

データマイグレーションの必要要件

継続的マイグレーションによる
デジタルデータの長期保存

はじめに

当分科会(テープストレージ専門委員会・長期保存分科会)ではこれまで、デジタルデータを100年以上の長期に亘って保存する「デジタルデータの長期保存」について議論・研究してきました。

そして、デジタルデータを次々と新しいストレージ装置に移し変える(マイグレーション)ことで、100年以上の長期保存も可能になると考えました。

テープストレージを専門とする当分科会では、マイグレーションの中でも特に、記録されたビット列をそっくりそのまま保存し続ける「ビットマイグレーション」に注目しています。

アーカイブシステムの内側でビットマイグレーションを継続的に繰り返すことで、単一のストレージ装置では実現できない程の長期のデジタルデータの保存を実現するシステム(以下、マイグレーションシステム)が構築できます。

この資料では、デジタルデータの長期保存を目的として継続的にマイグレーションを繰り返す際に、必要となる要件について説明しています。説明には、なぜ必要なのかという面も含めました。また、要件の対象として、「システム」「データ」「作業」の3種類に分類し、それぞれの要件名の最後に表記しました。

手に入れやすい装置の使用

マイグレーションシステムには手に入れやすい装置を使用すること [システム]

ここで言う「手に入れやすい装置」とは、一般的な個人・団体が容易に入手できるという意味で、市販されている装置のことである。入手困難な装置(例えばプロトタイプや研究レベルの装置など)を使用した場合、その装置が壊れてしまったら、継続的なマイグレーションを行うことが出来なくなる。

また、記録装置の媒体価格も重要な要素である。媒体のビット当たりの単価が高価であると、アーカイブするデジタルデータが増えるにつれてコストが増大してゆく。

信頼できる装置の使用

マイグレーションシステムは信頼できる装置を使用すること [システム]

信頼できる装置を定義することは困難であるが、一般的には多くのユーザーに広く使われている装置などはある程度の信頼性を備えていると考えられる。特にシステム全体の鍵となる部分のハードウェアが、時々故障するような信頼性の低いものでは、安全で確実なマイグレーションは期待できない。マイグレーション作業中にデジタルデータが失われる危険性も高くなってしまう。

一方、広く普及している安価な装置を大量に並列稼働させ、システム全体の信頼性を高めるという手法も考えられる。アーカイブの場合に置き換えると、デジタルデータの複数のコピーを作成し、複数のシステムで複数管理し、複数のマイグレーションを並列に行う、という形になる。

保存環境の改善

マイグレーション先のシステムが、保存環境を改善できること [システム]

今日現在アーカイブとして保存しているデジタルデータを失うリスクは年々高まってゆく。記録メディアの経年変化による読み取りエラー率の上昇や、システム全体がサポートできる記録容量の不足、記録メディアの格納場所の不足、など記録メディアに関わるものだけではなく、システム全体の信頼性の低下や故障時の代替装置の入手が困難になるなどアーカイブシステム・装置に関わるものがリスクとしてあげられる。これらのリスクをできるだけ回避するために、マイグレーション先のシステムは、単位体積あたりの記録容量がより大きい記録メディアや、よりエラー率の低い装置、より新しく入手しやすい装置などを選択することで、マイグレーション元の環境に比べ、デジタルデータ損失リスクがより少ない環境を提供できるシステムであることが望ましい。例えば、テープストレージ装置のLTOドライブであれば、より新しい世代のドライブ・記録メディアにマイグレートしていくことなどは、この要件に合致すると言える。

そのままの形でマイグレーション

マイグレーション対象デジタルデータを、そのままの形でマイグレーションすること [データ]

一般的にデジタルデータと総称しても、実際にそれぞれのデジタルデータがあらわす内容は文書・音声・映像など様々である。その内容はある決まった法則にしたがってビット列に変換され記録メディアに記録される。このとき、システムに記録されたビット列は記録したとおりの形で読み出されることが期待されている。

仮にマイグレーションの1過程において、デジタルデータの暗号化を実行したとする。このマイグレーション以降は、暗号化された状態のデジタルデータがマイグレーションされてゆく。後に、保存されているデジタルデータから元の内容を読み出す必要が生じた際、ビット列への変換法則の他に、暗号データの復号処理が必要となる。つまり、ビット列への変換法則しか知らない場合には、記録されたデジタルデータを完全に読み出すことができても、その内容を再利用することができなくなってしまう。

このような事態を避けるためには、記録したビット列に対してどんな種類の変換も加えることなく、そのままの形でマイグレーションを繰り返すことが重要である。

関連する情報を一緒にマイグレーション

マイグレーションの際には、対象デジタルデータと関連する情報を一緒にマイグレーションすること [データ]

デジタルデータの解釈に必要な情報のことをメタデータと呼ぶ。デジタルデータのマイグレーションにおいては、メタデータをデジタルデータ本体と一緒にマイグレーションすることが必要である。

たとえば、単純な例として、「ファイル名」というメタデータを考えてみる。システムはファイル名につけられた拡張子を参照してそのファイル内のデジタルデータの種別を判断している。仮にこのファイルのファイル名がわからなくなると、そもそもそれが何のデジタルデータなのかがわからなくなってしまい、結果として再利用することができなくなってしまう。

一方、マイグレーション作業に関するメタデータも考えられる。これらのメタデータはデジタルデータ本体の内容の解釈には必要ないが、これまでのマイグレーションが確実に間違いなく行われたことを記録し、長期保存されているデジタルデータ自体が正確であり信頼おけるものであることを保証するものとなる。

マイグレーションにおいては、これら2種類のメタデータをデジタルデータ本体と共にマイグレーションすることが必要となる。

正常に終了したことを確認

マイグレーションが正常に終了したことを確認すること [作業]

デジタルデータのマイグレーションが正常に終了していなければ、そのデジタルデータは完全とは言えず、結果的にデジタルデータを失ったことと同じことになってしまう。よってマイグレーションの正常終了を確認することは重要である。

マイグレーションが正常に終了したことは、マイグレーション前後のビット列が同じであることを比較検証することで確認できる。例えば、そのデジタルデータのチェックサム値を取得し、マイグレーション前後で比較することができる。

デジタルデータの一一致検証のようなマイグレーションが正常に行われたかどうかのチェックは、実際にマイグレーションを行う際のユーティリティやISV提供ソフトウェア、等が提供・管理することになる。

結果の記録

マイグレーションの実行とその結果を記録すること [作業]

デジタルデータのマイグレーションを実行した際に、その結果などを記録することで、記録したデジタルデータの完全性や原本性を確認することができる。

さらに、共に保存する情報として様々な用件が提案されている。例えば、いつどのような形でマイグレーションしたかを記録しておくことで、当時の再生技術を将来にわたっても保持すべきかどうかの判断材料とすることができるだろう。他に、実施日時・マイグレーション管理責任者・部門名、デジタルデータ総バイト数、マイグレーション時のエラーレート、マイグレーションに使用したツール、などの記録が考えられる。

これらの記録は、記録しているデジタルデータの内容や品質に直接的に関わるメタデータではないが、記録されているデジタルデータの品質を間接的に保証するために有用である。

マイグレーション先のシステムでも読み出せる

マイグレーション対象デジタルデータがマイグレーション先のシステムでも読み出せること [システム]

マイグレーションしたデジタルデータを読み出せる環境が常に存在することを保証する必要がある。そのためには、特定のソフトウェアやハードウェアに依存しないフォーマットで記録されるべきである。

記録フォーマットを標準化することは、特定機器への依存のない記録方式を実現するひとつの方法である。

戦略の重要性

マイグレーションにおける戦略の重要性 [作業]

ただ漫然とマイグレーションを繰り返すだけではデジタルデータの長期保存は実現できない。本資料で述べているそれぞれの必要要件を満たす記録方式を用いたシステムを構築することは、マイグレーションにおける正しい戦略のひとつであると考えられる。

またマイグレーションの周期についても、記録したデジタルデータの品質が劣化する前にマイグレーション処理を行うことができるように、それぞれのシステム構成に応じて適切に設定する必要がある。

例えば、テープストレージのLTOメディアの第三世代について本委員会の媒体SWGが加速試験を行ったレポート[下記参照]では、10年を目安としたマイグレーションを推奨している。

[資料]

テープメディア保管寿命評価(2009),テープストレージ専門委員会,
<http://home.jeita.or.jp/is/committee/tech-std/std/com02.html>

補足

この資料で紹介している要件全てが、マイグレーションシステムの構築に必須であるわけではありません。どの要件が対象システムで有効となるかはケースバイケースとなります。しかしながら、特に長期保存を対象としたシステムにおいては、ここで紹介している必要要件全てについて検討を加えることを推奨します。

本資料の中での「デジタルデータ」という表現は、コンピュータによって扱われる電子データを対象としています。