



**UNIVERSIDAD
ALBERTO HURTADO**

Facultad de Economía y Negocios

**PROPUESTA DE METODOLOGÍA DE APLICACIÓN Y OPTIMIZACIÓN DE
BIG DATA EN UNIDADES DE INFORMACIÓN**

**Seminario de Grado para optar al Grado de Licenciado en Gestión de
Información, Bibliotecología y Archivística**

Por

Luis Aguilar Vera

Marcos Briones Vera

Marcelo Burgos Aravena

Carolina Ruiz Cornejo

Profesor guía: Claudio Escobar Arriagada

Santiago, Chile

2016

Tabla de contenido

.....	1
Capítulo 1: Presentación	4
1.1 Problema:	4
1.2 Justificación:	7
1.3 Objetivos	8
1.3.1 Objetivos Generales:	8
1.3.2 Objetivos específicos:	8
1.4 Pasos metodológicos:	9
Capítulo 2: Marco Teórico	12
2.1 Inicios de la Tecnologías de Información y Comunicación (TIC) e Internet al servicio de las necesidades informacionales	12
2.1.1 Web 2.0 y su aplicación	14
2.1.2 Web 3.0 o Sensorial	16
2.1.3 Características de la web al servicio:	17
2.2 Big Data	19
2.2.1 Historia	19
2.2.2 El Big Data en Latino América	21
2.2.3 Definiciones	22
2.2.4 Tipos de datos	25
2.2.5 Características de Big Data: Volumen, variedad y velocidad	27
2.2.6 Fuentes de Big Data	31
2.2.7 Bases de datos y tecnologías	34
2.2.8 Analítica de Big Data	35
2.3 Unidades de Información	36
2.3.1 Biblioteca	41
2.3.2 Archivo	43
2.3.3. Centro de documentación	45
Capítulo 3: Metodología	47
3.1 ICAV como metodología base	47
3.2 Propuesta de metodología	49
3.2.1 Identificar	51

3.2.2 Consolidar	58
3.2.3 Analizar.....	62
3.2.4 Visualizar	64
Conclusión.....	69
Glosario	73
Referencias Bibliográficas:	76

Capítulo 1: Presentación

“Propuesta de metodología de aplicación y optimización de Big Data en unidades de información”

1.1 Problema:

La nueva era de la información ha tomado dominio de todo lo que hoy conocemos. No existen organizaciones que no utilicen la información como una fuente de progreso para su beneficio. El mundo en general crea millones y millones de datos que se van acumulando en este mundo virtual que ya se nos hace muy familiar pero que sin duda aún no conocemos en profundidad. Internet nos permitió acceder a un universo variado de datos que a cada minuto aumenta su volumen a una velocidad abismante. Toda esta avalancha de datos son producidos diariamente por las diversas plataformas que tienen una continua actividad digital en todo el mundo. Cada vez que visitamos un sitio web, hacemos una consulta en los buscadores, compartimos una foto, actualizamos nuestra red social preferida o simplemente cada vez que accedemos a un ordenador o a nuestros dispositivos móviles, estamos generando datos, y toda esta enorme masa de datos se conoce como Big Data.

La compañía norteamericana Domo, la cual se especializa en software, hizo un análisis de la información generada cada 1 minuto en internet. A través de una completa infografía denominada “Data Never Sleeps” (“Los datos nunca duermen”) lograron plasmar la enorme actividad digital de internet cada 60 segundos, con números realmente imponentes.

Dentro de sus datos detallan que se generan 48 horas de vídeo en Youtube, 2 millones de consultas en Google, 100.000 Tuits en Twitter, más de 685.000 publicaciones en Facebook y la cifra suma y sigue, considerando que la población digital sigue creciendo cada año.¹

¹ Domo. Data Never Sleeps. Recuperado de <https://www.domo.com/learn/infographic-data-never-sleeps>



Figura 1: Domo. Data Never Sleeps. Recuperado de <https://www.domo.com/learn/infographic-data-never-sleeps>

Debido a esta creciente cantidad de información, es que las organizaciones han tomado cartas en el asunto y actualmente están diseñando soluciones para el procesamiento, gestión y control de estos datos masivos. Las organizaciones están entendiendo e internalizando el real valor de la Big Data para generar ventajas competitivas que les permitan mejorar la toma de decisiones y un análisis concreto de sus clientes y sus necesidades además de abrir camino para encontrar nuevas fuentes de ingreso adelantándose a sucesos de forma predictiva gracias al uso de los datos.

Marcelo Román, gerente general de Asociación Chilena de Empresas de Tecnología de Información (ACTI), ejemplifica algunos posibles usos del Big Data en distintos tipos de organizaciones: “En el área financiera se pueden analizar en tiempo real las transacciones y con ello facilitar la detección de operaciones fraudulentas; en salud, al estudiar los patrones, se pueden prever riesgos sanitarios; el comercio puede saber qué piensa un consumidor y cuáles son sus necesidades, para así responderlas. Biométricamente, la policía y el poder judicial también hacen uso de los datos almacenados para detectar personas, verificar su identidad, etcétera”.²

Las unidades de información tales como bibliotecas, archivos, centros de documentación entre otros, al igual que las otras organizaciones también generan una enorme cantidad de datos de todo tipo y a gran velocidad (Big Data) y que también necesitan ser gestionados y valorados. Sin embargo, surgen las siguientes interrogantes, ¿están siendo aprovechados aquellos datos para generar nuevos servicios?, ¿Están utilizando los datos para la toma de decisiones dentro de la unidad de información? ¿Se está aprovechando la Big Data para optimizar procesos? ¿Cómo se puede utilizar Big Data en Unidades de información?

Según datos de la Asociación Chilena de Empresas de Tecnología de Información (ACTI), sólo el 17% de las empresas chilenas han implementado el Big Data o piensa hacerlo en el corto plazo. Sin embargo hay cifras alentadoras para nuestro país en esta materia pues la Corporación Internacional de Datos (IDC) adelanta que la inversión en Big Data para

² Román, M. (2014). El fenómeno del Big Data en Chile y el mundo. Recuperado de <http://www.telecheque.cl/noticias/el-fenomeno-del-big-data-en-chile-y-el-mundo/>

nuestro país será de 50 millones de dólares en los próximos años, siendo Chile la nación de mayor crecimiento en la región.³

La biblioteca está en la actualidad en disposición de generar sistemas de información complejos que doten de sentido las demandas de los lectores, las respuestas a los mismos y las necesidades de los diferentes agentes que participan en el ecosistema bibliotecario, en cuyo eje se sitúa la biblioteca, tanto desde el punto de vista del suministro de servicios como desde el de la generación de información estratégica. La toma de decisiones basada en datos, ha sido siempre una aspiración de cualquier biblioteca, ahora el problema no es la falta de datos, sino la necesidad de saber gestionar y valorar los mismos teniendo en cuenta la combinación de volumen, variedad y velocidad que ahora adquieren.⁴

1.2 Justificación:

Big Data es una gran oportunidad en nuestra actual era de la información, es una nueva revolución info-tecnológica de gran impacto, y promete cambiar por completo la manera de entender, de aprender de hacer negocios y de tomar decisiones, en resumen, es una nueva herramienta a nuestro alcance para la creación de conocimiento.

En unidades de información el Big Data puede ser aprovechado para conocer los documentos más solicitados, la cantidad de búsquedas realizadas por usuarios en los Opac's con resultado positivo o negativo, ranking de usuarios o libros en bibliotecas, gustos e intereses de los usuarios a través de redes sociales, informes de préstamos y devoluciones, cantidad de libros procesados cada mes, número de visitas en sitio web, clics realizados por usuarios para llegar a la información en bases de datos, webs,

³ (ACTI), A. C. (30 de Abril de 2015). Big Data toma fuerza en Chile con soluciones que analizan textos, voz, imágenes y video. Recuperado de (ACTI), Asociación Chilena de Empresas de Tecnología de Información: <http://www.acti.cl/detalle-noticia.php?mul=737>

⁴ Hernandez, J. (10 de Abril de 2015). *Nubeteca*. Obtenido de <http://www.nubeteca.info/blog/?p=25>

o bibliotecas digitales, establecer toma de decisiones en desarrollo de colecciones en base a los usuarios y sus datos, etc.

Con la revolución tecnológica y la entrada de la información y el conocimiento a la sociedad, la profesión de los especialistas en información ha sufrido una gran evolución siendo capaz de adaptarse a todos los cambios tecnológicos para poder satisfacer las necesidades de sus usuarios. Han sabido evolucionar en su profesión y detectar las necesidades informacionales de distintos sectores abriendo espacios para nuevas áreas de trabajo. El profesional de la información es experto en transformar las necesidades de información del usuario en estrategias y recursos para la generación de conocimiento a través de la investigación y el aprendizaje.

Considerando lo anterior, y teniendo en cuenta que el Big Data está aquí, y que tiene como fin la generación de conocimiento que ayude a tomar decisiones de valor, el profesional de la información que es experto en esta tarea, tiene una participación muy relevante pues el Big Data se integrará (o ya se está integrando) en nuestra práctica diaria.

1.3 Objetivos

1.3.1 Objetivos Generales:

- Diseñar metodología de aplicación de Big Data en unidades de información con el propósito de optimizar sus servicios y funciones.

1.3.2 Objetivos específicos:

- Presentar el estado del arte de la Big Data mencionando tecnologías existentes para iniciar un proyecto que permita la gestión de datos en unidades de información.
- Identificar las fuentes de datos que manejan las unidades de información, para posterior control y uso.
- Identificar los tipos de datos generados por la Unidad de Información y definir la relevancia de estos para su posterior uso.
- Definir posibles acciones en base al análisis de datos

1.4 Pasos metodológicos:

- Revisión de literatura pertinente a Big Data y sus componentes:

Revisión de textos académicos, artículos, noticias y fuentes de información de variada procedencia sobre Big Data y su desarrollo en el tiempo. Esto nos brindará un amplio panorama del tema a tratar considerando sus distintos enfoques y las posibilidades de tratamiento de datos en diferentes tipos de organizaciones.

- Clasificación y orden analítico de fuentes primarias y secundarias:

Reunir la bibliografía referente a la temática de Big Data a abordar, y analizar una a una su contenido para poder rescatar los datos más importantes y relevantes a la investigación que se va a realizar referente a las posibilidades que otorga el Big Data a las Unidades de Información. Esto conllevará crear un orden analítico de la información recaudada, separada por sus autores y destacando la fuente de la cual provienen.

Una vez realizado el análisis de la bibliografía, ésta será clasificada y valorada para luego ser destinada por temática a cada área en específica que tratará la investigación, estas temáticas son:

- Definiciones del término Big Data y sus características generales.
 - Tecnologías asociadas a Big Data.
 - Posibles beneficios de la gestión de datos en distintos tipos de organización.
 - Usos actuales de Big Data en organizaciones de Hispanoamérica.
 - Big Data en Unidades de Información en el mundo
 - Estado actual de Big Data en Chile
- Identificar y seleccionar metodología base para la implementación del proyecto Big Data.

Identificar metodologías que actualmente se usan en el manejo de Big Data, y seleccionar una que sea acorde a las necesidades planteadas en el

desarrollo del proyecto. Esto permitirá establecer una metodología base que luego se implementará en Unidades de Información.

- Presentar Hadoop como una posible herramienta para la implementación de un sistema Big Data en unidades de información chilenas.

Hadoop es la tecnología líder en código abierto dentro del dominio de Big Data. En complemento con el punto anterior, y en base a la información recolectada y observada, se analizará la implementación de Hadoop en Unidades de Información.

- Dar a conocer los beneficios y posibles riesgos del uso de Big Data

En base al análisis desarrollado en los puntos anteriores, determinar los beneficios y posibles riesgos que nos puede generar la implementación de la metodología en nuestra unidad de información es algo que debemos considerar. La implementación de Big Data en unidades de información traerá riesgos, ya sean en costos, en personal, en la cultura organizacional entre otros y también beneficios en optimización de tiempos, mejor toma de decisiones, creación de nuevos servicios en base a datos demostrables, etc.

- Definición de metodología Big Data para Unidades de Información

Recolección de la información resultante del análisis de los puntos anteriores para la creación de una metodología que permita utilizar el Big Data en beneficio del desarrollo de los servicios, procesos y toma de decisiones en las Unidades de Información.

Daremos a conocer de forma detallada los riesgos y beneficios a considerar al implementar Big Data en unidades de información.

- Conclusiones prácticas con respecto a la metodología de aplicación de Big Data en unidades de información basadas en el análisis realizado.

Conclusiones detalladas en base a la investigación y análisis realizado, destacando los pros y los contras de la implementación de Big Data en Unidades de Información. Comentaremos el aporte de esta herramienta a la profesión y como esta podría ser parte del futuro de nuestras Unidades de

Información al ganar una ventaja competitiva frente a diversos puntos como el Marketing, toma de decisiones, creación de nuevos servicios etc.

Capítulo 2: Marco Teórico

2.1 Inicios de la Tecnologías de Información y Comunicación (TIC) e Internet al servicio de las necesidades informacionales.

La tecnología cambia nuestra forma de vivir, aplicando un conjunto de conocimientos y habilidades con un claro objetivo, satisfacer las necesidades. Las Tecnologías de Información y Comunicación (TIC) producido en el ámbito de informática y telecomunicaciones, permiten el acceso, producción, tratamiento y comunicación de información en diversos códigos (texto, imagen, sonido, etc.), se perfila como el medio para conseguir entretenimiento, noticias y trabajar a distancia, inventos como la televisión, el automóvil o el teléfono, han cambiado radicalmente nuestra forma de interactuar con el mundo y nuestra visión del mismo, incluso la forma de comunicarse con otras personas.

En el siglo XX Internet es uno de los inventos tecnológicos más significativos. En la década del setenta el Departamento de Defensa Norteamericano desarrolla protocolos que especifican a las máquinas como debe transmitirse la información a centros de investigación, siendo los dos más importantes: TCP *Transmission Control Protocol* (Protocolo de Control de Transmisión) y IP *Internet Protocol* (Protocolo de Internet), con el fin de comunicar entre los que forman la red, parte el proyecto, denominado ARPANET (Advanced Research Project Agency Network). El ejército de EE.UU dividió la red en dos partes: MILNET para usos militares, restringida y a salvo del uso privado; y ARPANET para usos civiles, actualmente conocida como Internet es abierta a cualquiera que desease acceder.

Internet es un conjunto descentralizado de redes de comunicación interconectadas, los cuales garantizan que las redes físicas heterogéneas que la componen, actúan como una red lógica única de alcance mundial⁵. Al ser definido como medio de comunicación entre las subredes, comenzó la expansión global desarrollando la familia de diversos servicios y protocolos en internet, algunos conocidos:

-SMTP o *Simple Mail Transfer Protocol* (envío de correo electrónico).

-FTP o *File Transfer Protocol* (transmisión de archivos).

⁵ Carretero, J. [et al] (2001) Descubre Internet.

-VoIP o *Voice Over Internet Protocol* (transmisión de contenido y comunicación multimedia).



Figura 2: Taringa (2016) Macri lanzará un "Plan Federal de Internet" el 17 de Mayo. Taringa. Recuperado de <http://www.taringa.net/posts/noticias/19413805/Macri-lanzara-un-Plan-Federal-de-Internet-el-17-de-Mayo.html>

Uno de los servicios más exitosos que ha tenido internet ha sido la World Wide Web (WWW o Web), hasta tal punto que es habitual la confusión entre ambos términos. La WWW se desarrolló en Europa por Tim Berners Lee en el año 1992 y utiliza internet como medio de transmisión. Es un conjunto de recursos que permite, de forma sencilla, la consulta remota de archivos de hipertexto, como:

-HTML o Hypertext Markup Language (lenguaje de marcas de hipertexto).

-HTTP o Hypertext Transfer Protocol (protocolo de transferencia de hipertexto).

-Navegador o web browser (un programa cliente).

2.1.1 Web 2.0 y su aplicación

En el año 2004 dos empresas del sector informático, O'Reilly y MediaLive International introdujeron el concepto web 2.0 dándole cierta entidad teórica el 2005 a través del artículo "What is Web 2.0 Design Patterns and Business Models for the Next Generation of Software".⁶ El año 2000 marcó el inicio de una nueva era, marcada por la transición de la tecnología web 1.0 a la 2.0 Nace de la observación, en un entorno empresarial con una impronta tecnológica y de marketing. Existiendo una serie de características que pueden cumplir con un servicio web del lado 2.0, las que por lo demás obtienen gran éxito. Lo valioso de las aplicaciones Web 2.0 son los datos.

"Hoy sólo se puede pensar en versiones beta siempre dispuestas a ser modificadas, a tono con una gestión efímera y una circulación acelerada de la información. El espíritu de las empresas Web 2.0 radica en transformar prototipos en versión beta y colocarlos en línea, utilizando el método de ensayo y error para aprender del consumo de los usuarios".⁷

El usuario actúa de la manera que desee, en forma tradicional (navegando a través de los contenidos) o de forma activa (creando y aportando sus contenidos). O'Reilly afirmó que las aplicaciones web 2.0 son aquellas que sacan partido a las ventajas intrínsecas de la web, servicio que mejora cuando más gente lo use, actualizando y remezclado los datos de múltiples recursos, creando una "arquitectura de participación" en red, utilizando determinadas tecnologías:

-Ajax (asynchronous javascripts and xml): combinación de xml y javascript, aplicaciones que permiten reducir el tráfico de datos y la carga de trabajo del servidor con una mayor interactividad lenguaje utilizado por google maps, Gmail, flickr, etc.

⁶ O'Reilly y MediaLive (2005) "What is Web 2.0 Design Patterns and Business Models for the Next Generation of Software".

⁷ Pardo, Hugo. (2013). Planeta web 2.0. Capítulo 1. Nociones básicas alrededor de la Web 2.0. Recuperado de https://docs.google.com/document/d/1q6oul37iNKw3ELg0lhXlpag6wGwOcUyXIzx_A0YT0U8/edit?hl=es&pli=1

-Mashups o aplicaciones web híbridadas: integra de manera transparente los datos de otros servicios web gratuitamente. HousingMaps.com combina la información de una base de datos inmobiliaria con google maps.

-Software social: elemento clave de la web cuando se refiere al aprovechamiento de la inteligencia colectiva. Conjunto de aplicaciones que permite a los individuos comunicarse como blogs (weblogs o bitácora), foros, las wikis, etc.

-Rss: familia de formatos xml para difundir información y ser reutilizada por otros programas o sitios web, implementado en webs de instituciones, revistas, bases de datos, catálogos de bibliotecas, etc..

“Imaginemos una biblioteca pública, ha conseguido una serie de fotografías antiguas y decide digitalizar y publicarlas en internet. Esta publicación puede hacerse a través de una serie de páginas html o mediante un catálogo bibliográfico. Estas opciones, correctas y legítimas, serían consideradas web 1.0. Ahora bien, supongamos que se decide publicarlas a través de Flickr, se abre una cuenta y las publica en el servidor, permitiendo que los usuarios puedan verlas, añadir comentarios, reutilizarlas, etc., entonces estaríamos ante un servicio web 2.0”⁸

Ejemplos de servicios web 2.0:

- Blogs: conocidos web logs o bitácora permite al internauta escribir mensajes con información ordenada cronológicamente

- Wikipedia: enciclopedia libre, el usuario consulta gratuitamente, además crea, modifica y amplía libremente.

- Flickr: permite al usuario, compartir, comentar y guardar sus fotografías, con el provecho de las cámaras digitales.

- Google maps: sistema de búsqueda de localizaciones geográficas.

- Youtube: cualquier persona registrada puede compartir videos, añadir comentarios, etiquetas para describirlo, etc., el resto puede dar una puntuación (rating), añadir comentarios, seleccionarlos como favoritos, etc.

Esta atrapante “tela de araña” (World Wide Web) permite que las páginas web con su lenguaje de hipertexto funcionen en Internet, y sean más

⁸ Margaix, D. (2007). Conceptos de web 2.0 y bibliotecas 2.0: origen, definiciones y retos para las bibliotecas actuales. Recuperado de <http://eprints.rclis.org/9521/1/kx5j65q110j51203.pdf>

dinámicas, esto se conceptualiza como la aplicación de tecnologías interactivas, participativas y multimedia a los servicios y que aparte de texto, se incluya:

- Imagen
- Sonido
- Video
- Enlaces
- Programas (interpretes)
- Formato (html)

“Pero son los principios de compartir, reutilizar, mejora continua, consideración del usuario como fuente de información, confianza, aprovechamiento de la inteligencia colectiva, etc., los que han impulsado el establecimiento de la actitud 2.0, haciendo que la tecnología pasa a segundo plano”.⁹

2.1.2 Web 3.0 o Sensorial

En los últimos años O’Reilly hace alusión a la Web 3.0 conceptuándose como “web semántica” por su creciente capacidad para añadir metadatos semánticos que son tratados por máquinas de procesamiento, reconociendo la mejora de los servicios, más aún, el mayor volumen de información se consigue manejar mejor. Pero en la quinta edición del Foro Internacional de Contenidos Digitales (FICOD) reconoce la web como “sensorial” refiriéndose al estatus de los dispositivos móviles (smartphones) confieren a Internet (Internet sensitivo), tomando en cuenta el “tiempo real”, que cambia el status-quo de la información, negocios publicitarios y servicios para el ciudadano, destacando las redes sociales y la conexión entre la comunidad.

Este concepto inspiró nueva literatura profesional para su planteamiento y aplicación en distintos ámbitos: periodismo, educación, bibliotecas, etc.; es válido considerar una biblioteca o un archivo 2.0 dispuestos a los aspectos novedosos que aporta la web y a la modificación del trabajo que componen

⁹ Margaix, D. (2007). Conceptos de web 2.0 y bibliotecas 2.0: origen, definiciones y retos para las bibliotecas actuales. Recuperado de <http://eprints.rclis.org/9521/1/kx5j65q110j51203.pdf>

los profesionales de la información frente a sus conceptos y sus implicaciones; hacen hincapié en los aspectos tecnológicos y su relación con el usuario; es con este precisamente, que el nuevo modelo de vínculo ha tenido mayor calado gracias a sus servicios.

A la hora de seleccionar el software es importante considerar en primer lugar al usuario, las facilidades que ofrece el sistema en cuanto a tipos, formas y niveles de acceso que se habilitarán según los grupos especiales de usuarios definidos, navegación, ayudas incorporadas, alternativas de interrogación para las distintas bases de datos, así como la recuperación y presentación de los resultados en formatos que se concierne con las preferencias y necesidades de los usuarios en plataformas interactivas.

En ese sentido, González y Goyannes (2004) explica que las bases de datos referenciales y la literatura científica permiten establecer relaciones entre conocimientos, instituciones y científicos, que no son visibles para el investigador o el gerente organizacional.

“Contenidos con un nivel de complejidad que convierten la Recuperación de Información (RI) en un proceso, pero el proceso de RI no es posible sin la existencia de un fondo documental con las ofertas informacionales que permitan dar respuestas a las solicitudes de información, sólo tiene lugar a partir de la expresión de una necesidad de información”.¹⁰

2.1.3 Características de la web al servicio:

1. El software no se limita a un solo dispositivo: Interfaces a los servicios web, con la capacidad de acceso en todo lugar y momento, promueve la usabilidad, familiaridad y sencillez de las interfaces de los sistemas operativos (como los blog), como los teléfonos móviles de tercera generación (3G) los cuales permiten más plataformas para el entretenimiento, gestión de información, consumo de los media, producción de contenidos multimediales y el fortalecimientos de redes sociales. Un ejemplo es la lectura del libro digital conocidos como e-book, versión electrónica de un libro o un texto publicado en la World Wide Web u otros formatos electrónicos, con un modelo hipertextual.

¹⁰ Calderín Cruz, Mabel. (2007). Sistema de información documental en la Academia: Situación actual y perspectivas futuras, caso UCAB. Recuperado de http://www.cienciared.com.ar/ra/usr/3/333/n6_v1pp51_76.pdf

2. La world wide web como plataforma: ofrece software gratuito utilizando la web como plataforma. En esta existen las herramientas y contenidos del usuario, y no en su computadora, lo que permite trabajar par a par. Las ideas fluyen en formato post, promoviendo pensamiento efímero (como slogans), con fecha de caducidad, para públicos carentes de tiempo de atención.

3. Aprovecha la inteligencia colectiva: valora y potencia la relevancia de las aplicaciones de escritura colaborativa en el nuevo periodismo. Permite la revisión entre pares en todo tiempo, nuevas entradas representan el trabajo, que hace crecer la inteligencia colectiva en forma eficiente. La aplicación permite al usuario publicar, y la comunidad determinara la relevancia del contenido. No existen filtros en la entrada de contenidos sino de salida.

4. Modelos de programación ligera: busca la simplicidad de la Arquitectura de la información. La programación ligera, pone énfasis en la reducción de la complejidad (menos es más) evitando las excesivas especificaciones y funciones del software empaquetado (*transparentando el trabajo y el régimen republicano*) lo que permite ensamblar productos con creatividad (valor añadido), además aumenta la productividad y sinergia, ausente en productos por separado; debido a las aplicaciones híbridas denominadas mashups (punto de conexión entre funciones diferentes).¹¹

“La simplicidad genera formas recreativas de reutilizar la programación hacia la estandarización, integración y usos bajo demanda, siempre pensando en la inclusión del usuario como productor de contenidos”¹²

Toda esta revolución tecnológica hoy en día nos ha llevado a estar inmersos en un océano de información que no nos deja indiferentes pues a cada segundo se están generando datos en internet en todo el mundo llevándonos a poner nuestros ojos en lo que hoy se conoce como Big Data.

¹¹ Pardo, Hugo. (2013). Planeta web 2.0. Capítulo 1. Nociones básicas alrededor de la Web 2.0. Recuperado de https://docs.google.com/document/d/1q6oul37iNKw3ELg0lhXlpag6wGwOcUyXIzx_A0YT0U8/edit?hl=es&pli=1

¹² *Ibíd.*

2.2 Big Data

Big data es el análisis y gestión de grandes cantidades de información y que no pueden ser tratados de la manera convencional, ya que supera la capacidad de los software usados hasta ahora. “Hoy, recién iniciado el siglo XXI, se generan, según la Unión Europea, 1.700 nuevos billones de bytes por minuto. Equivale, dicen, a unos 360.000 DVD, lo que de media vienen a ser seis megabytes por persona y día (más o menos la cantidad de datos que generaba en toda su vida una persona del siglo XVI)”¹³. Por lo que la función de la Big Data es convertir toda la cantidad de datos en información útil para la organización que las use, y facilitar la toma de decisiones. Aunque más allá de su uso para manejar amplias cantidades de información, el uso de la Big Data se considera en gran medida como una forma de potenciar la organización, ya que las organizaciones o empresas al analizar los datos pueden saber cuál es el perfil de sus usuarios o clientes creando un perfil de ellos, en cuanto a los servicios o productos que estos requieren, dando así la información que la entidad necesita para saber cómo interactuar con sus usuarios y en cómo les hacen llegar el servicio de la forma más adecuada para cada persona.

“En 2004, Walmart echó un vistazo al contenido de sus gigantescas bases de datos de antiguas transacciones: qué artículo había comprado cada cliente y su coste total, qué más había en el carrito de la compra, la hora del día, e incluso el tiempo que hacía. Así, observó que antes de un huracán no sólo aumentaban las ventas de linternas, sino también las de Pop-Tarts, un dulce para el desayuno”¹⁴

2.2.1 Historia

Desde que aparecieron las primeras formas de escritura, la raza humana no ha cesado de recopilar información. Ha sido una evolución gradual pero constante, y este proceso se ha acelerado a finales del Siglo XX con la invención del almacenamiento digital y del internet provocando un aumento descomunal del volumen de datos.

¹³ Van Den Berg, Eva. (31 de marzo de 2015). ¿Qué es eso del “Big Data”? Recuperado de http://elpais.com/elpais/2015/03/26/buenavida/1427382655_646798.html

¹⁴ Mayer-Schönberger, Viktor; Cukier, Kenneth (2013) Big data: La revolución de los datos masivos. Turner Publicaciones S.L.

El término “Big Data” se empleó por primera vez en un artículo fechado en agosto de 1997 de los investigadores de la NASA Michael Cox y David Ellsworth. Ambos afirmaron que el ritmo de crecimiento de los datos empezaba a ser un problema para los sistemas informáticos actuales.¹⁵

Esto se denominó el “problema del Big Data”. A continuación se presenta una tabla donde se expone la evolución de los datos hasta llegar al Big Data y su analítica.

Periodo de tiempo	Hito
1930 - 1949	Se realizan trabajos para descodificar los mensajes alemanes en la segunda guerra mundial.
	Se automatiza la defensa antiaérea apuntando contra aviones enemigos.
	Se realizan simulaciones por ordenador para predecir el comportamiento de una reacción nuclear en cadena.
1950 - 1969	Se genera el primer modelo de predicción meteorológica.
	La analítica resuelve el “problema del camino más corto” mejorando el transporte aéreo y la logística,
1970 – 1999	Se crea el modelo para predecir el precio óptimo de las acciones en el futuro.
	Se comercializa la primera herramienta para apoyar a las decisiones guiadas por modelo.
	Lanzamiento de Amazon y eBay. Inicia personalización de la experiencia online.
	Se aplican algoritmos a las búsquedas

¹⁵ Cox, M., Ellsworth, D. (1997). Application – controlled demand aging for out – of -core visualization. Recuperado de <http://www.nas.nasa.gov/assets/pdf/techreports/1997/nas-97-010.pdf>

	web para maximizar la relevancia de los resultados.
2000 – Actualidad	Uso generalizado de la analítica. Recomendaciones de compra, gestión del tráfico, entre otros...
	Proceso del lenguaje natural. Trabajo con datos no estructurados. Ejemplo mensajes de facebook, páginas web, PDFs e emails.
	Llega de Big Data. Cada día se generan 2.5 trillones de bytes de datos.
	Se hace viable la computación distribuida y el análisis de Big Data para la mayoría de las organizaciones.
Futuro	Intercambio de analítica en la nube. Aumento de la economía colaborativa.
	Las campañas de marketing desaparecen ya que la interacción con los consumidores es personal.

Figura 3: Ardila Cañas, Esneyder; Gómez Díaz, Isabel (2014) Estrategias para la gestión de grandes volúmenes de datos por medio de Big Data en el contexto de la analítica de negocios: Caso MVM ingeniería de software S.A.S. Recuperado de http://bibliotecadigital.usbcali.edu.co/jspui/bitstream/10819/2702/1/Gestion_Volumen_Datos_Ardila_2014.pdf

2.2.2 El Big Data en Latino América

De acuerdo a IDC, International Data Corporation, 2014 (Llandez, Anesini, Crippa, & Villate, 2014) Latino América aún se encuentra en una fase de evangelización de las tecnologías, en la que los principales proveedores de tecnología están dedicando grandes esfuerzos en promover los beneficios de su implementación. Chile, Colombia y Perú son los países donde más se invertirá en Big Data, siendo Chile el país con mayor tasa de crecimiento de adopción de las tecnologías.

También se estima que más del 65% de los usuarios brasileños y mexicanos de sitios sociales gastan más de una hora por semana en Facebook, y un 24% lo hace en Twitter. Desde el punto de vista empresarial, el 56% de las

empresas latinoamericanas están usando sitios sociales o de comunidades virtuales para obtener retroalimentación de sus productos y servicios.

Sin embargo, aún existe mucho espacio para crecer, dado que la adopción de estas tecnologías aún no se considera prioritaria para las compañías y mucho menos para el de la región. Es preciso señalar que estas cifras se refieren sólo al ámbito empresarial, en el ámbito que nos compete, las Bibliotecas, Archivos y Centros de documentación son casi nulas las iniciativas de la utilización de Big Data, sabemos que en Chile la Biblioteca del congreso, el ex profesor de nuestra carrera en la Universidad Alberto Hurtado Patricio Pastor está liderando iniciativas en esa institución, por otra parte algunas Instituciones educacionales como la USACH con su Centro de Investigación, Sociedad, Economía y Cultura (Cisec) está realizando estudios ocupando las herramientas y beneficios del Big Data.

2.2.3 Definiciones

Algunas definiciones que se pueden encontrar de Big Data son las siguientes:

“Big Data se refiere a los conjuntos de datos cuyo tamaño está más allá de las capacidades de las herramientas típicas de bases de datos para capturar, almacenar y analizar”.¹⁶

“Big Data una nueva generación de tecnologías, arquitecturas y estrategias diseñadas para capturar y analizar grandes volúmenes de datos provenientes de múltiples fuentes heterogéneas a una alta velocidad con objeto de extraer valor económico de ellos”¹⁷

En febrero de 2001 Doug Laney, analista de Gartner, publicó un artículo donde definía a Big Data o Datos masivos como el crecimiento constante de datos como una oportunidad y un reto para investigar en el volumen, la velocidad y la variedad, y que en definitiva hace referencia a una cantidad de

¹⁶ McKinsey Global Institute (2011). Big data: The next frontier for innovation, competition, and productivity. Recuperado de file:///C:/Users/Alumno/Downloads/MGI_big_data_full_report.pdf

¹⁷ Consultora IDC (2012). De que forma Big Data transforma la protección y el almacenamiento de datos. Recuperado de <http://webdocs.commvault.com/assets/es/idc-big-data-transforms-data-protection-and-storage-analyst-report-es.pdf>

datos tal que supera la capacidad de software convencional para ser capturados, administrados y procesados en un tiempo razonable.¹⁸

Es importante señalar que hasta el día de hoy las 3V's (volumen, variedad y velocidad, explicadas más adelante) siguen siendo las dimensiones comúnmente aceptadas del Big Data.

Para continuar presentaremos otras definiciones de Big Data de las cuales se pueden encontrar muchas, pero con las siguientes, que son más actuales, intentaremos mostrar el Estado del Arte de Big Data.

“Big Data aplica a la información que no puede ser procesada o analizada mediante procesos tradicionales”.¹⁹

Big Data son “cantidades masivas de datos que se acumulan con el tiempo que son difíciles de analizar y manejar utilizando herramientas comunes de gestión de bases de datos”.²⁰

Big Data se refiere “al tratamiento y análisis de enormes repositorios de datos, tan desproporcionadamente grandes que resulta imposible tratarlos con las herramientas de bases de datos y analíticas convencionales”.²¹

Por otra parte, el analista Dan Kunsnetzky, del Grupo Kunsnetzky, señala que “La frase Big Data se refiere a las herramientas, procesos y procedimientos que permitan a una organización crear, manipular y administrar grandes conjuntos de datos e instalaciones de almacenamiento.”

22

También Forrester define Big Data “como las técnicas y tecnologías que hacen que sea económico hacer frente a los datos a una escala extrema. Big Data trata de tres cosas: 1) Las técnicas y la tecnología, lo que significa que la empresa tenga personal, el cual tenga gran representación y análisis de datos para tener un valor agregado con información que no ha sido

¹⁸ Laney, D.(2001). 3D Data Management : Controlling data volumen, velocity and variety. Recuperado de <http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>

¹⁹ ZDNet (2013). CBS Interactive, What is “Big Data?”. Recuperado de: <http://www.zdnet.com/topic-big-data/2013>.

²⁰ Thinkupapp.com(2012). Recuperado de <http://thinkupapp.com>

²¹ Dans, E (2011). Recuperado de <http://www.enriquedans.com/2011/10/big-data-una-pequena-introduccion.html>

²² Kunsnetzky, D. (2010). Recuperado de E. Plugge, P. Membrey & T.Hawkins, The Definitive Guide to MongoDB: NoSQL Database for cloud and Desktop Computing, Published Apress Media LLC, New York, 2010.

manejada. 2) Escala extrema de datos que supera a la tecnología actual debido a su volumen, velocidad y variedad. 3) El valor económico, haciendo que las soluciones sean asequibles y ayuden a la inversión de negocios”.²³

Según Zdnet.com²⁴ Big Data “se refiere a las herramientas, los procesos y procedimientos que permitan a una organización crear, manipular y gestionar conjuntos de datos muy grandes y las instalaciones de almacenamiento”.²⁵

Gartner también define Big Data como “un gran volumen, velocidad o variedad de información que demanda formas costeables e innovadoras de procesamiento de información que permitan ideas extendidas, toma de decisiones y automatización del proceso”.²⁶

Y también encontramos la definición que indica que “Big Data consiste en consolidar toda la información de una organización y ponerla al servicio del negocio”.²⁷

Y ahora citaremos autores fundamentales, quizás por encontrarnos más cercanos a su pensamiento, según Mayer – Schonberger, Viktor y Cukier, Kenneth “los datos masivos están en el punto de remodelar nuestro modo de vivir, trabajar y pensar”...y dado que el conocimiento se almacena en las Bibliotecas, los bibliotecarios son las personas más idóneas para descubrir y mostrar esas nuevas formas de vivir, trabajar y de pensar”.²⁸

Kenneth Cukier también señala con claridad: “Más no es solo más. Más es nuevo. Más es mejor. Más es distinto”²⁹

También es importante decir que Big Data no se refiere a alguna cantidad en específico, ya que es usualmente utilizado cuando se habla en términos de petabytes y exabytes de datos.

²³ Hopkins, B. (2011). Beyond the hype of Big Data. Recuperado de http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0121-11292015000100006

²⁴ ZDNet es una empresa ligada al rubro de la tecnología.

²⁵ ZDNet (2010). Recuperado de <http://www.zdnet.com/search?q=big+data>

²⁶ Salgado, M. (2014). Oracle apuesta por Big Data con tecnología y proyectos. Recuperado de <http://www.computerworld.es/big-data/oracle-apuesta-por-big-data-con-tecnologia-y-proyectos>, 2014.

²⁷ Russom, P. (2012). Big Data Analytics, TDWI (The Data Warehousing Institute. Recuperado de http://www.tableau.com/sites/default/files/whitepapers/tdwi_bpreport_q411_big_data_analytics_tableau.pdf

²⁸ Schonberger, V., Cukier, K. (2013, p.232)

²⁹ Cukier, K. (2014, p.37)

Entonces ¿cuánto es demasiada información de manera que sea escogida para ser procesada y analizada utilizando Big Data? Es por esto que mostraremos brevemente las unidades de información en términos informáticos.

La unidad más pequeña que se puede representar en informática es el Bit. Un Bit solo puede tomar dos valores 0 o 1. Este sistema de numeración en el que los números se representan utilizando solamente las cifras cero y uno, se denomina sistema Binario.

El Byte, es el conjunto formado por 8 Bites. Con un Byte puedes representar cualquier carácter.

Un Kilobyte, son 1024 Bytes, un Megabyte son 1024 Kilobytes, un Gigabyte son 1024 Megabyte, y así sucesivamente.

Para efectos de nuestro trabajo definiremos Big Data como: Enorme cantidad de volúmenes de datos generados diariamente por distintos agentes externos e internos de la organización captados desde plataformas tecnológicas. Estos pueden ser aprovechados a través del análisis y control de datos para generar una mejora en la toma de decisiones y servicios de la organización.

2.2.4 Tipos de datos

“Los Big Data son diferentes de las fuentes de datos tradicionales que almacenan datos estructurados en las bases de datos tradicionales”³⁰. Se suele dividir los datos en dos: estructurados y no estructurados, sin embargo con las nuevas tecnologías y herramientas para manipular el Big Data, han surgido también los datos semi-estructurados.

- Datos estructurados: Son la mayoría de los datos tradicionales, estos tienen longitud, formato, esquema y campos fijos. Dichos datos son las utilizadas en bases de datos relacionales, estos tienen piezas de información que se conocen previamente, con un formato y orden específico, estos son hojas de cálculo y archivos. “Formatos típicos son: fecha de nacimiento (DD, MM, AA); documento nacional de

³⁰ Joyanes, L. (2015). Sistemas de información en la empresa: el impacto de la nube, la movilidad y los medios sociales. Buenos Aires: Alfaomega.

identidad o pasaporte (por ejemplo, 8 dígitos y una letra); número de la cuenta corriente de un banco (20 dígitos), etcétera”³¹.

- Datos semi-estructurados: No están en bases de datos relacionales. Estos no tienen formato variable (no fijo) y flujo lógico, sin embargo, su comprensión no es simple. Tienen etiquetas y marcadores para separar sus elementos, y su lectura necesita reglas complejas para saber cómo proceder. Ejemplo de esto son los registros Web logs, que se compone de piezas de información, tales como documentos XML, HTML, JSON. También los datos almacenados en NoSQL son datos semi-estructurados³².
- Datos no estructurados: Estos no se pueden almacenar en estructuras de datos relacionales predefinidas. No tiene un formato específico, por lo que se almacenan como objetos sin estructura fija y sin control sobre ellos. No se puede analizar en una tabla ya que la información que tiene no se puede desmenuzar en datos básicos. Estos pueden ser datos de tipo texto como PDF o emails; o datos multimedia como audio y video, imágenes y fotografías ya que se clasifican en píxeles, mensajes SMS, libros, y mensajes de mensajería instantánea como Whatsapp. “Al menos, el 80% de la información de las organizaciones no residen en las bases de datos relacionales o archivos de datos, sino que se encuentran esparcidos a lo largo y ancho de la organización; todos estos datos se conocen como datos no estructurados”³³.

Hay un crecimiento exponencial de datos, en especial los no estructurados y semi-estructurados, que requieren herramientas diferentes a los tradicionales sistemas de bases de datos relacionales, por lo que su crecimiento ha provocado la creación de estas herramientas para manipularlos, tal es el caso de MapReduce, Hadoop o NoSQL. En el siguiente cuadro se muestra el aumento por año de la cantidad de datos estructurados vs. no estructurados.

³¹ *Ibíd.*

³² Joyanes, L. (2015). *Sistemas de información en la empresa: el impacto de la nube, la movilidad y los medios sociales*. Buenos Aires: Alfaomega.

³³ *Ibíd.*

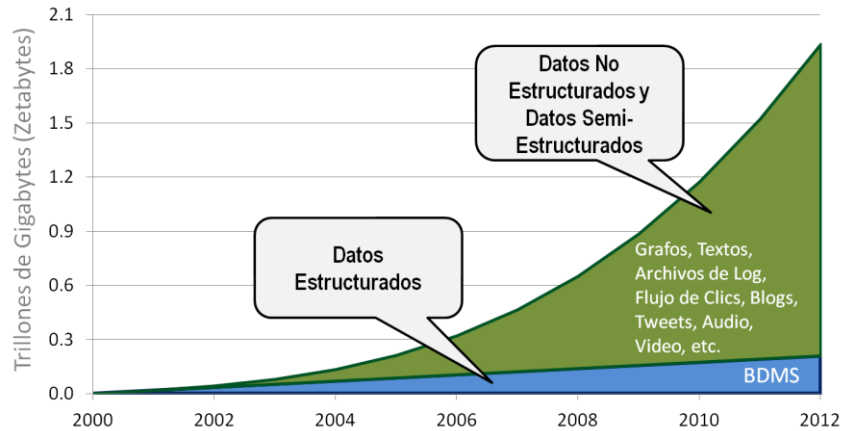


Figura 4: Oporto Diaz, Samuel (2014). Taller de introducción a la base de datos NoSQL. Kasperu. Recuperado de <http://www.kasperu.com/courses/BG011/II/default.html>

2.2.5 Características de Big Data: Volumen, variedad y velocidad

Cada día que pasa se crean más y más datos de forma masiva y sin detenerse, “Se crean 2,5 quintillones de bytes de datos en 24 horas, de forma que el 90% de los datos del mundo actual se han creado en los últimos dos años”³⁴, y “Según las previsiones de Gartner, en 2020 más de 25 mil millones de dispositivos estarán conectados a Internet, acrecentando un volumen de datos que a finales de 2013 ya se estimaba en 4,4 billones de GB y que llegará, según los pronósticos, a multiplicarse por 10 en tan solo 6 años”³⁵. Este aumento exponencial no solo impactará en el volumen de la información, ya que los datos tienen más de una característica, por lo que no se pueden tratar homogenizados y tratarlos a todos por igual, ya que estos datos proceden de todo tipo de sitios, tales como sensores climatológicos, redes sociales, transacciones, señales GPS, sitios de videos, etcétera. Estos datos generados no son todos iguales, ya que Big Data abarca tres dimensiones, conocidas en un modelo llamado las tres v, las que hacen referencia al volumen, velocidad y variedad.

³⁴ Lantares. Velocidad, variedad y volumen, las 3 magnitudes claves del Big Data. Lantares. Recuperado de <http://www.lantares.com/blog/velocidad-variedad-y-volumen-las-3-magnitudes-clave-de-big-data>

³⁵ *Ibíd.*



Figura 5: Calabuig, Oriol. (2004) ¿Qué es Big Data? Las entrañas de los datos. Portal comunicación. Recuperado de: http://portalcomunicacion.com/monograficos_det.asp?id=261

1. Volumen

Se habla de volúmenes Big Data cuando estos superan la capacidad de un software habitual. Esta magnitud se encuentra en constante movimiento por que los avances tecnológicos cada vez van a la par con el aumento de estos. Al decir grandes volúmenes de información, se hace referencia a petabytes o exabytes por ejemplo. Este concepto es bastante variable ya que cada día que pasa se van generando más datos y cambiando la concepción de lo que es grandes volúmenes. “Estamos pasando de la era del petabyte a la era del exabyte, y para 2015 a 2020, se espera entremos a la era del zettabyte. En el año 2000, se almacenaron en el mundo 800.000 petabytes. Se espera que el año 2020 se alcancen 35 zettabytes (ZB). Solo Twitter genera más de 7 terabytes (TB) de datos cada día. Facebook, 10 TB; y algunas empresas generan terabytes de datos cada hora de cada día del año”³⁶.

³⁶Joyanes, L. (2015). Sistemas de información en la empresa: el impacto de la nube, la movilidad y los medios sociales. Buenos Aires: Alfaomega.

Nombre	Simbolo	Potencias binarias y valores decimales
byte	b	$2^0 = 1$
Kbyte	KB	$2^{10} = 1\ 024$
Megabyte	MB	$2^{20} = 1\ 048\ 576$
Gigabyte	GB	$2^{30} = 1\ 073\ 741\ 824$
Terabyte	TB	$2^{40} = 1\ 099\ 511\ 627\ 776$
Petabyte	PB	$2^{50} = 1\ 125\ 899\ 906\ 842\ 624$
Exabyte	EB	$2^{60} = 1\ 152\ 921\ 504\ 606\ 846\ 976$
Zettabyte	ZB	$2^{70} = 1\ 180\ 591\ 620\ 717\ 411\ 303\ 424$
Yottabyte	YB	$2^{80} = 1\ 208\ 925\ 819\ 614\ 629\ 174\ 706\ 176$

Figura 6: Koday Games. (21 de diciembre de 2012). Unidades de medida informática. Koday Games. Recuperado de <http://kodaygames.blogspot.cl/2012/12/unidades-de-medida-en-informatica.html>

2. Variedad

Una de las grandes riquezas que poseen los datos es su variedad. Esto hace referencia a las distintas fuentes productoras de datos, estos pueden ser por ejemplo los producidos por las redes sociales, en los distintos dispositivos electrónicos donde se les puede hacer uso; a los sensores de movimiento o detectores de huella, de información creada por maquinas, etc. “La variedad representa todos los tipos de datos, y supone un desplazamiento fundamental en el análisis de requisitos desde los datos estructurados tradicionales hasta la inclusión de los datos en bruto, semi-estructurados y no estructurados como parte del proceso fundamental de la toma de decisiones.”³⁷

3. Velocidad

El concepto de velocidad se refiere a la rapidez que los datos se reciben y procesan, y luego cuando se toman las decisiones según los datos obtenidos. “A la mayoría de los sistemas tradicionales les es imposible analizar de forma inmediata los grandes volúmenes de datos que les llegan, sin embargo, incorporar el concepto de tiempo real es imprescindible para sistemas de detección del fraude o la realización de oferta personalizadas a

³⁷ Joyanes, L. (2015). Sistemas de información en la empresa: el impacto de la nube, la movilidad y los medios sociales. Buenos Aires: Alfaomega.

los clientes”³⁸. Los datos al producirse a gran velocidad necesitan un procesamiento y un análisis en tiempo real, esto quiere decir, al momento que son recibidos, y tomar decisiones al instante y mejorar estas sobre la base de la información generada. “La importancia de la velocidad de los datos se une a las características de volumen y variedad, de modo que la idea de velocidad no se asocia a la tarea de crecimiento de los depósitos o almacenes de datos, sino que se aplica la definición al concepto de los datos en movimiento, es decir, la velocidad a la cual fluyen los datos”³⁹.

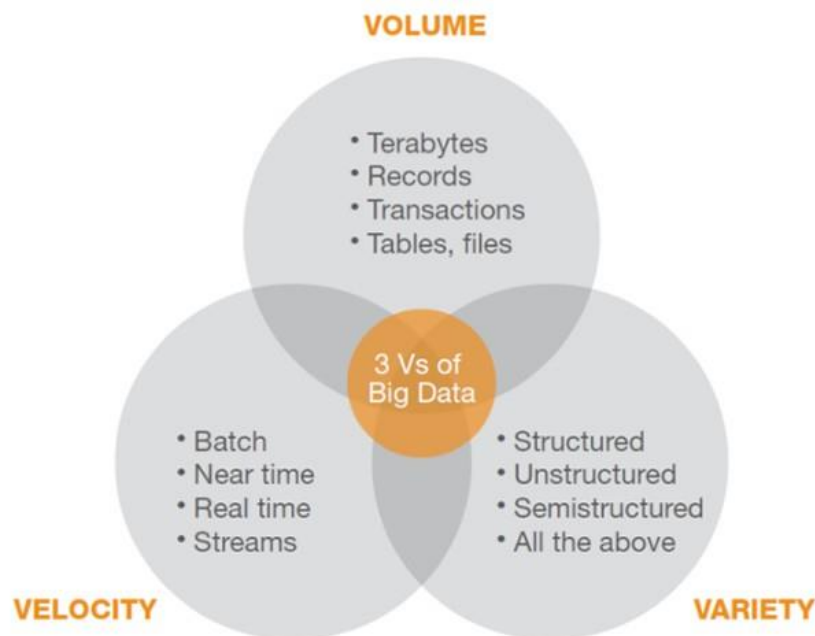


Figura 7: Walker, Michael (28 de noviembre de 2012). Data Veracity. Data Science Central. Recuperado de <http://www.datasciencecentral.com/profiles/blogs/data-veracity>

A estas tres principales magnitudes se han añadido dos más: La veracidad y el valor:

4. Veracidad

³⁸ López, José Carlos (27 de febrero de 2014). La moda del Big Data: ¿En qué consiste en realidad? El economista. Recuperado de <http://www.economista.es/tecnologia/noticias/5578707/02/14/La-moda-del-Big-Data-En-que-consiste-en-realidad.html>

³⁹ Joyanes, L. (2015). Sistemas de información en la empresa: el impacto de la nube, la movilidad y los medios sociales. Buenos Aires: Alfaomega.

Es la confianza en los datos, cuando se obtienen datos de calidad y fiabilidad, sacando fuera lo imprevisible antes de tomar una correcta toma de decisiones. “Según IBM uno de cada tres líderes de negocio (directivos) no se fía de las informaciones que se utilizan para tomar decisiones”⁴⁰, no se puede tomar decisiones plenas si no se confía en los datos presentados, por lo que la veracidad en los datos es un reto que se debe afrontar a medida que la cantidad de información crece.

5. Valor

Por último, encontramos el valor, donde se ve cual es la importancia del dato para la finalidad que se desee, saber que datos analizar de manera rentable y eficiente para las organizaciones. Aquí tecnologías como Hadoop, juega un papel importante ya que puede procesar millares de volúmenes de datos de forma eficaz y económica.

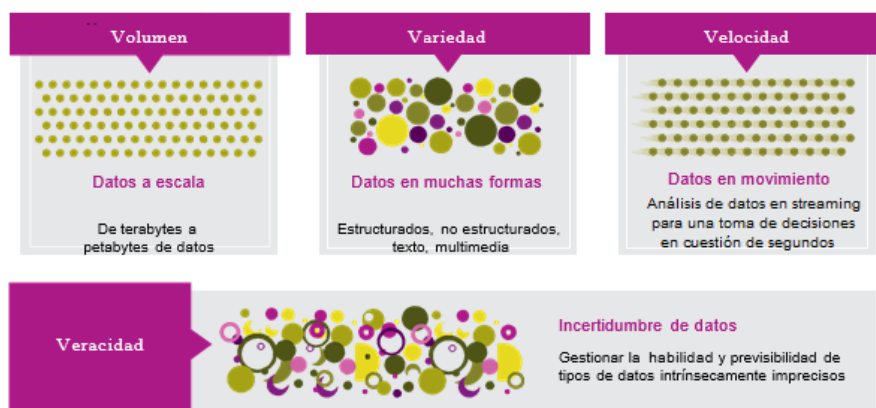


Figura 8: IBM Institute for Business Value (2002) Analytics: el uso de big data en el mundo real Cómo las empresas más innovadoras extraen valor de datos inciertos. IBM. Recuperado de http://www-05.ibm.com/services/es/gbs/consulting/pdf/El_uso_de_Big_Data_en_el_mundo_real.pdf

2.2.6 Fuentes de Big Data

Las fuentes de Big Data son las unidades productoras de estos y se clasifican en diferentes categorías, y cada una tiene otras variadas fuentes. Estas son según la clasificación presentada por Luis Joyanes, y expuesta en el siguiente esquema de Sunil Soares

⁴⁰ Joyanes, L. (2015). Sistemas de información en la empresa: el impacto de la nube, la movilidad y los medios sociales. Buenos Aires: Alfaomega.

Tipos de Big Data



Figura 9: Soares, Sunil (3 de junio de 2012) Not Your Type? Big Data Matchmaker On Five Data Types You Need To Explore Today. Dataversity. Recuperado de http://www.dataversity.net/not-your-type-big-data-matchmaker-on-five-data-types-you-need-to-explore-today/?cm_mc_uid=69942747005014585768846&cm_mc_sid_50200000=1465949454

- **Web y Social Media:** Incluye contenido web y los datos obtenidos de medios sociales. Los datos en esta categoría se capturan o almacenan según algunas características, tales como los datos procedentes de flujos de clics o *feeds* en el caso de Twitter o Tumblr; entradas o *posting* en Facebook; según sistemas de gestión de contenidos como YouTube; y sitios de almacenamiento de información como Dropbox. Otros tipos de medios sociales son wikis como MediaWiki y Wikipedia; marcadores sociales como Del.icio.us; agregadores de contenido como Digg y Meneame

- Máquina a máquina (M2M)/ Internet de las cosas: Hace referencia a los datos creados por máquinas, y las tecnologías que permiten conectarse con otros dispositivos entre sí. M2M hace uso de los dispositivos como sensores o captadores de eventos, tales como humedad, velocidad, temperatura, variables meteorológicas, etcétera, y estas se transmiten de variados medios (cable o señales móviles) a otro dispositivo que los almacena o analiza, y creando información significativa para los usuarios del sistema. Algunos dispositivos de captura de datos chips NFC; medidores inteligentes de temperatura, electricidad, presión; sensores; dispositivos GPS. De esta comunicación M2M surge el internet de las cosas, que se refiere a los millones de dispositivos que se comunican entre sí y que acceden por sí mismos a internet si es necesario.
- Transacciones de grandes datos: Son grandes datos de origen transaccional que vienen de operaciones de todo tipo, tales como facturaciones en telecomunicaciones, registros detallados, transacciones bancarias. Uno al mover dinero en un sistema bancario o telecomunicacional se generan datos de usuario, fecha hora y montos entre otros metadatos.
- Biometría: O reconocimiento biométrico, provienen de datos de seguridad, tales como lectores o escaneos de retina e iris, o huellas dactilares, reconocimiento facial, genética y ADN. Estos se refiere a la identificación automática de un individuo basándose en sus características anatómicas y genéticas.
- Creados por personas: Las personas generan gran variedad de datos, tales como llamadas telefónicas, notas de voz, e-mails, documentos electrónicos, estudios, registros médicos, documentos papel, faxes, ingresar manualmente información a una base de datos o a una hoja de cálculo, etcétera. Estos datos pueden contener información sensible que necesitara un tratamiento especial en cuanto a su privacidad, ya sea cifrándola u ocultándola, por lo que estos datos son protegidos por leyes nacionales y supranacionales relativas a la protección de datos y privacidad⁴¹.

⁴¹ Joyanes, L. (2015). Sistemas de información en la empresa: el impacto de la nube, la movilidad y los medios sociales. Buenos Aires: Alfaomega

Vale mencionar que todos los datos procedentes de estas fuentes pueden ser de tres tipos, mencionados anteriormente, ya sean estructurados, no estructurados y semi-estructurados.

2.2.7 Bases de datos y tecnologías

Según lo expuesto anteriormente, los datos en Big Data pueden ser de tres tipos: estructurados, no estructurados y semi-estructurados. Cada dato, más allá de su tipo, tiene características o dimensiones, como volumen, velocidad y variedad los que ofrecen puntos de vistas para comprender la naturaleza de los datos y para saber que plataformas usar para explotarlos. Los datos tienen distintas fuentes según su unidad creadora, ya sean en medios sociales, M2M, transacciones, biometría y generados por humanos. Comprendido ya lo anterior, las preguntas que surgen son ¿Cómo trabajar y donde se almacenan estos datos? ¿Qué plataforma se usa y como analizar los Big Data?

Los datos en las empresas y organizaciones almacenan su gran cantidad de datos en bases de datos, tales como las relacionales, NoSQL, in-memory así como las heredadas.

- SQL: Son las bases de datos tradicionales. Las plataformas relacionales siguen siendo utilizadas en las organizaciones y sus sistemas computacionales. Algunas de estas son Oracle Database, IBM DB2, Microsoft SQL Server, etcétera.
- NoSQL (Not only SQL): Estas son base de datos no relacionales, no utilizan SQL como lenguaje principal. Son de alto rendimiento y escalables.
- In-memory: Estas bases de datos hacen sus procesos en la memoria principal utilizándola para almacenaje de datos. Tienen una gran velocidad de proceso, ya que a diferencia de los sistemas tradicionales que guardan su información en discos, estos lo hacen en su memoria principal.
- Legacy (heredadas): Se le llama así cuando las organizaciones siguen ocupando sus bases de datos antiguas, aunque ahora funcionando integradas con bases de datos relacionales y analíticas.
- Hadoop: Apache Hadoop es un framework de código abierto que es capaz de procesar grandes cantidades de datos a través de muchas computadoras en clústers de servidores básicos. Esta plataforma permite

que “las empresas pueden explorar datos complejos mediante el análisis personalizado adaptado a sus datos y necesidades”⁴².

Este es un proyecto que nace en conjunto de Google y Yahoo!, los que trabajaron con grandes cantidades de datos. “Está diseñado para extender un sistema de servidor único a miles de máquinas, con un muy alto grado de tolerancia a las fallas. En lugar de depender del hardware de alta gama, la fortaleza de estos clústers se debe a la capacidad que tiene el software para detectar y manejar fallas al nivel de las aplicaciones”⁴³.

Este sistema tiene varias características tales como la de ser redimensionarle, es decir, que se le pueden ir agregando nuevos nodos cuantas veces el usuario estime conveniente. Es rentable, por que incorpora el uso de servidores básicos, y de este modo reduce el costo de terabyte de almacenamiento. También es flexible, ya que funciona sin esquema y asimila cualquier dato sea estructurado o no y de cualquier fuente.

2.2.8 Analítica de Big Data

El crecimiento de los datos en las organizaciones crece a niveles exponenciales, por lo que las organizaciones deben explotarlas de forma eficiente. El 80% de estos datos son del tipo no estructurados, y estos datos muchas veces son difíciles de analizar o pueden tardar mucho tiempo sin las herramientas y formación adecuadas. La analítica Big Data permite a los usuarios de los datos masivos, analizarlos de terabyte a petabyte, de forma eficaz, y el poder explorarlos y visualizarlos mediante graficas o cuadros y en cuadros en tiempo real cuando el usuario lo estime. Por lo tanto “El análisis de Big Data debe ayudar a tomar mejores decisiones y evaluar las medidas que se han de tomar del modo más eficiente y rentable posible”⁴⁴

⁴² Rodríguez Sánchez, Francisco (2014) Herramientas para Big Data: Entorno Hadoop. Escuela técnica superior de ingeniería de telecomunicaciones Universidad politécnica de Cartagena. Cartagena, España. Recuperado de <http://repositorio.upct.es/bitstream/handle/10317/4402/tfg482.pdf?sequence=1&isAllowed=y>

⁴³ IBM. ¿Qué es Hadoop?. IBM. Recuperado de: <http://www-01.ibm.com/software/cl/data/infosphere/hadoop/que-es.html>

⁴⁴ Joyanes, L. (2015). Sistemas de información en la empresa: el impacto de la nube, la movilidad y los medios sociales. Buenos Aires: Alfaomega

El análisis de la Big Data implica examinar a gran velocidad grandes volúmenes de datos y así conocer correlaciones e información útil, y con estos resultados obtener ventaja competitiva para nuestra organización. Es necesario organizar, refinar y convertir estos datos en información relevante para los fines que se tengan. “La información en crudo tiene solo valor potencial, es su análisis y sistematización lo que permite incrementar la capacidad de innovar de las organizaciones”⁴⁵. El tratamiento de grandes cantidades de datos se realiza en las siguientes etapas:

- Adquisición: Los datos proceden de distintas partes, tales como de datos tradicionales (en base de datos relacionales por ejemplo), así como de fuentes de datos no estructurados que se pueden almacenar en NoSQL por ejemplo.
- Organización de la información: Es preparar y tratar la información para obtener buenos resultados.
- Análisis: Analizar toda la información recabada con herramientas estadísticas avanzadas, utilizando la minería de datos por ejemplo.
- Decisión: Es tomar las decisiones acertadas en tiempo real y lo más rápido posible en base a los resultados obtenidos del análisis. De este modo convertir los datos brutos en conocimiento para ser aplicados en tableros de control y herramientas de visualización.

2.3 Unidades de Información

Una unidad de servicio y/o producto de información, es una entidad dedicada al tratamiento y recuperación del material en distinto soporte o léxico especializado, y es parte del Sistema de Información Documental (SID), el cual abarca un conjunto de subsistemas: Sistema de clasificación (Dewey, Biblioteca del Congreso, etc.), Sistema de ordenación (numéricos, cronológicos y simples), alfabéticos (onomásticos, geográficos, temáticos) y mixtos (alfanuméricos), etc.; encargados de funciones distintas (análisis, almacenamiento, distribución, etc.), pero interrelacionadas entre sí, cualquier modificación en uno de sus componentes también implica en los demás. Se basa en la Teoría General de Sistemas, cuyo modelo de análisis es aplicable a distintas ramas del saber humano y se diseñan en función de los objetivos específicos, con la convicción de satisfacer las necesidades de información y

⁴⁵ Joyanes, L. (2015). Sistemas de información en la empresa: el impacto de la nube, la movilidad y los medios sociales. Buenos Aires: Alfaomega

actualizaciones de conocimientos de los usuarios con documentos, procesos y materiales (incluso didácticos) que permite una mejor formación del ser humano, asimismo prospera la calidad del gobierno; a través de calidad de asesoramiento a los expertos.⁴⁶

Los procesos de organización se realizan para facilitar el acceso de manera progresiva a los contenidos informativos ya sea presencial o en línea, aprovechando los avances de las Tecnologías de información y Comunicación (TIC). Toman como desafíos el estudio de diversos programas, proyectos e iniciativas relacionadas con la organización y servicios de información, así como la infraestructura tecnológica y recursos disponibles para mejorar la calidad de sus distintas dependencias, instrumentos indispensables para la investigación y docencia, permite el flujo de conocimiento en el entorno. Suele ser el sector público el que lidera el desarrollo y estrategias, dirigidas a su disposición, aunque también surgen iniciativas del sector privado.

Estos servicios abarcan los desafíos (como utilidad, oportunidad, tiempo y esfuerzo) intrínseco de la pirámide informacional:

⁴⁶ Calderín Cruz, Mabel. (2007). Sistema de información documental en la Academia: Situación actual y perspectivas futuras, caso UCAB. Recuperado de http://www.cienciarred.com.ar/ra/usr/3/333/n6_v1pp51_76.pdf



Figura 10: Creación propia basada en “Teoría de la decisión y el aprendizaje organizacional (25 de abril de 2012). Aspectos del nuevo modelo organizacional. Recuperado de <https://chtedeapor20121908558.wordpress.com/category/bitacoras/>

1. Dato: elementos discontinuos o registros icónicos, simbólicos (fonémicos o numéricos) o sígnicos simbólicos (lingüísticos, lógicos, matemáticos), por medio de los cuales se representan hechos, conceptos o instrucciones.

2. Información: es el orden de los datos o materia relacionada, que se procesa de manera potencialmente significativa, adquiriendo valor informacional cuando los elementos se identifican.

Según la norma UNE 50-113-92/1. Documentación e información. Vocabulario. Parte 1:

Información: Mensaje utilizado para representar un hecho o un concepto en un proceso de comunicación, con el fin de aumentar los conocimientos.⁴⁷

3. Comunicación: Transmisión de señales mediante un código común, entre emisor y receptor.

4. Conocimiento: es la información como comprensión (información relevante). Es la información que puede ser entendida por cualquier persona

⁴⁷ Castillo, Lourdes (2001) Introducción a la información científica y técnica. Recuperado de <http://www.uv.es/macass/1.pdf>

que pueda leer el código lingüístico. Pero solo podrá ser comprendida por aquel con una base memorizada de conceptos, valoraciones, experiencias.

5. Sabiduría: es la información como oportunidad. Es decir, es la evaluación e internalización del conocimiento, que siendo contextualmente relevante permite la intervención ventajosa de la realidad.

Además se agrega valor en cada procedimiento, como lo demuestra la figura a continuación:



Figura 11: Del dato a la inteligencia. Actaf. Recuperado de http://www.actaf.co.cu/index.php?option=com_mtree&task=att_download&link_id=338&cf_id=24

Las unidades de información son atendidas por especialistas y se diferencian de acuerdo a las necesidades de los clientes que atiende, lo que conlleva a diferencias sustanciales en toda su estructura según la disciplina correspondiente.



Figura 12: Jaramillo, Johana. (27 de mayo de 2012). Unidades de información y la pirámide informacional. Recuperado de <http://es.slideshare.net/jaramillojohana/unidades-de-informacion-13094582>

Para conseguir los objetivos de este trabajo se abordaron tres unidades de información: Biblioteca, Archivo y Centro de Documentación.

Unidad de Información	Biblioteca	Archivo	Centro documentación
Fondo material	<ul style="list-style-type: none"> - Colección de materiales con distintos soportes (libros, revistas, periódicos, etc.) - Editoriales 	<ul style="list-style-type: none"> - Documentos que dan certeza de un hecho - Surgen por la actividad de una persona o entidad. 	<ul style="list-style-type: none"> - Literatura gris (separatas, informes, tesis, boletines, etc.) - Bases de datos - Publicaciones periódicas. - Documentos elaborados con la información extraída del doc. original.
Usuarios	<ul style="list-style-type: none"> - Incluye todo tipo de cliente activos (que buscan información) sin 	<ul style="list-style-type: none"> - Principalmente investigador. 	<ul style="list-style-type: none"> - Cliente pasivo: el documentalista ofrece la información al

	distinción de edad, sexo o condición.		usuario.
Servicios	<ul style="list-style-type: none"> - Consultas en sala - Préstamo - Servicios de información bibliográfica y referencia - Formación de usuarios - Extensión bibliotecaria - Actividades culturales - Animación a la lectura 	<ul style="list-style-type: none"> - La documentación está a disposición tanto de la empresa que la genera y como al resto de los usuarios. - No hay prestamos - Consulta en sala 	<ul style="list-style-type: none"> - Acceso a base de datos especializadas - Préstamo y reproducción (fotocopias) - Servicios de referencia
Organización y descripción	<ul style="list-style-type: none"> - La descripción del material es más superficial (título, editorial, fecha de edición, etc.) 	<ul style="list-style-type: none"> - Recoge el tipo de documento que es y la información que contiene (tipo y contenido) 	<ul style="list-style-type: none"> - Se resalta el contenido: incluye un resumen del documento.

Figura 13: Silva, Martha. (2 Julio-Diciembre de 1999). Gestión por procesos en las unidades de información. Recuperado de <http://tesis.udea.edu.co/bitstream/10495/1947/1/Gesti%C3%B3n%20por%20procesos%20en%20las%20unidades%20de%20informaci%C3%B3n.pdf>

2.3.1 Biblioteca

El término “biblioteca” procede del latín *bibliothēca* y éste, a su vez, de los vocablos griegos *biblion* (libro) y *theke* (caja),⁴⁸ abarca la formación de las colecciones bibliográficas (libros y publicaciones en serie) y audiovisuales, debidamente clasificados y ordenados, para una eficaz recuperación disponible a la consulta del público (amantes de la lectura, estudiantes, investigadores), el cual de forma autónoma es capaz de acceder y localizar la información.

⁴⁸ Concepto de biblioteca. Tipos y función. Perspectivas de futuro. Recuperado de <http://www.bibliopos.es/Biblion-A2-Biblioteconomia/01Concepto.pdf>

Todas las bibliotecas nos ofrecen recursos para la investigación o estudio de un trabajo y resolver sus respectivos problemas, y sobre todo satisfacer las necesidades de información, disfrutando de la lectura; ya que en esta se encuentra material de diversas áreas de conocimiento y cuenta con espacio disponibles al público para consultar o tomar en préstamo el material; estos son presenciales (salas amuebladas) o en línea (a través de dispositivos tecnológicos), potenciando en Internet la promoción de la colección; para ambos casos se encuentra a disposición el personal de la biblioteca (en esencia un bibliotecario).

Como se menciona anteriormente es un espacio de estudio, útil para todos, el cual se debe respetar (cuidarlo y conservarlo) y guardar silencio o hablar en voz baja para no molestar el trabajo de los demás. Las grandes bibliotecas cuentan también con departamentos especiales como hemeroteca (lugar donde se conservan y se prestan publicaciones periódicas), además, cuentan con salas o dependencias en las que el público puede tener acceso a fotografías, videos, discos musicales, discos de computación, Cd, etc. para su clasificación cuenta con fichas calcográficas, que actualmente se encuentran en formato apto para las computadoras, facilitando el flujo de los datos, esta ficha contiene el nombre del autor, título, y otros datos del material (páginas, editorial, materia, etc.), además de la cota o código que la biblioteca asigna al libro para la eficaz localización.

La norma ISO 2789-1991 (Norma UNE-EN ISO 2789) sobre estadísticas internacionales de bibliotecas, la define como: "Organización o parte de ella cuya principal función consiste en mantener una colección y facilitar, mediante los servicios del personal, el uso de los documentos necesarios para satisfacer las necesidades de información, de investigación, de educación y ocio de sus lectores.

Estas se pueden dividir dependiendo de qué fondos tengan y a quienes estén dirigidas, pueden ser:

Escolares: ejecutadas para apoyar el proceso de enseñanza, aprendizaje y desarrollo lector (ludoteca y literatura básica), son utilizadas por los alumnos (niños y jóvenes) y profesores.

Universitarias: apoyan el proceso de enseñanza, aprendizaje e investigación, utilizados por estudiantes y profesores universitarios, se nutren de textos académicos seleccionados de las diversas áreas de conocimiento.

Especializada: encargadas de trascender el conocimiento científico, concurren a textos específicos, estudiantes especialistas en el área.

Pública o popular: su origen es comunitario y organizacional, ciudadanos dispuestos a promover y recuperar la memoria, con colecciones culturales y recreativas.

2.3.2 Archivo

Según la escala evolutiva de la información es cimentada en el dato; palabra originaria del latín datum (“lo que se da”), le sigue el documento (contenido y soporte) y el documento de archivo (contenido, soporte y contexto), a diferencia, este nace por una razón, es un testimonio que permite llegar al conocimiento. A través de la aparición de los primeros testimonios escritos, se conoce la base de lo que hoy conocemos como “archivo”, palabra procedente del latín archivum, y éste del griego archeión que significa “principio u origen”, se encarga de los fondos documentales, producidos por las acciones o actividades inevitables de una entidad, tienen función informativa y probatoria, considerándola auténtica y única⁴⁹.

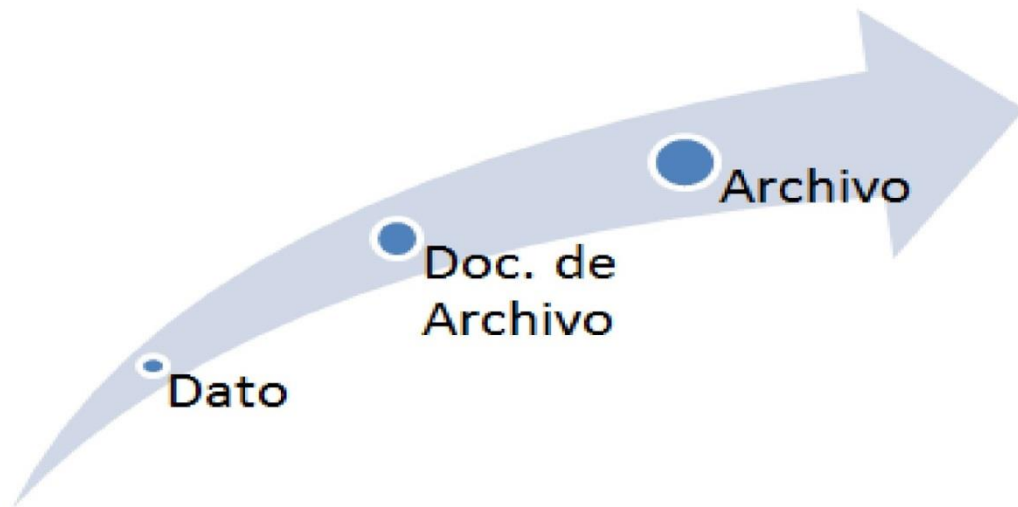


Figura 14: Creación propia

Hablar de este es hablar de la historia del hombre. Con el paso del tiempo, crece el interés por preservar documentos de origen histórico, produciendo

⁴⁹ Martín Gavilán, César (2009) Temas de Biblioteconomía: Concepto y función de archivo clases de archivos El sistema archivístico Español.
Recuperado de <http://eprints.rclis.org/14058/1/sisarchivesp.pdf>

un esfuerzo serio para lograr este fin. En la Edad Moderna se aprecia en España Carlos I quien salvaguarda numerosos documentos, en la Fortaleza de Simancas a cargo de un archivero. A fines del siglo XVIII se produce la apertura de los archivos a los ciudadanos. La Edad Contemporánea delimita la formulación de una legislación y sistemas de archivos modernos, que demandan: mayor número de usuarios, diversidad de soportes, expansión y diversificación de entidades productoras y multiplicación de la producción documental. Para lo cual, se aplica la disciplina correspondiente a esta unidad de información, la archivística. Según definiciones del Consejo Internacional de Archivos esta disciplina se “encarga del estudio teórico y práctico de los principios, procedimientos y problemas relativos a las funciones de los archivos.

- **Archivo (fondo documental)**

Si bien se considera una palabra polisémica, archivo es un “fondo documental” de diverso soporte que surge como registro de funciones producidas por personas naturales o jurídicas. La organización viene impuesta por la propia documentación el encargado del archivo la compone y disminuye las pérdidas, favorece la conservación y mejora la búsqueda de información. El archivo será eficaz cuando se encuentre rápidamente lo que se busca.

Las funciones de los archivos, se agrupan en tres bloques:



Figura 15: Creación propia.

2.3.3. Centro de documentación

La Documentación es una ciencia que utilizan métodos y técnicas de otras disciplinas como la estadísticas, psicología, biblioteconomía, archivística, etc. para cumplir sus objetivos, como fin último conseguir que la información llegue al usuario. Apoya la investigación científica y el desarrollo tecnológico, está compuesta principalmente por material gris, bases de datos de información científica y técnica, asisten investigadores y especialistas, la documentación científica es clave tanto para la investigación básica, como la aplicada.

Dos grandes objetivos:

1. Análisis científico: producción y consumo, así como la estructura y propiedades de la información.
2. Recogida: procesamiento analítico sintético, almacenamiento, recuperación y difusión de dicha información.
3. Suministrar: difundir la información, previamente transformada, facilitando documentos concretos el usuario pida.

Para Álvarez-Ossorio el centro de documentación (o información) es el elemento intermedio necesario para reunir toda la información disponible, en el tema de su interés, organizarla, tratarla y suministrarla a sus usuarios en forma directamente utilizable. En el centro de documentación se realizan todas las tareas de la cadena documental: seleccionar, recoger, analizar, almacenar, recuperar y difundir la información.⁵⁰

Los Sistemas de Información (SI) son considerados un componente vital para el desarrollo de cualquier organización. Según la “cadena de utilidad del servicio” desarrollada por investigadores de la Universidad de Harvard, Heskett y Schlesinger, quienes plantean que existe una relación directa entre las experiencias de servicio, la lealtad del cliente y rendimiento financiero.⁵¹ Busca la excelencia en la formación e investigación, enfocándose al entregar el servicio en sus pilares: colaboradores (empleados y accionistas) y

⁵⁰ Castillo, Lourdes (2004) Los centros de documentación.

Recuperado de <http://www.uv.es/macass/T1.pdf>

⁵¹ Heskett, L [et. al.] Poniendo a trabajar la caden servicio-ganancia.

consumidor, lo que produce la salida de personas informadas o con unas necesidades de información satisfechas. Abarcando los distintos niveles de gestión: alta dirección (se ocupa de la planificación y tareas de política y representación), mandos intermedios (concretan estrategias generales: programas y proyectos, elaboran informes para los mandos, coordinan los diferentes departamentos) y los gestores de “primera línea” (ejecutan programas y proyectos, además cotidianamente administran los recursos para conseguir objetivos a corto plazo).

Los centros de documentación fusionan disciplinas, instaurando una autentica gestión de los conjuntos de conocimientos y capacidades necesarias para la organización, registro, almacenamiento, recuperación y difusión de la información (manual o automatizada), los cuales son elementos necesarios para potenciar y proceder al registro (se centraliza la entra y salida de los documentos), objetivos, términos asociados a la gestión (índice, clasificación, descripción, etc.), consultas al catálogo bibliográfico y fondo documental.

Capítulo 3: Metodología

Como ya se ha mencionado, las Unidades de Información generan una gran cantidad de datos que se van acumulando diariamente como consecuencia de los procesos y servicios que en ellas se ofrecen y practican. La principal utilidad de esta metodología será saber qué hacer con aquellos datos generados y aprovechar los beneficios que estos nos pueden otorgar para el desarrollo y/o mejoramiento de las Unidades de Información.

Ben Shower dijo que “Este es un momento muy interesante para las Bibliotecas, Museos y el sector del Patrimonio en general. Hay algunos trabajos realmente innovadores realizados por bibliotecarios, conservadores, archiveros e informáticos para utilizar el potencial de los datos que están fluyendo a través de sus sistemas y servicios, en la captura y análisis para poder mejorar los servicios y satisfacer las necesidades del usuario”⁵²

3.1 ICAV como metodología base

Para efectos de este trabajo, desarrollaremos nuestra propuesta en base a la Metodología ICAV (identificar, consolidar, analizar y visualizar) utilizada comúnmente en el mundo empresarial. Se seleccionó esta metodología entre un abanico de opciones como: Big Data Consulting Methodology, CBIG Framework y Metodología Compañía Business Rio. Esta preferencia se optó principalmente porque se lleva a cabo de forma sintetizada en solo 4 pasos “desde la identificación de la necesidad hasta la visualización de resultados, pasando por un proceso de análisis e interpretación de información”⁵³. Debido a esto es que se puede realizar un plan de aplicación de utilización del Big Data de forma exitosa de principio a fin y de forma relativamente sencilla, para la comprensión de todos los sectores de la organización, sin importar su conocimiento previo en el tema. Aplicaremos esta metodología

⁵² Showers, Ben (2015) Library analytics and metrics: Using data to drive decisions and services. Recuperado de <http://www.facetpublishing.co.uk/title.php?id=049658#table-of-contents-tab>

⁵³ Ardila Cañas, Esneyder; Gómez Díaz, Isabel (2014) Estrategias para la gestión de grandes volúmenes de datos por medio de Big Data en el contexto de la analítica de negocios: Caso MVM ingeniería de software S.A.S. Recuperado de http://bibliotecadigital.usbcali.edu.co/jspui/bitstream/10819/2702/1/Gestion_Volumen_Datos_Ardila_2014.pdf

en Unidades de Información y con esto daremos inicio a la propuesta de metodología.



Figura 16: Big Data. Recuperado en: <http://www.bigdata.pe/web/index.php/metodologia>

1. Identificar: Antes de iniciar cualquier proyecto de Big Data se requiere identificar claramente las necesidades, tanto las necesidades internas (del personal a cargo) como las externas relacionadas directamente con el usuario de nuestra Unidad de Información. Se deben generar preguntas en base a necesidades y exponerlas para que estas puedan ser respondidas.

2. Consolidar: Una vez identificadas las necesidades de la Unidad de Información, se debe ubicar donde están las fuentes de datos requeridas que nos permitirán responder las preguntas anteriormente mencionadas. Estas fuentes de datos serán definidas según la Unidad de Información a la que se aplique y es importante considerar todas las fuentes de datos ya sean estructurados, así como también los semi-estructurados y no estructurados, los que ocupan más de las tres cuartas partes del total de la información de las organizaciones, y su utilización significa hacer más rico el análisis. Éstas pueden ser por ejemplo:

- Comentarios de redes sociales
- Textos
- Videos e imágenes

- Web logs
- Información de sensores
- Audio

Una vez identificadas las fuentes, estas se deben agrupar en un repositorio común donde sean clasificadas en áreas según la necesidad a responder.

3. Análisis: Una vez que se tiene la información consolidada en un repositorio común, se puede empezar a realizar el análisis de grandes volúmenes de información utilizando técnicas avanzadas de análisis predictivo y minería de datos. Es muy importante tener claro que esta labor debe estar a cargo de especialistas que determinarán cuáles son los procedimientos para analizar los datos y que herramientas de análisis estadísticos son los adecuados, así como los niveles de medición de las variables, etc.

Este proceso consta de seleccionar, explorar, y modelar grandes volúmenes de datos que develen información previamente desconocida, nuevas correlaciones, patrones y tendencias significativas, filtrando datos, utilizando técnicas de reconocimiento de comportamiento y aplicando métodos de predicción basados en el análisis de los datos que posteriormente serán representados estadísticamente.

4. Visualizar: Dar a conocer los resultados y proceder a la distribución y diseminación de esa información a todo el personal involucrado para tomar acciones que permitan dar respuestas a las preguntas realizadas en la fase 1 (Identificar).

3.2 Propuesta de metodología

Lo primero que debemos considerar es reunir a los responsables de cada área de la organización para dar a conocer el proyecto de implementación de Big Data y comenzar los siguientes pasos, para esto ofreceremos un cuadro resumen en donde se podrán apreciar las principales actividades y etapas de la metodología a aplicar.

Metodología	Etapas	Actividades
ICAV	Identificar	Identificar claramente las necesidades del negocio
		Identificar usuarios finales
		Conocer las preguntas que requieren responder para tomar mejores decisiones empresariales
	Consolidar	Ubicar donde están las fuentes de información requeridas para responder las preguntas del negocio
		Identificar que fuentes de información no estructuradas se requieren para hacer más rico el análisis
		Consolidar información e repositorio común donde se acumulen las agregaciones y transformaciones necesarias para poder responder de la mejor manera las respuestas del negocio
	Analizar	Realizar el análisis de grandes volúmenes de información utilizando técnicas avanzadas de análisis predictivo y minería de datos
	Visualizar	Diseminar esa información a los usuarios responsables para que puedan visualizar la información para la toma de decisiones empresariales

Figura 17: Ardila Cañas, Esneyder; Gómez Díaz, Isabel (2014) Estrategias para la gestión de grandes volúmenes de datos por medio de Big Data en el contexto de la analítica de negocios: Caso MVM ingeniería de software S.A.S. Recuperado de http://bibliotecadigital.usbcali.edu.co/jspui/bitstream/10819/2702/1/Gestion_Volumen_Datos_Ardila_2014.pdf

3.2.1 Identificar

3.2.1.1 Identificar las necesidades de la organización

Es necesario este procedimiento porque:

- a) Es necesario comprender a cabalidad la organización y su funcionamiento
- b) Se deben tener claro cuáles son los problemas de la organización mientras se describe los recursos y limitaciones de las áreas.
- c) Plantear las necesidades de cada área otorga la base para la creación de nuevo conocimiento a través del control de datos.
- d) Una vez se hayan evaluado las necesidades, se pueden tomar mejores decisiones en cuanto a las prioridades para implantar un sistema.

- Se debe identificar las necesidades y requisitos mediante entrevistas, estableciendo un diálogo entre entrevistador, jefe de área y empleados si se considera necesario para así recabar la información base para iniciar la metodología.

- Se debe determinar las necesidades para que la organización cumpla con crear y mantener los datos de forma correcta y para así evidenciar las actividades de la organización.

- Se deben analizar las necesidades de la organización, así como las obligaciones legales y rendición de cuentas entre partes internas (directivos y empleados) y partes externas (clientes, accionistas, etc.)

Algunas herramientas para recabar esta información pueden ser:

- a) Realizar un análisis FODA, lo que significa conocer las fortalezas, oportunidades, debilidades y amenazas de la organización. Esta es una metodología de estudio del estado de la organización, analizando características internas y su situación externa.
- b) Revisar o elaborar la misión y visión de la organización. Misión es la razón de ser de la organización enfocándose en su actividad; por otro lado, visión es la imagen que proyecta a largo plazo para su expectativa ideal.
- c) Elaborar un modelo de negocios, que es un esquema sencillo de 9 pasos, que reflejan la lógica de una organización para conseguir

ingresos. Los 9 módulos se agrupan en cuatro áreas principales: Clientes, oferta, infraestructura y viabilidad económica.

3.2.1.2 Identificar las necesidades internas de la organización

- **Diagnóstico**

Para mayor profundidad en esta etapa, se recomienda realizar un diagnóstico actual de la organización.

Se debe comprender el contexto de la organización y la actividad a la que se dedica, para saber los factores que inciden en cuanto a la creación de los datos y la mantención y análisis de estos. Se debe tener conocimiento del fin de la organización, la estructura, el entorno legal, normativo, económico y político. Además de conocer las debilidades y fortalezas en cuanto a la gestión de la información dentro de la organización.

Se pueden utilizar varias herramientas para realizar un diagnóstico en la organización

Herramienta	Descripción
Diagrama de afinidad	Es esencialmente un método de intuición, implica generar ideas por inspiración súbita y luego agruparla por temas afines.
Diagrama de relaciones	Es un método de inducción lógica que permite aclarar las causas y sus relaciones para identificar, confirmar y seleccionar las causas originales más importantes que afectan a un problema en análisis.
Diagrama matricial	Es un método para garantizar datos verbales con la finalidad de establecer conclusiones para resolver o prevenir problemas, a través de relacionar diferentes factores o elementos de ciertos avances.
Análisis matricial de variaciones	Método cuya finalidad es identificar y seleccionar causas potenciales para prevenir problemas o asegurar resultados de un proceso de sistema. Consiste en relacionar las diferentes variaciones de un proceso o sistema.

Diagrama de árbol	Método para definir los medios para lograr una meta y objetivo final. Implica desarrollar un objetivo en una serie de medios en multietapas: medios primarios, secundarios, etc. Y acciones específicas.
Graficas de procesos de decisiones programadas	Es un método de lógica que es utilizado para predecir el futuro, enfatizando en las situaciones no deseadas durante la realización de un evento, para diseñarlo y dirigirlo hacia un resultado deseable.
Diagrama de flechas	Es indispensable para proyectos de largo plazo, permite analizar el progreso del proyecto de acuerdo a su programa para mejorar o reducir el tiempo total y así poder optimizar el trabajo.

Figura 18: Sánchez, Jesús (2011) Método de diagnóstico. Recuperado de <http://es.slideshare.net/jesussanval/metodo-de-diagnostico>

Para hacer un diagnóstico funcional y organizacional, hay diversos métodos y técnicas:

- a) Entrevistas: Se complementa con un cuestionario, y puede recabar información valiosa y detallada en una conversación con miembros de la organización. Es importante seleccionar el criterio de selección de los entrevistados, según la finalidad de la entrevista.
- b) Cuestionario: Permite recoger información de más gente, de manera más rápida y económica y da la facilidad para análisis estadísticos.
- c) Datos secundarios: Son todos los datos disponibles en la organización de forma estadística, informes, estados financieros, etc.
- d) Análisis de redes de comunicación: Se estudia la comunicación en la organización y la efectividad de ella misma.
- e) Observación directa: es observar en terreno aspecto determinados de una organización.
- f) Entrevista grupal: Se selecciona un grupo de personas afines de la organización para entrevistarlos en grupo. Se suele tratar sobre aspectos críticos de la entidad.

A continuación se presentan las principales áreas de diagnóstico de interés para las organizaciones, las que idealmente deberían ser recopiladas y analizadas.

	Medio	Medio externo
Estructuras y procesos técnicos	Procesos productivos, administrativos y de servicios Estructura organizacional. Contabilidad y finanzas Tecnología disponible	Productos/mercados. Tecnología existente. Marco regulatorio y legal. Mercado laboral.
Personas y procesos humanos	Procesos de recursos humanos: selección, capacitación., evaluación del desempeño, remuneración. etc. Comportamiento organizacional: poder, liderazgo, trabajo en equipo, motivación, compromiso, cultura, clima, comunicación, etc.	Relaciones públicas. Consumidores. Proveedores. Conducta de la competencia.

Figura 19: Raineri, Andrés; Martínez, Andrea (1997) Diagnostico organizacional. Revista Academia, Universidad de Chile. Recuperado de [http://ww2.educarchile.cl/UserFiles/P0001/File/P0001_File_DO%20Un%20enfoque%20estrategico%20y%20pr%C3%A1ctico%20\(Rainieri\),%2019%20mayo.pdf](http://ww2.educarchile.cl/UserFiles/P0001/File/P0001_File_DO%20Un%20enfoque%20estrategico%20y%20pr%C3%A1ctico%20(Rainieri),%2019%20mayo.pdf)

- **Políticas**

La organización que desee implementar la metodología deberá establecer, documentar, mantener y promulgar una política de gestión del método ICAV, con el objetivo de que los datos e información que se gestione sirva en función de evidencia y rendición de cuentas; y de probar que las actividades se llevan a cabo bajo el marco normativo de la política. Esta política se elaborará a partir de un diagnóstico de las actividades de la organización y considerando el entorno legal y normativo. Es de suma necesidad que esta política sea respaldada por los altos cargos y comunicada a los demás niveles, y debe ser revisada periódicamente para ver que las necesidades de la organización se mantengan o varíen.

La dirección debe asegurarse que esta política:

a) Corresponde con los propósitos del sistema

- b) Cuenta con un marco de referencia para revisar los objetivos
- c) Será comunicada y comprendida dentro de la organización
- d) Debe ser revisada para adecuarla a los cambios de las realidades de la organización

La comunicación en la organización es importante ya que si es buena, mejora el clima y la productividad laboral. Por lo tanto para comunicar las políticas y normas, se pueden ocupar algunos de los siguientes medios según el criterio de la organización:

Herramienta	Descripción
Manual de empleado	Incluye información relacionada con la empresa que todo empleado debería conocer. Puede incluir la historia de la empresa, sus normas internas, su misión y valores, las políticas de recursos humanos, el organigrama, pautas de seguridad o procedimientos específicos.
Publicación institucional o "newsletter"	Con el objetivo de informar a los empleados, en forma de revista, periódico, o folleto interno —en papel o digital—, incluye datos de la operatividad de la empresa como de los mismos empleados.
Cartelera	Es un efectivo medio de comunicación siempre que se ubique en un lugar concurrido por los empleados, con fácil acceso visual. Es necesario actualizarla periódicamente y consignar la fecha de cada publicación. Puede incluir informaciones legales o sociales, resultados de la empresa, nuevas políticas o intercambios entre el personal.
Circulares	Es un mensaje que se envía a toda la empresa o a una parte de ella. Los medios para hacerla conocer son variados (correo electrónico, carta, intranet, blogs, etc). Puede dar a conocer información, órdenes, sugerencias o convocatorias.
Reuniones	Permiten la interacción personal dentro de los equipos de trabajo o entre las distintas áreas de la empresa.

	Pueden usarse para informar, capacitar, coordinar nuevas tareas y tomar decisiones.
Correo electrónico	Representa un medio rápido de comunicación con muchos interlocutores. Es, a la vez, de bajo costo y fácil de personalizar.
Buzón de sugerencias	El objetivo es conocer las opiniones y propuestas de los empleados en diversos aspectos: normas, organización, condiciones de trabajo, tareas, relaciones laborales, beneficios sociales.
Videoconferencia	Es una herramienta muy utilizada gracias a su rapidez y bajo costo. Permite conectar a dos o más personas que estén distanciadas geográficamente, evitando desplazamientos y ahorrando así tiempo y dinero.
Intranet	Es una red de comunicación dentro de la empresa. Su instalación es simple y económica, sólo requiere la inversión en infraestructura. Su ventaja principal ser un medio rápido e instantáneo: disponibilidad al instante de la información, material compartido o base de datos de la empresa.

Figura 20: Buenos negocios (2013) 9 herramientas útiles para la comunicación interna. Recuperado de <http://www.buenosnegocios.com/notas/448-9-herramientas-utiles-la-comunicacion-interna>

- **Responsabilidades y recursos humanos**

Se deben asignar las competencias y responsabilidades relacionadas con la gestión de los datos en la organización. Se tiene que tener en conocimiento mediante registros las competencias del personal en cuanto a su educación, formación, habilidades y experiencias apropiadas. Se debe tener en cuenta:

- a) Los profesionales en la gestión, deben asumir la responsabilidad de seguir y mejorar los procesos
- b) La dirección tiene la responsabilidad de garantizar el éxito del plan
- c) Los jefes de las áreas deben asegurar que el personal a cargo cumple con los requisitos y con las políticas

d) El personal que usa información durante sus tareas diarias es responsable de la creación y la mantención de estos y manteniendo la fiabilidad e integridad

e) Se deben determinar las competencias necesarias para el personal que realizan las tareas. Si no es así, se debe proporcionar la formación u otras acciones para lograr la competencia.

- **Infraestructura**

La organización debe determinar y proporcionar la infraestructura necesaria para llevar a cabo el plan. Por esto se debe evaluar la factibilidad de implementar el proyecto Big Data en lo que a infraestructura se refiere. La entidad a cargo debe ser la responsable de proporcionar:

a) Edificios, espacios laborales y servicios necesarios

b) Equipos para los procesos, computadores con capacidad suficiente (hardware) y programas necesarios para la totalidad de labores de la organización (software)

c) Servicios de apoyo

3.2.1.3 Identificar necesidades externas de la organización

La organización debe determinar:

a) Los requisitos de los usuarios de la unidad de información, en cuanto a la entrega del servicio

b) Los servicios posteriores a la entrega del servicio, como la garantía, obligaciones contractuales y servicios suplementarios

b) Requisitos reglamentarios y legales en cuanto a la entrega del servicio

c) Al usuario se le debe informar de forma clara información sobre el servicio y sus implicancias

d) Debe haber una comunicación fluida con el usuario en cuanto a consultas, contratos y atención al cliente

e) Se debe aplicar retroalimentación con el usuario, como por ejemplo sobre quejas y reclamos

3.2.1.4 Identificar las necesidades informacionales de cada área de la organización.

Es decir, identificar los datos o información que hace falta para mejorar el servicio que ofrece cada área.

Ej: “Necesitamos conocer cuáles son los libros más solicitados mensualmente para reorganizar la colección de alta demanda”

3.2.1.5 Exponer las necesidades informacionales.

Éstas se deben exponer a través de preguntas que nos lleven a reflexionar sobre cómo podemos solucionar lo anteriormente expuesto.

Ej: ¿Cómo puedo potenciar el desarrollo de colecciones a través de los datos?

Ej2: ¿Cómo podríamos reorganizar la colección de alta demanda?

3.2.2 Consolidar

Esta fase es la encargada de la recolección de los datos, por lo tanto se recomienda que el encargado de esta etapa, sea el informático de la organización o el personal dedicado a las TICs.

*En caso de no tener un encargado de las TICs, se debe capacitar al personal en esta área.

3.2.2.1 Identificar las fuentes de datos de la organización que nos servirán para responder nuestras preguntas.

En base al diagnóstico realizado en la etapa anterior, quedarán en evidencia las fuentes de datos internos y externos en las que la organización trabaja y

genera datos. Estas fuentes de datos detectadas deben ser utilizadas para la recolección de información relevante en base a las necesidades planteadas.

Estas fuentes de datos pueden ser:

- Softwares de gestión documental de la Unidad de Información

Ej: de datos: Cantidad de préstamos, tamaño de colección, transacciones realizadas

- Bases de datos relacionales

Ej: de datos: Número de consultas, cantidad de empleados, datos internos de archivo

- Sitios webs de la organización

Ej: de datos: Número de visitas, Número de clickeos para llegar a información, Alcance de la información publicada

- Sensor de visitas

Ej: de datos: Cantidad de personas por hora/mes/año que ingresan a la Unidad de Información, temporada con más visitas.

- Opac's

Ej: de datos: Información específica de la colección (títulos, año de publicación, editorial, etc),

- Computadoras locales

Ej: de datos: Cantidad de documentos que maneja, correos electrónicos, transferencia de archivos.

- Etc.

3.2.2.2 Recolectar los datos pertinentes en las fuentes de información en base a las necesidades planteadas.

En esta etapa se recolectarán los datos realmente útiles para nuestro propósito.

Para efectos de esta etapa es necesario utilizar tecnologías que nos ayuden con la tarea de recolección de datos. A continuación mencionaremos nuestra recomendación:

Tecnología recomendada:

La tecnología recomendada a este proceso es Apache Hadoop, framework de código abierto. Para estas labores se recomienda su uso por las siguientes dos razones principalmente:

- Procesa grandes volúmenes de datos mediante clústers, usando un modelo simplificado de programación.
- Su diseño permite pasar de pocos nodos a miles, y esto de forma expedita.

La arquitectura de Hadoop se divide en dos partes:

- HDFS (Hadoop Distributed File System): Es un Sistema de archivos que abarca la totalidad de nodos en un clúster para almacenar datos
- MapReduce: Perteneciente a la primera generación de Hadoop, actualmente forma parte de YARN (Yet Another Resource Negotiator). Este framework otorga un sistema donde los datos se procesan de forma paralela y distribuida. Este algoritmo tiene dos tareas, que son Mapa y Reducir, los cuales forman la metodología de trabajo de Hadoop.

MAP

En “Map” se procesan los datos de entrada, que lo más común es que vengan en forma de archivo o directorio y se almacenan en el sistema donde se almacenan los archivos, el HDFS. En esta función se reciben un conjunto de datos, y estos se transforman a un número de pares (clave, valor) y así cada elemento se encontrará ordenado por su respectiva clave, transformándose a una lista de pares. Esto se aplica a cada entrada de

datos, obteniéndose así una lista de pares según lo que se le solicite a Map.⁵⁴

(Entrada) Map <clave1, valor1> → lista <clave2, valor2>

REDUCE

Esta función funciona en paralelo para cada grupo creado en “Map Reduce” se usa una vez para cada clave de las salidas de la función Map. Con esta clave, se ordena una lista con todos los valores asociados a la clave, realizando una fusión, y produciendo un conjunto más pequeño de valores. Por lo que en la función Reduce se combinan los valores de la misma clave, en un mismo resultado.⁵⁵

(Entrada) Map <clave1, valor1> → lista <clave2, valor2>-> reduce -> <k3, v3> (salida)

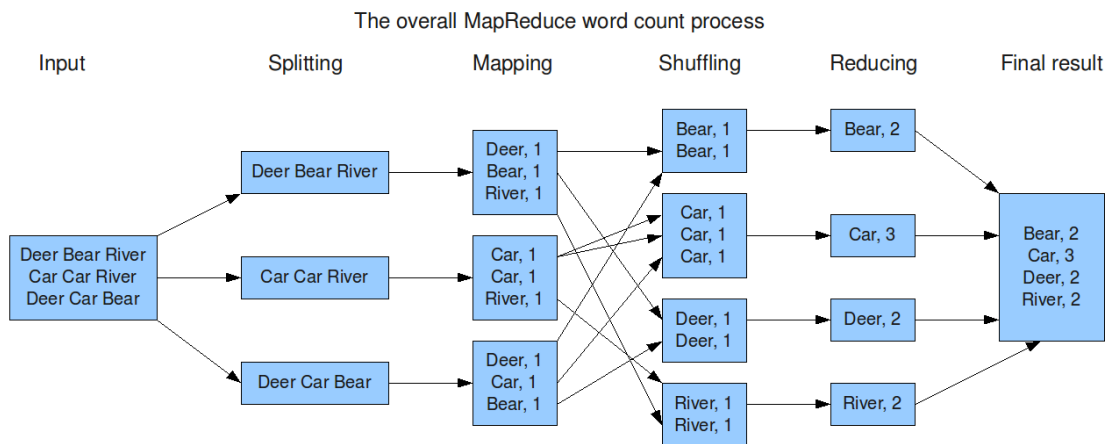


Figura 21: Usulli, Micheli (2013) An example of MapReduce with rmr2. Milano. Recuperado de <http://www.milano.net/blog/an-example-of-mapreduce-with-rmr2/>

⁵⁴ Olmedo, Yolanda (2012) ¿Qué es MapReduce? Solidq. Recuperado de <http://blogs.solidq.com/es/big-data/que-es-mapreduce/>

⁵⁵ *Ibíd.*

3.2.2.3 Agrupar los datos recolectados en un repositorio o almacén de datos

En esta etapa se deben agrupar los datos considerando las preguntas de la etapa 1. Es decir, agrupar los datos necesarios para dar respuesta a la necesidad planteada.

Estos datos pueden ser agrupados por:

- Rango etario
- Datos socioeconómicos
- Datos demográficos
- N° de Préstamos
- Fechas
- Cantidad de la colección
- Número de visitas en sitios webs
- Etc.

Se debe crear un repositorio o almacén de datos. Estos deben estar “limpios” y adaptados en su formato para poder facilitar su análisis, pues será una base de datos paralela a los sistemas utilizados para operar dentro de la institución.

3.2.3 Analizar

El análisis es quizás la etapa más interesante de ICAV, ya que es la etapa donde se debe conjugar las herramientas tecnológicas y el valor humano, es decir, el asimilar, el entender y el razonamiento por parte de las personas. A través de los datos recopilados, se debe incluir por ejemplo la descripción de patrones de datos, descubrimiento de relaciones entre los mismos, prepararse proactivamente para el futuro a través de técnicas predictivas para así optimizar los recursos y dar nuevas alternativas de servicios y mejoras.

El análisis de datos debe entregar nuevo conocimiento:

Algunos ejemplos:

- a) Satisfacción de los usuarios

- b) Posible creación de nuevos servicios, en base a la necesidad de usuarios.
- c) Características y tendencias de procesos y servicios, para así llevar a cabo plan de acciones preventivas y/o de crecimiento.
- d) Optimización de labores.
- e) Marketing

3.2.3.1 Analizar los datos recolectados.

El análisis de los datos está netamente relacionado a la interpretación y al cruce de datos que nos ayudarán a identificar patrones o tendencias que no estaban siendo consideradas. (Creación de nuevo conocimiento)

Existen distintos medios para la interpretación de resultados, uno de los más utilizadas son las matrices de confusión siendo muy empleadas en problemas de clasificación. Consiste en una tabla que indica cuantas clasificaciones se han hecho para cada tipo, en donde la diagonal de la tabla representa las clasificaciones correctas.⁵⁶

Predicción	Clase A	Clase B	Clase C
Clase A	45	2	3
Clase B	10	38	2
Clase C	4	6	40

Figura 22: Herrera Varela, Ricardo (2006) Bibliomining: minería de datos y descubrimiento de conocimiento en bases de datos aplicados al ámbito bibliotecario. Universidad Carlos III de Madrid.

Recuperado en: http://bibliotecarios.cl/conferencia_2006/C2006_019.pdf

Otra de las opciones disponibles es la minería de datos, esta técnica consta de dos categorías principales a tener en cuenta para la extracción del conocimiento; la predicción y la descripción, siendo el primera de ellos la forma que podemos pronosticar la forma que evolucionaran ciertos patrones de comportamiento de nuestra biblioteca y el segunda la manera de cómo podemos tener una situación actual del área de estudio determinada para el análisis.⁵⁷

⁵⁶ Varela, R. H. (Diciembre de 2006). Bibliomining: minería de datos y descubrimiento de conocimiento en bases de datos aplicados al ámbito bibliotecario.

⁵⁷ *Ibíd.*

Tecnologías

En lo que tecnologías se refiere, también hay opciones de análisis de datos ofrecidas por Hadoop. En él hay un gran y diverso ecosistema de sistemas y herramientas, que con el pasar del tiempo van mejorando y aumentando.

A continuación se mencionaran unos sistemas recomendados para la fase de análisis:

- Apache UIMA: (Unstructured Information Management Applications) Este framework tiene por función el analizar los datos no estructurados (que por lo general ocupan un 80% del total de información de una organización), los que suelen ser de texto, video audio, etc. Y de estos se puede obtener información relevante para el usuario final.
- Apache Mahout: Es creado para crear data mining y aprendizaje automático con el uso de Hadoop. De este modo se pueden descubrir patrones de datos y además de tener algoritmos de clasificación, clustering y recomendación.⁵⁸

3.2.4 Visualizar

3.2.4.1 Dar a conocer los resultados.

Los resultados obtenidos en el análisis, se deben exponer en una nueva reunión con todos los encargados de cada área. En esta reunión se deben asignar las responsabilidades y las acciones a seguir para cada uno de los involucrados en el manejo de información y los datos.

Entrega de resultados obtenidos

Todo proyecto de Big Data tiene como sentido o finalidad mejorar la toma de decisiones en las organizaciones. Para ello se comunica la información final de acuerdo a la necesidad de cada área usuaria; solución de inteligencia en los negocios, también abordable en unidades de información.

⁵⁸ Ticout (2013) Introducción a Hadoop y su ecosistema. Ticout Outsourcing Center. Recuperado de <http://www.ticout.com/blog/2013/04/02/introduccion-a-hadoop-y-su-ecosistema/>

Al momento de distribuir el resultado del análisis de Big Data se recomiendan los dispositivos móviles. Aparatos que servirían a la unidad de información como herramienta de interacción, agiliza los procesos internos de la organización; canal que da ventaja para ofrecer y contactar a sus usuarios. Tendencia rápida que permite: entregar la información (resultados obtenidos) y recolectar la opinión de los tres pilares de nuestra organización: usuarios, trabajadores y accionistas. Conceder información a través de un sistema de gestión eficiente, permite cumplir con la misión, visión y valores de cada organización, fomentar el desempeño y satisfacción de cada usuario.

Con disciplina, metodología y las características de Big Data se pretende entregar seguridad al momento de acceder a la información, frente al análisis de diversos escenarios, donde se podrá visualizar la información en diversos dispositivos móviles sean amenos y modificables por el usuario final.

La variedad con la que se consiguen los datos, se preñera regular bajo un marco legal, donde la velocidad y volumen con la que se recopilan, será encausada a los principios y finalidades por las que fueron tratadas, logrando valor para la toma de decisiones de la organización y paralelamente veracidad en nuestro servicio. Se puede ejemplificar este hecho en la figura 23.

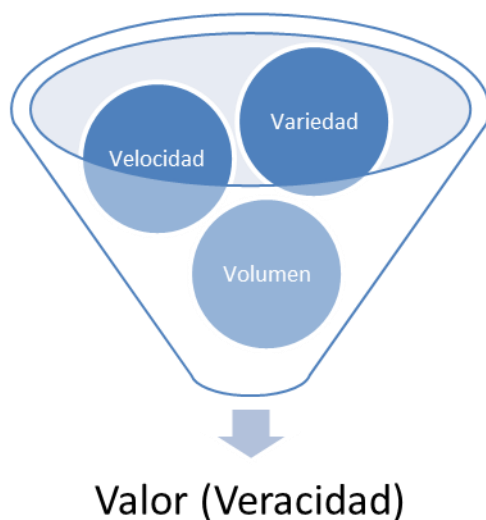


Figura 23: Creación propia.

Reuniones mensuales de retroalimentación

Se recomienda realizar una reunión mensual, con un número significativo de actores, para la retroalimentación. Esto tiene un doble propósito. Primero generar una sinergia cognitiva, definida como una capacidad colectiva para

intervenir sobre el proceso de desarrollo, capacidad basada no solo en la voluntad, si no en el conocimiento científico sobre la estructura y dinámica del proceso que convoca. Segundo provocar una forma de entender los procesos de crecimiento organizacional y hacer un seguimiento de las labores.

Informes mensuales por área

El seguimiento de las labores, consta de un informe mensual que contenga observaciones dirigidas a los avances que ha tenido el área en materia de control de datos. Estos informes deben responder a la necesidad planteada en un inicio, verificando los efectos que está causando el correcto manejo de los datos y como estos están optimizando el desarrollo en el área específica.

Con estos pasos, podremos mantener constantemente informados a todo el personal de los avances en materia de control de datos. Se formará un equipo sólido de trabajo y se desarrollará la organización en conocimiento de todos.

Es importante dejar en claro a cada uno de los involucrados:

- 1- Asegurarse que las actividades se encuentren dentro de los límites de lo legal y ético con el fin de mejorar el servicio. Dentro de las consideraciones éticas se destaca el seguimiento que se hace al usuario para conocer sus comportamientos.
- 2- Se debe desarrollar políticas de aislamiento de los datos a utilizar. Lo más recomendado es consultar e informar previamente al usuario sobre la finalidad de los datos analizados; ya sea por, teléfono, e-mails o persona a persona. De esta forma certificaremos que el grupo de usuarios ha autorizado el uso de los datos.
- 3- Tener especial atención con los usos de los datos personales, ya que estos se relacionan con la intimidad y privacidad y su uso indebido significa perjuicios en contra de los titulares de dichos datos. Siempre se debe tener presente que el tratamiento de datos personales es un tema complejo, y con múltiples aristas.

Es importante aclarar que los datos personales, son atribuibles a una persona determinada identificada o identificable. Valorando su sensibilidad, se reconocen dos clases de datos:

- Datos sensibles: se refieren a características físicas o morales, hechos o circunstancias de su vida íntima o privada de las personas, tales como:

- a) Hábitos personales,
- b) Origen racial,
- c) Ideologías y opiniones políticas,
- d) Creencias o convicciones religiosas,
- e) Estados de salud física o psíquicos y,
- f) Vida sexual.

- Datos inexorables: aquellos que no se pueden evitar, eludir o detener, que son los respectivos al asunto económico y comercial, relativo a personas naturales o jurídicas.

En la etapa de visualización se debe dejar en claro los propósitos por los cuales se aplicarán las acciones a seguir. Dentro de estos propósitos, considerando que aplicaremos esta metodología en Unidades de Información, es importante señalar el especial interés en satisfacer a los usuarios de nuestra Unidad.

La existencia de la organización (misión, visión y valores) se rinde al cumplimiento de los requisitos de sus clientes/usuarios, por ende este juega un papel significativo para definir exigencias. La apreciación del interesado hacia la organización, es vital, por lo que se requiere evaluar si se han cumplido sus requisitos, lo que concierne a información relativa a su percepción.

El seguimiento de la satisfacción será una herramienta para tomar medidas de desempeño (establecer métodos para obtener y utilizar dicha información). Se debe implementar el perfeccionamiento de la eficacia en un sistema de gestión, adoptando un enfoque basado en procesos; se lograra calidad y aumentar la satisfacción del usuario, y a su vez la del personal a cargo.

A continuación en la figura se ilustra el modelo de un sistema de gestión de calidad basado en procesos, donde el cliente/usuario juega un papel significativo para definir requisitos.

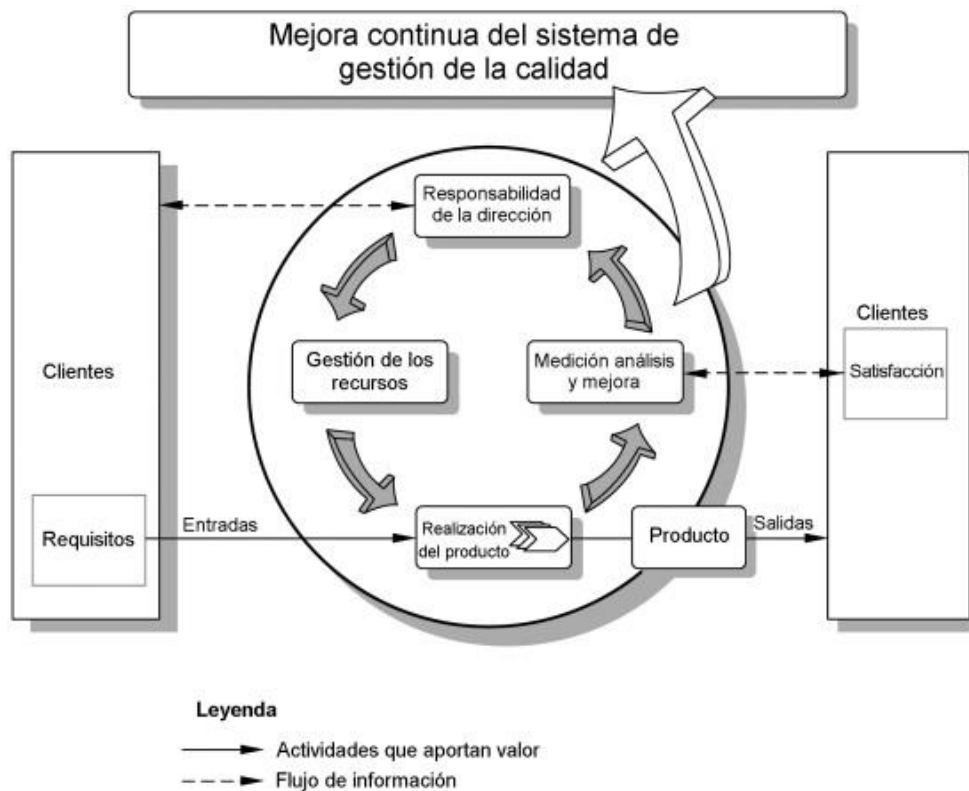


Figura 23: International Organization for Standardization (2008) Sistema de gestión de calidad.

Recuperado de <https://www.mct.es/sites/default/files/archivos/ISO-9001.pdf>

Conclusión

Las Bibliotecas, Archivos, Centros de Información, en general, la totalidad de las Unidades que manejan información, tienden más a adaptarse a las innovaciones tecnológicas que ser ellas mismas un ente generador de innovación. Esto lo comprobamos como investigadores desde el comienzo de nuestra búsqueda, ya que desde el principio tuvimos que lidiar con la escasa y casi inexistente bibliografía donde se abordara el uso del Big Data en unidades de Información.

La gran cantidad de información rescatada y analizada para nuestro trabajo proviene principalmente de otras áreas, por ejemplo, bibliografía elemental fue la de Luis Joyanes Aguilar Doctor en Informática y Doctor en Sociología, que ha sido fundamental en el estudio y publicación sobre el Big Data, abordado eso sí, desde un prisma empresarial. Como grupo tuvimos la capacidad para visualizar el vacío, la carencia de material donde los Gestores de Información pudiésemos tener un acercamiento al concepto del Big Data en Unidades de Información y mucho menos tener una herramienta práctica para la aplicación del uso del Big Data.

Actualmente la adopción del Big Data está en una fase inicial y quedan pocas dudas de que su uso no se agotará o desechará pronto. Esto recién comienza, ¿Por qué? Porque estamos frente a un fenómeno que es global y transversal, es decir, se puede aplicar a cualquier sector en el que se quiera obtener beneficios a partir de datos obtenidos, tratados y almacenados.

Big data es un fenómeno que ha sido formado a lo largo del tiempo a través de las tecnologías de información y comunicación. El desarrollo de las distintas plataformas a nivel digital, es uno de los principales alimentos para este fenómeno generando millones y millones de datos que no están siendo aprovechados. Big data se presenta como un cambio de orden en la organización, donde el usuario debe considerar su rol protagónico.

Con frecuencia se utilizan bases de datos para almacenar información referente: al usuario, recursos, transacciones, registros, etc. Estos se pueden rastrear y registrar en el sistema de información otorgando la capacidad de

revisar la visita completa de un usuario, y saber si obtuvo los recursos de información que necesitaba y cual le fue útil. Esto es usar Big Data.

Sin embargo, hay algunas cosas por resolver. El desarrollo de competencias TIC crea brechas tanto digitales como sociales entre las personas, facilitando oportunidades de interconexión y aprovechamiento de la información para unos y aislando a otros a un estado de restricciones cognitivas generando dificultades en el manejo de herramientas y recursos necesarios para desarrollarse intelectual y socialmente. Es aquí donde los bibliotecarios debemos ser capaces de educar y ofrecer recursos de aprendizaje de nuevas tecnologías de la información para que nadie quede fuera.

Como pudimos analizar en este trabajo, Big Data se puede convertir en una herramienta importante en la gestión de procesos o toma de decisiones de una organización tan solo teniendo en cuenta los datos que generamos a diario los usuarios y/o el personal a cargo. Recolectando los datos necesarios y analizándolos debidamente, podemos generar nuevo conocimiento que irá en apoyo al desarrollo de nuestra organización. Parece ser algo muy sencillo, incluso si somos más objetivos, lo único que estamos haciendo es tomar los datos que siempre han estado presentes y les damos utilidad.

Por lo tanto, con Big Data se puede generar valor, siempre y cuando haya una utilización correcta de la información. Estos datos ayudan a mejorar rendimiento y a comprender el contexto de la organización, así como a comprender mejor al usuario y a potenciales usuarios como también el medio en general de la unidad de información en la que se trabaje. De este modo, los datos pueden generar valiosa información para anticipar necesidades y conocer las necesidades informacionales de los usuarios, para así crear estrategias de acción resultantes de una fructuosa toma de decisiones. Para crear valor en estos datos se debe transparentar la información y hacerla utilizable. Esto si se almacenan los datos de forma más precisa y detallada mejorando así el rendimiento de la información resultante. De esta forma se puede obtener por ejemplo una segmentación más estrecha de usuarios, y así ofrecer el producto informacional de forma más precisa; y a largo plazo se podría mejorar el desarrollo de la próxima generación de servicios en las unidades de información.

Sin embargo, la cantidad de datos que hoy en día genera una organización a través de sus distintos procesos, productos y servicios, es tan grande que

dificulta la tarea y se hace muy necesario acceder a herramientas tecnológicas que nos ayuden con esta labor.

El trabajo de administrar Big Data para la organización, se ha convertido en una ventaja competitiva en las empresas de este siglo, ya lo pudimos observar en este trabajo de investigación en donde las empresas usan Big Data para crear planes de acción en base a las necesidades que los mismos datos les van revelando. Es por esto que como profesionales de la información, debemos estar muy alerta a los cambios que el mundo tecnológico le está dando al uso de los datos. El rol del bibliotecario cada día está tomando más protagonismo pues son los gestores de información aquellos que deben facilitar la tarea en este plano y ser las piezas claves en las empresas para el desarrollo de sus actividades.

Es importante que los bibliotecarios tengan conocimiento de lo que está sucediendo con el Big Data actualmente, y puedan desarrollar competencias que le capaciten en el manejo de grandes cantidades de datos.

Al realizar este trabajo, pudimos detectar una pequeña deficiencia en el actual manejo de la información que los bibliotecarios están teniendo hoy en día; El área tecnológica. Creemos que los profesionales de la información deben tener conocimientos más avanzados en tecnologías y que esto no sea una traba para la implementación de sistemas de gestión documental en cualquier organización.

En la realización de este trabajo, podemos tomar como lección aprendida el haber hecho el marco teórico (capítulo 2), que fue de bastante importancia al realizar esta propuesta metodológica, ya que hizo función de sustento teórico y conceptual; y como información base del tema de Big Data para trabajar en esta propuesta. De este modo se logró un mayor manejo del tema a tratar y por lo tanto, una realización fluida en la realización del capítulo 3.

Al desarrollar la metodología Big Data, pudimos darnos cuenta de la gran falencia tecnológica que tenemos los bibliotecarios a la hora de implementar cualquier forma de gestión y/o control de datos/información ya que dependemos de un informático para poder hacer efectiva la implementación. Creemos que es importante desarrollar en el futuro una malla curricular que aborde estas temáticas tecnológicas en las cuales se traten aspectos técnicos en la implementación de cualquier sistema de gestión de información y que ésta se vaya actualizando anualmente según la necesidad que el mundo de la informacional tenga.

Vemos como Gestores de Información el aporte al atisbar falencias de nuestra profesión, pero creemos también en que nuestra profesión es la más idónea para hacerse cargo y empoderarse en el uso y tratamiento del Big Data en una Institución. Y finalmente pensamos que somos un ente clave en valor cultural que podemos realizar con las herramientas adecuadas y sobretodo queremos ser un elemento clave también en la contribución hacia la Sociedad del Conocimiento.

Mantener a las nuevas generaciones actualizadas con los diversos fenómenos que la era informacional va generando, es una tarea que los directivos de las Universidades que imparten la carrera de Bibliotecología o Gestión de Información tienen que considerar, pues es clave para todos los profesionales que trabajan con tecnologías de la información, que sus competencias sean actualizadas. De lo contrario, tendremos profesionales no aptos para la época en la que vivimos, pues el mundo tecnológico está en constante cambio.

Glosario

ARPANET: (Advanced Research Projects Agency Network o Red de la Agencia de Proyectos de Investigación Avanzada) Fue una red de computadoras creada por encargo del Departamento de Defensa de los Estados Unidos para utilizarla como medio de comunicación entre las diferentes instituciones académicas y estatales. Pág: 12

Clúster: se aplica a los conjuntos o conglomerados de ordenadores unidos entre sí normalmente por una red de alta velocidad y que se comportan como si fuesen una única computadora. Pág: 34, 35, 60

Dewey: (Sistema de Clasificación Decimal Dewey) es un sistema de clasificación de bibliotecas. Pág: 36

Framework: (Infraestructura digital) es una estructura conceptual y tecnológica de soporte definido, normalmente con artefactos o módulos concretos de software, que puede servir de base para la organización y desarrollo de software. Pág: 34, 47, 60, 64

FTP: (File Transfer Protocol o transmisión de archivos) En informática, es un protocolo de red para la transferencia de archivos entre sistemas conectados a una red TCP (Transmission Control Protocol), basado en la arquitectura cliente-servidor. Pág: 12

GPS: (sistema de posicionamiento global) es un sistema que permite determinar en toda la Tierra la posición de un objeto (una persona, un vehículo) con una precisión de hasta centímetros (si se utiliza GPS diferencial), aunque lo habitual son unos pocos metros de precisión. Pág: 27, 33

HTML: (Hypertext Markup Lenguaje o lenguaje de marcas de hipertexto) lenguaje de marcado para la elaboración de páginas web. Págs: 15, 16, 26

HTTP: (Hypertext Transfer Protocol o protocolo de transferencia de hipertexto) es el protocolo de comunicación que permite las transferencias de información en la World Wide Web. Págs: 13

IP: (Internet Protocol o Protocolo de Internet) Es un número que identifica, de manera lógica y jerárquica, a una Interfaz en red (elemento de comunicación/conexión) de un dispositivo (computadora, tableta, portátil,

smartphone) que utilice el protocolo IP (Internet Protocol), que corresponde al nivel de red del modelo TCP/IP. Pág: 12

JSON: Acrónimo de JavaScript Object Notation, es un formato de texto ligero para el intercambio de datos. JSON es un subconjunto de la notación literal de objetos de JavaScript aunque hoy, debido a su amplia adopción como alternativa a XML, se considera un formato de lenguaje independiente. Pág: 26

Navegador web: (Web browser) es un software, aplicación o programa que permite el acceso a la Web, interpretando la información de distintos tipos de archivos y sitios web para que estos puedan ser visualizados. Pág: 13

Nodo: es un punto de intersección, conexión o unión de varios elementos que confluyen en el mismo lugar. Pág: 35, 60

Opac: (Online public access catalog) es un catálogo automatizado de acceso público en línea de los materiales de una biblioteca o una unidad de información. Pág: 7, 59

PDF: (Portable Document Format o formato de documento portátil) es un formato de almacenamiento para documentos digitales independiente de plataformas de software o hardware. Pág: 21, 26

Recuperación de información: (ISR o Information Search and Retrieval) es la ciencia de la búsqueda de información en documentos electrónicos y cualquier tipo de colección documental digital, encargada de la búsqueda dentro de éstos mismos, búsqueda de metadatos que describan documentos, o también la búsqueda en bases de datos relacionales. Pág: 17

Repositorios: Depósito o archivo es un sitio centralizado donde se almacena y mantiene información digital, habitualmente bases de datos o archivos informáticos. Pág: 23, 49, 50, 62

SMS: (Short Message Service o servicio de mensajes cortos) es un servicio disponible en los teléfonos móviles que permite el envío de mensajes cortos, conocidos como mensajes de texto, entre teléfonos móviles. Pág: 26

SMTP: (Simple Mail Transfer Protocol o protocolo para transferencia simple de correo) Es un protocolo de red utilizado para el intercambio de mensajes de correo electrónico entre computadoras u otros dispositivos (PDA, teléfonos móviles, impresoras, etc). Pág: 12

TCP: (Transmission Control Protocol o Protocolo de Control de Transmisión) Es uno de los protocolos fundamentales en Internet. Fue creado entre los años 1973 y 1974 por Vint Cerf y Robert Kahn. Pág: 12

Tecnologías de la información y la comunicación (TIC): son el conjunto de medios (radio, televisión y telefonía convencional) de comunicación y las aplicaciones de información que permiten la captura, producción, almacenamiento, tratamiento, y presentación de informaciones en forma de voz, imágenes y datos contenidos en señales de naturaleza acústica, óptica o electromagnética. Pág: 12, 37, 58, 70

VoIP: (Voice Over Internet Protocol o transmisión de contenido y comunicación multimedia) Es un conjunto de recursos que hacen posible que la señal de voz viaje a través de Internet empleando el protocolo IP (Protocolo de Internet). Pág: 13

World Wide Web: (WWW o Web) Es un sistema de distribución de documentos de hipertexto o hipermedios interconectados y accesibles vía Internet. Con un navegador web, un usuario visualiza sitios web compuestos de páginas web que pueden contener texto, imágenes, vídeos u otros contenidos multimedia, y navega a través de esas páginas usando hiperenlaces. Pág: 13, 15, 17, 18

3G: Es la abreviación de tercera generación de transmisión de voz y datos a través de telefonía móvil mediante UMTS (Universal Mobile Telecommunications System o servicio universal de telecomunicaciones móviles. Pág: 17

Referencias Bibliográficas:

(ACTI), A. C. (30 de Abril de 2015). *Big Data toma fuerza en Chile con soluciones que analizan textos, voz, imágenes y video*. Obtenido de (ACTI), Asociación Chilena de Empresas de Tecnología de Información: <http://www.acti.cl/detalle-noticia.php?mul=737>

Ardila Cañas, Esneyder; Gómez Díaz, Isabel (2014) Estrategias para la gestión de grandes volúmenes de datos por medio de Big Data en el contexto de la analítica de negocios: Caso MVM ingeniería de software S.A.S. Recuperado de http://bibliotecadigital.usbcali.edu.co/jspui/bitstream/10819/2702/1/Gestion_Volumen_Datos_Ardila_2014.pdf

Buenos negocios (2013) 9 herramientas útiles para la comunicación interna. Recuperado de <http://www.buenosnegocios.com/notas/448-9-herramientas-utiles-la-comunicacion-interna>

Calabuig, Oriol. (2004) ¿Qué es Big Data? Las entrañas de los datos. Portal comunicación. Recuperado de: http://portalcomunicacion.com/monograficos_det.asp?id=261

Calderín Cruz, Mabel. (2007). Sistema de información documental en la Academia: Situación actual y perspectivas futuras, caso UCAB. Recuperado de http://www.cienciared.com.ar/ra/usr/3/333/n6_v1pp51_76.pdf

Carretero, J. (2001). Descubre Internet.

Castillo, Lourdes (2001) Introducción a la información científica y técnica. Recuperado de <http://www.uv.es/macas/1.pdf>

Castillo, Lourdes (2004) Los centros de documentación. Recuperado de <http://www.uv.es/macas/T1.pdf>

Domo. (Agosto de 2015). Obtenido de Sitio Web Domo.com: <https://www.domo.com/blog/2015/08/data-never-sleeps-3-0/>

García Calero, Jesús. (2015, 13 abr.). Ana Santos: “El Big Data español está aquí, en la Biblioteca Nacional”. ABC.es, cultura.

Figura 22: Herrera Varela, Ricardo (2006) Bibliomining: minería de datos y descubrimiento de conocimiento en bases de datos aplicados al ámbito bibliotecario. Universidad Carlos III de Madrid. Recuperado en: http://bibliotecarios.cl/conferencia_2006/C2006_019.pdf

Hernandez, J. (10 de Abril de 2015). *Nubeteca*. Obtenido de <http://www.nubeteca.info/blog/?p=25>

Heskett, L [et. al.] Poniendo a trabajar la cadena servicio-ganancia.

IBM Institute for Business Value (2002) Analytics: el uso de big data en el mundo real Cómo las empresas más innovadoras extraen valor de datos inciertos. IBM. Recuperado de http://www-05.ibm.com/services/es/gbs/consulting/pdf/EI_uso_de_Big_Data_en_el_mundo_real.pdf

Koday Games. (2012, 21 dic). Unidades de medida informática. Disponible en: <http://kodaygames.blogspot.cl/2012/12/unidades-de-medida-en-informatica.html>

Lantares. Velocidad, variedad y volumen, las 3 magnitudes claves del big data. Disponible en: <http://www.lantares.com/blog/velocidad-variedad-y-volumen-las-3-magnitudes-clave-de-big-data>

Lopez, Jose Carlos (2014, 27 feb). La moda del Big Data: ¿En qué consiste en realidad? Disponible en: <http://www.eleconomista.es/tecnologia/noticias/5578707/02/14/La-moda-del-Big-Data-En-que-consiste-en-realidad.html>

Martín Gavilán, César (2009) Temas de Biblioteconomía: Concepto y función de archivo clases de archivos El sistema archivístico Español. Recuperado de <http://eprints.rclis.org/14058/1/sisarchivesp.pdf>

Martín, M., & Jorge, U. (Julio de 2014). *Big Latam: Quien es quién en Big Data en Latinoamérica*. Obtenido de http://blog.soydata.net/wp-content/uploads/InfogBIGLATAM_1.pdf

Margaix, D.(2007).Conceptos de web 2.0 y bibliotecas 2.0: origen, definiciones y retos para las bibliotecas actuales. Recuperado de <http://eprints.rclis.org/9521/1/kx5j65q110j51203.pdf>

Mayer-Schönberger, Viktor; Cukier, Kenneth (2013) Big data: La revolución de los datos masivos. Turner Publicaciones S.L.

Oporto Diaz, Samuel (2014). Taller de introducción a la base de datos NoSQL. Disponible en: <http://www.kasperu.com/courses/BG011/II/default.htm>

Pardo, Hugo. (2013). Planeta web 2.0. Capítulo 1. Nociones básicas alrededor de la Web 2.0. Recuperado de https://docs.google.com/document/d/1q6oul37iNKw3ELg0lhXlpaq6wGwOcUyXlzx_A0YT0U8/edit?hl=es&pli=1

Raineri, Andrés; Martínez, Andrea (1997) Diagnostico organizacional. Revista Academia, Universidad de Chile. Recuperado de [http://ww2.educarchile.cl/UserFiles/P0001/File/P0001_File_DO%20Un%20enfoque%20estrategico%20y%20pr%C3%A1ctico%20\(Raineri\),%2019%20mayo.pdf](http://ww2.educarchile.cl/UserFiles/P0001/File/P0001_File_DO%20Un%20enfoque%20estrategico%20y%20pr%C3%A1ctico%20(Raineri),%2019%20mayo.pdf)

Rodríguez Sánchez, Francisco (2014) Herramientas para Big Data: Entorno Hadoop. Escuela técnica superior de ingeniería de telecomunicaciones Universidad politécnica de Cartagena. Cartagena, España. Recuperado de <http://repositorio.upct.es/bitstream/handle/10317/4402/tfg482.pdf?sequence=1&isAllowed=y>

Román, M. (2014). *El fenómeno del Big Data en Chile y el mundo*. Obtenido de <http://www.telecheque.cl/noticias/el-fenomeno-del-big-data-en-chile-y-el-mundo/>

Sánchez, Jesús (2011) Método de diagnóstico. Recuperado de <http://es.slideshare.net/jesussanval/metodo-de-diagnostico>

Soares, Sunil (3 de junio de 2012) Not Your Type? Big Data Matchmaker On Five Data Types You Need To Explore Today. Dataversity. Recuperado de http://www.dataversity.net/not-your-type-big-data-matchmaker-on-five-data-types-you-need-to-explore-today/?cm_mc_uid=69942747005014585768846&cm_mc_sid_50200000=1465949454

Showers, Ben (2015) Library analytics and metrics: Using data to drive decisions and services. Recuperado de <http://www.facetpublishing.co.uk/title.php?id=049658#table-of-contents-tab>

Tilves, Mónica. (2013, 13 feb.). Que ocurre en internet cada 60 segundos [Infografía]. SILICON.es. Disponible en: <http://www.silicon.es/que-ocurre-en-internet-cada-60-segundos-infografia-33066>

Tuitt, Daniel (12 de octubre de 2013) A History of Big Data. Hcltech. Recuperado de <http://www.hcltech.com/blogs/history-big-data>

Van Den Berg, Eva. (2015, 31 mar). ¿Qué es eso del “big data”? Disponible en: http://elpais.com/elpais/2015/03/26/buenavida/1427382655_646798.html

Walker, Michael (2012, 28 nov). Data Veracity. Disponible en: <http://www.datasciencecentral.com/profiles/blogs/data-veracity>