



INSTRUMENTE UTILIZATE PENTRU INTEGRAREA DATELOR

Cursul 6

AGENDA

1. Sisteme de gestiune a bazelor de date
2. Sisteme de gestiune a depozitelor de date
3. Instrumente pentru federalizarea datelor
4. Instrumente pentru migrarea datelor



1. SGBD

- **un ansamblu complex de programe care asigură interfața între o bază de date și utilizatorii acesteia.**
- **Rolul unui SGBD**
 1. ***a defini și descrie*** structura BD, printr-un limbaj propriu specific, conform unui anumit model de date;
 2. ***a încărca/valida*** datele în BD respectând restricțiile de integritate impuse de modelul de date utilizat;
 3. ***a realiza accesul la date*** pentru diferite operații (consultare, interogare, actualizare, editare situații de ieșire), utilizând operatorii modelului de date;
 4. ***a întreține*** BD cu ajutorul unor instrumente specializate (editoare, utilitare (shells), navigatoare, convertoare etc.);
 5. ***a asigura protecția*** BD sub aspectul securității și integrității datelor.



EXEMPLE DE SGBD

- Oracle
- DB2
- Informix
- Paradox
- MySQL
- Progress
- SQL Server
- Ingress II
- MS Access
- Visual FoxPro



NOI TIPURI DE BAZE DE DATE

- Baze de date NoSQL
- Baze de date in-memory



BAZE DE DATE NoSQL

Not
Only SQL

- BD NoSQL reprezintă o tehnologie de răspunde cerințelor de aplicații în cloud și proiectate să rezolve problemele de scalabilitate, performanță, modelare și distribuire limitată a datelor din bazele de date relaționale.
- nu impun o anumită schemă (structură a datelor), au un API simplu, sunt "eventual consistente" și pot gestiona o cantitate foarte mare de date
- **Teorema CAP** (Eric Brewer) – orice sistem distribuit de management al datelor poate să îndeplinească maxim două din următoarele trei proprietăți:
 - C: Consistență - toate nodurile sistemului informatic stochează aceleași date
 - A: Availability / Disponibilitate - orice cerere va primi un răspuns
 - P: Partitionare - sistemul continuă să funcționeze în condiții de partitionare a rețelei



CASSANDRA

- Dezvoltat initial de Facebook ca un proiect intern (2008).
- SGBD distribuit, orientat pe coloane
- Gratuit și "open-source", printre primele alegeri atunci când vorbim de soluții de "Big Data", fiind scalabilă linear, proiectată să gestioneze cantități mari de date structurate
- Avem proprietatile: consistența slabă, disponibilitate, partitionare
- Optimizat să ofere performanță foarte bună pentru date distribuite în mai multe centre de date cu replicare asincronă fără master - cu latență foarte scăzută;
- Utilizari: proiectul CERN ATLAS, WebEx (Cisco), Reddit, Twitter



CASSANDRA: MODELUL DATELOR

- E un sistem de stocare cheie-valoare structurat, extins cu două nivele de imbricare
- Modelul poate fi descris ca o serie de hash-map-uri imbricate, iar perechile cheie-valoare nu sunt stocate ca doua valori individuale, ci sunt cuplate intr-o clasa Column. Coloanele sunt grupate in familii de coloane, fiecare coloana avand: nume, valoare, marca de timp
- Tabela este harta multidimensionala (hash map) indexata dupa o cheie (cheia randului)
- Modelarea începe de la funcționalitățile aplicației
- Limbajul de interogare oferit de Cassandra (CQL, Cassandra Query Language) nu suportă multe dintre operațiile comune în SQL, cum ar fi "table joins" și "data aggregation"
- Tabelele pot fi create, sterse sau modificate la runtime fara a bloca update-uri sau interogari.



MONGODB

- Baza de date orientata pe documente;
- Este scrisă în C++, iar ca sistem de stocare, MongoDB utilizează un format similar JSON3 , și anume BSON (Binary JSON).
- Distribuit sub o combinatie de licenta GNU si Apache;
- Se poate face cautare dupa campuri, expresii regulate;
- Orice camp dintr-un document poate fi indexat;
- Oferă un grad ridicat de disponibilitate, performanță și scalare automată
- Utilizat pentru backend de Craigslist, eBay, Forsquare, SourceForge;



MONGODB – MODELUL DATELOR

- O înregistrare în MongoDB este un document ce prezintă o structură de tip câmpuri și valori atribuite acestora, ca de exemplu:

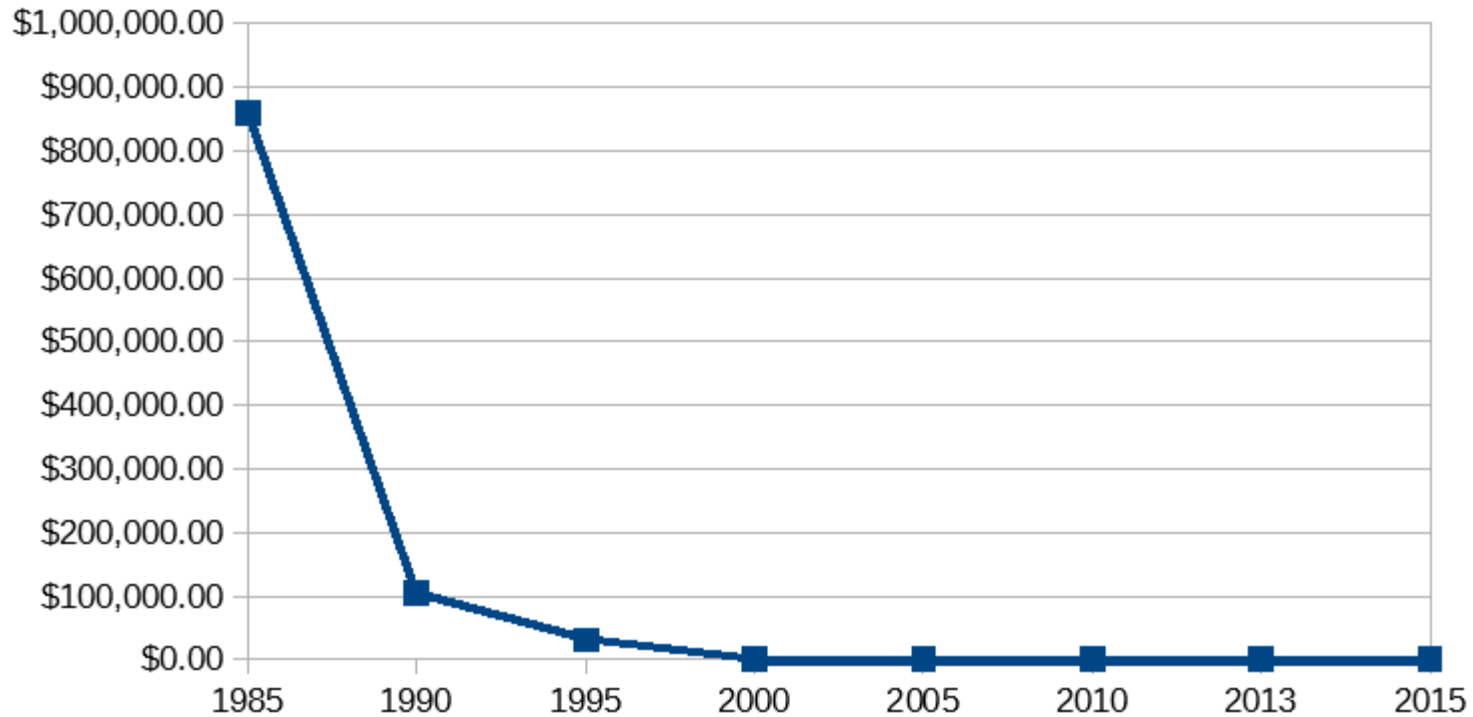
```
{  
  name: "sue",           ← field: value  
  age: 26,               ← field: value  
  status: "A",          ← field: value  
  groups: [ "news", "sports" ] ← field: value  
}
```

- Printre avantajele folosirii bazelor de date orientate pe documente se numără faptul că documentele corespund tipurilor de date native în mai multe limbaje de programare, iar schema dinamică a acestora poate suporta polimorfismul obiectelor.
- MongoDB asigură disponibilitatea datelor prin intermediul unor seturi de replici ale acestora, acestea asigurând totodată o redundanță crescută. Fiecare replica într-un set poate acționa ca replica primară sau secundară la un moment dat; atunci când replica primară devine disfuncțională, în mod automat e selectată acea replica secundară care să devină replica primară.



SISTEME DE DATE IN MEMORIE

Price per GB of RAM



DATA GRID VS DATABASE

	Data Grid	Database
Aplicație existentă	Necesită schimbări	Neschimbat
RBDMS existent	Neschimbat	Necesită schimbări
Viteză	Da	Da
Scalabilitate	Da	Nu



APACHE GEODE

- 📅 2002
- 📅 Java
- 📅 Domeniul bancar, folosit pentru a distribui modificări asupra portofoliilor traderilor
- 📅 Dezvoltat de Pivotal (Gemfire)
- 📅 Propriul limbaj de regasire a datelor (OQL)



APACHE IGNITE

📅 2007

📅 Java

📅 Dezvoltat de GridGain
Systems (GridGain)

📅 Suporta SQL(ANSI 99)

📅 Folosit de Apple (Apple
App Store)

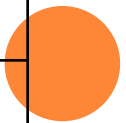


HAZELCAST

- 📁 Java
- 📁 Suporta partial SQL
- 📁 Oferă multiple structuri de date distribuite
- 📁 Se axează pe noi funcționalități



	Apache Ignite	Apache Geode	Hazelcast
Limbaj de programare	Java	Java	Java
Noduri omogene	Da	Da	Da
Stocarea datelor	Partitionat + Replicat	Partitionat + Replicat	Partitionat + Replicat
Off-Heap Memory management	Da	Da	Da
SQL	Da (ANSI SQL 99)	Nu (OQL)	Da (Partial)
J-Cache	Da	via Spring	Da



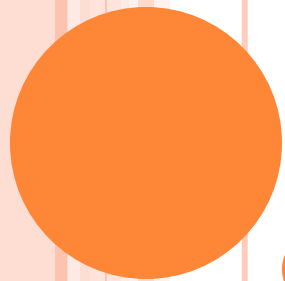
2. INSTRUMENTE DE GESTIUNE DW

- Oracle Data Integrator
- SAP BI



PIATA INSTRUMENTELOR BI





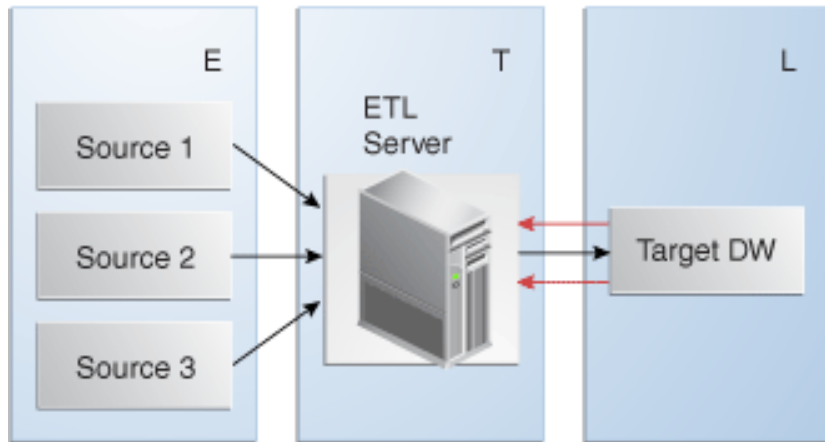
ORACLE DATA INTEGRATOR

ORACLE DATA INTEGRATOR ODI

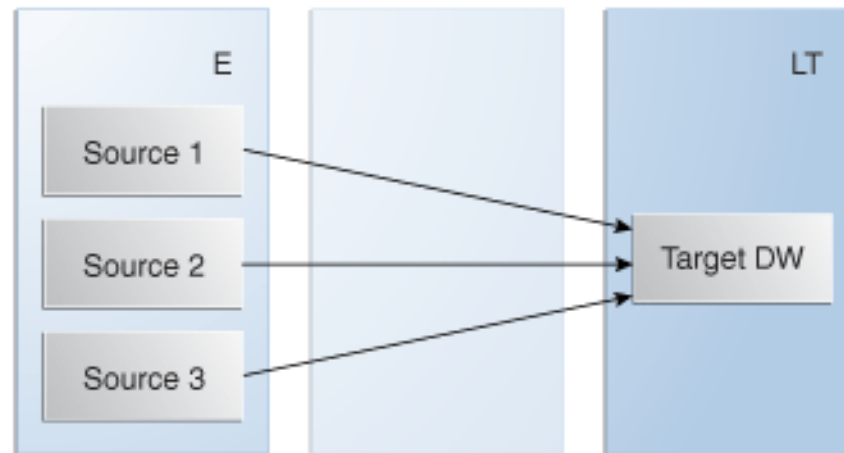
- Oracle Data Integrator ofera o tehnologie **Extract Load and Transform (ELT)** care imbunatateste performanta si reduce costurile de integrare a datelor.
- instrument performant de transfer, transformare si sincronizare a datelor intre sisteme informatice prin programe in timp real, in mod sincron si asincron;
- design modularizat, inovativ, conectivitate cu toate marile baze de date, aplicatii datawarehouse si aplicatii analitice



Traditional ETL Approach



Oracle Data Integrator E-LT Approach



KNOWLEDGE MODULES

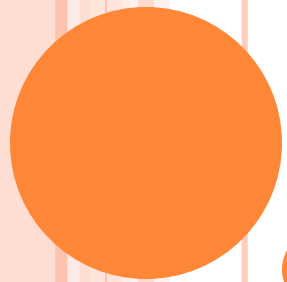
- componente ale tehnologiei Oracle Data Integrator Open Connector instruite sa proceseze un anumit set de task-uri pe o anumita tehnologie sau set de tehnologii
 - RKM (Reverse Knowledge Modules) – pentru reverse engineering al modelelor de date pentru o anumita tehnologie
 - LKM (Loading Knowledge Modules) – pentru a extrage date din tabellele sursa ale bazei de date si alte sisteme(fisiere, middleware, mainframe)
 - JKM (Journalizing Knowledge Modules) – folosite pentru a crea un log al modificarilor de date (insert, update si delete) a bazelor de date sursa pentru a contoriza modificarile
 - IKM (Integration Knowledge Modules)) – pentru a incarca date in tabellele tinta
 - CKM (Check Knowledge Modules) – pentru a verifica integritatea constrangerilor
 - SKM (Service Knowledge Modules) – pentru a genera cosul necesar pentru crearea serviciilor de date



MOD DE FUNCTIONARE

- **In faza de design:** se va alege o interfata si pentru fiecare dintre etapele ei (incarcare, verificare, etc) se vor specifica regulile functionale (mapari, constrangeri, etc.) si modulele de cunoastere care se vor folosi.
- **In faza de RUN :** ODI va folosi regulile functionale, KM-urile, optiunile KM-urilor si metadate continute in Repository(modele, topologii) pentru a genera automat un set de taskuri pentru a procesa jobul definit. Taskurile includ conexiune, managementul tranzactiei si codul adecvat pt task

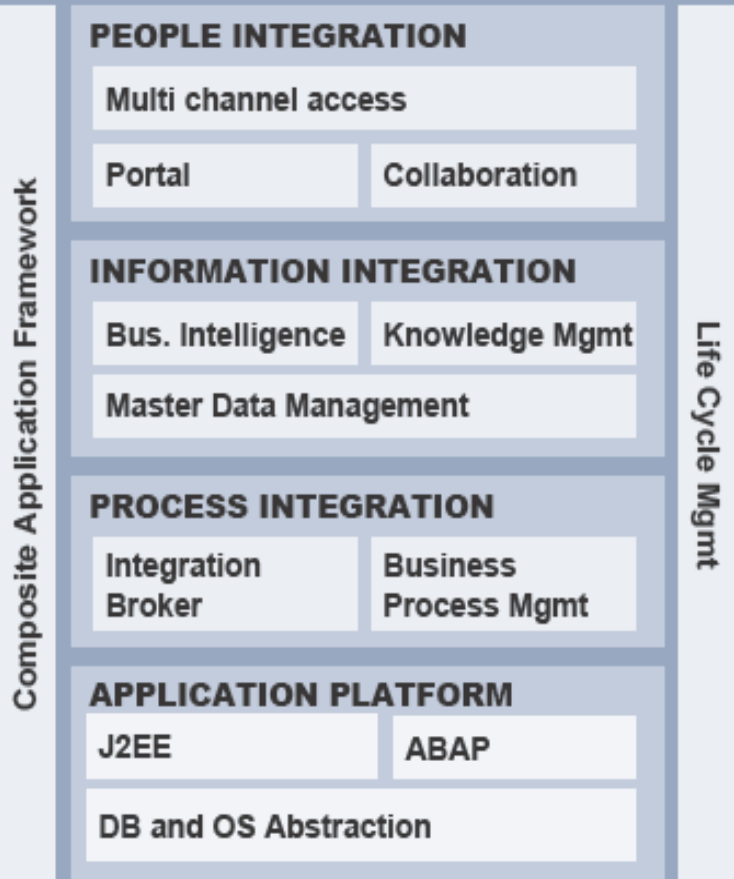




SAP BI

SAP NETWEAVER

SAP NetWeaver™



SAP Mobile Infrastructure

- Permite utilizarea aplicațiilor mobile să fie utilizate conectat sau deconectat

SAP Business Intelligence & Enterprise Portal

- Soluție completă pentru sistemul de analiză/raportare dintr-o companie

SAP Master Data Management

- Integritatea informației în ecosistemul unei companii

SAP Exchange Infrastructure

- Design, execuția și monitorizarea proceselor în cadrul aplicațiilor
- Workflows ad-hoc

- Interfața grafică: browser Web
- Scalabilitate și stabilitate, caching optimizat
- Independent de DB și SO

SAP BI (BUSINESS INFORMATION WAREHOUSE)

- **Business Content** = container ce cuprinde
 - Infocuburi (peste 420),
 - Query-uri (peste 1700),
 - Rapoarte si
 - Roluri utilizatorcu specific industrial si functional= solutii preconfigurate pentru diferite industrii
- **Extractori** („plug-in“) - extragerea datelor din SAP ERP si incarcarea in SAP BW
 - complet (full extraction) sau
 - partial (delta extraction).

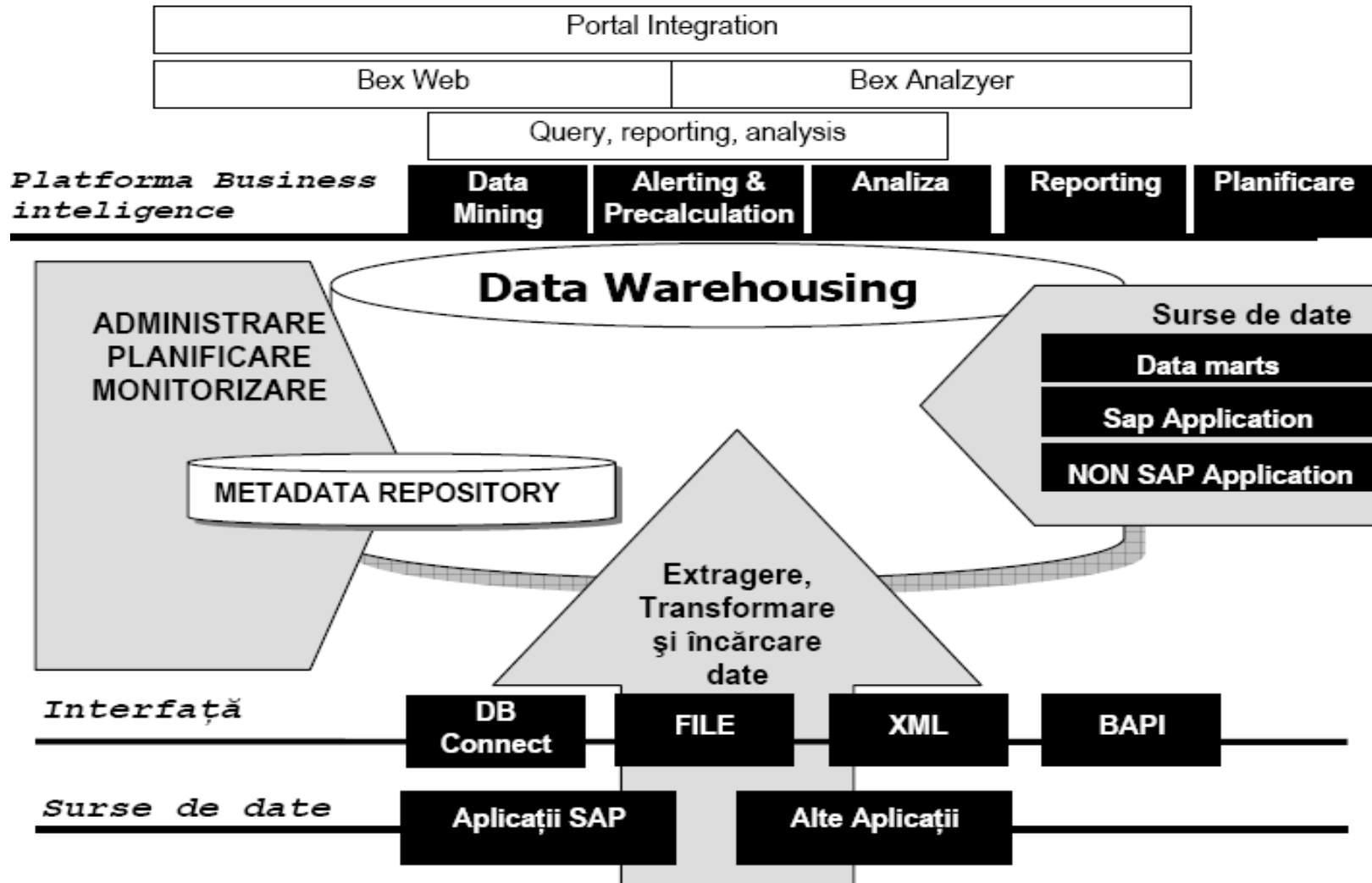


SAP BI ARHITECTURA PE 3 NIVELE

- **Business Explorer:** destinat interacțiunii cu utilizatorii de cunoștințe, prin interfața de raportare și analiză (Business Explorer)
 - BEx Browser (organizarea foilor de calcul în canale și clustere)
 - BEx Analyzer (crearea de rapoarte standard, personalizate și ad hoc).
- **Business Information Warehouse Server:**
 - **Administrator Workbench:** managementul SAP BW.
 - **Meta Data Repository** (Depozitul de metadate): metadate economice dar și tehnice.
 - **Staging Engine (Motorul pentru scenarii):** maparea datelor, coordonează procesul de transpunere în metadate.
 - **InfoCube (Infocubul):** depozitul central de date, unde sunt gestionate caracteristicile și indicatorii.
 - **Operațional Data Store (ODS):** depozitul central temporar al datelor interogate în cadrul sistemului BW.
 - **Procesorul OLAP :** analizează datele OLTP agregate.
- **Source Systems:** include sursele de date diverse: SAP ERP, sisteme non-SAP, fișiere de date, dar și sistemul BW



ARHITECTURA SAP BUSINESS INTELLIGENCE



INTEGRAREA DATELOR

- Integrare prin interfete: fisiere text, DB-Link & API.
- Procesul ETL in BW este greoi, dificil
- Adesea se prefera alte instrumente ETL
 - IBM DataStage
 - IBM Cognos



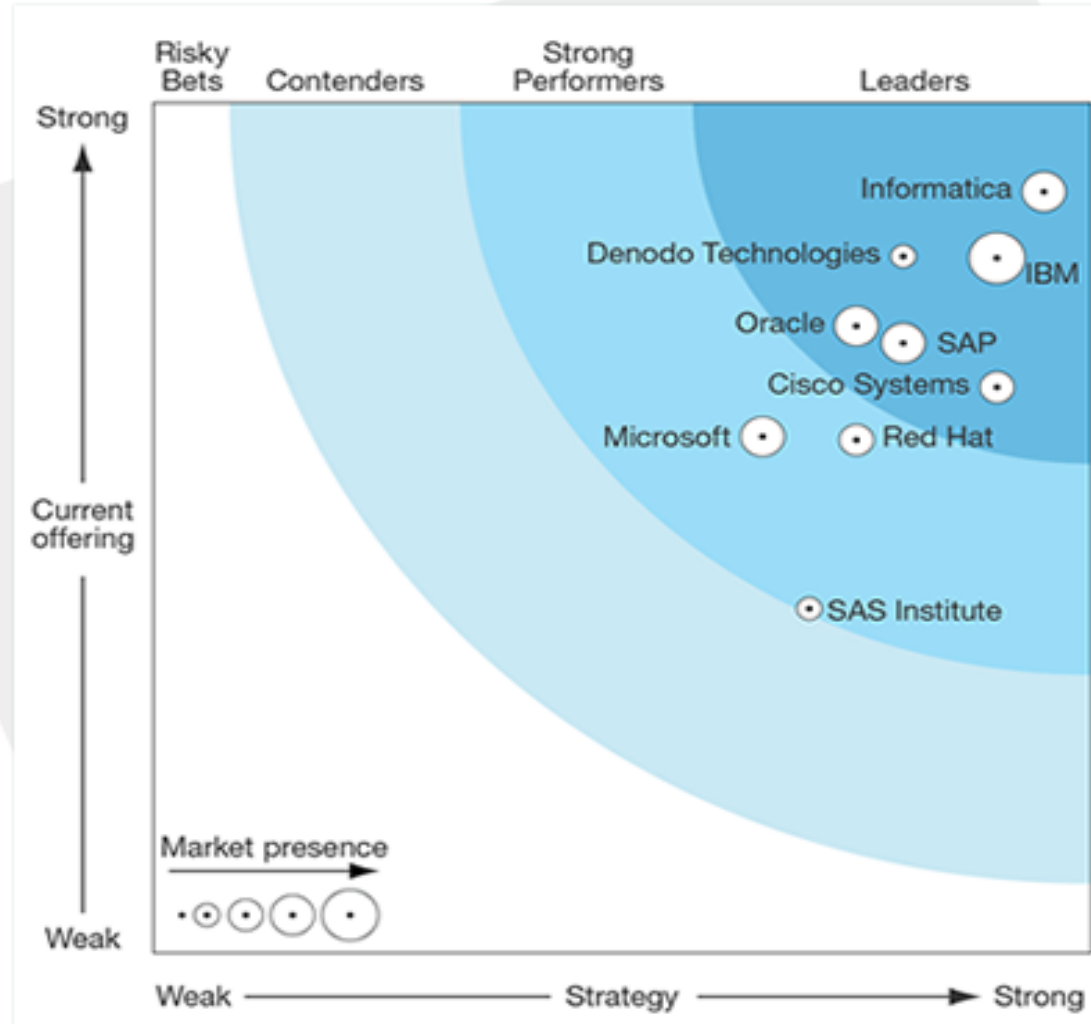
3. INSTRUMENTE DE FEDERALIZARE A DATELOR

Cele mai populare

- SAP BusinessObjects Data Federator;
- Sybase Data Federation;
- IBM InfoSphere Federation Server;
- Oracle Data Service Integrator;
- SAS Enterprise Data Integration Server.



CELE DE VARF ... (2015, FORRESTER REPORT)



METODE DE FEDERALIZARE

- **federated views** - view-uri create pe mai multe tabele sau surse de date cu ajutorul unor API; utilizate de aplicatii BI sau portaluri (folosind standarde JDBC, ODBC, ADO.NET)
- **data services** – ofera posibilitatea de federalizare atat pentru aplicatii SOA, cat si non-SOA
- **data mashups** – combinarea virtuala a datelor din diferite surse (rapoarte, dashboard, cuburi, structuri de date) si prezentarea rezultatelor ca obiecte reutilizabile afisate in mod interactiv intr-un browser (standarde deschise REST, SOAP OVER HTTP, JMS ETC).
- **caches** – patrare in memoria cache a rezultatelor interogarilor frecvente pt o viteza m. buna; pot fi sincronizate automat periodic
- **virtual data marts** – care folosesc abstractizarea pentru a obtine din depozitul de date exact datele pentru care ar fi fost construit un data mart fizic;



4. INSTRUMENTE PENTRU MIGRAREA DATELOR

- Oracle Migration Workbench
- Oracle SQL*Loader și Import/Export
- Ispirer SQLWays
- Embarcadero DT/Studio
- Microsoft Data Migration Assistant (DMA)
- Microsoft SQL Server Migration Assistant (SSMA)



ORACLE MIGRATION WORKBENCH

Obiectul	MS SQL Server & Sybase	MS Access	Informix	MySQL	DB2/400	DB2/UDB
Tabele	D	D	D	D	D	D
Viziuni	D	D (Interogari)	D	I	D	N
Indexes	D	D	D	D	D	D
Groupuri/Roluri	D	I	D	I	N	D
Utilizatori	D	N	D	D	D	D
Restrictii	D	D (Reguli)	D	D	D	D
Drepturi	D	N	D	D	D	D
Tipuri de date definite de utilizator	D	I	N/A	D	D	D
Proceduri stocate	D	I	D	I	N	N
Declanșatori	D	I	D	I	N	N
Embedded SQL	I	I	ESQL/C Pro*C	I	N	I
Alte opțiuni	N/A	Relații, tabele de legătură, reutilizarea codului aplicației.	Tipuri de date colecție, secvențe.	Support ENUM	N/A	Obiectele schemei migrate în obiecte utilizator

ORACLE SQL*LOADER ȘI IMPORT/EXPORT

