

J-PARC MLF BL17 におけるクラウドを使った実験状況モニタリングおよび解析システムの開発

総合科学研究機構(CROSS) 中性子科学センター,
日本原子力研究開発機構(JAEA)J-PARC センター^A,
笠井 聡, 青木 裕之^A, 永井 佑樹, 阿久津 和宏, 花島 隆泰, 宮田 登

Development of Monitoring and Data Analysis Systems using Cloud Applications at BL17 in J-PARC MLF

Neutron Science and Technology Center, Comprehensive Research Organization for Science and Society (CROSS), J-PARC Center, Japan Atomic Energy Agency (JAEA)^A,
Satoshi Kasai, Hiroyuki Aoki^A, Yuki Nagai, Kazuhiro Akutsu, Takayasu Hanashima, Noboru Miyata

e-mail: s_kasai@cross.or.jp

Published online: 28 June 2024

要約

大強度陽子加速器施設(J-PARC)の物質・生命科学実験施設(MLF)に設置されている偏極中性子反射率計 BL17 SHARAKU では、ユーザーが実験を実行中に J-PARC エリア外からその状況を確認するシステムが未開発であった。また、同じく J-PARC エリア外から実験データそのものを得たり、解析を行うことが出来なかった。これらの問題を解決するために、クラウドを用いた実験中のモニタリング機能と、解析システムを構築した。モニタリング機能は、現在の実験により得ているデータの状況を、Web システムによるグラフで確認出来る。また解析システムは、終了した実験に対し、Web を用い J-PARC 内にて解析するのと同じ機能で解析とデータ取得が出来る。クラウドを用い Web 機能によりデータを提供するシステムでは、ユーザー当事者のみにデータの提供を可能とするべく、認証機構を付加した。また Web の特性上短時間での応答が必要であり、解析機能開発ではベースの Python 言語の他に C 言語により作成されコンパイルされたコードを利用して大幅な速度向上を図った。

Abstract

Users at the polarized neutron reflectometer BL17 SHARAKU, which is installed at the Materials and Life Science Experimental Facility (MLF) of the Japan Proton Accelerator Research Complex (J-PARC), could not check the status of experiments from outside the J-PARC. In addition, users could not obtain or analyze experimental data from outside the J-PARC.

To solve these issues, we developed a monitoring function that enables users to check the status of the current experiment via a web system using cloud computing. We also implemented a remote analysis system that enables data analysis and acquisition outside the J-PARC.

We built an authentication mechanism into the system so that users could obtain only the experimental data they had performed. By using the C language in addition to the base Python language, the speed was greatly improved compared to existing analysis systems.

Keywords

J-PARC, SHARAKU, 偏極中性子反射率計、クラウド、リモート監視、リモートデータ解析

1. 緒言

茨城県東海村に位置する大強度陽子加速器施設(J-PARC^{*1})物質・生命科学実験施設(MLF^{*2})では、ユーザーが全世界から来訪し実験等を行っている。ユーザーは来訪期間に施設近隣に宿泊しており、その間 J-PARC イン트라ネットワーク外にも実験状況モニタリングおよびデータの解析を行いたいとの要望があった。MLF において J-PARC 施設外から実験状況を確認する手段は、IROHA2^{*3}による仕組み [1]や各ビームライン^{*4}個別の情報提供機能も使われていたが、今回、開発対象である偏極中性子反射率計 BL17 SHARAKU でイン트라ネットワーク内にて提供されていた、実験中データの可視化による実験状況モニタリングとデータ解析機能を改修し、図 1 に示すようにクラウド(AWS^{*5})を活用しインターネットで公開するシステムを構築することとなった。

インターネット公開においては、イン트라ネット利用では不要であった個人認証の機能、実験状況モ

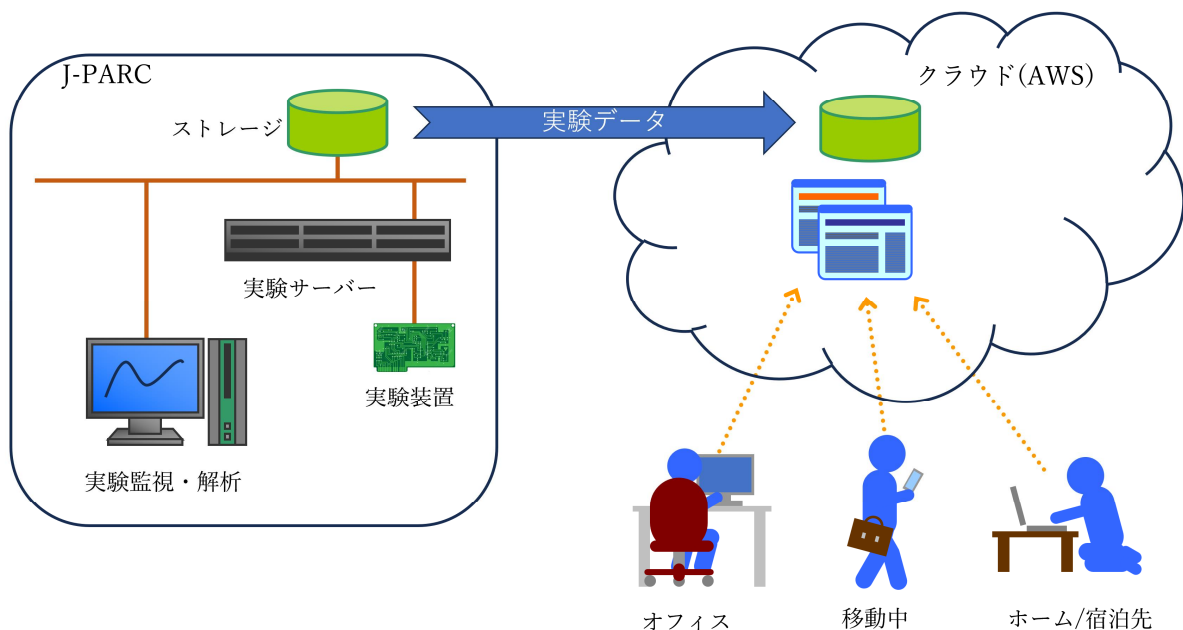


図 1. 実験監視・解析システム概略

ニタリングの即時性確保、Web セッションの時間制限をクリアする増速化などの技術課題を解決し、構築を行った。

^{*1} Japan Proton Accelerator Research Complex

^{*2} Material and Life science Experimental Facility

^{*3} IROHA2 : J-PARC MLF に設置された中性子実験装置の制御や自動測定を行うためのソフトウェアフレームワーク

^{*4} Beam Line : MLF にて中性子・ミュオンビームを様々な検出装置へと導く設備

^{*5} Amazon Web Services

2. 全体構成

2.1 データの流れ

MLF では、実験にて DAQ*6 計算機により得られたイベントデータ*7 はビームラインのローカルストレージに逐次保存される。今回 BL17 SHARAKU で開発したシステムでは、イベントデータは測定単位で終了後 30 分でクラウド*8 へ転送される。また実験中のデータは定時間周期で自動的に解析し、グラフ化されたデータは付帯情報と共にクラウドへ保存される。

2.2 実験状況モニタリング・解析機能

クラウドに保存されるイベントデータは、ひとつの測定が終了した時点で送られるため、その時点で実行中の測定データは表示出来ない。このため、実験状況モニタリングとしての表示データはビームラインに設置の解析計算機がモニタリング用の解析データを作成したものを利用する。完了した測定のイベントデータからは、種々の解析画面で閲覧・結果を得られる (図 2 参照)。解析機能に関しては MLF 来訪者が帰った後もデータの解析とダウンロードが出来る仕組みになっており、利便性は非常に高い。

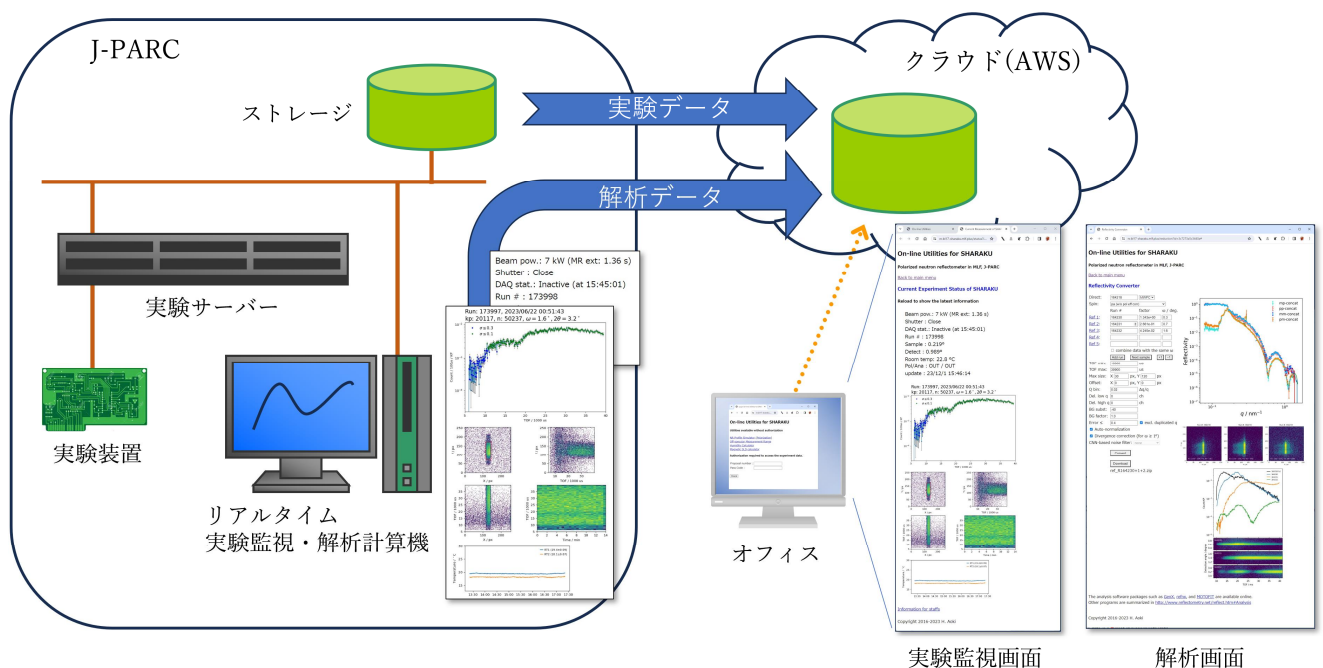


図 2. データの種類と流れ

*6 Data Acquisition System : データを収集し、保存・分析のためにデジタル形式に変換する仕組み

*7 Event Data : この場合は DAQ により収集された生データ (Raw Data)

*8 MLF で取得された中性子測定データ (1 次データ) は一般情報として管理されており、またユーザーは 3 年間優先して使用する権利を有するため、ユーザーやデータ処理の主体者 (ユーザー及び MLF スタッフ等) の許可の下に持ち出しも含めクラウドにおくことも問題はない。当然アクセスはデータ処理の主体者に制限される。

3. 開発のポイント

3.1 認証機能の導入

J-PARC へ来訪し実験を行うユーザーは、自身の実験に対してのみアクセス権が付与され、かつ他の実験にアクセス出来ない仕組みとする必要がある。一般に実験データへのアクセス権限可否を判断するには、測定ごとに付加されるランナンバー (run number) が使われる。しかし実験を開始してからランナンバーが発生するため、来訪したユーザーに許される参照可能なランナンバーは事前には決まらない。このため認証は、一連の実験の時刻範囲とランナンバーの範囲を複数組み、自由に組み合わせるものとした。ビームラインスタッフは、ユーザーの一連の実験単位であるプロポーザルナンバー (proposal number) に対し、実験期間の時刻範囲と共に過去に同じユーザーでの実験があればそのランナンバーの範囲に対して参照の許可を設定する。図3に示すように、この許可内容はスタッフ間で共有スプレッド

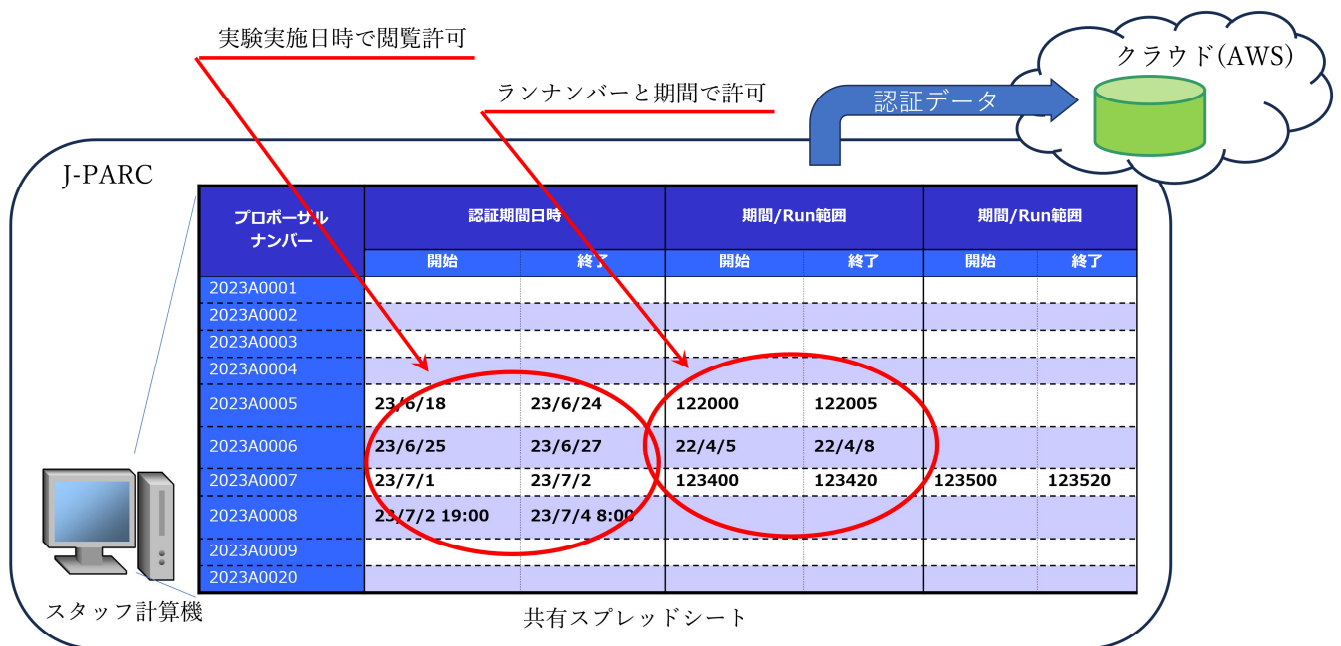


図3. ユーザー認証テーブル

シートで管理され、実験の日程変更時には即座に更新可能である。

3.2 Web 処理の処理高速化の工夫

旧来のビームラインでの解析は、解析計算機上でのアプリケーションにより解析を行っていた。MLFのビームラインで広く使われる空蟬(Utsusemi)解析フレームワーク[2]は、そのひとつである。これらは画面を通した GUI^{*9}を用いアプリケーションとの対話により解析を進めるため、解析作業の途中で計算機内に条件や途中経過データを持ちうる。しかし Web アプリケーションによるインターネットでの処理は、ひとつの画面応答内で処理を完了させる必要があり、画面応答の1セッションの時間にも制限がある。本開発では、Web システムは AWS Lambda^{*10}上にて Docker^{*11}コンテナを用い、ユーザーからの

^{*9} Graphic User Interface

^{*10} AWS Lambda : AWS 上のひとつのサービスで FaaS(Function as a Service)と呼ばれるもの

^{*11} Docker : Docker (ドッカー) は、コンテナ仮想化を用いてアプリケーションを開発・配置・実行するオープンプラットフォーム

接続により個別にコンテナが起動するため、それ以前の解析経緯やデータを持たない処理を前提とした。よって解析計算は1セッションの時間制約で完結することとした。

Web 解析では、イベントデータをまず1中性子ごとの各種数値・状態を線形配列に格納し、そこから各種の解析結果へと計算している。そのイベントデータを線形配列に格納するまでの計算を、システム全体を記述している Python^{*12}言語から C 言語^{*13}にて新たに作成した。C 言語で記述しコンパイルしたバイナリルーチンは、計算機の CPU が直接実行出来るコードで処理が行なわれ、最大のパフォーマンスが得られる。また一般に Python からバイナリーコード呼び出しの処理により変数の値受け渡しがあり、Python のインタプリタ特有のガベージコレクション^{*14}により処理時間が伸びる。対策として、データ受け渡しの方法について大きな線形配列を、データそのものではなく戻り値のポインター（アドレス引き渡し）とした。計算結果は中性子数に比例する大きさであるが、この線形配列の変数の先頭位置（アドレス）のみを返す事により Python からの呼び出しを高速化出来た。

標準的な大きさのイベントデータにて、Python 言語記述による解析と C 言語によるバイナリーコードで処理時間を比較した。結果を表 1 に示す。処理時間は検出器種類（0次元検出器、MWPC^{*15}）、偏極有無などの条件により左右されるため、比較は「Event Data Viewer」解析画面で異なる条件での実行例として示す。また処理時間は、Web の処理が開始し表示に至る直前までのファイル読み込みと計算の合計時間で計測した。その結果、合計時間が5分の1程度となり数秒という十分な増速効果を得られた。

表 1. バイナリルーチン速度比較

RunNo.	イベントサイズ (Kb)	実験環境	Python 実行[A] (sec.)	バイナリー実行[B] (sec.)	比 [A / B]
174571	14,612	0次元検出器 非偏極	5.892	1.223	4.82
168804	17,664	0次元検出器 偏極	3.309	0.776	4.27
168918	45,952	MWPC 偏極	15.609	2.774	5.63

(使用計算機 CPU Xeon Gold 5118 2.30GHz 仮想環境 12 コアで計測)

さらに、現時点でクラウドは2万以上のランナンバーを保持しており、イベントデータのメタデータ（付帯情報）を照会する Web 画面の大幅な速度低下になる。この対策にランナンバーをキーとするデータベースを、照会の都度自動で構成している。

3.3 安全性

Web による公開に伴い安全性に配慮した機能も用意した。ログイン管理では足跡が残り、かつ不審な連続ログイン失敗は拒否される。公開する Web サーバーはゲートウェイにより守られ、またアクセスログにより不審なアクセスを検知出来る。これらのゲートウェイ等を、AWS クラウドの様々なクラウドサービスを組み合わせ構築した。そのシステム構成図を図 4 に示す。

*12 Python：インタプリタ型の高水準汎用プログラミング言語

*13 C 言語：OS 等広く用いられているプログラミング言語で、処理速度が速く機械に直接アクセスできるなどの特徴がある

*14 garbage collection；動的に確保したメモリー領域が不要になると、自動的に開放する機能

*15 Multi-Wire Proportional Chamber：二次元中性子検出器

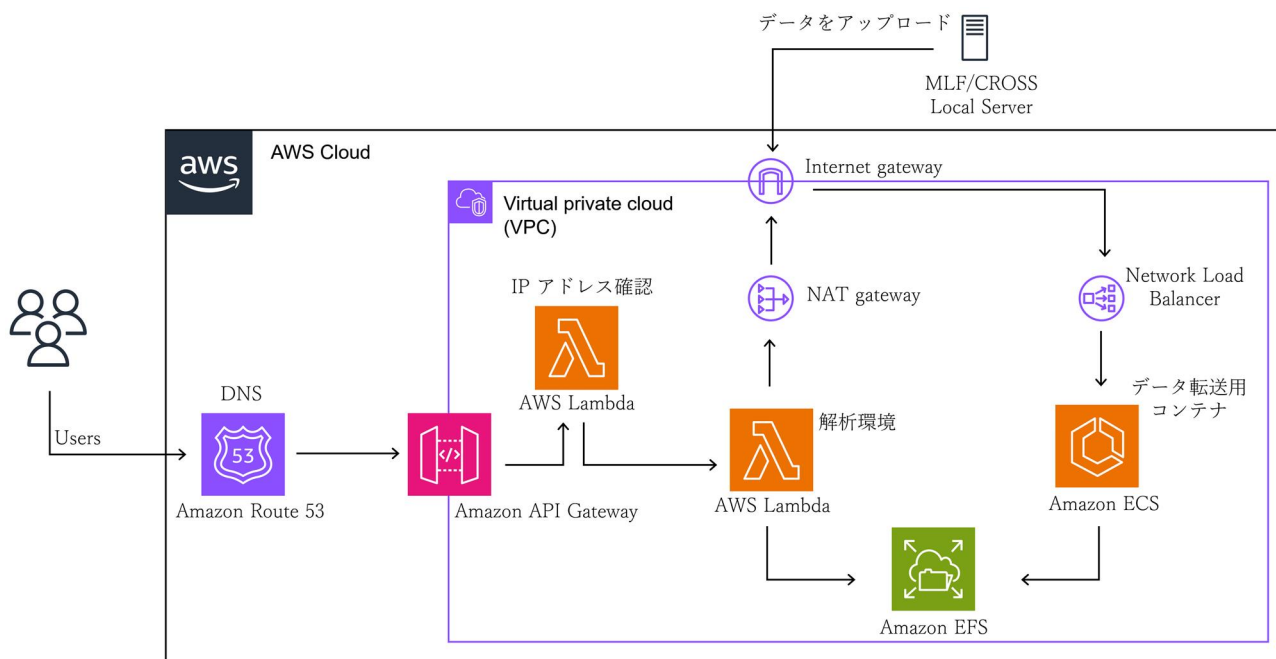


図4.システム構成図

AWS には、種々のサービスがあるが、ストレージとして Amazon EFS*16を利用し、解析計算として AWS Lambda 関数を Docker コンテナで使用した。これらをバーチャルプライベートクラウド（Virtual Private Cloud）の中に構成し、外部からは https の接続に制限して実現している。

このように、クラウド技術を自在に活用することは、今後の MLF へのユーザーサービス拡充に重要である。

4. 総括

実験状況モニタリング・解析計算のクラウド化には、種々の技術開発とサービス連携が必要不可欠である。開発は Python と関連ライブラリを使い行ったが、常時起動の解析計算機とは違い Web のリクエスト単位に起動されるアプリケーションの工夫、サービス起動のためのデータハンドリング設計と Docker 化、Docker 化のための利用ライブラリ管理が必要だった。また、データ管理のための認証機構、認証設定共有化の仕組み、MLF とクラウドの間のデータ同期機能も開発した。さらに十分なパフォーマンスを得るために C 言語による高速化開発も行い、その利用手法も工夫した。また、データベース機能を使い処理速度を向上させる工夫も行った。これら技術開発の積み重ねにより、ようやく BL17 SHARAKU でクラウドを用いた実験モニタリングと解析システムが構築出来た。

解析計算機能のクラウド化により、従来では MLF 内でのみ可能であった中性子のイベントデータを反射率に変換する作業は、インターネットを利用することによりユーザー自身で容易に実施可能となった。なお、解析計算 Web の1年間の利用実績(集計は 2022 年 12 月 20 日～2023 年 12 月 19 日)は 79,957 ページビュー、337 ログインとなり、このうち約 1/10 は日本国外からの接続であった。従って、海外のユーザーにもこの解析計算機能が活用されていることが明らかとなっている。

今後さらにフィッティング等の長時間におよぶ解析機能をどう実現するかなど、新たな課題にも挑んでいく予定である。

*16 Amazon Elastic File System

謝辞

岡崎 伸生(CROSS)さんには、クラウドアプリケーションの設定に関し、数々の助言を頂きました。
稲村 泰弘(JAEA)さんには、クラウドアプリケーションの外部公開に際し、支援と助言を頂きました。
この場を借りて感謝申し上げます。

Reference

- [1] J-PARC MLF IROHA2 ポータルサイト <https://mlfinfo.jp/groups/comp/ja/iroha2.html>
- [2] Yasuhiro Inamura, Takeshi Nakatani, Jiro Suzuki, and Toshiya Otomo, J. Phys. Soc. Jpn. 82, SA031 (2013), <https://journals.jps.jp/doi/10.7566/JPSJS.82SA.SA031>