

Cache memory الذاكرة المؤقتة

In addition to working with the main memory, processors also work with a special type of high-speed memory referred to as **cache** (pronounced cash). In fact, most of the time processors work directly with various types of cache memory and this cache memory, in turn, works with the main memory. Essentially, the cache memory acts as a high-speed buffer in between the processor and main memory, shuffling data into the processor as it needs it, or requests it. As a result, the processor takes advantage of the high-speed cache memory and therefore works faster, which, in turn, makes the computer that the processor drives, operate faster. Figure 5-1 illustrates the concept.

بالإضافة إلى العمل مع الذاكرة الرئيسية ، تعمل المعالجات أيضاً مع نوع خاص من الذاكرة عالية السرعة يشار إليها باسم ذاكرة التخزين المؤقت (النقد الواضح). في الواقع ، تعمل المعالجات في معظم الأوقات بشكل مباشر مع أنواع مختلفة من ذاكرة التخزين المؤقت ، وتعمل ذاكرة التخزين المؤقت هذه بدورها مع الذاكرة الرئيسية. بشكل أساسي ، تعمل ذاكرة التخزين المؤقت كمخزن مؤقت عالي السرعة بين المعالج والذاكرة الرئيسية ، حيث تقوم بتبديل البيانات في المعالج حسب حاجتها أو طلبها. نتيجة لذلك ، يستفيد المعالج من ذاكرة التخزين المؤقت عالية السرعة وبالتالي يعمل بشكل أسرع ، مما يجعل الكمبيوتر الذي يقوم المعالج بتشغيله يعمل بشكل أسرع. يوضح الشكل 5-1 المفهوم.

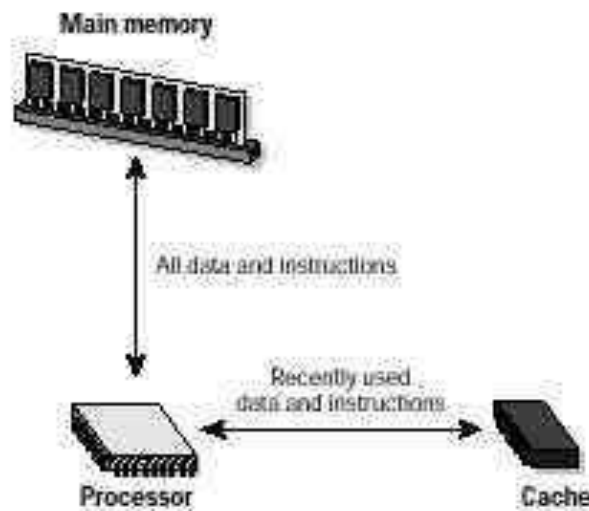


Figure 5-1: Cache helps make processors more efficient

Cache is a special form of high-speed memory that stores instructions and data the processor has recently used. Thanks to its proximity to the main computing engine inside the processor, and the fact that the processor often needs to reuse those instructions and data, cache keeps the processor busy and speeds up a computer's performance.

ذاكرة التخزين المؤقت هي شكل خاص من الذاكرة عالية السرعة تخزن التعليمات والبيانات التي استخدمها المعالج مؤخرًا. بفضل قربه من محرك الحوسبة الرئيسي داخل المعالج ، وحقيقة أن المعالج يحتاج في كثير من الأحيان إلى إعادة استخدام هذه التعليمات والبيانات ، فإن ذاكرة التخزين المؤقت تحافظ على انشغال المعالج وتزيد من سرعة أداء الكمبيوتر.

L1 and L2 cache

Cache memory, in its various forms, plays a particularly important role in a processor's performance. Cache can improve a processor's efficiency by offering it access to the data it needs more quickly than regular memory would. Not only are cache memory chips (**typically Static Random Access Memory, or SRAM**) faster than regular memory chips, but they also have a faster connection to the processor.

تلعب ذاكرة التخزين المؤقت ، بأشكالها المختلفة ، دورًا مهمًا بشكل خاص في أداء المعالج. يمكن أن تحسن ذاكرة التخزين المؤقت كفاءة المعالج من خلال منحه إمكانية الوصول إلى البيانات التي يحتاجها بسرعة أكبر من الذاكرة العادية. لا تعد رقائق ذاكرة التخزين المؤقت (عادةً ذاكرة الوصول العشوائي الثابتة أو SRAM) أسرع من رقائق الذاكرة العادية ، ولكنها تتمتع أيضًا باتصال أسرع بالمعالج.

The way cache works طريقة عمل ذاكرة التخزين المؤقت

Because of the way most software works, processors tend to spend a lot of their time either performing the same operations over and over or performing several different operations on the same set of data. Well, one day, somebody realized that if a processor could access used instructions and data more quickly, it could run much more efficiently. So, a clever designer came up with the idea to create a special "work area" right alongside the processor, called a **cache**, that temporarily stores the data and instructions the processor used most recently. The idea was (and is) that once the processor finishes what it is working on, it can "fetch" what it needs next from this nearby area instead of getting it from regular memory, which is further away and takes longer to get to.

نظرًا للطريقة التي تعمل بها معظم البرامج ، تميل المعالجات إلى قضاء الكثير من وقتها إما في إجراء نفس العمليات مرارًا وتكرارًا أو إجراء عدة عمليات مختلفة على نفس مجموعة البيانات. حسنًا ، في أحد الأيام ، أدرك شخص ما أنه إذا تمكن المعالج من الوصول إلى التعليمات والبيانات المستخدمة بسرعة أكبر ، فيمكن أن يعمل بكفاءة أكبر. لذلك ، جاء المصمم الذكي بفكرة إنشاء "منطقة عمل" خاصة بجوار المعالج ، تسمى ذاكرة التخزين المؤقت ، والتي تخزن مؤقتًا البيانات والتعليمات التي استخدمها المعالج مؤخرًا. كانت الفكرة (ولا تزال) أنه بمجرد انتهاء المعالج من العمل عليه ، يمكنه "جلب" ما يحتاجه بعد ذلك من هذه المنطقة المجاورة بدلاً من الحصول عليه من الذاكرة العادية ، والتي تكون بعيدة وتستغرق وقتًا أطول للوصول إليها.

Processors aren't the only computer-related component to use a cache. Many software programs, such as Web browsers, also use a cache. While a processor's cache and a browser's cache are not the same thing, they are conceptually similar. For both components, a cache speeds up access to recently used information.

المعالجات ليست المكون الوحيد المرتبط بالكمبيوتر الذي يستخدم ذاكرة التخزين المؤقت. العديد من البرامج ، مثل متصفحات الويب ، تستخدم أيضًا ذاكرة التخزين المؤقت. في حين أن ذاكرة التخزين المؤقت للمعالج وذاكرة التخزين المؤقت للمتصفح ليسا نفس الشيء ، إلا أنهما متشابهان من الناحية المفاهيمية.

بالنسبة لكلا المكونين ، تعمل ذاكرة التخزين المؤقت على تسريع الوصول إلى المعلومات المستخدمة مؤخرًا.

Web browsers, for example, set up memory and/or disk caches (which use RAM or space on the hard disk, respectively) to store recently used files. The thinking is that you probably will want to use the files you've accessed recently again. If they're stored either in RAM, or on disk, the browser will be able to get to them much more quickly than if it had to go out to the Internet again. For example, when you hit the Back button on your browser, the Web page loads almost instantly because the files the browser needs are nearby. If there wasn't a cache (or if you empty or clear the files in the cache), going back to the previous page would take just as long as when you called it up in the first place. This is similar to what happens with processors and their specialized cache; if the information the processor needs is close by in the cache, the processor operates quickly without waiting, but if the information isn't close by, the processor has to request it from main memory. (The main memory isn't as slow as the Internet, of course, but it is a lot slower than getting it from the cache.)

متصفحات الويب ، على سبيل المثال ، تقوم بإعداد الذاكرة و / أو ذاكرة التخزين المؤقت للقرص (التي تستخدم ذاكرة الوصول العشوائي أو مساحة على القرص الصلب ، على التوالي) لتخزين الملفات المستخدمة مؤخرًا. الفكرة هي أنك ربما تريد استخدام الملفات التي قمت بالوصول إليها مؤخرًا مرة أخرى. إذا تم تخزينها إما في ذاكرة الوصول العشوائي أو على القرص ، فسيكون المتصفح قادرًا على الوصول إليها بسرعة أكبر بكثير مما لو اضطر إلى الاتصال بالإنترنت مرة أخرى. على سبيل المثال ، عندما تضغط على الزر "رجوع" في متصفحك ، يتم تحميل صفحة الويب على الفور تقريبًا لأن الملفات التي يحتاجها المستعرض قريبة. إذا لم تكن هناك ذاكرة تخزين مؤقت (أو إذا قمت بإفراغ أو مسح الملفات الموجودة في ذاكرة التخزين المؤقت) ، فإن العودة إلى الصفحة السابقة سيستغرق نفس المدة التي استدعتها فيها في المقام الأول. هذا مشابه لما يحدث مع المعالجات وذاكرة التخزين المؤقت المتخصصة الخاصة بهم ؛ إذا كانت المعلومات التي تحتاجها المعالج قريبة في ذاكرة التخزين المؤقت ، فإن المعالج يعمل بسرعة دون انتظار ،

Lecture No. : 7

Cache memory

ولكن إذا لم تكن المعلومات قريبة ، فسيطلبها المعالج من الذاكرة الرئيسية. (الذاكرة الرئيسية ليست بطيئة مثل الإنترنت بالطبع ، لكنها أبطأ بكثير من الحصول عليها من ذاكرة التخزين المؤقت).

The process of going out to main memory to get more data and instructions also forces the cache to become "flushed," or emptied out. During the process of running programs, the processor regularly flushes the cache. This is important to know because it explains why it's possible to have too much of a good thing; that is, too much cache.

تؤدي عملية الخروج إلى الذاكرة الرئيسية للحصول على مزيد من البيانات والتعليمات أيضًا إلى إجبار ذاكرة التخزين المؤقت على "مسحها" أو تفرغها. أثناء عملية تشغيل البرامج ، يقوم المعالج بانتظام بمسح ذاكرة التخزين المؤقت. من المهم معرفة ذلك لأنه يشرح سبب إمكانية الحصول على الكثير من الأشياء الجيدة ؛ هذا هو ، الكثير من ذاكرة التخزين المؤقت.

The two most common types of cache are referred to as L1, or Level 1, and L2, or Level 2 cache. (It is possible to have Level 3 caches, but they are not very common.) Although technically speaking caches are a type of memory, in most cases the L1 and L2 cache are actually built into the processor chip or processor card itself. Thus, they're really more a feature of the processor than of memory.

Each level of cache is a separate chunk of memory and is treated independently by the processor. The two levels refer to how close the cache is physically located to the main number-crunching section of the processor. Figure 5-2 shows how the different caches work together with main memory.

يُشار إلى النوعين الأكثر شيوعًا من ذاكرة التخزين المؤقت باسم L1 ، أو المستوى 1 ، و L2 ، أو المستوى 2 من ذاكرة التخزين المؤقت. (من الممكن أن يكون لديك ذاكرة تخزين مؤقت من المستوى 3 ، لكنها ليست شائعة جدًا.) على الرغم من أن ذاكرات التخزين المؤقت من الناحية الفنية هي نوع من الذاكرة ، إلا أنه في معظم الحالات يتم تضمين ذاكرة التخزين المؤقت L1 و L2 في شريحة المعالج أو بطاقة المعالج نفسها. وبالتالي ، فهي في الحقيقة ميزة للمعالج أكثر من كونها ميزة للذاكرة.

Lecture No. : 7

Cache memory

كل مستوى من ذاكرة التخزين المؤقت هو جزء منفصل من الذاكرة ويتم معالجته بشكل مستقل بواسطة المعالج. يشير المستويان إلى مدى قرب ذاكرة التخزين المؤقت فعليًا من قسم طحن الأرقام الرئيسي في المعالج. يوضح الشكل 5-2 كيفية عمل ذاكرات التخزين المؤقت المختلفة مع الذاكرة الرئيسية.

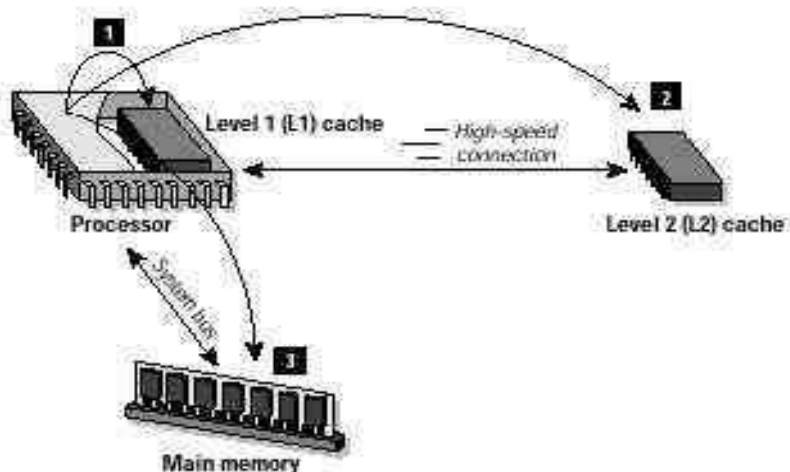


Figure 5-2: Multiple caches

The way a processor works with a system that has multiple caches is that the processor checks the L1 cache first, then the L2 cache, and then, finally, the main memory.

إن الطريقة التي يعمل بها المعالج مع نظام يحتوي على ذاكرة تخزين مؤقت متعددة هي أن المعالج يتحقق من ذاكرة التخزين المؤقت L1 أولاً ، ثم ذاكرة التخزين المؤقت L2 ، ثم أخيراً الذاكرة الرئيسية.

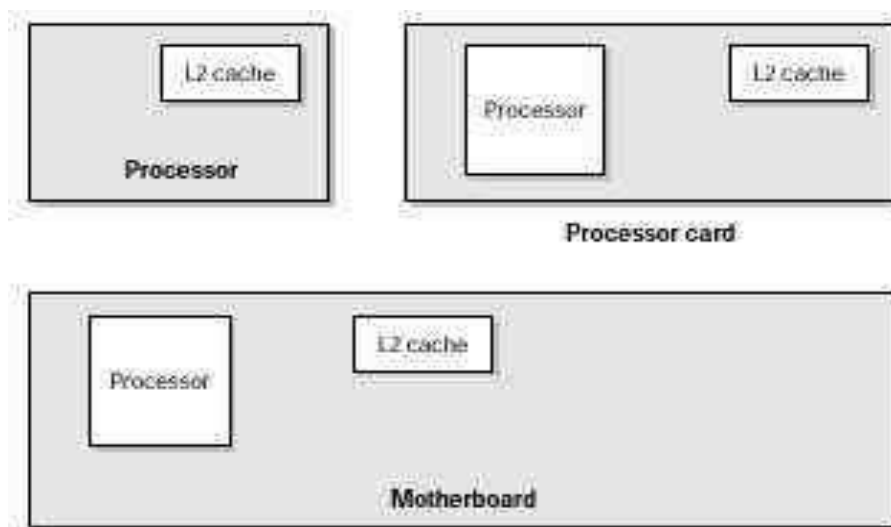
Traditionally, L1 cache, which is usually the smaller of the two, has been located on the processor itself and L2 cache has been located outside, but near the processor. Recent processor designs have begun to integrate L2 cache onto the processor card or into the CPU chip itself, much like L1 cache. This speeds up access to the larger L2 cache which, in turn, speeds up the computer's performance.

تقليدياً ، تم وضع ذاكرة التخزين المؤقت L1 ، والتي عادة ما تكون أصغر من الاثنين ، على المعالج نفسه وتم وضع ذاكرة التخزين المؤقت L2 بالخارج ، ولكن بالقرب من المعالج. بدأت تصميمات المعالج الحديثة في دمج ذاكرة التخزين المؤقت L2 في بطاقة المعالج أو في شريحة وحدة المعالجة المركزية نفسها

، مثل ذاكرة التخزين المؤقت L1. يؤدي ذلك إلى تسريع الوصول إلى ذاكرة التخزين المؤقت الأكبر L2 والتي بدورها تعمل على تسريع أداء الكمبيوتر.

Another traditional difference between L1 and L2 caches has been the speed at which the processor can access the different types of memory. Because L1 cache is integrated into the core of the microprocessor, it typically runs at the same speed as the CPU; so on a 500MHz processor, the connection speed to L1 cache is usually 500MHz. On older systems, the L2 cache often connected to the processor at the same speed as main memory. This speed is determined by a connecting route, called the computer's *system bus*, and typically runs at 66, 100, or 133MHz (although faster speeds are possible).

هناك اختلاف تقليدي آخر بين ذاكرات التخزين المؤقت L1 و L2 وهو السرعة التي يمكن للمعالج من خلالها الوصول إلى الأنواع المختلفة من الذاكرة. نظرًا لأن ذاكرة التخزين المؤقت L1 مدمجة في قلب المعالج الدقيق ، فإنها تعمل عادةً بنفس سرعة وحدة المعالجة المركزية ؛ لذلك في معالج 500 ميگاهرتز ، تكون سرعة الاتصال بذاكرة التخزين المؤقت L1 عادة 500 ميگاهرتز. في الأنظمة القديمة ، غالبًا ما يتم توصيل ذاكرة التخزين المؤقت L2 بالمعالج بنفس سرعة الذاكرة الرئيسية. يتم تحديد هذه السرعة من خلال مسار اتصال ، يسمى ناقل نظام الكمبيوتر ، ويعمل عادةً عند 66 أو 100 أو 133 ميگاهرتز (على الرغم من إمكانية الحصول على سرعات أعلى).



الشكل 6: مواقع التخزين المؤقت Figure 6: Cache locations

The L2 cache is located in different places on different processors. Some processors have the L2 cache integrated into the main chip itself, others have L2 cache on the circuit board that holds the processor, and still others work with L2 cache that's separate from the processor on the computer's motherboard.

توجد ذاكرة التخزين المؤقت L2 في أماكن مختلفة على معالجات مختلفة. تحتوي بعض المعالجات على ذاكرة تخزين مؤقت L2 مدمجة في الشريحة الرئيسية نفسها ، بينما تحتوي المعالجات الأخرى على ذاكرة تخزين مؤقت L2 على لوحة الدائرة التي تحتوي على المعالج ، ولا يزال البعض الآخر يعمل مع ذاكرة التخزين المؤقت L2 المنفصلة عن المعالج الموجود على اللوحة الأم للكمبيوتر.

On newer systems, however, where the L2 cache is located on a daughter card, such as most Pentium II and Pentium IIIs, or in the processor itself, as with the Celeron A, K6-3, and some mobile Pentium IIs (sometimes called Pentium II PEs, for performance enhanced) and Pentium IIIs (those designed for note-books), communication between the processor and the L2 cache occurs much more rapidly. On the Pentium II and III, for example, the processor-to-L2 cache connection is often via a *backside bus*; it runs faster than the system bus, but at half the speed of the processor. (This is sometimes referred to as a 1:2 ratio.) Again, with a 500MHz Pentium III processor, the processor-to-L2 cache connection speed is 250MHz. Additionally, systems that incorporate L2 cache on the chip itself feature a 1:1 ratio between the speed of the processor and the speed of the processor-to-L2 cache connection. So with a 500MHz processor, the connection to the L2 cache also runs at 500MHz.

ومع ذلك ، في الأنظمة الأحدث ، حيث توجد ذاكرة التخزين المؤقت L2 على بطاقة ابنة ، مثل معظم Pentium II و Pentium IIIs ، أو في المعالج نفسه ، كما هو الحال مع Celeron A و K6-3 وبعض أجهزة Pentium IIs المحمولة (تسمى أحياناً Pentium II PEs ، لتحسين الأداء) و Pentium IIIs (تلك المصممة لدفاتر الملاحظات) ، يحدث الاتصال بين المعالج وذاكرة التخزين المؤقت L2 بسرعة أكبر. في Pentium II و III ، على سبيل المثال ، غالباً ما يكون اتصال ذاكرة التخزين المؤقت من المعالج إلى L2 عبر ناقل خلفي

؛ يعمل بشكل أسرع من ناقل النظام ، ولكن بنصف سرعة المعالج. (يشار إلى هذا أحياناً على أنه نسبة 1:2). مرة أخرى ، باستخدام معالج Pentium III بسرعة 500 ميگاهرتز ، تكون سرعة اتصال ذاكرة التخزين المؤقت من المعالج إلى L2 هي 250 ميگاهرتز. بالإضافة إلى ذلك ، تتميز الأنظمة التي تتضمن ذاكرة التخزين المؤقت L2 على الشريحة نفسها بنسبة 1:1 بين سرعة المعالج وسرعة اتصال ذاكرة التخزين المؤقت من المعالج إلى L2. لذلك مع معالج 500 ميگاهرتز ، يعمل الاتصال بذاكرة التخزين المؤقت L2 أيضاً بسرعة 500 ميگاهرتز.

The faster your processor is, the more important it is to have a reasonable amount of L2 cache. In fact, without the proper amount of L2 cache, a processor often sits idle, "wasting cycles" as they say, which means your computer is not running as fast it can.

كلما كان المعالج أسرع ، زادت أهمية أن يكون لديك قدر معقول من ذاكرة التخزين المؤقت L2. في الواقع ، بدون المقدار المناسب من ذاكرة التخزين المؤقت L2 ، غالباً ما يجلس المعالج في وضع الخمول ، "يضيع الدورات" كما يقولون ، مما يعني أن جهاز الكمبيوتر الخاص بك لا يعمل بأسرع ما يمكن.

This lack of L2 cache explains why some of the early Celeron-based computers had relatively poor performance. The original Celeron essentially wasted a great deal of its processing power. The upgraded Celeron A chip and all current Celerons (including the mobile versions), however, incorporate some L2 cache on the processor itself, and dramatically improve the performance of computers using the "A" version of the Celeron.

يفسر هذا النقص في ذاكرة التخزين المؤقت L2 سبب ضعف أداء بعض أجهزة الكمبيوتر القديمة المستندة إلى سيليرون نسبياً. أهدرت Celeron الأصلية بشكل أساسي قدرًا كبيرًا من قدرتها على المعالجة. ومع ذلك ، فإن شريحة Celeron A التي تمت ترقيتها وجميع إصدارات Celerons الحالية (بما في ذلك إصدارات الأجهزة المحمولة) تتضمن بعض ذاكرة التخزين المؤقت L2 على المعالج نفسه ، وتحسن أداء أجهزة الكمبيوتر بشكل كبير باستخدام الإصدار "A" من Celeron.

You can tell whether or not a system uses the Celeron or Celeron A because all Celerons faster than 300MHz are Celeron As and all processors slower than 300MHz are original Celerons. (The only exception is that all Celerons designed for notebooks have the integrated L2 cache, regardless of speed.) Unfortunately, desktop PC-oriented 300MHz chips were available in both Celeron and Celeron A formats, so the only way to tell 300MHz Celerons apart is to look at the computer's documentation (or use a diagnostic program that lists the processor's type and speed).

يمكنك معرفة ما إذا كان النظام يستخدم Celeron أو Celeron A أم لا لأن جميع Celerons الأسرع من 300MHz هي Celeron As وجميع المعالجات الأبطأ من 300MHz هي Celerons الأصلية. (الاستثناء الوحيد هو أن جميع Celerons المصممة لأجهزة الكمبيوتر المحمولة تحتوي على ذاكرة تخزين مؤقت L2 مدمجة ، بغض النظر عن السرعة.) لسوء الحظ ، كانت رقائق 300 ميغاهرتز الموجهة للكمبيوتر المكتبي متوفرة في كل من تنسيقات Celeron و Celeron A ، وبالتالي فإن الطريقة الوحيدة لتمييز 300 MHz Celerons عن بعضها هو لإلقاء نظرة على وثائق الكمبيوتر (أو استخدام برنامج تشخيص يسرد نوع المعالج وسرعته).

Because most processors incorporate L2 cache into their basic design, you often don't have the option to choose more or less cache in the system you'd like to purchase or put together. (Older processors with standalone L2 cache are the one exception.) Instead, you get the amount of L2 cache a particular model of microprocessor includes. Therefore, when you decide which type of processor to get, make sure you find out how much L2 cache it includes.

نظراً لأن معظم المعالجات تدمج ذاكرة التخزين المؤقت L2 في تصميمها الأساسي ، فغالباً ما لا يكون لديك خيار اختيار ذاكرة تخزين مؤقت أكثر أو أقل في النظام الذي ترغب في شرائه أو تجميعه. (المعالجات الأقدم مع ذاكرة التخزين المؤقت L2 المستقلة هي الاستثناء الوحيد). بدلاً من ذلك ، تحصل على مقدار ذاكرة التخزين المؤقت L2 التي يشتمل عليها طراز معين من المعالجات الدقيقة. لذلك ، عندما تقرر نوع المعالج الذي تريد الحصول عليه ، تأكد من معرفة مقدار ذاكرة التخزين المؤقت L2 التي يتضمنها.

Computers that operate as *servers*, machines that sit at the center point of computer networks, typically need more cache than normal desktop machines because of the type of work they do. Due to this fact, several versions of the Pentium II and Pentium III Xeon, which is designed for servers, include 2MB (or more) of expensive L2 cache, as opposed to most desktop-oriented Pentium IIs and IIIs, which include only 512KB.

عادةً ما تحتاج أجهزة الكمبيوتر التي تعمل كخوادم ، وهي الأجهزة التي تقع في مركز شبكات الكمبيوتر ، إلى ذاكرة تخزين مؤقت أكثر من أجهزة سطح المكتب العادية بسبب نوع العمل الذي تقوم به. نتيجة لهذه الحقيقة ، تشتمل العديد من إصدارات Pentium II و Pentium III Xeon ، المصممة للخوادم ، على 2 ميجابايت (أو أكثر) من ذاكرة التخزين المؤقت L2 باهظة الثمن ، على عكس معظم Pentium IIs و IIIs الموجهة لسطح المكتب ، والتي تتضمن 512 كيلوبايت فقط.