archive Formats

- **Archiving only** formats only concatenate files.
- **Compression only** formats only compress files.
- **Multi-function** formats can concatenate, compress, encrypt, create error detection and recovery information, and repackage the archive into self-extracting/self-expanding files.
- **Software Packaging** formats are used to create software packages that may be self-installing files.
- **Disk Image** formats are used to create disk images or optical disk images of mass storage volumes.

Examples

Note: a comprehensive List of archive formats and Comparison of archive formats is available.

By Type

v · d · e	[hide]
Archive formats (comparison by type)	
Archiving only	ar · cpio · shar · tar · LBR
Compression only	bzip2 · gzip · xz · LZMA · SQ · compress
Archiving and compression	7z · ACE · AFA · ARC · Cabinet · cpt · DGCA · .dmg · GCA · kgb · LHA · LZX · RAR · qda · sit · SQX · zoo · ZIP · lzop · UDA
Software packaging and distribution	deb · pkg · RPM · RUNZ · MSI · JAR (WAR · RAR (Java) · EAR)
Document packaging and distribution	OEB Package Format · OEBPS Container Format · Open Packaging Conventions · PAQ

By Operating System

Unix operating systems utilize the tar file format, ar, and shar To concatenate files. These archive formats can then be compressed into gzip format.

On Windows platforms, the most widely-used archive format is ZIP; other formats are CAB, RAR (file format), and ACE. Windows Installer is a high-level archive format for distribution of software.

On Amiga computers the standard archive format is LHA.

on Apple Macintosh computers ZIP is now natively used in recent Mac OS X (10.3+), though StuffIt used to be the most common.

Linux often uses TAR, gz, and RPM package manager, a Package management system for distribution of software.

History

Origins

Ubiquitous amongst Unix and Unix-like operating systems is the tar file format ("**t**ape **a**rchive"). Originally intended for transferring files to and from tape, it is still used on disk-based storage to combine files before they are compressed.

Development

Historically, every major computer platform, every operating system, and every vendor had its own preferred archive format. Some formats became more commonly used because of licensing, feasibility, and popularity. Today the most common formats are supported by many platforms and vendors. New technologies continue to introduce new formats.