

Czasopismo: World Wide Web

Artykuł: "Pareto-based cache replacement for YouTube"

Opracowanie: Adrian Belciak

Opracowanie

Serwis społecznościowy YouTube.com to największa na świecie witryna tworzona przez użytkowników. Codziennie użytkownicy oglądają 2 miliardy filmów w tym serwisie.

Dla użytkownika końcowego najważniejszy jest szybki dostęp do interesującego go filmu, musi być również on szybko buforowany. Ze względu na wydajność dysków (nawet SSD) odczyt danych z dysku jest zbyt kosztowny. Z pomocą przychodzi technika **Memcached** – czyli trzymanie informacji w pamięci RAM (filmów Youtube). Rozwiązanie to zostało zaimplementowane przez firmę Danga Interactive, jest oparte na tablicy mieszającej i rozproszone jest na wielu maszynach. Dzięki temu można zmniejszyć liczbę żądań do zdalnego pobrania treści z dysków.

Wyróżniamy 4 rodzaje algorytmów Memcached

- **Pareto-based Least Frequently Used (PLFU)** - algorytm oparty na zasadzie Pareto, usuwamy najrzadziej używany film (najniższa częstotliwość odtworzeń),
- **Pareto-based Least Recently Used (PLRU)** - algorytm oparty na zasadzie Pareto, usuwamy najdawniej używany film (najdłuższy okres czasu od ostatniego użycia),
- **Least Frequently Used (LFU)** – usuwamy najrzadziej używany film,
- **Least Recently Used (LRU)** – usuwamy najdawniej używany film.

Zasada Pareto - znana również jako reguła 80/20, prawo mocy lub prawo Zipfa, 80% wyników wypływa tylko z 20% przyczyn. Mniejszymi środkami oraz wysiłkiem można osiągnąć dużo większe korzyści czy efekty. Celem jest zmniejszenie liczby przypadków, w których wideo żądane przez użytkowników nie może zostać znalezione w Memcached. 20% najpopularniejszych filmów to 80% całkowitego dostępu do filmów

Zasada działania

- Jeśli którykolwiek z popularnych filmów był już w Memcached, wideo będzie nadal przechowywane w Memcached.
- Jeśli którykolwiek z popularnych filmów nie jest jeszcze w Memcached, algorytm prosi PLFU lub PLRU o umieszczenie wideo w Memcached.
- Jeśli którykolwiek z niepopularnych filmów był już w Memcached, wideo zostanie potraktowane jako kandydat do usunięcia, gdy brakuje miejsca w Memcached.

Autorzy artykułu wykonali symulację działania algorytmów, aby wybrać najlepszy (testy dla dwóch innych kategorii oraz dla dwóch rozmiarów Memcached). W rezultacie ustalili, że kategoria filmów w serwisie Youtube nie ma wpływu na wydajność wszystkich algorytmów. Dzięki zastosowaniu zasady Pareto udało się utrzymać więcej potrzebnych danych w Memcached. Najlepszym algorytmem był PLFU - działa on niezależnie od testowanej kategorii wideo i wielkości Memcached. Zastosowanie PLFU może radykalnie zmniejszyć liczbę wyszukiwanych filmów z wewnętrznej bazy danych YouTube i skrócić czas żądania pobrania filmów. PLRU używany przez dłuższy czas, zapewnia drugą najlepszą wydajność