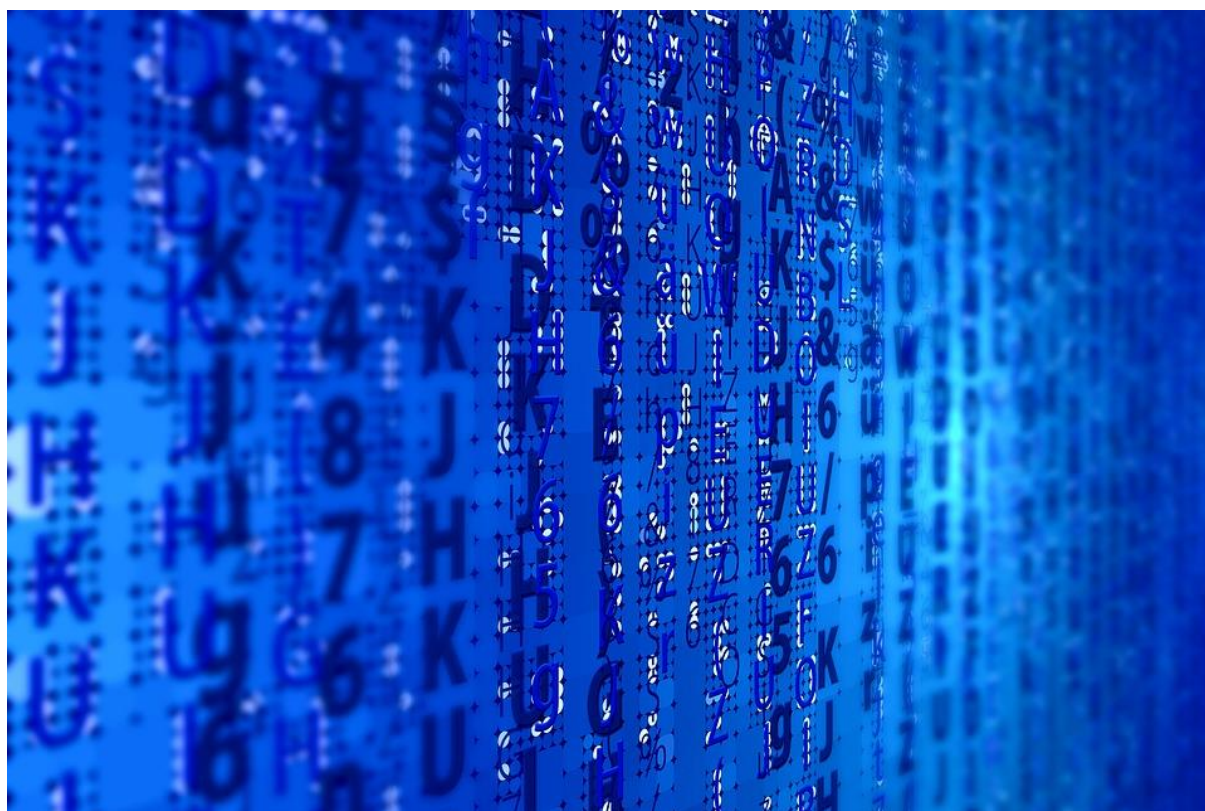


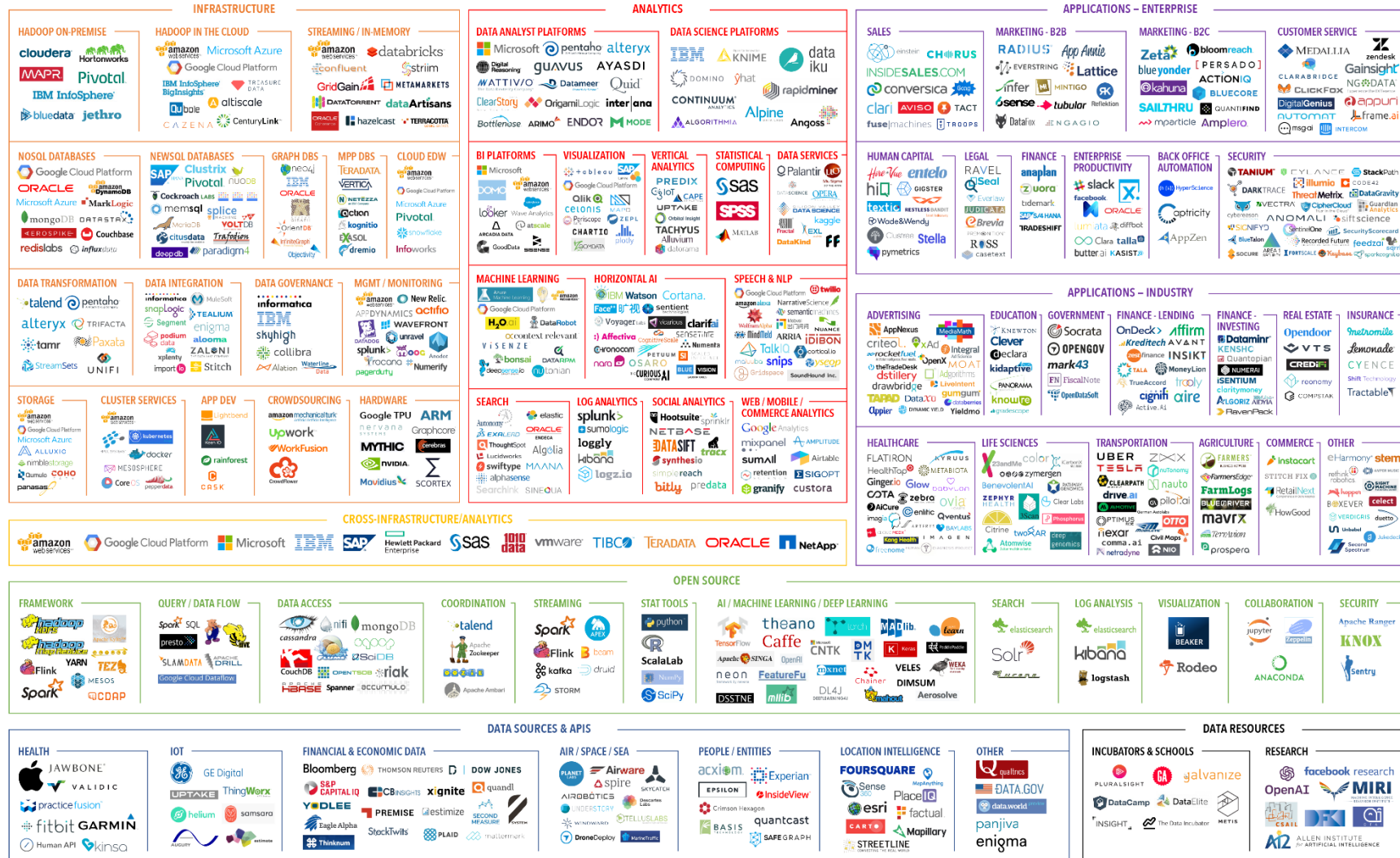
Panorama general de tecnologías Big Data

La intención de la lectura es reflejar al participante del curso la amplia variedad de sistemas que podrían formar parte en un proyecto de Big Data. La lectura no pretende exponer cada una de las tecnologías existentes, sino exponer -por grupos funcionales- la división de este mercado de tecnologías y resaltar algunos de los actores más relevantes del espacio



La siguiente imagen (ver siguiente página) condensa el *Big Data Landscape* (en español; “Panorama general de Big Data”). Se trata de una recopilación de herramientas populares usadas en distintos proyectos de analítica, basada en datos del mercado de las tecnologías de la información. Esta versión incluye más de 900 herramientas/proveedores. La publicación puede ser de utilidad por la amplia visión de la oferta actual en el sector.

BIG DATA LANDSCAPE 2017



El escenario habitual que presenta cada nueva versión de este *Big Data Landscape* es el nacimiento permanente de *Startups*, lo que está a la par de su constante absorción por grandes compañías. La última versión presenta a los gigantes tecnológicos con sus soluciones consolidadas de Big Data, que incluyen infraestructura (por ejemplo, herramientas para almacenar y procesar los datos) y componentes de analítica (por ejemplo, herramientas para la visualización de los datos).

La gran cantidad de herramientas existente responde al amplio abanico de posibilidades de participación que existe para ellas en proyectos de analítica de datos: en un proyecto pueden necesitarse herramientas que produzcan los datos, herramientas que soporten y propicien el almacenamiento de datos, aquellas que utilizan una técnica de analítica especial para obtener determinados resultados y otras que sirvan para la visualización y tratamiento de los datos. Entonces, por cada forma de participación, se puede encontrar una oferta variada y elevada.

Según el análisis de Matt Turck, director e inversionista de capital riesgo en la compañía FirstMark Capital, en este momento la democratización de la tecnología (acceso de las personas a la misma) ya se considera un hecho consolidado. Además, se puede afirmar que esta democratización se dirige hacia una oferta conjunta de “Big Data + AI (inteligencia artificial)”, auspiciada por la adopción de estas tecnologías por grandes compañías.

Como hemos comentado en varias ocasiones, el número de herramientas existentes en el entorno Big Data es año tras año superior. Como referencia comparativa para entender la magnitud del incremento de agentes del mercado, incluimos en esta lectura la versión 2012 del *Big Data Landscape*. Puedes verla en la siguiente página. En comparación a la de 2017, te dará cuenta de que es mucho más pequeña.

Big Data Landscape (Version 2.0)



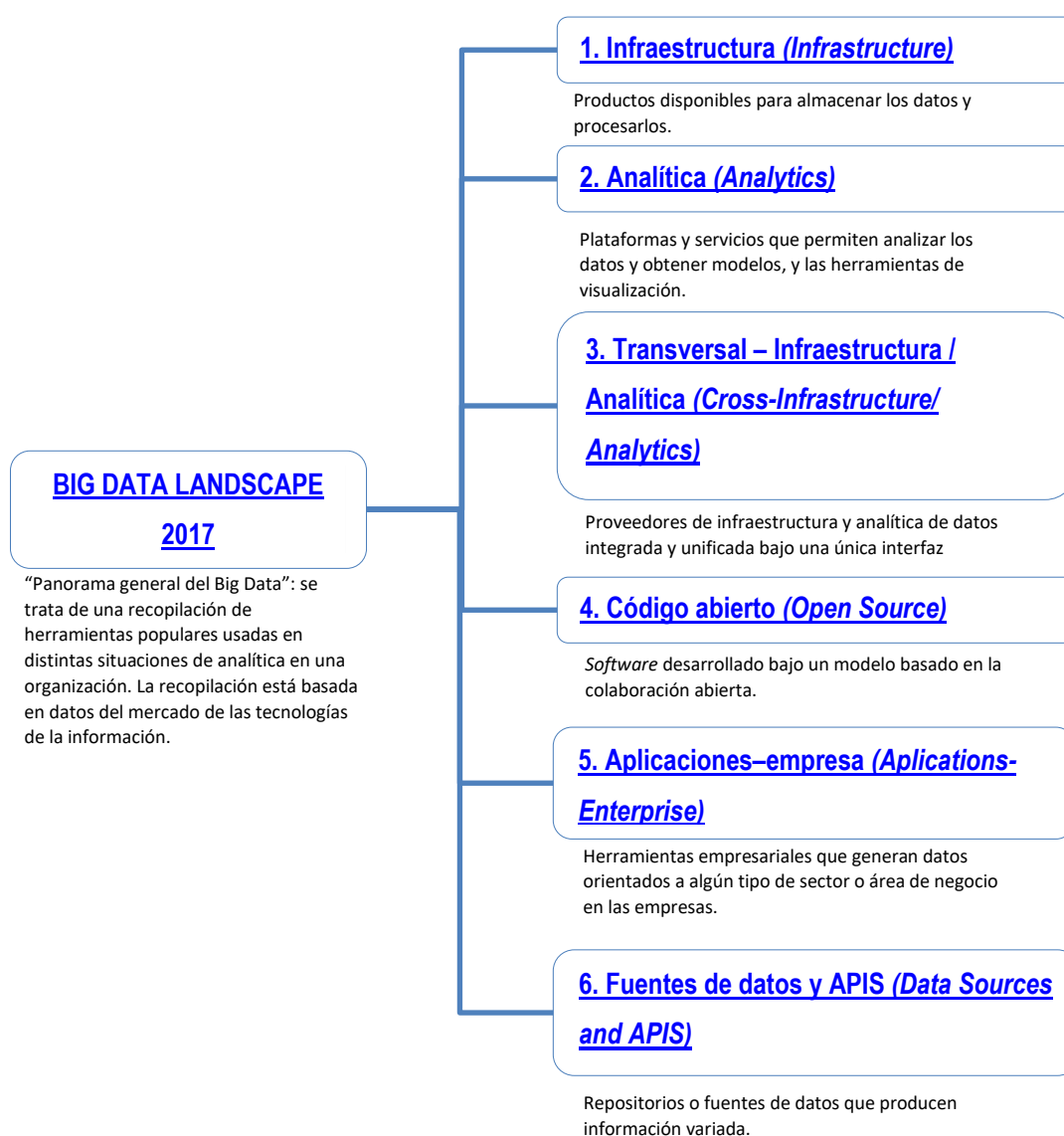
© Matt Turck (@mattturck) and Shimon Zilis (@shivonz) Bloomberg Ventures

En los siguientes subtítulos introduciremos a cada una de las agrupaciones del *Big Data Landscape 2017*.

1. Mapa de la lectura

NOTA: haz clic sobre los conceptos en el mapa para navegar la lectura. Puedes volver a este mapa cuando quieras, haciendo clic sobre el texto “Volver al mapa (clic aquí)”, ubicado al final de cada página del documento.

Además, recuerda que, si no conoces alguno de los conceptos mencionados en esta lectura, la mayoría están presentes en el Glosario del curso. Te recomendamos tenerlo a mano para facilitar la comprensión de esta lectura.



2. Infraestructura (Infrastructure)



Esta agrupación de herramientas se centra en los productos disponibles para almacenar los datos y procesarlos.

Podemos destacar los productos que proveen plataformas de Big Data, como por ejemplo *Hadoop*, un ecosistema de proyectos que trabajan para ofrecer un servicio coherente de tecnologías relacionadas con Big Data, y que puede desplegarse en una arquitectura basada en los servidores propios de una compañía (*on-premise*) o en la nube (*cloud*).

Además, se incluyen proveedores de bases de datos populares en sus distintas variantes, por ejemplo, *NoSQL* para almacenar datos semiestructurados (entre otros tipos de datos) y las *Graph dbs* indicadas para registrar relaciones entre entidades.

En esta agrupación también se han recogido las herramientas de transformación e integración que posibilitan una mejor gestión y administración de la información durante el desarrollo de los proyectos.

3. Analítica (Analytics)



En este nuevo grupo se exponen principalmente las plataformas y servicios que permiten analizar los datos y obtener modelos analíticos (por ejemplo, un algoritmo

para predecir las ventas de una compañía), reservando espacio para algunos con capacidades de visualización.

Se incluyen herramientas de análisis tradicional y *data mining*, cuya contraparte en el mundo del Big Data son las ubicadas en el subgrupo *Data Science Platforms*.

Además, se especifican distintos ámbitos de análisis de datos: desde servicios de inteligencia artificial, *machine learning* y *deep learning* generales; hasta aquéllos con propósito específico, centrados en interpretación del lenguaje o destinados a obtener resultados a partir de las redes sociales.

En este ámbito funcional de aplicaciones, se engloban las herramientas de visualización, es decir, aquellas que ejecutan la búsqueda, interpretación, contrastación y comparación de datos para transformarla en información comprensible para el usuario. Por ejemplo, herramientas que permiten transformar la información en gráficos.

4. Transversales – Infraestructura/analítica (*Cross-Infrastructure/ Analytics*)

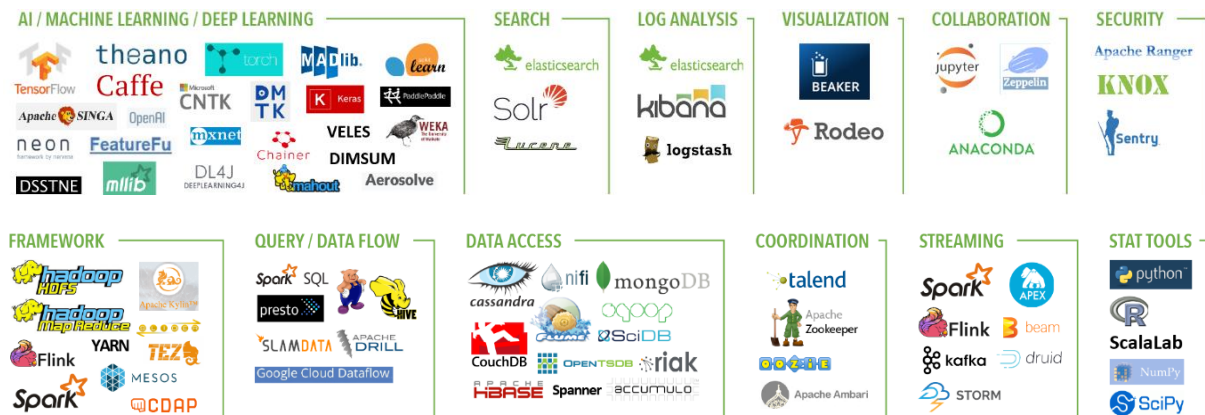
CROSS-INFRASTRUCTURE/ANALYTICS



El *Big Data Landscape* también reserva un espacio para destacar los proveedores que ofrecen las capacidades de infraestructura y analítica de datos de forma integrada y unificada bajo una única interfaz.

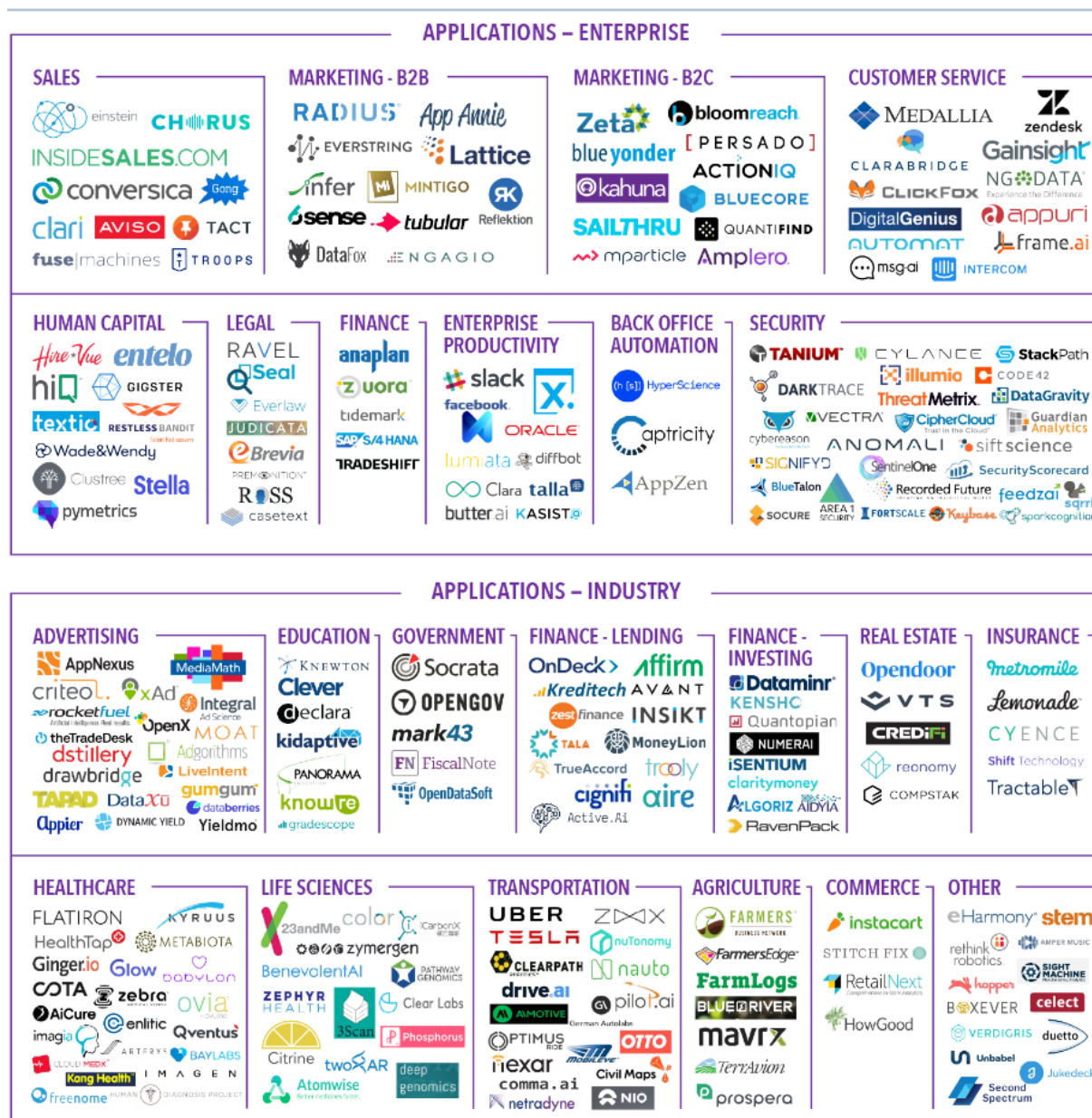
5. Código abierto (*Open Source*)

OPEN SOURCE



Bajo el término *open source* se engloba cualquier *software* (programa) desarrollado bajo un modelo basado en la colaboración abierta, es decir, que permite el acceso de cualquier persona diferente a los creadores al código de programación para modificarlo sin restricciones de licencia. Es como si una persona escribiera un libro y, en vez de publicar una versión definitiva y cerrada que nadie puede modificar, publicara una versión abierta para que todo el mundo pudiera editarla y mejorarla.

6. Aplicaciones–empresa (Applications-Enterprise)



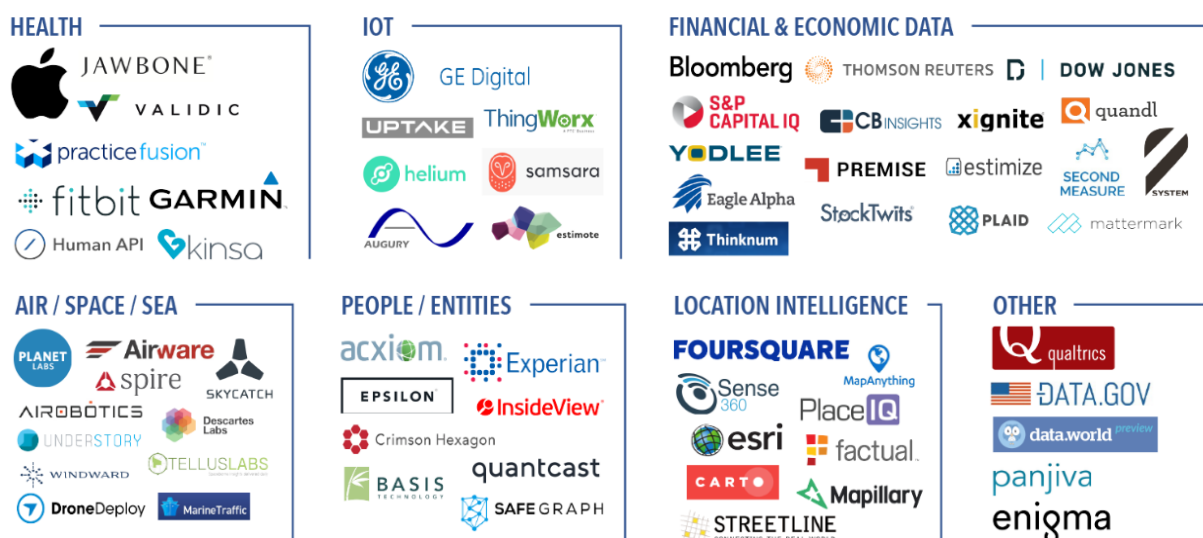
Este grupo presenta herramientas empresariales que generan datos orientados a algún tipo de sector o área de negocio específico en las empresas.

El interés por conocer a las herramientas que integran este grupo radica en poder enlazar alguna necesidad de información específica (ventas, seguridad, *marketing*, servicio al cliente, etc.) con algún proveedor indicado en el grupo, de forma que deba contemplarse su participación en esa área o en algún proyecto específico que esa

área quiera realizar. Por ejemplo, el *software* Socrata, especializado en el ámbito del gobierno, permite almacenar y gestionar todo tipo de información correspondiente al sector público, para -por ejemplo- realizar un control de los datos de una ciudad para mejorar la toma de decisiones y actuar más rápidamente.

7. Fuentes de datos y APIS (*Data Sources and APIS*)

DATA SOURCES & APIS



En el grupo anterior se consideraban herramientas que se pueden usar como fuentes de información en el contexto de un proyecto (por ejemplo, datos extraídos de un sistema transaccional), pero también deben contemplarse repositorios o fuentes de datos que producen información variada, y pueden ser de utilidad a la hora de poner en marcha una iniciativa de Big Data. Estas fuentes también pueden proporcionar información específica de un campo, o información de carácter general, y pueden ser de acceso público o disponibles tras su adquisición.

Por ejemplo, los índices Dow Jones reflejan el comportamiento del precio de las acciones de compañías importantes en ciertos sectores y es información que está siempre disponible, es decir, los índices existen independientemente de las necesidades de una organización o de un proyecto específico dentro de ella. Y así

como estos índices, existen muchas otras herramientas que ofrecen información especializada y siempre disponible; a las que las organizaciones pueden acceder; ya sea porque los datos son públicos, o bien, de pago. Esta información puede ayudar a enriquecer los datos que utiliza una empresa en su toma de decisiones.

Conclusiones

Como has podido contemplar a lo largo de la lectura, el ecosistema de aplicaciones y herramientas que pueden participar en los proyectos de Big Data es muy amplio y sigue en constante evolución. Si bien es imposible tener un conocimiento permanentemente actualizado en cada herramienta, es interesante tener un conocimiento general sobre las posibilidades existentes; para complementar al máximo los proyectos de analítica de datos en Big Data y obtener mejores resultados.

Referencias

Turck, M., (2017). *Firing on All Cylinders: The 2017 Big Data Landscape*. Recuperado de:

<http://mattturck.com/bigdata2017/>



Esta obra está sujeta a la Licencia Reconocimiento-NoComercial-SinObraDerivada 3.0 España de Creative Commons. Para ver una copia de esta licencia, visite <http://creativecommons.org/licenses/by-nc-nd/3.0/es/> o envíe una carta Creative Commons, PO Box 1866, Mountain View, CA 94042, USA.