

Nástroj `lsql-csv` na zpracování CSV souborů z příkazového řádku

Přemysl Šťastný

Univerzita Karlova

28.06.2024

Úvod

- ▶ `lsql-csv` je nástroj pro provádění krátkých dotazů nad daty malých CSV souborů z shellu
- ▶ Implementuje nový jazyk LSQL podobný SQL, speciálně navržený pro práci se CSV soubory v shellu
- ▶ Designovým cílem LSQL je být lapidárnějším jazykem než SQL
- ▶ Účelem jeho návrhu je umožnit uživateli rychle psát jednoduché dotazy přímo do terminálu

Co je to CSV soubor?

- ▶ CSV je jednoduchý souborový formát určený pro výměnu tabulkových dat

Adam,21

"Petra",23

Karel,25

Příklad: Hello World

```
hello-world.txt:
```

```
World,Hello
```

```
lsql-csv 'hello-world.txt, &1.2 &1.1'
```

```
Hello,World
```

Příklad: Jednoduché filtrování

```
lsq1-csv -d: '/etc/passwd, &1.*, if &1.3 >= 1000'
```

```
nobody:x:65534:65534:nobody:/var/empty:/bin/false
```

```
me:x:1000:1000:~/home/me:/bin/bash
```

Pozn.: 3. sloupec je UID uživatele.

Příklad: Jednoduchý join

```
lsq1-csv -d: '/etc/{passwd,group}, &1.1 &2.1,  
    if &1.4 == &2.3'
```

```
root:root
```

```
bin:bin
```

```
daemon:daemon
```

```
me:me
```

Pozn.: 4. sloupec /etc/passwd je defaultní GID uživatele a 3. sloupec /etc/group je GID skupiny.

Příklad: Jednoduché seskupení

```
lsq-l-csv -d: 'p=/etc/passwd, p.7 count(p.3), by p.7'
```

```
/bin/bash:7
```

```
/bin/false:7
```

```
/bin/sh:1
```

```
/bin/sync:1
```

```
/sbin/halt:1
```

```
/sbin/nologin:46
```

```
/sbin/shutdown:1
```

Příklad: Jednoduché třídění

```
lsq1-csv -d: '/etc/passwd, &1.*,  
if &1.3 >= 1000, sort &1.3'
```

```
me1:x:1000:1000::/home/me1:/bin/bash
```

```
me2:x:1001:1001::/home/me2:/bin/bash
```

```
me3:x:1002:1002::/home/me3:/bin/bash
```

```
nobody:x:65534:65534:nobody:/var/empty:/bin/false
```


Příklad: Komplikovanější join

```
lsq1-csv -d: 'p=/etc/passwd g=/etc/group,  
p.1 cat(g.1","), if p.1 in g.4, by p.1'
```

```
adm:adm,disk,sys,  
bin:bin,daemon,sys,  
daemon:adm,bin,daemon,  
lp:lp,  
mythtv:audio,cdrom,tty,video,  
news:news,
```

Pozn.: 1. sloupec /etc/passwd je název uživatele, 1. sloupec /etc/group je název skupiny a 4. sloupec /etc/group je seznam uživatelů skupiny.

Alternativní řešení

- ▶ SQL
- ▶ Standardní linuxové nástroje
- ▶ Jiné SQL implementace pro CSV (q, CSV SQL,...)
- ▶ Obecné programovací jazyky

Jak se `lsql-csv` liší od SQL?

- ▶ `lsql-csv` dotazy jsou v mnoha usecasech lapidárnější než SQL dotazy
- ▶ Data a jazyk SQL jsou typed. `lsql-csv` a CSV soubory jsou typeless.
- ▶ Import CSV souboru do SQL databáze může být vhodnější řešení, pokud jsou zapotřebí komplexnější queries, jsou zapotřebí joiny většího množství tables, nebo je zapotřebí provést velké množství queries.

Jak se `lsql-csv` liší od SQL?

```
SELECT dataX FROM data.txt WHERE dataX > 1000;
```

```
data.txt, &1.2, if &1.2 > 1000
```

Rozdíly v použití `lsq1-csv` od standardních linuxových nástrojů při zpracování CSV

- ▶ V některých případech dokáže být `lsq1-csv` čitelnější a kratší

Rozdíly v použití lsq1-csv od standardních linuxových nástrojů při zpracování CSV

```
lsq1-csv -d: '/etc/{passwd,group}, &1.1 &2.1,  
if &1.4 == &2.3'
```

```
sort -t: -k3,3 /etc/group >/tmp/group.sort  
sort -t: -k4,4 /etc/passwd >/tmp/passwd.sort
```

```
join -t: -14 -23 /tmp/passwd.sort /tmp/group.sort |  
cut -d: -f2,8
```

Kde je balíček nahraný?

- ▶ GitHub - zdrojové kódy
- ▶ Hackage - hlavní zdroj pro distribuci balíčku
- ▶ Stackage
- ▶ NixOS

```
cabal install lsq1-csv
```