

スプールファイルを作成しない データ処理結果のアーカイブファイルの提供

根本利弘^{†1} 喜連川優^{†2}
 東京大学^{†1†2} 国立情報学研究所^{†2}

1. はじめに

近年、WWW を通じて地球環境データなどの大規模なデータを加工し、結果を提供するサイトが構築・運用されるようになってきた。データ統合解析システム (Data Integration and Analysis System : DIAS) [1]においても、「地球温暖化対策に資するアンサンブル気候予測データベース (database for Policy Decision making for Future climate change : d4PDF)」[2]のオリジナルファイルから利用者に指定された一部分を抽出して提供する機能を公開している。このようなシステムにおいては、利用者のリクエストを受けると httpd は全体を統括するスクリプトをまず起動し、そのスクリプト内で複数のサブ処理プログラムが実行され、各サブ処理は各々の結果ファイルをスプール領域に出力し、統括スクリプトがそれらの結果ファイルを tar や zip などの汎用ファイルアーカイブによって 1 つにまとめて利用者に提供するというような構成をとるものも多い。オリジナルデータのデータ量が大きいとその処理結果も大きくなる場合が多く、必然的に大きなスプール領域が必要とされる。また、tar コマンドや zip コマンドはディスク上に存在するファイルをアーカイブ対象とするため、すべての結果ファイルの出力が完了した後でないとアーカイブファイルを作成できず、利用者へのデータ送込に遅延が生じる。全てのサブ処理の完了までに時間がかかる場合には、一度セッションを終了し、処理が完了した段階でメール等により利用者に処理の完了を伝え、利用者が処理結果をダウンロードするような非同期的なシステムとなる。利用者は通知を受けてからすぐに結果のダウンロードを行うとは限らず、システム側は一定期間、スプール上の結果ファイルを保持しておく必要が生じ、より一層大きなスプール領域が要求されることとなる。

筆者らは、データの切り出しなどの比較的軽量で結果の出力がすぐに開始されるようなサブ処理を複数回繰り返すようなサービスにおいて、個々のサブ処理コマンドからの出力を統括スクリプトがパイプを通じて読み込み、それらをディスク上に出力せずに tar フォーマット形式に逐次変換して httpd を通じて利用者へ送る手法を開発し、WWW による d4PDF の切り出しインタフェースに適用した。全ての処

理の完了を待たずに結果を利用者に送出することができるため、利用者はリクエストのレスポンスとして遅延なく結果を受け取ることが可能となる。また、サブ処理の結果をファイルとして出力する必要がなくなり、必要とされるスプール領域を大幅に削減することができる。本稿ではこの方式について説明を行うとともに、実際に d4PDF の切り出しインタフェースへのリクエストのログを用いてスプール領域の削減効果を示し、評価を行う。

2. スプールファイルを作成しないアーカイブファイルの送込

2.1 tar ファイルの構造

表 1 tar ファイルの構造

	長さ (byte)	内容
Header 1	512	ファイル名, 作成日時, ファイルサイズ等
Content 1	512×n ₁	データ本体 (512 の倍数に 0 パディング)
Header 2	512	ファイル名, 作成日時, ファイルサイズ等
Content 2	512×n ₂	データ本体 (512 の倍数に 0 パディング)
Header 3	512	ファイル名, 作成日時, ファイルサイズ等
Content 3	512×n ₃	データ本体 (512 の倍数に 0 パディング)
		...
Tailer	1024	0

表 1 はデフォルト値である 512 バイトをブロックサイズとした場合の tar アーカイブファイルの構造である[3]。tar ファイルは、アーカイブされる個々のファイルに対する 512 バイトのヘッダ後にその内容が続き、512 バイトの倍数となるよう 0 で埋められる。これがアーカイブされるファイルの数だけ繰り返され、最後に 1024 バイトの 0 で終端される。シンボリックリンクやデバイスファイルもアーカイブ可能であるが、アーカイブされるファイルが通常のファイルである場合、ヘッダには、ファイル名、パーミッション、ユーザ ID、グループ ID、ファイルサイズ、更新時刻、ファイル種別、ユーザ名、グループ名が含まれる。

2.2 スプールファイルを作成しないアーカイブファイルの送込法

利用者からのリクエストを受けると、httpd は統括スクリプトを実行する。統括スクリプトは、実際にサブ処理コマンドを実行する前に、そのサブ処理の結果ファイルが持つべきファイル名、パーミッション等の必要な情報を決定して tar ファイルのヘッダを構成し、httpd を通じて利用者へ送出する。その後サブ処理コマンドを実行し、ファイルを作成することなくパイプを通じてその結果を受け取り、tar ファイルのデータ本体として利用者へ送出する。これを繰

Provision of archive file of processing results without creating spool files

†1 TOSHIHIRO NEMOTO, The University of Tokyo.

†2 MASARU KITSUREGAWA, The University of Tokyo / National Institute of Informatics.

り返し行い、最後に 1024 の 0 を送出する。このようにして、スプールファイルを作成することなく利用者へアーカイブファイルを送る。データ切り出しインタフェースの場合、tar ファイルに含まれるヘッダ情報のうち、パーミッション、ユーザ ID、グループ ID、ファイル種別、ユーザ名、グループ名に関しては、システム管理者により事前に設定された値を用いる。ファイル名は利用者のリクエスト内容により決定し、また、更新時刻は処理開始時刻とする。ファイルサイズは、切り出し結果をプレーンバイナリ形式とすることにより、切り出す期間、空間領域より格子数が決まり、対象とする実験名、変数名から 1 格子あたりのバイト長が決まるため、実際に切り出し処理を実行する前にデータサイズを求めることができる。本方式ではサブ処理の実行と利用者へのアーカイブファイルの送出を同時に実行でき、利用者が結果を受け取るまでの時間を短縮できる。

3. スプール領域の削減効果の評価

3.1 d4PDF 切り出しインタフェース

スプールファイルを作成しないアーカイブファイルの送出は既に d4PDF 切り出しインタフェースにおいて実装され、運用が行われている。ここでは、この d4PDF 切り出しインタフェースのアクセスログを用いて、本方式によりスプール領域の削減効果の評価する。

d4PDF は多数のアンサンブル数を特徴とする高解像度大樹モデルを使用した、総容量約 2PB の高精度モデル出力実験出力データセットである。d4PDF 切り出しインタフェースは、利用者により指定された実験名、期間、変数名、アンサンブルおよび空間領域をオリジナルデータから切り出し、利用者へ送出する。アンサンブルおよび同じカテゴリに属する変数は複数指定可能である。結果は、アンサンブル、変数、および 1 ヶ月または 1 年単位に分割されたファイル群により構成され、これらが tar ファイルで送られる。

3.2 アクセスログによるスプール領域の削減効果の評価

d4PDF 切り出しインタフェースの 2019 年 4 月 1 日から 9 月 30 日までの約 31000 件のアクセスログを用い、d4PDF 切り出しインタフェースが通常の tar コマンドを使用し、スプールファイルを作成するとした場合のスプール領域の使用量を求めた。結果は利用者へ非同期に送信されることとなるため、スプールファイルは一定の期間、スプール領域に保持されるものとしている。図 1 はスプールファイルを 1 日間保持した場合、図 2 はスプールファイルを 1 週間保持した場合のスプール領域の使用量の推移を示したグラフである。保持期間を 1 日とした場合には、最大約 18TB、保持期間を 1 週間とした場合には最大約 43TB のスプール領域を使用する。すなわち、サービスを継続するためには最低でもこれらの容量のスプール領域の確保が要求される。実運用においては事前に必要とされる最低限の容量を決定することは難しいため、余裕をもって領域を確保しなければな

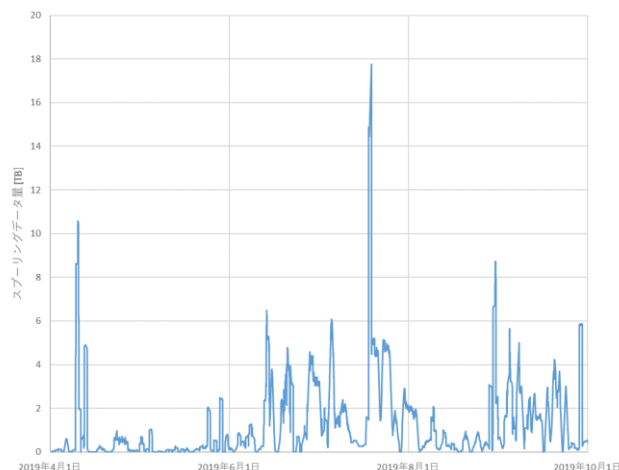


図 2 スプール領域上のデータ量（保持期間 1 日）



図 1 スプール領域上のデータ量（保持期間 1 週間）

らず、さらに大きな容量が必要とされる。このように、スプールファイルを作成せずにアーカイブファイルを構成して利用者へ送出することにより、必要とされるスプール領域を大幅に削減できたことが示された。

4. おわりに

本稿では、WWW を通じて大規模データの処理を行うシステムにおいて、スプールファイルを作成せずに結果をアーカイブファイル形式で送出する方式について述べた。また、スプール領域の削減効果を実際のアクセスログを用いて評価し、スプール領域の大幅な削減が可能であることを示した。

参考文献

- [1] “DIAS: データ統合・解析システム”. <https://diasjp.net/>, (参照 2019-10-31).
- [2] “地球温暖化対策に資するアンサンブル気候予測データベース”. <http://www.miroc-gcm.jp/~pub/d4PDF/>, (参照 2019-10-31).
- [3] “GNU tar 1.32: Basic Tar Format”. https://www.gnu.org/software/tar/manual/html_node/Standard.html, (参照 2019-10-31)