

# **Gestió d'Infraestructures per al Processament de Dades**

## **Introducció (1)**

Remo Suppi.

Departament Arquitectura d'Ordinadors i Sistemes Operatius

UAB (Remo.Suppi@uab.cat)

# Quins coneixement teniu?

SO i Linux

Programació, C i Python, OO, estructures de dades, aplicacions paral·leles (MPI, OpenMP, OpenAcc, Python)

Xarxes: TCP/IP, ICMP, adreçament, *routing*, dhcp, dns, ...

Bases de Dades (SQL i NoSQL)

Rendiment del codi/Anàlisis de prestacions.

Codificació, grafs, processament senyal, imatge, vídeo.

I a més a més en 3<sup>r</sup> tindreu:  
Optimització  
Computació Cloud i Aplicacions de Dades Massives  
Criptografia i seguretat  
Visualització de dades  
Xarxes Neuronal i DL  
Aprendentatge Computacional

SO i Linux

Programar, C i Python, OO, estructures de dades, aplicacions paral·leles (MPI, OpenMP, OpenCL, Python)

Xarxes: TCP/IP, ICMP, adreces IP, dns, ...

Bases de Dades

Rendiment del sistema

Codificació, Gràfics, processament de vídeo

Necessitem  
Infraestructura  
i saber gestionar-la

Quins  
coneixement  
teniu?

Optimització  
Cloud i Dades Massives  
Criptografia i seguretat  
Visualització de dades  
Xarxes Neuronal i DL  
Aprendentatge Computacional

FEBRUARY 27TH-MARCH 5TH 2010

Economist.com

Obama the warrior

Misgoverning Argentina

The economic shift from West to East

Genetically modified crops blossom

The right to eat cats and dogs

# The data deluge

AND HOW TO HANDLE IT: A 14-PAGE SPECIAL REPORT



2010

Què  
veurem?

**Canvi de paradigma: de dg a bd**

**Infraestructura: On?**

**Emmagatzemament: cóm?**

**Programari: Eines**

**Estratègies pel processament.**

# Canvi de paradigma: dades grans a Big Data

“Why big data  
is a big deal”

*InfoWorld* – 9/1/11

“Keeping Afloat  
in a Sea of 'Big  
Data”

*ITBusinessEdge* – 9/6/11

“The challenge–  
and opportunity–  
of big data”

*McKinsey Quarterly* – 5/11

“The promise of  
Big Data”

*Intelligent Utility* – 8/28/11

“What is big  
data  
analytics?”

*InfoWorld* – 2019

“Big data meets Big  
Brother as China  
moves to rate its  
citizens”

*Wired* – 2018

“Big Data  
Suggests Big  
Potential for  
Urban Farming”

*Wired* – 2018

“Data sheet:  
Advanced analytics”

*McKinsey Quarterly* – 2018

“2018 Most Wired  
Hospitals Embrace  
Big Data Analytics”

*Healthitanalytics* – 2018

“Google has  
access to  
detailed health  
records on tens  
of millions of  
Americans”

*Ars-Technica* 2019

2011 ... 2018 ...

# De Big Data a la IA

Artificial Intelligence and Cardiologists Join Forces to Predict Atrial Fibrillation (+)

Artificial Intelligence Enhances Tongue Diagnostic Systems(+)

AI, machine learning top health CIO priorities in 2023, survey finds (+)

Can Artificial Intelligence Coach You to Better Health?(+)

Artificial Intelligence in Medical Diagnostics Market Worth \$3.7 billion (+)

Latest Artificial Intelligence News (Forbes)  
(+)

Can Artificial Intelligence Coach You to Better Health?  
(+)

WHO says AI can transform health care if understood properly  
(+)

(+)Can artificial intelligence transform public health nutrition?

# 2023...

# Context del Big Data

Home a la lluna: ordenador de missió 32KB (1969), un TE té 64 GB (2023): en 50 anys l'increment és de **2.097.152 vegades.**

270 PB en un mes (283.000.000 GB), 40.000 consultes per segon, 3.5 mil milions de cerques per dia.

*January 2023: 5.16 billion internet users worldwide, which is 64.4 percent of the global population. Of this total, 4.76 billion, or 59.4 percent of the world's population, were social media users. (Statista)*

# Google

<https://influencermarketinghub.com/social-media-statistics/>



# Context del Big Data

2,5 exabytes/dia

1 exaB=  $10^{18}$ bytes =

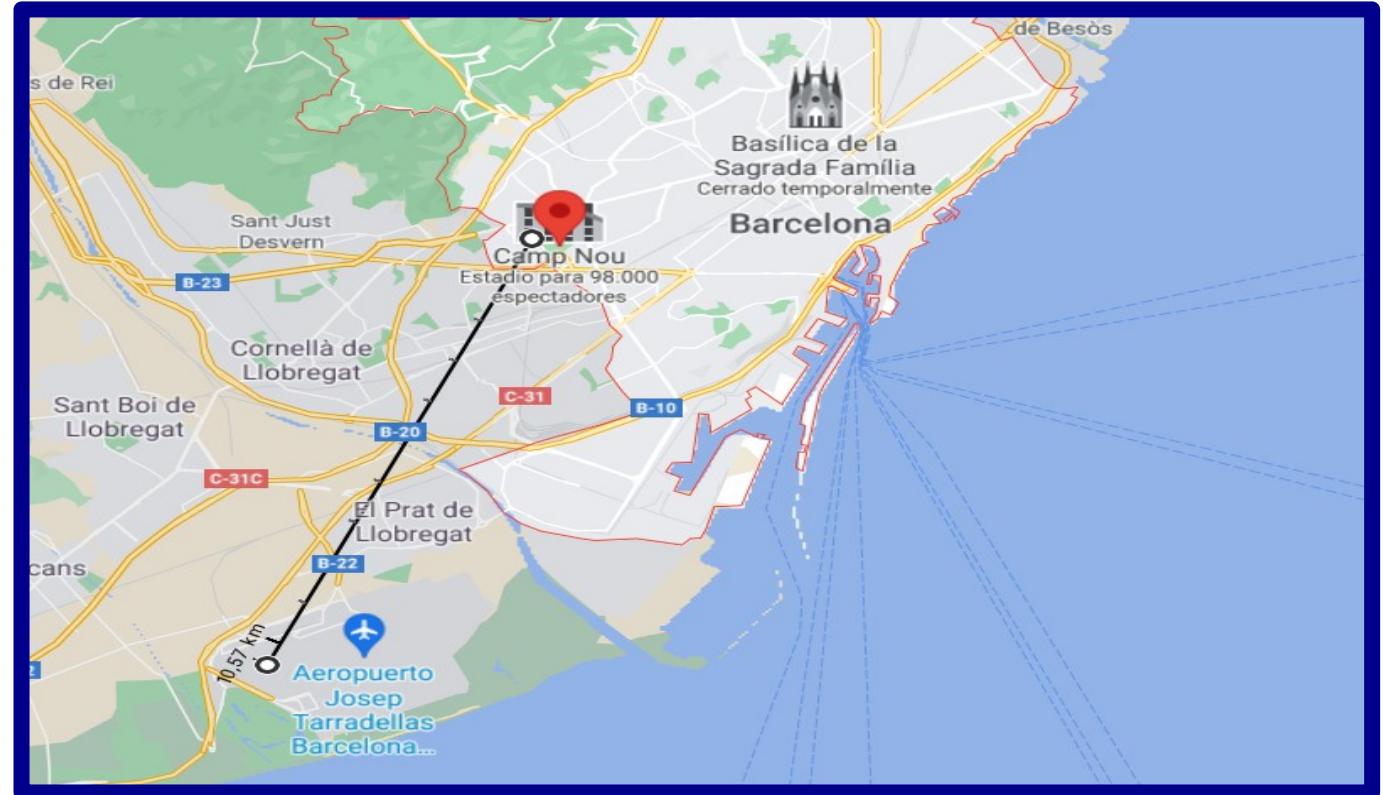
1 milió de terabytes

I el futur?

2024-5: 44 zettabytes

(1zetaB=  $10^{21}$ bytes =

1.000 milions de terabytes)



# Canvi de paradigma del BD

Anàlisis amb  
TOTES les dades

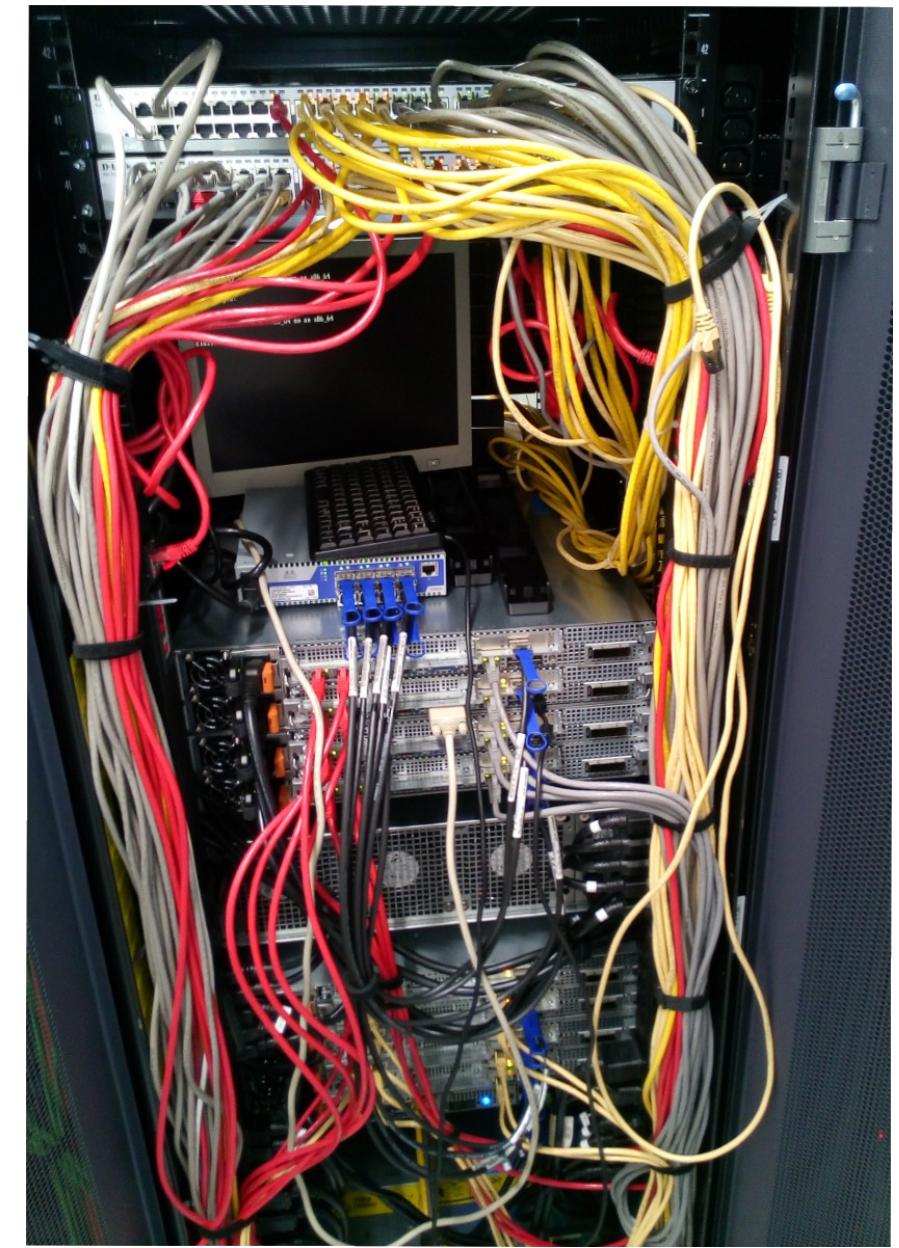
Anàlisis amb  
un conjunt de  
dades

610,377094	174,065664	187,343724	18,1275388	111,34813	911,828123	682,63981	235,579986	945,308139	903,886763	321,717985	740,739659	849,556901	522,031213	386,395272	887,831571	353,685399	769,841497
994,860923	798,428998	641,403092	963,071481	492,90734	577,823839	348,622665	373,520994	727,613451	914,260947	947,221386	392,945422	807,987774	425,553011	933,094725	865,509257	724,252832	979,23015
609,421871	510,160789	941,818439	415,968918	49,3042769	658,569688	719,678054	176,139555	343,826779	511,594249	117,271706	32,9902663	392,63532	544,222097	456,657699	395,529524	328,931355	749,395607
194,292226	494,55672	729,053162	263,805261	17,790357	554,202099	812,062099	113,006311	78,3274908	854,214361	168,703088	718,276065	77,3961937	776,642686	991,5103	353,264617	787,298433	971,929563
730,125198	729,112181	329,422045	819,528591	281,721861	890,975753	688,525234	591,617002	0,5116922	697,035359	475,066674	711,021806	973,397271	721,853593	55,2336648	305,984155	319,443933	145,750589
987,782501	42,247856	836,322241	703,229421	328,884669	925,163199	104,105808	323,208114	31,8409947	794,512526	992,933926	86,7564313	213,396439	941,163199	641,472417	430,281269	612,939107	534,572423
706,755371	445,249257	410,40261	186,596785	324,029093	216,539795	103,12409	79,2276057	465,523942	300,097287	262,840228	739,971325	851,577548	126,970662	90,9819849	851,207317	566,489528	673,760815
964,917678	908,14883	572,903739	666,825665	561,396526	334,112736	573,378834	198,551653	587,548356	400,540549	373,891808	143,061218	606,039387	246,763878	736,129159	695,122881	741,063211	327,346693
304,703417	307,000549	288,853517	653,87108	406,882354	224,454612	691,566009	744,536855	9,8186476	242,05731	58,07362	712,927601	851,405646	12,3164731	259,707194	432,45682	730,301896	
2,4785238	716,722303	270,656544	935,007173	293,8218	49,5355026	950,782654	647,446799	46,683385	147,12698	552,65769	31,0870393	652,540513	701,075237	283,260836	242,985197	74,9830834	493,751482
659,48579	820,151259	608,331581	966,468395	520,171416	867,834933	775,697501	325,869234	168,115958	343,805802	336,63234	158,122009	821,698635	586,968425	739,139401	900,890383	501,30543	481,135478
254,648727	313,358453	366,527161	857,381108	446,208982	623,334203	587,397349	446,37238	489,614159	521,840769	121,1948	482,212729	6,68582876	595,078461	220,212161			
357,752912	412,37946	137,585041	36,5148477	267,310317	648,03163	908,771788	285,026038	534,621079	277,112324	676,269081	617,817635	362,544615	442,199311	344,14814	417,542618	399,623225	747,8205
156,45739	775,6155132	620,822796	477,14134	409,312205	655,632776	789,955987	3,14077856	103,067448	641,527159	531,409238	831,203027	327,593924	322,09879	665,79516	642,166388	727,835036	7,9,059132
350,149635	627,7914513	592,706565	314,86355	816,707255	593,554535	360,691557	889,802534	204,325027	218,520501	30,116185	822,206299	461,720602	912,453676	389,08138	320,50804	187,53949	94,081179
928,587249	382,256979	604,121018	477,666133	726,854189	497,245352	635,620982	818,511703	852,181425	93,1349558	633,774476	563,601613	318,704752	67,7610194	90,075237	77,7610194	1,4,866,832733	
220,530012	861,329106	400,756869	600,23128	389,479633	43,801	825,51355	695,460162	462,402743	802,153545	311,678531	201,817729	621,668005	122,973423	553,094088	324,928054	487,835742	92,3,656687
675,217291	630,388494	969,911139	725,6865	23,062099	70,61642	85,61642	64,61642	71,61642	99,61642	77,61642	84,61642	57,61642	56,61642	53,61642	52,61642	51,61642	47,2,970811
404,315015	415,056334	600,868516	511,527161	23,712,562	33,819,8	8,61642	8,61642	8,61642	8,61642	8,61642	8,61642	8,61642	8,61642	8,61642	8,61642	8,61642	323,29431
749,973272	719,170778	389,147726	352,856,655	844,444,988	317,905208	730,61642	730,61642	730,61642	730,61642	730,61642	730,61642	730,61642	730,61642	730,61642	730,61642	730,61642	730,61642
817,629193	972,886062	969,399284	119,291053	275,399491	245,010469	890,46514	680,285359	736,182181	813,116947	383,83926	656,604419	618,164852	497,111228	377,04387	137,171852	230,490855	
715,094845	233,356284	258,045837	926,534914	112,137239	804,89951	52,3652196	451,476386	512,071872	63,6534579	451,602829	926,856924	420,66387	646,18989	447,324882	623,814981	99,868043	172,90853
275,479995	865,435733	523,043249	92,522,099	440,847178	661,984,008	551,56805	17,798996	436,547784	472,175329	5,249573	864,171572	230,000981	657,404718	251,268117	315,352666	271,732663	889,975611
961,415916	391,683343	273,440927	134,652,5	5,95,77	7,94,246	4,95,77	5,95,77	4,95,77	5,95,77	4,95,77	4,95,77	4,95,77	4,95,77	4,95,77	4,95,77	4,95,77	
735,412008	703,808456	313,360098	536,61,6	5,0,498,	27,59,2	45,89046	33,83,62	28,83,62	27,83,62	27,83,62	27,83,62	27,83,62	27,83,62	27,83,62	27,83,62	27,83,62	
235,678154	508,913692	215,935308	856,452,7	5,9,52	6,0,9,52	7,0,9,52	7,0,9,52	7,0,9,52	7,0,9,52	7,0,9,52	7,0,9,52	7,0,9,52	7,0,9,52	7,0,9,52	7,0,9,52	7,0,9,52	
113,483044	144,559945	843,840822	881,269421	817,986365	281,551413	820,575721	72,8485467	210,10172	724,661741	877,787125	117,15071	877,676088	641,129773	893,350449	704,024519	256,799392	354,744039
851,318868	579,337191	619,601073	76,826749	854,031014	849,456726	364,896655	219,321964	846,524821	273,053454	490,489716	708,364563	734,259649	147,635407	630,148427	302,649934	461,560534	877,367405
590,200017	340,764759	918,351727	602,765614	409,46055	594,237003	99,0892692	365,450571	228,77276	138,489105	733,301884	735,144892	379,380174	888,653145	766,48741	891,497105	782,124396	594,425645
336,982041	152,326177	331,208941	97,238311	584,666947	880,501649	580,550082	459,996242	151,270901	832,85447	80,0463667	135,686033	337,69231	325,015224	159,659608	989,542502	475,728937	166,587605
123,492898	344,54809	445,080405	102,839097	844,813933	34,65,058075	843,495401	443,017152	866,285258	680,284385	606,877959	824,453812	285,326613	899,454477	5,50677016	750,774163	731,410503	
362,055302	751,153941	124,852671	55,2117	965,832403	955,237056	811,918583	920,637356	293,877608	436,80413	366,583773	93,950326	891,781744	460,733622	541,418389	163,385823	128,623343	786,172588
843,117358	778,976374	260,308792	491,991438	485,481499	230,171557	169,989758	29,8402202	747,017149	954,723009	942,950782	241,766585	427,556579	953,700004	202,342059	986,840166	841,086386	199,82016
558,178008	197,6287391	197,505144	704,614228	494,737528	482,018183	167,058691	207,504045	802,529236	110,586768	372,184213	740,342961	376,664572	560,386777	794,076393	127,650223	757,109203	360,938979
95,0170271	637,404619	826,016634	672,65658	605,365316	908,362417	150,946111	822,080149	719,705189	370,043737	824,655201	236,272715	508,026676	731,851986	757,746408	163,758985	49,5959518	374,948517
829,413754	931,720538	901,693248	139,391897	266,780486	612,406314	501,586487	312,792938	722,396287	366,088906	870,942048	644,651299	925,400687	108,084666	934,452977	876,576508	694,869891	970,294655
471,914819	483,868003	811,606351	479,737029	177,490995	475,518806	757,147713	483,364174	499,471256	607,812692	604,331083	383,611968	598,687017	560,426711	655,425898	486,342961	407,153555	
54,0041656	963,168738	430,271179	243,217474	808,65397	201,662386	30,352854	419,044654	459,403844	306,836827	41,411403	469,726155	704,616466	491,307455	49,801768	561,147983	377,353073	
342,679163	764,337721	439,59878	430,156484	978,271168	781,330722	669,565988	822,356967	116,10874	658,502111	291,246634	487,461618	922,561652	279,096487	930,917992	171,721863	939,343729	995,786676
306,474278	623,41088	176,134425	327,353576	117,320483	980,034949	771,463417	167,914999	109,848503	491,848293	482,565342	8						

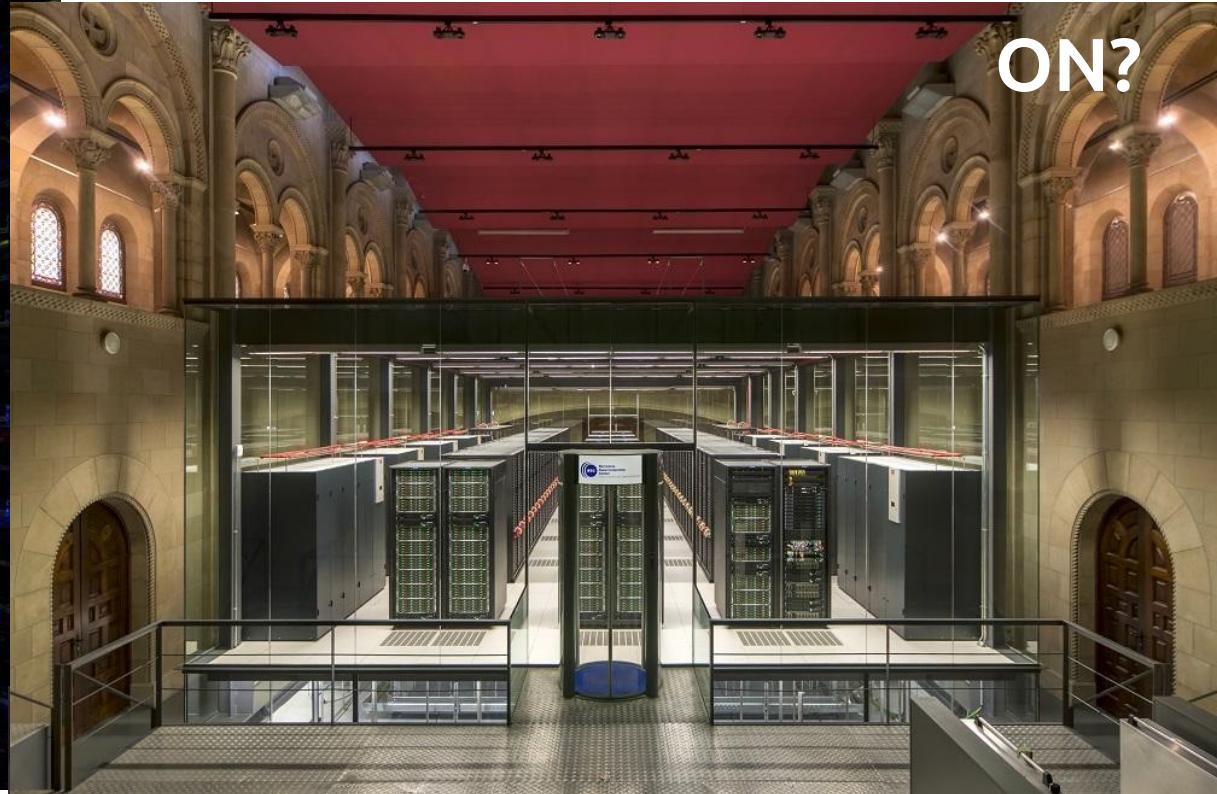
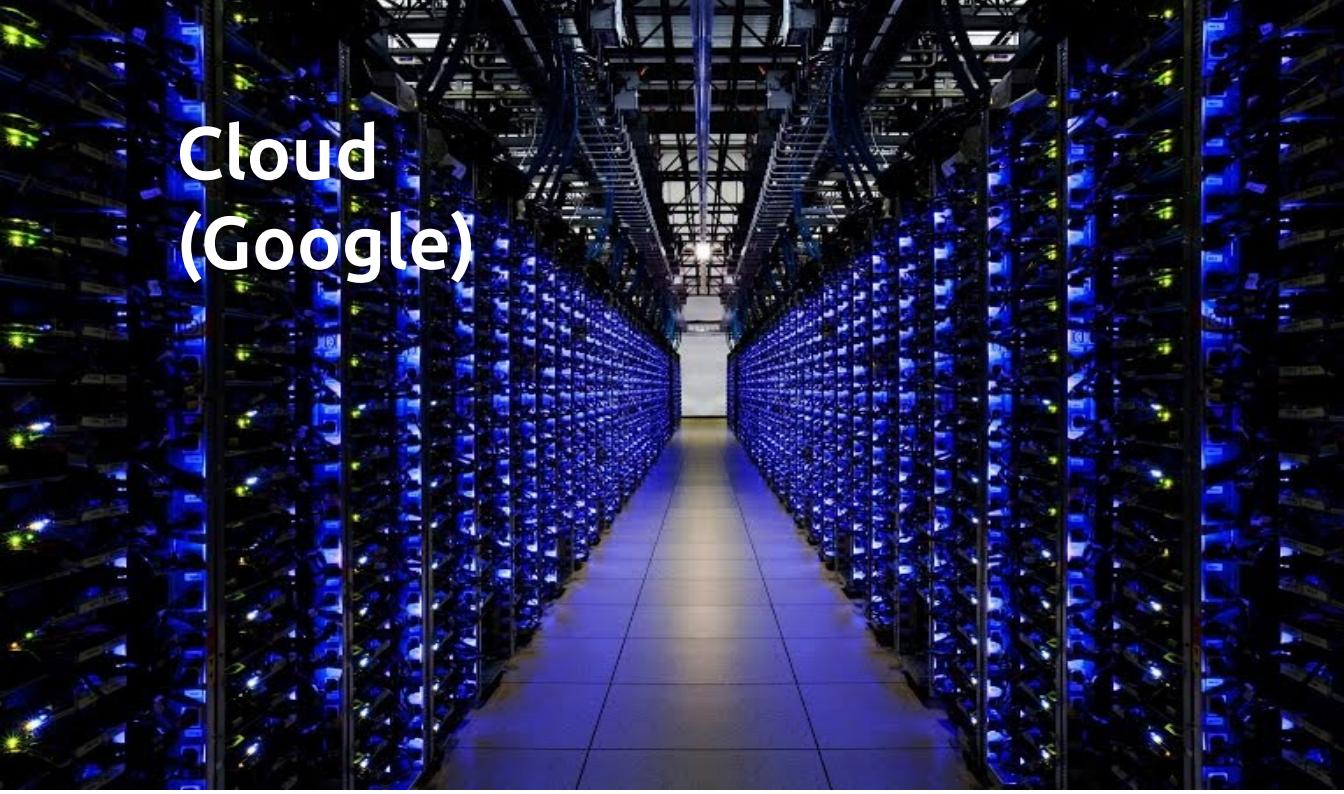
# Clústers



ON?



# Cloud (Google)



HP Desktop PC AMD  
Ryzen 7 4700G 8-cores  
Radeon 64GB RAM 3TB  
NVMe SSD (1.000\$)

ASUS ROG Strix Laptop, AMD Ryzen 9 5900HX 8 Cores,AMD  
Radeon RX 6800M 12G, 32GB RAM, 2TB SSD) 2.200\$

<https://www.google.com/about/datacenters/gallery/>  
[https://www.youtube.com/watch?v=Ik4GwE\\_gXKc](https://www.youtube.com/watch?v=Ik4GwE_gXKc)



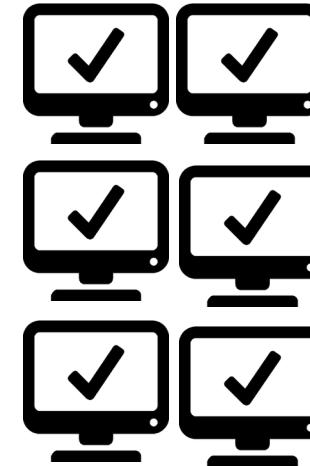
ON?

# Canvi de paradigma del BD



# centralitzat

410,40261	186,596785	324,029093	216,539793
572,903739	666,825655	561,396526	334,112736
288,855317	926,825644	653,87108	406,882354
270,656544	935,007173	293,8218	49,535502
608,331581	966,468395	520,171416	867,834933
336,552715	867,381108	381,833109	446,209828
137,585041	36,5418477	267,310317	648,038163
620,822796	477,14134	409,312205	655,632776
592,706568	314,86355	816,707255	593,554353
604,121018	477,666133	726,854189	497,245352
400,756869	600,231236	389,479633	143,878161
969,911139	725,688126	420,72533	572,098843
600,8685816	511,529293	236,720521	733,888988
389,147726	352,87965	188,44524	866,388044
369,339926	352,91053	275,399491	245,010469



# distribuït

410.40261	165.984785	324.02903	216.53979
572.903739	666.025642	561.385625	331.12736
288.853517	926.825644	653.817038	408.88254
270.658144	835.007173	793.82182	500.88254
608.158481	750.817038	631.82182	813.84062
396.527161	36.5418477	881.835109	446.20891
137.585041	36.5418477	267.310317	408.06363
572.706568	314.835586	815.707256	503.63776
604.121018	477.661133	782.85419	497.24552
400.754689	603.512316	479.82182	431.87611
949.371222	479.82182	479.82182	479.82182
600.685816	522.879635	276.20521	733.89888
389.147726	352.879635	188.445524	661.38045
969.398284	19.291205	275.39941	245.10499

410_40261	186_598785	324_020903	116_535975
572_903561	186_598785	324_020903	116_535975
228_853511	126_825444	635_871308	406_882544
270_655444	035_007173	793_8218	493_553626
608_331586	966_446389	530_171401	867_834383
394_621586	966_446389	530_171401	867_834383
137_585041	561_841877	267_310317	648_031863
620_822796	471_14134	409_312205	655_632776
592_705658	616_705658	616_705255	592_545354
634_705658	616_705658	616_705255	592_545354
400_754869	600_233216	380_479635	413_876116
949_911139	628_611216	420_72535	672_09843
369_147726	611_511126	373_73981	733_09843
369_147726	611_511126	373_73981	733_09843
969_399268	119_291053	275_399491	245_104069

415.46298	186.59678	324.02909	245.39397
872.00737	862.59565	561.38635	331.11776
288.053517	928.82544	653.87130	408.82534
270.65544	935.00717	793.8218	493.55362
308.13518	966.46389	530.17146	807.84384
394.11518	935.00717	793.8218	493.55362
137.585041	561.38477	267.31037	648.03163
620.22796	477.14139	409.31295	655.63277
592.70638	816.74139	816.70278	816.70278
604.10149	816.70278	816.70278	816.70278
400.745869	816.70278	816.70278	816.70278
704.574869	600.233216	589.47963	433.87611
949.91139	726.61126	420.72533	572.08433
369.14772	532.87695	188.55524	245.01049
969.39926	719.21053	275.39491	245.01049

416,49261 186,59785 324,07938 216,53979  
 572,90179 666,02565 561,86266 334,11278  
 288,85157 426,82544 653,87108 482,88254  
 270,65654 93,03757 392,819 428,535526  
 603,331581 964,46385 520,17146 867,84393  
 36,527181 85,37181 381,33019 446,20982  
 137,280541 59,314877 267,310317 489,03163  
 55,270568 31,48385 812,30230 655,63278  
 55,270568 31,48385 812,30230 655,63278  
 604,12108 477,661332 726,854189 497,24532  
 400,75686 627,661332 389,746733 143,876811  
 969,911139 726,681126 420,72533 572,09843  
 605,860156 511,529283 236,20251 733,89888  
 189,47726 352,89765 188,44524 866,38045  
 969,399284 119,290153 275,39941 245,10469

410,40261 186,59768 324,029093 216,539795  
 572,903739 686,825681 561,395625 344,112736  
 288,855157 826,855157 561,87108 406,882354  
 270,656182 561,656182 493,823546 535,823546  
 366,561821 493,561821 493,823546 535,823546  
 366,561821 857,188110 883,188110 446,209882  
 137,585041 364,584477 267,310317 648,031631  
 620,227986 477,14134 409,223055 695,632776  
 592,707568 314,86383 618,707552 593,554353  
 604,121018 477,66133 132,66133 497,425352  
 404,796018 512,796018 512,796018 497,425352  
 949,91126 680,91126 429,72539 127,089449  
 600,865106 511,865106 236,202507 713,889889  
 389,147726 352,897653 188,445425 686,388405  
 989,399284 119,199284 275,399451 245,010469

410.40261 186.59678 324.02903 216.53979  
 75.30739 666.82569 561.39862 334.11276  
 288.55262 100.17146 200.17146 200.17146  
 270.45526 100.17146 200.17146 200.17146  
 608.33158 666.48395 501.17146 867.84393  
 366.52617 85.31011 381.33109 446.20982  
 137.58504 345.54177 267.31031 648.03163  
 620.82796 47.41731 301.62055 655.632776  
 592.70658 311.8355 816.70255 595.53277  
 100.17146 200.17146 200.17146 200.17146  
 400.75889 620.17146 200.17146 457.84527  
 969.11139 768.68126 420.72535 572.08943  
 600.86516 511.52333 236.72635 733.88998  
 384.14772 320.87655 188.44524 666.38048  
 969.39928 119.12035 273.79541 249.01048

## Canvi de paradigma del BD

Impossible analitzar/emmagatzemar amb tècniques i BBDD tradicionals.

Anàlisis de dades: saber *què està passant*

Descobrir nova informació que no es 'visible'

Anàlisis global en lloc de local

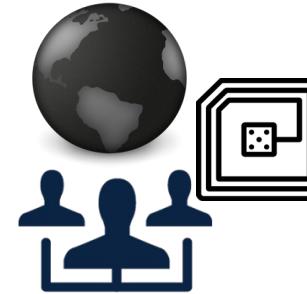
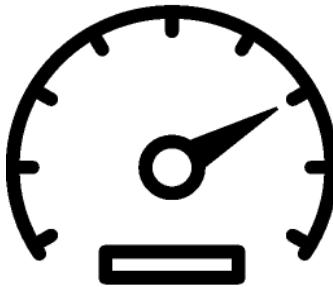
Ja no hi ha mostreig, tot el conjunt és vàlid

La inexactitud de les dades ja no és un problema

# On es fa servir el BD?

Entorn	'Noves' dades	Què és possible?
<b>Salut</b> Reduir el temps/No visites	Monitorització remota del pacient	Medicina preventiva, reducció hospitalització
<b>Manufactura</b> <i>Suport In person</i>	Sensors de producte	Diagnòstic automatitzat, suport
<b>Serveis basats en localització</b>	Dades en temps real	Trànsit, publicitat, cerques
<b>Sector públic</b>	Open data	Serveis personalitzats, reducció de despeses
<b>Ventes al detall</b>	Social media	<i>Sentiment analysis, micro-segmentació</i>

# Què caracteritza al BD?



5  
V

Volum

Velocitat

Varietat

Veracitat

Valor

# Impacte del BD?

## Volum

El 90% de les dades mundials ha estat creat només en els darrers dos anys.

El volum de dades creades només per empreses nord-americanes cada any és suficient per omplir deu mil biblioteques del Congrés o 10.000 sensors en l'ala d'A380.

Els cotxes connectades serà la nova plataforma digital més important, amb 220 milions en la dècada 20-30 (1,3M cotxes entren/surten al dia a BCN i cadascun inclou 100 sensors que generen dades cada 0,1 seg)

La majoria de les empreses només analitzen el 12% de les dades que en tenen.

## Velocitat

S'estima que pel període 2020-2025 hi haurà més de 50 milions de dispositius intel·ligents connectats al món, que recopilaran, analitzaran i compartiran informació.

## Varietat

Les dades errònies costen \$3.1 bilions anualment.

Les inversions en dades a la indústria de serveis financers representen gairebé \$9 mil milions

L'alt impacte de la Intel·ligència Artificial al màrqueting està creixent i es pronostica que arribarà gairebé a \$40 mil milions per a l'any 2025. IoT estalviarà els consumidors i les empreses \$1 bilió a l'any.

## Veracitat

Els salaris dels científics de dades han augmentat ràpidament en els darrers anys i s'estima que continuaran creixent a causa de la gran demanda de professionals de l'àrea.

## Per tant què és BD?

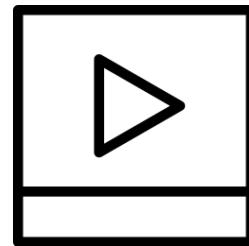
Big Data es refereix a conjunts de dades que creixen tant que és difícil capturar-los, emmagatzemar-los, processar-los, compartir-los, analitzar-los o visualitzar-los amb infraestructures (HW/SW) habituals.

¿Quan és Big?

**No és un nombre únic sinó un conjunt de paràmetres**

# És útil el BD?

**Decisions  
bassades  
en dades**



Vídeos, Social-  
Data, Documents,  
...

## Salut USA

Increments en la  
indústria/any

\$300 10<sup>9</sup>

## Manufactura

Decrement de les  
despeses de muntatge

-50%

## Dades Personals de Localització

Increment dels guanys

\$100 10<sup>9</sup>

## Sector Públic EU

Increment/any

€250 10<sup>9</sup>

## Vendes detall USA

Increments del marge  
després d'imposts

60+%

# És útil el BD?



Respondre a preguntes que abans no tenien resposta  
Formular noves preguntes i que sigui mol més àgil.  
Prendre decisions basades en evidències  
Democratitzar les dades  
Visualitzar el coneixement invisible

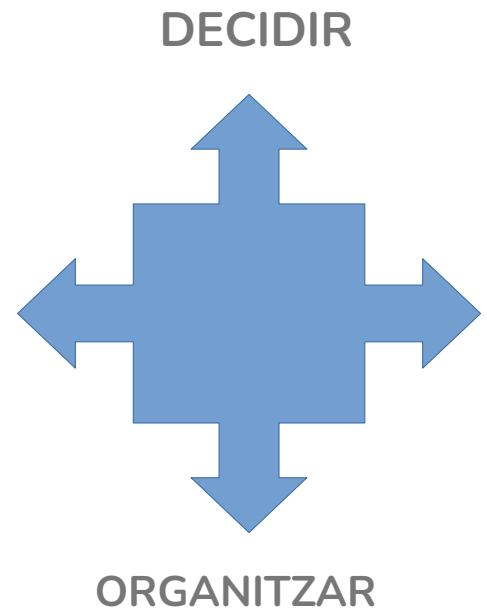


Big data: aquí i ara  
Mal ús de les dades  
Intrusió en la privacitat  
Ús no just de les dades

# És útil el BD?

Decidir en temps real sobre big data

Analitzar tots les seves dades ràpidament

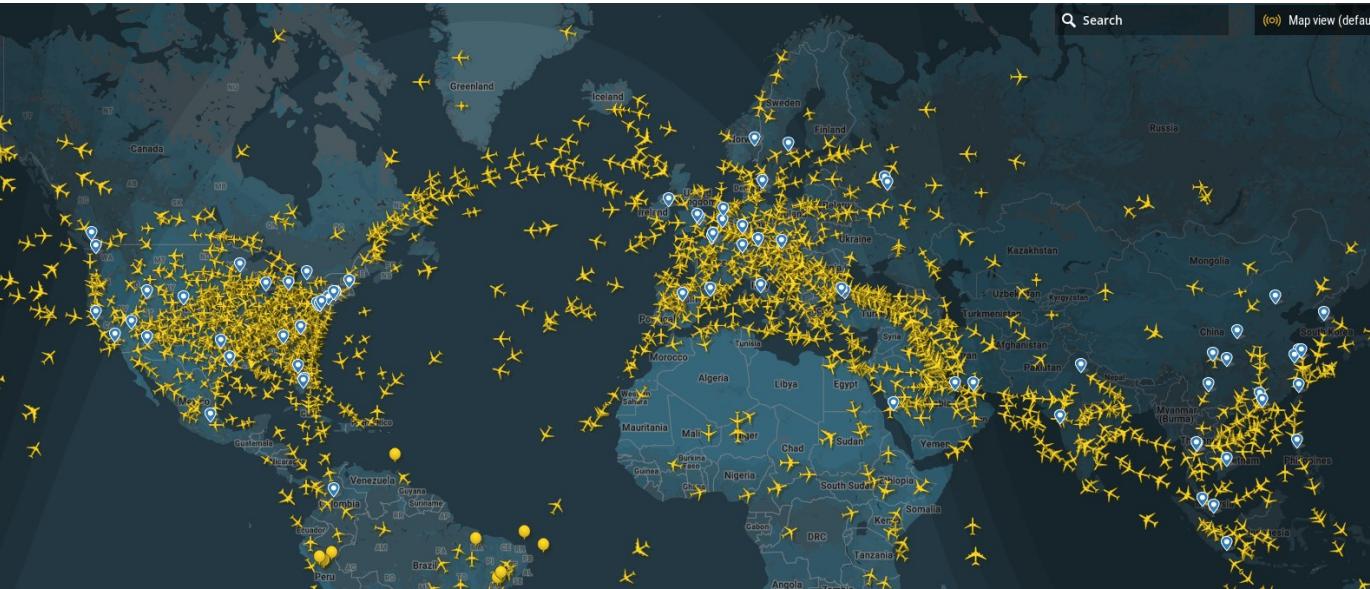


Adquisició:  
tots les dades  
disponibles

Organitzar i  
separar el big  
data fent servir  
parallelisme  
massiu

Prendre  
Millors  
Decisions  
Fent Servir  
Big Data

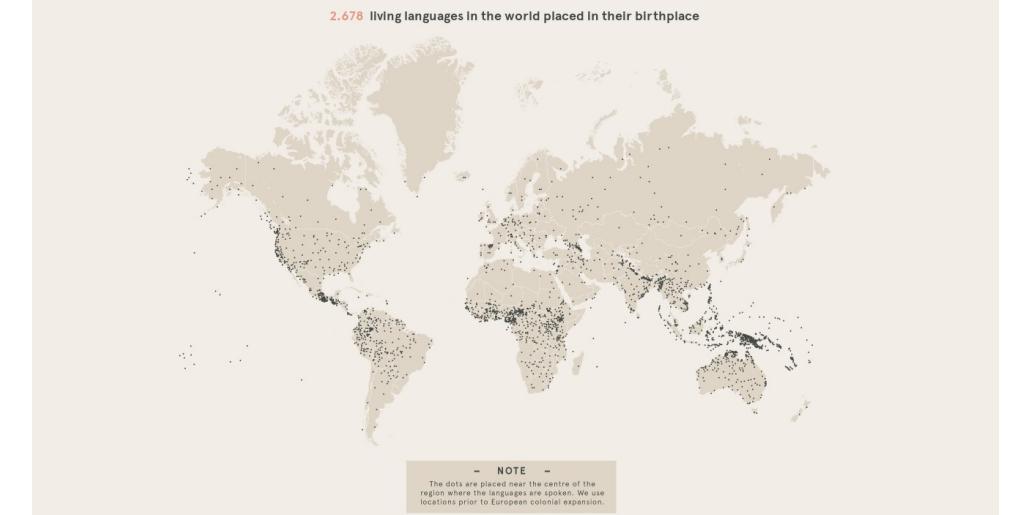
# Visualització del BD



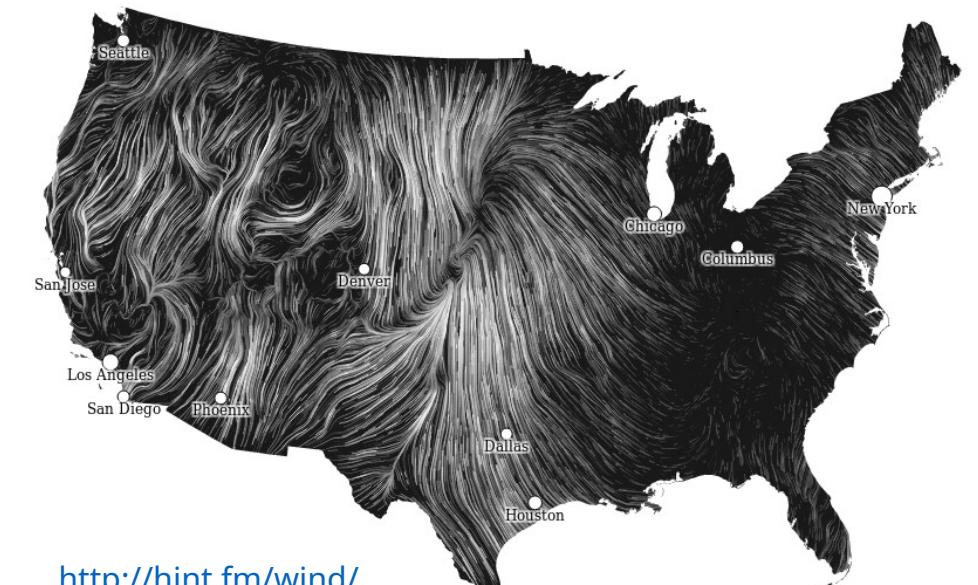
<https://www.flightradar24.com>



<https://cybermap.kaspersky.com/>



<http://www.puffpuffproject.com/languages.html>



<http://hint.fm/wind/>

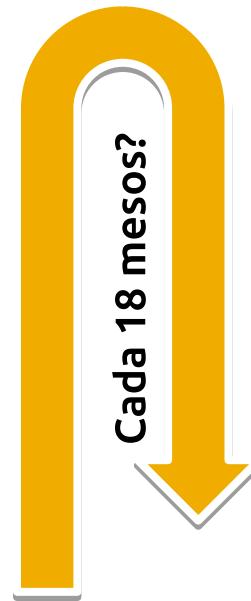
# On es processen les dades?

# Infraestructura

	Model de Còmput i Emmagatzemament	Característiques	Costs
<b>Mainframe</b>	Centralitzat	Aprofitar al màxim els recursos (cost alt)	Inversió i manteniment HW/SW
<b>Clúster</b>	Distribuït	PC i servidors distribuïts connectats per xarxes(primer local després Internet)	Inversió i manteniment HW/SW
<b>Cloud</b>	Centralitzat	Grans CPD amb recursos de baix cost (economia d'escala)	Models de pagament per ús
<b>Servidors</b>	Local/Remot	Propis o llogats (baix cost)	Inversió, reposició

# On es processen les dades?

## *Cloud Computing: canvi de model?*



### Model clàssic

- **Comprar & Propietat:**  
Hardware, Software, Aplicacions per satisfer les necessitats "punta".
- **Instal·lar, configurar, verificar, avaluar...**
- **Administrari**
- ..
- **Finalment:** Fer servir

### Infraestructura

### *Cloud Computing*

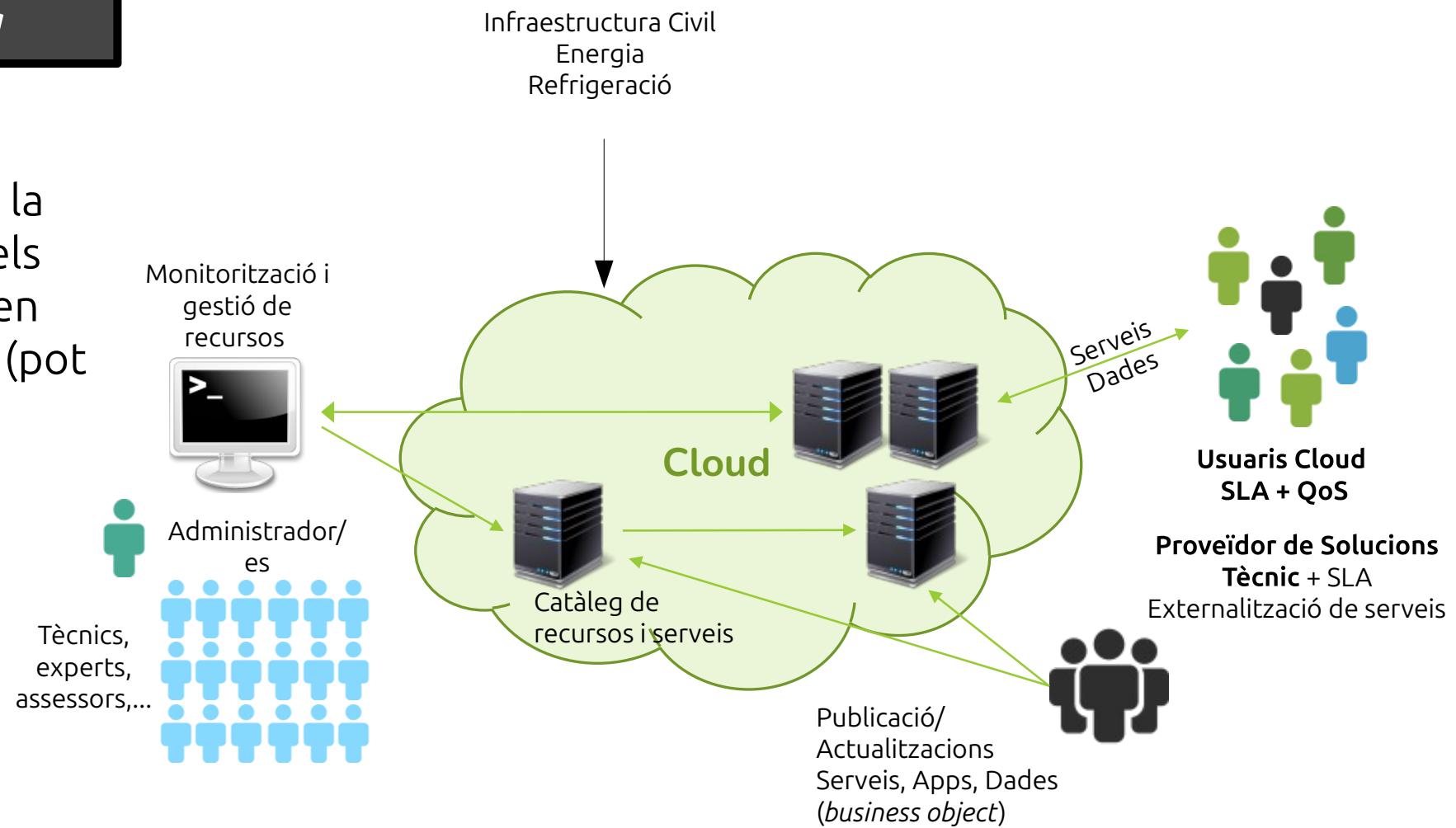
- Subscriure i contractar (SLA)
- Fer servir
- Pagar per ús i en base a una QoS.

# On es processen les dades?

# Infraestructura

## Cloud Computing

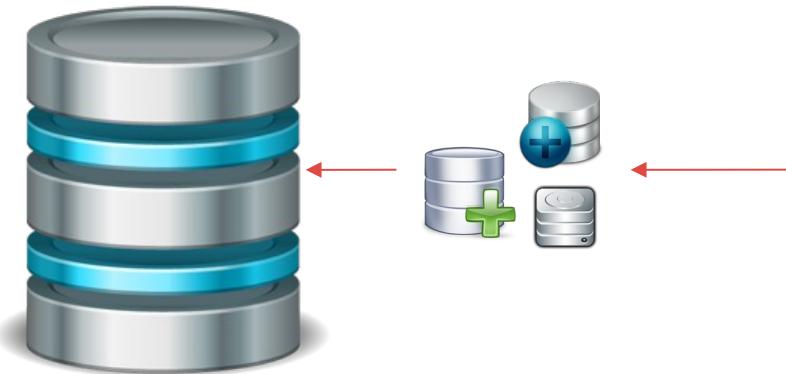
Arquitectura basada en la virtualització en el qual els clients (guests) s'executen sobre una màquina host (pot ser amb altres sistemes operatius) de forma concurrent.



# On es processen les dades?

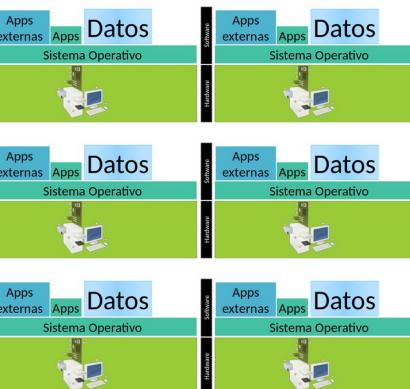


Local



CPD<sup>2</sup>

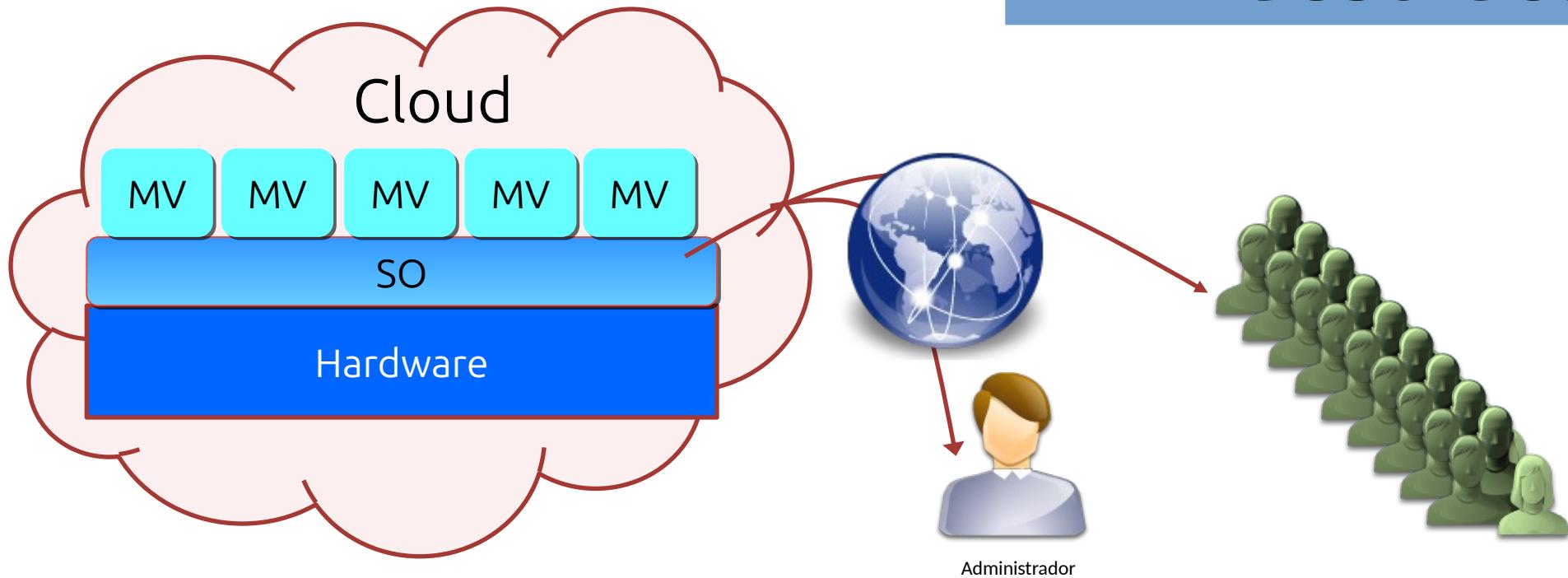
# Infraestructura



CPD

# On es processen les dades?

## Infraestructura

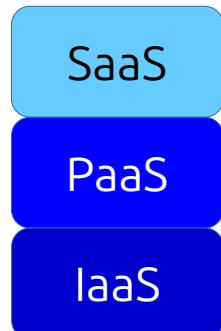


### Models de servei:

**Software com servei (SaaS):** fer servir les aplicacions del proveïdor que s'executen en una infraestructura cloud on es troben també les dades.

**Plataforma com servei (PaaS):** l'usuari disposa de tota la infraestructures per desenvolupar les seves aplicacions (p. ex. Eines, BD, llibreries, entorns, etc). NO es preocupa per la Infraestructura.

**Infraestructura com servei (IaaS):** l'usuari demana i desplega capacitat de processament, emmagatzematment, xarxes i altres recursos que pot incloure sistemes operatius i aplicacions.



# Beneficis del BD & Cloud?

## Infraestructura

**Cost:** no hi inversions d'infraestructura tant hardware com (en alguns casos) software (llicències software). Solo se paga por uso.

**Rapidesa:** accessibilitat i entorns connectats a alta velocitat, Disponibilitat per regions.

**Eficiència:** personal d'IT especialitzat mínim (versions, actualitzacions, problemes de seguretat, etc.).

**Personalització:** configuració està adaptada a les necessitats de cada client, escalabilitat sota demanda, elasticitat.

**Accessibilitat:** accés a la informació des de diferents dispositius des de qualsevol lloc amb accés a internet.

**Seguretat:** protocols de seguretat predefinits, garanties de QoS i serveis de còpies de seguretat adequades a les necessitats. Alta disponibilitat.

# Beneficis del BD & Cloud?

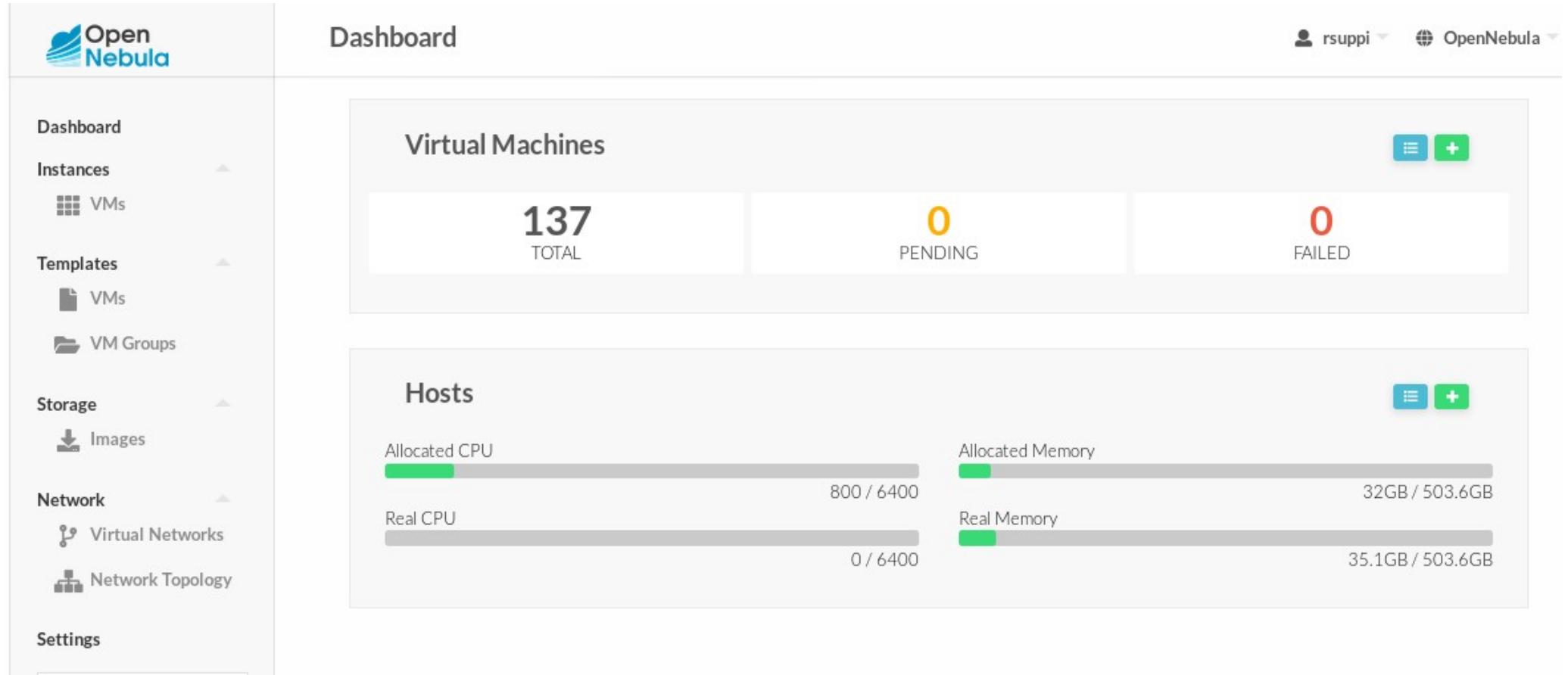
# Complex?

The screenshot shows the AWS Services console interface. The top navigation bar includes the AWS logo, a 'Services' dropdown, a 'Resource Groups' dropdown, and a search bar with placeholder text 'Find a service by name or feature (for example, EC2, S3 or VM, storage.)'. Below the search bar are two buttons: 'Group' and 'A-Z'. The main content area is organized into several categories:

- Compute**: EC2, Lightsail, ECR, ECS, EKS, Lambda, Batch, Elastic Beanstalk, Serverless Application Repository.
- Customer Enablement**: AWS IQ, Support, Managed Services.
- Analytics**: Athena, EMR, CloudSearch, Elasticsearch Service, Kinesis, QuickSight, Data Pipeline, AWS Data Exchange, AWS Glue, AWS Lake Formation, MSK.
- Business Applications**: Alexa for Business, Amazon Chime, WorkMail.
- Storage**: S3, EFS, FSx, S3 Glacier, Storage Gateway, AWS Backup.
- Blockchain**: Amazon Managed Blockchain.
- Satellite**: Ground Station.
- End User Computing**: WorkSpaces, AppStream 2.0, WorkDocs, WorkLink.
- Management & Governance**: AWS Organizations, CloudWatch, AWS Auto Scaling, CloudFormation, CloudTrail, Config, OpsWorks, Service Catalog, Systems Manager.
- Internet Of Things**: IoT Core, Amazon FreeRTOS, IoT 1-Click, IoT Analytics, IoT Device Defender, IoT Device Management, IoT Events, IoT Greengrass, IoT SiteWise, IoT Things Graph.
- Database**: RDS, DynamoDB, ElastiCache, Neptune, Amazon Redshift, Amazon QLDB, Amazon DocumentDB.
- Security, Identity, & Compliance**: IAM, Resource Access Manager, Cognito, Secrets Manager, GuardDuty, Inspector, Amazon Macie, AWS Single Sign-On, Certificate Manager, Key Management Service, CloudHSM, Directory Service, WAF & Shield, Artifact, Security Hub.
- Game Development**: Amazon GameLift.
- Migration & Transfer**.

# Beneficis del BD & Cloud?

# Complex?

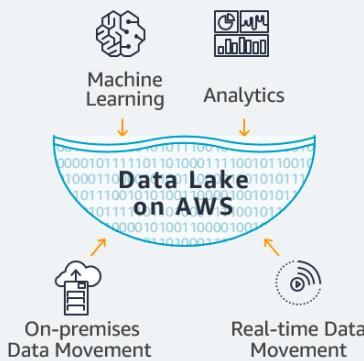


# Actors en BD & Cloud?

# Infraestructura

## Data lakes y analytics en AWS

AWS le ofrece el conjunto más amplio de servicios para migrar, almacenar y analizar datos con el fin de que pueda crear su solución de data lakes y analytics.



### Migración de datos

Importe sus datos de On-Premises en tiempo real.

### Data lakes

Almacene cualquier tipo de datos de manera segura, desde gigabytes a exabytes.

### Analytics

Analice datos con la selección más amplia de servicios de análisis.

### Machine Learning

Prediga resultados futuros e impone acciones para lograr una respuesta rápida.

Microsoft Azure

Información general Soluciones Productos Documentación Precios Formación Marketplace Asociados Soporte técnico Blog Más Cuenta gratuita >

Introducción a la guía para desarrolladores de Azure Guía de la arquitectura de aplicaciones en Azure Cursos gratis de Azure de Pluralsight

Producción Products SDK y herramientas Arquitectura

Automático

- Azure Synapse Analytics (anteriormente SQL DW)  
Servicio de análisis ilimitado que permite obtener conclusiones con una rapidez inigualable
- Azure Databricks  
Cree un área de trabajo de Azure Databricks con clústeres de Spark
- HDInsight  
Aprovisione clústeres de Hadoop, Spark, R Server, HBase y Storm en la nube

- Data Factory  
Coordine y administre el movimiento y la transformación de datos
- Explorador de datos de Azure  
Servicio de exploración de datos altamente escalable y rápido para datos de telemetría y registro
- Stream Analytics  
Análisis en tiempo real de flujos de datos rápidos de aplicaciones y dispositivos

Google Cloud Por qué elegir Google Soluciones Productos Precios Primeros pasos

Gartner

Analíticas de Big Data

Información general

Productos y soluciones

Inteligencia empresarial

Almacenamiento de datos nativo de la nube

Lago de datos

Integración de datos

Tratamiento de datos con Apache Hadoop y Apache Spark

Analíticas de streaming

Orquestación de flujos de trabajo

Preparación de datos

Internet de las cosas

Inteligencia artificial de Cloud y aprendizaje automático

Probar gratis Contactar con Ventas

Reinventa los usos de las analíticas de datos

Obtén información valiosa de forma más rápida y olvídate de las dificultades de las analíticas de datos con Google Cloud Platform. Los servicios de analíticas de datos completos, integrados y sin servidor superan las limitaciones tradicionales de escala, rendimiento y rentabilidad, por lo que ofrecen multitud de ventajas.

Alibaba Cloud Contact Sales Call Us (+44 20-7048-8226) Block the Attacks Search EU - English Cart Console Log In

Free Account

Why Us Olympics Products Solutions Pricing Marketplace Resources & Support Partners Documentation

Scale Up Your Business With Cloud Tech Talk Series

Become Internet Champion in Digital Age

Sign Up Now

Free Trials Enjoy free trials for more than 50 products

New User Discounts Enjoy up to 90% off on trending cloud products

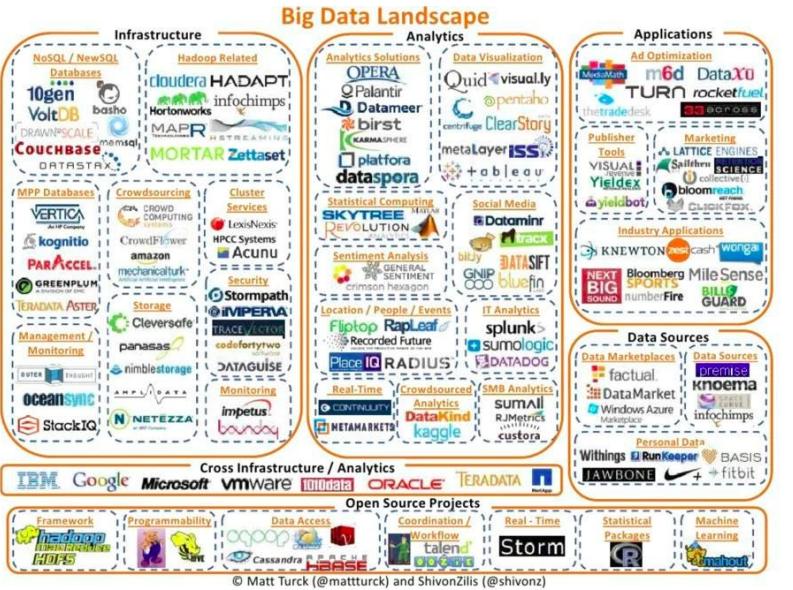
Light Space Start your virtual private server (VPS) journey with...

Cloud Servers Up to 95% Off Special Sale from \$0.43 per month

Selected, One-Stop Store for Enterprise Applications

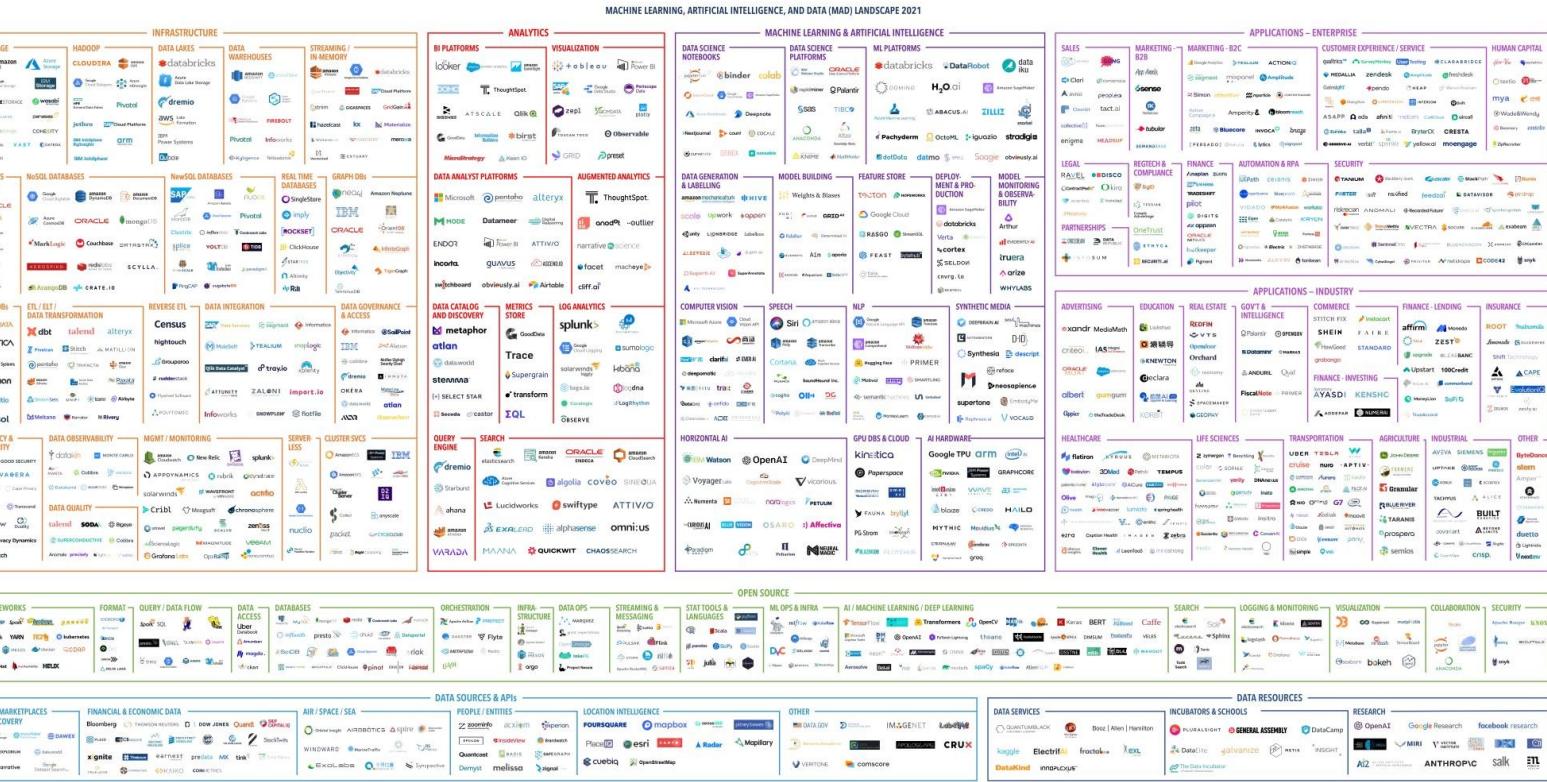
Support various scenarios to meet companies' needs at different stages of development

# *Big Data Landscape*



2012

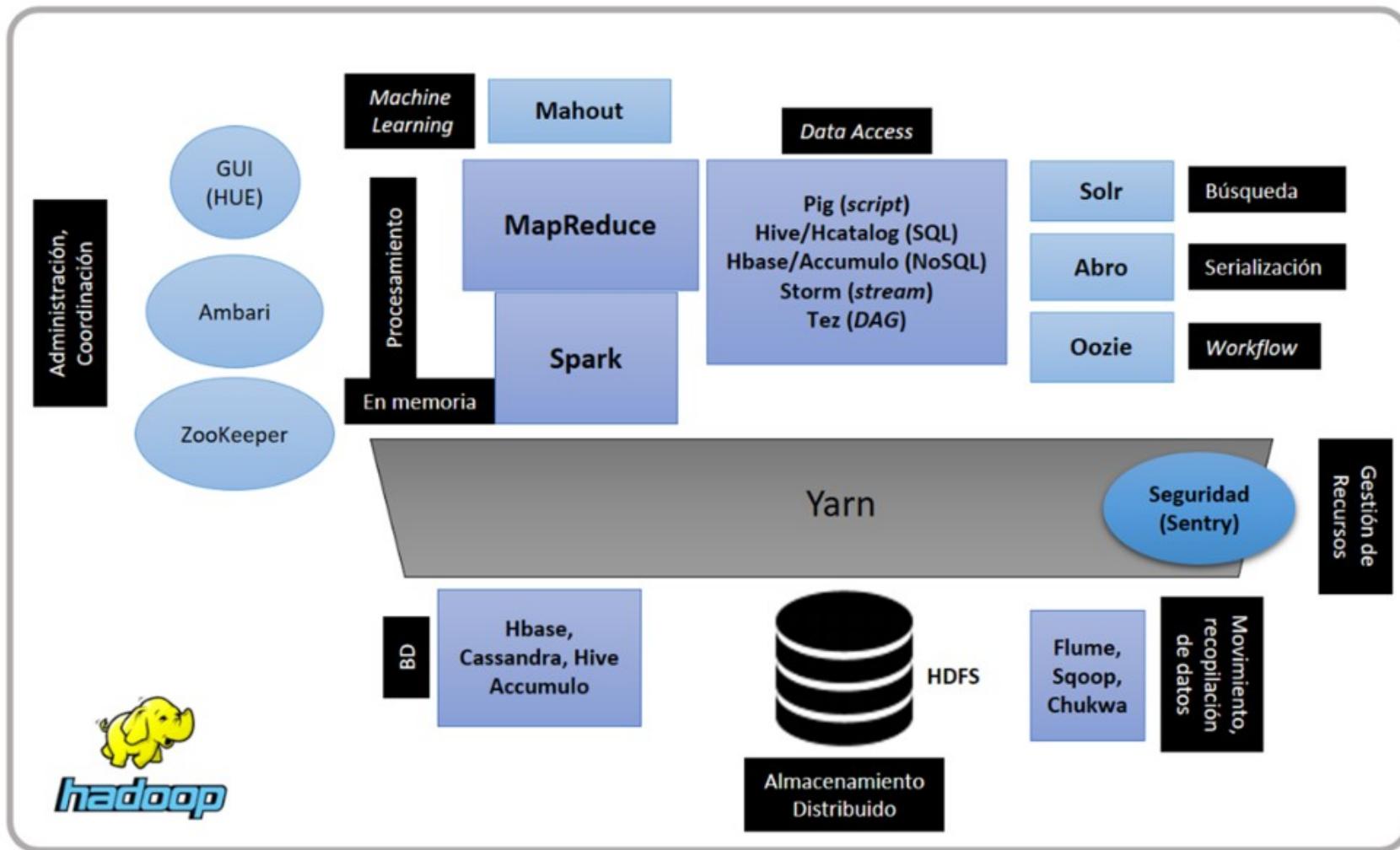
Noviembre 2021



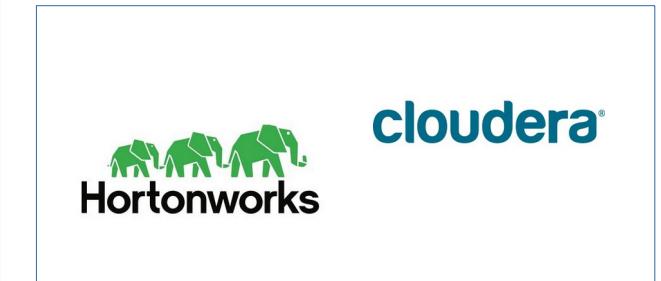
<https://www.forbes.com/sites/brunoaziza/2021/10/03/the-2021-data-landscape-is-out-now-what/>

# Què s'executarà sobre aquesta Infraestructura?

Hadoop (simplificat).

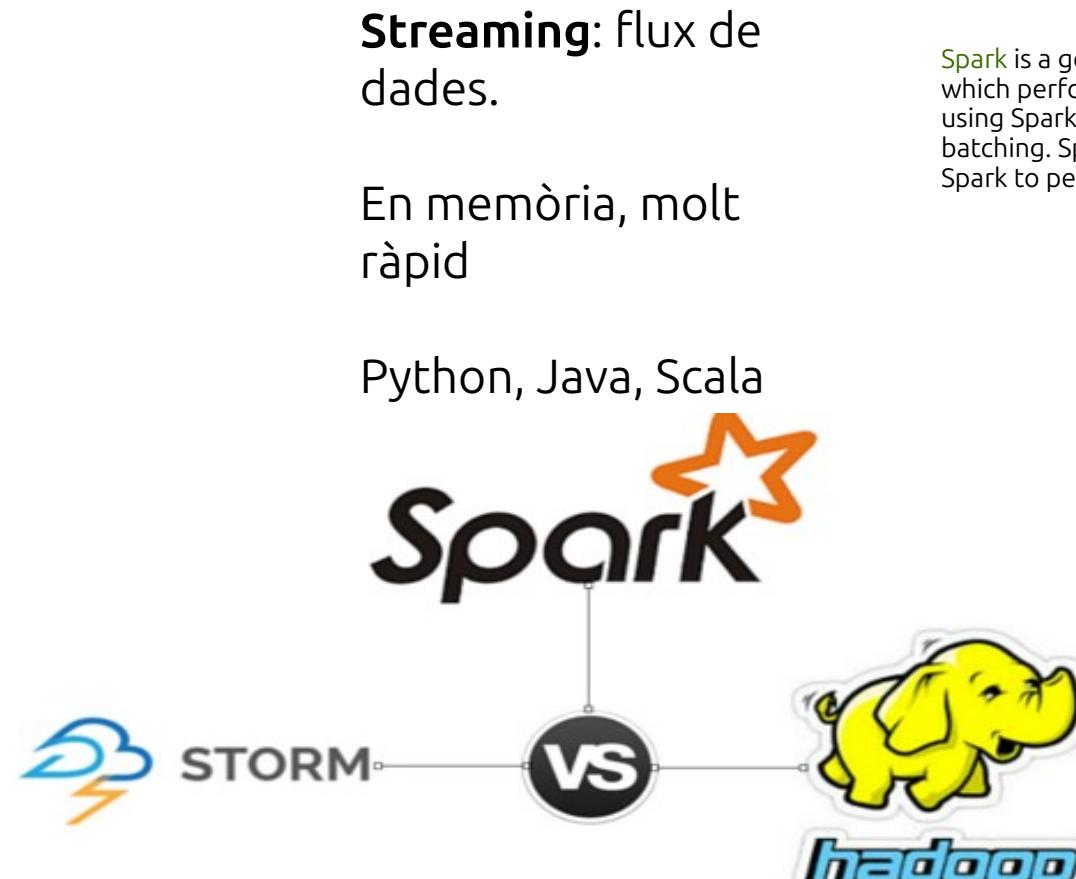


## Distribuciones Hadoop



**Streaming:** flux de dades.  
En funció com es van generant  
Java, Clojure, Scala  
Storm és molt complex i té un cost elevat per desenvolupar aplicacions

For processing real-time streaming data Apache Storm is the stream processing framework. Since it can do micro-batching using a trident. Also, “Trident” an abstraction on Storm to perform stateful stream processing in batches.



## **Streaming:** flux de dades.

En memòria, molt ràpid

Python, Java, Scala

Spark is a general purpose computing engine which performs batch processing. No doubt, by using Spark Streaming, it can also do micro-batching. Spark Streaming is an abstraction on Spark to perform stateful stream processing.

**Batch:** processament de dades emmagatzemades

**Map-Reduce**

# I les Bases de Dades?

**Bases de Dades relacionals:** fàcils de fer servir, coneixement, suport, atomicitat de les operacions, integritat de dades (tipus i compatibilitat).

**Desavantatges:** atomicitat afecta el rendiment, escalabilitat sol ser inferior a NoSQL.

**Bases de dades NoSQL:** escalabilitat i el seu caràcter descentralitzat (estructures distribuïdes), obertes i flexibles. Major adaptabilitat, canvis dels esquemes sense haver de parar les bases de dades, Escalabilitat horitzontal: execució amb pocs recursos. Optimització de consultes per a grans quantitats de dades.

**Desavantatges:** No totes contemplen la atomicitat de les instruccions i la integritat de les dades. Problemes de compatibilitat SQL. Linux si, però altres SO no tant. Eines d'administració no molt desenvolupades.

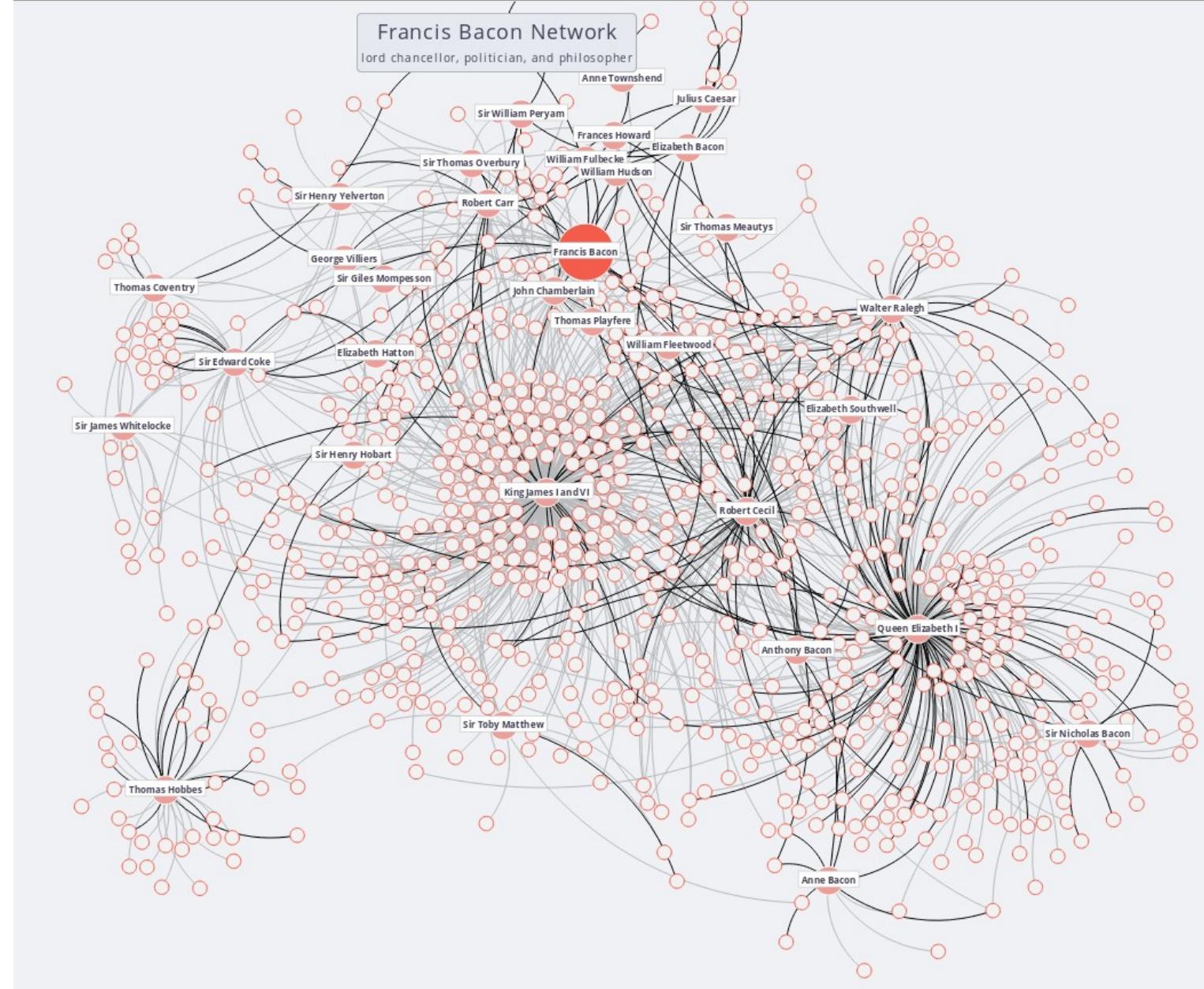
Rank			DBMS	Database Model
Sep 2022	Aug 2022	Sep 2021		
1.	1.	1.	Oracle +	Relational, Multi-model ⓘ
2.	2.	2.	MySQL +	Relational, Multi-model ⓘ
3.	3.	3.	Microsoft SQL Server +	Relational, Multi-model ⓘ
4.	4.	4.	PostgreSQL +	Relational, Multi-model ⓘ
5.	5.	5.	MongoDB +	Document, Multi-model ⓘ
6.	6.	6.	Redis +	Key-value, Multi-model ⓘ
7.	↑ 8.	↑ 8.	Elasticsearch	Search engine, Multi-model ⓘ
8.	↓ 7.	↓ 7.	IBM Db2	Relational, Multi-model ⓘ
9.	9.	↑ 11.	Microsoft Access	Relational
10.	10.	↓ 9.	SQLite +	Relational

# Exemple *Text Analytics*

Text Mining amb les relacions entre les persones, els documents i les institucions per crear una reconstrucció digital de les xarxa socials primerenca d'Anglaterra.

Es va utilitzar el reconeixement d'entitats amb nom per processar el text no estructurat, específicament una matriu de documents i entitats amb nom, que son susceptible d'anàlisi estadística.

A més es van aplicar mètodes estadístics d'aprenentatge de gràfics a les dades estructurats i el modelat per temes.

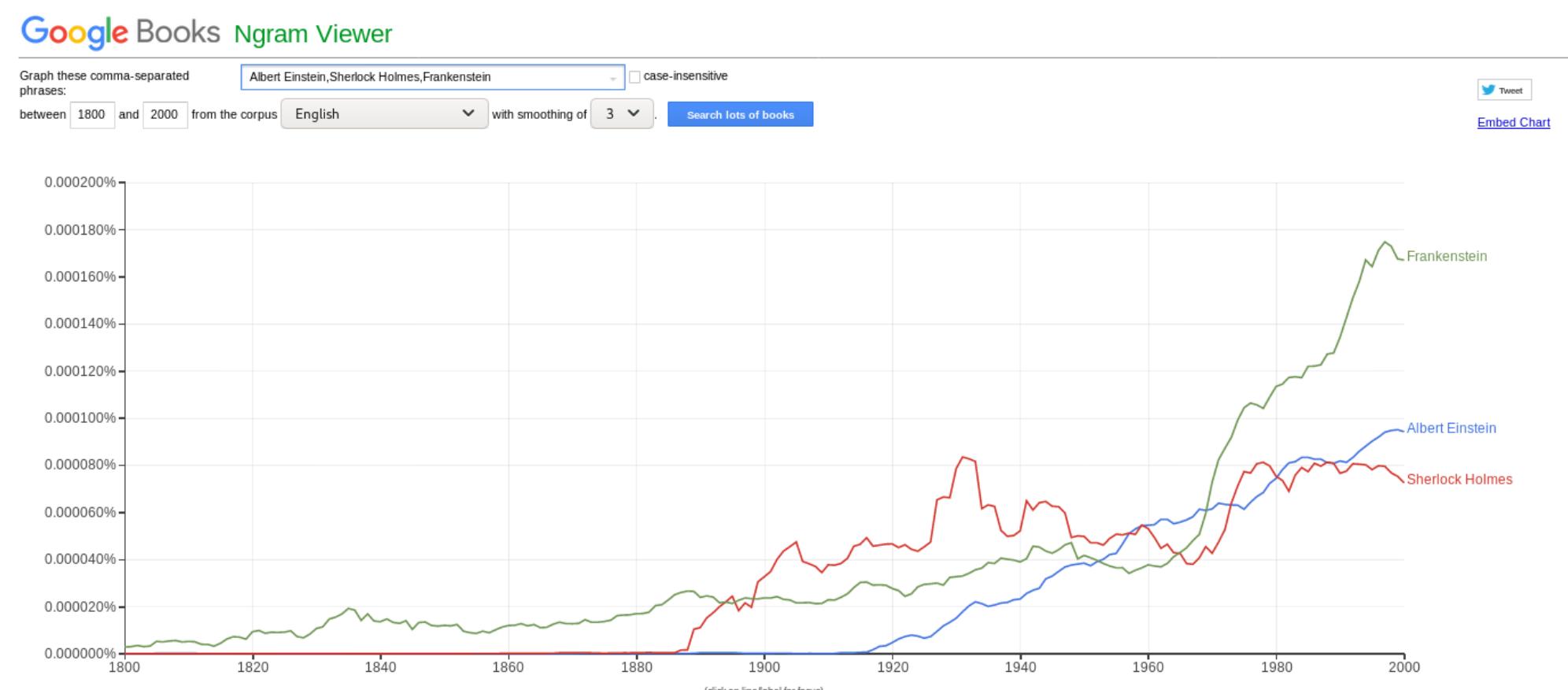


[http://www.sixdegreesoffrancisbacon.com/?ids=10000473&min\\_confidence=60&type=network](http://www.sixdegreesoffrancisbacon.com/?ids=10000473&min_confidence=60&type=network)

# Exemple *Text Analytics*

[Google Ngram Viewer](#), és un cercador en línia que mostra un gràfic de la freqüència de qualsevol grup de cerques amb format de text delimitat per comes, trobat en recursos impresos entre l'any 1500 i el 2000.

Analitza text de Google (95 B de línies) en anglès, xinès, francès, alemany, hebreu, italià, rus o espanyol





Tots els materials, enllaços, imatges, formats, protocols i informació utilitzada en aquesta presentació són propietat dels seus respectius autors i es mostren amb finalitat acadèmica i sense ànim de lucre, excepte tots aquells que tenen llicències o distribució d'ús lliure i/o cedides per tal finalitat. (Articles 32-37 de la llei 23/2006, Spain).

Sota cap concepte (en el cas que es mostrin) accions, ordres, exemples o qualsevol altre activitat es poden provar fora de l'àmbit acadèmic i que no sigui proves en màquines virtuals/xarxes internes protegides i amb finalitat d'aprenentatge ja que es podria incorre en activitats delictives i/o punibles.