Politiche di scheduling di threads

mercoledì 12 aprile 2023 23:02

Politiche di scheduling di threads

Attualmente siamo al kernel 5,
load sharing lo abbiamo in kernel 2.4

Load sharing (e.g. Linux 2.4)

Esiste un'unica coda globale di threads pronti ad eseguire.

**Abbiamo la possibilità quando si libera una CPU, ad esempio la 0, di andare a prendere qualsiasi thread, che ripetiamo, essi stanno all'interno di un'unica struttura.
Questa è una condivisione perfetta del carico perché ogni thread può andare in esercizio su ogni CPU ad ogni istante di tempo, quindi anche sulla CPU1.
Su tutte le CPU.
E dobbiamo anche avere la possibilità di identificare cose più prioritarie e cose meno prioritarie.
Inoltre devo stare attento a non mandare in esercizio un thread su due CPU, perché il percorso d'esecuzione è unico. A qualsiasi istante di tempo, il thread deve poter utilizzare una sola CPU.**

Supponiamo di avere un insieme di thread che devono essere eseguiti, e un thread che viene ad essere eseguito sulla CPU0, più avanti il medesimo thread passa ad eseguire sulla CPU1, qui abbiamo sicuramente un problema di caching: l'impatto che il thread ha avuto sulla cache durante l'esercizio in CPU0, può non essere visibile alla CPU1. Quindi appena abbiamo lo spostamento del thread dalla CPU0 alla CPU1 chiaramente le informazioni devono essere riportate all'interno dell'architettura di cache visibile alla CPU1, per permettere l'accesso a questo thread.

Load balancing (e.g. Linux 2.6 e successivi)

I threads non sono più all'interno di un'unica coda globale, ma sono registrati all'interno di code separate e queste code sono una per ogni CPU-CORE.

Abbiamo un insieme di Thread Control Block che sono gestiti dalla CPU0 e un insieme di Thread Control Block gestiti dalla CPU1, qualora avessimo solo due CPU-CORE.

· code di threads pronti ad eseguire separate, una per ogni CPU-core
· migliore scalabilità delle operazioni di scheduling su macchine altamente parallele
· spostamento periodico di thread da una coda ad un'altra in caso di sbilanciamento del carico