

問4 データ分析処理の並列化に関する次の記述を読んで、設問1~3に答えよ。

K社は、様々な商品の売行きを調査し分析した結果を、顧客企業に提供している。

K社で使用しているアプリケーションプログラムの全体の流れを図1に示す。変換処理には変換サーバを、分析処理には分析サーバを用いている。変換処理の結果は、ファイル転送処理によって、変換サーバから分析サーバにファイル転送される。

最近では、分析対象となるデータの種類や量が増加するだけでなく、分析結果をより短時間で提供することが求められている。そこでK社では、処理時間を大幅に短縮する方策を検討することにした。

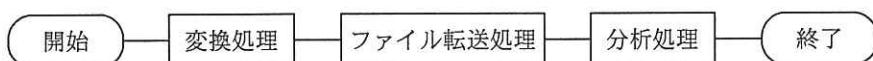


图1 アプリケーションプログラムの全体の流れ

[変換処理プログラムの改良]

変換処理は、分析処理がしやすいように入力ファイルを加工する処理であり、入力ファイルごとに異なるレコードフォーマットやコードを統一する。変換処理のステップを図2に示す。

変換処理の時間を短縮するために、プログラムの改良を実施した。改良前の各ステップの処理時間の比率は、フォーマット変換処理が20%，コード変換処理が30%であった。プログラムの改良によって、フォーマット変換処理の時間は $1/4$ に、コード変換処理の時間は $1/3$ に短縮された。その他の処理のプログラムは変更せず、処理時間の短縮もない。これによって変換処理は a 倍の性能向上が達成された。

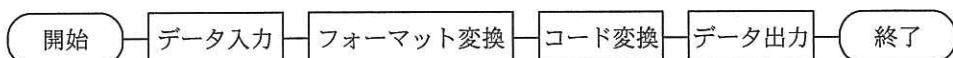


图2 変換処理のステップ

[分析処理の並列化]

分析処理は、変換処理が output した結果ファイルを基に、顧客に合わせた分析レポートを出力する。従来は、1台のサーバで全ての処理を逐次実行してきた。処理1件当

たりの処理時間は、顧客からの処理要件や、データの種類や量によってその都度異なるが、最短で1時間、最長で3時間である。

分析処理の時間を短縮するために、分析サーバを3台用意し、処理を並行して実行することにした。ただし、各分析サーバに割り当てられた処理は逐次実行し、1件の処理を複数のサーバに分散して実行することはない。複数の処理を1台のサーバで並行実行することもない。

3台の分析サーバに処理を割り当てる方法として、方法Aと方法Bを考えた。

(1) 方法A

各分析サーバで実行する件数が均等になるように処理を割り振る（図3）。



図3 分析処理の割当方法A

(2) 方法B

分析サーバ間の処理時間のばらつきが小さくなるように、bした分析サーバに、順次、次の処理を割り当てる（図4）。



図4 分析処理の割当方法B

方法Aの場合、①分析サーバ間の処理時間のばらつきが大きくなり過ぎて、予想していた時間内に分析処理が完了しないことがある。一方、方法Bの場合、割り当てられた全部の処理を完了する時間が最も長く掛かる分析サーバでも、他の分析サーバよりも最大でc時間長く掛かるだけで済む。

分析処理全体の処理時間の短縮を考え、方法 B を採用した。

[システム構成の検討]

システム構成案として、分析処理のデータの入出力方法が異なる構成案 1 と構成案 2 を考え、ファイル転送処理と分析処理における特徴について考察した。

(1) 構成案 1

個々の分析サーバがもつ内蔵ディスクでデータの入出力を行う（図 5）。

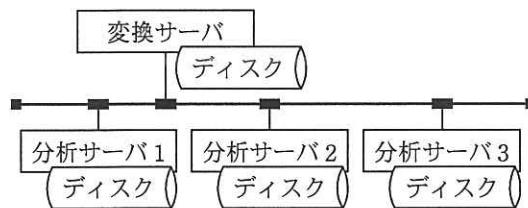


図 5 システム構成案 1

(2) 構成案 2

どの分析サーバからもアクセスできるファイルサーバでデータの入出力を行う（図 6）。

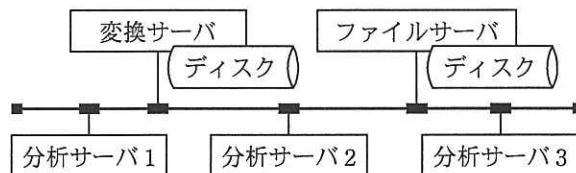


図 6 システム構成案 2

ファイル転送処理において、構成案 1 では、変換処理結果のファイルを、全分析サーバに対し、それぞれ別々にファイル転送する。変換サーバは、ファイル転送時間の短縮のために、複数のファイル転送処理を同時に実行する。しかし、分析サーバが増えた場合は、変換サーバの内蔵ディスクからの読み込みや d がボトルネックにならないか考慮する必要がある。一方、構成案 2 の場合、変換処理の結果は、ファイルサーバにだけファイル転送すればよいので、分析サーバが増えた場合でもこれらがボトルネックになることはない。

分析処理において、構成案 1 の場合、分析に必要なデータは個々の分析サーバの内蔵ディスクにあるので、分析サーバが増えても はボトルネックにならない。一方、構成案 2 の場合、データはファイルサーバにあるので、分析サーバが増えた場合は、 だけでなく についても、ボトルネックにならないか考慮する必要がある。

ある分析サーバに障害が起きたときは、別の分析サーバが処理を引き継ぐ。その際の処理時間の遅延は無視できるとした場合、構成案 2 におけるシステム全体の稼働率は、変換サーバ及び各分析サーバの稼働率を x 、ファイルサーバの稼働率を y としたとき、 と表すことができる。ファイルサーバの稼働率がシステムの稼働率に大きく影響するので、②可用性向上のために対策を施す必要がある。

設問 1 本文中の に入れる適切な数値を答えよ。答えは、小数第 2 位を切り捨てて、小数第 1 位まで求めよ。

設問 2 [分析処理の並列化] について、(1)~(3)に答えよ。

- (1) 本文中の下線①について、15 件の分析処理を行う場合、分析サーバごとの処理時間の差は最大で何時間になるか。数値を答えよ。
- (2) 本文中の に入れる適切な字句を 10 字以内で答えよ。
- (3) 本文中の に入れる適切な数値を答えよ。

設問 3 [システム構成の検討] について、(1)~(3)に答えよ。

- (1) 本文中の , に入れる適切な字句を解答群の中から選び、記号で答えよ。

解答群

- ア ネットワークの帯域
- イ ファイルサーバのディスクの入出力
- ウ 分析サーバの CPU 処理能力
- エ 分析サーバの内蔵ディスクからの読み込み
- オ 変換サーバの内蔵ディスクへの書き込み

- (2) 本文中の に入る適切な式を答えよ。ここで、ネットワークの稼働率は 1 とする。

(3) 本文中の下線②について、ファイルサーバの稼働率を高めるための対策として適切なものを、解答群の中から全て選び、記号で答えよ。

解答群

- ア ファイルサーバのディスクコントローラ装置を二重化する。
- イ ファイルサーバのディスクを、よりシークタイムが短いものに交換する。
- ウ ファイルサーバのディスクを RAID0 で構成する。
- エ ファイルサーバのディスクを RAID1 で構成する。
- オ ファイルサーバのディスクを RAID6 で構成する。