

DĚLAT
DOBRÝ SOFTWARE
NÁS BAVÍ

PROFINIT

A4M33BDT

Technologie pro velká data

Hive cvičeni

Jan Hučín
Sergii Stamenov

28.3.17

Osnova

1. Test
2. Stručné opakování Hive
3. Samostatná práce s Hive



TEST

Hive opakování

Hive cvičení

Vytvoření tabulek a nahrávání dat

- › Z adresáře na HDFS /user/pascep/ zkopírovat data do vlastního adresáře na HDFS.
- › Vytvořit dočasnou externí tabulku a nahrát data
- › Vytvořit interní tabulku uloženou jako parquet a nahrát data z dočasné tabulky. Zkonvertovat teplotu z Fahrenheita na Celsia.

Struktura datového souboru

- stanice string
- mesic int
- den int
- hodina int
- teplota double (F * 10)
- flag char(1)
- latitude double
- longitude double
- vyska double
- stat char(2)
- nazev string

Vytvoření dočasné tabulky

```
create external table pocasi_tmp (  
  stanice string,  
  mesic int, den int, hodina int,  
  teplota double,  
  flag string,  
  latitude double, longitude double,  
  vyska double, stat string, nazev string)  
ROW FORMAT  
DELIMITED FIELDS TERMINATED BY ',' LINES TERMINATED BY '\n'  
STORED AS TEXTFILE  
LOCATION "/user/username/teplota-usa/";
```

Vytvoření interní tabulky

```
CREATE TABLE IF NOT EXISTS
pocasi (
  stanice string,
  mesic int,
  den int,
  hodina int,
  teplota double,
  flag string,
  latitude double,
  longitude double,
  vyska double,
  stat string,
  nazev string
)
STORED AS parquet;
```

```
INSERT OVERWRITE TABLE pocasi
SELECT
  stanice,
  mesic,
  den,
  hodina,
  ((teplota / 10) - 32) * 5/9,
  flag,
  latitude,
  longitude,
  vyska,
  stat,
  nazev
FROM pocasi_tmp
WHERE mesic IS NOT NULL;
```


Ukoly

- › Najit stát s nejvyšší průměrnou teplotou v letě
| stat | avg_teplota |
- › Pro každou sezonu najit stát s nejvyšší průměrnou teplotou (jedním dotazem)
| sezona | stat | avg_ teplota |
- › Najit nejvyšší pokles/stoupaní teploty
stat | den | den + 1 | delta

Domácí úkol

- › Vytvořit tabulku s partitioningem po měsících
- › Udělat srovnání výkonu dotazu pro různé formáty ukládání tabulky (Parquet vs Avro vs CSV)
- › Udělat srovnání výkonu dotazu pro různé execution enginey: Spark, Tez, Map-Reduce

Děkujeme za pozornost

PROFINIT

Profinit, s.r.o.
Tychonova 2, 160 00 Praha 6



Telefon
+ 420 224 316 016



Web
www.profinit.eu



LinkedIn
linkedin.com/company/profinit



Twitter
twitter.com/Profinit_EU