

USO DE BIG DATA PARA MEJORAR LA MOVILIDAD



AUTOR

ALVARO ROJAS GUARNIZO

Trabajo de grado presentado como requisito para optar al título de:

INGENIERIA INFORMATICA

Director:

OSCAR IVAN VARELA VELEZ

UNIVERSIDAD MILITAR NUEVA GRANADA

FACULTAD DE ESTUDIOS A DISTANCIA

PROGRAMA INGENIERIA INFORMATICA

BOGOTÁ, 05 MAYO 2023

INTRODUCCION

En las última dos décadas Bogotá ha incrementado en un ritmo acelerado la cantidad de habitantes que residen en la capital del país siendo la ciudad con mayor cantidad de habitantes en Colombia. De acuerdo con el censo realizado por el DANE en el año 2005 la ciudad contaba con una población total censada de 6.778.691 con una densidad 4143 habitantes por km² pasando al año 2018 a tener una población 7.387.400 es decir una densidad 4515 habitantes por km² por ultimo esta misma entidad estima que el total de la población para 2023 será 7.936.532 sin contar la población migrante tanto nacionales como extranjeros que llegan a la capital del país, por diferentes fenómenos, como el desplazamiento forzado por grupos al margen de la ley en el caso de los nacionales o en busca de nuevas oportunidades en el caso de los extranjeros.

Este aumento en la cantidad de habitantes se ve reflejado en el incremento de usuarios que día a día utilizan el sistema de transporte público de la ciudad. Teniendo presente el actual contexto el concepto de Sistema Inteligente de Transporte planteado en el Plan Distrital de Desarrollo 2020 – 2024 por la Alcaldía Mayor de Bogotá tiene una especial relevancia donde el uso de la tecnología facilitara gestionar el trafico de forma inteligente y mejorar la calidad de vida de los capitalinos.

El concepto de Sistema Inteligente de Transporte nace con el propósito por consolidar una movilidad sostenible, empleando el uso de datos, innovación y tecnología con el fin de tomar las mejores decisiones para mejorar la movilidad de los diferentes actores viales disminuyendo el tráfico vehicular, así como también los tiempos que emplean los ciudadanos en los diferentes trayectos que deben hacer en su día a día.

El presente escrito versa especialmente sobre como el uso de los datos es aplicado para gestionar el sistema de transporte público masivo de la ciudad.

¿Qué es Big Data?

El termino de Big data es un concepto que ha transformado el ámbito de la tecnología y el análisis de datos este ha sido planteado por diferentes personas, así como organizaciones para referirse a las diferentes tendencias tecnológicas encargadas de manejar grandes y complejos volúmenes de datos los cuales en algunos casos eran difícil o imposible de ser procesados y almacenados con las herramientas de software y hardware tradicionales. Los datos son recopilados, almacenados y por último analizados en el menor tiempo posible con el fin de tomar decisiones informadas de manera ágil. El Big Data hace uso de tecnologías como el procesamiento - almacenamiento distribuido y en la nube, el aprendizaje automático entre otros, a continuación, se plantea algunas de las definiciones más importantes.

El concepto Big data fue utilizado por primera vez en el año de 1997 por los investigadores de la NASA Michael Cox y David Ellsworth en el artículo científico “Application-controlled demand paging for out-of-core visualization” en el cual describen el concepto Big Data para referirse a conjuntos de datos de gran tamaño, que superaban la capacidad de cómputo de los recursos de la época y plantean la necesidad de desarrollar técnicas y herramientas que permitan el análisis de estos conjuntos de datos (Cox & Ellsworth, 1997).

De acuerdo con (Adrian Merv, 2011) vicepresidente de la consultora Gartner Big Data es un término utilizado para referirse a colecciones de datos que exceden las herramientas de hardware y software de uso común para realizar procesos de captura, gestión y procesamiento de datos dentro de un tiempo razonable para los usuarios.

La compañía (IBM, 2022) define Big Data como el término adoptado para referirse a las colecciones de datos con mayor tamaño y complejidad que sobrepasan la capacidad de las bases

de datos tradicionales encargadas de almacenar, administrar y procesar los datos con el nivel más bajo de retraso posible. La data en su mayoría es producida en tiempo real y en una escala muy grande por diversas fuentes entre las cuales se destacan sensores y dispositivos conectados a internet conocidos como internet de las cosas, redes sociales, aplicaciones transaccionales, sitios web, entre otros.

Características del Big Data

El big data inicialmente fue definido por (Russom, 2011) con unas características básicas las cuales hacían referencia a las 3Vs del big data (Volumen, Variedad y Velocidad).

Volumen: El volumen hace referencia a las cantidades de datos utilizados en los procesos de big data para la toma de decisiones y que cada día se van incrementando de manera exponencial, por otra parte como lo mencionan (Aranega, Kerkach, Hernandez, & Naomi, 2018) al hablar de volumen las personas tienden asociarlo a petabytes y zetabytes que es lo más popular que se encuentra en algunas literaturas cuando realmente no es así, los autores describen que se hace uso de Big Data cuando el conjunto de datos son superiores a un Terabyte.

Variedad: Cada día que pasa se incrementa la gama de fuentes de datos con las cuales se trabaja en big data podemos encontrar fuentes internas como lo son las interacciones con los clientes, empleados, proveedores, procesos de negocio y otras externas como redes sociales, dispositivos de internet de las cosas (sensores, cámaras, etc), entre otros. Estos tipos de fuentes también entregan los datos en diferentes tipos como podrían ser estructurados, semiestructurados y no estructurados (Schroeck, Shockley, Smart, Romero, & Tufano, 2012).

Los datos pueden tener diferente tipificación de acuerdo con la forma como estos son almacenados y organizados como se observan en la Tabla 1.

Tabla 1 Datos estructurados, semiestructurados y no estructurados

Estructurado	Están organizados conforme a un modelo o esquema. Se almacenan en forma tabular y algunas veces su estructura también incluye la definición de las relaciones entre ellos. Típicamente están representados en bases de datos que hacen parte del funcionamiento de sistemas de información.
No estructurados	Su organización y presentación no está guiada por ningún modelo o esquema. En esta categoría se incluyen, por ejemplo, las imágenes, texto, audios, contenidos de redes sociales, videos.
Semiestructurados	Su organización y presentación tiene una estructura básica (etiquetas o marcadores), pero no tiene establecida una definición de relaciones en su contenido. En esta categoría se incluyen contenidos de e-mails, tweets, archivos XML.

Fuente: (PLANEACIÓN, 2018)

Velocidad: Los datos continúan aumentando su velocidad en la creación, procesamiento y análisis. Esto se debe a la naturaleza en tiempo real de la creación de datos y a la necesidad de incorporar datos en streaming a los procesos de negocio y toma de decisiones. La velocidad tiene un impacto en la latencia, que es el tiempo de espera entre la creación, captura y acceso de los datos. Hoy en día, los sistemas tradicionales no son capaces de capturar, almacenar y analizar datos a la velocidad a la que se generan de forma continua es por esto que se usan frameworks como el de FastApi expuesto por (Montorio, 2022) para el análisis de datos en memoria para los que el tiempo es fundamental, por ejemplo para detectar fraudes en tiempo real. En algunos

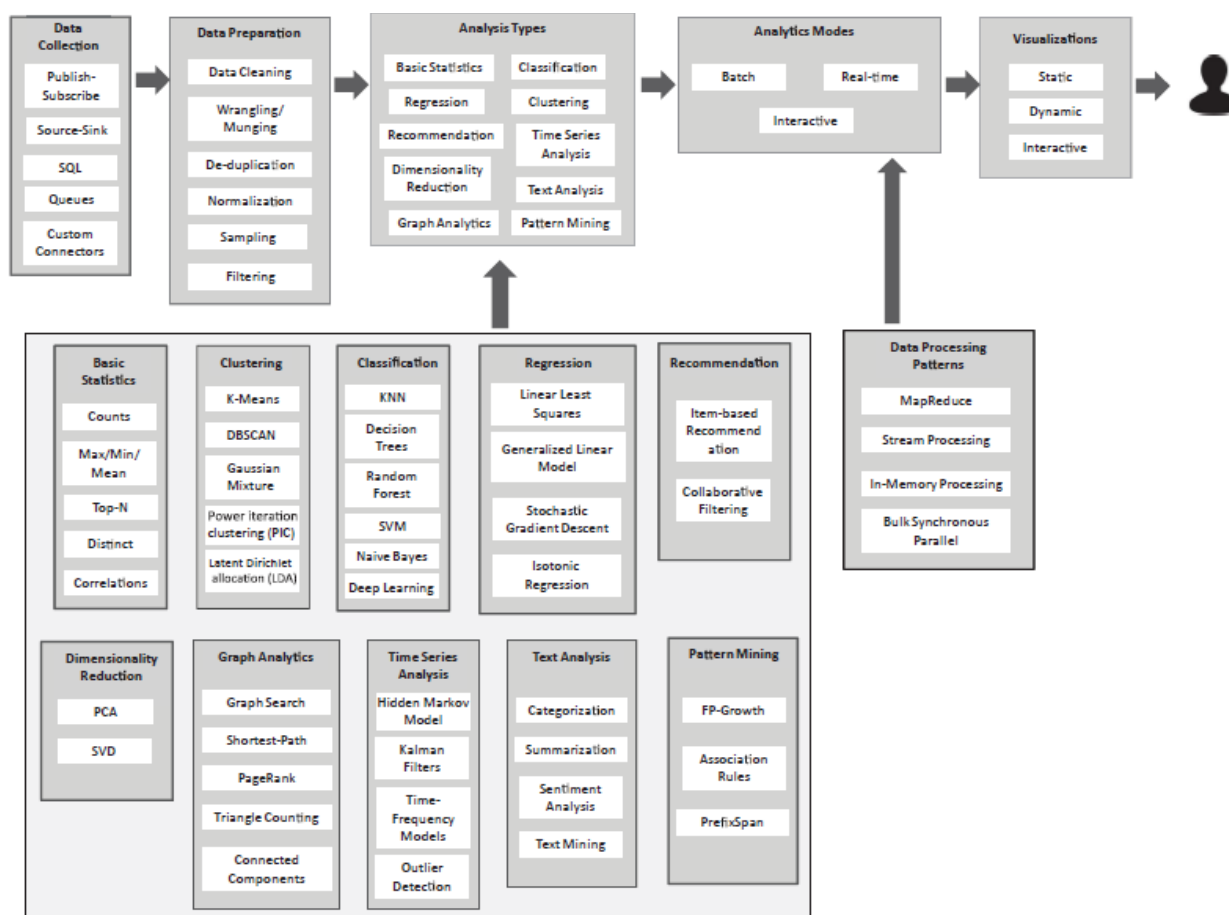
industrias, sectores económicos o entidades gubernamentales en las cuales la única forma de que los datos sean útiles es siempre y cuando estos sean procesados en tiempo real.

¿Como funciona Big Data?

(Bahga & Madisetti, 2019) Describen el flujo analítico del Big Data iniciando con la captura de los datos estos pueden ser producidos en tiempo real o consultados en bases de datos existentes, los datos son originados desde diferentes fuentes como redes sociales, sensores, dispositivos móviles, transacciones financieras, entre otros. Una vez se conocen la procedencia de los datos estos son almacenados en diferentes formatos como datos estructurados, no estructurados y semiestructurados dependiendo de su naturaleza, con el fin de que puedan ser accedidos y procesados por sistemas de almacenamiento distribuidos que permiten el almacenamiento de datos en múltiples nodos de clúster. Luego que los datos son almacenados se procede con el procesamiento de los datos con el fin de encontrar en esos datos información valiosa, el procesamiento se realiza con técnicas como procesamiento de lenguaje natural, minería de datos, análisis de redes sociales, entre otros. Con la información procesada se inicia el análisis de los datos con el fin de encontrar patrones y tendencias, así como la realización de análisis predictivos y el descubrimiento de relaciones ocultas entre los datos, la última fase del procesamiento consiste en visualizar los datos mediante visualizaciones como gráficos, tablas y mapas interactivos.

Los autores Bahga & Madisetti incluyen en su libro Big Data Analytics: A Hands-On Approach la figura 1 que describe el flujo del big data con un mayor nivel de detalle que el descrito anteriormente.

Figura 1 Flujo analítico del Big Data



Fuente: Big Data Analytics A Hands-On Approach

¿Ventajas y desventajas del big data?

El Big data proporciona una serie de ventajas y desventajas para los diferentes sectores que hacen uso de esta tecnología empresas públicas y privadas, así como la sociedad en general a continuación, se describen las que más inciden:

Ventajas:

Partiendo de la premisa que los datos son un activo fundamental de las organizaciones para conocer cuál es el mercado y su potencial (Nagore, 2023) identifica 8 ventajas que aporta el

Big data a las compañías con el fin de proporcionar una visión holística y resolver los principales retos a los cuales se ve abocada una compañía con menor tiempo y recursos:

- Mejora la toma de decisiones: El análisis de grandes volúmenes de datos permite obtener información más precisa y detallada, lo que ayuda a tomar decisiones más acertadas.
- Optimización de costos: El análisis de datos puede ayudar a detectar procesos ineficientes o áreas de la empresa donde se puede reducir costos relacionados con los procesos generando resultados positivos para las compañías.
- Mayor seguridad en los datos: Debido a la forma organizada en la cual se trabajan con los datos permite identificar riesgos internos, así como identificar datos sensibles que no estén asegurados de manera adecuada.
- Identifica problemas: El análisis de grandes volúmenes de datos permite identificar patrones y tendencias en el comportamiento de los clientes, lo que ayuda a predecir su comportamiento futuro y a adaptar la estrategia empresarial a sus necesidades.
- Mejor orientación al cliente y fidelización: El análisis de datos permite conocer mejor a los clientes y ofrecer servicios personalizados y adaptados a sus necesidades.
- Aumenta ingresos: El análisis de grandes volúmenes de datos puede ayudar a identificar nuevas oportunidades de negocio o nichos de mercado.
- Accesibilidad de la información: Una vez los datos se encuentran digitalizados y ordenados y se cuenta con herramientas que facilitan la búsqueda, se desarrollan dinámicas de trabajo fluidas con lo cual es más fácil llegar a la información más

relevante acerca de los clientes o usuarios de una compañía con el fin de poder establecer patrones de segmentación.

- Mejora gestión de los equipos: El desarrollo de Big Data desarrolla nuevas oportunidades para tecnologías como el aprendizaje automático (machine learning), minería de datos (data mining), análisis predictivo (predictive analytics), ciencia de datos (data science). Permitiendo identificar los colaboradores que generan mayor valor a la organización con el fin de aumentar su nivel de satisfacción.

Desventajas:

Según comentan (Mayer & Cukier, 2013) una de las desventajas que se presenta al trabajar con Big Data es la posible de la pérdida de privacidad y seguridad de la información personal debido a la exposición que puede ser ocasionada de manera involuntaria teniendo como resultado consecuencias negativas como las expuestas anteriormente.

Otra de las desventajas es señalada (O'Neil, 2016) quien describe que el procesamiento y análisis de grandes cantidades de datos pueden llevar a errores y conclusiones incorrectas si no se tienen en cuenta los sesgos y limitaciones de los datos utilizados.

Finalmente, otro desafío importante del big data es la necesidad de recursos tecnológicos y de personal capacitado para procesar y analizar grandes cantidades de datos. Esto puede ser costoso y dificultar el acceso al análisis de datos a pequeñas empresas y organizaciones con recursos limitados.

¿Sectores en que los que se utiliza big data?

El big data se puede utilizar en diferentes ámbitos empresas privadas y públicas, así como en organizaciones gubernamentales, algunos de los sectores donde más se utiliza Big Data son:

Comercio electrónico: (Linze & Jun, 2021) describen el uso de big data para realizar análisis de compras, patrones de comportamiento del cliente, recomendaciones de productos y personalización de la experiencia del cliente.

Salud: En la salud (Sayantan, Amandeep, Najmul, & Matti, 2020) hace referencia a las técnicas de big data para la investigación médica, la identificación de patrones de enfermedades y la predicción de brotes epidémicos. También se utiliza en el análisis de imágenes médicas y registros electrónicos de salud.

Finanzas: (Lopez, 2019) refiere el uso del big data en el análisis de riesgos, el comercio automatizado, el análisis de fraude y la toma de decisiones de inversión.

Transporte: El big data se utiliza en la gestión del tráfico, la predicción de la demanda de transporte público, la optimización de la gestión de flotas y por último en la Planificación y diseño de infraestructura de transporte de acuerdo con lo expuesto por (Ignacio, 2020), el sector del transporte público será el sector principal en el cual se aplique el desarrollo del presente escrito.

Agricultura: Con el fin de optimizar el uso de los recursos y mejorar los rendimientos de los cultivos a través del análisis de datos de sensores y sistemas de información geográfica se hace uso de Big Data (Chunling & Zunfeng, 2019).

¿Aplicaciones utilizadas en Big Data?

Entre las aplicaciones más utilizadas en Big Data se encuentran las mencionadas por (Dasgupta, 2018) en el libro Practical Big Data Analytics se describen las principales características, así como un paso a paso para su instalación a continuación, se mencionan las más importantes:

Hadoop: Es un framework de código abierto que permite el procesamiento distribuido de grandes conjuntos de datos en clusters de servidores haciendo uso de MapReduce y Pig. En cuanto al almacenamiento y análisis de datos lo realiza mediante herramientas como Hive y HBase.

Spark: Es un motor de procesamiento de datos en memoria que permite procesar grandes conjuntos de datos en tiempo real y en batch, así como también permite realizar análisis complejos.

NoSQL: El libro cubre el uso de bases de datos NoSQL, como MongoDB y Cassandra, para el almacenamiento y procesamiento de grandes conjuntos de datos no estructurados.

Caso de éxito de uso de Big Data en el transporte publico

Transport for London: La entidad encargada del transporte público de Londres (TFL, 2017) refiere hacer uso de los datos para mejorar la eficiencia y reducir los retrasos en su red de transporte. Utilizan sensores y datos de GPS para recopilar información en tiempo real sobre el tráfico y la ubicación de los vehículos, lo que les permite ajustar el tráfico de manera inteligente y optimizar la ruta y el horario de los autobuses y trenes.

¿Qué es el SITP?

De acuerdo con el Plan Maestro de Movilidad que es la ruta de navegación en los temas de movilidad, se establece la estructuración del SITP Sistema Integrado de Transporte Publico de Bogotá, con el fin de manejar un único medio con el cual se busca mejorar la calidad de vida de los bogotanos, mejorando los tiempos de servicio para los diferentes trayectos que se son realizados por los usuarios.

Entre las ventajas que tiene el SITP se encuentran poder realizar cambios entre los diferentes esquemas de rutas (Troncales, Urbanas, Alimentadoras, Complementarias y Especiales) en algunos sin costo adicional y en otros a bajo costo, todo con la misma tarjeta Tu llave (SITP, 2023).

¿Qué es el CGT?

Según el portal Tic Bogotá (TIC Bogota, 2023) el Centro de Gestion del Tránsito de Bogotá es un equipo que está conformado por profesionales expertos en movilidad y seguridad vial, que trabajan de manera mancomunada con otras entidades y organismos del gobierno para garantizar una movilidad eficiente y segura en la ciudad. Su principal objetivo es coordinar y controlar el tráfico vehicular en la ciudad mediante el uso de tecnología de última generación avanzada de software y hardware entre los que se destacan, sistema de semáforos inteligentes, sistema de cámaras, sensores de velocidad, sensores de conteo vehicular, entre otros dispositivos de acuerdo con lo expuesto por el Alto consejero Distrital de TIC Felipe Guzman.

Componentes del centro de gestión del tráfico

La secretaria de movilidad (Secretaria de movilidad, 2019) registra entre su inventario los siguientes elementos que hacen parte del centro de gestión del tránsito con el fin de que recopilen datos y sean gestionados:

- Sensores bicicletas
- Sensores de vehículos
- Circuito cerrado de televisión
- Herramientas geo-localizadas
- Sensorización inteligente
- Plataforma única de visualización
- Operación inteligente - Correlación de eventos
- Semaforización inteligente

Implementación Sistema Inteligente de Transporte (SIT) de Bogotá

La secretaria de Movilidad de Bogotá en alianza con la compañía de tecnología SKG realizaron la implementación del Sistema Inteligente de Transporte (SIT) de Bogotá con el cual se busca comprender cual es la dinámica de movilidad de los bogotanos, identificar cuáles son las principales dificultades y brindar soluciones mediante el uso de la tecnología, una vez realizados los procesos de análisis la administración decide optar por la solución Smart Helios el cual es una herramienta de inteligencia de negocios que permite obtener análisis de datos en tiempo real con el objetivo de tomar de manera inteligente y acertada decisiones desde el CGT.

Resultados de la implementación de SIT

En la página de la compañía SKG (SKG, 2023) describen el seguimiento a los resultados posterior a dos años de la implementación:

- Incremento del 12% en la velocidad promedio de tránsito en la ciudad
- Disminución del 42% en el tiempo empleado para dar respuesta a los incidentes de tránsito.
- Identificación de los principales problemas que reducen la movilidad en la ciudad
- Favoreció la toma de decisiones informadas con el propósito de priorizar el transporte sostenible.

Conclusiones

- Los datos y la tecnología son parte fundamental para desarrollar proyectos de gestión inteligente de la movilidad.
- El desarrollo de un sistema de gestión inteligente tiene la capacidad de tomar las mejores decisiones y así mejorar la calidad de vida de los ciudadanos, con la disminución del tiempo en la cual utilizan para desplazarse de un lugar a otro.
- Aunque se cuente con la mejor de las tecnologías no llegaran a ser suficientes en el caso de Bogotá, debido a la gran cantidad de obras que se están realizando en los diferentes sectores de la ciudad parte de las cuales a futuro permitirán tener una mejor movilidad.

Referencias

- Adrian Merv. (2011). Big Data. *Teradata Magazine*.
- Aranega, I., Kerkach, N., Hernandez, C., & Naomi, T. S. (2018). *BIG DATA ABSTRACT GRAPHIC*. Madrid.
- Bahga, A., & Madiseti, V. (2019). *Big Data Analytics: A Hands-On Approach*.
- Chunling, Z., & Zunfeng, L. (2019). Application of big data technology in agricultural internet of things. *International journal of distributed sensor networks*, 1 - 11.
- Cox, M., & Ellsworth, D. (1997). Application-controlled demand paging for out-of-core visualization. *ACM Transactions on Graphics*, 93-108.
- Dasgupta, N. (2018). *Practical Big Data Analytics*. Birmingham: Packt.
- IBM. (2022). *Análisis del Big Data*. Obtenido de IBM: <https://www.ibm.com/co-es/analytics/hadoop/big-data-analytics>
- Ignacio, L. (2020). *Aplicabilidad de herramientas de big data y business analytics a la mejora de la ecomovilidad urbana: el caso bicimad*. Madrid.
- Linze, L., & Jun, Z. (2021). Research and Analysis of an Enterprise E-Commerce Marketing System Under the Big Data Environment. *Journal of Organizational and End User Computing*, 5 - 6.
- Lopez, P. (2019). *Oportunidades estratégicas y riesgos en la implementación del big data en el sector financiero español*. Madrid.
- Mayer, V., & Cukier, K. (2013). *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. Houghton Mifflin Harcourt.
- Montorio, A. (2022). *Big Data para la comprensión de Métricas de E-Sports usando tecnología Web*. Madrid.
- Nagore, A. (10 de 01 de 2023). *apd*. Obtenido de <https://www.apd.es/big-data-ventajas/>
- O'Neil, C. (2016). *Weapons of math destruction*. New York: Crown.
- PLANEACIÓN, D. N. (2018). *POLÍTICA NACIONAL DE EXPLOTACIÓN DE DATOS*. Bogota.
- Russom, P. (2011). *Big Data Analytics*. Renton: tdwi.
- Sayantan, K., Amandeep, D., Najmul, I., & Matti, M. (2020). Big data analytics in healthcare: a systematic literature review. *Enterprise information Systems*, 878 - 912.
- Schroeck, M., Shockley, R., Smart, J., Romero, D., & Tufano, P. (2012). *Analytics: el uso de big data en el mundo real*. España: IBM.
- Secretaria de movilidad. (17 de 11 de 2019). *Secretaria de movilidad*. Obtenido de <https://www.movilidadbogota.gov.co/web/node/2416>
- SITP. (03 de 05 de 2023). *SITP*. Obtenido de <https://www.sitp.gov.co/publicaciones/40075/informacion-general/>

SKG. (03 de 05 de 2023). *SKG Tecnologia*. Obtenido de <https://skgtecnologia.com/tag/sistema-inteligente-de-transporte/>

TFL. (13 de 10 de 2017). *TFL*. Obtenido de <https://tfl.gov.uk/info-for/media/press-releases/2017/october/tfl-s-free-open-data-boosts-london-s-economy>