

**Tipo** : Lectura  
**Capítulo** : Exploración del Big Data

---

## I. OBJETIVO

Ampliar sus conocimientos sobre Big Data

## II. LECTURAS COMPLEMENTARIAS

Se sugiere revisar los siguientes enlaces para profundizar en los conceptos tratados en el presente capítulo.

### ***Tipos de implementación de Big Data (Fuente: PowerData)***

¿Cómo se puede hacer una implementación de Big Data? No existe una única respuesta a esta cuestión porque ello dependerá de:

- La **madurez del business intelligence** de la empresa.
- La infraestructura de que la organización disponga para la implementación.
- Su nivel de conocimiento del nuevo entorno, su expertise (**data science**).
- Su planteamiento, más innovador o más tradicional, que condicionará inequívocamente su elección.

En cualquier caso, sí que puede hablarse de que, generalmente, las empresas, atendiendo a su configuración y dependiendo de los factores precitados se agrupan atendiendo a tres **modelos diferentes de implementación de Big Data**:

- Revolucionario
- Evolutivo
- Híbrido

### **La implementación revolucionaria de Big Data**

Quienes optan por esta **forma de implementar Big Data** han decidido romper con todo y empezar de cero. Es obvio que todas las empresas ya están funcionando, han generado datos y los han ido guardando y procesando a su manera. Hasta su **primer contacto con Big Data** han tomado decisiones en función de datos, pero siempre con un alto componente de incertidumbre y una proporción de hechos que hoy día puede parecer arriesgadamente ridícula.

Esta implementación les pone en el camino de las mejores decisiones porque les confiere la capacidad para hacerlo, no sólo en lo relativo a volumen sino también en cuanto a poder de análisis. Aunque puede sonar a magia, dar este paso de gigante, cualitativamente hablando, requiere nada más y nada menos que de mover todos los datos de que se dispone al nuevo entorno. Y, junto con ellos, los informes, el **modelado y la integración con los procesos de negocio**. El pasado queda atrás y desde el primer momento todo comenzará de cero en esta nueva plataforma.

Hay 3 tres tecnologías que permiten llevarlo a cabo: hadoop, las **bases de datos paralelas** y las **bases de datos in memory**, para **análisis en tiempo real**. Mediante su táctica de "divide y vencerás" permiten trabajar con volúmenes inmensos de información en poco tiempo y a bajo coste y que se complementan con la posibilidad de moverlo todo al **cloud**.

\* Ventajas de este tipo de implementación:

- **Agilidad del procesamiento de la información:** ya que cada porción se procesa de forma mucho más rápida y lo que antes se procesaba en 6 horas, ahora se hace en minutos.
- Optimización de los recursos que implica una notable **reducción de costes** en comparación a otros métodos.
- **Flexibilidad** de uso gracias al sistema de nodos: que permite además poder calcular lo que se va a gastar y pagar sólo por lo que se usa.

**\* Inconvenientes:**

- Altos costos en lo referente a habilidades a lo que se sumaría la dificultad de encontrar candidatos con este perfil en el mercado, profesionales que conozcan y sepan manejar adecuadamente esta tecnología, sacándole todo el partido.
- Quizás también puede suponer un mayor tiempo de adaptación hasta que se pone en marcha, si se compara con otros sistemas, ya que se trata de un método totalmente nuevo.

**Implementación de Big Data: el método evolutivo**

El método evolutivo es otra forma de acceder a este mundo que suele elegirse por empresas, quizás no tan pioneras, pero que ya contaban con un **BI** bastante maduro. Hablamos de organizaciones que disponen de su **warehouse**, su **herramienta de visualización y reporting** y que llevan años analizando datos, a la manera antigua, pero con madurez.

Decantarse por el **método evolutivo de implementación de Big Data** supone que, manteniendo la estructura actual, simplemente se suman los datos que no tenían cabida en el sistema para pre-procesarlos desde la plataforma de **Big Data**. De esta forma, el sistema actual los puede ver y así se logran **análisis**, que siguen siendo del mismo tipo o parecidos, pero que ahora **incorporan más datos**.

La ventaja que tiene esta opción es que su umbral de entrada tiene un costo menor, ya que las empresas todavía pueden seguir empleando su herramienta, aunque de la **extracción de datos** y su estructuración se encargue **Big Data**. En concreto, Big Data se convierte en una entrada a la **plataforma de BI** existente. Los datos se acumulan y analizan, y los resultados se envían al **data warehouse**.

**\* Ventajas de este sistema:**

- Rendimiento. Los modelos ahora se nutren con muchos más datos y más diversificados.
- Volumen. Se **multiplica el volumen de información** que se obtiene, gracias a **Big Data**.
- Ahorro. Tanto en tiempo de **implantación**, como en costes, ya que se mantiene el **BI** existente.

**\* Inconvenientes:**

- La velocidad de extremo a extremo siempre estará limitada por el **entorno actual de BI** (que es mucho más lento).
- El nivel de percepción no es tan amplio, ya que fallaría la granularidad por el **BI** existente.
- Esta solución no es definitiva, porque llegará un momento en que el **BI** no tenga capacidad para hacer frente a los requerimientos informativos de la organización y se tenga que optar por hacer el traslado completo de los datos al exterior de la empresa.

**La alternativa híbrida a la implantación de Big Data**

Este sistema alterna el uso de una y otra tecnología en función del objetivo perseguido. Para determinado tipo de información, de análisis o de usuario se seguiría usando la **BI** existente, mientras que, para **análisis mucho más refinados**, como los de **tipo predictivo, simulaciones**, etc. se usaría **Big Data**. Para lograrlo, bastaría con establecer

un par de puntos de integración, consiguiendo que los datos del **data warehouse** se almacenen en el motor de análisis, lo que sería visto por el éste como un **DataMart**.

En esta opción **Big Data** no sólo es un anexo al **BI** tradicional que permite ahora ver datos que antes no podía, sino que además es una plataforma que sirve para hacer **análisis avanzado** mezclando datos del tradicional que siguen allí, con los nuevos como por ejemplo **Redes Sociales** o los **datos No-Estructurados** que los BI de antes no contemplaban a la hora de trabajar.

**\* Ventajas:**

- Precisión: los sistemas ahora se nutren con una información estructurada, cuando antes sólo podían acceder a datos no estructurados.
- **Big Analytics.** O lo que es lo mismo, posibilidad de hacer un análisis predictivo, creando un modelo estadístico con todos los datos e identificando las relaciones causales y las correlaciones; apoyándonos también en herramientas avanzadas de visualización.
- **El internet de las cosas:** que permite conocer en tiempo real todo lo que está sucediendo en cualquier parte y en relación a cualquier asunto.
- **Rentabilidad.** En definitiva, esta opción permite ofrecer mejores servicios que, claro está, también pueden cobrarse.

**\* Inconvenientes:**

- Resistencia: de las empresas menos evolucionadas, tecnológicamente hablando, a la implantación de un modelo de este tipo.
- Dificultad: de quienes están acostumbrados a limitarse a un tipo de análisis simplemente descriptivo, a la hora de lanzarse a las nuevas posibilidades.