
A nagyméretű és nagy mennyiségű adatok kezelésének adatbázis technológiai következményei Vállalati információrendszerek **GIKOF 2013**

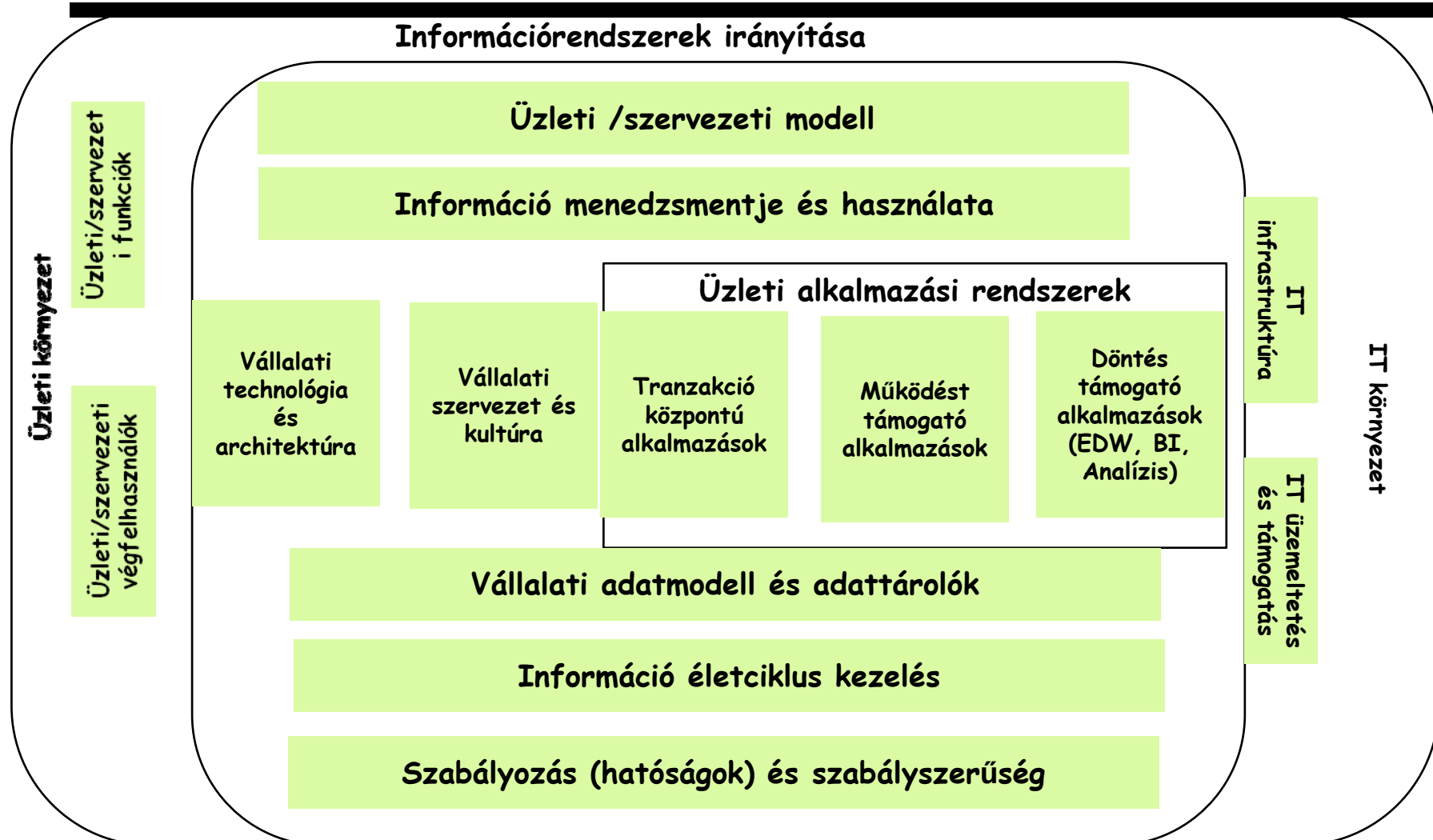
Molnár Bálint
ELTE, Információs rendszerek tanszék

A nagyméretű adatok fogalma



- Mik a nagyméretű adatok (Big data)
- 3V
 - (volumen, viharesebesség, változatosság)
 - Adatmennyiség= data volumes,
 - Nagysebességgel keletkezik = high velocity
 - Nagy változatosság variety;
- Amikor a hagyományos adatkezelési eljárások akadályokba ütköznek
- Inkább egy koncepció, mint egy pontos definíció, fogalom meghatározása

Vállalati információ kezelése (Vik)



A nagyméretű adatok kezelésének új megközelítései vállalati környezetben



- Hagyományos (vállalati) információrendszerek információkezelése
 - Strukturált adatok, tárolva:
 - Relációs adatbázis-kezelő;
 - Objektum-orientált adatbázis-kezelő;
 - Hálós adatbázis;
 - Hierarchikus adatbázis

A nagyméretű adatok kezelésének új megközelítései vállalati környezetben



- Az adatok nagy része ma strukturálatlan formában keletkezik;
- Az adatok 85%-90%
- Ezeket az adatokat nem aknázzák ki:
 - Nehéz elemezni (parsing);
 - Modellézni;
 - Értelmezni.

A nagyméretű adatok kezelésének új megközelítései vállalati környezetben



- A nagyméretű adatok keretében:
 - A strukturált;
 - Félig-strukturált;
 - A strukturálatlan;
 - A polistrukturált adatokat is kezelni kell.
 - Pl.: e-levél, web lap tartalom, videó, audió stb.
 - Adatbázis tartalom, napló állományok, XML állományok, strukturálatlan szöveges dokumentumok, web lapok, grafikák.

A nagyméretű adatok kezelésének új megközelítései vállalati környezetben



- Skálázhatóság
- Infrastruktúra menedzsment
- Végfelhasználó kielégítő kiszolgálása
- Adatmodellezés - *NoSQL adatbázisok kifejtése*
- ETL (Extract-Transform-Load) vállalati adattárházak hagyományos megoldásai - szűk keresztmetszet félig-strukturált adatok céljaira.
 - Adatok felfalása (data ingestion)
 - Map-reduce technológia a párhuzamos feldolgozásra (funkcionális programozás)

- Az információrendszerek dokumentum központúvá válnak
- Nem csak a nagy mennyiség és méret, hanem az adatszerkezet típusok magas fokú heterogenitása is probléma.
- Hagyományos rendszerszerkezési, elemzési és tervezési módszerek, információ kezelési eljárások nem kielégítőek

Adatbázisok evolúciója



- RDBMS (SQL, relációs adatbázis-kezelés)
- RDBMS oszlop-orientált (Sybase IQ/SAP)
 - Teljesítmény: lekérdezések kontra sor-orientált problémái
 - Üzleti intelligencia igényei, egyéb adat analitika igények;
 - Memória-alapú adatbázis kezelés (In-memory);
 - Grid (homogén csomópontok) illetve számítási felhő (heterogén csomópontok)

Adatbázisok evolúciója



- *Not Only SQL*, NoSQL
- DBMS
 - Teljesítmény:
 - Elosztott hálózati architektúra (distributed arch.)
 - Masszív párhuzamos, konkurens adatfeldolgozás (high concurrency)
 - Particionálással szembeni tűrés (partition tolerance)
 - Kiterjeszhetően skálázható arch. (Scale-out architecture)

Relációs kontra nem relációs adatbázisiskzelés



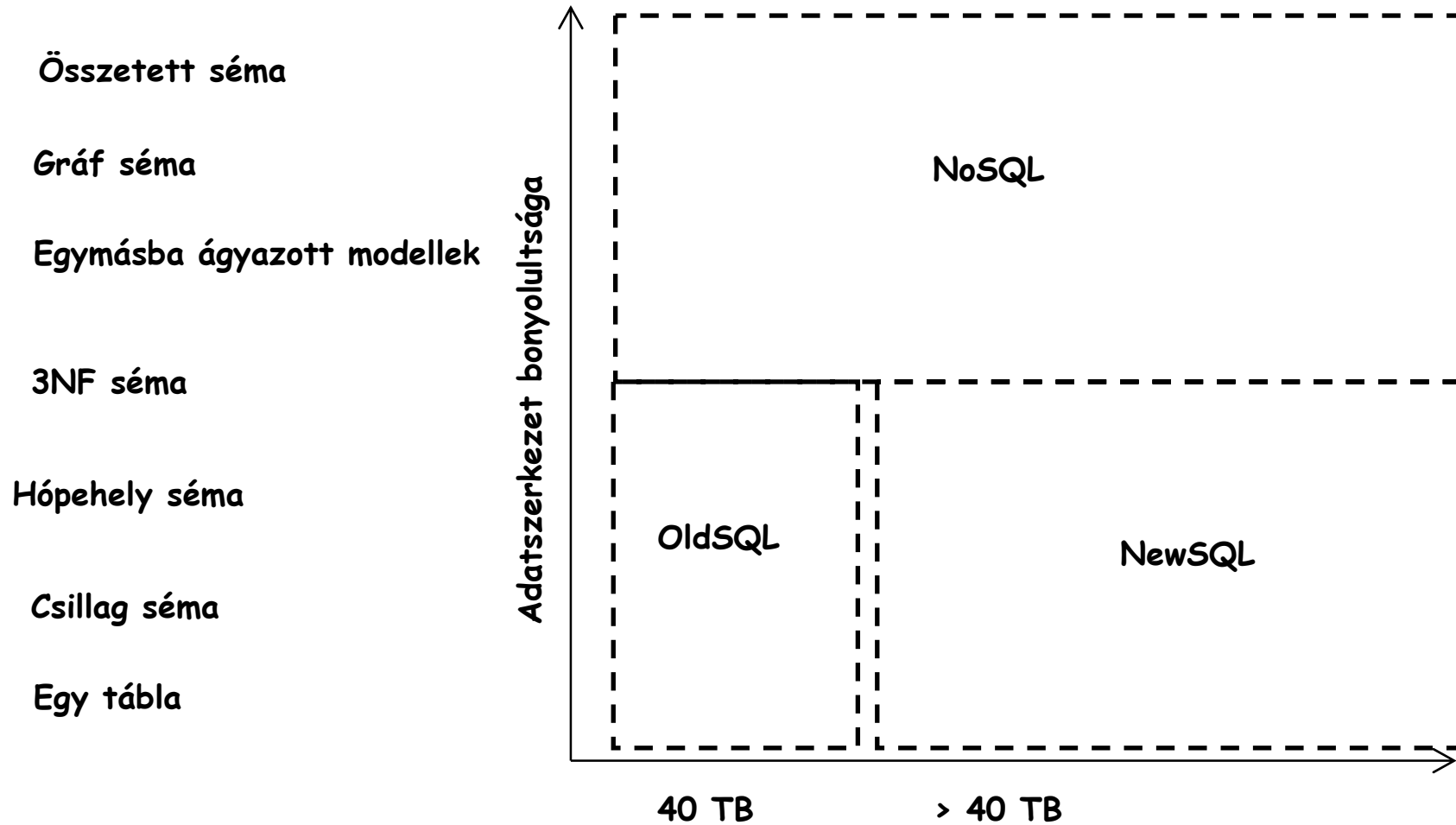
- Strukturált adat:
 - Minden adatelemre létezik meta-adat, SQL tárolás, elérés pontosan definiált.
- Strukturálatlan adat:
 - Az adatbázis séma nem írja le pontosan.

Relációs kontra nem relációs adatbázisiskzelés



- RDBMS :
 - 3NF, táblák, kapcsolatok, integritási feltételek. Oszlopok strukturált adat. Logikai adatmodell
- Nem relációs DBMS
 - Egymásba ágyazott,
 - hierarchikus adatszerkezetek
 - BoM, darabjegyzék, XML komplex dok.
 - A régi hierarchikus, és hálós adatbázisok, objektum, objektum-relációs adatszerkezetek, dokumentumok, XML adatszerkezetek, gráfok
- Új SQL (NewSQL)
- NoSQL

Alkalmazhatóság (OldSQL, NewSQLm NoSQL)



NoSQL

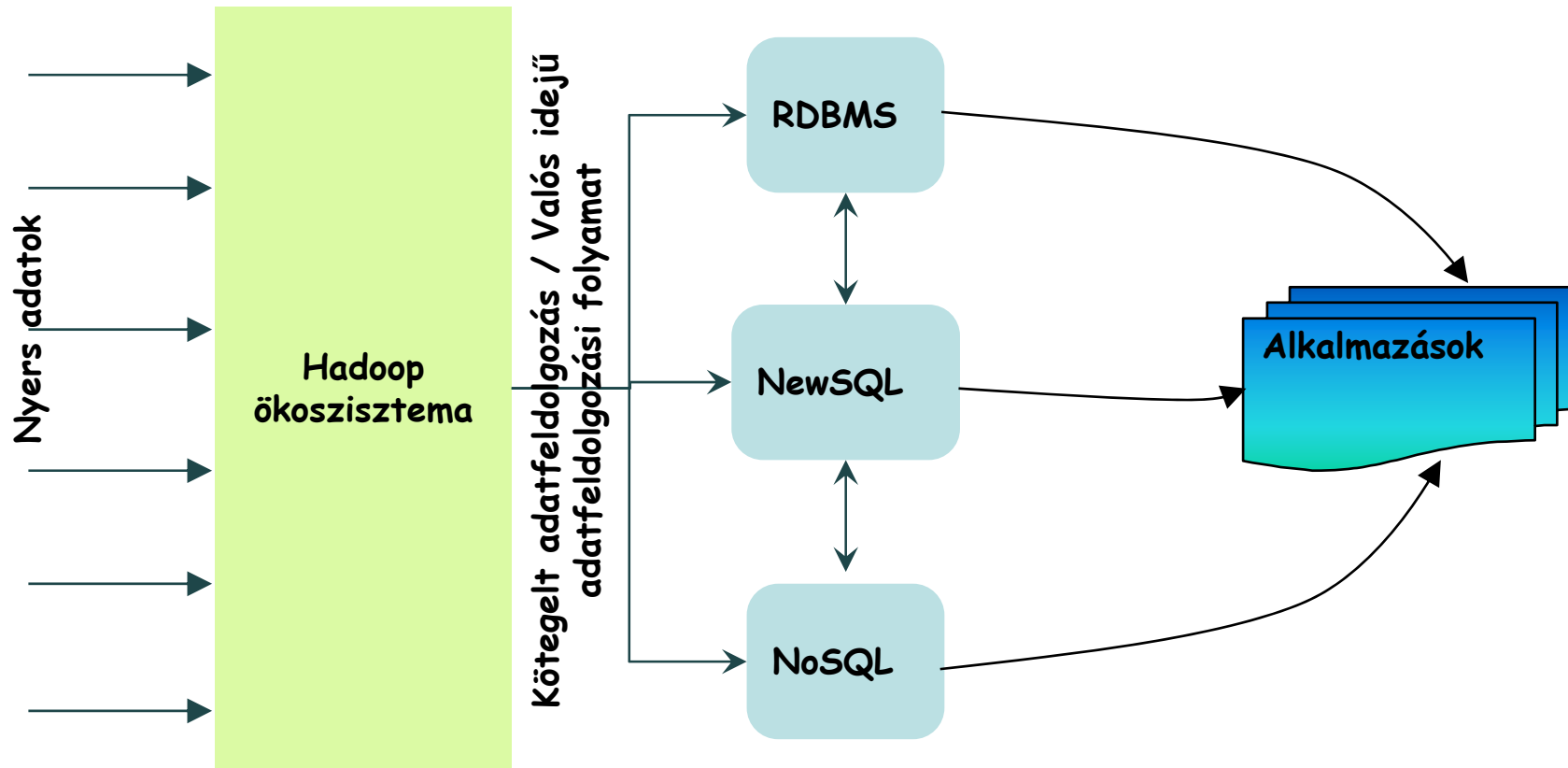
- Kulcs-érték pár (key-value pair)
- Gráf adatbázis
- Dokument adatbázis
- XML által definiált adatszerkezet
- JSON (JavaScript object notation)

Poliglott tartós tárolás (perzisztencia)

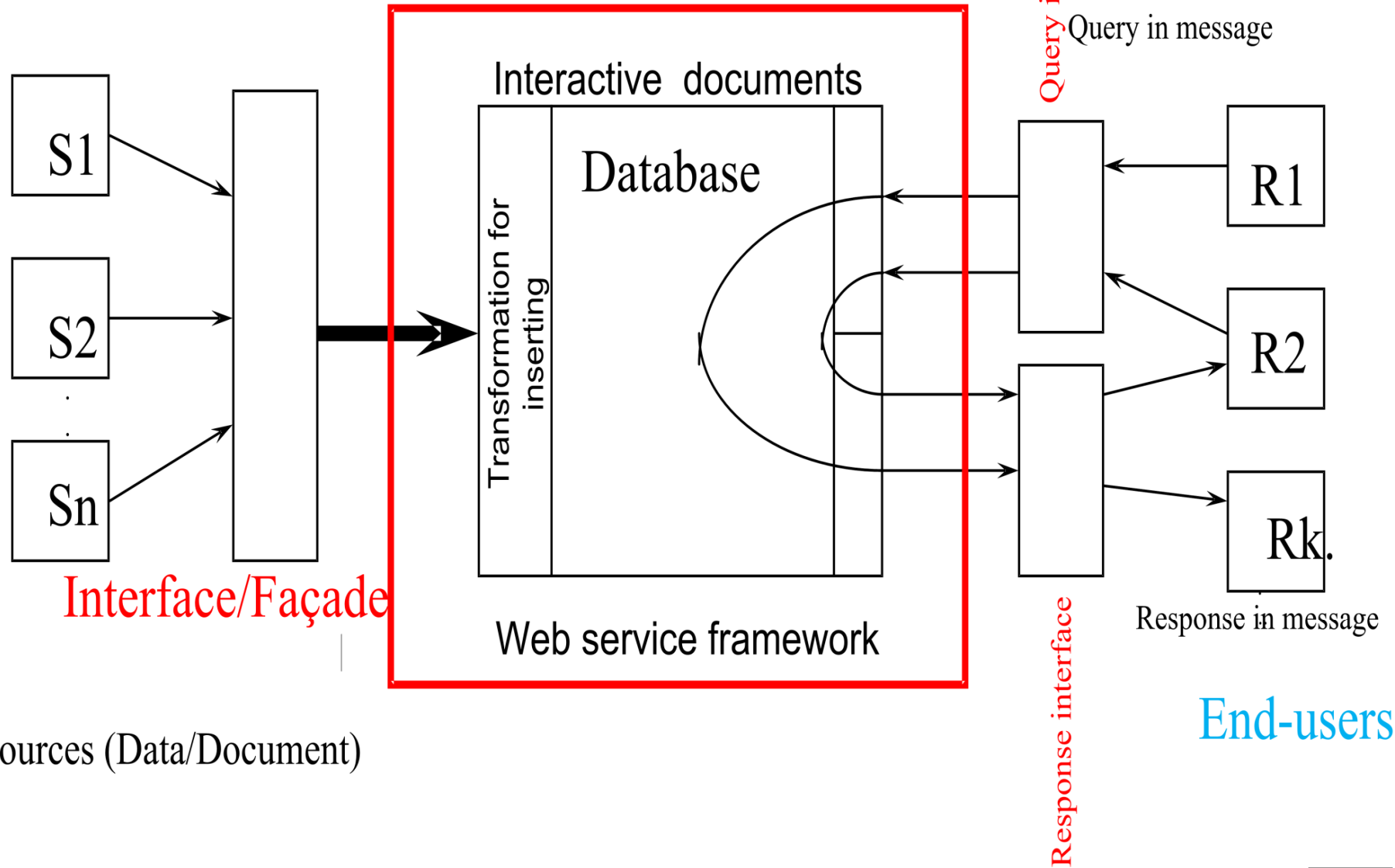


- RDBMS és egy vagy több NoSQL adatbázis kohabitációja az adatbázis-kezelés rétegében

Poliglott tartós tárolás (perzisztencia)







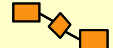
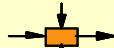
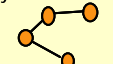
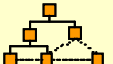


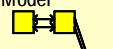
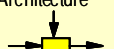
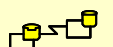

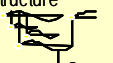

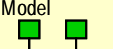


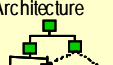
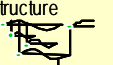









New message (Data /Document)



Sources (Data/Document)

End-users

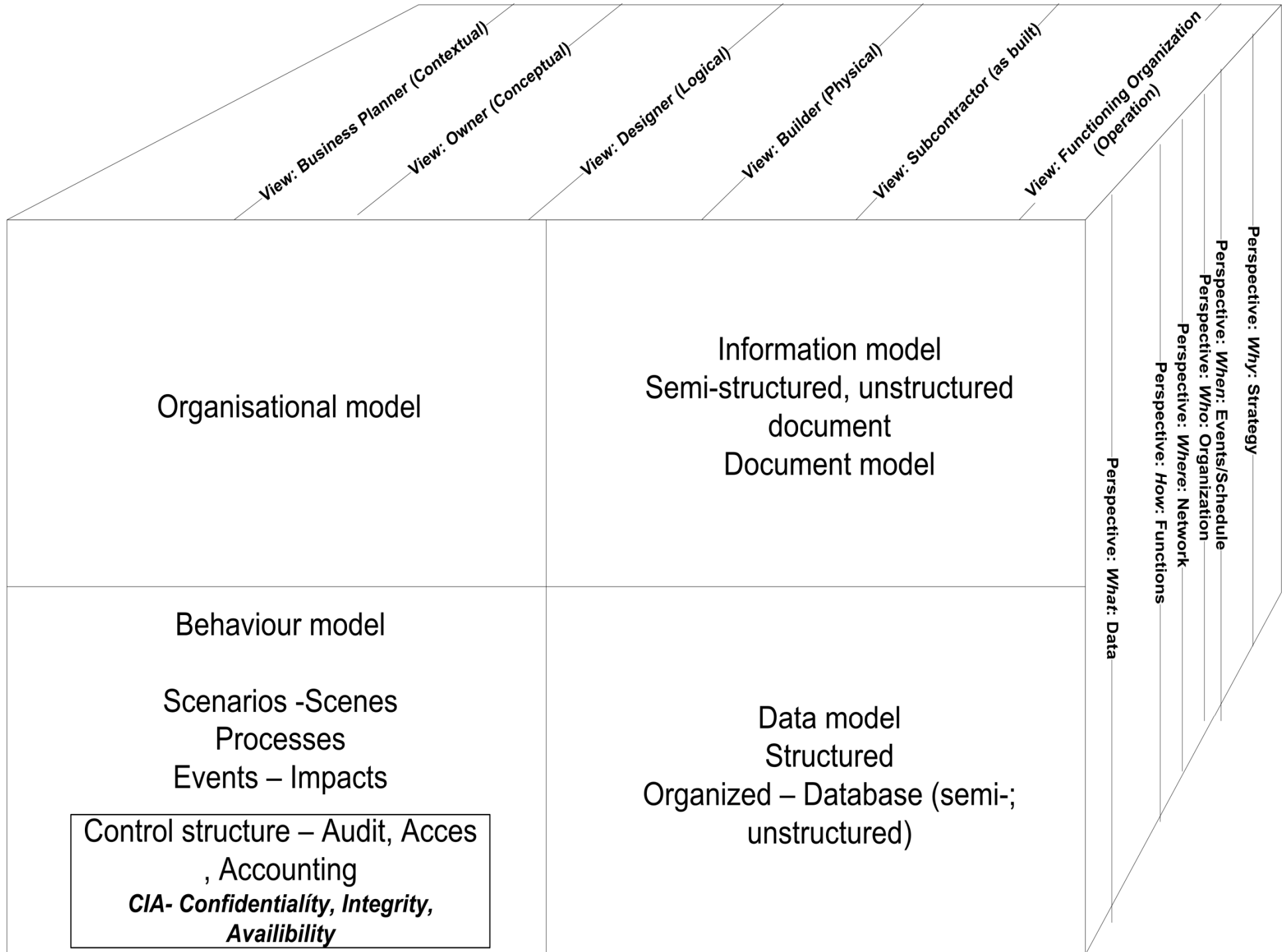
VA Enterprise Architecture	DATA <i>What</i>	FUNCTION <i>How</i>	NETWORK <i>Where</i>	PEOPLE <i>Who</i>	TIME <i>When</i>	MOTIVATION <i>Why</i>	Based on work by John A. Zachman
SCOPE (CONTEXTUAL) <i>Planner</i>	Things Important to the Business  Entity = Class of Business Thing	Processes Performed  Function = Class of Business Process	Business locations  Node = Major Business Locations	Important Organizations  People = Major Organizations	Events Significant to the Business  Time = Major Business Event	Business Goals and Strategy  Ends/Mears = Major Business Goals	SCOPE (CONTEXTUAL) <i>Planner</i>
ENTERPRISE MODEL (CONCEPTUAL) <i>Owner</i>	Semantic Model  Ent = Business Entity Rel = Business Relationship	Business Process Model  Proc = Business Process I/O = Business Resources	Business Logistics System  Node = Business Location Link = Business Linkage	Work Flow Model  People = Organization Unit Work = Work Product	Master Schedule  Time = Business Event Cycle = Business Cycle	Business Plan  End = Business Objective Means = Business Strategy	ENTERPRISE MODEL (CONCEPTUAL) <i>Owner</i>
SYSTEM MODEL (LOGICAL) <i>Designer</i>	Logical Data Model  Ent = Data Entity Rel = Data Relationship	Application Architecture  Proc = Application Function I/O = User Views	Distributed System Architecture  Node = IS Function Link = Line Characteristics	Human Interface Architecture  People = Role Work = Deliverable	Processing Structure  Time = System Event Cycle = Processing Cycle	Business Rule Model  End = Structural Assertion Means = Action Assertion	SYSTEM MODEL (LOGICAL) <i>Designer</i>
TECHNOLOGY MODEL (PHYSICAL) <i>Builder</i>	Physical Data Model  Ent = Segment/Table Rel = Pointer/Key	System Design  Proc = Computer Function I/O = Data Elements/Sets	Technology Architecture  Node = Hardware/Software Link = Line Specifications	Presentation Architecture  People = User Work = Screen Format	Control Structure  Time = Execute Cycle = Component Cycle	Rule Design  End = Condition Means = Action	TECHNOLOGY MODEL (PHYSICAL) <i>Builder</i>
DETAILED REPRESENTATIONS (OUT-OF-CONTEXT) <i>Sub-Contractor</i>	Data Definition  Ent = Field Rel = Address	Program  Proc = Language Statement I/O = Control Block	Network Architecture  Node = Addresses Link = Protocols	Security Architecture  People = Identity Work = Job	Timing Definition  Time = Interrupt Cycle = Machine Cycle	Rule Design  End = Sub-Condition Means = Step	DETAILED REPRESENTATIONS (OUT-OF-CONTEXT) <i>Sub-Contractor</i>
FUNCTIONING ENTERPRISE	Data Ent = Rel =	Function Proc = I/O =	Network Node = Link =	Organization People = Work =	Schedule Time = Cycle =	Strategy End = Means =	FUNCTIONING ENTERPRISE
	DATA <i>What</i>	FUNCTION <i>How</i>	NETWORK <i>Where</i>	PEOPLE <i>Who</i>	TIME <i>When</i>	MOTIVATION <i>Why</i>	

	Structure (What)	Activities (How)	Locations (Where)	People (Who)	Time (When)	Motivation (Why)
Objectives/ Scope (Planner's View)	Most significant business concepts	Mission	International view of where organization operates	Human resource philosophies and strategies	Annual planning	Enterprise vision
Enterprise Model (Business Owner's View)	Business language used	Strategies and high-level business processes	Offices and relationships between them	Positions and relationships between positions	Business events	Goals, objectives, business policies
Model of Fundamental Concepts (Architect's View)	Specific entities and relationships between them	Business functions and tactics	Roles played in each location and relationships between roles	Actual and potential interactions between people	System events	Detailed business rules
Technology Model (Designer's View)	System representation of entities and relationships	Program functions/ operations	Hardware, network, middleware	User interface design	System triggers	Business rule design
Detailed Representation (Builder's View)	Implementation strategy for entities and relationships	Implementation design of functions/ operations	Protocols, hardware components, deployed software items	Implementation of user interface	Implementation of system triggers	Implementation of business rules
Functioning System	Classes, components, tables, ...	Deployed functions/ operations	Deployed hardware, middleware, and software	Deployed user interface (including documentation)	Deployed systems	Deployed software

↑
Computation Independent Models

↑
Platform Independent Models (PIMs)

↓
Platform Specific Models (PSMs)



A schematic mapping between Zachman architecture and requirements and constraints in the framework of Axiomatic Design



Aspects	Entities	Activities	Locations	People	Time	Motivation	
Perspectives							
Contextual	UR Intensional document	UR	UR	UR Intensional document	UR	UR Intensional document	Scope
Conceptual	FR Document hierarchy	FR Document hierarchy	FR	FR Document hierarchy	FR	FR Document hierarchy	Enterprise Model
Logical	DC&P	DC&P	DC&P	DC&P	DC&P	DC&P	System Model
Physical	DC&P	DC&P	DC&P	DC&P	DC&P	DC&P	Technical Model
Detailed representation (Actual implementation and operation)	IF&OV	IF&OV	IF&OV	IF&OV	IF&OV	IF&OV	Components
Functioning enterprise/organizat ion	Data	Function	Network	Organization	Schedule	Strategy	

-Köszönöm a figyelmet



...Kérdések?