

Key-Valueストアを用いた GPVデータ閲覧サービスの構築

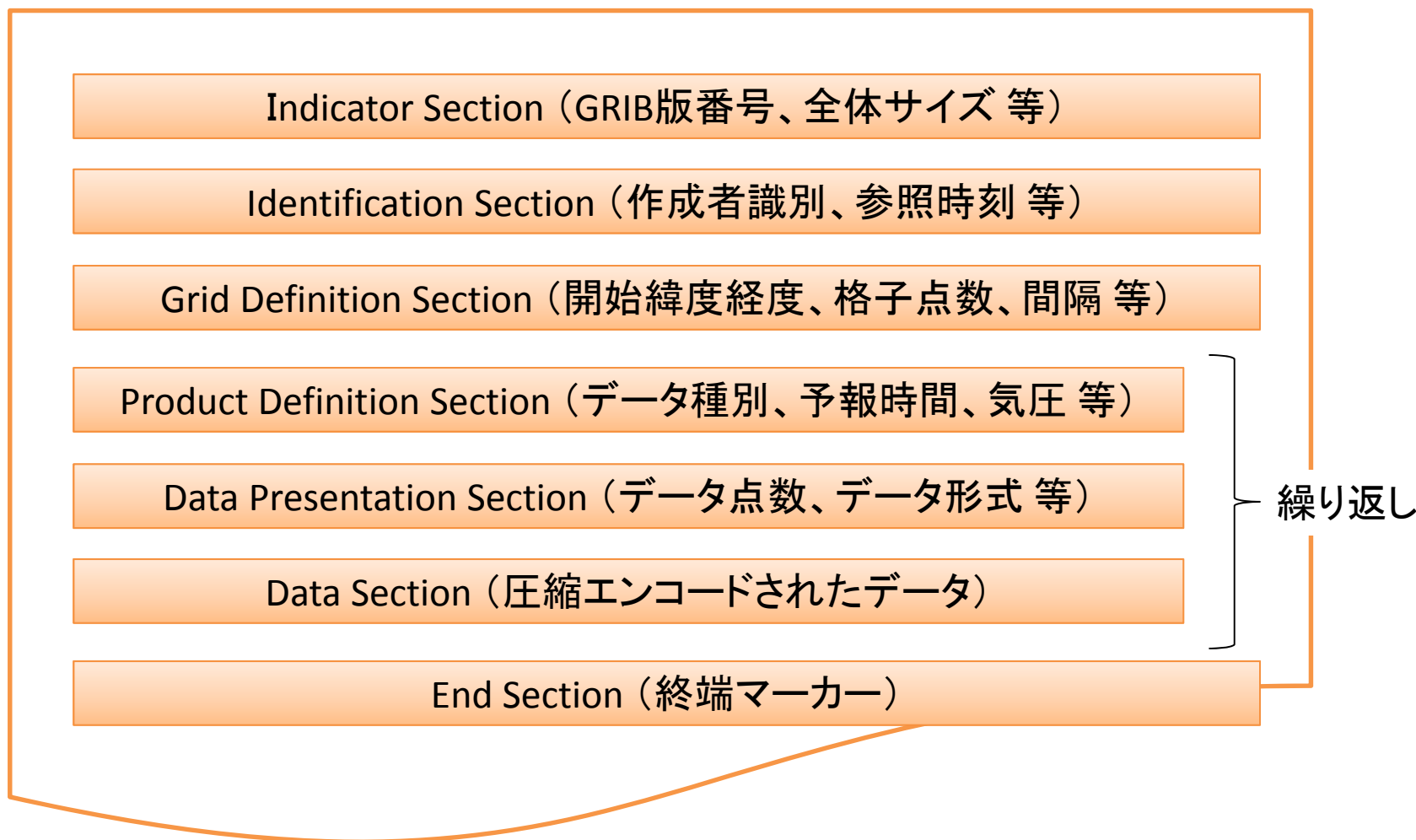
2011年10月21日

株式会社テクリエ
岡村 和英

- 気象情報をクラウド上で容易に利用できるようにしたい。
 - 特定のパラメータ値に基づいて事前に作成された画像を取得するのではなく、要求に応じて任意の画像を生成・表示するようにできないか？
 - パブリッククラウド等を用いることで急激なアクセス増加に応じて可変的にスケールの変更可能なシステムに出来るのでは？
 - 過去のデータへも容易かつ迅速にアクセスしたい。

GRIBファイルの構造

- MSM気圧面予報データの場合



- MSM気圧面予報データの場合

参照時刻における

1000hPa等圧面の高度、風速、気温、上昇流速度、相対湿度
(緯度方向0.1度、経度方向0.125度間隔)

950hPa等圧面の高度、.....

900hPa等圧面の.....

∩

100hPa等圧面の.....

3時間後の...

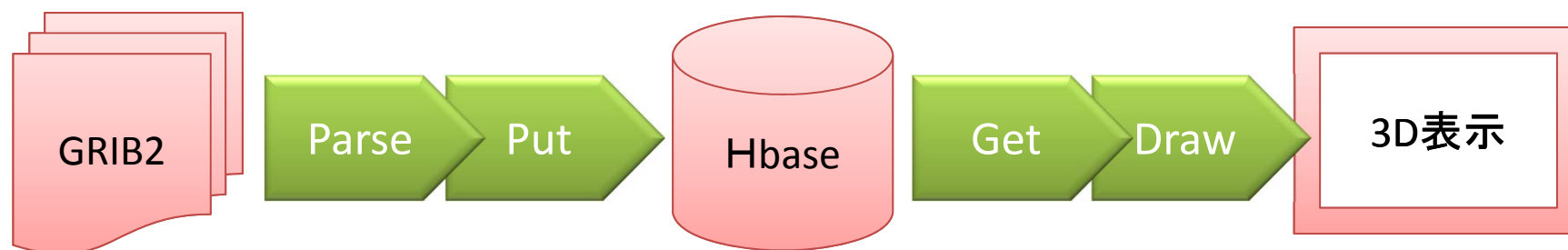
6時間後の...

∩

15時間後の...

今回作成したもの

- GRIB2データパーサー
 - GRIB2形式のバイナリファイルからGPVデータを読み込む。
- Key-ValueストアAPI
 - GRIB2ファイルから読み込んだGPVデータをエリア分割し、Hbaseに書き出す。
 - 指定されたエリアのGPVデータをHbaseから読み込む。
 - 今回はMSM(日本周辺)のデータのみに対応。
- GPVビューアー
 - GPVデータを3D表示する。



- Windows7
- Java SE 7
- Hbase 0.90.4
 - 今回はHDFSではなくローカルディスクを利用
- Tomcat 6.0.29
- JOGL (Java OpenGL)

HBaseテーブルレイアウト

- 単一テーブル”gpv”に全データを格納
- キー

対象日時	データ種別	緯度(5度単位)	経度(5度単位)
------	-------	----------	----------

- 指定された緯度経度を中心とした±5度(10度)分を1エリアとして取り扱う(隣のエリアと半分重なっている)。
- 古い予報値を新しい予報値で順次上書きしていき、最終的には解析値(観測値から導出された値)に置き換わる。
- データ
 - ”info”ファミリー: GRIBファイルの参照時刻、ファイル名。
 - ”griddef”ファミリー: データの開始緯度・経度、間隔、データ点数など。
 - ”data”ファミリー: `気圧値`をカラム名としてデコード済みのデータを格納する。

<< Demonstration >>

- Key-Valueストアにデータを格納することで、GRIB2から都度読みだすより迅速に目的のデータを取得することが出来る様になった。
- 複数のGRIB2ファイルに跨っていた過去データも容易に取得できるようになった。

しかし・・・

- GRIB2データに比べてかなり大きなリポジトリが必要になってしまうのでは？
 - エンコードされた予報値(12bit)を実数値(64bit)に展開・格納したことによる増加(約5倍)。
 - エリアの重ね合わせによる増加(約4倍)。
 - Hbaseの仕組みによるオーバーヘッド。

リポジットリサイズの比較

- 気圧面15時間予報データ
 - GRIB 48.1MB → Hbase 3.98GB !



.logs 0bytes
.oldlogs 1.49GB

試験ツール内でPUT完了直後に
Hbaseをshutdownしちゃってました



- 気圧面15時間予報データ

- GRIB 48.1MB → Hbase 1.98GB (COMPACTION完了後)
 - .logs 1.02GB
 - テーブル本体 990MB (COMPRESSION=NONE)

- エンコードされた予報値(12bit)を実数値(64bit)に展開・格納したことによる増加(約5倍)。
- エリアの重ね合わせによる増加(約4倍)。
- Hbaseのデータ保持方法によるオーバーヘッド。



COMPRESSION=GZ に設定したところ「172MB」に縮小。
(GRIBファイルは圧縮しても20～24MB程度)

レポジットリサイズの比較

- 気圧面15時間予報データ
 - GRIB 48.1MB → Hbase 1.98GB (COMPACTION完了後)
 - .logs 1.02GB
 - テーブル本体 990MB (COMPRESSION=NONE)
→ COMPRESSION=GZ すれば 134MB
- 1日分格納後・・・ 気圧面15時間予報 3時間毎 × 8
+ 気圧面33時間予報 6時間毎 × 4
 - GRIB 573MB → Hbase 5.66GB
 - .logs 1.00GB
 - テーブル本体 4.65GB (20倍 → 8倍)

テーブル作成時にmaxVersions=1に設定することにより、古い予報値を新しい予報値で上書きした分だけ容量が削減されている。

やり残したこと

- GPVデータの加工表示
 - 特定高度面における等圧線の表示
 - 特定経路線における高度、等圧面の表示
 - 異なる種類のデータの重ね合わせ表示
 - 地形、気圧、風向き、温度など
- GSM(全球モデル)データへの対応
 - 気圧面によってデータ点数が異なる。
 - テーブルレイアウトの再検討
- モバイルデバイスからの利用



まだまだ、当初の目標にはほど遠い状況

クラウド部会に参加して一緒に作りませんか？