Alta Disponibilidad 24/7: El Reto

Carlos Edgardo López Grande.[1]
Ricardo Salvador Guadrón G.[2]

Resumen

En la actualidad, las actividades diarias para los negocios, instituciones, instituciones educativas o nivel de usuario común están directamente vinculadas a la red de redes, la Internet. Su desarrollo ha permitido a la mayoría de los servicios y aplicaciones, que hace años eran sólo para el modo cliente-servidor, ahora acceder desde cualquier lugar en cualquier dispositivo con conexión a Internet.
Por lo tanto, ofrecer un ambiente que da conexiones y transacciones 7/24 es necesario para cada empresa e institución con servicios en línea. Sin embargo, es una premisa difícil de cumplir.

Palabras clave:

Http, servicio al cliente 24 / 7, conectividad, internet (red de computadores), monitoreo.

Introducción

Jaime, un estudiante de modalidad virtual se dispone a realizar su evaluación final a las 10:00 P.M. Relajado y después de haber estudiado enciende su computadora y busca en los favoritos de su navegador de Internet el enlace que lo conecta a su aula virtual, sin embargo, esta noche no es como todas las noches, recibe un mensaje en el navegador: "503 Service Unavailable". Inmediatamente y preocupado por la realización de su examen final decide contactar a su docente, pero éste le replica que desconoce si el servidor del aula virtual funciona o no funciona y que no puede hacer nada.
Existen códigos de estado HTTP [3] que nos orientan a determinar, desde la perspectiva del cliente, el tipo de problema que pueda estar sucediendo con el servidor y se clasifican de la siguiente manera:

- **Respuestas Informativas (1XX)**
- **Peticiones Correctas (2XX)**
- **Redirecciones (3XX)**
- **Errores de Cliente (4XX)**
- **Errores de Servidor (5XX)**

Para nuestro caso nos enfocaremos en algunos códigos de estado HTTP referentes a Errores de Servidor (5XX):

- **500 Internal Server Error:** Mensaje de error genérico que indica que existe una condición inesperada al intentar conectarse con el servidor.El mensaje no especifica más.
- **502 Bad Gateway / 504 Gateway Timeout:** Mensaje que se muestra cuando un servidor (posiblemente el webserver) esté trabajando como puerta de enlace o Proxy para cumplir la solicitud del cliente para acceder a la URL requerida y recibe una respuesta inválida del servidor.
- **503 Service Unavailable:** El servidor no está disponible para atender la solicitud HTTP debido a una sobrecarga temporal o mantenimiento al servidor.
- **520 Origin Error (Cloudflare[4]):** Especifica que hay un problema de conexión desconocido entre Cloudfare y el servidor al que quiere conectarse.
- **521 Web Server is Down (Cloudflare):** Indica cuando el servidor de origen rechaza la conexión debido a que se encuentra fuera de servicio.

(1) Técnico en Mantenimiento y Servicio de Computadoras. Docente Escuela Ingeniería Eléctrica y Electrónica, modalidad presencial y semipresencial. Escuela Especializada en Ingeniería ITCA-FEPADE, El Salvador, Centroamérica. ***email: carlos.lopez@itca.edu.sv.*** *(2) Ingeniero Electricista. Director Escuela Ingeniería Eléctrica, Electrónica y Computación. Escuela Especializada en Ingeniería ITCA-FEPADE, El Salvador, Centroamérica.* ***email: rguadron@itca.edu.sv*** *(3) Los códigos de estado HTTP están especificados por los estándares RFC 2616, RFC 2518, RFC 2817, RFC 2295, RFC 2274 y RFC 4918, otros no están estandarizados, pero son comúnmente utilizados.* ***(4)*** *CloudFlare es un servicio que funciona como intermediario entre los visitantes y el servidor web. Almacena una copia del sitio web en cache y lo distribuye por todo el mundo para permitir a los visitantes un acceso rápido y reduciendo la carga soportada por el servidor web.*

- ✓ **522 Connection Timed Out (Cloudflare):** Especifica que el tiempo de respuesta del servidor ha sido excedido debido a fallas de comunicación.

Como se mencionó antes, estos mensajes ayudan a orientar en cierta medida a encontrar el problema que pueda estar sucediendo. Es necesario aclarar que todos estos mensajes de error no están relacionados solamente a problemas de red como la mayoría cree.

Ahora, como administradores de nuestros servicios ¿Cómo hacemos para determinar la causa de la caída de nuestro servicio? ¿Podría determinarse la causa del problema antes que el cliente haga la solicitud de servicios? ¿Estos mensajes brindan la información adecuada para la solución del problema? Para responder estas interrogantes es necesario que conozcamos los conceptos que definiremos a continuación.

ALTA DISPONIBILIDAD

Es un modelo de diseño de infraestructura de red, servicios y sistemas que asegura en cierta medida la continuidad operacional en un período de tiempo determinado, es decir, la capacidad de brindar al usuario de forma ininterrumpida acceso al sistema o a los servicios a lo largo del tiempo. Cuando el usuario no tiene acceso a dichos servicios, entonces se dice que no está disponible (downtime [5]).

Un Datacenter [6] posee alta disponibilidad cuando tiene la capacidad de proveer los servicios 7/24 a sus usuarios. Sin embargo, para realmente garantizarla existen otros requerimientos descritos en los estándares TIA-942 de la Asociación de la Industria de las Telecomunicaciones (TIA por sus siglas en Inglés) y que están relacionados a aplicaciones y procedimientos tales como:

- **Arquitectura de red.**
- **Diseño del sistema eléctrico.**
- **Sistemas de redundancia.**
- **Control de acceso y seguridad en la red.**
- **Control ambiental.**
- **Protección contra amenazas físicas (incendios,inundaciones, entre otros)**
- **Administración de las alimentaciones de corriente eléctrica**

Bajo estos requerimientos, la TIA ha definido cuatro niveles de Datacenter y cada uno de ellos basada en la disponibilidad que ofrecen:

A) Nivel 1: Susceptible

Cumple con un 99.671% de disponibilidad, entre sus principales características tenemos:

a) Susceptible a interrupciones planeadas o no planeadas.

b) Ruta única de alimentación eléctrica y distribución de enfriamiento no redundante.

c) Con posibilidad de tener: piso elevado, UPS o generador.

d) Downtime anual de 28.8 horas (5 minutos diarios).

e) Debe ser apagado completamente para realizar mantenimiento preventivo.

B) Nivel 2: Componentes redundantes

Cumple con un 99.741% de disponibilidad, entre sus principales características tenemos:

a) Menos susceptible a interrupciones planeadas o no planeadas.

b) Ruta única de alimentación eléctrica y enfriamiento, incluye redundancia.

c) Incluye: piso elevado, UPS y generador.

d) Downtime anual de 22 horas (3.66 minutos diarios).

(5) Tiempo de inactividad usado para definir cuando el sistema no está disponible. (6) Área de las tecnologías de la información donde se encuentran los sistemas principales de procesamiento de datos, las telecomunicaciones y el almacenamiento, sistemas redundantes de energía, comunicaciones redundantes. La operación de estos servicios no puede parar.

C) Nivel 3: Mantenimiento concurrente

Cumple con un 99.982% de disponibilidad, entre sus principales características tenemos:

a) Permite actividades planeadas de mantenimiento sin interrumpir la operación, pero eventos sin planearse aún interrumpen la operación.

b) Existen múltiples rutas de alimentación eléctrica y enfriamiento. Hay redundancia, pero solo una ruta activa

c) Incluye: piso elevado y suficiente capacidad de llevar carga una ruta de distribución.

d) Downtime anual de 1.6 horas (0.3 minutos diarios).

D) Nivel 4: Tolerancia a fallos

Cumple con un 99.995% de disponibilidad, entre sus principales características tenemos:

a) Las actividades planeadas o no planeadas no afectan la operación, ininterrumpidamente ofrece los servicios.

b) Múltiples rutas activas de enfriamiento y alimentación eléctrica.

c) Downtime anual de 0.4 horas (0.06 minutos diarios).

Si bien es cierto que las características de diseño físicas y lógicas de estos Datacenter permiten mantener estos altos grados de disponibilidad, no significa que se hagan por arte de magia. Siempre existen amenazas que de suceder, comprometen la integridad en la entrega de los servicios a los usuarios.

AMENAZAS

Las amenazas a los Datacenter pueden clasificarse en dos grandes categorías, dependiendo del destino al que se dirijan:

1. Amenazas lógicas o digitales: Son todas aquéllas que atentan contra la infraestructura lógica del Datacenter, como por ejemplo: virus, hackers, cuellos de botella en las redes por sobrecarga de conexiones entrantes y otras actividades, ya sea accidentales o maliciosas que afecten directamente el flujo de datos.

2. Amenazas físicas: Son todas aquéllas que comprometen directamente a los equipos y su causa puede estar ligada a distintos ámbitos, como problemas de enfriamiento, de alimentación eléctrica, con los enlaces de datos, errores humanos, actividades maliciosas, incendios, inundaciones y toda actividad que afecte directamente la integridad física del Datacenter.

La alta disponibilidad que se requiere para ofrecer los servicios del Datacenter demanda el control de todas estas amenazas y que, de surgir alguna de ellas, se pueda tener el tiempo suficiente para solventarla antes que el usuario acceda a los servicios.

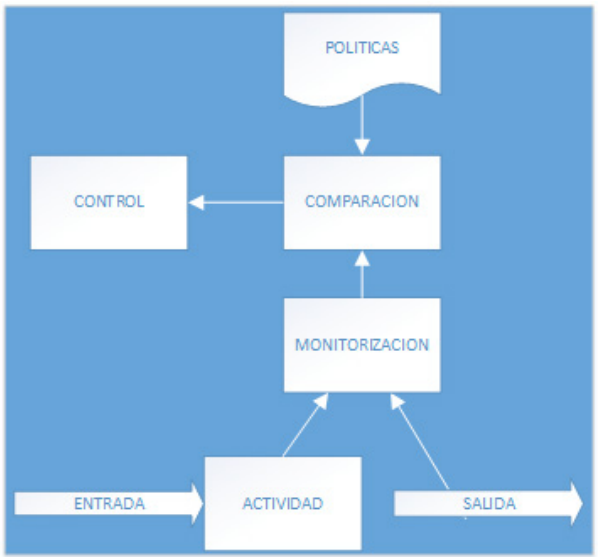

Fig. 1. Ciclo de vida monitorización

NETWORK OPERATION CENTER (NOC)

En nuestro país existe una cultura de “apaga fuegos” en la que estamos acostumbrados a que sea el usuario el primero en identificar los problemas y nosotros, los “bomberos”, los administradores de

nuestros servicios; esperamos que nos avisen para comenzar a trabajar en la solución al problema donde en muchos casos, el "fuego" se vuelve incontrolable y puede dejar pérdidas irreparables.

¿Por qué esperar a que sea el usuario el primero en enterase del problema? ¿Por qué no enfrentar estas amenazas de forma proactiva? Muchas de estas amenazas en algún momento pueden estar fuera de nuestro alcance, pero si logramos identificarlas con tiempo, podemos lograr que las consecuencias sí puedan controlarse.

Monitorizar nuestro Datacenter a través de un Centro de Operaciones de Red (NOC por sus siglas en Inglés) nos permitirá tener un mejor control de los servicios, equipos y amenazas que puedan en algún momento afectar el flujo de datos, pero, ¿Qué es un NOC?

Un NOC es la ubicación central responsable de la administración y monitorización del Datacenter, que al mismo tiempo actúa como punto de contacto para todas las solicitudes de servicio relacionadas con estos elementos de configuración. El alcance de sus tareas es relativamente amplia, ya que no sólo se enfoca en la monitorización de redes, sino que puede llegar a detectar la sobrecarga de los procesadores de un servidor, las capacidades de almacenamiento de sus discos duros o incluso, si la temperatura del Datacenter es la adecuada para su correcto funcionamiento.

La gran pregunta ahora es: ¿Qué y cómo debo monitorizar? Lo primero que debemos hacer es identificar cuáles son los servicios tecnológicos críticos en nuestra red, como por ejemplo el Sistema de Nombres de Dominio DNS por sus siglas en Inglés, Aplicaciones, Bases de Datos BD, Active Directory, entre otros. También, debe definirse el método de notificación que utilizaremos para identificar los eventos que se generen en la red y además, es necesario tener como buena práctica el resguardo de los logs de eventos. ¿Por qué? Porque cada dispositivo en nuestro Datacenter e infraestructura de red, como servidores Linux, Unix, Microsoft, Switches y Routers, permiten reportar eventos a través de Syslog[7]. Monitorizando y recolectando estos logs, los eventos que se generan en la red se tratan de forma proactiva y reactiva. Sin embargo, no se debe monitorizar sólo por monitorizar, el hecho de que estos equipos permitan la generación de logs, no significa que los problemas están resueltos. Debemos convertir estas herramientas en entidades inteligentes que nos permitan resolver las amenazas generadas en la infraestructura.

La monitorización cumple con un ciclo de vida como se ilustra en la Fig. 1.

Debe buscarse unificar esfuerzos para correlacionar los eventos generados en la infraestructura de la organización. Por ejemplo, si la BD ha fallado, identificar qué correlación tendrá si el área de BD tiene distintas reglas o políticas que los del área de servidores. Para ello primero se deben establecer políticas o reglas muy claras:

- ¿Quién va a monitorizar?
- ¿Cómo se debe monitorizar?
- ¿Cuáles son las reglas que todos los de IT deben estar siguiendo?
- ¿Dónde se van a almacenar los datos?
- ¿Cómo se harán las correlaciones cuando sucedan fallas en la infraestructura?

Una vez hechas las políticas, se deben crear los **mecanismos necesarios para comparar** los eventos que sucedan contra las políticas creadas y entender así qué es lo que está pasando en nuestra red.

Una vez que entendemos lo que está pasando en nuestra red deben definirse **mecanismos de control** para evitar que los eventos afecten la actividad normal de la infraestructura.

En base a estos controles, se debe establecer una **serie de actividades** que pueden ser proactivas y/o reactivas que surjan a partir de alguna solicitud de cambio.

Estas actividades nos permitirán que, a través de la **monitorización de sus entradas y salidas**, podamos obtener los insumos necesarios para definir correctamente los mecanismos de control.

(7) Por Syslog se conoce tanto al protocolo de red como a la aplicación o biblioteca que envía los mensajes de registro. Un mensaje de registro suele tener información sobre la seguridad del sistema, aunque puede contener cualquier información. Junto con cada mensaje se incluye la fecha y hora del envío.

Existen dos tipos de ciclos de monitorización:

1. **Sistema Abierto:** ejecutan una actividad específica sin importar las condiciones ambientales que la rodean. No importa qué esté alrededor, al detectar un evento, se ejecuta una acción. Por ejemplo: un backup puede ser iniciado en cualquier momento sin importar otras condiciones, solamente es necesario que se automatice y el backup se realizará según las configuraciones con las cuales fue definido (hora, cada cuanto tiempo, tipo de backup, entre otros).

2. **Sistema Cerrado:** son los que monitorizan un ambiente y responden a sus cambios. Por ejemplo: en cargas de red, balancear la monitorización evaluará el tráfico en un circuito. Si el tráfico de la red excede un cierto rango, el sistema de control empezará a redirigir el tráfico a través de un circuito de respaldo.

¿Cómo realizar esta monitorización? ¿Qué herramientas pueden utilizarse para poder realizarla con éxito en el NOC?

HERRAMIENTAS UTILIZADAS EN EL NOC

Existen distintos tipos de herramientas que pueden ser utilizadas en el NOC, tales como:

- **Herramientas de Monitorización:**
  Tienen la capacidad de poder trabajar en segundo plano y que nos ayudan principalmente a recolectar todo tipo de eventos. Estas herramientas tienen la característica de poder generar salidas de manera calendarizada (que nos envíen logs o reportes al correo electrónico o dispositivo móvil).

- **Herramientas de Diagnóstico:**
  Nos permiten comprobar la conectividad, como por ejemplo verificar si algún servidor, componente de la red o algún dispositivo está trabajando. Normalmente las herramientas de diagnóstico son herramientas activas, ya que trabajan 24/7 de manera constante con la finalidad de identificar patrones o comportamientos específicos para poder emitir algún tipo de alarma, alerta o comunicado para poder solucionar de manera preventiva todos estos incidentes.

Todos los sistemas operativos incluyen herramientas o comandos que en alguna medida permiten realizar un monitoreo básico con respecto a conectividad. Por ejemplo, el comando **"Ping"**, permite verificar la conexión desde un punto de la infraestructura de red a otro. Tenemos también el comando **"Tracert"** o **"Trace Route"**, que permite saber en qué punto específico de la ruta de red se ha perdido conectividad. Estos comandos son útiles a la hora de verificar si no se ha perdido la conexión con nuestros servicios. Sin embargo, así como se mencionó al inicio de este artículo, no todos los problemas generados en un Datacenter están relacionados a la conectividad. ¿Cómo detectar que el error de la transacción, del inicio de sesión, el envío de correo o cualquier otro servicio fue generado por el poco espacio de almacenamiento en el disco duro o por una sobrecarga en el microprocesador? Para dar respuesta a estas preguntas, podemos encontrar en el mercado actual distintas alternativas, algunas de uso gratuito y otras a través de la compra de una licencia.

Se presentan a continuación 2 alternativas de software para poder realizar la monitorización de nuestro Datacenter, el corazón de nuestro modelo de negocios, haciendo mucho más eficiente la entrega de los servicios a los usuarios finales:

1. **Cacti:** Haciendo uso del protocolo SNMP[8] y a través de una plataforma web diseñada con PHP, Cacti almacena en una base de datos diseñada en MySQL toda la información que generen los dispositivos de la infraestructura que tengan habilitado el protocolo SNMP gracias a la herramienta RRDTools en la que se basa para la administración de la información. Esta gestión ayuda a hacer más eficiente el monitoreo de la red, ya que a través de la captura de datos, los administradores del Datacenter y servicios pueden identificar los fallos o eventos que

(8) El Protocolo Simple de Administración de Red o SNMP es un protocolo de la capa de aplicación que facilita el intercambio de información de administración entre dispositivos de red. Los dispositivos que normalmente soportan SNMP incluyen routers, switches, servidores, estaciones de trabajo, impresoras, bastidores de módem y mucho más. Permite a los administradores supervisar el funcionamiento de la red, buscar y resolver sus problemas, y planear su crecimiento.

se generen, pudiendo encontrar la solución al problema de manera más rápida. A pesar de ser una aplicación de distribución libre, Cacti ofrece muchas más capacidades que Smokeping. Podemos monitorear sensores de temperatura, voltajes específicos de fuentes, UPS o sistemas de alimentación eléctrica, switches, routers, entre otros. Todo aquel dispositivo que pueda comunicarse a través del protocolo SNMP, puede ser monitoreado por CACTI Fig. 2.

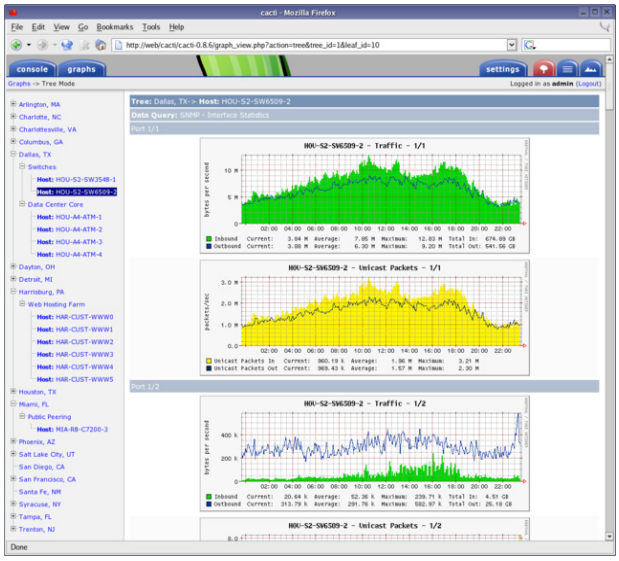

Fig. 2. Consola administración web de Cacti

2. **Paessler Router Traffic Grapher PRTG Network Monitor:** Es una aplicación desarrollada por la empresa Paessler. Una herramienta innovadora que permite la monitorización de redes desde cualquier dispositivo que tenga acceso a Internet. Tiene la posibilidad de manejar una consola de administración local, consola a través de la web y además tiene aplicaciones desarrolladas exclusivas para IOs y Android.

PRTG solamente puede ser instalado en plataformas Microsoft; sin embargo, permite monitorizar servidores Linux, Unix y MacOS. Utiliza distintos protocolos para monitorizar las redes, tales como SNMP, WMI, packet sniffing, etc., además, permite monitorizar QoS, sitios Web, correo electrónico, bases de datos y servidores virtuales.

Permite la monitorización de redes remotas y para redes de cualquier tamaño. Reconoce los dispositivos de la infraestructura de forma automática y monitoriza sistemas multiprocesador.

Permite además notificar las alertas oportunamente utilizando distintos canales: Correo electrónico, mensajes de texto al celular, notificaciones en los smartphones que tengan instalada la aplicación, peticiones HTTP, por scripts, syslogs, etc. Genera reportes personalizables y periódicos sobre las incidencias que se generen dentro de la infraestructura.

Existen diferentes versiones de la aplicación [9], cada una de ella brindando las características que correspondan a las mismas:

- PRTG 100 ($ 440.00)
- PRTG 500 ($ 1,600.00)
- PRTG 1000 ($ 2,700.00)
- PRTG 2500 ($ 5,600.00)
- PRTG 5000 ($ 9,500.00)
- PRTG Unlimited ($ 13,500.00)
- PRTG Corporate Country ($ 40,500.00)
- PRTG Corporate 5 Core Global ($ 47,250.00)

Paessler además brinda la posibilidad de tomar un plan de soporte y mantenimiento, que permite a los clientes mantenerse actualizado con las últimas versiones sin pagar montos adicionales de dinero mientras el plan de mantenimiento se encuentre vigente. Los planes están dispuestos para 12, 24 y 36 meses.

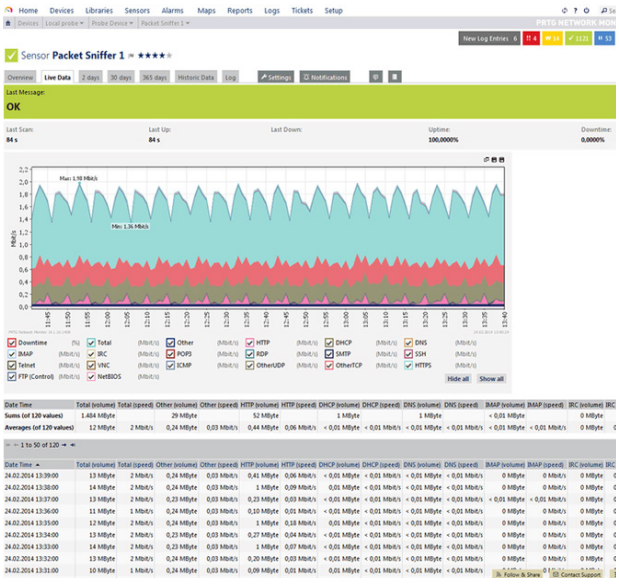

Fig. 3. Consola de PRTG – Consumo ancho de banda

(9) https://shop.paessler.com/shop/prtg/new/#

Hay opciones para quienes no pueden adquirir el software o que no son organizaciones como para hacer inversiones tan grandes.

Se puede descargar una versión de prueba de 30 días de la aplicación y probar así todas las funcionalidades de PRTG, o también se puede descargar una versión Freeware que permite administrar 10 dispositivos en nuestra infraestructura sin tener un límite de tiempo, para siempre.

ESCENARIO

Monitoreando un servidor Windows 2008 Server con la aplicación de monitorización RPTG se pueden obtener varios reportes no solamente de la conectividad a la red, sino también alertas y gráficas sobre el rendimiento del procesador, la salud del sistema o el espacio de almacenamiento físico. Se configuró el envío de correo electrónico notificando alertas sobre el disco duro. A continuación se presentan imágenes del resultado.

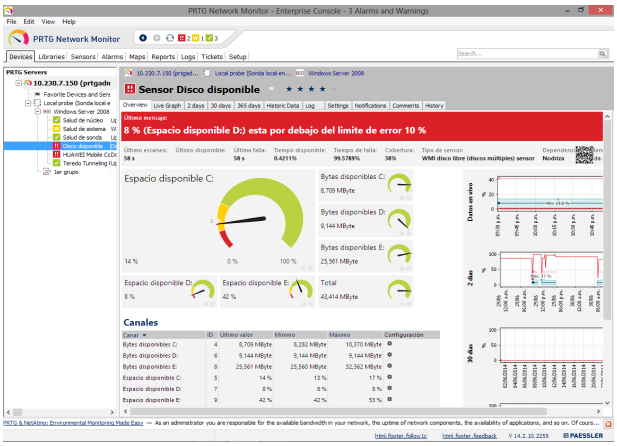

Fig. 4. Problemas espacio de almacenamiento Windows Server 2008

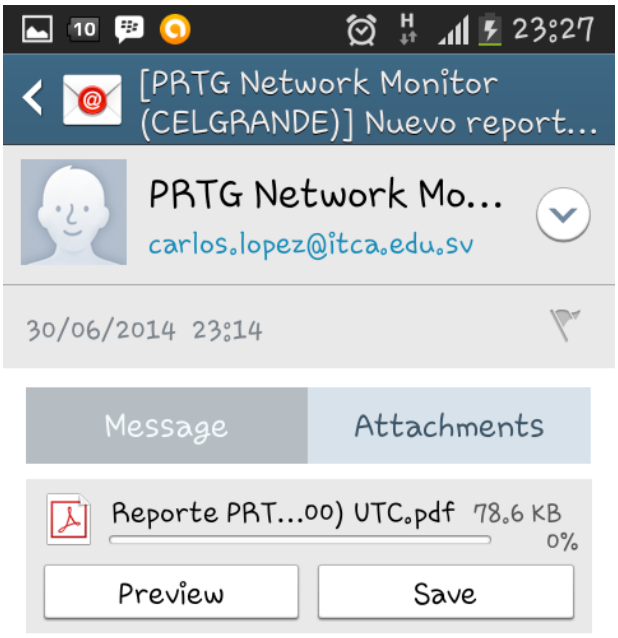

Fig. 5. Reporte de Servicios enviado a correo en PDF

Fig. 6. Correo de alerta sobre espacio de almacenamiento en disco duro.

Conclusiones

Conociendo estas herramientas, no es aceptable que los involucrados en IT desconozcan el estado de su Datacenter. No es posible que en estos tiempos seamos siempre los últimos en darnos cuenta que nuestros servicios no se están brindando de forma adecuada. No es posible que, teniendo a la mano soluciones, siempre seamos los que buscamos las excusas.

Independientemente del modelo, el negocio va creciendo junto a la satisfacción y el bienestar de los usuarios o clientes de los servicios.

Como clientes de cualquier servicio, siempre esperamos que cuando lo solicitemos, se nos brinde de forma eficiente. "Hagamos con los demás lo que nosotros queremos que hagan con nosotros", seamos oportunos para identificar los problemas que puedan generarse en nuestro Datacenter y resolvámoslos de forma proactiva. No hay excusas. Existe una variedad de herramientas que nos permiten mantener monitorizados todos nuestros servicios, que pueden ir desde las de distribución libre hasta aquéllas en las que haya que realizar una fuerte inversión económica. **El menú está servido ¿Qué vamos a ordenar?**

Bibliografía

- Checkupdown. 10 de abril 2014<http://www.checkupdown.com/default_es.html>
- Editor RFC (Request for Comments series). 15 de abril 2014<http://www.rfc-editor.org/>
- OpenWebCMS, ¿Qué es Cloudflare y cómo puede ayudarle?. 12 de abril 20114 <http://openwebcms.es/2012/que-es-cloudflare-y-como-puede-ayudarte/>
- Wikipedia, Estándares TIA-942. 1 de Enero de 2014 <http://en.wikipedia.org/wiki/TIA-942>
- Wikipedia, Latencia. 18 de Mayo de 2014 <http://es.wikipedia.org/wiki/Latencia>
- Smokeping. 18 de Mayo de 2014 <https://wiki.archlinux.org/index.php/smokeping>
- Wikipedia, RRDTool, 17 de Mayo de 2013 <http://es.wikipedia.org/wiki/RRDtool>