

Introducere in Business Intelligence



Prof. Ramona Bologna,
ASE Bucuresti

Agenda

1. Conceptul Business Intelligence
2. Rolul sistemelor de asistare a deciziei in analiza datelor: OLTP, OLAP, sisteme expert, depozite de date, data mining
3. Depozite de date: caracteristici si arhitectura

1. Conceptul Business Intelligence

- Termenul de BI este relativ nou – 1989, Gartner Group
- Termen “umbrela”: un ansamblu de tehnologii, platforme software, aplicatii si procese specifice.
- Gartner Group estima o creștere de 63% in 2015 si de doar 19% in 2019 pentru solutiile BI si analytics

Business Intelligence

- ❑ Principala valoare = capacitatea de **a transforma datele în informații.**
- ❑ Cine nu obține informații suficient de repede și nu le prelucrează eficient, rămâne în urmă sau dispare, într-un mediu concurențial din ce în ce mai agresiv.
- ❑ **Business Intelligence** se refera la sisteme informatice de identificare, extragere și analizare a datelor disponibile într-o companie, sisteme al caror **scop** este **de a oferi un suport real pentru luarea deciziilor de business.**
- ❑ O soluție de Business Intelligence integrează **datele curente** ale afacerii dar și **date prealabile**, provenind din mai multe programe și aplicații și le consolidează într-o singură **bază de date optimizată pentru regăsirea și analiza informației.**

Business Intelligence

□ Scurt istoric

- 1960 – stocare date pe hartie
- 1970 – aparitia mainframe-urilor
- 1970 – aparitia bazelor de date
- 1980 – aparitia aplicatiilor de business
- 1980 – aparitia conceptului de depozit de date
- 1990 – nasterea Business Intelligence
- 2000 – evolutie continua spre sol. BI accesibile
- 2010 – BI in cloud, analytics

Business Intelligence

□ Definitii

- Set de competente, procese, tehnici si aplicatii menite sa ajute la fundamentarea **deciziilor de business**
- Set de concepte si metode menite sa imbunatateasca procesul decizional
- Categorie de aplicatii si tehnologii menite sa colecteze, stocheze si analizeze date in vederea optimizarii procesului decizional
- Sistem de colectare, integrare si analiza a informatiilor generate de business venite in ajutorul companiilor pentru imbunatatirea deciziilor

Business Intelligence

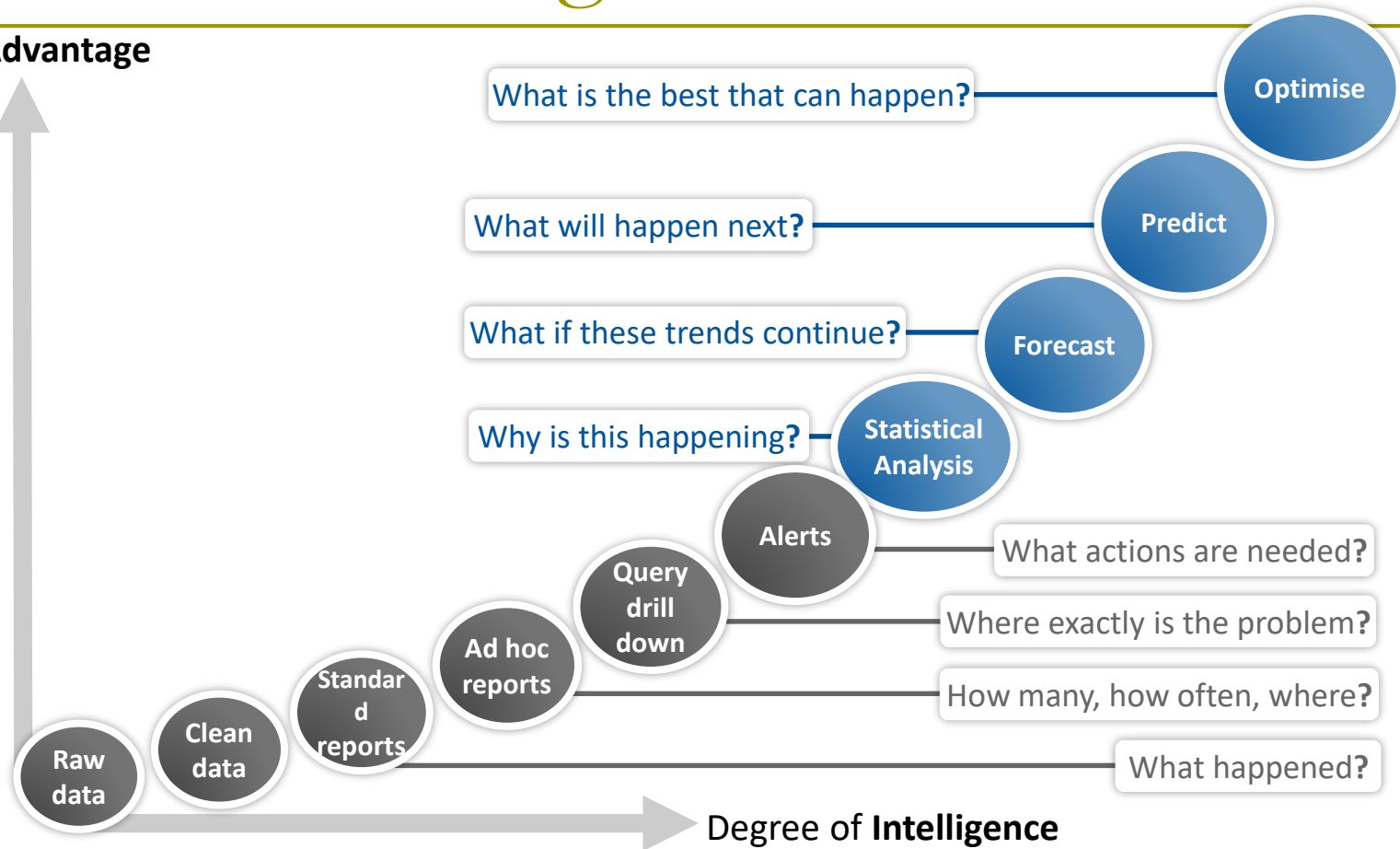
- Cerinte impuse de management
 - Punct unic de acces catre date, indiferent de sursa acestora
 - Acoperirea tuturor proceselor de business
 - Calitate inalta a informatiilor prezentate
 - Informatii structurate conform cerintelor operative si strategice ale managementului
 - Implementare rapida, eficienta

Ciclul de viata al unei solutii BI

1. Definire strategie de business
2. Executie procese de business
3. Colectare date generate de business
4. Analiza datelor colectate
5. Generarea de informatii si cunostinte pe baza datelor analizate
6. Redefinirea strategiei si proceselor de business

Business Intelligence

Advantage



Concepte de baza

- Elemente principale in BI
 - Colectarea datelor – regasirea datelor
 - Analiza datelor – intelegerea datelor si transformarea acestora in informatii si cunostinte
 - Distributia datelor – livrarea informatiilor in timp util catre cei care necesita informatia

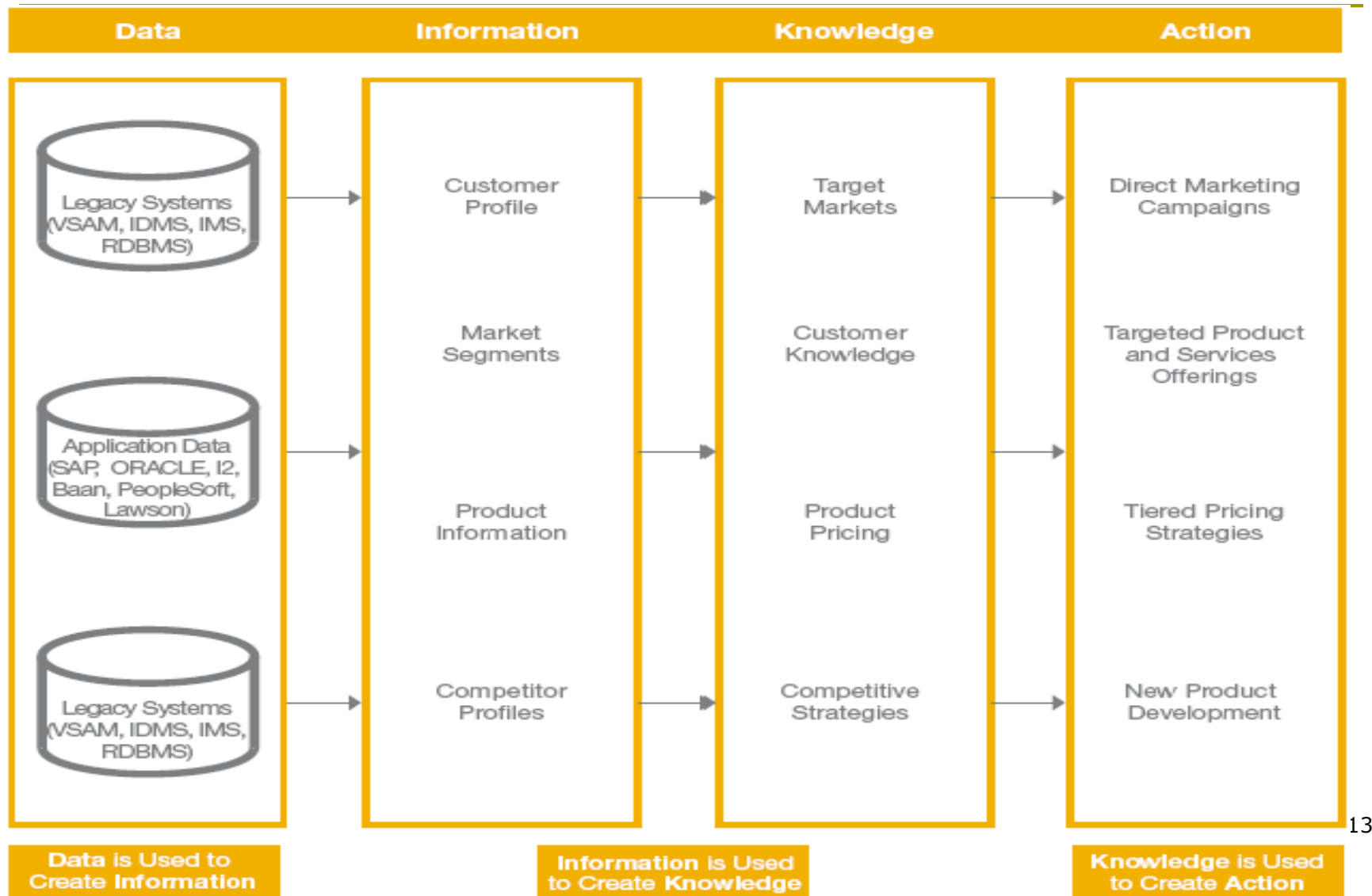
Termeni si acronime in BI

- BI – Business Intelligence
- ETL – Extract, Transform, Load
- DW – Data Warehouse
- KPI – Key Performance Indicator
- Fapt (fact) – masura a unui eveniment de business
- Masura – proces de business cuantificabil
- Dimensiune – impartire a masurilor si atributelor, in functie de interes

Termeni si acronime in BI

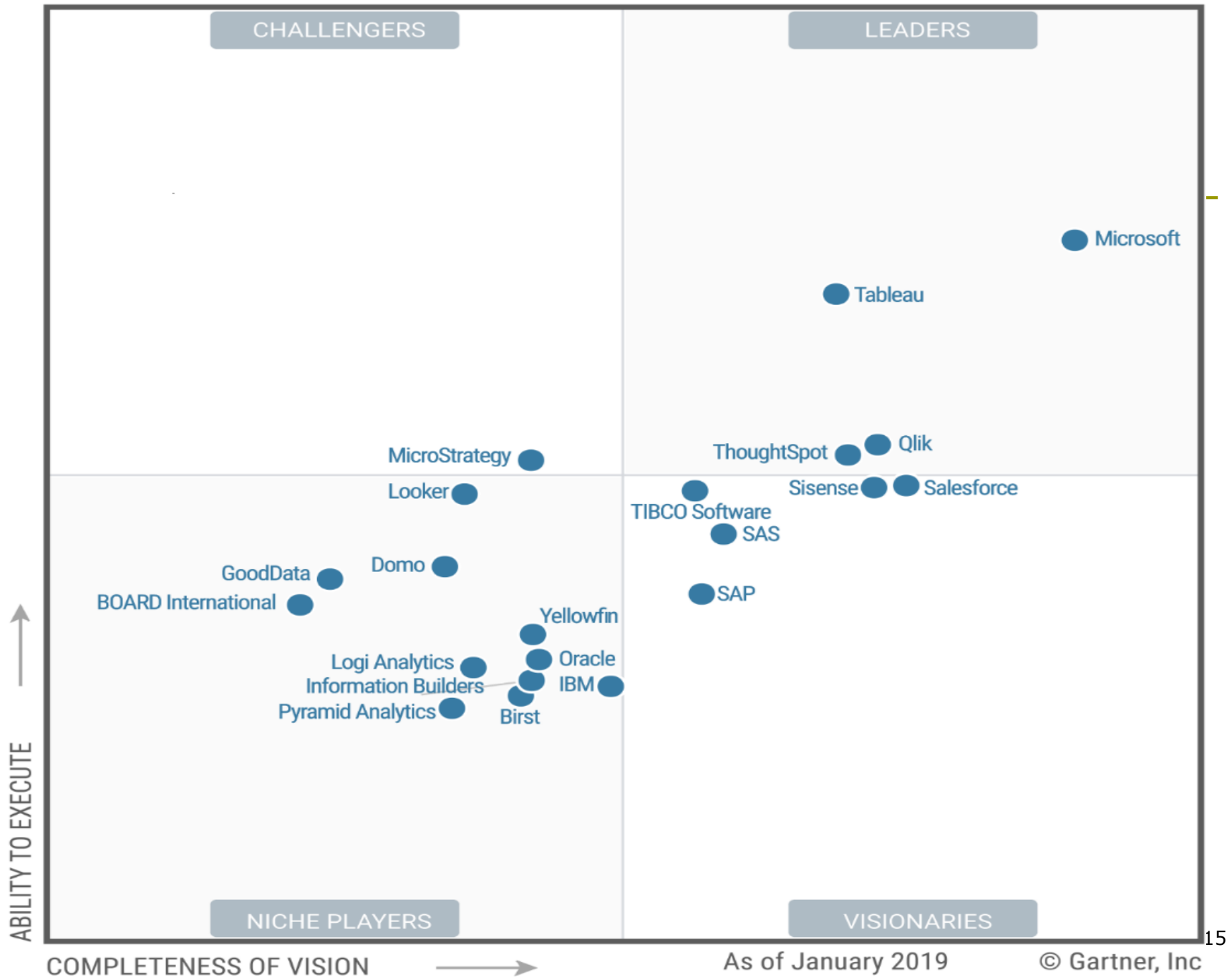
- Atribut – caracteristica care intra in componenta unei dimensiuni
- OLTP – On-line Transaction Processing
- OLAP – On-line Analytical Processing
- Metadata – date despre date
- Granularitate – nivelul de detaliere / sumarizare a datelor
- Schema stea – model de organizare a datelor intr-un sistem de DW

De la colectarea datelor la luarea deciziilor strategice



Elemente de BI

- ⇒ **tehnologia**
- ⇒ putere de procesare
- ⇒ volume de date
- ⇒ tehnologie de retea
- ⇒ standarde – interoperabilitatea BI
- ⇒ software de BI
- ⇒ **oameni**
- ⇒ **cultura companiei**





Probleme

- **Initiativele pentru BI sunt scumpe?**
 - date disparate trebuie extrase si reunite din sisteme OLTP, sisteme interne sau externe.
 - tehnologie noua,
 - sarcini noi,
 - transformari de roluri si responsabilitati,
 - aplicatii suport de decizie livrate rapid si la o calitate acceptabila.
- **60% dintre proiectele clasice de BI esuau datorita:**
 - planificarii inadecvate,
 - sarcinilor ratate,
 - termenelor depasite,
 - unui management de proiect prost,
 - nelivrarii rezultatelor proiectului, sau rezultate de proasta calitate.
- **DAR ACUM, CE E DIFERIT?**

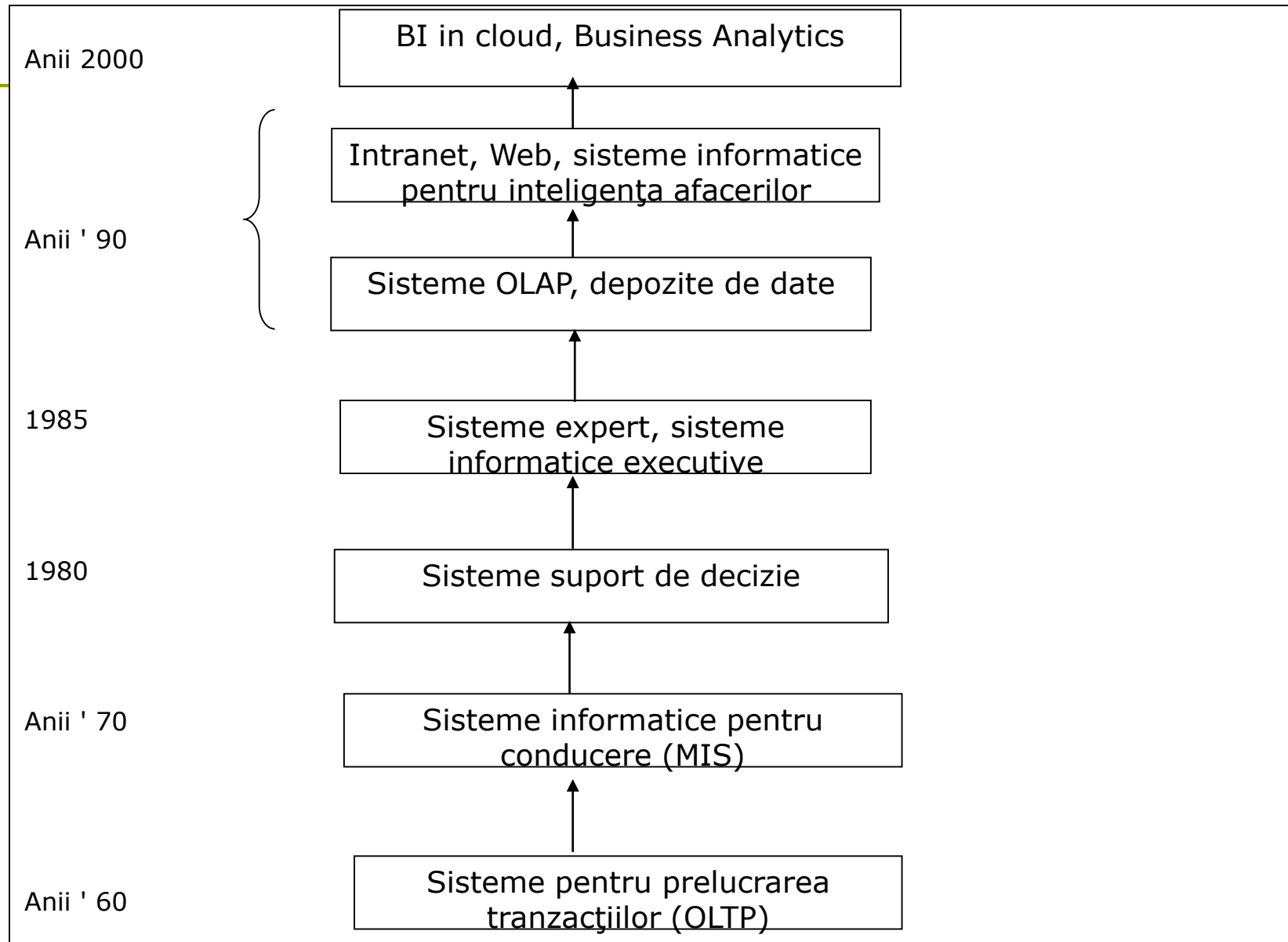
Factori de risc in proiecte BI

- ❑ Lipsa de înțelegere a **complexității** proiectelor de BI
- ❑ Lipsa de înțelegere a faptului că soluțiile de BI implică cel mai adesea **subunitati multiple ale companiei**, ceea ce le face diferite de solutiile stand-alone
- ❑ Reprezentantii companiei sunt **indisponibili** sau **neinteresati**
- ❑ **Lipsa de personal** pregatit disponibil sau utilizarea suboptimala a acestuia
- ❑ **Structura inadecvata a echipei** de proiect
- ❑ Lipsa unei **abordari iterative** in dezvoltarea solutiei
- ❑ **Management** de proiect **ineficient**
- ❑ **Lipsa de metodologie**
- ❑ Lipsa de apreciere asupra **impactului datelor necurate** asupra profitabilitatii
- ❑ Nu este inteleasa necesitatea **utilizarii metadatelor**
- ❑ Utilizarea de **metode si instrumente disparate**

2. Rolul sistemelor de asistare a deciziei in analiza datelor

OLTP, OLAP, sisteme expert, depozite de date, data mining

2.1. Sisteme de asistare a deciziei



a. Sisteme de prelucrare a tranzacțiilor

- **OLTP** - On-Line Transaction Processing Systems
- **Obiectiv:** Prelucrare automata de cantități mari de date rezultate din tranzacțiile zilnice
- **Tipul aplicațiilor:** Gestiunea stocurilor, a plăților, a livrărilor, a comenzilor
- **Operații tipice:** Actualizare
- **Posibilități de decizie:** Deloc sau modele de decizie simple

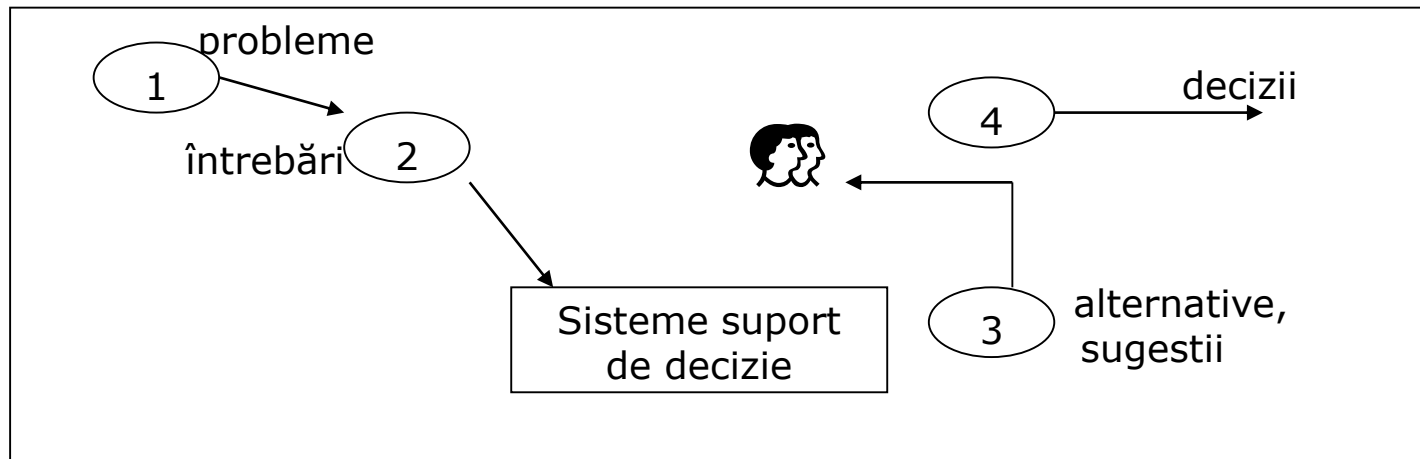
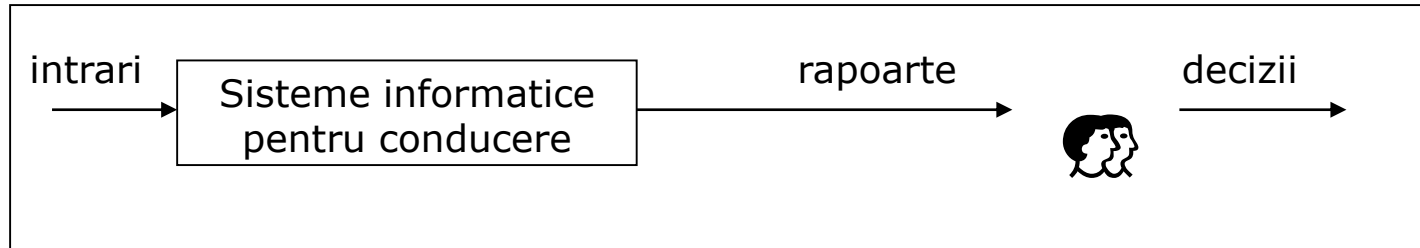
b. Sisteme informatice pentru conducere

- **MIS – Management Information Systems**
- obținerea și stocarea unor volume mari de informații detaliate, rezultate din procesarea tranzacțiilor.
- **Iesirile** sunt predeterminate, sunt sub formă de rapoarte (periodice, la cerere, de excepție), regăsiri de informații on-line
- **Obiectiv:** Informații
- **Tipul aplicațiilor:** Controlul producției, analiza bugetului, previziuni pe termen scurt
- **Operații tipice:** Raportare
- **Posibilități de decizie:** Suport pentru rezolvarea problemelor structurate

c. Sisteme suport de decizie

- ❑ **DSS – Decision Support Systems**
- ❑ interactiv, flexibil și adaptabil
- ❑ interfață simplă și ușor de utilizat
- ❑ **Obiectiv:** Asistarea în procesul decizional, implementarea deciziilor
- ❑ **Tipul aplicațiilor:** Evaluarea creditelor, planificarea asigurării calității, planificarea proiectelor
- ❑ **Operații tipice:** Analiză
- ❑ **Posibilități de decizie:** Suport pentru probleme semistructurate

Comparatie MIS-DSS



Comparatie OLTP-DSS

Caracteristici	OLTP	DSS
Obiectiv	Procesarea automată a tranzacțiilor	Asistarea în procesul decizional și implementarea deciziilor
Tipul de date utilizate	Curente	Istorice, curente
Surse de date	Interne	Interne, externe
Accentul pe	Integritatea și consistența datelor	Flexibilitate, suport pentru cereri ad-hoc și analize
Preocuparea majoră	Eficiența prelucrării datelor operaționale	Eficacitatea procesului decizional
Nivelul de conducere	Operațional	Tactic și strategic
Granularitatea datelor	Detaliu	Agregate
Operații tipice	Actualizare	Analiza
Tipul de utilizatori	Operatori	Manageri

d. Sistemele expert

- **sistem bazat pe cunoștințe**
- Utilizarea de cunoștințele expert pentru a rezolva o anumită problemă
- Succes în diferite domenii: diagnosticare, prognozari
- **Obiectiv:** înlocuire decident uman
- **Tipul aplicațiilor:** pt domenii specifice
- **Operații tipice:** inferente logice
- **Posibilități de decizie:** decizii complexe, nestructurate (reguli, euristici)

e. Sistem informatic pentru inteligența afacerilor

- ❑ BIS business intelligence system
- ❑ Sistemele suport de decizie
- ❑ Sistemele expert



sistem puternic care să îmbunătățească considerabil procesul decizional:

- ❑ Datawarehouse
- ❑ OLAP -Online Analytical Processing

f. Business analytics

BA= business intelligence + advanced analytics

BA	Description	Examples
Descriptiv	Urmareste KPI pentru a intelege starea curenta a firmei (Ce se intampla?)	<ul style="list-style-type: none">-enterprise reporting-Dashboards-Scorecards-OLAP-Data visualization-Simple statistical techniques
Predictiv	"Analizeaza trendul datelor pentru a evalua probabilitatea unor rezultate viitoare (Ce se va intampla?)	<ul style="list-style-type: none">-Simulation-Advanced statistics-Data mining-Machine learning-Text analytics-Artificial Intelligence-Spatial machine learning
Prescriptiv	Utilizeaza performantele trecute pentru a genera recomandari despre rezolvarea unei situatii similare in viitor (Ce ar trebui sa fac?).	<ul style="list-style-type: none">-Decision trees-Optimization and simulation algorithms-Fuzzy Rule-Based System-Switching Neural Networks

3. Depozite de date

- Arhitectura depozitului de date.
- Concepte de baza:
 - cub de date,
 - datamart,
 - dimensiuni,
 - fapte,
 - agregari,
 - granularitate.

Depozite de date

- **Consiliul OLAP 1995:** o stocare centralizată a datelor detaliate provenite din toate sursele relevante din cadrul unei organizații ce permite interogarea dinamică și analiza detaliată a tuturor informațiilor.
- **William Inmon:** *o colecție de date orientate pe subiecte, integrate, istorice și nevolatile destinată sprijinirii procesului de luare a deciziilor manageriale*
- **+instrumente de interogare, analiza și prezentare a informațiilor**
 - Instrum. de analiza on-line- OLAP
 - Instrum. de data mining

In cifre, un proiect clasic...

- **Dimensiunea DW** – TBytes
- **Costul implementarii** – peste 1 mil \$
 - Servicii profesionale
 - Software pentru extragere, transformarea, incarcarea si analiza datelor
 - Sisteme hardware si stocarea datelor

Scopul DW – integrarea datelor

□ **INTEGRAREA datelor**

- modalități unice de codificare, sistem de unități de măsură consistente,
- sistem stabil de reprezentare fizică a datelor,
- convenții clare privind modul de reprezentare a datelor calendaristice,
- convenții unice privind denumirile datelor.

□ **FLEXIBILITATE** – sa se conecteze la nivelul intregii organizatii a.i. servere de la furnizori diferiti sa se poata conecta la depozitul central existent

□ **ARHITECTURA** – adaptare usoara la modificarile de performante, capacitate si conectivitate

□ **Data mart** - >500 GB, <1mil \$, <3 luni

Ce este un depozit de date?

- ❑ **William Inmon:** este o colecție de date orientate pe subiecte, integrate, istorice și nevolatile, fiind destinat fundamentării deciziei manageriale.
- ❑ O BD pentru luarea deciziilor, **separata** de BD operationala a companiei
- ❑ Oferă suport pentru **procesarea informațiilor**, oferind o platforma de **date istorice consolidate** pentru analiza
- ❑ Structurile de date într-un depozit de date sunt optimizate pentru o **regăsire** și o **analiza rapida**.

a.DW – orientat pe subiecte (Inmon)

- ❑ Organizat pe subiecte importante: **client, produs, vanzari.**
- ❑ Accent pe modelarea si analiza datelor de catre decidenti
- ❑ Oferă o **perspectiva simpla si concisa** asupra anumitor subiecte, **excluzand datele care nu sunt utile in procesul de luare a deciziilor**

b.DW – integrat (Inmon)

- Integreaza surse de date **multiple**
- Tehnici de **curatare** si **integrare** a datelor.
- Consistenta in
 - conventiile de numire,
 - structura codurilor,
 - unitatile de masura folosite de diferitele surse
 - E.g., Pret hotel: moneda, taxe, mic dejun inclus, etc.

c.DW – istorice (Inmon)

- **Datele sunt istorice** și sunt actualizate la intervale regulate.
- Orizontul de timp este mult mai mare decât la sist. operationale (ex: 5-10 ani)
- Fiecare element structural cheie al depozitului:
 - Contine o **referire temporală**, implicita sau explicita, ceea ce nu se întâmplă la datele operationale

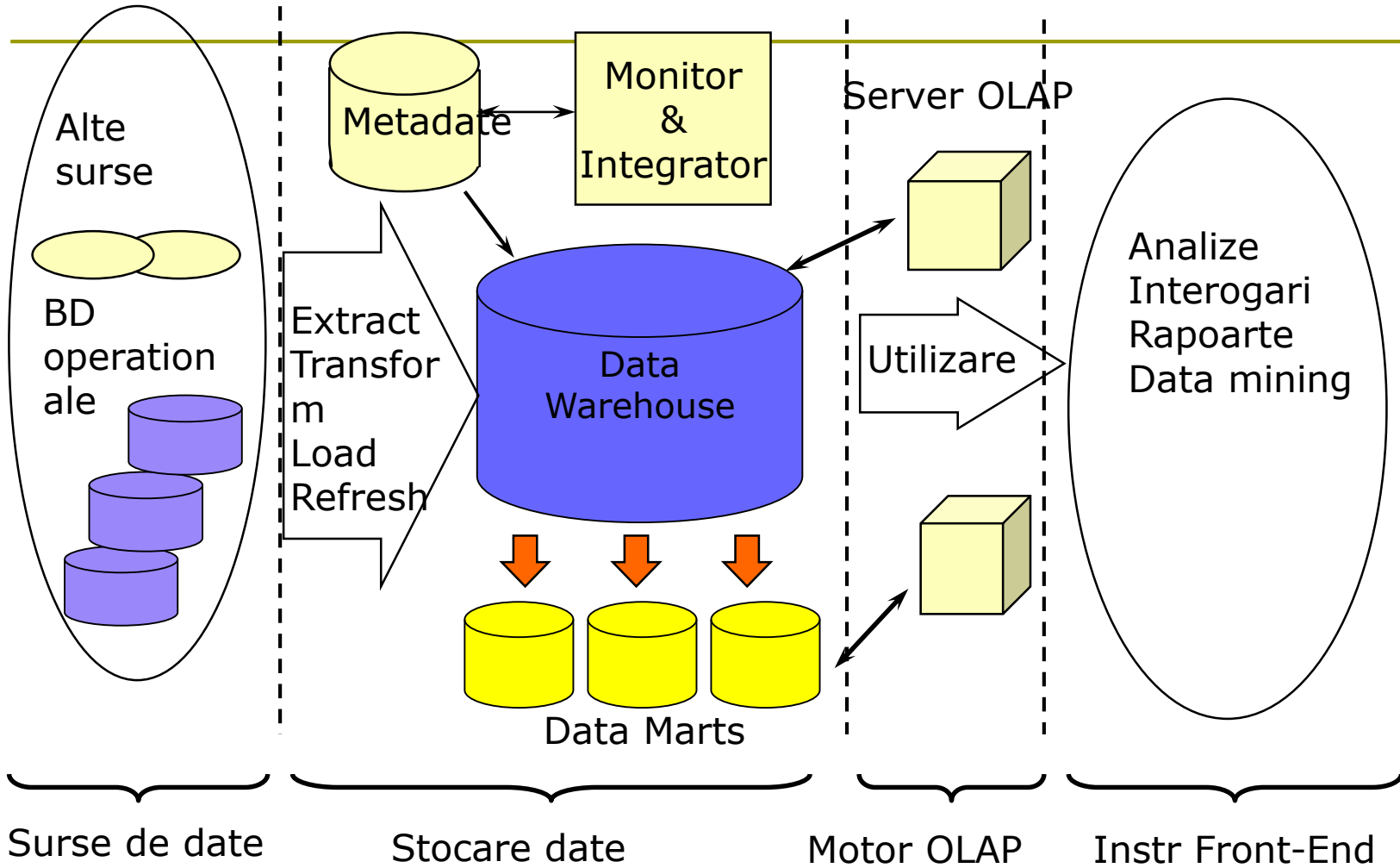
d.DW - nevolatil (Inmon)

- Un depozit separat fizic de date transformate din mediul operational
- In DW **nu au loc actualizari operationale ale datelor.**
 - Nu necesita mecanisme de procesarea tranzactiilor, recuperare si controlul concurentei
 - Sunt necesare doar 2 operatii pentru accesarea datelor :
 - **Incarcarea datelor si accesul la date.**
- **Actualizare** doar adăugarea periodică a unor date extrase din sistemele operationale
- Preocupare pt. **optimizarea accesului la date**:denormalizare, sumarizare, statistici ale accesării și reorganizare dinamică a indexării

Aplicatii ale depozitelor de date

- **Telecomunicatiile.**
 - folosirea retelei,
 - profilul clientilor care folosesc un anumit serviciu,
 - profitabilitatea produselor si serviciilor oferite.
- **Bancile** - gestionarea profitabilitatii generale, prin analiza profitabilitatii pe produs si pe client.
 - determinarea profilurilor clientilor pentru a directiona cât mai eficient campaniile de marketing.
- **Comertul cu amanuntul.**
 - trendul vânzarilor în functie de anotimp, vacante, campanii de publicitate, activitatea competitorilor.
 - mentalitatile si obiceiurile cumparatorilor = intrari in sistemul de dirijare a actiunilor promotionale si a altor campanii de marketing
 - Analiza trendului performantelor
 - Vânzari încrucisate
 - Profilul consumatorului si piata tinta.

Architectura multinivel



De ce un depozit de date separat?

□ Performante mai bune

- SGBD— potrivit pt OLTP: metode de acces, indexari, controlul concurenței, recuperare.
- Depozit—potrivit pt OLAP: cereri complexe, perspective multidimensionale, consolidare

□ Functii si date diferite

- Date: luarea deciziilor necesita date istorice
- Consolidarea datelor: luarea deciziilor necesita consolidari de date din surse eterogene
- Calitatea datelor: datele din surse diferite au reprezentari, codificari si formate diferite care trebuie reconciliate

Tipuri de DW

1 DEPOZITE DE ÎNTREPRINDERE (ENTERPRISE WAREHOUSE)

- întreaga structură organizațională;
- un volum extins de date: atât informații **detaliat**, cât și **agregate**;
- suporturi hardware performante;
- costurile și timpul de proiectare și implementare sunt considerabile.

2 DATA MART

- depozit de date specific unui anumit subset de cerințe sau unui departament din cadrul organizației;
- specifice unui domeniu de activitate;
- de regulă, datele conținute într-un data mart sunt **agregate**;
- costurile și termenele de implementare sunt considerabil reduse.

3 DEPOZIT VIRTUAL (VIRTUAL WAREHOUSE)

- o serie de **vederi (views)** realizate direct asupra BD operaționale;
- procesele de agregare pot afecta capacitățile de prelucrare ale serverelor utilizate în activitatea operațională;
- aparent ușor de implementat, necesită capacități de procesare deosebite;
- necesită curățare și consolidare în timpul rularii.