

ORGANIZACIONES Y SISTEMAS DE INFORMACIÓN

Daniel Plúa / Bernabé Ortega / Carlos Núñez
Oswaldo Villacrés / Wilson Chango / Jaime Sayago
Víctor Quiñonez / Ligia Jácome

Guillermo Pizarro Vásquez (Coordinador)

Universidad Politécnica Salesiana

Organizaciones
y Sistemas de Información

Colección de Libros de Investigación
en Ciencias de la Computación

*Daniel Plúa, Bernabé Ortega, Carlos Núñez,
Oswaldo Villacrés, Wilson Chango, Jaime Sayago,
Víctor Quiñonez, Ligia Jácome*

*Guillermo Pizarro Vásquez
(Coordinador)*

Organizaciones y Sistemas de Información

Colección de Libros de Investigación
en Ciencias de la Computación



ABYA | UNIVERSIDAD
YALA | POLITÉCNICA
SALESIANA

2019

Organizaciones y Sistemas de Información

Colección de Libros de Investigación en Ciencias de la Computación

© *Guillermo Pizarro Vásquez (Coordinador)*

*Daniel Plúa, Bernabé Ortega, Carlos Núñez, Oswaldo Villacrés,
Wilson Chango, Jaime Sayago, Víctor Quiñonez, Ligia Jácome*

1ra edición: Universidad Politécnica Salesiana
Av. Turuhuayco 3-69 y Calle Vieja
Cuenca-Ecuador
Casilla: 2074
P.B.X. (+593 7) 2050000
Fax: (+593 7) 4 088958
e-mail: rpublicas@ups.edu.ec
www.ups.edu.ec

CARRERA DE COMPUTACIÓN

Derechos de autor: 056709

Depósito legal: 006331

ISBN UPS: 978-9978-10-376-0

Diseño,
diagramación
e impresión: Editorial Universitaria Abya-Yala
Quito-Ecuador

Tiraje: 300 ejemplares

Impreso en Quito-Ecuador, julio 2019.

Publicación arbitrada de la Universidad Politécnica Salesiana

Índice

Prefacio	7
-----------------------	---

CAPÍTULO 1

Metodología para la detección temprana de problemas en sistemas de operadoras celulares	9
--	---

Fundamentación del problema de investigación.....	9
---	---

Estado del arte.....	21
----------------------	----

Metodología de Investigación	23
------------------------------------	----

Análisis e interpretación de resultados.....	23
--	----

Conclusiones y recomendaciones	30
--------------------------------------	----

Bibliografía	31
--------------------	----

CAPÍTULO 2

El flujo de información de los Sistemas de Información Académica sobre Bases de Datos Federadas	33
--	----

Fundamentación del problema	33
-----------------------------------	----

Estado del Arte	36
-----------------------	----

Metodología de la investigación	37
---------------------------------------	----

Análisis e interpretación de resultados.....	38
--	----

Conclusiones y recomendaciones	42
--------------------------------------	----

Propuesta tecnológica	44
-----------------------------	----

Conclusiones y recomendaciones de la propuesta tecnológica	77
--	----

Bibliografía	79
--------------------	----

CAPÍTULO 3

Bases de datos autónomas almacenadas en SGBD heterogéneos y su incidencia en el tiempo de obtención de la ponderación de indicadores en autoevaluación de carreras de la ESPOCH..... 81

Fundamentación del problema de investigación.....	81
Metodología de investigación.....	83
Análisis e interpretación de resultados.....	84
Conclusiones y recomendaciones	91
Propuesta	92
Metodología	106
Diagrama de casos de uso	108
Modelo conceptual	110
Diagramas de interacción.....	110
Esquema de base de datos	114
Diagrama de componentes.....	115
Conclusiones	123
Recomendaciones	123
Bibliografía	124

CAPÍTULO 4

Análisis diseño e implementación de un sistema de respaldo de datos y restauración de la información basado en software AVAMAR..... 127

Identificación del problema	127
Metodología de la investigación.....	129
Método de image backup.....	133
Análisis de resultado y propuesta.....	140
Conclusiones y recomendaciones	166
Bibliografía	169
Sobre los autores.....	173

Prefacio

Este libro es el esfuerzo de varios autores presentando resúmenes de sus trabajos de investigación sustentados como tesis de maestría, por cada capítulo.

En el primer capítulo, se hace mención a los departamentos de auditoría en empresas de telecomunicaciones que dedican mucho tiempo y esfuerzo para tratar de encontrar causas a problemas que ocurren en sus sistemas consumiendo mucho tiempo en esto y generando pérdidas económicas por la no detección. Debido a esto se desarrolla una metodología que la misma permite la construcción de alarmas o informes de manera automática así como la ayuda para determinar la causa raíz de cada problema encontrado.

El segundo capítulo, tiene como objetivo general proponer una solución de optimización en el flujo de la información académica institucional, aplicada sobre el Sistema de Información Académico Docente, donde se almacena la mayoría de registros relacionados con la generación de evidencias institucionales de acreditación universitaria, control interno de procesos académicos, investigación, vinculación con la comunidad y de gestión, con análisis de optimización del flujo de información basado en el uso del algoritmo KP.

El tercer capítulo aborda acerca de las bases de datos autónomas heterogéneas y describe un mecanismo para utilizar parcial o totalmente los datos almacenados en ellas, con el objetivo de establecer su incidencia para calcular varios indicadores del Modelo Genérico de Evaluación de Carreras, considerando como caso de estudio

las carreras de la Escuela Superior Politécnica de Chimborazo. Se presenta el desarrollo de una solución informática que interconecta bases de datos autónomas.

En el cuarto capítulo, se realiza un sistema para el respaldo de la información de ambientes virtuales y de computadores de usuarios finales a través de la red LAN, se cubrió, para este efecto, el análisis, diseño e implementación a nivel de laboratorio, de un sistema de respaldos especializado para estos ambientes. De esta manera se comprobó así sus prestaciones y su factibilidad de ejecución, desde un punto de vista técnico. AVAMAR, surge como una alternativa de solución para enfrentar los retos de aquellos ambientes de TI en donde los respaldos demandan respuestas más ágiles de recuperación.

CAPÍTULO 1

Metodología para la detección temprana de problemas en sistemas de operadoras celulares

Daniel Plúa

FUNDAMENTACIÓN DEL PROBLEMA DE INVESTIGACIÓN

Uno de los objetivos primordiales de toda operadora celular es evitar que se consuman servicios de telefonía sin que estos se facturen, así como asegurar que el cliente pueda usar dichos servicios, para ello es necesario que todas las configuraciones en las centrales y sistemas estén correctamente aprovisionadas, a esto se lo suele llamar procesos de aseguramientos de ingresos (Alexander *et al.*, 2000).

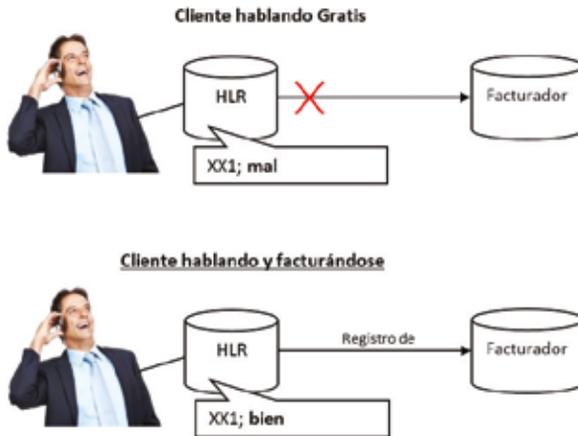
Los problemas de servicios sin facturar así como los que imposibilitan al cliente de una operadora celular poder usar dichos servicios deben ser controlados y monitoreados de manera efectiva (Mano *et al.*, 2004), para ello se deben automatizar estos controles utilizando metodologías que permitan generar alarmas automáticas e informes automáticos, y de esta manera obtener una respuesta oportuna que permita evitar futuros problemas (Madureira *et al.*, 2005; Maia *et al.*, 2003).

Estos controles deben ser capaces de corregir errores en línea sobre los clientes afectados así como detectar la causa del problema ahorrando de esta manera mucho dinero, tiempo y esfuerzo a las diferentes áreas encargadas de corregir los problemas presentados.

Definición de atributos facturadores

Se denominan atributos facturadores a aquellas características de líneas celulares que son necesarias para que exista facturación en el cliente, por ello necesitan que los valores que tienen estos atributos sean los correctos de acuerdo a lo establecido en documentación de cada proyecto tecnológico de las operadoras celulares, y que hace que el sistema le facture el o los servicios que tiene contratado el cliente de una empresa de telefonía celular.

Gráfico 1
Facturación vs. No facturación



Fuente: Plúa Morán, 2017

Elaboración: Autor

La función que realizan los atributos facturadores básicamente es enrutar el tráfico que se generan por los diferentes servicios en

telefonía celular como son: voz, mensajes escritos, mensajes multimedia, video llamada, navegación entre otros para que llegue hacia el sistema de facturación.

Los valores que poseen estos atributos para cada línea, son guardados en la base de datos de las centrales celulares que tienen el nombre de HLR; estos atributos se guardan a través de comandos generados por transacciones comúnmente desde el sistema de servicio a clientes, manejado por un asesor de servicio al cliente.

En las empresas de telecomunicaciones el área encargada de velar porque los valores de estos atributos sean correctos es decir que se encuentren correctamente aprovisionados para que no exista pérdida por no facturación de servicios es el área de aseguramiento de ingresos (Global Revenue Assurance Professional Association, 2009; Mano *et al.*, 2003).

Problemática actual en operadoras por errores de mal aprovisionamiento en central de atributos necesarios para la facturación

Existen operadoras de telefonía celular que actualmente no tienen un control robusto de características (*features*) necesarias para la correcta facturación.

Según cifras publicadas por ARAXXE, del 0.3% al 0.6% de las transacciones no se facturan o no se facturan de manera correcta al no tener controles, y esto se refleja en millonarias pérdidas para las empresas.

Este tipo de errores suelen permanecer durante meses sin detectarse, generando cuantiosas pérdidas por no facturación en operadoras celulares alrededor del mundo, y por el lado de los clientes explotan la debilidad incrementando drásticamente el tráfico sin facturar, lo que da como consecuencia el agravante de un crecimiento exponencial del tráfico que no se factura.

Gráfico 2
Tráfico normal vs. Tráfico con problemas



Fuente: Plúa Morán, 2017

Elaboración: Autor

La falta de controles robustos para ejercer un control eficiente sobre el correcto aprovisionamiento de estos atributos, genera más de un problema en las operadoras telefónicas, generando millones de dólares en pérdidas, ocasionando:

- Pérdidas por no facturación a clientes.
- Mala imagen de la empresa, ya que el cliente percibe la falta de calidad.
- Pagos a carriers internacionales que no serán cobrados.
- Fraudes por bypass.
- Llamado de atención a gerencias responsables del proceso y gerencias del departamento de aseguramientos de ingresos.

Pérdidas por no facturación a clientes

Cuando un atributo responsable de la facturación no está correctamente aprovisionado el CDR (Registro de detalle de llamada Call Detail Record) de la llamada no se envía hacia el sistema factu-

rador; como resultado de esto el evento que realiza el cliente no se cobra (Khairul Nizam Baharim *et al.*, 2008).

Gráfico 3
Llamadas sin facturar



Fuente: Plúa Morán, 2017
Elaboración: Autor

Pagos a carriers internacionales que no serán cobrados a clientes

Cuando un cliente puede generar eventos sin ser facturados también podría hacer llamadas internacionales o enviar mensajes de texto hacia países extranjeros, estos tipos de eventos generan pagos a carriers internacionales por utilizar este servicio. Es decir que estos valores que se tiene que pagar a los carriers, la operadora lo debe realizar sin recibir ningún ingreso por cobro a los clientes que utilizaron los servicios para llamadas al extranjero.

Gráfico 4
Pago a carrier por evento que no se factura



Fuente: Plúa Morán, 2017
Elaboración: Autor

Fraudes por mafias de bypass

Los mafias dedicadas al bypass en telecomunicaciones entre sus actividades tienen la de detectar errores en el aprovisionamiento de atributos al momento de activar algún servicio celular, la finalidad es evitar la facturación de los servicios (Xerandy *et al.*, 2006, Meza, 2008).

En una encuesta realizada por la CFCA a diversas compañías en 26 países, el 80% de las compañías contestaron que las pérdidas tipo bypass se habían incrementado y el 45% contestaron que habían tendido a incrementar el fraude dentro de ellas.

Gráfico 5
Esquema usado para el bypass telefónico



Fuente: Plúa Morán, 2017

Elaboración: Autor

Cuando los bypass detectan un error en el aprovisionamiento que no les generó cobro por los eventos, generalmente realizan los siguientes:

- Montan infraestructura para enrutar tráfico local hacia destinos internacionales, la mayoría de veces usando el internet (Ingeniería en procesamiento digital, 2013).
- Compran una cantidad considerable de líneas a nombre de la misma empresa.
- El dueño de la empresa que compra las líneas es ficticio, utilizando documentos falsos.
- La dirección a donde llegan las facturas son lugares deshabitados.
- Le cobran a clientes un valor mucho menor por llamadas al extranjero y todo este tráfico hacia el exterior lo pasan a través de las líneas con atributos mal aprovisionados.
- Este grupo de líneas genera un volumen muy grande de llamadas.

Llamados de atención a gerencia

Cuando los indicadores de tráfico normal en algún servicio se elevan o disminuye el ingreso percibido por algún servicio que la operadora presta, inmediatamente la gerencia de financiero, legal, técnico o incluso presidencia buscan al responsable de que no existan fugas de ingreso, esto es la “Gerencia de Aseguramiento de Ingresos”, que debe indicar porqué no se controla algún producto o que se está escapando de realizar un control eficiente, y dar la cara frente a alguna falta de control, incluso frente a un llamado de atención del órgano regulador.

Flujos de aprovisionamiento de las características (*features*) de facturadores

Las causas del mal aprovisionamiento de los atributos ocurren en un punto dentro del flujo por donde viaja la transacción hasta llegar al HLR (la base de datos de la central celular).

Gráfico 6 Gerencia siendo objeto de llamada la atención



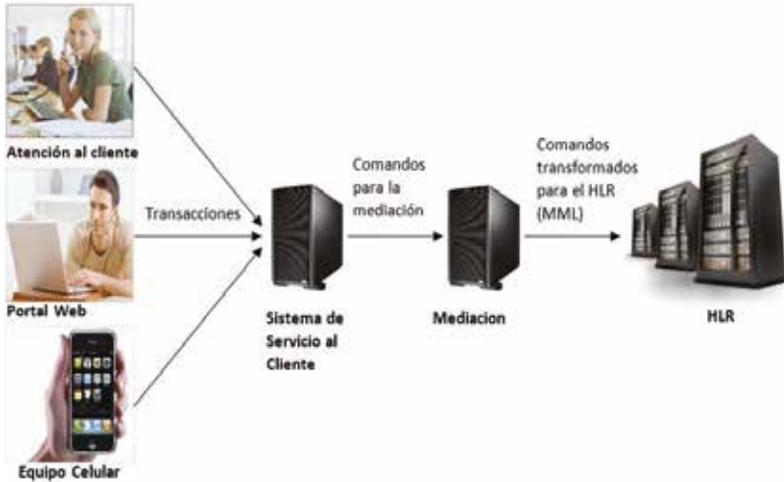
Fuente: Plúa Morán, 2017

Elaboración: Autor

La siguiente es una descripción del flujo por donde viaja la transacción:

- El proceso se inicia en la generación una transacción cuyo objetivo es el de modificar el valor del atributo de la línea, esta transacción por lo general está relacionada con algún cambio de plan, colocar algún servicio como video llamada, navegación, entre otros. Esta transacción se genera desde:
 - Asesor de servicios al cliente.
 - El cliente mismo desde el portal web de la operadora.
 - Desde su propio equipo podría lanzar transacciones.
- Esta transacción va hacia la plataforma mediación que es la que traduce el comando a el lenguaje que entiende el HLR.
- Finalmente llega la transacción al HLR donde se registra el atributo en esta base de datos.

Gráfico 7
Flujo de transacción para aprovisionar un atributo en Central



Fuente: Plúa Morán, 2017

Elaboración: Autor

Los problemas que suelen ocurrir en el flujo son los siguientes:

- Que la transacción salga errónea desde su origen, por ejemplo una mala programación en del sistema, que en primera instancia genera la transacción.
- Que la transacción salga errónea desde el sistema mediador (convierte comandos del sistema de servicio al cliente a comandos que entiende el HLR), por mala programación en la lógica de medicación o en la conversión de comandos que realiza este sistema.
- Que ocurra un error en la comunicación desde el Sistema de origen hasta el sistema mediador, puede ser errores de pérdida de conexión.
- Que ocurra errores en la comunicación desde del Sistema de mediación al HLR.

Esquemas de automatizaciones usados

Los esquemas de automatización para el aprovisionamiento de las características (*features*) del facturador es el flujo por el cual viaja la transacción de aprovisionamiento que modifica el atributo facturador de la línea, se diferencian entre sí por el origen, este origen de la transacción puede variar dependiendo los avances tecnológicos que existan, se citarán los principales.

Los atributos son aprovisionados siguiendo algún flujo de aprovisionamiento automático que varía en el origen de la transacción:

- Originado desde el sistema de servicio al cliente.
- Originado desde la página Web de la operadora celular.
- Originado desde el teléfono celular del usuario.
- Originado por una persona dentro de la empresa.
- Originado por sistemas automáticos dentro de la misma empresa por ejemplo relojes de cobranzas.

Gráfico 8

Flujo originado desde el sistema de servicio al cliente



Fuente: Plúa Morán, 2017

Elaboración: Autor

Originado desde el sistema de servicio al cliente

Es el flujo de aprovisionamiento más común y es aquel que se origina por una transacción que realiza un asesor de servicio al cliente desde el sistema de servicio al cliente. Por lo general son transacciones puntuales y queda registro de la transacción ejecutada así como el usuario que la ejecutó. La transacción se envía hacia el sistema de mediación y luego hacia el HLR.

Originado desde la página Web de la operadora celular

Ciertas operadoras permiten a sus clientes activar o desactivar atributos desde la página web con el objeto de disminuir la carga operativa de sus asesores, el riesgo de esto es que el cliente podría descubrir alguna debilidad en el sistema que le permita hacer uso de servicio de manera gratuita y explotarlos.

Gráfico 9
Flujo originado desde la página Web de la operadora Celular



Fuente: Plúa Morán, 2017

Elaboración: Autor

Originado por una persona dentro de la empresa

En ocasiones existen transacciones que no son originadas por ningún sistema sino por empleados internos de la empresa, a menudo personal de áreas técnicas, estas transacciones pueden ser:

- De aprovisionamientos masivos realizados por el área de Sistemas.
- Por un usuario no autorizado.

DE APROVISIONAMIENTOS MASIVOS REALIZADOS POR EL ÁREA DE SISTEMAS

Existen transacciones que se deben realizar de manera masiva a un conjunto grande de clientes por algún cambio que venga del área estratégica de la empresa para estos clientes. El área de sistemas debe ejecutar de manera masiva las transacciones, estas transacciones nacen de medicación y como resultado gran número de clientes tendrán modificaciones en sus atributos en la base HLR.

Gráfico 10
Transacciones generadas masivamente



Fuente: Plúa Morán, 2017

Elaboración: Autor

POR UN USUARIO NO AUTORIZADO

Ciertos usuarios realizan transacciones sobre líneas para modificar sus atributos, las empresas deben tener políticas bien definidas acerca de este tipo de transacciones y con todas las aprobaciones del caso, de lo contrario se trata de una transacción no autorizada que puede dejar mal aprovisionado algún atributo y que la línea genere tráfico sin que se facture.

Gráfico 11
Acceso no autorizado a los sistemas



Fuente: Plúa Morán, 2017

Elaboración: Autor

ESTADO DEL ARTE

A continuación, un análisis acerca de investigaciones previas sobre el problema de investigación.

Investigaciones previas acerca del problema de investigación

A continuación, una revisión bibliográfica de algunos libros científicos relacionados al problema de investigación que se propone:

Plan de predicción de problemas en sistemas

Es posible asignar una medición cualitativa a la probabilidad de ocurrencia de cualquier tipo de amenaza, y en esta escala la probabilidad se puede expresar como despreciable, baja, media y alta.

La política que seguirá la empresa u organización respecto al riesgo implica la actitud que se va a tomar una vez que se ha analizado la probabilidad de los riesgos. Esta actitud no siempre consiste en eliminar por completo todos los riesgos, eso es ideal pero muy costoso. Un riesgo se puede tomar, disminuir su probabilidad de ocurrencia o eliminar, hay riesgos tanto físico como lógicos que deben eliminarse casi a cualquier precio; no obstante, es la dirección general de la empresa la que decide cuál debe ser la acción correcta a seguir.

Debido a la naturaleza de la informática, que con seguridad es la ciencia que tiene más dinamismo en su evolución y, por tanto, en sus cambios e innovaciones, aunque se inviertan muchos recursos para eliminar en su totalidad algunos riesgos, siempre quedará un riesgo residual, que en el caso de la informática es un riesgo desconocido (Baca, 2016, p. 256).

Plan de corrección de problemas en sistemas

Debido a la naturaleza de la informática, que con seguridad es la ciencia que tiene más dinamismo en su evolución y, por tanto, en sus cambios e innovaciones, aunque se inviertan muchos recursos para eliminar en su totalidad algunos riesgos, siempre quedará un riesgo residual, que en el caso de la informática es un riesgo desconocido.

Dentro del plan general de contingencia, también se contempla el tercer sub-plan o plan de corrección, también llamado plan de continuidad del negocio, en el cual se define la forma en que la organización se recuperará ante el caso de un desastre informático. El plan se determina con base en los resultados de los análisis de la evaluación de riesgos, al determinar cuáles riesgos son de consecuencias

catastróficas, y de impacto en el negocio, en caso de suceder; en general, contempla ubicaciones alternativas, opciones para recuperación de datos, recuperación de recursos humanos, comunicaciones, equipamiento, gestión de proveedores, etcétera, así como todo tipo de actividades que en verdad sean críticas para que el negocio pueda volver a funcionar de manera normal (Baca, 2016, p. 259).

METODOLOGÍA DE INVESTIGACIÓN

Para la estrategia de mejora continua se utiliza el ciclo de Deming o más conocido como círculo PDCA por sus siglas. El ciclo Deming funciona como una rueda en constante movimiento, donde cada una de sus fases alimenta a la siguiente, por lo que la rueda se mueve de forma natural (García, Quispe, & Ráez, 2003).

Las fases/acciones son las siguientes:

- **Plan:** Aquí se define el estado actual de cómo se realizan actualmente la detección y corrección de errores en los sistemas, los tipos de sistemas existentes, su estructura, es decir se analiza la situación actual.
- **Do:** En esta etapa se seleccionan los indicadores que van a evaluar la eficiencia y eficacia del método a implementar.
- **Check:** Es crucial monitorizar la implementación del SGSI para determinar su eficacia y cumplimiento, siendo necesario un procedimiento. Esta fase se ejecuta mínimo una vez a la semana interviniendo el Comité de Dirección en la parte estratégica.
- **Act:** A partir de la revisión del SGSI, todos los registros serán usados para determinar e implementar planes de mejora al método (Ian Gregory *et al.*, 2011).

ANÁLISIS E INTERPRETACIÓN DE RESULTADOS

La recreación de casos trata de reproducir el escenario problema que genera una transacción, una vez que determinada cuál fue la

transacción que genera la inconsistencia, se ha obtenido la causa que genera estos problemas.

Es posible obtener un método que permita obtener conocimiento acerca de qué transacciones son las potenciales generadoras de inconsistencias para los distintos escenarios, basados en la cantidad de ocurrencias de las mismas. Conforme se vaya ejecutando el control y a mayor número de veces ejecutado se tendrá un conocimiento más preciso, que permitirá obtener la transacción que generó la inconsistencia dado el escenario del problema. Es decir, mientras más veces se haya ejecutado más experiencia adquirirá la base de conocimientos y el resultado será más aproximado al real.

Para la realización de la recreación del escenario de inconsistencia se debe tener:

- Chip de prueba.
- Acceso a las APIS del sistema de servicio al cliente.

Método para aprender la causa que genera el escenario X:

1. Tomar como entradas al conjunto de posibles transacciones ejecutadas desde el sistema de servicio al cliente ejecutadas desde la última corrida del control, esto se lo obtiene del archivo de LOGs.

2. Asignarle el peso respectivo a cada entrada haciendo uso de la base de conocimientos, mediante la fórmula:

$$P_n \equiv \frac{C_n}{\sum_{i=1}^N C_i}$$

Dónde:

P_n : Es el peso asociado a cada transacción de entrada

C_n : Es el contador, que se incrementa cada vez que la transacción genera la inconsistencia

N : Cantidad de transacciones encontradas en el Log del sistema de servicio al cliente

3. Encontrar transacción que tenga el mayor peso.

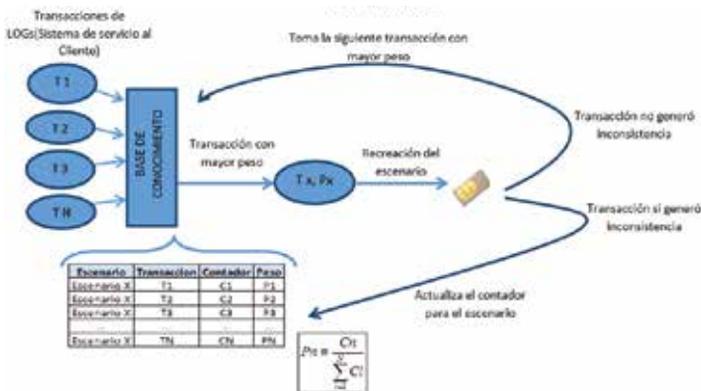
4. Probar la transacción con un número de prueba haciendo uso de las APIs del sistema de servicio al cliente.

5. Si se produjo la inconsistencia esperada se debe incrementar en 1 el contador de la transacción en la base de conocimientos y se termina el proceso.

6. Si no se produjo la inconsistencia esperada, se debe tomar el siguiente peso mayor, y repetir los pasos desde el 4 en adelante.

7. En caso de que ninguna transacción generó la inconsistencia, el problema no es por error en un proceso de sistemas sino un error operativo por parte de áreas de conciliación, activaciones masivas, activaciones no autorizadas, entre otras.

Gráfico 12
Método de aprendizaje para la búsqueda de transacción generadora de inconsistencia



Fuente: Plúa Morán, 2017

Elaboración: Autor

Se debe generar una bitácora de todas las transacciones que se prueban en la línea de prueba y enviarlos como un anexo en el informe. En el caso de que no se genere ningún error se envía lo ejecutado para que el área de Sistemas continúe el análisis de los casos.

En cada conclusión relacionada al escenario se debe citar la transacción que causa el error. De esta manera el informe tiene más datos que aportan a la pronta solución del problema ahorrando tiempo y dinero a la empresa.

Extracción y análisis de Logs

Cuando ocurre un problema o una falla en alguna plataforma la manera de encontrar las causas de la misma es mediante el uso y análisis de los archivos Logs generados por cada una de ellas.

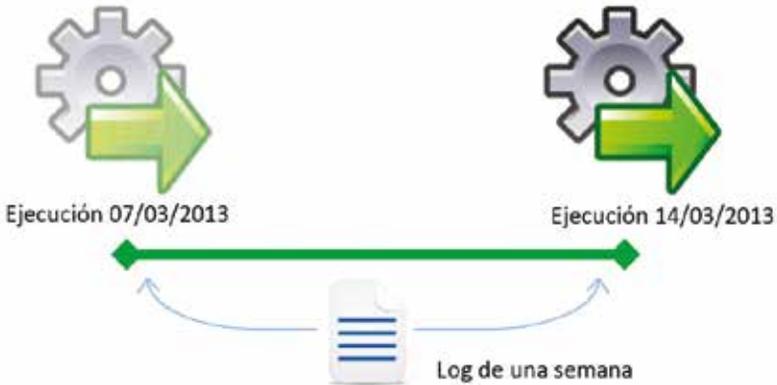
Se debe tener acceso a los Logs de todas las plataformas involucradas en el control, para el caso de informes automáticos de lo que puede tener de manera diaria. Para el caso de alarmas se debe tener acceso al sistema en línea o a una réplica con un retraso muy bajo.

Dada la frecuencia establecida para cada control, la extracción de los datos del Log debe ser a partir de la última ejecución del control hasta la presente ejecución, por ejemplo un control de frecuencia semanal ejecutado el día lunes 14 de Marzo a las 01h00, requerirá la extracción de los registros del archivo log correspondientes a la última semana, es decir desde el 07 de Marzo a las 01h00 al 14 de Marzo a las 01h00.

Con ayuda de estos Logs se puede llegar a determinar qué transacción fue la que generó la inconsistencia. O también si la transacción se ejecutó con algún tipo de error.

Esto es de gran ayuda para las áreas encargadas de corregir el origen del problema, de esta manera pueden resolverlo de manera más rápida y eficiente.

Gráfico 13
Extracción de Log



Fuente: Plúa Morán, 2017

Elaboración: Autor

Las centrales no soportan tener guardado Logs de mucho periodo de tiempo por el volumen de transacciones, habitualmente se suelen almacenar en la plataforma alrededor de un mes. Debido a esto lo que se recomienda es que en el momento de detectarse la inconsistencia extraer los Logs de la o las plataformas involucradas.

Investigación de hechos históricos que desencadenaron el problema

Se debe tener una base sólida y bien estructurada donde se encuentren todos los proyectos o cambios realizados en los sistemas o plataformas de la empresa, en particular en el sistema de servicio al cliente y en la central.

Una vez detectado el error se recurre a esta base para buscar el posible evento que desencadenó el problema, para ello se debe hacer una búsqueda teniendo filtros que involucren:

- Tipo de plan.
- Tipo de producto (Prepago, Autocontrol, Tarifario u otros).
- Tipo de servicio afectado (Voz, SMS, GPRS, entre otros).
- Fecha en que se puso en sistemas de producción el proyecto u otros campos.

Se debe hacer uso de los datos que se obtienen a partir de las inconsistencias encontradas, con estos datos efectuar la búsqueda en la base de proyectos, de esta manera se puede obtener en el mejor de los casos el proyecto o cambio que generó exactamente el problema o en su defecto un pequeño número de proyectos posibles generadores de causas dependiendo del nivel de afinamiento que se aplique en el filtro.

Alarmas automáticas

Este tipo de controles se utiliza cuando se quiere controlar un tipo de inconsistencia sencilla en particular y además se cuenta con datos en línea por ejemplo: escenario de no facturación por falta de un parámetro en central y además se cuenta con datos cuya frecuencia de actualización es cada hora. Cuando se tiene esto lo óptimo es crear una alarma, que consiste en la implementación de un control sencillo, corto y de fácil comprensión.

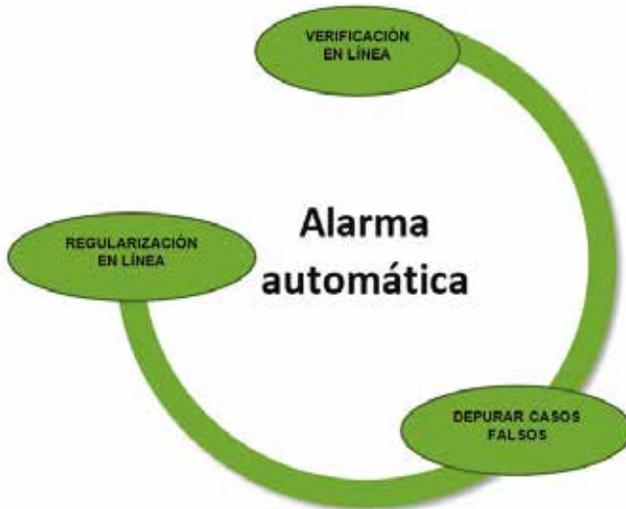
Estas alarmas se las deben enviar al área responsable de la regularización para que se proceda con la corrección de los casos.

Para la correcta realización de una alarma se debe desarrollar los siguientes tres pasos:

- Realizar una verificación en línea hacia los sistemas en producción de los casos.
- Depurar (eliminar) posibles casos falsos.
- Regularizar (corregir) los casos que proceden para la corrección.

Realizar una verificación en línea hacia los sistemas en producción de los casos. Para realizar este tipo de verificación en línea hacia los sistemas en producción se debe contar con el acceso respectivo hacia los sistemas, ya sean las centrales, el sistema de servicio al cliente y otro sistema.

Gráfico 14
Elementos de una alarma automática



Fuente: Plúa Morán, 2017

Elaboración: Autor

Se debe tener en cuenta que las consultas deben ser solo de los casos que resultan inconsistentes, se debe tener especial cuidado cuando se tiene un número exageradamente grande a consultar (>1000 casos), ya que por tratarse de sistemas en producción esto podría bajar el rendimiento de la red. Por esta razón cuando se tienen esos casos es preferible no hacer este tipo de verificaciones.

En el caso de realizar la verificación lo que se realiza es verificar si los datos que intervienen en la lógica del negocio tienen o no los valores correctos.

Depurar (eliminar) posibles casos falsos. Una vez realizada la verificación en línea se procede a eliminar los casos cuyos campos o atributos ya que se encuentran con los valores correctos de acuerdo a la lógica del negocio, es decir que se encuentran consistentes. Esta verificación se la realiza previo a la regularización en línea para asegurarnos que los casos que se envían a corregir sean exclusivamente datos inconsistentes.

Regularizar (corregir) los casos que proceden para la corrección. Se trata de enviar comandos a los sistemas que se encuentren en producción para la corrección de casos debidamente verificados, una buena práctica en este punto es la revisión de los logs o respuesta del sistema de que efectivamente el cambio se realizó con éxito antes de avanzar a la siguiente ejecución del sistema, para asegurarnos que realmente se ejecutó y se corrigió el caso. Se debe tener acceso exclusivamente a los comandos para realizar esta operación, una vez enviados los comandos se procede de igual manera a enviar la alarma de notificación a las áreas relacionadas al error, para que se corrija la causa aunque el caso ya haya sido modificado por la alarma online.

CONCLUSIONES Y RECOMENDACIONES

Conclusiones

- La metodología fue desarrollada y funciona bajo cualquier lenguaje de programación de alto nivel, la misma que ayuda de sobremanera al trabajo del departamento de aseguramiento de ingresos.
- Esta metodología fue diseñada para la automatización de controles, y se determina que puede dividirse en dos tipos de controles ya sea alarmas para controles sencillos o informes propiamente para el caso de controles complejos.
- El algoritmo fue diseñado para determinar la causa raíz de cada problema encontrado y mantener una base de cono-

cimientos que se utiliza para determinar las causas de los problemas de manera más rápida.

- Estas técnicas fueron diseñadas para la prevención de futuros errores y evitar que problemas del mismo tipo se vuelvan a presentar.

Recomendaciones

- Todo departamento de aseguramiento de ingresos deben contar con controles automatizados que permitan mejorar su rendimiento y así su respuesta sea mejor frente a los múltiples problemas que enfrenta una operadora de servicio celular.
- Se debe tener clasificados los controles por su complejidad, para generar ya sea alarmas o informes automáticos, debe tener en cuenta esta división en complejidad para no desgastar tiempo innecesario en la realización de los mismos.
- En controles conocidos y que se está totalmente seguros que son casos inconsistentes se recomienda incluir la corrección automática, esto representa un gran ahorro en dinero para la empresa además del impacto de la imagen de cara al cliente, ya que inmediatamente cuando se logra detectar un error se corrige, y no se tienen que esperar por la revisión y aprobación de algunas áreas.
- Incluir en la automatización de controles la determinación automática de los causales del problema, esto ayudara mucho para corregir de manera ágil la causa de los problemas encontrados.

BIBLIOGRAFÍA

- Alexander, H., Zhang, Q., Alexander, R.B., Zhang, J., Jiang, X., Zhang, B., & Zhang, Y. (2000). *The importance of telco's revenue assurance function study*. IEEE Xplore. Recuperado de <https://goo.gl/nN4zXw>
- Baca Urbina, G. (2016). *Introducción a la seguridad informática*. Grupo Editorial Patria, ProQuest Ebook Central. Recuperado de <https://goo.gl/C3Twhk>.

- Global Revenue Assurance Professional Association (2009). "Levels of Assurance", The GRAPA Standards. Recuperado de <https://goo.gl/fhKeBA> (20-10-2012)
- Gregory, I., Landryová, L., & Soldán, P. (2011, 28 de Mayo). *The monetization of highly automated systems in SMEs: A case study*. IEEE Xplore. Recuperado de <https://bibliotecas.ups.edu.ec:2095/Xplore/home.jsp>.
- Ingeniería en procesamiento digital (2013). "Legalidad del servicio de telefonía por Internet". Telefonía por Internet. Recuperado de <https://goo.gl/qW7KKc> (12-12-2012).
- Khairul Nizam Baharim; Mohd. Shafri Kamaruddin ; Faeizah Jusof (2008). *Leveraging Missing Values in Call Detail Record Using Naïve Bayes for Fraud Analysis*. IEEE Xplore. Recuperado de <https://goo.gl/a8ojYP>
- Madureira, M. A., Maia, D., Mano, R., & Céspedes, R. (2005, 9 de Junio). *Benefits and results of a revenue assurance and audit program for distribution utilities*. IEEE Xplore. Recuperado de <https://goo.gl/1KHns8>
- Maia, D., & Madureira, M. A. (2003). *Case study: an innovative revenue assurance process*. Metering Latin America. Sao Paulo.
- Mano, R., Céspedes, R., & Maia, D. (2003). *An ISO 9000 Certifiable Revenue Assurance Process*. Metering Europe, October. Paris, France.
- _____ (2004). *Protecting revenue in the distribution industry: a new approach with the revenue assurance and audit process*. IEEE Xplore. Recuperado de <https://goo.gl/Y4iw36>
- _____ (2004). *Revenue Assurance & Audit Process*. Metering Americas. San Diego, USA, March.
- Meza, M. (2008). *Fraudes a las plataformas. Fraudes en telecomunicaciones*. Recuperado de <https://goo.gl/fG4cMN> (3-08-2012).
- Xerandy, H., Suksmono, A.B., & Nugraha, T. (2006). *A research on usage pattern and analysis technique for communication fraud: SIM cloning and surfing*. IEEE Xplore. Recuperado de <https://goo.gl/sejjmM>

CAPÍTULO 2

El flujo de información de los Sistemas de Información Académica sobre Bases de Datos Federadas

Bernabé Ortega-Tenezaca
Carlos Israel Núñez Miranda

FUNDAMENTACIÓN DEL PROBLEMA

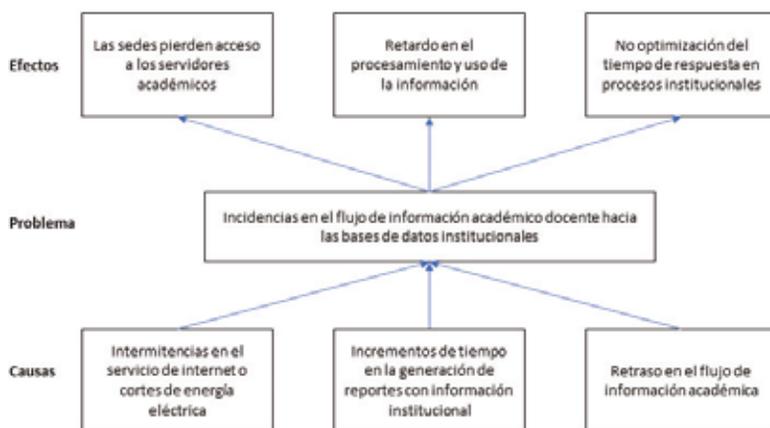
En el Ecuador, las instituciones de Educación Superior son evaluadas por entes gubernamentales, con la finalidad de categorizarlas, de acuerdo a criterios de evaluación aplicados sobre el componente académico, eficiencia, investigación, infraestructura y organización. Cada institución establece políticas propias para el mantenimiento y consolidación de la información, apoyados fundamentalmente sobre las TIC, cuya fortaleza principal es la de brindar las facilidades para el registro de información y obtención de informes de sus bases de datos en correspondencia de los indicadores de acreditación institucionales y de carrera (Antonio & De la Cuerda 2013).

Los sistemas informáticos universitarios se encuentran alojados en los servidores del Centro de Datos del campus principal, con la finalidad de obtener acceso inmediato a la información contenida en las bases de datos académicas y de gestión de forma centralizada

en dependencia de su contenido general y de forma local para el manejo y acceso inmediato para la generación de reportes, así también para la realización de tareas de mantenimiento de las bases de datos, respaldos, sincronización y gestión de la información almacenada.

Los servicios informáticos, brindan el apoyo en las tareas de gestión académica y administrativas, que permite tener información de calidad y oportuna. Sin embargo, la expansión de las sedes y centros de investigaciones, no siempre se encuentran en la misma provincia, donde se experimentan cortes continuos de energía y muchas veces prolongados; de igual forma el servicio de telecomunicaciones en las extensiones es deficiente provocando interrupciones en la continuidad de servicios, dificultando el flujo de información continua, en la generación de evidencias de cumplimiento académico, investigativo, entre otros, ocasionando retrasos y generación de trabajo administrativo adicional para la resolución de las incidencias.

Gráfico 1
Relación Causa-Efecto



Elaboración: Autores

Análisis crítico

Los servicios de transmisión de datos en las sedes presentan frecuentes intermitencias y cortes debido a problemas en la infraestructura de los diversos proveedores, lluvias eléctricas, accidentes de tránsito que involucran postes para cables de energía eléctrica, entre otros, lo cual impide que las sedes tengan acceso en tiempo real a los aplicativos informáticos de gestión de clase, información web, plataformas virtuales y demás servicios.

El incremento de población académica provoca lentitud en la generación de reportes de evidencias de investigación, académicas y administrativas frecuentemente con la finalidad de procesar la información obtenida y utilizarla en proceso de gestión en las diferentes entidades y organismos de regulación de acuerdo al organigrama de distribución universitaria. La obtención de la información es lenta y provoca retardos en la consolidación y cruce de la información.

Los procesos administrativos docentes toman mucho más tiempo por el crecimiento de docentes, carreras y estudiantes. Entre ellos el proceso de gestión y elaboración de distributivos dentro de la sede y extensiones se ha vuelto mucho más complicado por el tiempo que toma desarrollarlos en base a normativas vigentes que se convierten en el principal insumo para el proceso de parametrización de los sistemas de acuerdo a los periodos académicos.

La carga, tráfico y flujo de la información se encuentra soportada únicamente en un servidor centralizado, donde existen aplicaciones académicas y administrativas de gestión de la información, dependientes de la energía eléctrica y disponibilidad del servicio de Internet. Operativamente, existen tiempos de mayor carga como el inicio de cada semestre, cierre de mes, cierre de cada ciclo, peticiones de información para reuniones de Consejo Académico, Áreas de conocimiento, procesos de contraloría, términos de acreditación, reporte a las entidades de control entre otros.

ESTADO DEL ARTE

En el repositorio de la Universidad Autónoma de Asunción Paraguay, Facultad de Ciencias y Tecnología, existe un trabajo de tesis titulada “Sistema de bases de datos federadas para la gestión de la información”, en el cual se investiga la arquitectura del paradigma de un Sistema de Base de Federada, aplicados en un entorno de estudio real, que integra fuentes heterogéneas dispersas de una organización gubernamental de salud, de modo que bajo ciertos parámetros se integra totalmente y permite la incorporación de nuevas fuentes de datos, utilizando para ello, el despliegue de máquinas virtuales y postgres como gestor de base de datos, mismos que son integrados mediante el uso de la herramienta db-link que permite el acceso remoto a datos con lo cual se logra determinar que la cantidad de recursos y tareas a ser destinadas en una composición de fuentes heterogéneas de datos se reduce aplicando apropiadamente la tecnología enfocada al acceso integrado a la información, y denota que una adecuada administración del SBDF es directamente proporcional con el conocimiento del profesional que se encuentra a cargo (Agüero Torales, 2013).

En la Escuela Superior Politécnica del Ejército ESPOL, Facultad de Ingeniería en Electricidad y Computación se encuentra una tesis de maestría titulada “Implementación de un sistema de autenticación federada Single Sing On para una empresa de medicina pre-pagada”, que propone brindar una solución de gestión de identidad federada a través de Active Directory Federation Services ADFS; hace uso de un mecanismo intermedio denominado Web Application Proxy WAP cuyo principal aporte es la implementación de un sistema centralizado de administración de usuarios que demuestra que en el proceso de registro de datos no existe duplicidad de los mismos lo que provoca una reducción significativa de costos asociados a dicha administración, bajo consideraciones de seguridad de la identidad y control de los accesos a los sistemas propios del negocio bajo la implementación del protocolo WS-Federation que permite bloquear

el servicio a un cliente con tan solo dar de baja su cuenta sobre software privativo (Álvarez & Rigoberto, 2016).

En la Universidad Técnica de Ambato, en la Facultad de Ingeniería en Sistemas Electrónica e Industrial, en la Carrera de Ingeniería en Sistemas Computacionales e informáticos, se encuentra registrado un trabajo de investigación titulado “Sistema de Federaciones de identidades para la Facultad de Ingeniería en Sistemas, Electrónica e Industrial usando software de código abierto”, que mediante el uso de openLDAP y middleware de autenticación Shibboleth, sobre herramientas Open Source para la federación de aplicaciones, con lo cual se logra otorgar un sistema de identidad federada Single Sign-On, integrado al sistema nacional ecuatoriano MINGA que provee acceso a diferentes recursos académicos universitarios mediante la suscripción a servicios de internet avanzado de la Corporación Ecuatoriana para el Desarrollo de la Investigación y la Academia CEDIA (Guevara Aulestia & Cevallos Teneda, 2016).

METODOLOGÍA DE LA INVESTIGACIÓN

En el presente trabajo de investigación se planteó la siguiente hipótesis: Las bases de datos federadas inciden en la optimización del flujo de información académica institucional universitaria.

La presente investigación estará enmarcada dentro del paradigma crítico propositivo por lo tanto tendrá un enfoque de la investigación predominante cualitativo porque se obtendrá información directa de los investigados, en virtud de los cuales será factible desarrollar un análisis crítico de los resultados y proponer alternativas de solución.

Esta investigación permite desarrollar un estudio sistemático de los procesos de generación de información académica científica desarrolladas en el campus principal, extensiones y centro de investigaciones, en el cual se producen los acontecimientos, el investigador

toma contacto en forma directa con la realidad, para tener información de acuerdo con los objetivos del problema.

La investigación es bibliográfica porque se realizará mediante la revisión de documentos oficiales, leyes, normas, artículos científicos, libros y otros; mismos que permiten fundamentar la base legal y obtener información referente al tema correspondiente, de esta manera se recopilará información valiosa que servirá como sustento científico del proyecto.

Nivel o tipo de investigación

La investigación utilizará un nivel exploratorio que permite conocer las características actuales del problema para determinar si es factible o no solucionarse; se pasará al nivel descriptivo para analizar el problema, estableciendo sus causas y consecuencias, así como las dificultades por las que está atravesando; el nivel correlacional, llegará a establecer la relación de una variable con la otra y la incidencia que tiene en la solución del problema.

Población y muestra

El presente proyecto trabajará con la población del personal docente y administrativo encargado con relación a la generación de información académica en su totalidad debido a que su tamaño no es significativo.

ANÁLISIS E INTERPRETACIÓN DE RESULTADOS

Formulación de la hipótesis

H_0 : No existe incidencia.

H_1 : Si existe incidencia.

H_0 = Las bases de datos federadas NO inciden en la optimización del flujo de información académica institucional universitaria

H_1 = Las bases de datos federadas SI inciden en la optimización del flujo de información académica institucional universitaria

Modelo matemático

$H_0: O = E$

$H_1: O \neq E$

Modelo estadístico

En este trabajo se utilizó la prueba estadística Chi - Cuadrado, debido a que es aplicable para variables aleatorias discretas o continuas, y nos permitirá determinar si existe o no, relación entre las mismas, con la siguiente fórmula:

$$\chi^2 = \sum [(O - E)^2 / E]$$

- χ^2 = Chi - Cuadrado
- O = Frecuencias Observadas
- E = Frecuencias Esperadas
- Σ = Sumatoria

Nivel de significancia

$\alpha = 0,05$ Nivel de significancia y con un nivel de confianza del 95%

Para la verificación de la hipótesis, se procedió a seleccionar de dos preguntas por cada variable dependiente e independiente para determinar la relación que existe entre ellas.

- Bases de Datos Federadas

Pregunta 3: ¿La base de datos centralizada existente, garantiza la disponibilidad de información institucional?

Cuadro 1
Disponibilidad de información institucional

Opciones	Frecuencia	Porcentaje
SÍ	0	0%
NO	25	100%
TOTAL	25	100%

Elaboración: Autores

- Optimización del flujo de información académica institucional

Pregunta 9: ¿Se puede obtener la información consolidada requeridos en su departamento desde los sistemas informáticos?

Cuadro 2
SÍ para gestionar y consolidar informes

Opciones	Frecuencia	Porcentaje
SÍ	20	80%
NO	5	20%
TOTAL	25	100%

Elaboración: Autores

Cuadro 3
Chi-Cuadrado calculado

V	Alternativas	Cálculos				
		O	E	(O-E)	(O-E) ²	(O-E) ² /E
VI	SÍ	0	10	-10	100	10
	NO	25	15	10	100	6.66
VD	SÍ	20	10	10	100	10
	NO	5	15	-10	100	6.67
	TOTAL	100	50	0	100	33.33

Elaboración: Autores

Grados de libertad

$$gl = (f - 1)(c - 1)$$

$$gl = (2 - 1)(2 - 1)$$

$$gl = 1 * 1$$

$$gl = 1$$

Con los grados de libertad obtenidos el valor Chi-Cuadrado según la tabla de distribución es de: 3.8415

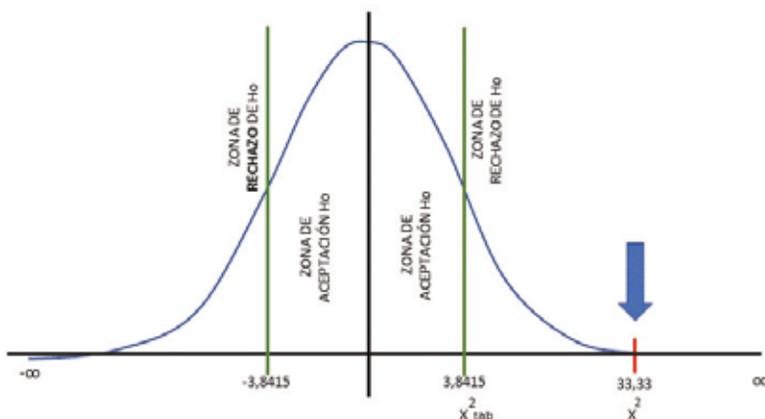
Cuadro 4
Valor Chi-Cuadrado según la tabla de distribución

v/p	0.001	0.0025	0.005	0.01	0.025	0.05
1	10.8274	9.1404	7.8794	6.6349	5.0239	3.8415
2	13.8150	11.9827	10.5965	9.2104	7.3778	5.9915
3	16.2660	14.3202	12.8381	11.3449	9.3484	7.8147

Fuente: <https://goo.gl/XKDXYA>

Decisión estadística

Gráfico 2
Reportes proporcionados validados o corroborados



Elaboración: Autores

X^2_{tab} = valor obtenido de la tabla de distribución de chi-cuadrado;
Tabla 4.16

X^2 = valor obtenido del cálculo de Chi-cuadrado; Tabla 4.15

$X^2 > X^2_{\text{tab}}$ Si se rechaza H_0 y se acepta H_1
 $33.33 > 3.84$

Entonces se rechaza y se acepta, lo que significa que las bases de datos federadas SÍ inciden en la optimización del flujo de información académica institucional universitaria.

CONCLUSIONES Y RECOMENDACIONES

Una vez concluida la etapa de trabajo en campo en contraste con los resultados del marco teórico estructurado, se puede determinar:

Los directivos y funcionarios de la Universidad Estatal Amazónica, poseen un alto grado de confianza en los sistemas informáticos de gestión de la información, dado que brindan un valor agregado en los procesos de registro de información.

En las sedes y centro de investigación se han experimentados intermitencias en el servicio de internet y suministro eléctrico, lo que provoca que no exista a disposición una conexión a internet o red de datos, y el acceso a los diferentes sistemas de gestión de la información.

Las bases de datos centralizada, provoca que quienes tienen acceso directo en la sede principal, no tengan inconvenientes en procesar la información, en contraste a la situación de las sedes, que, al experimentar interrupciones de acceso, se provoca trabajo adicional para resolución de incidencias.

Los reportes generados desde los sistemas de gestión de información, requieren ser revisados con la finalidad de corroborar que se hayan incluido completamente los registros de las sedes y centro de investigación.

Para ciertos procesos de gestión de la información existe grupos designados de funcionarios y docentes que evalúan la documentación registrada como evidencia del trabajo docente, investigativo y administrativo.

Se considera que el modelo de gestión actual es pertinente y óptimo para la generación de información académica y de registro de actividades no docentes, por lo que se cree conveniente extender las funcionalidades del sistema académico a las sedes, entre ellos, la posibilidad de registrar la asistencia directamente en cada sede.

En ciertos procesos de obtención y consolidación de información existen métodos manuales de revisión de información, previa su consolidación, tales como seguimiento del sílabo, evidencias de vinculación.

La información que maneja cada uno de los departamentos no es aislado, pues mantiene una dependencia del cumplimiento de otros procesos para poder procesarla, por lo que es necesario que exista una fluidez de los procesos y los datos correspondientes.

Se recomienda lo siguiente

Debido a la situación económica del Ecuador, que conlleva a una asignación presupuestaria que no contribuye a brindar soluciones definitivas en las que se requiere de actualización y aprovisionamiento tecnológico, es necesario establecer vías de solución alterna que permitan extender los servicios informáticos hacia las sedes y centro de investigación.

Las soluciones de red implican y de duplicidad de la información siguen dependiendo estrictamente de un canal de comunicación, por lo tanto, es necesario establecer soluciones a nivel de base de datos que permitan distribuir la información manteniendo un entorno que permita obtener la información de manera rápida y efectiva, pero a la vez que permita otorgar servicios informáticos a nivel local.

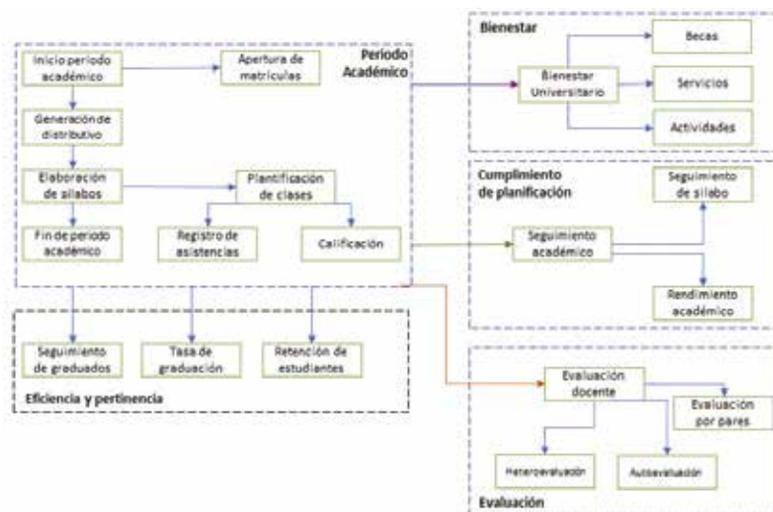
Extender los servicios informáticos, que permitan tener calidad de generación de evidencias en todos los ámbitos, especialmente para la implementación del control de asistencia a través del registro biométrico dactilar de los usuarios del sistema.

PROPUESTA TECNOLÓGICA

Descripción de procesos, sistemas y servicios informáticos institucionales que generan información académica

Procesos institucionales

Gráfico 3
Diagrama de procesos académicos



Elaboración: Autores

La Universidad Estatal Amazónica (UEA), mantiene procesos definidos enfocados en mantener procesos de calidad, en el aspecto administrativo, académico, de vinculación e investigación. La im-

portancia de la gestión académica radica en los procesos de seguimiento y cumplimiento de planificación.

Sistemas y servicios informáticos institucionales

La Universidad Estatal Amazónica (UEA) ha desarrollado sistemas informáticos de gestión de la información universitaria que permiten obtener de forma inmediata el estado actual de cumplimiento de varios indicadores y requisitos de acreditación universitaria, para lo cual se utiliza la infraestructura tecnológica ubicada en su centro de proceso de datos ubicado en el campus principal de Puyo.

Cada docente de la UEA tiene asignado un lugar de trabajo conformado por un escritorio, sillas de docente y visitantes, un archivador, acceso a una red de datos mediante un computador. De igual manera, todas y cada una de las aulas y laboratorios de la UEA, tienen como parte de su infraestructura y aprovisionamiento tecnológico, pizarrón de tiza líquida, una pantalla de proyección, acceso a red inalámbrica, un proyector, un equipo de registro biométrico dactilar, y un equipo de cómputo de acceso exclusivo para el profesor ubicado sobre el escritorio.

Sistema de Información Académico Docente (SIAD)

El Sistema de Información Académico Docente, SIAD, actualmente se encuentra en un estado de madurez óptimo acoplado a los requerimientos de las entidades de control universitario como CES, CEAACES, SENESCYT, desarrollado mediante el uso de software libre. Se compone de 170 tablas desplegadas sobre la versión 14.14 distrib 5.5.57 de MySQL con un sistema de almacenamiento gobernado por InnoDB, en un servidor CentOS 6.9 de 64 bits, y un entorno de desarrollo conformado por el lenguaje de programación PHP 5.4.45 zend engine v2.4.0 y la versión comunitaria del framework de desarrollo HTML5 ExtJS de la empresa Sencha (Ochoa Carrión, 2017).

Se caracteriza principalmente, porque mantiene un control de bloqueo de utilización del computador de cada una de las aulas, cuya activación depende de la planificación académica que realiza el docente en el sistema SIAD al iniciar el período académico, donde principalmente se calendariza y declara el nombre de la clase, los objetivos de la misma, entre otros datos importantes. A su vez permite realizar un seguimiento del sílabo tanto del profesor como del estudiante, llevando un registro de asistencia de manera eficiente y mucho más rápida, en comparación a la forma tradicional.

El Sistema de Información Académico Docente (SIAD) de la Universidad Estatal Amazónica cumple las exigencias del Modelo Genérico de Evaluación del Entorno de Aprendizaje de Carreras Presenciales y Semipresenciales de las Universidades y Escuelas Politécnicas del Ecuador, versión 2.0 con fecha marzo de 2015 (Ortega Tenezaca, & Núñez Miranda, 2017).

Área académica

Está relacionada directamente con la docencia y los estudiantes. Elaboración de su expediente, registrando y actualizando Datos Personales (cédula de identidad, usuario, apellidos y nombres, género, estado civil, grupo sanguíneo, país de origen, etnia, fecha de nacimiento, provincia, cantón, parroquia, tipo de discapacidad, ciudad de residencia, dirección actual, teléfonos de contacto y correos electrónicos personal e institucional), Curriculum Académico (nivel, grado, denominación, área, subárea, subárea específica, fecha título, Código SENESCYT, Fecha de registro, País, Institución y archivo adjunto), Estudios en Curso (nivel, grado, Número de Documento, denominación, área, subárea, subárea específica, fecha inicio, fecha termina, país, Institución, Institución financia, Culminado, y archivo adjunto), Producción Científica (corresponden al registro de los libros, artículos, patentes –Registros, Eventos– Seminarios con sus respectivos archivos adjuntos), Capacitaciones (Tipo, Nombre, Institución, Número de horas, Inicia, Termina y archivo adjunto), Contratos (Número de contrato, ingreso, culmina, categoría, cargo,

Relación de trabajo, unidad académica, dedicación, horas de dedicación, remuneración y archivo adjunto), Cargos Directivos (Número de documento, Cargo, descripción, Fecha inicia cargo, fecha finaliza cargo, estado y archivo adjunto), Cátedra (código, asignatura, carrera, año, semestre, paralelo, periodo lectivo) y Documentos Personales (documento, fecha de registro y archivo adjunto).

Elaboración del sílabo, registrando una breve descripción sobre la presentación de la asignatura y contribución a la formación del profesional, los objetivos generales de la asignatura, los resultados de aprendizaje del perfil a los que contribuye la asignatura, los resultados de aprendizaje de la asignatura, la programación del sílabo (donde los contenidos responden a los resultados del aprendizaje de la asignatura, fecha, hora de inicio y fin de la clase, duración en número de horas, lugar donde se desarrollará la clase (aula, auditorio, laboratorios, entre otros), bibliografía básica y complementaria, estrategias utilizadas por del docente y recursos que se utilizan para una determinada clase), el carácter de la investigación en los contextos de aplicación, evaluación de la asignatura (tipos de evaluación, breve descripción, fechas de evaluación y porcentaje sobre la nota final), compromisos de los estudiantes y del docente dentro del aula, aspectos relevantes del currículo del docente y bibliografía. Los sílabos pueden ser editados dentro del plazo establecido.

De acuerdo al distributivo asignado para cada docente, se debe planificar en la agenda las demás actividades de no docencia o de gestión, dando cumplimiento al tipo de docente con el número de horas diarias (docente a tiempo completo, a medio tiempo o tiempo parcial).

Una vez registrado su expediente, creados y transferidos los sílabos el docente puede iniciar su actividad docente según su agenda planificada, mediante dos métodos: el uso del sistema biométrico (registro de huella) o iniciando desde el Sistema SIAD desde la opción Impartir Clase. En ambos casos, el estudiante puede observar el cumplimiento de la programación del sílabo mediante la pantalla que se muestra a través del proyector luego del inicio de la actividad

del docente, como son: el tema, objetivos, estrategias y materiales a utilizarse en la clase.

El registro de asistencia de los estudiantes se realiza mediante el sistema biométrico, una vez finalizado se guarda y se cierra la ventana. El porcentaje de asistencia de los estudiantes se va incrementando cada vez que el estudiante asiste a clases debiendo cumplirse un 80% al finalizar el semestre, caso contrario el sistema no le tomará en cuenta para rendir el examen final. El docente dará cumplimiento con el tema de la clase según la programación, debiendo finalizar la clase registrando su huella digital en el biométrico.

Para el desarrollo de la planificación de la programación del sílabo el docente elaborará trabajos, lecciones, exámenes que serán registrados como notas frecuentes sobre 10 puntos, el sistema calculará sobre 3 puntos. El docente cumplirá con las fechas establecidas para los exámenes parciales, finales y supletorios debiendo registrar en el sistema hasta la fecha establecida en el sistema. El estudiante tendrá acceso a los sílabos de las asignaturas correspondientes al semestre que están cursando, realizar un seguimiento de su asistencia, así como también de las calificaciones frecuentes, parciales, final y si es el caso el supletorio, de igual forma podrá registrar y modificar su expediente en los campos habilitados.

Mediante las opciones de generación de reportes se obtiene la información correspondiente al cumplimiento de los indicadores de evaluación.

El docente puede visualizar la asistencia y el porcentaje proporcional de cada estudiante, punto clave para realizar un seguimiento al estudiante que no asiste con frecuencia. Se visualiza también las calificaciones y reportar el rendimiento académico al colectivo del semestre para la toma de decisiones; también puede registrar las evidencias de acuerdo a las actividades del docente que constan en el distributivo, tales evidencias deben ser escaneadas y subir al

sistema en formato PDF, siendo el coordinador de la carrera quien aprueba las horas del docente.

Para finalizar desde el perfil docente en el Sistema SIAD, puede visualizar el calendario académico para el periodo vigente, el calendario de evaluaciones para cada grupo de estudiantes y el calendario de feriados muy útil para no programar en el sílabo esos días.

Tiene acceso a la biblioteca donde se guardan todos los sílabos de todos los periodos elaborados por los docentes de todas las carreras, pudiendo ser copiado a la asignatura actual correspondiente y adelantar el trabajo de elaboración del sílabo, ésta es opcional.

Balcón de Servicios

El sistema de balcón de servicios, se conforma de subsistemas de gestión de información aplicados sobre los ámbitos de cumplimiento con indicadores de gestión universitaria de los departamentos de Bienestar universitario, Investigación, y actividades complementarias como Evaluación Docente, Seguimiento de Graduados, Bolsa de Empleo, Fichas estudiantiles y Prácticas pre-profesionales.

Sistema del Departamento de Bienestar Universitario

El Sistema del Departamento de Bienestar Universitario (BE-DBU), ha sido creado con la finalidad de brindar una herramienta tecnológica de gestión de ayudas económicas a los estudiantes en dependencia de los criterios determinados en la Ley Orgánica de Educación Superior, en su Capítulo del Principio de igualdad de oportunidades, por lo que se clasifican en becas de tipo socioeconómica, deportivas, de méritos académicos, concurso académico, investigación, cultural, IDI, residencia, económica tipo A.

Sistema del Departamento de Investigación

El Sistema del Departamento de Investigación, realiza el seguimiento de las publicaciones desarrolladas por los docentes con horas de investigación dentro de la UEA, quienes reportan el avance de su producción científica como libros o publicaciones en revistas indexadas con pertinencia local, regional o alto impacto.

Evaluación Docente

El Sistema de Evaluación permite realizar el proceso de evaluación docente realizado por los estudiantes a sus catedráticos, por sus pares académicos, por sí mismos, y por el coordinador de carrera. Este proceso se realiza por cada período académico.

Seguimiento a graduados

El sistema de seguimiento a graduados, permite llevar un registro activo de las fichas de quienes se han graduado en las distintas carreras de la UEA; permite obtener datos tabulados, estadísticos, realizar encuestas, obtener listados, realizar consultas por período, entre otros. Su importancia se refleja en que actualmente forman parte del cogobierno universitario.

Adicionalmente se cuenta con sistemas para la gestión del POA, PAC, Plan de mejoras, seguimiento de proyectos, solicitudes de compra, gestión de compras públicas, entre otros.

Despliegue de base de datos y sistemas informáticos de la UEA

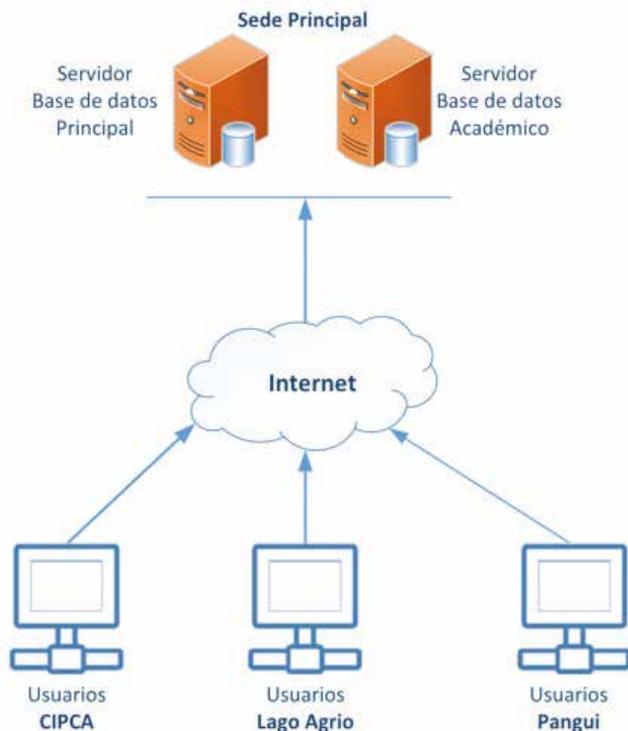
La UEA en su campus universitario de la ciudad de Puyo, mantiene un centro de procesamiento de datos y telecomunicaciones, desplegados en servidores físicos y virtualizados, por los cuales, proporcionan las condiciones necesarias para el despliegue de los servicios y sistemas informáticos, cuya gestión se realiza en la Unidad de Tecnologías de la Información bajo coordinación y aprobación de la máxima autoridad.

Cuadro 5
Descripción de los principales sistemas informáticos UEA

Sistema	Actividades	Componentes	Evidencias
SIAD (Sistema Académico Docente)	Docentes	Académico	Vicerrectorado Académico Departamento de Planificación y desarrollo Secretaría Académica Decanatos Dirección de escuela Coordinación de carrera Departamento Financiero
	No docentes	Investigación Gestión Vinculación	
BEDBU (Departamento de Bienestar Universitario)	Ayudas económicas	Becas	Vicerrectorado Administrativo Bienestar Universitario Financiero
Departamento de investigación	Resultados de investigación	Artículos y libros	Vicerrectorado académico Departamento de investigación Departamento de Planificación y desarrollo Financiero
Evaluación Docente	Administrativas	Evaluación Coevaluación Heteroevaluación	Vicerrectorado académico Coordinador de Carrera Docentes
Seguimiento a graduados	Administrativas	Directorio	Rectorado Departamento de planificación y desarrollo

Elaboración: Autores

Gráfico 4
Despliegue lógico de servidores de base de datos



Elaboración: Autores

Existen servicios complementarios que han sido migrados desde plataformas opensource a servicios educativos otorgados de forma gratuita como Office 365, G-suite y Workplace, donde se despliegan herramientas de gestión como correo electrónico, almacenamiento en nube, redes sociales, entre otras.

Adicionalmente a la tabla de descripción de servicios existen aplicativos de mucha importancia como soporte técnico, gestión de servicios de tecnología de la información, gestión de credenciales de

acceso a sistemas, phpmyadmin, sistema de reservas de uso de laboratorios, gestión de roles de pago, los cuales no se han incluido, debido a que su labor es de apoyo a la gestión y no de generación directa de indicadores de acreditación, sin embargo su aporte es vital en el desarrollo de las actividades universitarias en todos sus aspectos.

Cuadro 6
Servidores que generan información sobre indicadores

Servidor	Servicios	DBMS	Tipo	Ubicación
Master	Sitios web institucionales	MySQL	Información	Campus Puyo
	Entornos virtuales	MySQL	Gestión	Campus Puyo
	Revista científica	MySQL	Gestión	Campus Puyo
	Aplicativos de apoyo a la gestión	MySQL	Gestión	Campus Puyo
Goliat	Sistema Académico	MySQL	Gestión	Campus Puyo
	Balcón de servicios	MySQL	Gestión	Campus Puyo
Docs	Sistema de Publicación de Tesis	Postgres	Gestión	Campus Puyo
	Sistema de gestión bibliotecaria	MySQL	Gestión	Campus Puyo

Elaboración: Autores

Análisis del despliegue de base de datos institucionales

Base de datos Master

El servidor de base de datos principal soporta la mayoría de bases de datos informativas, de gestión y aplicativos de apoyo, implementadas sobre software libre.

Cada usuario está limitado únicamente a la base de datos asignada y con la política de privilegio mínimo y acceso permitido para localhost.

Cuadro 7
Despliegue de base de datos master

Software	Tipo despliegue	Sitios desplegados	Bases de datos
CMS	Página institucional	1	1
CMS	subdominios	27	27
Gestión	Desarrollo interno	2	2
LCMS	Entornos virtuales de aprendizaje	7	7
Otros	Aplicativos de apoyo a la gestión	7	7

Elaboración: Autores

Gráfico 5
Rendimiento de servidor master

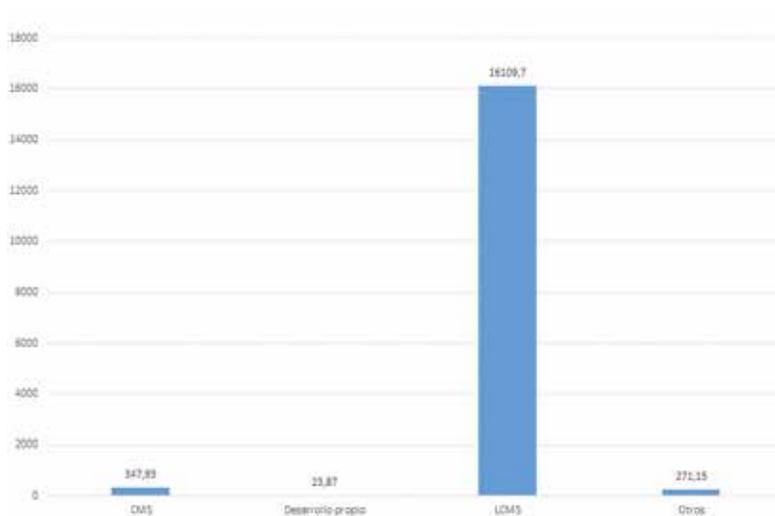


Fuente: Aplicativo Workbench

El gráfico 6 indica el espacio de disco utilizado por las bases de datos, de las cuales los LCMS (Learning Content Management System), tienen mayor uso que las bases de datos informativas y las

herramientas de apoyo a la gestión contenidas en el servidor, esto se debe a las mismas, se implementan como herramienta de apoyo a la docencia.

Gráfico 6
Espacio de disco de bases de datos-Servidor Maestro



Elaboración: Autores

Base de datos académicos

El servidor de base de datos académicos contiene las bases de datos de uso interno, las cuales se han desarrollado para gestionar el balcón de servicios y sistema académico SIAD, con un total de 9 bases de datos, en las cuales, se asigna un único usuario, con la política de privilegio mínimo.

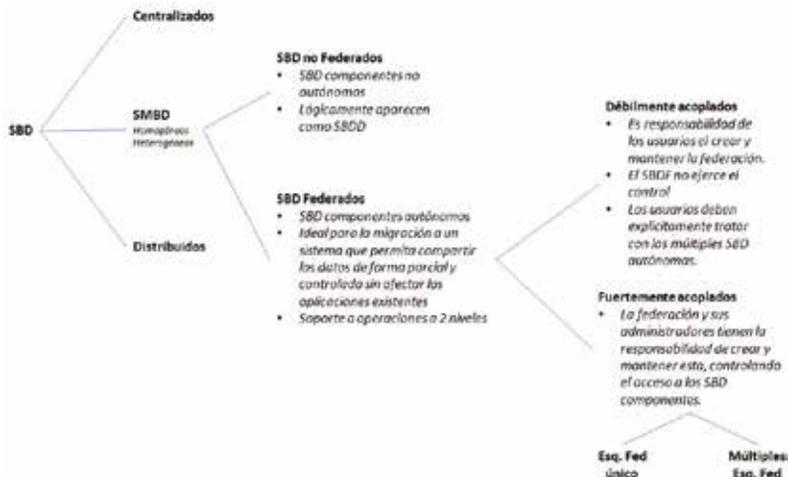
La base de datos académica, es la que utiliza mayor espacio debido a su uso diario, y el registro de asistencia y evidencia de trabajo docente y vinculación.

Mediante el gráfico 8 se puede observar que la tabla que tiene mayor parte de los registros se denomina asistencia_alumno, que en la actualidad utiliza 679.5MB, debido a la cantidad de registros diarios, lo cual influye en el tiempo que tarda la generación de la cartilla del control de asistencia, y su relación directa con las asignaturas, planificación académica, días de asistencia que toma en cuenta la fecha de matrícula legal del estudiante.

Gestión de la federación de tablas

Según el caso de estudio expuesto, se utilizará el sistema de almacenamiento federated, debido a que las fuentes de datos son homogéneas sobre el motor de base de datos MySQL, en tanto que la administración de despliegue remoto se desarrolla mediante conexión SSH.

Gráfico 9
Arquitectura de BD según contexto de un SBDF



Elaboración: Autores

Determinación de algoritmo de optimización del flujo de optimización sobre procesos de generación de información

Algoritmos heurísticos en optimización combinatoria

Según la Real Academia de la lengua, indica que la palabra griega *heuriskein* se define en algunas ciencias, como la manera de buscar la solución mediante métodos no rigurosos. El autor Rafael Martí añade que se utiliza en el ámbito de la optimización y se aplica en la industria y en la ciencia con una amplia gama de problemas que van desde los más sencillos a la complejidad de la ingeniería y define a los problemas de difícil solución como aquellos en los cuales no se puede garantizar el encontrar la mejor solución posible en un tiempo razonable.

Un método heurístico es un procedimiento para resolver un problema de optimización bien definido mediante una aproximación intuitiva, en la que la estructura del problema se utiliza de forma inteligente para obtener una buena solución (Martí, 2003).

...existen otras razones para utilizar métodos heurísticos, entre las que podemos destacar: Algoritmos Heurísticos.

- El problema es de una naturaleza tal que no se conoce ningún método exacto para su resolución.
- Aunque existe un método exacto para resolver el problema, su uso es computacionalmente muy costoso.
- El método heurístico es más flexible que un método exacto, permitiendo, por ejemplo, la incorporación de condiciones de difícil modelización.
- El método heurístico se utiliza como parte de un procedimiento global que garantiza el óptimo de un problema. Existen dos posibilidades:
 - El método heurístico proporciona una buena solución inicial de partida.

- El método heurístico participa en un paso intermedio del procedimiento, como por ejemplo las reglas de selección de la variable a entrar en la base en el método Simplex (Martí, 2003).

Los algoritmos heurísticos más conocidos pertenecen a las siguientes características:

- Métodos de descomposición
- Métodos inductivos
- Métodos de reducción
- Métodos constructivos
- Métodos de búsqueda local

Cuadro 8
Resumen de métodos estructurados

Algoritmo	Aplicación
KP (Mochila)	Cutting Stock, donde hay que cortar una plancha de acero en diferentes piezas. Determinar los artículos que puede almacenar un depósito para maximizar su valor total. Maximizar el beneficio en asignación de inversiones cuando solo hay una restricción.
Cubrimiento de conjuntos	Localización de servicios, tales como hospitales, bomberos, etc. La asignación de tripulaciones a vuelos.
Empaquetado de conjuntos	Uno de los ejemplos/aplicaciones es el problema del Acoplamiento Máximo o Matching. Un acoplamiento es un subconjunto de las aristas de un grafo de manera que cualquier vértice no sea incidente con más de una de esas aristas. El problema del acoplamiento máximo consiste en encontrar un acoplamiento de máximo cardinal.
Partición de conjuntos	Asignación de tripulaciones en una versión más restringida que la anteriormente mencionada. Creación de distritos electorales: Asignación de electores a un colegio electoral.
Viajante (Traveling Sales Problem TSP)	Fabricación de circuitos integrados. Rutas de vehículos. Recogida (robotizada) de material en almacenes. Instalación de componentes en ordenadores. Aparece como sub-problema en otras aplicaciones.

Fuente: Procedimientos Meta heurísticos en Optimización Combinatoria Rafael Martí

Criterios de selección

Según el autor Rafael Martí, los algoritmos idóneos a aplicarse son el de KP (mochila) y Viajante (Traveling Sales Problem TSP), sin embargo, se descarta este último por cuanto su objetivo es el de crear un ciclo hamiltoniano, lo cual no es procedente en nuestro entorno por cuanto no se puede garantizar un flujo continuo del suministro de internet entre las sedes y centro de investigaciones.

Algoritmos de optimización

Algoritmo Knapsack (Algoritmo de la mochila)

Se tienen n objetos donde cada objeto j tiene un peso un valor w_j . El problema consiste en seleccionar los objetos a incluir en una mochila sin exceder el peso máximo W , de modo que el valor total de los mismos sea máximo. Para formular el problema se define una variable x_j , por cada objeto j , de modo que vale 1 si el objeto es seleccionado y 0 en caso contrario (Martí, 2003).

$$\begin{aligned} & \text{MAX } v_1x_1 + v_2x_2 + \dots + v_nx_n \\ & \text{s: } a_j : \\ & w_1x_1 + w_2x_2 + \dots + w_nx_n \leq W \\ & x_j \geq 0, \text{ entero} \end{aligned}$$

Este problema tiene numerosas aplicaciones tales como:

- La denominada Cutting Stock, en donde hay que cortar una plancha de acero en diferentes piezas.
- Determinar los artículos que puede almacenar un depósito para maximizar su valor total.
- Maximizar el beneficio en asignación de inversiones cuando sólo hay una restricción.

Implementación del SGBDF

La implementación de la solución propuesta se realiza primeramente en un entorno de pruebas virtualizado, que emula el caso

de estudio propuesto (previa la implementación del SBDF definitiva), cuyo entorno es similar al estado actual, que contiene la base de datos del Sistema de Información Académico Docente SIAD, con un corte actualizado de datos correspondiente al cierre de periodo académico marzo 2017-julio 2017.

Vmware workstation 12.5.6, es una aplicación que funciona en entornos Windows y permite convertir al equipo anfitrión en proveedor de equipos virtuales, proporcionando características propias a cada sistema operativo seleccionado tales como número de procesadores, memoria RAM, espacio de almacenamiento, dispositivos de red virtuales y la posibilidad de configurar una red virtual entre los equipos seleccionados.

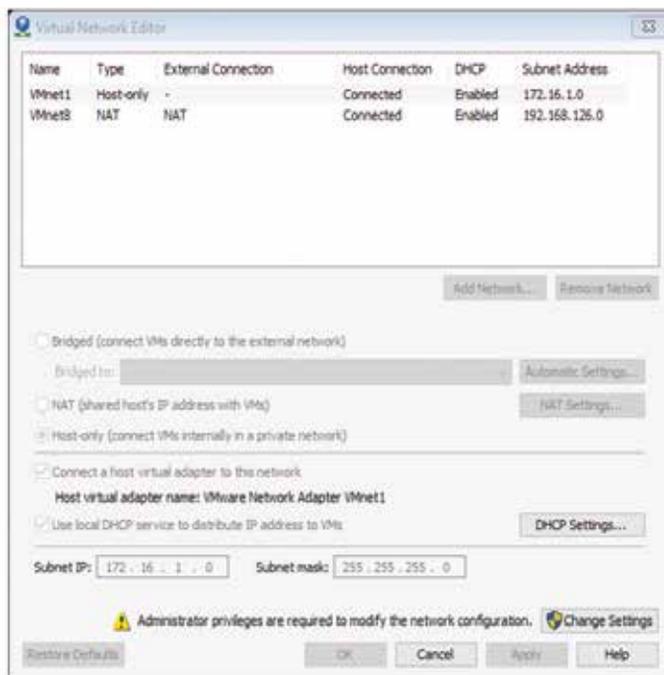
Configuración del entorno virtual

Inicialmente se realiza la configuración de la red virtual que consiste en la configuración de dos tarjetas de red virtuales que cumplen las funciones de conectividad LAN y acceso WAN para todas las máquinas virtuales que se configuraran para lo cual se designa a como red interna 172.16.1.0 sobre la tarjeta VMnet1, en tanto que la navegación WAN se realiza mediante la interfaz VMnet8 mediante NAT, es decir sobre la tarjeta física de conexión del equipo anfitrión en este caso corresponde a la tarjeta de red Intel Centrino Advanced-N 6235.

Una vez realizado el proceso, se puede observar el panel de control con las interfaces de red disponibles y configuradas que brindan servicio mediante DHCP.

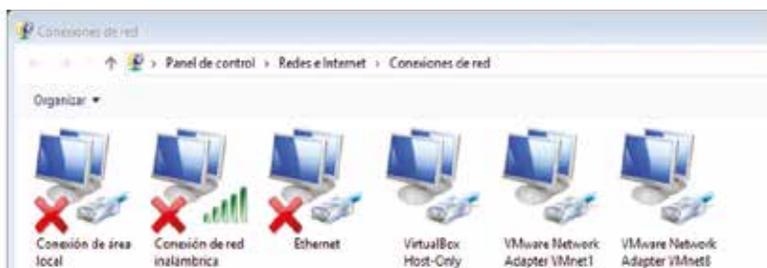
El proceso de virtualización ejecuta simultáneamente cuatro equipos con características similares de procesador, memoria RAM, espacio de almacenamiento, dos tarjetas de red virtuales (de LAN, y WAN), sobre los cuales corre un sistema operativo Centos 7 release 7.3.1611 (Core) de 64bits sin entorno visual, con servidor de base de datos MariaDB 10.2.7 en un entorno seguro, y acceso SSH.

Gráfico 10
Configuración de red virtual



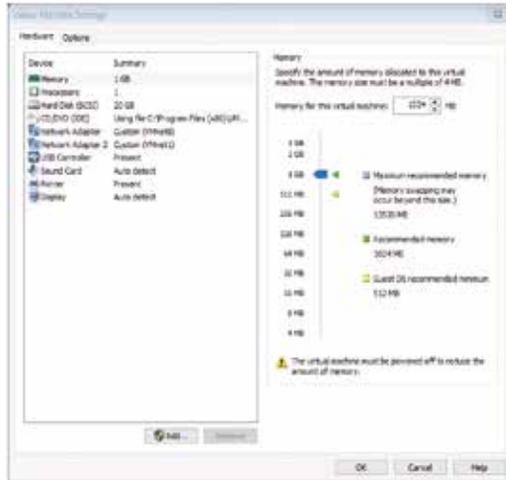
Elaboración: Autores

Gráfico 11
Conexiones de red disponibles



Elaboración: Autores

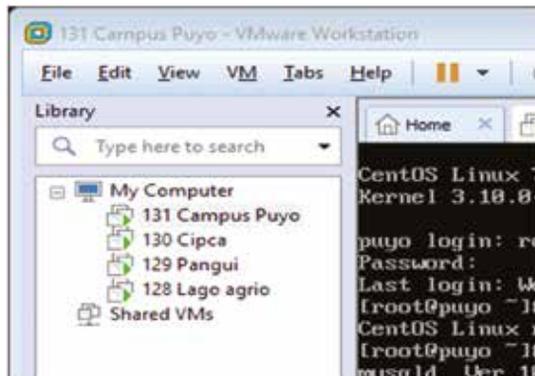
Gráfico 12
Configuración de máquinas virtuales



Elaboración: Autores

La configuración de las tarjetas de red permite que cada equipo se comporte de forma autónoma con acceso propio a Internet.

Gráfico 13
Despliegue de equipos



Elaboración: Autores

Gráfico 14

Recursos del servidor de base de datos principal

```

login as: root
root@172.16.1.131's password:
Last failed login: Wed Aug 16 15:01:38 -05 2017 from gateway on ssh:notty
There was 1 failed login attempt since the last successful login.
Last login: Wed Aug 14 18:53:47 2017 from gateway
[root@puyo ~]#
[root@puyo ~]# cat /etc/redhat-release
CentOS Linux release 7.3.1611 (Core)
[root@puyo ~]#
[root@puyo ~]# mysql -V
mysql Ver 14.14 Distrib 5.6.34, for Linux (x86_64) using EditLine wrapper
[root@puyo ~]#
[root@puyo ~]#
[root@puyo ~]# df -h
S.ficheros      Tamaño Usados  Disp Us% Montado en
/dev/mapper/cl-root    17G   5,0G    12G   34% /
devtmpfs          478M    0  478M    0% /dev
tmpfs             489M    0  489M    0% /dev/shm
tmpfs            489M   6,0M   482M    1% /run
tmpfs            489M    0  489M    0% /sys/fs/cgroup
/dev/sda1        1014M  184M   831M   18% /boot
tmpfs           50M    0   50M    0% /run/user/0
[root@puyo ~]#
[root@puyo ~]# free -m
              total        used         free       shared  buff/cache   available
Mem:           576          204          451           6         140         519
Swap:         2047           0         2047
[root@puyo ~]#

```

Elaboración: Autores

Gráfico 15

Tablas de la base de Datos SIAD

```

Database changed
mysql> show tables;
+-----+
| Tables_in_siad |
+-----+
| academico_alumno |
| academico_docente |
| acadotrotitulo |
| acaduea |
| actividad_empresa |
| actividad_podocencia |
| actividad_pyvc_estudiante |
| admision |
| agenda_academica |
| ..... |
| tipo_relacion_trabajo |
| tipo_seccion |
| titulo |
| ubicacion_geografica |
| usuario |
| vinculacion_preprofesional |
+-----+
170 rows in set (0,00 sec)

```

Elaboración: Autores

En el equipo 131 Campus Puyo, representa la sede principal de la Universidad Estatal Amazónica, ubicada en la ciudad de Puyo, por lo tanto, se representa la situación actual implementando una copia exacta del Sistema Académico Docente, para lo cual se extrae un backup de la base de datos que tiene un tamaño de 276MB, desplegadas en 170 tablas.

El espacio de producción de la base de datos es de 1088.22MB

Gráfico 16
Espacio de la base de Datos SIAD

```
mysql> SELECT table_schema "siad", sum(data_length + index_length) /1024/1024 "MB"
      FROM information_schema.tables GROUP BY table_Schema;
+-----+-----+
| siad      | MB      |
+-----+-----+
| information_schema | 0.00976563 |
| mysql     | 0.95384789 |
| performance_schema | 0.00000000 |
| siad      | 1088.20285320 |
+-----+-----+
4 rows in set (0,82 sec)

mysql>
```

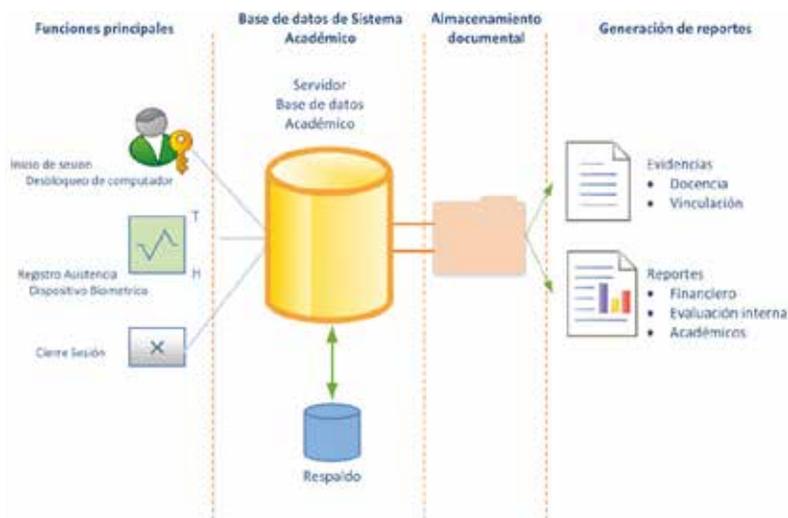
Elaboración: Autores

Flujo de información-Estado Actual

Para la realización de la presente investigación se utiliza el Sistema de Información Académico Docente SIAD, cuyo flujo actual de datos se despliega bajo un modelo centralizado, con un solo motor de base datos que brinda el servicio al campus principal, las sedes y el centro de investigaciones, lo cual representa un problema de generación de información debido a que en las sedes y centro de investigación, no se puede desplegar la solución completa que incluye la utilización de registro dactilar de huellas para el inicio de sesión en el computador del docente, de igual forma, durante las intermitencias de servicio de Internet y cortes de servicio de energía eléctrica,

el acceso externo no se podría realizar, sobre todo porque no existen administradores TIC en las extensiones y centro de investigación.

Gráfico 17
Procesos SIAD



Elaboración: Autores

La distribución de la infraestructura tecnológica es heterogénea, al igual que el servicio de red.

Cuadro 9
Distribución actual de la infraestructura tecnológica

IP pública	Localización	Ancho de banda	Proveedor ISP	Medio
186.3.44.x	Puyo	150MB	CEDIA	FO 1/1
186.47.83.x	Pangui	20MB	CNT	FO 8/1
181.112.218.x	Lago Agrio	20MB	CNT	FO 8/1
181.112.149.x	Km 42 vía Tena	20MB	CNT	FO 8/1

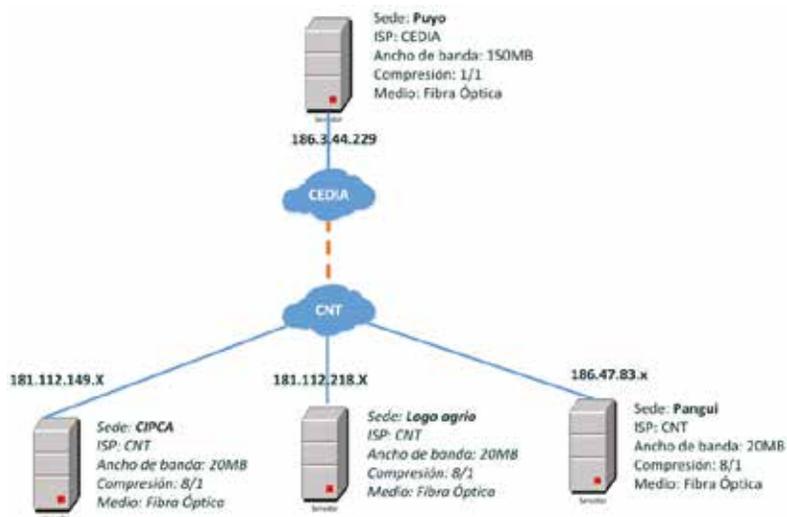
Elaboración: Autores

Dentro del análisis del Sistema de Información Académico Docente SIAD, resaltan las tablas correspondientes al registro de:

- Docentes
- Docentes Periodo asignatura
- Notas Alumno Asignatura
- Planifica Asignatura
- Asistencia Alumnos
- Agenda académica (controla las actividades de no docencia como investigación, gestión, vinculación)

Cuyo funcionamiento permitiría extender la gestión académica en las sedes y centro de investigación.

Gráfico 18
Distribución de red



Elaboración: Autores

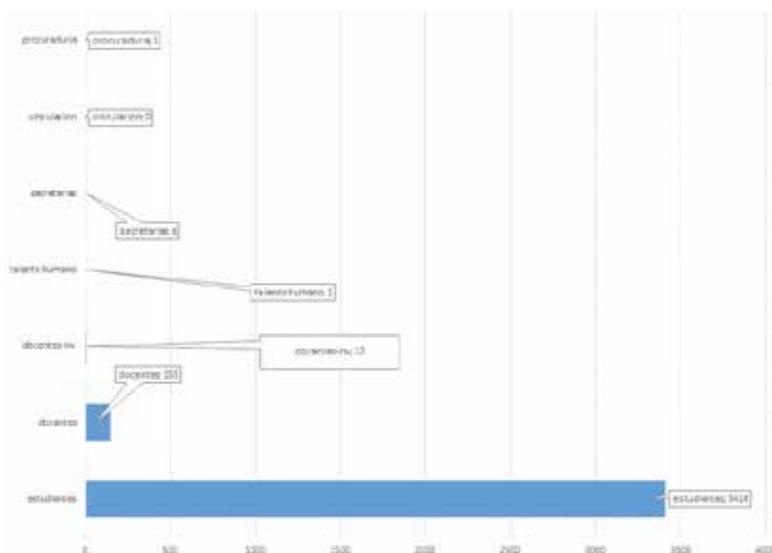
Cuadro 10
Distribución de la comunidad académica

Población	Campus	CIPCA	El Pangui	Lago Agrio
Estudiantes	2766	332	63	255
Docentes	136	0	7	12
Docentes-Investigadores	0	12	0	0
Talento Humano	1	0	0	0
Secretarías	7	0	0	1
Vinculación	1	0	0	0
Procuraduría	1	0	0	0

Elaboración: Autores

Se consideran las secretarías de: rectorado, vicerrectorado académico, Departamentos de Ciencias de la Vida, Departamento de Ciencias de la Tierra, Lago Agrio, y secretaría académica con tres funcionarios (la Sede del Pangui y CIPCA no registran secretarías).

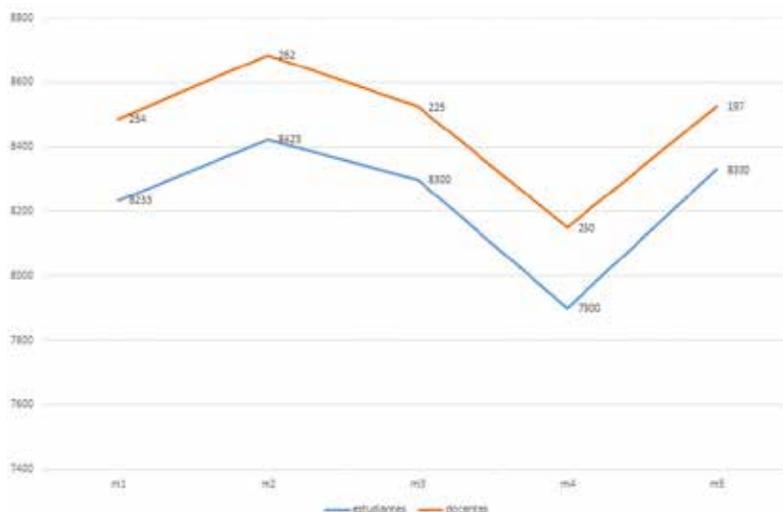
Gráfico 19
Población del sistema académico SIAD



Elaboración: Autores

La tasa transaccional diaria durante el periodo académico se encuentra en el rango de 8 233 y 8 423 registros generados por control de asistencia y entre 254 y 262 registros de evidencias docentes, generadas entre las 7h00 y 19h00 con un aproximado máximo de 30 estudiantes por curso.

Gráfico 20
Muestras aleatorias de generación de registros SIAD



Elaboración: Autores

Interpretación del problema

De acuerdo al flujo actual de información, se representa el problema mediante un modelo de optimización apoyado en algoritmos de optimización, representado en el gráfico 21 de la siguiente forma:

Modelo de federación

Como parte del proceso principal es necesario habilitar el engine federated en nuestro servidor de base de datos principal.

Para la implementación del algoritmo seleccionado, es necesario establecer los beneficios y los pesos de las rutas seleccionadas, por lo que se utiliza una escala máxima de 10 puntos sobre los siguientes criterios.

- Población universitaria por localidad
- Número de transacciones por hora
- Ancho de banda disponible
- Accesibilidad a la información
- Características de los servidores

Cuadro 11
Cálculo de peso beneficio

Sedes	Población	Transacciones	Peso
Puyo	5	5.2	1
CIPCA	3	1.1	1
Lago Agrio	1	2.8	1
El Pangui	1	1.1	2

Elaboración: Autores

La implementación del algoritmo KP, se lo aplica mediante el uso de Wolfram Development Platform a través del uso de la función Knapsack Solve en la cual se ingresa la distribución de los pesos y beneficios determinados en el Cuadro No. 11 y un valor de peso máximo de 25% que corresponde a cada una de las sedes, de donde se obtiene como vector de entrada los siguiente:

La solución óptima estimada de acuerdo a la estimación del algoritmo KP, asigna muy poco valor a la base de datos ubicada en el Pangui. Su explicación radica en el número de transacciones que maneja la base de datos, al momento solamente se tiene registro de asistencia y evidencia de docentes, con un número limitado de 63 estudiantes y 7 docentes, adicionalmente en esta sede no se genera evidencias de investigación ni vinculación puesto que, por su reciente inauguración, no se tienen cursos superiores que realicen dichas actividades, en tanto que la mayoría de docentes trabaja bajo la modalidad de contrato.

Gráfico 23
KnapsackSolve

```

In[85]:= UEA = {Entity["Sedes", "Puyo"] -> {Quantity[5, "Poblacion"], Quantity[5.2, "Transaccion"], 1},
Entity["Sedes", "Cipca"] -> {Quantity[3, "Poblacion"], Quantity[1.1, "Transaccion"], 1},
Entity["Sedes", "Lago"] -> {Quantity[1, "Poblacion"], Quantity[2.8, "Transaccion"], 1},
Entity["Sedes", "Pangui"] -> {Quantity[1, "Poblacion"], Quantity[1.1, "Transaccion"], 1}};

counts = KnapsackSolve[UEA, Quantity[25, "Transaccion"]]

UEA[[All, 1]] counts

UEA[[All, 1]] counts; Total[%]

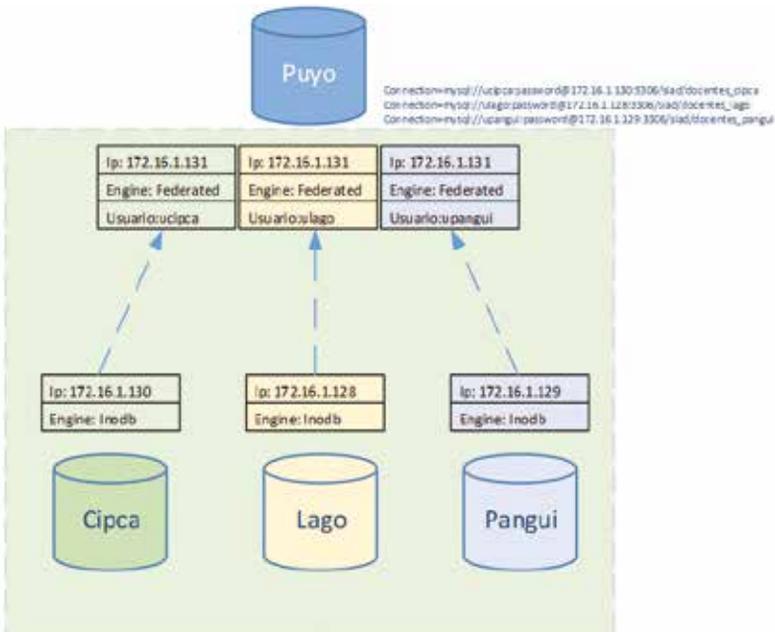
UEA[[All, 2]] counts      1

UEA[[All, 2]] counts; Total[%]

output[86]<= Puyo -> 3, lago ->2, cipca ->5, lago ->1 |>
output[87]<= Puyo -> 273 poblacion, lago ->142 poblacion, cipca ->52 poblacion, lago ->10 |>
output[88]= 1043 transaccion
output[89]<= Puyo -> 3, lago ->2, cipca ->5, lago ->1 |>
output[90]=24.54
    
```

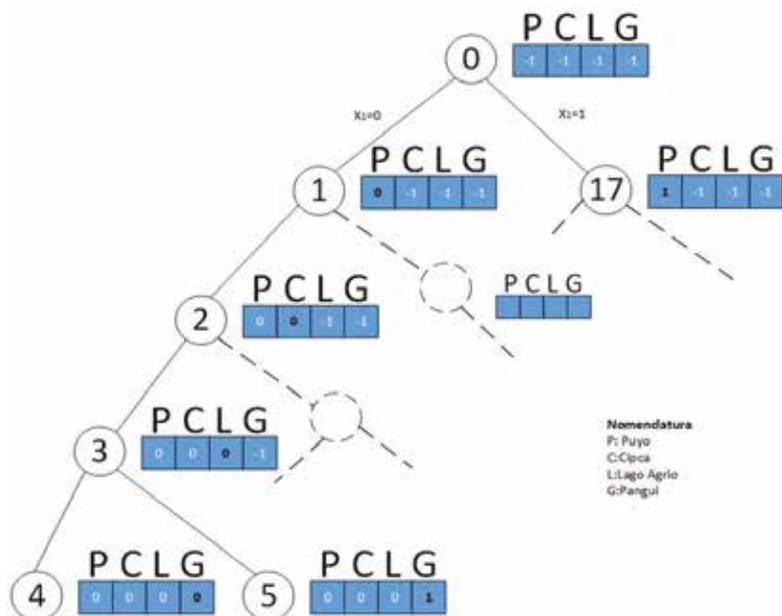
Fuente: Wolfram Development Platform

Gráfico 24
Modelo de bases de datos federadas



Elaboración: Autores

Gráfico 25
Generación del árbol de solución



Elaboración: Autores

Las pruebas realizadas en entorno simulado tienen el mismo comportamiento en un entorno de producción al realizar el ingreso de datos masivos de prueba, provocando intermitencias de conexión mediante el reinicio del servicio de networking, donde se observa que no existe afectación por cuanto el tratamiento de la data es local, y una vez restablecido el servicio, se actualiza el contenido de las mismas.

Incidencias de la implementación de la propuesta

Se implementa la solución en una base de datos de prueba alternativa conformada por un respaldo actualizado de información del sistema de producción denominado SIGPA, accedido desde la dirección sigpa.uea.edu.ec con vigencia limitada.

Cuadro 12
Comparativa de esquemas para el Flujo de información

Esquema	Disponibilidad	Trabajo administrativo extra	Incremento de espacio en disco	Renderización de distributivo
Centralizado	90%	10% - 15%	8.5%	Normal
Federado	100%	5%	7.8%	Normal

Elaboración: Autores

Mediante las pruebas de implementación de la propuesta, se evidencia un incremento significativo de productividad, por cuanto la solución de tablas federadas permitirá extender las funcionalidades del Sistema de Información Académico Docente, por cuanto el modelo de tablas permite guardar datos de las sedes y centro de investigación de forma local, permitiendo desplegar el aplicativo de control de asistencia mediante el registro biométrico de la huella digital, superando el problema de intermitencias del servicio de Internet.

El trabajo administrativo adicional disminuye significativamente debido a que, se deberá atender únicamente incidencias por fallas humanas como olvido en el cierre de sesión, la no planificación de una fecha, o por fallas propias de un equipo informático como cuelgue de sistema, infección por virus, entre otros.

Cuadro 13
Comparativa de esquemas en obtención de reportes

Esquema	Actividad docente	Gestión	Investigación	Documentación
Centralizado	90%	90%	90%	Normal
Federado	100%	100%	100%	Incremento 3%

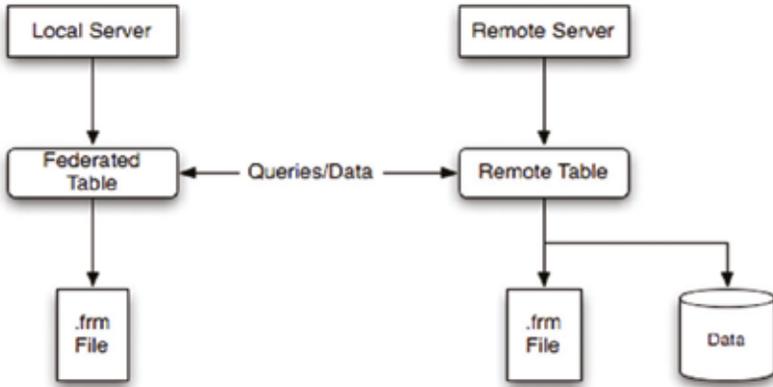
Elaboración: Autores

El registro de evidencias docentes mediante acceso federado disminuye la carga centralizada de archivos en el servidor académi-

co, con un ahorro de espacio, en tanto que no se muestran mejoras significativas en uno de los procesos que conlleva más tiempo correspondiente a la renderización del reporte para obtener el distributivo docente del periodo académico en curso.

Sobre la obtención de reportes, el acceso federado ofrece ventajas a la hora de obtener las evidencias docentes de cada sede a través del reporte de actividad mensual docente, informes de gestión e investigación por realizarse de forma local, sin embargo, representa un incremento del 3% en el tiempo de procesamiento centralizado, en dependencia de la disponibilidad de conexión por Internet, lo cual no es representativo porque afecta únicamente a la sede que no posea dicha conexión.

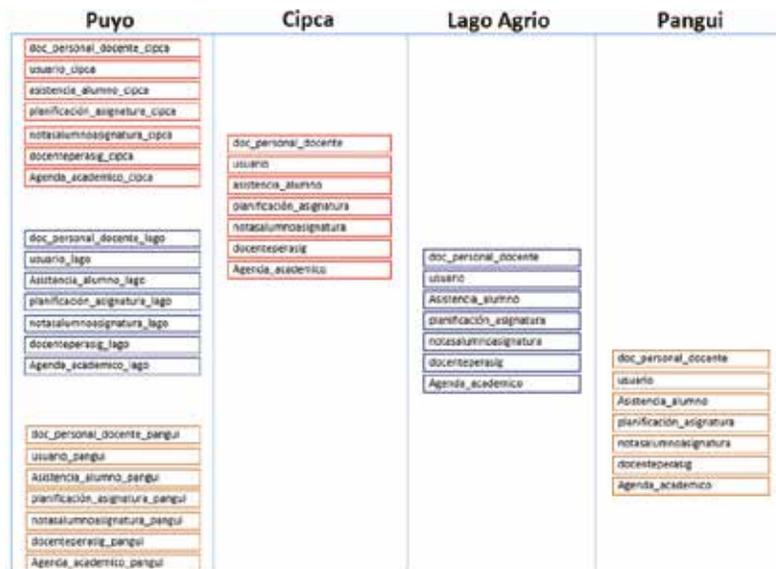
Gráfico 26
Funcionamiento de las tablas federadas



Fuente: <https://goo.gl/pTCUGn>

De acuerdo a la distribución de la figura se obtiene la siguiente estructura:

Gráfico 27
Distribución de Tablas



Elaboración: Autores

Pruebas de carga

El modelo propuesto permite registrar información de forma local en cada sede, correspondiente a cada periodo académico. Entre los procesos críticos, con mayor tiempo en la generación de resultados, se encuentra la cartilla de asistencia, que vincula una de las tablas con mayor cantidad de registros. Para la demostración de la prueba de carga, se utiliza un curso de prueba de la carrera de ingeniería agroindustrial, con las siguientes características: como punto de referencia se toma el estado actual y federado de las tablas de base de datos, en contraste con el nivel de concurrencia al aplicativo, caracterizado como alto medio y bajo según el número de usuarios conectados simultáneamente, el curso consta de 29 estudiantes, y se obtiene la cartilla de asistencia.

Cuadro 14
Comparativa de Tiempos

Estado	Concurrencia		
	Alta	Media	Baja
Centralizado	90s	60s	45s
Federado	50s	40s	30s

Elaboración: Autores

Según se muestra en la tabla de comparativa de tiempos se obtiene un porcentaje considerable de optimización del tiempo, lo que significa que existe una incidencia directa en la aplicación de la propuesta y el tiempo de respuesta.

CONCLUSIONES Y RECOMENDACIONES DE LA PROPUESTA TECNOLÓGICA

Una vez que se ha desarrollado la propuesta y comprobado el funcionamiento, se concluye que:

El método descriptivo utilizado sobre el funcionamiento de los sistemas y servicios informáticos institucionales, destaca al Sistema de Información Académico Docente SIAD como el responsable de la obtención de evidencias de acreditación, evaluación y seguimiento obtenida a partir de los reportes generados, con la particularidad de que el control de asistencia dactilar del estudiante, permite además hacer un seguimiento del silabo por parte del estudiante por cuanto en la pantalla de registro, se encuentra declarado el tema y el objetivo de la clase a dictarse.

El análisis sobre el despliegue de base de datos institucional permite determinar que la existe una carga de trabajo diaria considerable sobre la generación de evidencias durante el desarrollo de las actividades académica, de investigación, de gestión, entre otras, que permiten obtener información actualizada y pertinente a los requere-

rimientos de los organismos de control interno y gubernamentales para procesos de control, seguimiento y acreditación.

La adaptación del algoritmo KP (Knapsack Problem) para la estimación de la optimización del flujo de la información permite establecer el modelo de distribución de bases de datos aplicando, utilizando el motor de almacenamiento federado, lo cual incide positivamente en la disponibilidad de la información con la inclusión de los registros de las sedes y centro de investigación, comprobado en la generación de reportes y la continuidad y autonomía de los mismos al no tener un acceso centralizado dependiente de la conectividad a Internet.

A nivel nacional es necesario que se realicen los estudios por parte del gobierno ecuatoriano que permitan apoyar a las Instituciones de Educación Superior, en el aprovisionamiento y actualización tecnológica, por cuanto, son las mismas instituciones de Estado quienes exigen el cumplimiento de indicadores de acreditación, pero, el presupuesto asignado no siempre va acorde a la realidad de cada institución, lo cual desfavorece al avance y desarrollo institucional.

Es necesario realizar estudios sobre la implementación de soluciones de programación que permitan monitorear en tiempo real el flujo de la información en entornos federados, lo cual impulsaría a tomar en cuenta este tipo de motor de almacenamiento como alternativa de solución para entidades con características similares o de bajo presupuesto.

La variedad de algoritmos existentes, permite establecer mecanismos de comprobación de las soluciones planteadas, se debería tomar en cuenta su existencia como parte de las temáticas de estudio en las carreras de ingeniería en sistemas y su profundización y aplicación en los estudios de nivel superior.

BIBLIOGRAFÍA

- Agüero Torales, M. M. (s. f.). *Sistemas de Bases de Datos Federadas para la Gestión de la Información*.
- Álvarez, H., & Rigoberto, E. (2016). *Implementación de un sistema de autenticación federada, single sign on, para una empresa de medicina prepagada* (Tesis de posgrado). Guayaquil: Escuela Superior Politécnica del Litoral.
- Antonio, G., J., & De la Cuerda, D. (2013, octubre 7). *El modelo EVAM y la evaluación de la educación superior en Ecuador*. Universidad de Alicante.
- CEAACES (2015, marzo). Modelo genérico de evaluación del Entorno de Aprendizaje de Carreras Presenciales y Semipresenciales de las Universidades y Escuelas Politécnicas del Ecuador V2.0.
- CEDIA (s. f.). Miembros Plenos de CEDIA.
- CES (2013, octubre 30). RPC-SO-42-No.434-2013.
- Guevara Aulestia, D.O. & Cevallos Teneda, A. S. (2016). *Sistema de Federaciones de Identidades para la Facultad de Ingeniería en Sistemas, Electrónica e Industrial usando software de código abierto* (Tesis de posgrado). Universidad Técnica de Ambato, Ambato
- Marti, R. M. (2003). *Algoritmos heurísticos en optimización combinatoria*. Universidad de Valencia, Facultad de Ciencias Matemáticas.
- Ochoa Carrión, L. P. (2017). *Aplicación RIA basada en ontología para la gestión de procesos administrativos en la Universidad Estatal Amazónica*.
- Ortega Tenezaca, D. B., & Núñez Miranda, C. I. (2017). *Las bases de datos federadas y su incidencia en la optimización del flujo de información académica institucional universitaria* (Tesis de posgrado). Ambato: Universidad Técnica de Ambato.
- UEA. (s. f.). Centro de Investigación Posgrado y Conservación Amazónica.
- Wolfram, M. (s.f.). *Modelado de grafos y de redes: New in Mathematica 8*. <https://goo.gl/dwV18b>
- Wayner, P. (2001). *La ofensiva del software libre: cómo Linux y el movimiento del software libre se impusieron frente a los titanes de la alta tecnología*. Ediciones Granica S.A.
- Xu, F. (2009). Implementation of a Federated Search System: Resource Accessibility Issues. *Serials Review*, 35(4), 235-241. <https://doi.org/10.1080/00987913.2009.10765251>

Bases de datos autónomas almacenadas en SGBD heterogéneos y su incidencia en el tiempo de obtención de la ponderación de indicadores en autoevaluación de carreras de la ESPOCH

Oswaldo Villacrés Cáceres

FUNDAMENTACIÓN DEL PROBLEMA DE INVESTIGACIÓN

Las exigencias por almacenar grandes cantidades de información o datos y el incremento de archivos en las organizaciones, fueron algunos de los factores que favorecieron el desarrollo de las bases de datos en los años 60.

Actualmente, las bases de datos constituyen una de las herramientas más ampliamente utilizadas en todos los campos a nivel científico, académico, social, económico, político y cultural a nivel mundial (Simari *et al.*, 2017).

La estructura organizacional de las instituciones, ha ocasionado la proliferación de bases de datos autónomas almacenadas en

diferentes Sistemas Gestores de Bases de Datos (SGBD) dispersas en las distintas unidades que conforman la organización, y con ello, la necesidad de acceder a los datos de los distintos sistemas a través de una interfaz común. Para solucionar estas necesidades y ofrecer un adecuado acceso a bases de datos autónomas, en los años 90, surgen los sistemas gestores de bases de datos federadas (SGBDF), aunque, el concepto viene de Hammer y McLeod en 1979.

En Europa, Asia y Estados Unidos existen diversidad de sistemas informáticos que soportan sistemas gestores de bases de datos federadas (Codina, 2010) (Verger & Noguera, 2005) (Saavedra Places, 2003) (Crespo, Maldonado, Cano, Robles, & Casanova, 2002) (Saltor, 2000)

En América Latina, países como Argentina, Brasil, Colombia, Venezuela, Chile, México, entre otros, cuentan con aplicaciones que utilizan sistemas gestores de bases de datos federadas (González & Flores, 2013) (Vásquez Tapia & Razo Rodríguez, 2014).

En Ecuador, se ha ido incursionando en la implementación de sistemas informáticos que utilizan bases de datos federadas, un ejemplo de ello, el sistema integral de gestión hospitalaria en el Hospital Abel Gilbert Pontón de Guayaquil.

En la Escuela Superior Politécnica de Chimborazo (ESPOCH), la utilización de diversos SGBD para el almacenamiento y gestión de datos, ha originado, que no se pueda disponer de información centralizada, precisa y en su debido momento, para llevar a cabo los procesos de autoevaluación de carreras. Además, debido a la gran cantidad de información y evidencias requeridas, se han establecido mecanismos poco eficientes para su almacenamiento y gestión, ocasionando inconsistencia, duplicación y pérdida de evidencias. En consecuencia, afectando al normal desarrollo de los procesos de autoevaluación de carreras, que permitan conocer la ponderación de los criterios e indicadores propuestos en el modelo genérico de evaluación de carreras. Y a su vez, que faciliten la toma de decisiones oportunas y adecuadas en pos de la mejora continua y consolidación de los indicadores.

Se presenta en el Gráfico 1, las diversas causas que originan el problema y sus respectivos efectos:

Gráfico 1
Árbol del problema



Elaboración: Autor

METODOLOGÍA DE INVESTIGACIÓN

Se utiliza enfoque cuali-cuantitativo. Cuantitativo, pues se identificará la problemática aplicando encuestas, para finalizar con un análisis estadístico. Cualitativo, debido que, de la información recopilada se pretende establecer juicios de valor respecto a cada una de las preguntas recabadas.

La investigación es de tipo descriptiva y explicativa, ya que se caracterizará un problema en concreto indicando sus principales causas y se darán a conocer sus resultados. Será también, bibliográfica o documental porque utilizará fuentes como: artículos, libros, revistas, documentos, entre otros.

Se trabajará con la población total, como se detalla en la Tabla 1. La aplicación de encuestas se realizó a docentes coordinadores de las

Comisiones de Evaluación y Aseguramiento de la Calidad de las Carreras (CEACC). La encuesta fue estructurada por preguntas cerradas dicotómicas, mixtas y de elección múltiple, considerando indicadores establecidos en el modelo genérico de evaluación de carreras definido por el Consejo de Evaluación, Acreditación y Aseguramiento de la Calidad (CEAACES). Mientras, la entrevista fue aplicada a los administradores de los sistemas informáticos y de bases de datos.

Tabla 1
Población y muestra

Personal	Número	Porcentaje
Docentes coordinadores CEACC	28	84,85
Administradores de BD	5	15,15
TOTAL	33	100

Elaboración: Autor

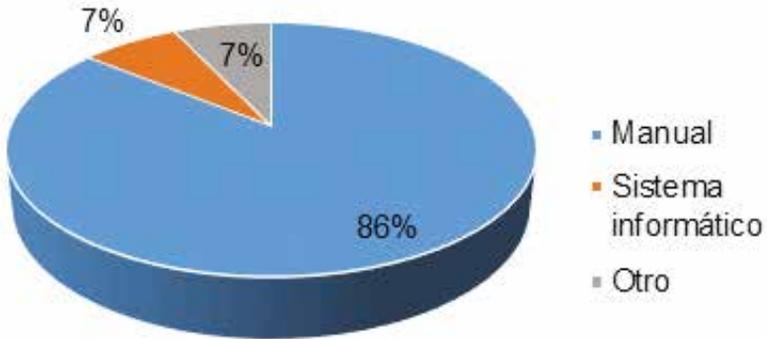
ANÁLISIS E INTERPRETACIÓN DE RESULTADOS

De los resultados obtenidos a través de las encuestas aplicadas a docentes coordinadores de Comisión de Evaluación y Aseguramiento de la Calidad de Carrera (CEACC) de la ESPOCH, utilizando formularios digitales a través de GoogleDrive (se puede acceder utilizando el enlace <http://goo.gl/forms/3QEHdn1Rmi>) se destacan los siguientes:

En el Gráfico 2, el 86% de docentes miembros de CEACC indican haber utilizado mecanismos manuales para dar seguimiento al proceso de autoevaluación, sólo el 7% mencionaron haber utilizado un sistema informático y también Otro tipo (físico y digital) de elemento para dicho proceso.

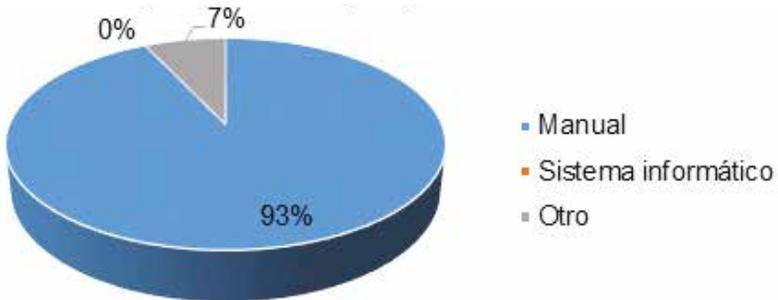
En cuanto a la forma de recopilar datos y organizar evidencias, se puede observar en el Gráfico 3, que 93% de docentes miembros de CEACC realizan de forma manual. Tan solo, 7% realiza a través de Otro mecanismo (manual y digital).

Gráfico 2
Mecanismos para seguimiento autoevaluación



Elaboración: Autor

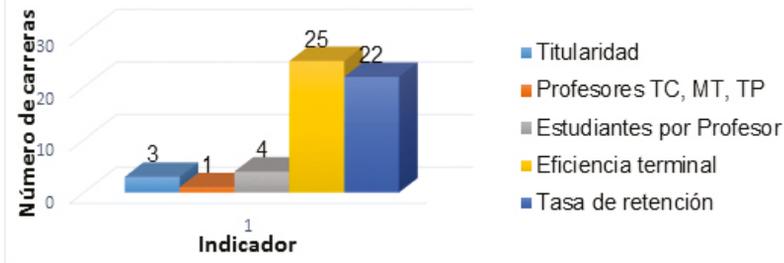
Gráfico 3
Recopilación datos y organización evidencias



Elaboración: Autor

De acuerdo con los resultados presentados en el Gráfico 4, los indicadores con mayor complejidad para su cálculo son Eficiencia terminal y Tasa de retención, esto debido a la documentación que es necesaria recopilar y los datos que deben ser generados para la valoración de mencionados indicadores.

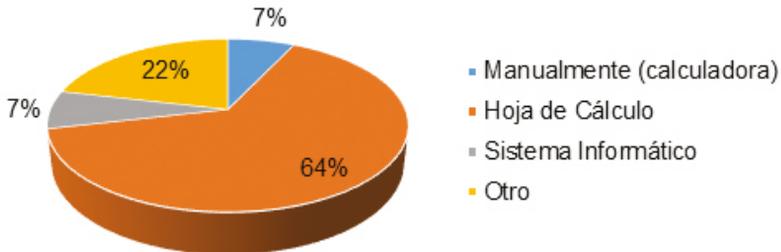
Gráfico 4
Indicador con mayor dificultad



Elaboración: Autor

Como se observa en el Gráfico 5, el mecanismo tecnológico más utilizado para valorar indicadores es la Hoja de Cálculo, con un 64%. Mientras, tan solo 7% lo realiza de forma manual y otro 7% lo realiza a través de un Sistema Informático.

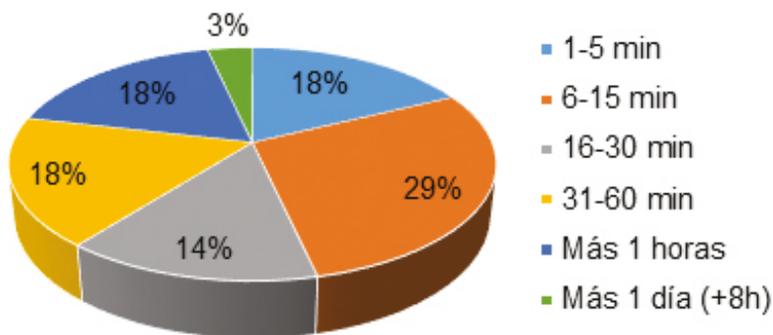
Gráfico 5
Mecanismos tecnológicos para valorar indicadores



Elaboración: Autor

Se observa en el Gráfico 6, que un 29% de comisiones destina entre 6-15 minutos para calcular la ponderación del indicador Titularidad. Mientras, al 3% de las CEACC, tardan más de 1 día (más de 8 horas) en desarrollar esta actividad.

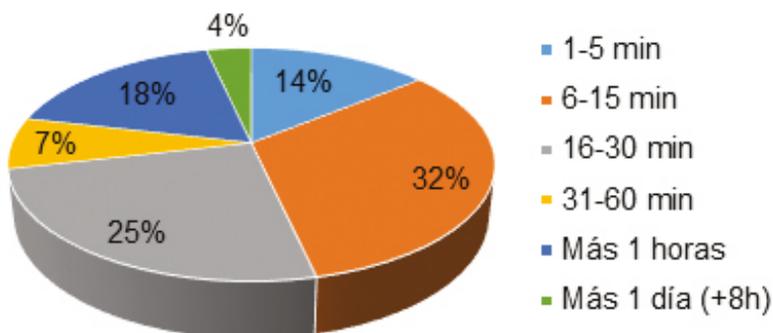
Gráfico 6
Cálculo “Titularidad”



Elaboración: Autor

Como presenta el Gráfico 7, a criterio del 32% de CEACC, el tiempo promedio para calcular la ponderación del indicador Profesores TC, MT, TP está entre 6-15 minutos. Mientras, el 25% consideran que es de 16-30 minutos.

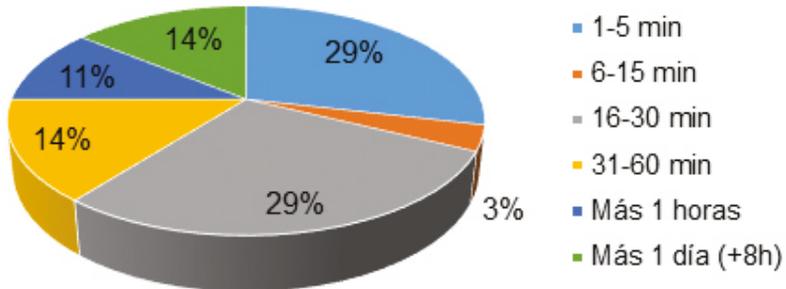
Gráfico 7
Cálculo “Profesores TC, MT, TP”



Elaboración: Autor

Para calcular el indicador Estudiantes por profesor, el 29% tardan entre 1-5 minutos al igual que entre 16-30 minutos. Solo, el 3% demora entre 6-15 minutos. Ver Gráfico 8.

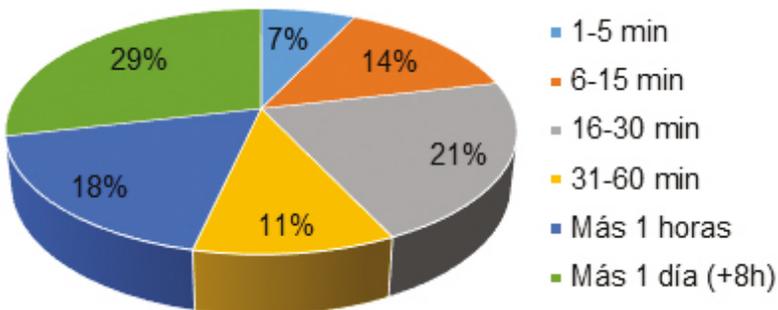
Gráfico 8
Cálculo “Estudiantes por profesor”



Elaboración: Autor

En el cálculo del indicador Eficiencia terminal, el 29% de los encuestados dedica más de un día. El 7%, tarda de 1-5 minutos, como se puede observar en el Gráfico 9.

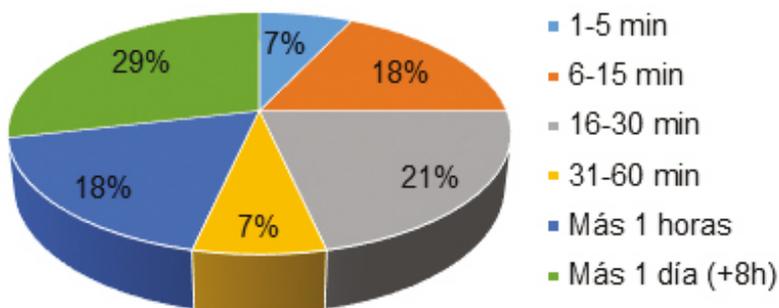
Gráfico 9
Cálculo “Eficiencia terminal”



Elaboración: Autor

Se puede observar en el Gráfico 10, para calcular el indicador Tasa de retención, el 29% de los encuestados dedica más de 8 horas. Tan solo, 7% tarda 1-5 minutos y en igual porcentaje de 31-60 minutos.

Gráfico 10
Cálculo “Tasa de retención”



Elaboración: Autor

De las entrevistas realizadas a los diferentes administradores de los sistemas de información y administradores de bases de datos, se pudo conocer que la institución cuenta con 18 sistemas informáticos y un total de 55 bases de datos.

Un 85,45% de bases de datos son autónomas, es decir, 47 de las 55 existentes. Adicionalmente, se pudo constatar que existe heterogeneidad en los SGDB utilizados para administrar las bases de datos, entre los SGDB empleados están Oracle XE, SQL Server en diferentes versiones, PostgreSQL y MySQL.

La Tabla 2, presenta mayor detalle de lo mencionado anteriormente y de cada uno de los sistemas informáticos existentes en la ESPOCH, el SGDB de las bases de datos y los sistemas informáticos con los que se integran o comparten datos.

Tabla 2
Descripción de Sistemas informáticos existentes en la ESPOCH

N.	Nombre del Sistema informático	Dependencia que utiliza	SGBD Versión	S. I. accede a datos de la BD
1	Sistema académico grado OASis (Una base de datos master y una por cada carrera)	Comunidad Politécnica	SQL Server 2012	3, 5, 8, 10, 16
2	Sistema para pagos de matrículas	Secretarías académicas, Estudiantes, Banco del Pacífico	SQL Server 2000	-
3	Sistema médico para reservar turnos	Comunidad Politécnica, Departamento Médico	SQL Server 2000	-
4	Seguimiento de proyectos	Dirección de investigación, Docentes, Autores	Postgre SQL 9.3	-
5	Seguimiento a graduados	Vinculación y Estudiantes	Postgre SQL 9.3	-
6	Seguimiento de eventos	Dirección de investigación	Postgre SQL 9.3	-
7	Empadronamiento de graduados	Secretaría general	Postgre SQL 9.3	-
8	Migración de moodle	Facultades, Postgrado	Postgre SQL 9.3	-
9	Sistema de recaudación	Tesorería institucional	Postgre SQL 9.3	13
10	Centralizada	Comunidad politécnica	Postgre SQL 9.3	2, 8, 9, 13
11	Sitio Web institucional	Público en general	Postgre SQL 9.3	-
12	Sistema de talento humano	Dirección de talento humano, Comunidad politécnica	MySQL 5.0.9	-
13	Comprobantes electrónicos	Comunidad politécnica, Público en general	Postgre SQL 9.3	-

14	Sistema para el control de ticket	Comunidad Politécnica	SQL Server 2012	-
15	Sistema de evaluación al desempeño docente	Estudiantes y Docentes.	SQL Server 2000 y Oracle XE	-
16	Sistema para el control de sílabos	Docentes, Estudiantes, Autoridades	Postgre SQL 9.3	-
17	Moodle	Estudiantes y docentes de la institución	Postgre SQL 9.3	-
18	SINFO	Dirección financiera	SQL Server 2008	-

Fuente: Administradores de sistemas informáticos y BD de la ESPOCH.

CONCLUSIONES Y RECOMENDACIONES

Conclusiones

- El 96% de CEACC, dedican tiempos excesivos para el cálculo de indicadores: Titularidad, Eficiencia terminal, Tasa de retención, Estudiantes por profesor y Profesores TC, MT, TP.
- Los indicadores con mayor dificultad para calcular son: Eficiencia terminal y Tasa de retención.
- El 29% de los docentes miembros de CEACC, dedican más de 8 horas para el cálculo de los indicadores Eficiencia terminal y Tasa de retención.
- El 29% de comisiones tardan entre 6-15 minutos para calcular el indicador Titularidad.
- En el cálculo del indicador Profesores TC, MT, TP, el 32% demora entre 6-15 minutos, y el 25% demora entre 16-30 minutos.
- Para calcular el indicador Estudiantes por profesor, el 29% de los miembros de CEACC necesita entre 16-30 minutos.

- La institución cuenta con un 85,45% de bases de datos autónomas.

Recomendaciones

- Es necesario contar con un sistema que facilita el procesamiento y almacenamiento de información necesaria para cada criterio e indicador.
- Es necesario contar con un sistema informático que permita calcular de manera eficiente y eficaz, la valoración de los indicadores (en dependencia de los datos disponibles en los sistemas de información de la institución) del modelo genérico de evaluación de carreras.
- El sistema informático a ser implementado, debe permitir integrar las bases de datos existentes.
- Debido a la existencia de bases de datos autónomas almacenadas en Sistemas Gestores de Bases de Datos Heterogéneas, se recomienda implementar un sistema informático que permita mantener la autonomía de cada uno de los sistemas existentes, a través de un sistema de bases de datos federadas.

PROPUESTA

Datos informativos

La propuesta se basa en implementación de un sistema informático que haga uso del sistema de bases de datos federadas para la integración de las bases de datos autónomas almacenadas en SGBD heterogéneas.

El sistema fue implementado en la Dirección de Evaluación y Aseguramiento de la Calidad (DEAC), Escuela Superior Politécnica de Chimborazo (ESPOCH), la cual tiene su origen en el Instituto tecnológico Superior de Chimborazo, creado mediante Ley No.6090, expedida por el Congreso Nacional, el 18 de abril de 1969. El cambio de de-

nominación a Escuela Superior Politécnica de Chimborazo ESPOCH, se produce mediante Ley No. 1223 del 29 de octubre de 1973 publicada en el Registro Oficial N° 425 del 6 de noviembre del mismo año.

Antecedentes de la propuesta

La Escuela Superior Politécnica de Chimborazo, cuenta con bases de datos autónomas almacenadas en SGBD heterogéneas. Las bases de datos disponen de datos suficientes para calcular indicadores: Titularidad, Profesores TC-MT-TP, Estudiantes por Profesor, Eficiencia Terminal y Tasa de retención.

Aunque, en la ESPOCH se ha constituido la Comisión de Evaluación y Aseguramiento de la Calidad en cada Carrera, conformada por cuatro docentes, quienes son los responsables de llevar a cabo el proceso de autoevaluación de la carrera y esto implica el calcular los indicadores del modelo de evaluación. La principal dificultad para calcular indicadores, conlleva un proceso de verificación y validación manual de información, provocando esfuerzo humano y en especial dedicar al menos 4 horas a la semana por parte de los docentes miembros de las CEACC para mencionado proceso.

Adicionalmente, los indicadores del modelo de evaluación genérico, son calculados en base a datos registrados manualmente en una hoja de cálculo, necesarios para la generación de los informes de autoevaluación de carrera.

Justificación

La propuesta es relevante, porque busca la aplicación de bases de datos federadas como mecanismo tecnológico para la integración de datos de bases de datos autónomas almacenadas en SGBD heterogéneas existentes en la Escuela Superior Politécnica de Chimborazo.

La importancia de la aplicación de bases de datos federadas para integrar datos optimiza el tiempo dedicado en la verificación y

validación de datos necesarios para calcular indicadores (Titularidad, Profesores TC-MT-TP, Estudiantes por Profesores, Eficiencia Terminal y Tasa de retención) del modelo genérico de evaluación de carreras.

Fue factible la propuesta ya que la institución cuenta con los recursos bibliográficos, tecnológicos, económicos y humanos necesarios para la solución del problema.

Fundamentación

La propuesta está enfocada al acceso integrado de las bases de datos autónomas almacenadas en SGBD heterogéneas existentes en la ESPOCH, para facilitar el cálculo de indicadores del modelo genérico de evaluación de carreras, para lo cual la tecnología a utilizar son las bases de datos federadas, siendo esencial la fundamentación científica técnica de mencionada tecnología.

Características

De acuerdo a Espino Barrios (2009), menciona ciertas características que debe cumplir un sistema de base de datos federadas, y se mencionan a continuación:

- *Distribución.* Los datos están localizados en múltiples bases de datos.
- *Heterogeneidad.* Existen diferencias o se debe permitir diferencias en el software, hardware y/o sistema de comunicación.
- *Heterogeneidad semántica.* Existe discrepancia respecto al significado, interpretación o pretensión de utilización de los datos o datos relacionados.
- *Autonomía.* Cada sistema de base de datos tiene un control separado e independiente.

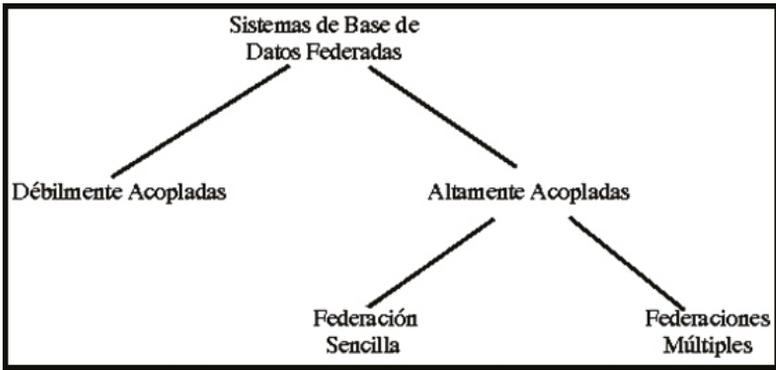
Sin embargo, existen características adicionales como menciona Aslan & McLeod (1999), respecto a la interoperabilidad, que

es una condición mediante la cual los sistemas heterogéneos interactúan entre sí, lo cual se alcanza por medio de la capacidad que tiene los componentes para compartir e intercambiar unidades de información de manera activa y cooperante con otros componentes del sistema de base de datos federadas.

Taxonomía

En 1990, Sheth & Larson, proponen la taxonomía presentada en el Gráfico 11, esta taxonomía se centra en la dimensión de la autonomía.

Gráfico 11
Taxonomía de sistemas de base de datos federadas



Fuente: Sheth & Larson (1990)

En el Gráfico 11, se pueden distinguir dos grandes categorías de sistemas de base de datos federadas: débilmente acopladas y altamente (fuertemente) acopladas.

Sistemas de base de datos federadas débilmente acoplados

Un sistema de bases de datos federadas es débilmente acoplado, si la responsabilidad de crear y mantener la federación recae en

el usuario y no hay control por parte del sistema federado y sus administradores. Para Litwin, Mark, & Roussopoulos (1990), este mismo concepto está relacionada a multibases de datos o bases de datos interoperables. En donde, los usuarios necesitan acceder a múltiples datos sin el beneficio de un esquema global y que el componente esencial de un sistema de este tipo es el lenguaje usado para manejar las bases de datos participantes. Otro requerimiento importante es que el usuario debe ser capaz de formular manipulaciones de multibases de datos no procedurales en la ausencia de un esquema global. El usuario es responsable de comprender la semántica de los objetos en los esquemas de exportación y resolver la heterogeneidad de los sistemas manejadores de base de datos y de la semántica.

Sistemas de base de datos federada fuertemente acoplados

Un SBDF es fuertemente acoplado, si su administrador tiene la responsabilidad de crear y mantener la federación y el control de acceso a los sistemas de bases de datos componentes. Una federación está compuesta por una integración selectiva y controlada de sus componentes. La actividad de desarrollar un SBDF fuertemente acoplado consiste en la creación de un esquema federado sobre el cual las operaciones (consultas y/o actualizaciones) son ejecutadas.

Un SBDF fuertemente acoplado puede tener uno o más esquemas federados; se dice, que tiene una federación sencilla si permite la creación y manejo de solamente un esquema federado. Tener un esquema federado sencillo ayuda a mantener la uniformidad en la interpretación semántica de los datos integrados, en tanto, un SBDF fuertemente acoplado se dice que tiene una federación múltiple si permite la creación y manejo de múltiples federaciones.

Un SBDF fuertemente acoplado provee localización, duplicación y transparencia de distribución. Esto es llevado a cabo al desarrollar un esquema federado que integra múltiples esquemas de

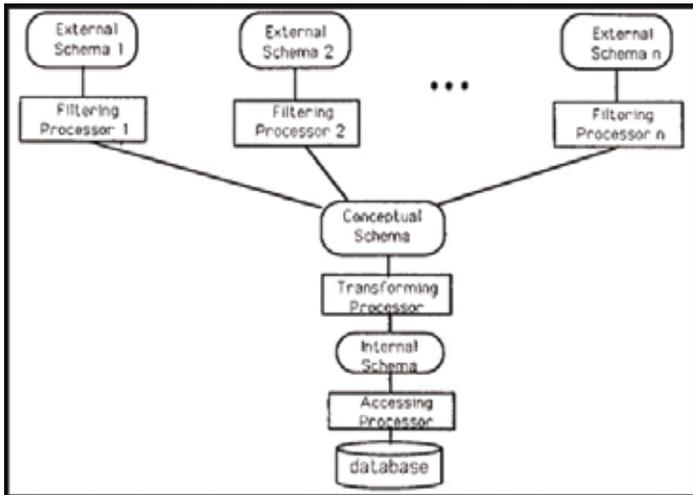
exportación. Las transparencias son manejadas por los mapeos entre el esquema federado y los esquemas de exportación, y un usuario de la federación puede hacer consultas a través de un *lenguaje de consultas clásico* contra el esquema federado con la ilusión de que se está accediendo a un solo sistema (Muñoz & Aguilar, 2011).

Arquitectura

Existen diversas arquitecturas para el manejo de bases de datos federadas, sin embargo, a continuación se presentan dos de las arquitecturas tradicionales; la arquitectura propuesta por el ANSI/SPARC, que generalmente se la aplica para bases de datos centralizadas y puede ser amplia para su utilización en sistemas de bases de datos federadas, y la arquitectura de Amith Sheth y James Larson.

Arquitectura ANSI/SPARC

Gráfico 12
Arquitectura ANSI/SPARC

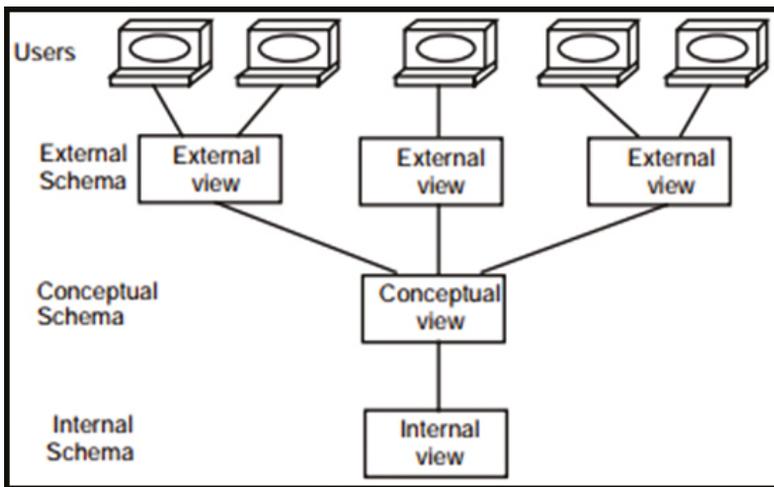


Fuente: Sheth & Larson (1990)

Como se observa en el Gráfico 12, la arquitectura ANSI/SPARC propuesta en 1975, es un estándar de diseño abstracto para un Sistema de Gestión de Base de Datos (SGBD).

Una representación más comprensible se presenta en el Gráfico 13.

Gráfico 13
Representación arquitectura ANSI/SPARC



Fuente: Ôzsu & Valduriez (2011)

La arquitectura ANSI/SPARC considera tres niveles:

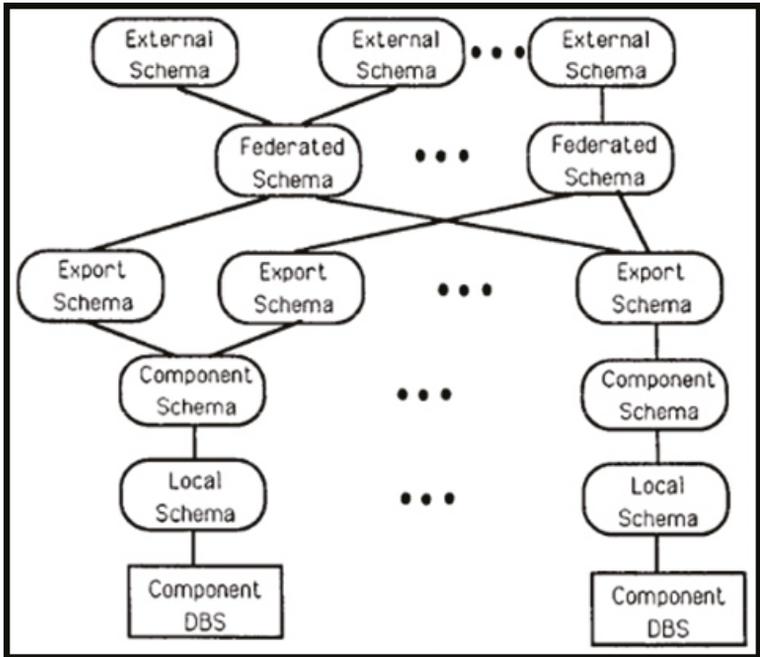
- Nivel externo. También denominada, vista de usuario o nivel lógico, describe una parte de la base de datos que es relevante para un usuario en particular. Excluyendo datos irrelevantes y datos que el usuario no está autorizado a acceder.
- Nivel conceptual. Es una manera de describir que datos son almacenados en la base de datos y cómo los datos están relacionados. En este nivel, no se especifica cómo se almacenan físicamente los datos.

- Nivel interno. Es el nivel más bajo, contiene definiciones de los registros almacenados, los métodos de representación, los campos de datos e índices. Especifica qué y cómo son almacenados los datos. Sólo hay un esquema interno por base de datos.

Arquitectura de Amith Sheth y James Larson

Como se observa en el Gráfico 14, la arquitectura propuesta por Amith Sheth y James Larson (1990), contempla 5 niveles de esquemas: esquema local, esquema componente, esquema de exportación, esquema federado y esquema externo.

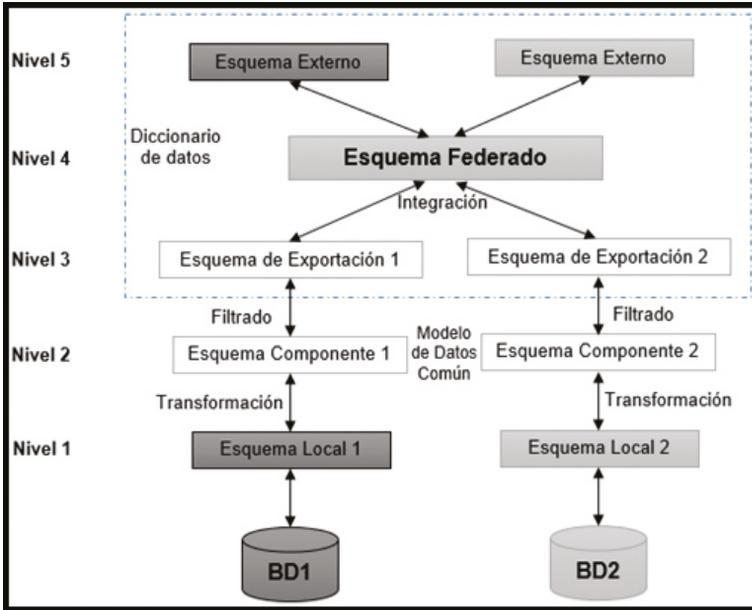
Gráfico 14
Arquitectura de sistemas de base de datos federadas



Fuente: Sheth & Larson (1990)

Para mayor comprensión, se ha elaborado en el Gráfico 15, una representación más clara de la arquitectura propuesta por Sheth y Larson.

Gráfico 15
Representación de la arquitectura de SBDF de Sheth y Larson



Fuente: Velásquez (2014)

A continuación, se describen cada uno de los niveles, según Espino (2009):

- **Esquema local.** Es un esquema conceptual de los componentes. Un esquema local está expresado en el modelo nativo de datos del componente del sistema manejador de base de datos, y por lo tanto, diferentes esquemas locales se pueden expresar en distintos modelos de datos.

- Esquema componente. Es derivado de trasladar el esquema local en un modelo de datos llamado Modelo de Datos Común (MDC) o Modelo Canónico, este modelo, facilita las tareas de negociación e integración realizados en el desarrollo de un SBDF fuertemente acoplado, o, facilita la negociación y especificación de vistas y consultas a múltiples bases de datos en un SBDF débilmente acoplado.
- Esquema de exportación. Representa un subconjunto de la totalidad de los datos que contiene el esquema componente. Este subconjunto de datos es el que se quiere compartir en la base de datos federada. No todos los datos de un sistema de base de datos componente podrá estar disponible a la federación y los usuarios. El propósito de definir esquemas de exportación es para facilitar el control y administración de la autonomía asociada.
- Un procesador de filtrado se puede utilizar para proporcionar el control de acceso como se especifica en un esquema de exportación mediante la limitación del conjunto de operaciones permitidas que se pueden presentar en el esquema de componente correspondiente. Estos procesadores de filtrado y los esquemas de exportación son compatibles con la función de la autonomía de un FDBS.
- Esquema federado. Es una integración de múltiples esquemas de exportación de cada base de datos componente. Un esquema federado también incluye la información sobre la distribución de los datos que son generados cuando se realiza la integración de esquemas de exportación. Pueden existir múltiples esquemas federados en un sistema de base de datos federado.
- Conceptos similares a esquema federado, representan los términos esquema importado, esquema global, esquema conceptual global, esquema unificado y esquema empresarial.
- Esquema externo. Representa un esquema para un usuario y/o aplicación o una clase de usuarios y/o aplicaciones.

Metodología para el desarrollo de un sistema de base de datos federada

Sheth & Larson, mencionan “el desarrollo de un nuevo sistema de base de datos federada, consiste principalmente en la integración de las bases de datos componentes existentes”. Proponen dos metodologías para el desarrollo de un sistema de base de datos federada: bottom-up (de abajo hacia arriba), puede ser utilizado para este propósito o para agregar un nuevo componente de base de datos de un sistema de base de datos federada; y, el proceso top-down (de arriba hacia abajo) es utilizado cuando nuevas aplicaciones son desarrolladas usando sistemas de base de datos federada existentes, que es una extensión del proceso tradicional de diseño de base de datos distribuido.

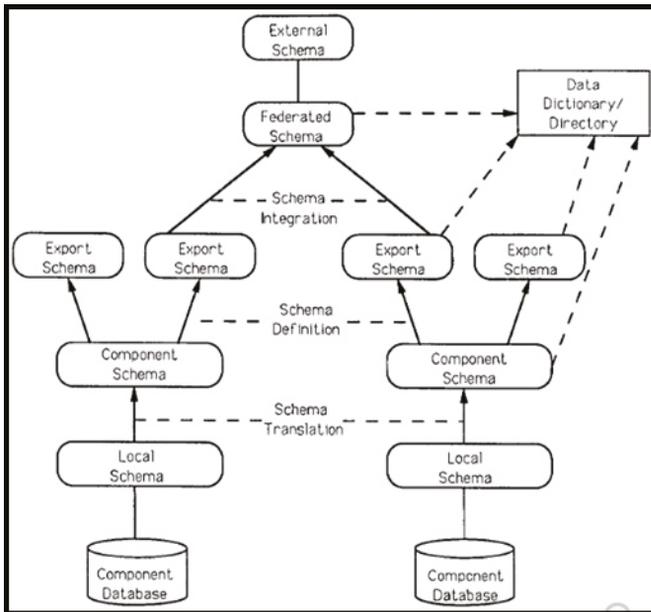
Proceso de desarrollo Bottom-Up

El proceso de desarrollo Bottom-Up, se utiliza para integrar varias bases de datos existentes para desarrollar un sistema de base de datos federada. El Gráfico 16, representa gráficamente el proceso Bottom-Up que se detalla a continuación:

- Traducir esquemas. Traducir el esquema local de una base de datos componente en un esquema componente expresado en el MDC. Generar las asignaciones entre los objetos en los dos esquemas. Desarrollar (o identificar si ya existe) el procesador de transformación que puede transformar comandos expresados en el esquema de componente en los comandos expresados en el esquema local correspondiente.
- Definir esquemas de exportación. Definir esquemas de exportación a partir de un esquema de componentes. Este paso se lleva a cabo por los administradores (DBAs componentes) de la respectiva base de datos componente para autorizar la parte de sus bases de datos que se incluirán en el FDDBS basado en las negociaciones con el DBA de la federación.

- Integración de esquemas. Seleccione un conjunto relacionado de esquemas de exportación para integrarse e integrarlos. La integración de cada conjunto de esquemas de exportación producirá un esquema federado. Desarrollar (o identificar si ya existe) un procesador de la construcción que transformaría los comandos expresados en esquemas federados en órdenes expresadas en los esquemas de exportación correspondientes. Esto incluye las asignaciones de mapeo con información de distribución apropiado.
- Definir esquemas externos. Si es necesario, se debe definir esquemas externos para cada usuario federado o grupos de usuarios federados. Construir o identificar los procesadores de transformación y filtrado necesarios. El procesador de transformación, realiza la traducción de esquema si el modelo de datos del esquema externo es diferente que el MDC.

Gráfico 16
Proceso Bottom-Up para desarrollo de SBDF



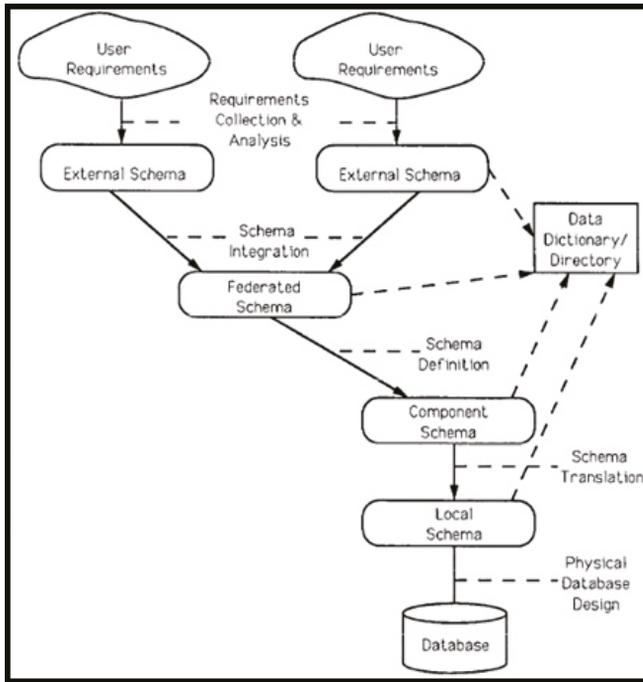
Proceso de desarrollo Top-Down

El proceso de desarrollo Top-Down (de arriba hacia abajo), se utiliza cuando un SBDF ya existe y se quieren incorporar nuevos requisitos de usuarios (por ejemplo, para soportar nuevas aplicaciones). El Gráfico 17 representa gráficamente el proceso Top-Down que se detalla a continuación:

- Definir o modificar el esquema externo. Reunir los requisitos del usuario de la federación y analizarlos para definir esquemas externos o extensiones al esquema externo existente.
- Analizar esquemas. Comparar esquemas federados pertinentes con los esquemas externos para identificar partes de los esquemas externos que ya están en el esquema federado y, por tanto, son compatibles con el SBDF. Si alguna parte de un esquema externo no está soportado por el SBDF, tendrá que ser extendido o desarrollado para incluir esa parte de un esquema federado; esta parte no soportada como un esquema temporal (que se elimina al final del proceso de integración) . Una o más bases de datos componentes tendrá que soportar el esquema temporal. Esto se puede lograr en una de tres maneras:
 - Existen los datos requeridos en una o más bases de datos componente. En este caso, identificar los esquemas componente que contienen la descripción de los datos necesarios y negociar con sus administradores para tener una descripción de estos datos se ubican en un esquema de exportación con los derechos de acceso adecuados.
 - Los datos requeridos no se implementan en cualquier base de datos componente y un administrador de base de datos componente (DBA) está dispuesto a situar los datos requeridos en su base de datos componente. En este caso, los esquemas local, componente y de exportación de la base de datos pertinente se modifican.

- Los datos requeridos no se implementan en cualquier base de datos componente, y ningún DBA componente está dispuesto a colocar los datos requeridos en sus bases de datos componente. En este caso, el esquema temporal se implementa como una base de datos separada de un DBMS componente existente. Alternativamente, un nuevo DBMS componente puede ser usado en un nuevo procesador de transformación. El esquema temporal se convierte en un nuevo esquema de componentes.
- Integrar esquemas. Integrar el esquema temporal con el esquema federal relevante y descartar el esquema temporal.

Gráfico 17
Proceso Top-Down para desarrollo de SBDF



METODOLOGÍA

Para llevar a cabo la presente propuesta, se ha tomado la opción de adoptar el Proceso Unificado de Racional, RUP (Rational Unified Process), con UML (Unified Modeling Language) como complemento ideal para integrar una metodología acorde a las necesidades del proyecto.

Modelo operativo

Análisis de bases de datos

Dado que es necesario utilizar los datos existentes en bases de datos de la ESPOCH, se analizaron todas las bases de datos, en busca de datos que faciliten el cálculo de indicadores: Titularidad, Profesores TC-MT-TP, Estudiantes por Profesor, Eficiencia Terminal y Tasa de retención.

Gráfico 18
Diagrama graduación del modelo físico BD OASis

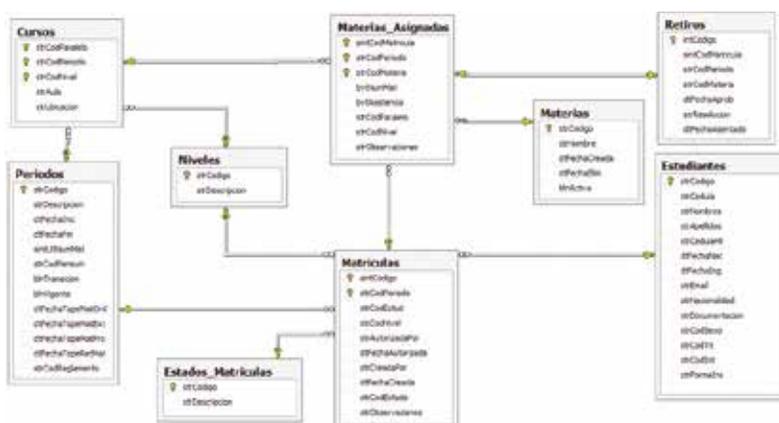


Fuente: Dirección de Tecnologías de la Información ESPOCH

Del análisis realizado para establecer el esquema local de la base de datos federada, se seleccionaron: 28 bases de datos autó-

nomas (una por cada carrera) utilizadas por el Sistema Académico OASis, y una base de datos autónoma utilizada por el Sistema para Administración de Recursos Humanos (SARH); las bases de datos se encuentran gestionadas con SQL Server 2013 y MySQL 5.0.9 respectivamente. Los gráficos siguientes, representan parte del modelo físico de la base de datos utilizada por el Sistema Académico OASis.

Gráfico 19
Diagrama matriculación del modelo físico BD OASis



Fuente: Dirección de Tecnologías de la Información ESPOCH

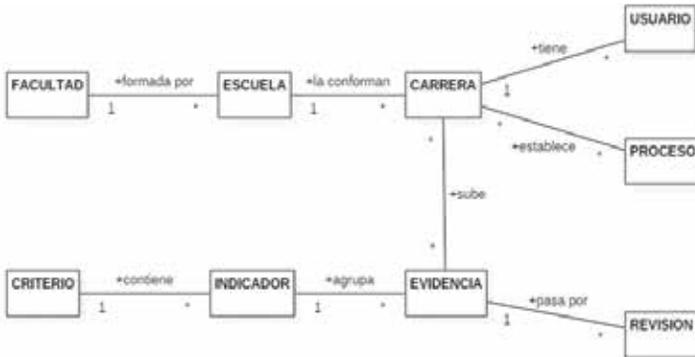
En el Gráfico 21, se presenta parte del modelo físico de la base de datos utilizada por el Sistema para Administración de Recursos Humanos (SARH).

Documentación técnica

La Documentación Técnica, presenta y detalla algunas de las disciplinas, como Requerimientos, Análisis y Diseño, e Implementación, que forman parte de la metodología de desarrollo RUP (Rational Unified Process); en la cual se puede revisar el análisis de requerimientos del sistema.

MODELO CONCEPTUAL

Gráfico 23
Modelo conceptual del SISAC



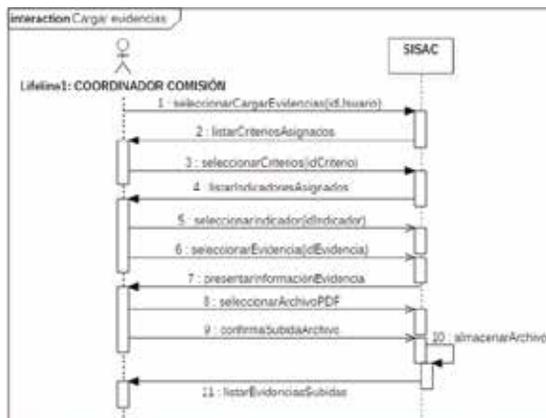
Elaboración: Autor

DIAGRAMAS DE INTERACCIÓN

Diagramas de secuencia

Cargar evidencias

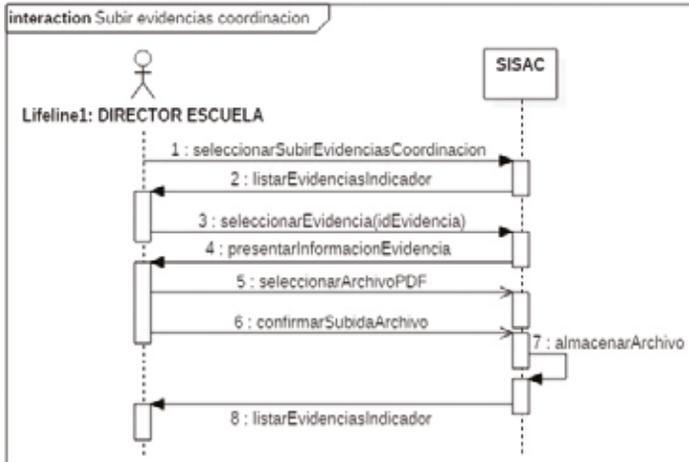
Gráfico 24
Diagrama de secuencia “Cargar evidencias”



Elaboración: Autor

Subir evidencias coordinación

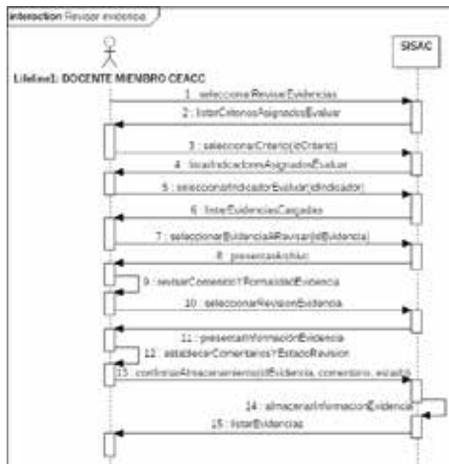
Gráfico 25
Diagrama de secuencia “Subir evidencias coordinación”



Elaboración: Autor

Revisar evidencia

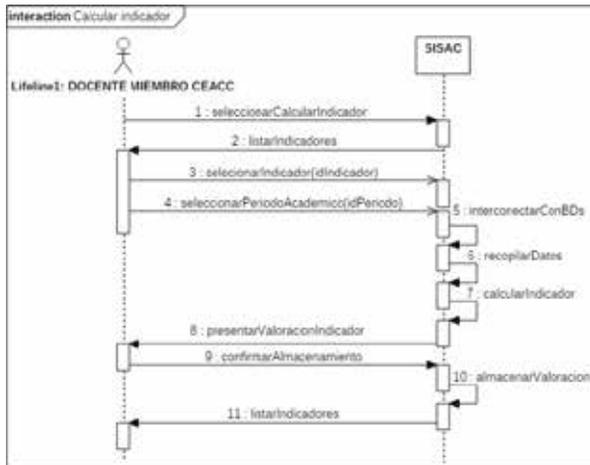
Gráfico 26
Diagrama de secuencia “Revisar evidencia”



Elaboración: Autor

Calcular indicador

Gráfico 27
Diagrama de secuencia “Calcular indicador”



Elaboración: Autor

Auditar evidencias

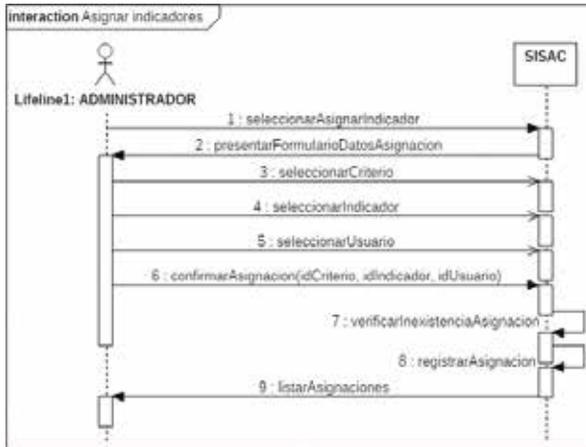
Gráfico 28
Diagrama de secuencia “Auditar evidencias”



Elaboración: Autor

Asignar indicadores

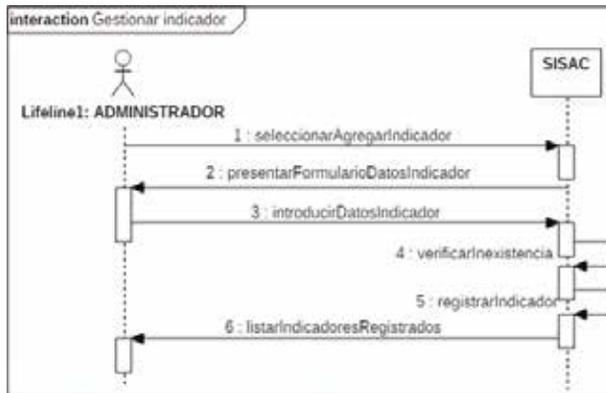
Gráfico 29
Diagrama de secuencia “Asignar indicadores”



Elaboración: Autor

Gestionar indicador

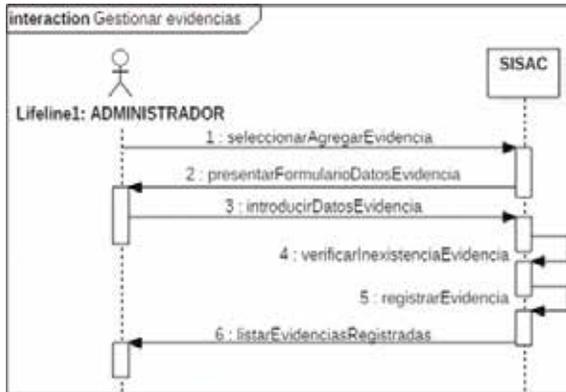
Gráfico 30
Diagrama de secuencia “Gestionar indicador”



Elaboración: Autor

Gestionar evidencias

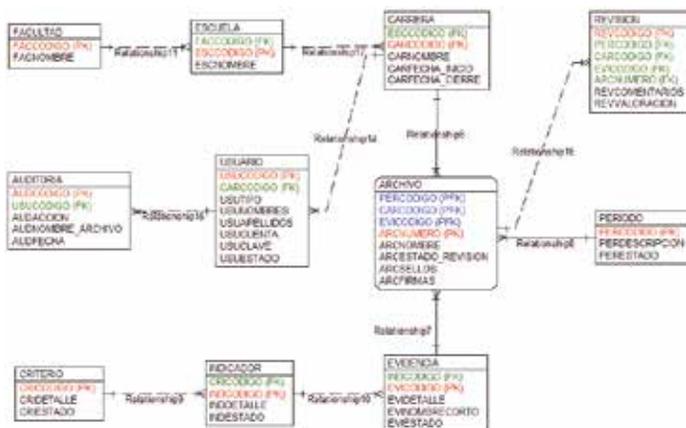
Gráfico 31
Diagrama de secuencia “Gestionar evidencias”



Elaboración: Autor

ESQUEMA DE BASE DE DATOS

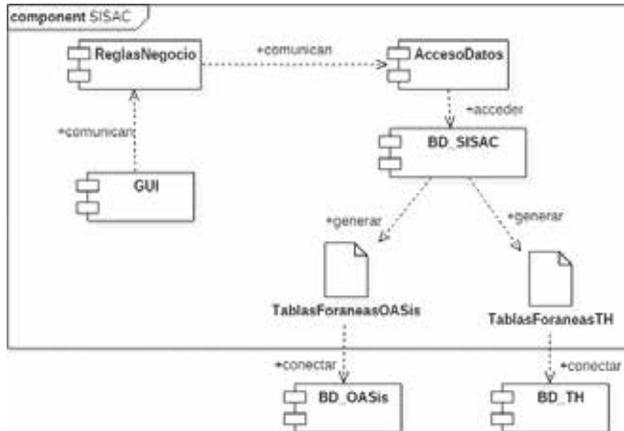
Gráfico 32
Esquema de base de datos



Elaboración: Autor

DIAGRAMA DE COMPONENTES

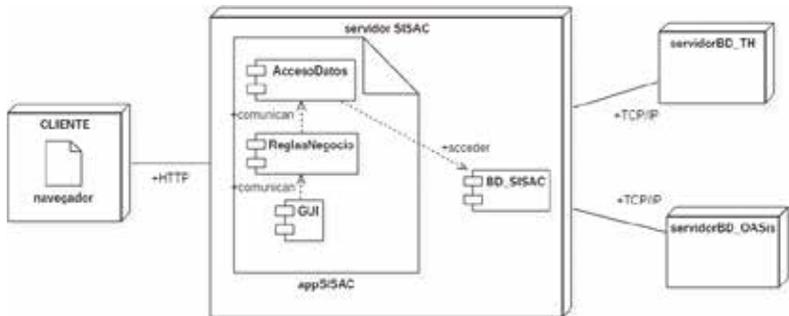
Gráfico 33
Diagrama de componentes SISAC



Elaboración: Autor

Diagrama de despliegue

Gráfico 34
Diagrama de despliegue SISAC



Elaboración: Autor

Implementación

Para la implementación del SISAC, se utilizó el framework Java Enterprise Edition (JEE), el cual tiene como lenguaje de programación base a JAVA, con las funcionalidades y portabilidades propios de esta herramienta. La herramienta de modelado StarUML y el IDE NetBeans. Como gestor de base de datos se utiliza PostgreSQL, por tratarse de una herramienta libre, y como contenedor o servidor de aplicaciones Apache Tomcat.

El Gráfico 35, presenta el formulario para gestión de evidencias, parte de la implementación del SISAC:

Gráfico 35
Gestión de evidencias

Gestión de Evidencias

Buscar por:

Para subir una nueva evidencia clic aqui >>>

CRITERIO	INDICADOR	EVIDENCIA	ARCHIVO	ACCIÓN
PERTINENCIA	ESTADO ACTUAL Y PROSPECTIVA	PEDI POA	1EIS2_1	
PERTINENCIA	SEGUIMIENTO A GRADUADOS	ORGANICO FUNCIONAL	4EIS4_1	

Elaboración: Autor

El Gráfico 36, muestra el formulario para cargar evidencias de los indicadores del modelo de evaluación.

Gráfico 36
Cargar evidencias

Gestión de Evidencias [Ayuda](#)

Periodo: **Marzo 2013 - Agosto 2013**

Criterio: **PERTINENCIA**

Indicador: **SEGUIMIENTO A GRADUADOS**

Evidencia: **ORGANICO FUNCIONAL**

Orgánico funcional de la carrera o facultad, vigente en el periodo de evaluación.

Detalle:

Imagen: Ningún archivo seleccionado
Solo se admiten archivos .pdf

Elaboración: Autor

Para la revisión de evidencias se cuenta con el formulario del Gráfico 37:

Gráfico 37
Revisión evidencias

Revisión de evidencias

Todo: Criterio: Indicador: Evidencia:

Facultad: **INFORMATICA Y ELECTRIC** Criterio: **PERTINENCIA**

Escuela: **INGENIERIA EN SISTEMAS** Indicador: **ESTADO ACTUAL Y PROSI**

Carrera: **INGENIERIA EN SISTEMAS** Evidencia: **ESTADO ACTUAL CARRER**

Detalle evidencia: Documento de estado actual y prospectiva de la carrera.

Periodo Académico	
<input checked="" type="radio"/>	Todos
<input type="radio"/>	Septiembre 2012 - Febrero 2013
<input type="radio"/>	Marzo 2013 - Agosto 2013
<input type="radio"/>	Septiembre 2013 - Febrero 2014

Elaboración: Autor

El listado de evidencias encontradas se presenta en el Gráfico 38:

Gráfico 38
Listado de evidencias buscadas

CRITERIO	INDICADOR	EVIDENCIA	ARCHIVO	REVISIÓN	
PERTINENCIA	ESTADO ACTUAL Y PROSPECTIVA	PEDI POA	1E152_1	<input checked="" type="checkbox"/>	
PERTINENCIA	SEGUIMIENTO A GRADUADOS	ORGANICO FUNCIONAL	4E154_1	<input type="checkbox"/>	
AMBIENTE INSTITUCIONAL	ADMINISTRACION - ENCUENTROS Y SEMINARIOS	PUBLICIDAD EVENTO	4E1510_1	<input checked="" type="checkbox"/>	

Elaboración: Autor

Una vez encontrada la evidencias, se puede revisar a través del formulario que se observa en el Gráfico 39:

Gráfico 39
Revisión evidencias

Revisión de evidencias

[Atrás](#)

Periodo Académico Marzo 2013 - Agosto 2013

Facultad: INFORMATICA Y ELECTRONICA Criterio: PERTINENCIA

Escuela: INGENIERIA EN SISTEMAS Indicador: SEGUIMIENTO A GRADUADOS

Carrera: INGENIERIA EN SISTEMAS Evidencia: ORGANICO FUNCIONAL

Detalle de la evidencia: Orgánico funcional de la carrera o facultad, vigente en el período de evaluación.

Estado revisión
 Sellos
 Firmas

Comentarios

No existen comentarios.

Elaboración: Autor

Código fuente

A continuación, fragmentos de código (en el orden que se deben crear y configurar las diferentes tablas) correspondiente a la implementación de la base de datos federada:

- Configuración Foreign Data Wrapper (FDW) PostgreSQL a MySQL

```
-- Cargar la biblioteca odbcc_fdw.sql
-- para manejo de Foreign Data Wrapper (FDW)
CREATE EXTENSION odbcc_fdw;
-- Creamos un Servidor Foráneo con la siguiente sentencia,
este servidor
-- es como una referencia a la conexión ODBC desde
el Postgres.
/* SINTAXIS
CREATE SERVER <NombreServerDesdePG>
FOREIGN DATA WRAPPER odbcc_fdw
OPTIONS (dsn '<NombreConexionODBC>');
*/

CREATE SERVER conexionPGMySQL
FOREIGN DATA WRAPPER odbcc_fdw OPTIONS (dsn
'PostgreSQL-MySQL');
-- Mapeamos un "Login Role" con un usuario de la BD SQL
SERVER.
-- En mi caso creé un único usuario.
/* SINTAXIS
CREATE USER MAPPING
FOR <UsuarioPostgres>
SERVER <NombreServerDesdePG>
OPTIONS (username '<UsuarioSQLServer>', password
'<PassSQLServer>');
*/
```

```
-- Con este usuario se debe acceder y ejecutar la consulta
-- pues se está creando el mapeo solo para este usuario
CREATE USER MAPPING
  FOR tesis
  SERVER conexionPGMySQL
  OPTIONS (username 'root', password 'ovcbcs');

-- Creación de un esquema donde podrán todas las ta-
-- blas foráneas.
/* SINTAXIS
CREATE SCHEMA <TablasForaneas>;
*/
CREATE SCHEMA TablasForaneasMySQL;

-- Creamos las “Tablas Foráneas” que son las que accederán
-- por debajo
-- a las tablas SQL SERVER según la Query que le asignemos,
-- y desde el
-- Postgres las llamamos como una tabla cualquiera.
/* SINTAXIS
CREATE FOREIGN TABLE <TablasForaneas>.
<NombreTablaPostgres>
  (orderid integer NOT NULL
  , customerid varchar(20) NOT NULL )
SERVER <NombreServerDesdePG>
  OPTIONS (database '<NombreDBSQLSERVER>', schema
'<SchemaSQLServer>', sql_query 'SELECT orderid, custo-
merid FROM orders', sql_count 'select count(orderid) from
orders');
*/
CREATE FOREIGN TABLE TablasForaneasMySQL.prueba
(
  codigo          char(5) NOT NULL,
  nombre          char(30)      NOT NULL,
  apellidochar(30)  NOT NULL
)
```

SERVER conexionPGMySQL

```
OPTIONS (database 'villacres', schema 'villacres', sql_query  
'SELECT codigo, nombre, apellido FROM tabla1', sql_count  
'select count(codigo) from tabla1');
```

-- Para acceder a través de FDW a la tabla de MySQL desde PostgreSQL,

-- realizar un SELECT utilizando el schema y en nombre de la tabla

```
SELECT *
```

```
FROM TablasForaneasMySQL.prueba
```

- Configuración FDW PostgreSQL a SQLServer

-- Cargar la biblioteca odbc_fdw.sql

-- para manejo de Foreign Data Wrapper

```
CREATE EXTENSION odbc_fdw;
```

-- Creamos un Servidor Foráneo con la siguiente sentencia, este servidor

-- es como una referencia a la conexión ODBC desde el Postgres.

```
CREATE SERVER conexionPGSQLServer
```

```
FOREIGN DATA WRAPPER odbc_fdw OPTIONS (dsn  
'PostgreSQL-SQLServer200811');
```

```
DROP SERVER conexionPGSQLServer CASCADE;
```

-- Mapeamos un "Login Role con un usuario de la BD SQL SERVER.

```
CREATE USER MAPPING
```

```
FOR postgres
```

```
SERVER conexionPGSQLServer
```

```
OPTIONS (username 'postgres', password 'ovcianmjlp');
```

-- Creación de un esquema para las tablas foráneas.

```
CREATE SCHEMA TablasForaneasSQLServer;
```

```
DROP SCHEMA TablasForaneasSQLServer;
```

```
-- Creamos las "Tablas Foraneas" que son las
-- que accederán por debajo a las tablas
-- SQL SERVER según la Query que le asignemos, y desde el
-- Postgres las llamamos como una tabla cualquiera.
CREATE FOREIGN TABLE TablasForaneasSQLServer.prueba
(
    codigo_mat          varchar(5)    NOT NULL,
    nombre_mat          varchar(50)   NOT NULL
)
SERVER conexionPGSQLServer
  OPTIONS (database 'bd_tesis', schema 'dbo', sql_query 'SE-
LECT codigo_mat, nombre_mat FROM materia', sql_count
'select count(codigo_mat) from materia');
```

- Conectividad entre distintas BD alojadas en distintos servidores

```
-- En PostgreSQL 9.2
-- Cargar libreria dblink
CREATE EXTENSION dblink;
```

```
-- Comprobando conectividad a PostgreSQL
select dblink_connect('dbname=proyecto hostaddr=127.0.0.1
user=postgres password=ovcianmjlpl port=5433');
```

```
-- Acceder desde PostgreSQL 9 a través de FDW
select * from
dblink ('dbname=proyecto hostaddr=127.0.0.1 user=tesis
password=tesis port=5433',
'select * from TablasForaneasMySQL.prueba') as t1(codigo
char(5), nombre char(30), apellido char(30));
```

CONCLUSIONES

Con la incorporación del Sistema Integrado para Seguimiento y Autoevaluación de Carreras (SISAC), se pudo demostrar su incidencia en el cálculo de indicadores de autoevaluación de carreras de la ESPOCH, con una mejora del 99,82%.

La implementación del SISAC, constituido por un sistema de bases de datos federadas permite el funcionamiento continuo de las aplicaciones existentes sin generar cambios y conservando la estructura de los sistemas y bases de datos en la institución, además, permitirá y facilitará la incorporación de nuevas aplicaciones y bases de datos.

Finalmente, utilizar los datos existentes en las bases de datos autónomas de la ESPOCH, a través de la implementación del SISAC, permite disminuir el tiempo dedicado para recolección y valoración de indicadores del modelo genérico de evaluación de carreras.

RECOMENDACIONES

Para la actualización del Sistema Integrado para Seguimiento y Autoevaluación de Carreras (SISAC), es necesario que el administrador del SISAC mantenga comunicación con cada uno de los administradores de bases de datos que constituyen al sistema de bases de datos federadas, pues ellos serán quienes concedan los privilegios para acceder a datos nuevos que reposan en cada base de datos.

Antes de iniciar la implementación de sistemas informáticos que utilicen sistemas de bases de datos federadas es importante analizar que bases de datos autónomas la constituirán y los datos existentes en cada una.

En caso de interconectarse a nuevas bases de datos u obtener datos de nuevas tablas, el administrador debe crear un esquema donde aloje todas las tablas foráneas que constituyen el sistema de bases de datos federadas.

Para interconectar nuevas bases de datos al SISAC, hay que tener en cuenta que este utiliza PostgreSQL como SGBD, el cual solo soporta conectividad a SQLServer, MySQL y Oracle.

BIBLIOGRAFÍA

- Aslan, G., & McLeod, D. (1999). Semantic heterogeneity resolution in federated databases by metadata implantation and stepwise evolution. *The VLDB Journal*, 8, 120-132.
- CEAACES (2014). *Reglamento para los Procesos de Autoevaluación de las Instituciones, Carreras y Programas de Sistema de Educación Superior*. Quito, Ecuador.
- _____ (2015). *Modelo Genérico de Evaluación del Entorno de Aprendizaje de carreras presenciales y semipresenciales de las Universidades y Escuelas Politécnicas del Ecuador. Versión 2.0* (Vol. 2.0). Quito, Ecuador.
- Codina, L. (2010). Nuevas fuentes de información académica. *Anuario ThinkEPI*, 4, 139-142.
- Crespo, P., Maldonado, J. A., Cano, C., Robles, M., & Casanova, J. C. (2002). Integración de Bases de datos hospitalarias: Gestión de las bases de datos Componentes.
- Espino Barrios, L. F. (2009). *Sistemas de Bases de Datos Federadas*. <https://goo.gl/yPjwma>
- González, A. P., & Flores, R. C. (2013). Implementando un proveedor de datos OAI para una colección federada de recursos educativos abiertos. *educonector. info*, 52-67.
- Heimbigner, D., & McLeod, D. (Julio de 1985). A federated architecture for information management. *ACM Transactions on Information Systems*, 3(3), 253-278. Recuperado de <https://goo.gl/5tEh6R25> (10-2015).
- IEEE (2008). *Especificación de Requisitos según el estándar*.
- Litwin, W., Mark, L., & Roussopoulos, N. (1990). Interoperability of multiple autonomous databases. *ACM Computing Surveys (CSUR)*, 267-293.
- Muñoz, A., & Aguilar, J. (17 de Diciembre de 2011). Ontología del proceso de integración de bases de datos federadas. *ResearchGate*, 18. Recuperado de <https://goo.gl/QMcoiX> (11-10-2015).
- Ôzsu, T., & Valduriez, P. (2011). *Principles of Distributed Database Systems* (Third ed.). New York: Springer. doi:10.1007/978-1-4419-8834-8

- Saavedra Places, Á. (2003). *Arquitectura para federación de bases de datos documentales basada en ontologías*. Tesis doctoral.
- Saltor, F. (2000). *Sistemas de Información Cooperativos y Bases de Datos Federadas*. Cuenca, Ecuador: Castilla-La Mancha.
- Sheth, A. P., & Larson, J. A. (3 de Septiembre de 1990). Federated database system for managing distributed, heterogeneous, and Autonomous Databases. *ACM Computing Surveys*, 22(3), 183-236.
- Simari, G., Pacífico, C., Silva, L., Denegri, L., Novello Pelayo, I., Teze, J., . . . Gottifredi, S. (2017). Desarrollo de Sistemas de Argumentación Masiva sobre Base de Datos Federadas. *Ciencia, Docencia y Tecnología*.
- Vázquez Tapia, R., & Razo Rodríguez, A. F. (2014). Indixe: el agregador nacional de la Red Mexicana de Repositorios Institucionales (REMERI) para la interoperabilidad con redes federadas de repositorios institucionales. *IV Conferência Internacional sobre Bibliotecas e Repositórios Digitais* (pp. 9-23). Porto Alegre.
- Velásquez, A. (9 de Diciembre de 2014). *Bases de Datos Distribuidas*. Recuperado de <https://goo.gl/H9Yz7d> (3-11-2015).
- Verger, F., & Noguera, P. F. (2005). *Creación de un Sistema de bases de datos federadas a través de Tecnología Java*.
- Zamora, C. E., & García, J. J. (11 de 2010). *Ingeniería de Software*. Recuperado de <https://goo.gl/YWE81E> (15-2-2016).

CAPÍTULO 4

Análisis, diseño e implementación de un sistema de respaldo de datos y restauración de la información basado en software AVAMAR

Wilson Chango
Jaime Sayago
Víctor Quiñonez
Ligia Jácome

IDENTIFICACIÓN DEL PROBLEMA

El ambiente de TI (Tecnología de Información) de las empresas está cambiando y evolucionando cada día. El resguardo de la información que generan las empresas ha sido una tarea importante para el equipo de TI. En sus inicios esa tarea, era muy rústica y tenía muy poco nivel de automatización y confiabilidad. Copias de información a una unidad de disco externo, memoria USB o cinta magnética han sido las soluciones típicas implementadas ante la imperiosa necesidad de respaldar la información. Sin embargo, ese tipo de soluciones ha convertido por sí mismas en un problema para las empresas modernas por sus limitaciones y escasas garantías al momento de la recuperación, y no son aplicables para empresas grandes las cuales incluyen cientos de máquinas virtuales y máquinas de usuarios finales que necesitan ser respaldadas (Solsona, Moya, & Calero, 2006).

Una de las posibles soluciones es el de disponer de una red completamente independiente de la de producción para el ambiente de respaldo y restauración de la información. Sin embargo, no todas las empresas pueden darse ese lujo ya que eso implica destinar recursos de hardware como tarjetas NICs, concentradores, etc., dedicados a la red de respaldo; resulta necesario el estudio de la problemática de los ambientes actuales y encontrar una alternativa de solución que contemple todos los factores críticos indicados y provea mecanismos que permitan realizar las tareas de respaldo y recuperación de una manera eficiente y optimizando el consumo de recursos de la infraestructura productiva de tal manera que cause el menor impacto posible.

AVAMAR es una de las tecnologías que facilita la integración con ambientes virtuales como VMware e Hyperv, la incorporación de módulos especializados para respaldos de máquinas virtuales y de usuarios finales, entre otras características, hacen que esta herramienta sea una de las ideales para el estudio de los sistemas de respaldos en estos escenarios particulares.

El escenario del análisis planteado, corresponde a un escenario típico de una empresa moderna, el cual cuenta con servidores ejecutándose sobre la plataforma virtual VMware y varios usuarios de esos servicios, trabajando desde la red LAN. Uno de los Objetivos es analizar las soluciones de respaldos más apropiada para el ambiente descrito, determinando primero las necesidades del ambiente virtual y de usuarios finales, diseñando y modelando la herramienta y finalmente implementándola en un ambiente de laboratorio(Linthicum, 2017).

Los ambientes virtuales VMware proveen características inherentes que deben ser aprovechadas por terceras herramientas para la optimización del respaldo, por ejemplo la capacidad conocida como CBT (Changed Block Tracking), de Vmware, cuya función es la de reconocer únicamente los bloques de información que cambian de una máquina virtual, la compresión de la información durante las tareas de respaldo es crítica para optimizar el consumo de ancho de banda que una herramienta de respaldos consumiría en su operación(Choo, Rana, & Rajarajan, 2017).

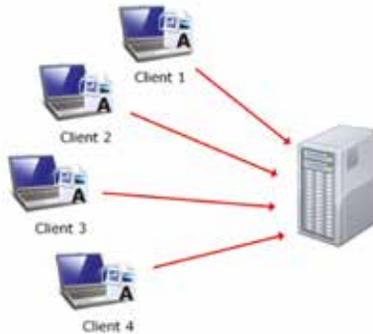
METODOLOGÍA DE LA INVESTIGACIÓN

La metodología que utiliza es basada en un enfoque cuantitativo, primero se llega a la elaboración de matrices para establecer los requerimientos básicos para la implementación de la solución y los pasos posteriores para realizar las pruebas, también realiza una investigación bibliográfica para profundizar teorías sobre respaldos de información, revisando artículos técnicos sobre visualización de servidores.

a. Arquitectura de la solución de respaldo y recuperación AVAMAR

La arquitectura incluye tanto hardware como software, puede escalar de manera vertical para acoplarse a las necesidades de empresas pequeñas, medianas y grandes. La clave de su arquitectura es el proceso de compresión de la información a través del análisis en línea y en el origen de la data antes de ser respaldada. Este análisis le permite identificar información redundante y solo guardar aquella información nueva, ahorrando así recursos de almacenamiento, CPU en el host y ancho de banda, el proceso de compresión, puede explicarse en la figura 1.

Figura 1
Proceso de respaldo en un ambiente tradicional



Elaboración: Autores

En la figura anterior, se observa el proceso de respaldo en un ambiente tradicional cliente servidor. Se observa que todos los clientes tienen exactamente el mismo documento para respaldarse, tal vez se trate de un documento adjunto que todos recibieron como parte de un correo electrónico, muy típico en los entornos reales. Un sistema de respaldo tradicional guardaría 4 copias exactas del documento.

b. Características de la solución de respaldo y recuperación AVAMARE

Se exponen las características más relevantes de la solución de respaldo escogida para el presente estudio, estos son: Detección de duplicidad global, tolerancia a fallos sistemática, uso de la red IP, arquitectura escalable, opciones flexibles de instalación, manejo centralizado.

c. Términos comunes de la solución de respaldo

Los siguientes términos guardan relación con las características técnicas de la solución de respaldo y recuperación estudiada (objetos, stripe, nodo, servidor, sistema).

d. Componentes de la solución de respaldo y recuperación

Existen tres componentes básicos (Avamar Server, Avamar Client, Avamar Administrator) en un sistema de respaldo como el propuesto. Estos pueden apreciarse en la figura 2.

e. Arquitectura de tolerancia a fallos sistemática

Para asegurar la integridad de la información respaldada, la solución de respaldo de ambientes virtuales y usuarios finales, provee una arquitectura de tolerancia a fallos sistemática a través de los siguientes mecanismos (RAID, RAIN, checkpoints, alta disponibilidad de enlaces de red).

Figura 2
Proceso de respaldo en un ambiente tradicional



Elaboración: Autores

f. Especificaciones de la solución de respaldo de máquinas virtuales y usuario final

El servidor de la solución de respaldo utiliza SUSE Linux Enterprise Server (SLES) como sistema operativo base para funcionamiento.

Existen dos versiones (Avamar Data Store, Avamar Virtual Edition) en las cuales puede operar la solución de respaldos:

g. Proceso de compresión de información

La característica fundamental, del sistema es la capacidad de reducir la cantidad de información a ser respaldada. Se muestra los porcentajes típicos de duplicidad de información que la solución es capaz de encontrar en un ambiente de respaldo:

- Duplicidad inicial encontrada a nivel de file systems: aproximadamente 65%.
- Duplicidad inicial encontrada a nivel de base de datos: aproximadamente 35%.
- Respalos diarios de file systems: aproximadamente 99.7%
- Respalos subsecuentes de las bases de datos: aproximadamente 97%.

h. Integración con ambientes virtuales para respaldo de máquinas virtuales

La solución de respaldos es compatible con ambientes VMware y Hyper-V, dos de las plataformas más utilizadas a nivel mundial.

Generalidades de la plataforma de virtualización VMware

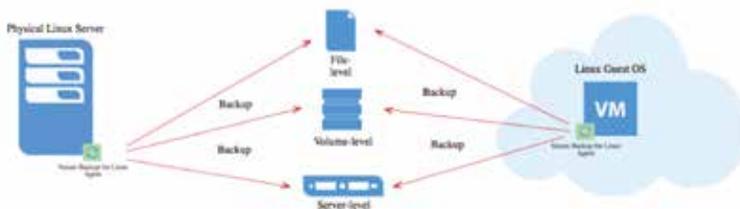
VMware VSphere es utilizada para construir infraestructura de servidores y escritorios virtuales que puedan mejorar la disponibilidad, seguridad y maniobra de las aplicaciones de misión crítica de las empresas. VSphere virtualiza servidores, almacenamiento y red, permitiendo que se ejecuten múltiples sistemas operativos y sus aplicaciones independientemente en VMs mientras comparten recursos físicos del host.

Métodos de respaldo de máquinas virtuales

Existen dos métodos básicos para realizar el respaldo de máquinas virtuales.

Método de Guest Backup: Es el que contempla la instalación de un agente en cada máquina virtual, el cual se comunicará con el servidor de respaldos y llevará a cabo la operación de respaldo o restauración, ver figura 3.

Figura 3
Estructura de la plataforma de virtualización VMware



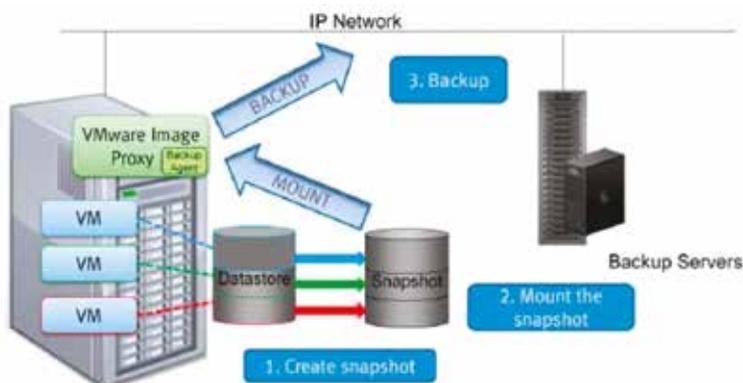
Fuente: EMC Corporation (2012). Recuperado de <http://www.emc.com/>

Una de las ventajas de este método es que permite manejar el ambiente virtual como si fuese físico, algo con lo que está familiarizado.

MÉTODO DE IMAGE BACKUP

Utiliza una máquina física o virtual como servidor proxy para manejar el proceso de respaldo, ver figura 4.

Figura 4
Esquema de respaldo de máquinas virtuales a nivel de imagen



Fuente: EMC Corporation. (2012). Recuperado de <http://www.emc.com/>

Esto es posible a través de la creación de un snapshot de la VM y luego respaldando esa imagen en el servidor de respaldos.

i. Respaldo de usuarios finales

La solución de respaldo AVAMAR incluye un módulo denominado Desktop/Laptop, el cual es activado durante el proceso de instalación de la herramienta y está destinado para habilitar el backup y restauración de ambientes de máquinas de usuario final.

j. Consideraciones de diseño del sistema de respaldo y recuperación de ambientes virtuales y usuarios finales

Se inicia con el levantamiento de información del ambiente a resguardar, esto es número y tipo de usuarios finales (sistemas operativos, cantidad de espacio a resguardar, etc.); y también del ambiente virtual.

K. Descripción del escenario

El escenario planteado es considerado a nivel de laboratorio. cuenta con todos los componentes de un escenario real, pero en un ambiente controlado. En la figura 5 se describe el escenario sobre el cual se trabaja.



Elaboración: Autores

Se trata de una infraestructura típica que consta de los siguientes elementos:

- Dos servidores de físicos Dell PowerEdge R720 con 4 CPUs Intel Xeon E5530 x 2.40 GHz y 12GB en RAM

- Un almacenamiento VNX 5200 con 24TB de capacidad
- Un ambiente virtual VMware VSphere 5 Enterprise Plus
- Una red LAN y una red SAN
- Usuarios de computadores de escritorio y laptops
- Máquinas virtuales con servidores

I. Levantamiento de la información necesaria para el diseño

Se define de forma específica las características de los elementos que componen el escenario. Se presenta las características de las máquinas de usuario final como por ejemplo su sistema operativo, cantidad y tipo de información, etc. De igual manera se procede con el ambiente virtual. En la tabla 1 se resume la información necesaria a nivel de usuarios finales para el diseño de la solución de respaldos y recuperación.

Tabla 1
Información de los usuarios finales
y sus requerimientos de espacio para respaldar

Tipo de Usuario	Cantidad de Usuarios	Cantidad de Información por usuario [GB]	Crecimiento anual %	Espacio Total [GB]	Espacio Total a 1 año
Oro	10	50	5	500	525
Plata	10	25	5	250	263
Bronce	20	10	2	200	204
Total	40	85	12	950	992

Nota. El resultado anual de capacidad para almacenamiento de información para los usuarios tipo Oro, Plata y Bronce es de 992 GB.

Elaboración: Autores

En cuanto al ambiente virtual, se tiene VMware VSphere 5.1 Enterprise Plus. Sobre él se están ejecutando tres VMs como lo muestra la Tabla No.2

Tabla 2
Información de las máquinas virtuales
y sus requerimientos de espacio para respaldar

Tipo de Servidor	Sistema Operativo	Cantidad de Información actual [GB]	Crecimiento anual %	Espacio Total [GB]	Espacio Total a 1 año
Proxy Server	Linux	60	5	60	63
Directorio Activo	Windows 2012	100	8	100	108
Base de Datos	SQL 2008	120	10	120	132
Total	40	280	23	280	303

Nota. Resultado anual de capacidad para almacenamiento.

Elaboración: Autores

m. Consideraciones de diseño de las políticas de respaldo

Es necesario definir políticas adecuadas en base a la realidad de la empresa o institución. Hay escenarios atípicos, por ejemplo, en las entidades bancarias, en las cuales por regulaciones legales se exige que su información sea respaldada por un período superior a 3 años. Para esos escenarios, dado el gran período de tiempo que la información debe guardarse, el sistema de respaldos seguramente será muy grande para poder manejar esa cantidad de información.

A continuación, se describen las principales consideraciones de diseño para las políticas de respaldo a tomar en cuenta:

- Definición de dominios.
- Definición de datasets.
- Definición de períodos de retención.
- Definición de la periodicidad de los respaldos.
- Definición del tipo de respaldo.
- Definición de la ventana de respaldo.

En la Tabla 3 se encuentra la definición para las políticas de respaldo, como resumen se puede mencionar que se han creado 4 dominios: Oro, Plata, Bronce y Virtual. Los tres primeros corresponden al ambiente de usuarios finales y el último al ambiente virtual.

Tabla 3
Definición de diseño de las políticas de respaldo
de usuarios finales y ambiente virtual

Dominio	DATASET		Source Data	Exclusiones	Inclusiones	PERIODICIDAD Y TIPO DE RESPALDO			RETENTION	
	Plataforma					Repetir	Incremental	Full	Número	Unidades
ORO	Windows	TODOS EL DI- RECTORIO ESPECÍFICO	*.mp3 *.Wav	CMIS BACKUPS	Diario	SI	2	Semanas		
		TODOS EL DI- RECTORIO ESPECÍFICO	*.mp3 *.Wav *.JPEG *.gif		Semanal	SI	1	Meses		
PLATA	Windows	TODOS EL DI- RECTORIO ESPECÍFICO	*.mp3 *.Wav *.JPEG *.gif	CMIS BACKUPS	Diario	SI	1	Semanas		
		TODOS EL DI- RECTORIO ESPECÍFICO	*.mp3 *.Wav *.JPEG *.gif		Semanal	SI	2	Meses		
BRONCE	Windows	TODOS EL DI- RECTORIO ESPECÍFICO	*.mp3 *.Wav *.JPEG *.gif	CMIS BACKUPS	Diario	SI	1	Semanas		
		TODOS EL DI- RECTORIO ESPECÍFICO	*.mp3 *.Wav *.JPEG *.gif		Semanal	SI	1	Meses		
VIRTUAL	Imagen Virtual	Toda la imagen de la máquina virtual	N/A	N/A	Diario	SI	1	Semanas		
		Toda la imagen de la máquina virtual	N/A		Semanal	SI	1	Meses		
					Mensual	SI	2	Meses		

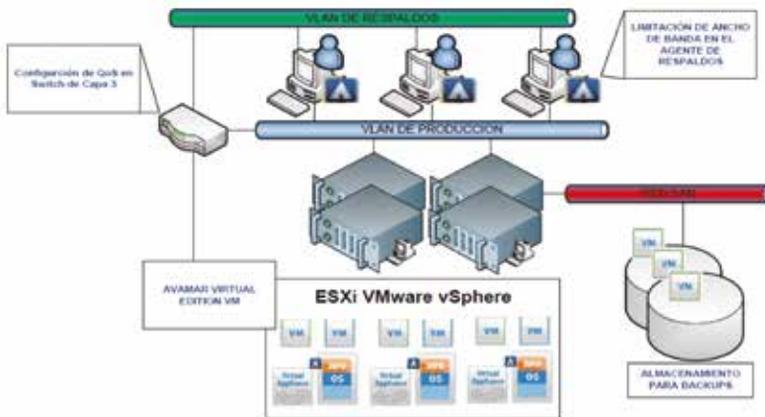
Nota. Políticas establecidas a los usuarios oro, plata bronce y ambiente virtual
 Elaboración: Autores

n. Consideraciones de diseño de la red

La implementación de una solución de respaldo es a través de la red LAN. Es decir, el tráfico de producción de la empresa como el de respaldos comparten la misma infraestructura de red.

La política presenta un desafío a nivel de control del tráfico de la red por obvias razones. Es por eso que a continuación se presentan algunas consideraciones de diseño (segmentación de las redes de backup y de producción, aplicación de calidad de servicio QoS, compresión del tráfico de backup, limitación a nivel del agente) que serán tomadas en cuentas para la fase de implementación (ver figura 6).

Figura 6
Diagrama lógico de red con parámetros de diseño para administración del tráfico de backup



Fuente: Los autores

En la figura anterior se muestra un diagrama lógico de la red de la solución de respaldos propuesta en la cual se puede apreciar la segmentación en dos redes del ambiente: una red de backups y otra de producción.

o. Dimensionamiento de la solución de respaldo

Para realizar el dimensionamiento se utiliza una herramienta conocida como “Backup and Recovery Manager System Sizer”; es un software que permite ingresar como parámetros las políticas definidas en la tabla 3, y junto con la información de los requerimientos de espacio de los usuarios finales y máquinas virtuales, arroja un resultado de un posible modelo de equipo y la proyección del espacio requerido después del período de interés.

Aquí se muestran los cálculos que realiza la herramienta respecto al requerimiento futuro de capacidad de almacenamiento, ver figura 7.

Figura 7
Solución de respaldo recomendado por el software EBSS

Recommendation Details
<p>Required for Storage</p> <p>1 M600(2TB) system with 3 Storage, 1 Utility, and 1 Spare nodes per system. Last node requires 51.00% of its capacity.</p>
<p>Required for Performance</p> <p>1 M600(2TB) system with 3 Storage, 1 Utility, and 1 Spare nodes per system. 6 TB of Capacity licensing required.</p>

Fuente: Los autores

Se puede observar el resultado final arrojado por la herramienta de dimensionamiento de la solución de respaldos. Dado los requerimientos y características de la información del escenario, EBSS indica que la solución más adecuada es un Avamar Store compuesto de 3 nodos (M600) con 2TB de capacidad cada uno, 1 nodo utility (para control de la solución), y 1 nodo como Spare (repuesto). Se indica que el último nodo contará con una ocupación de 51%.

ANÁLISIS DE RESULTADO Y PROPUESTA

Los ambientes virtuales y de usuarios finales son cada vez más críticos dentro de una empresa moderna. Prácticamente, la tendencia de los centros de datos es consolidar todos sus servicios de hardware y software en una infraestructura virtual aprovechando todas las ventajas que ésta provee.

La tarea de resguardar ese ambiente con técnicas de respaldo no debe ser trivial. Es importante contar con una herramienta que provea la facilidad de respaldar el ambiente virtual de una manera segura, rápida y eficiente.

AVAMAR, tiene puntos de integración muy completos con la plataforma de virtualización VMware que la convierte en una de las soluciones más apropiadas para el respaldo del ambiente virtual, otras de las ventajas analizadas de esta solución, es que además de la integración con VMware cuenta con un módulo denominado “Desktop/Laptop” que le permite instalar un agente en el computador del usuario final y desde éste controlar el respaldo automático de la información crítica del usuario.

Para la implementación se ha escogido la versión virtual de AVAMAR denominada AVE o Avamar Virtual Edition por sus siglas en inglés. La solución será instalada sobre un servidor ESXi y correrá su servicio como una VM más del ambiente.

En la implementación, se pretende a nivel de laboratorio, completar las siguientes actividades:

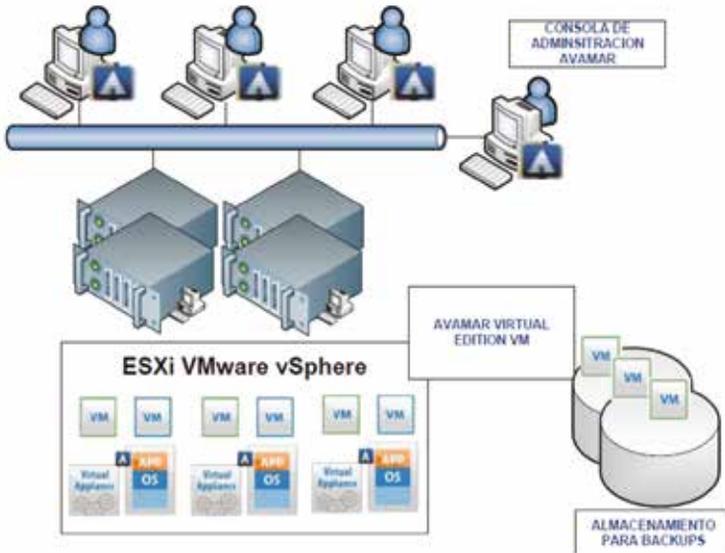
- Instalación de AVE en un ambiente virtual de laboratorio
- Instalación de los agentes en varios computadores de usuarios finales.
- Instalación del proxy de Avamar para respaldo del ambiente virtual.
- Configuración de la solución para dejar activos los respaldos y restauraciones.

- Configuración de los agentes de usuario final para optimizar el consumo de ancho de banda durante las operaciones de respaldo y recuperación.
- Pruebas de respaldo y recuperación de información de usuario final.
- Pruebas de respaldo y recuperación de VMs dentro del ambiente virtual.

a. Instalación de la solución de respaldo de ambiente virtual y usuario final

Los componentes que constituyen la solución de respaldo de ambientes virtuales y de usuario final son las que se muestran en la figura 8:

Figura 8
Componentes de la solución de respaldo de ambiente virtual y usuarios finales



Fuente: Los autores

Se puede apreciar que existen varios componentes que están presentes en la solución de respaldo. Básicamente existen los siguientes componentes (AVAMAR Virtual Edition VM, Virtual Appliance o Proxy, consola de administración, agente, almacenamiento).

b. Requerimientos para la instalación del sistema de respaldos AVAMAR virtual edition

Una vez descargado el software de instalación de AVE, que básicamente consiste en un archivo “.ovf” que debe ejecutarse desde VSphere se debe cumplir algunos requisitos sobre la plataforma física (servidor y almacenamiento) para que sea posible la instalación de la herramienta. Este requerimiento se puede apreciar en la tabla 4.

Tabla 4
Requerimientos mínimos para la instalación de la solución de respaldos de ambientes virtuales y usuario final (EMC2, 2014)

	0,5 TB AVE	1 TB AVE	2 TB AVE	4 TB AVE
Processors	Minimum two 2GHz processors	Minimum two 2GHz processors	Minimum two 2GHz processors	Minimum four 2GHz processors
Memory	6 GB	8 GB	16 GB	36 GB
Disk space	850 GB	1,600 GB	3,100 GB	6,100 GB
Network connection	1 GbE connection	1 GbE connection	1 GbE connection	1 GbE connection

Nota. Requerimientos mínimos para la instalación de la solución de ambientes virtuales y usuarios finales

Fuente: EMC Corporation (2012). Recuperado de <https://goo.gl/3jFHZA>

De acuerdo a la tabla anterior los requerimientos mínimos del hardware sobre el cual se instalará la solución son 2 procesadores de 2Ghz, 6GB en RAM, 850GB en disco y conexión de 1GBE.

c. Instalación del proxy para respaldo de máquinas virtuales

Se expone los requerimientos para la instalación del proxy que permite el respaldo de máquinas virtuales a nivel de imagen de VM.

Requerimientos de instalación para el proxy

Para la instalación del agente proxy dentro de la infraestructura virtual se necesita contar con los siguientes requerimientos de recursos dentro del host VMware ESXi.

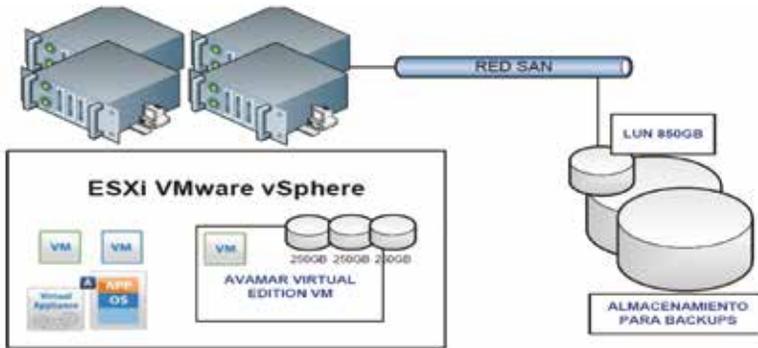
Proceso de instalación del sistema de respaldos

Se describe los pasos de instalación de manera general, ver figura 9:

- Descargar los archivos de instalación de AVE.
- Disponer de un host ESXi con las características mínimas descritas en la Tabla 4.
- Lanzar la consola vSphere client y conectarse al host ESXi.
- Ejecutar el template OVF sobre el host ESXi.
- Asignar el datastore apropiado para la instalación de AVE que cumpla con los requisitos mínimos de instalación. En este punto es importante aclarar que el aprovisionamiento de espacio en disco debe ser del tipo “Thick” o fijo, ya que esta es una exigencia de la herramienta para su correcto funcionamiento.
- Una vez establecidos todos los parámetros y requerimientos anteriores el proceso de despliegue de la VM comenzará y tardará unos cuantos minutos.

En la figura anterior se puede apreciar que el primero paso para iniciar el proceso de instalación es de disponer de un host físico con ESXi del cual se aprovisionará memoria, CPU y tarjeta NIC para la VM de Avamar.

Figura 9.
Componentes de la solución de respaldo
de ambiente virtual y usuarios finales



Fuente: Los autores

Proceso de instalación del proxy para respaldo de máquinas virtuales

El proceso de instalación del proxy para el respaldo de máquinas virtuales empieza con la descarga de la de la imagen en formato OVA1 del proxy que se obtiene desde la página web de la herramienta de respaldos. Posterior a eso se realiza el despliegue de la misma sobre el ambiente VMware verificando previamente que se cumplen los requisitos.

- Ingresar en la consola de VCenter para creación del proxy.
- Dentro de la opción “Host and Clusters” abrir la ventana file, deploy OVF template.

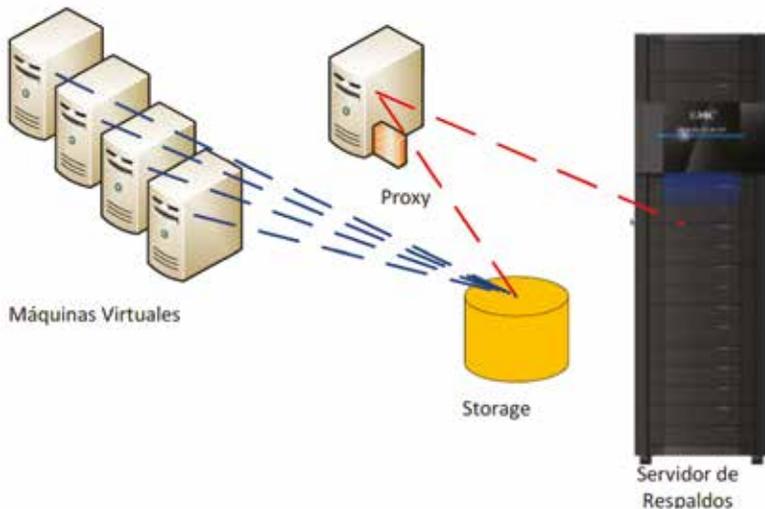
En las siguientes ventanas se podrá configurar la dirección IP, DNS y modificar el nombre del proxy. Con ese paso se culmina la configuración inicial que permite tener operativo el proxy. Posterior a eso se necesitará acceder al proxy desde el VCenter para terminar la configuración del equipo, en este se indica el nombre o IP del servidor de respaldos principal al que se conecta el proxy. Esos procesos se ven en las figuras 10 y 11.

Figura 10
Configuración del proxy con el servidor de respaldos

```
Press 1 to register the Proxy, otherwise press 2 to exit.
Main Menu
-----
1) Register Universal Proxy with an Avamar Management Console
2) quit
Your choice: 1
```

Elaboración: Autores

Figura 11
Esquema de respaldo de máquinas virtuales a través de un proxy



Elaboración: Autores

Proceso de configuración básica de la solución de respaldos

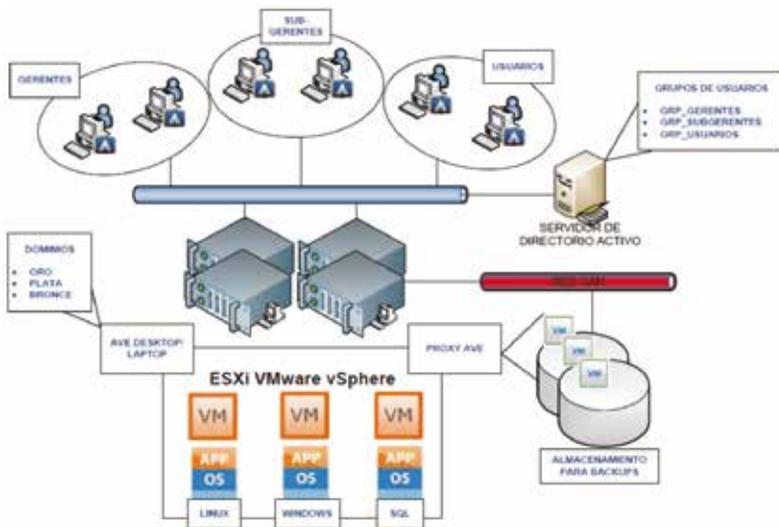
Luego de que la VM que contiene la solución de respaldo de ambientes virtuales y usuarios finales es instalada, se debe realizar ciertas configuraciones básicas, como, por ejemplo:

- Configuraciones del sistema
- Partición de los discos duros
- Creación de filesystems
- Montaje de los filesystems creados
- Configuración de la interfaz de red
- Configuración del acceso HTTP al software de administración
- Configuración de passwords de administradores
- Levantamiento de la interfaz

e. Configuración de la solución de respaldos de máquinas virtuales y usuarios finales

Se visualiza la configuración de la herramienta a alto nivel, la figura 12, presenta un diagrama del ambiente de laboratorio implementado y que emula el ambiente planteado.

Figura 12
Diagrama del ambiente de laboratorio



Elaboración: Autores

Los componentes de la infraestructura son los siguientes:

- 6 PCs de usuario final con Windows 7, se encuentran divididas en 3 grupos: El grupo de gerentes, de sub gerentes y de usuarios.
- 3 Máquinas virtuales: Un servidor linux, un windows 2012 y una base de datos SQL.
- Una Máquina virtual con el software de AVAMAR.
- Un almacenamiento desde el cual se aprovisionará los espacios de disco necesarios.
- 2 Servidores físicos con ESXi instalado.
- 1 Servidor de directorio activo y DNS.
- AVE desktop/laptop: este módulo es el encargado de las operaciones de respaldo y recuperación de las máquinas de usuario final.
- Proxy AVE: es el módulo es el encargado de las operaciones de respaldo y recuperación de las máquinas virtuales.

Configuración del módulo AVE desktop/laptop

A continuación, se exponen las configuraciones más importantes del módulo AVE desktop/laptop para respaldo y recuperación del ambiente de usuarios finales.

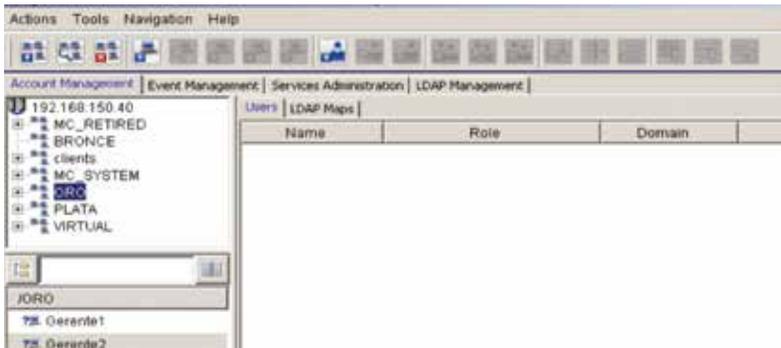
CONFIGURACIÓN DE DOMINIOS

Los dominios, son necesarios para diferenciar un grupo o perfil de usuarios de otros, con el fin de brindarles servicios distintos de backup de acuerdo a sus privilegios. Se definieron 4 perfiles, para eso es necesario crear estos grupos tanto en el servidor de directorio activo como en la herramienta de respaldos la cual llama a esos grupos dominios (Usuarios Oro, Usuarios Plata, Usuarios Bronce, Usuarios Virtuales).

CREACIÓN DE LOS USUARIOS

Luego de creados los dominios es necesario añadir los usuarios que pertenecerán a los dominios, ver figura 13.

Figura 13
Creación de usuario bajo el dominio Oro



Elaboración: Autores

En la figura anterior se observa los usuarios que han sido creados y añadidos al dominio ORO, de igual manera se procede para los demás dominios, quedando de la siguiente manera:

- Dominio ORO con dos usuarios: Gerente1, Gerente2
- Dominio PLATA con dos usuarios: Subgerente1, Subgerente2
- Dominio Bronce con dos usuarios: Usuario1, Usuarios

Configuración del agente para máquinas de usuario final

El proceso de instalación y configuración del agente de respaldo corre en las máquinas de usuario final. El proceso de instalación empieza con la descarga del software del cliente desde el mismo servidor de backup, a través del servidor web en la página <http://192.168.150.40>. Una vez descargado el software se ejecuta el instalador, el proceso tarda unos 4 minutos.

Configuración de los datasets

Los datasets son los archivos y/o carpetas que se guardan en el sistema de respaldos, en la tabla 5, se puede observar un resumen de cómo queda la configuración.

Tabla 5
Resumen de configuración de datasets

Nombre del dataset	Nombre del dominio	Directorio	Exclusiones
ORO_SET	ORO	C\MIS BACKUPS	.*.mp3 *.Wav
PLATA_SET	PLATA	C\MIS BACKUPS	.*.mp3 *.Wav *.JPEG *.gif multimedia
BRONCE_SET	BRONCE	C\MIS BACKUPS	.*.mp3 *.Wav *.JPEG *.gif multimedia

Nota. Exclusiones para respaldar información de los tipos de usuarios

Elaboración: Autores

Como se puede observar en la tabla, se ha creado 3 datasets, uno por cada dominio. Se puede apreciar que cada dataset tiene sus restricciones en cuanto a la extensión del archivo permitido para respaldar. Por ejemplo, el dataset ORO_SET no admite el backup de archivos tipo mp3 o wav.

CONFIGURACIÓN DEL SCHEDULE

El schedule u horario, es la programación de la periodicidad con la cual se ejecutarán las tareas de respaldo. También se define las ventanas de tiempo dentro de las cuales operará una política específica.

De igual manera que en los datasets, esos horarios serán programados de acuerdo a la Tabla 3, en la cual se definió la periodicidad de los respaldos por cada dominio creado. En la tabla 6 se muestra la configuración de la herramienta schedule.

Se ha creado tres schedules con las mismas características de periodicidad a pesar que en la tabla 3 se especifica horarios diferen-

ciados por cada dominio. Esto es porque AVAMAR al utilizar compresión de la información y registro de bytes duplicados en el origen es capaz de obtener respaldos FULL a diario. A diferencia de otras herramientas que lo deben realizar semanalmente y diarios solamente incrementales para no afectar el performance y el espacio requerido para los backups.

Tabla 6
Resumen de configuración de schedules

Nombre del schedule	Nombre del dominio	Periodicidad	Ventana de operación
ORO_SCH	ORO	Full Diario	9AM-19PM
PLATA_SCH	PLATA	Full Diario	9AM-19PM
BRONCE_SCH	BRONCE	Full Diario	9AM-19PM

Nota. Ventana de operación para backup de información

Elaboración: Autores

CONFIGURACIÓN DE LA POLÍTICA DE RETENCIÓN

Se muestra la configuración de la política de retención para cada dominio. Hay que recordar que la retención indica el tiempo que el respaldo estará guardado en el sistema. Luego de pasado el tiempo de retención el respaldo es borrado. Normalmente los sistemas de respaldo tienen esta rotación en tiempo, sería poco práctico guardar copias indefinidamente ya que no habría sistema que soporte tal cantidad de información.

Un resumen de cómo se configura las políticas de retención por cada uno de los dominios creados se puede observar en la tabla 7.

Se puede observar en la tabla anterior, es el período de retención el parámetro que varía de un dominio a otro. Se considera que el dominio ORO es de usuarios más críticos, en ese caso gerentes, por ello los períodos de tiempo durante los cuales sus backups permanecen disponibles es mayor respecto a los otros dominios.

Tabla 7
Resumen de configuración de las políticas de retención

Nombre del Schedule	Nombre del dominio	Retención para el backup diario	Retención para el backup semanal
ORO_RET	ORO	2 semanas	9AM-19PM
PLATA_RET	PLATA	1 semana	9AM-19PM
BRONCE_RET	BRONCE	1 semana	9AM-19PM

Nota. Ventana de operación para backup de información

Elaboración: Autores

El dominio BRONCE es considerado el de menor prioridad y por ende sus períodos de retención de información son menores. En general esa es una política que se aplica en escenarios reales. Es decir, segmentar los usuarios por grado de criticidad y de esta manera administrar los recursos del sistema de respaldo más eficientemente.

CONFIGURACIÓN DE GRUPO

El último paso de configuración para empezar a realizar los respaldos de las máquinas de usuario final, es el de creación de grupos de backup, a los cuales se asigna a cada una de las políticas configuradas de retención, schedule y datasets creadas anteriormente.

Se observa en la tabla 8 el resumen de los grupos creados y las políticas asignadas a cada uno.

Tabla 8
Resumen de configuración de grupos de respaldo

Grupo	Dominio	Dataset	Schedule	Política de retención
ORO_GRP	ORO	ORO_SET	ORO_RET	ORO_RET
PLATA_GRP	PLATA	PLATA_SET	PLATA_RET	PLATA_RET
BRONCE_GRP	BRONCE	BRONCE_SET	BRONCE_RET	BRONCE_RET

Nota. Resumen configuración de respaldo de los grupos de usuarios

Elaboración: Autores

Configuración del módulo de respaldo de máquinas virtuales

Una vez concluidas las tareas de instalación del proxy como se describió en los capítulos anteriores, es necesario realizar la integración con el VCenter para respaldar las máquinas virtuales y que éstas puedan entrar dentro del esquema de respaldos. Como primer paso se debe crear un nuevo cliente del tipo VMware VCenter y colocar los datos para que el servidor de respaldos pueda conectarse al VCenter.

Concluidos estos pasos se tiene el cliente configurado en la herramienta y es necesario definir las máquinas virtuales que se desean respaldar.

f. Pruebas del sistema de respaldo de máquinas virtuales y usuarios finales

Las pruebas presentadas a continuación se muestran divididas dos módulos, en primer lugar, es validar la operatividad de la herramienta. Para eso se realizarán básicamente dos tipos de pruebas:

- Pruebas de respaldo de información
- Pruebas de recuperación de información

Con esa información finalmente se analiza el rendimiento ofrecido por la herramienta en el escenario plateando y establecer si el sistema es factible para un escenario real.

g. Pruebas sobre el módulo de respaldo a máquinas de usuario final

En la figura 14 se muestra el estado de la solución de respaldos antes de empezar a operar. Se puede observar el estado activo de los principales “demonios” o procesos que se ejecutan en la herramienta y que permiten la operación de la misma: Scheduler State, Maintenance Activities State, etc.

Figura 14
Estado de la herramienta de respaldo
antes de iniciar el primer backup



Elaboración: Autores

i. Pruebas de respaldo de máquina de usuario final

Primera prueba: Se realiza un primer backup del dominio ORO, en la figura 15 se muestra el status del respaldo.

Como se puede observar en la figura anterior, el primer backup del dominio ORO, tuvo una duración de 1 minuto con 8 segundos para respaldar un total de 14.5MB.

Segunda prueba: Ahora bien, se realiza un segundo full backup del mismo grupo sin alterar la información de la carpeta C:\Mis Backups, se puede observar en la siguiente figura No 43, esa operación tardo apenas 10 segundos a pesar de que la carpeta a respaldar tiene los mismos 14.5MB de información. Si se observa la columna “New Bytes”, ésta indica 0%, lo que quiere decir que AVAMAR no identificó información nueva y no pasó un solo byte de información a través de la red. Este último hecho es una de las características más

importantes de la herramienta, es decir aquella característica que le permite identificar únicamente el delta de variación de la información y enviar solo ese diferencial a través de la red.

Figura 15
Status del primer respaldo

Job	Error	Start Time	End Time	Progress	Size	Status
Completed		2015-04-14 10:23:00	2015-04-14 10:23:00	100%	14,5 MB	Completed
Waiting-Client		2015-04-14 10:23:00	2015-04-14 10:23:00	0%	0	Waiting-Client
Waiting-Client		2015-04-14 10:23:00	2015-04-14 10:23:00	0%	0	Waiting-Client
Completed		2015-04-14 10:50:00	2015-04-14 10:50:00	100%	39,4 MB	Completed
Completed		2015-04-14 10:50:00	2015-04-14 10:50:00	100%	21 MB	Completed
No Proxy		2015-04-14 10:43:00	2015-04-14 10:43:00	0%	0	No Proxy
No Proxy		2015-04-14 10:43:00	2015-04-14 10:43:00	0%	0	No Proxy
No Proxy		2015-04-14 10:37:00	2015-04-14 10:37:00	0%	0	No Proxy
No Proxy		2015-04-14 10:37:00	2015-04-14 10:37:00	0%	0	No Proxy

Elaboración: Autores

Las mismas pruebas fueron realizadas con los otros dominios creados, es decir con PLATA y BRONCE; en la tabla 9 podemos observar un resumen de los resultados de esas pruebas.

Tabla 9
Resumen de Resultados de los dos primeros backups

Dominio	Tiempo de backup		Bytes a respaldar		Bytes reales (no duplicados)	
	1er Backup	2do Backup	1er Backup	2do Backup	1er Backup	2do Backup
ORO	1m:8s	0m:10s	14,5MB	14,5MB	99,70%	0%
PLATA	0m:57s	0m:10s	39,4MB	39,4MB	71,70%	0%
BRONCE	1m:45s	0m:10s	21MB	21MB	99,10%	0%

Nota. Resultados de los dos primeros backups de los grupos de usuarios

Elaboración: Autores

En la tabla anterior se puede notar la tendencia de este tipo de herramienta de backups: a medida que más información se tenga respaldada más posibilidad de encontrar data duplicada existe. Es por ello que en el ejercicio del segundo backup la herramienta, al tener la información del primer respaldo, fue capaz de identificar duplicidad en toda la información por lo tanto los bytes reales guardados en el segundo backup son 0.

Tercera prueba: En esta tercera prueba se añade mucha más información a la carpeta C:\Mis Backups de uno de los usuarios en el dominio ORO. Exactamente esa misma información (archivos, carpetas) fue añadida a otro usuario del dominio PLATA.

Se realiza una operación de respaldo sobre el usuario del dominio ORO y se obtiene los resultados de la figura 16.

Figura 16
Tercera prueba sobre el dominio ORO

Status	Start Date	Start Time (UTC)	Elapsed	End Time (UTC)	Type	Server	Progress Bytes	New Bytes	Client	Container	Domain
Completed	2015-04-15 11:26	09h:09m:11s		2015-04-15 11:26	On-Demand	server1	35,8 MB	0%	system32.exe	Sub	PLATA
Completed	2015-04-15 11:29	09h:09m:11s		2015-04-15 11:30	On-Demand	server1	21,0 MB	0%	lsass.exe	Sub	ORO
Completed	2015-04-15 19:21	09h:09m:29s		2015-04-15 19:40	On-Demand	server1	25,0 GB	0,8%	Lab_Users	Sub	ORO
Completed	2015-04-15 19:21	09h:09m:57s		2015-04-15 19:21	On-Demand	server1	85,0 GB	<0,0%	Lab_Active Directory	Sub	ORO
Completed	2015-04-15 19:42	09h:09m:03s		2015-04-15 19:42	On-Demand	server1	795,3 MB	88,6%	user1	Sub	ORO

Elaboración: Autores

En esta prueba el tiempo de duración de la operación de backup fue de 2 minutos con 3 segundos. Se respaldó un total de 795,3MB de información. La cantidad de información nueva calculada por la herramienta fue del 88.6%, si tomamos esta última prueba realizada sobre el dominio PLATA y lo comparamos con un sistema tradicional de respaldos tendríamos los resultados que se muestran en la tabla No 10.

Tabla 10
Resumen de resultados de los dos primeros backups

Sistema de almacenamiento	Información cruda para respaldar [MB]	Información respaldada [MB]	Tiempo de respaldo [S]
AVAMAR	820	2,46	23
TRADICIONAL	820	820	66

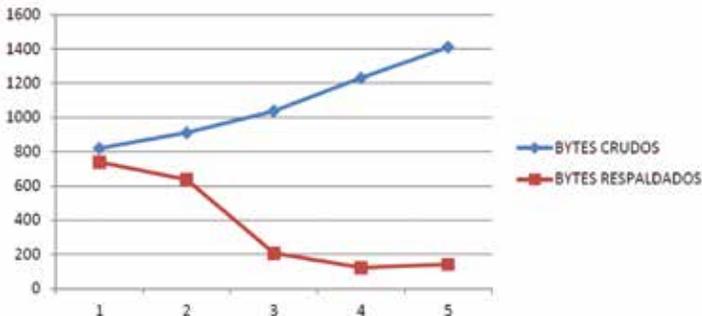
Nota. Resultados de los dos primeros backups

Elaboración: Autores

En una solución tradicional, la información total respaldada sería, para el ejercicio, de 820MB sobre 2.46MB que respalda AVAMAR. Es decir que el sistema fue aproximadamente 300 veces más eficiente un sistema tradicional.

Tendencia del sistema: Se realizaron varias pruebas más de respaldos, agregando más información a las carpetas C:\Mis Backups, de los usuarios del sistema, modificando el contenido de la misma, etc.; y se observa una tendencia en el comportamiento de la herramienta que se ve reflejada en la figura 17.

Figura 17
Tendencia de comportamiento
en bytes respaldados de la solución



Elaboración: Autores

En la Figura anterior, se puede observar una clara tendencia en el espacio de almacenamiento requerido a medida que las tareas de backup progresan, se puede apreciar mientras la curva azul, que representa los “Bytes Crudos” a respaldar, tiende a aumentar, la curva de “Bytes Respaldados” tiende a disminuir y a fijarse en un valor determinado.

PRUEBAS DE RESTAURACIONES DE INFORMACIÓN DE USUARIO FINAL

Se prueba la capacidad de recuperación granular de la herramienta, es decir recuperar un archivo o carpeta particular sin necesidad de restaurar todo el backup.

Prueba de restauración 1: La prueba simula la pérdida de un archivo en la carpeta C:\Mis Backups, de un usuario final en el dominio ORO. Se borra intencionalmente el archivo “velocity_provider_playbook.pdf” cuyo tamaño es de 27MB.

Desde la consola de administración se puede explorar un backup específico y navegar por las carpetas respaldadas hasta encontrar el archivo específico a respaldar. Se lo selecciona y se ejecuta la orden de “recuperar ahora”. Por defecto el archivo será recuperado en su ubicación original, ver figura 18.

Figura 18
Tarea de recuperación de un único archivo completada

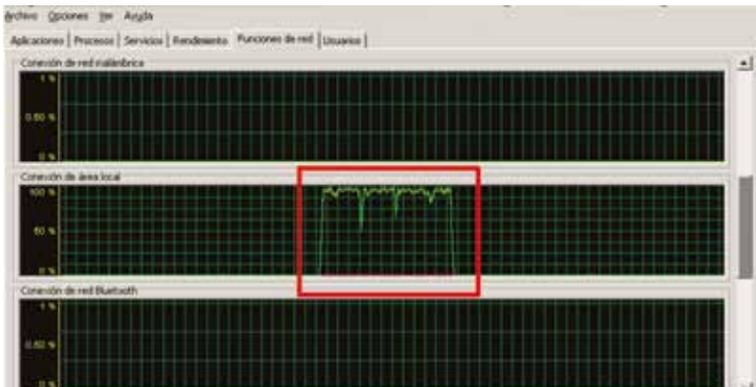
Status	Error Code	Start Time (COF)	Elapsed	End Time (COF)	Type	Server	Progress Bytes	New Bytes	Client	Container	Details
Client Not		2015-04-15 09:38	00:00:00	2015-04-15 09:38	On-Demand	Avast	0	0	Unavail	isa	SPONS
Completed		2015-04-15 09:38	00:01:45	2015-04-15 09:40	On-Demand	Avast	21.5 MB	99.7%	Unavail	isa	SPONS
Completed		2015-04-15 11:25	00:00:17	2015-04-15 11:25	On-Demand	Avast	26.4 MB	71.7%	rsynccmd2pc	isa	PLATA
Completed		2015-04-15 11:29	00:00:15	2015-04-15 11:30	On-Demand	Avast	21.5 MB	3%	Unavail	isa	SPONS
Completed		2015-04-15 16:30	00:00:29	2015-04-15 16:40	On-Demand	Avast	202.0 MB	8.6%	Lab_Users	isa	SHAREPOINT 163 163 454
Completed		2015-04-15 16:30	00:00:57	2015-04-15 16:30	On-Demand	Avast	60.0 MB	<0.00%	Lab_Active_Dr	isa	SHAREPOINT 163 163 454
Completed		2015-04-15 16:40	00:00:04	2015-04-15 16:40	On-Demand	Avast	736.3 MB	98.6%	isast-gc	isa	ORO
Completed		2015-04-15 17:00	00:00:26	2015-04-15 17:00	On-Demand	Avast	820.2 MB	8.3%	rsynccmd2pc	isa	PLATA
Running		2015-04-15 19:19	00:00:16	2015-04-15 19:19	Restore	Avast	2.0 MB	100%	Lab_Users	isa	SHAREPOINT 163 163 454
Completed		2015-04-15 19:24	00:00:00	2015-04-15 19:24	Restore	Avast	771.3 MB	3.7%	isast-gc	isa	ORO

Elaboración: Autores

En la figura anterior se muestra la operación de restauración del archivo concluida. El archivo tenía un tamaño de 27MB y la recuperación tardó 5 segundos el cual es un tiempo bastante aceptable.

Prueba de Restauración 2: Esta prueba se simulará haber perdido todo el contenido de la carpeta C:\Mis Backups. Para este ejercicio la carpeta tiene un tamaño de 800MB aproximadamente. Se mide la rapidez del respaldo en ese escenario, La recuperación de toda la carpeta de backups tardó 1 minuto con 13 segundos, ver figura 51.

Figura 19
Consumo de ancho de banda durante
la operación de recuperación



Elaboración: Autores

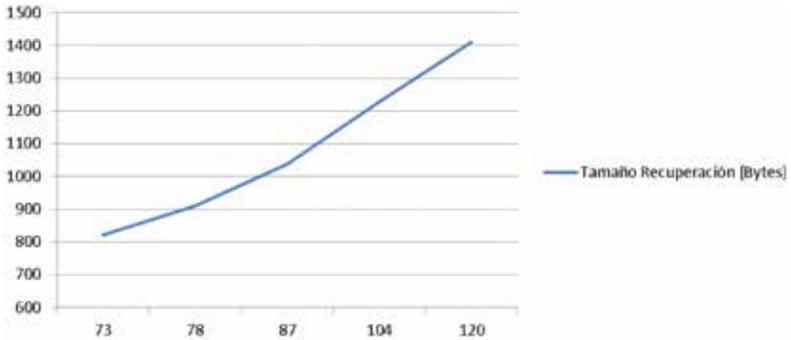
Como muestra la figura anterior, para esa prueba se mide el comportamiento de la tarjeta de red NIC. Comprobando que esa tarea de restauración llega a consumir casi el 100% del ancho de banda disponible en la interfaz, figura 20.

PRUEBAS DE RESTRICCIÓN DEL ANCHO DE BANDA

Uno de las características del agente encargado de las tareas de respaldo en AMAVAR puede generar problemas dependiendo del escenario y es aquella en la que, durante una tarea de respaldo o res-

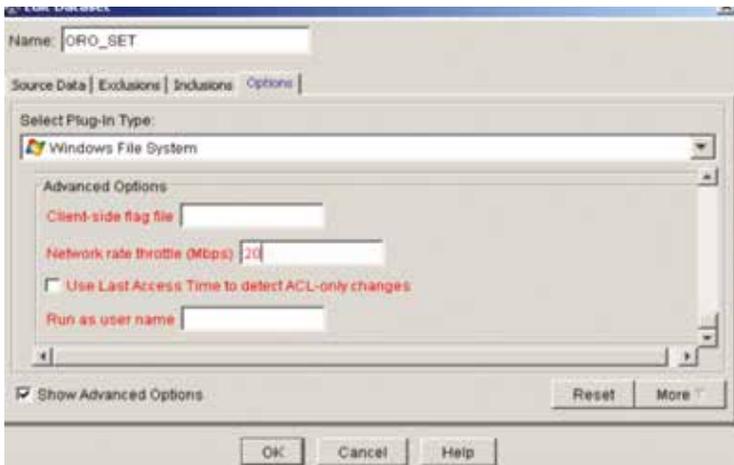
tauración, se tienen a utilizar todo el ancho de banda disponible de la NIC, Se realizará la misma prueba de backup sobre el mismo cliente, pero con la siguiente opción activada, ver figura 21.

Figura 20
Tiempo de la operación de restauración vs bytes restaurados



Elaboración: Autores

Figura 21
Configuración de la limitación del ancho de banda para los backups



Elaboración: Autores

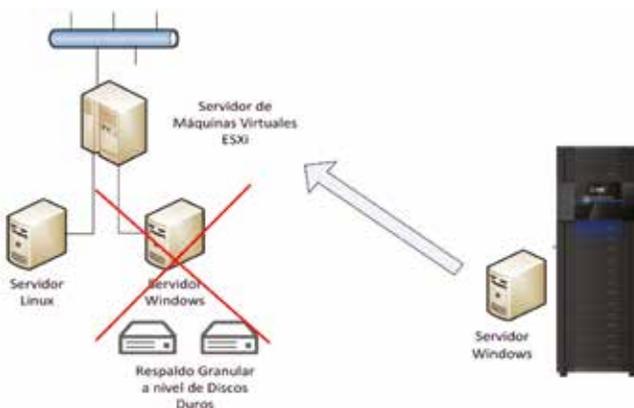
En la figura anterior se muestra la configuración del parámetro “network rate throttle” sobre el dataset del dominio ORO. Ese parámetro permite configurar el avatar del cliente para que mantenga un promedio determinado de uso del ancho de banda durante una operación de backup. Para el caso de la prueba se ha establecido el promedio en 20Mbps.

h. Pruebas sobre el módulo de respaldo de máquinas virtuales de máquinas virtuales

Para las pruebas de las herramienta de respaldo de máquinas virtuales se consideró un ambiente heterogéneo que se compone de máquinas con sistemas operativos Windows y Linux. En esos equipos se ejecuta deliberadamente acciones que degradan el funcionamiento de los equipos dejándolos inoperativos. A partir de ahí se realiza la recuperación de las máquinas virtuales mediante la integración de la herramienta de respaldo al hypervisor de VMware para la recuperación de las mismas, en la figura 22 se explica el escenario de pruebas que se utiliza.

Figura 22

Escenario de pruebas para respaldo de máquinas virtuales



Elaboración: Autores

Pruebas de recuperación a nivel de máquinas virtuales

Escenario 1: En el escenario 1 se ejecuta un respaldo inicial de las máquinas virtuales e inmediatamente se ejecuta un segundo respaldo para observar los tiempos que le toma a la herramienta completar la tarea y la optimización del tiempo de respaldo de mediante la tecnología de duplicación.

Escenario 2: En el escenario 2 se trabaja con una máquina virtual basada en Linux que sufre una falla que la dejó inoperativa y con daños a nivel de los bloques de archivos del sistema operativo. Posterior al fallo del equipo se ejecuta una recuperación de la máquina a un estado previo con el backup existente. Se realiza pruebas de consistencia y funcionamiento del equipo virtual.

Escenario 3: En el escenario 3 se trabaja sobre un servidor Windows que sufre un daño irreversible y es necesaria su restauración desde un respaldo de la herramienta de respaldos. Para ese ejercicio se realizará una eliminación completa de la máquina en VCenter y posterior restauración.

Escenario 4: En el escenario 4 se tiene una base de datos SQL que tiene su información albergada en la unidad D, el cual ha sufrido un fallo y pierde información de la misma. En ese escenario se busca recuperar la unidad D y realizar pruebas sobre la base de datos para verificar que la información se encuentra sin cambios al estado previo a la pérdida de datos.

Pruebas de duplicidad y optimización de respaldos

RESULTADOS PRUEBAS SOBRE EL ESCENARIO 1

Posterior a la configuración de la herramienta se realiza un respaldo inicial de las máquinas virtuales en las cuales se busca determinar la eficiencia de la herramienta de backup con su tecnología de compresión una vez que se realizan respaldos posteriores.

En la tabla 11 se presenta un comparativo en el escenario de pruebas 1 en el que se observa que al ejecutarse por segunda ocasión el respaldo una mejora en tiempo promedio 27 veces, adicionalmente la información nueva que ingresa en el sistema de respaldo es inferior al 1% y representa un ahorro de recursos de almacenamiento en un sistema en que su recurso más limitado es el espacio. Además, esa tecnología de compresión optimiza el uso de recursos de red LAN y WAN en ambientes dispersos en los que optar por un respaldos máquinas virtuales completas puede representar un gran peso para la red y ocasionar congestiones y saturación que degradan la experiencia de usuario al acceder a los servicios de la red del cliente.

Tabla 11
Resumen de resultados de los backups de máquinas virtuales

Máquina virtual	Tiempo de backup		Bytes a respaldar		Bytes reales (no duplicados)	
	1er Backup	2do Backup	1er Backup	2do Backup	1er Backup	2do Backup
LAB_ACTIVE_DIRECTORY	49m:14s	0m:54s	60GB	60GB	5,30%	0,05%
LAB:UBUNTU	33m:56s	8m:25s	20MB	20GB	10,50%	0,6%

Nota. Resultados de los dos primeros backups de máquinas virtuales.

Elaboración: Autores

Resultados pruebas sobre el escenario 2: La primera acción que se realiza sobre la máquina virtual consiste en ejecutar el comando indicado en la figura que deja inoperativa la máquina. Ante ese escenario se procede a forzar el apagado de la máquina y se procede a ejecutar una restauración de un respaldo previo de la máquina virtual.

Como aspecto adicional se puede resaltar que el proceso de apagado de la máquina virtual lanza errores ante la inconsistencia de la data que está en la misma, ante eso se forzó el apagado de la máquina y se procede con el respaldo desde la consola de “Backup & Restore” de la herramienta de respaldos.

Dentro de la herramienta se escoge el dominio de respaldos y desplegamos las opciones para el VCenter en el cual se puede encontrar las máquinas virtuales que están siendo respaldadas y el respaldo más reciente en el historial. Escogidas las opciones y el respaldo más reciente se procede a lanzar la recuperación de la máquina virtual, ver figura 23.

Figura 23
Entorno para recuperación de la máquina virtual



Elaboración: Autores

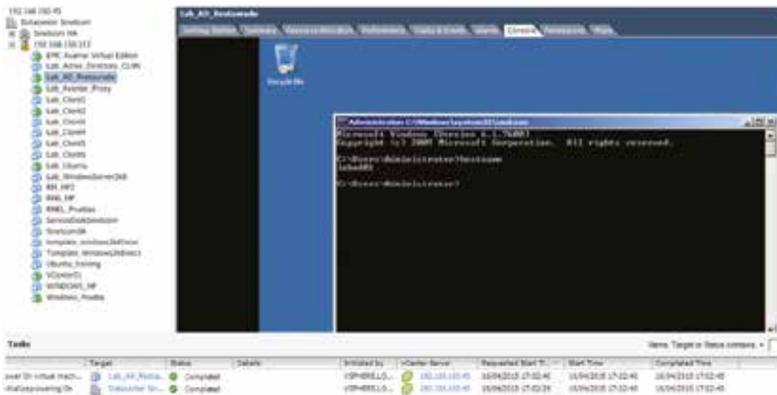
Posterior a iniciar el proceso de recuperación tendrá una ventana de confirmación en la que se puede revisar los parámetros para la restauración de la máquina virtual y confirmar el proceso.

Resultados pruebas sobre el escenario 3: El escenario se ha simulado la pérdida total de la máquina virtual desde el VCenter, para esto se elimina la máquina virtual del datastore de VMware y se verifica que no existe conectividad ni posibilidad de acceso a la misma. Posterior a eso se ejecuta la restauración completa de la máquina virtual y se verifica los tiempos de recuperación completa de la máquina virtual. Concluido el proceso de eliminación de la máquina se procede al proceso de recuperación.

Antes de proceder con la restauración de la máquina virtual es necesario crear una máquina virtual de similares características a la

máquina que se desea recuperar. Se puso especial cuidado en crear la máquina con la misma cantidad de datastores (unidades de disco) para posibilitar el proceso de restauración, una vez lanzado ese proceso se verifica que la restauración de la máquina virtual Windows demoró aproximadamente 24 minutos y se recuperó completamente la máquina. Se realizaron pruebas de acceso remoto y se comprobó que la máquina se restauró correctamente como se indica en la figura 24.

Figura 24
Verificación de la restauración de la máquina virtual



Elaboración: Autores

Resultados pruebas sobre el escenario 4: Para la ejecución de la prueba se realiza una verificación inicial de la base de datos SQL para verificar el estado inicial de la misma, posterior a esto se procede a borrar la base de datos y se realiza una recuperación granular del servidor virtual para recuperar la base de datos eliminada y se realiza una verificación final para verificar con los resultados iniciales, para el escenario se procede a bajar el motor de base de datos desde los servicios de Windows y se borra físicamente el archivo de testdb.mdf. Inmediatamente se ejecuta la restauración desde la consola de la herramienta de respaldos para la unidad D donde se encontraba la base de datos como se indica en las opciones de restauración de la figura 25.

CONCLUSIONES Y RECOMENDACIONES

Conclusiones

- Un sistema de respaldo y recuperación apropiado para ambientes virtuales debe tomar en cuenta lo expuesto en el punto anterior, por lo tanto, deberá utilizar mecanismos que permitan la no contención al momento de lanzar una operación de backup. La prioridad de estos sistemas es la optimización en la utilización de recursos del ambiente virtual.
- Existen dos estrategias básicas que una herramienta de respaldo debe manejar para operar en ambientes virtuales: utilización de un agente para respaldo y respaldo a través de imágenes. Las aplicaciones más críticas pueden utilizar esta última estrategia y las otras el agente.
- Mediante el uso del mecanismo de respaldo a través de snapshots o imágenes de las máquinas virtuales, se utilizan todas las facilidades que ofrece el hypervisor, VMware por ejemplo, para optimizar los backups. Por ejemplo, Avamar se integra con VMware a través de VADP para poder respaldar las VM de una manera no disruptiva sin tiempos de downtime del ambiente productivo.
- La estrategia para el respaldo del ambiente de usuarios finales, es la de colocar un agente o software en el computador del cliente final. Este agente puede manejar de manera automática las operaciones de respaldo y restauración y además aplicar mecanismos de optimización del respaldo, por ejemplo, compresión de información para encontrar duplicada en los datos y enviar por la red solo información nueva.
- AVAMAR, cumple con las exigencias básicas que un ambiente virtual y de usuarios finales requiere para un óptimo esquema de backup. Avamar cuenta con un módulo especializado para el respaldo de usuarios finales denominado Desktop/Laptop, y otro para el respaldo de máquinas virtuales a través de un proxy o intermediario.

- Una de las características claves de la solución de respaldo escogida, es su capacidad para hallar duplicidad de información en el origen de datos. Se pudo observar, en el capítulo de diseño y dimensionamiento de la solución, que Avamar logra niveles de compresión de más del 90%.
- La capacidad de compresión que logra la herramienta de respaldos permite ahorrar en recursos del host respaldado (CPU, RAM, etc.) y en ancho de banda. El cliente de Avamar solo envía información nueva por la red, por lo tanto, el ahorro de ancho de banda, tiempo, y recursos físicos es sustancial. Esto se pudo corroborar en las pruebas realizadas a la herramienta.
- El proceso de implementación e instalación de esta herramienta es bastante sencillo. Una persona con conocimientos básicos de Linux y de VSphere puede realizar la instalación de la herramienta en menos de tres horas. Esto le da una ventaja a esta solución frente a otras como TSM de IBM cuya instalación y configuración es mucho más compleja.
- En las pruebas de la funcionalidad de respaldo de usuario final, fue posible comprobar que mientras los primeros backups duraban algunos minutos, los siguientes tuvieron una duración del orden de segundos. Y esta tendencia se mantuvo para los respaldos subsecuentes. Este comportamiento, como se estudió, se manifiesta gracias a la habilidad de la herramienta de detectar información duplicada. A mayor información respaldada mayor índice de duplicidad de la información.
- Un factor importante a considerar en los sistemas de respaldo en general es el consumo del ancho de banda. Especialmente en los sistemas de respaldo a través de la LAN, ya que tanto la información de producción como la de backup compartirán el mismo medio y recursos. Se pudo constatar durante la fase de pruebas de la herramienta que, en una operación de respaldo el agente instalado en las

máquinas de usuario final tendía a consumir todo el ancho de banda disponible.

- Durante los escenarios de respaldo y recuperación de máquinas virtuales se pudo verificar que los respaldos iniciales demandan de gran cantidad de tiempo y espacio, pero al ejecutarse estas tareas una segunda vez en adelante los tiempos de respuesta y consumo de recursos de espacio disminuyen considerablemente por la tecnología de compresión en el origen de la herramienta de respaldos.
- La recuperación de máquinas virtuales que han sufrido daños irreversibles desde un respaldo de la herramienta requieren de una máquina virtual de similares características a la máquina dañada. Este procedimiento se asemeja bastante a las recuperaciones Bare Metal que consisten en recuperar una máquina física respaldada en un hardware de similares características que quedó inoperativo.
- Los respaldos granulares, como en el escenario de respaldo y recuperación de la base de datos SQL, permiten recuperar información de discos locales de la máquina virtual sin afectar discos específicos como el de sistema operativo que pudieron tener cambios y actualizaciones que pueden requerirse no regresar a puntos anteriores en el tiempo.

Recomendaciones

- Para el respaldo de ambientes virtuales especialmente, se recomienda tener las precauciones debidas respecto a la herramienta de backup seleccionada para este propósito. La solución de backup debería al menos manejar dos sistemas de respaldo (por agente y por imagen de VM), integrarse con la plataforma virtual y utilizar los APIs disponibles del hypervisor.
- Es muy importante pensar en los respaldos como en un sistema integral. Muchas de las empresas tienen sistemas de respaldo independientes para cada ambiente. Se reco-

mienda analizar las opciones del mercado y buscar una herramienta que integre todos estos escenarios en una sola solución centralizada desde la cual se pueda controlar todo el ambiente de backups, esto es usuarios finales, máquinas virtuales y servidores físicos.

- Para tener un ambiente de respaldos y recuperación confiable se recomienda en primer lugar tener una política sólida de backups en la empresa. Deben definirse las prioridades de las aplicaciones correctamente, la periodicidad de los respaldos, períodos de retención, etc. Estas políticas dependerán del RTO y RPO particular del negocio.
- Para el respaldo de máquinas virtuales críticas se recomienda realizarlo a través de la utilización del método de imágenes o snapshots usando un proxy. El proxy permitirá el respaldo no disruptivo de las máquinas virtuales que un agente sí provocaría. Además, la utilización de este esquema libera los recursos del host ya que la carga del proceso del backup la asumiría el proxy y el VCenter en el caso de VMware.
- Para el ambiente de respaldos de usuarios finales se recomienda tener presente el consumo de ancho de banda, ya que necesariamente este respaldo debe realizarse en horas laborables mientras el computador del cliente esté prendido. Por lo tanto, se debería considerar una herramienta que permita por una parte el ahorro de ancho de banda con comprensión, y por otro el control del ancho de banda que el agente pueda utilizar para limitarlo y que no lo monopolice.

BIBLIOGRAFÍA

- Agüero Torales, M. M. (2013). *Sistemas de Bases de Datos Federadas para la Gestión de la Información*. Recuperado de <https://goo.gl/FwSXJi>
- Chen, F., Li, P., Xu, H., & Xie, W. (2019). Research on Data Storage and Processing Optimization Based on Federation HDFS and Spark. En L. Barolli, N. Javaid, M. Ikeda, & M. Takizawa (Eds.), *Complex, Intelligent, and Software Intensive Systems* (Vol. 772, pp. 1047-1057). Cham: Springer International Publishing Ag.

- Cho, A. S., & Kang, T. H.-K. (2018). Load-Carrying Performance and Hydrostatic Tests of Encapsulated Anchor Systems for Unbonded Post-Tensioning Single-Strands. *International Journal of Concrete Structures and Materials*, 12(1), UNSP 15. <https://doi.org/10.1186/s40069-018-0233-z>
- Choo, K. R., Rana, O. F., & Rajarajan, M. (2017). Cloud Security Engineering: Theory, Practice and Future Research. *IEEE Transactions on Cloud Computing*, 5(3), 372-374. <https://doi.org/10.1109/TCC.2016.2582278>
- Deng, S., Gu, C., Cai, G., & Li, Y. (2018). Financing Multiple Heterogeneous Suppliers in Assembly Systems: Buyer Finance vs. Bank Finance. *McSom-Manufacturing & Service Operations Management*, 20(1), 53-69. <https://doi.org/10.1287/msom.2017.0677>
- DuBois, L., & Amatruda, R. (s. f.). Backup and Recovery: Accelerating Efficiency and Driving Down IT Costs Using Data Deduplication, 16. EMC Avamar 7.1 Administration Guide. (s. f.), 380.
- Gheza, F., Oginni, F. O., Crivellaro, S., Masrur, M. A., & Adisa, A. O. (2018). Affordable Laparoscopic Camera System (ALCS) Designed for Low- and Middle-Income Countries: A Feasibility Study. *World Journal of Surgery*, 42(11), 3501-3507. <https://doi.org/10.1007/s00268-018-4657-z>
- Hamilton, J., & Olsen, E. W. (2003). Design and implementation of a storage repository using commonality factoring. En F. Titsworth (Ed.), *20th Ieee/11th Nasa Goddard Conference on Mass Storage and Technologies (msst 2003), Proceedings* (pp. 178-182). Los Alamitos: Ieee Computer Soc.
- Linthicum, D. S. (2017). Cloud Computing Changes Data Integration Forever: What's Needed Right Now. *IEEE Cloud Computing*, 4(3), 50-53. <https://doi.org/10.1109/MCC.2017.47>
- Macías, M., Barría, C., Acuna, A., & Cubillos, C. (2016). SGSI support through malware's classification using a pattern analysis. En *2016 IEEE International Conference on Automatica (ICA-ACCA)* (pp. 1-4). <https://doi.org/10.1109/ICA-ACCA.2016.7778516>
- Muzammal, M., Qu, Q., & Nasrulin, B. (2019). Renovating blockchain with distributed databases: An open source system. *Future Generation Computer Systems-the International Journal of Escience*, 90, 105-117. <https://doi.org/10.1016/j.future.2018.07.042>
- Ochoa Carrión, L. P. (2017). *Aplicación ría basada en ontología para la gestión de procesos administrativos en la Universidad Estatal Amazónica*.

- Ortega Tenezaca, D. B., & Núñez Miranda, C. I. (2017). *Las bases de datos federadas y su incidencia en la optimización del flujo de información académica institucional universitaria*. Recuperado de <https://goo.gl/hVxM7e>
- Riaz, S., Rehan, M., Umer, T., Afzal, M. K., Rehan, W., Munir, E. U., & Iqbal, T. (2018). FRP: A novel fast rerouting protocol with multi-link-failure recovery for mission-critical WSN. *Future Generation Computer Systems-the International Journal of Escience*, 89, 148-165. <https://doi.org/10.1016/j.future.2018.06.029>
- Sailema, C., & Gistavo, W. (2015). *Análisis, consideraciones de diseño e implementación en laboratorio de un sistema de respaldo de datos de máquinas virtuales y usuario final a través de la red LAN. Caso de estudio AVAMAR*. Pontificia Universidad Católica del Ecuador. Recuperado de <https://goo.gl/3q5a9e>
- Su, L., Wan, H., Qin, Y., Zhao, X., Gao, Y., Song, X., ... Gu, M. (2019). Synthesizing Fault-Tolerant Schedule for Time-Triggered Network Without Hot Backup. *Ieee Transactions on Industrial Electronics*, 66(2), 1345-1355. <https://doi.org/10.1109/TIE.2018.2833022>
- Teneda, C., & Stefania, A. (2016). *Sistema de Federaciones de Identidades para la Facultad de Ingeniería en Sistemas, Electrónica e Industrial usando software de código abierto*. Recuperado de <https://goo.gl/rD2eJb>
- Yi-Wen, Z. (2019). Energy-aware mixed partitioning scheduling in standby-sparing systems. *Computer Standards & Interfaces*, 61, 129-136. <https://doi.org/10.1016/j.csi.2018.06.004>
- Zhang, L., Marzano, G., Sasaki, Y., Kurata, M., & Skalomenos, K. (2018). Force redistribution of steel moment-resisting frame retrofitted with a minimal disturbance arm damper. *Soil Dynamics and Earthquake Engineering*, 114, 159-173. <https://doi.org/10.1016/j.soildyn.2018.06.035>

Sobre los autores

Daniel Plúa

Docente Investigador de la Universidad Politécnica Salesiana, Sede Guayaquil, miembro del Grupo de Investigación (GIIAR). Graduado de Magíster en Sistemas de Información Gerencial de la Escuela Superior Politécnica del Litoral (ESPOL). Correo electrónico: dplua@ups.edu.ec

Bernabé Ortega

Docente Investigador de la Carrera de Ingeniería en Sistemas e Ingeniería en Software de la Universidad Regional Autónoma de los Andes - Uniandes Puyo. Graduado de Magister en Gestión de Bases de datos de la Universidad Técnica de Ambato y Magister en Docencia Universitaria y Administración Educativa de la Universidad Tecnológica Indoamérica. Estudiante de Doctorado en Tecnologías de la Información y las Comunicaciones en la Universidad de Coruña en España. Correo electrónico: up.delfinortega@uniandes.edu.ec

Carlos Núñez

Docente de la Universidad Técnica de Ambato. Graduado de Ingeniero en Sistemas Computacionales e Informáticos y Magister en Gestión de Bases de Datos obtenido en la Universidad Técnica de Ambato. Correo electrónico: ci.nunez@uta.edu.ec

Oswaldo Villacrés

Docente Investigador de la Escuela Superior Politécnica de Chimborazo (ESPOCH), miembro del Grupo de Investigación de Ciencia de Datos. Graduado de Ingeniero en Sistemas Informáticos de la ESPOCH y Magíster en Gestión de Bases de Datos de la Universidad Técnica de Ambato. Correo electrónico: ovillacres@esepoch.edu.ec

Wilson Chango

Docente Investigador de la Pontificia Universidad Católica del Ecuador (PUCE), Sede Esmeraldas, integrante del Grupo de Investigación de Ingeniería de Software y el Grupo de Investigación de Redes y Comunicaciones. Graduado de Ingeniero en Sistemas Informáticos de la Escuela Superior Politécnica de Chimborazo (ESPOCH) y Magíster en Redes de Comunicaciones de la Pontificia Universidad Católica del Ecuador (PUCE). Correo electrónico: wilson.chango@pucese.edu.ec

Jaime Sayago

Docente Investigador de la Pontificia Universidad Católica del Ecuador (PUCE), Sede Esmeraldas, integrante del Grupo de Investigación de Ingeniería de Software. Graduado de Ingeniero de Sistemas, Magíster en Sistemas de Información Geográfica de la Universidad San Francisco de Quito (USFQ) y Máster en Investigación en Ingeniería de Software de la Universidad Nacional de Educación a Distancia (UNED) en España. Correo electrónico: jaime.sayago@pucese.edu.ec

Víctor Quiñonez

Docente Investigador de la Pontificia Universidad Católica del Ecuador (PUCE), Sede Esmeraldas, integrante del Grupo de Investigación de Ingeniería de Software. Graduado de Ingeniero en Sistemas Informáticos, Magíster en Administración y Dirección de Empresas de la Universidad Tecnológica de Israel y Master Univer-

sitario en Investigación en Ingeniería de Software de la Universidad Nacional de Educación a Distancia (UNED) en España. Correo electrónico: xavier.quinonez@pucese.edu.ec

Ligia Jácome

Docente de la Universidad Tecnológica Indoamérica, miembro del Grupo de Investigación de Ingeniería de Software. Graduada de Ingeniera en Informática y Computación de la Universidad Central del Ecuador en Quito y Magíster en Tecnologías de la Información y Multimedia Educativa de la Universidad Técnica de Ambato. Correo electrónico: ligiajacome@uti.edu.ec

Este libro es el esfuerzo de varios autores y se compone de la recopilación de trabajos de investigación que han sido sustentados como tesis de Maestría.

En cuatro capítulos se tratan temas claves para entender el funcionamiento de los sistemas de información. En el primer capítulo se menciona la importancia de la auditoría en empresas de telecomunicaciones para evitar pérdidas económicas y se plantea una metodología que permite la construcción de alarmas o informes automáticos para determinar el origen de los problemas del sistema auditado.

El segundo capítulo tiene como objetivo proponer una solución de optimización en el flujo de la información académica institucional, aplicada sobre el Sistema de Información Académico Docente. En él se almacena la mayoría de registros relacionados con la generación de evidencias de acreditación universitaria, control interno de procesos académicos, investigación, vinculación con la comunidad y gestión, con análisis de optimización del flujo de información basado en el uso del algoritmo KP.

El tercer capítulo aborda el tema de las bases de datos autónomas heterogéneas y describe mecanismos para utilizar los datos almacenados para calcular indicadores del Modelo Genérico de Evaluación de Carreras. Como ejemplo, se estudia el caso de la Escuela Superior Politécnica de Chimborazo, que ha desarrollado una solución informática que interconecta bases de datos autónomas.

Por último, se analiza la creación de un sistema para el respaldo de la información de ambientes virtuales y de computadores de usuarios finales a través de la red LAN. Para ello se presenta el análisis, diseño e implementación a nivel de laboratorio, de un sistema de respaldos especializado para estos ambientes a la vez que se comprobó factibilidad de ejecución desde un punto de vista técnico.

