



Estructura de Datos II

BIG DATA

Integrantes:

David Reyes AR364551

Rubén González 8-1021-1604

Urbano Iguala 9-744-1120

Francisco Caballero 8-863-2214

Introducción

Big Data bien puede ser la próxima gran cosa en el mundo de TI.

Los grandes datos irrumpieron en escena en la primera década del siglo XXI.

Las primeras organizaciones en adoptarlo fueron empresas en línea y nuevas empresas (Google, eBay, LinkedIn y Facebook)

Al igual que muchas nuevas tecnologías de la información, los macro datos pueden generar reducciones drásticas de costos, mejoras sustanciales en el tiempo requerido para realizar una tarea informática o nuevas ofertas de productos y servicios.

BIG DATA



El acto de acceder y almacenar grandes cantidades de información para la analítica ha existido desde hace mucho tiempo.



Big Data – Las Tres V

Volumen

- Las organizaciones recopilan datos de diversas fuentes, como transacciones comerciales, dispositivos inteligentes (IO), equipo industrial, vídeos, medios sociales y más. En el pasado, su almacenamiento habría sido un problema - pero el almacenamiento más barato en plataformas como los data lakes y el Hadoop han aliviado la carga.

Velocidad

- Con el crecimiento del Internet de las Cosas, los datos llegan a las empresas a una velocidad sin precedentes y deben ser manejados de manera oportuna. Las etiquetas RFID, los sensores y los medidores inteligentes están impulsando la necesidad de manejar estos torrentes de datos en tiempo casi real.

Variedad

- Los datos se presentan en todo tipo de formatos: desde datos numéricos estructurados en bases de datos tradicionales hasta documentos de texto no estructurados, correos electrónicos, vídeos, audios, datos de teletipo y transacciones financieras.

Big Data – Fuente de Datos



Apps

Vertical Apps

AGILONE Atigeo ellucian.
emcien MYRRIX Placed.
PROBETIX PRACTICE Quantivo NG@DATA
wibi: data

Operational Intelligence

VITRIA loggly splunk
sumologic TIBCO

Ad / Media Apps

IPONWEB bloomreach
collective DataXu
Media Science METAMARKETS Recorded Future
TURN

Business Intelligence

ATTIVO Autonomy bime
birst Business Objects Chart.io
COGNOS DOMO GoodData
IBM JASPERSOFT MicroStrategy
pentaho RECOMMIND SAP
SiSense

Analytics And Visualization

1010data alteryx AYATA
centrifuge CETAS CIRRO
ClearStory Datameer dataspora
emcien KARMASHERE metaLayer
OPERA Palantir panopticon
platforma QlikView RJMetrics
saffron sas +tableau
visual ly

Data As A Service

bluekai DATASIFT Gnip
factual FICO Gnip
INRIX kaggle knoema
LexisNexis LOGATE SPACE CURVE

Infrastructure

Analytics Infrastructure

calpont cloudera EXASOL
GREENPLUM HADAPT Hortonworks
INFOBRIGHT INFORMATICA Kognitio
MAPR PARACCEL VERTICA

Operational Infrastructure

10gen KEROPIKE COUCHBASE
DATASTAX MarkLogic TERRACOTTA
VoltDB

Infrastructure As A Service

CONTINUITY infochimps
MORTAR Qubole

Structured Databases

ALTBASE IBM DB2 SQL Server
MySQL ORACLE PostgreSQL
SYBASE

Technologies

HBASE Cassandra Hadoop

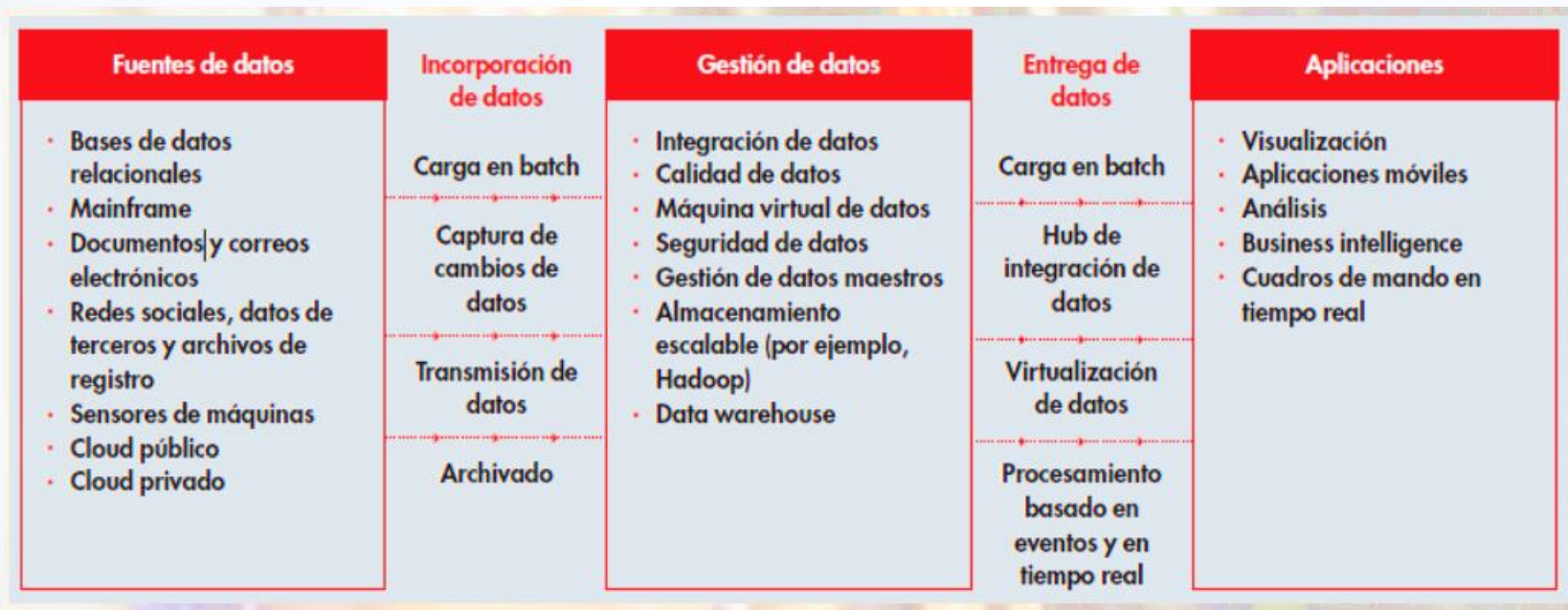
- Integración de datos: Preparar y consolidar estructuras y fuentes diversas en un conjunto de datos cohesionado para el análisis.
- Calidad de datos: Limpiar con fiabilidad los datos así como eliminar los duplicados y los errores.
- Seguridad de datos: Detectar y proteger los datos en todos los almacenes de datos asignando reglas relativas al uso, al acceso y a los permisos.
- Máquina virtual de datos: Crear una capa de abstracción para los datos que asocie con amplitud el procesamiento de datos desde el entorno de implantación subyacente.
- Gestión de datos maestros: Almacenar una única versión consolidada, completa y fidedigna de la realidad de los diversos dominios de datos.
- Estructura de datos distribuida: Emplear una tecnología como Hadoop para escalar de forma rentable las exigencias de almacenamiento y procesamiento.
- Data warehouse: Tecnología de data warehouse que admita los requisitos de rendimiento, uso y escalabilidad que tienen los análisis de Big Data y las integraciones con las infraestructuras de Hadoop.

Big Data – Entrega de Datos

- Carga en batch: Escalar con eficiencia la carga en batch de los datos entre los sistemas back-end operativos, los análisis y las fuentes.
- Transmisión en tiempo real: Transmisión de datos en tiempo real a las aplicaciones, los análisis y los sistemas back-end que la requieren.
- Hub de integración de datos: Permite que los datos estén disponibles mediante un enfoque como el del modelo de publicación y suscripción a fin de evitar la proliferación de integraciones de punto a punto.
- Virtualización de datos: Entregar datos de sus sistemas sin sobrecargarlos.
- Procesamiento basado en eventos: Detectar y analizar amenazas, oportunidades y otros eventos críticos para el negocio y darles una respuesta en tiempo real.



Big Data – Arquitectura de Big Data



Big Data – Ventajas del Big Data

Mejora en la toma de decisiones. Disponer de un gran volumen de datos estructurados que se puedan interpretar ayuda a las organizaciones a poder tomar una decisión.

Feedback a tiempo real. La tecnología Big Data permite no solo almacenar y procesar datos. También supone recibir datos a tiempo real y contar con la información necesaria rápidamente. El Big Data es por encima de todo una tecnología ágil y veloz que permite, por ejemplo, obtener información a tiempo real del lanzamiento de un producto o el resultado de una estrategia

Conocimiento del mercado. La obtención y seguimiento de datos del Big Data permite conocer a los consumidores y adelantarse a sus necesidades, así como detectar tendencias en consumo o fallos en la atención al cliente. La inclusión de profesionales con un Máster en Big data supone a permite a las empresas localizar de posibles oportunidades mediante el tratamiento de estos datos estructurados y comparables.

Tecnología del presente y del futuro. La ciencia de datos comienza a ser un elemento diferenciador y de éxito entre las empresas. También empiezan a experimentarse sus ventajas de aplicación en otros campos como la medicina. La tecnología del Big Data está en constante evolución y todo apunta a que jugará un papel todavía más importante en la toma de decisiones futuras.

1. El exceso de datos

La utilidad del Big Data surge de su capacidad para extraer información de los datos. Si estos datos **se almacenan y analizan indiscriminadamente solo se convierten en ruido que ralentizará y obstaculizará las tareas**. Los profesionales con Máster Big Data obtienen la capacidad de discernir y seleccionar las datos y patrones que serán de utilidad.

2. Ciberseguridad

La seguridad es el mayor reto para la tecnología Big Data. **La vulnerabilidad de esta enorme cantidad de datos almacenados es el objetivo de los nuevos ciberataques**. Reconocer la responsabilidad de trabajar con datos y dotar los sistemas de normas y herramientas, son aspectos clave de la formación en un Máster en Big data.

3. Las políticas de protección de datos

En relación con lo anterior, y como ocurre con muchos de los grandes avances tecnológicos, **son necesarios consensos y protocolos** que establezcan hasta qué punto esta tecnología podría considerarse intrusiva y cómo protegerla.

4. Tecnofobia

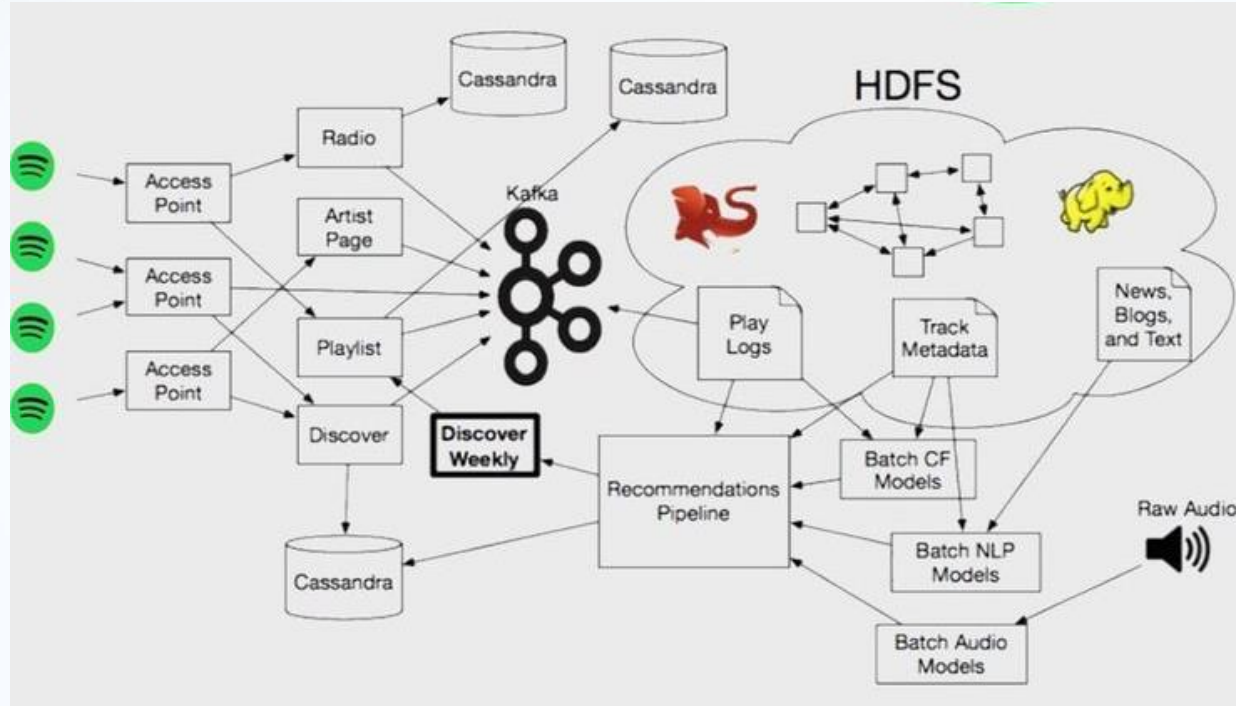
Más como consecuencia de los mitos sobre el Big Data que por sus desventajas, **existen reacciones desfavorables por parte de los usuarios, que consideran que vulnera su privacidad**. Lo que estos usuarios probablemente no sepan, es que la tecnología de datos lleva años entre sus acciones cotidianas permitiendo avances destinados a facilitar su día a día.

Big Data – Casos de Exito



Cuando un canal de TV lanza una nueva serie, esta tiene tan solo un 35% de posibilidades de triunfar, pero cuando lo hace Netflix, lo hace sus posibilidades ascienden al 70% (Apostando por el éxito: “House of Cards” o «Narcos»)

Big Data – Casos de Exito



Descubrimiento Semanal de Spotify. Una playlist con 30 canciones que la plataforma ofrece semanalmente con el objetivo de recomendar nuevas canciones a sus usuarios permitiéndoles explorar nuevos horizontes musicales.

- El uso de big data en el mundo real - IBM Global Business Services
- El gran cuaderno del Big Data – Informatica
- Inteligencia Artificial, Inteligencia Computacional y Big Data – Francisco Herrera
- “El análisis masivo de datos está transformando a la empresa”, John Veihmeyer, Presidente mundial de KPMG (Miguel Ángel García. El País, 25.01.2015)
http://economia.elpais.com/economia/2015/01/22/actualidad/1421926110_795741.html
- La ciencia de los datos y el conocimiento en la empresa (Fernando Piera) (CEDE, Cuadernos de Gestión del Conocimiento Empresarial nº50. Noviembre 2014)
http://www.directivoscede.com/sites/default/files/document/conocimiento/20-11-2014/1411_num50_piera.pdf
- Big Data y Social Media en retrospectiva (Francisco Rangel. Autoritas, 2.04.2015)
<http://www.autoritas.net/es/big-data-y-social-media-en-retrospectiva/>
- Tendencias de análisis Big Data en comunicación digital (Francisco López y Marina Pulido. Top Comunicación, 10.02.2015)
<http://www.topcomunicacion.com/noticia/6261/cuales-son-las-principales-tendenciasde-analisis-big-data-en-comunicacion-digital>

¡GRACIAS!

