

「巨大時系列データの高速アクセスに関する共同研究」 2年目

2021.02.19

山本幸生（宇宙科学研究所）

古庄晋二、生座本義勝、小林正英、早部秀一（エスペラントシステム）

飯沢篤志、長尾正

アジェンダ

- これまでの活動
- テレメトリデータ活用の課題
- Zap-Overによる解決
- 実現可能になったサービス形態
- 操作デモの説明
- デモムービー
- 100年間利用可能にするために
- まとめ

これまでの活動

- 1998年～
 - 外部のデータ構造を使わず、データに内在する性質を使う**自然数インデックス**の特許出願
 - インメモリデータベースエンジン **Zap-In** を開発
- 2013年～
 - ネットワーク越しにビッグデータファイルを接続して高速に検索・ブラウズする技術
... **Zap-Over** を開発
- 2019年 共同研究1年目（Zap-Over）
 - Apollo 11号～17号が月に設置した地震計データを用いてZap-Overの実用化研究
 - Zap-Overが定めるファイル形式 **DSA** の圧縮技術の開発
 - （多次元）座標の高速検索技術の開発
- 2020年 共同研究2年目（Zap-Over）
 - ビッグデータファイル群の組み合わせを簡便に行う改良
 - UNIONとJOINを一気に行うCOMBINE技術の開発

テーマ『100年後も使える（使われる）ビッグデータのアーカイブ技術』

テレメトリデータ活用の課題

1. 100年使える（使われる）フォーマットがない
2. インターネット経由で簡単にアクセスできない
3. データを組み合わせることが簡単でない
4. レコード数・カラム数が多く扱いにくい
5. 毎日データが発生して作業負担がある

Zap-Overで1～4の解決の目処ができた

Zap-Over (D5A) の利用方法

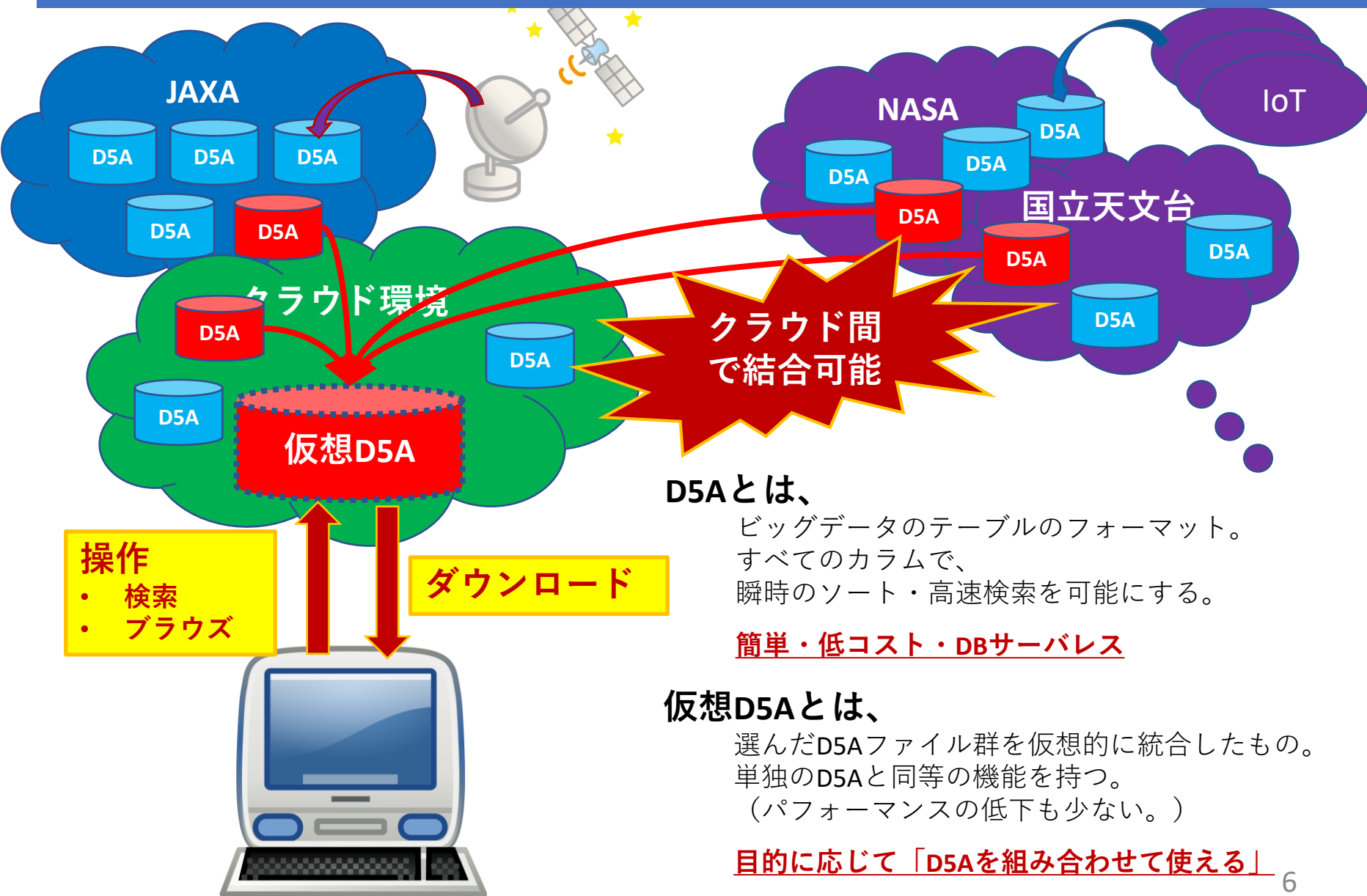
データ提供者： *... テレメトリデータをD5Aファイル化（コンテンツ化）する*
表形式（最大数兆レコード、10万カラム）データをD5Aファイルにして
ファイルサーバに配置する。

データ利用者： *... D5Aファイルを選んで組み合わせる*
ネットワーク上の所望のD5Aファイルを選んで組み合わせ、望むスキーマ
にする。
検索・ソートして必要な部分を取り出す。

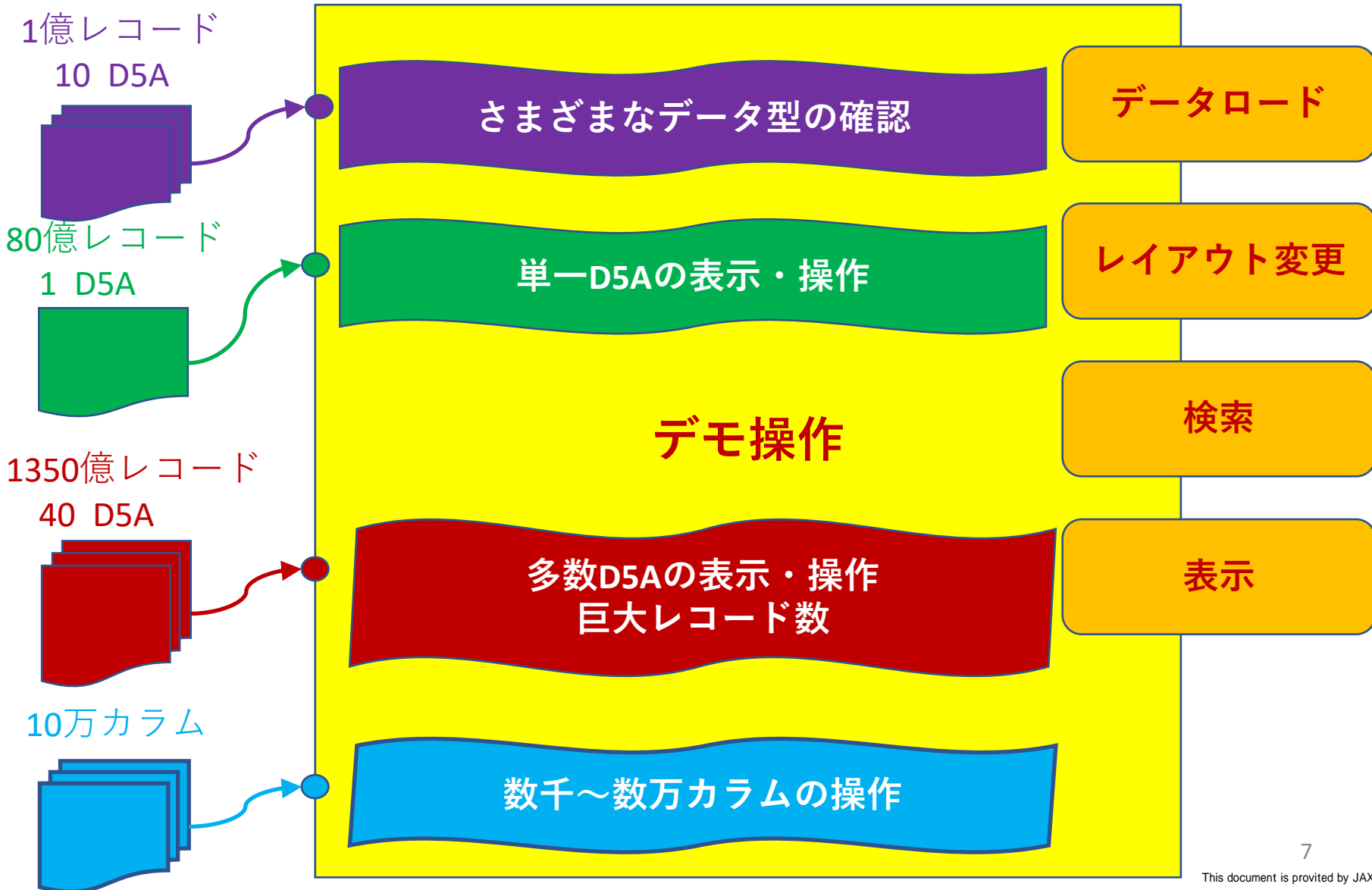
意義：

1. ビッグデータをインターネットのコンテンツとして利用できる。
2. ビッグデータを組み合わせて利用できる。
3. ビッグデータを低コストで利用可能にする。

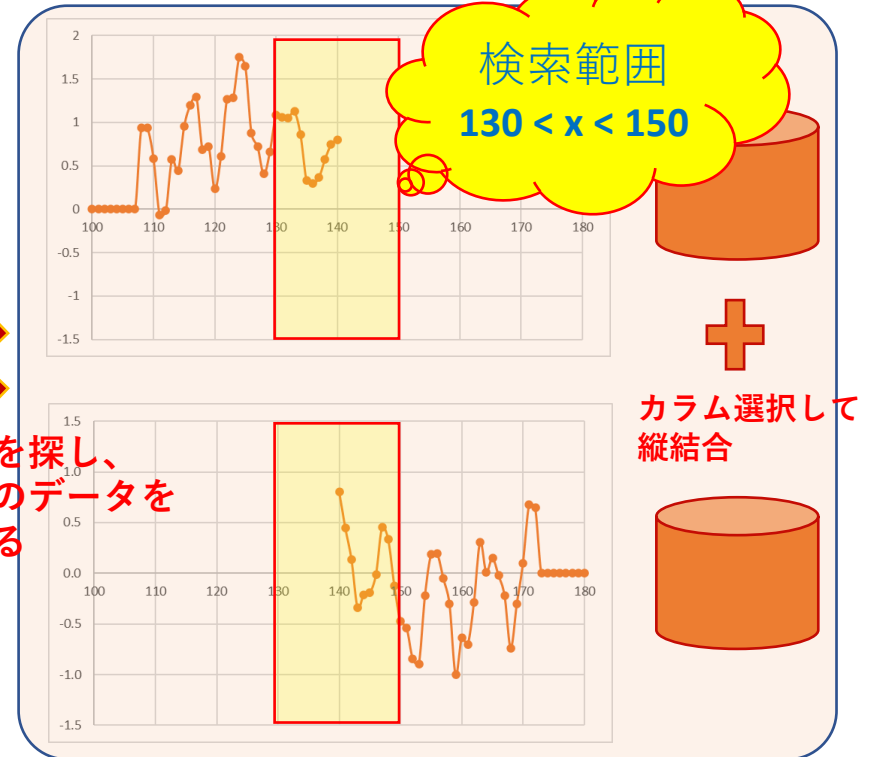
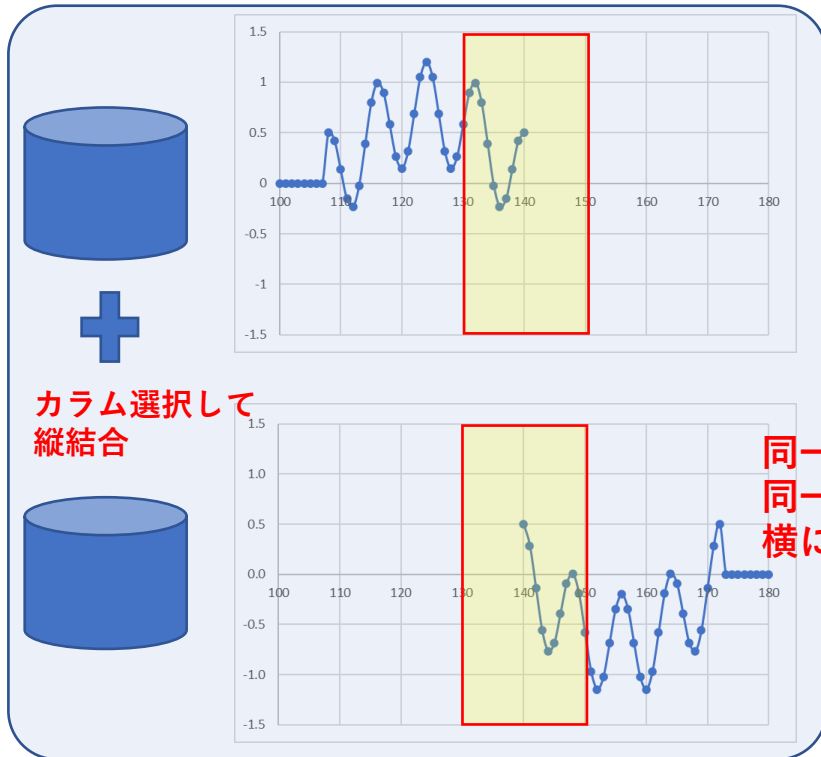
実現可能になったサービス形態



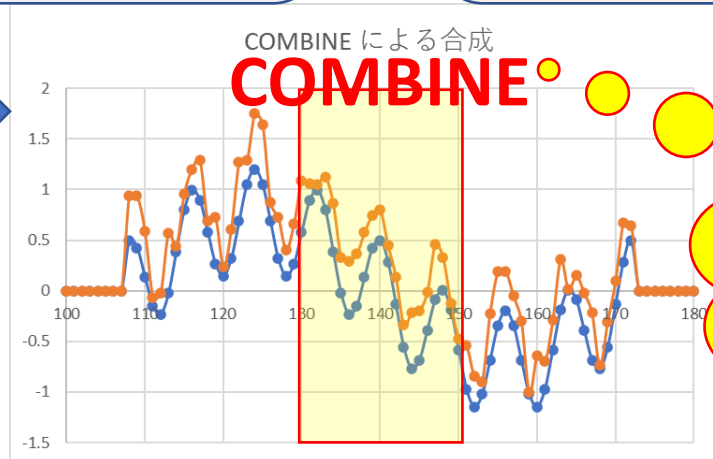
操作デモの説明



時刻をキーに一気に結合：COMBINE



同一時刻を探し、
同一時刻のデータを
横に並べる



COMBINE =
系列毎に
時刻で検索し、
時刻でソート。
そして、時刻をキー
に横に並べる。

デモムービー

AMAZON EC2 においたAPOLLO 1 1号の地震計データと、手元のSSDにおいたAPOLLO 1 2号～1 6号の地震計のデータ、全部で4 0個、リアルタイムに結合しました。

即座に約1 3 5 0億レコードのテーブルを作成。

それを検索・ソート・表示するデモを行いました。
また上記テーブルの所望の一部をCOMBINEで切り出すデモも行いました。

100年間利用可能にするために

- <https://zap100.github.io/>
- 100年間の継続的な技術供給体制の確立
 - オープンで非営利でアカデミックな形で活動し、持続性とその条件である世界的広がりを達成する。
- 当面の利用人口の確保
 - JAXAなどのデータを早期にD5A化し、多くの人にZap-Overに触れてもらう。

まとめ・謝辞・雑感

- インターネット上に散在するデータを組み合わせて使う研究が易しく便利になった。
- 宇宙科学情報解析シンポジウムという開かれた場があったおかげでここまでたどり続けることができた。特に山本幸生先生のご指導に感謝します。
- また、**Apollo 11号～17号**が観測した地震計のデータがあったことがこの2年間の研究を推進した。観測データが幅広い分野の進歩を促していると実感している。

本研究は、総務省戦略的情報通信研究開発推進事業（SCOPE）の独創的な人向け特別枠「異能vation」プログラムの支援を受けたものです。