UNIVERSITATEA BABEŞ–BOLYAI
FACULTATEA DE ŞTIINŢE ECONOMICE ŞI GESTIUNEA AFACERILOR
DOMENIU: CIBERNETICĂ ŞI STATISTICĂ

Teză de Doctorat

-Rezumat-

Proiectarea și Implementarea Depozitelor de Date pentru Business Intelligence aplicate în Economie

CONDUCĂTOR ŞTIINŢIFIC

PROF. UNIV. DR. NIŢCHI ŞTEFAN IOAN

DOCTORANDĂ

NAGY ILONA MARIANA

Cluj-Napoca
2012

Cuprins

Cuprins extins

Cuvinte cheie

Business Intelligence, depozite de date, metodologii de dezvoltare a soluțiilor software, metadate, arhitectura depozitelor de date, framework pentru implementare, prototip, automatizarea proceselor;

Introducere

Progresul înregistrat în domeniul tehnologiei informației a condus evoluția sistemelor de procesare a datelor de la primele stadii ale aplicațiilor autonome până la sistemele analitice avansate ale mediului informational din zilele noastre, anume sisteme de Business Intelligence. În cadrul acestui context extins de sisteme informaționale, depozitele de date definesc un ansamblu de technologii apărute la începutul anilor 1990 ca rezultat al progreselor înregistrate în domeniul procesării datelor, cu precădere a procesării unor volume însemnate de date.

Tehnologia depozitelor de date reprezintă o componentă a framework-ului general de Business Intelligence, care cuprinde o gamă amplă de aplicații și unelte utilizate pentru analiza unor volume mari de date și pentru transformarea acestora în informație inteligibilă și cunoștințe specifice domeniului vizat. Această tehnologie vastă permite gestionarea mediului informațional în cadrul căruia o serie de componente asigură culegerea și integrarea datelor din cadrul întreprinderii. Scopul acestor procese este determinat de facilitarea accesului la date consolidate și structurate, pe baza cărora ultilizatorii finali își pot îmbunătăți procesul de luare a deciziilor.

Mediul de stocare efectivă al technologiei depozitelor de date este cunoscut sub numele de depozit de date. Acesta reprezintă un model al datelor dintr-o intreprindere, structurate special pentru facilitarea procesele de analiză și interogare. Depozitul de date definește o componentă esențială și dominantă a sistemelor suport de decizie dictate de date, având ca scop principal facilitarea procesului de luare a deciziilor bazate pe date reale prin asigurarea răspunsurilor la întrebări de business într-un mod precis și oportun. Pentru îndeplinirea acestui scop, depozitul de date este definit de modele de date proprii care specifică structura datelor în cadrul mediului de stocare. Aceste modele de date, optimizate pentru interogare și analiză, sunt create într-o manieră stabilă, consistentă și predictibilă cu ajutorul diferitelor tehnici de modelare. Totodată, procesele de interogare și analiză sunt facilitate prin intermediul diferitelor tipuri de metadate, menite să descrie structura prin care o intreprindere folosește informația și, deasemenea, menite să atașeze semantică proceselor de business și datelor rezultate din acestea.

Având în vedere nivelul ridicat de complexite, soluțiile de dezvoltare a depozitelor de date presupun o abordare structurată și planificată, definită sub forma unei metodologii însoțită de un framework arhitectural adecvat. Metodologiile sunt destinate atingerii unor rezultate în conformitate cu specificații bine definite și asigurării unor procese repetitive și consistente ce pot fi învățate. Architecturile reprezintă structuri care integrează toate componentele depozitului de date și asigură o structură solidă pentru integrarea la nivelul întregii intreprinderi. Alegerea unor metodologii și arhitecturi potrivite determină succesul de ansamblu al implementării soluțiilor de depozite de date. Un alt aspect esențial îl reprezintă utilizarea unui framework capabil să asigure un set de principii pentru descrierea

componentelor sale și a interoperabilității dintre acestea, și care să sprijine existența unui mediu de procese re-folosibile, integrare, consistență și flexibilitate în livrarea informațiilor.

i. Enunțarea Problemei și a Obiectivelor Cercetării

Depozitul de date, componentă importantă a mediului informțional, este definit de o serie de concepte esențiale, anume model de date, metodologie de dezvoltare, arhitectură, și framework. Proiectarea și implementarea acestuia reprezintă în multe cazuri o provocare supusă, asemenător tuturor proiectelor complexe, unor riscuri de eșec ridicate.

Literatura de specialitate, dar mai ales lucrările și rapoartele provenite din afara mediului academic, prezintă o serie de cazuri în care ratele de succes ale dezvoltării soluțiilor de depozite de date sunt influențate negativ de costurile ridicate și de intervalul de timp necesar pentru activitățile specifice de planificare, proiectare și implementare. Conform Adelman et al. [6], trei din zece situații care duc în general la eșec în dezvoltarea depozitelor de date sunt determinate de următoarele motive: 1) proiectul este peste bugetul alocat; 2) termenul de livrare este depășit, și 3) o serie de costuri ale proiectului sunt nejustificate. Alte riscuri cunoscute includ: schimbarea frecventă a cerințelor prioiectului din partea utilizatorilor finali, activități de gestionare a proiectelor deficiente, dezvoltarea unor architecturi slabe pentru soluțiile software propuse, lipsa datelor calitative, etc. Intreprinderile care beneficiază de rezultatele dezvoltării acestor soluții complexe de depozite de date, precum și companiile care se ocupă de dezvoltarea propriu-zisă sunt vulnerabile acestor riscuri, deoarece activitățile de proiectare, implementare și mentenanță, etc. implică eforturi financiare considerabile și sunt văzute în general ca fiind îndelungate și extrem de laborioase.

Dezvoltarea depozitelor de date este ghidată de abordări metodologice și arhitecturale menite să faciliteze livrarea unor soluții de succes în cadrul limielor definite de proiect. În acest caz, literatura și numeroase practici de succes oferă direcții comprehensive pentru proiectarea și implementarea depozitelor de date, direcții pe care intreprinderile le pot folosi și adapta nevoilor specifice. Aceste direcții de îndrumare se concentrează în principal pe activități legate de gesiunea proiectelor, modele și tehnici de modelare specifice depozitelor de date, precum și arhitecturi de referință în domeniu. Optimizarea proceselor de implementare, de exemplu în cazul activităților repetitive și consumatoare de timp, nu este însă tratată suficient în literatură. Astfel, noi considerăm că propunând un framework care să gestioneze aceste activități în condiții bine determinate, putem obține reducerea semnificativă a costurilor ridicate implicate în implementarea soluțiilor de depozite de date.

Prin urmare, principalele obiective ale cercetării noastre sunt determinate de reducerea costurilor (i.e. costuri de ansamblu ale dezvoltării soluției de depozite de date) și eficientizarea implementării cerințelor venite din partea utilizatorilor finali. În vederea îndeplinirii acestor obiective, ne propunem să realizăm o prezentare extesivă a aspectelor teoretice legate de conceptele de bază ale domeniului, anume sisteme suport de decizie și importanța lor în mediul de afaceri, architectura informațională a intreprinderii și framework-uri specifice, etc. Întelegerea acestor concepte este esențială pentru dezvoltarea cu succes a soluțiilor de data warehouse, acoperind următoarele perspective: *definirea* modelelor de date logice și fizice ale diverselor structuri de stocare, diferite tipuri de metadate și activități de

gestionare a datelor; *metodologia de dezvoltare* a depozitelor de date (gestiunea proiectului și activități de planificare, practici de succes și diferite standarde impuse la nivelul întregii întreprinderi, etc.); *aspecte arhitecturale* (sisteme utilizate, date și procesele implicate); *framework-ul* definit pentru asigurarea unui set de principii pentru dezvoltarea unor componente și a interacțiunii acestora în mediul informațional; și *detaliile de implementare* (unelte și aplicații specifice, echipa de implementare, limitări de timp și buget, etc.).

Mai precis, obiectivele cercetării sunt definite pentru fiecare concept în parte, astfel:

- *Abordarea metodologică*

  Metodologia definește un set de principii care guvernează proiectarea și implementarea soluțiilor software. Noi ne propunem introducerea unor metodologii generice în cadrul tezei și discutarea potrivirii acestora în contextul specific al soluțiilor de depozite de date. Intenția noastra constă în selecția unei metodologii adecvate în concordanță cu cerințele proiectului, urmărirea principiilor de dezvoltare, și validarea rezultatelor obținute.

- *Modelul de date*

  Modelul de date descrie, din punct de vedere logic și fizic, schema și proprietățile structurilor de date elaborate pentru stocarea acestora în mediile operaționale și analitice. Obiectivul nostru privind modelul de date, este realizarea unei prezentări comprehensive a opțiunilor de modelelare și a tehnicilor specifice, precum și a altor concepte referitoare la depozitele de date. Intenționăm selectarea și îmbunătățirea unei metodologii de modelare multi-dimensionale, și utilizarea acesteia pentru dezvoltarea modelului soluției de depozite de date propuse la nivelul intreprinderii.

- *Arhitectură*

  Planul arhitectural stă la baza proiectării și implementării soluției de depozite de date, fiind astfel esențial în procesul global de dezvoltare. Obiectivul nostru constă în prezentarea celor mai comune tipuri architecturale și a caracteristicilor acestora, selectarea unei arhitecturi potrivite conform unor criterii bine definite și compatibile cu metodologia abordată, și utilizarea acesteia ca fundație pentru dezvoltarea soluției de depozite de date.

- *Framework*

  Framework-ul definește limitările sistemului de depozite de date, diversele componente și interacțiunea dintre acestea. Propunerea noastră vizează, din acest punct de vedere, dezvoltarea unui framework menit să asigure implementarea automată a modelului depozitului de date, ca parte a contribuției noastre practice la domeniul cercetat. Obiectivul nostru este condus de cerințele de reducere a costurilor și timpului implicat în procesul de dezvoltare a acestor soluții complexe.

- *Implementare*

  Referitor la implementarea soluției de depozite de date, intenția noastră este definită de proiectarea și implementarea unui prototip pentru crearea unor structuri de date specifice într-un mod automatizat, bazat pe framework-ul propus. Implementarea efectivă este realizată pornind de la diferite tipuri de metadata tehnice în mediul SAP Business Warehouse.

ii. Organizarea Generală a Tezei

Având în vedere contextul amplu acoperit, am stabilit separarea tezei de doctorat intitulată "Proiectarea și Implementarea Depozitelor de Date pentru Business Intelligence Aplicate în Economie" în șase capitole principale (vezi Figura 1). O secțiune inițială este dedicată unor noțiuni introductorii, în vreme ce o secțiune finală prezintă concluziile, diseminarea rezultatelor și potențiale direcții de cercetare ulterioară în domeniu. *Introducerea* evidențiază motivația tezei și obiectivele cercetării determinate de aceasta. Descriem deasemenea principalele provocări întâlnite în domeniul dezvoltării depozitelor de date, provocări ce dictează posibilitățile de îmbunătățire și determină astfel obiectivele noastre, și introducem organizarea generală a tezei. Secțiunea de concluzii concentrează atenția pe reușitele obținute și determină obiective de cercetare ulterioare. Cinci capitole principale descriu aspecte fundamentale ale tehnologiilor Business Intelligence și depozitelor de date, modele de date și tehnici de modelare, arhitecturi și framework-uri întâlnite în domeniul depozitelor de date, etc.

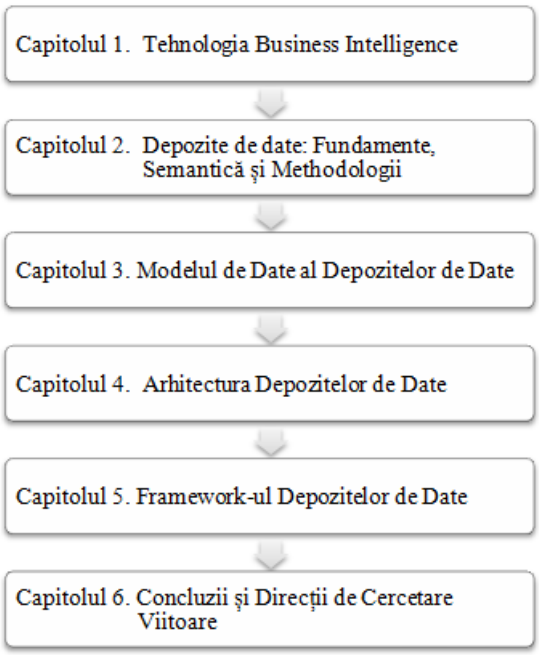

Figura 1. Organizarea Generală a Tezei

Capitolul 1. Tehnologia Business Intelligence

Primul capitol al tezei este dedicat introducerii contextului proceselor de luare a deciziilor în mediul de afaceri, precum și tehnologiilor care le sprijină. Obiectivul nostru principal constiuie plasarea tehnologiei depozitelor de date în cadrul framework-ului general de Business Intelligence. În acest scop ne propunem o scurtă prezentare a istoricului utilizării datelor în mediul sistemelor infomaționale, examinarea evoluției sistemelor inteligente de suport a deciziilor și discutarea beneficiilor acestora pentru procesele decizionale din cadrul intreprinderii. Deasemenea analizăm diferitele definiții, arhitecturi și cicluri de dezvoltare a sistemelor de Business Intelligence, relația dintre acestea și tehnologia depozitelor de date, precum și rolul depozitelor de date în mediul analitic.

Conceptul de "informație" este văzut în societatea informațională[1] din zilele noastre ca o componentă esențială pe care actorii mediului de afaceri (e.g organizații, intreprinderi, etc.) trebuie să o exploateze pentru a putea dobândi o întelegere mai profundă a proceselor desfășurate, cât și pentru a-și îmbunătăți procesul de luare a deciziilor și capacitatea de reacție la schimbări. Abundența datelor și informațiilor interne și externe intreprinderii pot fi exploatate eficient în beneficiul organizațiilor prin intermediul sistemelor inteligence, ca de exemplu a sistemelor suport de decizie și a celor de Business Intelligence.

Progresele înregistrate în domeniul tehnologiei informației au dus la extinderea sistemelor de procesare a datelor de la sisteme și aplicații autonome la sisteme complexe de tip Business Intelligence. Sistemele suport de decizie prezintă capabilități de colectare a datelor din diferite surse și de preparare a acestora pentru procese de analiză. Acestea facilitează accesul la date la nivelul intreprinderii, asigură capacități ridicate de procesare și permit analiza avantajelor și dezavantajelor diferitelor alternative, astfel încât managerii și analiștii pot lua decizii în condiții de informare pe baza unor date precise, oportune și de o calitate superioară.

Pe de altă parte, sistemele de Business Intelligence cuprind o cateorie largă de aplicații și unelte, de la cele destinate achiziției de date, transformării și stocării acestora, la unelte care oferă utilizatorilor finali capabilități de procesare analitică și interactivă. Rolul principal al acestor sisteme este reprezentat de oferirea unui framework integrat pentru îmbunătățirea procesului de luare a deciziilor, livrând utilizatorilor de business informații corecte la timpul potrivit. În timp ce Business Intelligence reprezintă o tehnologie cuprinzătoare, sistemele suport de decizie sunt mai reduse ca dimensiune, fiind definite în general de un program mai complex sau o aplicație de sine stătătoare. Deasemenea sistemele suport de decizie pot fi integrate în mediul de Business Intelligence, ca parte a framework-ului analitic extins. Din punct de vedere arhictectural, depozitele de date sunt definite ca elemente de stocare și ca fundația pe care tehnologia de Business Intelligence este construită. Astfel, deși

[1] O definiție exactă a termenului *societate informațională* nu a fost universal acceptată. Ne referim în acest caz la definiția dată de N. Moore: "o societate în care […] informația este folosită ca o resursă economică, este intensiv exploatată de publicul general în cadrul activităților desfășurate ca și consumatori; și pe baza ei se dezvoltă un sector informațional în mediul economic" [125].

complementare, depozitele de date și sistemele de Business Intelligence pot fi utilizate separat.

În general, dezvoltarea unui sistem de BI urmărește ciclul generic specific majorității sistemelor informaționale, incluzând activități de analiză, colectare a datelor, procesarea și stocarea informațiilor, analiză și diseminarea informațiilor rezultate către destinatarii vizați. Fazele ciclului de dezvoltare sunt menite să asigure livrarea unor informații calitative și precise utilizatorilor de busienss pe baza cerințelor exprimate, colectate la nivelul întregii intreprinderi.

Tehnologiile descrise în cadrul acestui capitol sunt menite să faciliteze procesarea datelor din surse multiple și transformarea acestora în informații inteligibile și valoroase care sprijină procesul de luare a deciziilor. Indiferent de gradul de complexitate prezentat, folosirea uneia dintre aceste tehnologii în concordanță cu nevoile informaționale ale intreprinderii duce la creșterea semnificativă a capacității de reacție la schimbările care au loc în mediul de afaceri.

Capitolul 2. Depozite de Date: Fundamente, Semantică și Metodologii

Al doilea capitol al tezei este dedicat unei prezentări extensive a fundamentelor, semanticii și metodologiilor tehnologiei depozitelor de date. Definim tehnologia depozitelor de date ca un amestec de diverse tehnologii din mediul informațional și analizăm conceptul de depozit de date și rolul acestuia ca și componentă de stocare în framework-ul general analitic. Ne propunem deasemenea prezentarea structurilor de stocare specifice depozitelor de date, anume depozitul de date și data mart-urile, principalele diferențe dintre acestea și rolul lor în cadrul tehnologiei reprezentate. Totodată tratăm extensiv aspectele semantice ale intregrării datelor, prezentând diferitele tipuri de metadate, gestiunea și importanța acestora în mediul depozitelor de date. În final, analizăm numeroasele metodologii de dezvoltare a soluțiilor software și discutăm potrivirea acestora în cazul depozitelor de date.

Tehnologia depozitelor de date a apărut la începutul anilor 1990 ca o consecință a progreselor înregistrate în domeniul tehnologiei informațiilor. Aceasta este destinată gestionării mediului informațional în cadrul căruia o serie de componente permit colectarea și integrarea datelor la nivelul întregii intreprinderi, astfel încat utilizatorii de business să se poată baza în procesul de luare a deciziilor pe date consolidate, structurate și calitative. Astfel, depozitele de date pot fi văzute ca o tehnologie cuprinzătoare utilizată pentru gestiunea mediului analitic al unei intreprinderi printr-o serie de componente care permit colectarea și integrarea datelor din surse diverse, cu scopul transformării lor în informații strategice pentru intreprindere. Datele colectate sunt consolidate și structurate sub forma unui model comun în cadrul acestui mediu, fiind preparate pentru a fi "consumate" de către utilizatorii de business. Un depozit de date este definit ca o componentă de stocare în cadrul tehnologiei, un depozit de date integrate, orientate pe subiecte, non-volatile și care variază în timp [79]. Deoarece datele sunt colectate și stocate pentru o perioadă îndelungată de timp, diverse unelte analitice și de minare de date pot fi folosite pentru a efectua calcule matematice și statistice cu scopul de a facilita întelegerea proceselor de business, de exemplu prin detectarea unor tendințe și modele economice. Mediul depozitelor de date se bazează pe structuri de stocare modelate cu ajutorul diferitelor tehnici de modelare (e.g. entitate-relație (ER), modelare multi-dimensională, etc.), care îmbunătățesc performanța analizelor complexe efectuate pe volume mari de date.

Pentru efectuarea acestor analize complexe este necesară o întelegere la nivel global a datelor stocate în structurile depozitelor de date de-a lungul întregului lor ciclu de viață, acest lucru fiind realizat prin intermediul metadatelor. Metadatele nu reprezintă doar "date despre date", ci au o conotație și un rol mai extinse, concentrând totalitatea informațiilor și cunoștințelor existente în cadrul întreprinderii. Metadatele captează caracteristici generale și specifice, oferă context și semnificație datelor brute și crează un nivel semantic pentru sistemele informaționale dintr-o intreprindere. Acest nivel semnatic asigură o interpretare și o înțelegere adevate a datelor de către toți actorii implicați în utilizarea și exploatarea lor. Sistemele informaționale ale intreprinderilor conțin diverse tipuri de metadate, de la cele de business și tehnice, statice și dinamice, la cele descriptive, structurale și administrative. Însă pentru a-și putea îndeplini funcția, metadatele trebuie sa fie gestionate într-un mod adecvat.

Gestiunea metadatelor are un rol esențial în asigurarea bunei funcționării a activităților intreprinderii, mai ales în cele patru domenii în care sunt utilizate cu precădere: proiectare, activități operaționale, gestiune și activități de guvernare. Metadatele ajută deasemenea la minimizarea eforturilor administrării depozitelor de date și la îmbunătățirea procesului de extragere a acestora din mediul operațional. În mediul analitic, metadatele colectate din diverse surse sunt stocate în structuri de stocare specifice, astfel facilitându-se un acces consistent și sigur la date, precum și operații de interogare și navigare din partea utilizatorilor finali.

Având în vedere niveul de complexitate ridicat al tehnologiei depozitelor de date, dezvoltarea unor astfel de soluții necesită o abordare structurată și planificată, definită sub forma unei metodologii. Metodologiile sunt menite să asigure un set de pași și principii repetitive, consistente și de încredere, pentru atingerea unor rezultate predictibile (e.g. un produs sau o soluție software). Metodologiile pot fi formale (i.e. caracterizate de o abordare structurată și un set bine definit de activități) sau informale (i.e. definite de practici de succes, diverse cursuri de specialitate, etc.); dirijate de date (i.e. bazate pe analiza modelului de date la nivelul corporației), dirijate de obiective (i.e bazate pe obiectivele companiei și pe analiza proceselor de business) sau dirijate de utilizatori (i.e. concentrate pe implementarea strategiilor de business), etc. Majoritatea metodologiilor generice de dezvoltare a soluțiilor software (e.g. modelul *waterfall*, modelul *incremental*, abordarea *spirală*, modelul *RAD*, etc.) sunt potrivite pentru un anumit nivel de complexitate al depozitelor de date. Două mari abordări sunt considerate de referință pentru dezvoltări ale acestor soluții de depozite de date, anume modelul Inmon și modelul Kimball. Modelul Inmon urmează o dezvoltare de tip spirală și recomandă construirea unei soluții de dimensiuni mari – depozitul de date la nivelul întregii intreprinderi, în timp ce modelul Kimball se concentrează pe o metodologie construită pe baza analizei cerințelor utilizatorilor finali, care are avantajul de a facilita livrarea de rezultate într-un mod rapid, conducând la un nivel ridicat de satisfacție din partea utilizatorilor de business. Cu toate acestea, alegerea unei metodologii potrivite depinde de nevoile intreprinderii și ale utilizatorilor finali. Urmarea anumitor modele în procesul de selecție al metodologiei duce la o creștere semnificativă a șanselor de succes în implementarea soluțiilor de depozite de date.

Capitolul 3. Modelul de Date al Depozitelor de Date

În al treilea capitol al tezei tratăm concepte și aspecte esențiale ale modelelor de date și ale tehnicilor de modelare în mediul depozitelor de date. Ne propunem astfel introducerea diferitelor tipuri de modele de date, a caracteristicilor fundamentale ale acestora, precum și două tehnici de modelare de referință folosite pentru dezvoltarea schemei depozitelor de date (i.e. tehnica entitate-relație (ER) și tehnica multi-dimensională (MD)). Deasemenea argumentăm importanța modelării de tip multi-dimensional pentru construirea strucurilor de date specifice mediului analitic, și trecem în revistă eforturile de cercetare în acest domeniu. Obiectivul principal în cadrul acestui capitol îl constituie determinarea unei metodologii potrivite pentru derivarea modelelor multi-dimensionale pornind de la schema entitate-relație a sistemelor operaționale, care reprezintă principala sursă de date pentru mediul analitic. În a doua parte a acestui capitol aplicăm în cadrul unui studiu de caz o metodologie de derivare considerată adecvată pentru un model de date din domeniul reasigurărilor, pe care o extindem cu opțiuni de modelare specifice mediului de afaceri. Implementarea modelului rezultat este prezentată în ultimul capitol al tezei, ca parte a contribuției noastre practice la domeniul cercetat.

Ca mediul de stocare a soluției generale a depozitelor de date, depozitul propriu-zis este definit de reprezentări specifice ale datelor și a relațiilor dintre acestea. Aceste reprezentări, cunoscute sub numele de modele de date, sunt menite să asigure o documentare completă a mediului informațional în ceea ce privește procesele existente, entitățile, relațiile, și fluxurile de date, etc. Modelele de date reprezintă rezultatul tehnicilor de modelare care definesc și analizează cerințele exprimate de utilizatorii finali, în scopul sprijinirii proceselor de business ale intreprinderii.

În mediul depozitelor de date sunt recunoscute diferite tipuri de modele de date și tehnici de modelare care stau la baza structurilor specifice și determină modul în care datele sunt stocate. Două modele în particular, anume modelul entitate-relație și modelul multi-dimensional, sunt folosite în cadrul celor cele două nivele de stocare diferite ale sistemelor de depozite de date. Soluția propusă în cadrul acestei teze utilizează ambele modele pentru a defini schema de date a modelul informațional la nivelul intreprinderii. În timp ce modelul ER se bazează pe o tehnică standardizată aplică predominant în mediul operațional, modelarea MD este specifică sistemelor analitice și nu este definită de o abordare acceptată ca standard în lumea modelării datelor. Cu toate acestea, literatura abundă în propuneri privind metode multi-dimensionale și derivarea modelelor multi-dimensionale din diverse surse, cum ar fi cerințele utilizatorilor, procese de business sau modele entitate-relație existente în sistemele sursă. Ne propunem așadar analiza câtorva dintre aceste abordări, discutarea avantajelor și dezavantajelor acestora, precum și formularea unor păreri personale cu privire la compatibilitatea acestor abordărilor în dezvoltarea depozitelor de date. Scopul nostru principal constituie prezentarea unui studiu de caz care prin care să se realizeze derivarea unui model multi-dimensional reprezentând un process din domeniul reasigurărilor, prin aplicarea unei metodologii considerate adecvate. Deoarece considerăm potrivită

proiectarea unei soluții analitice capabile să intregreze cu ușurință datele în sistemul depozitelor de date și fiindcă sistemele operaționale reprezintă sursa principală de date pentru mediul analitic, ne concentrăm eforturile asupra transformării modelului ER a acestora în model multi-dimensional. În final, utilizăm modelul rezultat pentru implementarea soluției de depozite de date printr-un prototip creat pe baza unui framework de automatizare pe care îl definim și descriem în ultimul capitol al tezei. Ca activități preliminare definim atât modelul entitate-relație cât și modelul multi-dimensional și tehnicile de modelare, și prezentăm pe larg toate conceptele de bază corespunzătoare.

Atât în mediul operațional cât și în cel analitic, datele sunt reprezentate prin intermediul unor diagrame, folosind texte și simboluri menite să faciliteze cititorilor întelegerea lor. Aceste diagrame, cunoscute sub numele de modele de date, sunt obținute prin diverse procese de inginerie software sau tehnici de modelare. Tehnicile de modelare sunt destinate definirii și analizei cerințelor exprimate de utilizatorii de business în scopul producerii unor modele de date definite la diferite nivele de abstractizare (e.g. conceptual, logic și fizic), capabile să sprijine procesele de business ale intreprinderii. Modele rezultate au un rol esențial în descrierea datelor și a caracteristicilor acestora, precum și în aplicarea regulilor de business. Acestea sunt utilizate pentru facilitarea utilizării datelor ca o resursă pentru intreprindere, pentru integrarea informațiilor din cadrul intreprinderii, pentru definirea unui model arhitectural comun pentru întregul mediu informațional și pentru proiectarea structurilor de stocare, anume a bazelor și depozitelor de date.

Tehnica entitate-relație prezintă caracteristici adecvate modelării datelor tranzacționale (e.g. modele de date normalizate cu un nivel de redundanță, dependență și inconsistență redus; un număr mare de entități; date calitative, etc.), potrivite pentru operații de inserare, actualizare și ștergere, și este așadar utilizată predominant în sistemele operaționale. Spre deosebire de aceasta, modelarea multi-dimensională este definită de caracteristici precum număr redus de entități, prezentare intuitivă a datelor, modele optimizate pentru analiză și interogare, etc., care ii determină compatibilitatea pentru mediul analitic. Cu toate acestea, ambele tehnici pot fi folosite pentru modelarea datelor în mediul analitic: tehnica ER este în mod obișnuit folosită pentru definirea modelelor de date ale depozitului de date central, în timp ce modelarea MD este folosită pentru definirea structurilor de tip data mart în cadrul nivelului de prezentare a datelor.

Modelele multi-dimensionale sunt văzute de numeroși autori ca fiind forme restricționate ale modelelor ER, ceea ce determină o mapare aproape directă între ele. Așadar, modelele multi-dimensionale din mediul depozitelor de date pot fi derivate în mod direct din schemele entitate-relație ale sistemelor sursă operaționale. Aceasta abordare este sprijinită de diferite metodologii dirijate de date și adecvate pentru dezvoltarea soluției propuse de depozit de date, care cuprinde un nivel de date intregrate (depozitul de date central) și mai multe structuri data mart aprovizionate cu date din acest nivel. Considerând caracterul complex al acestor modele, literatura prezintă numeroase abordări definite pentru producerea modelelor multi-dimensionale ca reprezentări abstracte ale datelor din intreprindere pentru mediul analitic. Majoritatea acestora însă, datorită diverșilor factori precum complexitate ridicată, folosirea unor sisteme de notații noi, reprezentări grafice diferite, etc., nu a fost aplicată în

practică în afara domeniului de cercetare din mediul academic. Datorită acestui fapt metodologiile de modelare multi-dimensională utilizate în industrie sunt în general bazate pe abordări informale și practici de succes.

Pentru a ne atinge obiectivele privind construirea unei soluții de depozite de date comprehensive, urmăm o abordare bine definită de tip top-down propusă de W.H. Inmon [81]. În cadrul acestui capitol descriem principiile oferite de această abordare pentru construirea nivelului de depozit de date central, precum și o metodologie propusă de Moody and Kortink [121] pe care o extindem și o aplicăm pentru dezvoltarea modelului multi-dimensional. Exemplificăm această metodologie prin studiul de caz al modelării unui process de business din domeniul reasigurărilor și contribuim la procesul de dezvoltare a modelului cu îmbunătățiri specifice domeniului (e.g. analiza utilității modelului și integrarea dimensiunilor potrivite în modelul multi-dimensional, reprezentarea datelor financiare prin intermediul diferitelor monede, manipularea unor modificări în cadrul dimensiunilor, etc.). Deasemenea evaluăm modelul rezultat pe baza unor caracteristici pe care modelele de date trebuie să le posede pentru a putea sprijini utilizări avansate în analiza datelor, și concluzionăm că acesta este conform cu majoritatea cerințelor, determinând astfel o reprezentare validă a datelor intreprinderii.

Capitolul 4. Architectura Depozitelor de Date

Al patrulea capitol al tezei este dedicat prezentării unei serii de aspecte arhitecturale și a principalelor abordări de dezvoltare din domeniul depozitelor de date. Ne propunem descrierea caracteristicilor fundamentale ale celor mai comune tipuri de arhitectură folosite în mediul depozitelor de date, precum și a două ahitecturi și framework-uri de implementare de referință (i.e. modelul Inmon și modelul Kimball). Prezentăm deasemenea o serie de factori care influențează procesul de selecție a arhitecturii potrivite, discutăm framework-ul adecvat fiecărei combinații dintre acești factori și introducem o serie de elemente care determină implementarea cu succes a soluțiilor de depozite de date.

Depozitele de date sunt definite în cadrul tehnologiei comprehensive care acoperă mediul informațional al intreprinderii, ca depozite de date colectate, integrate și consolidate din diverse surse de date eterogene. Complexitatea gestiunii, transformării și integrării acestor date, atât din cadrul intreprinderii cât și din afara acesteia, determină ca dezvoltarea unor astfel de soluții să fie considerată o provocare. Literatura și numeroase practici de succes prezintă principii bine definite care să ghideze proiectarea și implementarea soluțiilor de depozite de date, pe care intreprinderile le pot adapta nevoilor lor specifice. Procesul de dezvoltare a acestor soluții presupune selecția unui framework arhitectural și a unei metodologii compatibile care să asigure succesul acestui demers. Selecția lor este bazată pe diverși factori care includ infrastructura informațională, mediul de afaceri, capabilitatea mediului tehnic, implicarea părților interesate, precum și resursele financiare ale intreprinderii, etc. [13]. Arhitectura de implementare a soluțiilor de depozite de date este semnificativ diferită și mai complexă decât arhitectura clasică a bazelor de date. Aceasta este menită să asigure o fundație solidă pentru integrarea și consolidarea datelor de la nivelul întregii intreprinderi și un framework general pentru dezvoltarea și utilizarea eficientă a tuturor componentelor grupate în trei categorii principale: achiziția de date, depozitul de stocare și livrarea de informații.

Cele mai comune tipuri arhitecturale în domeniul depozitelor de date prezentate în literatură includ data mart-urile indepedente, arhitectura de tip “autobuz” a data mart-urilor, depozitul de date la nivel intreprinderii, arhitectura centralizată și arhitectura de tip federație. Data mart-urile independente sunt în general implementate în intreprinderi mici, fiind caracterizate de vederi de date departamentale autonome, de cele mai multe ori extrase din sistemele sursă operaționale. Deși sunt mai eficiente din punctul de vedere al resurselor utilizate, acestea duc la creșterea volumelor de date și la redundanța proceselor, având deasemenea o scalabilitate redusă, o limitare a integrării datelor și o deficiență majoră în prezentarea unei vederi integrate asupra datelor din intreprindere. Arhitectura de tip “autobuz” data mart-urilor diminuează o parte din dezavantajele menționate, oferind un framework comprehensiv pentru integrarea vederilor departamentale pe baza unei structuri arhitecturale de tip “autobuz”. Data mart-urile individuale sunt dezvoltate utilizând dimensiuni conformate pornind de la cerințele utilizatorilor și de la procesele de business, în timp ce structurile de stocare modelate cu tehnica multi-dimensională permit stocarea atât a datelor atomice cât și a celor sumarizate.

Depozitul de date la nivelul intreprinderii reprezintă cel mai complet și complex tip arhitectural din mediul depozitelor de date. Scopul său principal în constituie oferirea unei fundații de date integrate, definite la nivel atomic și stocate în structuri normalizate, capabile să permită definirea mai multor modele multi-dimensionale de date agregate. Arhitectura centralizată prezintă caracteristici similare cu depozitul de date la nivelul intreprinderii, fără a include însă nivelul superior de vederi departamentale. Arhitectura de tip federație constă dintr-un set de depozite de date organizate separat și dispersate geografic, care operează într-un mod semi-autonom. Aceasta este specifică organizațiilor mari formate prin achiziționarea și unificarea altor unități având propriile soluții de Business Intelligence care nu au fost înlăturate, ci folosite într-o manieră integrată.

Literatura de specialitate prezintă deasemenea diferite abordări privind implementarea acestor tipuri arhitecturale. Două dintre ele, anume abordarea de tip top-down (i.e. realizată de modelul Inmon) și abordarea de tip bottom-up (i.e. realizată de modelul Kimball) se evidențiază ca implementări arhitecturale și metodologice de referință. Ideea principală a modelului Inmon este dezvoltarea unei arhitecturi comprehensive de depozite de date, astfel: un depozit central (i.e. depozitul de date propriu-zis), care stochează date integrate și consolidate de la nivelul intreprinderii, și o serie de structuri de tip data mart, care oferă o vedere multi-dimensională a datelor în scopul facilitării proceselor de analiză și interogare. Construirea unei astfel de arhitecturi presupune realizarea unor activități de planificare și proiectare la începutul proiectului, menite să rezolve potențialele nelămuriri privind integrarea datelor, aspecte de securitate, calitate și standarde, precum și modelul de date general. Acestă arhitectură facilitează o definire uniformă a datelor și impunerea regulilor de business la nivelul întregii intreprinderi. Vederile departamentale dependente sunt dezvoltate ulterior cu ajutorul tehnicii multi-dimensionale și sunt încărcate cu date din depozitul central. În cazul abordării bottom-up, propusă de modelul Kimball, implementarea depozitului de date este bazată pe crearea de vederi multi-dimensionale ale datelor din intreprindere și integrarea lor pe baza unei structuri de tip "autobuz" (i.e. dimensiuni conformate) pentru a obține o vedere de ansamblu la nivelul întregii intreprinderi. Data mart-urile sunt create pe baza cerințelor specifice fiecărui departament. Deși oferă un grad de integrare mai redus, această abordare este mult mai des folosită pentru implementarea depozitelor de date deoarece necesită eforturi mai reduse din partea intreprinderii și asigură livrarea unor rezultate imediate. Un al treilea tip de abordare, anume cea hibridă, încearcă combinarea avantajelor celor două abordări de referință prin determinarea gradului de planificare și proiectarea necesar sprijinirii integrării datelor la nivel intreprinderii (i.e. modelul top-down), și construirea structurilor de tip data mart prin modelul bottom-up.

Având în vedere opțiunile de implementare prezentate, selecția unei abordări potrivite nu este o sarcină ușoară. Alegerea unui framework arhitectural care să safisfacă nevoile intreprinderii este influențată de diverși factori, cum ar fi inter-dependența informațională dintre departamentele intreprinderii, urgența finalizării proiectului, caracterul de rutină al sarcinilor, viziunea strategică a depozitului de date, cantitatea de resurse disponibilă și alocată pentru dezvoltarea soluției, etc. Combinația acestor factori favorizează selecția unui anumit tip arhitectural. Simpla selecție a acestuia nu garantează însă succesul implementării și al utilizării în producție. O serie de elemente care se referă la aspecte organizaționale, de mediu,

legate de proiect, tehnice și educaționale, determină gradul de acceptare a soluției de depozite de date de către utilizatorii finali, precum și capacitatea acesteia de la sprijini în mod efectiv procesul de luare a deciziilor.

Capitolul 5. Framework-ul Depozitelor de Date

Al cincilea și ultimul capitol al tezei este dedicat descrierii contribuției practice, anume propunerea unui framework și al unui prototip destinat automatizării procesului de implementare a schemei depozitului de date în mediul analitic. Pentru a ne justifica propunerea, introducem pe scurt o serie dintre cele mai utilizate framework-uri în dezvoltarea soluțiilor analitice și discutăm utilitatea automatizării în implementarea depozitelor de date. Ne structurăm contribuția în două parți principale: în prima parte prezentăm arhitectura framework-ului propus, împreună cu componentele sale, o descriere detaliată, precum și interacțiunea dintre acestea; în a doua parte descriem particularități de proiectare și implementare ale prototipului în mediul SAP Business Warehouse. Deasemenea evaluăm utilitatea framework-ului și prototipului propuse, precum și importanța acestora în facilitarea dezvoltării depozitului de date la nivelul intreprinderii prin care se dorește reducerea costurilor în crearea sistemelor inteligente pentru sprijinirea proceselor de luare a deciziilor.

Având în vedere caracteristicile soluțiilor de depozite de date, discutate în cadrul capitolelor anterioare, alocarea unor resurse substanțiale și variate din partea intreprinderilor, cât și un angajament susținut din partea părților interesate, este esențială pentru procesul de dezvoltare al soluțiilor analitice, considerat elaborat și costisitor. Activitățile de dezvoltare presupun și o examinare a mediului informațional al intreprinderii, astfel încât existența unei metodologii comprehensive care să ghideze proiectarea și implementarea depozitelor de date pe baza unei arhitecturi solide este aproape obligatorie. Costurile, reflectate în mare parte ca cheltuieli de forma resurselor financiare și a timpului de livrare din punct de vedere al gestiunii proiectelor, și mai ales reducerea lor, reprezintă o preocupare fundamentală pentru toate intreprinderile. Necesitatea diminuării costurilor a dus la realizarea unei automatizări parțiale sau complete a unor procese în proiectarea și utilizarea depozitelor de date, ca de exemplu în modelarea conceptuală și logică a schemei acestora, extragerea, transformarea și încărcarea datelor, etc. Această automatizarea nu acoperă însă toate fazele din dezvoltarea proiectelor, mai ales din cauza influenței aspectelor de afaceri în mediul analitic. Cu toate acestea, considerăm că procesul de automatizare poate fi extins, cu anumite limitări, la faza de implementare și creare a structurilor de date și a proceselor corespunzătoare de extragere, transformare și încărcare a datelor. Pentru realizarea acestei propuneri este necesară definirea unui framework arhitectural complex, pe care îl prezentăm în cadrul acestui capitol.

Conform [55], un framework este definit pentru a asigura existența unei filosofii și al unei îndrumări care să descrie aspectul, modul de funcționare și interoperabilitatea aplicațiilor software. Ne concentrăm așadar pe definirea unui framework pentru realizarea automatizată a proceselor de creare a structurilor de date pe baza metadatelor tehnice și începem cu prezentarea problemei pentru propunerea noastră. Deasemenea introducem principalele caracteristici ale framework-urilor pentru dezvoltarea depozitelor de date, enumerăm unele dintre cele mai utilizate framework-uri în domeniu și analizăm necesitatea automatizării în mediul analitic. Detalii specifice de implementare pentru prototipul propus, evaluarea utilității automatizării, precum și importanța contribuției noastre, sunt introduse în a doua parte a capitolului.

Construirea unei soluții analitice de depozite de date la nivelul intreprinderii reprezintă o activitate complexă care necesită utilizarea unui framework solid și a unor activități efective din domeniul planificării proiectelor. Literatura recunoaște numeroase framework-uri arhitecturale și metodologice utilizate în mediul depozitelor de date, fiecare descriind diferite structuri și procese, precum și secvențe de pași urmați pentru dezvoltarea acestor soluții comprehensive. Procese de proiectare și implementare de succes sunt sprijinite de arhitecturi și metodologii compatibile. În încercarea noastră de dezvoltare a unei soluții de depozite de date la nivelul intreprinderii, aderăm arhitectura și metodologia consistentă propusă de Inmon. Astfel, din punct de vedere arhitectural urmăm o proiectare de tip top-down pentru depozitul de date, realizând un depozit central de date integrate și consolidate (i.e. nivelul de data warehouse) și mai multe structuri de tip data mart încărcate cu date din acest depozit. Considerând numeroasele structuri și procese care definesc o soluție de depozite de date, propunem un framework pentru o implementare automatizată și un prototip corespunzător, bazate pe presupunerea că activități de dezvoltare repetitive și consumatoare de timp pot fi realizate în mod eficient și într-o perioadă mai scurtă de timp. Prototipul de implementare asigură crearea automată a structurilor de stocare specifice pentru nivelul de depozit de date central și nivelul de data mart-uri, precum și pentru procesele de extragere, transformare și încărcare corespunzătoare din metadate tehnice. Printre beneficiile automatizării, menite să justifice propunerea noastră, menționăm: crearea de componente software care se conformează cu o sintaxă și constrângeri bine definite, reducându-se astfel factorul de eroare umană; standardizarea componentelor software, care duce la îmbunătățirea lizibilității codului; reducerea forței de muncă, a costurilor și a timpului de dezvoltare aferente, etc.

Implementarea prototipului este realizată în mediul SAP Business Warehouse. Am ales această plarformă tehnologică deoarece SAP BW oferă o fundație comprehensivă pentru unelte de Business Intelligence prin componentele sale arhitecturale. Deasemenea, sprijină procesele de achiziție și de curățare și pregătire a datelor, menite să asigure o calitate superioară și integrarea la nivelul întregii intreprinderi, permite definirea unui nivel central al depozitului de date, care să stocheze date granulare, integrate, rezultate din procesul de curățare și pregătire; și sprijină crearea de vederi multi-dimensionale (i.e. data mart-uri) prin intermediul unei scheme extinse. Proiectare prototipului este realizată pentru fiecare nivel de stocare al arhitecturii, astfel încât să beneficiem de aceste avantaje și capabilități oferite de SAP BW (i.e. *Achiziția Datelor* (procesele de achiziție, curățire și pregătire a datelor), *Gestiunea Datelor Primare* (depozitul de date central), și *Livrarea Datelor* (data mart-urile)).

Prototipul propus permite implementarea automatizată a structurilor inițiale aferente depozitului de date central și data mart-urilor, precum și proceselor de extragere, transformare și încărcare corespunzătoare, pornind de la metadate tehnice. Prin componentele sale definite pentru fiecare nivel arhitectural de stocare, procesul de implementare acoperă: procesele de achiziție, curățire și pregătire a datelor, precum replicarea surselor de date, generarea și executarea pachetelor de extragere, generarea unităților informaționale de modelare, etc.; generarea structurilor depozitului de date central pentru asigurarea unei stocări permanente a datelor granulare; generarea schemei data mart-urilor; și generarea transformărilor tehnice și a regulilor de mapare dintre diferitele structuri și obiectele de metadate.

- Schema inițială rezultată poate fi extinsă ulterior prin interfața pentru utilizator oferită de SAP BW, aceste îmbunătățiri incluzând remodelarea unor structuri de date, realizarea de extrageri selective, transformări bazate pe logica de business, etc.). În final am demonstrat că prototipul propus pentru automatizarea acestor procese este benefic pentru reducerea costurilor în cazul dezvoltării unor soluții complexe, în cadrul cărora sunt create un număr mare de structuri de stocare și procese corespunzătoare, prin activități repetitive și consumatoare de timp.

Capitolul 6. Concluzii și Direcții de Cercetare Viitoare

Principalele obiective de cercetare tratate în cadrul acestei teze au fost determinate de dezvoltarea unei soluții comprehensive de depozite de date pornind de la cerințele de reducere a costurilor și de eficientizare a procesului de implementare. Îndeplinirea acestora a presupus o înțelegere temeinică a diferitelor aspecte legate de depozitele de date, anume: pozitionarea și rolul tehnologiei depozitelor de date în cadrul framework-ului de Business Intelligence; definirea modelelor de date specifice, care determină structurile de stocare a datelor în mediul analitic; metodologia de dezvoltare a soluției analitice, care ghidează procesele de proiectare și implementare efective; arhitectura care definește fundația dezvoltării soluției de depozite de date; framework-ul prin care este descris un set de pași pentru construirea componentelor și definirea interacțiunii dintre acestea; precum și desfășurarea procesului efectiv de implementare.

Cantitatea abundentă de informații existentă în mediul economic poate fi exploatată eficient prin intermediul unor aplicații și unelte specifice (e.g. sisteme support de decizie, tehnologia Business Intelligence, etc.). Acestea sunt esențiale în facilitarea accesului la datele interne și externe intreprinderii, asigurând capabilități avansate de procesare și analiză a acestora. Astfel, am început introducerea în mediul sistemelor analitice prin prezentarea unor caracteristici generale, analiza evoluției istorice a sistemelor suport de decizie și discutarea similarităților și diferențelor dintre primele faze ale acestor sisteme și tehnologiile comprehensive din zilele noastre.

Am prezentat depozitele de date ca fiind o tehnologie cuprinzătoare utilizată pentru manipularea mediului analitic al intreprinderilor, și depozitul de date propriu-zis ca o componentă de stocare a datelor și fundația pe care tehnologia de Business Intelligence este construită. Așadar, am argumentat existența unei diferențieri între conceptele de tehnologie a depozitelor de date și depozitul de date propriu-zis, și am examinat diversele perspective prezentate în literatură. În timp ce tehnologia depozitelor de date cuprinde o serie de componente și procese menite să permită colecționarea și integrarea datelor din diferite surse, cu scopul principal de a le transforma în informații strategice pentru intreprindere, depozitul de date definește componenta de stocare a tehnologiei, depozitul de date integrate, orientate pe subiecte, non-volatile și care variază în timp. Aceste volume mari de date integrate fac subiectul a numeroase calcule matematice și statistice, menite să producă o perspectivă de afaceri valoroasă. Din acest motiv depozitele de date se bazează pe structuri optimizate pentru interogări și analize performante, proiectate cu ajutorul unor tehnici de modelare specifice. În cadrul tezei am introdus particularitățile diverselor structuri din mediul depozitelor de date (e.g. depozitul de date central, data mart-uri) și le-am examinat diferențele. Aceste structuri precum și datele depozitului de date sunt descrise prin intermediul metadatelor. În scopul facilitării întelegerii rolului lor fundamental, am prezentat un studiu extins a numeroaselor definiții, clasificări și caracteristici de gestiune a metadatelor. Ne-am concentrat în principal pe metadatele de tip business și tehnice, insistând asupra caracterului descriptiv al metadatelor de business, esențial în întelegerea semanticii

proceselor de business, și asupra importanței metadatelor tehnice în favorizarea automatizării în mediul depozitelor date.

Având în vedere complexitatea soluțiilor de depozite de date, am prezentat o serie de metodologii de dezvoltare menite să ofere o abordare structurată și planificată numeroaselor proceselor desfășurate. Dintre aceste metodologii, atât generice cât și specifice domeniului depozitelor de date, am descris în detaliu două abordări de referință, anume modelul Inmon (i.e. de tip top-down) și modelul Kimball (i.e. de tip bottom-up). Deasemenea, am introdus două framework-uri definite pentru selectarea unei metodologii adecvate unei implementări de succes a depozitelor de date și am enunțat motivele adoptăriia metodologiei propusă de Inmon, de tip spirală și dirijată de date pentru dezvoltarea soluției propuse.

Structurile de date optimizate pentru mediul analitic sunt proiectate cu ajutorul unor tehnici de modelare specifice, menite să asigure determinarea și analiza cerințelor exprimate de către utilizatorii de business și să producă modele de date capabile să sprijine procesele de business ale intreprinderii. Considerând abordarea noastră în cazul dezvoltării depozitelor de date, definită de existența unui nivel de date granulare, consolidate și integrate și a unui nivel de structuri multi-dimensionale construite pentru optimizarea performațelor de analiză și interogare, am descris două dintre cele mai utilizate tehnici de modelare a datelor, anume tehnica entitate-relație și tehnica multi-dimensională. Tehnica ER prezintă caracteristici adecvate modelării structurilor normalizate, caracterizate de redundanță, dependență și inconsistență minime, capabile să stocheze date calitative la un nivel de historizare și granularitate ridicate, în timp ce tehnica multi-dimensională, specifică mediului analitic, produce prezentări intuitive ale date sub forma unor modele optimizate pentru interogare și analiză.

Modelele multi-dimensionale sunt definite ca forme restrânse ale modelelor de tip entitate-relație, obținute prin diverse metodologii nestandardizate (i.e. nu există o metodologie acceptată universal ca standard în modelarea multi-dimensională). Astfel, am prezentat un studiu al celor mai citate lucrări în domeniu, am discutat atât avantajele cât și dezavantajele acestora, și ne-am exprimat opiniile personale privind potrivirea lor pentru dezvoltarea soluții de depozite de date. Am selectat abordarea propusă de Moody and Kortink [121] pentru derivarea modelului multi-dimensional, justificându-ne decizia astfel: 1) obiectivul nostru de proiectare a unei soluții comprehensive de depozite de date, care să include un nivel de integrare (i.e. depozitul de date central) și un nivel de prezentare a datelor (i.e. data mart-urile) este suportată de această metodologie; 2) abordarea este bazată pe modelul de date al intreprinderii în care relații dintre date sunt descrise, simplificând astfel procesul de extragere, transformare și încărcare în mediul analitic; și 3) metodologia a fost validată în practică și deasemenea permite arhitectului soluției să rafineze pașii de dezvoltare pe baza cerințelor utilizatorilor sau a cunoștințelor de business. Am exemplificat metodologia printr-un studiu de caz reprezentând un process de business din domeniul reasigurărilor, prin care am descris derivarea modelului de date pornind de la schema entitate-relație a sistemului operațional. Deasemenea, am contribuit la rafinarea acestuia prin includerea unor aspecte specifice reasigurărilor, anume includerea reprezentării datelor financiare prin intermediul diferitelor monede, analiza utilității modelului de date și integrarea dimensiunilor adecvate în

cadrul acestuia, gestionarea modificărilor în dimensiunile modelului, etc. În final am evaluat modelul de date rezultat pe baza unor caracteristici considerate esențiale în sprijinirea proceselor de analiză avansată și am concluzionat că modelul propus se conformează majorității cerințelor, determinând astfel o reprezentare validă a datelor.

Un alt aspect important tratat în cadrul tezei a fost selecționarea unei arhitecturi adecvate pentru depozitul de date, cu atât mai mult cu cât aceasta este semnificativ mai diferită și mai complexă decât arhitectura clasică a bazelor de date. Am prezentat diferitele tipuri de arhitecturi recunoscute în literatură (e.g. data mart-urile independente, arhitectura de tip "autobuz" a data mart-urilor, depozitul de date la nivel intreprinderii, arhitectura centralizată și arhitectura de tip federație) și am detaliat două implementări arhitecturale de referință: implementarea de tip top-down (i.e. realizată de modelul Inmon) și cea de tip bottom-up (i.e. realizată de modelul Kimball). Deasemenea am prezentat un framework definit pentru facilitatea selecției tipului arhitectural adecvat, determinată de factori organizaționali, de mediu, referitori la proiect, tehnici și educaționali, etc., și am selectat astfel cel mai potrivit tip de arhitectură pentru dezvoltarea soluției de depozite de date propuse.

În ceea ce privește framework-ul de dezvoltare propus, am pornit de la ideea că un proces de proiectare și implementare de succes este întotdeauna sprijinit de o arhitectură și o metodologie compatibile. Considerând numeroasele structuri și procese definesc soluția de depozite de date, am definit un framework și un prototip corespunzător pentru implementarea automatizată a acestora, bazându-ne pe faptul că o serie de activități repetitive și consumatoare de timp pot fi realizate într-un mod mai eficient și într-o perioadă mai scurtă de timp. Ne-am justificat propunerea prin prezentarea numeroaselor beneficii ale automatizării în dezvoltarea soluțiilor software în general, precum și a celor de depozite de date în particular. Framework-ul propus este compus din cinci componente de bază, definite pentru gestiunea datelor și a metadatelor, procesul de curățire și pregătire a datelor, și procesele de generare a structurilor de date specifice depozitului de date central și data mart-urilor. Am descris amănunțit rolul și caracteristicile fiecărei componente, precum și interacțiunea dintre ele. Am proiectat prototipul corespunzător pentru crearea automatizată a structurilor de stocare în formă inițială pentru nivele de depozit de date central și data mart-uri, și a proceselor de extragere, transformare și încărcare a datelor, în mediul SAP Business Warehouse pornind de la metadate tehnice. Alegerea platformei tehnologice a fost determinată de capabilitățile avansate pe care SAP BW le oferă, anume o fundație comprehensivă pentru activitățile complexe de Business Intelligence. Am construit componentele prototipului pentru fiecare nivel arhitectural de stocare a datelor, astfel încât să putem exploata la maximum aceste capabilități (i.e. componenta de curățare și pregătire a datelor pentru nivelul *Achiziția Datelor*, componenta de depozit de date central pentru nivelul *Gestiunea Datelor Primare*, și componenta de data mart-uri pentru nivelul *Livrarea Datelor*).

Am realizat implementarea automatizată a schemei inițiale a depozitului de date prin utilizarea prototipului, după cum urmează: structurile normalizate ale depozitului de date, obținute ca o mapare 1:1 a surselor de date implementate pe sistemele sursă, și schema stea relațională a modelului de date pentru nivelul de data mart-uri, obținută prin generarea structurilor specifice pentru modelul derivat și prezentat ca studiu de caz. Deasemenea, am

definit limitările automatizării obținute prin intermediul prototipului și am subliniat faptul că schemele inițiale rezultate pot fi îmbunătățite ulterior prin interfața utilizator oferită de SAP BW (aceste dezvoltări ulterioare pot include remodelarea structurilor de date, extragerea selectivă din sistemele sursă, transformări pe baza logicii de business, crearea unor rutine pentru curățarea și integrarea datelor, etc.). Totodată am demonstrat că prototipul propus este benefic pentru reducerea costurilor în cazul depozitelor de date comprehensive dezvoltate la nivelul intreprinderii, în cazul creării unui număr mare de structuri de date și procese de extragere, transformare și încărcare corespunzătoare implementate prin intermediul activităților repetitive și consumatoare de timp.

Rezultatele prezentate în cadrul tezei au fost diseminate printr-o serie de articole prezentate la conferințe naționale și internaționale, și publicate în volume ale conferințelor și jurnale de specialitate de diferite categorii. Am obținut validarea prototipului de automatizare a creării structurilor de date și proceselor corespondente din mediul depozitelor de date printr-un articol prezentat la cea de-a doua Conferință Mondială pentru Inovare și Știința Calculatoarelor inclus în volumul Procedia Technology Journal al editurii Elsevier Publishing Ltd., indexat în ScienceDirect, Scopus și Thomson Reuters Conference Proceedings Citation Index (Web of Science) [132]. Deasemenea, am validat propunerea de framework pentru implementarea depozitelor de date la nivelul intreprinderii printr-un articol indexat BDI publicat în Database Systems Journal, editura ASE București [131]. Am discutat aspecte fundamentale prezentate în cadrul tezei (e.g. importanța automatizării în mediul depozitelor de date [133], aspecte privind modelarea metadatelor în cadrul depozitelor de date [137], o comparație a metodologiilor utilizate pentru construirea structurilor specifice depozitelor de date [135], probleme de securitate în mediul SAP BW [134] [136], etc.) într-o serie de alte articole, precum urmează:

- **I. M. Nagy**, *Automation prototype for the development of data warehousing data structures*, accepted for publishing in Procedia Technology Journal, Elsevier Publishing Ltd., ISSN: 2212-0173 (indexat ISI)
- **I. M. Nagy** și E. Tolea, *A Metamodel for Manipulating Business Knowledge Within a Data Warehouse*, Proceedings of the 6th International Conference On Virtual Learning, Editura Universității din București, ISSN: 1844-8933, pp. 255-261 (indexat ISI Proceedings)
- **I. M. Nagy**, *A Framework for Semi-Automated Implementation of Multidimensional Data Models*, Database Systems Journal, Volumul 3, Ediția 2, Editura ASE Bucuresti, 2012, ISSN: 2069-3230 (indexat BDI)
- **I. M. Nagy** și C. Stefanache, *Ensuring Data Protection in the SAP Business Information Warehouse: A Case Study*, Journal of Applied Computer Science & Mathematics, Volumul 9, Ediția 4, 2010, ISSN:1843-1046, pp. 83 – 87 (indexat BDI)
- **I. M. Nagy** și L. Feischmidt, Mobilizing Business Processes Security issues and advantages of using SAP Mobile Infrastructure in the development of mobile application, Economy Informatics, Volumul 10, Ediția 1, 2010, pp. 44 – 52 (indexat BDI)

- **I. M. Nagy**, *The Importance of Automation in the Data Warehousing Environment – A Case Study*, 19th International Economic Conference "The Persistence of the Global Economic Crisis: Causes, Implications, Solutions", Sibiu, 2012, pg. 201 - 208, ISBN 978-606-12-0323-9
- **I. M. Nagy** și A. Onaciu, *Two Methodologies for Deriving the Data Warehouse Structure*, Proceedings of the 2nd Symposium on Business Informatics, Austrian Computer Society Conference, pp. 198 –206, ISBN: 978-3-85403-280-9

Deasemenea am contribuit la o monografie despre sistemele inteligente de suport a deciziilor cu un sub-capitol în care am tratat aspecte teoretice ale tehnologiilor Business Intelligence și a depozitelor de date:

- Nițchi Ioan Ștefan, Airinei Dinu, Arba (Cordis-Herbil) Raluca, Benţa Dan, Brandas Claudiu, Buchmann Robert, Crisan Emil Lucian, Homocean Daniel, Jecan Sergiu, Kleinhempel Simona, Mihaila Adrian-Alin, Muntean Mihaela, **Nagy Ilona Mariana**, Petrusel Razvan, Podean Ioan Marius, Rusu Maria Lucia, Sitar-Taut Dan Andrei, book, *Sisteme inteligente de asistare a deciziilor*, Risoprint, Cluj-Napoca, 2010.

În privința direcțiilor de cercetare viitoare considerăm că prototipul prezentat ca parte a contribuției noastre la domeniul cercetat poate fi extins astfel încât să acopere automatizarea și a altor structuri de stocare și procese din mediul depozitelor de date. Prototipul poate fi deasemenea utilizat per sub-module pentru a asigura separarea anumitor procese care sunt executate, în contextul curent, pentru întregul model de date. De exemplu, sub-părți ale modulelor prototipului pot fi dezvoltate și utilizate exclusiv pentru generarea meta-obiectelor de modelare sau stocare a datelor, pornind de la metadate definite în documente tehnice, putându-se astfel reduce semnificativ timpul de implementare al depozitelor de date, precum și costurile aferente. Totodată, procedurile implementate pentru validarea tehnică automatizată a structurilor de date generate pot fi utilizate separat pentru efectuarea acestor acțiuni la nivelul obiectelor deja existente în mediul depozitelor de date.

Dezvoltări ulterioare adiționale pot cuprinde: includerea parțială sau completă a metadatelor de business în procesul de automatizare (e.g. în transformarea datelor dintre nivelele de *Gestiune Primară a Datelor* și *Livrarea Datelor*); generarea automată a obiectelor pentru structurile de tip master data, ca parte a nivelului de *Gestiune Primară a Datelor* în mediul SAP BW, etc. În final, considerăm că am reușit să propunem un framework coerent și un prototip flexibil ce pot aduce beneficii importante intreprinderilor care implementează sau oferă servicii de mentenanță pentru soluții de depozite de date.