

2^η Διάλεξη



❖ Παγκόσμιος Ιστός & Big Data

⦿ Τι είναι τα big data

- ✦ Τεράστια σύνολα δεδομένων
 - ✓ Χιλιάδες TB

⦿ Πώς προέκυψαν

- ✦ Ευρεία χρήση τεχνολογίας
- ✦ Εξάπλωση αισθητήρων & κινητών συσκευών
- ✦ Μείωση κόστους αποθήκευσης & επεξεργασίας

Ας θυμηθούμε τις μονάδες

<u>Decimal</u>		
<u>Value</u>	<u>Metric</u>	
1000	kB	<u>kilobyte</u>
1000 ²	MB	<u>megabyte</u>
1000 ³	GB	<u>gigabyte</u>
1000 ⁴	TB	<u>terabyte</u>
1000 ⁵	PB	<u>petabyte</u>
1000 ⁶	EB	<u>exabyte</u>
1000 ⁷	ZB	zettabyte
1000 ⁸	YB	<u>yottabyte</u>

❖ Τα 3 V (Volume, Variety, Velocity)

❖ Volume

- ✦ Terabytes έως και Petabytes δεδομένων
- ✦ Η ποσότητα των δεδομένων αυξάνεται συνεχώς
- ✦ Το τι θεωρείται «Big Data» σήμερα, αύριο θα είναι ακόμη μεγαλύτερο

⦿ YouTube

- ✦ 2 ΔΙΣ χρήστες κάθε μήνα
- ✦ 4 ΔΙΣ προβολές βίντεο τη μέρα
- ✦ 500 ώρες νέων video κάθε λεπτό!

⦿ Twitter

- ✦ 500 εκατομμύρια tweets τη μέρα

⦿ Facebook

- ✦ Μέχρι και 1 ΔΙΣ χρήστες τη μέρα
- ✦ 55 εκατομμύρια status updates

⦿ Instagram

- ✦ 95 εκατομμύρια φωτογραφίες και videos καθημερινά!

📍 Συναλλαγές πιστωτικών καρτών

Enabling Everywhere

Visa connects millions of consumers and businesses every day through the power of VisaNet.

46+ million
merchant locations⁵



3.3 billion cards
worldwide¹



200+
countries and
territories



15,900
financial
institutions²



160+
currencies



\$11.0 trillion
total volume³



65,000+
transaction messages
per second (capacity)⁴



Our vision is to be the best way to pay and be paid for everyone, everywhere.

VISA Fact Sheet

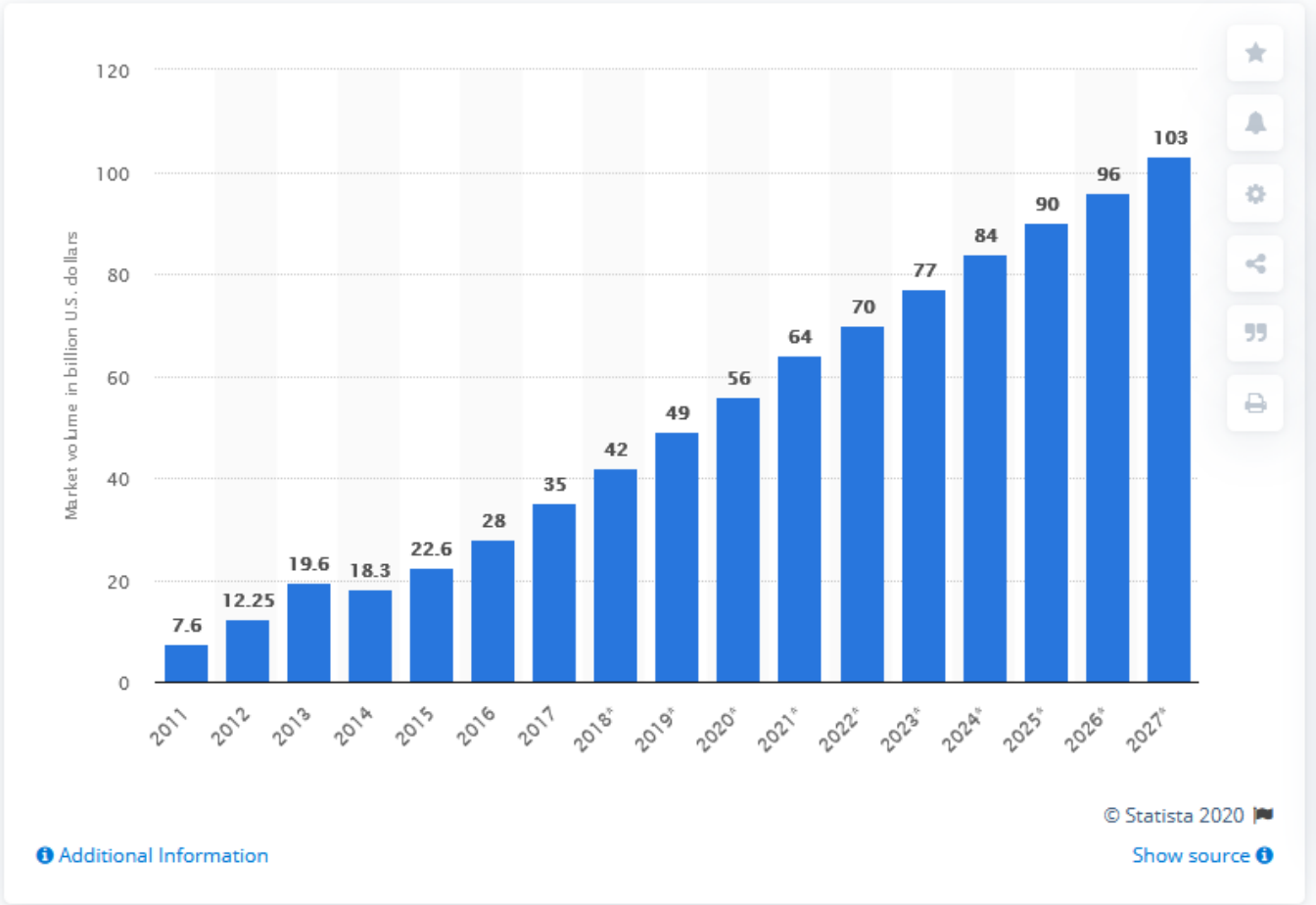
- ◉ Συγκέντρωση δεδομένων
 - * Από το διαδίκτυο
 - ✓ Κοινωνικά δίκτυα
 - ✓ Web 2.0
 - ✓ Οικονομικά δεδομένα
 - * Από αισθητήρες
 - * Από έξυπνες συσκευές
 - * Δορυφορικά δεδομένα - GPS
- ◉ Μορφές
 - * Κείμενο
 - * Δεδομένα πλοήγησης
 - * Δεδομένα αισθητήρων
 - * Ήχος
 - * Βίντεο
 - * Αρχεία καταγραφής

- ⊙ Η ταχύτητα με την οποία παράγονται τα δεδομένα
 - * Συνεχώς αυξάνεται
- ⊙ Υπάρχουν εφαρμογές που απαιτούν real time επεξεργασία των δεδομένων αυτών
 - * Ανίχνευση απάτης πιστωτικών καρτών
 - * Χρηματοοικονομικά εργαλεία
 - * Συστήματα προτάσεων
 - * Διαφημίσεις - \$\$\$

- ⊕ Αποθήκευση
 - ✦ Καταγραφή και αποθήκευση των δεδομένων
- ⊕ Επεξεργασία
 - ✦ Ξεκαθάρισμα των δεδομένων
 - ✦ Ανάλυση των δεδομένων
- ⊕ Πρόσβαση
 - ✦ Κατάλληλα εργαλεία για την πρόσβαση σε **πληροφορία**
 - ✦ Οπτικοποίηση
 - ✦ Δημιουργία διαίσθησης
 - ✦ Πρόσβαση μέσω Παγκόσμιου Ιστού
- ⊕ Η παγκόσμια οικονομία σχετικά με τα big data υπολογίζεται ότι θα φτάσει στα 103 δις \$ το 2027

Big Data Market Size

(in billion U.S. dollars)



Big Data - SAP

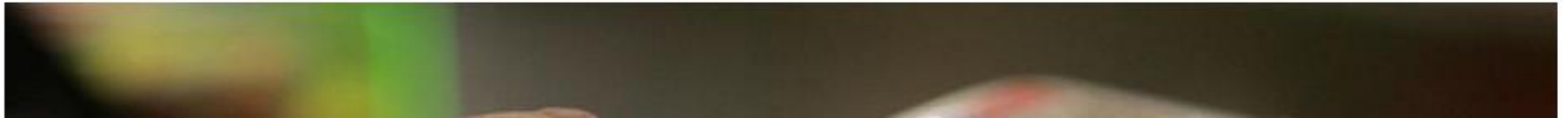


📍 Κυβερνητικές εφαρμογές

- ✦ Στατιστική υπηρεσία
- ✦ Κτηματολόγιο
- ✦ ΔΙΑΥΓΕΙΑ
- ✦ Υπουργείο Οικονομικών
 - ✓ Φοροδιαφυγή

ΟΙΚΟΝΟΜΙΑ | 🕒 21-07-2020 | 19:17

Παράταση για την online σύνδεση των ταμειακών μηχανών με το Taxisnet



United States of America [\[edit \]](#)

- In 2012, the [Obama administration](#) announced the Big Data Research and Development Initiative, to explore how big data could be used to address important problems faced by the government.^[51] The initiative is composed of 84 different big data programs spread across six departments.^[52]
- Big data analysis played a large role in [Barack Obama's successful 2012 re-election campaign](#).^[53]
- The [United States Federal Government](#) owns six of the ten most powerful [supercomputers](#) in the world.^[54]
- The [Utah Data Center](#) is a data center currently being constructed by the [United States National Security Agency](#). When finished, the facility will be able to handle a large amount of information collected by the NSA over the Internet. The exact amount of storage space is unknown, but more recent sources claim it will be on the order of a few [exabytes](#).^{[55][56][57]}

✦ Υγεία

- ✦ Στατιστικές αναλύσεις
- ✦ Δεδομένα
 - ✓ Αιματολογικές εξετάσεις
 - ✓ Απεικονιστικές εξετάσεις (μαγνητικές, αξονικές)
 - ✓ Αισθητήρες πίεσης, θερμοκρασίας, κορεσμού κλπ
- ✦ Μοντέλα διάγνωσης
 - ✓ Διάγνωση μέσω υπολογιστή (computer-aided diagnosis)
 - ✓ Π.χ. Σύνδρομο down – εξετάσεις εγκυμοσύνης
- ✦ Φαρμακευτικές
 - ✓ Έρευνα για νέα φάρμακα/ εμβόλια
- ✦ Βάσεις Δεδομένων DNA

TheScientist

EXPLORING LIFE, INSPIRING INNOVATION

Big Data and Collaboration Seek to Fight COVID-19

Researchers try unprecedented data sharing and cooperation to understand COVID-19—and develop a model for diseases beyond the coronavirus pandemic.



Emma Yasinski

Jul 21, 2020

Although cough and fever have been considered the most tell-tale signs of COVID-19, in May, researchers published a study suggesting that loss of smell and taste were [better able to predict](#) who would test positive for the disease. The insight came from data shared by millions of individuals who logged on to a phone app to report what, if any, symptoms they were experiencing on a given day.

ABOVE: © ISTOCK.COM,
METAMORWORKS

What's Even Creepier Than Target Guessing That You're Pregnant?

The age of big data is frightening to a lot of people, in part because of the implicit promise that algorithms, sufficiently supplied with data, are better at inference than we are. Superhuman powers are scary: Beings that can change their shape are scary, beings that rise from the dead are scary, and beings that can make inferences that we cannot are scary. It was scary when a statistical model deployed by the guest marketing analytics team at Target **correctly inferred based on purchasing data** that one of its customers—sorry, *guests*—a teenage girl in Minnesota, was pregnant, based on an arcane formula involving elevated rates of buying unscented lotion, mineral supplements, and cotton balls. Target started sending her coupons

📍 Τουρισμός

✦ Κρατήσεις

✓ Εισιτήρια

✓ Ξενοδοχεία

✓ AirBnB

✦ Reviews

✓ TripAdvisor

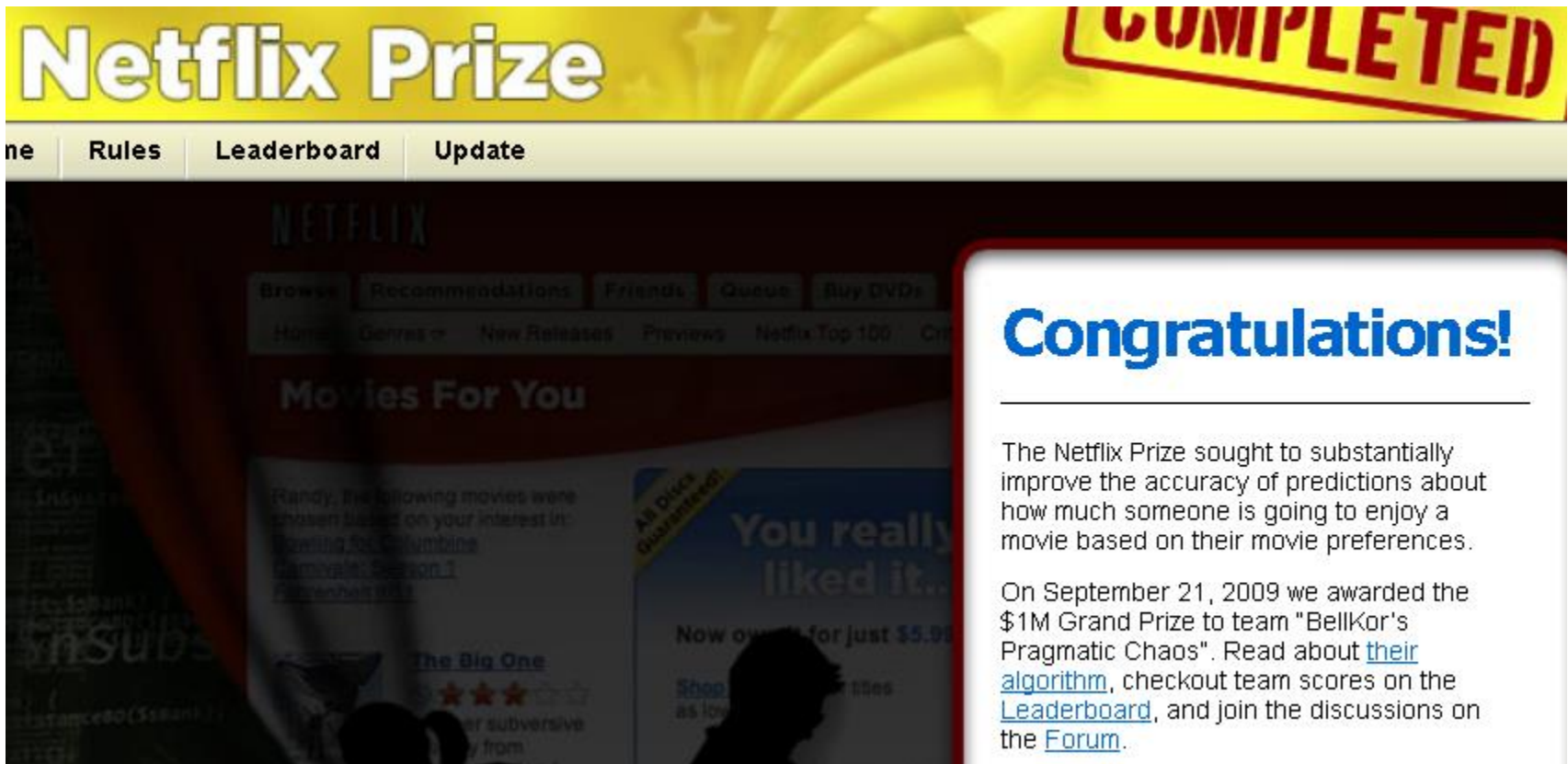
📍 Αθλητισμός

✦ Formula 1

✓ Σε κάθε αγώνα χιλιάδες σένσορες → TBs
δεδομένων

Netflix's New 'My List' Feature Knows You Better Than You Know Yourself (Because Algorithms)

The Huffington Post | By Dino Grandoni   



Technology [\[edit \]](#)

- [eBay.com](#) uses two data warehouses at 7.5 [petabytes](#) and 40PB as well as a 40PB [Hadoop](#) cluster for search, consumer recommendations, and merchandising. [Inside eBay's 90PB data warehouse](#) [↗](#)
- [Amazon.com](#) handles millions of back-end operations every day, as well as queries from more than half a million third-party sellers. The core technology that keeps Amazon running is Linux-based and as of 2005 they had the world's three largest Linux databases, with capacities of 7.8 TB, 18.5 TB, and 24.7 TB.^[67]
- [Facebook](#) handles 50 billion photos from its user base.^[68]
- As of August 2012, [Google](#) was handling roughly 100 billion searches per month.^[69]
- [Oracle NoSQL Database](#) has been tested to past the 1M ops/sec mark with 8 shards and proceeded to hit 1.2M ops/sec with 10 shards.^[70]

Private sector [\[edit \]](#)

Retail [\[edit \]](#)

- [Walmart](#) handles more than 1 million customer transactions every hour, which are imported into databases estimated to contain more than 2.5 petabytes (2560 terabytes) of data—the equivalent of 167 times the information contained in all the books in the US [Library of Congress](#).^[2]

Retail banking [\[edit \]](#)

- FICO Card Detection System protects accounts world-wide.^[71]
- The volume of business data worldwide, across all companies, doubles every 1.2 years, according to estimates.^{[72][73]}



Twitter can predict where FLU outbreaks will occur and could help doctors prepare for busy times



Τα tweets αποκαλύπτουν την οικονομική κατάσταση του χρήστη

In: MEDIA

Το twitter μπορεί να «μαρτυρήσει» χονδρικά το επίπεδο του εισοδήματος και μάλιστα με αρκετή ακρίβεια

CRUNCH NETWORK

The rise of big data policing

Posted Oct 22, 2017 by [Andrew Guthrie Ferguson \(@ProfFerguson\)](#)



Welcome to the Los Angeles Police Department's Real-Time Analysis Critical Response (RACR) Division. The RACR Division, in partnership with Palantir—a private technology company that began developing social network software to track terrorists—has jumped head first into the [big data age of policing](#).

Just as in the hunt for international terror networks, [Palantir's software system](#) integrates, analyzes, and shares otherwise-hidden clues from a multitude of ordinary law enforcement data sources. A detective investigating a robbery suspect types a first name and a physical description into the computer—two fragmented clues that would have remained paper scraps of unusable data in an earlier era. The database searches for possible suspects.

Big data in the UK Police Force

Feature 7 SEPTEMBER 2017

Of the 43 territorial forces in England and Wales, there are examples of individual forces that have recently piloted big data technology, with promising results

Age, description, address, tattoos, gang affiliations, vehicle ownership instantly pop up in sortable fields. By matching known attributes, the computer narrows the search to a few choices. A photograph of a possible suspect's car is identified from an automated license-plate reader scouring the city for data about

the location of every vehicle. Detectives follow up with a witness to identify the car used in the robbery. A match leads to an arrest and a closed case.

Roll call. Monday morning. Patrol officers receive [digital maps](#) of today's "crime forecast." Small red boxes signify areas of predicted crime. These boxes represent algorithmic forecasts of heightened criminal activity: years of accumulated crime data crunched by powerful computers to target precise city blocks. Informed by the data, "[predictive policing](#)" patrols will give additional attention to these "hot" areas during the shift.

- ⦿ Ερωτήσεις?
- ⦿ stefanis@upatras.gr

