

PoC 環境 利用案内書

(第2版)

1. 目的

本環境は、AI アプリケーションおよび AI サービスの評価・検証を目的として提供する PoC (Proof of Concept) 環境です。

利用用途や技術要件に応じて複数の利用形態を提供し、日単位で手軽に使えるワークステーション利用から、サーバーを一定期間占有して自由に構築できるサーバー利用まで、柔軟に対応します。

本環境は評価・検証用途を目的として提供されるため、商用サービスレベルの可用性や継続利用を保証するものではありません。

2. 利用形態

本環境では以下の2つの利用パターンを提供します。利用目的・期間・必要な権限に応じて適切なパターンをお選びください。いずれのパターンも root 権限のある環境を提供します。

GPU は仮想マシンへパススルーしているため、利用者ご自身のVM 内でNVIDIA Driver の更新も可能です。

項目	パターンA: ワークステーション利用	パターンB: サーバー利用
提供形態	占有型仮想マシン (予約時に自動作成)	占有型仮想マシン (node01~03 の3台セット)
権限	root 権限あり	root 権限あり
利用単位・期間	日単位	3週間程度
予約方法	予約サイトで予約→VM 自動生成→ログイン可	メール相談ベース (運営者が構築)
アクセス	現状は居室での現地利用のみ (今後 VPN 利用も許可予定)	Tailscale (VPN) のみ
事前導入	NVIDIA Driver	NVIDIA Driver
OS	Ubuntu Desktop	AlmaLinux / Ubuntu Server から選択可
GPU	ワークステーションにより異なる (RTX A1000 8GB / RTX 5000 24GB / RTX 4500 20GB。詳細は第3章)	NVIDIA L40S 48GB ×3 (node03 のみに搭載)

2.1 パターンA: ワークステーション利用

JAMBASE もしくは O-Nexus でご利用いただけます。

予約サイトで予約を行うと、仮想マシンが自動的に作成され、ログイン可能になります。利用者には root 権限が付与され、環境構築やアプリケーション実行を自由に行えます。日単位での手軽な利用を想定しています。

日貸しのため、7時から21時までの間に予約を入れると、即時に仮想マシンの作成が始まります。それ以外の時間帯に予約した場合は、翌朝7時を待って仮想マシンの作成が始まります。なお、居室の利用時間は9時から18時です。現状は居室での現地利用のみに対応しています（今後、VPN経由での利用も許可予定です）。

VMにはNVIDIA Driverがあらかじめインストールされています。既存のコンテナ資産（Dockerイメージ・Composeファイル等）を利用する場合は、Docker等のコンテナ関連パッケージを導入のうえご利用ください。

利用可能機能

- root 権限でのフル操作
- 任意アプリケーション実行
- Docker 等によるコンテナ実行
- 利用者専用ストレージ利用
- GPU 利用（ワークステーションにより異なる。詳細は第3章）

2.2 パターンB：サーバー利用

node01～03の3台をセットで、一定期間（3週間程度）占有利用いただく形態です。各ノードには占有型の仮想マシンを割り当て、利用者にはroot権限を付与します。予約はメール相談ベースとし、運営者が要件に合わせて環境を構築のうえ払い出します。

3台のノードはノードごとに搭載スペックが異なるため、割り当てられる仮想マシンのリソースもノードごとに異なります（想定構成：node01=Inject サーバ、node02=DB サーバ、node03=推論サーバ）。GPU（NVIDIA L40S）はnode03のみに搭載されており、node03の仮想マシンへパススルーされています。詳細は「3. 割り当てリソース（専有VM）」を参照してください。

本形態は、オンプレミスサーバ導入のリファレンス構成としても活用いただける3ノード構成となっています。

VMにはNVIDIA Driverがあらかじめインストールされています。Docker等のコンテナ関連パッケージは、必要に応じて導入してご利用ください。

利用可能機能

- root 権限でのフル操作
- SSH 接続 / 任意アプリケーション実行
- Docker 等によるコンテナ実行
- GPU 利用（node03、NVIDIA L40S）
- 3ノード間の内部通信（Inject⇄DB⇄推論のトポロジ）
- 利用者専用ストレージ利用

ネットワーク構成

- 内部通信：3台のVMは同一の内部ネットワークで接続され、Inject⇄DB⇄推論の内部通信が行えます。
- リモートアクセス：Tailscale（VPN）経由でアクセスします。auth keyは貸出ごとに発行します。

3. 割り当てリソース（専有 VM）

各ノードに割り当てる専有 VM のリソースは以下のとおりです。利用環境のハードウェアごとに異なります。

3.1 ワークステーション（パターン A）

ノード	拠点	vCPU	メモリ	ストレージ	GPU
Jambase01	JAMBASE	8	24GB	ルート 500GB	RTX A1000 8GB
Jambase02	JAMBASE	6	24GB	ルート 175GB ／/data 1TB	NVIDIA RTX 5000 24GB
Jambase03	JAMBASE	6	24GB	ルート 650GB ／/data 1TB	NVIDIA RTX 5000 24GB
Onx01	O-Nexus	8	24GB	ルート 500GB	RTX A1000 8GB
Onx02	O-Nexus	6	24GB	ルート 175GB ／/data 1TB	NVIDIA RTX 4500 20GB
Onx03	O-Nexus	6	24GB	ルート 650GB ／/data 1TB	NVIDIA RTX 4500 20GB

3.2 サーバー（パターン B）

ノード	役割	vCPU	メモリ	ストレージ	GPU
node01	Inject サーバ	16	240GB	ルート 300GB ／/data 15TB	なし
node02	DB サーバ	16	240GB	ルート 300GB ／/data 15TB	なし
node03	推論サーバ	32	500GB	ルート 10TB（/data 含む）	NVIDIA L40S 48GB ×3

※ 上記は各ノードに割り当てる専有 VM の目安です。確定値は貸出時の構成に合わせて記載します。

4. ストレージ利用方針

OS 領域とデータ領域の分離

OS ブート領域と data 領域を分離して提供します。領域を割り当てる場所のハードウェアにより性能が異なり、基本的にはルートディレクトリが高速、/data ディレクトリは I/O 性能が低くなります。

5. 更新・サポートの責任分界

各ノードの VM は利用者の専有であり、CUDA・Docker の導入、NVIDIA ドライバの更新は利用者ご自身の VM 内で自由に実施いただけます（他利用者や基盤へは影響しません）。

- 利用者領域：VM 内の CUDA / Docker / NVIDIA ドライバ / アプリケーション構成
- 運営者領域：物理サーバ・ハイパーバイザ・GPU パススルー基盤・初期イメージ（利用者は変更不可）

更新に失敗するなどして環境が復旧不能となった場合は、初期イメージへのリセット（再払い出し）で対応します。個別のソフトウェアスタックの詳細なトラブルシューティングは、PoC 環境の性質上ベストエフォートでの対応となります（可用性・継続利用は保証しません／第 6 章参照）。

6. データ管理・利用上の注意事項（免責）

本章は現行案内書から引き継ぐ免責・共通事項です。

利用者が持ち込むデータおよび作成したデータの管理は利用者責任とします。利用終了時には利用者ご自身で不要データの削除を実施してください。運営者は利用終了後に利用者領域の確認およびクリーンアップ（VM 破棄・領域の消去）を実施する場合があります。

本 PoC 環境では利用者データのバックアップは実施しません。必要なデータは利用者ご自身でバックアップを取得してください。

本環境は PoC（評価・検証）用途として提供されます。以下については保証を行いません。

- 可用性
- 性能
- 継続利用
- 利用者データ保全

セキュリティおよびデータ保全是、基本的に利用者ご自身の責任において実施いただくものとし、利用者は自己の責任において本環境を利用するものとします。

7. 事前確認事項

利用申請時（パターン B はメール相談時）に以下の内容をご連絡ください。

確認項目	内容・記載例
利用パターン	ワークステーション利用 / サーバー利用 のいずれか
利用期間	日単位（パターン A） / 3 週間程度（パターン B）
OS 選択	パターン A: Ubuntu Desktop / パターン B: AlmaLinux・Ubuntu Server
GPU 利用有無・枚数	使用する（L40S ○枚 等） / 使用しない
ストレージ利用要件	必要容量の目安、保存データの種別 等
特殊要件の有無と詳細	特定 CUDA バージョン、カスタム Driver 等 ある場合は詳細記載