

ABORDĂRI ȘI SOLUȚII SPECIFICE ÎN MANAGEMENTUL, GUVERNANȚA ȘI ANALIZA DATELOR DE MARI DIMENSIUNI (BIG DATA)

Vladimir Florian

vladimir@ici.ro

Gabriel Neagu

gneagu@ici.ro

Institutul Național de Cercetare-Dezvoltare în Informatică – ICI București

Rezumat: Conform documentelor cu caracter strategic la nivel național pentru perioada 2014-2020, domeniul TIC (Tehnologia Informației și Comunicațiilor) se remarcă prin dinamica sa competitivă în rândul sectoarelor de specializare inteligentă (SI) din economie și reprezintă în același timp o prioritate în cercetare-dezvoltare-inovare (CDI), datorită suportului pe care îl asigură pentru competitivitatea celorlalte sectoare SI. Lucrarea prezintă rezultatele analizei tematicii CDI privind managementul, guvernanta și analiza datelor de mari dimensiuni (Big Data), selectată prin prisma relevanței soluțiilor pe care le poate oferi pentru cele 10 sectoare SI. Principalele aspecte prezentate se referă la: explicitarea conceptelor de bază ale tematicii, evidențierea potențialului de impact pentru competitivitate și identificarea unor soluții care să illustreze acest potențial.

Cuvinte cheie: Big Data, guvernanta, știința datelor, date deschise, Analytics, Cloud, arhitectura de referință, specializare inteligentă.

Abstract: According to strategic documents at national level for the period 2014-2020, ICT (Information and Communication Technologies) is characterized by its competitive dynamics among sectors of smart specialization (SS) in the economy and is also a priority in research, development and innovation (RDI), due to the support it provides for the competitiveness of other sectors SI. The paper presents the results of analyzing the RDI topic on Management, Governance and Analytics of Big Data, which was selected through the relevance of the solutions it can offer for the 10 SS sectors. The main tackled issues regarding this topic are to explain its basic concepts, to emphasize its potential impact on economic competitiveness and to identify solutions that illustrate this potential.

Keywords: Big Data, governance, data science, open data, Analytics, Cloud, reference architecture, smart specialization.

1. Introducere

Strategia Națională privind Agenda Digitală pentru România 2020 [1] evidențiază rolul cercetării-dezvoltării-inovării (CDI) în tehnologiei informației și comunicațiilor (TIC) pentru susținerea dezvoltării economiei și societății, cu accent pe mediul de afaceri, prin identificarea „eCommerce, cercetare-dezvoltare și inovare în TIC” ca una din cele 4 priorități ale acestei strategii. În concordanță cu această prioritate, *Programul “Agenda Digitală pentru România”*, Secțiunea “Servicii electronice”, include proiectul „Cercetare-Dezvoltare și Inovare în TIC : Dezvoltarea de produse și servicii inovative care să deservească cele 10 sectoare identificate în domeniul Smart Specialization” (TIC-SI), având ca obiectiv investigarea și concretizarea acestui rol.

Cele 10 sectoare de specializare inteligentă (SI) sunt nominalizate în *Strategia Națională pentru Competitivitate 2014 – 2020* [2], structurate în trei categorii: după rolul economic și influența asupra ocupării forței de muncă, dinamica competitivă și, respectiv, potențialul de inovare, dezvoltare tehnologică și valoare adăugată. În cea de a doua grupă este inclus și sectorul TIC. Ca instrument de operaționalizare al acestei strategii, *Programul Operațional Competitivitate 2014-2020* [3] stabilește două axe prioritare: “CDI în sprijinul competitivității economice și dezvoltării afacerilor” și „TIC pentru o economie digitală competitivă”.

În ceea ce privește CDI, trebuie subliniat faptul că *Strategia Națională de Cercetare-Inovare 2014-2020* [4] are la bază tot principiul specializării inteligente, între cele 4 domenii SI pentru CDI fiind inclus și TIC, detaliat în 4 subdomenii și 21 de arii tematice.

Conform metodologiei de dezvoltare a proiectului TIC-SI, în prima etapă au fost selectate și analizate 4 tematici CDI-TIC considerate relevante pentru sectoarele SI, cu accent pe cele economice, care sunt orientate pe 4 dimensiuni esențiale ale soluțiilor informatice moderne: date, suport decizional, timp real - conectivitate, mobilitate. Rapoartele de analiză tematică au fost structurate pe trei secțiuni principale: *descrierea tematicii respective*, cu accent principal pe

conținutul principalelor concepte care guvernează evoluția domeniului tematic respectiv, așa cum sunt evidențiate de literatura de specialitate și de oferta recentă sau de perspectivă imediată a pieței de soluții pe plan internațional, pe de o parte, de cerințele de informatizare la nivelul organizațiilor, cu precădere al companiilor industriale de diverse mărimi, pe de altă parte; *potențialul de impact pentru competitivitate* al tematicii respective, evidențiat de conținutul ofertei de produse și servicii al unor firme reprezentative în domeniu, de analize de impact și de studii prospective efectuate de firme de consultață de prestigiu la nivel internațional, de interesul existent la nivelul comunităților de utilizatori privind implementarea unor asemenea soluții; *tipologia de soluții specifice* identificate pentru tematica respectivă.

Lucrarea de față prezintă, conform structurii menționate, rezultatelor analizei pentru prima tematică: **managementul, guvernarea și analiza datelor de mari dimensiuni (Big Data)**.

2. Descrierea tematicii

2.1 Managementul datelor și guvernarea datelor

Termenul de guvernare nu este prezent în dicționarul limbii române. Cu toate acestea el este utilizat din ce în ce mai frecvent în vorbirea curentă, fiind un echivalent pentru englezescul „governance”, care are o traducere clară în limba română: conducere, administrație, cărmuire, stăpânire, guvernare [5]. Conform acestui document, „guvernarea” este actul de a conduce în sens de „pilotare”. Poate fi un proces separat sau o parte din procesele de management sau conducere. În cazul unei întreprinderi sau al unei organizații non-profit, „guvernarea” se referă la management coerent, politici de coeziune, orientare, procese și drepturi de decizie pentru un anumit domeniu de responsabilitate. Termenul este utilizat pentru a face o distincție cât mai clară între actul de conducere exercitat de un guvern și conducerea exercitată în cazul unor structuri economice, sociale sau de altă natură.

Guvernarea TIC a existat ca subiect de cercetare, în diverse forme, în literatura care se ocupa de infrastructura TIC, valoarea economică a TIC și managementul de proiect, de peste două decenii.

„Guvernarea datelor” este un set de procese care asigură că managementul activelor de tip date este executat în mod formal, în conformitate cu regulile de bună practică, în cadrul unei organizații [6]. În lucrarea [7] guvernarea datelor este definită ca o abordare de management al datelor și al informațiilor la nivelul unei organizații, care formalizează un set de politici și proceduri ce cuprind întregul ciclu de viață al datelor, de la achiziție, utilizare și până la eliminarea acestora.

Dicționarul de management al datelor al DAMA (the Data Management Association International), definește „guvernarea datelor” ca "exercitarea autorității, a controlului și luarea deciziilor în comun (planificare, monitorizare și aplicare) asupra managementului activelor de date" (<http://blogs.perficient.com/healthcare/blog/2012/06/12/data-governance-vs-data-management/>). DAMA a identificat 10 funcții majore ale managementului datelor în DAMA-DMBOK (Data Management Body of Knowledge). Guvernarea datelor este componenta centrală a managementului datelor, care leagă împreună alte 9 discipline: managementul arhitecturii datelor, dezvoltarea datelor, managementul operării bazelor de date, managementul calității datelor, managementul metadatelor, managementul datelor de referință, managementul documentelor și al conținutului digital, managementul magaziiilor de date și inteligența în afaceri, managementul securității datelor.

Guvernarea informațiilor este constituită din ansamblul de capacități și practici care servesc la crearea, culegerea, evaluarea, stocarea, utilizarea, controlul, organizarea accesului, arhivarea și distrugerea informațiilor în decursul ciclului de viață al acestora [8]. Guvernarea informațiilor are două scopuri:

1. maximizarea valorii informațiilor pentru organizație, prin asigurarea îndeplinirii cerințelor de fiabilitate, siguranță și accesibilitate pentru luarea deciziilor;

2. protecția informațiilor, astfel încât valoarea lor să nu fie afectată de eventuale erori umane sau tehnologice, întreruperi ale accesului, condiții de utilizare neadecvate sau accidente.

Spre deosebire de governanța infrastructurilor TIC, governanța informațiilor ia în considerare aspectele și caracteristicile specifice artefactelor informaționale. De exemplu, spre deosebire de artefactele fizice, informațiile pot fi replicate și partajate la distanțe mari în mod rapid și facil. De asemenea, informațiile sunt bunuri a căror valoare nu scade în timp, spre deosebire de componentele infrastructurilor TIC ce se uzează moral sau fizic. Este posibil chiar ca valoarea acestora să crească în timp și pe măsura utilizării.

2.2 Fenomenul Big Data

Davenport, Barth și Bean [9] afirmă că organizațiile înoată într-un ocean de date în expansiune, care sunt fie prea voluminoase, fie prea nestructurate pentru a putea fi gestionate și analizate prin metode tradiționale. Printre sursele de date în plină dezvoltare pot fi enumerate datele de tip clickstream de pe Web, conținutul social media (tweet-uri, blog-uri, anunțurile de pe peretele Facebook etc.) și datele video din marketing-ul on-line, precum și cele de divertisment video. Big Data cuprind o gamă extrem de largă: de la date de voce generate în centre de tip „call center”, la date de genomică și proteomică din cercetarea biologică și medicină. De remarcat că doar o foarte mică parte a acestor informații este formatată în rânduri și coloane, conform bazelor de date convenționale.

Companiile orientate spre exploatarea comercială a Big Data se diferențiază în trei moduri principale:

a) Concentrarea pe fluxurile de date în detrimentul depozitelor de date: există mai multe tipuri de aplicații ale Big Data. Primul tip susține procesele de lucru ale organizației, cum sunt identificarea fraudelor în timp real sau evaluarea pacienților în medicină cu privire la riscurile pentru sănătate. Un al doilea tip implică monitorizarea continuă a procesului pentru a detecta evenimente sau situații ca: modificări ale percepției consumatorilor sau necesitatea intrării în service a unui motor cu reacție. Al treilea tip utilizează Big Data pentru a explora relațiile în rețele sociale, cum ar fi prietenii propuși pe LinkedIn și Facebook. În toate aceste aplicații, datele nu sunt constituite de "stocul" dintr-un depozit de date, ci dintr-un flux continuu. Devine mai importantă evaluarea de fluxuri și procese continue decât ceea ce a avut loc în trecut. Aceasta reprezintă o schimbare substanțială față de situația în care analiștii de date efectuau mai multe analize pentru a determina semnificația într-o cantitate fixă de date. „Streaming Analytics” permite prelucrarea datelor în timpul unui eveniment pentru îmbunătățirea rezultatelor [9].

Volumul și viteza crescute ale datelor în mediile de producție vor determina organizațiile să dezvolte procese continue pentru colectarea, analiza și interpretarea datelor. Deși "stocuri" mici de date situate în depozite sau baze de date vor continua să fie utile pentru dezvoltarea și rafinarea modelelor analitice folosite, odată dezvoltate modelele, acestea trebuie să proceseze fluxuri de date continue cu rapiditate și precizie. În contexte de monitorizare în timp real, organizațiile trebuie să adopte o abordare de tip flux continuu în analiza și luarea deciziilor pe baza a o serie de ipoteze și presupuneri. „Social Media Analytics”, de exemplu, preia tendințele rapid schimbătoare în sentimentele clienților despre produse, mărci și companii.

b) Utilizarea suportului experților în date (*Data scientists*) și al dezvoltatorilor de produse și procese, mai puțin al analiștilor de date (*Data analysts*): deoarece interacțiunea cu datele în sine - obținerea, extragerea, manipularea și structurarea acestora - este critică pentru orice analiză, personalul care lucrează cu Big Data trebuie să dețină abilități substanțiale și creative. Experții în date înțeleg Analytics, dar sunt experimentați și în TIC, având de multe ori studii avansate în informatică, fizică computațională, biologie ori științe sociale. Setul lor actualizat de calificări în gestionarea datelor - incluzând programare, competențe matematice și statistice, precum și înțelegere a afacerii și abilitatea de a comunica eficient cu factorii de decizie - merge mult dincolo de ceea ce era necesar pentru analiștii de date din trecut.

c) Mutarea Analytics de la sistemul informatic către activitatea de bază și funcțiunile

operaționale : volumele de date în creștere necesită îmbunătățiri majore în bazele de date și tehnologiile de Analytics. Preluarea, filtrarea, stocarea și analiza fluxurilor de Big Data pot bloca rețelele tradiționale și platformele de baze de date relaționale. Încercările de a reproduce și a scala tehnologiile existente nu vor ține pasul cu cerințele Big Data, determinând schimbarea tehnologiilor, abilităților și proceselor TIC. Produse noi destinate să prelucreze Big Data includ platformele open source, cum ar fi Hadoop, inventat pentru a sprijini gama largă de date generate și gestionate în Internet. Hadoop permite organizațiilor să încarce, stocheze și interogheze seturi masive de date pe o rețea mare de servere ieftine, precum și să execute operații de Analytics avansate, în paralel. Bazele de date relaționale au fost, de asemenea, transformate: noile produse au performanță de interogare crescută cu un factor de 1.000 și sunt capabile de a gestiona o mare varietate de surse de Big Data. Pachetele de analiză statistică evoluează în mod similar pentru a lucra cu aceste noi platforme de date, tipuri de date și algoritmi.

O altă tendință este furnizarea de capacități Big Data în Cloud. Multe aplicații de Big Data utilizează informații provenite din spațiul public, cum ar fi modelarea rețelelor sociale și analiza sentimentelor.

O abordare specifică în gestionarea Big Data este de a lăsa datele acolo unde se află. Așa-numitele "piețe de date virtuale" permit experților în date să partajeze datele existente, fără replicarea acestora.

Alinierea la Big Data determină organizațiile să-și regândească ipotezele de bază cu privire la relația dintre afaceri și TIC, precum și rolurile acestora. Un principiu cheie al Big Data este că lumea și datele care o descriu sunt în continuă schimbare; de aceea vor avea de câștigat organizațiile care pot să recunoască aceste modificări și să reacționeze rapid și inteligent. Noile avantaje se bazează pe descoperire și agilitate, capacitatea de a explora în mod continuu sursele de date existente și cele noi pentru a identifica modele predefinite, evenimente și oportunități. Organizațiile de succes vor instrui și recruta oameni cu un nou set de aptitudini care pot integra aceste noi capacități de Analytics în mediile lor de producție.

O altă modalitate prin care Big Data determină modificarea rolurilor tradiționale de afaceri și TIC este aceea că prezintă descoperirea și analiza ca prime comandamente ale afacerii. Următoarea generație de procese și sisteme TIC trebuie să fie proiectate pentru perspectivă, nu doar pentru automatizare.

Arhitecturile TIC tradiționale conțin aplicații (sau servicii) ca și "cutii negre" care realizează sarcini, fără expunerea datelor și a procedurilor interne. În contrast, mediile de Big Data trebuie să înțeleagă datele noi și deci, raportările rezumative nu mai sunt suficiente. În consecință, aplicațiile TIC trebuie să analizeze și să raporteze în mod transparent pe o mare varietate de dimensiuni, inclusiv interacțiunile clienților, utilizarea produselor, acțiunile de service și alte măsuri dinamice. Pe măsură ce Big Data evoluează, arhitecturile se vor dezvolta într-un ecosistem de informații: o rețea de servicii interne și externe partajând continuu informații, optimizând deciziile, comunicând rezultatele și generând noi perspective pentru afaceri.

2.3 Guvernanța datelor și Big Data

Guvernanța Big Data este într-un stadiu incipient. Cu toate acestea, guvernanța și integrarea informațiilor sunt factori esențiali pentru a obține valoarea maximă dintr-un proiect de tip Big Data. Fără garanția că informațiile pe care se bazează sunt de încredere, organizațiile nu pot lua decizii bazate pe tehnicile Analytics și Business Intelligence.

Volumul datelor care se acumulează într-o organizație crește continuu și cu viteze din ce în ce mai mari. Pentru a se evita situația în care devine imposibil ca datele să fie clasificate și interogate, iar informațiile imposibil de gestionat, este necesar a se adopta politici de guvernanță clare și bine definite. Față de abordarea tradițională (baze de date relaționale), guvernanța Big Data implică noi tipuri și forme de informații, cum ar fi: bazele de date non-relaționale sau incomplet relaționale, date nestructurate de tipul celor provenite din aplicațiile informatice de tip „social media” sau generate de senzori. Apar noi provocări pentru implementarea guvernanței, deoarece aceste noi

tipuri de date trebuie să fie integrate cu infrastructurile de governanță a informațiilor și tehnologice existente. Absența unei abordări coordonate a guvernantei Big Data sau neglijarea imperativelor de integrare în governanța și managementul datelor unei organizații conduce la apariția următoarelor situații de risc sau pericol:

- „inundarea” cu date care sunt dificil sau imposibil de regăsit și analizat;
- neîndeplinirea conformității cu reglementările și regulile interne, precum și cu standardele generale de calitate cum ar fi: *Sarbanes-Oxley (SOX)* sau *Basel 2*, *Federal Rules of Civil Procedure (FRCP)* sau *the Federal Rules of Evidence (FRE)*, *the Health Insurance Portability and Accountability Act (HIPAA)* sau reglementări similare ale Uniunii Europene;
- suportarea unor pierderi financiare sau de reputație;
- costuri sporite datorate unor politici neclare de retenție a informațiilor.

2.4 Big Data Analytics

„Analytics” este definit, în literatura de limbă engleză [10], ca o deliberare bazată pe fapte ce conduce la formularea de perspective de pătrundere (eng. „insights”), diagnostice, precum și la posibile implicații pentru planificarea viitoarelor acțiuni, într-un mediu organizațional. Aria de cuprindere a Analytics poate varia de la urmărirea de rutină și monitorizarea performanței în afaceri, până la o diagnosticare dirijată a cauzei principale a problemelor de afaceri, precum și o predicție strategică cu privire la inițiativele de afaceri viitoare. Caracterul comun în toate aceste activități este faptul că sunt conduse în mod semnificativ de fapte („raționale”, prin natură), obținute prin colectarea intenționată a datelor necesare.

Analytics este un ansamblu de procese de analiză a datelor la care contribuie în mod decisiv științe ca: statistica matematică, teoria algoritmilor, știința computațională. Preeminența sa în mediul organizațional se datorează progreselor făcute în domeniul TIC, conducând la infrastructuri de calcul performante și la apariția unor tehnici și instrumente software avansate pentru colectarea și prelucrarea informațiilor.

„Business Analytics” - BA este un termen care poate fi definit ca "o mulțime a tuturor competențelor, tehnologiilor, aplicațiilor și practicilor necesare pentru explorarea și investigarea în mod iterativ, continuu, a performanțelor anterioare în afaceri, în scopul obținerii unei perspective și conducerii planificării afacerii [10]. În funcție de rezultatele sale, acest proces poate fi de tip descriptiv, de diagnosticare, predictiv sau prescriptiv. Principalul avantaj al utilizării BA în procesul de luare a deciziilor de afaceri este posibila evitare a subiectivității. Deși creierul uman este capabil de prelucrarea mai multor dimensiuni de date la un moment dat, acestuia îi lipsește coerența, care este obținabilă printr-un proces științific rațional utilizând ajutor computațional.

Tranziția de la euristică la rezolvarea problemelor pe bază de fapte a fost stimulată de accesul mai facil la datele de afaceri și dezvoltarea unor capacități de procesare mai inteligente.

„Big Data Analytics” - BDA se referă la colectarea, organizarea și analiza unor seturi mari de date pentru a descoperi modele predefinite, șabloane și alte informații utile. Utilizarea BDA implică rezolvarea unor provocări: "spargerea" silozurilor de date pentru a accesa toate datele pe care o organizație le stochează în diferite locuri și, adesea, pe diferite sisteme; crearea de platforme care pot colecta date nestructurate la fel de ușor ca pe cele structurate.

Big Data necesită implementări de înaltă performanță ale tehnicilor Analytics. BDA se realizează de obicei folosind instrumente software specializate și aplicații pentru analiză predictivă, explorarea datelor, explorarea textului, prognoză și optimizare a datelor.

2.5 Datele deschise și valoarea lor economică

Datele deschise pot fi utilizate în mod liber, modificate și împărtășite de oricine, în orice scop (<http://opendefinition.org/>). Redifuzarea acestora este permisă cu respectarea cerințelor care

conservă proveniența (dreptul de autor) și deschiderea (partajarea mai departe). Există două dimensiuni în deschiderea datelor:

- legală: trebuie să fie plasate în domeniul public sau în condiții liberale de utilizare cu restricții minime;
- tehnică: trebuie să fie publicate în formate electronice care sunt ușor de citit în mod automat și, de preferință, non-proprietate. Astfel, oricine poate accesa și utiliza datele folosind instrumente software comune, disponibile în mod liber. Datele trebuie să fie, de asemenea, disponibile public și accesibile pe servere publice, fără restricții de parolă sau firewall.

În lucrarea [11] se evidențiază că seturile de date variază de la complet deschise la complet închise, în patru dimensiuni:

- a. accesibilitate (gama de utilizatori care au permisiunea de a accesa datele),
- b. lizibilitate (ușurința cu care datele pot fi prelucrate în mod automat),
- c. cost (prețul pentru a obține datele), și
- d. drepturi (limitări cu privire la utilizarea, transformarea și difuzarea/distribuția datelor).

Prin utilizarea datelor deschise sau publice provenite de la cele mai diverse surse și combinarea cu date interne, proprietatea unei companii (agregare) se pot realiza introspecții și aplicații inovative care pot fi exploatate pentru a face operațiile din cadrul firmei mai eficiente și mai eficiente, sau pot contribui la dezvoltarea de produse și servicii noi și inovative.

Cu cât mai multe date sunt deschise, cu atât mai mult ele pot fi folosite, reutilizate, readaptate în alte scopuri, combinate cu alte date pentru a crea valoare adăugată, prin:

- reducerea costurilor în furnizarea de servicii existente, atât de către organizațiile guvernamentale, cât și în sectorul privat (de exemplu a face același lucru pentru un cost mai mic);
- apariția unor noi servicii și îmbunătățirea calității serviciilor existente;
- contribuție indirectă la îmbunătățirea guvernării prin creșterea responsabilității și implicării cetățenilor, ambele generând o mai mare încredere în guvernare.

2.6 Știința datelor și expertii în știința datelor

Există în prezent, atât în domeniul academic, cât și în rândurile practicienilor, o dezbateră asupra definiției științei datelor. Una din abordări constă în a considera știința datelor ca fiind un pas evolutiv spre un domeniu interdisciplinar care include: știința calculatoarelor, informatica, modelarea, statistica matematică și Analytics.

În esența sa, știința datelor presupune utilizarea metodelor automatizate pentru a analiza cantități masive de date și pentru a extrage cunoștințe din acestea. Tendința este de așteptat să se accentueze în anii următori, pe măsură ce datele provenite de la senzori mobili, instrumente sofisticate, de pe Web și din diverse alte surse cresc continuu și se acumulează. Astfel, în cercetarea academică, se consideră că în cadrul unui număr din ce în ce mare de discipline tradiționale vor apărea noi subdomenii cu adjectivul "computațională" sau "cantitativă". În aproape toate domeniile de reflecție intelectuală, știința datelor oferă o abordare nouă, puternică, pentru a face descoperiri.

La nivel conceptual, știința datelor este un set de principii fundamentale care susțin și ghidează extragerea de informații și cunoștințe din date [12]. Probabil, conceptul cel mai strâns legat de știința datelor este explorarea datelor (*Data Mining*), extragerea cunoașterii din date, prin intermediul tehnologiilor care încorporează aceste principii.

În cadrul unei organizații, scopul fundamental al științei datelor este de a promova, sprijini și ameliora în permanență procesele de decizie conduse de date.

Decizia determinată de date (*Data Driven Decision Making*) este ansamblul de practici și tehnici de decizie asistată care se bazează pe analiza datelor și nu pe intuiție. De exemplu, în marketing deciziile de selecție a publicității se pot lua pe baza experienței specialiștilor în domeniu care decid „ce merge și ce nu” sau selecția se poate executa pe baza analizei datelor referitoare la modul în care consumatorii reacționează la diverse reclame.

Dar poate că cea mai sugestivă definiție a științei datelor se poate face prin clarificarea specializării celor care o practică. Conform [13], termenul de „expert în știința datelor” (*Data Scientist*), desemnează o persoană care știe să extragă sens din date și să interpreteze datele. Un astfel de expert stăpânește instrumentele și metodele statisticii și de învățare automată și, de asemenea, petrece mult timp în procesul de colectare, curățare și punere în evidență a datelor, dar are și cunoștințe de persistență, statistică, precum și de inginerie software.

După ce datele au primit o formă prezentabilă, o altă componentă esențială este analiza exploratorie a datelor, care combină metodele de vizualizare și sensul din date. Expertul în știința datelor va găsi modele predefinite, va construi modele și algoritmi, poate proiecta experimente și este implicat ca parte esențială a procesului decizional condus de date. Va comunica cu membri echipei, ingineri și persoane de conducere într-un limbaj clar pentru aceștia și cu vizualizări care să permită înțelegerea implicațiilor.

Așa cum se constată în lucrarea [14], importanța acestei specializări a devenit evidentă abia în momentul în care companiile au devenit conștiente de avantajul competitiv care poate fi obținut prin exploatarea datelor pe care le au la dispoziție, precum și de dificultățile cu care sunt confruntate, fiind practic inundate cu date.

Caracteristic pentru instrumentele și tehnologiile care fac posibil managementul datelor și dezvoltarea de aplicații intensive ca date, este faptul că majoritatea au fost create în cadrul unor companii ca Facebook, Google, Twitter sau LinkedIn, de către specialiști care intră în categoria experților în știința datelor și care au excelat în afaceri datorită modului inteligent în care au fructificat datele pe care le-au avut la dispoziție.

2.7 Tranziția către întreprinderea reactivă și condusă de date

Examinând topul celor mai valoroase companii, lucrarea [15] sesizează o ascensiune a firmelor „bazate pe software”, cum sunt Uber, FlixBus, Tesla sau Airbnb. Succesul lor se datorează faptului că adoptă un model de afaceri puternic bazat pe software, care le permite să fie reactive și „conduse de date” (*data-driven*) și astfel să reacționeze rapid la factorii externi.

Tendința este ca succesul și valoarea afacerii să se bazeze pe formula „business = date + algoritmi”. Iată un exemplu în acest sens. Compania Uber nu deține o flotă de mașini. Succesul său se bazează pe colectarea datelor în timp real și pe algoritmi pe care îi utilizează pentru a transforma aceste date în decizii. Sistemul Uber urmărește în permanență condițiile de trafic, cererea și oferta de servicii de transport, precum și istoricul elasticității prețurilor la consumator. Astfel este capabil să optimizeze tariful călătoriilor și să direcționeze mașinile către locurile cu cea mai mare cerere de transport. În mod similar companiile Tesla și Airbnb, care produc efecte disruptive pe piețele fabricării de autoturisme și, respectiv, industria hotelieră, sunt în esență companii bazate pe software.

A deveni companie condusă de date presupune cultivarea și adoptarea unei mentalități conform căreia desfășurarea afacerii este bazată pe utilizarea continuă a tehnicilor de tip Analytics în luarea deciziilor de afaceri pe bază de fapte. Scopul este acela de a se ajunge la un stadiu la care utilizarea datelor și a elementelor disciplinei Analytics de către personalul de decizie și de către angajați să devină o parte firească a fluxurilor de lucru zilnice ale acestora. Unul din aspectele care diferențiază companiile conduse de date față de competitorii lor este reprezentat de determinarea de a colecta datele relevante pentru toate aspectele afacerii lor, ceea ce le permite să exploreze în adâncime pentru a înțelege cauzele principale din spatele anumitor condiții specifice ale afacerii, cum sunt modificările în comportamentul clienților sau ale tendințelor pieței.

Procesul de tranziție către acest tip de companie include definirea acelor metrici ale succesului care vor fi evaluate și maparea acestor metrici pe seturile de date care vor contribui la evaluările respective, adoptarea utilizării datelor și tehnicilor Analytics în fluxurile zilnice de lucru, în întreaga organizație.

3. Potențialul de impact pentru competitivitate

3.1 Evaluări preliminare

La nivel macroeconomic, economia bazată pe date este definită ca un ecosistem de tipuri diferite de actori care interacționează în cadrul unei piețe unice digitale, conducând la mai multe oportunități de afaceri și la o mai mare disponibilitate de cunoștințe și de capital, în special pentru IMM-uri, precum și o stimulare mai eficace a cercetării și inovării relevante [16].

La nivel microeconomic, sondajul Economist Intelligence Unit [17] relevă o relație puternică între „centrarea pe date” și realizarea de performanțe superioare. O mai mare disponibilitate și utilizare a datelor la aceste companii duce la o mai bună partajare a cunoștințelor (70% față de 41%), o gestionare a riscurilor superioară (67% față de 43%) și o organizare mai colaborativă (59% față de 33%). Companiile conduse de date, de asemenea, raportează o creștere a calității și a vitezei de execuție (55% față de 24%), o mai rapidă luare a deciziilor (55% față de 28%) și satisfacția angajaților mai mare (44% față de 21%).

Studiul dedicat modului în care deciziile determinate de date (DDD) influențează performanțele firmelor [18], a concluzionat statistic că, cu cât gradul în care o firmă este condusă de date este mai ridicat, cu atât este mai productivă: diferențele înregistrate indică faptul că o deviație standard mai ridicată pe scara indicatorului DDD este asociată cu o creștere de 4-6% a productivității; de asemenea, indicatorul DDD este corelat cu rentabilitatea mai mare a activelor, rentabilitatea capitalului, utilizarea activelor și valoarea de piață, iar relația pare să fie cauzală.

3.2. Calitatea datelor

Asigurarea calității datelor este văzută ca o disciplină matură, în special atunci când în centrul atenției se află evaluarea seturilor de date și aplicarea acțiunilor de remediere sau corective asupra acestora [19]. La această percepție au avut o contribuție majoră două fenomene ce s-au manifestat recent. Primul este conștientizarea faptului că seturile de date create cu un anumit scop funcțional în cadrul unei organizații (cum ar fi vânzări, marketing, contabilitate, sau de achiziții publice pentru a numi doar câteva) sunt refolosite în contexte diferite, în special pentru raportare și analiză. În consecință, calitatea datelor nu mai poate fi exprimată și măsurată în funcție de cât de adecvate sunt unui anumit scop, ci trebuie să fie evaluate din perspectiva unor scopuri multiple, luând în considerare toate utilizările și cerințele de calitate din aval. Cel de al doilea, strâns legat de precedentul, este convingerea că asigurarea uzabilității datelor pentru toate scopurile necesită o supraveghere mai cuprinzătoare.

Cele două fenomene au consolidat poziția proeminentă a guvernancei datelor în medii caracterizate de Big Data.

Trebuie avut în vedere specificul guvernancei datelor și al asigurării calității acestora în cazul Big Data, comparativ cu abordările convenționale. Aplicațiile Big Data preiau fluxuri de intrare multiple, provenind din interiorul și din afara organizației, provenind dintr-o varietate de fluxuri de date, seturi de date publice sau open source, rețele de senzori sau alte surse de date nestructurate. Astfel de seturi de date nu pot fi guvernate separat sau în mod singular. În plus, cea mai dificilă este problema coerenței. Când seturile de date sunt create în interiorul organizației și un utilizator din aval sezează o potențială eroare, problema poate fi comunicată proprietarilor sistemului de origine, care au posibilitatea de a găsi cauza principală a problemelor și apoi corectarea proceselor care au condus la erori. În cazul sistemelor care lucrează cu Big Data, care absorb volume masive de date provenite din exterior, există oportunități limitate de a implica proprietarii proceselor în executarea de modificări la sursă. Pe de altă parte, în cazul în care se

optează pentru „corectarea” fluxului de date potențial, se introduce o inconsistență cu sursa originală, ceea ce poate duce la concluzii incorecte și decizii eronate.

3.3 Potențialul și limitele deciziei determinate de date

În mediul organizațional există un spectru larg de decizie. La un capăt al spectrului sunt deciziile operaționale, de rutină, pe termen scurt foarte structurate și din ce în ce mai încorporate în aplicații software sofisticate. La celălalt capăt al spectrului sunt deciziile strategice, care stabilesc direcțiile și politicile unei afaceri sau organizații pe termen lung. Acestea sunt de obicei adoptate la un nivel ridicat de management și tind să fie complexe și nestructurate datorită incertitudinii și a riscurilor care însoțesc în general deciziile pe termen mai lung. Între aceste extreme se află mai multe tipuri de decizii, inclusiv cele care nu sunt de rutină ca răspuns la situații noi sau neprevăzute, dincolo de domeniul de aplicare al proceselor operaționale, precum și deciziile tactice asupra adaptărilor necesare pentru a pune în aplicare strategii pe termen mai lung.

Pentru a pune în valoare avantajele acestei abordări trebuie înțeles când automatizarea devine limitativă și este necesar ca procesul decizional condus de date să fie privit ca un instrument ajutător pentru decizii mai eficiente și mai inteligente [20].

Pe măsură ce se colectează mai multe date noi și se aplică metode de analiză mai sofisticate, pot fi luate decizii cu intervenție umană limitată sau inexistentă. În timp, Big Data și aplicații avansate de știința datelor vor permite decizii operaționale la un nivel cu totul nou, într-o mare varietate de discipline.

Big Data sunt capabile să furnizeze informații despre comportamentul uman. Sunt utilizate elemente de date pe care le lasă în urmă oamenii în activitatea curentă, care pot proveni de la sistemul de localizare al telefonului mobil sau de la tranzacțiile de cumpărare executate cu cardul de credit. Dar generarea unor decizii / concluzii pe baza informațiilor cu caracter personal obținute din diverse surse poate genera suspiciuni serioase cu privire la confidențialitate, la probleme legate de proprietatea asupra datelor și de controlul datelor. Este important ca utilizatorii să fie conștienți de aceste riscuri, precum și de faptul că au ultimul cuvânt cu privire la utilizarea datelor colectate despre ei. Este necesar să se găsească echilibrul corect între utilizarea unor astfel de aplicații de luare a deciziilor determinate de date și confidențialitatea vieții private.

Aceleași probleme și dileme intervin în situații care impun decizii strategice cum ar fi securitatea națională sau ordinea publică. Utilizarea Big Data și a științei datelor în asemenea cazuri necesită luarea în considerare a contextului și este încă un domeniu de investigare.

Necesitatea existenței unui cadru care să sprijine personalul de decizie să perceapă și să descrie în mod rapid contextul în care se iau deciziile este subliniată în lucrarea [20]. Un context ordonat, fie că e simplu sau complicat, presupune un univers ordonat, în care relațiile cauză-efect sunt perceptibile, iar răspunsurile corecte pot fi determinate pe bază de fapte. Un context complex și haotic este neordonat și nu există o relație imediată între cauză și efect, iar calea de urmat este determinată pe baza unor șabloane care apar pe parcurs. Universul ordonat este subiectul managementului bazat pe fapte, iar contextul neordonat este gestionat pe bază de șabloane / modele. Una dintre cele mai mari provocări ale DDD constă în a evita să se considere în mod greșit că un context neordonat, imprevizibil, complex este de fapt unul ordonat, complicat, dar previzibil. Astfel de ipoteze încurajează simplificări care sunt utile doar în anumite circumstanțe. Un management eficient nu adoptă un singur mod de decizie (DDD sau decizia bazată pe modele), indiferent de situație. În cazul deciziilor operaționale, este necesar să se facă distincția între acele situații în care deciziile pot fi încorporate în procesele automate și cele care necesită intervenție umană. În cazul deciziilor strategice, trebuie făcută diferența dintre contexte complicate, dar previzibile și complexe, și cele intrinsec imprevizibile.

4. Tipologia de soluții TIC specifice

4.1 Tipuri de proiecte și guvernanta Big Data

Factorul determinant în adoptarea unei strategii de guvernanta este tipul de proiect Big Data care se urmărește a fi implementat [21]. Perioada de valabilitate și modul de utilizare a datelor sunt indicatori importanți care influențează modul de abordare. Astfel, anumite date au valoare pe perioadă scurtă, expirând rapid, altele dimpotrivă își păstrează valoarea un timp îndelungat și trebuie stocate pe perioade lungi. De asemenea, unele date sunt utilizate în mod individual, la nivel de înregistrare, iar altele sunt anonimizate și utilizate la nivel agregat. În Figura 1 sunt grupate tipurile de proiecte cu diferite cerințe impuse guvernantei și integrării informațiilor, în funcție de necesitățile de preservare, recunoaștere, percepție sau retenție a datelor.

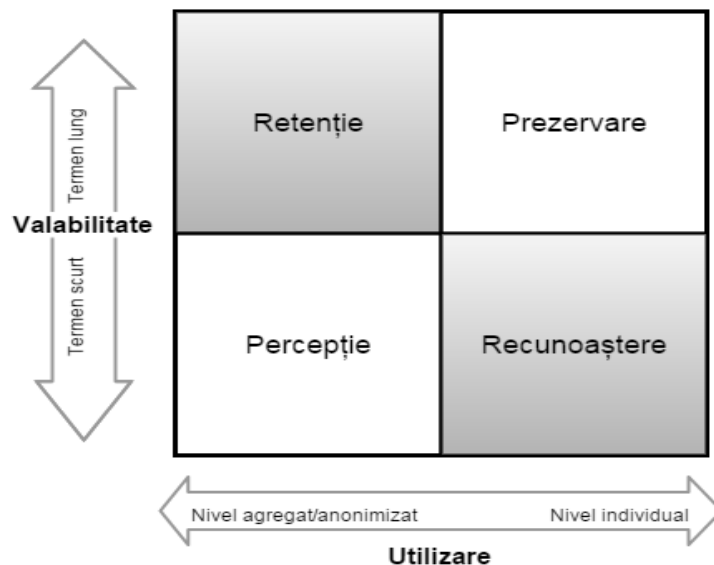


Figura 1. Cadranel tipurilor de proiecte Big Data (după [21]).

Percepția. Proiectele din această categorie colectează și assemblează date cu scopul de a identifica tendințe (de exemplu să identifice sentimentele consumatorilor pe baza utilizării analizei mijloacelor de difuzare de tip „social media”). Datele se acumulează rapid și au o durată de viață scurtă. În consecință, se pune accentul pe integrarea lor rapidă.

Rolul guvernantei și integrării informațiilor este de a asigura livrarea datelor, consistența impusă acestora, protecția datelor sensibile, precum și distrugerea sau arhivarea în timp util a acestora. Politicile de gestiune a ciclului de viață a datelor sunt aplicate la nivel agregat. Politicile de retenție și arhivare sunt importante pentru că asigură controlul creșterii volumului datelor. Datele sensibile trebuie mascate (nedivulgate) pentru a se asigura că rămân realiste, protejate și în siguranță. Se impune un anumit nivel al calității datelor pentru asigurarea consistenței și pentru a facilita analiza lor, dar nu este necesară aplicarea riguroasă a tuturor aspectelor calității.

Retenția. Această categorie de proiecte este similară celor din cadranul Percepție, cu excepția că datele sunt reținute (memorate) pe o perioadă mai lungă, în scopul executării unor analize istorice. În general cu cât datele sunt reținute mai mult, cu atât este necesară mai multă guvernanta. Exemple de aplicații specifice acestui cadran sunt prognozarea stocurilor necesare sau analize de tip demografic.

Proiectele din acest cadran se concentrează pe creșterea consistenței datelor. Managementul ciclului de viață al datelor este în continuare o capacitate importantă, pentru a păstra controlul asupra creșterii volumului de date. În acest cadran, guvernanta și integrarea informațiilor sprijină consistența, precum și includerea datelor provenite din surse multiple.

Recunoașterea. Proiectele de acest tip sunt similare celor din cadranul Percepție, în sensul că

perioada de valabilitate a datelor este foarte scurtă (de exemplu analiza unor date de campanie de marketing sau interpretarea datelor colectate de la dispozitive și senzori). Însă această categorie se distinge prin concentrarea pe date de tip individualizat, înregistrări separate.

Aria de cuprindere a calității datelor este mai largă în acest cadran. În consecință, guvernarea datelor trebuie să depășească consistența și să urmărească asigurarea corectitudinii datelor. Pentru validarea și recunoașterea datelor este utilizat managementul datelor principale (*Master Data Management* – MDM), care furnizează un set de entități principale, unice derivate din surse de date fragmentate.

Aspecte importante în acest cadran sunt: arhivarea datelor pentru a controla creșterea volumelor, managementul datelor de test, precum și integrarea diferitelor tipuri de stocare (replicare, pe loturi, federativă). În plus, se pune accentul pe asigurarea agilității, datorită timpului scurt cerut de analiza datelor.

Prezervarea. În acest cadran se află proiectele care au cele mai avansate cerințe pentru guvernarea ocupându-se de date individuale care trebuie prezervate un timp îndelungat (de exemplu aplicații la nivel de întreprindere de tip „mission critical” sau sisteme de Analytics și de raportare în domeniul financiar). Sistemul informatic trebuie să asigure atât corectitudinea, cât și siguranța și confidențialitatea datelor.

Principalele acțiuni avute în vedere în cadrul guvernării sunt: asigurarea acurateții menținerii datelor de încredere, asigurarea calității datelor și asigurarea standardizării și validării informațiilor. În acest scop este utilizat MDM. Obiectivul de interes major pentru managementul ciclului de viață al datelor se mută de la politicile agregate la înregistrări individuale (de ex. arhivarea unor înregistrări particulare ale consumatorilor). Asigurarea confidențialității și siguranței datelor personale sunt în centrul atenției proiectelor de acest tip.

4.2 Arhitectura de referință a sistemelor Big Data Analytics

Procesul general de extragere de perspective din Big Data poate fi descompus în 5 etape, grupate în 2 subprocese principale: managementul datelor și Analytics (Figura 2) [22].

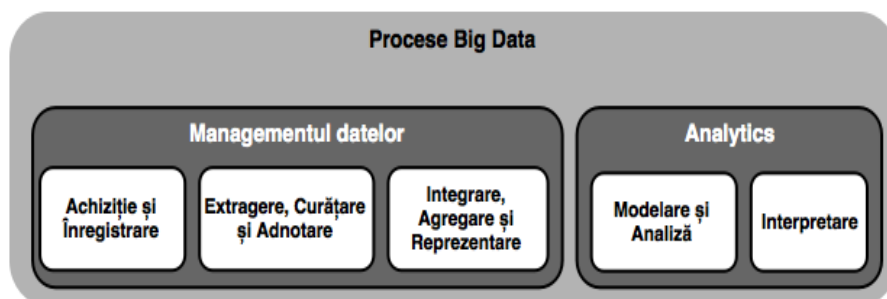


Figura 2. Procesele Big Data Analytics [22].

Managementul datelor cuprinde procesele și tehnologiile suport pentru achiziția și stocarea datelor, precum și pentru regăsirea și pregătirea acestora pentru a fi analizate. Analytics cuprinde tehnicile utilizate pentru a analiza datele și a extrage cunoștințe și soluții din date.

În mod similar, în lucrarea [23] se formulează o perspectivă asupra unei infrastructuri Big Data generice, subliniind că aceasta include o infrastructură generală de management al datelor, bazată în mod tipic pe Cloud, precum și o parte de Big Data Analytics care se bazează pe suportul obligatoriu al unui sistem distribuit și a unei rețele de mare performanță. Serviciile generale componente ale infrastructurii Big Data includ:

- instrumente de management al datelor, specifice Big Data (Big Data Management);
- servicii de evidență, indexare, căutare/regăsire, semantică și spațiu de nume;

- servicii de securitate și siguranță (controlul accesului, ranforsare a politicilor de acces, asigurarea confidențialității, a încrederii, a disponibilității și protecției datelor personale);
- medii suport pentru colaborare și management de grup.

Pe lângă serviciile generale de infrastructură Cloud de bază, în sprijinul aplicațiilor de Big Data Analytics sunt necesare servicii și instrumente specifice unei infrastructuri BDA:

- servicii de administrare a clusterelor;
- servicii și instrumente specifice mediului Hadoop;
- instrumente software de tip Data Analytics (jurnale, evenimente, explorarea datelor, învățare automată etc.);
- servere și sisteme de gestiune a bazelor de date;
- baze de date și sisteme de prelucrare paralelă.

Bazându-se pe investigarea unor cazuri de utilizare, precum și pe implementările de arhitecturi de Big Data Analytics la principalele companii mari cu activitate pe Web (Facebook, Google, Twitter, Netflix, LinkedIn etc), în lucrarea [24] se propune o arhitectură de referință de nivel înalt pentru sistemele de BDA, care evidențiază componentele funcționale, subsistemele de stocare și fluxurile de date din cadrul acestor sisteme (Figura 3). Fluxul datelor se desfășoară de la stânga la dreapta. Componentele funcționale ale arhitecturii de referință sunt interconectate de-a lungul acestui flux de date, formând o magistrală de prelucrare. Activitățile de specificare a joburilor și a modelelor sunt figurate separat pentru a ilustra caracterul distinct al acestora față de cel al funcțiilor on-line care constituie fluxul datelor.

Sursele de date sunt definite în două dimensiuni, mobilitate și structurare. Atributul in situ se referă la datele care nu se mișcă (de exemplu, un fișier Hadoop ce urmează a fi prelucrat). Datele de tip streaming sunt datele care aparțin fluxului, vin în mod continuu și trebuie prelucrate în timp real (de exemplu, fluxurile de date generate de Twitter).

Atributul structurare diferențiază datele astfel încât acestea pot fi considerate structurate dacă respectă un model strict (cazul bazelor de date relaționale care respectă o schemă), nestructurate dacă nu pot fi asociate cu nici un model (cazul paginilor Web sau al imaginilor) sau semistructurate (cazul documentelor în format XML sau JSON).

Extragerea datelor se referă la operațiile de preluare și introducere a datelor in situ în sistem. Aceste operații constau în extragerea datelor din structurile distribuite de stocare în care se află, stocarea lor temporară în depozite temporare sau transferul și încărcarea lor în spații de stocare specifice, denumite depozite de date brute. La rândul lor, datele de streaming pot fi extrase și stocate temporar în depozite temporare de date de streaming.

Urmează etape de prelucrare a datelor (combinare și curățare), executate pe datele brute, după care rezultatele sunt salvate în depozite temporare de date prelucrate. Operațiile de extragere a informațiilor constau în extragerea de informații noi din datele brute, precum și de structurare a acestora și stocarea în formate structurate în baze de date de nivel organizațional.

Operațiile de tip Analytics în profunzime (*Deep Analytics*) se referă la executarea de joburi în sistem pe loturi, asupra datelor in situ. Rezultatele acestor operații pot fi memorate în depozite dedicate sau în depozite de tip „Publish & Subscribe”, care au rolul de a facilita regăsirea rezultatelor analizelor în mod indirect, fără a exista o cuplare între componentele care publică (plasează) date și componentele care le preiau (abonați).

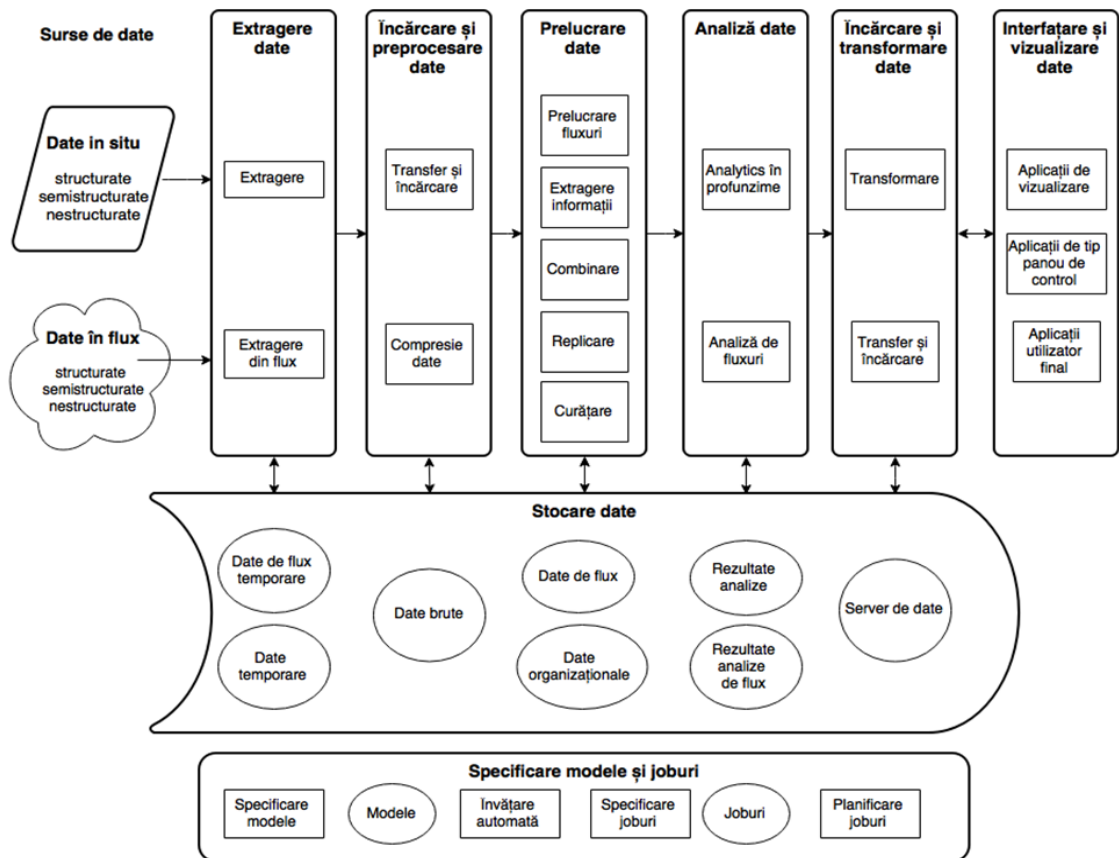


Figura 3. Arhitectura de referință BDA (după [24]).

Analiza datelor în flux (*Stream Analysis*) produce rezultate care sunt transformate și stocate în depozite de tip server pentru aplicațiile de vizualizare, panouri de bord și interfață utilizator. Aplicațiile de tip interfață utilizator, spre deosebire de cele de vizualizare, furnizează un set limitat de funcții de control, pentru a putea rula pe dispozitive mobile inteligente.

4.3 Calculul în Cloud, Big Data și Big Data Analytics

Calculul în Cloud și Big Data sunt strâns legate. Tehnicile Big Data oferă utilizatorilor posibilitatea de a folosi echipamentele obișnuite pentru a procesa cereri și interogări distribuite pe mai multe seturi de date și furnizează seturi rezultate în timp util. Calculul în Cloud oferă motorul care stă la baza prelucrărilor Big Data prin utilizarea Hadoop, o clasă de platforme de prelucrare a datelor distribuite.

O arhitectură pentru Big Data în Cloud, prezentată în Figura 4, este propusă în lucrarea [25]. Conform acestei arhitecturi, volume mari de date din Cloud și Web sunt stocate într-o bază de date distribuită tolerantă la defecte și procesate prin intermediul unui model de programare pentru seturi de date de mari dimensiuni, cu algoritmi paraleli, distribuiți într-un cluster. Scopul principal al vizualizării datelor, este de a prezenta rezultatele analitice într-o formă vizuală, prin diferite grafice, pentru luarea deciziilor.

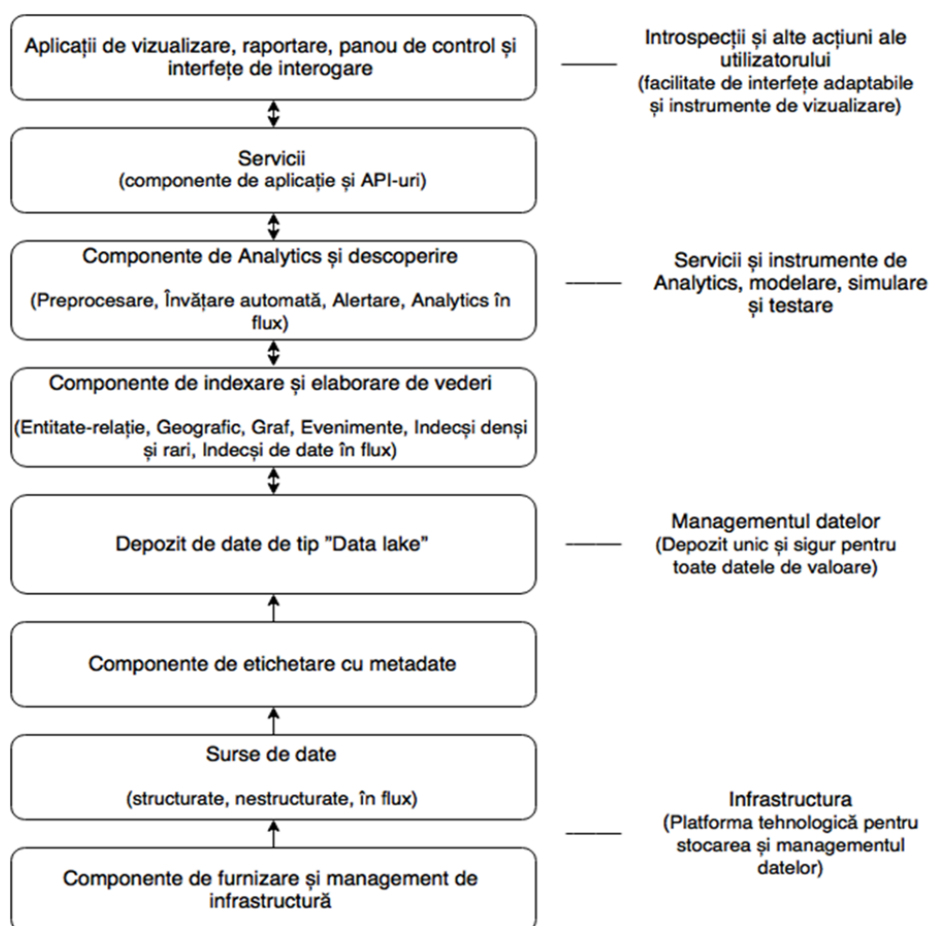


Figura 4. Arhitectura Big Data în Cloud [25].

Calculul în Cloud oferă nu numai facilități pentru calculul și prelucrarea Big Data, dar servește și ca un model de implementare și furnizare a serviciilor

În lucrarea [26] sunt prezentate rezultatele unei colaborări între compania Booz Allen Hamilton și guvernul SUA, ale căror eforturi au fost direcționate către utilizarea Big Data Analytics în căutarea de teroriști și amenințările acestora. Obiectivul a fost integrarea unui spectru larg de surse de date și aplicarea unor instrumente și metode sofisticate de Analytics pentru a descoperi legături și modele predefinite ascunse. De asemenea, s-a urmărit dezvoltarea unei infrastructuri de furnizare a informațiilor despre insurgenți sau alte entități care planifică plantarea de dispozitive explozibile improvizate sau similare. Rezultatele colaborării s-au concretizat sub forma unei arhitecturi de referință pentru Analytics în Cloud, prezentată în Figura 5.

Un element fundamental al arhitecturii de referință pentru Analytics în Cloud este constituit de depozitul central de informații denumit „lac de date”. Acesta este un spațiu de stocare vast, consolidat, care evită metodele rigide ale structurilor de date clasice. Orice căutare de informații poate utiliza acest spațiu de stocare în integralitatea lui și, în plus, surse de date externe. Utilizatorul nu mai este obligat să se mute de la o bază de date la alta și să extragă din acestea datele lor specifice pentru a realiza apoi o consolidare.

Un depozit de tip „lac de date” este mai mult decât un simplu mijloc de stocare, este un mediu proiectat cu intenția de a întări legăturile ce se află în date. Arhitectura de referință a fost elaborată astfel încât să fructifice și să exploreze aceste conexiuni, să identifice corelații și modele predefinite. Astfel se reduce complexitatea Big Data, măbind gradul de manevrabilitate și posibilitatea de administrare a acestora.

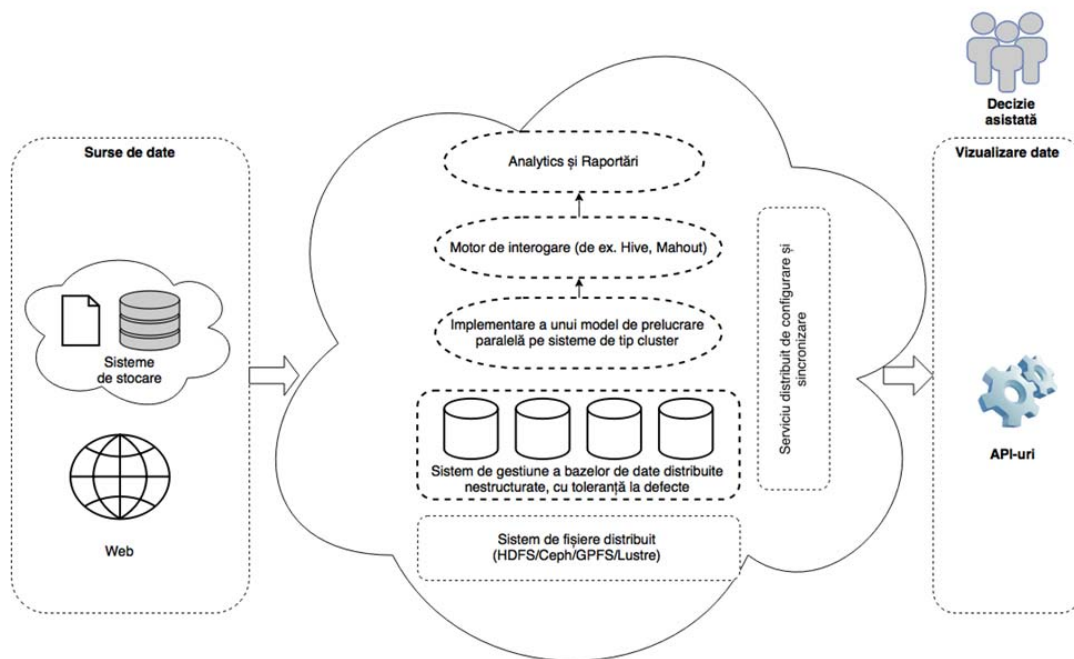


Figura 5. Arhitectură de referință pentru Analytics în Cloud [26].

5. Concluzii

Lucrarea analizează tematica managementului, guvernantei și analizei datelor de mari dimensiuni (Big Data) prin prisma relevanței sale pentru soluțiile pe care le oferă pentru creșterea competitivității în sectoarele de specializare inteligentă la nivel național. Dintre ideile evidențiate de această analiză menționăm:

- Organizațiile înnoată într-un ocean de date în expansiune care sunt fie prea voluminoase, fie prea nestructurate pentru a putea fi gestionate și analizate prin metode tradiționale.
- Fenomenul Big Data este dinamic, volumul datelor care se acumulează într-o organizație crește continuu, în mod exponențial și cu viteze din ce în ce mai mari.
- Guvernanța și integrarea informațiilor sunt factori esențiali pentru a obține valoarea maximă dintr-un proiect de tip Big Data.
- Spre deosebire de abordarea clasică, în gestionarea Big Data prelucrările sunt aduse în proximitatea datelor.
- Organizațiile care valorifică Big Data se diferențiază prin importanța acordată fluxurilor de date (în detrimentul datelor care se referă la trecut), promovarea experților în date, coroborată cu migrarea analizei avansate a datelor dinspre sistemul TIC spre activitățile de bază și implementarea schimbărilor organizatorice aferente acestei migrări;
- Volumul și viteza crescute ale datelor în mediile de producție vor determina organizațiile să dezvolte procese continue pentru colectarea, analiza și interpretarea datelor.
- Organizațiile de succes vor instrui și recruta oameni cu un nou set de aptitudini, capabili să integreze noile capacități de Analytics în mediile lor de producție.
- Pe măsură ce Big Data evoluează, arhitecturile se vor dezvolta într-un ecosistem de informații: o rețea de servicii interne și externe partajând continuu informații, optimizând deciziile, comunicând rezultatele și generând noi perspective pentru afaceri.
- Scopul fundamental al științei datelor, în cadrul unei organizații, este de a promova, sprijini și ameliora în permanență procesele de decizie determinate de date.

- Tendința este ca succesul și valoarea afacerii să se bazeze pe formula „business = date + algoritmi”. Companiile care adoptă un model de afaceri bazat pe software sunt reactive și „conduse de date” (data-driven).
- A deveni companie condusă de date (data driven) presupune cultivarea și adoptarea unei mentalități conform căreia desfășurarea afacerii este bazată pe utilizarea continuă a tehnicilor de tip Analytics în luarea deciziilor de afaceri pe bază de fapte.
- Pe măsură ce utilizarea Analytics de către companii se maturizează, adoptarea se va răspândi și colaborarea între echipe și departamente se va îmbunătăți în mod continuu.
- Calitatea datelor nu mai poate fi exprimată și măsurată în funcție de cât de adecvate sunt unui anumit scop, ci trebuie să fie evaluate din perspectiva unor scopuri multiple, luând în considerare toate utilizările și cerințele de calitate din aval.
- Tipologia proiectelor Big Data ia în considerare cerințele de utilizare a datelor (la nivel agregat / individualizat) și de valabilitate a datelor (pe termen scurt / lung).
- Calculul în Cloud și Big Data sunt strâns legate. Arhitectura de referință pentru Analytics în Cloud include ca element central conceptul de “lac de date”, un spațiu de stocare vast, consolidat, care evită limitările metodelor de administrare și prelucrare, specifice structurilor de date clasice.

Rezultatele acestei analize vor fi valorificate prin investigarea soluțiilor specifice identificate, în vederea demonstrării utilității acestora pentru realizarea de produse și servicii în beneficiul întreprinderilor din sectoarele de specializare inteligentă.

BIBLIOGRAFIE

1. **SN-ADR:** Strategia Națională privind Agenda Digitală pentru România 2020. Ministerul pentru Societatea Informațională, Februarie 2015. <https://www.google.ro/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0ahUKEwitjO2SrQHJAhVFOxoKHXERBcQQFggcMAA&url=http%3A%2F%2Fwww.mcsi.ro%2FCMSPage.s%2FGetFile.aspx%3Fnodeguid%3D0617c1d7-182f-44c0-a978-4d8653e2c31d&usq=AFQjCNGvMiD0UUKjm5Jr4Wgj0HpNUbXUXQ>
2. **SNC:** Strategia Națională pentru Competitivitate 2014 – 2020. Ministerul Economiei, iunie 2014. http://www.minind.ro/PROPUNERI_LEGISLATIVE/2014/SNC_2014_2020.pdf
3. **POC:** Programul Operațional Competitivitate 2014-2020. Ministerul Fondurilor Europene, 2014. <http://http://www.fonduri-ue.ro/files/programe/COMPETITIVITATE/POC/VO.POC.2014-2020.18.12.2014.pdf>
4. **SN-CDI 2020.** Strategia Națională de Cercetare-Inovare 2014-2020 – versiunea tehnică. Ministerul Educației și Cercetării Științifice, februarie 2014. http://www.cdi2020.ro/wp-content/uploads/2014/02/STRATEGIA_Versiunea-tehnica_Februarie-2014.pdf
5. **DIMITRIU, R.:** Ce înseamnă guvernanta? Ce sens are termenul în legătură cu dezvoltarea de proiecte? Biroul Virtual de Consultanță Antreprenorială (site cofinanțat din Fondul Social European prin Programul Operațional Sectorial Dezvoltarea Resurselor Umane 2007-2013 Investește în oameni!). <http://www.e-birouvirtual.ro/node/3726#comment-3731>
6. **OTTO, B.:** A Morphology of the Organisation of Data Governance. In: Proceedings of the 19th European Conference on Information Systems (ECIS 2011). Paper 272. Helsinki, Finland, 2011. <http://aisel.aisnet.org/ecis2011/272>
7. **KORHONEN, J.J.; MELLERI, I.; HIEKKANEN, K.; HELENIUS, M.:** Designing Data Governance Structure: An Organizational Perspective. În: GSTF Journal of Computing (JoC), 2(4), 2013. <http://dl6.globalstf.org/index.php/joc/article/viewFile/576/592>

8. **TALLON, P. P.; RAMIREZ, R. V.; SHORT, J. E.:** The information artifact in IT governance: Toward a theory of information governance. In: Journal of management information systems-JMIS, 30(3), 2013, pp. 141-177. ISSN 0742-1222. http://www.researchgate.net/profile/Paul_Tallon/publication/269500452_The_Information_Artifact_in_IT_Governance_Toward_a_Theory_of_Information_Governance/links/5547ce510cf2e2031b3849d7.pdf
9. **DAVENPORT, T. H.; BARTH P.; BEAN, R.:** How 'Big Data' Is Different. In: MIT Sloan Management Review 54(1), 2012. <http://sloanreview.mit.edu/article/how-big-data-is-different/>
10. **BANERJEE, A.; BANDYOPADHYAY, T.; ACHARYA, P.:** Data analytics: Hyped up aspirations or true potential. In: Vikalpa, 38(4), 2013, pp. 1-11. <http://www.vikalpa.com/pdf/articles/2013/04-Perspectives.pdf>
11. **CHUI, M.; MANYIKA, J.; KUIKEN, S. V.:** What executives should know about open data. In: McKinsey Quarterly, January, 2014. <https://digitalstrategy.nl/files/2014.01-H-What-executives-should-know-about-open-data.pdf>
12. **PROVOST, F.; FAWCETT, T.:** Data science and its relationship to big data and data-driven decision making. In: Big Data, 1(1), 2013, pp. 51-59. http://www.researchgate.net/profile/Tom_Fawcett/publication/256439081_Data_Science_and_its_relationship_to_Big_Data_and_data-driven_decision_making/links/02e7e5228cce561fd4000000.pdf
13. **O'NEIL, C.; SCHUTT R.:** Doing Data Science: Straight Talk from the Frontline, O'Reilly Media, Inc. 2013. http://cdn.oreillystatic.com/oreilly/booksamplers/9781449358655_sampler.pdf
14. **DAVENPORT, T. H.; PATIL, D. J.:** Data scientist: The sexiest job of the 21st century. In: Harvard Business Review, oct. 2012. <https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century/>
15. **MEIJER, E.; KAPOOR, V.:** The responsive enterprise: Embracing the hacker way. In: Communications of the ACM, 57(12), 2014, pp. 38-43.
16. **EUROPEAN COMMISSION:** Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions - Towards a thriving data-driven economy. COM(2014) 442, 2.7.2014. http://ec.europa.eu/information_society/newsroom/cf/dae/document.cfm?doc_id=6210
17. **TERADATA:** The virtuous circle of data - Engaging employees in data and transforming your business. Teradata White Paper, 2014.
18. **BRYNJOLFSSON, E.; HITT, L. M.; KIM, H. H.:** Strength in numbers: How does data-driven decision making affect firm performance? MIT - Sloan School of Management and University of Pennsylvania - Operations & Information Management Dept, 22.04.2011. http://papers.ssrn.com/sol3/Delivery.cfm/SSRN_ID1968725_code1376648.pdf?abstractid=1819486&mirid=1
19. **LOSHIN, D.:** Data Governance for Big Data Analytics: Considerations for Data Policies and Processes. Data Informed, 12.02.2013. <http://data-informed.com/data-governance-for-big-data-analytics-considerations-for-data-policies-and-processes/>
20. **WLADAWSKY-BERGER, I.:** Data-Driven Decision Making: Promises and Limits, 27.09.2013. <http://blogs.wsj.com/cio/2013/09/27/data-driven-decision-making-promises-and-limits/>
21. **CORRIGAN, D.:** Integrating and governing big data, IBM White Paper, 2013. [https://www-950.ibm.com/events/wwe/grp/grp037.nsf/vLookupPDFs/Integrating_Governing_BigData/\\$file/Integrating_Governing_BigData.pdf](https://www-950.ibm.com/events/wwe/grp/grp037.nsf/vLookupPDFs/Integrating_Governing_BigData/$file/Integrating_Governing_BigData.pdf)
22. **GANDOMI, A.; HAIDER, M.:** Beyond the hype: Big data concepts, methods, and analytics. In: International Journal of Information Management, 35(2), 2015, pp. 137-144. <http://ac.els->

cdn.com/S0268401214001066/1-s2.0-S0268401214001066-main.pdf?_tid=f28bfe04-7bc1-11e5-a2a4-00000aacb35e&acdnat=1445851043_78a530da1a1bc187fa3c2f31bf7ffa5e

23. **DEMCHENKO, Y.; DE LAAT, C.; MEMBREY, P.:** Defining architecture components of the Big Data Ecosystem. In Proceedings of the International Conference on Collaboration Technologies and Systems (CTS), mai 2014, IEEE, pp. 104-112. <http://uazone.org/demch/papers/bddac2014-bd-ecosystem-archi-v05.pdf>
24. **PÄÄKKÖNEN, P.; PAKKALA, D.:** Reference Architecture and Classification of Technologies, Products and Services for Big Data Systems. În: Big Data Research, 2015. <http://www.sciencedirect.com/science/article/pii/S2214579615000027>
25. **HASHEM, I. A. T.; YAQOUB, I.; ANUAR, N. B.; MOKHTAR, S.; GANI, A.; KHAN, S. U.:** The rise of “big data” on cloud computing: review and open research issues. In: Information Systems, 47, 2015, pp. 98-115. http://umexpert.um.edu.my/file/publication/00001293_117865.pdf
26. **SULLIVAN, J.; ESCARAVAGE, J.; GUERRA, P.:** The Cloud Analytics Reference Architecture: Harnessing Big Data to Solve Complex Problems, Booz Allen Hamilton White paper. McLean, VA, SUA, 2014. <http://www.boozallen.com/media/file/the-cloud-analytics-reference-architecture-vp.pdf>