

Solución tecnológica para el monitoreo del sistema de recuperación de información de la Plataforma C.U.B.A.

Tesis presentada en opción al título académico de
Máster en Informática Avanzada

Autor

Yadiel Pérez Villazón

Tutor

Dr. Ramón Santana Fernández

Consultor

Ing. Paúl Rodríguez Leyva

La Habana, 2018

Año 57 de la Revolución



Declaración de autoría

Declaro por este medio que yo Yadiel Pérez Villazón, con carné de identidad 89031525389, soy el autor principal del trabajo final de maestría “Solución tecnológica para el monitoreo del sistema de recuperación de información de la Plataforma C.U.B.A.”, desarrollada como parte de la Maestría en Informática Avanzada y que autorizo a la Universidad de las Ciencias Informáticas a hacer uso de la misma en su beneficio, así como los derechos patrimoniales con carácter exclusivo.

Y para que así conste, firmo la presente declaración jurada de autoría en La Habana a los ____ días del mes de _____ del año _____.

Yadiel Pérez Villazón

Síntesis

En la presente investigación se define una solución para monitorear el sistema de recuperación de información de la Plataforma C.U.B.A. Se analizaron diferentes fuentes bibliográficas relacionadas con los sistemas de recuperación de información y sus tecnologías, las estrategias, soluciones y tecnologías para el monitoreo, identificando una serie de métricas para evaluar los recursos y los servicios implicados en la recuperación de información de C.U.B.A. La solución de monitoreo está compuesta por 5 componentes: recopilación de mediciones de las métricas, almacenamiento, visualización, evaluación y alertas de las mediciones. También propone una solución tecnológica basada en series temporales donde están implicadas aplicaciones como Telegraf, InfluxDB, Grafana y Kapacitor. La solución se valida a través de la aplicación del criterio de expertos en su variante Delphi, lo que permitió evaluar de muy adecuado el nivel de influencia y la aplicabilidad de la solución. Finalmente se compara el seguimiento realizado al sistema de indexación aplicando la solución y sin aplicarla y se corrobora que la implantación de la solución de monitoreo posibilita la mejora de la pertinencia en el monitoreo del sistema de recuperación de información de la Plataforma C.U.B.A.

Palabras clave: métricas, monitoreo, recuperación de información, seguimiento.

Índice de figuras

Figura 1: Entrada a un sistema una necesidad de información. Fuente: Elaboración propia.....	10
Figura 2: Proceso de recuperación de información. Fuente: Elaboración propia.....	11
Figura 3: Proceso de adquisición – indexación. Elaboración propia.....	13
Figura 4: Componentes de la Plataforma C.U.B.A. Fuente: Elaboración propia.....	14
Figura 5: Despliegue de la Plataforma C.U.B.A. Fuente: Elaboración propia.....	15
Figura 6: Proceso de rastreo de Apache Nutch. Fuente: http://nutch.apache.org	18
Figura 7: Procesos que intervienen en el monitoreo. Fuente: Elaboración propia.....	24
Figura 8: Comunicación entre los elementos que intervienen en una solución de monitoreo. Fuente: Elaboración propia.....	43
Figura 9: Infraestructura tecnológica para el monitoreo. Fuente: Elaboración propia.....	49

Sumario

1 - Introducción.....	1
Capítulo 1: Marco teórico de la investigación.....	10
1.1 - Los motores de búsqueda.....	12
1.1.1 - Arquitectura de un motor de búsqueda.....	12
1.1 - Motor de Búsqueda de la Plataforma C.U.B.A.....	13
1.1.1 - Despliegue de la Plataforma C.U.B.A.....	15
1.1.2 - Principales tecnologías.....	16
1.1.2.1 - Apache Solr.....	16
1.1.2.2 - Apache Nutch.....	17
1.2 - Monitoreo y seguimiento.....	19
1.2.1 - Conceptos asociados al monitoreo.....	19
1.2.2 - Propósito del monitoreo.....	20
1.2.2.1 - Detección temprana de problemas.....	20
1.2.2.2 - Disponibilidad.....	21
1.2.2.3 - Desempeño.....	21
1.2.2.4 - Toma de decisiones.....	21
1.2.2.5 - Predicciones.....	21
1.2.3 - Tipos de monitoreo.....	22
1.2.4 - Procesos que intervienen en el monitoreo.....	23
1.3 - Métricas.....	24
1.3.1 - Recolección estadística.....	24
1.3.2 - Frecuencia de medición.....	25
1.3.3 - Interpretación de las métricas.....	25
1.3.4 - Unidades de las métricas.....	26
1.3.4.1 - Cantidad.....	26
1.3.4.2 - Tiempo de retardo.....	26
1.3.4.3 - Cantidad por tiempo.....	26
1.3.4.4 - Cantidad de entradas por punto de datos.....	27
1.4 - Soluciones tecnológicas para el monitoreo.....	27
1.4.1 - Nagios.....	27

1.4.2 - Zabbix.....	29
1.4.3 - Icinga.....	30
1.4.4 - Valoración de las tecnologías de monitoreo.....	32
1.5 - Monitoreo basado en series temporales.....	32
1.5.1 - Base de datos de series de tiempo.....	33
1.5.1 - Plataformas de series temporales.....	33
1.5.1.1 - TICK.....	33
1.5.1.2 - Graphite.....	34
1.5.1.3 - Prometheus.....	35
1.5.2 - Valoración de las plataformas de series temporales.....	36
1.6 - Conclusiones parciales.....	37
2 - Capítulo 2: Implementación de la solución de monitoreo.....	38
2.1 - Solución de monitoreo de la Plataforma C.U.B.A.....	38
2.1.1 - Procesos que intervienen en el monitoreo.....	38
2.1.1.1 - Recolección de datos.....	39
2.1.1.2 - Almacenamiento de datos.....	39
2.1.1.3 - Presentación de los datos.....	39
2.1.1.4 - Alertas ante incidencias.....	40
2.1.2 - Recolección de métricas.....	40
2.1.2.1 - Sistema de indexación.....	40
2.1.2.2 - Sistema de rastreo.....	42
2.1.2.3 - Experiencia de usuario.....	42
2.1.2.4 - Recursos.....	43
2.1.2.4.1 - Memoria RAM.....	44
2.1.2.4.2 - CPU.....	45
2.1.2.4.3 - Red.....	45
2.1.2.4.4 - Sistema operativo.....	46
2.1.2.4.5 - Almacenamiento.....	47
2.1.2.4.6 - Procesos.....	48
2.1.3 - Infraestructura tecnológica.....	48
2.2 - Conclusiones parciales.....	50
3 - Capítulo 3: Validación de la propuesta de solución.....	51

3.1 - Validación de la solución de monitoreo a través del criterio de expertos.....	51
3.2 - Validación de la solución de monitoreo por los expertos seleccionados.....	53
3.3 - Validación de la pertinencia a través de un análisis comparativo.....	57
3.3.1 - Objetivo de la prueba.....	57
3.3.2 - Escenario de la prueba.....	57
3.3.3 - Objetivos a cumplir.....	58
3.3.4 - Análisis de la pertinencia.....	58
3.4 - Satisfacción de los especialistas con respecto a la solución de monitoreo.....	61
3.4.1 - Conclusiones del capítulo.....	63
Conclusiones.....	64
4 - Recomendaciones.....	65
5 - Referencias bibliográficas.....	66
6 - Anexos.....	71
6.1 - Anexo 1: Operacionalización de la variable dependiente.....	71
6.2 - Anexo 2: Encuesta para determinar nivel de competencia de los expertos.....	71
6.3 - Anexo 3: Encuesta para medir el criterio de los expertos.....	72
6.4 - Anexo 4: Tablas del método Delphi.....	74

Introducción

El crecimiento de nombres de dominios a nivel mundial fue de aproximadamente 1.3 millones en el primer trimestre de 2017 y cerró con 330.6 millones en todos los dominios de nivel superior (TLDs, *por sus siglas en inglés*) (Domain Name Stat, n.d.). Según un informe sobre el panorama global digital en 2017, se revela que más de la mitad de la población mundial usa Internet, estos datos tienen implicaciones de interés para las empresas, los gobiernos y la sociedad en general ya que ofrece ventajas comunicacionales a muy bajos costos y con un mayor alcance, elemento que hace que esté presente en cualquier estrategia de mercado. Por la variedad de información que se puede encontrar es un pilar que las empresas la inserten como parte de la gestión del conocimiento (Statista, 2017).

En el año 1989 Tim Berners-Lee puso a disposición del auditorio la 'World Wide Web' , y en tan poco tiempo, Internet ya se ha convertido en una parte integral de la vida cotidiana de la mayoría de la población mundial. No es solo Internet lo que está creciendo rápidamente, de igual forma se han identificado un número considerable de hitos que se suman al crecimiento y tráfico de datos en la red de redes (Aghaei, Nematbakhsh, & Farsani, 2012; Berners-Lee & Cailliau, 1992; Statista, 2017):

- Más de la mitad de la población mundial posee al menos un teléfono inteligente.
- Casi dos tercios de la población mundial posee un teléfono móvil.
- Más de la mitad del tráfico web mundial proviene de teléfonos móviles.
- Más de la mitad de las conexiones móviles en todo el mundo ahora son de "banda ancha".
- Una de cada cinco personas de la población mundial realiza transacciones en línea.

Los datos anteriores reflejan la expansión y consolidación de Internet, como medio principal de comunicación electrónica de datos. Estos elementos además evidencian la cantidad de información que a cada instante se genera y que es indispensable procesar y colocar de manera oportuna para el consumo de aquellos que deseen consultarla.

Los motores de búsqueda son el principal medio para recuperar todo el contenido web que se encuentra físicamente distribuido y que por su alto índice de crecimiento es un desafío. El funcionamiento general de un motor de búsqueda se estudia desde dos perspectivas complementarias: la recopilación y la recuperación de información. Un motor compila

automáticamente las direcciones de las páginas que formarán parte de su índice tras indexarlas. Una vez estén estos registros depositados en la base de datos del motor, los usuarios buscarán en su índice por medio de una interface de consulta (más o menos avanzada en función del grado de desarrollo del motor) (Ryan, Ryan, Ryan, Munro, & Robinson, 2013; Social, 2018).

La mayoría de los motores de búsqueda basan su funcionamiento en técnicas de recuperación de información (RI) que es el proceso de recuperación que se lleva a cabo mediante consultas a la base de datos donde se almacena la información estructurada, mediante un lenguaje de preguntas. Es necesario tener en cuenta los elementos claves que permiten hacer las búsquedas, determinando un mayor grado de pertinencia y precisión, como son: los índices, palabras clave, tesauros¹ y los fenómenos que se pueden dar en el proceso como son el ruido y silencio documental (Langville & Meyer, 2011; Martínez Méndez, 2004).

La RI está compuesta por una serie de componentes que propician a través de su trabajo en conjunto, la disponibilidad oportuna de la información que se desea consultar. Entre los componentes principales se encuentran los rastreadores webs y los sistemas de indexación.

Un rastreador web es un robot de Internet que ayuda en la indexación web. Los rastreadores web ayudan a recopilar información sobre un sitio web y los enlaces relacionados con ellos, y también ayudan a validar el código HTML² y los hipervínculos. Un rastreador web también se conoce como araña web, indexador automático o simplemente rastreador (Christopher, Prabhakar, & Hinrich, 2008). La indexación se refiere de manera general a la organización de datos según un esquema o plan específico. En TI³, el término tiene varios usos similares, que incluye, hacer que la información sea más presentable y accesible (Maron, 1977; techopedia, 2018).

Cuba no se encuentra ajena del acceso a los medios digitales y cada día más cubanos se acercan a esta nueva forma del conocimiento. La mayoría de las instituciones de los diferentes sectores poseen conexión al menos a la intranet cubana y muchos de ellos acceso a red de redes. Otro dato importante es que a partir de 2014 ETECSA⁴, comenzó a habilitar zonas Wi-fi⁵ públicas y a mediados

¹ Es la lista de palabras o términos controlados empleados para representar conceptos.

² HTML: HyperText Markup Language (lenguaje de marcas de hipertexto)

³ La tecnología de la información (*TI*, o más conocida como *IT* por su significado en inglés: *information technology*) es la aplicación de ordenadores y equipos de telecomunicación para almacenar, recuperar, transmitir y manipular datos.

⁴ La Empresa de Telecomunicaciones de Cuba S.A., tiene la misión de "Brindar servicios de telecomunicaciones que satisfagan las necesidades de los clientes y la población, así como respaldar los requerimientos del desarrollo socio-económico del país".

⁵ Es una tecnología que permite la interconexión inalámbrica de dispositivos electrónicos.

de 2016 habían alcanzado al menos 4 plazas de las 15 principales ciudades cubanas, mientras se prevé que para 2020 el 50 % de los hogares cubanos tenga acceso de banda ancha (Manresa, 2016).

Otro elemento a resaltar es que se ha constatado un crecimiento de los dominios .cu, la Tabla 1 proporcionada por el monitor de dominios cubanos⁶, detalla un aproximado de la cantidad de direcciones web, así como el número de subdominios de los principales dominios de segundo nivel del país (García Torres, 2017):

Tabla 1: Cantidad de sitios por dominios. Tomado de <https://monitorcuba.redcuba.cu>

Direcciones web		2683			
com.cu	35	edu.cu	199	sld.cu	340
cult.cu	244	cuba.cu	693	bloguea.cu	39
uci.cu	27	gob.cu	36	rimed.cu	54

La mayoría de los sitios cubanos no están accesibles desde Internet por lo que buscadores como Google⁷ o Yahoo⁸ no pueden indexar las URLs⁹ generadas por estas aplicaciones, además, muchos usuarios que utilizan los medios digitales tienen limitada la navegación solamente a la intranet cubana. Por la necesidad de contar con un sistema que permita centralizar toda la información que se genera a cada instante por los sitios web cubanos y proveer una solución que brinde funcionalidades para la búsqueda en la web cubana, el Centro de Ideoinformática (CIDI) en el año 2015 despliega la Plataforma C.U.B.A¹⁰, una solución informática que permite la recuperación de información de todos los dominios que se encuentran subordinadas al TLD de primer nivel .cu. Por la complejidad que requiere el despliegue de las tecnologías en la que está soportada esta solución, fue desplegada en una infraestructura de 19 servidores. Entre los principales componentes que presenta la Plataforma C.U.B.A están: balanceadores web, servidores dedicados a la interfaz que permite realizar las búsquedas, servidores de indexación y rastreadores.

De manera general C.U.B.A cuenta con una aplicación web (www.redcuba.cu) que brinda las

⁶ Sistema para la evaluación de los sitios web cubanos que están prestando servicios.

⁷ Es una compañía estadounidense fundada en septiembre de 1998 cuyo producto principal es un motor de búsqueda.

⁸ Es una compañía norteamericana focalizada en brindar prestaciones relacionadas con Internet.

⁹ Uniform Resource Locator (Localizador Uniforme de Recursos).

¹⁰ Contenidos Unificados de Búsqueda Avanzada.

opciones básicas para la recuperación de información, entre las funcionalidades principales se encuentran:

- Búsqueda básica y avanzada en la web.
- Búsqueda de imágenes.
- Búsqueda de documentos de diferentes formatos.
- Directorio (principales sitios por ministerios e instituciones)

A pesar de que la plataforma ha sido estable durante los últimos dos años de explotación, la misma presenta algunas dificultades sobre el correcto funcionamiento de sus componentes debido a que el monitoreo sobre los recursos y servicios se torna complejo, pues, no existe un mecanismo centralizado para esta tarea y obliga a los operarios ejecutar órdenes en una línea de comandos para visualizar el estado de estos elementos. Entre los principales problemas están:

1. Funcionamiento inestable de los servicios implicados en la plataforma.
2. La no existencia de un seguimiento en tiempo real a los recursos de los servidores (RAM¹¹, CPU¹², Almacenamiento).
3. La no existencia de un mecanismo que permita alertar al instante ante la ocurrencia de una incidencia que afecte la disponibilidad de la plataforma.
4. No es posible conocer cuántos documentos han sido indexados en una determinada o rango de fechas, ni la calidad de la información que se almacena en el sistema de indexación.
5. No es posible conocer el funcionamiento de cada una de las etapas del sistema de rastreo ni cuántas direcciones visita en una determinada o rango de fechas.
6. No es posible conocer la utilización real de la plataforma por parte de los usuarios como por ejemplo:
 - Cantidad de visitas (Nacionales y extranjeras).
 - Cantidad de búsquedas en los distintos módulos (web, imágenes, documentos).

¹¹ *Random Access Memory* (Memoria de Acceso Aleatorio)

¹² *Central Processing Unit* (Unidad Central de Procesamiento)

- Cantidad de búsquedas provenientes de usuarios nacionales o extranjeros.
- Consultas en tiempo real; términos más buscados, tendencias de búsquedas.
- Tiempos de respuesta de las consultas y cantidad de resultados de los mismos.
- Cantidad de clic propiciados por el usuario o simplemente la página a donde el usuario llega para poder encontrar su búsqueda.

Las dificultades antes mencionadas tienen implicaciones en la disponibilidad y calidad del servicio que presta la Plataforma C.U.B.A y principalmente en la satisfacción de los usuarios que utilizan este recurso para la búsqueda de información en la red cubana, entre las situaciones que se pueden generar:

1. Dejar a los clientes por prolongadas horas sin servicio, pues al no detectarse a tiempo la deficiencia o el motivo de la situación, el problema es solucionado un tiempo después.
2. No es posible tomar decisión con respecto a la utilización de los recursos de hardware ya que no se puede evaluar en el tiempo si se están aprovechando de manera adecuada o si existen servidores que necesitan más recursos de los que presentan, además, en ocasiones se satura el almacenamiento y no es detectado a tiempo, esto tiene implicaciones en la disponibilidad de los servicios.
3. El equipo de desarrollo no puede mejorar los resultados de búsqueda si no tienen información relacionada con la satisfacción de los usuarios durante la búsqueda. Además es indispensable conocer el uso real de la plataforma para poder planificar una expansión de los recursos y servicios.
4. El rastreo de URL y la cantidad de documentos que se indexan a cada instante así como la calidad de indexación son puntos críticos de un sistema de recuperación de información y son elementos a los cuales es necesario darle un seguimiento riguroso. Actualmente no es posible conocer cuántas URLs son rastreadas en un tiempo determinado así como cuántos documentos son indexados ni la información que se está almacenando. Este es un elemento que se debe conocer en todo momento, pues del correcto funcionamiento de estos dos servicios depende en gran medida de la calidad de los resultados del sistema de recuperación de información.

Una vez planteado lo anterior, se identifica el siguiente **problema científico**:

¿Cómo mejorar la pertinencia en el monitoreo del sistema de recuperación de información de la Plataforma C.U.B.A?

El **objeto de estudio** del presente trabajo se centra en las tecnologías de monitoreo para sistemas de recuperación de información; enmarcándose el **campo de acción** en las tecnologías de monitoreo para el sistema de recuperación de la Plataforma C.U.B.A.

Como **objetivo general** del trabajo se plantea:

Desarrollar una solución tecnológica basada en la recopilación de métricas para mejorar la pertinencia del proceso de monitoreo en la Plataforma C.U.B.A.

Como **objetivos específicos** se establecieron los siguientes:

- Construir el marco teórico de la investigación relacionado con la infraestructura tecnológica de la Plataforma C.U.B.A y las soluciones para el monitoreo.
- Describir las métricas para el monitoreo de la Plataforma C.U.B.A.
- Describir la solución tecnológica para el monitoreo de la Plataforma C.U.B.A.
- Implementar la solución tecnológica para el monitoreo de la Plataforma C.U.B.A.
- Validar la solución tecnológica para el monitoreo de la Plataforma C.U.B.A.

Para dar cumplimiento a los objetivos específicos se formularon las siguientes **tareas de investigación**:

- Revisión y análisis de la infraestructura tecnológica que presenta la Plataforma C.U.B.A.
- Elaboración de reporte técnico sobre el estado actual de las tecnologías asociadas al monitoreo.
- Descripción de los elementos a tener en cuenta para el monitoreo de la Plataforma C.U.B.A.
- Descripción de las métricas a tener en cuenta para el monitoreo de la Plataforma C.U.B.A.
- Diseño de la solución tecnológica para el monitoreo de la Plataforma C.U.B.A.

- Implementación de la solución tecnológica para el monitoreo de la Plataforma C.U.B.A.
- Realización de pruebas a la solución tecnológica para el monitoreo de la Plataforma C.U.B.A.

Se plantea como **hipótesis**:

El desarrollo de una solución tecnológica basada en la recopilación de métricas permite mejorar la pertinencia del proceso de monitoreo en la Plataforma C.U.B.A.

Definiendo como pertinencia:

El grado en que las funciones facilitan el cumplimiento de tareas y objetivos específicos. Como variable independiente se tiene “solución tecnológica para el monitoreo en la Plataforma C.U.B.A.” y como variable dependiente “pertinencia del proceso de monitoreo en la Plataforma C.U.B.A.”. Los indicadores para la medición de la variable independiente fueron obtenidos a partir de la Norma Cubana de Requisitos de Calidad y Evaluación de Sistemas (25023:2017, 2016) y están explicados en la Tabla 23.

La investigación realizada como parte del trabajo se sustenta en el empleo de los **métodos científicos** siguientes:

Análisis Sintético: La aplicación de este método permitió la extracción de los elementos más importantes relacionados con las tecnologías de monitoreo y el funcionamiento de la infraestructura de la Plataforma C.U.B.A. Se analizaron teorías, documentos, bibliografías relacionadas con el monitoreo en el proceso de recuperación de información y a partir de esto se concretaron resultados que contribuyen a un mejor seguimiento de los servidores y servicios implicados en la Plataforma C.U.B.A.

Análisis Histórico-Lógico: Permite constatar teóricamente cómo ha evolucionado la infraestructura de la Plataforma C.U.B.A desde su creación y las tecnologías asociadas al monitoreo, lo que permitió inferir elementos para proponer una mejor solución.

Hipotético-Deductivo: Permite verificar la hipótesis planteada y establecer nuevas predicciones a partir del sistema de conocimientos que se tiene.

Observación: Desde un punto de vista contemplativo, la aplicación de esta técnica permitió observar el funcionamiento de todos los servidores y servicios de la Plataforma C.U.B.A.

Entrevista: Aplicada a los administradores de la Plataforma C.U.B.A y a usuarios que utilizan esta solución para evaluar y constatar cómo se realizaba el proceso de seguimiento de los elementos implicados en la Plataforma C.U.B.A y el tiempo promedio de resolución de problemas antes una incidencia.

Encuesta: Utilizada en la validación de la solución tecnológica para el monitoreo durante la aplicación del método Delphi, aplicada en la selección de los expertos y recopilación de sus criterios; y en la técnica de ladov, aplicada a especialistas de recuperación que utilizan la Plataforma C.U.B.A y sus administradores para determinar su grado de satisfacción con el mecanismo de seguimiento.

Como **aporte teórico** se plantean un grupo de métricas a utilizar en el proceso de monitoreo en el sistema de recuperación de información de la Plataforma C.U.B.A.

El **aporte práctico** estriba en la solución informática que automatiza todo el proceso de seguimiento y monitoreo de la Plataforma C.U.B.A.

Para mostrar el desarrollo de la investigación y los resultados, el trabajo se ha estructurado en tres capítulos, además de las **Conclusiones, Recomendaciones, Referencias Bibliográficas y Anexos.**

En el primer capítulo se realiza un estudio del estado del arte referente a las tecnologías de monitoreo que permitan principalmente recopilar, analizar y visualizar en tiempo real métricas asociadas al seguimiento de un sistema de recuperación de información. Se realiza una descripción de los principales elementos relacionados con la infraestructura de la Plataforma C.U.B.A.

En el segundo capítulo se fundamentan las tecnologías y métricas a utilizar para guiar el seguimiento en tiempo real del sistema de recuperación de información de la Plataforma C.U.B.A, así como la implementación de la solución tecnológica para el monitoreo.

En el tercer capítulo se realiza la validación de la solución tecnológica para el monitoreo del sistema de recuperación de información de la Plataforma C.U.B.A, a través de la aplicación del criterio de expertos en su variante Delphi para determinar la aplicabilidad y nivel de influencia de la estrategia. Además contempla el análisis comparativo, con relación al tiempo de atención a las incidencias y a la satisfacción de los usuarios con el servicio prestado.

Capítulo 1: Marco teórico de la investigación

En el presente capítulo se realiza una descripción del estado actual que presenta el sistema de recuperación de información de la Plataforma C.U.B.A, así como una descripción de la infraestructura de servidores donde se encuentra desplegada esa solución y las diferentes tecnologías que intervienen en el sistema de recuperación de información. Por otra parte se hace una valoración de los diferentes conceptos asociados al monitoreo y el seguimiento, componentes que intervienen en este proceso así como las principales soluciones existentes en la actualidad.

Motor de búsqueda: una aplicación de recuperación de información

En la actualidad la búsqueda constante de información en la web es una actividad casi diaria para la mayoría de las personas ya sean informáticos, médicos, maestros o cualquier otra profesión. La búsqueda y la comunicación en Internet, en conjunto con el entretenimiento, son los usos más populares que se les da a dispositivos como teléfonos inteligentes, tabletas, computadoras entre otros. El escenario de búsqueda más común implica utilizar un motor de búsqueda, donde se escribe una consulta y se reciben respuestas en forma de una lista de documentos en orden jerárquico. Los motores de búsqueda son las aplicaciones más comunes que involucran la recuperación de información (Baboescu, Tullsen, Rosu, & Singh, 2005; Muller, Hendel, & Yeung, 1999).

Gerard Salton, un pionero en el campo de la recuperación de información, planteó que la recuperación de información (*IR, por sus siglas en inglés*) es un campo relacionado con la estructura, el análisis, la organización, el almacenamiento, la búsqueda y la recuperación de información (Croft, Metzler, & Strohman, 2015).

Otros autores plantean que es la localización y presentación a un usuario de resultados relevantes a una necesidad de información (ver Figura 1) expresada como una pregunta, normalmente, en lenguaje natural, esta necesidad de información, se convierte en una consulta, que a su vez, se introduce en el sistema mediante una fórmula o ecuación de búsqueda (Salton, McGill, & 1983, 1983).

Figura 1: Entrada a un sistema una necesidad de información. Fuente: Elaboración propia.

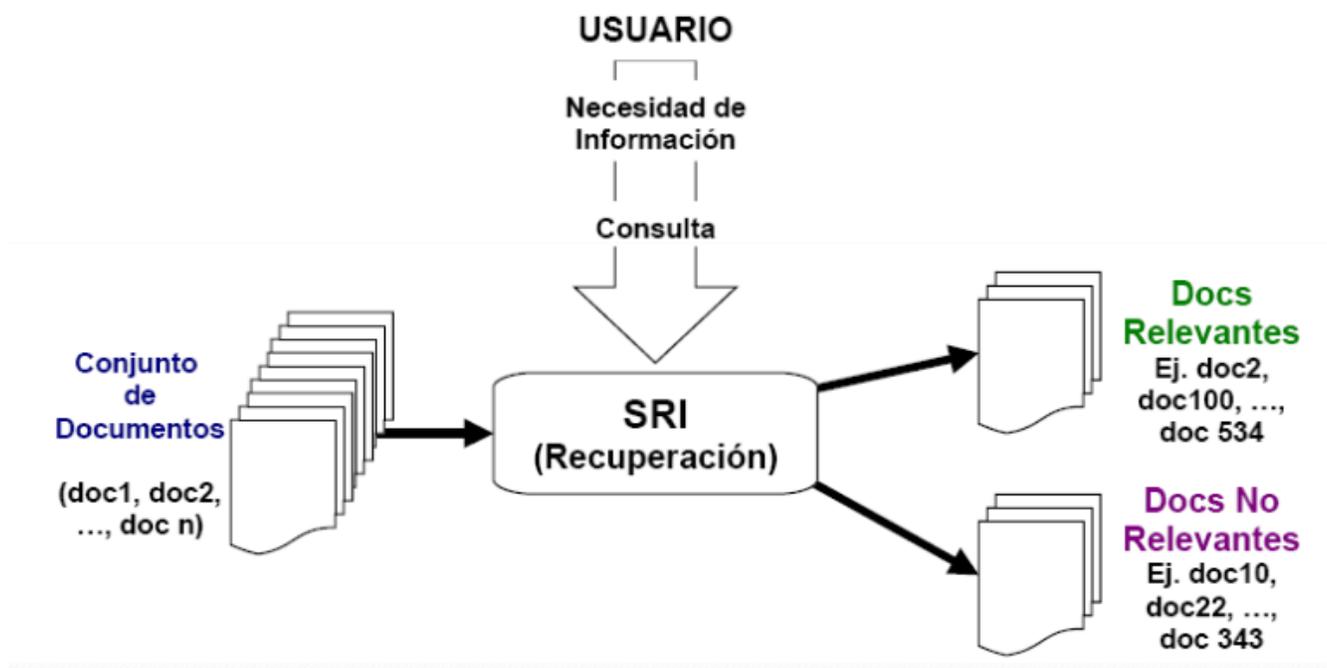


Proceso de recuperación de información

El proceso de recuperación de información inicia cuando existe una necesidad de información (ver Figura 2) por parte de un usuario o sistema, esta necesidad de información se convierte en una consulta introducida en el *IR*. En la recuperación de información una consulta no identifica únicamente a un objeto dentro de la colección de documentos, sino que múltiples objetos pueden ser respuesta a una consulta con diferentes grados de relevancia. Un objeto es una identidad que está representada por información en una base de datos. En dependencia de la aplicación estos objetos pueden ser archivos de texto, imágenes, audio, mapas, videos, entre otros. Los objetos están representados por documentos que muchas veces no están almacenados en el sistema de recuperación, sino que existe una representación de los mismos de manera lógica (Korfhage, 1997; Q. Wu, Burges, Svore, & Gao, 2010).

La mayoría de los sistemas de recuperación de información establecen una clasificación cuando los objetos son respondidos a las consultas introducidas, esto es una medida para identificar cuán relevante es el objeto para la consulta. Los objetos con mayor clasificación son mostrados y el proceso puede tener otras iteraciones si es reordenada o acotada la consulta (Baeza-Yates, Ribeiro, & others, 2011; Worth, 2010).

Figura 2: Proceso de recuperación de información. Fuente: Elaboración propia.



1.1 - Los motores de búsqueda

Un motor de búsqueda es la aplicación práctica de técnicas de recuperación de información para colecciones de texto a gran escala. El motor de búsqueda web es el ejemplo más común, pero también existen otras aplicaciones como búsqueda en el escritorio o la búsqueda empresarial. El término "motor de búsqueda" se usó originalmente para referirse al hardware especializado en la búsqueda de texto, sin embargo, a mediados de la década de 1980, se comenzó a utilizar el nombre de "sistema de recuperación de información" para referirse al software que compara las consultas con documentos y produce listas de resultados clasificadas de documentos (Ryan et al., 2013; Worth, 2010).

Los motores de búsqueda deben ser capaces de capturar o rastrear grandes volúmenes de datos y hacerlo con rapidez, pues el nivel de actualización y crecimiento de la información es una característica de la web en la actualidad. Además, deben proporcionar respuestas a decenas de consultas en instantes de tiempo. Cuando se implementan en entornos operacionales a gran escala se tiene que dar seguimiento a las principales características de rendimiento del motor de búsqueda, tales como, el tiempo de respuesta de las consultas y la velocidad de indexación. El tiempo de respuesta es el retraso entre enviar una consulta y recibir la lista de resultados, el rendimiento mide el número de consultas que se pueden procesar en un momento dado y la velocidad de indexación es la velocidad a la que los documentos de texto se pueden transformar en índices¹³ para buscar (Croft et al., 2015).

1.1.1 - Arquitectura de un motor de búsqueda

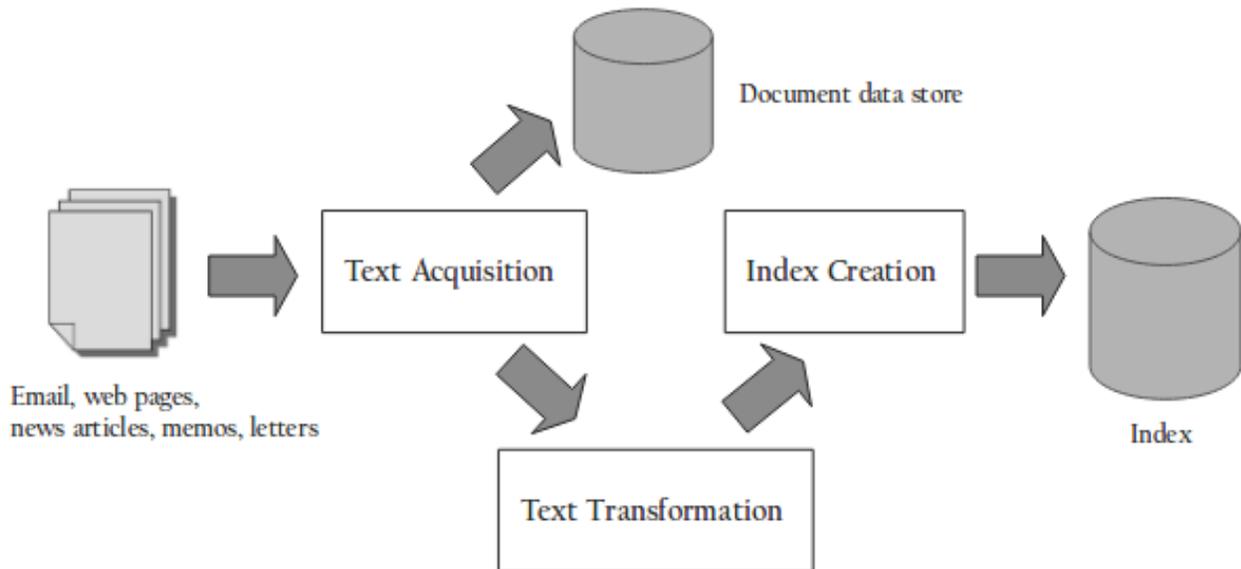
Los motores de búsqueda cuentan con tres componentes principales: *proceso de indexación*, *proceso de consulta* y *proceso adquisición de documentos*. El proceso de indexación construye las estructuras que permiten la búsqueda, y el proceso de consulta utiliza esas estructuras y una consulta previamente introducida para producir una lista clasificada de documentos. El proceso adquisición de documentos es el encargado de identificar y poner a disposición los documentos que se buscarán. Aunque en algunos casos esto implicará simplemente el uso de una colección existente, la adquisición de texto requerirá más a menudo crear una colección a partir del rastreo o escaneo de la web.

El proceso de adquisición de documentos (ver Figura 3), además de enviar los documentos

¹³Un índice es una estructura de datos que mejora la velocidad de búsqueda.

encontrados al proceso de indexación, crea un almacén de datos del documento, que contiene el texto y los metadatos de cada uno de los documentos. Los metadatos son información sobre un documento que no forma parte del contenido de texto, como el tipo de documento, la estructura y tamaño del documento entre otras características (Croft et al., 2015).

Figura 3: Proceso de adquisición – indexación. Elaboración propia.



1.1 - Motor de Búsqueda de la Plataforma C.U.B.A

La Plataforma C.U.B.A es una solución informática desarrollada en el Centro de Ideoinformática (CIDI) de la Universidad de las Ciencias Informáticas. Es un software perteneciente a la familia de los motores de búsqueda y el sistema de recuperación de información que presenta tiene como base el motor de búsqueda Orion¹⁴, producto de igual manera desarrollado por este Centro (Croft et al., 2015).

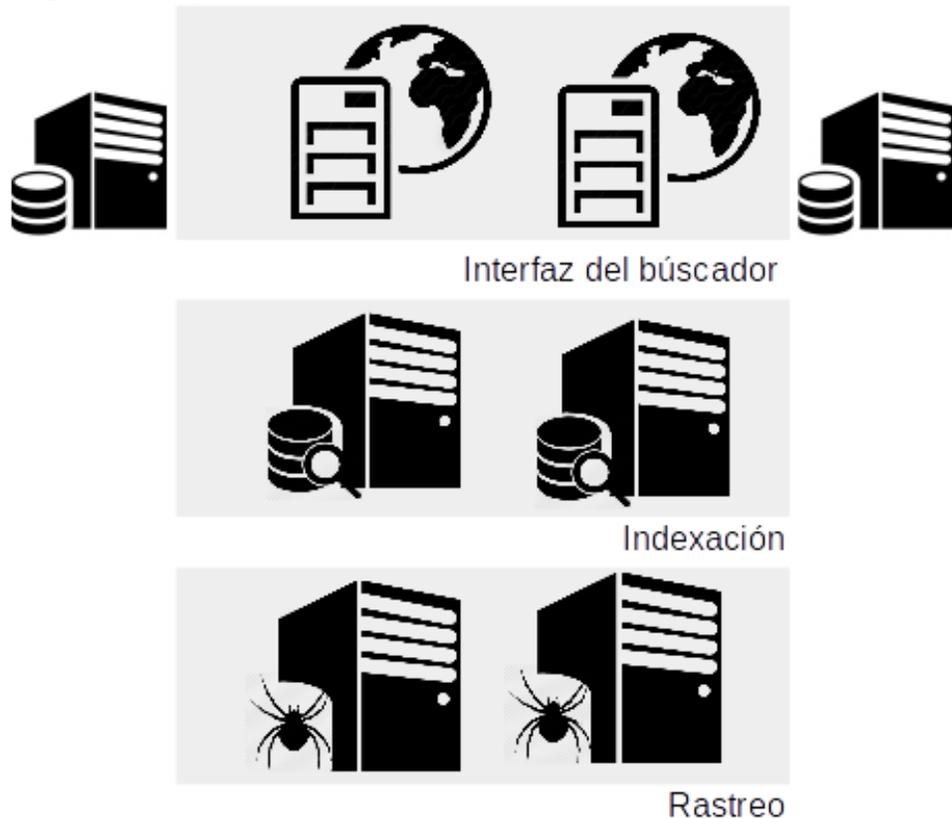
C.U.B.A, con más de 3 años de explotación como solución a la centralización y unificación de los contenidos que se encuentran bajo el dominio de Cuba (.cu), está compuesto por los principales

¹⁴ Sistema de recuperación de información desarrollado por la Universidad de las Ciencias informáticas y que tiene como característica principal su posibilidad de ser personalizada a las necesidades de una institución. Ver más sobre este producto en <http://www.uci.cu/investigacion-y-desarrollo/productos/xilema/orion>.

componentes de un motor de búsqueda (ver Figura 4):

- Servidor de indexación basado en la plataforma de código abierto Apache Solr¹⁵.
- Servicio de rastreo basado en Apache Nutch¹⁶.
- Interfaz que permite a los usuarios realizar búsquedas de contenidos en la web cubana, así como la recuperación de documentos en los distintos formatos (pdf, odt, doc, entre otros) e imágenes. La Interfaz cuenta además con una base de datos para guardar preferencias y datos del sitio.

Figura 4: Componentes de la Plataforma C.U.B.A. Fuente: Elaboración propia.



¹⁵ Es una plataforma de búsquedas basada en Apache Lucene, que funciona como un "servidor de búsquedas".

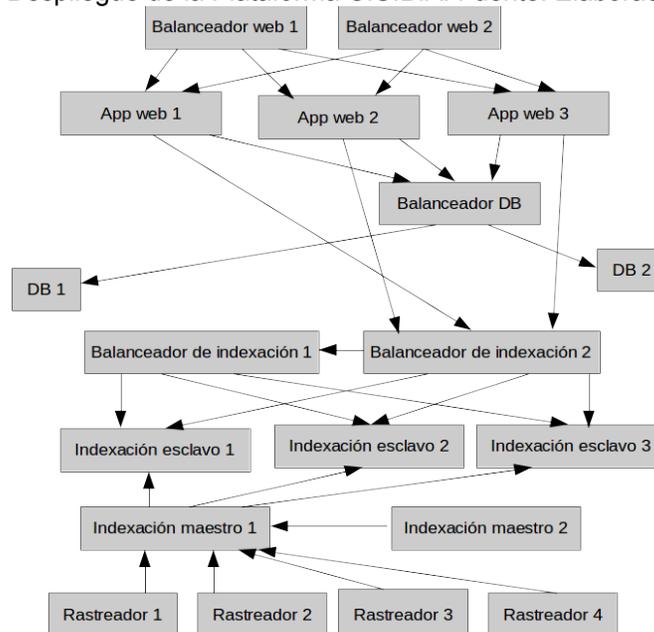
¹⁶ Es un proyecto de software de código abierto altamente extensible y escalable que permite rastrear la web.

1.1.1 - Despliegue de la Plataforma C.U.B.A

La Plataforma C.U.B.A se encuentra desplegada en 19 servidores (ver Figura 5), presentando como elemento significativo la redundancia en cada una de sus capas, entre las principales características que presenta actualmente esta infraestructura se encuentran:

- Balanceo de carga en el acceso de la interfaz del buscador, a la base de datos donde se almacenan las preferencias del sitio y al sistema de indexación, garantizando la alta disponibilidad en este sentido.
- Redundancia en los nodos de la interfaz del buscador, en la base de datos, en el sistema de indexación y en los nodos destinados al rastreo.

Figura 5: Despliegue de la Plataforma C.U.B.A. Fuente: Elaboración propia.



- Utilización de Centos¹⁷ en su versión para servidores como sistema operativo de código abierto en cada uno de sus nodos, así como todas las tecnologías presentadas en esta

¹⁷ Es un sistema operativo de código abierto, basado en la distribución Red Hat Enterprise Linux, operándose de manera similar, y cuyo objetivo es ofrecer al usuario un software de "clase empresarial" gratuito. Se define como robusto, estable y fácil de instalar y utilizar. Desde la versión 5, cada lanzamiento recibe soporte durante diez años, por lo que la actual versión 7 recibirá actualizaciones de seguridad hasta el 30 de junio de 2024.

solución.

1.1.2 - Principales tecnologías

Las tecnologías por las cuales está soportada la Plataforma C.U.B.A tienen la característica de que son de código abierto, entre las que más resaltan están Apache Solr como servidor de indexación, Apache Nutch como sistema para el rastreo web, Nginx¹⁸ como servidor web y que también está implementado como solución de balanceo de carga y Symfony¹⁹ como framework de PHP²⁰. A continuación se describirán con más detalles Apache Solr y Apache Nutch.

1.1.2.1 - Apache Solr

Apache Solr, es una plataforma de búsqueda altamente escalable construida con Apache Lucene²¹. Es un proyecto con una inmensa comunidad a nivel mundial que permite una constante actualización de nuevas características. Puede ser personalizado por parte de los clientes y desarrolladores. Puede ser usado en la nube y es una solución que se caracteriza por su robustez, tolerancia a fallas y confiabilidad (Falcón Márquez, 2015; Paz Saavedra, 2015).

Solr está escrito en Java y se ejecuta como un servidor independiente, está basado en configuraciones lo que garantiza una alta flexibilidad y proporciona una interfaz gráfica de administrador para su supervisión (Estrada Ramos, 2012) .

Entre los elementos significativos de Solr (Shahi, 2015):

Índice invertido: Apache Lucene crea un índice invertido de los documentos que agrega a Solr, y en el momento de la consulta, busca en el índice los documentos coincidentes.

Modelo de espacio vectorial: Utiliza el modelo de espacio vectorial (*VSM, por sus siglas en inglés*) de forma predeterminada junto con el modelo booleano para determinar la relevancia de un documento con respecto a una consulta del usuario.

¹⁸ Es un servidor web/proxy inverso ligero de alto rendimiento y un proxy para protocolos de correo electrónico.

¹⁹ Framework diseñado para optimizar el desarrollo de las aplicaciones web basado en el patrón Modelo Vista Controlador

²⁰ Es un lenguaje de programación de propósito general de código del lado del servidor originalmente diseñado para el desarrollo web de contenido dinámico.

²¹ Es una API de código abierto para recuperación de información, originalmente implementada en Java por Doug Cutting. Está apoyado por el Apache Software Foundation y se distribuye bajo la Apache Software License.

Basado en configuración: Entre sus archivos de configuración resaltan `solrconfig.xml` y `schema.xml`. El archivo `schema.xml` define principalmente los campos del esquema y el comportamiento de esos campos (cómo se combinará el texto al indexar y consultar). Por otra parte el archivo `solrconfig.xml` contiene configuraciones y definiciones específicas del núcleo relacionadas con el manejo de solicitudes, el formato de respuesta, la indexación, la configuración de otros componentes, la administración de memoria y el almacenamiento en caché. Las configuraciones pueden editarse manualmente o modificarse dinámicamente llamando a las respectivas API. A partir de Solr 5.0 se pueden cargar archivos JAR a través de una llamada API.

Cadena de análisis: la consulta de búsqueda de los documentos que se indexan pasan por una cadena de analizadores y tokenizadores²². Los tokens son utilizados para hallar la coincidencia ante una consulta introducida.

1.1.2.2 - Apache Nutch

Apache Nutch es un software de rastreo web de código abierto extensible y escalable. Proporciona instalaciones para análisis, indexación y filtros de puntuación para implementaciones personalizadas. Puede ser ejecutado en un solo ordenador o como en un entorno distribuido como Apache Hadoop²³. Permite la integración con el lenguaje de scripts Python para el rastreo web. Apache Nutch puede integrarse fácilmente con Apache Solr e indexar todas las páginas web rastreadas en Apache Solr (Lopez, Duerr, & Khalsa, 2015) .

Apache Nutch en su funcionamiento completa en cada ciclo una serie de etapas donde intervienen un grupo de componentes (ver Figura 6) (Laliwala, 2013, pp. 4–11) :

1. La primera etapa es el *Inyector*, todo comienza cuando se toman las URLs de un archivo denominado semilla que es donde se escriben las primeras URLs y las agrega a *Crawldb*. Como una parte central de Nutch, el *Crawldb* mantiene información sobre todas las URLs conocidas (calendario de búsqueda, estado de búsqueda, metadatos, entre otras características).
2. A partir de las URL inyectadas en el *Crawldb*, el generador (*GeneratorJob*) crea una lista de

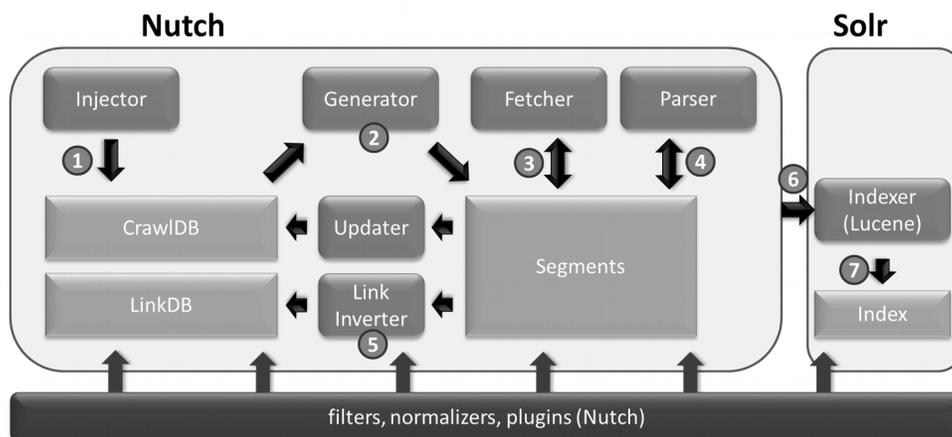
²² Es el proceso de partir un texto en elementos denominados tokens. Llamamos tokens a unidades indivisibles. Un tokenizador o segmentador es el primero de los componentes que se utiliza en el procesamiento de texto.

²³ Es un framework de software que soporta aplicaciones distribuidas bajo una licencia libre. Permite a las aplicaciones trabajar con miles de nodos y petabytes de datos.

búsqueda y la coloca en un directorio de segmento recién creado.

3. A continuación, se obtiene el contenido de las URLs que están en la lista (*Fetcher*) y lo vuelve a escribir en el directorio del segmento. Este paso por lo general es el más lento.
4. Luego le corresponde al analizador (*Parser*) procesar el contenido de cada página web (omitiendo las etiquetas HTML). Si el rastreo funciona como una actualización o una extensión se agregan las nuevas URL al *CrawlDB* para ser procesada en el siguiente ciclo.
5. Antes de ser indexada la información, todos los enlaces deben invertirse, los enlaces invertidos se guardan en el *Linkdb*.
6. Utilizando todas las fuentes de datos (*CrawlDB*, *Linkdb* y segmentos) el indexador crea un índice y lo guarda dentro del directorio de Solr.

Figura 6: Proceso de rastreo de Apache Nutch. Fuente: <http://nutch.apache.org>



La Figura 6 también refleja cómo Apache Nutch puede ser altamente modular, flexible y personalizable ya que se pueden adicionar filtros, normalizadores y complementos durante todo el proceso.

La evaluación y seguimiento de los sistemas y en especial de los motores de búsqueda es esencial para tener productos con mejor calidad. Mantener la disponibilidad y la prestación constante de los servicios son claves cuando se tiene una solución que es utilizada por miles de usuarios. En el caso del sistema de recuperación de información de la Plataforma C.U.B.A es importante conocer, desde el punto de vista de los recursos, si se están utilizando debidamente o si algún servidor necesita mejorar el hardware y en el caso de las tecnologías del motor de búsqueda hay que conocer si se está

utilizando efectivamente, si las respuestas a consultas de los usuarios responden en el tiempo establecido, si la indexación se está comportando de manera correcta o si cada una de las etapas de rastreo se desenvuelven como lo planificado. Datos como estos y otros no se pueden lograr si no se concibe una adecuada estrategia de monitoreo y seguimiento.

1.2 - Monitoreo y seguimiento

El monitoreo se ha convertido en un término general cuyo significado depende en gran medida del contexto. Más ampliamente, se refiere al proceso de tomar conciencia del estado de un sistema. Esto se hace de dos maneras, proactivo y reactivo. El primero implica observar los indicadores visuales, como las series temporales y los cuadros de mandos. El segundo implica formas automáticas de entregar notificaciones a los operadores con el fin de llamar su atención a un cambio grave en el estado del sistema; esto generalmente se conoce como alerta (Ligus, 2013).

El monitoreo es el proceso de mantener la vigilancia sobre la existencia y la magnitud del cambio de estado y el flujo de datos en un sistema. Además tiene como objetivo identificar las fallas y ayudar a su eliminación posterior. Las técnicas utilizadas en la supervisión de sistemas de información se cruzan con los campos de procesamiento en tiempo real, estadísticas y análisis de datos. Un conjunto de componentes de software utilizados para la recopilación de datos, su procesamiento y presentación se denomina sistema de supervisión.

Otro elemento de gran importancia y que va unido a los sistemas de monitoreo son las alertas. Una alerta es la capacidad de un sistema de monitoreo para detectar y notificar a los operadores sobre eventos significativos que denotan un grave cambio de estado. La notificación se conoce como alerta y es un mensaje simple que puede adoptar múltiples formas: correo electrónico, SMS²⁴, mensajes instantáneos (IM, *por sus siglas en inglés*) o una llamada telefónica. La alerta se transmite al destinatario apropiado, es decir, una parte responsable de tratar el evento.

1.2.1 - Conceptos asociados al monitoreo

Para una mejor comprensión acerca del monitoreo y las tecnologías asociadas al mismo se describen algunos conceptos.

Agente: En el contexto del monitoreo un agente es un software que recolecta continuamente las

²⁴ Mensaje corto de texto que se puede enviar entre teléfonos celulares o móviles.

entradas de datos y las envía a un sistema de monitoreo (Ligus, 2013).

Alarma: Una pieza de configuración que describe una condición indeseable y alertas emitidas en respuesta a ella (Ligus, 2013).

Alerta: Una alerta es un mensaje de notificación que informa sobre un cambio de estado, que generalmente indica un posible problema (Ligus, 2013).

Entrada de datos: Es un valor numérico con un conjunto de propiedades recopiladas en el origen de la medición por un agente de supervisión (Ligus, 2013).

Métrica: Una colección de entradas de datos descrita por un conjunto de propiedades. Las series de tiempo a menudo se conocen como métricas. Las métricas de supervisión tampoco se deben confundir con las métricas de rendimiento, que son un conjunto de indicadores de rendimiento empresarial de alto nivel.

Monitor: Un proceso que evalúa los puntos de datos más recientes en una serie de tiempo para el ajuste del umbral. Esta es una parte integral de una alarma.

Supervisión: El proceso de recopilación y recuperación de datos relevantes que describen un cambio de estado.

Serie de tiempo: Una lista de puntos de datos ordenados en orden temporal natural, más comúnmente presentado en una trama.

1.2.2 - Propósito del monitoreo

La implementación de un sistema de monitoreo está orientado principalmente a disminuir las afectaciones y contar con un seguimiento exhaustivo de un sistema o proceso, algunos de los propósitos son:

1.2.2.1 - Detección temprana de problemas

La detección rápida de problemas amenazantes es el objetivo más importante del monitoreo y es la función de las alertas en una solución de supervisión. La dificultad consiste en perseguir dos objetivos conflictivos: velocidad y precisión. Conocer si algo no está funcionando adecuadamente y además conocerlo lo más rápido posible. La precisión permite identificar la diferencia si una afectación provoca problemas temporales y transitorios de impacto insignificante. Detrás de cada valor de

umbral razonable acecha el riesgo de que problemas potencialmente desastrosos pasen desapercibidos (Ligus, 2013).

1.2.2.2 - Disponibilidad

La pérdida de disponibilidad es el tiempo de inactividad de un sistema. Sucede cuando el sistema está sujeto a la pérdida total de disponibilidad. La pérdida de disponibilidad también puede ser parcial o no estar disponible solo para una parte de los usuarios. La clave es la detección temprana y la prevención en entornos de producción ocupados. El tiempo de inactividad generalmente se traduce directamente en pérdidas. Una configuración de monitoreo completa que permite la identificación oportuna de los problemas resulta indispensable. A través de la supervisión se puede identificar la causa raíz detrás de la falla. La acción correctiva se basa en los resultados del análisis de causa raíz y se lleva a cabo para prevenir futuras apariciones del problema. La solución del problema evita la repetición de las mismas fallas a largo plazo (Ligus, 2013).

1.2.2.3 - Desempeño

Prestar atención al comportamiento anómalo en el sistema ayuda a detectar la saturación de recursos y defectos poco frecuentes. Una serie de fallas se consiguen mediante el Aseguramiento de la Calidad²⁵ (QA, *por sus siglas en inglés*), estas, son difíciles de detectar y es probable que surjan solo después de largas horas de pruebas. Otro grupo de errores emergen cuando el sistema se expone a una carga a gran escala. Aunque es difícil de aislar en entornos de prueba, son consistentemente reproducibles en producción. Una vez que se encuentran a través de una supervisión, son más fáciles de identificar y eliminar (Ligus, 2013).

1.2.2.4 - Toma de decisiones

Los operadores deben tomar decisiones rápidamente, y en situaciones críticas, conocer bien su sistema puede reducir errores y mejorar sus posibilidades de mitigación exitosa. Una adecuada supervisión permite a largo plazo tomar decisiones con respecto a la utilización de los recursos o mejoras en las características del sistema supervisado (Ligus, 2013).

1.2.2.5 - Predicciones

²⁵ Es cualquier proceso sistemático para determinar si un producto o servicio cumple con los requisitos especificados.

En el contexto del monitoreo, una predicción es un pronóstico cuantitativo que contiene un grado de incertidumbre sobre los niveles futuros de recursos o fenómenos que conducen a su utilización.

Supervisar el tráfico y los patrones de uso a lo largo del tiempo sirve como fuente de información para la toma de decisiones.

Ayuda a predecir, por ejemplo, los momentos de mayor tráfico de datos y el instante de mayor utilización de un determinado recurso de hardware. Cuando los patrones de uso se salen de los límites proyectados, es probable que haya una buena razón para ello, incluso si esta razón no depende directamente de la operación del sistema.

Además de predecir la carga de trabajo futura, una interacción estrecha con el monitoreo puede ayudar a predecir las tendencias comerciales. Los clientes pueden tener diferentes necesidades en diferentes momentos del año. La capacidad de predecir demandas y luego hacerlas coincidir en función de la estacionalidad se traduce directamente en ganancias de ingresos (Ligus, 2013).

1.2.3 - Tipos de monitoreo

Existen decenas de métodos y técnicas diferentes para monitorear sistemas distribuidos o procesos, la aplicación de una técnica o un método es distinto en cada contexto en dependencia de lo que se desea controlar y dar seguimiento. No existen claramente tipos de monitoreo, pero de manera general se pueden definir tres categorías básicas para el monitoreo: *monitoreo técnico*, *monitoreo funcional* y *monitoreo de procesos* (Kjkoster, 2014).

Estas tres categorías tienen una jerarquía muy clara y en un escenario donde existe una supervisión bien configurada, una alarma que indique que todo está funcionando bien en el nivel superior indica que todos los niveles inferiores funcionan bien. Sin embargo, si existen alarmas de dificultades en un nivel inferior, esto es una indicación de que hay un problema de bajo impacto.

En sentido inverso, una alarma que indique un correcto funcionamiento de uno de los niveles inferiores no garantiza nada sobre el rendimiento de los niveles superiores de supervisión. Simplemente sirve como guía para indicar que un determinado subsistema tiene un buen rendimiento y es poco probable que contribuya a una dificultad en un nivel superior (Kjkoster, 2014).

El contexto de la investigación se centrará en la aplicación de las tres categorías de monitoreo teniendo en cuenta como proceso principal o de nivel superior el monitoreo del sistema de

recuperación de información de la Plataforma C.U.B.A y los servidores y sistemas desplegados como elementos de niveles inferiores. El seguimiento y control de los niveles más bajo puede contribuir al correcto funcionamiento del sistema de recuperación de información de la Plataforma C.U.B.A.

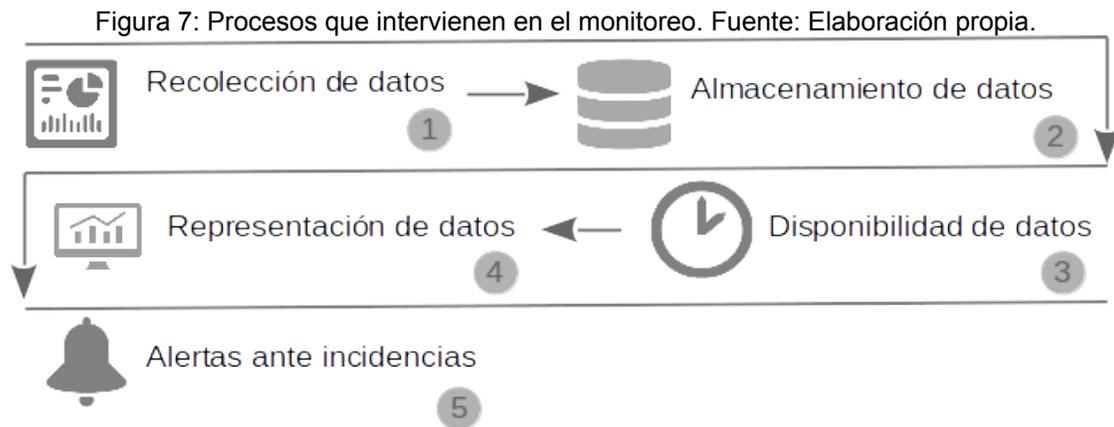
1.2.4 - Procesos que intervienen en el monitoreo

Durante la etapa de implementación de un sistema de monitoreo es necesario definir como primer paso un grupo de variables como parte de la estrategia de seguimiento de los sistemas o procesos, se pueden definir de manera general para todos los elementos o de manera individual, por lo general como los sistemas o procesos poseen características diferentes, pues la definición está sujeta a cada elemento. Entre las variables a definir se encuentran (Worthington, 2016):

- Métricas que se van a tomar.
- Frecuencia de medición.
- Medidas de evaluación de una medición.
- Responsable de analizar la medición.

Una vez definidas las métricas es necesario almacenar los valores recolectados en dependencia de la frecuencia de medición, de esta manera pueden ser consultadas por los responsables de evaluar las mismas. Además del valor, debe ser tomado el instante de tiempo de la medición, esto permite llevar un control y un seguimiento de los valores en una línea de tiempo. La persistencia de los datos garantiza contar con un histórico de esa información que puede ser útil para futuras planeaciones y tomas de decisiones.

No basta con implementar una correcta adquisición y persistencia de los datos, además se debe implementar una correcta forma de visualización de esa información mediante gráficas y tablas que muestren el comportamiento de los sistemas, si es posible, en tiempo real. Otro elemento de importancia es la implementación de mecanismos de alertas que notifiquen cuando ocurra una incidencia en la evaluación de una determinada medición. La Figura 7 muestra de manera gráfica los procesos que intervienen en una solución de monitoreo (Ligus, 2013).



1.3 - Métricas

Las métricas son colecciones de entradas de datos numéricos organizadas en grupos de listas ordenadas cronológicamente. Cada entrada de datos consta de un valor de medición registrado, la marca de tiempo en que se realizó la medición y un conjunto de propiedades que lo describen. Cuando las entradas de datos de una métrica se segmentan en intervalos de tiempo fijos y se resumen mediante una transformación matemática de alguna manera significativa, se pueden presentar como series de tiempo e interpretarse en gráficos bidimensionales (Ligus, 2013).

La longitud de los intervalos de puntos de datos, también conocida como granularidad de tiempo, depende de los tipos de mediciones y del tipo de información que se extraerá. Los intervalos comunes incluyen 1, 5, 15 y 60 minutos, pero también es posible representar intervalos tan pequeños como un segundo y grandes como un día.

La ventaja más importante de utilizar series de tiempo para el monitoreo es su propiedad de ilustrar con precisión el proceso de cambio en un contexto de datos históricos. Son una herramienta indispensable para notar con exactitud el cambio de estado. Las series de tiempo son bidimensionales con datos en el eje *y* y el tiempo en el eje *x*.

De esta forma, trazar datos de múltiples métricas uno contra el otro agrega una sola capa de complejidad a la vez al gráfico. Por esa razón, las series temporales proporcionan un medio eficiente para resaltar las relaciones correlativas entre los datos de muchas fuentes, como las dependencias entre capas en una pila de software (Ligus, 2013).

1.3.1 - Recolección estadística

Por cada métrica almacenada, se guardan además un conjunto de propiedades asociadas a la misma como es el caso del nombre del servidor, grupo a la que pertenece, entre otros. La recolección se realiza en intervalos de tiempo espaciados uniformemente y resumiendo matemáticamente las entradas de datos en cada intervalo, la Tabla 2 refleja los elementos matemáticos que se utilizan comúnmente.

Tabla 2: Resúmenes estadísticos para la recolección de mediciones

Conteo	N	El conteo de valores o entradas durante el intervalo de tiempo.
Suma	$sum(N)$	La suma de los valores o entradas durante el intervalo de tiempo.
Promedio	$avg \rightarrow sum(N) / N$	El valor medio para todas las entradas durante el intervalo de tiempo.
Porcentaje	$Percent (0 - 100)$	Porcentaje de los valores o entradas durante el intervalo de tiempo. Por ejemplo utilización del almacenamiento de un dispositivo con respecto al total presentado.
Desviación	dvt	Desviación estándar del promedio en la distribución de las entradas recolectadas.

1.3.2 - Frecuencia de medición.

La recolección de los datos se realizan con una frecuencia de medición, los intervalos de tiempos de la serie temporal se toman uniformemente. La frecuencia de medición puede ser corta o larga, cuanto más grande sea el intervalo de tiempo, más largo será el período. Las métricas recolectadas en tiempos cortos tienden a revelar la hora exacta de un evento y, por lo tanto, son útiles para encontrar una dirección en las relaciones causales, así como para describir las líneas de tiempo. Sin embargo, podrían ser más costosos en el proceso de almacenamiento. Las métricas recolectadas en tiempos largos, son mucho más adecuadas para ilustrar tendencias (Ligus, 2013). Seleccionar el intervalo de tiempo correcto para tomar mediciones de una métrica es importante para una interpretación precisa de los datos.

1.3.3 - Interpretación de las métricas

Una comprensión sólida de los datos presentados en las métricas es esencial para extraer conclusiones fiables de las tramas de series temporales. Saber a qué categoría pertenece una

métrica y dónde se originan sus datos es útil para darse cuenta del impacto de los cambios de patrones y presentar la información de las maneras más concluyentes. Algunas estadísticas resumen, enfatizan la información clave mejor que otras, y los mismos datos pueden aparecer disímiles cuando se muestran en un intervalo de recolección de tiempo diferente. Darse cuenta de las diferencias de ese tipo ayuda a construir tramas que transmiten claramente el punto (Ligus, 2013).

1.3.4 - Unidades de las métricas

Las métricas se pueden diferenciar en dependencia del tipo de la unidad en el proceso de identificación a la hora de realizar la recolección de sus mediciones, generalmente se evidencian tres tipos de unidades (Nayak, Nanda, Dwarakanath, Babu, & Selvakumar, 2014):

1.3.4.1 - Cantidad

Las cantidades son una colección de valores discretos o continuos que resultan de las entradas; los ejemplos incluyen el número de coincidencias en un resultado de búsqueda, tamaño del paquete, espacio libre en el disco, entre otros, es el grupo de métricas más común. Los datos resultantes almacenados como series temporales ilustran el flujo y los estados operacionales. Las unidades más comunes son: **bytes**, **kilogramos**, **precio**, **un entero (constante)**, **suma de elementos**, entre otros.

1.3.4.2 - Tiempo de retardo

Los tiempos de retardo son cantidades de tiempo requeridas para completar una acción. Los ejemplos incluyen latencia de la red, tiempo de respuesta de solicitud web, tiempo de ida y vuelta de ICMP²⁶, tiempo de un usuario visitando una página web.

Al igual que las cantidades, los retrasos generalmente se registran en todos los niveles y juegan un papel crucial en la supervisión del rendimiento debido al efecto inmediato de los tiempos de respuesta en la experiencia del usuario. Los puntos de datos resultantes son casi siempre una mezcla de múltiples entradas que ocurrieron en un período de tiempo dado. Los valores de porcentos y promedios deben ser observados más de cerca. Las unidades más comunes son: **milisegundos**, **segundos**, **minutos**, **horas**, **días**, **ciclos de CPU**, entre otros.

1.3.4.3 - Cantidad por tiempo

Son cantidades discretas o continuas que fluyen a través del sistema por unidad de tiempo,

²⁶ Internet Control Message Protocol (Protocolo de control de mensajes de Internet)

conocido como rendimiento; ejemplos incluyen transmisión de bits y operaciones de entrada/salida (*IOPS*, por sus siglas en inglés).

Esas métricas son adecuadas para monitorear bits de datos producidos en grandes cantidades con alta variabilidad de valores. Se usan con mayor frecuencia para controlar las métricas de nivel inferior, como las estadísticas de los dispositivos de hardware. Normalmente, el dispositivo de hardware subyacente tiene un mecanismo incorporado para realizar un seguimiento e informar sobre el flujo de rendimiento. En tales casos, la cantidad por métrica de tiempo representa una entrada por punto de datos: es decir, el dispositivo fue consultado por su estado una vez en un período de punto de datos determinado. En otros casos, cuando hay múltiples entradas disponibles por punto de datos, la variabilidad del rendimiento se puede observar a través de la distribución de entrada mediante el uso de porcentos, como en el caso de los dos tipos anteriores. Las unidades más comunes son: ***bits por segundo***, ***IOPS***, ***millas por hora***, entre otros.

1.3.4.4 - Cantidad de entradas por punto de datos

Las mediciones de métricas se pueden dividir en dos categorías en dependencia del número de entradas para construir un punto de datos y la naturaleza de la evaluación, las multiprocesadas y las de procesamiento simple (Ligus, 2013).

Los puntos de datos para métricas multiprocesadas son estadísticas de resumen combinadas de múltiples entradas. Las entradas que se almacenan son puntos de datos resultantes, contienen un conjunto completo de estadísticas resumen útil que describen efectos acumulativos (sumas (sum(N), cantidades (N)), los valores típicos (media, mediana) y de porcentos. Entre los ejemplos se pueden evidenciar: *los bytes transferidos*, *el número de solicitudes HTTP*, *el tiempo promedio de respuesta de un sitio*, entre otros. Los puntos de datos de procesamiento simple requieren una entrada para construir un punto de datos significativos. En la mayoría de los casos, la métrica ilustra un cambio de estado a lo largo del tiempo.

1.4 - Soluciones tecnológicas para el monitoreo

Existe un grupo considerable de soluciones para implementar un sistema de monitoreo y alertas ante incidencias, a partir de intercambios y encuestas a administradores de redes se analizan las principales aplicaciones que realizan esta tarea.

1.4.1 - Nagios

Nagios es una herramienta para el monitoreo de servicios, servidores y dispositivos de red, mantiene un seguimiento de estos elementos verificando en todo momento que estén funcionando como deberían. Nagios además monitorea los procesos e informa sobre su estado.

El monitoreo del sistema en Nagios se divide en dos categorías de objetos: hosts²⁷ y servicios. Los hosts representan un dispositivo físico o virtual en su red (servidores, enrutadores, estaciones de trabajo, impresoras, entre otros). Los servicios son funcionalidades particulares, por ejemplo, un servidor Secure Shell (SSH)²⁸ se puede definir como un servicio para ser monitoreado. Cada servicio está asociado con un host en el que se está ejecutando. Además, las máquinas se pueden agrupar en grupos de host (Kocjan & Beltowski, 2016).

Nagios usa cuatro estados distintos: Ok²⁹, Warning³⁰, Critical³¹ y Unknown³². El enfoque para ofrecer solo tres estados permite a los administradores ignorar los valores de monitoreo y decidir cuáles son los límites críticos o de advertencia. Los administradores del sistema tienden a ignorar la disminución gradual del espacio de almacenamiento. Los operadores ignoran el funcionamiento de un proceso hasta que entra en un estado crítico porque se está quedando sin espacio en disco. Tener un límite estricto para mirar es mucho mejor, porque siempre se detecta un problema, independientemente de si cambia de advertencia a crítica en 15 minutos o en una semana. Esto es exactamente lo que hace Nagios. Cada comprobación realizada por Nagios se basa en valores numéricos (como la cantidad de espacio de disco o uso de CPU) a uno de los tres estados posibles (Kocjan & Beltowski, 2016).

Otro de los beneficios de Nagios son los informes que se visualizan en su consola de presentación de datos donde indica el estado de todos los servicios, por ejemplo, cuáles tienen un funcionamiento normal, cuáles están en un estado de advertencia y los que presentan un punto crítico. También puede ayudar a priorizar lo que se debe tratar primero, y qué problemas se pueden manejar más adelante.

Nagios realiza todas sus comprobaciones utilizando complementos. Estos son componentes

²⁷ Un host o anfitrión es un ordenador que funciona como el punto de inicio y final de las transferencias de datos.

²⁸ *SSH (Secure Shell*, en español: intérprete de órdenes seguro) es el nombre de un protocolo y del programa que lo implementa, y sirve para acceder servidores privados a través de una puerta trasera.

²⁹ Se utiliza para indicar conformidad en las situaciones que se desea confirmar, aprobar o estar de acuerdo con algo.

³⁰ Se utiliza para indicar que existe una posibilidad de peligro, que se debe estar atento ante una proximidad de peligro o falla.

³¹ Se utiliza para indicar que existe una falla, que se está al límite de funcionamiento.

³² No se conoce el estado.

externos para los cuales Nagios pasa información sobre lo que debe verificarse y cuáles son los límites críticos y de advertencia. Los complementos son responsables de realizar los controles y analizar los resultados. El resultado de dicho control es el estado (operativo, cuestionable o fallido) y el texto adicional que describe la información del servicio en detalle. Este texto está destinado principalmente a que los administradores del sistema puedan leer el estado detallado de un servicio (Kocjan & Beltowski, 2016).

Características

Nagios posee como fortaleza principal su flexibilidad gracias a las múltiples formas que un administrador puede configurar el monitoreo en una infraestructura de red. Además cuenta con mecanismos para reaccionar automáticamente a problemas y un sistema de notificación. Entre los principales objetos de Nagios se encuentran: (Kocjan & Beltowski, 2016):

- ✓ **Comandos:** Son definiciones de cómo este sistema debe realizar tipos de chequeos en particular. Son una capa de abstracción sobre los complementos reales que le permite agrupar tipos de operaciones similares.
- ✓ **Períodos de tiempo:** Se trata de intervalos de fecha y hora en los que una operación debería o no debe ser realizada.
- ✓ **Host y grupos de hosts:** Son dispositivos que pueden pertenecer a un grupo y estar separados. Un host puede ser miembro de más de un grupo.
- ✓ **Servicios:** Son funcionalidades o recursos que pueden ser monitoreados en un host, por ejemplo, uso de CPU, espacio de almacenamiento o servidor web.
- ✓ **Notificaciones:** Las notificaciones son una parte esencial de Nagios, las mismas se pueden configurar en dependencia del tipo de error, se pueden enviar a un contacto o a contactos que pertenecen a un determinado grupo. Las notificaciones no son estrictamente un objeto, sino una combinación de todos los objetos anteriores. Las notificaciones tienen las características de ser escalables a otros contactos si lleva un tiempo prolongado en el mismo estado.

1.4.2 - Zabbix

Zabbix se puede caracterizar como un sistema de monitoreo semi-distribuido con administración centralizada. Si bien muchas instalaciones tienen una única base de datos central, es posible usar

monitoreo distribuido con nodos y proxies³³, y la mayoría de las instalaciones usarán agentes Zabbix (Vacche & Lee, 2015).

Características

Zabbix contiene dos componentes principales, Zabbix server, es el servidor central o servidor recolector y el agente Zabbix que es la aplicación que tiene la función de recolectar las distintas métricas de los dispositivos a supervisar. Entre las principales características están (Vacche & Lee, 2015) :

- Interfaz web centralizada y fácil de usar.
- El servidor se puede ejecutar en la mayoría de los sistemas operativos tipo Unix, incluidos Linux, AIX, FreeBSD, OpenBSD y Solaris.
- Agentes nativos para todos los sistemas Unix antes mencionados y para Microsoft Windows.
- Posibilidad de monitorear los dispositivos utilizando SNMP³⁴ y IPMI³⁵.
- Gráficos incorporados y otras capacidades de visualización.
- Notificaciones que permiten una fácil integración con otros sistemas.
- Configuración flexible, basada en plantillas.

El servidor Zabbix está escrito en C³⁶, y contiene una interfaz web escrita en PHP³⁷, ambos pueden convivir en la misma máquina o en servidores diferentes. En el segundo escenario, tanto el servidor Zabbix como la interfaz web necesitan acceso a la base de datos.

1.4.3 - Icinga

Icinga es un sistema de código abierto escalable cuya función principal es el monitoreo de

³³ Es un programa o dispositivo que realiza una tarea de acceso a Internet en lugar de otro ordenador. Un **proxy** es un punto intermedio entre un ordenador conectado a Internet y el servidor que está accediendo

³⁴ El Protocolo Simple de Administración de Red (SNMP, *por sus siglas en inglés*) es un protocolo de la capa de aplicación que facilita el intercambio de información de administración entre dispositivos de red.

³⁵ La Solución inteligente para Gestión de Servidores (IPMI, *por sus siglas en inglés*) es una interfaz de gestión de hardware basada en mensajes estandarizados. Permite gestionar servidores con independencia de sistemas operativos.

³⁶ Es un lenguaje de programación de propósito general que ofrece economía sintáctica, control de flujo y estructuras sencillas y un buen conjunto de operadores.

³⁷ Es un lenguaje de programación del lado del servidor gratuito e independiente de plataforma.

servidores e infraestructura de red, permite alertas sobre problemas y brinda informes de tiempo de actividad. Icinga es una bifurcación del proyecto de Nagios, que apunta a construir muchas características necesarias encima de él. Icinga es compatible con Nagios, por lo que todas las configuraciones y complementos de Nagios se pueden reutilizar con Icinga tal como está (Mehta, n.d.).

Entre los componentes de esta tecnología se encuentra Icinga Core la cual maneja la programación de las verificaciones y el procesamiento de sus resultados. Además determina los estados de los servicios, ya sea, chequeándolos periódicamente (verificaciones activas) o chequeando los resultados que están siendo reportados por un sistema remoto (verificaciones pasivas). Dependiendo de los resultados de los controles y del estado del servicio, se envían las notificaciones si es necesario.

El Core también proporciona todo tipo de información (tiempos de inactividad, resultados de verificación, historial de alertas, registros de ejecución de comandos, entre otros) a través de archivos de registro, que se pueden usar para generar informes en una interfaz web. Los archivos de ejecución y los complementos pueden usar esos archivos de registro para leer el estado actual de los servicios (Mehta, n.d.).

Características

Debido a que Icinga fue basado en Nagios posee muchas características en común con estas tecnologías (Mehta, n.d.) :

- Monitoreo de servicios de red (Correo electrónico, DHCP, SSH, entre otros)
- Monitoreo de recursos (RAM, CPU, almacenamiento).
- Monitoreo de dispositivos de red (interruptores, enrutadores, sensores de temperatura y humedad, etc.).
- Diseño basado en complementos.
- Notificaciones ante la detección de incidencias.
- Escalabilidad de alertas.
- Gráficos basados en complementos.

- Visualización de informes basados en plantillas.

El núcleo de Icinga está escrito en C y tiene una arquitectura modular con núcleo independiente, interfaz de usuario y base de datos en la que los operarios pueden integrar varios complementos. Estos últimos se comunican a través de una capa de abstracción de Doctrine de Icinga y APIs REST, que comunican los datos externos y las estructuras internas. Esta agrupación de componentes permite a los usuarios distribuir el sistema de Icinga para una supervisión redundante. También ofrece a los usuarios la libertad de personalizar Icinga para satisfacer sus necesidades (Mehta, n.d.).

1.4.4 - Valoración de las tecnologías de monitoreo

Los sistemas de monitoreo analizados permiten la supervisión de servidores e infraestructura de red, permiten las principales funcionalidades del monitoreo: notificaciones, alertas escalables, almacenamiento de métricas, generación de informes y visualización de los estados basados en gráficas. Estas tecnologías están orientadas al seguimiento de los servicios (Correo electrónico, SSH, servicio de Internet, entre otros) y recursos de los servidores (Estado de la CPU, memoria RAM, Almacenamiento, entre otros). En el caso de los sistemas de recuperación de información, donde intervienen un grupo numeroso de servidores y servicios, además existen las tecnologías involucradas en la recuperación de información donde es necesario darle un seguimiento al comportamiento de cada una de las etapas de funcionamiento de las mismas.

Por la necesidad de un seguimiento en tiempo real, la generación de mediciones de las distintas métricas puede producir grandes volúmenes de información, datos de los cuales algunos deben persistir en el tiempo, y otros deben permanecer temporalmente. Otra cuestión a resaltar es que no se puede personalizar la toma de mediciones a métricas de un sistema o proceso determinado, pues, la toma de datos a monitorear por estos sistemas está orientada al funcionamiento de los recursos y servicios en un servidor. Por último la flexibilidad en el uso de distintas tecnologías en los componentes de un sistema de monitoreo se limita por el alto acoplamiento que existe, no es posible utilizar otro sistema de notificación, otro tipo de base de datos u otro sistema de visualización de gráficas y tablas. Es por eso que no se deciden utilizar estas tecnologías de monitoreo.

1.5 - Monitoreo basado en series temporales

Los sistemas, las infraestructuras de servidores y las arquitecturas han evolucionado en función de las nuevas demandas y necesidades. Contar con almacenamiento de las mediciones basados en el

tiempo permite mejorar las respuestas a las consultas de determinadas mediciones, pues, las mismas se realizan basadas en el tiempo y esto garantiza que solamente se busquen los datos que se encuentran en el intervalo de tiempo consultado (Leighton et al., 2015).

1.5.1 - Base de datos de series de tiempo

Una Base de Datos de Series de Tiempo (*TSDB, por sus siglas en inglés*) es una base de datos optimizada para datos de series temporales o de tiempo. Las series temporales son mediciones o eventos que son rastreados, monitoreados, muestreados y agregados a lo largo del tiempo. Estos podrían ser métricas asociadas a recursos de un servidor, la supervisión del rendimiento de una aplicación, el tráfico de red, los datos de un sensor, eventos, los clics, los intercambios en un mercado y muchos otros tipos de datos analíticos. La diferencia entre los datos de series de tiempo y los datos regulares es que las consultas son con respecto al tiempo (Deri, Mainardi, & Fusco, 2012; Influxdata, 2018b).

Una base de datos de series de tiempo está diseñada específicamente para el manejo de métricas y eventos o mediciones que tienen marca de tiempo. Una TSDB está optimizada para medir el cambio a lo largo del tiempo. Las propiedades que hacen que los datos de series temporales sean muy diferentes a otras cargas de trabajo de datos son la gestión del ciclo de vida de los mismos, el resumen y escaneos para muchos registros (Dugan, 2014; Faloutsos, Ranganathan, & Manolopoulos, 1994; InfluxData, 2018).

1.5.1 - Plataformas de series temporales

1.5.1.1 - TICK

Es una plataforma de series temporales que proporciona servicios y funcionalidades para acumular, analizar y actuar sobre datos de series de tiempo. Está compuesta por Telegraf, sistema colector de mediciones, InfluxDB como base de datos de series temporales, Chronograf para la visualización de información y Kapacitor para el procesamiento de las alertas.

Telegraf

Telegraf es un agente de servidor basado en complementos para recopilar y generar informes de métricas. Posee complementos o integraciones para obtener métricas directamente del sistema en el que se ejecuta, para extraer métricas de APIs de terceros y para consumir métricas de tecnologías

como StatsD³⁸ y Kafka³⁹. También tiene complementos de salida para enviar métricas a un grupo grande de almacenes de datos, servicios y colas de mensajes, tales como: InfluxDB, Graphite, OpenTSDB, Datadog, Librato, Kafka, MQTT, NSQ y muchos otros (Influxdata, 2018a).

InfluxDB

InfluxDB es una base de datos de series de tiempo construida desde cero para manejar altas cargas de escritura y consulta. Es un almacén de datos de alto rendimiento para datos con marcas de tiempo para métricas de aplicaciones, datos de sensores de IoT⁴⁰ y análisis en tiempo real. Permite establecer la persistencia de los datos durante un tiempo definido. Ofrece un lenguaje de consulta similar a SQL para interactuar con datos (InfluxData, 2018b; Persen, Todd and Winslow, 2016).

Chronograf

Chronograf es una interfaz de administración y el motor de visualización de la plataforma. Permite configurar y mantener el monitoreo y las alertas para su infraestructura. Es simple de usar e incluye plantillas y bibliotecas que le permiten construir tableros rápidamente con visualizaciones en tiempo real de sus datos y crear fácilmente reglas de alerta y automatización (InfluxData, 2018a).

Kapacitor

Kapacitor es un motor de procesamiento de datos nativo. Puede procesar los datos de flujo y lote de InfluxDB. Kapacitor le permite conectar su propia lógica personalizada o funciones definidas por el usuario para procesar alertas con umbrales dinámicos, combinar métricas para patrones, calcular anomalías estadísticas y realizar acciones específicas basadas en estas alertas, como el reajuste de carga dinámico. Kapacitor se integra con HipChat, OpsGenie, Alerta, Sensu, PagerDuty, Slack y más.

1.5.1.2 - Graphite

Graphite es una herramienta de monitoreo que puede ser desplegada desde servidores con bajas prestaciones hasta en una infraestructura en la nube. Se puede utilizar Graphite para rastrear el rendimiento de los sitios web, aplicaciones, servicios comerciales y servidores en red. Marcó el comienzo de una nueva generación de herramientas de monitoreo, presenta tecnologías para almacenar, recuperar, compartir y visualizar datos de series de tiempo (Graphite, 2018).

³⁸ Es un demonio desarrollado y presentado por Etsy para agregar y resumir métricas de aplicación.

³⁹ Es un proyecto de intermediación de mensajes de código abierto desarrollado por la Apache Software Foundation escrito en Scala.

⁴⁰ Internet de las Cosas

Graphite fue diseñado y escrito originalmente por Chris Davis en Orbitz en 2006 como un proyecto paralelo que finalmente se convirtió en su herramienta de monitoreo fundamental. En 2008, Orbitz permitió que Graphite fuera lanzado bajo la licencia de código abierto Apache 2.0 (Graphite, 2018).

Componentes

Graphite presenta tres componentes de software (Graphite, 2018):

- **Carbon:** Es un servicio de alto rendimiento que escucha datos de series de tiempo.
- **Whisper:** Es una biblioteca de base de datos simple para almacenar datos de series de tiempo.
- **Graphite-web:** Es una interfaz de usuario y API de Graphite para renderizar gráficos y tableros.

Arquitectura

Las métricas se alimentan en la pila a través del servicio Carbon, que escribe los datos en las bases de datos de Whisper para el almacenamiento a largo plazo. Los usuarios interactúan con la interfaz de usuario web Graphite o API, que a su vez consulta Carbon y Whisper para obtener los datos necesarios para construir los gráficos solicitados.

La plataforma web de Graphite ofrece una variedad de estilos y formatos de salida, incluidas imágenes en bruto, CSV, XML y JSON, lo que permite a cualquier persona insertar fácilmente gráficos personalizados en otras páginas web o paneles.

1.5.1.3 - Prometheus

Prometheus es una plataforma compuesta por un conjunto de herramientas que permiten el monitoreo y las alertas, es un ecosistema de código abierto originalmente creado en SoundCloud⁴¹. Es un proyecto que presenta una comunidad de usuarios y desarrolladores muy activa. Actualmente es un proyecto independiente de código abierto que se unió a la Cloud Native Computing Foundation en 2016 como el segundo proyecto alojado, después de Kubernetes (Prometheus, 2018).

Características

⁴¹ Es una plataforma de distribución de audio en línea en la que sus usuarios pueden colaborar, promocionar y distribuir sus proyectos musicales.

Entre las características principales de Prometheus son (Prometheus, 2018):

- Presenta un modelo de datos multidimensional con datos de series de tiempo identificados por nombre de la métrica y pares de clave/valor.
- Posee alta flexibilidad en su lenguaje de consulta.
- Permite la recolección de las mediciones a través del protocolo HTTP.
- Presenta múltiples modos de soporte gráfico y de tablero.

Componentes

Entre los componentes que están presentes en el ecosistema están (Prometheus, 2018):

- **Exporters:** Son una serie de bibliotecas que permiten exportar las métricas existentes de sistemas de terceros tales como métricas asociadas al CPU, RAM, entre otros.
- **Prometheus server:** Permite el almacenamiento de los datos de series de tiempo.
- **Alertmanager:** Maneja las alertas enviadas por el servidor Prometheus, es el encargado de notificar a los operarios ante la ocurrencia de una incidencia.

1.5.2 - Valoración de las plataformas de series temporales

Cada una de las plataformas de series temporales cuenta con procesos de recopilación y almacenamiento basado en bases de datos de series temporales, así como herramientas para la evaluación de las mediciones y el envío de notificaciones. Además cuentan con soluciones para la visualización de gráficas y tablas. Luego de realizadas pruebas de conceptos a cada una de las plataformas se decide utilizar Telegraf como solución a la recopilación de métricas, por la flexibilidad que permite en la toma de las métricas asociadas a los recursos y aplicaciones, además que posee un bajo consumo de recursos, un elemento a resaltar, pues debe estar desplegada en cada uno de los servidores de la infraestructura.

Como solución al almacenamiento se decide utilizar InfluxDB por el rendimiento que posee al recibir grandes volúmenes de datos. Como sistema de alertas y notificaciones se decide utilizar Kapacitor, por la flexibilidad que brinda para esta tarea, además de ser parte de la Plataforma TICK en conjunto con Telegraf e InfluxDB. Como solución a la visualización de datos no se utiliza ninguna de las

herramientas presentes en las plataformas estudiadas por la poca flexibilidad que presentan a la hora de elaborar un tablero de datos. Se decide utilizar Grafana ya que permite las características antes mencionadas, además permite restringir el acceso a determinado tablero y la autenticación contra un controlador de dominio.

1.6 - Conclusiones parciales

- El análisis realizado de cada uno de los componentes de la Plataforma C.U.B.A, entre los cuales se encuentran Apache Solr como solución de indexación y Apache Nutch como solución de rastreo, permitieron describir cada una de sus características.
- El estudio de los conceptos y procesos relacionados con el monitoreo permitió determinar los componentes principales para la solución de monitoreo y los elementos esenciales para la descripción de las métricas.
- El estudio de las tecnologías y plataformas de series temporales permitió seleccionar la solución TICK como plataforma de series temporales a utilizar en el marco de la investigación.

Capítulo 2: Implementación de la solución de monitoreo

Con el objetivo de mejorar la pertinencia en el monitoreo del sistema de recuperación de información de la Plataforma C.U.B.A, en el presente capítulo se definen un grupo de métricas para monitorear los recursos de hardware de los servidores y de las aplicaciones desplegadas en la infraestructura de la Plataforma C.U.B.A. Además se proponen un grupo de procesos que intervienen en el ciclo de vida del seguimiento de esas métricas desde su recopilación hasta su visualización y evaluación. Por último se describe la infraestructura tecnológica soportada para esta solución.

2.1 - Solución de monitoreo de la Plataforma C.U.B.A

La solución para el monitoreo del sistema de recuperación de la Plataforma C.U.B.A define en primer lugar los procesos que intervienen en el monitoreo así como el flujo de comunicación de los mismos. Además explica cada una de las métricas que se deben recolectar para su posterior evaluación y visualización. Por último se explica una solución tecnológica basada en una plataforma de series temporales para automatizar toda esta tarea.

2.1.1 - Procesos que intervienen en el monitoreo

La solución de monitoreo de la Plataforma C.U.B.A está compuesta como punto de partida por un grupo de agentes distribuidos en los servidores encargados de la recolección de métricas según las necesidades y responsabilidades (recolección de las métricas). Toda la información se almacena en una base de datos de series de tiempo con políticas de retención de los datos (almacenamiento de las métricas). Por último se visualiza toda la información de manera centralizada donde se puede constatar una vista global sobre el funcionamiento de los servidores y aplicaciones del sistema de recuperación de información de la Plataforma C.U.B.A en tiempo real (visualización de las métricas). Unido a estos tres procesos se cuenta con un mecanismo que envía notificaciones ante incidencias encontradas durante la evaluación de las mediciones (alertas ante incidencias).

La Figura 8 muestra cómo estos elementos se comunican; los recolectores de datos envían las mediciones para un sistema cuya función es almacenar y mantener disponible esa información. Por otra parte está el sistema de alertas que constantemente evalúa las mediciones y en caso de detectar alguna falla, rápidamente notifica al operario o encargado de dar seguimiento al sistema. Por último interviene una solución que toma los datos de la tecnología de almacenamiento y los presenta al usuario mediante gráficas, tablas, etc.

2.1.1.1 - Recolección de datos

Los datos de las operaciones del sistema son recopilados por agentes de servidores, bases de datos y equipos de red. La fuente de datos son registros, estadísticas del dispositivo y mediciones del sistema. Los agentes de recopilación agrupan las entradas en las métricas y les dan un conjunto de propiedades que sirven como una dirección en el espacio y el tiempo. Las entradas se envían posteriormente al sistema de monitoreo a través de un protocolo acordado y se almacenan en la base de datos de métricas (Ligus, 2013).

2.1.1.2 - Almacenamiento de datos

Los datos procedentes de los agentes recolectores se agrupan en dependencia de sus propiedades y, posteriormente, se almacenan en sus respectivas métricas. Las mediciones tomadas se resumen mediante una estadística para generar una serie de tiempo. Los puntos de datos de las series temporales resultantes se envían uno por uno a un motor de evaluación de alarmas y se verifican en busca de anomalías. Cuando se detectan tales condiciones, se activa una alarma y envía una alerta al operador (Ligus, 2013).

2.1.1.3 - Presentación de los datos

El operador puede generar diagramas de series temporales de series de tiempo seleccionadas como una forma de obtener una visión general del estado actual o en respuesta a la recepción de una alerta. Cuando se identifica una falla y se lleva a cabo la acción mitigadora, los gráficos deben dar una respuesta inmediata y reflejar el grado en que la acción correctiva ha ayudado. Cuando no se observa mejora, puede ser necesaria una intervención adicional (Ligus, 2013).

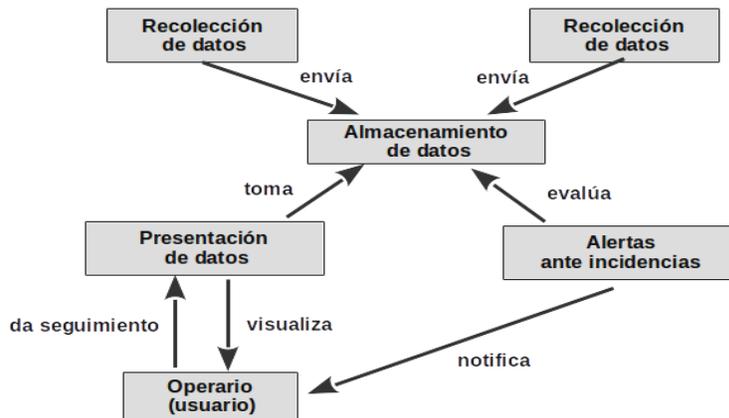
2.1.1.4 - Alertas ante incidencias

Una alerta es una notificación de un problema potencial, que puede tomar una o más de las siguientes formas: correo electrónico, SMS, llamada telefónica, entre otros. Una alarma emite una alerta cuando el sistema pasa por un umbral⁴² y este umbral es detectado por un monitor. Por ejemplo, puede configurar una alarma para que le avise cuando el sistema exceda el 80% de la utilización de la CPU por un período continuo de 10 minutos (Ligus, 2013).

⁴² Consisten en límites (expresados como el número de puntos de datos) y la duración de la infracción.

Un monitor métrico se adjunta a una serie de tiempo y lo evalúa contra un umbral Cuando los puntos

Figura 8: Comunicación entre los elementos que intervienen en una solución de monitoreo. Fuente: Elaboración propia.



de datos que llegan caen por debajo del umbral, superan el umbral o salen del rango definido durante el tiempo suficiente, se dice que el umbral se ha incumplido y el monitor pasa de un estado despejado a otro de alerta. Cuando los puntos de datos caen dentro de los límites del umbral definido, el monitor se recupera y vuelve al estado despejado. Los estados del monitor se utilizan como factores en la evaluación de estados de alarma (Ligus, 2013).

2.1.2 - Recolección de métricas

El monitoreo en la Plataforma C.U.B.A debe cubrir tres grupos principales: *sistema de indexación y rastreo*, *experiencia de usuario* y *disponibilidad de recursos*. Las mediciones para todos los grupos deben ser almacenadas como una serie de tiempo a través de una interfaz común que permite una efectiva identificación de las fuentes de problemas. La cobertura completa de monitoreo abarca el tráfico de red, recursos de hardware, sistema operativo, servicios y las aplicaciones que forma parte del sistema de recuperación de información de C.U.B.A.

2.1.2.1 - Sistema de indexación

El sistema de indexación es una de las herramientas claves dentro de un sistema de recuperación de información, pues, es el punto central entre la recuperación de los documentos y la indexación de los mismos. La evaluación de este proceso es fundamental para entender si el motor de búsqueda se está utilizando efectivamente.

La efectividad, en términos generales, mide la capacidad del motor de búsqueda para encontrar la

información correcta, y la eficiencia mide qué tan rápido se hace. Para una consulta determinada, y una definición específica de relevancia, se puede definir la efectividad como una medida de qué tan bien la clasificación producida por el motor de búsqueda corresponde a una clasificación basada en los juicios de relevancia del usuario.

La eficiencia se define en términos de los requisitos de tiempo y espacio para el algoritmo que produce el orden de relevancia. La búsqueda es un proceso interactivo que involucra diferentes tipos de usuarios con diferentes necesidades de información. En este entorno, la eficacia y la eficiencia se verán afectadas por muchos factores, como la interfaz utilizada para mostrar los resultados de búsqueda y las técnicas de refinamiento de las consultas, como la sugerencia de consulta y la retroalimentación de relevancia.

En el entorno de la investigación se miden las métricas asociadas al funcionamiento del sistema de indexación (Croft et al., 2015). La Tabla 3 relaciona las métricas relacionadas con el sistema de indexación.

Tabla 3: Métricas relacionadas con el sistema de indexación. Elaboración propia.

Nombre de la métrica	Descripción	Punto de datos	Unidad
<i>elapsed_indexing_time</i>	Cantidad de tiempo necesaria para generar un índice de documento en un sistema en particular	<i>indexing</i>	ms ⁴³
<i>indexing_processor_time</i>	Ciclos de CPU utilizados para generar un índice de documento.	<i>indexing</i>	cantidad
<i>query_latency</i>	La cantidad de tiempo que un usuario debe esperar después de emitir una consulta antes de recibir una respuesta.	<i>indexing</i>	ms
<i>query_throughput</i>	Número de consultas procesadas (por horas)	<i>indexing</i>	h ⁴⁴
<i>indexing_temporary_space</i>	Espacio de disco temporal utilizado al crear un índice.	<i>indexing</i>	bytes
<i>indexing_size</i>	Cantidad de almacenamiento utilizado en los archivos de índices.	<i>indexing</i>	bytes

⁴³ En milisegundos

⁴⁴ En horas

<i>indexing_count</i>	Cantidad de documentos indexados	<i>indexing</i>	cantidad
<i>suggestion_count</i>	Cantidad de sugerencias almacenadas	<i>indexing</i>	cantidad

2.1.2.2 - Sistema de rastreo

En un sistema de recuperación de información el rastreo de las web en búsqueda de nueva información o de actualización de los documentos existentes es tan importante como la indexación de los datos, pues, de la calidad del rastreo depende la información que el sistema pueda mostrar antes las necesidades de búsqueda de los usuarios. La web en la actualidad presenta la característica de estar en constante crecimiento y actualización, lo que lleva a contar con un sistema de rastreo que esté funcional todo el tiempo. La evaluación de todas las etapas de rastreo, así como las estadísticas brindadas al concluir cada etapa son de suma importancia para evaluar el buen desempeño de este proceso. La Tabla 4 muestra las métricas que se miden durante la tarea de rastreo.

Tabla 4: Métricas relacionadas con el sistema de rastreo. Elaboración propia.

Nombre de la métrica	Descripción	Punto de datos	Unidad
total_url	Total de URL rastreadas	crawling	cantidad
url_not_found	Total de URL no encontradas	crawling	cantidad
url_fetched	Total de URL visitadas	crawling	cantidad
url_nofetched	Total de URL no visitadas	crawling	cantidad
url_redir_perm	Total de URL redirigidas	crawling	cantidad
connection_failures	Total de conexiones fallidas	crawling	Cantidad
total_host	Total de sitios visitados	crawling	Cantidad
<i>elapsed_time</i>	Tiempo transcurrido (por cada etapa)	crawling	ms

2.1.2.3 - Experiencia de usuario

El monitoreo del usuario final permite a las organizaciones determinar cómo los usuarios reales interactúan con sus aplicaciones y cómo estas se están desempeñando para ellos. Las métricas clásicas de comportamiento del usuario que se usan en los sitios web son el tiempo promedio que se pasa en el sitio, la cantidad de visitantes, dónde el usuario da clic, entre otros (Roden, 2010). En el entorno de la Plataforma C.U.B.A se recopilan métricas relacionadas con las visitas (nacionales y extranjeras), tanto de personas como de otros sistemas⁴⁵, las consultas de los distintos módulos (métricas que evidencian si realmente le brinda utilidad un determinado módulo para los usuarios), entre otros elementos.

⁴⁵ Principalmente robots que indexan la interfaz de la Plataforma C.U.B.A.

Tabla 5: Métricas relacionadas con la experiencia de usuario. Elaboración propia.

Nombre de la métrica	Descripción	Punto de datos	Unidad
visit_count	Cantidad de visitantes.	user_experience	cantidad
visit_count_national	Cantidad de visitantes nacionales.	user_experience	cantidad
visit_count_foreign	Cantidad de visitantes extranjeros (divididos por países).	user_experience	cantidad
query_count	Cantidad de consultas (divididas por los distintos módulos web, imágenes y documentos)	user_experience	cantidad
query_count_national	Cantidad de consultas realizadas nacionalmente.	user_experience	cantidad
query_count_national	Cantidad de consultas realizadas desde el extranjero.	user_experience	cantidad
query_count_per_user	Cantidad de consultas por usuarios.	user_experience	cantidad
query_trend	Tendencias de búsquedas realizadas en un intervalo de tiempo.	user_experience	lista de consulta
query_result	Cantidad de resultados por consulta	user_experience	cantidad

2.1.2.4 - Recursos

Cada tarea o acción que se ejecuta en el sistema trae consigo utilización de ciclos de CPU, muchas de estas acciones o tareas también requieren de memoria. El intercambio de información entre los servidores ocupan ancho de banda y los datos necesitan estar almacenados, ocupando espacio en los dispositivos para este tipo. De igual forma las aplicaciones requieren constantemente efectuar escrituras y lecturas (*E/S, I/O por sus siglas en inglés*). Los patrones de uso de los recursos cambian en dependencia de la carga, es por ello la necesidad de darles seguimiento a esos patrones de cambio para identificar a tiempo patrones de uso y picos de utilización. El monitoreo de la utilización de los recursos ayuda a identificar esto, además requieren una atención cercana y constante (Ligus, 2013).

El costo de cualquier actividad del usuario se puede expresar en términos de los recursos que utiliza, por lo que es tan importante vigilar de cerca la saturación de los recursos en todo momento. Una acción computacional típica es una petición a un sitio web. Cada petición requiere recursos: como mínimo, consume memoria y ciclos de CPU, pero con frecuencia también lee y escribe algunos

datos en otros dispositivos, como unidades de disco, lo que introduce un mayor costo de E/S.

El agotamiento de cualquier recurso requerido para atender una solicitud lleva a la creación del llamado cuello de botella de rendimiento. Los patrones de uso no se pueden predecir con un 100% de confiabilidad y la escasez de recursos no siempre se puede prevenir con una planificación de capacidad precisa o una asignación dinámica. Considere los ataques de Denegación de Servicio (DoS), donde el objetivo del atacante es cerrar el servicio impulsando la saturación del recurso más escaso que puede manipular. El monitoreo de recursos computacionales en el contexto del uso del sistema es necesario para discernir patrones y reaccionar en consecuencia (Ligus, 2013).

En la solución de monitoreo de la Plataforma C.U.B.A asociada a los recursos de hardware se deben recolectar métricas relacionadas con la utilización de la memoria RAM, tráfico de la red, CPU, procesos, sistema, dispositivos de almacenamiento, entre otros. Los datos sobre la utilización de los recursos y la disponibilidad deben recopilarse directamente de los dispositivos que proporcionan los recursos. La frecuencia de recopilación de las mediciones debe ser entre 30 segundos y 1 minuto.

2.1.2.4.1 - Memoria RAM

La memoria de acceso aleatorio (*RAM, por sus siglas en inglés*) es un almacenamiento volátil donde se guardan datos del sistema operativo y aplicaciones para que el procesador del dispositivo pueda acceder rápidamente a ellos. La memoria RAM es mucho más rápida de leer y escribir que otros tipos de almacenamientos, como una unidad de disco duro (HDD) o unidad de estado sólido (SSD) (Techtarget, 2017). Por la utilización de este recurso por la mayoría de las aplicaciones y del sistema operativo es necesario darle seguimiento. La Tabla 6 muestra las métricas a tomar:

Tabla 6: Métricas para la memoria RAM. Elaboración propia.

Nombre de la métrica	Descripción	Punto de datos	Unidad
active	Memoria RAM activa	<i>mem</i>	bytes
cached	Memoria RAM en caché	<i>mem</i>	bytes
available	Memoria RAM disponible	<i>mem</i>	bytes
free	Memoria RAM libre	<i>mem</i>	bytes
inactive	Memoria RAM inactiva	<i>mem</i>	bytes
total	Memoria RAM total	<i>mem</i>	bytes
used	Memoria RAM usada	<i>mem</i>	bytes
available_percent	Porcentaje disponible de memoria RAM	<i>mem</i>	porcentaje

used_percent	Porcentaje utilizado de memoria RAM	<i>mem</i>	porcentaje
buffered	Memoria RAM Almacenada	<i>mem</i>	bytes

2.1.2.4.2 - CPU

La unidad de procesamiento central (CPU, por sus siglas en inglés) es la unidad que realiza la mayor parte del procesamiento dentro de una computadora. Para controlar las instrucciones y el flujo de datos hacia y desde otras partes de la computadora, la CPU depende en gran medida de un conjunto de chips, que es un grupo de microchips ubicados en la placa base (Techopedia, 2018).

La CPU tiene dos componentes típicos:

- Unidad de control: extrae las instrucciones de la memoria y las decodifica y ejecuta.
- Unidad de lógica aritmética (ALU): maneja operaciones aritméticas y lógicas.
- La Tabla 7 reúne las métricas que se van a recolectar asociadas a la CPU (Turnbull et al., 2017).

Tabla 7: Métricas relacionadas con la CPU. Elaboración propia.

Nombre de la métrica	Descripción	Punto de datos	Unidad
usage_user_percent	Porcentaje de uso de usuario	<i>cpu</i>	porcentaje
usage_idle_percent	Porcentaje de uso inactivo	<i>cpu</i>	porcentaje
usage_system_percent	Porcentaje de uso del sistema	<i>cpu</i>	porcentaje
usage_nice_percent	Porcentaje de uso aceptable	<i>cpu</i>	porcentaje
used_percent	Porcentaje utilizado	<i>cpu</i>	porcentaje

2.1.2.4.3 - Red

Los datos entregados a través de la red viajan con una latencia⁴⁶ a veces notable. Un enlace de red también tiene un rendimiento limitado, definido como la cantidad de información transmitida por unidad de tiempo. La latencia debe mantenerse al mínimo y cuanto mayor sea el rendimiento, mejor. Las interrupciones de transmisión se pueden expresar en términos de latencia aumentada y rendimiento reducido (Ligus, 2013).

Las tecnologías que intervienen en el sistema de recuperación de información de la Plataforma C.U.B.A se encuentran distribuidas en distintos servidores por lo que cualquier interrupción de la red

⁴⁶ Es el tiempo de demora causado por el procesamiento digital y los medios físicos de transporte.

se manifestará inevitablemente en el desempeño general del sistema. Las aplicaciones se diseñan con la suposición de que la conexión de la red siempre está activa, pero no siempre es así. Una inestabilidad de la red producirá la pérdida de paquetes, la latencia de la red afectará el rendimiento de la aplicación y el ancho de banda de la red será limitado. Por estas razones, la red debe ser monitoreada de cerca (Deutsch & Microsystems, 2013). La Tabla 8 muestra las métricas relacionadas con la red.

Tabla 8: Métricas relacionadas a la red. Elaboración propia.

Nombre de la métrica	Descripción	Punto de datos	Unidad
bytes_sent	Bytes enviados	net	bytes
bytes_recv	Bytes recibidos	net	bytes
err_in	Errores de entrada	net	cantidad
err_out	Errores de salida	net	cantidad
packers_recv	Paquetes recibidos	net	cantidad
packers_sent	Paquetes entregados	net	cantidad
conn_status	Estado de la conexión	net	booleano

2.1.2.4.4 - Sistema operativo

La supervisión del sistema operativo examina el uso de los recursos a nivel de sistema en general, toma mediciones asociadas a la carga media, baja y alta de todos los núcleos de CPU que están presentes en el servidor. Además proporciona datos generales asociados a cantidad de usuarios activos y números de núcleos presentes. Por último informa el tiempo que un servidor lleva encendido, un dato que evidencia si el servidor está apagado o fue reiniciado recientemente. La Tabla 9 describe las métricas a medir del sistema (Ligus, 2013; Massie, Li, Nicholes, & Vuksan, 2013).

Tabla 9: Métricas relacionadas con el sistema. Elaboración propia.

Nombre de la métrica	Descripción	Punto de datos	Unidad
load_cpu1	Promedio de carga de un minuto	system	bytes
load_cpu15	Promedio de carga de 15 minutos	system	bytes
load_cpu5	Promedio de carga de 5 minutos	system	bytes
n_cpus	Cantidad de núcleos presentes	system	cantidad
n_users	Cantidad de usuarios activos	system	cantidad
uptime	Tiempo de encendido	system	tiempo
cpu_speed	Velocidad de CPU en términos de MHz	system	cantidad

	(recopilada una vez)		
--	----------------------	--	--

2.1.2.4.5 - Almacenamiento

El rendimiento del almacenamiento tiene un efecto directo en el rendimiento de las aplicaciones y las experiencias de los usuarios, por lo que los administradores deben implementar una supervisión integral para darle seguimiento a la unidades de almacenamiento. Entre las ventajas que proporciona este tipo de monitoreo están (V. Wu & Eagle, 2015):

- Advertencias en tiempo real de problemas graves, por lo general notificando a un operario cuando se exceden umbrales relacionados con la utilización y las operaciones de lecturas y escrituras en disco (IOPS, por sus siglas en inglés).
- Ofrece datos históricos importantes sobre el almacenamiento que podrían revelar tendencias de uso y ayudar a predecir las necesidades futuras de recursos antes de que el rendimiento del almacenamiento se vea afectado.

La Tabla 10 explica las métricas relacionadas con el almacenamiento, se deben recolectar las mediciones por cada dispositivo existente en el servidor de manera independiente (Turnbull et al., 2017).

Tabla 10: Métricas relacionadas con el almacenamiento. Elaboración propia.

Nombre de la métrica	Descripción	Punto de datos	Unidad
free	Espacio libre (en bytes)	<i>disk</i>	bytes
total	Espacio total (en bytes)	<i>disk</i>	bytes
used	Espacio utilizado (en bytes)	<i>disk</i>	bytes
used_percent	Porcentaje de espacio utilizado.	<i>disk</i>	porcentaje
inodes_free	Cantidad de inodos libres	<i>disk</i>	cantidad
inodes_total	Cantidad de inodos totales	<i>disk</i>	cantidad
Inodes_used	Cantidad de inodos usados	<i>disk</i>	cantidad
reads	Cantidad de lecturas	<i>diskio</i>	cantidad
writes	Cantidad de escrituras	<i>diskio</i>	cantidad
read_bytes	Bytes leídos	<i>diskio</i>	bytes
write_bytes	Bytes escritos	<i>diskio</i>	bytes

iops_in_progress	Cantidad de operaciones de I/O	<i>diskio</i>	cantidad
read_time	Tiempo de lectura (NS ⁴⁷)	<i>diskio</i>	tiempo
write_time	Tiempo de escritura (NS)	<i>diskio</i>	tiempo

2.1.2.4.6 - Procesos

Los procesos son instancias de un programa en ejecución, la mayoría de los sistemas operativos modernos presentan la característica de ser multitarea lo que indica que múltiples instancias de un programa pueden ejecutarse simultáneamente. Cada proceso que se inicia es identificado con un número de identificación único conocido como *Process ID (PID)*, que es siempre un número natural (Turnbull et al., 2017). Tener el control de cuántos procesos se están ejecutando en cada momento es de suma importancia para darle seguimiento al rendimiento de los servidores. La Tabla 11 reúne las métricas relacionadas con los procesos que se encuentran en ejecución en el sistema, esto le brinda al operario una vista general de cuántos procesos e hilos de ejecución están presentes.

Tabla 11: Métricas relacionadas con los procesos. Elaboración propia.

Nombre de la métrica	Descripción	Punto de datos	Unidad
total	Cantidad de procesos	<i>procesos</i>	cantidad
running	Cantidad de procesos en ejecución.	<i>procesos</i>	cantidad
sleeping	Cantidad de procesos en espera.	<i>procesos</i>	cantidad
stopped	Cantidad de procesos detenidos.	<i>procesos</i>	cantidad
blocked	Cantidad de procesos bloqueados.	<i>procesos</i>	cantidad
zombies	Cantidad de procesos zombis.	<i>procesos</i>	cantidad
unknown	Cantidad de procesos desconocidos.	<i>procesos</i>	cantidad
total_threads	Cantidad de hilos en ejecución	<i>procesos</i>	cantidad

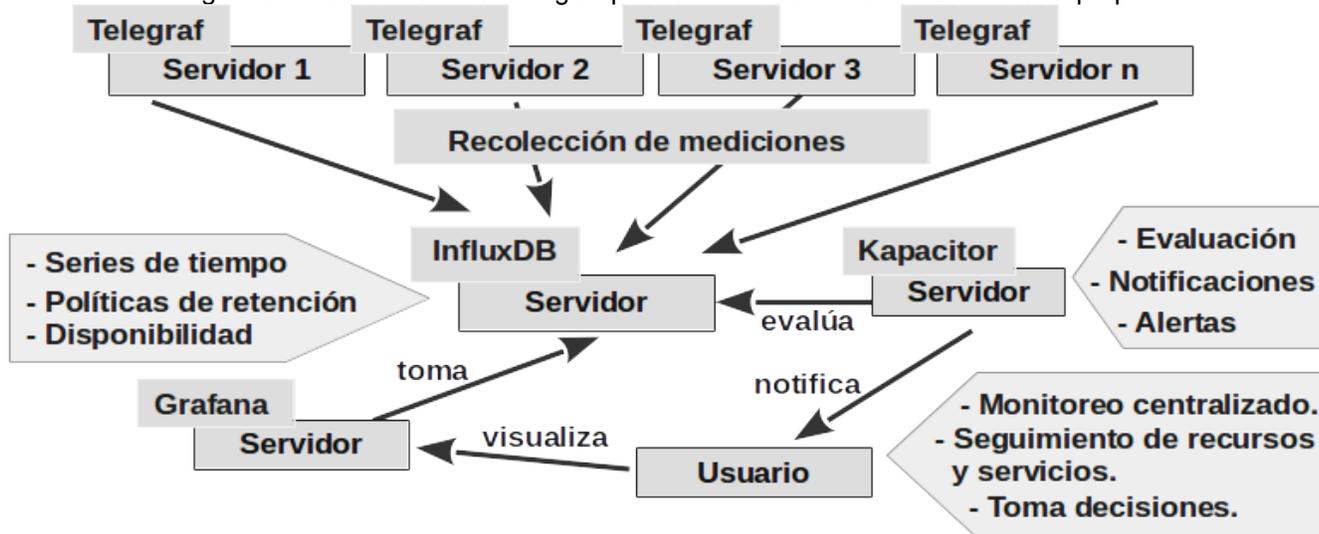
2.1.3 - Infraestructura tecnológica

Por la necesidad de llevar el seguimiento de los recursos de hardware de los 19 servidores que actualmente están presentes en el despliegue del sistema de recuperación de información de la Plataforma C.U.B.A y de sus aplicaciones se propone una infraestructura basada en las tecnologías presentes en las plataformas de series temporales. La Figura 9 describe el funcionamiento y las

⁴⁷ Nanosegundos

tecnologías que se utilizan, como agente recolector de métricas se utiliza Telegraf, como base de datos de series de tiempo InfluxDB, como tecnología para la visualización centralizada Grafana y como sistema de alertas y notificaciones Kapacitor. Cada una de estas tecnologías se despliega en distintos servidores, agregando a la Plataforma C.U.B.A cuatro nuevos nodos.

Figura 9: Infraestructura tecnológica para el monitoreo. Fuente: Elaboración propia.



2.2 - Conclusiones parciales

- La solución de monitoreo de la Plataforma C.U.B.A está compuesta por tres grupos de métricas: métricas para el seguimiento del sistema de indexación y rastreo, métricas para el seguimiento de la experiencia de usuario y métricas para el seguimiento de los recursos de hardware. De cada una de las métricas se definieron el nombre de la métrica, el punto de datos a la cual pertenece y la unidad de medición.
- La infraestructura tecnológica está basada en la plataforma de series de tiempo TICK teniendo como sistema recolector de mediciones a Telegraf, como base de datos de series de tiempo a InfluxDB, como sistema de visualización a Grafana y como sistema de alertas a Kapacitor.

Capítulo 3: Validación de la propuesta de solución

La validación permite exponer los resultados de una investigación, en el presente capítulo se realizan un conjunto de pruebas que permite la evaluación de la pertinencia de la solución de monitoreo del sistema de recuperación de información de la Plataforma C.U.B.A. Se toma como punto de partida un entorno real, aplicando esta solución en la infraestructura de servidores que se encarga de brindar los servicios de esta plataforma. Se muestran los resultados de la aplicación del criterio de expertos, a través del método Delphi, a expertos en sistemas de recuperación de información y a los administradores de redes y aplicaciones de RI de la Plataforma C.U.B.A. Además se describe un análisis comparativo en el seguimiento del sistema de indexación aplicando la solución de monitoreo en un escenario de pruebas.

3.1 - Validación de la solución de monitoreo a través del criterio de expertos

Con la recopilación de mediciones asociadas a todas las métricas descritas en la presente investigación se procede a aplicar el método Delphi (Okoli & Pawlowski, 2004, pp. 5–6) con el objetivo de ponderar un grupo de indicadores a fin de autoevaluar la solución de monitoreo. Este es un método que tiene su base en la subjetividad del individuo escrutador, particularmente en su experiencia sobre un tema determinado (Okoli & Pawlowski, 2004). Entre las ventajas que este método ofrece están (von der Gracht, 2012):

- El reconocimiento del buen juicio del ser humano, en particular de un grupo conocedor.
- Puede eliminar sesgos cuando las valoraciones se basan en un conjunto de opiniones fundadas en el conocimiento o la experiencia.
- El anonimato preserva la objetividad.
- Reduce o elimina coaliciones o sesgos de grupos de individuos.

Por la experiencia en el área de la recuperación de información, y a juicio del autor, se confeccionó un listado de 10 especialistas que cumplen con los requisitos de expertos y a los cuales se les aplicará el instrumento denominado *coeficiente de competencia de experto*, diseñado para calcular el grado de experticia sobre el tema en cuestión. Se tomaron en consideración los siguientes aspectos: título universitario, categoría docente y científica, años de experiencia en los temas de RI, el nivel de dominio sobre el tema que se encuesta y las fuentes de argumentación. A todos estos expertos les

fue enviado un cuestionario (ver Anexo 4), que fue respondido por 10 de ellos.

Para la determinación del coeficiente de competencia de los expertos (**Kcomp**) se utilizó la fórmula $Kcomp = (Kc + Ka) * 0.5$. Donde *Kc* representa el coeficiente de conocimiento que tiene el experto acerca del tema, y se calcula a partir de su propia valoración dentro de una escala del 0 (el experto no tiene ningún conocimiento) al 10 (el experto tiene pleno conocimiento) multiplicada por 0,1. En el Anexo 1 se muestra la Tabla 24 con la autovaloración de cada experto y el coeficiente de conocimiento.

Por otra parte *Ka* representa el coeficiente de argumentación. Es el que resulta de sumar los grados de influencia que el sujeto considera que distintas fuentes de argumentación han tenido, en el conocimiento acumulado por él respecto a un tema en particular. La ponderación de esos grados de influencia, están determinados por criterios relacionados con la importancia de esas fuentes, a partir de la siguiente Tabla 12.

Tabla 12: Tabla patrón para la estimación de las fuentes de argumentación.

Nº	Fuentes de Argumentación:	Grados de influencia de c/u de las fuentes en su conocimiento y criterios		
		Alta (A)	Media (M)	Baja (B)
1	Análisis teórico sobre la temática realizado por usted	0,3	0,2	0,1
2	Experiencia adquirida	0,5	0,4	0,2
3	Conocimiento sobre el tema generado por discusiones e intercambios académicos	0,05	0,04	0,03
4	Trabajo de autores nacionales que conoce y han trabajado la temática.	0,05	0,04	0,03
5	Participación en grupos diseñadores de programas, materiales e iniciativas.	0,05	0,04	0,03
6	Participación en proyectos de investigación y/o desarrollo de artículos o ponencias sobre el tema	0,05	0,04	0,03

Los valores de *Ka* correspondientes a cada experto se encuentran registrados en la Tabla 26 del Anexo 1. Finalmente se calcularon los coeficientes de competencias (*Kcomp*) según la fórmula antes descrita y se definieron los niveles de competencia según los intervalos que se muestran:

- Si $0.8 < Kcomp < 1.0$ coeficiente de competencia alto.

- Si $0.5 < K_{comp} < 0.8$ coeficiente de competencia medio
- Si $K_{comp} < 0.5$ coeficiente de competencia bajo.

Como se puede apreciar en los resultados de la Tabla 26 el 80% de los expertos obtuvo un alto nivel de competencia, el 20% nivel medio y ninguno con bajo nivel de competencia.

3.2 - Validación de la solución de monitoreo por los expertos seleccionados.

Para la validación de la solución de monitoreo del sistema de recuperación de información de la Plataforma C.U.B.A por los expertos se diseñó un cuestionario que se muestra en el Anexo 2 el cual les fue enviado en conjunto con un documento que explica la solución de monitoreo y un enlace donde los expertos pueden constatar el resultado de la investigación a través de la herramienta gráfica que muestra en tiempo real, las gráficas de las evaluaciones de las métricas descritas en la investigación. Se presentaron los diez aspectos definidos para valorar esta solución, cada uno de los cuales cada experto midió con una valoración de Muy adecuado (MA), Bastante adecuado (BA), Adecuado (A), Poco adecuado (PA) e Inadecuado (I), además se les permitió dar algún criterio u opinión acerca de la solución.

Los resultados de la evaluación realizada por los expertos a cada uno de los aspectos propuestos aparecen en la Tabla 25. Sobre la base de la tabla anterior, se determina la frecuencia absoluta por aspectos (ver Tabla 28) y luego la distribución de frecuencia acumulada de cada fila (ver Tabla 29). A partir de la Tabla 29 se calcula la distribución de frecuencias relativas acumuladas de cada fila (ver Tabla 30).

Una de las principales dificultades en la evaluación y validación de cuestionarios mediante el método Delphi proviene de la subjetividad de los criterios emitidos por el grupo de expertos y, por consiguiente, su dificultad a la adaptación de un modelo matemático para su análisis (Goodman, 1987; Hasson & Keeney, 2011).

En el desarrollo del cuestionario de evaluación remitido a los expertos se emplearon valores de escala nominales (Nada adecuado, Poco adecuado, Adecuado, Bastante Adecuado y Muy adecuado) asociados de forma automática a indicadores ordinales (1, 2, 3, 4 y 5 respectivamente).

Para comprobar el acierto en dicha asunción de equivalencia entre escalas se empleó el modelo de Torgerson (Guttman & Torgerson, 1959) a través de un reescalamiento de forma que se logre una

mayor objetividad en el tratamiento de la información al convertir la escala ordinal original (cualitativa) en una escala de intervalo (cuantitativa) (Medina et al., 2011) que permita la valoración de cada uno de los ítem de forma individual.

En la Tabla 13, se realiza el análisis estadístico final donde se calcula: los percentiles de la distribución normal estándar correspondientes a cada una de las frecuencias relativas acumuladas (que se consideran una aproximación de la probabilidad acumulada), la suma algebraica de todos los percentiles anteriores, los puntos de corte (media de los percentiles de cada categoría evaluativa), la suma algebraica de los percentiles dividida por el producto de la cantidad de aspectos sometidos a consulta y la cantidad de categorías evaluativas empleadas(N), la media de los percentiles de cada aspecto sometido a consulta (P), la diferencia (N-P) para cada aspecto analizado (filas). Cada uno de los resultados que se obtienen en la columna (N-P) se comparan con los puntos de corte y se determina en qué categoría evaluativa se encuentra cada aspecto sometido a consulta de los expertos.

Tabla 13: Determinación de la curva normal estándar inversa y los puntos de corte

Aspecto	5 (MA)	4 (BA)	3 (A)	2 (PA)	1 (NA)	Promedio (P)	Diferencia (N-P)	Clasific.
1	0.52	3.50	3.50	3.50	3.50	2.90	-0.37	MA
2	3.50	3.50	3.50	3.50	3.50	3.50	-0.97	MA
3	0.84	3.50	3.50	3.50	3.50	2.97	-0.44	MA
4	0.84	3.50	3.50	3.50	3.50	2.97	-0.44	MA
5	0.25	3.50	3.50	3.50	3.50	2.85	-0.32	MA
6	0.84	3.50	3.50	3.50	3.50	2.97	-0.44	MA
7	0.84	3.50	3.50	3.50	3.50	2.97	-0.44	MA
8	0.84	3.50	3.50	3.50	3.50	2.97	-0.44	MA
9	0.52	-0.84	3.50	3.50	3.50	2.04	0.49	BA
10	-0.52	-0.25	-0.52	3.50	3.50	1.14	1.39	BA
Puntos de Corte (PC)	0.85	2.69	3.09	3.50	3.50			
	Valor límite (N = 2.53)							

Finalmente se ubican en una recta numérica (ver Tabla 14) todos los puntos de corte para cada una de las categorías y la diferencia (N-P) para cada uno de los aspectos sometidos a consulta.

Tabla 14: Rangos y puntos de corte

1 (-0.37)	2 (-0.97)	3 (-0.44)	4 (-0.44)	5 (-0.32)	6 (-0.44)	7 (-0.44)	8 (-0.44)	9 (0.49)	10 (0.49)	-	-	-	
							0.85 (MA)			2.69 (BA)	3.09 (A)	3.50 (PA)	3.50 (NA)

A continuación se procede a determinar el grado de acuerdo (Astigarraga, 2006) entre los expertos a partir de los resultados anteriores utilizando la matriz correspondiente donde se determinó el coeficiente de concordancia (C) para cada uno de los aspectos, a partir de la expresión:

$$C = 100 * \left(1 - \frac{Ds}{Xm}\right)$$

donde Ds es la desviación estándar calculada mediante la fórmula:

$$Ds = \sqrt{\left(\frac{1}{n-1}\right) \sum_{i=1}^{10} (Xi - Xm)^2}$$

y Xm equivale a la media del criterio de los expertos por indicador y se calcula mediante la fórmula:

$$Xm = \frac{\sum_{i=1}^{10} CEi}{10}$$

Cada coeficiente de concordancia debe tener un valor superior a 75, los valores para cada indicador se muestran en la Tabla 15.

A partir de los resultados del análisis de concordancia (C) (ver Tabla 15), se estableció que en la población de los 10 expertos seleccionados y los 10 aspectos evaluados de acuerdo con las puntuaciones de las categorías arriba descritas, alcanzan la condición de Muy adecuado y de Bastante adecuado y se obtuvo un grado de concordancia superior a 80 en todos los aspectos, por lo que se considera que los resultados obtenidos son válidos y fundamentan los criterios dados por los expertos.

Con el acuerdo entre los expertos respecto a la evaluación de la solución de monitoreo permite confirmar su aplicabilidad y correcto nivel de influencia para darle seguimiento al sistema de recuperación de información de la Plataforma C.U.B.A. Este proceso de validación permitió

enriquecer la propuesta a partir de los criterios establecidos por los expertos seleccionados. De los resultados obtenidos se pueden llegar a las siguientes conclusiones:

Tabla 15: Coeficiente de concordancia por aspectos.

Expertos	Aspectos									
	1	2	3	4	5	6	7	8	9	10
1	5	5	5	5	5	4	5	4	5	4
2	4	5	5	5	5	5	5	5	5	5
3	5	5	5	5	4	4	5	5	5	5
4	5	5	5	5	4	5	5	5	5	3
5	5	5	4	5	4	5	5	4	5	4
6	5	5	5	4	5	5	4	5	5	3
7	4	5	5	5	5	5	5	5	4	4
8	5	5	4	5	4	5	5	5	5	5
9	4	5	5	4	5	5	4	5	4	3
10	5	5	5	5	5	5	5	5	3	4
Suma	47	50	48	48	46	48	48	48	46	40
Xm	4.7	5.0	4.8	4.8	4.6	4.8	4.8	4.8	4.6	4.0
Ds	0.48	0.73	0.42	0.42	0.52	0.42	0.42	0.42	0.73	0.81
Ds/Xm	0.10	0.14	0.09	0.09	0.11	0.09	0.09	0.09	0.16	0.20
C	90	86	91	91	89	91	91	91	84	80

- La solución de monitoreo para el sistema de recuperación de información de la Plataforma C.U.B.A a través de sus gráficas de tendencias de las métricas asociadas a los recursos permite tomar una decisión inmediata sobre el correcto/incorrecto uso de los mismos.
- Permite dar un seguimiento en tiempo real del mecanismo de indexación y de las diferentes etapas del sistema de rastreo, detectando de inmediato cualquier incidencia respecto al correcto funcionamiento de estas aplicaciones.
- El sistema de alerta permite notificar de forma inmediata a los implicados correspondientes sobre alguna deficiencia acerca de la disponibilidad de la plataforma u alguna incidencia.

Como conclusión de la evaluación de los expertos se puede arribar que existe una mejora notable

en la pertinencia del seguimiento de los recursos y servicios de la Plataforma C.U.B.A.

3.3 - Validación de la pertinencia a través de un análisis comparativo

Para ilustrar el funcionamiento de la solución de monitoreo del sistema de recuperación de información se propone la realización de un análisis comparativo en dos escenarios: aplicando la solución de monitoreo y sin aplicarla.

3.3.1 - Objetivo de la prueba

El objetivo de la prueba es validar la mejora de la pertinencia en el seguimiento de los recursos y aplicaciones de la Plataforma C.U.B.A aplicando la solución de monitoreo. Durante la prueba se medirán los indicadores de pertinencia (25023:2017, 2016) a través de los indicadores descritos en la Tabla 16:

Tabla 16: Indicadores de medición de la variable pertinencia

Nombre	Función de medición
Pertinencia funcional del objetivo de uso.	$X = 1 - A/B$ <p>A = Número de funciones faltantes o incorrectas entre aquellas que son necesarias para lograr un objetivo de uso específico. B = Número de funciones necesarias para alcanzar un objetivo de uso específico.</p>
Pertinencia funcional del sistema.	$X = \sum_{i=1}^n A_i/n$ <p>A_i Puntuación de aptitud para el objetivo de uso i, es decir, el valor medido de FAp-1-G para el i-ésimo objetivo de uso específico. n = Número de objetivos de uso.</p>

3.3.2 - Escenario de la prueba

Durante la prueba se medirá el cumplimiento de cinco objetivos relacionados con las visitas, las búsquedas, el servicio de indexación, los recursos de hardware y el sistema de alertas. Para el éxito de cada uno de los objetivos deben cumplirse un grupo de tareas. Para la prueba se medirán los indicadores durante 48 horas de seguimiento, en una primera etapa, sin aplicar la solución de monitoreo y en un segundo momento aplicándola. Para el objetivo del seguimiento de los recursos de

hardware se utilizará el servidor de indexación cuyas características están reflejadas en la Tabla 17.

Tabla 17: Características del servidor de indexación

Identificador del servidor	CPU	HDD	RAM	Servicios
SOLRE01	4 GHz	300 GB	8 GB	Apache Solr, Telegraf, Nginx

3.3.3 - Objetivos a cumplir

A continuación se describen los objetivos a cumplir.

Tabla 18: Objetivos a cumplirse para dar seguimiento a la Plataforma C.U.B.A

Nº	Objetivo	ID	Funciones	Periodicidad
O1	Seguimiento de visitantes	F1	Cantidad visitantes son nacionales	Cada 1 hora
		F2	Cantidad visitantes son extranjeros	Cada 1 hora
		F3	Cantidad visitantes en total	Cada 1 hora
		F4	Representación en un mapa de calor	-
O2	Seguimiento a las búsquedas	F5	Cantidad de búsquedas totales	Cada 1 horas
		F6	Cantidad de búsquedas en imágenes	Cada 1 hora/diario
		F7	Cantidad de búsquedas en documentos	Cada 1 hora/diario
		F8	Cantidad de búsquedas en la web	Cada 1 hora/diario
		F9	Tendencias de búsquedas (frases)	Cada 1 hora/diario
		F10	Resultados promedios	Cada 1 hora/diario
		F11	Tiempo de respuesta promedio	Cada 1 hora/diario
O3	Seguimiento al sistema de indexación.	F12	Cantidad de documentos indexados	Cada 1 hora/diario
		F13	Cantidad de imágenes indexadas	Cada 1 hora/diario
		F14	Cantidad de frases indexadas	Cada 1 hora/diario
O4	Seguimiento a los recursos del servidor de indexación.	F15	Comportamiento de la memoria RAM	Cada 5 minutos
		F16	Comportamiento de la carga de la CPU	Cada 5 minutos
		F17	Comportamiento del almacenamiento	Cada 5 minutos
O5	Respuesta ante una incidencia	F18	Simulación de la caída del servicio de indexación	
		F19	Simulación del completamiento del almacenamiento del servidor de indexación al 98%	

3.3.4 - Análisis de la pertinencia

Antes de la realización de los cálculos finales se recogieron (ver Tabla 19) observaciones relacionadas con algunas de las funciones que se debían cumplir.

Tabla 19: Observaciones de las funciones a cumplir

	F	C / PC/ NC	Observación
Aplicando la solución de monitoreo.	F1	C	A través de la visualización de una gráfica de tendencia fue posible darles seguimiento en tiempo real a los visitantes nacionales.
	F9	C	A través de la visualización de una tabla que mostraba la frase y la cantidad de veces que fue buscada, fue posible medir en tiempo real cuáles eran las tendencias de búsquedas durante el día.
	F12	C	A través de la visualización de las gráficas y tablas fue posible determinar en cualquier instante de tiempo la cantidad de documentos indexados.
	F17	C	A través de la visualización de las gráficas de los recursos fue posible darle seguimiento en tiempo real al comportamiento del almacenamiento y hacer análisis de picos y tendencias.
	F18	C	Con la solución de alertas y notificaciones, a través del servicio de correo, fue posible detectar al instante que el servicio de indexación no estaba disponible, lo que propició la toma de medidas de inmediato.
	F19	C	Con la solución de alertas y notificaciones, a través del servicio de correo, fue posible detectar al instante que uno de los servidores de indexación estaba al 98 % de datos, lo que propició la toma de medidas de inmediato.
Sin aplicar la solución de monitoreo.	F1	NC	No se puede obtener información relacionada con las visitas a la plataforma.
	F9	C	No se puede obtener información relacionada con las búsquedas.
	F12	C	No se puede obtener información relacionada con el

			proceso de indexación.
	F17	PC	Para obtener el comportamiento del almacenamiento del servidor se tuvo que ejecutar una serie de comandos en la consola lo que hacía engorroso el procedimiento. Además, implica permanencia total delante del servidor. Para un grupo grande de servidores, no es factible.
	F18	4h	Al no contar con un sistema que permita notificar ante una incidencia como esta, se pudo detectar al realizar la comprobación del estado de los servicios.
F19	12h	Al no contar con un sistema que permita notificar ante una incidencia como esta, se pudo detectar al realizar la comprobación del estado del almacenamiento.	

Una vez analizados los resultados es posible calcular la pertinencia funcional del objetivo de uso que no es más que la proporción de las funciones requeridas por el usuario para lograr un objetivo de uso específico y se puede determinar por la función de medida (25023:2017, 2016):

Los resultados de la pertinencia funcional del objetivo de uso para este escenario se resumen en la Tabla 20 y Tabla 21, y los resultados de la pertinencia funcional del sistema se resumen en la Tabla 22 concluyendo que la solución de monitoreo para el sistema de recuperación de la Plataforma C.U.B.A mejora la pertinencia en el seguimiento de sus recursos y sistemas.

Tabla 20: Resultados de la pertinencia funcional del objetivo de uso antes de ser aplicada la solución de monitoreo.

Pertinencia funcional del objetivo de uso (X)				
Nº	Número de objetivo	Funciones faltantes o incumplidas (A)	Funciones necesarias (B)	X
1	O1	4	4	0
2	O2	7	7	0
3	O3	2	3	0
4	O4	0	3	1
5	O5	2	2	0

Tabla 21: Resultados de la pertinencia funcional del objetivo de uso después de ser aplicada la solución de monitoreo.

Pertinencia funcional del objetivo de uso (X1)				
Nº	Número de objetivo	Funciones faltantes o incumplidas (A)	Funciones necesarias (B)	X1
1	O1	0	4	1
2	O2	0	7	1
3	O3	0	3	1
4	O4	0	3	1
5	O5	0	2	1

Tabla 22: Resultados de la pertinencia funcional del sistema (A) antes de aplicar la solución de monitoreo (B) después de aplicarla.

A	B
<p>Pertinencia funcional del sistema (X2)</p> $\left(\sum_{i=1}^5 X_1/5 \right)$	<p>Pertinencia funcional del sistema (X2)</p> $\left(\sum_{i=1}^5 X_1/5 \right)$
0.2	1

3.4 - Satisfacción de los especialistas con respecto a la solución de monitoreo.

El conocimiento del estado de satisfacción de los especialistas que dan seguimiento a los recursos de hardware y aplicaciones involucrados en la Plataforma C.U.B.A son de gran importancia para conocer si la solución de monitoreo para el sistema de recuperación de información de esta plataforma contribuye en el grado de satisfacción de los especialistas con respecto al seguimiento en tiempo real. La técnica de ladov constituye una vía para el estudio del grado de satisfacción de los implicados en el proceso objeto de análisis.

Esta técnica se basa en la aplicación de un cuestionario que tiene una estructura interna determinada, que sigue una relación entre tres preguntas cerradas y un análisis posterior de otro conjunto de preguntas abiertas. La relación entre las preguntas cerradas se establece a través del denominado Cuadro Lógico de ladov; el cual posibilita determinar posteriormente el nivel de satisfacción del usuario y del grupo (Kuzumina, 1970). Para aplicar el procedimiento se debe establecer una escala de satisfacción que responde a la siguiente estructura (Paz Saavedra, 2015):

(1) Clara satisfacción, (2) Más satisfecho que insatisfecho, (3) No definida, (4) Más insatisfecho que satisfecho, (5) Clara insatisfacción y (6) Contradictoria.

Luego de aplicado el cuestionario a los especialistas en recuperación de información y administradores en redes y haber triangulado las preguntas cerradas, el número resultante de la interrelación de las tres preguntas cerradas indica la posición de cada cual en dicha escala de satisfacción. El resultado final de esta técnica es el índice de satisfacción grupal (ISG), que refleja el grado de satisfacción de los encuestados. Para ponderar el ISG se establece una escala numérica entre +1 y -1.

(+1) Máximo de satisfacción, (+0.5) Más satisfecho que insatisfecho, (0) No definido y contradictorio, (-0.5) Más insatisfecho que satisfecho y (-1) Máxima insatisfacción. El cálculo del ISG se realiza mediante la siguiente fórmula:

$$ISG = \frac{A(+1) + B(+0.5) + C(0) + D(-0.5) + E(-1)}{N}$$

Donde:

N: cantidad de usuarios encuestados

A: cantidad de usuarios con clara satisfacción

B: cantidad de usuarios más satisfechos que insatisfechos

C: cantidad de usuarios no definidos

D: cantidad de usuarios más insatisfechos que satisfechos

E: cantidad de usuarios con clara insatisfacción

La solución de monitoreo para el sistema de recuperación de información de la Plataforma C.U.B.A tiene como principales beneficiarios los usuarios que se desempeñan como administradores de redes en las instituciones cubanas por lo que se decide aplicar la prueba a este grupo de clientes utilizando el cuadro lógico de ladov (ver Tabla 31).

Se seleccionaron un total de 10 usuarios para la aplicación del cuestionario entre ellos se encuentran especialistas en recuperación de información del centro CIDI de la Universidad de las Ciencias Informáticas y los administradores de red de la Plataforma C.U.B.A. A continuación se

detallan las preguntas cerradas del cuestionario:

1. Considera usted que se deba seguir dándole seguimiento a los recursos y aplicaciones al sistema de recuperación de información de la Plataforma C.U.B.A de la forma que se realiza actualmente de manera no centralizada.
2. ¿Utilizaría la solución de monitoreo para el sistema de recuperación de información de la Plataforma C.U.B.A?
3. ¿Le satisface los componentes y tecnologías que forman parte de la solución de monitoreo para el sistema de recuperación de información de la Plataforma C.U.B.A?

Luego de calcular el índice de satisfacción grupal (ISG) se obtuvo como resultado 0,8306, pudiéndose constatar que este valor se encuentra en el intervalo de satisfacción, por lo que se puede concluir que la satisfacción de los usuarios que se benefician con la solución de almacenamiento es alta.

3.4.1 - Conclusiones del capítulo

- La aplicación del criterio de expertos, a través del método Delphi, a expertos en recuperación de información permitió corroborar la aplicabilidad y correcto nivel de influencia de la solución de monitoreo durante su aplicación en el seguimiento de los recursos y servicios de la Plataforma C.U.B.A. El análisis comparativo establecido en el seguimiento de dos servidores de indexación, uno sin aplicar la solución y otro aplicándola, permitió comprobar que la misma mejora la pertinencia en el monitoreo del sistema de recuperación de la Plataforma C.U.B.A.

Conclusiones

El desarrollo de la presente investigación permitió arribar a las siguientes conclusiones:

- El análisis realizado de cada uno de los componentes de la Plataforma C.U.B.A, entre los cuales se encuentran: Apache Solr como solución de indexación y Apache Nutch como solución de rastreo, permitieron determinar las métricas esenciales para monitorear esta plataforma.
- El estudio de los conceptos y procesos relacionados con el monitoreo permitió determinar los componentes principales para la solución tecnológica de monitoreo de la Plataforma C.U.B.A, así como las tecnologías más idóneas para esta tarea.
- La solución de monitoreo de la Plataforma C.U.B.A está compuesta por tres grupos de métricas: métricas para el seguimiento del sistema de indexación y rastreo, métricas para el seguimiento de la experiencia de usuario y métricas para el seguimiento de los recursos de hardware. De cada una de las métricas se definieron el nombre de la métrica, el punto de datos a la cual pertenece y la unidad de medición.
- La infraestructura tecnológica está basada en la plataforma de series de tiempo TICK teniendo como sistema recolector de mediciones a Telegraf, como base de datos de series de tiempo a InfluxDB, como sistema de visualización a Grafana y como sistema de alertas a Kapacitor.
- La aplicación del criterio de expertos en su variante Delphi permitió validar la solución de monitoreo de muy adecuada para ser utilizada en el seguimiento de los recursos y servicios de la Plataforma C.U.B.A.
- El análisis comparativo establecido en el seguimiento de dos servidores de indexación, uno sin aplicar la solución y otro aplicándola, permitió comprobar que la misma mejora la pertinencia en el monitoreo del sistema de recuperación de la Plataforma C.U.B.A.

Recomendaciones

- Incluir en la solución de monitoreo algún mecanismo de monitoreo activo para los elementos críticos que afecten la disponibilidad de la plataforma.
- Incorporar la medición de otras métricas asociadas a los elementos de indexación o rastreo.
- Utilizar la solución de monitoreo propuesta para la Plataforma C.U.B.A en personalizaciones basadas en el motor de búsquedas de Orion.

Referencias bibliográficas

- 25023:2017, I. (2016). INGENIERÍA DE SOFTWARE Y SISTEMAS – REQUISITOS DE LA CALIDAD Y EVALUACIÓN DE SOFTWARE (SQuaRE) – MODELOS DE LA CALIDAD DE SOFTWARE Y SISTEMAS (ISO/IEC 25010: 2011, IDT).
- Aghaei, S., Nematbakhsh, M. A., & Farsani, H. K. (2012). Evolution of the world wide web: From WEB 1.0 TO WEB 4.0. *International Journal of Web & Semantic Technology*, 3(1), 1.
- Astigarraga, E. (2006). El método delphi. *Techniques*, 1–14. <https://doi.org/10.2139/ssrn.420040>
- Baboescu, F., Tullsen, D. M., Rosu, G., & Singh, S. (2005). A tree based router search engine architecture with single port memories. *ACM SIGARCH Computer Architecture News*, 33(2), 123–133.
- Baeza-Yates, R., Ribeiro, B. de A. N., & others. (2011). *Modern information retrieval*. New York: ACM Press; Harlow, England: Addison-Wesley,.
- Berners-Lee, T. J., & Cailliau, R. (1992). World-wide web.
- Christopher, D. M., Prabhakar, R., & Hinrich, S. (2008). *Introduction to Information Retrieval*. <https://doi.org/10.1109/LPT.2009.2020494>
- Croft, W. B., Metzler, D., & Strohman, T. (2015). *Information retrieval in practice*. <http://ciir.cs.umass.edu>. Retrieved from ciir.cs.umass.edu/downloads/SEIRiP.pdf
- Deri, L., Mainardi, S., & Fusco, F. (2012). tsdb: A compressed database for time series. In *International Workshop on Traffic Monitoring and Analysis* (pp. 143–156).
- Deutsch, P., & Microsystems, S. (2013). Fallacies of distributed computing, 20, 5–7.
- Domain Name Stat, L. (n.d.). Domain name registration's statistics. Retrieved November 28, 2018, from <https://domainnamestat.com/>
- Dugan, J. M. (2014). *Time Series Database*.
- Estrada Ramos, L. M. (2012). Apache Solr , un motor de búsqueda de código abierto. *Revista Digital Universitaria*, 13(11), 1–9.
- Falcón Márquez, O. R. (2015). Desarrollada en la UCI la Plataforma CUBA | Universidad de las

Ciencias Informáticas. Retrieved March 14, 2018, from <http://www.uci.cu/desarrollada-en-la-uci-la-plataforma-cuba>

Faloutsos, C., Ranganathan, M., & Manolopoulos, Y. (1994). *Fast subsequence matching in time-series databases* (Vol. 23). ACM.

García Torres, T. (2017). Primera Jornada Red Cuba 2017 abre debate a mejorar visibilidad y posicionamiento de sitios cubanos | Universidad de las Ciencias Informáticas. Retrieved March 6, 2018, from <http://www.uci.cu/universidad/noticias/primera-jornada-red-cuba-2017-abre-debate-mejorar-visibilidad-y-posicionamiento>

Goodman, C. M. (1987). The Delphi technique: a critique. *Journal of Advanced Nursing*, 12(6), 729–734. <https://doi.org/10.1111/j.1365-2648.1987.tb01376.x>

Graphite. (2018). About of Graphite. Retrieved March 27, 2018, from <https://graphiteapp.org/#overview>

Guttman, L., & Torgerson, W. S. (1959). Theory and Methods of Scaling. *American Sociological Review*, 24(4), 600. <https://doi.org/10.2307/2089571>

Hasson, F., & Keeney, S. (2011). Enhancing rigour in the Delphi technique research. *Technological Forecasting and Social Change*, 78(9), 1695–1704. <https://doi.org/10.1016/j.techfore.2011.04.005>

Influxdata. (2018a). Telegraf from InfluxData | Agent for Collecting & Reporting Metrics & Data. Retrieved March 16, 2018, from <https://www.influxdata.com/time-series-platform/telegraf/>

Influxdata. (2018b). Time Series Database. Retrieved March 16, 2018, from <https://www.influxdata.com/time-series-database/>

InfluxData. (2018). Chronograf | Complete Interface for the InfluxData Platform. Retrieved March 16, 2018, from <https://www.influxdata.com/time-series-platform/chronograf/>

InfluxData. (2018). InfluxDB | The Time Series Database in the TICK Stack. Retrieved March 16, 2018, from <https://www.influxdata.com/time-series-platform/influxdb/>

InfluxData. (2018). Kapacitor | Real-Time Stream Processing Engine for InfluxData. Retrieved March 16, 2018, from <https://www.influxdata.com/time-series-platform/kapacitor/>

Kjkoster. (2014). Types of Monitoring. Retrieved March 14, 2018, from http://www.kjkoster.org/Blog/Types_of_Monitoring.html

- Kocjan, W., & Beltowski, P. (2016). *Learning Nagios* (Third edit). Birmingham, UK: www.packtpub.com.
- Korfhage, R. R. (1997). *Information {S}torage and {R}etrieval*.
- Kuzumina, N. V. (1970). *Metódicas investigativas de la actividad pedagógica*. Editorial Leningrado.
- Laliwala, Z. (2013). *Web Crawling and Data Mining with Apache Nutch*. Cs_Konzepte.
- Langville, A. N., & Meyer, C. D. (2011). *Google's PageRank and beyond: The science of search engine rankings*. Princeton University Press.
- Leighton, B., Cox, S. J. D., Car, N. J., Stenson, M. P., Vleeshouwer, J., & Hodge, J. (2015). A best of both worlds approach to complex, efficient, time series data delivery. In *International Symposium on Environmental Software Systems* (pp. 371–379).
- Ligus, S. (2013). *Effective Monitoring and Alerting*. (M. Oram, Andy & Hendrickson, Ed.) (2013th ed.). Gravenstein Highway North, Sebastopol. <https://doi.org/10.1016/j.jcp.2015.01.030>
- Lopez, L. A., Duerr, R., & Khalsa, S. J. S. (2015). Optimizing apache nutch for domain specific crawling at large scale. In *Proceedings - 2015 IEEE International Conference on Big Data, IEEE Big Data 2015* (pp. 1967–1971). <https://doi.org/10.1109/BigData.2015.7363976>
- Manresa, H. I. (2016). ETECSA conectará a internet más hogares en Cuba. Retrieved from <http://www.cubahora.cu/ciencia-y-tecnologia/etecsa-conectara-a-internet-mas-hogares-en-cuba>
- Maron, M. E. (1977). On indexing, retrieval and the meaning of about. *Journal of the American Society for Information Science*, 28(1), 38–43.
- Martinez Méndez, F. J. (2004). *Recuperación de información: modelos, sistemas y evaluación*.
- Massie, M., Li, B., Nicholes, B., & Vuksan, V. (2013). *Monitoring with Ganglia*. (M. Loukides & M. Blanchette, Eds.), *Vasa* (First Edit). Gravenstein Highway North, Sebastopol. Retrieved from <http://medcontent.metapress.com/index/A65RM03P4874243N.pdf>
- Medina, A., Piloto-Fleitas, N., Nogueira, D., Hernández, A., Ricardo, A., & Viteri, J. (2011). Estudio de la construcción de Índices Integrales para el apoyo al Control de Gestión Empresarial. *Revista Eidos*, 3.
- Mehta, V. (n.d.). *Icinga Network Monitoring*.
- Muller, S., Hendel, A., & Yeung, L. (1999). Search engine architecture for a high performance multi-

layer switch element. Google Patents.

Nayak, K., Nanda, K., Dwarakanath, T., Babu, H., & Selvakumar, D. (2014). Data centre monitoring and alerting system using WSN. In *IEEE CONECCT 2014 - 2014 IEEE International Conference on Electronics, Computing and Communication Technologies*.
<https://doi.org/10.1109/CONECCT.2014.6740348>

Okoli, C., & Pawlowski, S. D. (2004). The Delphi method as a research tool: an example, design considerations and applications. *Inf. Manage.*, 42(1), 15–29.

Paz Saavedra, L. E. (2015). Competencias investigativas en los docentes beneficiados por la estrategia de formación y acceso para la apropiación pedagógica de las TIC. *Tendencias*, 16(1), 175. <https://doi.org/10.22267/rtend.151601.39>

Persen, Todd and Winslow, R. (2016). Benchmarking InfluxDB vs MongoDB for Time-Series Data, Metrics and Management.

Prometheus. (2018). Overview | Prometheus. Retrieved March 27, 2018, from <https://prometheus.io/docs/introduction/overview/>

Roden, T. (2010). *Building the Realtime User Experience*.

Ryan, G. J., Ryan, S. W., Ryan, C. M., Munro, W. A., & Robinson, D. (2013). Search engine. Google Patents.

Salton, G., McGill, M., & 1983. (1983). *Introduction to modern information retrieval*. Lavoisier.Fr. Retrieved from <http://www.lavoisier.fr/notice/frGWOX3AXAXOWLOO.html>

Shahi, D. (2015). *Apache Solr: A Practical Approach to Enterprise Search*. <http://link.springer.com>.
<https://doi.org/10.1007/978-1-4842-1070-3>

Social, W. A. (2018). Digital in 2017: Global Overview - We Are Social. Retrieved from <https://wearesocial.com/special-reports/digital-in-2017-global-overview>

Statista. (2017). • Global internet hosts in the domain name system 2017 | Statistic. Retrieved February 2, 2018, from <https://www.statista.com/statistics/264473/number-of-internet-hosts-in-the-domain-name-system/>

techopedia. (2018). What is Dictionary Attack? - Definition from Techopedia. Retrieved February 5, 2018, from <https://www.techopedia.com/definition/1774/dictionary-attack>

- Techopedia. (2018). What is a Central Processing Unit (CPU)? - Definition from Techopedia. Retrieved March 19, 2018, from <https://www.techopedia.com/definition/2851/central-processing-unit-cpu>
- Techtarget. (2017). Fast Guide to RAM. Retrieved March 19, 2018, from <http://whatis.techtarget.com/reference/Fast-Guide-to-RAM>
- Turnbull, J., Lieverdink, P., Matotek, D., Cornell, G., Gennick, J., Lowman, M., ... Welsh, T. (2017). *Pro Linux System Administration*. (B. M. N. Y. Springer Science, Ed.) (2nd ed.). New York: <http://www.apress.com>. <https://doi.org/10.1007/978-1-4842-2008-5>
- Vacche, A. D., & Lee, S. K. (2015). *Mastering Zabbix*. Birmingham, UK.: www.packtpub.com.
- von der Gracht, H. A. (2012). Consensus measurement in Delphi studies. Review and implications for future quality assurance. *Technological Forecasting and Social Change*, 79(8), 1525–1536. <https://doi.org/10.1016/j.techfore.2012.04.013>
- Worth, D. (2010). *Introduction to Modern Information Retrieval, 3rd Edition*. *Australian Academic & Research Libraries* (Vol. 41). <https://doi.org/10.1080/00048623.2010.10721488>
- Worthington, R. (2016). What are the different types of monitoring and evaluation technology? Retrieved March 14, 2018, from <http://kwantu.net/blog/2016/12/26/what-are-the-different-types-of-monitoring-and-evaluation-technology>
- Wu, Q., Burges, C. J. C., Svore, K. M., & Gao, J. (2010). Adapting boosting for information retrieval measures. *Information Retrieval*, 13(3), 254–270.
- Wu, V., & Eagle, H. (2015). *Mastering VMware vSphere 5 Storage*. (P. P. Ltd, Ed.), *Zhurnal Eksperimental'noi i Teoreticheskoi Fiziki*. Birmingham B3 2PB, UK.: www.packtpub.com. <https://doi.org/10.1017/CBO9781107415324.004>

Anexos

6.1 - Anexo 1: Operacionalización de la variable dependiente

Tabla 23: Operacionalización de la variable dependiente.

ID	Nombre	Descripción	Función de medición
FAP-1-G	Pertinencia funcional del objetivo de uso.	Proporción de las funciones requeridas por el usuario para proporcionar un resultado apropiado y lograr un objetivo de uso específico.	$X = 1 - A/B$ <p>A = Número de funciones faltantes o incorrectas entre aquellas que son necesarias para lograr un objetivo de uso específico.</p> <p>B = Número de funciones necesarias para alcanzar un objetivo de uso específico.</p>
FAP-2-G	Pertinencia funcional del sistema.	Proporción de las funciones requeridas por los usuarios para lograr sus objetivos para proporcionar resultados apropiados.	$X = \sum_{i=1}^n A_i/n$ <p>A_i Puntuación de aptitud para el objetivo de uso i, es decir, el valor medido de FAP-1-G para el i-ésimo objetivo de uso específico.</p> <p>n = Número de objetivos de uso.</p>

6.2 - Anexo 2: Encuesta para determinar nivel de competencia de los expertos

Estimado(a) compañero(a).

La presente encuesta forma parte de las acciones para validar la solución de monitoreo del sistema de recuperación de información de la Plataforma C.U.B.A. Su análisis y colaboración en cuanto a los aspectos que sometemos a su consideración serán de invaluable ayuda para el desarrollo de la investigación. Le solicitamos la mayor responsabilidad y sinceridad en la realización de la encuesta. Se necesita primeramente que evalúe su conocimiento acerca de aspectos relacionados con las tecnologías de recuperación de información de la Plataforma C.U.B.A y el grado de influencia de las fuentes de argumentación, según las indicaciones que se dan a continuación. Le agradecemos de antemano por su valiosa contribución.

Datos generales del encuestado

Título universitario: _____ - _____

Categoría científica: _____ Categoría docente: _____

Años de experiencia con tecnologías de recuperación de información: _____

Instrucciones

1. Según su criterio, marque con una X en la casilla que caracteriza su nivel de conocimiento sobre los temas referidos. “0” significa total desconocimiento del tema y “10” que tiene pleno conocimiento del mismo.

1	2	3	4	5	6	7	8	9	10

2. Entre las fuentes que le han posibilitado enriquecer su conocimiento sobre los temas, se someten a consideración algunas de ellas, para que las evalúe en las categorías de: Alto (A), Medio (M) y Bajo (B), colocando una X.

Fuentes de argumentación	Grado de influencia de las fuentes		
	A	M	B
Estudios teóricos realizados por usted.			
Experiencia adquirida durante su vida profesional.			
Conocimiento de investigaciones y/o publicaciones nacionales e internacionales.			
Conocimiento propio sobre el estado del tema de investigación.			
Actualización en cursos de posgrado, diplomados, maestrías, doctorado.			
Intuición.			

Gracias por su colaboración.

6.3 - Anexo 3: Encuesta para medir el criterio de los expertos

Estimado(a) compañero(a).

Con la finalidad de someter a su consideración como experto(a) la solución de monitoreo para el sistema de recuperación de información de la Plataforma C.U.B.A, solicitamos su valoración sobre diferentes aspectos que a continuación se presentan. De antemano, le agradecemos su valiosa contribución. Para expresar su evaluación, por favor, luego de analizar cuidadosamente el material que se adjunta, evalúe cada uno de los aspectos que se le presentan en la tabla, marcando con una cruz en la casilla correspondiente y teniendo en cuenta para ello el siguiente código de categorías de clasificación.

Muy adecuado	Bastante adecuado	Adecuado	Poco adecuado	Inadecuado
MA	BA	A	PA	I

Nº	Aspecto	MA	BA	A	PA	I
1	La solución de monitoreo para el sistema de recuperación de información de la Plataforma C.U.B.A a través de sus gráficas de tendencias de las métricas asociadas a los recursos permite tomar una decisión inmediata sobre el correcto/incorrecto uso de los mismos. ¿Cómo valora este aspecto?					
2	Con las métricas recopiladas de los recursos de hardware y la visualización en el tiempo de sus valores permite planificar cuándo es necesario hacer una expansión de la infraestructura. ¿Cómo valora este aspecto?					
3	Los valores de las métricas asociadas al proceso de indexación permiten identificar en tiempo real la cantidad de documentos e imágenes indexadas por hora y detectar alguna incidencia en cuanto a este aspecto. ¿Cómo valora este aspecto?					
4	Es posible darle seguimiento en tiempo real al comportamiento de cada una de las etapas del sistema de rastreo detectando cualquier incidencia en cuanto a este aspecto. ¿Cómo valora este aspecto?					
5	A través del sistema de alertas y notificaciones es posible detectar con inmediatez cualquier incidencia en la plataforma en cuando al uso de los recursos o disponibilidad de los servicios. ¿Cómo valora este aspecto?					

6	Es posible conocer en tiempo real la experiencia de uso de la Plataforma C.U.B.A en aspectos como: cantidad de visitantes nacionales y extranjeros, cantidad de búsquedas en los módulos: web, noticias, imágenes y otros. ¿Cómo valora este aspecto?					
8	A través de gráficas es posible conocer el comportamiento y disponibilidad de cada uno de los servicios telemáticos implementados en la plataforma como los servicios web, correo electrónico, balanceadores, entre otros. ¿Cómo valora este aspecto?					
9	Es posible darle seguimiento en tiempo real y hacer estudios de tendencias de aspectos relacionados con el rendimiento del sistema de recuperación en cuanto a tiempos de respuestas de las consultas y cantidad de resultados por petición de consultas. ¿Cómo valora este aspecto?					
10	Es posible hacer estudios de tendencias de uso de las principales funcionalidades del buscador de contenidos y detectar en el tiempo tendencias de búsquedas (frases más buscadas). ¿Cómo valora este aspecto?					

Si desea exponer cualquier otra opinión, por favor, exprese en el espacio disponible a continuación:

Gracias por su colaboración.

6.4 - Anexo 4: Tablas del método Delphi

Tabla 24: Cálculo del coeficiente de conocimiento para los expertos.

Experto	Escala										Kc
	1	2	3	4	5	6	7	8	9	10	
1								x			0.8
2							x				0.7
3									x		0.9
4									x		0.9

5								x			0.8
6								x			0.8
7							x				0.7
8									x		0.9
9								x			0.8
10						x					0.6

Tabla 25: Matriz de coeficientes de argumentación por experto.

Experto	Escala						Ka
	1	2	3	4	5	6	
1	0.3	0.5	0.04	0.05	0.04	0.05	0.98
2	0.3	0.5	0.05	0.04	0.05	0.05	0.99
3	0.3	0.4	0.05	0.05	0.05	0.04	0.89
4	0.3	0.4	0.05	0.05	0.04	0.05	0.89
5	0.2	0.5	0.04	0.05	0.04	0.05	0.88
6	0.2	0.5	0.05	0.04	0.03	0.04	0.86
7	0.2	0.4	0.04	0.05	0.05	0.04	0.76
8	0.3	0.5	0.04	0.05	0.04	0.03	0.96
9	0.3	0.5	0.05	0.05	0.04	0.05	0.99
10	0.3	0.4	0.04	0.04	0.05	0.03	0.86

Tabla 26: Resultado de la autoevaluación para determinar Kcomp

Número de experto	Kc	Ka	Kcomp	Nivel de competencia
1	0.8	0.98	0.89	Alto
2	0.7	0.99	0.845	Alto
3	0.9	0.89	0.895	Alto
4	0.9	0.89	0.895	Alto
5	0.8	0.88	0.84	Alto
6	0.8	0.86	0.83	Alto
7	0.7	0.78	0.74	Medio
8	0.9	0.96	0.93	Alto
9	0.8	0.99	0.895	Alto

10	0.6	0.86	0.73	Medio
Promedio de índice de competencia			0.849	ALTO

Tabla 27: Resultados de la evaluación por los expertos a los aspectos propuestos

N.º de experto	Aspectos									
	1	2	3	4	5	6	7	8	9	10
1	MA	MA	MA	MA	MA	BA	MA	BA	MA	BA
2	BA	MA								
3	MA	MA	MA	MA	BA	BA	MA	MA	MA	MA
4	MA	MA	MA	MA	BA	MA	MA	MA	MA	A
5	MA	MA	BA	MA	BA	MA	MA	BA	MA	BA
6	MA	MA	MA	BA	MA	MA	BA	MA	MA	A
7	BA	MA	BA	BA						
8	MA	MA	BA	MA	BA	MA	MA	MA	MA	MA
9	BA	MA	MA	BA	MA	MA	BA	MA	BA	A
10	MA	MA	MA	MA	MA	MA	MA	MA	A	BA

Tabla 28: Frecuencia absoluta por aspectos

N.º de aspecto	MA	BA	A	PA	I	Total
1	7	3	0	0	0	10
2	10	0	0	0	0	10
3	8	2	0	0	0	10
4	8	2	0	0	0	10
5	6	4	0	0	0	10
6	8	2	0	0	0	10
7	8	2	0	0	0	10
8	8	2	0	0	0	10
9	7	2	1	0	0	10
10	3	4	3	0	0	10

Tabla 29: Frecuencias relativas acumulativas.

N.º de aspecto	MA	BA	A	PA	I
1	7	10	10	10	10
2	10	10	10	10	10
3	8	10	10	10	10
4	8	10	10	10	10
5	6	10	10	10	10
6	8	10	10	10	10
7	8	10	10	10	10
8	8	10	10	10	10
9	7	9	10	10	10
10	3	7	10	10	10

Tabla 30: Frecuencias relativas de las evaluaciones por aspecto.

N.º de aspecto	MA	BA	A	PA	I
1	0.700	1.000	1.000	1.000	1.000
2	1	1.000	1.000	1.000	1.000
3	0.800	1.000	1.000	1.000	1.000
4	0.800	1.000	1.000	1.000	1.000
5	0.600	1.000	1.000	1.000	1.000
6	0.800	1.000	1.000	1.000	1.000
7	0.800	1.000	1.000	1.000	1.000
8	0.800	1.000	1.000	1.000	1.000
9	0.700	0.200	0.100	1.000	1.000
10	0.300	0.400	0.300	1.000	1.000

Tabla 31: Cuadro Lógico de ladov y preguntas abiertas para los especialistas de recuperación de información y administradores de redes. Fuente: elaboración propia.

¿Le satisface los componentes y tecnologías que forman parte de la solución de monitoreo para el sistema de recuperación de	Considera usted que se deba seguir dándole seguimiento a los recursos y aplicaciones al sistema de recuperación de información de la Plataforma C.U.B.A de la forma que se realiza actualmente de manera no centralizada.
-----------------------------------------------------------------------------------------------------------------------------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

información de la Plataforma C.U.B.A?	Si			No sé			No		
	¿Utilizaría la solución de monitoreo para el sistema de recuperación de información de la Plataforma C.U.B.A?								
	Sí	No sé	No	Sí	No sé	No	Sí	No sé	No
Me gusta mucho	1	2	6	2	2	6	6	6	6
Me gusta más de lo que me disgusta	2	2	3	2	3	3	6	3	6
Me da lo mismo	3	3	3	3	3	3	3	3	3
Me disgusta más de lo que me gusta	6	3	6	3	4	4	3	4	4
No me gusta nada	6	6	6	4	4	6	4	5	6
No sé qué decir	2	3	6	3	3	3	6	3	4
Preguntas abiertas:									
1. ¿Qué elementos consideras positivos en la solución de monitoreo para el sistema de recuperación de información de la Plataforma C.U.B.A?									
2. ¿Qué elementos consideras negativos en la solución de monitoreo para el sistema de recuperación de información de la Plataforma C.U.B.A?									
3. ¿Propondrías añadir alguna tecnología o componente a la solución de monitoreo para el sistema de recuperación de información de la Plataforma C.U.B.A?									