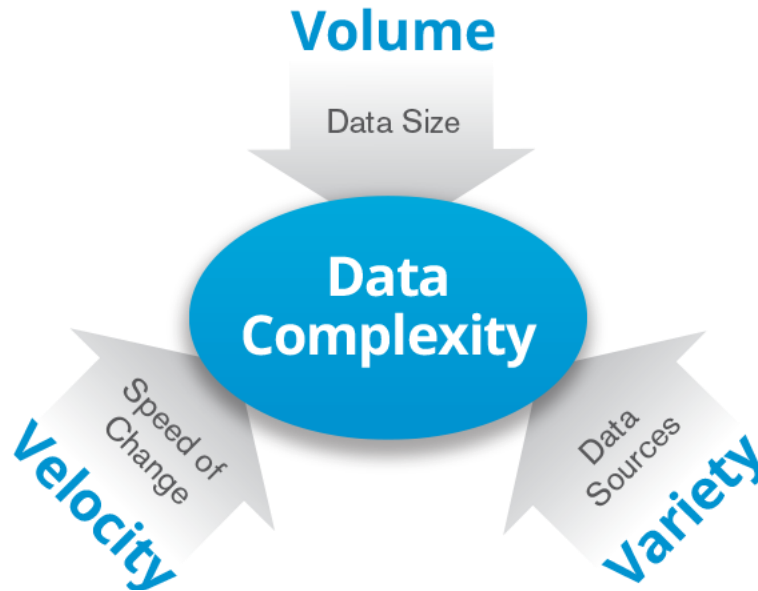




Cap. 12. Big Data-Very large data



Big Data

Kilobyte (KB)	1,000 bytes OR 10^3 bytes 2 Kilobytes: A Typewritten page. 100 Kilobytes: A low-resolution photograph.
Megabyte (MB)	1,000,000 bytes OR 10^6 bytes 1 Megabyte: A small novel OR a 3.5 inch floppy disk. 2 Megabytes: A high-resolution photograph. 5 Megabytes: The complete works of Shakespeare. Megabytes: A minute of high-fidelity sound. 100 Megabytes: 1 meter of shelved books. 500 Megabytes: A D-ROM.
Gigabyte (GB)	1,000,000,000 bytes OR 10^9 bytes 1 Gigabyte: a pickup truck filled with books. 20 A good collection of the works of Beethoven. 100 Gigabytes: A library floor of academic journals.
Terabyte (TB)	1,000,000,000,000 bytes OR 10^{12} bytes 1 Terabyte: 50000 trees made into paper and printed. 2 Terabytes: An academic research library. 10 Terabytes: The print collections of the U.S. Library of Congress. 400 Terabytes: National Climactic Data Center (NOAA) database.
Petabyte (PB)	1,000,000,000,000,000 bytes OR 10^{15} bytes 1 Petabyte: 3 years of EOS data (2001). 2 Petabytes: All U.S. academic research libraries. 20 Petabytes: Production of hard-disk drives in 1995. 200 Petabytes: All printed material.
Exabyte (EB)	1,000,000,000,000,000,000 bytes OR 10^{18} bytes 2 Exabytes: Total volume of information generated in 1999. 5 Exabytes: All words ever spoken by human beings.



Big data

Multiples of bytes <small>v · d · e</small>				
SI decimal prefixes		Binary	IEC binary prefixes	
Name (Symbol)	Value	usage	Name (Symbol)	Value
kilobyte (kB)	10^3	2^{10}	kibibyte (KiB)	2^{10}
megabyte (MB)	10^6	2^{20}	mebibyte (MiB)	2^{20}
gigabyte (GB)	10^9	2^{30}	gibibyte (GiB)	2^{30}
terabyte (TB)	10^{12}	2^{40}	tebibyte (TiB)	2^{40}
petabyte (PB)	10^{15}	2^{50}	pebibyte (PiB)	2^{50}
exabyte (EB)	10^{18}	2^{60}	exbibyte (EiB)	2^{60}
zettabyte (ZB)	10^{21}	2^{70}	zebibyte (ZiB)	2^{70}
yottabyte (YB)	10^{24}	2^{80}	yobibyte (YiB)	2^{80}

See also: [Multiples of bits](#) · [Orders of magnitude of data](#)

1 ZB (zettabyte)= o unitate egala
cu sextilion de bytes
1,000,000,000,000,000,000
bytes = $1000^7 = 10^{21}$

1ZB (zettabyte)=1 miliard de
terabytes



Big data

Explozia informatională 2010-> prezent

Membrii societății de tip occidental sunt supuși unui adevărat bombardament informațional: conform unui studiu american recent, fiecare primește, zilnic, o cantitate **de informație echivalentă cu cea cuprinsă în 147 de ziare!**

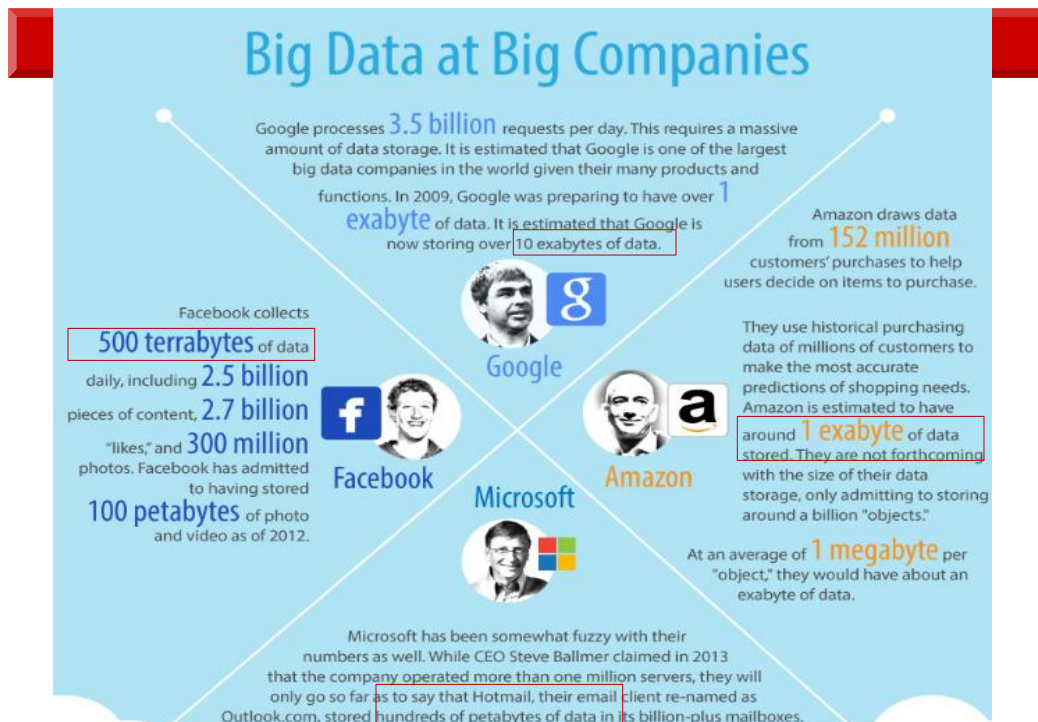
Dezvoltarea internetului, programele de televiziune disponibile 24 de ore din 24, precum și răspândirea telefoanelor mobile au făcut ca, în ziua de azi, o persoană să primească, în fiecare zi, de **200 ori mai multă informație decât primea în 1986.**



Big data

Explozia informationala 2010-> prezent

- ❑ se trimit aprox. **3 milioane email-uri / secunda**,
- ❑ **20 ore video** sunt uploaded in **YouTube** in **60 secunde**,
- ❑ Google proceseaza 24 petabytes de informatie
- ❑ se trimit aprox. **50 milioane SMS /zi**
- ❑ aprox. **73 produse** sunt **comandate pe Amazon /secunda**
- ❑ **zilnic, o persoană produce și transmite** altora, în medie, informație într-o cantitate echivalentă cu cea cuprinsă **în 6 ziare - de 200 ori mai mult decât în urmă cu 24 de ani**, când fiecare "genera" doar 2.5 pagini.
- ❑ in 2008 sau gestionat digital pana la 3.6 zettabytes sau 10,845 trilioane de cuvinte =34 gigabytes de persoana pe an
- ❑ *DACA s-ar stoca datele digitale existente pana la sfarsitul anului 2013 pe DVD se poate forma o stiva care sa acopere distanta de la luna si inapoi*



Other Big Data Companies

Target has focused attention on using customer buying histories, estimated incomes, ages, and marital statuses to predict potential buying patterns. Target caused some controversy with their efforts to determine whether or not female customers are pregnant in order to target their ads and properly stock their shelves.

VMware is a company specializing in cloud and virtualization software. VMware offers companies the ability to store and process big data through their server space and virtual machine software. They have recently received recognition for their expertise in the deployment of hybrid cloud computing systems. This allows companies to utilize multiple clouds, both private and public, simultaneously.



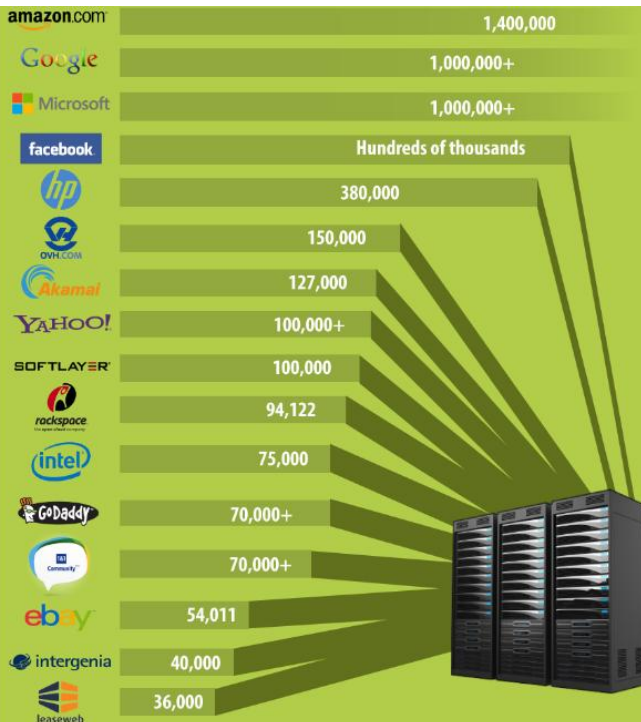
In 2013 UPS introduced ORION. This software utilizes customer, traffic, and driving data collected during deliveries to streamline routes, speed deliveries, and reduce emissions. By the end of 2013 they had saved **1.5 million** gallons of gasoline and reduced their carbon dioxide emissions by **14,000 metric tons**.

According to Amazon sources there are more than **60,000 companies** using AWS. AWS currently uses more bandwidth than Amazon.com proper. AWS represents **\$131 million** of Amazon's annual **\$5.7 billion** revenue. AWS fields more than **650,000** S3 requests every second, and hosts more than **900,000,000** objects.

Who Uses AWS?

1/3rd of all daily internet users will access a website built on AWS infrastructure at least once a day.
1% of all internet traffic is coming or going on AWS backed infrastructure.

Organizations/Companies Using AWS





▶ THE PAST

Digital storage grew annually by **23%** between 1986 and 2007.

Most data was stored on **videotapes** such as VHS cassettes in the pre-digital revolution world of the late 1980s, Vinyl LP records, audio cassette tapes, and photography accounted for significant portions as well.

Paper-based storage represented **33%** of all data storage on its own in 1986.

25% of all data stored in the world in 2000 was stored digitally.

2002 is the first year that digital storage capacity overtook analog capacity.

94% of all data was stored in digital format by 2007.



▶▶ PRESENT

Today, more than **2.5 exabytes** (2.5 billion gigabytes) of data is generated every single day. This is expected to continue growing at a significant rate with mobile devices accounting for much of this data.

Some experts have estimated that **90%** of all of the data the world today was produced within the last two years.

THE FUTURE
NEXT EXIT ↗

of Cloud Technologies

It is estimated that **40 zettabytes** will be created by 2020.

THE FUTURE NEXT EXIT

of Cloud Technologies

It is estimated that **40 zettabytes** will be created by 2020.

Increased usage by companies



A 2014 study found that **94%** of organizations either already are or want to make cloud computing a part of their operations.

Increased focus on security



Studies have shown that cloud users list security as one of their top five concerns for the future of cloud computing. Theft of intellectual property is the primary security threat.

Increased usage of private cloud computing



Currently, **7%** of companies use entirely private cloud computing, while **58%** use some combination of private and public cloud computing.

24% of respondents to a survey, however, claimed that they were interested in exploring private cloud adoption because of legal and regulatory challenges involved in public cloud computing.

Increased education and employment related to cloud computing



According to one survey, **66%** of U.S. and U.K. organizations were interested in increasing their organization's IT skills to better handle cloud computing, but **56%** reported that they were unaware of available courses in cloud computing.

42% of U.S. and U.K. organizations reported having hired IT professionals because of particular skills related to cloud computing, while 43% reported difficulty in finding candidates with necessary cloud computing skills.

79% of U.S. and U.K. companies reported that they believed that greater incorporation of cloud computing into college and university curriculums is necessary.



How Does Big Data Affect Our Daily Lives?

Sports Predictions



Big Data has been shown to be useful in predicting the outcomes of sporting events; big data was famously used in 2012 to predict that the U.S. would win 108 medals in that year's Summer Olympics in which the U.S. ended up winning 104 medals.

Voting Prediction



Big Data has been used to predict the outcomes of elections. Statistician Nate Silver managed to predict the outcome of the 2012 presidential election with perfect accuracy.

Smartphones



When a smartphone user gets directions, asks their phone a question out loud, or any number of other functions, it is the result of analyzing big data.

Personalized Advertising and Purchasing Recommendations



One of the primary uses for big data has been in the recommending of purchases and personalization of ads on websites. One study found that a person is more likely to complete Navy Seal training than to actually click a banner ad. Both customers and companies stand to benefit from more personalized and relevant ads.

Improved Traffic Flow



Several companies and cities have utilized big data to streamline the flow of traffic in their towns. Using data derived from drivers' GPS signals to react in real time to traffic conditions, weather, accidents, etc. in order to maintain smooth traffic flow.

Epidemic Detection and Prevention



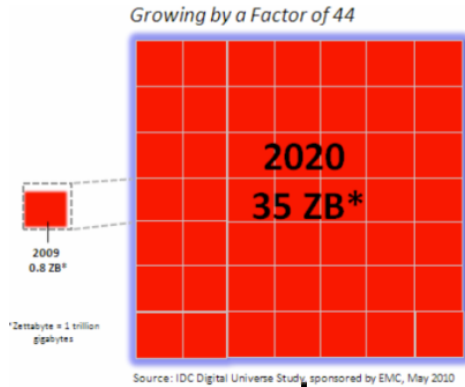
Big data has recently come into use by Google and more recently by the traditional medical establishment to predict where outbreaks of potentially epidemic viruses such as the flu are most likely to appear.



Cata informatie exista in web?

Spatiul Web format din 2 componente:

- “de suprafata” : situri publice cunoscute ca Web
- “de adancime” : situri specializate > 400 – 500 ori decat “suprafata”



Cantitatea de informatie digitala produsa a fost de :

- 0.8 zettabytes in 2009
- 1.2 zettabytes in 2010
- daca cresterea se mentine in 2020 se vor produce 35 -40ZB



Google servers & Data Centers

Americas

Berkeley County, South Carolina
Council Bluffs, Iowa
Douglas County, Georgia
Jackson County, Alabama
Lenoir, North Carolina
Mayes County, Oklahoma
Montgomery County, Tennessee
Quilicura, Chile
The Dalles, Oregon

Asia

Changhua County, Taiwan
Singapore

Europe

Dublin, Ireland
Eemshaven, Netherlands
Hamina, Finland
St Ghislain, Belgium





Google servers & Data Centers

Douglas County - 417,600 servers



The Dalles - 204,160 servers



Google servers & Data Centers

Council Bluffs - 241,280 servers



Lenoir - 139,200 servers



St. Ghislain - 296,960 servers





Big Data Mining



Big Data 3.0:
Intelligent



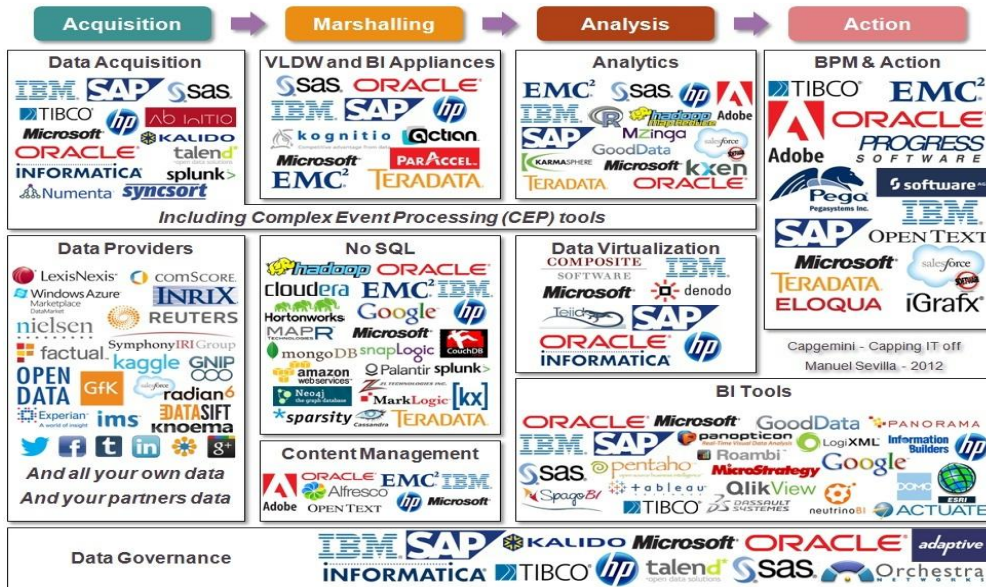
Big Data 2.0:
Networked



Big Data 1.0:
Transactional



Big Data Software





Supercalculatoare: Titan-cel mai rapid supercalculator din lume in 2013



Produs: Crop , USA

Lansat in 2013

Performanta: **20 petaflops** (*F*loating *p*oint *O*perations *P*er *S*econd), peste 18.000 de unitati de procesare grafica, urmatorul sistem si mai performant va avea 100 petaflops.



Supercalculatoare: Tianhe 2 cel mai rapid supercalculator din lume in 2014



Operat din orașul Guangzhou cu ajutorul unei echipe de 1300 ingineri

În total, Tianhe-2 dispune de **3.120.000 nuclee de procesare și 1375 Terabytes memorie RAM.**

Comparabil unui mic oraș, consumul de energie poate atinge 17.6MW/hr la încărcare maximă, respectiv **24MW/hr** după includerea sistemelor de climatizare folosite pentru ținerea sub control a temperaturilor.

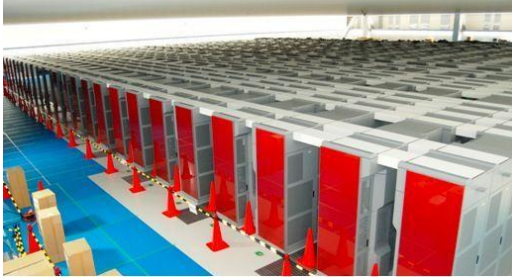
Lansat in 2014

Performanta: **33.86 petaflop/s** (33.86 miliarde de miliarde operații de calcul cu virgulă mobilă pe secundă)

Tianhe-2, e deținut de guvernul chinez și folosit de Universitatea Națională de Tehnologie din China, în domeniul apărării păstrează titlul de cel mai rapid supercomputer din lume.



Supercalculatoare: K- cel mai puternic calculator din lume



Domenii de utilizare: simulări de cutremure, previziuni ale schimbărilor climatice, cercetare nucleară, tranzacții pe bursa, exploatarea petrolului, etc.

•consumul energetic al K computer = consumul a 10.000 de case.

Produs: Fujitsu și RIKEN, este pe locul întâi în topul supercomputerelor din lume. Computerul este capabil să facă 8.2 catralioane de operații pe secundă, adică 8.200.000 miliarde, de 3 ori mai multe decât oricare alt calculator.

Calculatorul are 672 rack-uri care conțin în total 68 544 procesoare.

Supercomputer =conectarea a mii de calculatoare într-un centru de date



Top tari producatoare de supercomputere

- Loc I Japonia
- Loc II America , a pierdut locul I , dar domină autoritar top-ul celor mai performante computere din lume ocupând **233 dintre primele 500 poziții**.
- Loc III, China deține doar 76 dintre ocupanții pozițiilor rămase,
- Loc IV Marea Britanie (30),
- Loc V Franța (27) și
- Loc VI Germania (23).