



Ευφυή Συστήματα Λήψης Απόφασης στις Επιστήμες Υγείας

Χασάπη Μαρία Κωνσταντίνα

Υπ. Διδάκτωρ

BIG DATA

Σύνολα δεδομένων τα οποία δεν μπορεί να αντιληφθεί και να διαχειριστεί η παραδοσιακή πληροφορική και τα εργαλεία λογισμικού μέσα σε ένα ανεκτό επίπεδο χρόνου

BIG DATA

SMALL DATA

VS

in

Χαρακτηριστικά των Big Data (5Vs)



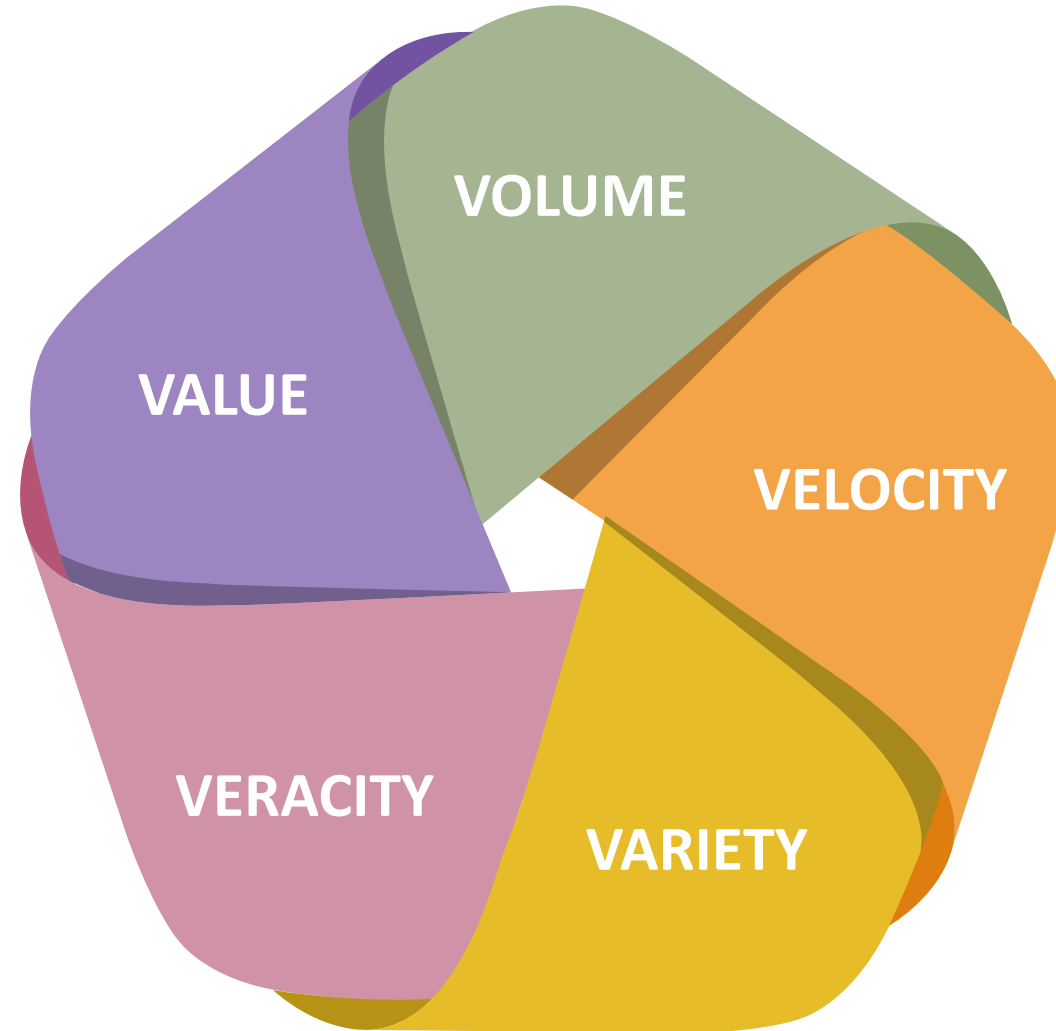
ΑΞΙΑ

Η πρόσβαση σε μεγάλα δεδομένα αν δε μπορούμε να τη μετατρέψουμε σε αξία είναι άχρηστη. Οι οργανισμοί καλούνται να επιλέξουν την πιο αποτελεσματική από πλευράς κόστους λύση με στόχο την αξιοποίηση της πληροφορίας που θα οδηγήσει στην έγκαιρη και όσο το δυνατόν πιο σωστή λήψη αποφάσεων, δίνοντας τη μεγαλύτερη δυνατή αξία στην επιχείρηση.



ΕΙΛΙΚΡΙΝΕΙΑ

Η ειλικρίνεια των πληροφοριών βασίζεται στο βαθμό συμφωνίας με την αλήθεια ή την πραγματικότητα. Τυχόν αβεβαιότητες στα δεδομένα μπορούν να προκληθούν από λανθασμένες εφαρμογές μοντέλων, παραπλάνηση, επικάλυψη πληροφοριών.



ΟΓΚΟΣ

Ο όγκος αναφέρεται στο πλήθος των δεδομένων που δημιουργούνται. Ο τεράστιος όγκος είναι βασική ιδιότητα των Big Data. Γενικά ακολουθούν εκθετική αύξηση ενώ δημιουργούνται πολύ γρήγορα.



ΤΑΧΥΤΗΤΑ

Η ταχύτητα είναι ένα χαρακτηριστικό των Big Data που αφορά το χρόνο δημιουργίας και αποθήκευσης των δεδομένων, τη διαθεσιμότητα και την πρόσβαση. Η ταχύτητα είναι μια βασική ιδιότητα των μεγάλων δεδομένων που παράγονται πολύ γρήγορα καθημερινά στον ψηφιακό κόσμο.



ΠΟΙΚΙΛΙΑ

Τα δεδομένα των Big Data δεν έχουν μια σταθερή δομή και σπάνια παρουσιάζονται σε μια απόλυτα επεξεργάσιμη μορφή. Τα δεδομένα αυτά μπορεί να είναι ιδιαίτερα δομημένα (δεδομένα από σχεσιακές βάσεις δεδομένων), ημιδομημένα (αρχεία καταγραφής ιστού, social media feeds κλπ) ή μη δομημένα (βίντεο, φωτογραφίες). Το εύρος και η ποικιλία των δεδομένων είναι τέτοια ώστε να επηρεάζουν τη σημασιολογία, ή τη μεταβλητότητα του νοήματος.

Χαρακτηριστικά των Big Data στην Υγεία

ΑΕΙΑ

Το κόστος της υγειονομικής περίθαλψης είναι μη βιώσιμο και εμφανίζει συνεχή αύξηση.

Τα πολλαπλά οφέλη όμως που προσφέρει η χρήση και αξιοποίηση των Big Data στον τομέα της υγείας είναι πολύ περισσότερα.

ΕΓΚΥΡΟΤΗΤΑ

Καθορίζουν την ανθρώπινη ύπαρξη.

Πληροφορίες σχετικές με τη διάγνωση, τη θεραπεία, τη χορήγηση φαρμάκων, την περίθαλψη και όποια άλλη πληροφορία, κρίνεται απαραίτητο να καταγραφεί. Η εγκυρότητα των στοιχείων αυτών είναι πάρα πολύ σημαντική.



ΟΓΚΟΣ

Ψηφιοποίηση ήδη διαθέσιμων δεδομένων, και δημιουργίας νέων μορφών δεδομένων. Προσωπικά ιατρικά αρχεία, εικόνες ραδιολογίας και ακτινολογίας, κλινικές δοκιμές, έρευνες, δημογραφικά στοιχεία, ανθρώπινα γονιδιώματα, γενετικές ακολουθίες, εικόνες τριών διαστάσεων (3D), δεδομένα γονιδιωματικών, βιομετρικών αισθητήρων.

ΤΑΧΥΤΗΤΑ

Στατικές πηγές, όπως ακτινογραφίες, έγγραφα νοσοκομείων, δελτία ασθενών, βιβλιάρια υγείας.

Σε **πραγματικό χρόνο**, όπως για παράδειγμα η επίβλεψη της αρτηριακής πίεσης και της καρδιακής λειτουργίας σε κάποια εγχείρηση.

Αργούς σχετικά ρυθμούς, όπως ο προσδιορισμός των επιπέδων της γλυκόζης διαβητικών ατόμων σε καθημερινή βάση.

Πληροφορίες που αφορούν κάποια γνωστή ασθένεια, **αυξάνονται ποσοστιαία** με πολύ **μικρότερο ρυθμό** σε σύγκριση με μία επιδημία σε έξαρση, όπου τα δεδομένα καταφθάνουν με **μεγάλο ρυθμό** και συνιστούν **“νέα” πληροφορία**, ενώ επιτακτική είναι η ανάγκη **άμεσης** επεξεργασίας τους.

ΠΟΙΚΙΛΙΑ

Μη δομημένα δεδομένα, όπως ιατρικές καταγραφές, χειρόγραφες σημειώσεις ιατρικού και νοσηλευτικού προσωπικού για την περιγραφή συμπτωμάτων – ενδείξεων – συμπεριφοράς, ιατρικές εικόνες.

Δομημένα δεδομένα, όπως πληροφορίες σχετικά με ηλεκτρονική συνταγογράφηση φαρμάκων, ποσοτικά δεδομένα μετρήσεων οργάνων και εξετάσεων.

Νέων πηγών, όπως συσκευών ευεξίας που καταγράφουν τους παλμούς ή τη διάρκεια ύπνου των ασθενών, κοινωνικών δικτύων, γονιδιωματικής έρευνας.

Δημιουργία Δεδομένων

Δεδομένα επιχειρήσεων: Τα εσωτερικά δεδομένα των επιχειρήσεων είναι η κύρια πηγή δημιουργίας Big Data. Τα δεδομένα των επιχειρήσεων είναι αποτέλεσμα των εμπορικών της συναλλαγών, της παραγωγής, στοιχείων αποθεμάτων, δεδομένα πωλήσεων, οικονομικά στοιχεία και δεδομένα με γνώμονα τις δραστηριότητες. Εκτιμάται ότι ο όγκος των επιχειρηματικών δεδομένων παγκοσμίως διπλασιάζεται κάθε 1 - 2 χρόνια, καθιστώντας περισσότερο αναγκαία την αποτελεσματική σε πραγματικό χρόνο ανάλυση των δεδομένων.

Internet of Things (IoT): Το IoT είναι μια σημαντική πηγή των Big Data. Αξιοποιώντας τις διεργασίες απόκτησης και μετάδοσης των δεδομένων στο IoT, η αρχιτεκτονική του δικτύου μπορεί να διαστρωματωθεί ως εξής: ανίχνευση, συγκέντρωση και αξιοποίηση των δεδομένων. Τα δεδομένα που δημιουργούνται από το IoT έχουν μεγάλη κλίμακα, ετερογένεια, ισχυρή συσχέτιση χώρου και χρόνου.

Βιοϊατρικά δεδομένα: Στοιχεία βιοϊατρικών καταγραφών, μετρήσεις, ερευνητικά στοιχεία στον τομέα της βιοϊατρικής έχουν εισέλθει στην εποχή των Big Data.

Διάφορες πηγές: Η δημιουργία των Big Data βασίζεται σε στοιχεία επιστημονικών εφαρμογών, στοιχεία αστρονομίας, φυσικών καταγραφών, εμπορικών συναλλαγών, διαδίκτυο κλπ. Τα στοιχεία από διαφορετικές πηγές παρουσιάζουν αυξημένη ετερογένεια ενώ είναι πολλές φορές αρκετά πολύπλοκα.

Διαχείριση Δεδομένων

Η αναγκαιότητα διαχείρισης των δεδομένων σε εφαρμογές μεγάλων δεδομένων επέβαλε τη δημιουργία μίας νέας γενιάς **συστημάτων, μοντέλων και προγραμματιστικών εργαλείων, τεχνολογίες** που επιτρέπουν την παράλληλη **επεξεργασία** δεδομένων σε μεγάλη κλίμακα και με ανεκτικό στα σφάλματα τρόπο.

Σε πολλές περιπτώσεις είναι χρήσιμο να πραγματοποιείται **εκκαθάριση και μετατροπή** των μη-δομημένων δεδομένων σε δομημένα, ιδιαίτερα όταν αυτά προέρχονται από πηγές του διαδικτύου και διαθέτουν μεγάλη μορφολογική ποικιλία.

Η ποικιλία των Big Data σε συνδυασμό με τον **τεράστιο όγκο** τους, καθώς και όλες οι **ιδιομορφίες** που συνήθως τα συνοδεύουν, καταδεικνύουν την ανάγκη αποτελεσματικής απεικόνισής τους. Για το λόγο αυτό είναι ιδιαίτερα σημαντική η συνεισφορά των εργαλείων που παρέχουν αυτήν τη δυνατότητα και κατατάσσονται στην κατηγορία **Οπτικής Αναλυτικής** (Visual Analytics). Μέσω αυτών τα Big Data γίνονται κατανοητά από τον ενδιαφερόμενο, δίχως να απαιτείται στις περισσότερες περιπτώσεις, τεχνική γνώση.

Βήματα Ανάλυσης Big Data



Βήματα Ανάλυσης Big Data

1 Απόκτηση- Καταγραφή

Οι άνθρωποι καθώς και όλες οι έξυπνες συσκευές αφήνουν συνέχεια το ίχνος τους. Αρκετές φορές με την θέληση και εν γνώσει τους άλλες φορές όχι. Δεδομένα από συνομιλίες, SMS, e-mail, κινήσεις στα social media, online αναζητήσεις και πληρωμές με κάρτες καθώς και γεωγραφικά δεδομένα συλλέγονται με χρήση των GPS. Όλα αυτά τα δεδομένα συλλέγονται και αποθηκεύονται στους server του αντίστοιχου παρόχου.

Βήματα Ανάλυσης Big Data

2

Καθαρισμός- Σχολιασμός- Αποθήκευση

Ένας πολύ σημαντικός παράγοντας είναι το κατά πόσο είναι εφικτή η ανάλυσή του (data analysis). Όταν στην ανάλυση εμπεριέχονται δεδομένα από διαφορετικές πηγές, τότε τα δεδομένα είναι ανομοιόμορφα και απαιτείται μεγάλη επένδυση σε χρόνο και πόρους ώστε αυτά να «καθαριστούν» και να συνδεθούν.

Οι τεχνικές data analysis μπορούν να ομαδοποιηθούν στις παρακάτω κατηγορίες:

- Data fusion (συγχώνευση δεδομένων): τεχνικές που χρησιμοποιούνται για να συγκεντρωθούν τα δεδομένα από διαφορετικές πηγές όπως γεωγραφικά δεδομένα που παράγονται από τα κινητά τηλέφωνα και από τα οχήματα που έχουν ενσωματωμένο GPS.
- Data mining («εξόρυξη» δεδομένων): οι τεχνικές που εφαρμόζονται ώστε να εντοπίζονται τυχόν ακολουθίες – μοτίβα (patterns) ή σχέση μεταξύ των δεδομένων ώστε να επιτυγχάνεται γρηγορότερη και αποτελεσματικότερη εξαγωγή συμπερασμάτων.
- Optimization (βελτιστοποίηση): τεχνικές που χρησιμοποιούνται ώστε να βελτιωθεί η απόδοση σύμφωνα με ένα πλήθος παραμέτρων.
- Visualization (απεικόνιση): οι τεχνικές που χρησιμοποιούνται ώστε να αναπαραχθούν εικόνες, διαγράμματα των αποτελεσμάτων της ανάλυσης των δεδομένων. Τεχνικές απεικόνισης χρησιμοποιούνται κατά την διάρκεια και μετά την ανάλυση των δεδομένων για να δώσουν μία πιο εύκολη και κατανοητή μορφή στα δεδομένα ή στην συνέχεια στις πληροφορίες.

Βήματα Ανάλυσης Big Data

3 Ενσωμάτωση
στα υπάρχοντα
data

Καινούργια σημαντικά στοιχεία ή νέες ιδέες προέρχονται από την ανάλυση ενός συνόλου δεδομένων, αλλά η αληθινή γνώση και πρόκληση είναι η ικανότητα εφαρμογής αναλυτικών μεθόδων και τεχνικών πάνω σε διαφορετικά δεδομένα από πολλές πηγές. Με την χρήση τεχνικών συγχώνευσης δεδομένων ανομοιόμορφα δεδομένα ενώνονται από διάφορες πηγές δημιουργώντας ή βελτιώνοντας τα υπάρχοντα αποτελέσματα.

Βήματα Ανάλυσης Big Data

4 Απεικόνιση-Ανάλυση

Παραδοσιακές στατιστικές τεχνικές ή μοντέλα απεικόνισης δεν είναι αποτελεσματικά στην επίλυση προβλημάτων Big Data εξαιτίας της πολυπλοκότητας και των χαρακτηριστικών των δεδομένων.

Αντιθέτως, προσεγγίσεις με τεχνικές εξόρυξης δεδομένων και απεικόνισης είναι πιο αποδοτικές στην μελέτη των BD. Οι τεχνικές εξόρυξης δεδομένων δεν βασίζονται σε προκαθορισμένες υποθέσεις ή δεν χρησιμοποιούν συγκεκριμένα ερωτήματα για να αναλύσουν ένα σύνολο δεδομένων. Ακόμη, προβάλλουν μοτίβα τα οποία ενδεχομένως θα οδηγήσουν σε κάποιο αποτέλεσμα. Οι αλγόριθμοι της εξόρυξης δεδομένων χρησιμοποιούν ορισμένες από τις παρακάτω διαδικασίες:

- Ταξινόμηση (classification): ταξινόμηση των δεδομένων ή των συμβάντων σύμφωνα με γνωστές κατηγορίες.
- Ομαδοποίηση (clustering): ομαδοποίηση δεδομένων που ακολουθούν παρόμοια μοτίβα.
- Μοντέλα πρόβλεψης (prediction): εντοπισμός πιθανών σχέσεων μεταξύ των στοιχείων βάση μοντέλων πρόβλεψης ή ανάλυση παλινδρόμησης (regression analysis).
- Σχέσεις μεταξύ των δεδομένων (association): εντοπισμός σχέσεων μεταξύ των τιμών των δεδομένων από ένα ή πολλά σύνολα δεδομένων.
- Εντοπισμός ανωμαλιών (anomaly detection): εντοπισμός των ακραίων τιμών ή περιπτώσεις διακοπής μίας ακολουθίας μέσα σε ένα σύνολο δεδομένων.
- Περίληψη: καταγραφή των διαφόρων χαρακτηριστικών που εμφανίζονται σε ένα ή περισσότερα σύνολα δεδομένων.

Βήματα Ανάλυσης Big Data

5

Ερμηνεία-
διάδοση
αποτελεσμάτων

Με την απεικόνιση των αποτελεσμάτων σε πραγματικό χρόνο δίνεται η δυνατότητα στους πολίτες αλλά και στις αρχές τις κάθε πόλης λήψης πιο αποτελεσματικών και λιγότερο ζημιογόνων αποφάσεων.

Μελετώντας κάποιος τα real-time δεδομένα από τα γραφήματα αποκτά γνώση για το παρόν αλλά ταυτόχρονα αντιλαμβάνεται και μελλοντικές καταστάσεις.

edureka!

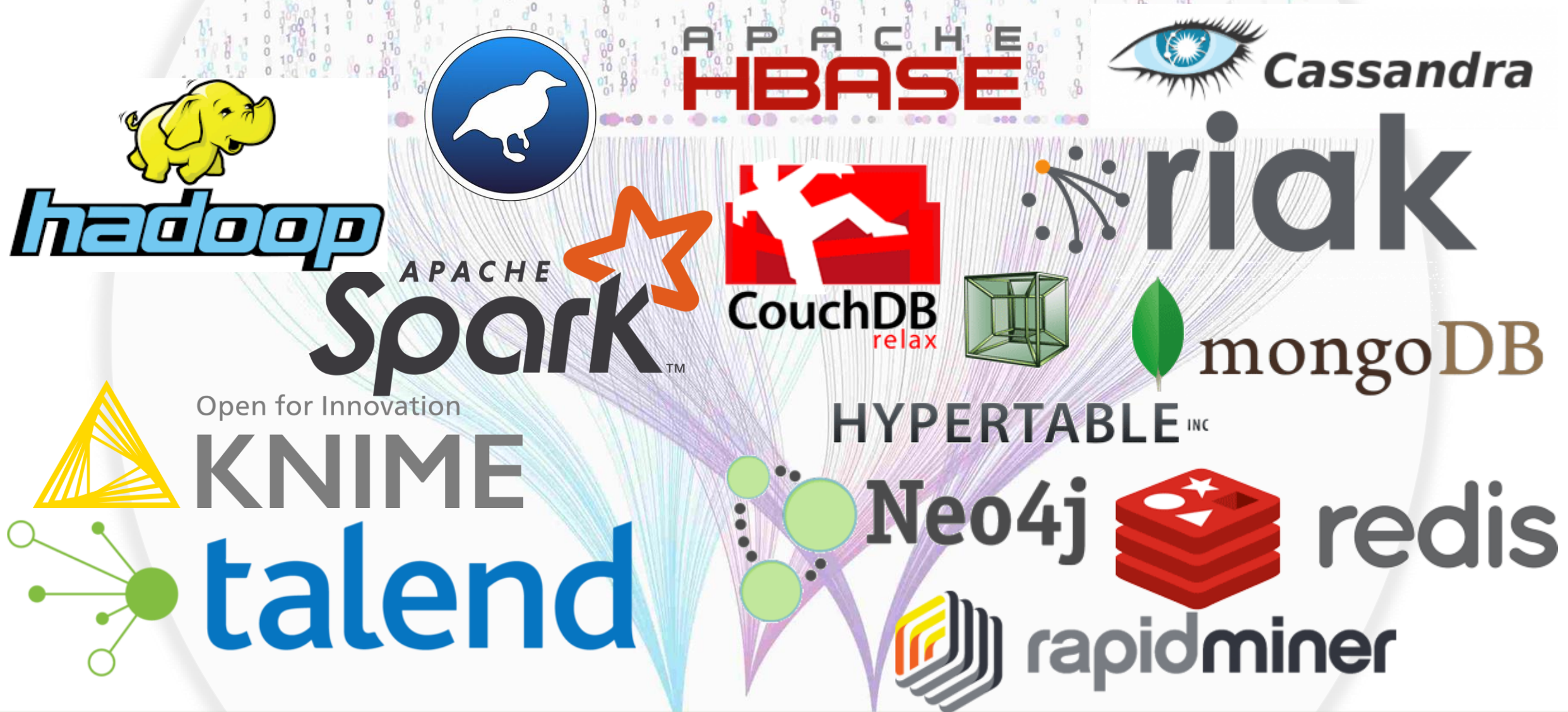




©Simplilearn. All rights reserved.



Εργαλεία Συλλογής, Διαχείρισης, Ανάλυσης Big Data



Γλώσσες Προγραμματισμού



Προβλήματα των Big Data

Προστασίας της ιδιωτικής ζωής και της ιδιοκτησίας των δεδομένων.

Το συγκεκριμένο πρόβλημα ανακύπτει από ορισμένες δυσκολίες των υπολογιστικών συστημάτων να ορίσουν τις πληροφορίες που θεωρούνται προσωπικά δεδομένα.

Δυνατότητες **επηρεασμού** της καταναλωτικής συμπεριφοράς.

Εκβιασμό, διακρίσεις, παρενόχληση, οικονομικές βλάβες.

Ένα άλλο πρόβλημα με τα Big Data είναι ότι μπορεί να είναι **παραπλανητικές** οι πληροφορίες που παρουσιάζουν εξαιτίας των αρνητικών συσχετίσεων τους, οι οποίες όχι μόνο δεν συνεισφέρουν στην πληροφόρηση αλλά αποτελούν βασικό εμπόδιο.

Η λειτουργία των Big Data απαιτεί κατάλληλες υποδομές και υπολογιστικά συστήματα τα οποία καταλαμβάνουν **μεγάλο χώρο** και απαιτούν αντίστοιχα **μεγάλη ισχύ** και **μεγάλη δικτύωση**. Η εκθετική αύξηση των δεδομένων έχει οδηγήσει σε αύξηση των απαιτήσεων ενέργειας καθιστώντας τα Big Data ως ιδιαίτερα **ενεργοβόρα**.



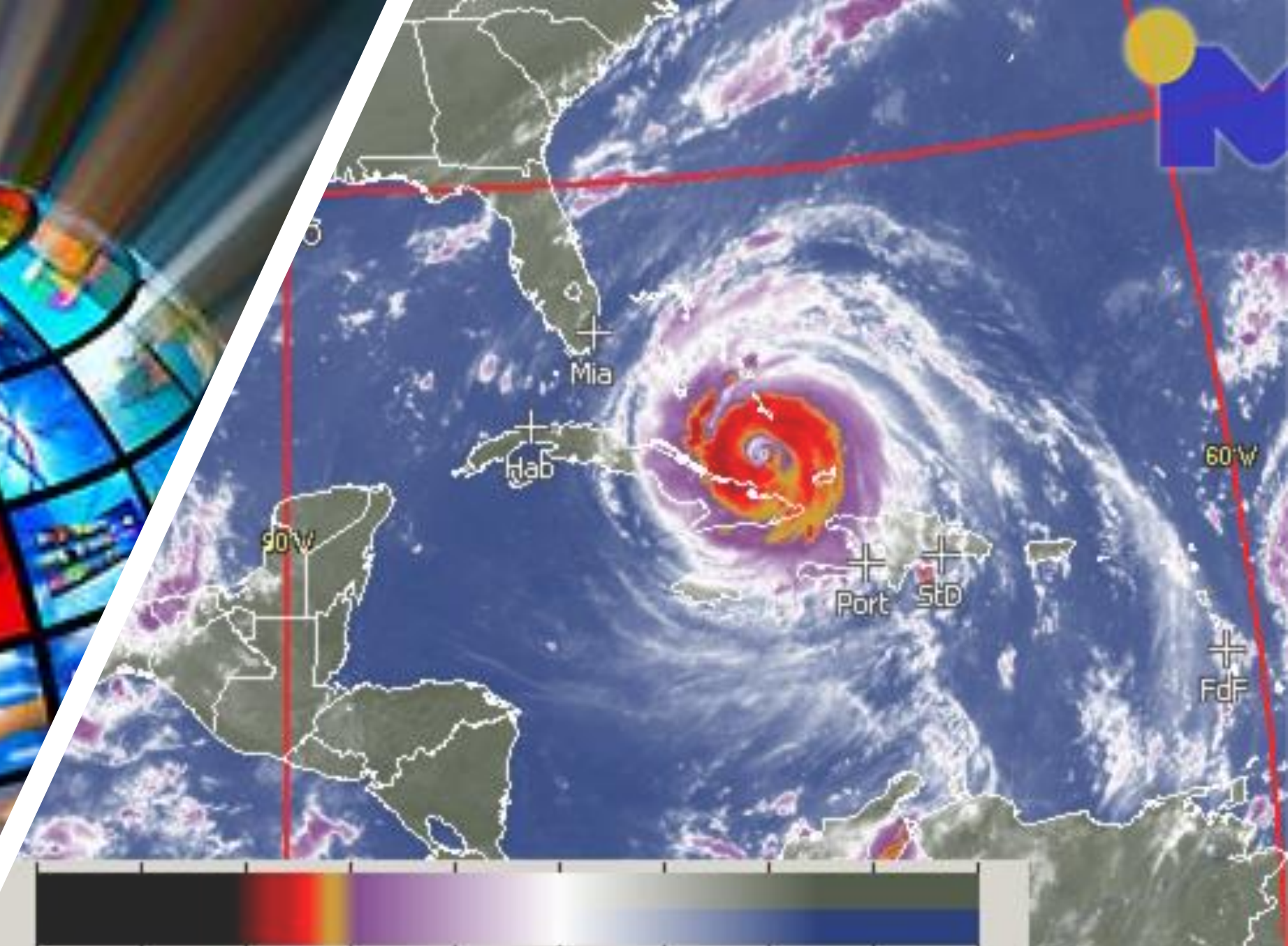
ΕΦΑΡΜΟΓΕΣ big data

The eBay logo is displayed in its characteristic multi-colored font (red, blue, yellow, green) against a white background.A dark blue smartphone graphic is shown on the left side of the image. The text "BEST SMARTPHONE TO BUY" is written on it in a light blue, sans-serif font.

**BEST
SMARTPHONE
TO BUY**

A dark grey arrow-shaped graphic pointing to the right, containing the text "ΤΗΛΕΠΙΚΟΙΝΩΝΙΕΣ" in white, bold, uppercase letters.

ΤΗΛΕΠΙΚΟΙΝΩΝΙΕΣ





ΔΙΑΔΙΚΤΥΟ



BANK

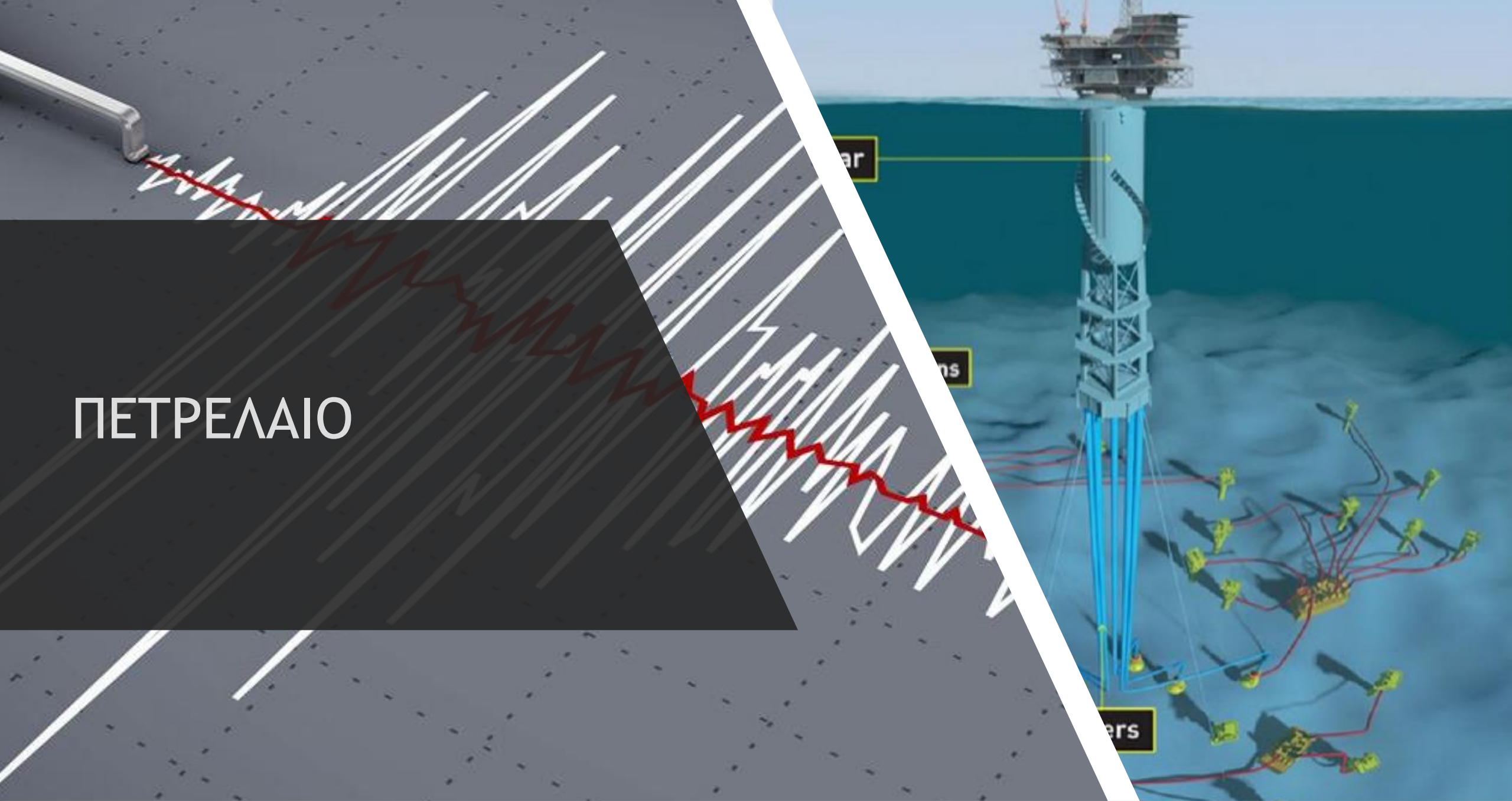
ΟΙΚΟΝΟΜΙΑ



MARKETING



ΠΕΤΡΕΛΑΙΟ



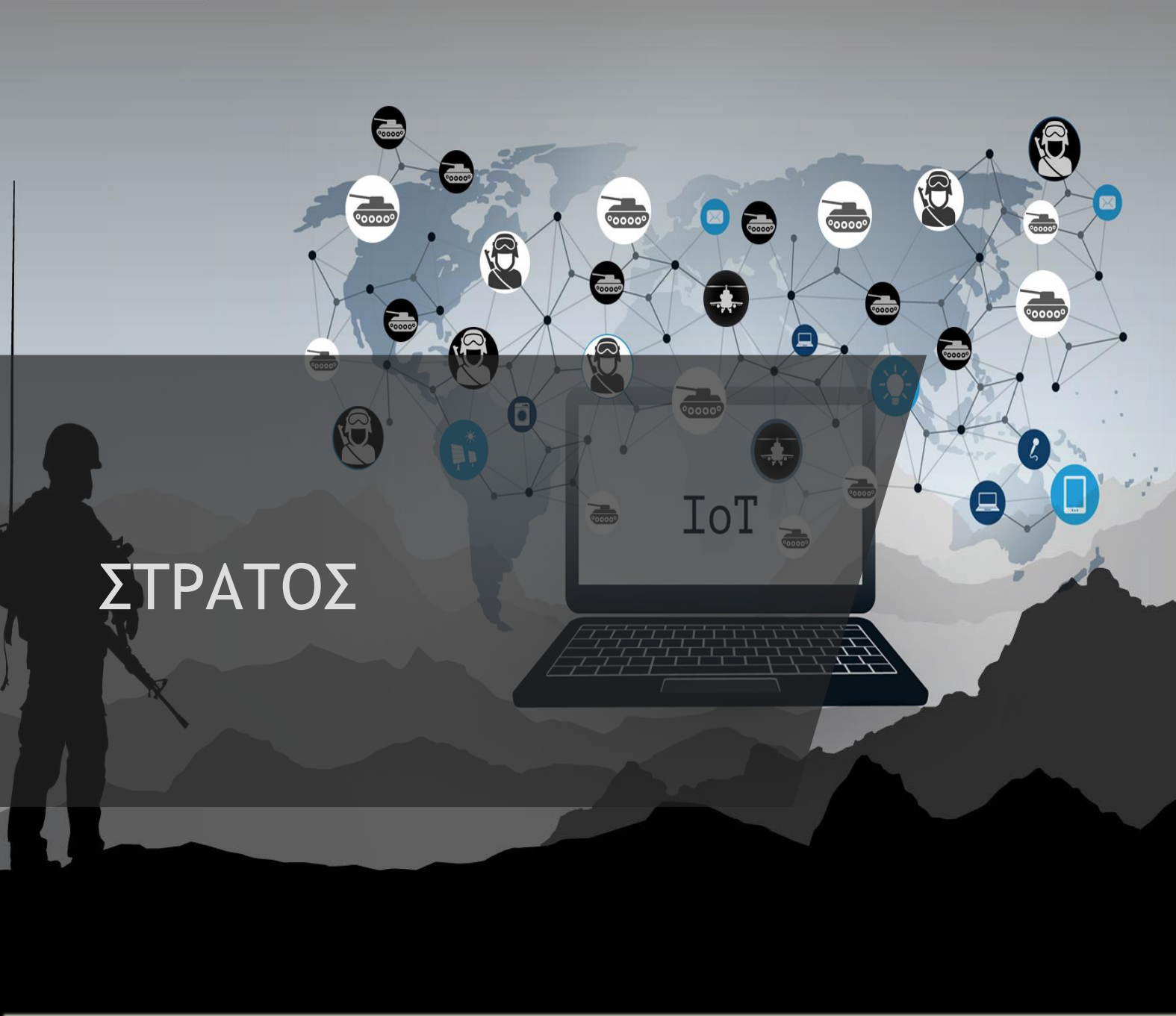


ΕΚΠΑΙΔΕΥΣΗ





ΠΛΟΗΓΗΣΗ



AMERICA'S ARMY: STRENGTH OF THE NATION

Extending Cloud-Enabled Analytics

Continuous asynchronous processes enrich data using the Cloud

Aggressively Index

Statistical Analysis of "Big Data" Enables Change Detection + Population-Centric Analysis

All-Source Analytics for 'Big Data' with Advances in Geospatial Indexing, Voice Index and Search, Biometric Entity Management, Motion Imagery Tracks, Multi-INT Visualization, Collection Management, Support for Mobile Devices, Powerful Compute Platforms, Multi-Level Security,

Mobile





Εκτιμάται ότι περίπου έως και το 30% των νοσοκομειακών λοιμώξεων μπορούν να προληφθούν με κατάλληλα προγράμματα πρόληψης. Η αξιοποίηση των Big Data σε πραγματικό χρόνο είναι απαραίτητη ώστε να ενσωματωθούν γονιδιωμικά με επιδημιολογικά δεδομένα, έτσι ώστε να επιτευχθεί η απαραίτητη πρόληψη και να μειωθεί το μέγεθος του φαινομένου.



Υπάρχουν συσκευές που επιτρέπουν την παρακολούθηση του ασθενούς σε πραγματικό χρόνο, όπως τα «έξυπνα» κινητά τηλέφωνα, ιατρικά όργανα που μπορεί να διαθέτει ο ασθενής (πιεσόμετρο, μετρητής σακχάρου κ.α.) ή ακόμη και πιο εξειδικευμένα που υπάρχουν στις κινητές μονάδες **τηλεϊατρικής** μέσω των οποίων η τεχνολογία των Big Data έχει επιτρέψει να διαμοιράζονται και να αναλύονται αυτοματοποιημένα σε πραγματικό χρόνο τα δεδομένα που συλλέγονται.





Η χρήση των Big Data, μπορεί να συνδράμει στην Ιατρική και φαρμακευτική **έρευνα**, καθώς επιτρέπει:

- Ανακάλυψη κρυφών σχέσεων μεταξύ ετερογενών δεδομένων.
- Ανάπτυξη μοντέλων πρόγνωσης της εξέλιξης των ασθενειών και έλεγχος.
- Ανάλυση δεδομένων μεγάλων πληθυσμών, ώστε να ελέγχονται σε πραγματικό χρόνο πιθανά μοντέλα εξατομικευμένης καταλληλότητας φαρμάκων.

