

国際化に向けたiDolphinの再構築  
-ISO13606準拠医療情報規格をHBaseにインプリメントする-  
＜HBase説明とデモ＞

デュアルカナム株式会社  
大松 重尚

【はじめに】

本プログラムでは、京都大学EHR共同研究講座 小林慎治先生、桑 直人先生、吉原博幸先生の説明を受け、その後に、ISO13606準拠医療情報規格のデータに関するHBaseの概略説明と、デモを実施して実現性、動作性を披露するものである。

【デモシステムのHBaseについて】

下記にて説明(資料1)・

Seagaia Meeting 2014

## 1/HBaseとは

Apache HBaseはオープンソースの、列指向、分散データベースである。

GoogleのBigTableをモデルとし、Javaにより書かれている。

Hadoopプロジェクトの一部として開発され、HDFS (Hadoop Distributed File System)の上で実行され、Hadoopに対しBigTableのような機能を提供する。

HBaseはBigTable論文に従い圧縮、インメモリ処理の機能、および各列ごとにブルームフィルタを持っている。

HBaseにおけるテーブルはHadoop上のMapReduceジョブの入出力として機能し、Java APIのほか、REST、Avro、ThriftといったゲートウェイAPIを通じアクセスが可能である。  
近年ではパフォーマンスが向上し、FacebookのMessaging Platformなど、データ駆動型のWebサイトでも使用されている。

出典: [http://ja.wikipedia.org/wiki/Apache\\_HBase](http://ja.wikipedia.org/wiki/Apache_HBase)

©ubiquitousid@gmail.com

2

### 資料1

【非構造データと構造化データ】

画面にて説明。

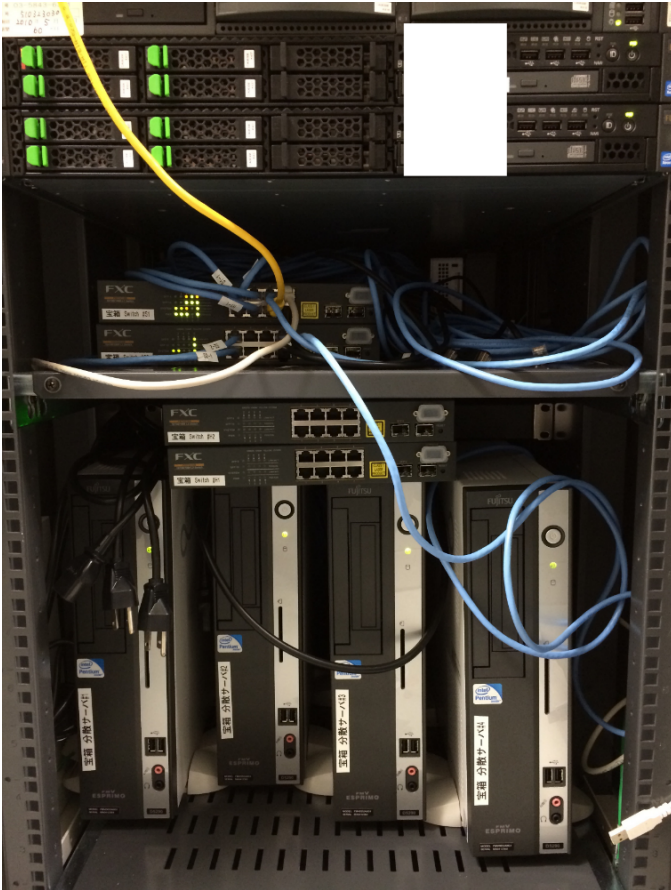
【デモデータについて】

研究用のデモとして、プレゼン内での説明に留める。

- ① HBaseに格納されているデータの説明 ※デモを一部実施。
- ② HBase格納データ詳細(構造化して表現)

【デモ環境について】

当社使用の大手通信キャリア内データセンター内に仮設置したPC4台分でのハードウェアに設定した。  
(資料2)



資料2

### 【デモについて】

Wifi経由でデモ環境と接続し、Web画面での動作を投影し、実施内容とその高速処理とを説明する。  
デモの項目については、以下の通り。

- ① ISO13606とHBaseについて  
ISO13606規格データをHBaseに取り込み、データ状態を画面表示。
- ② 非構造データ取り込み構造化データへ  
検歴データのCSV(非構造データ)をHBase上で構造化して取り込んでいる状態を参照。
- ③ HBase/ImpalaでのSQLライク処理について  
HBaseに取込んだISO準拠のデータを、Hiveを経由してImpalaで検索。  
(デモ 患者毎に受けた検査数をカウント)
- ④ Hbase/Impalaでの高速処理について  
HBaseに取込んだISO準拠のデータについてHiveを経由してImpalaで検索。  
患者検索、検査結果表示を行い、HBaseのデータをImpalaで高速に処理可能を実証。  
(患者情報:100件、検歴情報:1,000件) ※検歴一覧／あいまい検索を実行
- ⑤ HBase/Impalaによる大量データ処理について  
HBaseに取込んだ2億件のデータを特定条件でHiveを経由してImpalaで検索。  
一意なデータ(2億件)に、複数の特定マスタを結合して、条件選択毎にグルーピングした件数を取得。

以上。