# LTFSArchiver

|  |  |  |  |
| --- | --- | --- | --- |
| Z:\backups\progetti\PrestoPrime\Dissemination\Rai.jpg | LTFSArchiver is distributed by RAI under the terms of the GNU Affero General Public License version 3.0 | http://10.2.7.57/rightsdraw2/Logo33.JPG  FP7-gen-RGB | Software developed in PrestoPRIME (2009-2012) – project funded by the European Commission in the Seventh Framework program.  FP7-ICT-2007-3 231161 |

# Introduction

***LTFSArchiver*** is a software which exposes a set of services for archiving media files over LTO/LTFS storage technology.

LTFS (Linear Tape File System) is an open format permitting the usage of LTO (Linear Tape Open) data tapes for persistent storage of generic files with almost normal file access modalities, and access latency lower than older data tape technologies. This feature is available for LTO tapes and drives since generation 5. This storage technology is a very cost effective solution for long term preservation of large files, as it is the case for audiovisual files at master quality level. The uncompressed capacity of the single LTO tape ranges from 1.5TB for LTO5 to 2.5TB for LTO6, and the use of tape libraries allows the creation of high capacity and scalable storage systems. The use of LTFS is a protection against interoperability problems usually found with proprietary solutions.

Despite these recognized advantages, the use of LTO equipments, drives and libraries, with LTFS is not safe and straightforward enough, unless specific software tools are also used to ensure the correctness of operations.

***LTFSArchiver*** provides the means for the fruition of LTO/LTFS storage technology in digital audiovisual preservation contexts. It can be used independently through its web interface or can be integrated in wider frameworks, such as the PrestoPRIME Preservation Platform (P4) or other client applications, through the defined interfaces.

# Architecture and system requirements

The services offered by ***LTFSArchiver*** are deployed on Apache/http service. The persistency of service data is provided by PostgreSQL database. The ***LTFSArchiver*** requires a Linux host, tested on CentOS and Ubuntu, connected to LTO libraries (under the control of ***LTFSArchiver***) and/or LTO drives manually operated for tape change. The Linux host must have installed the LTFS software driver (available from most LTO equipment vendors).

The Linux system must have access to the disc based storage areas used as source for archiving operations and as destination of restoring operations. Such storage areas can either be local to the system or mounted from network resources through any suitable protocol, such as NFS, CIFS, DAVS or other.

# Tape management

The tapes registered for use by ***LTFSArchiver*** can be grouped into named *Pools*, so that each pool of tapes makes a defined context of storage capacity. The client applications don’t need to care about which LTO tape will be used for archiving their files, as they only have to ask for a target tape pool and ***LTFSArchiver*** will select a tape with suitable free capacity for the request.

A typical use case is to define a Pool for the “master copy” and another Pool for the “backup copy”. Pools can also be associated to different Users or Customers, so the tape content is homogenous to a context and tapes can then be exchanged/delivered easily.

# Writing to LTO

***LTFSArchiver*** supports the archiving of both single files and full folder hierarchies. All the requests get back a TaskID and are queued for being dispatched and served, optimizing the use of LTO resources. The client applications can use those TaskID for tracking the status of their requests and getting back the result on their completion.

Any archived resource –file or folder – gets a URL to be used for subsequent restore, which is named **FLocat**. Such identifier includes the **TapeId** and the full path of the resource within the LTFS, according to the following pattern:

lto-ltfs:<tapeid>:<resourcepath>

example: lto-ltfs:DBK002L5:d4564f09-9034-4360-b2a2-b992fcb17ab7/MyFolderName

Besides, ***LTFSArchiver*** supports integrity checks, on client request. File checksums can be computed on tape destination and verified against checksums computed at source or provided by the client application. The produced archiving report is both recorded on tape and returned both to the client as task result.

# Restoring from LTO

***LTFSArchiver*** supports the restore of both single files and full folder hierarchies. The client application must simply provide the **FLocat** of the resource on tape and the restore destination path, which must be accessible for writing by the Linux controlling server.

The management of the requests is the same than for Writing to LTO.

# Other processing on LTFS

For supporting the processing on resources stored on LTFS tapes, without the need of creating a copy on disc, ***LTFSArchiver*** provides a service for making the tape file system directly available. The tape is in this case mounted in read-only modality, for access by the client application or related services. After use the mounted tape must be released by request from the client application, in order to let the LTO drive available for other tasks.

The use cases requiring the use of this service include:

* regular data integrity checks
* storage or format migration processes
* partial restore activities, e.g. extraction of a media fragment from a large media file

inspection of a tape for any other reason also without knowing the **FLocats**

# LTFSArchiver

|  |  |  |  |
| --- | --- | --- | --- |
| Z:\backups\progetti\PrestoPrime\Dissemination\Rai.jpg | LTFSArchiver è distribuito da RAI secondo i termini della “GNU Affero General Public License” versione 3.0 | http://10.2.7.57/rightsdraw2/Logo33.JPG  FP7-gen-RGB | Software sviluppato in PrestoPRIME (2009-2012) – progetto finanziato dalla Commissione Europea nel Settimo Programma Quadro  FP7-ICT-2007-3 231161 |

# Introduzione

***LTFSArchiver*** è un software che espone un insieme di servizi per l’archiviazione di file audiovisivi su tecnologia di stoccaggio LTO/LTFS.

LTFS (Linear Tape File System) è un formato con specifiche note ed aperte che permette l’utilizzo dei nastri dati LTO (Linear Tape Open) per la memorizzazione permanente di file generici, con modalità d’accesso di tipo file e tempi di latenza d’accesso inferiori a quelli delle tecnologie di nastri dati precedenti. Questa caratteristica è disponibile per i nastri e i drive LTO a partire dalla generazione 5. Questa tecnologia di memorizzazione è una soluzione economicamente vantaggiosa nel caso di salvaguardia a lungo termine di file di grandi dimensioni, come è il caso per i file audiovisivi al livello master di qualità. La capacità senza compressione del singolo nastro LTO va da 1,5TB degli LTO5 fino a 2,5TB degli LTO6 e l’utilizzo di librerie robotizzate permette la realizzazione di sistemi di stoccaggio scalabili di grande capacità. L’uso di LTFS è inoltre una protezione contro i problemi di interoperabilità spesso riscontrabili con le soluzioni proprietarie.

Nonostante questi vantaggi riconosciuti, l’utilizzo di apparati LTO, drive e librerie, con LTFS può non risultare sufficientemente sicuro e diretto, a meno di non utilizzare degli strumenti software specifici come quello in oggetto, che ne assicurino la correttezza delle operazioni.

***LTFSArchiver*** fornisce i mezzi per usufruire della tecnologia di memorizzazione dati LTO/LTFS nei contesti di conservazione numerica di contenuti audiovisivi. Può essere utilizzato in modo indipendente tramite la propria interfaccia web oppure essere integrato in ambiti più ampi, come la PrestoPRIME Preservation Platform (P4) o altre applicazioni cliente, per mezzo delle interfacce programmatiche (API) definite.

# Architettura e requisiti di sistema

I servizi offerti da ***LTFSArchiver*** sono ospitati da un server Apache/http e sono accessibili con le usuali tecnologie web. La persistenza dei dati relativi alle richieste ed alla loro gestione è garantita da un database PostgreSQL. ***LTFSArchiver*** richiede un server Linux (testato su distribuzioni CentOS e Ubuntu) che si collega e controlla le librerie e/o i drive LTO a caricamento manuale. Su tale server si deve installare il driver software LTFS reso disponibile dai principali costruttori di drive LTO.

Il server Linux deve avere accesso in lettura/scrittura alle aree disco utilizzate come sorgente nelle operazioni di archiviazione o come destinazione nelle operazioni di ripristino. Tali aree possono appartenere a dischi locali al server o essere *montate* a partire da risorse remote tramite i protocolli dedicati a questo scopo quali NFS, CIFS, DAVS o altri.

# Gestione dei Tape

I nastri utilizzabili da ***LTFSArchiver*** possono essere raggruppati in *Pool,* in modo che ciascuno rappresenti un contesto di archiviazione separato. Le applicazioni cliente non dovranno preoccuparsi di quale nastro LTO utilizzare ma dovranno solo indicare un pool di destinazione. Sarà ***LTFSArchiver*** a scegliere in base alla dimensione di ciò che si archivia ed alla capacità residua dei singoli nastri disponibili nel pool.

Un caso tipico di utilizzo è la definizione di un Pool per la “qualità master” e di uno per le “copie di backup”. Un altro è l’associazione di differenti Pool a specifici Utenti o Clienti in modo che il contenuto dei nastri dati sia omogeneo al contesto e faciliti le operazioni di scambio e consegna.

# Scrittura su LTO

***LTFSArchiver*** supporta sia l’archiviazione di file singoli, sia l’archiviazione di intere gerarchie di cartelle. Tutte le richieste inviate, ottengono in risposta un TaskID e sono accodate per poi essere servite con logiche di ottimizzazione delle risorse. Le applicazioni cliente possono usare questi TaskID per tracciare lo stato delle richieste e per ottenere i risultati una volta evase.

Ogni risorsa archiviata – file o cartella – riceve un URL particolare, chiamato **FLocat,** da usarsi per il successivo ripristino. Tale identificatore include il **TapeId**,che idenditifca il nastro, ed il percorso completo della risorsa nel file system LTFS, secondo il modello seguente:

lto-ltfs:<tapeid>:<resourcepath>

esempio: lto-ltfs:DBK002L5:d4564f09-9034-4360-b2a2-b992fcb17ab7/NomeMiaCartella

***LTFSArchiver*** supporta il controllo di integrità, a richiesta del chiamante. Le checksum sui file possono essere calcolate sulla destinazione nastro e verificate rispetto a quelle calcolate all’origine o date dall’applicazioni cliente. Il rapporto di archiviazione prodotto è sia memorizzato su nastro che restituito al cliente come risultato dell’operazione.

# Ripristino da LTO

***LTFSArchiver*** supporta il ripristino di file e intere gerarchie di cartelle. E’ sufficiente che il cliente fornisca il locator **FLocat** della risorsa su nastro e il percorso di destinazione che deve essere accessibile in scrittura da parte del server di controllo Linux.

La gestione di questo tipo di richieste è del tutto simile a quella delle richieste di scrittura, è quindi possibile seguire lo stato ed ottenere un report al termine dell’operazione.

# Altre modalità di accesso LTFS

Al fine di poter rendere accessibile i dati su nastro senza avere la necessità di farne una copia, ***LTFSArchiver*** fornisce un servizio di “make available”. In questo caso il singolo nastro è montato in modalità solo lettura e il suo file system può essere acceduto direttamente dal cliente che ne ha fatto richiesta. Al termine dell’utilizzo il cliente dovrà rilasciare la risorsa con apposita operazione di smontaggio, pena il blocco della medesima e relativa impossibilità di utilizzo da parte di altri clienti.

I casi tipici di utilizzo di questa modalità di accesso includono:

* controlli periodici di integrità
* processi di migrazione di formato
* ripristino parziale di un file, per esempio estrazione di un intervallo di multimediale a partire da un filmato di grandi dimensioni
* ispezione del nastro dati anche senza conoscere gli FLocat