Str. Italiană nr. 28, sect. 2, București – România
weh@spiruharet.ro weh.spiruharet.ro

**Categoria de servicii:** ***WP 3. Servicii de dezvoltare a competențelor și de formare profesională (Skills Development and Training Services)***

**Subcategoria de servicii:** ***Sesiuni de instruire pe soluții digitale avansate (Training sessions on high performance digital solutions)***

# *SESIUNE DE INSTRUIRE PENTRU BUSINESS INTELLIGENCE ȘI ENTERPRISE RESOURCE PLANNING*

## - Big data. Asigurarea securității datelor organizațiilor -

**Partener WEH: Universitatea *Spiru Haret***

**Trainer de coordonare a transformarii digitale:**

**Andronie Mihai**

**2024**

# CUPRINS

# 1. Masivele de date (Big Data) – volume mari de date cu caracter complex

Pe parcursul ultimilor ani tehnologia informației a atins un nivel de dezvoltare fără precedent, pătrunzând pe toate nivelele de activitate umană, având ca rezultat generarea și colectarea unor volume de date din ce în ce mai mari, în formate complexe și variate. Acest tip de date a fost denumit de către specialiștii în domeniu *big data*, termen ce ar putea fi tradus în limba română prin *masiv de date* sau *date mari*. Pentru a sugera complexitatea și volumul acestor tipuri de date, va fi utilizată în continuare denumirea de *masiv de date*.

Gartner, unul din actorii principali la nivel global pe piața tehnologiilor informaționale, a definit masivele de date ca fiind "date în volum mare, care se schimbă cu viteză mare și care sunt de obicei în formate foarte variate și necesită sisteme de procesare a informației eficiente, inovatoare în vederea îmbunătățirii înțelegerii contextului de afaceri și a adoptării deciziilor" [Gartner big data].

Specialiștii în domeniul tehnologiei informației consideră că masivele de date au trei proprietăți specifice care sunt integrate într-um model denumit "modelul 3V" al masivelor de date, denumirea de "3V" provenind de la inițialele celor trei proprietăți esențiale, acestea fiind [Techtarget 3V]:

- *Volum* - datele mari sunt stocate în cantități mari, ceea ce face imposibil (sau foarte dificil) procesul de analiză sau procesare a acestora fără aplicații software dedicate care, la rândul acestora, necesită putere de calcul foarte mare; pentru a putea gestiona datele mari este nevoie, de asemenea, de dispozitive de stocare mari, cantitatea acestora fiind limitată în general de capacitatea de stocare de pe aceste dispozitive;

- *Varietate* - datele mari se prezintă de obicei într-o mare varietate de formate; varietatea datelor mari poate reprezenta o problemă importantă, dificil de depășit pentru cei care dezvoltă algoritmi de explorare sau procesare a acestora; de asemenea, varietatea datelor mari, face ca procesul de explorare a acestora să fie cu mult mai dificil decât procese similare utilizate în cazul surselor de date convenționale, cum ar fi depozitele de date structurate sau bazele de date relaționale;

- *Viteză* - datele mari sunt generate aproape continuu din diferite procese tehnologice sau economice, cum ar fi cele găsite în industrie, afaceri, economie: date legate de producţie, date provenind de la diferiţi senzori, date economice etc.

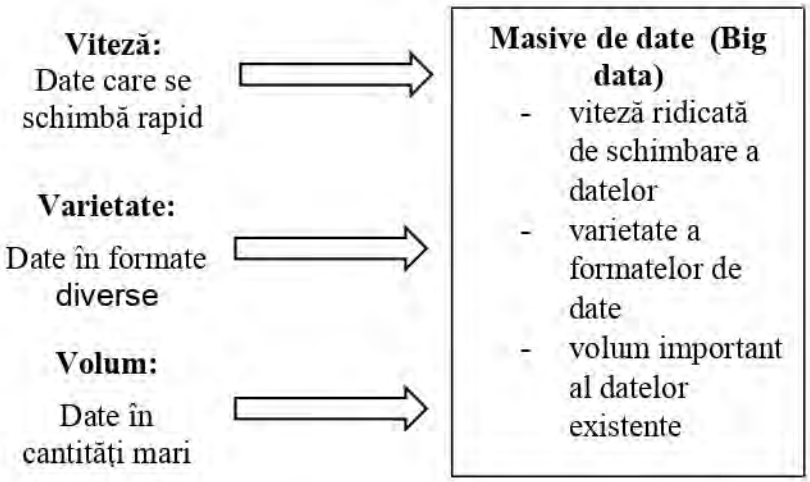


Figura 1. Modelul 3V al masivelor de date

Există anumiţi specialişti care consideră că se poate identifica şi o a patra caracteristică importantă a masivelor de date, şi anume veracitatea acestora [IBM big data], conducând către un "model 4V" al masivelor de date. Veracitatea datelor din masivele de date se referă la caracteristica acestora de a fi de slabă calitate. Sursele de date de care dispun companiile sunt de multe ori neverificate, rezultatele obţinute în urma analizei acestora putând fi la rândul lor greu de verificat. Din această cauză, instrumentele software de analiză a datelor din masive de date trebuie să aibă şi funcţia de asigurare a consistenţei acestora.

De obicei datele pentru care sunt întâmpinate cele mai mari probleme în asigurarea consistenţei sunt datele de pe internet, ştirile, paginile web etc.

Conform IBM Big Data & Analytics Hub [IBM big data], "unul din trei lideri de afaceri nu are încredere în informaţiile furnizate de către sistemele specializate de analiză a masivelor de date (inclusiv a sistemelor business intelligence) pentru a lua decizii, estimându-se că anual peste trei miliarde de dolari sunt cheltuiţi defectuăs din cauza datelor inexacte". Din aceste considerente, "identificarea de soluţii eficace de analiză a datelor în formate nestructurate la nivelul companiilor

este văzută ca un factor important în ceea ce privește dezvoltarea acestora" [Chaudhuri S. 2013]. Având în vedere slaba calitate a datelor de care companiile dispun, proiectarea și realizarea de instrumente performante de analiză a datelor reprezintă o provocare pentru dezvoltatorii software.

Pentru a ilustra mai bine contextul colectării, stocării și prelucrării masivelor de date va fi luat ca exemplu un domeniu economic de activitate unde sunt generate volume semnificative de date: *domeniul aviației*.

Acumularea unor volume mari de date în industria aeronautică poate fi văzută ca o oportunitate de a exploata aceste date prin intermediul unor instrumente specializate, în vederea obținerii de informații valoroase care pot fi utilizate de către manageri și alte persoane responsabile pentru a dezvolta și a îmbunătăți procesele care sunt desfășurate de companiile aeriene [Andronie M. 2015 A]. Potrivit unui document publicat în 2013, "liniile aeriene, aeroporturile, producătorii de aeronave, furnizorii, guvernele și alte părți interesate în domeniul aviației, depind de date pentru planificarea operațională și la nivel executiv. Seturile de date complexe și concurente crează provocări enorme din punct de vedere tehnic și uman în ceea ce privește colectarea, sortarea, și extragerea de cunoștințe din volumele de date disponibile. Seturile de date din domeniul aviației depășesc capacitățile de calcul ale unui calculator personal [Tulinda L. 2013] ".

Datele din domeniul aviației sunt de obicei în cantități mari, având formate variate și schimbându-se cu rapiditate, având astfel toate caracteristicile masivelor de date definite anterior. Masivele de date care provin din industria aeronautică, conform lucrării *Cross-Platform Aviation Analytics Using Big-Data Methods*, pot avea multiple surse [Tulinda L. 2013]:

- Date de urmărire a curselor la nivel local sau global;
- Informații aferente pasagerilor;
- Informații asociate diverselor operațiuni aeroportuare;
- Informații despre aeronavele disponibile;
- Date referitoare la condițiile meteorologice;
- Informații privind companiile aeriene care activează;

- Informații economice referitoare la activitățile companiilor aeriene;
- Rapoarte de securitate aeriană.

Sistemele *business intelligence* trebuie să fie capabile pentru analiza unor date cu complexitate crescută, oferind rezultate cu acurateţe ridicată pentru a oferi suport decizional adecvat pentru managerii companiilor. Din aceste considerente este important ca datele existente să fie analizate în ansamblul acestora. Doar după integrarea tuturor surselor de date devine posibilă realizarea de corelaţii între acestea şi este posibil să se obţină cele mai bune rezultate.

În figura 2. este prezentat un circuit de date propus pentru extragerea de cunoştinţe din datele specifice industriei aviatice.

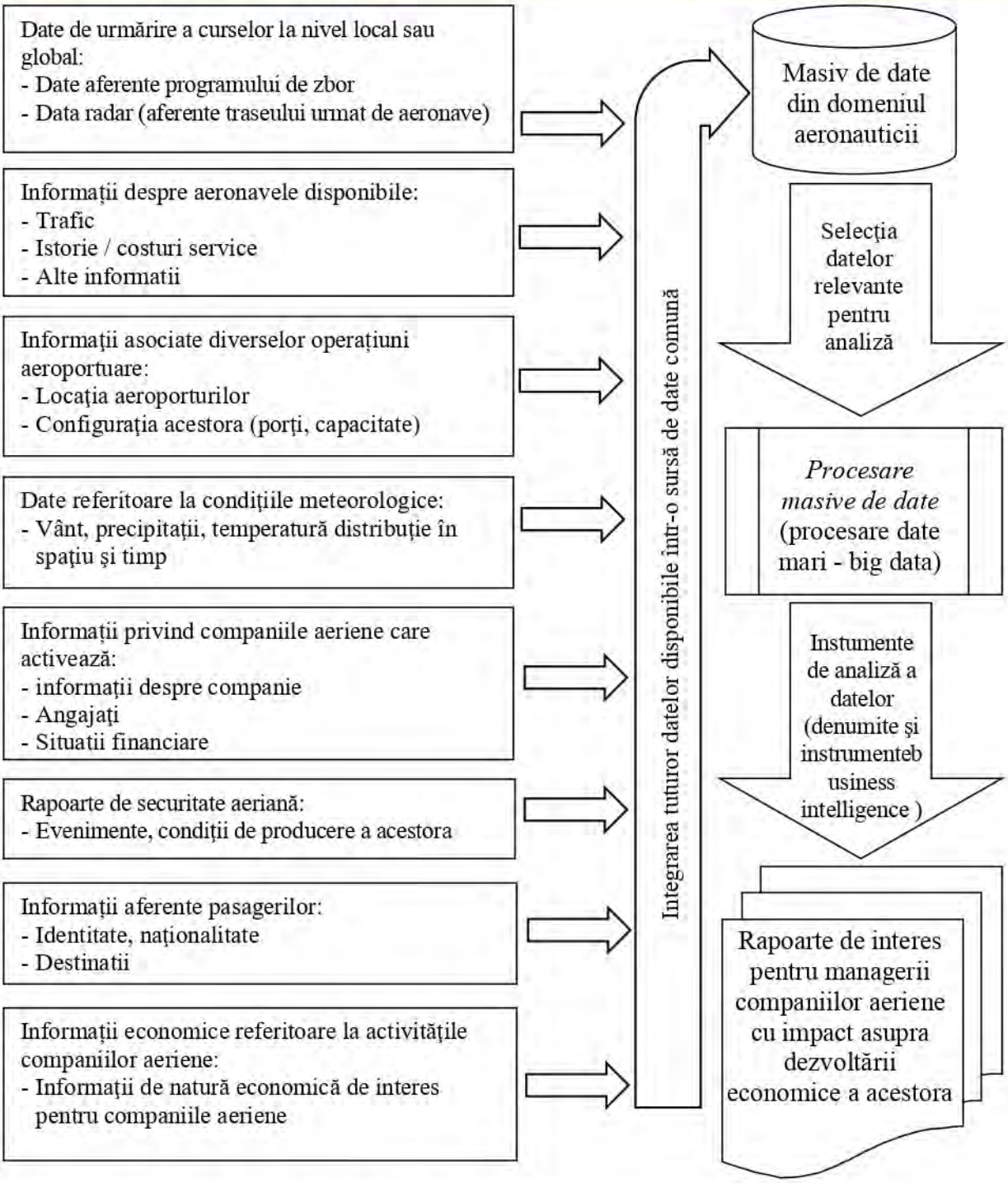


Figura 2. Exemplu privind sursele de date din industria aeronautică şi procesarea acestora

Se constată necesitatea integrării datelor disponibile într-un masiv de date unitar, înainte ca acestea să poată fi analizate cu ajutorul unor sisteme software specializate, cunoscute şi sub numele de sisteme de tip *business intelligence* (inteligenţa afacerii).

# 2. Asigurarea securităţii datelor de care dispun companiile

## 2.1. Probleme care pot afecta datele deţinute de către organizaţiile economice

Toate tipurile de instituții economice au potențialul de a colecta, în timp, cantități mari de date legate de activitățile desfășurate de acestea. De obicei, este posibil ca managementul unor astfel de instituții să utilizeze cantitatea de date disponibilă pentru a adopta decizii informate, bazate pe cunoștințele acumulate pe perioade lungi de timp, pe lângă experiența personală a factorilor decizionali. În acest fel, succesul pe termen lung poate fi atins, iar instituțiile își pot îmbunătăți calitatea ofertei pentru clienți [Miller G. 2006].

Datele suport decizional sunt de obicei stocate în baze de date specializate sau depozite de date care sunt găzduite pe servere specializate care trebuie să fie protejate împotriva tuturor tipurilor de evenimente care pot împiedica funcționarea corespunzătoare a acestora [Techopedia].

Problemele care pot afecta datele deținute de organizațiile economice pot avea mai multe surse, având cauze atât la nivel local cât și la nivel global [Andronie M. 2015 B]. Aceste amenințări sunt reprezentate în Figura 3.

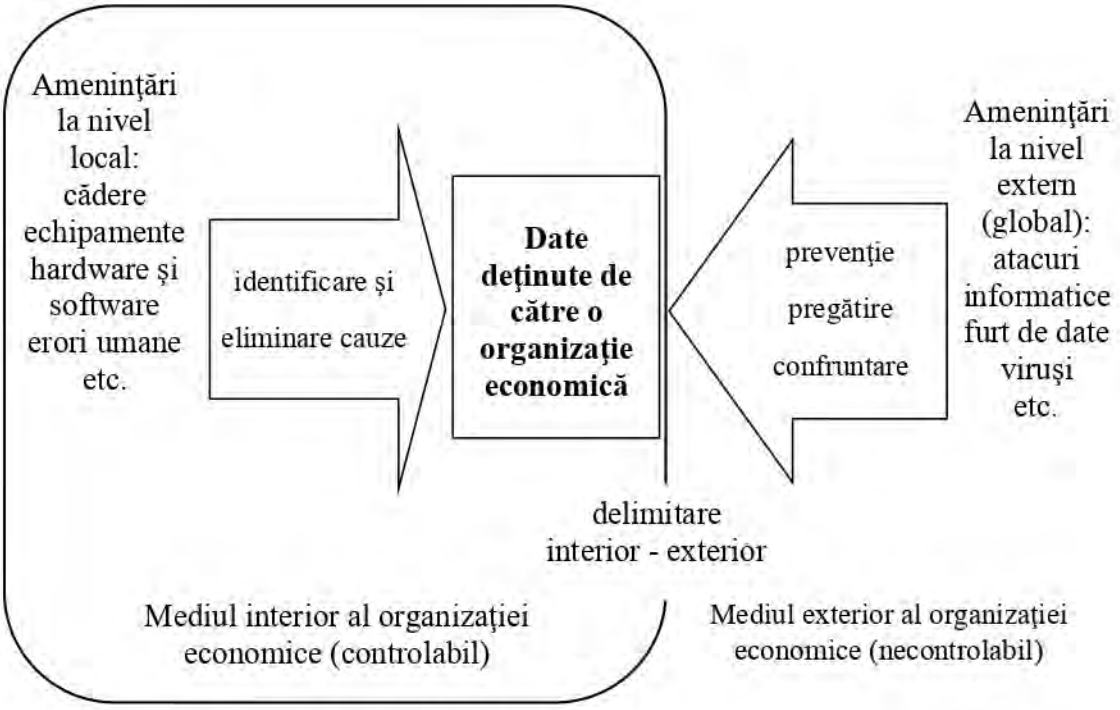


Figura 3. Ameninţări care afectează datele unei companii

Conform celor prezentate în figura precedentă, a fost definită o limită între amaninţările la nivel global, sau ameninţări externe companiei şi ameninţările la nivel local, care de cele mai multe ori sunt ameninţări interne companiei. Această limită separă două zone distincte.

- În interiorul acestei limite este un mediu local unde lucrurile pot fi mai mult sau mai puțin controlate și cauzele problemelor pot fi de obicei eliminate inainte de a conduce la consecințe grave pentru organizația economică;
- În exteriorul limitei respective, cauzele ameninţările nu pot fi de obicei controlate de către organizația economică și, în consecință, aceste ameninţări trebuie să fie depăşite, compania trebuind să fie pregătită în permanenţă pentru a le face față.

*Ameninţări la nivel global (ameninţări externe)*

Potrivit *Global Risks Report* din 2015 [Global Risks Report 2015] emis de către World Economic Forum, la nivel global, pot fi identificate o serie de ameninţări cu privire la domeniul tehnologic,

care pot fi relevante pentru companiile care au volume mai și foarte mari de date și care le exploatează prin utilizarea de instrumente de tip *business intelligence*. Cele mai importante probleme care pot apărea la nivel global sunt prezentate în tabelul 1, împreună cu probabilitatea și impactul potențial, reprezentate pe o scală numerică.

Tabelul 1. Riscuri la nivel global [Global Risks Report 2015]

| ***Posibilă problemă tehnologică*** | ***Şansa de apariţie*** | ***Impact Potenţial*** |
|---|---|---|
| *Cedarea infrastructurii critice* | *4.3* | *5.1* |
| *Atacuri informatice* | *5.1* | *4.9* |
| *Fraudarea sau furtul datelor* | *5.2* | *4.5* |
| *Utilizarea eronată a tehnologiilor* | *4.3* | *4.4* |

S-a constatat că nu toate ameninţările prezentate în tabelul precedent afectează o companie în acelaşi fel. Deşi aceste ameninţări pot avea un impact semnificativ asupra societăţii, doar unele dintre ele afectează şi companiile.

Dintre cele patru tipuri de ameninţări la nivel informatic, unele sunt mai întâlnite în cadrul companiilor, cum ar fi: căderea infrastructurilor critice, atacurile informatice şi furtul datelor.

O companie poate lua o serie de măsuri pentru a se confrunta cu ameninţările externe, printre acestea cele mai eficiente din punct de vedere economic fiind cele de pregătire pentru a le întâmpina, cu consecinţe minime.

*Ameninţări la nivel local (ameninţări interne companiei)*

Ameninţările la nivel local afectează compania de obicei din interiorul acesteia. Din acest motiv se poate considera că o companie are control de obicei asupra cauzelor care pot conduce la aceste ameninţări. Aceasta este principala diferenţă între ameninţările la nivel local (interne) şi cele la nivel global (externe).

Ameninţările la nivel local pot fi gestionate de către o companie, în cele mai multe cazuri putând fi eliminate cauzele acestora.

În acest context, cea mai bună politică pe care o companie o poate adopta este identificarea ameninţărilor locale cât mai devreme pentru a le putea înlătura cauzele cu costuri minime înainte de a se ajunge la consecinţe costisitoare.

*Cele mai comune probleme cu care companiile se confruntă cu privire la securitatea datelor deţinute*

În ceea ce privește metodele de stocare a datelor, pot fi identificate o serie de probleme diferite, care pot apărea în cazul datelor colectate de către organizațiile economice. Conform lucrării *Ensuring Security of Data Used by Economic Organizations for Decision Support* problemele cele mai frecvent întâlnite în cadrul intreprinderilor sunt:

- Pierderea datelor (temporară sau permanentă) - pierderile de date, din punct de vedere al instituțiilor care se bazează pe acestea, înseamnă că operațiunile sunt împiedicate, cu consecințe cum ar fi deciziile nepotrivite, eficiența redusă etc.; pierderile de date pot fi ireversibile sau nu, în funcție de măsurile de securitate a datelor adoptate de către proprietarii acestora; posibile cauze ale pierderilor de date pot fi: probleme tehnice ale server-elor, probleme hardware, dezastre (incendiu, furt etc.), probleme de software, erori de operare ale utilizatorilor şi/ sau ale administratorilor sistemelor informatice etc.

- Pierderea accesului la date (temporar sau permanent) - pierderea accesului la date este diferită de pierderea datelor, principala diferență fiind aceea că datele nu sunt distruse, și, cel puțin în teorie, acestea ar putea fi accesate din nou. Cele mai multe dintre problemele de acces la date sunt temporare, cauzate de evenimente, cum ar fi întreruperea alimentării cu energie electrică, probleme de rețea/ de conectare la Internet, etc. Consecințele pierderii accesului la date pot fi la fel de serioase ca cele ale pierderii datelor, dar de obicei sunt mai puțin grave decât acestea pentru că în cele mai multe cazuri, după ce problemele sunt rezolvate, operațiunile companiilor pot reveni la normal.

- Accesul neautorizat la date de către alte persoane sau organizații - accesul neautorizat la date poate fi foarte grav și poate fi potențial cauza tuturor celelalte probleme. Dacă persoane neautorizate au acces la date, ele pot distruge sau pot modifica datele, uneori fără ştiinţa proprietarilor acestora. O consecință nedorită a accesului neautorizat la datele unei companii

este și furtul de date. Datele deținute de o companie pot fi utilizate de către concurenții acesteia pentru a obține un avantaj pe piață.

- Alterarea datelor - modificarea datelor poate avea consecințe grave pentru o companie, cu consecințe cum ar fi adoptarea de decizii nepotrivite, satisfacția redusă a clienților, produse defecte etc. Alterarea datelor poate avea o multitudine de cauze, cum ar fi erorile umane, erori de software, căderea echipamentelor hardware, accesul neautorizat (interferențe externe) etc. O companie poate implementa politici diferite pentru a preveni alterarea datelor.

Figura 4 rezumă principalele probleme care afectează datele companiilor, împreună cu cauzele și consecințele acestora. Fiecare companie își poate particulariza această schemă pentru problemele sale legate de securitatea datelor, în scopul de a urmări cauzele acestor probleme și de a găsi soluții.

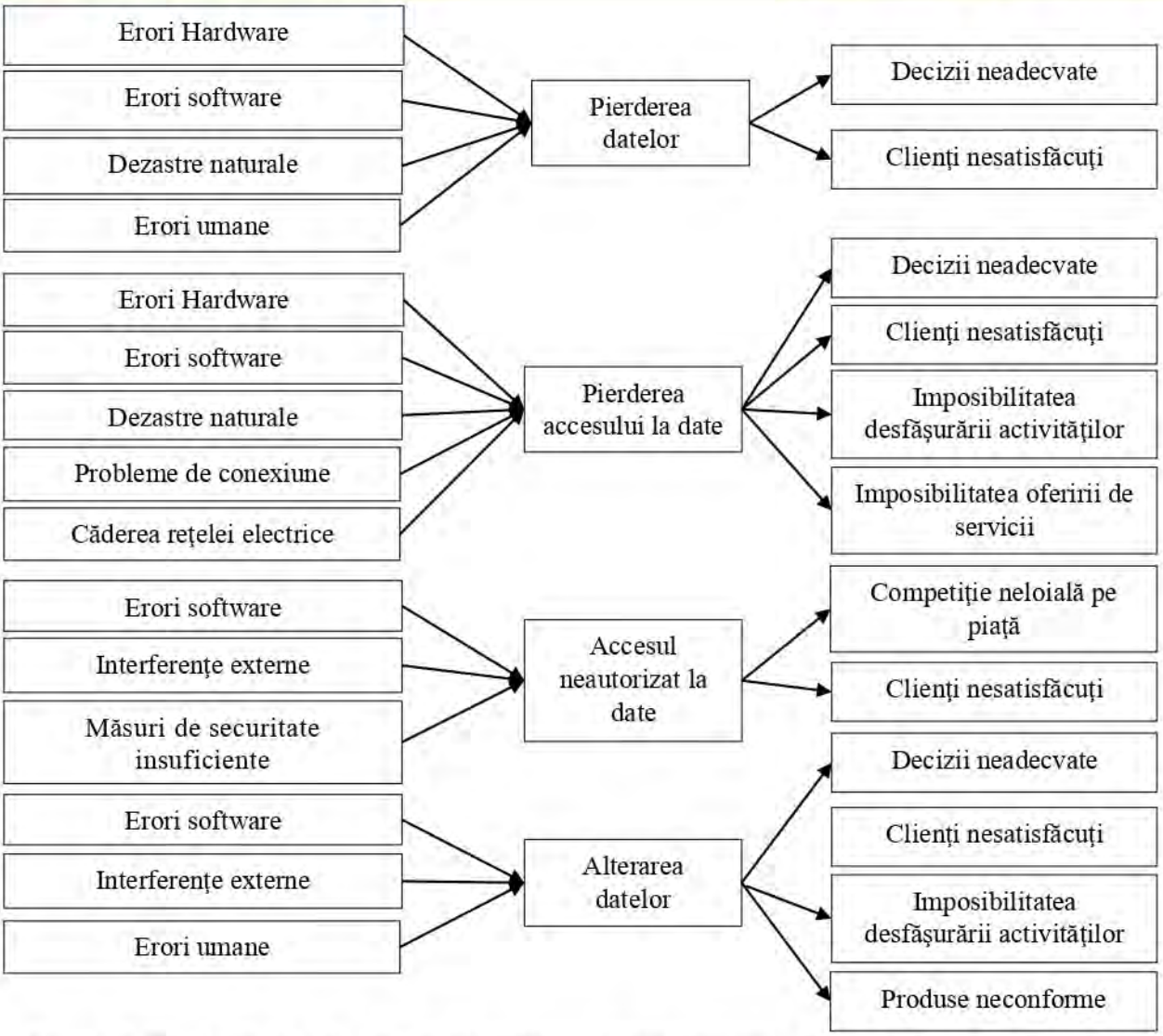


Figura 4. Probleme care pot afecta securitatea datelor companiilor, cauze, factori contributivi și consecințe ale acestora

## 2.2. Niveluri de securitate ale datelor companiilor

Datele deținute de organizații economice trebuie să fie stocate într-un mod securizat, astfel încât probabilitatea de compromitere a acestora să fie cât mai redusă posibil, crescând astfel capacitatea organizației de a profita de acestea.

În conformitate cu cele precizate în lucrarea *Ensuring Security of Data Used by Economic Organizations for Decision Support,* pot fi identificate o serie de niveluri la care trebuie să se asigure un nivel de securitate a datelor adecvat:

- Nivelul de securitate a accesului fizic - este vorba de controlul accesului neautorizat la echipamentele de stocare a datelor deținute de către organizația economică; atunci când datele sunt stocate în cloud (serviciul de depozitare a acestora este externalizat), asigurarea accesului la dispozitive de stocare a datelor nu mai revine organizației economice;

- Nivelul de securitate hardware - presupune utilizarea de echipamente hardware fiabile, concepute pentru a fi rezistente la erori / defecte (de exemplu o politică de securitate des întâlnită este dublarea resurselor critice pentru a asigura integritatea datelor);

- Nivelul de securitate software - implică utilizarea de programe care permit accesul doar persoanelor autorizate (de exemplu, folosind parole și alte sisteme de securitate care împiedică accesul persoanelor neautorizate); este de asemenea esențial să se utilizeze sisteme software care să aibă facilități de recuperare a datelor în cay că apar probleme;

- Nivelul de securitate a datelor - se pot pune în aplicare o serie de politici opționale, cum ar fi crearea periodică de copii ale datelor (backup).

Fiecare dintre cele patru niveluri prezentate mai sus este prezentat în Figura 5.

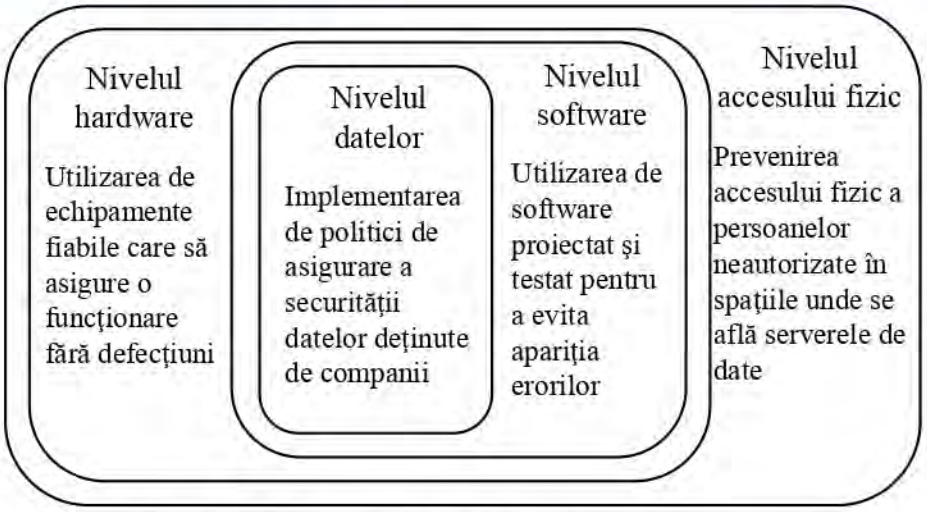


Figura 5. Nivelele de securitate a datelorunei organizaţii economice

## 2.3. Model pentru asigurarea securităţii datelor unei companii

În secţiunea curentă se propune un model de asigurare a securităţii datelor companiilor economice, având în vedere faptul că acestea sunt baza sistemelor de tip *business intelligence* pe care aceste companii le pot utiliza pentru fundamentarea deciziilor economice.

Este necesar ca modelul propus să îndeplinească simultan două cerinţe: să acopere cele mai comune probleme care afectează datele de care dispune o companie; să ofere soluţii pe multiple nnivele de apariţie a acestor probleme.

Pentru îndeplinirea simultană a celor două cerinţe enumerate în paragraful precedent, modelul propus este sub formă matriceală, având următoarele două dimensiuni:

- Tipuri de probleme ce afectează datele de care dispune o companie;

- Niveluri la care aceste tipuri de probleme pot apărea.

În tabelul 2 este prezentat sub formă de tablou bidimensional modelul pentru asigurarea securității datelor unei companii.

Tabelul 2. Model matriceal pentru asigurarea securității datelor unei companii

| | ***Nivelul securității accesului fizic*** | ***Nivelul securității hardware*** | ***Nivelul securității software*** | ***Nivelul securității datelor*** |
|---|---|---|---|---|
| ***Pierderea datelor*** | • Asigurarea securității fizice a spațiilor unde se află server-ele<br>• Păstrarea unor copii ale datelor importante în locații multiple | • Utilizarea de echipamente cu fiabilitate ridicată pentru a minimiza riscul defectării<br>• Crearea de planuri de recuperare în caz de dezastre | • Utilizartea de produse software specializate pentru recuperarea datelor în caz de necesitate | • Implementare a unor politici de tip *back-up/ recovery* eficace<br>• Crearea unor planuri de escaladare a problemelor pentru rezolvarea cât mai rapidă a acestora |
| ***Pierderea accesului la date*** | • Asigurarea unei conexiuni la date care să fie stabilă<br>• Asigurarea redundanței conexiunii la date | • Utilizarea de echipamente de teecomunicații fiabile<br>• Utilizarea de servicii de stocare date/ telecomunicați i acreditate | • Utilizarea unor produse software care să creeze automat copii ale datelor cele mai utilizate | • Păstrarea de copii ale datelor în locații multiple, ușor accesibile<br>• Crearea de planuri de escaladare a problemelor pentru rezolvarea rapidă a acestora |

| | | | | |
|---|---|---|---|---|
| ***Accesul neautorizat la date de către alte persoane sau organizaţii*** | • Asigurarea securităţii locaţiei unde sunt stocate datele<br>• Instalarea se alarme, sisteme antiefracţie<br>• Asigurarea de servicii de pază | • Instalarea echipamentele or hardware critice în locaţii unde doar persoanele autorizate au acces | • Criptarea datelor<br>• Utilizarea de produse software care implementează politici de securitate | • Implamantare a unor protocoale şi politici de securitate stricte<br>• Crearea de planuri de escaladare a problemelor pentru rezolvarea rapidă a acestora |
| ***Alterarea datelor*** | • Controlul accesului în locurile unde se află echipamentele de stocare a datelor | • Utilizarea de echipamente fiabile astfel încât să nu există posibilitatea compromiterii integrităţii datelor datorită unor erori | • Utilizarea de produse software testate corespunzător în vederea asigurării integrităţii datelor | • Păstrarea unor copii a datelor critice<br>• Crearea unor planuri de continuare a operaţiilor afacerii în caz de compromitere a datelor |

Sursa: Ensuring Security of Data Used by Economic Organizations for Decision Support

Având în vedere modelul propus, o companie poate să îşi proiecteze sisteme care să ofere performanţă pe fiecare nivel, creându-şi astfel propria politică de securitate. Modelul propus poate fi particularizat de fiecare companie în funcţie de problemele concrete cu care aceasta se confruntă.

Cu un model coerent de asigurare a securităţii datelor, o companie se poate asigura că rezultatele ce vor fi obţinute în urma analizei acestora cu ajutorul sistemelor de tip *business intelligence* vor fi unele pe care se poate baza în adoptarea deciziilor de afaceri.

Cercetări viitoare pot fi întreprinse pentru realizarea unui model de asigurarea a securităţii datelor îmbunătăţit, eventual cu mai multe dimensiuni. Un astfel de model ar putea fi ulterior particularizat pe domenii de activitate pentru a răspunde necesităţilor de asigurare a securtăţii datelor cât mai multor companii.