第5章　ホットレプリケーション：高アクセス頻度データの複製クラスタリングによる三次記憶システムの高速化

5.1　はじめに

一般に，テープドライブ装置においてはデータへアクセスする場合にはシークが必要であり，そのために要する時間がリクエストが発行されてからデータの読み書きが終了するまでの時間に対して大きな割合を占める場合も少くない．データを記録する際にアクセス頻度の高いデータをシークが短くなる位置に配置することで応答時間を短縮することが可能であるが，一般にデータが生成された時点においてそのデータのアクセス頻度を予測することは困難である．一方，一度テープ上へ記録されたデータの再配置のためには時間的，空間的に大きなコストを要するため，各データのアクセス頻度が判明した後に再配列を行うことは必ずしも得策とは言えない．

本章では，テープ上に予め確保した領域へ高アクセス頻度データを複製し，クラスタリングをすることで応答性能の向上を図るホットレプリケーションと名付けた手法を提案する．ホットレプリケーションでは，複製されたデータへのアクセスが不利になることがないよう，テープ途中でロード/イジェクト可能なテープドライブ装置を用いる．ホットレプリケーションでは，データが多重化されていることによるアクセシビリティの向上[1]，および高アクセス頻度データがクラスタリングされていることによるシーク時間の短縮により応答時間の短縮を図る．

以降，5.2節において，ホットレプリケーション手法について説明を行い，5.3節におい

[1]アクセシビリティが向上するとは，複製を生成することにより原データおよび複製データの両者に対してアクセスが可能となり，アクセス多重度が向上することを意味する．

て，ホットレプリケーションの基本性能についてシミュレーションにより評価を行う．

5.2 ホットレプリケーション

5.2.1 ホットレプリケーション手法

ホットレプリケーションとは，予めテープ上に確保しておいた領域へ高アクセス頻度データ（ホットデータ）の複製を作成してクラスタリングを行う手法である．ホットレプリケーションは，オリジナルデータ記録時にテープ全体にオリジナルデータを記録せず，高アクセス頻度データの複製のための領域を確保する．各データに対してある程度のアクセスが行われ各データのアクセス頻度が判明した時点で，当該領域に高アクセス頻度データの複製を作成し，クラスタリングする（図5.1）．多くの商用テープドライブ装置ではテープ上の既存のデータを破壊せずに新たなデータを書き込むためには新たに書き込むデータを既存データの後に追記する以外の方法がないため，複製のための領域をテープの終端部に確保する．

予め確保しておいた領域にホットデータの複製がクラスタリングされることにより，ホットデータが連続してアクセスされる場合のシーク長が短縮され，応答時間が短縮される．また，データが複製を持たない場合，アクセス要求されたデータが記録されているテープが使用されているときにはそのリクエストを直ちにサービスすることはできないが，複製を作成することによりオリジナルデータにアクセスできない場合においても複製をアクセスすることが可能となるため，応答時間が短縮される．すなわち，ホットレプリケーションでは，ホットデータのクラスタリングによるシーク長の短縮，および複製の作成によるアクセシビリティの向上により，応答時間の短縮が期待される．

また，ホットレプリケーションでは少数のホットデータのみを複製の作成対象とする．全データの複製を作成すれば，データのアクセシビリティは更に向上すると考えられるが，それだけオリジナルデータの読み込み，複製の書き込みにより長い時間を要し，複製を格納するための空間もより多く必要となる．また，アクセス頻度の低いデータの複製を作成してもそのデータがアクセスされることは少なく，応答時間短縮の効果は小さいことが予想される．

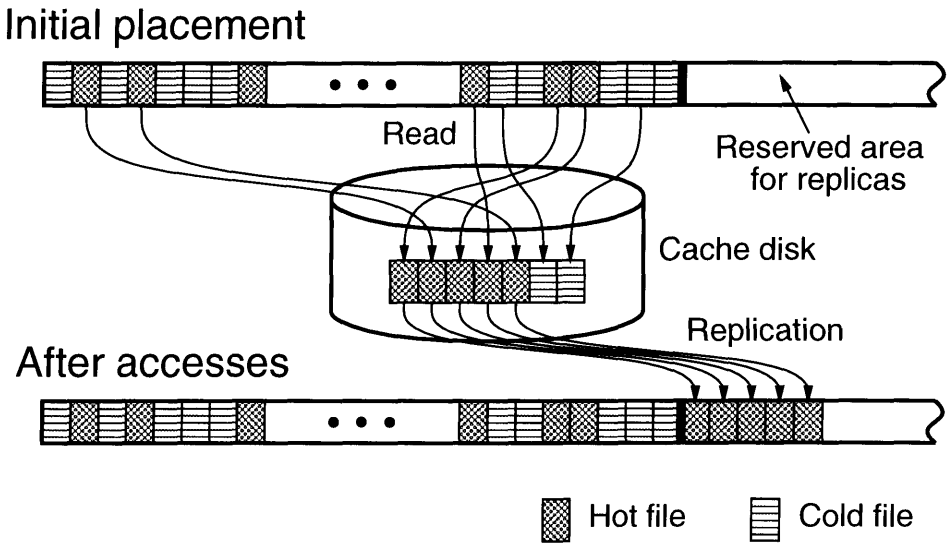

図 5.1: テープ途中でロード/イジェクト可能なテープドライブ装置によるホットレプリケーション

さらに，一般にテープドライブ装置ではテープのイジェクト/ロード，シークに要する時間的コストが大きいため，ホットレプリケーションでは，通常のアクセスリクエストへの影響を最小限に抑えるべく，複製作成時には適宜，以下の方針を採用する．

- キャッシュディスク上に存在するホットデータを複製する．
- 使用されていないテープドライブ装置を用い，そのドライブ内に存在するテープに複製を作成する．
- 上記方針によって複製を作成することができない場合には，新たなデータの読み込みやテープのマウントは行わず，複製の作成そのものを行わない．

キャッシュディスク上のホットデータのみを複製の作成対象とすることにより，複製の作成対象となるデータを新たに読み込む必要がなく，また，既にテープドライブ内に存在するテープを用いることにより，新たにテープをロードする必要がなくなるため，複製の作成に要するコストを最小限に抑えることができる．

5.2.2 テープ途中でロード/イジェクト可能なテープドライブ装置

一般的な商用テープドライブ装置では，テープイジェクト時には先頭まで巻戻す必要があり，テープをマウントした直後はテープ後方に存在するデータへのアクセス時には長いシークを必要とする．すなわち，テープ終端部に確保された領域のホットデータの複製へのアクセスは時間的なコストの面で不利となってしまう．ホットレプリケーションでは，複製へのアクセスが不利となることがないようにするため，テープを巻戻すことなくテープの途中でロード/イジェクトすることが可能なテープドライブ装置，テープメディアを用いる．

多くの商用テープドライブ装置では，テープメディア先頭にディレクトリ情報やメディアに関する情報などを記録しており，ロード時にはこの情報を読み込むようになっている．このため，イジェクト時にはテープを巻戻すようになっており，少数のテープドライブ装置によって多数のテープを扱うテープアーカイブシステムでは，テープの交換を繰り返すためにロード/イジェクト時のシークに多くの時間が費やされる．このロード/イジェクト時のシークを削減するために，テープ先頭のみではなくテープ上の複数の位置，あるいは任意の位置にディレクトリ情報やメディア情報などを記録できるようにしたり，テープカセットに不揮発性メモリ素子を設け，これにディレクトリ情報やメディア情報などを記録することで，テープを巻戻すことなく，ヘッドがテープの途中にあるときでもロード/イジェクトが可能であるドライブ装置が開発され，商用化されている．例えば，AMPEX 社製 DST312 や SONY 社製 GY-2120[12] ではテープ先頭以外の位置にディレクトリ情報やメディア情報などを記録できるようにすることにより，また，SONY 社製 AIT-S100[11] では，テープカセットに 64KB の EEPROM を搭載することにより，イジェクト/ロード時にテープを完全に巻戻す必要がなく，ヘッドがテープの先頭以外の位置に存在するときでもロード/イジェクトが可能となっている．

テープを巻戻すことなくロード/イジェクト可能なテープドライブ装置は既に商用のものとなっており，アーカイブシステムでの利用において機能面，性能面で大きな利点を持つため，今後の普及が期待される．

5.2.3 ホットレプリケーションによるシーク長の短縮

本節ではホットレプリケーションによるシーク長の短縮効果を解析的に求める．解析結果が複雑になるのを避けるため，全体のp（$p < 0.5$）の割合を占め，$1-p$の割合のリクエストを受ける高アクセス頻度データと，$1-p$の割合を占め，pの割合のリクエストを受ける低アクセス頻度データの2種類のデータ構成され，それらがテープ上にランダムに配置されている場合の平均シーク長を求める．リクエストは，発行された順に1つずつサービスされるものとし，複製が存在するデータに対するリクエストに対しては常に複製がアクセスされるものとする．また，複製用領域よりもホットデータの量が多い場合には，複製用領域に収まる分の複製を作成するものとしている．また，複製を作製せずにオリジナルデータをクラスタリングした場合に関して，全ホットデータをテープ中央に配置し，両側に均等にコールドデータを配置した場合，および端部に全ホットデータを配置した場合の相対シーク長も求める．

まず，テープ上の長さlの領域$[0,l]$のある任意の点xから他の任意の点yまでの平均シーク長$\lambda_1(l)$を求める．$\lambda_1(l)$は，

$$\begin{aligned} \lambda_1(l) &= \int_0^l \int_0^l \frac{|x-y|}{l^2} dxdy \\ &= \frac{l}{3} \end{aligned} \tag{5.1}$$

となる．また，$[0,l_1]$内の任意の位置xから$[l_1+l_2, l_1+l_2+l_3]$内の任意の位置yへの平均シーク長を$\lambda_2(l_1,l_2,l_3)$とすると，

$$\begin{aligned} \lambda_2(l_1,l_2,l_3) &= \int_0^{l_1} \int_{l_1+l_2}^{l_1+l_2+l_3} \frac{y-x}{l_1 l_3} dydx \\ &= \frac{l_1+2l_2+l_3}{2} \end{aligned} \tag{5.2}$$

となる．

複製が存在しない場合

テープ長Lに対し，ϕの割合の高アクセス頻度データ用領域を確保した場合，オリジナルデータ用領域のテープ長は$(1-\phi)L$となる．従って，オリジナル領域全てにデータが

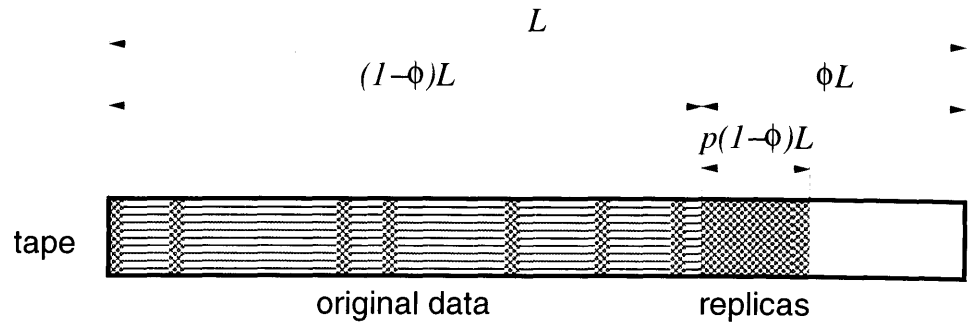

図 5.2: ホットレプリケーションにおけるデータ配置

記録され，複製が存在しない場合の平均シーク長は

$$\lambda_1((1-\phi)L) = \frac{(1-\phi)L}{3} \tag{5.3}$$

となる．

複製が存在する場合

全テープ長 L に対し，オリジナルデータ用領域全てにデータを記録した場合，$(1-\phi)L$ に相当するデータ量になり，高アクセス頻度データは，$p(1-\phi)L$ に相当するデータ量になる（図 5.2）．

高アクセス頻度データが複製用領域に全て複製される場合（$\phi \geq p(1-\phi)$）　複製容量域が十分に存在し，全ての高アクセス頻度データが複製されている場合，複製用領域へのアクセス確率 P_r は $P_r = 1-p$，オリジナルデータ用領域へのアクセス確率 P_o は $P_o = p$ となる．従って，平均シーク長は

$$\begin{aligned} & P_o^2\lambda_1((1-\phi)L) + 2P_oP_r\lambda_2((1-\phi)L, 0, p(1-\phi)L) + P_r^2\lambda_1(p(1-\phi)L) \\ = \ & (-2p^3 - p^2 + 4p)\frac{(1-\phi)L}{3} \end{aligned} \tag{5.4}$$

となる．

複製用領域より高アクセス頻度データが多く存在する場合（$\phi < p(1-\phi)$）　複製領域以上に高アクセス頻度データが存在し，全ての高アクセス頻度データが複製できない場合を

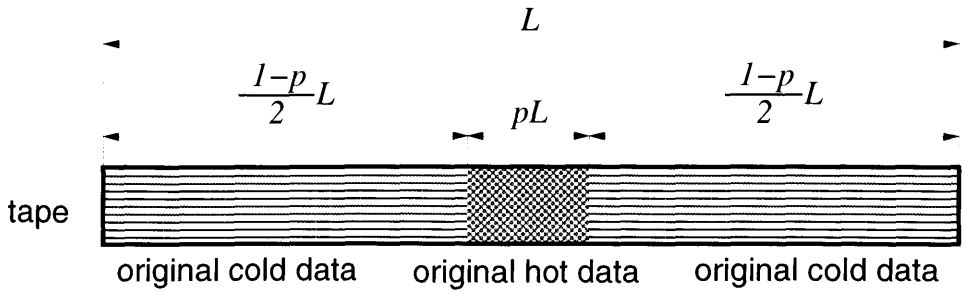

図 5.3: 全ホットデータをテープ中央に配置した場合

考える．複製用領域に可能な限り高アクセス頻度データの複製が作成され，複製が作成されなかった高アクセス頻度データについては，オリジナル領域のデータがアクセスされるものとする．このとき，複製用領域へのアクセス確率は $P_r = (1-p)\frac{\phi}{p(1-\phi)}$，オリジナルデータ用領域へのアクセス確率は $P_o = p + (1-p)\{1 - \frac{\phi}{p(1-\phi)}\}$ である．従って，平均シーク長は，

$$\begin{aligned} & P_o^2\lambda_1((1-\phi)L) + 2P_oP_r\lambda_2((1-\phi)L, 0, p(1-\phi)L) + P_r^2\lambda_1(p(1-\phi)L) \\ = \; & \{\phi^3 - 4\phi + 1 + (-2\phi^3 + 5\phi^2 + \phi)\frac{1}{p} - 2\phi^2\frac{1}{p^2}\}\frac{L}{3(1-\phi)^2} \end{aligned} \tag{5.5}$$

となる．

複製を作製せず，全ホットデータをテープ中央に配置し，両側に均等に低アクセス頻度データを配置した場合（図 5.3）

高アクセス頻度データへのアクセス確率 P_h は $P_h = 1-p$，低アクセス頻度データへのアクセス確率 P_c は p である．従って，全データ量が L に相当する場合，平均シーク長は

$$\begin{aligned} & P_h^2\lambda_1(pL) + 2(\frac{P_c}{2})^2\lambda_1(\frac{1-p}{2}L) \\ & +4P_h\frac{P_c}{2}\lambda_2(pL, 0, \frac{1-p}{2}L) + 2(\frac{P_c}{2})^2\lambda_2(\frac{1-p}{2}L, pL, \frac{1-p}{2}L) \\ = \; & (5p - 2p^2)\frac{L}{6} \end{aligned} \tag{5.6}$$

となる．

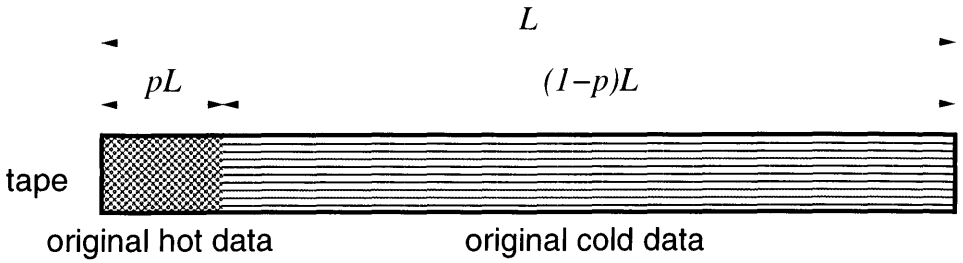

図 5.4: 全ホットデータをテープ端部に配置した場合

複製を作製せず，端部に全高アクセス頻度データを配置した場合（図 5.4）

高アクセス頻度データへのアクセス確率は $P_h = 1 - p$，低アクセス頻度データへのアクセス確率は p である．テープ先頭に全高アクセス頻度データを配置し，その後に低アクセス頻度データを配置した場合の平均シーク長は

$$\begin{aligned} & P_h^2\lambda_1(pL) + 2P_hP_c\lambda_2(pL, 0, (1-p)L) + P_c^2\lambda_1((1-p)L) \\ = \ & (4p - 4p^2)\frac{L}{3} \end{aligned} \tag{5.7}$$

となる．

ホットレプリケーションによるシーク長の短縮効果

図 5.5 は，複製が存在せず，複製用の領域を除くオリジナルデータ用の領域上にランダムにデータを配置した場合の平均シーク長を 1 としたときの，複製が存在する場合の相対シーク長を示している．オリジナル領域が全体に占める割合 $1 - \phi$ は，テープ 1 本に記録できるオリジナルデータ量の割合であり，従って，単位データあたりに要するメディアのコストは $\frac{1}{1-\phi}$ となる．一方，複製が存在しない場合の平均シーク長は $1 - \phi$ と比例関係にある．従って，複製が存在しない場合の平均シーク長を 1 とした図 5.5 は，メディアコストで正規化したシーク長の短縮効果と言うこともできる．

図 5.5 によると，複製用の領域が小さく，その領域以上にホットデータが存在する場合，すなわちすべてのホットデータの複製を作成できない場合には，平均シーク長は急激に長くなる．クラスタリングの効果を保つためには，ホットデータの量よりも複製用の領域を大きくとる必要がある．一方で，アクセス頻度が 70/30 則（全体の 70%のデータが全リ

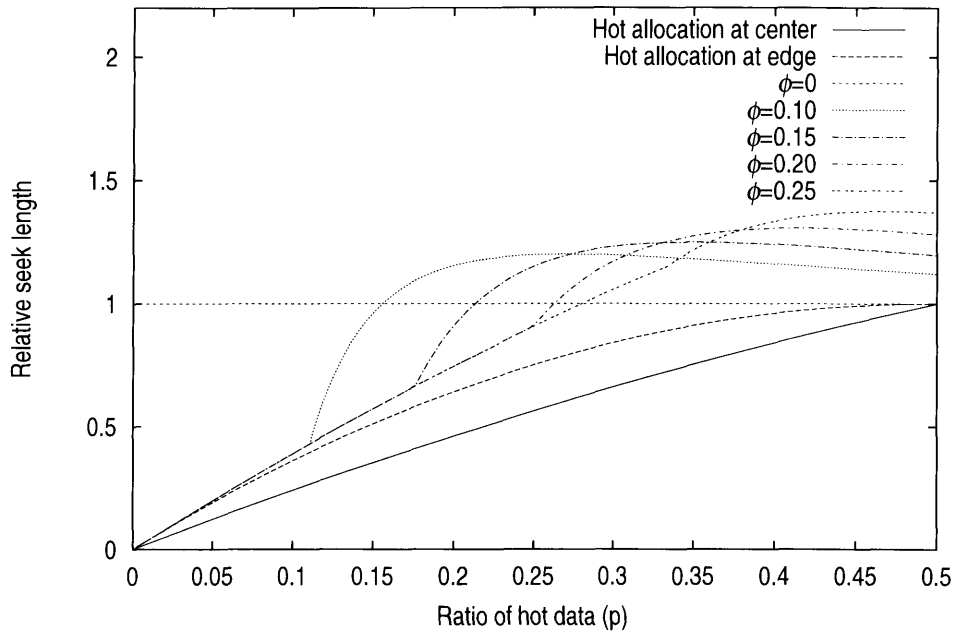

図 5.5: 高アクセス頻度データをテープ終端部に複製したホットレプリケーションにおけるシーク短縮効果

クエストの 30%を受ける）よりもアクセスの偏りが緩やかになると，たとえ複製用領域が十分でも，ホットデータのクラスタリングの効果は薄れ，コールドデータをアクセスするためにオリジナルデータ領域へシークする回数が増えるために平均シーク長は延びる．すなわち，シーク短縮の観点のみから言えば，70/30 則よりも緩い分布をなすデータのために，ホットデータのすべての複製を作成できるように複製用領域をテープ全体の 20~25%以上としても，その効果は望めないことがわかる．複製を作製せずにオリジナルデータをクラスタリングした場合には，複製をクラスタリングした場合と比べ，よりいっそう平均シーク長を短縮することが可能である．しかしながら，オリジナルデータをクラスタリングするためには，一度テープ上の全データをディスク上へ読み込み，その後，目的の配置となるようにデータをテープへ書き戻す必要がある．すなわち，複製を用いずにオリジナルデータをクラスタリングするためには，今日の典型的なテープ装置ではテープ 1 本あたり数時間を要することとなる．一方，ホットデクラスタリングでは少数のホットデータのみをテープへ書き込むだけであるため，複製を作製せずにクラスタリングする場合に比べて，書き込みに要する時間は極めて小さい．また，ホットレプリケーションは，

表5.1: シミュレーションパラメータ

エレメントアーカイバ	
全エレメントアーカイバ数	16台
最大テープ数	200本/台
テープドライブ数	2台/台
テープドライブ	
ロード時間	35秒
シーク速度	25MB/秒
リード/ライト速度	0.5MB/秒
イジェクト時間	20秒
テープハンドラロボット	
移動時間（テープの操作なし）	2秒
移動時間（テープの操作あり）	14秒
テープマイグレーション装置	
ワゴンの移動時間	9秒

複製のための領域を必要とするが，今日の一般的なテープメディアの容量当りのコストは小さく，従って複製のための領域に要するコストはわずかである．ホットレプリケーションはアクセスの偏りが大きい場合にはクラスタリングを行わない場合と比べ充分にシーク長を短縮することが可能であり，現実的な手法であると言える．

5.3 基本性能評価

5.3.1 シミュレーション条件

本節では，ホットレプリケーションの基本性能をシミュレーションにより評価する．16台のエレメントアーカイバにより構成されるスケーラブルテープアーカイバにおいてホットデクラスタリング（テープマイグレーション）を導入した場合と導入しない場合のそれぞれについて，テープの途中でロード/イジェクトが可能なテープドライブ装置を用いてホットデータの複製をテープ終端部に作成した場合のシミュレーションを行う．スケーラブルテープアーカイバのパラメータを表5.1に示す．各テープの容量は7GBとし，先頭より5.5GBまでをオリジナルデータ用，終端部の1.5GBの領域をホットデータの複製用領

域とする．リクエスト到着間隔は負の指数分布に従う．リクエストサービススケジューリングに関しては，ホットデータのクラスタリングによるシーク長の短縮の効果と複製によるアクセシビリティの向上による効果の両者が得られるよう，次のスケジューリングを採用する．

1. リクエストキュー内のリクエストを発行順にソートする．

2. 複製が存在するデータに関しては，複製が優先的にアクセスされるようにするため，まずクラスタリングされた複製のみが存在すると仮定し，リクエストキュー内のサービス可能なもののうちで最も先に発行されたリクエストを選択し，サービスを行うテープを決定する．

3. 2において，リクエストされているテープが使用されている，あるいはリクエストされているテープをサービスするためのドライブ装置が，他のリクエストのサービスのために使用されているためにサービス可能なリクエストがない場合には，複製の存在するデータに関し，複製だけではなくオリジナルデータが存在するものとしてサービス可能なリクエストを選択し，サービスを行うテープを決定する．

4. 選択されたテープ上に記録されたデータに対するすべてのリクエストをテープの先頭方向から順にまとめてサービスする．

ホットデクラスタリングは，リクエストされたテープが存在するエレメントアーカイバ内のドライブ装置がすべて使用されている場合には，他のエレメントアーカイバへテープを移送してサービスが行われる．従って，ホットデクラスタリングを採用している場合には，上記スケジューリング3が応答性能に与える影響は小さいと考えられる．

なお，C言語の表記に準じた詳細なスケジューリング手順を図5.6に示す．

```
Q = sort_by_time(Q)

D = { }
t = NULL

r = first(Q)
while (r != NULL && t == NULL) {
    if (have_replica(r) &&
        is_accessible(tape_id(original_id(r)))) {
        t = tape_id(original_id(r))
        D = {original_id(r)}
        Q = Q − {r}
    }
    else if (!have_replica(r) &&
            is_accessible(tape_id(replica_id(r)))) {
        t = tape_id(replica_id(r))
        D = {replica_id(r)}
        Q = Q − {r}
    }
    r = next(Q)
}

if (t == NULL) {
    r = first(Q)
    while (r != NULL && t == NULL) {
        if (is_accessible(tape_id(replica_id(r)))) {
            t = tape_id(replica_id(r))
            D = {replica_id(r)}
            Q = Q − {r}
        }
        r = next(Q)
    }
}

if (t != NULL) {
    r = first(Q)
    while (r != NULL) {
        if (have_replica(r) &&
            tape_id(original_id(r))== t) {
            D = D + {original_id(r)}
            Q = Q − {r}
        }
        else if (tape_id(replica_id(r))== t) {
            D = D + {replica_id(r)}
            Q = Q − {r}
        }
        r = next(Q)
    }
}
D = sort_by_position(D )
```

ただし，

Q　リクエストキュー

D　アクセスデータリスト

sort_by_time(**Q**)　**Q** を発行時間順にソートする．

first(**Q**)　**Q** の先頭リクエスト ID を返す．

next(**Q**)　直前に実行された *first*(**Q**) または *next*(**Q**) によって返されたリクエストの次のリクエスト ID を返す．次のリクエストがない場合は NULL を返す．

have_replica(r)　リクエスト r により要求されているデータが複製を持つ場合は真，持たない場合は偽を返す．

replica_id(r)　リクエスト r により要求されているデータのオリジナルのデータ ID を返す．

original_id(r)　リクエスト r により要求されているデータの複製のデータ ID を返す．

tape_id(d)　データ d が記録されているテープ ID を返す．

is_accessible(t)　テープ t がアクセス可能な場合は真，テープドライブ装置が使用されている，テープハンドラロボットが使用されている，テープが他のリクエストにより使用されてるなどの理由によりアクセスできない場合は偽を返す．

sort_by_position(**D**)　**D** をテープ上の位置順にソートする．

図 5.6: スケジューリング手順

5.3.2 シーク時間の短縮による効果

本シミュレーションでは，各データのアクセス頻度分布はZipf分布[2]に従うとし，それらをテープ上のオリジナル領域にランダムに配置した状態をオリジナルデータの初期配置とする．ホットデータの複製もまた，テープ終端部の複製用領域にランダムに配置した状態を初期状態とする．本シミュレーションにおいては，動的な複製の作成は行わない．各データをランダムに配置しているため，各テープのヒートはほぼ均一となり，従って複数のリクエストが1本のテープに集中する可能性は低く，複製によるアクセシビリティの向上による性能向上は期待できず，シーク長の短縮による応答性能の向上の効果が明確になる．本シミュレーションにおいてはアクセス頻度分布がほぼ90/10則に従う（全体の10%のデータが全リクエストの90%を受ける）ようにZipf分布のパラメータzを設定している．

図5.7は，データのアクセス頻度分布が$z = 2.0$のZipf分布に従い，各データのサイズが100MBであるときの，初期状態から50000アクセスまでの平均応答時間である．データ全体の約10%が全リクエストの約90%を受けることとなるため，アクセス頻度の高い順に10%のデータをホットデータとみなし，複製を作成している．図5.7には，ホットデータの複製が存在しない場合，テープ終端部にホットデータの複製が存在する場合，およびテープ終端部にホットデータの複製が存在し，かつホットデータにのみに対してアクセスリクエストが発行される場合の平均応答時間を示している．応答時間はリクエストが発行されてからデータの読み込みが完了するまでの時間である．ホットデータのみに対してリクエストが発行される場合は，ホットレプリケーションによる応答性能向上の限界値に相当する．図5.7より，テープ終端部にホットデータの複製を作成することにより，平均応答時間が短縮されていることがわかる．ホットデクラスタリングを用いている場合には，用いていない場合に比べて平均応答時間の短縮効果は小さい．これは，複製が応答時間

[2]全要素数をNとしたとき，i番目の要素の確率p_iが

$$p_i = \frac{1}{\sum_{n=1}^{N} \frac{1}{n^z}} \cdot \frac{1}{i^z}$$

により表される分布．Zipf分布において，$N = 167200$の場合には$z \approx 2.0$，$N = 1672000$の場合には$z \approx 1.4$とすることにより90/10則に従う分布が得られる．

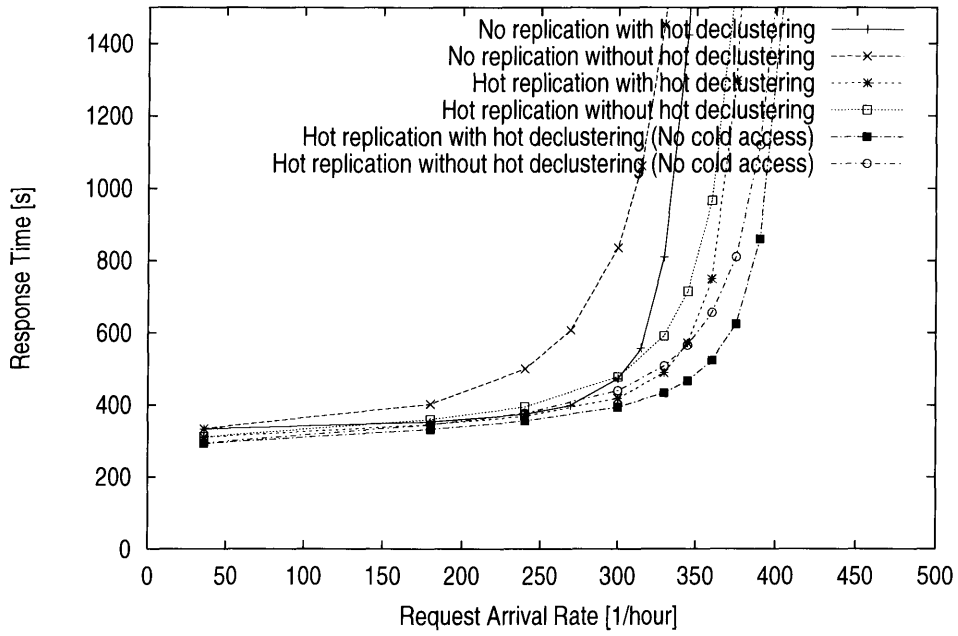

図 5.7: ホットレプリケーションによる応答時間（ファイルサイズ 100MB）

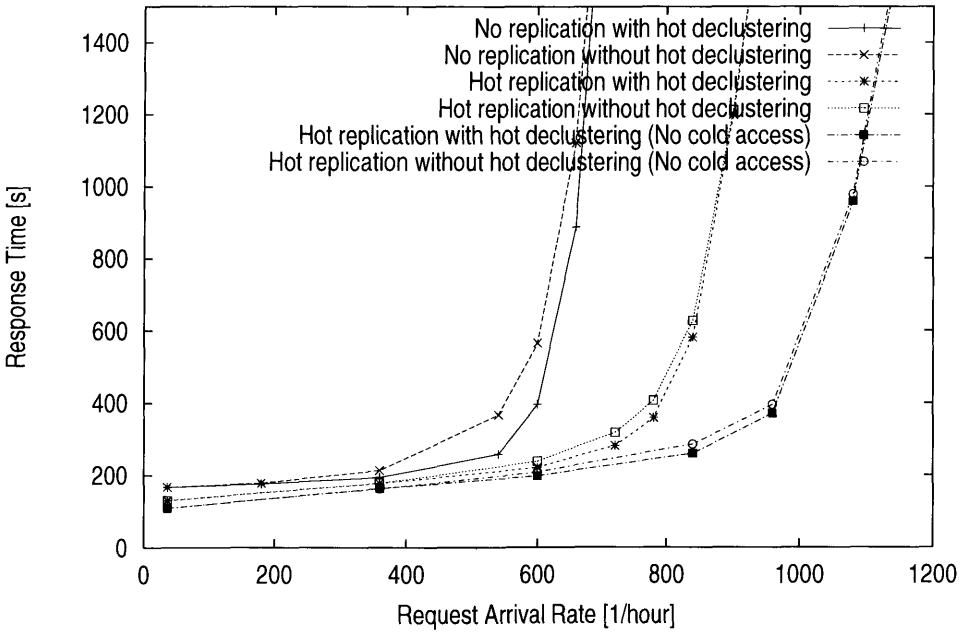

図 5.8: ホットレプリケーションによる応答時間（ファイルサイズ 10MB）

の短縮効果を示す状況とホットデクラスタリングが効果を示す状況が一部重なっており，ホットデクラスタリングが用いられていない場合において，ホットデータの複製のクラスタリングにより効果が得られる状況の一部が，ホットでクラスタリングによって解消されているためである．例えば，あるテープに対してリクエストが発行されたとき，そのテープが存在するエレメントアーカイバ内のテープドライブ装置がすべて使用されていることによるリクエストのブロックは，ホットデクラスタリングにより隣接するエレメントアーカイバへそのテープを移送することで解消することも，異なるエレメントアーカイバ内のテープ上の複製をアクセスすることで解消することもできる．しかしながら，ホットデクラスタリングはコールドデータへのアクセスリクエストに対しても適用可能であり，一方，ホットレプリケーションは同一テープ上の異なるホットデータへのリクエストを同時にサービス可能であるなど，ホットレプリケーションとホットデクラスタリングの適用範囲は異なる部分もあり，そのために同時に用いた場合には，わずかではあるが応答性能が向上している．

図5.8は，データのアクセス頻度分布が$z = 1.4$のZipf分布に従い，ファイルサイズが10MBの場合の初期状態から50000アクセスまでの平均応答時間である．データ全体の約10%が全リクエストの約90%を受けることとなるため，アクセス頻度の高い順に10%のデータをホットデータとみなし，複製を作成している．ファイルサイズが100MBの場合と比較し，ホットデータの複製の作成がより効果的であることがわかる．これは，ファイルサイズが小さくなったためにデータの読み込み時間が短縮され，そのためリクエストが発行されてから読み込みが終了するまでの全応答時間に対してシーク時間の占める割合が大きくなったことによる．ホットレプリケーションは，ホットデータの複製をクラスタリングすることでシーク時間が短縮されるため，ファイルサイズが小さくなるに従いその有効性が向上する．

5.3.3 複製によるアクセシビリティの向上による効果

本節では，各テープに対するアクセス頻度がZipf分布に従い，更に各テープ内のデータのアクセス頻度分布もまたZipf分布に従うとしてデータをランダムに配置したものをオリジナルデータの初期配置とした場合，および全体としては同じアクセス頻度分布を持

たせたデータを全テープ上にランダムに配置したものをオリジナルデータの初期配置とした場合のシミュレーションを行う．両者の違いはテープ間のアクセス頻度の偏りの有無であり，テープ間にアクセス頻度の偏りがある場合には少数のテープにアクセスが集中することとなるため，これらの結果を比較することで複製によるアクセシビリティの向上による効果を見ることができる．本シミュレーションにおいてもアクセス頻度分布が90/10則に従うようにZipf分布のパラメータzを設定している．

図5.9は各テープに対するアクセス頻度が$z = 1.15$のZipf分布に従い，更に各テープ上のデータそれぞれに対するアクセス頻度もまた，$z = 1.15$のZipf分布に従うとした場合の初期状態から50000アクセスまでの平均応答時間である．また，図5.10は，データ全体としては図5.9のデータと同じ分布を持たせ，それらを全テープ上にランダムに配置した場合の初期状態から50000アクセスまでの平均応答時間である．各データのサイズは100MBであり，データ全体の約10%が全リクエストの約90%を受けることとなるため，アクセス頻度の高い順に10%のデータをホットデータとしている．また，複製は全テープ上の複製用領域にランダムに配置したものを初期状態とし，動的な複製の作成は行っていない．

テープ間のアクセス頻度の偏りがある場合には，テープ間のアクセス頻度の偏りがない場合に比べ，ホットデクラスタリング，ホットレプリケーションのいずれも用いないと大幅に応答性能は劣化している．これは，少数のテープにリクエストが集中してしまうため，あるリクエストによってアクセス要求されたテープが別のリクエストによって使用されているためにブロックされてしまったり，あるいはリクエストされたテープが存在するエレメントアーカイバ内のテープドライブが別のリクエストにより常に使用中となってしまうためにサービスを行うことができず，テープドライブを効率的に使用することができないためである．ホットデクラスタリングは，このような場合，空いているテープドライブが存在するエレメントアーカイバへテープをマイグレートし，リクエストをサービスすることで応答時間の劣化を抑えるが，ホットレプリケーションでは，オリジナルデータと複製の2つが存在するため，どちらか一方の存在するテープが使用されていたり，そのテープが存在するエレメントアーカイバのテープドライブが使用されていても，他の異なるエレメントアーカイバ内に存在するもう一方のデータへアクセスすることによりテー

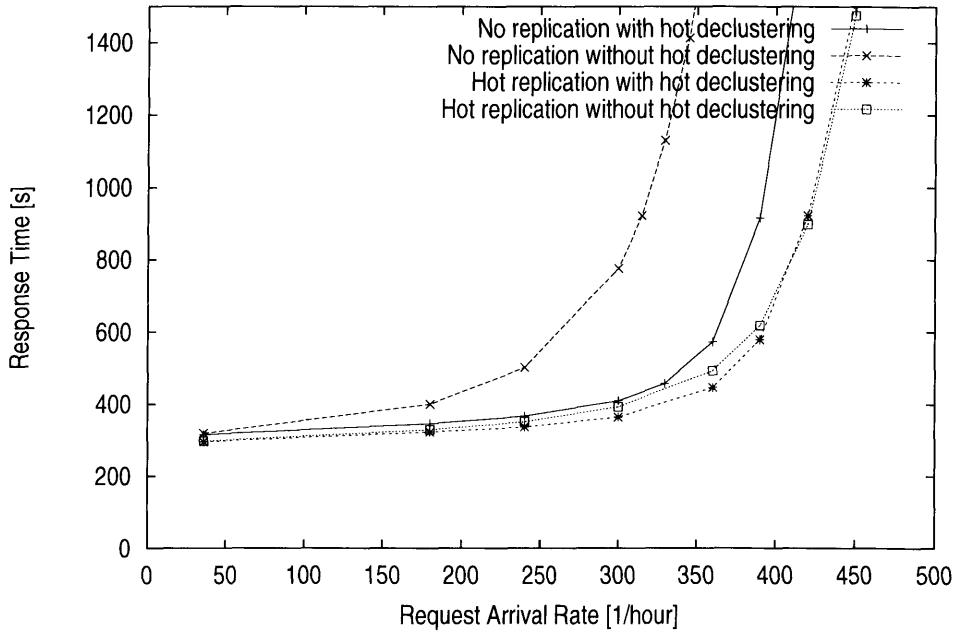

図 5.9: ホットレプリケーションによる応答時間（テープ間のアクセス頻度の偏りあり，ファイルサイズ 100MB）

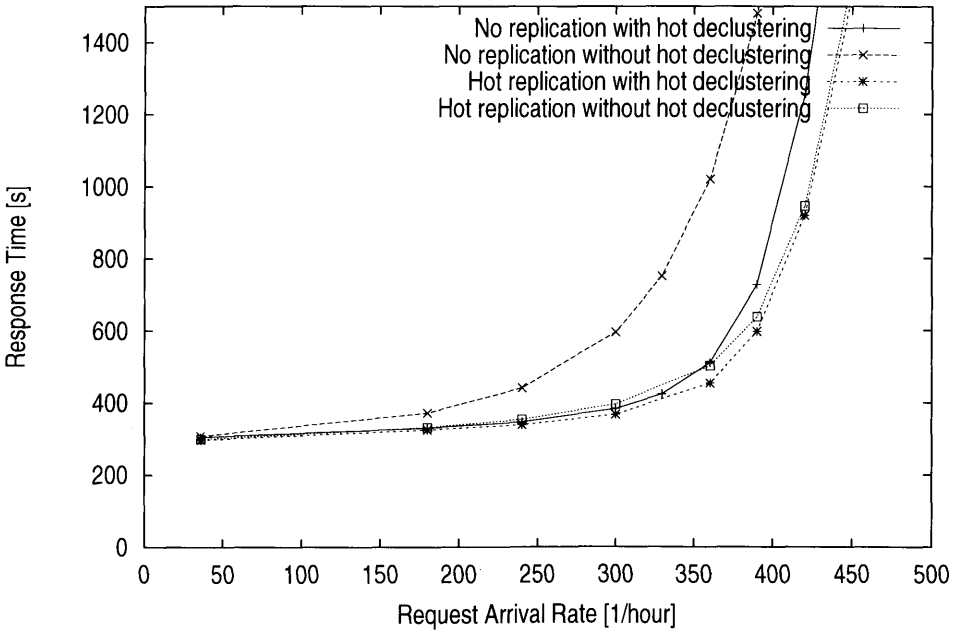

図 5.10: ホットレプリケーションによる応答時間（テープ間のアクセス頻度の偏りなし，ファイルサイズ 100MB）

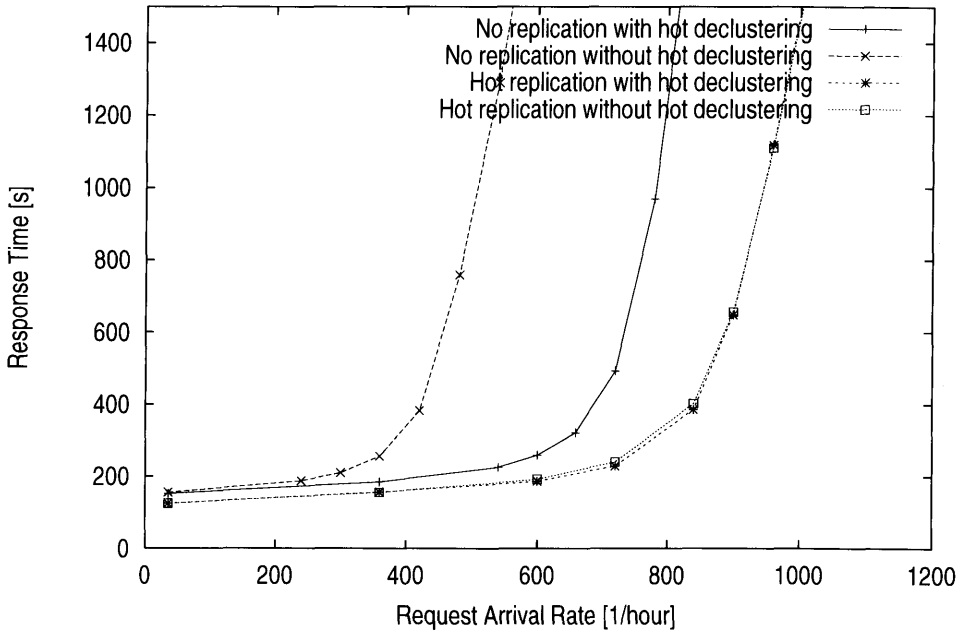

図 5.11: ホットレプリケーションによる応答時間（テープ間のアクセス頻度の偏りあり，ファイルサイズ 10MB）

プドライブを効率的に使用し，応答性能の劣化を防ぐ．加えて，ホットレプリケーションでは，シーク時間の短縮の効果もあり，ホットデクラスタリングに対してより有意な応答性能の向上が得られる．

図 5.11 は各テープに対するアクセス頻度が $z = 1.05$ の Zipf 分布に従い，更に各テープ上のデータそれぞれに対するアクセス頻度もまた，$z = 1.05$ の Zipf 分布に従うとした場合の初期状態から 50000 アクセスまでの平均応答時間である．また，図 5.12 は，データ全体としては図 5.11 のデータと同じ分布を持たせ，それらを全テープ上にランダムに配置した場合の初期状態から 50000 アクセスまでの平均応答時間である．各データのサイズは 10MB であり，データ全体の約 10% が全リクエストの約 90%を受けることとなるため，アクセス頻度の高い順に 10%のデータをホットデータとしている．

図 5.9，図 5.10 に示したファイルサイズが 100MB の場合と比較し，図 5.11，図 5.12 に示すファイルサイズが 10MB の場合には，ホットレプリケーション，ホットデクラスタリングの両者とも用いないときには，応答性能の悪化がより顕著である．これは，ファイルサイズが小さくなったためにデータの読み込み時間が短縮され，ホットレプリケーション

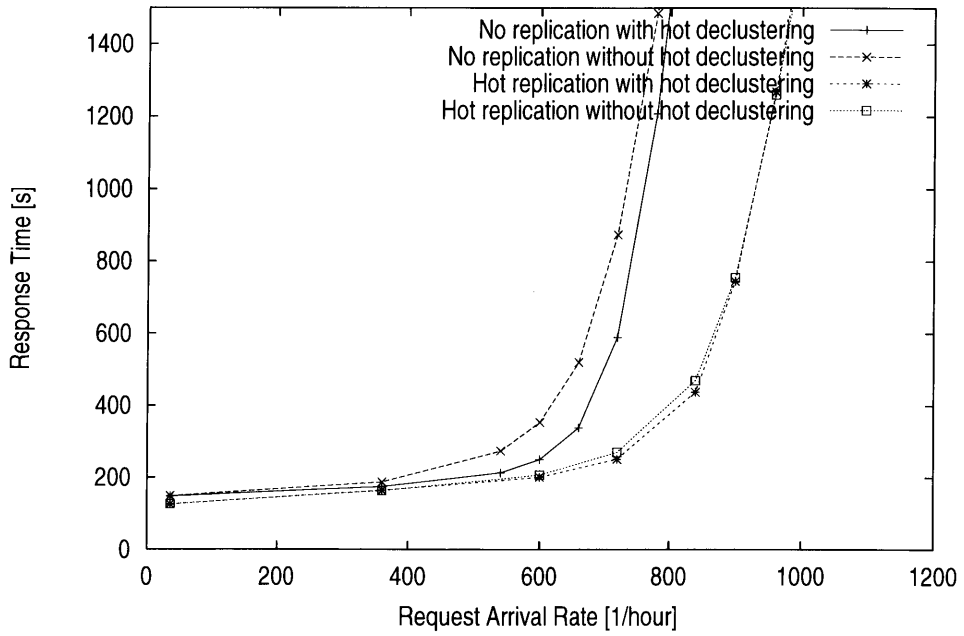

図 5.12: ホットレプリケーションによる応答時間（テープ間のアクセス頻度の偏りなし，ファイルサイズ 10MB）

の効果が際立たせられたためである．アクセシビリティの向上による応答時間短縮効果においても，ファイルサイズが小さくなるに従って，ホットレプリケーションの有効性は向上する．

5.4 まとめ

本章では，高アクセス頻度データの複製を作成し，それらを予め確保しておいたテープ上の領域にクラスタリングするホットレプリケーションの提案を行い，シミュレーションによりその有効性を示した．ホットレプリケーションは，複製によるアクセシビリティの向上，および複製のクラスタリングによるシーク長の短縮により応答性能を向上させる．

スケーラブルテープアーカイバ環境を想定したシミュレーションにおいて，ランダムに配置したデータを用いてシーク時間の短縮による効果を示すとともに，テープ間のアクセス頻度に偏りがある場合とない場合の比較を行うことで複製によるアクセシビリティの効果を示し，基本性能を明らかにした．

第6章　衛星画像データベースシステムへのアクセス履歴を用いた評価

6.1　はじめに

本章では，東京大学生産技術研究所においてWorld Wide Web，gopher，ftpを通じて公開している衛星画像データベースのアクセス履歴を用い，ホットデクラスタリングおよびホットレプリケーションの性能を評価する．以降，まず6.2節において，衛星画像データベースのアクセス履歴についての説明を行い，6.3節，6.4節において，衛星画像データベースのアクセス履歴を用いたシミュレーションを実行し，それぞれホットデクラスタリング，ホットレプリケーションの性能評価を行い，ホットデクラスタリング，ホットレプリケーションが実システムにおいても有効であることを明らかにする．

6.2　アクセス履歴

図6.1は，1996年4月から1998年10月半ばまでの衛星画像データベースシステム上のクイックルック画像に対するアクセスの分布である．リクエスト数はftpによるアクセス約49000件，gopherによるアクセス約215000件，WWWによるアクセス約196000件の合計460000件である．横軸は1996年4月1日からの経過日数，縦軸はデータ番号を表し，グラフの各点は当該データに対し当該日にアクセスがあったことを示している．データ番号はデータ全体をNOAA，GMSに分割し，それぞれ古いデータから順に番号を付け，0~29799がNOAAデータ，29800~58636がGMSデータである．図6.1において，NOAA，GMSとも最上部に斜めの線状の分布が存在しており，受信されたばかりの最新の画像にアクセスが集中している傾向が伺える．また，縦方向の線状の分布が見られ，短期間に多くのデータがまとめてアクセスされたことがわかる．これは，図6.1の縦方向の線状の分

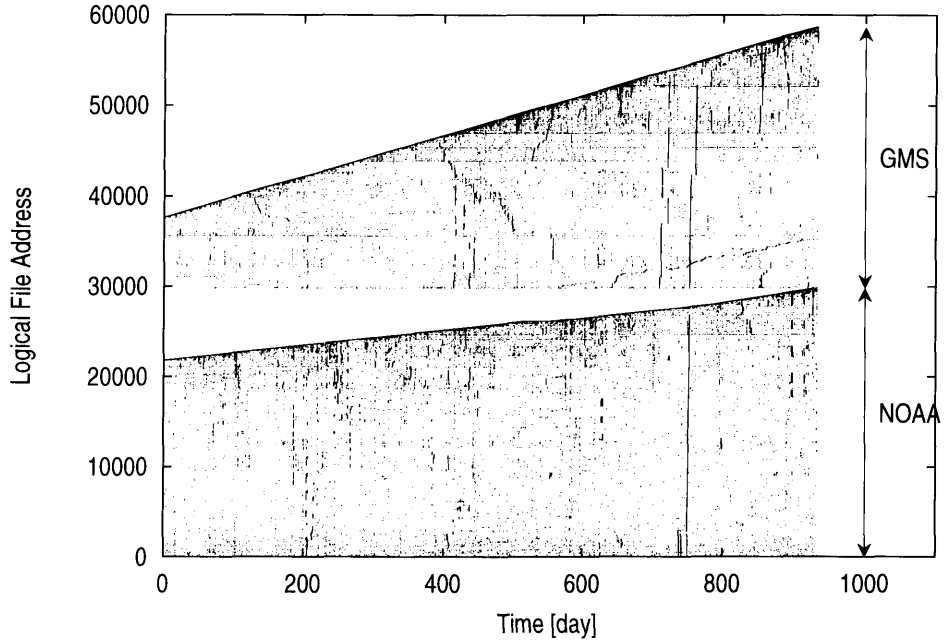

図 6.1: リクエスト分布

布に対応しており，特定のユーザがある一定期間の画像データを一括して転送したことによるものである．

図 6.2 は 1 日ごとのリクエスト件数を示している．1 日に 10000 件以上のリクエストを受けている日もあるが，これらは特定の利用者が一定期間の画像データにまとめてアクセスをしたことによるものが多い．

図 6.3 は，衛星画像データベースにおいてアクセス頻度上位のデータに対するリクエストが全リクエストに対して占める割合を表したグラフである．比較のために論文 [6] において示された 70/30 則に従う曲線[1]も示している．衛星画像データベースに対するリクエストは全体の 30%のデータが 70%のリクエストを受けているが，論文 [6] による分布とはやや異なり，アクセス頻度が極端に高いデータはなく，より緩やかな分布をしていることがわかる．

[1]Fraction of Total Requests = Fraction of Total Data$^{\frac{\log\beta}{\log\alpha}}$ によって表され，点 $(0,0),(\alpha,\beta),(1,1)$ を通り，区間 $(0,1)$ において連続かつ微分可能．Zipf 分布の要素数 $\to\infty$ とした場合の極限である．

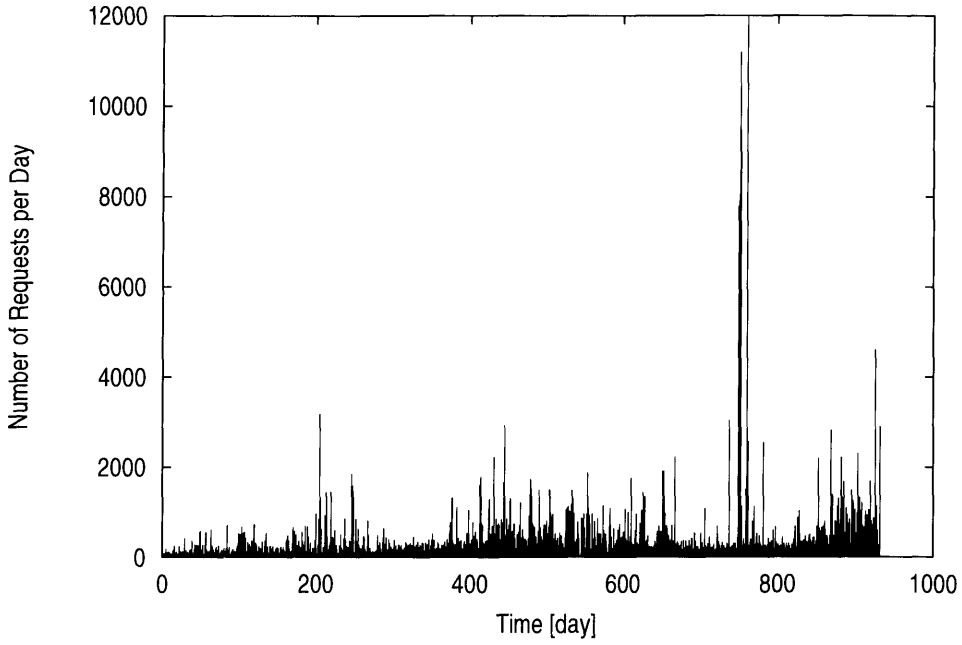

図 6.2: 1 日毎のリクエスト数

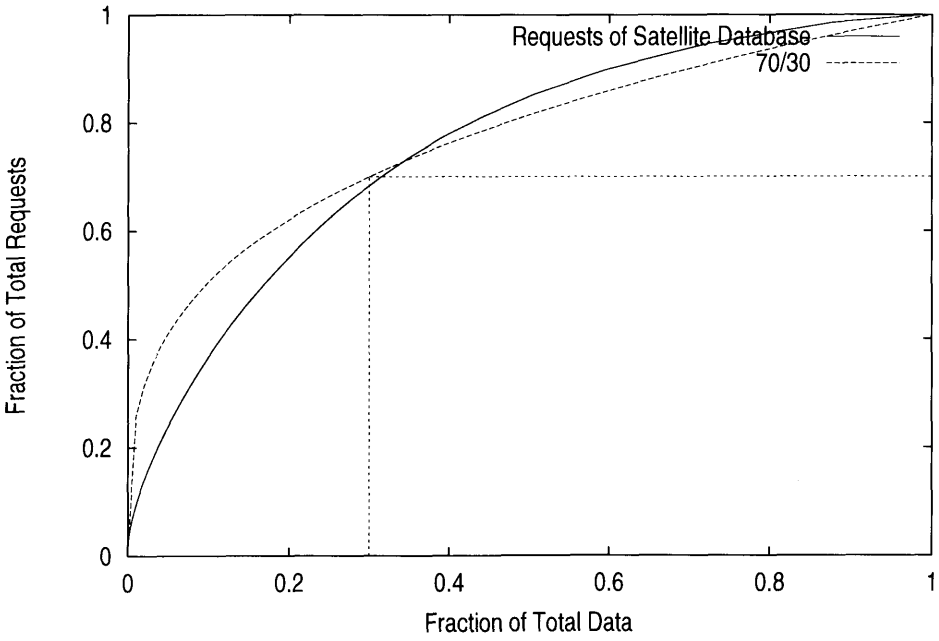

図 6.3: アクセスローカリティ

6.3 ホットデクラスタリングの評価

6.2節で示したクイックルック画像に対するアクセス履歴を用い，これらのリクエストが対応する衛星原画像へのリクエストであると仮定してシミュレーションを実行し，スケーラブルテープアーカイバにおけるホットデクラスタリングの効果を評価する．クイックルック画像のアクセス分布と原画像へのアクセス分布とは必ずしも一致しないが，原画像とクイックルック画像は一対一に対応しており，また，最新画像に対するアクセスが多い，特定ユーザが短期間に一括してデータにアクセスしている，などの特徴は原画像に対するアクセスにも共通すると考えられることからこのデータを用いて評価することとした．

シミュレーションでは，WWW，gopher，ftpによるクイックルック画像へのアクセスを衛星原画像に対する読み出しリクエストと仮定した460000件のアクセスの他に，新たに受信されたデータの書き込みリクエスト約29000件を加えた計489000件を用いる．シミュレーションは489000リクエスト終了時まで実行するが，性能評価には初期状態から450000までのリクエストを使用する．クイックルック画像に対するアクセス系列を原画像に対するアクセス系列とするにはリクエスト間隔が短いため，各リクエスト間の時間を均等に延ばすことで時間の経過を減速し，リクエスト到着率を下げた場合のシミュレーションも実行した．初期状態における各テープのデータ配置，圧縮の有無は，実際の衛星画像データベースシステムの8mmテープアーカイバにおけるデータ配置に基づいており，データはNOAA，GMSそれぞれ時間順に記録されている．新たに書き込まれるデータも実際のデータ配置に基づき記録されるものとした．NOAAデータは，初期のデータはテープ長112m（非圧縮時容量5GB）の548本の8mmテープに非圧縮で記録され，後期のデータはテープ長160m（非圧縮時容量7GB）の21本の8mmテープに圧縮されて記録される．GMSデータはすべてテープ長160m（非圧縮時容量7GB）の94本の8mmテープに圧縮されて記録される．データ圧縮は8mmテープドライブに備わっているハードウェア圧縮機能を用いており，個々のデータに対する圧縮率を得ることは困難であるため，非圧縮時のテープ容量と圧縮機能を用いて実際に記録されたデータ量より求めた各データの圧縮率の平均値に基づき，NOAAデータは一律に67%，GMSデータは一律20%に圧縮されて

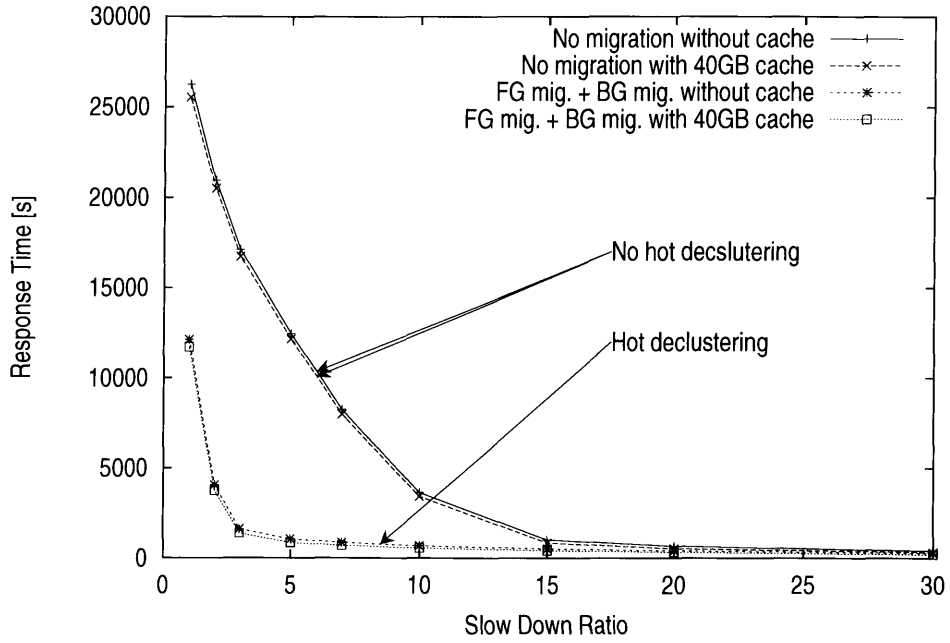

図 6.4: 平均応答時間

いるものとした．即ち，例えばGMS データは，テープ上では見かけ上約 20MB のデータとなり，読み出し時間，書き込み時間とも 20%に短縮される．スケーラブルテープアーカイバはそれぞれが 2 台のテープドライブを有する 4 台のエレメントアーカイバ NTH-200B により構成され，初期状態では第 1 エレメントアーカイバから第 3 エレメントアーカイバの 3 台に NOAA データを記録したテープが古いものから順に，第 4 エレメントアーカイバにはGMS データのテープが配置されている．スケーラブルテープアーカイバの各エレメントアーカイバのパラメータは 4.5 節の表 4.1 に従う．

シミュレーションはキャッシュとしてのディスクがない場合，およびテープアーカイバ上のデータをキャッシュするために 10MB/s のデータ転送速度を有し，LRU によりデータを管理する 40GB のディスクを備えている場合について行う．また，スケーラブルテープアーカイバでのアクセススケジューリングとしては，まずリクエストを受けているサービス可能なテープの中で最も先にリクエストを受けたものを選択し，そのテープ上のリクエストを受けているすべてのデータをテープの先頭方向より順にアクセスするというスケジューリングを採用した．

図 6.4 は，初期状態から 450000 アクセスまでの平均応答時間である．横軸はリクエスト

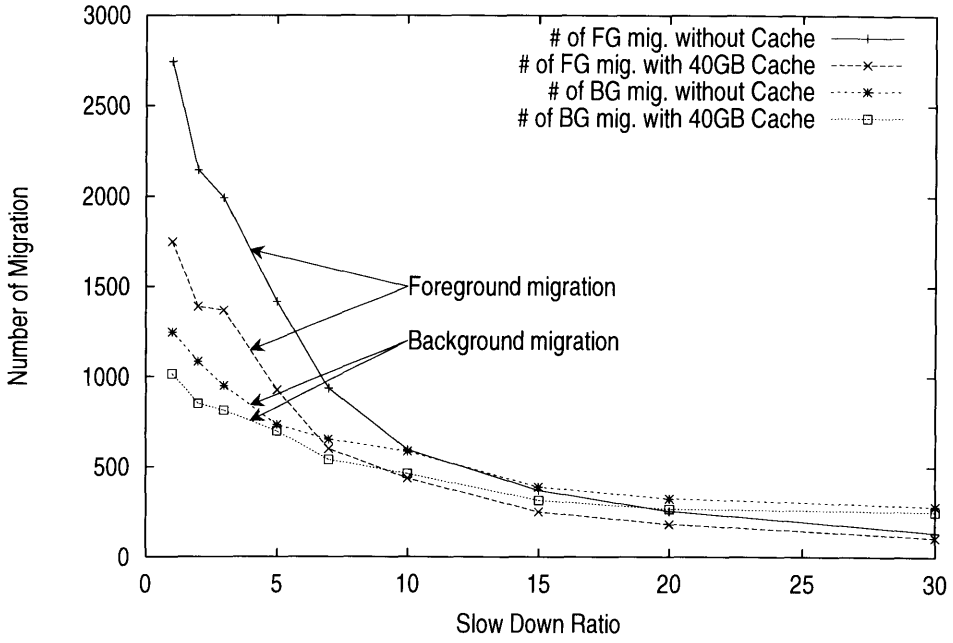

図 6.5: マイグレーション数

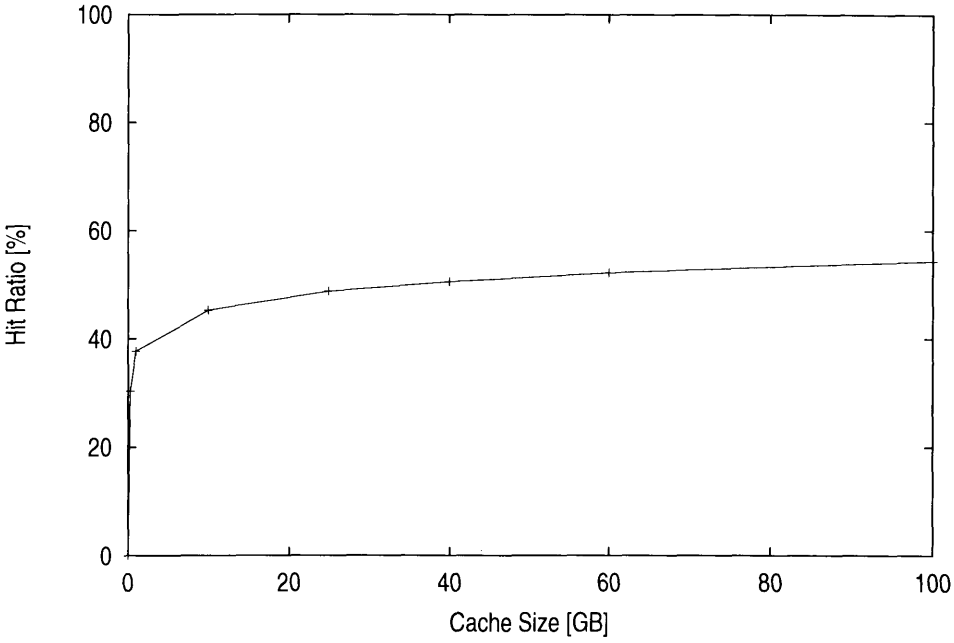

図 6.6: ディスクサイズとキャッシュのヒット率の関係

到着率を低下させるための各リクエスト到着時間間隔を延ばす際の倍率（slow down ratio）を示す．ディスクによるキャッシュの有無にかかわらず，ホットデクラスタリングは応答時間の短縮に有効であることが示されている．ディスクによるキャッシュのみを用いた場合に比べ，ホットデクラスタリングのみを用いた場合の方が応答時間は大きく短縮されており，ホットデクラスタリングがキャッシュディスクより有効であることが示された．図6.6は，キャッシュ用ディスクのサイズを変えたときのディスク容量とキャッシュのヒット率の関係を表しているが，ディスク容量を40GB以上に大きくしてもヒット率はほとんど向上せず，キャッシュによる応答時間短縮はほぼ限界に達していることがわかる．即ち，ここで取り上げた衛星画像データベースではアクセスローカリティは必ずしも高くなく，このような場合にはマイグレーション機構により大きく性能向上が可能であることがわかる．また，ディスクによるキャッシュとホットデクラスタリングを併用することで更に応答時間の短縮が図れることがわかる．

図6.5は，初期状態から450000アクセスまでのマイグレーション数を表している．リクエスト間隔が長い場合にはテープドライブが使用されているときに新たにリクエストを受けることが減るためにフォアグランドマイグレーション数は減少する．また，キャッシュを用いてもフォアグランドマイグレーション数はほとんど減少しないが，これはキャッシュにヒットするのは最新データのような短期間にアクセスが集中するデータであるのに対し，フォアグランドマイグレーションが生じるのはほぼ同時に複数のテープ上のデータに対してリクエストが発行された場合であり，これはあるユーザが過去のある期間のデータを一括してアクセスするような場合であるが，このときにアクセスされるデータがその直前にアクセスされていることはほとんどないためと考えられる．一方，バックグランドマイグレーション数はリクエスト到着時間間隔が長い場合にもフォアグランドマイグレーション数ほど減少しない．これは大部分のバックグランドマイグレーションは初期状態のテープ数の偏りを平衡化するために実行されているためである．リクエスト間隔が短い場合にバックグランドマイグレーション数が増加しているが，リクエスト到着時間間隔が短いときにはフォアグランドマイグレーション数が増加し，このために生じたテープ数の偏りを平衡化するためのバックグランドマイグレーションも増加するためである．

図6.7は，リクエストの到着時間間隔を5倍に延ばした場合の20000リクエストごとの

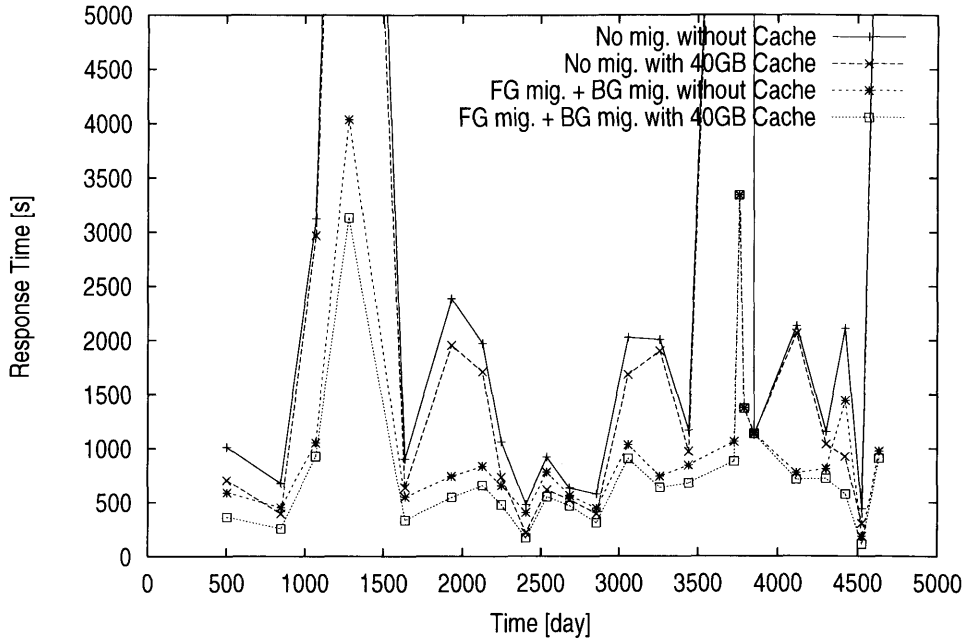

図 6.7: 平均応答時間の変化

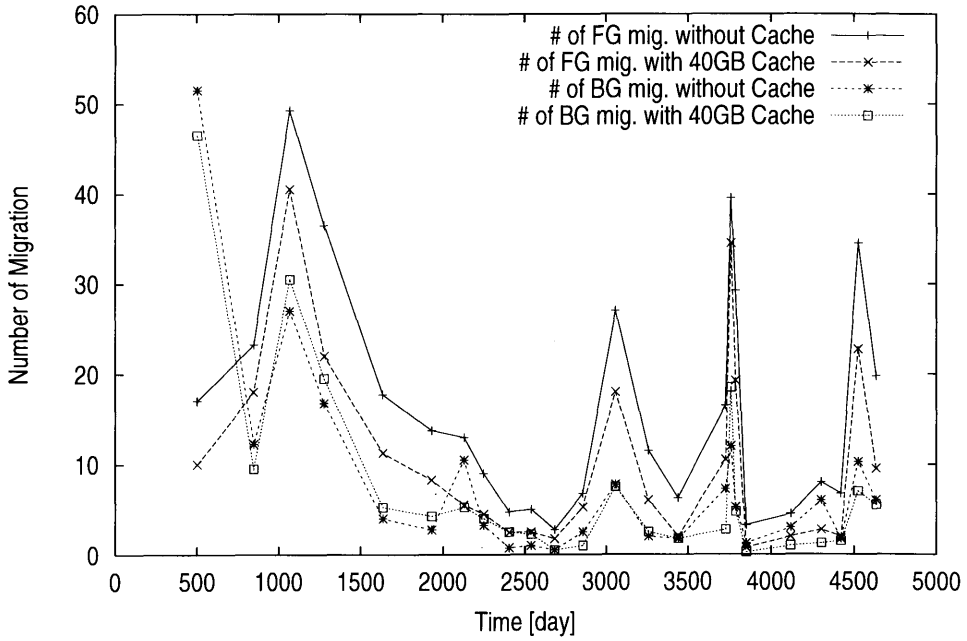

図 6.8: マイグレーション数の変化

平均応答時間を表している．1200日，2000日，3200日，4000日付近（オリジナルのリクエストにおいてはそれぞれ240日，400日，640日，800日付近）ではリクエストの到着率が高いため，ホットデクラスタリングを実行しない場合には応答時間が極めて長くなっているが，ホットデクラスタリングを用いることにより大幅に短縮されることがわかる．また，このときにはキャッシュが存在しても応答時間は余り短縮されていないこともわかる．これは，図6.1においてアクセスが縦方向に線状に分布するアクセス，即ち多くの種類のデータに対するアクセスが集中する期間であり，キャッシュが有効でないためである．一方，2500日，4400日付近は逆にホットデクラスタリングを用いても応答時間はほとんど短縮されないのに対し，キャッシュを用いると応答時間が短縮されている．この期間では小数のデータに繰り返しリクエストが発行されるためキャッシュは有効であり，逆にこれらのデータは同一テープ上に存在するためにホットデクラスタリングは機能しない．このようにディスクによるキャッシュとホットデクラスタリングは相補的に動作するために，図6.4に見られるように両者を併せて使用することでより大きな性能向上が得られる．

図6.8は，図6.7と同じくリクエストの到着時間間隔を5倍に延ばした場合の20000リクエストごとのマイグレーション数を表している．この図からも1200日，2000日，3200日，4000日付近では，フォアグランドマイグレーションが多く実行されており，これによって応答時間が短縮されることがわかる．バックグランドマイグレーションはシミュレーション開始後に多く実行されているが，これは初期状態のテープ数の偏りを平衡化するために実行されたマイグレーションである．その後はフォアグランドマイグレーションにより生じたテープ数の偏りを解消するためにバックグランドマイグレーションが実行されるため，フォアグランドマイグレーション数の増減と同期してバックグランドマイグレーションも増減する．

6.4 ホットレプリケーションの評価

本節では，シミュレーションによりホットレプリケーションの評価を行う．6.3節で述べたホットデクラスタリングの評価の場合と同様，クイックルック画像に対するアクセス履歴を用い，これらのリクエストが対応する衛星原画像へのリクエストであると仮定し，

シミュレーションを実行する．

6.3節同様，シミュレーションには6.2節で示したクイックルック画像に対するリクエストに新たに受信されたデータの書き込みリクエストを加えた489000のリクエストを用いる．アーカイブシステムはなるべく実システムに即した環境を想定するが，実システムでは各テープの全領域にオリジナルデータが記録されており，複製用の領域は確保されていないため，シミュレーションにおいては次のデータ配置を初期配置とする．各テープの容量は7GB（非圧縮時）であるとし，そのうち先頭より5.5GBの領域をオリジナルデータ用の領域として，衛星原データをデータ番号順に配置する．データはNOAA衛星画像データ，GMS衛星画像データともテープドライブ装置に備え付けられている圧縮機能を用いて圧縮されているものとする．テープはNOAA衛星画像データ用328本，GMS衛星画像データ用108本の計436本で構成される．スケーラブルテープアーカイバは4台のエレメントアーカイバNTH-200Bで構成され，初期状態では各エレメントアーカイバに均等に109本ずつテープを配置する．また，テープアーカイバ上のデータをキャッシュするためのディスクは10MB/sのデータ転送速度をもち，LRUによりデータを管理する．ディスク容量は，十分な容量である40GB，および平均的なファイルサイズの約3倍と極めて小さい容量である300MBとする．アクセススケジューリングとしては，5.3節と同様のスケジューリングを採用する．その他のパラメータは5.3節の表5.1に従う．

本シミュレーションでは，シミュレーション開始時には複製は存在せず，シミュレーション開始後に10度以上アクセスされたデータをホットデータとみなし，その複製をテープ終端部の複製用領域に動的に作成する．複製の作成対象となるテープは，複製用の領域を除くオリジナルデータ用の領域がすべて記録されたテープである．シミュレーション開始後にデータが記録されるテープに関しては，オリジナルデータ用の全領域にオリジナルデータが記録された時点で複製作成対象のテープとなる．

図6.9は，ホットデクラスタリングを用いたときの初期状態から450000アクセスまでの平均応答時間である．キャッスディスクサイズが300MB，および40GBの場合それぞれに対し，ホットレプリケーションを用いた場合，およびホットレプリケーションを用いない場合についての結果を示している．横軸はリクエスト到着率を変化させるための各リクエスト間隔を延ばす際の倍率を表すリクエスト遅延率である．図6.10はホットデク

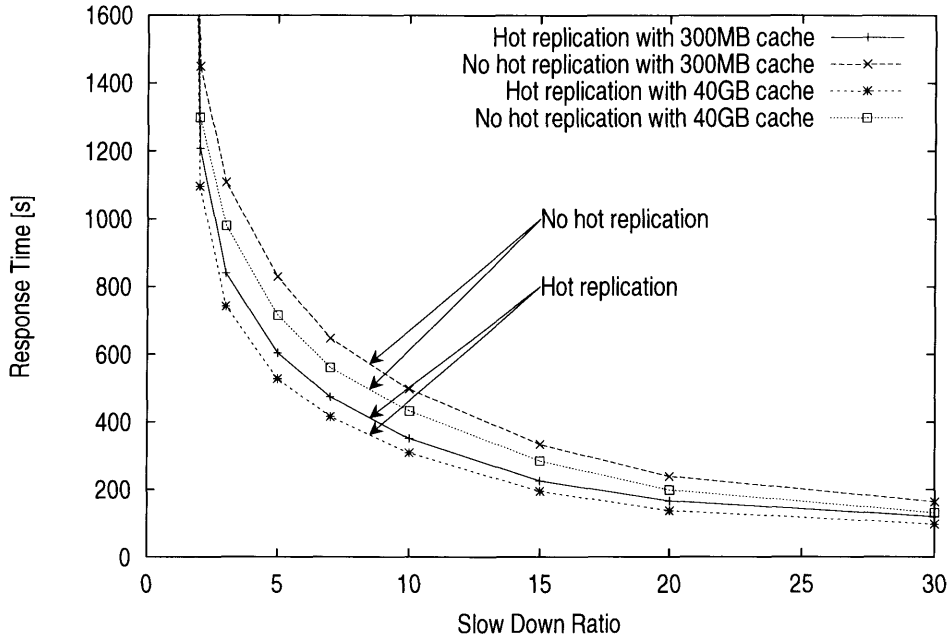

図 6.9: ホットレプリケーションによる平均応答時間（ホットデクラスタリングあり）

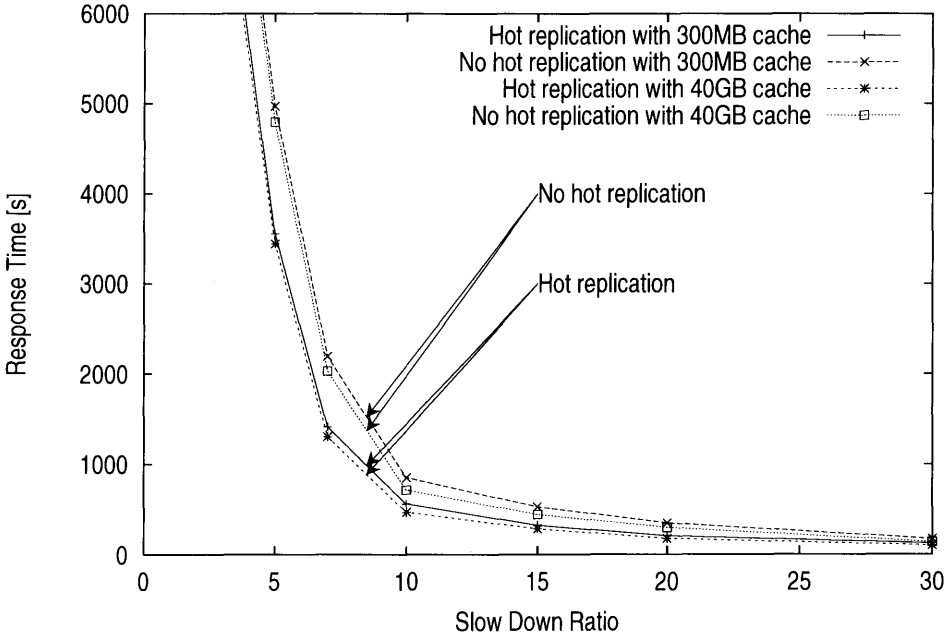

図 6.10: ホットレプリケーションによる平均応答時間（ホットデクラスタリングなし）

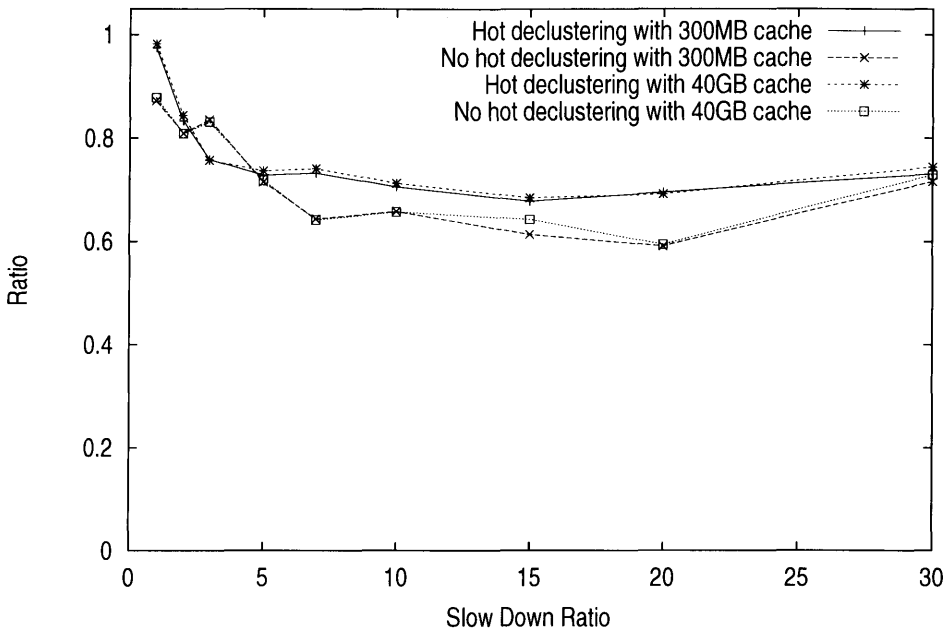

図 6.11: ホットレプリケーションによる平均応答時間の短縮率

ラスタリングを用いない場合の初期状態から 450000 アクセスまでの平均応答時間である．図 6.9 と同様にキャッスディスクサイズが 300MB，および 40GB の場合それぞれに対し，ホットレプリケーションを用いた場合，およびホットレプリケーションを用いない場合についての結果を示している．また，図 6.11 は各シミュレーション条件におけるホットレプリケーションを用いない場合の応答時間を 1 としたときのホットレプリケーションを用いた場合の相対平均応答時間を表している．

ホットデクラスタリングの有無，ディスクによるキャッシュのサイズに関わらず，ホットデータの複製を作成することにより平均応答時間が短縮されることがわかる．ホットデクラスタリングを用いた場合，用いない場合とも，40GB のディスクによるキャッシュ以上にホットレプリケーションは平均応答時間を短縮している．図 6.6 はキャッシュサイズとヒット率の関係をしているが，キャッシュサイズを 40GB 以上にしてもヒット率はほとんど向上せず，40GB が十分なサイズであることがわかる．すなわち，ホットレプリケーションは十分なサイズのキャッシュ以上に応答性能を向上させることがわかる．また，ホットレプリケーションはキャッシュサイズが 300MB の場合においても平均応答時間を短縮し，キャッシュディスクが極めて小さくても十分に複製が作成され，平均応答時間が短縮され

ることがわかる．シミュレーション開始から450000アクセスまでに作成された複製データ数は，キャッシュサイズが300MBの場合では9337~9762，40GBの場合では9675~10066であり，これは全データの約16~17%にあたる．この結果からも，キャッシュディスクが極めて小さい場合においても，十分にキャッシュディスクが存在する場合とほぼ同程度に複製が作成されており，ホットレプリケーションは有効であることが示されている．

また，図6.11より，リクエスト遅延率が大きい場合，ホットデクラスタリングを用いていないときは，用いたときよりも平均応答時間を短縮している．これは，テープドライブがすべて使用されているエレメントアーカイバ内のテープに新たにリクエストが発行された場合に，ホットデクラスタリングではそのテープの移動を行い，ホットレプリケーションでは別のエレメントアーカイバ内の複製を参照するというように，リクエストがブロックされることを避ける場合など，ホットデクラスタリングとホットレプリケーションが効果を示す状況が一部重複しているためである．すなわち，ホットデクラスタリング適用時には，ホットレプリケーションが効果を発揮する状況の一部を既にホットデクラスタリングが解消してしまっているため，その分，ホットレプリケーションの効果は小さくなる．また，リクエスト遅延率が小さくなるとアクセスリクエストへの対応のためにテープドライブ装置の使用率が上昇し，複製の作成が効果的に行われなくなるためにホットレプリケーションの効果が低下する．

図6.12，図6.13はそれぞれホットデクラスタリングを用いた場合，およびホットデクラスタリングを用いない場合のリクエスト遅延率が5のときの20000アクセス毎の平均応答時間を表している．1600日（リクエスト遅延率1では320日）以前では，まだ十分に複製が作成されておらず，キャッシュがホットレプリケーション以上に効果を示しているが，1600日以降はホットレプリケーションはキャッシュ以上に応答時間を短縮している．また，図6.1において垂直方向の線状の分布が見られ，過去のある一定期間の連続する多数のデータへのアクセスが行われる場合においてもホットレプリケーションは効果を示している．これは，ホットレプリケーションを用いていない場合，ほぼ同時に同一テープ上の連続する一連のデータに対してリクエストが発行されることが多いが，それらを同時に読み込むことはできず，順次読み込まなければならないため，多数のリクエストが待たされることとなり応答性能が悪化する．また，アクセスされるデータは最新画像データで

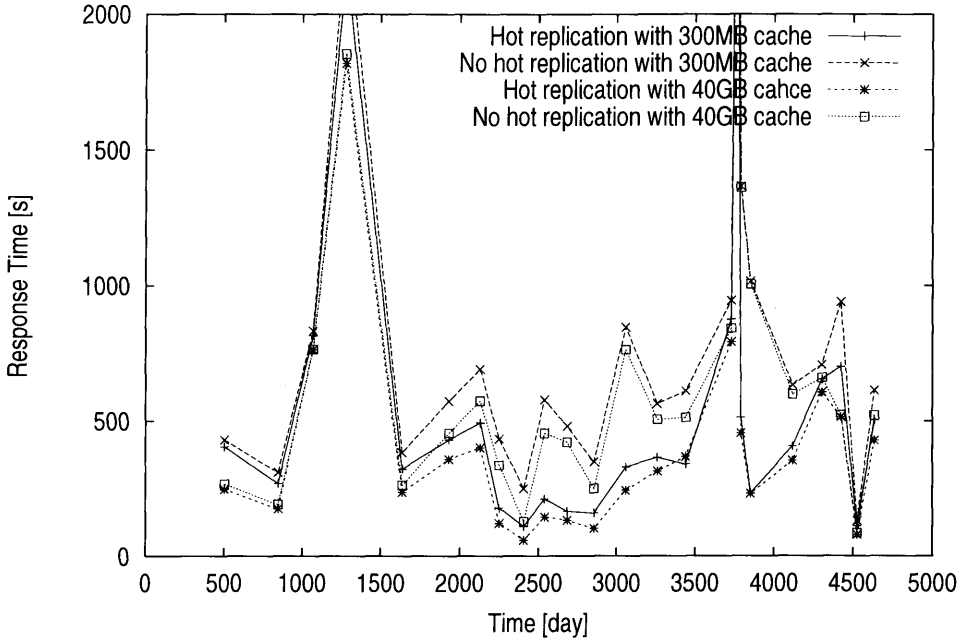

図 6.12: ホットレプリケーションによる平均応答時間の変化（ホットデクラスタリングあり）

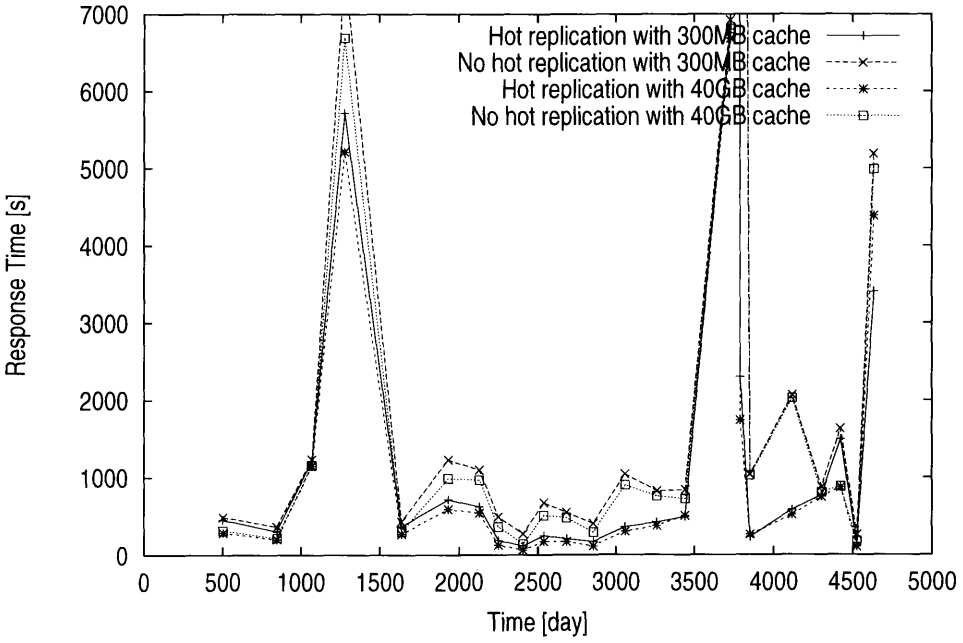

図 6.13: ホットレプリケーションによる平均応答時間の変化（ホットデクラスタリングなし）

はないためにその直前にアクセスされていることはほとんどなく，キャッシュも有効ではない．一方，ホットレプリケーションでは，このような場合，異なるテープ上の複製にアクセスすることが可能であるため，応答時間が短縮される．大規模アーカイブシステムにおいては，過去のある一定期間の連続する多数のデータへのアクセスのように，少数のメディア上の複数のデータをまとめてアクセスするということは多々あるが，ホットレプリケーションはこのようなアクセスに対しては，異なるテープ上の複製を用いて並列アクセスが行われるため，極めて効果的である．

6.5 まとめ

本節では，生産技術研究所においてWWW，gopher，ftpにより公開している衛星データのクイックルック画像に対するアクセス履歴を用い，これを衛星原画像へのアクセスであると仮定してシミュレーションを行い，ホットデクラスタリング，ホットレプリケーションの評価を行った．クイックルック画像へのアクセスパターンは原画像に対するアクセスパターンに近い傾向をもつと考えられるが，このアクセス履歴を用いたシミュレーションにおいても，ホットデクラスタリングは応答時間の短縮に有効であることを示した．特に，ホットデクラスタリングとディスクによるキャッシュは相補的であり，短期間に多数の異なるデータにアクセスするような，ディスクによるキャッシュが有効でない場合において応答時間が短縮され，ディスクによるキャッシュと併用することでよりいっそうの応答性能の向上が図れるることを示した．また，ホットレプリケーションを適用することで，さらに平均応答時間の短縮が図れることを明らかにした．特に，キャッシュ，ホットデクラスタリングをともに導入した場合においても，さらにホットレプリケーションを行うことで，応答時間の短縮可能であることを示した．

第7章　結論

7.1　本論文のまとめ

本稿では，三次記憶システム上のファイルへのアクセスの2つの特徴，ファイル内の部分参照性とファイル間の参照局所性を利用した高速化手法について述べた．

ファイル内の部分参照性に関しては，必要される部分のみを三次記憶と二次記憶の間でマイグレートする部分マイグレーション機能を提案し，さらに試作システムにおいて，衛星画像処理において用いられる2つの実アプリケーション，放射量・幾何補正プログラム，NDVI生成プログラムを実行し，処理時間が大幅に短縮されることを示し，部分マイグレーション機能の有効性を示した．

ファイル間の参照局所性に関しては，テープのアクセス頻度に応じてテープを筐体間でマイグレートするホットデクラスタリング，およびアクセス頻度の高いデータの複製ををあらかじめ用意したテープ上の領域へ作成してクラスタリングするホットレプリケーションを提案した．ホットデクラスタリングに関しては，シミュレーションにより応答時間が短縮されることを示した．さらに，テープドライブ故障時，ストライピング環境下においても有効であることを示した．ホットレプリケーションに関しては，シーク長の短縮の効果，複製による多重アクセスの効果により応答性能が向上することをシミュレーションにより明らかにした．加えて，東京大学生産技術研究所において構築中である衛星画像データベースへのインターネットを通じたアクセスの履歴を用いてシミュレーションを行い，ホットデクラスタリング，ホットレプリケーションが実システムにおいても，極めて有効であることを示すとともに，キャッシュを含め，これらの手法を同時に用いることでさらなる応答時間の短縮が図れることを示した．

7.2 今後の展開

本論文では，ホットデクラスタリングおよび，ホットレプリケーション単体の性能を明確にするため，単純なI/Oスケジューリングを採用しシミュレーションを行ったが，より高度なI/Oスケジューリングを用いることでさらに性能を向上させることが可能であると考えられる．具体的には，ホットデクラスタリング使用時には，マイグレーションを行って他のエレメントアーカイバ内の空きドライブを使用しアクセスすべきか，マイグレーションをせずに現在行われているサービスの終了後にそのドライブを用いてアクセスするべきかの判断，ホットレプリケーションではオリジナルデータと複製のいずれをアクセスすべきかの選択などの点を検討する必要がある．このようなより緻密なI/Oスケジューリングを用いることでさらなる応答性能の向上が得られると考えられる．

また，ホットレプリケーションにおけるガベージコレクションも考慮する必要があろう．地球環境情報データをはじめとする大規模ファイルにおいては，一般的に各ファイルのアクセス頻度は時々刻々変化する．すなわち，ホットレプリケーションにより作成された複製データは作成時においては高アクセス頻度データであるが，時間の経過とともにそれらへのアクセス頻度は低下することも考えられる．テープ上の複製データ用の領域は限られており，アクセス頻度の低いデータの複製を保持することにより新たな高アクセス頻度データの複製のための領域は減少するため，ホットレプリケーションの効果は低下してしまうこととなる．このような状況を避けるためには，アクセス頻度が低下したデータの複製を削除するガベージコレクションが必要となる．一般的な磁気テープドライブにおいては，途中のデータのみを削除し，その領域を再利用することは困難である．あるデータを消去する場合にはそのデータよりも後方に存在する全データも同時に削除されることとなる．この点を考慮したガベージコレクション手法を検討する必要がある．具体的には，後方の複製データとともにアクセス頻度の低下した複製を消去すべきか，あるいはアクセス頻度低下したデータの存在を容認するかの判断基準の検討，実行のタイミング，さらにはガベージコレクションを容易に行えるファイル編成に基づくホットレプリケーション手法も検討する必要があろう．

広く三次記憶システム一般の将来については，二次記憶装置の急速な低価格化，大容

量化を考慮する必要がある．ディスクドライブ単体では，パーソナルコンピュータ用の最も廉価なもの単位容量あたりの価格は既にテープメディアと同程度となっている．製品化されたディスクアレイとしては，ディスクアレイにコンポーネントとして用いられるディスクドライブは一般に信頼性の高く高速なものが使用され，また，ディスクアレイコントローラなども必要となるため，必ずしも容量単価はテープメディアと同程度とはならないが，数十TB程度の容量をディスクアレイのみで実現することはもはや困難なことではない．しかしながら，ディスクアレイはすぐにデータが書き換え可能な状態でホストコンピュータと接続されているため，誤動作や誤操作でデータを消去しやすいというような信頼性の問題や，大量のデータの保存のためには多数のドライブを稼働させる必要があり，そのために膨大な電力を必要とするなど，重要なデータの最終保管をディスクアレイのみで行うということは考えにくく，当面は，現在のディスクアレイがそのままテープライブラリ装置のような三次記憶システムに置き換わるということはないと考えられる．

現在のディスクアレイのもつ問題点に対しては，ディスクアレイなどの二次記憶装置を三次記憶装置的に使用するというアプローチが考えられる．例えば，アクセスリクエストに応じて必要なドライブのみを稼働させ，他のドライブは停止させることで電力消費量を抑えることは可能であろう．この方法では，起動，停止を繰り返すこととなり，ディスクドライブの信頼性を低下させることとなってしまうが，二次記憶装置の三次記憶装置的使用については今後，検討する必要があると思われる．

また，現在の三次記憶に対する二次記憶の利用，すなわち三次記憶のキャッシュとしての二次記憶という観点では，三次記憶の容量に対する二次記憶の容量の割合は増大しているが，キャッシュとしては充分な容量になりつつあり，もはやキャッシュとしては大幅な性能向上は望めなくなってきている．広大な二次記憶領域を活かすためには，適切なキャッシュアルゴリズムと併せて積極的なプリフェッチが有効であると考えられる．今後は，三次記憶システムに適した高性能プリフェッチアルゴリズムの検討が必要である．

謝辞

本研究を進めるにあたり，多くの方々のご指導，ご鞭撻を頂き，深く感謝致します．

指導教官である喜連川優教授には，日頃から研究全般にわたってのご指導を賜り，研究を進めていく上での環境面についても多大なご配慮を頂きました．本研究は，衛星画像データを例とし，大規模ファイルのアーカイブシステムの高性能化を主眼とした研究であり，ややもすると現実の制約に捕らわれ大局的な目標を見失いがちでしたが，喜連川教授の助言により，現実のシステムと研究とのバランスをとることができました．

また，現東京理科大学の高木幹雄教授には，衛星画像受信システムの整備にご尽力を頂き，現アーカイブシステムの礎を築いて頂くとともに，衛星画像データの取り扱いをはじめ，数々の貴重なアドバイスを頂きました．

現国立情報学研究所の羽鳥光俊教授，浜田喬教授，安達淳教授，現東京大学情報理工学系研究科の田中英彦教授，同大学新領域創成科学研究科の相田仁教授には，論文審査を通じ有益なご指導，ご意見を頂きました．

現株式会社日立製作所の迫和彦氏には部分マイグレーションの実験システムへの実装，および実験システム上での性能測定をして頂き，大変お世話になりました．

中野美由紀女史には，研究室の計算機環境ばかりでなく，研究室の生活環境の整備に多大な労力をつぎ込んで下さいました．

最後になりましたが，喜連川研究室の方々，および高木・喜連川両研究室のOB，OGの方々には，研究環境の整備や日常生活の面で大変お世話になりました．ここに改めて深く感謝の意を表します．

参考文献

[1] L. T. Chen, R. Drach, M. Keating, S. Louis, D. Rotem, and A. Shoshani. "Efficient organization and access of multi-dimensional datasets on tertiary storage systems". *Information Systems journal*, 1995.

[2] L. T. Chen and D. Rotem. "Optimizaing storage of objects on mass storage systems with robotic devices". In *Proceedings of EDBT'94*, pp. 273–286, Cambridge, U.K., 1994.

[3] L. T. Chen, D. Rotem, A. Shoshani, B. Drach, M. Keating, and S. Louis. "Optimizing tertiary storage organization and access for spacio-temporal datasets". In *Proceedings of Fourth NASA Goddard Conference on Mass Storage Systems and Technologies*, Collage Park, Maryland, March 1995.

[4] S. Christodoulakis, P. Triantafillou, and F. A. Zioga. "Principles of optimally placing data in tertiary storage libraries". In *Proceedings of the Twenty-third Very Large Database Conference*, pp. 236–245, Athenes, Greece, August 1997.

[5] Hewlett-Packard Company. "dds/dat drives". `http://www.products.storage.hp.com/eprise/main/storage/DisplayPages/ddsdat.html`.

[6] G. Copeland, W. Alexander, E. Boughter, and T. Keller. "Data placement in Bubba". In *Proceedings of the 1988 ACM SIGMOD International Conference on Management of Data*, pp. 99–109, Chicago, Illinois, June 1988.

[7] Ampex Corporation. "Ampex products". `http://www.ampexdata.com/products/dst/312.html`.

[8] Exabyte Corporation. "Exabyte products". http://www.exabyte.com/home/products.html.

[9] International Business Machines Corporation. "IBM LTO linear tape-open family". http://www.storage.ibm.com/hardsoft/tape/lto/index.html.

[10] Quantum Corporation. "Quantum 1 DLT tape". http://www.quantum.com/Products/Quantum+1+DLTtape/Default.htm.

[11] Sony Corporation. "AIT-2 テープドライブ AIT-S100". http://www.sony.co.jp/sd/ProductsPark/Professional/DataArchive/BC2/BC2-2/AIT-S100/index.html.

[12] Sony Corporation. "DTF-1 テープドライブ GY-2120". http://www.sony.co.jp/sd/ProductsPark/Professional/DataArchive/BC2/BC2-1/GY2120/index.html.

[13] Sony Corporation. "PetaServ". http://www.sony.co.jp/sd/ProductsPark/Professional/DataArchive/BC2/BC2-1/PetaServe/index.html.

[14] Sony Corporation. "PetaSite". http://www.sony.co.jp/sd/ProductsPark/Professional/DataArchive/BC2/BC2-1/PetaSite/index.html.

[15] Storage Technology Corporation. "StorageTek products". http://www.storagetek.com/products/.

[16] A. L. Drapeau and R. H. Katz. "Striped tape arrays". In *Proceedings of Twelfth IEEE Symposium on Mass Storage Systems*, pp. 257–265, Montrey, California, April 1993.

[17] E-Systems, Inc. *"EMASS Volume Server Technical Summary"*, 1993.

[18] D. A. Ford and J. Myllymaki. "A log-structured organization for teriary storage". In *Proceedings of the Twelfth International Conference on Data Engineering*, pp. 20–27, New Orleans, Louisiana, February 1996.

[19] General Atomics. *"UniTree Central File Manager User Guide"*, 1992.

[20] L. Golubchik, R. Muntz, and R. W. Watson. "Analysis of striping techniques in robotic storage libraries". In *Proceedings of Fourteenth IEEE Symposium on Mass Storage Systems*, pp. 225–238, Monterey, California, September 1995.

[21] B. K. Hillyer, R. Rastogi, and A. Silberschatz. "Scheduling and data replication to improve tape jukebox performace". In *Proceedings of the Fifteenth International Conference on Data Engineering*, pp. 532–541, Sydney, Australia, March 1999.

[22] B. K. Hillyer and A. Silberschatz. "On the modeling and performance characteristics of a serpentine tape drive". In *Proceedings of the 1996 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems*, pp. 170–179, Philadelphia, Pennsylvania, May 1996.

[23] B. K. Hillyer and A. Silberschatz. "Random I/O scheduling in online tertiary storage". In *Proceedings of the 1996 ACM SIGMOD International Conference on Management of Data*, pp. 195–204, Montreal, Canada, June 1996.

[24] B. K. Hillyer and A. Silberschatz. "Storage technology: Status, issues, and opportunities". `http://www.bell-labs.com/user/hillyer/papers/tsurv.ps`, June 1996.

[25] B. K. Hillyer and A. Silberschatz. "Scheduling non-contiguous tape retrievals". In *Proceedings of Sixth Goddard Conference on Mass Storage Systems and Technologies in cooperation with the Fifteenth IEEE Symposium on Mass Storage Systems*, pp. 113–123, College Park, Maryland, March 1998.

[26] Ltd. Hitachi. "DVD library system 製品情報". `http://www.hitachi.co.jp/Prod/comp/OSD/ram/prod/index.htm`.

[27] M. Kitsuregawa, T. Nemoto, and M. Takagi. "File management and migration on scalable tape archiver". In *Proceedings of the Second SEIKEN Symposium on Global Environmental Monitoring from Space*, pp. 17–25, February 1996.

[28] J. T. Kohl, C. Staelin, and M. Stonebraker. "HightLight:using a log-structured file system for tertiary storage management". In *Proceedings of Winter 1993 USENIX Conference*, pp. 435–447, San Diego, California, January 1993.

[29] A. Kraiss and G. Weikum. "Vertical data migration in large near-line document archives based on Markov-chain predictions". In *Proceedings of the Twenty-third International Conference on Very Large Data Bases*, pp. 246–255, Athens, Greece, August 1997.

[30] Fujitsu Limited. "MO ソリューション > 製品情報". `http://mo.fujitsu.com/products/`.

[31] Matsushita Electric Industrial Co. Ltd. "DVD RAM & PD". `http://www.panasonic.co.jp/dvdram/index.html`.

[32] J. Myllymaki and M. Livny. "Disk-tape joins: Synchronizing disk and tape access". In *Proceedings of Joint International Conference on Measurement and Modeling of Computer Systems SIGMETRICS '95/PERFORMANCE '95*, pp. 279–290, Ottawa, Canada, May 1995.

[33] J. Myllymaki and M. Livny. "Relational joins for data on tertiary storage". In *Proceedings of Thirteenth International Conference on Data Engineering*, pp. 159–168, Birmingham, UK, April 1997.

[34] T. Nemoto and M. Kitsuregawa. "Effectiveness of tape migration for satellite image database". In *Proceedings of International Symposium on Remote Sensing*, pp. 52–57, Pusan, Korea, November 1997.

[35] T. Nemoto and M. Kitsuregawa. "Scalable tape archiver for satellite image database and its performance analysis with access logs – hot declustering and hot replication –". In

Proceedings of 16th IEEE Symposium on Mass Storage Systems in cooperation with the 7th NASA GSFC Conference on Mass Storage Systems and Technologies, pp. 59–71, San Diego, California, March 1999.

[36] T. Nemoto, M. Kitsuregawa, and M. Takagi. "Design and implementation of scalable tape archiver". In *Proceedings of Fifth NASA GSFC Conference on Mass Storage Systems and Technologies*, pp. 229–237, Collage Park, Maryland, September 1996.

[37] T. Nemoto, M. Kitsuregawa, and M. Takagi. "Scalable tape archiver and its application to satellite image database". In *Proceedings of International Symposium on Remote Sensing*, pp. 185–189, Cheju, Korea, October 1996.

[38] T. Nemoto, M. Kitsuregawa, and M. Takagi. "Simulation studies of the cassette migration activities in a scalable tape archiver". In *Proceedings of The Fifth International Conference on Database Systems for Advanced Applications*, pp. 461–470, Melbourne, Australia, April 1997.

[39] T. Nemoto, K. Sako, M. Kitsuregawa, and M. Takagi. "Partial migratable hierarchical file system for satellite image database". In *Proceedings of International Symposium on Remote Sensing*, pp. 45–49, Taejon, Korea, October 1995.

[40] T. Nemoto, Y. Sato, K. Mogi, K. Ayukawa, M. Kitsuregawa, and M. Takagi. "Performance evaluation of cassette migration mechanism for scalable tape archiver". In *SPIE, Digital Image Storage and Archiving System*, pp. 48–58, Philadelphia, Pennsylvania, October 1995.

[41] OPENVISION Technologies, Inc. *"AXXiON-HSM AXXiON-Enterprise Extension System Adiministrator's Guide"*, 1996.

[42] S. Prabhakar, D. Agrawal, A. E. Abbadi, and A. Singh. "Tertiary storage: Current status and future trends". In *Proceedings of the 1997 ACM Symposium on Applied Computing*, pp. 3–19, San Jose, California, February 1997.

[43] M. Rosenblum and J. K. Ousterhout. "The design and implementation of a log-structured file system". *Proceedings of the Thirteenth ACM Symposium on Operating Systems Principles*, pp. 1–15, October 1991.

[44] K. Sako, T. Nemoto, M. Kitsuregawa, and M. Takagi. "Partial migration in an 8mm tape based tertiary storage file system and its performance evaluation through satellite image processing applications". In *Proceedings of 6th International Conference on Information Systems and Management of Data*, pp. 178–190, Bombey, India, 1995.

[45] O. Sandstå and R. Midtstraum. "Improving the access time performance of serpentine tape drives". In *Proceedings of the Fifteenth International Conference on Data Engineering*, pp. 542–551, Sydney, Australia, March 1999.

[46] S. Sarawagi. "Database systems for efficient access to tertiary memory". In *Proceedings of the Fourteenth IEEE Symposium on Mass Storage Systems*, pp. 120–126, Monterey, California, September 1995.

[47] S. Sarawagi. "Query processing in tertiary memory databases". In *Proceedings of the Twenty-first International Conference on Very Large Data Bases*, pp. 585–596, Zurich, Swizerland, September 1995.

[48] S. Sarawagi and M. Stonebraker. "Efficient organization of large multidimensional arrays". In *Proceedings of Tenth International Conference on Data Engineering*, pp. 328–336, Houston, Texas, February 1994.

[49] S. Sarawagi and M. Stonebraker. "Single query optimization for tertiary memory". SEQUOIA 2000 Technical Report 94/45, University of California, Berkeley, California, 1994.

[50] S. Sarawagi and M. Stonebraker. "Reordering query execution in tertiary memory databases. In *Proceedings of the Twenty-second International Conference on Very Large Databases*, pp. 156–167, Mumbai (Bombay), India, September 1996.

[51] M. Stonebraker, J. Frew, and J. Dozier. "The SEQUOIA 2000 architecture and implementation strategy". SEQUOIA 2000 Technical Report 93/23, University of California, Berkeley, California, 1993.

[52] M. Stonebraker and M. Olson. "Large object support in POSTGRES". In *Proceedings of Ninth International Conference on Data Engineering*, pp. 355–362, April 1993.

[53] S. Takamura, T. Nemoto, and M. Takagi. "Remote sensing data receiving and processing". In *Proceedings of International Seminar on Space Informatics and Sustainable Development: Grassland Monitoring and Management*, 1995.

[54] G. Weikum, P. Zabback, and P Scheuermann. "Dynamic file allocation in disk arrays". In *Proceedings of the 1991 ACM SIGMOD International Conference on Management of Data*, pp. 406–415, Denver, Colorado, May 1991.

[55] T. S. Woodrow. "Hierarchical storage management system evaluation". In *Proceedinsg of Third NASA Goddard Conference on Mass Storage Systems and Technologies*, pp. 187–216, October 1993.

[56] B. Zhai. "Fetching data from tertiary storage system". Master's thesis, National Unversity of Singapore, 1997.

[57] 向井景洋, 根本利弘, 喜連川優. "高性能ディスクにおけるアクセスプランを用いたプリフェッチ機構に関する評価". 第 11 回データ工学ワークショップ, March 2000.

[58] 鮎川健一郎, 根本利弘, 喜連川優, 高木幹雄. "大規模テープ・アーカイバにおけるマイグレーションによる負 荷分散". 情報処理学会第 52 回全国大会講演論文集, 1995. 3Y-9.

[59] 鮎川健一郎, 根本利弘, 茂木和彦, 喜連川優, 高木幹雄. "大規模テープ・アーカイバにおけるマイグレーションのシミュレーションによる評価". 情報処理学会第 51 回全国大会講演論文集, 1995. 4D-8.

[60] 根本利弘, 鮎川健一郎, 茂木和彦, 喜連川優, 高木幹雄. “スケーラブルテープアーカイバにおけるカセットマイグレーション方式とその評価”. 並列処理シンポジウム JSPP'96 論文集, pp. 275–282, June 1996.

[61] 根本利弘, 喜連川優. “ホットレプリケーション : 三次記憶システムにおける高アクセス頻度データの複製クラスタリング手法”. 情報処理学会論文誌. 投稿中.

[62] 根本利弘, 喜連川優. “衛星画像データベースのアクセス履歴を用いたスケーラブルテープアーカイバの性能評価”. 第 9 回データ工学ワークショップ, March 1998.

[63] 根本利弘, 喜連川優. “衛星画像データベースのアクセス履歴を用いたテープアーカイバにおけるメディアマイグレーション機構の性能評価”. 情報処理学会第 56 回全国大会講演論文集, 1998. 6Aa-8.

[64] 根本利弘, 喜連川優. “衛星画像データベースのアクセス履歴を用いたホットレプリケーションの評価”. 電子情報通信学会 1998 年情報・システムソサイエティ大会講演論文集, 1998. D-4-8.

[65] 根本利弘, 喜連川優. “スケーラブルテープアーカイバにおけるテープマイグレーションを用いた負荷分散手法とその性能評価”. 電子情報通信学会論文誌, Vol. J82-D-I, No. 1, pp. 53–69, January 1999.

[66] 根本利弘, 喜連川優. “テープアーカイブシステムにおけるホットレプリケーションの性能評価”. 電子情報通信学会技術研究報告, Vol. 100, No. 226, pp. 105–112, 2000.

[67] 根本利弘, 喜連川優. ‘dvd-ram ドライブを用いたスケーラブルアーカイバにおけるホットデクラスタリングの性能評価’. 第 12 回データ工学ワークショップ, March 2001.

[68] 根本利弘, 喜連川優, 高木幹雄. “三次記憶システムにおけるファイル編成に関する一考察”. 情報処理学会第 52 回全国大会講演論文集, 1995. 1Y-3.

[69] 根本利弘, 喜連川優, 高木幹雄. “スケーラブルテープアーカイバを用いた大規模ファイルシステムにおけるファイル編成方式の検討”. 情報処理学会第 53 回全国大会講演論文集, 1996. 1R-6.

[70] 根本利弘, 喜連川優, 高木幹雄. “リモートセンシング画像データベースの構築とその利用法”. 生研フォーラム「宇宙からの地球環境モニタリング」, pp. 10–19, November 1996.

[71] 根本利弘, 喜連川優, 高木幹雄. “スケーラブルテープアーカイバにおけるテープ上での動的データ再配置”. 情報処理学会第 55 回全国大会講演論文集, 1997. 4AC-6.

[72] 根本利弘, 喜連川優, 高木幹雄. “ネットワークによる衛星画像データアーカイブシステムの利用法”. 生研フォーラム「宇宙からの地球環境モニタリング」, pp. 97–106, February 1997.

[73] 根本利弘, 喜連川優, 高木幹雄. “大規模テープアーカイバにおけるデータ再配置手法の検討”. 情報処理学会第 54 回全国大会講演論文集, 1997. 3R-4.

[74] 根本利弘, 迫和彦, 喜連川優, 高木幹雄. “衛星データの高速アクセスを目的とした階層ファイルシステム の設計”. 生研フォーラム「宇宙からの地球環境モニタリング」, pp. 104–111, June 1995.

[75] 根本利弘, 迫和彦, 喜連川優, 高木幹雄. “衛星データを対象とした超大規模画像データベースの構想”. 情報処理学会第 50 回全国大会講演論文集, 1995. 2F-8.

[76] 根本利弘, 迫和彦, 喜連川優, 高木幹雄. “衛星画像の格納を目的とした大規模階層ファイルシステムの設計”. 情報処理学会研究報告, Vol. 95, No. 65, pp. 65–71, 1995.

[77] 根本利弘, 迫和彦, 喜連川優, 高木幹雄. “部分マイグレーション機能を有する大規模階層ファイルシステムの試作”. 情報処理学会第 51 回全国大会講演論文集, 1995. 5L-4.

[78] 根本利弘, 迫和彦, 喜連川優, 高木幹雄. “部分マイグレーション機構を有する三次記憶ファイルシステム PFS の 8 ミリテープアーカイブ装置への実装とその性能評価”. 電子情報通信学会論文誌, Vol. J81-D-I, No. 6, pp. 747–754, June 1998.

付録A　生産技術研究所における衛星画像データベースシステム

A.1　概要

東京大学生産技術研究所では1983年に気象衛星NOAAによる地表面観測画像の受信，1995年より気象衛星GMS（ひまわり）による観測画像の受信，さらに2001年5月よりTerra衛星のMODIS画像の受信を開始した．これらの受信した画像はすべてアーカイブされ，国内外の研究者への配布が行われる．また，1999年1月以降，タイのAsian Institute of Technology（AIT）において受信されたNOAA衛星による画像のアーカイブも行っており，今後，AITによるMODIS画像のアーカイブも予定している．2001年11月時点で，東京にて受信されたNOAAによる画像約39000シーン，計3.6TB，GMSによる画像約54000シーン，計5.4TB，タイにおいて受信されたNOAA画像約8000シーン，計800GB，MODISによる画像約800シーン，計900GBの合計約100000シーン，10.7TBの画像がアーカイブされている．これらの画像はテープアーカイバに格納されており，衛星名，観測日時等の情報とともに各画像のインデックスがデータベース化され，WWWによる検索，データの取得が可能となっている．

A.2　アーカイブデータ

A.2.1　NOAA AVHRRデータ

気象衛星TIROS-N/NOAAシリーズは，米国海洋大気庁NOAA（National Oceanic and Atmospheric Administration）による衛星であり，1978年にTIROS-N号が打ち上げられて以来，NOAA-6～16号が順次打ち上げられている．衛星NOAAの軌道は，高度約850kmの太陽同期極軌道であり，1つの衛星により地上の同一地点が1日に昼，夜の2回観測さ

表 A.1: NOAA AVHRR センサのチャネル構成

チャネル	主な利用法	波長	解像度
1	雲，氷，雪等の分布	0.58μm~0.68μm	1.1km
2	雪と氷，海水面と海氷等の識別	0.725μm~1.10μm	1.1km
3	海面，陸，雲等の温度分布	3.55μm~3.93μm	1.1km
4	海面，陸，雲等の温度分布	10.3μm~11.3μm	1.1km
5	海面，陸，雲等の温度分布	11.4μm~12.4μm	1.1km

れる．データ受信局からは，同一時間帯において，衛星が受信局の東側を通過するパスと西側を通過するパスの2回のデータを取得することができることが多く，従って1日あたりに受信されるデータは2機の衛星で計8シーン程度となる．

衛星NOAAに搭載されているAVHRR（Advanced Very High Resolusion Radiometer）は，表A.1に示す5チャネル（5バンド）より構成されている．各チャネルの1画素あたりのデータは10bitであり，1シーンあたりのデータ量は90MB~130MBとなる．チャネル1は可視，チャネル2は近赤外を観測しており，計測されるのは太陽光の地表面での反射であり，雲，雪，氷等の分布が得られる．チャネル2では，開水面での吸収が大きいため，陸と海，湖等との識別にも利用することができる．チャネル3~5はいわゆる大気の窓と呼ばれる，大気による吸収が小さい赤外領域の波長帯の観測を行い，地表面および大気の熱放射を計測し，地球上，特に海面，雲頂の温度分布を得ることができる．チャネル3は，太陽光反射の影響が大きく，昼間は利用できない．一方，チャネル4，チャネル5は大気中の水蒸気による減衰が大きく，その影響を補正する必要がある．

A.2.2 GMS S-VISSR データ

静止気象衛星GMS（Geostationary Meteoroloical Satellite）は世界気象監視（World Weather Watch : WWW）計画の一環として，日本により打ち上げられている衛星である．WWW計画では，地球全体を同時に観測できるように，GMS，GOES-E，GOES-W，METEOSAT，INSATの5つの衛星を約70度おきに赤道上空に配置しており，GMSは東経140度の赤道上空約35800kmの静止軌道上にある．1977年にGMS-1号が打ち上げられ，その後，1981

表 A.2: GMS-4 VISSR センサのチャネル構成

チャネル	主な利用法	波長	解像度
VIS	雲等の分布	0.5μm~0.75μm	1.25km
IR	雲等の分布	10.5μm~12.5μm	5km

表 A.3: GMS-5 VISSR センサのチャネル構成

チャネル	主な利用法	波長	解像度
VIS	雲等の分布	0.5μm~0.75μm	1.25km
IR1	雲等の分布	10.5μm~11.5μm	5km
IR2	雲等の分布	11.5μm~12.5μm	5km
IR3	水蒸気量	6.5μm~ 7.5μm	5km

年に GMS-2 号，1984 年に GMS-3 号，1989 に GMS-4 号が順次打ち上げられ，現在，1995 年に打ち上げられた GMS-5 号が稼働している．衛星は，おおよそ，北緯 60 度から南緯 60 度，東経 80 度から西経 160 度の範囲を 1 時間毎に観測する．VISSR データは衛星から一度地上の CDAS （Command and Data Aquisition Stataion）へ転送される．その後，簡易処理や各種文字情報の付加が行われて S-VISSR （Stretched VISSR）データとして GMS 衛星に転送され，その S-VISSR データが各地上局へ配信される．

GMS-4 号に搭載されている VISSR（Visible and Infrared Spin Scan Radiometer）は，表 A.2 に示されるように可視の波長帯の観測を行う VIS チャネルと，熱赤外の波長帯の観測を行う IR チャネルの計 2 チャネル（2 バンド）で構成される．一方，GMS-5 号に搭載されている VISSR では，表 A.3 に示すように，GMS-4 号の IR チャネルが 2 つの波長帯に分割されるとともに，水蒸気分布のための赤外チャネルがあらたに搭載され，VIS チャネル，IR1~IR3 チャネルの 4 チャネルより構成される．GMS-4 号，GMS-5 号のいずれにおいても，VIS チャネルの 1 画素のデータは 6bit，IR/IR1~IR3 チャネルは 8bit であり，1 シーンあたりのデータ量は約 100MB である．IR1，IR2 チャネルは衛星 NOAA 同様，大気の窓と呼ばれる帯域に設定されており，それぞれ衛星 NOAA のチャネル 4，チャネル 5 とほぼ同じ帯域であり，VIS チャネルとともに雲の分布等の観測に用いられる．IR3 チャネルは水蒸気による吸収の強い帯域に設定されており，水蒸気の分布の観測に用いられる．

A.2.3 Terra MODIS データ

地球観測衛星 Terra（EOS AM-1）は，EOS（Earth Observing System）計画の一環としてNASAにより打ち上げられた衛星であり，2000年に運用が開始された．Terraの軌道は，高度約705kmの太陽同期極軌道であり，衛星NOAAと近い軌道を持つ．衛星NOAAと同様，受信局付近を通過時にデータを取得することが可能であり，Terraは現在1機のみが運用されているため，1日あたりに受信されるデータは4シーン程度である．2002年にはTerraとペアをなす衛星Aqua（EOS PM-1）の打ち上げが予定されている．

Terraには，ASTER（Advanced Spaceborne Thermal Emission and Reflection Radiometer），CERES（Clouds and Earth's Radiant Energy System），MISR （Multi-angle Imaging Spectro Radiometer），MODIS（Moderate Resolution Imaging Spectro-Radiometer），MOPITT（Measurements of Pollution in the Troposphere）といった複数のセンサ群が搭載されている．MODISは，陸地，および海面上の生物学的，物理的現象を観測することを目的としたセンサであり，表A.4に示す解像度の異なる36バンドによって構成されている．MODISの各バンドの1画素あたりのデータは12bitであり，1シーンあたりのデータ量は900MB～1.4GBである．

A.3 システム構成

A.3.1 ハードウエア構成

図A.1は衛星画像データベースシステムのハードウエアの構成である．ファイルサーバ，テープアーカイブシステム，衛星データ受信システムは全て駒場キャンパスに設置されており，バックアップシステムが千葉実験所に設置されている．4GBの主記憶，8つの336MHz UltraSPARC-II CPUを備えたSUN社製Ultra Enterprise 6500をサーバとし，これにデータ格納用ディスクアレイ，テープアーカイブシステムが接続されている．ディスクアレイはSCSIまたはFibreChannelによって接続されており，合計1TB以上の容量を持つ．テープアーカイブシステムはStorage Technology社製SD-3テープドライブ装置1機，および2機の9840テープドライブ装置を備えたPowderHorn 9310ライブラリ装置1機，Exabyte社製EXB8505テープドライブを2機備えたNCLコミュニケーション社製

表 A.4: Terra MODIS センサのバンド構成

バンド	主な利用法	波長	解像度
1	Land/Cloud/Aerosols Boundaries	620nm～ 670nm	250m
2		841nm～ 876nm	
3	Land/Cloud/Aerosols Properties	459nm～ 479nm	500m
4		545nm～ 565nm	
5		1230nm～ 1250nm	
6		1628nm～ 1652nm	
7		2105nm～ 2155nm	
8	Ocean Color/Phytoplankton/Biogeochemistry	405nm～ 420nm	1000m
9		438nm～ 448nm	
10		483nm～ 493nm	
11		526nm～ 536nm	
12		546nm～ 556nm	
13		662nm～ 672nm	
14		673nm～ 683nm	
15		743nm～ 753nm	
16		862nm～ 877nm	
17	Atmospheric Water Vapor	890nm～ 920nm	
18		931nm～ 941nm	
19		915nm～ 965nm	
20	Surface/Cloud Temperature	3.660μm～ 3.840μm	
21		3.929μm～ 3.989μm	
22		3.929μm～ 3.989μm	
23		4.020μm～ 4.080μm	
24	Atmospheric Temperature	4.433μm～ 4.498μm	
25		4.482μm～ 4.549μm	
26	Cirrus Clouds Water Vapor	1.360μm～ 1.390μm	
27		6.535μm～ 6.895μm	
28		7.175μm～ 7.475μm	
29	Cloud Properties	8.400μm～ 8.700μm	
30	Ozone	9.580μm～ 9.880μm	
31	Surface/Cloud Temerature	10.780μm～11.280μm	
32		11.770μm～12.270μm	
33	Cloud Top Altitude	13.185μm～13.485μm	
34		13.485μm～13.785μm	
35		13.785μm～14.085μm	
36		14.085μm～14.385μm	

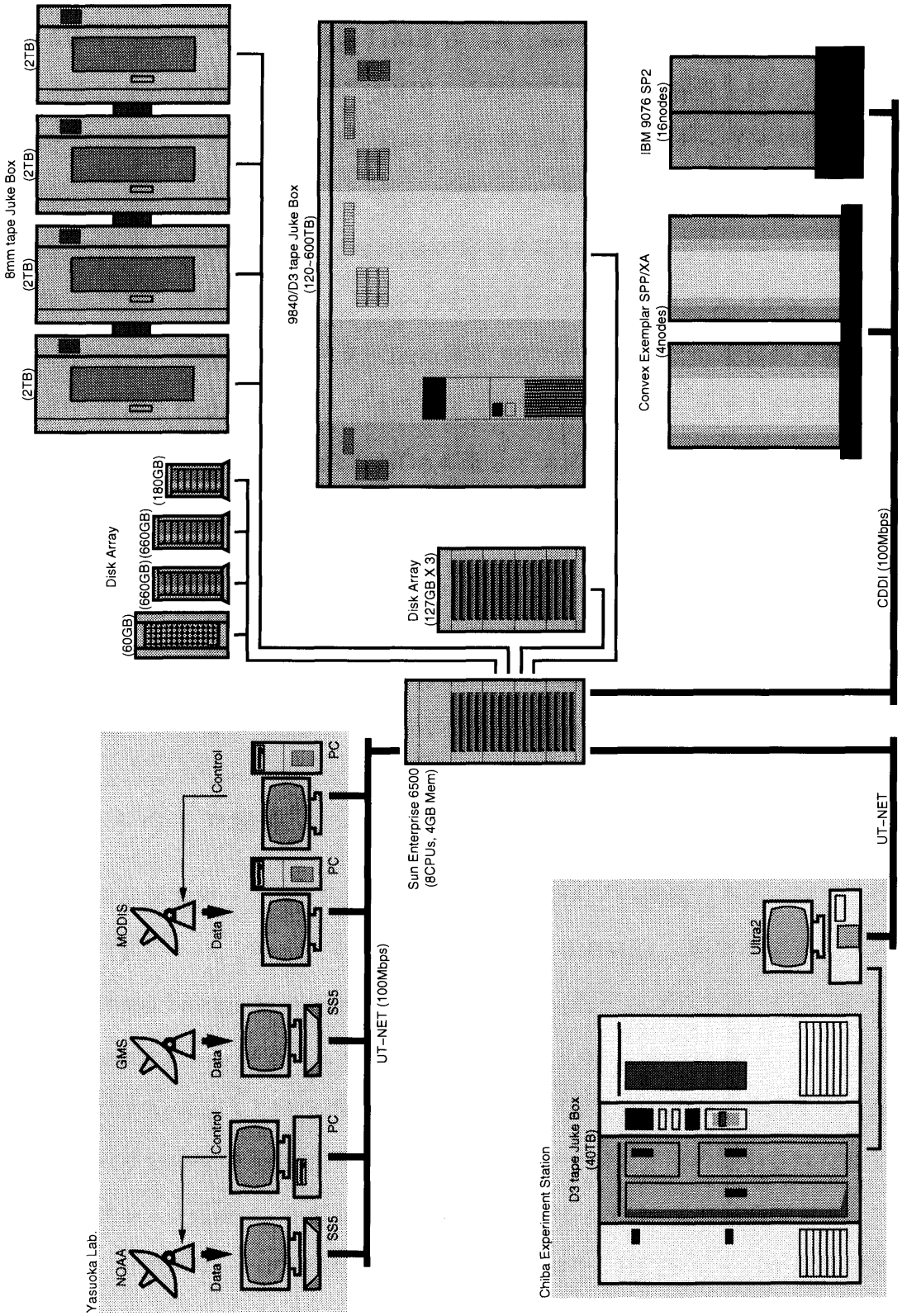

図 A.1: システム構成（ハードウエア）

NTH-200B ライブラリ装置4機である．4機の NTH-200B はテープ移送装置より接続されている．SD-3 テープドライブ装置は 11MB/秒（非圧縮時）の転送速度を持ち，テープメディア1本あたりの容量は非圧縮時で 50GB，圧縮時で 100GB の容量を持つ．一方，9840 テープドライブ装置は 10MB/秒（非圧縮時）の転送速度を持ち，テープメディア1本あたりの容量は非圧縮時で 20GB である．9310 ライブラリは約 6000 本のメディアを格納することが可能であり，従って，最大で 600TB の容量を持つ．

衛星データ受信システムとファイルサーバとは東京大学内 LAN により接続されており，NOAA 衛星データ，GMS 衛星データ，Terra 衛星 MODIS データとも受信終了直後にファイルサーバへ転送される．また，タイ AIT とも UT-NET，インターネットを経由して接続されており，AIT において受信された NOAA 衛星データもネットワークにより転送される．ファイルサーバに転送されたデータはまず，ディスク上に蓄えられ，転送完了直後にデータベースへ登録される．テープのマウント，シークに要する時間は，データの実際の書き込み時間に比べて大きく，従って1シーンあたりのマウント，シークのオーバーヘッドを減少させるために，一定期間分のディスク上の衛星データ数シーン分をまとめて一度にテープへ記録する．この処理は，リクエストの少ない夜間に実行している．

A.3.2 ソフトウエア構成

図 A.2 は，衛星画像データベースシステムのソフトウエアの構成である．ソフトウエアは大きく，HTTP サーバ，データベース管理システム（DBMS），階層記憶管理システム（HSM : Hierachical Storage Management），ダウンローダ/アップローダ，CGI（Common Gateway Interface）スクリプトにより構成される．

HTTP サーバは Apache 1.3.14 を用いており，クライアントからのリクエストの受信，クライアントへの検索結果の送信などを行う．また，ブラウジング用の縮小された衛星画像であるクイックルック画像は，一般に広く用いられている JPEG，または GIF フォーマットの画像であり，HTTP サーバの管理下に直接置かれており，HTTP サーバが直接画像を送信する．DBMS は Sybase 11.0.2 を用いている．DBMS は，各衛星画像のカタログデータベースを管理しており，ユーザの要求に従って検索を実行する．カタログデータベースについては，A.4 節において詳しく述べる．アップローダ/ダウンローダおよび HSM はテー

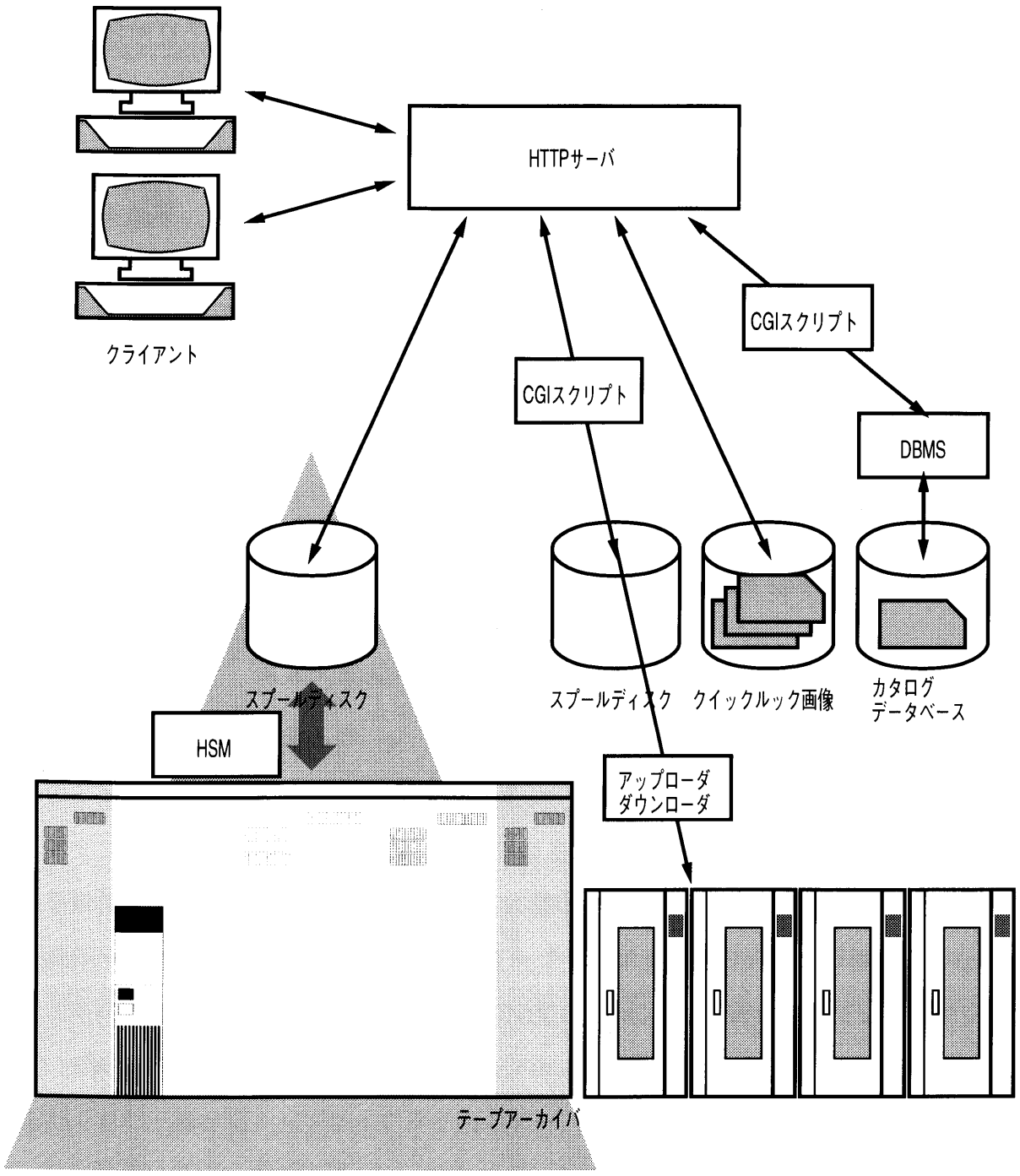

図 A.2: システム構成（ソフトウエア）

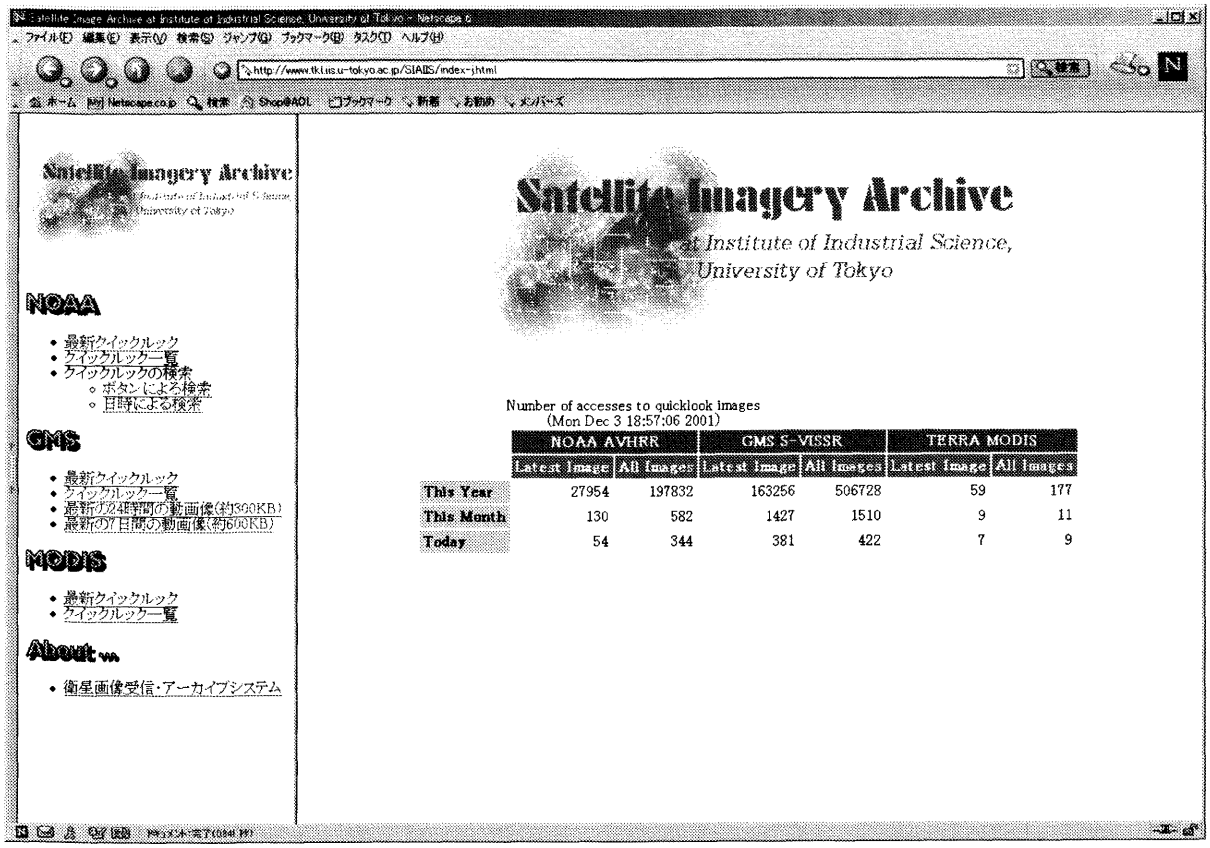

図 A.3: 衛星画像データベースの先頭ページ

プライブラリとディスク間のデータの移動を行う．データは，アップローダ/ダウンローダによる通常ファイルと，HSM によるファイルシステム上のファイルと二重に記録される．アップローダ/ダウンローダでは，データは通常のファイルとして管理され，アーカイバを操作するコマンド，および Solaris 2.6 に標準で提供されている mt，tar などのテープを操作するコマンドを用い，ディスク，テープライブラリ間のデータの移動を行なう．アップローダ/ダウンローダによるデータは UNIX 系 OS に標準で提供されているコマンドを用いて記録されており，特別なアプリケーションを必要としないため，データの可搬性が高く，また，細かくテープライブラリ内のデータを操作できるという利点がある反面，データの操作にはテープの状態，テープ上のデータのファイルの位置などを把握し，適切なパラメータを用いる必要がある．また，スプールディスク上に空き領域がない場合の不要なファイルの消去など，ファイルの管理も必要である．一方，HSM では，ファイルは HSM ソフトウエアによって提供されているファイルシステム上のファイルとして，一般

のディスク上のデータと同様にアクセスすることが可能であり，データのディスクとテープライブラリ間の移動は全てHSMソフトウェアが自動的に行う．しかしながら，HSMを用いた場合には，ブラックボックス化されたHSMソフトウエアが全てのファイルを管理することになるため，テープ上でのデータの配置やスプールディスク上のファイルのリプレースアルゴリズムの変更などが困難であるなど，細かな操作はできない．また，HSMソフトウエアのオーバーヘッドのため，アップローダ/ダウンローダによる場合と比べて若干転送時間が長くなる．さらに，HSMソフトウエアは，テープ上に独自フォーマットでデータを記録するため，HSMソフトウエアなしにデータを読み書きすることは困難である．このように，アップローダ/ダウンローダによる通常ファイルおよびHSMによるファイルにはそれぞれ利点があり，また，衛星データは一度失われると二度とのそのデータを作成，取得することができないため，異なる形態で二重にアーカイブしている．CGIスクリプトは，ユーザからのリクエストによりHTTPサーバから起動され，DBMS，アップローダ/ダウンローダとのインタフェースを受け持つ．ユーザにより指定された検索条件に従ったSQL文の作成とDBMSへの問い合わせ，指定されたデータに応じたアップローダ/ダウンローダの起動を行う．

A.4 カタログデータベース

A.4.1 スキーマ

図A.4，図A.6，図A.7はそれぞれ，NOAA衛星AVHRR画像データ，GMS衛星S-VISSR画像データおよびTerra衛星MODIS画像データのカタログデータベースのスキーマである．NOAA衛星画像では，衛星名，衛星進行方向（南 → 北，北 → 南），観測開始時刻，観測終了時刻，受信時のアンテナの最大仰角，および最大仰角となる時のアンテナの方位角，衛星画像データのファイル名，ファイルサイズ，クイックルック画像のファイル名，8mmテープアーカイバにおけるテープ番号，位置，D3テープアーカイバにおけるテープ名，位置，観測範囲，受信局名により構成される．NOAA衛星では，画像毎に観測されている領域が異なるため，図A.5に示される，受信された画像の観測開始時，中間時，終了時における，走査開始点，直下点，走査終了点の9地点の緯度経度もデータベース化し

```
create table hrpt
(
    satname varchar(8),          -- 衛星名
    direction bit,               -- 衛星進行方向
    starttime datetime,          -- 観測開始時刻
    endtime datetime,            -- 観測終了時刻
    elevation real,              -- アンテナ最大仰角
    azimuth real,                -- アンテナ方位角（仰角最大時）
    orgfile varchar(32),         -- 画像ファイル名
    filesize int,                -- 画像ファイルサイズ
    qlfile varchar(32),          -- クイックルック画像ファイル名
    nthnum int,                  -- 8mm テープ番号
    nthpos int,                  -- 8mm テープ位置
    stkvol varchar(8),           -- D3 テープ番号
    stkpos int,                  -- D3 テープ位置
    tl_latitude real,            -- 観測範囲（緯度，経度）
    tl_longitude real,
    tc_latitude real,
    tc_longitude real,
    tr_latitude real,
    tr_longitude real,
    cl_latitude real,
    cl_longitude real,
    cc_latitude real,
    cc_longitude real,
    cr_latitude real,
    cr_longitude real,
    bl_latitude real,
    bl_longitude real,
    bc_latitude real,
    bc_longitude real,
    br_latitude real,
    br_longitude real,
    station varchar(8),          -- 受信局名
    primary key nonclustered (orgfile)
)
```

図 A.4: NOAA HRPT 画像データのカタログのスキーマ

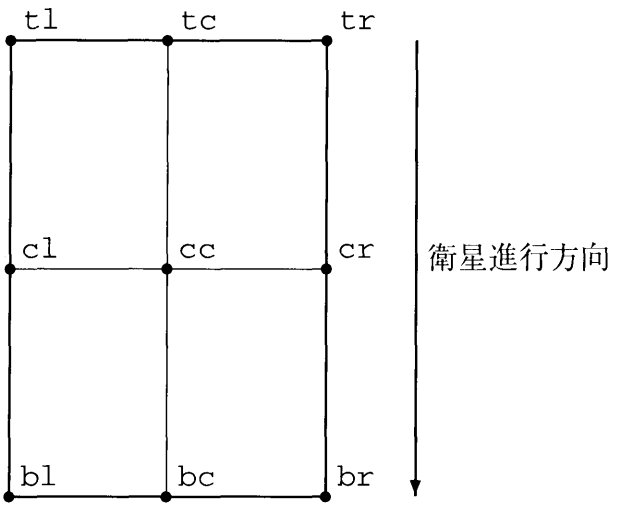

図 A.5: NOAA 画像の観測範囲指定点

```
create table vissr
(
    satname varchar(8),              -- 衛星名
    rectime datetime,                -- 観測時刻
    filebase varchar(32),            -- 画像ファイル名（ベース部）
    filesuffix varchar(8),           -- 画像ファイル名（拡張子部）
    filesize int,                    -- 画像ファイルサイズ
    qlfile varchar(32),              -- クイックルック画像ファイル名
    nthnum int,                      -- 8mm テープ番号
    nthpos int,                      -- 8mm テープ位置
    stkvol varchar(8),               -- D3 テープ名
    stkpos int,                      -- D3 テープ位置
    unique nonclustered (filebase, filesuffix)
)
```

図 A.6: GMS VISSR 画像データのカタログのスキーマ

```
create table modis
(
    satname varchar(32),          -- 衛星名
    starttime datetime,           -- 観測時刻
    l0file varchar(32),           -- Level 0 ファイル名
    l0filesize int,               -- Level 0 ファイルサイズ
    l1bfile_1000m varchar(32),    -- Level 1b(1000m) ファイル名
    l1bfilesize_1000m int,        -- Level 1b(1000m) ファイルサイズ
    l1bfile_500m varchar(32),     -- Level 1b(500m) ファイル名
    l1bfilesize_500m int,         -- Level 1b(500m) ファイルサイズ
    l1bfile_250m varchar(32),     -- Level 1b(250m) ファイル名
    l1bfilesize_250m int,         -- Level 1b(250m) ファイルサイズ
    l1bfile_geo varchar(32),      -- Level 1b(幾何情報) ファイル名
    l1bfilesize_geo int,          -- Level 1b(幾何情報) ファイルサイズ
    qlfile varchar(32),           -- クイックルック画像ファイル名
    station varchar(8),           -- 受信局名
    primary key nonclustered (l0file)
)
```

図 A.7: Terra MODIS 画像データのカタログのスキーマ

ている．GMS 画像に関しては，静止衛星であるためにアンテナは固定され，観測範囲も常に同じであるため，これらの情報のカラムはない．また，NOAA 画像では，全てのセンサの画像と付加情報は 1 つのファイルにまとめられているため 1 シーン当たり 1 ファイルであるが，GMS 画像ではセンサ毎に別々のファイルに分割されており 1 シーン当たり 3 ファイル（GMS-4）または 5 ファイル（GMS-5）になる．それらを区別するためにファイル名をベース部と拡張子部に分けており，ベース部はシーン毎の名称が，拡張子部にはセンサ毎の名称が与えられている．MODIS 画像では，Level 0 と呼ばれる原データと，各解像度毎に Level 1b と呼ばれる補正画像データのファイル名，ファイルサイズにより構成される．

A.4.2 データ検索

SQL インタプリタによってカタログデータベースをアクセスし，データを検索することは可能であるが，衛星画像の利用者は必ずしもデータベース管理システム，SQL 言語

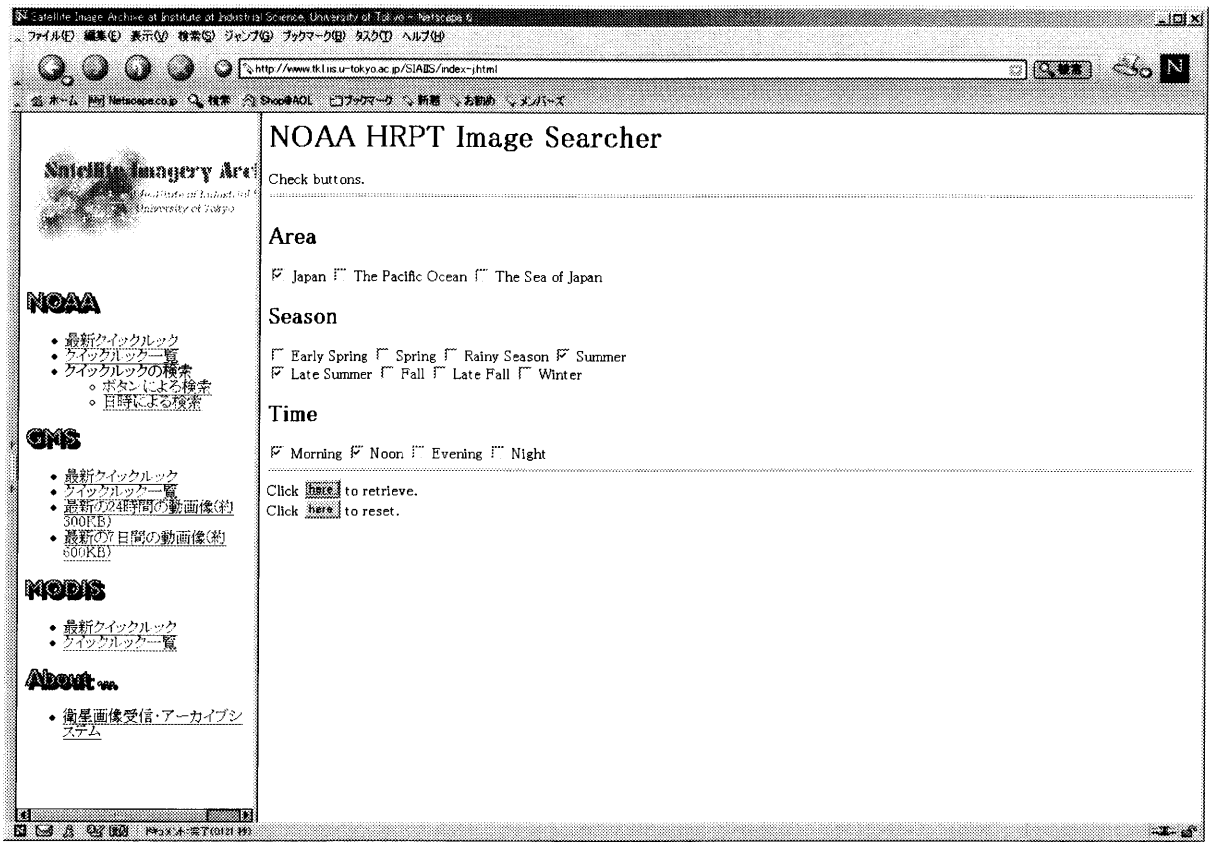

図 A.8: WWW による検索条件入力画面（条件ボタン）

に関しての知識を十分にもっているとは限らず，SQL を全く知らないという利用者も少なくない．このため，SQL を知らなくても容易に検索が可能となるように WWW を利用した検索システムを用意している．

図 A.8 は，WWW を利用した，観測された地域（日本，太平洋，日本海），季節，時間帯に関して，それぞれを大まかに分類したボタンを用意し，検索対象とするボタンを選択することによって NOAA 画像を検索するためのインタフェースである．各ボタンの検索条件を表 A.5 に示す．地域，季節，時間帯の各条件群内で複数のボタンが押された場合は，それらは OR により接続される．一方，各条件群間は AND で接続される．すなわち，図 A.8 では，日本中心が中心の画像であり，かつ，夏または晩夏の画像であり，かつ，朝または昼に観測された画像の検索が指定されている．retrieve ボタンをクリックすることにより，CGI スクリプトが起動され，指定された検索条件に対応した SQL 文が作成されて DBMS へ問い合わせが行われ，条件に合致する画像のリストが DBMS から CGI スクリプ

表 A.5: 条件ボタンの検索項目の内容

コマンドボタン	意味	検索条件
The Pacific Ocean	太平洋中心の画像	アンテナ最大仰角 $< 35°$ かつ その時の方位角 $< 100°$
Japan	日本中心の画像	アンテナ最大仰角 $> 60°$
The Sea of Japan	日本海中心の画像	アンテナ最大仰角 $< 35°$ かつ その時の方位角 $> 290°$
Early Spring	早春の画像	3 月 21 日 ~4 月 7 日
Spring	春の画像	4 月 8 日 ~5 月 24 日
Rainy Season	梅雨の画像	5 月 25 日 ~7 月 20 日
Summer	夏の画像	7 月 21 日 ~9 月 7 日
Late Summer	晩夏の画像	9 月 8 日 ~10 月 5 日
Fall	秋の画像	10 月 6 日 ~11 月 15 日
Late Fall	晩秋の画像	11 月 16 日 ~12 月 15 日
Winter	冬の画像	12 月 16 日 ~3 月 20 日
Morning	朝の画像	5 時 ~9 時
Noon	昼の画像	9 時 ~16 時
Evening	夕方の画像	16 時 ~20 時
Night	夜の画像	20 時 ~5 時

トへ返される．その後，CGI スクリプトは検索結果を HTML へ変換して HTTP サーバへ送り，さらに HTTP サーバがクライアントへ転送する．

図 A.9 は観測日時，衛星名，受信局，および観測された地点の緯度・経度を指定して検索を行うためのインタフェースである．観測年，月，日，時，衛星名，受信局，観測された地点の緯度，経度を指定するフィールドが用意されており，このフィールドへ数値等を入力することで検索条件を指定する．観測年，月，日，時に関しては，範囲を指定することも可能である．フィールド内を空欄とした場合，その条件を指定しないことを意味する．また，各フィールドは AND で結ばれる．例えば，図 A.9 では，日本標準時で 1999 年または 2001 年の 8 月の 20 日以降の 6 時から 18 時に生産技術研究所（IIS）において観測された画像で，かつ，北緯 35 度，東経 140 度の地点が観測された画像が指定されている．衛星名は指定されていないため，全ての衛星による画像が検索対象となる．Sybase 11.0.2 では空間情報の検索をサポートしておらず，ある領域に指定された点が含まれているかどうかの判断を DBMS で行うことは困難である．従って，検索はまず，日時，衛星名，受信局の条件が CGI スクリプトにより SQL 文に変換され，DBMS へ問い合わせが行われる．その後，CGI スクリプト自身が，DBMS からの検索結果として得られる観測範囲指定点の情報により，利用者に指定された地点が観測範囲内に存在するかどうかを判断する．そ

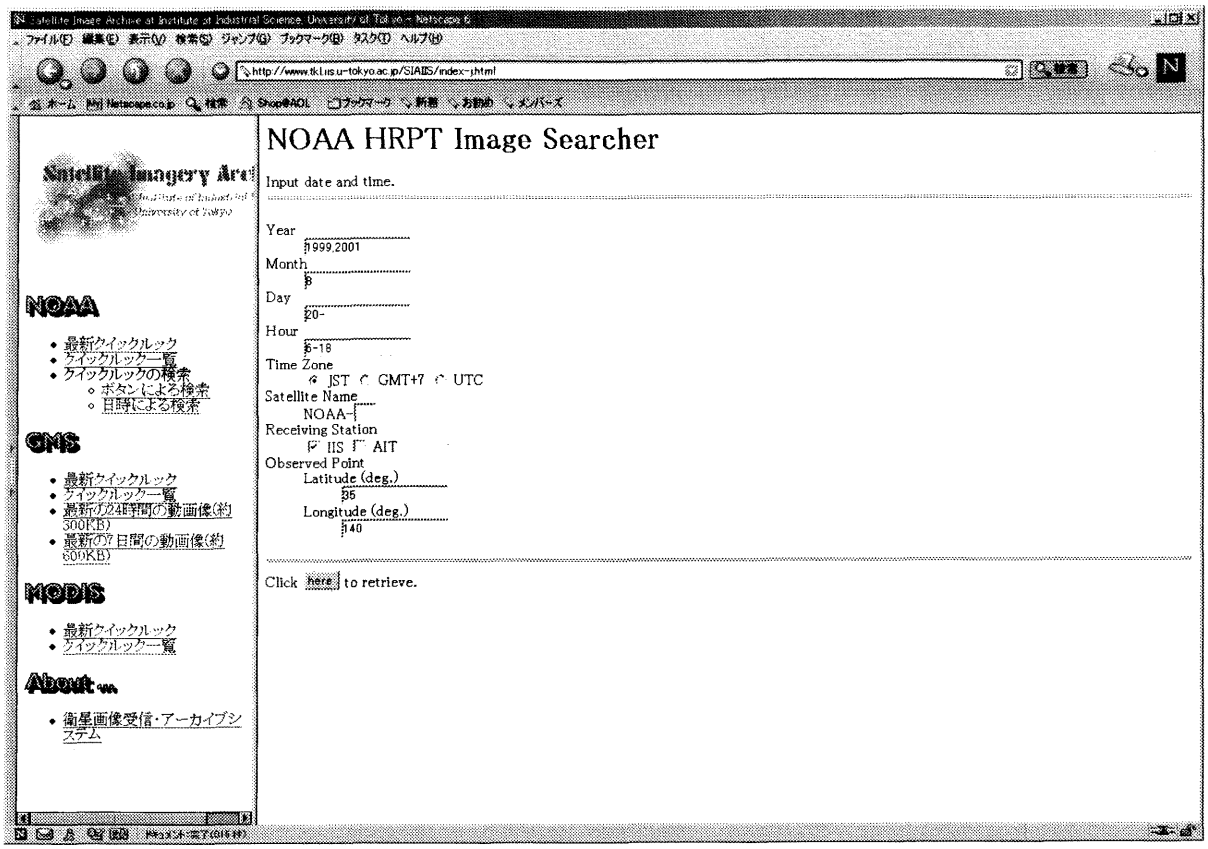

図 A.9: WWW による検索条件入力画面（日時，観測地点指定）

の後，CGI スクリプトにより条件に合致した画像のリストが HTML へ変換されて HTTP サーバへ送られ，さらにクライアントへ転送される．

利用者には指定された条件に合致した画像のリストが提示され，リスト内のアイテムを選択することで，図 A.10，図 A.11，図 A.12 に示されるように，観測日時，縮小画像，観測範囲画像等の情報が表示される．ここで，図 A.10，図 A.11，図 A.12 の原画像名をクリックすることにより，自動的にテープ上のデータがディスクアレイ上へマイグレートされ，さらにクライアントへデータが転送される．この他に，複数の衛星画像のディスクアレイ上への一括転送，ディスクアレイ上のスプール内の不要となった衛星データのユーザによる明示的な消去などが WWW のインタフェースを通して可能となっている．

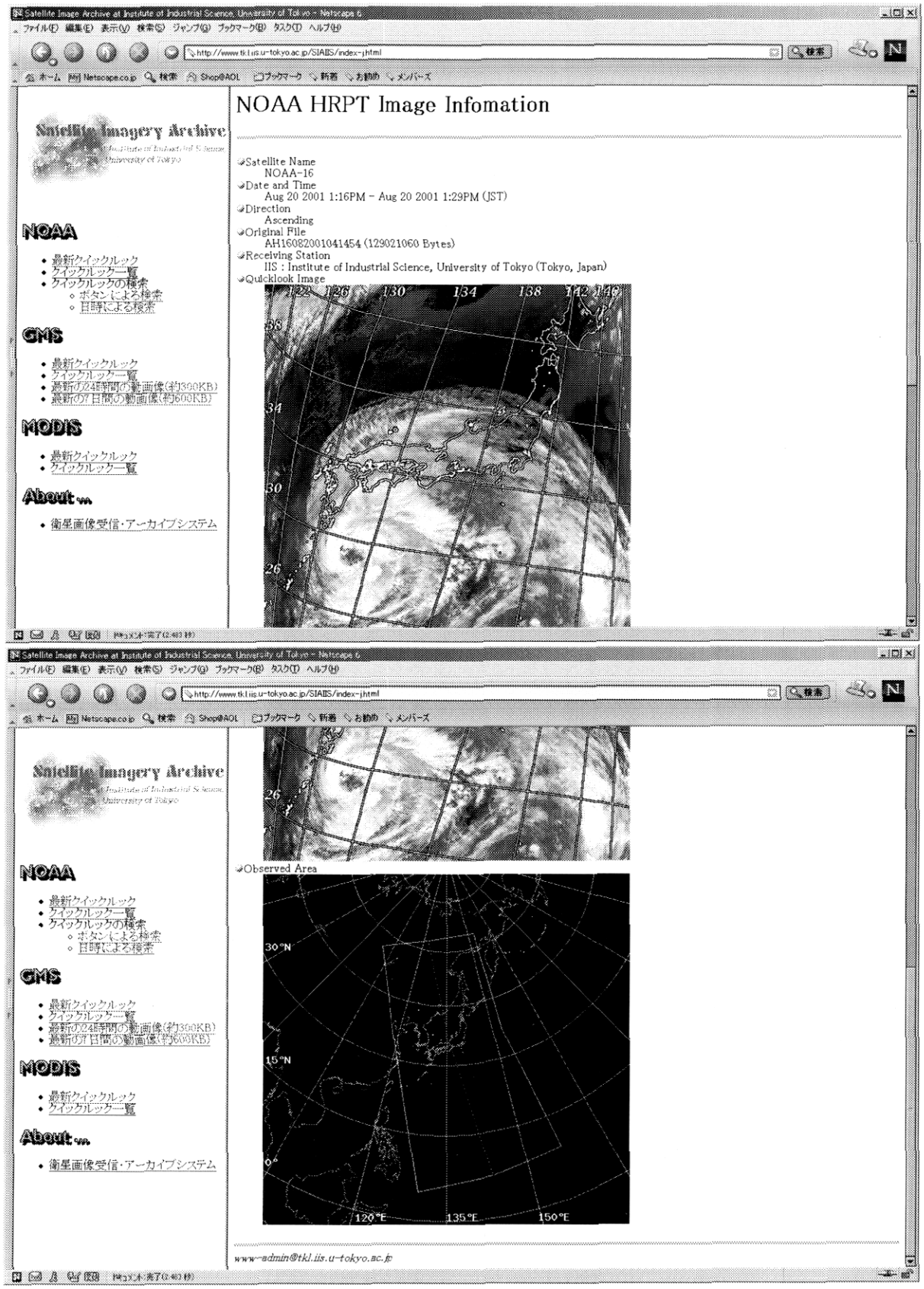

図 A.10: WWW による NOAA 画像情報表示画面

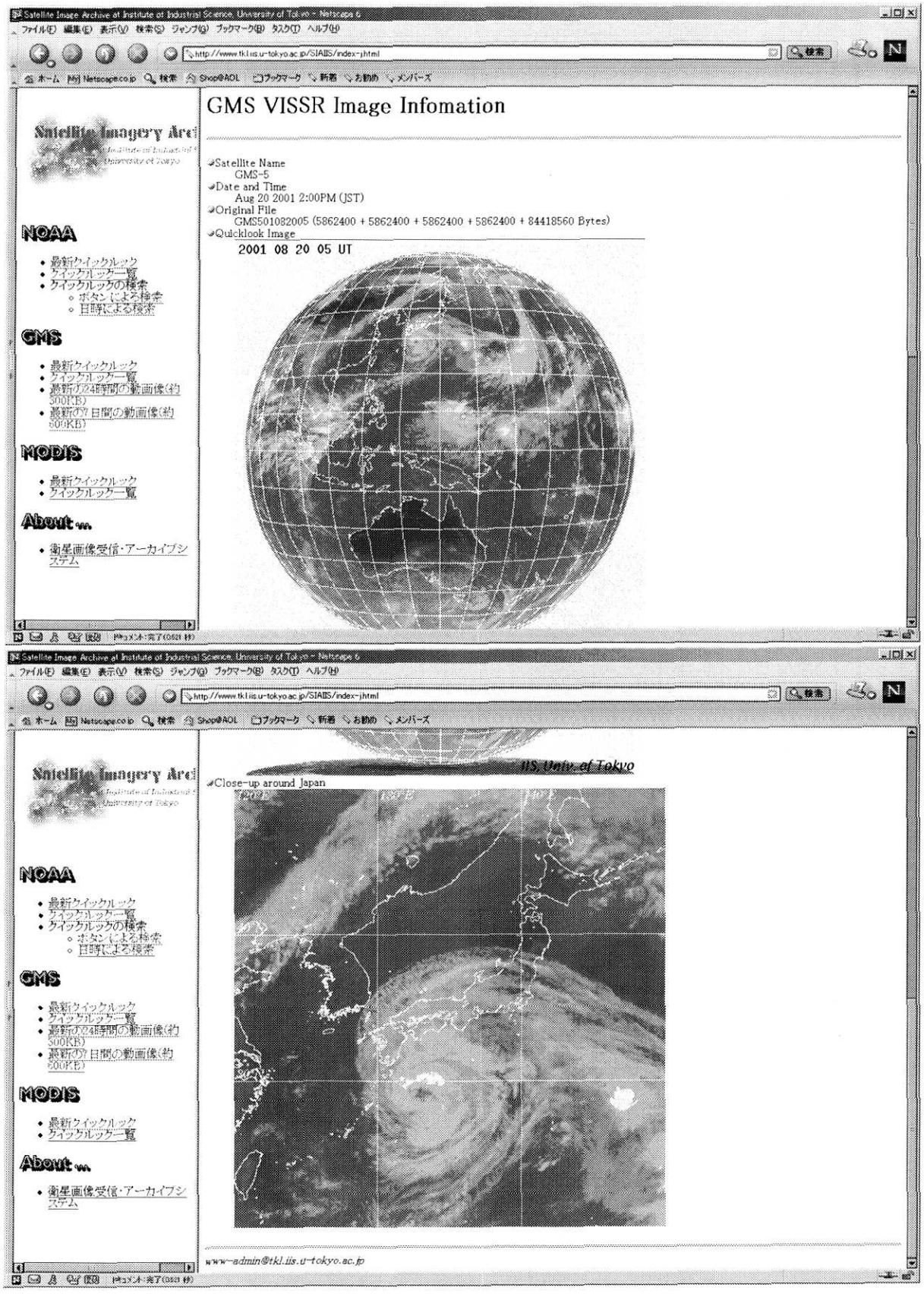

図 A.11: WWW による GMS 画像情報表示画面

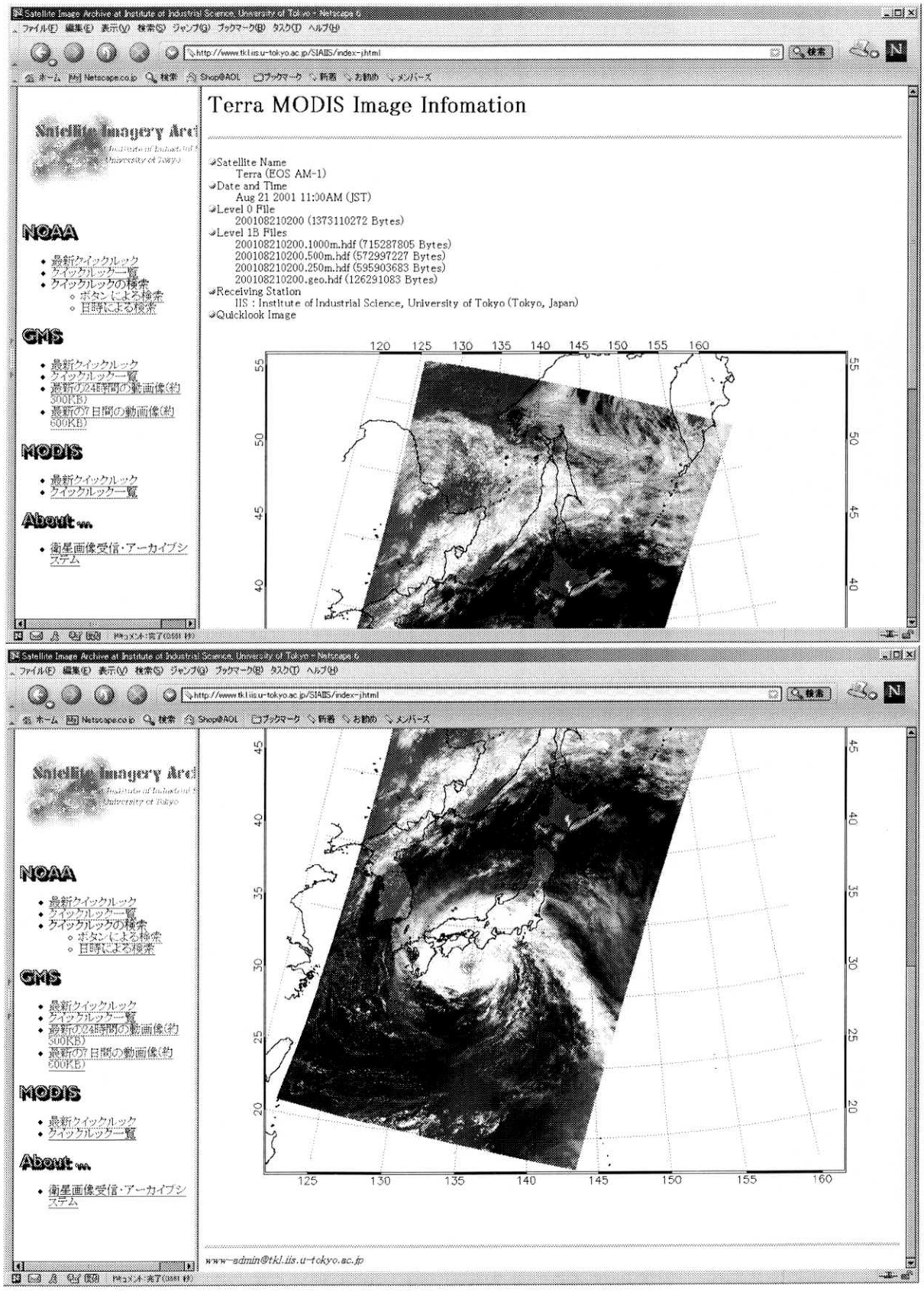

図 A.12: WWW による MODIS 画像情報表示画面

発表文献

学会論文誌

- 根本利弘, 喜連川優. “ホットレプリケーション：三次記憶システムにおける高アクセス頻度データの複製クラスタリング手法”. 情報処理学会論文誌. 投稿中.

- 根本利弘, 喜連川優. “スケーラブルテープアーカイバにおけるテープマイグレーションを用いた負荷分散手法とその性能評価”. 電子情報通信学会論文誌, Vol. J82-D-I, No. 1, pp. 53–69, January 1999.

- 根本利弘, 迫和彦, 喜連川優, 高木幹雄. “部分マイグレーション機構を有する三次記憶ファイルシステム PFS の 8 ミリテープアーカイブ装置への実装とその性能評価”. 電子情報通信学会論文誌, Vol. J81-D-I, No. 6, pp. 747–754, June 1998.

国際会議

- T. Nemoto and M. Kitsuregawa. “Scalable tape archiver for satellite image database and its performance analysis with access logs – hot declustering and hot replication –”. In *Proceedings of 16th IEEE Symposium on Mass Storage Systems in cooperation with the 7th NASA GSFC Conference on Mass Storage Systems and Technologies*, pp. 59–71, San Diego, California, March 1999.

- T. Nemoto, and M. Kitsuregawa. “Effectiveness of tape migration for satellite image database”. In *Proceedings of International Symposium on Remote Sensing*, pp. 52–57, Pusan, Korea, November, 1997.

- T. Nemoto, M. Kitsuregawa, and M. Takagi. “Simulation studies of the cassette migra-

tion activities in a scalable tape archiver". In *Proceedings of The Fifth International Conference on Database Systems for Advanced Applications*, pp. 461–470, Melbourne, Australia, April 1997.

- T. Nemoto, M. Kitsuregawa, and M. Takagi. "Scalable tape archiver and its application to satellite image database". In *Proceedings of International Symposium on Remote Sensing*, pp. 185–189, Cheju, Korea, October 1996.

- T. Nemoto, M. Kitsuregawa, and M. Takagi. Design and implementation of scalable tape archiver. In *Proceedings of Fifth NASA GSFC Conference on Mass Storage Systems and Technologies*, pp. 229–237, Collage Park, Maryland, September 1996.

- K. Sako, T. Nemoto, M. Kitsuregawa, and M. Takagi. "Partial migration in an 8mm tape based tertiary storage file system and its performance evaluation through satellite image processing applications". In *Proceedings of 6th International Conference on Information Systems and Management of Data*, pp. 178–190, Bombey, India, 1995.

- T. Nemoto, Y. Sato, K. Mogi, K. Ayukawa, M. Kitsuregawa, and M. Takagi. "Performance evaluation of cassette migration mechanism for scalable tape archiver". In *SPIE, Digital Image Storage and Archiving System*, pp. 48–58, Philadelphia, Pennsylvania, October 1995.

- T. Nemoto, K. Sako, M. Kitsuregawa, and M. Takagi. "Partial migratable hierarchical file system for satellite image database". In *Proceedings of International Symposium on Remote Sensing*, Taejon, Korea, October 1995.

- S. Takamura, T. Nemoto, and M. Takagi. "Remote sensing data receiving and processing". In *Proceedings of International Seminar on Space Informatics and Sustainable Development: Grassland Monitoring and Management*, 1995.

研究会・シンポシウム等

- 根本利弘, 喜連川優. “DVD-RAM ドライブを用いたスケーラブルアーカイバにおけるホットデクラスタリングの性能評価”. 第 12 回データ工学ワークショップ, March 2001.
- 根本利弘, 喜連川優. “テープアーカイブシステムにおけるホットレプリケーションの性能評価”. 電子情報通信学会技術研究報告, Vol. 100, No. 226, pp. 105–112, 2000.
- 向井景洋, 根本利弘, 喜連川優. “高性能ディスクにおけるアクセスプランを用いたプリフェッチ機構に関する評価”. 第 11 回データ工学ワークショップ, March 2000.
- 根本利弘, 喜連川優. “衛星画像データベースのアクセス履歴を用いたスケーラブルテープアーカイバの性能評価”. 第 9 回データ工学ワークショップ, March 1998.
- 根本利弘, 喜連川優, 高木幹雄. “ネットワークによる衛星画像データアーカイブシステムの利用法”. 生研フォーラム「宇宙からの地球環境モニタリング」, pp. 97–106, February 1997.
- 根本利弘, 喜連川優, 高木幹雄. “リモートセンシング画像データベースの構築とその利用法”. 生研フォーラム「宇宙からの地球環境モニタリング」, pp. 10–19, November 1996.
- 根本利弘, 喜連川優, 高木幹雄. “スケーラブルテープアーカイバにおけるカセットマイグレーション方式とその評価”. 並列処理シンポジウム JSPP’96 論文集, pp. 275–282, June 1996.
- M. Kitsuregawa, T. Nemoto, and M. Takagi. File management and migration on scalable tape archiver. In *Proceedings of the Second SEIKEN Symposium on Global Environmental Monitoring from Space*, pp. 17–25, February 1996.
- 根本利弘, 迫和彦, 喜連川優, 高木幹雄. “衛星データの高速アクセスを目的とした階層ファイルシステムの設計”. 生研フォーラム「宇宙からの地球環境モニタリング」, pp. 104–111, June 1995.

- 根本利弘, 迫和彦, 喜連川優, 高木幹雄. “衛星画像の格納を目的とした大規模階層ファイルシステムの設計”. 情報処理学会研究報告, Vol. 95, No. 65, pp. 65–71, 1995.

学会全国大会

- 根本利弘, 喜連川優. “衛星画像データベースのアクセス履歴を用いたホットレプリケーションの評価”. 電子情報通信学会 1998 年情報・システムソサイエティ大会講演論文集, 1998. D-4-8.

- 根本利弘, 喜連川優. “衛星画像データベースのアクセス履歴を用いたテープアーカイバにおけるメディアマイグレーション機構の性能評価”. 情報処理学会第 56 回全国大会講演論文集, 1998. 6Aa-8.

- 根本利弘, 喜連川優, 高木幹雄. “スケーラブルテープアーカイバにおけるテープ上での動的データ再配置”. 情報処理学会第 55 回全国大会講演論文集, 1997. 4AC-6.

- 根本利弘, 喜連川優, 高木幹雄. “大規模テープアーカイバにおけるデータ再配置手法の検討”. 情報処理学会第 54 回全国大会講演論文集, 1997. 3R-4.

- 根本利弘, 喜連川優, 高木幹雄. “スケーラブルテープアーカイバを用いた大規模ファイルシステムにおけるファイル編成方式の検討”. 情報処理学会第 53 回全国大会講演論文集, 1996. 1R-6.

- 根本利弘, 喜連川優, 高木幹雄. “三次記憶システムにおけるファイル編成に関する一考察”. 情報処理学会第 52 回全国大会講演論文集, 1995. 1Y-3.

- 鮎川健一郎, 根本利弘, 喜連川優, 高木幹雄. “大規模テープ・アーカイバにおけるマイグレーションによる負荷分散”. 情報処理学会第 52 回全国大会講演論文集, 1995. 3Y-9.

- 根本利弘, 迫和彦, 喜連川優, 高木幹雄. “部分マイグレーション機能を有する大規模階層ファイルシステムの試作”. 情報処理学会第 51 回全国大会講演論文集, 1995. 5L-4.

- 鮎川健一郎, 根本利弘, 茂木和彦, 喜連川優, 高木幹雄. “大規模テープ・アーカイバにおけるマイグレーションのシミュレーションによる評価”. 情報処理学会第 51 回全国大会講演論文集, 1995. 4D-8.

- 根本利弘, 迫和彦, 喜連川優, 高木幹雄. “衛星データを対象とした超大規模画像データベースの構想”. 情報処理学会第 50 回全国大会講演論文集, 1995. 2F-8.