

BASES DE DATOS PARA EL MANEJO DE DOCUMENTACIÓN EN ENTIDADES ESTATALES: BASES DE DATOS DOCUMENTALES

Sandra Nayibe Ospina Bernal

RESUMEN

Este documento pretende mostrar el impacto que se genera al momento de que se pueden manipular grandes cantidades de datos para el manejo de documentación e información en las entidades estatales, con el fin de exponer distintas técnicas encaminadas a lograr una efectiva organización teniendo en cuenta las 3 V que se manejan en Big Data como son volumen, velocidad y variedad en la recolección de la información, esto se logra a partir de una base de datos no relacional de código abierto como lo es NoSQL, esto lo logra a partir del estudio de la evolución, arquitectura, características y beneficios de cada una las bases de datos desde la exigencia que genera las entidades estatales por manipular grandes volúmenes de datos, ya que existe un crecimiento exponencial de datos y para dichas entidades es necesario y vital que su información sea administrada de tal manera que alteren los protocolos de seguridad y los atributos de calidad como consistencia, veracidad, confiabilidad, seguridad, replicación, factores como la transformación tecnológica y la expansión comercial que se deben soportar con herramientas que se adapten a la necesidad de las entidades.

Palabras Clave: *Bases De Datos Documentales, Base de Datos NoSQL, Minería de Datos, Protección de los Datos, Documentos*

INTRODUCCIÓN

El objetivo de este documento es proyectarlo a las entidades estatales que se encuentren interesadas en implementar en cada una el manejo de bases de datos en las cuales puedan gestionar cada dato que es ingresado al momento de que se genere alguna solicitud ante la entidad.

El tratamiento de la información es un factor importante en cada entidad, teniendo en cuenta que en Colombia existen entidades que procuran que su información sea manejada de forma planificada y controlada, ya que estas continúan en un proceso de normalización de la información y un manejo total de la misma, pero en la actualidad existen tratamientos errados de la información, lo que dificulta que se ejerza un control, por esta razón se debe concientizar a los profesionales que trabajan en el manejo de la información de base de datos documentales sobre el fácil manejo de la información teniendo en cuenta que aplicando el big data, este mejorará el flujo de la misma.

Las bases de datos , teniendo en cuenta su aplicabilidad, tiene la tendencia a mejorar el tratamiento del gran mundo de los datos, por lo cual a lo largo de la historia se han desarrollado algunas herramientas que provee soluciones eficaces y efectivas.

Actualmente, estos programas a los que hacemos referencia tienen comercializadas aplicaciones informáticas (denominadas vulgarmente pasarelas web) que permiten la consulta, desde un navegador web, de las bases de datos creadas con ellos. Esto permite ampliar notablemente el espectro de usuarios potenciales de las bases de datos ya que no es necesario utilizar redes de área local para compartir el uso de las bases de datos ni, mucho menos, desplazarse a la ubicación física donde éstas residen.

De manera personal e implícita dentro del presente documento se procura ejercer una exploración para resolver los cuestionamientos ¿Son las bases de datos NoSQL una evolución de las bases de datos documentales? ¿Las industrias colombianas están preparada para dar el salto de bases de datos documentales? Se perseguirá la respuesta a estas preguntas y comparativas con base en el conocimiento adquirido en el curso de Big Data.

DESARROLLO

Desde sus orígenes las bases de datos se remontan a la antigüedad donde se apreciaba que la información se sostiene por medio de registros que fueran guardados de manera cuidadosa, por esta razón al momento de la búsqueda esta se volvía lenta y poco eficaz y no existía una ayuda de las máquinas por lo tanto esto se realizaba de manera manual.

En 1884 se creó la máquina automática de tarjeta perforada, es decir que para esta época se realizaban los censos de forma manual. Pero en la década de los sesenta se dio inicio a las bases de datos jerárquicos y de redes, por lo cual ya era posible guardar estructuras de datos de listas y de árboles.

Ya en la década de los setenta un científico se dio a conocer por sus teorías de bases de datos relacionales, por esto después de tantos ires y venires se evidencia los sistemas de bases de datos relacionales, desarrolló el Relational Software System, o lo que es lo mismo, lo que actualmente se conoce como Oracle Corporation, desarrollando así un sistema de gestión de bases de datos relacional con el mismo nombre que dicha compañía.



Ilustración 1. Historia de las Bases de Datos

Por su parte en la década de los 80 se empieza hablar de unos temas de bases de datos en los cuales se comienza a escuchar el término SQL y la industria de los datos se muestran que se utilizan las columnas y la filas, así ya se puede concursar con otros motores de bases de red y/o jerárquicas. lo cual su programación se mostraba de forma sencilla y de un nivel bajo.

La necesidad de ingresar los datos tomo un giro inexplicable, el cual Microsoft office entregó dos herramientas (Access y Excel), el cual hizo que las bases de datos tomará un giro a nivel de la datos orientados a objetos. Cabe resaltar que esta etapa surge la tercera generación de gestores de bases de datos. Aunque para esta época se da el nacimiento de World Wide Web, lo cual facilita la consulta a la base de datos.

En la actualidad existen un sin número de alternativas que nos permite hacer búsquedas orientadas a las necesidades que presentan los usuarios una de las tendencias más amplias son las bases de datos que cumplan con el protocolo Open Archives Initiative – Protocol for Metadata Harvesting (OAI-PMH) los cuales permiten el almacenamiento de gran cantidad de artículos que permiten una mayor visibilidad y acceso en el ámbito científico y general.

BASES DE DATOS NO RELACIONALES NoSQL

Las bases de datos relacionales no presentan negativas al respecto pero al transcurrir de los años, se ha logrado aprender técnicas que son comunes para normalizarlas lo mejor posible, escalarlas según crece la demanda, y utilizarlas como sistema de persistencia para almacenar información desde un lenguaje u orientado a objetos (entre otros). La cuota de uso de software como SQLite, MySQL, PostgreSQL u Oracle, por poner cuatro ejemplos, es muy alta, encontrándose en la mayor parte de los desarrollos modernos.

Hablar de NoSQL o de bases de datos no relacionales abarca una amplitud a nivel de tecnología y arquitecturas, el cual busca resolver ciertos problemas que se presentan a nivel de escalabilidad y rendimiento en un concepto que ha tenido una gran aceptación como lo es Big Data, es decir que NoSQL es útil para las empresas que necesitan acceder y analizar grandes cantidades de datos que no son estructurado, pero llega la web, el software como servicio y los servicios en la nube y las startups de éxito en millones de usuarios y a su vez con todo ello llegaron los problemas de alta escalabilidad. Si bien los modelos relacionales se pueden adaptar para hacerlos escalar incluso en los entornos alta dificultad, sí que es cierto que a menudo, se hacen cada vez menos sagaz a medida que aumenta la complejidad.

En este tiempo es extraño que un ingeniero no conozca el término de NoSQL, por lo tanto los desarrolladores han hecho que está bases de datos no relacionales se empiece a trabajar con más fuerza, ya que muchos sistemas se basan en este concepto. Igualmente se debe aclarar que algunos sistemas NoSQL no son en su totalidad no relacionales y otros evitan que existan funciones relacionales que muestran tablas fijas y operaciones conjuntas, Por ejemplo, en lugar de utilizar tablas, una base de datos NoSQL podría organizar los datos en objetos, pares clave/valor o tuplas.

Los sistemas NoSQL intentan combatir este problema proponiendo una estructura de almacenamiento más variable, aunque se corra el riesgo de arriesgar ciertas funcionalidades como las transacciones que engloban operaciones en más de una colección de datos, o el no poder ejecutar el producto de dos tablas, también llamado JOIN, con esto se tiene que recurrir a la desnormalización de datos.

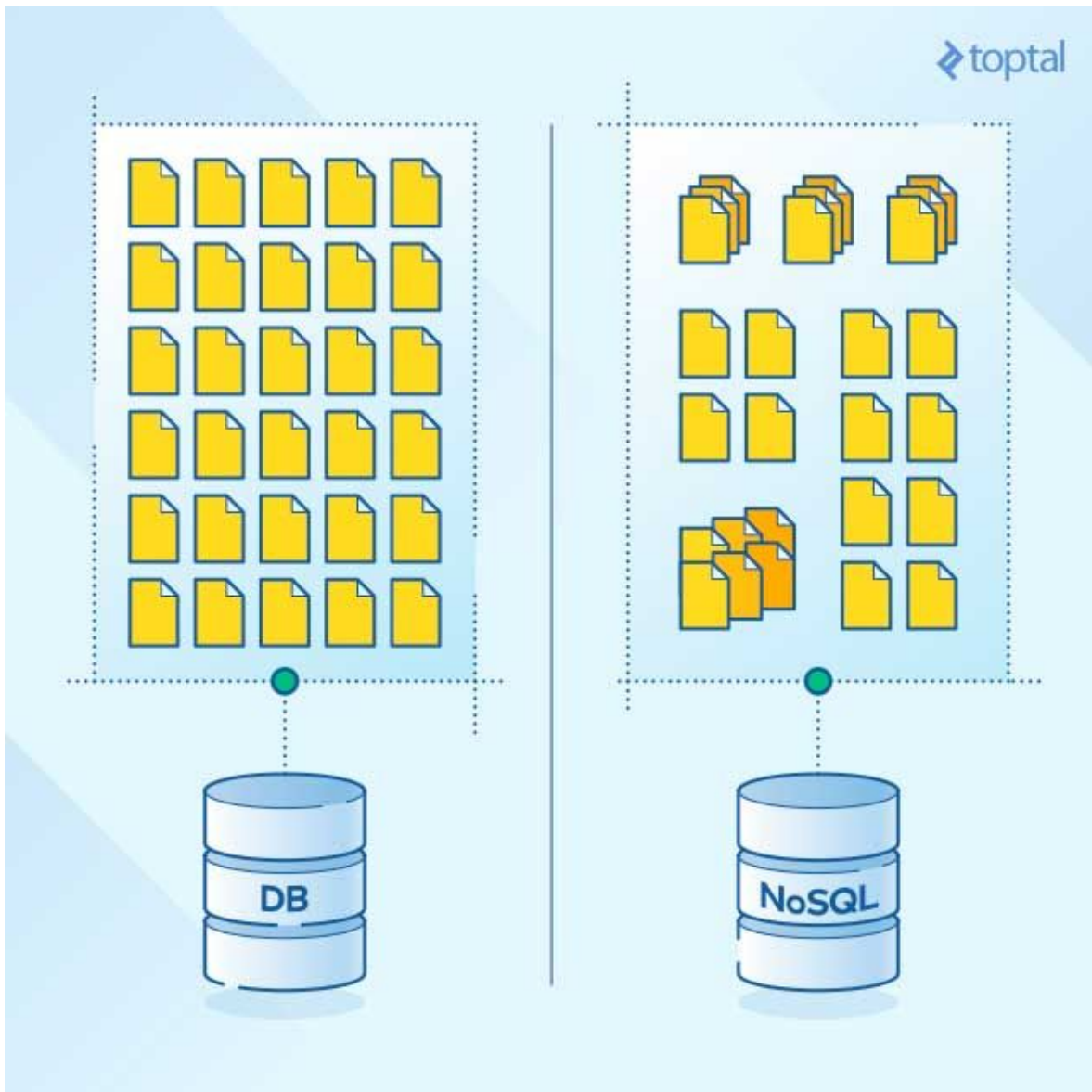


Ilustración 2. Diferencia entre una base de datos relacionales y NoSQL.

VENTAJAS Y DESVENTAJAS DE LAS BASES DE DATOS NOSQL

VENTAJAS

- A diferencia de las bases de datos relacionales, las bases de datos NoSQL están basadas en key-value pairs
- Algunos tipos de almacén de bases de datos NoSQL incluyen diferentes tipos de almacenes como por ejemplo el almacén de columnas, de

documentos, de key value store, de gráficos, de objetos, de XML y otros modos de almacén de datos.

- Algunos tipos de almacén de bases de datos NoSQL incluyen almacenes de columnas, de documentos, de valores de claves, de gráficos, de objetos, de XML y otros modos de almacén de datos.

DESVENTAJAS

- La mayoría de las bases de datos NoSQL no admiten funciones de fiabilidad, que son soportadas por sistemas de bases de datos relacionales. Estas características de fiabilidad pueden resumirse en: “atomicidad, consistencia, aislamiento y durabilidad.” Esto también significa que las bases de datos NoSQL, que no soportan esas características, ofrecen consistencia para el rendimiento y la escalabilidad.
- Con el fin de apoyar las características de fiabilidad y coherencia, los desarrolladores deben implementar su propio código, lo que agrega más complejidad al sistema.
- Esto podría limitar el número de aplicaciones en las que podemos confiar para realizar transacciones seguras y confiables, como por ejemplo los sistemas bancarios.

Feature	NoSQL Databases	Relational Databases
Performance	High	Low
Reliability	Poor	Good
Availability	Good	Good
Consistency	Poor	Good
Data Storage	Optimized for huge data	Medium sized to large
Scalability	High	High (but more expensive)

Ilustración 3. Cuadro comparativo entre las funcionalidades entre NoSQL y Bases Relacionales

En ese orden de ideas, se puede decir que la base de datos más popular es Apache Cassandra.

BASES DE DATOS DOCUMENTALES

A diferencia de las bases de datos relacionales tradicionales, el esquema para cada documento no relacional (NoSQL) puede variar, lo cual ofrece a los desarrolladores, administradores de bases de datos y profesionales de TI más flexibilidad en la organización y almacenamiento de datos de aplicaciones, así como una reducción del almacenamiento requerido para valores opcionales. Las bases de datos documentales son una forma moderna de almacenar datos en formato JSON en lugar de las simples filas y columnas de las bases de datos relacionales. Esto permite expresar los datos en su forma natural.

Por lo tanto, se puede definir como conjunto de información que se muestra de manera estructurada en registro y se almacena en soportes electrónicos legibles desde el computador. Es de aclarar que cada registro se compone de campos donde se alojan los datos. Cada uno de los registros tiene una unidad independiente de información en la cual puede ser estructurada en diferentes campos que se encuentran en la base de datos.

El Sistema de Gestión Documental (SGD) son un tipo de programas muy conocidos en el ámbito de la información y documentación, ya que están especialmente pensados para la gestión de información textual y de documentos cognitivos. Sus principales características se pueden sintetizar en disponer de un modelo de registro flexible (campos de longitud variable, campos multivalor, etc.), facilitan el acceso a los registros a través del fichero inverso, contienen un conjunto de variadas prestaciones de recuperación de la información, y están dotados de diversos instrumentos para el control terminológico. Algunos de los sistemas más conocidos y extendidos son CDS/ISIS, FileMaker, Knosys, e Inmagic DB/Text.

Según Rodríguez Yunta en su libro “Introducción a las bases de datos documentales” y a su vez teniendo en cuenta su definición, se encuentran tres bases de datos de contenidos de registros:

- **Bases de datos con información factual** (Información concreta y actual de carácter numerico)
- **Directorios** (Recolección de datos de personas o instituciones)
- **Bases de datos documentales** (Cada registro corresponde a un documento)

Para esto se debe crear una base de datos que pueda cubrir las necesidades de la empresa estatal así como lo requiera. La idea principal de consulta de esta información debe ser de modo electrónico. Por lo tanto los registros que existen en las bases de datos documentales que pueden incluir o no contenido y por esto se divide en tres categorías:

Bases de Datos de Texto Completo: Son las que contienen sus propios documentos digitalizados, por un volcado completo de su texto.

Archivos Electrónicos con Imágenes: Son aquellos que por medio de las referencias permite llegar a los enlaces de la imagen, con el fin de llegar al documento original.

Bases de Documentos Referenciales: Son aquellos que no contienen la información original sino la fundamental, la cual define en pocas palabras pero permite su localización de forma efectiva.

Ya contando con estas tres categorías, en la actualidad existen programas que tienen comercializadas aplicaciones informáticas (denominadas pasarelas web) que permiten la consulta, desde un navegador web, de las bases de datos creadas con ellos. Esto permite ampliar notoriamente el espectro de usuarios potenciales de las bases de datos ya que no es necesario utilizar redes de área

local para compartir el uso de las bases de datos y tampoco desplazarse a la ubicación física donde éstas laboran.

Para realizar la publicación de la base de datos, es necesario llevar un proceso generalizado, con esto los usuarios pueden ingresar por medio de un navegador a la información y así consultarla y en algunos casos actualizarla, en este caso se daría como ejemplo los servicios en línea prestados por las entidades del estado.



Ilustración 3. Elementos básicos en los cuales intervienen en el proceso y funcionamiento.

La pasarela web como bien se conoce es un conjunto de páginas web o aplicaciones que son capaces de leer e interpretar las órdenes que se le transmiten desde un formulario html, lo cual alguna de ella son enviadas por el usuario y las otras son ejecutadas programadamente.

La interfaz el mecanismo o herramienta que posibilita esta comunicación mediante la representación de un conjunto de objetos, iconos y elementos

gráficos que vienen a funcionar como metáforas o símbolos de las acciones o tareas que el usuario puede realizar en la computadora. Por ejemplo, introducir datos en una hoja en blanco, enviar un documento a la papelera, modificar un archivo, cerrar un programa.

Debemos tener en cuenta que se han establecido ciertas tipologías para caracterizar los diferentes modelos de las bases de datos documentales tales como según el organismo productor, modo de acceso, cobertura temática y destinatario, cobertura documental y modelo de tratamiento documental. Vale aclarar que las tres primeras se pueden aplicar a cualquier base de datos y cada una de ellas maneja especificaciones que describen cada uno de los procesos que interactúan en los procedimientos de las bases de datos.

La estructura interna de una base de datos documentales tiene ciertas características que describiremos mediante un ejemplo de una base de datos documentales que se puede manejar en una entidad. Debemos empezar mostrando la vista funcional que deberá incorporar nuestro software.

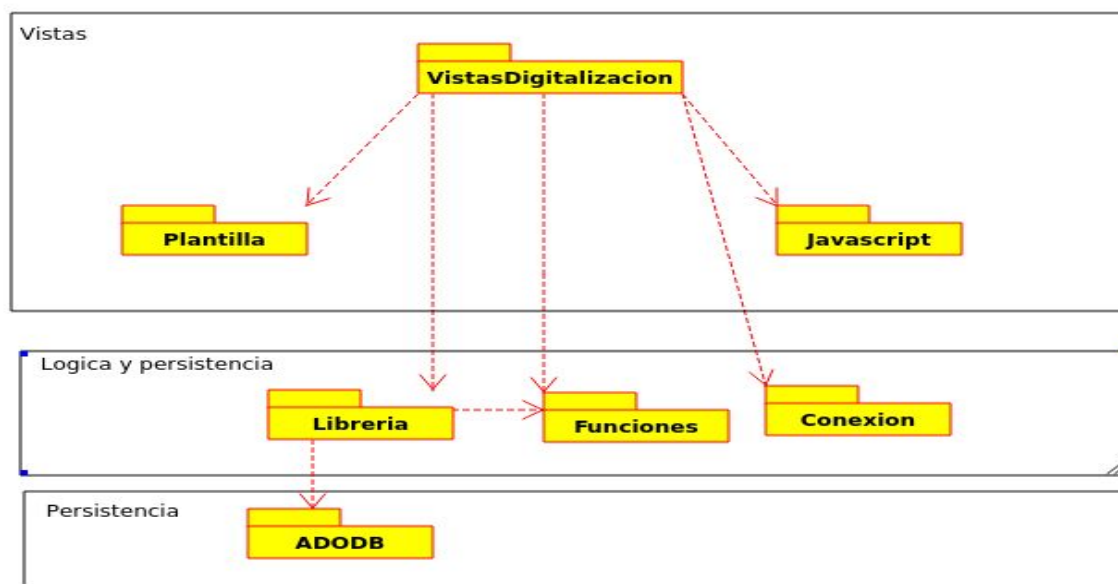


Ilustración 4. Vista funcional que debe tener un software que maneja bases de datos documentales.

En este diagrama de paquetes se representa el conjunto de directorios que se encuentran para la aplicación y su correspondiente relación. En el modelo se encuentran tres capas que representan el tipo de paquetes que lo conforman. El conjunto de vistas son donde se encuentran los programas que tienen el javascript y html es decir la capa de presentación de la aplicación. Lógica y persistencia tiene todos los programas que realizan consultas o transacciones a la base de datos y además tiene los algoritmos relacionados a la lógica del negocio. Por último se encuentra la capa de persistencia que denota la relación que hay sobre la librería ADODB, todo programa que realice una transacción a las base de datos debe utilizar una instancia de esta librería.

Las instituciones ha venido manejando un procesos donde se digitalizan todos los trámites que ingresan a las entidades El mismo día que ingresan deben ser digitalizados en su totalidad. El producto final es el conjunto de documentos publicados en la consulta de trámites . El formato de estos documentos es TIF y se encuentra un archivo por cada página del documento. Esto genera inconvenientes a la hora en que los usuarios , sin conocimiento técnico suficiente tengan que ingeniárselas para poder visualizarlos instalando complementos adicionales en el navegador y navegar por el documento se hace dispendioso.

Para el uso interno del documento, usuarios de las distintas dependencias en algunos casos , requieren el texto contenido de los documentos y no hay otra alternativa que transcribirlo. Esto acarrea demoras y que los procesos se vean afectados por tareas operativa.

Dentro del proceso interno de digitalización se encuentran ciertas problemáticas. Los administradores funcionales tienen dificultades a la hora de hacer seguimiento y control de los documentos y operarios a su cargo. Las alertas sobre los errores de indexación son confusas y en algunos casos interrumpen las asignaciones lo que conlleva al retrabajo y pérdida de tiempo.

En la descripción del sistema el software a desarrollar permitirá apoyar el proceso de digitalización de los documentos de radicación . Soporta las tareas de escaneo ,separación e indexación, control de calidad y correcciones. Registrará desde las actividades de escaneo los usuarios responsables de los cambios que tendrá el documento hasta su publicación final. Permitirá realizar la indexación ordenadamente con las validaciones pertinentes sin incurrir en demoras adicionales. El control de calidad se podrá realizar integradamente dentro de la aplicación y no como en este momento que lo realizan mediante archivos planos en formato en Excel. Las correcciones serán más controladas mediante estados para registrar controles y perfilar permisos para tener seguridad sobre los cambios que se realicen al documento después de publicado.

Para los requerimientos funcionales, de acuerdo a los requerimientos se representa a un alto nivel el diagrama de casos de uso. Se tienen los actores, escaneador ,indexador, calidad, Administrador funcional, Administrador técnico. El escaneador es el encargado de definir las imágenes escaneadas y de crear un lote de documentos que será asociado y consultado por la aplicación.El usuario indexador será el encargado de separar los documentos y de asignar el número del radicado a los cada documento separado. Además puede asumir las labores del usuario escaneado. El usuario de calidad es encargado de realizar control de calidad después de la indexación y registrar correcciones en caso de que se presenten. Administrador funcional: Será encargado de monitorear el proceso de digitalización con ayuda de reportes consolidados. Administrador técnico: Encargado de realizar soporte sobre las cuentas de usuarios. Creando nuevos usuarios y administrando claves. Asignar los grupos de usuarios.

En la vista de procesos El proceso de digitalización es en el que los documentó provenientes de la radicación de un trámite son escaneados y almacenados con el fin de que sean consultados por usuarios internos y externos. El proceso contiene las siguientes actividades:

Alistamiento: Al ser radicados los documentos físicos son asegurados con ganchos , a medida que van llegando van siendo acumulados en el puesto de cada radicador. El radicador es la persona encargada de registrar el trámite y recibir el documento relacionado.

Cada hora hay un recorrido en que un operador va recogiendo los documentos de los radicadores y los lleva al área de alistamiento. Personal encargado retira los ganchos y legajadores y deja los hojas sueltas acumulando los documentos uno encima del otro. De esta manera se crea un aproximado 400 hojas que formarán el lote para ser escaneado.

Escaneo: Se reciben los lotes de hojas y son escaneados. Esta tarea se realiza con ayuda de un software de escritorio dedicado especialmente al escaneo y que es lo más compatible posible a la máquina escáner. En el software se crea un nuevo lote, se ejecuta el inicio de escaneo y el escáner recibe las hojas, hasta que no se complete un aproximado de 400 imágenes no se interrumpa la ejecución. Al momento que se finalizan las 400 imágenes aproximadas , se para el proceso y se crea una nueva carpeta, con el conjunto de imágenes. Esta carpeta se encuentra en ruta correspondiente a la configurada para compartir con la aplicación. Para cada carpeta creada con imágenes, el operador debe ingresar a la aplicación y transferirlas por medio de la opción de escaneo. De esta manera se crea un nuevo lote de imágenes lista para indexar.

Control De Calidad 1: Se verificará inicialmente la calidad de las imágenes. En esta actividad se corrigen problemas de rotación y claridad de la imagen. Esta revisión no es tan minuciosa por razones de tiempo.

Indexación: En esta actividad se realiza la separación de documentos identificando el límite de cada documento ya que se encontraran todas las imágenes de distintos documentos en conjunto. Después de separados el usuario encargado ingresa a la opción indexación , visualiza los documentos separados ingresando a uno por uno para con ayuda de un formulario asigna el

número de radicado correspondiente. Los documentos serán almacenados en un repositorio de forma segura y convertidos a formato pdf inicial. En el siguiente modelo se describe con más detalle el proceso de indexación.

Organización y Distribución: Los documentos físicos deben ser de nuevo legados y entregados a las dependencias correspondientes. Esto se realiza en nuevo recorrido que reorganiza los documentos y los empacan en válidas destinados a las dependencias correspondientes.

Control de calidad 2: En esta actividad se realiza una revisión más minuciosa de cada documento ya indexado. En caso que se requiera una corrección el documento se devolverá y es muy posible que se requiera alistar y escanear de nuevo.

Dentro del proceso de indexación hay unas actividades dedicadas a la conversión para lograr un PDF final con texto seleccionable. Inicialmente el PDF será con solo imágenes y luego es enviado a una carpeta común donde una herramienta externa lo procesa y lo devuelve a una carpeta de salida de documentos. El PDF resultante es recogido por un programa ejecutado desde una tarea programada. El programa que extrae el documento reemplaza el PDF inicial en la ruta de almacenamiento definitivo.

En el punto de vista del despliegue, en este modelo de alto nivel se muestran los nodos que intervienen para toda la aplicación. Encontramos como en el modelo anterior El cliente de usuario y de escáner y el nodo de la aplicación principal. Adicionalmente tenemos el nodo donde se encuentra la suite de herramientas ABBYY que realizan el reconocimiento de texto de los PDF iniciales. Se pueden encontrar las carpetas de entrada y salida, en donde se deja el PDF inicial y en donde se extrae el PDF convertido correspondientemente. Se encuentran la relación al nodo de la base de datos y el nodo de Área de almacenamiento SAN.

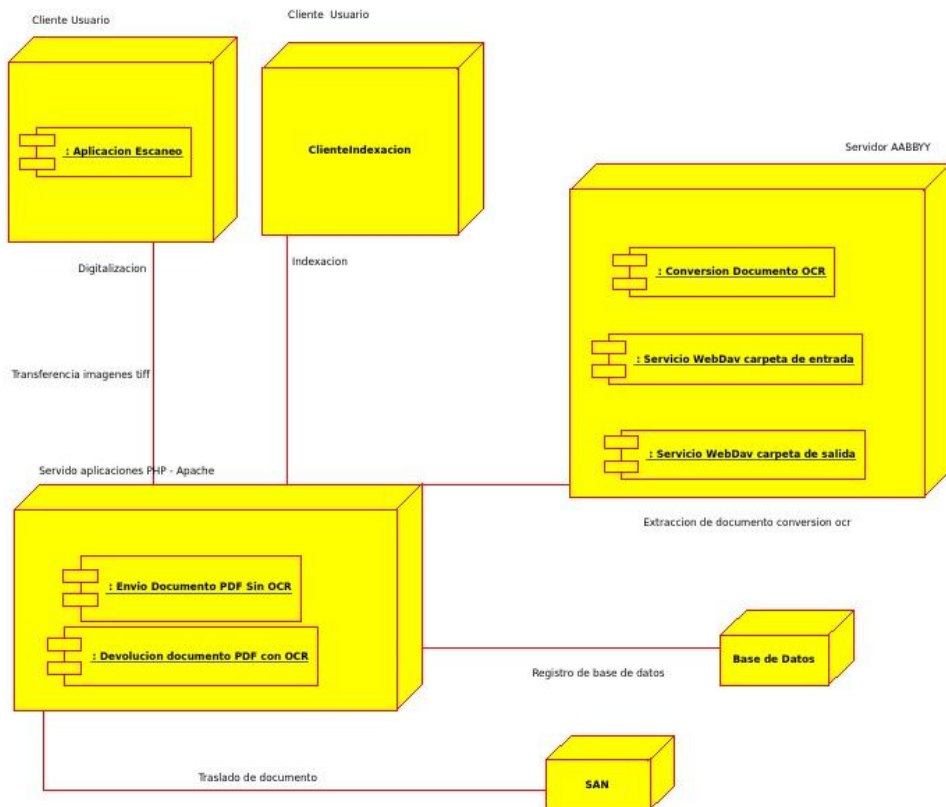


Ilustración 5. Vista de Despliegue

También se encontrará un diccionario de datos donde se evidenciara cuantos registros y que tipo de campos se tendrá en la base de datos.

Proyecto	Sistema de Gestión Documental
Nombre Tabla	cles_digi

Descripción	Comentarios de la tabla: Registrar información de la conexión y configuración del escaneo
Autor	Ingeniero Desarrollador
Fecha	
Versión	1.0
Número de Registros	30

Columna	Tipo	Nulo	Predeterminado	Comentarios
<u>idcl_esdi</u>	int(11)	No		Llave autonumerica de la tabla
ipcl_esdi	varchar(45)	No		dirección ip del cliente conectado al escáner
drcl_esdi	varchar(45)	No		Referencia del driver que identifica al escáner en el cliente
nomb_cles	varchar(45)	No		Identificador de la conexión escáner
uscl_esdi	varchar(45)	Sí	<i>NULL</i>	usuario del sistema operativo en el cliente conectado al escáner
opes_esdi	varchar(45)	Sí	<i>NULL</i>	Opciones adicionales que se ejecutan al

				momento de escanear
bacl_esdi	smallint(6)	Sí	NULL	Booleano que identifica si el proceso de escaneo se ejecuta en batch para varias imágenes por ejecución
ides_cles	varchar(3)	No	100	Estado para identificar si la conexión del escáner se encuentra activa
ruma_cles	varchar(100)	Sí	NULL	Ruta en el servidor para almacenar temporalmente las imágenes en caso que no funcione la ejecución remota del escaneo
rure_cles	varchar(100)	Sí	NULL	Ruta en cliente conectado al escáner donde se almacenan temporalmente las imágenes
rulo_cles	varchar(100)	Sí	NULL	Ruta en servidor conectado al escáner donde se almacenan temporalmente las imágenes
decl_esdi	varchar(255)	Sí	NULL	Descripción de última conexión en el cliente
feco_cles	varchar(45)	Sí	NULL	Fecha de última conexión en el cliente

En el servidor es necesario contar con los siguientes componentes para configurar el entorno y los datos se puedan almacenar.

Componente	Descripción	Enlace documentación de instalación
Servidor de aplicaciones	PHP 5.1.x o 5.2.x y puede ser adaptable a php 4	http://www.ibm.com/developerworks/data/library/techarticle/dm-0606bombardier/
Servidor Web	Apache 2.*, ApacheTomcat 6.x para reportes en BIRT	http://www.ibm.com/developerworks/data/library/techarticle/dm-0606bombardier/
Servidor de base de datos	Informix	http://www.ibm.com/developerworks/data/library/techarticle/dm-0606bombardier/
Sistema Operativo	Linux Red Hat 5.x	
Almacenamiento	>= 1GB para los documentos fuente	
Servicios o aplicaciones para reportes	Apache Tomcat 6.x , BIRT aplicación web para publicación de reportes	http://www.eclipse.org/birt/phenix/build/
Ruta en producción	http://10.20.5.212/GestionDocumentalProd2/Digitalizacion/Administracion/login.php	

Ruta en desarrollo	http://10.20.4.42/jlopezm/GestionDocumental/Digitalizacion/Ingreso.php	
Suite ABBYY para conversión OCR		

Se debe instalar las siguientes herramientas para el tratamiento de imágenes en el servidor principal.

Herramienta	Descripción
pdftk	Se utiliza para manipular archivos pdf
ImageMagick	Es una suite de herramientas que sirven para editar imágenes de muchos formatos entre los cuales es soportado el formato TIFF
libtiff	Librería de herramientas para manipular imágenes formato tiff
Poppler-tools	Librería de herramientas que apoyan la manipulación de pdf dentro de las cuales son necesarias pdftimages y xpdf

El componente para escaneo requiere la instalación de una máquina virtual de VirtualBox, que debe ser instalada y configurada en la máquina cliente conectada

al escáner. En la entrega de toda la documentación y los documentos fuente se encuentra una copia de la máquina virtual. Para instalarla se deben seguir las instrucciones del documento adjunto [Instalacion Maquina Virtual Cliente .docx]

CONCLUSIONES

Las bases de datos forman el núcleo de las principales aplicaciones, sitio web y servicios corporativos. En definitiva, el término NoSQL nos referimos a una multitud de bases de datos que intentan solventar las limitaciones que el modelo relacional se encuentra en entornos de almacenamiento masivo de datos, y concretamente en las que tiene en el momento de escalar, donde es necesario disponer de servidores muy potentes y de balanceo de carga.

Este documento tiene las especificaciones técnicas de instalación inicial así como también las referencias a las tecnologías aplicadas con el fin de tener el conocimiento necesario para dar el soporte requerido de último nivel. No se encontrará en este documento las referencias o el detalle del conocimiento necesario para comprender tecnologías que pueden afectar el funcionamiento de la aplicación pero que son externas a la misma, como pueden ser explicaciones del mantenimiento de la red o el mantenimiento de los servidores.

Mejorar procedimiento de digitalización de documentos asociados a radicados que ingresan a las entidades estatales. Hacer más accesible el documento digitalizado a los usuarios involucrados en los trámites de los radicados. Permitir que los textos de los documentos digitalizados pueden ser buscados por el detalle de su contenido.

Por lo tanto una base de datos conlleva la existencia de tres tipos de usuarios que son:

- Diseño el cual administra los datos
- Desarrollador implementa las transacciones e interfaces.
- Usuarios finales los cuales consultan y editan los datos

Los documentos digitalizados no son adecuadamente accesibles a los interesados en consultar y tramitar los radicados asociados. Deben desplegar pagina por pagina el documento y no les es posible seleccionar texto. Es necesario instalar un driver especial en el explorador para poder visualizarlo , lo cual a no todas las personas es fácil de asimilar.

BIBLIOGRAFÍA

[1] RODRÍGUEZ YUNTA, Luis. Bases de datos documentales: estructura y uso. En: MALDONADO, Ángeles (coord.). La información especializada en Internet. Madrid: CINDOC, 2001

[2] JORDI SERRA (2008). Los Documentos Electrónicos. Que Son Como Se Tratan Home Page, [online]. Disponible: <https://goo.gl/Wyzeuz>

[3] I. ANTONIO MUÑOZ CAÑAVATE. Sistemas de información en las empresas [en línea]. “Hipertext.net”, núm. 1, 2003. / <http://www.hipertext.net/web/pag251.htm>

[4] Oracle (2017, 11 Julio). Modelo relacional [online]. Disponible: <https://docs.oracle.com/database/122/CNCPT/introduction-to-oracle-database.htm#CNCPT 958>

[5] Big Data. La Revolución de los datos masivos. Viktor Mayer-Schönberger, Kenneth Cukier. Turner Publicaciones, 2013.

[6] Mastering Apache Cassandra – Second Edition. Nishant Neeraj. Pack Publishing. 2015.

[7] Historia de las Bases de Datos (2011, 4 Enero). Bases de Datos, Codd, Hollerith[Online].Disponible:<http://histinf.blogs.upv.es/2011/01/04/historia-de-las-bases-de-datos/>

[8] NoSQL vs SQL: Principales diferencias y cuándo elegir cada una de ellas (2015, 18 Noviembre). Bases de datos [online]. Disponible: <https://blog.pandorafms.org/es/nosql-vs-sql-diferencias-y-cuando-elegir-cada-una/>

[9] Bases de datos NoSQL : Guía definitiva (2017, 20 Abril). Bases de datos [online]. Disponible: <https://blog.pandorafms.org/es/bases-de-datos-nosql/>

[10] Abadal, Ernest (2002). "Elementos para la evaluación de interfaces de consulta de bases de datos". El profesional de la información , septiembre-octubre, 11:

[11] Rodríguez Muñoz, José V.; Saorín, Tomàs (1998). "Modelado documental de servicios de información en web". El profesional de la información , septiembre, 7