

ABCIでのOpen OnDemand活用事例と今後の展望

産業技術総合研究所 滝澤 真一郎



- 経産省「人工知能に関するグローバル研究拠点整備事業」(2016年)の一環として整備
 - 大規模なAIモデル開発の場を提供
 - 産総研が培ったスパコン構築・運用技術をAI計算基盤に応用
- 2018年8月1日：**ABCI運用開始** NVIDIA V100
 - 産学官によるA I 研究開発を加速するオープンイノベーションプラットフォーム
 - 高い計算能力を活用したAI技術の研究開発・実証、社会実装の推進、AI分野の最重要課題への挑戦が目的
- 2021年5月10日：**ABCI 2.0運用開始** NVIDIA A100
 - 経産省「人工知能に関する橋渡しインフラ拡張」によりアップグレードを実現
 - 2024年10月31日運用終了



ABCIの進化

生成AIや量子コンピューティングなど、重要な産業分野での研究開発を推進すべく、ABCIは進化し続けます

2011 D-Wave



1208 servers (5312 GPU) :
0.85Exa AI-FLOPS

2019 Google Quantum Supremacy

2023 QuEra

2025年度早期に
運用開始

2025 ABCI-Q

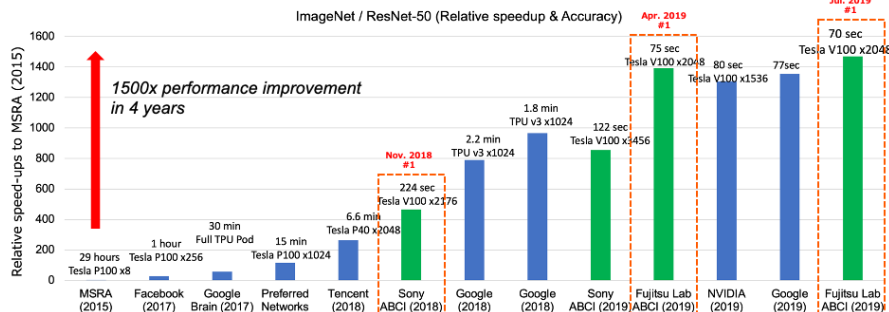
2024 ABCI3.0

2024年11月: Closed Beta運用
2025年1月20日: 正式運用開始

2012 AlexNet

2015 AlphaGo

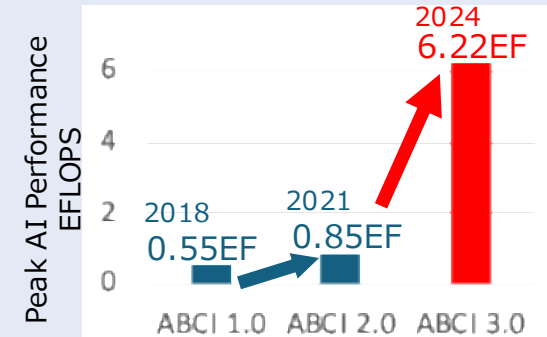
World records in "MLPerf training" benchmark



LLM Building Support Program in 2023



ABCI 3.0



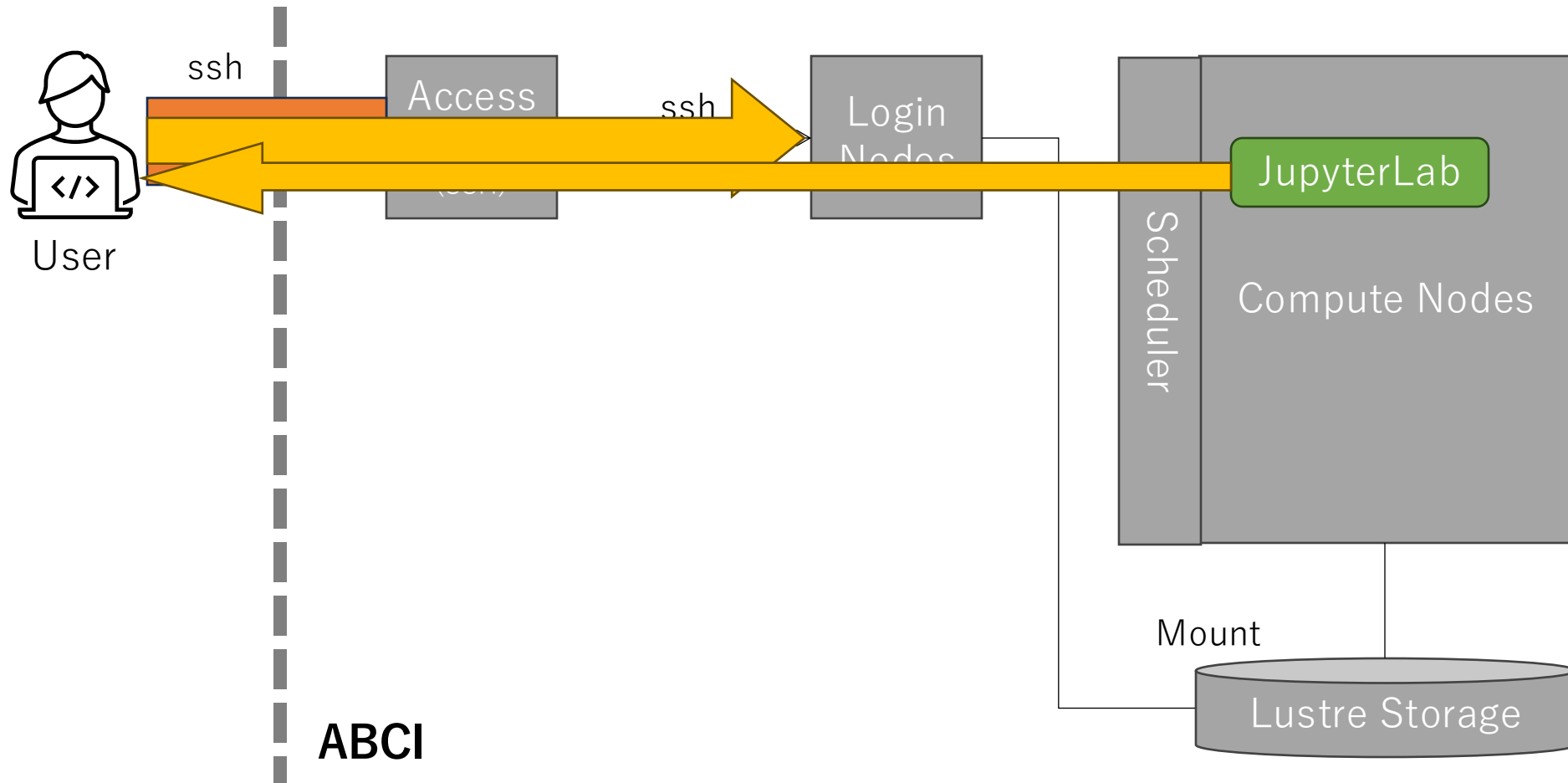
**ABCI-Qは未稼動、ABCI 3.0は稼動直後なので、
ABCI 2.0の時のOpen OnDemandの話をします**

最後に少しだけ、今後の展望を

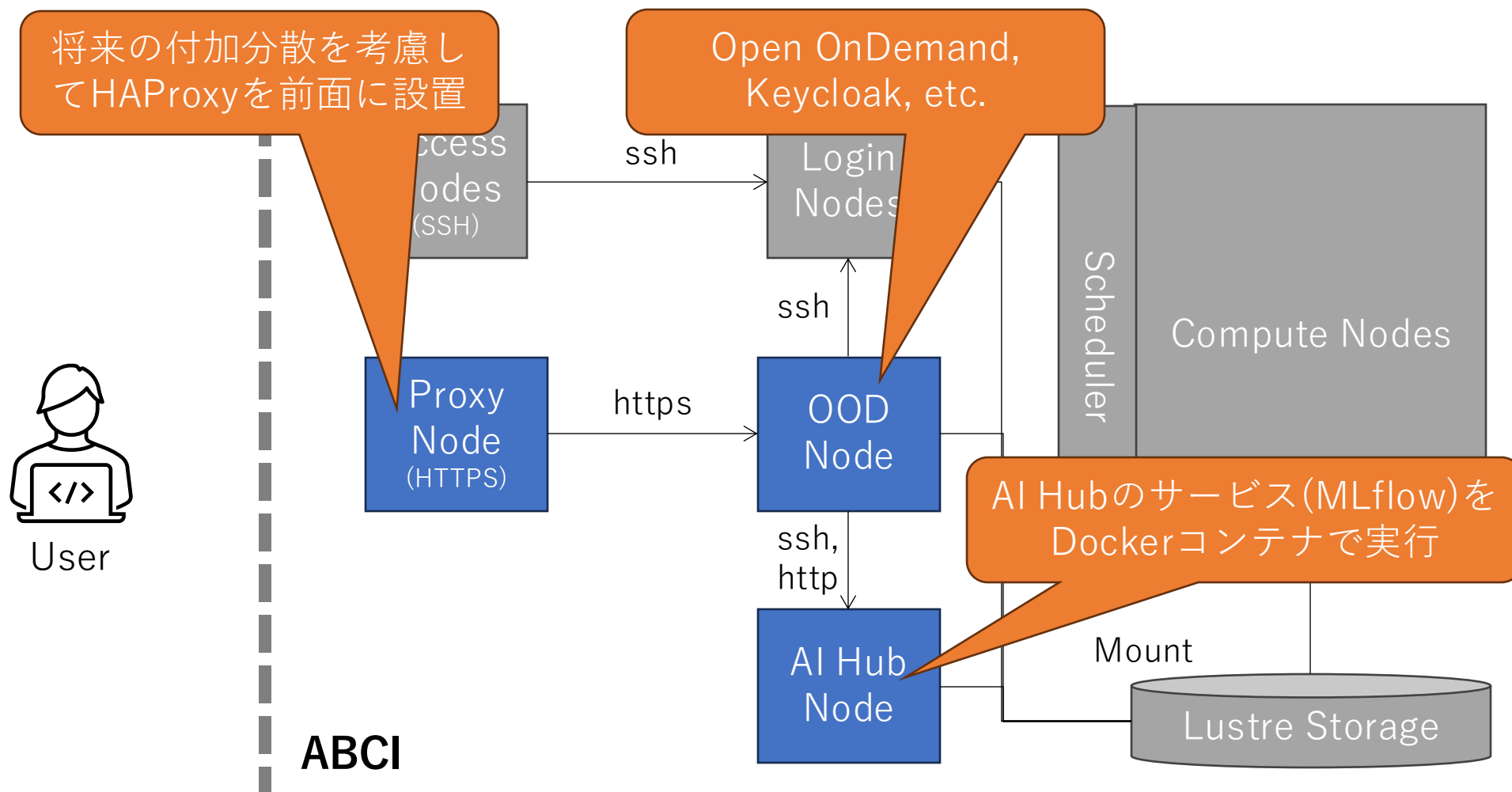
ABCIにおけるOpen OnDemand導入の歴史

- 2022年9月：Open OnDemandに注目
 - 第186回ハイパフォーマンスコンピューティング研究発表会での中尾さんの発表を受けて
 - 従来のABCIでは、2段階のSSHでのログインのみ。慣れていない利用者には利用障壁となっていた
 - Jupyter Notebook/Lab 利用需要は多かったが、計算ノードで起動したサービスにSSHトンネルでアクセスする手段を伝えていた
- 2022年12月：Open OnDemandの構築開始
- 2023年2月：初期バージョンの配備完了
 - JupyterLab App、Job Composer導入
- 2023年4月：ABCI固有アプリ開発開始
 - 量子コンピューティング向け: Qni, CUDA-Q
 - AI向け: AI Hub (チーム内で連携したAIモデル開発ツール)
- 2024年8月：試験運用開始、全利用者に開放

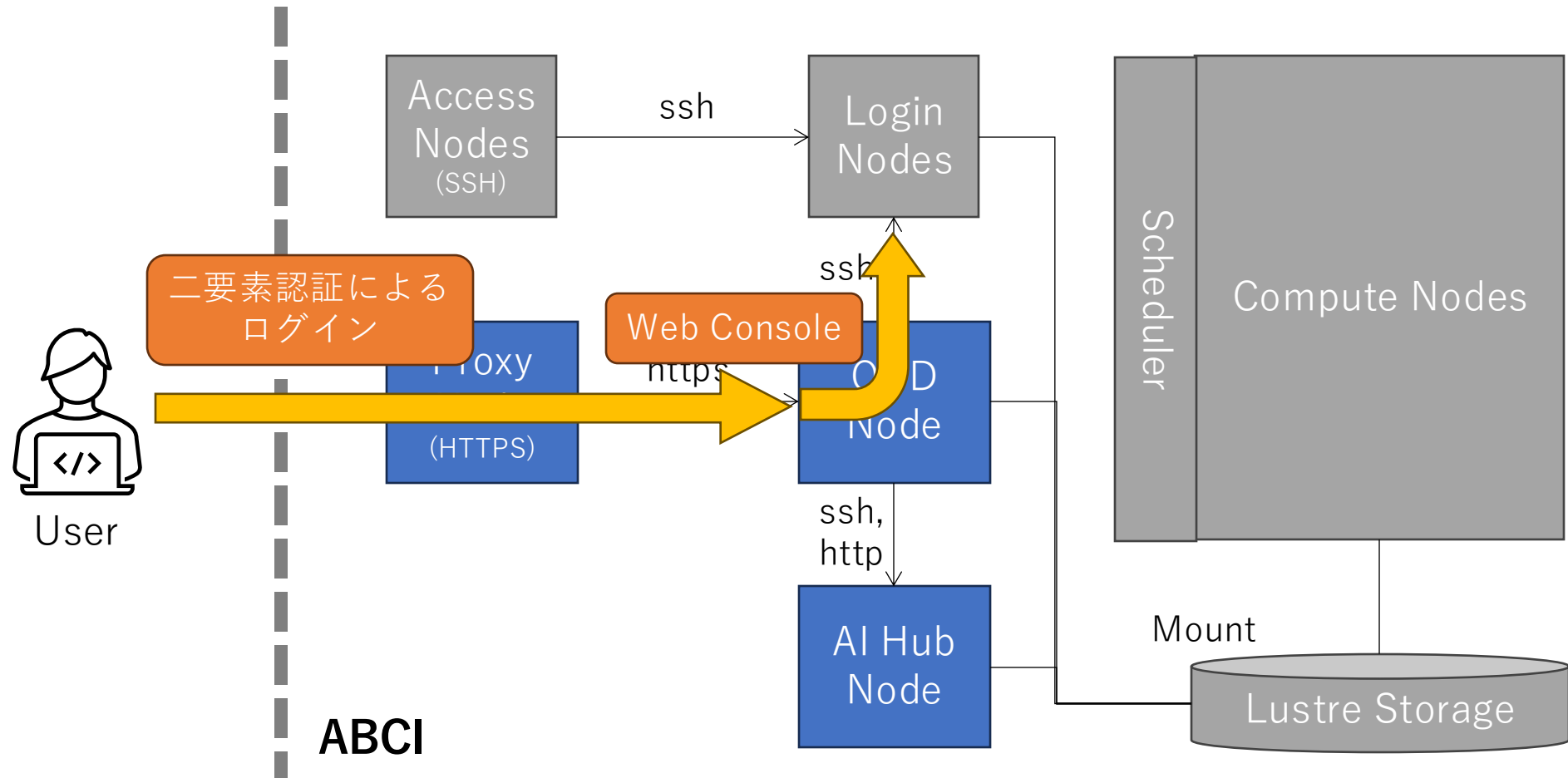
従来のABCI 2.0利用方法



ABCI 2.0におけるOpen OnDemand配備図

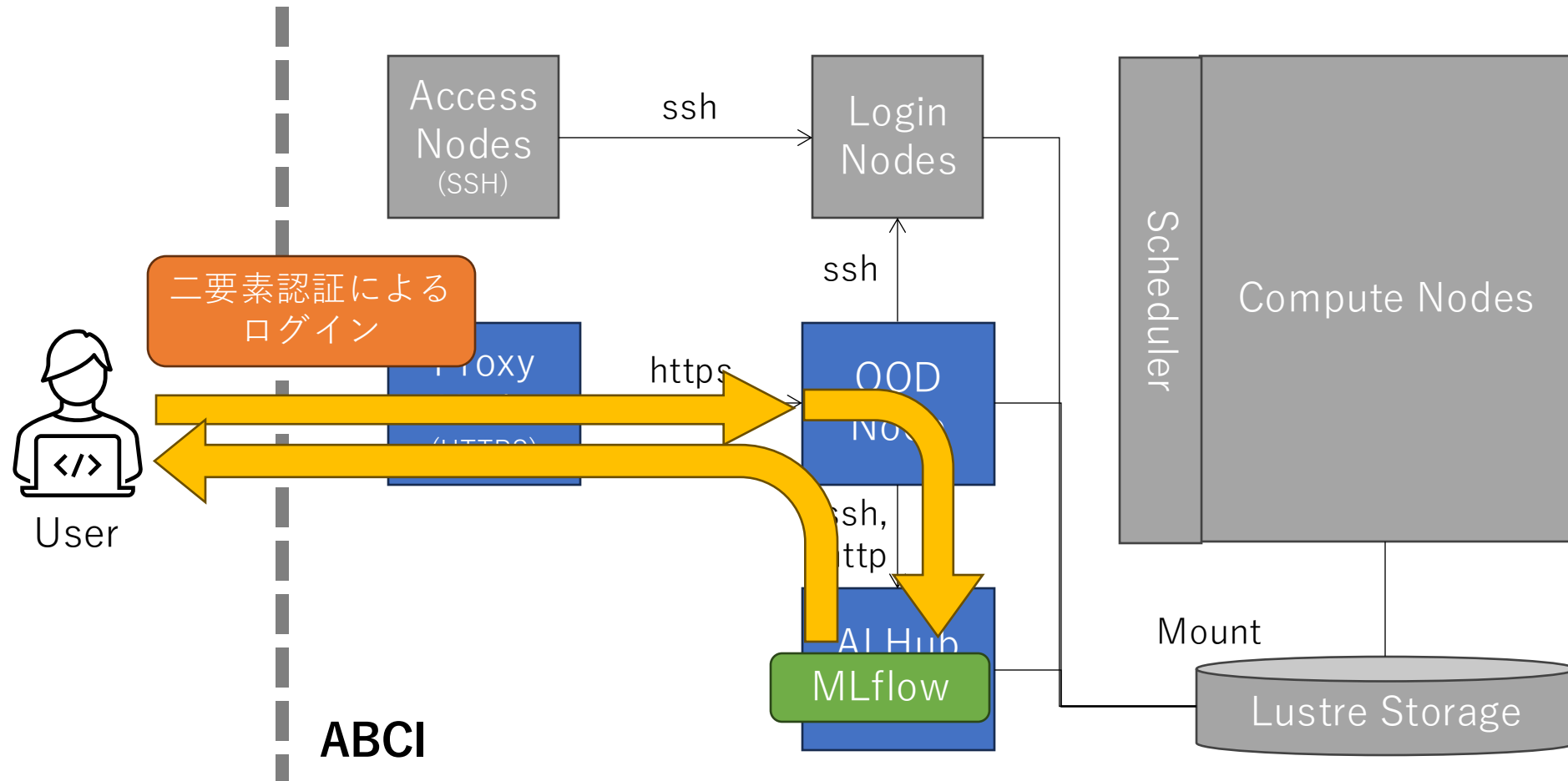


ABCI 2.0におけるOpen OnDemandによる利用方法





ABCI 2.0におけるOpen OnDemandによる利用方法



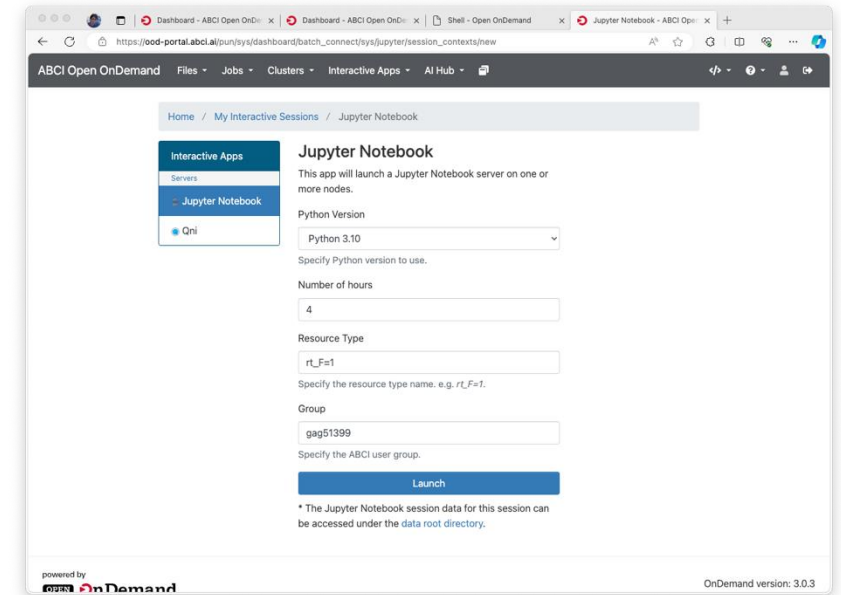
ABCI 2.0 Open OnDemandの設定

- Open OnDemand 3.0.3を導入
- 二要素認証を有効：パスワード + Email OTP
 - Keycloak設定で、TOTPへの設定変更可
 - ABCIアカウントと統合、アカウントとパスワードはldap参照
 - Emailアドレスは別DBで管理していたため、Emailアドレスは定期的にDBと同期
- セキュリティ対策
 - Keycloakとの通信は、ABCI内部でも暗号化
 - Proxy Node, OOD Nodeには改竄検知にOSS版 Tripwireを導入
 - OODにより起動されるアプリごとに認証機構を導入

ABCI 2.0 Interactive Apps

• JupyterLab

- OODのJupyter Notebookサンプルコードをベースに、JupyterLabを起動
 するようカスタマイズ
 - https://github.com/OSC/bc_example_jupyter
- アプリ起動の都度、JupyterLabを導入したPython仮想環境を構築
 - ABCIが提供する複数のPythonモジュールから1つを選択

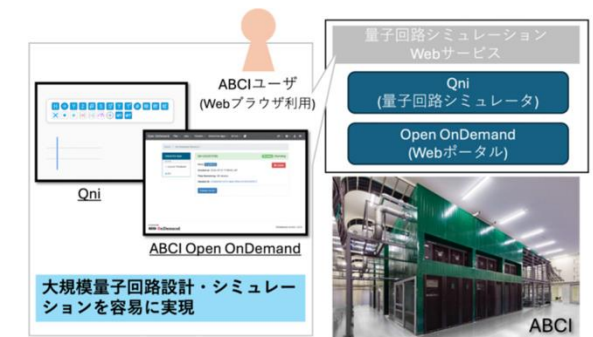
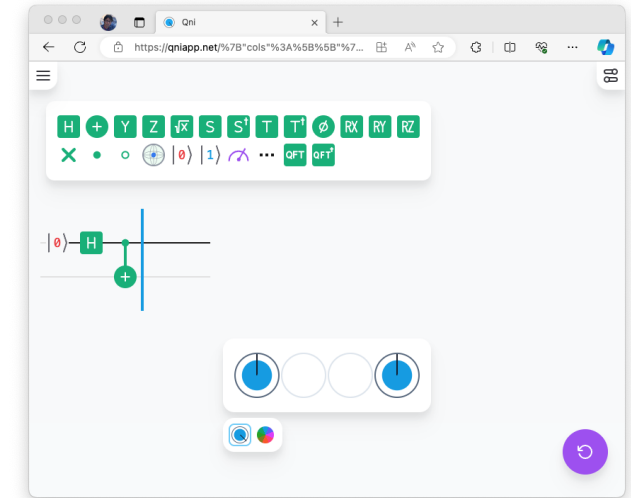


• CUDA-Q

- OODのJupyter Notebookサンプルコードをベースに、NVIDIA CUDA-Qコンテナイメージを用いたJupyterLabを起動
 - 当コンテナイメージにはCUDA-QとJupyterLabが導入されており、JupyterLab上でCUDA-Qプログラミングが可能
- 2024年9月4日開催の「量子・古典ハイブリッドコンピューティングハンズオンセミナー」にて使用

ABCI 2.0 Interactive Apps: Qni

- Webブラウザ上で動作する量子回路エディタ、兼、シミュレータ
 - 株式会社TISがオープンソースとして開発 (<https://qniapp.net/>)
- エディタのUIはそのまま、ABCI計算ノードのGPUを使ったシミュレーションを行うInteractive AppとしてTIS社と共同開発、ABCI 2.0に配備
 - 2種類のLinuxコンテナを計算ノード上で起動し、フロントエンドのWebサービスをOODに接続
 - フロントエンド: Qni UIサーバ (nginx)、BASIC認証
 - バックエンド: NVIDIA cuQuantum Applianceベースのシミュレーションサーバ(WSGI)、Qiskit Aerで1 GPUを使用して実行



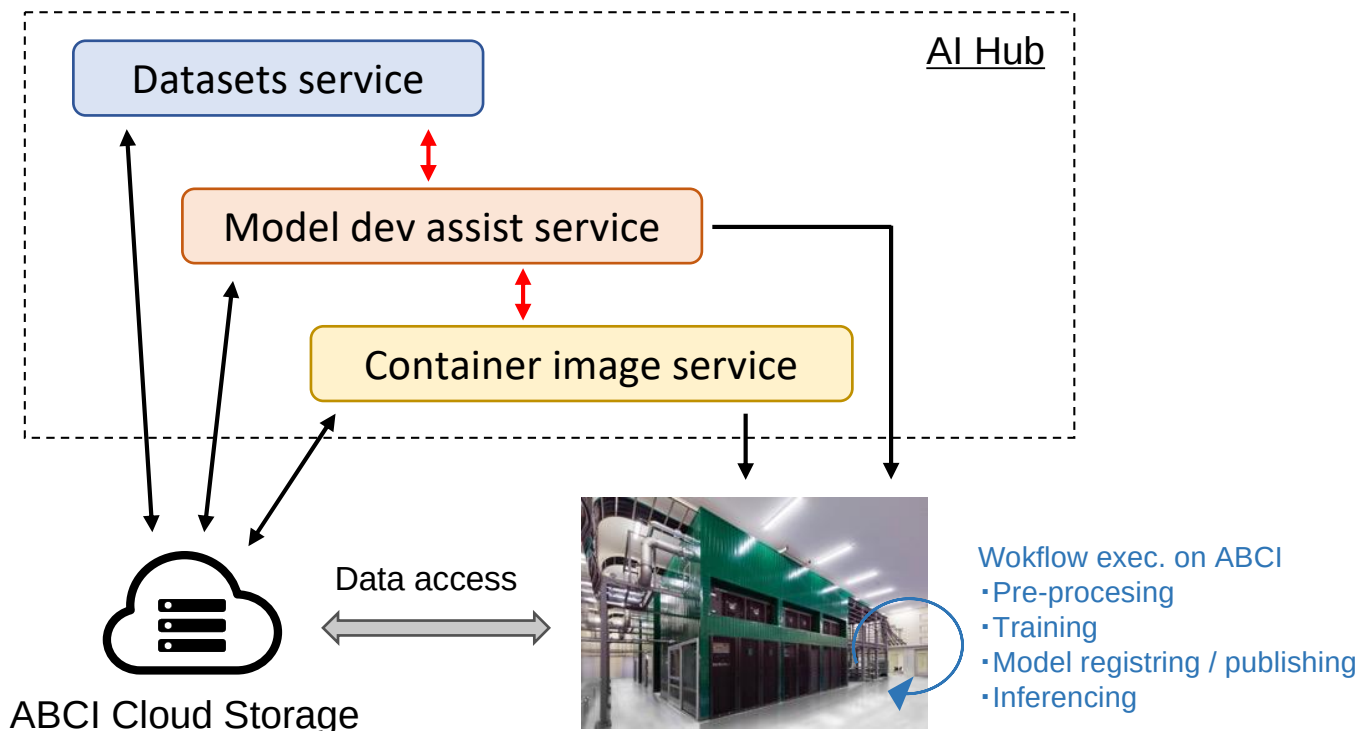
産総研 ABCI上でのQniによる量子回路シミュレーションWebサービス

AI Hub App (1/2)

AI資源リポジトリを中心に、大規模な学習済みモデルの再利用を促進することを目的に、AI研究開発コミュニティでの協力したAIモデル開発、データ・AIモデルの共有を実現するプラットフォーム

AI資源リポジトリ: AI資源を一元的に管理する場、ABCIクラウドストレージ

AI資源: 学習データ、AIモデル、コンテナイメージ、Notebook等



データセットサービス

- ABCI 利用者のデータセットの公開・共有を支援するカタログサービス

モデル構築支援サービス

MLflow Tracking Serverを実行・管理する
Open OnDemandアプリを導入

コンテナイメージサービス

- ABCI 利用者向けのコンテナイメージライブラリ・ビルダー

AI Hub App (2/2)

デモ動画

2024年9月のABCI 2.0 Open OnDemand利用状況

2024年9月分利用状況:

項目	値
総ログイン回数	246
ユニークなログイン人数	91

試験運用開始直後のため、
利用数は多くない。
イベント開催日（9/4）前後の
ログインは多い。

日毎のログイン回数:

日付	値	日付	値	日付	値
2024-09-01	2	2024-09-08	1	2024-09-24	4
2024-09-02	21	2024-09-09	6	2024-09-25	7
2024-09-03	63	2024-09-10	7	2024-09-26	7
2024-09-04	64	2024-09-11	3	2024-09-27	5
2024-09-05	15	2024-09-12	11	2024-09-28	1
2024-09-06	13	2024-09-13	7	2024-09-29	2
2024-09-07	1	2024-09-14	1	2024-09-30	5

- 9/17 ~ 9/24はメンテナンスのためデータがありません。

Open OnDemandを導入して良かった点

- ABCIへのログイン障壁が下がった
 - ハンズオン/ハッカソンイベントでのABCI初回利用者のトラブル減
 - 2段階SSHの設定につまづく人、SSHトンネルの設定がうまくいかない人が結構いた
- Web UIを持ったサービスをバッチジョブ管理でき、容易に起動できること
 - 利用者にとっての利便性拡大
 - QniやAI Hub以外にもWebベースアプリをいくつか開発中、バッチジョブ管理が容易にできるのでサービス提供の敷居が下がる

今後の展望

- ABCI 3.0、ABCI-QともにOpen OnDemandを導入、正式サービスとして公開
 - ABCI 3.0: 1/20より運用開始、OODも同時にサービス開始
 - 現在はコンソール、ファイル操作、JupyterLabのみ
 - Job Composerは公開に向けて開発中、AI Hubは数ヶ月以内を目処に提供予定
 - ABCI-Q: 2025年度上旬の運用開始と同時にOODもサービス開始
- 主たる利用者層を考慮し、アプリには差別化
 - ABCI 3.0: AI Hub等、最新のAI向けアプリの開発の場、順次サービス公開
 - ABCI-Q: 量子コンピューティング向けアプリを公開、AI向けアプリの一部を取り込む