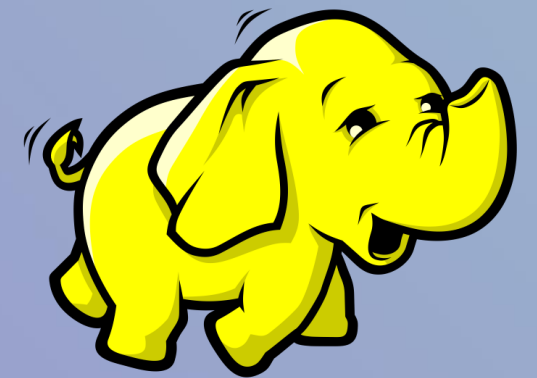


**DATA
DISCOVERY**



Introducción al Big Data

MIGUEL GARCÍA MENA



Presentación



Telefónica

8 años y 7 meses

Experto Análisis Big Data

sept. 2016 – actualidad · 4 años y 5 meses
Madrid

Análisis de Clientes GP & Big Data

jun. 2015 – sept. 2016 · 1 año y 4 meses
Madrid

Análisis e Investigación Comercial Empresas

jul. 2012 – jun. 2015 · 3 años
Madrid



Profesor

datahack school

oct. 2019 – actualidad · 1 año y 4 meses
Madrid y alrededores, España

Profesor de Introducción a R , en el MASTER BIG DATA & ANALYTICS



Becario de Investigación

Universidad Complutense de Madrid

nov. de 2009 – dic. de 2011 · 2 años y 2 meses

Becario de Investigación en la Cátedra del Departamento de Marketing de la UCM,
patrocinada por el Banco Santander



Universidad Politécnica de Madrid

Máster en Tratamiento Estadístico y Computacional de la Información, Matemáticas y estadística

2014 – 2016



Universidad Nacional de Educación a Distancia - U.N.E.D.

Máster en Métodos Estadísticos Avanzados, Estadística

2011 – 2012



Universidad Complutense de Madrid

Doctorando, Marketing

2010 – 2012



Universidad Complutense de Madrid

Postgrado en Marketing, Marketing

2010 – 2011



Universidad Complutense de Madrid

Postgrado en Dirección de Empresas, Administración y Dirección de Empresas

2007 – 2009



Universidad Complutense de Madrid

Diplomatura en Ciencias Empresariales, Administración y Dirección de Empresas

2004 – 2007

<https://www.linkedin.com/in/miguel-garcia-mena/>

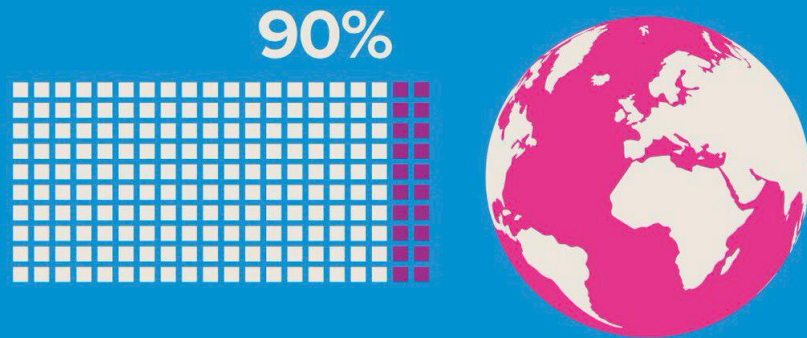


¿Qué es realmente el Big Data?



Cada día generamos más datos...

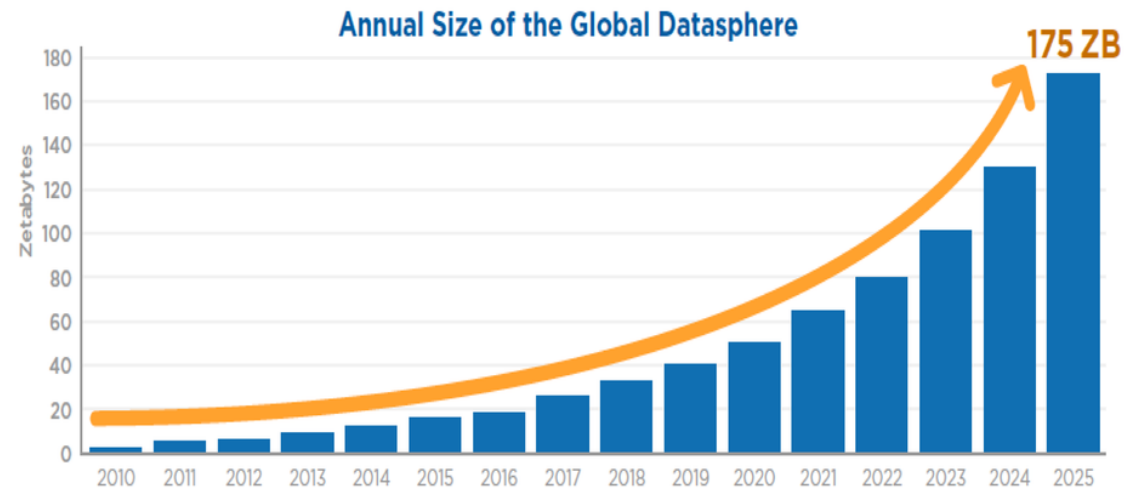
IBM estimates that 90% of all the world's data has been created in the last two years.



... y más rápido:

- Automatización
- Acceso a Internet universal
- Contenido generado por los usuarios

- Transacciones financieras
- Redes de sensores
- Logs de servidores y aplicaciones
- Redes sociales

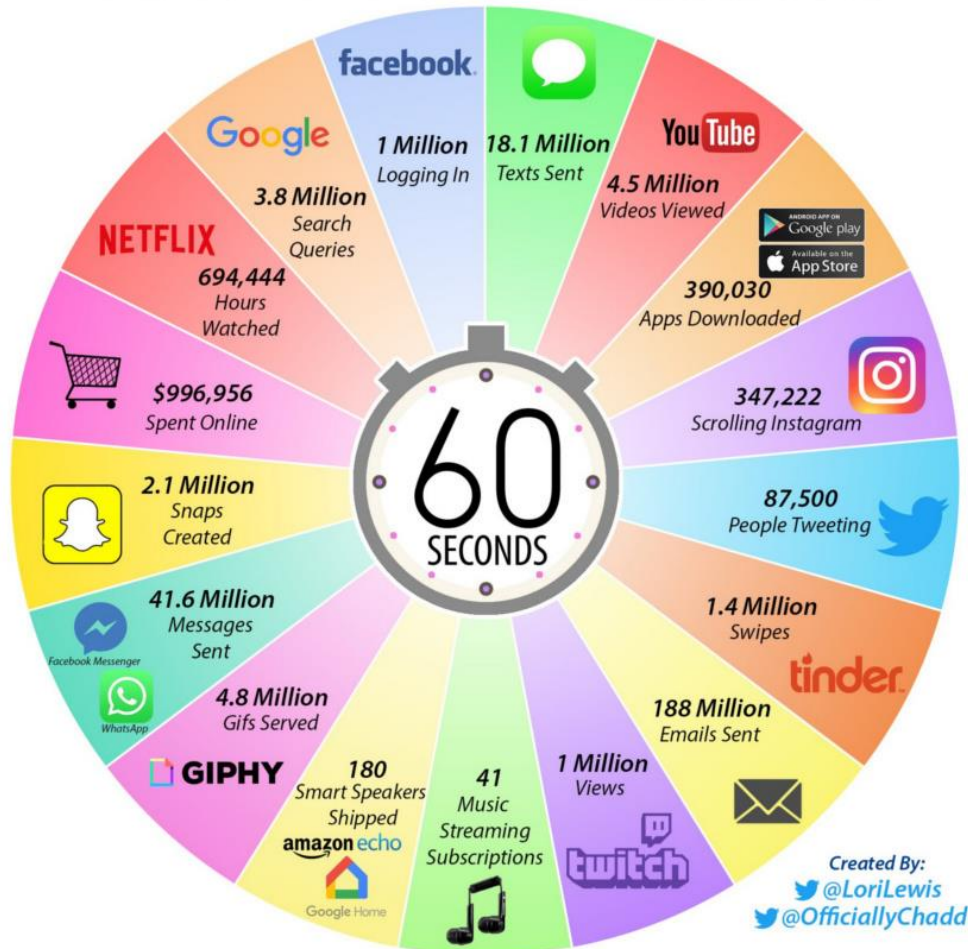


Source: Data Age 2025, sponsored by Seagate with data from IDC Global DataSphere, Nov 2018



... cada vez más variados

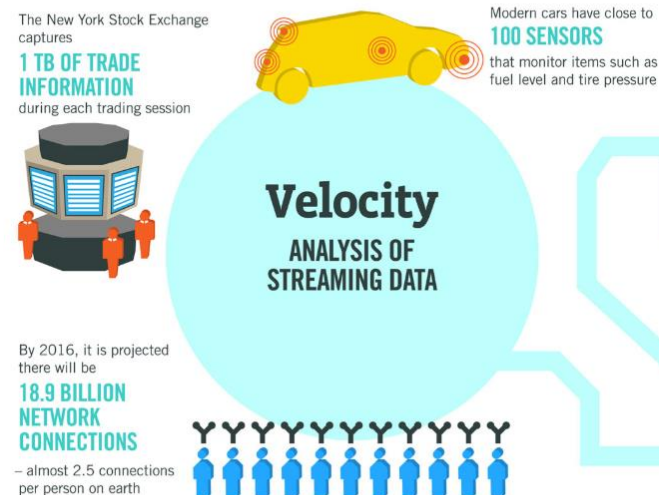
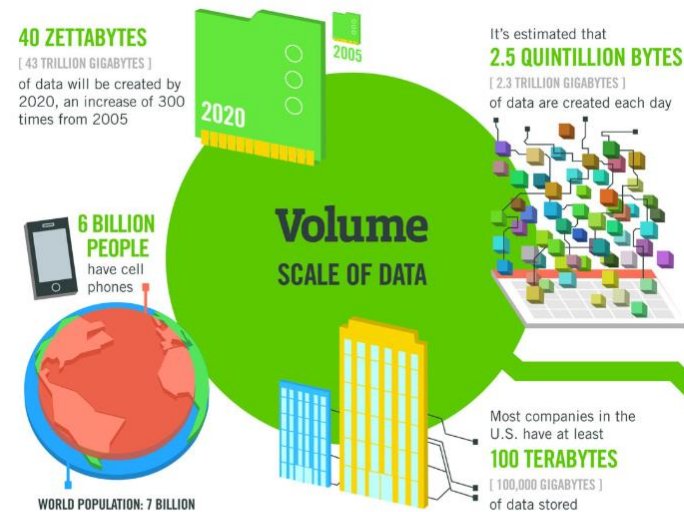
2019 *This Is What Happens In An Internet Minute*



2020 *This Is What Happens In An Internet Minute*



Las V's del Big Data



The FOUR V's of Big Data

From traffic patterns and music downloads to web history and medical records, data is recorded, stored, and analyzed to enable the technology and services that the world relies on every day. But what exactly is big data, and how can these massive amounts of data be used?

As a leader in the sector, IBM data scientists break big data into four dimensions: **Volume, Velocity, Variety and Veracity**

Depending on the industry and organization, big data encompasses information from multiple internal and external sources such as transactions, social media, enterprise content, sensors and mobile devices. Companies can leverage data to adapt their products and services to better meet customer needs, optimize operations and infrastructure, and find new sources of revenue.

By 2015
4.4 MILLION IT JOBS
will be created globally to support big data,
with 1.9 million in the United States

As of 2011, the global size of data in healthcare was estimated to be

150 EXABYTES
[161 BILLION GIGABYTES]



30 BILLION PIECES OF CONTENT
are shared on Facebook every month



By 2014, it's anticipated there will be

420 MILLION WEARABLE, WIRELESS HEALTH MONITORS



4 BILLION+ HOURS OF VIDEO
are watched on YouTube each month



400 MILLION TWEETS
are sent per day by about 200 million monthly active users



Variety

DIFFERENT FORMS OF DATA

1 IN 3 BUSINESS LEADERS

don't trust the information they use to make decisions



Poor data quality costs the US economy around

\$3.1 TRILLION A YEAR



27% OF RESPONDENTS

in one survey were unsure of how much of their data was inaccurate

Veracity

UNCERTAINTY OF DATA



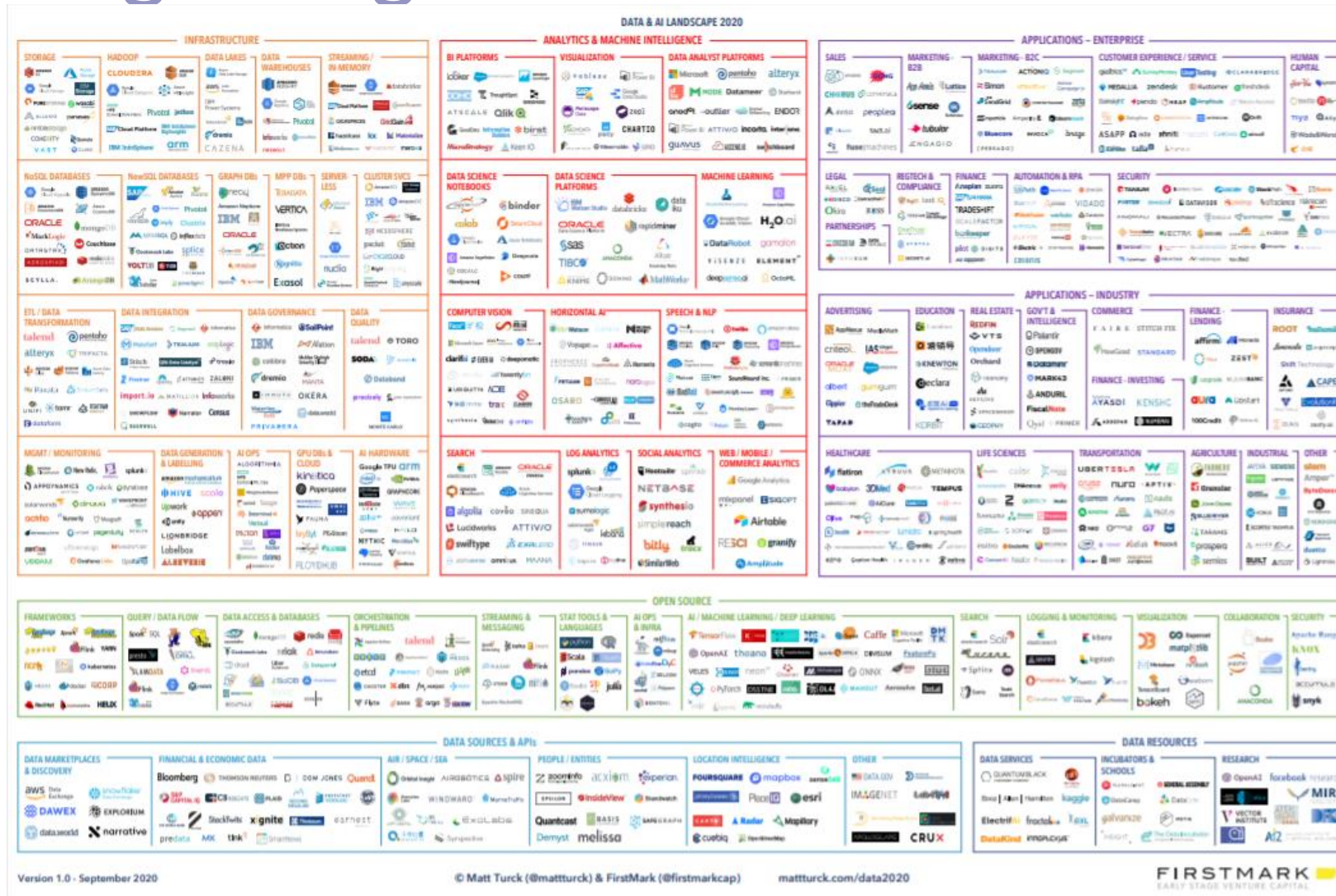
Con un objetivo...



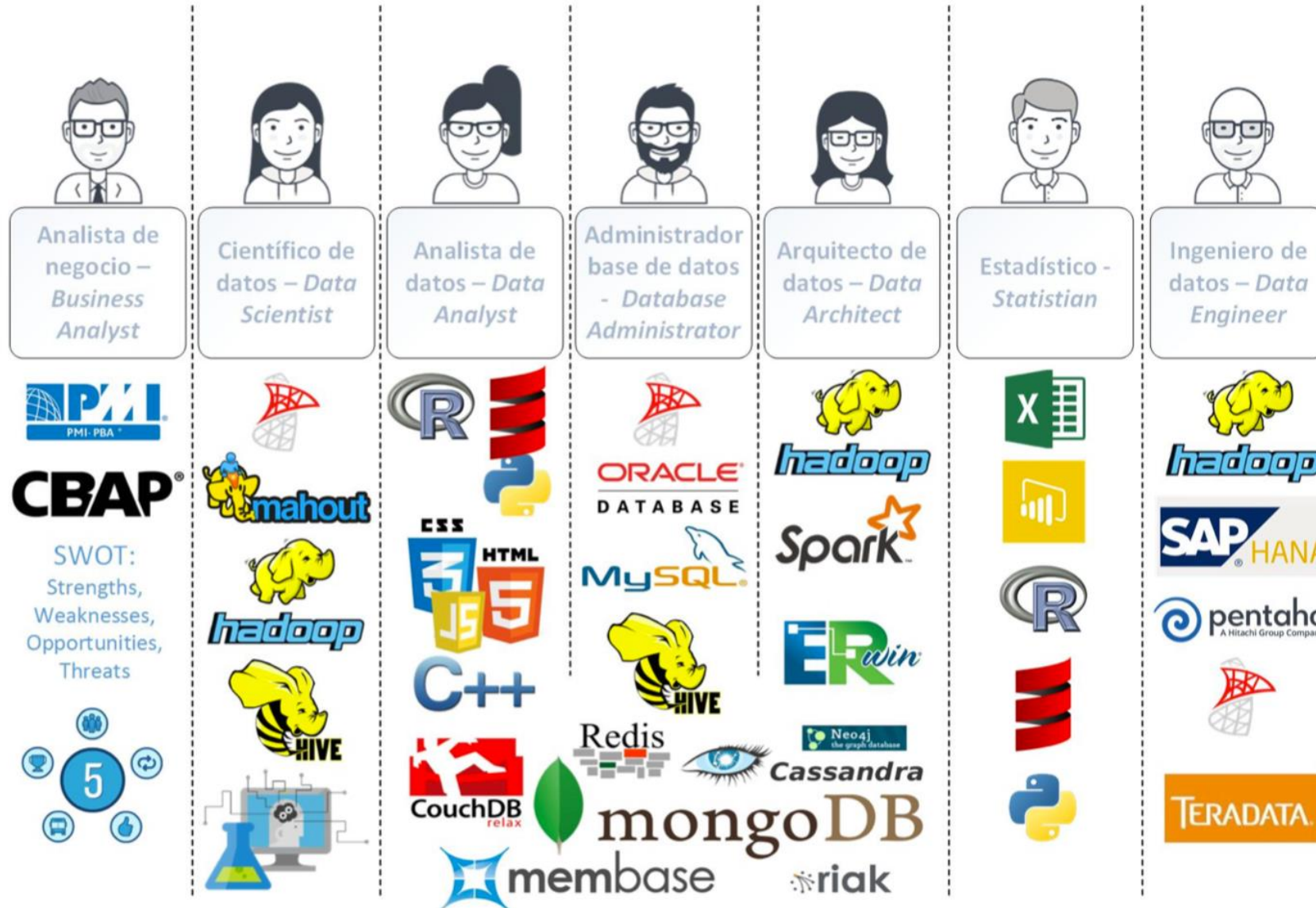
- Establecer procesos **extracción de información**
- Incrementar el **conocimiento** de negocio
- Mejorar la capacidad de **decisión**: medir las acciones y resultado
- **Ahorrar** grandes cantidades de dinero
- Abrir **nuevas oportunidades** de negocios



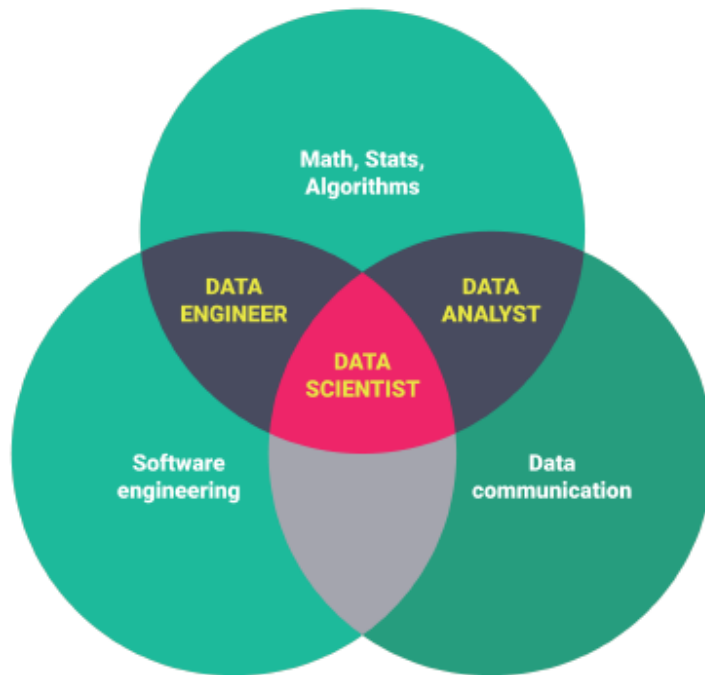
Tecnologías Big Data



Perfiles Big Data



¿Qué es un Data Scientist?



MODERN DATA SCIENTIST

Data Scientist, the sexiest job of 21st century requires a mixture of multidisciplinary skills ranging from an intersection of mathematics, statistics, computer science, communication and business. Finding a data scientist is hard. Finding people who understand who a data scientist is, is equally hard. So here is a little cheat sheet on who the modern data scientist really is.

MATH & STATISTICS

- ☆ Machine learning
- ☆ Statistical modeling
- ☆ Experiment design
- ☆ Bayesian inference
- ☆ Supervised learning: decision trees, random forests, logistic regression
- ☆ Unsupervised learning: clustering, dimensionality reduction
- ☆ Optimization: gradient descent and variants

PROGRAMMING & DATABASE

- ☆ Computer science fundamentals
- ☆ Scripting language e.g. Python
- ☆ Statistical computing package e.g. R
- ☆ Databases SQL and NoSQL
- ☆ Relational algebra
- ☆ Parallel databases and parallel query processing
- ☆ MapReduce concepts
- ☆ Hadoop and Hive/Pig
- ☆ Custom reducers
- ☆ Experience with xaaS like AWS

DOMAIN KNOWLEDGE & SOFT SKILLS

- ☆ Passionate about the business
- ☆ Curious about data
- ☆ Influence without authority
- ☆ Hacker mindset
- ☆ Problem solver
- ☆ Strategic, proactive, creative, innovative and collaborative

COMMUNICATION & VISUALIZATION

- ☆ Able to engage with senior management
- ☆ Story telling skills
- ☆ Translate data-driven insights into decisions and actions
- ☆ Visual art design
- ☆ R packages like ggplot or lattice
- ☆ Knowledge of any of visualization tools e.g. Flare, D3.js, Tableau

A cartoon illustration of a man and a woman standing side-by-side. The man is wearing a blue blazer, a white shirt, a red tie, and red trousers. The woman is wearing a red top, a dark blue skirt, and red shoes. Both are wearing glasses and holding a tablet.

MarketingDistillery.com is a group of practitioners in the area of e-commerce marketing. Our fields of expertise include: marketing strategy and optimization; customer tracking and on-site analytics; predictive analytics and econometrics; data warehousing and big data systems; marketing channel insights in Paid Search, SEO, Social, CRM and brand.

Marketing
DISTILLERY



Son además perfiles muy demandados

Profesiones más demandadas en un futuro próximo (de 2 a 3 años)			
	Sector	Profesión	En % del total
1	TECNOLOGÍA	'Big data'	11,24
2	TECNOLOGÍA	'Data science'	9,55
3	COMERCIAL	Comercial digital	5,06
4	INGENIERÍA	Ingeniero informático	4,49
	TECNOLOGÍA	Programador informático	4,49
5	INGENIERÍA	Ingeniero industrial	3,93
6	OPERARIOS	Operarios cualificados	3,37
	TECNOLOGÍA	Responsable de ciberseguridad	3,37
7	RR HH	Responsable 'HR analytics'	2,81
8	DIRECCIÓN	CTO - 'Chief transformation officer'	2,25
	TECNOLOGÍA	Desarrollador aplicaciones multimedia	2,25
	TECNOLOGÍA	Especialistas 'e-commerce development'	2,25
	COMERCIAL	'Senior key account manager'	2,25
	INGENIERÍA	Ingeniero de proyecto	2,25
	SALUD	Médico	2,25
	TECNOLOGÍA	'User experience (UX) manager'	2,25
	COMERCIAL	Consultor	2,25

https://cincodias.elpais.com/cincodias/2020/02/27/fortunas/1582830031_321411.html



DATA

Data Scientist: The Sexiest Job of the 21st Century

by [Thomas H. Davenport](#) and [D.J. Patil](#)

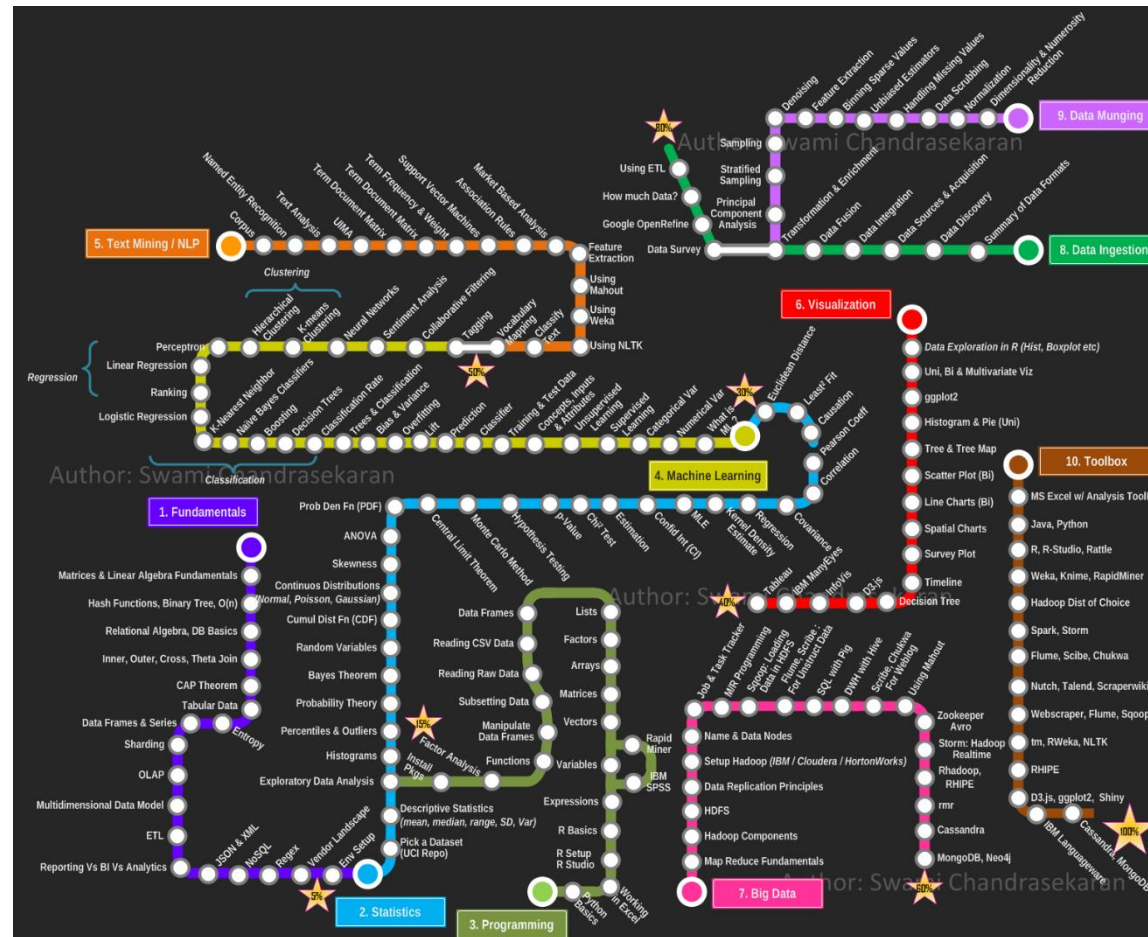
FROM THE OCTOBER 2012 ISSUE

 Summary  Save  Share  16 Comment  Text Size  Print **\$8,95** Buy Copies

When Jonathan Goldman arrived for work in June 2006 at LinkedIn, the business networking site, the place still felt like a start-up. The company had just under 8 million accounts, and the number was growing quickly as existing members invited their friends and colleagues to join. But users weren't seeking out connections with the people who were already on



Hoy comenzáis un camino largo...



MUCHO ÁNIMO A TODOS EN ESTA NUEVA ETAPA





datahack & datahack_school