



XV Congreso
internacional

ALASA 2018

Lima, Perú.

➤ 15 al 18 de mayo de 2018.
The Westin Hotel
& Convention Center

“ Big Data: disrupción económica y sector asegurador ”

Gonzalo de Cadenas Santiago, PhD.



Asociación Latinoamericana para
el desarrollo del Seguro Agropecuario



***“THE CONTRIBUTION OF THE
MANY TO THE MISFORTUNES
OF THE FEW”***

Objetivo

- NO solo es comentar “*Big Data*” y negocio asegurador
- El negocio asegurador es “*model driven*” → “*data driven (*)*”
- “*Big Data*” incluye a la “*Ciencia de Datos*”
- La *Ciencia de Datos* puede hacer frente a problemas fundamentales inherentes al negocio asegurador, reducir ineficiencias, externalidades y costes sociales (contribuir al sentido del aseguramiento, mantra)
- La solución de dichas ineficiencias son especialmente aplicables al negocio del seguro agrícola

Problemas económicos en la naturaleza del aseguramiento

1. *In-observabilidad de la acción del asegurado (selección adversa ~ abuso)*
2. *Desconocimiento de la distribución real de las pérdidas y de las frecuencias de suceso (sesgo & varianza ~solidaridad)*
3. *Información parcial sobre homogeneidad de la cartera aseguradora (pooling/discriminación ~ineficiencia)*
4. *Alta “fluidez” de eventos conduce impredecibilidad (desorden o aleatoriedad ~ capacidad para evolucionar)*

Ineficiencias/externalidades

- 1) No observabilidad **genera Incentivos a tomar mayor riesgo** (MH).
Distorsión de coberturas e incremento primas. El coste se comparte entre todos asegurados, la aseguradora y el estado. (*Welfare loss*)
- 2) **Diseño ineficiente** de las coberturas, mala estimación del riesgo emergente (*unfair premiums*)
- 3) **Sub-óptimo** (*mutualización / segmentación*) en el modelo de **negocio** al desconocer la población asegurada y sus vínculos reales
- 4) Impredictibilidad riesgo cambiante con entorno **hace el seguro inefectivo**

Ciencia de Datos & Aseguramiento

- Importan los datos porque condicionan el modelo probabilístico y el mecanismo de incentivos y preferencias del asegurado, y por extensión son claves para el correcto cálculo de la prima optima
- Un negocio que es *Model Driven* necesariamente es *Data Driven*, todo lo que sea mejorar la explotación de estos mejora el ejercicio de nuestro cometido
- Ciencia de Datos una manera de reducir ineficiencias y externalidades inherentes a los problemas conceptuales que subyacen al ejercicio de aseguramiento

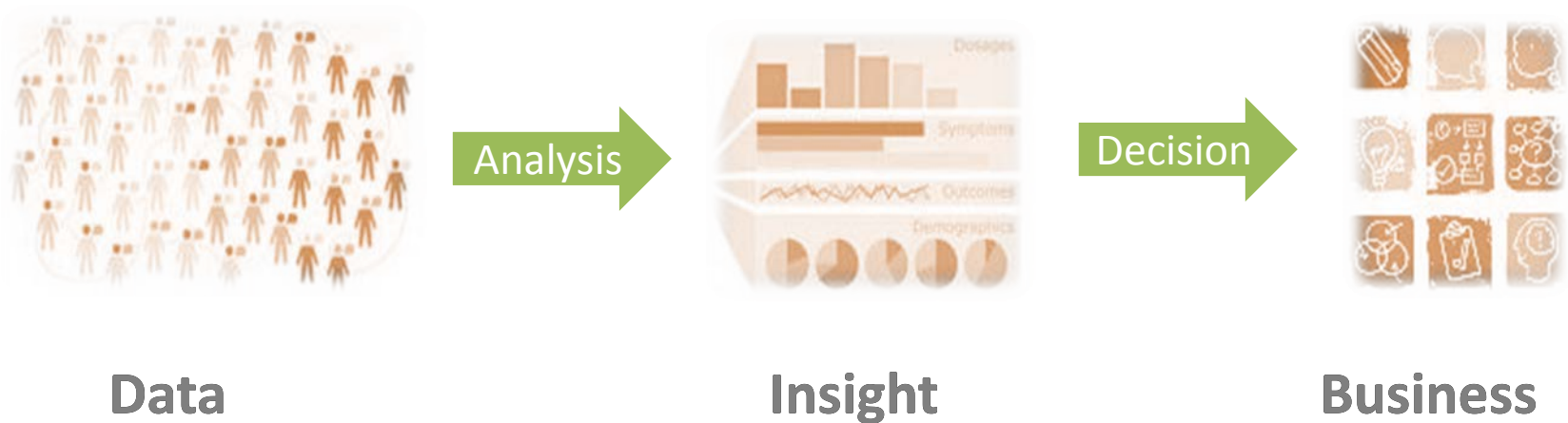
Necesitamos

Mejorar la cobertura --> mejorar la información disponible acerca del riesgo en entorno de creciente complejidad e impredecibilidad

- 1) Mejor información sobre la pérdida potencial
- 2) Conocer la frecuencia de ocurrencia
- 3) Decidir entre granularidad del riesgo y su coste
- 4) Características de la cartera para correcto *pooling*
- 5) Información sobre el comportamiento de los agentes
- 6) Mayor información sobre la variabilidad del riesgo (*solli.cost*)

Qué es Big Data

- *Gartner: “Big data is high-volume, high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation (Data Science)”*

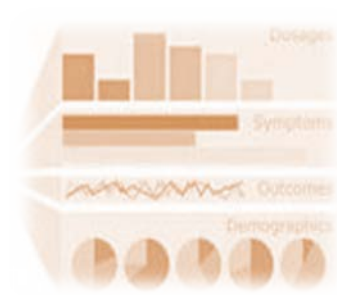


Qué es Big Data

...as to anticipate, improve and accelerate decision making in the context of vast amounts of data, applying intelligence to generate competitive advantage.



Data

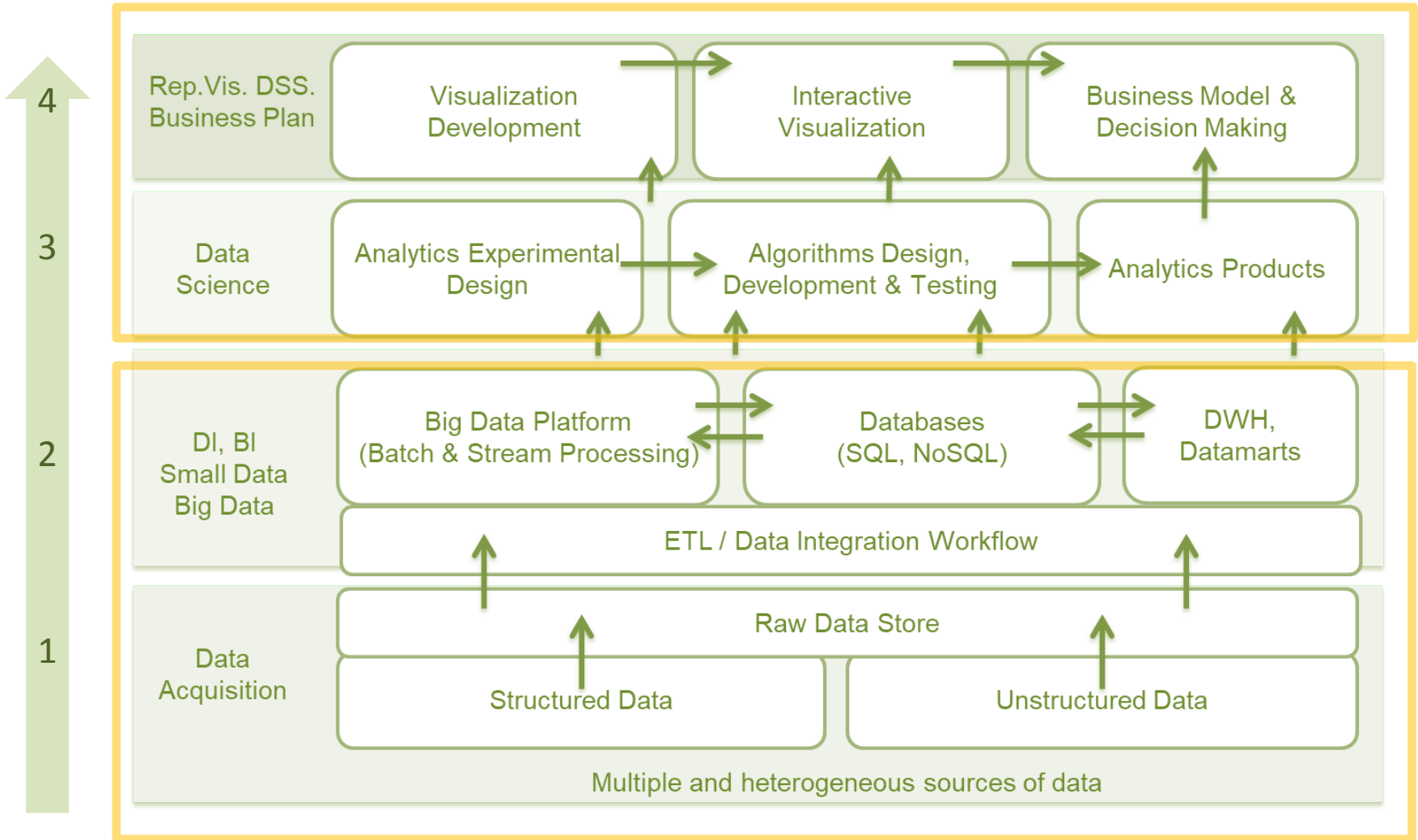


Insight



Business

Ecosistema de Big Data



Ecosistema de Big Data

Big Data Stacks y Eco. de Datos

Data Mining & Análisis

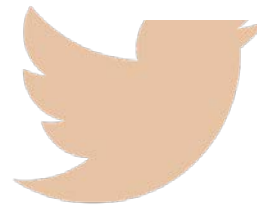
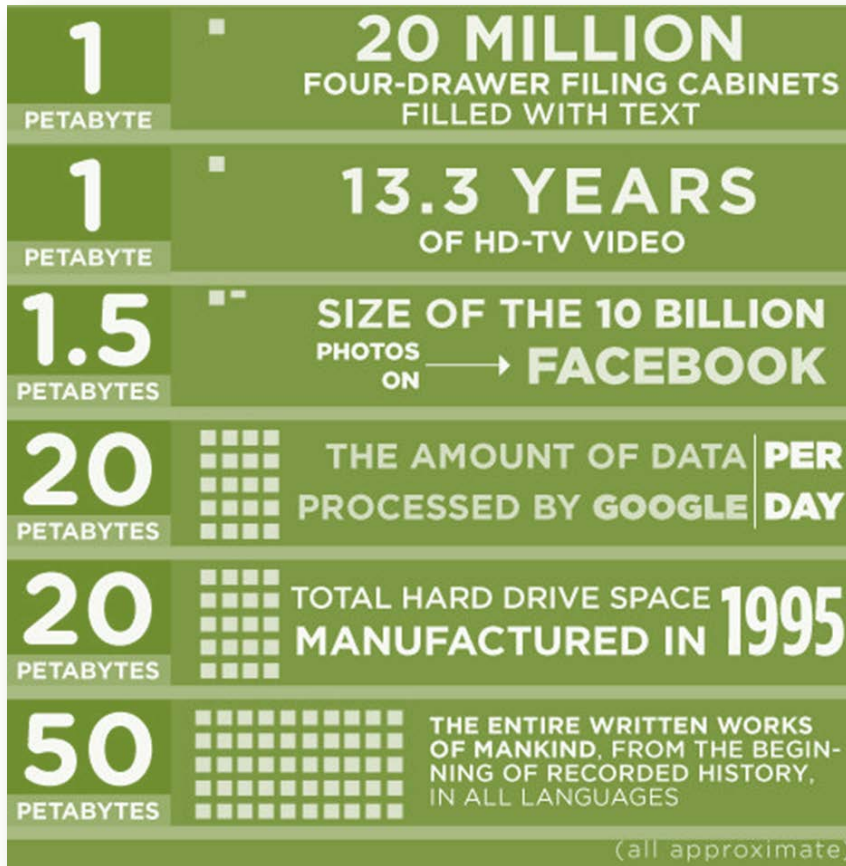


Visualización

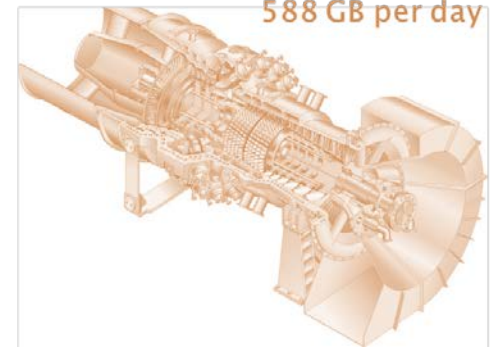
Instrumentos de *M. Learning*



¿Cuánto es “grande y variado”?



270 million users /
100 GB per day

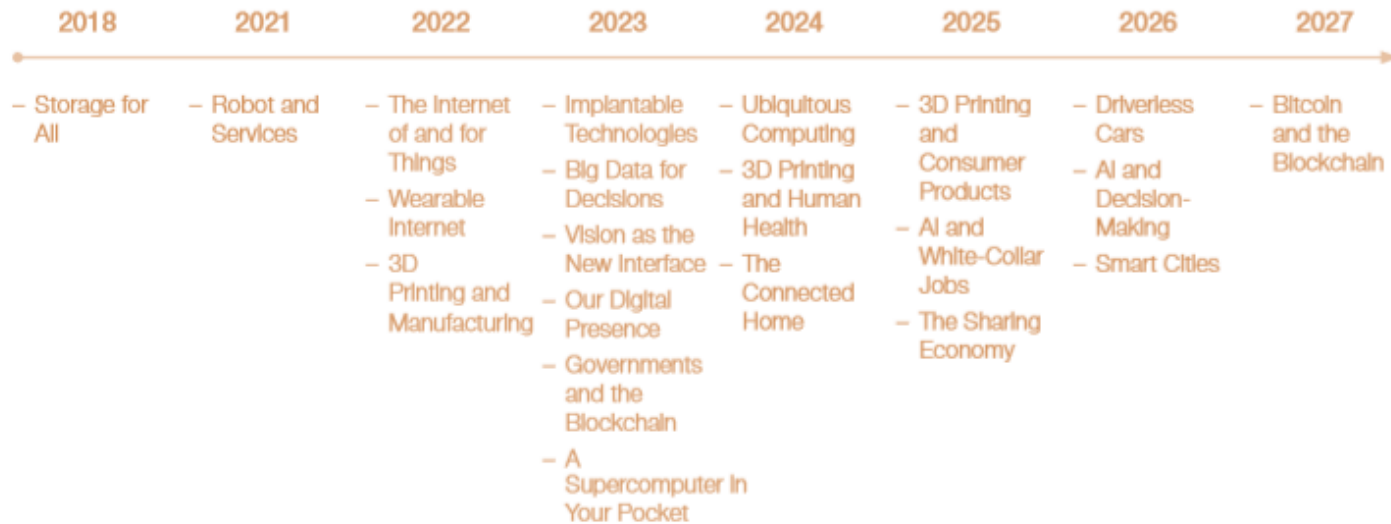


1 turbine /
588 GB per day

byte	b	$2^0 = 1$
Kbyte	KB	$2^{10} = 1\ 024$
Megabyte	MB	$2^{20} = 1\ 048\ 576$
Gigabyte	GB	$2^{30} = 1\ 073\ 741\ 824$
Terabyte	TB	$2^{40} = 1\ 099\ 511\ 627\ 776$
Petabyte	PB	$2^{50} = 1\ 125\ 899\ 906\ 842\ 624$
Exabyte	EB	$2^{60} = 1\ 152\ 921\ 504\ 606\ 846\ 976$
Zettabyte	ZB	$2^{70} = 1\ 180\ 591\ 620\ 717\ 411\ 303\ 424$
Yottabyte	YB	$2^{80} = 1\ 208\ 925\ 819\ 614\ 629\ 174\ 706\ 176$

¿Cómo de relevante?

Figure 1: Average Year Each Tipping Point Is Expected to Occur ¹



The survey results were also analysed to see what percentage of the respondents expected the tipping point to have occurred by 2025, or ten years from now. Eleven of the 21 transition points had a high expectation (over 80%) of occurring by this date.

¿Cuánto es “grande y variado”?

<Lo que pide la nueva conectividad>

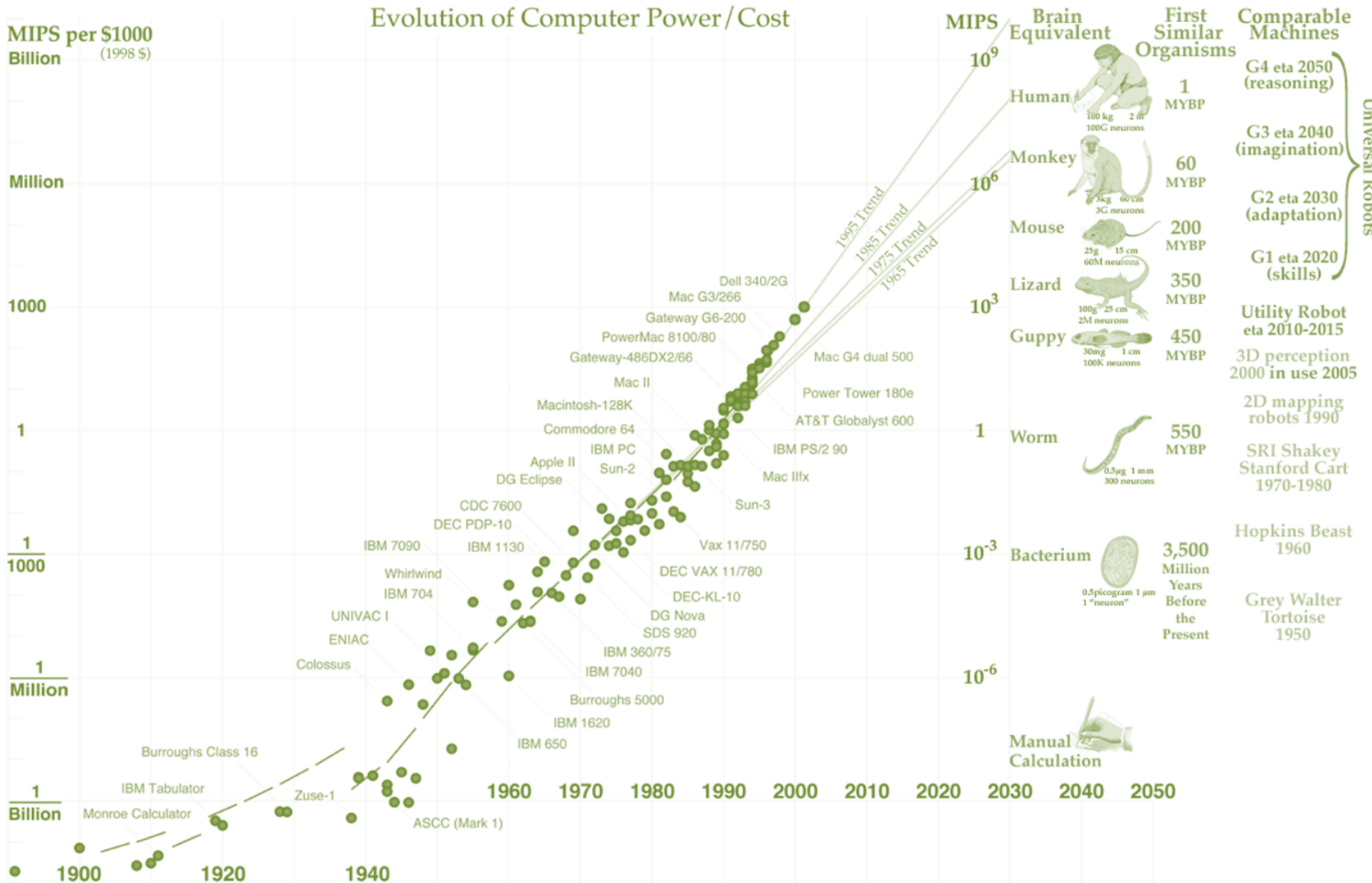
Facebook (use)

12 TB added per day.
800 TB processed per day.
25.000 MapReduce tasks per day.
65 Million files in HDFS.
30.000 concurrent clients.

Facebook (infrastructure)

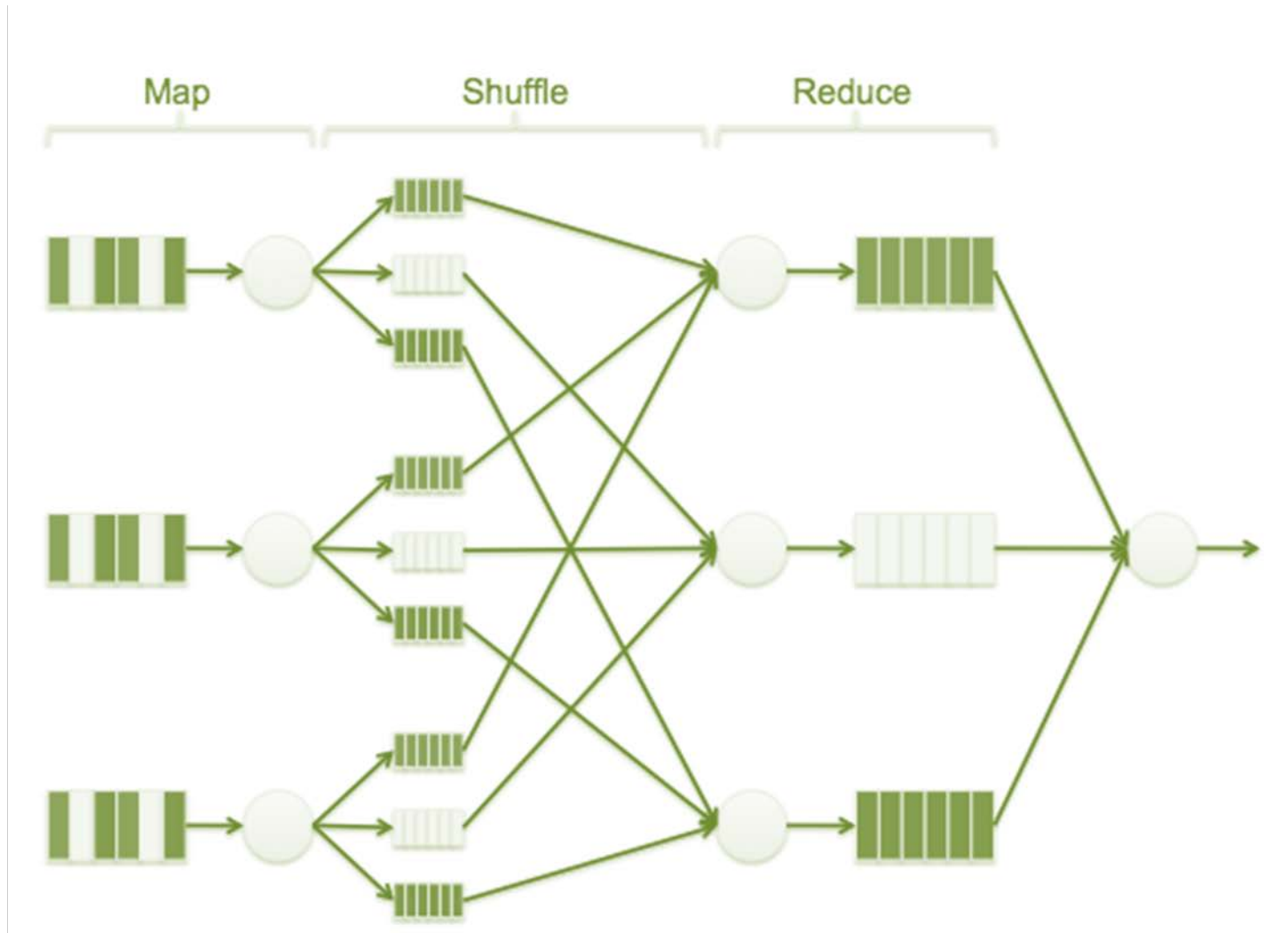
21 PB of data.
2.000 servers.
1.200 machines with 8 cores.
800 machines with 16 cores.
12 TB per machine.
32 GB RAM per machine.
15 MapReduce tasks per machine

¿Por qué ahora?



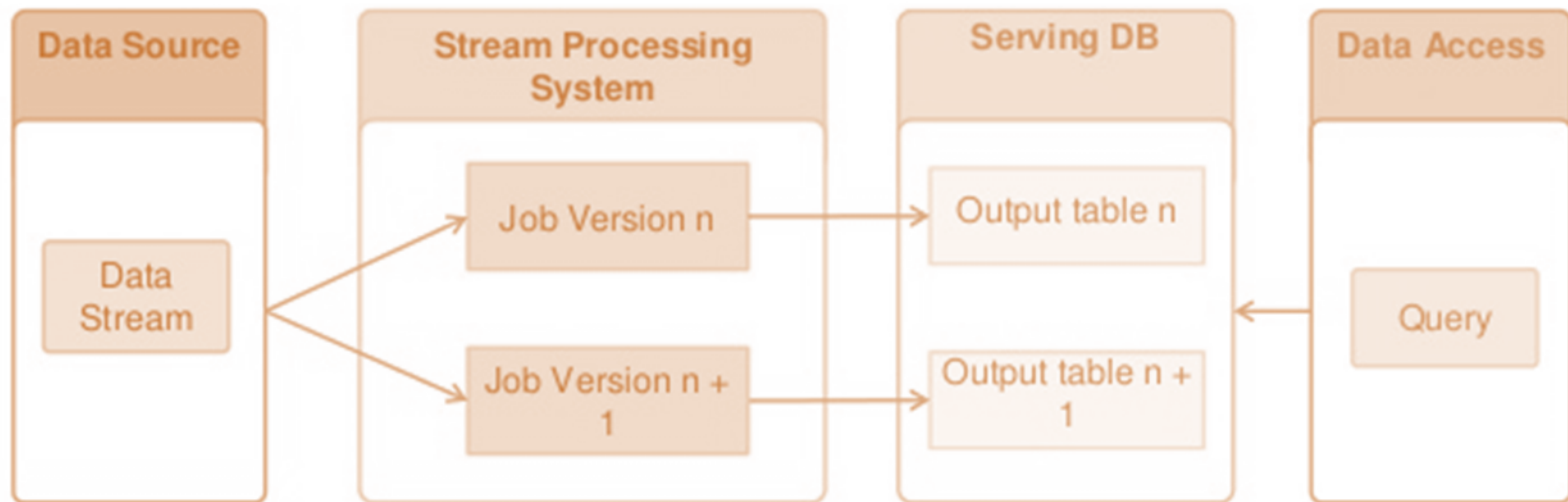
La verdadera disrupción: la horizontalidad

MapReduce (Divide and Conquer)



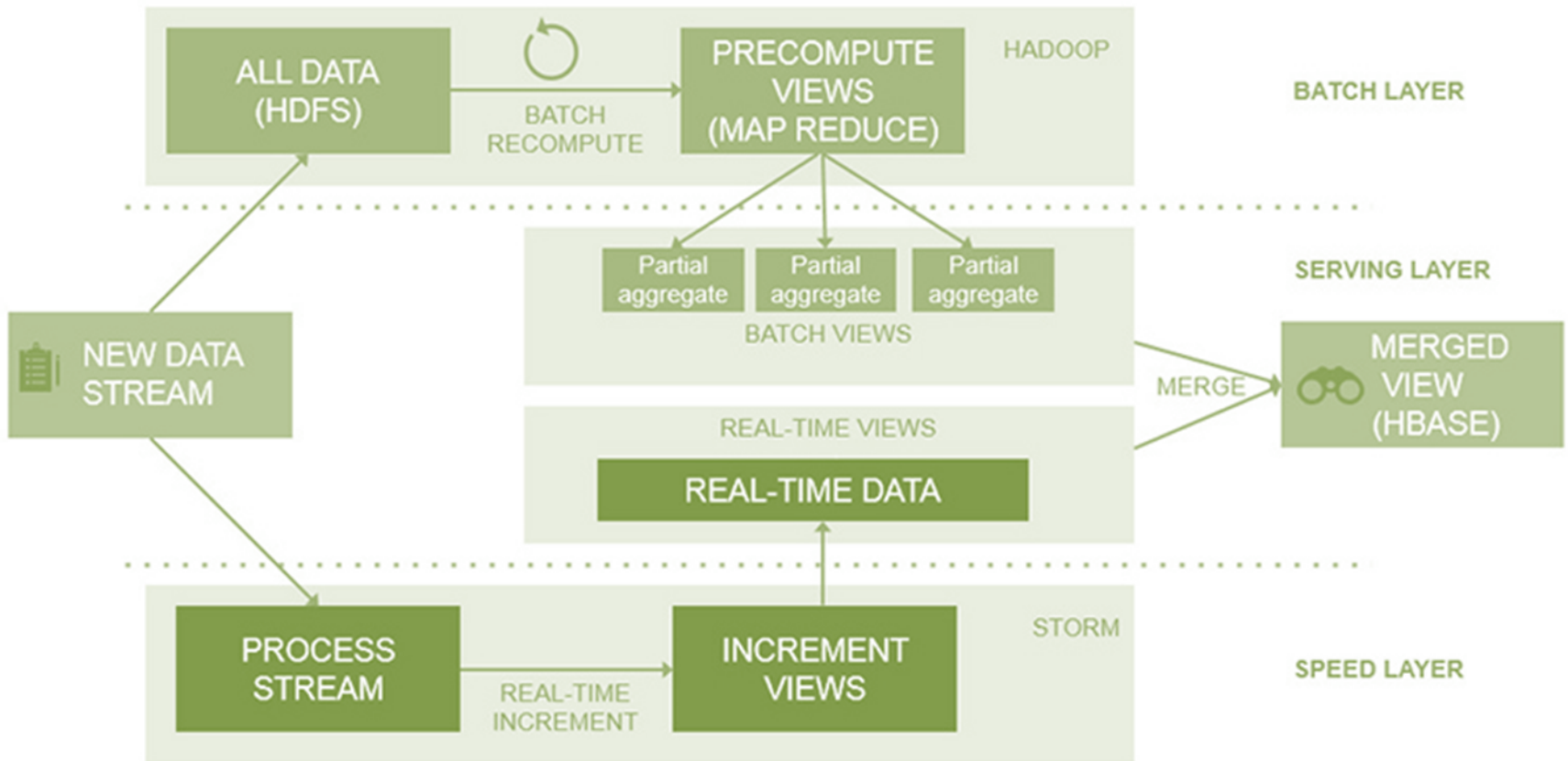
MAPREDUCE formas primitivas (Batch)

Kappa Architecture



MAPREDUCE formas avanzadas (real time)

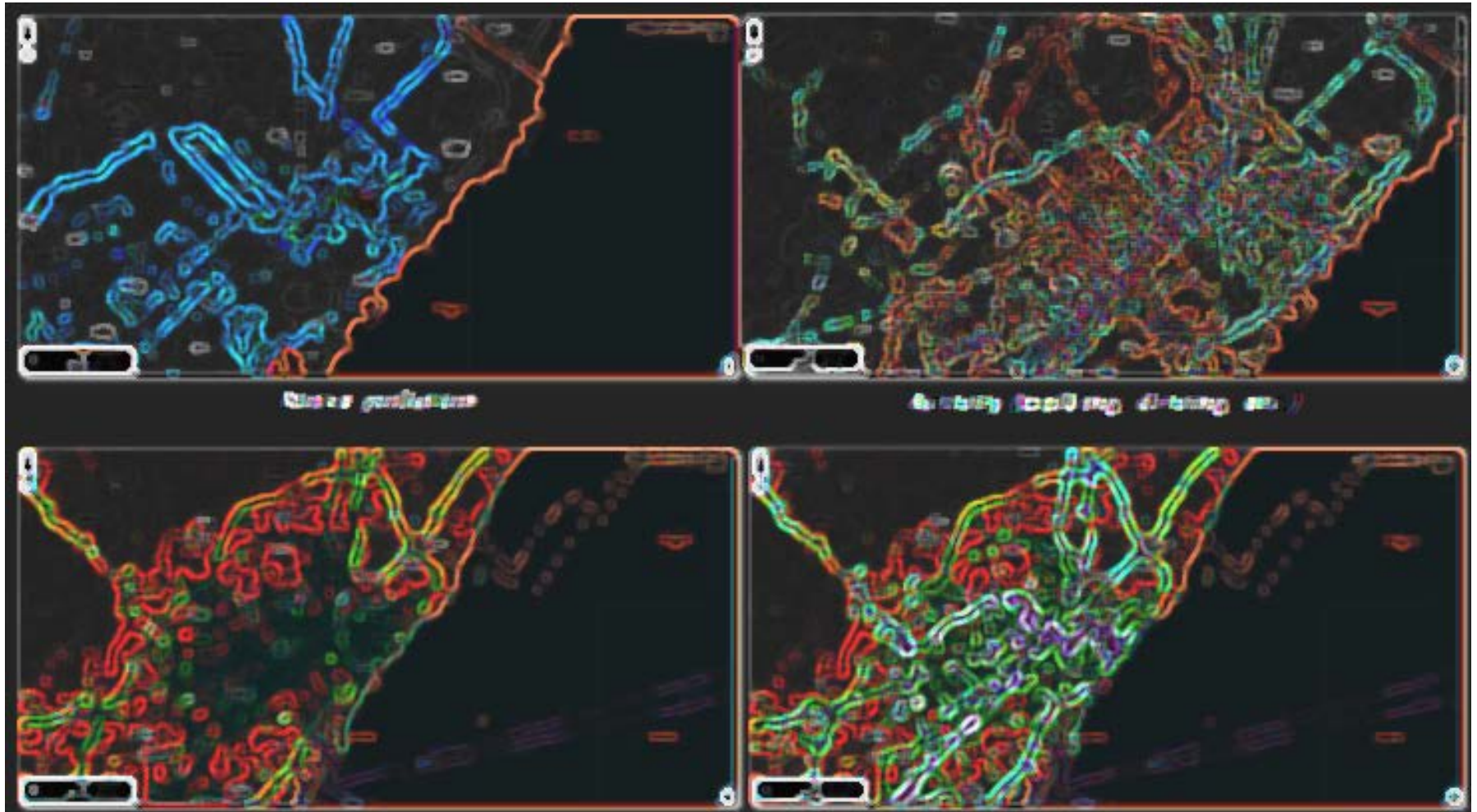
Lambda Architecture



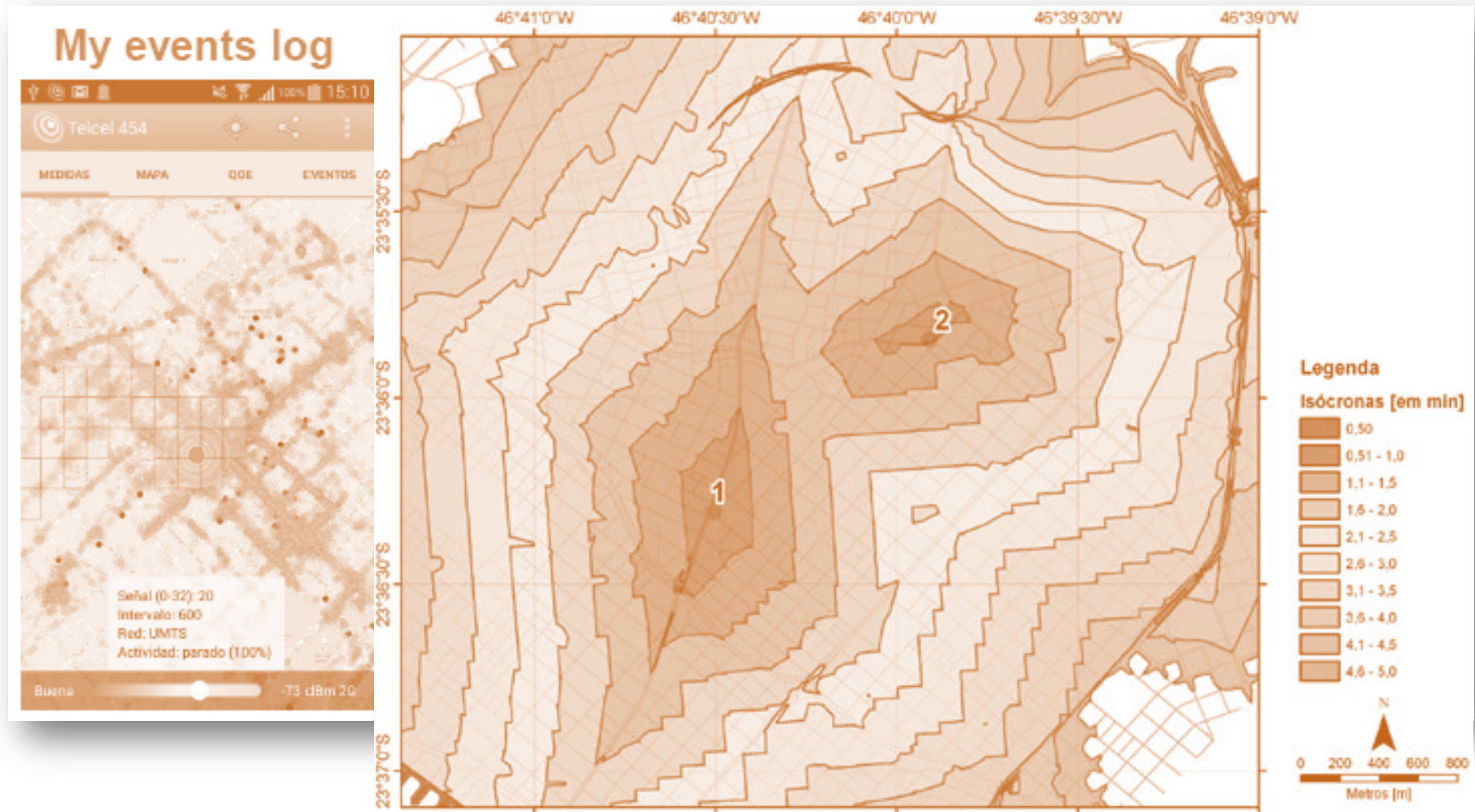
Aplicaciones: Análisis Espectral



Aplicaciones: Análisis Crowdsence



Aplicaciones: Distribución empírica de aversión al riesgo



Aplicaciones: Aggro Big Data Base

Combinación de datos no estructurados

Crop data + predicciones de cambio climático + genetic marker data + satellite + social media

Objetivo, crop yield distribution inference conditional to unpredictable (weather) before only proxied

Nuevas técnicas estadísticas necesarias porque tenemos FAT data ($K \gg N$) Lasso, Bayesian Shrinkage, PCA, ML etc

[Joshua Woodard](#), (2016) "Big data and Ag-Analytics: An open source, open data platform for agricultural & environmental finance, insurance, and risk", Agricultural Finance Review, Vol. 76 Issue: 1, pp.15-26, <https://doi.org/10.1108/AFR-03-2016-0018>

Aplicaciones: NPL & Sentimiento

Row ID	S Title	S Author	 D attitude
Row0	...if you can afford it	- Grandpa284	"... 0.241
Row1	2 michelin stars well deserved	- hoggat	"2... 0.133
Row2	A Michelin Star experience maybe?	- 4QueenMum2	"... 0.3
Row3	A Unique Place to Visit	- HughRogersSon	"... 0.2
Row4	A fun place to have dinner with friends in Berlin	- ParisGlenn	"... 0.3
Row5	A great Thai Experience	- Scottish_Lion	"... 0.267
Row6	A great disappointment: we made a reservation f...	- alelovesmiami	"... -0
Row7	A great experience	- Lillehopp	"... 0.2
Row8	A great neighbourhood restaurant	- Ryan3	"... 0.381
Row9	A little Thai Gem in Berlin	- xaxalala	"... 0.167
Row10	A nice surprise	- stabreim	"... 0.179
Row11	A piece of old Berlin	Eric W	"... 0.2
Row12	A tourist trap, but near to east side gallery	- CeceVenezia	"... 0
Row13	AMAZING BURGERS AND CHEESECAKE!!	- SAVB1	"... 0.1
Row14	AWESOME DISCOVERY IN BERLIN	- Favina80	"... 0.3
Row15	Affordable and yet so delicious	- ShawnKuek	"... 0.2
Row16	Amazing Asian Food!	- mdcwilson	"... 0.133
Row17	Amazing Thai food even better presentation!	- kkw142	"... 0.167

Aplicaciones: NPL & Sentimiento

[Click → Ejemplo de Global Topic Conversation “Insurance & Peru”](#)

condescend
crave light complaint
critic lover queer knife comfort
cute low rave suggest pricei importantli classic
despit loyal refresh superb enjoi strong posit import chao
doubt manic refus sweet expans tire effici stand poor highlight care
downsid memor relax tantal fat top terribl stun easili speedi polit happili break
move rid tast filthi treat true open sure slow decent sound pleas glad
ridicul tender fuss want ultim pretti well fantast star reserv clearli smart gem
grill although uniqu simpl cheap just long disappoint spot reason clear
crowd brilliant love worth good friendli even definit real
thrive hard clean down right littl cut problem prepar charm
fine super back nice great best excel valu
happi will recommend busi delici like fun ye noisi perfectli attent
truli heart easi hot fresh amaz awesom wonder need miss wrong
unabl hope excit impress friend understand incred help interest perfect special might mind wonderfulli
nonsens rude unfortun lack hungri larg lost live wide sluggish plain frozen
entertain normal sadli unpretenti least kind limit whatev slack phenomen free
abov especi odd scarc vibrant less welcom sickli paradis fiend beauti
adequ fair okai seriou vile shake overwhelmingli fell basic
advanc fatti origin serious overbear fault attract

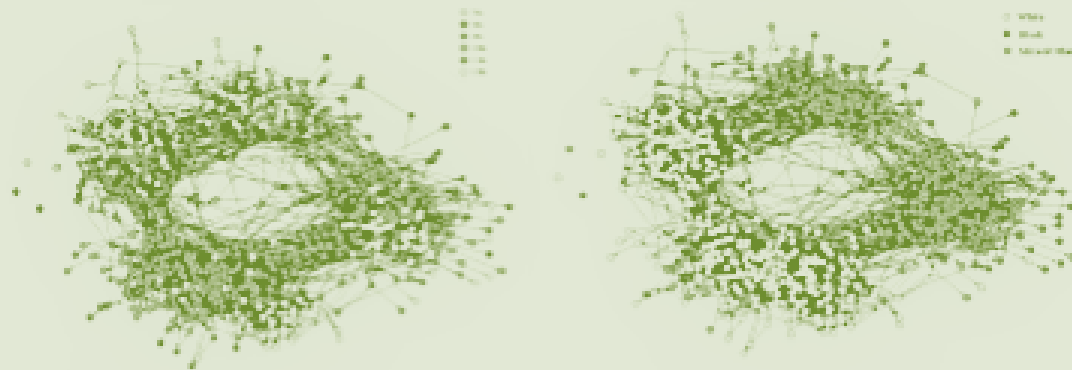
Aplicaciones: Segmentación de colectivos afines



Segmentación vs Pooling

Cluster, Segmentation and (Social) Networks

Homophily is the tendency of individuals to associate and bond with similar others, “birds of a feather flock together”



from Moody (2001) Race, School Integration and Friendship Segregation in America

Homophilia



BUSINESS

CULTURE

MARKETS

FUTURE

STARTUPS



Facebook friends could change your credit score

by [Katie Lobosco](#) @KatieLobosco

April 27, 2013 11:04 AM ET



DIGITAL TRENDS



[Home](#) > [Money](#) > Banks may soon scan Facebook and cell records to...

BANKS MAY SOON SCAN FACEBOOK AND CALL RECORDS TO SEE IF YOU DESERVE A LOAN

By [Kyle Wiggers](#) - Posted on May 7, 2013 2:04 pm

Claves en el sector

- La evaluación y gestión de riesgos e incertidumbres es clave para una toma de decisiones óptima en la producción agrícola.
- La creciente variabilidad del clima está acelerando aún más la incertidumbre en todas las etapas de las cadenas de valor.
- Los recientes avances de Big Data pueden mejorar la gestión de riesgos en la producción agrícola. (Datos de alta resolución de satélites, plataformas de datos abiertas y datos de máquinas agrícolas)
- Una toma de decisiones basada en el riesgo que incorpore estos avances implicará un entendimiento común entre las empresas, los responsables políticos y profesionales de finanzas y seguros.



XV Congreso
internacional

ALASA 2018

Lima, Perú.

➤ **15 al 18 de mayo de 2018.**
The Westin Hotel
& Convention Center



Asociación Latinoamericana para
el desarrollo del Seguro Agropecuario

