

Tipo de artículo: Artículo original

Temática: Soluciones informáticas

Recibido: 20/07/2020 | Aceptado: 24/08/2020 | Publicado: 01/11/2020

## Modelo de implementación de inteligencia de negocios basado en sistemas de código abierto

### *Business intelligence implementation model based on open-source systems*

Enrique Condor <sup>1\*</sup>, Andy Reyes <sup>2</sup>, Alex Pacheco <sup>3</sup>, Marco de la Cruz <sup>4</sup>, Carlos Castro <sup>5</sup>

<sup>1</sup> Universidad Nacional José María Arguedas. Andahuaylas - Perú [enricoti@unajma.edu.pe](mailto:enricoti@unajma.edu.pe)

<sup>2</sup> Open Nova IT Consulting. Lima – Perú. [andy.reyes@opennova.pe](mailto:andy.reyes@opennova.pe)

<sup>3</sup> Universidad Nacional de Cañete. San Vicente de Cañete - Perú. [apacheco@undc.edu.pe](mailto:apacheco@undc.edu.pe)

<sup>4</sup> Universidad Nacional Daniel Alcides Carrión. Cerro de Pasco - Perú. [ocmar55i@gmail.com](mailto:ocmar55i@gmail.com)

<sup>5</sup> Universidad Nacional José María Arguedas. Andahuaylas - Perú [ccastro@unajma.edu.pe](mailto:ccastro@unajma.edu.pe)

\* Autor para correspondencia: [enricoti@unajma.edu.pe](mailto:enricoti@unajma.edu.pe)

---

#### Resumen

Esta investigación propone un modelo de implementación de soluciones de software libre para el análisis de datos de la organización. Este proceso de análisis se realiza sobre plataforma Linux lo cual garantiza el uso de estándares abiertos, por lo tanto, ofrece un nivel alto de integración independientemente de los fabricantes o proveedores de tecnología. El aporte principal de esta investigación es ofrecer el modelo que guíe el proceso de implementación desde el sistema operativo hasta el sistema de análisis de datos y la creación de los reportes claves que los tomadores de decisiones de la organización necesitan. Se realiza un proceso de evaluación técnico y económico para comparar funcionalidades que la organización requiere y determinar que el modelo propuesto es una alternativa viable a los sistemas propietarios que dominan el mercado.

**Palabras clave:** Pentaho; Linux; inteligencia de negocios; open-source; servicios de TI

#### Abstract

*This research proposes an implementation model of open-source solutions for the analysis of company data. This analysis process is performed on Linux platform which guarantees the use of open standards, therefore, offers a high level of integration regardless of the manufacturers or technology providers. The main contribution of this research is to offer the model that guides the implementation process from the operating system to the data analysis system until*

*the creation of the key reports that the organization's decision makers need. A technical and economic evaluation process is carried out to compare functionalities that the organization requires and determine that the proposed model is a viable alternative to proprietary systems that dominate the market.*

**Keywords:** *Pentaho; Linux; business intelligence; open-source; IT services*

## Introducción

Las empresas que generan gran cantidad de información requieren soluciones informáticas que les permitan llevar el registro de esta para luego procesarla de tal manera que estos datos puedan ser transformados en conocimiento. El conocimiento generado tiene el objetivo de facilitar la toma de mejores decisiones en el momento adecuado. Las empresas que implementan un sistema de inteligencia de negocios pueden identificar nuevas oportunidades de negocio y finalmente mejorar su desempeño.

Existen soluciones empresariales para la inteligencia de negocios en el mercado como Microsoft, esta solución cuenta según (Dutta, 2019) con gran parte del mercado mundial, las funcionalidades que ofrece esta solución son muy amplias y que ofrecen grandes ventajas como señala (Cordero & Rodríguez, 2017), sin embargo, no todas son requerimientos de las empresas que se inician en el proceso de implementación de una solución de inteligencia de negocios, es en este punto que una tecnología basada en estándares abiertos es una opción viable. Como parte de esta investigación se propone soluciones empresariales, pero de alcance libre como Pentaho en la versión comunitaria.

El modelo propuesto define una implementación sobre plataforma de virtualización considerando el principio de disponibilidad que define (Reyes, 2020) donde un sistema virtualizado con componentes redundantes en sus hipervisores hereda dicha disponibilidad a sus sistemas huéspedes. Como segundo nivel se define al sistema operativo orientado a servicios (Romero & Figueroa & Vera & Álava & Castillo, 2018), el sistema deberá ser Linux en cualquiera de sus distribuciones que sea certificado sobre la plataforma de virtualización.

Finalmente se implementa Pentaho Community sobre el sistema operativo sin embargo el modelo propone una cuarta capa de aseguramiento y fortalecimiento la cual garantiza según (Rosenthal & Zulu & Kudzai, 2016) el buen funcionamiento de la plataforma al reducir las intrusiones que conlleven a una vulneración de la plataforma.

Una vez implementada y asegurada la plataforma, el sistema contará con conexiones a las bases de datos de la organización la cual le permitirá alimentar los reportes que los tomadores de decisiones requieren. En este modelo se evalúa el proceso de implementación de esta tecnología sin llegar a ser una guía técnica ya que las opciones de sistemas operativos son diversas, en esta investigación evaluaremos el sistema de virtualización oVirt con sistema operativo CentOS como caso específico de prueba.

También se realiza una evaluación comparativa entre funcionalidades de la plataforma sobre software propietario vs software libre y determinar que la opción propuesta por el modelo es una alternativa viable. Finalmente se evalúan los precios tentativos de adquisición de ambas plataformas considerando que ante funcionalidades similares el precio es un factor determinante en la decisión de adquisición.

## Materiales y métodos

Nivel 1 – Virtualización de componentes: El modelo propuesto sugiere implementar la solución sobre plataforma de virtualización, principalmente por algunas características clave que serán heredadas al sistema de inteligencia de negocios como:

- Alta disponibilidad
- Recuperación ante desastres
- Escalabilidad

Para cumplir con estas 3 características el sistema propuesto es oVirt para mantener estándares abiertos (Poat & Lauret, 2018), sin embargo, el sistema puede ser vmware, citrix, hyper-V, Oracle VM, entre otros (Graniszewski & Arciszewski, 2016) sin ningún inconveniente, se sugiere oVirt por ser un sistema abierto y en caso de contar con la infraestructura de servidores disponibles no incurrir en el costo de licencias nuevas.

La arquitectura base como se muestra en la Figura 1 cumple con los requerimientos básicos de una plataforma de virtualización con redundancia en los servidores de virtualización y el acceso a los sistemas externos como el almacenamiento y las bases de datos consumidas por el sistema de inteligencia de negocios.

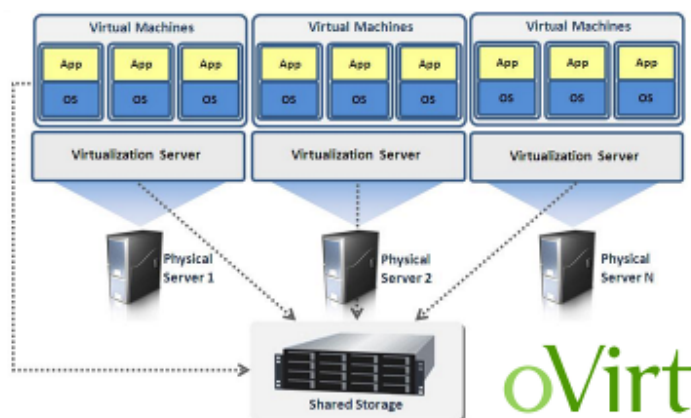


Figura 1: Plataforma de virtualización

Nivel 2 – Sistemas operativos: Para implementar el sistema operativo se debe evaluar cuales son los sistemas certificados sobre la plataforma de virtualización definido en el Nivel 1 validando no solo el sistema sino las versiones de Kernel el cual le da las funcionalidades core al sistema (Stewart, 2018), considerando que el modelo propone oVirt como plataforma de virtualización, también proponemos un sistema abierto en este nivel como el sistema operativo CentOS disponible en (CentOS, 2020), sin embargo, esto no descarta la posibilidad de usar Red Hat, Ubuntu, Oracle Linux o cualquier otro sistema empresarial con certificación vigente.

El sistema operativo deberá ser implementado siguiendo la buena práctica de despliegue, es decir no se permiten implementaciones por default que solo cumplen con un despliegue inicial, sino un despliegue orientado a servicios, en este caso al sistema de inteligencias de negocios, estos requerimientos particionamiento se muestran en la Tabla 1.

**Tabla 1: Particionamiento orientado a servicios**

VG/LV	Partición	Espacio sugerido
vg_centos/lv_root	/	20GB para perfil Minimal
vg_centos/lv_swap	<swap>	8GB
vg_centos/lv_tmp	/tmp	5 GB
vg_centos/lv_var	/var	10 GB
vg_centos/lv_log	/var/log	10 GB
vg_centos/lv_audit	/var/log/audit	5 GB
vg_datos/lv_opt	/opt	100 GB con posibilidad de ampliar
vg_backup/lv_backup	/backup	250 GB con depuración periodica

Nivel 3 – Inteligencia de negocios: En este nivel se implementa la solución Pentaho en versión comunitaria (Hitachi, 2020) se instala la plataforma sobre el sistema operativo con un dimensionamiento correcto como el definido en la Tabla 1. Los reportes claves de la solución ayudaran al tomador de decisiones a establecer las estrategias adecuadas para la empresa al contar con la información clave (Rodríguez-Cruz & Pinto, 2017) como se ve en la Figura 2 donde se resume la información.



**Figura 2: Reporte en plataforma Pentaho**

Estos reportes pueden programarse para actualizarse periódicamente según los nuevos datos de la empresa, para ello se sugiere que las conexiones a estas bases de datos estén siempre disponibles, sin embargo, estos reportes no según ANDY no se deben guardar en estas bases de datos sino en una dedicada al sistema de inteligencia de negocios.

Considerando que Pentaho comunitario es una solución basada de open-source, esta tiene la capacidad de integrarse a otras soluciones vía Rest APIs (Heredia & Sailema, 2018), es decir su información puede ser consumida por otros sistemas como se muestra en la Figura 3.

Finalmente, el sistema implementado mantiene la alta disponibilidad como lo define (Moniruzzaman & Hossain, 2014) con un tiempo de caída calculado desde el instante de la caída y el instante que el sistema operativo inicie en un hipervisor disponible, siendo este tiempo menor a 1 minuto.

Nivel 4 – Aseguramiento de plataforma: En este nivel se afina el sistema operativo en coordinación con el administrador de redes y aplicaciones, en un primer nivel se siguen algunas buenas prácticas del sistema operativo con enfoque a redes (Moreira & Alcívar, 2018) como las mostradas en la Tabla 2.

**Tabla 2: Lineamientos de implementación de redes**

Nro	Lineamiento
-----	-------------

1	Activación del sistema SELinux en modo enforcing
2	Restringir el sistema firewalld para restringir todos los puertos de gestión para permitir el acceso solo a administradores
3	Permitir el acceso público a través de firewalld a los puertos de servicios http y https pero con redirección de http a https
4	Activar el respaldo programado

En el nivel de aplicaciones se sugiere seguir los lineamientos de la Tabla 3 para asegurar la capa web como especifica (Von & Ahlborn & Mason & Mansmann, 2018).

**Tabla 3: Lineamientos de implementación de seguridad**

Nro	Lineamiento
1	Implementar certificados digitales para forzar la conexión segura
2	Activar el módulo de seguridad de apache para prevenir acceso via urls con parámetros avanzados
3	Activar la auditoria y generación de logs enfocados al acceso a la plataforma web

## Resultados y discusiones

A fin de justificar que la solución propuesta basada en estándares abiertos sea una alternativa viable, se realiza un análisis de funcionalidades básicas que una empresa que recién está iniciándose en soluciones de inteligencia de negocios requiere, como se muestra en la Tabla 4.

**Tabla 4: Comparación de funcionalidades**

Funcionalidad	Pentaho Comunitario	Microsoft BI
Multiplataforma	Si	Si
Base de conocimiento	Comunidad Pentaho	Soporte Microsoft
Reportes customizados	Si	Si
Integración via APIs	Si	Si
Conexión a bases de datos	Si	Si
Importar y exportar datos	Si	Si

Ahora se procede a realizar el análisis económico como se muestra en la Tabla 5 para 100 usuarios en el periodo de 1 año en la nube en dólares americanos.

**Tabla 5: Análisis económico en dólares americanos**

Opción	Detalle	Costo unitario	Total
Microsoft BI	Licencia mensual	10	12,000
	Profesional por usuario		
		Total	12,000
Pentaho	Sistema operativo	100	1,200
	CentOS en AWS mensual		
	Pentaho comunitario	0	0
		Total	1,200

## Conclusiones

En caso este modelo de implementación sea aplicado en una empresa pública, se mantiene los principios de multiplataforma, además de considerar que en caso las diferentes alternativas cumplen los requerimientos, el análisis económico es determinando en la adquisición.

El modelo está enfocado a empresas que inician en los sistemas de inteligencia de negocios, ya que las funcionalidades muy avanzadas que solo están disponibles en versiones empresariales de soporte comercial pagado no son necesarias los primeros años, sin embargo, el uso de estándares abiertos permite una futura migración transparente a soluciones propietarias.

La posibilidad de integración con otras plataformas a través de APIs le da gran valor a la solución propuesta ya que permite darles valor a las implementaciones previas.

Aunque el soporte es comunitario, la comunidad de Pentaho cuenta con millones de usuarios testeando y sobre todo publicando código que puede ser utilizado de forma libre, de tal manera que cuando la organización lo requiere puede pasar a la versión pagada de Pentaho y complementarlo con el soporte pagado.

## Agradecimientos

A las comunidades de software libre que invierten mucho tiempo en el desarrollo y pruebas de soluciones sin las cuales esta investigación no sería posible.

Al Ing. Andy Reyes de la empresa Open Nova por enfocarse en soluciones de software libre en Perú formando verdaderos especialistas.

## Referencias

Dutta, P. (2019). Business Analytics using Microsoft Power BI and AWS Redshift. *International Journal of Trend in Scientific Research and Development*, Volume-3(Issue-2), 984-986. <https://doi.org/10.31142/ijtsrd21545>

Cordero Guzman, D. M., & Rodríguez López, G. (2017). La inteligencia de negocios: una estrategia para la gestión de las empresas productivas. // *Business intelligence: a strategy for the management of productive enterprises*. CIENCIA UNEMI, 10(23), 40-48. <https://doi.org/10.29076/issn.2528-7737vol10iss23.2017pp40-48p>

Reyes, A. (2020). Implementing model applied to a virtualized data center based on open source projects. *Test Engineering and Managementm* Volume 83, Pages 9243-9249. <http://www.testmagazine.biz/index.php/testmagazine/article/view/5280>

Romero Castro, M. I., Figueroa Morán, G. L., Vera Navarrete, D. S., Álava Cruzatty, J. E., Parrales Anzúles, G. R., Álava Mero, C. J., ... Castillo Merino, M. A. (2018). Introducción a la seguridad informática y el análisis de vulnerabilidades. <https://doi.org/10.17993/ingytec.2018.46>

Rosenthal, R., Zulu, D., & Kudzai Chiwome. (2016). Implementing efficient security on Linux based systems in Zimbabwe: a case study on the shellshock vulnerability. Unpublished. <https://doi.org/10.13140/RG.2.1.3453.0327>

Poat, M. D., & Lauret, J. (2018). Enabling High Availability Service with oVirt Virtualization and CephFS. *Journal of Physics: Conference Series*, 1085, 32015. <https://doi.org/10.1088/1742-6596/1085/3/032015>

Graniszewski, W., & Arciszewski, A. (2016). Performance analysis of selected hypervisors (Virtual Machine Monitors - VMMs). *International Journal of Electronics and Telecommunications*, 62(3), 231-236. <https://doi.org/10.1515/eletel-2016-0031>

Stewart, K. (2018). Lessons learned from the linux kernel. *Proceedings of the 1st International Workshop on Software Health - SoHeal '18*. the 1st International Workshop. <https://doi.org/10.1145/3194124.3194132>

Centos. (200). The CentOS Project. Retrieved July 19, 2020, from <https://centos.org/>

Hitachi. (2020). Pentaho Community Edition. Retrieved July 19, 2020, from <https://community.hitachivantara.com/s/pentaho>

Rodríguez-Cruz, Y., & Pinto, M. (2017). Requerimientos informacionales para la toma de decisiones estratégicas en organizaciones de información. *Transinformação*, 29(2), 175-189. <https://doi.org/10.1590/2318-08892017000200005>

Heredia, J. S., & Sailema, G. C. (2018). Comparative Analysis for Web Applications Based on REST Services: MEAN Stack and Java EE Stack. *KnE Engineering*, 3(9), 82. <https://doi.org/10.18502/keg.v3i9.3647>

Moniruzzaman, A. B. M., & Hossain, S. A. (2014). A Low Cost Two-Tier Architecture Model for High Availability Clusters Application Load Balancing. *International Journal of Grid and Distributed Computing*, 7(1), 89-



98. <https://doi.org/10.14257/ijgdc.2014.7.1.09> Moreira Santos, M. G., & Alcívar Marcillo, P. A. (2018). Security in the data link layer of the OSI model on LANs wired Cisco. *Journal of Science and Research: Revista Ciencia E Investigación*, 3(CITT2017), 106-112. <https://doi.org/10.26910/issn.2528-8083vol3isscitt2017.2018pp106-112>

Von Bomhard, N., Ahlborn, B., Mason, C., & Mansmann, U. (2018). The Trusted Server: A secure computational environment for privacy compliant evaluations on plain personal data. *PLOS ONE*, 13(9), e0202752. <https://doi.org/10.1371/journal.pone.0202752>

Moreira Santos, M. G., & Alcívar Marcillo, P. A. (2018). Security in the data link layer of the OSI model on LANs wired Cisco. *Journal of Science and Research: Revista Ciencia E Investigación*, 3(CITT2017), 106-112. <https://doi.org/10.26910/issn.2528-8083vol3isscitt2017.2018pp106-112>

Von Bomhard, N., Ahlborn, B., Mason, C., & Mansmann, U. (2018). The Trusted Server: A secure computational environment for privacy compliant evaluations on plain personal data. *PLOS ONE*, 13(9), e0202752. <https://doi.org/10.1371/journal.pone.0202752>