



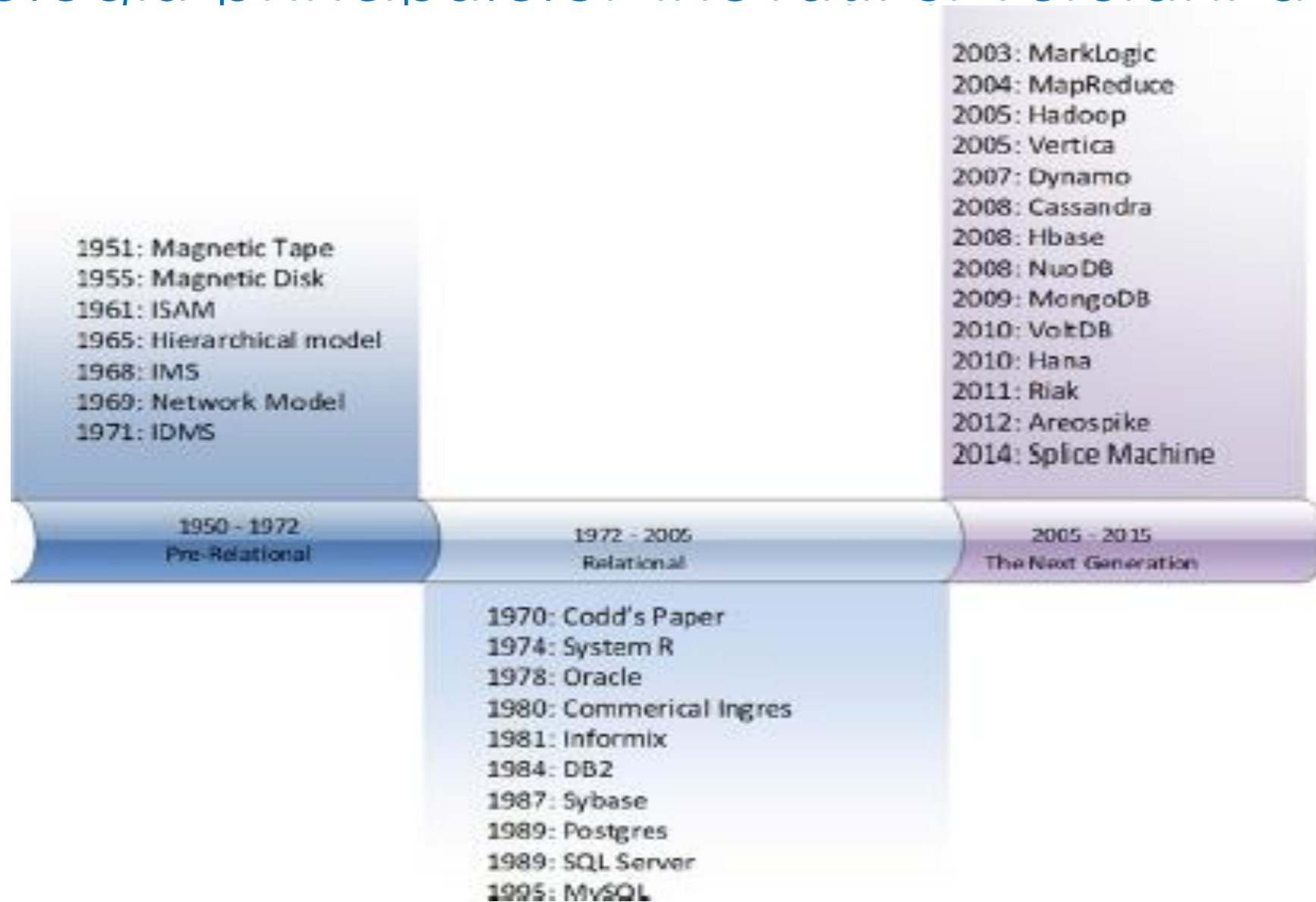
Azure CosmosDB

Conf.dr. Cristian KEVORCHIAN

Universitatea din București

Facultatea de Matematică și Informatică

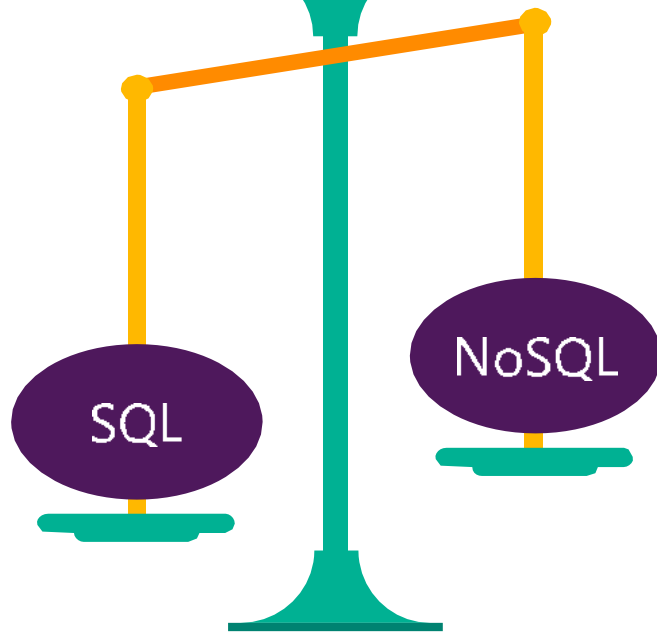
Cronologia principalelor inovatii si versiuni de SGBD



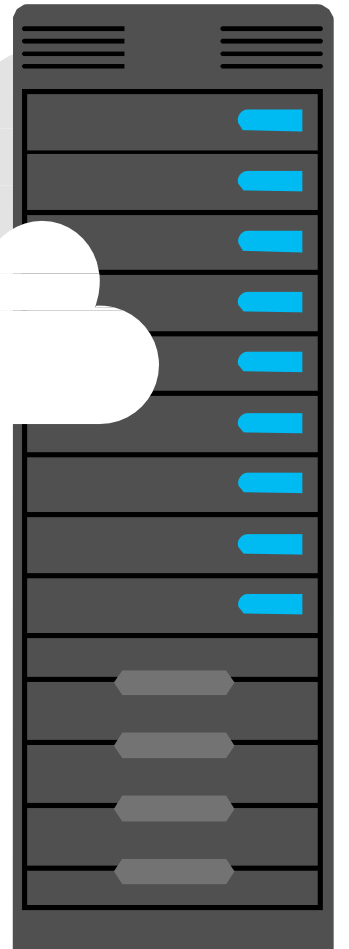
NoSQL-Definitie(IBM)

"NoSQL, care înseamnă "nu numai SQL", este o nouă abordare a proiectării bazelor de date care oferă o variantă de schemă flexibilă pentru stocarea și regăsirea datelor dincolo de structurile tradiționale de tabelă indentificate în bazele de date relaționale (Codd, 1970).

În timp ce bazele de date NoSQL există de mai mulți ani, **epoca cloud, big data și noua generație de aplicații web și mobile** le-a conferit reala consacrare tehnologică. Alegerea unei soluții noSQL se bazează pe attribute cum ar fi scalarea, performanța și ușurința implementării. Cele mai frecvente tipuri de baze de date NoSQL sunt bazele de date de tip cheie-valoare, document, coloană și grafic.



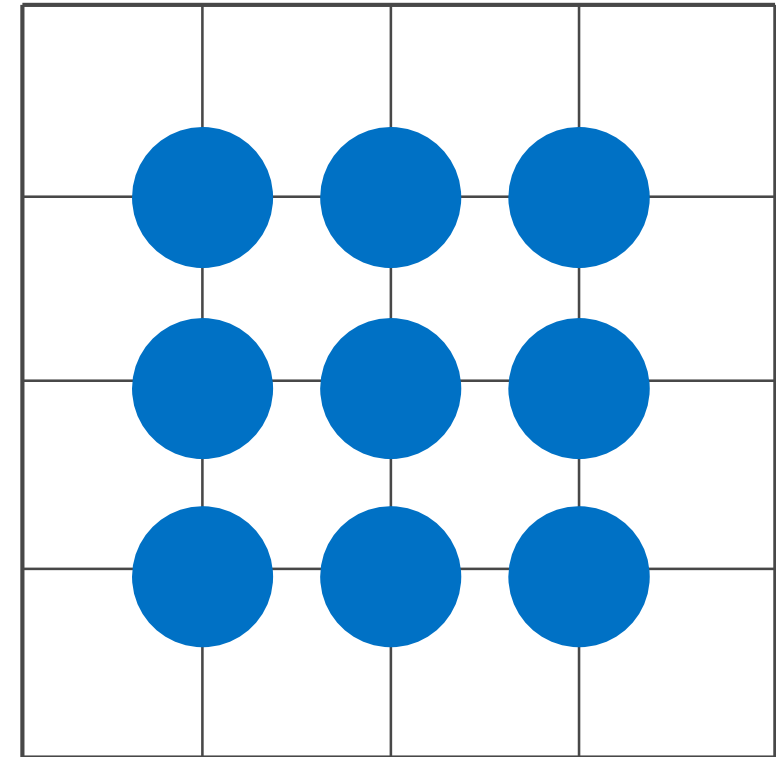
"Choose the store that best fits your research needs"



Abordarea Relațională

Datele sunt încărcate în tabele care includ:

- Schema
- Coloanele
- Liniile



Abordarea NoSQL: varietatea tipurilor de stocare

O bază de date NoSQL utilizează patru categorii de stocare :

Key value

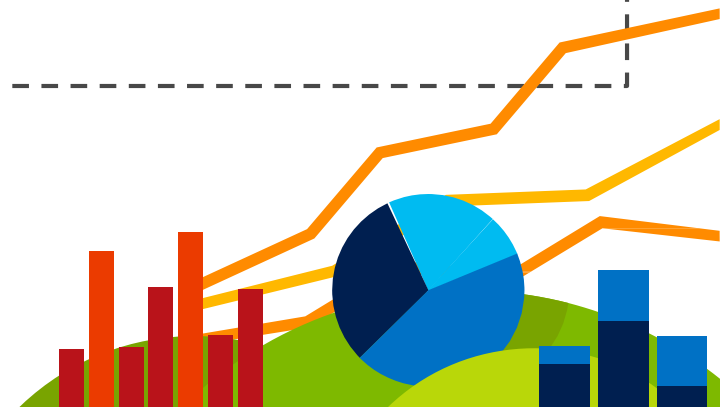
Wide column

Document

Graph

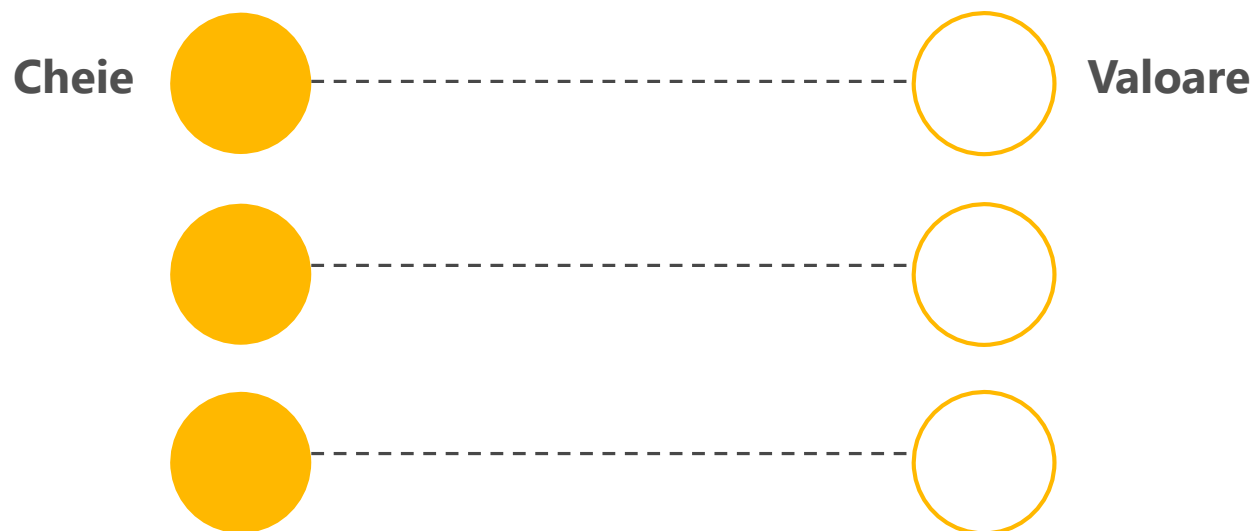
Azure CosmosDB(fostul DocumentDB)

Are în componere: cheie-valoare, columnar, document și graf



Stocare cheie-valoare

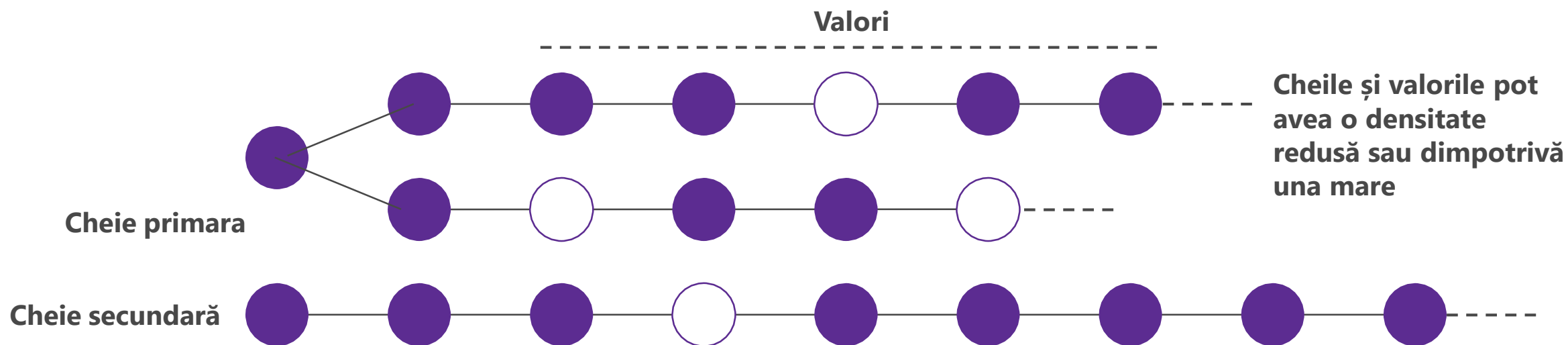
Stocarea cheie-valoare oferă o variantă de stocare rapidă prin intermediul unui model de date de complexitate redusă - orice poate fi stocat drept valoare, atât timp cât fiecare valoare este asociată unei chei sau nume.



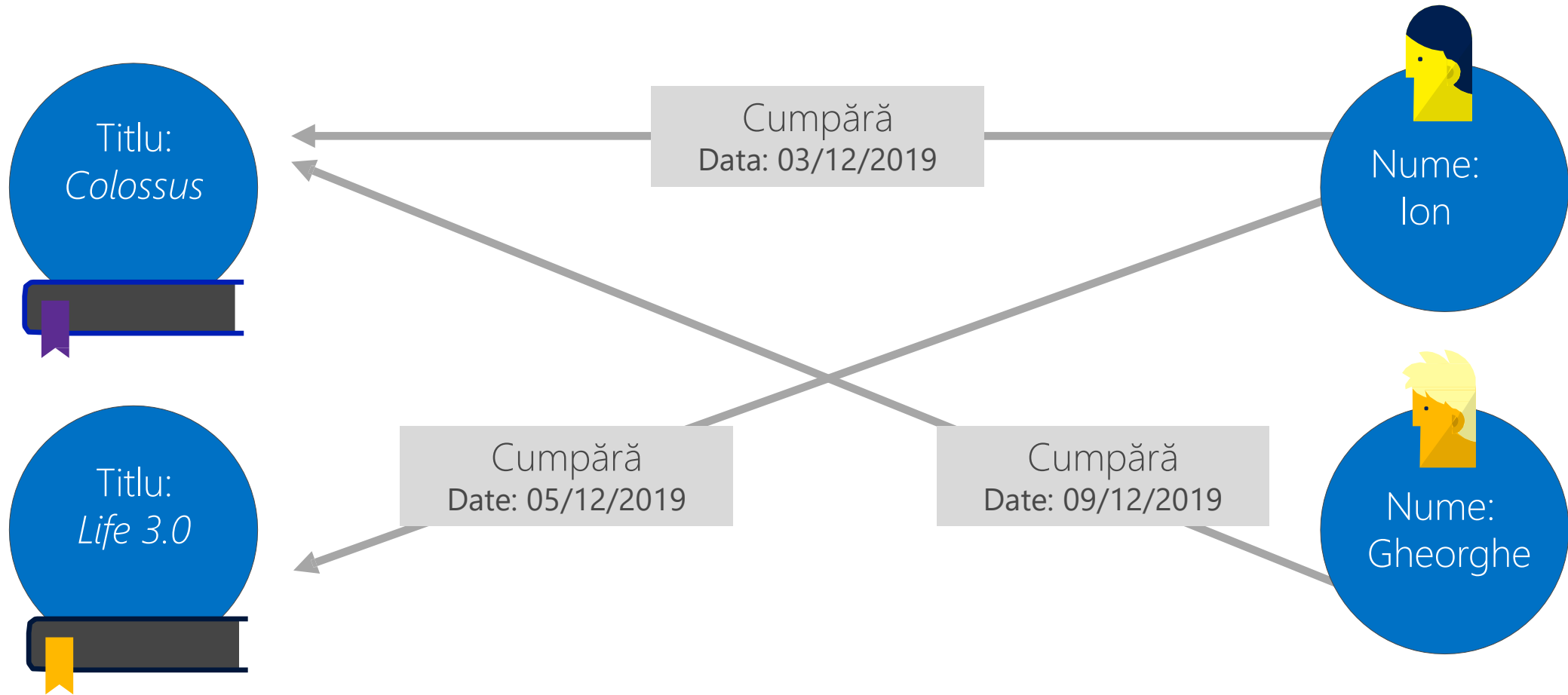
Stocare de tip columnar

Bazele de date din categoria BigTable precum HBase sau Hypertable sunt de tip columnar și au o schemă ce trebuie predefinită.

Datele sunt stocate în celule grupate pe coloane. Coloanele sunt grupate logic în "column families". Acestea pot conține un număr teoretic nelimitat (limitat în funcție de implementarea specifică) de coloane ce pot fi create la runtime sau la definirea schemei.



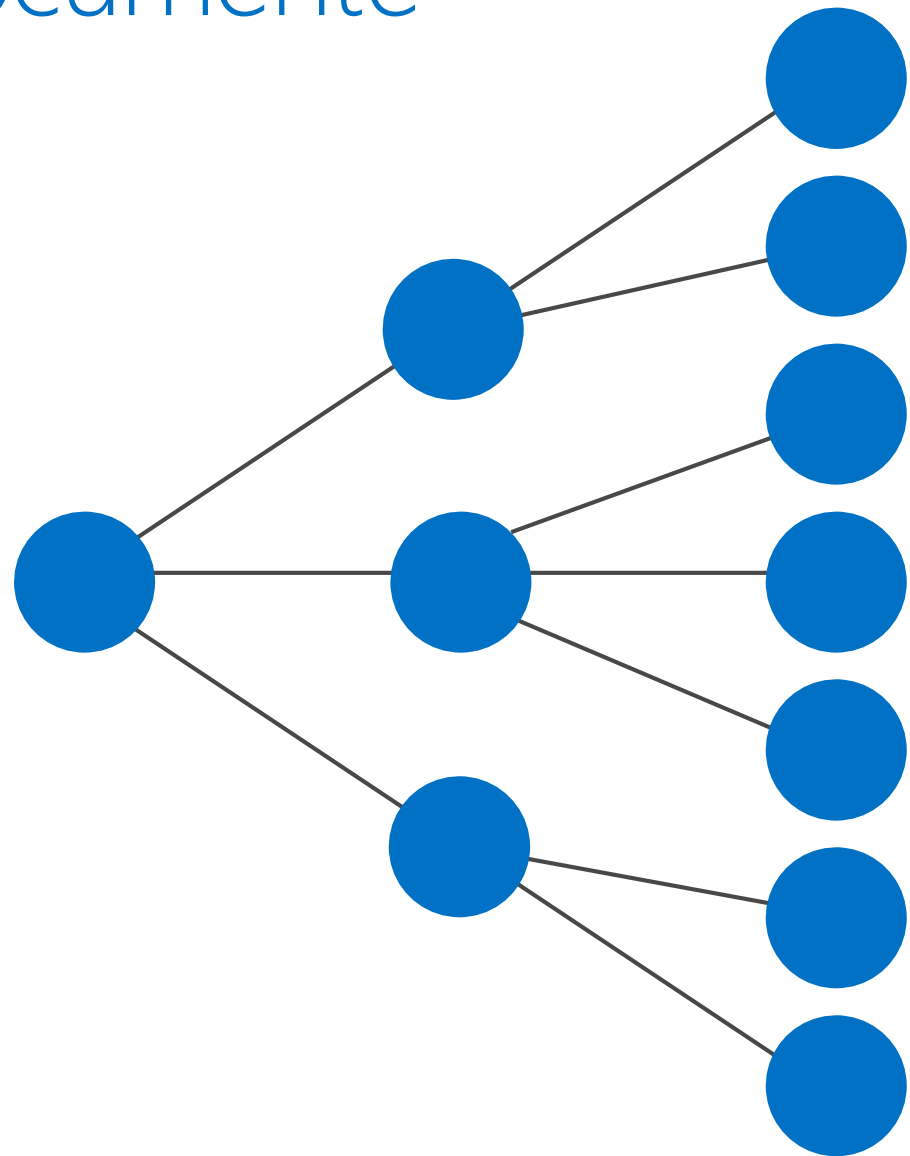
Baze de date Graf



Spații de stocare pentru documente

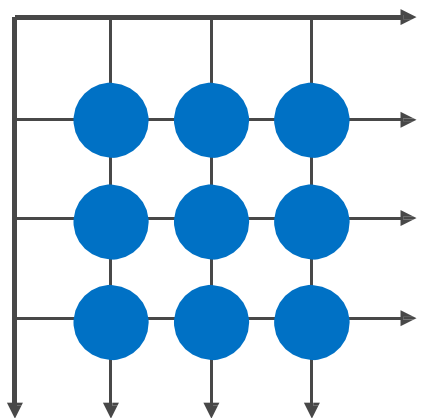
Spațiile de stocare documente conțin obiecte ca date ierarhizate în structuri arborescente(cele mai cunoscute fiind JavaScript Object Notation [JSON] și Extensible Markup Language [XML]).

De notat faptul că aceste nu sunt documente Microsoft Word!



Rațiunea evoluției NoSQL

Drivers



Operational efficiency






Resiliency

Scalability

**Speed and variety of
cloud-based data**

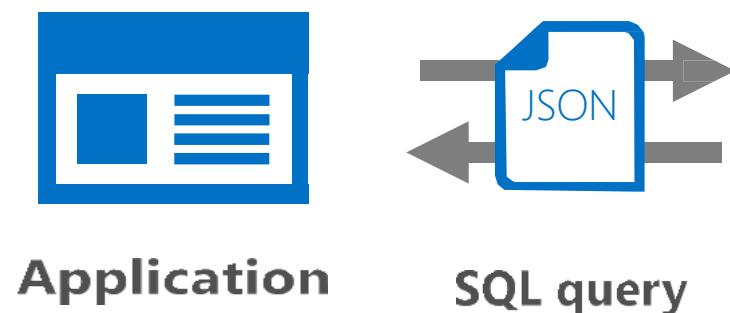


SQL și NoSQL: each has its place

SQL Server in a VM	Azure SQL Database	Azure DocumentDB	Apache HBase	Tables/Blobs
				
Func.Complete DBMS				
"RichQuery"				
Proces Transacțional				
				"Managed as a service"
				Elastic scale
				Schema-free data model
				Internet-accessible http/rest
				Arbitrary data formats

Azure DocumentDB

O bază de date document NoSQL livrată ca serviciu cu management complet oferit de Azure



Adaptat **arhitecți și dezvoltatori cloud** care operează soluții NODQL enterprise-ready document database

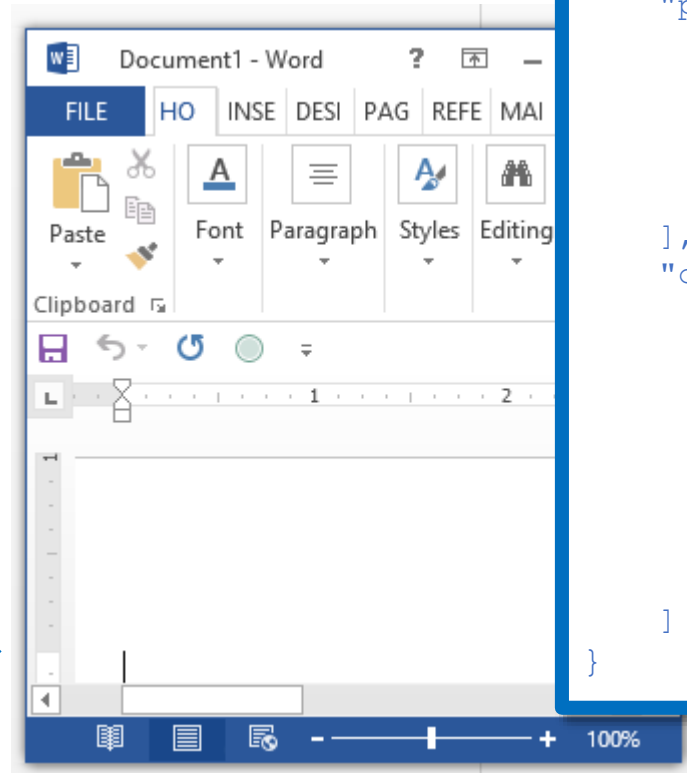
Documente

Optimizat pentru: stocare scheme-agnostică JSON pentru scalare date ierarhice și denormalizate.



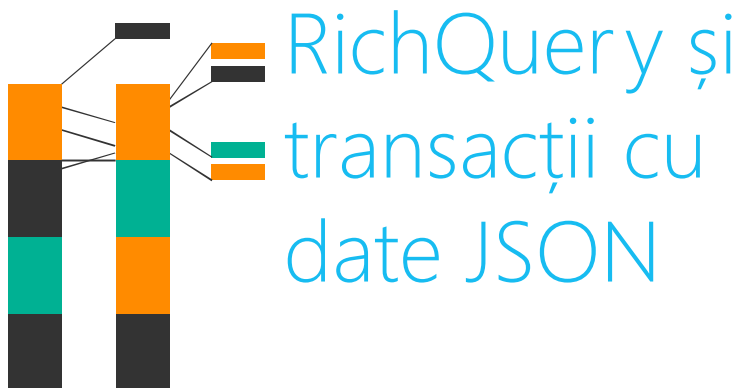
```
{
  "name": "SmugMug",
  "permalink":
  "smugmug",
  "homepage_url":
  "http://www.smugmug.com", "blog_url":
  "http://blogs.smugmug.com/",
  "category_code": "photo_video",
  "products": [
    {
      "name": "SmugMug",
      "permalink":
      "smugmug"
    }
  ],
  "office_address": "67 E. Evelyn Ave",
  { : Evelyn "",
    "address_line1": "94041",
    "city": "Mountain View",
    "state_code": "CA", "country_code": "USA", "latitude": 37.390056,
    "longitude": -122.067692
  }
}
```

NU este vorba de documente Word



Detalii privind Azure DocumentDB

Ideal pentru aplicații proiectate pentru cloud pentru următoarele tipuri de implementări:



- **Query JSON doc fără indici secundari**



- Monitorizarea consistenței
- Scalare elastică



RAD

- **Tool-uri cunoscute ușor de utilizat**—REST, JSON, JavaScript

Suportă nativ JavaScript, SQL query, și transacții peste documente JSON

Proprietăți importante

Indexare Automată

- Toate proprietățile documentului sunt disponibile pentru interogări
- Nu este necesară schema sau indexul secundar

Managed service

- Spin-up la cerere fără nicio configurare
- SLA de 99.95%
- Integrare cu Azure HDInsight și Azure Search

Auto-scalare/sharding

- Scalabilitate și fiabilitate îmbunătățite datorită distribuției seturilor de date mari pe mai multe mașini

SQL query language

- Poate fi utilizat atât SQL cât și .NET LINQ

Proprietăți de top

Tranzacții ACID (Atomicitate, Consistență, Izolare și Durabilitate)

- Modele simple de implementat
- JavaScript pentru acțiuni CRUD

Drepturi de acces granulare

- Permite accesul la toate documentele din cadrul colecțiilor

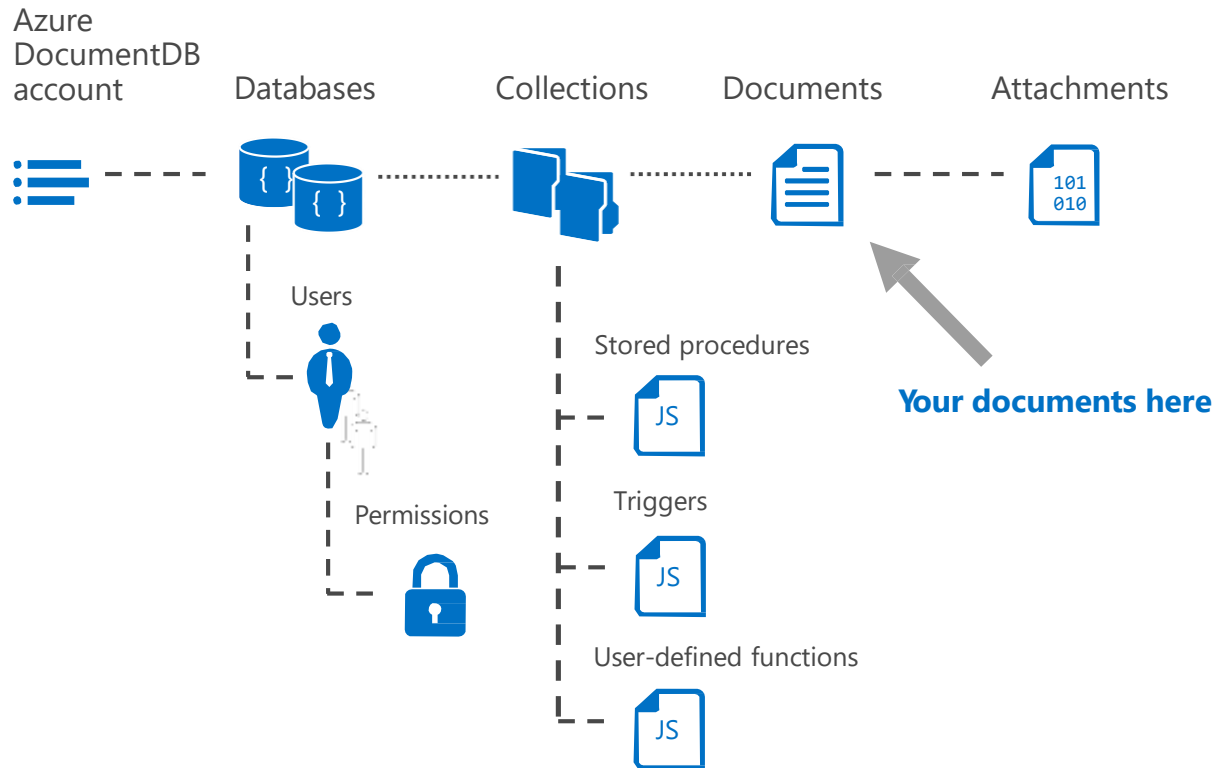
Controlul Consistenței

- Patru niveluri de consistență oferă mai multe opțiuni pentru coerență, disponibilitate și cerințe de performanță

Standards-based open API with RESTful HTTP

- Utilizează standarde JSON—fără necesitatea mapeării Binary JSON (BSON) to JSON

Azure DocumentDB basics



Modelul Resurselor

- Entități adresabile logic prin URI (Uniform Resource Identifier)
- Partiționare pentru scalare în sus
- Replicare pentru îmbunătățirea nivelului de disponibilitate
- Entități reprezentate ca JSON
- Scalare on-the-fly a contului

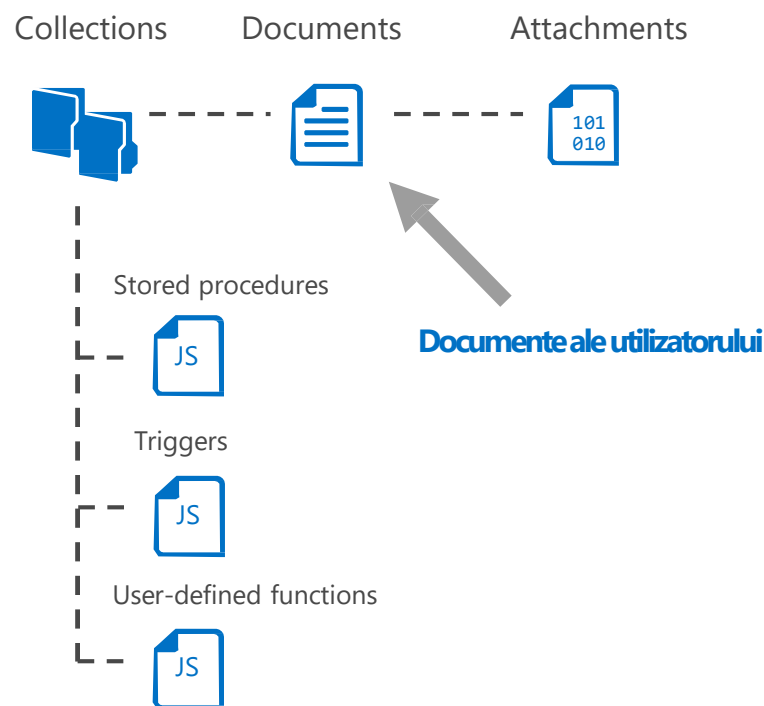
Modelul Interacțiunii

- Interacțiuni RESTful peste HTTPS
- Conectivitate HTTPS și TCP
- Verbe și semantică asociate standardului HTTPS

Dezvoltare

- Clienți .NET, Node.js, Python, Java, and JavaScript
- SQL pentru expresii de interogare, .NET LINQ
- JavaScript pentru server-side

Colecții Azure DocumentDB



- Collections != tabele
- Unitate de partiționare
- Limitarea transacției
- Interogarea sau actualizarea sunt constituite la nivel de colecție
- Elasticitate până la 10 GB
- RU-uri distribuite în mod egal dealungul partițiilor

"Rich Query" peste date JSON

Construcția de app, scalabile cu interogări transacționabile și procesarea datelor ca documente JSON



Interogarea datelor
ca documente JSON
fără specificarea
indexului secundar
sau construcția view-
urilor



JavaScript nativ
pentru procese
transacționale

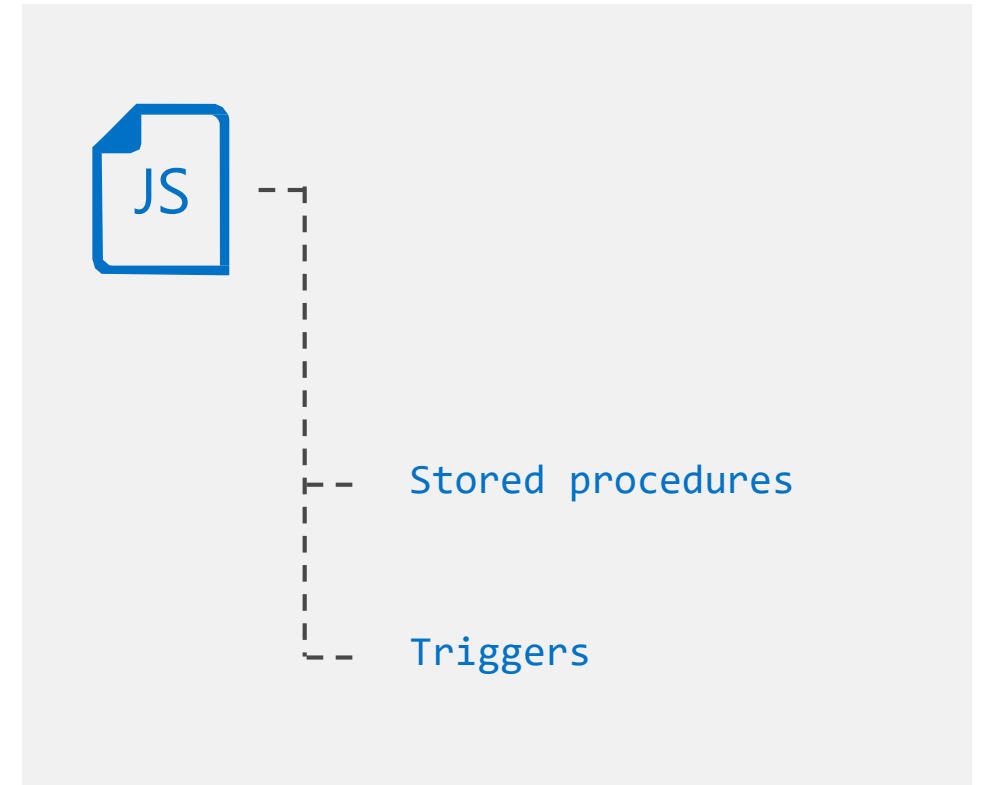


SQL ca limbaj classic
de interogare

Tranzacții JavaScript

Procese transacționale pe documente multiple cu procedure stocate și trigger-e

- JavaScript ca limbaj procedural
- Limbaj integrat
- Execution impachetată într-o tranzacție implicită
- Preinregistrate și orientate către o colecție
- Executate cu garanții ACID
- Trigger-ele invocate ca pre- sau post-operații



Patru nivele de consistență

Consistență slabă la nivelul operațiilor de citire

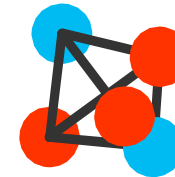
```
Document myDoc = await  
client.ReadDocumentAsync(documentLink, new  
RequestOptions { ConsistencyLevel =  
ConsistencyLevel.Eventual });
```



Puternică



Sesiune



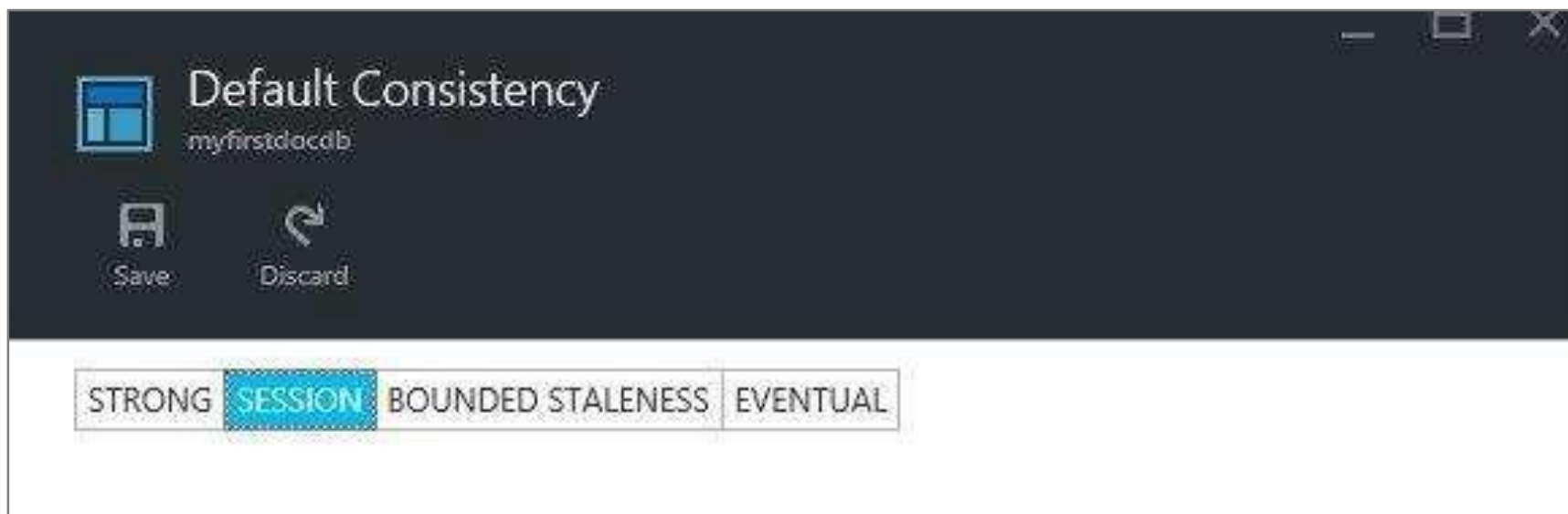
**Staleness
Limitat**



Eventual

Nivelele de consistență permit acordarea de garanții

Alegeți nivelul de consistență și efectuați compromisul previzibil între coerență, disponibilitate și performanță



Alegerea
nivelului

Strong
Consistența datelor

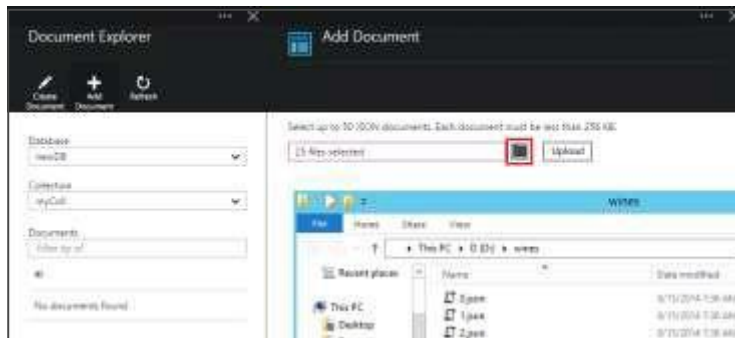
Sesiune
Citiri monotone
(pe o citire explicit
solicată) și scrie

Bounded Staleness
Ordine totală a
propășării scrierii

Eventual
Cea mai mica
latență la read și
write

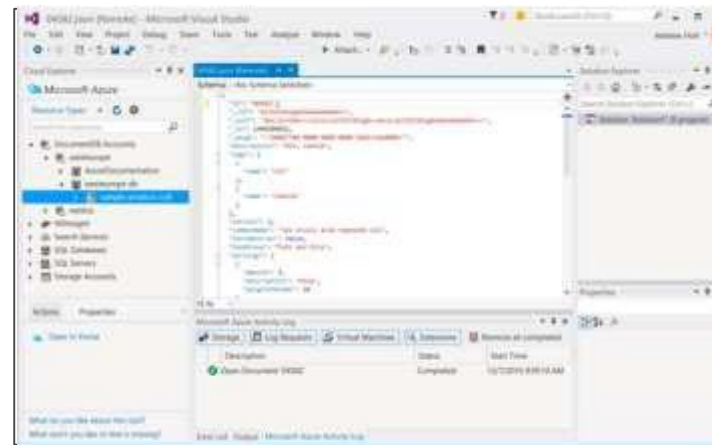
Tools

Document Explorer in Azure portal



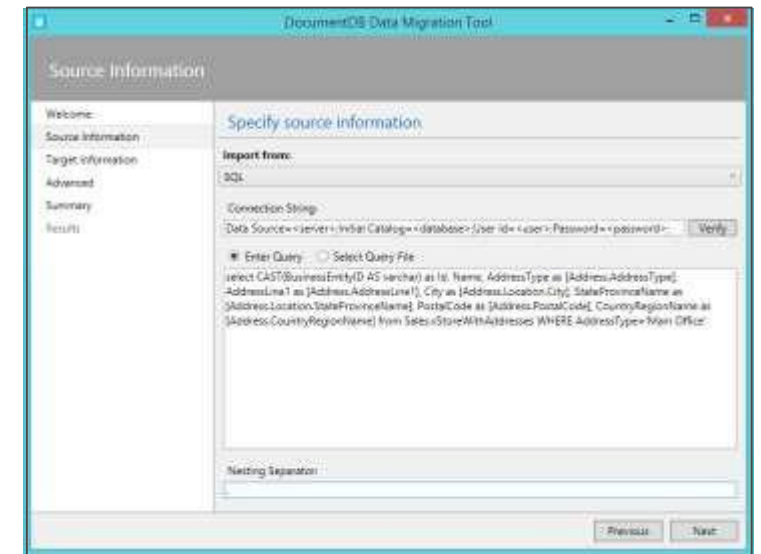
<http://portal.azure.com>

Microsoft Cloud Explorer for Visual Studio



<https://azure.microsoft.com/en-us/blog/exploring-azure-documentdb-in-visual-studio/>

Azure DocumentDB data-migration tool



<https://azure.microsoft.com/en-us/documentation/articles/documentdb-import-data/>

Câteva Resurse de documentare....

David Chappell NoSQL overview paper on Infopedia

http://www.davidchappell.com/writing/white_papers/Azure-NoSQL-Technologies-v2.0--Chappell.pdf

Seven Databases in Seven Weeks: A Guide to Modern Databases and the NoSQL Movement [book]

http://www.pdfbooks.com/pdf/files/English/Databases/Seven_Databases_In_Seven_Weeks.pdf

Replicated Data Consistency Explained Through Baseball [paper]

<http://research.microsoft.com/apps/pubs/default.aspx?id=206913>

