Zadanie 1

Czy Hive wspiera indeksy? Jeśli staniesz przed problemem – czy da się szybko wyciągać dane?

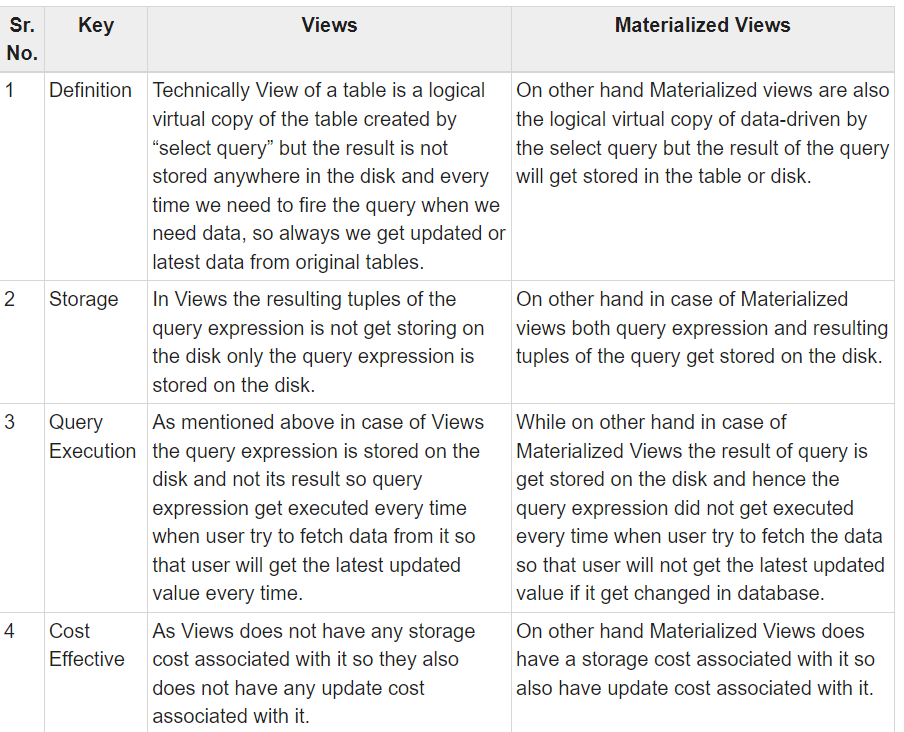
Odpowiedz na pytanie i podaj przykłady jak rozwiążesz problem indeksów w Hive.

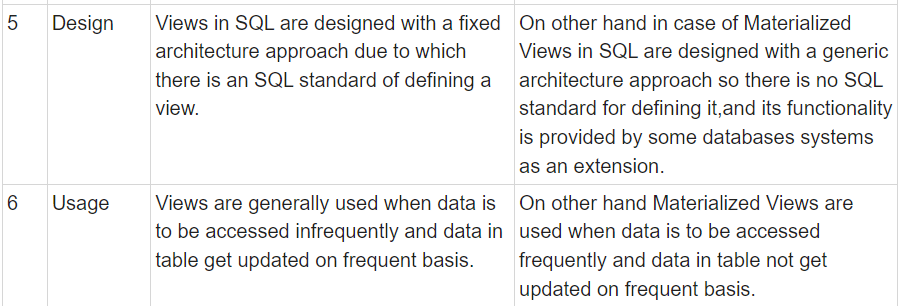
Celem indeksów było zwiększenie szybkości wykonywania zapytań na danych kolumnach tabeli.

Bez stosowania indeksów zapytania zawierające np. 'WHERE tab1.col1 = 10' ładowały całą tabelę lub jej partycję i przeszukiwały wszystkie wiersze. W odróżnieniu, jeśli indeks był na kolumnie col1, wtedy wystarczyło załadować i przeszukać tylko część pliku.

* Indeksy są usunięte od wersji 3.0
* Inne sposoby na szybkie wyciąganie danych:

1. materialized view with automatic rewriting:



****

1. Using columnar file formats ([Parquet](https://cwiki.apache.org/confluence/display/Hive/Parquet), [ORC](https://orc.apache.org/docs/indexes.html)) – they can do selective scanning; they may even skip entire files/blocks.