大容量データのアーカイブに適したライブラリ装置
-データ分析基盤の構築のための-

IDEMA JAPAN アーカイブ WG
〒105-0003　東京都港区西新橋 2-11-9

Library device suitable for large-capacity data storage for construction of data analysis base

IDEMA JAPAN Archive WG
2-11-9, Nishi-Shinbashi, Minato-ku, Tokyo 105-0003, Japan

（要旨）
ビッグデータ、AI、IOT 等によりデータ活用が進むと、データ保存に必要なストレージの消費電力とビットコストの削減が深刻な課題となる。この課題克服のためには、現状の HDD や SSD の中心としたストレージシステムにストレージの階層化技術を用いて、光ディスクやテープといった Removable Media のライブラリ装置を活用することが有効である。

（キーワード）
Removable Media、データ利活用、データ分析プラットフォーム、データ蓄積、大容量ストレージ、ストレージコスト、データアーカイブ、ライブラリ装置、長期保管、階層化、情報ライフサイクル管理

目次

1　はじめに

デジタルトランスフォーメンションの広がりによって、日々、大量のデジタルデータが生み出されている。それらのデータをITシステムで分析することで、これまでの人間による解釈、推測では考えが及ばなかった新たな知見を得ることができる。

Big Dataの分析による潜在知の発見、IoTシステムによる生産性の向上、Deep LearningによるAI（人工知能）の高度化など社会の利便性、ビジネスの効率性や収益性の向上につなげる取り組みが多くの分野で行われている。

これらの取り組みに共通していることは「より大量のデータ」を入力することがデータ分析の精度向上につながると言うことである。このためデータ蓄積機能は、データ分析IT基盤における重要な検討ポイントといえる。データ分析の試行段階ではHDDやSSDを中心としたストレージ装置にデータを蓄積するのが一般的である。しかし、実証あるいは実用段階に移行するとデータ量は急激に増大し、ストレージはシステムのコスト、運用上の大きな課題となる。

増大するデータ量、それに伴う ストレージ装置のコスト、運用の課題に対するベストプラクティスとして広く認識されているのが図1の記憶階層モデルに基づくストレージの階層化である。図2には階層制御を用いたデータ蓄積システムの構成例を示す。

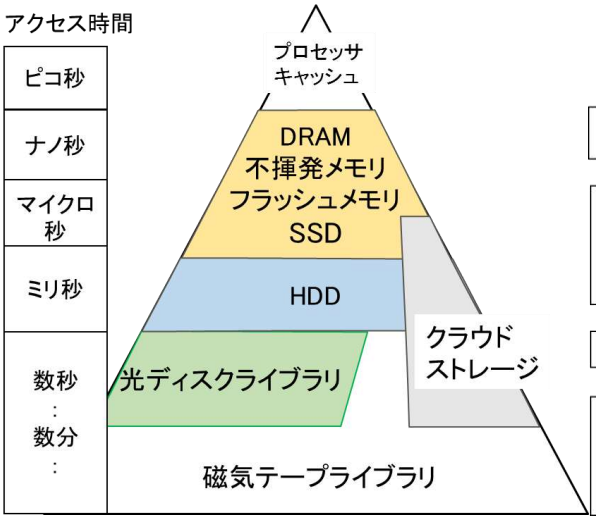


図1　ストレージ階層モデル

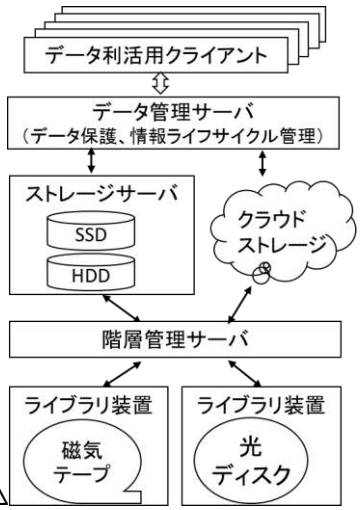


図2　データ蓄積システム構成の例

HDDより上位の階層は、現状のITシステムでも一般的なものであり、「キャッシュ」や「先読みバッファ」などのソフトウェア技術によって、最適なコストで、より高速な処理に実現に貢献している。下層のライブラリ装置について馴染みのある方は少ないだろうが、省電力性，長期保存性など、データ蓄積システムにおいて有利な多くの特長を持っている。

本文では、大容量のデータ蓄積に適したライブラリ装置の特徴、活用方法、さらに導入にあたっての留意事項、記録メディア、ドライブ、ロボットといったライブラリ装置で用いられる構成要素について説明する。さらに、活用する際に必要となるマイグレーション、階層管理、冗長性、シームレス記録、ライブラリ装置対応コマンド等のデータ運用方法とともにライブラリ装置のシステム構成を解説する。

ライブラリ装置の活用により、データ利活用の鍵である「より大量のデータ」をあきらめないデータ分析基盤の構築の一助になれば幸いである。

2　Removable Media 対応システム

データをアーカイブするためのメディアとしては、HDD や SSD のように読み書きの機構と記録媒体が一体化しており、常時オンラインで使うことを前提としたものと、テープや光ディスクなどドライブから取出して棚管理も可能なメディア（Removable Media）に二分される。

さらにデータのアーカイブ先としては、この 10 年で急浮上して、今では多くのシステムにとって接続が必須となりつつあるクラウドサービスがある。これらは一般的にサブスクリプション方式のサービスであり、短期間で準備できて、小さな初期投資で開始（Small Start）が可能である。一方で、長期アーカイブのコストに関しては高額になる傾向があり、利活状況次第では比較的短期間でも想定外に大きく課金されるなど不透明感があるため、オンプレミスのアーカイブシステムと単純にコスト比較することを困難にしている。

しかしながら、クラウドサービスであっても、裏ではストレージメディアを使用しており、その多くは現在も HDD をベースとしたストレージであると考えられるが、長期保存用サービスには、テープやオプティカル・ライブラリ装置による階層管理を取り入れて運用コストを抑えている例もある。

それぞれのメディアの特徴を表 1 に示す。

メディア	HDD	テープ	SSD	光ディスク		
	7200RPM	LTO	3D QLC	Blu-ray	アーカイバルディスク	M-DISC ※1
電力不要のデータ耐久性	4 年未満	30 年以上	1 年未満	最大 50 年	最大 100 年	最大 1000 年
カートリッジ/ドライブ単位の物理容量	12TB	12TB/30TB ※2	最大 128TB	1.5TB ※4	3.3TB ※4	100GB
目標最大物理容量	20TB	330TB	最大 256TB	1.5TB	11TB	100GB
フォームファクター	3.5 インチ	4.1 インチ	2.5 インチ	4.7 インチ (ディスク直径)		
不変性	SW-WORM	SW-WORM	SW-WORM	メディア特性 (TRUE-WORM)		
最大転送速度（MBps）	最大 261	最大 360~900 ※2	最大 500	最大 146 ※4	最大 250 ※4	最大 30
下位互換性	無制限	1～2 世代	無制限	無制限		
取り外し可能なカートリッジ/ドライブ	×	○	×	○		
GB 単位相対の入手コスト	低	最低	中 ※3	中	低	中
GB 単位相対の TCO	中	低	低~中	低	最低	低

出典:MARC STAIMAR ※1:Blu-ray テクノロジーがベース。 ※2:LTO8 の例、 圧縮比 2.5:1。 ※3:3D QLC の生産量、需要、工場生産能力に応じてコストと価格が低下する見込み。

※4 ソニー・オプティカルディスクアーカイブ・カートリッジとして

表 1　各種ストレージの比較[3)]

2.1　Removal Media

ストレージメディアの中で一般的に長期アーカイブに向いているのは、テープや光ディスクのような Removable Media であり、基本的に読み書き時以外は電力が不要という特長がある。

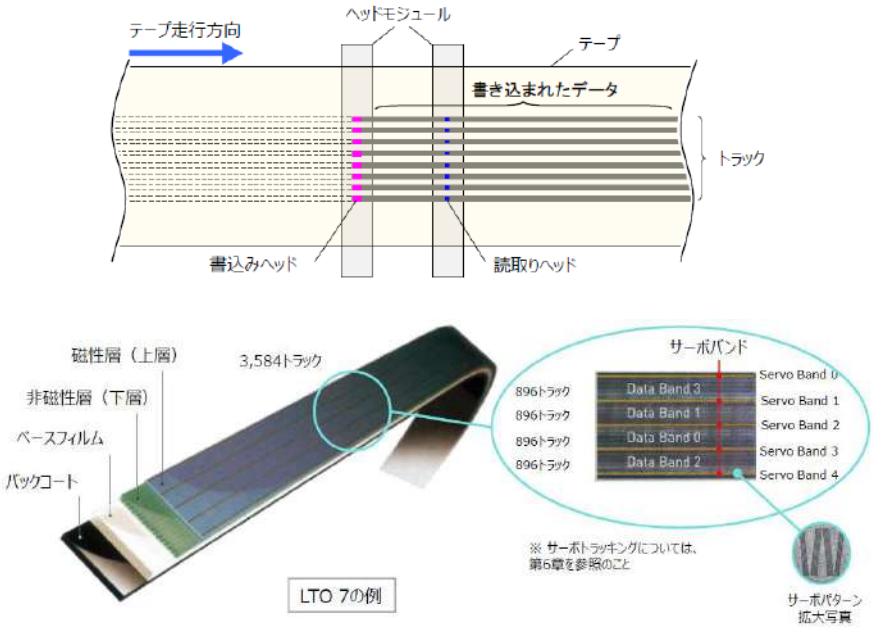
2.1.1　テープ

磁気テープの歴史はコンピュータが登場した約 70 年前に遡り、様々な形状、フォーマットのテープストレージ装置が開発されてきたが、現在、データアーカイブ用としては

LTO（Linear Tape Open）が一般的にも用いられている。LTO は、2017 年に発売された 8 世代目（LTO Ultrium 8）で一巻あたりの非圧縮容量 12TB を達成している。デバイスとしての成熟度は非常に高く、世代を追う毎に記録容量、読み書き速度、信頼性などが向上を遂げている。今後も大容量データを安価にアーカイブするためには無くてはならない技術と言える。

図 2　LTO テープカートリッジ外観

LTO では 1/2 インチ幅のテープに微細な帯状トラックを往復記録することで大容量化を実現している。ドライブ側はデータ転送速度向上のために世代を追うごとに多チャンネル化してきており、最新の LTO8 では 32ch（チャンネル）のヘッドが上下しながら最大 6,656 本のトラックを書き込む。記録・再生用で別々のヘッドを搭載することで、書き込み中のデータを書き込みと同時にベリファイすることで信頼性を担保している。メディア側には製造時に複数のサーボバンドが記録されており、全長 900m 前後のテープをドライブが常時正確にトラッキングすることを助けている。



JEITA, テープストレージ動向＜2018年版＞, https://home.jeita.or.jp

図 3　テープドライブの記録・再生動作と磁気テープ構造の概要

2.1.1.1 光ディスク

光ディスクの歴史は、1982 年の CD から始まり、DVD、Blu-ray（BD）、さらにそれを多層化した BDXL と進化してきた。その間に様々な派生バージョンも開発されてきたが同じ半径の円盤状メディアをレーザで読み書きする点は共通であるため、互換性が保ちやすいという特長がある。しかも温湿度変化に対する耐性が非常に高いことから長期保管に最適なメディアと言える。光ディスクは、空調費用が最も少なくて済むだけでなく、自然災害などによる一時的に極端な状況に置かれてもデータの読み出せる可能性が高い。

一方で、光ディスクの課題はデータ容量と転送速度であった。これらの課題を解決するものとして、2014 年にアーカイバル・ディスク（AD）という規格がパナソニックとソニーにより発表された。AD は両面 6 層の記録層を持ち、ディスク 1 枚当たりの記録容量が 300GB である。

転送速度に関しても両面同時アクセスに加えてレーザ（BD と同じ波長 405nm）を多チャンネル化することで 300MB/s 以上の転送レートを実現している。AD は、今後の数年でディスク容量を 1TB まで伸ばすと同時に、さらなる転送速度向上が計画されている。

図 4　光ディスクと光ディスクカートリッジ

光ディスクドライブは、データの書き込みや、記録された情報の読み出しにレーザ光等を使用する光ディスクの記録再生装置である。スピンドルモーターによって光ディスクを回転させ、光ピックアップ（レーザー・対物レンズを含む）によって光ディスク内のデータを読み書きする。光ピックアップ全体を、ディスク回転と同時に、半径方向に移動（シーク）させることで対象アドレスにアクセスするが、その際に物理的接触はない。また、レーザの焦点を結ぶ位置を変えることで、複数の記録層を持つメディアにも対応する。

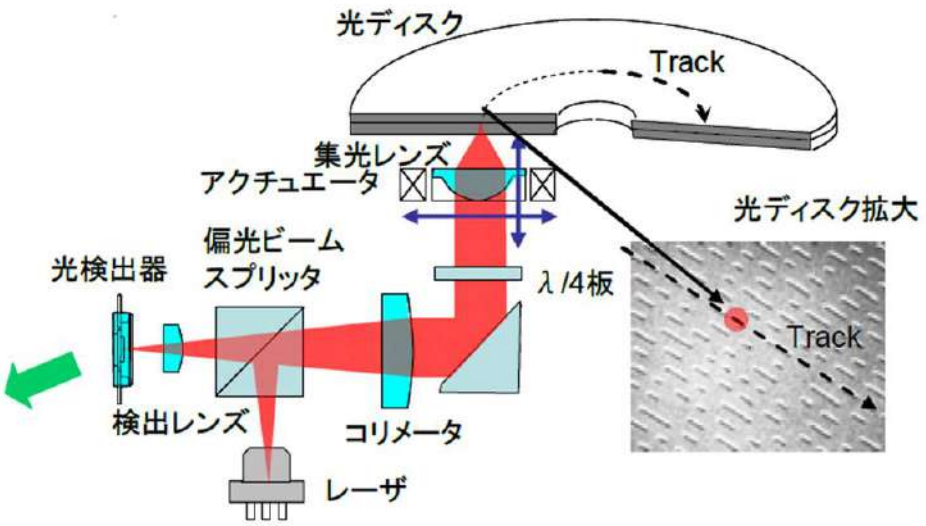
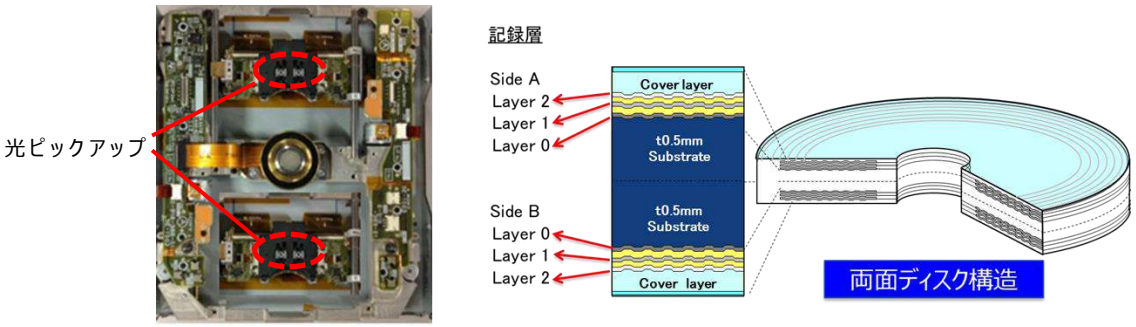


図5　光ディスクドライブの原理図[4]



(a)　マルチピックアップの例
ソニー製 Optical Disc Archive
ドライブ
8ch 光ピックアップ　(片面分)

(b)　複数の記録層を持つ光ディスクの例
アーカイバルディスク
6 層の記録層を持ち両面同時にアクセスする

図6　読み書き速度や容量を向上させるための技術例

2.2　オンラインストレージデバイス

一方、特に利活用の多い、所謂アクティブアーカイブに関しては、保存期間が長期であっても HDD や SSD だけで構成するケースが多く、その場合は Software-defined Storage (SDS)*の技術を組み合わせることで拡張性や冗長性を担保するのが昨今のトレンドである。ただし、大容量になるにつれ、SDS 用の Node（サーバ）の数が増え消費電力が激増してしまうため、今後はなんらかの自動階層化技術 (Auto-tiering) を使って、アクセスが極端に少ないデータ、すなわちコールドデータを選別して長期アーカイブ用ストレージに移動することで、より適切にデータを配置することが望まれる。

また、HDD、SSD に関しては以下のような技術を用いてビット単価を下げる努力がされてきた。

2.2.1　HDD

従来の LMR（面内磁気記録）や PMR（垂直磁気記）では既に高密度化は頭打ちであったが、ランダムアクセス性を犠牲にして高密度化を図った SMR（瓦記録）や、PMR をアシストする技術として、HAMR（熱アシスト磁気記録）や MAMR（マイクロ波アシスト記録）といった新技術の出現により、未だ緩やかに高密度化は進んでおり、2020 年までに 16TB から 20TB, 特に MAMR では 2025 年までに 40TB に達するという予測がある。

2.2.2 SSD

2D から 3D NAND 型フラッシュメモリに移行することで、この数年で劇的に大容量化すると共にビット単価も下がってきており、今年（2019 年）中にはモバイル・コンピュータの半数以上で SSD が採用されるとされている。また、HDD ではスループット低下が大きく、あまり広がらなかった冗長排除の技術も、昨今の AFA（All Flash Array）とは相性がよく、データ種類によっては高圧縮が効いてビット単価が一気に下がり、少なくとも High Performance HDD（10000 回転以上 SAS HDD 等）に比肩できるレベルとなってきた。

*Software-defined storage
サービス管理インターフェースを持つ仮想化されたストレージ。
ソフトの例：オープンソースの OpenStack Swift や Cloudian 社の HyperStore。

2.3 ロボット

2.3.1 光ディスクライブラリ（チェンジャー）の歴史

1985 年、オーディオ CD から CD-ROM が出現し、1989 年には太陽誘電から CD-R が発売された。当時は、5 インチ MO が光ディスクとして既に普及しており、各種ライブラリ装置が HP、SIGNET、Plasmon 等から発売されていた。ライブラリ装置の制御コマンドは、SCSI の Changer Device 規格でほぼ統一されていたため、各社の UNIX マシン、OS で標準的に稼動していた。1995 年頃には、CD-R ライターを搭載し、書き込み可能なメディアが使用できるライブラリ装置が登場した。その後、DVD-R/RAM が発売され、さらに Blu-ray が発売されたが、その時点で残っていたライブラリ装置の国内メーカはアサカだけだった。そのアサカも、2013 年 4 月でライブラリ事業を終息した。

一方で、2016 年、Facebook が、自社データセンターで使用するコールドデータ保管用のストレージとして、10,000 枚の Blu-ray ディスクを格納可能なライブラリを発表。最初はメディア 1 枚で 100GB の Blu-ray ディスクで開始し、その後は Panasonic と Sony が共同開発した Archival Disc（1 枚で 300GB）を使用したものも納入されているとされるが詳細は公表されていない。

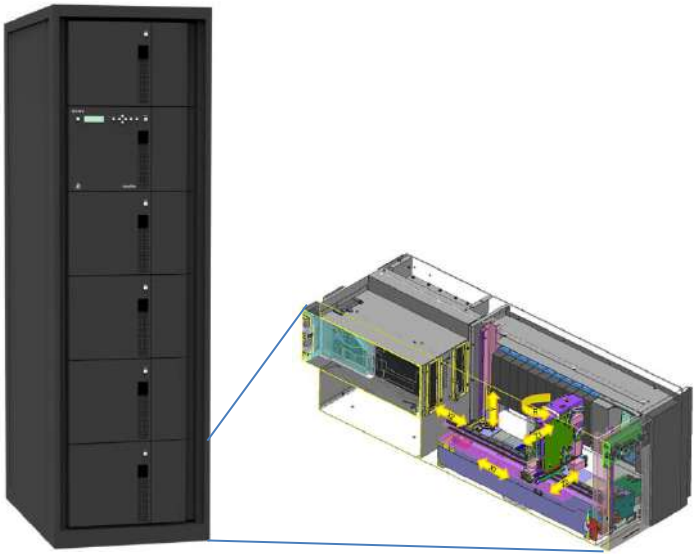


図 7　光ディスクライブラリの例（ソニー・ペタサイト ODS-L30M）

2.3.2　光ディスクライブラリ・システムの制御

光ディスクやテープのライブラリは共通コマンドで制御できるようにインターフェース仕様が共通化されている。ここではコマンドの概要を紹介し、ライブラリがどのように動作するかを説明する。

(a)　SCSI の誕生と Medium Changer コマンド

SCSI（Small Computer System Interface）が最初に規格化されたのが 1986 年で、Medium Changer コマンドは 1998 年 3 月に SCSI-3 として最終の仕様が公開された。インターフェースは、当初のパラレル通信からシリアル通信に切り替わったが、ライブラリ制御コマンドは同じままである。ライブラリ内のハードウェアを機能別にエレメントとして定義し、個々のエレメントに対してアドレスが付加される。例えば、ストレージ・スロットが 24 個あるライブラリでは、ストレージエレメントのアドレスが 1 から 24 まで割り当てられて、ライブラリ内の物理的な位置に対応する。データ転送エレメント（ドライブ）も同様で、物理的な位置順にアドレスが割り振られる。

(b)　Media Transport Element（メディア搬送部）

カートリッジやメディアをスロットからドライブ等へ移動する搬送メカ部分。カートリッジ・メディアはそのまま掴んで移動するが、ベア・ディスクではトレイに載せたり、センターホールで保持したりして移動する。

(c)　Storage Element（ストレージ・スロット）

メディアを格納する場所のことで、ステータス取得コマンド（Read Element Status）でメディアの有無を検出する。各社ライブラリはストレージ密度を向上させ、アクセス時間を短縮するために様々な工夫をしているが、移動する際の論理的な違いはエレメント数のみである。カートリッジ・メディアのシリアル番号（Medium Auxiliary Memory に格納）　や、添付されたバーコードの読み取りをサポートしているライブラリも同様で、移動には Storage Element 番号しか使用せず、これらの

附帯情報は、Read Element Status で読み取る。

(d) Import/Export Element（メディアの出し入れ）

カートリッジやメディアをライブラリから出し入れする部分で、I/O スロットとかメールスロット等とも呼ばれる。

(e) Data Transfer Element

ライブラリ内のドライブ搭載位置（スロット）を示す。通常は内蔵ドライブ自体と同義であるが、ドライブ非搭載のスロットにもアドレスは割り振られる。メディアをドライブに移動するには、Move Medium コマンドに、ストレージエレメントとデータ転送エレメントのアドレスを指定する。

(f) ライブラリ装置の運用上の課題

メディアのドライブへの移動時間と、移動後にデータアクセス可能になるまでの時間は、ハードディスク等の固定メディアに比較して著しく長く、ライブラリ装置の種類、規模、構成、及びドライブとメディアの種類で決定される。通常は、上位ソフトウェアでこれらを考慮して、タイムアウト等のエラーが発生しないようにすると同時に、メディアのアクセス順を最適化することで効率的に使用できるようにする。

2.3.3 オフラインライブラリ（メディア保存用の棚）

アクセス頻度の少ないデータを保存したメディアをライブラリ装置から取り出し、廉価なオフライライブラリに保管することもできる。前章(c)で述べた RFID に記録したカートリッジ・メディアのシリアル番号や添付されたバーコードを読み取る機能を備えたオフラインシェルフもあり、これを用いることにより、読み出したいデータが保存されているメディアがどこに保管されているかがオンラインで分かり、メディアの管理を容易に行うことができる。

(a) オフラインライブラリ

(b) RFID タグ内臓 35 枚入りマガジン

(c)

（c） 360 度回転可能なマガジンラック

図８ オフライライブラリの例（Netzon 社 HMS Soff シリーズ）

2.4 システム構成

2.4.1 マイグレーション

データを記録したメディアにはそれぞれに平均的な寿命が存在するため、保存状態を適切なタイミングでチェックしたり、その時点で一番相応しいメディアにデータを移行することは長期アーカイブにおいては避けられない。メディアの寿命を超えて長期保管する際には、マイグレーション計画を立てる必要がある。

2.4.2 階層管理ソフトウェア

階層管理ソフトウェアを使用することで，簡単かつ安全に低コストの Removable Media を使用することが可能となる。ポリシー設定に従ってキャッシュと Removable Media の間でファイルの移動を階層管理ソフトウェアが自動的に行うため，ユーザは Removable Media の存在を意識することなく，使用することができる。頻繁に使用するデータはキャッシュに置き，使用頻度の少ないデータは Removable Media のみに保存するようにポリシーを設定することで，高速なキャッシュと低コストな Removable Media の双方のメリットを有効に活用することが可能となる。

階層管理ソフトウェアを効果的に使用するためには以下のポリシー設定を行う。

(a) 記録ポリシー

キャッシュ上に存在するファイルを Removable Media に書き込む条件を設定する。

（設定例）

- ファイル生成からの経過時間
- ファイル更新からの経過時間
- スケジュール設定
- 即時

(b) 降格ポリシー

キャッシュ上にファイルの情報を残しファイルの実体を削除する条件を設定する。記録が完了している事が前提条件となる。

（設定例）

- ファイル更新からの経過時間
- ファイルの最終アクセス時間からの経過時間
- キャッシュの使用率

(c) 昇格ポリシー

Removable Media からキャッシュ上にファイルを読み戻す条件を設定する。

（設定例）

- ファイルアクセス時
- スケジュール設定（事前読込）

(d) 削除ポリシー

Removable Media からもキャッシュ上からもファイルを削除する条件を設定する。

（設定例）

- ファイル生成からの経過時間
- ファイル更新からの経過時間
- スケジュール設定

2.4.3 情報ライフサイクル管理（ILM）

情報ライフサイクル管理とは、情報が作成されてから廃棄されるまでの期間中のデータの管理、プロセス、ポリシーなどの事であるが，その重要な要素として、アクセス頻度の低いデータは低コストのストレージに、アクセス頻度の高いデータは高速なストレージを使用するべきとされている。また、データ保持期間を過ぎたデータは自動的に削除されることが望まれる。階層化ソフトウェアは，これをユーザの操作なしに自動的に行うことができるものであり、Removable Media を有効に利用するための重要な要素となる。

2.4.4 冗長性（レプリケーション、イレージャーコーディング）

データアーカイブにおいては、通常は、階層化ソフトによって複数メディアにレプリケーションされる。データのレプリケーションは、ストレージのオーバーヘッドコストを伴うが、障害に対応するシンプルかつ効果的な方法である。データのレプリケーションと比較して、容量効率が良いということで、イレージャーコーディング（Erasure Coding）が使用される例もある。この技術は、データの冗長性を確保するためのパリティデータを演算し、元のデータを拡張し、さらに分割して、複数の媒体に記録する。

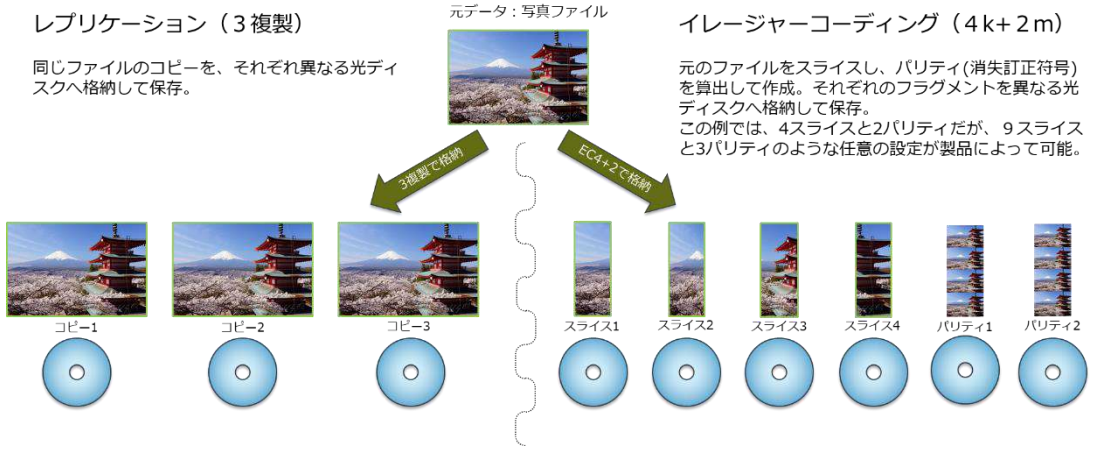


図９　レプリケーション方式とイレージャーコーディング方式

この例においては、どちらの方式もサーバが２台故障しても、データを維持できる（消失しない）。イレージャーコーディング方式は常に計算が伴うため、性能的に複製方式が有利な面もあるが、イレージャーコーディング方式は容量効率やサイズの大きなファイルを扱う際に効果的といった面を持つ。上の例で見ると、複製方式は３倍の格納領域が必要な事に対し、EC 方式は約 1.5 倍の領域で済む。

2.4.5 シームレス記録（カートリッジ・スパニング）

記録メディア容量を超えるファイルを複数の記録メディアに記録することが可能である。小さなサイズ多数のファイルをキャッシュディスク上で一つにまとめて、記録再生効率の良い大きさのファイルにして記録することも行われている。

2.4.6 クラウド連携アーカイブシステム

クラウド同様の環境をオンプレミスにも構築できる OpenStack Swift や、クラウドサービスに代表される Amazon S3 が提供する既存のクライアントや SDK を活用することで、膨大なデータを取り扱うクラウドの技術をプライベートクラウドに持ち込み、オンプレミスに存在するライブラリ装置に膨大なデータをアーカイブすることができる。これを実現するために、ライブラリ装置とクラウドを連携するゲートウェイソフト*を用いる。

*ゲートウェイソフトの例

・PoINT 社製 PoINT Archival Gateway
・Qstar 社製 QS2
・日本テクノ・ラボ社製 MNEMOS Gateway

2.5 Removable Media の対応での注意点

前提条件として、既存のスケールアップ*型の Removable Media をどのようにスケールアウト**型の世界に落とし込めるか、どうやって/どこにバッファを持たせるかといった観点での検討が必要で、下記のような注意点（システム特性）をカバーできるような視点が必要である。

・データの格納先が多段化され、状況（例えば古くなったデータ）によっては、読み出しに時間が掛かる可能性がある事を認識していないとシステムとして成立しない。
・多重処理を許容する場合は、多重度の上限設定が必要。
・データの分割機能や集約機能。異なる特性のシステムなので、最適なデータ I/O 特性も異なるため、最適なデータサイズに加工することで効率化を図ることを推奨する。あまりに掛け離れた用途には使用してはならない。

*スケールアップ
コンピュータシステムの性能を増強する手法の一つで、コンピュータの構成部品をより高い性能や容量のものに交換・増設することにより拡張すること。コンピュータ自体を丸ごとより高い性能のものに入れ替える場合もある。[1)]

**スケールアウト
コンピュータシステムの性能を増強する手法の一つで、コンピュータの台数を増やすことでシステム全体の性能を向上させること。処理を並列化、分散化できるシステムで適用される。[1)]

引用文献

1) IT 用語辞典 e-Words http://e-words.jp/
2) SNIA ストレージネット和キング用語集
3) TechTarget ジャパン ホームページ
https://techtarget.itmedia.co.jp/tt/news/1802/12/news02.html
4) 日野泰守著、電子通信情報学会 知識ベース 知識の森 3-2 光ディスク装置
http://www.ieice-hbkb.org/files/08/08gun_02hen_03.pdf#page=7