

大容量連続メディアデータの管理方式と性能評価

4 R - 3

小久保 成人* 山田 雅彦* 北村 隆* 五味 弘* 保田 真人** 長坂 篤**

*(株) 沖テクノシステムズ ラボラトリ

**沖電気工業(株)

1 はじめに

大容量かつ連続的なメディアデータに対して同時に多数のアクセスを行うときに高速応答が可能なディスク管理方式を提案し、その実現について報告する。また、この管理方式による応答速度の性能評価を行い、考察を述べる。

2 ビデオ・オン・デマンド システム (VOD)

VOD システムは大容量かつ連続的なメディアデータである動画を多数同時に実時間でアクセスするシステムである。この VOD システムで動画を効率よく管理する方式が必要とされている。

3 管理方式

VOD で連続した大容量データ (動画・音声等) を効率よく管理するために、以下に示す単位でデータを細分化して管理する。また、多重アクセスに対して細分化されたデータを効率よく配送する方式について述べる。

3.1 連続メディアデータの処理に関する管理単位

● CMSegment

MPEG ストリームなどの連続メディア全体を CM-Segment に分割し管理する。また、同時アクセスの負荷分散のために動的に連続メディアを複製する必要があり、その管理のためにも CMSegment を使用する。ディスクへのアクセスは CMSegment に対して行われ、この CMSegment に複数の Extent (後述) を保持させることにより、動的複製による負荷分散の機能に対応する。動的複製については後述する。

● ContinuousMedia

MPEG ストリームなどの連続メディア全体を管理するために ContinuousMedia という単位を設けた。図 1 に連続メディアデータの処理に関する管理単位の概要を示す。

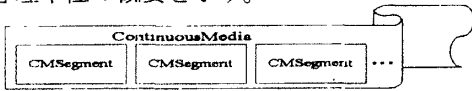


図 1: 連続メディアデータの管理

3.2 保管場所に関する管理単位

● Block

データを扱う最小の単位で、1Block のサイズは 376kbyte である。

● Extent

大容量の記憶領域を論理的に管理し、要求に応じて読み出し、書き込みを行う必要がある。その単位として Extent を設けた。Extent は通常 32Block 分のディスク上の連続した領域を確保しており、メディアデータの原本保管領域、動的複製領域、DVD キャッシュ領域はこの Extent 単位で割り当てられる。

● HardDisk

ハードディスク 1 台を管理するために HardDisk という単位を設けた。HardDisk は複数の Extent を保持し、この HardDisk を通して Block に対する読み込み、書き込みの処理を行う。

● StripingLogicalDisk

大容量データを保管し、多数のアクセスに対して均等に配送処理を行うには、後述のストライピング処理が必要になる。その機能を実現するために StripingLogicalDisk という単位を設けた。StripingLogicalDisk では、複数の HardDisk を保持し、それらを用いたストライピング処理を行う。

● VolumeSet

VolumeSet はディスク装置を保持し、記憶領域を論理的に管理する単位として設けた。VolumeSet は HardDisk または StripingLogicalDisk を保持し、それらのディスク上の領域を Extent 単位で管理している。

● CMDomain

複数の連続メディアを保管し、これに対するアクセス手段を提供するために CMDomain という単位を設けた。CMDomain では連続メディアの保管に使用する VolumeSet の集合を管理する。図 2 に保管場所に関する管理単位の概要を示す。

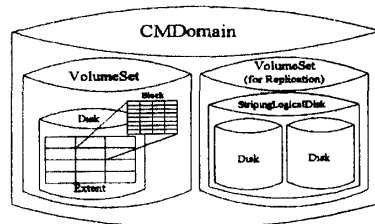


図 2: 保管場所の管理

"A Management Scheme of Large Continuous Media Data and its Performance Evaluation"

Naruhito KOKUBO*, Masahiko YAMADA*, Takashi KITAMURA*, Hiroshi GOMI*, Makoto HOTA**, Atsushi NAGASAKA**

*Oki Technosystems Laboratory, Inc.

**Oki Electric Industry Co., Ltd.

4 負荷分散方式

VOD で人気のあるタイトルにアクセスが集中すると、特定の資源に対して負荷がかかり、データの配送効率に影響が出る。本章ではその負荷を分散する方式について述べる。

4.1 ストライピングによる負荷分散

StripingLogicalDisk 内で保持している HardDisk に対してストライピング処理を行う。ストライピングの単位は Block 単位で、保持している HardDisk の順番に 1Block ずつの書き込み、あるいは読み込みの処理を行う。この処理により、同一ストリームに対する多重アクセス時に一つのディスクにアクセスが集中するのを防ぐことができる。図3にストライピング処理の概要を示す。

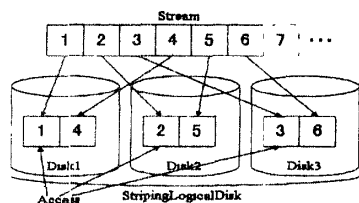


図3: ストライピング処理

4.2 動的複製による負荷分散

1つの CMSegment に対してアクセスが集中すると、結果的に1台の HardDisk に対しての負荷が増大しデータの転送効率が悪化する。この現象を防ぐため同一 CMSegment にアクセスが集中する前に、それを予測し事前に CMSegment の分身となる新たな Extent を別の HardDisk 上に複製する。この処理後の新たな CMSegment へのアクセスは複製された Extent に割り振られ、同一 HardDisk に対するアクセスの集中を防止することができる。図4に動的複製を行う条件を示すためにセグメントの状態を示す。

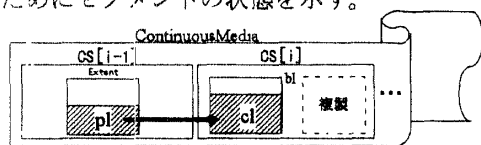


図4: 動的複製開始時の状態

CS[i-1] から CS[i] にアクセスが移動した場合に増える負荷を $pl(i=0$ の時 $pl=0)$ 、CS[i] が保持する Extent が属する VolumeSet の閾値の合計を bl 、CS[i] が保持する Extent が属する VolumeSet の現在の負荷の合計を cl とすると、動的複製は以下のような条件を満たすときに実行される。

$$(bl \leq cl + pl) \text{ かつ}$$

(複製が作成可能な VolumeSet が存在する) かつ
(動的複製中でない)

5 性能評価

前章の管理方式を用いた性能評価を行った。それぞれの評価は OKITAC9000/715/50、2Gbyte ハードディスク (SCSI2)2 台で実施した。

5.1 ストライピングにおける性能評価

2 台の HardDisk で StripingLogicalDisk を作成し読み出し要求の負荷をかけ、サーバーのデータ転送速度を求めた。

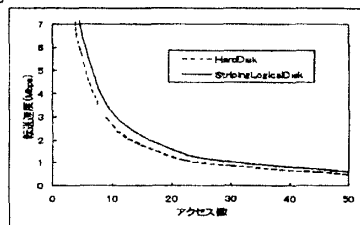


図5: ストライピングにおける性能評価

1 台の HardDisk と比較して 2 台の HardDisk をストライピングして動作させると転送速度の落ち込みが緩やかであることが分かる。今回の計測は同じ SCSI インターフェースに 2 台のディスクを接続したが、異なるインターフェースを用いた場合、その差がより明確になる。

5.2 動的複製における性能評価

1 台の HardDisk を原本用、もう 1 台を複製用とし動的複製が行われる状況を作り、アクセスを 3CMSegment にわたって移動した場合の HardDisk の負荷の状況とサーバーの転送速度を求めた。

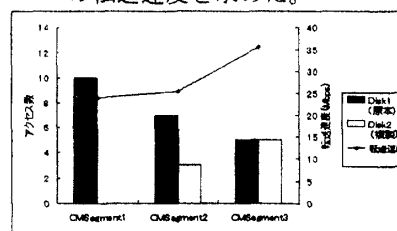


図6: 動的複製における性能評価

CMSegment1 では 1 台の HardDisk にアクセスが集中し転送速度が落ちていたが、CMSegment2 で少し負荷が分散され転送速度が上がっている（この時点では動的複製中に移動するアクセスもあるので完全に負荷は分散されていない）。複製の完了した CMSegment3 で完全に負荷は分散し、転送速度が上がっている。

6 おわりに

大容量かつ連続的なメディアデータに対して同時に多数のアクセスを行うときに高速応答が可能なディスク管理方式を提案した。応答速度の性能評価の結果、その効果を実証することができた。またこの方式を採用した OKI MediaServer で実際の有効性を確認できた。今後の課題としては、シークレベルでの応答速度の向上を行い、データ転送能力を強化する予定である。

参考文献

- Asit Dan and Martin Kienzle and Dinkar Sitaram : "A dynamic policy of segment replication for load-balancing in video-on-demand servers", Multimedia Systems, 3(3)93-103 July 1995
- 長坂 篤, 新谷 義弘, 中松 芳樹, 日比 孝, 畠中 啓, 保田 真人: DAVIC 規格準拠メディアサーバ (OKI MediaServer) の開発, 電子情報通信学会 MVE96-58, 1996
- 久保田 俊郎, 保田 真人, 新谷 義弘, 長坂 篤: DVD チェンジャを用いた低コスト大容量ビデオファイルシステムの構築, 情報処理学会 4R-02, 1997