



Maestría en Finanzas

FI-75301 Macrodatos y
Fintech

M.Sc. Walter Jeremías López.

Introducción al Big Data.



Maestría en Finanzas

FI-75301 Macrodatos y Fintech

M.Sc. Walter Jeremías López.

Objetivos de aprendizaje:

- Describir los macrodatos a nivel general con aplicaciones prácticas en negocios.
- Explicar las características del Big Data y su procesamiento.
- Analizar las implicaciones de generar datos masivos en las empresas.



Maestría en Finanzas

FI-75301 Macrodatos y Fintech

M.Sc. Walter Jeremías López.

Competencias a desarrollar:

- El alumno conoce la manera en que puede analizar macrodatos en su empresa.
- El alumno explica las características y funcionamiento de Big Data para incorporarlo en los negocios.



Maestría en Finanzas

FI-75301 Macrodatos y Fintech.

M.Sc. Walter Jeremías López.

Agenda:

- Componentes, historia y definición de Big Data.
- Características, fuentes de datos y arquitecturas.
- Competencias, habilidades y perfiles para la gestión de datos.
- Tendencias tecnológicas y sociales del Big Data.
- Conclusiones.

¿Qué es el Big Data?

En qué consisten los datos masivos o macrodatos.

1984 – 2008 – Hoy.

Las palabras datos masivos remiten a la acumulación de un considerable volumen de datos, el cuál marca su primera característica y una de las limitaciones de las aplicaciones que en las últimas décadas gestionaban datos. (García-Alsina, 2017).

Son un conjunto de datos que exceden la capacidad de procesamiento de las bases de datos convencionales dentro de un tiempo adecuado. (Chen et al, 2014).

Discusión:

¿Qué ha facilitado toda esta explosión y generación masiva de datos?

60 segundos en el universo online

Acciones que sucedieron en Internet en 60 segundos en 2017



Tipos de fuentes de Big Data.

Web y medios sociales

- Datos de flujos de clicks
- *Feeds* de Twitter
- Entradas de Facebook
- Contenido Web

Máquina a máquina

- Lecturas medidores inteligentes
- Lecturas RFID
- Lecturas sensores plataformas petroleras
- Señales GPS

Datos de transacciones grandes

- Demandas de salud
- Llamadas de telecomunicaciones
- Registros de detalles
- Registros de facturación

Biometría

- Reconocimiento facial
- Genética

Generado por los humanos

- Registros de voz de centros de llamada
- Correo electrónico
- Registros médicos electrónicos

Tipos de datos obtenidos y aplicaciones:

Datos de la Web:

- Historiales de navegación.
- Comportamientos y preferencias de los clientes: compras, visitas, reproducciones de videos y audios, descargas, lecturas, comentarios, búsquedas.

Datos de Texto:

- Correos, mensajería, chats, tweets y publicaciones en social media.
- Análisis de texto: sintáctico (parsing) y semántico (significados de palabras o frases).
- NPL (Natural Processing Language).
- Análisis de texto: sentiment análisis o minería de opinión.
- Reconocimiento de patrones y acciones legales.

Tipos de datos obtenidos y aplicaciones:

Datos de sensores:

- Temperatura, fuerza, velocidad, caudal, táctiles, presión, magnéticos, de presencia, de proximidad.
- En máquinas, medios de transporte, dispositivos, electrodomésticos, ropa y accesorios (weareables)

Datos de Geolocalización (Posición y tiempo):

- Registran la ubicación de un usuario y sus movimientos.
- Localizar clientes.
- Monitorear rutas de contactos en tiempo real e histórico.
- Enviar mensajes y recibir notificaciones de contactos cercanos.
- Son sensibles a la privacidad de los usuarios, clientes o proveedores.

Tipos de datos obtenidos y aplicaciones:

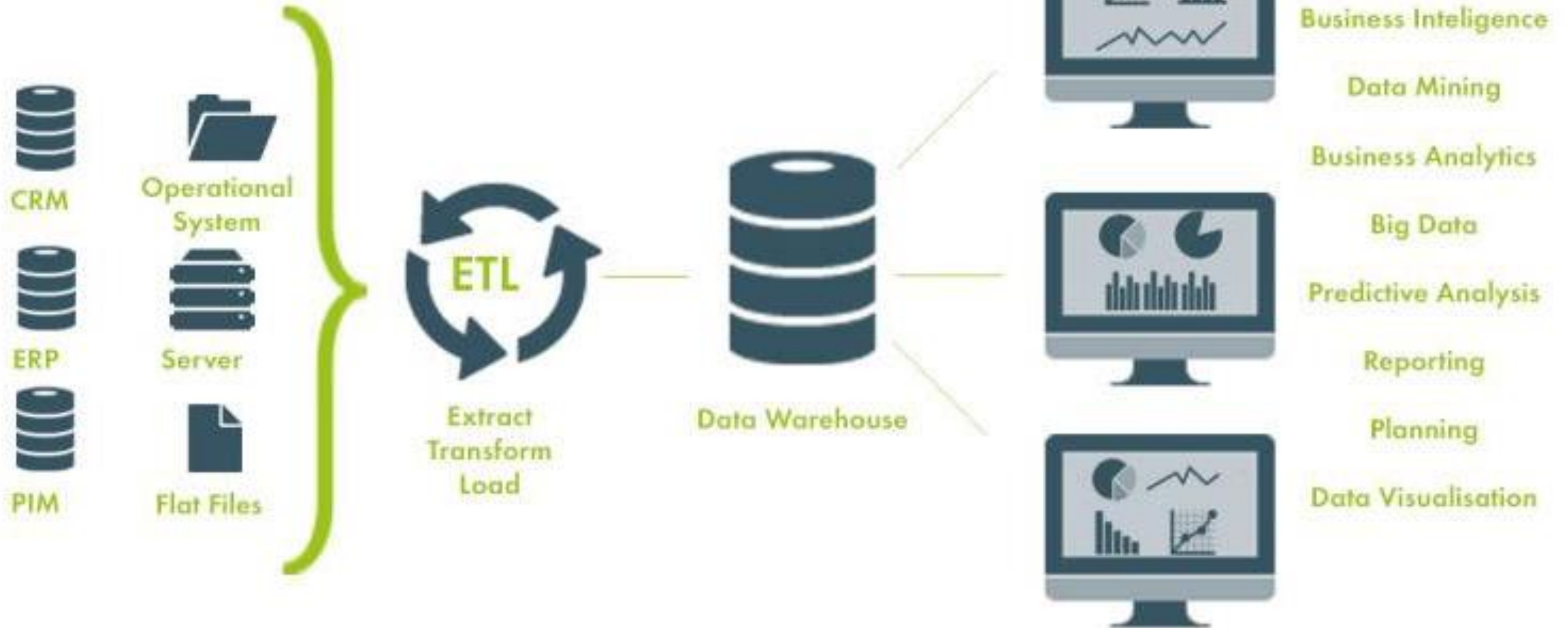
Datos de RFID:

- Etiquetado de productos y mercancías.
- Control de calidad y de inventarios (stocks).
- Logística, localización y seguimiento de objetos de paquetes y mensajería.
- Seguridad, detección de falsificaciones y control de robos.

Datos de NFC:

- Transferencia de datos entre dispositivos: envío de tarjetas personales, sincronización de aplicaciones, compartir fotos, videos, libros, documentos y otros archivos.
- Sistemas de pago por móviles (emulando tarjetas de crédito).
- Pagos entre personas.
- Control de accesos.

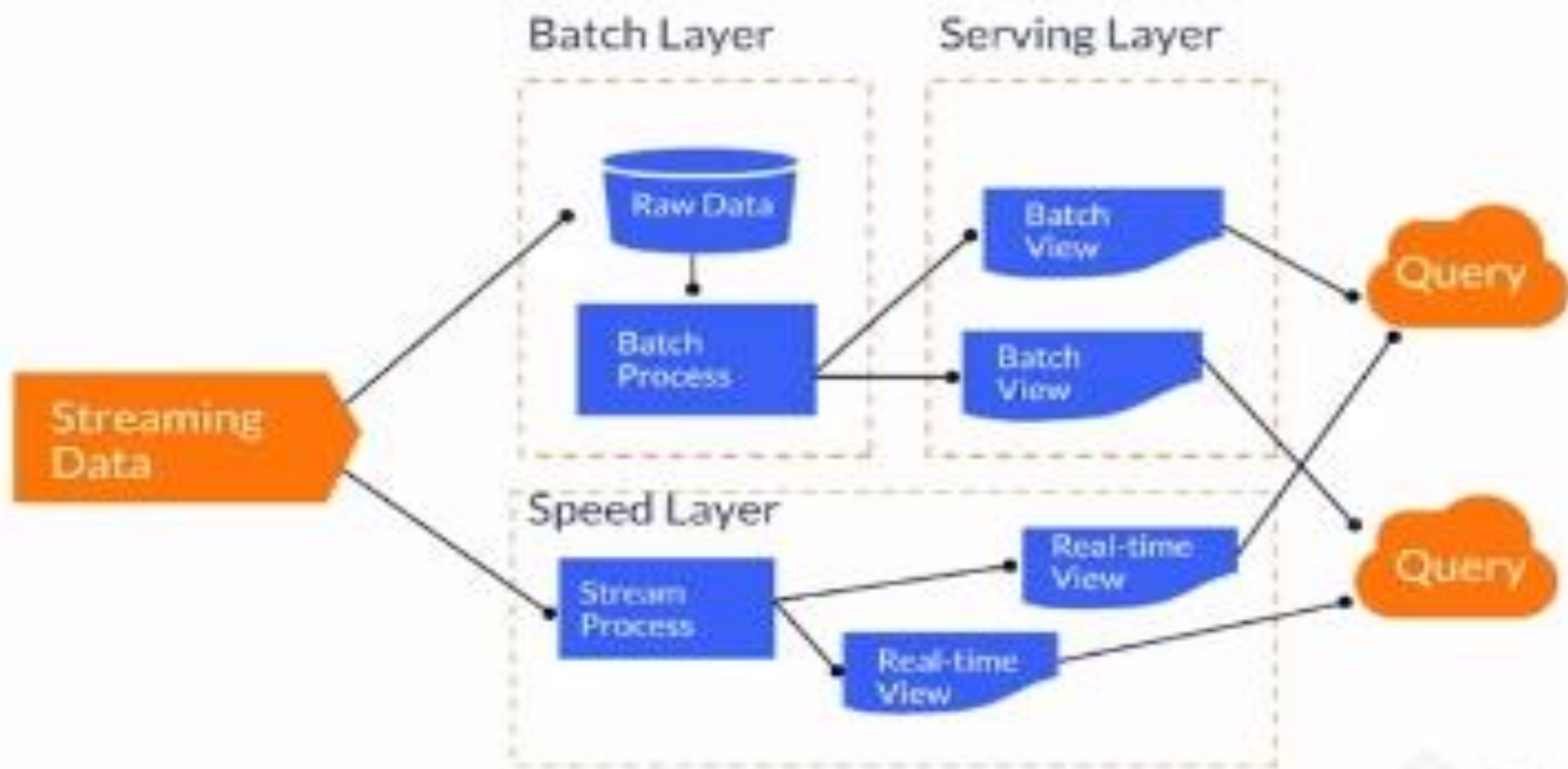
Fases del Big Data



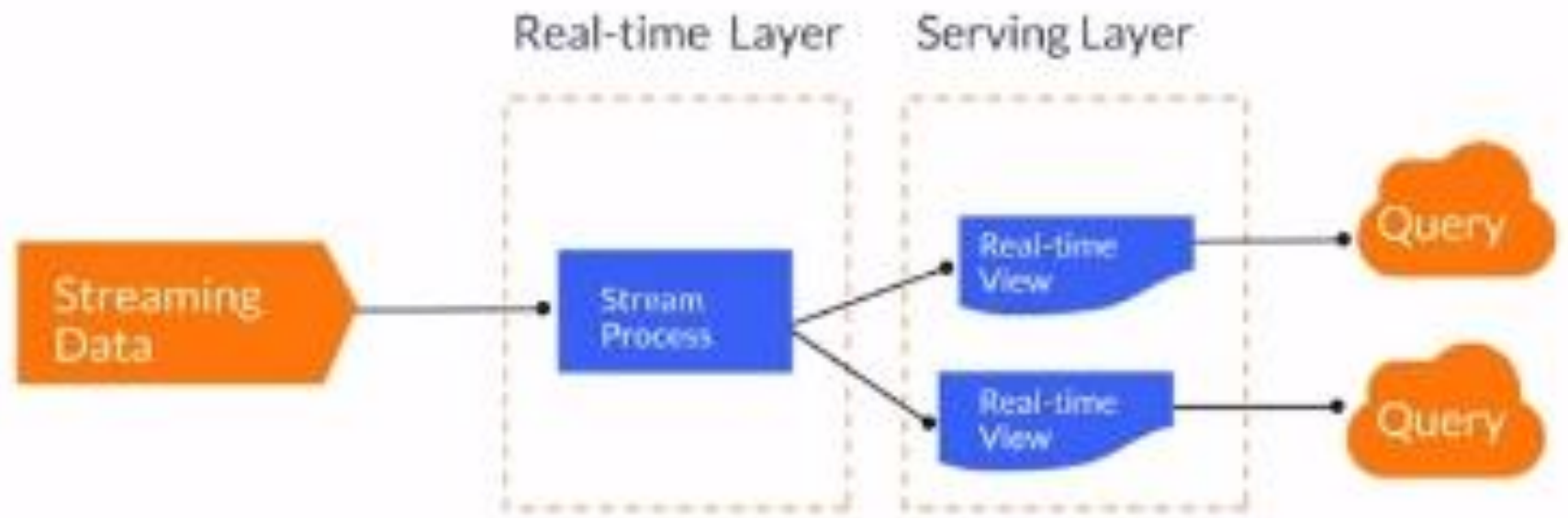
Proceso ETL del Big Data



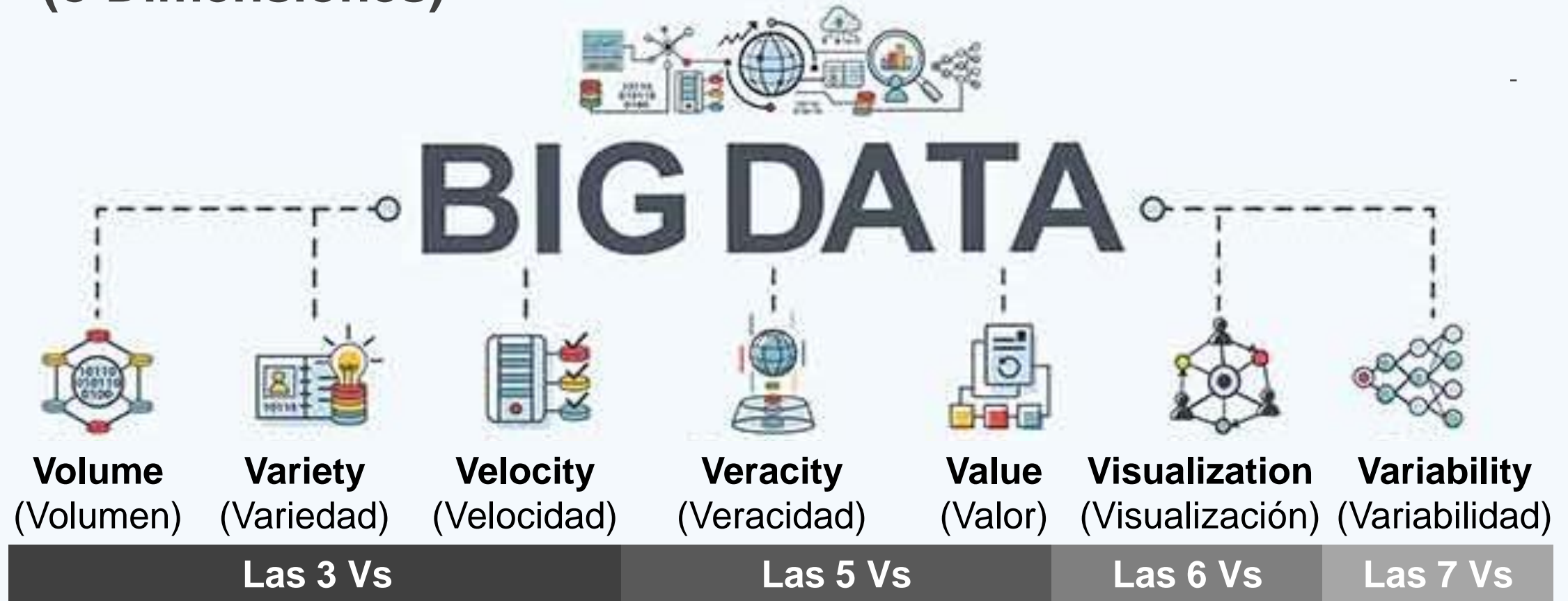
Arquitectura Lambda



Arquitectura Kappa



Características del Big Data: Su evolución. (o Dimensiones)



Y ya se está discutiendo la octava V y aún más: Verificación, Viabilidad, Viralidad y Viscosidad.

Dimensiones: Volumen

Características de los datos masivos, conocidas como: las Vs del Big Data.

Se refiere a dos aspectos:

Al incremento exponencial de los datos, fruto de las nuevas tecnologías y la facilidad de generar datos digitales.

Los retos que implica recoger, almacenar, procesar e integrar grandes volúmenes de datos de fuentes variadas y distribuidas.

Dimensiones: Volumen

Características de
los datos masivos,
conocidas como: las
Vs del Big Data.

Entendamos los tamaños:

Unidades y Prefijos para base 10 y 2:

- bit y Byte.
- Kilo/Kibi.
- Mega/Mebi.
- Giga/Gibi.
- Tera/Tebi.
- Peta/Pebi.
- Exa/Exbi.
- Zeta/Zebi.
- Yotta/Yobi.

La diferencia está en la base del sistema numérico:

- Si es decimal es base 10.
- Si es binario es base 2.
- Ejemplos:
- 1 Kilobyte = 10^3 bytes
1,000 bytes.
- 1 Kibibyte = 2^{10} bytes
1,024 bytes.

Dimensiones: Variedad

Características de los datos masivos, conocidas como: las Vs del Big Data.

1) Los macrodatos son variados en cuanto a:

Tipologías, formatos y estructuras para su organización y presentación.

2) Proviene de y van hacia diversas fuentes:

Móviles, audio, video, GPS, sensores y medidores, RFID, NFC, tweets, social media, logs, web, IoT, M2M.

Dimensiones: Variedad

Características de los datos masivos, conocidas como: las Vs del Big Data.

3) Según su nivel de estructuración pueden ser:

- Estructurados: son datos elementales, organizados en bases de datos relacionales.
- Semiestructurados: Tienen algún tipo de estructura implícita autodefinida.
- No estructurados: Se documentan en el formato que fueron creados, sin estructura predefinida.

Dimensiones: Velocidad

Características de los datos masivos, conocidas como: las Vs del Big Data.

Se refiere a dos aspectos:

A la velocidad con la que son generados.

A la velocidad de carga y procesamiento para la demanda del usuario.

“Lo más antes posible, o en tiempo real”.

Dimensiones: Veracidad

Características de los datos masivos, conocidas como: las Vs del Big Data.

- 1) Deben ser fiables, pues sirven para tomar decisiones y crear conocimiento para la acción.
- 2) Deben proporcionar confianza, exactitud y certeza. Asegurarse que no hay datos erróneos.
- 3) Hay que hacer una actualización, limpieza y revisión preliminar de los datos.

Dimensiones: Valor

Características de los datos masivos, conocidas como: las Vs del Big Data.

- 1) Crean conocimiento que es fuente de innovación, competitividad y productividad.
- 2) Deben crear valor a la organización para justificar los esfuerzos invertidos.
- 3) ¡Son para ocuparlos!

Dimensiones: Visualización

Características de los datos masivos, conocidas como: las Vs del Big Data.

- 1) Presentación visual, práctica, dinámica, interactiva y comprensible.
- 2) El campo que explora soluciones de visualización se conoce como *visual analytics* (VA). Busca convertir el exceso de información en una oportunidad.

Dimensiones: Variabilidad

Características de los datos masivos, conocidas como: las Vs del Big Data.

Relacionada con la velocidad con la que se producen los datos:

La generación de nuevos datos puede ocasionar la obsolescencia de otros.

Se debe gestionar en términos de eficiencia, al mismo nivel que la veracidad.

Dimensiones: Verificación

Características de los datos masivos, conocidas como: las Vs del Big Data.

Relacionada con la seguridad e integridad de los datos, especialmente cuando tienen procedencia externa.

Se trata de asegurar la integridad de los datos mediante: mensajes de autenticación, firmas digitales, certificados de terceros, etc.

Dimensiones: Viabilidad

Características de los datos masivos, conocidas como: las Vs del Big Data.

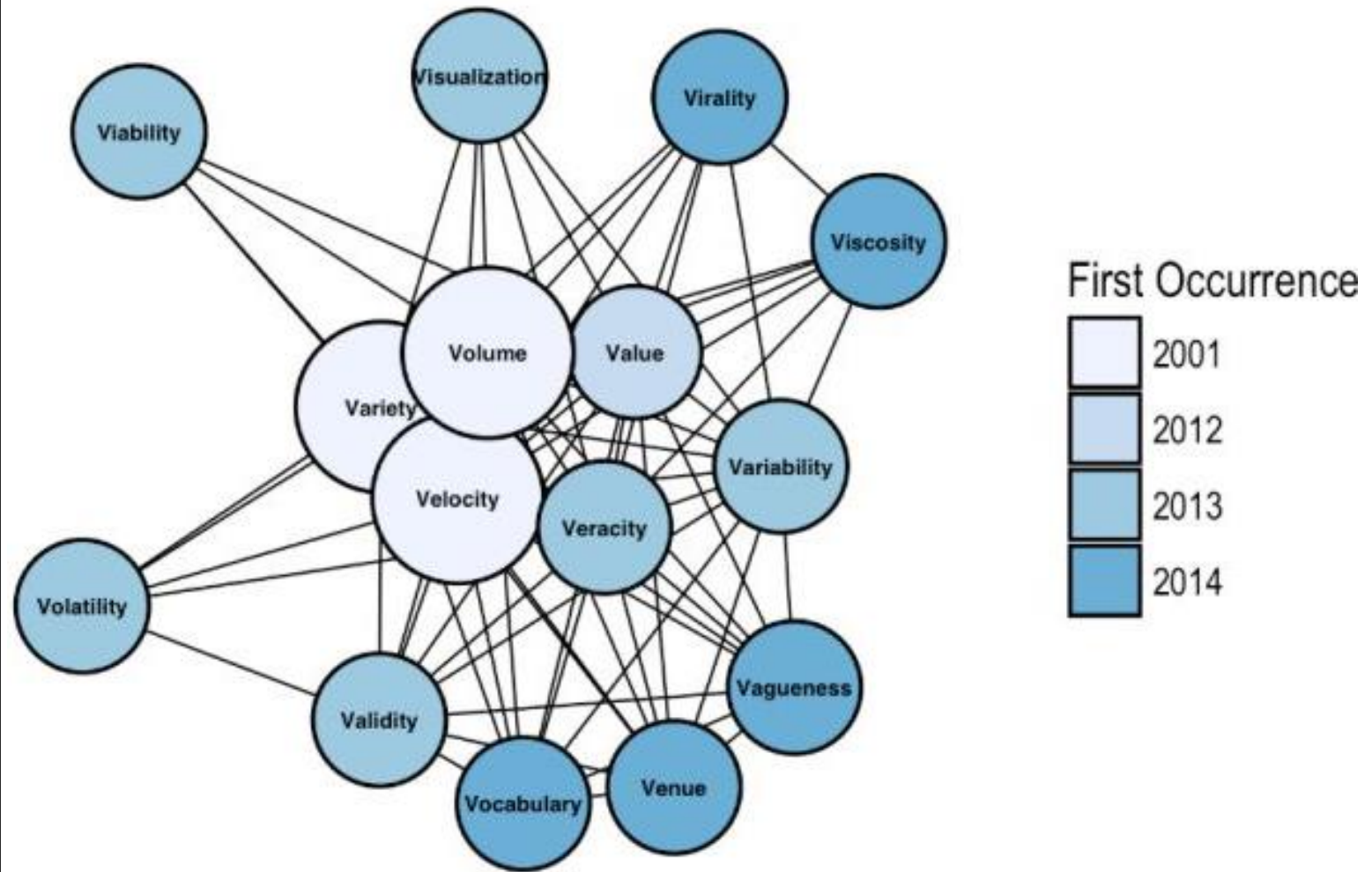
Dada la capacidad y complejidad de las plataformas para gestionar datos masivos se tiene los siguiente:

- Los presupuestos pueden llegar a ser muy elevados.
- Se deben adaptar al tamaño y necesidades de la organización.
- Se debe observar el costo.

¿Qué tan viable es hacerlo?

Dimensiones: Big Data

¿Cuántas Vs serán
necesarias para
categorizar los datos
masivos?



Lean el siguiente artículo:

[Las 42 Vs del Big Data.](#)

Habilidades y conocimientos

Competencias de los profesionales responsables para la gestión de Big Data.

Análisis de datos:

- Aplicar métodos estadísticos a los datos.
- Usar análisis predictivo para descubrir nuevas relaciones.
- Analizar conjuntos de datos complejos.
- Aplicar machine learning.
- Aplicar inteligencia de negocios.

Habilidades y conocimientos

Competencias de los profesionales responsables para la gestión de Big Data.

Desarrollo de SW y HW:

- Aplicar ingeniería para diseñar y desarrollar estructuras, instrumentos, máquinas, procesos y sistemas.
- Diseñar, construir y gestionar bases de datos relacionales o no relacionales.
- Aplicar soluciones informáticas a los problemas.
- Desarrollar algoritmos para analizar varias fuentes de datos de forma segura y fiable.

Habilidades y conocimientos

Competencias de los profesionales responsables para la gestión de Big Data.

Ciencia de datos (Data Science):

- Trasladar problemas de negocio desestructurados a un marco matemático abstracto.
- Usar datos para mejorar servicios existentes o desarrollar nuevos.
- Dar servicio de apoyo científico, técnico y analítico a las organizaciones.
- Analizar datos de clientes para identificar y optimizar acciones.

Habilidades y conocimientos

Competencias de los profesionales responsables para la gestión de Big Data.

Gestión de datos (Data Curation):

- Integrar y custodiar diferentes fuentes de datos y suministrarlas para análisis futuros.
- Mantener repositorios de análisis de datos históricos.
- Recoger y gestionar diversas fuentes de datos.
- Visualizar datos variables y complejos.

Habilidades y conocimientos

Competencias de los profesionales responsables para la gestión de Big Data.

Métodos de Investigación:

- Crear nuevos conocimientos y capacidades empleando métodos científicos.
- Estudiar sistemáticamente.
- Descubrir nuevos enfoques y aplicaciones.
- Realizar trabajo creativo y de innovación.
- Aplicar ingenio a problemas complejos y desarrollar ideas innovadoras.

Qué equipo de profesionales necesitas...

Para la transformación digital de tu empresa



EL CHIEF
INFORMATION OFFICER



EL TÉCNICO DE DATOS



EL DATA ARCHITECT



EL DATA SCIENTIST

Para diseñar estrategias de captación más eficientes



EL DATA ANALYST



EL CAMPAIGN EXPERT

Para aportar insights de negocio fiables a los distintos departamentos



EL DATA SCIENTIST



EL CHIEF
INFORMATION OFFICER

Perfiles profesionales para Big Data:

Perfiles	Detalles	Competencias	
		Técnicas	Metodológicas
Desarrollador de	Business Intelligence. Web. Software. Negocio. Analista. Aplicaciones. Bases de datos. Front-End.	No SQL. Java. SQL. Javascript. MySQL. Linux. Oracle. Hadoop. HTML.	Programación y diseño Web con CSS. OOP (Object Oriented). SOA (Service Oriented Architecture). Metodologías ágiles: TDD (Test Driven Development). BDD (Behavior-Driven Development). Agile. Scrum.
Arquitecto de	Soluciones. Datos. Business Intelligence.	Oracle. Java. SQL. Hadoop.	Data modelling. ETL. Enterprise architecture. Open Source.

Perfiles profesionales para Big Data:

Perfiles	Detalles	Competencias	
		Técnicas	Metodológicas
Analista de	Negocio. Datos. Business Intelligence. Apoyo.	Oracle. BI e informes. SQL. Java.	Data modelling. ETL. Análisis de datos.
Administrador de	Sistemas (SysAdmin). Bases de datos (DBA).	Linux. MySQL. Puppet. Hadoop. Oracle	Configuración de sistemas. Virtualización y Clustering. ETL. Recuperación de desastres: DRP (Disaster Recovery Plan) BCP (Business Continuity Plan)

Perfiles profesionales para Big Data:

Perfiles	Detalles	Competencias	
		Técnicas	Metodológicas
Gestor de proyectos.	Oracle. Technical Project. Managers. Business Intelligence.	Oracle (BI EE, EBS, R12). Netezza, Business Objects. Hyperion.	ETL. Desarrollo de Software ágil. PRINCE2 (PROjects IN Controlled Environments v2). Gestión de stakeholders.
Diseñador		Oracle (BIEE). SQL. Netezza. UNIX	ETL. Data modelling. CSS. Data integration, mining y warehouse. Migración y analítica.

Perfiles profesionales para Big Data:

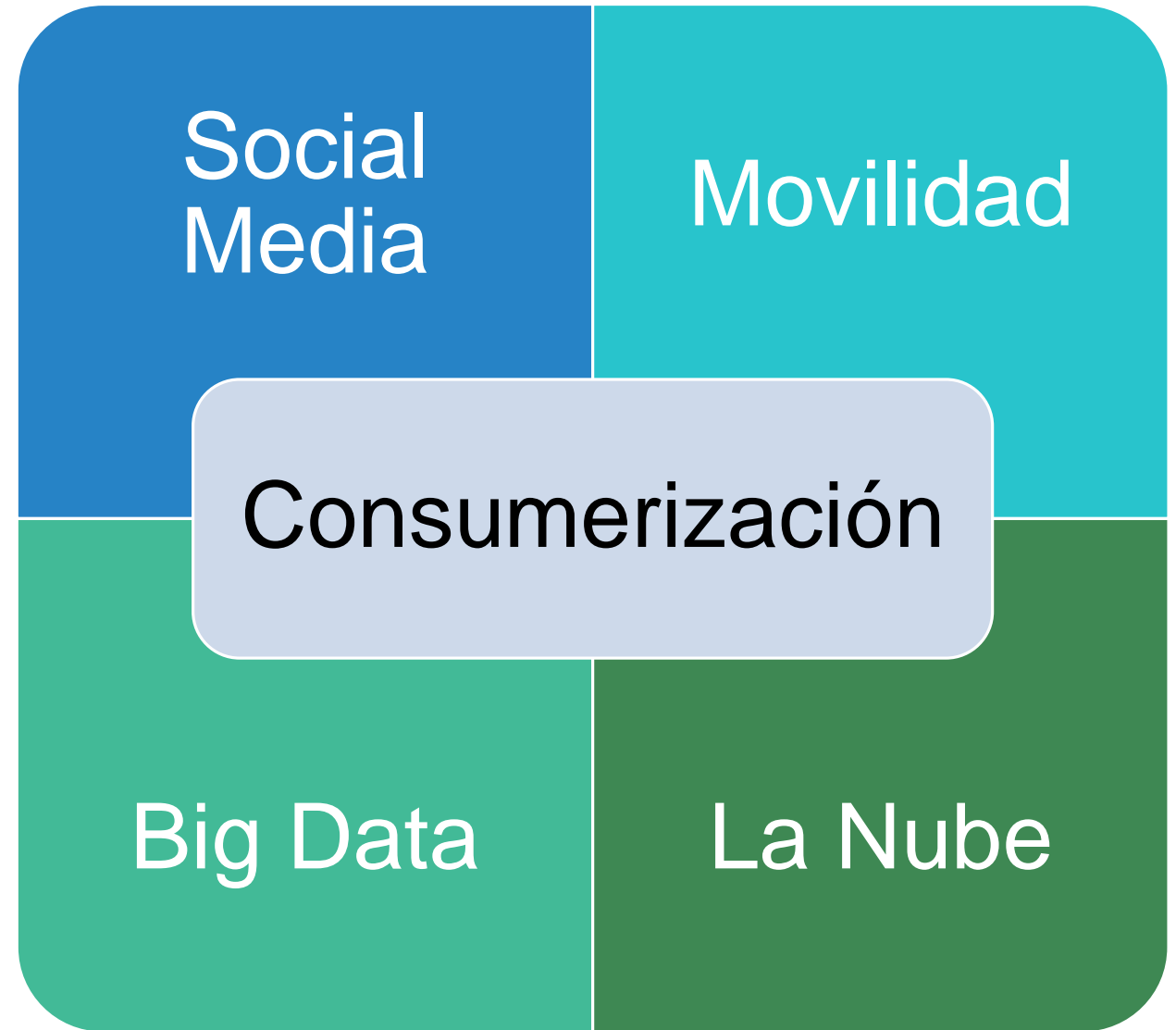
Perfiles	Detalles	Competencias	
		Técnicas	Metodológicas
Científico de datos		MapReduce. Hadoop. Java. No SQL. Python R	Estadística. Análisis. Matemáticas. Data analytics. Inteligencia Artificial. Machine y Deep Learning. Minería de datos. Creación de modelos. MPP (Massively Parallel Procesing).

Perfiles profesionales para Big Data:



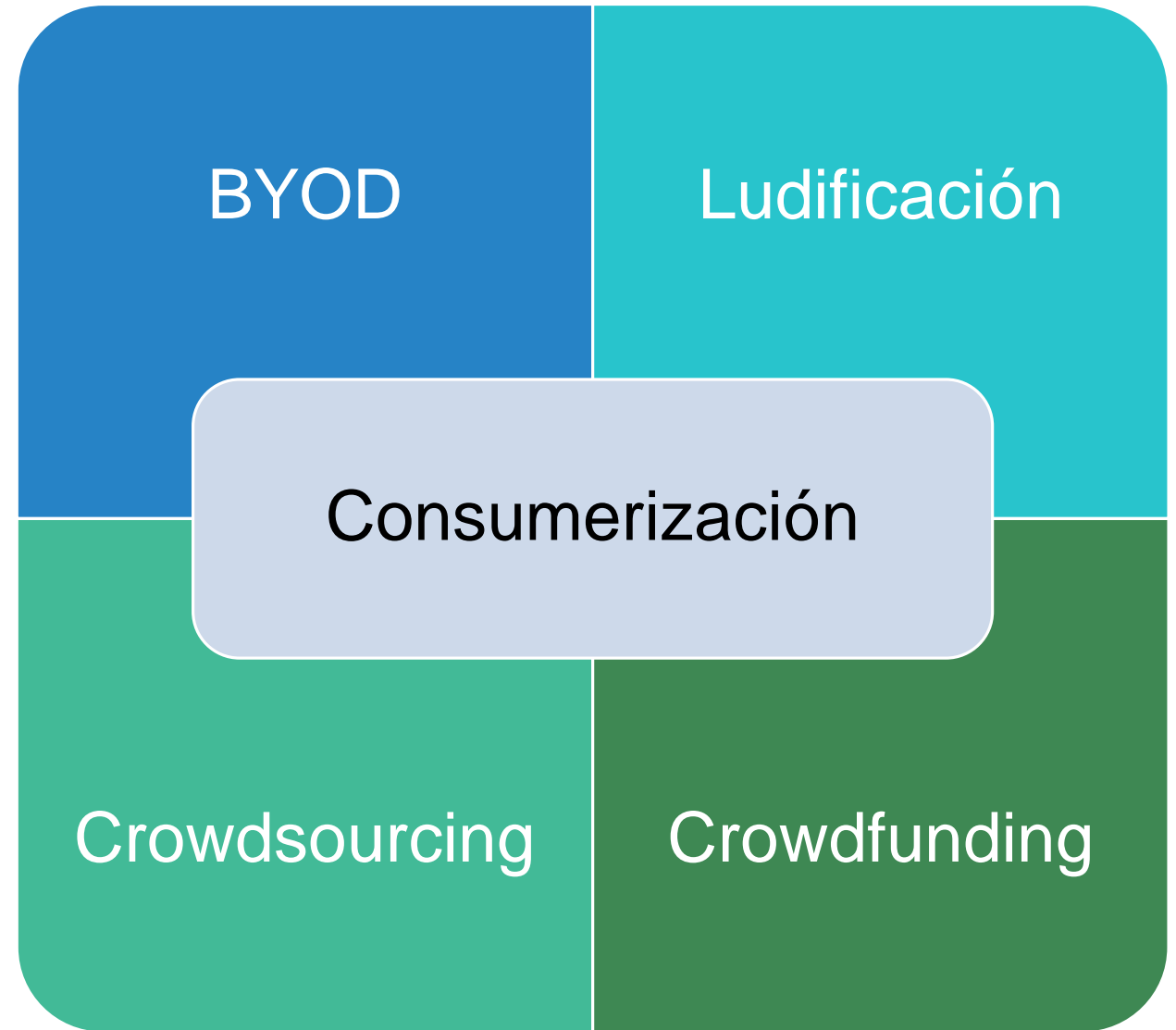
Tendencias Tecnológicas

¿Cuáles son las 4 grandes tecnologías del futuro y cómo convergen?



Tendencias Tecnológicas

¿Cuáles son las principales tendencias que se desprenden de la consumerización de TI?



Tendencias en Big Data

El futuro de los datos masivos.

Dominios sin explorar:

- Datos sin analizar y exceso de información.

Protección de datos:

- Seguridad de los datos, leyes y políticas de privacidad.

Países emergentes:

- Comenzarán a generar y analizar macrodatos en empresas y gobierno.

Gobierno de datos:

- Plataformas llave en mano para analítica end-to-end: predictiva, M2M y de sentimientos.

Big Data trata de cantidades masivas de datos provenientes de distintas fuentes y cuyas características se resumen en las 7Vs: Volumen, Variedad, Velocidad, Veracidad, Valor, Visualización y Variabilidad. Aunque están surgiendo más dimensiones.

Conclusiones

Las Fases del Big Data se conocen como el proceso ETL: Extraer, Transformar y Cargar.

Las arquitecturas del Big Data son: Lambda y Kappa.

Conclusiones

Los perfiles profesionales que están surgiendo son: analistas de datos, desarrolladores de SW y HW (con énfasis en aplicaciones para Big Data) y científicos de datos, que es la nueva carrera de este siglo: Data Science – Investigación y Modelos.

Conclusiones

Estos perfiles requieren muchas competencias y habilidades especializadas.

En el futuro las principales tendencias serán: Big Data, Movilidad, Social Media y La Nube, orientadas hacia la consumerización de TI, lo cual cambiará la forma en que vivimos.

Conclusiones



Maestría en Finanzas

FI-75301 Macrodatos y Fintech.

M.Sc. Walter Jeremías López.



¡Gracias por su atención!

¿Preguntas o comentarios?