



≪みなとみらい駅直結≫ インフラプラットフォームエンジニア | AI課題を解決するエンジニア | クラウド・AI開発経験歓迎

Job Information

Hiring Company

Unsung Fields Corp.

Job ID

1573182

Industry

Software

Company Type

Small/Medium Company (300 employees or less)

Job Type

Permanent Full-time

Location

Kanagawa Prefecture, Yokohama-shi Nishi-ku

Train Description

Minatomirai Line, Minatomirai Station

Salary

8 million yen ~ 14 million yen

Refreshed

February 5th, 2026 13:00

General Requirements

Minimum Experience Level

Over 6 years

Career Level

Mid Career

Minimum English Level

Daily Conversation

Minimum Japanese Level

Native

Minimum Education Level

Bachelor's Degree

Visa Status

No permission to work in Japan required

Job Description

インフラプラットフォームエンジニア

【職務概要】

インフラストラクチャエンジニアとして、本番環境の推論を実行するGPUプラットフォームを担当します。クラスターアーキテクチャ、デプロイの信頼性、オブザーバビリティ、キャパシティ管理、インシデント対応メカニズムまでを含め、ハードウェア、モデル、テナント、トラフィックパターンのスケールアップ時でも、プラットフォームを予測可能かつ高信頼に保つことがミッションです。

サービング/ランタイムチームやゲートウェイチームと密接に連携し、適切なアイソレーションが強制され、必要なテレメ

トリが提供され、ダウンタイムなしで安全な変更が可能となるようプラットフォームを設計・運用します。本ポジションは、強いシステム直感と実運用に根ざした規律を併せ持ち、信頼性の高いロールアウト、整備された運用ツール、迅速なインシデント対応を実現する役割です。

【業務内容】

- GPU クラスタのアーキテクチャおよび運用を担当：プロビジョニング、ノードイメージ管理、ドライバ／ランタイムのライフサイクル管理、GPU プラグイン／オペレーターのライフサイクル、サービングプールおよびシステムサービス向けの標準化されたデプロイパターン。
- 本番環境のベースラインを定義・維持：ゴールデンノード構成、クラスタ・ハードニング、アップグレードパス、「既知の良好な」互換性マトリクス（ドライバ ↔ CUDA ↔ ランタイム ↔ カーネル）。
- プラットフォームに信頼性を組み込む：SLO／SLI、アラート品質、運用手順書、インシデント対応ツール、そして再発防止までを含む実効性のあるポストモーテム（自動化、ガードレール整備、再発要因の排除）。
- 安全なデリバリーを実現：カナリアデプロイ、段階的ロールアウト、ロールバック経路、設定の安全性（検証、ガードレール、変更管理、セーフ・デフォルト）。
- フリート全体の健全性および保守ワークフローの管理を担当：ノードドレイン、GPU の隔離、自動修復、計画メンテナンス、監査可能な「ブレイクグラス」手順。
- キャパシティおよび利用効率の管理：スケジューリング制約、ピンパッキング／断片化管理、ウォームプール、オートスケーリングのプリミティブ、プロダクトのティアや公平性目標と整合したクォータ適用フック。
- オペザバビリティの確立：ゲートウェイ → サービング → GPU にまたがるメトリクス／ログ／トレーシング、レイテンシ内訳、飽和シグナル、キュー深度、GPU メモリ／計算メトリクス、顧客の症状と根本原因を結び付けるフリート健全性ダッシュボード。
- 異種混在環境における本番運用に耐える状態の確保：ハードウェア世代や進化するサーバープラットフォーム間の差異を管理し、信頼性リスクを最小化しながら利用効率を向上。
- セキュリティベースラインの確立：シークレット管理、最小権限アクセス、オペレーター操作の監査証跡、安全な運用ワークフロー。
- ネットワークチームとの連携：トポロジー、障害ドメイン、ロードバランシング、テールレイテンシや可用性に影響する性能重視のトラフィック経路設計。
- 運用ツールの構築：フリート管理、デバッグワークフロー、安全な管理操作、キャパシティ管理ツール、MTTR を削減しオペレーター効率を高めるメンテナンス自動化。
- チーム横断での協業：ロールアウト計画、ヘルスセマンティクス、キャパシティシグナル、障害対応方針を整合させ、全体プラットフォームが高負荷下でも予測可能に振る舞うようにする。

【雇用形態】

正社員

※試用期間あり、3ヶ月

【給与】

年収：800万円～1,400万円

月収：66.6万円～116.6万円（月額基本給：66.6万円～116.6万円）

■昇給：あり

【就業時間】

09:00～18:00（休憩時間 60分）

【勤務地】

〒220-6010 神奈川県横浜西区みなとみらい2丁目3番1号 クイーンズタワーA 10階

■アクセス：各線 桜木町 駅から徒歩7分、東横線・みなとみらい線 みなとみらい 駅から直結

■就業場所全面禁煙

■勤務地変更範囲：会社の定める事業所

■転勤・出向：無し

【休日休暇】

- 年間休日 120 日
- 完全週休二日制
- 年間有給休暇（入社7ヶ月目には最低10日以上）

【待遇・福利厚生】

- 交通費 一部支給（上限月1万5千円）
- 社会保険（健康保険、厚生年金、雇用保険、労災保険）
- 残業手当：通常の残業代

Required Skills

【必須要件】

- 本番分散システムを対象としたインフラ／SRE／プラットフォームエンジニアリングの経験5年以上。
- 本番環境におけるKubernetes（または同等のオーケストレーション基盤）の豊富な運用経験と、実運用のオーナーシップ。
- GPU クラスタやその他の高性能コンピューティング基盤（または同等に敏感なインフラ）の運用経験。
- Linux、ネットワーク、分散システムの障害モードにまたがる高いデバッグ能力。
- 高い運用規律：自動化ファーストの思考、定量的に測定可能な信頼性、慎重な変更管理、インシデント時の明確なコミュニケーション。
- 担当システムに対するオンコールローテーションへの参加意欲。

【歓迎要件】

- 高スループットなゲートウェイ／サービスメッシュの利用・運用経験（例：Envoy）、ならびに OpenTelemetry やマルチリージョンアーキテクチャの経験。
- SlurmなどのHPC系スケジューリング、RDMA／InfiniBand、または性能要件に敏感なネットワーキングに関する経験。
- 社内向けデベロッパープラットフォームの構築経験および、一貫したデプロイ／ロールバックを実現する「ゴールデンパス」の整備経験。
- GPUドライバ／ランタイムのアップグレードをフリート全体で安全に管理した経験（互換性テスト＋段階的ロールアウト）。
- レイテンシに敏感なシステム向けのオブザーバビリティ設計パターンへの理解（リクエスト相関、サンプリング戦略、高カーディナリティメトリクスの制御）。

Company Description