



Introducción a BigData II

ANDRES TRIANA
Microsoft Certified Trainer (MCT)



CONTACTO
andres.triana@qualitycode.com.co
319 506 0662
www.qualitycode.com.co



Temario del Curso

Modulo 1

Fundamentos

Modulo 2

Hadoop

Modulo 3

Data mining y Machine Learning

Modulo 4

NoSql - MongoDB

BIG DATA



Los 11 elementos de una solución Big Data

- Cómo prepararse, Recopilación de datos, Análisis, Agregación de valor y Componentes que no pueden faltar



Cómo preparase para implementar una solución Big Data

. Fase inicial - Estudio de viabilidad y propuesta

- **Detección de necesidades:** las carencias se ponen de manifiesto bien por comparación entre las nuevas tecnologías y las existentes en la empresa, bien cuando en el curso operativo se perciben limitaciones que impiden avanzar. Estas necesidades suelen tener que ver con el volumen de datos a almacenar, su variedad, la velocidad de recogida, almacenamiento y procesamiento; y su escalabilidad horizontal.
- **Justificación de la inversión:** la inversión se justifica por cuestiones técnicas, que con la implementación de una solución big data se pretende atajar, al hacerse posible la evolución y la innovación, en un entorno de más alto rendimiento que a la vez consigue garantizar el ahorro de costes.
- **Evaluación de limitaciones:** además de tomar en consideración la infraestructura de la empresa, su madurez tecnológica y sus posibilidades (presupuesto y recursos, entre otros); también hay que valorar las cuestiones legales, en especial en lo concerniente a las exigencias en materia de privacidad de datos.

Cómo preparase para implementar una solución Big Data

Fase de planificación - determinación de los recursos intervinientes

Cada organización habrá de determinar el presupuesto destinado al proyecto, que se establecerá en función de su capacidad económica y necesidades

Roles:

- *Gestores:* sponsors, director de proyecto, coordinadores y gestores de calidad.
- *Diseñadores y arquitectos de datos:* perfiles muy técnicos y con una visión clara de los objetivos del proyecto de implementación.
- *Implementadores:* analistas y desarrolladores, personal cualificado con conocimientos en tecnología y en el negocio.
- *Operadores de datos:* tanto de entrada, como intermedios y de resultado.

Cómo preparase para implementar una solución Big Data

Fase de diseño de la infraestructura de la solución - arquitectura

Hay que tener muy en cuenta las distintas opciones del mercado, su escalabilidad y coste. Pero, sobre todo, es necesario garantizar la escalabilidad y condiciones de calidad

- *Infraestructura*: es el soporte físico de la solución y consta de elementos como servidores, redes o equipos.
- *Arquitectura*: es su esqueleto lógico, que se compone de comunicaciones, procedimientos y protocolos, entre otros.

Cómo preparase para implementar una solución Big Data

Fase de implementación y puesta en marcha

En esta etapa se lleva a cabo el despliegue y se ponen en marcha la gestión y gobierno de la solución

A) Instalación y puesta en marcha de la infraestructura: servidores y sus componentes.

B) Implementación de la arquitectura: configurando la infraestructura para su correcto funcionamiento y capacidad de interacción con los sistemas externos ya existentes en la empresa (comunicación con orígenes de datos, destinos de datos y herramientas de reporting).

C) Puesta en marcha operativa: ingesta, transformaciones de datos y explotación de datos.

Recopilación de datos

La parte más importante de una solución big data son los datos que contiene. Estos datos se transformarán en información y, mediante las técnicas adecuadas serán capaces de proporcionar conocimiento.

A) Evaluación de datos: comienza por el estudio de la utilidad de los datos (valorando especialmente si conviene incorporar los que no tienen una utilidad inmediata); continúa con la evaluación de su volumen y frecuencia de explotación y concluye con la definición de accesos y restricciones, necesarios para proteger la información sensible

B) Ingestión de datos para su explotación: el núcleo de esta acción es la preparación y estandarización de los datos.

C) Gestión de Datos: en lo relativo a su visibilidad, disponibilidad, mantenimiento y seguridad.

Analysis

La analítica es la parte nuclear de una solución big data. Para llevarla a cabo correctamente son necesarias dos acciones:

Generar los cálculos que se esperan de la solución: diseñando los algoritmos que se precisen.

Detectar datos y correlaciones de interés: mediante la intervención de especialistas que puedan descubrir tendencias, patrones, oportunidades y/o amenazas tras los datos

Agregación de valor

Partiendo del trabajo de los analistas en su investigación se pueden elaborar conclusiones que permitan identificar nuevas vías de avance para el negocio, fuente de innovación

- *Incorporando nuevos datos:* autogenerados, ajenos a la solución o incluso provenientes del exterior de la organización.
- *Diseñando nuevos proyectos:* para áreas distintas de la empresa o enfocados a clientes. Estas iniciativas pueden resultar en el reemplazo de antiguas tecnologías o en el acceso a resultados a distinto nivel, lo que permitirían técnicas como el scoring de riesgos (especialmente indicadas para los sectores de banca y seguros), el análisis de logs (que hace posible un mejor control, optimización de la seguridad y accesos o de el aumento de efectividad en la gestión de errores) o el marketing personalizado (muy indicado para empresas del sector retail).
- *Explotando los resultados obtenidos:* enfocados a nuevas campañas de marketing.

Componentes que no pueden faltar en una solución Big Data

- Fuentes de big data
- Capa de Almacenamiento
- Capa de Analisis
- Capa de Consumo







Estrategias de Big Data en el sector Retail

Descubre algunas ventajas competitivas en un mercado competitivo y saturado

La unión de los avances tecnológicos y las nuevas técnicas de analítica predictiva permiten:

- Conocer todos los tickets que un gran supermercado con presencia a nivel nacional ha generado en un día.
- Saber lo que va en la cesta de la compra de cada persona.
- Conocer a qué hora se compra qué.
- Descubrir qué más productos adquieren quienes van a comprar cerveza.
- Monitorizar los picos de entrada al parking de un determinado centro comercial

El sector retail y los grandes datos



Se obtienen de:

- Sus teléfonos.
- Las compras presenciales.
- El shopping en línea.
- Los pagos que realizan.
- Las publicaciones en las redes sociales.

El sector retail y los grandes datos

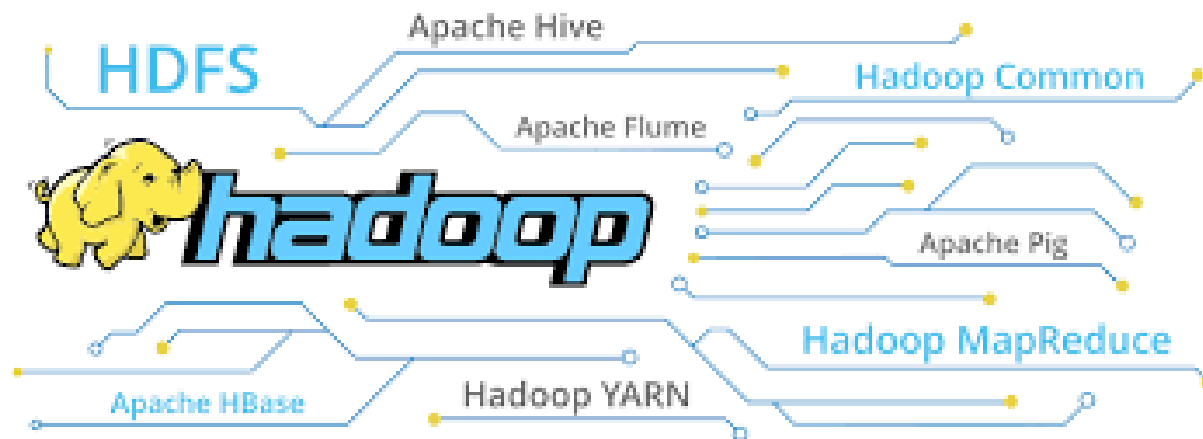
Esta gran masa de datos en aumento se traduce en conocimiento.

Quienes pueden beneficiarse de él están en disposición de:

- Alcanzar una visión más completa del cliente.
- Lograr una mayor penetración de los mercados.
- Contar con más criterios para identificar y clasificar los riesgos.

Big data retail y la tecnología

La aparición de **Hadoop** marca una nueva forma de enfrentarse a big data. Los desafíos que plantean el procesamiento, la recogida e incluso el almacenamiento de grandes datos se ven mitigados por esta tecnología y sus posibilidades.



Big data retail y la tecnología

La forma de poner en práctica este proceso en la realidad del sector retail es:

1. Implantar big data siguiendo una hoja de ruta lógica: hadoop en combinación con el lenguaje de programación R, que es open source, frente a SPSS o SAS que obligan a pagar licencias.
2. Almacenar datos (en condiciones de coste óptimo).
3. Guardar todos los datos que se considere necesario, sin desechar ninguno, en un entorno de tecnología Hadoop, que permite almacenar datos, tanto estructurados, como no estructurados.
4. Aplicar técnicas de perfilado, limpieza y modelado de datos.
5. Comenzar el trabajo con los datos, previa aplicación de métodos de análisis predictivo y minería de datos.



Big data retail y la tecnología

¿cómo se puede diferenciar cuál es el orden óptimo de las estanterías o cuántos niveles deben presentar?

¿De qué forma se pueden configurar las rutas de los clientes dentro de la tienda?

¿Cómo descubrir el tiempo que se necesita mantener una oferta 3x2 para que un producto se consolide en la cesta de compra de los clientes?

¿Qué técnicas ayudan a aumentar las ventas de las marcas blancas?



Big data retail y la tecnología

Planificación de la disposición de los productos

Determinar el tamaño óptimo de cada departamento.

Calcular su ubicación y distancia idónea con respecto a la entrada y a otros departamentos.

Big data retail y la tecnología

Optimización del surtido

Los diferentes productos y referencias se deben clasificar en función de:

- El beneficio que reportan.
- Su volumen de ventas.
- Las preferencias de los clientes en base a segmentos o en función de criterios individuales de clasificación.

Es necesario llevar a cabo un análisis en detalle para clasificar los productos y obtener información acerca de la función de cada uno. En este sentido, se pueden distinguir dos grandes grupos:

- Producto central: es el producto primario que los clientes han ido a comprar.
- Productos accesorios: se trata de complementos del producto central, en su mayoría productos de impulso y artículos que no son importantes, pero con potencial para llegar a convertirse en productos centrales.

Big data retail y la tecnología

Análisis de fidelidad

1. Identificación de los **clientes más fieles** a una marca o a un producto: para ello es necesario llevar a cabo un análisis de los datos del CRM.
2. Análisis de las transacciones de los usuarios: para confirmar su pertenencia a la categoría e clientes fieles y afianzar el conocimiento adquirido sobre sus hábitos.
3. Detección de **patrones de comportamiento**: serán la base para diseñar estrategias efectivas de retención de clientes.
4. **Segmentación de clientes**: en base al estudio de los datos demográficos, psicográficos y de conducta de compra.

Esta segmentación se aplica para estrategias en una doble vertiente

1. El **diseño de ofertas y estrategias de marketing** personalizadas.
2. La identificación de forma prematura de los primeros **signos de abandono** por parte del cliente, que marcan el inicio del lanzamiento de campañas de **retención de clientes** insatisfechos que les proporcionan un mejor servicio antes de que alcancen el punto de no retorno.

Big data retail y la tecnología

Análisis de la cesta de la compra

1. **Productos relacionados:** consiste en identificar qué artículos se adquieren de forma conjunta en cada compra.
2. **Modelos de cesta de la compra:** es el punto de partida para elaborar perfiles de clientes tipo.



Sector Financiero

Proyectos en los que se justifica una inversión

Los 5 desafíos que big data plantea a los bancos

- 1. Falta de integración en datos y metadatos:**
proviene de la existencia de demasiados almacenes de información, que provocan que su aprovechamiento no sea óptimo y sus niveles de rendimiento se queden en cifras muy inferiores a las que deberían alcanzarse

Los 5 desafíos que big data plantea a los bancos

- 2. Carencia de perfiles cualificados:** aunque la figura del data scientist es relativamente nueva, sorprende la ausencia de profesionales cualificados para la minería de datos, la estadística avanzada o el análisis de negocio de última generación

Los 5 desafíos que big data plantea a los bancos

3. Dificultades técnicas y de infraestructura:

que provocan que el tiempo dedicado al procesamiento de grandes conjuntos de datos se alargue innecesariamente, que no se cuente con todos los datos que se debería por la complejidad de su recogida y almacenamiento y que sigan existiendo limitaciones a la hora de enfrentarse al análisis de datos no estructurados.

Los 5 desafíos que big data plantea a los bancos

- 4. Costes elevados:** los que implican las licencias, su mantenimiento, los servidores... y todo ello con capacidades de almacenamiento y respuesta muy limitadas, las que corresponden a un modelo no optimizado distinto al que se puede experimentar gracias a la aparición de tecnologías como Hadoop, que minimizan el gasto a cambio de un rendimiento óptimo y escalable.

Los 5 desafíos que big data plantea a los bancos

- 5. Falta de apoyo:** a veces es cuestión de motivación, otras de visión, lo que está claro es que la alta gerencia de la empresa debe brindar el soporte necesario a las iniciativas big data, de otra forma, es imposible que prosperen hasta aportar los resultados esperados

Hacia dónde se mueve big data en el sector financiero

1. Analizar el comportamiento de los clientes a través de los distintos canales permite predecir cuáles serán las tendencias.

Hacia dónde se mueve big data en el sector financiero

1. Analizar el comportamiento de los clientes a través de los distintos canales permite predecir cuáles serán las tendencias.
2. Identificar los signos de desgaste o abandono, que pueden terminar desembocando en la pérdida de clientes.

Hacia dónde se mueve big data en el sector financiero

1. Analizar el comportamiento de los clientes a través de los distintos canales permite predecir cuáles serán las tendencias.
2. Identificar los signos de desgaste o abandono, que pueden terminar desembocando en la pérdida de clientes.
3. Conocer realmente a los clientes (y clientes potenciales)



Hacia dónde se mueve big data en el sector financiero

Una reciente encuesta de satisfacción lanzada por Capgemini-Consulting, que revela que:

- Sólo el **37%** de los clientes de entidades financieras tienen la percepción de que se les conoce bien.
- El **45%** no están seguros de que se haya establecido un vínculo entre ellos y su banco.
- El **44%** no está totalmente satisfecho con la experiencia multicanal ofrecida por su entidad bancaria.
- El **13%** no perciben que el banco les transmita la confianza y credibilidad suficientes.

La implementación efectiva de big data en el sector banca

1. Definición de objetivos e identificación de procesos de negocio.
2. Determinación de los datos relevantes y necesarios.
3. Diseño del marco de trabajo analítico.
4. Recogida de datos y aplicación de las técnicas de análisis.
5. Monitorización del esfuerzo, para garantizar alineación con los objetivos.
6. Validación del desempeño analítico desde la perspectiva de negocio, para optimizar el proceso y realizar los ajustes necesarios.



¡GRACIAS!

