xAPI規格に準拠した学習履歴の長期保存及び活用の試行

齊藤智也[1)] 矢吹渓悟[1)] リーラーラクスキット[1)] 王躍[1)] 爲末隆弘[1)]

1) 山口大学ICT基盤センター

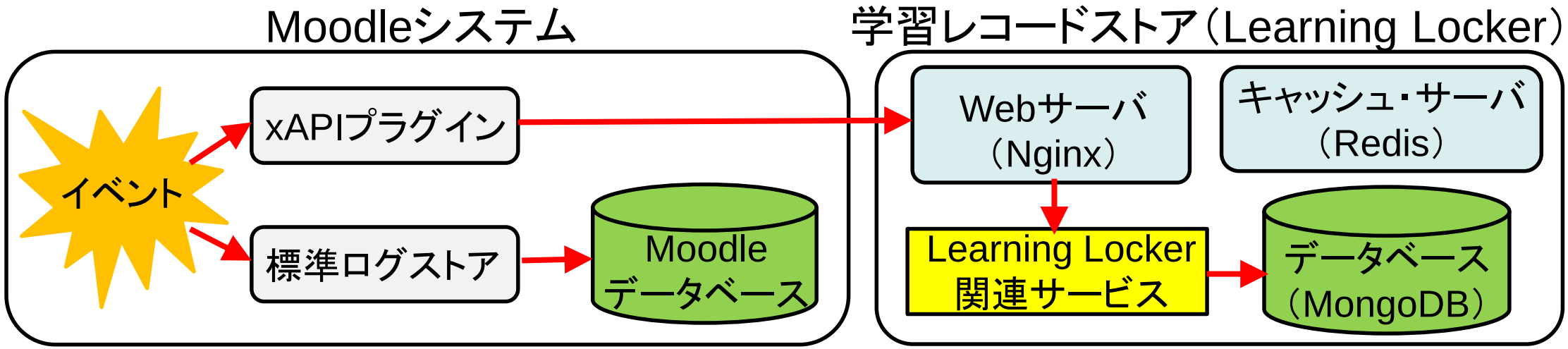
山口大学におけるxAPI準拠の学習履歴の保存

■Learning Locker（オープンソースの学習レコードストア）を利用。

- xAPI形式の学習履歴をMongoDB（オープンソースのNoSQL型のDBMS）に保存。

■学習履歴の蓄積に関する問題点

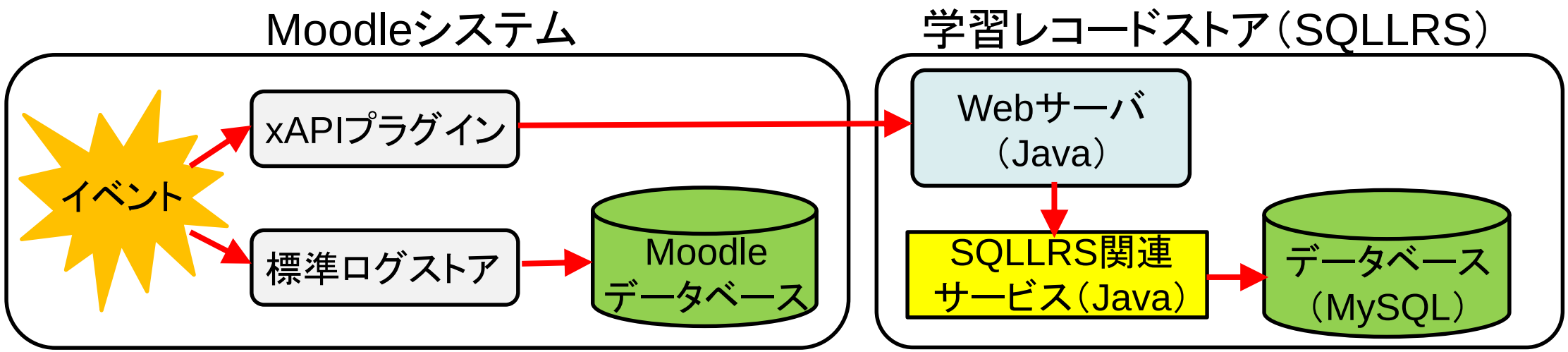
- 保存されているレコード数の増加に伴ってMongoDBの応答速度が低下。
  - ➢ 半年程度（1000～1300万件）の履歴を保存すると、データの追加のタイムアウトが頻繁に発生。
- データの検索が低速。
  - ➢ 検索キーに相当するフィールド（学籍番号、コースID、コンテンツのID等）の値を抽出するためには、xAPI形式（JSON形式）のテキストデータの探索が必要。



SQLLRSの採用(1/2)

■学習レコードストアのソフトウェアをSQLLRSに変更。

- リレーショナル型データベースを使用するオープンソースの学習レコードストア。
  - ➢山口大学ではデータベースとしてMySQLを採用。

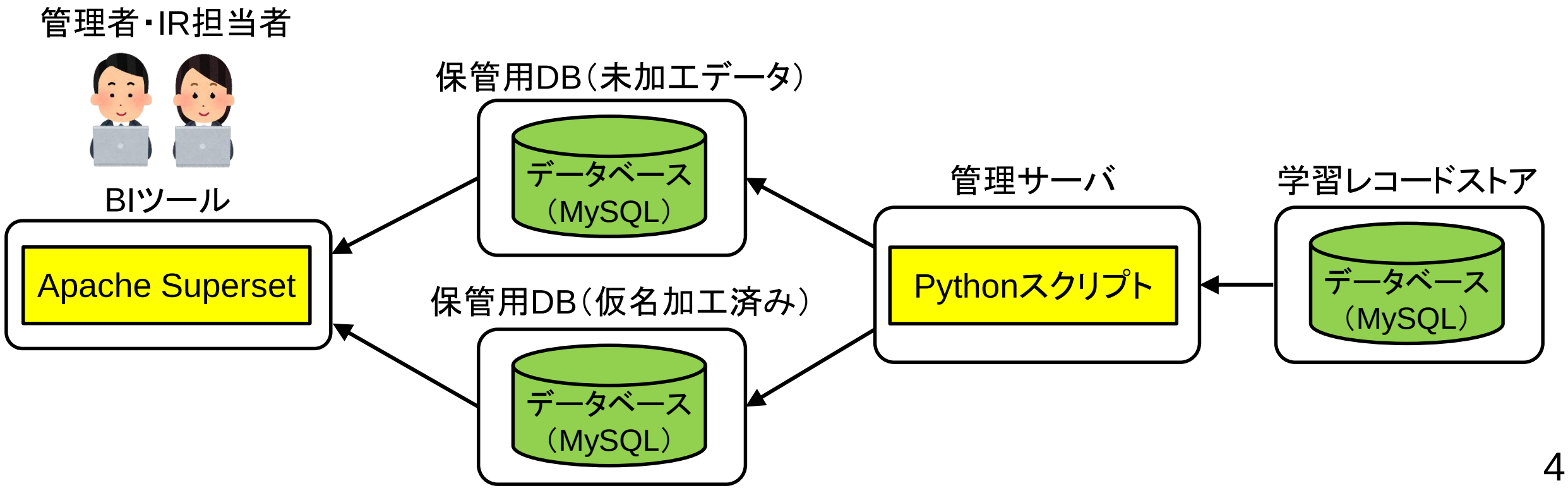


SQLLRSの採用(2/2)

■Pythonスクリプトによって定期的に学習履歴を保管用データベースに移動。

■BIツール(Apache Superset)を用いて可視化。

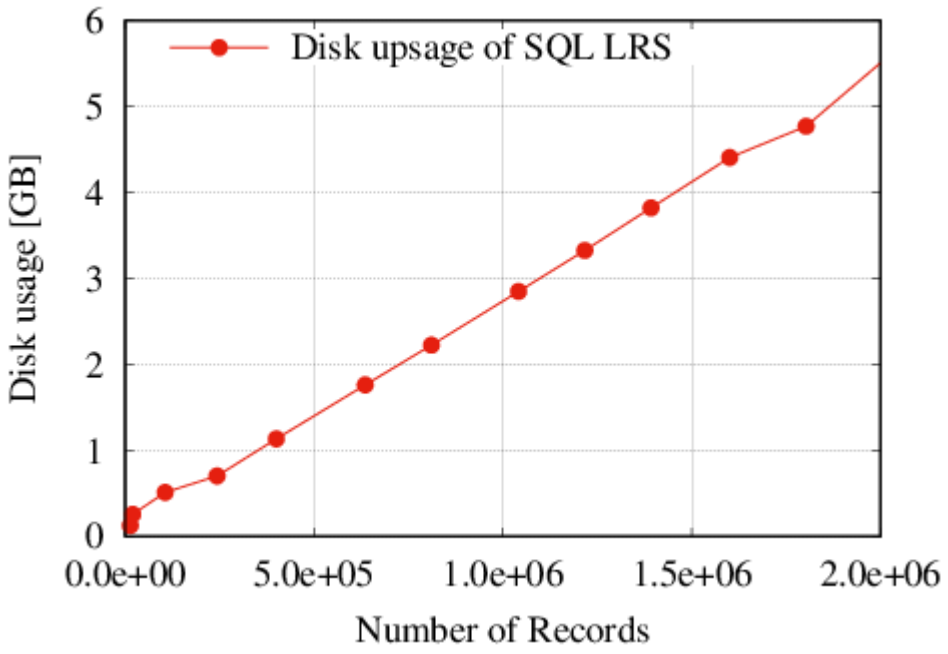
- 一部のアカウントは仮名加工済みデータのみアクセス可能。



SQLLRSのファイル使用量の削減

■SQLLRSのデータベースのファイル使用量が3年分（約9000万件）で1TB前後に到達すると推測。

- 山口大学のMoodleの利用科目数は1600科目（全体の30パーセント）程度。



■ 初期設定ではデータベースのROWフォーマットが「DYNAMIC」なので、データベースの内容をダンプした後、「COMPRESSED」に変更してリストア。
  - 3年分で250GB程度までディスク使用量を削減。
  - 100万件で2.8GB程度（左側のグラフ）。

■ 1日1回、長期保管用データベースに学習履歴を移動。
  - 長期保管用データベースでもROWフォーマットには「COMPRESSED」を使用。

検索キーになり得る属性項目をフィールドとして追加

YAMAGUCHI UNIVERSITY

■長期保管用データベースでは、検索キーになり得る属性項目をテーブルのフィールド（カラム）として採用。

- タイムスタンプ
- ステートメントID
- サイトURL（山口大学では複数のMoodleサイトがあるため）
- メールアドレス（ユーザの識別）
- イベントソースのURL
- コースID
- モジュールID
- Moodleのイベント名
- 関係者のメールアドレス（教師や採点者の識別）
- xAPIの本文

これまでに判明している注意事項

■長期保管用データベースのファイル使用量は3年分で約80GB。

- mysqldumpコマンド等でダンプデータを出力すると420GB以上になる。
- 物理バックアップの機能を使用してデータをバックアップ。

■MySQLではSQLクエリの実行結果のデータサイズは1GBが上限。

- 検索結果のデータサイズが1GBを超える場合、実行結果を取得できない。
- SQLクエリの設計や実行結果の読み出しに工夫が必要。

■非圧縮時に100GBを超えるデータベースの検索は極めて低速。

- 年度ごと、もしくは学期ごとにデータベースを分割する対策が必要。

まとめ

■xAPI形式の学習履歴の長期保存を試行。

- 学習レコードストアとしてSQLLRSを採用。
- 定期的に学習履歴を保管用データベースに移動。
- Apache Supersetを用いた可視化を試行。

■今後の課題

- 長期保管用データベースを年度ごと／学期ごとに分割して、データベースごとのレコード件数及びディスク使用量を縮小。
- LMSのログで「何が分かるのか？」の調査・検討が必要。
  - ➢山口大学ではようやく学習分析の活動を開始したところ。