

社内文書検索AIアシスタント

電通国際情報サービス (ISiD)

X (クロス) イノベーション本部 AIテクノロジー部

AIトランスフォーメーションセンター

AITC

AI TRANSFORMATION
CENTER トランスフォーメーションセンター

- シナリオ概要
- アプリUI
- アーキテクチャ
- 考慮事項

01

シナリオ概要

- 社内に膨大なドキュメントが散在し、「社内ルールが分からない」「社内システムの申請方法が分からない」といった課題をお持ちではないでしょうか？
- このシステムでは、AIアシスタントがユーザーからの質問をもとに社内ドキュメントを検索し、その内容を元に回答する仕組みを提供します。
- また、ユーザーはドキュメントのうち特にAIが参照した箇所を確認したり、あるいはドキュメントそのものをダウンロードすることもできます。
- 以下のような状況で役立つシステムを作成することができます
 - ▶ 社内に多数のドキュメントがあり、どれを参照すればいいかが分からない。
 - ▶ ドキュメントの分量が非常に多く、読みこなすのに時間がかかる
 - ▶ 様々なドキュメントを参照に業務を行う業種・業態において利用可能

02

デモUI

Chrome File Edit View History Bookmarks Profiles Tab Window Help 15.04GB Thu Sep 7 18:46

Know Narrator Search demo × 育児・介護休業等に関する規則の... × モデル就業規則について | 厚生労働省 × +

QA履歴 設定

子供が何歳になるまで育児... ⓘ

子供が何歳になるまで育児... ⓘ

■ 良い質問の仕方
言葉の曖昧性を除き、検索の意図が伝わるようにしましょう。
×: 有休 取得日数
○: 勤続1年目の社員が取得可能な有休の日数を教えてください。

質問テンプレート ⓘ

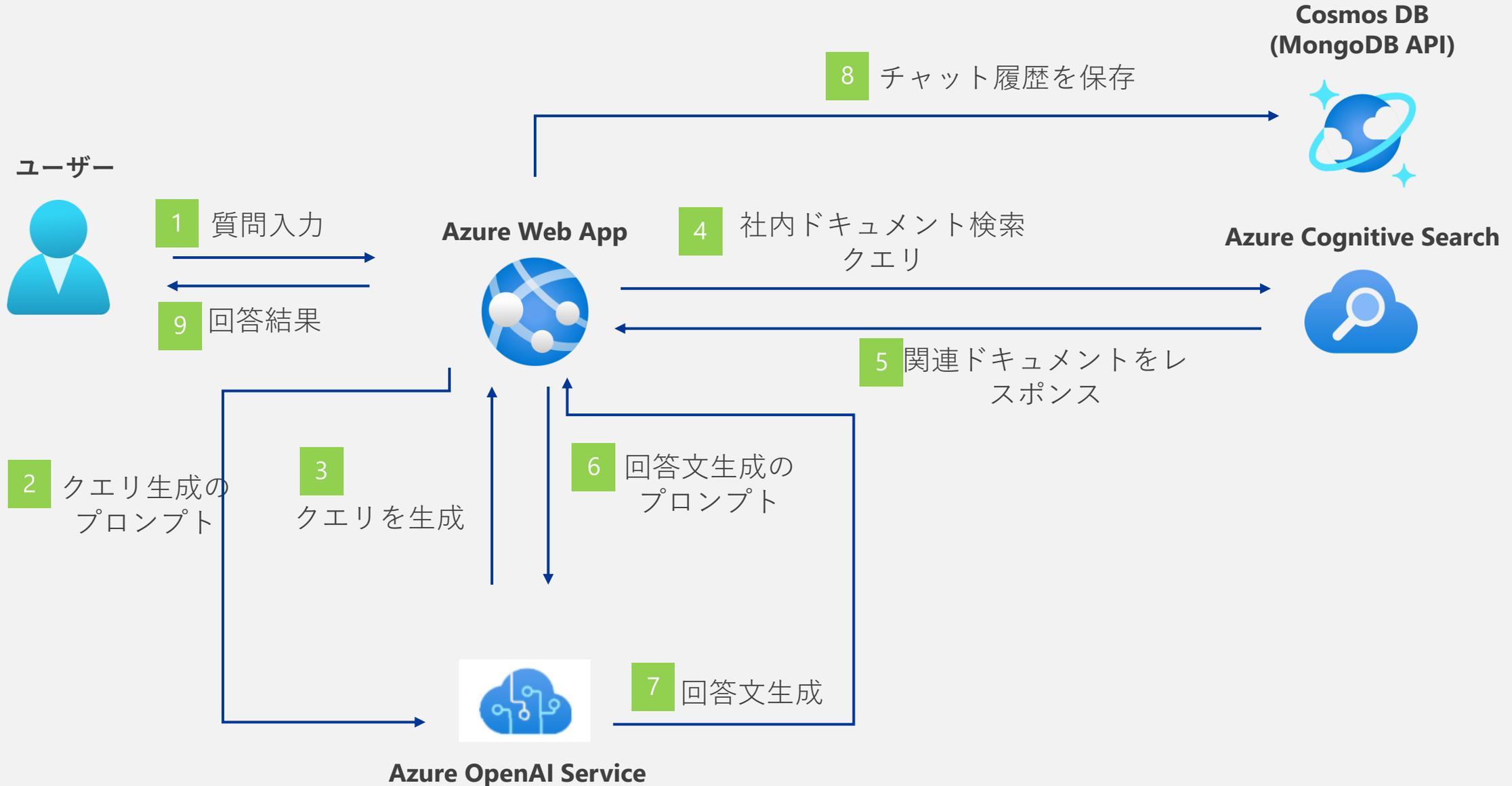
質問を入力してください。 送信

バージョン情報: sch.v23.07.24

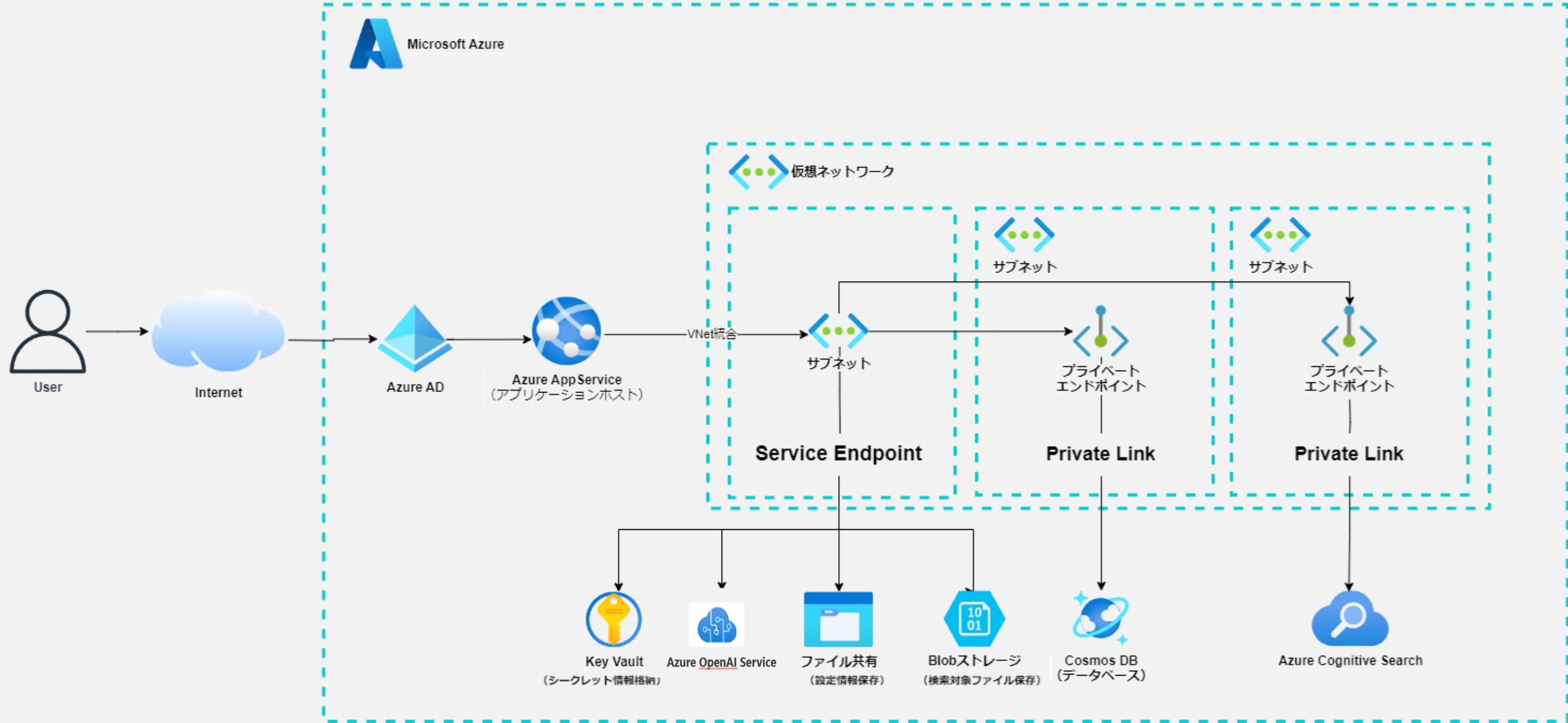
03

アーキテクチャ

ワークフロー



アーキテクチャ詳細



04

考慮事項

信頼性(可用性)

- Azureのサービスでは「可用性ゾーン」や「リージョン」といった単位で可用性を設計しており、これらを適切に組み合わせることで、ビジネスクリティカルなワークロードの信頼性を実現するように設計することが可能です。詳細は「[Azureリージョンと可用性ゾーンとは](#)」をご参照ください。
- このシナリオで用いられているApp Service、Azure OpenAI Service等のコンポーネントはそれぞれゾーン冗長、リージョン冗長、geoレプリケーションなど高可用性のオプションや構成を利用可能です。必要となる可用性に応じて導入を検討してください。複数リージョン間/複数ゾーン間でAct-Act構成を取る場合にはAzure Front Door、Azure Application Gatewayなどの利用をご推奨します。

信頼性(回復性)

- アプリケーションの正常性を監視するために、Application Insightsを使用すると、カスタマーエクスペリエンスや可用性に影響を及ぼすパフォーマンスの問題についてアラートを生成し、対応することができます。詳細については、「[Application Insightsの概要](#)」を参照してください。
- 回復性に関するその他の記事については、「[信頼性の高い Azure アプリケーションを設計する](#)」を参照してください。

セキュリティ

- セキュリティは、重要なデータやシステムの意図的な攻撃や悪用に対する保証を提供します。詳細については、「[セキュリティの重要な要素の概要](#)」を参照してください。セキュリティで保護されたソリューションの設計に関する一般的なガイダンスについては、「[Azureのセキュリティのドキュメント](#)」を参照してください。

コスト最適化

- 不要な費用を削減し、運用効率を向上させる方法を検討することです。詳しくは、[「コスト最適化の柱の概要」](#)に関する記事をご覧ください。

オペレーショナルエクセレンス

- システムの健全性の担保、トラブルの解決、利用動向の監視を行うためには適切な監視とログ収集が必要となります。
詳細は「[ワークロードの監視](#)」をご参照ください。API Managementを利用することで、API利用の監視やトレースを行うことが容易になります。
- ソフトウェアのアップデートや脆弱性への対応などソフトウェア/インフラ設計の改修を円滑で安全に進められるよう、DevOpsプロセスの確立について「[リリース エンジニアリングの継続的インテグレーション](#)」をご参照ください

パフォーマンス効率

- アプリケーションの負荷が高まることを見越し、スケーラビリティの確保をあらかじめ検討することは重要です。詳細は「[スケーリング用のアプリケーションを設計する](#)」をご参照ください
- App Serviceは負荷に応じて水平にスケールさせることが可能です。詳細については「[Azure App Service での自動スケーリング](#)」をご参照ください。
- 頻出のクエリについてはアプリ側でキャッシュする等のキャッシュ戦略もご検討ください。詳細は「[キャッシュを使用する](#)」をご参照ください。
- 特定のユーザーにAzure OpenAIの利用が集中することを避けたい場合にはAPI Managementによるスロットリング導入などをご検討ください。詳細は「[Azure API Management を使用した高度な要求スロットル](#)」をご参照ください。