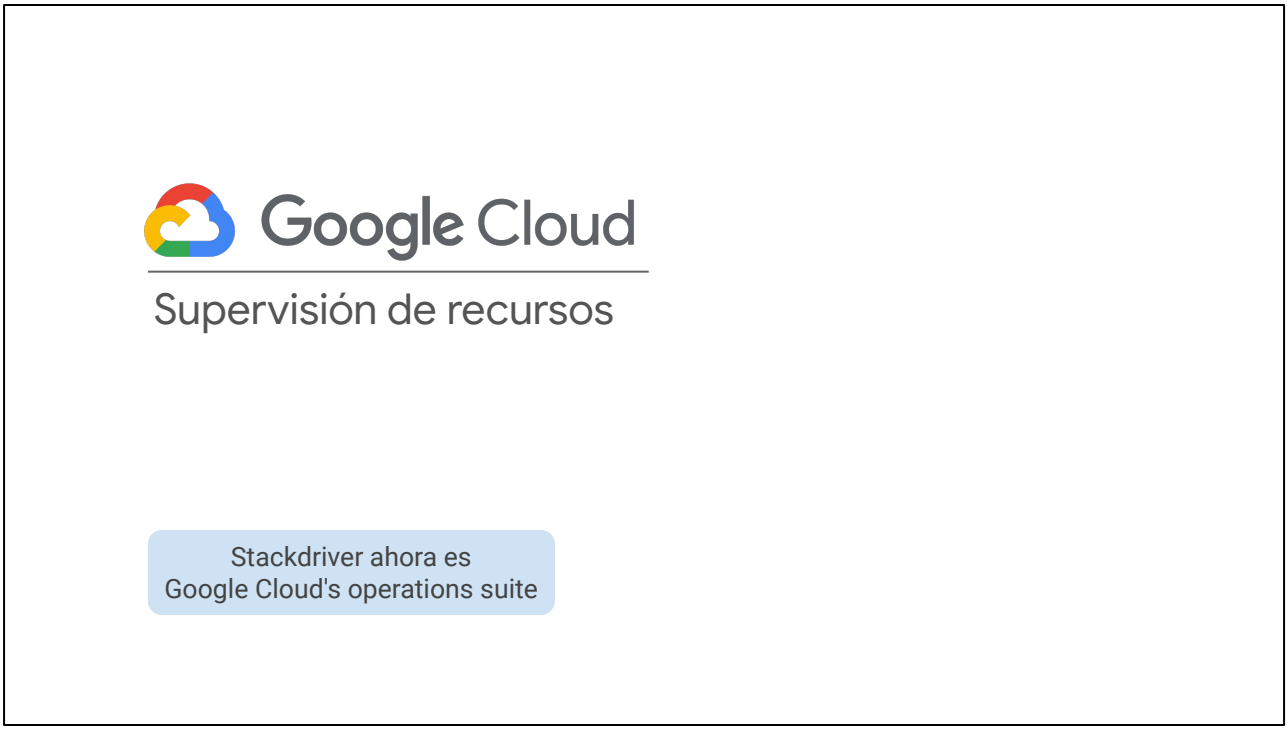


En este módulo, presentaré una descripción general de las opciones para supervisar recursos en Google Cloud.

Las funciones que analizaremos en este módulo se basan en Google Cloud's operations suite, un servicio que ofrece supervisión, registro y diagnóstico para sus aplicaciones.

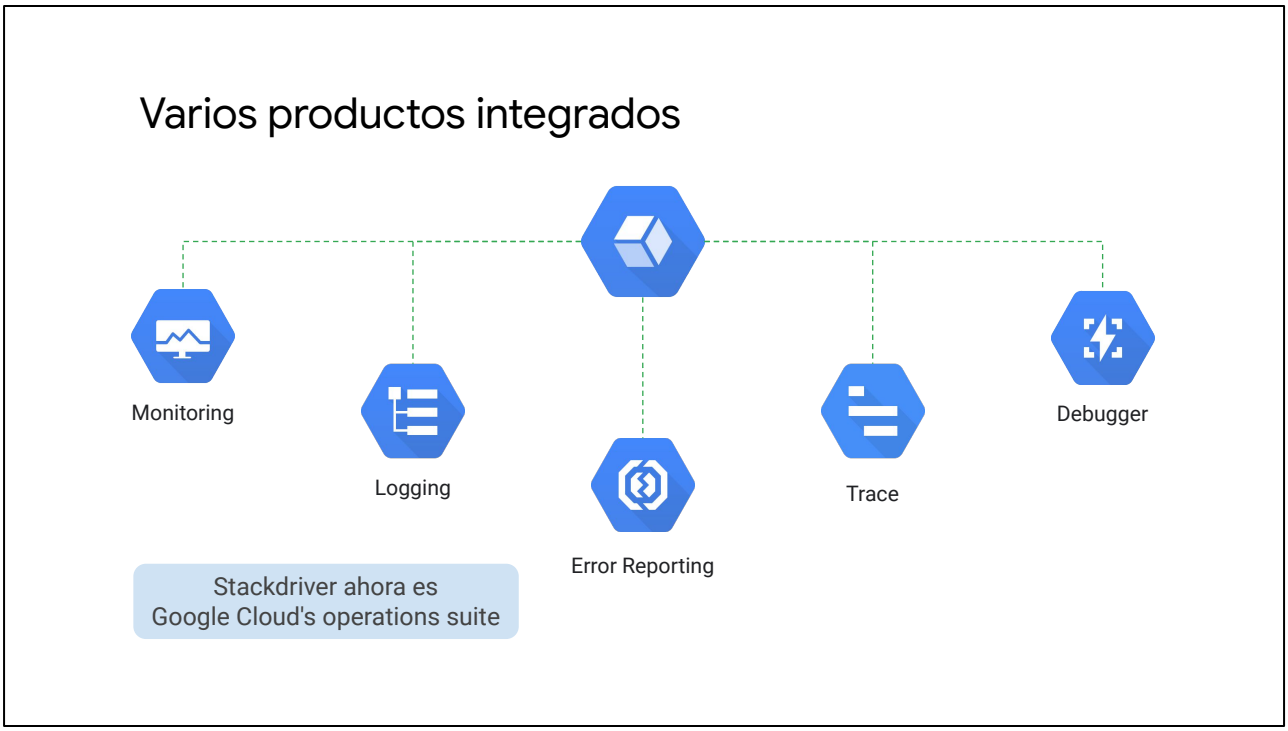
Descripción general de Google Cloud's operations suite

- Supervisión, registro y diagnóstico integrados
- Administra varias plataformas
  - Google Cloud y AWS
  - Descubrimiento dinámico de Google Cloud con valores predeterminados inteligentes
  - Integraciones y agentes de código abierto
- Acceso a herramientas potentes de datos y analítica
- Colaboración con software de terceros

Google Cloud's operations suite
(anteriormente Stackdriver)

Google Cloud's operations suite descubre de forma dinámica los recursos de la nube y los servicios de aplicación según una integración profunda a Google Cloud y Amazon Web Services. Debido a que cuenta con valores predeterminados inteligentes, puede tener visibilidad central hacia la plataforma de la nube en cuestión de minutos.

Esta ventaja le brinda acceso a las potentes herramientas de datos y analítica, así como le permite colaborar con diversos proveedores de software de terceros.



Como mencionamos antes, Google Cloud's operations suite dispone de servicios de supervisión, registro, generación de informes de errores, seguimiento de fallas y depuración. Usted solo paga por lo que usa, y hay asignaciones de uso gratuito para que pueda comenzar sin necesidad de pagar tarifas o compromisos por adelantado. Para obtener más información sobre los precios, consulte la sección de vínculos del siguiente video: [https://cloud.google.com/stackdriver/pricing]

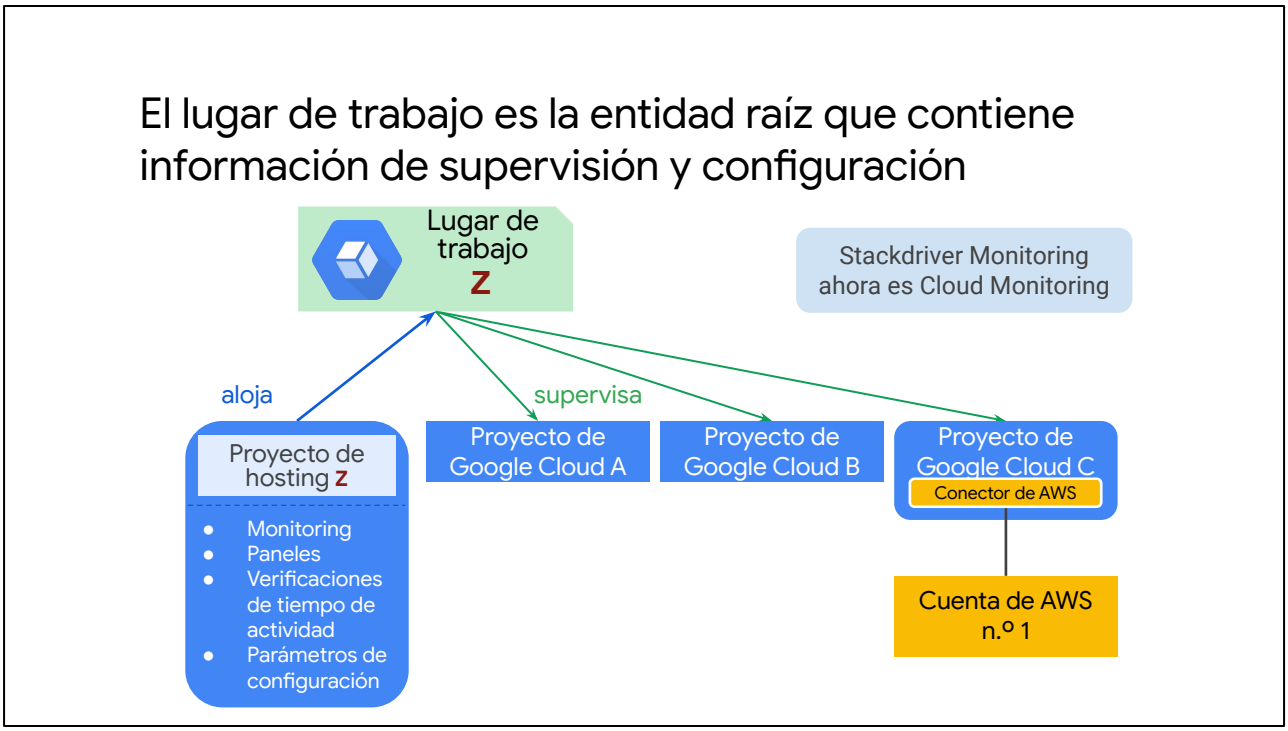
Ahora, en la mayoría de los entornos, paquetes completamente diferentes, o bien una colección de software con baja integración, administran estos servicios. Cuando vea estas funciones trabajar en conjunto en un único servicio integrado y completo, se dará cuenta de lo importante que es crear aplicaciones confiables, estables y que se puedan mantener.

Ingeniería de confiabilidad de sitios

La supervisión es importante para Google, ya que se encuentra en la base de la ingeniería de confiabilidad de sitios (SRE).

La SRE es una disciplina que aplica diferentes aspectos de la ingeniería de software a operaciones cuyo objetivo es crear sistemas de software ultraescalables y altamente confiables. Esta disciplina le ha permitido a Google crear, implementar, supervisar y mantener algunos de los sistemas de software más grandes del mundo.

Si desea obtener más información sobre la SRE, le recomendamos explorar el libro gratuito que escribieron los miembros del equipo de SRE de Google. Se encuentra en la sección de vínculos del siguiente video:
[https://landing.google.com/sre/book.html]

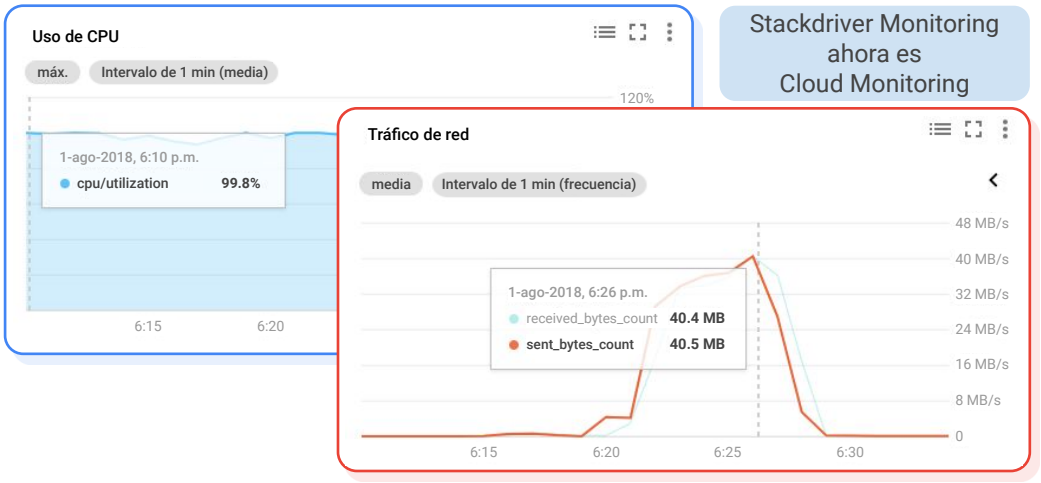


Un lugar de trabajo es la entidad raíz que contiene la información de supervisión y configuración en Cloud Monitoring. Cada lugar de trabajo puede tener entre 1 y 100 proyectos supervisados, incluidos uno o más proyectos de Google Cloud y cualquier cantidad de cuentas de AWS. Puede tener todos los lugares de trabajo que desee, pero los proyectos de Google Cloud y las cuentas de AWS no se pueden supervisar en más de un lugar de trabajo.

Un lugar de trabajo contiene los paneles personalizados, las políticas de alertas, las verificaciones de tiempo de actividad, los canales de notificaciones y las definiciones de grupo que usa con sus proyectos supervisados. Un lugar de trabajo puede acceder a los datos de métricas de los proyectos supervisados, pero las entradas de registros y los datos de métricas permanecen en los proyectos individuales.

El primer proyecto supervisado de Google Cloud en un lugar de trabajo se llama proyecto de hosting y debe especificarse cuando se crea el lugar de trabajo. El nombre de ese proyecto se convierte en el nombre del lugar de trabajo. Si desea acceder a una cuenta de AWS, debe configurar un proyecto en Google Cloud para que contenga el conector de AWS.

En los paneles, se visualiza la utilización y el tráfico de red



Cloud Monitoring permite crear paneles personalizados que contienen gráficos de las métricas que desea supervisar. Por ejemplo, puede crear gráficos en los que se muestre el uso de CPU de sus instancias, los paquetes o bytes que envían y reciben esas instancias, y los paquetes o bytes que rechaza el firewall en ellas.

En otras palabras, los gráficos brindan visibilidad a la utilización y el tráfico de red de sus instancias de VM, como se muestra en esta diapositiva. Estos gráficos se pueden personalizar con filtros para quitar el ruido, grupos para reducir la cantidad de series temporales y conjuntos para agrupar varias series temporales.

Para obtener una lista completa de las métricas admitidas, consulte el vínculo a la documentación del siguiente video:
[https://cloud.google.com/monitoring/api/metrics_gcp]

Las políticas de alertas pueden notificar ciertas condiciones

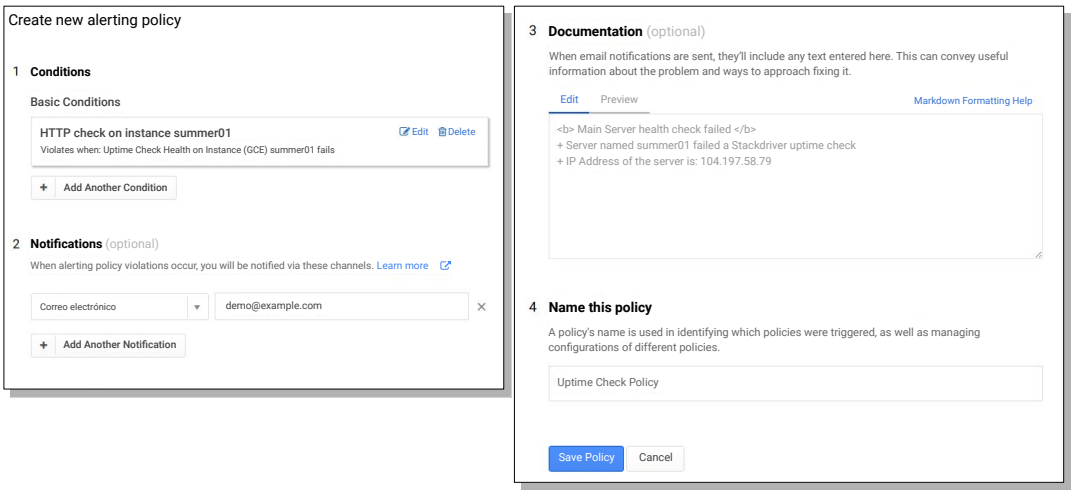
Si bien los gráficos son extremadamente útiles, solo pueden brindar estadísticas mientras alguien los revisa, pero ¿qué sucede si el servidor se apaga en medio de la noche o durante el fin de semana? ¿Espera que alguien siempre esté pendiente de los paneles para determinar si sus servidores están disponibles o si tienen suficiente capacidad o ancho de banda?

Si no es así, le recomendamos crear políticas de alertas que le notifiquen cuándo se cumplen condiciones específicas.

Por ejemplo, como se muestra en esta diapositiva, puede crear una política de alertas cuando la salida de red de su instancia de VM supera algún umbral por un período específico. Cuando se cumpla esta condición, usted o alguien más recibirá automáticamente una notificación por correo electrónico, SMS o cualquier otro canal con el fin de solucionar este problema.

También puede crear una política de alertas que supervise su uso de Google Cloud's operations suite y le envíe alertas cuando se acerque al umbral de facturación. Para obtener más información sobre esto, consulte la sección de vínculos del siguiente video:
[https://cloud.google.com/stackdriver/pricing#alert-usage]

Cree una política de alertas

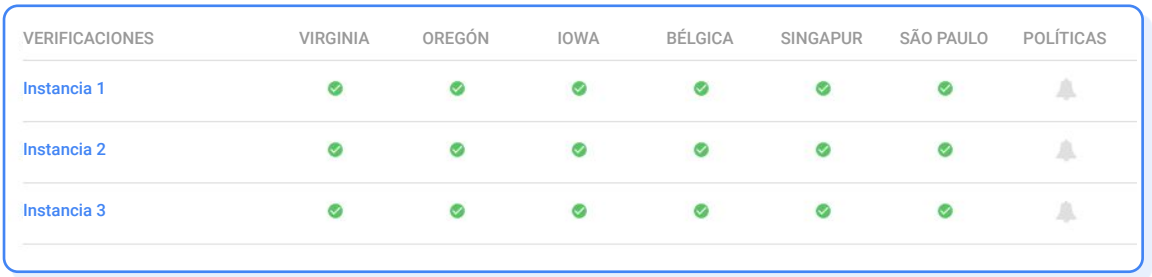


Este es un ejemplo de cómo luce el proceso para crear una política de alertas. A la izquierda, puede ver una condición de verificación de HTTP en la instancia summer01. Esta condición enviará un correo electrónico que se personaliza con el contenido de la sección de documentación a la derecha.

Analicemos algunas de las siguientes prácticas recomendadas para crear alertas:

- Recomendamos generar alertas sobre síntomas, no necesariamente sobre causas. Por ejemplo, se recomienda supervisar las consultas con errores de una base de datos y, luego, identificar si la base de datos dejó de funcionar.
- Posteriormente, asegúrese de que use varios canales de notificaciones, como correo electrónico y SMS, para evitar un punto único de fallo en su estrategia de alertas.
- Además, también recomendamos personalizar sus alertas según las necesidades del público describiendo las acciones necesarias que deben tomarse o los recursos requeridos que deben examinarse.
- Finalmente, evite el ruido, ya que este podría causar alertas que deberán descartarse con el tiempo. Específicamente, ajuste las alertas de supervisión de modo que sean prácticas. No configure alertas sobre todos los eventos posibles.

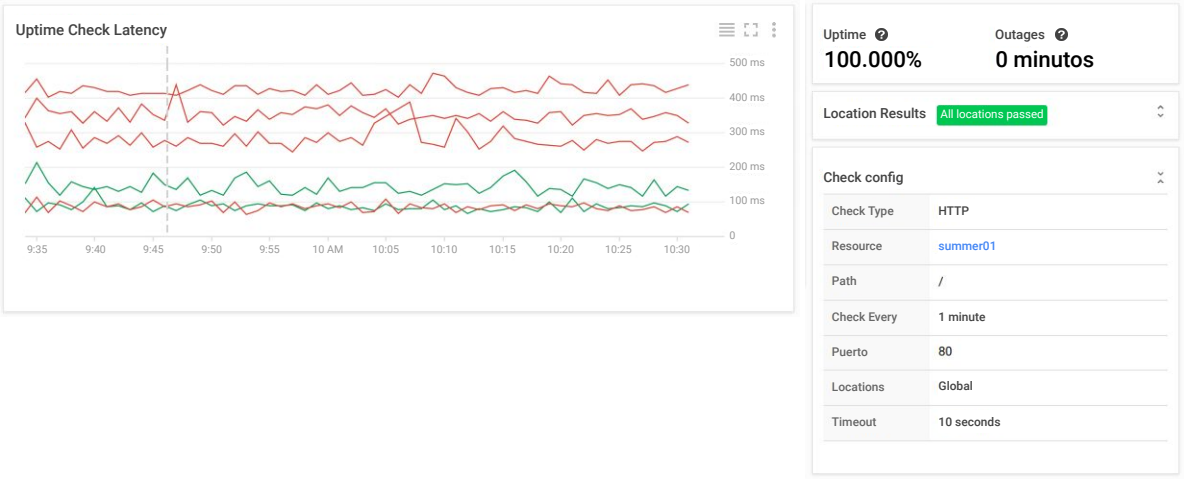
Las verificaciones de tiempo de actividad prueban la disponibilidad de sus servicios públicos



Las verificaciones de tiempo de actividad se pueden configurar para probar la disponibilidad de sus servicios públicos en diferentes ubicaciones en el mundo, como puede ver en esta diapositiva. El tipo de verificación de tiempo de actividad se puede configurar en HTTP, HTTPS o TCP. El recurso que se verificará puede ser una aplicación de App Engine, una instancia de Compute Engine, una URL de un host, o una instancia o un balanceador de cargas de AWS.

Para cada verificación de tiempo de actividad, puede crear una política de alertas y ver la latencia de cada ubicación global.

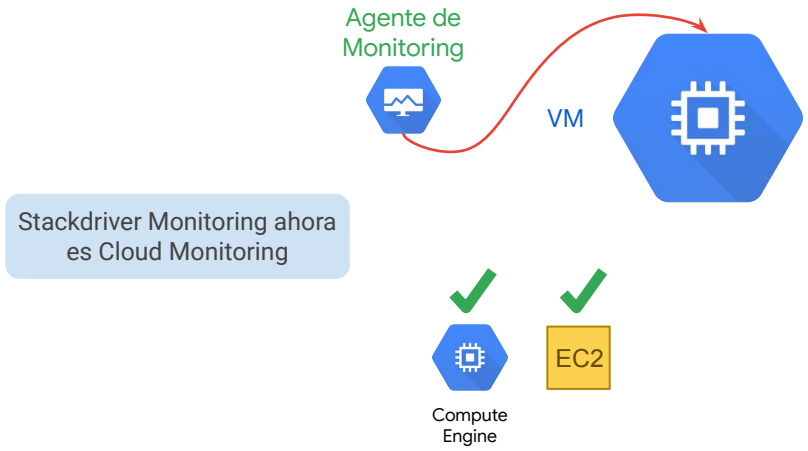
Ejemplo de verificación de tiempo de actividad



Este es un ejemplo de una verificación de tiempo de actividad de HTTP. El recurso se verifica cada minuto con un tiempo de espera de 10 segundos. Las verificaciones de tiempo de actividad que no reciban una respuesta dentro de este tiempo de espera se considerarán fallas.

Hasta ahora, hay un 100% de tiempo de actividad sin interrupciones.

Agente de Monitoring



Cloud Monitoring puede acceder a algunas métricas sin este agente, incluidos el uso de CPU, algunas métricas de tráfico de disco, información sobre el tiempo de actividad y tráfico de red.

Sin embargo, para acceder a recursos del sistema y servicios de aplicaciones adicionales, debe instalar el agente de Monitoring.

El agente de Monitoring es compatible con instancias de EC2 y Compute Engine.