

ストレージ階層仮想化機能におけるデータ配置先決定法の研究 Determining Allocation Tiers in Virtual Storage Tiering Functions

大平 良徳†
Yoshinori Ohira

今崎 美保†
Miho Imazaki

須藤 梓†
Azusa Sudo

塚 弘明†
Hiroaki Akutsu

江口 賢哲†
Yoshiaki Eguchi

1. はじめに

データ量が年々増加する一方で、IT投資額の増加が見込めない状況が続いている。このため、保有データをアクセス頻度に応じた階層（高速なメディア（記憶媒体）や安価で低速なメディア）に配置することで、コストを削減する方式が注目されている。

これに対し我々は、ストレージ領域のアクセス頻度の偏りに着目し、データのアクセス頻度の採取と、適切な階層への再配置をストレージ装置が実行する、ストレージ階層仮想化機能を開発した。本研究では、複数階層を組み合わせた構成にて、スループット性能を最大化するデータ配置先決定法を提案し、その評価結果を述べる。

2. 背景と目的

ある企業における、サーバからのストレージに対するデータアクセス頻度の分布を分析した結果、20%の領域に約80%のI/O負荷が集中していることが報告されている[1]。このように、ストレージ内のアクセス頻度には偏りがあるため、企業では、アクセス頻度に応じて適切なストレージ階層にデータを配置する「階層管理」を行っている。この階層管理は従来、ストレージ装置ではボリューム単位で行っていたが、サーバ仮想化の普及などによりボリューム内でアクセス頻度の異なるデータが混在するようになった。

本要求に対し、ストレージ装置が動的にボリュームより小さな単位で階層管理を行う機能として、ストレージ階層仮想化機能を開発した。本機能は、容量の利用効率を向上させることを目的とした容量仮想化機能[2]の拡張機能として提供する。容量仮想化機能とは、サーバからの書き込みに応じて、ボリュームより小さい実領域（ページ）をストレージプール（以下、プール）から仮想ボリュームに動的に割当てる機能である。ストレージ階層仮想化機能は、プール内で複数の階層を管理し、ページを管理単位として、アクセス頻度の採取と、アクセス頻度に応じた適切な階層への再配置を実行する機能を持つ。

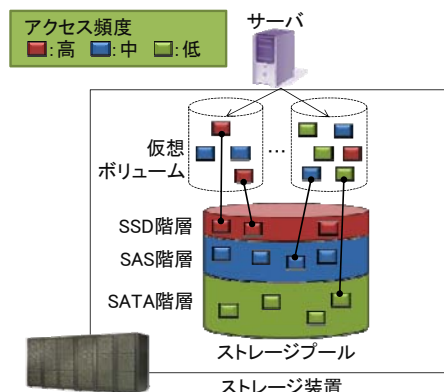


図1 ストレージ階層仮想化機能概要

† (株)日立製作所 横浜研究所
Hitachi, LTD., Yokohama Research Laboratory

この複数階層で構成したプールの最大スループット性能は、各階層のスループット性能限界値の単純和とはならず、各階層 i のスループット性能限界値 (P_i) と、各階層 i に対する I/O 比率 (R_i) とによって以下の値となる性能特性を持っている。[3]

$$\min_{i \in T} \frac{P_i}{R_i} \dots (a)$$

しかし、本性能特性によって求められる最大スループット性能は、実際にページをどのように配置するかによって変化する。本研究では、複数階層で構成したプールの最大スループット性能を向上させるページ配置先決定法を提案すること目的とする。

3. ページ配置先決定法

本節では、本研究にて考案したページ配置先決定法である、性能制限付き上詰め配置法について述べる。その前にまず、比較対象とする、単純上詰め配置法について述べる。

3.1 単純上詰め配置法

本方式は、広く知られた貪欲法を使った方式であり、各ページを I/O 負荷で降順ソートし、I/O 負荷が高いページから順に高位の階層に配置する。

以下、本方式の手順を示す。I/O 負荷の降順でソートした N 個のページの集合を $\{p_1, p_2, \dots, p_N\}$ 、1 ページ当たりの容量を c 、階層 i に配置したページの集合を T_i 、階層 i の容量限界値を C_i とした時、単純上詰め配置法は、 p_1 より、各ページを配置する階層を以下の手順で決定する。

1. 未配置のページ p_k を階層 i に配置する ($p_k \in T_i$)
2. 階層 i が容量限界に達したかどうか判定する

$$\sum_{p_k \in T_i} c > C_i$$

3. 階層 i が容量限界に達していた場合、下位の階層 ($i+1$) に配置先を変更する。

3.2 性能制限付き上詰め配置法

本方式は、単純上詰め配置法を拡張した方式であり、各階層の容量限界だけでなく、性能限界も考慮する。

以下、本方式の手順を示す。前節で定義に加え、ページ p_k の性能を $P(p_k)$ 、階層 i のスループット性能限界値を P_i とした時、性能制限付き上詰め配置法では、 p_1 より、各ページを配置する階層を以下の手順で決定する。

1. 未配置のページ p_k を階層 i に配置する。 ($p_k \in T_i$)
2. 階層 i が容量限界もしくは性能限界に達したかどうか判定する。

$$\sum_{p_k \in T_i} c > C_i, \text{ または, } \sum_{p_k \in T_i} P(p_k) > P_i$$

3. 階層 i が容量限界、もしくは性能限界に達していた場合、下位の階層 ($i+1$) に配置先を変更する。

4. 評価

4.1. 方式比較

本節では、どの階層が先に性能限界となるかによって場合分けを行い、各方式の最大スループット性能を算出することにより、両方式の比較を行う。尚、説明を簡単化するため、階層数は2階層とする。

(1) 階層2が性能限界となる場合 ($P_1/R_1 > P_2/R_2$)

階層2が性能限界となる場合の、単純上詰め配置法の最大スループット性能は、以下の値となる。これは、式(a)と等しい。

$$P_2 \left(\frac{\sum_{k=1}^N P(p_k)}{\sum_{k=m+1}^N P(p_k)} \right) = \frac{P_2}{R_2} \quad (\text{ただし, } m = \frac{C_1}{c})$$

一方、性能制限付き上詰め配置法の最大スループット性能も、同じ値となる。すなわち、この場合は、単純上詰め配置法と性能制限付き上詰め配置法のスループット性能は同等である。

(2) 階層1が性能限界となる場合 ($P_1/R_1 < P_2/R_2$)

階層1が性能限界となる場合の、単純上詰め配置法の最大スループット性能は、以下の値となる。これも、式(a)と等しい。

$$P_1 \left(\frac{\sum_{k=1}^N P(p_k)}{\sum_{k=1}^m P(p_k)} \right) = \frac{P_1}{R_1} \quad (\text{ただし, } m = \frac{C_1}{c}) \cdots (b)$$

一方で、性能制限付き上詰め配置法の最大スループット性能は、以下の値となる。

$$P_1 + P_2 \cdots (c)$$

これは、性能限界に達した階層1に配置できなかったページを、階層2に配置することによって、階層2の性能限界に到達することができるためである。この式(c)は、条件式($P_1/R_1 < P_2/R_2$)を使って変形すると、以下のようになる。すなわち、この場合は、性能制限付き上詰め配置法の方が、単純上詰め配置法よりも、最大スループット性能が高くなる。

$$P_1 + P_2 > P_1 + \frac{R_2}{R_1} \cdot P_1 = \frac{P_1}{R_1}$$

(3) 比較結果

(1)(2)の結果より、性能制限付き上詰め配置法の方が、単純上詰め配置法よりも、最大スループット性能が高いことが示された。

4.2 スループット性能見積もり

本節では、実際に階層1が性能限界となる階層構成にて、各方式の最大スループット性能を見積もり、性能制限付き上詰め配置法の効果を示す。

まず、前提とする階層構成を示す。階層はSAS, SATAの2階層とし、各階層はそれぞれSAS15K 450GB×8台, SATA7.2K 1TB×16台で構成する。見積もり簡略化のため、RAID構成は組まないこととする。

本構成における、各階層の容量比、性能比は表1の通りとなる。特に、スループット性能の限界値は、 $P_1 + P_2 = P_1 + 53/42P_1 \approx 2.3P_1$ となる。尚、各メディアの性能値は、カタログ仕様[4]を基にした。

表1 階層構成

階層#	メディア	台数(台)	性能比(%)	容量比(%)
1	SAS	8	54	18
2	SATA	16	46	82

図2は、ストレージの標準ベンチマークであるSPC-1[5]の仕様を基に作成したアクセス頻度の偏りを示すグラフである。SPC-1を本見積もりのワークロードと仮定すると、各階層の容量比がSAS:SATA=18:82であることから、図2より $R_1:R_2 = 86:14$ であることが分かる。すなわち、 $P_1/R_1 < P_2/R_2$ であり、本構成は階層1が先に性能限界に達する。

この時、単純上詰め配置法のスループット性能は、式(b)より $1.16P_1$ となり、スループット性能の限界値の54%に制限される。一方、性能制限付き上詰め配置法は、スループット性能の限界値である $2.3P_1$ となる。以上より、性能制限付き上詰め配置法が、単純上詰め配置法と比較した最大スループット性能を1.9倍向上することができるという効果を確認した。

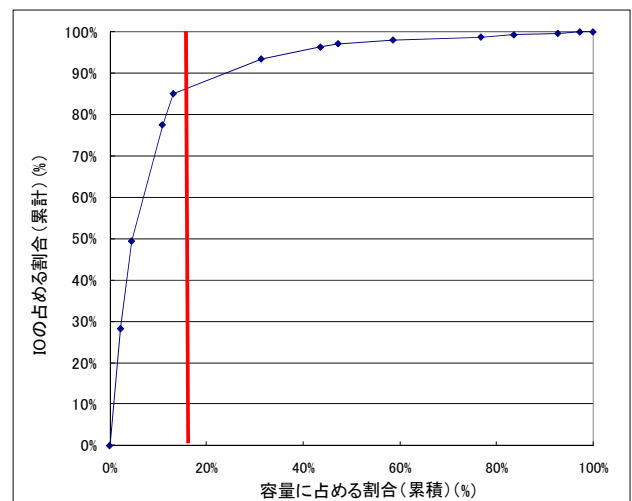


図2 SPC-1の負荷の偏り

5 まとめ

複数階層を組み合わせた構成でのページ配置先決定法である、単純上詰め配置法と性能制限付き上詰め配置法とで最大スループット性能の比較を行った。結果、性能が低い階層が性能限界になる場合は、単純上詰め配置法と性能制限付き上詰め配置法とは同等であり、性能が高い階層が性能限界となる場合は、性能制限付き上詰め配置法の方が、単純上詰め配置法に比べて、最大スループット性能が1.9倍向上することを示した。

参考文献

- [1] L. Cherkasova, M. Gupta, "Characterizing locality, evolution, and life span of accesses in enterprise media server workloads", NOSSDAV'02, May 2002
- [2] 江口賢哲, "大規模ストレージシステムにおける動的容量割り当て(Dynamic Provisioning)機能の研究", FIT2008, C-019, September 2008
- [3] 須藤 梓 他, ストレージ階層仮想化機能の実現方式検討, 情報処理学会 2011
- [4] Seagate: <http://www.seagate.com/www/en-us/products/enterprise-hard-drives/>
- [5] SPC-1: <http://www.storageperformance.org/specs>