4.1 ¿Qué se necesita para trabajar big data?

Al iniciar una carrera en el mundo del análisis de datos, todo profesional se ve a sí mismo con incertidumbre. Esto se da en un contexto en el que la humanidad atraviesa una nueva era parecida a la revolución industrial, pero en esta ocasión relacionada con la revolución del conocimiento y la explosión de los datos que el mundo genera. Las nuevas profesiones relacionadas con la industria del *big data* han sido incluso nombradas por la revista "The Economist" como el nuevo trabajo más sexy del siglo XXI (http://www.eleconomista.es/ecoaula/noticias/8883560/01/18/Cientifico-de-datos-la-profesion-mas-sexy-del-siglo-XXI.html).

Tan solo pensando que el 98% de los datos generados a lo largo de la historia de la humanidad se produjo en los últimos 2 años se dimensiona lo que este campo representa para el funcionamiento del mundo en este nuevo siglo.

El impacto que los datos generan en los resultados de cualquier compañía, Estado, institución o carrera profesional es innegable y cada vez más evidente. Las normativas y legislaciones de todos los países —tarde o temprano— tendrán que ser actualizadas para abarcar en su proceder legal a las nuevas formas de comunicación, interacción, comportamiento y compromiso social que existe alrededor de los datos.

Estas situaciones hacen que la industria de los datos sea atractiva y compleja a la vez. En ocasiones se idealiza a las disciplinas de este rubro mostrándolas como altamente futuristas o de ciencia ficción (en películas y otros materiales audiovisuales) cuando son muy reales, alcanzables y posibles.

Por su complejidad es necesario contar con profesionales capacitados para ocupar las distintas posiciones que requiere esta industria y desempeñar los diferentes roles comprendiendo su complementariedad (los que antes se veían como elementos completamente separados) aprovechando la sinergia que brinda el abordaje desde distintas disciplinas. Sin embargo, la especificidad de la tarea dificulta encontrar perfiles idóneos al mismo tiempo que su demanda es cada vez mayor por el fuerte crecimiento de la industria.

Este módulo fue diseñado como una guía inicial esclarecedora para mostrar qué caminos existen y analizar sus implicancias, determinar cuáles son los conocimientos básicos necesarios y ver qué opciones tienen los profesionales que quieren trabajar con big data de manera eficiente.

4.1.1 Principales caminos para dedicarse al big data

Trabajar con big data implica abordar las complejidades técnicas, procedimentales, estadísticas, matemáticas y de expertise de la disciplina referida a la cual se emprende cada proyecto.

Por ello, en materia de profesionales capacitados y eficientes para desempeñar estas labores es trascendental pensar en las competencias que una persona debe tener para trabajar en esta área.

En más de 10 años trabajando en este rubro y desarrollando proyectos para 23 países desde mi empresa Analytics Town, me he dado cuenta que el mismo patrón se repite en las necesidades que las empresas tienen a la hora de contratar profesionales que trabajen en sus equipos de BI, Big Data o Analytics.

Básicamente existen dos grandes ramas para las personas que quieren trabajar con big data, y que determinan el enfoque de carrera:

1. Infraestructura y arquitectura de soluciones tecnológicas.
2. *Analytics* y *data science*.

El enfoque en infraestructura y arquitectura está vinculado con las áreas de sistemas que montan un ecosistema capaz de responder a las necesidades de almacenamiento, disponibilidad, velocidad de procesamiento, hardware, software, administración de bases de datos y soporte técnico, entre otros. Su rol principalmente consiste en lograr que todo funcione.

El enfoque en analytics y data science tiene más que ver con el entendimiento de tendencias, patrones, variables de segmentación y formulación de soluciones pensadas para el negocio con base en las necesidades y oportunidades que de él emanan. Quienes se dedican a esta rama del big data deben extraer el conocimiento existente en los datos y conocer las técnicas de perfilado de usuarios, análisis del comportamiento y evaluación estadística de variables para determinar con el mayor nivel de certeza posible, conclusiones válidas, representativas y significantes.

Este material en su conjunto está mayormente dedicado a este grupo, los analistas y científicos de datos.

4.1.2 Enfoque en analytics y data science

A diferencia de otras carreras, trabajar con big data bajo el enfoque de analytics y data science requiere un compilado de habilidades y disciplinas múltiples como parte del perfil profesional.

El principal rol de estos profesionales es el de encontrar el conocimiento escondido entre los datos, el cual puede tener un uso práctico y aplicable en la generación de valor, y la detección de patrones y tendencias e incluso la predicción de escenarios que permitan apoyar la toma de decisiones. A su vez deben brindar una base sólida para soportar la complejidad de los datos y moldearlos de forma tal, que puedan ser consumibles por quien debe utilizarlos como insumo para sus actividades.

Los roles relacionados con analytics y data science requieren al menos estos tres conocimientos y habilidades:

- **Conocimiento estadístico y matemático:** para entender cómo manipular los datos y cómo hacer que todas las conclusiones sean válidas y representativas.

- **Conocimiento técnico y programático:** para hacer "ejecutable" el tipo de investigación o análisis que se desee realizar. Desde configurar un *dashboard* hasta programar un algoritmo, requiere un alto nivel de experticia por lo menos en parte del equipo de trabajo.

- **Conocimiento de la disciplina o negocio que se desea analizar:** para marcar el rumbo sobre qué buscar, para qué usarlo (conocimiento accionable) y en dónde se encontrará el retorno de inversión. Las capacidades de análisis estadístico o de programación de un algoritmo de predicción, clasificación o cualquier otro tipo, aportan elementos parciales y posiblemente desordenados si no se sabe qué buscar. Es necesario, además, el conocimiento de la disciplina en la que se está trabajando —y sobre la cual se desean generar los *insights* apropiados—para saber qué detalles específicos tiene este vertical para reconocer los patrones y tendencias, así como entender su significado y la forma como representan un fin concreto.

Un buen profesional del análisis de datos debe tener al menos dos de estos tres tipos de conocimiento para ser considerado un analista de calidad y su especialización le dará un valor diferencial en el mercado profesional, ya que quien esté a cargo del área —al conformar su equipo de trabajo— buscará perfiles complementarios que se apoyen y generen sinergia, que no solo enriquecerá la interacción y eficiencia del equipo sino también la calidad de los análisis que realicen.

Así mismo, deben tener las habilidades y capacidades necesarias para transmitir este conocimiento apropiadamente, de lo contrario, puede no ser adoptado. Cualquier proceso de análisis de datos que no se convierta en conocimiento accionable carece de retorno de inversión y por lo tanto de viabilidad.

4.1.3 Enfoque en arquitectura e infraestructura

El conocimiento asociado a los profesionales de arquitectura e infraestructura técnica del mundo big data, combina distintas habilidades para lograr que los datos sean manejables y útiles desde el punto de vista de su tamaño, la capacidad de procesarlos y almacenarlos y la velocidad con la que fluyen y se analizan, así como también la capacidad para analizar datos de distintos tipos y nivel de complejidad.

Por otro lado, debe conocer y manejar distintos aspectos relacionados a sistemas (IT), *software* y *hardware*, entre otros. Entre las competencias necesarias, se destacan:

- **Capacidad para garantizar una infraestructura física redundante:** Esto implica el manejo de servidores, *switches* y *data centers* que aseguren el desempeño correcto de esta infraestructura. Así mismo, debe garantizar la redundancia de los datos, esto es, que estén almacenados y disponibles en distintas ubicaciones y se repliquen en tiempo real para asegurar su integridad y disponibilidad.
- **Habilidades técnicas para trabajar con infraestructura:** Están relacionadas con el conocimiento de la performance de los equipos que gestiona (tanto en el data center como en la nube), su nivel de disponibilidad y escalabilidad, la flexibilidad que estos proporcionan en picos de demanda y los costos asociados (funcionamiento, gestión y ampliación o reducción de sus capacidades).
- **Conocimientos de seguridad Informática, encriptación y cifrado:** Están orientados a garantizar la seguridad y confidencialidad de los datos y la forma apropiada de comunicarse con las distintas fuentes de información, el software y las aplicaciones (segura, moderna, escalable y apropiada para la difusión ágil y veloz de información).
- **Detección de amenazas y aseguramiento de los recursos:** Estas habilidades tienen mucho que ver con la capacidad deductiva y de resolución de problemas, especialmente ante una situación que requiera una alerta o en una contingencia.

La combinación de estas capacidades, habilidades y conocimientos es una parte vital de la estabilidad, escalabilidad y capacidad de ejecución de proyectos de big data.

Conceptos clave

- Existen dos ramas principales para trabajar en big data, una enfocada en el análisis y otra enfocada en la infraestructura y los sistemas.
- Para enfocarse en el análisis de datos es necesario poseer habilidades y conocimientos relacionados con estadística y matemática, la utilización de herramientas y la programación, el dominio del negocio o vertical sobre el que se está generando la información y la capacidad de comunicar cada uno de estos conocimientos apropiadamente.
- El profesional que se siente más cómodo con la parte técnica de sistemas y desea garantizar la seguridad y el acceso a la información desde un punto de vista de infraestructura, hardware y software, debe tener conocimientos sobre infraestructura física redundante, la habilidad de manipulación de *storage* y bases de datos, y conocimientos de seguridad informática y detección de amenazas.

Referencias

The Economist. (2018). Científico de datos, la profesión más sexy del siglo XXI. Recuperado de http://www.eleconomista.es/ecoaula/noticias/8883560/01/18/Cientifico-de-datos-la-profesion-mas-sexy-del-siglo-XXI.html