



Semestrální práce z předmětu KIV/DB2

DATABÁZOVÁ TECHNOLOGIE XML (ORACLE)

STANISLAV KRÁL

A20N0091P

ZÁPADOČESKÁ UNIVERZITA V PLZNI
FAKULTA APLIKOVANÝCH VĚD

Obsah

1	Popis řešeného tématu	2
2	Popis datové struktury	3
2.1	Obchodní páry	3
2.2	Portfólia	4
2.3	Uskutečněné obchody	5
3	Popis dotazů nad databázovou strukturou	6
3.1	Vyhledání všech nákupů v daném portfóliu	6
3.2	Vyhledání všech obchodů s kryptoměnou o větším než daném množství	6
3.3	Vyhledání všech obchodních párů skládajících se z konkrétní měny	7
3.4	Vyhledání posledních obchodů ze všech importů nákupů . . .	7
3.5	Vyhledání obchodů z konkrétní burzy	8
4	Diskuze nad řešením v dané technologii	9

1 Popis řešeného tématu

Zvolené téma představuje situaci investora, který nakupuje a prodává kryptoměny na různých směnárnách či burzách. Předpokladem tohoto tématu je to, že služby realizující samotný nákup či prodej kryptoměn umožňují export všech uskutečněných obchodů ve formátu XML se specifickým schématem.

Pro takového investora je důležité, aby všechny informace o nákupech byly jednoduše dohledatelné a aby jednotlivé kryptoměny mohl třídit do portfolií. Tento investor by si mohl například chtít třídit nákupy do portfolia, ve kterém si bude vést obchody s významnějšími kryptoměnami (např. Bitcoin či Ethereum) a obchody s méně významnými kryptoměnami (dle tržní kapitalizace např. Basic Attention Token či Solana) do jiného portfolia.

Dále je také nutné, aby ke každému obchodu byla přiřazena informace v jaké měně byl uskutečněn. Toto se nejčastěji reprezentuje pomocí tzv. *obchodních párů*, které se skládají z měny (např. USD) a dané komodity (např. BTC).

Pomocí takto popsaného nástroje (databáze), který tento problém řeší, si investor může např. svá portfolia porovnat oproti aktuálnímu kurzu kryptoměn, a zjistit jak se jeho obchodům daří. Takový nástroj by také našel využití i při vyplňování daňového přiznání.

2 Popis datové struktury

V databázi jsou definovány obchodní páry skládající se ze jména komodity, zkratky komodity a měně, ve které se bude daná komodita obchodovat. Dále se v databázi nachází seznam portfólií, které jsou popsány identifikátorem, názvem a popisem. V těchto portfóliích se dále mohou vyskytovat jednotlivé obchody, které nesou základní informace o uskutečněném nákupu či prodeji a odkazují se na obchodní pár. Kořenovým prvkem každého XML souboru je prvek `crypto-tracker`.

Jednotlivé XML soubory jsou uloženy v tabulce `crypto_orders_import_xml`, která má následující strukturu:

- `id` – identifikátor importu
- `import_data` – obsah souboru obsahující data pro import obchodů z burzy/směnárný ve formátu XML
- `import_timestamp` – časové razítko, kdy byl proveden import
- `exchange` – název burzy/směnárný, ze které je prováděn import

2.1 Obchodní páry

Seznam obchodních páru se vyskytuje v prvku `trading-pairs`. Jednotlivé obchodní páry jsou reprezentovány prvky `trading-pair` a v atributu `id` se nachází jejich identifikátory. V tomto prvku se pak dále vyskytují následující prvky popisující obchodní pár:

- `name` - název kryptoměny v obchodním páru
- `symbol` - zkratka názvu kryptoměny v obchodním páru
- `currency` - zkratka měny v obchodním páru

```
<trading-pairs>
  <trading-pair id="1">
    <name>Bitcoin</name>
    <symbol>BTC</symbol>
    <currency>USD</currency>
  </trading-pair>
  <trading-pair id="2">
    <name>Litecoin</name>
    <symbol>LTC</symbol>
```

```
        <currency>EUR</currency>
    </trading-pair>
</trading-pairs>
```

2.2 Portfólia

Seznam portfólií se vyskytuje v prvku **portfolios**. Jednotlivá portfólia jsou reprezentována prvky **portfolio** a v atributu **id** se nachází jejich identifikátory. Tyto prvky dále obsahují následující atributy:

- **name** – pojmenování portfólia
- **description** – vlastní popis portfólia

Pokud se v daném portfóliu vyskytují nějaké obchody, tak jsou reprezentovány prvky **order** ve vnořeném prvku **orders**.

Důležité je zmínit, že napříč více XML soubory v databázi se může vyskytovat prvek popisující jedno portfólio i více než jednou. Toto je způsobené tím, že jednotlivé nákupy portfólia se nachází v různých XML souborech. Avšak název a popis jednoho portfólia by se měl vyskytovat v XML souborech pouze jednou.

```
<portfolios>
  <portfolio id="1" name="Main Portfolio" description="Tracking
    Bitcoin/Ethereum transactions">
    <orders>
      ...
    </orders>
  </portfolio>
  <portfolio id="2" name="Altcoin Portfolio"
    description="Tracking altcoin portfolio">
    <orders>
      ...
    </orders>
  </portfolio>
  <portfolio id="3" name="Parents portfolio"
    description="Tracking investments of my parents"/>
</portfolios>
```

2.3 Uskutečněné obchody

Seznam obchodů daného portfolia se nachází v prvku **orders** vyskytujícím se v prvku **portfolio**, jež označuje portfolio, do kterého patří. Jednotlivé položky seznamu obchodů jsou reprezentovány prvky typu **order**, jehož atribut **trading-pair** obsahuje identifikátor obchodního páru, kterého se týká. Prvek **order** obsahuje následující atributy:

- **size** – množství kryptoměny, která byla obchodována
- **price** – cena za jednu minci kryptoměny
- **fee** – poplatek za uskutečněný nákup
- **timestamp** – datum a čas, kdy se nákup uskutečnil ve formátu UNIX Timestamp

Pokud daný obchod představoval prodej, tak prvek **order** obsahuje atribut **sell** nastavený na hodnotu **true**.

```
<orders>
  <order trading-pair="4">
    <size>10000</size>
    <price>0.16</price>
    <fee>1</fee>
    <timestamp>1618675561</timestamp>
  </order>
  <order trading-pair="2" sell="true">
    <size>5</size>
    <price>230</price>
    <fee>3</fee>
    <timestamp>1618675531</timestamp>
  </order>
</orders>
```

3 Popis dotazů nad databázovou strukturou

Ke zjištění užitečných informací o obchodování bylo vytvořeno 5 dotazů.

3.1 Vyhledání všech nákupů v daném portfóliu

Pro vyhledání všech nákupů patřících do nějakého portfólia slouží jednoduchý dotaz, ve kterém je vybrané portfólio specifikováno pomocí jeho identifikátoru. Výsledek dotazu pak obsahuje následující sloupce:

- `order_size` – množství kryptoměny, která byla obchodována
- `price` – cena za jednu minci kryptoměny
- `order_fee` – poplatek za uskutečněný nákup
- `timestamp` – datum a čas, kdy se nákup uskutečnil ve formátu UNIX Timestamp

Dotaz je formulován následovně:

```
SELECT query.*
FROM crypto_orders_import_xml,
     XMLTABLE('crypto-tracker/portfolios/portfolio[@id=1]/orders/order'
              PASSING import_data
              COLUMNS trading_pair_id NUMBER PATH '@trading-pair',
                      order_size NUMBER PATH 'size',
                      price NUMBER PATH 'price',
                      order_fee NUMBER PATH 'fee',
                      timestamp NUMBER PATH 'timestamp'
              ) query
;
```

3.2 Vyhledání všech obchodů s kryptoměnou o větším než daném množství

Pro vyhledání všech nákupů patřících do nějakého portfólia, ve kterých se jednalo o minimální množství vybrané kryptoměny, slouží následující dotaz:

```
SELECT query.*
FROM crypto_orders_import_xml,
     XMLTABLE('//portfolios/portfolio[@id=1]/orders/order[@trading-pair=1
and size >= 0.015]' PASSING import_data
```

```

        COLUMNS trading_pair_id NUMBER PATH '@trading-pair',
        order_size NUMBER PATH 'size',
        price NUMBER PATH 'price',
        order_fee NUMBER PATH 'fee',
        timestamp NUMBER PATH 'timestamp'
    ) query
;

```

Vybraná kryptoměna je referencována identifikátorem obchodního páru. Struktura výsledku dotazu je stejná jako dotazu pro vyhledání všech nákupů v daném portfoliu.

3.3 Vyhledání všech obchodních párů skládajících se z konkrétní měny

Pro vyhledání všech obchodních párů, které se skládají z měny USD slouží následující dotaz:

```

SELECT query.*
FROM crypto_orders_import_xml,
    XMLTABLE('//trading-pairs/trading-pair[contains(currency,
        "USD")]') PASSING import_data
        COLUMNS currency VARCHAR(128) PATH 'currency',
        symbol VARCHAR(128) PATH 'symbol',
        name VARCHAR(128) PATH 'name'
    ) query
;

```

Výsledek takového dotazu se pak skládá z následujících sloupců:

- `currency` – zkratka měny v páru
- `symbol` – zkratka kryptoměny v páru
- `name` – název kryptoměny v páru

3.4 Vyhledání posledních obchodů ze všech importů nákupů

Pro vyhledání všech posledních obchodů ve všech importech slouží následující dotaz, jehož výsledek má stejnou strukturu jako výsledek dotazu na všechny obchody:

```

SELECT query.*
FROM crypto_orders_import_xml,
     XMLTABLE('//portfolios/portfolio/orders/order[last()]' PASSING
              import_data
              COLUMNS trading_pair_id NUMBER PATH '@trading-pair',
                      order_size NUMBER PATH 'size',
                      price NUMBER PATH 'price',
                      order_fee NUMBER PATH 'fee',
                      timestamp NUMBER PATH 'timestamp'
              ) query
;

```

3.5 Vyhledání obchodů z konkrétní burzy

Pro vyhledání všech obchodů z vybrané burzy slouží následující dotaz, jehož výsledek má stejnou strukturu jako výsledek dotazu na všechny obchody:

```

SELECT query.*
FROM crypto_orders_import_xml,
     XMLTABLE('crypto-tracker/portfolios/portfolio/orders/order'
              PASSING import_data
              COLUMNS trading_pair_id NUMBER PATH '@trading-pair',
                      order_size NUMBER PATH 'size',
                      price NUMBER PATH 'price',
                      order_fee NUMBER PATH 'fee',
                      timestamp NUMBER PATH 'timestamp'
              ) query WHERE exchange = 'binance'
;

```

Filtrování obchodů konkrétní burzy je realizováno pomocí SQL konstrukce WHERE.

4 Diskuze nad řešením v dané technologii

Jednotlivé soubory představují exportované informace o nákupech z různých burz. Problematické je to, že na jednom místě jsou definované všechny informace o portfoliu (název a popis), avšak v dalších XML souborech, ve kterých se nachází další nákupy patřící do daného portfolia, je toto portfolio referencováno stejným prvkem, avšak atributy pro specifikaci názvu a popisu jsou vynechány. Portfolio je referencováno pomocí jeho identifikátoru. XML databáze pak pomalu degraduje na relační databázi a mohou vznikat konflikty, kdy jedno portfolio bude mít název a popis specifikovaný na různých místech.

Struktura řešení je navržena tak, že identifikátory, reference a popisy portfolií se nacházejí v attributech.

Datový formát pro uložení dat z této domény je vhodný spíše pro jednorázové zpracování a následné uložení do jiného formátu, nikoliv jako pro implementaci hlavního úložiště.

Limitace dotazování spočívá v tom, že obchodovací páry v jednotlivých obchodech jsou referencovány přes ID, a tak není možné v rámci jednoho dotazu získat obchody včetně příslušných obchodních párů.

Mezi vhodná metadata při ukládání do relační databáze patří datum exportu z burzy a zdroj, ze kterého export pochází.

Jako jediný nepovinný údaj se v databázové struktuře používá atribut `sell`, který označuje prodeje. Jeho přítomnost není problematická.