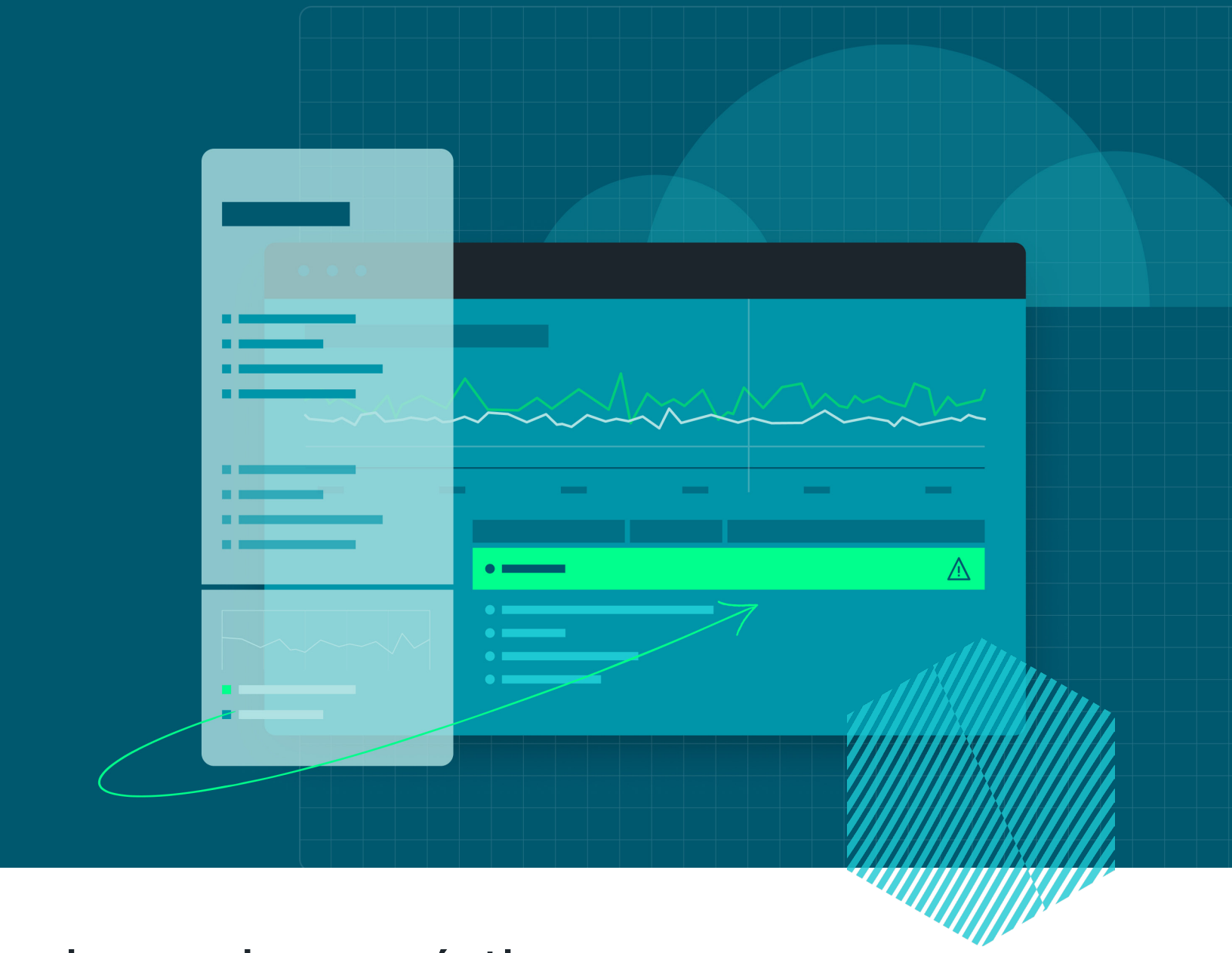
new relic



Las mejores prácticas de la administración de logs

Adopta el logging eficaz para la observabilidad de todo el stack

Whitepaper / Agosto 2022

Contenido

Introducción

La administración de logs ha evolucionado. Las organizaciones ya no tienen que rebuscar en montañas de logs sin procesar de aplicaciones e infraestructura cada vez que algo deja de funcionar. El registro de logs o logging ahora juega un papel crucial en las operaciones, la inteligencia empresarial y el marketing de una organización. Los logs impulsan la observabilidad. Los logs bien estructurados son un súperpoder que permite que las organizaciones puedan comprender rápida y fácilmente cómo funciona todo su sistema, e incluso adelantarse a los problemas.

Emplear los logs para una observabilidad eficaz exige más atención y cuidado que el simple hecho de volcar enormes cantidades de logs mal formateados en una base de datos o un archivo. ¿Cómo pueden las organizaciones cambiar de manera inteligente sus prácticas de logging para que los logs detallados puedan mejorar su capacidad de correlacionar los incidentes en todas las aplicaciones y toda la infraestructura, en tiempo real, sin tener que alternar entre distintas aplicaciones y herramientas? ¿Cómo pueden lograr una mejor observabilidad de extremo a extremo? ¿Cómo pueden estar más cerca de conseguir la observabilidad de todo el stack para beneficio de todos sus negocios?

Es fácil cambiar las prácticas de logging para asegurarse de que los logs refuercen la observabilidad de todo el stack. En este whitepaper, analizamos algunas de las mejores prácticas de logging para las organizaciones modernas.

Logging tradicional

Tradicionalmente, la actividad de logging tiene lugar en un silo de datos que se almacena aparte de otros sistemas. En el pasado, la observabilidad dependía del monitoreo del rendimiento de aplicaciones (APM) y del monitoreo de infraestructura. Con todo lo importante que es el monitoreo, no da a conocer todos los detalles de lo que sucede en el interior de los logs de aplicaciones y de dispositivos de infraestructura. Muchas herramientas de monitoreo y logging independientes y aisladas giran alrededor de una aplicación, se centran solamente en un segmento del stack y no pueden ofrecer información completa sobre lo que está pasando y por qué. Para los equipos es vital contar con la información necesaria para acelerar el tiempo de comercialización, comprender mejor el comportamiento del cliente y reducir el tiempo de respuesta a los incidentes.

Muchas organizaciones optan por no saber hasta el último detalle de sus logs y se les dificulta determinar la causa subyacente de los problemas, o bien utilizan herramientas aparte y tratan de mapear los detalles de los logs con los errores y las trazas. Mantener logs detallados en silos separados es un obstáculo para que los equipos puedan ver el panorama completo, lo que aumenta los costos y el tiempo de comercialización para los productos, reduce la visibilidad de la experiencia del cliente y aumenta el tiempo medio de resolución (MTTR).

Observabilidad de todo el stack

La capacidad de ver todo lo que hay en el stack tecnológico que pudiera afectar la experiencia del cliente se conoce como observabilidad de todo el stack u observabilidad de punta a punta. Se basa en una vista integral de todos los datos de telemetría (métricas, eventos, logs y trazas).

La observabilidad de todo el stack proporciona una visibilidad completa de cómo rinden las aplicaciones y los sistemas complejos —idealmente desde una única solución integrada— para poder ubicar y corregir los incidentes, reducir el MTTR y comprender la experiencia del cliente.

Con la observabilidad de todo el stack, los ingenieros y desarrolladores no tienen que lidiar con datos de muestra ni sacrificar la visibilidad del stack tecnológico ni perder tiempo juntando datos aislados. Por el contrario, tienen oportunidad de poner toda su atención en el desarrollo de código de más alta prioridad, más creativo, más satisfactorio para ellos y más positivo para el negocio.

Logging para la observabilidad de todo el stack

Generar logs para todo el stack puede ser una tarea abrumadora. Los desarrolladores e ingenieros pueden tener preguntas en relación a qué incluir en los logs, cuánto detalle incluir y si un exceso de datos puede elevar los costos. Muchas organizaciones pagan el alto precio de centralizar la administración de logs en una plataforma distinta y finalmente se ven forzados a limitar los datos de los logs enviados en función del rendimiento y el precio. Este tipo de muestreo de datos limita la visibilidad y el valor para el negocio. Teniendo en cuenta lo anterior, examinamos algunas de las mejores prácticas de la administración de logs para la observabilidad de todo el stack.

Incluir la información correcta en los logs

Los logs se generan escribiendo texto en una salida estándar o un archivo. La decisión más importante es seleccionar qué incluir en ellos. Los logs deben incluir todos los metadatos necesarios para ayudar a identificar los eventos y las causas raíz cuando se hace una investigación. Los elementos de metadatos de los logs pueden incluir mensajes de error o trazas del stack y los valores, las métricas o los eventos relacionados.

Todo lo que una organización incluye en los logs debe tener un propósito. Puede tratarse de datos de uso, eventos de usuario o errores y excepciones de la aplicación, da igual. Lo importante es que sea valioso para el equipo. La información de los datos de logs debe:

- Ser útil de inmediato y de alguna forma
- Proporcionar los detalles necesarios para comprender las causas subyacentes y tomar decisiones

Anticipar escenarios frecuentes

Los logs no son solo para responder a los incidentes. Los logs también pueden servir en otras áreas del negocio, como la creación de perfiles de rendimiento o la recopilación de datos estadísticos.

Si a la hora de generar logs se tienen en cuenta los escenarios que se dan con frecuencia, es posible asegurar que los logs ofrezcan un valor directo a la organización. Por ejemplo, los logs relacionados con la interacción del usuario pueden ofrecer información crucial sobre la experiencia del cliente.

Los logs que tienen que ver con el sistema pueden monitorear problemas o fallas de hardware. Los logs que ofrecen detalles de la aplicación pueden arrojar luz sobre el rendimiento y problemas potenciales como las fugas de memoria. Todo lo anterior puede ser importante para tomar decisiones de negocios.

Incluir mensajes útiles en los logs

Los mensajes de los logs son valiosos únicamente en la medida en que ofrecen información y contexto. Cuando los detalles son suficientes y los logs son comprensibles, los equipos pueden usarlos con eficiencia. La infraestructura de terceros suele capturar los detalles granulares necesarios. Aún así, para las aplicaciones desarrolladas en la propia empresa, los equipos siempre deben capturar los detalles en los logs que les permitan diagnosticar y determinar por qué ocurrió un error o evento para poder tomar las medidas necesarias que afectan al negocio.

Cuando se trata de errores de la aplicación, el mensaje debe informar lo que está pasando con la línea de código. Por ejemplo, un mensaje de error que diga `Transacción fallida` no ayuda tanto como un mensaje de error que diga `Transacción fallida: no se pudo crear usuario ${path/to/file:line-number}`. Los logs que incluyen datos sobre las transacciones ayudan a los desarrolladores e ingenieros a comprender por qué fallaron las transacciones.

Normalmente, los códigos de error o los códigos de estado pueden indicar el tipo de problema de la aplicación. En lugar de solo especificar el texto o el número del código de error, si se agrega una breve descripción en el log, se les puede ahorrar a otros desarrolladores o ingenieros el tiempo y el esfuerzo de tener que averiguarlo ellos mismos para resolver el problema.

Los logs deben aportar información vital a la organización. Los desarrolladores y los ingenieros deben evitar los mensajes crípticos o poco descriptivos que solo un número reducido de personas puede comprender.

Mantener los logs sencillos y concisos

Aunque es imprescindible incluir suficiente información dentro del mensaje del log, también es importante que no sea demasiado. Los datos innecesarios y excesivos en el mensaje pueden inflar el tamaño del almacenamiento de los logs y los costos, ralentizar las búsquedas en los logs y distraer del problema principal, lo que dificulta la depuración.

Los equipos deben mantener los logs concisos para captar solo la información más crítica. Los logs deben explicar por qué ocurrió un error y al mismo tiempo evitar los excesos.

Deben proporcionar información sobre la causa raíz de un error sin incluir hasta el más mínimo detalle del entorno. Por ejemplo, si a una aplicación se le hizo imposible conectarse y recuperar datos de una API interna, puede ser útil incluir los mensajes de error de la API o la información del estado de la red del entorno. Pero probablemente es innecesario incluir cuánta memoria utiliza la aplicación o cuántas aplicaciones se están ejecutando.

No olvide la fecha y hora

Los equipos deben incluir un timestamp para los logs. Quizás parezca obvio, pero los desarrolladores y los ingenieros que están acostumbrados a escribir logs en una base de datos que automáticamente incluye la fecha y la hora, podrían no pensar en agregar un timestamp en sus mensajes de logs. Se recomienda seleccionar el nivel más granular que tenga sentido incluir en los logs. Las tareas muy frecuentes podrían necesitar un seguimiento del tiempo que incluya hasta los milisegundos; en cambio el seguimiento de las tareas poco frecuentes podría llegar solo a los minutos (o incluso solo al día). Lo que importa no es solo el grado de detalle, sino la aplicación sistemática de un estándar en toda la organización.

Otra observación vital que también podría parecer obvia es que hay que sincronizar todos los sistemas a la misma vez para que una plataforma de observabilidad pueda usar el timestamp para correlacionar los eventos del log con otros datos de telemetría.

[1] (New Relic, Inc., n.d.)

Usar un formato de logs analizable

Si una plataforma de observabilidad no puede extraer datos de los logs, no es de mucha ayuda. Los equipos deben usar un formato de logs que los desarrolladores e ingenieros puedan analizar y mantener una estructura de logs homogénea para facilitar la recopilación y agregación. Por ejemplo, la administración de logs de New Relic permite que la definición de reglas de análisis de logs personalizadas sea cosa fácil,[1] pero las reglas de análisis no pueden hacer su magia si los datos de logs son ininteligibles.

Un buen ejemplo de un formato de logs sin analizar es un log de acceso NGINX predeterminado que contiene texto sin estructura. Es útil para hacer búsquedas, pero nada más. En un formato sin analizar, los equipos tendrían que hacer una búsqueda de texto completo para poder responder a la mayoría de las preguntas. El siguiente es un ejemplo de una línea típica:

```
127.180.71.3 - - [10/May/2022:08:05:32 +0000]
"GET /downloads/product_1 HTTP/1.1" 304 0 "-"
"Debian APT-HTTP/1.3 (0.8.16~exp12ubuntu10.21)"
```

Después del análisis, el log se organiza en atributos, como `código de respuesta` y `URL de solucitud` . El siguiente es un ejemplo de la misma información de log en un formato de logs analizable:

```
{
  "remote_addr":"93.180.71.3",
  "time":"1586514731",
  "method":"GET",
  "path":"/downloads/product_1",
  "version":"HTTP/1.1",
  "response":"304",
  "bytesSent": 0,
  "user_agent": "Debian APT-HTTP/1.3
  (0.8.16~exp12ubuntu10.21)"
}
```

Si el formato es completamente personalizado, configurar el tipo de log activa las reglas de análisis definidas por el cliente.

Si una organización tiene varias aplicaciones que sirven un propósito común, los equipos deben ocuparse de estandarizar un formato de logs que se pueda usar con todas las aplicaciones. Esto facilita la incorporación de datos en su plataforma de observabilidad, incluso cuando el equipo asociado con cada aplicación desea tener visibilidad de los distintos atributos.

Un análisis a fondo de los formatos de logs

Normalmente se usan tres categorías de formato para estructurar el texto que afectan su facilidad de uso una vez que la herramienta de agregación de logs recopila los datos. Las tres categorías de formato son las siguientes:

- **Estructurado**. Uno de los formatos estructurados más comunes para registrar logs es JSON: muchas herramientas lo pueden analizar rápidamente, y es muy flexible y liviano. Lo ideal es que todos los logs se generen con un formato estructurado. Si bien JSON ayuda a organizar los datos jerárquicos, hay otros ejemplos de datos de logs estructurados que incluyen formatos de uso común como los valores separados por coma (CSV) y los valores separados por tabulador (TSV).
- **Común**. Un formato común no es estructurado pero es muy conocido, definido e invariable. El formato de logs común de Apache para logs de acceso es un ejemplo. La ventaja de un formato común es que muchas herramientas pueden analizar los datos sin necesidad de configurar nada.
- **Personalizado**. Si una aplicación no utiliza un formato estructurado o uno común, escribe los logs empleando un formato personalizado. Para reconocer el inicio y el fin de una línea de log individual durante el reenvío de logs, es posible que los equipos tengan que realizar análisis. Si se crean reglas de análisis definidas por el cliente, se contribuye a que los datos sean más valiosos.

Categorizar y agrupar logs

La especificación de un modelo de datos para los logs permite que los equipos puedan hacer búsquedas más eficaces. Deben definir e incluir atributos siempre que sea posible para categorizar y agrupar los logs según corresponda.

Las normas de OpenTelemetry para logs creadas por una coalición de líderes de la industria (entre ellos New Relic), abarcan muchos elementos como las convenciones de nombres y las definiciones de valores de campos.[2] Aunque no todos las estructuras admiten de manera nativa los logs formateados estrictamente según esas normas, pueden servir de guía.

Algunos atributos comunes que pueden ser útiles en un modelo de datos de logs son los recursos, los logs en contexto y los niveles de logs.

Recursos
Los recursos definen cuándo y de dónde vinieron los logs, por ejemplo:

- Fecha y hora
- Nombre de host o identificador de la máquina
- Nombre de la aplicación o el servicio

El nombre de host puede ser significativo en los logs de las aplicaciones basadas en host clásicas con entornos que tienen nombre. Un ID de pod o de contenedor sería mejor para organizar los logs de entornos contenerizados u orquestados.

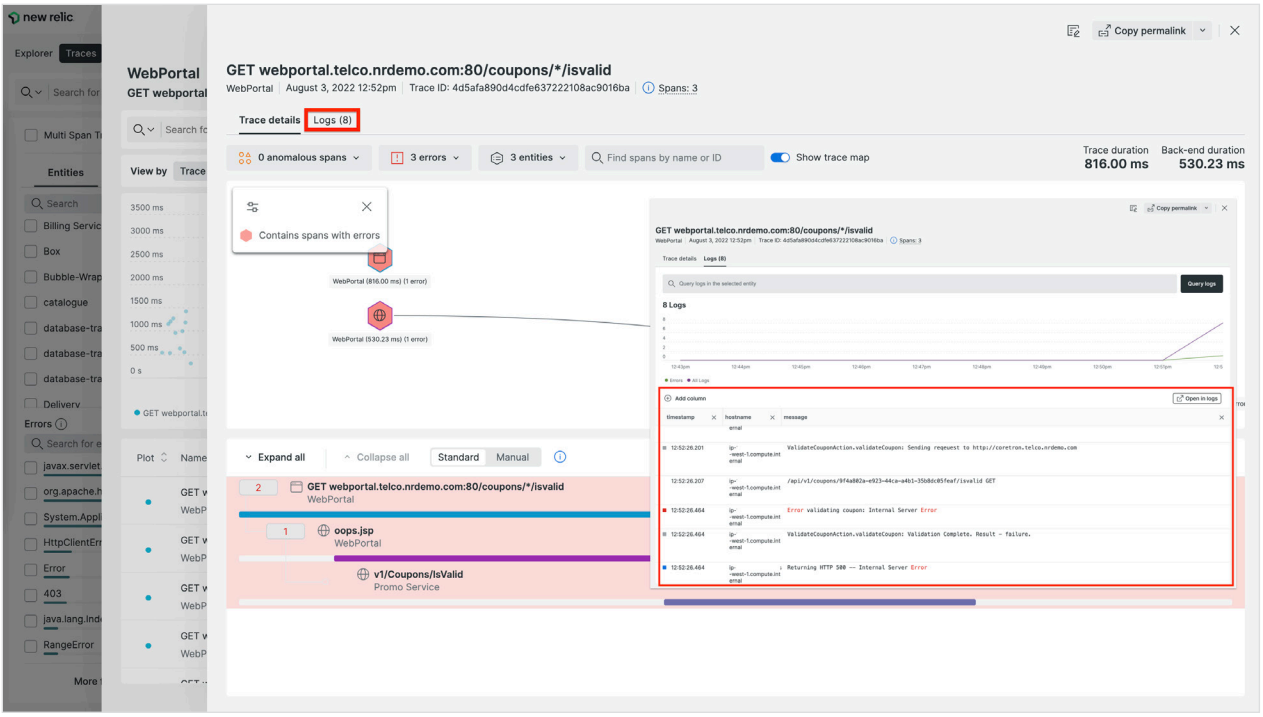
A menudo, los entornos orquestados o de plataformas como servicio (PaaS) rellenan automáticamente los logs con muchos metadatos. Eso es excelente para organizarse, pero también es importante incluir anotaciones en los logs que proporcionen calificadores útiles que un sistema no puede conocer, como los números de versión de un producto, los entornos de prueba o de producción, las ramas de prueba o las versiones de prueba A/B. La agregación de logs significa que se recopilan todos los logs de distintas fuentes en un mismo sistema. Sin los metadatos correctos, los equipos no pueden distinguir entre un log de error real en producción y una transacción que falló como parte de una ejecución de prueba.

Otro recurso que puede ayudar a los equipos a identificar el origen del log es el reenvío de logs. Por ejemplo, la mayoría de las soluciones de reenvío de logs que proporciona New Relic agrega anotaciones automáticamente a los datos para especificar el tipo y la versión de la herramienta utilizada para enviar los datos.

[2] (OpenTelemetry, n.d.)

Logs en contexto

A los equipos les resulta útil ver los logs en contexto en sus aplicaciones y hosts. Por ejemplo, la función Logs in context (Logs en contexto) de New Relic es capaz de agregar automáticamente información de la aplicación a los logs. El agente APM de New Relic aporta datos de la administración del rendimiento de la aplicación a la estructura de logging y los incluye en los logs de la aplicación. Esto se traduce en que la función Logs in context correlaciona automáticamente los datos de los logs con los eventos y las trazas de las aplicaciones asociadas. Los errores de APM y las trazas distribuidas se vinculan directamente con los logs creados durante la misma transacción que cada error o traza. La función Logs in context establece esta correlación insertando una ID de span, una ID de traza y un nombre de aplicación en los mensajes de logs. De esta forma, los equipos pueden unir los datos de la aplicación y de los logs y acelerar la resolución de problemas.



Logs filtrados para mostrar los errores en un contexto de trazas en la plataforma de observabilidad de New Relic

Niveles de logs
Los desarrolladores, los especialistas de DevOps y los gerentes a veces se refieren a los niveles de logs como niveles de severidad. Los niveles de logs describen la importancia relativa de un evento (con términos como depuración, información, advertencia, error y fatal) y el nivel de densidad de la información del framework de logging. Un atributo de severidad ayuda a filtrar o descartar información menos crítica para que los equipos puedan buscar únicamente los errores críticos.

Usados de manera eficaz, los niveles de logs pueden restringir la cantidad de datos, reducir el costo de usar una herramienta de administración de logs centralizada y asegurarse de que las búsquedas sean siempre rápidas. Idealmente, el sistema de administración de logs se encargaría de descartar los datos no deseados en los casos en que no es posible controlar cómo las aplicaciones generan los logs. Por ejemplo, con New Relic, los equipos pueden mostrar valores atípicos usando patrones impulsados por el aprendizaje automático en función del nivel del log. Los niveles de logs codificados por color también ofrecen una indicación visual que ayuda a centrar la atención en las áreas más críticas.

Los equipos deben usar los niveles de logs con cuidado, especialmente el nivel de log de depuración. La depuración puede ayudar a captar los mensajes muy extensos asociados con un comportamiento concreto. Sin embargo, una depuración innecesaria puede crear un volumen de logs significativamente más alto y ralentizar las funciones de ingesta y búsqueda sin aportar valor adicional. Los equipos y proyectos más grandes pueden beneficiarse de las normas de nivel de logs para que la agrupación, la categorización y los métodos de logging se apliquen de manera coherente.

Utilizar herramientas y estructuras de logging

En lugar de invertir tiempo y recursos en implementar una solución de logging desde cero, utilizar una herramienta y un framework de logging ya establecidos y bien probados puede ahorrar tiempo y problemas a los equipos. Por ejemplo, los agentes de lenguajes APM de New Relic decoran los logs con los metadatos necesarios para dar acceso a la función automática Logs in context (logs en contexto) y reenviar los logs sin necesidad de instalar o mantener software de terceros, todo en un solo despliegue.

Cuando se usa una estructura de logging invariable, se simplifica la adopción del equipo de ingeniería, se normaliza la salida de los logs y se garantiza que los equipos puedan activar la función Logs in context de una manera uniforme. Los equipos deben tener cuidado al introducir frameworks de logging y probar el impacto en el rendimiento, al igual que lo harían con cualquier código nuevo.

Hacer referencia a los valores grandes, no incluirlos

Hay ocasiones en que los equipos necesitan una parte significativa de los datos del log para poder ofrecer un contexto más exhaustivo, como un volcado de memoria o un conjunto de archivos o imágenes. En general, resulta más conveniente guardar ese tipo de datos por separado o incluso cargarlos en un servidor designado y hacer referencia a su ubicación en el log en lugar de guardarlos como un amasijo de datos dentro del log. Los equipos deben mantener los logs lo más ligeros posible y acceder a los datos por separado.

Compartir vistas, consultas y alertas útiles

Los equipos deben crear y compartir las visualizaciones, consultas y alertas estándar para que sus logs puedan brindar una perspectiva más amplia del estado actual de su organización, y mejorar la visibilidad y la comunicación entre los equipos. En eso radica el poder de la observabilidad de todo el stack.

¿Qué no incluir en los logs?

Aunque puede ser tentador incluir todo tipo de cosa potencialmente útil, existen ciertas excepciones que los equipos deben tomar en cuenta y trampas que deben evitar.

Información confidencial

Los equipos deben tratar la información confidencial con cuidado. Es vital proteger los datos regulados, como la información de identificación personal (PII) y los números de tarjetas de crédito de acuerdo con los reglamentos y las leyes aplicables, como el Reglamento General de Protección de Datos (GDPR) de la Unión Europea [3] y la ley de Responsabilidad y Portabilidad del Seguro Médico (HIPAA) en Estados Unidos.[4]

La guía de logging del Open Web Application Security Project (OWASP) especifica qué no debe incluirse en los logs, como los tokens de acceso, las contraseñas, la información confidencial y la información que las personas desean que se mantenga en privado.[5]

En los logs que se almacenan en un servidor privado o una base de datos privada, es fácil que sin quererlo se incluya PII, como nombres y direcciones de correo electrónico. Para hacer seguimiento de las acciones o los eventos específicos de un usuario, los equipos deben utilizar identificadores anónimos. Aunque los datos de los logs están seguros en una plataforma de observabilidad como New Relic, es importante tener mucho cuidado para evitar la transmisión de PII fuera de la organización.

Código fuente y datos de propiedad exclusiva

Aparte de la información reglamentaria y de cumplimiento, los equipos también deben evitar almacenar otra información confidencial dentro de los logs, como el código fuente de las aplicaciones o los datos protegidos dentro de la organización.

Además de almacenar los logs de manera segura, también es importante garantizar el acceso a ellos. La información que puede revelar secretos comerciales o proyectos y funciones que aún no se han publicado o anunciado no pertenece dentro de los logs. Por ese motivo, los equipos deben eliminar esta información de los logs, especialmente si los logs se almacenan externamente, en un servicio de terceros.

Información duplicada

Agregar información duplicada a los logs no ocasiona problemas, y tener demasiada información por lo general es mejor que no tener suficiente. Sin embargo, incluir información duplicada puede crear logs innecesarios que no tienen ningún propósito, lo que se traduce en costos más altos.

[3] (European Commission, n.d.)
[4] (U.S. Department of Health and Human Services (HHS), n.d.)
[5] (Open Web Application Security Project (OWASP), n.d.)

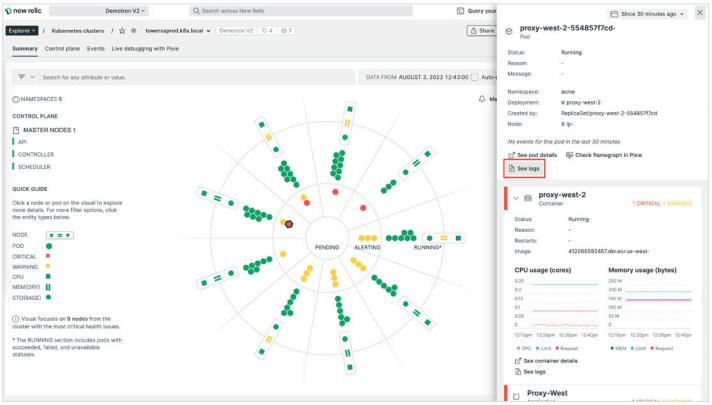
Conclusión

Hacer que los logs sirvan para mejorar la observabilidad de todo el stack abre la oportunidad de tomar decisiones en tiempo real con un impacto en el negocio a la vez que se garantiza que los desarrolladores e ingenieros pasen menos tiempo depurando y respondiendo a incidentes y más tiempo innovando.

Una vez implementadas estas mejores prácticas, los logs pueden aportar los detalles necesarios para hacer que todo funcione bien para los clientes, dar una mayor visibilidad de todo el stack para resolver problemas con más rapidez y acelerar el desarrollo.

Plataforma de observabilidad de New Relic

New Relic proporciona una sola plataforma unificada para todos los datos de telemetría, incluidos los logs detallados. La plataforma de observabilidad de New Relic incorpora administración de logs, APM, monitoreo de infraestructura, monitoreo serverless, monitoreo de móviles, monitoreo de browser, monitoreo sintético, rastreo distribuido y monitoreo de Kubernetes, entre otras cosas. Estas capacidades permiten a las organizaciones visualizar, analizar y resolver problemas en todo su stack de software. Como parte de esta plataforma, la administración de logs de New Relic permite a las organizaciones combinar los datos de logs con los datos de monitoreo de aplicaciones y de monitoreo de infraestructura, lo que resulta en una plataforma de observabilidad completa y potente.



APM, infraestructura, eventos y acceso a los logs combinados en una sola vista

New Relic conecta las métricas, los eventos, los logs y las trazas de todo el stack de software integrado con AIOps (inteligencia artificial para las operaciones de TI), lo que permite a las organizaciones hacer búsquedas más rápidas en los logs y resulta más económico que las distintas soluciones heredadas. En lugar de usar herramientas separadas en distintas partes del stack, los desarrolladores e ingenieros pueden ver fácilmente todos los logs detallados de un error en particular en una vista unificada.

Los problemas de velocidad y escalabilidad en las soluciones de logging heredadas dificultan hacer consultas en los logs detallados porque ejecutarlos con datos demorados puede tardar minutos o incluso horas. Por el contrario, una búsqueda con la administración de logs de New Relic demora apenas unos segundos, permitiendo investigar y responder a los incidentes en todo el stack de software a la mayor brevedad.

La plataforma de observabilidad de New Relic incluye la administración de logs, un nivel gratuito para los clientes de volumen bajo y un precio reducido por GB que permite a los equipos ingerir todos los logs detallados que necesitan.

Para comenzar a utilizar la administración de logs de New Relic, regístrate para una cuenta gratuita. Las cuentas gratuitas incluyen 100 GB de ingesta de datos al mes, un usuario Full Platform y un número ilimitado de usuarios Basic.

Registrarse ahora

Referencias

European Commission. n.d. “EU data protection rules.” European Commission. Accessed July 19, 2022. https://ec.europa.eu/info/law/law-topic/data-protection/eu-data-protection-rules_en.

New Relic, Inc. n.d. “Parsing log data.” New Relic Documentation. Accessed July 28, 2022. https://docs.newrelic.com/docs/logs/ui-data/parsing/#custom-parsing.

OpenTelemetry. n.d. “OpenTelemetry Logging Overview.” OpenTelemetry. Accessed July 18, 2022. https://opentelemetry.io/docs/reference/specification/logs/overview/.

Open Web Application Security Project (OWASP). n.d. “OWASP Logging Guide.” https://owasp.org/www-pdf-archive/OWASP_Logging_Guide.pdf.

U.S. Department of Health and Human Services (HHS). n.d. “Summary of the HIPAA Security Rule.” HHS.gov. Accessed July 19, 2022. https://www.hhs.gov/hipaa/for-professionals/security/laws-regulations/index.html.

new relic.