DEPARTAMENTO DE CIENCIAS DE LA COMPUTACIÓN

CARRERA DE INGENIERÍA EN SISTEMAS E INFORMÁTICA

PROYECTO DE TITULACIÓN PREVIO LA OBTENCIÓN DEL TÍTULO DE INGENIERO EN SISTEMAS E INFORMÁTICA

TEMA: PLATAFORMA DE GESTIÓN Y MITIGACIÓN A FAVOR DE INFRAESTRUCTURAS CRÍTICAS DE LAS UNIVERSIDADES MIEMBROS DE CEDIA

AUTOR: VALLADARES RUIZ BRYAN PAÚL

DIRECTOR: ING. WALTER FUERTES, PHD

SANGOLQUÍ, MAYO DE 2017

CERTIFICADO

DEPARTAMENTO DE CIENCIAS DE LA COMPUTACIÓN
CARRERA DE INGENIERÍA EN SISTEMAS E INFORMÁTICA

Certifico que el trabajo de titulación, "Plataforma de Gestión y Mitigación a Favor de Infraestructuras Críticas de las Universidades miembros de CEDIA" realizado por el señor Bryan Paúl Valladares Ruiz, ha sido revisado en su totalidad y analizado por el software anti-plagio, el mismo cumple con los requisitos teóricos, científicos, técnicos, metodológicos y legales establecidos por la Universidad de Fuerzas Armadas ESPE, por lo tanto me permito acreditarlo y autorizar al señor Bryan Paúl Valladares Ruiz para que lo sustente públicamente.

Sangolquí, 9 de Mayo de 2017

Ing. Walter Fuertes, PhD.

DIRECTOR

AUTORÍA DE RESPONSABILIDAD

DEPARTAMENTO DE CIENCIAS DE LA COMPUTACIÓN
CARRERA DE INGENIERÍA EN SISTEMAS E INFORMÁTICA

Yo, Bryan Paúl Valladares Ruiz, con cédula de identidad N° 1718117706, declaro que este trabajo de titulación "Plataforma de Gestión y Mitigación a Favor de Infraestructuras Críticas de las Universidades miembros de CEDIA", ha sido desarrollado considerando los métodos de investigación existentes, así como también se ha respetado los derechos intelectuales de terceros considerándose en las citas bibliográficas.

Consecuentemente declaro que este trabajo es de mi autoría, en virtud de ello me declaro responsable del contenido, veracidad y alcance de la investigación mencionada.

Sangolquí, 9 de Mayo de 2017

Bryan Paúl Valladares Ruiz

C.C 1718117706

AUTORIZACIÓN

DEPARTAMENTO DE CIENCIAS DE LA COMPUTACIÓN
CARRERA DE INGENIERÍA EN SISTEMAS E INFORMÁTICA

Yo, Bryan Paúl Valladares Ruiz, autorizo a la Universidad de las Fuerzas Armadas ESPE publicar en la biblioteca Virtual de la institución el presente trabajo de titulación "Plataforma de Gestión y Mitigación a Favor de Infraestructuras Críticas de las Universidades miembros de CEDIA", cuyo contenido, ideas y criterios son de mi autoría y responsabilidad.

Sangolquí, 8 de Mayo de 2017

Bryan Paúl Valladares Ruiz

C.C. 1718117706

DEDICATORIA

La presente tesis la dedico a Dios, por guiarme en cada paso del camino, ayudarme, darme fuerza para continuar y mejorar cada día. Siento sus bendiciones en cada momento de mi vida.

A mis padres cuyo ejemplo de trabajo y sacrificio me han hecho salir adelante y dar lo mejor de mí en cada tarea desempeñada. A mi madre Gladys Ruiz por siempre estar pendiente de mi formación, sus consejos y cariño han hecho de mí una persona de bien. A mi padre Renan Valladares, su ejemplo diario de trabajo y esfuerzo me motivan para siempre actuar con responsabilidad, rectitud y honestidad pese a las dificultades que se presenten.

A mis hermanos que siempre estan apoyándome, su ejemplo de perseverancia y superación me inspira a llegar siempre a lo más alto. A Sandy por estar conmigo en cada momento, su amor y dedicación llenan mi corazón.

AGRADECIMIENTO

Quiero agradecer primero a Dios, por llenarme de bendiciones, regarlarme una familia maravillosa, darme siempre fuerza para seguir adelante sin desmayar. Gracias por darme la vida y ayudarme a obtener este nuevo triunfo.

A mis amados padres, sin su apoyo nunca lo habría logrado, gracias por todo su esfuerzo y sacrificio para hacer de mí un hombre de bien. A mi padre por siempre dar el ejemplo de trabajo, honradez y rectitud. A mi madre por su amor y eterno esfuerzo, por sus consejos que siempre los llevaré conmigo.

A mis hermanos por su apoyo incondicional y por su ayuda, son un ejemplo para mí.

A mi director de tesis, Ing. Walter Fuertes, PhD, por confiar en mí para desarrollar este proyecto, por sus consejos, guía, conocimientos y ejemplo de valores para poder cumplir todos los objetivos con calidad.

A todas las personas que me han guiado y han sido amigos de verdad, a los profesores, por impartir no solo conocimientos sino también valores, todos quienes me acompañaron hasta culminar este objetivo.

ÍNDICE

ÍNDICE DE TABLAS

ÍNDICE DE FIGURAS

RESUMEN

El crecimiento y la evolución de las amenazas, las vulnerabilidades y los ciberataques aumentan los incidentes de seguridad y generan impactos negativos en las organizaciones. Este estudio presenta un sistema de procesamiento analítico en línea (OLAP) para alertas tempranas de actividades maliciosas. El objetivo de esta plataforma es sistematizar el apoyo a la ciberseguridad provisto por un Equipo de Respuesta a Incidentes de Seguridad Informática (CSIRT) y establecer un mecanismo de análisis con el fin de mejorar el nivel general de seguridad de redes y equipos mediante servicios de alerta temprana. Para cumplir este objetivo, se ha desarrollado una solución de inteligencia de negocios adaptando la metodología de desarrollo de Ralph Kimball para apoyar el análisis de incidentes de seguridad informática. Esta metodología genera un almacén de datos de información recopilada de alertas y eventos grabados de una transmisión continua de datos de varias fuentes de seguridad de Internet que recopilan, rastrean y reportan malware, botnet y fraude electrónico. Además, con Pentaho BI se desarrolla procesos de carga de dimensiones, medidas y hechos, cubos OLAP, informes y cuadros de mando. Los resultados obtenidos demuestran claramente la funcionalidad de la aplicación, donde es posible visualizar con certeza, tanto las advertencias tempranas como el nivel de seguridad de las Instituciones miembros sobre las amenazas y vulnerabilidades registradas.

PALABRAS CLAVES:

- **BUSINESS INTELLIGENCE**
- **ALERTAS TEMPRANAS DE ATAQUES COMPUTACIONALES**
- **CUBOS OLAP**
- **MODELO DE DATOS DIMENSIONAL**
- **METODOLOGÍA DE RALPH KIMBALL**

ABSTRACT

The growth and evolution of threats, vulnerabilities and cyber-attacks increase security incidents and generate negative impacts on organizations. We present an online analytical processing (OLAP) system for early alerts of upcoming malicious activities. This study aims to systematize the support of cybersecurity granted by a Computer Security Incident Response Team (CSIRT) and shall help to establish a mechanism to analyze and improve the overall level of security of networks and equipment by providing early warning services. In order to accomplish this task, a business intelligence solution has been developed adapting the methodology of Ralph Kimball to support the analysis of computer security incidents. This generates a data warehouse of information collected from alerts and events recorded from a continuous transmission of data from various Internet security sources that gather, trace and report malware, botnet, and electronic fraud. Furthermore, we constructed with Pentaho BI load data into the dimensions, measures and facts, OLAP cubes, reports and dashboards. The acquired results clearly demonstrate the functionality of the application where it is possible to visualize with certainty of both, the early warnings, as well as the level of security of the participant Institutions, about the registered threats and vulnerabilities.

KEYWORDS:

- **BUSINESS INTELLIGENCE**
- **EARLY WARNING TO COMPUTER ATTACKS**
- **OLAP CUBES**
- **DIMENSIONAL DATA MODEL**
- **RALPH KIMBALL METHODOLOGY**

CAPÍTULO I
INTRODUCCIÓN

1.1 Antecedentes

De acuerdo a la recomendación X.1205 de la Unión Internacional de Telecomunicaciones (UIT), la "Ciberseguridad es el conjunto de herramientas, mejores prácticas, políticas, salvaguardas, normas, capacitación, seguros y tecnologías cuyo objetivo es proteger la infraestructura e información sensible de una institución y sus usuarios" (Unión Internacional de Telecomunicaciones, 2007). En esta recomendación se concluye que los activos de una institución y los usuarios son los dispositivos informáticos conectados, los servicios/aplicaciones, los sistemas de comunicaciones y la totalidad de la información transmitida y/o almacenada en la nube. Sin embargo, la realidad muestra que en el Ecuador y el mundo aún no se garantiza la seguridad de los activos de la organización y los usuarios contra los riesgos de seguridad en el ciberespacio.

De conformidad con la guía de seguridad para los países en desarrollo de la misma UIT (2007), se deben crear soluciones de seguridad confiables en las redes y equipos . Por lo tanto, la implantación de soluciones adecuadas de seguridad y confianza, constituye uno de los principales pilares para ayudar a los países a explotar los recursos de telecomunicaciones y de las TIC.

Manejar un nivel de seguridad adecuado para prevenir vulnerabilidades, riesgos tecnológicos e incidentes que pudiesen ocurrir, es primordial para el correcto desempeño de las instituciones. Al adoptar nuevas metodologías y tecnologías se depende más de las mismas y de las infraestructuras críticas que las alojan (Unión Internacional de Telecomunicaciones, 2007).

La seguridad informática o seguridad de tecnologías de la información es el área de la informática que utiliza procesos y prácticas diseñados para proteger las redes, equipos, programas, datos e información sensible y prevenir daños o accesos no autorizados. Incluye tanto la seguridad cibernética como la seguridad física (Mansur, 2012). Contribuye a la preservación de todos los medios tecnológicos,

informáticos y financieros, adquiridos por una organización, para cumplir sus objetivos (Unión Internacional de Telecomunicaciones, 2007).

Los ataques cibernéticos se han incrementado en los últimos años y uno de sus destinos son las entidades gubernamentales a nivel mundial (UIT, 2007). Motivo por el cual muchos investigadores están dedicando grandes esfuerzos para detectarlos y mitigarlos. Como consecuencia, estos ataques se ejecutan en infraestructuras críticas de las naciones, tales como los sistemas educativos, de salud y gubernamentales, afectando a la calidad de vida de los ciudadanos y por lo tanto oponiéndose a los principios marcados en el plan del buen vivir.

1.2 Problemática

Al plantear el proceso de la seguridad de la información, es importante identificar correctamente los activos y recursos que han de resguardarse, para poder definir con precisión el alcance de la seguridad y asegurar que la protección sea eficaz. Se debe desarrollar normas que dirijan la ética, transparencia y responsabilidad, integradas a un marco legal que contenga procedimientos y normas. Se los debe aplicar en la legislación individual y ampliar internacionalmente.

Para disminuir y evitar los ataques a la seguridad informática, las infraestructuras deben tener medidas de prevención técnicas y jurídicas. Los incidentes se presentan de distintas maneras como: denegación de servicios, virus, ataques de fuerza bruta, pishing, robo o destrucción de datos sensibles. Todos esto tiene consecuencias que ponen en riesgo a las instituciones que son víctimas.

Dados los anteriores indicadores, se desconoce cuáles serían las consecuencias futuras en caso de perpetrarse ataques a la seguridad de la información. Tampoco se conoce la frecuencia ni el tipo de los ataques que se llevarán a cabo, el tipo de técnicas que utilizarán los atacantes para apropiarse de la información confidencial de las empresas y de sus activos fijos, ya que no se cuenta con las herramientas necesarias para llegar a una solución.

En el panorama actual de ciberseguridad, no es posible prevenir ataques o brechas, los atacantes de ahora tienen una financiación importante, son pacientes y sofisticados y tienen puesta la mira en personas y procesos así como en la tecnología.

Las organizaciones cada vez confían más en la información digital y comparten una vasta cantidad de datos alrededor del globo, se han convertido en blancos fáciles para diferentes formas de ataques. Como resultado, las operaciones que se realizan día a día en las instituciones, datos y propiedad intelectual están en serio riesgo. En un contexto corporativo, un ataque puede no solo dañar la imagen y la reputación de la institución, sino que puede resultar en pérdida de ventaja competitiva, crear incumplimientos legales/regulatorios y causar grave daño financiero.

Se debe considerar que un negocio actual sin ningún registro o información computarizada o cualquier presencia en línea, tendrá un éxito muy limitado, si es que lo tiene. Tomando ventaja de todos los sistemas computarizados y lo que Internet tiene que ofrecer, puede verdaderamente elevar el status del negocio para que se vuelva eficiente y exitoso. Sin embargo, esto también significa que se están exponiendo a los potenciales efectos perjudiciales de las amenazas de seguridad cibernética. Por lo tanto, con el fin de expandir ampliamente un negocio necesita poner atención a la inteligencia de seguridad, incluso internamente, ya que las infracciones no intencionales de seguridad informática pueden ser increíblemente devastadoras.

El sesenta y cinco por ciento de los que respondieron la encuesta Global Information Security Survey (GISS) en el año 2014 ven amenazas en su entorno de seguridad de la información.

En un entorno de amenazas en constante evolución, de ataques rápidos, cibercriminalidad y actividades de espionaje, los enfoques tradicionales serán cada vez más importantes de mantener, no obstante no serán suficientes para abordar adecuadamente el riesgo. El entorno seguro de hoy, tendrá vulnerabilidades en el mañana, por lo que una organización no puede ser complaciente.

El panorama tecnológico esta evolucionado rápidamente y aquellas organizaciones que no se mantienen al día con él, se quedarán atrás. Las tecnologías avanzadas ofrecen nuevas capacidades y beneficios, sin embargo también introducen nuevos riesgos, superando la capacidad de evaluarlos adecuadamente.

1.3 Justificación

La organización y sistematización de una capacidad efectiva de respuesta a incidentes de seguridad informática implica varias decisiones e innovaciones. Según informó el NIST (Cichonski, 2012), un equipo de respuesta a incidentes, también conocido como Equipo de Respuesta a Incidentes de Seguridad Informática (CSIRT) o Equipo de Respuesta para Emergencias Informáticas (CERT), es responsable de proporcionar servicios de respuesta a incidentes de una organización. Conforme al NIST (Johnson, 2016), el CSIRT / CERT recibe información sobre posibles incidentes, los investiga y toma medidas para asegurar que los daños causados por los incidentes puedan ser minimizados. En los últimos años, se han utilizado técnicas relacionadas con el almacenamiento de datos y la inteligencia de negocios para proporcionar un buen apoyo como parte de las técnicas modernas utilizadas por los CSIRT.

Debido a la necesidad expuesta, la comunidad científica ha desarrollado soluciones importantes como por ejemplo (Hellwig, 2016), (Yang, 2016), (Bollinger, 2015) 2011), (Belsis, 2005) se refieren a las nuevas técnicas que se utilizan en CSIRT para mejorar el tratamiento de incidentes de seguridad informática y cómo analizar la información obtenida. Otros documentos de investigación recientes como (Elmellas, 2016), (Sharkov, 2016), (Mejía, 2016) denotan cómo la seguridad de la información puede alinearse con los objetivos empresariales para permitir mejores decisiones con la información obtenida. Otros estudios (Wu, 2014), (Rajasekharaiah, 2016), analizan la importancia de la Inteligencia de Negocios en la gestión de riesgos en las instituciones y las alertas tempranas para prevenir futuros desastres. Por último, destacan los beneficios de la analítica de Big Data y la inteligencia de negocios para las revisiones de los desafíos de seguridad y privacidad en las grandes empresas de la industria de la informática (Gahi, 2016). A pesar de todos estos

esfuerzos, los estudios no muestran claramente cómo las tecnologías modernas han interactuado, ni explícitamente cómo articular un modelo de datos dimensional e inteligencia de negocios para obtener alertas tempranas de CSIRT.

Los eventos de seguridad informática tienen efectos fatales en instituciones, las que están empezando a darse cuenta de que necesitan "sensores de riesgo", para evitar que el cibercrimen se convierta en un hecho.

Las organizaciones carecen de las soluciones disponibles y diferenciadoras necesarias para ayudarles a prevenir las pérdidas ante las amenazas cibernéticas. Para todas las soluciones altamente técnicas y de bajo nivel de software y hardware, herramientas de seguridad, personas y ciberpolíticas adoptadas hoy en el mercado, no existe una herramienta consistente y de "tamaño adecuado" para tomar decisiones con respecto a los eventos de seguridad informática.

Entonces, ¿Cómo se obtiene información actualizada y contínua sobre lo que podría estar esperando a la vuelta de la esquina para dañarlos? Y lo que es más importante, ¿Qué se planifica para evitarlo? La respuesta es Cyber Business Intelligence.

La mayoría de las empresas usan la inteligencia de negocios tradicional, como una parte vital de sus operaciones para rastrear los Indicadores de Desempeño Clave, más conocidos simplemente como KPIs. Para todo, desde el rendimiento de ventas, el compromiso de los clientes, la penetración de la marca y la efectividad del marketing hasta la retención de empleados y por supuesto, el rendimiento financiero, la inteligencia de negocios es una verdadera mercancía comercial. Los ejecutivos obtienen información clave y la utilizan para mantener todos los aspectos de su negocio en buen camino.

Hoy en día, la mayoría de las empresas no tienen casi ninguna forma de inteligencia de negocio para eventos de seguridad informática, para el seguimiento de KPIs que podrían mantenerlos más seguros y libres de riesgo.

Por lo anteriormente señalado, este proyecto resulta útil porque persigue investigar, diseñar e innovar soluciones a los Eventos de Ciberseguridad, con el fin de alertar y disminuir las vulnerabilidades y amenazas a los sistemas de información y las redes de datos a las Universidades miembros de CEDIA, pudiendo ser utilizados los resultados posteriormente para mejorar los sistemas de gestión de riesgos de las mismas. Los objetivos propuestos para el presente proyecto de titulación se describen a continuación.

1.4 Objetivos

1.4.1 Objetivo General

Realizar la sistematización del soporte de Ciberseguridad otorgado por el CSIRT de CEDIA a las Universidades miembros.

1.4.2 Objetivos Específicos

- Investigar las prácticas actuales del soporte que brindan los CSIRT del Ecuador a la comunidad.
- Realizar la definición de procedimientos de soporte del CSIRT de CEDIA
- Implementar una solución de Business Intelligence para analizar las alertas reportadas y determinar el nivel de seguridad de las Universidades miembros de CEDIA.
- Realizar la evaluación y validación de resultados.

1.5 Alcance

Realizar una investigación para determinar las prácticas actuales de soporte que brindan los CSIRT del Ecuador y definir los procedimientos de soporte de CSIRT de CEDIA.

• Desarrollar una solución de Business Intelligence utilizando la metodología de Kimball para analizar las alertas reportadas al CSIRT de CEDIA, en base a la información recopilada de alertas y eventos registrados a partir de flujos continuos de

datos (feeds) de escáneres tales como: Zone-h, TurkBot, Clean MX Foundation, CERT.br, Team Cymru, Nessus, Netcraft, Shadowserver.

• Investigar conjuntamente con los especialistas de CEDIA, el nivel de sensitividad de los eventos registrados y la clasificación de la información, para determinar el nivel de seguridad de las Instituciones miembros de CEDIA y mejorarla.

• Para realizar la investigación y clasificación de los eventos, considerar los parámetros dados por el FIRST (Forum of Incident Response and Security Teams) y la experiencia de los especialistas del CSIRT de CEDIA que proporcione una referencia confiable para cualquier tipo de organización, pública o privada.

Para el cumplimiento de los objetivos y el alcance de la sistematización del soporte de Ciberseguridad otorgado por el CSIRT de CEDIA a las Universidades miembros, se requiere una metodología experimental, en este caso aplicar la Metodología De Kimball que conducirá hacia la solución del problema expuesto anteriormente y también requerir una fundamentación teórica que contribuye a dar una visión global de los conceptos necesarios que ayudan a respaldar y abordar el problema del presente trabajo de titulación. En el siguiente capítulo se puntualiza más específicamente el marco teórico y la metodología usada en este proyecto.

CAPÍTULO II
MARCO TEÓRICO

En este capítulo se especfica los fundamentos teóricos / técnicos que se aplicaron en el desarrollo de la plataforma de Business Intelligence, en el cual se describe el marco teórico y las herramientas utilizadas.

2.1 Modelo de Datos Dimensional

De acuerdo con Kimball (Kimball R. , Kimball Group, 2008), el modelamiento dimensional es una técnica de diseño lógico que busca presentar los datos en un marco estándar e intuitivo. Es inherentemente dimensional y se adhiere a una disciplina que utiliza el modelo relacional con algunas restricciones importantes. Cada modelo dimensional está compuesto por una tabla de hechos con una clave multiparte y un conjunto de dimensiones, las que tienen una clave primaria que corresponde a uno de los componentes de la clave multiparte de la tabla de hechos. Esta característica "tipo estrella" a menudo se llama una estrella de unión.

Los modelos dimensionales se pueden instanciar tanto en bases de datos relacionales, denominados esquemas en estrella, como también en multidimensionales, conocidas como cubos de procesamiento analítico en línea (OLAP). Los esquemas de estrella, consisten en tablas de hechos vinculadas a las de dimensiones asociadas a través de relaciones de claves primarias. Los cubos OLAP contienen atributos y hechos dimensionales. Se accede a través de lenguajes con capacidad analítica superior a la de SQL, como XMLA. Los cubos OLAP son a menudo el paso de despliegue final de un sistema DW / BI dimensional, o puede existir como una estructura agregada basada en un esquema de estrella relacional más atómica (Kimball R. , 2005).

2.2 Procesos ETL (Extract, Transform and Load)

Según Kimball, el sistema de extracción, transformación y carga (ETL) consume una proporción extensa de tiempo y esfuerzo requeridos para construir un entorno de data warehouse y business intelligence (DW / BI). El desarrollo del sistema ETL es un desafío, porque muchas restricciones externas ejercen presión

sobre su diseño, tales como: los requisitos de negocio, las realidades de los datos de origen, el presupuesto, las ventanas de procesamiento y los conjuntos de habilidades del personal disponible. Sin embargo, puede ser difícil comprender por qué el sistema ETL es tan complejo y requiere muchos recursos (Kimball R. , 2013).

El proceso ETL inicia con la obtención de los datos de su ubicación original, se transforman, adaptan y dan formato a los mismos. Posteriormente se carga en un conjunto final de tablas para consulta de los usuarios. El diseño y construcción del sistema ETL depende de la fuente, las limitaciones de los datos, los lenguajes de script y las herramientas ETL y de BI disponibles y las habilidades del personal. Se debe adoptar un enfoque estructurado para el desarrollo.

2.3 Dashboard

De acuerdo con Rouse, un dashboard es una herramienta de visualización de datos que muestra el estado actual de las métricas y los indicadores clave de rendimiento (KPI) de una institución. Los Dashboards consolidan y organizan números, información esencial y scorecards de rendimiento en una sola pantalla. Las características esenciales de un dashboard de BI incluyen: una interfaz personalizable y la capacidad de extraer datos en tiempo real de múltiples fuentes. Es como el tablero de instrumentos de un automóvil, indica el estado en un punto específico en el tiempo, muestra el progreso en el tiempo hacia objetivos específicos (Rouse, 2012).

2.4 Modelo de Datos Tipo Estrella

La arquitectura de datos tipo estrella es un almacén de datos cuyo diagrama se asemeja a una estrella. En el centro consta de una tabla de hechos y los puntos de la estrella son dimensiones. Generalmente las tablas de hechos en un esquema de tipo estrella, están en tercera forma normal (3NF), esto quiere decir que no debe existir dependencia transitiva. Mientras que las de dimensionales son des-normalizadas. Las principales características de este modelo son:

Estructura simple: Posee efectividad en consultas, el tiempo de carga de datos en las tablas de dimensiones es relativamente largo, la redundancia de datos causa que el tamaño pueda ser grande.

Comúnmente utilizado en las implementaciones de data warehouse, apoyado por un gran número de herramientas de inteligencia de negocios (Kimball, 2013).

2.5 Análisis Ad-Hoc

De acuerdo con Guzik (Guzik, 2011), Ad-Hoc es un análisis de datos en donde hay una flexibilidad en cuanto a los formatos, consultas predeterminadas, valores preseleccionados, dimensiones, hechos. El objetivo de estos análisis es dejar la mayor libertad posible a los usuarios y analistas para realizar consultas de forma abierta, sin ningún tipo de restricciones o limitaciones previas de modelos ya predefinidos o construidos.

Una consulta Ad-Hoc se crea con el fin de obtener información cuando surge la necesidad y se compone de SQL construido dinámicamente que suele ser creado por las herramientas de consulta. En contraste con cualquier consulta que se predefine y se realiza rutinariamente.

Los usuarios analizan diversos tipos de datos, múltiples conjuntos de consultas que están predefinidas bajo la administración de una base de datos o desarrollador del sistema y por lo tanto existe una barrera entre las necesidades de los usuarios y la información. Los recursos de TI también reciben un alto costo, ya que un usuario puede tener que ejecutar varias consultas diferentes en un período determinado. Hoy en día, los almacenes de datos aceleran la recuperación de información vital para responder a las consultas interactivas en una aplicación.

La mayoría de los usuarios de datos son de hecho personas no técnicas. Existen muchas herramientas de consulta ad hoc para que los usuarios puedan ejecutar consultas muy complejas sin intentar saber qué sucede en el backend. El software de consultas ad hoc incluyen características que admiten todos los tipos de relaciones de consulta. Los usuarios finales pueden construir fácilmente consultas complejas

utilizando una interfaz de usuario (GUI) a través de estructuras de objetos de forma arrastrar y soltar (Guzik, 2011).

2.6 CSIRT (Computer Security Incident Response Team)

"CSIRT (Computer Security Incident Response Team) ofrece servicios de respuesta a incidentes de seguridad informática 24x7 a cualquier usuario, compañía, agencia gubernamental u organización. Proporciona un punto de contacto único y confiable para reportar incidentes de seguridad informática en todo el mundo. Provee los medios para reportar incidentes y para difundir información importante relacionada con eventos de seguridad." (CSIRT, 2006).

Muchas empresas no han considerado adecuadamente los problemas de seguridad durante las operaciones diarias normales. El CSIRT sirve para sensibilizar a sus clientes sobre los problemas de seguridad informática y proporciona información para la protección segura de la infraestructura y el equipo informático crítico frente a posibles ataques informáticos organizados.

Las organizaciones deben compartir la responsabilidad de coordinar sus esfuerzos de respuesta con otras instituciones similares. Recopilar información de todas las fuentes y escáneres es crítico en la protección de la infraestructura de la información. Crear redes en un entorno seguro y de confianza, el intercambio de información sobre incidentes y técnicas de detección y respuesta desempeñan un papel importante en la identificación y corrección de vulnerabilidades.

La Figura 1 se describe de la siguiente manera. Según el (US-CERT, 2005) los procesos de Preparar y Proteger se muestran continuos por encima y por debajo de los procesos de detección, clasificación y respuesta. Estos involucran la implementación de todo el personal necesario, tecnología, infraestructura, políticas y procedimientos necesarios para que las actividades de manejo de incidentes ocurran de manera oportuna, coordinada y efectiva. Las pequeñas flechas que entran en el proceso de Preparar y Proteger indican requisitos, políticas o reglas que regirán la

estructura y función de los mismos. La flecha dirigida desde preparar a proteger se traduce en recomendaciones de mejora. Al reportar o detectar un evento se lo clasifica y se envía una notificación a la institución afectada.

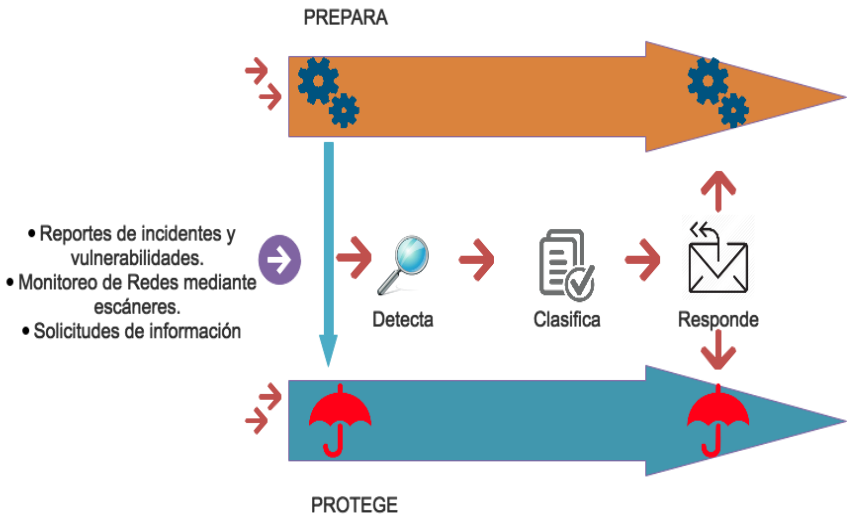


Figura 1. Manejo de Incidentes del CSIRT

2.7 Business Intelligence (BI)

Stackowiak (2007) define Business Intelligence como el proceso de tomar grandes cantidades de datos, analizar esos datos y presentar un conjunto de informes de alto nivel que condensan la esencia de esos datos en base a las acciones del negocio, permitiendo a la gerencia tomar las decisiones empresariales fundamentales diarias. (Cuiet, 2007) considera el BI como una manera y un método de mejorar el desempeño del negocio proporcionando asistencias poderosas para que al encargado de tomar decisiones ejecutivas le permita tener información procesable a mano. Las herramientas de BI son vistas como una tecnología que permite la eficiencia de la operación en una institución.

BI, se define como la aplicación de un conjunto de metodologías y tecnologías, tales como J2EE, DOTNET, Servicios Web, XML, data warehouse, OLAP, Data Mining, Tecnologías de representación, entre otras, para mejorar la eficacia de la operación de la empresa, la gestión de apoyo y decisión para lograr ventajas competitivas. Business Intelligence es una nueva tecnología en lugar de una

solución integrada para las empresas, dentro de la cual el requisito de negocio es definitivamente el factor clave que impulsa la innovación tecnológica. Por lo tanto, identificar y abordar de forma creativa las cuestiones clave de negocio, es el principal reto de una aplicación de BI para lograr un impacto real (Grupo Gartner, 2010).

BI ofrece beneficios a las instituciones que lo utilizan. Puede eliminar muchas conjeturas dentro de una organización, mejorar la comunicación entre las diferentes áreas, mientras se coordinan las actividades y permitir a las empresas responder rápidamente a los cambios en diferentes condiciones. BI mejora el rendimiento general de la institución que lo utiliza.

La información se considera a menudo como el segundo recurso más importante de una empresa (los activos más valiosos de una empresa son sus personas). Cuando una empresa puede tomar decisiones basadas en información oportuna y precisa, la institución puede mejorar su rendimiento. BI también acelera la toma de decisiones, para actuar rápida y correctamente con la información antes que las empresas competidoras. También puede mejorar la experiencia del cliente, al permitir una respuesta oportuna y adecuada a los problemas (RANJAN, 2009).

2.8 Metodología de Kimball

La metodología del ciclo de vida de Kimball fue concebida a mediados de la década de 1980 por miembros del Grupo Kimball y otros colegas de Metaphor Computer Systems, empresa pionera en apoyo a la toma de decisiones, desde entonces, ha sido utilizado con éxito por muchos equipos de proyecto de data warehouse y business intelligence (DW / BI) en prácticamente todas las industrias, áreas de aplicación, funciones empresariales y plataformas técnicas.

Originalmente denominado enfoque del ciclo de vida empresarial dimensional, esta denominación reforzó los principios básicos del método:

- Concentrarse en agregar valor empresarial.
- Dimensionar la estructura de los datos que se entregan a la empresa

- Desarrollar iterativamente el entorno DW / BI en incrementos de ciclo de vida manejables.

El enfoque del ciclo de vida de Kimball se ilustra en la Figura 2, que proporciona una hoja de ruta global que representa la secuencia de tareas de alto nivel requeridas para los proyectos de DW / BI exitosos (Kimball R. , Kimball Group, 2008).

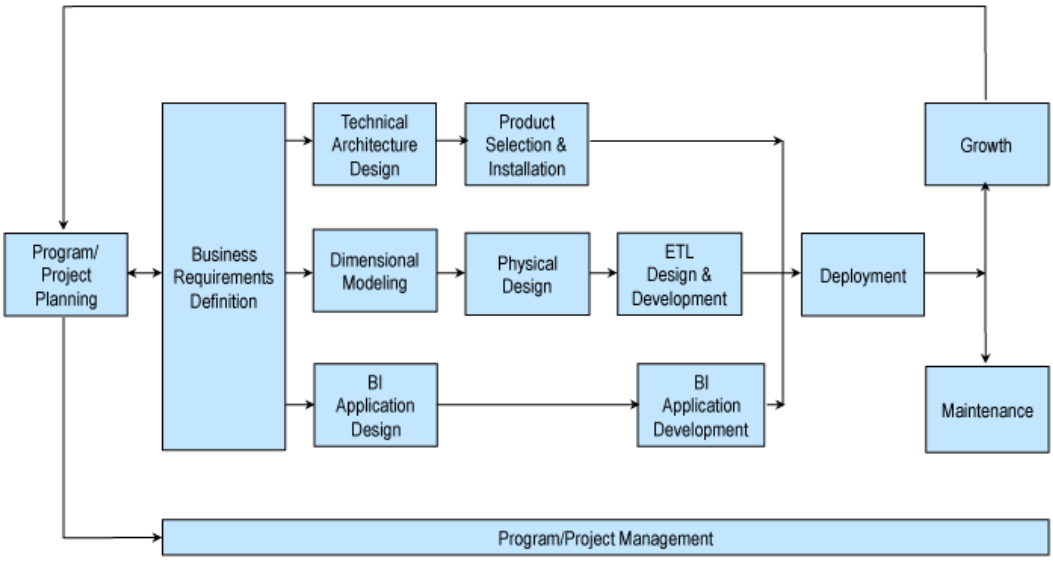


Figura 2. Metodología de Kimball

Fuente: (Kimball, 2008)

Cuando el enfoque fue publicado por primera vez en la década de 1990, la mayoría de los enfoques alternativos no enfatizaban los principios propuestos por Kimball. Sin embargo, desde entonces, han sido ampliamente adoptados y se han convertido en las mejores prácticas de la industria. Una mejor explicación de la metodología de Kimball se la puede encontrar en el Capítulo V del presente documento.

CAPÍTULO III
PRÁCTICAS Y PROCEDIMIENTOS DE LOS CSIRT EN EL ECUADOR

En este capítulo se describe las prácticas que realizan los diferentes CSIRT que están implementados en Ecuador, sus procedimientos y el soporte que realizan a favor de la comunidad para atender eventos de seguridad informática. Se detallan los procedimientos que realiza el CSIRT de CEDIA a las instituciones miembros y los tipos de eventos que registra.

3.1 Prácticas actuales del soporte que brindan los CSIRT del Ecuador a la comunidad

En Ecuador existen actualmente tres CSIRT implementados; el CSIRT de la Universidad Técnica Particular de Loja, el ECUCERT (Centro de Respuesta a Incidentes Informáticos del Ecuador) de carácter público y el CSIRT de CEDIA los que realizan procedimientos similares para recibir, atender y procesar los eventos de seguridad que ocurran.

Entre las principales actividades que realizan los CSIRT en el Ecuador se encuentran:

- Establecimiento de criterios generales y específicos para garantizar la seguridad de los servicios de telecomunicaciones, la información transmitida y la invulnerabilidad de la red, mediante la coordinación de la gestión de vulnerabilidades e incidentes de seguridad de la información.
- Establecimiento y mantenimiento de un vínculo fluido y una relación colaborativa con sus equivalentes en otros países, así como con organismos internacionales involucrados en ciberseguridad.
- Definición del impacto, alcance y naturaleza del evento o incidente.
- Entendimiento de la causa técnica del evento o incidente.

- Identificación de las causas alternativas u otras amenazas potenciales como resultado del evento o incidente.
- Investigación y recomendación de soluciones.
- Coordinación y apoyo la implementación de estrategias de respuesta con otras partes de las instituciones incluyendo grupos de TI y especialistas, grupos de seguridad física, oficiales de seguridad de la información (ISO), gerentes de negocios, ejecutivos, relaciones públicas, recursos humanos y asesoría legal.
- Difusión de información sobre riesgos, amenazas, ataques, exploits y estrategias de mitigación correspondientes a través de alertas, avisos, páginas web y otras publicaciones técnicas.
- Coordinación y colaboración con partes externas tales como vendedores, ISPs, otros grupos de seguridad y CSIRTs.
- Mantenimiento de un repositorio de datos sobre incidentes, vulnerabilidades y actividades relacionadas con la circunscripción que pueden utilizarse para la correlación, la elaboración de tendencias y el desarrollo de las lecciones aprendidas y así de esta forma mejorar la postura de seguridad y los procesos de gestión de incidentes de una organización.

3.2 Definición de procedimientos del soporte del CSIRT de CEDIA

El CSIRT de CEDIA ofrece los siguientes servicios de soporte a las instituciones miembros:

3.2.1 Respuesta a incidentes

CSIRT CEDIA apoya a los administradores en el manejo de aspectos técnicos y organizacionales de los eventos. Provee asistencia o aviso referente a los siguientes aspectos del manejo de incidentes:

- Investigación inicial
- Investigar si en efecto un evento ha ocurrido.
- Determinar el alcance.

3.2.2 Coordinación de incidentes

El CSIRT de CEDIA coordina los incidentes o eventos registrados en las redes de las Universidades miembros, otorga un seguimiento a los mismos y realiza las siguientes actividades:

- Determinar la causa inicial (vulnerabilidad explotada).
- Contactarse con otros sitios que pueden haber estado involucrados y con departamentos de seguridad de miembros del CEDIA y/o entidades a cargo del manejo de investigaciones judiciales y legales.
- Crear reportes para otros CSIRTs.
- Preparar anuncios a usuarios, si aplicara.

3.2.3 Resolución de incidentes

CSIRT CEDIA apoya en la resolución de eventos, otorgando soporte cuando las Universidades lo soliciten, entre las actividades realizadas para la solución de incidentes se encuentran las siguientes:

- Eliminar la vulnerabilidad
- Apoyo en el aseguramiento de sistemas derivados de lo aprendido en el incidente.
- Evaluar si ciertas acciones pueden arrojar resultados en proporción con su costo y riesgo, en particular acciones dirigidas a un eventual proceso judicial o acción disciplinaria, tales como: recolección de evidencias luego del hecho, observación de un incidente en progreso, plantando trampas al intruso, etc.
- Recolectar evidencia cuando se contemplen acciones judiciales, policiales o disciplinarias dentro de la organización donde ocurre el evento.

Además, CSIRT-CEDIA recolecta estadísticas referentes a incidentes que ocurran dentro de CEDIA o esté involucrado uno o varios de sus miembros, si es

preciso notifica a la comunidad para apoyar en la protección y disminuir vulnerabilidades contra ataques conocidos.

3.3 Tipos de eventos que maneja el CSIRT de CEDIA

A continuación, se describe los principales eventos que maneja el CSIRT de CEDIA, los que son detectados por sensores que envían flujos continuos de datos como: Team Cymru, Nessus, Netcraft, TurkBot y Zone-h, Shadowserver Foundation, CERT.br, Clean MX.

3.3.1 Botnets

Un botnet es una colección de computadoras, conectadas a Internet, que interactúan para lograr alguna tarea distribuida. Puede ser usada para aplicaciones útiles y constructivas, sin embargo, el término botnet se refiere típicamente a un sistema diseñado y usado para propósitos ilegales. Tales sistemas se componen de máquinas comprometidas que son controladas sin el conocimiento de su propietario. Las máquinas comprometidas se conocen como drones o zombies, el software malicioso que se ejecuta en ellos se llama "bot". Una colección de computadoras es inútil sin algún mecanismo de control, que constituye la interfaz entre la botnet y el guía. Con el control de tantos sistemas comprometidos, los pastores ahora pueden participar en varios tipos de actividades perjudiciales, por ejemplo: “click fraud”, ataques DDoS, “keylogging”, spam, entre otros (Shadowserver, 2017).

3.3.2 DNS Open Resolver

El Sistema de Nombres de Dominio (DNS) ha sido el objetivo de muchos tipos de ataques en los últimos años. Los servidores DNS autorizados están expuestos a Internet y generalmente permiten consultas de todas las direcciones IP. Sin embargo, los resolvers de DNS normalmente son internos a una organización y permiten consultas sólo desde los clientes internos que sirven. Los resolvers de DNS que permiten consultas de todas las direcciones IP y están expuestos a Internet, pueden ser atacados y utilizados para realizar ataques de Denegación de Servicio (DoS) (Infoblox Experts Community, 2014).

3.3.3 Sinkhole

Un sinkhole es un sistema que es capaz de recibir diferentes tipos de tráfico como HTTP o SMTP (E-Mail) y registrar ese tráfico. Su objetivo es reunir tanta información como sea posible sobre las conexiones y equipos. Esta información se puede utilizar para una variedad de propósitos como la corrección de vulnerabilidades (Shadowserver, 2016).

3.3.4 Ataques de amplificación basados en UDP

Un ataque de denegación de servicio reflexivo distribuido (DRDoS) es una forma de denegación de servicio distribuida (DDoS) que se basa en el uso de servidores UDP de acceso público, así como factores de amplificación de ancho de banda, para colapsar un sistema de víctimas con tráfico UDP (US-CERT, 2014).

3.3.5 Honeypots

Un honeypot es un recurso informático cuyo único propósito es ser explotado. Es una trampa para los delincuentes informáticos. Un honeypot atacado y debidamente investigado puede proporcionar información valiosa tanto sobre el ataque como sobre el atacante. Aunque los honeypots desempeñan un papel especializado en la red, están disfrazados como un recurso de red normal. Esto lo convierte en un objetivo más atractivo si el atacante lo ve como un activo valioso para aprovechar y no una trampa disfrazada y controlada (Shadowserver, 2015).

3.3.6 Ingeniería Social

La ingeniería social abarca una serie de técnicas destinadas a manipular a la "víctima" para revelar más de lo que debe divulgar sin conocimiento de ello. El ingeniero social intentará frecuentemente presionar a la víctima para que actúe de inmediato (por ejemplo, “Su cuenta se cerrará en 24 horas si no hace clic en el enlace de abajo para actualizar la información de su cuenta”), o alentar a la víctima a actuar instintivamente, también puede utilizar información obtenida previamente para engañar y ganarse la confianza de la víctima (Shadowserver, 2016).

3.3.7 Phishing

Phishing se refiere al uso de correos electrónicos presentados de tal manera que parecen comunicaciones oficiales de una organización bancaria, de servicios o minoristas, lo que lleva a la víctima a "confirmar" sus datos confidenciales. La información dirigida frecuentemente incluye detalles de acceso, datos de la tarjeta de crédito o de la cuenta bancaria, fecha de nacimiento y número de seguro social. Típicamente, el pretexto de la comunicación es algún tipo de medida de seguridad que se está implementando o respuesta a algún uso posiblemente fraudulento de la cuenta, la estafa es apoyada por la inclusión de imágenes oficiales y se presenta con cierto sentido de urgencia. El mecanismo utilizado para obtener los detalles del objetivo puede diferir, pero con frecuencia hacen uso de un enlace manipulado en el correo electrónico que aparentemente conduce al sitio oficial, pero que en realidad conduce a un servidor hackeado idéntico al original. Los datos recopilados se almacenan en un área oculta en el mismo servidor o en un servidor hackeado diferente o se envían a una cuenta de correo electrónico que el hacker supervisará durante la duración de la estafa (Shadowserver, 2016).

3.3.8 Gusanos

Los gusanos son variantes de malware que pueden propagarse por su cuenta. Contienen funcionalidades incorporadas que aprovechan redes informáticas y los mecanismos de transferencia de archivos que les permiten autocopiar e infectar otras máquinas. Para obtener acceso a los equipos de destino, los gusanos no necesitan ninguna interacción humana. Ellos penetran e infectan puramente a través de vulnerabilidades que son inherentes al propio sistema (Shadowserver, 2016).

3.3.9 Degradación de sitios web

La degradación es el cambio de apariencia de los sitios web agregando imágenes o palabras en el sitio web defectuoso. Se lo realiza por diferentes motivaciones: motivación política, transmitir puntos de vista. Los atacantes que están en contra de un gobierno o un movimiento en particular pueden optar por desfigurar

sitios web relacionados para transmitir sus ideales, estos son conocidos como “hacktivistas” (Trendmicro, 2017).

3.3.10 Ataque de fuerza bruta

Ataque de fuerza bruta es cuando un atacante utiliza un conjunto de valores predefinidos para atacar a un objetivo y analiza la respuesta hasta que tenga éxito, depende del conjunto de valores utilizado. Si es más grande, tomará más tiempo, pero hay una mejor probabilidad de éxito. Un ejemplo de fuerza bruta es el ataque de diccionario para romper una contraseña. En donde el atacante intenta combinar letras y números para generar la clave de forma secuencial. Estos ataques pueden tomar mucho tiempo, dependiendo del sistema utilizado y la longitud de la contraseña (Infosec, 2016).

Con los tipos de incidentes registrados y descritos en este capítulo y la información proporcionada por CEDIA se procede a realizar la clasificación de eventos de seguridad informática que se detalla en el siguiente capítulo de la presente tesis.

CAPÍTULO IV
CLASIFICACIÓN DE EVENTOS DE SEGURIDAD INFORMÁTICA

Posteriormente de describir los eventos registrados, se realiza una clasificación utilizando las guías provistas por el FIRST y los especialistas del CSIRT de CEDIA para obtener el nivel de seguridad de cada evento.

4.1 Alcance de la clasificación de eventos

De acuerdo al FIRST (Forum of Incident Response and Security Teams) es fundamental que el CSIRT proporcione una respuesta consistente y oportuna al cliente, y que la información sensible sea manejada adecuadamente. Basado en un documento publicado por el FIRST que proporciona las directrices necesarias para que los Administradores de Incidentes clasifiquen cada categoría, nivel de criticidad y nivel de sensibilidad para cada caso. Esta información es ingresada en el Sistema de Seguimiento de Incidentes cuando se crea un caso. En el presente proyecto son ingresados en la plataforma de Business Intelligence. Se requiere una clasificación de caso consistente para que el CSIRT proporcione informes precisos de manera regular. Además, las clasificaciones proporcionan a los Administradores de Incidentes del CSIRT los procedimientos adecuados de manejo de casos.

4.2 Clasificación por criticidad y sensibilidad

La matriz de criticidad y sensibilidad (Tabla 1). Define el mínimo tiempo de respuesta del cliente y los requisitos de comunicación para el caso. El nivel de criticidad debe ser ingresado en el ITS (Incident Tracking System) cuando el caso es creado y no debe ser alterado durante el ciclo de vida del caso, excepto cuando ha sido clasificado incorrectamente. Tipicamente el Administrador de Incidentes determinará el nivel de criticidad (FIRST, 2005).

Tabla 1

Clasificación de eventos por criticidad y sensibilidad (FIRST, 2005)

Nivel de Criticidad	Definición	Ejemplos típicos de Incidentes	Tiempo Inicial de Respuesta
1	Incidente que afecte sistemas críticos o información con potencial impacto en los clientes o las ganancias.	• Denegación de servicios • Activo Comprometido • Hacking Interno (Activo) • Hacking Externo (Activo) • Virus / Gusanos • Destrucción de Propiedad	60 minutos
2	Incidente que afecta sistemas no críticos o información, sin impacto a los clientes o las ganancias	• Hacking Interno (No Activo) • Hacking Externo (No Activo) • Acceso no autorizado. • Violación a las Políticas • Actividad fuera de la ley • Información Comprometida • Activo no crítico comprometido • Destrucción de propiedad no crítica.	4 horas
3	Posible incidente, sistemas no críticos.	• Email • Uso inapropiado de la propiedad • Violación a las políticas	48 horas
Definiciones:			

Continua

- **Tiempo Inicial de Respuesta: Especifica la cantidad máxima de tiempo que debe pasar antes de que el Administrador de Incidentes del CSIRT responda al cliente. En la mayoría de los casos el Administrador responderá antes del tiempo especificado. Como mínimo lo que debe ocurrir en este período de tiempo es:**
  1. **Evaluación inicial y selección**
  2. **Clasificación del evento**
  3. **Ingresar el caso al ITS (Incident Tracking System)**
  4. **Establecer la pertenencia del evento**
  5. **Enviar un email al cliente. Este correo incluye alguna información como la fecha/hora de la solicitud, número de caso, teléfono y email del administrador del incidente, el nivel de criticidad y sensitividad del evento y una indicación de cuando el cliente recibirá actualizaciones del caso.**

4.3 Tipos de eventos y su nivel de seguridad

De acuerdo con la Tabla 2, se clasifica a los eventos del CSIRT de CEDIA obteniendo una puntuación de 1 a 3: Alto(1), Medio(2), Bajo(3), para poder medir el nivel de seguridad.

Tabla 2

Tipo de Eventos y Nivel de Seguridad

Nivel	Tipo de Evento	
1	dns_openresolver	tc-openresolvers
	scan_ssl_poodle	tc-phishing
	compromised_website	tc-proxy
	scan_snmp	tc-routers
	scan_ssl_freak	tc-scanners
	spam_url	cleanmx-portals
	open_proxy	cleanmx-phishing

Continua

	tc-bruteforce tc-ddosreport tc-defacement tc-dipnet tc-fastflux tc-malwareurl tc-mydoom	hma-openproxy cediahp-wget zone-h-accepted scan_isakmp sh-cyc sh-botnetcc tc-nachi
2	botnet_drone microsoft_sinkhole sinkhole_http_drone scan_netbios scan_ntp scan_portmapper scan_ssdp scan_tftp cwsandbox_url certbrglobal sh-bots tc-slammer tc-spam tc-spreaders tc-stormworm tc-toxbox certbrspampot	scan_elasticsearch scan_ipmi scan_memcached scan_mongodb scan_ntpmonitor ssl_scan tc-beagle tc-blaster tc-bots tc-dameware tc-phabot tc-sinit cleanmx-viruses n6 cediahp-scan cediahp-fail cediahp-success
3	scan_mdns scan_mssql scan_xdmcp	scan_chargen scan_qotd scsummary

Una vez obtenido el nivel de seguridad de los eventos, se procede al desarrollo de la plataforma de Business Intelligence utilizando la metodología de Ralph Kimball, que se detalla en el siguiente capítulo.

CAPÍTULO V
DESARROLLO E IMPLEMENTACIÓN DE UNA SOLUCIÓN DE BUSINESS INTELLIGENCE PARA ANALIZAR LOS EVENTOS REPORTADOS AL CSIRT DE CEDIA

Luego de la clasificación de eventos, se procede al desarrollo de la plataforma de Business Intelligence utilizando la metodología de Ralph Kimball para obtener un modelo dimensional de datos, procesos ETL, cubos OLAP, reportes y dashboards de eventos reportados al CSIRT de CEDIA, con el fin de obtener los eventos e incidentes registrados y proveer de alertas tempranas a las Universidades miembro.

5.1 Marco conceptual

De conformidad con la metodología de Kimball, se detallan a continuación los procesos previos al desarrollo: planificación del proyecto, sustentación de la solución, descripción de la institución y sustento de la solución de Business Intelligence.

5.1.1 Planificación del Proyecto

En esta etapa se divide el desarrollo del proyecto en los subsiguientes capítulos principales, de acuerdo a los lineamientos de la metodología de Kimball: Marco Conceptual, Análisis de la solución, diseño, desarrollo y pruebas. (i) En el marco conceptual, se describe el problema y la solución propuesta. (ii) En el análisis se detalla la metodología a ser utilizada, elicitación de requerimientos y pruebas de concepto. (iii) En lo referente a diseño, se establece un modelo dimensional, arquitectura de la solución y procesos ETL(Extract, Transform and Load). (iv) En la parte de desarrollo, se realizan los procesos ETL, dashboards y reportes. La Figura 3 representa las fases del proyecto.

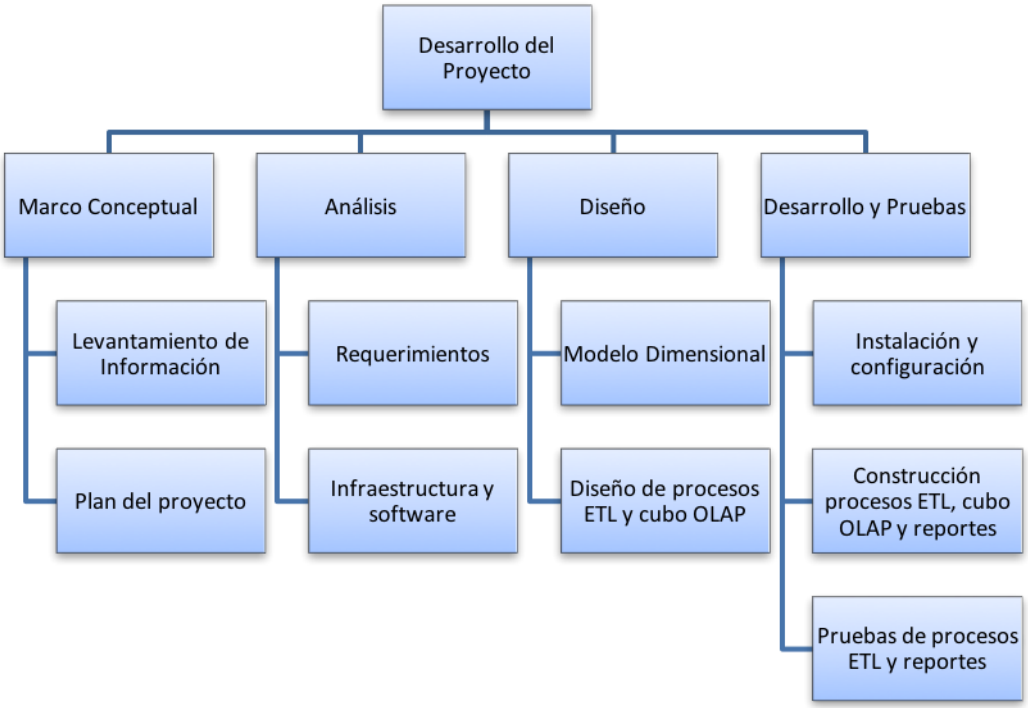


Figura 3. Desarrollo del proyecto utilizando la metodología de Kimball

5.1.2 Detalle y sustentación de la solución

En esta etapa se describe la solución a desarrollar para CEDIA (Fundación Consorcio Ecuatoriano para el desarrollo de Internet Avanzado). En base al levantamiento preliminar, obtenido en la fase de planificación del proyecto de acuerdo a la metodología de Kimball, se elabora el análisis de la situación actual, diseño y desarrollo del proyecto. La solución propuesta contempla el modelo de negocio, flujos continuos de datos (feeds), principales funcionalidades (análisis de eventos de seguridad informática), modelamiento de datos y necesidades de información.

5.1.3 Descripción de CEDIA

CEDIA (Fundación Consorcio Ecuatoriano para el desarrollo de Internet Avanzado) sus principales objetivos son: estimular, promover y coordinar el desarrollo de las TIC (Tecnologías de Información y Comunicación) y las redes de telecomunicaciones. Se enfoca en el desarrollo científico, tecnológico, innovador y educativo en el Ecuador. (CEDIA, 2015).

CEDIA promueve a los académicos e investigadores para que accedan a gran variedad de servicios orientados a impulsar y facilitar sus labores de enseñanza e investigación. Los servicios de CEDIA apoyan la conectividad, capacitaciones, infraestructura, repositorios, proyectos, colaboración, eventos, financiamiento y publicación de resultados (CEDIA, 2015).

5.1.4 Sustento de la solución

Para dar sustento a la solución existen dos principales enfoques al abordar una solución de Business Intelligence: Paradigma de Bill Inmon y paradigma de Ralph Kimball. La Tabla 3 muestra las diferencias entre los paradigmas:

Tabla 3

Comparación de Paradigmas de Desarrollo

	Inmon	Kimball
Desarrollo de un data warehouse	Orientado a temas / objetos.- Se organizan los registros para que todos los elementos y sus variaciones en el mundo real queden incorporados entre sí. Integración.- Debe contener todos los datos de la institución y la información debe ser consistente. No volátil.- No se debe modificar la información, debe mantenerse en modo de solo lectura, manteniéndose para consultas posteriores.	Una copia de bases de datos transaccionales específicamente estructurados para consulta y análisis.

Continua

	Variante en el tiempo.- Si se realiza algún cambio debe quedar registrado en un informe.	
Descripción de paradigmas	Orientado al Tema Fuertemente integrado Resistir las variantes del tiempo. Integración lograda a través de un modelo de datos empresariales asumido. Caracteriza los data marts como agregados	Orientado a procesos empresariales. Debe ser evolutivo. Acentúa el modelo dimensional. Integración lograda a través de dimensiones conformadas. Esquemas de estrella realzan la semántica de los Querys.
Desarrollo por Etapas	Es un software progresivo de las áreas temáticas, de acuerdo con las prioridades establecidas.	Se basan en procesos específicos del negocio y se vinculan a las dimensiones.

La metodología de Kimball propone una solución que da soporte a las decisiones. Su objetivo es diseñar y desarrollar bases de datos dimensionales que satisfagan las necesidades específicas de un área de una institución, lo que permite una mejor calidad de los datos y controlar con eficacia la información que se va a analizar.

El data warehouse de PYMES (pequeñas y medianas empresas) se acerca más a la idea de Kimball, debido a que la mayoría de proyectos de Business Intelligence comienzan con un esfuerzo dentro de un área específica en este caso el área del CSIRT que recibe flujos continuos de datos de varios sensores. Así se contempla en el futuro agregar más fuentes de datos al Data warehouse de eventos. Con lo

anteriormente descrito, se llega a la conclusión de seguir la metodología de Kimball, por ser la que abarca mejor las necesidades del CSIRT.

Se construye un Data Warehouse, que contiene información de cada escáner o fuente considerados data marts. La unión y homogenización de los orígenes de datos representan una solución estratégica de Business Intelligence dentro del CSIRT, por abarcar la principal actividad de esta institución.

El data warehouse de eventos contiene la información necesaria para el análisis de eventos e incidentes de seguridad registrados por el CSIRT, por fecha, puerto, institución, provincia, ciudad, hostname, tipos de eventos, sistema operativo, redes.

Además, por medio de reportes los usuarios podrán verificar los eventos registrados, el estado de eventos, los eventos que más se han demorado en solucionarse, nivel de seguridad de las instituciones, eventos más ocurridos, instituciones con mayor número de eventos registrados, sistemas operativos con la mayor cantidad de eventos, puerto que más eventos registra, entre otros.

5.2 Análisis

Se describe a continuación la metodología a usarse para desarrollar el proyecto, se detallan los requerimientos y necesidades de información por último se muestra el plan de pruebas que mide la eficacia del proyecto.

5.2.1 Metodología de Desarrollo

En la figura 4, se describe la metodología utilizada que tiene como base el ciclo de vida de data warehouse que fue realizado por Ralph Kimball.

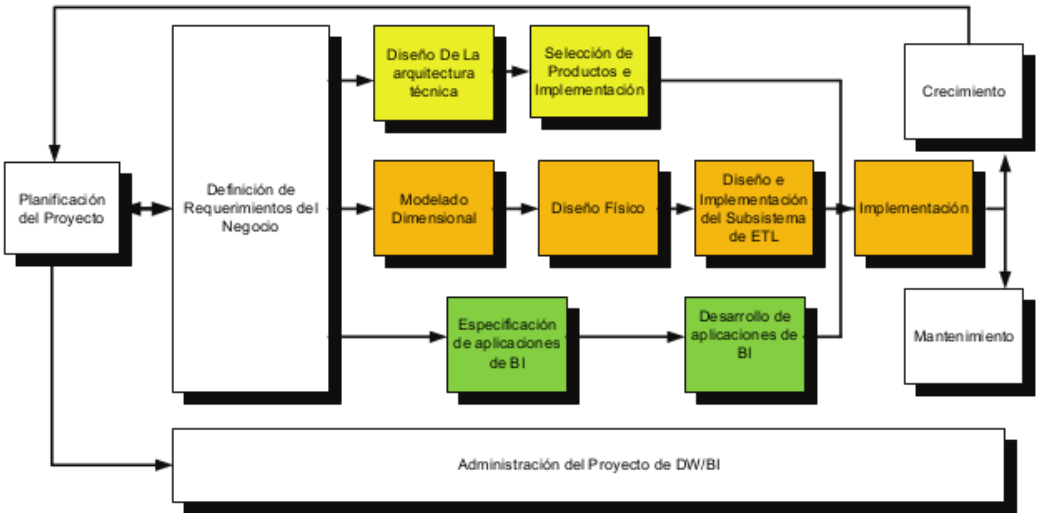


Figura 4. Gráfico de metodología de Ralph Kimball (Kimball, Mundy & Thornthwaite, 2008)

Como se ilustra en la Figura 4 se destaca la tarea de requerimientos, que son el soporte inicial y principal a las siguientes tareas, en donde la segunda tarea más importante es la de planificación. En la metodología de Kimball se puede observar tres rutas claramente definidas:

- Tecnología (Ruta Superior): Implica tareas con software y arquitectura.
- Datos (Ruta Central): En esta ruta se diseña e implementa el modelo dimensional y los procesos ETL.
- Aplicaciones de Inteligencia de Negocios (Ruta Inferior): Se visualizan las tareas de desarrollo las aplicaciones de inteligencia de negocios.

Todas las rutas convergen en un solo punto cuando se instala e implementa la plataforma de Business Intelligence.

5.2.1.1 Marco Conceptual de Metodología de Ralph Kimball

El marco conceptual describe la planificación del proyecto y las diferentes técnicas de recopilación de información que se utilizaron para levantar requerimientos. A continuación, se describe las fases de esta etapa.

Planificación

Es la primera etapa del desarrollo del proyecto, en el que se realiza el levantamiento de información, se identifica las necesidades, se determina el problema y se encuentra una solución. El desarrollo exitoso del proyecto se logra a partir de diversas reuniones y entrevistas con los especialistas de CEDIA y técnicos del CSIRT.

En esta etapa se determina las actividades a realizarse, cronogramas y recursos. Desde aquí se establece una relación con los miembros de CEDIA del área del CSIRT, quienes interactúan con el desarrollador.

5.2.1.2 Análisis y requerimientos

Esta fase describe la definición de requerimientos que según Kimball es la principal etapa para un desarrollo exitoso.

Definición de Requerimientos

La definición de requerimientos es la raíz del proyecto y es el cimiento para las etapas posteriores.

En esta etapa se realizan reuniones y entrevistas a los usuarios lo que permite entender los procesos de negocio, sus necesidades y requerimientos. Las herramientas para la definición de requerimientos son las entrevistas y reuniones llevadas periódicamente con miembros del CSIRT, quienes ayudan a descubrir los requerimientos, entender los factores clave que guían al negocio y determinar su alcance.

Se identifica la fuente de la información, su arquitectura y se verifica la calidad de los datos de origen.

5.2.1.3 Diseño

Esta fase contempla el diseño dimensional de la base de datos, procesos ETL y reportes que serán publicados en la plataforma de Business Intelligence. A continuación, se describe en breve detalle las etapas que comprende esta fase.

Diseño Dimensional

Como referente se ha utilizado el libro publicado por Kimball "The Data Warehouse Toolkit". Con la definición de requerimientos se determinan los orígenes y fuentes de datos necesarios para cumplir los requerimientos. Para diseñar el modelo de datos se especifica el nivel de detalle, que son los atributos de cada dimensión y el nivel de granularidad de las métricas, se analizan las dimensiones necesarias para cumplir con los requerimientos y dar forma al modelo dimensional del negocio.

Diseño Físico

Se selecciona las estructuras de almacenamiento necesarias, así como los métodos que garanticen un mejor acceso a los datos. Se decide utilizar como DBMS (Data Base Manage System) MySQL ya que es Open Source, no requiere de mayores recursos para su ejecución, es robusto y el esquema físico se adapta a él.

Diseño de ETL y Reportes

Se realizan procesos ETL desde orígenes de datos para dar formato, orden, coherencia a los datos de los flujos continuos de información provisto por los escáneres de CEDIA. Para que el usuario final los interprete fácilmente en los reportes y dashboards.

5.2.1.4 Desarrollo

El desarrollo contempla la construcción, instalación e implementación de la plataforma y el plan pruebas. A continuación, se presenta un breve detalle de lo que comprende esta fase.

Instalación de software

Se instala la base de datos MySQL, así como la plataforma de Business Intelligence de software libre Pentaho.

Diseño y construcción de procesos ETL

El principal objetivo de los procesos ETL y de un proyecto de Business Intelligence es la calidad de los datos; en esta etapa se debe asegurar su consistencia, facilidad de lectura y formato. Kimball describe lo siguiente "el proceso de Data Staging es el iceberg de un proyecto de data warehousing" (Kimball, 2008). En general esta es una de las etapas que más conlleva retrasos en un proyecto.

Los procesos ETL se dividen en tres sub procesos principales. (i) Extracción: Se obtienen los registros desde los flujos continuos de datos provisto por los escáneres de CEDIA. (ii) Transformación: Son procesos para convertir, formatear y dejar la información lo más limpia posible para la interpretación de los usuarios finales. (iii) Carga: Son procesos para poblar el data warehouse.

Diseño y construcción de Cubos OLAP

Para la construcción de cubos, se define la estructura de tabla de hechos, medidas, miembros calculados y dimensiones. Las dimensiones y sus jerarquías se definen dentro de cada cubo o de una forma general dentro del esquema. Esto evita tener que definir varias veces lo mismo, así como reutilizar elementos ya definidos; por tanto, antes de crearlos, se va establece las dimensiones compartidas con sus correspondientes jerarquías.

Construcción de reportes y dashboards

Para la construcción de reportes se utiliza Pentaho Report Designer con los requerimientos de datos especificados. Para la creación de dashboards se emplea la herramienta Community Chart Components instalada dentro de la plataforma de Pentaho.

Pruebas

Posterior a la construcción de procesos ETL, dashboards y reportes, se realiza un plan de pruebas que permite identificar errores generados en alguna etapa preliminar.

5.2.1.5 Implantación

Se realiza la implantación para utilizar el producto final dentro de las instalaciones del negocio. Existen algunos factores que afectan el correcto funcionamiento del producto final: la capacitación, soporte a usuarios, entendimiento del negocio y procesos.

Kimball plantea las siguientes tareas al momento de la implantación: configuración de hardware, conexión a las bases, acceso a intranet o internet, direcciones LAN (si no son dinámicamente asignadas), auditorias de tecnología sobre las configuraciones en las que se encuentran las PCs, proveer actualizaciones de hardware y software (determinando responsables, proyecto o área de usuario), verificaciones de seguridad, prueba de procedimientos de instalación, planificación de instalación, capacitación de usuarios, entre otros.

5.2.2 Requerimientos funcionales del CSIRT de CEDIA

La Tabla 4 describe los requerimientos funcionales que se levantaron junto con los especialistas de CEDIA en la gestión del CSIRT.

Tabla 4

Requerimientos Funcionales

Nº	Descripción	Prio	Ex
1	Reporte de eventos por institución.- contiene los datos más importantes de eventos registrados en cada institución. Este reporte contiene los siguientes parámetros: Institución, año, tipo de evento en donde se puede escoger uno o varios. Se presentará la información agrupada por tipo de evento.	1	E
2	Reporte de estado de eventos.- contiene eventos que están abiertos, se considera un evento abierto cuando ocurre hasta siete días antes de la fecha actual. El reporte contiene los siguientes parámetros: Institución, año, tipo de evento y estado. Se presenta la información agrupada por tipo de evento.	1	E
3	Reporte Eventos que más se han demorado en solucionarse.-	1	E

Continua

	presenta información acerca de los eventos que más se han demorado en solucionarse. Contiene el parámetro año donde se puede seleccionar uno o varios.		
4	Reporte Número de Eventos por año.- presenta el número de eventos que han sucedido en el año, se considera el mismo evento cuando ocurre un incidente consecutivo dentro de los siete días posteriores. Contiene los siguientes parámetros: Año, Institución y Tipo de Evento.	2	D
5	Dashboard Consolidado por Año.- presenta información referente a la cantidad de eventos por tipo y cantidad de eventos por institución al año. Se puede escoger el año a desplegar la información	1	E
6	Dashboard Comportamiento en el tiempo de Eventos.- contiene información acerca de la cantidad de eventos generados por tipo, su comportamiento a través del tiempo incluyendo el mes. Se puede escoger los siguientes parámetros: el tipo de evento y el año.	1	E
7	Dashboard Nivel de Seguridad de Instituciones.- presenta información acerca del nivel de seguridad de las instituciones y el porcentaje que representa cada mes para poder ser comparado. Se puede escoger el mes y el año a ser analizados.	2	D
8	Los usuarios tienen la opción de imprimir los reportes y dashboards si lo requieren.	1	E
9	Todos los reportes pueden ser exportados a otros archivos con formato: PDF	2	D
10	El administrador puede crear análisis, tablas cruzadas y gráficos, define los campos y filtros que muestra el reporte mediante Ad-Hoc y la utilización del cubo en la plataforma BI.	2	D

5.2.3 Requerimientos no funcionales

La Tabla 5 describe los requerimientos no funcionales que levantados con los especialistas de CEDIA en la gestión del CSIRT.

Tabla 5

Requerimientos no funcionales

Nº	Descripción	Prio	Ex
1	Debe ser desarrollado en software libre.	1	E
2	El sistema debe tener como motor de base de datos MySQL	3	D
3	El sistema cuenta con una interfaz gráfica en web para mostrar los reportes y dashboards	1	E
4	La plataforma es amigable con el usuario, de fácil manejo, permite al administrador generar sus propios análisis y gráficos.	1	E

Nomenclatura

Pri: Prioridad

Valor	Descripción
1	Alta
2	Media
3	Baja

Ex:

Valor	Descripción
E	Exigible
D	Deseable

5.2.6 Plan de Pruebas

Para el plan de pruebas se verifica la salida de los reportes creados usando caja negra o entrada/salida. Estas pruebas se centran en lo que se espera de un proceso y sus posibles resultados; se denomina también pruebas funcionales, en la cual el tester provee datos de entrada y estudia la salida, sin preocuparse del proceso que se realiza por detrás. Estas pruebas además se basan en la especificación de requerimientos y se da una medida del número de requisitos aprobados.

5.2.6.1 Plan de pruebas de reportes y dashboards

El siguiente plan de pruebas mide la eficacia de los dashboards y reportes. El mismo se detalla para cada reporte y dashboard.

Reporte de eventos por Institución

Este reporte, (Tabla 6) ayuda a los administradores de cada institución miembro a visualizar los tipos de eventos que ocurren, la fecha, ciudad, provincia. Para efecto de la prueba se ingresan como parámetros de inicio Institución: Escuela Politécnica del Ejército, Año: 2016, Tipo de Evento: botnet_drone. El resultado esperado es el siguiente:

Tabla 6

Resultado esperado para el reporte de eventos por institución

AÑO	MES	DÍA	HORA	EVENTO	INSTITUCIÓN	PROVINCIA	CIUDAD
2016	4	16	09:55:30	botnet_drone	Escuela Politécnica del Ejército	PICHINCHA	QUITO

Reporte de estado de eventos

El reporte de estado de eventos (Tabla 7), ayuda a visualizar eventos que se encuentran abiertos o cerrados. Se considera abierto un evento cuando ocurre dentro de los siete días anteriores a la fecha actual. Para la prueba se consideran los siguientes parámetros, Institución: Institución, Año: 2017, Tipo de Evento: botnet_drone, Estado: Abierto. El resultado esperado es el siguiente:

Tabla 7

Resultado esperado para el reporte estado de eventos

AÑO	MES	DÍA	HORA	EVENTO	INSTITUCIÓN	CIUDAD	ESTADO
2017	3	22	13:23:29	botnet_drone	Institución	RIOBAMBA	ABIERTO

Reporte Eventos que más se han demorado en solucionarse

El reporte de eventos que más se han demorado en solucionarse (Tabla 8), ayuda a visualizar una tendencia entre los eventos que no han tenido atención por

parte de los administradores de cada institución. Se ingresan como parámetros Año: 2015, 2016. El resultado esperado es el siguiente:

Tabla 8

Resultado esperado para el reporte eventos que más se han demorado en solucionarse

Año	Evento	Institución	Número de Eventos
2015	microsoft_sinkhole	Institución	22
2016	botnet_drone	Institución 2	19

Reporte Número de Eventos por Año

El reporte de número de eventos por Año descrito en la Tabla 9, ayuda a calcular la cantidad de eventos por año de cada institución, se considera un evento que se repite dentro de los siete días posteriores. Los parámetros de entrada son Año: 2016, Institución: Institución, Tipo de Evento: dns_openresolver. El resultado esperado es el siguiente:

Tabla 9

Resultado esperado número de eventos por año

Año	Evento	Institución	Número de Eventos
2016	dns_openresolver	Institución	1
El Evento estuvo activo: 1 mes y 15 días			

Consolidado por Año

El reporte consolidado por año (Tabla 10), ayuda a visualizar la cantidad de eventos generados en total en cada año y cuáles fueron las instituciones que tienen más cantidad de eventos registrados. A continuación, se verifica la mayor cantidad

de eventos registrados, para ello se ingresa como parámetro Año: 2017. El resultado esperado es el siguiente:

Tabla 10

Resultados reporte consolidado por año

Evento	Cantidad
bots	3178
scan_ssl_poddle	1773
bruteforce	440

Institución	Cantidad
Institución 1	3092
Institución 2	1236
Institución 3	583

Comportamiento en el tiempo de eventos

Este reporte ayuda a visualizar el comportamiento de eventos al pasar los años y como ha ido creciendo o decreciendo. Además, la cantidad de eventos por cada mes (Tabla 11). Como parámetros de entrada se selecciona Año: 2016 y Tipo de Evento: botnet_drone

Tabla 11

Resultados reporte comportamiento en el tiempo

Evento	Año				
botnet_drone	2013	2014	2015	2016	2017
	84	1959	1254	784	195
Meses **Año: 2016**	Enero	Febrero	Marzo		
	121	52	22		

Reporte Nivel de Seguridad

Este reporte ayuda a visualizar el nivel de seguridad de las instituciones, se ha dado valores entre 1 y 3 a los eventos para obtener el nivel de seguridad se suman los eventos y se obtiene un porcentaje de la mayor institución, a partir de ahí se obtienen

los porcentajes de las demás instituciones (Tabla 12). Como parámetros de entrada se selecciona. Año: 2016, Mes: 6. El resultado esperado es el siguiente:

Tabla 12

Resultado esperado para el reporte nivel de seguridad.

Institución	Porcentaje
Institución 1	100%
Institución 2	95,92%
Institución 3	62,59%

Nivel	Cantidad
1	331
2	375
3	48

5.2.7 Software a utilizar

Se describe el software que se va a utilizar durante el ciclo de vida de desarrollo del proyecto. Se comienza por el modelamiento de datos, elección del repositorio, procesos ETL (Extract, Transform and Load) y finalmente elaboración de dashboards y reportes. El detalle de la instalación y configuración se lo detalla en el Manual Técnico. Se describe en la siguiente sección el uso de las herramientas utilizadas.

5.2.7.1 Modelado de datos

El modelado de datos es un proceso utilizado para definir y analizar los requisitos de datos necesarios para soportar los procesos de negocio, dentro del alcance de los sistemas de información correspondientes en las organizaciones. Para realizar el diseño de la base de datos se ha utilizado Power Designer versión Nº 16.5.

PowerDesigner es una herramienta de modelado empresarial colaborativa producida por Sybase, actualmente propiedad de SAP. Se ejecuta bajo Microsoft Windows como una aplicación nativa y se ejecuta bajo Eclipse a través de un complemento. Admite el diseño de software de arquitectura basado en modelos. Almacena modelos que utilizan una variedad de extensiones de archivo, como .bpm, .cdm y .pdm. La estructura interna del archivo puede ser XML o un formato de

archivo binario comprimido. También puede guardar modelos en un repositorio de bases de datos. (SAP, 2016)

5.2.7.2 Gestor de Base de Datos

Para el desarrollo del proyecto se decidió utilizar un sistema de gestión de base de datos en software libre. Se eligió MySQL, ya que su diseño y arquitectura robusta es ideal para el número de registros y transacciones que se ejecutarán en el proyecto.

MySQL es una base de datos de código abierto de uso libre que facilita la gestión eficaz de las bases de datos conectándolas al software. Se trata de una solución estable, fiable y potente con características avanzadas como: Seguridad de los datos, escalabilidad, alto rendimiento, entre otros. (Oracle, 2016)

5.2.7.3 Plataforma de Business Intelligence

Pentaho está conformada por componentes OpenSource, es una plataforma completa unificada, simplifica la preparación y mezcla de cualquier información y proporciona un amplio espectro de análisis de negocio de autoservicio, incluyendo dashboards, informes, visualización de datos.

Fundamental para la estrategia de Pentaho es el concepto de entrega de datos gobernada, definida como la capacidad de combinar datos confiables y oportunos para hacer análisis a escala para todos los usuarios y entornos.

La plataforma de Pentaho es una arquitectura multi-tenant, en donde varios clientes consumen el servicio desde una misma plataforma (Figura 5), que facilita la incorporación de análisis en cualquier aplicación.

La plataforma de integración de datos Pentaho proporciona datos precisos y "analíticos listos" a los usuarios finales desde cualquier fuente de datos.

Independientemente del origen de datos, los requisitos de análisis o el entorno de despliegue, Pentaho permite convertir grandes datos en conocimiento (Pentaho, 2016).

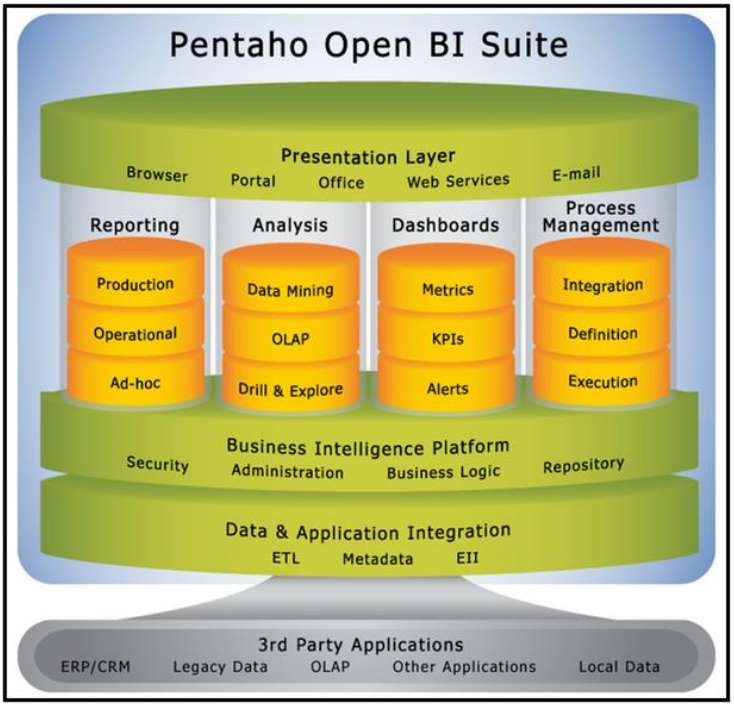


Figura 5. Arquitectura de Pentaho Community (Kimball, 2008)

5.2.7.4 Pentaho Data Integration

Pentaho Data Integration prepara y combina datos para crear una imagen completa del negocio que impulsa las ideas útiles. La plataforma proporciona datos precisos y listos para análisis a los usuarios finales desde cualquier fuente. Con herramientas visuales para eliminar la codificación y la complejidad, Pentaho pone grandes datos y todas las fuentes de datos al alcance de los usuarios de TI (Pentaho, 2016).

5.2.7.5 Saiku Business Analytics

Saiku permite a los usuarios de negocios explorar fuentes de datos complejas, utilizando una interfaz de arrastrar y soltar, una terminología comercial fácil de entender, todo dentro de un navegador. Se puede seleccionar los datos que le interesan, mirarlo desde diferentes perspectivas, profundizar en el detalle. Una vez que tenga su respuesta, guardar sus resultados, compartirlos, exportarlos a Excel o PDF (Meteorite, 2016).

5.2.7.6 Pentaho Report Designer

Pentaho Report Designer (PRD) es una herramienta de informes o reportes. Es de código abierto y contiene una fuente rica en características: posee una interfaz gráfica de usuario, es fácil de usar, una de sus funcionalidades es crear informes relacionales y analíticos de una amplia gama de fuentes de datos, se conecta con cualquier tipo de fuentes de datos, soporta sub-informes, gráficos y reportes basados en API que pueden ser incorporados en cualquier aplicación, el motor de informes Pentaho ejecuta el reporte diseñado por PRD (Edureka, 2014).

5.2.7.7 Pentaho Schema Workbench

Mondrian Schema Workbench es una interfaz de diseño que permite crear y probar esquemas de cubos Mondrian OLAP visualmente. El motor de Mondrian procesa solicitudes MDX con los esquemas ROLAP (Relational OLAP); los archivos de esquema son modelos de metadatos XML que se crean en una estructura específica utilizada por el motor Mondrian. Los modelos XML pueden considerarse estructuras de tipo cubo que utilizan las tablas de hechos y dimensiones existentes en el RDBMS (Pentaho, 2009).

5.3. Diseño

El diseño de la base de datos dimensional, incluye el diseño, arquitectura, procesos ETL.

5.3.1 Modelo de Datos Dimensional

Para el almacenamiento de datos existen dos modelos. (i) Modelo Estrella (Desnormalizado): Es más simple, utiliza una cantidad de datos moderada y optimiza el tiempo de respuesta en las consultas. (ii) Modelo Copo Nieve (Normalizado): Es más complejo, forma normalizada de las dimensiones, rompe el análisis dimensional, se utiliza con mayor cantidad de datos y las consultas se realizan en más tiempo.

Luego de realizar una comparación entre ambos modelos, se decide usar el modelo estrella con excepción de una tabla, dado que la solución está dirigida a una mediana empresa por el volumen de información. Con el modelo escogido se optimiza el tiempo de respuesta de los reportes.

5.3.1.1 Diagrama del Modelo Dimensional

El modelo dimensional está conformado por nueve dimensiones y una tabla de hechos (Figura 6), está basado en el esquema estrella o copo de nieve parcial. Se lo realizó en base a los requerimientos definidos por los especialistas del CSIRT, se diseña el modelo dimensional para abarcar todas las necesidades de datos que deben mostrarse en reportes y dashboards.

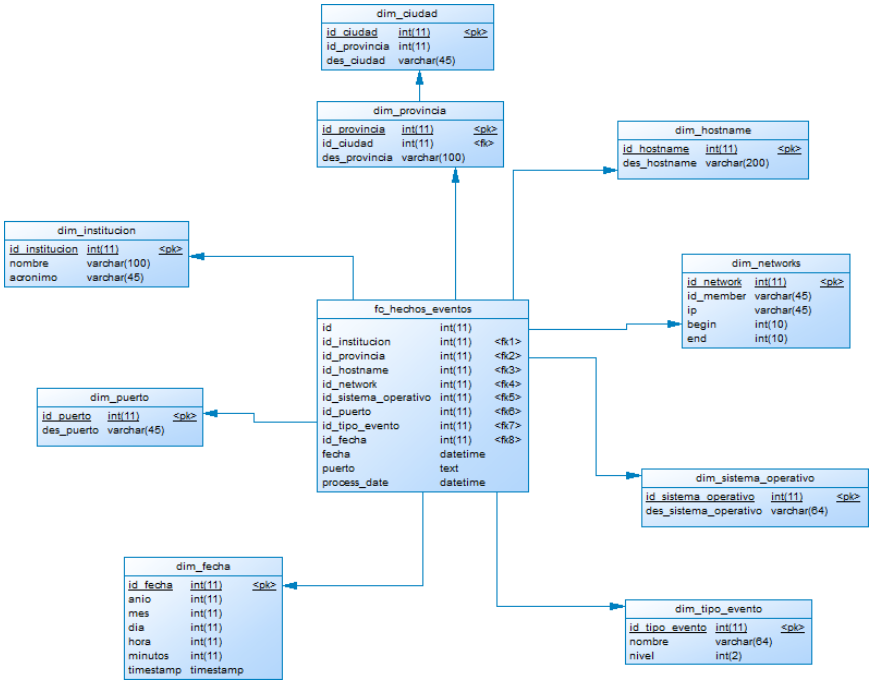


Figura 6 Diagrama del Modelo Dimensional

5.3.1.2 Dimensiones

Según Kimball (Kimball, 2008), una dimensión incluye los atributos que se van a analizar y se estructuran de forma jerárquica. El modelo dimensional está constituido por las siguientes dimensiones:

Dimensión Institución: Comprende la tabla de Institución y se refiere a las Universidades e instituciones pertenecientes a CEDIA, esta dimensión tiene los siguientes niveles de asociación de datos.

- Nombre
    - Acrónimo

Dimensión Fecha: Comprende la tabla de Fecha, mostrando los siguientes niveles jerárquicos.

- Minutos
    - Hora
        - Día
            - Mes
                - Año

Dimensión Tipo de Evento: Comprende la tabla de Tipo de Evento y se refiere a los eventos que tiene registrado y que analiza el CSIRT de CEDIA, mostrando los siguientes niveles jerárquicos.

- Nombre
    - Nivel de seguridad

Dimensión Sistema Operativo: Comprende la tabla Sistema Operativo y se refiere a los sistemas operativos descritos en los flujos de datos de algunos tipos de eventos.

Dimensión Puerto: La dimensión contiene los puertos en donde han ocurrido eventos de seguridad.

Dimensión Networks: Contiene las IPs y redes de cada institución.

Dimensión Hostname: Contiene los nombres de los hostname que más vulnerabilidades poseen.

Dimensión Provincia: Es la dimensión que almacena la provincia que llega de los flujos de datos de eventos ocurridos.

Dimensión Ciudad: Contiene las diferentes ciudades organizadas por provincias para su mejor despliegue.

Tabla de Hechos Eventos: Contiene las características de los eventos ocurridos en las instituciones, cuantificándolas por cantidad. En esta tabla se registra la fecha en que se procesó el evento. La información de eventos permite conocer los tipos de evento que más ocurren por provincia o por ciudad, la institución con mayor cantidad de eventos, los puertos y host más afectados.

5.3.1.3 Estándares del Modelo – Nomenclatura

Dimensiones: Las tablas que representan dimensiones tienen la siguiente nomenclatura.

DIM + NOMBRE DE DIMENSIÓN

Tabla de Hechos: La tabla que representa los hechos tiene la siguiente nomenclatura.

FC + NOMBRE DE TABLA DE HECHOS

Clave Primaria: Las claves primarias llevan la siguiente nomenclatura.

Id + nombre de la dimensión

5.3.2 Arquitectura

Los sistemas de Business Intelligence y toma de decisiones utilizan sistemas OLAP en su arquitectura, que permite analizar la información de diferentes ángulos y perspectivas, se caracteriza por ser un análisis multidimensional de datos, análisis de usuarios, selección de información. En la siguiente sección (Tabla 13), se realiza una comparación entre arquitectura ROLAP y MOLAP.

Tabla 13

Comparación arquitectura ROLAP y MOLAP

	ROLAP	MOLAP
VENTAJAS	Almacenamiento de grandes cantidades de datos	Buen rendimiento, rápida recuperación de información. Optimizado para operaciones de datos.

Continua

	Cubre funcionalidades inherentes a bases de datos relacionales.	Realiza cálculos complejos y rápidos
DESVENTAJAS	Bajo rendimiento, múltiples consultas de base de datos relacional, largo tiempo de respuesta. Limitado a SQL, no aporta todas las necesidades para consultas multidimensionales.	Se limita a la cantidad de datos que maneja. Maneja menores cantidades de datos. Se debe incluir solo información de alto nivel.

Se utiliza para el proyecto la arquitectura ROLAP, que soporta cálculos complejos, agregación y categorización de diferentes tipos de bases de datos. Se descarta utilizar MOLAP debido al limitante de datos que no soporta el volumen de datos que se va a manejar en el proyecto.

Se utiliza como herramienta Saiku Business Analytics ya que permite a los usuarios de negocios explorar fuentes de datos complejas, utilizando una interfaz de arrastrar y soltar familiar y una terminología fácil de entender, todo dentro de un navegador.

La arquitectura conceptual del presente proyecto comprende componentes como la extracción y limpieza de datos para la construcción de la base de datos dimensional de eventos. En la Figura 7 se describe la arquitectura a utilizar.

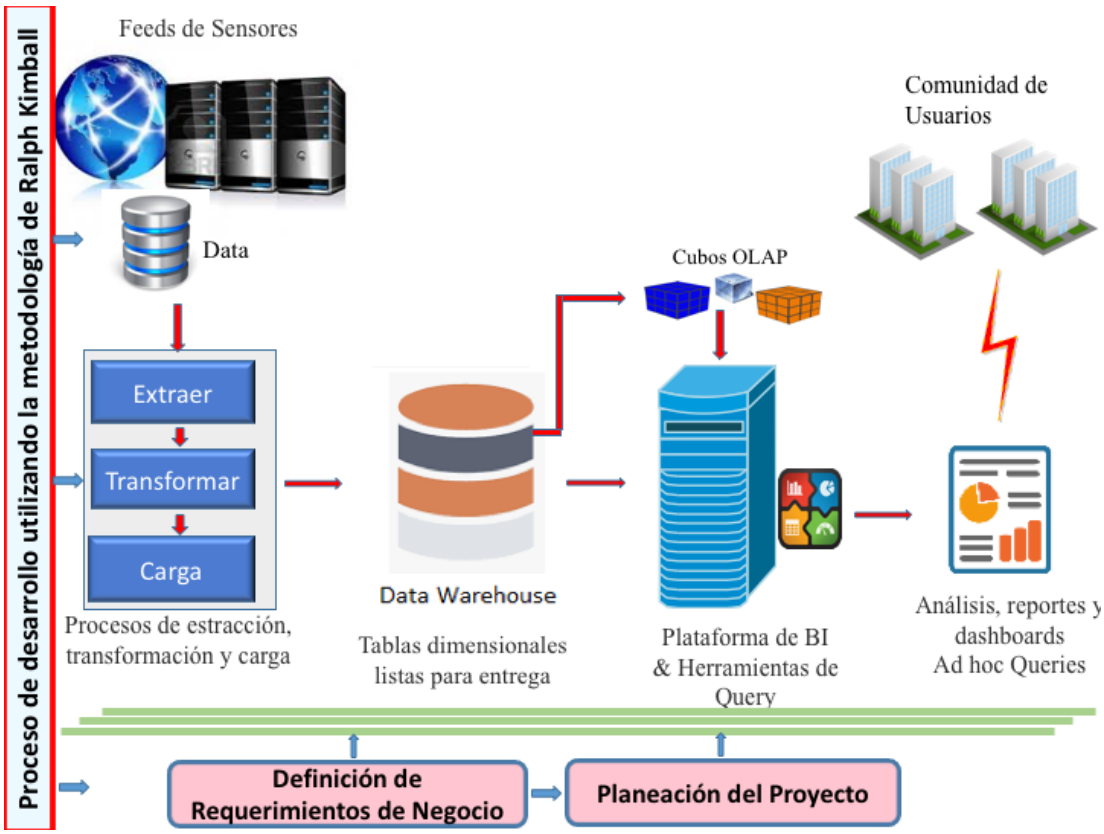


Figura 7. Arquitectura del proyecto

A continuación, se describen los componentes de la arquitectura:

Feeds de sensores desde los escáneres de CEDIA: Son las fuentes de orígenes de datos CEDIA recibe feeds(fuentes de datos) de escáneres tales como: Zone-h, TurkBot, Clean MX Foundation, CERT.br, Team Cymru, Nessus, Netcraft, Shadowserver. Estas fuentes son procesadas y almacenadas en bases de datos.

La Tabla 14 describe las fuentes de datos, de donde proviene la información de eventos del CSIRT de CEDIA.

Tabla 14

Fuente de Datos

Fuente de Datos	Tipo	Conexión
Base de Datos	Base de datos relacional	OLDB
Archivos planos	Hojas de cálculo	

ETL (Extract, Transform and Load): La construcción del proceso ETL es teóricamente una de las tareas más importantes en el desarrollo de una base de datos dimensional. Es complicado, consume mucho tiempo y utiliza la mayor parte de los esfuerzos de implementación del proyecto, costos y recursos (El-Sappagh, 2011). Los procesos de ETL son la base del DW. Aquí, el sistema extrae la información de los sistemas fuente, asegura la calidad y la consistencia de los datos, homogeneiza los datos de las alimentaciones divergentes para que puedan ser utilizados juntos (es decir, procesar y transformar la información si es necesario). Finalmente, genera los datos en el formato apropiado para que las herramientas de análisis puedan usarlo.

En consonancia con Kimball y Caserta, los sistemas ETL montan o "cargan" la base de datos dimensional. La construcción de tal sistema es una actividad que no es visible para los usuarios finales. Sin embargo, consume el 70% de los requerimientos de recursos para el desarrollo y mantenimiento de un sistema de data warehousing. Además, estos procesos no son simplemente una mera transferencia de información de un sistema u otro. Son mucho más, ya que son capaces de dar un valor significativo a los datos. Por lo tanto, procesos mal definidos o mal validados pueden afectar a un sistema de BI perfectamente diseñado.

Para el diseño de los procesos ETL, se ha utilizado algoritmos SQL para extraer los datos, que son almacenados en campos varchar (1024) y automatizados en la extracción de las dimensiones y hechos.

Base de Datos Dimensional: RDBMS en la que reside la estructura de datos dimensional, en el caso del proyecto se utiliza MySQL. Optimizado para acceso OLAP.

Cubos OLAP: Procesamiento de consultas MDX que retorna resultados multidimensionales. Se permite el uso de consultas SQL al RDBMS, maneja cachés y optimiza el rendimiento.

Plataforma de BI: Plataforma donde se centra la administración de archivos, reportes, dashboards, usuarios y roles.

Reportes y dashboards: Son realizados en diferentes componentes y subidos al BI para ser analizados por los usuarios.

Comunidad Usuarios: Usuarios finales que tendrán acceso al BI en el caso del proyecto, las Universidades miembro de CEDIA.

5.3.3 Diseño de Extracción de Datos

En esta etapa comprende la carga de la base de datos dimensional desde las diferentes fuentes de datos.

5.3.3.1 Carga Tabla de Hechos

En el flujo de trabajo del Job Tabla de hechos (Figura 8), primero se debe respaldar la información de la actual tabla de hechos para que no se pierdan registros por algún error posterior. El procedimiento es el siguiente: (i) Eventos Data Estructurado: Se extrae y se estructura los flujos continuos de datos de los escáneres, las fuentes son diferentes dependiendo del tipo de incidente, por lo que se debe extraer los datos relevantes y homogenizarlos para poderlos usar. (ii) Hechos Eventos: Se homogeniza y se acoplan todos los eventos en la tabla de hechos. (iii) Case Ciudad: Algunas fuentes no contemplan ciudad por lo que se realiza un proceso para incluir ciudad de acuerdo a la provincia a la que pertenecen. (iv) Create Hechos Eventos: Se almacena toda la información en la tabla final que formará parte del Data Warehouse.

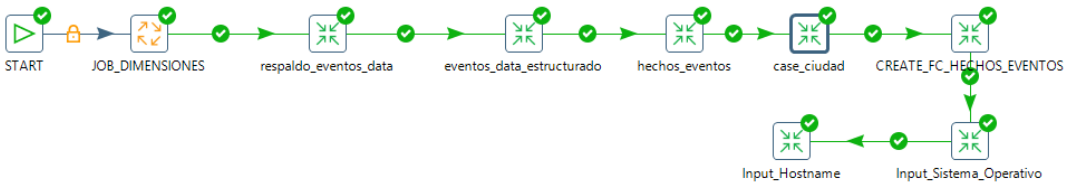


Figura 8. Job Carga de Tabla de Hechos

Para cargar la tabla de hechos se realizan varias transformaciones, la primera de ellas es cargar las dimensiones. El proceso de carga se muestra a continuación (Figura 9) y se describe en la siguiente etapa.

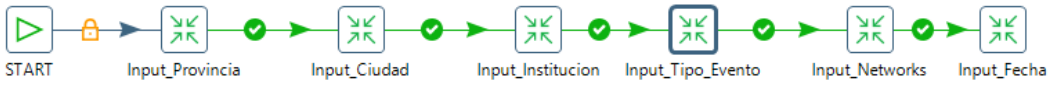


Figura 9. Job de carga de dimensiones

5.3.3.2 Carga de Dimensiones

Carga de Provincia

La dimensión provincia contiene atributos de las diferentes provincias del país. En la Tabla 15 se describe la fuente de la dimensión Provincia.

Descripción de Tablas Fuentes

Tabla 15

Fuente Dimensión Provincia

Tipo de Fuente	Nombre de la Fuente	Descripción
Hoja de Cálculo	Dim_Provincia	Este archivo plano contiene el listado de provincias para su fácil modificación

Estandarización de Datos y Limpieza de Datos Dimensión Provincia

Se especifica la nomenclatura y transformación que deben tener los datos antes de guardarlos en la base de datos dimensional, (Tabla 16).

Tabla 16

Limpieza de Datos Dimensión Provincia

Nombre	Key	Tipo	Formato	Limpieza	Valor por Defecto
Id_provincia	PK	Integer	Número Entero	No Nulo	1
Id_ciudad	FK	Integer	Número Entero	No Nulo	1
Des_pronvincia		Varchar(100)	Texto		

Carga de Ciudad

La dimensión ciudad contiene atributos de las diferentes ciudades del país (Tabla 17).

Descripción de Tablas Fuentes

Tabla 17

Fuente Dimensión Ciudad

Tipo de Fuente	Nombre de la Fuente	Descripción
Hoja de Cálculo	Dim_Ciudad	Este archivo plano contiene el listado de ciudades para su fácil modificación

Estandarización de Datos y Limpieza de Datos Dimensión Ciudad

Se especifica la nomenclatura y transformación que deben tener los datos antes de guardarlos en la base de datos dimensional, (Tabla 18).

Tabla 18

Limpieza de Datos Dimensión Ciudad

Nombre	Key	Tipo	Formato	Limpieza	Valor por Defecto
Id_ciudad	FK	Integer	Número Entero	No Nulo	1
Des_ciudad		Varchar(100)	Texto		

Carga de Institución

La dimensión institución contiene atributos de las instituciones miembro de CEDIA.

Descripción de Tablas Fuentes

La tabla fuente es de donde se extrae los datos para alimentar a la dimensión Institución (Tabla 19).

Tabla 19

Fuente Dimensión Institución

Tipo de Fuente	Nombre de la Fuente	Descripción
Base de Datos	Sys_members	Esta tabla contiene el listado de los miembros de CEDIA.

Estandarización de Datos y Limpieza de Datos dimensión institución

Se especifica la nomenclatura y transformación que deben tener los datos antes de guardarlos en la base de datos dimensional, (Tabla 20).

Tabla 20

Limpieza de Datos Dimensión Institución

Nombre	Key	Tipo	Formato	Limpieza	Valor por Defecto
Id_institucion	FK	Integer	Número Entero	No Nulo	1
nombre		Varchar(100)	Texto		
acronimo		Varchar(100)	Texto		

Carga de Tipo de Evento

La dimensión tipo de evento contiene atributos de los eventos registrados por CEDIA.

Descripción de Tablas Fuente

La tabla fuente es de donde se extrae los datos para alimentar a la dimensión tipo de evento (Tabla 21).

Tabla 21

Fuente Dimensión Tipo de Evento

Tipo de Fuente	Nombre de la Fuente	Descripción
Base de Datos	Alrt_categories	Esta tabla contiene el tipo de eventos que registra CEDIA.

Estandarización de Datos y Limpieza de Datos Dimensión Tipo Evento

Se especifica la nomenclatura y transformación que deben tener los datos antes de guardarlos en la base de datos dimensional, (Tabla 22).

Tabla 22

Limpieza de Datos Dimensión Tipo de Evento

Nombre	Key	Tipo	Formato	Limpieza	Valor por Defecto
Id_tipo_evento	PK	Integer	Número Entero	No Nulo	1
nombre		Varchar(100)	Texto		
nivel		int(2)	Número Entero		1,2,3

Carga de Networks

La dimensión networks contiene los atributos de las redes de CEDIA, descrito en la Tabla 23.

Descripción de Tablas Fuente

La tabla fuente es de donde se extrae los datos para alimentar a la dimensión networks (Tabla 23).

Tabla 23

Fuente Dimensión Networks

Tipo de Fuente	Nombre de la Fuente	Descripción
Base de Datos	Sys_networks	Esta tabla contiene el listado de redes de CEDIA.

Estandarización de Datos y Limpieza de Datos dimensión Networks

Se especifica la nomenclatura y transformación que deben tener los datos antes de guardarlos en la base de datos dimensional, (Tabla 24).

Tabla 24

Limpieza de Datos Dimensión Networks

Nombre	Key	Tipo	Formato	Limpieza	Valor por Defecto
Id_network	PK	Integer	Número Entero	No Nulo	1
Id_member		Varchar(45)	Texto		
ip		Varchar(45)	Texto		
begin		Int(10)	Número Entero		
end		Int(10)	Número Entero		

Carga de Fecha

La dimensión fecha contiene todos los atributos y niveles jerárquicos de fechas descrito en la Tabla 25.

Descripción de Tablas Fuente

Tabla 25

Fuente Dimensión Fecha

Tipo de Fuente	Nombre de la Fuente	Descripción
Base de Datos	Alrt_alerts	Esta tabla contiene las alertas registradas.

Estandarización de Datos y Limpieza de Datos dimensión fecha

Se especifica la nomenclatura y transformación que deben tener los datos antes de guardarlos en la base de datos dimensional, (Tabla 26).

Tabla 26

Limpieza de Datos Dimensión Fecha

Nombre	Key	Tipo	Formato	Limpieza	Valor por Defecto
Id_fecha	PK	Integer	Número Entero	No Nulo	1
Anio		int(11)	Número Entero		
Mes		int(11)	Número Entero		
Día		int(11)	Número Entero		
Hora		int(11)	Número Entero		
Minutos		int(11)	Número Entero		

Carga de Hostname

La dimensión hostname contiene todos los nombres de hostname de las instituciones que han registrado eventos descrito en la Tabla 27.

Descripción de Tablas Fuente

Tabla 27

Fuente Dimensión Hostname

Tipo de Fuente	Nombre de la Fuente	Descripción
Base de Datos	Alrt_alerts	Esta tabla contiene las alertas registradas.

Estandarización de Datos y Limpieza de Datos Dimensión Hostname

Se especifica la nomenclatura y transformación que deben tener los datos antes de guardarlos en la base de datos dimensional, (Tabla 28).

Tabla 28

Limpieza de Datos Dimensión Hostname

Nombre	Key	Tipo	Formato	Limpieza	Valor por Defecto
Id_hostname	PK	Integer	Número Entero	No Nulo	1
Des_hostname		Varchar(200)	Texto		

Carga de Sistema Operativo

La dimensión sistema operativo contiene todos los sistemas operativos extraídos de los eventos registrados (Tabla 29).

Descripción de Tablas Fuente

Tabla 29

Fuente Dimensión Sistema Operativo

Tipo de Fuente	Nombre de la Fuente	Descripción
Base de Datos	Alrt_alerts	Esta tabla contiene las alertas registradas.

Estandarización de Datos y Limpieza de Datos Dimensión Sistema Operativo

Se especifica la nomenclatura y transformación que deben tener los datos antes de guardarlos en la base de datos dimensional, (Tabla 30).

Tabla 30

Limpieza de Datos Dimensión Sistema Operativo

Nombre	Key	Tipo	Formato	Limpieza	Valor por Defecto
Id_sistema_operativo	PK	Integer	Número Entero	No Nulo	1
Des_sistema_operativo		Varchar(200)	Texto		

5.3.3.3 Esquema de Extracción

El esquema descrito en la Tabla 31 permite conocer el orden en que se realiza la carga de dimensiones y tabla de hechos.

Tabla 31

Esquema de Extracción

Nº	Proceso	Dependencia
1	Job_Dimensiones	
2	Carga Provincia	1
3	Carga Ciudad	1, 2
4	Carga Institución	1
5	Carga Tipo Evento	1
6	Carga Networks	1
7	Carga Fecha	1
8	Respaldo de Hechos	
9	Estructuración de tabla de hechos	
10	Carga hechos eventos	1,2,3,4,5,6,7

5.4 Desarrollo y pruebas

Esta fase describe la construcción y prueba de la base de datos dimensional, que incluye la instalación y configuración de todo el software, procesos ETL, reportes y dashboards.

5.4.1 Desarrollo

A continuación, se describe el proceso de instalación de la plataforma de BI, utilizando los procesos ETL diseñados previamente y la construcción de reportes, dashboards y consultas Ad-Hoc.

5.4.1.1 Configuración e instalación del software

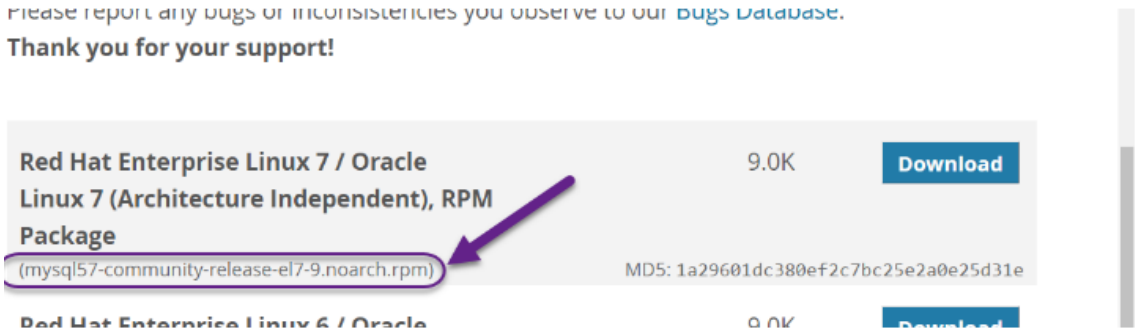
A continuación, se describe las instrucciones a tomar en cuenta en la configuración de la plataforma de BI.

Instalación y configuración de MySQL

Para instalar MySQL en el servidor se debe primero ingresar a un navegador web e ingresar la siguiente dirección:

https://dev.mysql.com/downloads/repo/yum/

Se debe localizar la versión de MySQL que se va a utilizar.



Luego utilizar los siguientes comandos:

```
$ wget https://dev.mysql.com/get/mysql57-community-release-el7-9.noarch.rpm
```

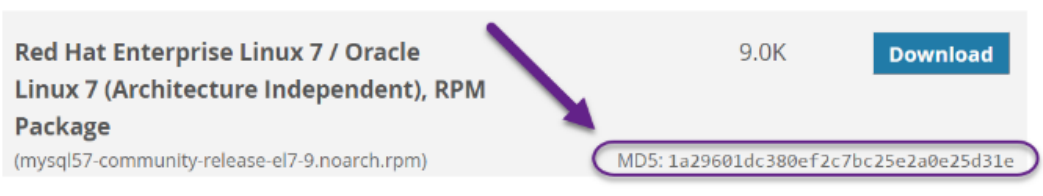
Una vez que el archivo .rpm está guardado, se debe verificar la integridad de la descarga corriendo md5sum y comparándolo con el valor MD5 que está en el sitio.

```
$ md5sum mysql57-community-release-el7-9.noarch.rpm
```

Este comando da como salida lo siguiente:

```
1a29601dc380ef2c7bc25e2a0e25d31e  mysql57-community-release-el7-
9.noarch.rpm
```

Comparar esta salida con el valor MD5 del sitio:



Ahora que ya está verificado que el archivo no está corrupto o dañado, se instala el paquete.

```
$ sudo rpm -ivh mysql57-community-release-el7-9.noarch.rpm
```

Esto agrega dos nuevos repositorios YUM, ahora se puede instalar MySQL server.

```
$ sudo yum install mysql-server
```

Presionar (Y) para confirmar

Iniciar MySQL

Iniciar el demonio con el siguiente comando.

```
$ sudo systemctl start mysqld
```

Para asegurar que se inició, se ejecuta el siguiente comando.

```
$ sudo systemctl status mysqld
```

Si MySQL se ejecutó correctamente debe salir lo siguiente:

```
$ Dec 01 19:02:20 centos-512mb-sfo2-02 systemd[1]: Started MySQL
Server.
```

Durante el proceso de instalación, una clave temporal es generada para usuario root.

Configurar MySQL

Para asegurar la seguridad de MySQL se debe configurar con los siguientes comandos:

```
$ sudo mysql_secure_installation
```

Con este comando se cambiará la clave de root la salida es la siguiente:

```
The existing password for the user account root has expired.
Please set a new password.

New password:
```

Se ingresa una clave de 12 caracteres que contenga al menos una mayúscula, una minúscula, un número y un carácter especial.

La salida es la siguiente:

```
Estimated strength of the password: 100
Change the password for root ? (Press y|Y for Yes, any other key
for No) :
```

PROBANDO MySQL

Para verificar la instalación se utiliza el siguiente comando:

```
$ mysqladmin -u root -p version
```

La salida será la siguiente:

```
mysqladmin  Ver 8.42 Distrib 5.7.16, for Linux on x86_64
Copyright (c) 2000, 2016, Oracle and/or its affiliates. All
rights reserved.

Oracle is a registered trademark of Oracle Corporation and/or its
affiliates. Other names may be trademarks of their respective
owners.

Server version          5.7.16
Protocol version        10
Connection              Localhost via UNIX socket
UNIX socket             /var/lib/mysql/mysql.sock
Uptime:                 2 min 17 sec

Threads: 1  Questions: 6  Slow queries: 0  Opens: 107  Flush
tables: 1  Open tables: 100  Queries per second avg: 0.043
```

Lo que indica que se ha instalado correctamente.

Instalación de Java

Descargar la última versión de Java

Para 64Bit utilizar el siguiente comando:

```
# cd /opt/

# wget --no-cookies --no-check-certificate --header "Cookie:
gpw_e24=http%3A%2F%2Fwww.oracle.com%2F; oraclelicense=accept-
securebackup-cookie" "http://download.oracle.com/otn-
pub/java/jdk/8u121-b13/e9e7ea248e2c4826b92b3f075a80e441/jdk-
8u121-linux-x64.tar.gz"



# tar xzf jdk-8u121-linux-x64.tar.gz
```

Instalar Java con sus alternativas

Después de extraer el archivo, se utiliza el comando alternatives para instalarlo

```
# cd /opt/jdk1.8.0_121/

# alternatives --install /usr/bin/java java
/opt/jdk1.8.0_121/bin/java 2

# alternatives --config java




There are 3 programs which provide 'java'.


  Selection    Command

-----------------------------------------------

*  1           /opt/jdk1.7.0_71/bin/java

 + 2           /opt/jdk1.8.0_45/bin/java

   3           /opt/jdk1.8.0_91/bin/java

   4           /opt/jdk1.8.0_121/bin/java


Enter to keep the current selection[+], or type selection
number: 4
```

En este punto Java se ha instalado correctamente en el sistema. Se configura lo siguiente.

```
# alternatives --install /usr/bin/jar jar
/opt/jdk1.8.0_121/bin/jar 2

# alternatives --install /usr/bin/javac javac
/opt/jdk1.8.0_121/bin/javac 2

# alternatives --set jar /opt/jdk1.8.0_121/bin/jar

# alternatives --set javac /opt/jdk1.8.0_121/bin/javac
```

Comprobar la versión de Java Instalada

```
root@tecadmin ~# java -version

java version "1.8.0_121"

Java(TM) SE Runtime Environment (build 1.8.0_121-b13)

Java HotSpot(TM) 64-Bit Server VM (build 25.121-b13, mixed
mode, sharing)
```

Configurar Variables de Entorno

Utilizar los siguientes comandos:

- Setup **JAVA_HOME** Variable

```
# export JAVA_HOME=/opt/jdk1.8.0_121
```

- Setup **JRE_HOME** Variable

```
# export JRE_HOME=/opt/jdk1.8.0_121/jre
```

Instalación de la plataforma Pentaho Business Intelligence

PERMISOS Y USUARIOS

Es recomendable utilizar un usuario para ejecutar Pentaho. Llegados a este caso se debe crear un usuario pentaho y traspasarle la propiedad del BI.

```
adduser pentaho
chown -R pentaho:pentaho /opt/pentaho/biserver-ce
```

Se debe modificar el archivo start-pentaho.sh el cual esta descrito en el Manual Técnico.

PROBANDO LA INSTALACIÓN

Se ejecuta el script y se verifica que todo funciona. Utilizando el usuario pentaho que se ha creado para esta función.

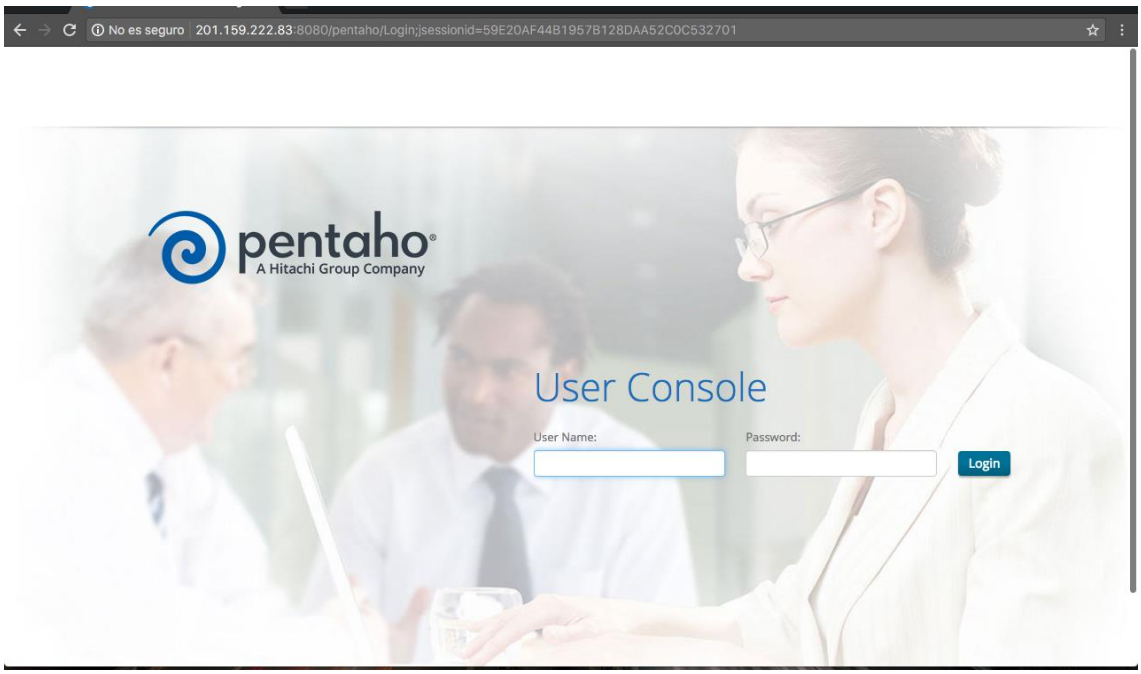
```
sudo su pentaho
cd /opt/pentaho/biserver-ce
./start-pentaho.sh
```

Se puede comprobar ingresando en un navegador a la siguiente dirección localhost:8080/pentaho

Centos y firewalld

Se debe abrir el puerto 8080 para que desde otra máquina puedan acceder al portal.

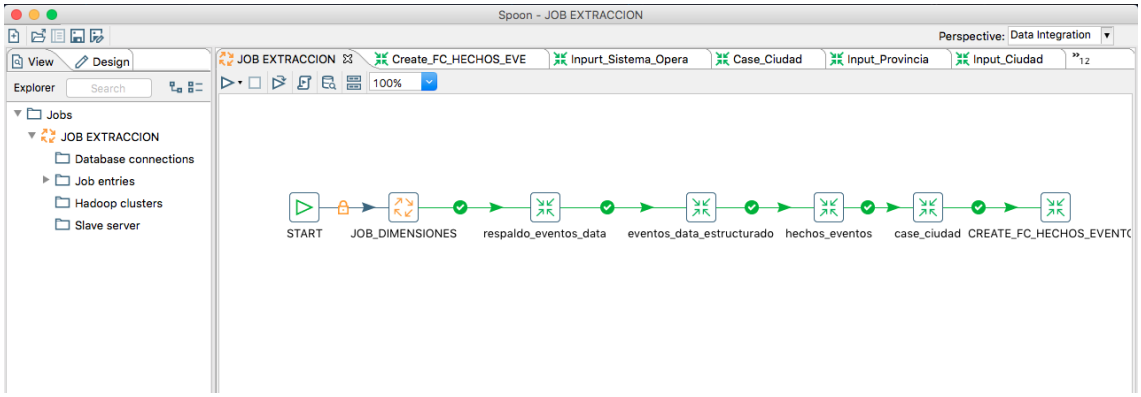
```
firewall-cmd --zone=public --add-port=8080/tcp --permanent
firewall-cmd --reload
```



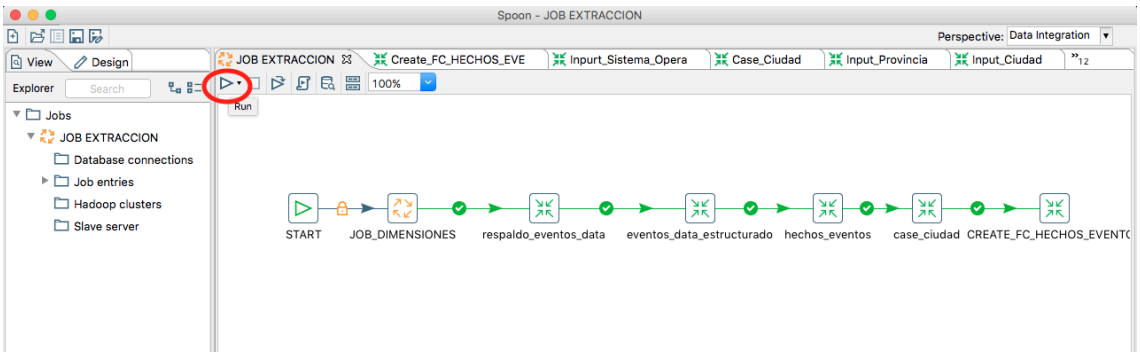
Instalación de Pentaho Data Integration

El instalador se puede utilizar en todos los sistemas operativos.

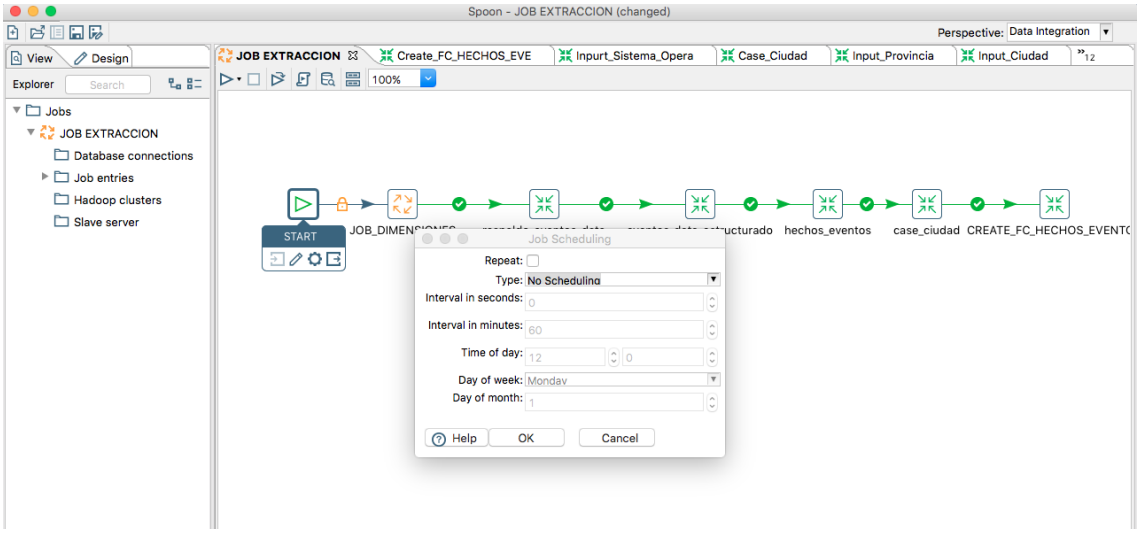
Al abrir Data Integration se tiene lo siguiente.



Para comenzar a llenar el Data Warehouse se debe hacer click en el siguiente botón:



Para calendarizar el Job que se ejecute diariamente se debe realizar lo siguiente:



Doble Click en START y llenar la información de la ventana dependiendo de cada cuanto se desea ejecutar los procesos de llenado del Data Warehouse.

Instalación de Pentaho Report Designer

Pentaho Report Designer(PRD) es una herramienta de reporting fácil de utilizar y con multitud de aplicaciones. Los informes que genera se dividen en secciones o grupos de datos en los que los elementos del informe pueden ser posicionados.

Para descargar se debe acceder a: http://community.pentaho.com/, (Figura 10)

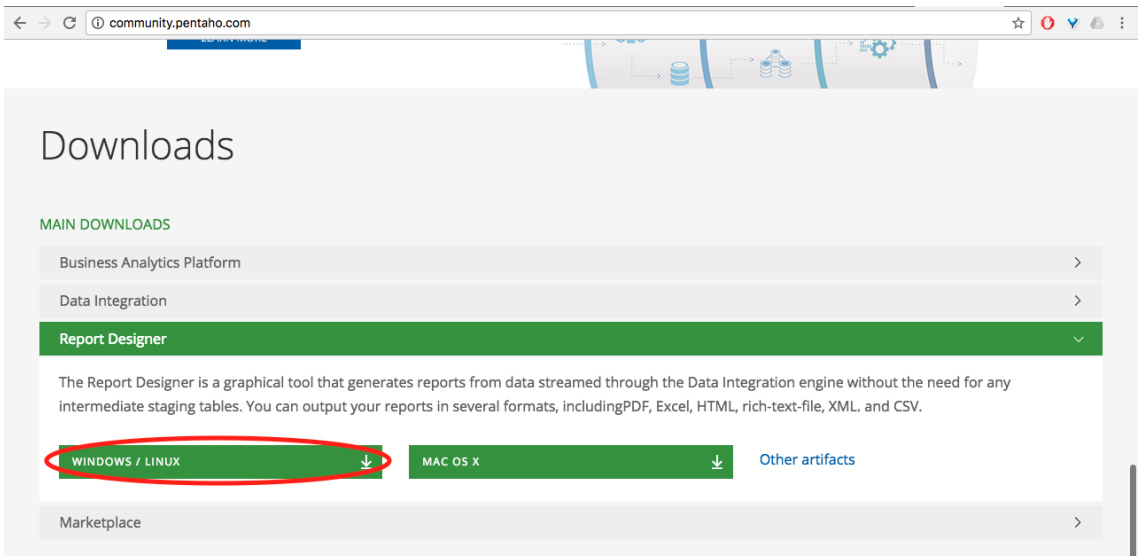


Figura 10. Descarga de Pentaho Report Designer

Para publicar un reporte en Pentaho BI se debe realizar lo siguiente:

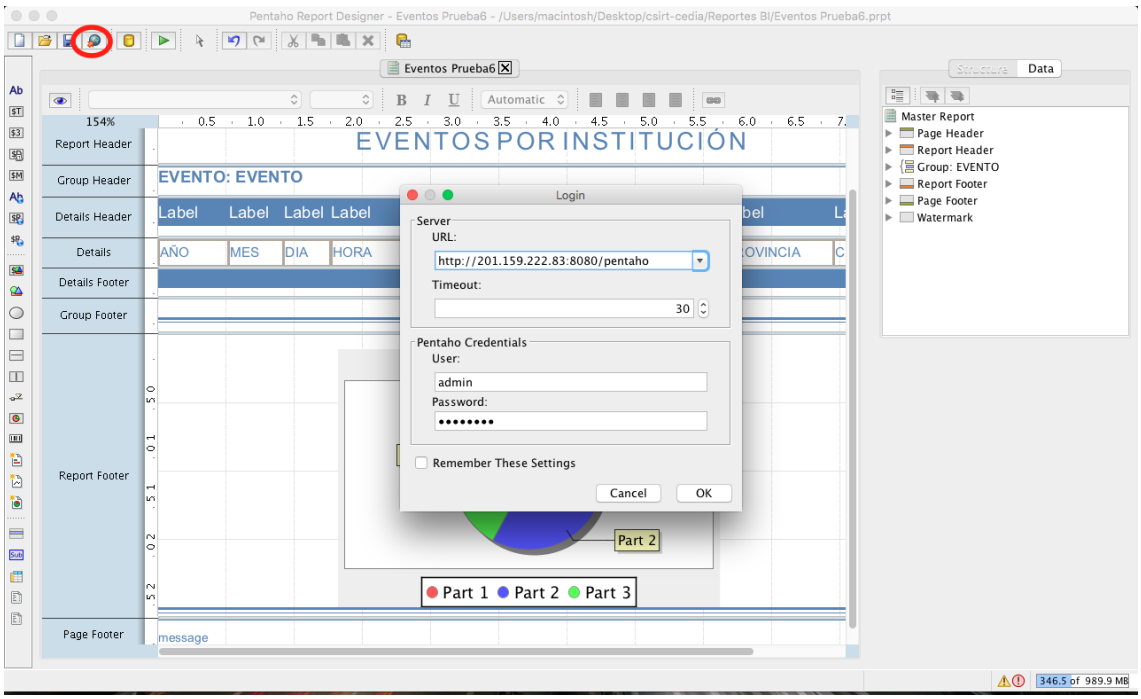


Figura 11. Publicación de Reportes

Click en Publish (Figura 11) y luego llenar los datos de la ventana.

Creación y publicación de cubo OLAP

Se puede instalar en cualquier sistema operativo. Al abrir Schema Workbench se puede ver el cubo realizado, (Figura 12).

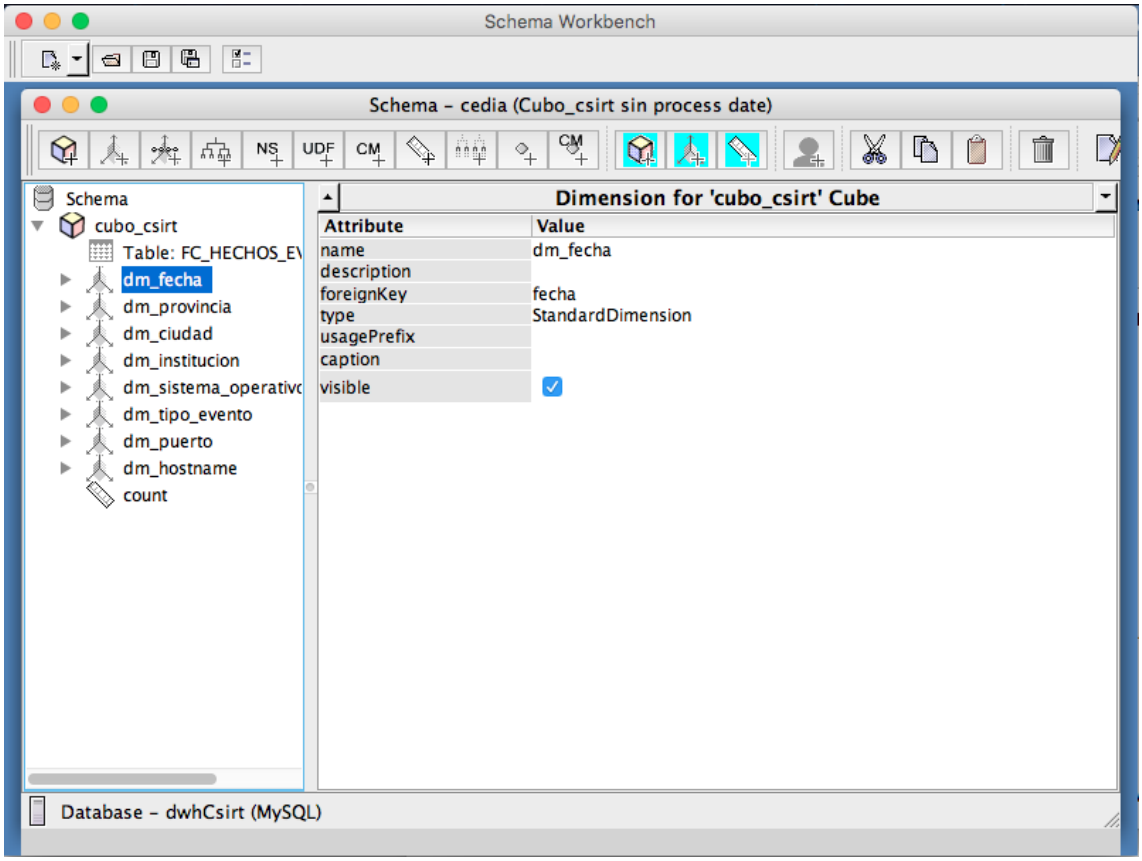
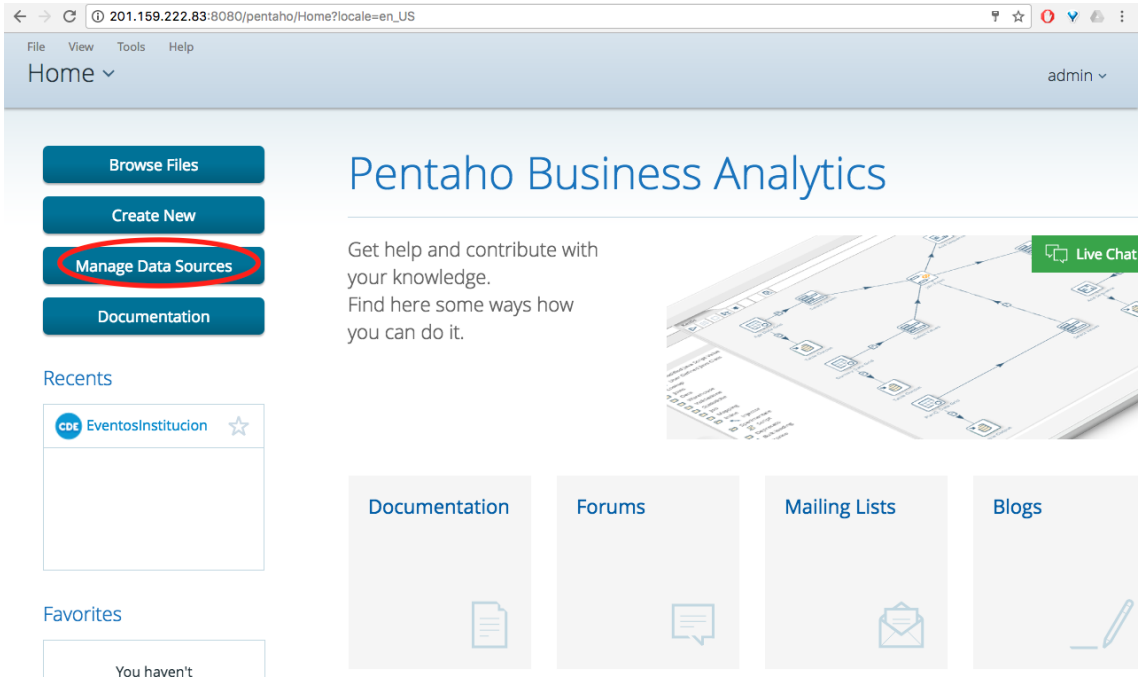
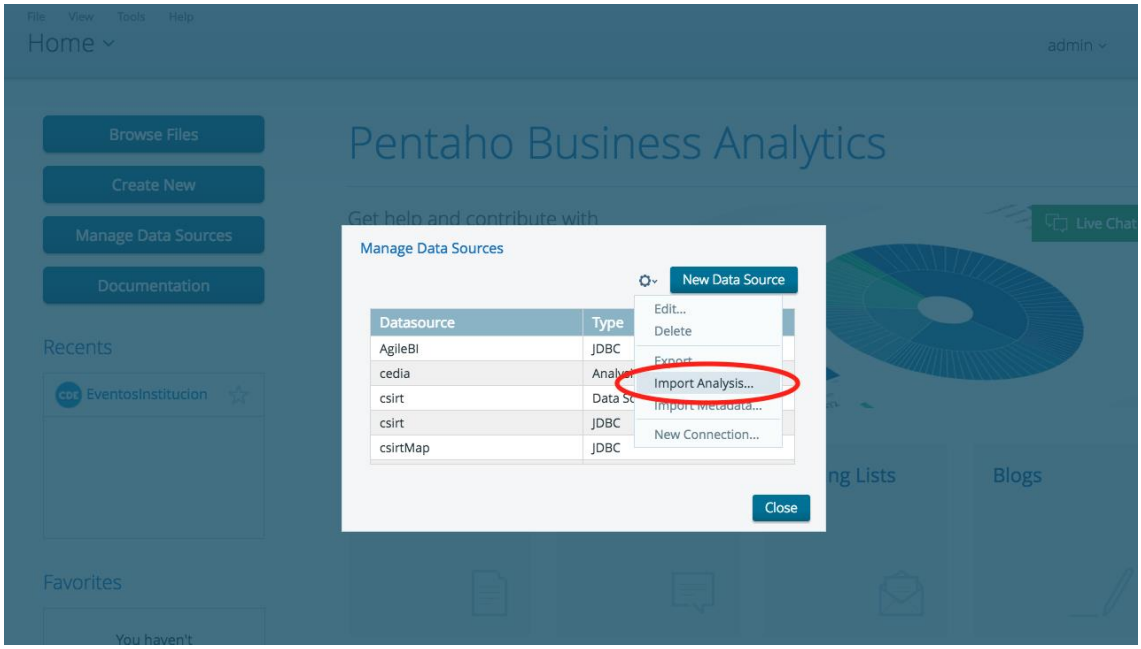


Figura 12. Cubo OLAP en Schema Workbench

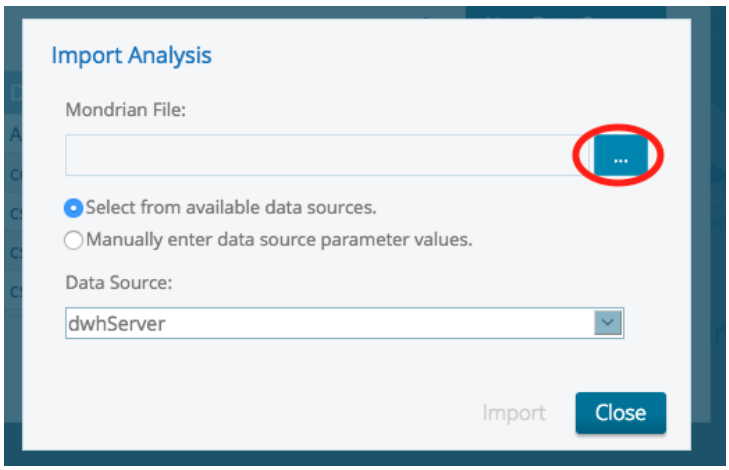
Esta aplicación genera un archivo XML con la definición del cubo. Para publicar el Cubo OLAP en Pentaho BI se debe realizar lo siguiente: Se debe entrar a Pentaho BI como administrador.



Dar click en Manage Data Sources.



En la ventana que aparece hacer click en Configuración y luego Import Analysis.



En la nueva ventana dar click en el botón señalado y escoger el archivo del cubo .xml, Escoger el Data Source donde se encuentre la base de datos.

5.4.2 Pruebas

Los tipos de pruebas aplicadas se describen a continuación:

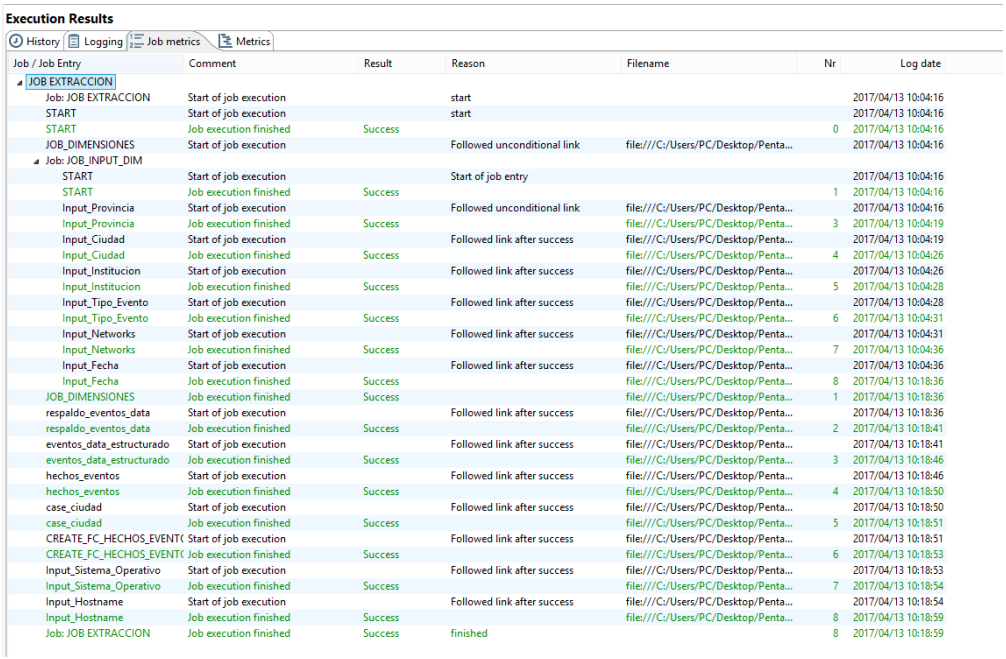
5.4.2.1 Pruebas de Procesos ETL

Se realizaron las siguientes pruebas en los procesos ETL, descritos en la Tabla 32.

Tabla 32. Pruebas Procesos ETL

Prueba	Descripción
Unitarias	Para cada proceso de extracción, transformación y carga se aplicó pruebas de caja blanca para comprobar el funcionamiento.
Integración	Se verificó la integración e interacción de todos los procesos para completar la carga de la base de datos dimensional.
Funcional	Se comprobó que los procesos ETL cumplan con las funcionalidades descritas en los requerimientos.

A continuación, se ejecutan los procesos ETL y se verifica su salida sin errores (Figura 13).



Job / Job Entry	Comment	Result	Reason	Filename	Nr	Log date
JOB EXTRACCION						
Job: JOB EXTRACCION	Start of job execution		start			2017/04/13 10:04:16
START	Start of job execution		start			2017/04/13 10:04:16
START	Job execution finished	Success			0	2017/04/13 10:04:16
JOB_DIMENSIONES	Start of job execution		Followed unconditional link	file:///C:/Users/PC/Desktop/Penta...		2017/04/13 10:04:16
Job: JOB_INPUT_DIM						
START	Start of job execution		Start of job entry			2017/04/13 10:04:16
START	Job execution finished	Success			1	2017/04/13 10:04:16
Input_Provincia	Start of job execution		Followed unconditional link	file:///C:/Users/PC/Desktop/Penta...		2017/04/13 10:04:16
Input_Provincia	Job execution finished	Success		file:///C:/Users/PC/Desktop/Penta...	3	2017/04/13 10:04:19
Input_Ciudad	Start of job execution		Followed link after success	file:///C:/Users/PC/Desktop/Penta...		2017/04/13 10:04:19
Input_Ciudad	Job execution finished	Success		file:///C:/Users/PC/Desktop/Penta...	4	2017/04/13 10:04:26
Input_Institucion	Start of job execution		Followed link after success	file:///C:/Users/PC/Desktop/Penta...		2017/04/13 10:04:26
Input_Institucion	Job execution finished	Success		file:///C:/Users/PC/Desktop/Penta...	5	2017/04/13 10:04:28
Input_Tipo_Evento	Start of job execution		Followed link after success	file:///C:/Users/PC/Desktop/Penta...		2017/04/13 10:04:28
Input_Tipo_Evento	Job execution finished	Success		file:///C:/Users/PC/Desktop/Penta...	6	2017/04/13 10:04:31
Input_Networks	Start of job execution		Followed link after success	file:///C:/Users/PC/Desktop/Penta...		2017/04/13 10:04:31
Input_Networks	Job execution finished	Success		file:///C:/Users/PC/Desktop/Penta...	7	2017/04/13 10:04:36
Input_Fecha	Start of job execution		Followed link after success	file:///C:/Users/PC/Desktop/Penta...		2017/04/13 10:04:36
Input_Fecha	Job execution finished	Success		file:///C:/Users/PC/Desktop/Penta...	8	2017/04/13 10:18:36
JOB_DIMENSIONES	Job execution finished	Success		file:///C:/Users/PC/Desktop/Penta...	1	2017/04/13 10:18:36
respaldo_eventos_data	Start of job execution		Followed link after success	file:///C:/Users/PC/Desktop/Penta...		2017/04/13 10:18:36
respaldo_eventos_data	Job execution finished	Success		file:///C:/Users/PC/Desktop/Penta...	2	2017/04/13 10:18:41
eventos_data_estructurado	Start of job execution		Followed link after success	file:///C:/Users/PC/Desktop/Penta...		2017/04/13 10:18:41
eventos_data_estructurado	Job execution finished	Success		file:///C:/Users/PC/Desktop/Penta...	3	2017/04/13 10:18:46
hechos_eventos	Start of job execution		Followed link after success	file:///C:/Users/PC/Desktop/Penta...		2017/04/13 10:18:46
hechos_eventos	Job execution finished	Success		file:///C:/Users/PC/Desktop/Penta...	4	2017/04/13 10:18:50
case_ciudad	Start of job execution		Followed link after success	file:///C:/Users/PC/Desktop/Penta...		2017/04/13 10:18:50
case_ciudad	Job execution finished	Success		file:///C:/Users/PC/Desktop/Penta...	5	2017/04/13 10:18:51
CREATE_FC_HECHOS_EVENT(	Start of job execution		Followed link after success	file:///C:/Users/PC/Desktop/Penta...		2017/04/13 10:18:51
CREATE_FC_HECHOS_EVENT(	Job execution finished	Success		file:///C:/Users/PC/Desktop/Penta...	6	2017/04/13 10:18:53
Input_Sistema_Operativo	Start of job execution		Followed link after success	file:///C:/Users/PC/Desktop/Penta...		2017/04/13 10:18:53
Input_Sistema_Operativo	Job execution finished	Success		file:///C:/Users/PC/Desktop/Penta...	7	2017/04/13 10:18:54
Input_Hostname	Start of job execution		Followed link after success	file:///C:/Users/PC/Desktop/Penta...		2017/04/13 10:18:54
Input_Hostname	Job execution finished	Success		file:///C:/Users/PC/Desktop/Penta...	8	2017/04/13 10:18:59
Job: JOB EXTRACCION	Job execution finished	Success	finished		8	2017/04/13 10:18:59

Figura 13: Proceso de carga satisfactoria

5.4.2.2 Pruebas a la Plataforma de Business Intelligence

La Tabla 33 describe las pruebas realizadas a la plataforma de Pentaho BI

Tabla 33. Pruebas Plataforma de BI

Prueba	Descripción
Funcional	Se evaluó las funcionalidades de: creación de cubos OLAP, reportes, dashboards.

De acuerdo a la Tabla 34 se especifica los casos de pruebas realizados a la plataforma y sus resultados.

Tabla 34. Casos de Pruebas

Módulo	Prueba	Resultado
Plataforma de BI	Login de Usuarios	Aprobado
	Ejecución de Análisis Ad-Hoc	Aprobado
Procesos ETL	Ejecución de Job_Dimensiones	Aprobado
	Carga Provincia	Aprobado

Continua

	Carga Ciudad	Aprobado
	Carga Institución	Aprobado
	Carga Tipo Evento	Aprobado
	Carga Networks	Aprobado
	Carga Fecha	Aprobado
	Respaldo de Hechos	Aprobado
	Estructuración de tabla de hechos	Aprobado
	Carga hechos eventos	Aprobado
Base de Datos Dimensional	Prueba de Consistencia de Datos	Aprobado
	Integración con Procesos ETL	Aprobado
Análisis	Ejecución de Reportes	Aprobado
	Ejecución de Dashboards	Aprobado

La descripción de las pruebas descritas en la anterior tabla se encuentra en los anexos.

5.5 Evaluación y resultados

Posteriormente del desarrollo de la plataforma utilizando la metodología de Kimball, inicia la etapa de evaluación de resultados que serán descritos a continuación.

5.5.1 Prueba de Concepto

La Figura 14 ilustra la plataforma de BI donde se ha ejecutado la prueba de concepto. Para levantar la plataforma de BI, se ha instalado un servidor CentOS versión 7, donde se encuentra el Pentaho BI, así como un servidor MySQL, que contiene el almacén de datos. El administrador de la plataforma crea, edita o ejecuta los procesos ETL, creando nuevos cubos OLAP mediante Schema Workbench o nuevos informes de Pentaho Report Designer y los publica directamente en el BI desde una PC. La comunidad de usuarios, en este caso la institución miembro de CSIRT, puede acceder para ver sus informes y los cuadros de mando.

Se realizan pruebas de consistencia de datos para verificar que la información del almacén de datos es válida a la de sus fuentes y no aparecen datos repetidos o inconsistencias en los datos presentados. Esto se hace con consultas directas a las fuentes y se comparan con el DW. Además, se verifican los procesos ETL creados, se verifican y se prueban los trabajos para autenticar el llenado del almacén de datos. Además, se realizan pruebas de roles de usuario en la plataforma Pentaho BI para que cada Institución sólo pueda acceder a la información correspondiente.

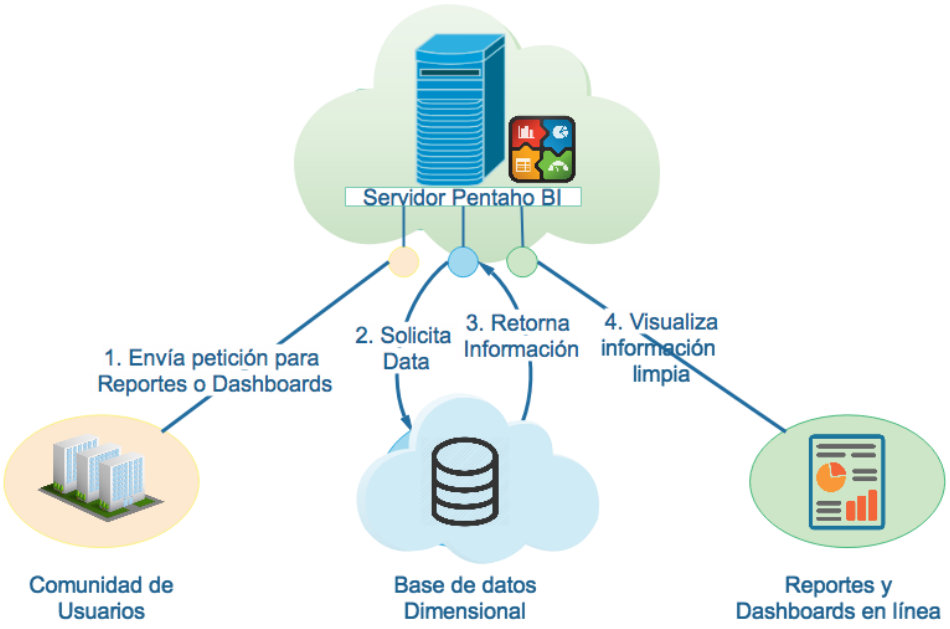


Figura 14. Topología para la prueba de concepto

5.5.2 Evaluación de Resultados

Después de completar la metodología de Kimball, el sistema de BI puede mostrar la cantidad de datos almacenados en el DW usando dashboards construidos en Pentaho BI. La mayoría de estos resultados se analizan tanto para la satisfacción de los gestores de incidentes de CSIRT como para los miembros de la comunidad de usuarios. A continuación, se muestra los reportes, dashboards, consultas ad-hoc y datos obtenidos en la plataforma de BI.

Reporte Eventos por Institución

Este reporte (Figura 15), contiene varios parámetros que pueden ser seleccionados. Se debe escoger la Institución a consultar, año y tipo de evento. Se despliega el reporte que contiene los siguientes datos: Año, Mes, Día, Nombre de Evento, Institución, Provincia y Ciudad. El informe ayuda a los administradores de cada institución a visualizar los tipos de eventos que ocurren en sus redes para realizar seguimiento de los mismos. Se presenta información esencial de cada incidente y una suma total de los mismos. Con los parámetros de entrada se facilita la búsqueda.

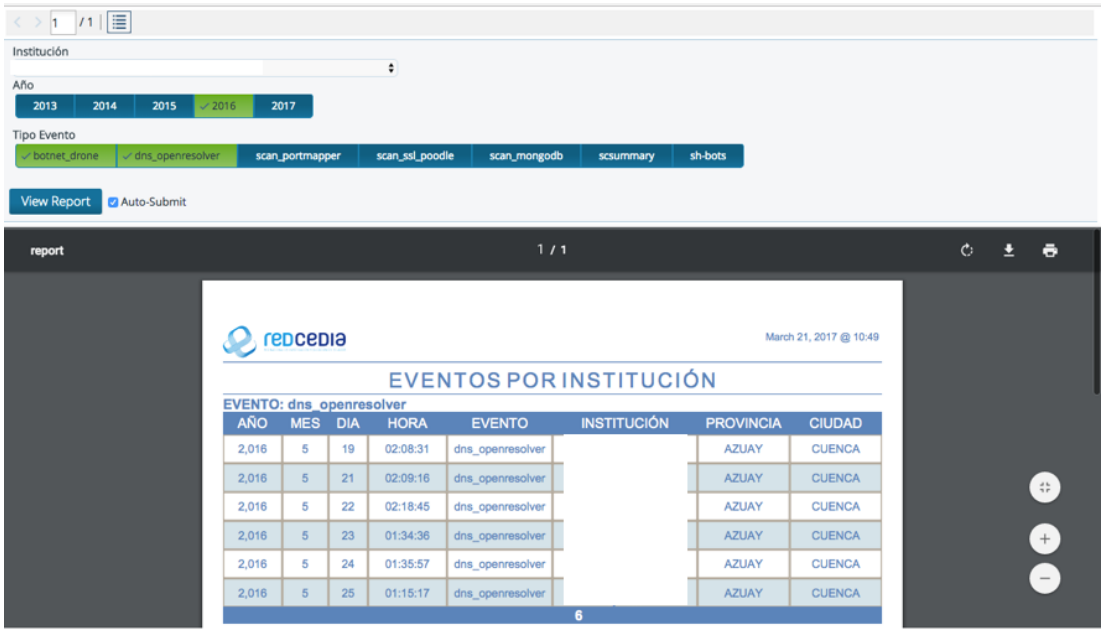


Figura 15. Reporte Eventos por Institución

Se puede guardar o imprimir directamente el reporte haciendo click en la flecha señalada (Figura 16).

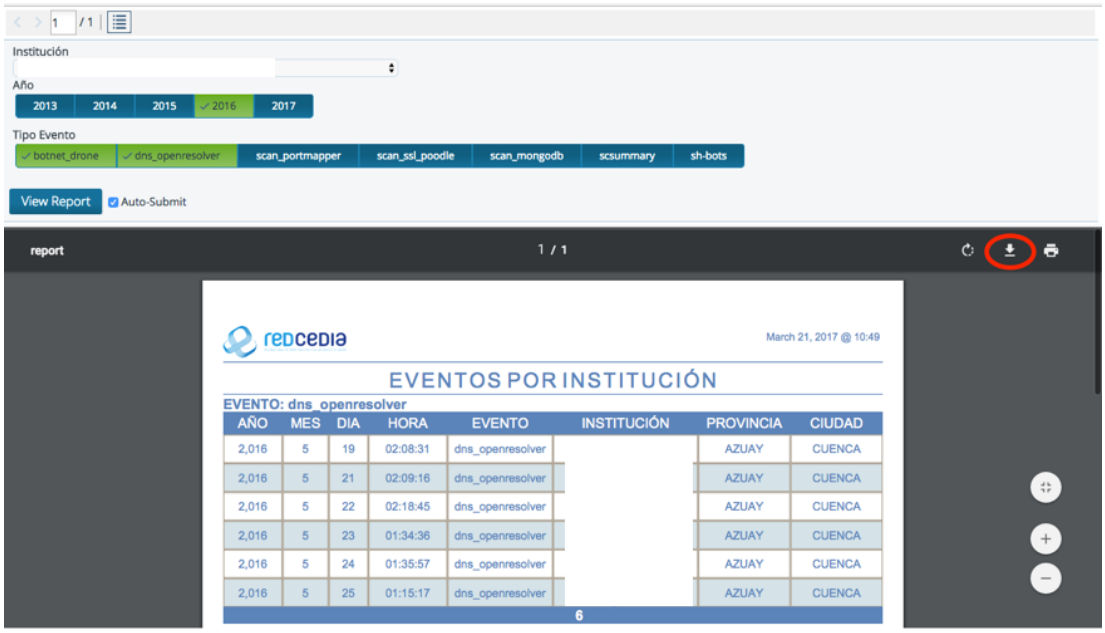


Figura 16. Guardar Reporte

Reporte Estado de Eventos

El siguiente reporte (Figura 17), muestra la información de eventos que están abiertos o cerrados, se puede seleccionar los parámetros de Institución, Año, Tipo de Evento y Estado. Se visualiza eventos que se encuentran abiertos o cerrados, un incidente se considera abierto o activo cuando ocurre dentro de los siete días anteriores a la fecha actual, las instituciones pueden observar los eventos de los que están siendo víctimas en ese momento para que tomen las correcciones necesarias.

Figura 17. Reporte Estado de Eventos

Reporte Cantidad de Eventos

En este reporte (Figura 18), se puede visualizar la cantidad de eventos de una institución al año. Un evento se considera el mismo si se repite en los posteriores siete días. Se debe seleccionar los parámetros, Año, Institución y Tipo de Evento. La institución revisa el número de incidentes generados en el año y la cantidad de tiempo que estuvieron activos. El reporte ayuda a contrastar la información con años anteriores y verificar si ha existido una mejora en seguridad de sus redes.

Figura 18. Reporte Cantidad de Eventos

Reporte Eventos que más se han demorado en Solucionarse

Este reporte (Figura 19), describe los eventos que más se han demorado en solucionarse. Se puede escoger uno o más años para visualizar la información. Los datos que despliega el reporte son: Año, Nombre de Evento, Institución, Número de Eventos en el año. El resultado del informe genera una tendencia entre los eventos que no han tenido atención por parte de los administradores de cada institución

Figura 19. Reporte Eventos que más se han demorado en Solucionarse

5.4.2.4 Pruebas de Dashboards

Dashboard Consolidado por Año

En este dashboard (Figura 20), se visualiza las instituciones que poseen la mayor cantidad eventos registrados y los eventos que más se repitieron en el año. Para cambiar el año dar click en el parámetro encerrado en un círculo. Se despliega una lista de años. El resultado del Dashboard ayuda a los especialistas de CEDIA a conocer cuáles son los eventos que más ocurren por año y generar las alertas necesarias a las instituciones que más vulnerabilidades presentan.

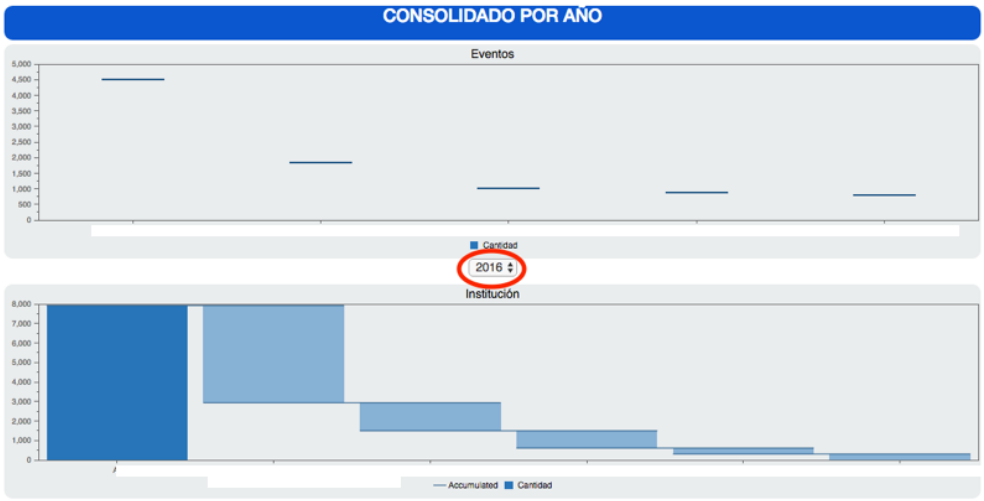


Figura 20. Dashboard Consolidado por Año

Dashboard Comportamiento en el Tiempo de Eventos

Este dashboard (Figura 21), describe el comportamiento de eventos en el tiempo, se puede seleccionar el tipo de evento en la parte superior y el año en la parte central. El gráfico superior muestra cómo ha evolucionado el evento a través de los años y en gráfico inferior se muestra la cantidad de eventos ocurridos en cada mes.

Figura 21. Dashboard Comportamiento en el Tiempo de Eventos

Dashboard Nivel de Seguridad

En la parte izquierda del Dashboard (Figura 22), se puede visualizar el porcentaje de las instituciones según el nivel de seguridad (instituciones que más incidentes presentan) y en la parte derecha la cantidad de eventos de distintos niveles de sensibilidad y criticidad. Los parámetros que se pueden escoger son el año y el mes.

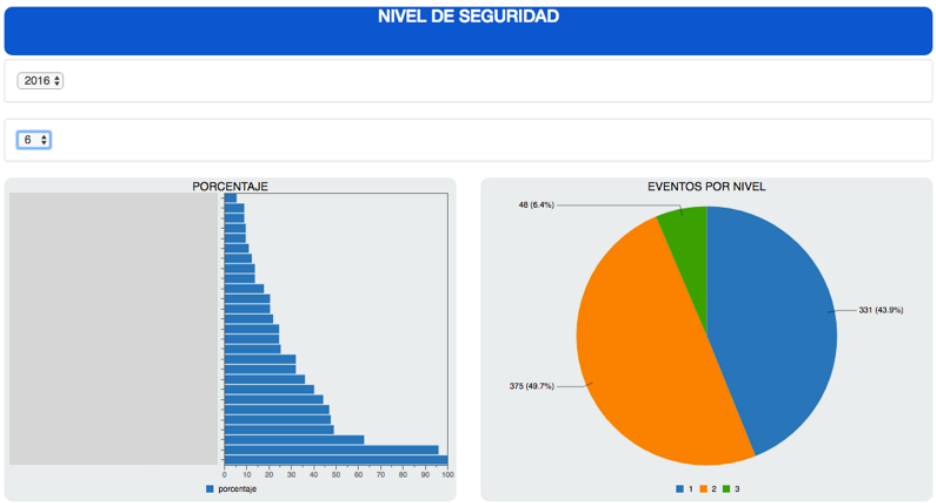


Figura 22. Dashboard Nivel de Seguridad

Análisis con Cubo OLAP

Para realizar análisis y gráficos utilizando el cubo OLAP, ingresar como administrador al Pentaho BI, (Figura 23).

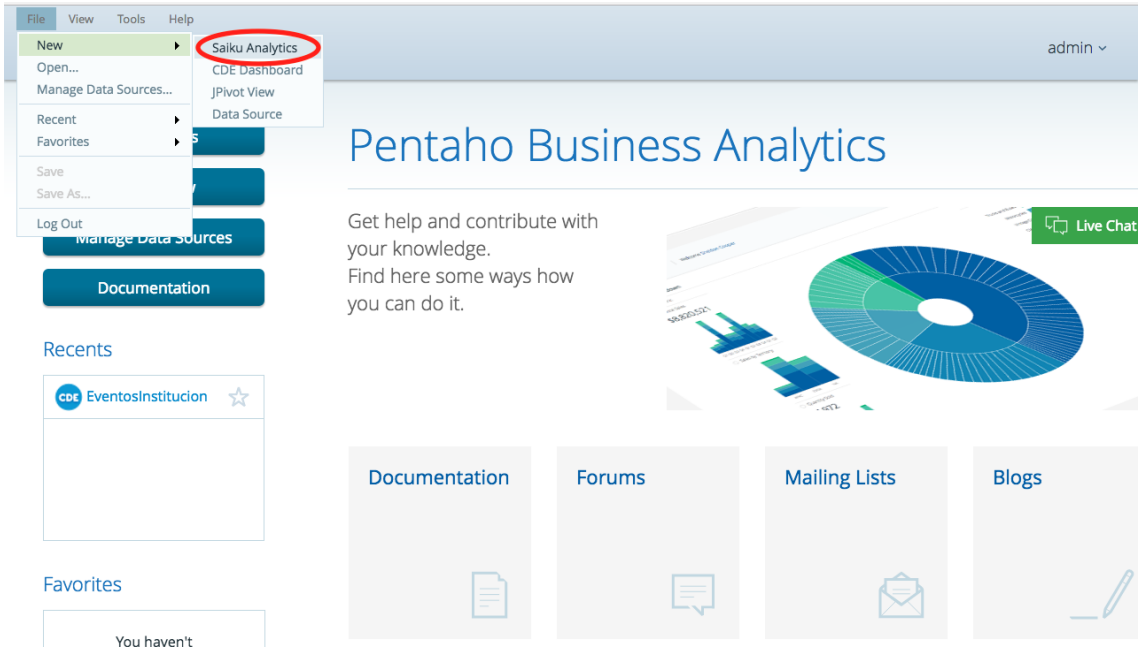


Figura 23. Utilizar cubo OLAP

Se debe dar click en File -> New -> Saiky Analytics, Figura 22, seleccionar el cubo que se va a utilizar, (Figura 24).

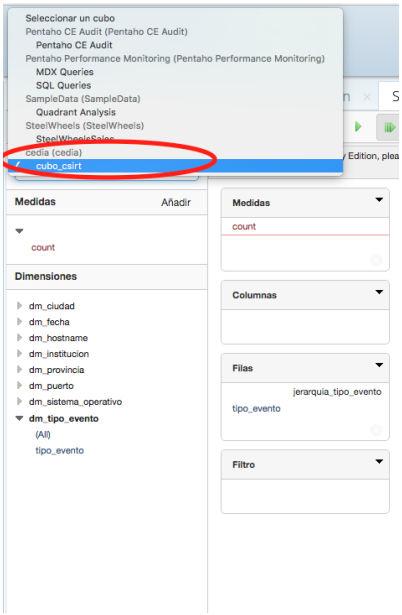


Figura 24. Selección de Cubo OLAP

Una vez seleccionado se tienen las dimensiones que se pueden analizar. Se deben arrastrar las dimensiones a las filas, columnas o medidas, (Figura 25).

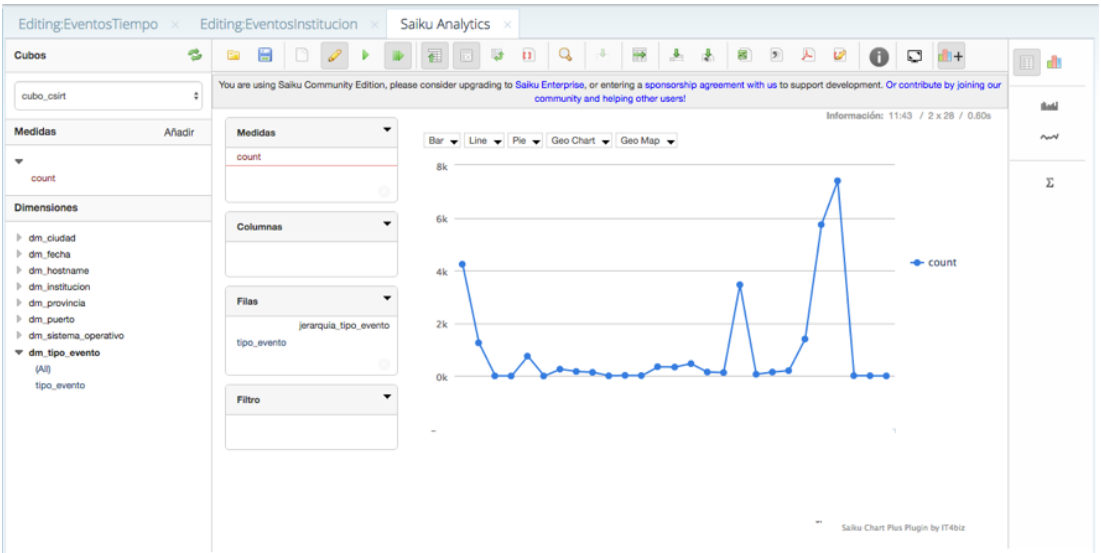


Figura 25. Análisis con Cubo OLAP

La Figura 26 ilustra los eventos más significativos de 2016. Con esta información, se pueden tomar decisiones sobre las medidas correctivas que se deben imponer para mejorar la seguridad de la información en las Instituciones miembros del CSIRT.

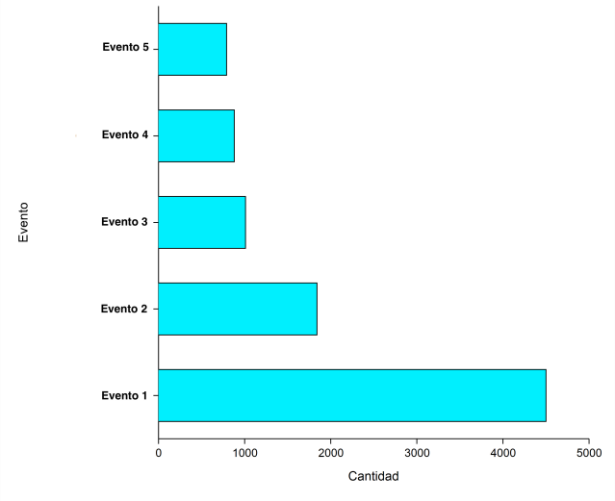


Figura 26. Eventos que más ocurrieron en el año 2016

La Tabla 35 enumera la evolución de la ocurrencia de eventos en las Instituciones. Hubo 5569 ciberataques registrados en 2014, mientras que en 2015 fueron apenas 3067. Sin embargo, en 2016 aumentaron hasta 11.675 ataques, mientras que en el año en curso aproximadamente 6190 ataques ya han ocurrido. Estos datos se han obtenido a través de consultas ad-hoc de los cubos OLAP de información.

Tabla 35

Evolución de ocurrencia de eventos

Año	Cantidad de eventos
2013	251
2014	5569
2015	3067
2016	11675
2017	6190

En la clasificación de los tipos de eventos, la Figura 28 indica un mayor número de eventos bots, que es un tipo de código malicioso. Este malware incluye funcionalidad de replicación y mecanismos de comunicación con el atacante que le permite ser controlado remotamente. El proceso de infección y propagación de un bot es similar al de un gusano. Además, es capaz de propagarse automáticamente, explotando las vulnerabilidades existentes en los ordenadores.

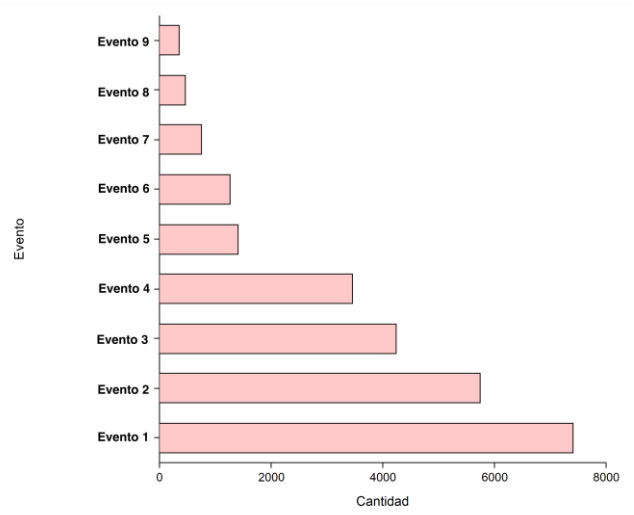


Figura 27. Número de eventos por tipo

Con respecto al grado de criticidad y sensibilidad de los eventos, según FIRST, hay una mayoría del nivel dos, que es de impacto medio para las Instituciones (Figura 29). Sin embargo, no deben pasarse por alto, sino que deben analizarse y buscarse correctivos.

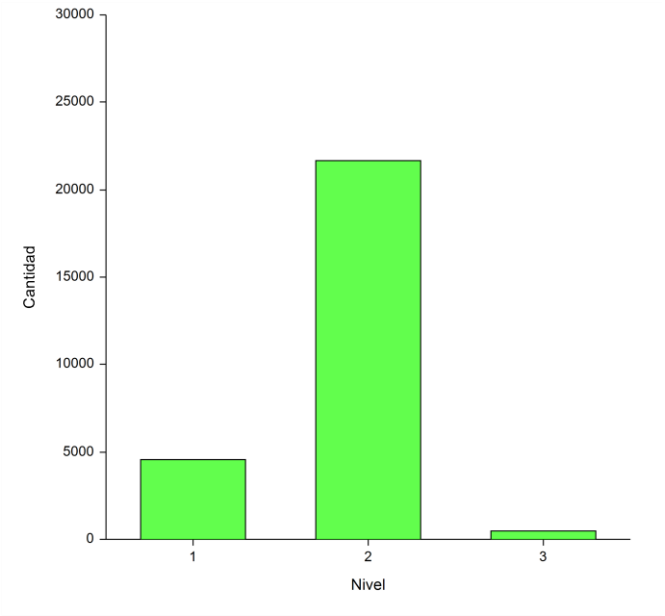


Figura 28. Número de eventos clasificado por nivel de criticidad y sensibilidad

CAPÍTULO VI
CONCLUSIONES Y RECOMENDACIONES

6.1 Conclusiones

Este estudio tiene como objetivo establecer un sistema de base de datos dimensional e inteligencia de negocios para analizar y mejorar el nivel de seguridad general de las redes y equipos mediante la prestación de servicios de alerta temprana, el análisis y la evaluación de los incidentes de un CSIRT ejecutado. Se ha investigado las prácticas de los CSIRT en Ecuador y los procedimientos que realiza CEDIA con los eventos registrados y su tratamientos, considerando que las instituciones no han realizado las correcciones necesarias para enmendar vulnerabilidades y eventos que ocurren en sus redes.

Se ha extraído información de diferentes fuentes de escáneres que captan información sobre vulnerabilidades, malware y botnets, entre otros. La información más relevante de cada evento ha sido formateada, para que las Instituciones puedan visualizar sus vulnerabilidades para corregirlas. Se ha montado una solución de base de datos dimensional y Business Intelligence, utilizando los pasos de la metodología Ralph Kimball, que genera un almacén de datos de información recopilada de alertas y eventos grabados de una transmisión continua de datos de varias fuentes de seguridad de Internet. Entre las fases realizadas, se ha diseñado un modelo de datos dimensionales con procesos ETL, mientras que se han implementado algoritmos de extracción y se han construido cubos OLAP. Se otorgaron puntuaciones a los eventos para dar prioridad a aquellos que son capaces de causar un daño más severo a las Instituciones. Por último, también se han construido informes y dashboards con Pentaho BI, consultas ad-hoc e informes.

El aporte de este proyecto es utilizar técnicas de Business Intelligence para el manejo de eventos en un CSIRT, con el objetivo de homogeneizar las fuentes de datos y obtener información actualizada y continua sobre incidentes. Constituye un gran beneficio para las instituciones miembro conocer los eventos ocurridos,

fundamental para planificar y ejecutar planes de corrección y mejoramiento de la seguridad en sus redes.

6.2 Recomendaciones

La utilización de software libre en pequeñas y medianas intituciones es favorable para reducir gastos, no obstante se recomienda en un futuro obtener licencia de Pentaho para poder utilizar nuevas funcionalidades en la plataforma que complementen el trabajo realizado y faciliten aun más la obtención de información.

Obtener información de eventos ocurridos debe ir de la mano con la resolución de los mismos. Se recomienda revisar constantemente la información proporcionada por la plataforma y actuar en caso de eventos graves, ponerse en contacto con el CSIRT para el seguimiento de incidentes y resolución de los mismos.

Como trabajo futuro se planea integrar la aplicación con diferentes sistemas que forman parte del CSIRT.

REFERENCIAS BIBLIOGRÁFICAS

CSIRT. (Noviembre de 2006). CSIRT. Obtenido de http://www.csirt.org/

datawarehouse4u. (02 de 09 de 2009). Obtenido de http://datawarehouse4u.info/Data-warehouse-schema-architecture-star-schema.html

Administración Electrónica España. (2013). Recuperado el 2016, de http://administracionelectronica.gob.es/pae_Home/pae_Documentacion/pae_Metodolog/pae_Magerit.html#.VsNU95PhC34

Edureka. (02 de Septiembre de 2014). Obtenido de https://www.edureka.co/blog/introduction-to-prd-pentaho-report-designer-pentaho-bi/

FIRST. (03 de Agosto de 2005). Obtenido de https://www.first.org/_assets/resources/guides/csirt_case_classification.html

GeekInterview. (09 de 01 de 2008). Obtenido de http://www.learn.geekinterview.com/data-warehouse/dw-basics/what-is-an-ad-hoc-query.html

Infoblox Experts Community. (13 de Mayo de 2014). Infoblox Experts Community. Obtenido de https://community.infoblox.com/t5/IPv6-Center-of-Excellence/Finding-and-Fixing-Open-DNS-Resolvers/ba-p/3405

ISO 27000. (2014). ISO 27000. Recuperado el 2016, de http://www.iso27000.es/download/doc_sgsi_all.pdf

Kimball, R. (02 de 08 de 2005). Obtenido de http://www.kimballgroup.com/data-warehouse-business-intelligence-resources/kimball-techniques/dimensional-modeling-techniques/star-schema-olap-cube/

Kimball, R. (02 de Agosto de 2005). Kimball Group. Obtenido de http://www.kimballgroup.com/1997/08/a-dimensional-modeling-manifesto/

Kimball, R. (2008). Kimball Group. Obtenido de http://www.kimballgroup.com/data-warehouse-business-intelligence-resources/kimball-techniques/dw-bi-lifecycle-method/

Kimball, R. (02 de 08 de 2013). Obtenido de http://www.kimballgroup.com/data-warehouse-business-intelligence-resources/kimball-techniques/etl-architecture-34-subsystems/

Martinez, C. H. (2014). Recuperado el 2016, de http://www.suarez-menendez.com/Publicaciones/BrochuresServicios/Analisis%20de%20Riesgo%20de%20TI.pdf

Meteorite. (25 de Marzo de 2016). Obtenido de http://meteorite.bi/products/saiku

Nielsen, J. (2013). Usability Engineering. Morgan Kaufman, San Francisco.

Oracle. (02 de Febrero de 2016). Obtenido de https://www.oracle.com/mysql/index.html

Pentaho. (21 de Marzo de 2009). Obtenido de http://mondrian.pentaho.com/documentation/workbench.php

Pentaho. (20 de Febrero de 2016). Obtenido de http://www.pentaho.com/about

Pentaho. (20 de Febrero de 2016). Obtenido de http://www.pentaho.com/product/data-integration

pmg-ssi. (31 de marzo de 2015). pmg-ssi. Recuperado el 2016, de http://www.pmg-ssi.com/2015/03/iso-27001-establecer-los-objetivos-para-la-ciberseguridad/

RANJAN, J. (2009). Journal of Theoretical and Applied Information Technology (Vol. 9).

Rouse, M. (12 de 08 de 2012). TechTarget. Obtenido de http://searchbusinessanalytics.techtarget.com/definition/business-intelligence-dashboard

SAP. (02 de Agosto de 2016). Obtenido de https://www.sap.com/product/data-mgmt/powerdesigner-data-modeling-tools.html

Shneiderman , B., & Plaisant, C. (2004). Designing the Use Interdace: Strategies for Effective Human-Computer Interaction. Pearson/Addison-Wesley, Boston.

Unión Internacional de Telecomunicaciones. (2007). Unión Internacional de Telecomunicaciones. Obtenido de http://www.itu.int/ITU-D/cyb/publications/2007/cgdc-2007-s.pdf

Wamala, F. (2011). ITU national cybersecurity strategy guide. International Telecommunications Union, 11.

Hellwig, O., Quirchmayr, G., Huber, E., Goluch, G., Vock, F., & Pospisil, B. (2016, August). Major Challenges in Structuring and Institutionalizing CERT-Communication. In Availability, Reliability and Security (ARES), 2016 11th International Conference on (pp. 661-667). IEEE.

Kruidhof, Olaf. "Evolution of National and Corporate CERTs-Trust, the Key Factor." (2014): 81-96.

Qian, Y., Fang, Y., Jaatun, M. G., Johnsen, S. O., & Gonzalez, J. J. (2010, January). Managing emerging information security risks during transitions to Integrated Operations. In System Sciences (HICSS), 2010 43rd Hawaii International Conference on (pp. 1-11). IEEE.

Yang, J., Ryu, D., & Baik, J. (2016, January). Improving vulnerability prediction accuracy with secure coding standard violation measures. In Big Data and Smart Computing (BigComp), 2016 International Conference on (pp. 115-122). IEEE.

Bollinger, J., Enright, B., & Valites, M. (2015). Crafting the InfoSec Playbook: Security Monitoring and Incident Response Master Plan. " O'Reilly Media, Inc.". ISBN: 9781491913598.

Ruefle, R., Dorofee, A., Mundie, D., Householder, A. D., Murray, M., & Perl, S. J. (2014). Computer security incident response team development and evolution. IEEE Security & Privacy, 12(5), 16-26.

Bhatt, S., Manadhata, P. K., & Zomlot, L. (2014). The operational role of security information and event management systems. IEEE Security & Privacy, 12(5), 35-41.

Osorno, M., Millar, T., & Rager, D. (2011). Coordinated Cybersecurity Incident Handling: Roles, Processes, and Coordination Networks for Crosscutting Incidents. Johns Hopkins Univ Laurel Md Applied Physics Lab.

Belsis, M. A., Simitsis, A., & Gritzalis, S. (2005, November). Workflow based security incident management. In Panhellenic Conference on Informatics (pp. 684-694). Springer Berlin Heidelberg.

Patrick, H., & Fields, Z. (2017). A Need for Cyber Security Creativity. In Collective Creativity for Responsible and Sustainable Business Practice (pp. 42-61). IGI Global.

Cichonski, P., Millar, T., Grance, T., & Scarfone, K. (2012). Computer security incident handling guide. NIST Special Publication, 800, 61.

Johnson, C., Badger, L., Waltermire, D., Snyder, J., & Skorupka, C. (2016). Guide to cyber threat information sharing. NIST Special Publication, 800, 150.

Wu, D. D., Chen, S. H., & Olson, D. L. (2014). Business intelligence in risk management: Some recent progresses. Information Sciences, 256, 1-7.

Elmellas, J. (2016). Knowledge is power: the evolution of threat intelligence. Computer Fraud & Security, 2016(7), 5-9.

Grobler, M., Jacobs, P., & van Niekerk, B. (2016). Cyber Security Centres for Threat Detection and Mitigation. Threat Mitigation and Detection of Cyber Warfare and Terrorism Activities, 21.

Sharkov, G. (2016). From Cybersecurity to Collaborative Resiliency. In Proceedings of the 2016 ACM Workshop on Automated Decision Making for Active Cyber Defense (pp. 3-9). ACM.

Mejía, J., Muñoz, M., Ramírez, H., & Peña, A. (2016). Proposal of Content and Security Controls for a CSIRT Website. In New Advances in Information Systems and Technologies (pp. 421-430). Springer International Publishing.

Rajasekharaiah, K. M., Dule, C. S., & Srimani, P. K. (2016, March). CRSA cryptosystem based secure data mining model for business intelligence applications. In Electrical, Electronics, and Optimization Techniques (ICEEOT), International Conference on (pp. 879-884). IEEE.

Gahi, Y., Guennoun, M., & Mouftah, H. T. (2016, June). Big Data Analytics: Security and privacy challenges. In Computers and Communication (ISCC), 2016 IEEE Symposium on (pp. 952-957). IEEE.

Zuech, R., Khoshgoftaar, T. M., & Wald, R. (2015). Intrusion detection and big heterogeneous data: a survey. Journal of Big Data, 2(1), 3.

Mahmood, T., & Afzal, U. (2013, December). Security analytics: Big data analytics for cybersecurity: A review of trends, techniques and tools. In Information assurance (ncia), 2013 2nd national conference on (pp. 129-134). IEEE.

Jaramillo, E., Munier, M., & Aniorté, P. (2013). Information security in business intelligence based on cloud: A survey of key issues and the premises of a proposal. WOSIS.

Ahmad, A., Hadgkiss, J., & Ruighaver, A. B. (2012). Incident response teams–Challenges in supporting the organisational security function. Computers & Security, 31(5), 643-652.

Kimball, R., Ross, M., Becker, B., Thornthwaite, W., & Mundy, J. (2015). The Kimball Group Reader: Relentlessly Practical Tools for Data Warehousing and Business Intelligence Remastered Collection. John Wiley & Sons.

Kimball, R., Margy R. (2013). The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling (3rd. edition). Wiley. ISBN 978-1-118-53080-1.

Kimball, R., & Ross, M. (2011). The data warehouse toolkit: the complete guide to dimensional modelling. John Wiley & Sons.

Bouman, R., & Van Dongen, J. (2009). Pentaho solutions. Business Intelligence and Data Warehousing with Pentaho and MYSQL.

El-Sappagh, S. H. A., Hendawi, A. M. A., & El Bastawissy, A. H. (2011). A proposed model for data warehouse ETL processes. Journal of King Saud University-Computer and Information Sciences, 23(2), 91-104.

Kimball, R., & Caserta, J. (2011). The Data Warehouse? ETL Toolkit: Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data. John Wiley & Sons.

Mansur, A. (2012). Seguridad Informatica y administración de datos de los sistemas de información contable para la productividad y competitividad de las PYME España: Eumed. Eumed. net.