

BIG DATA Y CUSTOMER EXPERIENCE

Big Data y Customer Experience (CX) son dos términos que hoy en día están muy de moda. Muchísimas empresas a lo largo del mundo empiezan a mirar cada vez más hacia estos nuevos métodos y tecnología.

Pero que significan estos dos términos. El primero, Big Data, se refiere a los conjuntos de datos cuyo tamaño está más allá de las capacidades de las herramientas típicas de software de bases de datos para capturar, almacenar, gestionar y analizar. El segundo, Customer Experience (CX), es el producto de las percepciones de un cliente después de interactuar racional, física, emocional y/o psicológicamente con cualquier parte de una empresa. Ésta percepción afecta los comportamientos del cliente y genera recuerdos que impulsan la lealtad y afectan el valor económico que genera una organización y que alcanza a todas las interacciones desde la atracción y el interés que generó el marketing, el descubrimiento, la compra, el uso, el servicio al cliente y hasta la baja, sin importar el canal por el que ocurran. El objetivo final de CX es optimizar las interacciones, siempre desde el punto de vista del cliente, para conseguir su completa fidelización.

Para poder mejorar la experiencia del cliente, primero debemos medirla y aquí es donde entra a jugar el Big Data.

Lo cierto es que con la definición de Big Data sólo nos hacemos una pequeña idea de que estamos hablando, así que vamos ir poco a poco tratando de entender todo lo que abarca.

Actualmente los datos se pueden dividir en tres categorías:

1. Datos Estructurados: datos con formato o esquema fijo que poseen campos fijos. Los podemos obtener de las bases de datos relacionales, hojas de cálculo y archivos principalmente.

2. Datos Semiestructurados: datos que no tienen formatos fijos, pero contienen etiquetas y otros marcadores que permiten separar los datos. Los podemos obtener, por ejemplo, del texto de etiquetas XML y HTML.
3. Datos no Estructurados: datos sin tipos predefinidos. Se almacenan como documentos u objetos sin estructura uniforme. Como ejemplo de datos sin campos fijos tenemos: audio, video, imágenes, mensajes SMS o de whatsapp, emails, libros, artículos, etc...

Las fuentes de las que obtenemos estos datos son numerosas, a continuación mostraré una clasificación muy habitual:

1. Web y medios sociales: contenido web, flujos de clicks, contenido de redes sociales tipo Twitter, Facebook, Google my business, etc, datos de imágenes en redes tipo Instagram, pinterest, etc...
2. Maquina a Maquina: lecturas de sensores, lecturas de datos procedentes de chips RFID o NFC, de códigos QR, señales GPS, etc...
3. Transacción de grandes datos: datos procedentes de operaciones transaccionales de todo tipo como registros de facturación, registros detallados de las llamadas (CDR), etc...
4. Biometría: datos procedentes de reconocimiento facial, ocular, de huellas, genética, etc...
5. Generados por los humanos: Correo electrónico, datos de voz de llamadas, documentos electrónicos, en papel, etc...

Teniendo en cuenta todos estos datos que somos capaces de obtener y para hacernos una idea de la cantidad de datos existentes, decir que una imagen en hd ocupa menos de 300Megabytes(Mb). Actualmente la información se mide en zettabytes(Zb). Un Zb equivale a 1000 exabytes(Eb), un Eb a 1000 petabytes(Pb), un Pb a 1000

terabytes(Tb), un Tb 1000Gigabyte(Gb), y un Gb equivale a 1000Mb. Así que nos podemos imaginar cual es la dimensión de los datos existentes.

Es evidente que para poder trabajar con todos estos datos ha sido necesario el desarrollo de la tecnología que permita tratarlos y gestionarlos. Gracias a la evolución de los ordenadores, a su capacidad de procesamiento y almacenamiento, se ha podido desarrollar sistemas como Hadoop que subyacen bajo todos estos grandes datos y evolucionar las bases de datos desarrollando dos nuevos tipos: nosql y en memoria.

Hadoop es un marco de trabajo/framework) que permite procesar grandes cantidades de datos a muy bajo coste, y que incluye una colección de componentes de procesamiento de datos distribuidos para almacenamiento y proceso de datos estructurados, semiestructurados o no estructurados con una alta escalabilidad. Se compone de:

1. Hadoop Distributed File System(HDFS): un sistema de archivos distribuidos.
2. Hadoop YARN: un marco de trabajo para la planificación de trabajos y gestión de recursos en cluster.
3. Hadoop MapReduce: sistema de procesamiento paralelo de grandes conjuntos de datos. Lo que hace este sistema es dividir las tareas complejas en cientos/miles de otras subtarefas más sencillas que envía a ejecutar en diferentes nodos del cluster, y luego combina los resultados en un conjunto de datos más pequeño y fácil de analizar.
4. Hadoop Common: utilidades típicas que soportan los módulos anteriores.

Hadoop permite acceder tanto a datos de archivos del sistema como a datos en bases de datos.

Las bases de datos en memoria usan la memoria principal o central del sistema para almacenar los datos, estando optimizadas en velocidad. Están siendo cada vez más

utilizadas para analizar grandes volúmenes de datos por empresas como SAP, Oracle, Microsoft, SAS e IBM.

Las bases de datos nosql son bases de datos sin esquemas de tablas fijas, y están optimizadas para la lectura/escritura de grandes datos. Existen varios tipos según el formato en que almacenan los datos:

1. Clave-Valor: Asocian una clave única al valor que se quiera guardar lo que permite una gran flexibilidad, permitiendo guardar en cada clave diferentes tipos de datos(estructurados, no estructurados, ...). Como ej estaría Cassandra desarrollada por Facebook y luego donada a la fundación Apache que es actualmente la responsable de su desarrollo. La usan además de Facebook, empresas como Netflix, Twitter, Reddit, Cisco, IBM, etc...
2. Grafos: Como su nombre indica, almacenan la información en grafos. Lo que hacen es dividir la información en partes más básicas(nodos) y establece relaciones entre estos(edges). Como ej tenemos a Neo4j, que es una base de código abierto muy popular.
3. Tabulares/Columnares: Basadas en la tecnología BigTable de Google usan una estructura de claves multidimensional formada por una clave fila, una clave columna y tiempo con fecha y hora(timeline). Como ej tenemos apache Hbase que tiene una estrecha relación con Hadoop ya que se contruye en la parte superior de sus sistema de archivos(HDFS).
4. Documentos: Organizan sus datos en documentos desestructurados. Al acceder a los documentos se obtiene un número no especificado de campos con sus respectivos valores, que depende de cada documento. Como ej tenemos MongoDB, una de las bases de datos más utilizadas del mundo. Como ejemplo de empresas que la usan tenemos Foursquare , The New York Times o bit.ly .

Actualmente grandes empresas como Google, Amazon, Microsoft , IBM o Cloudera han creado grandes sistemas en cloud(nube) que usan esta tecnología, uniendo

cientos/miles de ordenadores, con sus procesadores y sistemas de almacenamiento, como si fuese un gran superordenador que tuviese todos esos procesadores y sistemas de almacenamientos disponibles para cualquier proceso/programa. Dichas empresas han puesto, estas capacidades, al alcance de las empresas en forma de alquiler, que los podrán usar para el análisis de Big Data, permitiendo que muchas de ellas puedan acceder a esta tecnología sin poseer los recursos para montar este tipo de sistemas.

Destacar que para un adecuado análisis de la información es necesario integrar la información de los sistemas de la empresa con estos Big Data. Para esta tarea, existen varias herramientas que nos permitirán importar los datos, ya que este tipo de tecnologías, como Hadoop, están preparadas para trabajar con datos estructurados.

Una vez sabiendo la gran cantidad de datos existente, y conociendo la existencia de tecnología que nos permite almacenarla y analizarla. ¿ Como nos puede ayudar a mejorar la experiencia del cliente en el ámbito de los CACT en particular?.

Para responder esta pregunta iré mostrando diferentes índices/KPIs de la experiencia de cliente(CX) y como el análisis de grandes datos puede ayudar a medirlos. Estos índices/KPIs han sido adaptados a los centros basándome en los más habituales que se usan en CX. :

1. Gasto medio por cliente: Cuanto más gastan mejor les parece la oferta de los CACT, lo que presupone una mejor experiencia.

Los bancos, hoy en día, ofrecen a las empresas información anónima del gasto de los clientes por tarjeta: diaria, mensual y por localización, obtenidos a través de su Big Data. Si analizamos los datos de venta de la empresa y le unimos los datos ofrecidos por los bancos, podemos tener una medida más real del gasto medio por cliente, pudiendo hacerle un seguimiento en todos los centros que visite.

2. Centros Visitados: Cuanto mayor sea la media de centros visitados por cliente mayor es el interés generado, lo que debería indicar una mejor experiencia.

Uniendo la información de venta de entradas del centro, los datos obtenidos por la Wifi y la app(pueden rastrear un mismo dispositivo en diferentes centros), más los datos proporcionados por compañías de telefonía usando su Big Data(nos pueden dar información de las rutas hechas por lo clientes), obtendremos un valor más real sobre el número de centros visitados por cliente.

3. Número de reclamaciones/quejas: Cuanto menor sea el número de reclamaciones/quejas, mejor será la experiencia.

Si a las reclamaciones/quejas realizadas a través de los diferentes canales que ofrece la empresa(teléfono, app, física, web), le unimos los datos obtenidos a través de redes sociales, páginas clave como tripadvisor, etc... ofrecidos por empresas especializadas que usan análisis de Big Data, podremos tener una mejor medida del número de quejas, ya que podremos incluir aquellas quejas de clientes que se sienten más cómodos usando los medios sociales.

4. Tiempos de espera: Cuanto menor sean los tiempos de espera, tanto en la entrada al centro como en la atención de los bares y tienda, mejor será la experiencia.

La información obtenida por la instalación de sensores que permitan medir el número de coches/personas en las colas, el número de personas en el interior de los centros, junto a información del istac sobre el número de visitantes a la isla, así como la información de aena y los puertos sobre el número de pasajeros, nos permitirá contrastar de manera más real el número de posibles visitantes a los centros junto con los días de mayor tiempo de espera y poder tomar mejores medidas para atender a los clientes.

5. Net Promotor Score (NPS): Es una medida estandar para evaluar los gustos de un cliente. Saber en nuestro caso si le ha gustado un centro, que partes del

centro le han gustado más y cuales menos. Si le ha gustado la comida o está contento con el servicio, la limpieza, etc...

Usando información obtenida de la app, cómo el tiempo que permanecen en los centros, en cada zona de los centros, datos obtenidos por las Wifis y los sensores sobre el número medio de personas que hay dentro de cada centro, los datos obtenidos por otros índices como el gasto medio, el número de reclamaciones/quejas, junto a información de análisis de sentimientos realizadas en las redes sociales y páginas como tripadvisor suministrada por empresas especializadas usando Big Data, podemos contrastar los sentimientos positivos/negativos del cliente con el tiempo que permanecen en el centro y la posible saturación que puedan tener los centros.

Como nota final, espero que este ensayo haya ayudado a entender la importancia de CX para las empresas y como el uso de Big Data junto con los datos corporativos permitirá obtener medidas más reales que ayudarán a las empresas a tomar decisiones más acertadas para mejorar sus negocios.

José Juan Molina Toledo