

# Bigdata Conceptos, definiciones y aplicaciones

**Hugo Andrés Dorado**

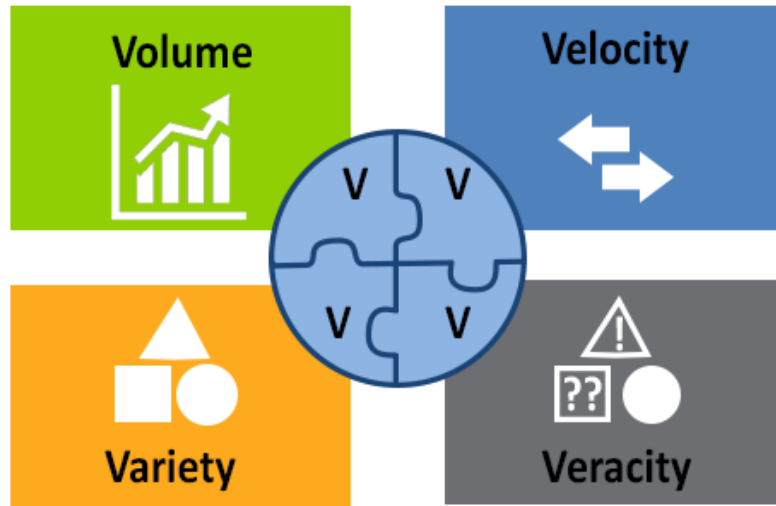
Científico de datos

[hugo.doradob@gmail.com](mailto:hugo.doradob@gmail.com)

# Bigdata – Grandes volúmenes de datos

Una tendencia a almacenar, vincular, procesar y analizar grandes cantidad de información con el propósito de tomar decisiones.

# Las cuatro Vs del bigdata



# V - Volumen

$$n \rightarrow \infty$$

Almacenamiento

Pentabytes or exabytes

Complejidad



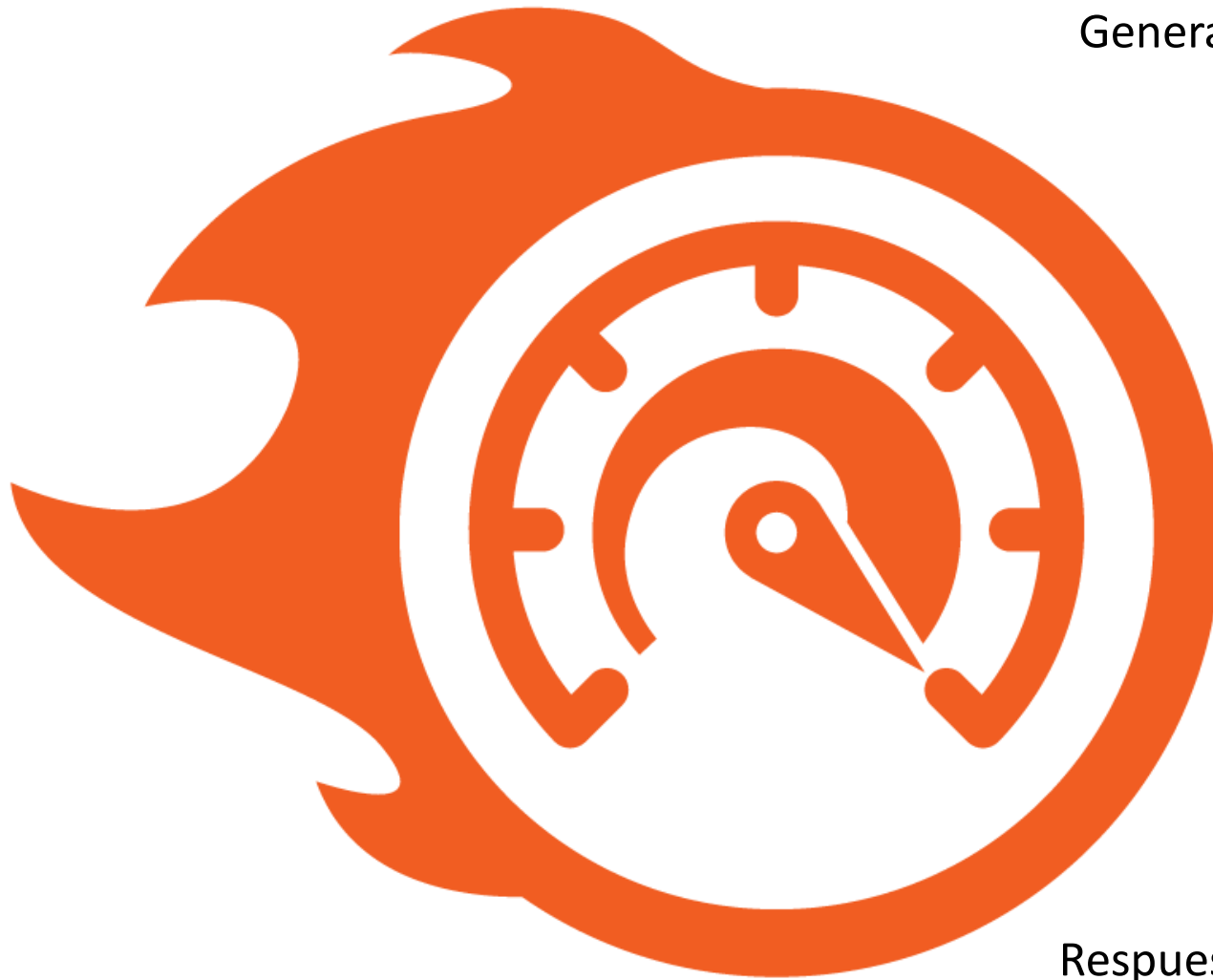
Small data



zettabyte is a trillion gigabytes (GB)

<https://www.bbvaopenmind.com/en/technology/digital-world/small-data-vs-big-data-back-to-the-basics/>

# V - Velocidad



Generación de datos

Procesamiento de datos

Respuestas en tiempo real

# V - Veracidad

Confiabilidad

Sesgos

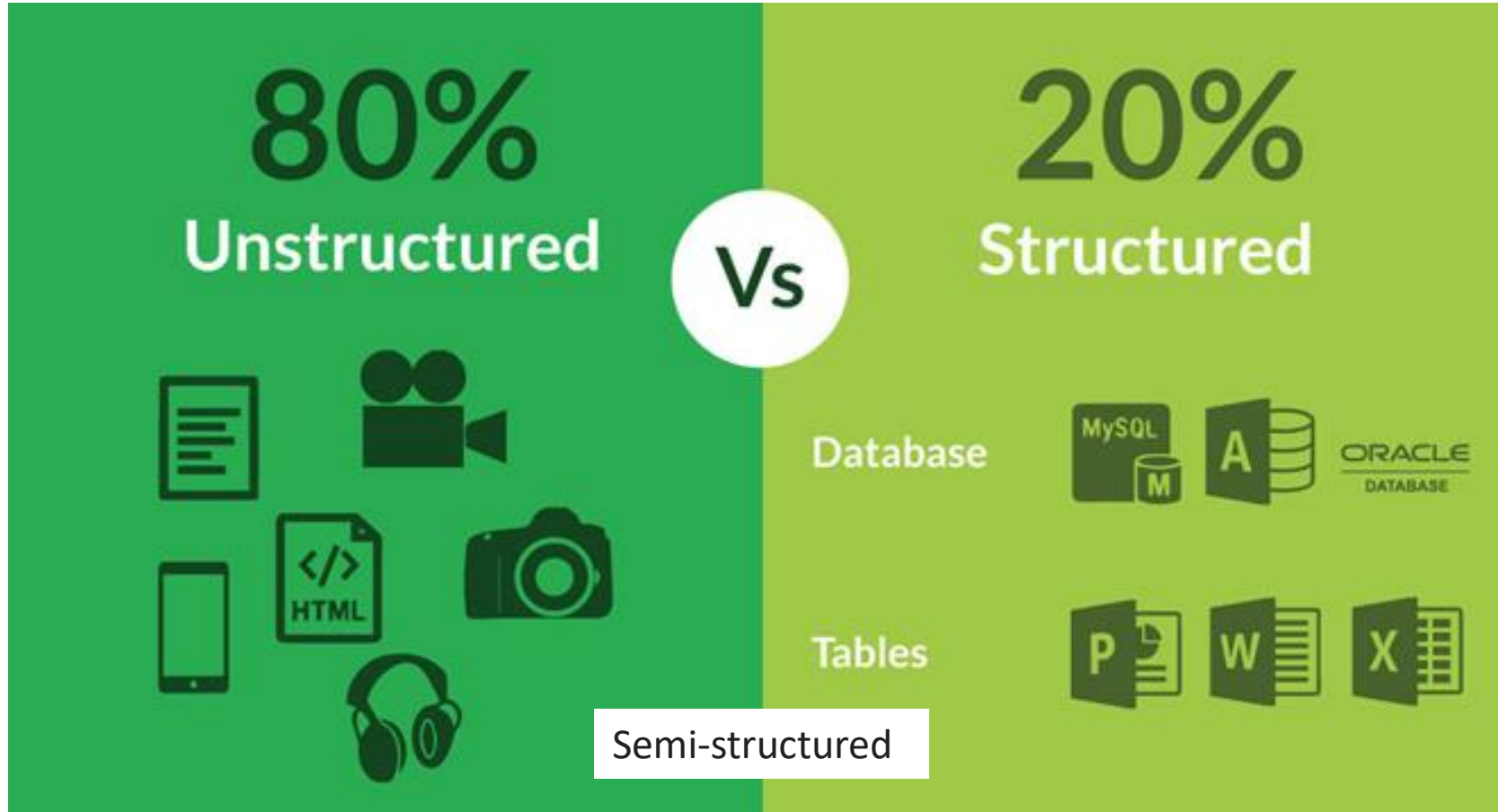


Precisión

Expertos

Pertinencia

# V - Variety



# V - Variety

Clasificación

Emails

XML

Jerarquías

JSON







Server Logs



Sensor Outputs

Geolocalización

# La plataforma de Big Data debe incluir seis aspectos clave

	Big Data Platform Imperatives	Technology Capability
1	Discover, explore, and navigate Big Data sources 	Federated Discovery, Search, and Navigation
2	Extreme performance—run analytics closer to data 	Massively Parallel Processing Analytic appliances
3	Manage and analyze unstructured data 	Hadoop File System/MapReduce Text Analytics
4	Analyze data in motion 	Stream Computing
5	Rich library of analytical functions and tools 	In-Database Analytics Libraries Big Data Visualization
6	Integrate and govern all data sources 	Integration, Data Quality, Security, Lifecycle Management, MDM, etc



# Pasos para el bigdata



# Habilidades requeridas para análisis en bigdata

## Store and process

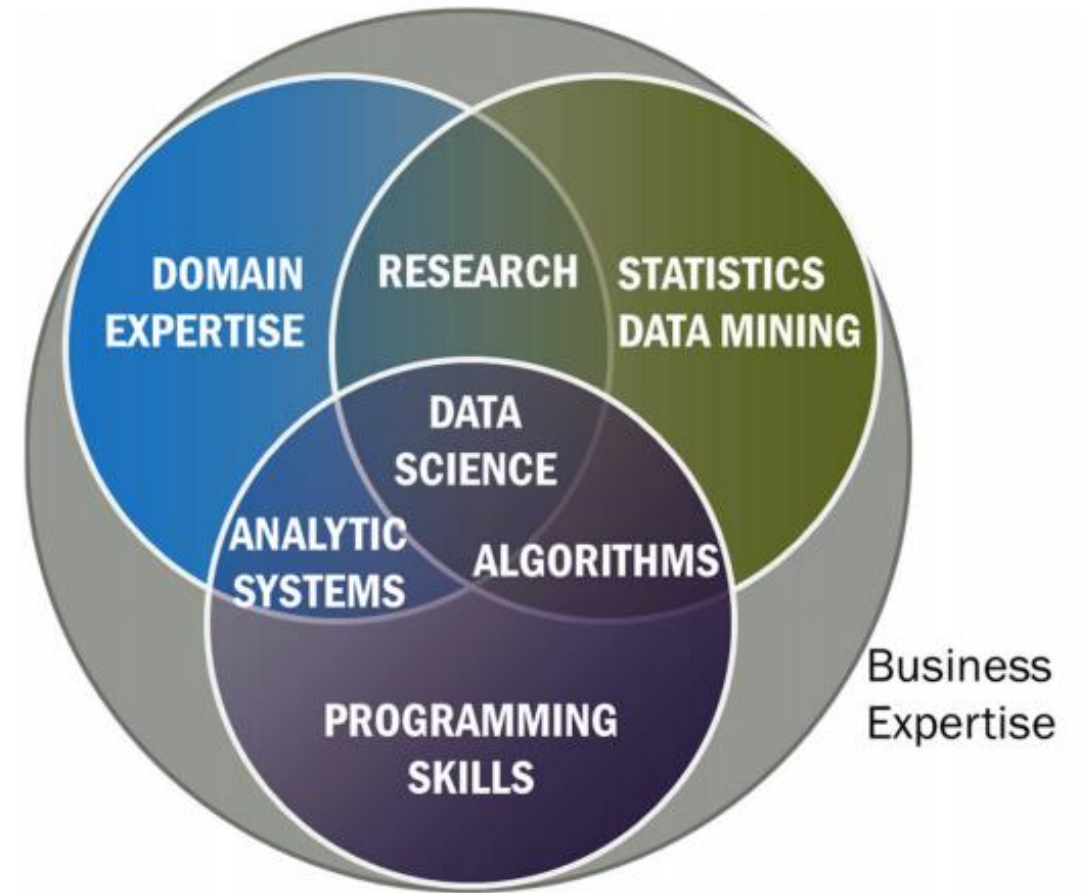
- Large scale databases
- Software Engineering
- System/network Engineering

## Analyse and model

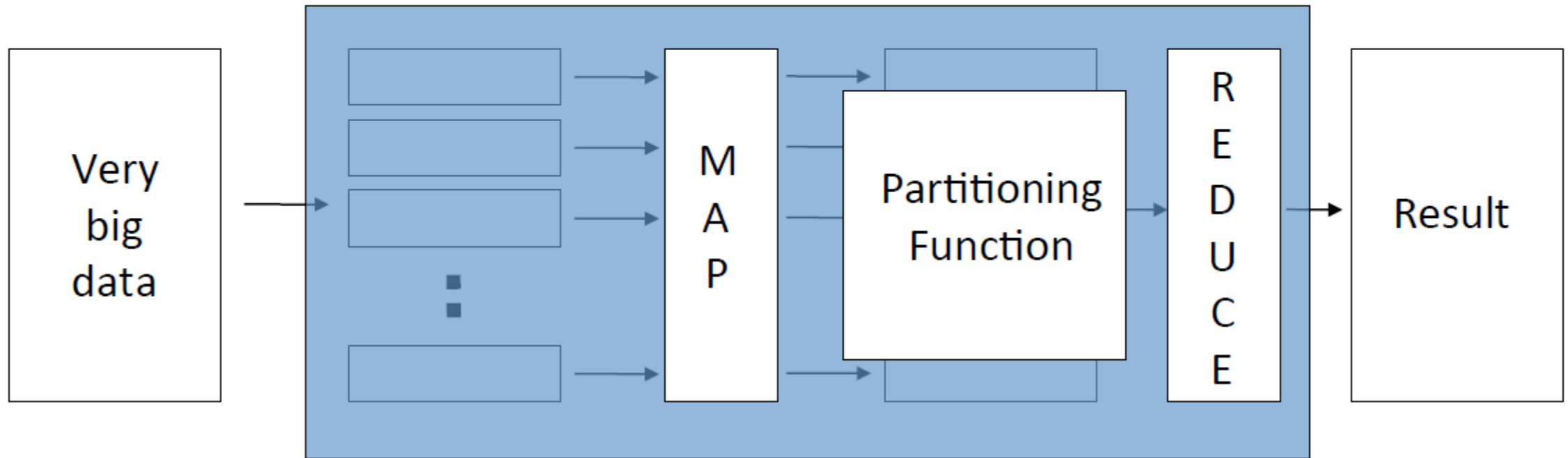
- Reasoning
- Knowledge Representation
- Multimedia Retrieval
- Modelling and Simulation
- Machine Learning
- Information Retrieval

## Understand and design

- Decision theory
- Visual analytics
- Perception Cognition



# Map Reduce



# Cuándo considerar una solución de Big Data

- ☐ Necesita procesar datos lo más rápido posible y necesita trabajar con un esquema bajo demanda.
- ☐ Desea analizar no solo datos estructurados sin procesar, sino también datos semiestructurados y no estructurados de una amplia variedad de fuentes.
- ☐ No está satisfecho con la eficacia de sus algoritmos o modelos
- ☐ No está completamente seguro de a dónde lo llevará la investigación y desea elasticidad de procesamiento, almacenamiento y los tipos de análisis.

Si sus respuestas a cualquiera de estas preguntas son "sí",  
Necesita considerar una solución de Big Data.

# El big data en el tiempo del covid - 19



## Datos:

- Telefonía.
- Transacciones en bancos.
- Compras en línea.
- Twitter

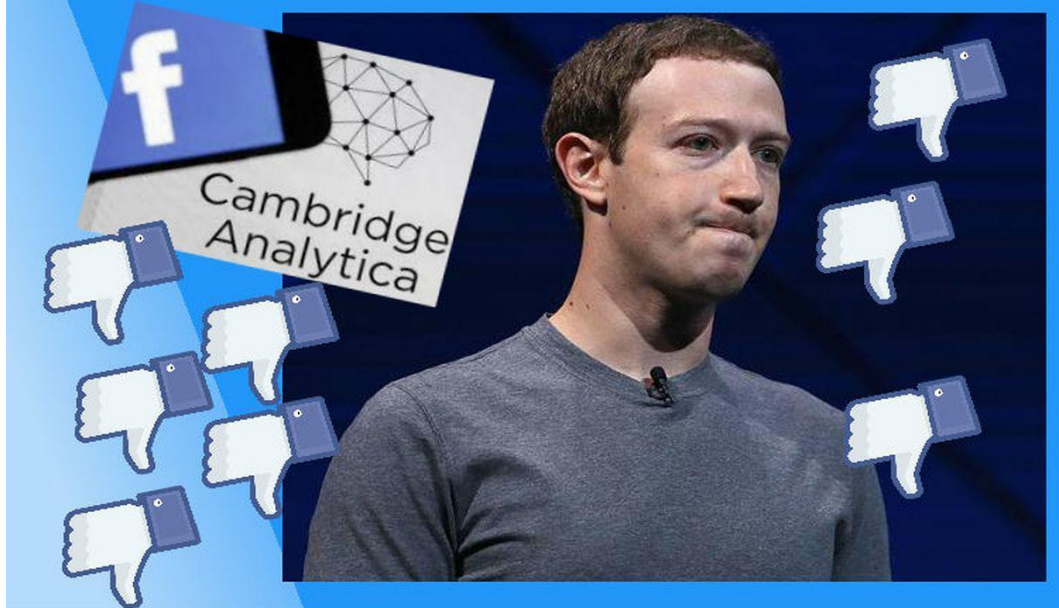
## Seguimiento a:

- Predicción de contagios por localidad.
- Monitoreo en tiempo real.
- Respuestas en la economía
- Conocimiento de la población sobre la enfermedad.

<https://blogs.iadb.org/ideas-que-cuentan/es/el-big-data-en-los-tiempos-del-coronavirus/>



# Ejemplos del poder los datos – Caso de cambridge analytical



"Explotamos Facebook para acceder a millones de perfiles de usuarios. Y **construimos modelos para explotar lo que sabíamos de ellos y apuntar a sus demonios internos**. Esa era la base sobre la cual la compañía se fundó", ex miembro de Cambridge Analytical

<https://www.bbc.com/mundo/noticias-43472797>

# El bigdata en el deporte

Cámaras y Sensores



Entorno competitivo

Video análisis

Predecir esfuerzos

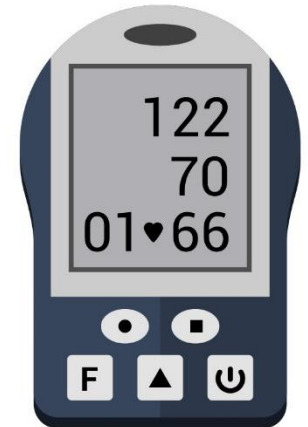
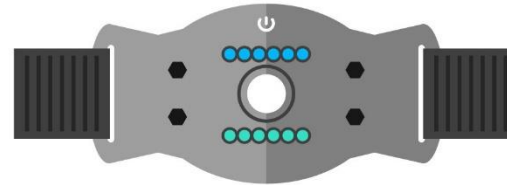
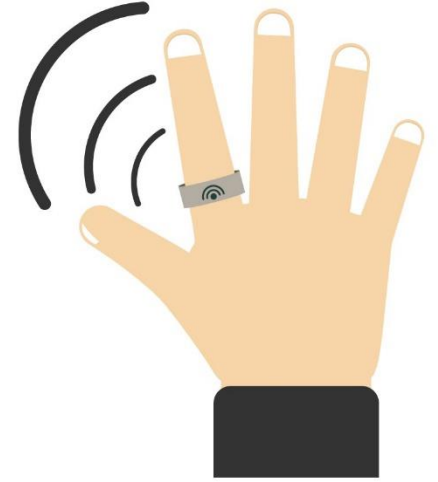
Evitar lesiones

Modelos de entrenamiento

Captación de talento

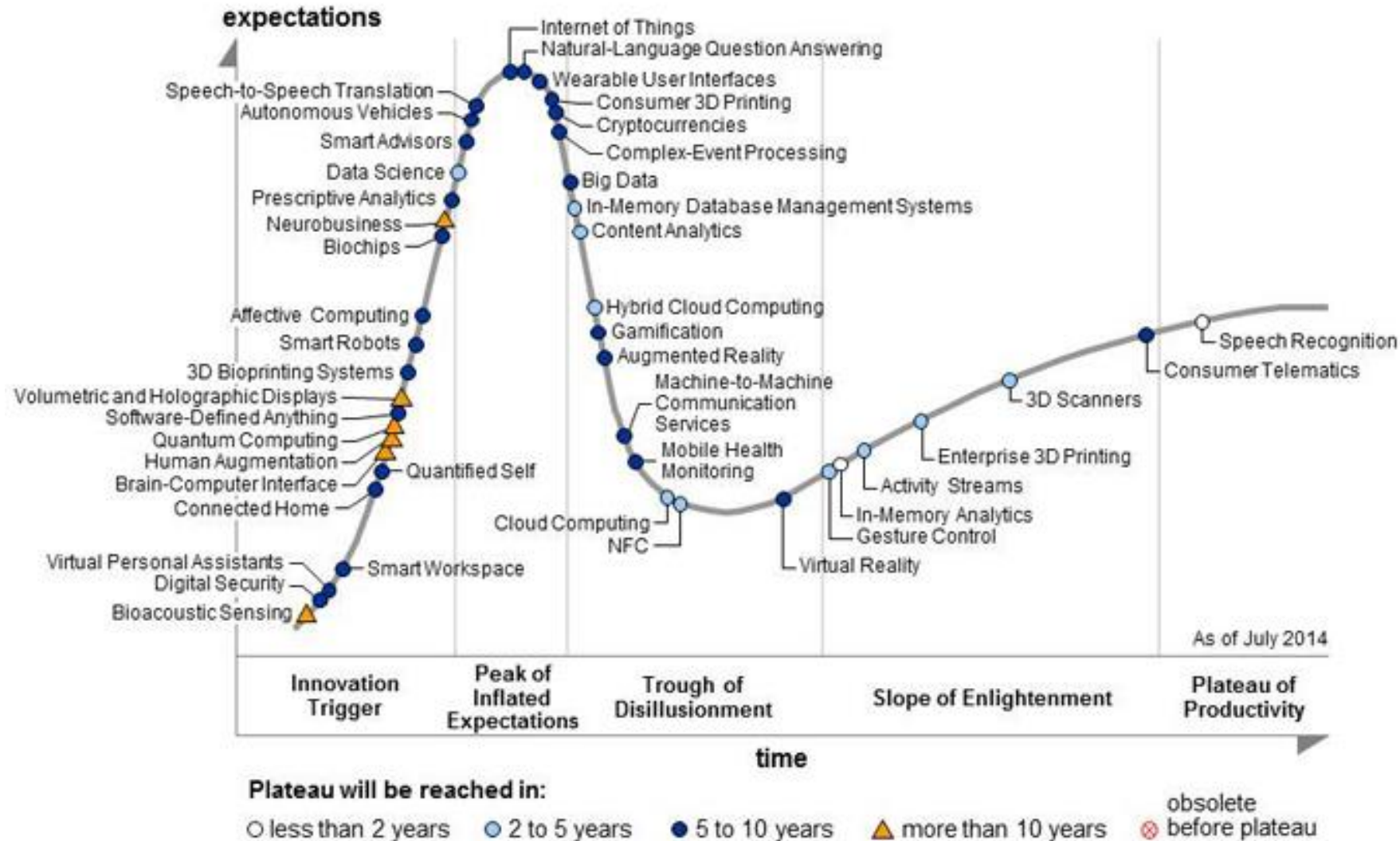
El Arsenal en el 2015 instaló en su estadio ocho cámaras que siguen en todo momento a sus jugadores y recogen 1,4 millones de datos por partido

# El bigdata en el deporte - Wereables



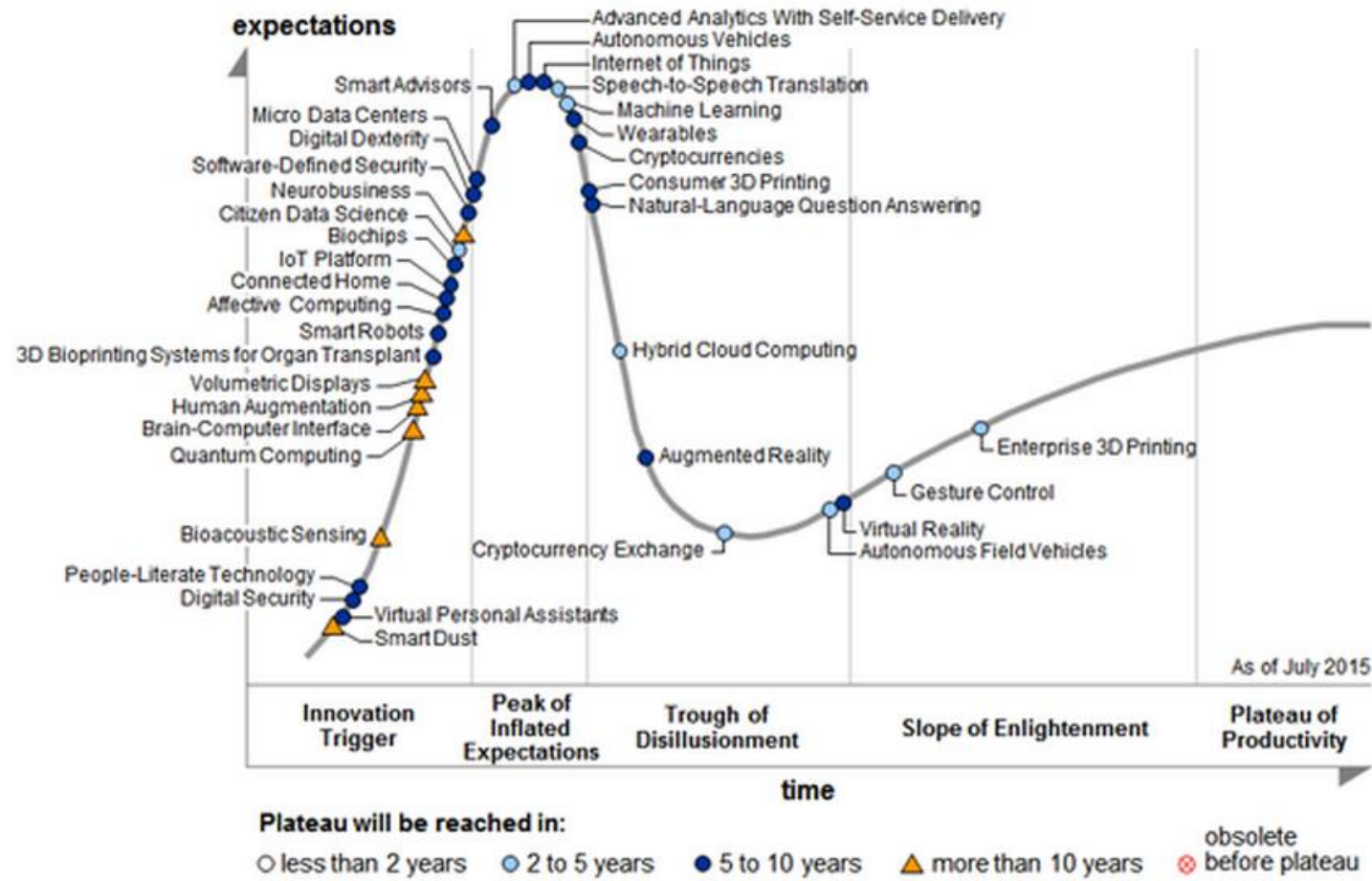


# Bigdata en el Gartner (2014)



# Bigdata en el Gartner

Figure 1. Hype Cycle for Emerging Technologies, 2015



# Herramientas y software

## Procesamiento de datos



## Datos espaciales

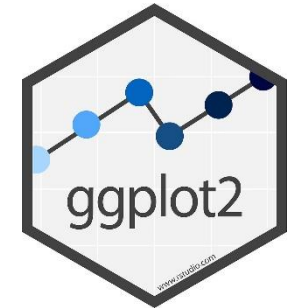


*Quantum Gis*

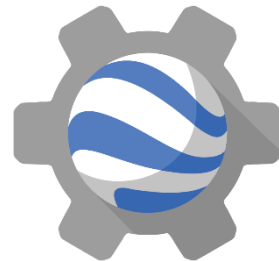
## Computación en la nube



## Visualización de datos



## Análisis de datos



*Google earth  
engine*



# Recursos en línea.



<https://www.coursera.org/specializations/big-data>

<https://www.coursera.org/specializations/jhu-data-science>

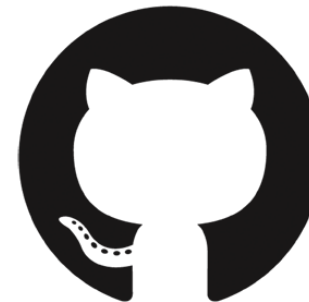


<https://www.edx.org/es/aprende/ciencia-de-datos>

<https://www.edx.org/es/micromasters/big-data>



<https://www.datacamp.com/courses/data-science-for-managers>



<https://github.com/hdorado>

<https://github.com/CIAT-DAPA>