

***BIG DATA* Y SU APLICACIÓN EN EL SECTOR DEL TURISMO**

Antonio López de Avila, Enrique Lancis, Susana García y Calixto Mellén*

Resumen: El artículo describe las nuevas contribuciones que aporta el *Big Data* con respecto a otros sistemas de información predecesores. Los autores profundizan en las características reales del sistema, el funcionamiento, fases que se han de acometer en proyectos, ejemplos concretos y las novedades que incorpora esta tecnología. Asimismo se hace hincapié en las aportaciones de este sistema al sector turístico, su potencial influencia en los destinos turísticos y la descripción de casos reales desarrollados dentro de la industria turística.

Palabras clave: *Big Data*, sistema de información, inteligencia de negocio, turismo, destinos turísticos y destinos turísticos inteligentes.

Abstract: The article describes the new contributions that Big Data brings with respect to other preceding information systems. The authors delve into the real characteristics of the system, operation, phases that have to undertake projects, concrete examples and innovations that incorporate this technology. Also emphasizes the contributions of this system to the tourism sector, its potential influence on tourist destinations and the description of real cases developed within the tourism industry.

Key words: Big Data, information system, business intelligence, tourism, tourist destinations and intelligent tourist destinations.

I. INTRODUCCIÓN

Desde la escritura cuneiforme, una de las formas más antiguas de expresión escrita, hasta las tecnologías actuales, el ser humano no ha dejado de producir y recopilar información.

Vivimos en un mundo globalizado, inmersos en la sociedad de la información, donde la implantación de las TIC está permitiendo replantear las relaciones sociales, económicas, políticas y culturales. El crecimiento del sector tecnológico y su asimilación en la vida diaria de personas y empresas ha provocado un aumento desmesurado del volumen de datos que requiere sistemas de almacenamiento cada vez más sofisticados,

si bien su coste ha disminuido mucho en los últimos años.

Actualmente, la información que se genera día a día, registra un incremento casi exponencial. De acuerdo con la consultora IDC, la información generada en 2011 alcanzó la cifra de 1,8 zettabytes (1), Google procesa al día 20 petabytes por segundo, según Mc Kinsey Global Institute, el 90% de los datos se han generado en los últimos dos años, la mayoría provienen de redes sociales (más de 900 millones de usuarios de Facebook y más de 500 millones de seguidores de Twitter).

La velocidad e intensidad de los cambios de las nuevas tecnologías, han influido en la

* Sociedad Estatal para la Gestión de la Información y las Tecnologías Turísticas (SEGITTUR). Ministerio de Industria, Energía y Turismo.

mayor parte de las actividades económicas, entre ellas, el turismo, repercutiendo en la forma de viajar, de visitar lugares, dinamizar sus búsquedas, gestionar reservas, comprar, opinar y compartir información, etc. **El pasado año 2013, el sector turístico, registró la cifra de 1.087 millones de turistas internacionales que demandan y generan información en sus viajes.** El auge de la tecnología digital y móvil, la comunicación entre los objetos inteligentes a través de sensores y la interacción del visitante con el destino a través de las redes sociales, generará un volumen de datos que será necesario almacenar, analizar y difundir entre todos los actores del destino.

Actualmente, la tecnología abre una gama de posibilidades que facilitan el logro de este objetivo, así han surgido nuevos modelos de gestión de la información, entre ellos, los sistemas de “Big Data”, cuyo valor añadido es fundamentalmente la capacidad de extraer y gestionar el conocimiento. Esta tecnología ofrece una solución eficiente para la gestión, mantenimiento y análisis de la información que se genera en los destinos y de la participación de los distintos actores en el mismo a través de los diferentes canales y redes sociales.

II. LOS PILARES DE *BIG DATA*

II.1. Antecedentes

La necesidad de almacenar y consultar datos se remonta a la Antigüedad, cuando se realizaban todo tipo de registros manuales para recoger información sobre las cosechas y censos.

No es hasta 1884 cuando se logra por primera vez el tratamiento automático de la información gracias a la creación de la máquina automática de tarjetas perforadas de Herman Hollerith. Esta máquina tabuladora se utilizó con gran éxito para elaborar el censo de EE.UU de 1890 y, por ello, se considera a Hollerith el primer ingeniero estadístico de la historia.

Pero no es hasta la década de los sesenta, cuando se generaliza el uso de las bases de datos y la necesidad de almacenar grandes cantidades de información. En esta época surgen las primeras generaciones de bases de datos de red y las bases de datos jerárquicas.

Durante los años 70-80 se define el primer modelo de datos relacional, surgen los SGBDR (2) y se desarrolla el SQL que se convierte en un estándar en la industria. Son sistemas rápidos que alberga información estructurada del orden de terabytes.

A partir de la década de los 90, surgen los sistemas de *Business Intelligence* (BI). Sistemas orientados al análisis de información que integran otro tipo de información (información semi-estructurada, además de la información estructurada). Durante el comienzo del siglo XXI se consolidan estas aplicaciones como un conjunto integral de estrategias y herramientas que permiten analizar los datos de modo automático, extraer patrones de información y facilitar la toma de decisiones empresarial.

Actualmente, la proliferación de datos no estructurados procedentes fundamentalmente de los medios sociales (entornos de

trabajo colaborativos, entornos de *social media*, etc.) y la necesidad de considerar otras fuentes de datos (las cámaras, móviles, sensorización, etc.) ha fomentado la popularidad del uso de Bases de Datos No SQL, optimizadas para ofrecer una respuesta rápida y facilitar el trabajo de forma distribuida.

II.2. El concepto de *Big Data*

Según Francis X. Diebold (2012), el origen del término *Big Data* (un anglicismo de uso popular hoy en día a nivel mundial) se remonta a finales de los años 90. Afirma que el término “actualmente en boga en las ciencias estadísticas, econométricas y en las tecnológicas, probablemente se originó, en la década de 1990, en las conversaciones de Silicon Graphics, en la que John Mashey (experto en ciencias de la computación) ocupó un lugar destacado”.

Aunque no hay documentos académicos que confirmen taxativamente la atribución del término a Mashey o la atribución del uso del término en el sentido actual, según Steve Lohr (2013), se ha confirmado que, en el plano no académico, John Mashey, a mediados/finales de la década de los 90, impartió centenares de coloquios a pequeños grupos para explicar el significado de dicho término. De hecho, algunas de las diapositivas que utilizaba en sus presentaciones se publicaron en 1998, bajo el título de “Big Data and the Next Wave of Infra Stress”. Por último mencionar, que en una conversación de Lohr con Mashey, éste último relata que cuando optó por utilizar ese término, lo que pretendía simplemente era “usar una etiqueta que explicara una serie de cuestiones y que,

a su vez, seleccionando la frase más corta y más sencilla, transmitiera que los límites de la tecnología seguían avanzando”.

Finalmente, más adelante, Laney (2001), propone las tres V de *Big Data* (volumen, variedad y velocidad) en una nota de investigación inédita en META Group (2001).

En la actualidad, cuando se habla de *Big Data*, (traducida por la Fundeu como “datos masivos”, “datos a gran escala” o “macrodatos”) nos referimos a grandes volúmenes de datos. No obstante, este factor no es el único condicionante a la hora de caracterizar el sistema ya que muchas compañías llevan tiempo desarrollando complejos almacenes de datos (data warehouses) y utilizando potentes herramientas analíticas (data mining) que les permiten estudiar estos volúmenes de datos.

No obstante, **uno de los principales valores añadidos del sistema de *Big Data*, es la incorporación de los datos no estructurados al sistema** (procedentes principalmente de los entornos de social media), de modo que actualmente, debido al avance de las tecnologías, es posible capturarlos, almacenarlos, procesarlos y analizarlos.

Tradicionalmente, las tecnologías no daban cabida al análisis de datos no estructurados. Los sistemas de tratamiento de información se organizaban en torno a modelos de datos relacionales o multidimensionales, facilitando el procesamiento de datos de forma periódica y automática, siempre que los datos fueran de carácter estructurado o semi-estructurado. Por otro lado, la mayoría de las herramientas tampoco ofrecían una

solución adecuada para realizar análisis en tiempo real, que es otro de los valores que aportan los sistemas de *Big Data* y que facilita la celeridad en la toma de decisiones.

Actualmente, la popularización de este término ha llevado a intentar acotarlo en base a la propuesta de distintas definiciones, entre ellas se citan las siguientes:

“El análisis a alta velocidad, de un gran volumen y de una gran variedad de datos mediante formas innovadoras y rentables de procesamiento, con el objetivo de mejorar la comprensión y la toma de decisiones”. (Gartner, 2012)

“Conjuntos de datos cuyo tamaño va más allá de la capacidad de captura, almacenamiento, gestión y análisis de las herramientas de base de datos tradicionales”. (McKinsey Global Institute, 2011)

No obstante, para realizar una aproximación más precisa a la definición del sistema cabe proceder a su descripción en base a sus principales características, inicialmente, categorizadas bajo la denominación de las tres “V”, a saber: volumen, velocidad y variedad:

Volumen

El sistema de *Big Data* almacena un gran volumen de datos mediante una infraestructura distribuida y escalable. La contribución al aumento del volumen de datos cada vez es mayor. Los datos de transacciones almacenados a lo largo de los años, los datos de texto que constantemente generan las redes sociales, la creciente cantidad

de datos recopilados de sensores, etc. En el pasado, un volumen de datos excesivo planteaba un problema de almacenamiento. Hoy en día la preocupación se centra en la forma de determinar la relevancia de los ingentes volúmenes de datos y la creación de valor a partir de la información pertinente.

Velocidad

Una de las características más relevantes del sistema es el tiempo de respuesta respecto al procesamiento y análisis de los datos. La velocidad con que se generan nuevos datos en menor tiempo, incrementa la importancia de disponer de herramientas capaces de realizar el procesamiento de los datos en tiempo reducido o tiempo real para poder satisfacer la demanda. **La velocidad refleja la frecuencia con la que los datos se almacenan, procesan y comparten.** Las empresas y la sociedad pueden beneficiarse en mayor medida de los datos si pueden ser capturados y analizados que tiempo real. Este es el punto en el que muchas herramientas de análisis, CRM, puntos de venta o sistemas similares se quedan cortos.

Variedad

El sistema es capaz de analizar y procesar una gran diversidad de tipos y formatos de datos proporcionados, en parte, por la consideración de las nuevas fuentes de datos. Los datos ya no se pueden encajar en estructuras sencillas y fáciles de usar. Los nuevos tipos incluyen contenido no estructurado, datos geo-espaciales, información del hardware, localización, datos de velocidad de acceso, métricas, móviles, RFIDs, resultados de búsqueda, análisis de sentimiento,

streaming, redes sociales, texto, web, transacciones comerciales o financieras, datos de mediciones, sensores, etc. Además según algunos cálculos, el 80% de los datos ya son alfanuméricos. Estos también deben incluirse en los análisis y en el proceso de toma de decisiones.

Así, de acuerdo a estas características, se podría afirmar que el sistema de *Big Data* representa el tratamiento de grandes **Volúmenes** de información que se procesan y analizan a alta **Velocidad** y que pueden presentar una compleja **Variabilidad** en cuanto a su composición.

Posteriormente, algunos autores han considerado los “V” más en la caracterización del sistema, que son igualmente relevantes para definir la complejidad del mismo:

Veracidad

Una característica que hace referencia a la confiabilidad de los datos, la **calidad de la información y su precisión**. El sistema es capaz de tratar y analizar inteligentemente el gran volumen de datos, obteniendo una información veraz y útil que permite mejorar la toma de decisiones.

Valor

Hace referencia a la generación de valor a través de los datos almacenados. El proceso de conversión de los datos en información que facilite la toma de decisiones y la posibilidad de que la gestión del conocimiento se convierta en una realidad. **Se hace hincapié en la importancia de saber diferenciar y generar datos, información y conocimiento.**

Estas cinco “V” aplicadas a un destino turístico podrían ser interpretadas de la siguiente manera: “Dispongo de una gran información de lo que consultan mis turistas en las web de mi destino, lo que compran, consumen, visitan y opinan de mí en las redes sociales. Así mismo dispongo de información de los servicios disponibles y los recursos del destino, el gasto, perfiles de población y origen, etc. Todo ello me genera un gran Volumen proveniente de fuentes y formatos muy Variados, además puedo asegurar que la información de origen es Veraz, porque la he tratado y filtrado. Por otro lado, una vez puesto en marcha exigiremos obtener una información de Valor para la toma de decisiones con una Velocidad de proceso aceptable”.

II.3. Soluciones tecnológicas

La aparición del paradigma de *Big Data*, ha fomentado el desarrollo de distintos modelos de programación que ofrecen distintas soluciones para la extracción, almacenamiento y análisis inteligente de los datos.

Actualmente uno de los principales modelos de programación para el *Big Data* es: **Mapreduce y Hadoop.**

Es un paradigma de programación que actualmente cuenta con multitud de implementaciones.

La popularización de este modelo comenzó con las primeras implementaciones de Google. Su necesidad de realizar complejas operaciones de para calcular el PageRank de las páginas Web y manipular

grandes volúmenes de datos, fomentó el desarrollo de frameworks (3) capaces de poder procesar terabytes de datos.

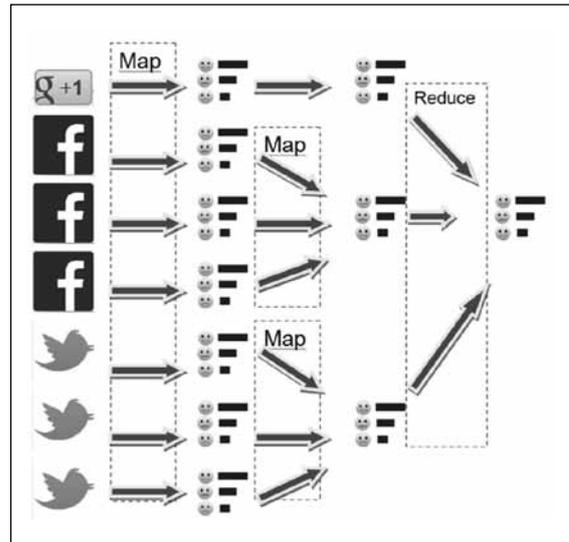
Mapreduce facilita el desarrollo de programas que procesan en paralelo (4) grandes cantidades de datos a través de un grupo de procesadores distribuidos (5). En otras palabras, permite una escalabilidad masiva a través de cientos de miles de servidores.

El funcionamiento de Mapreduce se basa en la división de un problema en otros menores. Los archivos se dividen en bloques y éstos se procesan en múltiples nodos (6) de forma paralela. Se parte de un problema inicial, el nodo maestro lo divide en sub-problemas y lo distribuye entre los distintos nodos. A modo de ejemplo, se parte de un caso de análisis de información que procede de distintos medios sociales.

Tal y como se observa en la figura 1 existen varios nodos analizando los distintos medios sociales. Así, en una primera fase (“fase map”), existe un nodo analizando google plus, tres nodos analizando Facebook y otros tres analizando Twitter. A continuación se unen los datos y en la fase de “reduce” se presenta el resultado final.

Mapreduce y HDFS (7) constituyen el corazón de Hadoop, actualmente una de las implementaciones con mayor éxito. Hadoop es un proyecto de código abierto administrado por **Apache Software Foundation**, que soporta aplicaciones distribuidas bajo licencia libre, facilita el procesamiento de grandes volúmenes de información (petabytes) y permite trabajar con miles de nodos.

Figura 1
Funcionamiento de MapReduce



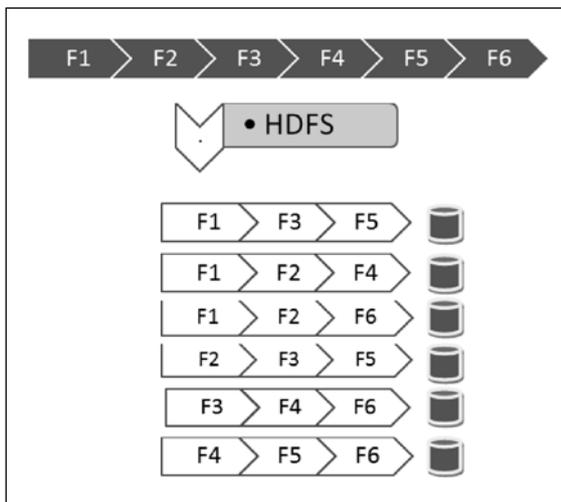
El desarrollo de Hadoop se asienta sobre la base de los documentos de Google para Mapreduce y de Google File System.

Consta de un *framework* de computación y persistencia distribuida que tiene su origen en dos conceptos: almacenamiento de datos en Hadoop Distributed File System (HDFS) y procesamiento a través de trabajos MapReduce.

HDFS proporciona un sistema de archivos distribuidos que facilita el crecimiento del sistema mediante la simple adhesión de nuevas máquinas al clúster (8). De este modo, es posible que multitud de máquinas de escasa capacidad ofrezcan excelentes resultados, eliminando la necesidad de adquirir costosos equipos.

Además, el sistema de ficheros distribuidos de Google es capaz de ejecutar trabajos

Figura 2
Proceso de distribución de ficheros
garantizando la redundancia



a gran escala y alto rendimiento, detectando y compensando los problemas de hardware en cualquier servidor. La figura 2 presenta el proceso de distribución de ficheros en distintos equipos, garantizando la redundancia de información para seis ficheros de datos.

Más allá de la información transaccional y operacional, de las cargas de trabajo OLTP (9) y OLAP (10), ideales para datos estructurados y semiestructurados, **uno de los valores añadidos de Hadoop es que posibilita la ejecución de análisis en tiempo real y otorga cabida a los datos no estructurados.**

II.4. Fases del proceso de *Big Data*

El sistema de *Big Data* sigue un esquema de funcionamiento acorde al considerado en

las soluciones de Inteligencia de Negocio (*Business Intelligence*), no obstante cada una de las fases que componen el esquema de flujo de la información ha ido evolucionando y mejorando en concordancia con el avance de las tecnologías. De este modo el funcionamiento del sistema puede describirse de acuerdo a las siguientes fases:

II.4.1. Captura de datos

En la etapa inicial de la captura de datos se pueden distinguir fundamentalmente dos tipos de herramientas en función de la conexión que se establezca con los distintos orígenes de datos, de este modo es posible diferenciar entre:

- Conexión en modo *batch* (lotes de datos): Se establece una conexión periódica al origen de datos, habitualmente a sistemas de archivos o bases de datos, facilitando la apertura, importación, edición y actualización de los datos de forma automática.
- Conexión por *streaming* (en tiempo real): Conexión permanente al origen de datos posibilitando la extracción de los mismos cada vez que estos se generan.

II.4.2. Almacenamiento de la información

La asociación del concepto *Big Data* a grandes volúmenes de datos no es nueva. La gran mayoría de las empresas llevan mucho tiempo manejando grandes volúmenes de datos y desarrollando *Data Warehouses* y potentes herramientas analíticas (*Data-*

Mining) que les permiten analizar de forma adecuada estos grandes volúmenes.

Realmente, una de las principales novedades es la consideración de los datos no estructurados (procedentes principalmente de las redes sociales), que en la actualidad, gracias a la evolución de la tecnología, es posible almacenar y analizar.

Tradicionalmente los sistemas de tratamiento de información se estructuraban en bases de datos relacionales o multidimensionales, facilitando el procesamiento y análisis de datos (datos estructurados o semi-estructurados fundamentalmente) de forma periódica y automática. Asimismo se utilizaban las tecnologías de *DataMining* para explorar y analizar de forma automática grandes bases de datos, con el objetivo de identificar patrones o tendencias que explicaran el comportamiento de los datos en un determinado contexto y realizar pronósticos y predicciones.

Sin embargo, por un lado, las tecnologías no daban cabida al procesamiento y análisis de datos no estructurados procedentes de entornos de social media y, por otro, la mayoría de las herramientas no ofrecía una solución adecuada para realizar análisis en tiempo real.

Actualmente las tecnologías de almacenamiento y procesamiento de la información proporcionan una solución adecuada para manejar la complejidad de los diversos formatos, ofreciendo además capacidad de respuesta en “tiempo real”.

Así, han surgido nuevos modelos de almacenamiento de la información que se

basan fundamentalmente en sistemas de ficheros distribuidos (nodos configurados para conformar un único sistema de ficheros lógicos) y bases de datos No-SQL o Not-Only-SQL1.

Los sistemas de bases de datos NoSQL recibieron un fuerte impulso por parte de las principales compañías que basan su actividad en Internet y en entornos de redes sociales como Google, Amazon, Twitter, Facebook, ya que facilitaban la resolución de problemas de almacenamiento y procesamiento de grandes volúmenes de datos en tiempo real. La ventaja de estos sistemas de bases de datos es que obtienen mayor rendimiento para ciertos modelos de datos y escalan bien horizontalmente. La pérdida de coherencia y flexibilidad en tiempo de ejecución, con respecto a los sistemas clásicos, se ven compensados por sus propiedades óptimas para el procesamiento en tiempo real.

II.4.3. Análisis de los datos

Las herramientas de procesamiento de datos han evolucionado para integrar junto a los procesos tradicionales de depuración, normalización y validación, otros procesos que facilitan el tratamiento de datos no estructurados.

Por otro lado, los sistemas de análisis de datos han ido adaptándose de acuerdo a las necesidades de la demanda. Inicialmente el análisis estadístico dependía del conocimiento y manejo de programas estadísticos que permitían realizar analítica predictiva, prever los distintos sucesos, tomar decisiones más inteligentes, resolver problemas y

mejorar los resultados. El mayor inconveniente de estos sistemas era la necesidad de ejecutar los análisis ad-hoc, habitualmente llevados a cabo mediante equipos de expertos en estadística, que debían proceder a realizar nuevamente todo el estudio (depuración, validación y análisis de datos), cada vez que recibían nuevas actualizaciones de los datos o cambiaba algún código alfanumérico de identificación. Para paliar este inconveniente, surgieron las herramientas de minería de datos (*Data Mining*). Estos procedimientos, integrados ya en los sistemas de *Business Intelligence*, utilizan los métodos de la inteligencia artificial y la estadística para descubrir y analizar los patrones a partir de las bases de datos con las que trabaja. Su valor añadido es la posibilidad de realizar análisis automáticos o semi-automáticos a través de grandes volúmenes de datos para identificar relaciones o tendencias significativas.

II.4.4. Presentación de la información

Como hemos indicado, el avance de las tecnologías ha fomentado el desarrollo de nuevas herramientas que permiten presentar el análisis de la información de modo interactivo y en tiempo real y esto ha supuesto un gran avance. **Los tradicionales cuadros de mando, en muchas ocasiones vinculados a bases de datos relacionales, ceden el paso al uso de cuadros de mando en tiempo real cuyos gráficos dinámicos, se actualizan automáticamente cada vez que se actualizan los datos.**

Actualmente, la capa de presentación cuenta con la capacidad de ofrecer una respuesta en tiempo real, posibilitando un

análisis dinámico e intuitivo que facilita la comprensión de la información y contribuye a generar una valiosa herramienta para la toma de decisiones. A modo de resumen, se citan algunas características que se consideran deseables para realizar una buena presentación de datos, a saber:

- Información dinámica (textual, tabular, gráfica, geoespacial, cronológica etc.) vinculada a la actualización de los datos en tiempo real.
- Posibilidad de interactuar con los contenidos.
- Posibilidad de personalizar las herramientas de visualización haciendo hincapié en las preferencias personales de visualización, además de considerar los distintos niveles de autorización o acceso de los usuarios.
- Posibilidad de navegación dinámica entre jerarquías de las dimensiones agregando y desagregando datos (*drill down y drill up*)
- Priorizar la usabilidad y la simplicidad en el diseño en aras a facilitar la comprensión de la información y acelerar la toma de decisiones.

III. APORTACIONES DEL SISTEMA AL SECTOR TURÍSTICO

Las principales aportaciones del *Big Data* al sector turístico identificadas en la actualidad son, entre otras:

- Previsiones del comportamiento de los visitantes en base a la identificación de patrones y tendencias.
- Mejora de los programas de fidelización, fomentando la ejecución de análisis multi-variantes a través del estudio de distintas variables, como por ejemplo: determinar las características de los clientes que asisten a distintos eventos, generando un determinado gasto y un retorno de la inversión en el destino.
- Incremento del conocimiento de nuestros clientes mediante el uso de técnicas de segmentación basadas en sus características socio-demográficas y sus preferencias. De este modo, será posible diseñar programas de recompensa en tiempo real, que premien la fidelidad de los turistas en relación al consumo de distintos bienes o servicios.
- Identificación de mercados potenciales que puedan estar interesados en viajar a nuestros destinos, análisis de estados de reservas de vuelos y previsión de llegadas de distintos puntos del planeta y desarrollo de modelos predictivos.
- Caracterizar los bienes y servicios más demandados, utilizando la información para desarrollar productos diferenciales y a medida.
- Mejorar los sistemas de *Revenue Management* (gestión de ingresos) actuales que fueron desarrollados en base a una tecnología que actualmente presenta limitaciones respecto al formato de datos, la capacidad de proceso, variables de análisis y coste.
- Disminución del coste de computación ganando en velocidad de proceso.
- Mejoras en distintas áreas de actividad económica relacionadas con el sector turístico como:
 - Gestión del flujo e impacto del turista sobre el territorio y posibilidad de abrir zonas alternativas de atracción, poniendo en valor otros recursos turísticos, para descongestionar zonas concretas de gran afluencia
 - Sistemas de control del tráfico
 - Mejora de la seguridad mediante la predicción de posibles anomalías
 - Incremento del gasto en comercio en base al conocimiento de las preferencias de la demanda

IV. EL SISTEMA DE **BIG DATA** A DISPOSICIÓN DE UN DESTINO TURÍSTICO

Dentro del marco del proyecto de Destinos Turísticos Inteligentes, y gracias a la financiación del Programa de Cooperación Transfronteriza España-Portugal (POCTEP), se va a poner en marcha el primer exponente de un **Sistema de Inteligencia Turística Territorial**, dentro del área transfronteriza Badajoz-Elvas. **El objetivo es recoger, procesar y analizar en tiempo real grandes flujos de información con el fin de mejorar la competitividad y la eficiencia del destino a través de sus entes de gestión y empresas privadas radicadas en la zona, ofreciendo al turista una mejor**

experiencia en el destino y al residente una mayor calidad de vida, asegurando así la sostenibilidad y rentabilidad de la actividad turística.

En el caso que presentamos, el objetivo principal es fomentar la eficiencia en la gestión pública y en la oferta privada de los espacios urbanos transfronterizos de Badajoz-Elvas, mediante el estudio de los principales factores de influencia, entre ellos:

- Monitorización de los consumos turísticos de los visitantes
- Análisis de datos en tiempo real
- Evolución del tráfico urbano
- Evolución del tránsito peatonal
- Evolución del acceso de vehículos a parkings
- Evolución de la ocupación hotelera y la restauración
- Acceso a centros sanitarios privados
- Caracterización de visitantes de oficinas de información turística

La ejecución del sistema de información en Badajoz y Elvas contribuirá a la conversión de la zona en un destino inteligente, pues la incorporación de esta tecnología fomentará la capacidad innovadora y productiva de la empresa privada, facilitará la recuperación de la competitividad del destino, permitirá evaluar el retorno de las acciones de la actividad en el destino (turismo, hostelería, comercio, sanidad, ocio, etc.), im-

pulsará la difusión de las mejores prácticas en el sector, adaptando la oferta a las nuevas exigencias y generará nuevas oportunidades de empleo.

V. CASOS DE *BIG DATA* EN LA INDUSTRIA TURÍSTICA

Actualmente comienzan a desarrollarse casos de implementaciones de los sistemas de *Big Data*, relacionadas con las empresas y las actividades de la industria turística, entre ellos, cabe destacar los siguientes:

CASO 1. Características, comportamiento y gasto de turistas en Madrid y Barcelona

El estudio “*Big Data* y turismo: nuevos indicadores para la gestión turística” presenta la potencialidad que ofrece el análisis de grandes volúmenes de datos de cara a facilitar la toma de decisiones dentro de un sector especialmente relevante para la economía española.

El estudio analiza el comportamiento de los visitantes extranjeros a las ciudades de Madrid y Barcelona utilizando datos procedentes de terminales que se conectaron a la infraestructura de Telefónica entre el 7 y el 21 de octubre de 2012 (facilitados por Telefónica Móviles España) y de pagos electrónicos realizados mediante tarjetas de crédito transmitidas por la red de terminales del banco durante el mismo periodo y en las mismas ciudades (facilitados por BBVA). El número de terminales móviles y tarjetas objeto de estudio han sido, respectivamente, de 680.928 y 168.921.

Los indicadores que se seleccionaron se basaron en dos tipologías diferenciadas. Por un lado, se estudiaron aquellos indicadores considerados tradicionalmente en los estudios de investigación dentro del sector y, por el otro, se seleccionaron otros indicadores cuya cuantificación precisaba del uso de un sistema de *Big Data*.

A partir del análisis realizado fue posible extraer una serie de recomendaciones dirigidas fundamentalmente a gestores hoteleros. Entre ellas, se pueden citar:

- Incremento de la captación de clientes e identificación de los países en los que es preciso focalizar la acción comercial.
- Delimitación de las áreas de la ciudad donde se realizan la mayor parte de las transacciones comerciales.
- Definición y garantía de un producto atractivo adaptado a las necesidades del cliente.

CASO 2. Kayak

KAYAK es un sitio web de búsqueda de viajes que gestiona más del mil millones de operaciones al año. Su actividad depende en gran medida de la posibilidad de generar análisis de datos a gran velocidad para facilitar la toma de decisiones y proporcionar las mejores ofertas al cliente.

KAYAK utiliza modelos analíticos para garantizar que los precios que aparecen en su página web sean consistentes con los de

otros sitios web, evitando los posibles problemas de sincronización a través de las diversas fuentes de datos.

Además, respecto a la gestión de los vuelos, KAYAK ha introducido recientemente un algoritmo de pronósticos de precios de vuelos que predice en base a un determinado nivel de confianza, si el precio de un vuelo variará en los próximos siete días.

Debido a la cantidad de datos que KAYAK procesa y analiza, la compañía ha apostado por el uso de una variedad de herramientas de código abierto de análisis estadístico, haciendo uso intensivo de Hadoop, por su rapidez y bajo costo en relación a tecnologías alternativas.

CASO 3. British Airways

British Airways ha apostado por incrementar su competitividad en el mercado fomentando el uso de técnicas que facilitan la obtención de datos relativos a las preferencias clave de sus clientes. Recientemente, la compañía decidió incorporar información de sus clientes a través de su programa denominado “Know Me” con el objetivo de analizar las características de sus clientes mediante una solución de *Big Data*, para lograr una ventaja competitiva frente a cualquier otra compañía aérea.

Los tres pilares básicos del programa “Know Me” son los siguientes:

- Premiar la fidelidad de los clientes ofreciéndoles beneficios específicos.
- Mejorar la experiencia del vuelo y paliar los posibles inconvenientes.

- Incrementar la excelencia en el servicio a través del análisis de los problemas y las preferencias de sus clientes.
- Desarrollar ofertas que inspiren y motiven a sus clientes, desechando aquellas que sean irrelevantes, centrándose en las propuestas que aportan verdadero valor al viajero e impulsan la imagen de calidad de la compañía.

La información generada mediante el programa "Know Me" se traduce en la propuesta de ofertas en tiempo real que se ponen a disposición del clientes durante el check-in, en salas de espera, mediante los iPADS del personal de la compañía, etc.).

CASO 4. Marriott, pionera en analítica de datos para *Revenue Management*.

Marriott fue una de las primeras cadenas hoteleras en adoptar la analítica como herramienta para gestionar sus ingresos. La gestión de ingresos es el proceso por el cual los hoteles establecen el precio óptimo para sus habitaciones en función de las plazas disponibles. Marriott establece algoritmos para predecir el precio óptimo de la habitación, persuadiendo de este modo a los propietarios de hoteles independientes de que van a obtener más ingresos con la marca Marriott que con otras enseñas de sus competidores.

Con el fin de mejorar sus capacidades de gestión de los ingresos, Marriott mejoró sus algoritmos de gestión, incrementó la velocidad de procesamiento de datos y extendió el *Revenue Management* a los restaurantes, catering y salas de reuniones.

Actualmente ha centrado su foco de atención en la caracterización de las distintas ofertas y actividades de marketing para identificar aquellas que realmente impulsan las ventas.

CASO 5. Nsight, una consultora especializada en minería de datos.

Nsight es una de las empresas que está logrando un mayor valor diferencial en el sector de los viajes mediante el uso del sistema de *Big Data*.

Nsight se ha especializado en la minería y análisis de datos para la industria, agregando diariamente datos de más de 80 millones de consumidores. Esta información se obtiene a partir de una red de 5.000 sitios web de viajes y se estima que se analizará en un año alrededor de 29.000 millones de registros. Con este volumen de datos será posible extraer interesantes y precisas conclusiones.

Nsight no se limita a recoger y evaluar los datos, sino que también establece una serie de categorías de análisis. Una de las partes más importantes de este proceso ha sido la creación de un índice de eficiencia que compara el impacto de las distintas tarifas y permite mejorar el *Revenue Management*.

Además ofrece una serie de pronósticos de reservas en hoteles, analizando las reservas de sus clientes en un horizontes de 30, 60 y 90 días. Mediante el análisis de datos que ofrece esta empresa, predecir el futuro y el perfil de clientes puede dejar de ser ciencia ficción.

CASO 6. BBVA y el Ayuntamiento de Madrid

El estudio realizado por BVA y el ayuntamiento de Madrid “Dinámicas del turismo en la ciudad de Madrid” se basa en la actividad comercial real del año 2012. Muestra las posibilidades que nacen de la unión de los conceptos *Smart City* y *Big Data*. El estudio analiza el comportamiento de los visitantes de Madrid, en base a la traza de los pagos realizados con sus tarjetas de crédito y débito en los comercios de la ciudad. Los resultados presentan en detalle lo que sucede en las ocho zonas de Madrid que más visitantes reciben.

- Anticiparse al flujo de reservas y analizar el comportamiento del mercado en función de la variación de precios, logrando de este modo incrementar su beneficio.
- Controlar los gastos, conociendo las necesidades del servicio y optimizando las horas extraordinarias de los trabajadores según las previsiones de ocupación.
- Desarrollar una guía para la dirección estratégica de los establecimientos, basada en el análisis de datos procedente del estudio del comportamiento de los clientes y de las respuestas obtenidas a través de encuestas.

CASO 7. Air France-KLM

Air France-KLM ha desarrollado un nuevo sistema de *Revenue Management*, común a todas sus aerolíneas, utilizando Hadoop como plataforma de *Big Data*. Este sistema incorpora nueva información referente a sus pasajeros, optimiza sus itinerarios para incrementar el beneficio y estima las posibilidades de cancelación y no-shows para conocer el *overbooking* (sobre venta) lícito en cada vuelo.

CASO 9. Hipmunk

Esta agencia de viajes on-line ha asentado su éxito sobre la base de la valoración asignada a vuelos y alojamientos en una serie de criterios, clasificándolos y categorizándolos de acuerdo a las opiniones de los usuarios. El sistema de *Big Data* ha facilitado la posibilidad de mejorar su sistema de venta, ajustar las ofertas y desarrollar sus propios algoritmos para evaluar los servicios en función de diversas variables, entre ellas: precio, duración de un vuelo, número de escalas, *amenities* del hotel, etc.

CASO 8. IBM y Denihan Hospitality Group

IBM y Denihan Hospitality Group, propietario y operador de hoteles boutique muy reconocido en los EE.UU, están colaborando en un proyecto de *Big Data* aplicando soluciones de analítica avanzada. A través de este sistema han conseguido:

BIBLIOGRAFÍA

- BBVA (2012) *Dinámicas del turismo en la Dinámicas del turismo en la ciudad de Madrid, un estudio ciudad de Madrid, un estudio basado en la actividad comercial*. Madrid, BBVA.
- CHEN, M. et al., (2014) *Big Data: Related Technologies, Challenges and Future Prospects*. Springer.

CHEN, H.; CHIANG, R.H.L. y V.C. STOREY., (2012) "Business Intelligence and Analytics: From Big Data to Big Impact" en *MIS Quarterly*. Vol. 36, No. 4, pp. 1165-1188.

DAVENPORT, T. (2013): *At the Big Data Crossroads: turning towards a smarter travel experience*. Amadeus IT Group.

DIEBOLD, F.X., (2012) *A Personal Perspective on the Origin(s) and Development of "Big Data": The Phenomenon, the Term, and the Discipline*. University of Pennsylvania.

LITH, A. y J MATTSSON, (2010) *Investigating storage solutions for large data*. Master of Science Thesis. Sweden, Department of Computer Science and Engineering, Chalmers University of Technology, Göteborg.

FUJITSU, (2012) *The White book of Big Data*. Fujitsu Services

LOHR, S., (2013) *The Origins of "Big Data": An Etymological Detective Story*. The New York Times.

MANYIKA, J. ET AL., (2011) *Big Data: The next frontier for innovation, competition, and productivity*. McKinsey Global Institute.

SUST, V.O. et al., (2014): *Big Data y turismo: nuevos indicadores para la gestión turística*. Barcelona, Telefónica I+D y RocaSalvatella.

ZIKOPOULOS, P. y C. EATON, (2011) *Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data*. McGraw-Hill Osborne Media

WEBGRAFÍA

<http://hadoop.apache.org/docs/r2.5.1/hadoop-project-dist/hadoop-hdfs/HdfsUserGuide.html>

<http://hadoop.apache.org/docs/r2.5.1/hadoop-mapreduce-client/hadoop-mapreduce-client-core/MapReduceTutorial.html>

<http://thinkdigital.travel/knowledgestream/>

<https://www.sita.aero/content/empower-your-crew>

<https://www-03.ibm.com/press/us/en/pressrelease/42343.wss>

<https://www.hipmunk.com/>

NOTAS

(1) kilobyte (kB): 10^3 ; megabyte (MB): 10^6 ; gigabyte (GB): 10^9 ; terabyte (TB): 10^{12} ; petabyte (PB): 10^{15} ; exabyte (EB): 10^{18} ; zettabyte (ZB): 10^{21} ; yottabyte (YB): 10^{24}

(2) SGBDR: Sistema de gestión de bases de datos. Un conjunto de programas que facilitan la de definición (especificación de tipos de datos, estructuras y restricciones), construcción (proceso de almacenamiento) y manipulación (funciones de consulta, extracción de datos, modificación, borrado, etc.) de una base de datos.

(3) Framework: Esquema de trabajo que define un conjunto de estandarizado de conceptos, prácticas y criterios que facilitan el desarrollo y/o la implementación de un software o aplicación.

(4) Procesamiento en paralelo: Modo de procesamiento basado en el principio de la división de los grandes problemas en otros más pequeños, de este modo, se procede a la ejecución simultánea de múltiples instrucciones que posteriormente se resuelven simultáneamente (en paralelo)

(5) Procesamiento distribuido: Una colección de máquinas físicamente independientes pero conectadas entre sí por una red de comunicaciones, logrando que una sola tarea de procesamiento de datos pueda ser procesada o ejecutada entre varias máquinas de la red.

(6) Nodo: un punto de intersección en donde varios elementos de una red que comparten las mismas características se vinculan e interactúan.

(7) HDFS: Hadoop Distributed file System. Sistema de archivos sobre el que se ejecutan la mayor parte de las herramientas del ecosistema Hadoop

(8) Clúster: Conjunto de ordenadores contruidos en base a hardware común que se comportan como si fueran una única máquina.

(9) OLTP: Procesamiento de transacciones en línea. Un tipo de procesamiento que facilita operaciones transaccionales. Una transacción genera un pro-

ceso que puede involucrar operaciones de inserción, modificación y borrado de datos. El proceso transaccional es típico de las bases de datos operacionales.

(10) OLAP: Procesamiento analítico en línea. Son sistemas orientados al proceso analítico. Agiliza la consulta de grandes volúmenes de datos utilizando estructuras multidimensionales. Muy utilizado para analizar tendencias de ventas, patrones de comportamiento de consumidores, informes de dirección, etc.