

Implementación de una infraestructura para Analítica de Datos y Big Data en una fábrica de productos de belleza

Jhojan Arredondo Silva, Licenciado¹, Jhon Sovero Sovero, Licenciado²

¹TECSUP, Perú, hugo.arredondo@tecsup.edu.pe

²TECSUP, Perú, jhon.sovero@tecsup.edu.pe

Resumen– Este artículo se basa en el desarrollo de una infraestructura que permite manipular, gestionar y gestionar el conjunto de datos estructurados y no estructurados dentro de la empresa COSMETIC S.A.C. La principal conclusión de la investigación es proporcionar una infraestructura para el análisis de datos a cargo de transformarlos en valor comercial que beneficie a la empresa frente a sus competidores, así como ayudar a mitigar los riesgos, el tiempo y los gastos de los productos con los que trabaja.

Palabras claves-- Infraestructura, Big Data, Microsoft Azure, Cloud Computing, MongoDB.

Abstract– This article is based on developing an infrastructure that allows manipulating, managing and managing the set of structured and unstructured data within the company COSMETIC S.A.C. The main conclusion of the research is to provide an infrastructure for data analytics in charge of transforming them into business value that benefits the company against its competitors, as well as helping to mitigate the risks, time and expenses of the products with which it works.

Keywords-- Infrastructure, Big Data, Microsoft Azure, Cloud Computing, MongoDB.

I. INTRODUCCIÓN

El mundo se inunda de datos a un ritmo de incremento del 40% al año. [1] Esto ocasiona que, en la actualidad para tomar excelentes decisiones por parte de la gerencia de cualquier empresa, es imprescindible que tal empresa cuente con un sistema de gestión de datos de Big Data que ayude a sintetizar y analizar toda la información reunida día a día.

Big Data no solo beneficia al momento de tomar decisiones, sino también, se pueden encontrar otros beneficios tales como: Identificar las causas de fallos y problemas en tiempo real, mejorar la comprensión del potencial del marketing basado en datos, generar ofertas a clientes basadas en sus hábitos de compra, mejorar el compromiso del cliente y aumentar su lealtad, reevaluar carteras de riesgos rápidamente, personalizar la experiencia del cliente y agregar valor a las interacciones del cliente con la empresa. [2]

Por ello, en el presente artículo se expone en forma breve la implementación de una infraestructura para analítica de datos y Big Data dentro la empresa COSMETIC S.A.C. Dicha infraestructura servirá para la gestión y administración de los grandes volúmenes de información que genera diariamente y que nos son utilizados inteligentemente.

II. PLANTEAMIENTO DEL PROBLEMA

El crecimiento de los datos exponencialmente y el aumento de capacidades necesarias para su almacenamiento, gestión y procesamiento, es un problema actual al que se enfrentan muchas organizaciones y empresas, debido a que existe demasiada información acumulada de la cual no se saca mucho provecho a la hora de tomar decisiones.

En el Perú, existen organizaciones y empresas que no pueden prestar los servicios de una consultoría sobre el manejo de Big Data, que les permita escoger las herramientas adecuadas y la forma de configurar un ambiente para adaptarlo a su modelo de negocio, debido a los elevados precios al momento de pagar.

¿Qué tecnologías se deben utilizar para la construcción de una infraestructura de Big Data en la empresa Cosmetic SAC, que permita el análisis de datos en tiempo real para la toma de decisiones?

¿Qué aspectos se debe tener en cuenta para diseñar e implementar las herramientas escogidas? [3]

III. OBJETIVO

Diseñar e implementar una infraestructura para Analítica de Datos y Big Data en la empresa Cosmetic S.A.C.

IV. JUSTIFICACIÓN

Este proyecto nace con iniciativas de Big data para las empresas que cuentan con grandes recursos que requieran administración. Se pretende brindar una nueva realidad dirigida y dar otro enfoque al mercado generando nuevas oportunidades y beneficios.



Figura 1. Beneficios del Big Data

Por otra parte, el proyecto contribuye a estudiar los volúmenes de información dentro de la sociedad actual y empresarial gracias a la adecuada gestión y análisis de la información que ayuda a las empresas un mejor conocimiento de su propia organización y el entorno en donde desarrollan sus actividades para apoyar con el seguimiento y anticiparse al requerimiento de sus clientes.

V. DESARROLLO DE LA SOLUCIÓN

A. Infraestructura de Hardware

Para poder lograr la alta disponibilidad de los servicios y el mejor uso de los recursos computacionales, se ha optado primero por comparar dos alternativas principales de solución que son la Virtualización en VMware y Cloud Computing.

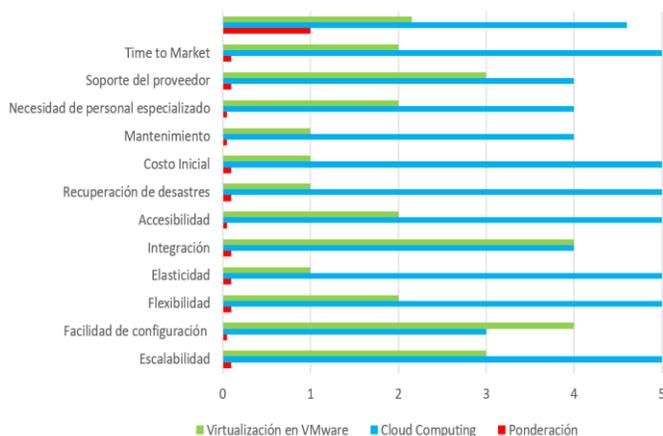


Figura 2. Comparación entre Cloud Computing y Virtualización en VMware [4]

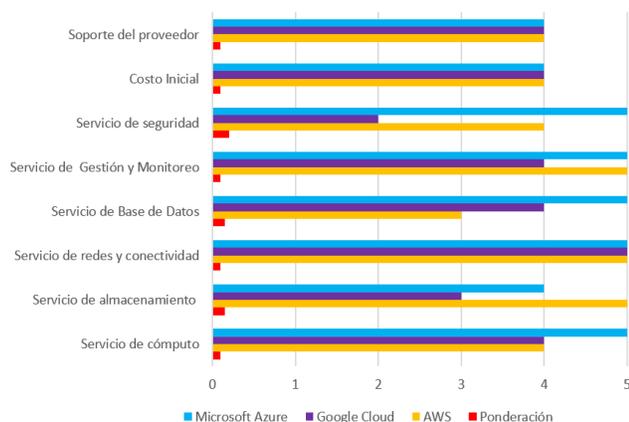


Figura 3. Comparación entre Microsoft Azure, Google Cloud y Amazon AWS [5]

En general, la virtualización y la nube operan en un modelo de uno a muchos. En pocas palabras, la virtualización puede hacer que un solo sistema funcione como diferentes sistemas, la computación en nube lo trae a todos en un lugar como un

servicio listo para usar a través de la red, lo que permite un acceso más fácil como una solución única. [6]

Por ello, La solución preferida para este proyecto en infraestructura de hardware, se eligió Cloud Computing como infraestructura de soporte para los servidores.

Por otro lado, se escogió adquirir los servicios de Microsoft frente a las otras soluciones porque, Microsoft Azure tiene una cobertura de cumplimiento normativo más completa, con más de 70 ofertas. [7]

B. Tecnología de Software

Para el funcionamiento del Big Data es necesario utilizar softwares que funcionen dentro de la infraestructura de hardware elegida anteriormente, por ello a continuación se realizaron comparación entre alternativas de solución.

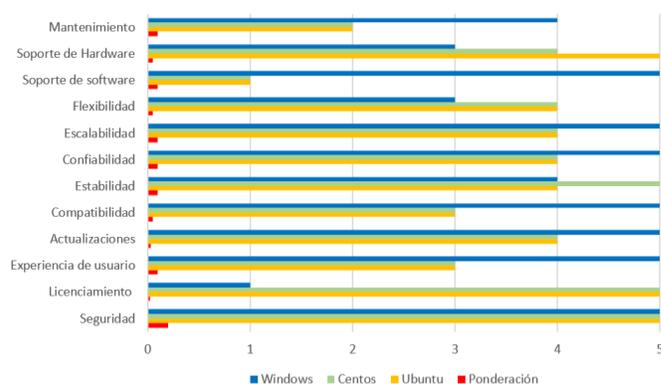


Figura 4. Comparación entre Windows, Ubuntu y Centos 7 [8]

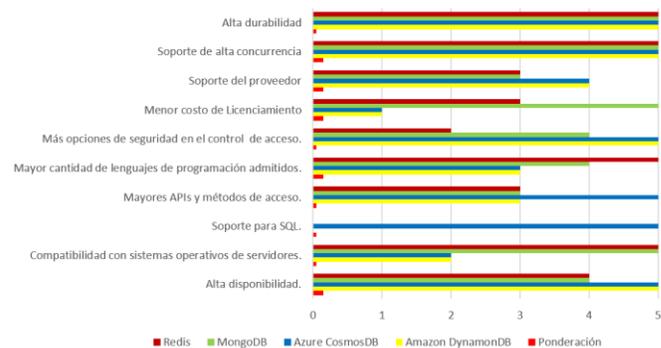


Figura 5. Comparación entre Redis, MongoDB, CosmosDB y DynamoDB [9]

La solución preferida para este proyecto en tecnologías de software, se eligió Windows como sistema operativo de los servidores puesto que más allá de los detalles, Windows ofrece los grandes beneficios de la confiabilidad y la familiaridad. [10] Por otro lado se escogió MongoDB como motor de base de datos, porque MongoDB destaca en la gestión de datos geoespaciales de gran volumen. [11]

VI. IMPLEMENTACIÓN

La aplicación será implementada según los diagramas propuestos a continuación:

Cantidad	Nombre	CAPA	Servicio	ROL
1	VM01_Webtier	WEB TIER	IaaS	Servidor web de acceso a los gerentes de la empresa COSMETIC S.A.C.
1	VM01_Managment	WEB TIER	IaaS	Servidor web de administración de la Base de Datos VM01_Datattier.
1	VM01_Apptier	APP TIER	IaaS	Servidor que procesa la lógica del Big Data.
1	VM02_Apptier	APP TIER	IaaS	Servidor que sirve para realizar una conexión segura a la base de datos VM01_Datattier.
1	VM01_Datattier	DATA TIER	IaaS	Servidor de Base de Datos MongoDB, que recolecta la información de Base de Datos externas y ERP.

Figura 6. Descripción de los servidores involucrados

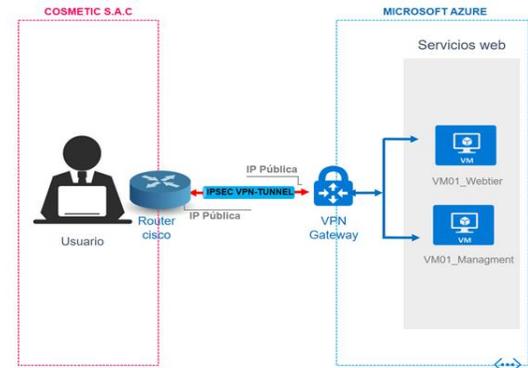


Figura 9. Diagrama de acceso del usuario



Figura 7. Diagrama de alta disponibilidad de los servidores en Azure

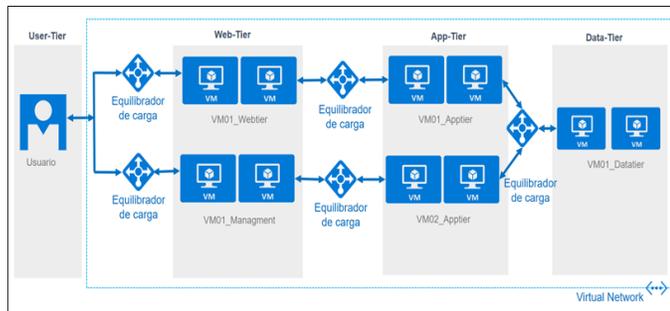


Figura 8. Diagrama de niveles de capa de la implementación

VII. RESULTADOS

El resultado mostrado en la siguiente figura nos permite comprobar la visibilidad resumida de la administración, gestión, ventas y otras tareas de la empresa COSMETIC S.A.C, mediante una interfaz gráfica que facilita la fácil comprensión del gerente o personal encargado de analizar los resultados de Big Data.



Figura 10. Interfaz gráfica de administración gerencial

No sólo debemos hacer Analítica de datos que generan las empresas de manera interna sino también explotar todos los datos que se generan en el exterior, que opinan los clientes en RRSS, qué hace la competencia, nuevas patentes y nuevos eventos. [12]

VII. CONCLUSIONES

- En este proyecto MongoDB es el principal motor de base de datos encargado de recolectar la información en tiempo real de las bases de datos externas, ya que tiene mayor compatibilidad con los sistemas operativos y menor costo de licenciamiento.
- La creación de una infraestructura de Big Data se realizó en servidores almacenados en Microsoft Azure. Gracias al fácil acceso y baja necesidad de mantenimiento.
- Para que Microsoft Azure ofrezca una alta disponibilidad de 99.95% a sus servicios se deberá activar o utilizar el Availability Set en la infraestructura implementada en la nube.

BIBLIOGRAFÍA

- [1] BBVA Open4U, «BBVAOPEN4U,» 6 junio 2016. [En línea]. Available: <https://bbvaopen4u.com/es/actualidad/infografia-big-data-presente-y-futuro>. [Último acceso: 01 julio 2018].
- [2] Corporación Colombia Digital, «Colombia Digital,» 16 mayo 2018. [En línea]. Available: <https://goo.gl/6cQZsF>. [Último acceso: 19 junio 2018].
- [3] A. González, «Repositorio Institucional Universidad Católica de Colombia,» 7 Agosto 2013. [En línea]. Available: <https://goo.gl/VDb2Uk>. [Último acceso: 6 junio 2018].
- [4] N. Bista, «EDUCBA,» 12 marzo 2018. [En línea]. Available: <https://www.educba.com/cloud-computing-vs-virtualization/>. [Último acceso: 7 junio 2018].
- [5] S. THORPE, «Caylent,» 25 enero 2018. [En línea]. Available: <https://caylent.com/aws-google-azure-cloud-comparison/>. [Último acceso: 27 junio 2018].
- [6] Veritis, «Veritis,» 19 febrero 2018. [En línea]. Available: <https://www.veritis.com/blog/virtualization-vs-cloud-computing-glance-comparison-collaboration/>. [Último acceso: 23 junio 2018].
- [7] Microsoft , «Microsoft Azure,» junio 23 2017. [En línea]. Available: <https://azure.microsoft.com/es-es/overview/azure-vs-aws/>. [Último acceso: 19 junio 2018].
- [8] Q-Success, «w3techs,» 4 junio 2018. [En línea]. Available: <https://w3techs.com/technologies/comparison/os-linux,os-windows>. [Último acceso: 20 junio 2018].
- [9] SOLID IT, «DB-ENGINES,» 10 junio 2018. [En línea]. Available: <https://goo.gl/ooV9ev>. [Último acceso: 20 junio 2018].
- [10] A. Cady, «ThinkIT by SINGLEHOP,» 24 enero 2017. [En línea]. Available: <https://www.singlehop.com/blog/linux-servers-vs-microsoft-windows-servers/>. [Último acceso: 22 junio 2018].
- [11] A. A. Morales, «MappingGIS,» 15 diciembre 2017. [En línea]. Available: <https://mappinggis.com/2014/07/mongodb-y-gis/>. [Último acceso: 21 junio 2018].
- [12] Clavei, «Clavei expertos en transformación digital,» 25 abril 2018. [En línea]. Available: <https://www.clavei.es/blog/conclusiones-sobre-big-data-y-business-analytics-de-calzatic/>. [Último acceso: 24 junio 2018].