

بسم الله الرحمن الرحيم

Computer Architecture

- 3 -

Characteristics of Memory Systems

التنظيم الهرمي للذاكرة Memory Hierarchy

تمهيد

لذاكرة الحاسوب تنظيم هرمي (Hierarchy)، توجد في أعلاه مسجلات المعالج (Processor Registers)، يلي ذلك مستوى أو أكثر من ذاكرة الكاش (Cache Memory). في حالة وجود أكثر من مستوى واحد من الكاش يطلق عليها تسمية L1 Cache، L2 Cache ... وهكذا. تأتي بعد ذلك الذاكرة الرئيسية (Main memory)، المكونة أساساً، كما نعلم، من ذاكرة DRAM. كل المستويات السابقة توجد داخل جهاز الحاسوب، لذلك يطلق عليها جميعاً تسمية الذاكرة الداخلية (Internal Memory)، و في إمكان المعالج (Processor) الدخول على أي منها مباشرة. و يلي ذلك في التنظيم الذاكرة الخارجية (External Memory)، التي تبدأ في مستواها الأول بقرص صلب (Hard Disk) أو أكثر، ثم عدة مستويات من وسائط تخزين ثانوي أخرى مثل أقراص الـ ZIP و الأقراص الضوئية (Optical Disks) و الأشرطة (Tapes). و ليس في إمكان المعالج الدخول على الذاكرة الخارجية مباشرة بل يجب أن يتم الدخول عبر I/O Module.

عند الانتقال من أعلى إلى أسفل في التنظيم الهرمي للذاكرة نلاحظ انخفاضاً في التكلفة، زيادة في السعة، و بطئاً في السرعة. قد تكون لدينا الرغبة في استخدام أسرع أنواع الذاكرة المتاحة و لكن تواجهنا هنا مشكلة التكلفة العالية لهذا النوع مما يمنعنا من استخدام ساعات كبيرة منه، فنضطر للتضحية بالسرعة في سبيل تقليل التكلفة و نستخدم الذاكرة الأقل سرعة بسعات عالية. أما مشكلة البطء فيتم التحايل عليها عن طريق التأكد من أن الجزء من محتويات الذاكرة الذي توجد حاجة له في الوقت الحالي موضوع في الذاكرة الأسرع.

Characteristics of Memory Systems

لتسهيل فهم نظم الذاكرة يجب تصنيف تلك النظم حسب خصائصها الأساسية. و الجدول التالي يوضح أهم تلك الخصائص:

Location	Performance
Processor	Access Time
Internal (Main)	Cycle Time
External (Secondary)	Transfer Rate
Capacity	Physical Type
Number of Words	Semiconductor
Word Size	Magnetic
Unit of Transfer	Optical
Word	Physical Characteristics
Block	Volatile/Nonvolatile
Access Method	Erasable/Nonerasable
Sequential	Organization
Direct	One-Dimensional
Random	Two-Dimensional
Associative	

1- موقع الذاكرة Location :

يشير إلى موقع الذاكرة و ما إذا كانت داخلية (Internal) أو خارجية (External). و المقصود بالذاكرة الداخلية هنا أساساً هو الذاكرة الرئيسية (Main Memory)، و لكن توجد أنواع أخرى من الذاكرة الداخلية. حيث أن المعالج، مثلاً، يحتاج إلى ذاكرته المحلية الخاصة في صورة مسجلات (Registers) يستخدمها أثناء المعالجة، كما قد تحتاج وحدة التحكم (Control Unit) داخل المعالج نفسه إلى ذاكرتها الداخلية الخاصة أيضاً. كما تعتبر ذاكرة الكاش (Cache Memory) نوعاً من أنواع الذاكرة الداخلية أيضاً. أما الذاكرة الخارجية فالمقصود بها هو أجهزة التخزين الثانوي (Secondary Storage) مثل الأقراص الممغنطة (Magnetic Disks) و الأشرطة الممغنطة (Magnetic Tapes) و الأقراص الضوئية (Optical Disks). و الفرق الأساسي ما بين الذاكرة الداخلية و الذاكرة الخارجية يتمثل في أنه في حين يمكن للمعالج (Processor) الدخول إلى الذاكرة الداخلية بصورة مباشرة و لكنه لا يستطيع الدخول على الذاكرة الخارجية إلا من خلال وحدة إدخال و إخراج (I/O Module).

2- سعة الذاكرة Capacity :

تُقاس سعة الذاكرة بكل من الطول (Length) أي عدد مواقع الذاكرة، و العرض (Width) أي طول الموقع الواحد أو عدد خاناته.

3- وحدة النقل Unit of Transfer :

تعني كمية البيانات المنقولة من أو إلى الذاكرة عند إجراء عملية قراءة أو كتابة واحدة. فقد يتم قراءة موقع واحد (Word) من الذاكرة أو قراءة عدة مواقع متتالية في شكل كتلة (Block) من البيانات. ففي حين أن وحدة النقل عند التعامل مع الذاكرة الرئيسية هي الـ Word فإن وحدة النقل عند التعامل مع الذاكرة الثانوية هي الـ Block.

4- طريقة الدخول Access Method : تعني كيفية الوصول للبيانات المخزنة في الذاكرة. و توجد عدة طرق للدخول هي:

- **Sequential Access**، أي الدخول التتابعي، كما هو الحال في الأشرطة الممغنطة (Magnetic Tapes)، حيث يجب عند البحث عن بيانات معينة أن نبدأ من بداية الشريط و نبحث بالترتيب حتى نصل إلى البيانات المطلوبة.
- **Direct Access**، أي الدخول المباشر، كما هو الحال في الأقراص الممغنطة (Magnetic Disks) و الأقراص الضوئية (Optical Disks)، حيث يتم تحريك رأس القراءة (Head) إلى موقع تقريبي للبيانات المطلوبة ثم يتم بعد ذلك إجراء بحث تنابعي للوصول للبيانات.
- **Random Access**، أي الدخول العشوائي، كما هو الحال في الذاكرة الرئيسية (Main Memory)، حيث يتم الوصول لأي بيانات مباشرة عن طريق العنوان.
- **Associative Access**، كما هو الحال في ذاكرة الكاش (Cache Memory)، حيث يتم البحث في محتويات الذاكرة عن نمط ثنائي (Bit Pattern) معين، و ذلك في وقت واحد لكل مواقع الذاكرة لإيجاد أي تطابقات (Matches). أي أن الوصول للموقع هنا يتم عن طريق محتوياته و ليس عن طريق عنوانه.

لاحظ أن طرق الدخول المذكورة أعلاه مرتبة تصاعدياً من حيث السرعة. فالدخول التتابعي هو أبطأها و الدخول الـ Associative هو أسرعها.

5- الأداء Performance :

يستخدم للإشارة إلى سرعة الذاكرة التي يتم قياسها باستخدام:

- Access Time (Latency) و المقصود به الزمن اللازم للقيام بعملية قراءة أو كتابة واحدة. ففي ال RAM هو الزمن ما بين وضع العنوان على أطراف العنوان و ظهور البيانات على أطراف البيانات، و في القرص الصلب هو الزمن اللازم لوضع رأس القراءة (Head) في المكان المطلوب.
- Cycle Time و هو الزمن ما بين عمليتي دخول متتابعتين. و هو يساوي ال Access Time زائد أي زمن إضافي لازم قبل البدء بعملية دخول أخرى. و هذا الزمن الإضافي قد يكون مطلوباً مثلاً لإتباع عملية القراءة بعملية إعادة كتابة للمحتويات في ال DRAM لأن عملية القراءة مدمرة لمحتويات الموقع.
- Transfer Rate أي معدل أو سرعة نقل البيانات من الذاكرة و هو يساوي مقلوب ال Cycle Time.

التنظيم الهرمي للذاكرة Memory Hierarchy

عند تصميم نظام الذاكرة يجب الإجابة على التساؤلات التالية:

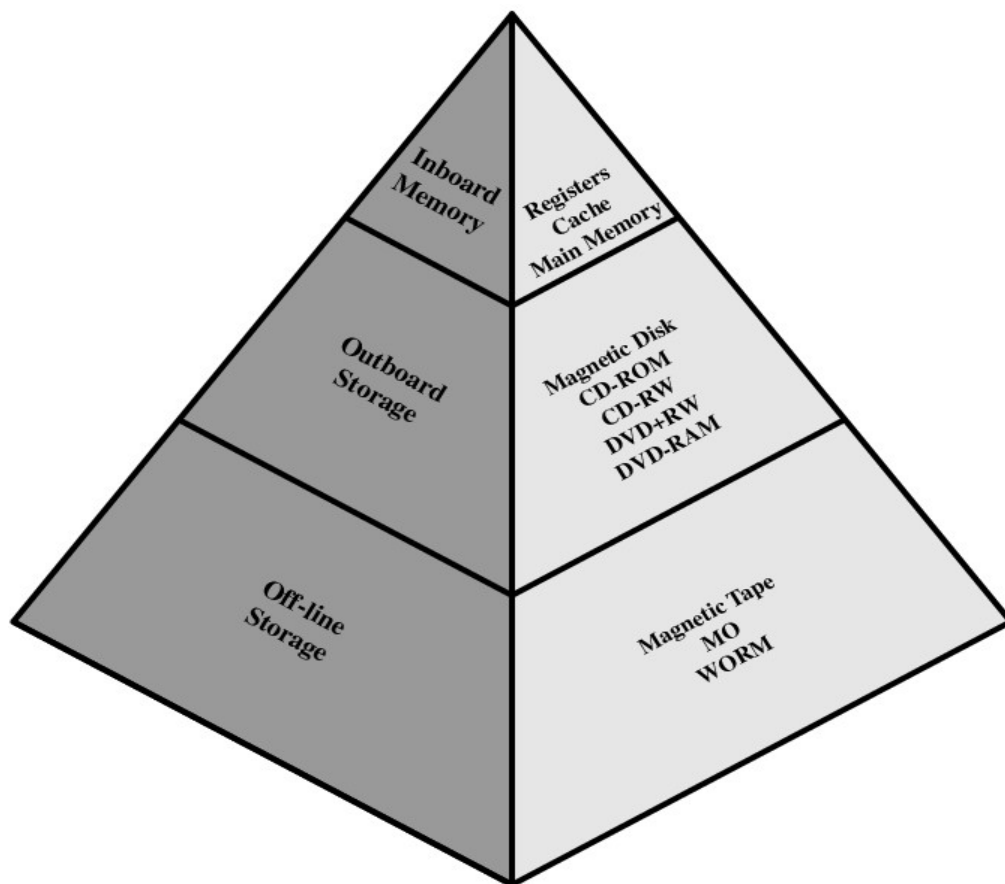
- How much?
- How fast?
- How expensive?

ليس من السهل الإجابة على التساؤل الأول، حيث أنه كلما توفرت ساعات أكبر من الذاكرة كلما تم تطوير تطبيقات جديدة للإستفادة من هذه السعة المتاحة. أما التساؤل الثاني فإجابته سهلة نسبياً حيث أن سرعة الذاكرة يجب أن تتناسب مع سرعة المعالج حتى لا يضطر المعالج للإنتظار أثناء المعالجة مما يقلل من كفاءته. و في ما يتعلق بالتساؤل الثالث فيجب أن تتناسب تكلفة الذاكرة مع التكلفة الكلية لجهاز الحاسوب.

لا مناص عند تصميم نظام الذاكرة من المفاضلة ما بين سعة الذاكرة و سرعتها و تكلفتها. حيث من الملاحظ أنه:

- كلما زادت سرعة الذاكرة كلما زادت تكلفة ال bit الواحد فيها
- كلما زادت سعة الذاكرة كلما قلت تكلفة ال bit الواحد فيها
- كلما زادت سعة الذاكرة كلما قلت سرعتها

و عليه فإن مصمم الذاكرة يرغب في استخدام التقنية التي تتيح له سعة ذاكرة كبيرة، و ذلك نظراً إلى أن السعة الكبيرة في الذاكرة مطلوبة أساساً و لأن مثل هذه التقنية ذات تكلفة أقل لل bit الواحد. و لكن لتحقيق متطلبات الأداء (Performance) يضطر المصمم لاستخدام التقنيات التي تتيح له السرعة العالية رغم سعاتها المنخفضة و تكلفتها الأعلى. و حل هذه الإشكالية هو في عدم استخدام تقنية واحدة و إنما عدة تقنيات في شكل تنظيم هرمي (Hierarchy)، كما هو موضح بالشكل التالي:



نلاحظ عند الإنتقال من قمة الهرم إلى قاعدته:

- انخفاض التكلفة للـ bit الواحد
- زيادة السعة
- انخفاض السرعة
- انخفاض عدد مرات الدخول

و عليه يتم دعم الذاكرة السريعة قليلة السعة و مرتفعة التكلفة بالذاكرة البطيئة كبيرة السعة و قليلة التكلفة. و سر نجاح هذه الإستراتيجية هو تقليل عدد مرات الدخول للذاكرة البطيئة قدر الإمكان، و ذلك بنقل البيانات في شكل كتل (Blocks) من الذاكرة البطيئة و وضعها في الذاكرة السريعة بحيث يجد المعالج كل ما يحتاج إليه في الذاكرة السريعة و لا يحتاج في أغلب الأحيان الرجوع للذاكرة البطيئة. و ما يجعل هذا ممكناً هو مبدأ يُعرف بـ Locality of Reference أي تَمَوُّضُ المراجع. و المقصود بالمراجع هنا هو مواقع الذاكرة التي يحتاج المعالج للرجوع إليها أثناء تنفيذ البرامج. حيث لوحظ أن تلك المراجع، سواء كانت تعليمات (Instructions) أو بيانات (Data) تميل للتكتل و التَمَوُّضُ. حيث عادة ما تحتوي البرامج على حلقات (Loops) و برامج فرعية (Subroutines)، فعند الدخول في هذه الحلقات أو البرامج الفرعية يتم الرجوع لمجموعة محددة من التعليمات بصورة متكررة. كما أنه عند التعامل مع الجداول (Tables) أو المصفوفات (Arrays) يتم الرجوع بصورة متكررة لمجموعة محددة من البيانات. لاحظ أنه بعد مرور فترات زمنية طويلة تتغير تلك التكتلات، و لكن

خلال فترة زمنية قصيرة فإن المعالج يتعامل أساساً مع كتلة ثابتة تقريباً من المراجع.

الآخيرة