Fecha de recepción: 12 de noviembre de 2019
Fecha de aceptación: 18 de noviembre de 2019

05 [Agregación de datos para la toma de decisiones y su despliegue en servidores con diferente infraestructura]

Mgs. Galo Hurtado
Mgs. Marco Guamán

gphurtado@sudamericano.edu.ec
mguaman@sudamericano.edu.ec

RESUMEN

Las instituciones con sedes en diferentes ciudades afrontan el desafío de tener información en tiempo real debido a que cada sede tiene un sistema de información independiente; la solución en este tipo de situaciones es centralizar la información, pero esto demanda una inversión en tecnología. Para este trabajo se considera el caso de la Unidad Educativa de las Fuerzas Armadas del Ecuador, la cual cuenta con sedes en Quito, Machala y Cuenca. El objetivo de esta investigación consiste en analizar el despliegue de un entorno de agregación de datos a través de una interfaz de programación de aplicaciones (API), para lo cual se configuran dos servidores con características diferentes que permitan el despliegue de la interfaz. El primero, existente en una oficina central, y el segundo, alquilado en una plataforma en la nube; se analizan en cada servidor los tiempos de respuesta y el uso de memoria. La elección sobre la infraestructura de servidores radica en el volumen de datos. La API permite obtener los datos más relevantes de cada una de las sedes, para su posterior análisis con cualquier herramienta. Finalmente, el conocimiento generado a raíz de la centralización de datos posibilita a la oficina central de la institución tomar decisiones con un menor grado de incertidumbre.

Palabras Clave:

Agregación, Datos, Interfaz, Servidores

ABSTRACT

Institutions with headquarters in different cities face the challenge of having information in real time because each site has an independent information system; the solution in this type of situation is to centralize information, but this demands investment in technology. For this work, it is considered the case of the Educational Center of the *Armada del Ecuador*, which has offices in Quito, Machala and Cuenca. The goal of this research involves analyzing the deployment of a data aggregation environment through an application program interface (API) that configures two servers with different characteristics and allows deployment of the interface. The first one, existing in a central office, and the second one, rented in an on-line platform; response times are analyzed on each server and memory usage. The choice for server infrastructure lies in the data volume. The API allows obtaining the most relevant data from each one of the venues, for subsequent analysis with any tool. Finally, the knowledge generated by the centralization of data enables the institutional central office to make decisions with a lower degree of uncertainty.

Keywords:

Aggregation, data, interface, servers.

Introducción

En el país existen instituciones educativas públicas de niveles Básico y Bachillerato que operan en diferentes ciudades. Lastimosamente, las limitaciones económicas que se han revelado como una constante en el sector público se han convertido en carencias que limitan la operatividad de estas instituciones. Una de esas privaciones tiene que ver con el aspecto tecnológico: se carece de conectividad en tiempo real entre las diferentes sucursales de una misma institución. En este escenario, el proceso de centralización de los datos de rendimiento académico se realiza a través de intercambio de archivos, usando medios tradicionales como el correo electrónico. Por lo tanto, este proceso presenta inconvenientes como los siguientes: la incompatibilidad de formato de los archivos, datos llenados de forma incorrecta, entregas tardías... En consecuencia, analizar la información recopilada y tomar decisiones para (de ser necesario) un cambio de estrategia es difícil, puesto que la integridad de la información se ve afectada.

Tomando en cuenta este hecho, la presente investigación se refiere a la implementación de una interfaz de programación de aplicaciones (API), capaz de centralizar los datos y generar información agregación de datos de diferentes sitios, de manera efectiva (Pourghbleh y Jafari, 2017); vinculado a esto, se estudia el rendimiento de dos entornos de servidores donde se despliegan las API.

Así pues, el presente caso de estudio se desarrolla en el ámbito educativo, específicamente en la unidad educativa de las Fuerzas Armadas, que actualmente cuenta con tres sedes a nivel del Ecuador: una en Cuenca, otra en Machala y, finalmente la sede matriz, ubicada en Quito.

La oferta de estudio de esta institución va desde el nivel inicial hasta el bachillerato, y cuenta en concreto con seis mil estudiantes aproximadamente.

En cuanto al funcionamiento administrativo de la Institución es de central importancia decir que, si bien cada sucursal es considerada una institución educativa autónoma, todas están sujetas a las disposiciones de su oficina central, por lo que dicha oficina frecuentemente requiere recopilar información del historial y de la situación académica de los estudiantes. Sin embargo, los sistemas de información que guardan este tipo de información no están integrados; es decir, cada sucursal tiene su propio sistema de gestión académica; situación que impide analizar datos para la toma de decisiones y una posterior mejora en el rendimiento académico de la institución. Dicho de otro modo, tal y como se presentan las condiciones, no es posible extraer conocimiento de los datos.

La adopción de las bases de datos para la persistencia de los datos es fundamental en los sistemas de información. Y los avances tecnológicos relacionados a Internet sólo han aumentado la disponibilidad y el acceso a la información (Elmasri, 2012). La administración y el análisis adecuados de los datos permiten obtener información; en el ámbito empresarial, por ejemplo, esta información ayuda en la toma de decisiones estratégicas. Por consiguiente, el acceso y la disponibilidad de los datos se han convertido en algunos de los elementos principales para la competitividad hoy en día. De hecho, en la actualidad se obtiene información en tiempo real de varias fuentes: redes sociales, blogs o medios digitales; luego, estos datos son condensados, agrupados y filtrados por los sistemas de información de cada empresa (Grozea et al, 2017). Este proceso demuestra que, si bien los datos se encuentran separados, es posible hacerlos confluir en un mismo lugar y extraer conocimiento a partir de su análisis correspondiente.

En este contexto, es necesario determinar si una API de agregación facilita la convergencia de información para tomar decisiones en la empresa, para lo cual es preciso construir un tipo de arquitectura denominada transferencia de estado representacional (REST, por sus siglas en inglés) (Guevara et al., 2017). Esta arquitectura es un estilo de servicios web que brinda eficiencia y facilidad de uso, usando el protocolo http para obtener datos en diferentes formatos (Arsaute et al., 2018). Posteriormente, es necesario evaluar el rendimiento de los servidores en diferentes entornos tecnológicos.

En este sentido, no deja de ser importante señalar que:

- las fuentes de datos usados para la agregación tuvieron un tiempo de respuesta aceptable al momento de la alimentación de la API;
- el experimento realizado en la investigación reveló que la diferencia de respuestas de los servidores no fue significativa.
- el volumen de datos analizado fue procesado por las dos infraestructuras de servidores analizadas.

Mencionado esto, nos está demás insistir en que la API de agregación se revela como una solución pertinente para las instituciones educativas con dependencias en lugares geográficamente distantes, ya que les hace posible la recopilación de datos de diferentes fuentes y les proporciona el soporte necesario para la toma de decisiones.

Metodología

La investigación tiene un carácter correlacional que permite la medición de los tiempos de respuestas de una API sobre servidores con diferentes características de cómputo. Se analizan, así, los resultados, implementando una base de datos en cada una de las fuentes. Adicionalmente, se usa un enfoque cualitativo para el levantamiento de información.

El proceso que se realiza es el siguiente:

1

Se preparan los entornos para el despliegue de las aplicaciones (en la figura 1 se detalla el funcionamiento y el despliegue sobre los servidores). Los insumos se inician con la preparación de un entorno de servicios web de consumo o API de agregación. La primera, en un servidor instalado en una intranet, con una infraestructura básica, y ubicado en la oficina central. La segunda, en un servidor virtual privado (VPS, por sus siglas en inglés) contratado en la nube.

Figura 1. Arquitectura API. Fuente: elaboración de los autores.

2

A continuación, se configura una arquitectura de transferencia de estado representacional (REST) en cada una de las sedes, que permitirá alimentar de datos actualizados a la API de agregación que se ejecuta en la oficina central, tal como se indica en la figura 2. Estas distintas arquitecturas permiten comparar las prestaciones en distintos ambientes de ejecución.

Figura 2. Diagrama de conexiones para los colegios militares.

Fuente: elaboración de los autores.

3

Acto seguido, se despliegan cada una de las API sobre contenedores de aplicaciones, los mismos que utilizan menos recursos que la implementación de una máquina virtual (Rouse, 2015). Para el presente caso de estudio se opta por la plataforma de contenedores Docker; desde el sitio web hub.docker.com (repositorio de contenedores de aplicaciones o sistemas operativos tanto públicos como privados compatibles con Docker), se descargan los contenedores Wildfly (para publicar el servicio web) y MySQLQ (como servidor de base de datos).

Vale señalar que los servicios web publicados en los servidores de aplicación fueron desarrollados utilizando la arquitectura de JavaEE y la de REST –que son soportadas completamente por el servidor Wildfly (Costa, 2014)–. El formato de comunicación utilizado entre una API y otra es la notación de objetos de JavaScript (JSON, por sus siglas en inglés), la cual es la notación más usada actualmente para transferencias de datos en la web (Bourhis et al., 2017).

Hecho esto, los dos contenedores (Wildfly y MySQLQ) utilizados trabajan con balanceo de carga para optimizar el rendimiento, tanto del servidor VPS contratado, como del servidor local.

4

Se realizan mediciones de rendimiento de cada infraestructura. (La correlación en la investigación se halla en la comparación de los resultados de las API, tanto en un servidor de la oficina central, como en un servidor dedicado contratado para el efecto.)

5

Finalmente, se realiza una discusión en torno a la información obtenida.

Resultados

Como resultados del informe, se presenta la comparativa de rendimiento de la ejecución de la API de agregación, tanto en el servidor instalado localmente, como en el VPS contratado. Así pues, la figura 3 indica el uso de la memoria en el momento de ejecución de la API de agregación en el servidor VPS contratado; mientras que la figura 4, muestra la gráfica de uso de memoria en el servidor local.

Figura 3. Log de memoria de la ejecución de la api en el servidor VPS.

Fuente: Software de pruebas aplicaciones (SOAPUI)

Figura 4. Registro de la memoria de la ejecución de la api en el servidor de la oficina central.

Fuente: Software de pruebas aplicaciones (SOAPUI)

Ahora bien, tanto en el VPS como en el servidor local, el uso de memoria es similar debido a que la cantidad de datos es mínima en relación con la capacidad de procesamiento de los datos. Es decir, no existe mayor aplicación del cómputo masivo, por cuanto no es notable la diferencia.

También se realizaron métricas del tiempo de respuesta en la ejecución de la API para la misma petición en ambos servidores. En la figura 5 se denota que en el VPS el tiempo de respuesta fue de 5 605 milisegundos, mientras que en el servidor local fue de 8 563 milisegundos, con una carga de 3 738 150 bytes.

Figura 5. Tiempos de respuesta para la misma petición
(Servidor VPS versus local respectivamente)

Fuente: Software de pruebas aplicaciones (SOAPUI)

Características de los servidores

Se realizó una comparativa de los servidores con los que se trabaja en el desarrollo de la propuesta de solución.

El servidor usado en el entorno local es un servidor HP Proliant G5 con 10 años de uso y el de la nube es un servidor de tecnología actual disponible en las empresas de hosting, específicamente para pruebas con características limitadas de procesamiento. A continuación, se grafica una comparativa:

Figura 6. Comparativa de servidor VPS frente al servidor local

Fuente: elaboración de los autores.

Conclusiones

Las APIs de agregación permiten una fácil integración de datos de diferentes fuentes externas a una aplicación, lo que mejora la experiencia del usuario. Actualmente, tienen un mayor auge en el sector financiero.

El estudio realizado plantea la utilización de una API de agregación en el sector educativo, ya que su función de integrar datos para disponer de mayor información que facilite el análisis de estrategias académicas conjuntas en el sistema educativo representan un aporte para la toma de decisiones a nivel directivo.

A partir de este caso de estudio, se puede señalar para trabajos futuros la posibilidad de ampliar su uso a nuevos procesos académicos, generando nuevas API de agregación para recopilar información, por ejemplo, sobre los procesos de evaluación no sólo de estudiantes sino también de docentes.

Referencias bibliográficas

Arsuate, A., Zorzan, F., Daniele, M., González, A. y Frutos, M. (2018). Generación automática de API REST a partir de API Java, basada en transformación de Modelos (MDD). Workshop de Investigadores en Ciencias de la Computación, (20), 6229-633. Recuperado de: <http://sedici.unlp.edu.ar/handle/10915/67777>

Bourhis, P., Reutter, J., Suárez, F., Vrgoč, D. (2017). JSON: Data model, Query languages and Schema specification. ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database, (17). 123–135. doi: <https://doi.org/10.1145/3034786.3056120>

Costa, F. (2014). WildFly: New Features. Birmingham, UK: Packt Publishing

Grozea, C., Cercel, D., Onose, C. y Trausan-Matu, S. "Atlas: News aggregation service," 2017 16th RoEduNet Conference: Networking in Education and Research (RoEduNet), Targu Mures, 2017, pp. 1-6. doi: 10.1109/ROEDUNET.2017.8123756

Guevara, A., Juma, A., Valencia, G. (2016). APIS agregadas y cómputo masivo. Recuperado de: <https://documents.tips/engineering/apis-agregadas-computomasivo.html>

Pourghhebleh, B. y Jafari, N. (2017). Data aggregation mechanisms in the Internet of things: A systematic review of the literature and recommendations for future research. Journal of Network and Computer Applications, (97) 1, 23-24. doi: 10.1016/j.jnca.2017.08.006

Rouse, M. (2015). Application containerization [Mensaje en un blog]. Recuperado de: <https://searchitoperations.techtarget.com/definition/application-containerization-app-containerization>