

次世代月惑星探査を担う巨大時系列 データへの高速なアクセス手法の開発

APOLLO 月震計データ配信を支える検索技術とその応用

- 山本 幸生¹ 平田 成² 古庄 晋二³ 飯沢 篤志⁴ 生座本 義勝³ 佐藤 悠³ 星野 達也³ 成川 絃太³ 伊藤 秀典³ 長尾 正⁵
- ¹JAXA ²会津大学 ³株式会社エスペラントシステム
- ⁴リコーITソリューションズ ⁵Layman's Admin

月惑星探査で求められるデータハンドリングとは？

- 探査機側
 - 高速なCPUや大規模なメモリを搭載しにくい
 - 高密度な半導体製品は放射線耐性が弱い
 - 電力に制約がある
- 地上側
 - 時系列データが大規模になりやすい
 - 1秒に1回のデータでも一年経つと3000万レコード以上になる
 - カラム数が多い
 - 探査機はセンサーの塊なので数1000~数万のカラム数となる
 - ある程度の期間を経過した後は更新がほとんどない
 - 直近ではデータレコーダの再生により古いデータが届く
 - 異なるサンプリングタイミングの時系列データが並ぶ
 - 機器ごとに異なるサンプリングタイミングでデータを出力する
 - 複数のサブシステムでデータ処理が行われデータソースがサーバに分散される

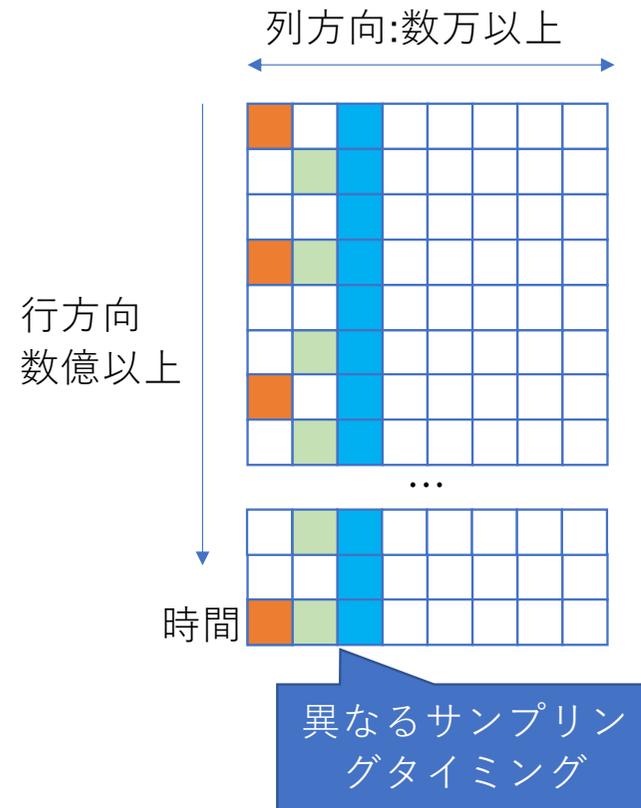
必要とされるデータ形式とは？

- 必須の要求

- 行方向に数億以上、列方向に数万以上を扱えること
- 異なるサンプリングタイミングのデータを格納できること
- (更新系は遅くて構わないが)取得は高速であること

- オプションな要求

- オーバーヘッドが軽いこと
- 副問い合わせが高速なこと
- 過剰なチューニングを避けられること



会津大学 ARC-Space 公募型共同研究 令和3年度
「巨大時系列テレメトリデータへの高速アクセス手法の開発業務」

- Apolloの要件

- 巨大なデータを高速に取り扱えること
- チューニングは最小限(専門家不要)

+

- JAXAの衛星運用としての要件

- 異なるサンプリングタイミングの高速な結合が可能なこと
- 1万以上のカラム数を扱えること

(株)エスペラントが持つD5Aというシステムを拡張

D5Aシステムの特徴

- 特徴

- 全てのカラムが内部的にソートされ(転置)インデックスがついている
 - 自然数インデックス

- メリット

- オーバーヘッド小(問い合わせ開始から応答開始までの時間小)
- 少ないリソースのマシンで稼働
- データベースのメジャーな機能はサポート (SORT, UNION, SELECT, 一部JOIN)
- サンプルング周期の異なるデータの結合 (COMBINE)
- チューニング不要 (CSVファイルに型情報があればD5Aに変換可能)

- デメリット

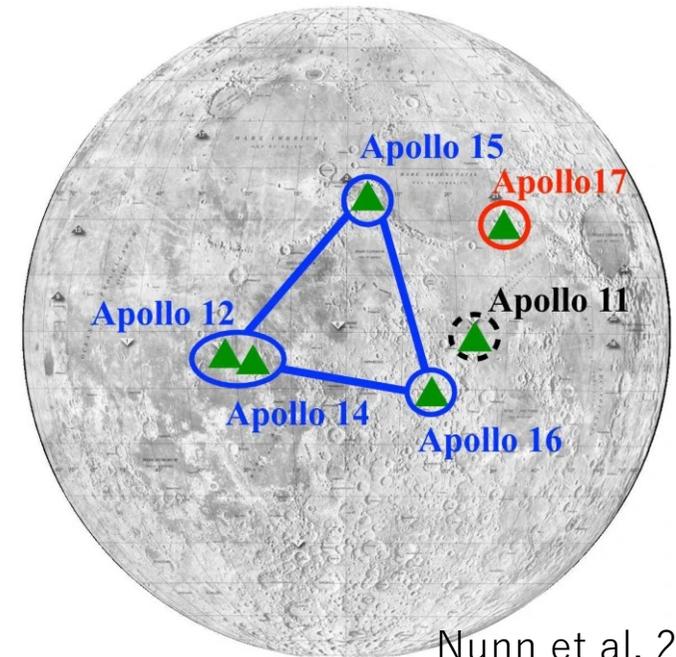
- D5Aを作成するのに時間がかかり更新頻度の高いデータには不向き
- ブロックサイズが大きく小さいデータが苦手
- 型の種類が少ない

時間方向に長いデータの例

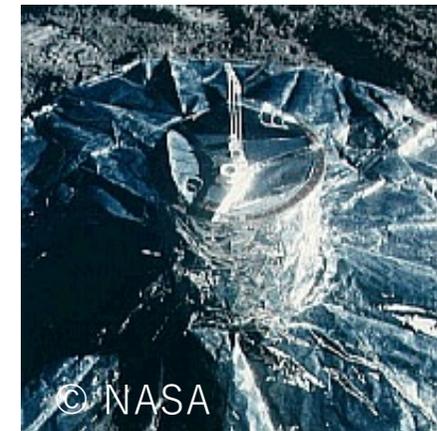
Apollo月震計データ

Apollo月地震データとは?

- Apolloミッションのサイエンスパッケージ(ALSEP)の一つである地震計等による
人類が唯一保有する月震データ
- テキサス大中村教授によりアーカイブされ、その後ISASに送付
- 独自フォーマットで総計110GBのデータ
- 生データがJAXAのDARTSウェブサイトからダウンロード可能
- CSV形式でデータをダウンロード可能なビューワMoon Seismic Monitorを開発(2012~2020)



Nunn et al. 2020



2012年に開発したシステムの問題点

- User Interfaceの問題
 - Flashを用いていた → 後に脆弱性のためアプリごと終了
 - 画面上で波形をポイントしても値表示が困難だった
- Databaseの問題
 - データベースに汎用のPostgreSQLを用いておりオーバーヘッド大
 - 月震データの区間を3時間とするとデータ転送が10分以上経っても始まらない
 - さらに伸ばすとハングアップ
 - 高速化のために専門家を雇用し多大なチューニングを実施
 - Array型とPL/pgSQLによる展開

たかだか110GBでなんでこんなことになるのか？

Apollo月地震データのデータ構造

ALSEP word 1 CONT. 905	ALSEP word 2 CONT. 59	ALSEP word 3 CONT. 380	ALSEP word 4 SPZ 0	ALSEP word 5 LSM 1023	ALSEP word 6 SPZ 0	ALSEP word 7 SWS -	ALSEP word 8 SPZ 0
ALSEP word 9 LPX 524	ALSEP word 10 SPZ 0	ALSEP word 11 LPY 495	ALSEP word 12 SPZ 0	ALSEP word 13 LPZ 497	ALSEP word 14 SPZ 0	ALSEP word 15 SIDE -	ALSEP word 16 SPZ 0
ALSEP word 17 LSM -	ALSEP word 18 SPZ 0	ALSEP word 19 LSM -	ALSEP word 20 SPZ 0	ALSEP word 21 LSM -	ALSEP word 22 SPZ 0	ALSEP word 23 SWS -	ALSEP word 24 SPZ 0
ALSEP word 25 LPX 525	ALSEP word 26 SPZ 0	ALSEP word 27 LPY 494	ALSEP word 28 SPZ 0	ALSEP word 29 LPZ 497	ALSEP word 30 SPZ 0	ALSEP word 31 SIDE -	ALSEP word 32 SPZ 0
ALSEP word 33 HK 35	ALSEP word 34 SPZ 0	ALSEP word 35 TidX 331	ALSEP word 36 SPZ 0	ALSEP word 37 TidY 234	ALSEP word 38 SPZ 0	ALSEP word 39 SWS -	ALSEP word 40 SPZ 0
ALSEP word 41 LPX 525	ALSEP word 42 SPZ 0	ALSEP word 43 LPY 495	ALSEP word 44 SPZ 0	ALSEP word 45 LPZ 496	ALSEP word 46 CV 0	ALSEP word 47 SIDE -	ALSEP word 48 SPZ 0
ALSEP word 49 LSM -	ALSEP word 50 SPZ 0	ALSEP word 51 LSM -	ALSEP word 52 SPZ 0	ALSEP word 53 LSM -	ALSEP word 54 SPZ 0	ALSEP word 55 SWS -	ALSEP word 56 SIDE -
ALSEP word 57 LPX 525	ALSEP word 58 SPZ 0	ALSEP word 59 LPY 495	ALSEP word 60 SPZ 0	ALSEP word 61 LPZ 496	ALSEP word 62 SPZ 0	ALSEP word 63 SIDE -	ALSEP word 64 SPZ 0

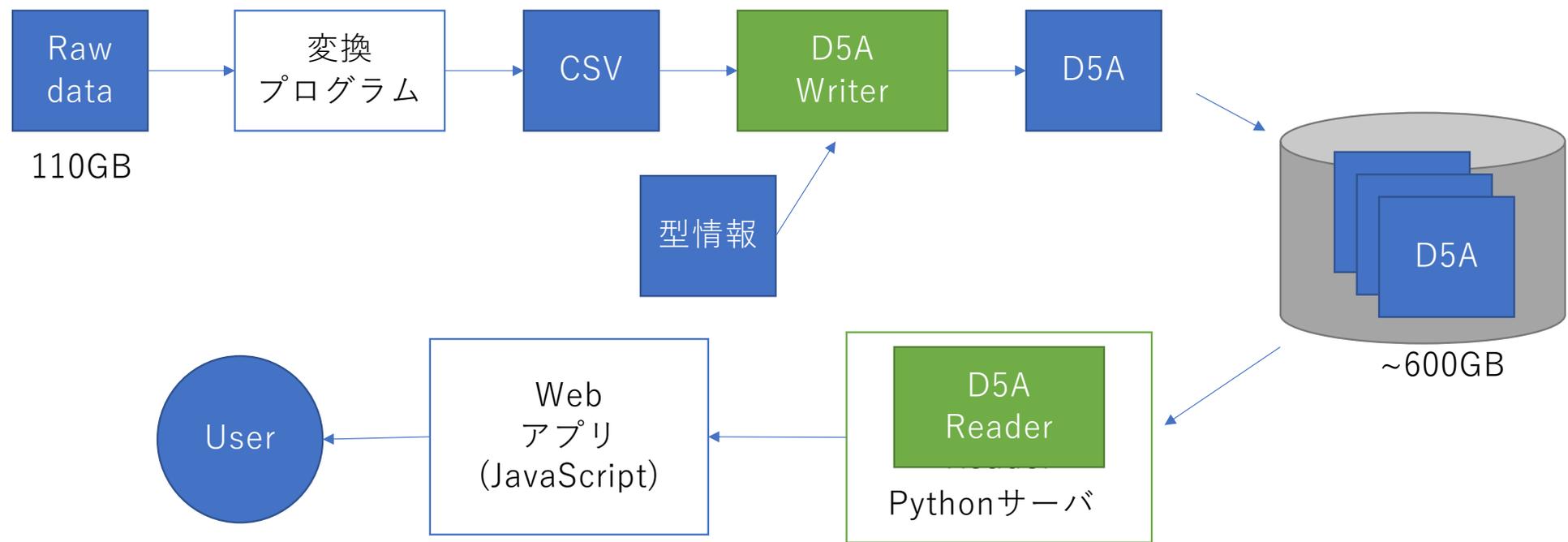
- 1フレームを0.604秒で伝送
- SPZが31個 ~ 53Hz
- LPX/LPY/LPZが4個 ~ 6.625Hz

53Hzを1年間運用すると
 $53 * 86400 * 365 \sim 16.7$ 億

実際の運用期間は7.5年
 さらに号数Apollo12,14,15,16,17

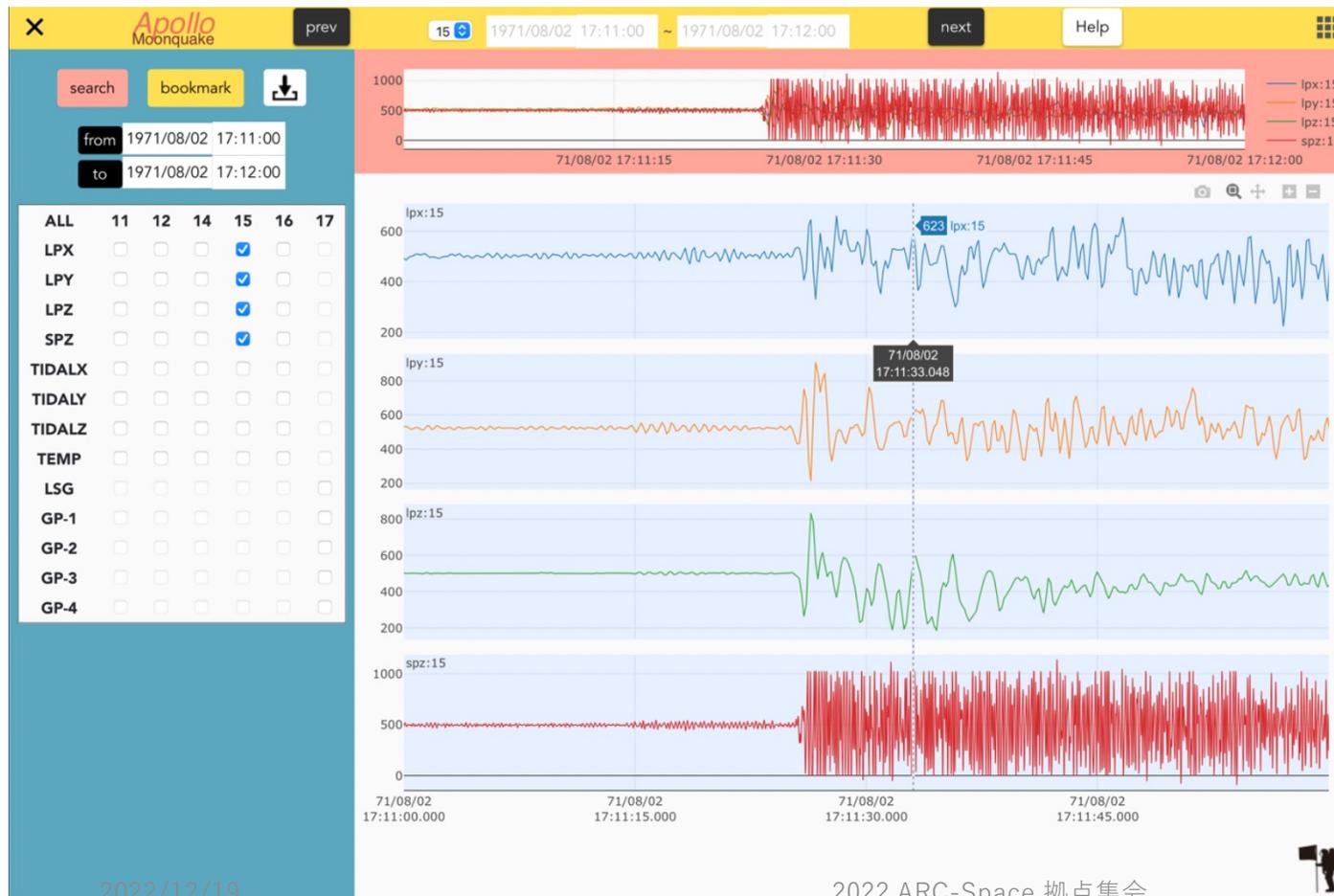
~10億レコードをほどほどのマシンで
 高速に扱いたい

D5Aフォーマットへの変換と利用



新しいApollo月震システム

<https://app.darts.isas.jaxa.jp/apollo-moonquake>



- ウェブアプリとして十分な速さ
- 表示範囲を1日にしても軽快
- CSVでダウンロード可
- D5Aでもダウンロード可
- マウスオーバーで値表示

カラム数が多いデータの例

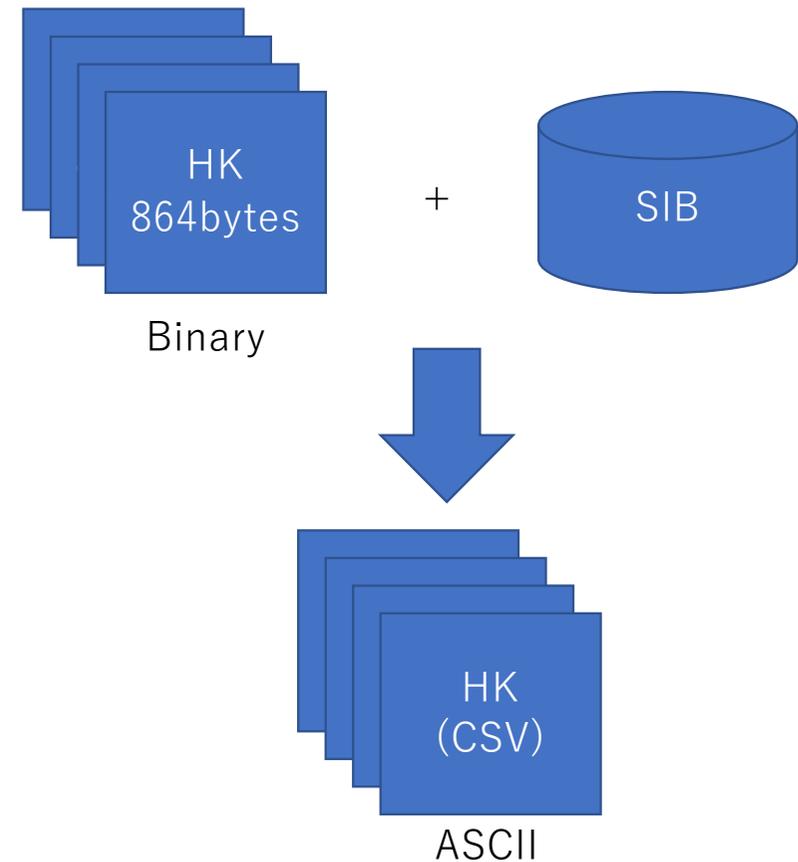
探査機のHKテレメトリ

HKパケット

- House Keeping Packetの略
- 特徴
 - どの**宇宙機でも必ず出力**する
 - 出力周期は宇宙機によって様々
 - 含まれる内容・パケットへの詰め方は宇宙機によって様々
- ON/OFFなどのステータスや、電圧や温度などのアナログ値をデジタル値にした値が含まれ、これらの設計情報はSatellite Information Base(SIB)として管理されている

宇宙科学研究所でのHKの管理

- EDISONと呼ばれるシステムでSIBを元にバイナリのHKデータをCSVに変換
- CSVデータを一定期間ごとに区切り圧縮して提供
- データを部分的に切り出して提供する機能あり
- データ量の少ない宇宙機についてはRDBMSによる検索機能あり

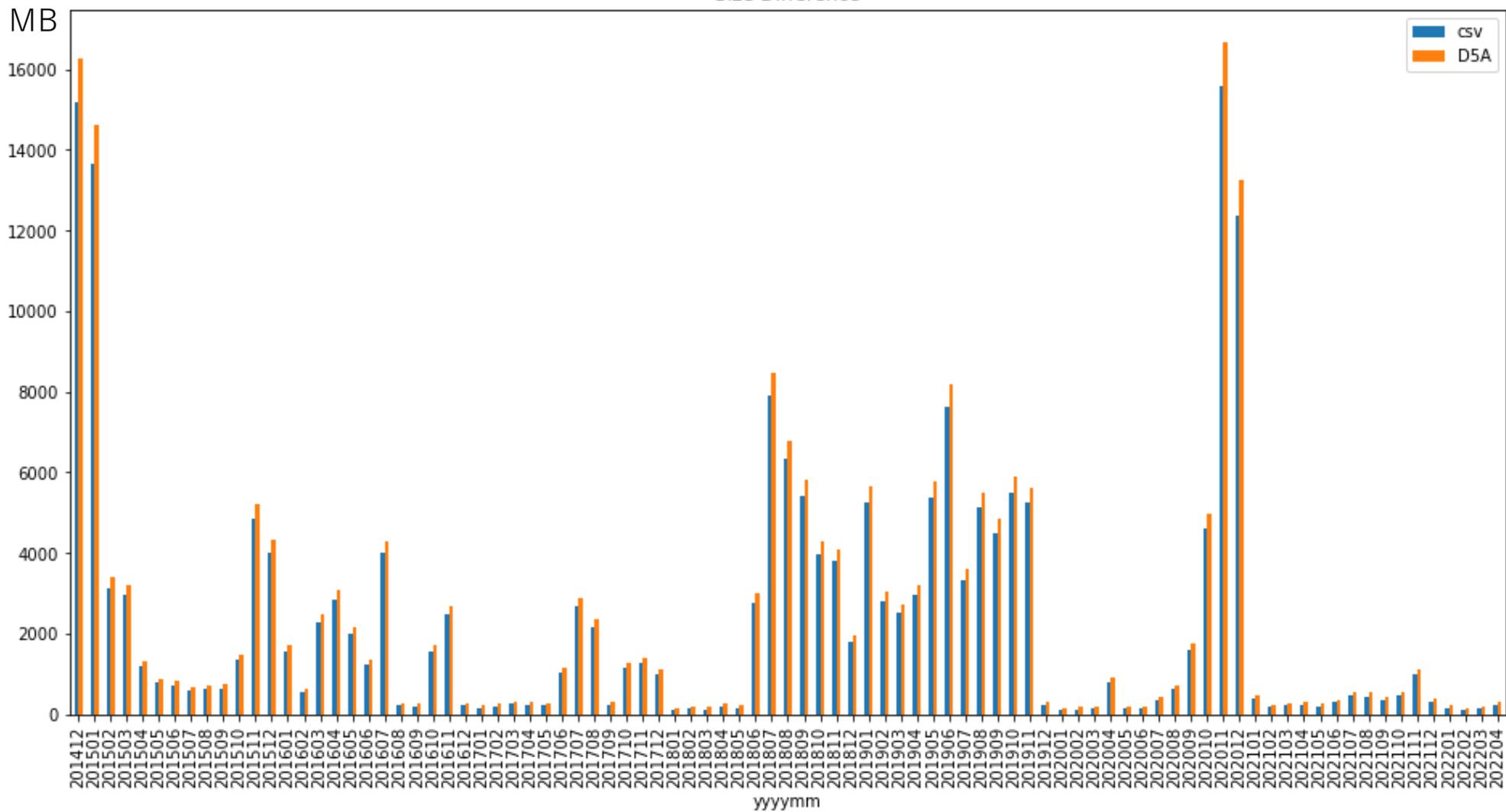


Hayabusa2のHKをD5A化

- データ内容
 - 1683カラム(時刻情報含む)
 - timestamp/string/int/double
- データ準備
 1. CSVを用意する
 2. Confファイルを記述する
 3. D5Aに変換する
- 使い方
 - PythonのWrapperから呼び出して使用

```
source "csv/201412.csv"
dest   "d5a/201412.d5a"
work   "work"
VNo    -
ORG    +
COMPRESS +
timestamp TIME
string AOCs_HK.AOCU_THR_F_ENA
string AOCs_HK.AOCU_THR_O_ENA
string AOCs_HK.AOCU_THR_F_DIS_PERMIT
string AOCs_HK.AOCU_THR_O_DIS_PERMIT
string AOCs_HK.AOCU_ON
string AOCs_HK.AOCU_OFF_PERMIT
string AOCs_HK.AOCU_RES_PERMIT
string AOCs_HK.AOCU_ACM_X_ON
string AOCs_HK.AOCU_ACM_Y_ON
string AOCs_HK.AOCU_ACM_Z_ON
string AOCs_HK.AOCU_ACM_S_ON
string AOCs_HK.AOCU_RW_A_ON
string AOCs_HK.AOCU_RW_B_ON
string AOCs_HK.AOCU_RW_C_ON
string AOCs_HK.AOCU_RW_D_ON
```

Size Difference



まとめ

- 本開発では、データ量の多い例としてApollo月震データを、カラム数の多いデータ例として探査機のHKテレメトリを用いた。
- Apollo月震データではアプリケーションを作成し、バックエンドとして動作するエンジンをPostgreSQLからD5Aへと変更した。ウェブアプリケーションとして十分な動作速度を実現した。
- 探査機HKテレメトリでは1ヶ月単位ごとにデータをまとめ、20GB弱のデータをD5Aに変換した。

巨大なデータをお持ちで遅くて困っているという方はご相談ください