



Západočeská univerzita v Plzni

Fakulta aplikovaných věd

Semestrální práce z KIV/DB2

Semestrální práce č.2 - JSON

Jakub Schenk

A21N0069P

schenkj@students.zcu.cz

13. dubna 2022

Obsah

1	Zadání	2
2	Úvod	3
3	Struktura XML	3
3.1	Struktura Hráče	3
3.2	Struktura Videohry	4
3.3	Struktura tabulky Hraje	4
4	Dotazy nad databází	5
4.1	Dotaz obsahující logický operátor, porovnávající operátor, výčtový operátor a projekci	5
4.2	Dotaz s využitím regulárního výrazu a podmínka na (ne)existenci hodnoty	5
4.3	Dotaz využívající JSON array	5
4.4	Dotaz s účelem aktualizovat u záznamu vydavatele na novou hodnotu	5
4.5	Dotaz pro zmměnu několika záznamů	5
4.6	Dotaz pro vyzkoušení agregace	6
5	Závěr	6
5.1	Body k zamyšlení a diskuzi	6
5.2	Shrnutí	6

1 Zadání

Připravte realistické datové JSON soubory týkající se nahlášeného tématu. Vložte soubory do dokumentové databáze MongoDB. Vytvořte sadu příkazů zodpovídající reálné otázky. Vytvořte dokumentaci vysvětlující použitou datovou strukturu a popisující dotazy přirozeným jazykem. Diskutujte svoje řešení. Požadavky:

- Alespoň 10 samostatných JSON souborů rozdělených do alespoň 2 kolekcí.
- JSON soubory obsahují zanořené JSON objekty.
- JSON soubory obsahují JSON Array, tedy vícenásobnou hodnotu
- Alespoň 3 vyhledávací příkazy. `db.xyz.find()`
- V sadě příkazů je alespoň jednou použitý (lze kombinovat):
 - logický operátor;
 - operátor porovnávající numerické hodnoty;
 - výčtový operátor (`$in`);
 - regex výraz pro textovou hodnotu;
 - podmínka na (ne)existenci hodnoty;
 - práce s JSON Array;
 - nastavení projekce.
- Alespoň 2 update příkazy pro změnu obsahu některých dokumentů. `db.xyz.update()`
- Alespoň 1 agregační dotaz pro spojení dat z různých kolekcí. `db.xyz.aggregate($lookup)`

2 Úvod

Za své téma pro tabulky v relaci m:n jsem si vybral Hráč-hraje-Videohry. Toto téma jsem si zvolil, protože je mi relativně blízké, splňuje zadání a přijde mi poměrně názorné.

Jako atributy hráče jsem zvolil id jakožto primární klíč, jeho jméno a příjmení, věk a výpis videoherních konzolí na kterých potencionálně hraje videohry. Pro videohry jsem zvolil opět id jako primární klíč, název videohry, vydavatele, věkovou hranici a výčet videoherních konzolí, na kterých si hráč může videohru zahrát. Spojující tabulka se nazývá "Hraje" a obsahuje virtuální záznam nějakého časového období, během kterého hráči z tabulky Hráči hrají hry z tabulky Hry. Tato tabulka má dva foreign keys, kterými jsou id z tabulky Hráči a id z tabulky Hry.

Důvodem, proč jsem si vybral session (neboli něco jako záznam neurčitého časového období) namísto vlastnictví nějaké hry je, že z mého pohledu session spíše odpovídá zadání (hráč-hraje-videohry), protože tyto hry jsou často zdarma a tedy by byla téměř každá hra u téměř každého hráče, a protože hráči u her nezůstávají pořád, a tedy by se tabulka musela poměrně často měnit. Jedná se pouze o něco na názorové úrovni a nebylo by tak jiné, kdyby tam bylo "vlastnictví" nebo "pravidelnější hraní" videoher, ale session je dle mého názoru nejreálnější a pro účely této práce stačí.

3 Struktura XML

3.1 Struktura Hráče

Podobně jako u XML jsem se rozhodl u jednotlivých hráčů uvádět "id" jakožto primární klíč a celočíselná hodnota, která by měla být unikátní. "jmeno" je obsahující dva podzáznamy "krestne" a "prijmeni". Ty předvídatelně obsahují křestné jméno a příjmení hráče. "vek" je celočíselná hodnota, která není povinná a obsahuje věk hráče. Nakonec "typ_konzole" je set hodnot, který obalují jednotlivé konzole, na kterých hráč může hrát své videohry.

```
"hrac" : {  
    "id" : 1,  
    "jmeno" : {  
        "krestne" : "Ales",  
        "prijmeni" : "Bodry"  
    },  
    "vek" : 7,  
    "typ_konzole" : ["PC", "Android"]  
}
```

```

    }
    ...

```

3.2 Struktura Videohry

Podobně jako u XML jsem se rozhodl u jednotlivých videoher uvádět. "id" jakožto primární klíč a celočíselná hodnota, která by měla být unikátní. "nazev" je představuje název titulu. "vydavatel" pak představuje vydavatele/společnost která vydala/provozuje tuto videohru. "vekova_hranice" je celočíselná hodnota, která není povinná a jedná se o doporučení pro hráče, které říká jakého věku by měl hráč být před tím, než bude tuto hru hrát. Nakonec "podpora_konzole" je set hodnot, které značují jednotlivé konzole, na kterých si hráč může zahrát tuto videohru.

Pro jednoduchost jsem rozhodl o jednoduchých a unikátních "id" a "vydavatelích" videoher. Někteří vydavatelé jsou vymyšlení a myslím, že neexistuje seznam všech videoher, s přidělenými unikátními ID.

```

"videohra" : {
    "id" : 3,
    "nazev" : "Counter Strike",
    "vydavatel" : "Valve",
    "vekova_hranice" : 12,
    "podpora_konzole" : ["PC"]
}

```

3.3 Struktura tabulky Hraje

Poslední tabulku skládají data o session, neboli jaký hráč hraje/hrál jakou hru (v nějakém období). Tato tabulka obsahuje dvojici cizích klíčů, a to sice primární klíče ostatních tabulek, a to sice id hrace a id hry. Tato dvojice pak představuje složený primární klíč. Celý soubor nese informaci o nějakém časovém období během kterého kteří hráči hráli které hry.

```

"hraje" : {
    "id_hrace" : 1,
    "id_hry" : 18
}

```

4 Dotazy nad databází

Po nahrání jednotlivých JSON souborů do databáze (upravené příkazy pro MongoDB jsou v oddělených souborech začínajícími na "insert-"), jsem se začal dotazovat na data v tabulkách následujícími dotazy.

4.1 Dotaz obsahující logický operátor, porovnávající operátor, výčtový operátor a projekci

```
db.hraci.find({$or:[{vek:{gte:30}},{id:{in:[1,2,3]}]}},{jmeno:1,vek:1})
```

Tímto dotazem lze získat křestná jména a příjmení všech hráčů, kteří jsou starší 30 let, nebo mají $id = 1, 2$, nebo 3 .

4.2 Dotaz s využitím regulárního výrazu a podmínka na (ne)existenci hodnoty

```
db.hraci.find({$or:[{"jmeno.prijmeni":{"regex:/^H.*/"}},{vek:{exists:false}}]})
```

Jelikož v datech není žádný záznam s nevyplněným věkem, výsledkem jsou jména, věky a id hráčů, jejichž příjmení začíná písmenem "H".

4.3 Dotaz využívající JSON array

```
db.hry.find({podpora_konzole:{$elemMatch:{seq:'Android'}}})
```

Výsledkem dotazu by měl být seznam her, které podporují Android jakožto konzole, na které jde videohra hrát.

4.4 Dotaz s účelem aktualizovat u záznamu vydavatele na novou hodnotu

```
db.hry.updateOne({id:13},{set:{vydavatel:"Mojang"}})
```

Po provedení dotazu by se měl vydavatel u videohry s id 13 změnit z "Microsoft" na "Mojang".

4.5 Dotaz pro změnu několika záznamů

```
db.hry.updateMany({id:{in:[3,4,6]}},{set:{vekova_hranice:15}})
```

Po provedení dotazu je tabulka her aktualizována a hry s id 3, 4, a 6 mají přenastavenou hodnotu u věkové hranice na 15 let.

4.6 Dotaz pro vyzkoušení agregace

```
db.hraci.aggregate([{$lookup:{from:"hry",localField:"vek",foreignField:"vek"}}
```

Poměrně nesmyslný dotaz, který slouží jen pro edukativní účely a názornou ukázkou agregovaného dotazu.

Dotazem by měl uživatel získat seznam záznamů hráčů, který je u nich případně rozšířený o hry, které mají stejnou věkovou hranici, jako je věk hráče.

5 Závěr

5.1 Body k zamyšlení a diskuzi

Členění do souborů; jaký semantický význam má jeden soubor?

Při vytváření souborů pro mě měl jeden soubor význam jednoho většího celku zaobalujícího několik záznamů, které chci do tabulky uložit. Jedná se tedy o obalový prvek několika záznamů, které můžou mít něco společného

Způsob uložení relace mezi entitami; embedded JSON objekt vs uložit odděleně do jiné kolekce.

Limitace datového formátu JSON pro uložení dat z této domény.

Dotazy nad daty dokáží být poměrně nepřehledné a pokud jsou komplikovanější, tak začíná být náročnější jejich konstrukce.

Limitace v dotazování – na jaké otázky není možné nebo je těžké odpovědět? Opět nevím jestli jsem otázku správně pochopil, ale z mého pohledu je největší limitací uživatel a jeho schopnost vyjádřit co za data od systému potřebuje. To je ale poměrně obecný problém, když uživatel nedokáže zformulovat jasný dotaz.

Je problematická existence nepovinných údajů? Existenci nepovinných údajů nepovažuji za problematickou. Pokud je nad databází zkonstruován dotaz, který vyžaduje neexistující hodnotu, je vrácen prázdný výsledek.

5.2 Shrnutí

Práci s touto problematikou jsem se naučil jak pracovat s JSON formátem a MongoDB, která je myslím poměrně rozšířená. Vypracováním semestrální práce jsem si snad dostatečně připomenul látku z přednášek a cvičení.