

# Universidad CENFOTEC

Maestría en Tecnología de Bases de Datos

Proyecto Integrado 2

## **Proyecto de Graduación**

**“ANÁLISIS DE EVENTOS E INCIDENTES EN EL DEPARTAMENTO GLOBAL SYSTEM MONITORING (GSM) EN INTEL COSTA RICA POR MEDIO DE TÉCNICA DE MINERÍA DE TEXTO MEDIANTE LA IMPLEMENTACIÓN DE UN DATA WAREHOUSE 2.0”**

### **Profesor**

Msc. José Cabezas Jaikel

### **Integrantes**

Ing. Alexandra Aguirre Rojas

Msc. Jairo Brenes Delgado

**Abril, 2015**

**San José, Costa Rica**

## Resumen Ejecutivo

El presente documento muestra la documentación respectiva a una herramienta desarrollada para la utilización de la minería de datos en un Data Warehouse 2.0 mediante escenarios con información de alertas e incidentes del departamento Global System Monitoring (GSM) en Costa Rica y Malasia respectivamente, de la compañía Intel. La utilización de algoritmos en la minería de texto, son fundamentales para la búsqueda de tendencias mediante palabras claves, lo cual, es y será, un pilar fundamental en las organizaciones pues les permitirá transformar en conocimiento los grandes volúmenes de datos existentes hoy en día.

A partir de ahí, el presente proyecto tiene como objetivo principal la implementación de un Data Warehouse 2.0 y su respectiva aplicación de técnicas de minería de texto que permitan recuperar las descripciones de tickets relacionados al soporte de la plataforma tecnológica de manejo de tickets de Intel, en tiempo real. Se manejarán tres categorías de tickets: *incidentes*, los cuales son problemas específicos no esperados del cliente los cuales no pueden ser asociados a un guion de resolución, los “¿cómo puedo?” concernientes a preguntas frecuentes a resolver mediante un script y los tickets tipo *solicitud* donde el cliente pide un servicio en específico, por ejemplo el aumento de espacio en disco a nivel de servidor.

Las descripciones de los tickets no poseen ningún formato estándar por lo cual es difícil recuperar manualmente cuales son los problemas más frecuentes. Por esta razón, se utilizarán técnicas de minado de texto, para otorgar la oportunidad de mostrar a los usuarios aquellas palabras claves que podrían considerarse como relevantes para su análisis, ya sea en períodos de una hora para alertar de una posible degradación de servicios en una aplicación determinada, o en tiempos iguales o menores a un año a fin de poder analizar proactiva y reactivamente los problemas enfrentados por un servidor y aquellos que podrían llegar a generarse.

Como resultado, se implementarán planes de acción que permitan disminuir el volumen de tickets relacionados a una misma área de soporte; para evitar escalaciones<sup>1</sup> futuras, visualizar elementos recurrentes que afecten la operaciones de negocio, incrementar los niveles de resolución de incidentes y finalmente aumentar el nivel de confianza de los clientes.

---

<sup>1</sup> **Escalación:** se refiere al proceso de requerir un especialista de más alto nivel para resolver un problema.

**Palabras clave:** Data Warehouse 2.0. Intel. Minería de Datos. ITIL v3. Técnicas de pre-procesado de datos. Herramienta de Gestión de Incidentes y alertas, Análisis de Minado de Texto (SSIS Term Extraction) . Identificación de Patrones Recurrentes de Texto.

## Tabla de Contenidos

<b>Capítulo 1: Introducción</b> .....	1
Antecedentes .....	1
Justificación .....	3
Marco Institucional .....	4
¿Qué abarca el departamento GSM?.....	5
Planteamiento del Problema .....	7
Problema General.....	8
Objetivos .....	8
Objetivo General .....	8
Objetivos Específicos.....	9
Alcances.....	9
Limitaciones.....	10
<b>Capítulo II: Marco Metodológico</b> .....	12
Tipo de Investigación.....	12
Área de Estudio.....	13
Objeto y Sujeto de estudio .....	13
Población y muestra.....	14
Fuentes de Información.....	14
Diseño de técnicas e instrumentos .....	15
Estado de la Cuestión.....	16
Situación Actual.....	18
Capacidad actual en el análisis de la información .....	18
Requerimientos esperados para el análisis de la información.....	18
Análisis de la brecha de estados: actual vs deseado.....	21
Análisis de la información deseada.....	21
<b>Capítulo III: Desarrollo de la Solución</b> .....	23
Análisis del Negocio .....	23
Selección de Técnicas para la Implementación del proyecto.....	23
Arquitectura Data Warehouse 2.0 .....	23
Modelo Tabular (SSAS).....	24

Minería de Datos: Minería de Texto .....	26
Análisis de Herramientas de Minería de Texto.....	27
Análisis Técnico de las estructuras fuentes.....	29
Creación del Data Warehouse 2.0.....	32
Análisis de Infraestructura y arquitectura de datos .....	32
Diseño del Modelado Multidimensional.....	33
Arquitectura Propuesta.....	34
Identificación de las tablas fuentes de los Sistemas Transaccionales .....	36
Modelo Data Warehouse: Esquema estrella del Data Warehouse .....	36
Carga de datos al Staging Area.....	37
Descripción de cada Dimensión y Tabla de Hechos.....	40
Carga de datos al Data Warehouse .....	41
Carga de datos al Modelo Tabular .....	43
Modelo tabular.....	43
Creación de las particiones para el Data Warehouse 2.0 .....	45
<b>Capítulo IV: Análisis de Resultados.....</b>	<b>47</b>
Herramientas utilizadas para la presentación de datos.....	47
Presentación del Dashboard según fases del Data Warehouse 2.0 .....	47
Sector Interactivo .....	47
Sector Integrado .....	49
Sector Archivo .....	52
Unido a esto el usuario final puede crear estos mismos reportes de forma local utilizando power view con Excel. A continuación se muestra cuáles pasos se deben seguir.....	53
Visualización de la solución mediante Excel.....	53
Análisis e interpretación de los resultados .....	59
Sector Interactivo .....	59
Sector Integrado .....	61
Sector Archivo .....	62
Análisis mediante minería de texto.....	64
<b>Capítulo V: Conclusiones y Recomendaciones .....</b>	<b>69</b>
Conclusiones .....	69
Recomendaciones .....	71
<b>Capítulo VI: Referencias Bibliográficas .....</b>	<b>72</b>

Bibliografía .....	72
Capítulo VII: Anexos .....	74
Anexos .....	74
Work Breakdown Structure (WBS) .....	74
Scripts para el Staging Area.....	<b>Error! Bookmark not defined.</b>

## Tabla de Figuras

<i>Figura 1 Flujo de manejo de incidentes en GSM</i> .....	6
<i>Figura 2 Organigrama GSM</i> .....	7
<i>Figura 3 Interfaz Service Now</i> .....	30
<i>Figura 4 Impact Manager</i> .....	31
<i>Figura 5 Diseño de la Arquitectura Propuesta</i> .....	35
<i>Figura 6 Esquema Estrella del Data Warehouse</i> .....	37
<i>Figura 7 Staging Area-Fuentes de Datos</i> .....	38
<i>Figura 8 Carga de Datos al Data Warehouse</i> .....	42
<i>Figura 9 Carga de datos al tabular satisfactoria</i> .....	43
<i>Figura 10 Solución Modelo Tabular</i> .....	44
<i>Figura 11 Carga de datos de la Dimensión de Alertas</i> .....	44
<i>Figura 12 Creación de Particiones para el manejo de datos en el DW 2.0</i> .....	45
<i>Figura 13 Partición del Sector Integrado</i> .....	46
<i>Figura 14 Partición del Sector Archivo</i> .....	46
<i>Figura 15 Gráficos del Sector Interactivo</i> .....	48
<i>Figura 16 Gráficos del Sector Interactivo</i> .....	48
<i>Figura 17 Gráficos del Sector Interactivo</i> .....	49
<i>Figura 18 Información del Servidor - Problemas e Incidentes</i> .....	51
<i>Figura 19 Información del Servidor - Parches y Alertas</i> .....	52
<i>Figura 20 Salud del Servidor</i> .....	53
<i>Figura 21 Power View Excel</i> .....	54
<i>Figura 22 Power View Excel</i> .....	54
<i>Figura 23 Power View Excel</i> .....	54
<i>Figura 24 Power View Excel</i> .....	55
<i>Figura 25 Power View Excel</i> .....	55
<i>Figura 26 Power View Excel</i> .....	56
<i>Figura 27 Power View Excel</i> .....	56
<i>Figura 28 Incidetes por Región y Tipo</i> .....	57

<i>Figura 29 Problemas por Servicio</i> .....	57
<i>Figura 30 Alertas por Tipo</i> .....	58
<i>Figura 31 Espacio en Disco</i> .....	58
<i>Figura 32 Resultado de la aplicación del Algoritmo "Term Extracction"</i> .....	59
<i>Figura 33 Recurrencias de clases de alertas</i> .....	60
<i>Figura 34 Recurrencias de clases de alertas</i> .....	61
<i>Figura 35 Alertas generadas por colectores de alertas</i> .....	61
<i>Figura 36 Espacio en Disco</i> .....	63
<i>Figura 37 Parches por Estado</i> .....	63
<i>Figura 38 Term Extraction</i> .....	64
<i>Figura 39 Estructura interna del flujo de datos del Algoritmo de Minería de Texto llamado "Term Extraction" .....</i>	64
<i>Figura 40 Paquete que extrae información de la descripción de Eventos para ser analizados por "Term Extraction"</i> .....	65
<i>Figura 41 Selección de la columna a ser analizada por "Term Extraction"</i> .....	65
<i>Figura 42 Exclusión de términos a ser analizados por "Term Extraction"</i> .....	66
<i>Figura 43 Definición de términos a ser utilizados por el algoritmo de Text Mining "Term Extraction"</i> .....	67
<i>Figura 44 Selección de la Tabla con los resultados analizada por "Term Extraction"</i> .....	67
<i>Figura 45 Resultado del análisis de recurrencia de términos encontrados por el algoritmo "Term Extraction" .....</i>	68
<i>Figura 46 Work Breakdown Structure</i> .....	74

# Capítulo 1: Introducción

## Antecedentes

Durante los últimos años las empresas han pasado de tener grandes volúmenes de información en forma física (impresa) a tener el doble o más de esa información guardada en forma digital.

Esta evolución se ha venido dando de forma iterativa, donde las organizaciones cayeron en cuenta de que era insostenible manejar tanto papel de forma ordenada, de ahí surge la necesidad de migrar a medios digitales de almacenamiento. Sin embargo, actualmente ni siquiera los medios digitales están dando la respuesta requerida a las empresas, pues se tiene una exorbitante cantidad información subutilizada, la cual con el uso adecuado podría tener gran valor para la toma de decisiones. Aquí en este punto, es donde la información debe convertirse en conocimiento, como un proceso natural de tener las herramientas y los datos, pero a sabiendas de que es requerido procesarlo de alguna forma para sacarle el mayor provecho.

Por ejemplo Google, el motor de búsqueda más conocido a nivel mundial, es una de las primeras herramientas la cual ha permitido buscar y devolver contenidos. Esto mediante algoritmos, midiendo la relevancia de los resultados con técnicas de categorización, peso de palabras, relaciones entre palabras, entre otros.

En un contexto similar el presente trabajo pretende mostrar la utilidad de herramientas para la minería de texto en el campo de las Tecnologías de Información, en un área sensitiva de toda organización como lo es la Gestión de Alertas e Incidentes. En el caso de Intel Corporation se utiliza ITIL como la metodología de Gestión de Tecnologías de Información.

La mayoría de empresas, utilizan alguna herramienta que les permita manejar internamente su soporte a aplicaciones y el monitoreo del mismo. Esto representa una inversión de alto costo, donde el ideal de toda empresa es llegar a identificar todos los posibles problemas generados y evitar la resolución repetitiva e innecesaria de incidentes los cuales podrían ser resueltos en la primera instancia de soporte conocida como Nivel 1, ahorrando paralelamente en costos al no tener que escalarlo a niveles más altos (usualmente de mayor costo) o identificando causas raíz para evitar que esos problemas no se vuelvan a presentar.

Intel Corporation, utiliza la herramienta Service Now para el manejo de tiquetes: ya sean incidentes o solicitudes de servicio y para manejar un catálogo básico de elementos clave como lo son el nombres de las aplicaciones y servidores. Esto le permite mantener una base de datos de todos los tiquetes para análisis posteriores en la etapa de Gestión de problemas.

La compañía Intel posee una amplia variedad de departamentos con tareas específicas, sin embargo existen algunos en particular que cumplen una función estándar para cualquier servicio o aplicación dentro de la empresa. Un ejemplo de ello son los departamentos de monitoreo, los cuales como su nombre lo especifica tienen como objetivo monitorear y darle seguimiento a cualquier alerta o comportamiento no común de los servidores.

En Intel, éste departamento se conoce como: Global System Monitoring reconocido por el acrónimo GSM, en español se identificaría como: Sistema de Monitoreo Global. Su cobertura es de tipo 24/7, con presencia en Costa Rica y Malasia. Éste departamento utiliza un repositorio de alertas llamada BEM Impact Manager, allí son capturadas todas las alertas de todos los servidores de Intel.

Los agentes toman cada una de esas alertas y procuran buscar la solución más óptima para su pronta resolución, algunas de ellas representan falsas alarmas, otras pueden tener un impacto de tipo: bajo, medio, alto o crítico. Las mismas son atendidas por el nivel de criticidad más alto al más bajo.

Se puede identificar que el volumen de alertas a generarse en una hora o menos puede ser bastante alto, lo cual incurre en que los agentes tengan como prioridad atender las alertas más críticas. Pero ¿qué sucede si hay una serie de alertas recurrentes con impacto bajo que no han sido observadas durante horas?, ¿qué pasaría si un servidor está alertando gradualmente que algo puede estar ocurriendo?, de una forma manual éstas interrogantes son complicadas de contestar, ya que podemos tener diez agentes trabajando en una misma alerta, la cual, al ser repartida entre ellos no determina que puede estarse generando una recurrencia considerable de un mismo síntoma. ¿cómo podemos llegar a resolver un problema de esta índole? Se podría pensar que es simplemente un problema de comunicación entre agentes, sin embargo, en departamentos de éste tipo los Service Level Agreement<sup>2</sup> con los clientes son de minutos, es decir ellos no posee el tiempo para estar preguntando si alguien ha visto un mismo caso.

El presente proyecto, pretende minimizar situaciones como ésta, donde al utilizar minería de texto se llega a analizar las descripciones de las alertas en tiempo real y enviar correos de aviso por medio de reportería,

---

<sup>2</sup> **Service Level Agreement**, conocido por el acrónimo: SLA, es el contrato entre el departamento de Tecnologías de Información y el cliente donde se especifica en cuánto tiempo va estar resuelto un caso según su prioridad. Si el contrato no se cumple generalmente hay sanciones para el departamento de TI.

a los agentes de cuando alguna de éstas alertas se ésta convirtiendo en recurrente para tomar acciones inmediatas.

Para su desarrollo se propone un Data Warehouse 2.0 el cual nos permite tener un sector interactivo ligado a un ambiente transaccional representado por la consola de alertas. A fin de aprovechar al máximo la solución de un Data Warehouse 2.0 se ofrece dentro del alcance un sector integrado y otro de archivo, los cuáles permitirán al departamento de monitoreo realizar análisis más profundos de la salud actual de los servidores, uniendo diversas variables como lo son: rendimiento del servidor (uso del CPU, memoria, I/O), espacio en disco, problemas e incidentes asociados, aplicaciones asociadas, solicitudes y parches. La forma proyectada para realizar dichos análisis será mediante *dashboards* y reportes donde se ofrezcan éstas variables y se les permita a los agentes interactuar a fin de obtener conclusiones dependiendo de sus necesidades.

## Justificación

La justificación del presente proyecto, se basa en la búsqueda de una solución que permita identificar los problemas más recurrentes en el contexto de Alertas e Incidentes en el grupo Global System Monitoring en Costa Rica y Malasia, mediante la implementación de patrones recurrentes en tiempo real. La hipótesis propuesta se basa en el análisis del entorno actual, para lo cual se utilizarán como punto de partida los indicadores de incidentes y alertas más frecuentes provenientes de la Base de Datos de Gestión de Incidentes (conocida como “Servicie Now”) y de Impact Manager, respectivamente.

Tomando en cuenta esta información y por medio de la adecuada extracción de datos y análisis por medio de técnicas de Minado de Texto (SSIS - Term Extraction), se proporcionarán las guías elementales para la identificación de las áreas de mejora que permitan brindar una mejor calidad en el servicio ofrecido.

El punto clave para ello es identificar cuáles son aquellos problemas comunes y generados de forma recurrente, sin embargo no es una tarea fácil cuando los volúmenes son altos y no se puede invertir tiempo en la limpieza y extracción de datos claves de forma manual.

Se espera que con las acciones correctivas a implementar de forma posterior al análisis del presente proyecto, generen de forma positiva una mejora en el tiempo de resolución de los incidentes y alertas que

pueda ir más alineada al actual nivel de servicio ofrecido a los clientes incrementando consecuentemente el nivel de satisfacción y percepción del servicio.

Optimizando los tiempos de resolución, se tiene como expectativa poder invertir más tiempo en el análisis de problemas, los cuáles si son repetitivos están indicando factores aún no resueltos; contando con un Data Warehouse 2.0 se desea utilizar toda esa data histórica para poder comprender con mayor facilidad cuál o cuáles son los factores que podrían estar involucrados y como a futuro se puede evitar o mitigar incidentes de mayor calibre.

Se considera como un imperativo el hecho de que toda empresa quiera conocer el significado de su información y la importancia que ésta pueda aportar a sus objetivos estratégicos al corto, mediano o largo plazo.

## Marco Institucional

En este trabajo se presenta un análisis de las alertas e incidentes generados mediante consola al departamento de Global System Monitoring en Intel, por medio del sistema de Gestión de Incidentes “Service Now” e Impact Manager como repositorio de alertas. En ambas se registran los problemas técnicos de los empleados de la empresa, aplicaciones e infraestructura a través de los servidores que alojan las aplicaciones utilizadas a nivel empresarial.

La base de datos Service Now está considerada como el principal sistema de información de tiquetes y el mayor sistema utilizado en Intel, diferenciándose gradualmente de otras bases de datos utilizadas en la empresa.

Para identificar los valores más concernientes en esta base de datos, se ha elaborado una estrategia de búsqueda de términos relevantes de la propia base de datos y consola de alertas, la cual es altamente fiable, se actualiza constantemente y recoge todas las posibles variantes y sinonimias de los términos empleados.

Se realizó una selección de la terminología más apropiada, identificándose los términos que comprenden el contexto de incidentes relacionados al servicio ofrecido por el departamento Global System Monitoring.

Para identificar todos los posibles problemas relacionados es necesario un paso previo de limpieza de los datos en el que necesitara los distintos términos relacionados para una identificación óptima en las alertas

e incidentes. En total se analizará el total de incidentes y alertas generadas durante todo el 2014 y los meses de Enero, Febrero, Marzo y Abril del 2015.

### ¿Qué abarca el departamento Global System Monitoring?

Para lograr una mejor noción del departamento, su función y la relación que mantienen sus grupos funcionales, se describirán a continuación sus principales características.

GSM (Global System Monitoring, fue conocido anteriormente como el OSC (Operation Service Center). Actualmente Global System Monitoring ofrece un modelo de soporte de 24 horas por 7 días y cuenta con dos centros de soporte localizados en las siguientes regiones:

- Penang en Malasia (edificio: PG12).
- Belén en Costa Rica (edificio: CR2).

Global System Monitoring cuenta con cuatro grupos funcionales que ofrecen resolución remota de incidentes generados por la tecnología de MOM (Microsoft Operación Manager) en las siguientes áreas:

- **Aplicaciones:** Provee soporte especialmente a las principales aplicaciones e infraestructura de los clientes de negocio más importantes para Intel. Entre las aplicaciones más comunes que Global System Monitoring soporta se encuentran las siguientes: SAP, Tico, Fingís, entre otros.
- **Infraestructura:** Soporta principalmente la infraestructura tecnológica de los sistemas de correos corporativos y configuraciones de redes. Entre las áreas de soporte más importantes se encuentran las siguientes: Exchange, DNS, DHCP.
- **Network (Redes):** Provee servicios de monitoreo los circuitos de LAN y WAN entre los diferentes sitios geográficos donde Intel opera.
- **Telecom:** Brinda servicio de monitoreo de las principales infraestructuras de telefonía y de sus equipos de comunicación (PBX).

La visión de este departamento establece que “Global System Monitoring provee liderazgo global para la administración de sistemas remotos a las operaciones de Intel a nivel mundial”.

En la Figura 1 se muestra un diagrama del flujo de manejo de incidentes del departamento Global System Monitoring. Donde MOM (Servicio instalado en los servidores) genera una alerta y ésta se envía a Impact Manager (consola que almacena las alertas). Si es una alerta valida el agente de Global System Monitoring la analizará y generará un ticket tipo incidente en Service Now, si el primer nivel (Global System Monitoring) no tiene una solución para la misma debe escalarla a su siguiente nivel de soporte lo más antes posible, si se da el caso de que Global System Monitoring tiene la solución efectiva, le dira a su cliente, en éste caso el dueño del servidor que el problema fue resuelto.

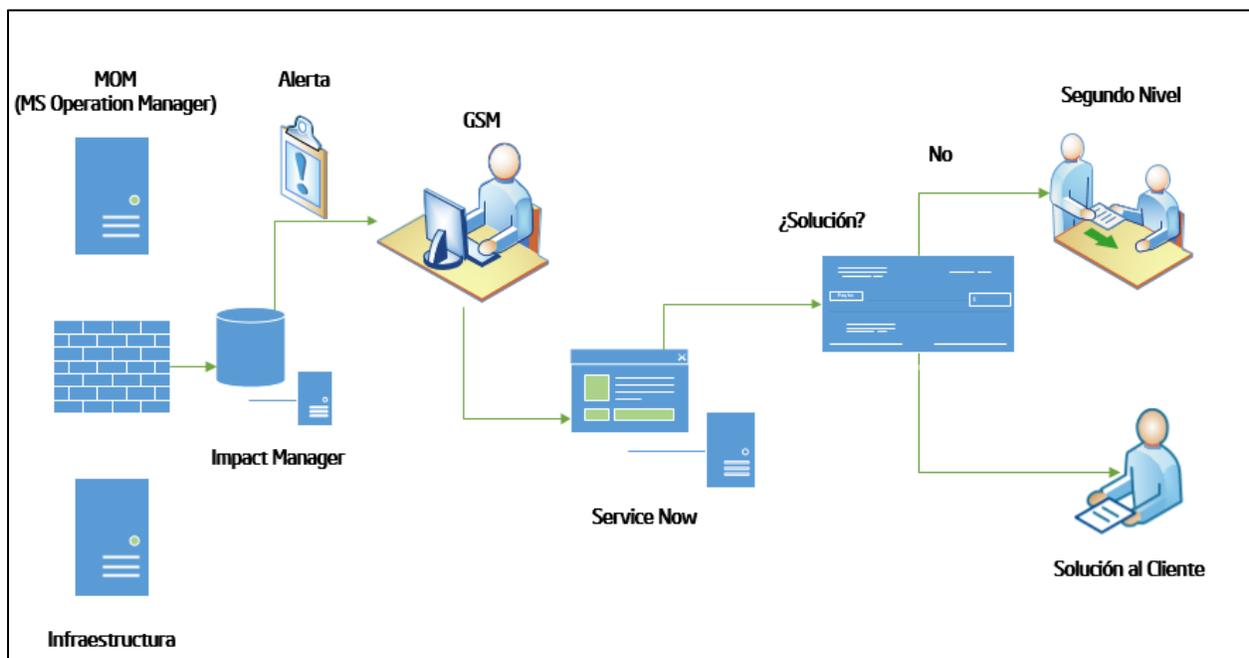


Figura 1 Flujo de manejo de incidentes en Global System Monitoring

La Figura 2 presenta el organigrama de Global System Monitoring. El mismo no se describe a detalle pues no es el objetivo, como si lo es la visualización de sus procesos. Nótese que Global System Monitoring soporta múltiples departamentos y el nivel de criticidad de estos departamentos son muy altos y complejos a la vez, como por ejemplo aquellos de los cuales dependen las operaciones de ensamble y muestra de chips desarrollados en Intel a nivel mundial. Por lo tanto la disponibilidad y efectividad de los equipos es altamente necesaria, siendo Global System Monitoring el grupo de soporte que pueda responder a las solicitudes de múltiples departamentos en términos de disponibilidad de los equipos.

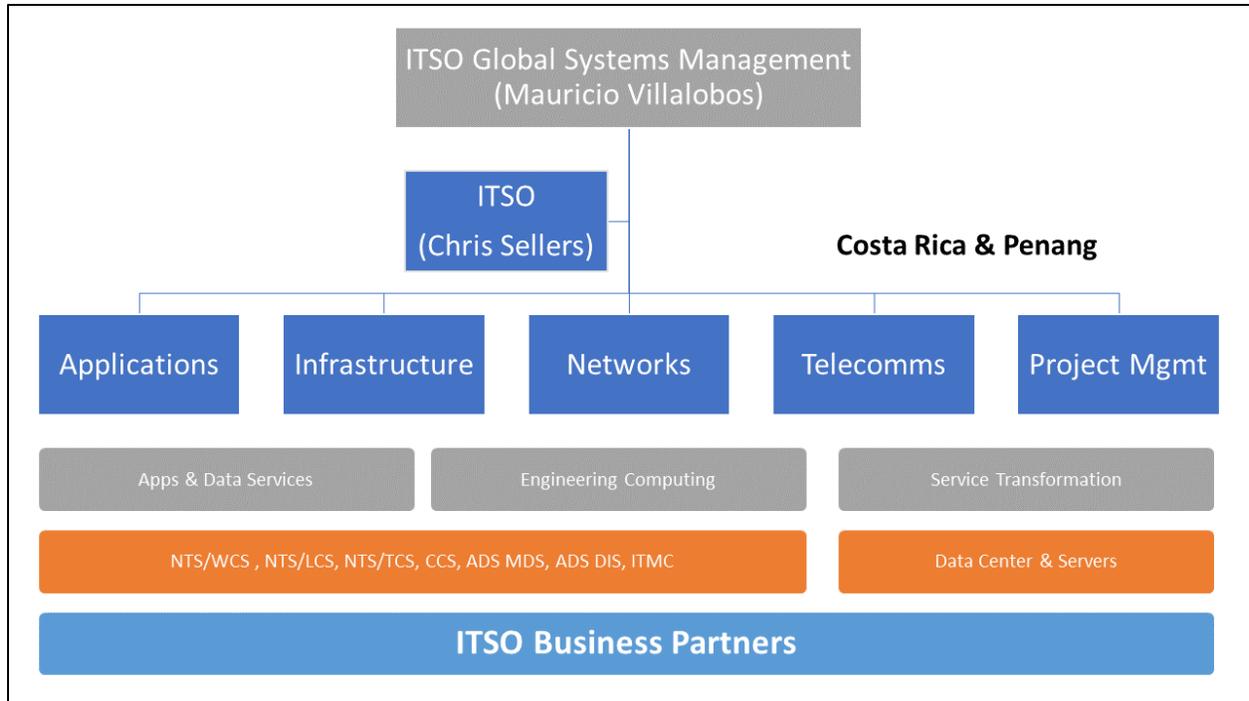


Figura 2 Organigrama Global System Monitoring

## Planteamiento del Problema

Actualmente el tiempo de resolución de los tickets es muy prolongado y no está cumpliendo con los niveles de servicio pactado con los clientes (Service level agreement), además al Global System Monitoring recibir diariamente un alto volumen de alertas, incide en brindar un servicio de baja calidad medido en los tiempos de respuesta no óptimos, poca resolución a problemas recurrentes, y predisposición a generar incidentes mayores por no atacar a tiempo alguna alerta recurrente en tiempo real, es decir un manejo reactivo de los eventos. Todos estos puntos llegan a generar un descontento generalizado en los clientes.

Cuando la calidad en la resolución de los problemas no es la esperada, se traduce en que un cliente vuelva a abrir otro ticket por el mismo problema o que el servidor vuelva a generar la misma alerta, lo cual llega a incrementar el volumen de los incidentes y alertas. Estas situaciones pueden estar ligadas a falta de entrenamientos o poca experiencia por parte de los agentes de primer nivel y a insuficientes acciones concretas y proactivas para la eliminación de recurrencias en incidentes y alertas.

## Problema General

La problemática actual en el departamento de Global System Monitoring en Intel, radica principalmente en la falta de instrumentos que permitan identificar la recurrencia de eventos generados en tiempo real por equipos monitoreados por Global System Monitoring, provocando en muchos casos que múltiples representantes de Global System Monitoring estén atendiendo el mismo caso generando duplicación de tareas.

### *Sub-Problemas*

- La información de eventos generados en la consola de monitoreo en Global System Monitoring son analizados reactivamente, lo cual produce que en muchas ocasiones ciertas áreas funcionales en Intel se vean impactados por no haberse anticipado en el momento oportuno.
- Falta de herramientas de visualización que le permitan a Global System Monitoring identificar de una forma expedita la recurrencia de eventos de los servidores en Intel generadores de mayor ruido.
- El análisis de los eventos generados en Global System Monitoring no son basados en tiempo real lo cual genera que la gerencia de Global System Monitoring tome decisiones tardías en la gestión apropiada de eventos e incidentes.
- Global System Monitoring requiere de un medio para centralizar información de diversas fuentes: nombre de la aplicación, datos del servidor, registro de incidentes y problemas, alertas, rendimiento de servidores, espacio en disco, parches, entre otros, que le permitan interrelacionar datos y de allí obtener análisis de posibles situaciones a presentarse en los servidores.

## Objetivos

### Objetivo General

Analizar la información procedente de tiquetes y alertas del departamento Global System Monitoring de la Empresa Intel, Costa Rica, con el fin de identificar los patrones de incidentes y alertas más significativos

por medio de una técnica de minería de texto, visualizándola en tiempo real mediante la implementación de un Data Warehouse 2.0 .

### Objetivos Específicos

1. Definir las fuentes de información a analizar clasificándolas por el tipo de alerta.
2. Revisar la veracidad de los datos obtenidos a partir de los métodos de minería de texto aplicados.
3. Implementar una solución de Data Warehouse 2.0.
4. Usar un estándar para aplicar calidad de datos.
5. Comparar los métodos de minería de texto que permitan el análisis de datos variados.
6. Preparar un modelo de reporte para el manejo de problemas recurrentes a partir de datos obtenidos.
7. Organizar los datos obtenidos en busca de tipos específicos de patrones o tendencias.

### Alcances

El proyecto desarrollado tiene como alcance la aplicación de un modelo de minería de texto que permita analizar la información procedente de tiquetes y alertas del departamento Global System Monitoring de la Empresa Intel en Costa Rica, con el fin de identificar los patrones de incidentes más significativos para su posterior resolución y análisis de tendencias.

Los aspectos puntuales que comprende el presente proyecto están relacionados a la extracción de los datos (incidentes y alertas) provenientes de la Base de Datos de Intel conocida con el nombre de “Service Now” que almacena información de los incidentes y de Impact Manager, la cual contiene las alertas y eventos asociados a los servidores de la empresa.

Esto permitirá realizar tareas de clasificación y limpieza de datos apropiadas para así poder aplicar la técnica de minería de texto más propicia que permita analizar la recurrencia de patrones de texto en alertas e

incidentes con el fin de identificar áreas de mejora en términos de capacitación a los niveles inferiores de soporte (Nivel 1) y proveer una propuesta estándar de resolución y estudio de los principales incidentes generados, basado en el cruce de información de diversas fuentes relacionadas como lo son la base de datos de empleados, manejo de problemas, parches, rendimiento y espacio en disco del servidor, entre otras.

Los datos obtenidos serán procesados en un Data Warehouse 2.0, donde se abarcarán 3 sectores: el interactivo (datos menores o iguales a una hora), el integrado (datos iguales o menores a un mes pero mayores a una hora) y el de archivo (datos mayores o iguales a un mes pero menor a un año).

Para mostrar la información de cada sector se utilizarán reportes en Reporting Services, dashboards por medio de PowerView, la herramienta para publicar la reportería será SharePoint Business Intelligence 2013. Asimismo se otorgará la facilidad de poder conectarse a los datos por medio de Excel.

No serán contemplados el análisis de incidentes y alertas de otros departamentos fuera del departamento “Global System Monitoring Costa Rica y Malasia”, además no están autorizados desembolsos adicionales para la compra de programas de cómputo fuera de los que existen hoy en día en Intel.

## Limitaciones

- La información utilizada en el presente proyecto se limita solamente a aquella que son generadas por las herramientas relacionadas y relevantes a Global System Monitoring, por lo tanto no se tomará en cuenta cualquier otra herramienta o fuente de datos que no esté citada en el presente documento.
- Para el procesamiento de los datos se utilizarán solamente aquellos que cumplan con los criterios necesarios para poder hacer las interrelaciones necesarias, por ejemplo, si un tiquete no posee un nombre válido de servidor, el mismo no podrá ser tomado en cuenta ya que no podrá ser relacionado con las otras fuentes de datos.
- Los datos expuestos en el presente proyecto son clasificados como críticos, por lo cual se requiere alta confidencialidad en los datos presentados.
- El presente proyecto se limitará en esta fase a la creación de la solución del Data Warehouse 2.0 y la visualización de los aspectos más relevantes que el negocio desee analizar. Las demás consideraciones o requerimientos adicionales se contemplarán en etapas futuras definidas como parte de la mejora continua del actual proyecto.

- El software y otras herramientas requeridas para la elaboración del presente proyecto se basaran en las propuestas y estándares ya definidos por Intel, no habrá presupuesto definido para la compra de software adicional.

# Capítulo II: Marco Metodológico

## Tipo de Investigación

En el presente proyecto se utilizará el enfoque cualitativo basado en Investigación en Acción pues permite resolver problemas cotidianos e inmediatos y mejorar prácticas concretas basadas en problemas prácticos vinculados a un entorno en particular, donde se propicia el cambio y además motiva a las personas a participar en la detección de necesidades.

Además, se contemplará el presente proyecto con el enfoque cuantitativo del Método Correlacional basado en Proceso de Investigación que utiliza instrumentos de recogida de datos para agruparlos, analizarlos y finalmente emitir un informe donde interprete todos los aspectos relacionados con el problema (Naranjo Zeledón, 2014).

Una vez identificado el marco de investigación a ser aplicado en el presente proyecto se procederá a implementar técnicas de minado de texto sobre los incidentes generados en la herramienta de Gestión de Incidentes llamada “Service Now” institucional e Impact Manager para el manejo de alertas. Esto permitirá identificar los problemas más frecuentes materializados por escalaciones de los incidentes que el Primer Nivel de Soporte (Nivel 1) no está apto a resolver por falta de capacitación o por un volumen alto de generación de incidentes en poco tiempo, que pueden ser recurrentes pero no han sido identificados por el modelo de trabajo actual, en el cual cada agente resuelve aleatoriamente un incidente o alerta sin poder darse cuenta de un patrón de comportamiento repetitivo .

Para ello, se utilizará como fuente de información la base de datos de “Service Now” e Impact Manager las cuales registran todos los incidentes y alertas generadas por los sistemas informáticos de la empresa y relacionados al departamento Global System Monitoring en Costa Rica.

Service Now está considerada como el principal sistema de información de tiquetes de mayor utilización, siendo una de sus características distintivas, frente a otras bases de datos, el utilizar una indexación manual realizada por usuarios expertos para la recuperación de información.

Para identificar los valores más relativos en esta base de datos, se elaborará una estrategia de búsqueda de términos relevantes y recurrentes de la propia base de datos, la cual es altamente fiable, y que se actualiza constantemente recogiendo todas las posibles variantes y sinónimos de los términos empleados. Se realizó una selección de la terminología más apropiada, identificándose los términos que comprenden el contexto de incidentes relacionados al departamento de Global System Monitoring.

Para identificar todos los posibles problemas relacionados al departamento antes mencionado será necesario un paso previo de limpieza de los datos (descripción de los incidentes y alertas) en el que se utilizará los distintos términos relacionados para una identificación óptima. En total se colectaran todos los incidentes y solicitudes de servicio generados durante el 2014 y Enero, Febrero, Marzo y Abril del 2015 en el departamento Global System Monitoring en Intel.

Una vez que toda la información sea colectada y clasificada de la mejor forma mediante medios de limpieza de datos, se buscarán técnicas de minado de texto para la creación de redes de palabras-clave en función de su recurrencia, con el fin de determinar patrones específicos en términos de incidentes y solicitudes de servicio que nos permita identificar las posibles acciones correctivas. Estas redes nos permitirán establecer relaciones de cercanía entre las palabras contenidas en la descripción de los incidentes para determinar relaciones de semejanza entre ambas.

## Área de Estudio

En el presente estudio se enfocará en las alertas e incidentes reportados en las consolas de monitoreo del departamento GSM (Global System Monitoring) en Costa Rica generados por equipos tecnológicos de Intel a nivel mundial y que son registrados en las Bases de Datos de Service Now por parte del personal de Global System Monitoring en las instalaciones de Intel en Costa Rica.

## Objeto y Sujeto de estudio

El principal objeto de estudio del presente proyecto se basa en el análisis del comportamiento de los incidentes y alertas generados por el mal funcionamiento o eventos inesperados en la infraestructura tecnológica de servidores en Intel. Se pretende buscar tendencias, patrones comunes y análisis del contenido

de texto provenientes en las descripciones de dichos incidentes, que permitan una adecuada visualización e interpretación de las áreas encontradas.

Por otro lado, el sujeto de estudio se basa en los “servidores” e “incidentes” que son los componentes determinantes del presente análisis, ya que contienen los elementos de gran relevancia en el estudio en cuestión. Se pretende analizar los servidores que generen eventos recurrentes, condiciones actuales de los servidores, así como el análisis de texto de las descripciones de los incidentes, de tal forma que permitan poner en evidencia el comportamiento normal o anormal de los servidores que generan dichos incidentes y que son registrados por el personal de Global System Monitoring.

## **Población y muestra**

La población en estudio se compone de todos los servidores que componen la infraestructura tecnológica de Intel que son soportados por el departamento de Global System Monitoring, que generan tanto incidentes y problemas registrados en “Service Now”, como alertas en Impact Manager.

El sistema de gestión de incidentes conocido como “Service Now” tiene aproximadamente alrededor de 6 años de haberse implementado como una solución general de ITIL, para el presente análisis se recuperará datos desde el 2014 hasta la fecha, por lo cual la cantidad de registros analizados estará delimitada al tiempo definido anteriormente debido a que se pretende analizar los incidentes en tiempo real.

Es importante recalcar que no se integrará eventos o incidentes de otros departamentos fuera de Global System Monitoring que pudiera potencialmente tener registros relevantes para el presente análisis, lo cual se establece como una limitante inicial del presente proyecto.

## **Fuentes de Información**

La fuente primaria del presente estudio se basará en la Base de Datos de “Service Now”, la cual registra todos los eventos e incidentes generados por los servidores de Intel y que son generados en las consolas de monitoreo de GSM (Global System Management), además se identificaron las siguientes fuentes:

- **BEM Impact Manager:** consola para el manejo de alertas generadas por MOM (servicio de Microsoft que genera eventos de los servidores)
- **iMODS:** Muestra detalle del rendimiento y espacio en disco de los servidores.
- **CDIS:** Maneja la base de datos de los empleados de Intel, esto para su correlación con los datos del servidor para poder determinar el nombre del dueño del mismo.
- **IAP:** Aloja la información relacionada con las aplicaciones existentes en Intel, asociadas en su mayoría a algún servidor.

## Diseño de técnicas e instrumentos

La presente investigación se apoyará en la utilización de técnicas de minería de texto para analizar las tendencias de los incidentes y alertas mediante la recuperación de registros en tiempo real, a través de la implementación de un Data Warehouse 2.0 que permita recuperar datos actualizados a fin de mostrar los datos con mayor relevancia para la Gerencia del Departamento de GSM. Por otro lado, se permitiría la realización de otros tipos análisis de los incidentes derivados de la solución tabular que se construirá, a fin responder a las necesidades de Negocio.

Para la realización de estos análisis, se utilizará una técnica de Minería de Texto basadas en un Data Warehouse 2.0. Para el proyecto en general, se utilizaran las siguientes herramientas:

- **SQL Server Data Tools 2013 y SQL Server Integration Services (SSIS):** Para los algoritmos de transformación de “Extracción de Términos”
- **SQL Server Datar Tools 2013 y SQL Server Analysis Services (SSAS):** para el desarrollo del Proyecto Tabular del Data Warehouse 2.0.
- **SQL Server 2014 – SQL Server Management Studio:** como repositorio del Staging Area y para soluciones técnicas en general necesarias de la implementación de la solución propuesta en el presente proyecto.
- **Sharepoint Business Intelligence 2013:** para la creación de dashboards mediante PowerView y la publicación de reportes creados en Reporting Services.
- **Excel 2013:** para la utilización de la solución estilo “*self-service*” por parte de los usuarios finales.

## Estado de la Cuestión

Para la realización del Estado de la Cuestión se utilizó el Proceso de mapeo sistemático por “Kai Pettersen” (Petersen, 2008) sobre el tema “Uso de la minería de texto para el análisis de incidentes en el departamento Global System Monitoring en Intel” las cuales serán introducidas como elementos claves al inicio de cada párrafo como elemento introductorio de cada proceso.

El estado de la cuestión alrededor del tema de la “Aplicación de un modelo de minería de texto que permita analizar la información real procedente de tiquetes” en el departamento Global System Monitoring en Intel nos lleva a la siguiente pregunta: ¿Cuáles herramientas de minería de texto existentes en el mercado puedan ayudar a la identificación de patrones recurrentes en la descripción de los incidentes generados por medio de la herramienta Service Now e Impact Manager en el departamento Global System Monitoring de Intel?

El alcance actual nos lleva al análisis específico del Departamento Global System Monitoring de Intel, donde se analizarán específicamente los incidentes, alertas y solicitudes de servicio (service request) almacenados en la Base de Datos de Gestión de Incidentes, que a su vez nos permita analizar tendencias específicas que permitan resolver problemas recurrentes de forma eficiente, con el fin de tomar decisiones adecuadas para mejorar el nivel de servicio para los clientes del departamento Global System Monitoring y para los grupos de primer nivel (L1).

A nivel de contexto inicial es importante definir los conceptos esenciales de la presente investigación, se puede decir que la minería de texto “Es un análisis de los datos compartidos por todos los textos de la colección que se ofrece de manera indirecta, es decir, son informaciones que la colección dará a los especialistas pero que no fue específicamente incluida en esa colección en el momento de su creación para su posterior difusión a los usuarios.” Sobre Data Warehouse 2.0 podemos decir que nos muestra una perspectiva en tiempo real de los datos de negocio”. (García, 2014) y sobre el tema de Gestión de Incidentes tenemos que “...Tiene como objetivo resolver cualquier incidente que cause una interrupción en el servicio de la manera más rápida y eficaz posible” (Kempter, 2014). Dentro de las palabras claves se encuentran las siguientes: Data Warehouse 2.0, Minería de Texto, Gestión de Incidentes y Solicitud de Servicios (Service Request), algoritmos de minería de texto, modelos de minería de texto.

Según David Parcell “Se estima que hoy en día el 80% de las fuentes de las empresas son provenientes de fuentes de formato en texto, voz y video” (Parcell, 2012), lo cual hace difícil la interpretación y análisis de dichos datos de forma propicia, con el fin de poder descifrar las tendencias y toma de decisiones adecuadas.

Por lo tanto es de suma importancia determinar las técnicas de minería de texto adecuadas que permitan explorar y definir patrones de gran relevancia de los incidentes generados por la infraestructura de Global System Monitoring en Intel.

Basándose en las fuentes bibliográficas encontradas, se ofrecen escenarios reales de la aplicación de Minería de texto permitiendo así a los representantes del Departamento de Global System Monitoring a tomar decisiones instantáneas e idóneas para resolver en menor tiempo los incidentes más recurrentes y descifrar elementos claves para el manejo del problema planteado en el presente proyecto.

La principal hipótesis propuesta en esta investigación, se basa en que a mayor identificación y control de los incidentes más recurrentes en el departamento de Global System Monitoring habrá una mayor reducción de las escalaciones y el nivel de satisfacción de los clientes se incrementará debido a que habrá menos interrupciones en los servicios de Tecnologías de Información. Donde se podrá conocer a plenitud los principales aspectos de los incidentes mayoritariamente generados.

Aunado a esto se pretende proveer una plataforma donde a futuro se pueda ejercer la gestión de problemas, al identificar recurrencias en cantidad de alertas, incidentes y solicitudes de servicio proveniente de los servidores. Actualmente se asocian números de problemas a los incidentes, sin embargo no siempre se le da la debida continuación. A partir de ello se desea mostrar mediante reportes ésta información y así recomendar al negocio darle la debida continuación a los mismos.

Según la definición de ITIL (Foundation, s.f.), la gestión de problemas tiene como función mejorar el funcionamiento de la infraestructura de Tecnologías de Información, donde es imprescindible realizar un seguimiento continuo de los problemas basados en tendencias. Dentro de sus principales objetivos tenemos:

- Disminución del número de incidentes y una rápida resolución de los mismos.
- Mayor eficacia en la resolución de problemas.
- Gestión proactiva, que permita identificar problemas potenciales antes de que éstos se manifiesten o provoquen una seria degradación de la calidad del servicio.

## Situación Actual

### Capacidad actual en el análisis de la información

Hoy en día el departamento de Global System Monitoring no cuenta con una herramienta de visualización de incidentes en tiempo real y no se cuenta un análisis de las descripciones de los incidentes. Únicamente se cuenta con sistemas transaccionales que registran los incidentes (Impact Manager, Service Now). A continuación se presenta una breve descripción de las fuentes de información que dispone el departamento de Global System Monitoring:

#### *Service Now*

Es una base de datos de gran importancia para Intel y en Global System Monitoring ésta no es la excepción. Por medio de este repositorio se permite registrar el flujo de incidentes generados en la gestión de servicio en Intel y basándose en las mejores prácticas de ITIL.

Service Now posee múltiples tablas para la operación normal, de las cuales se tomarán aquellas tablas que sean de mayor interés para el presente proyecto, entre las que se encuentran: Gestión Incidentes, Solicitudes de Servicio, Gestión de Problemas, Artículos de Conocimiento, Nombres de Servidores identificados mediante el campo: Configuración del Item y nombre de Aplicaciones.

Ésta plataforma es de suma importancia para Global System Monitoring (Nivel 1) ya que permite gestionar los incidentes de forma centralizada, permitiéndole seguir con el flujo de acción de los incidentes por resolver y traquear el historial de los mismos.

#### *Impact Manager*

Es la base de datos que registra eventos (alertas) generados por equipos de Intel y que son generados por agentes de monitoreo (MOM- Microsoft) que son instalados en los equipos tecnológicos de Intel. Esta base de datos es relevante para Global System Monitoring (Nivel 1), ya que les permite mostrar en la consola de monitoreo aquellas alertas que son relevantes y que necesitan pronta acción. Estas alertas son generadas en tiempo real.

### Requerimientos esperados para el análisis de la información

La solución a implementar ha utilizado las siguientes técnicas para la recolección de requerimientos.

1. **Entrevistas y encuestas a los interesados:** Encargado del departamento de manejo de eventos, representantes de los Servicios de Soporte, representantes del manejo de Base de Datos del sistema Service Now, representantes del sistema Service Now.
2. **Sesiones de trabajo:** para entender los procesos que involucra el proyecto y definición de requerimientos.
3. **Grupos de enfoque:** donde se realizan discusiones interactivas con los principales involucrados del proyecto.

Mediante éstas técnicas se pretende entender las necesidades del negocio, a su vez ir conociendo con cuáles datos se cuenta para ir construyendo el Data Mart, a partir del cual se diseñara la solución. A continuación se mostrarán los requerimientos principales:

Gerencia ITIL (Incidentes) INTINC– 001 (Incidentes Históricos y en Tiempo Real)	
<b>Áreas</b>	Gerencia de ITIL, Gerencia TAC
<b>Responsables del requerimiento</b>	Gerente ITIL TI
<b>Nombre del requerimiento</b>	INTINC– 001. Cuantificar la cantidad de incidentes y eventos generados por la infraestructura de Intel en tiempo real por medio del Data Warehouse 2.0 y además mantener un histórico disponible para la identificación de tendencias
<b>Descripción detallada del requerimiento</b>	Es necesario identificar la cantidad de incidentes y eventos generados en tiempo real con el fin de tener detalle de los incidentes recurrentes en tiempo real y a nivel histórico.

Gerencia ITIL (Incidentes) INTINC– 002 (Incidentes Recurrentes)	
<b>Áreas</b>	Gerencia de ITIL, Gerencia TAC
<b>Responsables del requerimiento</b>	Agentes de Primer Nivel Gerente ITIL TI
<b>Nombre del requerimiento</b>	INTINC– 002. Identificar los eventos e incidentes más recurrentes generados en la herramienta Service Now.
<b>Descripción detallada del requerimiento</b>	Se necesita identificar aquellos eventos recurrentes en la infraestructura de Intel tanto a nivel histórico como en tiempo real, con el fin de realizar un proceso de identificación y clasificación de dichos eventos, que a su vez permita tomar decisiones precisas para cada situación en particular

	Además, se contará con la cantidad de eventos históricos que permitan realizar análisis de tendencia de los mismos.
--	---

Gerencia ITIL (Incidentes)	
INTINC– 003 (Generación de Sistemas de Alerta por Incidentes Recurrentes)	
<b>Áreas</b>	Gerencia de ITIL, Gerencia TAC
<b>Responsables del requerimiento</b>	Gerente ITIL Agentes de Primer Nivel TI
<b>Nombre del requerimiento</b>	INTINC– 003. Generar un sistema de alerta para aquellos eventos e incidentes más recurrentes generados en tiempo real.
<b>Descripción detallada del requerimiento</b>	Se necesita alertar a los agentes de Primer Nivel de soporte sobre aquellos eventos recurrentes que se generen en tiempo real, esto con el fin de facilitar el proceso de toma de decisiones y evitar que dichos eventos actuales generen un mayor impacto. Se investigará sobre aquellas herramientas que permitan ser integradas con Data Warehouse 2.0 para la generación de dichas alertas sobre eventos recurrentes.

Gerencia ITIL (Incidentes)	
INTINC– 004 (Análisis de la descripción de los eventos e incidentes)	
<b>Áreas</b>	Gerencia de ITIL, Gerencia TAC
<b>Responsables del requerimiento</b>	Agentes de Primer Nivel Gerente ITIL TI
<b>Nombre del requerimiento</b>	INTINC– 004. Implementar técnicas de análisis de texto basadas en las descripciones de los eventos e incidentes.
<b>Descripción detallada del requerimiento</b>	Las descripciones de los tickets no poseen ningún formato estándar por lo cual es difícil recuperar manualmente cuales son los problemas más frecuentes. Por esta razón se pretende buscar técnicas de minado de texto, la cual otorguen la posibilidad de mostrar a los usuarios aquellas palabras claves que podrían considerarse como relevantes para el análisis de los incidentes recurrentes.

## **Análisis de la brecha de estados: actual vs deseado**

Como se mencionó anteriormente, el departamento Global System Monitoring cuenta con sistemas transaccionales (OLTP) que le permite únicamente realizar un monitoreo reactivo sobre los incidentes que reciben en sus consolas de monitoreo.

Una vez realizado el análisis de las reglas de negocio, de los sistemas transaccionales y de las estructuras de datos existentes, se deberá proceder a realizar la construcción del Data Warehouse 2.0, mediante la integración de los elementos más representativos para el departamento de Global System Monitoring y la posterior implementación del algoritmo de minería de texto sobre las descripciones de los incidentes y alertas recibidos en la consola de monitoreo de Global System Monitoring, con el fin de generar resultados y tomar posteriores decisiones únicamente con el fin de mejorar el servicio brindado a sus clientes. Ésta solución proveerá posteriormente reportes para la visualización de la información encontrada y se podrán realizar lecturas directamente al Data Warehouse 2.0 construido.

## **Análisis de la información deseada**

La implementación del Data Warehouse 2.0 y el posterior análisis de los incidentes y alertas a través de la aplicación de algoritmos de Minería de Texto tienen el único fin de generar decisiones que le permitan al departamento de Global System Monitoring actuar proactivamente y no de manera reactiva (como se ha realizado hasta la fecha).

El escenario ideal que se pretende lograr es el de generar un sistema de reporte en tiempo real de las tendencias de los eventos e incidentes generados por los sistemas de monitoreo implementados en la infraestructura tecnológica de Intel, y la potencial identificación de factores recurrentes que generen incidentes en un futuro, con el fin de minimizarlos en su debido tiempo.

El objetivo primordial de la aplicación de la técnica de minería de texto sobre incidentes y solicitudes de servicio, es de poder llegar a identificar e interpretar aquellos patrones recurrentes en las descripciones de

dichos eventos, con el fin de determinar comportamientos específicos en los incidentes que serán analizados. El propósito de identificar este conocimiento le permitiría a Global System Monitoring poder accionar de forma proactiva ante potenciales incidentes e informar de forma apropiada a los dueños de los equipos sobre potenciales fallas que afecten el comportamiento normal de sus sistemas.

Además la Minería de Texto se utiliza en la identificación de relaciones entre palabras significativas encontradas en el campo de descripción de los incidentes, lo cual permitiría generar un nuevo conocimiento a partir de las múltiples “descripciones” que inicialmente podrían creerse que no tienen relación alguna entre sí. El fin de este tipo de análisis radicaría en la generación de deducciones a partir de conexiones ocultas entre distintas líneas de “descripciones” de los incidentes.

# Capítulo III: Desarrollo de la Solución

## Análisis del Negocio

Para entender mejor las reglas de negocio del departamento Global System Monitoring se realizaron entrevistas con los líderes del departamento para traducir los requerimientos de este departamento en una solución propicia, con el fin de identificar de primera mano los incidentes y alertas recurrentes y áreas de acción específicas que le permitan al departamento de Global System Monitoring ofrecer un servicio proactivo y de calidad a sus clientes.

Basados en los requerimientos provistos se ha determinado que este esfuerzo es parte de un proyecto a largo plazo que se desarrollará a lo largo del año, por lo cual nos enfocaremos en la aplicación de técnicas de minería de texto para extraer palabras claves de gran importancia en el campo de “Descripción” de cada incidente y solicitud de servicio; además de la visualización de los equipos que generan mayor recurrencias y gráficas que identifiquen la salud del servidor al interrelacionar diversas fuentes de información como parches por ejemplo.

## Selección de Técnicas para la Implementación del proyecto

### Arquitectura Data Warehouse 2.0

El presente proyecto utilizará las ventajas ofrecidas por el Data Warehouse 2.0. En este punto cabe mencionar que el mismo está basado en la metodología de Bill Inmon del Data Warehouse. Data Warehouse 2.0, consiste en la evolución del modelo tradicional del Data Warehouse a un modelo más adaptable y eficiente. Conforme se extiende el uso de Data Warehouse dentro de la organización se van presentando situaciones a resolver sobre la marcha por ejemplo, el gran crecimiento de la información, la necesidad de integrar información no estructurada, necesidad de nuevas formas de explotación, entre otros, para estas situaciones principalmente está basada la arquitectura del Data Warehouse 2.0. (Inmon, 2014).

Los metadatos son la clave de la arquitectura Data Warehouse 2.0, esto permite la adaptabilidad e integración de información estructurada y no estructurada, así mismo, se clasifican diferentes niveles de metadatos, para cubrir todas las necesidades desde el mismo Data Warehouse, en el modelo tradicional los metadatos se administraban en cada herramienta que explotaba la información, ahora es parte del Data Warehouse y permitirá la fácil adaptabilidad de cualquier herramienta de explotación (Kobielus, 2008).

Entre las ventajas ofrecidas por la arquitectura de Data Warehouse 2.0 que benefician al presente proyecto se encuentran las siguientes:

- Contar con datos a un nivel bajo de detalle.
- Unir datos estructurados y no estructurados.
- Soportar diferentes tipos de procesamiento, sin afectar el rendimiento.
- Soportar cambios en los datos a través del tiempo, mediante la existencia de sectores según las necesidades del negocio.
- Positiva evolución y crecimiento del modelo actual (Data Warehouse tradicional) y no será necesario una reingeniería del Data Warehouse.

### **Modelo Tabular (SSAS)**

Será desarrollado un Modelo Tabular utilizando Data Warehouse 2.0, conformando cada uno de los sectores mediante particiones en el modelo: Interactivo (datos menores o iguales a una hora), Integrado (Datos mayores a una hora pero menores a un mes) y el de Archivo (datos mayores a un mes pero menores a un año).

Los modelos tabulares son bases de datos “en memoria” de Analysis Services, utilizan algoritmos de compresión avanzados y un procesador de consultas multiproceso; el motor analítico en memoria xVelocity (VertiPaq) ofrece un acceso rápido a los objetos y los datos de los modelos tabulares para aplicaciones cliente de reportes como Microsoft Excel y Microsoft Power View (Microsoft, Modelo Tabular, 2014).

Los modelos tabulares admiten el acceso a los datos mediante dos modos: modo de almacenamiento en caché y modo DirectQuery. En el modo de almacenamiento en caché, puede integrar datos de varios orígenes como bases de datos relacionales, fuentes de distribución de datos y archivos de texto planos. En el modo DirectQuery, puede omitir el modelo en memoria, lo que permite a las aplicaciones cliente

consultar los datos directamente en el origen relacional, por ejemplo SQL Server. (Microsoft, Analysis Services, 2014).

### *Ventajas del modelo tabular*

A continuación se mencionan las principales ventajas de los modelos tabulares, tomadas en cuenta al seleccionar un modelo tabular en lugar de un modelo multidimensional.

- Al almacenar los datos en memoria, ofrece mayor velocidad en las consultas.
- No requiere generar Agregaciones<sup>3</sup> por lo que se simplifica el tiempo de procesamiento y se ahorra espacio en disco.
- DAX<sup>4</sup> (el lenguaje para acceder a los datos equivalentes al MDX<sup>5</sup>), tiene mayor flexibilidad para obtener información, al ser un lenguaje más familiarizado con la forma de trabajo de las hojas de cálculo de Excel.
- Su estructura es intuitiva por lo cual es más rápido y sencillo de entender, implementar y mantener.
- Se basa en modelos relacionales.

### *Problemas con el modelo tabular*

Los modelos tabulares de igual forma presentan algunas desventajas respecto a los modelos multidimensionales, dentro de éstas tenemos las siguientes:

- Las particiones no se procesan en paralelo si no secuencialmente, lo cual ocasiona un procesamiento más lento en comparación a un modelo multidimensional.
- Los modelos multidimensionales permiten crear perspectivas para realizar traducciones de los campos deseados a idiomas en específico, en los modelos tabulares no se pueden usar múltiples idiomas.

---

<sup>3</sup> Una agregación es un resumen de datos pre calculados almacenados en disco, con el fin de mejorar el tiempo de respuesta. (Microsoft, Developer Network, 2014)

<sup>4</sup> DAX: Data Analysis Expressions, lenguaje de fórmulas utilizado para crear cálculos personalizados sobre datos de tablas y columnas. (Microsoft, Developer Network, 2014)

<sup>5</sup> MDX: Multidimensional Expressions, lenguaje utilizado para crear expresiones dentro de un cubo. (Microsoft, Developer Network, 2014)

- Si el volumen de datos es bastante alto, los tabulares tardan un tiempo considerable en manejar configuraciones de diferentes particiones.
- El modelo tabular, al procesarse en memoria retiene excesiva memoria **RAM** y a su vez es dependiente de ella, de tal manera que puede llegar a afectar otras aplicaciones.

### Minería de Datos: Minería de Texto

Uno de los objetivos importantes del presente proyecto es la identificación de patrones recurrentes en los campos de descripción de cada Incidente y alerta respectivamente, razón por la cual es de suma importancia la aplicación de técnicas de Minería de Texto y la relación con Minería de Datos como elemento introductorio.

La Minería de Datos se ha definido como "la integración de un conjunto de áreas que tienen como propósito la identificación de un conocimiento obtenido a partir de las bases de datos que aporten un sesgo hacia la toma de decisión". Entendido lo que es la Minería de Datos, podemos extrapolar la misma idea a la Minería de Textos. Los datos a tratar con esta técnica serán, en lugar de los datos de las bases de datos, los documentos y textos de las organizaciones, administraciones, compañías (Textmining, 2014).

La Minería de Texto no se debe confundir con la recuperación de la información, que es la recuperación automática de documentos relevantes mediante indexaciones de textos, clasificación, categorización. La información que realmente le interesaría a la minería de textos es aquella contenida en esos documentos pero de manera general, es decir, no está contenida en un texto en concreto sino que es la información global que tienen todos los registros, textos, documentos de la colección en común. Es un análisis de los datos compartidos por todos los textos de la colección que se ofrece de manera indirecta, es decir, son informaciones que la colección dará a los especialistas pero que no fue específicamente incluida en esa colección en el momento de su creación para su posterior difusión a los usuarios. Por tanto, podemos decir que la Minería de Textos comprende tres actividades fundamentales (Statsoft, 2014):

1. Recuperación de información, es decir, seleccionar los textos pertinentes.
2. Extracción de la información incluida en esos textos: hechos, acontecimientos, datos clave, relaciones entre ellos, etc.
3. Por último se realizaría lo que antes definíamos como minería de datos para encontrar asociaciones entre esos datos claves previamente extraídos de entre los textos.

## Análisis de Herramientas de Minería de Texto

A continuación se detallan las herramientas de Minería de Texto disponibles en el mercado a fin de conocer las ventajas y desventajas de las mismas. A partir de ello se selecciona la mejor herramienta que se adecue a las necesidades del departamento de Global System Monitoring y que se alinee a los requerimientos de Intel.

### *SAS Text Analytics*

Esta aplicación incorpora las funcionalidades características de la minería textual, descrita en los apartados anteriores. Entre ellas:

- Capacidad de procesar documentos en distintos formatos (pdf, ascii, html, *Microsoft Office*, etc.), y extraer términos simples y compuestos (formados por más de una palabra), eliminar palabras vacías y reducir las palabras a lexemas. El sistema está optimizado para los idiomas inglés, francés y alemán.
- Identificar la función gramatical de la palabra en el texto con el objeto de evitar posibles ambigüedades (*part-of-speech tagging*).
- En el caso del idioma alemán, posibilidad para descomponer palabras en los distintos “lexemas” que la forman.
- Representación de documentos mediante un vector de términos ponderados según su frecuencia estadística.
- *SAS Text Miner* considera las tareas de categorización y agrupación automática como el principal objetivo de la minería textual; las actividades previas (identificación de términos simples y compuestos, extracción de nombres propios, ponderación, etc.) son tareas de pre-procesamiento, que deben realizarse preliminarmente para hacer posible la categorización y la clasificación.

### *Spss LixiQuest*

En este caso se trata de tres productos: *LexiQuest Mine*, *LexiQuest Categorize* y *Lexi-Quest Guide*. Los dos primeros ofrecen funcionalidad propia de la minería textual. El último consiste en tecnología de indexación y búsqueda de documentos en intranets e internet.

- *LexiQuest Mine* se presenta como una herramienta cuyo propósito es “automatizar el proceso de leer documentos para así descubrir su contenido”. En la presentación de esta herramienta se recurre al hecho de que, de un documento de treinta páginas, tan sólo tres párrafos pueden resultarnos relevantes. El problema se acentúa cuando tenemos que leer un gran número de documentos.
- *LexiQuest Mine* ofrece la posibilidad de procesar un gran número de documentos e identificar los términos y nombres propios que aparecen en ellos, así como mantener información sobre términos relacionados a partir de su co-ocurrencia.
- *LexiQuest Categorize* ofrece la función de categorización automática de documentos a partir de un entrenamiento previo (no se trata de clustering).

### *Megaputer TextAnalyst*

La aplicación funciona de la siguiente forma: a partir de un texto o colección de textos en formato ascii o rtf se identifican los principales términos, se le asigna una ponderación que representa en qué medida es significativo en la colección y en el texto procesado. Los términos pueden estar formados por dos o más palabras. Para identificar términos formados por varias palabras la aplicación también se basa en la frecuencia con la que aparecen juntas.

No se aplican reglas de tipo sintáctico o gramatical. A su vez, a cada par de términos se les asocia un valor que representa la relación existente entre ellos. Con esta información se forma una “red semántica”, que se visualiza mediante una estructura jerárquica. Cada concepto representa un nodo en el árbol. Sus nodos hijo representan los conceptos con los que está relacionado.

### *Microsoft SSIS – Term Extracction.*

Microsoft Integration Services es una plataforma para la construcción de la integración de datos de nivel empresarial y de datos transformaciones soluciones. Integration Services se utiliza para resolver problemas complejos de negocios mediante la copia o descarga de archivos, enviar mensajes de correo electrónico en respuesta a eventos, la actualización de los almacenes de datos, limpieza y minería de datos, y la gestión de objetos y datos de SQL Server. Los paquetes pueden trabajar solos o en armonía con otros paquetes para hacer frente a las necesidades empresariales complejas.

Esta herramienta pueden extraer y transformar datos de una amplia variedad de fuentes, tales como archivos XML de datos, archivos planos, y fuentes de datos relacionales, y luego cargar los datos en uno o más destinos.

La transformación “Term Extracction” extrae términos de texto en una columna de entrada de transformación, y luego escribe los términos a una columna de salida de la transformación. La transformación sólo funciona con texto Inglés y utiliza su propio diccionario de inglés y la información lingüística sobre inglés.

Se puede usar la transformación Extracción de términos para descubrir el contenido de un conjunto de datos. Por ejemplo, el texto que contiene los mensajes de correo electrónico puede proporcionar información útil sobre los productos, por lo que se puede utilizar la transformación Extracción de términos para extraer los temas de discusión en los mensajes, como una forma de analizar las votaciones.

Basándose en la anterior información, se elige Microsoft SSIS Term Extraction ya que es parte del suite de aplicaciones de Microsoft existentes dentro de Intel Corporation, lo cual facilita el trabajo con las herramientas ya utilizadas para la elaboración del Data Warehouse 2.0. Asimismo ésta herramienta cumple con la minería de texto deseada para la solución.

## **Análisis Técnico de las estructuras fuentes**

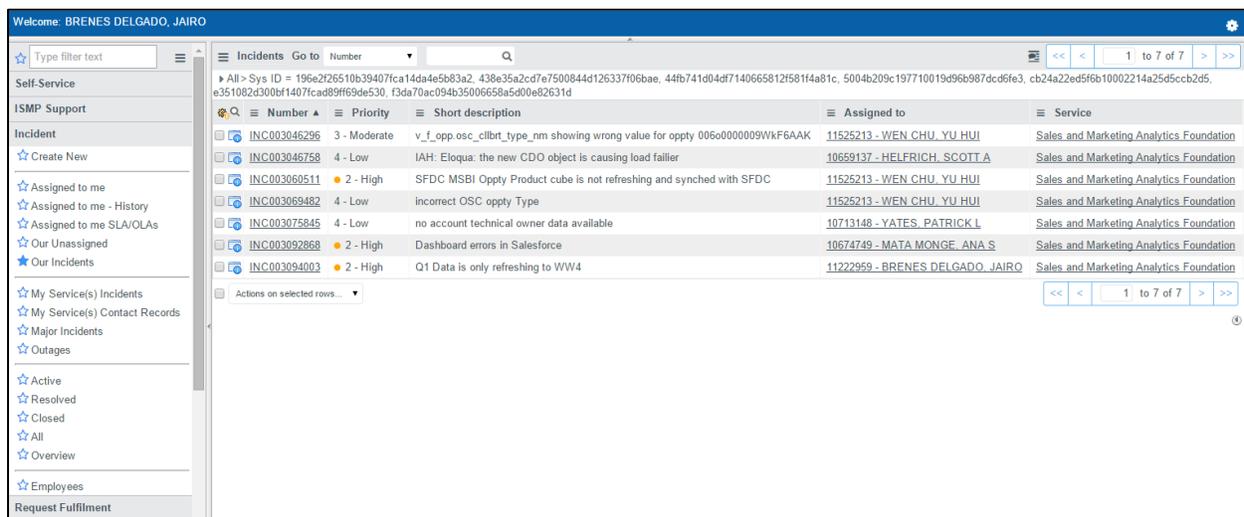
Los sistemas utilizados por Global System Monitoring, proceden de Bases de Datos transaccionales sobre MySQL (Service Now) y SQL Server (Impact Manager e iMODS). Se identificó que en algunos sistemas transaccionales posee información limitada a la hora de la recopilación de los datos, en especial en Service Now donde la información más reciente que se puede recuperar es de las últimas 8 horas. Por otro lado en los sistemas transaccionales de Impact Manager e iMODs se puede tener información en tiempo real lo cual permite tener datos de alertas e indicadores de rendimientos en tiempo real de los servidores.

Las principales fuentes de información provienen de los sistemas transaccionales utilizados por Intel hoy en día, los cuáles son relevantes para el Departamento de Global System Monitoring.

### *Service Now*

Como se ha detallado anteriormente, Service Now es una base de datos de gran importancia para Intel, por medio de este repositorio se permite registrar el flujo de incidentes generados en la gestión de servicio en Intel basados en las mejores prácticas de ITIL. Service Now posee múltiples tablas para la operación normal, de las cuales se tomarán aquellas tablas con mayor relevancia para el presente proyecto, entre ellas: Gestión Incidentes, Solicitudes de Servicio, Gestión de Problemas, Artículos de Conocimiento, servidores y nombres de Aplicaciones.

Se procedió a revisar el sistema transaccional de Service Now para determinar la información que se registra en el proceso de Gestiones de Eventos e Incidentes, por lo cual se solicitó acceso al sistema transaccional, por favor referirse a la Figura 3



Welcome: BRENES DELGADO, JAIRO

Incidents Go to Number

All > Sys ID = 196e2f26510b39407fca14da4e5b83a2\_438e35a2cd7e7500844d126337f06bae\_44fb741d04df7140665812f58114a81c\_5004b209c197710019d96b987dcdf63\_cb24a22ed5f6b10002214a25d5ccb2d5\_e351082d300bf1407f7cad89ff69de530\_f3da70ac094b35006658a5d00e82631d

Number	Priority	Short description	Assigned to	Service
INC003046296	3 - Moderate	v_f_opp osc_cllbrt_type_nm showing wrong value for oppty 0060000009WkF6AAK	11525213 - WEN CHU, YU HUI	Sales and Marketing Analytics Foundation
INC003046758	4 - Low	IAH: Eloqua: the new CDO object is causing load failler	10659137 - HELFRICH, SCOTT A	Sales and Marketing Analytics Foundation
INC003060511	2 - High	SFDC MSBI Oppty Product cube is not refreshing and synched with SFDC	11525213 - WEN CHU, YU HUI	Sales and Marketing Analytics Foundation
INC003069482	4 - Low	incorrect OSC oppty Type	11525213 - WEN CHU, YU HUI	Sales and Marketing Analytics Foundation
INC003075845	4 - Low	no account technical owner data available	10713148 - YATES, PATRICK L	Sales and Marketing Analytics Foundation
INC003092868	2 - High	Dashboard errors in Salesforce	10674749 - MATA MONGE, ANA S	Sales and Marketing Analytics Foundation
INC003094003	2 - High	Q1 Data is only refreshing to WW4	11222959 - BRENES DELGADO, JAIRO	Sales and Marketing Analytics Foundation

Actions on selected rows

Figura 3 Interfaz Service Now

La base de datos cuenta con una cantidad considerable de tablas y relaciones, las cuales permiten el registro de cada incidente durante todo su ciclo de vida y alineado con las mejores prácticas de ITIL. Este sistema permite registrar de forma completa el proceso de control de cada incidente y su respectivo flujo dentro de la metodología de ITIL. Entre los módulos de ésta metodología tenemos:

- Gestión de Incidentes
- Gestión de Solicitud de Servicios
- Gestión de Problemas
- Gestión de Cambio
- Gestión de Conocimiento

- Gestión de Lanzamiento
- Registro de Equipos Computacionales

Como se había mencionado anteriormente el presente proyecto se enfocará en el análisis de la Gestión de Eventos e Incidentes, tocando una parte preliminar de Gestión de Problemas a utilizar en el sector integrado y de archivo.

### BEM- Impact Manager

Es la base de datos que registra eventos (alertas) generados por equipos de Intel generados por agentes de monitoreo (MOM- Microsoft) instalados en los respectivos servidores. Esta base de datos es relevante para Global System Monitoring (Nivel 1), ya que les permite mostrar en la consola de monitoreo (Figura 4) aquellas alertas que son relevantes y que necesitan pronta acción.

The screenshot displays the BMC Impact Explorer interface. On the left, a tree view shows the hierarchy of monitored components, including 'CoreInfrastructure@cellmrd200'. The main pane shows a table of events with columns for Tier, Modified, Occurred, hostname, Message, Repeated, Last Modifier, trouble\_ticket, and CI. A specific event is highlighted in red, showing a critical alert at 4/8/15 11:02 AM from 'f28scorp202.corp.intel.com' with the message 'The Domain Controller's free disk is low on th...'. Below the table, the 'Event Information' section provides details for the selected event, including its severity (Critical), message, and classification (APP\_SQL\_ABS).

Tier	Modified	Occurred	hostname	Message	Repeated	Last Modifier	trouble_ticket	CI
1	4/8/15 6:01 PM	4/8/15 5:41 PM	vmspduperf2.amr.corp.intel.com	Database VM_VCDB in SQL Server instance...	1	bcardbba	unassigned	APP_SQL
1	4/8/15 12:07 PM	4/8/15 11:27 AM	dgger201.ger.corp.intel.com	The Health service on computer dgger201.g...	1	aeastr2	INC00397862	SYS_AVAI
1	4/8/15 11:16 AM	4/8/15 11:02 AM	f28scorp202.corp.intel.com	The Domain Controller's free disk is low on th...	1	qcastr2	INC00397747	APP_Activ
1	4/8/15 9:15 AM	4/6/15 11:01 PM	lsgger201.ger.corp.intel.com	Time Created: 04/07/2015 07:59:02 Log Na...	1	slm83	INC003093105	APP_Activ
1	4/8/15 6:44 AM	4/8/15 6:32 AM	dgger201.ger.corp.intel.com	The Health service on computer dgger201.g...	1	rhalla	unassigned	SYS_AVAI
1	4/8/15 2:52 AM	4/8/15 2:39 AM	cdscdbkimb.ccr.corp.intel.com	The computer cdscdbkimb.ccr.corp.intel.co...	1	slm83	unassigned	SYS_AVAI
1	4/8/15 2:25 AM	4/8/15 2:20 AM	cdscdprcmb.ccr.corp.intel.com	Multiple Alerts Site (CD) - The computer cds...	2	slm83	unassigned	SYS_AVAI
1	4/8/15 2:24 AM	4/8/15 1:35 AM	cdspftcb.ccr.corp.intel.com	The computer cdspftcb.ccr.corp.intel.com w...	1	slm83	unassigned	SYS_AVAI
1	4/8/15 1:34 AM	4/8/15 1:30 AM	cdscdbkimb.ccr.corp.intel.com	Multiple Alerts Site (CD) - The computer cds...	2	slm83	unassigned	SYS_AVAI
1	4/8/15 1:29 AM	4/8/15 1:23 AM	cdspftcb.ccr.corp.intel.com	Multiple Alerts Site (CD) - The computer cds...	2	slm83	unassigned	SYS_AVAI
1	4/7/15 7:37 PM	4/7/15 7:31 PM	lsgger201.ger.corp.intel.com	Multiple Alerts Site (IS) - The computer lsg...	2	lgh2	unassigned	SYS_AVAI
1	4/7/15 7:37 PM	4/7/15 7:31 PM	irsam201.amr.corp.intel.com	Multiple Alerts Site (IR) - The computer irs...	2	lgh2	unassigned	SYS_AVAI
1	4/7/15 2:24 PM	4/6/15 8:30 PM	fmsvds02.ed.cps.intel.com	Multiple Alerts -- 999: The VDS configura...	3	slm83	INC003092880	APP_VDS

Figura 4 Impact Manager

### IMODS

Base de Datos que recolecta información de rendimiento de los equipos tecnológicos de Intel, permitiendo en este momento obtener información sobre el CPU, Memoria y Espacio en Disco de las unidades de disco

lógicas existentes. Esta fuente de datos nos brinda el estado de los equipos de Intel, con el fin de identificar patrones de potenciales problemas que se presentan en los equipos.

### *IAP*

Base de Datos que recolecta información de las Aplicaciones utilizadas en Intel. En ella se registra el identificador único de la aplicación, su nombre, descripción, información de contacto y el nombre de los servidores asociados.

## Creación del Data Warehouse 2.0

### Análisis de Infraestructura y arquitectura de datos

Global System Monitoring utiliza el sistema “Service Now” e “Impact Manager”. Ambos representan la principal fuente de datos y representan la base para la extracción de datos cargados en el área de *Staging*.

La infraestructura actual se basa en servidores físicos para la base de datos con SQL Server versión 2014. Para la creación del modelo de Data Warehouse 2.0 se utilizaron tanto máquinas virtuales como servidores físicos otorgadas por la empresa. Para el despliegue (*deploy*) y procesamiento del modelo tabular se utiliza un servidor físico, el cual es parte también de la infraestructura de Intel. Para el Staging Area se utiliza una máquina virtual.

En cuanto a la arquitectura de datos fue necesario definir las tablas del sistema de Base de Datos transaccional requeridas, ya que como fue mencionado el sistema de tiquetes se encuentra basado en ITIL, por lo tanto abarca todas sus áreas. En nuestro caso el enfoque estará en la Gestión de Eventos, Incidentes y en menor proporción en la Gestión de Problemas, ya que son las áreas principales que Global System Monitoring opera.

Para la implementación del presente proyecto se requirió contar con dos servidores (HP ML580 G8, con 8 procesadores y 36 GB de RAM) con Microsoft SQL Server 2014 configurado correctamente, con el fin de contar con los servicios de Bases de Datos, SSIS (SQL Server Integration Services), SSAS (SQL Server Analysis Services) en modalidad Tabular. Estos recursos son suficientes para la construcción de la solución de Data Warehouse 2.0 y la implementación del algoritmo de Minería de Texto para el análisis de Eventos e Incidentes generados en las consolas de monitoreo de Global System Monitoring.

## Diseño del Modelado Tabular

1. Para el modelado tabular se debe tener un análisis previo de las tablas a utilizar del sistema transaccional a fin de seleccionar cuáles son las que proveerán los datos requeridos, cuál será la granularidad, identificar las dimensiones y hechos.
2. Se debe definir la nomenclatura a utilizar, los estándares para ubicar la documentación, el mapeo de nombres técnicos con los nombres de negocio, los roles y permisos, la creación de perspectivas si son necesarias.
3. A partir de ello se comienza con la definición, diseño y construcción de los ETLs que permitirán la extracción, limpieza y transformación de los datos hacia los DataMart o Staging Area que tendremos. Esto forma parte del Back End. Los ETLs deben tener sus conexiones definidas, las tareas, el flujo de ejecución, las variables o parámetros necesarios y la frecuencia de ejecución.
4. Iniciando las tareas del “Front End”, se creó el modelo tabular, definiendo las siguientes tablas de hechos: Incidentes (*Incidents*), Solicitudes de Servicio (*Service Requests*), Alertas/Eventos(*Alerts*) y Rendimiento del Servidor (Server Performance) relacionada a las siguientes dimensiones:
  - Fecha (Date)
  - Servidor (Server)
  - Aplicación (Applications)
  - Empleados (Employees)
  - Espacio en Disco en Servidores (Disk Space)
  - Tiquetes relacionados a Problemas (Problems)
  - Parches de Sistemas Operativos (Patches)

**Nota:** el nombre entre paréntesis es el nombre a utilizar en el tabular, ya que Intel maneja como idioma primario el inglés.

5. Para la capa de presentación de la aplicación de Business Intelligence, se requiere un Dashboard interactivo para lo cual se utilizó la tecnología de PowerView y SSRS (SQL Server Reporting Services), sobre la plataforma SharePoint Business Intelligence 2013.

6. Para el análisis de datos en tiempo real e histórico se usaron técnicas de minería de textos, usando la herramienta: Microsoft SSIS – Term Extracction.

### Arquitectura Propuesta

El modelo transaccional se encuentra en una base de datos SQL Server 2014. Como parte de la solución de Data Warehouse 2.0, se utilizaron los sistemas transaccionales existentes, para extraer la información requerida y así proceder a cargar al Staging Area (basada en SQL Server 2014). A partir de ello se realizan las tareas de ETLs necesarias para incorporar los eventos históricos y los más recientes en tiempo real, para ellos se utilizaran paquetes de “Integration Services” para la carga de datos al área de Staging.

Además se cuenta con un servidor adicional encargado de realizar las tareas de Data Warehouse 2.0 (Sector Interactivo, Sector Integrado y Sector Archivo) donde se integraran las actividades necesarias.

Las herramientas de ETLs serán el SSIS (SQL Server Integration Services) y las herramientas para la construcción de la solución Data Warehouse 2.0 (tabular) será el SSAS (SQL Server Analysis Services) ambas de Microsoft. La posterior visualización de los datos se realizará medio de PowerView, Excel y SSRS (SQL Server Reporting Services).

A continuación se mostrará el diseño de la arquitectura propuesta:

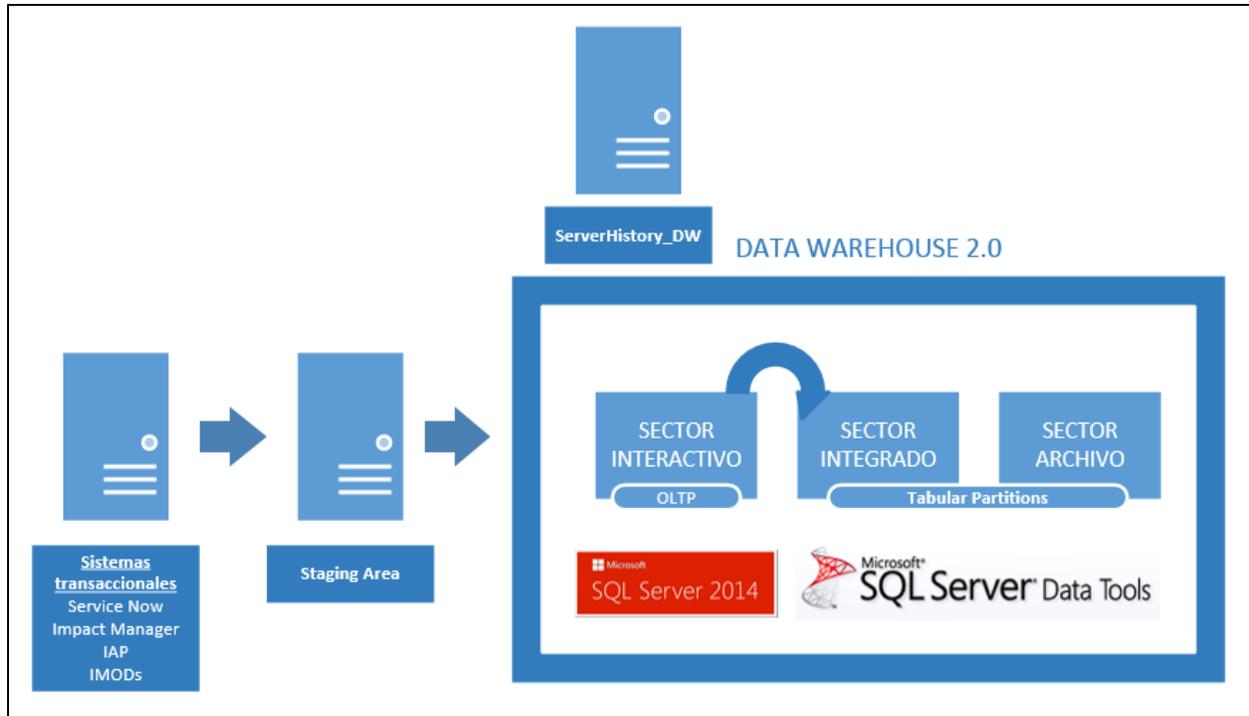


Figura 5 Diseño de la Arquitectura Propuesta

A continuación se describen las consideraciones relacionadas a la antigüedad de los datos, dichas cargas se realizan mediante particiones sobre la solución del modelo tabular para ajustar los siguientes requerimientos:

- El sector interactivo del Data Warehouse 2.0 contiene datos en tiempo real, con una antigüedad no mayor a 60 min. Ésta es actualizada constantemente de los sistemas transaccionales
- El sector Integrado del Data Warehouse 2.0 contiene datos de un mes.
- En el sector Archivo del Data Warehouse 2.0 tiene una antigüedad no mayor a un año.

**Nota:** la herramienta puede contener todos los datos generados por las fuentes de datos mayores a un año, sin embargo basándose en los requerimientos del negocio se tiene como límite datos no mayores a un año. Si a futuro si necesitara incluir mayor cantidad de datos, para otro tipo de análisis, la solución sería escalable a ello, mediante la modificación de la partición que abarca el sector de Archivo.

## Identificación de las tablas fuentes de los Sistemas Transaccionales

Las tablas cargadas en el “*Staging Area*” (SA) son aquellas que se identificaron en las fuentes de datos primarias que tienen gran importancia en el análisis de recurrencias de servidores monitoreados por Global System Monitoring. A continuación se detallan las tablas que conformarán el SA provenientes de la fuente de datos primarias de Global System Monitoring:

- **dbo.ServerInventory:** muestra el inventario de todos los equipos (servidores) monitoreados por el departamento de Global System Monitoring.
- **dbo.imalerts\_all:** registra todas las alertas provenientes de la BEM- Impact Manager.
- **dbo.DiskUsage:** reporta información del espacio en disco disponible en los 18.000 servers soportados por Global System Monitoring.
- **dbo.ServersTickets:** registro de tiquetes de Service Now generados por cada servidor.
- **dbo.ServerUptimeReport:** brinda un Informe del tiempo en que ha estado funcionando cada servidor desde que fueron reiniciados.
- **dbo.ServiceNow\_AllIncidents:** muestra todos los incidentes registrados en Service Now por cada servidor, estos pueden ser generados por Impact Manager o creados por los usuarios.
- **dbo.ServiceNow\_AllRequest:** informe de todas las solicitudes de servicio creadas por parte de los usuarios y que son registrados en Service Now.
- **dbo.ServiceNowProblems:** detalla todos los tiquetes de problemas creados en Service Now debido a repetitivos incidentes sobre un mismo equipo.
- **[HostingPerf] dbo.ServerPerfData:** muestra indicadores de rendimiento por cada servidor tiempo de procesador, espacio libre, memoria utilizada (*Processor Time, Free Space, Committed Memory Bytes in Use*).

### Modelo Data Warehouse 2.0: Esquema estrella del Data Warehouse 2.0

A continuación se mostrará el detalle del Esquema Estrella basado en los requerimientos de Global System Monitoring:

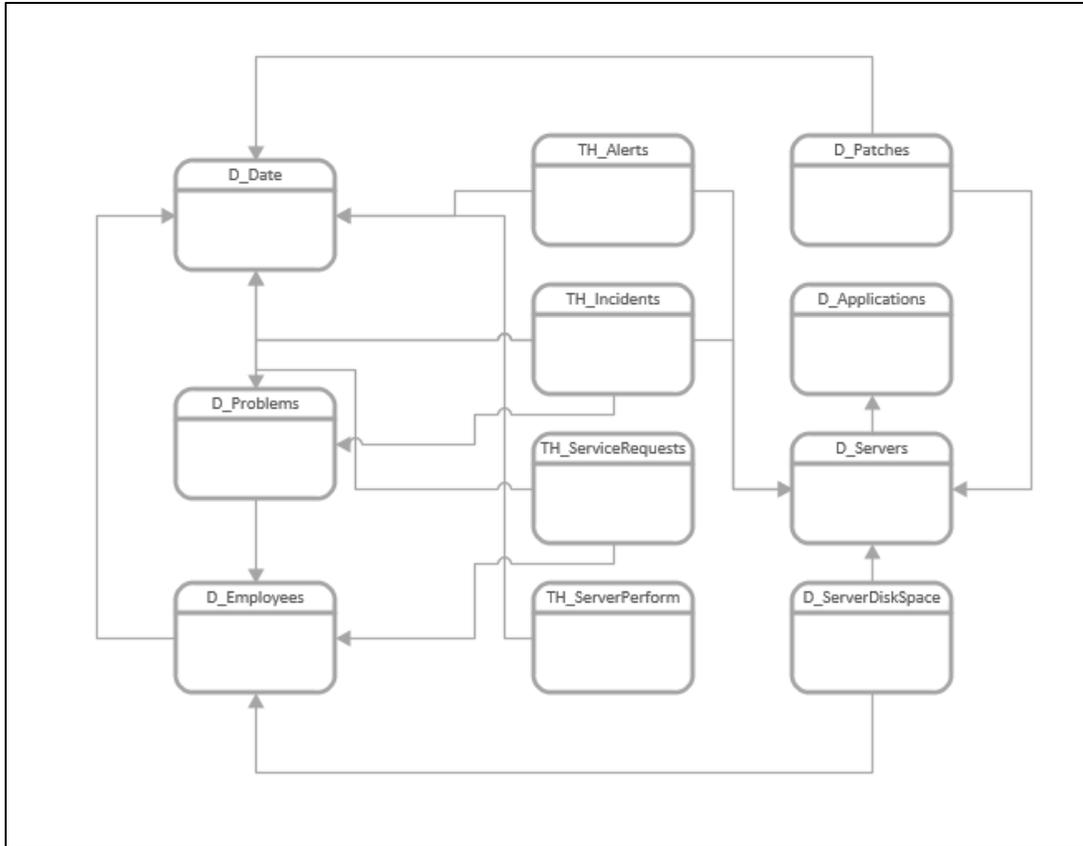


Figura 6 Esquema Estrella del Data Warehouse

### Carga de datos al Staging Area

La carga de los datos de los Sistemas Transaccionales hacia al área de “Staging” se realizó a través de la herramienta SSIS (SQL Server Integration Services), se definieron contenedores separados para la carga de datos al área de Staging de acuerdo a la naturaleza de los datos y se calendarizaron tareas de refrescamiento para obtener datos en tiempo real. En la Figura 6 se muestra los contenedores en SSIS (SQL Server Integration Services) que cargan los datos de las fuentes de datos externas:

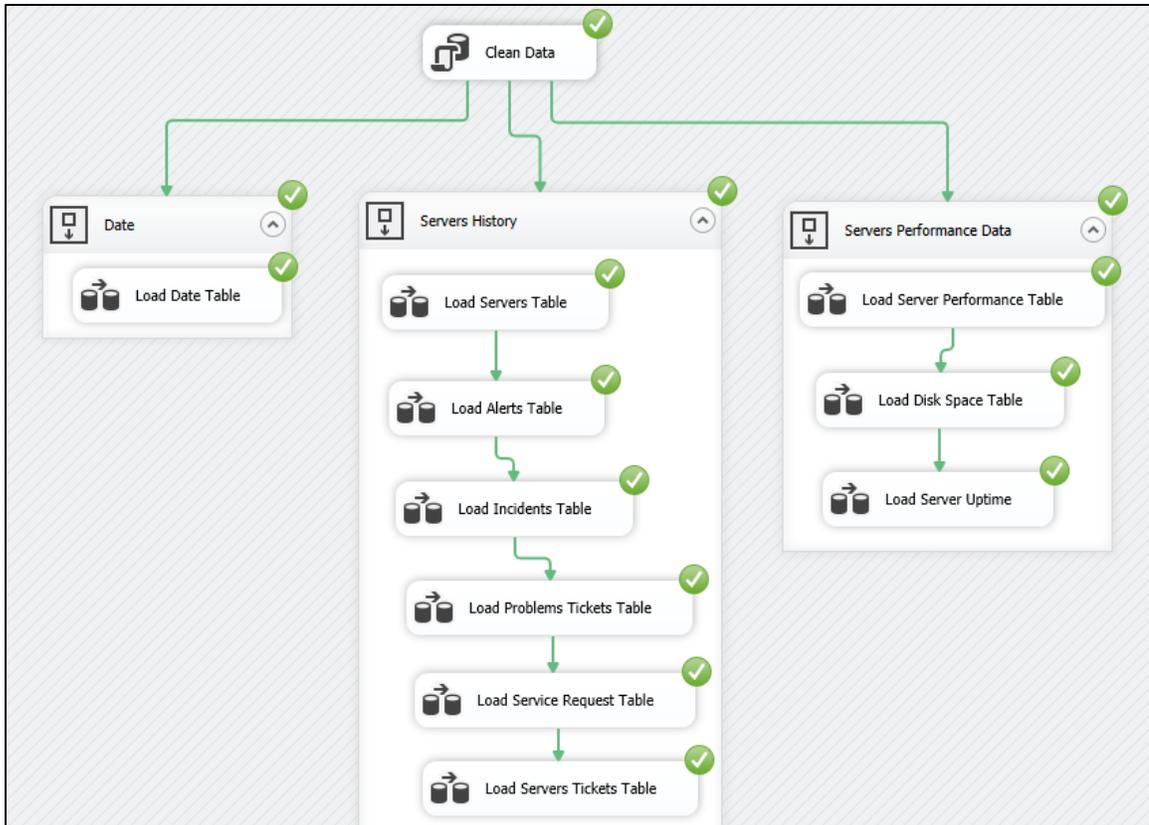


Figura 7 Staging Area-Fuentes de Datos

A continuación se explicará el proceso de carga de las tablas en el área de “Staging” provenientes de las diferentes de datos:

1. *Limpiar Datos*

Este proceso se encarga de truncar (Truncate) las tablas involucradas durante la carga de las tablas en el área de Staging.

2. *Contenedor Fecha*

**Tabla Fecha:** Se encarga de cargar los datos de fechas desde el 2010 hasta el 2030 utilizadas por Intel en términos de fechas fiscales y de calendario. Estos datos provienen de un servidor compartido para usos internos y el destino final de estos datos estará en la Base de Datos de Staging “ServerHistory\_SA” .

3. *Contenedor Historia del Servidor*

**Tabla Servidor:** Se cargan los datos relacionados de los servidores existentes en la infraestructura de Intel. Estos datos provienen la Base de Datos de Service Now y el destino final de estos datos estarán en la Base de Datos de Staging “ServerHistory\_SA”.

**Tabla Alertas:** Se traen datos relacionados a las alertas/eventos generados por los paquetes de monitoreo instalados en la infraestructura de Intel. Estos datos provienen la Base de Datos de BEM Impact Manager y el destino final de estos datos estarán en la Base de Datos de Staging “ServerHistory\_SA”.

**Tabla Incidente:** Carga de los incidentes generados por los usuarios y por paquetes de monitoreo. Estos datos provienen la Base de Datos de Service Now y el destino final de estos datos estarán en la Base de Datos de Staging “ServerHistory\_SA”.

**Tabla Solicitudes de Servicio:** Proceso que se encarga de extraer información de las solicitudes de servicio generadas por los usuarios. Estos datos provienen la Base de Datos de Service Now y el destino final de estos datos estarán en la Base de Datos de Staging “ServerHistory\_SA”.

**Tabla Empleados:** Se cargan los datos relacionados a los empleados de Global System Monitoring que trabajan en Intel. Estos datos provienen la Base de Datos CDIS y el destino final de estos datos estarán en la Base de Datos de Staging “ServerHistory\_SA”.

**Tabla Aplicaciones:** Esta tarea se encarga de extraer la información de las Aplicaciones en Intel a las que tienen los servidores registrados. Estos datos provienen la Base de Datos de IAP y el destino final de estos datos estarán en la Base de Datos de Staging “ServerHistory\_SA” .

**Tabla Problemas:** Se extraen la información de los tiquetes relacionados a los Problemas relacionados a los servidores y que son producto de la recurrencia de muchos incidentes sobre la misma infraestructura. Estos datos provienen la Base de Datos de Service Now y el destino final de estos datos estarán en la Base de Datos de Staging “ServerHistory\_SA”.

#### 4. *Contenedor Datos de Rendimiento del Servidor:*

**Tabla Rendimiento de Servidor:** Se encarga de extraer información sobre el uso de CPU, Memoria y Network de los servidores. Estos datos provienen la Base de Datos de IMODs y el destino final de estos datos estarán en la Base de Datos de Staging “ServerHistory\_SA”

**Tabla de Espacio en Disco en el Servidor:** Proceso que se encarga de extraer información del espacio en disco disponible en la infraestructura de servidores de Intel. Estos datos provienen la Base de Datos de IMODs y el destino final de estos datos estarán en la Base de Datos de Staging “ServerHistory\_SA”.

**Tabla Parches:** Proceso que se encarga de extraer información de los parches aplicados a los servidores. Estos datos provienen la Base de Datos IMODs y el destino final de estos datos estarán en la Base de Datos de Staging “ServerHistory\_SA”

5. Para las tablas mencionadas anteriormente se realizaron las siguientes tareas de transformación (CAST) del tipo de datos para asegurar la consistencia e integridad de los datos a ser analizados e interpretados.
  - Tipo de dato para campos relacionados a “Fechas” fueron convertidos al tipo “Date”, ya que muchos de los atributos utilizaban diferentes tipos.
  - Tipo de dato para campos relacionados a “Nombre de Servidor” fueron convertidos “Varchar (50)”.

## Descripción de cada Dimensión y Tabla de Hechos

A continuación se mostraran las tablas de dimensiones y de hechos implementadas en el presente proyecto que permiten el análisis respectivo de recurrencias por servidor y áreas específicas en las que se deba poner mayor atención:

- **Fecha:** Dimensión de Fechas, contiene detalle de los años, trimestres, meses y semanas basadas en el calendario de Intel. Se incluye una jerarquía por año, cuatrimestre, mes y semana de trabajo.
- **Servidor:** Dimensión de Servidores que muestra detalle de todos los servidores monitoreados por Global System Monitoring.
- **Aplicación:** Dimensión de Aplicaciones que contiene información de las aplicaciones asociadas a cada servidor que genera los eventos e incidentes.
- **Empleados:** Dimensión de Empleados, se detallan los atributos de los empleados que son asociados a los servidores y agentes de monitoreo de Global System Monitoring.

- **Problemas:** Dimensión de Tiquetes de Problemas, se detalla la información de los tiquetes relacionados a los Problemas asociados a los servidores que Global System Monitoring monitorea.
- **Parches:** Dimensión de Parches de Sistema Operativo, se guarda información de los Parches instalados en los servidores.
- **Espacio en Disco:** Dimensión de Espacio en Disco por cada equipo de Intel.
- **Incidentes:** Tabla de Hechos que detallan los incidentes generados y registrados por Global System Monitoring.
- **Rendimiento de Servidor:** Tabla de Hechos que detallan el rendimiento por cada equipo de Intel en términos de uso de Memoria y CPU.
- **Alertas:** Tabla de Hechos que detallan los eventos generados y registrados por Global System Monitoring.
- **Solicitudes de Servicio:** Tabla de Hechos que detallan las solicitudes de servicio generados por los usuarios de Global System Monitoring.

### Carga de datos al Data Warehouse

La carga de los datos del área de “Staging” al Data Warehouse se realizó a través de la herramienta SSIS (SQL Server Integration Services), se definieron las tareas calendarizadas para el refrescamiento apropiado de los datos en tiempo real de los elementos principales (Eventos y Rendimiento de Servidores).

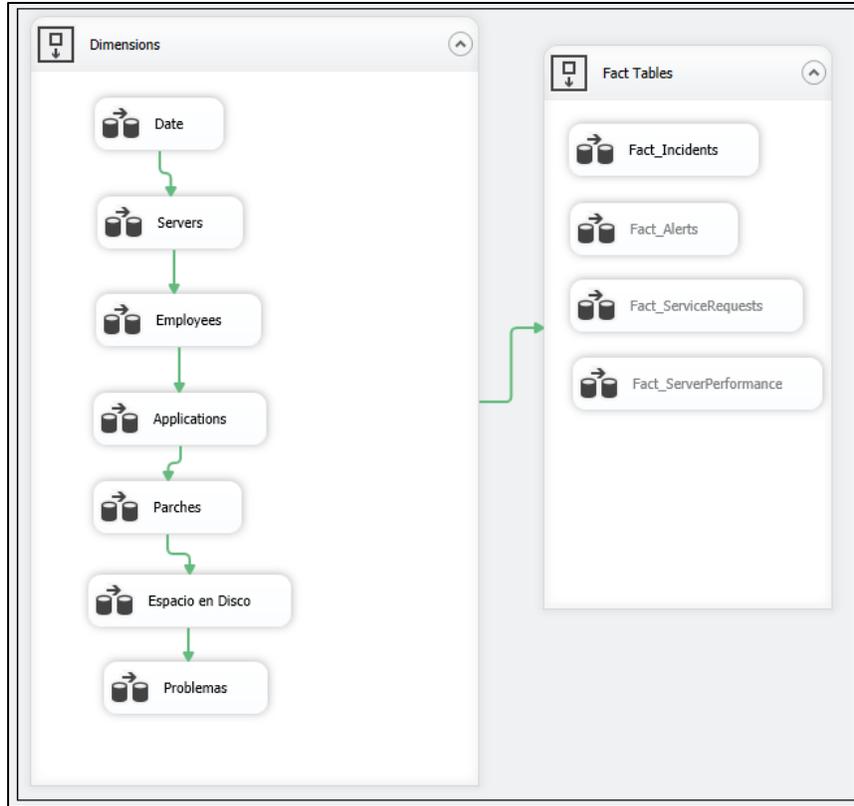


Figura 8 Carga de Datos al Data Warehouse

A continuación se explicará el proceso de carga de las tablas de dimensiones y tablas de hechos en el área de “Data Warehouse” provenientes de las tablas del área de “Staging”:

**Sector Interactivo:** Para este sector los datos son extraídos directamente de los sistemas transaccionales. La antigüedad máxima de los datos en esta etapa es de 60 minutos.

#### **Sector Integrado y Archivo**

- **Contenedor de tablas de Dimensiones:**

Las Tablas de Dimensiones son cargadas del área de “Staging” (ServerHistory\_SA) y son refrescadas cada 24 horas en la Base de Datos del Data Warehouse 2.0 (ServerHistory\_DW) a través de una tarea calendarizada de SQL.

- **Contenedor de tablas de Hechos:**

Las Tablas de Hechos son cargadas del área de “Staging” (ServerHistory\_SA) y son refrescadas cada 24 horas en la Base de Datos del Data Warehouse 2.0 (ServerHistory\_DW) a través de una tarea calendarizada de SQL.

### Carga de datos al Modelo Tabular

A continuación se mostrará el detalle del proceso de carga de las tablas de dimensiones y de hechos:

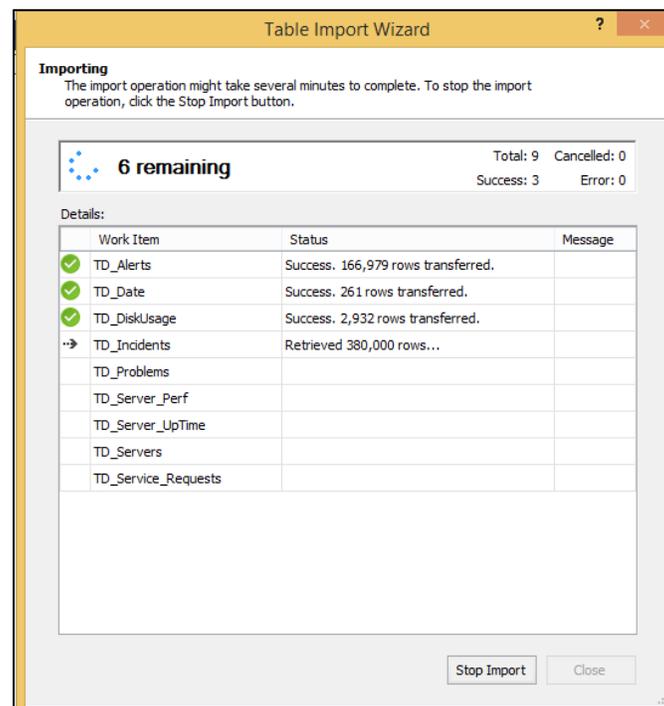


Figura 9 Carga de datos al tabular satisfactoria

### Modelo tabular

Como se mencionó anteriormente sobre la utilización de Modelos Multidimensionales de tipo Tabular en el presente proyecto, se muestra a continuación diagrama de las relaciones multidimensionales en el modelo Tabular:

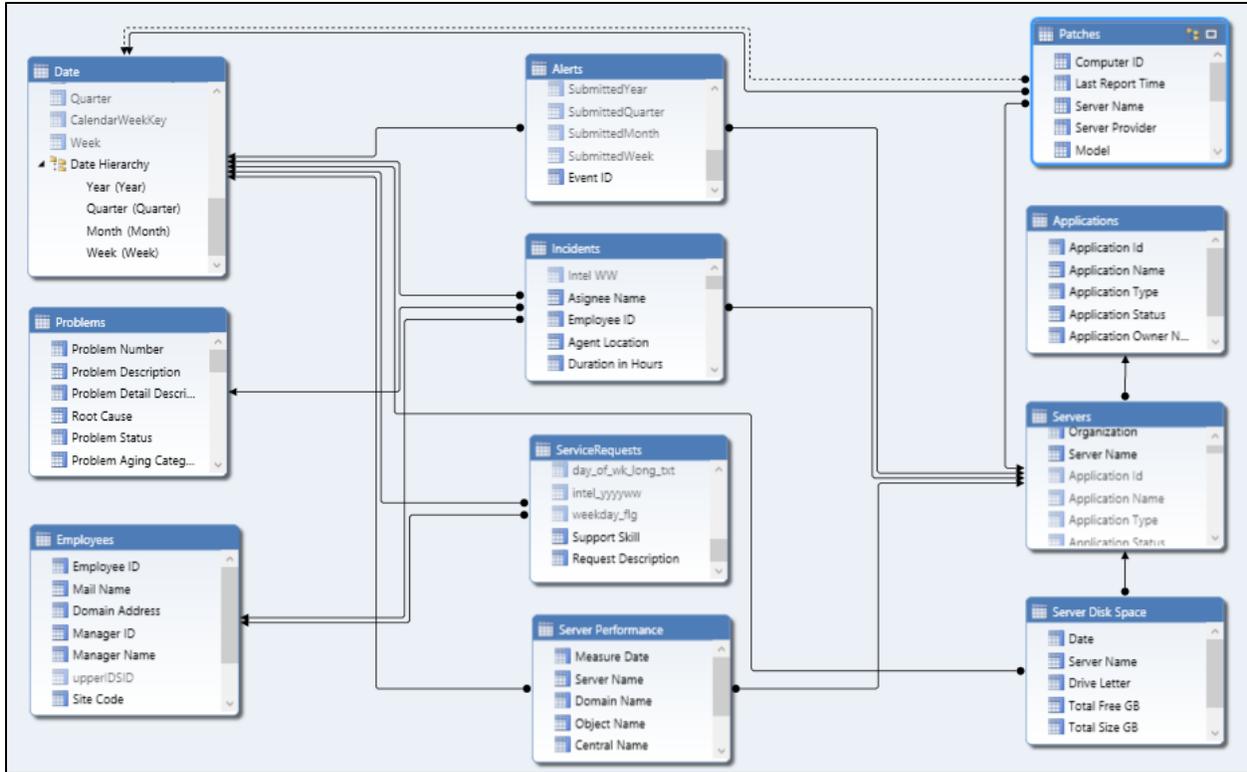


Figura 10 Solución Modelo Tabular

Job...	hostname	EventID	Resource_ID	Reporting_TEC	Date_Time	Status	Class	Severity	Message	Source	Sub_Source	Origin	Sub_Origin	adapter_host	page
3564	afbc	353329322	145069	CELLFMRD001	8/18/2014 2...	OPEN	NW_WAN	CRITICAL	(WAN) CO...			0.0.0.0		looku...	
3565	afbc	353330037	145069	CELLFMRD001	8/18/2014 3...	OPEN	NW_WAN	CRITICAL	(WAN) CO...			0.0.0.0		looku...	
4624	amcaar	353129195	125882	CELLFMRD001	8/6/2014 3:...	OPEN	NW_WAN	CRITICAL	BGP-5-AD...			0.0.0.0		looku...	
4637	amcaar	353238190	125882	CELLFMRD001	8/12/2014 1...	OPEN	NW_WAN	CRITICAL	A minor IP...			0.0.0.0		looku...	
4656	amcaar	353408929	125882	CELLFMRD001	8/23/2014 3...	OPEN	NW_WAN	CRITICAL	BGP-5-AD...			0.0.0.0		looku...	
4674	amcaar	353620563	125882	CELLFMRD001	9/3/2014 8:...	OPEN	NW_WAN	CRITICAL	BGP-5-AD...			0.0.0.0		looku...	
4677	amcaar	353621982	125882	CELLFMRD001	9/3/2014 10...	OPEN	NW_WAN	CRITICAL	A minor IP...			0.0.0.0		looku...	
4678	amcaar	353621983	125882	CELLFMRD001	9/3/2014 10...	OPEN	NW_WAN	CRITICAL	A minor IP...			0.0.0.0		looku...	
4679	amcaar	353621997	125882	CELLFMRD001	9/3/2014 10...	OPEN	NW_WAN	CRITICAL	A minor IP...			0.0.0.0		looku...	
4680	amcaar	353622002	125882	CELLFMRD001	9/3/2014 10...	OPEN	NW_WAN	CRITICAL	A minor IP...			0.0.0.0		looku...	
4682	amcaar	353627380	125882	CELLFMRD001	9/3/2014 4:...	OPEN	NW_WAN	CRITICAL	LINK-3-LP...			0.0.0.0		looku...	
4684	amcaar	353628013	125882	CELLFMRD001	9/3/2014 6:...	OPEN	NW_WAN	CRITICAL	LINK-3-LP...			0.0.0.0		looku...	
4699	amcaar	353677263	125882	CELLFMRD001	9/6/2014 9:...	OPEN	NW_WAN	CRITICAL	LINK-3-LP...			0.0.0.0		looku...	
4713	amcaar	353703075	125882	CELLFMRD001	9/7/2014 6:...	OPEN	NW_WAN	CRITICAL	LINK-3-LP...			0.0.0.0		looku...	
4736	amcaar	353814055	125882	CELLFMRD001	9/13/2014 1...	OPEN	NW_WAN	CRITICAL	BGP-5-AD...			0.0.0.0		looku...	
4768	amcaar	353976165	125882	CELLFMRD001	9/21/2014 5...	OPEN	NW_WAN	CRITICAL	A minor IP...			0.0.0.0		looku...	
4771	amcaar	353978924	125882	CELLFMRD001	9/21/2014 8...	OPEN	NW_WAN	CRITICAL	A minor IP...			0.0.0.0		looku...	
4772	amcaar	353978925	125882	CELLFMRD001	9/21/2014 8...	OPEN	NW_WAN	CRITICAL	A minor IP...			0.0.0.0		looku...	
4814	amcaar	354134218	125882	CELLFMRD001	9/29/2014 3...	OPEN	NW_WAN	CRITICAL	A minor IP...			0.0.0.0		looku...	
4819	amcaar	354148742	125882	CELLFMRD001	9/29/2014 6...	OPEN	NW_WAN	CRITICAL	A minor IP...			0.0.0.0		looku...	
5183	amcbor	352977241	125884	CELLFMRD001	7/26/2014 8...	OPEN	NW_WAN	CRITICAL	BAD LINK ...			0.0.0.0		looku...	
5184	amcbor	352977244	125884	CELLFMRD001	7/26/2014 8...	OPEN	NW_WAN	CRITICAL	BAD LINK ...			0.0.0.0		looku...	
5187	amcbor	353013458	125884	CELLFMRD001	7/29/2014 3...	OPEN	NW_WAN	CRITICAL	BAD LINK ...			0.0.0.0		looku...	
5217	amcbor	353363017	125884	CELLFMRD001	8/20/2014 9...	OPEN	NW_WAN	CRITICAL	BAD LINK ...			0.0.0.0		looku...	
5239	amcbor	353363037	125884	CELLFMRD001	8/20/2014 1...	OPEN	NW_WAN	CRITICAL	BAD LINK ...			0.0.0.0		looku...	
5249	amcbor	353371934	125884	CELLFMRD001	8/20/2014 1...	OPEN	NW_WAN	CRITICAL	BAD LINK ...			0.0.0.0		looku...	
5251	amcbor	353372830	125884	CELLFMRD001	8/20/2014 1...	OPEN	NW_WAN	CRITICAL	BAD LINK ...			0.0.0.0		looku...	

Figura 11 Carga de datos de la Dimensión de Alertas

## Creación de las particiones para el Data Warehouse 2.0

La estrategia implementada en el presente proyecto para la creación del Data Warehouse 2.0 se basó en la utilización de particiones a nivel del Modelo Tabular en las tablas de dimensiones y de hechos, de tal forma que permita la carga de datos de acuerdo a las reglas de negocio establecidas por Global System Monitoring. La Figura 12 muestra la opción para la creación de “Particiones” en la herramienta de “SQL Server Data Tools para Visual Studio 2013”

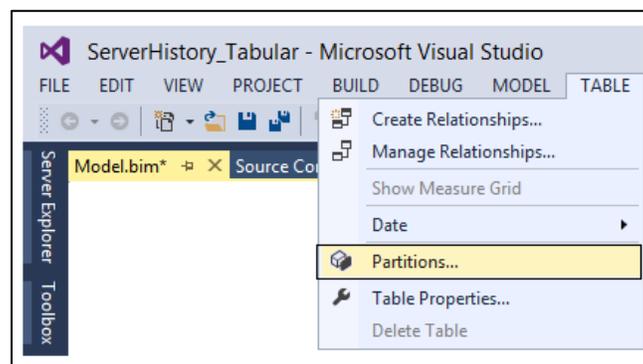


Figura 12 Creación de Particiones para el manejo de datos en el Data Warehouse 2.0

A continuación se detallan las reglas de negocio pactadas con Global System Monitoring:

### **Sector Interactivo:**

Para este sector los datos son consultados directamente de los sistemas transaccionales. La antigüedad máxima de los datos en esta etapa es de 60 minutos.

### **Sector Integrado:**

Este sector retiene información de las tablas de dimensiones y hechos de las últimas 4 semanas. En la Figura 13 se muestra a modo de ejemplo la partición creada en la tabla de Incidentes y el punto importante a mencionar gira alrededor de la condición del cálculo de la fecha para que cumpla los requerimientos de negocio

Ex: WHERE [SubmitDate] >= Convert (DATE, DATEADD(week, -2, GETDATE()))

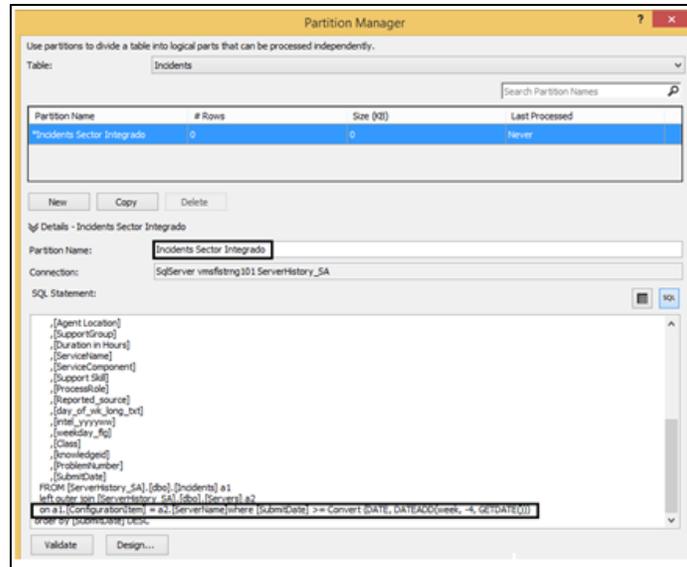


Figura 13 Partición del Sector Integrado

**Sector Archivo:**

Este sector retiene información de las tablas de dimensiones y hechos del último año. En la figura 14 se muestra a modo de ejemplo se muestra la partición creada en la tabla de Incidentes y el punto importante a mencionar gira alrededor de la condición del cálculo de la fecha para que cumpla los requerimientos de negocio

Ex: WHERE [SubmitDate] >= Convert (DATE, DATEADD(year, -1, GETDATE()))

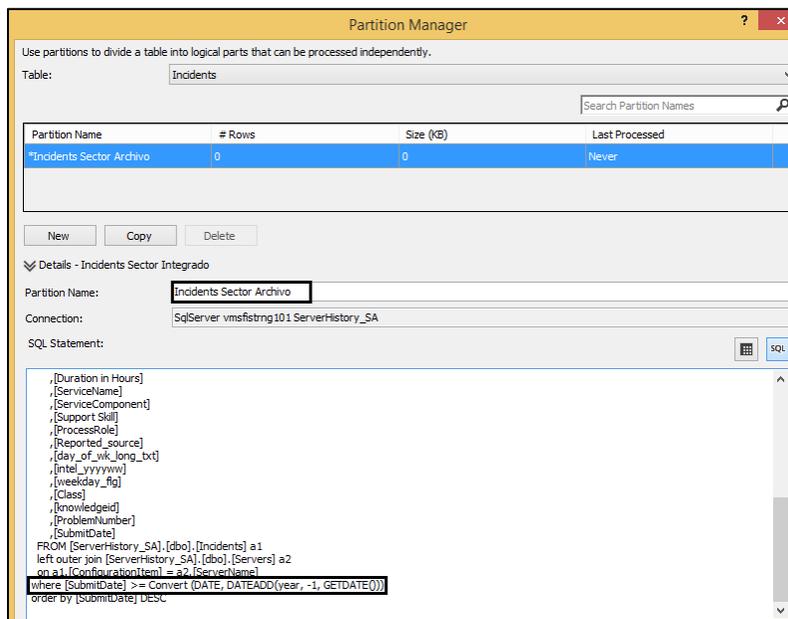


Figura 14 Partición del Sector Archivo

# Capítulo IV: Análisis de Resultados

## Herramientas utilizadas para la presentación de datos

Los reportes y *dashboards* de cada uno de los sectores del Data Warehouse 2.0 fueron creados utilizando Reporting Services y Power View.

**Reporting Services** (Microsoft, Developer Network, s.f.), es una herramienta de Microsoft caracterizada por su facilidad de uso y por ser una plataforma basada en servidor, es decir los reportes pueden ser publicados en web.

Para el presente proyecto fue utilizado en el sector interactivo para la generación de reportes utilizando distintos orígenes de datos, ya que debíamos contar con los datos de alertas (BEM Impact Manager) y Service Now.

**Power View**: Power View (Microsoft, Office, s.f.) es una experiencia interactiva de exploración, visualización y presentación de datos para la creación de reportes y dashboards. Power View es una característica de Microsoft Excel 2013 y de Microsoft SharePoint Business Intelligence 2010 y 2013.

Power View fue utilizado para el sector Integrado y de Archivo.

## Presentación del Reporte o Dashboard según las fases del Data Warehouse 2.0

### Sector Interactivo

El reporte implementado en la fase del sector interactivo se basa en los requerimientos establecidos por el departamento Global System Monitoring (GSM), en los que se definen que la visualización de los datos en tiempo real no debe tener una antigüedad mayor a 60 minutos. Esto con el fin de proveer a los analistas de Global System Monitoring una solución que les permita actuar en forma preventiva sobre potenciales tendencias y recurrencias en la consola de monitoreo.

El primer conjunto de gráficos mostrados en la Figura 15, permiten identificar las clases de alertas recurrentes que son generadas por los eventos que reciben en la consola de monitoreo, por otro lado le permiten identificar los colectores o destinos donde llegan más las alertas.

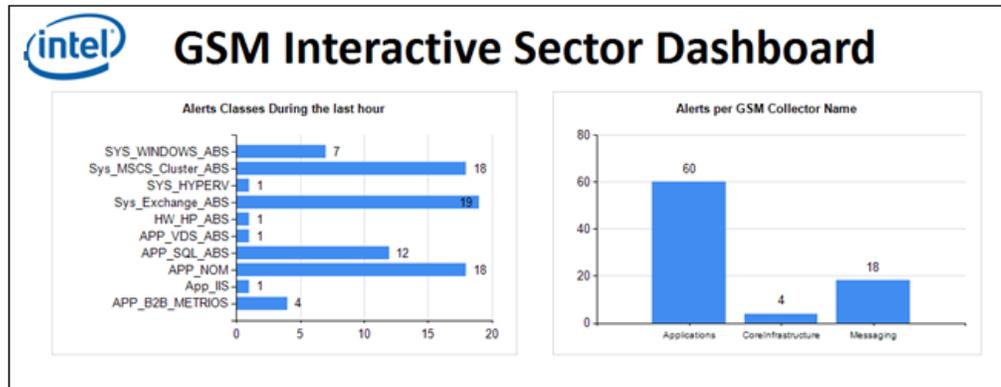


Figura 15 Gráficos del Sector Interactivo

El segundo conjunto de gráficos en la Figura 16 muestra los servidores que han generado mayor recurrencia durante esa hora y el segundo grafico muestra la clasificación de los eventos de acuerdo a la severidad.

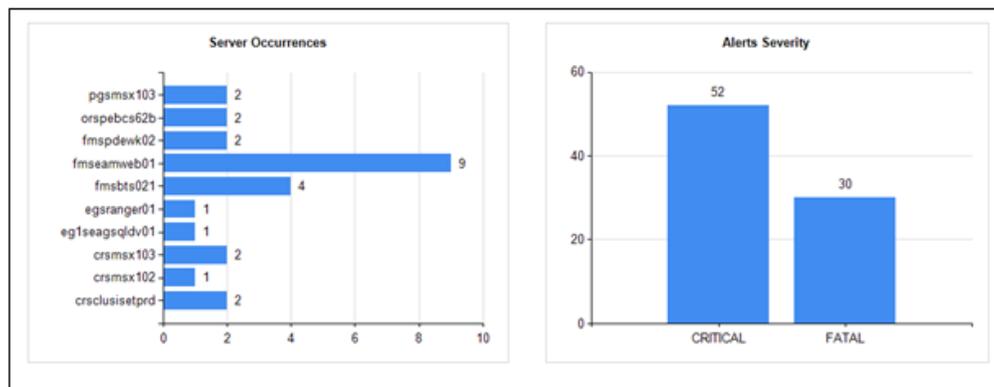


Figura 16 Gráficos del Sector Interactivo

El tercer conjunto de gráficos en la Figura 17 muestra el análisis de Minería de Texto en base a los campos de descripciones que vienen en los eventos / incidentes que llegan a la consola de monitoreo y el segundo grafico muestra los top 10 servers que tienes problemas en espacio en disco.

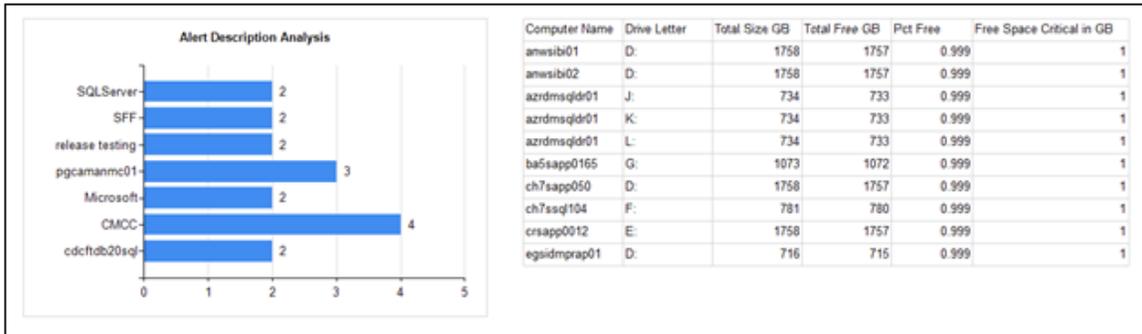


Figura 17 Gráficos del Sector Interactivo

Cabe destacar, que como respuesta a uno de los principales requerimientos se establecieron alertas para cada gráfico, donde el trigger va depender de los parámetros utilizados por Global System Monitoring.

**Ejemplo:** se desea conocer cuáles clases de alertas obtenidas a través de minería de texto tienen más de 5 ocurrencias en un intervalo de 5 minutos, donde la alerta se va enviar nuevamente si el estado de las recurrencias cambió en algo, sino la alerta no se vuelve a generar.

En la Figura 18 podemos ver que el reporte generado a través de Reporting Services en SharePoint 2013, genera un extracto de los ítems que cumplen la regla pre-establecida.

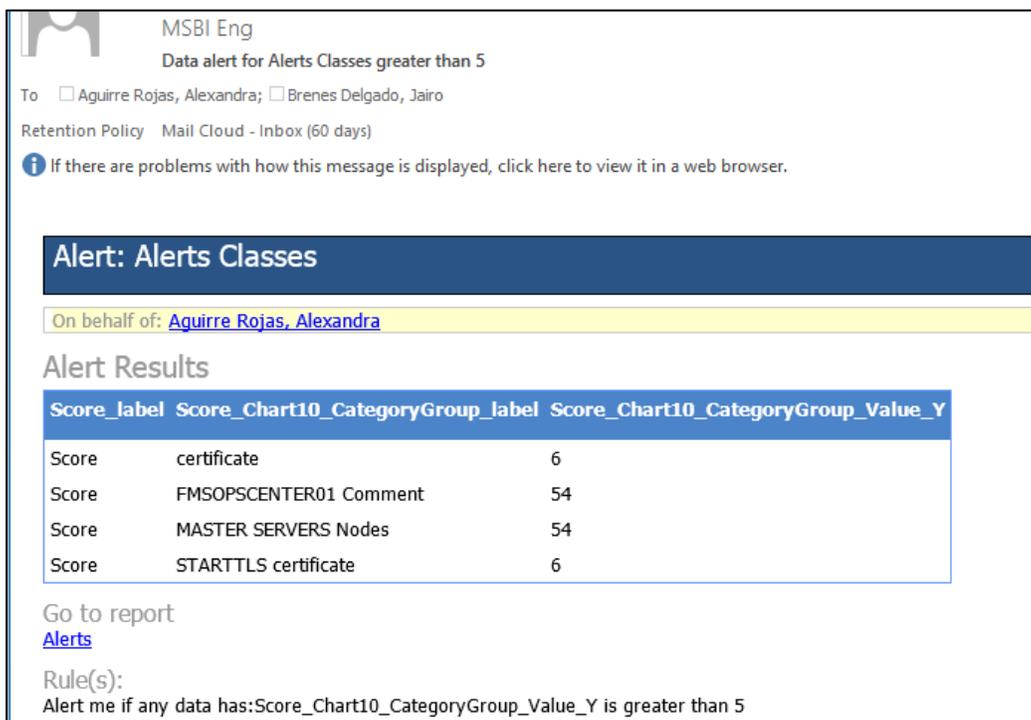


Figura 18 Generación de alertas del Sector Interactivo

## Sector Integrado

El sector integrado de la solución provee datos no mayores a un mes. Estos datos se obtienen directamente del tabular donde está implementada la partición respectiva.

El siguiente dashboard fue creado mediante la herramienta PowerView en SharePoint Business Intelligence 2013.

El requerimiento principal por parte del departamento Global System Monitoring es tener acceso a información de los servidores en sólo lugar, evitando tener que ir a diferentes consolas buscando datos uno por uno del servidor. A partir de ello, las presentes diapositivas del dashboard pretenden dar una fotografía de diversas variables que puedan estar afectando a un servidor o servidores en particular. Los datos son mayores a una hora pero menores o iguales a un mes. Sin embargo algunos de los datos se pueden visualizar mediante semanas o días, según los filtros a aplicar.

La Figura 19 muestra información general de los servidores seleccionados, como Nombre de la Aplicación instalada en el servidor, el nombre del dueño del servidor, el modelo y sistema operativo, se indica si es un servidor físico o virtual porque dependiendo de ello el *troubleshooting* a realizar es diferente y en cuál región está ubicado. Cabe destacar que hay una cantidad considerable de otros datos que podrían irse agregando al dashboard según las necesidades.

En éste primer slide se muestra la forma de de rutear la alerta con el departamento Global System Monitoring, también la cantidad de incidentes y problemas asociados a incidentes que ha tenido el servidor elegido durante ese mes.

Se especifica si alguno de ellos fue un Major Incident y el nivel en que se encuentra el Status.

intel Server Information - Problems & Incidents

fmsidssqtab1 fmsidssqtab2

Microsoft BI Application Name	SHESHADRI, RAMA P Server Owner	ProLiant DL380p Gen8 Server Model	Windows Server 2008 64-bit Server OS	No Server Is Virtual	AMR Server Location Region	
Core Infrastructure Monitor GSM Team	Production Monitor Status	Impact Manager Monitor Service	Yes Monitor Alert Fwd	in_use Asset Status	4 Count of Incident Number	3 Count of Problem Number

Problem Number	Incident Number	Short Description	MI Flag	SLA Status
PR800038104	INC002957816	fmsidssqtab1.amr.corp.intel.com 10.18.112.22 Multiple Alerts -- An error occurred when instantia	False	Within the Service Target
	INC002959502	fmsidssqtab1.amr.corp.intel.com 10.18.112.22 An error occurred when instantiating a metadata obj	False	Within the Service Target
	INC002960377	fmsidssqtab1.amr.corp.intel.com 10.18.112.22 An error occurred when loading the Read.	False	Within the Service Target
	INC003008697	fmsidssqtab1.amr.corp.intel.com 10.18.112.22 Multiple Alerts -- An error occurred when loading t	False	Within the Service Target

Figura 19 Información del Servidor - Problemas e Incidentes

A continuación, la Figura 20 muestra el detalle de los parches instalados durante el mes y las alertas emitidas por el servidor durante ese período de tiempo. Con ello se pretende mostrar cuáles parches están pendientes y tomar acciones inmediatas, evitando generar alertas por inestabilidad del servidor.

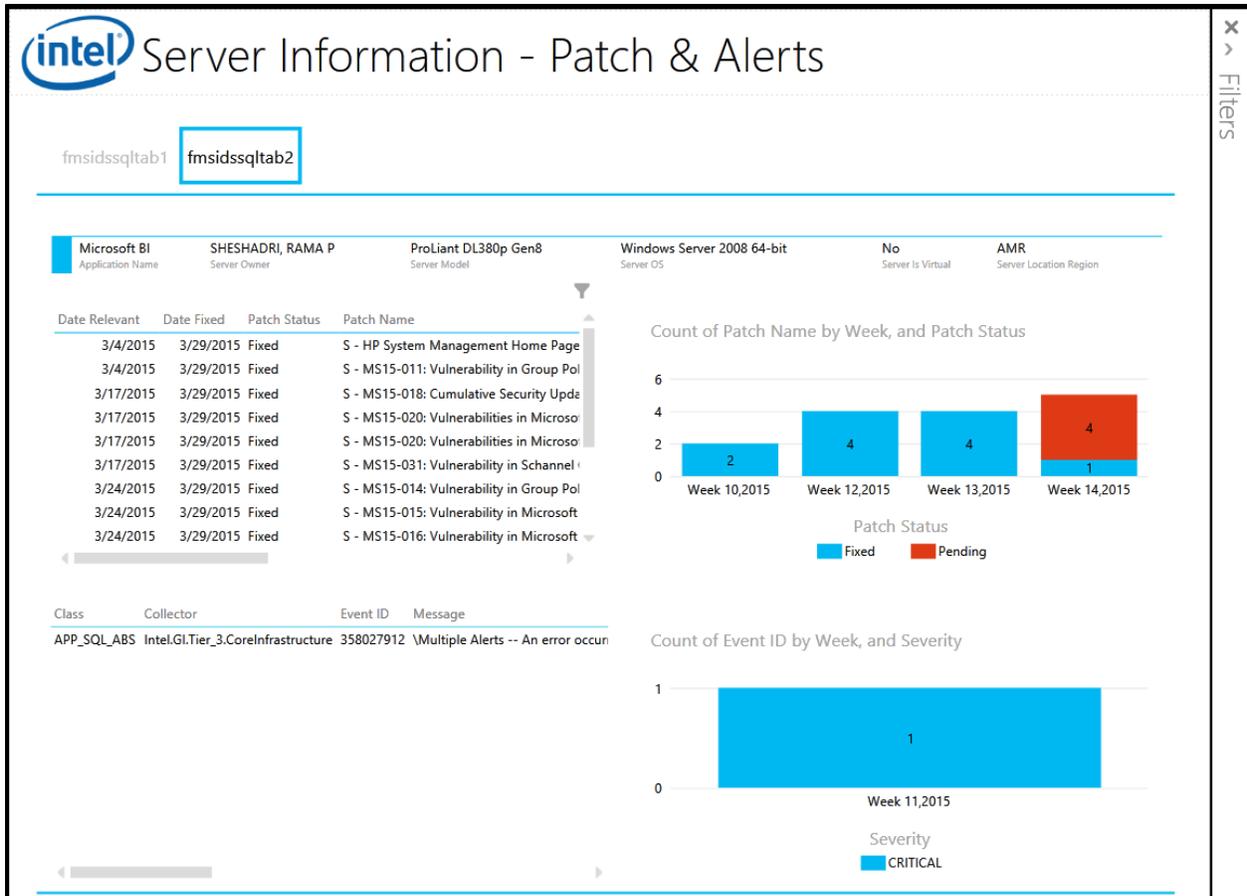


Figura 20 Información del Servidor - Parches y Alertas

### Sector Archivo

La presente solución de visualización está basada en la tecnología de Microsoft llamada PowerView mediante SharePoint Business Intelligence 2013, ya que permite tener un dashboard online.

El sector de Archivo presenta los datos de más de un mes pero menos de un año. En la Figura 21 vemos una presentación de diferentes aspectos relevantes para la salud de un servidor y ver sus tendencias relacionado a problemas, incidentes, espacio en disco y parches.

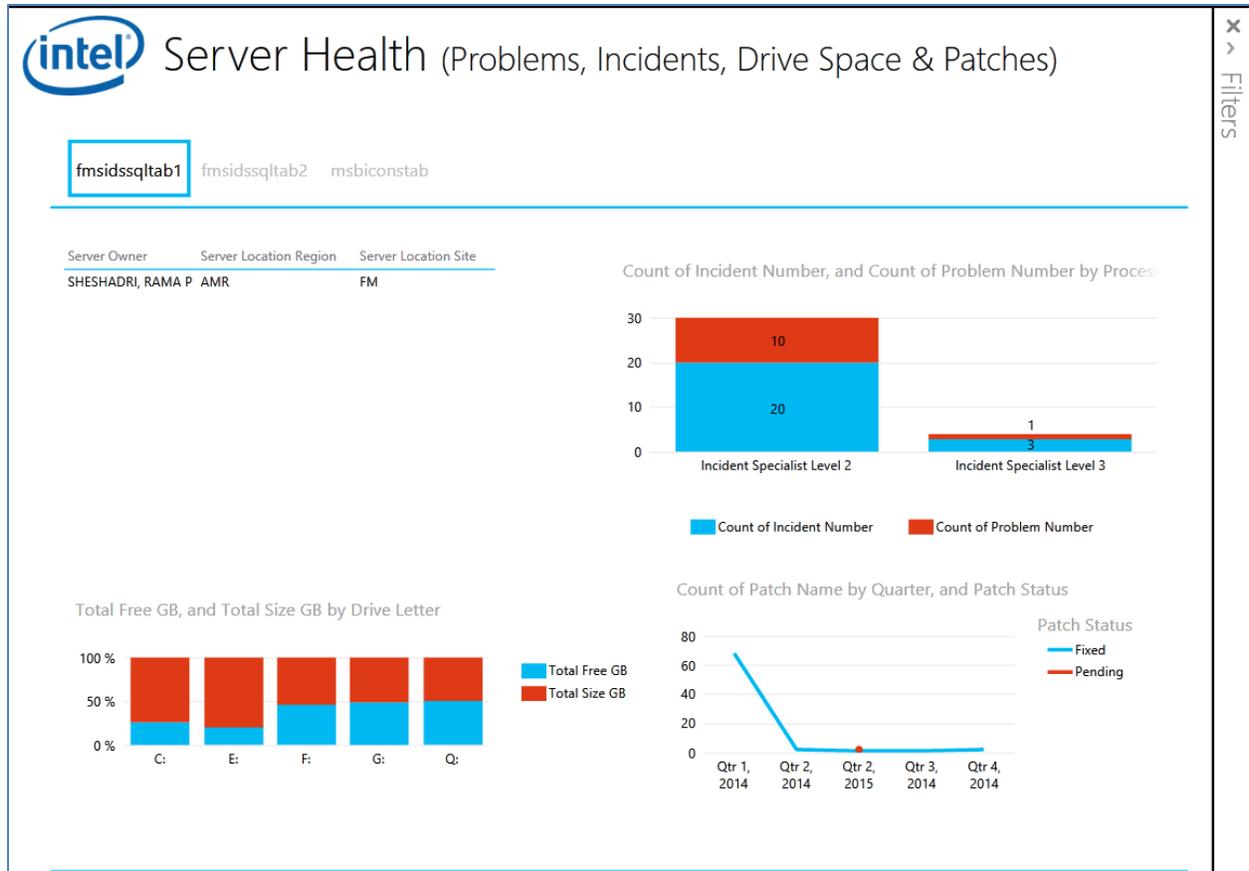


Figura 21 Salud del Servidor

Unido a esto el usuario final puede crear estos mismos reportes de forma local utilizando power view con Excel. A continuación se muestra cuáles pasos se deben seguir.

### Visualización de la solución mediante Excel

#### ¿Cómo crear la conexión?

1. Abrir Excel
2. Ir a Data > Connections > Add > Browse for More > New Source > Data Connection Wizard: Microsoft SQL Server Analysis Services > Connect to Database Server: Server Name (*Nombre del Servidor*) > Select Database and Table (*Seleccionar la base de datos y la perspectiva*) > Next > Save Data Connection File and Finish (*Escribir el nombre deseado de la conexión*) > Close
3. Ir a Data > Connections > Existing Connections > *Seleccionar la conexión recién creada* > Open

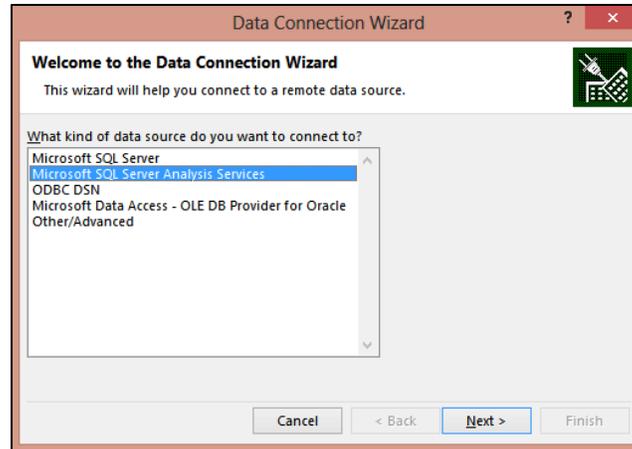


Figura 22 Power View Excel - Data Connection Wizard

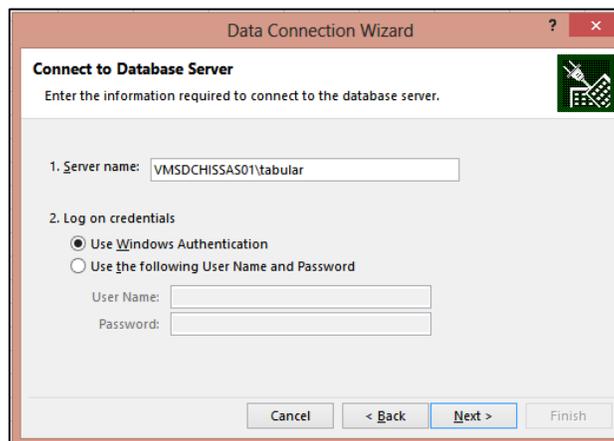


Figura 23 Power View Excel - Data Connection Wizard

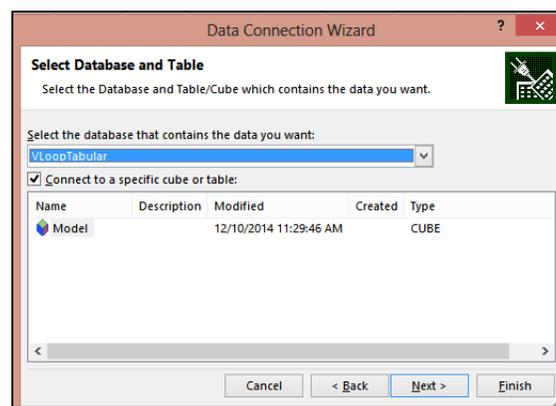


Figura 24 Power View Excel – Selección del Tabular

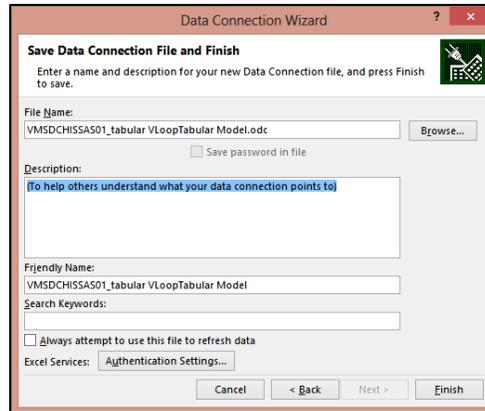


Figura 25 Power View Excel – Nombre de la conexión

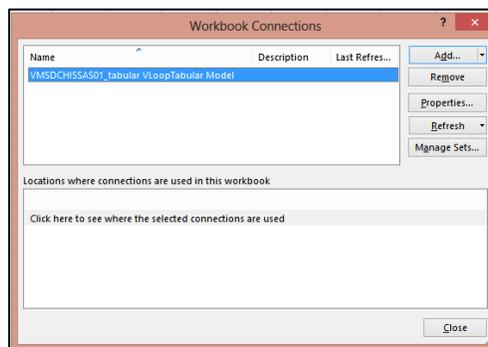


Figura 26 Power View Excel – Conexión creada

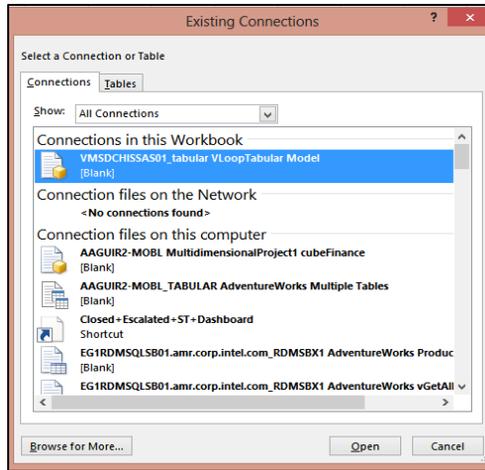


Figura 27 Power View Excel – Abrir conexión

4. Seleccionar la forma de presentación de datos, en este caso: PowerView Report

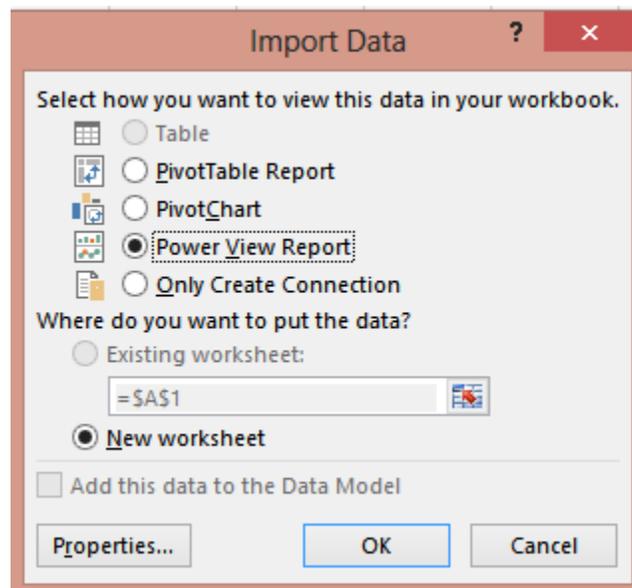


Figura 28 Power View Excel

5. Comenzar a crear el reporte según las necesidades, seleccionando los filtros requeridos, las medidas acorde a los datos que se quieren mostrar y los atributos.

Ejemplos



Figura 29 Incidentes por Región y Tipo

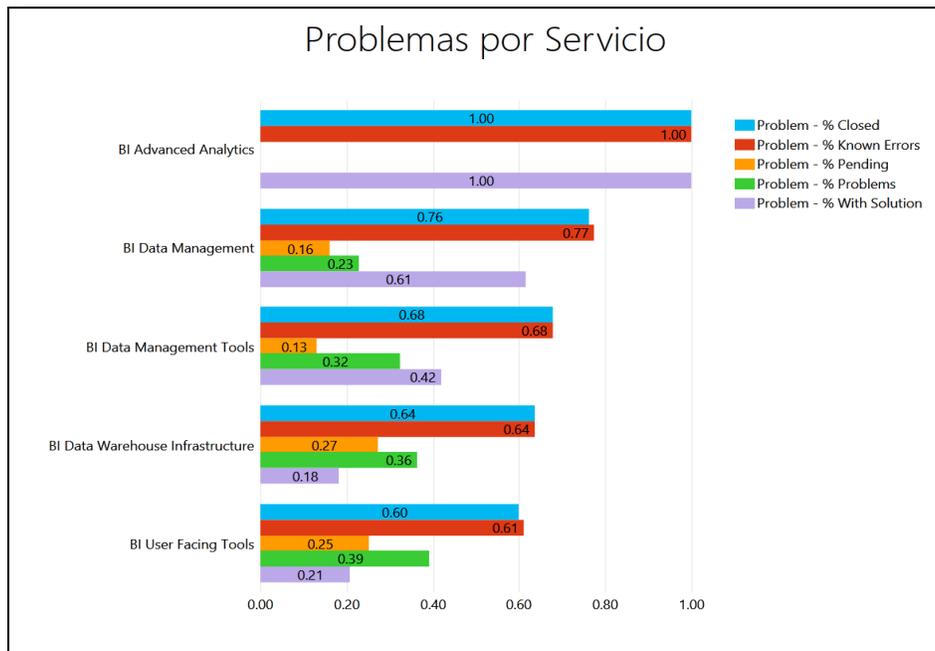


Figura 30 Problemas por Servicio

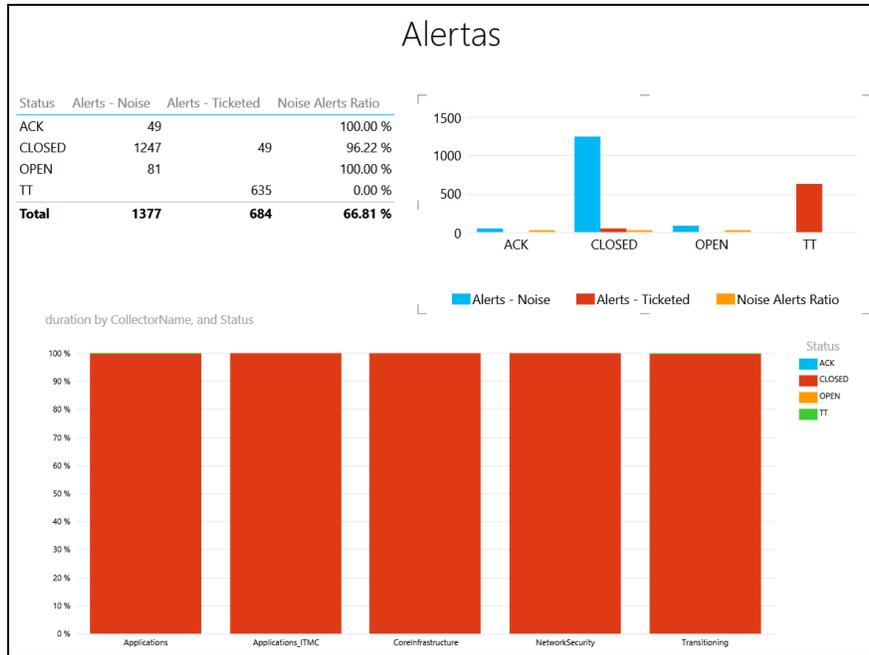


Figura 31 Alertas por Tipo

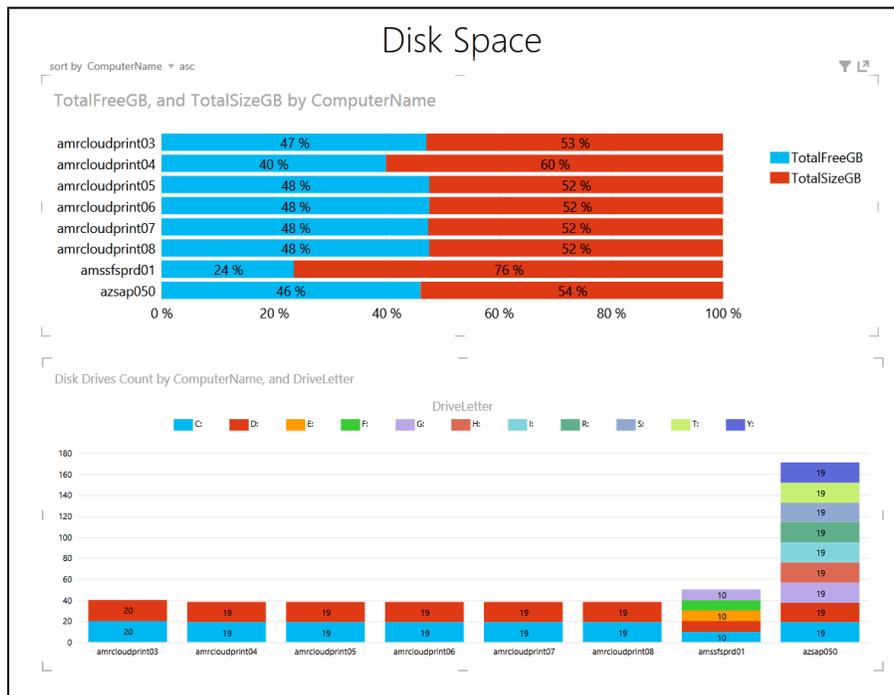
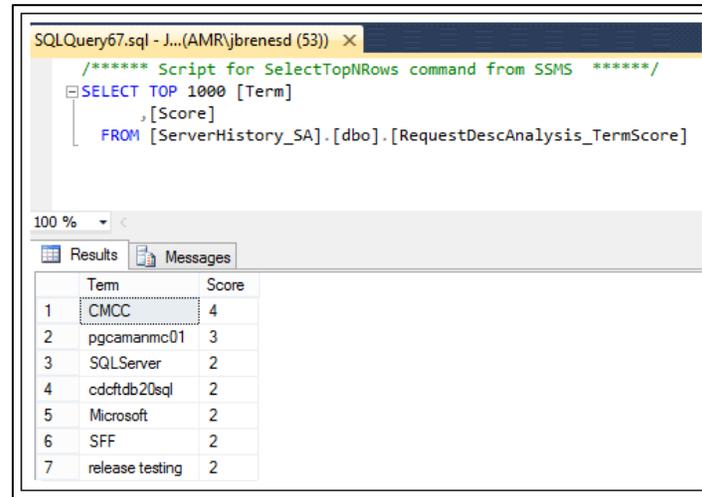


Figura 32 Espacio en Disco

## Análisis e interpretación de los resultados

### Sector Interactivo

El análisis de los resultados en el sector interactivo, sobre la aplicación del algoritmo de Minería de Texto en los campos descriptivos de cada evento o incidente en este momento generan los siguientes resultados:



The screenshot shows a SQL query window with the following code:

```

/***** Script for SelectTopNRows command from SSMS *****/
SELECT TOP 1000 [Term]
      [Score]
FROM [ServerHistory_SA].[dbo].[RequestDescAnalysis_TermScore]
  
```

Below the query, the 'Results' pane displays a table with the following data:

	Term	Score
1	CMCC	4
2	pgcamanmc01	3
3	SQLServer	2
4	cdcftdb20sql	2
5	Microsoft	2
6	SFF	2
7	release testing	2

Figura 33 Resultado de la aplicación del Algoritmo "Term Extraccion"

En la figura 17, muestra los resultados analizados por el algoritmo de Minería de Texto "Term Extraccion", donde se pueden obtener el resultado de los términos compuestos más importantes que fueron identificados por el algoritmo. En este caso en particular se puede identificar la aplicación "CMCC" como uno de los términos o frases más relevantes obtenidas por el algoritmo y que forma parte de un patrón recurrente encontrado en ese lapso de tiempo en las descripciones analizadas cada Evento. El poder identificar estos segmentos repetidos es de gran valor para el departamento Global System Monitoring (GSM) ya que permite identificar de forma rápida y oportuna las categorías gramaticales repetidas y que fueron encontradas en el campo de descripción de cada incidente.

Por otro lado, es de suma importancia obtener otro tipo de resultados que son de suma relevancia operativa para Global System Monitoring, ya que permiten visualizar e interpretar resultados de forma expedita, lo cual viene a contribuir de forma positiva en la continuidad del negocio de cada una de las aplicaciones que han confiado en Global System Monitoring (GSM), ya que le permite responder de una manera apropiada a las necesidades existentes en la infraestructura de servidores de Intel.

Por ejemplo, en la Figura 18 se puede determinar de forma rápida que algo está sucediendo de forma anormal en el siguiente servidor “fmseamweb01”, ya que en este lapso de tiempo ha generado 9 eventos. Por lo cual requiere una pronta atención por parte de Global System Monitoring (GSM) y notificación respectiva al dueño de ese servidor para informarle de que algo anómalo le está sucediendo a ese equipo.

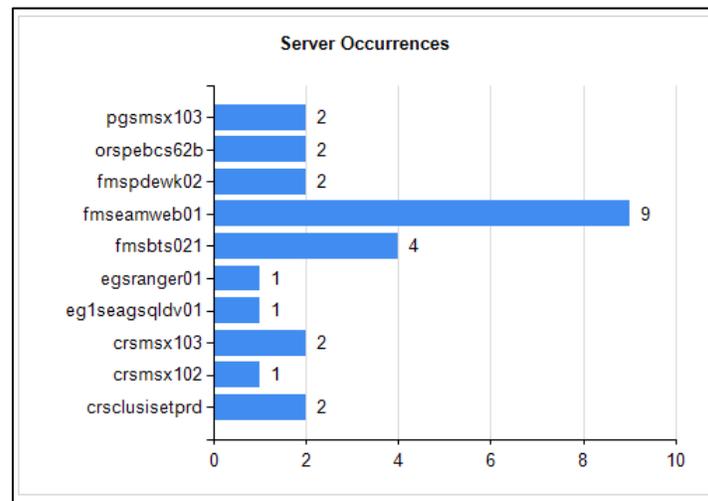


Figura 34 Recurrencias de clases de alertas

Por otro lado, el Sector Integrado al ser actualizado constantemente (~60 min) le permite a Global System Monitoring tener una visión clara de que tipo de clases de eventos están generando mayor recurrencia. En la figura 34, se identifica la siguiente clase de alerta: “Sys\_Exchange\_ABS” ha generado 19 eventos recurrentes y que necesita ser manejado apropiadamente. En este punto en particular, se podrían crear varias hipótesis de este caso, podría ser en primer lugar como si dicha recurrencia sea resultado de una inadecuada resolución por parte de los analistas de Global System Monitoring y que por lo tanto requieran una mayor capacitación u orientación al respecto. O que por lo contrario, existe un problema real en la infraestructura de la aplicación de “Exchange” que requiera una pronta acción. Es por tal motivo que esta visualización de datos en tiempo real es de suma importancia ya que permite generar varios campos de acción al respecto.

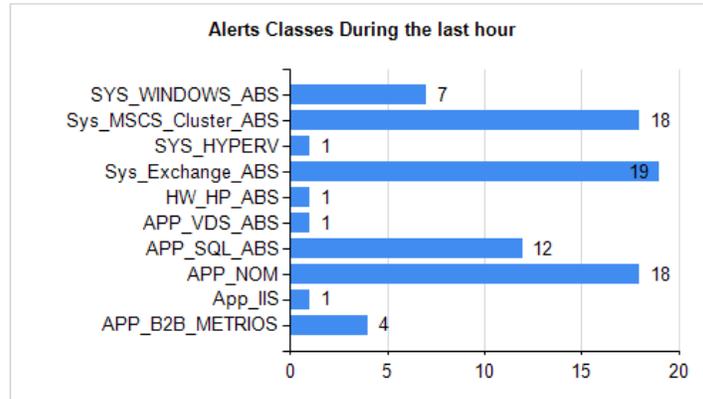


Figura 35 Recurrencias de clases de alertas

También permite apreciar que tipo de alertas están llegando en mayor cantidad a los grupos de soporte, lo cual le permite a Global System Monitoring tomar acción pronta y canalizada de los recursos específicos que debe asignar de acuerdo a la recurrencia de eventos, con fin de balancear recursos humanos en base a las necesidades existentes. En la figura 20 se puede observar que el colector de alertas llamado “Applications” ha recibido en este momento la mayor cantidad de eventos.

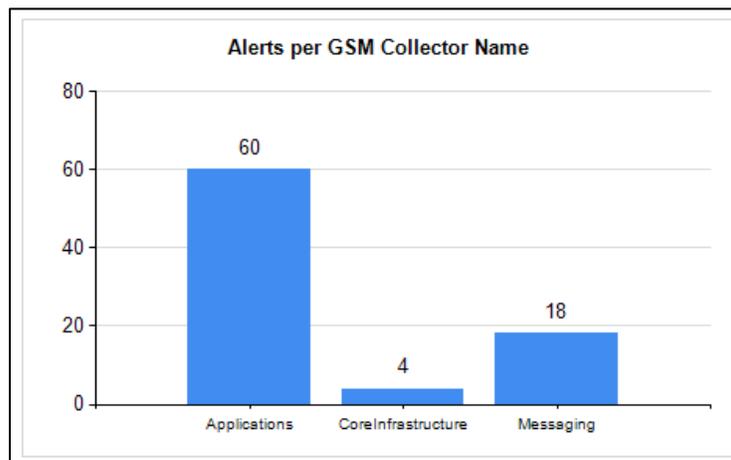


Figura 36 Alertas generadas por colectores de alertas

## Sector Integrado

Para el sector integrado mediante el dashboard se pretende ver cuál ha sido el comportamiento del servidor en el último mes, por ejemplo si hubo algún tipo de Major Incident se puede determinar si en algún punto

se estuvieron generando un volumen de alertas poco común y a su vez de incidentes. Este comportamiento podría darse por varias razones.

Por ejemplo, si un server se llegó a quedar sin espacio en el disco D: las bases de datos pueden llegar a caerse y estar corruptas, esto inmediatamente generaría un número considerable de alertas por medio de la consola y de incidentes por parte de clientes que no pueden acceder la aplicación.

Los agentes manejan un gran volumen de servidores, por lo cual es difícil saber a que aplicación puede estar afectando un problema como éste, es por ello que en la información general del servidor se incluye el nombre de la aplicación y el nombre del dueño del server; con estos datos se pretende perder minutos valiosos tratando de buscar al contacto correcto para mandar las respectivas comunicaciones y evitar que los clientes finales sigan creando más incidentes.

Como en toda organización, basada en ITIL, las métricas de manejo de incidentes, problemas, Major Incidents y cambios es de suma importancia. Es un reporte como éste podemos ver también cuáles parches fueron aplicados, los cuáles pueden tener dos efectos:

- a. Resolver alguna alerta recurrente.
- b. Generar un problema no conocido, lo cual puede llegar a convertirse en un problema ticket o en un roll back de un parche aplicado, ya que quizás no fue probado como se debía en el proceso de Change and Release Management que establece ITIL.

## Sector Archivo

El sector Archivo permite al agente de Global System Monitoring tomar todas aquellas variables relevantes para análisis de tendencias y comportamientos.

El principal objetivo del sector archivo es llegar a analizar cuáles problemas fueron correctamente cerrados e identificados con un root cause pues esto llegaría a mitigar cualquier recurrencia de un problema ya conocido evitando re trabajo.

A su vez nos permite ver una fotografía del servidor en el tiempo, por ejemplo cuál ha sido la tendencia de los discos en cuanto a espacio libre vrs espacio utilizado.

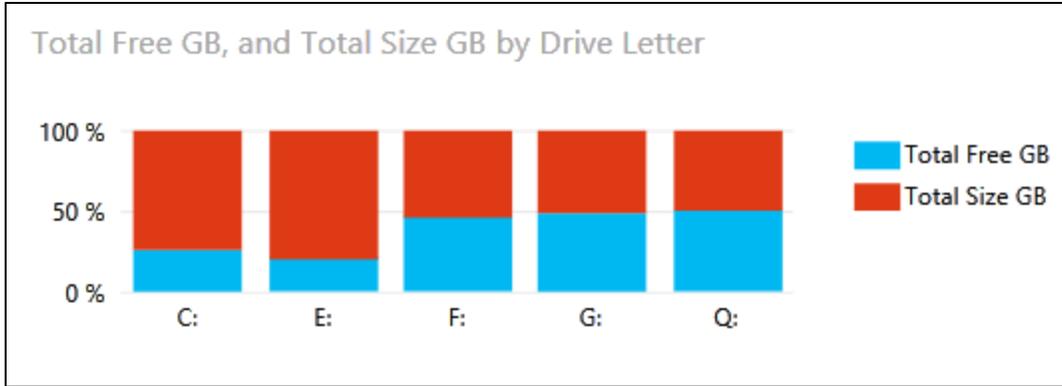


Figura 37 Espacio en Disco

El dashboard nos permite identificar cuáles parches aún están pendientes, esto significa que son servidores que no están cumpliendo con los requerimientos establecidos a nivel corporativo.

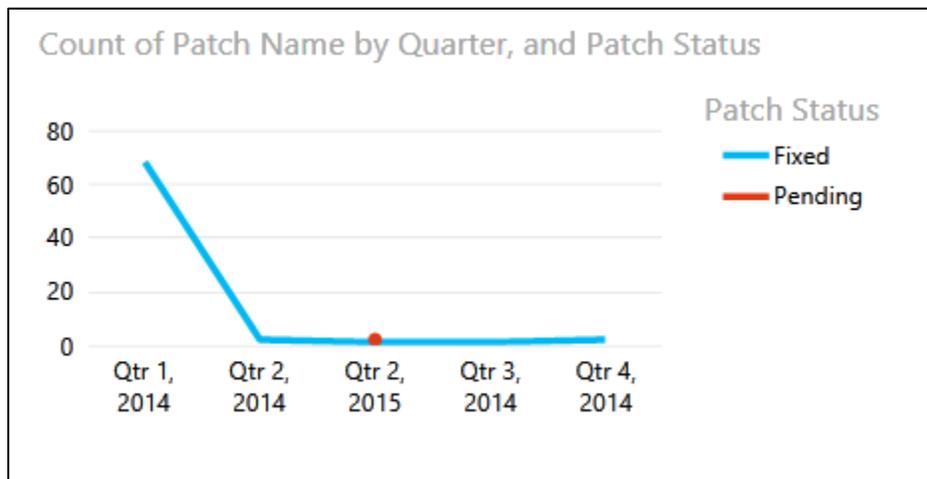
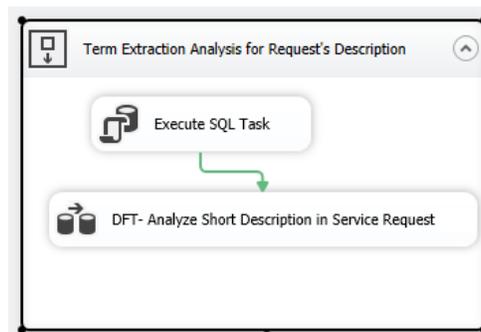


Figura 38 Parches por Estado

## Análisis mediante minería de texto

Para obtener los resultados sobre el análisis de incidentes utilizando Minería de Texto es que se utiliza en el presente proyecto la técnica llamada “Pre-procesamiento de Texto”. Esta técnica consiste en extraer las palabras utilizadas en un documento, el pre-procesamiento incluye la eliminación de los signos de puntuación y la extracción de las palabras separadas entre sí por espacios en blanco o signos de puntuación (si éstos no se han eliminado en el paso previo), fue implementada mediante el algoritmo “Term Extraction” incorporado en SSIS (SQL Server Integration Services).

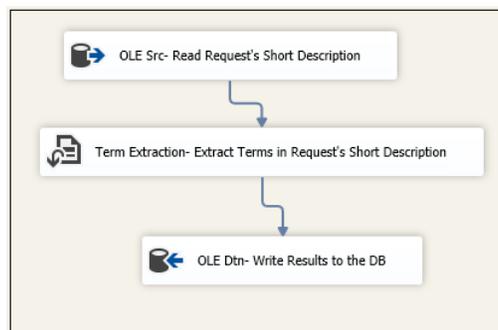
La figura 21 muestra el contenedor creado para la implementación del algoritmo de minería de texto seleccionado “Term Extracción”:



**Figura 39 Term Extraction**

**Figura 21 Paquete de Ejecución del Algoritmo de Minería de Texto llamado “Term Extraction”**

La figura 22 muestra los tres pasos mínimos requeridos para la aplicación del algoritmo “Term Extraction”: Definir Fuente de Datos, Aplicar el Algoritmos, Definir Destino con los resultados.



**Figura 40 Estructura interna del flujo de datos del Algoritmo de Minería de Texto llamado “Term Extraction”**

La figura 23 muestra la sentencia “SELECT” utilizada para la extracción de la fuente de datos, donde se especifica que recupere las descripciones de los eventos durante los últimos 60 minutos:

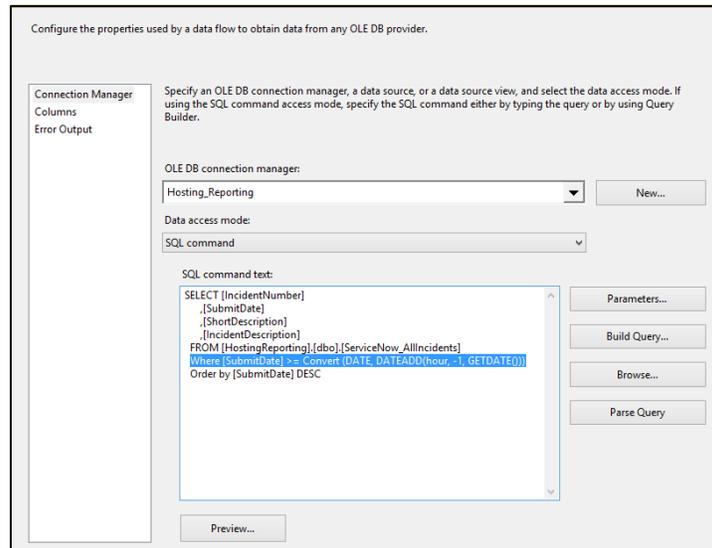


Figura 41 Paquete que extrae información de la descripción de Eventos para ser analizados por “Term Extraction”

La figura 24 muestra los campos requeridos por el algoritmo de Minería de Texto “Term Extracction”, en este paso se define el nombre de los campos necesarios para mostrar los resultados:

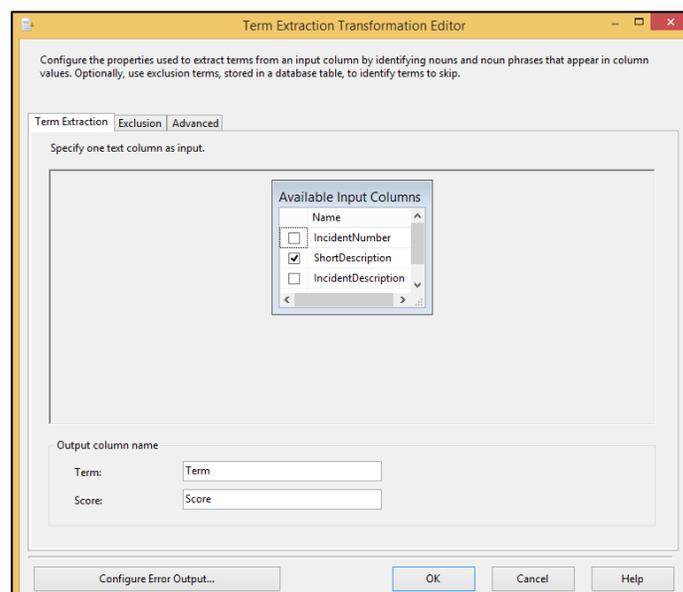
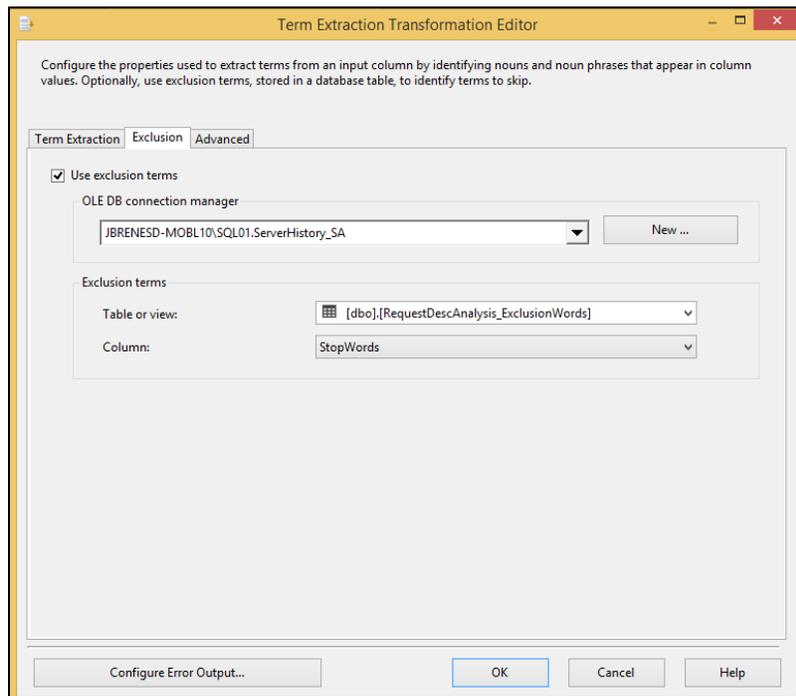


Figura 42 Selección de la columna a ser analizada por “Term Extraction”

Un aspecto importante en el pre-procesamiento es la identificación de los llamados “segmentos repetidos” o “frases” que no sean relevantes para el análisis en las Descripciones de los Incidentes generados en las

consolas de monitoreo de Global System Monitoring, por tal motivo que Global System Monitoring estableció una lista con aquellas palabras que no tienen relevancia alguna en el presente proyecto y que podrían generar ruido durante la ejecución del presente análisis. A continuación se muestra un extracto de las palabras que son consideradas como excluyentes para Global System Monitoring. La figura 25 muestra la opción donde se define la tabla fuente de los términos que van a ser excluidos.



**Figura 43** Exclusión de términos a ser analizados por “Term Extraction”

A la hora de aplicar el algoritmo de “Extracción de Términos” mediante la técnica “Pre-procesamiento de Texto” y que han permitido obtener secuencias de palabras y frases claves que tienen un significado especial para Global System Monitoring. A continuación se muestra un extracto de las palabras y frases irrelevantes para Global System Monitoring y que son producto del análisis realizado a las descripciones de requerimientos:

	StopWords
1	Server
2	Intel
3	Ping
4	VM
5	request
6	check
7	student
8	support
9	account
10	configure
11	administration
12	adm
13	status
14	loan
15	fom
16	data
17	v2
18	acct
19	service
20	gsm
21	user
22	asset
23	group
24	answer
25	non-server
26	order
27	Know

Figura 44 Definición de términos a ser utilizados por el algoritmo de Text Mining “Term Extraction”

La figura 44 permite la definición del destino donde los datos analizados por el algoritmo van a ser mostrados:

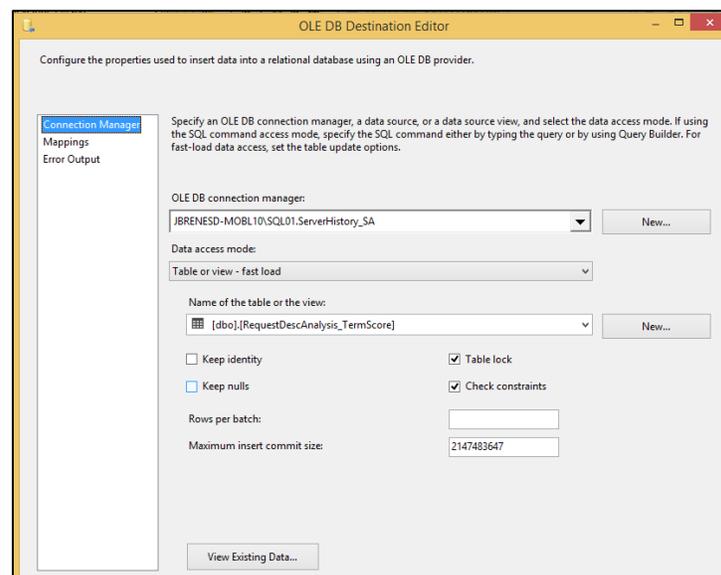


Figura 45 Selección de la Tabla con los resultados analizada por “Term Extraction”

La figura 28 muestra los resultados analizados por el algoritmo de Minería de Texto “Term Extracction”, donde se pueden visualizar el análisis de los términos compuestos importantes, que permitieron buscar el conjunto de términos o frases más relevantes que representan un contenido especial por el algoritmo y recurrente en las descripciones de cada Evento. Identificar los segmentos repetidos que aparecen en los incidentes pudo hacerse realidad mediante la implementación del algoritmo de “Extracción de Términos” que identifica las categorías gramaticales de cada palabra (sustantivo, adjetivo, preposición, verbo, etc.) encontradas en el campo de descripción de cada incidente.

Global System Monitoring puede inferir cosas importantes que son registradas en cada evento o tiquete y que en muchas ocasiones son obviadas por sus analistas, ya que en el pasado no existía un procedimiento de análisis de descripciones. Como se mencionaba anteriormente Global System Monitoring podría inferir que la aplicación llamada “Quality Center” es uno de los principales términos que se encuentran en las descripciones de los eventos y que viene a ser el principal generador de recurrencias. Lo cual le permitiría a Global System Monitoring alertar y notificar a los encargados de esta aplicación (Quality Center) sobre cualquier anomalía o situación especial que los usuarios están reportando.

	Term	Score
1	Quality Center	13165
2	FDO	11838
3	CORP Worker	10384
4	GID	4969
5	DCU	4758
6	MESC	4733
7	EMP Machine	4420
8	FLEX	4377
9	PSD	4004
10	rialto	3703
11	EAM	3519
12	ID	3304
13	Linux	2956
14	storage	2941
15	BAR	2536
16	Lab Operating System	2476

**Figura 46** Resultado del análisis de recurrencia de términos encontrados por el algoritmo “Term Extraction

# Capítulo V: Conclusiones y Recomendaciones

La Minería de Texto en un ambiente de Data Warehouse 2.0 se convierte en una herramienta de suma importancia hoy en día, ya que nos sirve para obtener información relevante y muy útil para cualquier tipo de organización en tiempo real, en especial para ser utilizadas para la toma de decisiones y análisis de tendencias. Todo esto es consecuencia del desarrollo tan abrumador que la sociedad de la información ha supuesto en cuanto a la generación de información y nuestra capacidad para almacenarla. Cada vez es más fácil recabar datos y guardarlos adecuadamente.

Global System Monitoring al ser un departamento cuya misión es la de proveer servicios de monitoreo y aseguramiento de la continuidad operativa de la infraestructura de servidores a nivel mundial, es por tal motivo considerado como un departamento esencial y de gran importancia para muchas otras departamentos. Por lo cual la aplicación de técnicas de minado de texto e implementación de arquitecturas de Data Warehouse 2.0 se perciba una nueva concepción del servicio que Global System Monitoring ofrece, ya que se están detectando prontamente patrones ocultos y elementos recurrentes en la infraestructura de Intel, con el fin de evitar que la continuidad del servicio se vea interrumpida. Es en este momento donde el proyecto ha tomado una mayor relevancia, ya que los resultados obtenidos han logrado que Global System Monitoring pueda enfocar sus esfuerzos en el análisis de eventos e incidentes en tiempo real.

Al finalizar la implementación del Data Warehouse 2.0, la aplicación del algoritmo de minado de texto y la posterior utilización para generar análisis sobre la información obtenida, es que llegamos a las siguientes conclusiones y brindar algunas recomendaciones:

## Conclusiones

- Gracias a técnicas como la Minería de Texto se ha podido ver más allá de lo que a simple vista existen en los campos descriptivos en los eventos monitoreados por Global System Monitoring. Con esta y otras técnicas de recuperación y organización de la información se ha podido explorar, analizar, comprender y aplicar el conocimiento encerrado en los textos. El reto es saber aprovechar el potencial de conocimiento escondido en ellos.

- La Minería de Texto se convierte en una herramienta de suma importancia ayudando a Global System Monitoring encontrar patrones ocultos y recurrentes en campos textos que anteriormente no se analizaban. Asimismo, nos ayuda a predecir el comportamiento de un futuro servidor, basándose en los datos históricos de incidentes que presentaron un mismo perfil.
- El campo de la minería de textos, su aplicación no siempre es sencilla y las organizaciones o administraciones que necesitan de estos estudios necesitaran contar con conjuntos de datos depurados y bien definidos para llevar a cabo estos estudios.
- Dada la existencia de varias inconsistencias de campos nulos, cause que ciertos eventos e incidentes no puedan ser tomados en consideración dentro del presente análisis, por lo tanto tienen que ser descartados del análisis hasta que los sistemas transaccionales sean corregidos, ya que podría generar problemas de inconsistencia y calidad en la información generada.
- Este esfuerzo nos permite identificar aquellos eventos críticos en términos de impacto para el negocio. Esto con el fin de documentar y capacitar a los primeros niveles de soporte en aquellos elementos que aunque no tenga mucho significado podría convertirse en un potencial elemento crítico para la organización. El modelo de DataWarehouse 2.0 ofrece una manera eficiente para el desarrollo de ésta aplicación de Business Intelligence, ya que además de un análisis histórico, permite un análisis de datos en tiempo real (quasi).
- Las técnicas de minado de texto en una plataforma Data Warehouse 2.0 han permitido a Global System Monitoring identificar patrones recurrentes dentro de un esquema de incidentes en algunos servicios, ya que por su complejidad y pocas opciones de clasificación se turnan complejos de interpretar y analizar.
- La constante actualización por parte del programa de la Maestría nos permitió identificar que un Data Warehouse 2.0 era la solución que se adaptaba al escenario requerido, ya que como hemos mencionado la parte transaccional es de suma importancia para Global System Monitoring, pero también lo era poder analizar tendencias. Es decir, el curso de Depósito de Datos Avanzados nos permitió conocer de ésta tecnología y explorarla con un modelo tabular.

## Recomendaciones

- Dada la innovación de este proyecto en Global System Monitoring, es que se solicita a la gerencia el soporte en términos de la continuidad y seguimiento de las etapas siguientes, con el fin de que la solución actual tenga un mantenimiento adecuado y de posteriores mejoras que puedan utilizarse. Es por tal motivo que se recomienda trasladar el soporte de esta iniciativa al departamento de BI de ITSO.
- Se recomienda la revisión constante de los sistemas transaccionales y de los ETL del Data Warehouse 2.0, esto con el fin de satisfacer las necesidades y reglas de negocio. Por otro lado se solicitan validaciones exhaustivas en los sistemas transaccionales para evitar el ingreso de datos que no cumplan con la calidad respectiva. Además se debe entrenar a los usuarios de dichos sistemas de manera constante para reforzar la forma adecuada en la que se debe generar los incidentes y eventos en el sistema transaccional.
- Basado en los elementos recurrentes identificados por la presente solución es que se recomienda a Global System Monitoring la generación de actividades proactivas para la reducción de eventos a través de la creación de documentos técnicos y capacitación a los niveles de primer soporte para que resuelvan esos incidentes a primera instancia.
- Las tareas de pre-procesamiento de la información (incidentes) tomaron un valor determinante en el minado de texto, ya que nos permitieron depurar la calidad de la información de los datos a ser analizados.

# Capítulo VI: Referencias Bibliográficas

## Bibliografía

- Foundation, I. (s.f.). *ITIL v3*. Obtenido de ITIL Foundation Gestión de Servicios de TI: [http://itilv3.osiatis.es/operacion\\_servicios\\_TI/gestion\\_problemas/control\\_proceso.php](http://itilv3.osiatis.es/operacion_servicios_TI/gestion_problemas/control_proceso.php)
- García, J. (2014). *Introducción de la minería de textos como técnica de recuperación y organización de la información*. Obtenido de Text Mining: <http://textmining.galeon.com>
- Inmon, B. (2014). *Corporate Information Factory*. Obtenido de Inmon CIF: <http://inmoncif.com/home/>
- Kempter, A. (2014). *Gestión de Incidentes - Visión General*. Obtenido de ITIL: [http://itil.osiatis.es/Curso\\_ITIL/Gestion\\_Servicios\\_TI/gestion\\_de\\_incidentes/vision\\_general\\_gestion\\_de\\_incidentes/vision\\_general\\_gestion\\_de\\_incidentes.php](http://itil.osiatis.es/Curso_ITIL/Gestion_Servicios_TI/gestion_de_incidentes/vision_general_gestion_de_incidentes/vision_general_gestion_de_incidentes.php).
- Kobiellus, J. (01 de Mayo de 2008). *TechWorld*. Obtenido de Data Warehouse 2.0: [http://www.techworld.com.au/article/214248/data\\_warehouse\\_2\\_0/?fp=4&pp=1&fpid=128](http://www.techworld.com.au/article/214248/data_warehouse_2_0/?fp=4&pp=1&fpid=128)
- Microsoft. (2014). *Analysis Services*. Obtenido de Microsoft Technet: <https://technet.microsoft.com/es-es/library/ms143708.aspx>
- Microsoft. (2014). *Developer Network*. Obtenido de MSDN Microsoft: <https://msdn.microsoft.com/es-es/library/ms174758.aspx>
- Microsoft. (2014). *Developer Network*. Obtenido de MSDN Microsoft: <https://msdn.microsoft.com/es-es/library/hh213101.aspx>
- Microsoft. (2014). *Developer Network*. Obtenido de Microsoft: <https://msdn.microsoft.com/es-es/library/ms144884.aspx>
- Microsoft. (2014). *Modelo Tabular*. Obtenido de Microsoft MSDN: <https://msdn.microsoft.com/es-es/library/hh212945.aspx>
- Microsoft. (s.f.). *Developer Network*. Obtenido de <https://msdn.microsoft.com/es-es/library/ms159106.aspx>
- Microsoft. (s.f.). *Office*. Obtenido de <https://support.office.com/es-hn/article/Power-View-explorar-visualizar-y-presentar-los-datos-98268d31-97e2-42aa-a52b-a68cf460472e>

Naranjo Zeledón, L. (2014). Metodologías de Investigación Introducción a los abordajes tradicionales y alternativos.

Parcell, D. (2012). *Datos no estructurados: más allá del CRM*. Obtenido de PuroMarketing: <http://www.puromarketing.com/53/11699/datos-estructurados-alla.html>

Petersen, K. (2008). Systematic Mapping Studies in Software Engineering. . En K. Petersen.

Statsoft. (2014). *Text Mining*. Obtenido de Statsoft: <http://www.statsoft.com/textbook/text-mining>

Textmining. (2014). *Minería de Textos*. Obtenido de Textmining: <http://textmining.galeon.com/>

# Capítulo VII: Anexos

## Anexos

### Work Breakdown Structure (WBS)

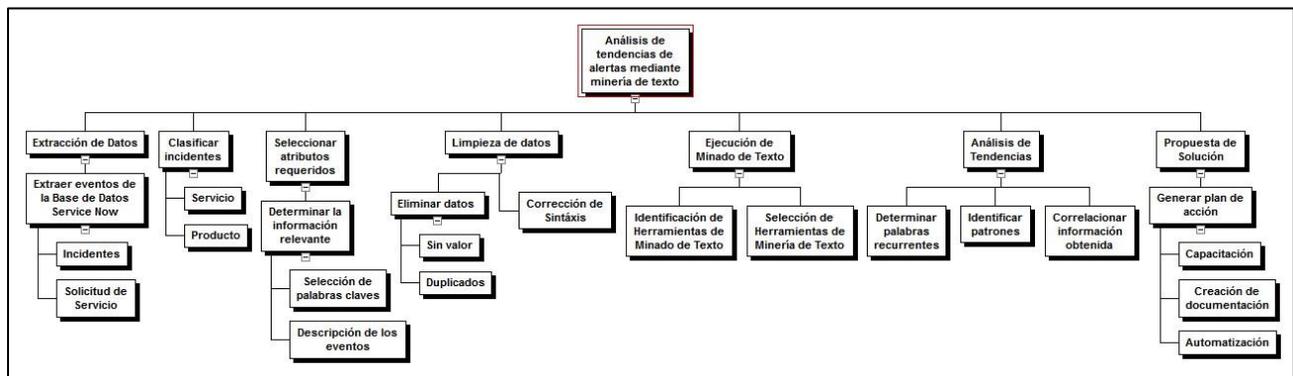


Figura 47 Work Breakdown Structure