Artículo de Investigación

Monitorización de Sistemas con Tecnologías de Big Data

Systems Monitoring with Big Data Technologies

Herrera, Yogledis; Heredia, Freddy

Yogledis Herrera
yogledys1981@gmail.com
Universidad de la Rioja, España
Freddy Heredia
freddyheredia4@gmail.com
Universidad Centroccidental , Venezuela

Ecuadorian Science Journal
GDEON, Ecuador
ISSN-e: 2602-8077
Periodicidad: Semestral
vol. 5, núm. Esp.3, 2021
esj@gdeon.org

Recepción: 31 Agosto 2021
Aprobación: 04 Octubre 2021

URL: http://portal.amelica.org/ameli/jatsRepo/606/6062738014/index.html

DOI: https://doi.org/10.46480/esj.5.3.151

Como citar: Herrera , Y., & Heredia , F. (2021). Monitorización de Sistemas con Tecnologías de Big Data. Ecuadorian Science Journal, 5(3), 152-161. DOI: https://doi.org/10.46480/esj.5.3.151

Resumen: El siguiente trabajo está orientado a proveer información relevante sobre el desempeño de un aplicativo ERP a partir de los archivos de registros de eventos de mismo (logs) que pertenece a una empresa dedicada a la consultoría de sistemas empresariales ERP, con el fin de apoyar a la empresa en la búsqueda de las casusas de baja disponibilidad de su aplicativo. Los logs que son capturados por los sistemas informáticos, por lo general tienden a ser borrados, a pesar de que estos pueden ser transformados en nuevos conocimientos. En su mayoría estos son revisados cuando existe algún tipo de problema, como es el caso de la empresa seleccionada como piloto para el presente trabajo. El objetivo principal de la empresa en estudio es garantizar la disponibilidad del software ERP que comercializa, actualmente cuenta con varias implementaciones del sistema, a los que se conectan sus clientes a cualquier hora del día y desde cualquier parte del país, por lo que juega un papel importante la disponibilidad las 24 horas del día y los 365 días del año, por lo tanto requiere una visión global pero también detallada del comportamiento de sus aplicativos. Entre varias herramientas existentes en la industria del software para el procesamiento de grandes volúmenes de datos con capacidad de entregar la información requerida por la empresa antes mencionada, se seleccionó una plataforma de Big Data llama-da Elastic Stack, que permite estructurar, enriquecer, almacenar y visualizar los datos contenidos en los logs provenientes de los aplicativos de la empresa, apoyando a los tomadores de decisiones al análisis y generación de nuevos conocimientos sobre el comportamiento de los aplicativos para identificación de patrones, fallas, entre otros.

Palabras clave: Análisis de Datos, Big Data, Elk, Elasticsearch, Log, Visualización de Datos.

Abstract: The following research work is aimed to provide relevant information on the performance of an ERP application from its event log files (logs) owned by a company dedicated to consulting ERP business systems, in order to support the company in finding the causes of low availability of its ERP application. The logs that are captured by computer systems, in general tend to be erased, although these can be transformed into new knowledge. Most of these are reviewed when there is some kind of problem, as is the case with the company selected as a pilot for this work. The main objective of estudied company under study is to guarantee the availability of the ERP software that it sells, currently it has several implementations of the system, to which its clients connect at any time of the day and

from anywhere in the country, so availability 24 hours a day and 365 days a year plays a important role, therefore it requires a global but also detailed vision of the behavior of its applications. Among several existing tools in the software industry for the processing of large volumes of data with the capacity to deliver the information required by the mentioned company, a Big Data platform called Elastic Stack was selected, which allows to structure, enrich, store and visualize the data contained in the logs from the company's applications, supporting decision makers for the analysi and generation of new knowledge about the behavior of the applications to identify patterns and failures.

Keywords: Big Data, Data Analysis, Data Visualization, Elk, Elasticsearch, Log.

Introducción

Existen muchas razones que pueden ocasionar fallas en los servicios de los sistemas informáticos de una empresa, entre ellas, causas de hardware y/o de software. en términos de hardware se puede mencionar agotamiento de memoria en el servidor, insuficiencia de espacio de almacenamiento, alguna falla de conexión en la red, entre otros, y si nos referimos a causas originadas en software se puede mencionar algún software malicioso que consume los recursos del servidor, código fuente con errores en los sistemas instalados, entre otros, que de igual manera afecta los recursos del hardware y ocasionan caídas de los servidores. Indistintamente de las causas de las caídas de los sistemas, el desconocimiento del comportamiento de los mismos ocasiona grandes pérdidas de dinero y tiempo para cualquier organización, es por ello que este tema es de suma preocupación para los responsables de las organizaciones al momento de presentarse. Identificar las causas en tiempo real o predecir las caídas de los aplicativos, muchas veces se hacen difíciles, en especial si no se tiene una visualización global de lo que está ocurriendo en la aplicación.

La presente investigación tiene por finalidad dar a conocer los resultados de la implementación de una herramienta Big Data que permitió procesar datos capturados en logs de diferentes fuentes y visualizarlos en tiempo real, obteniendo una visión amplia del comportamiento de varios aplicativos. La implementación se realizó en una empresa del área tecnológica ubicada en Ambato-Ecuador, el cual tiene nueve implementaciones de un sistema, que diariamente presentan problemas de disponibilidad, por tanto el trabajo se enfocó en identificar errores desde el punto de vista de programación, en visualizar eventos, procesos y errores capturados por los logs generados por un sistema de planificación de recursos empresariales (ERP), instalados en los servidores de la empresa, con el fin de tener una idea clara del comportamiento de los mismos al momento de una falla en los aplicativos y poder tomar decisiones para solventar la problemática.

Los logs generados por los sistemas de información de la empresa en estudio emanan un gran número de líneas de registro, además de poseer variadas estructuras de logs. Estos capturan información de lo que está ocurriendo con el sistema en un determinado momento, cabe indicar que entre más transacciones se realicen y más usuarios manipulen el sistema, los logs generados se hacen más voluminosos y más difíciles de analizar, produciendo miles de líneas de logs por minuto, por tanto, realizar un análisis de logs manualmente es una tarea muy difícil, surgiendo la necesidad de implementar una herramienta de big data.

Estado del Arte

Jaramillo Valbuena y Londoño (2014) realiza una investigación sobre los sistemas de almacenamiento para grandes volúmenes de datos, incluyendo un comparativo entre los Sistemas de Administración de Bases

de Datos (DBMS) tradicionales y los nuevos enfoques NoSQL (Not Only SQL). Esta información es confirmada por (Armbrust et al., 2010) quienes afirman que los sistemas NoSQL se adecuan a casos en los que se necesita atender a muchos usuarios sin perder rendimiento, recomiendan los sistemas de bases de datos relacionales cuando se trata de garantizar integridad referencial de datos.

Elastic Stack marca la diferencia con respecto a otras herramientas que permite el análisis de logs, ya que se ajusta a lo propuesto por Jaramillo Valbuena y Londoño; Jorge Mario (2014) permitiendo la escalabilidad horizontal, fiabilidad de datos, tiempos de respuesta rápidos, proporciona una interfaz para realizar consulta, además es confiable y seguro, recibe logs en cualquier estructura, extrayendo, analizando y visualizando en tiempo real los datos capturados en los logs. Además proporciona un mecanismo sólido para realizar un registro centralizado que juega un papel importante en la identificación de servidores y/o problemas relacionados con la aplicación. Le permite buscar en todos los registros en un solo lugar e identificar los problemas que abarcan varios servidores mediante la correlación de sus registros dentro de un marco de tiempo específico que se encuentra en los entornos de tecnología de información (TI), incluidos los casos de uso de analítica web, inteligencia comercial, cumplimiento y seguridad.

Hamilton et al., (2018) presenta el resultado del trabajo del grupo de sistemas de control y seguridad industrial del CERN, empresa que produce gran cantidad de datos en sus aplicativos en diferentes áreas, el cual ha desarrollado alrededor de 200 aplicaciones de control que incluyen diversos dominios, tales como, protección magnética al LHC, criogenia y sistemas de supervisión de redes eléctricas, millones de cambios de valor y alarmas de muchos dispositivos son archivados en una base de datos Oracle centralizada siendo difícil obtener estadísticas de alto nivel de dicho archivo, por lo que se implementó un sistema soportado por ElasticSearch, Logstach y Kibana para proporcionar fácil acceso a estas estadísticas. El resultado fue un sistema que proporciona estadísticas agregadas basadas en la cantidad de cambios de valor y alarmas que son clasificadas según criterios como el tiempo, el dominio de la aplicación, el sistema y el dispositivo, además de estas estadísticas, cada aplicación genera archivos de registro basados en texto que se analizan, recopilan y se muestran mediante ElasticStack para proporcionar un acceso a todos los registros de la aplicación, el sistema implementado resultó que puede ser usado también para detectar situaciones anormales.

En el trabajo de Andreassen et al., (2015), se muestra cómo se puede procesar casi cualquier tipo de datos estructurados o no estructurados gracias a herramientas como Logstash junto con ElasticSearch, enseñando además cómo se puede visualizar la información personalizada por el usuario y cómo el sistema se adapta a medida que el volumen de datos crece. Uno de los primeros aspectos a considerar durante su investigación fue el formato de los datos debido a la variedad de fuentes como aplicaciones de LabView ejecutándose en CompactRIO y PXI, Apache Tomcat, servicios de JAVA y aplicaciones C++, donde cada uno de ellos tiene un formato y propósito propio, estableciendo como punto de partida para definir el formato principal del estándar syslog utilizado para el registro de mensajes. Gracias a la introducción de ELK, el registro de mensajes se pudo unificar en un formato común con un almacenamiento centralizado y los usuarios han creado tableros según sus propias necesidades gracias a las mejoras realizadas en la gestión y análisis de datos. La detección de errores fue otro aspecto mejorado gracias a la implementación de estas herramientas, identificándolos fácilmente y reduciendo el tiempo de detección por parte de los programadores.

Por otra parte, Prakash et al.,(2016) muestran el funcionamiento del ecosistema Elasticsearch, Logstash y Kibana (ELK) para analizar de manera eficiente los archivos de registro, resultando una herramienta de gestión de monitoreo muy útil debido a su facilidad de uso, donde se pueden interpretar los resultados y obtener información de forma interactiva con el fin de identificar geográficamente a los usuarios del sistema en función de los registros de acceso utilizando ELK, también para el seguimiento de actividades fraudulentas y violaciones de seguridad resultando prometedor para su uso en IoT.

Bajer (2017) , se propuso la implementación de una estrategia utilizando ELK para procesar datos de IoT, a pesar de que estas herramientas fueron diseñadas principalmente para manejar grandes cantidades de datos de registro, aplicadas para almacenar, buscar y visualizar datos de IoT. Esta investigación logró demostrar

que ElasticSearch no solo proporciona una base de datos de alto rendimiento y un mecanismo para recopilar datos versátiles, sino que también brinda una plataforma para visualización de datos de manera sencilla y con alto rendimiento.

La investigación presentada por Langi et al., (2015), se llevó a cabo un análisis de la red social Twitter, que puede ser utilizada para mostrar información sobre una persona, un servicio o un producto desde la perspectiva la perspectiva del usuario de la red. Twitter comparte una API que permite obtener datos en tiempo real, usando ElasticSearch se puede procesar y analizar la información que ofrece la API. Como resultado de esta investigación se obtuvo un análisis comparativo de rendimiento entre ElasticSearch y Logstash como fuente de entrada para análisis de dat a de social media y una herramienta propia de la red social llamada Twitter River mostrando que Logstash utiliza más recursos que Twitter River incluyendo 0.99% más procesador CPU, 7.2% más RAM por día, 112MB más de espacio de disco duro usado semanalmente.

Mateos Mohíno (2017), en su trabajo denominado "Logs Analyzer: Herramienta para la evaluación en tiempo real de registros log con tecnologías Big Data", realiza "un estudio y análisis en profundidad del archivo de registro de acceso (Acces log file), el cual recoge actividades relacionada con las peticiones llevadas a cabo sobre un sistema". Para lograr su objetivo utiliza la plataforma Apache Hadoop y desarrolla en el framework "stack MEAN" para la interfaz gráfica con el usuario, incluye MongoDB como base de datos, para la extracción de información el autor utiliza Apache Flume, y finaliza con la programación de las visualizaciones de los requerimientos de su problemática.

De igual manera Amaducci Szwarc (2016), en su trabajo "Monitorización de Sistema con Bluemix " desarrolla una aplicación para el análisis de logs a través de varios lenguajes de programación. El trabajo de Amaducci se diferencia al de este trabajo de investigación en que no se requerirá de conocimientos avanzados de programación ya que Elastick stack requiere sólo configuraciones de varias herramientas que incorpora la tecnología y que integradas trabajan muy bien.

Desde el punto de vista de la finalidad que tiene el análisis de logs, un gran número de trabajos que realizan análisis de logs están orientados a resolver problemas de seguridad, en detectar vulnerabilidades y fallos de software enfocados a la seguridad informática (Ma Begoña, 2016), diferenciándose del presente trabajo que está orientado a determinar anomalías en los aplicativos de una empresa.

Desde el punto de vista de arquitectura para la gestión de logs, González (2015) en su trabajo menciona que "la infraestructura necesaria para poder realizar una gestión de logs dentro de una organización suele describirse mediante una arquitectura compuesta por capas", refiriéndose a las capa de generación, recepción, análisis, almacenamiento y la monitorización.

González (2015) Realiza un análisis entre varias arquitecturas para la gestión de logs (Sistemas Syslog , Rsyslog, Syslog-ng, Nxlog, Graylog2 y ELK) proponiendo ELK Stack como una herramienta completa para el análisis de logs, conclusión que llegó, luego de probar con varios escenarios que comprobaron que la tecnología seleccionada es apropiada para la visualización de datos almacenados en logs. Diferenciándose del presente trabajo, ya que aparte de visualizar información extraída de los logs se enfoca en que la visualización responda a ciertos requerimientos identificados y que sea fáciles de interpretar.

Partiendo de los resultados presentados por González (2015) de seleccionar una herramienta de análisis de logs que cumpla con las características de una herramienta Big Data, esto quiere decir que permite analizar distintos tipos de fuentes, almacenar y visualizar en tiempo real los datos extraídos en los logs, este trabajo complementa como se pueden realizar visualizaciones que sean fácil de interpretar por el usuario, gráficas que muestre en tiempo real, que permita monitorizar los aplicativos instalados en varios servidores etc.

Ligus (2013) define la monitorización como el proceso de mantener la vigilancia sobre la existencia y la magnitud del cambio de estado y el flujo de datos en un sistema, y que tiene como objetivo identificar los fallos y ayudar en su posterior eliminación. Jason (2017), define a la monitorización como "el conjunto de software y procesos que son utilizados para asegurar la disponibilidad de uno o más sistemas", en un

sentido abstracto afirma que puede ser dividido en tres grandes categorías: detección de fallas, producción de alertas y planificación de crecimiento. Este concepto es de gran relevancia para este trabajo ya que se pretende implementar una herramienta que permita visualizar los datos capturados en los logs con el fin de determinar fallas a través de la monitorización de sistemas, buscando tener una visión amplia del comportamiento de los aplicativos de la empresa en estudio, que maneja grandes volúmenes de datos, ya que los mismos constantemente dejan de funcionar y no se conocen las causas.

Lauriac (2016), indica que la monitorización "sirve para apreciar el avance de un proyecto, para asegurarse de que éste se sitúa sobre el buen camino para alcanzar los resultados esperados, o para observar y comprender las brechas, las dificultades o incluso las nuevas oportunidades", con esta definición podemos adatar la palabra proyecto a un sistema y/o aplicativo, por tanto adaptando este concepto al presente trabajo se puede indicar que un aplicativo puede alcanzar los resultados si se mantiene monitorizado.

Cuando los aplicativos se detienen y se desconocen las causas, nace la necesidad de contar con una herramienta que ayude a identificar de manera rápida las causas del fallo para tomar decisiones inteligentes. Con un sistema de monitoreo se dará facilidades a los administradores de sistemas para que den soluciones óptimas en caso que ocurra un error en los aplicativos.

A pesar de que hoy en día existen variadas herramientas que permite la monitorización de datos capturados masivamente en logs, es importante señalar que existen algunas desventajas que presentan las herramientas, algunas son: su desarrollo requiere de mucho tiempo, la implementación es compleja, requiere mucho personal para su implementación, es costosa la implementación, entre otros.

La tecnología Elastic Stack suele ser una herramienta atractiva para monitorear el comportamiento de aplicativos a través del análisis y procesamiento de datos almacenados en logs, además que tiene la capacidad de almacenar un alto volumen de datos, siendo Elasticsearch la herramienta utilizada para el almacenamiento.

Las investigaciones planteadas anteriormente confirman que Elastic Stack es una herramienta que supera las desventajas indicadas anteriormente, además es una herramienta Big Data, fácil de usar que permite procesar diversas fuentes de datos, gestionar grandes volúmenes de líneas de logs, almacenarla en una base de datos no SQL cumpliendo con escalabilidad, seguridad, procesamiento en tiempo real.

Metodología de trabajo

La presente investigación es de tipo aplicada, porque en ella se muestra la aplicación de los conocimientos teóricos centrándose en la resolución práctica de un problema.

La metodología de trabajo se centró en las siguientes etapas:

Selección de la tecnología a implementar: Se define los parámetros a evaluar en las herramientas, se evalúa en varias herramientas y se selecciona la herramienta que más se ajuste a los parámetros seleccionado.

Comprensión de la estructura de los log: Se realiza un análisis de los logs generados por los aplicativos de la empresa, que permita identificar la estructura de cada logs y la información que se pueda extraer.

Recolección de logs: En cada servidor donde se encuentre instalado un aplicativo que requiere ser monitorizado, se debe instalar una herramienta que facilite la recolección de logs y envío al lugar donde será transformado.

Preparación y almacenamiento de datos: A cada log receptado se debe realizar una limpieza de datos, eliminando y/o agregando datos, adaptando su estructura a un formato adecuado para su almacenamiento automático y en tiempo real en la base de datos.

Visualización de datos: Se extrae datos almacenado en la base de datos y se muestra en gráficas de manera automática y en tiempo real.

Análisis de resultados: Se presenta un caso de uso real de la monitorización de los aplicativos para un lapso de tiempo determinado.

Resultados

La gestión de logs consiste en el proceso de generar, transmitir, almacenar, analizar y eliminación de datos de registro(Kent & Souppaya, 2006).

Existen diversas herramientas disponibles en el mercado tanto comerciales como de software libre para la gestión de logs. Algunas de las características consideradas para seleccionar una herramienta de gestión de logs es la capacidad de búsqueda, visibilidad en tiempo real con tableros, análisis históricos, informes, notificaciones de alertas, monitoreo del rendimiento de la aplicación y creación de perfiles y seguimiento de eventos.

Se evaluó 3 herramientas, Splunk, Graylog y Elastic Stack, seleccionando esta última como herramienta a implementar.

Elastic Stack es una colección de tres herramientas de código abierto, permite realizar un registro centralizado que ayuda a identificar los problemas con los servidores web o las aplicaciones, le permite buscar a través de todos los registros en un solo lugar e identificar los problemas que abarcan múltiples servidores al correlacionar sus registros dentro de un marco de tiempo específico (Elastic, 2019). Permitiendo recoger los logs generados por varios aplicativos alojados en distintos servidores, transformarlos de forma automática y almacenarlos en una base de datos con escalabilidad horizontal, disponibilidad y rapidez en las búsquedas de datos, permitiendo integrarse a una herramienta de visualización, mostrando tiempo real la información capturada a través de gráficos.

A modo de resumen la Figura 1 representa la arquitectura implementada:

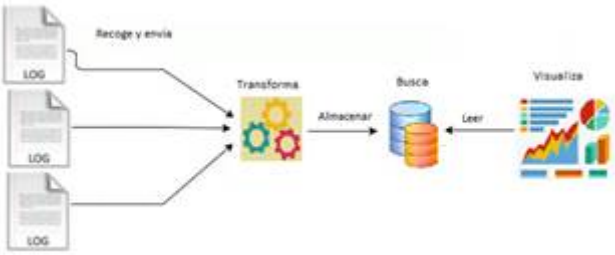


FIGURA 1.
Arquitectura propuesta.
Elastic (2019)

Para las visualizaciones se utilizó el componente de Elastic Stack llamado Kibana que ofrece una variedad de gráficas, con facilidad de configurar y ajustar a los requerimientos de la empresa.

La visualización de la propuesta presentará en forma condensada y en tiempo real información enfocada a responder a las siguientes interrogantes:

¿Cuáles son los 10 procesos más demandados a una fecha específica?

Se creó una visualización dinámica que permita seleccionar la fecha y muestre los 10 proceso más demandado entre todos los servidores.

Se creó un gráfico de barra acumulado, esta grafica en el eje X se visualiza los procesos que se están ejecutando en los aplicativos, en el eje Y las de llamadas del aplicativo, el color de las barras representa el servidor, al colocar el cursor sobre el color se muestra la información detallada. (Ver Figura 2)

FIGURA 2.
Procesos más demandados entre todos los servidor
Autores

En esta visualización se observa que el proceso 1000216 tiene mayor demanda con respecto a los otros procesos, y está siendo utilizado por el servidor de compras.

¿Cuáles son los 10 procesos más demandados en una fecha específica y en un servidor específico?

En el ERP en estudio un proceso puede ser un reporte o la ejecución de una tarea específica, cabe indicar que los logs no tenían registrado los nombres de los procesos, sino los códigos de los mismos, para esta visualización se debe seleccionar el servidor y la fecha, luego se presenta un diagrama de barra, el cual en el eje x representa los procesos y el eje Y la cantidad de solicitudes.

FIGURA 3.
Procesos más demandados por servidor
propia

La gráfica anterior se visualiza con mayor facilidad que en el servidor de contabilidad el proceso más demandado es el 183, al consultar de forma manual el ERP se determinó que corresponde al "Reporte Diario Asiento Contable", este reporte es de suma importancia ya que es reporte muy usado por todos los usuarios del departamento de contabilidad de todas las sucursales a nivel nacional.

¿Cuáles son las 10 clases java con mayor duración entre todos los servidores?

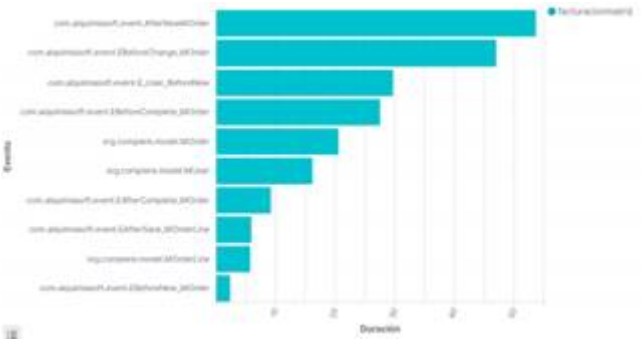
Para el cumplimiento de este requerimiento se seleccionó una gráfica de barras horizontal apiladas donde cada servidor es representado por un color, el eje "x" representa la duración en milisegundos, el eje "y"

representa los nombres de las clases Java, cada barra está ordenado por eventos más solicitados. Esto le permite al visualizador identificar con poco esfuerzo cual es la clase Java con mayor duración y cuales aplicativos lo están solicitando, los aplicativos se representa a través de una leyenda de color. Además, está configurada de manera tal, que cuando pase el mouse sobre cualquier área de las barras se muestre el nombre del servidor la duración y la clase Java, también se configuró para que permita realizar filtros por fechas y por servidor.

FIGURA 4.
Rendimiento de las clases Java entre los servidores
propia

¿Cuáles son las 10 clases java con mayor duración en un servidor específico?

Para visualizar la duración de las clases Java que tiene un aplicativo se realiza el filtro por servidor, el resultado de la búsqueda se muestra en la figura 5.



PROPIA
Rendimiento de las clases Java por servidor
propia

¿Cuáles son las tendencias de consumo de los recursos del servidor?

Se creó un dashboard el cual permite visualizar la tendencia de consumo del CPU, de memoria RAM y el tráfico de red, (Ver Figura 6)

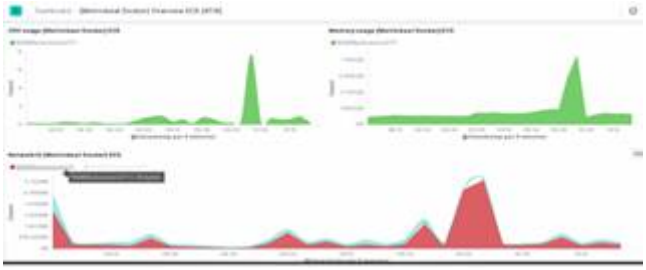


FIGURA 6.
Dashboard consumo de recurso de Servidor
propia

¿Cuáles son los SQL más demandados por servidor?

La siguiente gráfica permite visualizar el rendimiento de los SQL, se seleccionó una gráfica de líneas ya que permite visualizar con rapidez el comportamiento del SQL y determinar los puntos picos, en el eje "x" representa el tiempo y el "y" representa la cantidad de peticiones, la gráfica se configuró de manera tal que al posicionar el mouse sobre un punto se muestre la consulta SQL y el tiempo de la consulta, como se visualiza en la Figura 7.

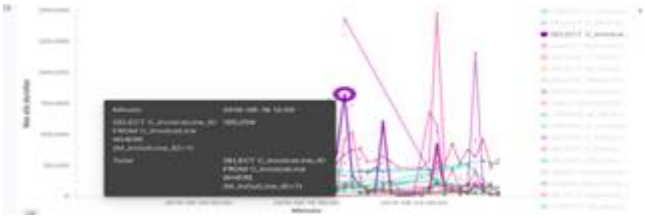


FIGURA 7.
SQL más demandados
propia

Estudio de Caso

La Figura 6 muestra el comportamiento de unos de los servidores al momento que se detuvieran los servicios, se puede ver que hubo un alto consumo de CPU, memoria Ram y tráfico de Red.

Posterior a la caída del servidor se elaboró un informe técnico con las evidencias reflejadas por varios gráficos extraídos de Elastic Stack y se remitió al área de desarrollo, donde determinaron rápidamente que el proceso correspondía a un reporte que presentaba errores de programación (redundancias cíclicas), por tanto, al invocarse el reporte consumía todos los recursos del servidor, afectando la disponibilidad del mismo.

Este caso de uso se considera exitoso ya que al igual que en párrafo anterior, la empresa en estudio pudo en repetidas oportunidades a través de la información extraída de Elastic Stack identificar varias causas que ralentizaban el desempeño del aplicativo ERP y corregir las causas de las caídas de sus servicios.

Conclusiones

En un escenario donde no se conoce con precisión el comportamiento de los sistemas críticos, es recomendable la implementación de una herramienta de monitorización de sistemas, que proporcione información relevante para la temprana detección de anomalías.

Las gráficas creadas en Kibana fueron sencillas de entender y permitieron a los desarrolladores de la empresa en estudio, identificar problemas relacionados a las caídas de los servidores, por lo que luego de su ajuste se incrementó notoriamente la disponibilidad y rendimiento de los sistemas.

En esta implementación se observó que es muy rentable usar una plataforma de monitorización sobre aplicativos críticos, ya que permite tomar medidas acertadas para aumentar la satisfacción del cliente final a través de sistemas estables, ahorrar dinero en búsqueda de errores, ahorrar tiempo de soporte y además proyectar el uso de recursos físicos de los servidores (memoria, almacenamiento, velocidad de procesamiento, entre otros).

Implementar un proyecto Big data, requiere muchos recursos de hardware, por lo que se recomienda a la empresa en estudio aumentar los recursos del servidor asignado.

Referencias Bibliográficas

Amaducci Szwarc, R. V. (2016). MONITORIZACIÓN DE SISTEMAS CON BLUEMIX. In Inuversida Autónoma de Madrid.

Andreassen, O., Charrondière, C., & De Dios Fuente, A. (2015). Monitoring Mixed-Language Applications with Elastic Search, Logstash and Kibana (ELK). 15th International Conference on Accelerator and Large Experimental Physics Control Systems (ICALEPCS 2015), WEPGF041.

Armbrust, M., Katz, R., Fox, A., Konwinski, A., Griffith, R., Joseph, A., Lee, G., Patterson, D., Rabkin, A., Stoica, I., & Zaharia, M. (2010). A view of cloud computing. Communications of the Acm, 53(4), 50–58.

Bajer, M. (2017). Building an IoT data hub with elasticsearch, Logstash and Kibana. Proceedings - 2017 5th International Conference on Future Internet of Things and Cloud Workshops, W-FiCloud 2017, 2017-Janua, 63–68. https://doi.org/10.1109/FiCloudW.2017.101

Elastic. (2019). Elastic Stack. Elastic.

González, A. C. (2015). Propuesta de Arquitectura Distribuida para la gestión de Logs. Universidad Carlos III de Madrid. https://e-archivo.uc3m.es/bitstream/handle/10016/22278/PFC_Abel_Cal_González.pdf

Hamilton, J., Gonzalez Berges, M., Tournier, J.-C., Schofield, B., Tournier, J.-C., CERN, Geneva, & Switzerland. (2018). SCADA Statistics monitoring using the elastic stack (Elasticsearch, Logstash, Kibana). 16th Int. Conf. on Accelerator and Large Experimental Control Systems, 451–455. https://doi.org/10.18429/JACoW-ICALEPCS2017-TUPHA034

Jaramillo Valbuena, S., & Londoño, J. M. (2014). Sistemas Para Almacenar Grandes Volúmenes De Datos. Revista Gti, 13(37), 17–28.

Jason, D. (2017). Monitoring with Graphite. O'Reilly Media, Inc.

Kent, K., & Souppaya, M. (2006). Guide to Computer Security Log Management. Nist Special Publication.

Langi, P. P. I., Widyawan, Warsun, N., & Aji, T. B. (2015). An evaluation of Twitter river and Logstash performances as elasticsearch inputs for social media analysis of Twitter. Proceedings of 2015 International Conference on Information and Communication Technology and Systems, ICTS 2015, 181–186. https://doi.org/10.1109/ICTS.2015.7379895

Lauriac, N. (2016). Diseño e implementación de un sistema de monitoreo. Terre Des Hommes, 46.

Ligus, S. (2013). Effective Monitoring and Alerting: For Web Operations (I. O'Reilly Media (ed.)).

Mateos Mohíno, J. C. (2017). LogsAnalyzer: Herramienta para la evaluación en tiempo real de registros log con tecnología Big Data. In UNIVERSIDAD DE CASTILLA-LA MANCHA. UCLM. (pp. 1–129).

Prakash, T., Kakkar, M., & Patel, K. (2016). Geo-identification of web users through logs using ELK stack. Proceedings of the 2016 6th International Conference - Cloud System and Big Data Engineering, Confluence 2016, 606–610. https://doi.org/10.1109/CONFLUENCE.2016.7508191