



TECHNICAL UNIVERSITY OF LIBEREC  
Faculty of Mechatronics, Informatics  
and Interdisciplinary Studies



# DOKUMENTOVÉ DATABÁZE

*Lukáš Matějů*

6.3.2024 | DPB



# ČÁST I.: OPAKOVÁNÍ



# OPAKOVÁNÍ

- NoSQL databáze
  - význam
    - často chybně – no SQL (žádné SQL)
    - správně – not only SQL (ne jen SQL)
      - u některých databází je možné SQL používat
  - nerelační databáze
    - žádné relace (tabulky)
  - technologie známá již od 60. let minulého století
    - snaha vyřešit problémy relačních databází (jaké jsou?)
    - jedno řešení pro všechno
  - ale až v současné době prudký nárůst popularity
    - obrovské objemy a široká škála dat
      - generované cloudy, mobilními zařízeními, sociálními médií, ...
    - cloud computing



# OPAKOVÁNÍ

- NoSQL databáze
  - prošly vývojem
    - dokáží zpracovávat obrovské objemy rychle se měnících nestrukturovaných dat
    - pomáhají vývojářům rychle vytvářet databázové systémy pro ukládání nových informací a jejich přípravu pro vyhledávání a analýzu
  - flexibilní vývoj
  - flexibilní zpracování dat
  - provoz v libovolném měřítku
- cílem tedy ve výsledku není nahradit relační databáze
  - jedná se spíš o alternativu s jinou aplikací (kladivo vs. sekera)
  - vhodné např. pro big data a webové aplikace v reálném čase



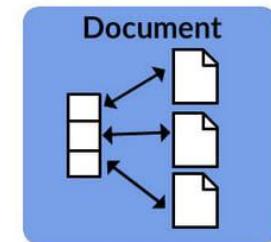
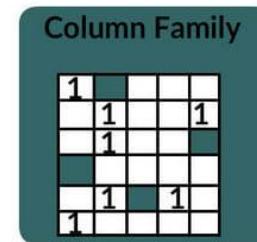
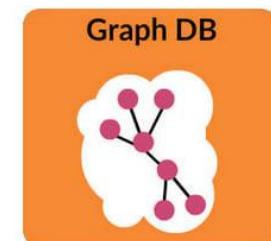
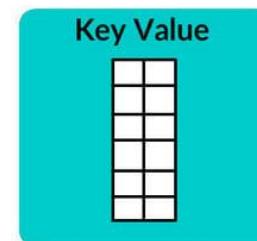
# OPAKOVÁNÍ

- relační databáze
  - provázaná data
  - pevné schéma
  - strukturovaná data
  - vertikální škálování
  - ACID
  - SQL
  - jednoduché složité dotazy
  - lepší podpora
  - transakce
- NoSQL databáze
  - velké množství dat
  - flexibilní schéma
  - veškerá data
  - horizontální škálování
  - CAP
  - velké množství jazyků
  - komplikované složité dotazy
  - převážně open source
  - analýza



# OPAKOVÁNÍ

- typy NoSQL databází
  - key-value úložiště
    - Riak, Redis, Couchbase Server
  - dokumentové databáze
    - MongoDB, CouchDB, Elasticsearch
  - sloupcová úložiště
    - Apache Cassandra, Druid, HBase
  - grafová úložiště
    - Neo4j
  - databáze časových řad
    - InfluxDB



<https://dev.to/lmolivera/everything-you-need-to-know-about-nosql-databases-3o3h#key>

# OPAKOVÁNÍ

- key-value úložiště
  - nejjednodušší typ NoSQL databází
  - data ukládána pomocí klíče (key) a přiřazené hodnoty (value)
    - každá hodnota přidružena k jedinečnému klíči
    - úložiště použije tento klíč k uložení dat pomocí odpovídající hash funkce
  - podobné asociativnímu poli / hashi
  - většinou podpora jen základních operací
  - optimalizované pro velké objemy dat
    - ale jednoduchých (žádné vazby)
    - velmi rychlé (podle klíče)
    - špatně upravitelné
  - dotazy vrací celou hodnotu
- např. ukládání uživatelských profilů



**ORACLE®**  
NOSQL DATABASE

# OPAKOVÁNÍ

- Redis
  - distribuované paměťové open-source key-value úložiště
    - plní roli databáze, mezipaměti, zprostředkovatele zpráv a streamovacího enginu
  - v současné době nejpopulárnější zástupce key-value úložišť
  - úložiště je celé udržováno v paměti
    - extrémně rychlé čtení a zápisy s nízkou latencí
    - v základu zapisuje kopie dat do souborového systému každé dvě vteřiny
  - podporuje různé abstraktní datové typy
    - řetězce, listy, sety, seřazené sety, hashe
    - bitmapy, hyperlogy, proudy, geoprostorové indexy
  - replikace, shardování, klastrování
  - využívané celou řadou společností
    - Twitter, Airbnb, Tinder, Yahoo, Adobe, Hulu, Amazon, OpenAI, Github, Snapchat, ...



# ČÁST II.: DOKUMENTOVÉ DATABÁZE



„Relační databáze je jako garáž, která vás nutí rozebrat vaše auto na součástky a uložit je do jednotlivých šuplíků.“

„Objektová databáze je jako skříň, do které musíte dát všechno své oblečení svázané.“



# DOKUMENTOVÉ DATABÁZE

- označovány také jako
  - dokumentově orientované databáze
  - úložiště dokumentů
- v současnosti asi nejpoužívanější typ NoSQL databází
- v principu podobné key-value úložištím
  - zachován princip key-value (klíč-hodnota)
    - key jednoznačným identifikátorem value
  - ale value obsahují strukturovaná nebo částečně strukturovaná data
    - tzv. dokumenty
  - samotná data mohou být indexována a dotazována
    - indexy nad atributy dat
    - dotazy na strukturu dat i na prvky v této struktuře
  - je možné získat jen požadované části dokumentů



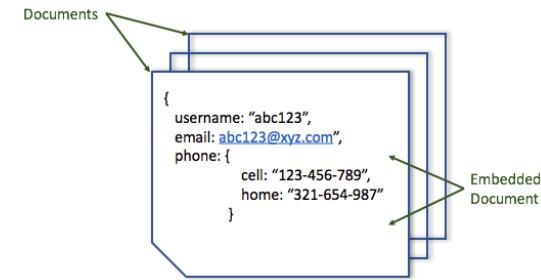
# DOKUMENTY

- základní stavební prvky dokumentových databází
  - zapouzdřují a kódují data v definovaném formátu (kódování)
    - implementace se liší databázi od databáze
  - používaná kódování
    - textová forma
      - XML, YAML, JSON
    - binární forma
      - BSON, PDF, MS Office dokumenty
  - identifikovány jednoznačným identifikátorem (klíčem)
    - typicky řetězec, URI nebo cesta
    - slouží pro přístup k dokumentům
      - ale i pro vkládání
    - často indexovány
      - rychlejší přístup k dokumentům



# DOKUMENTY

- základní stavební prvky dokumentových databází
  - koncepčně odpovídají objektům v OOP
  - volné schéma
    - dokumenty mohou být velmi komplexní
    - mohou obsahovat vnořené (embedded) dokumenty
    - nemusí obsahovat stejné oddíly, atributy, části nebo klíče
    - podobně jako objekty
    - vysoká míra flexibility
  - obecně data patřící k sobě ukládána do jednoho dokumentu
    - na rozdíl od relačních databází
    - usnadňuje přístup a práci s daty
  - k dokumentům často přidružena a uložena metadata



<https://expert.data.blog/category/bigdata/>



# DOKUMENT

- ukázka ve formátu XML

```
<artist>
  <artistname>Iron Maiden</artistname>
  <albums>
    <album>
      <albumname>The Book of Souls</albumname>
      <datereleased>2015</datereleased>
      <genre>Hard Rock</genre>
    </album>
    <album>
      <albumname>Killers</albumname>
      <datereleased>1981</datereleased>
      <genre>Hard Rock</genre>
    </album>
    <album>
      <albumname>Powerslave</albumname>
      <datereleased>1984</datereleased>
      <genre>Hard Rock</genre>
    </album>
    <album>
      <albumname>Somewhere in Time</albumname>
      <datereleased>1986</datereleased>
      <genre>Hard Rock</genre>
    </album>
  </albums>
</artist>
```

<https://database.guide/what-is-a-document-store-database/>



# DOKUMENT

- ukázka ve formátu JSON

```
{  
  "_id" : 1,  
  "artistName" : "Iron Maiden",  
  "albums" : [  
    {  
      "albumname" : "The Book of Souls",  
      "datereleased" : 2015,  
      "genre" : "Hard Rock"  
    }, {  
      "albumname" : "Killers",  
      "datereleased" : 1981,  
      "genre" : "Hard Rock"  
    }, {  
      "albumname" : "Powerslave",  
      "datereleased" : 1984,  
      "genre" : "Hard Rock"  
    }, {  
      "albumname" : "Somewhere in Time",  
      "datereleased" : 1986,  
      "genre" : "Hard Rock"  
    }  
  ]  
}
```

<https://database.guide/what-is-a-document-store-database/>

# DOKUMENT

- ukázka volného schématu ve formátu JSON

```
[  
  {  
    "_id" : 1,  
    "artistName" : "Iron Maiden",  
    "albums" : [  
      {  
        "albumname" : "The Book of Souls",  
        "datereleased" : 2015,  
        "songs" : "11"  
      },  
      {  
        "albumname" : "Killers",  
        "datereleased" : 1981,  
        "genre" : "Hard Rock"  
      },  
      {  
        "albumname" : "Powerslave"  
      },  
      {  
        "albumname" : "Somewhere in Time",  
        "datereleased" : [ 1986, 1995 ],  
        "genre" : "Hard Rock"  
      }  
    ]  
  },  
  {  
    "_id" : 2,  
    "artistName" : "Taylor Swift",  
    "instruments" : [ "vocals", "guitar", "piano" ]  
  }  
]
```

<https://database.guide/what-is-a-document-store-database/>



# DOKUMENT

- ukázka volného schématu ve formátu JSON

```
_id: ObjectId("5f8ef175c43ece2db0230f85")
title: "Post One"
body: "Body of post one"
category: "News"
likes: 4
tags: Array
  0: "news"
  1: "events"
user: Object
  name: "John Doe"
  status: "author"
  date: "Date()"

_id: ObjectId("5f8ef175c43ece2db0230f88")
title: "Post Four"
category: "Entertainment"

_id: ObjectId("5f8fe759c43ece2db0230f8a")
title: "Post Five"
likes: 81
tags: Array
  0: "news"
  1: "events"
user: Object
  name: "John Doe"
  status: "author"
  date: "Date()"

_id: ObjectId("5f8ef175c43ece2db0230f86")
title: "Post Two"
body: "Body of post two"
tags: "news"
date: "Date()"

_id: ObjectId("5f8ef175c43ece2db0230f87")
source: "id1"
title: "Post Three"
views: "80"

_id: ObjectId("5f9026800cbc092824d7e420")
source: "id2"
title: "Unknown"
user: Object
  name: "Jane Doe"
  gender: "female"
```



# DOKUMENTOVÉ DATABÁZE

- podporují standardní operace s daty (dokumenty)
  - operace CRUD
  - implementace se liší databázi od databáze
  - vytvoření (vložení) [**creation**]
  - čtení (dotazování, vyhledávání) [**retrieval**]
    - kromě vyhledávání podle klíče také podpora dotazovacího jazyka
      - vyhledávání v závislosti na obsahu (nebo metadatech)
  - aktualizace [**update**]
    - i jen části dokumentu
  - smazání [**deletion**]
  - mohou podporovat transakce
    - ACID
    - není ale pravidlem



<https://medium.com/@hau12a1/golang-http-crud-i-the-create-part-ae42c962c557>



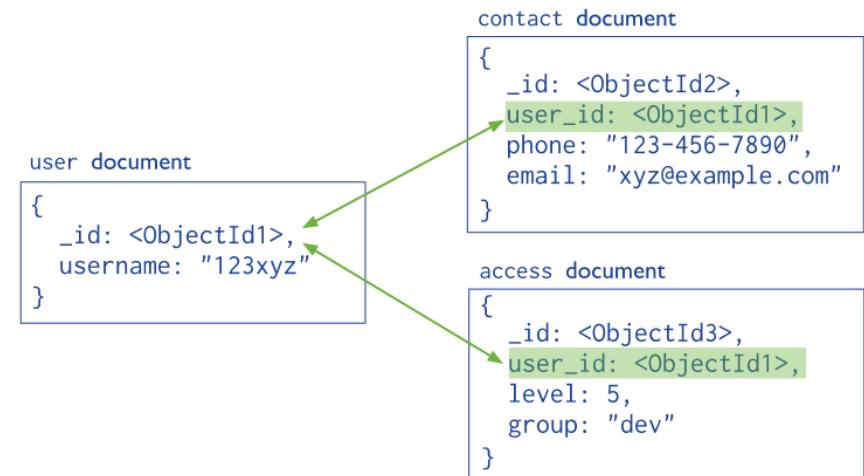
# DOKUMENTOVÉ DATABÁZE

- obecně se vyhýbají vazbám mezi dokumenty
  - případně dvě varianty řešení
    - embedded dokumenty
    - reference

```
{  
  _id: <ObjectId1>,  
  username: "123xyz",  
  contact: {  
    phone: "123-456-7890",  
    email: "xyz@example.com"  
  },  
  access: {  
    level: 5,  
    group: "dev"  
  }  
}
```

Embedded sub-document

Embedded sub-document



<https://docs.mongodb.com/manual/core/data-model-design/>



# DOKUMENTOVÉ DATABÁZE

```
{  
    title: "MongoDB: The Definitive Guide",  
    author: [ "Kristina Chodorow", "Mike Dirolf" ],  
    published_date: ISODate("2010-09-24"),  
    pages: 216,  
    language: "English",  
    publisher: {  
        name: "O'Reilly Media",  
        founded: 1980,  
        location: "CA"  
    }  
}  
  
{  
    title: "50 Tips and Tricks for MongoDB Developer",  
    author: "Kristina Chodorow",  
    published_date: ISODate("2011-05-06"),  
    pages: 68,  
    language: "English",  
    publisher: {  
        name: "O'Reilly Media",  
        founded: 1980,  
        location: "CA"  
    }  
}  
  
{  
    name: "O'Reilly Media",  
    founded: 1980,  
    location: "CA",  
    books: [123456789, 234567890, ...]  
}  
  
{  
    _id: 123456789,  
    title: "MongoDB: The Definitive Guide",  
    author: [ "Kristina Chodorow", "Mike Dirolf" ],  
    published_date: ISODate("2010-09-24"),  
    pages: 216,  
    language: "English"  
}  
  
{  
    _id: 234567890,  
    title: "50 Tips and Tricks for MongoDB Developer",  
    author: "Kristina Chodorow",  
    published_date: ISODate("2011-05-06"),  
    pages: 68,  
    language: "English"  
}  
  
{  
    _id: "oreilly",  
    name: "O'Reilly Media",  
    founded: 1980,  
    location: "CA"  
}  
  
{  
    _id: 123456789,  
    title: "MongoDB: The Definitive Guide",  
    author: [ "Kristina Chodorow", "Mike Dirolf" ],  
    published_date: ISODate("2010-09-24"),  
    pages: 216,  
    language: "English",  
    publisher_id: "oreilly"  
}  
  
{  
    _id: 234567890,  
    title: "50 Tips and Tricks for MongoDB Developer",  
    author: "Kristina Chodorow",  
    published_date: ISODate("2011-05-06"),  
    pages: 68,  
    language: "English",  
    publisher_id: "oreilly"  
}
```

<https://docs.mongodb.com/>



# DOKUMENTOVÉ DATABÁZE

- organizace
  - liší se databázi od databáze
  - kolekce
    - skupiny dokumentů
      - dokument může být jen v jedné nebo i více kolekcích
      - závislé na implementaci
    - tagy a metadata
      - dodatečné informace mimo dokumenty
  - hierarchie adresářů
    - skupiny dokumentů uspořádané do stromové skupiny
    - většinou založené na cestách nebo URI
  - fyzická a logická reprezentace
    - uložené na disku nebo v paměti
- distribuované databáze



# INDEXOVÁNÍ

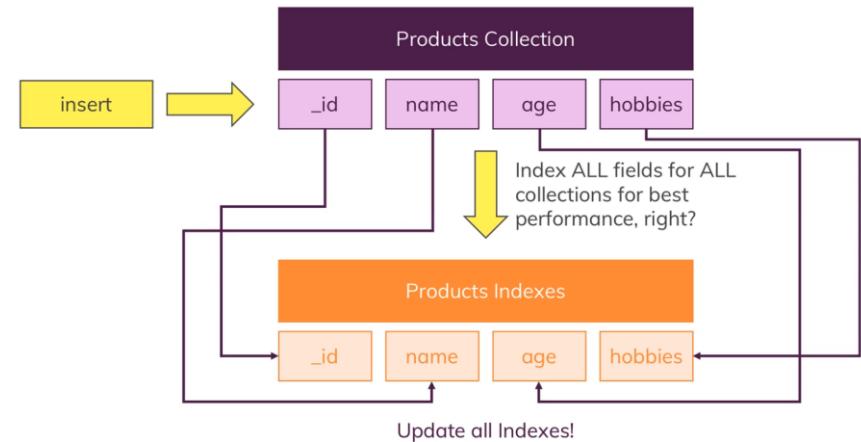
- vylepšuje rychlosť vyhľadávania
  - bez indexovania sa provádzí sken celé kolekcia
  - vhodný index výrazne omezuje dokumenty, ktoré je potreba skenovať
- index
  - speciálna datová struktúra definovaná na úrovni kolekcii
  - ukladá hodnotu specifického pole v seřazené formě
    - snadné procházení a porovnávanie
    - obsahuje také pointer na celý dokument pre snadný prístup
  - možnosť definovať nad libovolným polem ale i nad jejich kombináciami
    - v základe index\_id
  - využitie i pre rýchle řazení
  - základ pre sharding
    - len jeden môže byť použit



# INDEXOVÁNÍ

- vylepšuje rychlosť vyhľadávania
- proč tedy nedefinovať indexy nad všemi polí?
  - teoreticky vylepší veškeré vyhľadávanie
  - v praxi ale index není zadarmo
- cenou je vkládanie / aktualizácia
  - pri každém vložení je potreba vložiť prvek i do seřazeného indexu
    - potreba vložiť do všech indexov
    - pomalé?
- indexy je potreba řádně promyslet
- indexování je pomalejší, když dotaz vrací velkou část kolekce
  - způsobeno přechodem index – pointer – dokument
  - naopak při kompletním skenu už jsou dokumenty načteny v paměti

<https://www.udemy.com/course/mongodb-the-complete-developers-guide>

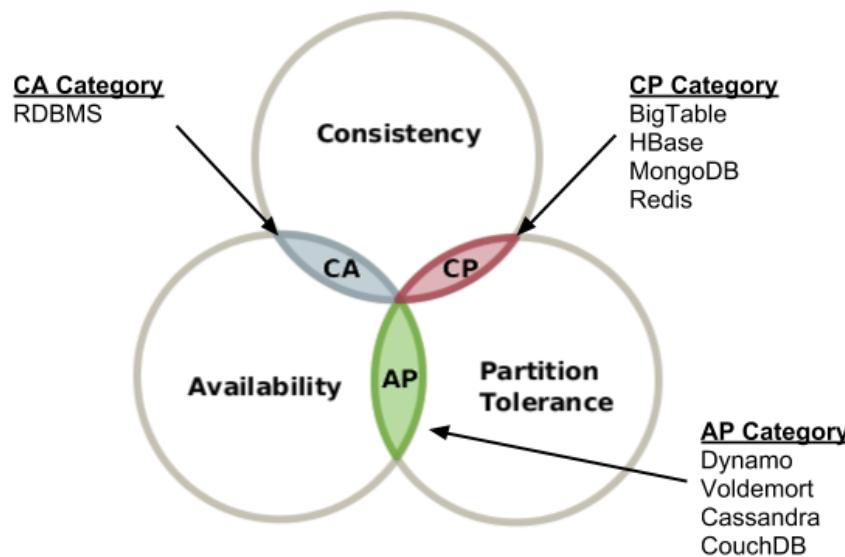


# CAP TEORÉM

- říká, že pro distribuovaný datový sklad nelze zajistit více jak dvě z následujících tří vlastnosti
  - konzistence – consistency (C)
    - všechny uzly mají stejná data
    - každé čtení vrací poslední výsledek, nebo chybu
  - dostupnost – availability (A)
    - na každý dotaz je vrácena (nechybová) odpověď
  - odolnost k přerušení – partition tolerance (P)
    - systém funguje dál i při zdržení či ztrátě zprávy
    - nezbytná pro big data
- dokumentové databáze
  - CP & AP



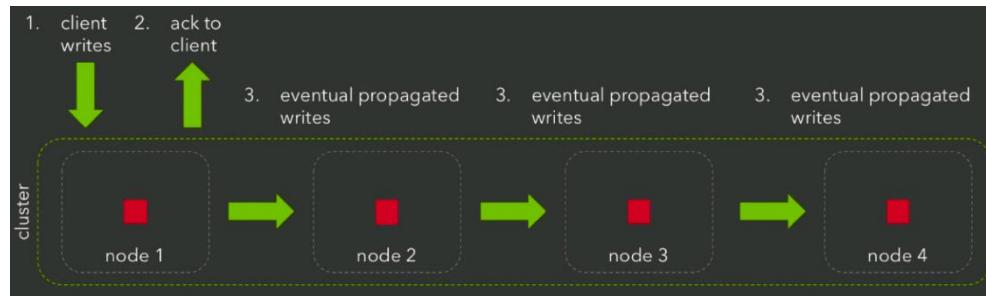
# CAP TEORÉM



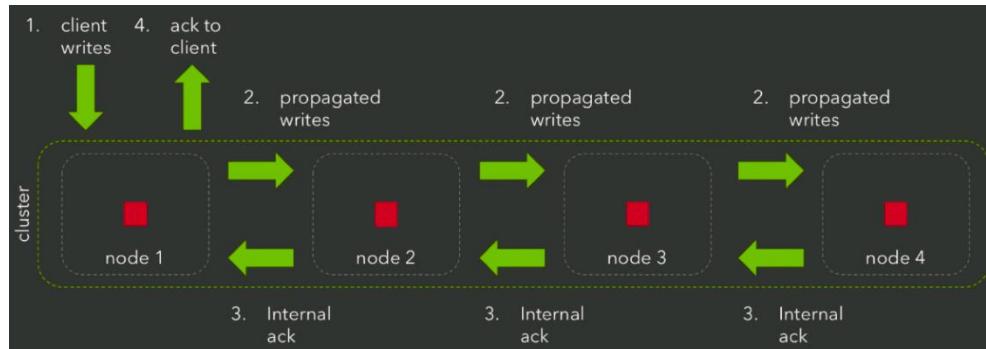
<https://changeaas.wordpress.com/tag/nosql/>

# CAP TEORÉM

- eventuálně konzistentní



- striktně konzistentní

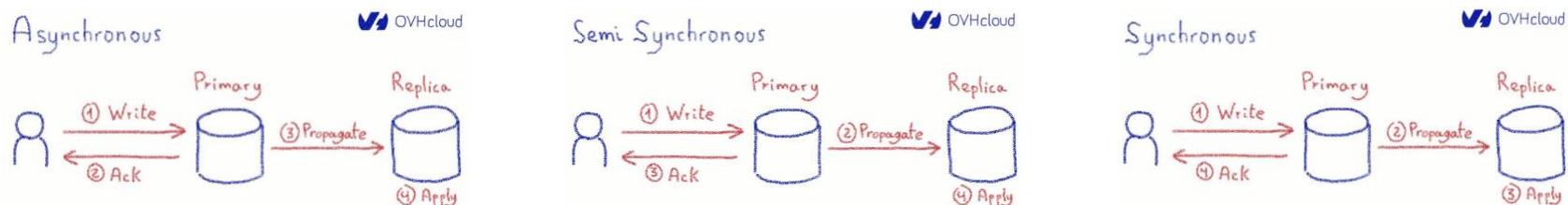


<https://www.cohesity.com/blogs/strict-vs-eventual-consistency/>



# REPLIKACE

- automatická distribuce změn v originále do jeho kopií
- u dokumentových databází nejčastěji typu master-slave
  - obvyklá vazba mezi originálem a kopíemi
  - master zaznamenává změny, které předává slaves
  - slaves potvrdí přijetí změn, čímž umožní další aktualizace
  - asynchronní (eventuálně k.), semi synchronní, synchronní (striktně k.)



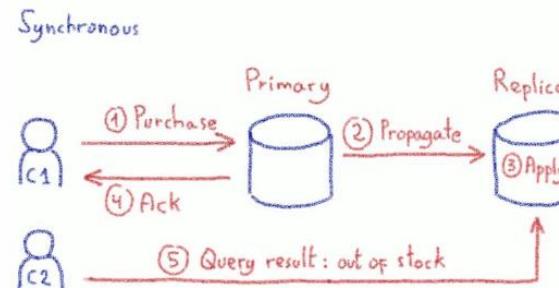
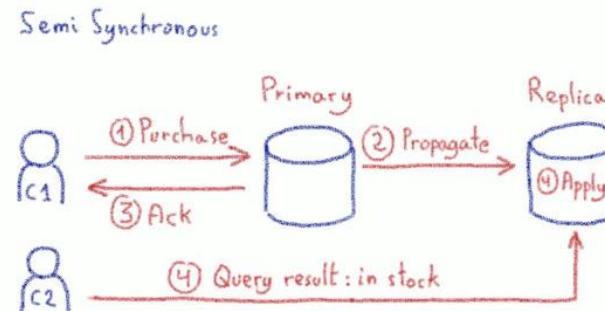
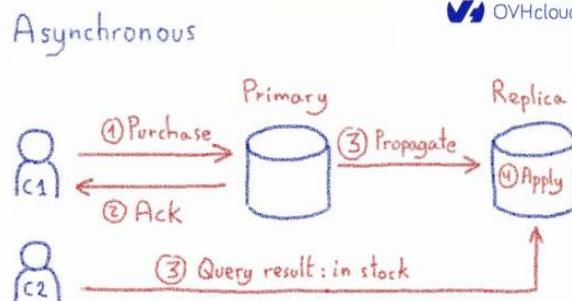
<https://www.ovh.com/blog/database-replication-101/>

- zajišťuje vysokou dostupnost



# REPLIKACE

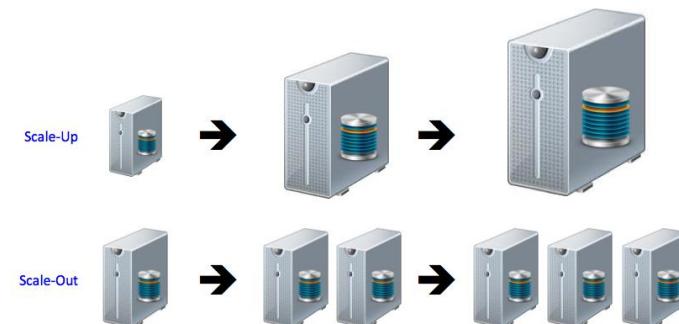
- automatická distribuce změn v originále do jeho kopíí



<https://www.ovh.com/blog/database-replication-101/>

# ŠKÁLOVÁNÍ

- škálovatelnost je vlastnost systému zvládat rostoucí množství požadavků přidáváním dalších prostředků
  - u dokumentových databází nejčastěji horizontální škálování (ven)
    - přidání (nebo ubrání) prvků
    - přidání výpočetních uzlů
  - přesun k distribuovanému paralelnímu zpracování
    - rozdelení dat mezi uzly
    - horizontální sharding
  - zvýšení kapacity
    - nové komponenty jsou levné
      - základní HW
    - distribuované clustery
  - cloudové služby řeší za uživatele
    - big data

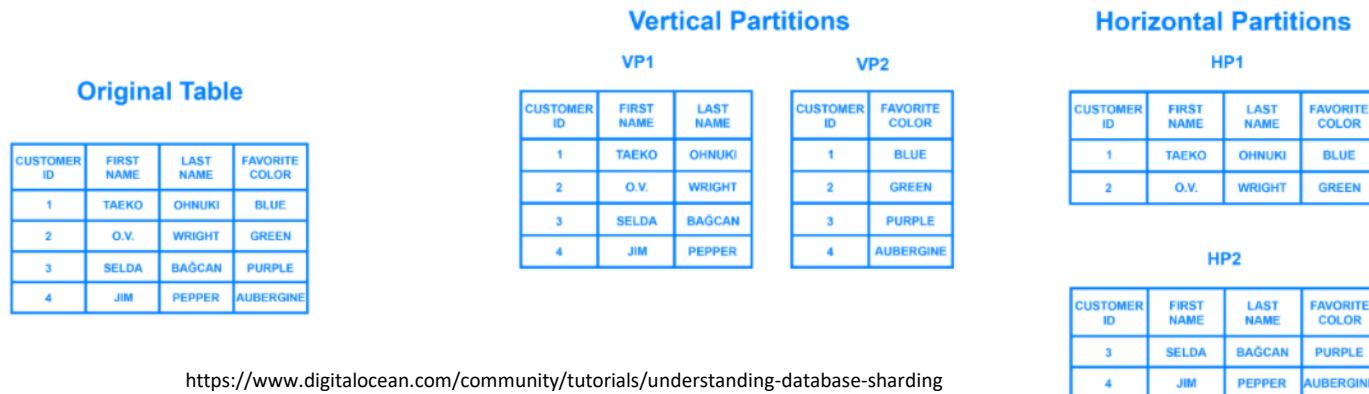


<https://medium.com/faun/scalability-248019b918ed>



# SHARDING

- databázový vzor pro horizontální škálování na více serverů
- rozdělení záznamů na části (partitions, shards) umístěné na různých serverech
  - např. u relačních databází rozdělení tabulky podle řádků, ne sloupců
  - např. u dokumentových databází rozdělení podle dokumentů, ne atributů

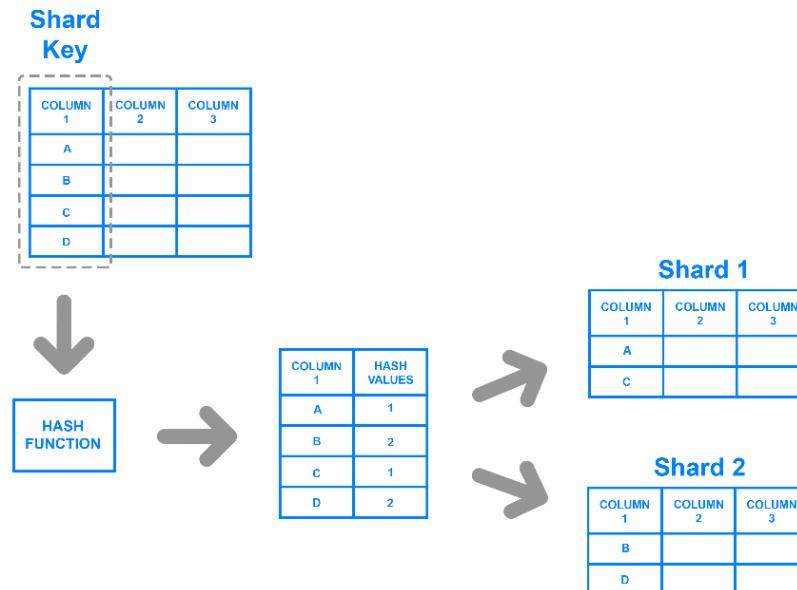


- data mezi shardy nejsou sdílena



# SHARDING

- architektury
  - založená na klíči (hashi)
  - důležitý shard key

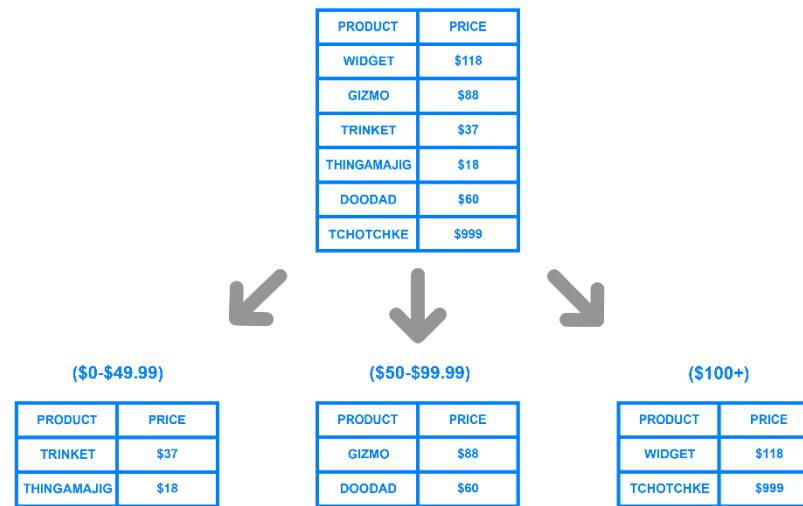


<https://www.digitalocean.com/community/tutorials/understanding-database-sharding>



# SHARDING

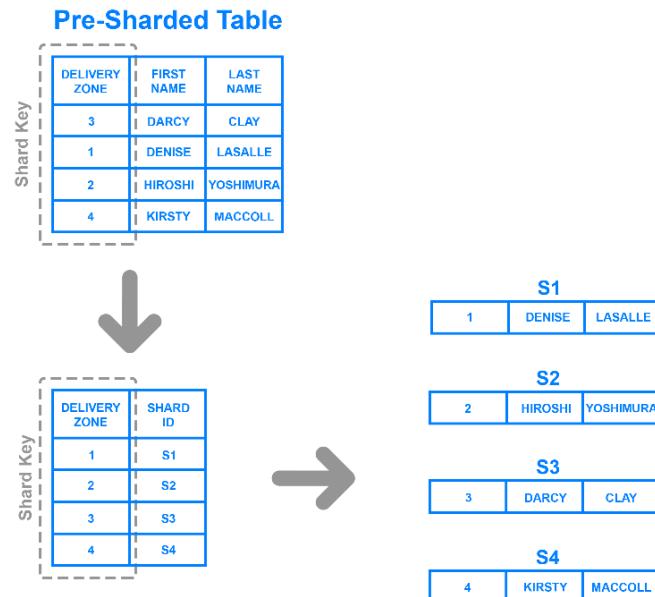
- architektury
  - založená na rozsahu



<https://www.digitalocean.com/community/tutorials/understanding-database-sharding>

# SHARDING

- architektury
  - založená na adresářích
  - lookup tabulka odkazující na shardy



<https://www.digitalocean.com/community/tutorials/understanding-database-sharding>





# POUŽITÍ DOKUMENTOVÝCH DATABÁZÍ

- široké možnosti použití (ukládání)
  - webové aplikace
    - blogovací platformy, analytická data, nastavení uživatelů, e-reklamy, ...
  - data generovaná uživateli
    - chaty, tweety, příspěvky, hodnocení, komentáře, ...
  - katalogy
    - uživatelské účty, produkty, preference, ...
  - počítačové hry
    - herní statistiky, žebříčky, vestavěné chaty, splněné úkoly, integrace social media, ...
  - networking
    - data ze senzorů, logy, real-time analýza, ...
  - ...





# SROVNÁNÍ S RELAČNÍMI DATABÁZAMI

```
{  
    "_id" : 1,  
    "artistName" : "Iron Maiden",  
    "albums" : [  
        {  
            "albumname" : "The Book of Souls",  
            "datereleased" : 2015,  
            "genre" : "Hard Rock"  
        }, {  
            "albumname" : "Killers",  
            "datereleased" : 1981,  
            "genre" : "Hard Rock"  
        }, {  
            "albumname" : "Powerslave",  
            "datereleased" : 1984,  
            "genre" : "Hard Rock"  
        }, {  
            "albumname" : "Somewhere in Time",  
            "datereleased" : 1986,  
            "genre" : "Hard Rock"  
        }  
    ]  
}
```

ARTISTS

ArtistId	ArtistName
1	Iron Maiden
2	Devin Townsend

ALBUMS

AlbumId	AlbumName	DateReleased	ArtistId	GenreId
1	The Book of Souls	2015	1	3
2	Killers	1981	1	3
3	Powerslave	1984	1	3
4	Somewhere in Time	1986	1	3

GENRE

GenreId	Genre
1	Country
2	Blues
3	Hard Rock
4	...

<https://database.guide/what-is-a-document-store-database/>



# SROVNÁNÍ S RELAČNÍMI DATABÁZAMI

## ARTISTS

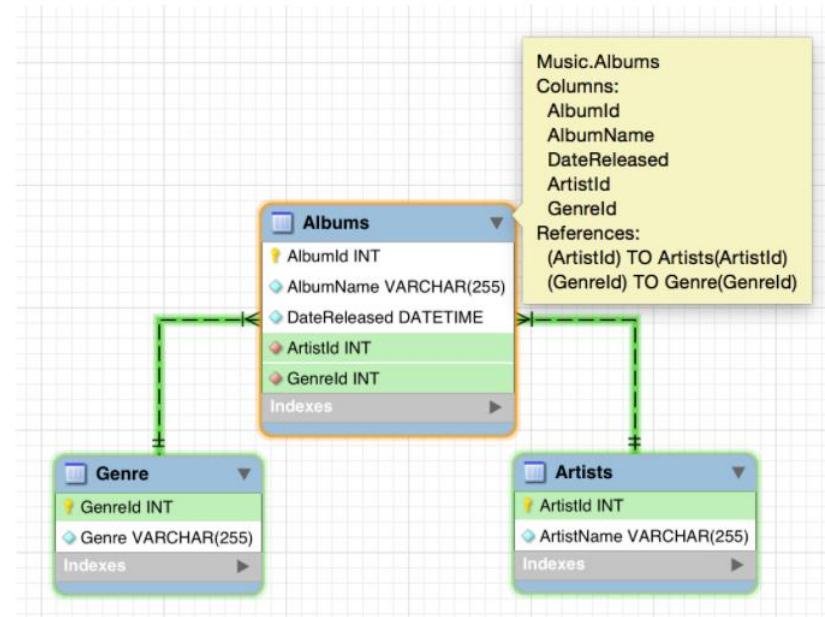
ArtistId	ArtistName
1	Iron Maiden
2	Devin Townsend

## ALBUMS

AlbumId	AlbumName	DateReleased	ArtistId	GenreId
1	The Book of Souls	2015	1	3
2	Killers	1981	1	3
3	Powerslave	1984	1	3
4	Somewhere in Time	1986	1	3

## GENRE

GenreId	Genre	...
1	Country	
2	Blues	
3	Hard Rock	
4	...	



<https://database.guide/what-is-a-document-store-database/>



# SROVNÁNÍ S RELAČNÍMI DATABÁZAMI

```
[  
  {  
    "_id" : 1,  
    "artistName" : "Iron Maiden",  
    "albums" : [  
      {  
        "albumname" : "The Book of Souls",  
        "datereleased" : 2015,  
        "songs" : "11"  
      }, {  
        "albumname" : "Killers",  
        "datereleased" : 1981,  
        "genre" : "Hard Rock"  
      }, {  
        "albumname" : "Powerslave"  
      }, {  
        "albumname" : "Somewhere in Time",  
        "datereleased" : [ 1986, 1995 ],  
        "genre" : "Hard Rock"  
      }  
    ]  
  }, {  
    "_id" : 2,  
    "artistName" : "Taylor Swift",  
    "instruments" : [ "vocals", "guitar", "piano" ]  
  }  
]
```



<https://database.guide/what-is-a-document-store-database/>



# SROVNÁNÍ S RELAČNÍMI DATABÁZAMI

- relační databáze
  - pevně definované schéma
    - relace, atributy, datové typy, ...
  - data ke konkrétnímu záznamu uložena v mnoha propojených relacích
  - propojení (vztah) mezi relacemi zajišťují cizí klíče
  - vertikální škálování
  - SQL
- dokumentové databáze
  - volné schéma
    - různá struktura, různé datové typy, ...
  - data ke konkrétnímu záznamu většinou ukládána v jediném dokumentu
  - cizí klíče neexistují, veškerá data uložena v daném dokumentu
  - horizontální škálování
    - big data
  - různé dotazovací jazyky



# ČÁST III.: ZÁSTUPCI DOKUMENTOVÝCH DATABÁZÍ

# ZÁSTUPCI DOKUMENTOVÝCH DATABÁZÍ

- poměrně velké množství...

Name	Publisher	License	Languages supported	Notes	RESTful API
AllegroGraph	Lotus Notes (IBM Lotus Domino)	IBM	Proprietary	LotusScript, Java, Lotus @Formula MultiValue	Yes
ArangoDB	MarkLogic	MarkLogic Corporation	Free Developer license or Commercial <sup>[13]</sup>	Java, JavaScript, Node.js, XQuery, SPARQL, XSLT, C++ Distributed document-oriented database for JSON, XML, and RDF triples. Built-in full-text search, ACID transactions, high availability and disaster recovery, certified security.	Yes
BaseX					
Caché	MongoDB, Inc	MongoDB, Inc	Server Side Public License for the DBMS, Apache 2 License for the client drivers <sup>[14]</sup>	C, C++, C#, Java, Perl, PHP, Python, Go, Node.js, Ruby, Rust, <sup>[15]</sup> Scala <sup>[16]</sup> Document database with replication and sharding, BSON store (binary format JSON).	Yes <sup>[17][18]</sup>
Cloudant	MUMPS Database	?	Proprietary and Affero GPL <sup>[19]</sup>	MUMPS Commonly used in health applications.	?
Clusterpoint Database	ObjectDatabase++	Ekky Software	Proprietary	C++, C#, TScript Binary Native C++ class structures	?
Couchbase Server	OpenLink Virtuoso	OpenLink Software	GPLv2[1] and proprietary	C++, C#, Java, SPARQL Middleware and database engine hybrid	Yes
	OrientDB	Orient Technologies	Apache License	Java JSON over HTTP, SQL support, ACID transactions	Yes
CouchDB	Oracle NoSQL Database	Oracle Corp	Apache and proprietary	C, C#, Java, Python, node.js, Go Shared nothing, horizontally scalable database with support for schema-less JSON, fixed schema tables, and key/value pairs. Also supports ACID transactions.	Yes
CrateIO	PostgreSQL	PostgreSQL	PostgreSQL Free License <sup>[20]</sup>	C HStore, JSON store (9.2+), JSON function (9.3+), HStore2 (9.4+), JSONB (9.4+)	No
	Qizx	Qualcomm	Proprietary	REST, Java, XQuery, XSLT, C, C++, Python Distributed document-oriented XML database with integrated full-text search; support for JSON, text, and binaries.	Yes
Cosmos DB	ReJSON <sup>[21]</sup>	Redis Labs	Redis Source Available License	Node.js, Java, Python, Go and all Redis clients. <sup>[22]</sup> Native in-memory data type packaged as Redis Module.	?
DocumentDB	RethinkDB	?	Apache License <sup>[23]</sup>	C++, Python, JavaScript, Ruby, Java Distributed document-oriented JSON database with replication and sharding.	No
Elasticsearch	SAP HANA	SAP	Proprietary	SQL-like language ACID transaction supported, JSON only	Yes
eXist	Sedna	sedna.org	Apache License	C++, XQuery XML database	No
Informix	SimpleDB	Amazon Web Services	Proprietary online service	Erlang	?
Jackrabbit	Solr	Apache	Apache License	Java Search engine	Yes
	TokuMX	Tokutek	GNU Affero General Public License	C++, C#, Go MongoDB with Fractal Tree indexing	?

[https://en.wikipedia.org/wiki/Document-oriented\\_database](https://en.wikipedia.org/wiki/Document-oriented_database)





# VÝZNAMNĚJŠÍ ZÁSTUPCI

- MongoDB
- Amazon DynamoDB
- Azure Cosmos DB
- Couchbase Server
- CouchDB
- Cloud Firestore
- MarkLogic Server
- Elasticsearch
- Solark



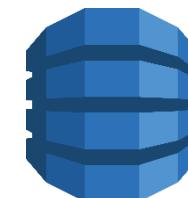
# VÝZNAMNĚJŠÍ ZÁSTUPCI

- MongoDB (od 2009)
  - v současnosti nejpopulárnější dokumentová databáze
  - licence Server Side Public License (SSPL)
    - velmi podobná GNU GPLv3
  - hlavní funkce
    - volné schéma
    - ad-hoc dotazy
      - dotazy neznámé v době vytvoření databáze
    - indexování
    - replikace
      - vysoká dostupnost
    - sharding
    - agregace
  - dokumenty ve formátu BSON (binární JSON)
  - cloudová služba MongoDB Atlas



# VÝZNAMNĚJŠÍ ZÁSTUPCI

- Amazon DynamoDB (od 2012)
  - dokumentová databáze nabízená v rámci Amazon Web Services
    - on-demand cloud computing
  - nástupce Amazon SimpleDB
  - synchronní replikace přes datová centra
    - vysoká odolnost a dostupnost
  - auto škálování
  - dokumenty ve formátu JSON



NETFLIX

CapitalOne®

[ROCKSET]

Nike

Shape  
your future  
START HERE > | United States  
Census  
2020

The Pokémon Company  
INTERNATIONAL

SAMSUNG

# VÝZNAMNĚJŠÍ ZÁSTUPCI

- Azure Cosmos DB (od 2017)
  - globálně distribuovaná cloudová služba Microsoftu
  - více modelová databáze
    - API – SQL, MongoDB, Cassandra, Azure Table Storage, Gremlin a etcd
  - elastická škálovatelnost propustnosti
  - možnost volby konzistence
    - až pět úrovní
  - automaticky indexovaná všechna data
  - terminologie – dokument -> item



MARS



CHIPOTLE

asos

Symantec

# VÝZNAMNĚJŠÍ ZÁSTUPCI

- Cloud Firestore (od 2017)
  - cloudová databáze od Google
  - vychází z původní Firebase Realtime Database
  - aktualizace v reálném čase
    - datová synchronizace aktualizující data na všech zařízeních
  - podpora pro off-line
  - automatická replikace
  - automatické škálování
  - podpora transakcí
  - dokumenty v JSON



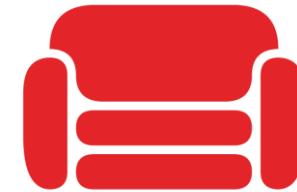
# VÝZNAMNĚJŠÍ ZÁSTUPCI

- Couchbase Server (od 2010)
  - původně Membase
  - open-source cloudová databáze
  - více modelová
    - key-value přístup
    - dokumentový přístup
      - JSON
      - ale i jiné formáty jsou podporovány
- Couchbase Mobile
  - pro mobilní a embedded zařízení



# VÝZNAMNĚJŠÍ ZÁSTUPCI

- CouchDB (od 2005)
  - Apache
  - open-source dokumentová databáze
  - dokumenty ve formátu JSON
  - ACID na úrovni dokumentů
  - eventuálně konzistentní
  - dotazy pomocí JavaScriptu
  - HTTP API
  - multi-master replikace
  - off-line provoz
  - horizontální škálování napříč různými zařízeními



# VÝZNAMNĚJŠÍ ZÁSTUPCI

- MarkLogic Server (od 2003)
  - dokumentová databáze od MarkLogic
  - umožňuje ukládání
    - dokumentů ve formátu XML, JSON
    - RDF trojic
      - pro sémantickou analýzu
  - ACID pro transakce
  - fulltextové vyhledávání
  - vestavěné strojové učení



# VÝZNAMNĚJŠÍ ZÁSTUPCI

- Elasticsearch (od 2010)
  - vyhledávací engine
  - umožňuje distribuované fulltextové vyhledávání v JSON dokumentech
    - škálovatelné
    - téměř v reálném čase
  - vyvíjen společně s
    - Logstash – pipeline pro zpracování dat z různých zdrojů na straně serveru
    - Kibana – platforma pro analýzu a vizualizaci
    - Beats – platforma pro data shippers
  - tvoří dohromady Elastic Stack
  - Elastic Cloud



# VÝZNAMNĚJŠÍ ZÁSTUPCI

- Solr (od 2004)
  - Apache
  - open-source vyhledávací engine
  - postaven na Apache Lucene
  - umožňuje distribuované fulltextové vyhledávání
    - škálovatelnost
    - odolnost vůči chybám
  - hlavní funkce
    - fulltextové vyhledávání, zvýrazňování výsledků, indexace v reálném čase, dynamické klastrování, podpora různých formátů dokumentů



# ONLINE KURZY

- MongoDB

<https://www.coursera.org/learn/introduction-mongodb>

<https://www.udemy.com/course/mongodb-the-complete-developers-guide/>

<https://www.udacity.com/course/data-wrangling-with-mongodb--ud032>

- Amazon DynamoDB

<https://www.coursera.org/learn/dynamodb-nosql-database-driven-apps>

- Azure Cosmos DB

<https://www.udemy.com/course/azure-cosmosdb/>

- CouchDB

<https://www.udemy.com/course/understanding-couchdb/>

# ONLINE KURZY

- Cloud Firestore

<https://www.udemy.com/course/firebase-course/>

- MarkLogic Server

<https://www.udemy.com/course/marklogic-fundamentals/>

- Elasticsearch

<https://www.udemy.com/course/elasticsearch-complete-guide/>

# LITERATURA

- HARRISON, Guy. *Next generation databases: NoSQL, NewSQL, and Big Data*. ISBN 978-1484213308.
- SULLIVAN, Dean. *NoSQL for Mere Mortals*. ISBN 978-0134023212.
- TIWARI, Shashank. *Professional NoSQL*. ISBN 978-0470942246.



# A PŘÍŠTĚ?

- MongoDB





Děkuji za pozornost.  
Otázky?

