

リモートサーバ上でのデータ変換・転送業務

○瀬川朋紀^{A)}、ERG サイエンスセンター

A) 教育・研究技術支援室 計測・制御技術系

概要

現在、科学観測データを公開する際、利便性向上のため、関連する研究分野における共通データフォーマットへ公開したいデータを変換する必要がある。この度、学外の計算機環境において、観測データフォーマットを共通データフォーマットに変換し、それを学内の web site で公開する業務を担うことになった。ここでは、枯れたツール(Python 2.7 系, SQLite3)を組み合わせて構築したプログラムを紹介する。

1 はじめに

宇宙航空研究開発機構宇宙科学研究所と名古屋大学宇宙地球環境研究所が共同運用する宇宙科学連携拠点として設置された ERG サイエンスセンター(ERG-SC)では、科学観測データの web 公開にあたり共通的なデータフォーマットとして NASA Common Data Format (CDF) [1]を採用している。その活動の一環として共同研究機関のひとつである A 大学の科学観測データ[2]を A 大学の計算機環境上で CDF に変換し、A 大学から CDF ファイルを ERG-SC に転送、データポリシーなど必要なメタ情報を付与した上で ERG-SC の web site から CDF ファイルを公開するという必要が生じ、それに伴う定常業務の確立と自動化を行うシステムを構築した。

2 利用ソフトウェア

本業務で利用した主要ソフトウェアを表 1 に示す。

A 大学側で Linux 系 OS 環境とアプリケーションとして Harris 社の Interactive Data Language (IDL)が用意された。A 大学の科学観測データを CDF に変換する IDL プログラムや、CDF の作成に必要なメタ情報などは研究者側から提供された。

また、多くの Linux ディストリビューションに付属している python や SQLite3 を利用し、業務に必要とされる処理の自動化をめざすこととした。多くの利用者がある python や SQLite3 ならばソフトウェアとしての寿命が長いこと、UNIX/Linux 環境ならば移植が簡単になることも期待して選定した。

SQLite3 で構築したデータベ

表 1. 主要ソフトウェア

名称	バージョン
Python	2.7 系
SQLite	3.7 系
OpenSSH	
sh	
運用環境 OS	Scientific Linux 7 系
Harris 社 IDL	8.4
CDF	V3.6.3.1

表 2. フィールド名とデータ型

Field Name	Data Type	Comment
id	INTEGER	
filename	TEXT	科学観測データファイル名
mtime	REAL	科学観測データファイルの更新時間 (UNIX TIME STAMP)[sec]
CDFtime	REAL	CDF ファイル作成時間(UNIX TIME STAMP)[sec]

ースの TABLE を表 2 に示す。表 2 の「filename」には科学観測データファイル名を記録し、「mtime」にはそのファイルの最終変更時刻(last modification time)を UNIX 時間(秒)で記録する。

3 運用

定常業務の自動化を実現するために、ERG-SC の計算機においては、**crontab** を用いて定期的にプログラムを実行させている。プログラムの流れを図 1 に示す。図 1 の背景が白色の処理は名古屋大学にある ERG-SC

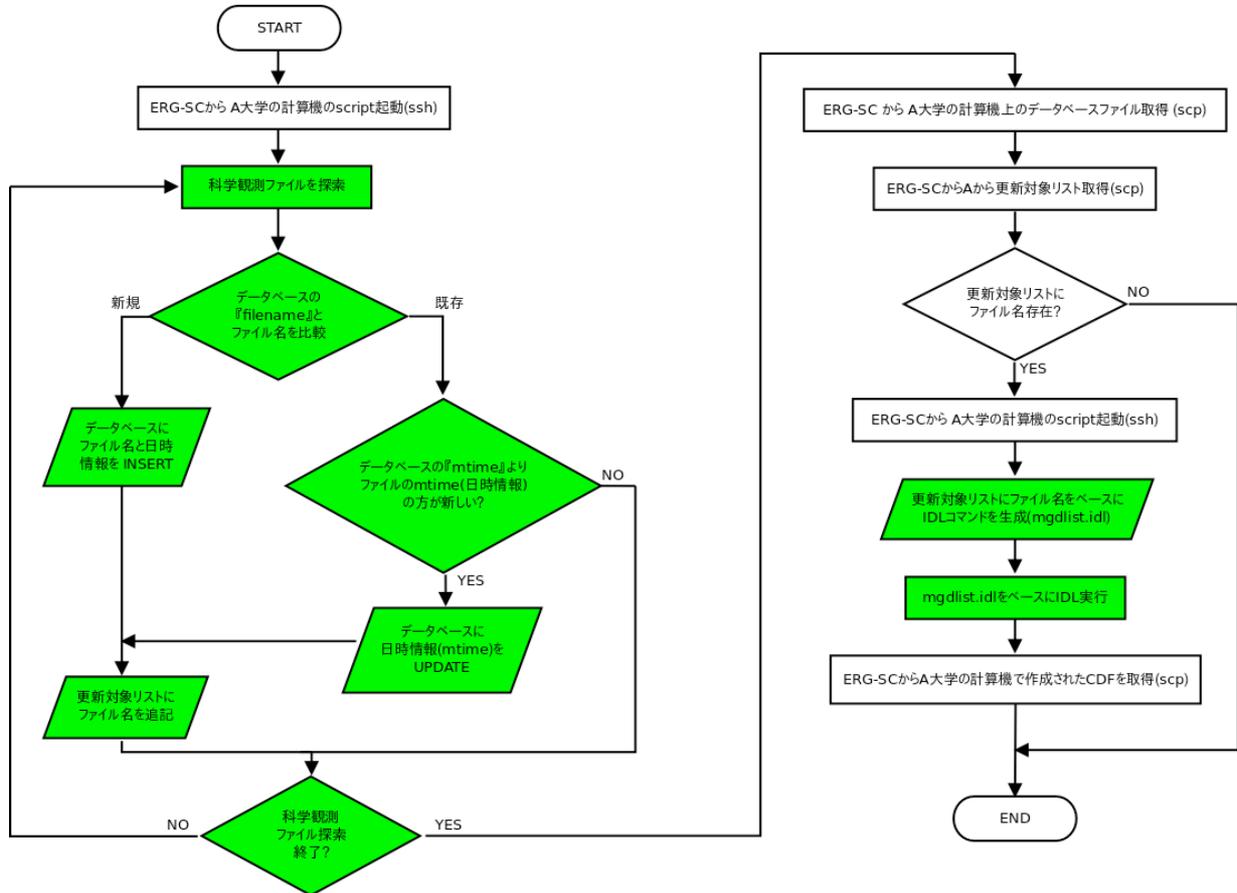


図 1 フローチャート

の計算機が主に実行しているところを示す。背景が緑色の処理は A 大学の計算機が実行しているところを示す。名古屋大学と A 大学間の通信内容は公開鍵認証による SSH によって保護されており、ERG-SC の計算機から A 大学の計算機へ SSH 経由で命令を発行している。一方、A 大学の計算機から ERG-SC の計算機への接続は許可されていない。

プログラムの基本的な流れは、科学観測データファイル名やファイルの更新日時情報をデータベースと比較し、レコードと一致しない場合、IDL プログラムを実行し、科学観測データファイルを CDF 形式へと変換する。変換された CDF ファイルのみを取得し、ERG-SC の web site に配置する。源泉の科学観測データファイルは取得しない。

4 今後の課題

今後、長期的にプログラムを保守、発展させること、さらに移植性を高めるために、プログラム中のコメントの充実が重要であると考えている。また、事前検証に必ずしも時間をあてられなかったため、運用の中

で見いだされる問題も多く、その都度改修を行っている。このような問題の一例として、A 大学側で源泉となる科学観測データファイルの削除があった際の対応方法の検討があげられ、今後、問題が深刻化する前に対応していきたいと考えている。

参考文献

- [1] NASA/Goddard Space Flight Center, Space Physics Data Facility, <http://cdf.gsfc.nasa.gov/>
- [2] Hori, T. et al., “CDF data archive and integrated data analysis platform for ERG-related ground data developed by ERG Science Center (ERG-SC)”, J. Sp. Sci. Info. Jpn., 4(JAXA-RR-14-009 (ISSN 1349-1113)), 75-89, 2015