

STUDIU PRIVIND PERSPECTIVELE BIG DATA ȘI BIG DATA *ANALYTICS*

TITI PARASCHIV, VIOREL IULIAN TĂNASE,
ANA MARIA COSTIN

Study on Big Data and Big Data Analytics Perspectives. Researchers now realize the importance of using historical data to support strategy decisions. It has been shown through case studies that "usually more data beat good algorithms". With this statement, the researchers realized that they can choose to invest more in processing larger data sets than investing in expensive algorithms. The large amount of data is better used due to possible correlations on a larger amount, correlations that can never be found if the data is analyzed on separate or small sets. A larger amount of data provides a better solution, but working with them is a challenge due to processing limitations.

Keywords: Big Data, Big Data Analytics, machine learning, Python, database, internet, Hadoop project.

1. Introducere

Astăzi, Internetul reprezintă un spațiu virtual unde se adaugă zilnic cantități mari de informații. Infografia IBM Big Data Flood arată că în universul digital există astăzi 2,7 Zettabytes de date. De asemenea, conform acestui studiu, există 100 de Terabyti actualizați zilnic prin activități pe rețelele de socializare ceea ce duce la o estimare a 35 Zettabytes de date generate anual până în 2020. Doar pentru a avea o idee despre cantitatea de date generate, un zettabyte (ZB) este de 10^{21} octeți, ceea ce înseamnă 10^{12} GB.

Astăzi trăim într-o societate informațională și ne îndreptăm spre o societate bazată pe cunoaștere, odată cu cea de-a patra revoluție industrială care s-a impus începând cu 2016. Pentru a extrage cunoștințe mai bune, avem nevoie de o cantitate mai mare de date. Societatea informației este o societate în care informația joacă un rol major în activitatea economică, culturală și politică. În societatea cunoașterii avantajul competitiv este obținut prin înțelegerea informațiilor și precizarea evoluției faptelor, pe baza datelor. La fel se întâmplă și cu Big Data. Fiecare organizație trebuie să colecteze un set mare de date pentru a-și susține decizia și extrage corelații prin analiza datelor ca bază pentru decizii. Vom defini conceptul Big Data, importanța acestuia și perspective diferite asupra utilizării acestuia. În plus, vom sublinia importanța analizei Big Data și vom arăta cum analiza Big Data va îmbunătăți deciziile în viitor.

2. Caracteristicile Big Data

Termenul „Big Data” a fost introdus pentru prima dată de către Roger Magoulas de la mass-media O'Reilly, în 2005, pentru a defini o cantitate mare de date pe care tehnicile tradiționale de gestionare a datelor nu le pot gestiona și prelucra din cauza complexității și dimensiunii acestor date. Un studiu privind evoluția Big Data ca cercetare și subiect științific arată că termenul „Big Data” a fost prezent în cercetare, începând cu anii 1970, dar a fost subiect în publicațiile științifice din 2008.[2] În prezent, conceptul Big Data este tratat din mai multe perspective care acoperă implicațiile sale în multe domenii.

Conform MiKE 2.0, standardul open source pentru managementul informațiilor Big Data este definit prin dimensiunea sa, cuprinzând o colecție mare, complexă și independentă de seturi de date, fiecare cu potențial de interacțiune. În plus, un aspect important al Big Data este faptul că nu poate fi tratat cu tehnici standard de gestionare a datelor, datorită inconsistenței și imprevizibilității combinațiilor posibile.

În viziunea IBM, Big Data are patru caracteristici: 1. Volum: se referă la cantitatea de date culese de o companie. Aceste date trebuie utilizate în continuare pentru a obține cunoștințe importante; 2. Viteză: se referă la perioada în care Big Data poate fi procesată. Unele activități sunt foarte importante și au nevoie de răspunsuri imediate, de aceea procesarea rapidă maximizează eficiența; 3. Varietate: se referă la tipul de date pe care Big Data le poate cuprinde. Aceste date pot fi structurate sau nestructurate; 4. Veridicitate: se referă la gradul în care un lider are încredere în informațiile utilizate pentru a lua o decizie. Așadar, obținerea corelațiilor corecte în Big Data este foarte importantă pentru faceri sau cercetare.

În plus, în IT-ul *Gartner Glossary Big Data* sunt definite drept active de informații de volum, viteză și varietate ridicate, care necesită forme inovatoare de procesare a informațiilor, eficiente din punct de vedere al costurilor, pentru o bună predictibilitate necesară în luarea deciziilor.

Potrivit președintelui Ed Dumbill în cadrul Conferinței Strata O'Reilly, Big Data poate fi descrisă drept date care depășesc capacitatea de procesare a sistemelor convenționale de baze de date. Datele sunt prea mari, se mișcă prea repede sau nu se încadrează în structurile arhitecturilor de baze de date. Pentru a obține valoare din aceste date, trebuie să alegeți o modalitate alternativă de procesare a acestora.

Într-o definiție mai simplă, considerăm că Big Data este o expresie care cuprinde diferite seturi de date foarte mari, extrem de complexe, nestructurate, organizate, stocate și procesate folosind metode și tehnici specifice utilizate pentru procesele de afaceri. Există o mulțime de definiții cu privire la Big Data care circulă în lume, dar considerăm că cea mai importantă este cea pe care fiecare lider le oferă datelor companiei sale. Modul în care este definit Big Data are implicații în strategia unei afaceri.

3. Importanța Big Data

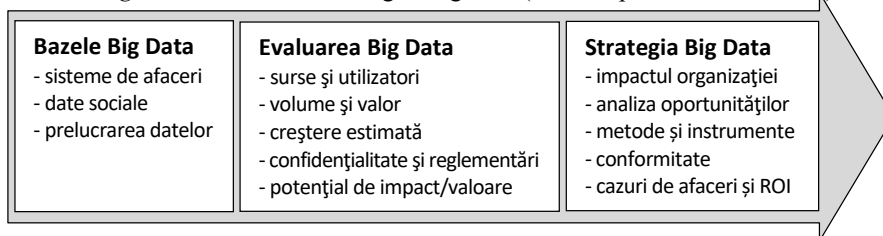
Principala importanță a Big Data constă în potențialul de a îmbunătăți eficiența în contextul utilizării unui volum mare de date, de tipuri diferite. Dacă Big Data este

definit în mod corespunzător și utilizat în consecință, organizațiile pot avea o viziune mai bună asupra activității lor, conducând astfel la eficiență în diferite domenii precum vânzările, fabricație, marketing, cercetare tehnică, studii de piață. Big Data poate fi utilizat eficient în următoarele domenii: - tehnologia informației, pentru a îmbunătăți securitatea și rezolvarea problemelor, analizând modelele din jurnalele existente; - în serviciul clienților prin utilizarea informațiilor de la centrele de apeluri pentru a obține modelul de client și, astfel, a spori satisfacția clienților prin personalizarea serviciilor; - îmbunătățirea serviciilor și produselor prin utilizarea conținutului de social media. Cunoșcând preferințele clienților potențiali, compania își poate modifica produsul pentru a se adresa unei zone mai mari de oameni; - detectarea fraudei în tranzacțiile online pentru orice industrie; - evaluarea riscurilor prin analiza informațiilor din tranzacțiile de pe piața financiară. În acest studiu voi analiza potențialul Big Data și puterea care poate fi activată prin analiza Big Data. Pentru a determina cea mai bună strategie pentru o companie, este esențial ca datele pe care contați să fie analizate în mod corespunzător. De asemenea, intervalul de timp al acestei analize este important, deoarece unele dintre ele trebuie efectuate frecvent pentru a determina orice schimbare în mediul de afaceri.

Un alt aspect este reprezentat de noile tehnologii care sunt dezvoltate în fiecare zi. Având în vedere faptul că Big Data este nouă pentru organizații, este necesar ca aceste organizații să învețe cum să utilizeze noile tehnologii, imediat ce sunt pe piață. Acesta este un aspect important care va aduce un avantaj competitiv unei afaceri.

Nevoia de specialiști este o provocare pentru Big Data. Conform studiului lui McKinsey, Big Data reprezintă următoarea frontieră pentru inovare, este nevoie de până la 190.000 de lucrători cu experiență analitică și 1,5 milioane de manageri de alfabetizare a datelor numai în Statele Unite. Aceste statistici sunt o dovadă că pentru o companie să ia inițiativa Big Data trebuie să angajeze experți sau să instruiască angajați existenți pe noul domeniu. Confidențialitatea și securitatea sunt, de asemenea, provocări importante pentru Big Data. Deoarece Big Data constă într-o cantitate mare de date complexe, este foarte dificil pentru o companie să sorteze aceste date pe niveluri de confidențialitate și să aplice securitatea respectivă. În plus, multe companii din zilele noastre fac afaceri între țări și continente, iar diferențele în legile privind confidențialitatea sunt considerabile și trebuie luate în considerare la începerea inițiativei Big Data. Pentru ca o organizație să obțină un avantaj competitiv în urma utilizării Big Data, trebuie să aibă grijă de toți factorii atunci când se implementează. O opțiune de dezvoltare a unei strategii Big Data este prezentată mai jos. În plus, pentru a aduce capabilități complete Big Data, fiecare companie trebuie să ia în considerare propriile caracteristici ale afacerii.

Figura 1: Dezvoltarea strategiei Big Data (Sursa <http://www.navint.com>)



4 Analiza Big Data

Lumea de astăzi este construită pe temelia datelor. Viețile de astăzi sunt afectate de capacitatea companiilor de a dispune, interoga și gestiona date. Dezvoltarea infrastructurii tehnologice este adaptată pentru a ajuta la generarea de date, astfel încât toate serviciile oferite să poată fi îmbunătățite pe măsură ce sunt utilizate. Ca exemplu, internetul a devenit astăzi o platformă uriașă de colectare de informații datorită rețelelor de socializare și serviciilor online. În orice minut sunt adăugate date. Explozia de date nu mai poate fi măsurată în gigabyte, deoarece datele sunt mai mari, sunt utilizate etabytes, exabytes, zettabytes și yottabytes.

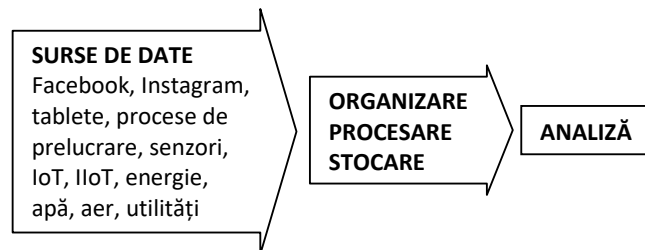
Pentru a gestiona volumul uriaș de date nestructurate stocate, au apărut fenomenele „Big Data”. Este important să se considere că în sectorul comercial Big-Data a fost adoptat mai rapid în industriile bazate pe date, precum serviciile financiare și telecomunicațiile, despre care se poate afirma că au cunoscut o creștere mai rapidă a volumului de date în comparație cu alte sectoare de piață. La început, Big Data a fost văzută ca o modalitate de a reuși să reducă din costurile de gestionare a datelor. Acum, companiile se concentrează asupra potențialului de creare de plusvaloare. Pentru a beneficia de informații suplimentare, este necesară evaluarea capacităților de analiză și procesare ale Big Data.

Pentru a transforma datele mari într-un avantaj pentru cercetare și afaceri, companiile examinează modul în care gestionează datele din centrul de date. Datele sunt preluate dintr-o multitudine de surse, atât din interior, cât și din afara organizației. Poate include conținut din videoclipuri, date sociale, documente și date generate de mașini, dintr-o varietate de aplicații și platforme. Întreprinderile au nevoie de un sistem optimizat pentru achiziționarea, organizarea și încărcarea acestor date nestructurate în bazele lor de date, pentru a putea fi redată și analizată în mod eficient. Analiza datelor trebuie să fie profundă, rapidă și în conformitate cu obiectivele companiei.

Scalabilitatea soluțiilor de date mari din centrele de date este o caracteristică esențială. Datele sunt vaste astăzi și vor crește doar mai mult. Dacă un centru de date nu poate face față decât nivelurilor de date preconizate pe termen scurt sau mediu, companiile vor alocă resurse pentru actualizări și modernizări ale sistemului. Prin urmare, planificarea anterioară și scalabilitatea sunt importante.

Pentru a lua fiecare decizie după dorință, este necesar să aducem rezultatele descoperirii cunoștințelor în activitatea de bază și, în același timp, să urmărim orice impact în diferitele tablouri de bord, rapoarte și analize de excepție monitorizate. Noile cunoștințe descoperite prin analiză pot avea, de asemenea, o influență asupra strategiei de afaceri, strategiei CRM și strategiei financiare.

Figura 2: Big Data Management



Până la jumătatea anului 2009, peisajul de gestionare a datelor era simplu: sistemele de procesare online a tranzacțiilor (OLTP) (în special bazele de date) au susținut procesele de afaceri ale întreprinderii; Stocurile de date operaționale (ODS) au acumulat tranzacțiile de afaceri pentru a sprijini raportarea operațională, iar depozitele de date despre întreprinderi (EDW) au acumulat și transformat tranzacțiile de afaceri pentru a sprijini decizia operațională și strategică.

Managementul Big Data se bazează pe captarea și organizarea datelor relevante. Analiza datelor presupune să înțelegă ce s-a întâmplat, de ce și să prezicem ce se va întâmpla. O analiză mai profundă înseamnă noi metode analitice pentru perspective mai aprofundate. Analizele de date mari și proiectul open source Apache Hadoop apar rapid ca soluție preferată pentru tendințele de business și tehnologie care perturbă tradiționalul management al datelor și procesarea datelor. Întreprinderile pot obține un avantaj competitiv prin faptul că sunt adoptatori timpurii ai analizelor de date mari. Chiar dacă analiza datelor mari poate fi dificilă din punct de vedere tehnic, întreprinderile nu ar trebui să întârzie implementarea. Pe măsură ce proiectele Hadoop se maturizează și suportul instrumentelor de informații de afaceri (BI) se îmbunătățește, complexitatea de implementare a analizelor de date mari se va reduce, dar avantajul competitiv al adoptatorilor timpurii va scădea. Riscul de implementare a tehnologiei poate fi redus prin adaptarea principiilor și modelelor arhitecturale existente la noile tehnologii și schimbarea cerințelor, mai degrabă decât prin respingerea acestora. [10]

Analiza de date mari poate fi diferențiată de arhitecturile tradiționale de prelucrare a datelor de-a lungul unui număr de dimensiuni: - Viteza de luare a deciziilor fiind foarte importantă pentru factorii de decizie; - Complexitatea procesării, deoarece ușurează procesul de luare a deciziilor; - Volumele de date tranzacționale care sunt foarte mari; - Datele pot fi structurate și nestructurate; - Flexibilitatea procesării / analizei constând în cantitatea de analiză care poate fi efectuată pe aceasta; - Concurență.

Inițiativa de analiză a datelor de mari dimensiuni ar trebui să fie un proiect comun care implică atât IT, cât și business. IT ar trebui să fie responsabil pentru implementarea instrumentelor de analiză a datelor de mari dimensiuni și pentru implementarea unor practici solide de gestionare a datelor. Ambele grupuri ar trebui să înțeleagă că succesul va fi măsurat prin valoarea adăugată prin îmbunătățirile de afaceri care sunt aduse de inițiativă.

Managementul datelor și analiza Big Data, Oracle oferă sisteme proiectate ca soluții de date mari, precum Oracle Big Data Appliance, Oracle Exadata și Oracle Exalytics. Soluțiile Big Data combină cele mai bune instrumente pentru fiecare parte a problemei. Instrumentele tradiționale de business intelligence se bazează pe baze de date relaționale pentru stocarea și execuția de interogare și nu au vizat Hadoop. Oracle BI combinat cu Oracle Big Data Connectors. Arhitectura presupune încărcarea elementelor cheie de informații din surse Big Data în SGBD. Conectoarele Oracle Big Data, ETL și Hadoop sunt conștiente de ETL, cum ar fi ODI, care oferă capacitățile necesare de integrare a datelor. Avantajele cheie sunt

că investițiile și abilitățile de business intelligence sunt puse la punct, există idei de la Big Data consumabile pentru utilizatorii de afaceri, există date Big combinate cu date de aplicație și OLTP. În figura 3 prezentăm comparativ modelele tradiționale și cele Big Data.

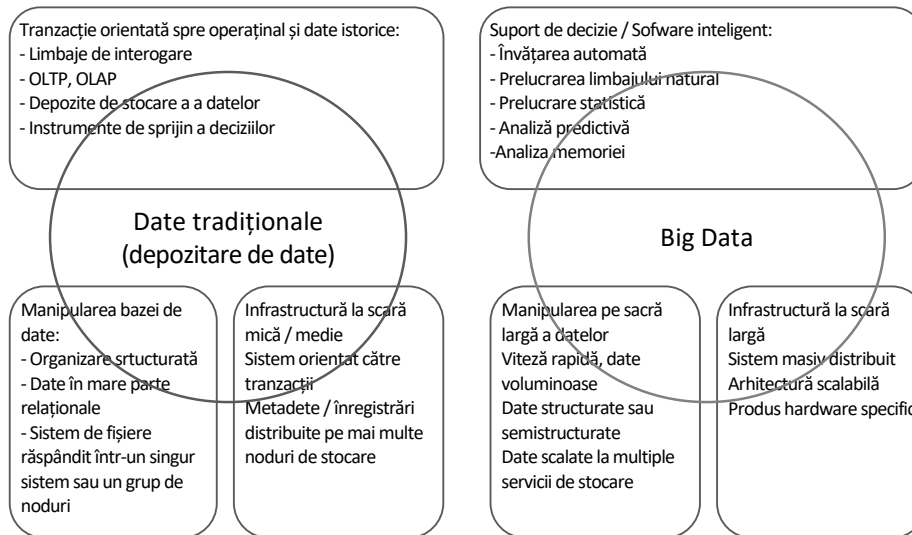


Figura 3: Modelele Big Data Vs Date tradiționale (Data Warehousing)

Big Data oferă multe oportunități pentru informații detaliate prin extragerea datelor. Descoperă relațiile dintre sentimentele sociale și datele despre vânzări. Previzionați problemele produsului pe baza datelor senzorilor de diagnosticare generate de produsele din domeniu. De fapt, problemele semnal-zgomot semnifică adesea analitice profunde pentru o perspectivă minieră ascunsă în zgomot, fiind esențiale, deoarece multe forme de Big Data nu sunt pur și simplu consumabile în forma brută „Big Data” este un Management Data

Oportunitatea de piață și Analytics determinată de noile cerințe ale pieței. Analiza în baza de date – Data Mining, sunt folosiți conectori Big Data pentru a combina date Hadoop și DBMS pentru analize profunde. De asemenea, este necesară reutilizarea abilităților SQL pentru aplicarea tehnicilor de exploatare a datelor mai aprofundate sau reutilizarea abilităților pentru analiza statistică. Totul se referă la Big Data în loc de date la scară RAM. Așa se face învățarea predictivă a relațiilor dintre conceptele de cunoaștere și evenimentele de afaceri.

Big Data prezintă o oportunitate semnificativă de a crea o nouă valoare din datele gigantice. Este important să se stabilească procedurile de guvernare adecvate pentru a gestiona dezvoltarea și implementările pe toată durata de viață a tehnologiei și a datelor. Nerespectarea implicațiilor pe termen lung ale dezvoltării va duce la probleme de productivitate și escaladarea costurilor

În această privință, costul stocării fizice a unor cantități mari de date este redus dramatic prin simplitatea prin care datele pot fi încărcate într-un cluster Big-Data, deoarece nu mai este necesar un strat ETL complex văzut în niciun Data Warehouse mai tradițional. Clusterul în sine este, de asemenea, construit în mod obișnuit folosind hardware-ul de mărfuri low cost, iar analiștii sunt liberi să scrie cod în aproape orice limbă contemporană prin intermediul API-ului de streaming disponibil în Hadoop.

Logica de afaceri folosită într-un flux ETL pentru a simboliza un flux de date și pentru a aplica standardele de calitate a datelor trebuie să fie codată (de obicei folosind Java) în cadrul fiecărui program Map-Reduce care prelucrează datele și orice modificări în sintaxa sursă sau semantică. Deși nodurile de stocare dintr-un cluster Hadoop pot fi construite folosind servere x86 cu mărfuri cu costuri reduse, nodurile master (Nume Nume, Nod secundar Nume și Job Tracker) care necesită niveluri mai ridicate de reziliență pentru a fi încorporate în servere, dacă se evită dezastrul. Operațiunile Map-Reduce, de asemenea, generează o mulțime de chat-uri de rețea, astfel încât o rețea privată rapidă este recomandată. Aceste cerințe se combină pentru a adăuga un cost semnificativ la un grup de producție utilizat într-un cadru comercial.

Capabilitățile de compresie din Hadoop sunt limitate datorită structurii blocului HDFS și necesită o înțelegere a tehnologiei de date și compresie pentru a implementa adăugarea la complexitatea implementării cu impact limitat asupra volumelor de stocare.

Alte aspecte care trebuie luate în considerare includ costul adevărat al dreptului de proprietate al clusterilor de preproducție și producție, cum ar fi construirea și întreținerea designului clusterilor înșiși, trecerea la producția codului Map-Reduce la clusterul de producție în conformitate cu procedurile operaționale standard și dezvoltarea acestor proceduri.

Oricare ar fi costul adevărat al Big-Data în comparație cu o abordare relațională de stocare a datelor, este important ca dezvoltarea strategiei Big-Data să fie făcută în mod conștient, înțelegând adevărata natură a costurilor și complexității infrastructurii, practicilor și procedurilor puse în aplicare. la loc.

5. Software de analiză a Big Data

În prezent, tendința este ca întreprinderile să își reevalueze abordarea privind stocarea, gestionarea și analiza datelor, deoarece volumul și complexitatea datelor crește atât de rapid, iar contabilitatea datelor nestructurate constituie 90% din datele de astăzi.

În fiecare zi, sunt create 2,5 octeți de cvintilete de date - atât de mult încât 90% din datele de astăzi au fost create doar în ultimii doi ani. Aceste date provin din diverse surse, cum ar fi: senzori folosiți pentru a colecta informații despre climă, postări pe site-urile de socializare, imagini și videoclipuri digitale, înregistrări de tranzacții de achiziție și semnale GPS ale telefonului mobil, jurnalele web și software, camere de luat vederi, dispozitive mobile de detectare a

informațiilor, tehnologii senzoriale aeriene și genomică. Aceste date sunt denumite date mari. Sistemele moștenite vor rămâne necesare pentru sarcini de muncă specifice de mare valoare, cu volum redus și vor complimenta utilizarea Hadoop - optimizând structura de gestionare a datelor din organizație, introducând volumele de lucru Big Data în sistemele potrivite.

Așa cum a fost menționat în Introducere Big Data are patru dimensiuni: volum, viteză, varietate și veridicitate:

- Volum: Întreprinderile sunt neplăcute cu date în creștere de tot felul, amasând cu ușurință terabyte - chiar și petabytes - de informații (de exemplu, transformă 12 terabyte de Tweet-uri create în fiecare zi în analiză îmbunătățită a sentimentului produsului; consum);

- Viteză: pentru procesele sensibile la timp, precum înlăturarea fraudei, fluxurile de date mari trebuie analizate și utilizate pe măsură ce se orientează către organizații pentru a maximiza valoarea informațiilor (de exemplu, cercetarea a 5 milioane de evenimente comerciale create în fiecare zi pentru a identifica o fraudă potențială ; analizați 500 de milioane de înregistrări zilnice de detalii ale apelurilor în timp real pentru a prezice rapid creșterea clienților).

-Varietate: Date mari constă în orice tip de date - date structurate și nestructurate, cum ar fi text, date senzor, audio, video, fluxuri de clicuri, fișiere jurnal și multe altele. Analiza tipurilor de date combinate aduce un aspect nou pentru probleme, situații etc. (de exemplu, monitorizați 100 de fluxuri video în direct de la camerele de supraveghere până la obiectivele de interes; exploatați creșterea datelor cu 80% din imagini, video și documente pentru a îmbunătăți satisfacția clienților);

Veracitate: Deoarece unul dintre cei trei lideri de afaceri nu au încredere în informațiile pe care le folosesc pentru a lua decizii, stabilirea încrederii în datele mari prezintă o provocare uriașă pe măsură ce varietatea și numărul de surse crește.

Datele mari reprezintă mai mult decât o simplă problemă de mărime; este o oportunitate de a găsi informații cu privire la tipurile de date și conținut noi și emergente, de a face întreprinderile mai agile și de a răspunde la întrebări care au fost considerate în prealabil. Întreprinderile care își construiesc soluția Big Data își pot permite să stocheze literalmente toate datele din organizația lor și să le păstreze toate online pentru interogare interactivă în timp real, informații de afaceri, analiză și vizualizare.

6. Concluzii

Această lucrare evidențiază modul în care accentul emergent pe Big Data semnalează apariția unei abordări centrate pe date a cercetării, în care eforturile de mobilizare, integrare, diseminare și vizualizare a datelor sunt privite ca fiind contribuții centrale la descoperire. Centrarea pe date evidențiază provocările implicate în colectarea, clasificarea și interpretarea lor, precum și conceptele, tehnologiile și instituțiile care realizează aceste procese. Instrumente precum dispozitivele de măsurare cu randament ridicat, conectate la IoT și aplicații pentru

sistemele mobile de comunicații generează, rapid, volume mari de date în format digital. În principiu, aceste date sunt disponibile imediat pentru diseminare prin intermediul platformelor internetului, care le pot face accesibile oricui are o conexiune de bandă largă, în câteva secunde. În realitate, accesul la date are implicații conceptuale, tehnice, juridice și etice. Instrumentele matematice și de calcul, dezvoltate pentru a analiza datele mari, sunt adesea opace în funcționarea și presupunerile lor, ducând la rezultate a căror semnificație științifică și credibilitate pot fi dificil de evaluat.

Îngrijorările privind Big Data țin de măsura în care aceasta se abate de la înțelegerea rațională bazată pe individ ca agenție (acțiune) și abilitățile cognitive (pe care se bazează filosofia contemporană a științei). Puterea oricărui set de date de a oferi cunoștințe constă în măsura în care acesta poate fi corelat cu alte seturi de date: aceasta conferă o valoare epistemică ridicată obiectelor digitale. Canalele de producție și diseminare a datelor, cum ar fi social media, baze de date guvernamentale și depozite de cercetare, funcționează într-o rețea globalizată, interconectată și distribuită, a cărei funcționare solicită o mare varietate de competențe și expertiză. Natura distribuită a procesului decizional, implicat în dezvoltarea de infrastructuri de date și analiză a Big Data, face imposibil ca o persoană să poată supraveghea calitatea, semnificația științifică și potențialul impact social al cunoștințelor produse.

Prin urmare, opinia autorilor este că, analiza datelor mari constituie instanța finală a unui sistem cognitiv distribuit care presupune îmbinarea metodelor științelor clasice cu cele ale științei datelor. Multe persoane, grupuri și instituții împart responsabilitatea interpretării conceptuale și rezultatelor sociale ale utilizării datelor. O provocare cheie pentru guvernanta Big Data este găsirea mecanismelor de repartizare a responsabilităților în cadrul acestei rețele complexe, astfel încât deciziile eronate și nejustificate - precum și acțiunile frauduloase, neetice, abuzive, discriminatorii sau greșite - să poată fi identificate, corectate și sancționate corespunzător. Prelucrarea și utilizarea datelor îi poate încuraja pe filosofi să evite abordările anistorice, necontextualizate în studiul probelor și, în schimb, să considere metodele, abilitățile, tehnologiile și practicile implicate în manipularea datelor mari, ca fiind cruciale pentru înțelegere și realizarea de cunoștințe empirice.

Anul 2016 este anul în care companiile încep să se orienteze spre utilizarea Big Data, întemeind o nouă știință numită știința datelor. Este anul în care s-a declanșat, în mod oficial, ce-a de-a patra revoluție industrială caracterizată prin transformarea digitală a societății. Acesta este motivul pentru care această lucrare prezintă conceptul Big Data și tehnologiile asociate pentru a înțelege mai bine beneficiile acestor noțiuni, concepte, tehnologii și metode care vor transforma cercetarea, societatea, economia și chiar ființa umană.

Bibliografie

- G. Noseworthy, Infographic: Managing the Big Flood of Big Data in Digital Marketing, 2012 <http://analyzingmedia.com/2012/infographic-hic-big-flood-of-big-data-in-digital-marketing>;
- H. Moed, The Evolution of Big Data as a Research and Scientific Topic: Overview of the Literature, 2012, ResearchTrends, <http://www.researchtrends.com>;
- MIKE 2.0, Big Data Definition, [http://mike2.openmethodology.org/wiki/Big Data Definition](http://mike2.openmethodology.org/wiki/Big_Data_Definition);
- P. Zikopoulos, T. Deutsch, D. Deroos, Harness the Power of Big Data, 2012, <http://www.ibmbigdatahub.com/blog/harness-power-big-data-book-excerpt>;
- Gartner, Big Data Definition, <http://www.gartner.com/it-glossary/big-data>;
- E. Dumhill, "What is big data?", 2012, <http://strata.oreilly.com/2012/01/what-is-big-data.html>;
- A Navint Partners White Paper, "Why is BIG Data Important?" May 2012, <http://www.navint.com/images/Big.Data.pdf>;
- Greenplum. A unified engine for RDBMS and Map Reduce, 2009. <http://www.greenplum.com/resources/mapreduce>;
- For Big Data Analytics There's No Such Thing as Too Big The Compelling Economics and Technology of Big Data Computing, White Paper, March 2012, By: 4syth.com, Emerging big data thought leaders;
- Big data: The next frontier for innovation, competition, and productivity. James Manyika, Michael Chui, Brad Brown, Jacques Bughin, Richard Dobbs, Charles Roxburgh, and Angela Hung Byers. McKinsey Global Institute. May 2011;
- Oracle Information Architecture: An Architect's Guide to Big Data, An Oracle White Paper in Enterprise Architecture August 2012; <http://bigdataarchitecture.com>;
- <http://www.oracle.com/us/corporate/presentations/1453796>;
- <http://www.informationweek.com/software/business-intelligence/sas-gets-hip-to-hadoop-for-big-data/240009035?pgno=2>;
- http://en.wikipedia.org/wiki/Apache_Hadoop.