



Universitatea
BABEȘ-BOLYAI

UBB FSEGA
Universitatea Babeș-Bolyai Facultatea de Științe Economice și Științe ale Afacerilor



CENTRUL DE FORMARE CONTINUĂ, ÎNVĂȚĂMÂNT
LA DISTANȚĂ ȘI CU FRECVENȚĂ REDUSĂ

OIDD FR OFICIUL PENTRU ÎNVĂȚĂMÂNT LA DISTANȚĂ
ȘI CU FRECVENȚĂ REDUSĂ

Specializarea: Informatică Economică

SUPOORT DE CURS

Business Intelligence

ANUL III Semestrul 6



Cluj-Napoca 2021

I. Informații generale

• Date de identificare a cursului

Date de contact ale titularului de curs:

Nume: Conf.Dr. Liana Stanca
Birou: 432, sediul FSEGA
Telefon: 0264-418 655
Fax: 0264-418 655
E-mail:
liana.stanca@econ.ubbcluj.ro
Consultații: Conform cu orarul afișat la sala
432

Date de identificare curs și contact tutori:

Numele cursului: Business Intelligence
Cod: ELR0261
Pagina web a cursului: <http://www.econ.ubbcluj.ro>
An studiu: Anul 3, semestru 6
Tip: Obligatoriu
Tutore: Conf.Dr. Liana Stanca
E-mail secundar: liana.stanca@econ.ubbcluj.ro

• Condiționări și cunoștințe prerechizite

Cursul nu are condiționări prerechizite. Cunoștințele prerechizite care pot facilita asimilarea materialului sunt legate de programare la nivel de bază și rețele de calculatoare. Sugerăm ca înainte de parcurgerea materialului să se identifice următoarele cunoștințele prerechizite:

- conceptele de teoria bazelor de date relaționale
- conceptele de programare de bază: rețele de calculatoare,
- modelele de procese de producție software din perspectiva proiectării;
- documentația necesară proiectării sistemelor informatice în domeniul economic.

• Descrierea cursului

Cursul prezintă aspecte teoretice și practice ale domeniului Business Intelligence (BI). Business Intelligence se referă la tehnologii, aplicații și practici pentru colectarea, integrarea, analiza și prezentarea informațiilor de afaceri. Scopul cursului de business intelligence este acela de a induce studenților skills-uri necesare procesului de proiectare, implementare și dezvoltare a instrumentelor de BI ce sprijină procesul decizional al oricărei afaceri. Cursul va examina tehnologiile de Business Intelligence (BI) care ajută o companie să-și îmbunătățească afacerea. Acesta discută subiecte BI atât din perspectivă managerială, cât și din punct de vedere tehnic. Perspectivile manageriale discută modul în care BI afectează procesul decizional al organizației, în timp ce perspectivele tehnice discută fundamentul unui sistem inteligent. În cadrul cursului se vor discuta aspecte cheie începând de la BI ca proces și arhitectură, Warehousing, Online Analytical Processing, Data Mining. În cadrul cursului vor fi atribuite exerciții practice și proiecte pentru a îmbunătăți experiența studenților în domeniul inteligenței de afaceri.

Sistemele de Business Intelligence au devenit din ce în ce mai importante în mediul competitiv de astăzi. Potrivit unor studii recente, companiile care folosesc BI și își gestionează datele ca o resursă strategică și investesc în calitatea acesteia avansează deja în ceea ce privește eputația și profitabilitatea. Acest curs va examina tehnologiile de Business Intelligence (BI) care ajută o companie să-și îmbunătățească afacerea. Scopul cursului, este de a oferi studenților o privire de ansamblu asupra subiectelor de Business Intelligence. La finalizarea cursului, studenții vor înțelege elementele fundamentale ale Business Intelligence și vor fi capabili să evalueze tehnicile BI care vor fi utilizate pentru anumite aplicații. Competențe dobândite prin absolvirea disciplinei:

1. absolvenții disciplinei sunt specialiști în informatică economică cu abilități de programare și analiză de business;
2. însușirea abilităților de concepere, proiectare și implementare a modelelor de analiză vizuală, ceea ce le crează posibilitatea încadrării în domeniul statisticii și informaticii economice
3. fundamentarea necesarului de resurse umane, materiale și financiare în raport cu cerințele

- obiectivelor fixate de activitățile specifice programării afacerilor on-line;
4. au competența necesară de a dezvolta aplicații specifice afacerilor electronice;
 5. au competența necesară de a optimiza metodele de selecție a alternativelor de dezvoltare tooluri de BI;
 6. dobândesc deprinderi de analiză proiectare și testare a modelelor de BI specifice putând deveni consultanți în acest domeniu;
 7. dobândesc deprinderi de măsurare a eficienței instrumentelor de BI;
 8. pot concepe, proiecta, implementa și testa instrumente de BI;
 9. utilizarea instrumentelor de business intelligence, cunoașterea principalele metodologii asociate și tendințele actuale din acest domeniu.
 10. utilizarea și dezvoltarea metodelor și modelelor specifice inteligenței artificiale aplicabile într-un sistem informatic integrat
 11. dezvoltarea abilităților de a proiecta, implementa tool-uri de vizualizare a datelor
 12. dezvoltarea abilităților de a proiecta, implementa tool-uri de extragerea de cunoștințe din date și de analiza respectiv interpretare.
 13. dezvoltarea abilităților de a concepe proiecta și implementa tooluri de analiza și interpretare a datelor (Experimental analysis techniques; Predictive Modeling; Data Mining; Vizualizare; Social Network Analysis)
 14. dezvoltarea abilităților de a analiza proceselor sociale și economice, cu componentele lor spațiale și temporale.

• Organizarea temelor în cadrul cursului

Ordinea temelor abordate de curs, conform structurii materialului didactic ce va fi disponibil pe platforma software de învățământ la distanță: <https://portal.portalid.ubbcluj.ro/>

Modul 1. Business Intelligence: concept, trăsături, analiză, vizualizare, tendințe

- 1.1. Definiția conceptului de business intelligence
- 1.2. Scop și structură proiecte business intelligence
- 1.3. Introducere în vizualizarea datelor
- 1.4. Tendințele pieței BI.

Modul 2. Arhitectura sistemelor de BI

- 2.1 Informații despre corporații și depozit de date.
- 2.2 Design multidimensional. OLAP. Tabele de control.

Modul 3. Analiza de afaceri: Vizuală

- 3.1. Definirea conceptului;
- 3.2. Proiectare și analiză vizualizare.
- 3.3. Tehnici de vizualizare a informațiilor cu algoritmi automatici pentru a sprijini raționamentul analitic prin interacțiunea om-computer
- 3.4. Clasificare**
- 3.5. Analiza Cluster**

Bibliografie obligatorie

Bibliografie opțională

Pentru paginarea temelor, recomandăm consultarea cuprinsului din partea a doua a materialului de față.

• Formatul și tipul activităților implicate de curs

Cursul va fi prezentat prin activități tutoriale periodice programate conform orarului facultății, afișat pe site-ul <http://econ.ubbcluj.ro>.

Prin adresele de e-mail oferită sau la sediul facultății, titularul și tutorii cursului stau la dispoziția studenților pentru consultații on-line sau față în față în afara activităților periodice preprogramate. Se încurajează studiile de caz legate de locul de muncă al acelor studenți care sunt deja angajați în domeniul testării software.

Activitățile tutoriale sunt, pentru studentul la distanță, facultative și nu afectează nota acestuia, obținută strict prin forma indicată: examen scris și prezentarea unui proiect ce are ca temă conceperea, proiectarea și implementarea unei afaceri virtuale pe baza cunoștințelor asimilate în cadrul cursului de față. Totuși, încurajăm participarea interactivă la activitățile tutoriale în special pentru dezvoltarea incrementală a dosarului de testare care va fi notat.

• **Materiale bibliografice obligatorii**

1. Business Intelligence: A Managerial Approach (2014) Turban, Sharda, Delen, King, Publisher: Prentice Hall, Edition: 2nd, ISBN: 13-978-0-136- 10066-9
 2. Turban, Efraim, Ramesh Sharda, and Dursun Delen. "Decision support and business intelligence systems (required)." Google Scholar (2010).
 3. Chen, Hsinchun, Roger HL Chiang, and Veda C. Storey. "Business intelligence and analytics: From big data to big impact." MIS quarterly 36.4 (2012).
 4. Turban, Efraim, Ramesh Sharda, and Dursun Delen. Business intelligence and analytics: systems for decision support. Pearson Higher Ed, 2014.
 5. Satriadi, K. A., Ens, B., Czauderna, T., Cordeil, M., & Jenny, B. (2021). Quantitative Data Visualisation on Virtual Globes.
 6. Zhou, Z., Wen, X., Wang, Y., & Gotz, D. (2021). Modeling and Leveraging Analytic Focus During Exploratory Visual Analysis. arXiv preprint arXiv:2101.08856.
- Materialele sunt accesibile la biblioteca facultății, la biblioteca catedrei sau pot fi puse la dispoziție de către titularul de curs.

• **Materiale și instrumente necesare pentru curs**

Calculator, materialul bibliografic, software licențiat și free necesar cursului și anume Tableau puse la dispoziție de titularul cursului.

• **Calendar al cursului**

Sunt estimate întâlniri preprogramate pe semestru, cu datele și locațiile afișate pe site-ul facultății la începutul semestrului. Premergător fiecărei întâlniri se recomandă parcurgerea materialului de față, pe module pentru a asigura cursivitatea discuțiilor. Conținutul acestor întâlniri va fi, în ordine: modulelor I, II; discuții pe marginea modulelor II, III; discuții pe marginea modului IV.

• **Politica de evaluare și notare** (orientativ - 1 pagină)

Evaluare practică – 50% din notă:

Conținut:

Dezvoltarea unui proiect care să conțină o aplicație software necesară derulării unei afaceri electronice la alegere. Proiectul va conține:

- o aplicație destinată curățării datelor;
- o aplicație destinată analizei vizuale;
- o bază de date care conține toate datele necesare pentru realizarea celor 2 aplicații.

Evaluare teoretică – 50% din notă

Conținut:

Test grilă cu întrebări, de dificultate și pondere în notă echitabile (30 întrebări de 0,3 punct pe întrebare).

Nivelul minim pentru promovarea examenului este dat de obținerea notei 5 la fiecare din cele două părți (practic și teoretic).

Ambele evaluări vor avea loc la datele programate pentru examen, la sediul facultății. Notele vor fi acordate în aceeași zi, comunicate personal fiecărui student cu posibilitate de contestare imediată.

Nu se vor accepta proiecte practice sau participări la examen la alte date decât cele programate.

• Elemente de deontologie academică

Tentativele de fraudare atât la examen scris cât și în dezvoltarea proiectului practic vor fi pedepsite prin anularea examenului și aplicarea regulamentului instituțional. Nu este admisă în timpul examenului utilizarea mijloacelor de comunicație. Contestațiile se vor soluționa în maxim 24 de ore de la afișarea rezultatelor.

• Studenți cu dizabilități

Titularul cursului și tutorii își afirmă disponibilitatea, în limita posibilităților, de a adapta la cerere, conținutul și metodele de transmitere a informațiilor, precum și modalitățile de evaluare (examen oral, examen on line) în funcție de tipul dizabilității cursantului. Vom urmări facilitarea accesului egal al tuturor cursanților la activitățile didactice.

E-mail de contact pentru situații deosebite și suport acordat studenților cu dizabilități:

liana.stanca@econ.ubbcluj.ro

• Strategii de studiu recomandate

Recomandăm în ordine:

- parcurgerea materialului de față și contactarea tutorilor pentru orice nelămuriri;
- parcurgerea bibliografiei obligatorii;
- cercetarea individuală pe tema cursului, folosind Internetul;
- parcurgerea documentațiilor on-line Tableau;

Cuprins

Table of Contents

MODULUL 1 BUSINESS INTELLIGENCE: CONCEPT, TRĂSĂTURI, ANALIZĂ, VIZUALIZARE, TENDINȚE	9
2.1.3. Conținutul informațional detaliat	10
2.1.3 .1 Definiția Conceptului de Business Intelligence.....	10
MODULUL 2. Arhitectura sistemelor de BI.....	18
3.4. Conținutul informațional detaliat	19
3.4.1. Data warehouse.....	19
3.4.2. Modele depozite de date	20
2.4.3. Model multidimensional.....	22
2.4.4. Extract Transform Load.....	24
2.4.5. Procesare Analitică Online	25
2.1.6. Tradițional vs Cloud Data Warehouse~concluzie	27
MODULUL 3 Analiza vizuală a datelor	33
3.4. Conținutul informațional detaliat	34
3.4.1 Visual Analytics~noțiuni fundamentale	34
3.4.2. <i>Decision analysis, Visualisation and visual analytics</i> Tabloul de bord ~tool în vizualizarea dinamica a analizei datelor	37
3.4.3 Clasificarea.....	39
3.4.4. Clustering.....	41

Obiectivele cursului/disciplinei:

- Scopul cursului, este de a oferi studenților o privire de ansamblu asupra subiectelor de Business Intelligence. La finalizarea cursului, studenții vor înțelege elementele fundamentale ale Business Intelligence și vor fi capabili să evalueze tehnicile BI care vor fi utilizate pentru anumite aplicații.
- Se oferă programatorului toate informațiile necesare pentru stăpânirea tehnicilor de analiză vizuală.

2. Suportul de curs

MODULUL 1 BUSINESS INTELLIGENCE: CONCEPT, TRĂSĂTURI, ANALIZĂ, VIZUALIZARE, TENDINȚE

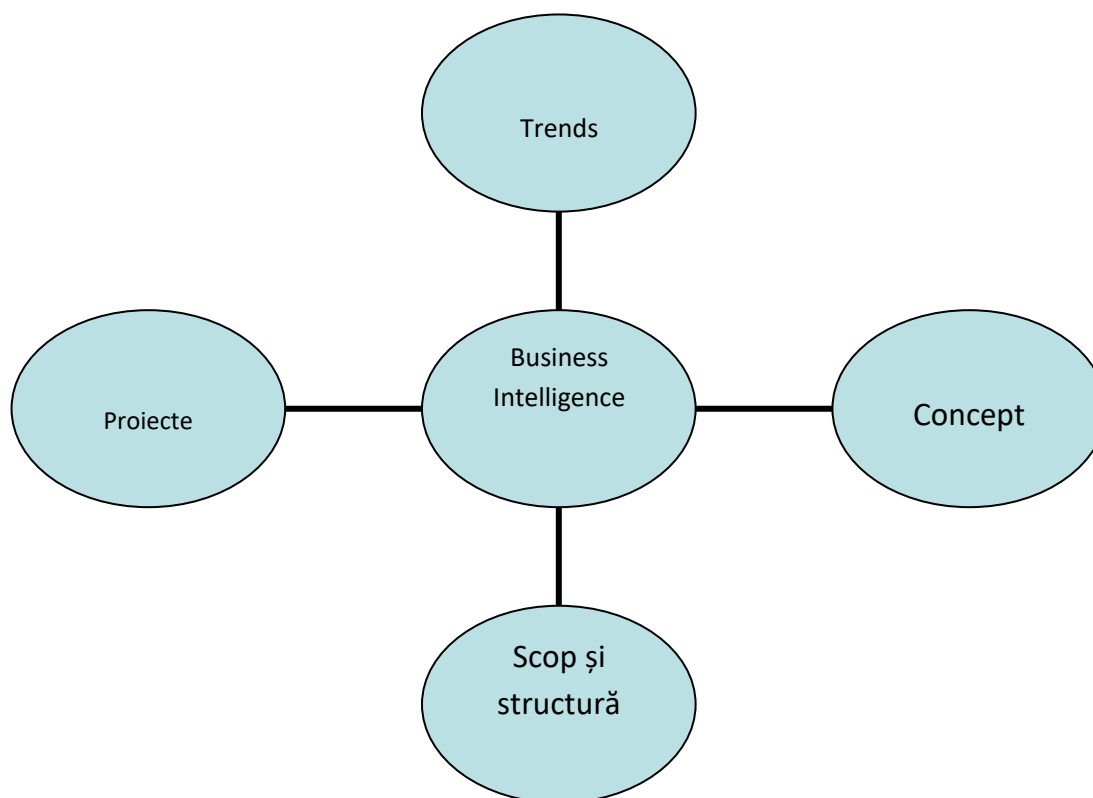
2.1.1. Scopul și obiectivele modului

Scopul acestui modul este familiarizarea cursantului cu termenul de business Intelligence, cu principiul de funcționare a aplicațiilor de business intelligence..

Ca obiective, se vor atinge următoarele:

1. Familiarizarea cu conceptele folosite în analiza vizuală a datelor
2. Oferirea unei imagini de ansamblu disciplinei Business Intelligence
3. Asimilarea noțiunii de Business Intelligence

2.1.2. Schema logică a modului



Recomandări privind studiul:

1. Sunt necesare cunoștințe teoretice de algoritmică și programare;
2. Studiarea conceptelor prezentate în vederea utilizării lor pe parcursul celorlalte capitole

Rezultate așteptate:

1. Asimilarea conceptelor de bază legate de analiza vizuală a datelor.
2. Studenții trebuie să înțeleagă și să asimileze principiile de bază ale lui Business Intelligence

2.1.3. Conținutul informațional detaliat

2.1.3 .1 Definiția Conceptului de Business Intelligence

Business intelligence este definită ca un proces de colectare și procesare a informațiilor de afaceri pentru a obține informații și a lua decizii de afaceri în scopul ca aceste afaceri să devină profitabile. Business Intelligence (BI) a fost definit ca tehnici bazate pe computer utilizate în identificarea, extragerea și analiza datelor de afaceri, cum ar fi veniturile din vânzări pe produse și/sau departamente sau costurile și veniturile asociate (Business Intelligence, 2011). Business intelligence este un termen umbrelă, care reunește software-ul de business intelligence împreună cu alte instrumente, aplicații și practici. (Questionpro, 2022) Practicile utilizate atât în etapa de colectare a informațiilor inteligente de afaceri, cât și în etapa de analiză a informațiilor au ca miză finală optimizarea performanței generale a afacerii. Pe baza acestei definiții se observă că business intelligence este un process secvențial în trei etape: (Questionpro, 2022)

1. etapa de colectare a datelor în business Intelligence depinde de procesul de cercetare în afaceri și se realizează cu ajutorul instrumentelor tehnologice de exemplu chestionare on-line, feedback-uri, colectarea de sentimente etc.
2. etapa de analiză a datelor colectate la etapa 1 facilitează obținerea de informații utile.
3. etapa de analiză și prezentare a informațiilor legate de afaceri permite crearea de acțiuni de afaceri cu caracter inteligent.

Business intelligence (BI) se referă la infrastructura procedurală și tehnică care colectează, stochează și analizează datele produse de activitățile unei companii. (Frankenfield, 2021)

Deci, pe baza definițiilor prezentate mai sus business intelligence (Frankenfield, 2021) (Larissa Terpeluk Moss, 2003) (Questionpro, 2022) reprezintă atât infrastructura tehnică care colectează, stochează și analizează datele companiei cât și analizează datele cu scopul de a produce rapoarte și informații care ajută pe manageri să ia decizii.

Practicile BI continuă să evolueze și oferă afacerilor din multe sectoare instrumente și informații din ce în ce mai puternice pentru luarea deciziilor. Cantități mari de informații disponibile în companii trebuie valorificate și interpretate - „prin inițiative de BI, companiile obțin perspective din volumele tot mai mari de date privind tranzacții, produse, stocuri, clienți, concurenți și industrie generate de aplicațiile la nivel de întreprindere” (Chen, Chiang, & Storey, 2010, p. 201) Watson și Wixom (2007) afirmă că „BI apare ca un factor cheie pentru creșterea valorii și a performanței... viitorul business intelligence este luminos” (p. 28). Exemple de utilizare a BI:

1. în departamentului de poliție din Richmond Virginia cu rol în predicție, când și unde vor avea loc crimele și a ajutat la alocarea resurselor în consecință (Smalltree, 2007). Supraveghetorii de tură au acces la informații detaliate la începutul turei, precum și la alerte în timp real din sistem atunci când este detectat un tipar de infracțiune. Instrumentele BI au permis departamentului de poliție să fie mai degrabă proactiv decât reactiv în combaterea criminalității.

2. Euro Disney folosește business intelligence pentru a optimiza distribuția personalului și pentru a evita supraaglomerarea parcurilor de distracții (Laursen & Thorlund, 2010). Alături de exemplele menționate mai sus, nenumărate organizații utilizează BI pentru a îmbunătăți practicile de afaceri.

Deci, nu este vorba doar despre transformarea datelor în informații, deși Business Intelligence implică în mod semnificativ acest proces. Business Intelligence transformă datele în perspective semnificative, acționabile, care permit organizațiilor să ia strategii de afaceri informate și decizii tactice.

Termenul de Business Intelligence se referă de obicei la o varietate de instrumente care oferă acces rapid și ușor de înțeles la informații despre starea actuală a unei organizații pe baza datelor existente.

Business Intelligence (Site2, 2021), este o infrastructură procedurală și tehnică care include colectarea, stocarea și analiza datelor produse de activitățile dintr-o companie. Acesta cuprinde procese precum analiza proceselor, extragerea datelor (text), evaluarea comparativă a performanței și, de asemenea,

analiza descriptivă. BI implică, de asemenea, analizarea datelor pentru a prezenta rapoarte de performanță și tendințe care informează deciziile de management.

2.1.3.2 Scop și structură proiecte business intelligence

Scopul business intelligence (Frankenfield, 2021) este de a colecta și a analiza informații pentru a facilita procesul de luare a deciziilor de afaceri care să genereze profit. În lumea competitivă a afacerilor secolului 21, capacitatea de a utiliza datele și tehnologia în timp real este unul dintre cei mai importanți parametri pentru a reuși a unei afaceri în scopul de a facilita accesul la informații rapide în scopul de a concepe decizii productive, analitice și de impact. Pe baza scopului se obțin tendințele de creștere a importanței business intelligence în diferite domenii, mai mult se poate constata că numai colectarea datelor nu este suficientă pentru tendințe ci mai exact procesarea și analiza datelor este ceea ce duce la perspective acționabile care conduc deciziile de afaceri. Domeniile care sunt în prezent afectate de business intelligence:

- Investiții sporite în tehnologia AI
- Date mare
- Importanța tot mai mare a guvernării datelor
- Creșterea software-ului și instrumentelor de business intelligence cu autoservire
- Interpretarea datelor prin povestire
- Business intelligence colaborativ
- Inteligență de afaceri încorporată
- Analiza cloud

Instrumentele și software-ul BI au o varietate de forme, de la foi de calcul (Microsoft Excel și Google Docs), software de raportare/interogare, software de vizualizare a datelor până la instrumente de extragere a datelor și procesare analitică online (OLAP). Instrumentele OLAP permit utilizatorilor să analizeze seturi de date dintr-o mare varietate de unghiuri, bazate pe diferite perspective de afaceri. (Frankenfield, 2021) Aplicațiile de tip business intelligence facilitează multe activități, inclusiv cele enumerate mai jos (Larissa Terpeluk Moss, 2003):

- Analiză multidimensională, de exemplu, procesare analitică online (OLAP)
- Click-stream analysis
- Data mining
- Forecasting
- Business analysis
- Balanced scorecard preparation
- Visualization
- Querying, reporting, and charting (including just-in-time and agent-based alerts)
- Geospatial analysis
- Knowledge management
- Enterprise portal implementation
- Mining for text, content, and voice
- Digital dashboard access
- Other cross-functional activities
- Exemple de baze de date de sprijin pentru decizii BI includ următoarele:
- Enterprise-wide data warehouses
- Data marts (functional and departmental)
- Exploration warehouses (statistical)
- Data mining databases
- Web warehouses (for click-stream data)
- Operational data stores (ODSs)
- Operational marts (oper marts)
- Other cross-functional decision-support databases

BI este un mediu în evoluție la nivel de întreprindere, îmbunătățit continuu pe baza feedback-ului din partea comunității de afaceri, practicile de dezvoltare și înlocuire a sistemelor din trecut considerate inadecvate (Larissa Terpeluk Moss, 2003). În trecut, fiecare sistem a avut un început și un sfârșit și

fiecare sistem a fost conceput pentru a rezolva o singură problemă izolată pentru un set de oameni de afaceri dintr-o linie de afaceri. Vechile practici de dezvoltare „cu o singură bandă” erau potrivite pentru astfel de sisteme statice de sine stătătoare. Cu toate acestea, ele nu sunt potrivite pentru inițiativele integrate de Business Intelligence, deoarece vechile practici nu includ activități interorganizaționale necesare pentru a susține un mediu de sprijin decizional la nivelul întregii întreprinderi. În trecut, activitățile interorganizaționale nu numai că erau considerate inutile, ci erau percepute ca nu facilitează progresul, deoarece încetineau proiectele. Pentru dezvoltarea unui sistem neintegrat, metodologiile convenționale în cascadă sunt suficiente deoarece oferă suficientă îndrumare pentru planificarea, construirea și implementarea sistemelor de sine stătătoare. Spre deosebire de sistemele statice de sine stătătoare, un mediu de asistență de tip business intelligence dinamic și integrat nu poate fi construit într-un singur big bang. Datele și funcționalitățile trebuie să fie lansate în versiuni iterative, iar fiecare implementare este probabil să declanșeze noi cerințe (Larissa Terpeluk Moss, 2003)

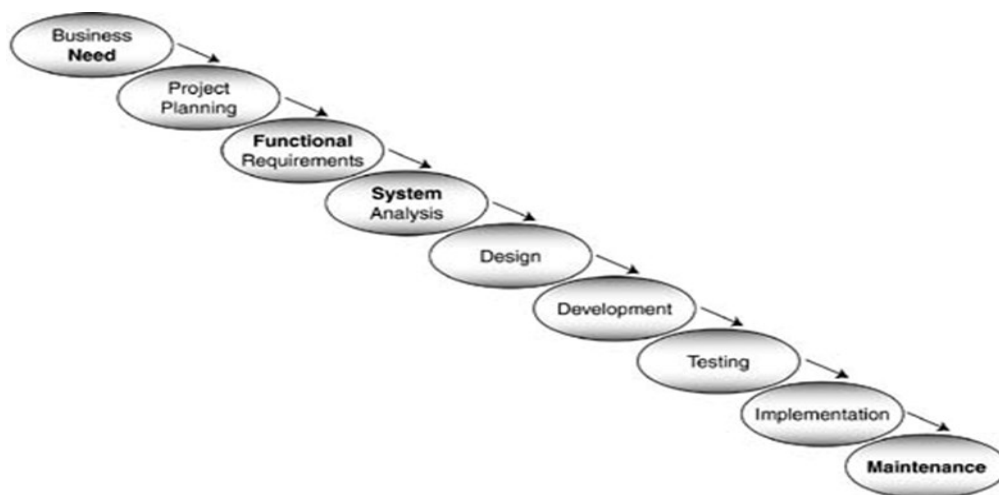


Figura 2.1.3.2. Waterfall Deployment (Larissa Terpeluk Moss, 2003)

Conform lui (Larissa Terpeluk Moss, 2003) Waterfall Deployment este metodologie de dezvoltare a vechilor sisteme

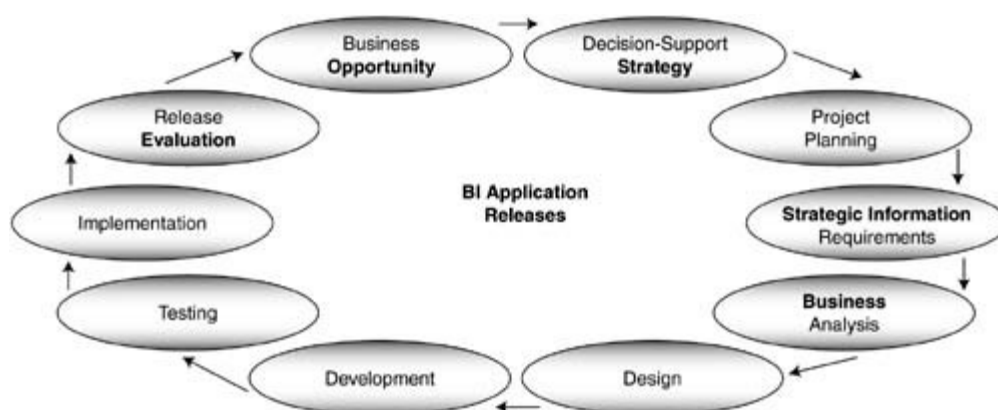


Figura 2.1.3.2. Conceptul de lansare a aplicației de tip Business Intelligence (Larissa Terpeluk Moss, 2003)

Proiectele de tip Business Intelligence sunt organizate în conformitate cu aceleași șase etape comune

fiecărui proiect de inginerie. În cadrul fiecărei etape de inginerie, se parcurg anumiți pași pentru a duce proiectul de inginerie până la finalizarea acestuia. Foaia de parcurs de Business Intelligence descrie 16 pași de dezvoltare în cadrul acestor etape, după cum este prezentat mai jos (Larissa Terpeluk Moss, 2003).

Pasul 1: Evaluarea cazului de afaceri

Se definește problema de afaceri sau oportunitatea de afaceri și se propune o soluție BI. Fiecare versiune de aplicație BI ar trebui să fie justificată din punct de vedere al costurilor și ar trebui să definească în mod clar beneficiile fie a rezolvării unei probleme de afaceri, fie a unei oportunități de afaceri.

Pasul 2: Evaluarea infrastructurii întreprinderii

Aplicațiile BI sunt inițiative interorganizaționale, se impune crearea unei infrastructuri de întreprindere care să le susțină. O infrastructură de întreprindere are două componente:

1. Infrastructură tehnică, care include hardware, software, middleware, sisteme de gestionare a bazelor de date, sisteme de operare, componente de rețea, depozite de metadata, utilități și așa mai departe.
2. Infrastructură nontehnică, care include standarde de metadata, standarde de denumire a datelor, modelul de date logic al întreprinderii (în evoluție), metodologii, linii directe, proceduri de testare, procese de control al schimbărilor, proceduri pentru gestionarea problemelor și soluționarea litigiilor și așa mai departe.

Pasul 3: Planificarea proiectului

Proiectele de sprijin pentru decizii BI sunt extrem de dinamice. Modificările aduse domeniului, personalului, bugetului, tehnologiei, reprezentanților afacerilor și sponsorilor pot afecta grav succesul unui proiect. Prin urmare, planificarea proiectului trebuie să fie detaliată, iar progresul real trebuie urmărit îndeaproape și raportat.

Etapa de analiză a afacerii cuprinde pași

Pasul 4: Definirea cerințelor proiectului

Cerințele au caracter fluctuant parcursul ciclului de dezvoltare, pe măsură ce oamenii de afaceri învață mai multe despre posibilitățile și limitările tehnologiei BI în timpul proiectului.

Pasul 5: Analiza datelor

Cea mai mare provocare pentru toate proiectele de sprijin decizional BI este calitatea datelor sursă. Analiza datelor în trecut a fost limitată la viziunea unei linii de afaceri și nu a fost niciodată consolidată sau reconciliată cu alte opinii din organizație.

Pasul 6: Prototiparea aplicației

Analiza rezultatelor funcționale, se face cel mai bine prin prototipare, astfel încât să poată fi combinată cu proiectarea aplicației. Prototiparea le permite oamenilor de afaceri să vadă potențialul și limitele tehnologiei, ceea ce le oferă posibilitatea de a-și ajusta cerințele proiectului și așteptările lor.

Pasul 7: Analiza depozitului de metadata

A avea mai multe instrumente înseamnă a avea mai multe metadata tehnice în plus față de metadatale de afaceri, care sunt de obicei capturate într-un instrument de modelare de inginerie software asistată de computer (CASE). Metadatale tehnice trebuie mapate la metadatale de afaceri și toate metadatale trebuie stocate într-un depozit de metadata. Arhivele de metadata pot fi licențiate (cumpărate) sau construite. În ambele cazuri, cerințele pentru ce tip de metadata de captat și stocat ar trebui să fie

documentate într-un metamodel logic.

Etapa de proiectare cuprinde pașii:

Pasul 8: Proiectarea bazei de date

Una sau mai multe baze de date țintă BI vor stoca datele de afaceri în formă detaliată sau agregată, în funcție de cerințele de raportare ale comunității de afaceri. Nu toate cerințele de raportare sunt strategice și nu toate sunt multidimensionale. Schemele de proiectare a bazei de date trebuie să corespundă cerințelor de acces la informații ale comunității de afaceri.

Pasul 9: Extrageți/Transformați/Încărcați Designul

Procesul ETL este cel mai complicat proces al întregului proiect de sprijinire a deciziilor BI. Ferestrele de procesare ETL (ferestrele lot) sunt de obicei mici, dar calitatea slabă a datelor sursă necesită, de obicei, mult timp pentru a rula transformarea.

Pasul 12: Dezvoltarea aplicației

Odată ce efortul de prototipare a consolidat cerințele funcționale, poate începe dezvoltarea adevărată a aplicației de acces și analiză. Dezvoltarea aplicației poate fi o simplă chestiune de finalizare a unui prototip operațional sau poate fi un efort de dezvoltare mai implicat folosind instrumente de acces și analiză diferite, mai robuste. În ambele cazuri, activitățile de dezvoltare a aplicațiilor front-end sunt de obicei efectuate în paralel cu activitățile de dezvoltare ETL back-end și dezvoltarea depozitului de metadate.

Pasul 13: Exploatarea datelor

Aplicațiile BI sunt adesea limitate la rapoarte prescrise, dintre care unele nu sunt nici măcar tipuri noi de rapoarte, ci înlocuiri ale rapoartelor vechi. Rambursarea reală provine din informațiile ascunse în datele organizației, care pot fi descoperite doar cu instrumente de data mining.

Pasul 14: Dezvoltarea depozitului de metadate

Dacă se ia decizia de a construi un depozit de metadate, mai degrabă decât de a licenția unul, o echipă separată este de obicei însărcinată cu procesul de dezvoltare. Acesta devine un subproiect considerabil în proiectul general BI.

Etapa de implementare cuprinde pașii:

Pasul 15: Implementare

Odată ce echipa a testat temeinic toate componentele aplicației BI, echipa lansează bazele de date și aplicațiile. Instruirea este programată pentru personalul de afaceri și pentru alte părți interesate care vor folosi aplicația BI și depozitul de metadate. Încep funcțiile de asistență, care includ operarea biroului de asistență, menținerea bazelor de date țintă BI, programarea și rularea joburilor batch ETL, monitorizarea performanței și reglarea bazelor de date.

Pasul 16: Evaluarea lansării

Cu un concept de lansare a aplicației, este foarte important să beneficiezi de lecțiile învățate din proiectele anterioare. Orice termene limită nerespectate, depășiri de costuri, dispute și soluționări ale litigiilor ar trebui să fie examinate, iar ajustările procesului trebuie făcute înainte de începerea următoarei ediții. Orice instrumente, tehnici, linii directoare și procese care nu au fost utile ar trebui reevaluate și ajustate, eventual chiar aruncate.

2.1.3.3 Exemple de Business Intelligence

Diverse instrumente BI pot accesa și analiza seturi de date și pot reprezenta constatările analitice sub formă de rezumate, rapoarte, tablouri de bord, diagrame, grafice și hărți pentru a oferi utilizatorilor informații detaliate cu privire la starea afacerii. De exemplu, instrumente precum Power BI și Tableau pot fi folosite pentru vizualizarea datelor și pentru business intelligence, facilitând cu succes punerea în practică a următoarelor scenarii:

Scenariu 1. Lowe's Corp, care operează al doilea cel mai mare lanț de vânzare cu amănuntul din țară, au adoptat instrumentele BI pentru a-și optimiza lanțul de aprovizionare, pentru a analiza produsele pentru a identifica potențialele fraude și pentru a rezolva problemele cu taxele colective de livrare din magazinele sale.

Scenariu 2. Compania de îmbuteliere Coca-Cola a avut o problemă cu procesele sale zilnice de raportare manual, mai exact au restricționat accesul la datele de vânzări și operațiuni în timp real. Înlocuirea procesului manual cu un sistem automatizat de BI, compania a simplificat complet procesul și a economisit 260 de ore pe an (sau mai mult de șase săptămâni de lucru de 40 de ore), echipa companiei poate analiza rapid valori precum operațiunile de livrare, buget și profitabilitate cu doar câteva clicuri.

Scenariu 3. Un proprietar de pizzerie dorește să înțeleagă preferințele clienților săi, tipul de pizza care le place și de ce le place. Pizzeria a continuat să efectueze un sondaj folosind software de business intelligence, iar feedback-ul este colectat de la clienți despre diferite tipuri de pizza. Proprietarul pizzeriei, care utilizează funcții proeminente din software, analizează mentalitatea clientului și obține informații valoroase despre pizzelle servite în restaurant. Toolul de BI îi oferă proprietarului informații despre pizza care este pe placul oamenilor și ce pizza este comandată cel mai mult în ce zi etc., astfel pe baza acestor informații proprietarul poate să creeze strategii de prețuri satisfăcătoare pentru clienții săi și profitabile pentru afacerea sa.

Deci, majoritatea companiilor pot beneficia de încorporarea soluțiilor BI în care managerii cu informații inexacte sau incomplete vor avea tendința, în medie, să ia decizii mai proaste decât dacă ar avea informații mai bune.

2.1.3.4. Business intelligence trends

În prezent, supraviețuirea afacerilor, a impus aplicarea soluțiilor de business intelligence ca o oportunitate de menținere, creștere, extindere și durabilitate a lor în viitor. BI trebuie să depună eforturi pentru a îmbunătăți acuratețea, actualitatea și volumul datelor pentru a fi semnificative. Așadar, trebuie găsite mai multe modalități de a capta informații pentru a verifica erorile, cea care nu este înregistrat, și de a organiza informațiile în așa fel încât să devină ușor de efectuat o analiză amplă.

Formularea deciziilor strategice de afaceri impune ca o organizație să cunoască tendințele pieței, nevoile clienților și opiniile consumatorilor pe baza informațiilor. Instrumentele de Business Intelligence facilitează numeroase moduri de a colecta informații despre afaceri. Metodologiile recomandate pe scară largă pentru a colecta informații fiabile pentru analiza statistică ce facilitează crearea de decizii bazate pe date, sunt: (Frankenfield, 2021) În prezent, sistemele de business intelligence au devenit capabile să proceseze date operaționale în timp real. Datele operaționale pentru o afacere pot fi disponibile dintr-un sistem de resurse Enterprise (ERS), un sistem de management al relațiilor cu clienții (CRM), un sistem de management al lanțului de aprovizionare (SCM) sau diferite baze de date din cadrul organizației (Bataweel, 2015). „Inteligența operațională de afaceri” (Bataweel, 2015). adaugă multe alte funcționalități practicilor comune de business intelligence. Folosind inteligența operațională de afaceri, organizațiile pot gestiona performanța, procesele de afaceri și tablourile de bord și tablourile de bord bazate pe roluri. Exemple de luare a deciziilor care pot fi făcute cu inteligența operațională de afaceri sunt realizarea unei evaluări a pieței pe mai multe zone geografice și realizarea unei liste de recomandări de produse personalizate pentru un client. (Bataweel, 2015). Sec 21, este marcat de schimbări care pot fi asociate cu dezvoltarea business intelligence mai exact: livrarea la timp a datelor reale pentru luarea deciziilor, o schimbare de paradigmă poate fi recunoscută în

mediile de operare actuale în care obiectivele sistemului de business intelligence s-au schimbat semnificativ. Există diferențe notabile între obiectivele anterioare ale business intelligence și situațiile actuale dictate de procesele externe și interne ale unei afaceri. Datele brute(Vine, 2000). sunt asamblate și manipulate în informații, iar ulterior informațiile trebuie cernute și concepute în modele semnificative. Informațiile extrase din analiza datelor servesc la structurarea business intelligence a companiei. Nevoile de business intelligence se schimbă transversal, vânzări și marketing, resurse umane, operațiuni și finanțe. Departamentele de vânzări și marketing urmăresc produsele, clienții, datele demografice, promoțiile, forța de vânzări, tipul de comandă și alte domenii conexe. Resursele umane caută în mod regulat să măsoare problemele angajaților, organizaționale și departamentale. Ritmul de adunare, stocul de depozit, costul producătorului și furnizorului și beneficiul mișcării sunt domeniile de operațiuni administrare. Companiile(Vine, 2000). care și-au anunțat nevoia unei strategii de big data, care implică o aplicație complexă de set de date, s-au gândit că li s-a părut o investiție inteligentă și informată. Încep să se întrebe de ce colectează date în primul rând. În acest peisaj Microsoft este considerată cea mai bună companie de business intelligence, urmată de Tableau Software, Qliktech și Logi Analytics (Vine, 2000). Business Intelligence revoluționată de schimbările în furnizarea de informații, sub formă de rapoarte (tablouri de board) (Bataweel,2015). De obicei, companiile(Site2, 2021) au date care sunt în diferite formate diferite și chiar nestructurate, ceea ce face dificilă colectarea și analizarea. Pentru a simplifica procesul, firmele de software oferă soluții de business intelligence pentru a optimiza informațiile obținute din date.

Cele mai cunoscute trenduri în BI în prezent sunt:

- Inteligență artificială: raportul Gartner indică faptul că AI și învățarea automată preiau acum sarcini complexe realizate de inteligența umană. Această capacitate este valorificată pentru a realiza analize de date în timp real și raportare tablou de bord.
- BI colaborativ: software-ul BI combinat cu instrumente de colaborare, inclusiv rețelele sociale și alte tehnologii de ultimă oră, îmbunătățesc lucrul și partajarea de către echipe pentru luarea deciziilor în colaborare.
- BI încorporat: BI încorporat permite integrarea software-ului BI sau a unora dintre caracteristicile acestuia într-o altă aplicație de afaceri pentru îmbunătățirea și extinderea funcționalității sale de raportare.
- Cloud Analytics: aplicațiile BI vor fi oferite în curând în cloud, iar mai multe companii vor trece la această tehnologie. Conform predicțiilor lor, în câțiva ani, cheltuielile pentru analizele bazate pe cloud vor crește de 4,5 ori mai repede.

Cele mai cunoscute instrumente respective categorii de software care intră sub incidența BI sunt:

- Tablouri de bord
- Vizualizarea datelor
- Crearea rapoartelor
- Exploatarea datelor
- Instrumentele ETL sau Extract-Transform-Load sunt menite să importe date dintr-un depozit de date în altul.
- OLAP sau procesare analitică online.

Business Intelligence, sau mai exact, Tableau, este clasat ca lider de către Gartner pentru al paisprezecelea an consecutiv în 2021.

Teme pentru verificarea cunoștințelor

1. Business Intelligence este definit ca:

- ☐ este definită ca un proces de colectare și procesare a informațiilor de afaceri pentru a obține informații și a lua decizii de afaceri în scopul ca aceste afaceri să devină profitabile. ca o ca o colecție largă rețea de rețele
- ☐ ca o reuniune de calculatoare pe arie medie de extindere
- ☐ ca o platformă pentru o colecție de calculatoare

2. Scopul business intelligence:

- ☐ corectarea erorilor și secvențierea transportului.
- ☐ este de a colecta și a analiza informații pentru a facilita procesul de luare a deciziilor de afaceri care să genereze profit.
- ☐ transmiterea de pachete de la sursă la destinație
- ☐ definirea modului prin care vor dialoga clientul și serverul

3. Trenduri în business intelligence sunt:

- ☐ Dependența de platformă, Structurarea riguroasă a documentelor, Hipertext
- ☐ Inteligență artificială, BI colaborativ, BI încorporat, Cloud Analytics
- ☐ Hipertext, Independența de platformă, Structurarea riguroasă a documentelor
- ☐ Interferența cu platforma , Structurarea riguroasă a serverelor, Hiperdocument

Răspunsuri

1. Business Intelligence este definit ca:

- ☐ este definită ca un proces de colectare și procesare a informațiilor de afaceri pentru a obține informații și a lua decizii de afaceri în scopul ca aceste afaceri să devină profitabile. ca o ca o colecție largă rețea de rețele
- ☐ ca o reuniune de calculatoare pe arie medie de extindere
- ☐ ca o platformă pentru o colecție de calculatoare

2. Scopul business intelligence:

- ☐ corectarea erorilor și secvențierea transportului.
- ☐ este de a colecta și a analiza informații pentru a facilita procesul de luare a deciziilor de afaceri care să genereze profit.
- ☐ transmiterea de pachete de la sursă la destinație
- ☐ definirea modului prin care vor dialoga clientul și serverul

3. Trenduri în business intelligence sunt:

- ☐ Dependența de platformă, Structurarea riguroasă a documentelor, Hipertext
- ☐ Inteligență artificială, BI colaborativ, BI încorporat, Cloud Analytics
- ☐ Hipertext, Independența de platformă, Structurarea riguroasă a documentelor
- ☐ Interferența cu platforma , Structurarea riguroasă a serverelor, Hiperdocument

Bibliografie

[Stanca_07] Stanca L., Suportul tehnologic al afacerilor electronice, Editura Presa Universitară Clujeană, 2007 pag:5-16, Biblioteca Facultății

[Stanca_04] Stanca L., Suportul economic al afacerilor electronice, Editura Presa Universitară Clujeană, 2004 pag:5-20, Biblioteca Facultății

[Buraga_05] Buraga S, Proiectarea site-urilor WEB-Design și funcționalitate, Editura Polirom, 2005, Biblioteca facultății, pag.57-90

MODULUL 2. Arhitectura sistemelor de BI

3.1.Scopul și obiectivele modului

Scopul acestui modul este de a prezenta toate noțiunile teoretice și practice necesare pentru învățarea principiilor de bază în lucrul data warehouse. Conceptele de bază prezentate în cadrul prezentului modul sunt:

- Modalitatea de funcționare a unui Data warehouse
- Data warehouse -definiție și proiectare
- Procesare Analitică Online
- Operații eliminare, transformare încărcare

Obiectivele modului sunt:

1. Familiarizarea cu concepte ca Data warehouse, OLAP, ETL, DATA MART;
2. Asimilarea principiului de funcționare a unui Data warehouse
3. Cunoașterea arhitecturii I
4. Operații de eliminare transformare și încărcare
5. Însușirea operațiilor de online analitică process.

Recomandări privind studiul:

1. Studiarea conceptelor prezentate în vederea utilizării lor pe parcursul celorlalte capitole.
2. Asimilarea conceptelor multidimensional data legate de conceperea, proiectarea și implementarea data warehouse care sta la baza dezvoltării proiectelor de business intelligence prezentate în capitolele următoare

Rezultate așteptate:

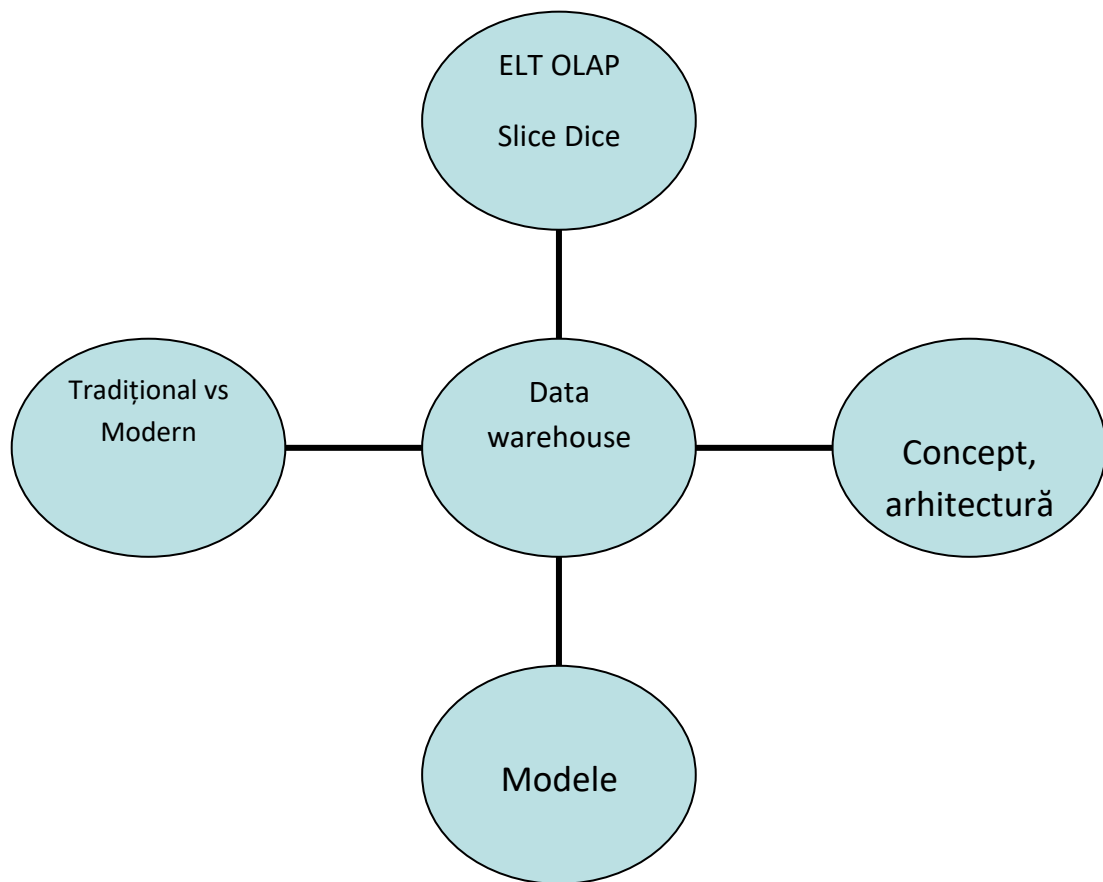
1. Studenții trebuie să fie capabili să își prezinte cunoștințele teoretice de natură coerentă și consistentă.
2. Studenții trebuie să poată proiecta și implementa un data warehouse.

3.2. Scurtă recapitulare a conceptelor prezentate anterior

În cadrul modulelor anterioare s-au prezentat atât noțiuni legate de business intelligence și modul de funcționare al acestuia.

3.3. Schema logică a modului

Prezentarea produsului informatic DataWarehouse a impus structurarea acestui modul ca în figura de mai jos:



3.4. Conținutul informațional detaliat

3.4.1. Data warehouse

Un depozit de date (Amaresan, 2022) este un sistem de business intelligence care reunește volume mari de date din mai multe surse într-un depozit centralizat pentru o organizare, analiză și raportare mai eficiente. Sursele ar putea include instrumente de captare a datelor de site-uri web, achiziții și tranzacții, sisteme de urmărire a inventarului, un sistem de planificare a resurselor întreprinderii (ERP) și software de marketing și vânzări.

Un depozit de date este o sursă de date doar pentru citire, în care utilizatorii finali nu au voie să modifice valorile sau elementele de date. Strategia de arhitectură a depozitului de date a lui Inmon (Inmon, 1999) este diferită de cea a lui Kimball (Kimball, 1996). Modelul de depozit de date al Inmon împarte magazinele de date ca o copie și distribuite ca o interfață între depozitul de date și utilizatorii finali. Kimball vede depozitul de date ca pe o uniune de marturi de date. Depozitul de date reprezintă colecțiile de magazine de date combinate într-un singur depozit central. Ambii considerau că arhitectura de depozit de date dependentă este necesară pentru a îndeplini cerințele utilizatorilor finali de întreprindere în ceea ce privește precizia, sincronizarea și relevanța datelor.

Un depozit de date a devenit mai sofisticat odată cu apariția tehnologiilor cloud. De la nivel scăzut la

nivel înalt, un depozit de date include de obicei o bază de date pentru a păstra datele brute, software pentru extragerea datelor din baza de date, pregătirea acestora și instrumente pentru analiză, raportare și vizualizare a datelor. Într-un depozit de date se pot aplica, tehnici avansate de inteligență artificială și de învățare automată(Site 1). Bill Inmon a definit un depozit de date ca o colecție de date care sprijină procesele de luare a deciziilor (2005). În mod tradițional, un depozit de date este construit pentru a colecta și organiza date istorice de afaceri pe baza cărora pot fi luate decizii. Organizația care deține aceste informații le poate analiza pentru a găsi modele istorice sau conexiuni care să le permită să ia decizii importante de afaceri (Inmon, 2005).

Sarcina principală (Amaresan, 2022)a unui depozit de date este de a sintetiza cantitatea mare de date produse de toate sistemele unei afaceri într-o singură locație accesibilă. Într-un depozit de date, informațiile circulă continuu în timp ce analiștii le examinează. Acest lucru face posibil ca companiile să creeze rapoarte și tablouri de bord care monitorizează și îmbunătățesc continuu funcțiile de afaceri.

Scopul principal (Amaresan, 2022)al unui depozit de date este de a agrega activele de date ale unei companii într-o singură sursă de adevăr pentru analiză și perspective, în loc să solicite analiștilor să urmărească datele din zeci de surse izolate. Beneficiile suplimentare sunt(Site 1):

- Informații mai utile: (Amaresan, 2022) un depozit de date nu numai compilează volume mari de date, dar facilitează analiza și raportarea pentru a ajuta utilizatorii finali să identifice tendințele de ansamblu și să ia decizii bazate pe date.
- Integritate îmbunătățită a datelor: a avea toate datele într-un singur loc este un început bun, dar nu merită prea mult dacă se lucrează cu date inexacte. Un depozit de date lucrează pentru a-și standardiza și curăța datele, deci se lucrează cu informații de încredere. (Amaresan, 2022)
- Date în timp real: depozitele de date se actualizează în timp real. (Site 1)
- Performanță îmbunătățită: depozitele de date sunt concepute pentru a gestiona seturi de date foarte mari fără timp de întârziere sau impact asupra restului infrastructurii tehnologice a unei companii. Depozitele de date pot manipula datele foarte rapid, chiar dacă volumul de date crește.
- Inteligență Artificială: Depozitele de date sunt capabile să integreze inteligența artificială și învățarea automată cu seturile lor de date pentru a descoperi modele ascunse prin data mining. Inteligența artificială este un avantaj competitiv major, deoarece devine mai larg adoptată în business intelligence.

Depozitele de date sunt diferite de bazele de date, ambele stochează date dar scopurile lor diferă semnificativ. Bazele de date sunt structuri care organizează datele în rânduri și coloane, făcând informațiile mai ușor de citit. În comparație cu depozitele de date, bazele de date sunt structuri simple destinate doar stocării. Un depozit de date depășește o simplă bază de date deoarece compilează date din mai multe surse în scopul de a facilita analiza lor. Depozitele de date nu stochează doar date, ci le agregă pentru utilizare pe termen lung în afaceri.

3.4.2. Modele depozite de date

În literatură de specialitate există trei modele de depozite de date:

Depozitele de date ale întreprinderilor(Amaresan, 2022) sunt baze de date centrale în care datele sunt organizate, clasificate și utilizate pentru luarea deciziilor. Aceste sisteme vor eticheta datele și le vor clasifica pentru un acces mai ușor. Deci, un depozit de date pentru întreprinderi este ideal pentru procesul de luare a deciziilor de afaceri pe termen lung. Un depozit de întreprindere colectează toate informațiile despre subiecte care acoperă întreaga organizație. Oferă integrarea datelor la nivelul întregii întreprinderi, în genereal de la unul sau mai multe sisteme operaționale la furnizorii externi de informații și are un domeniu de aplicare transversal. În general, conține date detaliate, precum și date rezumate și poate varia în dimensiune de la câțiva gigaocteti la sute de gigaocteti, terabytes sau mai mult (Site 1). Un depozit de date de întreprindere poate fi implementat pe mainframe tradiționale, computere superservere sau platforme de arhitectură paralelă. Necesită modelare extinsă de afaceri și

poate dura ani pentru a proiecta și construirea lui.(Site 1)

Depozit de date operaționale (ODS) (Amaresan, 2022) are rol important în analiza activităților zilnice, de rutină deoarece este actualizat în timp real și stochează date specifice unei activități alese.

Data mart (Amaresan, 2022) este o parte a unui depozit de date care susține un anumit departament de afaceri, o echipă sau o funcție. Orice informație care trece printr-un magazin de date este stocată și organizată automat pentru o utilizare ulterioară. De exemplu, un data mart de marketing își poate limita subiecții la client, articol și vânzări. Datele conținute în magazinele de date tind să fie rezumate. Data mart-urile sunt de obicei implementate pe servere departamentale low-cost bazate pe UNIX /LINUX sau Windows(Site 1). Ciclul de implementare a unui data mart este mai probabil să fie măsurat în săptămâni și nu în luni sau ani. Cu toate acestea, poate implica o integrare complexă pe termen lung dacă proiectarea și planificarea acestora nu ar fi la nivelul întregii întreprinderi. Un data mart are aceleași beneficii și funcții ca un depozit de date, doar la o scară mai mică.

(Thilini și Hugh, 2005) și (Eckerson, 2003) oferă o modalitate semnificativă de a vizualiza și analiza arhitectura depozitului de date. Eckerson afirmă că un sistem de depozit de date de succes depinde de procesul de pregătire a bazei de date care derivă date din diferite sisteme integrate de procesare tranzacțională online (OLTP). În acest caz, procesul ETL joacă un rol crucial pentru a face funcțional procesul de punere în scenă a bazei de date. Sondajul asupra factorilor care au influențat selecția arhitecturii depozitului de date de către (Thilini, 2005) identifică cinci arhitecturi de depozit de date care sunt utilizate în mod obișnuit:

1. (Kimball, 1996) a fost pionier în proiectarea și arhitectura depozitului de date cu uniuni de marturi de date cunoscute sub numele de arhitectură de magistrală sau depozit de date virtual. Arhitectura autobuzului permite magazinelor de date nu numai să fie situate într-un singur server, dar pot fi și amplasate pe un server diferit. Acest lucru permite depozitului de date să funcționeze mai mult în modul virtual și să combine toate magazinele de date și să proceseze ca un singur depozit de date.
2. (Inmon, 1999) a dezvoltat arhitectura hub and spoke. Hub-ul este serverul central care se ocupă de schimbul de informații și transformarea datelor mânerului spițelor pentru toate depozitele regionale de date de operare. Hub and speak s-a concentrat în principal pe construirea unei infrastructuri scalabile și care poate fi întreținută pentru depozitul de date.
3. Arhitectura depozitului central de date construită pe baza arhitecturii hub-and-spoke, dar fără componenta dependentă de data mart. Această arhitectură copiază și stochează date operaționale și externe eterogene într-un depozit de date unic și consistent. Această arhitectură are un singur model de date care este consecvent și complet din toate sursele de date. Potrivit (Inmon, 1999) și (Kimball, 1996), depozitul central de date ar trebui să fie constituit din punerea în scenă a bazei de date sau cunoscut ca depozit de date operaționale ca o etapă intermediară pentru procesarea operațională a integrării datelor înainte de transformarea în depozitul de date.
4. (Hackney, 2000) demonstrează că depozitul de date federat este o integrare a mai multor marturi de date eterogene, staging al bazei de date sau depozit de date operaționale, combinație de aplicații analitice și sisteme de raportare. Conceptul de concentrare federală pe cadrul integrat pentru a face depozitul de date mai fiabil. (Jindal, 2004) concluzionează că depozitele de date federate reprezintă o abordare practică, deoarece se concentrează pe o fiabilitate mai mare și oferă o valoare excelentă.
5. data mart independent este cel mai bine utilizat ca un depozit de date ad-hoc. și, de asemenea, să fie folosit ca prototip înainte de a construi un depozit de date real. Conform lui (Thilini și Hugh, 2005), fiecare unitate organizațională are tendința de a-și construi propria bază de date, care funcționează ca un magazin de date independent.

Conform lui (Thilini și Hugh, 2005), (Eckerson, 2003) și (Mailvaganam, 2007) din cele 5 arhitecturi 2 sunt cele mai populare mai exact hub-and-spoke propus de (Inmon, 1999), depozitul de date este sursa de date care ajunge în diferite Data Marts. Depozitul de date asigură integritatea și coerența datelor în întreaga organizație, iar în al doilea rând este arhitectura magistralei de date mart propusă de (Kimball, 1996), care spune că depozitul de date nu este altceva decât unirea tuturor data marts”. Aplicate corect, ambele strategii conduc la o implementare bună de Data Warehouse.

Până în prezent, numai experții în baze de date au putut să acceseze date pentru a crea

interogările complicate necesare pentru a prelua, formata și rezuma informațiile pentru a fi utilizate de către analiști și factorii de decizie manageriali. Pe măsură ce nivelurile inferioare de management sunt mai mult implicate în procesul de luare a deciziilor, nevoia de depozit de date a crescut pentru accesul direct al utilizatorilor finali la datele din depozitul de date de către persoane cu expertiză limitată în bazele de date. Un depozit de date poate oferi un singur depozit pentru toate datele, datele sale sunt stabile în timp, iar DW permite partajarea acelorași date între diferite unități de afaceri (Reeves, 2009). Datele sunt încărcate în intervalul de timp programat de la furnizorii de date.

Structura Data Warehouse este concepută special pentru un timp de răspuns rapid, ceea ce este potrivit ca instrument de raportare. Chiar și rapoartele care solicită un număr mare de rânduri sunt de obicei returnate în câteva minute (Reeves, 2009). Un depozit de date oferă informații de bază și feedback la momentul potrivit prin diverse instrumente de business intelligence. Un depozit de date este adesea o componentă de bază a unei infrastructuri de Business Intelligence în cadrul unei organizații. BI este instrumentul pentru a simplifica descoperirea și analiza informațiilor, precum și pentru a furniza informații utile factorilor de decizie corespunzători în intervalul de timp necesar pentru a sprijini luarea eficientă a deciziilor.

2.4.3. Model multidimensional

Depozitele de date care acționează ca o componentă a sistemelor de afaceri inteligente sunt utilizate pentru a furniza date necesare pentru măsurarea indicatorilor cheie de performanță într-o configurație organizațională. Prin urmare, depozitele trebuie să fie proiectate într-o manieră care să îndeplinească aceste cerințe corporative. De exemplu, în literatură se precizează că există o preocupare față de modelul multidimensional ca cel mai bun concept aplicat în multe depozite de date. Modelul multidimensional este astfel analizat folosind diverse idei pentru a determina viabilitatea și aplicabilitatea acestuia în contextul modelării depozitului de date. Proiectarea acestui model este analizat urmând mai mulți pași ai procesului de proiectare, cum ar fi logica, specificația cerințelor, conceptuală și fizică. Designul conceptual este privit ca nu are aplicabilitate în proiectarea modelului multidimensional, deoarece nu este acceptat pe scară largă. Ca urmare, modelul logic este utilizat pe scară largă în scopuri de modelare, deși duce la mai multe scheme care fac dificilă înțelegerea de către un utilizator obișnuit. Modelul logic reprezintă astfel modelul multidimensional în tabele care sunt relaționale și dispuse în structuri particulare precum schemele fulgilor de zăpadă și schemele stelare (Devlin și Cote, 1996). Prin urmare, aceste scheme arată o relație între diverse tabele dimensionale și tabele de fapte. Schemele stele sunt văzute pentru a utiliza mai multe tabele diferite care au ca rezultat tabele dimensionale denormalizate, în timp ce fulgul de zăpadă este considerat că folosește tabele normalizate pentru ierarhiile și dimensiunile lor. În această reprezentare relațională de depozit de date, un server cunoscut sub numele de OLAP oferă un cub de date care, în schimb, aduc o vedere asupra modelului multidimensional al unui depozit de date.

Proiectarea fizică a modelului multidimensional se referă la aspecte de implementare. Odată cu furnizarea unei dimensiuni tipice a unui depozit de date, proiectarea fizică este o componentă esențială pentru a se asigura că există suficient timp pentru a răspunde la interogări hoc și complexe care trebuie să fie acceptate. Mai multe tehnici sunt astfel utilizate pentru a îmbunătăți performanța unui sistem care include partiționarea datelor, vederi materializate și indexare. În special, indecșii utilizați în mod obișnuit sunt indici bitmap care sunt utilizați în principal în contextul depozitării datelor, comparativ cu arborii B care sunt utilizați în depozitarea operațională (Sahama și Croll, 2007). Diferența dintre depozitarea operațională și cea de date este că depozitele de date se găsesc de obicei la sfârșit și, prin urmare, acest factor ajută la distingerea celor două, oferind îndrumări relevante în timpul construcției acestor depozite. De asemenea, în depozitele de date datele sunt de obicei colectate din diverse surse; prin urmare datele pot fi modelate după ce sunt primite și transformate pentru a se potrivi cu modelul depozitului și apoi stocate în depozit. Astfel, acest proces este denumit ETL (extracție, transformare și încărcare) care ajută la analiza modelului de depozit de date folosind proiectarea multidimensională.

ETL este o componentă esențială a unui depozit de date care ajută utilizatorii să se asigure că datele sunt curate, deoarece sunt obținute din mai multe sisteme OLTP. Deoarece depozitul colectează date din diverse sisteme, este evident că există o mulțime de dificultăți în sortarea datelor și stocarea lor într-o locație centrală și astfel procesul ETL devine esențial în asistarea procesului de sortare (Kesner & Russell, 2009). Procesul de extracție implică multă curățare și curățare a datelor care asigură validarea tuturor datelor, în timp ce procesul de transformare implică transformarea multiplă a datelor, astfel încât să îndeplinească standardele și cerințele de date din depozit, iar ultima metodă de încărcare implică stocarea datelor transformate. Procesul de transformare asigură, de asemenea, integritatea datelor făcând astfel datele stocate în depozitele care utilizează acest model fiabile pentru utilizatori. De asemenea,

ETL ajută la exportul și importul de date operaționale complexe între încorporare și arhitectura de legătură orientată pe obiecte, prin care datele sunt transformate pentru a asigura stocarea datelor validate în depozit. Datele care sunt stocate sub formă de schemă stea sunt văzute ca fiind formate din tabele de fapte și dimensiuni. Prin urmare, procesul ETL face din modelul multidimensional un design practic pentru depozitele de date, deoarece asigură eficiența indicatorilor cheie de performanță. Conform lui Devlin (1996), în ciuda faptului că ETL este o componentă semnificativă pentru asigurarea integrității datelor, este de asemenea fundamental ca problema scalabilității și complexității să fie luată în considerare, deoarece joacă un rol critic în deciderea tipurilor de date. Să fie depozitate în depozit folosind arhitectura depozitului. Cea mai bună modalitate de a obține o soluție necomplexă și scalabilă este prin adoptarea „arhitecturii hub-and-spoke” pentru procesul de ETL. Această arhitectură ajută procesul ETL să funcționeze în mod flexibil și eficient. Un design centralizat pentru depozitul de date ajută la influențarea întreținerii.

Modelarea dimensională este o tehnică de proiectare a bazelor de date care sprijină utilizatorii de afaceri să interogheze date într-un sistem de depozit de date. Modelarea dimensională este dezvoltată pentru a fi orientată spre îmbunătățirea performanței interogărilor și ușurința în utilizare. Abordarea modelării dimensionale, la nivel logic, poate fi aplicată oricăror forme fizice, cum ar fi bazele de date relaționale și multidimensionale. În modelarea dimensională, există două concepte importante: (site_multid)

- Faptele sunt măsurători de afaceri, valori numerice care ar putea fi agregate. de exemplu, numărul de produse vândute pe trimestru.
- Dimensiunile se numesc contexte sunt descriptori de afaceri care specifică faptele, de exemplu, numele produsului, marca, trimestrul etc.

Modelul de date dimensionale este construit pe baza unei scheme stea cu un tabel de fapte în centru înconjurat de un număr de tabele de dimensiuni. Proiectarea modelării dimensionale presupune realizarea următorilor pași: (site_multid)

- Selectarea procesului de afaceri procesul de afaceri reprezintă o listă de activități zilnice desfășurate într-o firmă, susținute de un sistem de tranzacții online (OLTP) sau un sistem sursă. În acest pas, se adunăm cerințele de la utilizatorii de afaceri pentru a selecta procesul de afaceri sau sursa de măsurare pentru model. Exemple procesarea comenzilor, expedierile, achiziționarea de material etc.
- Declara granularității exprimă nivelul de detaliu asociat cu faptele din tabelul de fapte.
- Identificare dimensiuni în acest pas, adăugarea unui număr de dimensiuni care reprezintă toate descrierile posibile care pot lua valori individuale în contextul fiecărui fapt din tabelul de fapte. Data, ora, produsul, clientul, magazinul etc., sunt câteva exemple bune de dimensiuni comune.
- Identificare Fapte se selectează faptele numerice care se încarcă în tabelul de fapte, pentru a identifica faptele, trebuie identificate KPI-urile procesului de afaceri sau ce se încercă să se măsoare.

Arhitecturi pe care se structurează un depozit de date (colecție de tabele, vederi, indecsi, sinonime):

- Schema stea, cel mai des utilizat model de organizare al depozitelor de date. Tabela de fapte cuprinde, fără redundanțe, marea parte a datelor iar aceasta este conectată la tabelele dimensiune pe baza cheilor externe pe care acestea le conțin. În cadrul schemei apare ideea de star join ce este o legătură stabilită între un tabel de fapte și tabelele dimensiune respective star query ce este

- jonctiunea dintre un tabel de fapte și mai multe tabele dimensiune. Avantajul acestei scheme este generat de obținerea de performanțe optime pentru interogările dintr-un depozit de date
- Schema fulg de zăpadă “seminormalizat” obținând astfel avantajele modelului relațional. Tabelele dimensiune respectă regulile de normalizare din modelul relațional deci, se realizează o economisire de spațiu dar nu conduce la reducerea spațiului pentru tabela de fapte. Avantajele acestui model sunt redundanță redusă respective este ușor de întreținut. Dezavantajul acestei scheme este că la cereri de interogare complexe(join) se crește timpul de răspuns
 - Schema constelație de fapte sau schema galaxie, mai multe tabele de fapte, conectate ce utilizează aceleași tabele dimensiune. Aranjamentul tabelor de date și al tabelor de dimensiuni arată ca o colecție de stele în modelul schemei Galaxy. Dimensiunile partajate din acest model sunt cunoscute sub numele de dimensiuni conforme. Acest tip de schemă este utilizat pentru cerințe sofisticate și pentru tabele de date agregate care sunt mai complexe pentru a fi acceptate de schema Star (sau) schema Snowflake. Această schemă este dificil de întreținut datorită complexității sale.
 - Schema Stea (site agile) conține mai multe tabele de dimensiuni. O schemă stea cu mai puține tabele de dimensiuni prezintă mai multă redundanță. Această schema este baza pentru a proiecta o schemă de cluster stelar și câteva tabele de dimensiuni esențiale din schema stelară sunt fulgi de zăpadă și aceasta, la rândul său, formează o structură a schemei mai stabilă. Schema de stele este preferată de instrumentele BI ce permit utilizatorilor să interacționeze cu ușurință cu structurile de tabelare complexe prin interogări complexe. Schema fulg de zăpadă.(starschema) constă dintr-un tabel de fapte înconjurat de tabele cu mai multe dimensiuni care pot fi conectate la alte tabele cu dimensiuni prin relație multi-la-unu. Schema fulg de zăpadă este un fel de schemă stea, totuși este mai complexă decât o schemă stea în ceea ce privește modelul de date. O schemă fulg de zăpadă este concepută din schema stea prin normalizarea suplimentară a tabelor de dimensiuni pentru a elimina redundanța datelor. Prin urmare, în schema fulgilor de zăpadă, în loc să avem tabele cu dimensiuni mari conectate la un tabel de fapte, avem un grup de tabele cu mai multe dimensiuni. În schema fulgilor de zăpadă, tabelele de dimensiuni sunt în mod normal în a treia formă normală (3NF) (starschema).. Schema fulgilor de zăpadă ajută la economisirea spațiului de stocare, dar crește numărul de tabele de dimensiuni.(starschema).

2.4.4. Extract Transform Load

Instrumentele ETL sunt instrumente specializate care se ocupă de eterogenitatea depozitului de date, curățarea și încărcarea. Faza de extracție convertește datele într-un singur format necesar procesării transformării. Prin curățarea datelor, înregistrările duplicate sunt eliminate, consecvențele detectate și sursele de erori găsite în date.

Datele utilizate în procesele ETL pot proveni din orice sursă: o aplicație mainframe, tabele de baze de date, o aplicație ERP, un instrument CRM, internet, un fișier plat sau o foaie de calcul Excel. Zona proceselor ETL este foarte importantă, dar dificilă datorită lipsei unui model formal de reprezentare a activităților ETL care mapează datele primite din diferite surse de date pentru a fi reunite într-un format adecvat pentru încărcare în depozitul de date țintă (Mawilmada 2011). Construcția unui depozit de date, impune utilizarea instrumentul ETL pentru următoarele trei sarcini:

- (1) datele sunt extrase din diferite surse de date,
- (2) procesarea datelor mai exact acțiuni de transformat și curățare
- (3) încărcarea datelor în depozitul de date.

Instrumentele ETL sunt o categorie de instrumente specializate cu sarcini de omogenitate, curățare, transformare și încărcare („Shilakes și Tylman”, n.d.) în depozitul de date.

Bill Inmon (1993) a numit depozitul de date (DW) ca o „colecție de baze de date integrate, orientate pe subiecte, destinate să sprijine procesul de luare a deciziilor”, integrarea acestei colecții de surse se realizează prin utilizarea ETL (Extract , Transformare și Încărcare).

În timpul unui proces ETL, datele sunt extrase dintr-o bază de date OLTP, transformate pentru a se

potrivi cu schema de depozit de date și încărcate în baza de date a depozitului de date (Sharma & Gupta 2012). Multe depozite de date au incorporate, date din sisteme non-OLTP, cum ar fi fișiere text, sisteme vechi și foi de calcul. ETL a fost adesea o combinație complexă de proces și tehnologie care a consumat o parte semnificativă a eforturilor de dezvoltare a depozitului de date și a necesitat abilitățile analiștilor de afaceri, designerilor de baze de date și dezvoltatorilor de aplicații.

Procesul ETL nu este un eveniment unic, deoarece pe măsură ce sursele de date se schimbă, depozitul de date trebuie actualizat. De asemenea, pe măsură ce afacerea se schimbă, sistemul de depozit de date trebuia să se schimbe ca urmare, ETL se schimbă/evoluează. Procesele ETL au fost concepute pentru a ușura modificarea. Un sistem ETL solid, bine proiectat și documentat este necesar pentru succesul unui proiect de depozit de date. Un sistem ETL este alcătuit din trei procese consecutive:

1. Extragere – proces de citire a datelor dintr-o bază de date sursă specificată, urmat de extragerea unui subset de date dorit.
2. Transformare - proces de conversie a datelor extrase/dobândite din forma anterioară în forma în care trebuia să fie, astfel încât să poată fi plasate într-o altă bază de date. Transformarea are loc prin utilizarea regulilor sau a tabelor de căutare sau prin combinarea cu alte date.
3. Încărcare - proces de scriere a datelor în baza de date țintă.

Procesele ETL sunt responsabile pentru extragerea datelor din diverse surse, iar în procesarea încărcării, este posibil ca datele să fie transformate pentru a se potrivi nevoilor afacerii și, în cele din urmă, să le încarce într-un depozit de date țintă (Kimball, 2008). Datele utilizate în procesele ETL pot proveni din sistemele de baze de date vechi și de tranzacții și pot fi transformate în informații organizate într-un format ușor de utilizat pentru a încuraja analiza datelor și a sprijini luarea deciziilor de afaceri bazate pe fapte (Kimball, 2008). Odată ce datele au fost extrase și convertite în formatul așteptat, setul de reguli de afaceri este aplicat în etapa de transformare a procesului ETL. Transformarea datelor poate include diverse operațiuni filtrarea, sortarea, agregarea, unirea datelor, curățarea datelor, generarea de date calculate pe baza valorilor existente, validarea datelor etc (Kimball, 2008). Pasul final al ETL implică încărcarea datelor transformate în depozitul de date. Există o mulțime de instrumente ETL, inclusiv IBM DB2 Warehouse Manager, Oracle Warehouse Builder și Microsoft cu serviciile lor de integrare SQL Server.

Aplicațiile cu baze de date tranzacționale pot îndeplini funcții de rutină de procesare a datelor, baza de date tranzacțională nu este capabilă să facă procesări analitice, din următoarele motive (Kimball, 2005):

- Bazele de date tranzacționale conțin doar date brute, astfel încât viteza de procesare este semnificativ mai mică.
- Bazele de date tranzacționale sunt concepute pentru procesarea solicitărilor utilizatorilor, cum ar fi preluarea datelor sau actualizarea bazei de date, nu sunt concepute pentru analize, cum ar fi interogări, rapoarte și analize.
- Repetarea aceleiași analize pe datele live duce la un rezultat diferit față de rularea anterioară, deoarece datele sunt actualizate continuu.

2.4.5. Procesare Analitică Online

Procesarea analitică online (OLAP)¹ este o categorie de software care permite utilizatorilor să analizeze informații din mai multe sisteme de baze de date în același timp. Este o tehnologie care permite analiștilor să extragă și să vizualizeze datele de afaceri din diferite puncte de vedere. Bazele de date OLAP² sunt împărțite în unul sau mai multe cuburi. Cuburile sunt proiectate în așa fel încât crearea

¹ <https://www.guru99.com/online-analytical-processing.html>

² <https://www.guru99.com/online-analytical-processing.html>

și vizualizarea rapoartelor devin ușoare. OLAP înseamnă procesare analitică online.

Consiliul OLAP (1997) definește OLAP ca un grup de sisteme de sprijinire a deciziilor care facilitează accesul rapid, consecvent și interactiv la informații care au fost obținute prin reformulate, transformate și rezumate din setul de date relaționale, în principal din depozitul de date, care permit regăsirea optimă a datelor și realizarea tendinței.

În 1993, Codd a introdus termenul de procesare analitică online (OLAP) și a propus 12 criterii pentru a defini o bază de date OLAP. Pe baza criteriilor OLAP, OLAP este o tehnologie optimizată pentru interogări complexe ad-hoc, rezumate la nivel înalt de „detaliere” prin niveluri de detaliu și analiză interactivă prin dice și slice pentru a obține vizualizări diferite ale datelor. OLAP permite, de asemenea, accesul rapid la date și executarea rapidă a interogărilor de baze de date complexe în timp real (Wrembel și Koncilia, 2007).

Conform lui Chaudhuri (1997), Burdick, D. et al. (2006) și Vassiladis, P. (1999), OLAP este un concept de analiza strategică a bazelor de date, cu capacitatea de a analiza cantități mari de date pentru extrage informații valoroase. OLAP permite descoperirea modelelor și a relațiilor conținute în activitatea de afaceri prin interogarea a tone de date din mai multe sisteme surse de baze de date simultan (Nigel. P., 2008). Procesarea informațiilor bazei de date folosind OLAP necesită un server OLAP pentru a organiza și transforma și construi multidimensional data base. Multidimensional data base sunt separate în cuburi pentru ca instrumentele OLAP ale clientului să efectueze analize de date care urmăresc să descopere noi relații de tipar între cuburi. Unele programe software pentru server OLAP populare includ Oracle (C), IBM (C) și Microsoft (C). Madeira (2003) susține faptul că OLAP și depozitul de date sunt tehnologii complementare care se îmbină. Depozitul de date stochează și gestionează datele, în timp ce OLAP transformă seturile de date din depozitul de date în informații strategice. Funcția OLAP variază de la navigare („slice și dice”), la calcule, analize complexe (seriile de timp și modelarea complex). Pe măsură ce factorii de decizie implementează capabilități OLAP mai avansate, ei trec de la accesul de bază la date la crearea de informații și la descoperirea de noi cunoștințe.

Structura de date pe care OLAP o creează din datele relaționale se numește cub OLAP. Cuburile OLAP pot fi gândite ca o matrice multidimensională. Un administrator ar putea dori să-și analizeze datele de înscriere în funcție: de școală, de specialitate, de geografie etc. Aceste criterii diferite de analiză sunt dimensiunile cubului OLAP (Wrembel și Koncilia, 2007).

OLAP oferă un răspuns rapid în mod constant (Wrembel și Koncilia, 2007). Crearea de noi rapoarte care să răspundă la întrebări este un proces lent și costisitor din cauza multiplelor sisteme implicate și în fiecare sistem informațiile sunt răspândite în mai multe tabele. Wrembel și Koncilia (2007) descriu practica obișnuită la locurile de muncă, astfel încât există administratori în orice organizație care petrec o mulțime de ore în fiecare lună sau trimestru exersându-și abilitățile Excel pentru a uni datele manual din diferite sisteme pentru a crea rapoarte simple. Această muncă este optimizată cu un instrument OLAP. OLAP facilitează dezvoltarea de interogări bazate pe metadata și formulele tip foi de calcul (Wrembel și Koncilia, 2007). Interogările bazate pe metadata (numite interogare MDX) permit o tabulare încrucișată cu titluri de coloane și titluri de rând. Interogările bazate pe metadata permit, de asemenea, imbricarea mai multor straturi de attribute ca titluri de coloană (Wrembel și Koncilia, 2007). Formulele în stil tabelar oferite de OLAP sunt similare cu foaia de calcul Excel, dar sunt mai puternice decât o simplă formulă de tabel.

Sistemele OLAP au fost clasificate în mod tradițional folosind următoarea taxonomie (Nigel Pendse, 2008)

- MOLAP (procesare analitică multidimensională online) este forma clasică a OLAP ce stochează date într-un spațiu de stocare multidimensional optimizat, mai degrabă decât într-o bază de date relațională.
- OLAP relațional (ROLAP) funcționează direct cu baze de date relaționale și nu necesită pre-calcul. Datele de bază și tabelele de dimensiuni sunt stocate ca tabele relaționale și sunt create noi tabele pentru a păstra informațiile agregate. ROLAP se bazează pe manipularea datelor stocate în baza de date relațională pentru a da aspectul funcționalității OLAP tradiționale de feliere și decupare. În esență, fiecare acțiune de tăiere și tăiere este

echivalentă cu adăugarea unei clauze „WHERE” în instrucțiunea SQL. Instrumentele ROLAP nu folosesc cuburi de date precalculate, ci interogarea este realizată direct către baza de date relațională standard. Instrumentele ROLAP prezintă capacitatea de a pune orice întrebare, datorită faptului că metodologia nu se limitează la conținutul unui cub. ROLAP are, de asemenea, capacitatea de a accesa până la cel mai mic nivel de detaliu din baza de date.

- OLAP hibrid (HOLAP) este un compromis nedorit între suplimentar ETL și performanța lentă a interogării au asigurat că mai mulți instrumente OLAP comerciale acum o abordare „OLAP hibrid” (HOLAP), care permite proiectantului modelului să decidă în ce porțiuni a datelor vor fi stocate MOLAP și care porțiuni în ROLAP.

2.1.6. Tradițional vs Cloud Data Warehouse~concluzie

O arhitectură de depozit de date bazată pe cloud este concepută pentru a aborda limitările bazelor de date tradiționale. Trecerea la un depozit de date cloud va oferi unei întreprinderi oportunitatea de a profita de multe dintre beneficiile ale cloud-ului pentru gestionarea datelor. Arhitecturi tradiționale de depozit de date includ diferite abordări arhitecturale, design, modele, componente, procese și roluri care influențează eficiența arhitecturii. Pe baza informațiilor anterioare rezultă că abordarea arhitecturii pe trei niveluri este una dintre cele mai frecvent întâlnite abordări ale depozitării de date la nivel tradițional. Cele trei niveluri include: (Sitet)

- Nivelul de jos alcătuit din serverul de bază de date real folosit pentru a elimina datele din sursele de origine.
- Nivelul mediu are un server pentru procesarea analitică online (OLAP) care este responsabil pentru transformarea datelor.
- Nivelul superior similar cu un strat de interfață cu utilizatorul. Acesta constă în instrumente pentru analiza obișnuită a depozitării de date, cum ar fi raportarea.

Două dintre cele mai frecvent utilizate abordări ale designului de depozitare de date au fost create de Ralph Kimball și Bill Inmon. Abordarea lui Inmon (Inmon, 1999) este considerată de sus în jos, tratează depozitul ca un depozit centralizat pentru toate datele unei organizații. Odată ce există un model de date centralizat pentru acel depozit, organizațiile pot folosi magazine de date dimensionale bazate pe acel model. Data mart-urile sunt depozite pentru linii de afaceri individuale.

Abordarea lui Kimball (Kimball, 2005) se bazează pe o metodă de jos în sus, în care platformele de date sunt principalele metode de stocare a datelor. Depozitul de date este practic o colecție a acelor magazine de date care permite lucrări de analiză uniforme, raportare și alte elemente esențiale de business intelligence. Modele tradiționale de arhitectură de depozit de date se bazează pe una din următoarele: (Sitet)

- Depozitul virtual de date: se bazează pe depozitul care funcționează ca centru al activelor de date ale unei organizații. Acesta integrează datele din fiecare linie de afaceri pentru un acces ușor în întreaga întreprindere.
- Data mart: Subliniază datele unităților de afaceri individuale pentru analiză și raportare. Aceasta implică agregarea datelor din mai multe surse pentru un domeniu de interes, cum ar fi marketingul.
- Depozitul de date pentru întreprinderi

Toate depozitele de date au un strat de utilizator pentru sarcinile specifice de analiză a datelor sau de extragere a datelor. Sursele de date diferite, conțin în mare parte aceleași tipuri de date, pot fi introduse în structura depozitului de date și analizate direct prin stratul utilizator. Datele sunt modificate într-un format structurat rezumat, astfel încât să poată fi analizate holistic la nivelul utilizatorului. În funcție de cazul de utilizare în afaceri, această structură ar putea să nu fie necesară dacă organizațiile analizează doar date de tipuri similare. Data mart-urile sunt utile pentru a găzdui datele rezumate ale unei anumite linii de afaceri pentru interogări foarte specifice. De exemplu, echipele departamentului de marketing ar putea accesa această structură de date pentru analize predictive detaliate ale vânzărilor din diferite locații.

Arhitectura de depozit de date bazată pe cloud înseamnă că depozitele de date reale sunt accesate prin cloud. Există mai multe opțiuni de depozite de date bazate pe cloud, fiecare dintre ele având arhitecturi diferite pentru aceleași beneficii de integrare, analiză și acțiune asupra datelor din surse diferite. Abordare bazată pe depozit de date în cloud diferă de abordările tradiționale prin: (Sitet)

1. Costuri inițiale sunt impuse de diferitele componente necesare pentru depozitele tradiționale de date locale în vreme ce componentele arhitecturii cloud sunt accesate prin cloud, aceste cheltuieli nu se aplică.
2. Costuri continue sunt costuri de actualizare și întreținere pentru arhitecturi tradiționale, cloud-ul oferă un model redus, cu plata pe măsură.
3. Viteză este adecvată în arhitectura de depozit de date bazată pe cloud.
4. Flexibilitatea este prezentă în depozitele de date cloud ce sunt concepute pentru a ține cont de varietatea de formate și structuri găsite în big data în timp ce în tradițional sunt concepute pur și simplu pentru a integra date structurate similar.
5. Scala este facilitată de resursele elastice ale cloud-ului, ideală pentru scara necesară pentru seturile mari de date, opțiunile de stocare a datelor bazate pe cloud reduc dacă este necesar, ceea ce este dificil de realizat cu alte abordări.

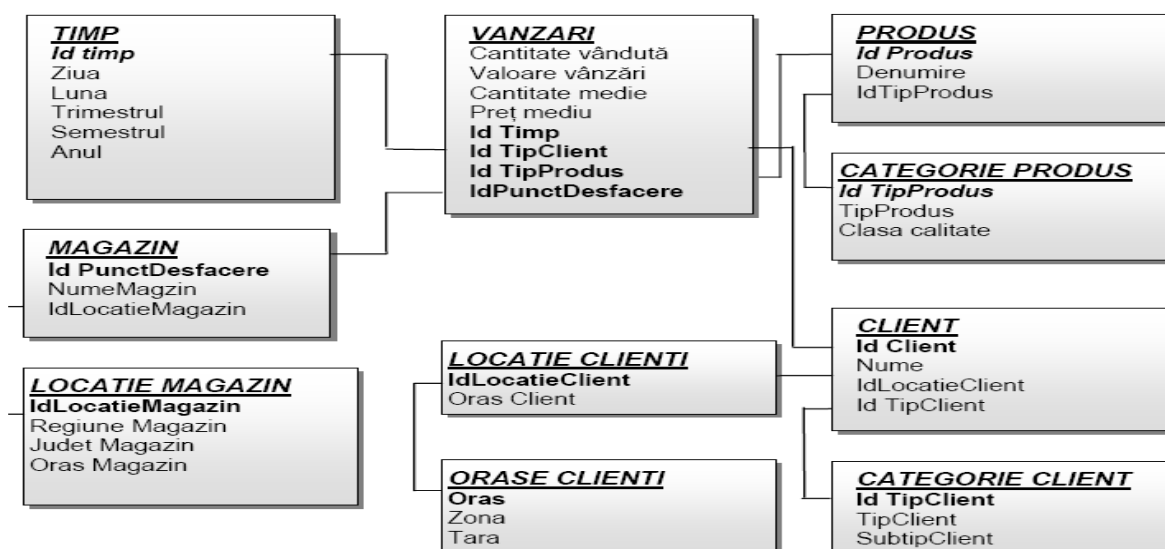
Unele dintre cele mai notabile depozite de date cloud de pe piață includ Amazon Redshift, Google BigQuery, Snowflake și Microsoft Azure SQL Data Warehouse. (Sitet)

Redshift utilizează arhitectura de procesare masiv paralelă în care nodurile stochează date în felii printr-un format de coloană. Fiecare nod are propria memorie, memorie RAM și putere de calcul. Principalele tipuri de noduri sunt nodurile lider și de calcul; primul primește interogări și le atribuie nodurilor de calcul pentru a efectua interogările. Nodurile de calcul pot procesa date în diferite secțiuni în același timp, Redshift are performanțe robuste de interogare. Nodurile de calcul returnează rezultatele la nodurile lider, care le adună pentru aplicații de la partea clientului. Utilizatorii se pot conecta direct la Redshift cu un sortiment de instrumente de BI sau de analiză pentru a interoga datele direct acolo unde locuiesc. Google BigQuery se bazează pe o arhitectură fără server în care furnizorul folosește diferite mașini pentru a gestiona resursele. Arhitectura BigQuery acceptă atât încărcarea tradițională a datelor, cât și fluxul de date, cea din urmă fiind concepută pentru a ingera date în timp real.

În concluzie, arhitectura tradițională a bazelor de date poate să fie folosită cu succes în scenario în care se lucrează cu integrări strânse ale unor tipuri de date structurate similar mai mult arhitectura locală este costisitoare de întreținut, nu funcționează la viteza și flexibilitatea necesare pentru seturile de date moderne în era actuală a datelor mari. În timp ce arhitectura de depozit de date bazată pe cloud este proiectată pentru scalabilitatea extremă a nevoilor actuale de integrare și analiză a datelor. Nu numai că produce beneficii semnificative de performanță și integrare, dar depozitele de date în cloud sunt mult mai rentabile, scalabile și flexibile pentru varietatea de formate de date utilizate de organizații în prezent. În cele din urmă, arhitectura de depozit de date bazată pe cloud este cea mai eficientă utilizare a resurselor de depozit de date. Cantități mari de date disponibile în mediul de business, a determinat dezvoltă conceptul de Data Lakes, care au ca scop stocarea și analiza Big Data. Apariția unor noi soluții de acest tip în depozitare nu le înlocuiește pe cele existente, ci doar umple golul. Coexistența și integrarea sistemelor de stocare a datelor (atât structurate, cât și nestructurate) în organizație permite controlul haosul și obținerea de cunoștințele necesare în afaceri.

Test de verificare a cunoștințelor:

1. Ce este Data warehouse? Definiți acest concept.
2. Enumerați câteva facilități oferite de datawarehouse.
3. Enumerați principalele motive pentru folosirea pe scară largă a datawarehouse.
4. Explicați mecanismul de funcționare a schemei constelație de fapte sau schema galaxie.
5. Ce este Procesarea analitică online (OLAP).
6. Enumerați tipurile de date calendaristice cunoscute.
7. Enumerați grupurile în care se împart de șirurile de caractere?
8. Definiți conceptul de fapt.
9. Descrieți Schema Stea
10. Enumerați și explicați tipurile de relații ce se pot stabili între două sau mai multe tabele ale unei baze de date.
11. Creați schema fulg de zăpadă următoare



1. Modificarea datelor unei înregistrări dintr-o tabelă a unei baze de date se face cu comanda :

- ☐ INSERT
- ☐ UPDATE
- ☐ SELECT
- ☐ DELETE

Răspunsuri

1. Modificarea datelor unei înregistrări dintr-o tabelă a unei baze de date se face cu comanda :

- ☐ INSERT
- ☒ UPDATE
- ☐ SELECT
- ☐ DELETE

1. Chan, K.C.C., Course Notes and Lab Manuals for COMP417, 2009. A77
2. Inmon, W.H., Strauss, D., and Neushloss, G., DW 2.0: The Architecture for the Next Generation of Data Warehousing, Morgan Kaufmann, 2008.
3. „Cartea albă a Consiliului OLAP” (PDF). Consiliul OLAP. 1997. Adus 2008-03-18.

4. Golfarelli, M., and Rizzi, S., Data Warehouse Design: Modern Principles and Methodologies, McGraw-Hill, 2009.
5. Rokach, L., and Maimon, O., Data Mining with Decision Trees: Theory and Applications, World Scientific, 2008.
6. Witten, I.H., Frank, E., Hall, M.A., Data Mining, Third Edition: Practical Machine Learning Tools and Techniques, Morgan Kaufmann, 2011.
7. Westphal, C., Data Mining for Intelligence, Fraud & Criminal Detection: Advanced Analytics & Information Sharing Technologies, CRC Press, 2008.
8. Cox, E., Fuzzy Modeling and Genetic Algorithms for Data Mining and Exploration, Morgan Kaufmann, 2005.
9. Liu, B., Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data, Springer, Berlin Heidelberg, 2009.
10. Tsitsis, K., and Chorianopoulos, A., Data Mining Techniques in CRM: Inside Customer Segmentation, Wiley, 2010.
11. Han, J. and Kamber, M., Data Mining: Concepts and Techniques, 2nd Edition, Morgan Kaufmann, 2005
12. Shapiro, A.F., and Jain, L.C., Intelligent and Other Computational Techniques in Insurance: Theory and Applications, World Scientific, 2003
13. Bevanda, V. (2018). Decision Engineering: Settling A Lean Decision Modeling Approach. Varazdin: Varazdin Development and Entrepreneurship Agency (VADA).
14. Eskelinen, T., Räsänen, T., Santti, U., Happonen, A., & Kajanus, M. (2017). Designing a Business Model for Environmental Monitoring Services using Fast MCDS Innovation Support Tools. Technology Innovation Management Review, 7(11), 36-46.
15. Kesner, R. M., & Russell, B. (2009). Enabling business processes through information management and IT systems: The FastFit and winter gear distributors case studies. Journal of Information Systems Education, 20(4), 401-405.
16. Roberts, B., & Koumpis, A. (2004). A framework for situation room analysis and exploration of its application potential in the information technologies market. Management Decision, 42(7), 882-891.
17. Suen, J., Wu, T., & Lin, K. (2018). A System Framework Design for Virtual Reality Game Using Gameplay Big Data Technology. International Journal of Organizational Innovation (Online), 11(2), 230-239.
18. Devlin, B., & Cote, L. D. (1996). Data warehouse: from architecture to implementation. Addison-Wesley Longman Publishing Co., Inc.
19. Sahama, T. R., & Croll, P. R. (2007, January). A data warehouse architecture for clinical data warehousing. In Proceedings of the fifth Australasian Symposium on ACSW frontiers-Volume 68 (pp. 227-232). Australian Computer Society, Inc.
20. Alenazi, S.R.A. et al., 2014. Prototyping Data Warehouse System for Ministry of Higher Education in Saudi Arabia. , 7(4), pp.74–81.
22. Anon, 2015. Architecture For Real-Time Analytical Data Integration Anddata. , (July).
23. Ferreira, N. & Furtado, P., Near Real-Time with Traditional Data Warehouse Architectures : Factors and How-to.
24. Mawilmada, P.K., 2011. Impact Of A Data Warehouse Model For Improved Decision – Making.
25. Oketunji, T. & Omodara, O., 2011. Design of Data Warehouse and Business Intelligence System. ,
26. Parmar, V., Yadav, R. & Sharma, M., 2016. Review Article A Data Cleaning Model For Dataware House. , 4(1).
Quality, D. & Line, B., Data quality.
27. Russom, P., Tdwi Checklist Report Data Integration For Real-Time Data Warehousing And Data Virtualization Tdwi Checklist Report Data Integration For Real-Time Data Warehousing And Data Virtualization.

28. Senapati, R. & Kumar, D.A., 2014. A Survey on Data Warehouse Architecture. , pp.5235–5239.
29. Sharma, N. & Gupta, S.K., 2012. Design And Implementation Of Access The Contents In The. , 6(1), pp.61–64.
30. Simitsis, A. & Vassiliadis, P., A Methodology for the Conceptual Modeling of ETL Processes.
31. Bernstein, P., Rahm, E., 2000. Data warehouse scenarios for model management. In: Proceedings of the
32. 19th International Conference on Conceptual Modeling (ER'00), LNCS, vol. 1920, Salt Lake City, USA, pp. 1–15.
33. Berson and Smith, 1997A. Berson, S.J. Smith Data Warehousing, Data Mining, and OLAP.
34. Dobre, A., Hakimpour, F., Dittrich, K.R., 2003. Operators and classification for data mapping in semantic integration. In: Proceedings of the 22nd International Conference on Conceptual Modeling (ER'03), LNCS, vol. 2813, Chicago, USA, pp. 534–547.
35. El Bastawesy, A., Boshra, M., Hendawi, A., 2005. Entity mapping diagram for modeling ETL processes. In: Proceedings of the Third International Conference on Informatics and Systems (INFOS), Cairo.
36. Muneer Alsurori, Juhana Salim,” Information and Communication Technology for Decision-Making in
37. the Higher Education in Yemen: A Review” 2009 International Conference on Electrical Engineering and Informatics, 5-7 August 2009, Selangor, Malaysia.
38. Wang Aihua, Guo Wenge, Xu Guoxiong, Jia Jiyu, Wen Dongmao, 2009. “GIS-Based Educational 52 Decision- Making System” Proceedings of 2009 IEEE International Conference on Grey Systems and Intelligent Services, November 10-12, 2009, Nanjing, China., 2009 IEEE, pp 1198-1202.
39. Qiusheng Liu, Guofang Liu,” Research on the Framework of Decision Support System Based on ERP
40. Systems”, 2010 Second International Workshop on Education Technology and Computer Science, 2010 IEEE.
41. Manjunath T.N, Ravindra S Hegadi, Ravikumar G K."Analysis of Data Quality Aspects in DataWarehouse Systems", (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 2 (1) , 2010, 477-485.
42. W. H. Inmon. 2012.“Building the Data Warehouse.” John Wiley & Sons, Third edition.
43. Rajan Vohra1 & Nripendra Narayan Das Intelligent decision support systems for admission management In higher education institutes- International Journal of Artificial Intelligence & Applications (IJAIA), Vol.2, No.4, October 2011.
44. Glorio, O., Mazon, J. N., Garrigoz, I., & Trujillo, J. (2010). Using Web-based Personalization on Spatial Data Warehouses (pp. 1-8). EDBT-ACM.
45. Hamdan, A. (2005). Women and Education in Saudi Arabia: Challenges and Achievements. International Education Journal, 6(1), 42-64.
46. Han, J., & Kamber, M. (2006). Data Mining: Concepts and Techniques (2nd ed.). Morgan Kaufmann Publishers.
47. Inmon, W. H. (1992). Building the Data Bridge: The Ten Critical Success Factors of Building a Data Warehouse. Database Programming & Design.
48. Inmon W. H. (1996), "The data warehouse and data mining", Communications of the ACM, 39(11), 49 –50.
49. Mohammed, M. A., Hasson, A. R., Shawkat, A. R., & Al-Khafaji, N. J. (2012). E-Government Architecture Uses Data Warehouse Techniques to Increase Information Sharing in Iraqi Universities (pp. 1-5). IEEE Explore.
50. Wang, F. (2009). Application Research of Data Warehouse and Its Model Design (pp. 798-801). IEEE computer society.
51. Yebai, L., & Li, Z. (2009). Interactive Modeling of Data Warehouse on E-business System (pp. 80-83). IEEE computer society.
52. Yen, D.C., Chou, D.C. and Chang, J. (2002), “A synergic analysis for Web-based enterprise

- resourcesplanning systems”, Computer Standards & Interfaces, 24(4), 337-46.
53. Anne Wanjiku Kibugu.2016, A Methodology For The Implementation Of A Data Warehouse Using An Etl Process Model For Improved Decision Support http://erepository.uonbi.ac.ke/bitstream/handle/11295/98685/Kibugu%20Anne_A%20Methodology%20for%20the%20Implementation%20of%20a%20Data%20Warehouse%20Using%20an%20Etl%20Process%20Model%20for%20Improved%20Decision%20Support.pdf?sequence=1&isAllowed=y
 54. S. Amaresan, 2022, The Plain-English Guide to Data Warehouses [+ Examples], <https://blog.hubspot.com/website/data-warehouse-concepts>
 55. UKEssays. (November 2018). Literature review about data warehouse. Retrieved from <https://www.ukessays.com/essays/computer-science/literature-review-about-data-warehouse.php?vref=1>
 56. (Site 1) https://www.vssut.ac.in/lecture_notes/lecture1428550844.pdf
 57. <https://www.questionpro.com/blog/business-intelligence/>
 58. (site_multid) <https://www.zentut.com/data-warehouse/dimensional-modeling/>
 59. (site agile) <https://ro.myservername.com/devops-practice-based-agile-manifesto>
 60. (starschema) <https://www.zentut.com/data-warehouse/snowflake-schema/>
 61. Adamson, Michael And Venerable, Michael, Data Warehouse Design Solutions, Wiley
 62. Adleman, Sid And Terpeluk Moss,Larissa, [Data Warehouse Project Management](#), Addison-Wesley Information Technology Series, Isbn 0-201-61635-1
 63. Bischoff, Joyce And Alexander, Ted. Data Warehouse: Practical Advice From The Experts, Englewood Cliffs, Nj: Prentice-Hall, 1997.
 64. Brackett, Michael. The Data Warehouse Challenge: Taming Data Chaos, New York: John Wiley & Sons, 1996. Isbn 0-471-12744-2
 65. Dyche, Jill, E-Data: Turning Data Into Information With Data Warehousing, Addison- Wesley English, Larry, Improving Data Warehouse And Business Information Quality: Methods For Reducing Costs And Increasing Profits, John Wiley & Sons
 66. Gill, Harjinder S. And Rao, Prekash, C. The Official Guide To Data Warehousing, Que, 1996. Isbn 0-7897-0714-4
 67. Nigel Pendse „Arhitecturi OLAP”. Raport OLAP. 2008
 68. Ibm, Discovering Data Mining, Prentice Hall (Available Through Red Book Publications)
 69. Inmon, W. H. Using Db2 To Build Decision Support Systems, Wiley, 1993. Isbn 0-471-56778-7
 70. Inmon, W.H. Building The Data Warehouse 3rd Edition, New York: John Wiley & Sons, 2002. Isbn 0-471-08130-2
 71. Inmon, W.H. And Hackathorn, R. D. Using The Data Warehouse, New York: Wiley-Qed,1994. Isbn 0-471-05966-8
 72. Inmon, W.H., Imhoff, C., And Battas, G. Building The Operational Data Store 2nd Edition, New York: John Wiley, 1999. Isbn 0-471-32888-X
 73. Inmon, Zachman, And Geiger, Data Stores, Data Warehousing And The Zachman Framework, McGraw-Hill, 1997
 74. Kimball, Ralph. The Data Warehouse Toolkit, New York: John Wiley & Sons, 1996. Isbn 0-471-15337-0
 75. Mattison, Rob, Web Warehousing & Knowledge Management, McGraw Hill
 76. Poe, Vidette. Building A Data Warehouse For Decision Support, Prentice-Hall: Englewood Cliffs, Nj, 1995. Isbn 0-13-371121-8
 77. Westerman, Paul, Data Warehousing: Using The Wal-Mart Model, August 2000
 78. (Sitet)Modern Data Warehouse Architecture: Traditional Vs Cloud Data Warehouse, <https://www.talend.com/resources/cloud-data-warehouse-architecture/>

MODULUL 3 Analiza vizuală a datelor

3.1. Scopul și obiectivele modului

În cadrul acestui modul se vor prezenta noțiunile teoretice și practice ale analizei vizuale. Conceptele de bază ale analizei vizuale, pe care le vom aborda în cadrul acestui modul sunt :

- Definire concept
- Tablouri de bord, analiza deciziilor
- Decizii dinamice
- Algoritmi de clasificare
- Algoritmi de grupare

Obiective acestui modul sunt:

1. Familiarizarea cu conceptele analizei vizuale.;
2. Asimilarea principiului de funcționare a unui algoritm de data mining
3. Cunoașterea structurilor și modului de funcționare a tablourilor de bord
4. Însușirea modului de lucru cu algoritmi de clasificare și clasterizare
5. Crearea de tablouri de bord complexe

Recomandări privind studiul:

1. Studiarea conceptelor prezentate în vederea utilizării lor în practică.
2. Asimilarea conceptelor de bază legate de conceperea, proiectarea și implementarea unor proiectelor de business intelligence.

Rezultate așteptate:

1. Studenții trebuie să fie capabili să își prezinte cunoștințele teoretice de natură coerentă și consistentă.
2. Studenții trebuie să poată concepe, proiecta și implementa un proiect amplu de pagină web dinamică

3.2. Scurtă recapitulare a conceptelor prezentate anterior

În cadrul modulelor anterioare s-au prezentat noțiuni legate de business intelligence și modul de funcționare al acestuia. Pe baza acestor cunoștințe în cadrul acestui modul se pot prezenta noțiunile de bază și avansate ale analizei vizuale.

3.3. Schema logică a modului

Prezentarea analizei vizuale a impus structurarea acestui modul ca în figura de mai jos:

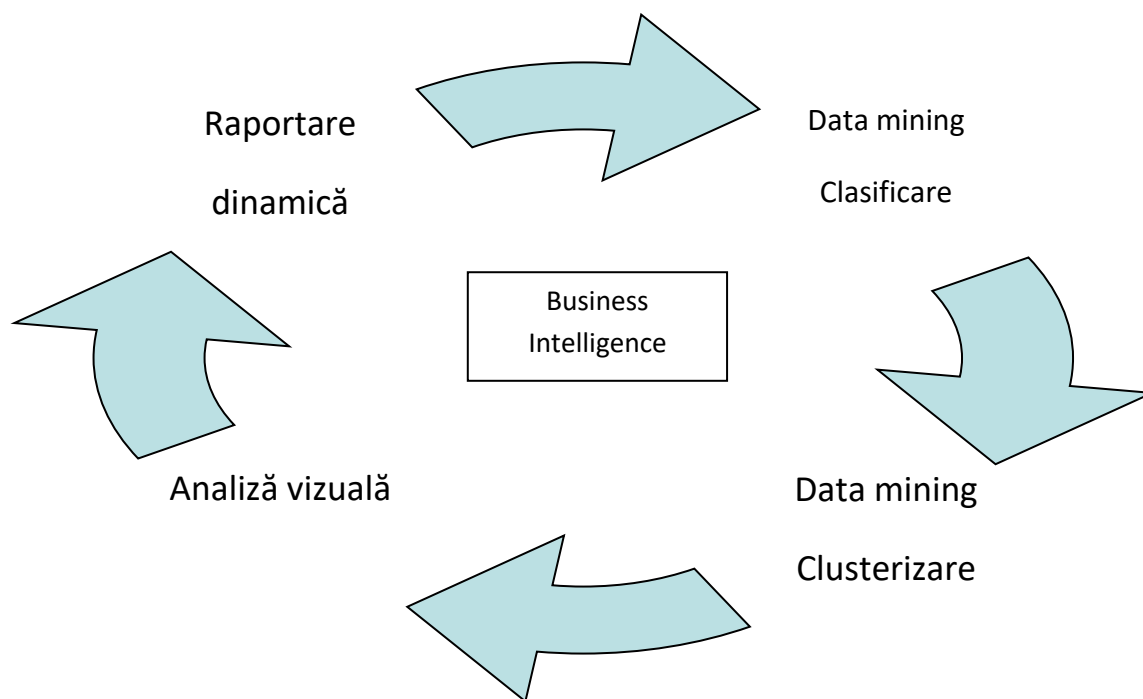


Fig.4.3.1 Structura modului 3

3.4. Conținutul informațional detaliat

3.4.1 Visual Analytics~noțiuni fundamentale

Analiza vizuală este o formă de anchetă în care datele care oferă o perspectivă asupra rezolvării unei probleme sunt afișate într-o manieră interactivă, grafică. Datele pot fi ingerate, analizate și traduse într-o reprezentare vizuală pentru a face interpretarea acestora mai ușoară și mai accesibilă jucătorilor cheie dintr-o organizație, pe baza conceptele cheie ale Visual Analytics. Datele provenite din diferite sisteme sursă pot fi integrate, unificate și standardizate prin aplicațiile de analiză vizuală ce încorporează strategii eficiente de gestionare a lor. Analiza vizuală este deosebit de utilă în aplicațiile care implică seturi de date mari și complexe și procese analitice care necesită un grad înalt de monitorizare și interacțiune -- de exemplu, analize de date mari și utilizări de data mining.

Visual Analytic este știința raționamentului analitic susținută de interfețe vizuale interactive pentru a permite formularea deciziilor bine informate destinate situațiilor complexe. Metodele de analiză vizuală îi ajută pe factorii de decizie în afaceri să-și combine capacitățile umane cu capacitățile de stocare și procesare ale metodelor îmbunătățite pentru a obține o perspectivă asupra problemelor complexe.

Deci, visual analytic(Bose, 2020) s gravitează în jurul ideii de tehnologii de vizualizare a datelor pentru a ajuta oamenii de știință ai datelor și alți profesioniști să identifice tendințele, modelele și relațiile în datele cu care lucrează. Instrumentele software de analiză vizuală sunt

deductibile facilitând ca utilizatorii non-tehnici, să facă analize complexe prin operațiuni simple.

Conform lui (Thomas, Cook,2005) Visual analytics este știința raționamentului analitic susținut de interfețe vizuale interactive. Analiza vizuală poate fi percepută ca o abordare integrată care combină vizualizarea, factorii umani și analiza datelor.

Visual Analytics (Bose, 2020) în contextul vizualizării se referă la domeniile de vizualizare a informațiilor și grafică pe computer și, în ceea ce privește analiza datelor, beneficiază în mare măsură de metodologii de curățare, de gestionare a datelor în scopul de a identifica informații și moduri ideale de reprezentare a vizuală a rezultatelor traduse ca și cunoștințe sub formă de tablouri de bord. Ecranele tabloului de bord pot avea diferite tipuri de motoare care implică grafice vizuale, diagrame circulare sau instrumente infografice, unde, după ce funcționează algoritmi de calcul, rezultatele sunt afișate pe ecran. Scopul (Bose, 2020) unei interfețe de analiză vizuală de tip tablou de bord, este acela de a face ca un utilizator uman să înțeleagă cu ușurință rezultatele și, de asemenea, să facă modificări simultan, care direcționează în continuare procesul algoritmic al deciziei.

Analiza vizuală prin tabloul de bord facilitează înțelegerea colecțiilor masive și în continuă creștere de date de mai multe tipuri(Thomas, Cook,2005)

- furnizează cadre pentru analiza datelor spațiale și temporale
- sprijină înțelegerea informațiilor incerte, incomplete și adesea înșelătoare
- oferă reprezentări ghidate adaptabile utilizatorului și sarcinilor, care permit conștientizarea completă a situației, sprijinind în același timp dezvoltarea acțiunilor detaliate
- suportă mai multe niveluri de abstractizare a datelor și a informațiilor
- facilitează descoperirea cunoștințelor prin sinteza informațiilor, care este integrarea datelor pe baza semnificației lor, mai degrabă decât pe tipul de date original.

Deci, analiza vizuală este știința care utilizează vizualizarea și tehnici de interacțiune pentru a integra raționamentul uman expert în procesul de analiză a datelor(Keim, et al.,2008; (Thomas, Cook,2006). (Sun et al.,2013) a clasificat aplicațiile de analiză vizuală într-un set de categorii, inclusiv spațiu și timp, multivariate, text, grafic și rețea și alte aplicații. Această clasificare diferențiază în mod natural aplicațiile de analiză vizuală de un anumit tip de date sau domeniu de aplicație. Analiza vizuală a fost definită ca „formarea de metafore vizuale abstracte în combinație cu un discurs informațional uman (interacțiune) care permite detectarea așteptărilor și descoperirea neașteptatului în spații informaționale masive, în schimbare dinamică.” Recent, prin combinarea analizei algoritmice a datelor și a vizualizării, analiza vizuală a început să utilizeze vizualizarea ca mediu și interacțiunea ca mijloc de a implica judecata umană în procesul de analiză a datelor (Keim,et al,2009)

Analiza vizuală a fost propusă pentru a transforma supraîncărcarea de informații într-o oportunitate prin crearea de instrumente și tehnici pentru a facilita judecata umană în procesul de extragere a informației. Analiza vizuală, ca abordare integrată combină vizualizarea, analiza algoritmică a datelor, interacțiunea umană cu computerul și raționamentul analitic, ceea ce a atras un interes din ce în ce mai mare dintr-o gamă largă de domenii și discipline. Odată cu dezvoltarea sa, cercetătorii din medii diferite au dat definiții detaliate ale acestuia, cu diferite obiective:

Analiza vizuală este „o metodă de a sintetiza informații și de a obține o perspectivă din date masive, dinamice, ambigue și adesea conflictuale; detectează ceea ce este așteptat și descoperă ceva nou, neașteptat în date, furnizând evaluări în timp util, defensibile și ușor de înțeles capabile să comunice evaluarea eficient” (Thomas, Cook,2006).

Analiza vizuală „combină tehnici de analiză automată cu vizualizări interactive pentru o înțelegere eficientă, a raționamentului în scopul de a sprijini procesul de formare a deciziilor pe baza unor seturi de date foarte mari și complexe (Keim,et al,2008)

Analiza vizuală este un domeniu de cercetare multidisciplinar bazat în principal pe vizualizare, analiza algoritmică a datelor și raționament analitic, care profită de vizualizare și interacțiuni cu instrumente adecvate pentru a integra raționamentul uman în procesul de

extragere a cunoștinșelor pentru a descoperi modele vizuale explicabile (cunoștințe) și pentru a obține informații despre seturi de date mari și complexe. Deci, analiza vizuală are același scop final de cercetare ca EDA(Cui,2019)., mai exact de a identifica cunoștințe și de a identifica perspectiva din seturile de date. Cu toate acestea, analiza vizuală exploatează vizualizarea ca un instrument pentru a integra cogniția umană, abilitățile de percepție și inteligența umană în procesul de analiză a datelor pentru a obține rezultate explicabile. În raport cu vizualizarea, analiza vizuală acordă o prioritate mai mare analizei datelor și descoperirii cunoștinșelor în date, mai degrabă decât doar prezentarea și înțelegerea datelor. Între timp, bazată pe vizualizare, analiza vizuală abordează provocarea de a identifica modelele complexe greu de interpretat într-o manieră intuitivă și semnificativă(Cui,2019).

Analiza vizuală integrează abilitățile cognitive, perceptuale și de raționament într-un proces de analiză pentru a obține o perspectivă din datele care sunt dificil de explorat prin vizualizare pură sau tehnici de analiză. Raționamentul analitic cuprinde diferite tipuri de raționament, cum ar fi deductiv, inductiv și analogic, care se bazează pe o analiză și o evaluare rațională, logică a datelor (Sedig, Parsons;2010) și este facilitat prin crearea de vizualizări și interacțiuni adecvate care maximizează capacitatea umană de a percepe și explora datele. Capacitatea umană adaptează procesele de analiză existente prin integrarea vizualizării și analizei algoritmice a datelor (Keim,2010) pentru a obține informații pe baza cărora se creează cunoștințe prin interacțiuni dinamice.

Teoriile de analiză vizuală trebuie să treacă dincolo de HITL la analitica „human-is the-loop” pentru a integra cogniția umană și proces de raționament cu analiză(Endert, et al. 2014)

Plecând de la mantra lui Shneiderman Privire de ansamblu în primul rând, Filtrare și zoom, Detalii la cerere” (Cui,2019) a subliniat în mod clar rolul vizualizării în procesul de descoperire a cunoștinșelor (Shneiderman,1996) creionând pașii tipici a unui procesul de analiză vizuală astfel:

Pasul 1 Preprocesează (curăță, transformă, integrează) datele pentru a le pregăti pentru procesarea ulterioară.

Pasul 2 Aplicare metode de analiză algoritmică datelor.

Pasul 3 Vizualizarea datele (procesate) cu tehnici de vizualizare adecvate.

Pasul 4 Utilizatorii generează cunoștințe perspicace prin activități de percepție umană, cunoaștere și raționament.

Pasul 5 Utilizatorii fac noi ipoteze și integrează cunoștințele nou generate în analiză și vizualizare prin interacțiuni.

Pasul 6 Regenerați o vizualizare actualizată pe baza interacțiunilor pentru a reflecta înțelegerea datelor de către utilizator.

(Cui,2019) explică mai în detaliu acest proces ca algoritm de lucru astfel: că pasul 1 ar putea fi curățarea datelor, normalizarea, transformarea, gruparea și/sau integrarea datelor eterogene într-o schemă comună. În procesul de analiză vizuală, cunoștințele pot fi obținute la fiecare pas. Cu toate acestea, analiza algoritmică inițială (Pasul 2) și vizualizarea (Pasul 3) a datelor nu sunt adesea suficiente pentru rezolvarea problemelor și luarea deciziilor. În consecință, intervenția umană presupune efectuarea activităților de percepție, cunoaștere și raționament pentru ca în pasul 4 se generează cunoștințe perspicace. Între timp, cunoștințele sunt folosite pentru a formula noi ipoteze. Pe baza acestor noi cunoștințe și ipoteze sunt integrate în procesele de analiză și vizualizare a datelor prin interacțiuni realizat de utilizator. Apoi, algoritmi de analiză a datelor și vizualizările sunt actualizate în funcție de interacțiunile utilizatorului, după prima buclă a procesului de analiză vizuală, se repetă continuu de la Pasul 4 la Pasul 6 până când este suficient. Astfel se obține o perspectivă asupra datelor pentru luarea deciziilor sau rezolvarea problemelor asociate datelor. În unele aplicații de analiză vizuală, Pasul 2 poate fi eliminat, deoarece nu este obligatoriu pentru toate tipurile de date și scenarii. Acest proces iterativ ilustrează bine filosofia „human-is-the-loop” Cunoștințele generate sunt stocate în procesul de analiză vizuală prin bucla de feedback din Pasul 5, ceea ce permite utilizatorului să tragă continuu concluzii mai rapide și mai bune și să obțină o perspectivă din date.

3.4.2. *Decision analysis, Visualisation and visual analytics*

Tabloul de bord ~tool în vizualizarea dinamica a analizei datelor

Începând cu anul 1986 putem vorbi de o literatură de specialitate ce se ocupă cu măsurarea performanței și calității deciziei pe baza analizei datelor economice doar în formă tabelară urmat apoi de prezentarea în format tabelar și grafic. Rezultatele prezentate în literatura sunt contradictorii:

1. autorii Dilla și Steinbart (2005) concluzionează că modul de prezentare a datelor contabile este irrelevant în procesul de luare a deciziei contabile într-un cadru complex BSC.
2. autorii Blocher și colab. (1986), concluzionează că introducerea graficelor în raportarea contabilă are efect benefic în procesul decizional. Blocher și colab. (1986) a arătat că suplimentarea tabelului cu grafice au îmbunătățit considerațiile auditorului într-o sarcină de prognoză. (Cardinaels și van Veen-Dirks, 2010) au constatat că prezentarea măsurilor de performanță afectează modul în care factorii de decizie influențează măsurile financiare și nefinanciare atunci când evaluează performanța bazată pe un BSC. În Hirsch, B., Seubert, A., Sohn, M), se demonstrează că: prezența graficelor în rapoartele de management contabil îmbunătățesc calitatea deciziei contabile dar nu se validează interdependența calității deciziilor și încrederea în luarea deciziilor, și nici obțin diferențe în ceea ce privește performanța decizională în funcție de interacțiunea dintre stilul cognitiv și formatul de prezentare.

Conform lui (Epstein și Manzoni, 1997;) tabloul de bord are scopul de măsurare a performanței. Tabloul de bord a apărut atât din necesitatea combaterii insuficiențelor sistemului contabil cât și din nevoia corelării celorlalte instrumente de gestiune (costuri, bugete), fiind utilizat în special în Franța și în restul Europei încă din perioada interbelică. (Cohen, 1994) considera că tabloul de bord este un instrument de informare și de semnalizare. (Bouquin, 2001) susține că este un instrument suport în luarea deciziilor și în previzionare, tabloul de bord reprezintă un ansamblu de indicatori, construiți pentru a permite managerilor să cunoască stadiul de evoluție a sistemelor pe care le pilotează și să identifice tendințele viitoare. Conform lui (Kerviller și Kerviller, 2000) tabloul de bord are rolul de instrument de diagnostic și de progres, tabloul de bord trebuie să scoată în evidență abaterile semnificative; instrument de dialog între managerii diferitelor niveluri ierarhice; în același timp, este un suport de apreciere obiectivă a performanțelor deoarece permite, în funcție de evoluția rezultatelor, degajarea de concluzii referitoare la acțiunile corective întreprinse și la oportunitatea acestora. În [site1] se afirmă că introducerea dashboardurilor în procesul de formare a deciziei manageriale a avut ca efect apariția unei noi culturi organizatorice, mai exact un management bazat pe obiective clare, măsurabile, asumate de către companie și angajați la fiecare nivel execuțional. Asistăm la descentralizarea deciziei, mai exact acces la informație relevantă la fiecare nivel are impact direct asupra vitezei decizionale. Deciziile se iau la fiecare nivel iar managerul va valida un cumul de decizii. Informația obținută la intervale de timp oferă angajatului posibilitatea de a observa tendințe, probleme care le comunică managerului. Odată cu descentralizarea procesului decizional profilul angajatului se schimbă, necesitând abilități și cunoștințe care până în prezent le aveau doar liderii. Fiecare domeniu al vieții este afectat de introducerea "vizualizării". În prezent, provocările generate de prelucrarea cantitatilor mari de date în timp real au impus noi contexte și provocări. Dorința de a lua decizii cu o acuratețe crescută a generat dezvoltarea rapidă a unui nou domeniu vizual analytics (VA). Visual analytics (VA) caută să pună la dispoziția specialiștilor din diverse domenii instrumente eficiente pentru analiză și înțelegerea seturilor mari de date, permițându-le astfel să decidă și să acționeze în mod corespunzător în timp real. Bazele vizualization theory sunt prezentate în lucrările (Vessey, 1991, 1994; Meyer, 2000). Vessey a realizat un model cognitiv ce îl indică să se utilizeze în studiile care vorbesc despre impactul diferitelor formate de vizualizare în

procesul de luare a deciziilor contabile pe baza rapoartelor contabile. Modelul lui Vessey (Vessey, 1991) reunește trei dimensiuni decision making, information processing, and visualisation research. Modelul lui Vessey (Vessey, 1991). Rezultatele studiilor (Vessey, 1994; Dilla and Steinbart, 2005; Dilla et al., 2010, 2013; Bernhard Hirsch Anna Seubert Matthias Sohn , (2015),") demonstrează: calitatea deciziei contabile crește ca urmare a utilizării formatelor grafice adecvate. În literatura de specialitate s-au obținut rezultate concludente relativ la tema impactului asupra calității deciziei pe baza formatelor de vizualizare adecvate dar conform (Bernhard Hirsch Anna Seubert Matthias Sohn , (2015)) nu sunt remarcate rezultate relevante legate de temele: interrelația calității și a încrederii în procesul de luarea deciziilor; decizii performante în funcție de interacțiunea dintre stilul cognitiv și vizualizarea.

(AICPA, 2014;) afirma ca software-ul folosit pentru a analiza mari volumele de date (adică instrumentele de extragere a datelor), precum și instrumente sofisticate de vizualizare a datelor cresc capacitatea persoanelor de a înțelege povestea pe care datele o prezintă în cercetările empirice sau în viața reală. (Wang și Cuthbertson) afirma ca profesorii trebuie să ofere informația în aria contabilă astfel încât să creeze un scenariu ce să includă dimensiunile: rolul decision analysis în analiza de risc, ce proceduri ar trebui să fie efectuate, implicațiile de testare 100% pe populației, dacă ar trebui utilizate date externe, rolul auditorilor interni în utilizarea decision analysis, interpretarea consecințele utilizării decision analysis. [Brown-Liburd și colab] susțin ca trebuie făcute cercetări astfel încât să se obțină seturile de calificare și caracteristicile ale auditorilor care efectuează activitatea decision analysis pe baza cărora să se obțină modelele mentale, skills-urile, competențele auditorilor în smart era. (Brown-Liburd et al) afirma ca este necesar să se identifice tehnici, metode atât de procesare a cantitatilor mari de date cât și formulare a deciziei contabile corespunzătoare. (Vandervelde și colab. 2008) afirma ca studentul trebuie să cunoască tehnicile, metodele de analiză a datelor economice și din mass-media socială pe baza cărora să poată să modeleze pentru a face predicții legate de factorii care influențează un client în afaceri. Pe baza acestor modele auditorii vor putea planifica și evalua riscul afacerilor. (PwC, 2015) În (PwC, 2015) se precizează ca este necesar ca în curricula contabilă să se introducă cursuri de programare (Python sau Java), structurat și nestructurat baze de date, multivariate și inferențiale statistici (inclusiv limba de programare R), și instrumente de vizualizare a datelor, printre altele dorite aptitudini. În concluzie conform lui (Earley, 2015) studentul secției de contabilitate și audit trebuie să dețină abilități de nivel superior pattern recognition, critical thinking, analytical process, visual analytics, să înțeleagă relațiile dintre finanțe, conturile de declarații, procesele de afaceri și factorii externi pe baza cărora să identifice riscul în afaceri și să fie capabili să înțeleagă modul în care modele de informație financiară pot spune o poveste despre performanța unei companii. În (Vrejoiu_2017) se afirmă că visual analytics este în esență interdisciplinară, combinând diferite domenii de cercetare relaționate, cum sunt: vizualizarea, explorarea datelor (data mining), managementul datelor (data management), fuzionarea datelor (data fusion), statistica și știința cognitivă (cognitive science), interacțiunea om-calculator (human-computer interaction - HCI). În (Vrejoiu_2017) se afirmă că scopul visual analytics este de a transforma supraîncărcarea cu informație într-o oportunitate, utilizând vizualizarea pentru a face modul de procesare a datelor și informațiilor transparent pentru un discurs analitic. Visual analytics facilitează astfel evaluarea constructivă, corecția și îmbunătățirea rapidă a proceselor și modelelor și, în final, extragerea cunoștințelor din date și luarea deciziilor bazate pe acestea. Tehnici de vizualizare eficace și eficiente au făcut posibilă integrarea utilizatorului în procesele de descoperire a cunoștințelor și de explorare a datelor dezvoltându-se explorarea vizuală a datelor (visual data mining). În literatura de specialitate au apărut un număr ridicat de definiții a lui visual data dintre care am decis ca cea enunțată în (Keim_Andrienko_et al) se potrivește cel mai bine cercetării noastre realizată în aria auditului și contabilității „visual analytics combină tehnici de analiză automată cu vizualizări interactive pentru înțelegerea efectivă, elaborarea de raționamente și luarea deciziilor pe baza seturilor de date masive și complexe”. Conform lui (Vrejoiu_2017) în modelarea vizuală a unei analize a datelor contabile se face pe 3 dimensiuni: vizualizare specifică a spațiului reprezentării datelor, interacțiunea cu datele,

identificarea modelelor / tiparelor vizuale și preluarea deciziei contabile. Conform lui (Keim) visual analytics permite investigarea și detectarea similarităților și/sau a anomaliilor în seturi foarte mari de date. În (KEIM) vizualizarea datelor în big data are ca principal dezavantaj generat de abilitățile cognitive umane limitate, mai exact ființa umană nu poate vizualiza layer-uri a datelor mari.

(Elmqvist, et al, 2011). afirmă că în mod obișnuit, se acordă un accent redus cercetării vizualizării și implicit interacțiunea în vizualizare. Interacțiunea în vizualizare este catalizatorul dialogului utilizatorului cu datele și, în cele din urmă, înțelegerea acestor date, dar greu de cuantificat. Interacțiunea este un concept intangibil care este dificil de proiectat, cuantificat și evaluat și sta la baza procesului de luare a deciziilor pe baza tool-urilor de vizualizare numite dashboard. Principalele obiective de business pentru vizualizarea datelor (William Dilla, Diane J. Janvrin, and Robyn Raschke 2010), (Elmqvist, et al, 2011) prin dashboard sunt: descoperirea datelor lipsă sau eronate în seturile de date; susținerea dezvoltării modelelor de Business.

Sistemul de indicatori pe baza cărora se creează dashboard-urile oferă posibilitatea managerilor de a ghida atenția angajaților/ clienților asupra punctelor cheie de urmărit în vederea îndeplinirii obiectivelor fixate, ajută conducerea întreprinderii, oferindu-i posibilitatea măsurării drumului parcurs pentru atingerea obiectivelor și permite luarea deciziilor, punerea în aplicare a planurilor de acțiuni și luarea măsurilor corective, mai exact o analiză completă asupra modului de desfășurare al activităților.

Scenariu de curs elaborat de noi pe baza aspectelor teoretice și practice prezentate în rândurile de mai sus și prezentat mai jos, îi oferă studentului în contabilitate cunoștințe legate de: strategiile de procesare a informațiilor utilizate de manageri pentru rezolvarea problemelor; identificarea prototipurilor de vizualizare adecvate problemei contabile ce se prezintă. (cognitive model fit); conform rezultatelor (William Dilla, Diane J. Janvrin, and Robyn Raschke (2010) înțelegerea cognitivă între sarcină – reprezentare ei între caracteristicile a factoriilor de decizie și reprezentarea adecvată a fenomenelor studiate sprijină obținerea de informații mai eficiente și o performanță mai precisă a factorilor de decizie".

3.4.3 Clasification

Data mining (Sharma, 2021) este procesul de înțelegere a datelor prin curățarea datelor brute, găsirea de modele, crearea de modele și testarea acestor modele. Include statistici, învățare automată și sisteme de baze de date. Exploatarea datelor include adesea mai multe proiecte de date, așa că este ușor să-l confundați cu analiza, guvernarea datelor și alte procese de date. Exploatarea datelor a apărut odată cu calculul în anii 1960 până în anii 1980. Din punct de vedere istoric, data mining a fost un proces manual intensiv de codare pentru a curăța, procesa și interpreta rezultatele extragerii de date astăzi. Data mining a permis ca procesele manuale să poată fi automatizate cu fluxuri repetabile, sisteme de învățare automată (ML) și inteligență artificială (AI). Învățarea automată (engl. "machine learning") își propune descoperirea unor modele interesante în date. Prin utilizarea algoritmilor:

- Clasificarea și regresia: presupun învățarea unei relații între intrări (vectorul x) și ieșire (y) din date. Pentru clasificare, ieșirea este discretă, iar pentru regresie, ieșirea este continuă. Pentru fiecare instanță de antrenare, valoarea dorită a ieșirii este cunoscută (învățare supervizată);
- Gruparea (clustering, partiționare, clusterizare): are ca scop găsirea unor grupuri astfel încât instanțele din același grup să fie mai asemănătoare între ele decât cu instanțele altor grupuri, este nesupervizată;
- Determinarea regulilor de asociere: identifică relații interesante între date tranzacționale aparent neînrudite;
- Selecția trăsăturilor: reprezintă determinarea unei submulțimi de atribute relevante din

mulțimea de antrenare. Data mining implică analizarea tiparelor descoperite pentru a vedea cum pot fi utilizate în mod eficient. În data mining, se sortează seturi mari de date, se identifică modelele necesare și se stabilesc relații pentru a efectua analiza datelor.

Clasificarea în data mining este o tehnică comună care separă punctele de date în diferite clase. Permite organizarea seturilor de date de tot felul, inclusiv seturi de date complexe și mari, precum și cele mici și simple. Clasificarea este un algoritm de învățare supravegheată, scopul principal este de a conecta o variabilă de interes cu variabilele necesare. Variabila de interes ar trebui să fie de tip calitativ. Algoritmul utilizat pentru clasificare în data mining se numește clasificator, iar observațiile care se realizează prin intermediul acestuia sunt numite instanțe. Tehnici de clasificare în data mining lucrează cu variabile calitative. În literatură, există mai multe tipuri de algoritmi de clasificare, fiecare cu funcționalitatea și aplicația sa unică. Toți acești algoritmi sunt utilizați pentru a extrage date dintr-un set de date. Algoritmii de clasificare se divid în două categorii: generative și discriminatoriu. (Sharma, 2021)

Un algoritm de clasificare generativă modelează distribuția claselor individuale, încercând să învețe modelul care creează datele prin estimarea distribuțiilor și a ipotezelor modelului. Acești algoritmi generativi prezic date nevăzute. Un algoritm generativ proeminent este Naive Bayes Clasifier.

Algoritm de clasificare discriminatorie (Sharma, 2021) este rudimentar determină o clasă pentru baza unui rând de date și modelează folosind datele observate, depinzând de calitatea datelor în loc de distribuțiile acestora. Exemplu de algoritm discriminatoriu este regresia logistică. Regresia logistică permite să se modeleze probabilitatea unui anumit eveniment sau clasă. În această regresie se utilizează o logistică pentru a modela o variabilă dependentă binară ce oferă probabilitățile unei singure încercări. Deoarece regresia logistică a fost construită pentru clasificare ajută la înțelegerea impactului mai multor variabile independente asupra unei singure variabile de rezultat. Problema cu regresia logistică este că funcționează numai atunci când variabila prezisă este binară și toți predictorii sunt independenți. De asemenea, se presupune că datele nu au valori lipsă.

Clasificarea (Leonel, 2019) este o abordare de învățare automată supravegheată, în care algoritmul învață din datele introduse care îi sunt furnizate - și apoi utilizează această învățare pentru a clasifica noile observații. Setul de date de antrenament este folosit pentru a obține condiții de limită mai bune care pot fi utilizate pentru a determina fiecare clasă țintă; odată ce astfel de condiții la limită sunt determinate, următoarea sarcină este de a prezice clasa țintă.

O problemă de clasificare este atunci când variabila de ieșire este o categorie, cum ar fi „roșu” sau „albastru” sau „boală” și „fără boală”. Un model de clasificare încearcă să tragă o concluzie din valorile observate. Având în vedere una sau mai multe intrări, un model de clasificare va încerca să prezică valoarea unuia sau mai multor rezultate.

De exemplu, atunci când filtrați e-mailurile „spam” sau „nu spam”, când se uită la datele tranzacțiilor, „frauduloase” sau „autorizate”. Pe scurt, Clasificarea fie prezice etichetele de clase categorice, fie clasifică datele (construiește un model) pe baza setului de antrenament și a valorilor (etichete de clasă) în clasificarea atributelor și le folosește în clasificarea datelor noi. Există o serie de modele de clasificare. Modelele de clasificare includ regresia logistică, arborele de decizie, pădure aleatoare, arbore cu gradient, perceptron multistrat, one-vs-rest și Naive Bayes.

Clasificatoarele binare funcționează cu doar două clase sau rezultate posibile (de exemplu: sentiment pozitiv sau negativ; dacă împrumutul va plăti sau nu împrumutul; etc), iar clasificatorii multclasă funcționează cu mai multe clase (de exemplu: cărei țări îi aparține un steag, dacă o imagine este un măr sau banană sau portocală; etc). Cei mai populari algoritmi utilizați pentru clasificarea binară sunt: (Leonel, 2019)

- K-Nearest Neighbors este un clasificator neliniar ce prezice clasa unui nou punct de date de testare prin găsirea clasei k a celor mai apropiați vecini ai acestuia. Se selectează k vecini cei mai apropiați ai unui punct de date de testare utilizând distanța euclidiană. Un algoritm costisitor, deoarece găsirea valorii lui k necesită o mulțime de resurse, trebuie să

se calculeze distanța fiecărei instanțe până la fiecare eșantion de antrenament, ceea ce mărește și mai mult costul de calcul.

- Logistic Regression
- Vector Machine Support reprezintă datele de antrenament în spațiu diferențiat. Noile puncte de date sunt apoi mapate în același spațiu, iar categoriile lor sunt prezise în funcție de partea decalajului în care se încadrează. Acest algoritm este util în special în spațiile cu dimensiuni mari și este destul de eficient în memorie, deoarece folosește doar un subset de puncte de antrenament în funcția sa de decizie.
- Decision Trees este cea mai robustă tehnică de clasificare în data mining, mai exact este o diagramă similară cu o structură arborescentă în care fiecare nod intern se referă la un test pe o condiție, iar fiecare ramură reprezintă un rezultat al testului (fie că este adevărat sau fals). Fiecare nod frunză dintr-un arbore de decizie deține o etichetă de clasă. Datele se împart în clase diferite în funcție de arborele de decizie. Algoritmul prezice clasele cărora ar aparține un nou punct de date conform arborelui de decizie creat. Limitele sale de predicție sunt linii verticale și orizontale.
- Naive Bayes

Multiclass(Leonel,2019) presupune că fiecare probă este atribuită unei singure etichete. Unul dintre primii algoritmi populari de clasificare în învățarea automată a fost Naive Bayes, un clasificator probabilistic inspirat de teorema Bayes (care ne permite să facem deducție argumentată a evenimentelor care au loc în lumea reală pe baza cunoștințelor anterioare ale observațiilor care ar putea implica acest lucru). Numele ("Naiv") derivă din faptul că algoritmul presupune că atributele sunt independente condiționat. Naive Bayes poate suferi de o problemă cunoscută ca „problema cu probabilitate zero”, când probabilitatea condiționată este zero pentru un anumit atribut, nereușind să furnizeze o predicție validă. O soluție este să folosiți o procedură de netezire (ex: metoda Laplace). Primul pas al algoritmului este de a calcula probabilitatea anterioară pentru etichetele de clasă date. Apoi găsirea probabilității de probabilitate cu fiecare atribut, pentru fiecare clasă. Ulterior, punând aceste valori în formula Bayes și calculând probabilitatea posterioară, și apoi identificarea clasei ce are o probabilitate mai mare, având în vedere că intrarea aparține clasei de probabilitate mai mare.

Algoritmii de clasificare se utilizează de către(Sharma,2021)

- specialiștii în marketing pentru segmentarea publicului. Aceștia își clasifică publicul țintă în diferite categorii utilizând pe baza acestor algoritmi pentru a concepe strategii de marketing mai precise și mai eficiente.
- Meteorologii pentru a prezice condițiile meteorologice în funcție de diverși parametri precum umiditatea, temperatura etc.
- Experții în sănătate publică pentru a prezice riscul apariției diferitelor boli și creează strategii pentru a atenua răspândirea acestora.
- Instituțiile financiare pentru a găsi debitorii în plată pentru a determina ale cui carduri și împrumuturi ar trebui să aprobe, în detectarea fraudelor.

Clasificarea este printre cele mai populare secțiuni ale data mining-ului având o mulțime de aplicații în viața noastră de zi cu zi.

3.4.4. Clustering

Algoritmii de grupare din Machine Learning sunt tehnici nesupravegheate (cele care au date de intrare fără răspunsuri etichetate). Obiectivul lor este de a atrage modele de date și de a grupa observațiile de date în diferite grupuri pe baza asemănărilor lor. Clustering este o altă tehnică comună, destinată grupării înregistrărilor, observațiilor sau cazurilor în funcție de similitudine, nu va exista o variabilă țintă ca în clasificare. În schimb, gruparea înseamnă doar separarea setului de date în subgrupuri. Această metodă poate include gruparea înregistrărilor utilizatorilor după zonă geografică sau grupă de vârstă. De obicei, gruparea datelor în subgrupuri este pregătirea pentru analiză. Subgrupurile devin intrări pentru o tehnică diferită. Clustering este

un proces care analizează setul de date și creează grupuri din puncte de date. Un cluster nu este altceva decât o grupare de astfel de puncte de date similare. În procesarea grupării, punctele de date sunt mai întâi grupate pentru a forma clustere și apoi etichete cu atribuitele acestor clustere.

Realizarea grupării pe setul de date, se realizează cu algoritmi de învățare nesupravegheați, deoarece etichetele de ieșire nu sunt cunoscute în setul de date. Clusteringul poate fi folosit ca parte a analizei exploratorii a datelor. Clusterele trebuie optimizate astfel încât distanța dintre punctele de date din interiorul unui cluster să fie minimă, iar distanța dintre diferitele clustere să fie cât mai mare posibil.

K-Means Clustering este o modalitate de implementare a unui algoritm de clustering care rezumă cu succes date cu dimensiuni mari.

Algoritmul k-medii (k-means) partiționează o mulțime de instanțe numerice în k grupuri (clustere). k reprezintă numărul de grupuri și este un parametru de intrare ales de utilizator. Modul de funcționare:

- Se inițializează aleatoriu cele k centre inițiale;
- Se atribuie fiecărui centru instanțele cele mai apropiate de el:

$$C(i) = \underset{j}{\operatorname{argmin}} \|x_i - c_j\|^2$$

- Se calculează centrul de greutate (media aritmetică) a tuturor instanțelor atribuite unui centru și se actualizează poziția centrului grupului respectiv:

$$c_j = \frac{\sum_{i: x_i \in C_j} x_i}{n_j}$$

Se repetă cei doi pași până când nu se mai modifică poziția niciunui centru

Algoritmul converge, dar găsește de obicei un minim local al funcției de eroare. În general, se recomandă mai multe rulări și alegerea rezultatului celui mai potrivit. Algoritmul dă rezultate bune dacă grupurile sunt convexe și mai ales de formă aproximativ sferică. Valoarea optimă a lui k nu se cunoaște apriori. În majoritatea cazurilor, se încearcă mai multe valori pentru numărul de grupuri k și se alege valoarea care dă cele mai bune rezultate. O metodă automată de alegere a lui k este metoda cotului (engl. “elbow method”). Se alege valoarea lui k pentru care funcția SSE(k) nu mai scade prea mult odată cu creșterea lui k. În funcție de alegerea centrelor inițiale, rezultatele finale pot diferi mult. O măsură a calității grupării este coeficientul de siluetă. Scopul său este maximizarea similarității intra-grup și minimizarea similarității inter-grup. Pentru o instanță, se calculează: ai (distanța medie față de instanțele din același grup) și bi (distanța minimă față de orice instanță din orice alt grup). Coeficientul de siluetă al instanței i este:

$$s_i = (b_i - a_i) / \max(a_i, b_i) \in [-1, 1]$$

Gruparea este mai bună când si este mai mare, aproape de 1 ($a_i \ll b_i$). Coeficientul de siluetă al unui grup este media coeficienților instanțelor.

Metoda coeficientului de siluetă este potrivită pentru algoritmul k-medii, dar nu reflect calitatea partiționării și pentru grupuri de forme arbitrare.

Algoritmul EM (Expectation-Maximization) o variantă flexibilă a algoritmului k-medii, în care grupurile nu mai sunt clar separate, ci o instanță aparține unui grup cu o anumită probabilitate. O instanță aparține de fapt tuturor grupurilor, dar în diverse grade. Dacă se consideră doar

probabilitatea cea mai mare, EM se reduce la k-medii. În continuare, presupunem că grupurile urmează o distribuție gaussiană multidimensională, cu următoarea funcție de densitate de probabilitate:

$$f(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{n/2} |\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})\right)$$

unde \mathbf{x} și $\boldsymbol{\mu}$ sunt vectori, iar $|\boldsymbol{\Sigma}|$ este determinantul matricei de covarianță. În cazul particular 1D, funcția de densitate de probabilitate este:

$$f(x|\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2} \frac{(x - \mu)^2}{\sigma^2}\right)$$

Din punct de vedere practic, trebuie evitată situația în care $|\boldsymbol{\Sigma}| = 0$. Astfel, se pot adăuga niște numere mici la diagonală principală sau la elementele egale cu 0. De asemenea, pentru a evita un număr mare de parametri, dimensiunile se pot considera independente; în acest caz, matricea $\boldsymbol{\Sigma}$ va fi o matrice diagonală.

Gruparea ierarhică-Dendograma, se poate realiza în două moduri: (Hastie, et al., 2009):

- Grupare aglomerativă (strategie bottom-up): inițial fiecare punct reprezintă un grup și apoi se combină acestea în grupuri din ce în ce mai mari. De exemplu: Agglomerative Nesting, AGNES. Cele mai multe metode ierarhice sunt de acest tip; Clusterizarea divizivă (top down), se pornește de la un singur cluster care cuprinde toate datele, clusterul rădăcină, se împarte într-un set de clustere, fiecare cluster fiind apoi divizat recursiv până se obține clustere atomice care conțin o singură dată. (Hastie, et al., 2009):

Algoritmul Aglomerativ(D)

```
Inițializează fiecare dată din setul de date  $D$  ca fiind un cluster,
Compute all pair-wise distances of  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n \in D$ ;
Calculează toate distanțele pereche ale  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n \in D$ ;
repeat
    Găsește două clustere care sunt cele mai apropiate unul de altul;
    Combină cele două clustere și formează un cluster nou  $c$ ;
    Calculează distanța de la clusterul  $c$  la toate celelalte clustere;
until există doar un singur cluster rămas
```

- Grupare divizivă (strategie top-down): inițial toate punctele aparțin unui singur grup și acesta este partiționat treptat în grupuri mai mici. De exemplu: Divisive Analysis

O modalitate mai simplă este aplicarea repetată a algoritmului k-medii cu $k = 2$. Utilizatorul trebuie să specifice un criteriu de terminare, precum numărul de grupuri dorit, un prag impus asupra diametrelor grupurilor etc.

Gruparea aglomerativă, la un moment dat trebuie determinate două grupuri care vor fi combinate. Acest lucru se face prin minimizarea unui anumit criteriu. Există mai multe criterii, dintre care menționăm: (Hastie, et al., 2009):

- Legătură simplă (single link): distanța minimă între oricare două instanțe. Produce „lanțuri” lungi de grupuri.
- Legătură completă (complete link): distanța maximă între oricare două instanțe. Produce grupuri sferice.
- Legătură medie (average link): media tuturor perechilor de distanțe. Scade sensibilitatea la valori extreme

Avantaje grupării ierarhică-Dendograma sunt

- Clusterizarea ierarhică are posibilitatea să utilizeze orice funcție distanță sau similaritate posibilitatea de a explora datele la diferite niveluri de granularitate, deoarece se reține toată ierarhia de clustere și utilizatorul poate alege să vizualizeze clusterele la orice nivel

al arborelui.

- Unele studii au demonstrat că clusterizarea ierarhică aglomerativă produce rezultate mai bune decât metoda k-means, putând detecta clustere de forme arbitrare.

Dezavantaje grupării ierarhice-Dendograma sunt

- calitatea clusterizării ierarhice poate fi afectată de efectul de lanț și de datele marginale
- date de complexitate de calcul reprezintă cerințele de spațiu ce sunt neeficiente și nepractice pentru seturi mari de date (Web-ul). Soluția este extragerea de eșantioane asupra cărora se aplică clusterizarea obținându-se distribuția datelor în clustere fie pe baza distanței, fie prin învățare supervizată.

Această metodă grupează sau împarte grupurile în funcție de metrica distanței selectate, este reprezentată folosind o dendrogramă. Dendrogramă creează o matrice de distanță între toate clusterelor care indică distanța dintre ele. Folosind o metrică de distanță, legătura dintre grupuri se face pe baza tipului de legătură. O problemă importantă în clustering este modul de determinare a asemănării între două obiecte, astfel încât grupurile să poată fi formate din obiecte cu similaritate ridicată în clustere și similaritate scăzută între clustere. Măsură de distanță: (Sharma,2021)

- Euclidiene,
- Manhattan
- Minkowski.
- Jacard

O funcție distanță întoarce o valoare mai mică pentru perechile de obiecte care sunt mai similare între ele.

Algoritmul Lance-Williams identifică o formulă unică pentru mai multe astfel de criterii, prin setarea parametrilor formulei, se alege un criteriu specific. Deciziile de combinare sunt critice deoarece procesul este greedy deci nu permite revizuirea unor decizii deja luate. Gruparea aglomerativă nu scalează bine, deoarece presupune analiza unui număr mare de instanțe sau grupuri.

Abordare divizionară se mai numește și abordare de sus în jos, toate punctele de date sunt considerate ca un singur cluster în faza inițială și apoi aceste puncte de date sunt împărțite pentru a crea mai multe clustere.

Algoritmul K-Medoid, funcționarea acestui algoritm este similară cu cea a K-Means dar diferă de K-Means în ceea ce privește modul în care este atribuit centrul clusterului. Medoidul clusterului este un punct de date real, în timp ce în K-Means calculează centroidul punctelor de date care ar putea să nu fie coordonatele unui punct de date real. În algoritm, k puncte de date sunt selectate aleatoriu ca medoizi ai clusterelor și distanța este calculată între toate punctele de date și medoizii clusterelor. (Sharma,2021)

Clusterelor se utilizează, modelele care sunt extrase din setul de date pot fi ușor de înțeles de către neprofioniști și, prin urmare, pot fi interpretate cu ușurință.

Perspective din date cu dimensiuni mari, seturile de date cu dimensiuni mari nu sunt ușor de analizat doar privind caracteristicile sale. Utilizarea grupării poate ajuta la furnizarea unor perspective și la extragerea unor modele din datele uriașe. Poate oferi un rezumat care ar putea fi util în rezolvarea unor întrebări.

Descoperirea clusterelor arbitrare cu ajutorul diferitelor metode de grupare, putem găsi clustere care pot lua orice formă aleatoare. Acest lucru poate ajuta la obținerea caracteristicilor de bază ale setului de date.

În concluzie, în timp ce analiza este asociată cu generarea și diseminarea de perspective la o scară industrializată, extragerea datelor pentru analiza de afaceri are loc în etapele de inovare și experimentare, în care oamenii de știință și analiștii de date găsesc modele ascunse în date și testează ipotezele de afaceri. În acest context, indiferent dacă întreprinderile aleg să utilizeze software de exploatare a datelor sau să contracteze cu companii de exploatare a datelor (experți în modul de scalare a datelor), exploatarea datelor oferă multiple beneficii comerciale, inclusiv:

[site7]

- Îmbunătățirea procesului decizional, planificare și prognoză.
- Ajută la reducerea riscului, expunerii și costurilor.
- Deschiderea de noi fluxuri de venituri și îmbunătățirea relațiilor cu clienții.

În plus, diferitele tipuri de organizații pot experimenta diferite tipuri de transformare de data mining - cu management și tehnici discrete de data mining - și pot obține beneficii de afaceri specifice industriei.

Teme pentru verificarea cunoștințelor

1. Definiți noțiunea analizei vizuale.
2. Enumerați facilitățile oferite de analizei vizuale.
3. Enumerați caracteristicile analizei vizuale.
4. Descrieți principiul de funcționare a lui Naïve Base.
5. Care este efectul regresiei logistice?
6. Există diferențe între clasificare și clusterizare
7. Ce sunt variabilele PHP? Scrieți sintaxa și exemplificați.
8. Definiți noțiunea de data mining?
9. Enumerați și explicați 2 algoritmi de clasificare.
10. Care este efectul algoritmului SVM?
11. Care este efectul algoritmului arbori de decizie?
12. În ce domenii se utilizează algoritmii de clasificare? Justificați.

K-Means Clustering este

- ☐ modalitate de implementare a unui algoritm de clustering care rezumă cu succes date cu dimensiuni mari
- ☐ variantă flexibilă a algoritmului k-medii, în care grupurile nu mai sunt clar separate, ci o instanță aparține unui grup cu o anumită probabilitate
- ☐ funcție distanță ce întoarce o valoare mai mică pentru perechile de obiecte care sunt mai similare între ele
- ☐ este printre cele mai populare secțiuni ale data mining-ului având o mulțime de aplicații în viața noastră de zi cu zi.

Data mining este

- ☐ procesul de înțelegere a datelor prin curățarea datelor brute, găsirea de modele, crearea de modele și testarea acelor modele.
- ☐ printre cele mai populare secțiuni ale data mining-ului având o mulțime de aplicații în viața noastră de zi cu zi
- ☐ funcționarea acestui algoritm este similară cu cea a K-Means dar diferă de K-Means în ceea ce privește modul în care este atribuit centrul clusterului.
- ☐ variantă flexibilă a algoritmului k-medii, în care grupurile nu mai sunt clar separate, ci o instanță aparține unui grup cu o anumită probabilitate

Răspunsuri

K-Means Clustering este

- modalitate de implementare a unui algoritm de clustering care rezumă cu succes date cu dimensiuni mari
- variantă flexibilă a algoritmului k-medii, în care grupurile nu mai sunt clar separate, ci o instanță aparține unui grup cu o anumită probabilitate
- funcție distanță ce întoarce o valoare mai mică pentru perechile de obiecte care sunt mai similare între ele
- este printre cele mai populare secțiuni ale data mining-ului având o mulțime de aplicații în viața noastră de zi cu zi.

Data mining este

- procesul de înțelegere a datelor prin curățarea datelor brute, găsirea de modele, crearea de modele și testarea acelor modele.
- printre cele mai populare secțiuni ale data mining-ului având o mulțime de aplicații în viața noastră de zi cu zi
- funcționarea acestui algoritm este similară cu cea a K-Means dar diferă de K-Means în ceea ce privește modul în care este atribuit centrul clusterului.
- variantă flexibilă a algoritmului k-medii, în care grupurile nu mai sunt clar separate, ci o instanță aparține unui grup cu o anumită probabilitate

Bibliografie completă a cursului

1. Business Intelligence (BI). (2011). In BusinessDictionary.com online dictionary. Retrieved from <http://www.businessdictionary.com/definition/business-intelligence-BI.html>
2. Bataweel, D. S. (2015). Business intelligence: Evolution and future trends, <https://core.ac.uk/download/pdf/327255786.pdf>
3. Chen, H., Chiang, R. H. L., & Storey, V. C. (2010). Business intelligence research. *MIS Quarterly*, 34(1), 201-203.
4. Conway, M., & Vasseur, G. (2009). The new imperative for business schools. *Business Intelligence Journal*, 14(3), 13-17.
5. Gagne, R. M., Briggs, L. J., & Wager, W. W. (1992). *Principles of instructional design* (4th ed.). Belmont, CA: Wadsworth/Thomson.
6. Gartner. (2006, February 7). Gartner says business intelligence software market to reach \$3 billion in 2009. Retrieved from http://www.gartner.com/press_releases/asset_144782_11.html
7. IBM. (2009). The new voice of the CIO: Insights from the global chief information officer study. Retrieved from http://www-304.ibm.com/businesscenter/cpe/download0/183490/MM_CIO_Study.pdf
8. IEEE (2011). Gagne's nine events. Retrieved from https://www.ieee.org/publications_standards/publications/subscriptions/prod/mdl/gagnes_nine_events.html
9. Inmon, W. H., Strauss, D., & Neushloss, G. (2008). *DW 2.0: The architecture for the next generation of data warehousing*. Burlington, MA: Elsevier.
10. Jourdan, Z., Rainer, R. K., & Marshall, T. E. (2008). Business intelligence: An analysis of the literature. *Information Systems Management*, 25(2), 121-131.
11. Laursen, G. H. N., & Thorlund, J. (2010). *Business analytics for managers: Take business intelligence beyond reporting*. Hoboken, NJ: John Wiley and Sons.
12. Liu, G. (2008). Innovating research topics in learning technology: Where are the new blue oceans? *British Journal of Educational Technology*, 39(4), 738-747.
13. Morris, H. et al. (2002). The financial impact of business analytics. Framingham, MA: IDC. International Data Group. Retrieved from http://www.remycorp.com/documents/IDC_ROIwpinal.pdf
14. Ranjan, J. (2008). Business justification with business intelligence. *VINE*, 38(4), 461 - 475.
15. Reisslein, J., Seeling, P., & Reisslein, M. (2005). Computer-based instruction on multimedia networking fundamentals: Equational versus graphical representation. *IEEE Transactions on Education*, 48(3), 438- 447.
16. Sircar, S. (2009). Business intelligence in the business curriculum. *Communications of the Association for Information Systems*, 24(17), 289-302.
17. Smalltree, H. (2007). Business intelligence case study: Gartner lauds police for crime-fighting BI. Retrieved from <http://searchbusinessanalytics.techtarget.com/news/1507220/Business-intelligence-casestudy-Gartner-lauds-police-for-crime-fighting-BI>
18. Turban, E., Sharda, R., Dursun, D., & King, D. (2010). *Business intelligence: A managerial approach* (2nd ed.). Upper Saddle River, NJ: Pearson Prentice Hall.
19. Watson, H. J. (2008). Business schools need to change what they teach. *Business Intelligence Journal*, 13(4), 4-7.
20. Watson H. J., & Wixom, B. H. (2007). The current state of business intelligence. *IEEE Computer*, 40(9), 96-99.
21. Wixom, B. H. (2010). The state of business intelligence in academia. Retrieved from <http://www2.commerce.virginia.edu/bicongress/images/BI%20in%20Academia%202009.pdf>
22. Xu, Z. et al. (2005). Business intelligence – A case study in life insurance industry. *IEEE International Conference on e-Business Engineering*, 129-132.
23. Zhu, P., & St. Amant, K. (2010). An application of Robert Gagné's nine events of instruction to the teaching of website localization. *Journal of Technical Writing and Communication*, 40(3), 337-362.

24. David Taylor, 2022, What is Business Intelligence? BI Definition, Meaning & Example, <https://www.guru99.com/business-intelligence-definition-example.html>
25. Vine, D. (2000). Internet Business Intelligence: How to build a big company system on a small company budget. Medford, N.J: CyberAge Books.
26. Site2, 2021, What is Business Intelligence (BI)?, <https://www.analyticssteps.com/blogs/what-business-intelligence-bi>
27. Thomas, J., Cook, K, Illuminating the Path: Research and Development Agenda for Visual Analytics (2005)
28. Bonani Bose, 2020, What is Visual Analytics: Key Concepts, <https://www.digitalvidya.com/blog/visual-analytics/>
29. D. A. Keim, F. Mansmann, J. Schneidewind, J. Thomas, and H. Ziegler, "Visual analytics: Scope and challenges," in Visual Data Mining (Lecture Notes in Computer Science), vol. 4404, S. J. Simoff, M. H. Böhlen, and A. Mazeika, Eds. Berlin, Germany: Springer, 2008.
30. J. J. Thomas and K. A. Cook, "A visual analytics agenda," IEEE Comput. Graph. Appl., vol. 26, no. 1, pp. 10–13, Jan. 2006.
31. Cui, W. (2019). Visual Analytics: A Comprehensive Overview. IEEE Access, 1–1. doi:10.1109/access.2019.2923736
32. G.-D. Sun, Y.-C. Wu, R.-H. Liang, and S.-X. Liu, "A survey of visual analytics techniques and applications: State-of-the-art research and future challenges," J. Comput. Sci. Technol., vol. 28, no. 5, pp. 852–867, 2013
33. D. A. Keim, F. Mansmann, A. Stoffel, and H. Ziegler, "Visual analytics," in Encyclopedia of Database Systems, L. Liu and M. T. Özsu, Eds. Boston, MA, USA: Springer, 2009.
34. D. Keim, G. Andrienko, J.-D. Fekete, C. Görg, J. Kohlhammer, and G. Melançon, "Visual analytics: Definition, process, and challenges," in Information Visualization (Lecture Notes in Computer Science), vol. 4950, A. Kerren, J. T. Stasko, J. D. Fekete, and C. North, Eds. Berlin, Germany: Springer, 2008.
35. D. A. Keim, F. Mansmann, and J. Thomas, "Visual analytics: How much visualization and how much analytics?" ACM SIGKDD Explor. Newslett., vol. 11, no. 2, pp. 5–8, 2010.
36. [49] M. Ankerst, "Visual data mining," Ph.D. dissertation, 2001, pp. 1–216.
37. [50] S. J. Simoff, M. H. Böhlen, and A. Mazeika, Eds., Visual Data Mining: Theory, Techniques and Tools for Visual Analytics. Heidelberg, Germany: Springer-Verlag, 2008.
38. [51] K. Sedig and P. Parsons, "Interaction design for complex cognitive activities with visual representations: A pattern-based approach," AIS Trans.Hum.-Comput. Interact., vol. 5, no. 2, pp. 84–133, 2013.
39. A. Endert, M. S. Hossain, N. Ramakrishnan, C. North, P. Fiaux, and C. Andrews, "The human is the loop: New directions for visual analytics," J. Intell. Inf. Syst., vol. 43, no. 3, pp. 411–435, 2014.
40. B. Shneiderman, "The eyes have it: A task by data type taxonomy for information visualizations," in Proc. IEEE Symp. Vis. Lang., Sep. 1996, pp. 336–343.
41. Jorge Leonel, 2019, Classification Methods in Machine Learning, <https://medium.com/@jorgesleonel/classification-methods-in-machine-learning-58ce63173db8>
42. Rohit Sharma, 2021, Classification in Data Mining Explained: Types, Classifiers & Applications [2022] <https://www.upgrad.com/blog/classification-in-data-mining/>
43. http://florinleon.byethost24.com/Curs_ML/SintezaML01C.pdf
44. Rohit Sharma, 2021, Explanatory Guide to Clustering in Data Mining – Definition, Applications & Algorithms, <https://www.upgrad.com/blog/clustering-in-data-mining/>
45. [site5] <https://www.investopedia.com/terms/k/kpi.asp>
46. [site 6] <https://insightsoftware.com/ultimate-glossary-bi-terms/>
47. [site7] <https://www.cognizant.com/us/en/glossary/data-mining>

Bibliografie opțională

1. Dedić N. & Stanier C. (2016). "Measuring the Success of Changes to Existing Business Intelligence Solutions to Improve Business Intelligence Reporting" (PDF). Measuring the Success of Changes to Existing Business Intelligence Solutions to Improve Business Intelligence Reporting. Lecture Notes in Business Information Processing. Lecture Notes in Business Information Processing. Vol. 268. Springer International Publishing. pp. 225–236. doi:10.1007/978-3-319-49944-4_17. ISBN 978-3-319-49943-7. closed access
2. (Rud, Olivia (2009). Business Intelligence Success Factors: Tools for Aligning Your Business in the Global Economy. Hoboken, N.J: Wiley & Sons. ISBN 978-0-470-39240-9.)
3. Coker, Frank (2014). Pulse: Understanding the Vital Signs of Your Business. Ambient

Light Publishing. pp. 41–42. ISBN 978-0-9893086-0-1.

4. Chugh, R & Grandhi, S 2013, 'Why Business Intelligence? Significance of Business Intelligence tools and integrating BI governance with corporate governance', International Journal of E-Entrepreneurship and Innovation, vol. 4, no.2, pp. 1-14. https://www.researchgate.net/publication/273861123_Why_Business_Intelligence_Significance_of_Business_Intelligence_Tools_and_Integrating_BI_Governance_with_Corporate_Governance
5. Golden, Bernard (2013). Amazon Web Services For Dummies. For dummies. John Wiley & Sons. p. 234. ISBN 9781118652268. Retrieved 6 July 2014. [...] traditional business intelligence or data warehousing tools (the terms are used so interchangeably that they're often referred to as BI/DW) are extremely expensive [...]
6. Miller Devens, Richard (1865). Cyclopaedia of Commercial and Business Anecdotes; Comprising Interesting Reminiscences and Facts, Remarkable Traits and Humors of Merchants, Traders, Bankers Etc. in All Ages and Countries. D. Appleton and company. p. 210. Retrieved 15 February 2014. business intelligence.

Glosar

KPI[site5] -Key Performance Indicator sau indicatorul cheie de performanta KPI este un instrument de evaluare a performantei care ilustreaza gradul de atingere a unui parametru dorit.

Data Warehousing: [site 6] Procesul de agregare a datelor din surse dispartate în scopul construirii unui depozit de date. Depozitarea datelor implică proiectarea, dezvoltarea, testarea, implementarea, operațiunile, analiza impactului și managementul schimbărilor.

Curățarea datelor: [site 6] Procesul de detectare și corectare a înregistrărilor defecte, care duce la decizii extrem de precise informate de BI, deoarece bazele de date enorme și achiziția rapidă de date pot duce la date inexacte sau defecte care influențează BI și analiza rezultată. Corectarea erorilor de tipar, deduplicarea înregistrărilor și standardizarea sintaxei sunt toate exemple de curățare a datelor.

Data Visualization: [site 6] practica de structurare și aranjare a datelor într-un context vizual pentru a ajuta utilizatorii să le înțeleagă. Modelele și tendințele care ar putea fi de nerecunoscut pentru neprofesionist în datele bazate pe text pot fi ușor vizualizate și digerate de utilizatorii finali cu ajutorul software-ului de vizualizare a datelor.

Extract, Transform, Load (ETL): [site 6] Se referă la procesul prin care datele sunt extras din surse, transformat sau standardizat pentru stocare în formatul eterogen adecvat și încărcat în magazinul sau depozitul final. Procesul ETL este de obicei rulat în paralel cu procesele de transformare care se execută pe măsură ce datele sunt extrase din surse. Sistemele ETL integrează în mod obișnuit date din mai multe aplicații (sisteme) gestionate și operate de diferiți angajați. De exemplu, un sistem de contabilitate a costurilor poate combina date de la salarizare, vânzări și achiziții.

OLAP cube: [site 6] O metodă de stocare a datelor într-o formă multidimensională, în general în scopuri de raportare. În cuburile OLAP, măsurile de date sunt clasificate în funcție de dimensiuni. Cuburile OLAP sunt adesea pre-rezumat pe dimensiuni pentru a îmbunătăți drastic timpul de interogare și acuratețea din bazele de date relaționale.

Scorecard: [site 6] O reprezentare grafică a progresului de-a lungul timpului a unei întreprinderi, angajat sau unitate de afaceri, către un anumit scop sau obiective care evidențiază KPI-uri

relevante. Tabelele de performanță sunt utilizate pe scară largă în multe industrii, atât în sectorul public, cât și în cel privat.

Scurtă biografie a titularului de curs

Conf. Dr. Liana Stanca doctor în Economie Cibernetică și Statistică Economică din anul 2005. Din 2001 licențiată în Informatică (Facultatea de Matematică și Informatică, secția Informatică din cadrul UBB Cluj-Napoca,), iar din anul 1999 licențiată în Studii Europene. Studii de specializare: master în Informatică Economică și Societate Informațională din cadrul Universității Babeș Bolyai în anul 2002-2003.

Domenii de competență: baze de date și programare, informatică economică, programare web statistică și matematică.. Competențele de cercetare științifică sunt demonstrate prin publicarea a 2 cărți unic autor și 8 în colaborare, a peste 35 de lucrări științifice în publicații recunoscute și participarea la peste 36 de comunicări științifice în cadrul unor sesiuni din țară și străinătate. Membru în comitete de organizare sau științifice ale unor conferințe Rewier în cadrul conferinței internaționale IAWTC și WSEAS. Membru în 6 proiecte de cercetare.

ȘEF CATEDRĂ

TITULAR DE DISCIPLINĂ

Lect.Dr.Cristian Bologa

Conf.Dr. Liana Stanca