طرق تنظيم الملف في الحاسبات الألكترونية

ترجمة : غسان عدنان البراك
المركز القومي للحاسبات الالكترونية

الملفـــات (Files)

الملف هو مجموعة منظمة من المعلومات المترابطة الموضوعة للاستخدام في تحقيق اهـداف معينة . وتلك المعلومات منظمة في قيود (Records) مترابطة تتيح توفير معلومات محددة حول مجال معين . وتخزن تلـك المعلومات اعتياديا على وسائط الخزن المعروفة كالبطاقات والاشـــرطة المثقبة او الاشرطة والاقراص المغنطة وغيرها .

1— اصناف الملفـات ووظائفهـا :

ـ الملف الاســـتاذ :

ويحتوي على الحال الحاضر لمجموعة قيود وفقراتها . ويكـــون اعتياديا ذاحجم ثابت تقريبا حيث توازن القيود المضافة القيود المحذوفة . ولكن ذلك لايعني عدم وجود تغييرات موسمية او دورية حسب نـوع التطبيق . ويعتبر الملف الاستاذ المصدر الاساسي للمعلومات ســـواء كانت للاستعمال في البرامج (داخليا) او لتلبية احتياجات الادارة (خارجيا).

ـ ملف الترحيـــل (Transaction File)

ويحوي على قيود الترحيل المستخدمة للاستفسار عن معلومات او لتحديث تلك المعلومـات .

ـ الملف التاريخـي (History File)

وهو اما يكون ملف استاذ قديم او ملف الحركات المتراكمة عبر فترة من الزمن . ويستخدم لتوفير معلومات احصائية او لاعادة ملف استاذ جديد عنـد الضرورة .

- **ملف الخلاصة** (Summary File)

وهو صورة مختصرة لملف اخر ، حيث يتم استخلاص وتجميع معلومات من قيود عديدة موجودة في الملف الاصلي ووضعها في قيد واحد من قيود ملف الخلاصة .

- **الملف الملحق** (Trailer File)

ويحتوي على قيود تفصيلية لقيود رئيسية في ملف رئيسي . حيث يوفر معلومات اضافية مفصلة لمعلومات القيود الرئيسية .

2-1 مكونات الملف

2-2-1 الحقول والحقول الفرعية :

يعتبر الحقل اصغر عنصر في الملف . اما الحقل الفرعي فهو جزء كامل المعنى من الحقل لتيسير تعريف البيانات ومعالجتها .

اصناف الحقول ووظائفها :

- **حقل السيطرة** : لتعريف ومعاملة قيد او قيد فرعي ، ويقوم حقل السيطرة من القيد بمقام المفتاح الذي يعطي صفة مميزة ضمن الملف لضمان تمييزه عن غيره من قيود الملف ، بينما يقوم حقل السيطرة الترميزية Coding Control بتصنيف القيد او القيد الفرعي لاغراض المعالجة

- **الحقول الدلالية** (Indicative Fields)

وهي بصورة عامة حقول تحتوي على بيانات متفرقة لاتتصف بسمة التغيير . وبصورة اكثر تحديدا نذكر الحقول الاحصائية التي توفر معلومات اضافية لغرض التحليل الاحصائي للملف ، والحقول الثابتة التي تحتوي على بيانات ثابتة لتلافي عملية اعادة توليد البيانات عند المعالجة ، وحقول المرجع التي توفر التمييز بين الحركة Transaction ووثيقتها المصدرية .

ـ **الحقول الكمية** : وتستعمل اساسا للعمليات الحسابية .

3 – 1 مواصفـــات الحقـــول :

ـ **الطول** : ان كان ثابتا او متغيرا لاحتواء اقصى حد من البيانـات المحتملـــــة .

ـ **النـــوع** : ان كان رقميا او ابجديا او مختلطا .

ـ **المغـزى او الاهميـة والمعنى** Significance

في الارقام فان الارقام المعنوية هي تلك الارقام الضرورية التي تجعل للعدد معنى . وقد يشمل ذلك عدد محدد من المراتب العشرية . وفي هذا المجال تعتبر الفراغات في الرقم شيء غير مقبول . اما في حالة الابجدية او المختلطة فالمعنى هنا كل الرموز من اقصى اليسار الى اقصى اليميـن ولاتقبل الفراغات في هذه الحالـة .

ـ **جهة ابتداء البيانات (يمينا ام يسارا)** : واعتياديا فان الحقـــول الابجدية او المختلطة تكون بياناتها بادئة من اليسار ومتضمنة فراغات ، اما الحقول الرقمية فتبدأ اعتياديا من اليمين تليها اصفار عند نهاية الرقم .

وهناك صفات خاصة للحقول الرقمية تشتمل على :

ـ **الاشارة** : (+) او (–)

ـ **المراتب العشرية** : وتكـــون مفترضة اعتياديا .

ـ **شكل التسجيل** : لاغراض الاحتساب او لاختصار الخزن ، فـــأن البيانات قد تكون بالنظام الثنائي، او السادس عشري (Hexadecimal) او مكبوسة (Packed) ، وفي هذه الحالة تصبح الترجمة مـــن والى النظام العشري مسألة ضـرورية .

2 – القيـود (Records)

القيد هو مجموعة من الحقول منظمة وفق صيغة محدودة ومرتبطة بمعرف او مفتاح مشـــترك .

2–1 اصناف القيود ووظائفها :

– **قيود البيانات** : وتمثل الحجم الرئيسي في الملف وتحتوي على معلومات محددة تخص تطبيق معين . ويحتوي كل قيد اعتياديا على حقول خاصة بالسيطرة على القيد والترميز بالاضافة الى حقول المعلومات . .

– **نقطة الفحص** (Check Point)

ويتم تكوينها على فترات لاستعادة محتويات مخزن البيانات لغرض تيسير اعادة المعالجة في حالات توقفها لسبب او اكثر ، من نقطة التوقف .

– **قيود الدلالة** (Label Record)

وتستعمل لاغراض تعريف الملف والتدقيق . وهناك نوعين من تلك القيود في الملف ، النوع الاول ويسمى قيد البداية او العنوان (Header Labels) وقد يحتوي على اسم الملف ورقمه ودورة وضعه ونفاذه وغيرها من المعلومات . ويتم اعتياديا تدقيق بيانات ذلك القيد قبل البدء بمعالجة الملف . اما النوع الثاني ويدعى بالمقطور او الملحق او النهاية (Trailer Label) فهو يدل على نهاية الملف (EOF) او البكرة (EOR) او جزء من الملف ، وقد يحتوي على مجاميع معينة لاغراض التدقيق .

– **قيود الفيض** (0ver-Flow Records)

وتحتوي على بيانات اضافية ذات علاقة بقيد رئيسي لم يسمح حجم ذلك القيد بخزنها فيه وتكون تلك المعلومات عادة موسمية او تحدث في مناسبات فقط . ويكون مفتاح القيد هذا هو نفس مفتاح القيد الرئيسي المتعلق به ويحوي على اشارة لذلك القيد .

2–2 مواصفات القيود

– **الطول** : يكون طول القيد هو مجموع اطوال حقوله . وقد يكون هذا الطول ثابت او متغير . ولاسباب تتعلق بتحديدات الاجهزة فان الطول يختلف بين الخزن الداخلي والخارجي .

– **الشكل** : هناك ثلاثة اقسام اساسية قد تحتويها اي من القيود حسب نوعها (كما سيأتي لاحقا) وهي :

أ ـ القسم الثابت : وهو ثابت الطول ويحتوي كل قيد على واحدة منه اعتياديا ، ويتكون من حقول القيد والسيطرة اضافة الى حقول دالة وكمية . وقد يأخذ كامل القيد او الجزء الاول من قيد متعدد الاجزاء.

ب ـ القسم المتغير : وهو مجموعة متغيرة العدد من المقاطع وتتكون تلك المقاطع اما من حقل مستقل او مجموعة من الحقول المتعددة على بعضها . وفي الوقت الذي يمكن فيه تواجد الحقل المستقل بمفرده . فان الحقول المترابطة يجب ان تتواجد معا . اضافة الى ان كل مقطع يكون ثابت الطول ولايتواجد الا بتواجد البيانات .

جـ ـ قسم السيطرة : ويمكن استخدامه في حالة وجود القسم المتغير. ويوفر هذا القسم معلومات عن وجود المقاطع من عدمه في القسم المتغير مع اطوالها وبالتالي فهو يستعمل لتحديد موقع البيانات ومعاملتها هناك .

وعلى هذا الاساس تقسم القيود الى نوعين :

ـ **القيود ثابتة الطول** : وهي القيود التي تحتوي على القسم الثابت فقط المنوه عنه سابقا وفيها يكون طول كل حقل وموقعه النسبي ثابت .

ـ **القيود متغيرة الطول** : وهي تحتوي على القسم الثابت كالنوع السابق اضافة الى القسم المتغير حيث يتم تحديد الطول والموقع النسبي لكل حقل بالاستخدام المبرمج لقسم السيطرة او بواسطة شفرات خاصة يتم تضمينها في كل قيد .

كما يوجد هناك اسلوب لغرض المعاملة المثلى للقيود يسمى باسلوب التجميع (Blocking) ونعني به تجميع قيدين او اكثر ومعاملتهم كوحدة لاغراض الادخال والاخراج . اما عدد القيود في المجمع الواحد فيدعى (معامل التجميع) ويدعى (المجمع) وفي بعض الاحيان (بالقيد الفيزياوي) وتسمى القيود الفعلية المكونة له (بالقيود المنطقية) اي ان كل قيد فيزياوي يتكون من واحد او اكثر من القيود المنطقية .

3 – تنظيم الملف

ان تلك العملية تعنى بالعلاقة بين حقول السيطرة في القيود وبين الموقع الفيزيائي لتلك القيود في الوسط المخزونة فيه . (Storage Medium)

3–1 اصناف التنظيم :

– التنظيم التتابعي او السياقي (Sequential)

وهو ذلك النوع من التنظيم الذي تتسلسل فيه القيود المتجاورة في الملف وفق تسلسل رقمي او ابجدي – رقمي استنادا الى حقول سيطرة محدودة لغرض تثبيت ذلك التسلسل .

– التنظيم العشوائي (Random)

وهو التنظيم الذي يتم وفقه خزن قيود الملف بصورة عشوائية بغض النظر عن تسلسل حقول السيطرة فيها .

3–2 اصناف المعالجة :

عند استخدام مصطلحات (تتابعي) و (عشوائي) في مجال المعالجة فالمعنى بذلك هو وضع قيود الحركات المدخلة او المهيئة للادخال او طريقة الرجوع الى القيود في الملف الرئيسي . وعلى هذا الاساس ، ومن الناحية النظرية على الاقل ، يمكن تطبيق المعالجة التتابعية والعشوائية على اي من الملفات المنظمة تتابعيا او عشوائيا . ومع ذلك فان طبيعة واسطة الخزن تؤثر على طريقة التنظيم والمعالجة وبذلك فليست جميع الطرق فعالة او ممكنة التطبيق مع واسطة خزن معينة .

3–3 تنظيم ومعالجة الاشرطة والبطاقات المثقبة :

بسبب طبيعة الاشرطة المغنطة والبطاقات المثقبة ، يصبح لزاما معالجة القيود بتتابعها ، وعليه فان الملفات الرئيسية على الاشرطة والبطاقات اضافة الى ملفات الحركات ذات العلاقة ، تكون اعتياديا ذات تنظيم تتابعي . ومع ذلك فهي تحتاج في معظم الاحيان الى عملية فرز (Sorting) قبل المعالجة . واستنتاجا من ذلك فان الملفات الرئيسية المنظمة عشوائيا لاتوضع مطلقا على الاشرطة المغنطة لاسباب بديهية . ومع ذلك فمن الشائع ان يكون لدينا (ملفات حركات) منظمة

عشوائيا على مثل تلك الوسائط في حالة وجود الملف الرئيسي مخزنا على وسائط خزن مباشرة الوصول (DASD) كالاقراص المغنطة والاسطوانات وغيرها .

3 -4 تنظيم ومعالجة اجهزة الخزن المباشرة

3 - 4 - 1 **التنظيم التتابعي** : تخزن القيود وفق تسلسل مفاتيها ، ولايكون المفتاح اعتياديا هو نفسه عنوان الموقع المخزون فيه القيد . وفي هذا النطاق يمثل الحذف والاضافات مشكلة يتم تلافيها بايجاد مساحات للفيض (Overflow) لغرض وضع القيود التي لايمكن حشرها في الملف لعدم توفر مساحة لها . وبازدياد مساحات الفيض تلك يزداد زمن المعالجة ، فاذا وصل ذلك الى الحد الذي يؤدي الى انقاص كفاءة الاداء توجب حينئذ اعادة تنظيم الملف بأكمله بالشكل الذي يعالج تلك المساحات ويقلل من زمن المعالجة .

ان تنظيم الملف بالشكل المبين اعلاه يعني معالجته تتابعيا ولكننا في هذه الحالة سوف لانستعمل مزية الوصول العشوائي التي تتمتع بها اجهزة الخزن المباشر والتي يمكن بواسطتها الوصول الى اي قيد بصورة سريعة نسبيا بغض النظر عن موقعه . وعليه ومن اجل تقليل واختصار زمن البحث (Search Time) بالنسبة للملف التتابعي ، يتم وضع فهرست لذلك الملف وبالتالي فاننا نصل الى مايدعى بطريقة التتابع المفهرست (Indexed Sequential Methos) حيث يرتبط الفهرست الرئيسي بفهرست الاسطوانة الذي بدوره يدلنا على فهرست المسار او الطوق ويتم البحث في ذلك المسار لايجاد القيد المطلوب . وعلى هذا الاساس فان التنظيم التتابعي المفهرس يسمح لنا بكلا المعالجات التتابعية العشوائية بشكل كفوء .

ـ **التنظيم العشوائي** : ويتم فيه خزن القيود في مواقع يتم تحديد عناوينها بواسطة احدى دوال مفتاح القيد . وعلى هذا الاساس لانحتاج الى فهرست .

4 - ملفات النظم المباشره (ON - LINE SYST

لقد استعرضنا فيما سبق حاله بسيطه تقتصي الوصول الى او استرجاع قيد معرف بمفتاح مفرد محدد ، وهذا مايدعى (بالاسترجاع مفرد المفتاح) ولكن هناك حالات عديدة من الضروري معها استرجاع قيود

تحتوي على اكثر من مفتاح وان يكون ذلك الاسترجاع سريعا ، فان كانت الحالة كذلك فلابد لنا من استعراض نوعين اخرين من طرق تنظيم الملفات المفيـدة في هذا المجـال :

– الملف المعكـوس (Inverted File)

وهنا يتم وضع قائمة بكل المفاتيح التي يحتويها الملف ومع كل مفتاح يخزن عنوان كل قيد يحتوي ذلك المفتاح . وهكذا فان الوصول يتـم بواسطة القائمة الى المفتاح المطلوب ثم الى العناوين المحددة لذلك المفتاح .

– الملف متعدد القوائم (Multi-List File)

وهنا يتم وضع قائمة بالمفاتيح كما جاء في النوع السابق مع فارق يتلخص بأن لكل مفتاح نخزن عنوان قيد واحد فقط من القيود التـي تحتوي ذلك المفتاح . ويتضمن ذلك القيـد المحدد عنوان قيد اخـر يحتوي نفس المفتاح وهكذا تتشكل لدينـا قائمة مترابطة من القيـود .