

xAPI規格に準拠した学習履歴の長期保存及び 活用のための基礎的検討

齊藤智也¹⁾ 王躍¹⁾ 河野綸華²⁾ レールマルク¹⁾

1) 山口大学情報基盤センター

2) 山口大学創成科学研究科

背景

- Moodleに蓄積される学習履歴の抽出・分析のため、オープンソースの学習レコードストアとして普及しているLearning Lockerを採用。
 - Learning Lockerは、LMSの学習履歴をxAPIに準拠した形式で蓄積。
- 既存のxAPIプラグインにはいくつかの不具合が見られるほか、収集可能なイベントの種別や問題タイプが少ないため、プラグインを修正して対応。
- 学習レコードストアに蓄積される学習履歴のレコード件数の増加に伴い、データベース・サーバの負荷が増大。
 - MoodleからLearning Lockerへのデータの書き込みが失敗するようになる。

今回の発表内容

- 山口大学の学習レコードストアの現状, 及びレコード件数の増加に伴う負荷の増大の状況。
- 複数年度にわたる長期保存への対応について, 現時点での構想。
- 昨年度の発表以降の, xAPIプラグインの修正内容。

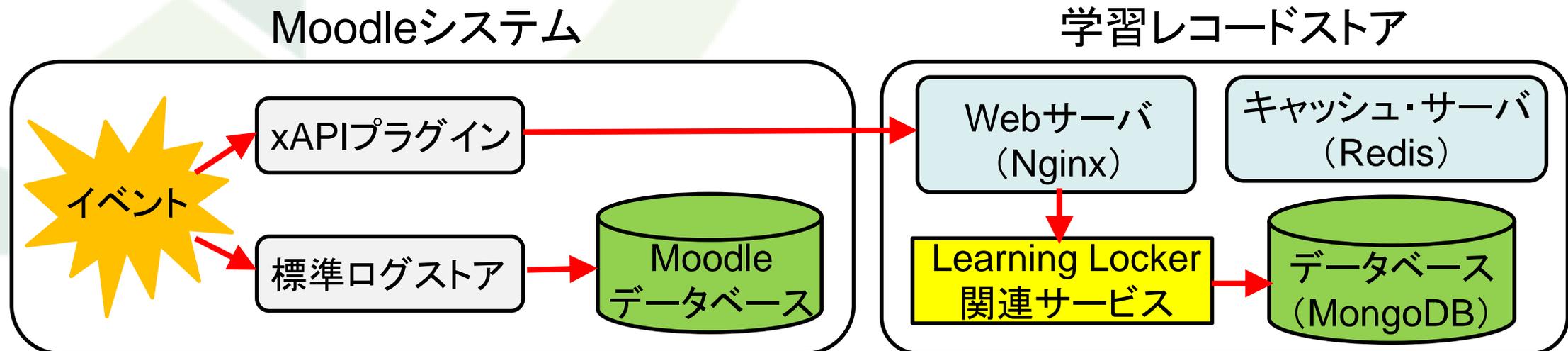
Learning Lockerの概要

■ Learning Pool社が提供するオープンソースの学習レコードストア。

- 山口大学ではオンプレミス環境に無償版を導入したが、有償のSaaS版もあり。

■ xAPI形式の学習履歴をHTTP/HTTPS経由で受信し、データベース(MongoDB)に保存。

- Moodle側のxAPIプラグイン(ログストアの一種)により、学習履歴をxAPI形式に変換。
- Moodleのプラグイン(の発生イベント)ごとに変換用のPHPスクリプトが必要。

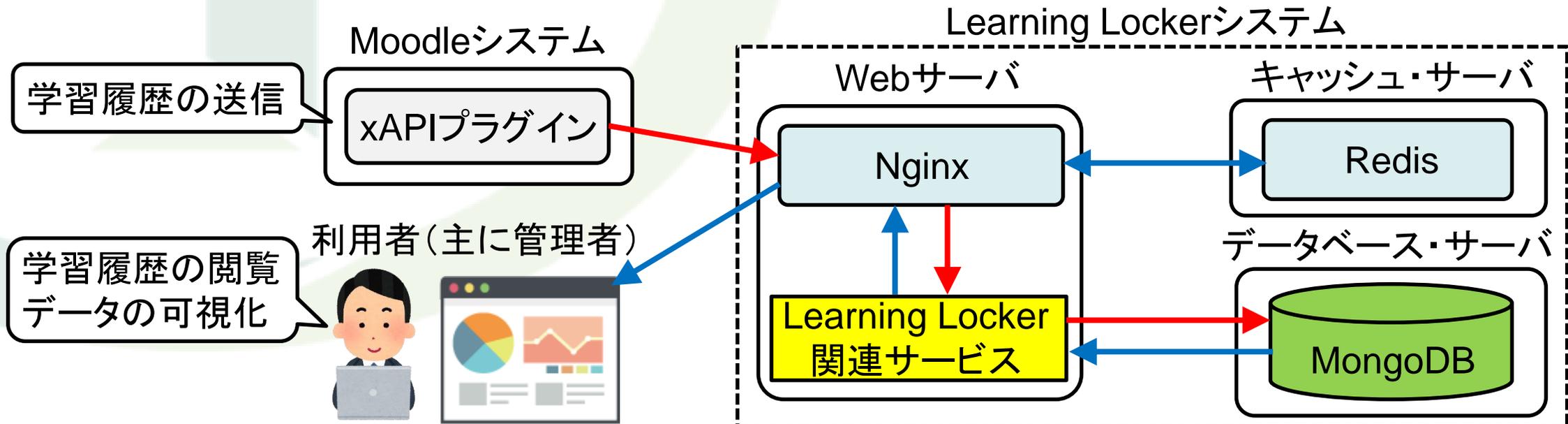


Learning Lockerの構成(山口大学の場合)

■WebサーバにLearning Lockerをインストールした後に機能を分離。

- RedisやMongoDBに格納されているデータを別のサーバにコピー。
- WebサーバにおいてLearning Lockerの設定ファイルを編集。
 - データベース・サーバ及びキャッシュ・サーバの参照先を変更。

■Learning Locker関連のサービスもそれぞれ別々のサーバに分離可能。



データベース・サーバの負荷の増大

- Learning Lockerではデータベース管理システムとしてMongoDBを採用。
- MongoDBに蓄積される学習履歴のレコード件数の増加に伴い、データベース・サーバの負荷が増大。
 - Learning Lockerはデータの挿入時に重複チェックを行っているらしく、MongoDBにとっては、保有するレコード数に応じてチェックにかかる時間が増大。
 - データの書き込みに要する処理時間が徐々に長くなり、最終的にはMongoDBへの書き込みの失敗が発生するようになる。
 - 山口大学のMongoDBの場合、保有レコード件数が1300万件を超えたあたりからエラーが発生するようになる。
 - 1年間に蓄積されるレコード件数が1800～2000万件なので、1年分も蓄積できない。
 - このあたりの件数になると、MongoDBのメモリ使用量が30GB以上になる。

現行のデータベース・サーバのスペック

■ベースとなるブレードサーバのスペック。

- CPU: Intel Xeon Silver 4216 (16コア/32スレッド) 1個。
- メモリ: DDR4 256GB。
- SSD: SATA-3接続, 1.92TB (2個のディスクによるRAID構成)。
- NIC: 10Gbps対応。
- 仮想化のハイパーバイザ: VMware ESXi 6.0。

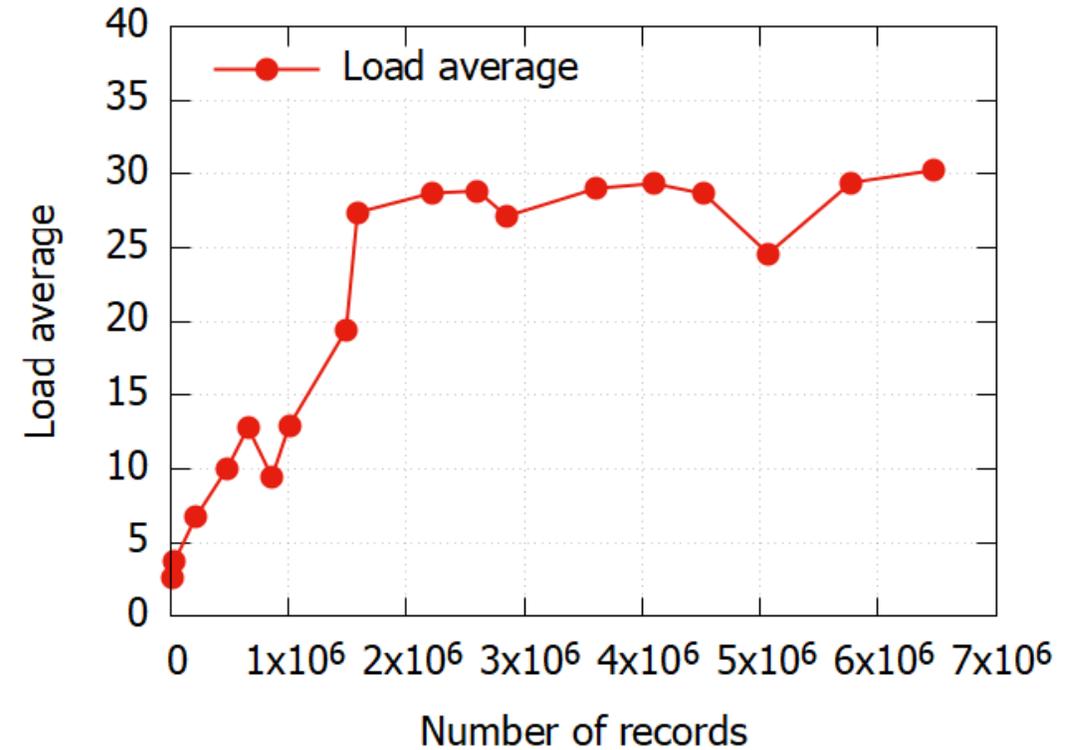
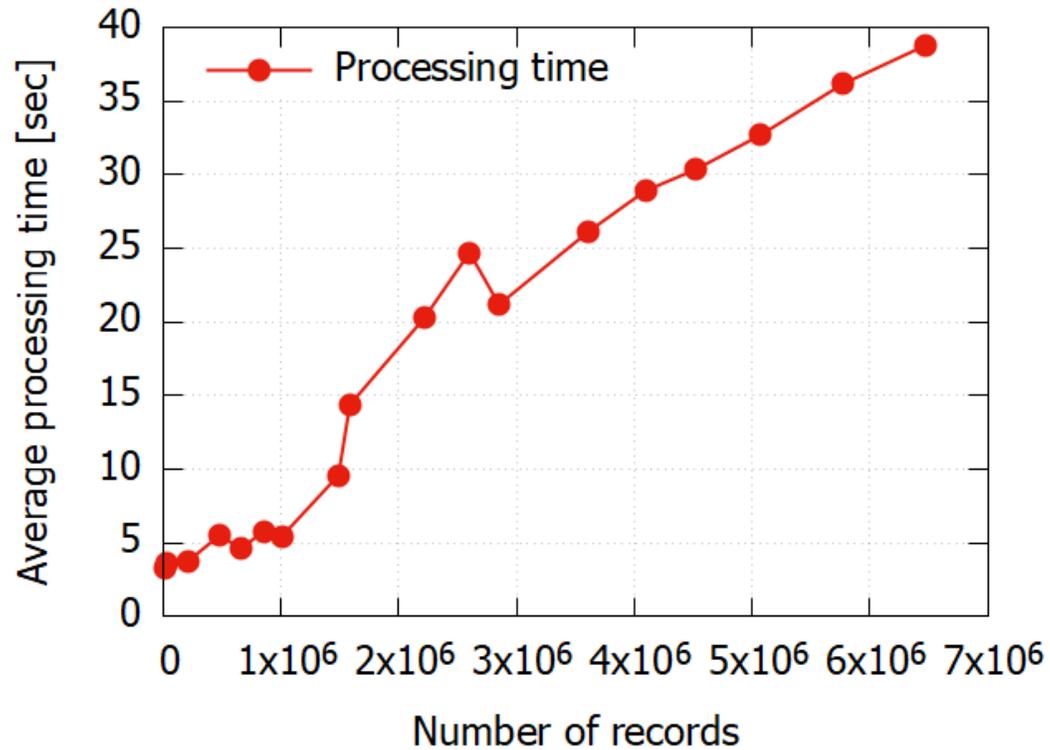
■MongoDBを稼働させている仮想サーバのスペック。

- 仮想CPUは8コア, 仮想メモリは64GB, 仮想ディスクは200GB。
- 3台のブレードサーバにWebサーバ, DBサーバ, キャッシュサーバを分散させているが, 各サーバはブレード内では他の仮想サーバと同居。

レコード件数に対するDBサーバの負荷の変動を計測

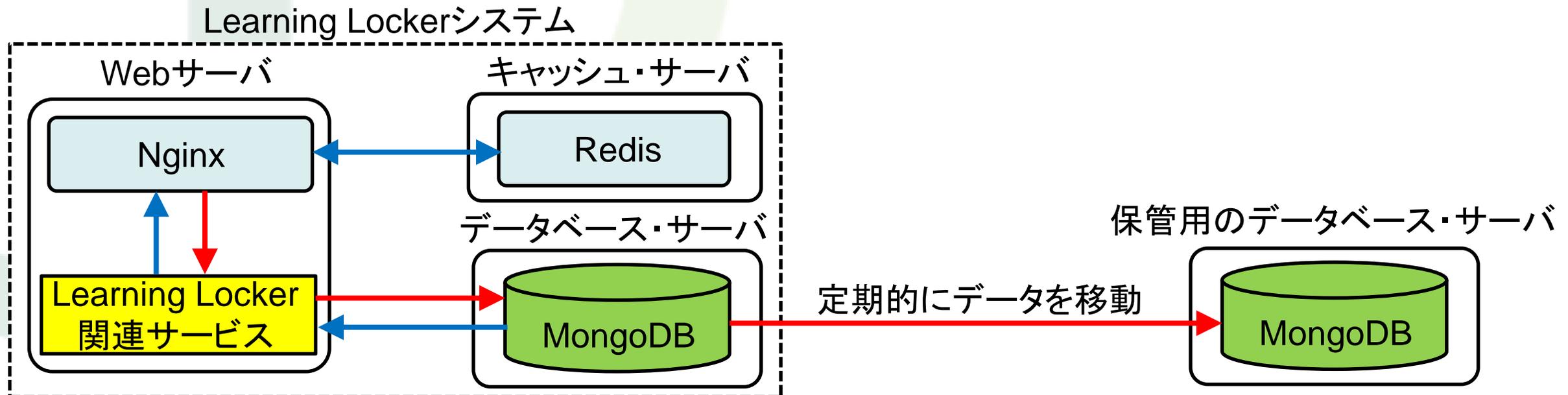
■ Moodleのスケジュールタスクにより1分ごとに学習履歴を書き込み。

- 1回あたりの最大書き込み件数は300件に設定。
- 書き込み件数が300件に達している回の処理時間を5つ抽出し、平均値を算出。
- 授業の開始・終了時刻に負荷が増大するため、それらの時刻のロード・アベレージの最大値を計測。



対応策の構想

- 学習レコードストア内のデータベース・サーバは、学習履歴を受け取るだけで高負荷となるため、学習分析への活用が困難。
 - 別のデータベース・サーバに定期的にデータを移動する必要がある。
 - 保管用のデータベース・サーバをどのような構成にするか？



現状の構成案は2つ

■長期的な学習履歴の保管に十分なスペックのデータベース・サーバを整備。

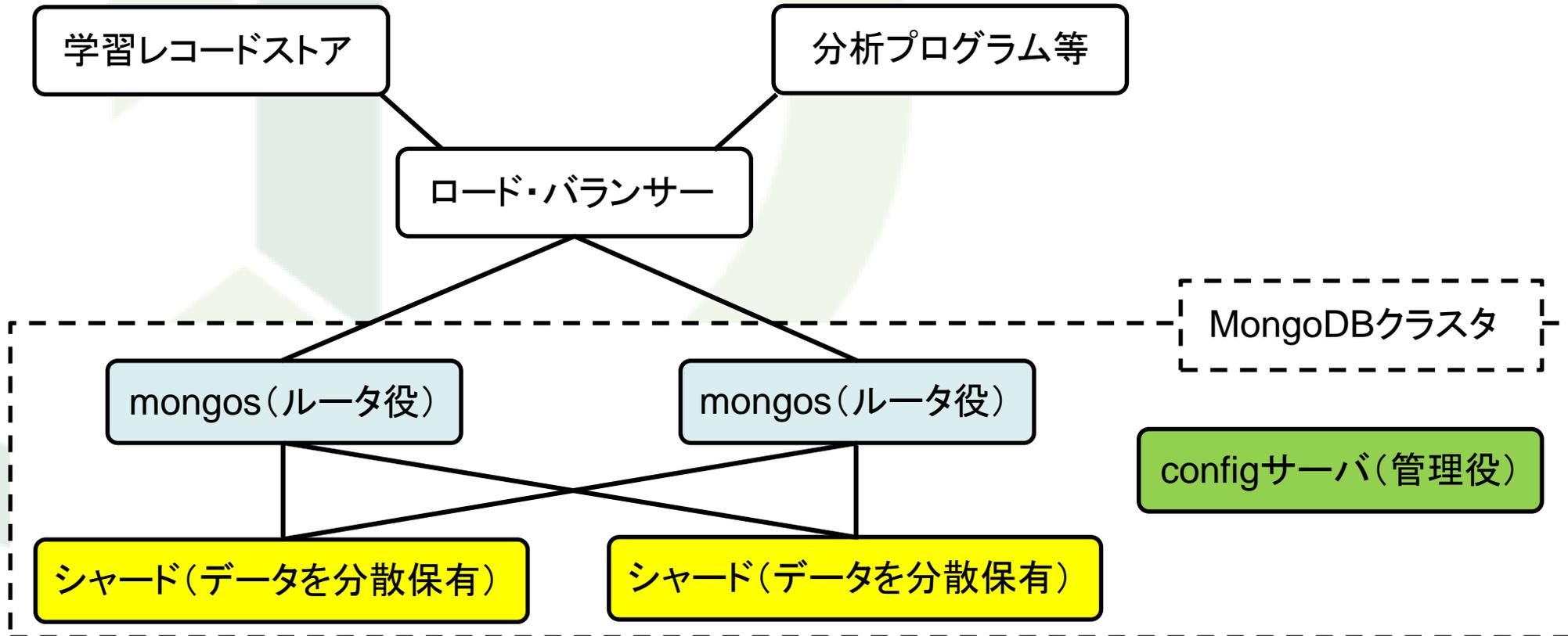
- 学習レコードストアのデータベース・サーバに対し、数倍のスペックが必要だろう。
- サーバの整備は比較的容易だが、ハードウェアに障害が発生した際に、それだけ大きな仮想サーバの引っ越し先が見つかるかどうか・・・

■MongoDBクラスタを整備。

- 構成要素となる各サーバはそれほど大規模ではないが、最小構成でも5台のサーバが必要。

MongoDBクラスタ

- 実験としては5つの構成要素を1台のサーバ内で稼働させることも可能。
- 実運用環境では最小構成で5台のサーバが必要。

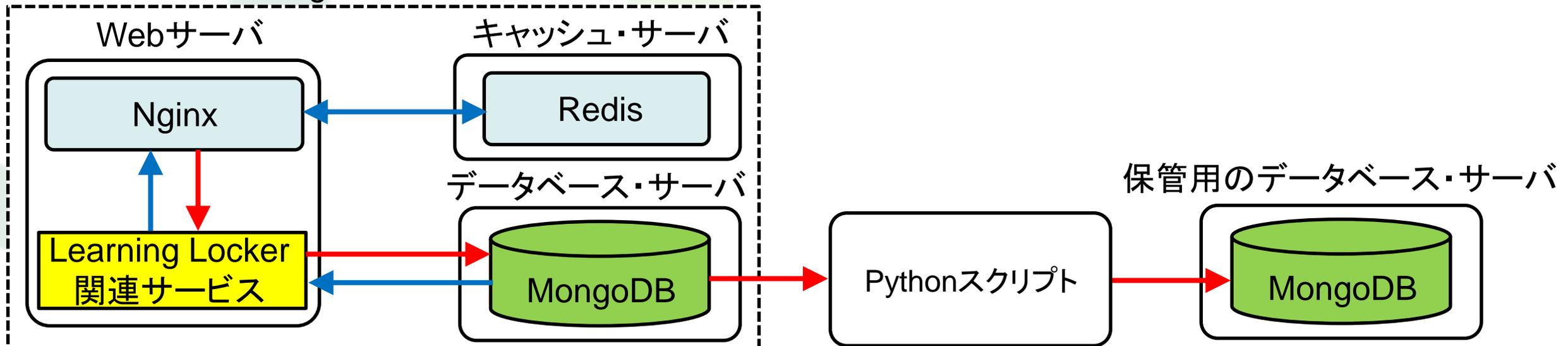


現在の応急措置

■メモリが64GBある仮想サーバでMongoDBを稼働させ、学習レコードストアのMongoDBにあるデータを定期的に移動。

- 「learninglocker_v2」データベース内の、「statements」コレクション。
- その他にユーザ関連のコレクション。
 - 学習レコードストア内での各ユーザのID(仮名のようなもの)が記録。

Learning Lockerシステム

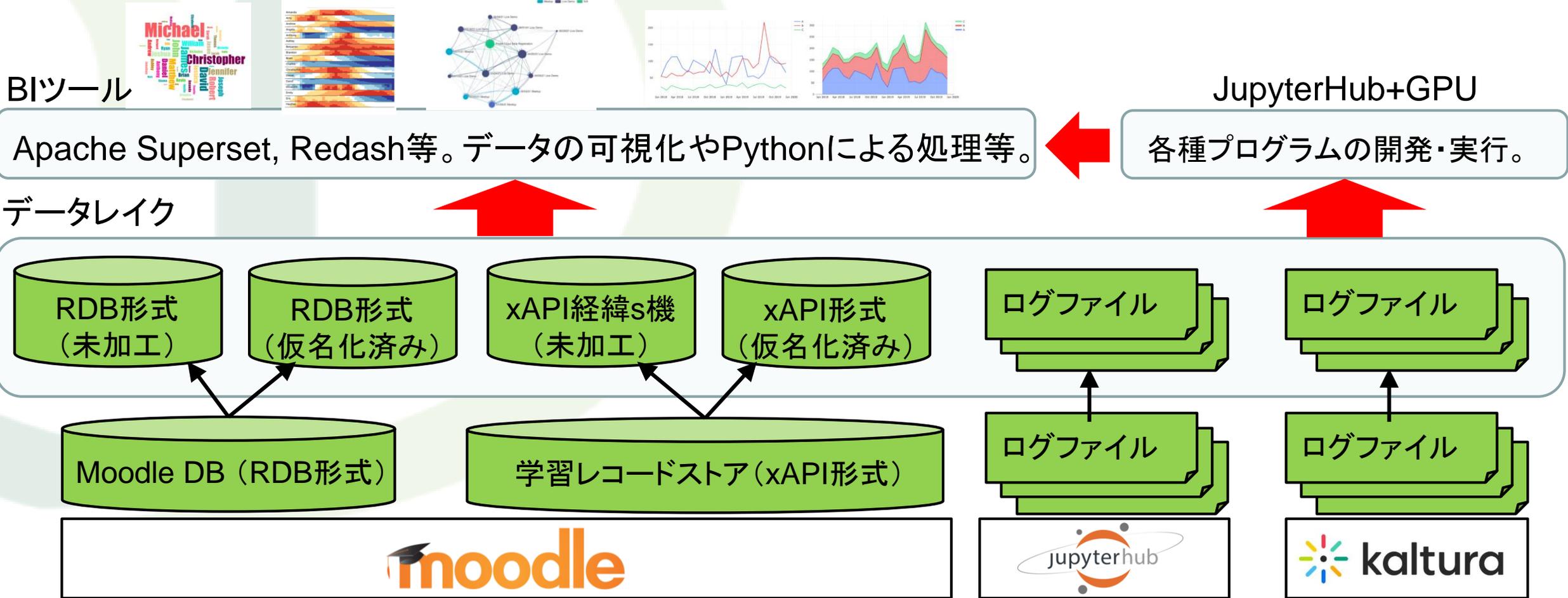


山口大学の事情

- 学習分析の内容や用途，着目するデータによっては，MongoDBのようなキーバリューストアにJSON形式のデータを格納して利用するのではなく，従来のリレーショナル・データベースを利用するほうが適切な場合もある。
- 山口大学の場合，情報基盤センターは教育データの分析に関与していない。
 - 教育データの分析の担当は，教育・学生支援機構の中の各部署。
 - しかし，そちらの各部署の方針が定まっておらず，活動も開始されていない。
 - 教育DXに関連して学習分析を進める必要がある？という漠然としたニーズだけ。
- 学習分析の担当者が各種データを閲覧・分析する環境の整備が必要。
 - テスト版を構築して，担当者に体験してもらわないと，明確なニーズが出てこない。
 - おそらく，リレーショナル・データベースとJSON形式の双方が必要になる。

教育データを提供する基盤 (NIIの学習分析基盤をヒント)

- 相変わらずソフトウェア関連の予算がゼロなので、オンプレに自作する？
- 大学内で研究・開発に活用するには、仮名化／匿名化済みのデータも必要。



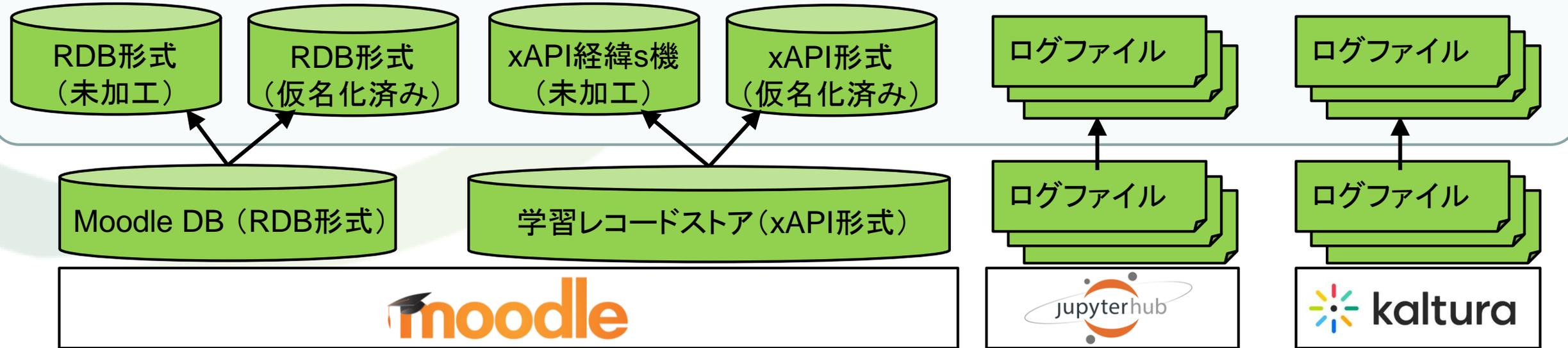
BIツール

JupyterHub+GPU

Apache Superset, Redash等。データの可視化やPythonによる処理等。

各種プログラムの開発・実行。

データレイク



xAPIプラグインの修正

- 以前には2020032801をベースにしていたが、現在は2022101800をベースにしたものに変更。
- 「作文」の問題タイプの変換プログラムを修正。
 - 回答が空の場合に発生するエラーへの対処がされていなかったため。
 - 昨年度にはすべての問題タイプについて、空の場合のエラーに対処したと思っていたが忘れていた。
- 独自に修正したプラグインはGitHub上で近日中に公開予定。
https://github.com/YU-MITC/moodle-logstore_xapi
 - 上記のリポジトリの内容を近日中に更新予定。

まとめ

- 学習レコードストアとしてLearningLockerを利用する場合、データベース(MongoDB)の負荷の増大への対応が必要。
 - 長期保管用のデータベース・サーバを構築することとし、その構成について検討中。
 - レコード件数の増加に伴うサーバ負荷の増大の状況を計測。
- 教育担当部署や学内のAI関連の研究者向けに教育・学習データを提供するデータ基盤の整備が必要。
 - 大学としての方針や、担当者のニーズが定かではない。
 - 情報基盤センターも、関連の各種ソフトウェアの動作や問題点を検証している段階。
 - GPUサーバ及びファイルサーバを整備した。その他の設備の調達を計画中。
- xAPIプラグインを若干修正したので、最新版を近日中に公開予定。

今後の課題

- 長期保管用のデータベース・サーバの構成を確定させ、学習履歴の保管・閲覧のためのデータ基盤の整備を遂行。
- 教育データの分析の担当者とのさらなる打合せ。

情報基盤センターの教員を募集中

- インフラ関係と教育システム関係で、助教を1名ずつ。
- 応募期限は2024年5月7日。着任は10月1日までのなるべく早い時期。
- 詳細はJREC-INをご覧ください。

- <https://jrecin.jst.go.jp/seek/SeekJorDetail?id=D124010633>
- <https://jrecin.jst.go.jp/seek/SeekJorDetail?id=D124010632>