Session Report

Cloud Run の本番運用に向けて考慮すべきポイントと
Cloud Ops のロギング、モニタリングの上手な活用を学ぶ

Cloud Ops で踏み出す
Cloud Run 本番運用への第一歩

Google Cloud
カスタマー エンジニア　岩成 祐樹

Google Cloud

Google Cloud

セッション レポート概要

マネージド コンピューティング プラットフォームである Cloud Run を利用することで、アプリケーションのオートスケールやログ、メトリクスの収集を任せることができますが、ロギングやモニタリングなど、本番運用に向けて検討すべき事項も少しだけあります。Cloud Run を使ったサービス運用に向けて、まず検討しておきたいポイント、Cloud Operations（Cloud Ops）の上手な活用方法について紹介します。

プレゼンター紹介

Google Cloud
カスタマー エンジニア　岩成 祐樹

Google Cloud の導入支援やセミナー等の登壇を通して Google Cloud を日本のお客様に広めるための活動を行なっている。また、Jenkins に関するコミュニティ活動などにも取り組んでいる。

目次

コンテナ化したアプリをサーバーレスで利用できる Cloud Run

Google Cloud では、Cloud Functions、App Engine、および Cloud Run の 3 つのマネージド コンピューティング プラットフォームを提供しています。Cloud Run は、Knative をベースに、コンテナ化したアプリケーションをサーバーレスで利用することができます。Cloud Run の特長は、大きく以下の 3 つです。

（1）高速なデプロイ
ステートレスなコンテナを、ゼロから n まで高速にスケールし、数秒でデプロイして URL を付与したり、HTTPS の対応やバージョニング、トラフィック スピリッティングをしたりなど、さまざまな機能がビルトインで提供されています。

（2）サーバーレス ネイティブ
VM などのインフラを気にすることなく、スケールやセキュリティも考える必要はありません。言語やライブラリの制約もなく、コンテナを用意して動かすだけなので、ロックインを避けることもできます。インスタンスの起動時間ではなく、インスタンスがリクエストを処理した時間の費用で利用できます。

（3）高いポータビリティ
Knative の API の一貫性を提供しているので、コンテナのレイヤでも、Knative のレイヤでも高いポータビリティを提供します。

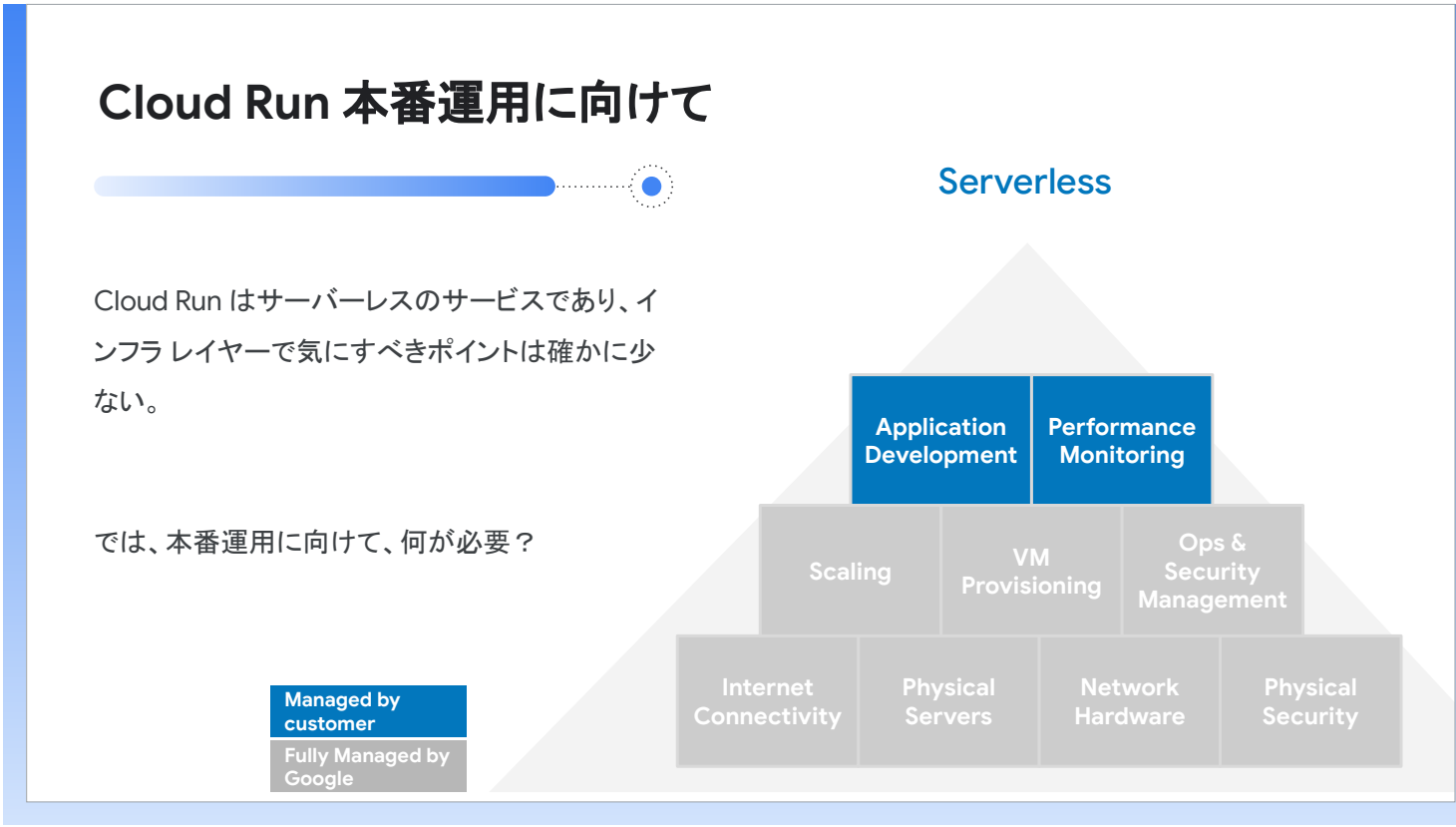
コンテナを利用するとき、Kubernetes か Cloud Run かという比較になりますが、Kubernetes はフレキシビリティが高い一方で仕様や使い方を理解する必要があります。Cloud Run は、コンテナを非常にシンプルに扱えるので、コンテナを用意してデプロイするだけで、オートスケールでシンプルに利用できます。

また Cloud Run は、常に機能がアップデートされています。サービスがリリースされた当初は、限定的な要件のみでコンテナを動かす仕組みでしたが、プロトコルやインスタンスのサイズなど、多くのワークロードに対応できるアップデートが繰り返されています。

Cloud Run では、今後も数多くのアップデートが予定されている

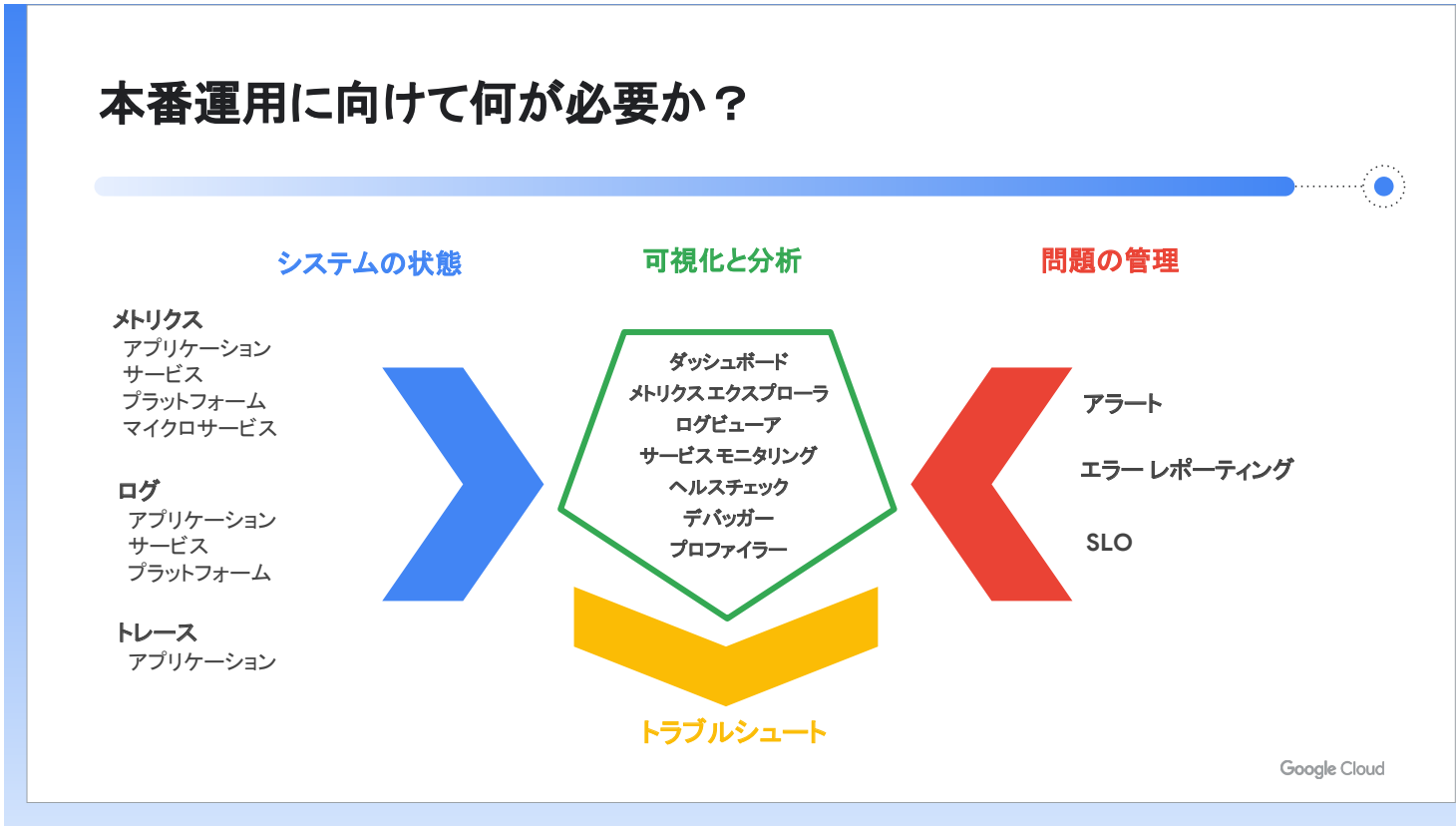
Cloud Run の本番運用で考慮すべきいくつかのポイント

本番運用において Cloud Run は、NoOps だと考えている人もいます。「Yes」と言いたいところですが、いくつかの考慮すべきポイントがあります。たとえば、サーバーレスのプロダクトを利用するとき、どの部分をプラットフォーム側が管理して、どの部分をユーザー側が管理するかですが、アプリケーション開発と、それが適切に動いているか、パフォーマンスやエラーなどに問題はないかなどのモニタリングは、ユーザーが設計しておく必要があります。



サーバーレスでは、何をプラットフォーム側が管理し、何をユーザー側が管理するか

本番環境に向け、何が必要でしょうか。まずアプリケーションが問題なく動いていることを把握するためには、メトリクスやログ、トレースの管理が必要です。また、問題が発生したときには、アラートやエラー レポーティング、サービスレベル目標（SLO）の管理が必要です。さらに、システムの状態や問題の管理を、可視化 / 分析するためのダッシュボードやエクスプローラなどが必要です。



システムの運用には、システムの状態把握と問題の管理、可視化と分析が必要

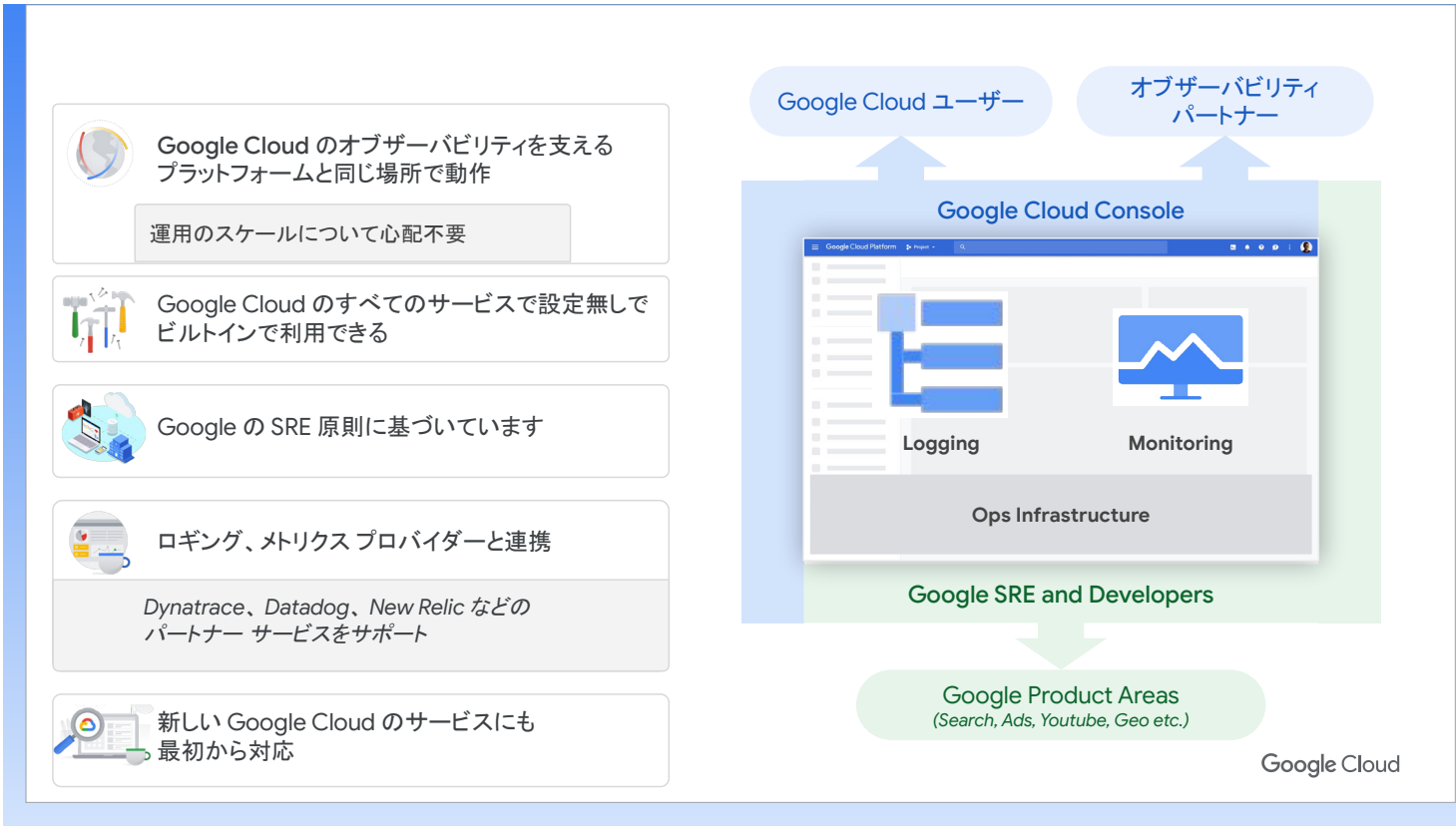
そのためのプロダクトが、Google Cloud のオペレーション スイート（Cloud Ops）です。Cloud Ops は、システム運用に必要な機能が統合された製品群で、2020 年に Stackdriver から Cloud Ops に名称が変更されました。リアルタイムのログ管理 / 分析の Cloud Logging、およびフルスタック モニタリングの Cloud Monitoring で構成されています。

システムのログを管理し、何かあったときに分析するための機能が Cloud Logging です。一方、システムの状況を常に把握しておくための機能が Cloud Monitoring です。2 つの機能を有効活用することが、安定した本番運用のための重要なポイントになります。

運用管理のために大事なのは、ロギングとモニタリングの 2 つ

Cloud Logging や Cloud Monitoring のような仕組みを独自に実装する場合、基盤のキャパシティや監視から考えることが必要です。Cloud Ops は、Google Cloud と同じプラットフォームで動いており、運用のスケールに合わせてプラットフォームもスケールします。Google Cloud にインテグレーションされているので、すべてのサービスで設定なしにすぐに利用できます。

Cloud Run をデプロイすると、すぐに Cloud Logging によりログ情報が収集され、Cloud Monitoring のためのメトリクスが収集されます。Cloud Ops 自体は、Google の Site Reliability Engineering（SRE）の原則に基づいて提供されており、SLO を簡単にチェックする機能などのベスト プラクティスがプロダクトに生かされています。もちろん、サードパーティの製品との連携もできます。



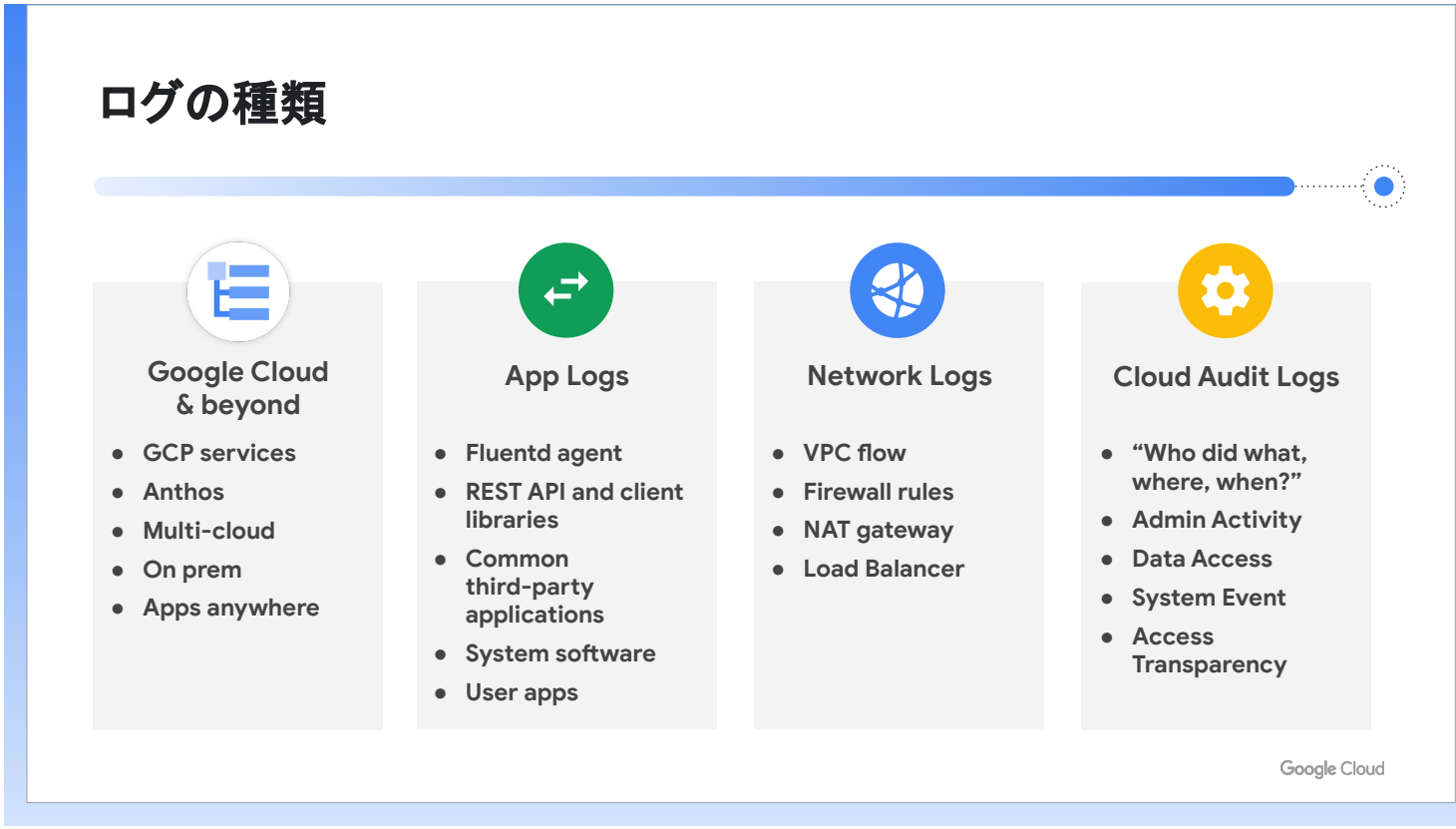
ログやモニタリングの仕組みを独自に実装する場合、基盤のキャパシティや監視を考えることが必要

リアルタイムのログ管理 / 分析を可能にする Cloud Logging

開発や運用において、さまざまなシチュエーションで必要になるのがログです。監査対応、分析用途で、さまざまなログをアーカイブしたい、開発時 / 運用時におけるデバッグで利用したい、ログの内容をもとにアラートを通知したいなどです。こうした機能を提供するのが Cloud Logging です。Cloud Logging を利用することで、ログの収集、ほかのサービスとの連携、分析、保存が簡単にできます。

ログの収集、分析、連携、保存を簡単にする Cloud Logging

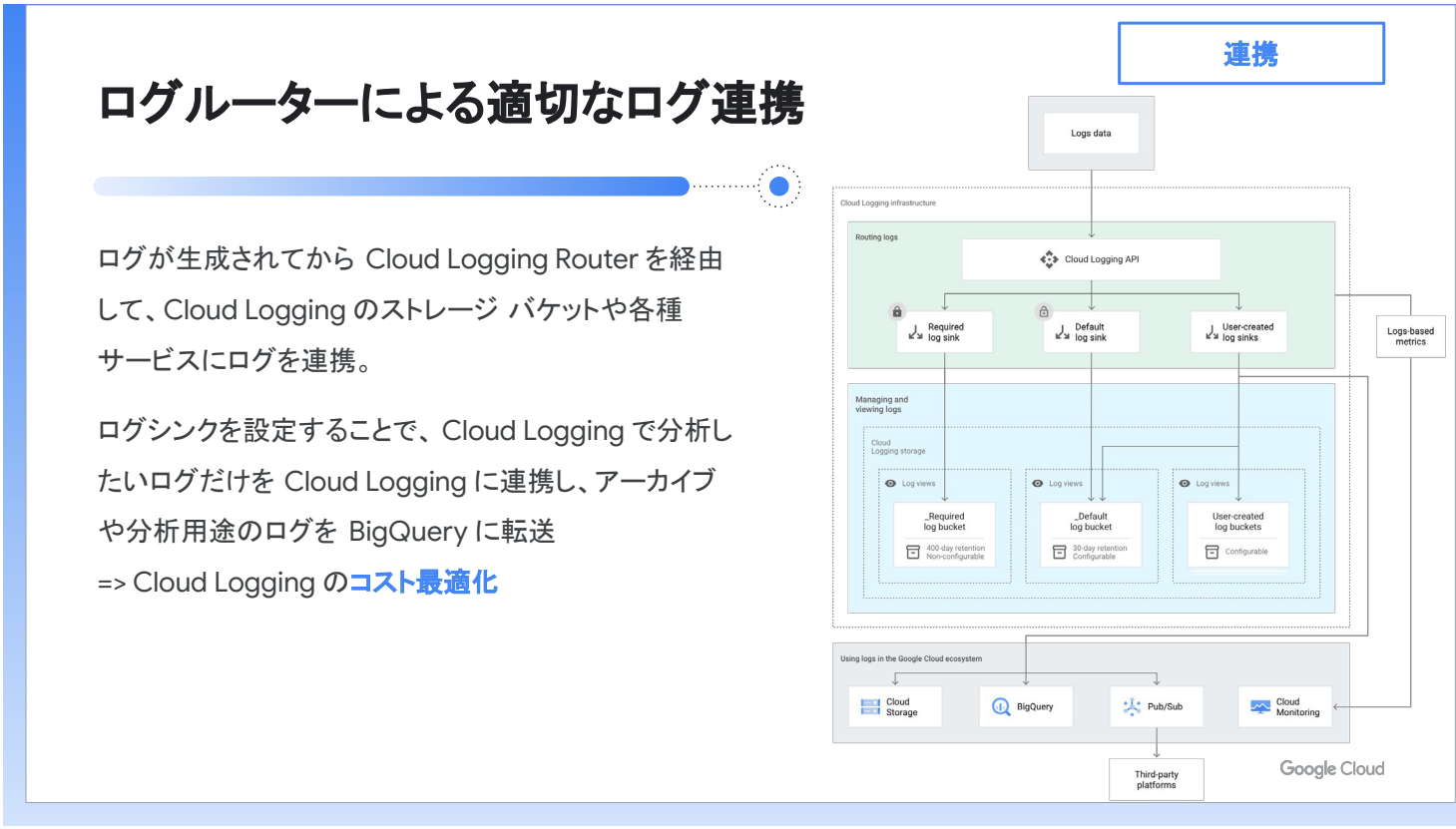
Google Cloud のログ、アプリケーションのログ、ネットワークのログ、クラウドの監査ログの 4 つを収集します。



Cloud Logging では 4 つのログを収集

Cloud Run を利用する場合、Cloud Logging はコンテナログとリクエストログの 2 つをデフォルトで収集します。コンテナログを収集する場合、構造化ログにすることで、検索しやすいログを出力できます。Cloud Logging では、いくつかの特殊フィールドが定義されており、jsonPayload から削除され、対応フィールドに書き込まれます。コンテナログとリクエストログの関連づけも意識しておくことも必要です。関連づけをしておかないと、問題が発生したときに、どのリクエストで問題が発生したのかをデバッグするのが困難になります。

Cloud Logging では、ログデータが生成されたタイミングで、ロギング API にログが送信されます。ロギング API に送信されたログは、シンクの設定により適切な場所にログが保存されます。たとえば不要なログを Cloud Logging のストレージに送信しないとか、アーカイブ用途や分析用途のログは BigQuery や Cloud Storage に送信するといった設定が可能です。適切な設定により、Cloud Logging の利用コストを最適化できます。ログバケットに保持期間を適切に設定することもポイントの 1 つ。デフォルトの 30 日では足りない場合、必要な機能を設定することができます。



ログシンクの適切な設定により Cloud Logging の利用コストを最適化できる

監査対応で特定の組織に関連するログを改ざんされない形式で保存する必要がある場合、新規のプロジェクトが発生するたびに、転送の設定をするのは手間がかかり、オペレーション ミスのリスクもあります。集約シンクと呼ばれる機能を利用することで、特定の組織に関連するログをすべて決められたストレージに集約することができます。

Cloud Run の Console で、特定の Cloud Run サービスに関連した各種リソース状況やログを参照できます。あらかじめフィルタリングされた状態で、ログ エクスプローラに遷移することも可能。ログ エクスプローラを使用したサンプルクエリも提供されているので、必要に応じて利用できます。

ログ エクスプローラを使用したサンプルクエリを利用できる

フィールド レベルのアクセス制御も可能。これまでは、たとえば開発チームのみ詳細情報を表示して、運用チームには限定情報のみ表示するといった制御はできませんでした。エラーレポートでは、エラー情報を自動的に集約し、確認することができます。エラーレポートからログに遷移することもできるので、効率的な分析が可能です。

ログベースのアラート機能もプレビューで提供されています。これまでログの内容でアラートを通知する場合、ログベースのメトリクスを定義し、しきい値を設定して、アラートを通知することが必要でした。この機能がアップデートされ、Cloud Logging からダイレクトにアラート通知を設定できるようになっています。

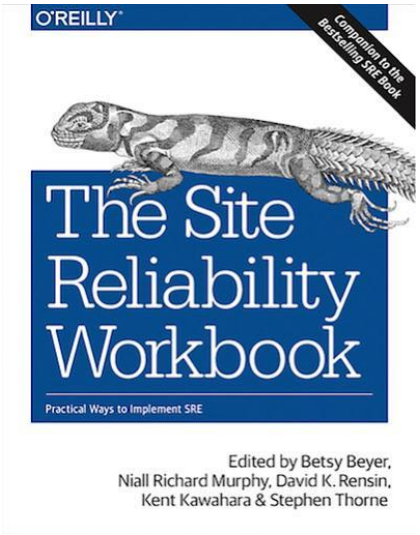
実際の開発では、ローカルでログを見る、デバッグをするというケースがあります。CLOUD ONAIR ソリューション チャンネル App Modernization OnAir では、Cloud Code を使ったローカル デバッグが紹介されていますので、ぜひご視聴ください。

フルスタックのモニタリングを提供する Cloud Monitoring

運用において、正しくサービスが稼働しているか、提供したい品質のサービスを提供できているか、近い将来に問題が発生する兆候がないかなどのモニタリングは必須です。モニタリングを考えるとき、『SRE Reliability book』という書籍で、ゴールデン シグナルという 4 つの指標について語られています。

注目すべき 4 つの指標 - ゴールデン シグナル

- レイテンシ
  - サービスがリクエストの処理にかける時間
- トラフィック
  - サービスに対する要求の量
- エラー
  - サービスが失敗する割合
- 飽和度
  - サービスのリソースがフル使用にどれだけ近いかを示す尺度



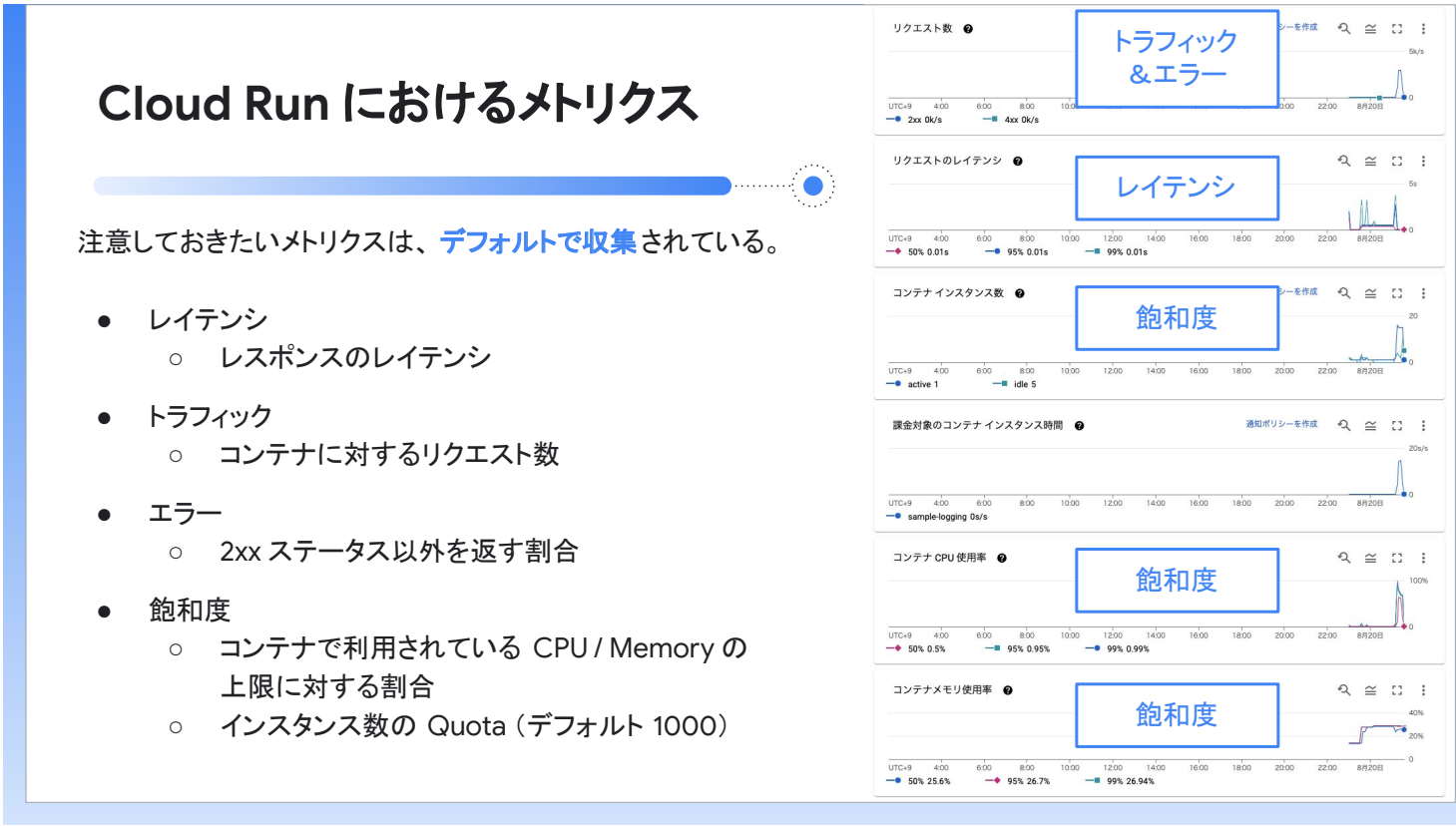
https://sre.google/books/

Google Cloud

4 つの指標は、実際にユーザーに影響を与える指標として説明されている

4 つの指標は、ユーザー体験やサービスの直帰率、コンバージョン レートなどに影響するので、どの程度許容できるかを考えてアラートを設定することが必要です。

Cloud Run では、Cloud Monitoring と連携して、リクエスト数やレイテンシーなどのメトリクスがデフォルトで収集されます。4 つの指標に対しては、以下の図のように対応しています。



注意しておきたいメトリクスはデフォルトで収集されている

飽和度に関しては、Cloud Run では VM などのリソースを気にする必要はありませんが、コンテナで利用されている CPU / メモリの上限に対する割合とインスタンス数の Quota（デフォルト 1,000）に関しては注意が必要です。

収集したメトリクスから、アラートの通知のポリシーを簡単に設定できます。Cloud Run のモニタリング画面から、通知ポリシーの作成を選択して、対象のメトリクスにフィルタされた状態でアラートの通知を設定できます。より複雑な条件を設定する場合は、Monitoring Query Language（MQL）を利用します。

Cloud Run の本番運用に向け、いくつかの考慮すべきポイントを紹介しました。Cloud Run は、多くの部分を Google Cloud にお任せできるフルマネージドのサービスです。独自に検討すべきポイントは多くはありませんが、適切に設定することで、もし問題が発生しても、状況を迅速に把握できる、デバッグがしやすいなどのメリットを享受できます。

その上で、自組織において本番運用に向けて必要なことが何かを検討することが重要です。Cloud Ops を活用し、ロギング、モニタリングを行うことで、本番運用で必要な要素を埋めることができます。ぜひ、トライしてみてください。

参照リンク

1. Cloud Functions 製品紹介ページ
2. App Engine 製品紹介ページ
3. Cloud Run 製品紹介ページ
4. オペレーション スイート（旧 Stackdriver）製品紹介ページ
5. Cloud Logging 製品紹介ページ
6. Cloud Monitoring 製品紹介ページ
7. サイト信頼性エンジニアリング（SRE）
8. Monitoring Query Language（MQL）の概要

製品、サービスに関するお問い合わせ

goo.gl/CCZL78

Google Cloud の詳細については、上記 URL もしくは QR コードからアクセスしていただくか、同ページ「お問い合わせ」よりお問い合わせください。

Google Cloud