**PROIECTAREA DD**

***Depozitarea datelor nu se referă la instrumente.*** Mai degrabă, este vorba despre ***crearea unei strategii de planificare, proiectare și construire a unui magazin de date capabil să răspundă la întrebările unei afaceri***.

În acest context o strategie ***bună este un proces*** ***care nu este niciodată terminat***, mai degrabă este un ***ciclu continuu de perfecționări***.

***Instrumentele care există* funcționează bine pentru a ajuta la realizarea proiectului informational pentru business** prin încărcarea tabelelor țintă în depozitul de date, ***nu pot crea un plan pentru depozit****.*

**PRIN URMARE, ESTE IMPORTANT SĂ *ÎNȚELEGEM PAȘII CARE TREBUIE REALIZATI* MAI ÎNAINTE DE UTILIZAREA INSTRUMENTULUI.**

**Iată acesti pasi:**

1. ***Clarificăm cerințele sarcinei Dvs***. (Instrumentul principal: ***document*** de definire a cerințelor, a ***rapoartelor*** pentru procesul decizional), prin ***examinarea întrebărilor și răspunsurilor adresate utilizatorilor de afaceri în timpul procesului de cercetare***. Adițional este important să examinăm orice rapoarte sau analize existente pe subiectul abordat pentru crearea unui DD.

***TRECEM AICI***!!!

***0\_LAB\_2\_2\_DWH\_\_DEDUCTOR\_\_MI\_\_201\_202.docx***

1. ***Clarificăm datele sursă (dacă ele există) pentru sarcina formulată***. ***Clarificăm relațiile*** care există între datele sursă. Încercăm să înțelegem datele sursă și conținutul acestora, efectuînd o analiză a ***lacunelor(lipselor)*** cu privire la ceea ***ce urmează să facem și nu avem în ceea ce privește datele existente. Identificăm sursele de găsire a acestor date sau de generare a lor***.
2. ***Proiectăm modelul de date pentru depozitul de date.*** Modelăm DD folosind una din „schemele” existente.
3. ***Definim regulile de mapare,*** adică regulile ce definesc de obicei ***ce trebuie de făcut în condițiile specifice ale proiectului concret formulat* *de sarcină*.** Creăm reguli noi pentru ceea ce urmează să facem, pornind de la tabelele sursă existente la tabelele țintă. ***Aceste reguli nu prezic viitorul***, dar urmează să ajute să luăm măsuri semnificative în present ***pentru a realiza proiectul formulat cu scopul lui țintă pentru activitățile de business***.
4. ***Convertim regulile Proiectului sarcinii sursă studiat, în scopul final,*** prin prezentarea regulilor pentru a atinge Scopul descriind metadatele lui (instrument principal al utlizării datelor și rezolvării practice a Proiectului formulat folosind SSAS (sau SAS) împreună cu proiectarea multidimensională și tehnologia cuburilor de date).

**PAȘI PENTRU PLANIFICAREA DD**

***Înțelegerea cerințelor***

În timpul fazei de planificare și proiectare a proiectului de depozit de date, este necesar de a crea un **Document de Definire a Cerințelor** / ***Requirements Definition Document*** (mai are si astfel de denumiri cum ar fi și ***Cerințe de Sistem sau Specificații de Cerințe Funcționale*** / ***System Requirements or Functional Requirements Specification***).

***Acest document descrie*** așteptările și nevoile utilizatorului final, așteptările și nevoile sistemei IT ce urmează a fi utilizată pentru efectuarea scopului de cercetare.

***Acest document definește*** ce dorim să abordăm, să creăm elementele unui domeniu de lucru, sau al unui plan de lucru și să prezentăm pentru acesta modele concepute care să ne ajute să înțelegem mai bine problema noastră, întrebările la care dorim să primim răspuns și datele de care dispunem pentru a realiza toate acestea.

***Cerințele sunt simple* – ele descriu ce ar urma să facă sistemul elaborat și sunt adesea exprimate în acești termeni.**

***De exemplu,***

***Sistemul ar trebui să imprime rezultatele sau să le prezinte la ecran la cererea utilizatorului.***

***Vezi aici***

***0\_LAB\_2\_2\_DWH\_\_DEDUCTOR\_\_MI\_\_201\_202.docx***

***Cerințele pot fi cele mai variate.*** Prezența lor sunt un Ghid pentru Proiectul examinat, procesul prin care știm exact unde ne aflăm și încotro ne orientăm, astfel încât să reușim atingerea Scopului formulat.

***Cerințele bine formulate*** pot servi ca ***un „Contract”*** între dvs., programator și Business-ul care a înaintat Proiectul pentru crearea DD, deoarece ele ***devin un real Ghid pentru modul în care intenționăm să realizăm sarcina formulată, pentru utilizatorul final***.

Important este să menționăm că acest ***Ghid urmează doar să descrie ce dorește să obțină userul*** si ***nu elemente tehnice sau de programare***, ***el trebuie să descrie rezultatele așteptate de user***.

Odată ce ***avem elaborate setul de cerințe***, putem trece la următorul pas.

***Culegerea datelor***

Acest pas se începe cu culegerea datelor ***disponibile suficiente*** pentru a crea un model bine structurat și a obține răspunsurile dorite. ***Putem obține aceste date si informație despre ele din***:

1. ***cerințele din interviuri cu userul,***
2. ***examinarea rapoartelor existente și***
3. ***evaluarea sistemelor informaționale operaționale existente.***

**Vom examina in continuare *modelul unui process de business,* cel mai des întilnit în practica realizării unor proiecte de creare a DD.**

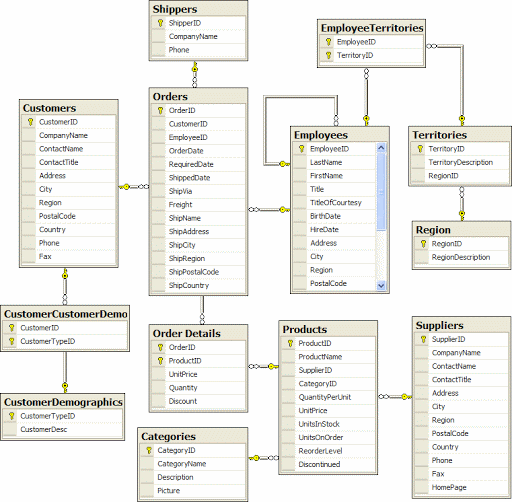
**Obiectivul cercetării:**

* ***Definim scopul și decidem asupra subiectului (subiectelor) pentru a proiecta un DD***
* ***Identificăm întrebările de interes ce urmează a fi clarificate***

Cea mai ***relevantă prezentare*** a „***modelului unui proces de business***”, ***este modelul logic al BDO prezentat prin schema ER***. El poate mai bine ajuta la înțelegerea procesului de cercetare a Proiectului/sarcinii formulate pentru proiectarea DD și poate răspunde la multiplele intrebari pentru ce urmează să realizati DD.

***În acest caz urmează*** a ne ***concentra asupra*** informațiilor disponibile (adică) despre atributele și relațiile dintre ele, ***si nu despre*** ***modul în care datele*** ar trebui să fie accesate sau organizate și nici pentru ce vor fi utilizate.

***De exemplu,*** pentru compania ***Northwinds Trading* /vezi Anexa 1/** cu BDO prezentată mai jos, urmează, în timpul utilizării **modelului unui process de business,**  să răspundem la un șir de ***întrebări/interogări*** cum ar fi:



***1. Cine a cumpărat produsele (clienții și structura acestora)?***

***2. Cine a vândut produsul (organizația de vânzări etc.)?***

***3. Ce s-a vândut (structura produsului)?***

***4. Când a fost vândut produsul (structura timpului)?***

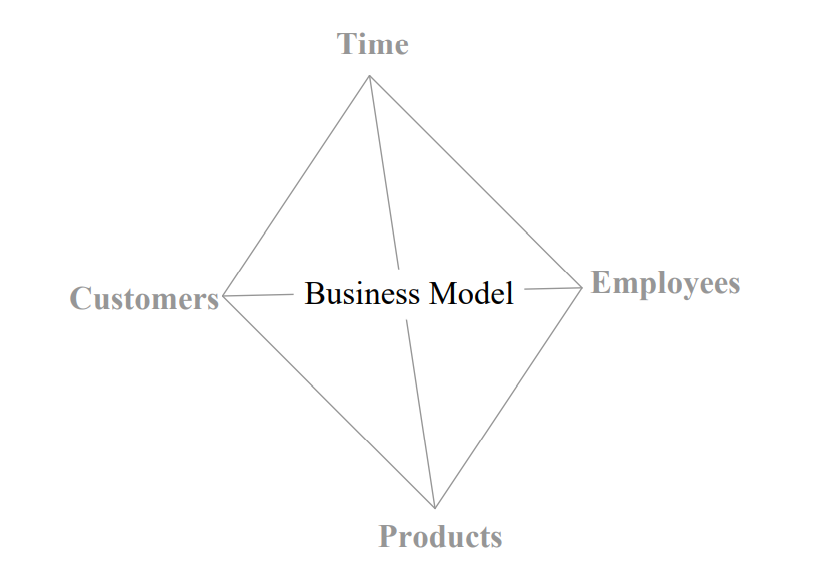
***5. Care sunt caracteristicile vânzării (reducere etc.)? etc…etc***

Astfel, ***Scopul sarcinii este de a construi o structură de date care să ofere răspunsuri la aceste, si nu numai…, întrebări/interogări***.

Pentru a înțelege mai bine procesele de afaceri urmează să cercetăm sarcina, domeniul de studiu ***formulînd cite mai multe întrebări/interogări.***

Astfel de fapt începe ***dezvoltarea cerințelor funcționale*** în raport cu sarcina studiată.

In figura ce urmează sunt prezentate ***subiectele de bază identificate în datele operaționale*** pentru exemplul mentionat:



* **Angajații**
* **Clienți**
* **Produse**
* **Timp**

Reținem că aceste subiecte sunt principalele în cazul nostru și sunt „subiectele” pentru analiză conform modelului abordat - **modelului unui process de business**. ***Intersecția fiecărei linii răspunde*** la ***întrebările / interogările*** despre datele problemei in modelul abordat in calitate de exemplu ***Northwinds Trading***.

***Acum este momentul să luăm în considerare datele despre clienții sarcinii sau despre produsele ei care se află deja în sistem.*** Astfel, am putea de exemplu să formulăm următoarea întrebare/interogare:

***Care a fost venitul total anul trecut de la persoanele angajate în process care erau singure versus cei căsătoriți?***

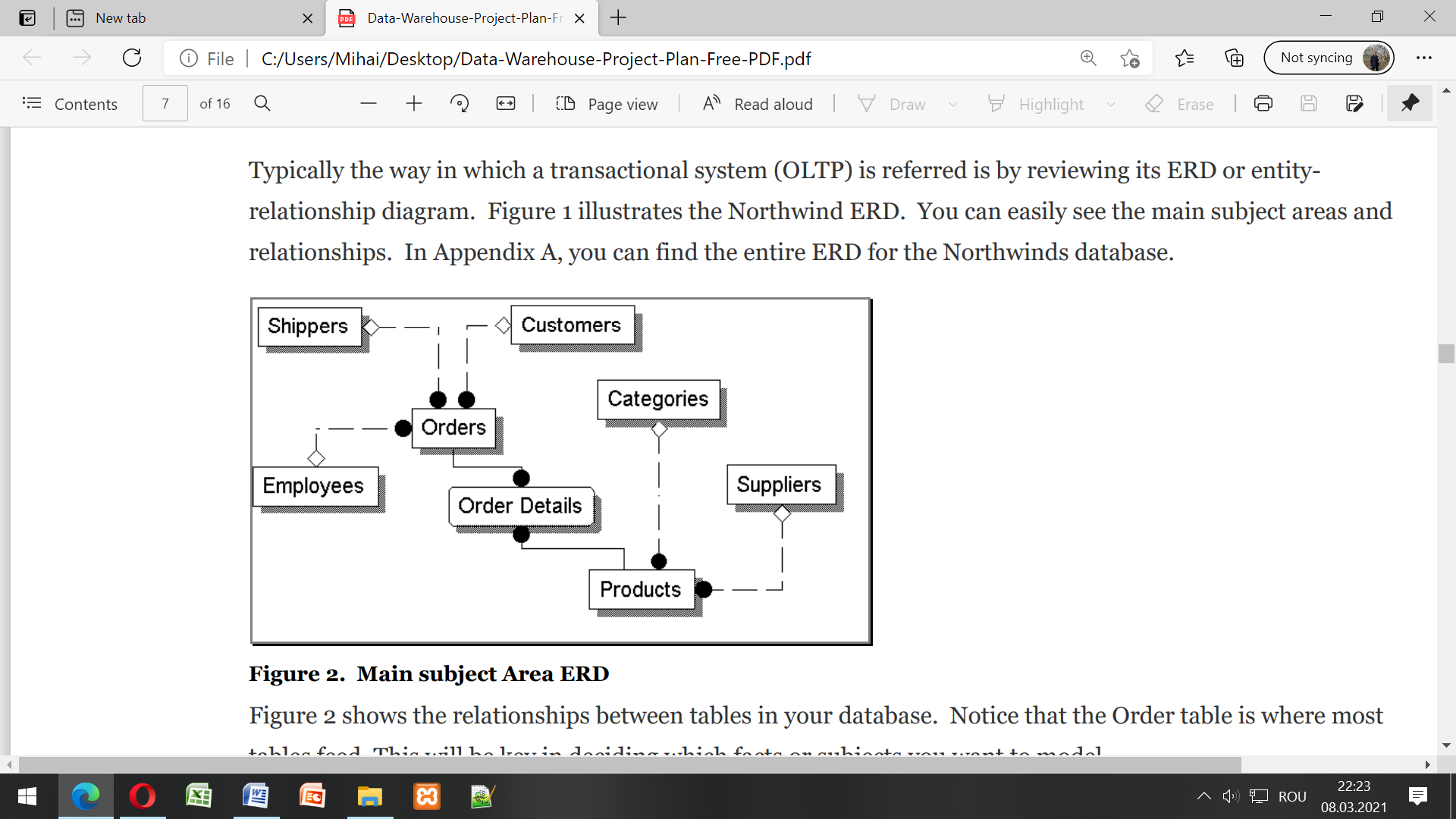
Pentru a răspunde la această întrebare, trebuie să ne dăm seama de ***nivelul de granularitate al datelor*** - sau cât de mult sunt detaliate datele în partea descententă, adică *“în jos”,* pînă la ce nivel.

Atunci când luăm în considerare ***o măsură***, cum ar fi ***veniturile*** sau ***vânzările***, ***este important să luăm în considerare ce date sunt disponibile***. **Ar putea să se întâmple ca aceste date să nu fie?!**

În acest context ne interesează răspunsul la un șir de **alte** întrebări cum ar fi:

* ***există informații despre vânzări pentru fiecare dintre produsele precăutate?***
* ***există informații despre fiecare client?***
* ***există date despre vânzările pentru ultimii cinci ani?***

**Gândiți-vă la date multidimensionale**, așa cum sunt prezentate in Figura ce urmează./ *shippers, customers, suppliers-expeditori, clienți, furnizori/*



***Masurile***, care ar ***exprima vânzările***(volum, costuri de ex!), ***nu există izolat***, ***dar mai degrabă în contextul dimensiunilor stratificate***, cum ar fi: **Produse, Angajați, Clienți și Timpul.** ***Împreună*** ***aceste dimensiuni definesc*** ce tip de date sunt disponibile.

Prin urmare, ***venitul din vânzări*** este măsură și este calificată/determinată de astfel de dimensiuni cum ar fi ***Produse, Angajați, Clienți și Timpul.***

O altă problemă importantă este cea a ***datelor effective*** (care există în realitate). ***O dată efectivă în sistemul operațional de reper va spune ceva despre client la o dată anumită***. In exemplul de mai sus, **datele de care dispunem /ca masuri, fapte!/** nu au nimic cu ceea ce ține de starea civilă codificată pentru fiecare dintre clienții luați la control de aplicația elaborată. Prin urmare, chiar dacă am avea starea civilă a clientului pentru moment, nu am fi cunoscut acest lucru pentru perioada în care am solicitat datele.

***Înțelegerea datelor sursă***

Înainte de a petrece prea mult timp la proiectare și a trece direct la programarea rezultatelor, **ar trebui să obținem** ***profilul* –*datelor***, ***profilarea datelor,*** adică ***generarea unei liste***, ***evedența unor frecvențe și statistici sumare pe câmpurile sursei sistemei de date***. În plus, trebuie să înțelegem în continuare și relațiile dintre tabelele sursă a sistemului de referință, pentru a limita greșeli în definirea JOIN-urilor.

În noua lume a instrumentelor bazate pe interfața grafică, putem obține o înțelegere a datelor chiar cu mult mai înainte de a crea premise informaționale pentru decizii fundamentale de proiectare. Azi sunt cunoscute o sumendenie de instrumente, în totalitate bazate pe SSAS si SAS (***Statistical Analysis System***), ***pe care trebuie doar să știm să le utilizăm. Aceste instrumente permit rapid să inspectăm vizual datele și fără programare si sunt concepute pentru utilizatorul de afaceri sau analistul IT*** pentru a sprijini ***profilarea datelor***, adică să obținem ***profilul* –*datelor***.

***Această soluție*** **ne permite** ***să ne conectăm direct la sursele de date*** și ***să descoperim*** și ***să corectăm automat problemele legate de calitatea datelor***.

Mai mult decît atît, aceste instrumente, ***ne ajută să înțelegem informații importante despre calitatea datelor care include statistici*** la nivel de coloană (de exemplu) numărare, numărare date nule și necompletată, min, maxim etc., distribuții de frecvență, incluzând numărul de modele și unicitatea utilă în determinarea strategiilor de duplicare / curățare. ***Analiza relației ajută la descoperirea valorilor redundante in tabelele multiple*** (adică) evidențiază înregistrări fără legaturi.

***PROIECTAREA MODELULUI DE DATE PENTRU DEPOZITUL DE DATE***

**După ce examinăm datele și considerăm că avem tot** de ce avem nevoie pentru a livra informațiile necesare, ***suntem gata să începem modelarea datelor.***

***Modul în care prelucrăm informațiile într-un sistem tranzacțional*** ***este foarte diferit*** de modul în care dorim să analizăm aceste date utilizînd ***tehnologia OLAP***. În OLTP dorim să primim răspunsuri rapide, fără ***redundanță /exces, surplus de date***/ și ***ne oprim doar la date curente online si operăm doar cu ele***.

În schimb, datele solicitate de analiștii de asistență decizională, au un ***orizont de timp îndelungat***, **sunt redundante** pentru acceptarea diferitelor vizualizări de date, ***sunt adesea rezumate și nu pot fi actualizate***. ***Pentru a furniza date/informatii*** către analiști de susținere a deciziilor, datele operaționale relevante sunt extrase din sistemele OLTP, curățate, codificate si incarcate in tabelele relavante de lucru, respectiv in tebelele de dimensiuni sau fapte.

**Procedurile de realizare a DD de date si de incarcare a datelor in DD sunt diferite de la instrument la instrument de gestionare a DD.**

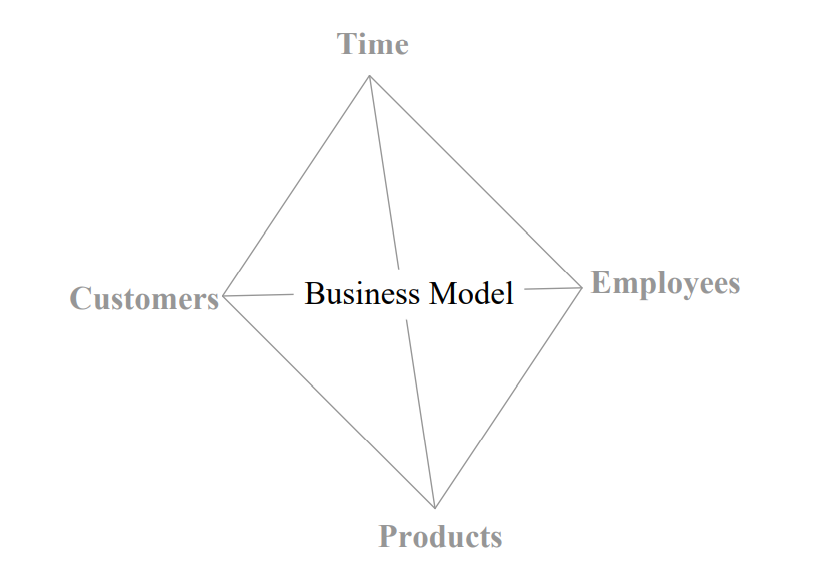
Un lucru este cert - datele sunt organizate foarte diferit în cele două scenarii mentionate anterior OLTP si OLAP.

***Printre altele***, ***OLTP*** folosește un model relațional în care toate datele sunt împărțite în tabele separate ***pentru a reduce redundanța și elimina informațiile cheie duplicate***.  Procesul de divizare/separare/decompozitie a informațiilor in tabele separate este cunoscută sub denumirea ***- normalizarea datelor***.

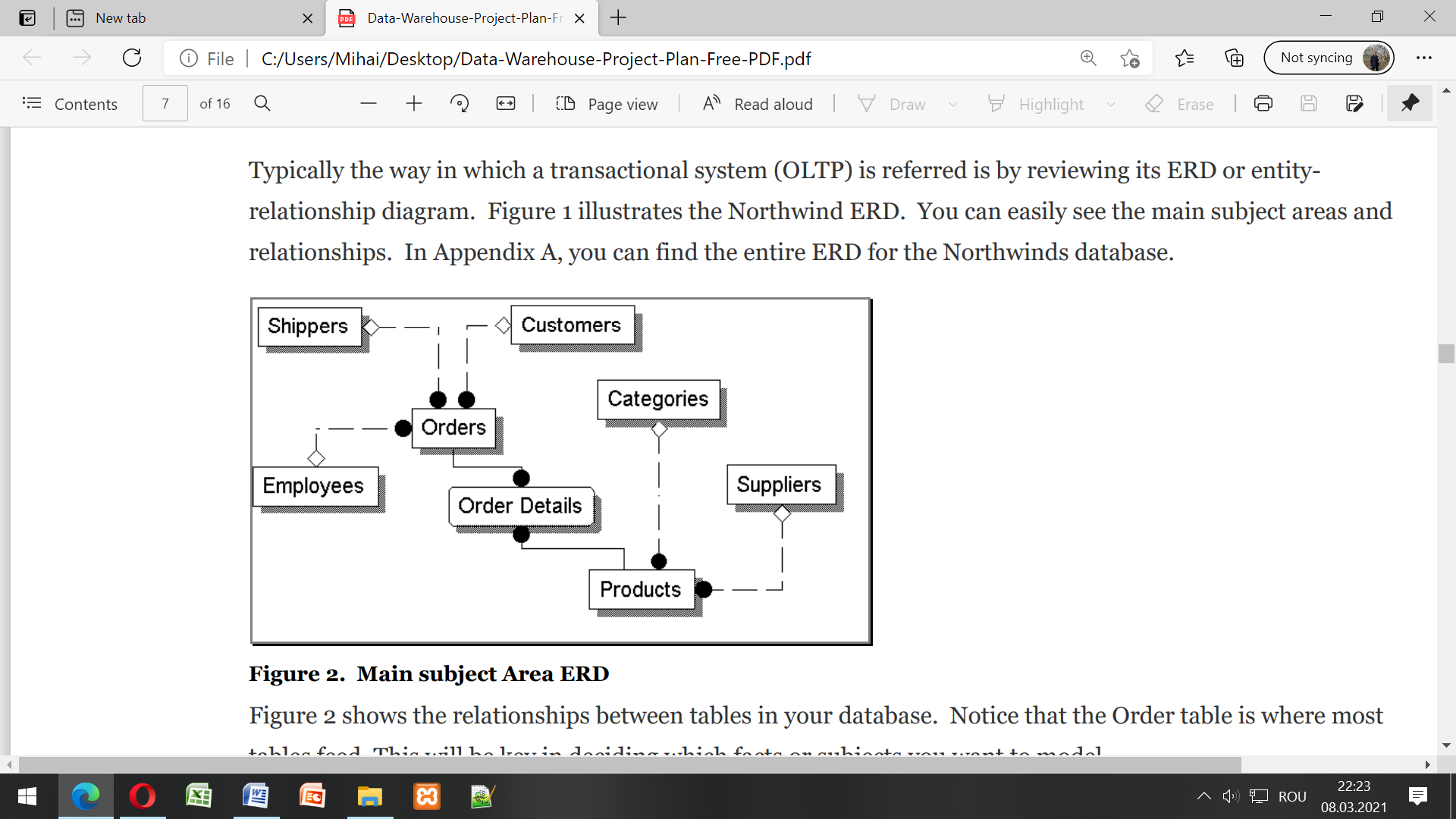
***OLAP,*** pe de altă parte, are nevoie de date ***structurate în mod eficient*** pentru ***raportarea utilizatorilor finali*** și pentru sprijinirea deciziilor. Ca urmare, datele tind să fie redundante. Acest proces se numește ***de-normalizare.***

Să aruncăm o privire asupra sistemului OLTP a ***Proiectului Northwind*** si modelului conceptual al lui prezentat prin schema ER.

**Mai intii de toate vom urmari modelul conceptual al domeniului de studiu cu subiectele/entitățile abordate**

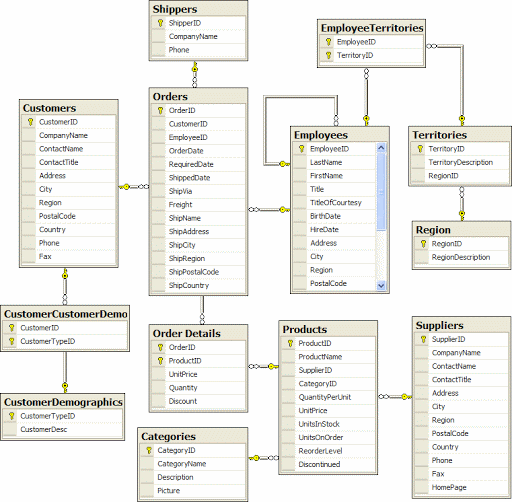


**Sau mai detaliat, prin imaginea ce urmează,**



ce prezenta diagramic relatiile dintre entitățile de bază, ce descriu subiectele mentionate într-un mod mai relevant, conform procesului Proiectului descries pentru compania ***Northwind Traders***.

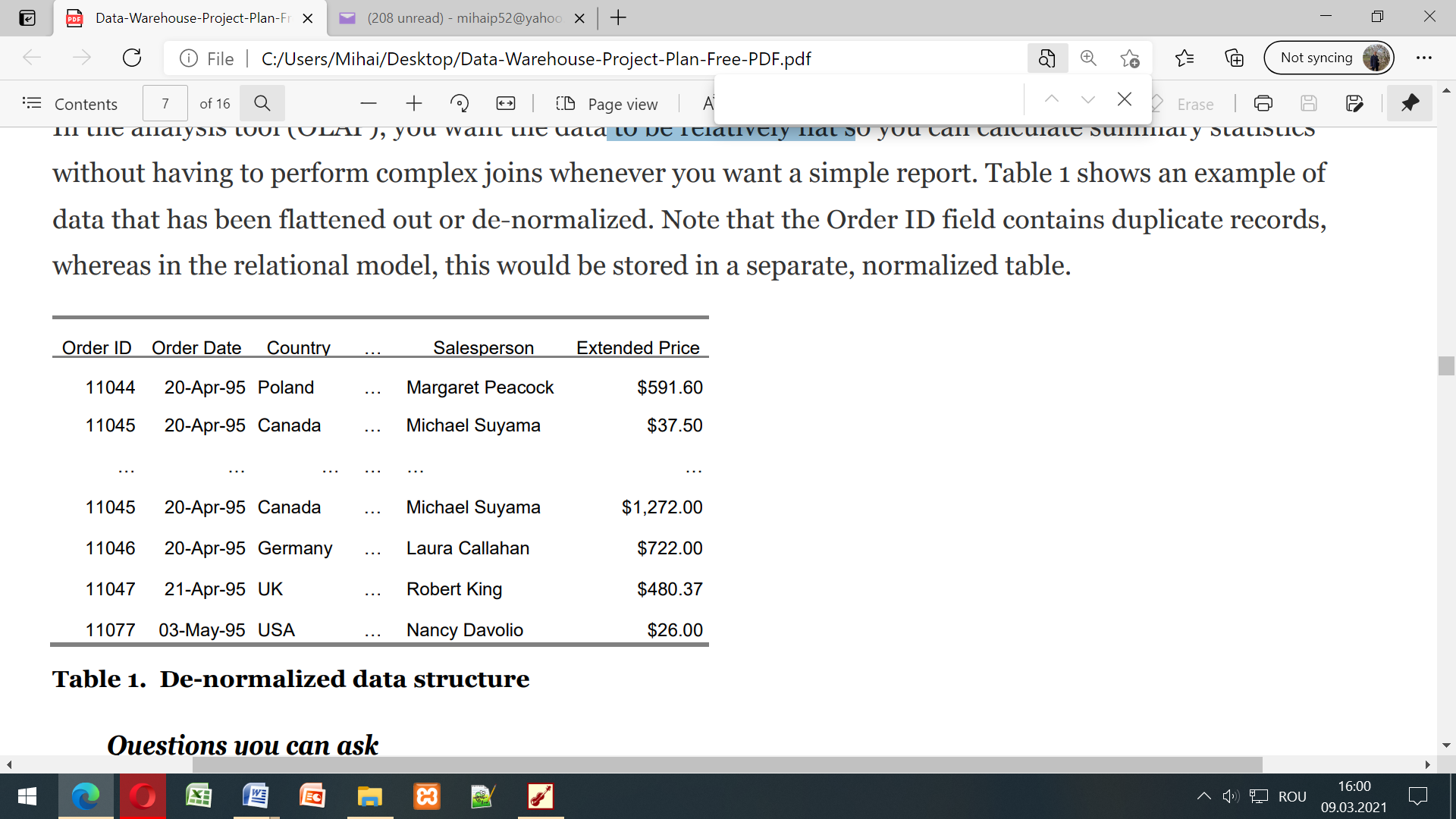
Observăm că tabelul ***Comenzi/Orders*** are cele mai multe legături cu celelalte tabele ce descriu Proiectul. Acest lucru se observa si din modelul logic al BDO de mai jos.



**El, impreună cu cel de mai sus, va fi esențial pentru a decide ce fapte sau subiecte dorim să modelăm în DD, utilizînd modelul multidimensional de date.**

Este evident că în instrumentul de analiză (OLAP), dorim ca datele să fie ***relativ plate***/***uniforme***, astfel ***încât să putem calcula statistici rezumative*** fără a fi nevoie să efectuăm îmbinări complexe (gen JOIN-uri) ori de câte ori dorim un raport simplu.

Tabelul ce urmează prezintă un exemplu de date care au fost aplatizate/uniformizate sau de fapt ***de-normalizate***. Observăm că câmpul ***ID comandă***/***ID\_Order*** conține înregistrări duplicate, fapt care în ​​modelul relațional, ar fi fost stocat într-un tabel separat, normalizat.



***Modelul logic multidimensional***

Să ne reamintim că în timpul analizei modelului logic, o importanță mare o au **cerințele funcționale**. Or, este important să găsim răspunsul la întrebarea: ***Ce dorim cu adevărat de la informațiile pe care le avem despre activitățile stipulate în Proiect cu referință la procesul studiat și cum este ea structurată?***

**Ori, pentru a raspunde la această intrebare, a atinge scopul formulat, este necesar să atingem următoarele obiective:**

1. ***Definim*** cerințele funcționale
2. ***Concretizăm*** subiectele/entitățile cu care vom lucra
3. ***Determinăm*** ce însemnătate/ințeles are dimensiunea Timp pentru Proiectul nostru
4. ***Identificăm*** granularitatea/detaliile (cât de adânc putem coborî) în cercetările noastre pentru subiectele concretizate mai sus.
5. ***Creăm fapte și dimensiuni*** „***reale” pornind de la subiectele/entitățile pe care le-m concretizat mai sus***
6. ***Definim*** cerințele funcționale

**Reamintim**

**Că folosind modelul unui process de business,**  am formulat un sir șir de ***întrebări/interogări*** cum ar fi:

***1. Cine a cumpărat produsele (clienții și structura acestora)?***

***2. Cine a vândut produsul (organizația de vânzări etc.)?***

***3. Ce s-a vândut (structura produsului)?***

***4. Când a fost vândut (structura timpului)?***

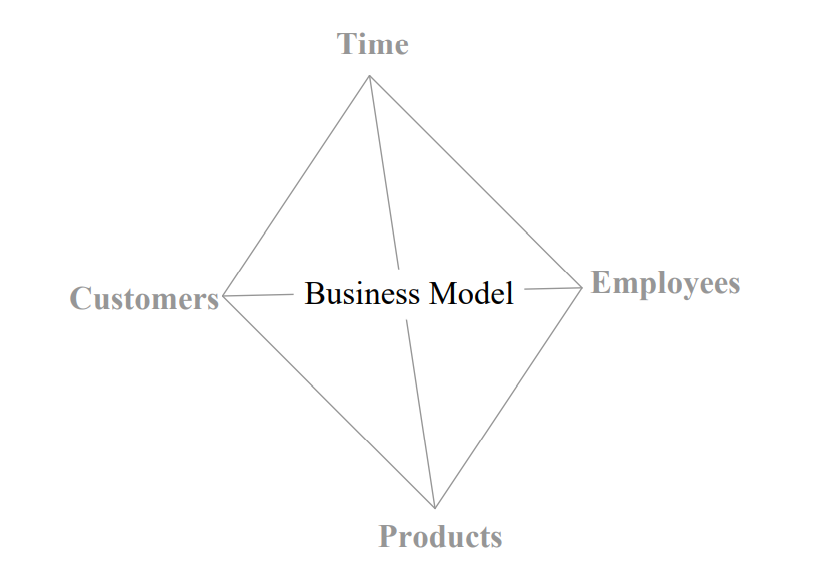
***5. Care sunt caracteristicile vânzării (reducere etc.)? etc…etc***

**si nu numai, adică p**entru a înțelege mai bine procesele de afaceri urmează să cercetăm sarcina, domeniul de studiu ***formulînd cite mai multe întrebări/interogări***. Ori, de fapt începem să dezvoltăm cerințele funcționale în raport cu sarcina studiată. Stăm de vorbă cu sarcina, cu cel ce ne-o formulează. ***Această etapă ne-a condus la aceea că urmează să***

1. ***Concretizăm*** subiectele/entitățile cu care vom lucre

**Reamintim**

In figura ce urmează sunt prezentate ***subiectle de bază identificate în datele operaționale*** pentru exemplul mentionat în formatul ***modelului conceptual***:



* **Angajații**
* **Clienți**
* **Produse**
* **Timp**

1. ***Determinăm*** ce însemnătate/ințeles are dimensiunea Timp pentru Proiectul nostru

**Reamintim**

Aici avem nevoie să înțelegem de ***ce informații temporale*** ***dispunem***, si ***cît de detaliat/adînc*** este necesar ***să descompunem subiectele/entitățile vizate***, pentru a obține răspunsurile formulate la pasul 1?

1. ***Identificăm*** granularitatea/detaliile (cât de adânc putem coborî) în cercetările noastre pentru subiectele concretizate mai sus.

**Reamintim**

Reieșind din pasii 1-3, vedem care sunt nivelele de granularitate a sarcinii formulate.

1. ***Creăm fapte și dimensiuni*** „reale” pornind de la subiectele/entitățile pe care le-m concretizat mai sus si granularitatea identificată

**Reamintim**

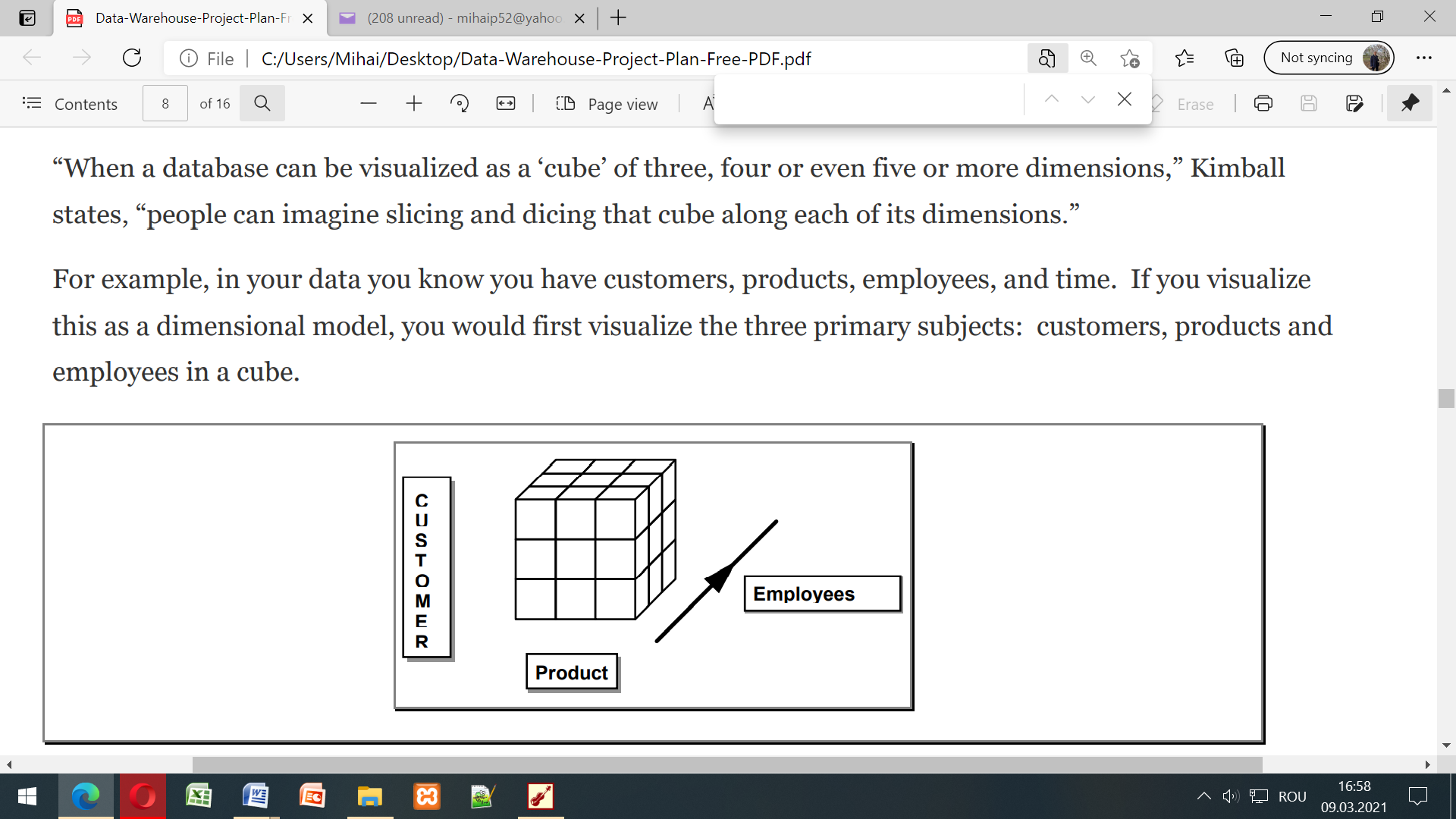
Astfel vom descopune subiectele/identitățile menționate în subiecte noi denumite ***FAPTE*** si ***DIMENSIUNI***, specifice modelului multidimensional.

**Reamintim cite ceva despre modelul multidimensional.**

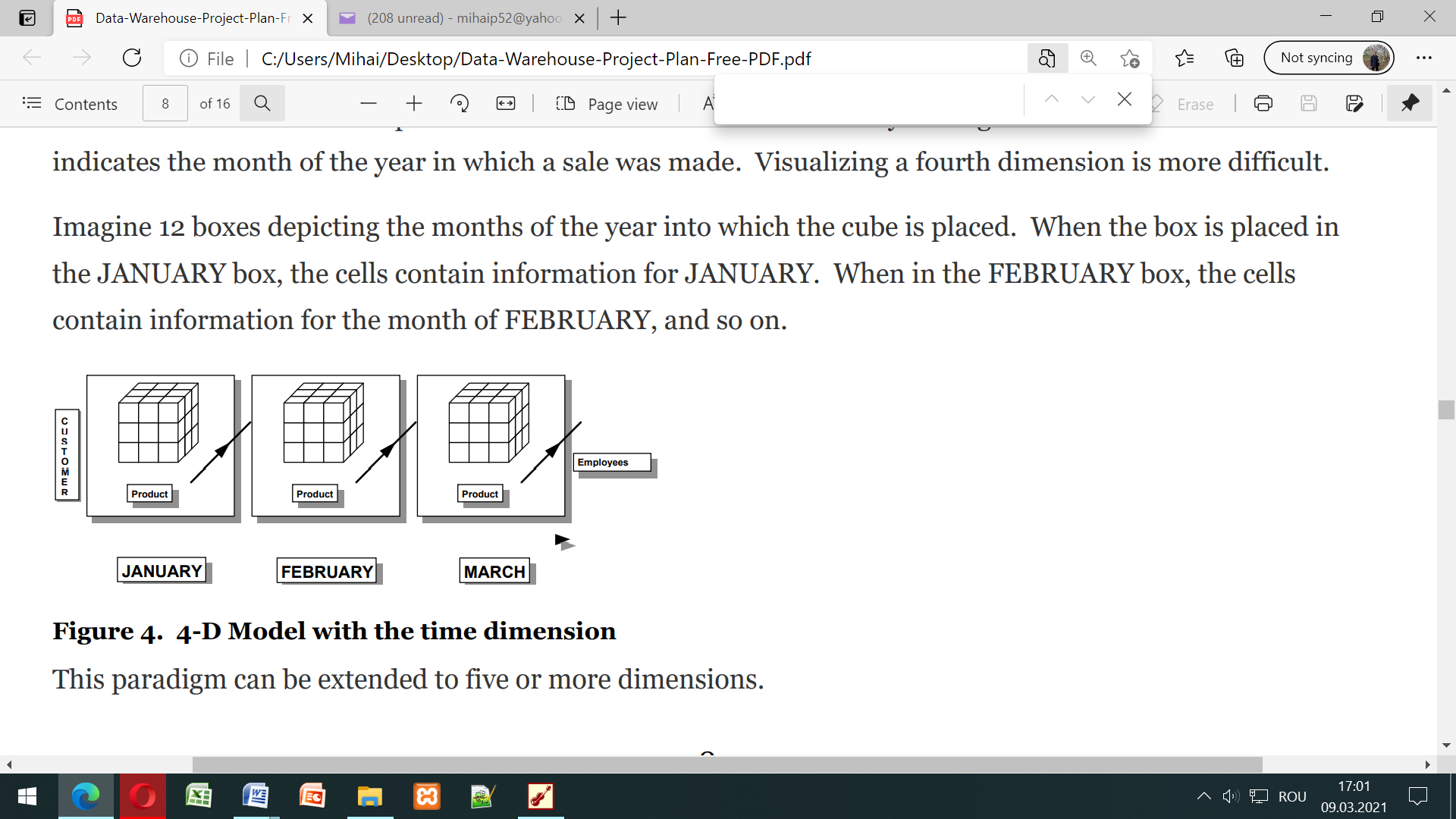
Despre modelul multidimensional a vorbit Ralph Kimball (Kimball, 1996) si a mentionat, că pentru a fi mai ușor de înțeles, el este necesar de asociat cu schema unui cub 3-dimensional, iar in cazul mai multor dimensiuni să-l asociem unui cub n-dimensional si deci multidimensional. Dimensiunile cubului alcătuiesc dimensiunile depozitului de date al Proiectului. ***Prin utilizarea imaginii unui cub, modelul DD poate fi mai bine înțeles***.

„*Când o bază de date poate fi vizualizată ca un„ cub ”de trei, patru sau chiar cinci sau mai multe dimensiuni*,” afirmă Kimball „*oamenii mai usor își pot imagina felierea și tăierea cubului respectiv de-a lungul fiecăreia dintre dimensiunile sale*” si deci mai usor inșeleg ce informații pot fi obținute, pentru a defini clar ***cerintele funcționale ale proiectului***.

De exemplu, datele ce descriu ***clienții, produsele, angajații și timpul***, usor pot fi prezentate printr-un cub după cum urmează:



Ori, modelul 4-dimensional, ce include si dimensiunea ***Timp*** poate fi prezentat după cum urmează:



Această abordare poate fi prelungită si pentru orice **n**, ***ori de cite ori dorim să alcătuim un cub pentru a obține informatii relevante ce va caracteriza punctul 1***, ***Definirea*** cerințelor funcționale ale Proiectului.

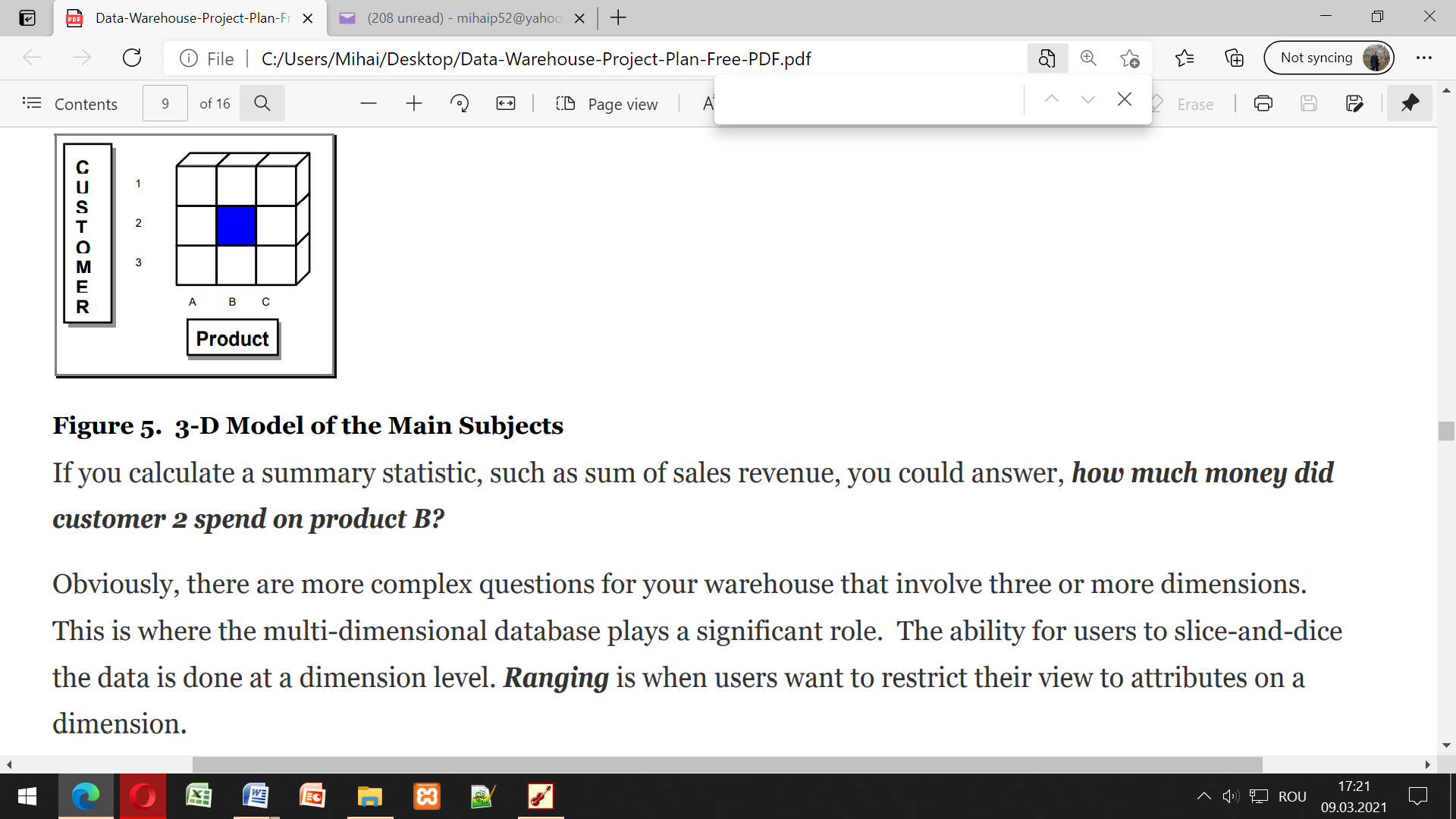
**În acest context, Kimball sugerează că datele încărcate într-un depozit de date sunt una din două tipuri:**

* **fapte sau**
* **dimensiune.**

**Un fapt** ***este o caracteristică*** a unei entități generalizate/***ce descrie procesul de business studiat***/ ***ce poate fi măsurat***; ***sunt numerice și, în general***, ***continue*** (de exemplu) vânzări și sunt utilizate pentru a ***evalua anumite caracteristici statistice ale datelor***.

**Dimensiunile**, pe de altă parte, ***sunt caracteristici a unor entități/ce descrie procesul de business studiat/ sub aspect informațional ce categorizează lucrurile (de exemplu) produse, angajat,*** *etc.*

In exemplul de mai sus, ***veniturile*** sunt ***măsuri*** ce sunt incluse în ***entitatea de Fapte***. Dacă am urmări modelul conceptual prezentat in formă de cub 4-dimensional de mai sus, ***la intersectia a două sau mai multe dimensiuni se află o faptă.*** /mai jos este prezentată intersectia a doua dimensiuni, Produsul/Product B si Clientul/Customer 2

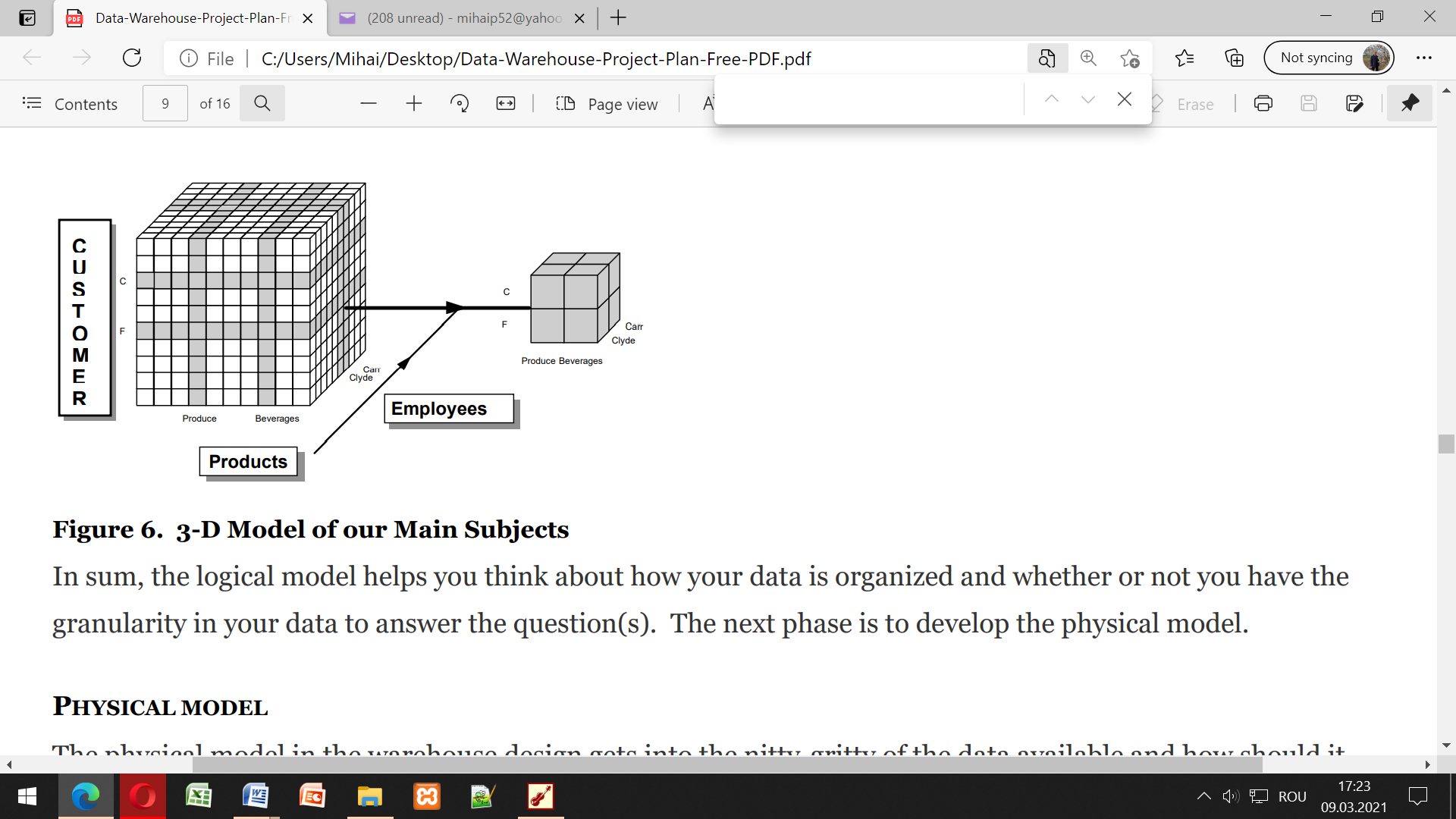


***Dacă calculăm o caracteristică statistică a datelor***, cum ar fi de exemplu SUM, si anume SUMA veniturilor din vânzări, am putea răspunde la următoarea intrebare/interogare,

***câți bani a cheltuit Clientul/Customer 2 pentru produsul B?***

***Evident că există întrebări mult mai complexe pentru un DD care implică trei sau mai multe dimensiuni.*** În acest context, DD multidimensional are o importanță foarte mare si joacă un rol semnificativ in asemenea operatii cu datele.

***Abilitatea utilizatorilor/analistilor de a tăia și felia datele într-un Cub de date se face la nivel de dimensiune. Aranjamentul acestor operatii*** este cazul în care utilizatorii doresc să își restricționeze vizualizarea la atributele unei dimensiuni. În scenariul precăutat, ***este posibil ca un utilizator să dorească doar câteva valori din fiecare dimensiune menționată***. ***În cazul în care dorim să știm*** cât de multe venituri au fost generate de doi dintre angajații de la cei mai mari doi clienți din categoriile de produse băuturi, răspunsul îl găsim in figura ce urmează pentru această situație.

****

**În concluzie,** **modelul logic** indică ***modul în care datele sunt organizate și dacă este sau nu granularitatea în datele proiectului pentru a răspunde la întrebări/interogări si a prezenta cerințele functionale.***

Următoarea etapă este dezvoltarea ***modelului fizic***.

**MODELUL FIZIC**

**Modelul fizic** in proiectarea DD conține detaliile datelor disponibile și cum urmează ele a fi stocate.

***Proiectarea fizică a datelor se prezintă printr-o schemă***. O schemă pentru DD poate fi reprezentată de unul sau mai multe moduri constructive de proiectare, cum ar fi:

1. **Modelul entitate-relație**
2. **Schema de tip stea a faptelor**
3. **Schemă de tip fulg de zăpadă a faptelor**
4. **Schemă de tip constelație a faptelor**
5. **Magazin multidimensional persistent /ce rămîne neschimbat mult timp/**
6. **Tabelele rezumative**

**Scop:**

* **Definirea arhitecturii reale de stocare in DD**
* **Deciderea faptului cum urmează să fie accesate datele și cum sunt aranjate**

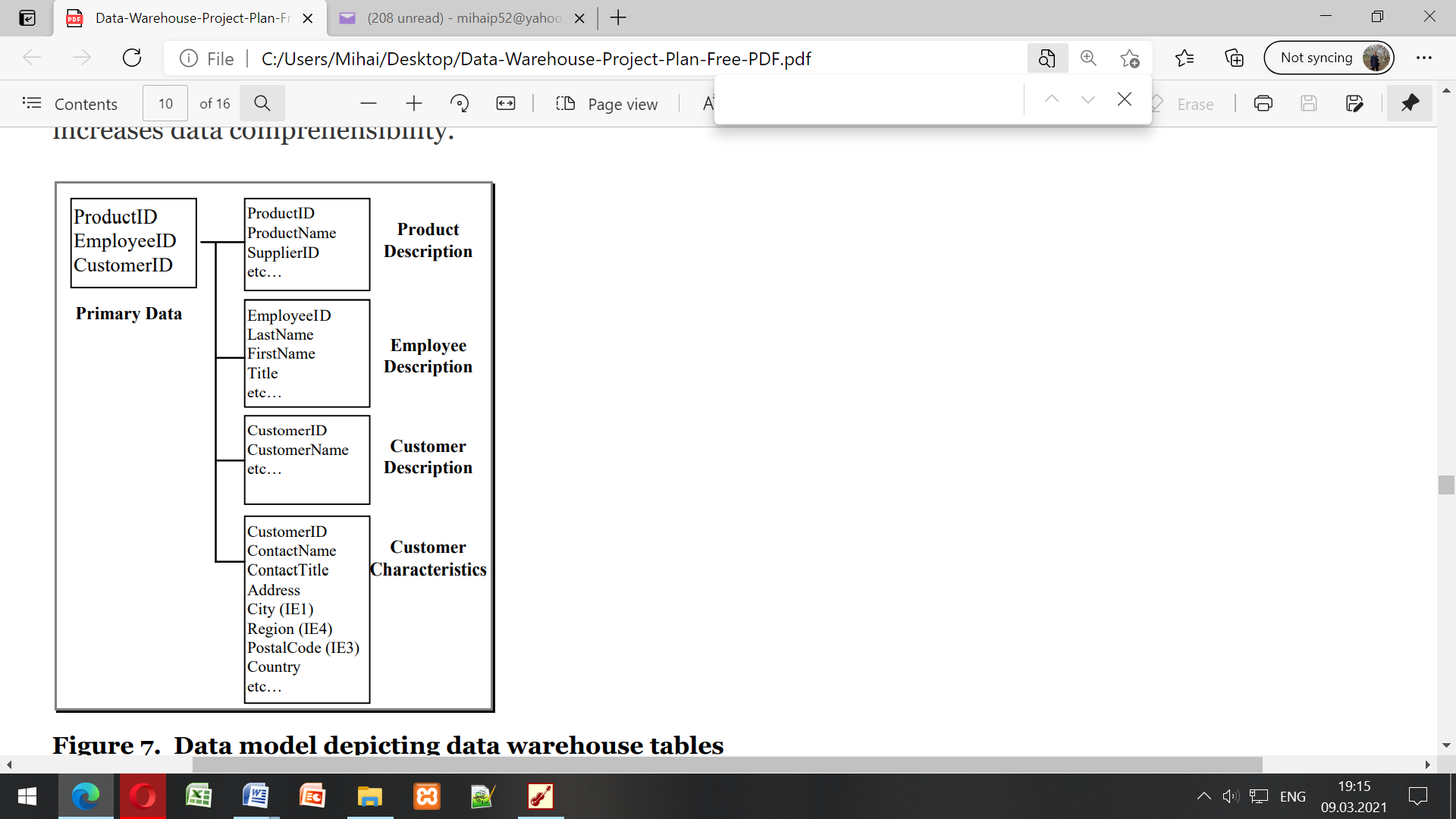
Pentru a clarifica modelul de date al unui DD, este util să examinăm mai întâi tipurile de tabele ce se regăsesc în el.

**Cele trei tipuri de tabele sunt:**

* ***Tabelele de date primare***
* ***Tabelele de descriptori***
* ***Tabele de caracteristici***

**Tabelele de date primare** conțin atât ***măsuri*** cât și ***atribute*** și conțin datele pe care utilizatorii finali le caută. În imaginea ce urmează, “*Model de date care prezintă tabele din DD*”, **Tabelele de date primare** continatributele: ***ID produs, ID client, ID angajat și***

***măsura Vânzări.***



***În depozitele mari de date***, descrierea atributelor integrale de tip text nu este stocată în ***Tabelele de datele primare (liste / dictionare/справочник)***, ele vor fi mai degrabă într-un ***Tabel descriptor***. ***Codurile numerice ale elementelor ID sunt stocate în Tabelul de date primare***. Aceste coduri numerice de identificare se indexează mai repede, generează indici mai mici și oferă o potrivire mai rapidă a clauzei „WHERE”.

***Tabelele descriptori***  conțin adesea doar două coloane, codul ID al atributului și descrierea in engleza a atributului. Există o relație unu-la-unu între ID și descriere. Aceste tabele înlocuiesc codurile de identificare utilizate în interogări cu termeni de afaceri obișnuiți familiari cu utilizatorul.

În aceasi figură există trei ***tabele descriptori*** care mapează codurile ***ProductID, CustomerID și EmployeeID***, împreună cu termenii lor de afaceri care pot fi înțelese de utilizator.

***În depozite mai mici,*** unde performanța de încărcare și problemele legate de stocare sunt mai puțin pronunțate, descriptori de text pot apărea în ***Tabelele de date primare***. Aceasta crește înțelegerea datelor.

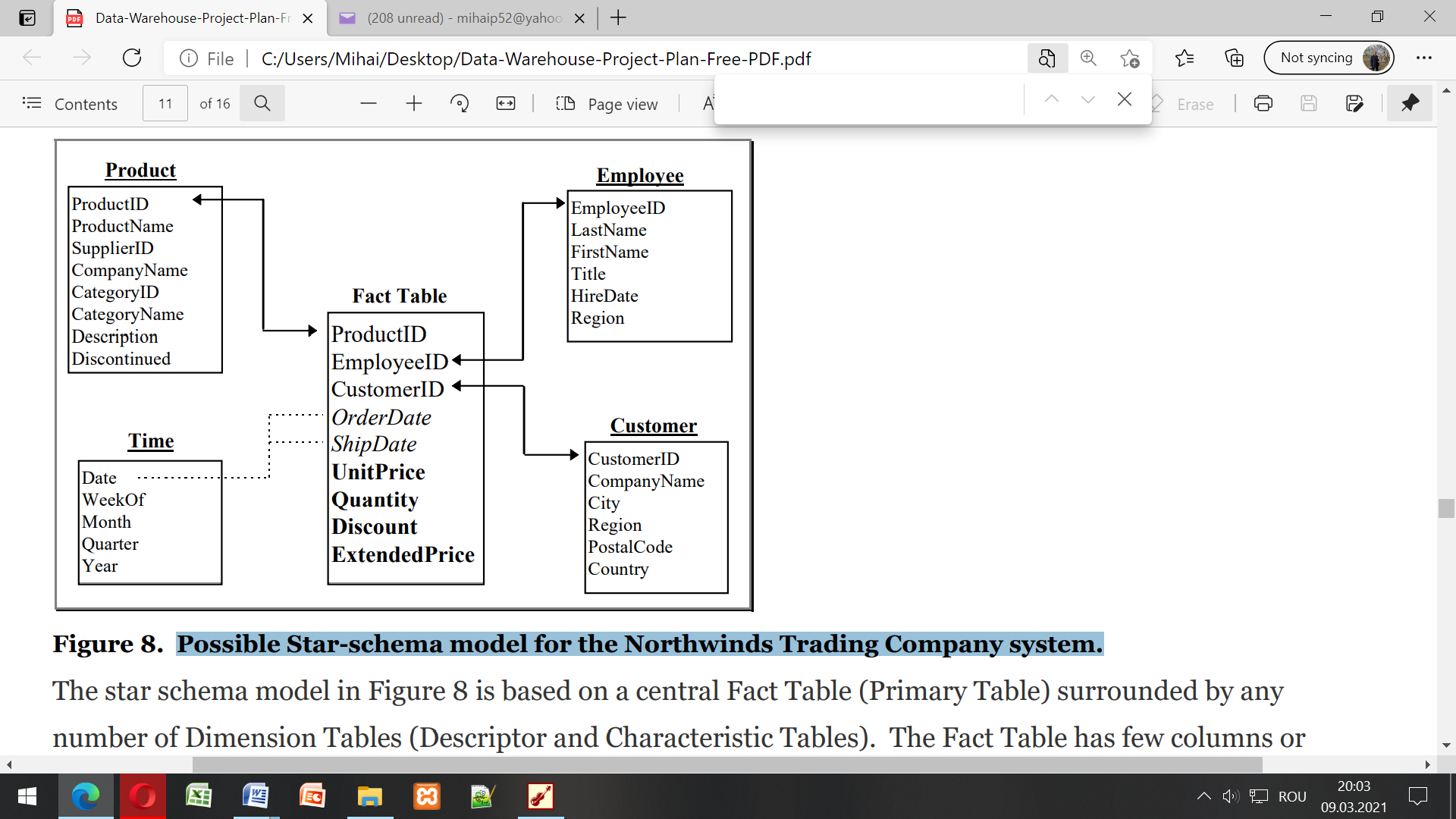
***Tabele de caracteristici*** conțin informații suplimentare despre un atribut și pot fi utilizate pentru a segmenta datele în un mod ad-hoc. Fiecare coloană dintr-un ***tabel de caracteristici*** reprezintă un atribut suplimentar utilizat ca criteriul de filtrare în interogări.

***În figura de mai sus, există un singur*** ***Tabel de caracteristici*** care conține atribute suplimentare legate de atributul Client. Folosind acest ***Tabel de caracteristici***, ***vânzările ar putea fi segmentate în funcție de locația clientului sau de oricare alt atribut*** din ***Tabelul de caracteristici.***

***Este evident că sarcina de bază este a dezvolta/elebora schema DD*** pentru ***a ajuta să răspundem la întrebări/interogări*** ***si să asigurăm cerințele funcționale.***

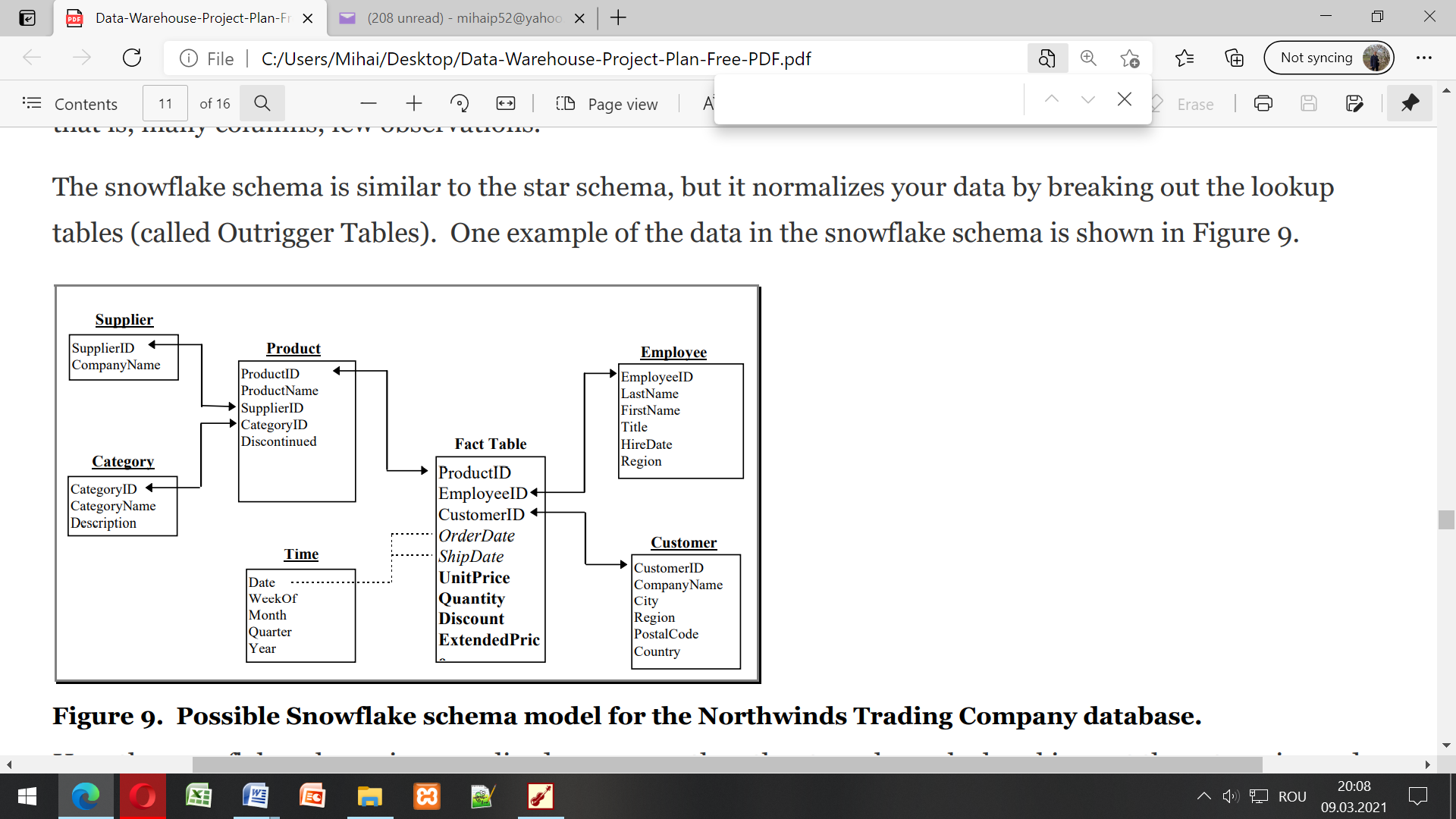
***Schema pe care urmează s-o utilizăm v-a ghida persoana care va realiza sarcina de programare dacă va fi una formulată, sau va utilize instrumental direct de procesare si administrare a unui DD.***

***Schema de tip Stea și Fulg*** de zăpadă sunt schemele cele mai utilizate. Mai jos prezentăm un posibil ***Model de schemă*** ***de tip Stea*** pentru sistemul ***Northwinds Trading Company***.



***Modelul schemei de tip Stea*** de mai sus se bazează pe un Tabel central de Fapte (***Tabel primar***) înconjurat de oricare număr de tabele de dimensiuni (***Tabel descriptor și Tabele caracteristice***). Tabelul de Fapte are câteva coloane sau variabile ***dar multe observații***. Acestea sunt legate de tabelele de dimensiuni, care ***sunt de obicei scurte și largi***, ***adică*** ***multe coloane, puține observații***.

***Schema Fulgului de Zăpadă*** este similară cu Schema de tip Stea, ***dar ea normalizează datele*** prin ruperea căutării tabelare (numite tabele Outrigger). Un exemplu al datelor din Schema de tip Fulg de Zăpadă este prezentat în figura ce urmează. Mai jos prezentăm un posibil ***Model de schemă*** de tip Fulg de Zăpadă, pentru sistemul ***Northwinds Trading Company.***



Aici ***Schema de tip Fulg de Zăpadă*** ***este normalizată*** chiar mai mult decât ***Schema de tip Stea*** prin divizarea categoriilor și informațiilor despre furnizori în tabelele de stabilizare.

***Normalizarea tabelelor de dimensiuni*** reduce stocarea excesivă a datelor prin eliminarea valorilor de date redundante din tabelul de dimensiuni.

***Normalizarea vine cu prețul unor interogări complexe care sunt adesea mai dificil de utilizat și implementat și necesită mai mult timp de procesare.***

**Anexa 1**

**Ce este baza de date Northwind?**

Baza de date Northwind este preinstalată cu aplicația Microsoft Access 2003 și se bazează pe o companie fictivă numită Northwind Traders, care importă și exportă alimente de specialitate din întreaga lume.

Baza de date conține câteva tabele de mostre, interogări, rapoarte și alte caracteristici [baze de date](https://ro.eyewated.com/ce-este-o-baza-de-date/) . Acesta include tranzacțiile de vânzare între companie și clienții săi, împreună cu detaliile de cumpărare între companie și vânzătorii săi.

Această bază de date are, de asemenea, tabele pentru inventar, comenzi, clienți, angajați și multe altele, ceea ce reprezintă o modalitate perfectă de a afla mai multe despre cum să utilizați MS Access.

S-ar putea să doriți să utilizați această bază de date specială pentru a experimenta proiectarea interogărilor folosind tabelul Comenzi și alte tabele conexe pentru analiza tendințelor, deoarece acestea conțin înregistrări pentru o perioadă de trei ani.

Cu baza de date de probă Northwind, puteți practica cu tabele, formulare, [rapoarte](https://ro.eyewated.com/tutorial-rapoarte-de-baze-de-date-microsoft-access/) , macrocomenzi, inventar, facturi și module VBA.

**Anexa 2**

1. SAS stands for the **Statistical Analysis System**, a software system for data analysis and report writing.
2. SAS is a software suite developed by SAS Institute for advanced analytics, business intelligence, data management, and predictive analytics. It is the largest market-share holder for advanced analytics. SAS was developed at North Carolina State University from 1966 until 1976, when SAS Institute was incorporated. SAS was further developed in the 1980s and 1990s with the addition of new statistical procedures, additional components and the introduction of JMP. A point-and-click interface was added in version 9 in 2004. A social media analytics product was added in 2010.
3. **SAS** (previously "Statistical Analysis System") is a statistical software suite developed by **SAS** Institute for data management, advanced analytics, multivariate analysis, business intelligence, criminal investigation, and predictive analytics.