

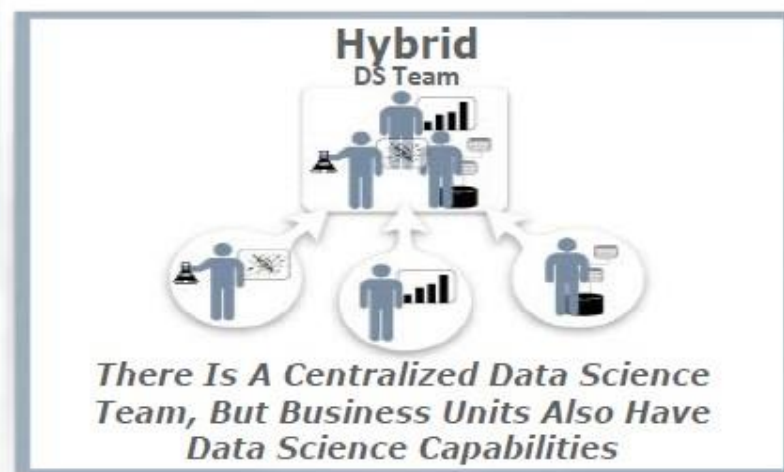
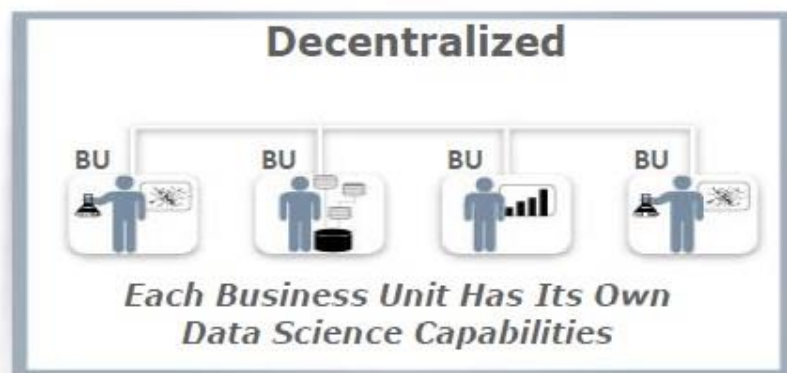
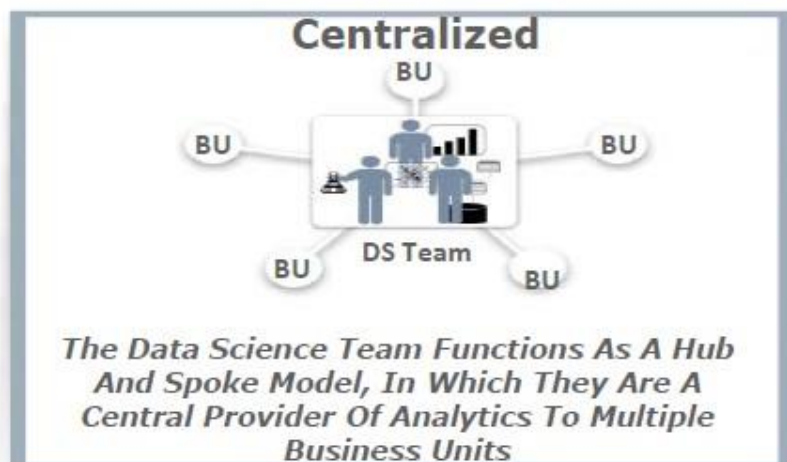
Sois los máximos responsables de vuestra “Compañía” y queréis poner en marcha un grupo de “Data Scientist” en vuestra organización. ¿Dónde le situaríais en el organigrama? ¿Departamento propio? ¿Dependiendo de otro grupo ya existente? ¿Repartidos en varios departamentos?...

Dependiendo de las necesidades de la empresa y de cómo esté estructurada, existen tres tipos diferentes de modelos operativos:

- El enfoque descentralizado: con este enfoque las organizaciones tienen científicos de datos que informan a diferentes funciones o unidades de negocio en toda la empresa. Esta organización surge comúnmente en organizaciones más grandes donde las iniciativas de ciencia de datos han surgido orgánicamente en varias partes del negocio. El beneficio de esta estructura es que los analistas son expertos en conocimiento y están completamente inmersos en las actividades cotidianas de sus departamentos. Sin embargo, debido a que están diseminados por toda la compañía, a menudo hay muy poca comunicación entre ellos, y como resultado, hay una falta de consistencia (especialmente en lo que respecta a métodos y definiciones) en todos los ámbitos. Asimismo, la movilidad de los científicos de datos es más restringida en una organización descentralizada, lo que a menudo lleva a nichos de conocimiento, a menos oportunidades de intercambio de conocimientos o a oportunidades limitadas de crecimiento profesional
- El enfoque centralizado: utilizando este método, hay un grupo central de analistas que informan desde un único departamento. La ventaja de este modelo es que los analistas pueden aprovechar sus diferentes conjuntos de habilidades, realizar entrenamientos y emplear los niveles más altos de coherencia en toda la organización. Estructuralmente, la centralización también simplifica la contratación y el reclutamiento, crea una agencia para impulsar iniciativas analíticas en toda la empresa y reduce los nichos de conocimiento. La desventaja es que este grupo a menudo está aislado de otros departamentos y no está afianzado en lo que está sucediendo día a día. En algunos casos, esto puede conducir a una dinámica no deseable donde la ciencia de datos se trata como una función de soporte en lugar de ser proactivos
- El enfoque híbrido que combina elementos de las dos estrategias anteriores. En general, los científicos de datos informan de forma centralizada, ya que reclutar y retener el talento generalmente es el principal obstáculo para formar un equipo de ciencia de datos en las primeras etapas. Esto ayuda a las organizaciones a analizar de manera eficiente los datos de una manera consistente. Y para el éxito, se posiciona a los científicos de datos para trabajar estrechamente con las unidades de negocios, una práctica conocida como *incrustación (embedding)*. En equipos pequeños de ciencia de datos, la incrustación puede tener el efecto secundario de dejar a los científicos de datos demasiado aislados. Una solución aplicada (Coursera) consistiría en crear pequeños subgrupos (clústeres) que se asociarían con diferentes unidades de negocio, interactuando con otros clústeres también

Organizational Models for Data Science Teams

Organizational Model	Centralized	Decentralized	Hybrid
Executive Sponsorship	Required	Required	Required
Business Unit Capabilities	None	High	Medium
DS Team Capabilities	High	None	High



Regardless Of Which Approach, They All Need Executive Sponsorship To Succeed

Parece entonces que la mejor manera de crear un equipo de Data Science es esta última, la híbrida, pero ese es el paso final al que hay que llegar. Primero debemos preguntarnos desde donde partimos. Suponemos que partimos de cero. Las preguntas que nos debemos hacer son:

- ¿Con qué datos trabajarán sus equipos?
- ¿Qué quieres hacer con esta información?
- ¿Quién se va a beneficiar con el valor generado de estos datos? directamente (el usuario o consumidor), Indirectamente (en los resultados de la organización)
- ¿Cuánto tiempo y dinero puede invertir en este proyecto?
- ¿Qué retorno de inversión se espera?

Asimismo, hay que tener en cuenta las expectativas de futuro que se pueden resumir en:

- **Escalabilidad:** Intentar anticipar cuánto aumentarán sus datos y qué nuevos procesos habrá que configurar, o que nuevos datos de productos se construirán
- **Accesibilidad y estabilidad:** Siempre hay que tener un compromiso entre accesibilidad y estabilidad. Por ejemplo, si se desea que los datos estén extremadamente disponibles para realizar el aprendizaje automático, hay que aceptar cierta pérdida de datos y precisión
- **Herramientas:** la escalabilidad es importante en este caso también. Se desea una herramienta que funcione con las diferentes tecnologías que estén disponibles ahora o en el futuro. Implementar una herramienta que ayude a diferentes tecnologías a trabajar juntas también ayudará a que diferentes personas y perfiles de datos puedan trabajar juntos

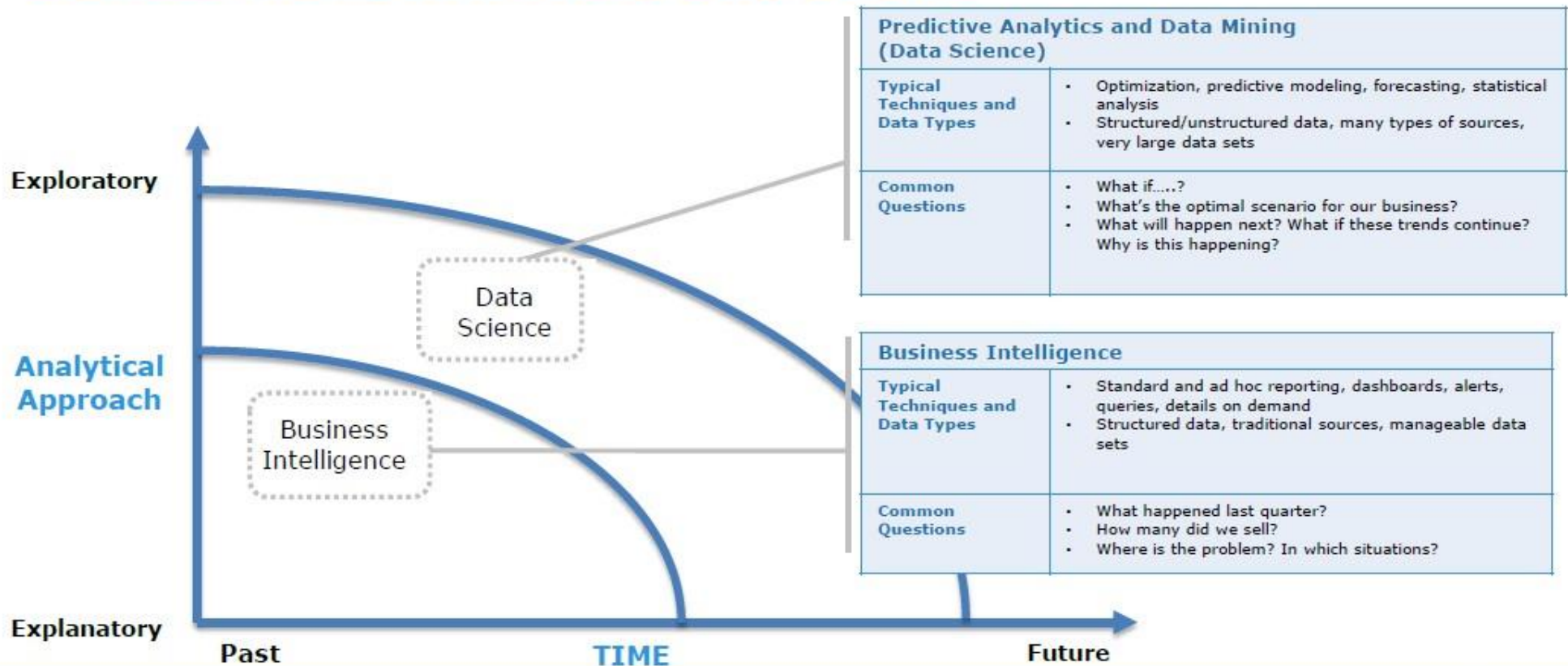
Parece razonable pensar que sobre todo al principio del establecimiento de un equipo de Data Science debemos adoptar la estructura centralizada. Establecer unos plazos y unos objetivos generales de preparación: identificación de datos y herramientas, definición de procesos y procedimientos generales, reclutamiento de personal,

En un segundo momento tener una estructura más distribuida entre los distintos departamentos para así tener un conocimiento más adecuado de la necesidades y problemas de cada parte de la empresa. Pero manteniendo un núcleo central fuerte donde recoger toda la información generada en cada grupo y homogeneizar los objetivos y herramientas necesarios. Estos han de ser lo suficientemente flexibles para que sean válidos para todos los grupos, pero al mismo tiempo el máximo estandarizado posible a nivel global de la empresa. Un equilibrio delicado.

En un tercer momento y dependiendo mucho de: el nivel de penetración de la cultura del dato en la empresa, el nivel de madurez de los equipos, el nivel de importancia de los datos en los productos/servicios de la empresa, ... podríamos llegar a un nivel mayor de descentralización.

Big Data Requires New Approaches to Analytics

Business Intelligence Versus Data Science



¿Cuál sería vuestra propuesta para una gran empresa?

¿y para una mediana? ¿y para una pequeña?

En el caso de las grandes empresas, no hay excusa para dejar de lado el Data Science

El análisis de datos, tanto en tiempo real como en la anticipación de tendencias, puede aportarnos información realmente útil. Otro espacio en el que el Data Science es clave es el diseño de estrategias de productos y servicios.

Pero incluso en aquellas empresas que podríamos llamar más tradicionales, o en las que parece que el Big data no podría tener ninguna repercusión, siempre generan datos que son valiosos, como todos los que tratan más de los aspectos operativos y contextuales de la organización. Estos que en el cuadro anterior están marcados en la zona Business Intelligence que pueden explicar lo que ocurrió en el pasado y sacar conclusiones.

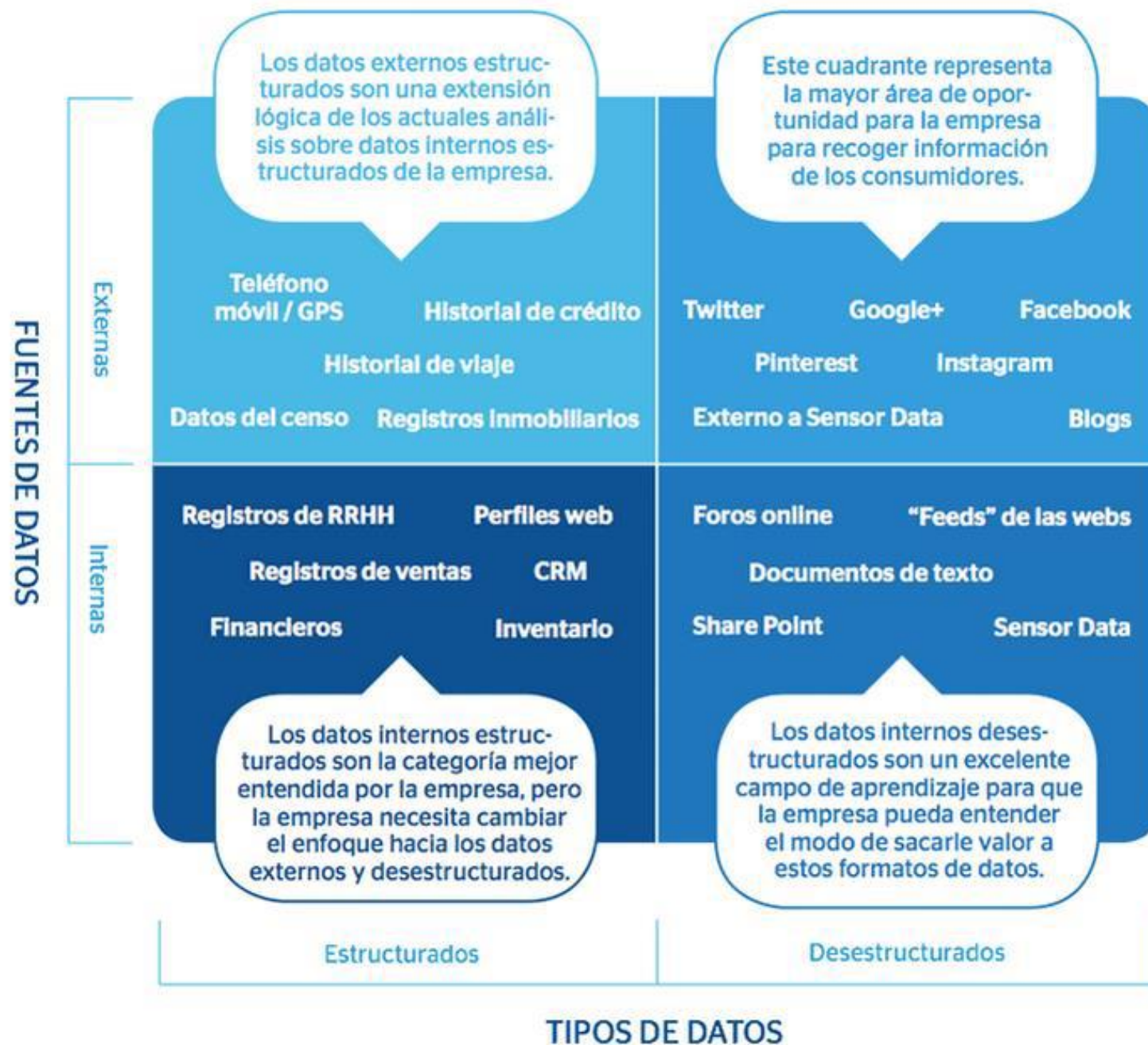
Y que sirven para establecer estadísticas, cuadro de mando, informes, alertas, consultas, etc., pero que requieren de una infraestructura que al menos debe estar formada por un equipo de almacenamiento y otro de análisis y/o inteligencia empresarial (BI).

Las tecnologías Big Data pueden hacer todo lo que un producto de almacenamiento de datos puede hacer y mucho más. El conjunto de habilidades y conocimientos requerido es muy superior, y redundará en beneficio de la empresa

Se cree que se requieren muchos datos para utilizar estas tecnologías. Ese no es el caso en absoluto. El volumen de datos es el objetivo, sí, pero no necesita millones de clientes para extraer información.

IBM define ciencia de datos como compuesta por cuatro "V" esenciales: Volumen, Velocidad, Variedad y Veracidad. Si puede estructurar sus datos en una de estas categorías o conceptos, son valiosos. La veracidad de los datos actuales, la gran variedad y la gran velocidad de los datos entrantes marcan la diferencia, además del volumen.

La inteligencia empresarial y la ciencia de datos a menudo se confunden, especialmente por aquellos que no están familiarizados con la industria. Sin embargo, no son sinónimos. La inteligencia empresarial implica datos, sí, pero se trata más de los aspectos operativos de la empresa. Por otro lado, la ciencia de los datos tiene más que ver con el análisis predictivo. El objetivo es recopilar suficiente información que pueda usar para generar patrones y perspectivas discernibles. Además, la ciencia de datos puede responder lo que sucederá si cambia varios aspectos de un proceso o plan de negocios.



Para las pequeñas y medianas empresas, en mi opinión, los criterios fundamentales a tener en cuenta a la hora de decidir si construir un equipo propio de Data Science son dos:

1. Que datos necesitamos usar para nuestro negocio

- 1.1. Asumimos que todas las empresas están situadas como mínimo dentro del primer cuadrante del cuadro anterior: datos internos estructurados. Eso requiere al menos aplicaciones de inteligencia empresarial sobre todo data warehouse.
- 1.2. Si además también tenemos que trabajar con datos situados en los cuadrantes interno/desestructurados y/o externo/estructurado es plausible pensar en adecuar el organigrama de la empresa a una interacción con un equipo de Data Science
- 1.3. Si nuestro negocio está orientado claramente hacia productos y/o servicios cuyos datos podríamos encuadrar en el cuadrante externo/desestructurados, debemos preparar una estrategia tendente al establecimiento de un equipo de Data Science en la empresa.

En los dos primeros casos el tamaño de la empresa, y sobre todo las perspectivas de crecimiento (punto 2) nos darán una orientación si se gestionan los datos de manera externa, generalmente por consultorías, o si se implementa a nivel interno, y hasta qué punto.

2. Perspectivas de crecimiento de la empresa

Ventajas e inconvenientes de la externalización

- Una gestión externa de los datos ofrece a las organizaciones pequeñas y medianas una alternativa para realizar análisis de negocios, “just in time”, en lugar de tener que construir una infraestructura propia con el coste que requiere
- El problema con los analistas externos es que no tienen el conocimiento profundo ni los detalles sobre los clientes que pueden tener sus equipos internos. Estos tienen un acceso mucho más amplio y una base de conocimiento más rica con la que trabajar.
- Con los informes de las consultoras, los empresarios pueden ponerse al día de las tendencias de distintos sectores. Los datos ya han sido recogidos, filtrados y analizados por otros, y aparecen en sencillos gráficos auto explicativos. Además, suele ser contenido gratuito. Sin embargo, es poco profundo, con más conclusiones de encuestas que Data Science como tal. Si buscamos información de último minuto de nuestro modelo de negocio, sector o micro nicho, tendremos que recurrir a revistas especializadas, y estas suelen ser costosas

Bibliografía

Data Engineering Teams. Creating Successful Big Data Teams and Products Jesse Anderson

<https://insidebigdata.com/2017/12/28/5-misconceptions-data-science/>

<https://insidebigdata.com/2016/11/03/how-to-grow-your-data-science-team/>

<https://hablemosdeempresas.com/empresa/data-science-ciencia-dato-negocio/>

<https://www.cio.com/article/3011648/analytics/dont-look-for-unicorns-build-a-data-science-team.html>

<https://es.slideshare.net/emcacademics/building-data-science-teams-31057129>

<https://www.linkedin.com/pulse/5-steps-build-data-science-team-christopher-doyle/>

<https://towardsdatascience.com/what-is-the-most-effective-way-to-structure-a-data-science-team-498041b88dae>

Lecturas complementarias

<https://www.oreilly.com/ideas/data-engineers-vs-data-scientists>

<https://www.slideshare.net/0xdata/intro-to-data-science-for-nondata-scientists>

<https://venturebeat.com/2015/01/25/so-you-want-to-build-a-data-science-team/>