

# Módulo 4. ¿Qué se necesita para trabajar Big Data?

## Unidad 1. Perfiles y *skills* necesarios en el mundo del 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 *El Economista* como el nuevo trabajo más sexy del siglo XXI. Para más detalles, puede hacer clic [aquí](#).

Con el correr de los años y la maduración en el campo de Big Data, fueron apareciendo nuevos perfiles y habilidades específicas con los que debemos contar a la hora de encarar un proyecto. Esto no quiere decir que necesitemos una persona de cada uno de los perfiles requeridos, sino que, dependiendo del estadio del proyecto y de la organización, serán necesarios nuevos roles para ir perfeccionando el proyecto.

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

El impacto que los datos genera 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, 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 y se muestran como altamente futuristas o de ciencia ficción (en películas y otros materiales audiovisuales), pero 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) y 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 Perfiles dentro del mundo de Big Data

Dentro del mundo de Big Data existen dos grandes ramas: la rama enfocada en la arquitectura e ingeniería y la rama dedicada a la explotación de los datos de manera analítica. Esta línea, que muchas veces se ve entrecruzada, nos sirve para separar en dos los perfiles necesarios para encarar un proyecto de Big Data.

Como comentamos anteriormente, a medida que el área se va profesionalizando, comienzan a aparecer nuevos perfiles y puestos más especializados. Por lo cual, si hace dos años un puesto era para un especialista, tal vez hoy ya se haya dividido y requiera tres perfiles distintos mucho más especializados.

Dentro de la rama de arquitectura e ingeniería, en el mundo del Big Data, podemos encontrar los siguientes perfiles: ingeniero de datos, *Data Ops* y arquitecto de datos.

Esta primera rama que vamos a detallar es clave respecto a la fundación de lo que luego explotaremos. Sentar una base sólida de arquitectura y disponibilidad de los datos es fundamental para etapas posteriores. A continuación, detallaremos las principales funciones de cada uno de estos perfiles.

- **Arquitecto de datos:** es el responsable de la definición de la estructura óptima para la recolección, almacenamiento y análisis de los datos. Los arquitectos de datos suelen ser perfiles *senior*, con más de 4 años de experiencia. Escuchan a sus clientes internos para poder definir la mejor infraestructura necesaria para cada uno de los proyectos. Suelen ser los responsables de las definiciones para construir soluciones *on-premise* o en la nube, construir un *data lake* o un *data warehouse*, además de elegir a los proveedores o herramientas para utilizar en cada caso, entre otras responsabilidades. Sus principales habilidades y conocimientos tienen que ver con la visión estratégica respecto de hacia dónde va la organización para poder diseñar la mejor estructura que sirva a futuro, también deben contar con fuertes conocimientos de las distintas nubes (AWS, Google, Microsoft, IBM, etc.) y de infraestructura *on-premise*.
- **Ingeniero de datos:** suele ser el responsable de llevar a cabo las definiciones que toma el arquitecto de datos. Estos ingenieros son los que realizan las extracciones de datos (de la definición y ejecución de los ETL), los almacenan y, en ciertas ocasiones, los analizan. Son responsables de todo el *pipeline* y el flujo de datos. Deben trabajar junto con el equipo que explota los datos para poder priorizar su *backlog*. Deben velar por la optimización del flujo de datos para que estos sean escalables en el tiempo y cada vez más eficientes. Suelen ser requisitos para esta posición tener fuertes conocimientos en bases de datos relacionales y no relacionales, ETL, SQL, Python, Spark, Hadoop, conocimientos de las principales herramientas y servicios de AWS, entre otros.
- **Data Ops:** esta es una nueva carrera dentro del mundo de Big Data; su objetivo principal es el de automatizar los *pipelines* y flujos que diseñan los ingenieros de datos. Al igual que estos, suelen ser perfiles que antes estaban relacionados con ejecutar ETL o mantener un *data warehouse* que se reconvierten en *data ops*. Se necesitan conocimientos técnicos muy fuertes en bases de datos, conocimientos

de herramientas o aplicaciones de algún proveedor de servicios en la nube (como AWS, por ejemplo), conocimientos en Spark, Hadoop, etc.

- **Big Data Specialist:** es una de las posiciones generalistas dentro del mundo del Big Data. Esta posición les sirve a las organizaciones que se están iniciando en el tema, ya que en una sola persona pueden tener a los tres perfiles que vimos con anterioridad. Obviamente, en un solo perfil no vamos a tener la especialidad y conocimientos que tienen los otros tres, pero, sin duda, nos podrá ayudar para sentar las bases de un proyecto de estas magnitudes.

Del otro lado, tenemos la rama analítica y de explotación de los datos. Estos perfiles vienen de una gran transformación en los últimos años. Podemos ver personas enfocadas en BI, en analítica comercial o inclusive personas con grandes conocimientos de negocio reconvertidas en estos nuevos perfiles. Para poder explotar y capitalizar todo el conocimiento de estos perfiles, es fundamental contar con una base sólida respecto a la arquitectura e infraestructura de Big Data, como lo detallamos previamente. Ya que, si no tenemos una buena base, los perfiles analíticos gastarán mucho tiempo en poder contar con los datos necesarios para cada uno de los análisis y modelos que el negocio les requiere. A continuación, detallaremos los principales perfiles que encontramos de este lado del Big Data.

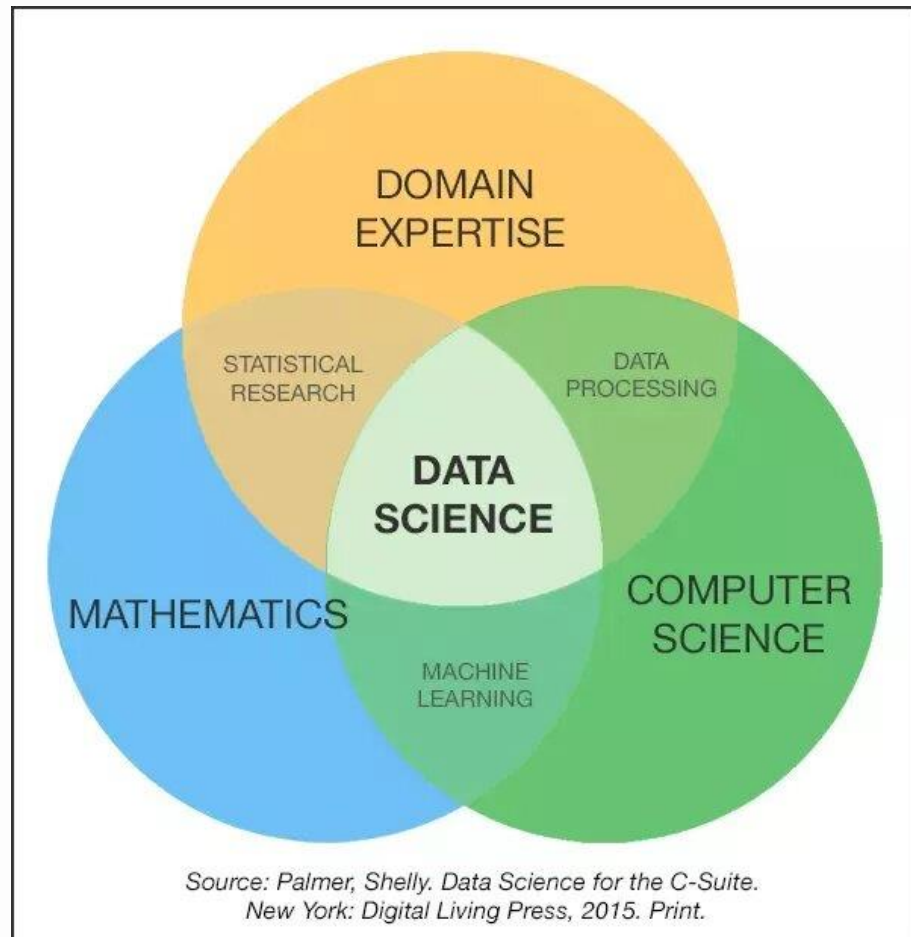
- **Data Visualization:** estos perfiles son los más similares a los conocidos dentro del mundo del BI. Su principal responsabilidad es la construcción de tableros para monitorear alertas o tendencias. Requieren de grandes conocimientos de manipulación de datos, de herramientas de visualización (como Power BI y Tableau, entre otros) y un poco de entendimiento del negocio. Su principal *skill* está relacionada con poder contar historias con los datos, con poder hacerlos entendibles y convertirlos en información importante para la toma de decisiones. Dentro de este campo, viene creciendo el *storytelling*, una rama muy enfocada en mostrar historias con los datos como si fuera un cuento. Recomiendo la lectura del libro *Storytelling with Data: a data visualization guide for business professionals*, de Cole Nussbaumer Knaffic.
- **Data Analyst:** estos perfiles vienen a potenciar y complementar el anterior. Se les suele requerir los mismos conocimientos que a un *Data Visualization*, pero se agrega la capacidad de generar *insights* que ayuden a la toma de decisiones.

Son especialistas en *data mining* y manipulación de grandes volúmenes de datos. Su objetivo es responder preguntas con datos, alertar sobre cambios en comportamientos y tendencias, realizar simulaciones de escenarios posibles, etc. Sus principales *skills* suelen ser las capacidades analíticas, manejo de SQL y conocimiento o entendimiento del negocio.

- **Data Scientist:** al igual que el anterior, este rol viene a potenciar los dos anteriores. A este perfil se le suelen pedir las habilidades y conocimientos de los *Data Visualization* y de los *Data Analyst*. Además de todo eso, son perfiles con tres grandes áreas de conocimiento: estadístico/matemático, programación y negocio. Al tener tres pilares tan importantes y diversos como estos, encontramos en el mercado muchos perfiles de distintas formaciones. Podemos encontrar físicos, matemáticos o estadísticos, también ingenieros en sistemas, licenciados en computación, ingenieros en *software* o cualquier carrera afín a la programación, como también podemos encontrar profesionales de la rama de las ciencias económicas, sociales, o cualquiera sea la carrera afín a la materia que se esté trabajando. Por ejemplo, en la rama de la inteligencia artificial, dentro del mundo de la medicina, es común encontrar biomédicos en esta posición.

Su principal función suele ser la de realizar modelos de *machine learning*, *deep learning*, o cualquier otra rama dentro de la inteligencia artificial. Deben tener conocimientos estadísticos, de programación y de entendimiento del negocio. Suelen programar en R o Python y conocer las distintas librerías de *machine learning*, como Sikit-learn, TensorFlow, Keras, y un sinfín más de librerías. Sin duda, *Data Scientist* es una de las carreras más demandadas dentro del mundo del Big Data de los últimos 2 años y, probablemente, por los próximos 2-5 años. Como comenté antes, este perfil debería dominar las tres áreas que detallaba: conocimientos estadísticos, conocimientos de programación y conocimientos del dominio (o negocio). En la realidad, suele ser difícil encontrar una persona que domine completamente estas tres áreas. Por lo general se encuentran, pero solo con el conocimiento de dos de estos tres dominios. A las personas que cumplen con las tres áreas se las conoce como *Data Scientist Unicorn*, como una paradoja de lo imposible de hallarlas.

**Figura 1: Ciencia de datos**



Fuente: Hussain, Haddad y Lester, 2018, p. 16.

- **Machine Learning Engineer:** este es un perfil relativamente nuevo dentro del mundo del Big Data, ya que es muy especializado en la parte de implementación y optimización de modelos operativos. En un contexto en el cual crecen los modelos robustos de inteligencia artificial, con predicciones *real time*, es cada vez más importante la *performance* del modelo. Esto quiere decir que la orquestación de todos los procesos que componen una predicción o recomendación de un modelo debe funcionar de la manera más rápida y óptima posible. Para ello, es necesario conocer muy bien cómo diseñar ese flujo

de datos, por lo que su principal responsabilidad pasa por implementar los modelos definidos por el *Data Scientist* de la mejor forma posible. Esta relación con el *Data Scientist* suele ser de ida y vuelta, ya que, una vez que este ha definido el mejor modelo posible en términos de la calidad de la predicción, puede que no sea el mejor modelo para la implementación. En ese momento se empieza a ajustar el modelo para llegar al óptimo en predicción y tiempos de respuesta.

- **Data Governance:** este perfil se encuentra dentro de la rama del Big Data, pero lo podríamos poner en cualquiera de los dos lados (dentro de la rama de arquitectura o dentro de la rama analítica). Tiene como objetivo gobernar los datos que son utilizados. Velan por la disponibilidad y calidad de los datos que se utilizan. Se suelen encontrar distintos perfiles según la empresa. Hay organizaciones que valoran más el conocimiento de los datos que los conocimientos técnicos, otras que velan por tener perfiles con habilidades de liderazgo y gestión para poder hacer frente a los distintos desafíos del puesto.

Como podemos ver en el cuadro siguiente, la rama de carreras relacionadas con Data/AI y la rama relacionada con Engineering y Cloud Computing serán las ramas con mayor crecimiento para 2022, según el reciente reporte sobre los trabajos del futuro del Foro Económico Mundial. Y año a año hemos visto cómo todas estas carreras han ido en aumento y especializándose cada vez más. Es de esperar que en los próximos años continúen apareciendo nuevas carreras más enfocadas que las que hoy conocemos. Pero las necesidades de las organizaciones van a seguir dependiendo del grado de madurez en la materia. Probablemente, el rol del *Data Scientist* va a pasar a ser un “generalista” dentro del mundo de *machine learning* e inteligencia artificial, y luego, dependiendo de la madurez, se requerirán perfiles especializados según cada caso.

## Figura 2: Surgimiento de las profesiones del futuro

Figure 1: Emergence of clusters of professions of the future, 2020-2022

Professional Cluster	Number of opportunities (per 10,000)	
	2020	2022
figures extrapolated from data for 20 economies (LinkedIn)		
Data and AI	78	123
Engineering and Cloud Computing	60	91
People and Culture	47	58
Product Development	32	44
Sales, Marketing and Content	87	125

Fuente: "Mapping Emerging Occupations", 2020, p. 8.

### 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 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 debe enfocarse en conocimientos de ingesta de datos, disponibilidad de datos y algunos conocimientos de seguridad informática. Recomiendo comenzar con capacitaciones orientadas al manejo y conocimiento de algún proveedor de servicios *online* (como AWS).

# Referencias

**Hussain, N. Haddad, F. y Lester, R.** (2018). *Data Science & Deep Learning ABC...*  
Recuperado de [https://meetnavpk.github.io/resources/ebook/pdf/Data\\_Science\\_&\\_Deep\\_Learning\\_ABC.pdf](https://meetnavpk.github.io/resources/ebook/pdf/Data_Science_&_Deep_Learning_ABC.pdf)

**Mapping Emerging Occupations.** (2020). *Jobs of Tomorrow. Mapping Opportunity in the New Economy.* Recuperado de [http://www3.weforum.org/docs/WEF\\_Jobs\\_of\\_Tomorrow\\_2020.pdf](http://www3.weforum.org/docs/WEF_Jobs_of_Tomorrow_2020.pdf)