

PO12: Przetwarzanie Big Data 22/23-Z

Rozpoczęto

środa, 9 listopada 2022, 13:29

Stan

Ukończone

Ukończono

środa, 9 listopada 2022, 13:39

Wykorzystany czas

9 min. 32 sek.

Punkty

25,00/26,00

Ocena

9,62 pkt. na 10,00 pkt. możliwych do uzyskania (96,15%)

06

Korzystając z poniższego polecenia uzyskaj odpowiedź na poniższe pytania:

set `{option_name}`;

a. Jaki silnik wykonawczy jest wykorzystywany (hive.execution.engine)?
tez ✓

b. Czy wykorzystywany jest optymalizator kosztowy (hive.cbo.enable)?
tak ✓

c. Is the file compaction mechanism running (hive.compactor.initiator.on)?
nie ✓

d. Are statistics gathered for new tables? (hive.stats.autogather)?
tak ✓

Pytanie 1

Poprawnie

Punkty: 4,00 z 4,00

Oflaguj pytanie

10

Za pomocą polecenia load data załaduj dane do nowoutworzonej tabeli

load data local inpath "/home/(your_username)/beers.csv"
into table beers_sf;

Sprawdź czy Twoje dane znalazły się w tabeli beers_sf. Wykonaj w tym celu polecenie

select count(*) from beers_sf;

Kilka pytań:
a. Dlaczego to tak długo trwało?
b. Co w tym czasie się działo?
c. Rzuć okiem na interfejs menadżera zasobów klastra YARN, jaki typ aplikacji jest obecnie uruchomiony?
d. Przejdź do interfejsu silnika TEZ. Czy zauważasz DAG, który został wykonany w ramach silnika TEZ?
e. Przejdź do szczegółów tego DAG. Z ilu węzłów składa się ten DAG?
2 ✓

f. Jaka klasa przetwarzania (Processor Class) odpowiadała za implementację pierwszego z węzłów? MapTezProcessor ✓
g. Jaka klasa przetwarzania odpowiadała za implementację drugiego z węzłów? ReduceTezProcessor ✓
h. Czy jest dla Ciebie jasne dlaczego akurat takie klasy zostały użyte i jak wyglądało to przetwarzanie?

Pytanie 2

Poprawnie

Punkty: 3,00 z 3,00

Oflaguj pytanie

1. 11

Kontynuujemy naszą pracę z danymi. Utwórz tabelę docelową w formacie ORC, a następnie przepisz dane z tabeli tekstowej do docelowej korzystając z polecenia DML.

create table beers_orc(line int, abv float, ibu float,
id int, name string, style string, brewery_id int, ounces float)
STORED AS ORC;

insert into beers_orc select * from beers_sf
where id is not null;

Znowu trochę to trwało. Zaglądnij do interfejsu silnika TEZ. Zobacz do szczegółów nowego DAG.
a. Ile tym razem było węzłów?
2 ✓
b. Czy taka liczba węzłów była przez Ciebie oczekiwana? Zastanów się do czego wykorzystywane były poszczególne węzły. Jeśli masz z tym problem zapytaj prowadzącego

Pytanie 3

Poprawnie

Punkty: 1,00 z 1,00

Oflaguj pytanie

20

20. Zanim zaczniemy przetwarzanie naszych danych sprawdźmy jeszcze organizację tabeli beers_orc. Uruchom ponownie polecenie wyświetlające statystyki na poziomie tabeli
A następnie poszukaj stosownych informacji i odpowiedz na następujące pytania:
a. Ile plików wykorzystywanych jest dla danych tabeli beers_orc?
Zastanów się nad tym, kiedy ta liczba plików może wzrosnąć, oraz czy gdy wzrośnie, to czy może także zmaleć?
1 ✓
b. Korzystając z przeglądarki plików HDFS dostępnej w ramach interfejsu węzła nazw znajdź katalog zawierający dane tej tabeli.
Jak nazywa się ten katalog?
beers_orc ✓
c. Ile jest w nim plików? (Czy jest to zgodne z liczbą uzyskaną ze statystyk?)
1 ✓
d. Jaki jest typ tego pliku (bazowy? zawierający deltę zmian?)
bazowy ✓
e. Jaka jest nazwa pliku?
000000_0 ✓
f. Czy zawartość tabeli beers_orc jest skompresowana?
Nie ✓
g. Czy tabela została podzielona na kubelki?
Nie ✓
h. Czy dane zostały posortowane?
Nie ✓
i. Jaka biblioteka SerDe odpowiada za serializację i deserializację danych? Podaj jej pełną nazwę.
org.apache.hadoop.hive ql.io.orc.OrcSerde ✓

Pytanie 4

Poprawnie

Punkty: 9,00 z 9,00

Oflaguj pytanie

21

Na początek coś prostego. Jakie 3 craftowe marki piw w USA mają największą wartość ibu?
(3 piwa o największej goryczce).

Wybierz wszystkie poprawne:
☐ a. Rise of the Phoenix
☒ b. Bitter Bitch Imperial IPA ✓
☒ c. Troopers Alley IPA ✓
☒ d. Dead-Eye DIPA ✓
☐ e. Pub Beer
☐ f. Rail Yard Ale (2009)

Twoja odpowiedź jest poprawna.

Poprawnymi odpowiedziami są: Bitter Bitch Imperial IPA, Troopers Alley IPA, Dead-Eye DIPA

Pytanie 5

Poprawnie

Punkty: 1,00 z 1,00

Oflaguj pytanie

22

Jaka jest średnia ilość alkoholu w craftowych markach piw w USA?
Podaj wartość w procentach z dokładnością do jednej dziesiątej.

Odpowiedź: 5,9 ✓

Poprawna odpowiedź to: 5,9

Pytanie 6

Poprawnie

Punkty: 1,00 z 1,00

Oflaguj pytanie

23

Jakie cztery typy/style piw charakteryzują się największą średnią ilością alkoholu?

Wybierz wszystkie poprawne:
☒ a. English Barleywine ✓
☒ b. American Barleywine ✓
☐ c. American Double / Imperial Stout
☐ d. Russian Imperial Stout
☐ e. Tripel
☐ f. Baltic Porter
☒ g. American Malt Liquor ✓
☐ h. American Double / Imperial IPA
☒ i. Quadrupel (Quad) ✓
☐ j. Belgian Strong Pale Ale

Twoja odpowiedź jest poprawna.

Poprawnymi odpowiedziami są: English Barleywine, Quadrupel (Quad), American Barleywine, American Malt Liquor

Pytanie 7

Poprawnie

Punkty: 1,00 z 1,00

Oflaguj pytanie

24

W którym mieście warzy się markę piwa o największej ilości alkoholu?

Wybierz jedną odpowiedź:
☐ a. New York
☐ b. Brooklyn
☐ c. Evansville
☒ d. Boulder ✓
☐ e. Louisville

Twoja odpowiedź jest poprawna.

Poprawna odpowiedź to: Boulder

Pytanie 8

Poprawnie

Punkty: 1,00 z 1,00

Oflaguj pytanie

25

Uporządkuj nazwy stanów pod względem malejącej największej wartości ibu w piwach warzonych na ich terenie. Podaj nazwy trzech pierwszych.

Wybierz wszystkie poprawne:
☒ a. Massachusetts ✓
☐ b. Vermont
☐ c. Ohio
☒ d. Oregon ✓
☒ e. Virginia ✓

Twoja odpowiedź jest poprawna.

Poprawnymi odpowiedziami są: Oregon, Virginia, Massachusetts

Pytanie 9

Poprawnie

Punkty: 1,00 z 1,00

Oflaguj pytanie

27

Wygeneruj za pomocą polecenia EXPLAIN realizację tego zapytania.
Zauważasz elementy optymalizacji?
Np. filter down albo projekcja na wczesnym etapie przetwarzania?
a. Która z tabel została rozesłana w całości (broadcasted) do wszystkich węzłów w celu wykonania operacji połączenia?
states_orc ✓
b. Ile operacji mapowania zostało wykorzystanych do realizacji tego zapytania.
3 ✓
c. Ile operacji redukcji było wykorzystanych w tym samym zapytaniu?
2 ✓
d. Które klauzule w zapytaniu odpowiadały za operację redukcji?
☐ SELECT ✗
☐ FROM
☐ WHERE
☒ GROUP BY ✓
☒ ORDER BY (LIMIT) ✓

Punkty: 0,00 z 1,00
Poprawna odpowiedź to:
• GROUP BY
• ORDER BY (LIMIT)

Jeśli czegoś nie rozumiesz podyskutuj o tym z prowadzącym.

Pytanie 10

Częściowo poprawnie

Punkty: 3,00 z 4,00

Oflaguj pytanie

Nawigacja w teście

1	2	3	4	5	6
✓	✓	✓	✓	✓	✓
7	8	9	10		
✓	✓	✓	✗		

Pokaż jedną stronę na raz
Zakończ przegląd

Jesteś zalogowany(a) jako Andrzej Kapczyński (Wyloguj)
WIIiT-Inf-st-I-sem7-PrzBigDat-22/23-Z

Zakończ przegląd