



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 I

Modulo 2

Fundamentos II

Modulo 3

Hadoop - Taller Práctico

Modulo 4

Hive - Taller Práctico

Modulo 5

HBase - Taller Práctico

Modulo 6

Spark - Taller Práctico

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, análisis y agregación de valor

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





 **quality.code>**
SOLUTIONS FOR YOUR BUSINESS

¡GRACIAS!

