

IMPORTANCIA DE LA SOLUCION BIG DATA EN LA APLICACIÓN DE
MOVILIDAD UBER MOVEMENT

Monografía Diplomado Big Data



Autor:

ANDRES CAMILO MONTEALEGRE GALLO

COD 066082025

UNIVERSIDAD LIBRE
FACULTAD DE INGENIERÍA
PROGRAMA DE INGENIERÍA DE SISTEMAS
BOGOTÁ D.C.
SEPTIEMBRE 2017



**UNIVERSIDAD
LIBRE®**

TITULO

Importancia de la solución Big Data en la aplicación de movilidad Uber Movement.

RESUMEN

Nos encontramos en un mundo que cambia muy rápido, donde las innovaciones disruptivas son cada vez más frecuentes por lo que la sociedad demanda soluciones rápidas y eficaces para los nuevos retos.

Desde los tiempos, la importancia de la solución Big Data ha cobrado una mayor relevancia a la hora de obtener análisis sobre los datos en diferentes sectores empresariales. En esta monografía estaré enfocado en el sector de movilidad y para ello analizaré la plataforma llamada Uber Movement lanzada el mes pasado en la ciudad de Bogotá, la cual ha sido creada para descubrir patrones y analizar el impacto de eventos generados en las vías de la ciudad, basado en información de 2016 y del primer semestre del presente año.

En razón a ello, este trabajo cuenta con el propósito de exponer las metodologías que en la actualidad propone Big Data para el análisis de información de los datos que se adquieren a través de la aplicación de movilidad Uber Movement.

Palabras Claves: Metodologías, Patrones, Movilidad, Big Data, Uber y Uber Movement.

INTRODUCCIÓN

Esta monografía se despliega usando como fuentes de información lo aprendido en el diplomado de Big Data, realizado en convenio con la Universidad de Guadalajara - México y la Universidad Libre de Colombia. Hoy en día Big Data se describe como un conjunto de recursos que permiten la gestión y análisis de cantidades masivas de datos, con un alcance y dimensiones en constante crecimiento.

El propósito de este documento es evidenciar como Uber ha implementado las tecnologías que ofrece Big Data para crear la plataforma gratuita llamada Uber Movement que cuenta con datos sobre la forma en la que se comporta la movilidad en Bogotá y la cual es utilizada como alternativa de transporte útil para el ciudadano, permitiendo medir el impacto que tengan las vías de la ciudad y tomar decisiones que ayuden a el usuario a elegir la mejor opción a la hora de movilizarse.

La tecnología Big Data comprende las V's (Volumen, Velocidad, Variedad, Veracidad, Viabilidad, Visualización de los datos, Valor de los datos), las cuales se enmarcan dentro de la aplicación de Uber Movement al momento de gestionar los datos suministrados por el usuario.

Uber Movement es una aplicación creada pensando en la agilidad de movilización de los ciudadanos con el fin de tomar la ruta más corta o más rápida para llegar a su destino.

Como conclusiones hablaremos de la importancia de Uber Movement en las grandes ciudades, aplicando los pilares de las mismas y sus diferentes variables.

DESARROLLO

BIG DATA

Un concepto¹ que significa muchas cosas para muchas personas, ha dejado de estar limitado al mundo de la tecnología. Hoy en día se trata de una prioridad empresarial dada su capacidad para influir profundamente en el comercio de una economía integrada a escala global. Además de proporcionar soluciones a antiguos retos empresariales, big data inspira nuevas formas de transformar procesos, empresas, sectores enteros e incluso la propia sociedad. Aun así, la amplia cobertura mediática que está recibiendo no nos permite distinguir claramente el mito de la realidad: ¿qué está ocurriendo realmente? Tras varias investigaciones (IBM Institute for Business Value, 2012) se ha descubierto que las empresas utilizan big data para obtener resultados centrados en el cliente, aprovechar los datos internos y crear un mejor ecosistema de información.

Hoy en día el término “big data” es una constante y, aun así, la mera noción sigue generando confusión. Esta palabra se ha utilizado para trasladar al público todo tipo de conceptos entre los que se incluyen grandes cantidades de datos, analítica de redes sociales, herramientas de última generación para gestionar los datos, datos en tiempo real y mucho más. Independientemente de la etiqueta que le colguemos, las empresas comienzan a comprender y explorar cómo procesar y analizar de nuevas formas una amplia variedad de información. Al hacerlo, un pequeño pero creciente grupo de pioneros está logrando resultados empresariales importantísimos.

¹ http://www-05.ibm.com/services/es/gbs/consulting/pdf/El_uso_de_Big_Data_en_el_mundo_real.pdf

Una vez se haya considerado un conjunto de datos como Big Data, será necesario definir la arquitectura del sistema, basada en las cinco capas típicas de los sistemas de análisis de información, mostradas en el siguiente diagrama.

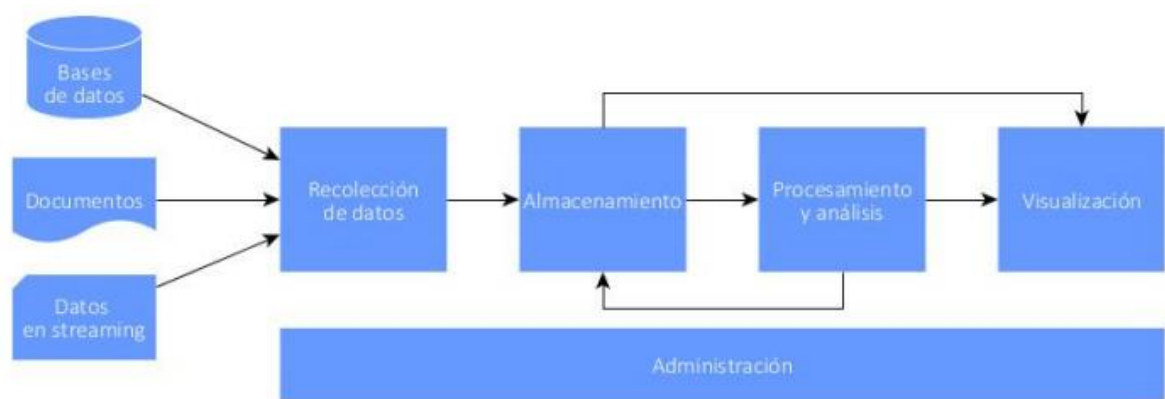


Figura 1. Arquitectura Big Data

Fuente: (Análisis y desarrollo plataforma Big Data, 2015)

Dentro de la tecnología que se ha desarrollado para analizar Big Data se encuentra la más conocida, Hadoop. Se trata de una plataforma software que provee almacenamiento distribuido y capacidades computacionales bajo una licencia libre. Hadoop se divide en dos componentes principales: HDFS para almacenamiento distribuido y MapReduce para capacidades computacionales, ambos basados en un modelo maestro-esclavo.

LAS V'S DE BIG DATA

Las características más importantes del Big Data perfectamente se pueden clasificar en cuatro magnitudes, más conocidas como las cuatro V del Big Data, relativas a volumen, velocidad, variedad y veracidad. A estas cuatro V, podemos añadir tres más, como pueden ser la de Viabilidad y Visualización. Pero si

hablamos de V en Big Data no podemos dejar pasar la principal característica del análisis de datos que es la V de Valor de los datos. Así pues, se empieza a hablar, ya no de las tradicionales cuatro V de Big Data, sino de las 7 “V” del Big Data²:

1. Volumen
2. Velocidad
3. Variedad de los datos
4. Veracidad de los datos
5. Variabilidad
6. Visualización de los datos
7. Valor de los datos



² <http://www.iic.uam.es/innovacion/big-data-caracteristicas-mas-importantes-7-v/>

Figura 2. Las 7 V del Big Data
Fuente: Francisco Javier Antón Vique (Fundación Big Data, 2013)

Volumen de información

El volumen se refiere a la cantidad de datos que son generados cada segundo, minuto y días en nuestro entorno. Es la característica más asociada al Big Data, ya que hace referencia a las cantidades masivas de datos que se almacenan con la finalidad de procesar dicha información, transformando los datos en acciones.

Cada vez se está más conectado al mundo por lo que se ha generado más y más datos. Para algunas empresas, el estar en el mundo digital es algo obligatorio, por lo que la cantidad de datos generados es aún mayor. Por ejemplo, una empresa que vende sus productos únicamente a través de un canal online, le convendría implantar tecnología Big Data para procesar toda aquella información que recoge su página web rastreando todas las acciones que lleva a cabo el cliente; conocer donde cliquee más veces, cuántas veces ha pasado por el carrito de la compra, cuáles son los productos más vistos, las páginas más visitadas, etc.

Velocidad de los datos

La velocidad se refiere a los datos en movimiento por las constantes interconexiones que realizamos, es decir, a la rapidez en la que son creados, almacenados y procesados en tiempo real.

Para los procesos en los que el tiempo resulta fundamental, tales como la detección de fraude en una transacción bancaria o la monitorización de un

evento en redes sociales, estos tipos de datos deben estudiarse en tiempo real para que resulten útiles para el negocio y se consigan conclusiones efectivas.

Variedad de los datos

La variedad se refiere a las formas, tipos y fuentes en las que se registran los datos. Estos datos pueden ser datos estructurados y fáciles de gestionar como son las bases de datos, o datos no estructurados, entre los que se incluyen documentos de texto, correos electrónicos, datos de sensores, audios, vídeos o imágenes que se tienen en los dispositivos móviles, hasta publicaciones en los perfiles de redes sociales, artículos que se leen en blogs, las secuencias de clic en una misma página, formularios de registro e infinidad de acciones más que se realizan desde un Smartphone, Tablet y ordenador.

Estos últimos datos requieren de una herramienta específica, debido a que el tratamiento de la información es totalmente diferente con respecto a los datos estructurados. Para ello, las empresas necesitan integrar, observar y procesar datos que son recogidos a través de múltiples fuentes de información con herramientas cualificadas.

Veracidad de los datos

Cuando se trata de veracidad es pertinente hacer referencia a la incertidumbre de los datos, es decir, al grado de fiabilidad de la información recibida.

Es necesario invertir tiempo para conseguir datos de calidad, aplicando soluciones y métodos que puedan eliminar datos imprevisibles que puedan surgir como datos económicos, comportamientos de los consumidores que puedan influir en las decisiones de compra.

La necesidad de explorar y planificar la incertidumbre es un reto para el Big Data que está a la orden del día en las compañías dedicadas al análisis de datos.

Variabilidad

La inteligencia empresarial es un componente fundamental para la variabilidad de un proyecto y el éxito empresarial. Se trata de la capacidad que tienen las compañías en generar un uso eficaz del gran volumen de datos que manejan.

La inteligencia competitiva también se asocia con la innovación de los equipos de trabajo y el uso de tecnologías empleadas. Una empresa inteligente analiza, selecciona y monitoriza la información con el fin de conocer mejor el mercado en el que opera, a sus clientes y diseñar estrategias eficaces.

Es necesario filtrar a través de esta información y seleccionar cuidadosamente los atributos y factores que son capaces de predecir los resultados que más interesan a las empresas. El secreto del éxito es descubrir las relaciones entre las variables ocultas.

Una vez conocida la viabilidad de cada empresa, es el momento de detallar el proyecto en una hoja de ruta, y desarrollar el plan de negocio.

Visualización de los datos

La visualización de datos se refiere al modo en el que los datos son presentados. Una vez que los datos son procesados (los datos están en tablas y hojas de cálculo), es necesario representarlos visualmente de manera que sean legibles y accesibles, para encontrar patrones y claves ocultas en el tema a investigar. Para que los datos sean comprendidos existen herramientas de visualización que ayudarán a comprender los datos gráficamente y en perspectiva contextual.

Valor de los datos

El dato no es valor. Tampoco tienes valor por el hecho de recopilar gran cantidad de información. El valor se obtiene de datos que se transforman en información; esta a su vez se convierte en conocimiento, y este en acción o en decisión. El valor de los datos está en que sean accionables, es decir, que los directivos de las empresas puedan tomar una decisión (la mejor decisión) en base a estos datos.

No todos los datos de los que se inicia se convierten en acción o decisión. Para ello, es necesario tener tecnologías aplicadas. Por ejemplo, una publicación en una red social, que gracias al uso de tecnologías de procesamiento de lenguaje natural, puede medir el sentimiento positivo o negativo, con la ayuda de un algoritmo de análisis de redes sociales o herramientas que permitan obtener de esto información.

En definitiva, el Big Data es una combinación de estas siete³ características donde las empresas pueden obtener una ventaja competitiva frente a sus competidores contribuyendo al éxito asegurado con las tres últimas V de variabilidad, visualización y la más importante, la de valor.

No todas las compañías optaran por la misma metodología con respecto al desarrollo y la creación de sus capacidades con tecnologías Big Data. Sin embargo, en todos los sectores existe la posibilidad de recurrir a estas nuevas tecnologías y analíticas para mejorar la toma de decisiones y el rendimiento, tanto a nivel interno como en el mercado.

UBER

³ <http://www.iic.uam.es/innovacion/big-data-caracteristicas-mas-importantes-7-v/>

Uber⁴ Technologies Inc. es una empresa internacional que proporciona a sus clientes una red de transporte privado, a través de su software de aplicación móvil (App), que conecta los pasajeros con los conductores de vehículos registrados en su servicio, los cuales ofrecen un servicio de transporte a particulares. La empresa organiza recogidas en cientos de ciudades de todo el mundo y tiene su sede en San Francisco (California).

UBER MOVEMENT

Esta plataforma ofrecerá de manera gratuita y abierta al público información sobre los tiempos de desplazamiento de una zona a otra, lo cual se traducirá en análisis sobre la movilidad de la capital del país, basado en datos históricos recientes. El tiempo que arroja la aplicación es un estimado basándose en los datos obtenidos de Uber, no es en tiempo real.

Además de explorar los tiempos de viaje, al ingresar a movement.uber.com con una cuenta de Uber los usuarios pueden consultar información de 2016 y del primer semestre del 2017. Esta información permite medir el impacto que tuvieron sobre el tráfico obras, cierres de vías, eventos masivos, entre otros. Posteriormente se hacen análisis sobre el transporte y la movilidad de la capital del país de manera gratuita.

De acuerdo con información recopilada por Uber Movement⁵, un ciudadano que realiza un recorrido de lunes a viernes en horas pico desde la zona de Unicentro hasta el Centro Internacional, que utiliza el Deprimido de la NQS inaugurado en el mes de marzo, se ahorra en promedio 3,3 horas al mes comparado con el mismo recorrido en 2016. En promedio, en esta ruta los capitalinos en 2016 se

⁴ <https://es.wikipedia.org/wiki/Uber>

⁵ <http://www.enter.co/chips-bits/apps-software/uber-movement-movilidad/>

demoraban 50 minutos. Hoy con esta obra pueden hacer sus recorridos en un tiempo promedio de 39 minutos.

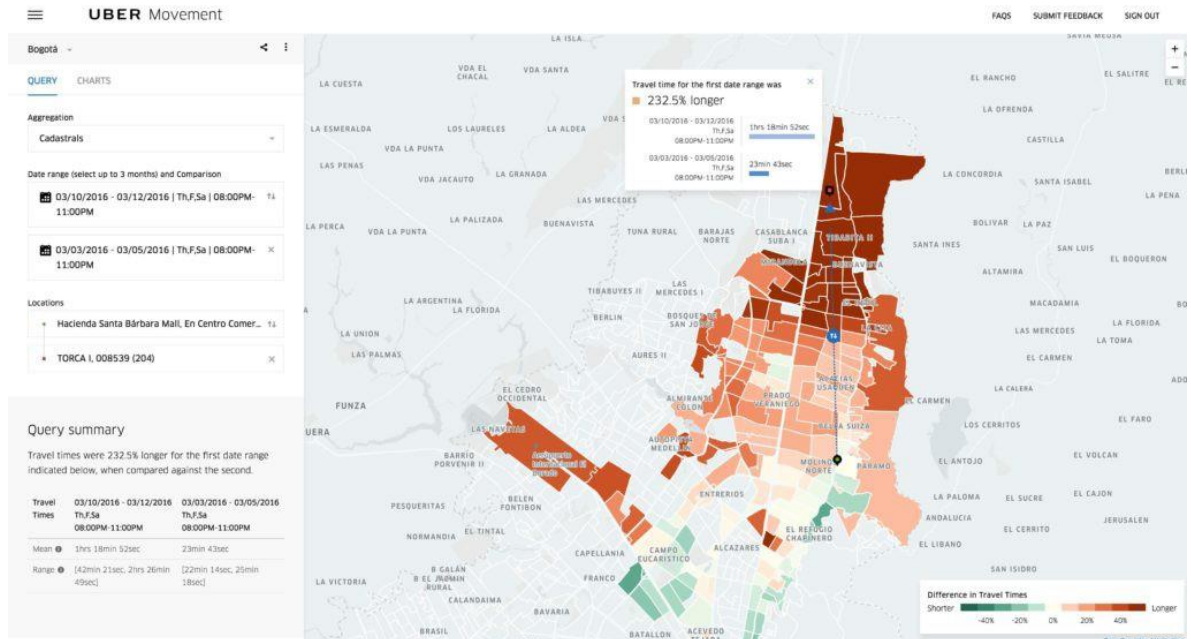
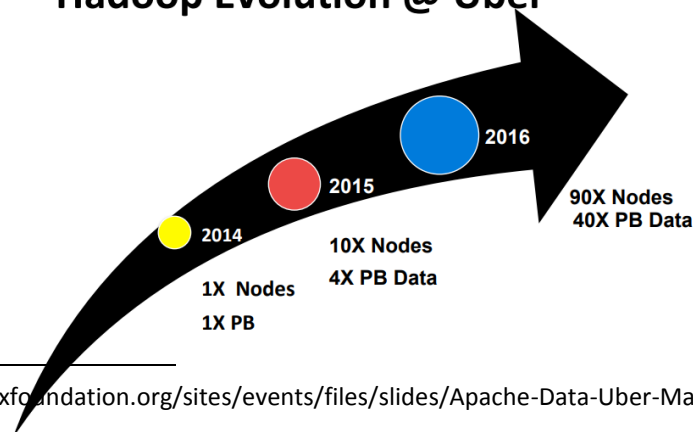


Figura 3. Uber Movement mide el impacto de obras y reparaciones sobre el tráfico. Fuente: (Captura de Pantalla)

Es importante destacar que en la actualidad la Uber Movement utiliza como framework a Hadoop y teniendo una exponencial de evolución en cuanto al procesamiento en los datos según (Mayank Bansal⁶, 2017) como lo muestra la siguiente imagen.

Hadoop Evolution @ Uber



⁶ <http://events.linuxfoundation.org/sites/events/files/slides/Apache-Data-Uber-Mayank-Bansal.pdf>

*Figura 4. Evolución Hadoop.
Fuente: (Mayank Bansal, 2015)*

Por otro lado, Bogotá es la primera ciudad en América Latina en contar con esta herramienta. También es una de las cinco primeras ciudades del mundo con Uber Movement. Con esto, la compañía quiere ratificar su compromiso con el desarrollo del país. Uber Movement quiere poner a disposición data segura y confiable que contribuya a la consolidación de datos e información para el desarrollo y planeación de movilidad en Bogotá.

Camila Escallón, del equipo de Asociada Regional de Asuntos Públicos Regional de Uber afirmó: *“Como parte de nuestro compromiso con la construcción de ciudades inteligentes, Uber Movement pone a disposición de todos una herramienta que permite de manera fácil y visual monitorear cómo se mueve Bogotá. La toma de decisiones basada en datos recientes y confiables podría ahorrarle a la ciudad millones de pesos con grandes efectos en la calidad de vida de los bogotanos.”*

Uber Movement utiliza información anonimizada y agregada de los GPS de la plataforma de Uber. Con esto entrega información sobre el comportamiento del tráfico de las ciudades donde está disponible. La aplicación registra la latitud y longitud de los viajes con una marca de tiempo cada cuatro segundos. Estos registros del GPS establecen tiempos promedio de viaje en zonas específicas de una ciudad.

“En Uber queremos convertirnos en aliados de las ciudades para que sean cada vez más modernas. Creemos en el valor del trabajo conjunto entre el sector público y privado en la construcción de mejores ciudades. Esperamos que Bogotá experimente la contribución de Movement a la eficiencia, innovación y precisión de las decisiones de ciudad basadas en información confiable y segura”, concluyó Escallón.

De acuerdo a esta información, se puede vislumbrar la importancia que ha generado la aplicación de Uber Movement en la gestión del tráfico y movilidad de

la ciudad. Consolidar datos y analizarlos a través de determinados patrones, se ha convertido en una ayuda trascendental en la planeación de rutas en la movilidad de Bogotá y otras grandes ciudades del mundo.

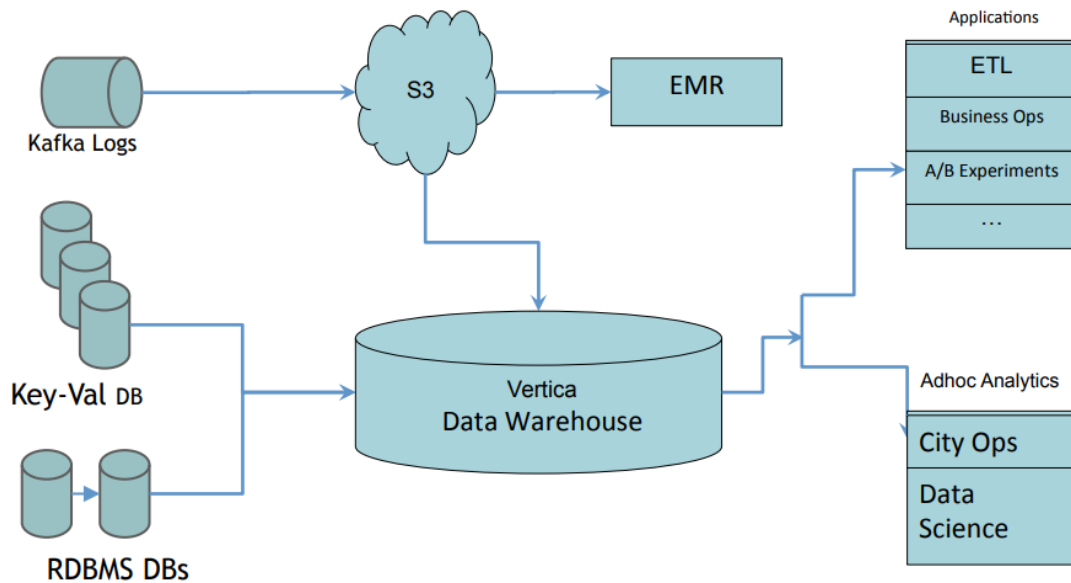
IMPORTANCIA BIG DATA EN UBER MOVEMENT

Es importante mencionar que por medio de esta plataforma se crea una nueva oportunidad para convertirse en aliado para la movilidad de la ciudad y del país y según manifestó Camila Escallon, asociada regional de asuntos públicos Colombia es un mercado interesante con un crecimiento positivo y es que las soluciones Big Data son una herramienta tan ambicionada por las empresas que hasta el momento sigue en alza silenciosamente debido a que no toda persona sabe lo que significa este concepto pero que sigue dándose a conocer de a poco con ayuda de las redes sociales que gobiernan hoy por hoy las sociedades modernas.

Uber es uno de esos servicios que nació de la necesidad de una mejor movilidad en las ciudades, y como tal, parece haber decidido ayudar con su experiencia a que esta sea mejorada y con Uber Movement una plataforma donde las instituciones públicas y del gobierno podrán acceder a los datos que la aplicación ha ido generando en los últimos años.

Y es que sin la aplicación de soluciones Big Data, quizás Uber no hubiera podido crear esta plataforma que por medio de los datos históricos predice por medio de patrones y eventos que se reportan en las vías, la mejor alternativa para llegar a su destino.

A continuación se muestra en este del diagrama, la infraestructura que Uber tiene hoy en día en la administración de sus bases de datos:



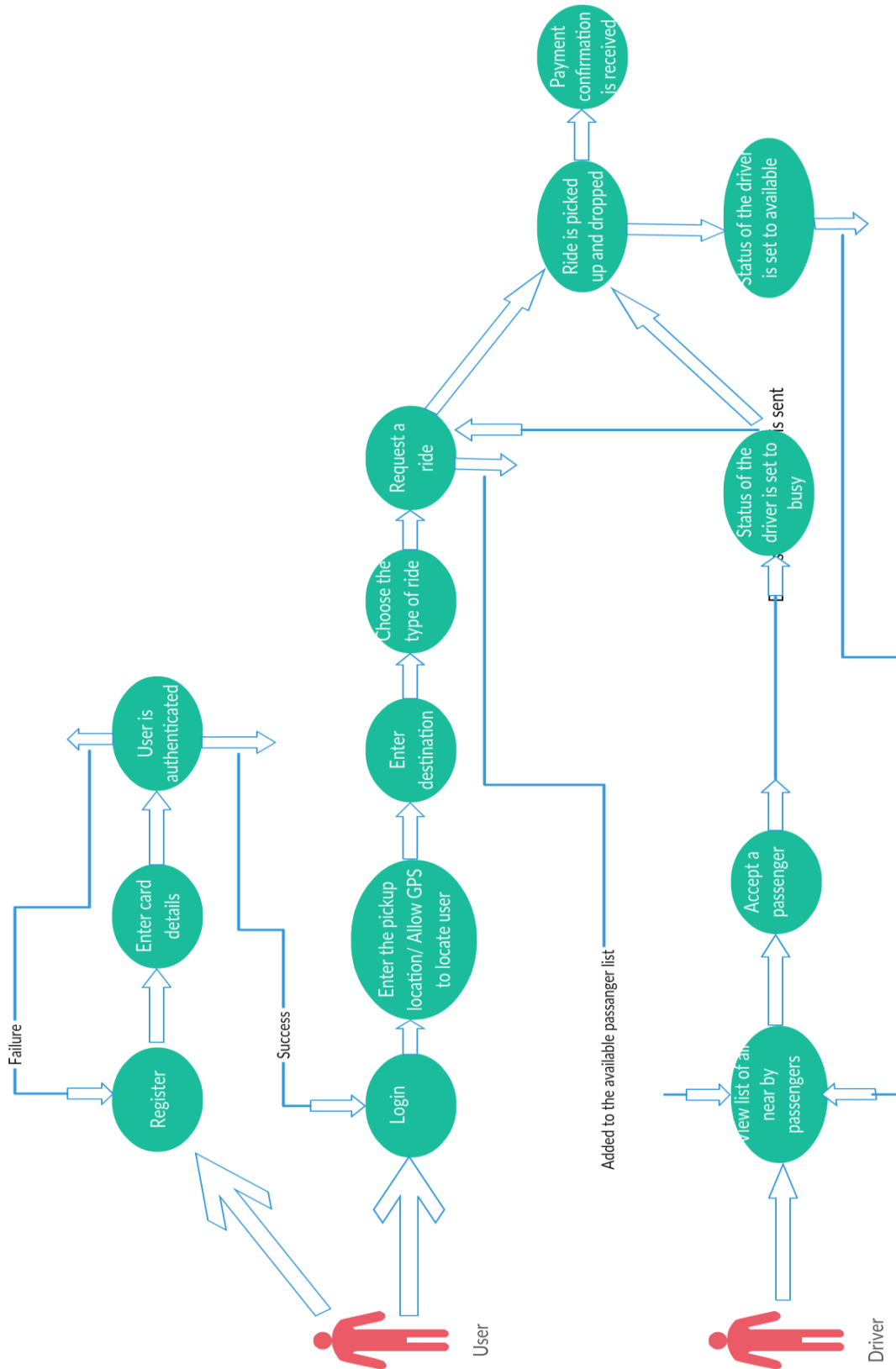
*Figura 4. Data Infrastructure Uber
Fuente: (Mayank Bansal, 2015)*

En el diagrama podemos ver que la infraestructura se centraliza en un Data Warehouse conectado con bases de datos llave valor y RDBMS así mismo con aplicaciones ETL y Datascience. Y es que Uber dedica tiempo y esfuerzos para conocer el impacto positivo que sus iniciativas tienen sobre las ciudades en las que operan. Los datos procedentes de los miles y miles de carros que trabajan en su plataforma generan miles de millones de posiciones GPS cada día. Realizan cinco millones de viajes al día en 450 ciudades de todo el mundo, todo un despliegue de Big Data (Transformación digital, muycomputerpro, 2017)⁷.

Y es que en cuanto al crecimiento que ha tenido Uber como compañía, es importante ilustrar con un diagrama de caso de uso su funcionamiento a grandes rasgos, como lo muestra la siguiente imagen.

⁷ <http://www.muycomputerpro.com/zona-transformacion-digital/uber-uso-big-data/>

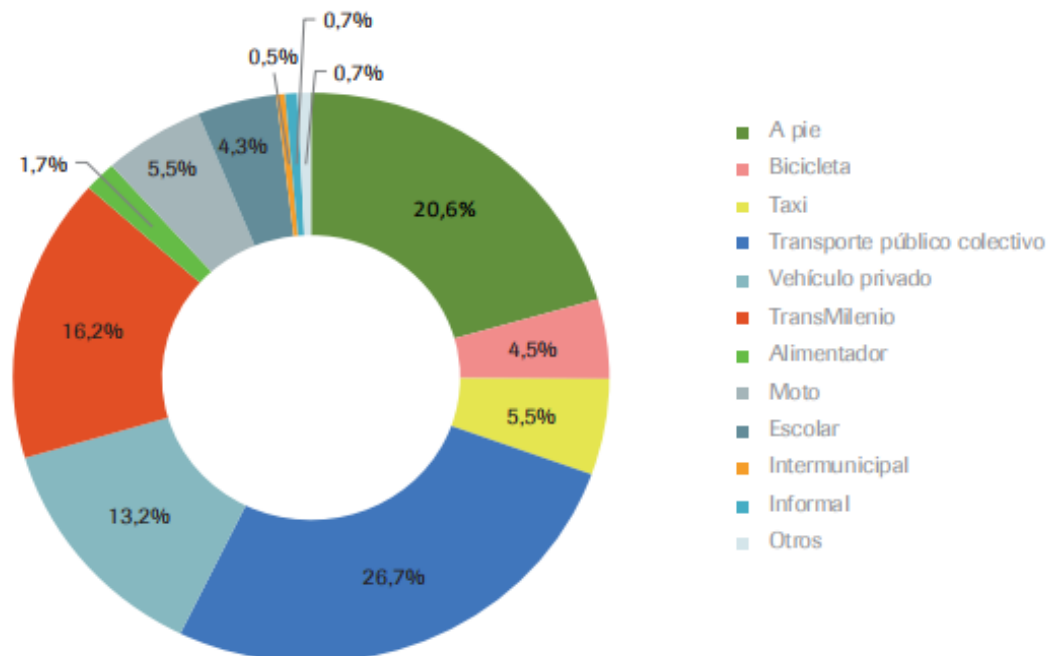
Figura 5. Diagrama Caso de Uso Uber



Y es que en Uber buscan las implicaciones que tienen los datos de los que disponen, por ejemplo utiliza los picos de demanda hora a hora los sábados por la noche en varias ciudades. Gracias a esta visualización simultánea, se puede “adivinar” la salida de los bares un sábado noche en dichas ciudades. Y eso es información muy efectiva sobre la forma en la que Big Data sirve para conocer la dinámica de la ciudad y las costumbres de sus ciudadanos.

INDICADORES DE MOVILIDAD EN BOGOTÁ

De acuerdo con la movilidad aplicada en el 2015, a diario en la ciudad se hacen 12.755.826 viajes mayores a 15 minutos. Predominan los viajes en transporte público (transporte público colectivo, Transmilenio, alimentador) con 45% seguido de los viajes en transporte no motorizado (a pie y bicicleta) con 25% y los viajes en vehículo privado con 24% (moto, automóvil, taxi) según la siguiente gráfica.



*Figura 6. Porcentaje de viajes diarios mayores a 15 minutos
Fuente: Secretaria Distrital de Movilidad (SDM), 2015.*

Y es que teniendo en cuenta la proyección de Bogotá como una Smart City⁸ para la actual administración y como un logro de la política de Movilidad de la Administración de la Bogotá Humana, con esfuerzo y dedicación Uber Movement lograra crear grandes alianzas con las instituciones públicas con el fin de hacer de Bogotá una ciudad moderna como las grandes ciudades del mundo a nivel de movilidad de sus ciudadanos.

⁸ Una Smart City, o ciudad inteligente, se puede describir como aquella ciudad que aplica las tecnologías de la información y de la comunicación (TIC) con el objetivo de proveerla de una infraestructura que garantice un desarrollo sostenible

CONCLUSIONES

Al finalizar el diplomado, y luego de analizar el uso de Big Data que le da la plataforma de Uber Movement se puede concluir lo siguiente:

- Las soluciones big data como conjunto de recursos que permiten la gestión y análisis de cantidades masivas de datos, juegan hoy en día un papel importante en las empresas que tienen la necesidad de la información para planear u organizar estrategias de mercado que impacten en la sociedad generando un valor positivo en sus clientes.
- Implantar tecnología Big Data para procesar toda aquella información que recoja una aplicación móvil, un sensor de movimiento, una cámara de video, un GPS, una página web, rastreando todas las acciones que lleva a cabo el cliente; conocer donde cliquea más veces, cuántas veces ha pasado por el carrito de la compra, cuáles son los productos más vistos, las páginas más visitadas, etc.
- Es importante resaltar que en la actualidad la Uber Movement se apoya en el framework Hadoop y que sigue evolucionando a medida que los usuarios aceptan la plataforma y la adoptan a su movilidad beneficiando de manera útil a los ciudadanos en la gestión del tráfico y movilidad de la ciudad.
- Consolidar datos y analizarlos a través de determinados patrones, se ha convertido en una ayuda trascendental en la planeación de rutas en la movilidad de Bogotá y otras grandes ciudades.
- La necesidad de explorar y planificar la incertidumbre es un reto para el Big Data que está a la orden del día en las compañías dedicadas al análisis de datos.
- En definitiva, el Big Data es una combinación de las V's donde las empresas pueden obtener una ventaja competitiva frente a sus competidores contribuyendo al éxito asegurado con viabilidad, visualización y valor.

BIBLIOGRAFIA

Infraestructura Hadoop implementada por Uber (2016) disponible en <http://events.linuxfoundation.org/sites/events/files/slides/Apache-Data-Uber-Mayank-Bansal.pdf>

Portafolio Artículo donde Uber lanza nueva plataforma (2017) disponible en: <http://www.portafolio.co/negocios/uber-lanza-plataforma-que-da-informacion-sobre-la-movilidad-en-bogota-509173>

Enter.co El google maps de Uber llega a Bogotá (2017) disponible en: <http://www.enter.co/chips-bits/apps-software/uber-movement-movilidad/>

Instituto de ingeniería del conocimiento las V's del Big Data disponible en: <http://www.iic.uam.es/innovacion/big-data-caracteristicas-mas-importantes-7-v/>

Big Data para solucionar problemas de movilidad disponible en: <https://blogs.iadb.org/gobernarte/2016/03/21/big-data-para-solucionar-problemas-de-movilidad/>

El futuro de la movilidad disponible en: https://www.opinno.com/sites/default/files/el_futuro_de_la_movilidad_0.pdf

Big Data en Español disponible en: <http://www.aacarreteras.org.ar/pdf/Big-Data-Espanol.pdf>