

Modelo de éxito de un data warehouse

A data warehouse success model

ISABEL CANDAL VICENTE

Doctora en Administración de Empresas. Docente de la Universidad del Este SUAGM. San Juan, Puerto Rico.

Contacto: icandal@suagm.edu

Fecha de recepción: 14 de noviembre de 2011

Fecha de aceptación: 16 de octubre de 2012

Clasificación del artículo: Estudio de Caso

Financiamiento: Universidad del Este SUAGM

Palabras clave: data mart, data warehouse, economía del conocimiento, sistemas de información.

Key words: data mart, data warehouse, knowledge economy, information systems.

RESUMEN

La revisión de literatura realizada para el presente artículo, sugiere que el estudio académico sistémico relacionado con data warehouse es limitado. El propósito de este estudio es examinar la asociación entre factores de implementación, viabilidad de éxito de la implementación y viabilidad de éxito del sistema que afecta la implementación exitosa de un data warehouse. Se administró un cuestionario mediante correo electrónico a un total de 103 personas de una compañía aseguradora. La población estuvo compuesta por patrocinadores, líderes de proyecto, analistas de negocios y desarrolladores y, además, por usuarios de negocio. Para establecer la asociación entre las variables del estudio, se utilizó el coeficiente de correlación de Kendall. El éxito de los factores de la implementación organizacional y operacional se asocia significativamente a la calidad de datos, calidad de sistema y calidad del servicio.

ABSTRACT

Most authors suggest that academic studies related to data warehouse systems are limited. The purpose of this study is to examine the association between three aspects, namely implementation factors, feasibility of successful implementation and feasibility of success of the system, which affect the successful implementation of a data warehouse. A questionnaire was administered via email and questions were answered by a total of 103 people working for an insurance company. The population included sponsors, project leaders, business analysts and developers and also business users. To establish the existing relations between the variables of our study, we used the Kendall correlation coefficient. The success of the organizational and operational factors is closely associated with data quality, system quality and service quality.

* * *

1. INTRODUCCIÓN

La convergencia de las tecnologías, la disminución de barreras comerciales y políticas, el crecimiento de los avances técnicos y tecnológicos de la revolución digital, han hecho posible hacer negocios y llegar a acuerdos de manera instantánea con miles de millones de personas a lo largo y ancho del planeta. El mundo se ha conectado creando una plataforma mundial que permite competir, colaborar, compartir conocimientos y compartir el trabajo de manera histórica.

Como consecuencia de la globalización, los países desarrollados han experimentado la transformación de una economía basada en la fabricación e industria, a una economía basada en el conocimiento y en la información. Esta nueva economía se ha denominado economía del conocimiento. La economía de conocimiento se caracteriza por utilizar el conocimiento como elemento fundamental para generar valor y riqueza por medio de su transformación a información. El ambiente de negocio se ha convertido en un ambiente global, competitivo, complejo y volátil [1]. La globalización, ha alterado los patrones de competitividad, ha encontrado su principal apoyo en el conocimiento y la capacitación profesional. La nueva economía demanda la explotación de nuevos modelos y paradigmas. La economía de conocimiento se caracteriza por utilizar el conocimiento como elemento fundamental para generar valor y riqueza por medio de su transformación a información. Esta es una tendencia global y una estrategia que genera riqueza para los países que la adoptan. El objetivo de este estudio es examinar la relación entre los factores de implementación, viabilidad de éxito de la implementación y viabilidad éxito del sistema que afectan la implementación exitosa de un DW en una compañía aseguradora en Puerto Rico.

Actualmente, el ámbito de las organizaciones corporativas está haciendo frente a los retos con mayor eficacia, al manejar sus operaciones de negocio, reduciendo costes, aumentando los créditos y aprovechando al máximo las tecnologías de la información (TI). Las organizaciones corporativas han implementado proyectos de *data warehouse* (DW) para mejorar la capacidad de medir, de entender y de analizar las operaciones de negocio [2]. Inmon define el DW como “un conjunto de datos orientados a un dominio, integrado, no volátil, no varía en tiempo y ayuda a la toma de decisiones de la empresa u organización” [3]. Esto es, los datos se organizan por temas para facilitar su acceso y entendimiento por parte de los usuarios finales. Los datos almacenados deben integrarse en una estructura consistente. La información suele estructurarse en distintos niveles de detalle para adaptarse a las distintas necesidades de los usuarios. La información de un DW es para ser leída, y no modificada. La información es, por tanto, permanente. El DW se carga con los distintos valores que toma una variable en el tiempo para permitir comparaciones. Kimball y Ross establecen que el DW es la herramienta que provee el acceso a los datos corporativos u organizacionales que permite cuestionar, analizar y presentar información; donde se publican los datos usados [4].

En paralelo con la rápida adopción de los DW, ha surgido el concepto relacionado denominado *data mart*. Un data mart es un subconjunto del DW relativo a un departamento o área de negocio concreto. El data mart puede funcionar de manera autónoma o estar enlazado al DW corporativo central. Marakas presenta un diagrama del flujo de datos a través de una organización [5]. Los sistemas de herencia de la organización y los almacenes relevantes de los datos de sistemas externos, proporcionan las fuentes de datos para el

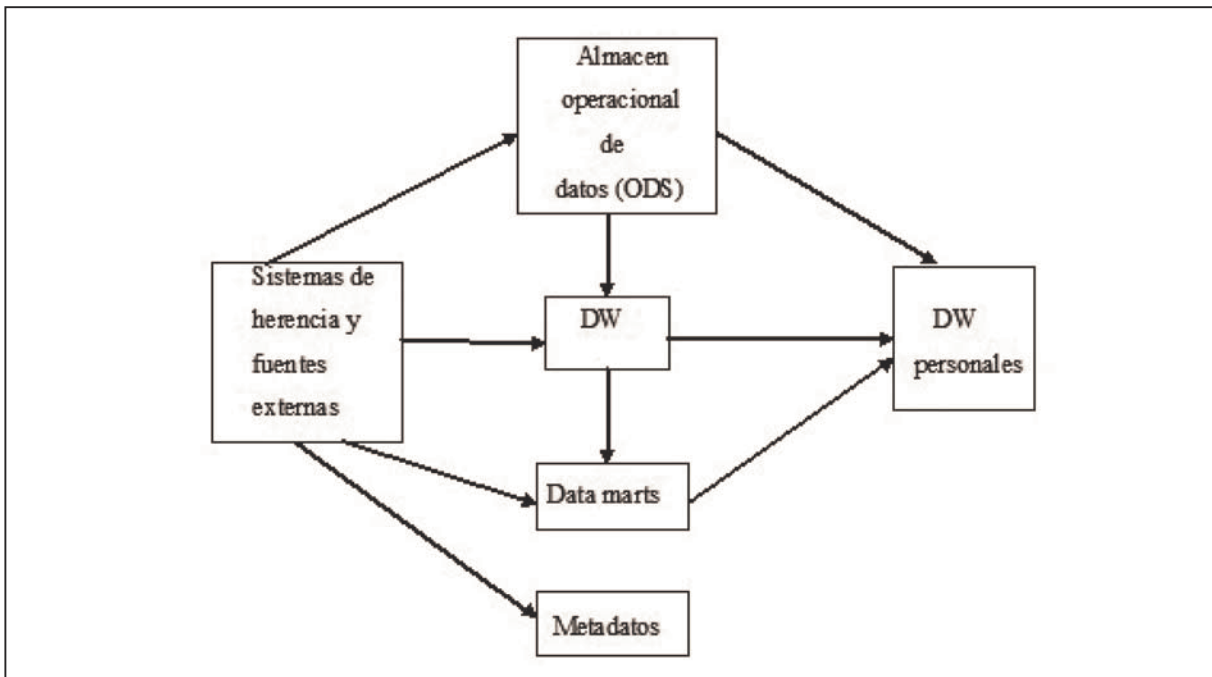


Figura 1. Componentes del flujo de datos de una organización.
Fuente: elaboración propia

DW y el data mart. Los almacenes operacionales de datos (ODS) son los que proveen el procesamiento operacional básico y pueden ser usados para alimentar el DW. Durante la transferencia de datos, ocurre un proceso de limpieza y de transformación de los mismos. Simultáneamente, los metadatos se recogen y se asocian a los datos del DW, de tal manera que los usuarios potenciales puedan determinar la fuente y las características generales de los datos del DW. Finalmente, el DW se puede emplear para crear unos o más DW personales previstos para el análisis aislado [5].

2. MODELO TEÓRICO DE INVESTIGACIÓN

Para explorar factores potenciales de éxito, se utilizó la versión revisada del modelo de éxito para sistemas de información (SI) de Delone y Mclean [6]. Un sistema de información es un conjunto de

componentes interrelacionados que acceden, procesan, almacenan y distribuyen información para apoyar la toma de decisiones, coordinación y control en una empresa [7].

La literatura sugiere que el modelo de Delone y Mclean ha sido el más influyente en investigaciones sobre factores de éxito de SI y, por lo tanto, se ha convertido en el marco conceptual para identificar factores del éxito de DW. La versión actualizada del modelo de Delone y Mclean añade una nueva métrica a las medidas de éxito: comercio electrónico. Este nuevo modelo establece seis dimensiones de éxito para SI: calidad del sistema, calidad de los datos, calidad del servicio, uso, satisfacción del usuario y beneficios netos. La calidad del sistema, calidad de los datos y la calidad del servicio afectan el uso y la satisfacción del usuario. Según Delone y Mclean, el uso y la satisfacción del usuario anteceden los beneficios netos percibidos (ver figura 2).

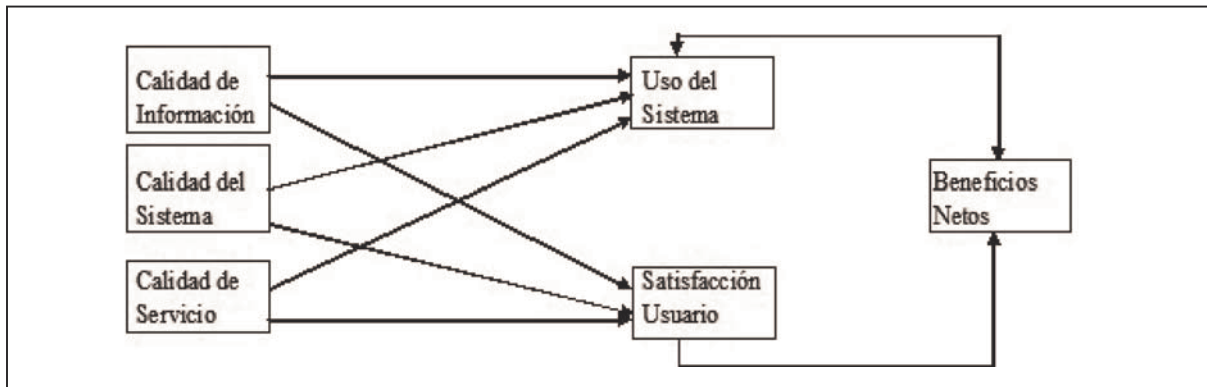


Figura 2. Modelo de éxito para sistemas de información
Fuente: tomada de Delone, W.H. & Mclean, E.R. (2003)

El modelo de investigación propuesto por este estudio utilizó, como marco conceptual, la versión actualizada del modelo de éxito para SI de Delone y Mclean, la revisión de la literatura, la de expertos en DW. Para el desarrollo del modelo teórico de investigación se utilizó, como proyecto piloto, el estudio *Factores Que Afectan la Implementación Exitosa de un Data Warehouse* [8] (ver figura 3).

3. METODOLOGÍA

Este estudio consiste de un diseño de investigación explicativo o correlacional. Su propósito es examinar la relación entre los factores de implementación, éxito de la implementación y éxito del sistema que afectan la implementación exitosa de un DW en una empresa aseguradora en Puerto

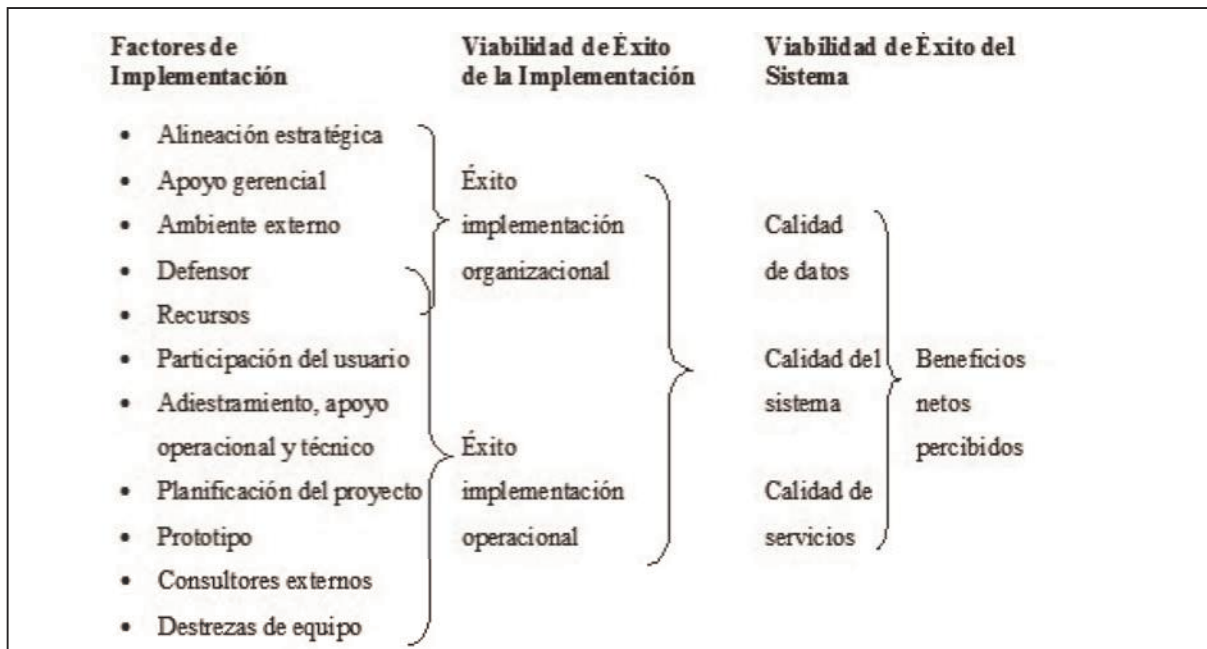


Figura 3. Modelo teórico de investigación.
Fuente: elaboración propia

Rico. Esta investigación es un estudio de caso no experimental, transeccional.

Para la recopilación de datos se utilizó como instrumento un cuestionario con dos versiones. El cuestionario se desarrolló basándose en el instrumento validado en su estudio *An empirical investigation of the factors affecting data warehousing success* de Wixom y Watson [1]. La validez de contenido se evaluó mediante consenso de expertos constituido por cuatro especialistas de TI, quienes evaluaron el instrumento de forma independiente. Una vez concluida la revisión del instrumento por el panel de expertos, se hicieron diversas modificaciones en los ítems.

La muestra se obtuvo de los participantes voluntarios que aceptaron completar y entregar el instrumento. En este caso, la elección de los individuos que son sujetos al análisis dependió de circunstancias fortuitas. La población se dividió en dos grupos. El primero estuvo compuesto por patrocinadores, líderes de proyecto, analistas de negocios y desarrolladores. El segundo, por usuarios de negocios.

Se utilizó la escala Likert de 5 puntos para medir la posición de los participantes con respecto a las afirmaciones elaboradas en el cuestionario. Se establecen niveles de actitudes desde: *muy de acuerdo, de acuerdo, indeciso, en desacuerdo y muy en desacuerdo*. Se añadieron las columnas *no sé* y *no aplica* para los participantes que no tuvieran suficiente información o ninguna información del ítem que se está midiendo. Se sometieron ambas versiones del cuestionario al departamento legal de la compañía aseguradora, la cual finalmente permitió la distribución del instrumento. La compañía aseguradora envió un correo electrónico invitando a ambos grupos a que participaran del estudio. Las personas que voluntariamente participaron, imprimieron el cuestionario, lo contestaron y lo depositaron en una urna sellada localizada en el área central del vestíbulo princi-

pal. La encuesta se administró a un total de 103 personas de una compañía aseguradora en Puerto Rico. El número total de cuestionarios contestados y recibidos fue de 63. Se recibieron 18 cuestionarios de la primera versión, administrados a los patrocinadores, líderes de proyectos, analistas de negocios y desarrolladores. Se recibieron 45 cuestionarios administrados a usuarios de negocio. La tasa de respuesta global fue de un 61%.

4. ANÁLISIS DE LOS DATOS

Se recogió evidencia de confiabilidad de la versión preparada para los usuarios de negocio mediante el cálculo del coeficiente alfa de Cronbach (μ), método para estimar la consistencia interna del cuestionario, ecuación (1) (correlación que sería obtenida entre dos formas perfectamente paralelas de pruebas si no hubiera cambios en los examinandos) [9].

$$P_{cc} \geq \frac{K}{k-1} \left(1 - \frac{\sum \sigma^2_i}{\sigma^2_c} \right) \quad (1)$$

El cálculo del coeficiente alfa para la muestra de patrocinadores, líderes de proyecto, analistas de negocio y desarrolladores fue de 726. El cálculo del coeficiente alfa para el instrumento utilizado con la muestra de usuarios de negocio fue de .841. Estos coeficientes de confiabilidad, tanto para la muestra de patrocinadores, líderes de proyecto, analistas de negocio y desarrolladores como para la muestra de usuarios de negocio son considerados altos. Nunnally, sugiere, para los tipos de investigación exploratorios, los niveles del coeficiente alfa mayores de 0.7 (considerados altos) [10]. Esto implica que existe asociación entre los ítems del instrumento con relación a los constructos que lo constituyen (consistencia interna) [10] (ver tabla 1).

Se utilizó el coeficiente de correlación de Kendall con el propósito de examinar la relación entre factores de implementación, éxito de la implemen-

Tabla 1. Análisis de confiabilidad de la prueba.

Muestra	Cronbach's Alpha	N
Patrocinadores, líderes de proyecto, analistas de negocio y desarrolladores	.726	46
Usuarios de negocio	.841	27

Nota: N es el número de cuestionarios para cada una de las muestras.

Fuente: elaboración propia

tación y éxito del sistema, que afecta la implementación exitosa de un DW [9]. El coeficiente de correlación tau de Kendall se usa cuando las variables estudiadas son medidas en escala ordinal y se ordenan por rangos o jerarquías. La fórmula es la ecuación (2):

$$T(\text{tau}) = \frac{S}{\frac{1}{2}N(N-1)} \quad (2)$$

T (tau) = coeficiente de correlación de Kendall

S = puntuación efectiva de los rangos

N = tamaño de la muestra en parejas de variables

Solamente se calcularon coeficientes de correlación para la muestra administrada a los usuarios de negocio. La muestra de patrocinadores, líderes de proyecto, analista de negocios y desarrolladores, no se utilizó debido a la homogeneidad de la muestra obtenida.

El coeficiente de correlación es un número (entre -1.00 y +1.00 inclusive) que indica el grado de la relación entre las dos variables. El signo indica la dirección de la relación. Cuando el valor es positivo, significa que existe una relación directa entre ambas variables, esto es, si las dos aumentan al mismo tiempo. El valor negativo indica que la relación es inversa, es decir, cuando una variable disminuye a medida que la otra aumenta. Cuando no hay correlación entre dos variables el coefi-

Tabla 2. Reglas para interpretar el tamaño del coeficiente de correlación.

Tamaño de la Correlación	Interpretación
.90 a 1.00 (-.90 a - 1.00)	Correlación bien alta positiva (negativa)
.70 a .90 (-.70 a -.90)	Correlación alta positiva (negativa)
.50 a .70 (-.50 a -.70)	Correlación moderada positiva (negativa)
.30 a .50 (-.30 a -.50)	Correlación baja positiva (negativa)
.00 a .30 (.00 a -.30)	Si existe correlación, es pequeña

Fuente: elaboración propia a partir de Hinkle, D.E., Wiersma, W. & Jurs, S.G. (2003)

ciente de correlación es 0. Para interpretar el nivel de asociación entre dos variables se utiliza el tamaño del coeficiente de correlación teniendo en consideración los criterios establecidos por [11] (ver tabla 2).

5. ANÁLISIS DEL TRABAJO DE INVESTIGACIÓN

Al cotejar los resultados obtenidos con los rangos de la tabla 2, se puede observar que, contrario a lo que se esperaba, la hipótesis H1a, que plantea una alta asociación entre los beneficios netos percibidos con la calidad de los datos, no es consecuente con el valor del coeficiente correlación en la asociación de estas variables, pues este fue 296. El anterior coeficiente es considerado bajo positivo, lo que implica que la asociación entre las variables anteriormente mencionadas es baja, si es que existe alguna relación, y no alta, como la presenta la hipótesis H1a. La relación es positiva entre las dos variables, es decir, mientras el valor de una variable aumenta, el valor de la otra también incrementa. La hipótesis H1b que indica una relación entre la calidad del sistema y los beneficios netos percibidos obtuvo un coeficiente

de 246. Esto implica que la asociación entre estas variables es baja. La hipótesis H1b plantea una alta asociación entre los beneficios netos percibidos con la con la calidad del sistema, esto no guarda relación con el valor del coeficiente de correlación en la asociación de estas variables. La hipótesis H1c plantea un nivel alto de calidad de servicio con un nivel alto de beneficios netos percibidos. Obtuvo un coeficiente de .270, si existe relación entre las variables. Se puede interpretar que el grado de asociación entre estas variables es pequeño positivo. La magnitud del coeficiente de correlación que asocia las variables de los factores organizacionales con calidad de datos fue de .252, lo que indica que las variables presentan una asociación pequeña. Esto implica que estos valores no concuerdan con lo indicado por las hipótesis H2a, que asocia un nivel alto de éxito en la implementación con un nivel alto de calidad de datos. La hipótesis H2b plantea una alta asociación entre los factores organizacionales con la calidad de sistema, no es cónsona con el coeficiente de correlación calculado para esas variables. El coeficiente de correlación para las variables anteriores fue de .323; lo que indica que las variables presentan una asociación baja positiva. La hipótesis H2c indica una alta asociación entre los factores organizacionales con un nivel alto de calidad de servicio está en relación con el valor del coeficiente de correlación en la asociación de estas variables. El coeficiente de correlación para estas variables resultó ser .410. Esto indica que la correlación es baja positiva. La hipótesis H3a plantea un nivel alto de éxito en la implementación de factores operacionales con un nivel alto de calidad de datos. Esto no concuerda con la magnitud del coeficiente de correlación para esas variables. Este coeficiente fue de .330. Esto implica una correlación baja positiva. La hipótesis H3b, que indica una alta asociación entre los factores operacionales con la calidad de sistema, no concuerda con la magnitud del coeficiente de correlación para esas variables. Este coeficiente fue de .289. Esto implica una correlación pequeña,

si es que existe alguna correlación. La hipótesis H3c plantea una alta asociación entre los factores operacionales con un nivel alto de calidad de servicios, lo cual no guarda relación con la magnitud del coeficiente de correlación para esas variables. Este coeficiente fue de .386 y esto implica una correlación baja positiva (ver tabla 3).

6. RESULTADOS Y RECTIFICACIÓN DE VARIABLES Y PARÁMETROS

Al cotejar los resultados obtenidos en la tabla 2, siguiendo los criterios establecidos por Hinkle para interpretar el tamaño del coeficiente de correlación, ninguna de las premisas obtuvo un tamaño de correlación de .70, o mayor, para que se interpretara como una correlación alta. Sin embargo, teniendo en cuenta la literatura existente, no se tiene evidencia de coeficientes mayores que los obtenidos.

Por otro lado, para aceptar o rechazar las hipótesis formuladas por este estudio, se utilizó la significancia estadística. La significancia estadística es el grado de riesgo asociado a no tener una certeza de 100% de que la diferencia se debe a lo que se puede pensar que se debe, pues podría deberse a un factor imprevisto. Se rechaza la hipótesis nula si el valor de la probabilidad (valor P) asociado al resultado observado es igual o menor que el nivel de significación establecido, convencionalmente 0.05 ó 0.01 [11].

Todas las hipótesis fueron aceptadas. Se encontró que los factores de éxito de la implementación organizacional y operacional se asocian significativamente a la calidad de datos, calidad de sistema y a calidad del servicio. Este estudio examinó los efectos de dos grupos en la implementación exitosa de un DW.

Investigaciones anteriores en el área de SI, han encontrado que los factores organizacionales

Tabla 3. Resultados obtenidos y análisis del trabajo de investigación.

Resultados de las Hipótesis Factores de Éxito del Sistema		Correlación Kendall	Resultado
H1a	Se asocia un nivel alto de calidad de datos con un nivel alto de beneficios netos percibidos.	.296(*)	Aceptada
H1b	Se asocia un nivel alto de calidad del sistema con un nivel alto de beneficios netos percibidos.	.246(*)	Aceptada
H1c	Se asocia un nivel alto de calidad de servicio con un nivel alto de beneficios netos percibidos.	.270(*)	Aceptada
H2a	Se asocia un nivel alto de éxito en la implementación de factores organizacionales con un nivel alto de calidad de datos.	.252(**)	Aceptada
H2b	Se asocia un nivel alto de éxito en la implementación de factores organizacionales con un nivel alto de calidad del sistema.	.323(*)	Aceptada
H2c	Se asocia un nivel alto de éxito en la implementación de factores organizacionales con un nivel alto de calidad de servicio.	.410(**)	Aceptada
H3a	Se asocia un nivel alto de éxito en la implementación de factores operacionales con un nivel alto de calidad de datos.	.330(**)	Aceptada
H3b	Se asocia un nivel alto de éxito en la implementación de factores operacionales con un nivel alto de calidad del sistema.	.289(**)	Aceptada
H3c	Se asocia un nivel alto de éxito en la implementación de factores operacionales con un nivel alto de calidad de servicios.	.386(**)	Aceptada

* Indica que la correlación es significativa al nivel de $p < 0.05$ (2-colas).

** Indica que la correlación es significativa al nivel de $p < 0.01$ (2-colas).

Fuente: elaboración propia

afectan el éxito de la implementación del DW. Esto se basa en el conocimiento de que los factores organizacionales pueden promover cambios gerenciales para ayudar a vencer la resistencia política y los problemas del cambio de gestión. Este estudio confirma, de forma empírica, este entendimiento mediante la relación positiva significativa entre los factores de organizacionales y los resultados obtenidos para la muestra de usuarios de negocio.

La Internet y el comercio electrónico están relacionados con la calidad del servicio por la demanda de los clientes para proveedor información y servicio. La medida de la calidad de servicio se

asocia generalmente a la satisfacción del cliente y a la percepción que éste tiene del servicio que se presta.

El modelo resultado del análisis incluye cinco variables independientes distribuidas en dos categorías: organizacional y operacional. La categoría de factores organizacionales está, a su vez, constituida por cuatro variables: alineación estratégica, apoyo gerencial, ambiente externo y defensor. La categoría de factores operacionales está constituida, por su parte, por dos variables: defensor y adiestramiento, apoyo operacional y técnico. Las variables dependientes son: calidad de datos, calidad del sistema y calidad de servicios (ver figura 4).

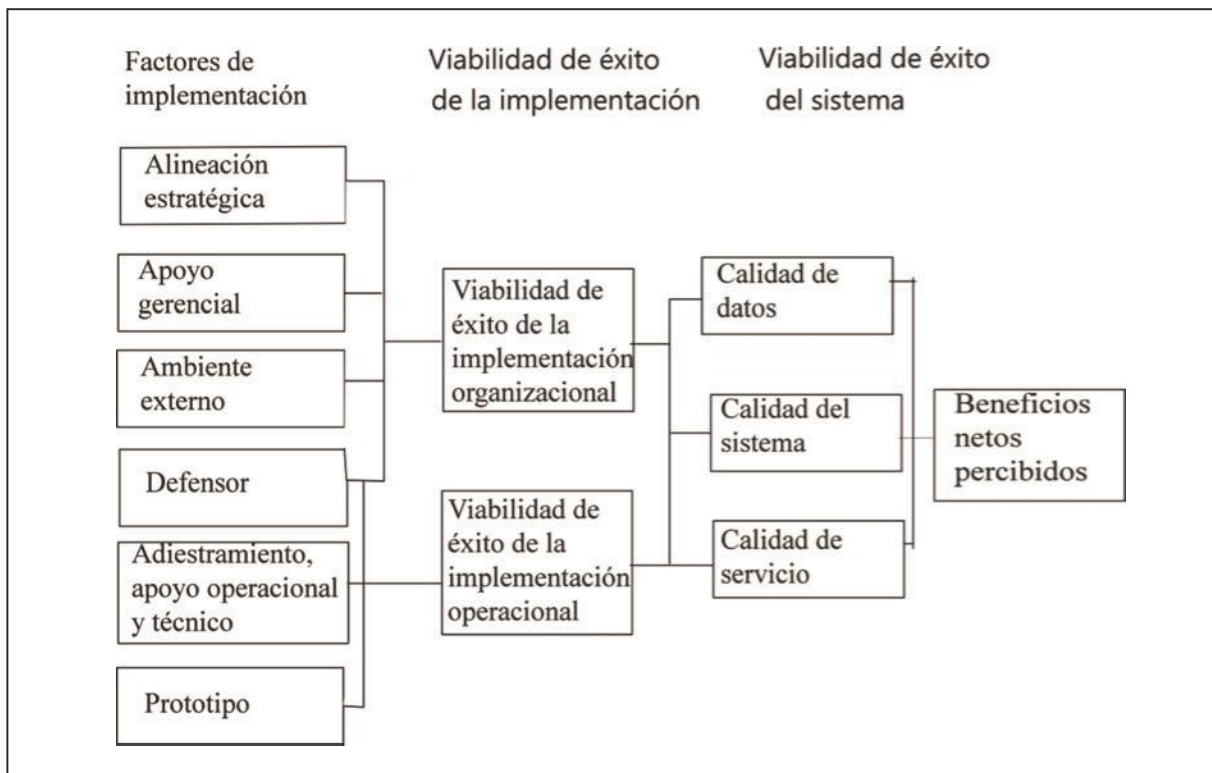


Figura 4. Modelo producto del análisis para la muestra de usuarios de negocio.

Fuente: elaboración propia

7. CONCLUSIONES

Los resultados de este estudio confirman, de forma empírica, que los factores organizacionales y operacionales afectan la viabilidad del éxito de la implementación del DW mediante la constatación del nivel de significancia que existe entre las variables. Por ende, la asociación entre los factores de viabilidad de éxito de la implementación y los factores de viabilidad de éxito del sistema es significativa. Siguiendo los criterios establecidos por Hinkle para interpretar el tamaño del coeficiente del coeficiente de correlación, ninguna de las dos muestras obtuvo un tamaño de correlación mayor que .70 [11].

Se asocian, significativamente, la calidad de datos, calidad del sistema y calidad de servicio con

los beneficios netos percibidos. La viabilidad del éxito de los factores organizacionales y operacionales se asocia a la calidad de datos, a la calidad del sistema y a la calidad de servicio.

Tomando en consideración los niveles de significancia iguales o menores de 0.05 y 0.01, se aceptan todas las hipótesis establecidas para la muestra de usuarios de negocio. Esta tesis confirma que la relación entre las variables del estudio tiene una relación significativa.

Para la muestra de usuarios de negocios, los resultados del tamaño de los coeficientes de correlación para las variables fueron: bajos o pequeños, basándose en la interpretación del tamaño del coeficiente de correlación según los criterios establecidos por Hinkle.

Referencias

- [1] B. H. Wixom and H. J. Watson, "An empirical investigation of the factors affecting data warehousing success", *MIS Quarterly*, Vol. 25, No. 1, pp 17-41, 2001.
- [2] TDWI, *Overcoming organizational challenges in data warehousing*, 2002, [Online]. Available:
<http://www.tdwi.org/Publications/BIJournal/display.aspx?ID=6514>
- [3] W. H. Inmon, *Building the data warehouse*, 4th. ed. Indiana: Wiley Computer, 2005.
- [4] R. Kimball and M. Ross, *The data warehouse toolkit*, Canada: Wiley Computer, 2002.
- [5] G. M. Marakas, *Decision support systems in the 21st century*, 2nd. ed. New Jersey: Prentice Hall, 2003.
- [6] W. H. DeLone and E. R. Mclean, "Measuring e-commerce success: Applying the DeLone & McLean information systems success model", *International Journal of Electronic Commerce*, Vol. 9, No. 1, pp. 31-47, 2004.
- [7] K. C. Laudon and J. P. Laudon, *Essentials of management information systems*, 5 th. ed. New Jersey: Prentice Hall, 2003.
- [8] I. Candal. "Factores que afectan la implementación exitosa de un data warehouse", *Vinculos*, Vol. 5, No. 1, pp.11-22, 2008.
- [9] L. Crocker and J. Algina, *Introduction to classical and modern theory*, Wadsworth, USA: Thompson/Learning, 2006.
- [10] J. C. Nunnally, *Teoría psicométrica*, Mexico: Mc Graw-Hill, 1987
- [11] D. E. Hinkle, W. Wiersma and S. G. Jurs, *Applied statistics for the behavioral sciences*, 5 th. ed. Boston: Houghton Mifflin Company, 2003