



# Sun™ One Grid Engine，企業版

## 5.3 管理和使用者指南

---

Sun Microsystems, Inc.  
4150 Network Circle  
Santa Clara, CA 95054 U.S.A.  
650-960-1300

文件編號：816-7474-10  
2002 年 9 月，修訂版 A

請將您對本文件的意見寄到：[docfeedback@sun.com](mailto:docfeedback@sun.com)

Copyright 2002 Sun Microsystems, Inc., 4150 Network Circle, Santa Clara, CA 95054 U.S.A. 版權所有。

本產品或文件按照限制其使用、複製、分發和反編譯的授權許可進行分發。未經 Sun 及其授權許可頒發機構的書面授權，不得以任何方式、任何形式複製本產品或本文件的任何部分。協力廠商軟體，包括字型技術，由 Sun 供應商提供許可和版權。

本產品的某些部分從 Berkeley BSD 系統衍生而來，經 University of California 許可授權。UNIX 是在美國和其他國家註冊的商標，經 X/Open Company, Ltd. 獨家許可授權。

Sun、Sun Microsystems、Sun 標誌、AnswerBook2、docs.sun.com 和 Solaris 是 Sun Microsystems, Inc. 在美國和其他國家的商標、註冊商標或服務標記。所有的 SPARC 商標都按授權許可使用，是 SPARC International, Inc. 在美國和其他國家的商標或註冊商標。具有 SPARC 商標的產品都基於 Sun Microsystems, Inc. 開發的架構。「能源之星」標誌是 EPA 的註冊商標。OPEN LOOK 和 Sun™ 圖形使用者介面是 Sun Microsystems, Inc. 為其使用者和授權許可持有人開發的。Sun 承認 Xerox 在為電腦行業研究和開發可視或圖形使用者介面方面所作出的先行努力。Sun 以非獨佔方式從 Xerox 獲得 Xerox 圖形使用者介面的授權許可，該授權許可涵蓋實施 OPEN LOOK GUI 且遵守 Sun 的書面許可協議的授權許可持有人。

本資料按「現有形式」提供，不承擔明確或隱含的條件、陳述和保證，包括對特定目的或非侵害性的商業活動和適用性的任何隱含保證，除非這種不承擔責任的聲明是不合法的。



請回收



Adobe PostScript

# 目錄

---

|               |       |
|---------------|-------|
| 前言            | xvii  |
| 本書編排架構        | xvii  |
| 使用 UNIX 指令    | xviii |
| 排印慣例          | xviii |
| Shell 提示符號    | xix   |
| 相關說明文件        | xix   |
| 線上存取 Sun 出版文件 | xix   |
| Sun 歡迎您的指教    | xx    |

## 部分 I. 背景和定義

1. Sun Grid Engine，企業版 5.3 簡介 1
  - 什麼是格線運算？ 1
  - 藉由管理資源和策略來管理工作量 3
  - 系統作業的方式 3
    - 使資源與需求相符 4
    - 作業和佇列：Sun Grid Engine 世界 4
    - 使用策略的多樣性 5
      - 使用票卷範例 (Ticket Paradigm) 的策略管理 5
  - Sun Grid Engine，企業版 5.3 組件 6

|                                   |    |
|-----------------------------------|----|
| 主機                                | 6  |
| 主要主機                              | 7  |
| 執行主機                              | 7  |
| 管理主機                              | 7  |
| 提交主機                              | 7  |
| 常駐程式                              | 7  |
| sge_qmaster – 主要常駐程式              | 7  |
| sge_schedd – 排程器常駐程式              | 8  |
| sge_execd – 執行常駐程式                | 8  |
| sge_commd – 通訊常駐程式                | 8  |
| 佇列                                | 8  |
| 客戶端指令                             | 8  |
| QMON, Sun Grid Engine, 企業版圖形使用者介面 | 10 |
| 自訂 QMON                           | 11 |
| Sun Grid Engine 專有名詞的詞彙           | 12 |

## 部分 II. 入門

### 2. 安裝 17

|                         |    |
|-------------------------|----|
| 基礎安裝概觀                  | 17 |
| 階段 1 – 計劃               | 18 |
| 階段 2 – 安裝軟體             | 18 |
| 階段 3 – 驗證安裝             | 18 |
| 計劃安裝                    | 19 |
| 先決條件作業                  | 19 |
| 安裝目錄 < <i>sge</i> 根目錄 > | 19 |
| Root 目錄下的多工緩衝處理目錄       | 19 |
| 目錄組織                    | 20 |

|  |    |
|--|----|
| 磁碟空間需求                                   | 21 |
| 安裝帳號                                     | 21 |
| 檔案存取權限                                   | 21 |
| 網路服務                                     | 21 |
| 主要主機                                     | 22 |
| 備用主要主機                                   | 22 |
| 執行主機                                     | 22 |
| 管理主機                                     | 23 |
| 提交主機                                     | 23 |
| 儲存格                                      | 23 |
| 使用者名稱                                    | 23 |
| 佇列                                       | 23 |
| ▼ 如何計劃安裝                                 | 24 |
| ▼ 如何讀取分發媒體                               | 25 |
| pkgadd 方法                                | 25 |
| tar 方法                                   | 26 |
| 執行基本安裝                                   | 26 |
| ▼ 如何安裝主要主機                               | 27 |
| ▼ 如何安裝執行主機                               | 27 |
| ▼ 如何安裝管理和提交主機                            | 28 |
| 用加強的安全性安裝                                | 29 |
| 所需的額外設定                                  | 29 |
| ▼ 如何安裝和設定 CSP 安全化系統 (CSP-Secured System) | 30 |
| ▼ 如何產生使用者證明文件和私密金鑰                       | 39 |
| ▼ 如何檢查證明文件                               | 41 |
| 顯示證明文件                                   | 41 |
| 檢查發出者                                    | 41 |

|             |    |
|-------------|----|
| 檢查主題        | 41 |
| 顯示證明文件的電子郵件 | 41 |
| 顯示有效性       | 42 |
| 顯示指紋        | 42 |
| 驗證安裝        | 42 |
| ▼ 如何驗證安裝    | 42 |

### 部分 III. 使用 Sun Grid Engine 企業版 5.3 軟體

|                             |    |
|-----------------------------|----|
| 3. 導覽 Sun Grid Engine，企業版   | 49 |
| Sun Grid Engine，企業版使用者種類和操作 | 49 |
| 佇列和佇列屬性                     | 50 |
| QMON 瀏覽器                    | 50 |
| ▼ 如何啓動 QMON 瀏覽器             | 51 |
| 「佇列控制」QMON 對話方塊             | 51 |
| ▼ 如何顯示佇列清單                  | 52 |
| ▼ 如何顯示佇列屬性                  | 52 |
| 使用 QMON 瀏覽器                 | 52 |
| 從指令行                        | 54 |
| 解譯佇列屬性資訊                    | 54 |
| 主機功能                        | 55 |
| ▼ 如何找出主要主機名稱                | 55 |
| ▼ 如何顯示執行主機清單                | 55 |
| ▼ 如何顯示管理主機清單                | 55 |
| ▼ 如何顯示提交主機清單                | 56 |
| 可要求的屬性                      | 56 |
| ▼ 如何顯示可要求屬性清單               | 57 |
| 使用者存取權限                     | 59 |

管理員、操作員和所有人 61

#### 4. 提交作業 63

執行簡單作業 63

▼ 如何從指令行執行簡單作業 64

▼ 如何提交作業，從圖形使用者介面，QMON 65

提交批次作業 69

關於 Shell Script 69

Script 檔的範例 70

使用 QMON 提交延伸和進階作業 71

延伸範例 71

進階範例 74

資源需求定義 79

Sun Grid Engine，企業版如何配置資源 80

延伸至一般的 Shell Script 81

如何選擇指令解譯器 81

輸出重新導向 82

作用中的 Sun Grid Engine，企業版註釋 82

環境變數 83

▼ 如何從指令行提交作業 85

預設要求 86

陣列作業 87

▼ 如何從指令行提交陣列作業 87

▼ 如何使用 QMON 提交陣列作業 87

提交互動作業 88

使用 QMON 提交互動式作業 89

▼ 如何使用 QMON 提交互動作業 89

使用 qsh 提交互動式作業 91

- ▼ 如何使用 qsh 提交互動作業 92
- 使用 qlogin 提交互動式作業 92
- ▼ 如何使用 qlogin 提交互動作業 92
- 透明遠端執行 92
  - 使用 qrsh 遠端執行 93
  - ▼ 如何 qrsh 使用呼叫透明遠端執行 93
  - 使用 qtcsh 的透明作業分配 94
    - qtcsh 用法 94
  - 帶有 qmake 的平行檔案製作過程 (Parallel Makefile Processing) 96
    - qmake 用法 97
  - Sun Grid Engine, Enterprise Edition 作業的排程方式 98
    - 作業優先權 98
    - 票卷 99
    - 佇列選擇 99
- 5. 檢查點、監視和控制作業 101
  - 關於檢查點作業 101
    - 使用者階層檢查點 102
    - Kernel 階層檢查點 102
    - 檢查點作業的遷移 102
    - 組成檢查點作業 Script 102
    - ▼ 如何從指令行提交、監視或刪除檢查點作業 103
    - ▼ 如何使用 QMON 提交檢查點作業 104
    - 檔案系統需求 105
  - 監視和控制 Sun Grid Engine，企業版  
作業 106
    - ▼ 如何使用 QMON 監視和控制作業 106
    - QMON 物件瀏覽器的額外資訊 115

- ▼ 如何使用 `qstat` 監視作業 115
- ▼ 如何使用電子郵件監視作業 118
- 從指令行控制 Sun Grid Engine, Enterprise Edition 作業 119
  - ▼ 如何從指令行控制作業 119

作業相依 120

控制佇列 120

- ▼ 如何用 `QMON` 控制佇列 120
- ▼ 如何用 `qmod` 控制佇列 124

自訂 `QMON` 125

## 部分 IV. 管理

### 6. 主機和叢集配置 129

關於主要和備用主要配置 130

關於常駐程式和主機 131

關於配置主機 131

無效的主機名稱 132

- ▼ 如何以 `QMON` 配置管理主機 132
- ▼ 如何刪除管理主機 134
- ▼ 如何新增管理主機 134
- ▼ 如何從指令行配置管理主機 134
- ▼ 如何以 `QMON` 配置提交主機 134
- ▼ 如何刪除提交主機 136
- ▼ 如何新增提交主機 136
- ▼ 如何從指令行配置提交主機 136
- ▼ 如何以 `QMON` 配置執行主機 136
- ▼ 如何刪除執行主機 138
- ▼ 如何關閉執行主機常駐程式 138

- ▼ 如何新增或修改執行主機 138
- ▼ 如何從指令行配置執行主機 142
- ▼ 如何使用 qhost 監視執行主機 143
- ▼ 如何從指令行刪除常駐程式 144
- ▼ 如何從指令行重新啟動常駐程式 145

#### 基礎叢集配置 145

- ▼ 如何從指令行顯示基礎叢集配置 145
- ▼ 如何從指令行修改基礎叢集配置 146
- ▼ 如何使用 QMON 顯示叢集配置 147
- ▼ 如何使用 QMON 刪除叢集配置 147
- ▼ 如何使用 QMON 顯示全域叢集配置 147
- ▼ 如何使用 QMON 修改全域和主機配置 148

## 7. 配置佇列與佇列行事曆 151

### 關於佇列配置 151

- ▼ 如何以 QMON 配置佇列 152
- ▼ 如何配置一般參數 153
- ▼ 如何配置執行法參數 154
- ▼ 如何配置檢查點參數 155
- ▼ 如何配置載入 / 暫停臨界值 156
- ▼ 如何配置限制 157
- ▼ 如何配置使用者聚合體 159
- ▼ 如何配置次級佇列 160
- ▼ 如何配置使用者存取 161
- ▼ 如何配置專案存取 162
- ▼ 如何配置所有人參數 163
- ▼ 如何從指令行配置佇列 164

### 關於佇列行事曆 165

- ▼ 如何以 QMON 配置佇列行事曆 165
- ▼ 如何從指令行配置行事曆 167
- 8. 管理使用者存取與策略 169
  - 關於設定使用者 170
  - 關於使用者存取 171
    - ▼ 如何以 QMON 配置帳號 171
    - ▼ 如何以 QMON 配置管理員 (Manager) 帳號 172
    - ▼ 如何從指令行配置管理員帳號 173
      - 可用的開關 173
    - ▼ 如何以 QMON 配置操作員帳號 173
    - ▼ 如何從指令行配置操作員帳號 174
      - 可用的開關 174
    - 關於佇列所有人帳號 175
    - 關於使用者存取權限 175
      - ▼ 如何以 QMON 配置使用者存取清單 176
      - ▼ 如何從指令行配置使用者存取清單 178
        - 可用的選項 178
    - 關於使用「使用者集」來定義「專案」與「部門」 178
    - 關於使用者物件配置 179
      - ▼ 如何以 QMON 配置使用者物件 179
      - ▼ 如何指定預設專案 180
      - ▼ 如何從指令行配置使用者物件 181
        - 可用的選項 181
    - 關於專案 182
      - ▼ 如何以 QMON 定義專案 182
      - ▼ 如何從指令行定義專案 185
        - 可用的選項 185

|                            |     |
|----------------------------|-----|
| 關於排程                       | 185 |
| 排程策略                       | 186 |
| 動態資源管理                     | 186 |
| 佇列分類                       | 187 |
| 作業分類                       | 188 |
| 在排程器間隔時間內發生的事              | 188 |
| 排程器監視                      | 188 |
| 排程器配置                      | 189 |
| 預設排程                       | 189 |
| 排程選擇                       | 189 |
| ▼ 如何以 QMON 配置變更排程器配置       | 191 |
| ▼ 如何以 QMON 管理策略 / 票卷進階資源管理 | 194 |
| 編輯票卷 (Edit Tickets) 區      | 195 |
| 「策略」按鈕區                    | 195 |
| 關於共享式的策略                   | 195 |
| ▼ 如何從 QMON 編輯共享樹策略         | 198 |
| 節點屬性顯示區                    | 198 |
| 共享樹策略參數                    | 201 |
| 關於特殊使用者 default            | 202 |
| ▼ 如何從指令行配置共享策略             | 203 |
| 關於功能 (Functional) 策略       | 204 |
| 功能共享                       | 204 |
| share_functional_shares 參數 | 204 |
| ▼ 如何從 QMON 配置功能共享策略        | 205 |
| ▼ 如何從指令行配置功能共享策略           | 208 |
| 關於截止日期策略                   | 208 |
| 截止日期票卷                     | 209 |

|                           |            |
|---------------------------|------------|
| share_deadline_tickets 參數 | 209        |
| 關於置換策略                    | 211        |
| share_override_tickets 參數 | 211        |
| ▼ 如何配置置換策略                | 212        |
| ▼ 如何從指令行配置置換策略            | 213        |
| 關於策略階層 (Policy Hierarchy) | 214        |
| 關於路徑別名                    | 215        |
| 檔案格式                      | 216        |
| 如何解讀路徑更名檔                 | 216        |
| 路徑別名檔之範例                  | 217        |
| 關於配置預設要求                  | 217        |
| 預設要求檔案之格式                 | 218        |
| 預設要求檔案範例                  | 218        |
| 關於收集計算與利用數據               | 219        |
| 關於檢查點支援                   | 219        |
| 檢查點環境                     | 220        |
| ▼ 如何以 QMON 配置檢查點環境        | 220        |
| 檢視配置的檢查點環境                | 221        |
| 刪除配置的檢查點環境                | 221        |
| 修改配置的檢查點環境                | 222        |
| 新增檢查點環境                   | 224        |
| ▼ 如何從指令行配置檢查點環境           | 224        |
| qconf 檢查點選項               | 224        |
| <b>9. 聚合體概念</b>           | <b>225</b> |
| 關於聚合體                     | 225        |
| ▼ 如何新增或修改聚合體配置            | 226        |
| 聚合體種類                     | 227        |

- 佇列聚合體 228
  - 主機聚合體 228
  - 全域聚合體 230
  - 使用者定義聚合體 231
- 可耗損資源 234
  - ▼ 如何設定可耗損資源 234
    - 設定可耗損資源範例 236
- 配置聚合體 244
  - ▼ 如何從指令行修改聚合體配置 244
    - qconf 指令範例 245
- 載入參數 245
  - 預設載入參數 245
  - 新增站點特定載入參數 246
    - ▼ 如何自行撰寫載入感應器 246
      - 規定 246
      - Script 範例 247
- 10. 管理平行環境 251
  - 有關平行環境 251
    - ▼ 如何以 QMON 配置 PE 252
      - ▼ 顯示 PE 的內容 252
      - ▼ 刪除 PE 252
      - ▼ 修改 PE 253
      - ▼ 新增 PE 253
    - ▼ 如何從指令行配置 PE 256
      - qconf PE 選項 256
    - ▼ 如何從指令行顯示配置的 PE 介面 257
    - ▼ 如何以 QMON 顯示配置的 PE 介面 257

|                                    |            |
|------------------------------------|------------|
| PE 啓動程序                            | 258        |
| 終止 PE                              | 259        |
| PE 與 Sun Grid Engine，企業版軟體之緊密整合    | 260        |
| <b>11. 錯誤訊息與疑難排解</b>               | <b>261</b> |
| Sun Grid Engine，企業版 5.3 軟體如何得到錯誤回報 | 261        |
| 不同錯誤或退出代碼之結果                       | 262        |
| 於除錯模式中執行 Sun Grid Engine，企業版程式     | 264        |
| 問題診斷                               | 265        |
| 沒有分派之擱置作業                          | 265        |
| 在錯誤狀態 E 報告的作業或佇列                   | 266        |
| 一般問題疑難排解                           | 267        |



# 前言

---

《*Sun Grid Engine*，企業版 5.3 管理和使用者指南》是一本提供產品背景資訊、安裝指示，以及如何完整使用產品的指示的綜合手冊。

---

## 本書編排架構

因為本指南是針對 Sun Grid Engine，企業版 5.3 產品的使用者和產品責任與使用者的產品責任不盡相同的系統管理者使用，所以本指南分為四部分。每個文件都包含對使用者和管理者有特殊利益的資訊。

這些部分和其目標讀者的說明如下。

- 第 1 部分 — 背景和定義

針對給使用者和管理員使用，指南的這個部分對產品使用、組件、專有名詞等等有詳盡的解釋。

- 第 2 部分 — 入門

針對給安裝產品的人使用 — 一般來說是管理者 — 指南的這個部分對完整的初次和升級安裝有詳細的指示。

- 第 3 部分 — 使用 Sun Grid Engine，企業版 5.3 軟體

指南的這個部分是針對給使用者和管理者使用。包含有涵蓋許多作業的指示和背景資訊。

- 第 4 部分 — 管理

指南的這個部分裡的背景資訊和指示是針對給有經驗的系統管理員使用。

---

# 使用 UNIX 指令

本文件可能不包含基本 UNIX® 指令及程序的相關資訊，如關閉系統、啓動系統及配置裝置等。

請參考下列資訊：

- 《Solaris Handbook for Sun Peripherals》
- 關於 Solaris™ 作業環境的 AnswerBook2™ 線上說明文件
- 其他隨著系統一起收到的軟體說明文件

---

## 排印慣例

| 字型               | 意義                                 | 範例   |
|------------------|------------------------------------|--|
| AaBbCc123        | 指令、檔案和目錄名稱；電腦螢幕輸出。                 | 編輯您的 .login 檔案。<br>使用 <code>ls -a</code> 列出所有檔案。<br>% You have mail。                     |
| <b>AaBbCc123</b> | 您鍵入的內容，與電腦螢幕輸出不同                   | % <b>su</b><br>Password:   |
| <i>AaBbCc123</i> | 書名、新字或專有名詞、要強調的文字。用真實名稱或數值替換指令行變數。 | 請閱讀《使用者指南》的第六章。<br>這些是類別選項。<br>您必須是超級使用者才能執行此項操作。<br>若要刪除某個檔案，請鍵入 <code>rm 檔案名稱</code> 。 |

---

## Shell 提示符號

| Shell                           | 提示符號   |
|---------------------------------|--------|
| C shell                         | 機器名稱 % |
| C shell 超級使用者                   | 機器名稱 # |
| Bourne shell 和 Korn shell       | \$     |
| Bourne shell 和 Korn shell 超級使用者 | #      |

---

## 相關說明文件

| 適用文件 | 標題   | 文件編號        |
|------|--|-------------|
| 參考文件 | 《Sun Grid Engine 5.3 和 Sun Grid Engine，企業版 5.3 參考手冊》 | 816-7479-10 |

---

## 線上存取 Sun 出版文件

部分精選之 Sun 系統說明文件位於：

<http://www.sun.com/products-n-solutions/hardware/docs>

整套的 Solaris 說明文件及其他主題的說明文件則位於：

<http://docs.sun.com>

在該網站上，您也可以找到如何訂購本指南印刷版的相關資訊。

---

# Sun 歡迎您的指教

Sun 一直致力於改善相關的說明文件，因此竭誠歡迎您提出批評和建議。您可以將意見透過電子郵件寄給 Sun：

`docfeedback@sun.com`

請將文件編號 (816-7474-10) 放到您 email 的主題中。

## 部分 I | 背景和定義

---

《*Sun Grid Engine，企業版 5.3 管理和使用者指南*》的這個部分包含單一章節：

- 第一章 – 第 1 頁的「Sun Grid Engine，企業版 5.3 簡介」。

本章很簡單扼要，可能會讓讀者誤解其對使用者和管理者的重要性，如果能熟悉本章內容，兩者的需求就能獲得滿足。本章包含的內容如下。

- 在複雜的計算環境內，對 Sun Grid Engine，企業版 5.3 軟體的主要角色說明
- 產品主要組件和每項組件的功能定義清單
- 在 Sun Grid Engine，企業版 5.3 環境中，須知的重要詞彙



# Sun Grid Engine，企業版 5.3 簡介

本章提供關於 Sun Grid Engine，企業版 5.3 系統的背景資訊，該資訊對使用者和管理員同樣有用。除了說明產品在管理叢集化電腦之混亂世界的角色外，本章包含了以下的主題。

- 簡單解釋格線計算
- QMON 的簡介，Sun Grid Engine，企業版 5.3 圖形使用者介面
- 每個產品重要組件的解釋
- 使用者和管理員可使用的詳細客戶端指令清單
- Sun Grid Engine，企業版 5.3 專有名詞的完整詞彙

## 什麼是格線運算？

格線在概念上是很簡單的。那是執行作業之計算資源的集合。在其最簡單的形式中，對使用者來說，格線就像提供強力分配資源的單一存取點的大型系統。在其較複雜的形式中（本節稍後會做解釋），格線可以提供使用者許多存取點。在任何情況下，使用者把格線視為單一計算資源。資源管理軟體，如 Sun Grid Engine，企業版，會接受使用者所提出的工作，並會依照資源管理政策，排在格線中、適當的系統上執行的時間。使用者可以一次提出數百萬的工作，而不用擔心它們執行的地方。

沒有兩個格線是相同的；一種尺寸不能符合所有的情況。共有三個主要格線類別，階層從單一系統到使用數千個處理器的超級電腦類別的運算中心 (compute farm)。

- **叢集格線** 是最簡單的格線，包含一起運作的電腦主機，提供使用者在單一專案或部門中單一的存取點。
- **Campus grids** 能使組織裡的多重專案或部門分享計算資源。組織可以使用 campus grids 來處理各式各樣的作業，從週期的企業過程到產生資料、資料探勘 (data mining) 和更多的作業。
- **全域格線** 為跨越組織界線的 campus grids 的集合，以建立龐大的虛擬系統。使用者能存取遠超過其組織範圍內能使用資源之計算能力。

圖 1-1 是叢集三種類別的圖形呈現。在叢集格線中，使用者的工作是由叢集中的一個系統所處理的。然而，如果使用者的叢集格線曾經是較複雜的 Campus grid 的一部份—而如果 Campus grid 曾是最大的全域格線的一部份—使用者的工作可能被任何位於世界各地的成員執行主機所處理。

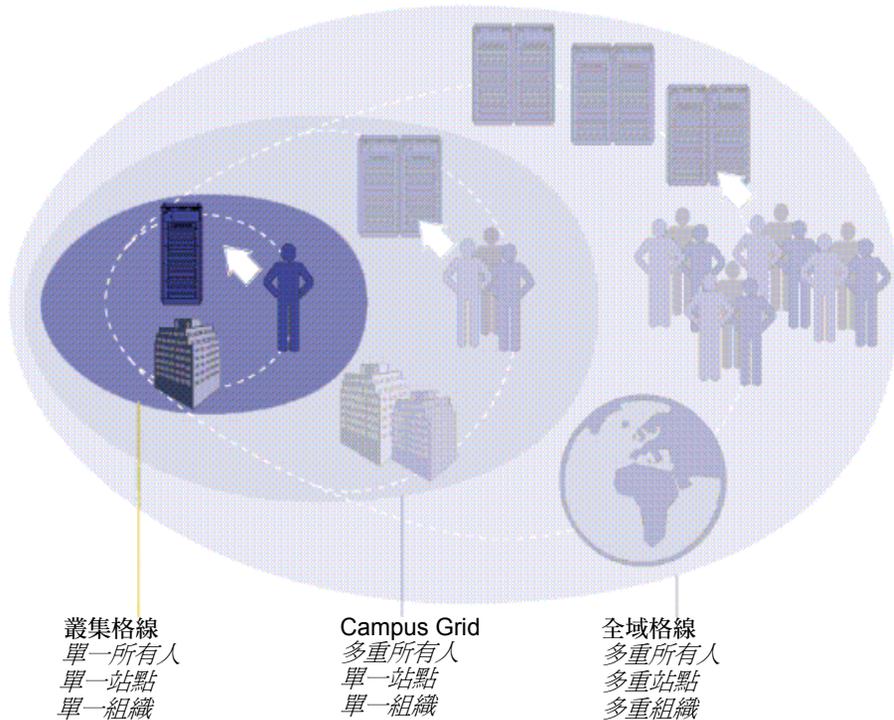


圖 1-1 三種類別的格線

Sun Grid Engine，企業版 5.3 軟體，最新版本的 Sun 資源管理軟體解決方案，提供 Campus grids 所需的能力和靈活性。此產品對其親戚，Sun Grid Engine，所啟動的現存叢集格線非常有用，因為它促成平穩的轉變，並藉由聯合 campus 上所有現存的 Sun Grid Engine 叢集格線來建立 Campus grid。此外，Sun Grid Engine，企業版是企業 campus 第一次對格線計算模式採取行動的好的開始。

Sun Grid Engine，企業版 5.3 軟體會妥善安排以組織技術和管理人員設定之企業資源策略為基礎的運算能力的傳送。Sun Grid Engine 系統使用這些策略檢查 Campus grid 裡的可用運算資源、收集這些資源，然後以將橫跨 Campus grid 用法最佳化的方式自動配置並傳送資源。

若要在 Campus grid 裡啟動合作，使用格線的專案所有人必須協商策略、對獨一無二的專案需求的手動置換策略能有靈活性、並使策略能自動受到監視且執行。

Sun Grid Engine，企業版 5.3 軟體可以消弭許多為競爭計算資源的部門和專案間的應得的權利。

---

## 藉由管理資源和策略來管理工作量

Sun Grid Engine，企業版系統是針對異質性、分配計算環境的先進資源管理工具。工作量管理 — 控制共享資源的使用，使其最能達到企業的目標（例如生產力、時效性、服務水準等等）— 都能透過**資源管理**和**策略管理**而達成。站點配置系統，使其在支援不同級別的時效性（工作截止期限 (deadline)）和重要性（工作優先性和使用者分享）的時候，能將利用和總處理能力最大化。

Sun Grid Engine，企業版軟體提供由多重共享資源組成的 UNIX 環境所用的先進資源管理和策略管理。Sun Grid Engine，企業版系統在以下的主要功能中優於標準的**載入管理**工具。

- 讓 Sun Grid Engine，企業版軟體執行特定站點管理策略的創新動態排程和資源管理。
- 可提供排程者最新工作階層資源消耗及系統裝載資訊的動態效能資料收集。
- 藉由以**認證安全性通訊協定 (Certificate Security Protocol) (CSP)** 式的加密 (encryption) 所增加的安全性可用性。除了以明文方式傳送訊息外，在這個較安全的系統內的訊息都以**私密金鑰 (secret key)** 加密。
- 企業目標，如生產力、時效性、服務水準的定義和實行的高階策略管理。

Sun Grid Engine，企業版軟體提供使用者提交需要計算的工作予 Sun Grid Engine，企業版系統做為相關工作量透明分配的方式。使用者可提交批次作業 (batch job)、互動作業和平行作業予 Sun Grid Engine，企業版系統。

本產品也支援檢查點程式 (checkpointing program)。檢查點作業會在不受使用者載入需求的介入下在工作站間遷移。

對管理員來說，此軟體提供了各式各樣監視和控制 Sun Grid Engine，企業版作業的工具。

---

## 系統作業的方式

Sun Grid Engine，企業版系統能從外界接收工作 — 使用者對電腦資源的需求、把這些作業放置在等候處 (holding area) 等候執行、將作業從等候處傳送到執行裝置、在作業執行中進行管理，並在完成時記錄執行的紀錄。

想像世界上首都的大型「金錢中心」的銀行。

## 使資源與需求相符

在銀行建築物的大廳有很多顧客，每個都有不同的需求，每個都等著被服務。有個客戶只想從他的戶頭提出少量的金錢。在他之後的另一個客戶跟銀行的一個投資專員有約；她希望在進行複雜的投資之前能夠獲得一些建議。在長長的隊伍之中，他們倆個的前面還有另一個客戶想要申請大型的貸款—如同在**她**前面的八個客戶一樣。

不同客戶和不同需求需要不同類型和層級的銀行資源。或許，在這特別的一天，銀行有許多員工有充足的時間來應付顧客從帳戶的簡單提款。但在同一天，只有一兩個貸款行員能夠協助處理許多貸款申請人。情況可能在別的日子裡會有不同。

影響當然是顧客必須等候服務 — 即使當中有許多人的需求符合可用資源而能夠獲得立即的服務。

如果 Sun Grid Engine，企業版系統是銀行經理，它就會以不同方式組織服務。

- 在進入銀行大廳時，顧客就會被要求宣告其名稱、單位（如代表的公司）及其需求。
- 顧客到達的時間會紀錄下來。
- 根據顧客在大廳所提供的資訊，需求和立即可用資源相符的顧客、需求有最高優先權的顧客，以及在大廳等候時間最久的顧客可以獲得服務。
- 當然，在「Sun Grid Engine，企業版銀行」裡，一個銀行行員可以在同一時間為數個顧客提供服務。Sun Grid Engine，企業版系統可能會把新的顧客分派到工作量最少且最適合的銀行行員。
- 做為銀行經理，Sun Grid Engine，企業版系統會讓銀行定義服務策略。典型的服務策略是「提供優先的服務給商業顧客，因為他們可以帶來較多利益」、「確定某些顧客群組獲得良好的服務，因為他們到目前所獲得的服務不佳」、「確認有約的顧客能獲得即時的回應」或「喜好某些對銀行主管有直接要求的顧客」。
- Sun Grid Engine，企業版管理員會自動實行、監視及調整這些策略。有優先存取的顧客會較快獲得服務，需協助其他顧客的銀行行員會給予這些顧客更多的注意力、而如果顧客不能達到預期的進展，Sun Grid Engine，企業版管理員就會察覺，並立刻以調整服務水平做為回應，以符合銀行的服務策略。

## 作業和佇列：Sun Grid Engine 世界

在 Sun Grid Engine，企業版系統中，*作業*對應銀行顧客，作業在電腦等候處而不是在大廳等候，而位在電腦伺服器的*佇列*取代銀行行員的位置，並提供服務給作業。就如同類比中銀行顧客的例子，每個工作的需求—通常包含了可用的記憶體、執行速度、可用軟體版權和類似需求—可能大不相同，而僅有某些佇列可以提供對應的服務。

對應此類比，Sun Grid Engine，企業版軟體會以以下的方式對可用資源和工作需求進行仲裁 (arbitrate)。

- 透過 Sun Grid Engine，企業版系統提交工作的使用者會宣告作業的需求簡介。此外，使用者的身分和其**專案**或**使用者群組**的單位會由系統取得。使用者提交作業的時間也會儲存起來。

- 在佇列排定給新作業執行的時程時，Sun Grid Engine，企業版系統會判定適合佇列的作業，並會立刻以最高的優先權或最長的等候時間分配作業。
- Sun Grid Engine，企業版佇列可允許許多作業同時執行。Sun Grid Engine，企業版系統會嘗試在負荷最少和最適合的佇列啟動新作業。

## 使用策略的多樣性

Sun Grid Engine，企業版叢集的管理員能定義高階的使用者策略，根據任何適合站點的事物自訂。有四個這樣的策略。

- **功能的** — 使用這個策略，管理員能依照使用者或作業與某些使用者群組、專案等等的單位提供特別的處理。
- **共享式的** — 在此策略下，服務的階層是根據指定的共享權利、對應其他使用者和使用者群組的共享資源、過去所有使用者使用資源的方法和目前系統內存在的使用者所訂定的。
- **截止日期** — 每當作業必須在某個時間點前或同一時間完成時，就需呼叫此策略，因此可能需要特別的處理才能達成。
- **置換** — 這個策略需要 Sun Grid Engine，企業版叢集管理員的手動介入，他會修改自動化策略執行。

Sun Grid Engine，企業版的策略管理會自動控制叢集內共享資源的使用，以達到管理的最佳目標。高度優先權的作業會優先分配並在與其他作業爭取資源時會獲得較高的 CPU 使用權。Sun Grid Engine，企業版會監視所有作業的進度並依照策略所定義的目標調整作業相關的優先權。

## 使用票卷範例 (Ticket Paradigm) 的策略管理

這些策略透過獨一無二、名為票卷的 Sun Grid Engine，企業版概念所定義。票卷可以比做上市公司股票的股份。您持有的股票愈多，您對公司愈重要。如果股東 A 持有的股份是股東 B 的兩倍，則 A 的選票也會是 B 的兩倍，因此對公司的重要性也是兩倍。Sun Grid Engine，企業版作業擁有的票卷愈多，則該作業愈重要。如果作業 A 的票卷數是作業 B 的兩倍，則作業 A 所能獲得的資源使用是作業 B 的兩倍。

Sun Grid Engine，企業版作業可以從全部四個策略取得票卷，而票卷的總數 — 以及從每個策略取得的數目 — 經常會隨著時間而改變。

Sun Grid Engine，企業版叢集管理員控制每個策略配置的票卷總數。就如同配置對作業的判定，配置也會判定策略之間相關的重要性。透過指定給特定策略票卷儲存區，管理可以只以分享的模式執行 Sun Grid Engine，企業版系統。管理也可以執行混合模式；例如，90% 為分享，10% 為功能。圖 1-2 表示策略和票卷之間的相互關係。

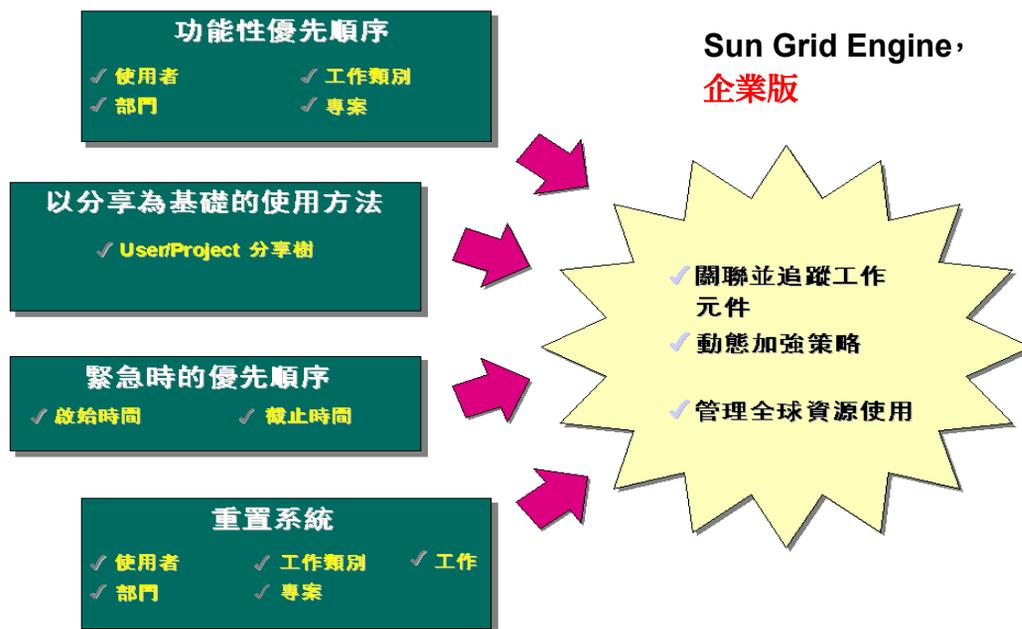


圖 1-2 Sun Grid Engine，企業版 5.3 系統內策略和票卷的相互關係

## Sun Grid Engine，企業版 5.3 組件

圖 1-3 顯示最重要的 Sun Grid Engine，企業版組件和在系統內的互動關係。以下各節解釋組件的功能。

### 主機

有四類型的主機對 Sun Grid Engine，企業版 5.3 系統是十分重要的。

- 主要
- 執行
- 管理

- 提交

## 主要主機

主要主機是全部叢集活動的中心。它會執行主要常駐程式，`sge_qmaster`，和排程器常駐程式，`sge_schedd`。這兩個常駐程式控制所有的 Sun Grid Engine，企業版組件，如佇列和作業，並維護組件狀態、使用者存取權等等的表格。

根據預設值，主要主機也是管理主機和提交主機。請參閱與這些主機相關的各節。

## 執行主機

執行主機是擁有執行 Sun Grid Engine，企業版作業權限的節點。因此，它們管理 Sun Grid Engine，企業版佇列並執行 Sun Grid Engine，企業版執行常駐程式，`sge_execd`。

## 管理主機

主機能被賦予執行 Sun Grid Engine，企業版系統任何種類的管理活動之權限。

## 提交主機

提交主機只對批次作業的提交和控制做考慮。尤其是，登入提交主機的使用者可以透過 `qsub` 提交作業、可以透過 `qstat` 控制作業狀態，而且可以使用 Sun Grid Engine，企業版 OSF/1 Motif 圖形使用者介面，`QMON`，在第 10 頁的「`QMON`，Sun Grid Engine，企業版圖形使用者介面」一節中有其說明。

---

**注意** – 主機在上述的類別中可以歸類到一個以上的類別。

---

## 常駐程式

四個常駐程式提供 Sun Grid Engine，企業版 5.3 系統的功能性。

### `sge_qmaster` – 主要常駐程式

叢集管理和排程活動的中心，`sge_qmaster` 維護主機、佇列、作業、系統負載和使用權限的表格。它會從 `sge_schedd` 接收排程決定，而從適當執行主機上的 `sge_execd` 接收需求活動。

## sgc\_schedd — 排程器常駐程式

排程常駐程式在 `sgc_qmaster` 的協助下維護最新叢集狀態的視區。它會做出以下的排程決定：

- 哪項作業分派到哪個佇列
- 作業如何重新排列並重新排列優先順序，以維護共享、優先權或截止日期

然後它會將這些決定轉寄到 `sgc_qmaster`，其會初始啓動所要求的活動。

## sgc\_execd — 執行常駐程式

執行常駐程式負責主機上的佇列和這些佇列中作業的執行。它會定期將資訊，如作業狀態或主機上的負載，轉寄到 `sgc_qmaster`。

## sgc\_commd — 通訊常駐程式

通訊常駐程式在著名的 TCP 連接埠上通訊。它用於所有 Sun Grid Engine，企業版組件之間的通訊。

## 佇列

Sun Grid Engine，企業版佇列是允許在特別主機上同時執行的作業類別的容器 (container)。佇列決定某些作業屬性；例如，是否能夠遷移。執行作業在它們的生命週期期間與佇列有關。與佇列的關聯性會影響可能發生於作業的某些事項。例如，如果佇列已經暫停，所有與該佇列相關的作業也會跟著暫停。

在 Sun Grid Engine，企業版系統中，您不需將作業直接提交給佇列。您只需要指定作業的需求簡介（例如，記憶體、作業系統、可用軟體等等），而 Sun Grid Engine，企業版軟體會自動將作業分派到低負載主機上的合適佇列。如果作業提交到特別的佇列，則作業會受限於此佇列和其主機，因此 Sun Grid Engine，企業版常駐程式將無法選擇低負載或較合適的裝置。

## 客戶端指令

Sun Grid Engine，企業版的指令行使用者介面是可以讓您管理佇列、提交及刪除作業、檢查作業狀態和暫停 / 啓動佇列和作業的輔助程式（指令）集。Sun Grid Engine，企業版系統 利用以下的輔助程式集。

- `qacct` — 此指令會從叢級日誌檔案取得仲裁計算 (arbitrary accounting) 的資訊。
- `qalter` — 此指令會變更已經提交但在擱置狀態的作業的屬性。
- `qconf` — 此指令提供叢集和佇列配置使用者介面。

- `qdel` — 此指令提供使用者、操作者或管理者傳送信號給作業或其子集的方法。
- `qhold` — 此指令阻止已提交的作業執行。
- `qhost` — 此指令顯示 Sun Grid Engine，企業版執行主機的狀態資訊。
- `qlogin` — 此指令以自動選擇低負載和合適的主機來初始啟動 `telnet` 或類似的登入階段作業。
- `qmake` — 此指令是標準 `make` 功能的替代品。它會依照其功能而延伸 `make`，好將獨立的 `make` 步驟分配到合適機器的叢集。
- `qmod` — 此指令能使所有人暫停或啟動佇列（也會將信息傳道所有目前與此佇列有關，且在作用中的過程）。
- `qmon` — 此指令提供 X-windows Motif 指令介面和監視功能。
- `qresub` — 此指令藉由複製執行或擱置的作業來建立新的作業。
- `qrls` — 此指令會將作業從之前指定的暫停狀態中釋放出來；例如，透過 `qhold`（請見上述）。
- `qrsh` — 此指令可以用於多種用途如下。
  - 欲透過 Sun Grid Engine，企業版系統提供互動應用程式的遠端執行 — 相當於標準 UNIX 功能，`rsh`
  - 欲考慮提交再執行時支援終端機 I/O（標準 / 錯誤輸出和標準輸入）和終端機控制的批次作業
  - 欲提供作業結束前一直維持在作用中的批次作業提交客戶端
  - 欲考慮 Sun Grid Engine，企業版軟體控制之平行作業的遠端作業執行
- `qselect` — 此指令列出對應指定選擇標準的佇列名稱清單。`qselect` 的輸出通常會傳輸到其他 Sun Grid Engine，企業版的指令，好將動作套用在選取的佇列集上。
- `qsh` — 此指令會在低負載的主機上開啓互動的 shell（在 `xterm` 之中）。任何類型的互動作業都可在此 shell 執行。
- `qstat` — 此指令所有與叢集相關的作業和佇列的狀態清單。
- `qsub` — 此指令是用於提交批次作業於 Sun Grid Engine，企業版系統的使用者介面。
- `qtcsch` — 此指令是眾所皆知並廣泛使用的 Unix C-Shell (`csch`) 導數，`tcsch` 完全相容的替代品。它透過 Sun Grid Engine，企業版提供指令 shell 給適當且低負載的主機指定應用程式之透明分配執行的延伸。

所有的程式會透過 `sge_commd` 與 `sge_qmaster` 通訊。這會反映在圖 1-3 所描述，Sun Grid Engine，企業版系統中的組件互動的圖解視區中。

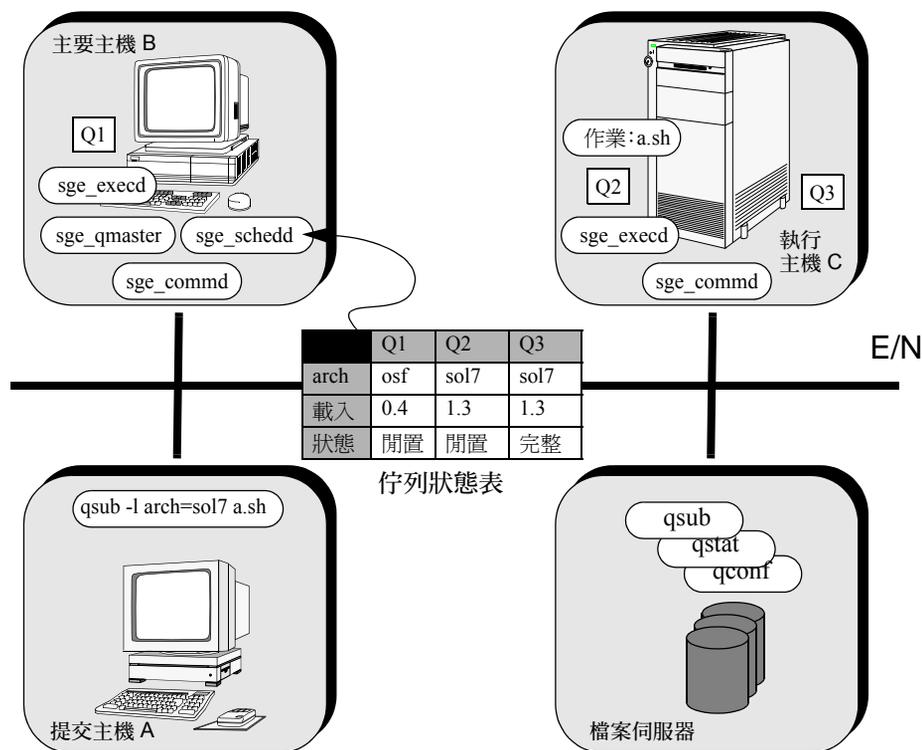


圖 1-3 Sun Grid Engine，企業版系統的組件互動

## QMON，Sun Grid Engine，企業版圖形使用者介面

使用 QMON，圖形使用者介面 (GUI) 工具，您可以完成大部分 — 如果不是全部 — 的 Sun Grid Engine，企業版 5.3 作業。圖 1-4 會顯示 QMON 主功能表，其通常為使用者和管理者功能的起始點。每個主要功能表上的圖示都是您可以按下初始啟動各種作業的 GUI 按鈕。每個按鈕的名稱（當您將滑鼠指標移到按鈕上方會在螢幕上出現文字）也會對其功能有所敘述。

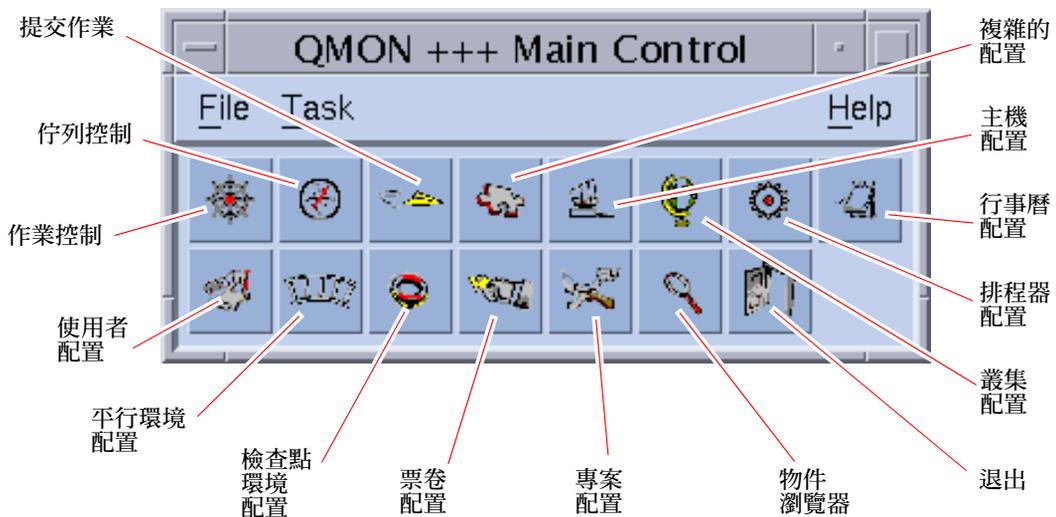


圖 1-4 QMON 主要功能表，定義的

## 自訂 QMON

qmon 的觀感大部分由特別設計的資源檔案所定義。參考的資源檔是放在 `<sgs 根目錄>/qmon/Qmon`，而合理的預設值已經編譯在其中。

叢集管理可以藉由將 qmon 特定資源定義包含到標準 `.Xdefaults` 或 `.Xresources` 檔案中，或是將站點特定 Qmon 檔案放到標準搜尋路徑，如 `XAPPLRESDIR`，所提及的位置上，來把站點特定預設值安裝在標準位置上，如 `/usr/lib/X11/app-defaults/Qmon`。詢問您的管理員看看上述的任何事物是否和您的例子有關，

此外，使用者可以藉由將 Qmon 檔案複製並修改至起始目錄（或至另一個私人 `XAPPLRESDIR` 搜尋路徑所指定的位置）或將必要的資源定義包含進使用者私人的 `.Xdefaults` 或 `.Xresources` 檔案中，來配置個人的喜好設定。私人的 Qmon 資源檔案也可以在作業或 X11 環境的（例如在 `.xinitrc` 資源檔案中）啟動期間透過 `xrdb` 指令安裝。

欲取得可能自訂 (customization) 的詳細資訊，請參閱參考 Qmon 檔案裡的注釋行。

另一個自訂 qmon 的方式已為圖 5-3 和圖 5-13 裡的「作業控制」和「佇列控制」自訂對話方塊做了解釋。在兩個對話方塊中，您可用「儲存」鈕儲存過濾並顯示配置自訂對話方塊至使用者起始目錄下的檔案，.qmon\_preferences 的定義。重新開始之際，qmon 會讀取這個檔案並重新啟動先前定義的運作方式。

---

## Sun Grid Engine 專有名詞的詞彙

詞彙提供了在 Sun Grid Engine，企業版上下文和一般的資源管理中經常使用的專有名詞的簡介。目前有許多專有名詞尚未使用，但將會出現在 Sun Grid Engine，企業版文件的其他部分。

|        |   |
|--------|---|
| 存取清單   | 被授與或拒絕資源的存取，如佇列或某一主機，的使用者和 UNIX 群組清單。使用者和群組可能屬於多重存取清單，而相同的存取清單可以用在不同的情況下。                           |
| 陣列作業   | 包含一個範圍的獨立相同作業的作業。每個作業與個別的作業非常類似。陣列作業的作業唯一不同處是獨特作業辨識器（整數）。   |
| 儲存格    | 有著個別配置和主要機器的個別 Sun Grid Engine，企業版叢集。儲存格可用來鬆散地連接個別的管理單元。  |
| 檢查點    | 將作業執行狀態儲存到所謂的檢查點，並使作業中斷，而在之後恢復後不會遺失資訊和已經完成的工作的程序。若檢查點再執行恢復前就已經移到另一個主機，則此過程稱為遷移。                     |
| 檢查點環境  | Sun Grid Engine，企業版配置實體，其定義與檢查點方法相關的事件、介面和動作。   |
| 叢集     | 機器的集合，名為主機，主機上會有 Sun Grid Engine，企業版功能產生。   |
| 聚合體    | 一套與佇列、主機或整個叢集相關的屬性。   |
| 截止期限策略 | 向必須在截止期限前或截止期限時完成的作業保證優先存取資源的 Sun Grid Engine，企業版策略。管理員可以決定截止期限作業可達到的重要性層級和可以提交截止期限作業的使用者階層。        |
| 部門     | 功能視為相同並置換 Sun Grid Engine，企業版排程策略的使用者和群組清單。使用者和群組可能僅屬於一個部門。   |
| 權利     | 與共享相同（請參閱以下解釋）。僅適用於 Sun Grid Engine，企業版。某個作業、使用者、使用者群組或專案所計劃消耗的資源數量。                                |
| 功能策略   | 指定重要性層級給作業、使用者、使用者群組、專案和作業類別的 Sun Grid Engine，企業版策略。例如，高度優先權的專案（及其全部的作業）可能會透過功能性策略獲得比低優先權專案更高的資源分享。 |
| 群組     | UNIX 群組。  |

## 硬資源 (hard resource)

|      |   |
|------|---|
| 需求   | 在作業啟動前必須配置好的資源。與軟資源 (soft resource) 需求相反。   |
| 主機   | 會有 Sun Grid Engine，企業版功能發生的機器。  |
| 作業   | 批次作業是可以不受使用者介入而執行並且不需要存取終端機的 UNIX shell script。<br><br>互動作業是以 Sun Grid Engine，企業版指令，qrsh、qsh 或 qlogin 啟動的階段作業，其會分別為使用者互動開啓 xterm 視窗或提供遠端登入階段作業的對等物。 |
| 作業類別 | 某種意義上相同並做類似處理的一套作業。在 Sun Grid Engine，企業版之中，作業類別是由對應作業的相同需求和適合這些佇列的佇列特性所定義的。   |
| 管理員  | 可以利用 Sun Grid Engine，企業版所有層面的使用者。主要主機和其他宣告為管理主機的超級使用者擁有管理員的權限。管理員權限也可以指定給非 root 使用者帳號。  |
| 遷移   | 在作業恢復執行前，將檢查點從一個主機移往另一個主機的過程。   |
| 操作員  | 除了無法變更配置卻可以維護作業之外，能執行和管理員一樣指令的使用者。  |
| 置換策略 | 通常用來置換功能性、共享和截止日期策略的自動化資源權利管理的 Sun Grid Engine，企業版策略。Sun Grid Engine，企業版可以將置換指定給作業、使用者、使用者群組、作業類別和專案。   |
| 所有人  | 可以暫停 / 解除暫停和停用 / 啟動他們所擁有的佇列的使用者。通常使用者是常駐在工作站的佇列的所有人。  |
| 平行環境 | Sun Grid Engine，企業版配置實體，其為 Sun Grid Engine，企業版定義必要的介面，以正確處理平行作業。  |
| 平行作業 | 包含一個以上有緊密關係之作業的作業。作業可跨多重主機分配。平行作業通常使用通訊工具，如共享記憶體或訊息傳遞 (MPI, PVM)，將作業同步化並使其產生關聯性。  |
| 策略   | 一套 Sun Grid Engine，企業版管理者可以用來定義 Sun Grid Engine，企業版運作方式的規則和配置。策略將自動由 Sun Grid Engine，企業版實行。   |
| 優先權  | Sun Grid Engine，企業版作業比較其他作業的相關重要性的層級。   |
| 專案   | Sun Grid Engine，企業版專案。  |
| 佇列   | 能同時在 Sun Grid Engine，企業版執行主機上執行的某些類別和數目的作業的容器。  |
| 資源   | 執行作業所消耗或佔用的計算裝置。典型的例子是記憶體、CPU、I/O 頻寬、檔案空間、軟體授權等等。   |
| 共享   | 與權利相同 (請參閱上面解釋)。僅適用於 Sun Grid Engine，企業版。某個作業、使用者、使用者群組或專案所計劃消耗的資源數量。   |

|                |  |
|----------------|--|
| 共享式的策略         | 允許使用者和專案和仲裁群組以階層方式定義權利的 Sun Grid Engine，企業版策略。例如，企業可以細分為處、部門、部門作業中的專案、進行這些專案的使用者群組和這些使用者群組的使用者。共享式的階層稱為共享樹，而一旦共享樹定義之後，它的權利分配就會自動由 Sun Grid Engine，企業版實行。 |
| 共享樹            | Sun Grid Engine，企業版共享策略的階層式定義。   |
| 軟資源            |  |
| 需求             | 作業所需但不必在作業啟動前配置的資源。依照可用的原則配置給作業。與硬資源需求相反。  |
| 暫停             | 暫停執行中的作業但使其維持在執行機器上的過程（和檢查點相反，其中作業已經中止）。暫停的作業仍會消耗一些資源，如交換記憶體或檔案空間。   |
| 票卷             | 在 Sun Grid Engine，企業版的一般資源共享定義單位。Sun Grid Engine，企業版作業、使用者、專案等等的共享愈多，則其就愈重要。例如，如果某作業的票卷是另一個作業票卷的兩倍，則該作業有權利取得兩倍的資源消耗。                                     |
| 用法             | 「消耗的資源」的另一個說法。在 Sun Grid Engine，企業版系統中，用法是由管理員可配置的 CPU 消耗時間、記憶體所佔用時間和 I/O 執行的數量的加權總和所決定的。  |
| 使用者            | 如果該人員至少在一個提交主機或執行主機上有一個有效的登入，則可以使用 Sun Grid Engine，企業版來提交作業或執行作業。  |
| 使用者集 (userset) | 存取清單（請參考以上解釋）或部門（請參考以上解釋）。   |

## 部分 II 入門

---

《*Sun Grid Engine*，企業版 5.3 管理和使用者指南》的這個部分包含單一章節。

- 第二章 – 第 17 頁的「安裝」。

本章包含 Sun Grid Engine，企業版 5.3 產品的首次安裝指示，以及將舊版的產品升級到新版次的指示。



# 安裝

---

本章說明並提供三個安裝作業的詳細指示：

- Sun Grid Engine，企業版 5.3 軟體的完整，首次安裝
- 附有特殊加密功能的安全性安裝
- 安裝驗證

---

**注意** – 本章的指示假設您正在執行 Solaris™ 作業環境的電腦上安裝軟體。任何由其他 Sun Grid Engine，企業版所執行的作業環境架構建立之功能的差異都紀錄在 `<sge 根目錄>/doc` 目錄中，以字串 `arc_depend_` 開頭的檔案。檔案名稱剩餘的部分表示檔案中注釋所應用的作業系統結構。

---

---

## 基礎安裝概觀

---

**注意** – 這些指示只用於首次基礎 Sun Grid Engine、企業版 5.3 安裝。要取得如何安裝新系統並加上額外的安全保護的指示，請參閱第 30 頁的「如何安裝和設定 CSP 安全化系統 (CSP-Secured System)」一節。要取得如何對舊版的 Sun Grid Engine 產品進行現存安裝的升級指示，請參閱《*Sun Grid Engine，企業版 5.3 發行手冊*》。

---

完整安裝包含以下廣義的作業。

- 計劃 Sun Grid Engine，企業版配置和環境
- 從外部媒體讀取 Sun Grid Engine，企業版分配檔案到工作站上。
- 在主要主機和每個 Sun Grid Engine，企業版系統中的執行主機上執行安裝 script。
- 關於管理和提交主機的註冊資訊
- 驗證安裝

安裝應該由熟悉 Solaris 作業環境的人來執行。整個過程由三個階段所合成。

## 階段 1 — 計劃

安裝的計劃階段包含以下作業。

- 決定您的 Sun Grid Engine，企業版環境是否為單一叢集或名為儲存格的子叢集集合
- 選擇將為 Sun Grid Engine，企業版主機的機器。決定每個機器將成為的主機種類 — 主要主機、備用主要主機、管理主機、提交主機、執行主機或主機的組合
- 確認所有的 Sun Grid Engine，企業版使用者在所有提交和執行主機上都有相同的使用者名稱
- 決定 Sun Grid Engine，企業版目錄組織為何。例如，您可以決定組織目錄為每個工作站的完整樹 (complete tree)，或者 cross-mount 目錄，或您可以在某些工作站上設定部分的目錄樹。您也必須決定每個 Sun Grid Engine，企業版 root 目錄的位置為何。
- 決定站點的佇列結構
- 決定網路服務是否定義為 NIS 檔案或為 `/etc/services` 中每個工作站的本端
- 完成安裝工作表（請參考第 24 頁的「在開始安裝前，在類似下表的表格中寫下您的安裝計劃。」的步驟 1），以用在之後的安裝步驟

## 階段 2 — 安裝軟體

安裝階段包含以下作業。

- 建立安裝目錄，並將分配檔案載入
- 安裝主要主機
- 安裝所有執行主機
- 註冊所有管理主機
- 註冊所有提交主機

## 階段 3 — 驗證安裝

驗證階段包含以下作業。

- 檢查常駐程式在主要主機上執行
- 檢查常駐程式在所有的執行主機上執行
- 檢查 Sun Grid Engine，企業版執行簡單指令
- 提交測試作業

# 計劃安裝

在您開始安裝 Sun Grid Engine，企業版 5.3 軟體之前，您必須仔細計劃如何達到完成符合您的環境的結果。本節會幫助您做出會影響本程序剩餘部分的重重大決定。

## 先決條件作業

以下各節說明您需用來安裝生產 Sun Grid Engine，企業版系統的資訊。

### 安裝目錄 <sg\_e 根目錄>

準備好目錄讀進 Sun Grid Engine，企業版分發媒體 (distribution media) 的內容。此目錄稱為 Sun Grid Engine，企業版 *root 目錄*，而之後當 Sun Grid Engine，企業版系統在作業中時，此目錄將用來儲存目前叢集配置和所有需進一步多工緩衝至磁碟的資料。

使用所有主機上正確參照的目錄之路徑名稱。例如，若使用自動裝載器 (automounter) 裝載檔案系統，請將 <sg\_e 根目錄> 設為 /usr/SGE，而非 /tmp\_mnt/usr/SGE。（在整份文件中，在參照安裝目錄時，就會使用 <sg\_e 根目錄> 環境變數。）

<sg\_e 根目錄> 是 Sun Grid Engine，企業版目錄樹的最高層級。每個儲存格裡的 Sun Grid Engine，企業版組件（請參閱第 23 頁的「儲存格」一節）需要對啟動上面的 <sg\_e 根目錄>/<儲存格>/common 有讀取存取。要取得所需權限的說明，請參閱第 21 頁的「檔案存取權限」一節。

為了安裝和管理的簡便，在您想執行 Sun Grid Engine，企業版安裝程序的所有主機上都需可讀取此目錄。例如，您可以透過網路檔案系統（如 NFS）選擇可用的目錄。若您選擇選取主機的本端檔案系統，您將必須在為特定機器啟動安裝程序前把安裝目錄複製到每個主機。

### Root 目錄下的多工緩衝處理目錄

- 在 Sun Grid Engine，企業版主要主機上，多工緩衝處理目錄維護在 <sg\_e 根目錄>/<儲存格>/spool/qmaster 和 <sg\_e 根目錄>/<儲存格>/spool/schedd 之下。
- 在每個執行主機上，維護名為 <sg\_e 根目錄>/<儲存格>/spool/<exec\_host> 的多工緩衝處理目錄。

您不需要將這些目錄匯出至其他機器。然而，匯出整個 <sg\_e 根目錄> 樹並使主要和所有執行主機可寫入存取可增加管理的簡便性。

## 目錄組織

決定 Sun Grid Engine，企業版目錄組織的角色為何（例如，每個工作站上的完整樹、cross-mounted 目錄、某些工作站上的部分目錄樹），以及每個 Sun Grid Engine，企業版 root 目錄，<sgc 根目錄>，的所在位置。

---

**注意** – 因為安裝目錄和 / 或多工緩衝處理目錄的變更基本上需要新的系統安裝（儘管能保存所有前次安裝的重要資訊），所以您必須更小心選擇適合的前置 (upfront) 安裝目錄。

---

根據預設值，Sun Grid Engine，企業版安裝程序將在安裝目錄下的目錄階層（請參閱第 20 頁的「樣本目錄階層」，圖 2-1）安裝 Sun Grid Engine，企業版系統、手冊、多工緩衝處理區和配置檔案。若您接受此預設運作方式，您應該安裝 / 選擇允許第 21 頁的「檔案存取權限」所述之存取權限的目錄。

您可以選擇多工緩衝處理區域來放置主要安裝期間的其他位置（相關指示請參閱第六章第 129 頁的「主機和叢集配置」）。

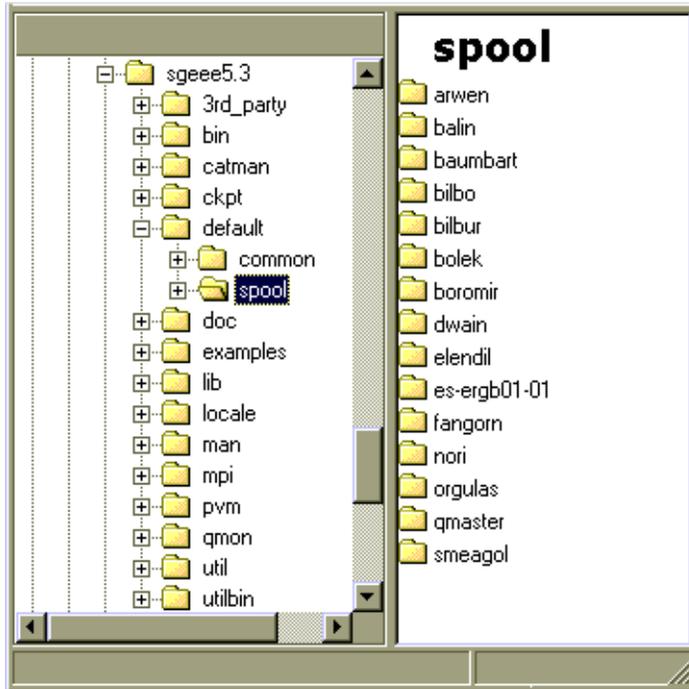


圖 2-1 樣本目錄階層

## 磁碟空間需求

Sun Grid Engine，企業版目錄樹有以下固定的磁碟空間需求。

- 沒有任何二進位的安裝工具箱（包含說明文件）需要 40 MB
- 每一套的二進位（架構 Cray (architecture Cray) 除外，其中二進位消耗約 35 MB）需要 10 到 15 MB

Sun Grid Engine，企業版日誌檔案的理想磁碟空間如下。

- 主要主機多工緩衝處理目錄需要 30 到 200 MB，依據叢集的大小而定
- 每個執行主機需要 10 到 20 MB

---

**注意** – 可以配置主要主機和執行主機的多工緩衝處理目錄，而這些目錄不需常駐在 `<sgc 根目錄>` 之下（就是它們的預設位置）。變更多工緩衝處理目錄的位置應該在主要安裝後完成（相關指示請參閱第六章第 129 頁的「主機和叢集配置」）。

---

## 安裝帳號

您可以將 Sun Grid Engine，企業版安裝在 root 帳號或無權限的（例如您自己的）帳號下。若您安裝在無權限的帳號下，該安裝僅會允許特定使用者執行 Sun Grid Engine，企業版作業。存取所有其他帳號將被拒絕。安裝在 root 帳號下將解決這項限制；但是，完整安裝程序需要 root 權限。

## 檔案存取權限

若您安裝為 root，您可能在共享的檔案系統上有為全部主機配置 root 讀取 / 寫入的問題，因此，您可能會有將 `<sgc 根目錄>` 放到全網路檔案系統的問題。您可以強制 Sun Grid Engine，企業版軟體透過非 root 的管理使用者帳號（例如，名為 `sgcadmin`）執行所有 Sun Grid Engine，企業版元件的完整檔案處理。因此，您只需給特定的使用者共享 root 檔案系統讀取 / 寫入存取權。Sun Grid Engine，企業版安裝程序會詢問您是否要在管理使用者帳號下放置檔案處理。若您回答「是」並提供有效的 *使用者名稱*，檔案處理將透過該使用者名稱執行。否則，將使用您執行安裝程序的使用者名稱。

您必須確定在所有的範例中，用做檔案處理的帳號必須在所有主機上有 Sun Grid Engine，企業版 root 目錄的讀取 / 寫入存取權。此外，安裝程序會假設您要讀進 Sun Grid Engine，企業版分發媒體的主機能夠存取該目錄。

## 網路服務

決定您站點的網路服務要定義為 NIS 檔案或在 `/etc/services` 之工作站的本端。若您的站點使用 NIS，請找出 NIS 伺服器主機，好讓您將項目新增至服務 NIS 對映。

Sun Grid Engine、企業版服務為 `sge_commd`。若要新增服務到 NIS 對映，請選擇一個保留、未使用過的連接埠號碼 – 1024 之下的任一數字。以下是一個 `sge_commd` 項目的範例。

```
sge_commd 536/tcp
```

## 主要主機

這是控制 Sun Grid Engine，企業版的主機，其執行主要常駐程式，`sge_qmaster`。主要主機以 Sun Grid Engine，企業版功能的作業為中心，所以必須遵守以下的需求。

- 它必須是個穩定的平台。
- 它不能因其他的處理過程而過度忙碌。
- 它至少要有 20 MB 的未使用的主要記憶體來執行 Sun Grid Engine，企業版常駐程式。對超大型的叢集來說 – 就是任何時間在系統內包含數百個或數千個主機和上萬個作業的叢集 – 則需要 1 GB 或更多的未使用之主要記憶體，而有兩個 CPU 會有更多好處。
- 它可選擇是否要有 Sun Grid Engine，企業版目錄，`<sge 根目錄>`，為其本端以減少網路流量。

## 備用主要主機

這些主機會備份 `sge_qmaster` 的功能，以防主要主機或主要常駐程式故障。做為備份主要主機，該機器必須有以下的特性。

- 它必須執行 `sge_shadowd`。
- 它必須分享 `sge_qmaster` 的狀態、作業和紀錄在磁碟的佇列配置資訊。特別是，備份主要主機需要至 `sge_qmaster` 的多工緩衝處理目錄和 `<sge 根目錄>/<儲存格>/common` 目錄的讀取/寫入 root 或管理使用者存取權。
- `<sge 根目錄>/<儲存格>/common/shadow_masters` 檔案必須包含一行定義主機為備用主要主機的數字。

這些條件符合之後就會啟動主機的備份主要主機功能。所以您不需要重新啟動 Sun Grid Engine，企業版常駐程式來將主機變成備用主機。

## 執行主機

這些主機執行提交給 Sun Grid Engine，企業版的作業。您將在每個執行主機上執行安裝 script。

## 管理主機

Sun Grid Engine，企業版操作者和管理者執行管理作業，如重新配置佇列或從這些主機新增 Sun Grid Engine，企業版使用者。主要主機安裝 script 會自動將主要主機變成管理主機。

## 提交主機

Sun Grid Engine，企業版作業可能從提交主機提交及控制。主要主機安裝 script 會自動將主要主機變成提交主機。

## 儲存格

您可以將 Sun Grid Engine，企業版設定為單一叢集或名為 *儲存格* 的鬆散組合叢集的集合。SGE\_CELL 環境變數指出參照的叢集。當 Sun Grid Engine，企業版安裝為單一叢集時，SGE\_CELL 並無設定，而儲存格數值假定為數值 default。

## 使用者名稱

為了讓 Sun Grid Engine，企業版驗證提交作業的使用者有提交作業和使用他們所需的執行主機的權限，使用者在相關的提交和執行主機的名稱必須相同。這個需求可能讓變更某些機器上的使用者名稱變成不可避免。

---

**注意** – 主要主機上的使用者名稱和檢查權限沒有關係，名稱不需相同，甚至不用存在。

---

## 佇列

計劃符合您站點需求的佇列結構。這表示決定哪個佇列要放在哪個執行主機上、您的連續、互動、平行和其他作業類型是否需要佇列、每個佇列需要多少作業插槽和其他佇列配置決定。

Sun Grid Engine，企業版管理員也可能讓安裝程序建立預設的佇列結構，其適合做為系統熟悉以及日後調整的起點。

---

**注意** – 除了安裝進 Sun Grid Engine，企業版軟體的目錄，大部分 Sun Grid Engine，企業版安裝程序所建立的設定可以在系統作業期間進行同步變更。

---

除非您已經熟悉 Sun Grid Engine，企業版或是您之前已經決定您要在叢集使用的佇列結構，否則您不該允許安裝程序替您安裝預設的佇列結構。反之，您應該準備一個指定該佇列結構的文件，並在完成安裝過程後直接前進到第七章第 151 頁的「配置佇列與佇列行事曆」。

## ▼ 如何計劃安裝

1. 在開始安裝前，在類似下表的表格中寫下您的安裝計劃。

| 參數              | 數值 |
|-----------------|----|
| <sgc 根目錄>       |    |
| 管理使用者           |    |
| 管理群組            |    |
| sgc_commd 連接埠編號 |    |
| 主要主機            |    |
| 備用主要主機          |    |
| 執行主機            |    |
| 管理主機            |    |
| 提交主機            |    |

圖 2-2 安裝前需填寫的範本表格

2. 確認檔案系統和目錄將包含 Sun Grid Engine，企業版分發，而透過上述的定義設定存取權限，多工緩衝處理與配置檔案已正確完成設定。

## ▼ 如何讀取分發媒體

Sun Grid Engine，企業版在 CD-ROM 上分發。詢問您的系統管理員或參考本端系統說明文件，以取得如何存取 CD-ROM 的資訊。CD-ROM 分發包含一個名為 Sun\_Grid\_Engine\_Enterprise\_5.3 的目錄。產品分發位在這個目錄中，分別為 tar 格式和 Sun Microsystems pkgadd 格式。pkgadd 格式是優先格式。

1. 建立管理者帳號（請參閱第 21 頁的「檔案存取權限」一節）。
2. 提供至分發媒體的存取並登入系統中 — 最好是直接連接檔案伺服器的系統。
3. 請依照第 19 頁的「安裝目錄 <sge 根目錄>」一節所述，建立安裝目錄以讀進 Sun Grid Engine，企業版安裝工具箱，以確定安裝目錄的存取權限已經正確設定。  
在這些指示中，安裝目錄縮寫為 <安裝目錄>。
4. 為 Sun Grid Engine，企業版叢集的 qmaster、執行和提交主機所要使用的二進位架構安裝二進位。  
根據您使用的安裝方法，執行下列事項之一。

### pkgadd 方法

當您輸入以下的指令，您必須準備回應關於基礎目錄的 script 問題（預設值為 /gridware/sge）、管理使用者（預設值為 sgeadmin）和管理使用者群組（預設值為 adm）。script 會要求您在這個安裝的計劃步驟期間所做的選擇（請參閱第 24 頁的「如何計劃安裝」一節）。

- a. 請在指令提示符號處輸入以下指令，回應產生的 script 問題。

```
# cd <cdrom 裝載點>/Sun_Grid_Engine_Enterprise_5.3/Packages
# pkgadd -d . SDRMEcomm
# pkgadd -d . SDRMEDoc
# pkgadd -d . SDRMEsp32 (此為可有可無；至少需要一個二進位集)
# pkgadd -d . SDRMEsp64 (此為可有可無；至少需要一個二進位集)
```

這些指令安裝以下的套裝軟體。

- SDRMEcomm — 安裝架構獨立檔案
- SDRMEDoc — 安裝說明文件
- SDRMEsp32 — 為 Solaris 2.6、Solaris 7、Solaris 8 和 Solaris 9 作業環境安裝 Solaris (SPARC® 平台) 32 位元二進位
- SDRMEsp64 — 為 Solaris 7、Solaris 8 和 Solaris 9 作業環境安裝 Solaris (SPARC 平台) 64 位元二進位

## tar 方法

- b. 在指令提示符號處輸入以下指令（在此範例中 <tar 目錄> 是 <cdrom 裝載點> /Sun\_Grid\_Engine\_Enterprise\_5.3/tar 完整目錄的縮寫）。

```
# cd <sgc 根目錄>
# gzip -dc <tar_dir>/sgccc-5_3-common.tar.gz | tar xvpf -
# gzip -dc <tar 目錄>/sgccc-5_3-doc | tar xvpf -
# gzip -dc <tar 目錄>/sgccc-5_3-bin-solsparc32.tar.gz | tar xvpf -
# gzip -dc <tar 目錄>/sgccc-5_3-bin-solsparc64.tar.gz | tar xvpf -
# util/setfileperm.sh <管理使用者> <管理群組> <sgc 根目錄>
```

- solsparc32 tar 檔案包含 Solaris 2.6、Solaris 7、Solaris 8 和 Solaris 9 作業環境所需的 Solaris（SPARC® 平台）32 位元二進位。
  - solsparc64 tar 檔案包含 Solaris 7、Solaris 8 和 Solaris 9 作業環境所需的 Solaris（SPARC 平台）64 位元二進位。
5. 從指令提示符號執行以下程序。

```
% cd <安裝目錄>
% tar -xvpf 分發來源
```

其中 <安裝目錄> 是安裝目錄的路徑名稱，而分發來源是 CD-ROM 上磁帶文件資料庫檔案的名稱。這會讀進 Sun Grid Engine，企業版安裝工具箱。

---

## 執行基本安裝

以下各節說明如何安裝 Sun Grid Engine、企業版 5.3 系統的所有元件，包含主要、執行、管理和提交主機。

---

**注意** – 若您想用加強的安全性安裝系統，請在您繼續安裝前，參閱第 29 頁的「用加強的安全性安裝」一節。

---

## ▼ 如何安裝主要主機

---

**注意** – Sun Grid Engine，企業版安裝程序為系統建立在該系統上執行的預設配置。它會詢問管理安裝的作業系統類型，並根據此項資訊做出有意義的設定。

---

1. 以 root 身分登入主要主機。
2. 根據安裝工具箱常駐的目錄是否能從主要主機看見，請執行以下任一作業。
  - a. 若安裝工具箱常駐的目錄能從主要主機看見，請將目錄 (cd) 變更至安裝目錄，然後再繼續進行步驟 3。
  - b. 如果不能看見目錄而且不能變為可見，請執行以下事項。
    - i. 在主要主機上建立本端安裝目錄。
    - ii. 透過網路（例如，使用 ftp 或 rcp）將安裝工具箱複製到本端安裝目錄。
    - iii. 將目錄 (cd) 變更至本端安裝目錄。
3. 執行以下指示

---

**注意** – 若您要透過認證安全性通訊協定方法執行安裝，您必須將 `-csp` 旗號新增至以下指令中（請參閱第 30 頁的「如何安裝和設定 CSP 安全化系統 (CSP-Secured System)」）。

---

```
% ./install_qmaster
```

這會初始啟動主要安裝程序。您將被詢問數個問題，並可能需要執行幾個管理動作。這些問題和動作項目是能自我解釋的。

---

**注意** – 有第二個作用中的終端機階段作業來執行管理作業會很簡便。

---

主要安裝程序會建立 `sge_qmaster` 和 `sge_schedd` 所要求的合適目錄階層。該程序會在主要主機上啟動 Sun Grid Engine，企業版組件 `sge_commd`、`sge_qmaster` 和 `sge_schedd`。主要主機也會登錄為具有管理和提交權限的主機。

若您確信有異狀發生，您可以隨時中斷並重複安裝程序。

## ▼ 如何安裝執行主機

1. 以 root 身分登入執行主機。

2. 至於主要安裝，請將安裝工具箱複製到本端安裝目錄，或使用網路安裝目錄。
3. 將目錄 (cd) 變更至安裝目錄，並執行以下指令。

---

**注意** – 若您要透過認證安全性通訊協定方法執行安裝，您必須將 `-csp` 旗號新增至以下指令中（請參閱第 30 頁的「如何安裝和設定 CSP 安全化系統 (CSP-Secured System)」）。

---

```
% ./install_execd
```

這會初始啟動執行主機安裝程序。執行主機安裝程序的運作方式和處理和主要主機的非常像。

4. 回應來自安裝 script 的提示符號。

---

**注意** – 您也可以使用主要主機執行作業。您只要實行主要機器的執行主機安裝即可。另外，若您使用一台非常慢的機器做為主要主機，或如果您的叢集相當龐大，您應該只使用主要機器執行主要作業。

---

執行主機安裝程序會建立 `sge_execd` 所要求的合適目錄階層。該程序會在執行主機上啟動 Sun Grid Engine，企業版組件 `sge_commd` 和 `sge_execd`。

## ▼ 如何安裝管理和提交主機

主要主機暗中允許執行管理作業，並提交、監視和刪除作業。它不需要任何種類的額外安裝做為管理或提交主機。相反地、純粹的管理和提交主機則需要登記。

- 從管理主機（如主要主機）到透過管理帳號（如 `superuser` 帳號），請輸入以下指令。

```
% qconf -ah 管理主機名稱 [...]  
% qconf -as 提交主機名稱 [...]
```

要取得配置其他主機類型的詳細細節和其他方法，請參考第 131 頁的「關於常駐程式和主機」一節。

---

# 用加強的安全性安裝

您可以使用以下指示更安全設定系統。這些指示會協助您使用**認證安全性通訊協定 (CSP)** 式的加密來設定系統。

Sun Grid Engine 5.3 和 Sun Grid Engine，企業版 5.3 產品都能利用這個安全設定，而這些指示適用於這兩個產品。爲了達成簡單扼要的目的，這些指示只引用 Sun Grid Engine 產品。

除了以明文方式傳送訊息外，在這個較安全的系統內的訊息都以私密金鑰 (secret key) 加密。私密金鑰透過公共 / 私人金鑰通訊協定交換。使用者透過 Sun Grid Engine 系統提出證明文件以證明身分，並從 Sun Grid Engine 系統接收證明文件來證實使用者正與正確的系統通訊。在初始宣佈階段之後，通訊會以加密的形式繼續透明進行。階段作業的有效性只維持一段時間，在這之後，階段作業必須重新宣佈。

## 所需的額外設定

設定 Sun Grid Engine 系統的認證安全性通訊協定加強版的步驟非常類似標準設定。您通常依循第 24 頁的「如何計劃安裝」、第 25 頁的「如何讀取分發媒體」、第 27 頁的「如何安裝主要主機」、第 27 頁的「如何安裝執行主機」和第 28 頁的「如何安裝管理和提交主機」等節的指示。

然而，以下額外的作業是必要的。

- 主要主機上認證權限 (CA) 系統金鑰和證明文件的產生  
這是透過呼叫具有 `-csp` 旗號的安裝 script 而完成的。
- 分發系統金鑰和證明文件至執行和提交主機  
安全執行是系統管理員的工作；即金鑰必須以安全的方式傳輸到執行主機和提交主機，如透過 `ssh`。
- 使用者金鑰和證明文件的產生  
這可以在主要安裝後自動由系統管理員執行。
- 系統管理員允許新使用者進入

## ▼ 如何安裝和設定 CSP 安全化系統 (CSP-Secured System)

1. 用第 17 頁的「基礎安裝概觀」、第 19 頁的「計劃安裝」和第 26 頁的「執行基本安裝」各節所規劃的方式安裝 Sun Grid Engine 系統，但有以下例外：在呼叫不同安裝 script 時，使用額外的旗號，`-csp`。

例如，在安裝主要主機的基本安裝指示告訴您輸入 `./install_qmaster` 來呼叫 script 之處，您可加入 `-csp` 旗號來修改該指示。因此，要安裝 CSP 安全性系統，您可輸入以下指令變更主要主機安裝程序。

```
% ./install_qmaster -csp
```

2. 回應來自安裝 script 的提示符號。

若要產生 CSP 證明文件和金鑰，您必須提供以下的資訊。

- 兩個字母的國碼 — 例如，US 代表美國
- 州
- 位置 — 如城市
- 組織
- 組織單位
- CA 電子郵件地址

當安裝繼續進行時，認證權限也會建立。Sun Grid Engine 特定 CA 建立在主要主機上。含有安全性相關資訊的目錄如下。

- 在 `$SGE_ROOT/{預設| $SGE 儲存格}/common/sgeCA` 之下，儲存可公開存取的 CA 和常駐程式證明文件。
- 在 `/var/sgeCA/{sge 服務| 連接埠 $COMM 連接埠}/{預設| $SGE 儲存格}/private` 之下，儲存對應的私密金鑰。
- 在 `/var/sgeCA/{sge 服務| 連接埠 $COMM 連接埠}/{預設| $SGE 儲存格}/userkeys/$USER` 之下，儲存使用者金鑰和證明文件。

在這個過程中，script 輸出看起來類似程式範例 2-1 的範例。

## 程式範例 2-1 CSP 安裝 Script – 目錄建立

為 OpenSSL 安全性架構初始啟動認證權限 (CA)

```
-----  
Creating /scratch2/eddy/sge_sec/default/common/sgeCA  
Creating /var/sgeCA/port6789/default  
Creating /scratch2/eddy/sge_sec/default/common/sgeCA/certs  
Creating /scratch2/eddy/sge_sec/default/common/sgeCA/crl  
Creating /scratch2/eddy/sge_sec/default/common/sgeCA/newcerts  
Creating /scratch2/eddy/sge_sec/default/common/sgeCA/serial  
Creating /scratch2/eddy/sge_sec/default/common/sgeCA/index.txt  
Creating /var/sgeCA/port6789/default/userkeys  
Creating /var/sgeCA/port6789/default/private  
Hit Return to continue >>
```

設定目錄之後，會產生 CA 特定的證明文件和私密金鑰。Sun Grid Engine 系統使用來自特別檔案的虛擬隨機資料 (pseudo random data) 或用 /dev/random (如果有的話)，來催化虛擬隨機編號產生器 (pseudo random number generator/PRNG)。(要取得更多隨機編號的詳細資訊，請參閱 <http://www.openssl.org/support/faq.html> 和 <http://www.cosy.sbg.ac.at/~andi>。)

在安裝 CA 基礎架構後，應用程式證明文件、使用者證明文件和私密金鑰由 CA 建立並簽署給管理使用者、虛擬常駐程式使用者和使用者，root。script – 其輸出類似程式範例 2-2 中的範例 – 對網站資訊的第一個查詢。

## 程式範例 2-2 CSP 安裝 Script – 資訊蒐集

```
Creating CA certificate and private key
-----

Please give some basic parameters to create the distinguished name (DN)
for the certificates.

We will ask for

    - the two letter country code
    - the state
    - the location, e.g city or your buildingcode
    - the organization (e.g. your company name)
    - the organizational unit, e.g. your department
    - the email address of the CA administrator (you!)

Hit Return to continue >>

Please enter your two letter country code, e.g. >US< >> DE
Please enter your state >> Bavaria
Please enter your location, e.g city or buildingcode >> Regensburg
Please enter the name of your organization >> Myorg
Please enter your organizational unit, e.g. your department >> Mydept
Please enter the email address of the CA administrator >> admin@my.org

You selected the following basic data for the distinguished name of
your certificates:

Country code:          C=DE
State:                 ST=Bavaria
Location:              L=Regensburg
Organization:          O=Myorg
Organizational unit:   OU=Mydept
CA email address:      emailAddress=admin@my.org

Do you want to use these data (y/n) [y] >>
```

在確認您提供的資訊正確後，安裝程式會繼續 CA 證明文件和私密金鑰的產生，從設定 CA 基礎架構開始。script 輸出類似程式範例 2-3 中的範例。

### 程式範例 2-3 CSP 安裝 Script – CA 基礎架構建立

```
Creating RANDFILE from >/kernel/genunix< in
>/var/sgeCA/port6789/default/private/rand.seed<

1513428 semi-random bytes loaded
Creating CA certificate and private key

Using configuration from /tmp/sge_ca14364.tmp
Generating a 1024 bit RSA private key
.....+++++
.....+++++
writing new private key to '/var/sgeCA/port6789/default/private/cakey.pem'
-----
Hit Return to continue >>
```

在安裝 CA 基礎架構後，CA 會建立並簽署應用程式、使用者證明文件和私密金鑰給虛擬常駐程式使用者和 root 使用者。Script 輸出與下幾頁所顯示的輸出類似。請注意，範例中有些行列都會進行縮寫，好放進這幾頁的每個單一行列中。縮寫的部份以 (...) 表示。

### 程式範例 2-4 CSP 安裝 Script – 證明文件和私密金鑰建立

```
Creating Daemon certificate and key
-----

Creating RANDFILE from >/kernel/genunix< in >/var/sgeCA/(...)/rand.seed<

1513428 semi-random bytes loaded
Using configuration from /tmp/sge_ca14364.tmp
Generating a 1024 bit RSA private key
.....+++++
.....+++++
writing new private key to '/var/sgeCA/port6789/default/private/key.pem'
-----
Using configuration from /tmp/sge_ca14364.tmp
Check that the request matches the signature
Signature ok
The Subjects Distinguished Name is as follows
countryName           :PRINTABLE:'DE'
stateOrProvinceName   :PRINTABLE:'Bavaria'
localityName          :PRINTABLE:'Regensburg'
```

程式範例 2-4 CSP 安裝 Script – 證明文件和私密金鑰建立 (接上頁)

```
organizationName      :PRINTABLE:'Myorg'
organizationalUnitName :PRINTABLE:'Mydept'
uniqueIdentifier       :PRINTABLE:'root'
commonName             :PRINTABLE:'SGE Daemon'
emailAddress           :IA5STRING:'none'
Certificate is to be certified until Mar  5 13:50:57 2003 GMT (365 days)

Write out database with 1 new entries
Data Base Updated
created and signed certificate for SGE daemons
Creating RANDFILE from >/kernel/genunix< in>/var/(...)/userkeys/root/rand.seed<

1513428 semi-random bytes loaded
Using configuration from /tmp/sge_ca14364.tmp
Generating a 1024 bit RSA private key
.....++++++
.....++++++
writing new private key to '/var/sgeCA/port6789/default/userkeys/root/key.pem'
-----
Using configuration from /tmp/sge_ca14364.tmp
Check that the request matches the signature
Signature ok
The Subjects Distinguished Name is as follows
countryName           :PRINTABLE:'DE'
stateOrProvinceName   :PRINTABLE:'Bavaria'
localityName          :PRINTABLE:'Regensburg'
organizationName      :PRINTABLE:'Myorg'
organizationalUnitName :PRINTABLE:'Mydept'
uniqueIdentifier       :PRINTABLE:'root'
commonName             :PRINTABLE:'SGE install user'
emailAddress           :IA5STRING:'none'
Certificate is to be certified until Mar  5 13:50:59 2003 GMT (365 days)

Write out database with 1 new entries
Data Base Updated
created and signed certificate for user >root< in >/var/(...)/userkeys/root<
Creating RANDFILE from >/kernel/genunix< in >/var/(...)/userkeys/eddy/rand.seed<
1513428 semi-random bytes loaded
Using configuration from /tmp/sge_ca14364.tmp
Generating a 1024 bit RSA private key
.....++++++
.....++++++
writing new private key to '/var/sgeCA/port6789/default/userkeys/eddy/key.pem'
-----
Using configuration from /tmp/sge_ca14364.tmp
Check that the request matches the signature
Signature ok
```

程式範例 2-4 CSP 安裝 Script – 證明文件和私密金鑰建立 (接上頁)

```
The Subjects Distinguished Name is as follows
countryName      :PRINTABLE:'DE'
stateOrProvinceName  :PRINTABLE:'Bavaria'
localityName      :PRINTABLE:'Regensburg'
organizationName   :PRINTABLE:'Myorg'
organizationalUnitName :PRINTABLE:'Mydept'
uniqueIdentifier   :PRINTABLE:'root'
commonName        :PRINTABLE:'SGE install user'
emailAddress       :IA5STRING:'none'
Certificate is to be certified until Mar  5 13:50:59 2003 GMT (365 days)

Write out database with 1 new entries
Data Base Updated
created and signed certificate for user >root< in >/var/(...)/userkeys/root<
Creating RANDFILE from >/kernel/genunix< in >/var/(...)/userkeys/eddy/rand.seed<

1513428 semi-random bytes loaded
Using configuration from /tmp/sge_cal4364.tmp
Generating a 1024 bit RSA private key
.....++++++
.....++++++
writing new private key to '/var/sgeCA/port6789/default/userkeys/eddy/key.pem'
-----
Using configuration from /tmp/sge_cal4364.tmp
Check that the request matches the signature
Signature ok
The Subjects Distinguished Name is as follows
countryName      :PRINTABLE:'DE'
stateOrProvinceName  :PRINTABLE:'Bavaria'
localityName      :PRINTABLE:'Regensburg'
organizationName   :PRINTABLE:'Myorg'
organizationalUnitName :PRINTABLE:'Mydept'
uniqueIdentifier   :PRINTABLE:'root'
commonName        :PRINTABLE:'SGE admin user'
emailAddress       :IA5STRING:'none'
Certificate is to be certified until Mar  5 13:51:02 2003 GMT (365 days)

Write out database with 1 new entries
Data Base Updated
created and signed certificate for user >root< in >/var/(...)/userkeys/root<
Creating RANDFILE from >/kernel/genunix< in >/var/(...)/userkeys/eddy/rand.seed<

1513428 semi-random bytes loaded
Using configuration from /tmp/sge_cal4364.tmp
Generating a 1024 bit RSA private key
.....++++++
.....++++++
```

## 程式範例 2-4 CSP 安裝 Script – 證明文件和私密金鑰建立 (接上頁)

```
writing new private key to '/var/sgeCA/port6789/default/userkeys/eddy/key.pem'
-----
Using configuration from /tmp/sge_ca14364.tmp
Check that the request matches the signature
Signature ok
The Subjects Distinguished Name is as follows
countryName          :PRINTABLE:'DE'
stateOrProvinceName  :PRINTABLE:'Bavaria'
localityName         :PRINTABLE:'Regensburg'
organizationName     :PRINTABLE:'Myorg'
organizationalUnitName :PRINTABLE:'Mydept'
uniqueIdentifier     :PRINTABLE:'root'
commonName           :PRINTABLE:'SGE admin user'
emailAddress         :IA5STRING:'none'
Certificate is to be certified until Mar  5 13:51:02 2003 GMT (365 days

Write out database with 1 new entries
Data Base Updated
created and signed certificate for user >eddy< in >/var/(...)/userkeys/eddy<
Hit Return to continue >>
```

在主要主機，sge\_qmaster，的安全性相關設定完成後，script 會提示繼續進行安裝程序的剩餘部分，其與程式範例 2-5 中的範例類似。

## 程式範例 2-5 CSP 安裝 Script – 安裝的延續

```
SGEEE startup script
-----

Your system wide SGEEE startup script is installed as:

    "/scratch2/eddy/sge_sec/default/common/rcsge"

Hit Return to continue >>
```

### 3. 請執行以下一項。

- a. 若您認為共享檔案系統的安全性不足以將 CSP 安全性資訊維持在一個執行常駐程式可存取的地方，則請前進至步驟 4。
- b. 若您認為共享檔案系統的安全性是足夠的，請繼續第 27 頁的「如何安裝執行主機」所規劃的基礎安裝程序。  
當您為執行主機安裝呼叫 `./install_execd script` 時，請記得加入 `-csp` 旗號。  
在完成所有其餘的安裝步驟後，請參照第 39 頁的「如何產生使用者證明文件和私密金鑰」一節的指示。

4. (選擇性的) 若共享檔案系統的安全性也不足以將 CSP 安全性資訊維持在一個執行常駐程式可存取的地方，您就必須將包含常駐程式私密金鑰和隨機檔案的目錄傳輸到**執行主機**。

- a. 做為**主要主機**上的 root，請輸入以下指令來準備將私密金鑰複製到您將設定為執行主機的機器上。

```
# umask 077
# cd /
# tar cvpf /var/sgeCA/port6789.tar /var/sgeCA/port6789/default
```

- b. 做為每個**執行主機**上的 root，請輸入以下指令來複製檔案。

```
# umask 077
# cd /
# scp <主要主機>:/var/sgeCA/port6789.tar .
# umask 022
# tar xvpf /port6789.tar
# rm /port6789.tar
```

- c. 輸入以下指令驗證檔案權限。

```
# ls -lR /var/sgeCA/port6789/
```

輸出應該類似程式範例 2-6 中的範例。

#### 程式範例 2-6 檔案權限驗證

```
/var/sgeCA/port6789/:
total 2
drwxr-xr-x  4 eddy      other      512 Mar  6 10:52 default
/var/sgeCA/port6789/default:
total 4
drwx----- 2 eddy      staff      512 Mar  6 10:53 private
drwxr-xr-x  4 eddy      staff      512 Mar  6 10:54 userkeys
/var/sgeCA/port6789/default/private:
total 8
-rw-----  1 eddy      staff      887 Mar  6 10:53 cakey.pem
-rw-----  1 eddy      staff      887 Mar  6 10:53 key.pem
-rw-----  1 eddy      staff     1024 Mar  6 10:54 rand.seed
-rw-----  1 eddy      staff      761 Mar  6 10:53 req.pem
/var/sgeCA/port6789/default/userkeys:
total 4
dr-x----- 2 eddy      staff      512 Mar  6 10:54 eddy
dr-x----- 2 root      staff      512 Mar  6 10:54 root
/var/sgeCA/port6789/default/userkeys/eddy:
total 16
-r-----  1 eddy      staff     3811 Mar  6 10:54 cert.pem
-r-----  1 eddy      staff      887 Mar  6 10:54 key.pem
-r-----  1 eddy      staff     2048 Mar  6 10:54 rand.seed
-r-----  1 eddy      staff      769 Mar  6 10:54 req.pem
/var/sgeCA/port6789/default/userkeys/root:
total 16
-r-----  1 root      staff     3805 Mar  6 10:54 cert.pem
-r-----  1 root      staff      887 Mar  6 10:54 key.pem
-r-----  1 root      staff     2048 Mar  6 10:53 rand.seed
-r-----  1 root      staff      769 Mar  6 10:54 req.pem
```

d. 請輸入以下指令繼續進行 Sun Grid Engine 安裝。

```
# cd $SGE_ROOT
# ./install_execd -csp
```

e. 請依照以第 27 頁的「如何安裝執行主機」一節的步驟 4 為開頭的安裝指示。

在完成所有其餘的安裝步驟後，請參照第 39 頁的「如何產生使用者證明文件和私密金鑰」一節的指示。

## ▼ 如何產生使用者證明文件和私密金鑰

要使用 CSP 安全性系統，使用者必須能存取使用者特定證明文件和私密金鑰。執行此步驟最方便的方法是建立辨識使用者的文字檔案。

1. 建立並儲存辨識使用者的文字檔案。

使用 `myusers.txt` 檔案的格式，如以下範例所示。（檔案的欄位是 `UNIX_username:Gecos_field:email_address`。）

```
eddy:Eddy Smith:eddy@my.org
sarah:Sarah Miller:sarah@my.org
leo:Leo Lion:leo@my.org
```

2. 做為主要主機上的 `root`，請輸入以下指令。

```
# $SGE_ROOT/util/sgeCA/sge_ca -usercert myusers.txt
```

3. 請輸入以下指令確認。

```
# ls -l /var/sgeCA/port6789/default/userkeys
```

此目錄列表應該產生跟以下類似的輸出。

```
dr-x----- 2 eddy  staff          512 Mar  5 16:13 eddy
dr-x----- 2 sarah  staff          512 Mar  5 16:13 sarah
dr-x----- 2 leo   staff          512 Mar  5 16:13 leo
```

4. 請跟每個您列於檔案（在本範例是 `myusers.txt`）的使用者輸入下列指令，將安全性相關檔案安裝進 `$HOME/.sge` 目錄中。

```
% source $SGE_ROOT/default/common/settings.csh
% $SGE_ROOT/util/sgeCA/sge_ca -copy
```

使用者應該看到以下的確認事項（在本範例是使用者 `eddy`）。

```
Certificate and private key for user eddy have been installed
```

每個 Sun Grid Engine 的安裝都有安裝對應 `COMMD_PORT` 編號的子目錄。以下的範例，以 `myusers.txt` 檔案為基礎，起因於在輸出前發出指令。

```
% ls -lR $HOME/.sge
/home/eddy/.sge:
total 2
drwxr-xr-x  3 eddy staff      512 Mar  5 16:20 port6789

/home/eddy/.sge/port6789:
total 2
drwxr-xr-x  4 eddy staff      512 Mar  5 16:20 default

/home/eddy/.sge/port6789/default:
total 4
drwxr-xr-x  2 eddy staff      512 Mar  5 16:20 certs
drwx-----  2 eddy staff      512 Mar  5 16:20 private

/home/eddy/.sge/port6789/default/certs:
total 8
-r--r--r--  1 eddy staff      3859 Mar  5 16:20 cert.pem

/home/eddy/.sge/port6789/default/private:
total 6
-r-----  1 eddy staff        887 Mar  5 16:20 key.pem
-r-----  1 eddy staff      2048 Mar  5 16:20 rand.seed
```

## ▼ 如何檢查證明文件

- 依照您要執行的事項，輸入以下一至數個指令。

### 顯示證明文件

以一個字串鍵入下列文字（該指令在本指南中因為太長而無法放到一行內），並在 `-in` 和 `~/ .sge` 元件間加入空格。

```
% $SGE_ROOT/utilbin/$ARCH/openssl x509 -in  
~/ .sge/port6789/default/certs/cert.pem -text
```

### 檢查發出者

以一個字串鍵入下列文字（該指令在本指南中因為太長而無法放到一行內），並在 `-in` 和 `~/ .sge` 元件間加入空格。

```
% $SGE_ROOT/utilbin/$ARCH/openssl x509 -issuer -in  
~/ .sge/port6789/default/certs/cert.pem -noout
```

### 檢查主題

以一個字串鍵入下列文字（該指令在本指南中因為太長而無法放到一行內），並在 `-in` 和 `~/ .sge` 元件間加入空格。

```
% $SGE_ROOT/utilbin/$ARCH/openssl x509 -subject -in  
~/ .sge/port6789/default/certs/cert.pem -noout
```

### 顯示證明文件的電子郵件

以一個字串鍵入下列文字（該指令在本指南中因為太長而無法放到一行內），並在 `-in` 和 `~/ .sge` 元件間加入空格。

```
% $SGE_ROOT/utilbin/$ARCH/openssl x509 -email -in  
~/ .sge/default/port6789/certs/cert.pem -noout
```

## 顯示有效性

以一個字串鍵入下列文字（該指令在本指南中因為太長而無法放到一行內），並在 `-in` 和 `~/ .sge` 元件間加入空格。

```
% $SGE_ROOT/utilbin/$ARCH/openssl x509 -dates -in  
~/ .sge/default/port6789/certs/cert.pem -noout
```

## 顯示指紋

以一個字串鍵入下列文字（該指令在本指南中因為太長而無法放到一行內），並在 `-in` 和 `~/ .sge` 元件間加入空格。

```
% $SGE_ROOT/utilbin/$ARCH/openssl x509 -fingerprint -in  
~/ .sge/port6789/default/certs/cert.pem -noout
```

---

# 驗證安裝

要確認 Sun Grid Engine，企業版常駐程式在執行中，您必須在主要主機然後在執行主機上尋找 `sge_qmaster`、`sge_schedd` 和 `sge_commd` 常駐程式。然後您要嘗試使用 Sun Grid Engine，企業版 5.3 指令並最後準備提交作業。

## ▼ 如何驗證安裝

### 主要主機上

1. 登入主要主機。
2. 根據您執行的作業系統，執行以下指令之一。
  - a. 在 BSD 式 UNIX 系統上，輸入以下指令。

```
% ps -ax
```

- b. 在執行 UNIX System 5 式之作業系統的系統上（如 Solaris 作業環境），輸入以下指令。

```
% ps -ef
```

3. 透過輸出尋找類似以下範例的 sge 字串。

在 BSD 式 UNIX 系統上，您應該可以看到與下面類似的輸出。

```
14673 p1 S < 2:12 /gridware/sge/bin/solaris/sge_commd
14676 p1 S < 4:47 /gridware/sge/bin/solaris/sge_qmaster
14678 p1 S < 9:22 /gridware/sge/bin/solaris/sge_schedd
```

若是 UNIX System 5 式的系統，您應該可以看到與下面類似的輸出。

```
root 439 1 0 Jun 2 ? 3:37 /gridware/sge/bin/solaris/sge_commd
root 439 1 0 Jun 2 ? 3:37 /gridware/sge/bin/solaris/sge_qmaster
root 446 1 0 Jun 2 ? 3:37 /gridware/sge/bin/solaris/sge_schedd
```

若您沒看到合適的字串，則主要主機上所需之一或數個 Sun Grid Engine，企業版常駐程式就沒有在這機器上執行（不管您是否真的在主要主機上，您都可以查閱檔案 <sge 根目錄>/<儲存格>/common/act\_qmaster）。前進至下一步。

4. （選擇性的）手動重新啟動常駐程式。

要取得如何繼續進行的指示，請參閱第 131 頁的「關於常駐程式和主機」一節。

## 執行主機上

1. 登入您執行 Sun Grid Engine，企業版執行主機安裝程序的執行主機。
2. 請在指令主機程序參考步驟 2，以決定適合您系統的 ps 指令，然後輸入該指令。
3. 在輸出中尋找 sge 字串。

在 BSD 式 UNIX 系統上，您應該可以看到與下面類似的輸出。

```
14685 p1 S < 1:13 /gridware/sge/bin//sge_commd
14688 p1 S < 4:27 /gridware/sge/bin/solaris/sge_execd
```

若是 UNIX System 5 式的系統，如 Solaris 作業環境，您應該可以看到與下面類似的輸出。

```
root 169 1 0 Jun 22 ? 2:04 /gridware/sge/bin/solaris/sge_commd
root 171 1 0 Jun 22 ? 7:11 /gridware/sge/bin/solaris/sge_execd
```

如果您沒看到類似輸出，執行主機上所需的一或數個常駐程式就沒有在執行。前進至下一步。

#### 4. (選擇性的) 手動重新啟動常駐程式。

要取得如何繼續進行的指示，請參閱第 131 頁的「關於常駐程式和主機」一節。

### 試驗指令

若兩個必要的常駐程式在主要和執行主機上執行，則 Sun Grid Engine，企業版系統應該為選擇性的。發出試驗指令檢查。

#### 1. 登入主要主機或另一個管理主機。

確定將您安裝 Sun Grid Engine，企業版二進位的路徑加入您的標準搜尋路徑中。

#### 2. 從指令行輸入指令。

```
% qconf -sconf
```

這個 qconf 指令顯示目前的全域叢集配置（請參閱第 145 頁的「基礎叢集配置」一節）。若該指令失敗，則最可能是您的 SGE\_ROOT 環境變數設定不適當或 qconf 未能聯絡與 sge\_qmaster 相關的 sge\_commd。前進至下一步。

#### 3. 檢查是否 script 檔，<sge 根目錄>/<儲存格>/common/settings.csh 或

<sge 根目錄>/<儲存格>/common/settings.sh 設定環境變數，COMMD\_PORT。

如果是這樣，請確定環境變數 COMMD\_PORT 已經設為您再度嘗試上述指令前的特別數值。如果 COMMD\_PORT 變數沒有用在設定檔中，則從您執行指令之機器上的服務資料庫（如，/etc/services 或是 NIS 服務對映）必須提供一個 sge\_commd 項目。若不是這回事，請將這樣的項目新增至機器的服務資料庫中，並給予其配置在 Sun Grid Engine，企業版主要主機上的數值，然後繼續進行至下一步。

#### 4. 重試 qconf 指令。

### 準備提交作業

在您開始提交批次 script 至 Sun Grid Engine，企業版系統前，請檢查您站點的標準及您個人的 shell 資源檔（.cshrc、.profile 或 .kshrc）是否包含指令，如 stty（批次作業沒有預設的終端機連線，因此對 stty 的呼叫會造成錯誤）。

1. 登入主要主機。
2. 輸入以下指令。

```
% rsh 執行主機 date
```

*執行主機* 表示其中一個您將使用之已安裝的執行主機（您應該檢查所有的執行主機看看您的登入或初始目錄是否各個主機都不同）。`rsh` 指令應該給您一個非常類似於在主要主機上本端執行之 `date` 指令的輸出。若有任何額外包含錯誤訊息的行列，您必須消除錯誤的原因才能成功執行批次作業。

對所有的指令解譯器來說，您可以在執行指令，如 `stty`，之前，檢查實際的終端機連線。以下為如何執行的 Bourne-/Korn-Shell 範例：

```
tty -s
if [ $?= 0 ]; then
    stty erase ^H
fi
```

C-Shell 語法非常類似：

```
tty -s
if ( $status = 0 ) then
    stty erase ^H
endif
```

3. 提交包含在 `<sgc 根目錄>/examples/jobs` 目錄其中一個樣本 script。
- 輸入以下指令。

```
% qsub script 路徑
```

4. 使用 Sun Grid Engine，企業版 `qstat` 指令監視作業的運作模式。  
要取得更多提交和監視批次作業的資訊，請參閱第 69 頁的「提交批次作業」。
5. 在作業完成執行後，請在初始目錄檢查重新導向的 `stdout/stderr` 檔案，`<script 名稱>.e<作業id>` 和 `<script 名稱>.o<作業id>`，而 `<作業id>` 為指定給每個作業之連續獨一無二的整數。

如果有任何問題，請參閱第十一章第 261 頁的「錯誤訊息與疑難排解」。



## 部分 III 使用 Sun Grid Engine 企業版 5.3 軟體

---

主要針對給使用者使用—也就是，不執行系統管理員工作的人員（請參閱第四部分第 127 頁的「管理」），《*Sun Grid Engine，企業版 5.3 管理和使用者指南*》的這個部分包含三個章節。

- **第三章 – 第 49 頁的「導覽 Sun Grid Engine，企業版」**

本章介紹一些 Sun Grid Engine，企業版 5.3 的基礎，並包含如何列出不同資源的指示。

- **第四章 – 第 63 頁的「提交作業」**

本章提供藉由 Sun Grid Engine，企業版 5.3 系統提交作業的完整指示，並以作業提交「練習」開始，讓您能熟悉此程序。

- **第五章 – 第 101 頁的「檢查點、監視和控制作業」**

本章解釋作業控制的概念並包含完成不同作業控制作業的指示。

在第三部分的每個章節包含透過 Sun Grid Engine，企業版 5.3 系統完成各種作業的背景資訊和詳盡指示。



# 導覽 Sun Grid Engine，企業版

---

本章介紹一些協助您使用本軟體的基礎 Sun Grid Engine，企業版 5.3 概念和專有名詞。要取得完整的產品背景資訊，包括綜合的字彙，請參閱第一章第 1 頁的「Sun Grid Engine，企業版 5.3 簡介」。

本章也包含完成以下作業的指示。

- 第 51 頁的「如何啓動 QMON 瀏覽器」
- 第 52 頁的「如何顯示佇列清單」
- 第 52 頁的「如何顯示佇列屬性」
- 第 55 頁的「如何找出主要主機名稱」
- 第 55 頁的「如何顯示執行主機清單」
- 第 55 頁的「如何顯示管理主機清單」
- 第 56 頁的「如何顯示提交主機清單」
- 第 57 頁的「如何顯示可要求屬性清單」

---

## Sun Grid Engine，企業版使用者種類和操作

使用者種類在 Sun Grid Engine，企業版分爲四個類別。

- **管理員** — 管理員可全權操作 Sun Grid Engine，企業版。根據預設值，所有管理主機的超級使用者都有管理員權限。
- **操作員** — 操作員可以執行許多和管理員相同的指令，除了進行配置變更，如新增、刪除或修改佇列之外。
- **所有人** — 佇列所有人被允許暫停或啓動他們所擁有的佇列或作業，但沒有更進一步的管理權限。
- **使用者** — 使用者有特定權限，請參見第 59 頁的「使用者存取權限」的說明，但沒有叢集或佇列管理的功能。

表 3-1 顯示不同使用者類別可用的 Sun Grid Engine，企業版 5.3 指令功能。

表 3-1 使用者類別和相關指令功能

| 指令      | 管理員 | 操作員         | 所有人          | 使用者        |
|---------|-----|-------------|--------------|------------|
| qacct   | 完整的 | 完整的         | 僅擁有作業        | 僅擁有作業      |
| qalter  | 完整的 | 完整的         | 僅擁有作業        | 僅擁有作業      |
| qconf   | 完整的 | 無系統設定<br>修改 | 僅顯示配置和存取權限   | 僅顯示配置和存取權限 |
| qdel    | 完整的 | 完整的         | 僅擁有作業        | 僅擁有作業      |
| qhold   | 完整的 | 完整的         | 僅擁有作業        | 僅擁有作業      |
| qhost   | 完整的 | 完整的         | 完整的          | 完整的        |
| qlogin  | 完整的 | 完整的         | 完整的          | 完整的        |
| qmod    | 完整的 | 完整的         | 僅擁有作業和所擁有的佇列 | 僅擁有作業      |
| qmon    | 完整的 | 無系統設定<br>修改 | 無配置<br>變更    | 無配置變更      |
| qrexec  | 完整的 | 完整的         | 完整的          | 完整的        |
| qselect | 完整的 | 完整的         | 完整的          | 完整的        |
| qsh     | 完整的 | 完整的         | 完整的          | 完整的        |
| qstat   | 完整的 | 完整的         | 完整的          | 完整的        |
| qsub    | 完整的 | 完整的         | 完整的          | 完整的        |

## 佇列和佇列屬性

爲了能在您的站點以最理想狀態利用 Sun Grid Engine，企業版系統，您應該熟悉佇列結構和配置給您 Sun Grid Engine，企業版系統的佇列屬性。

## QMON 瀏覽器

Sun Grid Engine，企業版主要功能是圖形使用者介面 (GUI) 指令工具，QMON 瀏覽器。QMON 瀏覽器提供各種的 Sun Grid Engine，企業版功能，包含作業提交、作業控制和重要資訊蒐集。

## ▼ 如何啓動 QMON 瀏覽器

- 從指令行輸入以下指令

```
% qmon
```

在顯示訊息視窗之後，會出現 QMON 主控制面板，其與以下圖示類似（請參閱圖 1-4 來辨識圖示的意義）。

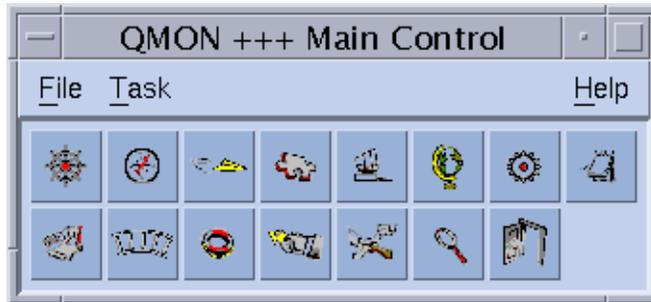


圖 3-1 QMON 主要控制功能表

本手冊的許多指示需要使用 QMON 瀏覽器。圖示按鈕的名稱，其說明了它們的功能，在您的滑鼠游標經過時會出現在螢幕上。

（要取得如何自訂 QMON 瀏覽器的指示，請參閱第 11 頁的「自訂 QMON」。）

### 「佇列控制」 QMON 對話方塊

第 120 頁的「如何用 QMON 控制佇列」一節所顯示和說明的 QMON 「佇列控制」對話方塊提供已安裝的佇列及其目前狀態的快速簡介。

## ▼ 如何顯示佇列清單

- 輸入以下指令。

```
% qconf -sql
```

## ▼ 如何顯示佇列屬性

您可以使用 QMON 或指令行顯示佇列屬性。

### 使用 QMON 瀏覽器

1. 從主要 QMON 功能表，按一下「瀏覽器」圖示。
2. 按一下「佇列」按鈕。
3. 在「佇列控制」對話中，將滑鼠游標移到圖示之上以取得合適的佇列。  
圖 3-2 是顯示的佇列屬性資訊的部分範例。

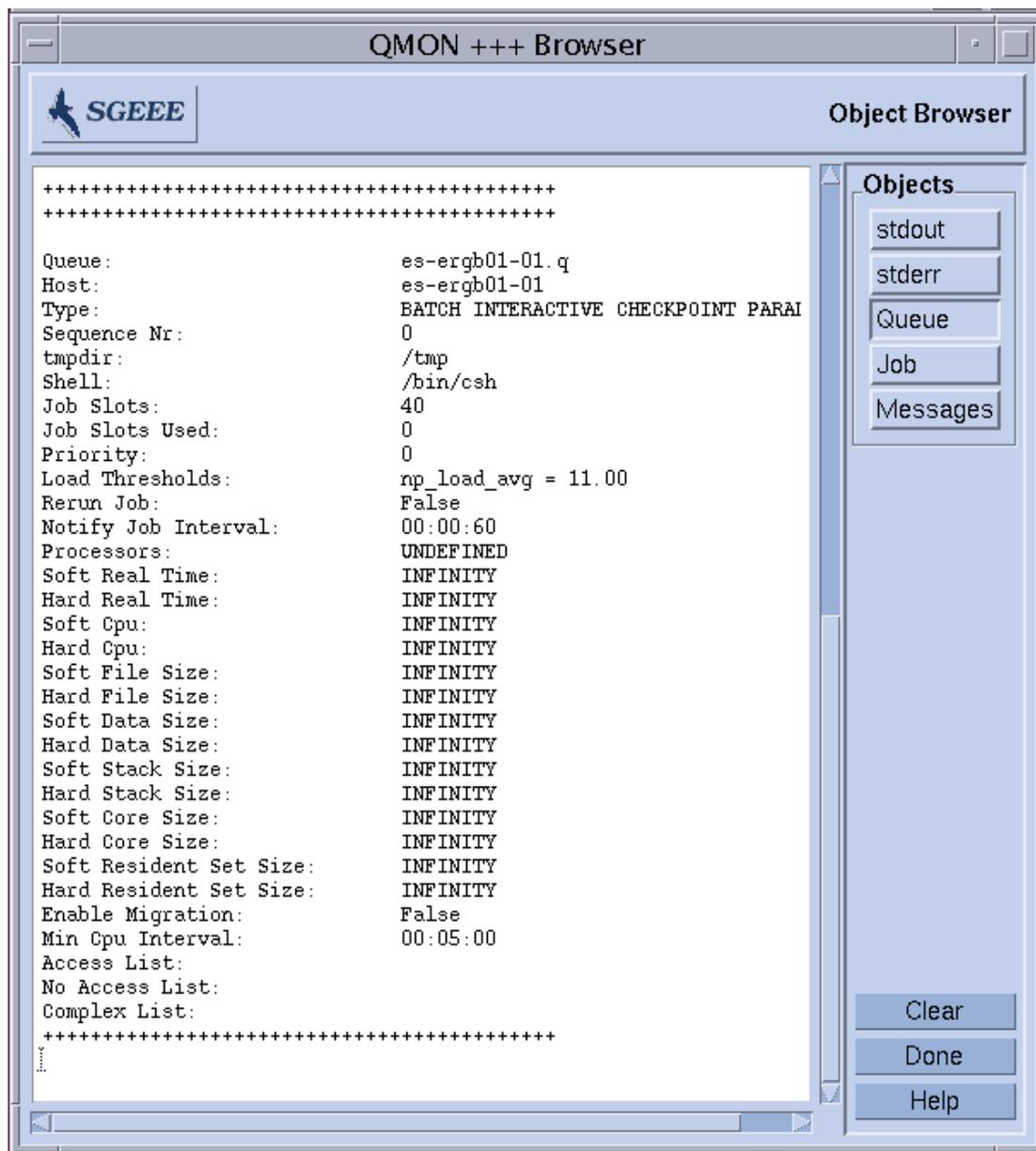


圖 3-2 QMON 瀏覽器的佇列屬性顯示

## 從指令行

- 輸入以下指令。

```
% qconf -sq 佇列名稱
```

螢幕出現類似顯示於圖 3-2 的資訊。

## 解譯佇列屬性資訊

您可以在 `queue_conf` 線上說明手冊和《*Sun Grid Engine 5.3 和 Sun Grid Engine，企業版 5.3 參考手冊*》的 `queue_conf` 一節找到每個佇列屬性的詳細說明。

以下是一些最重要的參數的清單。

- `qname` — 要求的佇列名稱。
- 主機名稱 — 佇列的主機。
- 處理器 — 多重處理器系統的處理器，佇列對其有存取權。
- `qtype` — 允許在此作業執行的作業種類。目前，此為批次、互動、檢查點、平行或以上的任何組合或是傳輸
- 插槽 — 可以同時在該佇列執行的作業數目。
- 所有人清單 — 如第 61 頁的「管理員、操作員和所有人」一節所解釋的佇列所有人
- 使用者清單 — 列在此參數的使用者存取清單（請參閱第 59 頁的「使用者存取權限」）中的使用者或群組辨識器可存取佇列。
- `x` 使用者清單 — 列在此參數的使用者存取清單（請參閱第 59 頁的「使用者存取權限」）中的使用者或群組辨識器無法存取佇列。
- 專案清單 — 使用列於此參數的專案辨識器（請參閱第 182 頁的「關於專案」）所提交的作業可存取佇列。
- `x` 專案清單 — 使用列於此參數的專案辨識器（請參閱第 182 頁的「關於專案」）所提交的作業無法存取佇列。
- 聚合體 (`complex`) 清單 — 列在此參數的聚合體與此佇列有關，而包含在這些聚合體的屬性組成一套佇列的可要求屬性（請參閱第 56 頁的「可要求的屬性」）。
- 聚合體數值 — 指定為特定聚合體屬性提供給該佇列的容量（請參閱第 56 頁的「可要求的屬性」）。

---

# 主機功能

按一下 QMON 主功能表的「主機配置」紐會顯示您 Sun Grid Engine，企業版叢集中與主機相關的功能簡介。然而，沒有 Sun Grid Engine，企業版管理者權限，您可能無法將任何變更套用到呈現的配置。

主機配置對話在第 131 頁的「關於常駐程式和主機」一節有說明。以下各節提供從指令行擷取這類資訊的指令。

## ▼ 如何找出主要主機名稱

主要主機的位置對使用者來說應該是透明的，因為主要主機可能隨時會在目前的主要主機和其中一個備用主要主機之間進行遷徙。

- 使用文字編輯器打開 `<sgc 根目錄>/< 儲存格>/common/act_qmaster` 檔案。目前主要主機的名稱位在此檔案中。

## ▼ 如何顯示執行主機清單

要顯示叢集中配置為執行主機的主機清單，請使用指令：

```
% qconf -sel
% qconf -se 主機名稱
% qhost
```

第一個指令顯示目前所有配置為執行主機的名稱清單。第二個指令顯示指定執行主機的詳細資訊。第三個指令顯示執行主機的狀態和載入資訊。要取得透過 `qconf` 顯示的資訊細節，請參考 `host_conf` 線上說明手冊，而如要取得其輸出和進一步選項的細節，請參考 `qhost` 線上說明手冊。

## ▼ 如何顯示管理主機清單

具管理權限的主機清單可以使用以下指令顯示：

```
% qconf -sh
```

## ▼ 如何顯示提交主機清單

提交主機清單可以使用以下指令顯示：

```
% qconf -ss
```

---

## 可要求的屬性

提交 Sun Grid Engine，企業版作業時，您可指定作業的需求簡介。使用者可指定要使作業順利執行的主機或佇列的屬性或特性。Sun Grid Engine，企業版會將這些作業需求對映到 Sun Grid Engine，企業版叢集的主機和叢集配置，並因此為作業找到合適的主機。

用來定作業需求的屬性與 Sun Grid Engine，企業版叢集（如網路共享磁碟上所需的空間）、主機（如作業系統架構）或佇列（如允許的 CPU 時間）有關，或是屬性是由站點策略衍生而來，如僅在某些主機上安裝軟體的可用性。

可用的屬性包含佇列屬性清單（請參閱第 50 頁的「佇列和佇列屬性」）、全域和主機相關屬性清單（請參閱第 227 頁的「聚合體種類」），以及管理員定義的屬性。但是，為了方便，Sun Grid Engine，企業版管理員通常會選擇只定義所有可要求可用屬性的子集。

目前可要求的屬性顯示於 QMON 提交對話方塊的「要求資源」子對話方塊（請參閱圖 3-3）中（要取得如何提交作業的詳細資訊，請參考第 69 頁的「提交批次作業」一節）。它們列在「可用資源」選擇清單中。

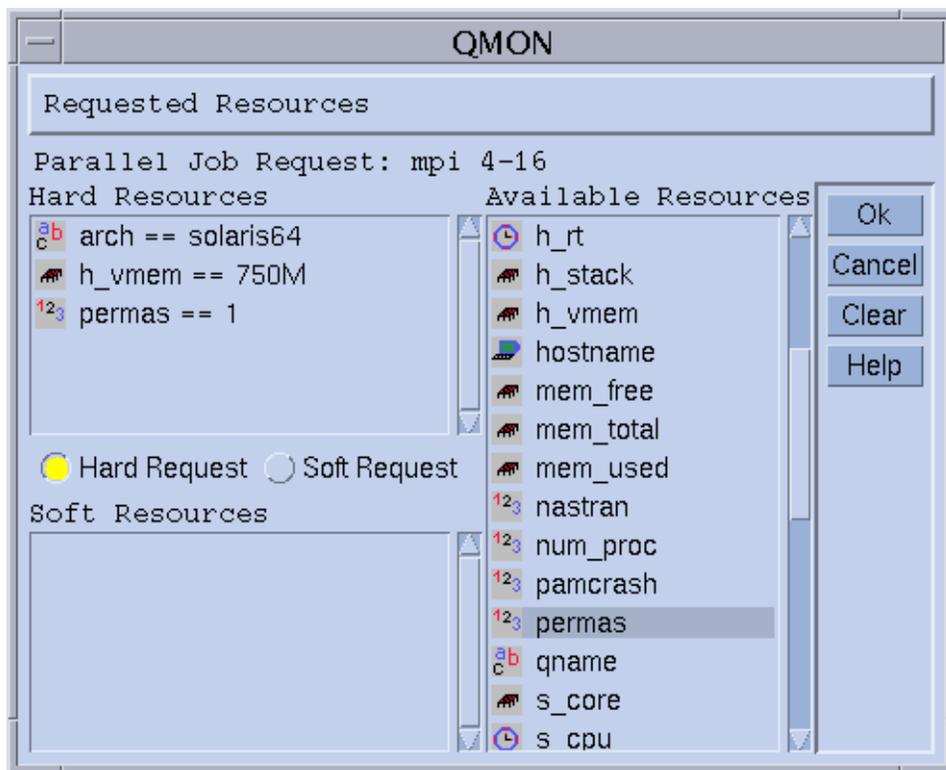


圖 3-3 要求資源對話方塊

## ▼ 如何顯示可要求屬性清單

1. 從指令行輸入以下指令來顯示配置的聚合體清單：

```
% qconf -scl
```

聚合體包含一套屬性的定義。有三個標準的聚合體：

- global – 對（選擇性的）叢集全域屬性來說
- host – 對主機特定屬性來說
- queue – 對佇列屬性的屬性來說

任何因上述指令而列印出之進一步的聚合體名稱指的是管理員定義的聚合體（請參閱第八章第 225 頁的「聚合體概念」，或《Sun Grid Engine 5.3 和 Sun Grid Engine，企業版 5.3 參考手冊》的聚合體形式說明，以取得更多聚合體的資訊）。

2. 以下指令顯示特定聚合體的屬性。

```
% qconf -sc 聚合體名稱 [...]
```

佇列聚合體的輸出可能看起來如表 3-2 所顯示。

表 3-2 顯示的 queue 聚合體屬性

| #Name            | Shortcut | Type   | Value   | Relop | Requestable | Consumable | Default |
|------------------|----------|--------|---------|-------|-------------|------------|---------|
| qname            | q        | STRING | NONE    | ==    | YES         | NO         | NONE    |
| hostname         | h        | HOST   | unknown | ==    | YES         | NO         | NONE    |
| tmpdir           | tmp      | STRING | NONE    | ==    | NO          | NO         | NONE    |
| calendar         | c        | STRING | NONE    | ==    | YES         | NO         | NONE    |
| priority         | pr       | INT    | 0       | >=    | NO          | NO         | 0       |
| seq_no           | seq      | INT    | 0       | ==    | NO          | NO         | 0       |
| resun            | re       | INT    | 0       | ==    | NO          | NO         | 0       |
| s_rt             | s_rt     | TIME   | 0:0:0   | <=    | NO          | NO         | 0:0:0   |
| h_rt             | h_rt     | TIME   | 0:0:0   | <=    | YES         | NO         | 0:0:0   |
| s_cpu            | s_cpu    | TIME   | 0:0:0   | <=    | NO          | NO         | 0:0:0   |
| h_cpu            | h_cpu    | TIME   | 0:0:0   | <=    | YES         | NO         | 0:0:0   |
| s_data           | s_data   | MEMORY | 0       | <=    | NO          | NO         | 0       |
| h_data           | h_data   | MEMORY | 0       | <=    | YES         | NO         | 0       |
| s_stack          | s_stack  | MEMORY | 0       | <=    | NO          | NO         | 0       |
| h_stack          | h_stack  | MEMORY | 0       | <=    | NO          | NO         | 0       |
| s_core           | s_core   | MEMORY | 0       | <=    | NO          | NO         | 0       |
| h_core           | h_core   | MEMORY | 0       | <=    | NO          | NO         | 0       |
| s_rss            | s_rss    | MEMORY | 0       | <=    | NO          | NO         | 0       |
| h_rss            | h_rss    | MEMORY | 0       | <=    | YES         | NO         | 0       |
| min_cpu_interval | mci      | TIME   | 0:0:0   | <=    | NO          | NO         | 0:0:0   |
| max_migr_time    | mmt      | TIME   | 0:0:0   | <=    | NO          | NO         | 0:0:0   |
| max_no_migr      | mmn      | TIME   | 0:0:0   | <=    | NO          | NO         | 0:0:0   |

`name` 一欄基本上和 `qconf -sq` 指令所顯示的第一欄一樣。佇列屬性包含大部分的 Sun Grid Engine，企業版佇列屬性。`shortcut` 一欄包含第一欄完整名稱的管理員可定義縮寫。使用者可在 `qsub` 指令的要求選項中填入完整名稱或捷徑。

`requestable` 一欄會告知對應的項目是否在 `qsub` 使用。因此，例如管理員只需將 `qname` 和 / 或 `qhostname` 項目設為不可要求就可以禁止叢集的使用者直接為其作業要求某機器 / 佇列。這麼做暗示簡單的使用者要求一般可以透過多重佇列達成，其會執行 Sun Grid Engine，企業版的載入平衡功能。

`relop` 一欄定義所用的關係操作以計算佇列是否符合使用者要求。執行的對照 (comparison) 為：

■ `User_Request relop Queue/Host/...-Property`

如果對照的結果是錯的，則使用者的作業不能在考慮的佇列中執行。例如，讓佇列 `q1` 配置 100 秒的軟 `cpu` 時間限制（要取得使用者過程限制的說明，請參閱 `queue_conf` 和 `setrlimit` 線上說明手冊），而佇列 `q2` 配置提供 1000 的軟 `cpu` 時間限制。

`consumables` 和 `default` 欄位對管理員是有意義的，可讓其宣告所謂的可消耗資源（請參閱第 234 頁的「可耗損資源」一節）。使用者會像其他屬性要求消耗品。然而，資源的 Sun Grid Engine，企業版內部簿記 (bookkeeping) 是不同的。

假設使用者提交以下要求。

```
% qsub -l s_cpu=0:5:0 nastran.sh
```

`s_cpu=0:5:0` 要求（有關語法的細節，請參閱 `qsub` 線上說明手冊）會要求至少授與 5 分鐘軟限制 `cpu` 時間的佇列。因此，只有至少提供 5 分鐘軟 CPU 執行時間限制的佇列才能正確配置以執行作業。

---

**注意** – 如果有一個以上的佇列可以執行作業，則 Sun Grid Engine，企業版只會考慮排程過程中的工作量資訊。

---

## 使用者存取權限

Sun Grid Engine，企業版管理員可以限制某些使用者或使用者群組存取佇列或其他 Sun Grid Engine，企業版功能（如，平行環境介面；請參閱第 251 頁的「有關平行環境」）。

---

**注意** – Sun Grid Engine，企業版會自動考慮由叢集管理配置的存取限制。以下各節在您想要查詢個人的存取權限時才會對您有重要性。

---

爲了限制存取權限，管理員建立並維護所謂的存取清單（或簡稱 *ACL*）。*ACL* 包含仲裁使用者和 UNIX 群組名稱。然後，在佇列或平行環境介面配置中將 *ACL* 加入存取允許或存取拒絕清單（請分別參閱《*Sun Grid Engine 5.3 和 Sun Grid Engine，企業版 5.3 參考手冊*》第五節的 *queue\_conf* 或 *sgc\_pe*）。

屬於列於存取允許清單的 *ACL* 之使用者有存取佇列或平行環境介面的權限。做爲存取拒絕清單中 *ACL* 的成員之使用者無法存取相關的資源。

*ACL* 也用來定義 Sun Grid Engine，企業版專案，而對應的使用者可以存取這些專案，例如，使用者可以將其作業從屬於這些專案。管理者可以用依照各個專案限制叢集資源的存取。

透過 QMON 主功能表的「使用者配置」圖示鈕開啓的「UserSet Configuration」對話方塊可以讓您透過「UserSet Configuration」對話方塊查詢您擁有存取權的 *ACL*。相關細節請參考第 169 頁的「管理使用者存取與策略」的第九章。

Sun Grid Engine，企業版專案存取可以透過 QMON 主功能表的「專案配置」圖示顯示。相關細節在第 182 頁的「關於專案」一節有說明。

您可用以下的指令從指令行取得目前配置的 *ACL* 清單：

```
% qconf -sul
```

在一或多重存取清單的項目可用以下指令列印出來：

```
% qconf -su acl 名稱[...]
```

*ACL* 由使用者帳號名稱和以字首「@」符號辨識的 UNIX 群組名稱之 UNIX 群組名稱所組成。如此您可以決定您的帳號要屬於哪個 *ACL*。

---

**注意** – 如果您有使用 *newgrp* 指令交換主 UNIX 群組的權限，則您的存取權可能會改變。

---

您現在可以檢查您有存取權或存取權被拒絕的佇列或平行環境介面。請查詢第 50 頁的「佇列和佇列屬性」和第 252 頁的「如何以 QMON 配置 PE」所述的佇列或平行環境介面配置。存取允許清單名爲 *user\_lists*。存取拒絕清單名爲 *xuser\_lists*。若您的使用者帳號或主要 UNIX 群組和存取允許清單有關，則您被允許存取相關資源。若您與存取拒絕清單有關，則您無法存取佇列或平行環境介面。若兩個清單都是空的話，則每個擁有有效帳號的使用者都可以存取相關資源。

Sun Grid Engine，企業版專案配置可以透過以下指令從指令行控制：

```
% qconf -sprjl  
% qconf -sprj <專案名稱>
```

分別顯示已定義的專案和特別的專案配置清單的清單。專案透過 ACL 定義，而您將依上面所述查詢 ACL 配置。

若您可以存取專案，您就被允許提交從屬於專案的作業。您可以從指令行執行，需透過：

```
% qsub -p < 專案名稱 > < 其他選項 >
```

叢集、主機和佇列配置定義專案存取的方式和透過 `project_lists` 和 `xproject_lists` 參數定義 ACL 的存取是一樣的。

## 管理員、操作員和所有人

Sun Grid Engine，企業版管理員清單可以透過以下方式取得：

```
% qconf -sm
```

操作員清單則可透過以下方式取得：

```
% qconf -so
```

---

**注意** – Sun Grid Engine，企業版管理主機的超級使用者被視為預設的管理員。

---

使用者，其為某個佇列的所有人，包含在第 50 頁的「佇列和佇列屬性」一節所述的佇列配置資料庫。要取得該資料庫您可執行：

```
% qconf -sq 佇列名稱
```

相關的佇列配置項目稱為 `owners`。



# 提交作業

---

本章提供使用 Sun Grid Engine，企業版 5.3 提交作業以供處理的相關背景資訊和指示。本章以執行簡單作業的範例開始，然後會繼續介紹執行複雜作業的指令。

本章包含完成以下作業的指令。

- 第 64 頁的「如何從指令行執行簡單作業」
- 第 65 頁的「如何提交作業，從圖形使用者介面，QMON」
- 第 85 頁的「如何從指令行提交作業」
- 第 87 頁的「如何從指令行提交陣列作業」
- 第 87 頁的「如何使用 QMON 提交陣列作業」
- 第 89 頁的「如何使用 QMON 提交互動作業」
- 第 92 頁的「如何使用 qsh 提交互動作業」
- 第 92 頁的「如何使用 qllogin 提交互動作業」
- 第 93 頁的「如何 qrsh 使用呼叫透明遠端執行」

---

## 執行簡單作業

使用本節的資訊和指令來熟悉提交 Sun Grid Engine，企業版 5.3 作業的基本程序。

---

**注意** – 若您已在未經授權的帳號 (unprivileged account) 下安裝 Sun Grid Engine, Enterprise Edition 程式，您必須以該特殊使用者的身分登入才能執行作業（詳情請參閱第 19 頁的「先決條件作業」）。

---

## ▼ 如何從指令行執行簡單作業

在執行任何 Sun Grid Engine, Enterprise Edition 指令之前，您必須先正確設定可執行搜尋路徑和其他環境條件。

1. 根據您的指令解譯器，輸入以下兩個指令之一：
  - a. 若您使用 `csh` 或 `tcsh` 做為指令解譯器：

```
% source sge_root_dir/default/common/settings.csh
```

`sge_root_dir` 指定在安裝程序之初，所選取的 Sun Grid Engine, Enterprise Edition root 目錄的位置。

- b. 若您使用 `sh`、`ksh` 或 `bash` 做為指令解譯器：

```
# . sge_root_dir/default/common/settings.sh
```

---

**注意** – 您可以將以上的指令加到 `.login`、`.cshrc` 或 `.profile` 檔中（哪一個都適合），以確保您之後啟動所有的互動階段作業會有正確的 Sun Grid Engine, Enterprise Edition 設定。

---

2. 將以下的簡單作業 `script` 提交到 Sun Grid Engine, Enterprise Edition 叢集中。

您可以在 Sun Grid Engine, Enterprise Edition root 目錄中的 `examples/jobs/simple.sh` 檔案找到以下的作業。

```
#!/bin/sh
#This is a simple example of a Sun Grid Engine batch script
#
# Print date and time
date
# Sleep for 20 seconds
sleep 20
# Print date and time again
date
# End of script file
```

輸入以下指令，其假設 `simple.sh` 是儲存上述 `script` 的 `script` 檔的名稱，而該檔案位於您目前工作的目錄下。

```
% qsub simple.sh
```

`qsub` 指令應該以下列方式確認成功的作業提交。

```
your job 1 ("simple.sh") has been submitted
```

3. 輸入以下指令以擷取作業上的狀態資訊。

```
% qstat
```

您應該收到包含目前所有 Sun Grid Engine, Enterprise Edition 系統所辨識的作業相關的狀態報告，以及每個作業所謂的 *作業 ID*（包含在提交確認中之獨一無二的號碼）、作業 `script` 的名稱、作業所有人、狀態資訊（`r` 表示執行中）、提交或啟動時間，和最後，作業執行的佇列名稱。

如果 `qstat` 指令沒有產生輸出的話，則系統無法實際辨識作業。例如，您的作業可能已經結束。您可以檢查 `stdout` 和 `stderr` 重新導向 (`redirection`) 檔案來控制完成作業的輸出。根據預設值，這些檔案是在執行作業的主機上之作業所有人起始目錄中產生的。這些檔案名稱是由作業 `script` 檔案名稱、附加的點符號，此符號之後的「`o`」表示 `stdout` 檔案、「`e`」表示 `stderr` 檔案，而最後是獨一無二的作業 ID。因此，您可以分別在 `simple.sh.o1` 和 `simple.sh.e1` 的名稱下找到作業的 `stdout` 和 `stderr` 檔案（如果該作業是在新安裝的 Sun Grid Engine，企業版系統上首次執行的話）。

## ▼ 如何提交作業，從圖形使用者介面，QMON

提交和控制 Sun Grid Engine, Enterprise Edition 作業以及取得 Sun Grid Engine, Enterprise Edition 系統簡介的簡單方法就是圖形使用者介面，QMON。相較其他的軟體，QMON 為提交和監視作業的作業提供了一個作業提交功能表和「作業控制」對話方塊。

在指令提示下，鍵入以下指令。

```
% qmon
```

在電腦啟動期間，會顯示一個訊息視窗並出現 QMON 主要功能表。

4. 按「作業控制」鈕的左邊，然後按「提交」鈕。

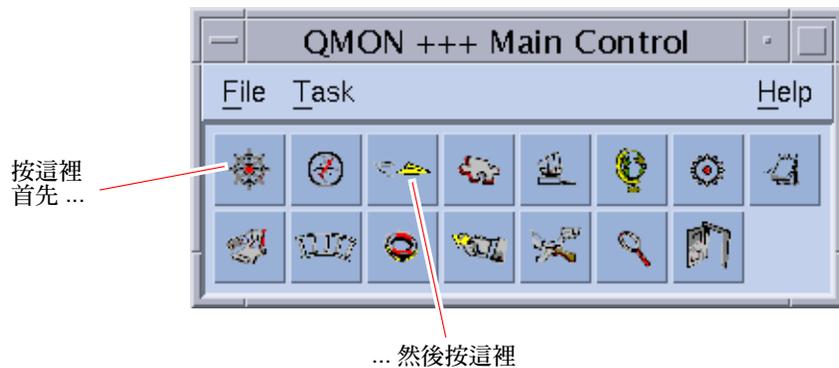


圖 4-1 QMON 主要功能表

會出現「作業提交」和「作業控制」對話方塊（請分別參閱圖 4-2 和圖 4-3）。當您將滑鼠指標移到按鈕上時，畫面就會顯示按鈕名稱（如「作業控制」）。

首先，先按這裡來  
選擇 script 檔案 ...

... 然後按「提交」  
以提交作業。

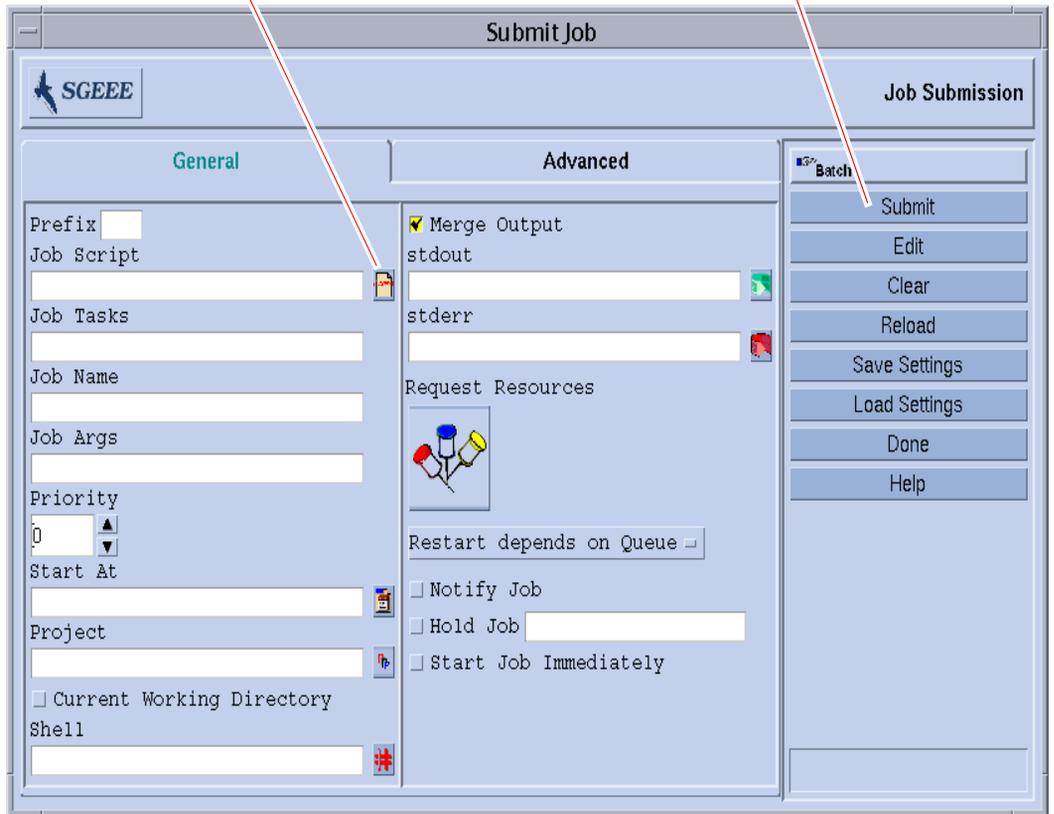


圖 4-2 QMON 作業提交對話方塊

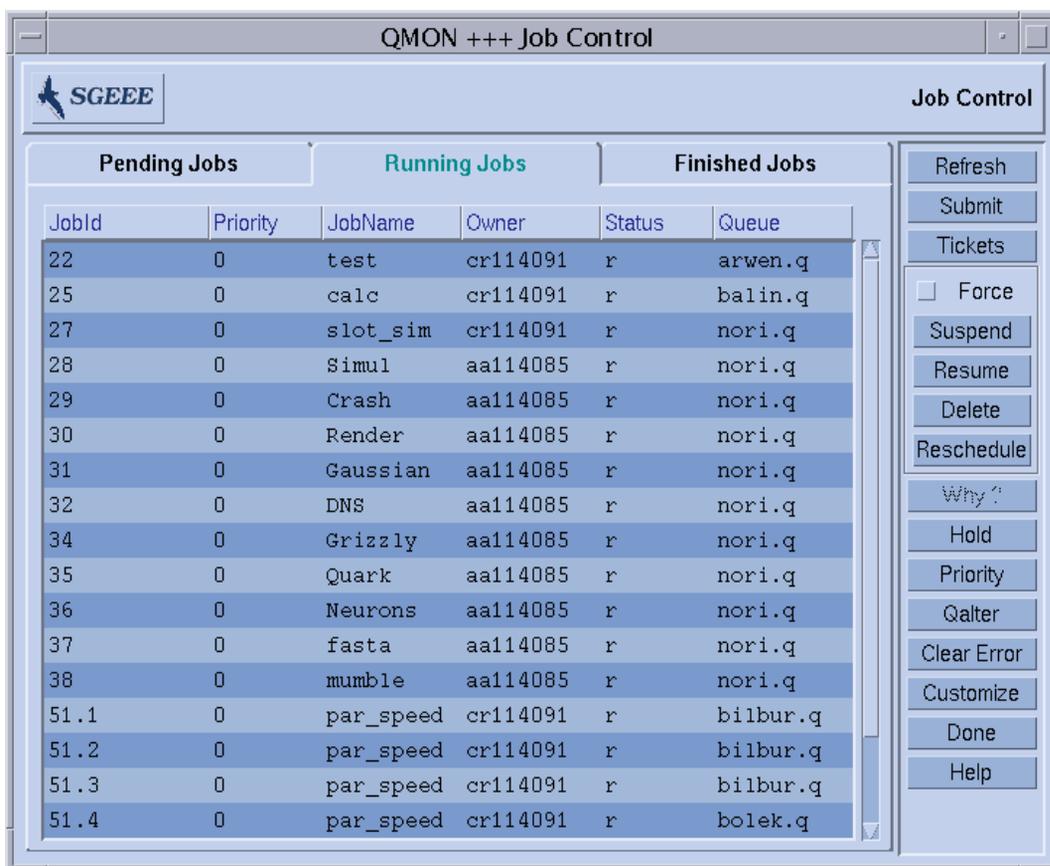


圖 4-3 QMON 作業控制對話方塊

5. 在作業提交功能表中，按「作業 Script」檔案選擇圖示來開啓檔案選擇方塊。  
顯示作業 Script 選擇方塊。

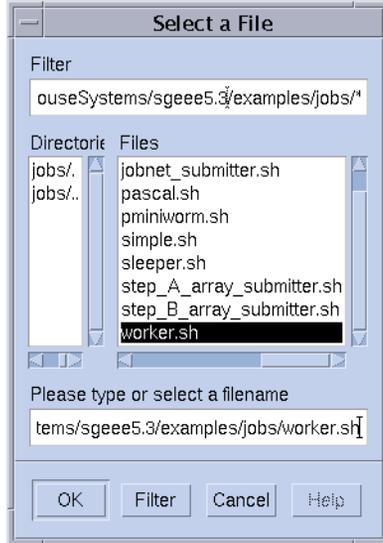


圖 4-4 作業 Script 選擇方塊。

6. 按下適當的檔名，以選擇 script 檔案（例如，來自命令行範例的檔案 *simple.sh*）。
7. 按下「作業提交」功能表底部的「提交」鈕。

過了幾秒後，您應該就可以在「作業控制」畫面監視您的作業了。您會先在「擱置中作業」下看到作業；而當作業啟動時，它會迅速移動到「執行中作業」。

---

## 提交批次作業

以下各節說明如何透過 Sun Grid Engine，企業版 5.3 系統提交較複雜的作業。

### 關於 Shell Script

Shell script，又稱為批次作業，原則是組合在檔案中命令行指令的序列。Script 檔是因 `chmod` 指令變成可以執行。若呼叫指令，就會啟動正確的指令解譯器（例如，`csh`、`tcsh`、`sh` 或 `ksh`），而每個指令都由都解譯為使用者手動鍵入執行 script。您可以從 shell script 中呼叫隨意的指令、應用程式和其他 shell script。

適當的指令解譯器是否呼叫為 `login-shell`，是根據它的名稱（`csh`、`tcsh`、`sh`、`ksh`、...）是否事實上包含在執行作業的特定主機和佇列的 Sun Grid Engine, Enterprise Edition 配置 `login_shells` 選項的價值清單 (value list) 中。

---

**注意** – 配置於叢集的不同主機和佇列的 Sun Grid Engine, Enterprise Edition 配置可能不同。您可以透過 `qconf` 指令的 `-sconf` 和 `-sq` 選項來顯示有效的配置（詳細的資訊請參考《*Sun Grid Engine 5.3 和 Sun Grid Engine, 企業版 5.3 參考手冊*》）。

---

如果指令解譯器被呼叫為 `login-shell`，您作業的環境將和您登入並執行 `script` 的環境完全相同。例如，在使用 `csh` 時，除了系統預設啟動資源檔之外，將執行 `.login` 和 `.cshrc`（就好像 `/etc/login`），但如果 `csh` 不是呼叫為 `login-shell`，則僅會執行 `.cshrc`。是否呼叫為 `login-shell` 的不同處之相關說明，請參考您選擇的指令解譯器的手冊。

## Script 檔的範例

程式範例 4-1 是簡單 shell script 的範例，其會先從 Fortran77 來源編譯應用程式，`flow`，然後加以執行：

```
#!/bin/csh

# This is a sample script file for compiling and
# running a sample FORTRAN program under Sun Grid Engine,
# Enterprise Edition.

cd TEST

# Now we need to compile the program 'flow.f' and
# name the executable 'flow'.

f77 flow.f -o flow
```

### 程式範例 4-1 簡單 Shell Script

您的本端系統使用者指南會提供建立和自訂 shell script 的詳細資訊（您可能也想看看 `sh`、`ksh`、`csh` 或 `tcsh` 線上說明手冊）。在以下各節中，重點是在需要考慮的特性，好為 Sun Grid Engine, Enterprise Edition 準備批次 script。

一般來說，您可以把您要手動從指令提示符號執行的所有 shell script 提交到 Sun Grid Engine, Enterprise Edition，只要他們不需要終端機連接（除了標準錯誤和輸出裝置外，其可以自動重新導向）且不需要互動使用者介入。因此，程式範例 4-1 已經準備好要提交到 Sun Grid Engine, Enterprise Edition，並會執行所要的動作。

# 使用 QMON 提交延伸和進階作業

在嘗試較複雜的作業提交形式前－*延伸*或*進階*－了解一些重要的程序背景資訊是有用的。以下各節提供該資訊。

## 延伸範例

「作業提交」對話方塊的標準形式（請參閱圖 4-2）提供了配置以下延伸作業的參數的方法：

- 用於嵌入式 script Sun Grid Engine, Enterprise Edition 提交選項的字首字串（愈取得詳細資訊請參閱第 82 頁的「作用中的 Sun Grid Engine，企業版註釋」一節）
- 要使用的作業 script  
按下相關檔案鈕會開啓檔案選項方塊（請參閱圖 4-4）。
- 提交陣列作業的作業 ID 範圍（請參閱第 87 頁的「陣列作業」）
- 作業名稱（選擇作業 script 之後就會設定預設值）
- 作業 script 的引數
- 設定作業的初始優先權之計算方塊

在 Sun Grid Engine，企業版中，本優先權會在單一使用者作業中排列順序。其會告訴 Sun Grid Engine，企業版排程器若同時有數個作業在系統內時，如何從中選擇單一使用者作業。

---

**注意**－管理員必須分配票卷功能性策略，並分配共享資源給功能性作業類別，讓使用者能加權自己的作業。

---

- 作業認為適合執行的時候  
若按下相關的檔案按鈕，就會出現對話方塊，以便輸入正確格式的時間（請參閱圖 4-5）
- 作業從屬的 Sun Grid Engine，企業版專案  
輸入欄位旁的按鈕讓您選擇可用的專案（請參閱圖 4-6）。
- 表示作業是否在目前工作的目錄下執行的旗號（僅針對提交和可能的執行主機間相同的目錄階層）
- 用於執行作業 script 的指令解譯器（請參閱第 81 頁的「如何選擇指令解譯器」）  
若按下相關的按鈕，就會出現對話方塊，以便輸入作業的指令解譯器規格（請參閱圖 4-7）。
- 表示作業的標準輸出和標準錯誤輸出是否合併成標準輸出串流的旗號

- 要使用的標準輸出重新導向（請參閱第 82 頁的「輸出重新導向」）  
如果沒指定，將使用預設值。若按下相關的檔案按鈕，就會出現說明者對話方塊，以便輸入輸出重新導向的替代方式（請參閱第 82 頁的「輸出重新導向」）。
- 要使用的標準錯誤輸出重新導向—和標準輸出重新導向非常相似
- 作業的資源需求  
若要定義您作業所需的資源，請按對應的圖像鈕。如果作業要求資源，則圖像鈕的顏色會改變。
- 定義作業在系統當機會類似事件後而中止的作業是否能重新啓動，或重新啓動運作方式是根據佇列或由作業所要求的選擇清單鈕
- 表示作業在即將暫停或取消時，是否作業由 SIGUSR1 或 SIGUSR2 信號通知的旗號
- 表示要分派給作業的使用者保留 (user hold) 或作業相依 (job dependency) 的旗號  
只要被指定任何類型的保留，作業就不適合執行（若要取得更多關於保留的資訊，請參閱第 106 頁的「監視和控制 Sun Grid Engine，企業版 作業」一節）。附接於「保留」旗號的輸入欄位僅將保留限制在特定陣列作業的作業範圍中（請參閱第 87 頁的「陣列作業」）。
- 強迫作業立刻啓動（如果可能的話）或被拒絕的旗號  
如果選擇此旗號，則作業不會排在佇列中。



圖 4-5 在時間輸入方塊中



圖 4-6 專案選擇方塊



圖 4-7 Shell 選擇方塊



圖 4-8 輸出重新導向方塊

「作業提交」螢幕右側的按鈕讓您初始啓動不同的動作：

- **提交** — 依照提交對話方塊的指定來提交作業。
- **修改** — 使用 `vi` 或 `$EDITOR` 環境變數所定義的編輯器修改 X 終端機裡選擇的 script 檔。
- **清除** — 清除在「作業提交」對話方塊中的所有設定值，包括任何的指定資源需求。
- **重新載入** — 重新載入指定的 script 檔、語法解析 (parse) 任何嵌入式 script 選項（請參閱第 82 頁的「作用中的 Sun Grid Engine，企業版註釋」一節）、語法解析預設設定值（請參閱第 86 頁的「預設要求」一節）並刪除這些設定值的中間手動變更。該動作相當於「清除」動作，其規格為前一個 script 檔的規格。若已經選擇 script 檔，則該選項只會顯示一種效果。
- **儲存設定值** — 將目前的設定值儲存到檔案。檔案選擇方塊已開啓，以便選擇檔案。儲存的檔案可能之後會明確載入（請參閱下文）或可能用為預設要求（請參閱第 86 頁的「預設要求」一節）。
- **載入設定值** — 載入之前用「儲存設定值」鈕所儲存的設定值（請參閱上文）。載入的設定值會覆寫目前的設定值。
- **完成** — 關閉「作業提交」對話方塊。
- **說明** — 顯示對話方塊特定說明。

圖 4-9 顯示已設定大部分參數的「作業提交」對話方塊。

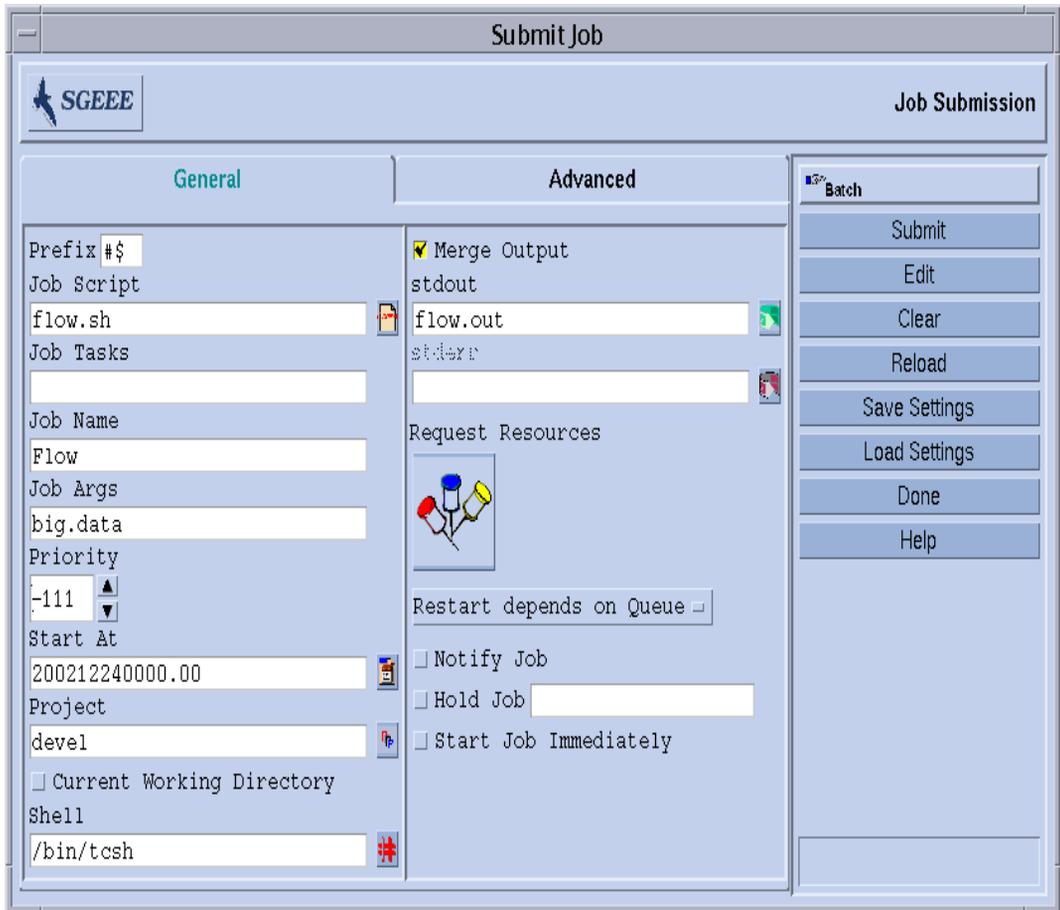


圖 4-9 延伸作業提交範例

配置在範例中的作業擁有 script 檔，`flow.sh`，其必須常駐在 QMON 工作目錄中。此作業叫做 Flow，script 檔採用單一引數，`big.data`。本作業以優先權 `-111` 啟動，並在 2002 年 12 月 24 日午夜前都適合執行。Sun Grid Engine，企業版 特定的專案定義表示作業附屬於專案 `devel`。作業將在提交工作目錄下執行，並將使用 `tcsh` 指令解譯器。最後，標準輸出和標準錯誤輸出將合併進檔案，`flow.out`，該檔案也會建立在目前的工作目錄中。

## 進階範例

「進階」提交螢幕允許以下額外的參數定義：

- 要使用的平行環境介面

- 作業執行前需為作業設定的一套環境變數。  
若按下相關的按鈕，就會出現說明者對話方塊，供要輸出的環境變數定義使用（請參閱圖 4-10）。環境變數能從 QMON 的執行時間環境取得，或可以定義仲裁的環境變數。
- 名為環境 (Context) 的名稱 / 數值配對的清單（請參閱圖 4-11），可以用來儲存並傳送可以從 Sun Grid Engine，企業版叢集中存取的作業相關資訊。  
環境變數可以透過 `-ac/-dc/-sc` 選項至 `qsub`、`qrsh`、`qsh`、`qlogin` 或 `qalter` 從指令行修改，並可透過 `qstat -j` 取得。
- 檢查點適合作業的情況下，要使用的檢查點環境（請參閱第 101 頁的「關於檢查點作業」一節）
- 與作業相關的帳號字串  
帳號字串將加到為作業而做的帳號紀錄，並在之後可用於帳號分析。
- 「驗證」旗號，其判定作業一致性檢查模式  
若要檢查作業要求的一致性，Sun Grid Engine，企業版會採用空白且未載入的叢集，並嘗試至少找出一個作業可執行的佇列。可能的檢查模式為：
  - 略過 — 完全沒有一致性的檢查。
  - 警告 — 雖然有不一致性的報告，但仍接受作業（如果叢集配置在作業提交後進行變更會比較符合要求）。
  - 錯誤 — 有不一致性的報告，如碰上任何作業，該作業會遭拒絕。
  - 剛驗證 (Just verify) — 將不會提交作業，但會產生一份作業對叢集中每個主機和佇列之適合性的廣泛報告。
- 關於使用者是透過電子郵件通知的事件  
目前定義的作業事件為啟動 / 結束 / 中斷 / 暫停。
- 通知郵件寄送的電子郵件位址清單  
如果按下相關按鈕，就會出現說明者對話來定義寄信清單（請參閱圖 4-12）。
- 要求強制選擇做為作業執行之佇列名稱清單。  
在符號清單項目 (bulleted list item)，第 72 頁的「作業的資源需求」，所述的對應資源需求將硬佇列 (Hard Queue) 清單和軟佇列 (Soft Queue) 清單看做同樣的事物。
- 適合做為平行作業的主要佇列的佇列名稱清單。  
平行作業在主要佇列啟動。所有作業產生至平行作業的其他佇列稱為副佇列 (*slave queues*)。
- 需要在啟動提交作業前順利完成的作業之 ID 清單  
剛建立的作業取決於這些作業的順利完成。
- 截止日期作業的截止日期初始啟動時間

截止期限初始啟動會定義在指定的截止期限前，截止期限作業必須達成最高優先權的時間點。建議從截止期限作業的理想截止期限扣除執行時間（於最高的優先權）的保守預估，以決定截止期限初始啟動時間。按一下「截止期限」輸入視窗旁的按鈕可以開啓圖 4-13 所顯示的說明者對話方塊。

---

**注意** – 並非全部的 Sun Grid Engine，企業版使用者都可以提交截止期限作業。請洽詢您的系統管理員您是否有提交截止期限作業的權限。另外有關截止期限作業的最高優先權的資訊，請聯絡叢集管理員。

---



圖 4-10 作業環境定義



圖 4-11 作業上下文定義



圖 4-12 郵件地址規格

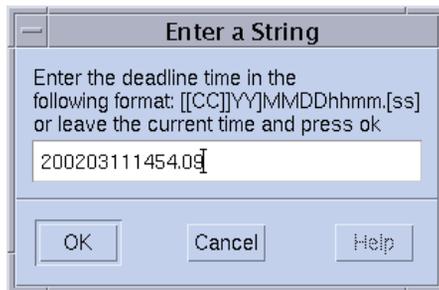


圖 4-13 截止期限時間輸入方塊

定義在圖 4-14 的作業和第 71 頁的「延伸範例」一節的作業定義相比，有以下額外的特性。

- 作業需要使用平行環境 `mpi`。它需要建立至少 4 個平行過程，而最多可以使用 16 個過程（如果有的話）。
- 已設定並輸出兩個作業環境變數。
- 已設定兩個上下文變數。
- 帳號字串 `FLOW` 將新增至作業帳號紀錄中。
- 如果因為系統當機造成作業失敗，則該作業將重新啟動。
- 如果偵測到作業要求和叢集配置間有不一致之處，則應該列印出警告。
- 在作業開始和結束時，郵件應該傳送到兩個電子郵件地址的清單。
- 作業最好是在佇列 `big_q` 中執行。

圖 4-14 顯示進階作業提交的範例。

**Submit Job**

SGE Job Submission

**General**

Parallel Environment  
mpi 4-16

Environment  
CLEAN\_SEMAPHORE=FALSE, MODEL\_S...

Context  
JOB\_STEP=preprocessing, PORT=1...

Checkpoint Object

Account  
FLOW

**Advanced**

Verify Mode  
Warning

Mail  
 Start of Job  
 End of Job  
 Abort of Job  
 Suspend of Job

Mail To  
me@myhost.com, me@another.addr...

Hard Queue List

Soft Queue List  
big\_q

Master Queue List

Job Dependencies

Deadline

**Batch**

Submit  
Edit  
Clear  
Reload  
Save Settings  
Load Settings  
Done  
Help

圖 4-14 進階作業提交範例。

## 資源需求定義

在目前的範例中，使用的提交選項沒有表達任何對執行作業的主機的需求。Sun Grid Engine，企業版會假設這樣的作業可以在任何主機上執行。然而在練習中，大多數的作業要求在執行的主機上能達成一定的先決條件，才能夠順利完成。這樣的先決條件為足夠的可用記憶體、要求軟體的安裝或一定的作業系統架構。另外，叢集管理通常會限制叢集內機器的使用。例如，作業所允許消耗的 CPU 時間通常會有限制。

在使用者對叢集的裝備及其使用策略不了解的情況下，Sun Grid Engine，企業版會提供使用者找出適合使用者作業之主機的方法。使用者必須做的是指定使用者作業的需求，並讓 Sun Grid Engine，企業版管理找尋適合且低負載主機的作業。

資源需求透過第 56 頁的「可要求的屬性」一節所說明的 *可要求屬性* 來指定。QMON 提供一個非常簡便的指定作業需求的方法。在按下「作業提交」對話方塊中的「要求資源」鈕後會開啓「要求資源」對話方塊（請參閱圖 4-15 的範例）只會顯示目前在「可用資源」選擇清單中合適的屬性。在屬性上連按兩下可使該屬性新增至作業的硬或軟（請參閱以下解釋）資源清單中，而且（除了設為 True 的 BOOLEAN 類型屬性之外）會開啓說明者對話方塊，以引導您輸入相關屬性的數值規格。

顯示在圖 4-15 的範例「要求資源」對話方塊作業的資源簡介，其中要求一個 solaris64 主機和至少提供 750 MB 記憶體的可用 `permas` 授權。如果找到一個以上符合此規格的佇列，則任何定義的軟資源需求都算在內（本範例中沒有）。然後，如果沒找到符合硬和軟需求的佇列，任何認可硬資源的佇列都可視為合適的佇列。

---

**注意** - 只有在一個以上的佇列適合作業時，載入標準會決定從哪裡啓動作業。

---

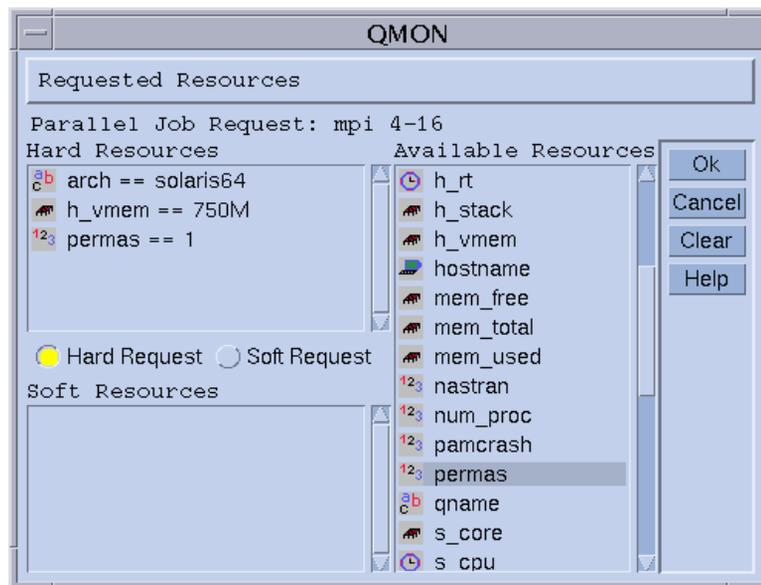


圖 4-15 要求資源對話方塊

---

**注意** – INTEGER 屬性 `permas` 是透過管理員延伸引進到「全域」聚合體，STRING 屬性 `arch` 是從「主機」聚合體匯入的，而 MEMORY 屬性 `h_vmem` 是從「佇列」聚合體匯入的。

---

相等的資源需求簡介也可以從 `qsub` 指令行提交：

```
% qsub -l arch=solaris64,h_vmem=750M,permas=1 \  
permas.sh
```

---

**注意** – 在第一個 `-l` 選項之前的內在 (implicit) `-hard` 開關已被略過。

---

750 MB 的記號 `750M` 是 Sun Grid Engine，企業版數量語法 (quantity syntax) 的範例。對要求記憶體消耗的屬性來說，您可以指定由所謂的乘數由後新增的 (appended) 整數小數 (integer decimal)、浮點小數 (floating point decimal)、八進位整數 (integer octal) 和十六進位整數 (integer hexadecimal)。

- `k` – 將數值乘以 1000。
- `K` – 將數值乘以 1024。
- `m` – 將數值乘以 1000 次 1000。
- `M` – 將數值乘以 1024 次 1024。

八進位常數由開頭的 0 (zero) 和 0 到 7 之間的數字所指定。指定十六進位的常數需要由前新增 (prepend) `0x` 的數字，然後使用 0 到 9、a 到 f 和 A 到 F 等數字。如果沒有乘數由後新增 (append)，則數值是以位元組來計算的。若使用浮點小數 (floating point decimal)，結果的數值在整數後的數值將遭刪除。

對那些有加上時間限制的屬性來說，您可以使用時數、分數或秒數和三者任何的組合來指定時間數值。時數、分數和秒數是以冒號分隔的十進位數字所指定的。`3:5:11` 的時間即表示 11111 秒。如果指定時數、分數或秒數的數字是 0，則該數字在保有冒號的情況下可以留空。因此 `:5:` 的數值即表示 5 分鐘。用於上述的「要求資源」對話方塊的表格是為延伸的部分，只在 `QMON` 有效。

## Sun Grid Engine，企業版如何配置資源

如上一節所示，了解 Sun Grid Engine，企業版軟體如何處理資源要求和如何配置資源是很重要的。以下提供了 Sun Grid Engine，企業版軟體的資源配置演算法的概略檢視。

1. 在所有預設要求檔案中讀取並進行語法分析（請參閱第 86 頁的「預設要求」一節）。
2. 為嵌入式選項處理 `script` 檔（請參閱第 82 頁的「作用中的 Sun Grid Engine，企業版註釋」一節）。

3. 當作業提交時，讀取所有 script 嵌入式選項，不管它們在 script 檔的位置。
4. 從命令行讀取所有的要求並進行語法分析。

一旦收集所有的 `qsub` 要求，硬和軟要求就會分開處理（會先處理硬的）。這些要求會對應以下的優先順序進行評估。

1. script/ 預設要求檔案從左到右
2. script/ 預設要求檔案從上到下
3. 命令行從左到右

換句話說，命令行可以用來置換嵌入式旗號。

已配置要求為硬的資源。如果要求無效，提交將遭拒絕。如果在提交時間無法符合一或多個要求（例如，要求的佇列忙碌中），此作業就會被置於佇列中並在之後重新排程。如果符合所有的硬要求，這些要求就會配置而作業就可以執行。

已檢查要求為軟的資源。即使沒有符合一些或全部的要求，作業仍然可以執行。如果多重佇列（已經符合硬需求）提供部分的軟體資源清單（重疊或不同的部分），Sun Grid Engine，企業版軟體將選擇提供最多軟要求的佇列。

此作業將啟動，並涵蓋配置資源。

使用執行 UNIX 指令如 `hostname` 或 `date` 的小型測試 script 檔來實驗取得引數清單選項和嵌入式選項或硬與軟需求如何影響對方的經驗是有用的。

## 延伸至一般的 Shell Script

如果在 Sun Grid Engine，企業版的控制下執行，則有些延伸部分會到影響 script 運作模式一般 shell script 的部分。以下幾節說明這些延伸部分。

### 如何選擇指令解譯器

用來處理作業 script 檔案的指令解譯器可以在提交時間指定（請參閱，例如，圖 4-9）。然而，如果沒有指定，則配置變數，`shell_start_mode`，會決定如何選擇指定解譯器：

- 如果 `shell_start_mode` 設為 `unix_behavior`，script 檔的第一行—如果以「#!」序列開頭—將評估決定指令解譯器。如果第一行沒有「#!」序列，則根據預設值會使用 Bourne Shell `sh`。
- 對其他的 `shell_start_mode` 設定值來說，使用的是用 `shell` 參數為作業啟動的佇列配置的預設指令解譯器（請參閱第 50 頁的「佇列和佇列屬性」一節和 `queue_conf` 線上說明文件）。

## 輸出重新導向

因為批次作業沒有裝端機連線，所以其標準輸出和標準錯誤輸出必須重新導入檔案中。Sun Grid Engine，企業版允許使用者定義輸出重新導入檔案的位置，但如果沒有指定，就會使用預設值。

檔案的標準位置在目前作業執行的工作目錄中。預設標準輸出檔案名稱爲 < 作業名稱 > .o < 作業 id >，預設標準錯誤輸出重新導向至 < 作業名稱 > .e < 作業 id >。 < 作業名稱 > 是從 script 檔案名稱建立或可以由使用者定義（範例請參閱 qsub 線上說明手冊的 -N 選項）。 < 作業 id > 是由 Sun Grid Engine，企業版指定給作業的獨一無二的辨識器。

如果是陣列作業的作業（請參閱第 87 頁的「陣列作業」一節），作業辨識器會加到以句點符號分隔的檔案名稱中。因此，產生的標準重新導向路徑爲 < 作業名稱 > .o < 作業 id > . < 作業 id > 和 < 作業名稱 > .e < 作業 id > . < 作業 id >。

如果標準位置不適合，使用者可使用圖 4-14 和圖 4-8 所顯示的 QMON 來指定輸出方向，或使用 -e 和 -o qsub 選項。標準輸出和標準錯誤輸出可以合併到一個檔案，而更改方向是可以以執行主機爲單位而指定的。例如，根據作業執行的主機，輸出更改方向的檔案位置也變的不同。要建立自訂但獨一無二的更改方向檔案路徑，虛擬環境變數可用來和 qsub -e 與 -o 選項一起使用。這些變數的清單如下。

- \$HOME — 執行機器上的起始目錄
- \$USER — 作業所有人的使用者 ID
- \$JOB\_ID — 目前的作業 ID
- \$JOB\_NAME — 目前的作業名稱（請參閱 -N 選項）
- \$HOSTNAME — 執行主機名稱
- \$TASK\_ID — 陣列作業作業索引編號

這些變數在作業執行期間擴張至真正的數值，而更新路徑會隨著它們建立。

欲知更進一步的細節，請參閱《Sun Grid Engine 5.3 和 Sun Grid Engine，企業版 5.3 參考手冊》的 qsub 項目。

## 作用中的 Sun Grid Engine，企業版註釋

之前有 # 符號的行被視爲 shell script 裡的註釋。然而，Sun Grid Engine，企業版會辨識特殊的註釋行，並以特別的方式使用它們。這種 script 行的其他部分會被視爲 Sun Grid Engine，企業版提交指令 qsub 的指令行引數清單的一部份。在這些特殊註釋行裡所供應的 qsub 選項也是由 QMON 「作業提交」對話方塊所解譯的，而當選取 script 檔時，對應的參數就會預先設定。

根據預設值的特殊註釋行是靠「#\$」字首字串辨識的。字首字串可以使用 qsub -C 選項重新定義。

說明的機制稱為提交引數的 script 嵌入。以下為使用 script 嵌入指令行選項的 script 檔案範例。

```
#!/bin/csh
#Force csh if not Sun Grid Engine, Enterprise Edition default
#shell
#$ -S /bin/csh
# This is a sample script file for compiling and
# running a sample FORTRAN program under Sun Grid Engine,
# Enterprise Edition.
# We want Sun Grid Engine, Enterprise Edition to send mail
# when the job begins
# and when it ends.
#$ -M EmailAddress
#$ -m b,e
# We want to name the file for the standard output
# and standard error.
#$ -o flow.out -j y
# Change to the directory where the files are located.
cd TEST
# Now we need to compile the program 'flow.f' and
# name the executable 'flow'.
f77 flow.f -o flow
# Once it is compiled, we can run the program.
flow
```

程式範例 4-2 使用 Script 嵌入式指令行選項

## 環境變數

在執行 Sun Grid Engine，企業版作業時，預設到作業環境的變數數目如下所列。

- ARC — 作業執行節點的 Sun Grid Engine，企業版架構名稱；此名稱會編碼進 sge\_execd 二進位
- COMMD\_PORT — 指定 TCP 埠，而 sge\_commd(8) 預料將在其上聽取通訊需求
- SGE\_ROOT — 在啟動或預設 /usr/SGE 之前，Sun Grid Engine，企業版根目錄設為 sge\_execd
- SGE\_CELL — 作業執行的 Sun Grid Engine，企業版儲存格
- SGE\_JOB\_SPOOL\_DIR — sge\_shepherd(8) 所用來在作業執行期間儲存作業相關資料的目錄

- SGE\_O\_HOME – 作業提交主機上之作業所有人的起始目錄
  - SGE\_O\_HOST – 作業提交的主機
  - SGE\_O\_LOGNAME – 作業提交主機上之作業所有人的登入名稱
  - SGE\_O\_MAIL – 在作業提交指令上下文中 MAIL 環境變數的內容
  - SGE\_O\_PATH – 在作業提交指令上下文中 PATH 環境變數的內容
  - SGE\_O\_SHELL – 在作業提交指令上下文中 SHELL 環境變數的內容
  - SGE\_O\_TZ – 在作業提交指令上下文中 TZ 環境變數的內容
  - SGE\_O\_WORKDIR – 作業提交指令的工作目錄
  - SGE\_CKPT\_ENV – 指定檢查點作業執行的檢查點環境（如使用 `qsub -ckpt` 選項選擇一般）
  - SGE\_CKPT\_DIR – 只為檢查點作業設定；包含檢查點介面的路徑 `ckpt_dir`（請參閱 checkpoint 線上說明手冊）
  - SGE\_STDERR\_PATH – 作業的標準錯誤串流轉移的檔案路徑名稱；一般用來增加來自 `prolog`、`epilog`、平行環境啟動 / 停止或檢查點 `script` 的錯誤訊息的輸出
  - SGE\_STDOUT\_PATH – 作業的標準錯誤串流轉移的檔案路徑名稱；一般用來增加來自 `prolog`、`epilog`、平行環境啟動 / 停止或檢查點 `script` 的訊息的輸出
  - SGE\_TASK\_ID – 此作業代表的陣列作業裡的作業辨識器
  - ENVIRONMENT – 請每次都設為 `BATCH`；此變數表示 `script` 在批次模式下執行
  - HOME – 來自 `passwd` 檔案的使用者起始目錄路徑
  - HOSTNAME – 作業執行節點的主機名稱
  - JOB\_ID – 當作業提交時，由 `sge_qmaster` 指令的獨一無二的辨識器；作業 ID 是範圍到 99999 的十進位整數
  - JOB\_NAME – 作業名稱，是由 `qsub script 檔案名稱`、時期 (`period`) 和作業 ID 的數字所建立的；此預設值可能會被 `qsub -N` 所覆寫
  - LOGNAME – 來自 `passwd` 檔案的使用者登入名稱
  - NHOSTS – 平行作業所使用的主機數目
  - NQUEUES – 配置給作業的佇列數目（序列作業請永遠配置一個）
  - NSLOTS – 平行作業所使用的佇列插槽數目
  - PATH – 預設的 `shell` 搜尋路徑：`/usr/local/bin:/usr/ucb:/bin:/usr/bin`
  - PE – 作業執行的平行環境（僅限平行作業）
  - PE\_HOSTFILE – 包含由 Sun Grid Engine，企業版指定給平行作業的虛擬平行機器之定義的檔案路徑
- 有關檔案格式的細節，請參閱 `sge_pe` 裡 `$pe_hostfile` 參數的說明。環境變數僅供平行作業使用。
- QUEUE – 作業執行的佇列名稱
  - REQUEST – 作業的要求名稱，該名稱是作業 `script` 檔案名稱或是透過 `qsub -N` 選項清楚指定給作業的名稱

- `RESTARTED` – 表示，不管檢查點作業是否已經重新啓動；如果設定（為數值 1），作業至少已中斷一次，並因此重新啓動。
- `SHELL` – 來自 `passwd` 檔案的使用者登入 shell

---

**注意** – 這不見得就是作業使用中的 shell。

---

- `TMPDIR` – 至作業暫時工作目錄的完整路徑
- `TMP` – 和 `TMPDIR` 相同；為 NQS 相容性之用
- `TZ` – 從 `sge_execd` 匯入的時區變數，如設定的話
- `USER` – 來自 `passwd` 檔案的使用者登入名稱

## ▼ 如何從指令行提交作業

- 輸入 `qsub` 指令，和適當的引數。

例如，使用 `script` 檔案名稱 `flow.sh` 的簡單作業 – 如第 64 頁的「如何從指令行執行簡單作業」一節所說明 – 可以使用以下指令提交：

```
% qsub flow.sh
```

然而，放棄延伸 `QMON` 作業提交的對等結果 – 如圖 4-9 所示 – 看起來如下：

```
% qsub -N Flow -p -l11 -P devel -a 200012240000.00 -cwd \  
-S /bin/tcsh -o flow.out -j y flow.sh big.data
```

可增加進一步的指令行選項，以組成更複雜的要求。例如，顯示在圖 4-14 的進階作業要求看起來如下：

```
% qsub -N Flow -p -l11 -P devel -a 200012240000.00 -cwd \  
-S /bin/tcsh -o flow.out -j y -pe mpi 4-16 \  
-v SHARED_MEM=TRUE,MODEL_SIZE=LARGE \  
-ac JOB_STEP=preprocessing,PORT=1234 \  
-A FLOW -w w -r y -m s,e -q big_q\  
-M me@myhost.com,me@other.address \  
flow.sh big.data
```

## 預設要求

上一節的最後一個範例顯示進階的作業要求可能會變的複雜而不方便，特別是如果類似的需求需要經常提交的話。為避免輸入這種指令行而造成難以處理且容易產生錯誤的作業，使用者可在 `script` 檔插入 `qsub` 選項（請參閱第 82 頁的「作用中的 Sun Grid Engine，企業版註釋」）或使用所謂的預設要求。

叢集管理可能為所有的 Sun Grid Engine，企業版使用者設定預設要求檔案。另一方面，使用者能在使用者的起始目錄建立私人的預設要求檔案，並能在工作目錄中建立應用程式專用預設要求檔案。

預設要求檔案只包含根據預設值在單一或多重行上應用到 Sun Grid Engine，企業版作業的 `qsub` 選項。叢集全域預設要求檔案的位置為 `<sge_root>/<儲存格>/common/sge_request`。私人的一般預設要求檔案位在 `$HOME/.sge_request` 之下，而應用程式專用的預設要求檔案預期放在 `$cwd/.sge_request`。

如果這些檔案中可使用一個以上，它們就可以依照優先順序合併成一個預設的要求：

1. 全域預設要求檔案。
2. 一般私人預設要求檔案。
3. 應用程式專用的預設要求檔案。

---

**注意** – `Script` 嵌入和 `qsub` 指令行比預設要求檔案有更高的優先權。因此，`script` 嵌入會覆寫掉預設要求檔案的設定值，而 `qsub` 指令行選項可能會再次覆寫掉這些設定值。

---

---

**注意** – `qsub -clear` 選項可以在任何時候在預設要求檔案、嵌入式 `script` 指令和 `qsub` 指令中使用，來刪除任何先前的設定值。

---

下面是一個私人預設要求檔案的範例。

```
-A 我的專案 -cwd -M me@myhost.com -m b,e
-r y -j y -S /bin/ksh
```

除非已經覆寫，特定使用者的所有作業之帳號字串將為我的專案、作業將在目前的工作目錄執行、在作業開始和結束時，郵件通知將傳送到 `me@myhost.com`、作業將在系統當機後重新啟動、標準輸出和標準錯誤輸出將合併，而 `ksh` 將被當成指令解譯器使用。

## 陣列作業

參數化並重複執行相同的一套操作（包含在作業 script 中）對 Sun Grid Engine，企業版陣列作業功能是理想的應用程式。這種應用程式的典型範例位在作業（如描繪）可在 Digital Content Creation 產業找到。在此範例中，動畫演算會分割成數個畫面，而相同的描繪運算可為每個畫面單獨執行。

陣列作業功能提供了一個方便提交、監視和控制這類應用程式的方法。另一方面，Sun Grid Engine，企業版提供陣列作業有效的實行方法，在獨立作業的陣列加入單一作業時處理運算。陣列作業的作業是經陣列索引編號提供做為參考的。所有作業的索引橫跨整個陣列作業的索引範圍，此範圍是在單一 `qsub` 指令提交陣列作業期間所定義的。

陣列作業可以以一個總數或由獨立作業或作業子集來監視和控制（如暫停、恢復或取消），在此情況下，對應的索引編號尾端會加上作業 ID 以供做作業參考。當作業執行時（跟一般作業非常相似），它們可以使用環境變數 `$SGE_TASK_ID` 來取得它們自己的索引編號，並存取指定給該作業辨識器的輸入資料集。

### ▼ 如何從指令行提交陣列作業

- 輸入 `qsub` 指令和適當的引數。

以下為提交陣列作業的範例。

```
% qsub -l h_cpu=0:45:0 -t 2-10:2 render.sh data.in
```

`-t` 選項定義作業索引範圍。在此情況下，使用的是每兩個索引時（規格的 `:2` 部分），`2-10:2` 會指定 2 是最低而 10 是最高的索引編號。因此，陣列作業包含 5 個作業，其作業索引為 2、4、6、8 和 10。每個作業會要求 45 分鐘的硬 CPU 時間。（`-l` 選項），而一旦由 Sun Grid Engine，企業版發送並啟動，將會執行作業 script `render.sh`。作業可使用 `$SGE_TASK_ID` 找出它們是否為作業 2、4、6、8 或 10，而作業可以使用其索引編號在資料檔案 `data.in` 中找出它們的輸入資料記錄。

### ▼ 如何使用 QMON 提交陣列作業

- 依照第 65 頁的「如何提交作業，從圖形使用者介面，QMON」的指示，並另外考慮以下的注意事項。

---

**注意** – 從 QMON 提交的陣列作業事實上與第 65 頁的「如何提交作業，從圖形使用者介面，QMON」所述的運作方式一致。唯一的不同是在圖 4-9 顯示的作業的作業輸入視窗需要包含和 `qsub -t` 選項一樣語法的作業範圍規格。若要取得陣列索引語法的詳細資訊，請參考《*Sun Grid Engine 5.3 和 Sun Grid Engine，企業版 5.3 參考手冊*》中的 `qsub` 項目。

---

第 106 頁的「監視和控制 Sun Grid Engine，企業版 作業」與第 119 頁的「從命令行控制 Sun Grid Engine, Enterprise Edition 作業」，和有關 `qstat`、`qhold`、`qrls`、`qmod` 和 `qdel` 的《*Sun Grid Engine 5.3 和 Sun Grid Engine，企業版 5.3 參考手冊*》各節，包含一般監視和控制 Sun Grid Engine, Enterprise Edition 作業和特別的陣列作業之持續資訊。

---

**注意** – 陣列作業提供所有一般作業可辨識的 Sun Grid Engine, Enterprise Edition 功能的完整存取。特別是，它們同時可以是平行作業，或可與其他作業產生相互依賴。

---

## 提交互動作業

提交互動作業而非批次作業在您的作業需要您的直接輸入來影響作業結果的情況時有用。這個例子通常是針對 X 視窗應用程式（其定義為互動），或針對需要立即結果的說明，來控制進一步運算的作業。

三個存在於 Sun Grid Engine, Enterprise Edition 系統以建立互動作業的方法。

- `qlogin` – 這是個類似 telnet 的階段作業，其在由 Sun Grid Engine, Enterprise Edition 軟體選取的主機上啟動。
- `qrsh` – 這相當於標準的 UNIX `rsh` 功能。指令不是在 Sun Grid Engine, Enterprise Edition 系統選取的主機上遠端執行，就是如果沒有指定執行指令，則遠端 `rlogin` (`rlogin`) 階段作業就會在遠端主機上啟動。
- `qsh` – 這是從執行作業的機器上叫出的 `xterm`，其顯示根據您的 `DISPLAY` 環境變數的規格或設定值而設定。如果沒有設定 `DISPLAY` 變數，而顯示目的也有明確定義，Sun Grid Engine, Enterprise Edition 會在提交互動作業的主機上將 `xterm` 導向 X 伺服器的 0.0 畫面。

---

**注意** – 如要正確運作，所有功能都需要正確的 Sun Grid Engine, Enterprise Edition 叢集參數配置。`qsh` 需定義正確的 `xterm` 執行路徑，而此類型的作業需能使用互動佇列。請聯絡您的系統管理員，看看您的叢集是否已經準備好執行互動作業了。

---

互動作業的預設處理和批次作業的處理不同，因為互動作業如果在提交時不能執行，它們就不會進入排程。這會立刻顯示沒有足夠的適當資源來分派提交後的互動作業。使用者在這樣的情況下會收到 Sun Grid Engine, Enterprise Edition 叢集目前太過忙碌的通知。

這套預設的運作方式可以使用 `-now no` 選項變更至 `qsh`、`qlogin` 和 `qrsh`。若指定該選項，則互動作業會像批次作業一樣進入排程。使用 `-now yes`，使用 `qsub` 提交的批次作業也可以像互動作業般處理，並會立刻分派執行或被拒絕。

---

**注意** – 互動作業只能在 INTERACTIVE（互動）類型的佇列中執行（詳情請參考第 151 頁的「關於佇列配置」）。

---

接下來的各節簡述 `qlogin` 和 `qsh` 功能的用法。`qrsh` 指令在第 92 頁的「透明遠端執行」一節較廣義的上下文 (broader context) 中有解釋。

## 使用 QMON 提交互動式作業

唯一一種可以從 QMON 提交的互動式作業是那些在 Sun Grid Engine, Enterprise Edition 選取的主機上叫出 `xterm` 的作業。

### ▼ 如何使用 QMON 提交互動作業

- 按住「作業提交」對話方塊右邊按鈕欄上的圖示，直到出現「互動」標示。

這會使「提交作業」對話方塊準備好提交互動作業（請參閱圖 4-16 和圖 4-17）。

對話方塊中選項的意義和用法和第 69 頁的「提交批次作業」一節中批次作業的意義與用法相同。根本的不同處是幾個輸入欄位設為不回應，因為它們不需套用到互動作業

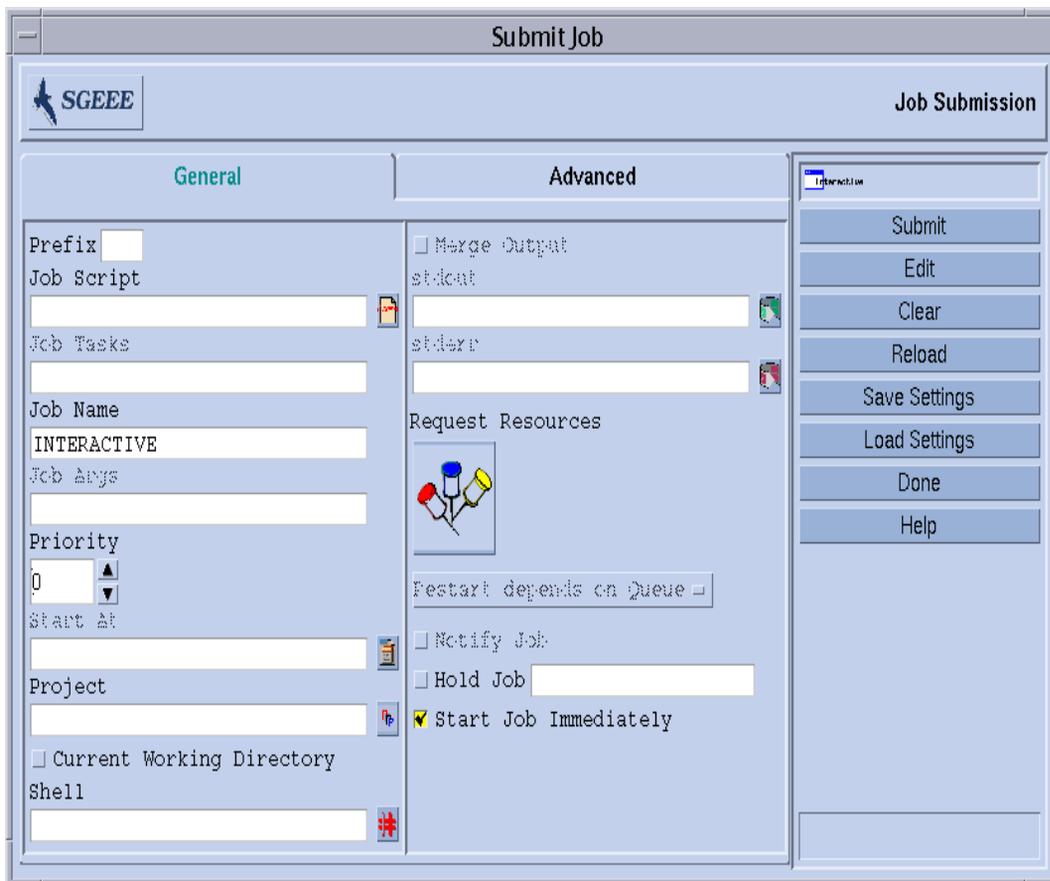


圖 4-16 互動作業提交對話方塊，一般

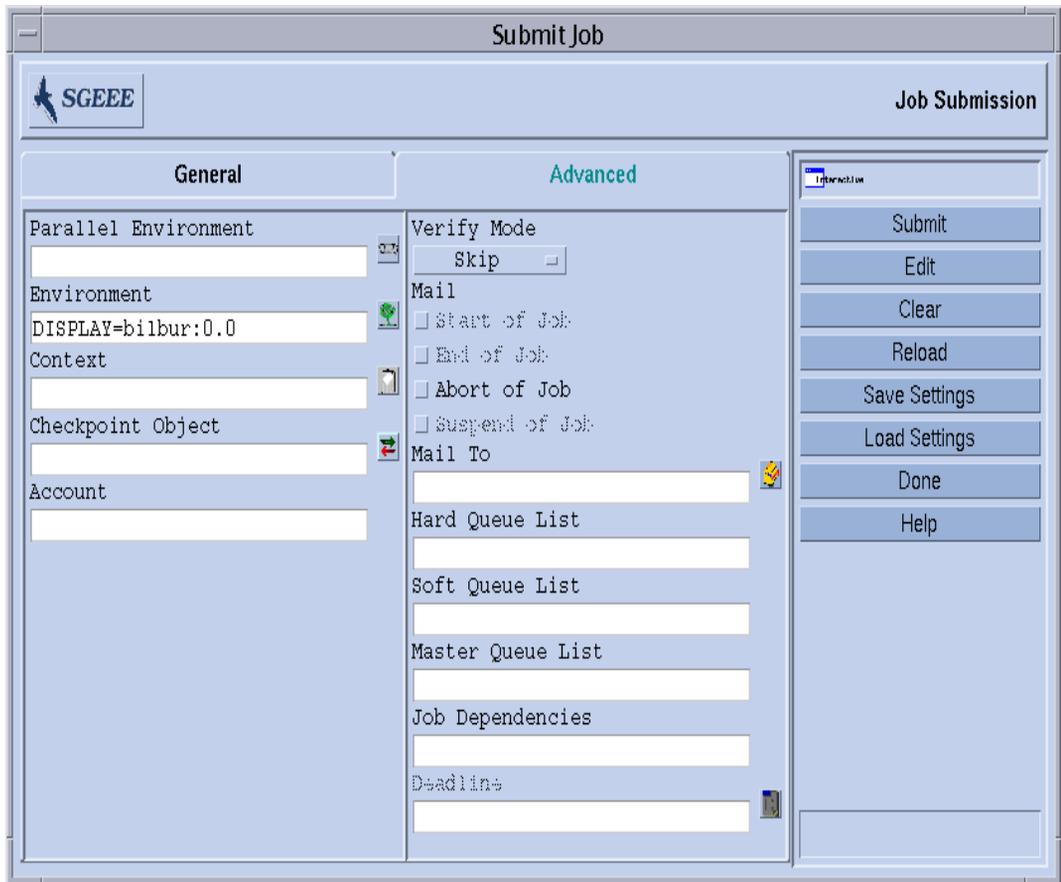


圖 4-17 互動作業提交對話方塊，進階

## 使用 qsh 提交互動式作業

Qsh 和 qsub 非常類似，並支援數個 qsub 選項，以及額外的開關 `-display` 來指向呼叫的 xterm 的顯示（詳細資訊請參考《*Sun Grid Engine 5.3 和 Sun Grid Engine，企業版 5.3 參考手冊*》中的 qsh 項目）

## ▼ 如何使用 qsh 提交互動作業

- 輸入以下指令在任何可用的 Sun Solaris 64 位元作業系統主機上啟動 xterm。

```
% qsh -l arch=solaris64
```

## 使用 qlogin 提交互動式作業

您可以從任何終端機或終端模擬使用 qlogin 指令從的 Sun Grid Engine, Enterprise Edition 控制下初始啟動互動階段作業。

## ▼ 如何使用 qlogin 提交互動作業

- 輸入以下指令來找出有 Star-CD 授權和一個至少可提供 6 小時硬 CPU 時間限制的低負載主機。

```
% qlogin -l star-cd=1,h_cpu=6:0:0
```

---

**注意** – 根據配置給 Sun Grid Engine, Enterprise Edition 系統使用的遠端登入功能，您可能必須在登入提示符號輸入使用者名稱、密碼或以上兩者。

---

## 透明遠端執行

Sun Grid Engine, Enterprise Edition 提供一套支援某些運算作業之透明遠端執行的緊密相關的功能。此功能的核心工具是第 93 頁的「使用 qrsh 遠端執行」一節所說明的 qrsh 指令。兩個建立在 qrsh 之上的高階功能 – qtcsch 和 qmake – 允許透過 Sun Grid Engine, Enterprise Edition 的隱含運算作業做透明分配，因此加強了標準 UNIX 功能 make 和 csh。qtcsch 的解釋位於第 94 頁的「使用 qtcsch 的透明作業分配」一節中，而 qmake 的解釋位於第 96 頁的「帶有 qmake 的平行檔案製作過程 (Parallel Makefile Processing)」一節中。

## 使用 qrsh 遠端執行

Qrsh 建立在 rsh 功能的周圍（有關 rsh 範圍的細節可參閱 `<sgc_root>/3rd_party` 裡的資訊），並可以用於多種用途。

- 透過可與標準 UNIX 功能 rsh（又稱為 HP-UX 的 remsh）相比的 Sun Grid Engine, Enterprise Edition 來提供互動應用程式的遠端執行。
- 透過可類似標準 UNIX 功能 rlogin 的 Sun Grid Engine, Enterprise Edition，來提供互動登入階段作業功能（請注意，qlogin 仍需要為 UNIX telnet 功能的 Sun Grid Engine, Enterprise Edition 代表）。
- 欲考慮提交再執行時支援終端機 I/O（標準 / 錯誤輸出和標準輸入）和終端機控制的批次作業。
- 提供一個方法來提交沒有嵌入 shell-script 的獨立程式。
- 提供批次作業提交客戶端，其在作業擱置或執行時會保持活動，而當作業完成或取消時才會結束。
- 允許 Sun Grid Engine, Enterprise Edition 系統控制的作業遠端執行作業（如平行作業的同步作業）在平行作業配置的分散資源架構內（請參閱第 260 頁的「PE 與 Sun Grid Engine，企業版軟體之緊密整合」一節）。

藉由這些功能，qrsh 是 qtcsh 和 qmake 功能實施，以及 Sun Grid Engine, Enterprise Edition 與平行環境（如 MPI 或 PVM）所謂的緊密整合的主要啟動基本架構。

### ▼ 如何 qrsh 使用呼叫透明遠端執行

- 輸入 qrsh 指令，依照以下大綱的引導來增加選項和引數。

```
% qrsh [ 選項 ] 程式 | shell-script [ 引/數 ] \  
          [ > stdout 檔 ] [ >&2 stderr 檔 ] [ < stdin 檔 ]
```

qrsh 幾乎了解全部的 qsub，並提供幾個額外的選項。

- `-now yes|no` — 該選項控制作業是否立刻接受排程，並在沒有適當資源時就會拒絕該作業，因此通常適用於互動作業（因此此為預設值），或如果作業無法在提交時間啟動時，是否該作業會像批次作業般進入排程。
- `-inherit` — qrsh 不會通過 Sun Grid Engine, Enterprise Edition 排程過程來啟動作業，但它會假設其已經嵌入在平行作業的上下文中，而該平行作業在指定的遠端執行主機上已經有配置的適當資源。這個 qrsh 形式通常用在 qmake 和緊密平行環境整合之中。預設值不會繼承外部的作業資源。
- `-noshell` — 有了此選項，您不用啟動在使用者登入 shell 中獲得 qrsh 的指令行，但要在沒有包裝 shell (wrapping shell) 的情況下執行該指令行。此選項可以用來加快執行的速度，因為避免掉了一些零碎空間，如 shell 啟動和取得 shell 資源檔。

- `-nostdin` — 壓縮輸入串流，STDIN。該選項設定之後，`qrsh` 會傳送 `-n` 選項給 `rsh(1)` 指令。如果在 `make(1)` 過程中使用 `qrsh` 來平行執行多重作業，則會特別有用。將不定義哪個過程會取得輸入。
- `-verbose` — 這個選項呈現排程過程上的輸出。這主要用做除錯的用途，因此根據預設值是關閉的。

## 使用 `qtcsch` 的透明作業分配

`qtcsch` 是大家熟知並廣為使用的 UNIX 之完全相容的替代品。C-Shell (`csch`) 導數 `tcsh` (`qrsh` 建立於 `tcsh` 的周圍 - 有關 `tcsh` 範圍的細節可參閱 [<sgc 根目錄>/3rd\\_party](#) 裡的資訊)。它透過 Sun Grid Engine, Enterprise Edition 提供指令 `shell` 給適當且低負載的主機指定應用程式之透明分配執行的延伸。哪個應用程式要遠端執行而哪些條件適用在執行主機的選取都定義在名為 `.qtask` 的配置檔案。

這樣的應用程式對使用者是透明的，但可透過 `qrsh` 功能提交 Sun Grid Engine, Enterprise Edition 執行。因為 `qrsh` 提供標準輸出、錯誤輸出和標準輸入處理以及到遠端執行應用程式的終端控制連線，所以遠端執行此類的應用程式和在 `shell` 相同主機上執行這類的應用程式只有三點明顯的差異處。

- 遠端主機可能比本端主機更加合適（功能更強、負載更低、已安裝需求硬 / 軟體資源），而本端主機可能連應用程式的執行都不允許。當然這是預想到的不同處。
- 遠端啟動作業或透過 Sun Grid Engine, Enterprise Edition 的處理會導致小小的延遲。
- 管理員可以透過互動作業 (`qrsh`) 和 `qtcsch` 限制資源的使用。如果沒有足夠的適當資源來透過 `qrsh` 功能啟動應用程式，或如果所有的系統都已過度載入，則隱含的 `qrsh` 提交將會失敗，並傳回對應的錯誤訊息 (`Not enough resources ... try later`)。

除了標準的使用外，`qtcsch` 也是協力廠商代碼和工具整合的適當平台。在單一執行表格中使用 `qtcsch` 在整合環境中的 `qtcsch -c 應用程式名稱` 呈現一個幾乎不需要變更的持續介面。所有所需的應用程式、工具、整合、站點，甚至是使用者特定配置都包含在適當定義的 `.qtask` 檔中。更進一步的優點是該介面可以從任何類型的 `shell script`、C 程式，甚至是 Java 應用程式之內使用。

## `qtcsch` 用法

`qtcsch` 的呼叫和對 `tcsh` 的呼叫完全相同。`Qtcsch` 藉由對 `.qtask` 檔的支援和提供一套特別化的 `shell` 內建模式來延伸 `tcsh`。

`.qtask` 檔定義如下。檔案中的每一行都有以下的格式：

```
% [!]appl_name qrsh_options
```

開頭選擇性的驚嘆號 (!) 定義叢集全域 .qtask 檔和 qtcsh 使用者的個人 .qtask 檔之中衝突定義間的優先權。若在叢集全域檔案中沒有驚嘆號的話，使用者檔案中最後衝突的定義將被拒絕。若驚嘆號在叢集全域檔案中，則不能覆寫對應的定義。

行的其他部分指定應用程式名稱，在指令行鍵入 qtcsh 時，該名稱會提交 Sun Grid Engine, Enterprise Edition 做為遠端執行之用，以及做為 qrsh 功能的選項之用，該功能將為應用程式所用，並為應用程式定義資源需求。

---

**注意** – 出現在指令行的應用程式名稱必須和 .qtask 檔所定義的一致。若有絕對值或相對的目錄規格做為字首，這是假設已輸入本端二進位且不進行遠端執行。

---

---

**注意** – 然而，在執行帶有應用程式名稱的對照前，Csh 別名已經擴展。用於遠端執行的應用程式也可能在 qtcsh 指令行的任何地方出現，特別是在標準 I/O 重新導向的前後。

---

因此，以下的範例為有效且有意義的語法。

```
# .qtask file
netscape -v DISPLAY=myhost:0
grep -l h=filesurfer
```

若為 .qtask 檔，則以下 qtcsh 指令行：

```
netscape
~/mybin/netscape
cat very_big_file | grep pattern | sort | uniq
```

將暗中造成：

```
qrsh -v DISPLAY=myhost:0 netscape
~/mybin/netscape
cat very_big_file | qrsh -l h=filesurfer grep pattern | sort | uniq
```

qtcsh 能在開關所影響的不同模式下作業，其中這些開關可以是開或關：

- 本端或遠端指令執行（遠端為預設值）
- 立即或批次遠端執行（立即為預設值）
- 冗長或非冗長輸入（非冗長為預設值）

這些模式的設定可以在啓動時間使用 `qtcs` 的選項引數，或在執行期間，使用 shell 內建指令 `qrshmode` 進行變更。要取得更多的資訊，請參閱《*Sun Grid Engine 5.3 和 Sun Grid Engine，企業版 5.3 參考手冊*》裡的 `qtcs` 項目。

## 帶有 `qmake` 的平行檔案製作過程 (Parallel Makefile Processing)

`qmake` 是標準 UNIX `make` 功能的替代品。它會延伸 `make` 的功能將獨立的 `make` 步驟分配至合適機器的叢集。`qmake` 建立在熱門的 GNU-`make` 功能，`qmake` 的周圍。要取得 `qmake` 所包含的細節，請參閱 `<sg_e_root>/3rd_party` 所提供的資訊。

要確認聚合體分散 `make` 過程可以執行到完成，`qmake` 會先以類似平行作業的類比形式配置所需的資源。然後，`qmake` 不需和 Sun Grid Engine, Enterprise Edition 排程產生互動就能管理這一套資源。當資源因啓動具有 `-inherit` 選項的 `qrsh` 功能而變成可用時，它會分配 `make` 步驟。

因為 `qrsh` 提供標準輸出、錯誤輸出和標準輸入處理，以及終端機控制連線至遠端執行 `make` 步驟，所以在本地執行 `make` 程序或使用 `qmake` 之間只有三個顯著的不同：

- 假設獨立的 `make` 步驟有一定的有效期，且有足夠的獨立 `make` 步驟可供處理，則 `make` 過程的平行化明顯加快速度。當然這是預想到的不同處。
- 在遠端啓動的 `make` 步驟中，將會有 `qrsh` 所造成的隱含小型零碎空間和這類的遠端執行。
- 要利用 `qmake` 的 `make` 步驟分配，使用者必須指定為最小程度的平行化；如可同時執行的 `make` 步驟的數量。此外，使用者可指定 `make` 步驟所要求的資源特性，如可用軟體授權、機器架構、記憶體或 CPU 時間等需求。

`make` 最常用的用法是聚合體套裝軟體的編譯。然而，這可能不是 `qmake` 的主要應用程式。程式檔案通常都很小（因為好的程式練習的緣故），因此單一程式檔案的編譯（其為單一 `make` 步驟）通常只花幾秒鐘。此外，編譯通常會暗示有很多無法加速的檔案存取（包含巢狀檔案）（如果這些檔案存取是以平行方式為多重 `make` 步驟而製作），因為檔案伺服器可以有效成為將所有檔案存取序列化的障礙。所以，有時您無法期待編譯過程能有令人滿意的加速。

其他可能的 `qmake` 應用程式更為適合。例如，相互依賴的控制和透過檔案製作之聚合體分析作業的工作流程。在某些領域中，如 EDA，這是很常見的事，而這樣的環境中，每個 `make` 步驟都是帶有非常重要的資源和運算時間需求的模擬或資料分析作業。這樣的例子中，可達成相當程度的加速。

## qmake 用法

qmake 的指令行語法看起來跟 qrsh 的指令行語法很像：

```
% qmake [-pe pe 名稱 pe 範圍][further options] \  
-- [gnu-make-options][ 目標 ]
```

---

**注意** – qmake 也支援 `-inherit` 選項，如本節稍後所述。

---

您必須特別注意 `-pe` 選項的用法及其與 `gmake -j` 選項的關係。兩個選項都可用來表示要達到的平行之數量。不同處在於 `gmake` 不會用 `-j` 提供指定像是平行環境做為使用的可能性。因此，`qmake` 會假定已為平行製作配置了名為 `make` 的預設的環境。此外，`gmake` 的 `-j` 不允許一個範圍的規格，只允許單一數字。`Qmake` 將用 `-j` 解譯所給的數字，做為 `1-< 給予的數字 >` 的範圍。相反地，`-pe` 允許所有這些參數的詳細規格。因此，以下的指令行範例是相同的。

```
% qmake -- -j 10  
% qmake -pe make 1-10 --
```

當以下的指令行無法透過 `-j` 選項表示時：

```
% qmake -pe make 5-10,16 --  
% qmake -pe mpi 1-99999 --
```

除了該語法，`qmake` 還支援兩種模式的呼叫：從指令行（沒有 `-inherit`）或在批次作業內（有 `-inherit`）的交互方式。這兩個模式初始啟動不同的動作序列：

- **互動式** – 當 `qmake` 在指令行呼叫時，這類的 `make` 過程會透過將 `qmake` 指令行指定的資源需求列入考慮的 `qrsh` 暗中提交給 Sun Grid Engine, Enterprise Edition。Sun Grid Engine, Enterprise Edition 然後會選擇一個執行與平行 `make` 作業相關的平行作業的 **主機**，然後在那裡啟動 `make` 程序。這個是必要的，因為 `make` 過程架構相依的，而所需的架構指定於 `qmake` 指令行。主機上的 `qmake` 過程然後會將獨立 `make` 步驟委託其他由 Sun Grid Engine, Enterprise Edition 配置工作並透過平行環境主機檔案傳遞 `qmake` 的主機執行。
- **批次** – 在此情況下，`qmake` 會與 `-inherit` 選項出現在批次 `script` 之內（如果 `-inherit` 選項不存在，將如上述的第一種情況，產生新的作業）。這會造成 `qmake` 利用已分配至嵌入 `qmake` 的作業的資源。它將直接使用 `qrsh -inherit` 來啟動 `make` 步驟。在批次模式呼叫 `qmake` 時，資源要求或 `-pe` 和 `-j` 選項就會被忽略。

---

**注意** – 單一的 CPU 作業也必須要求平行環境 (`qmake -pe make 1 --`)。若不需要平行執行，請呼叫具 `gmake` 指令行語法的 `qmake`（沒有 Sun Grid Engine, Enterprise Edition 選項和「--」），其運作方式將像 `gmake`。

---

若要取得 `qmake` 的進一步的詳細資料，請參考《*Sun Grid Engine 5.3 和 Sun Grid Engine, 企業版 5.3 參考手冊*》裡的 `qmake` 項目。

## Sun Grid Engine, Enterprise Edition 作業的排程方式

Sun Grid Engine，企業版軟體的策略管理會自動控制叢集內共享資源的使用，以達到管理的最佳目標。高度優先權作業將優先分派，並將獲得較佳的資源存取。Sun Grid Engine，企業版叢集的管理可以定義高階使用策略。可用的策略如下。

- **功能的** – 依照某些使用者群組、專案提供特別的處理。
- **共享為基礎的** – 服務的階層是根據指定的共享權利、對應其他使用者和使用者群組的共享資源、過去所有使用者使用資源的方法和目前系統存在的使用者所訂定的。
- **截止日期** – 作業必須在某個時間點前或同一時間完成，並可能需要特別的處理才能達成。
- **置換** – Sun Grid Engine，企業版叢集管理員的手動介入會修改資源策略執行。

Sun Grid Engine，企業版軟體能設定定期使用共享為基礎的策略、功能的策略或這兩種策略。這兩個策略可以用任何比例結合，您可以不用其中一個而只用第二個，也可以給予兩個相同的加權。

除了例行策略外，作業還必須提交初始啟動截止日期（請參閱第 74 頁的「進階範例」之下的截止日期提交參數說明）。截止日期作業會擾亂例行排程。管理員也可以暫時**置換**共享為基礎的、功能的和初始啟動截止日期的排程。置換可能會套用在獨立的作業或是所有和使用者、部門、專案或作業類別相關的作業上。

### 作業優先權

除了四個調停所有作業的策略外，Sun Grid Engine，企業版有時候會讓使用者自己設定本身作業的優先權。例如，一個提交數個作業的使用者可能會說作業 3 是最重要的作業，而作業 1 和 2 同等重要，但沒有作業 3 那麼重要。

---

**注意** – 這只有在 Sun Grid Engine，企業版軟體的策略結合包含具授與功能類別「作業」的共享資源的功能策略才有可能。

---

作業的優先權是透過 QMON 一般作業提交螢幕參數 (screen parameter) 優先權 (請參閱圖 4-9) 或透過 `-p` 選項至 `qsub` 而設定的。優先權的範圍為 -1024 (最低) 到 1023 (最高)。該優先權在單一使用者作業之間排名。其會告訴 Sun Grid Engine，企業版排程器若同時有數個作業在系統內時，如何從中選擇單一使用者作業。指定給特別作業的相關重要性根據給予使用者作業的最大和最小優先權，以及根據特定作業的優先權數值。

## 票卷

排程策略會以票卷執行。每個策略都有票卷儲存區，從儲存區中，策略可以配置票卷給進入多重機器 Sun Grid Engine，企業版系統的作業中。每個有效的例行策略會配置一些票卷給新作業，並可能在每個排程間隔時間，重新配置票卷給執行作業。每個策略用來分配票卷之標準說明如下。

票卷加權此四個策略。舉例來說，如果票卷沒有分配至功能策略，現在就不會使用此策略。如果指派相同數目的票卷給功能和共享為基礎的功能，則這兩個策略在決定作業的重要性時有相同的比重。

票卷由 Sun Grid Engine，企業版管理員配置給系統配置中的例行策略。管理員和作業員可以隨時變更票卷配置。新增之票卷會暫時置於系統中，代表一個截止期限或置換。策略由票卷的分派所結合—當票卷配置給多重策略時，作業會獲得部分的票卷，其會從每個有效的策略指出其重要性。

Sun Grid Engine，企業版授與票卷給進入系統的作業，來指出每個有效策略下的重要性。每個執行中的作業會於每個排程間隔時間得到 (例如，從一個置換中得到，或者因為截止期限已逼迫)、失去 (例如，因為擁有太多共享資源)、或保留相同數量的票卷。一個作業保有的票卷數目代表 Sun Grid Engine，企業版嘗試在每個排程間隔時間授與該作業的資源共享。

一個作業保有的票卷數目可以透過 QMON (第 106 頁的「如何使用 QMON 監視和控制作業」) 或透過 `qstat -ext` 顯示。`qstat` 指令也可以顯示指定給作業的優先權數值，例如，透過 `qsub -p` (要取得 `qstat` 的進一步的細節，請參閱《Sun Grid Engine 5.3 和 Sun Grid Engine，企業版 5.3 參考手冊》)。

## 佇列選擇

Sun Grid Engine, Enterprise Edition 系統不會分派要求非指定佇列的作業，如果這些作業不能立刻啟動。這樣的作業會在 `sge_qmaster` 標示多工緩衝處理，並會不時重新排程。因此，這樣的作業會分派到下個可用的適合佇列。

相反地，由名稱要求至某的佇列的作業會直接前往這個佇列，不管它們是否會啟動或進入多工緩衝處理。因此，以資訊工程 *批次佇列* 檢視 Sun Grid Engine, Enterprise Edition 佇列是對名稱要求的作業唯一有效的方法。用非特定要求提交的作業會使用 `sge_qmaster` 的多工緩衝處理機制做為佇列排序，然後利用一個較抽象和較有彈性的佇列排序概念。

如果作業已經過排程，而多重、未使用的佇列符合其資源要求，作業通常會分派給屬於承載量最低的主機之佇列（最合適的）。藉由將 Sun Grid Engine, Enterprise Edition 排程器配置項目 `queue_sort_method` 設定為 `seq_no`，叢集管理可能變更這個載入相依計劃到固定順序演算法中：佇列配置項目 `seq_no` 是用最低序列編號來定義指定佇列最高優先權的在佇列間的優先順序。

# 檢查點、監視和控制作業

---

在您使用 Sun Grid Engine，企業版 5.3 系統提交作業後，您需要能夠監視並控制這些作業。本章提供完成這些作業的背景資訊和指示。

本章包含以下特定作業的指示：

- 第 103 頁的「如何從指令行提交、監視或刪除檢查點作業」
- 第 104 頁的「如何使用 QMON 提交檢查點作業」
- 第 106 頁的「如何使用 QMON 監視和控制作業」
- 第 115 頁的「如何使用 qstat 監視作業」
- 第 118 頁的「如何使用電子郵件監視作業」
- 第 119 頁的「如何從指令行控制作業」
- 第 120 頁的「如何用 QMON 控制佇列」
- 第 124 頁的「如何用 qmod 控制佇列」

---

## 關於檢查點作業

本節探索兩個不同種類的作業檢查點。

- *使用者階層*
- *Kernel 階層*

## 使用者階層檢查點

許多應用程式，特別是通常消耗很多 CPU 時間的程式，已實行檢查點並重新啓動機制以增加容錯。狀態資訊和已處理資料的重要部分再運算式的某些階段重複寫入一或數個檔案。如果比較檢查點之前的狀況，應用程式可以中斷，然後稍後重新啓動，並達成一致的狀態，則就可處理這些檔案（稱為重新啓動檔案）。當使用者大部分必須處理重新啓動檔案，好將這些檔案移至正確的位置時，這種的檢查點稱為**使用者階層**檢查點。

對沒有整合的（使用者階層）檢查點之應用程式來說，您可以使用公共領域（例如，請參閱 University of Wisconsin 的 *Condor* 專案）或一些硬體廠商所提供所謂的**檢查點檔案庫**做為替代。重新將應用程式連結這樣的檔案庫會把檢查點機制安裝到應用程式中，而不需變更來源碼。

## Kernel 階層檢查點

有些作業系統在作業體統 kernel 之內提供檢查點支援。在此情況下，應用程式不需準備也不需重新連結。Kernel 階層檢查點通常適用於單一過程也可用於完整過程階層。例如，相依過程的階層會隨時進行檢查點檢查並重新啓動。通常可用使用者指令和 C 資料庫 (C-library) 介面初始啓動檢查點。

Sun Grid Engine，企業版支援作業系統檢查點，如果有的話。要取得目前支援的 kernel 階層檢查點功能的資訊，請參考《Sun Grid Engine, Enterprise Edition Release Notes》。

## 檢查點作業的遷移

檢查點作業可隨時中斷，因為它們重新啓動的功能確定僅少數已完成工作必須重複。此功能用來建立 Sun Grid Engine，企業版的遷移和動態裝載平衡機制。若有要求，檢查點 Sun Grid Engine，企業版作業會依照要求中斷並遷移到 Sun Grid Engine，企業版儲存區的其他機器，並藉動態的方式平衡叢集中的裝載。檢查點作業會因下列原因中斷和遷移。

- 執行佇列或作業應由 `qmod` 或 `qmon` 指令明確暫停。
- 執行佇列或作業會自動暫停，因為已超過佇列的暫停臨界值（請參閱第 156 頁的「如何配置載入 / 暫停臨界值」一節）而作業的檢查點事件規格包含暫停的例子（請參閱第 103 頁的「如何從指令行提交、監視或刪除檢查點作業」一節）。

遷移作業會移回 `sge_qmaster` 並再之後分配到另一個合適的佇列，如果有任何合適佇列的話。在這樣的例子中，`qstat` 輸出會顯示 R 做為狀態。

## 組成檢查點作業 Script

kernel 階層檢查點的 Shell script 和一般 shell script 沒有不同。

使用者階層檢查點作業的 Shell script 與一般 Sun Grid Engine, Enterprise Edition 批次 script 間唯一的不同在於一旦重新啓動，它們能正確處理該案的能力。環境變數 RESTARTED 設定在作業重新啓動時，會對作業進行檢查點檢查。它可以用來跳過作業 script 的區段，其僅可以在初始呼叫時執行。

因此，透明檢查點的作業 script 看起來可能類似程式範例 5-1。

```
#!/bin/sh
#Force /bin/sh in Sun Grid Engine, Enterprise Edition
#$ -S /bin/sh

# Test if restarted/migrated
if [ $RESTARTED = 0 ]; then
    # 0 = not restarted
    # Parts to be executed only during the first
    # start go in here
    set_up_grid
fi

# Start the checkpointing executable
fem
#End of scriptfile
```

#### 程式範例 5-1 檢查點作業 Script 範例

如果使用者階層檢查點作業已遷移，記下作業 script 從頭開始重新啓動是很重要的。使用者負責將 shell-script 的程式流程導向作業中斷的位置，因此需跳過 script 中這些執行一次以上的重要行列。

---

**注意** – Kernel 階層檢查點作業是可隨時中斷的，而包含的 (embracing) shell script 正從上個檢查點發生的地方重新啓動。因此，RESTARTED 環境變數和 kernel 階層檢查點作業沒有相關。

---

## ▼ 如何從命令行提交、監視或刪除檢查點作業

使用合適的開關輸入以下指令。

```
#qsub 選項引數
```

提交檢查點作業的運作方式和一般批次 script 的作業方式相同，qsub -ckpt 和 -c 開關除外，這些開關要求檢查點機制並定義為作業產生的檢查點事件。-ckpt 選項採用為檢查點環境 (第 219 頁的「關於檢查點支援」) 的名稱的引數。-c 選項並非強制

性，並也採用一個引數。它可以用來覆寫檢查點環境配置裡的 when 參數的定義（細節請參閱《Sun Grid Engine 5.3 和 Sun Grid Engine，企業版 5.3 參考手冊》中的 checkpoint 項目）。

-c 選項的引數可以是以下單字母選項（或字母的任何組合）中的一個或替代的時間值：

- n – 沒有執行檢查點。這有最高的優先權
- s – 檢查點只會在關閉作業主機上的 sge\_execd 才會產生。
- m – 以對應佇列配置所定義的所少 CPU 間隔時間產生檢查點（請參閱 queue\_conf 線上說明手冊的 min\_cpu\_interval 參數）。
- x – 若作業暫停，檢查點就會產生。
- interval – 在指定的間隔時間產生檢查點，但該間隔時間不能比 min\_cpu\_interval（請上述）所定義的還要頻繁。時間數值必須指定為 hh:mm:ss（以冒號分隔的二位數的時數、分數和秒數）。

檢查點作業 *監視* 與一般作業的不同處是這些作業會不時地遷移，因此，這些作業不能受限於單一佇列。然而，唯一的作業辨識編號與作業名稱都維持相同。

*刪除* 檢查點作業的運作方式如同第 119 頁的「從指令行控制 Sun Grid Engine, Enterprise Edition 作業」一節所述。

## ▼ 如何使用 QMON 提交檢查點作業

- 依照第 74 頁的「進階範例」的指示，注意以下的額外資訊。

透過 QMON 的檢查點作業提交與一般加上指定合適的檢查點環境之批次作業的提交相同。如同第 74 頁的「進階範例」程序中所解釋的，「作業提交」對話方塊提供與作業相關的檢查點環境一個輸入視窗。輸入視窗之外還有一個圖示鈕，其可以開啓顯示於圖 5-1 的「選擇」對話方塊。您可以從其可用的清單選擇合適的檢查點環境。有關安裝於您站點之檢查點環境屬性的資訊，請向系統管理員查詢，或參考第 219 頁的「關於檢查點支援」一節。



圖 5-1 檢查點物件選擇

## 檔案系統需求

當寫入檢查點檔案庫式的使用者階層或 `kernel` 階層檢查點時，進行檢查點之過程或作業涵蓋的完整虛擬記憶體影像則需轉儲。要達到這個目的，您需要足夠的磁碟空間。若檢查點環境配置參數 `ckpt_dir` 已設定，則檢查點資訊會轉儲到 `ckpt_dir` 之下的作業私人位置。若 `ckpt_dir` 設為 `NONE`，則會使用檢查點作業啟動的目錄。要取得檢查點環境配置的詳細資訊，請參考《*Sun Grid Engine 5.3 和 Sun Grid Engine，企業版 5.3 參考手冊*》中的 `checkpoint` 項目。

---

**注意** – 若 `ckpt_dir` 設為 `NONE`，您應該使用 `qsub -cwd script` 啟動檢查點作業。

---

檔案系統如何組織的額外需求所引起的的原因是，為了順利遷移並重新啓動作業，所有的機器上都必須看的到檢查點檔案和重新啓動檔案。因此需要 NFS 或類似的檔案系統。請詢問您的叢集管理看看這個需求是否符合您的站點。

若您的站點沒有執行 NFS 或如果因某些因素而沒使用，您應該能在使用者階層檢查點作業的情控下，從 `shell script` 的開端（例如，透過 `rcp` 或 `ftp`）明確傳輸重新啓動檔案。

---

# 監視和控制 Sun Grid Engine，企業版 作業

原則上有三種監視提交作業的方式。

- 使用 Sun Grid Engine, Enterprise Edition 圖形使用者介面，QMON
- 從指令行使用 `qstat` 指令
- 使用電子郵件

## ▼ 如何使用 QMON 監視和控制作業

Sun Grid Engine, Enterprise Edition 圖形使用者介面，QMON，提供特別為控制作業設計的對話方塊。

- 在 QMON 主功能表中，按一下「作業控制」鈕，然後根據以下各節詳細記載的額外資訊繼續進行。

此對話方塊的一般目的是提供監視所有執行、擱置和因此系統或零件能辨識之完成作業的可配置數目。對話方塊也可以用來操作作業，如變更其優先權、暫停、恢復或取消作業。共顯示三個環境，一個用於執行作業、另一個用於等待分派至合適資源的擱置作業，而第三個用於最近完成的作業。您可以按螢幕上方的對應 `tab` 標籤在三個列出環境中做選擇。

在其預設的格式中（圖 5-2），它會顯示每個執行和擱置作業的作業 Id、優先權、作業名稱、和佇列欄位。

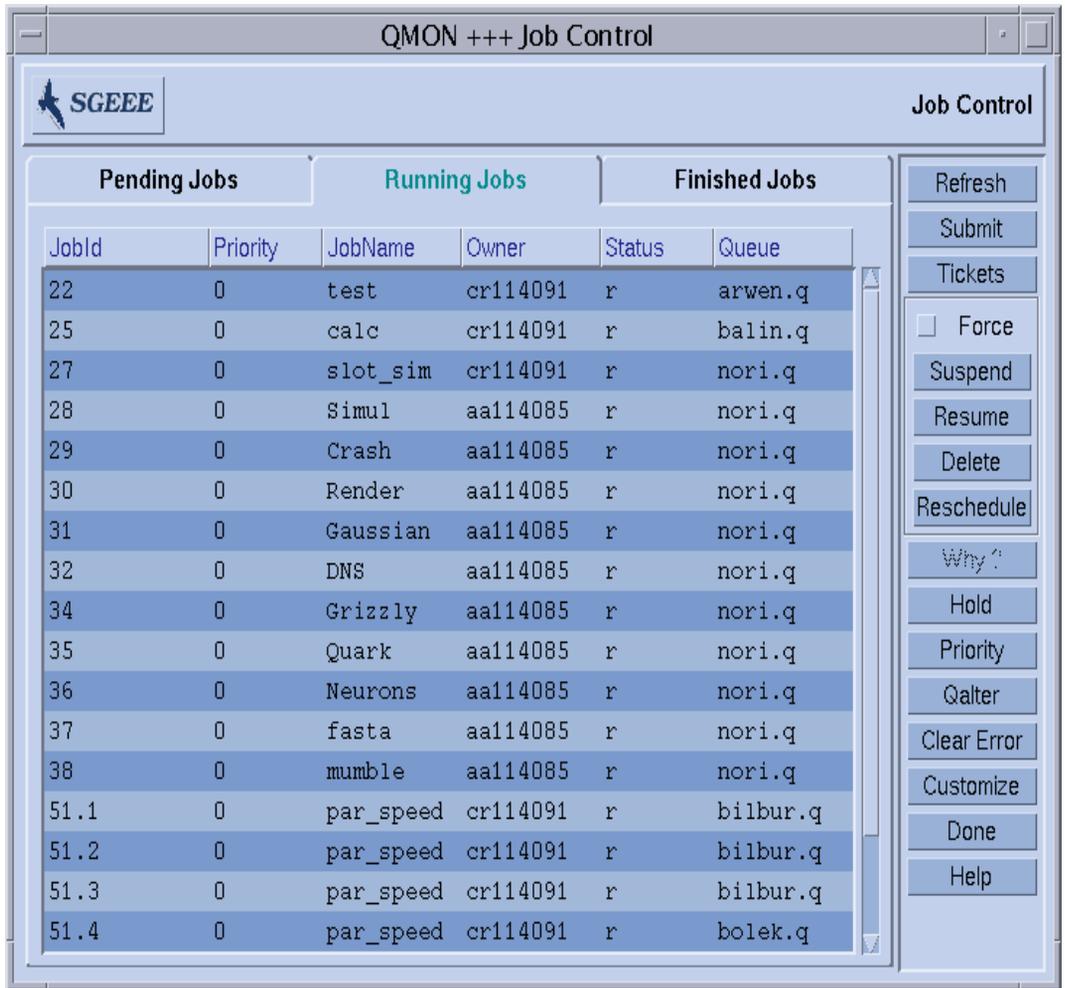


圖 5-2 作業控制對話方塊－標準格式

您可以配置與「自訂」對話方塊一同顯示的資訊集，（請參閱圖 5-3），您只需按「作業控制」對話方塊中的「自訂」鈕即可開啓此對話方塊。

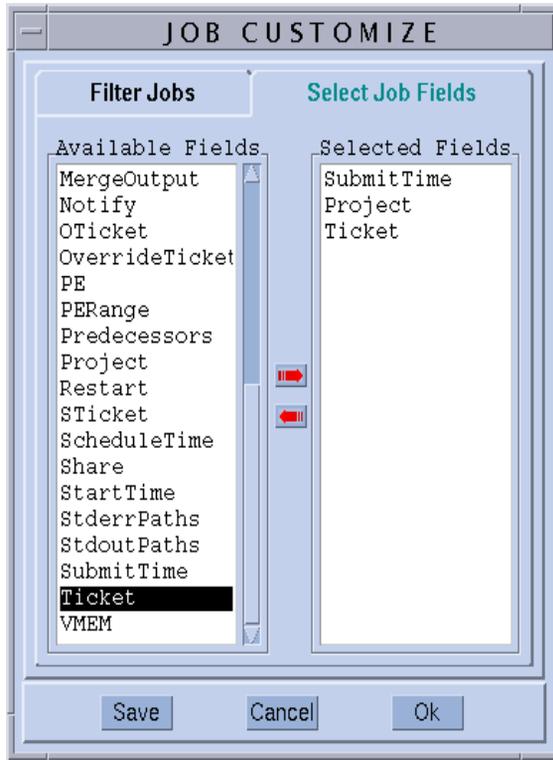


圖 5-3 作業控制自訂對話方塊

有了「自訂」對話方塊，您就可以選擇將顯示並過濾利益作業之 Sun Grid Engine, Enterprise Edition 作業物件的進一步的項目。圖 5-3 的範例選擇專案、票卷、和提交時間等額外的欄位。

顯示於圖 5-4 的「作業控制」對話方塊描述在「完成作業」清單使用自訂後之加強的外觀。



圖 5-4 作業控制對話方塊完成作業－加強

圖 5-5 中過濾功能的範例只會選擇 chaubal 所擁有的作業，其可在 solaris 架構執行或適合該架構。

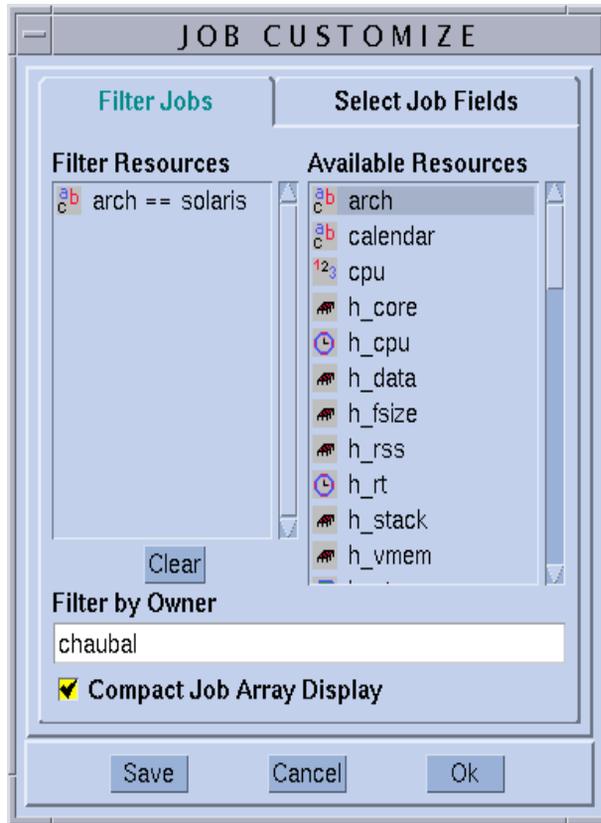


圖 5-5 作業控制過濾

產生顯示「擱置作業」的「作業控制」對話方塊顯示於圖 5-6。

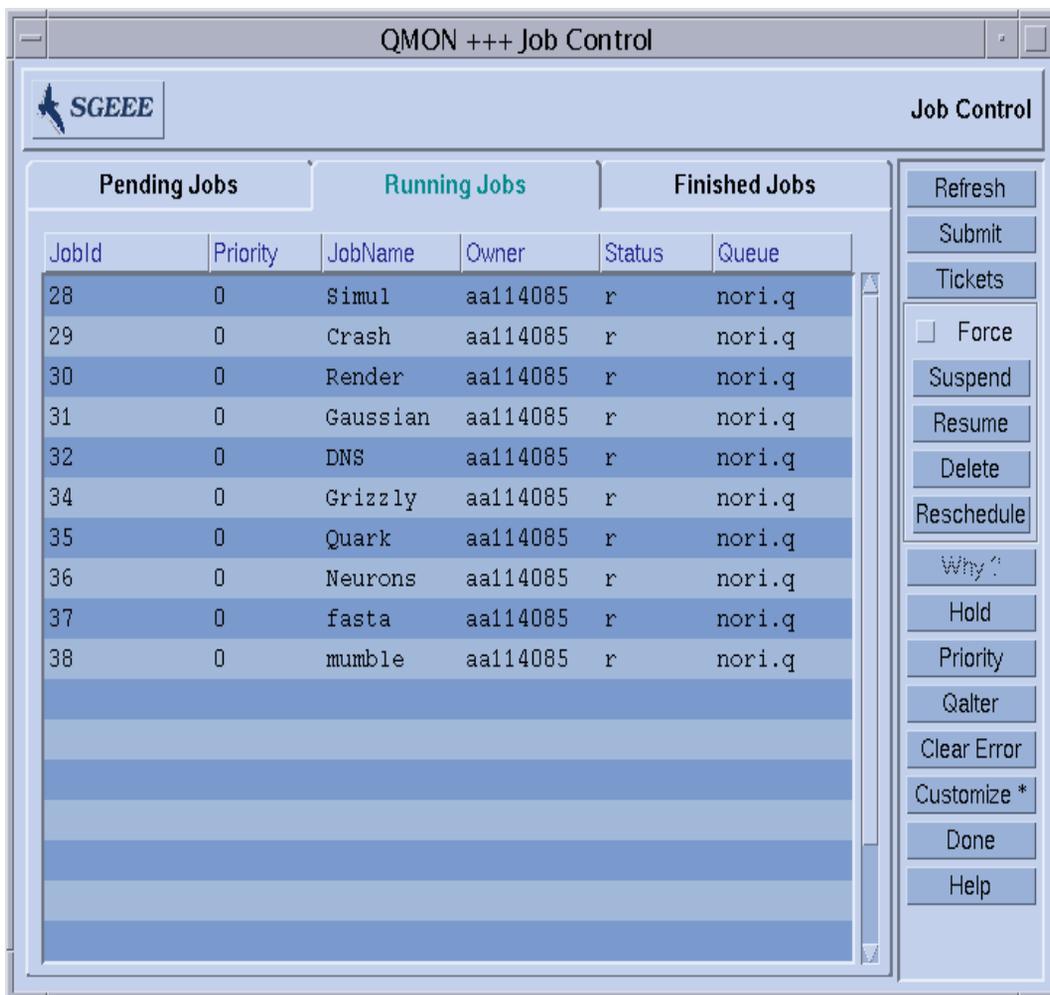


圖 5-6 作業控制對話方塊—過濾後

**注意** – 例如，顯示於圖 5-3 的「自訂」對話方塊中的 Save 鈕會將自訂儲存於使用者初始目錄的 .qmon\_preferences 檔案中，並因此重新定義「作業控制」對話方塊的預設外觀。

圖 5-6 中的「作業控制」對話方塊也是陣列作業如何在 QMON 顯示的範例。

作業可用以下滑鼠 / 鍵的組合選擇（做為日後操作之用）：

- 在按住「控制」鍵時，用滑鼠左鍵按一下作業可啟動多重作業的選項。

- 在按住「Shift」鍵時，用滑鼠左鍵按一下另一個作業可以選擇之間、包含選項開始的作業和目前作業之所有作業。
- 在按住「控制」鍵時，用滑鼠左鍵按一下作業可切換單一作業的選擇狀態。

您可以透過螢幕右邊的對應鈕暫停、恢復（解除暫停）、刪除、阻止（和放開）、重設優先權並修改 (Qalter) 選取的作業。

暫停、解除暫停、刪除、保留、修改優先權和修改作業的動作只適用於作業所有人或 Sun Grid Engine, Enterprise Edition 管理員和操作員的作業（請參閱第 61 頁的「管理員、操作員和所有人」）。只有執行的作業才能暫停 / 恢復；只有擱置作業可以阻止和修改（在優先權和其他屬性中）。

暫停作業表示相當於傳送帶有 UNIX kill 指令的 SIGSTOP 訊號給作業的處理群組，該指令會中止作業且不再消耗 CPU 時間。解除暫停作業會傳送 SIGCONT 訊號，並因此恢復作業（欲取得更多發出訊號過程的資訊，請參閱系統的 kill 線上說明手冊）。

---

**注意** – 您可以強制暫停、解除暫停和刪除；例如，註冊 sge\_qmaster 但因網路問題而不通知控制作業的 sge\_execd，例如，在對應的 sge\_execd 無法達成時。使用 Force 旗號來達到此目的。

---

若您在選取的擱置作業使用保留鈕，就會開啓「設定保留」(Set Hold) 子對話方塊（請參閱圖 5-7）。

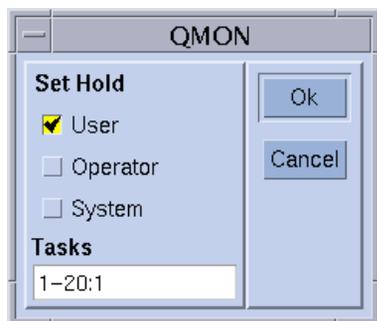


圖 5-7 作業控制保留

「設定保留」子對話方塊能啟動設定和重設使用者、系統和操作員保留。作業所有人和 Sun Grid Engine, Enterprise Edition 操作員和管理員可以設定或重設使用者保留 (User holds)。管理者和操作者可以設定或重設操作者保留 (Operator holds)，而管理者保留 (managers holds) 只能由管理者設定或重設。只要作業指定有任何的保留，該作業就不適合執行。設定或重設保留的替代方法為 qalter、qhold 和 qrls 指令（請參閱《Sun Grid Engine 5.3 和 Sun Grid Engine, 企業版 5.3 參考手冊》的對應項目）。

「設定保留」鈕的作業欄位適用於「陣列」作業。您可以使用這個按鈕將保留放置於陣列作業的特定子作業之上。請注意圖 5-7 中「作業」欄位中的文字格式。指定在本欄位的作業 id 範圍可能是單一數字、*n-m* 形式的簡單範圍、或有一定步長 (step size) 的範圍。因此，例如由 2-10:2 所指定的作業 id 範圍會產生作業 id 指數 2、4、6、8 和

10；（如，在總數五個相同作業中，每個作業都具有五個指數之一的環境變數 *SGE 作業 ID*）。要取得作業保留的詳細資訊，請參閱《*Sun Grid Engine 5.3 和 Sun Grid Engine, 企業版 5.3 參考手冊*》的 `qsub` 項目，或 `qsub(1)` 線上說明手冊。

若按住優先權鈕，就會開啓另一個子對話方塊（圖 5-8），其可以啓動輸入選取的 Sun Grid Engine, 企業版擱置和執行作業的新優先權。在 Sun Grid Engine, 企業版中，優先權在單一使用者作業之間排名。其會告訴 Sun Grid Engine, Enterprise Edition 排程器若同時有數個作業在系統內時，如何從中選擇單一使用者作業。

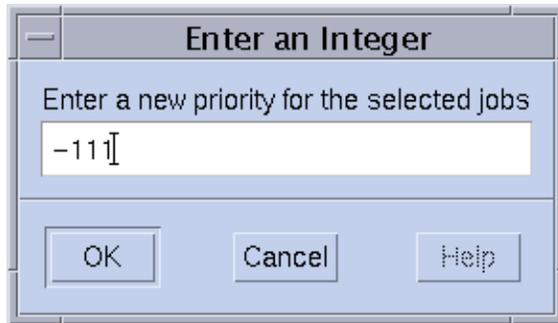


圖 5-8 作業控制優先權定義

為擱置作業按 *Qalter* 鈕時，就會開啓第 65 頁的「如何提交作業，從圖形使用者介面，*QMON*」說明的「作業提交」畫面，以及對應提交時定義之作業屬性的所有對話方塊集的項目。這些無法變更的項目設為無回應。其他的您可以進行編輯，並在「作業提交」對話方塊按 *Qalter* 鈕（可用來取代「提交」鈕）來用 Sun Grid Engine, Enterprise Edition 紀錄變更。

在「作業提交」畫面中，*驗證* 旗號使用於 *Qalter* 模式時有特殊的意義。您可以檢查擱置作業的一致性，並調查它們還沒排程的原因。您只需為「驗證」旗號選擇想要的一致性檢查模式，然後按 *Qalter* 鈕。系統會根據選擇的檢查模式顯示發生不一致時的警告。要取得更進一步的資訊，請參閱第 74 頁的「進階範例」一節和 `qalter` 線上說明手冊的 `-w` 選項。

另一個檢查為何作業仍舊擱置的方法是選擇作業，然後按「作業控制」對話方塊的 why? 鈕。這會開啓「物件瀏覽器」對話方塊，並顯示阻止 Sun Grid Engine, Enterprise Edition 排程器在最近一次的 pass 分派作業的原因清單。顯示這類訊息的範例瀏覽器畫面顯示在圖 5-9。

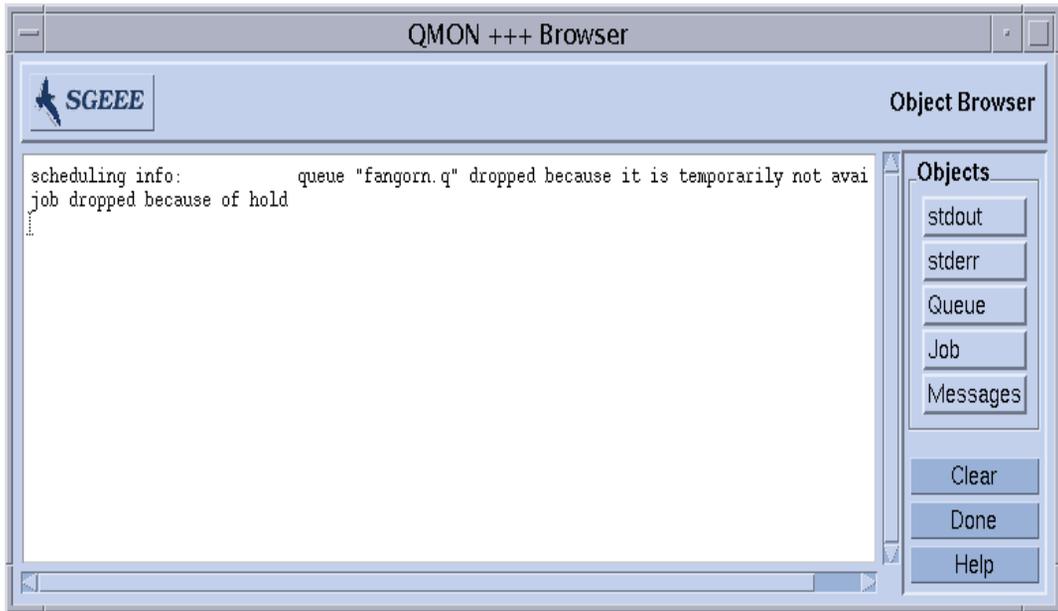


圖 5-9 顯示排程資訊的瀏覽器

---

**注意** – why? 鈕只會傳送有意義的輸出，若排程器配置參數 `schedd_job_info` 設為 `true` 的話（請參閱《*Sun Grid Engine 5.3 和 Sun Grid Engine, 企業版 5.3 參考手冊*》的 `sched_conf`）。顯示的排程資訊與上次的排程間隔時間有關。在您調查作業沒有進行排程時，該資訊可能已經不正確了。

---

「清除錯誤」按鈕可以用來移除選取之擱置作業的錯誤狀態，該作業曾在之前嘗試啓動，可是因為作業相依問題而失敗（如，權限不足以寫入指定的作業輸出檔案）。

---

**注意** – 錯誤狀態使用擱置作業清單中的紅色字體顯示，並應該在修正錯誤狀況後移除；如透過 `qalter`。如果作業要求在作業中斷時要傳送電子郵件，則這樣的錯誤狀況會自動透過電子郵件報告（如透過 `qsub -m a` 選項）。

---

為使顯示的資訊保持最新，QMON 使用輪詢計劃從 `sge_qmaster` 取得作業狀態。按「重新顯示」鈕可以強制更新。

最後，該鈕提供 QMON 「作業提交」對話方塊的連結（請參閱圖 5-10 的範例）。

## QMON 物件瀏覽器的額外資訊

QMON 物件瀏覽器可以用來快速取得 Sun Grid Engine, Enterprise Edition 作業的額外資訊，而不需要如第 106 頁的「如何使用 QMON 監視和控制作業」一節所述的自訂「作業控制」對話方塊。

按一下 QMON 主功能表的「瀏覽器」圖示鈕就可開啓「物件瀏覽器」。如果選取瀏覽器的「作業」鈕，且如果滑鼠指標移到「作業控制」對話方塊中的作業行上，則瀏覽器會顯示 Sun Grid Engine, Enterprise Edition 作業的相關資訊（請參閱圖 5-2 的範例）。

圖 5-10 的瀏覽器畫面會提供在此情況下所顯示的資訊的範例。

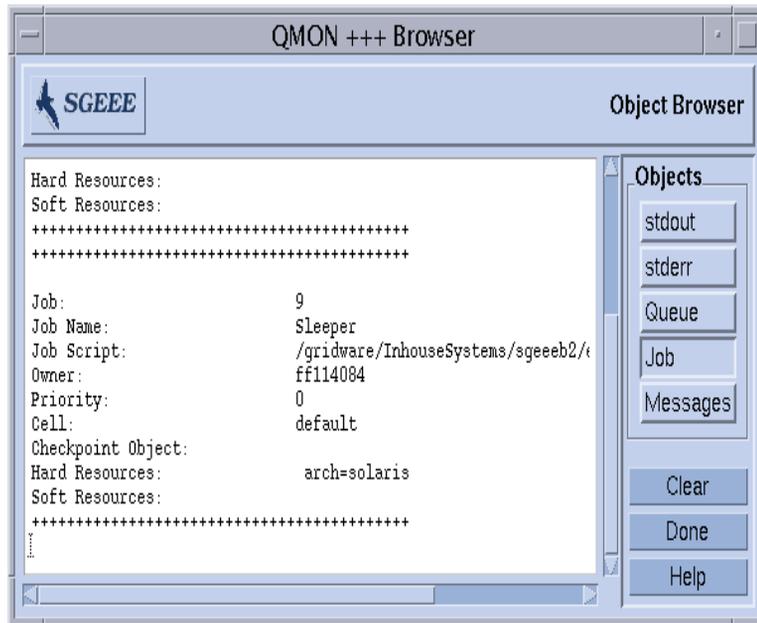


圖 5-10 物件瀏覽器－作業

### ▼ 如何使用 `qstat` 監視作業

- 從指令行透過詳述於以下各節的資訊的引導，使用以下其中一個指令。

```
% qstat
% qstat -f
% qstat -ext
```

第一個表格只提供提交作業的簡介（請參閱表 5-1）。第二個表格包含目前另外配置的佇列資訊（請參閱表 5-2）。第二個表格包含詳細資訊，如最新作業用法和指定給作業的票卷。

在第一個表格中，標題行表示欄位的意義。大部分欄位的目的應該可以自我解釋。然而，`state` 一欄包含具以下意義的單一字元碼：`r` 表示執行、`s` 表示暫停、`q` 表示已進入排程而 `w` 表示等待（要取得 `qstat` 輸出格式的詳細解釋，請參閱《*Sun Grid Engine 5.3 和 Sun Grid Engine，企業版 5.3 參考手冊*》中的 `qstat` 項目）。

第二個表格分為兩個部分，第一個顯示所有可用佇列的狀態，第二個（以 `- PENDING JOBS - ...` 分隔元為名）顯示 `sge_qmaster` 作業多工緩衝處理區域的狀態。佇列部分的第一行定義與列出與佇列相關欄位的意義。佇列由水平線分隔線分隔。若作業在佇列中執行，則作業會以第一個表格中以和 `qstat` 指令相同的格式，列印於相關佇列之下。第二個輸出部分的擱置作業也會如同 `qstat` 第一個表格的擱置作業的列印方式印出。

以下各欄的佇列說明需要更多的解釋。

- `qtype` — 此為佇列種類 — `B(atch)`、`I(nteractive)`、`P(arallel)` 和 `C(heckpointing)` 或組合之一。
- `used/free` — 此為佇列中已使用 / 未使用作業插槽的計算。
- `states` — 此為佇列的狀態 — `u(nknown)`、`a(larm)`、`s(uspended)`、`d(isabled)`、`E(rror)` 或組合之一。

此外，`qstat` 線上說明手冊包含更詳細的 `qstat` 輸出格式說明。

在第三個 Sun Grid Engine，企業版特定格式中，指定給作業的用法和票卷值包含在以下各欄中。

- `cpu/mem/io` — 這是目前累積的 CPU、記憶體和 I/O 用法。
- `tckts/ovrts/otckt/dtckt/ftckt/stckt` — 這些數值與總共指定給作業的票卷有關，透過 `qalter -ot`、透過置換策略、透過截止日期、透過功能性策略並透過共享式策略。

此外，截止日期初始啟動時間顯示於截止日期一欄中（如果有的話），而共享欄會顯示每個作業所擁有，目前與叢集中所有作業產生的用法相關的資源共享。要取得進一步的細節，請參閱 `qstat` 線上說明手冊。

`qstat` 指令的各種額外選項會增加兩個版本的功能。`-r` 選項可用來顯示提交作業的資源需求。此外，輸出可以限制於某個使用者，或限制於特定的佇列，而 `-l` 選項可以用來指定 `qsub` 指令的資源需求，如第 79 頁的「資源需求定義」一節所述。若使用資源需求，只有符合 `qstat` 指令行中的資源需求規格的佇列（而作業在這些佇列中執行）才會顯示。

表 5-1 和表 5-2 顯示輸出表格的範例 `qstat` 和 `qstat -f` 指令。

表 5-1 `qstat` 輸出的範例

| job-ID | prior | name      | user    | state | submit/start at      | queue   | function |
|--------|-------|-----------|---------|-------|----------------------|---------|----------|
| 231    | 0     | hydra     | craig   | r     | 07/13/96<br>20:27:15 | durin.q | MASTER   |
| 232    | 0     | compile   | penny   | r     | 07/13/96<br>20:30:40 | durin.q | MASTER   |
| 230    | 0     | blackhole | don     | r     | 07/13/96<br>20:26:10 | dwain.q | MASTER   |
| 233    | 0     | mac       | elaine  | r     | 07/13/96<br>20:30:40 | dwain.q | MASTER   |
| 234    | 0     | golf      | shannon | r     | 07/13/96<br>20:31:44 | dwain.q | MASTER   |
| 236    | 5     | word      | elaine  | qw    | 07/13/96<br>20:32:07 |         |          |
| 235    | 0     | andrun    | penny   | qw    | 07/13/96<br>20:31:43 |         |          |

表 5-2 qstat -f 輸出的範例

| queuename  | qtype | used/free | load_avg | arch | states                   |
|--|-------|-----------|----------|------|--------------------------|
| dq   | BIP   | 0/1       | 99.99    | sun4 | au                       |
| durin.q  | BIP   | 2/2       | 0.36     | sun4 |                          |
| 231  | 0     | hydra     | craig    | r    | 07/13/96 20:27:15 MASTER |
| 232  | 0     | compile   | penny    | r    | 07/13/96 20:30:40 MASTER |
| dwain.q  | BIP   | 3/3       | 0.36     | sun4 |                          |
| 230  | 0     | blackhole | don      | r    | 07/13/96 20:26:10 MASTER |
| 233  | 0     | mac       | elaine   | r    | 07/13/96 20:30:40 MASTER |
| 234  | 0     | golf      | shannon  | r    | 07/13/96 20:31:44 MASTER |
| fq   | BIP   | 0/3       | 0.36     | sun4 |                          |
| #####  |       |           |          |      |                          |
| - PENDING JOBS - |       |           |          |      |                          |
| #####  |       |           |          |      |                          |
| 236  | 5     | word      | elaine   | qw   | 07/13/96 20:32:07        |
| 235  | 0     | andrun    | penny    | qw   | 07/13/96 20:31:43        |

## ▼ 如何使用電子郵件監視作業

- 從指令行透過詳述於以下各節的資訊的引導，用合適的引數輸入以下指令。

```
% qsub 引數
```

qsub -m 開關要求如果有事件發生時，需傳送電子郵件給提交作業的使用者或 -M 旗號所指定的電子郵件地址（有關旗號的說明，請參閱 qsub 線上說明手冊）。-m 選項的引數指定事件。有以下選項可用：

- b – 郵件在作業開始時寄送。
- e – 郵件在作業結束時寄送。
- a – 郵件在作業中斷時寄送（如，透過 qdel 指令）。
- s – 郵件在作業暫停時寄送。
- n – 不寄送任何郵件（預設值）。

多重選項可以在以逗點分隔的清單中使用單一 `-m` 選項選取。

相同的郵件事件可由 QMON 「作業提交」對話方塊的說明配置。請參閱第 74 頁的「進階範例」一節。

## 從指令行控制 Sun Grid Engine, Enterprise Edition 作業

第 106 頁的「如何使用 QMON 監視和控制作業」一節解釋如何使用 Sun Grid Engine, Enterprise Edition 圖形使用者介面 QMON 刪除、暫停和恢復 Sun Grid Engine, Enterprise Edition 作業。

相同的功能也可以從指令行取得，本節有相關說明。

### ▼ 如何從指令行控制作業

- 從指令行透過詳述於以下各節的資訊的引導，用合適的引數輸入以下其中一個指令。

```
% qdel 引數  
% qmod 引數
```

您可使用 `qdel` 指令取消 Sun Grid Engine, Enterprise Edition 作業，不管它們在執行或在多工緩衝處理中。`qmod` 指令提供執行中作業的暫停或解除暫停（恢復）的方法。

對兩個指令來說，您需要知道作業辨識號碼，該號碼是回應成功的 `qsub` 指令而顯示的。若您忘記該號碼，則您可以透過 `qstat` 取得（請參閱第 115 頁的「如何使用 `qstat` 監視作業」一節）。

以下包含兩個指令的數個範例：

```
% qdel 作業 id  
% qdel -f 作業 id1、作業 id2  
% qmod -s 作業 id  
% qmod -us -f 作業 id1、作業 id2  
% qmod -s 作業 id. 作業 id 範圍
```

要刪除、暫停或解除暫停作業，您必須是作業的所有人、Sun Grid Engine, Enterprise Edition 管理員或操作員（請參閱第 61 頁的「管理員、操作員和所有人」）。

對兩個指令來說，`-f` 強制選項可以用來記錄作業在 `sge_qmaster` 的狀態變更而不通知 `sge_execd`（例如，`sge_execd` 因為網路問題之類的緣故而無法達成時）。`-f` 選項是用來給管理員使用的。然而，如果是 `qdel`，則使用者可以強制刪除他們自己的作業，若叢集配置 `qmaster_params` 項目中的旗號 `ENABLE_FORCED_QDEL` 已設定（要取得更多資訊，請參閱《*Sun Grid Engine 5.3 和 Sun Grid Engine，企業版 5.3 參考手冊*》中的 `sge_conf` 線上說明手冊）。

---

## 作業相依

建立聚合體作業最方便的方法通常是將作業分割為子作業。在這些例子中，子作業需靠其他子作業成功完成才能啟動。有一個範例是之前的作業產生後繼作業才能讀取和處理的輸出檔案。

Sun Grid Engine, Enterprise Edition 使用期作業相依功能來支援相依作業。作業可以配置來依靠一個或多重作業的成功完成。此功能由 `qsub -hold_jid` 選項實行。作業清單可指定讓提交作業依靠。作業清單也可以包含陣列作業的子集。除非相依清單中的所有作業都已經順利完成，否則提交作業將無法執行。

---

## 控制佇列

如第 50 頁的「佇列和佇列屬性」一節所述，佇列所有人有暫停 / 解除暫停或停用 / 啟動佇列的權限。如果這些使用者不時需要某些機器執行重要工作且如果他們會被在背景執行的 Sun Grid Engine, Enterprise Edition 作業強烈影響，這就會非常理想。

暫停或啟動佇列有兩個方法。

- QMON 佇列控制對話方塊
- `qmod` 指令

### ▼ 如何用 QMON 控制佇列

- 從 QMON 主要功能表上按下「佇列控制」鈕。

螢幕就會顯示「佇列控制」對話方塊，類似圖 5-11 所示的對話方塊。

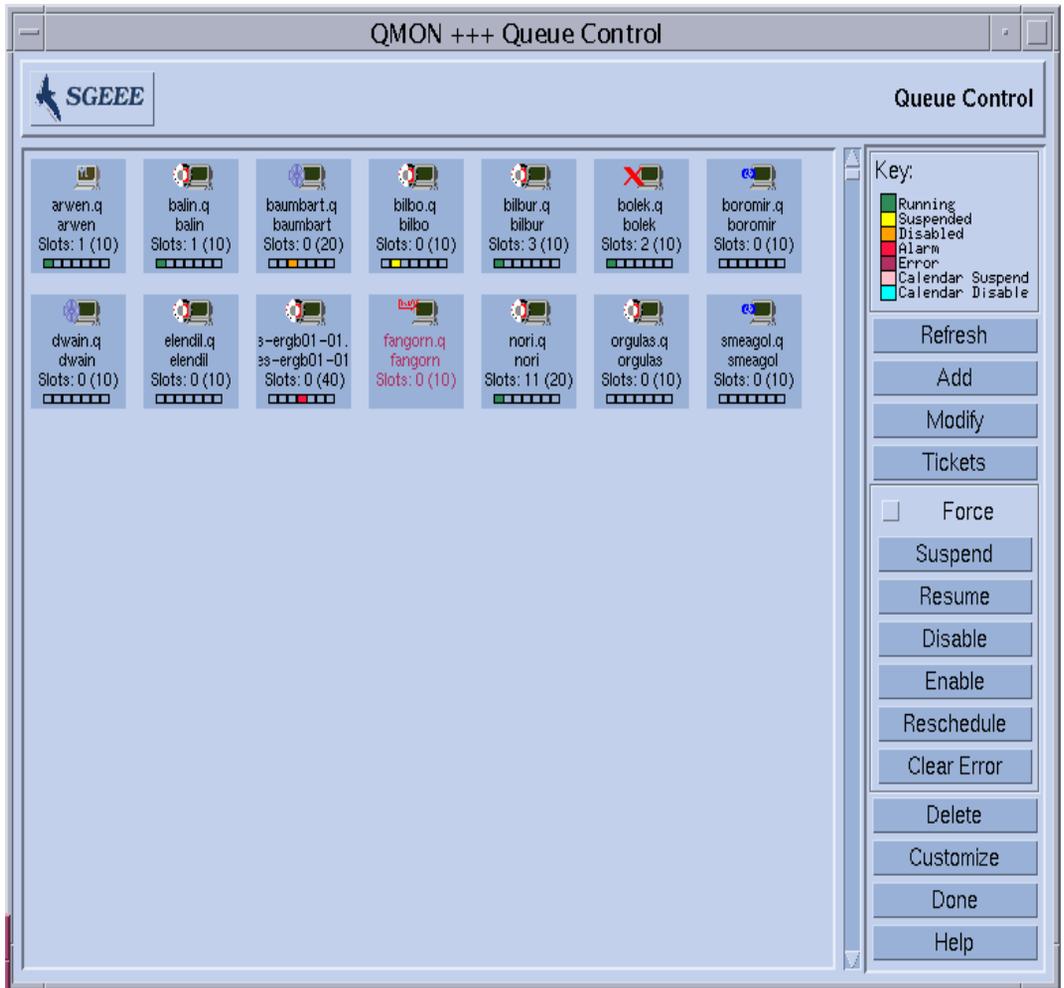


圖 5-11 佇列控制對話方塊

「佇列控制」對話方塊的目的是提供叢集中可用資源和活動的快速簡介。它也提供暫停 / 解除暫停、停用 / 啟動佇列，以及配置佇列的方法。每個顯示的圖示代表一個佇列。若主要顯示區域是空的，則表示沒有配置佇列。每個佇列圖示都貼上佇列名稱、佇列常駐的主機之名稱、和佔用的作業插槽數。如果 `sge_execd` 在佇列主機上執行，並已經用 `sge_qmaster` 登錄，則佇列圖示上的圖片表示佇列主機的作業系統架構，而圖示底部的指示條說明佇列的狀態。「佇列控制」對話方塊右邊的圖示顯示顏色代表的意義。

對這些佇列來說，使用者可以按住鍵盤上的「Shift」鍵，然後在佇列圖示上按一下滑鼠左鍵，就可取得佇列並隱含取得管理佇列的機器的目前屬性、負載和資源消耗的資訊。螢幕上將呈現類似於顯示於圖 5-12 的資訊畫面。

按一下滑鼠左鍵或點進一個包圍佇列圖示鈕的長方形區域就可選擇佇列。「刪除」、「暫停」、「解除暫停」、「停用」或「啟動」鈕可用來在選擇的佇列上執行對應的操作。暫停 / 解除暫停和停用 / 啟動操作需要通知對應的 `sge_execd`。如果這不可能做到（如，因為主機故障），則如果打開「強制」切換鈕，就可強制 `sge_qmaster` 內部狀態變更。

如果佇列暫停，則該佇列就會關閉不能用於以後的作業，而已經在佇列執行的作業也會如同一節第 106 頁的「如何使用 QMON 監視和控制作業」解釋一般暫停。佇列及其作業在佇列解除暫停時就會恢復。

---

**注意** – 如果在暫停佇列中的作業已另外明確暫停，則該作業在佇列解除暫停時也不會恢復。該作業需要再度明確解除暫停。

---

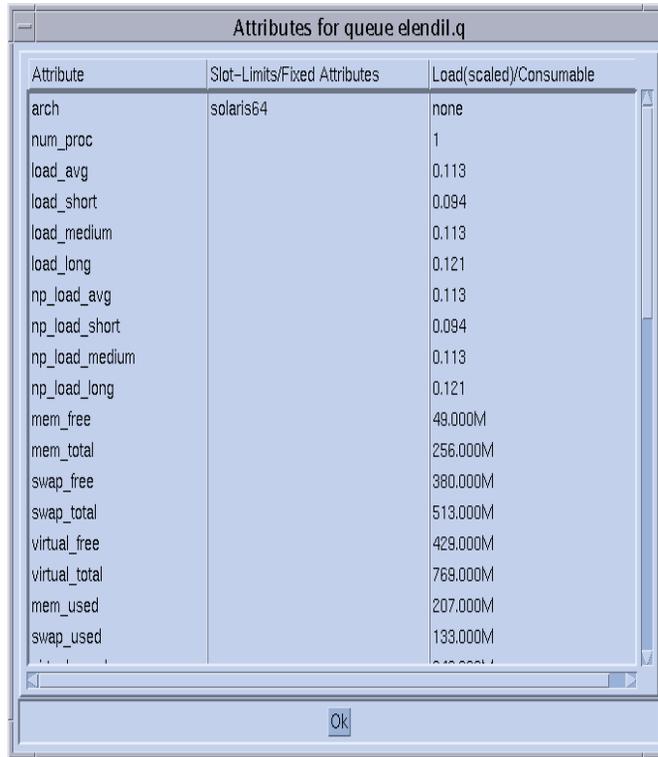
停用的佇列會關閉，但是，在這些佇列執行的作業被允許繼續進行。停用佇列通常是用來「耗盡」（「`drain`」）佇列。在啟動佇列後，佇列就可以再度用於作業執行。在仍舊執行的作業上沒有執行動作。

暫停 / 解除暫停和停用 / 啟動操作需要佇列所有人或 Sun Grid Engine, Enterprise Edition 管理者或操作者權限（請參閱第 61 頁的「管理員、操作員和所有人」一節）。

顯示於「佇列控制」對話方塊的資訊會定期更新。按「重新顯示」鈕可以強制更新。「完成」鈕會關閉對話方塊。

「自訂」鈕能讓您選擇透過過濾操作而顯示的佇列。圖 5-13 中的樣本畫面顯示只屬於架構 `osf4`（如，Compaq UNIX version 4）的主機上執行的佇列選項。「自訂」對話方塊的「儲存」鈕讓您將設定儲存在初始目錄中的 `.qmon_preferences` 檔中，在之後呼叫 QMON 做為標準重新啟動之用。

對配置的目的來說，在您按「佇列控制」畫面右邊的「新增」或「修改」鈕時，子對話方塊就會開啓（相關細節請參閱第 152 頁的「如何以 QMON 配置佇列」一節）。



| Attribute      | Slot-Limits/Fixed Attributes | Load(scaled)/Consumable |
|----------------|------------------------------|-------------------------|
| arch           | solaris64                    | none                    |
| num_proc       |                              | 1                       |
| load_avg       |                              | 0.113                   |
| load_short     |                              | 0.094                   |
| load_medium    |                              | 0.113                   |
| load_long      |                              | 0.121                   |
| np_load_avg    |                              | 0.113                   |
| np_load_short  |                              | 0.094                   |
| np_load_medium |                              | 0.113                   |
| np_load_long   |                              | 0.121                   |
| mem_free       |                              | 49.000M                 |
| mem_total      |                              | 256.000M                |
| swap_free      |                              | 380.000M                |
| swap_total     |                              | 513.000M                |
| virtual_free   |                              | 429.000M                |
| virtual_total  |                              | 769.000M                |
| mem_used       |                              | 207.000M                |
| swap_used      |                              | 133.000M                |

圖 5-12 佇列屬性顯示

所有附接佇列的屬性（包含從主機或叢集得到的）都列在「屬性」欄。插槽限制 / 固定屬性 (Slot-Limits/Fixed Attributes) 欄顯示已每個佇列插槽限制或以固定聚合體屬性定義的屬性值。負載（依比例） / 可消耗的欄位說明報告的（如依比例配置）負載參數（請參閱第 245 頁的「載入參數」一節）及根據 Sun Grid Engine, Enterprise Edition 可消耗資源功能的可用資源容量（請參閱第 234 頁的「可耗損資源」一節）。

---

**注意** – 若負載屬性配置為可消耗的資源，則負載報告和可消耗容量可能相互覆寫。顯示的是兩者的最小值，通常用於作業分派演算法。

---

**注意** – 顯示的負載和可消耗數值目前不考慮如第 22 頁的「執行主機」一節所述的負載調整修正。

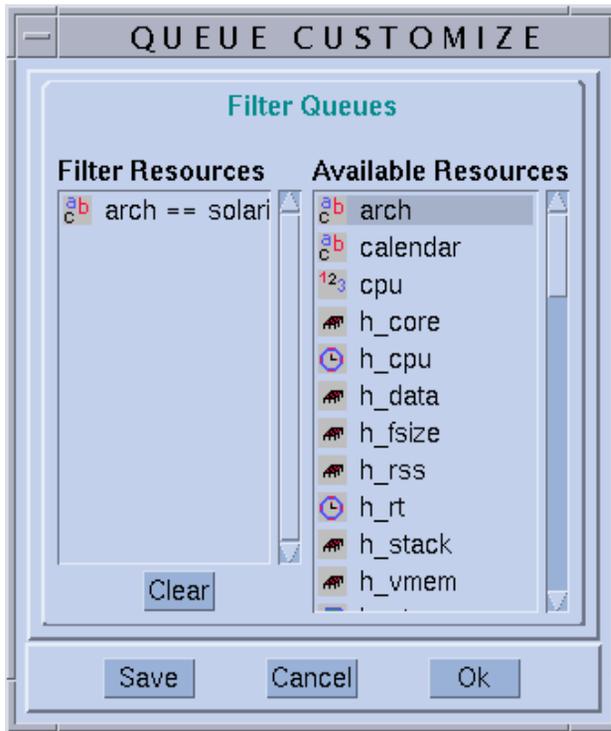


圖 5-13 佇列控制自訂

## ▼ 如何用 qmod 控制佇列

第 119 頁的「如何從指令行控制作業」說明 Sun Grid Engine, Enterprise Edition 指令 qmod 如何用來暫停 / 解除暫停 Sun Grid Engine, Enterprise Edition 作業。然而，qmod 指令會額外提供使用者暫停 / 解除暫停或停用 / 啟動佇列的方法。

- 透過詳述於以下各節的資訊的引導，用合適的引數輸入以下指令。

```
% qmod 引數
```

以下指令為 `qmod` 用於此目的的範例：

```
% qmod -s q 名稱
% qmod -us -f q 名稱1、q 名稱2
% qmod -d q 名稱
% qmod -e q 名稱1、q 名稱2、q 名稱3
```

前兩個指令暫停或解除暫停佇列，而第三和第四個指令停用並啟動佇列。第二個指令除了強制登錄 `sge_qmaster` 的狀態變更外，還使用 `qmod -f` 選項（萬一對應的 `sge_execd` 因網路問題而無法達成時。）

---

**注意** – 暫停 / 解除暫停和停用 / 啟動佇列需要佇列所有人或 Sun Grid Engine, Enterprise Edition 管理者或操作者權限（請參閱第 61 頁的「管理員、操作員和所有人」一節）。

---

---

**注意** – 您可以用 `crontab` 或 `at` 作業使用 `qmod` 指令。

---

## 自訂 QMON

QMON 的觀感大部分由特別設計的資源檔案所定義。參考的資源檔是放在 `<sge_root>/qmon/Qmon`，而合理的預設值已經編譯在其中。

叢集管理可以藉由將 QMON 特定資源定義包含到標準 `.Xdefaults` 或 `.Xresources` 檔案中，或是將站點特定 `Qmon` 檔案放到標準搜尋路徑，如 `XAPPLRESDIR`，所提及的位置上，來把站點特定預設值安裝在標準位置上，如 `/usr/lib/X11/app-defaults/Qmon`。詢問您的管理員看看上述的任何事物是否和您的例子有關，

此外，使用者可以藉由將 `Qmon` 檔案複製並修改至起始目錄（或至另一個私人 `XAPPLRESDIR` 搜尋路徑所指定的位置）或將必要的資源定義包含進使用者私人的 `.Xdefaults` 或 `.Xresources` 檔案中，來配置個人的喜好設定。私人的 `Qmon` 資源檔案也可以在作業或 `X11` 環境的（例如在 `.xinitrc` 資源檔案中）啟動期間透過 `xrdb` 指令安裝。

欲取得可能自訂的詳細資訊，請參閱參考 `Qmon` 檔案裡的注釋行。

另一個自訂 QMON 的方式已為圖 5-2 和圖 5-13 裡的「作業控制」和「佇列控制自訂」對話方塊做了解釋。在兩個對話方塊中，您可用「儲存」鈕儲存過濾並顯示配置自訂對話方塊至使用者起始目錄下的檔案，`.qmon_preferences` 的定義。重新開始之際，QMON 會讀取這個檔案並重新啟動先前定義的運作方式。



## 部分 IV 管理

---

《Sun Grid Engine，企業版 5.3 管理和使用者指南》的這個部分包含六個章節，是針對給管理者所使用的。

■ 第六章 — 第 129 頁的「主機和叢集配置」

本章提供配置 Sun Grid Engine，企業版 5.3 主機和叢集的一般背景和詳盡的指示。

■ 第七章 — 第 151 頁的「配置佇列與佇列行事曆」

本章包含佇列的重要概念的說明 — 佇列的功用是不同種類的 Sun Grid Engine，企業版 5.3 作業的「容器」。包含完整的配置佇列的指示。

■ 第八章 — 第 225 頁的「聚合體概念」

本章解釋 Sun Grid Engine，企業版 5.3 系統如何使用聚合體 (*complex*) 來定義所有使用者可能要求的作業資源屬性之相關資訊。管理員會配置不同的聚合體以符合環境的需求，而本章提供此做法的詳盡指示。

■ 第九章 — 第 169 頁的「管理使用者存取與策略」

本章提供透過 Sun Grid Engine，企業版 5.3 系統取得的使用者策略類型的完整背景資訊，並提供如何將這些策略配對到計算環境的指示。

■ 第十章 — 第 251 頁的「管理平行環境」

除了說明 Sun Grid Engine，企業版 5.3 系統如何與平行環境一致之外，本章還提供處理它們的完整配置指示。

■ 第十一章 — 第 261 頁的「錯誤訊息與疑難排解」

本章解釋 Sun Grid Engine，企業版 5.3 取得錯誤訊息的程序，並說明如何在除錯模式中執行軟體。



# 主機和叢集配置

---

本章提供配置系統 Sun Grid Engine，企業版 5.3 不同層面的背景資訊和指示。您會在本章找到以下作業的指示。

- 第 132 頁的「如何以 QMON 配置管理主機」
- 第 134 頁的「如何刪除管理主機」
- 第 134 頁的「如何新增管理主機」
- 第 134 頁的「如何從指令行配置管理主機」
- 第 134 頁的「如何以 QMON 配置提交主機」
- 第 136 頁的「如何刪除提交主機」
- 第 136 頁的「如何新增提交主機」
- 第 136 頁的「如何從指令行配置提交主機」
- 第 136 頁的「如何以 QMON 配置執行主機」
- 第 138 頁的「如何刪除執行主機」
- 第 138 頁的「如何關閉執行主機常駐程式」
- 第 138 頁的「如何新增或修改執行主機」
- 第 142 頁的「如何從指令行配置執行主機」
- 第 143 頁的「如何使用 qhost 監視執行主機」
- 第 144 頁的「如何從指令行刪除常駐程式」
- 第 145 頁的「如何從指令行重新啓動常駐程式」
- 第 145 頁的「如何從指令行顯示基礎叢集配置」
- 第 146 頁的「如何從指令行修改基礎叢集配置」
- 第 147 頁的「如何使用 QMON 顯示叢集配置」
- 第 147 頁的「如何使用 QMON 刪除叢集配置」
- 第 147 頁的「如何使用 QMON 顯示全域叢集配置」
- 第 148 頁的「如何使用 QMON 修改全域和主機配置」

## 關於主要和備用主要配置

備用主要主機名稱檔案，`<sgc 根目錄>/<儲存格>/common/shadow_masters` 包含主要的主要主機（Sun Grid Engine，企業版主要常駐程式 `sgc_qmaster` 初始啟動的機器）和備用主要主機的名稱。主要主機名稱檔案的格式如下。

- 檔案的第一行定義主要的主要主機
- 以下各行指定備用主要主機，每行一個

（備用）主要主機出現的順序是很重要的。如果主要的主要主機（檔案的第一行）無法繼續進行，則定義在第二行的備用主要主機將會接替。如果這個也故障，則定義在第三行的則會進行工作，以此類推。

若要將主機準備成 Sun Grid Engine，企業版備用主機，需先符合以下的需求：

- 備用主機需要執行 `sgc_shadowd`。
- 備用主要主機需要共享 `sgc_qmaster` 的狀態資訊、作業和紀錄至磁碟的佇列配置。特別是，（備用）主要主機需要至主機的多工緩衝處理目錄和 `<sgc 根目錄>/<儲存格>/common` 目錄的讀取/寫入 `root` 存取權。
- 備用主要主機名稱檔案必須包含定義主機為備用主要主機的行列。

一旦符合這些要求，則該主機的備用主要主機功能就會啟動。要取動這個功能不需重新啟動 Sun Grid Engine，企業版常駐程式。

備用主要主機上 `sgc_qmaster` 的自動錯誤修復啟動需要花費一些時間（已一分鐘為順序）。在此同時，每當執行 Sun Grid Engine，企業版指令時，您就會得到對應的錯誤訊息。

---

**注意** - `<sgc 根目錄>/<儲存格>/common/act_qmaster` 檔案包含實際執行 `sgc_qmaster` 常駐程式的主機名稱。

---

爲了要能啟動備用 `sgc_qmaster`，Sun Grid Engine，企業版必須確定原來的 `sgc_qmaster` 已經終止或它將會終止，且不會執行動作來干預剛啟動的備用 `sgc_qmaster`。這個很少無法達成。在這些例子中，對應的錯誤訊息將紀錄在備用主要主機上 `sgc_shadowd` 的訊息日誌檔（請參閱第十一章第 261 頁的「錯誤訊息與疑難排解」）而任何開啓至 `sgc_qmaster` 常駐程式的 `tcp` 連線都會永久失敗。如果發生的話，請確認主要常駐程式現在沒有在執行，並在任何的備用主要機器上手動重新啟動 `sgc_qmaster`（請參閱第 144 頁的「如何從指令行刪除常駐程式」一節）。

---

## 關於常駐程式和主機

Sun Grid Engine，企業版 根據系統上執行的常駐程式以及主機登錄方法，主機可以分為四個群組 `sge_qmaster`

- **主要主機** — 主要主機是全部叢集活動的中心。它會執行主要常駐程式 `sge_qmaster`。`sge_qmaster` 會控制所有 Sun Grid Engine，企業版元件，如佇列和作業，並維護組件狀態、使用者存取權等等的表格。第 27 頁的「如何安裝主要主機」一節說明如何初始設定主要主機，而第 130 頁的「關於主要和備用主要配置」一節顯示如何配置動態主要主機變更。主要主機通常執行 Sun Grid Engine，企業版排程器 `sge_schedd`。主要主機不需要由安裝程序之外所執行的進一步配置。
- **執行主機** — 執行主機通常是有執行 Sun Grid Engine，企業版作業權限的節點。因此，它們管理 Sun Grid Engine，企業版佇列並執行 Sun Grid Engine，企業版執行常駐程式，`sge_execd`。執行主機由第 27 頁的「如何安裝執行主機」一節所述的執行主機安裝程序進行初始設定。
- **管理主機** — 權限可以給予主要主機之外的其他主機，好在 Sun Grid Engine，企業版執行任何種類的管理活動。管理主機可以用以下的指令設定：

```
qconf -ah 主機名稱
```

細節請參閱 `qconf` 線上說明手冊。

- **提交主機** — 提交主機只允許提交和控制批次作業。特別是登入提交主機的使用者可以透過 `qsub` 提交作業、可以透過 `qstat` 控制作業狀態或執行 Sun Grid Engine，企業版的 OS/1 Motif 圖形使用者介面，`QMON`。提交主機可以用以下的指令設定：

```
qconf -as 主機名稱
```

細節請參閱 `qconf` 線上說明手冊。

---

**注意** — 主機在上述的類別中可以歸類到一個以上的類別。主要主機是預設的管理和提交主機。

---

## 關於配置主機

Sun Grid Engine，企業版維護主要主機之外所有主機種類的物件清單。若是管理和提交主機，則這些清單只提供主機是否有管理或提交權限的資訊。對於執行主機物件而言，其它的引數，如主機所在處的 `sge_execd` 提供的下載資訊，及 Sun Grid Engine，企業版管理員提供的下載引數檢核因素。

以下各節解釋如何使用 Sun Grid Engine，企業版圖形使用者介面，`QMON`，的說明並從指令行配置不同的主機物件。

GUI 管理由一套主機配置對話方塊所提供，其可以按 QMON 主功能表的「主機配置」圖示鈕呼叫。可用的對話方塊為管理主機配置（請參閱圖 6-1）、提交主機配置（請參閱圖 6-2）和執行主機配置（請參閱圖 6-3）。對話方塊可使用螢幕頂端的選擇清單鈕切換。

qconf 指令為主機物件管理提供指令行介面。

## 無效的主機名稱

以下為無效、保留或其他不准使用的主機名稱清單。

- global
- 平台
- all
- default
- unknown
- none

## ▼ 如何以 QMON 配置管理主機

1. 按一下 QMON 主功能表頂端的「管理主機」標籤。  
「管理主機配置」對話方塊（與下圖類似）開啓。

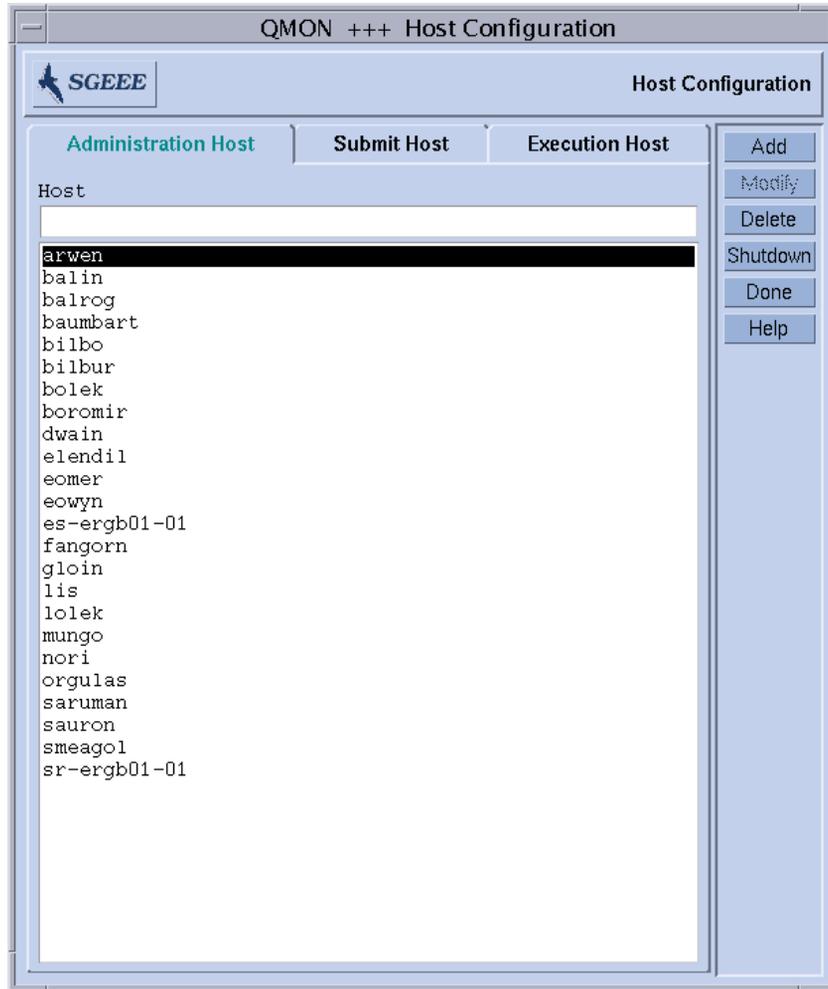


圖 6-1 管理主機配置對話方塊

---

**注意** - 當第一次按下「主機配置」鈕時，預設就會開啓「管理主機配置」對話方塊。

---

2. 根據您要配置主機的方式，依照以下各節的指導進行。

有了這個對話方塊，您可以從允許 Sun Grid Engine，企業版管理指令的主機配置主機。在螢幕中間的選擇清單會顯示已宣告可提供管理權限之主機。

## ▼ 如何刪除管理主機

- 在現存主機名稱上按一下滑鼠左鍵，然後按對話方塊底部的「刪除」鈕就可從該清單刪除現存主機。

## ▼ 如何新增管理主機

- 將名稱輸入「主機名稱」輸入視窗，然後按「新增」鈕或「Return」鍵增加新的主機。

## ▼ 如何從指令行配置管理主機

- 依照您想如何配置主機的方式，用合適的引數輸入以下指令。

```
% qconf 引數
```

qconf 指令的引數和其結果如下。

- qconf -ah *主機名稱*  
新增管理主機 — 將指定主機新增至管理主機的清單。
- qconf -dh *主機名稱*  
刪除管理主機 — 將指定主機從管理主機的清單刪除。
- qconf -sh  
顯示管理主機 — 顯示所有目前配置的管理主機清單。

## ▼ 如何以 QMON 配置提交主機

1. 按一下 QMON 主功能表頂端的「提交主機」標籤。

「提交主機配置」對話方塊（與下圖類似）開啓。

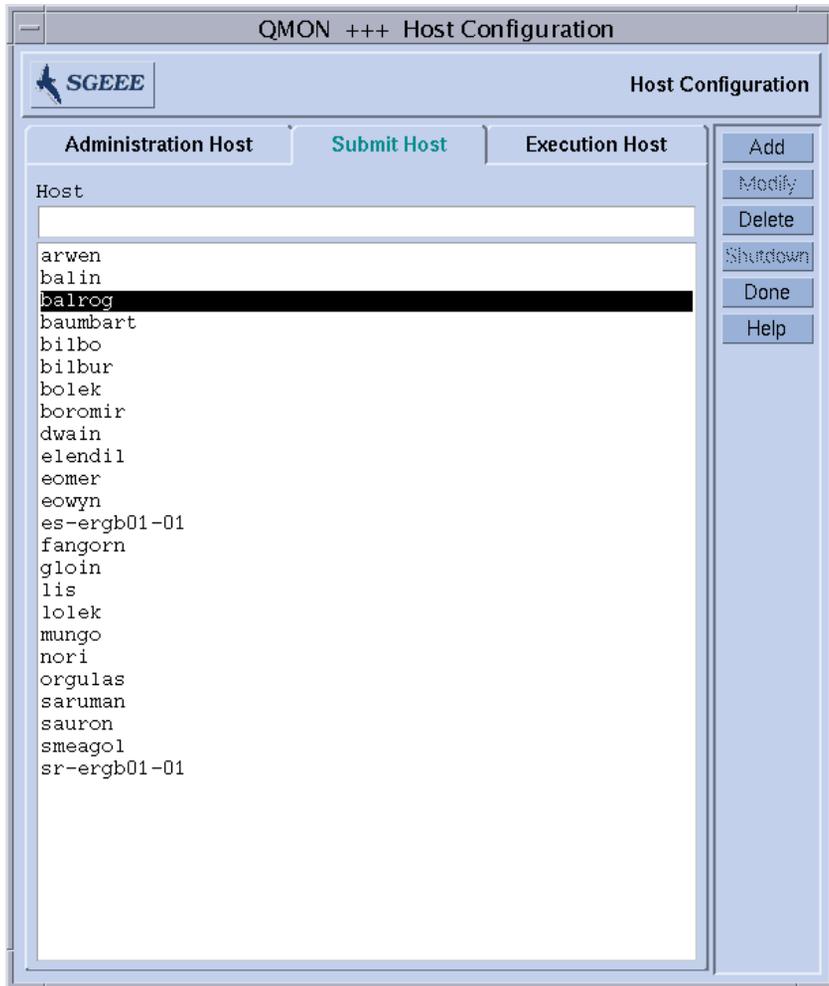


圖 6-2 提交主機配置

## 2. 根據您要配置主機的方式，依照以下各節的指導進行。

使用此對話方塊，您可以從提交、監視和控制作業的主機宣告主機。這些主機不允許管理 Sun Grid Engine，企業版指令，除非他們也被宣告為管理主機（請參閱第 132 頁的「如何以 QMON 配置管理主機」）。在螢幕中間的選擇清單會顯示已宣告可提供提交權限之主機。

## ▼ 如何刪除提交主機

- 按一下「提交主機」對話方塊中現存主機的名稱，然後按對話方塊底部的「刪除」鈕就可從該清單刪除現存主機。

## ▼ 如何新增提交主機

- 將名稱輸入「提交主機」對話方塊中的「主機名稱」輸入視窗，然後按「新增」鈕或「Return」鍵增加新的主機。

## ▼ 如何從命令行配置提交主機

- 依照您想如何配置主機的方式，用合適的引數輸入以下指令。

```
% qconf 引數
```

qconf 指令的引數和其結果如下。

- `qconf -as 主機名稱`  
新增提交主機 — 將指定主機新增至提交主機的清單。
- `qconf -ds 主機名稱`  
刪除提交主機 — 將指定主機從提交主機的清單刪除。
- `qconf -ss`  
顯示提交主機 — 顯示所有目前配置提供提交權限的主機名稱清單。

## ▼ 如何以 QMON 配置執行主機

1. 按一下 QMON 主功能表頂端的「執行主機」標籤。  
「執行主機配置」對話方塊（與下圖類似）開啓。

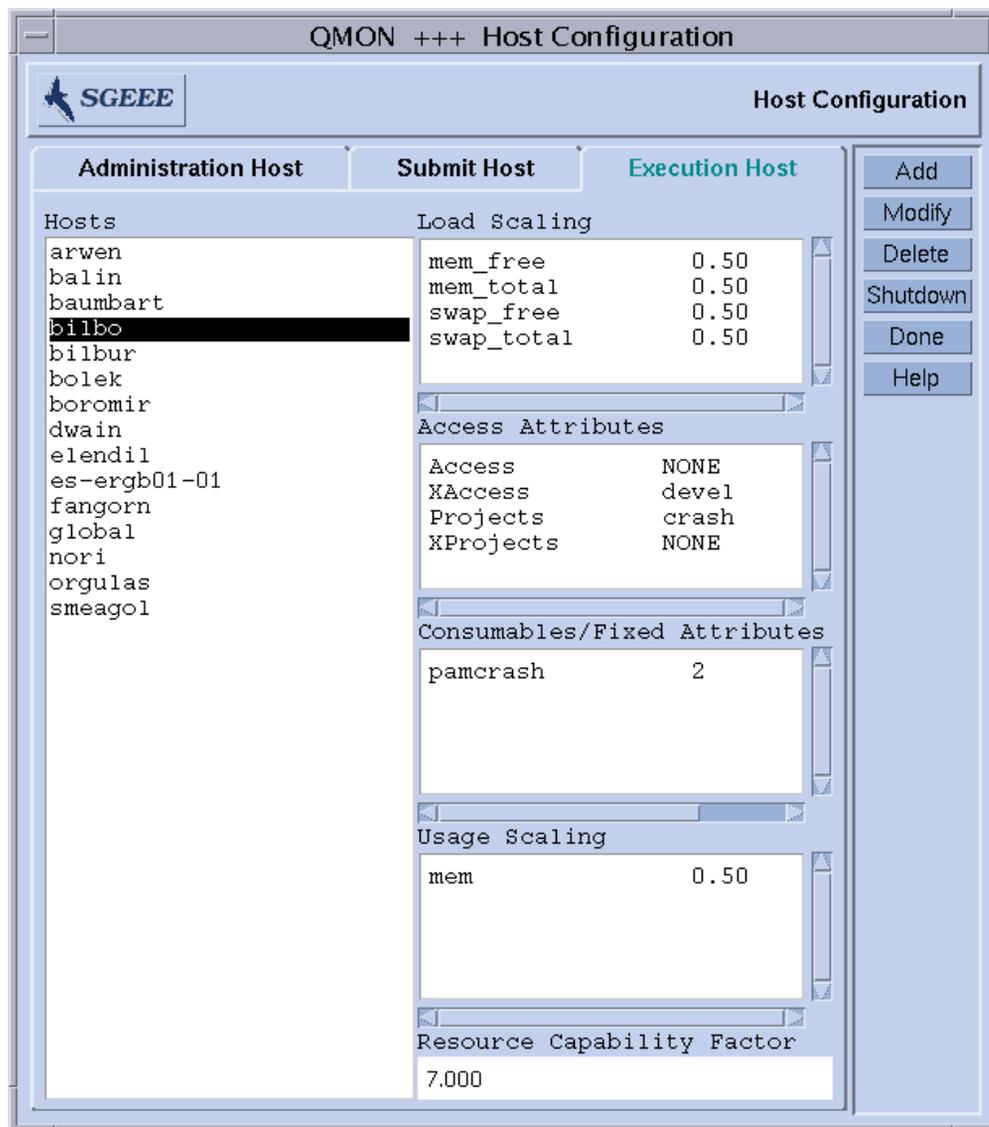


圖 6-3 執行主機配置

2. 根據您要配置主機的方式，依照以下各節的指導進行。

Sun Grid Engine，企業版執行主機可以從此對話方塊配置。這些主機不會自動允許管理或提交指令，除非他們也被宣告為管理或提交主機（請參閱第 132 頁的「如何以 QMON 配置管理主機」和第 134 頁的「如何以 QMON 配置提交主機」）。

主機選擇清單顯示已經定義的執行主機。目前配置給主機相關的消耗和固定聚合體屬性的負載換算因數 (load scaling factors)、存取權限和資源可用性顯示在「負載換算」(Load Scaling)，「存取屬性」和「可消耗 / 固定屬性」會顯示選取執行主機的視窗。要取得聚合體屬性、使用者權限和負載參數的細節，請參考第 225 頁的「關於聚合體」、第 59 頁的「使用者存取權限」和第 245 頁的「載入參數」等節。

對 Sun Grid Engine，企業版來說，額外的「用法換算」(Usage Scaling) 顯示視窗包含目前不同機器的獨立用法矩陣 CPU、記憶體和 I/O 的換算因數 (scaling factor)。

sge\_execd 會定期報告每個目前執行的作業之資源使用。換算因數指出執行作業的使用者或專案的特定機器之資源使用的相關花費。例如，它們可以用來比較在 400 MHz 處理器和 600 MHz CPU 之間 CPU 時間的每秒費用。沒在「用法換算」視窗顯示的矩陣的換算因數為「1」。

「資源容量因數」欄位在 Sun Grid Engine，企業版也是額外的，可由排程器在作業放置 (job placement) 時使用。那是與指出排程目的的全部相關電源之主機有關的單一數字。可能組成選為資源容量因數的數值之因數包括 CPU 的數目、CPU 時脈速度、CPU 種類、可用記憶體的數量、連接裝置的速度等等。

## ▼ 如何刪除執行主機

- 在「執行主機」對話方塊中，按一下要刪除的「執行主機」名稱，然後按對話方塊右邊按鈕欄中的「刪除」鈕。

## ▼ 如何關閉執行主機常駐程式

- 對任何選取的主機來說，請按「執行主機」對話方塊的「關閉」鈕。

## ▼ 如何新增或修改執行主機

1. 請按「執行主機」對話方塊按鈕欄中的「新增」或「修改」鈕。  
螢幕上將出現類似於顯示於圖 6-4 的對話方塊。

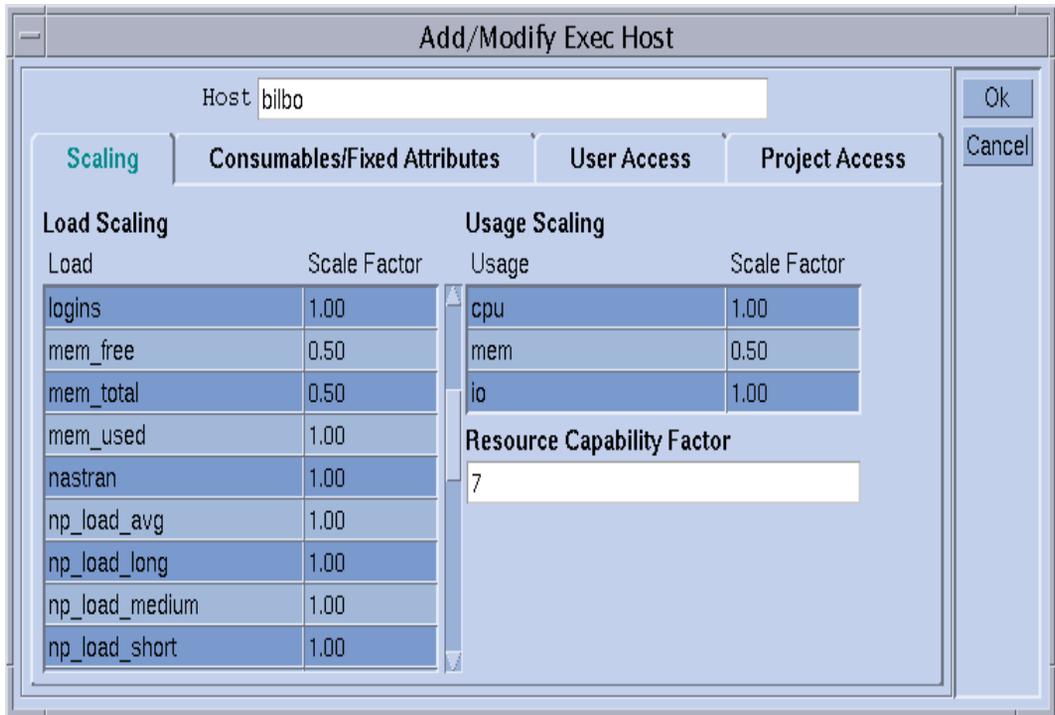


圖 6-4 修改負載換算

2. 根據您要修改主機的方式，依照以下各節的指導進行。

用來新增新的執行主機或修改現存執行主機的配置的對話方塊可進行所有與主機相關之屬性修改。執行主機的名稱可以顯示或新增進「主機」輸入視窗。您可以選擇對話方塊中的「換算」標籤來定義換算因數（請參閱圖 6-4）。

所有可用的負載參數都顯示於「負載換算」表的「負載」欄中，而對應的換算定義可在「換算因數」欄找到。您可以編輯「換算因數」欄。有效的換算因數是固定點表示法或科學記號中正值浮點記號。

對 Sun Grid Engine，企業版來說，目前用於用法矩陣 CPU、記憶體和 I/O 的換算因數顯示於「用法換算」表的「用法」欄中，而對應的換算定義可在「換算因數」欄找到。您可以編輯「換算因數」欄。有效的換算因數是固定點表示法或科學記號中正值浮點記號。

此外，資源容量因數可以指定給 Sun Grid Engine，企業版「資源容量因數」輸入欄位中的主機。有效的因數會再度會是科學記號固定點表示法中的正值浮點記號。

若在 tab widget 中選擇「可消耗 / 固定屬性」，則可定義與主機相關的聚合體屬性（請參閱圖 6-5）。與主機相關的聚合體（請參閱第 225 頁的「關於聚合體」一節）是全域和主機聚合體或是透過對話方塊左側底部的「聚合體選擇」區而連接主機的**管理員定義聚合體**。可用的管理員定義聚合體顯示在左邊，而其可以透過紅色箭頭連接或脫離。若您需要目前聚合體配置的進一步資訊或您想進行修改，「聚合體」配置圖示鈕可打開最上層的「聚合體配置」對話方塊。

對話方塊右側底部區域的「可消耗 / 固定屬性」表列出目前所有數值定義的聚合體屬性。按一下上方的「名稱」或「數值」鈕來增加清單。之後會開啓一選擇清單，列出所有隨附於主機的屬性（如上述之配置於全域、主機、與附接於該主機之管理員定義聚合體的所有屬性的交集 (union)）。「屬性選擇」對話方塊圖顯示於圖 6-6。選擇其中一個屬性，並按「確定」鈕確認選項，就會將此屬性新增至「可消耗 / 固定屬性」表中的「名稱」欄，並將指標放入對應的「數值」欄位中。您可以在「數值」欄位上連接兩下滑鼠左鍵來修改現存的數值。您可用滑鼠左鍵先選擇對應表格的行列來刪除屬性。鍵入 CTRL-D 或按滑鼠右鍵開啓刪除方塊並確認刪除動作，就可刪除選取的清單記錄。

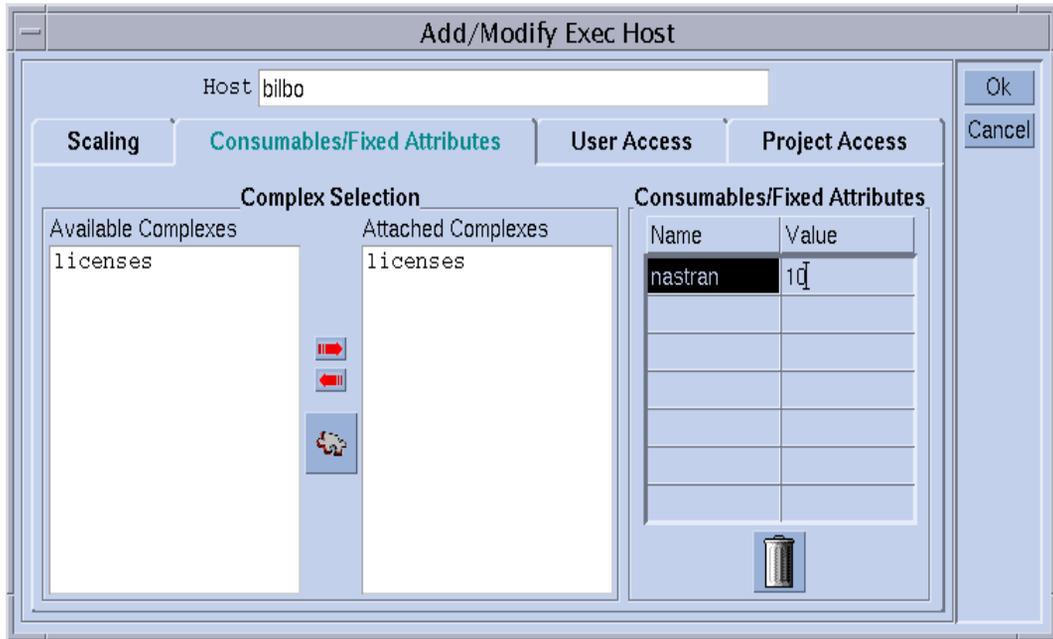


圖 6-5 修改可消耗 / 固定屬性

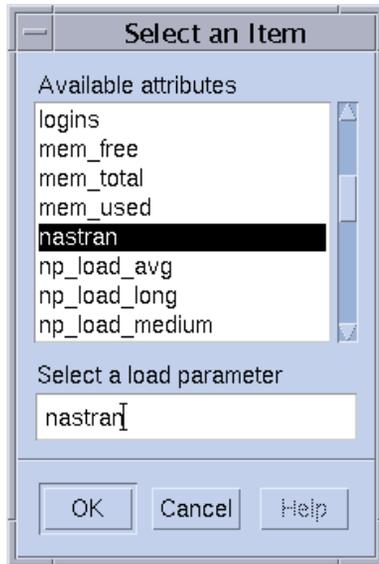


圖 6-6 可用的聚合體屬性

藉選擇「使用者存取」標籤（圖 6-7），您可以定義以之前配置的使用者存取清單為基礎之執行主機的存取權。

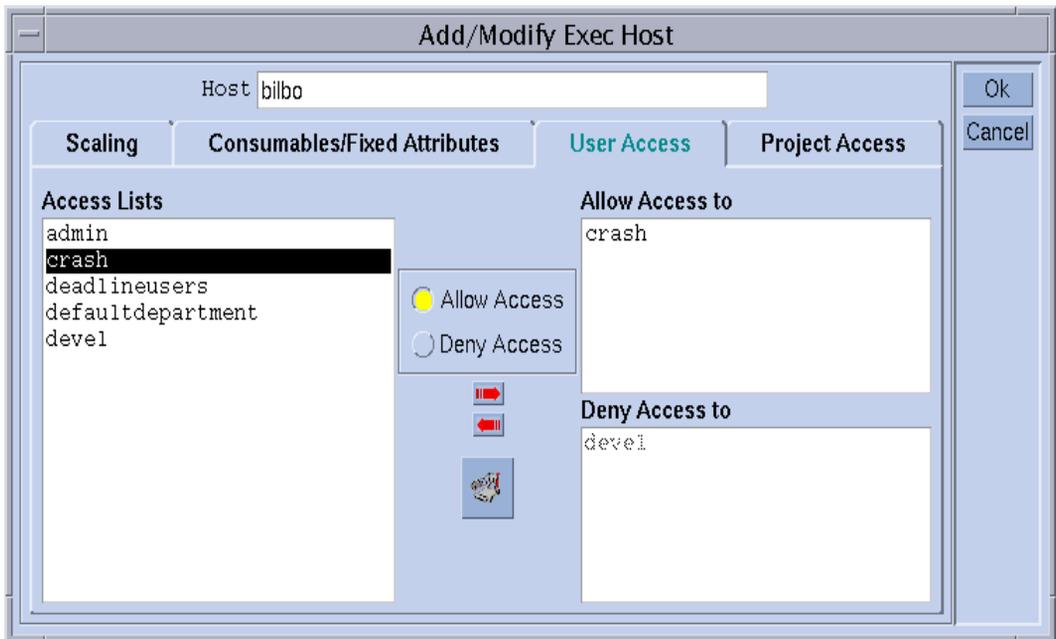


圖 6-7 修改使用者存取

藉選擇「專案存取」標籤（圖 6-8），您可以定義以之前配置的專案為基礎之執行主機的存取權。

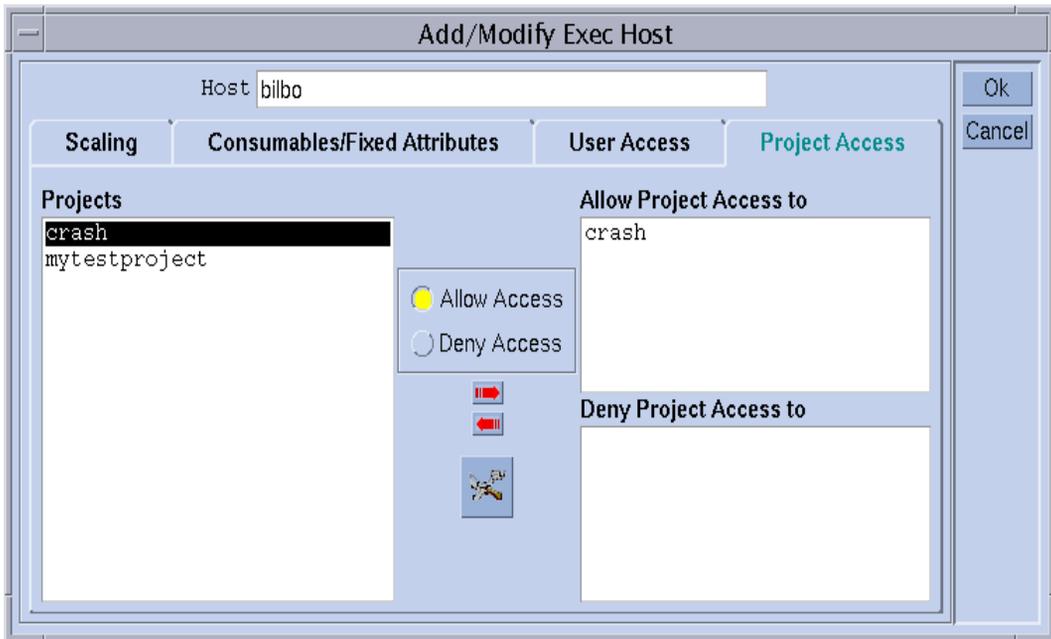


圖 6-8 修改專案存取

## ▼ 如何從指令行配置執行主機

- 依照您想如何配置主機的方式，用合適的引數輸入以下指令。

```
% qconf 引數
```

維護執行主機清單的指令行介面由以下選項提供給 qconf 指令。

- qconf -ae [ 執行主機範本 ]

新增執行主機 — 使用執行主機配置樣板叫出編輯器（預設的 vi 或對應至 \$EDITOR 環境變數）。若選擇性參數 *執行主機樣板*（已配置執行主機的名稱）已經存在，則該執行主機的配置可當成樣板使用。變更樣板並儲存至磁碟中可配置執行主機。請參閱《Sun Grid Engine 5.3 和 Sun Grid Engine，企業版 5.3 參考手冊》的 host\_conf 項目以取得將變更的樣板項目之詳細說明。

- qconf -de *主機名稱*

刪除執行主機 — 將指定主機從執行主機的清單刪除。執行主機配置中的所有項目都已遺失。

- `qconf -me 主機名稱`

修改執行主機 — 使用指定執行主機的配置做為樣板叫出編輯器（預設的 `vi` 或對應至 `$EDITOR` 環境變數）。變更樣板並儲存至磁碟中就可修改執行主機配置。請參閱《*Sun Grid Engine 5.3 和 Sun Grid Engine, 企業版 5.3 參考手冊*》的 `host_conf` 線上說明手冊以取得將變更的樣板項目之詳細說明。

- `qconf -Me 檔案名稱`

修改執行主機 — 使用 *檔案名稱* 的內容做為執行主機配置樣板。指定檔案中的配置必須參照現存的執行主機。該執行主機的配置可用檔案內容置換。這個 `qconf` 選項對離線的執行主機配置變更更有效；如在 `cron` 作業中，因為其不需要手動互動 (`manual interaction`)。

- `qconf -se 主機名稱`

顯示執行主機 — 如 `host_conf` 所定義的顯示指定執行主機的配置。

- `qconf -sel`

顯示執行主機清單 — 顯示配置為執行主機的主機名稱清單。

## ▼ 如何使用 `qhost` 監視執行主機

`qhost` 指令提供簡單取得執行主機狀態之快速簡介的方法。

- 輸入以下指令。

```
% qhost
```

該指令產出類似下列的輸出。

表 6-1 範例 qhost 輸出

| HOSTNAME          | ARCH     | NPROC | LOAD | MEMTOT | MEMUSE | SWAPTO | SWAPUS |
|-------------------|----------|-------|------|--------|--------|--------|--------|
| global            | -        | -     | -    | -      | -      | -      | -      |
| BALROG.genias.de  | solaris6 | 2     | 0.38 | 1.0G   | 994.0M | 900.0M | 891.0M |
| BILBUR.genias.de  | solaris  | 1     | 0.18 | 96.0M  | 70.0M  | 164.0M | 9.0M   |
| DWAIN.genias.de   | irix6    | 1     | 1.13 | 149.0M | 55.8M  | 40.0M  | 0.0    |
| GLOIN.genias.de   | osf4     | 2     | 0.05 | 768.0M | 701.0M | 1.9G   | 13.5M  |
| SPEEDY.genias.de  | alinux   | 1     | 0.08 | 248.8M | 60.6M  | 125.7M | 232.0K |
| SARUMAN.genias.de | solaris  | 1     | 0.11 | 96.0M  | 77.0M  | 192.0M | 9.0M   |
| FANGORN.genias.de | linux    | 1     | 2.01 | 124.8M | 49.9M  | 127.7M | 4.3M   |

要取得輸出格式的說明和更多的選項，請參考《*Sun Grid Engine 5.3 和 Sun Grid Engine*，企業版 5.3 參考手冊》的 qhost 項目。

## ▼ 如何從命令行刪除常駐程式

- 請使用以下其中一項指令。請注意，您需要 Sun Grid Engine，企業版管理員或操作者權限才能執行這些作業（請參閱第九章第 169 頁的「管理使用者存取與策略」）。

```
% qconf -kej  
% qconf -ks  
% qconf -km
```

- 第一個指令將刪除所有目前活動中的作業並結束所有的 Sun Grid Engine，企業版執行常駐程式。

---

**注意** – 若使用 qconf -ke 置換該指令，則 Sun Grid Engine，企業版執行常駐程式將中斷，但是活動中的作業不會取消。當 sge\_execd 沒在該系統上執行時，結束的作業會到 sge\_execd 再度重新啟動後才會報告給 sge\_qmaster。然而，作業報告不會遺失。

---

- 第二個指令會關閉 Sun Grid Engine，企業版排程器 sge\_schedd。
- 第三個指令會強制 sge\_qmaster 過程終止。

若您有執行作業而您想用 Sun Grid Engine，企業版的關機程序等到目前活動的作業都結束，您可以在執行上述的 `qconf` 序列前，針對每個佇列使用以下的指令。

```
% qmod -d 佇列名稱
```

`qmod` 停用指令會防止新作業排程進停用的佇列。然後您應該等候刪除常駐程式，直到佇列中不再有作業執行。

## ▼ 如何從指令行重新啟動常駐程式

1. 以 `root` 身分登入您想要 Sun Grid Engine，企業版 5.3 常駐程式重新啟動的機器。
2. 執行以下的 `script`。

```
% <sge 根目錄>/<儲存格>/common/rcsgc
```

該 `script` 會尋找在該主機正常執行的常駐程式，然後啟動對應的常駐程式。

---

## 基礎叢集配置

基礎的 Sun Grid Engine，企業版叢集配置是一套像程式有效路徑如 `mail` 或 `xterm` 配置來反映站點相依和影響 Sun Grid Engine，企業版運作方式的資訊。提供的全域配置是用於 Sun Grid Engine，企業版儲存區中的 Sun Grid Engine，企業版主要主機和每個主機。此外，Sun Grid Engine，企業版系統可以配置使用每個主機的本端配置來置換全域配置中的特別項目。

《*Sun Grid Engine 5.3* 和 *Sun Grid Engine，企業版 5.3 參考手冊*》的 `sge_conf` 項目包含配置項目的詳細說明。Sun Grid Engine，企業版叢集管理員應在安裝後直接根據站點需求調整全域和本端配置，並在之後維持更新狀態。

## ▼ 如何從指令行顯示基礎叢集配置

顯示目前配置的 Sun Grid Engine，企業版指令是 `qconf` 程式的顯示配置選項。以下為幾個範例（詳細說明請參閱《*Sun Grid Engine 5.3* 和 *Sun Grid Engine，企業版 5.3 參考手冊*》）。

- 請輸入以下其中一項指令。

```
% qconf -sconf
% qconf -sconf global
% qconf -sconf < 主機 >
```

前兩個指令是相等的，並顯示全域配置。第三個指令顯示主機的本端配置。

## ▼ 如何從指令行修改基礎叢集配置

---

**注意** – 用來變更叢集配置的 Sun Grid Engine，企業版指令 `qconf` 只能為 Sun Grid Engine，企業版管理員使用。

---

- 請輸入以下其中一項指令。

```
% qconf -mconf global
% qconf -mconf < 主機 >
```

- 第一個指令範例修改全域配置。
- 第二個範例在指定的執行或主要主機的本端配置上操作。

以上兩個指令為許多可用 `qconf` 指令的範例。其他指令請參考 《*Sun Grid Engine 5.3 和 Sun Grid Engine，企業版 5.3 參考手冊*》。

## ▼ 如何使用 QMON 顯示叢集配置

1. 在 QMON 主要功能表中按下「叢集配置」鈕。  
螢幕上顯示類似圖 6-9 範例的「叢集配置」對話方塊。

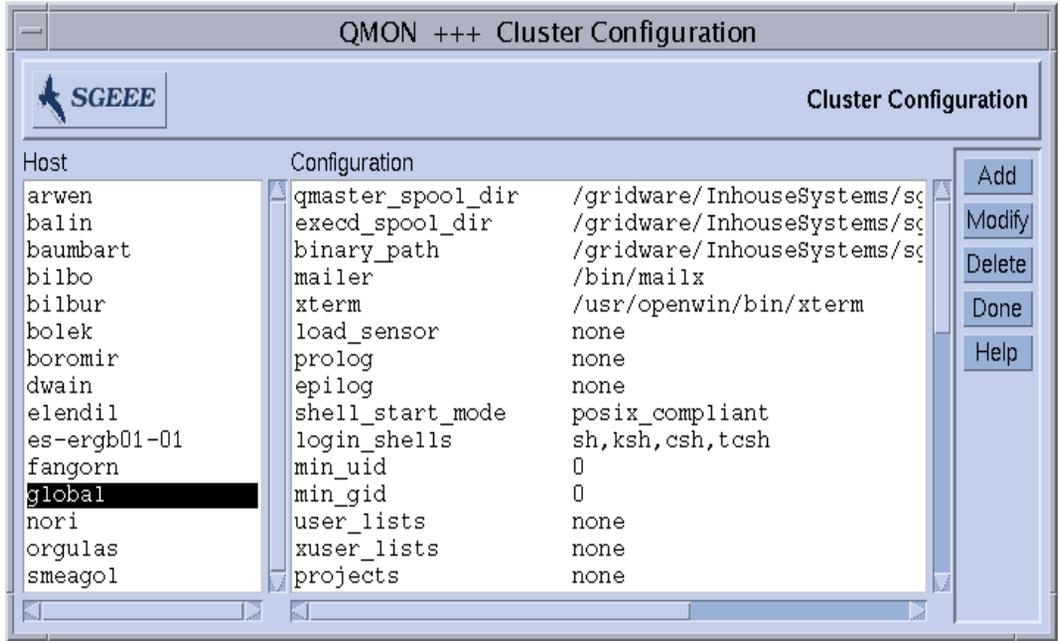


圖 6-9 「叢集配置」對話方塊

2. 在螢幕左側的「主機」選擇清單中，按一下主機名稱可顯示該主機目前的配置。

## ▼ 如何使用 QMON 刪除叢集配置

1. 在 QMON 主要功能表中按一下「叢集配置」鈕。
2. 在螢幕左側的「主機」選擇清單中，按一下您想刪除配置之主機的名稱。
3. 請按「刪除」鈕。

## ▼ 如何使用 QMON 顯示全域叢集配置

- 在「主機」選擇清單中，選擇名稱，global。

配置以 sge\_conf 線上說明手冊所述的格式顯示。使用「修改」鈕來修改選取的全域或主機本端配置。使用「新增」鈕為特定主機新增新的配置。

## ▼ 如何使用 QMON 修改全域和主機配置

1. 在「叢集配置」對話方塊中（如第 147 頁的「如何使用 QMON 顯示叢集配置」一節所述），按一下「新增」鈕或「修改」鈕。

類似圖 6-10 範例的「叢集設定」對話方塊就會開啓。

The screenshot shows the 'Cluster Settings' dialog box with the 'Configuration for Host' section. The 'General Settings' tab is selected, displaying various configuration fields. The 'Advanced Settings' tab is also visible, showing fields for Min UID, Min GID, Finished Jobs, Loadreport Time, Max Unheard, Stat Log Time, Reschedule Unknown, User Lists, Xuser Lists, Projects, XProjects, Ignore FQDN, Enforce Project, Enforce User, and GID Range for Job Monitoring.

| General Settings            |                                 | Advanced Settings            |                    |               |
|-----------------------------|---------------------------------|------------------------------|--------------------|---------------|
| Master Spool Dir            | /gridware/inhouseSystems/sgeee1 | Min UID                      | Min GID            | Finished Jobs |
| Execd Spool Dir             | /gridware/inhouseSystems/sgeee1 | 0                            | 0                  | 0             |
| Binary Path                 | /gridware/inhouseSystems/sgeee1 | Loadreport Time              | Max Unheard        |               |
| Mailer                      | /bin/mailx                      | 00:00:15                     | 00:05:00           |               |
| Xterm                       | /usr/openwin/bin/xterm          | Stat Log Time                | Reschedule Unknown |               |
| Load Sensor                 | none                            | 48:00:00                     | 00:00:00           |               |
| Admin User                  | ccadmin                         | User Lists                   |                    |               |
| Admin Mail                  | none                            |                              |                    |               |
| Prolog                      | none                            | Xuser Lists                  |                    |               |
| Epilog                      | none                            |                              |                    |               |
| Login Shells                | sh,ksh,csh,tcsh                 | Projects                     |                    |               |
| Maximal Array Job Instances | 2000                            |                              |                    |               |
| Maximal Array Job Tasks     | 75000                           | XProjects                    |                    |               |
| Maximal Jobs Per User       | 0                               |                              |                    |               |
| Shell Start Mode            | posix_compliant                 | Ignore FQDN True             |                    |               |
| Log Level                   | log_info                        | Enforce Project False        |                    |               |
|                             |                                 | Enforce User False           |                    |               |
|                             |                                 | GID Range for Job Monitoring |                    |               |
|                             |                                 | 33801-34000                  |                    |               |

圖 6-10 叢集設定對話方塊 — 一般設定

2. 請參照以下各節的詳細資訊進行變更。

「叢集設定」對話方塊提供變更全域或主機本端配置的所有參數的方法。所有的項目欄位只在變更全域配置時才能存取；即若您選取主機、全域且如果您按「修改」。如果修改一般主機，則其實際的配置會反映在對話方塊中，而只有適合主機本端變更的參數才能修改。如果增加新的主機本端配置，則對話方塊項目會是空的欄位。

「進階設定」標籤（圖 6-11）會根據變更的是全域、主機本端或新的配置來顯示對應的運作方式。它會提供較少用到的叢集配置參數的存取。

| Additional Parameters |  |
|-----------------------|--|
| Master Parameters     | none                                   |
| Schedd Parameters     | none                                   |
| Execd Parameters      | PTF_MIN_PRIORITY=20,PTF_MAX_PRIORITY=0 |
| Shepherd Command      | none                                   |
| Default Domain        | none                                   |

| Interactive Parameters |                      |
|------------------------|----------------------|
| Qlogin Daemon          | /usr/sbin/in.telnetd |
| Qlogin Command         | telnet               |
| rsh Daemon             |                      |
| rsh Command            |                      |
| rlogin Daemon          | /usr/sbin/in.rlogind |
| rlogin Command         |                      |

圖 6-11 叢集設定對話方塊 — 進階設定

在完成修改後，右上角的「確定」鈕會紀錄修改的配置。按「取消」可刪除任何變更。這兩個例子都會關閉對話方塊。

要取得所有叢集配置參數的完整說明，請參考 `sge_conf` 線上說明手冊。



# 配置佇列與佇列行事曆

---

本章提供配置 Sun Grid Engine，企業版 5.3 佇列與佇列行事曆的相關資訊與指示。

以下為本章涵蓋指示的特定作業清單。

- 第 152 頁的「如何以 QMON 配置佇列」
- 第 153 頁的「如何配置一般參數」
- 第 154 頁的「如何配置執行法參數」
- 第 155 頁的「如何配置檢查點參數」
- 第 156 頁的「如何配置載入 / 暫停臨界值」
- 第 157 頁的「如何配置限制」
- 第 159 頁的「如何配置使用者聚合體」
- 第 160 頁的「如何配置次級佇列」
- 第 161 頁的「如何配置使用者存取」
- 第 162 頁的「如何配置專案存取」
- 第 163 頁的「如何配置所有人參數」
- 第 164 頁的「如何從指令行配置佇列」
- 第 165 頁的「如何以 QMON 配置佇列行事曆」
- 第 167 頁的「如何從指令行配置行事曆」

---

## 關於佇列配置

Sun Grid Engine，企業版佇列為不同種類作業的容器，為屬於同一種類但同時執行之不同作業提供對應資源。作業不會停留在 Sun Grid Engine，企業版佇列中，反而會在分配至佇列當下立刻執行。Sun Grid Engine，企業版排程器的作業擱置清單是 Sun Grid Engine，企業版作業的唯一等候欄位。

配置 Sun Grid Engine，企業版佇列將會以 `sge_qmaster` 記錄佇列屬性。一旦佇列配置完畢，就會立即顯示在全部叢集上，且位於 Sun Grid Engine，企業版儲存欄位的所有主機上之全部 Sun Grid Engine，企業版使用者皆可看到。

## ▼ 如何以 QMON 配置佇列

1. 從 QMON 主要功能表上按下「佇列控制」鈕。
2. 在「佇列控制」對話方塊中按下「新增」或「修改」鈕。

就會出現「佇列配置」的對話方塊。第 120 頁的「如何用 QMON 控制佇列」一節會詳加說明「佇列控制」對話方塊，和其監視與操作佇列狀態的工具。如果是第一次開啓「佇列配置」對話方塊，就會出現「一般參數」表（請參閱第 153 頁的「如何配置一般參數」）。

3. 請參照下一節的詳細資訊決定如何配置。

受到想要之操作影響的佇列會顯示並定義於螢幕上方的「佇列」與「主機名稱」視窗。如果一個佇列被修改了，在「佇列配置」對話方塊開啓前，「佇列控制」對話方塊裡必須選出另一個現有佇列。如果要新增一個新的佇列，就必須定義佇列名稱與佇列位於的主機。

在「主機名稱」視窗下有三個按鈕可讓您輕鬆使用「佇列配置」對話方塊。「複製」鈕可透過佇列選取清單輸入一現有佇列的所有參數，「重新設定」鈕可以載入平台佇列的配置，而「重新顯示」鈕則可在對話方塊開啓時載入其餘修改過的物件配置（請參閱第 159 頁的「如何配置使用者聚合體」與第 161 頁的「如何配置使用者存取」以得知更多「重新顯示」鈕的詳細資訊）。

「佇列配置」對話方塊右上方的「確定」鈕會以 `sge_qmaster` 記錄這些變更，而右下方的「取消」按鈕則會捨棄任何變更。以上兩個按鈕都會關閉對話方塊。

有十組參數組可用來定義一佇列。

- 一般（請參閱第 153 頁的「如何配置一般參數」）
- 執行法（請參閱第 154 頁的「如何配置執行法參數」）
- 檢查點（請參閱第 155 頁的「如何配置檢查點參數」）
- 載入 / 暫停臨界值（請參閱第 156 頁的「如何配置載入 / 暫停臨界值」）
- 限制（請參閱第 157 頁的「如何配置限制」）
- 聚合體（請參閱第 159 頁的「如何配置使用者聚合體」）
- 次級（請參閱第 160 頁的「如何配置次級佇列」）
- 使用者存取（請參閱第 161 頁的「如何配置使用者存取」）
- 專案存取（請參閱第 162 頁的「如何配置專案存取」）
- 所有人（請參閱第 163 頁的「如何配置所有人參數」）

您可透過「佇列參數」標籤選擇理想的參數組。

## ▼ 如何配置一般參數

- 選擇「一般」參數組。

以下顯示的圖 7-1 為範例圖示。

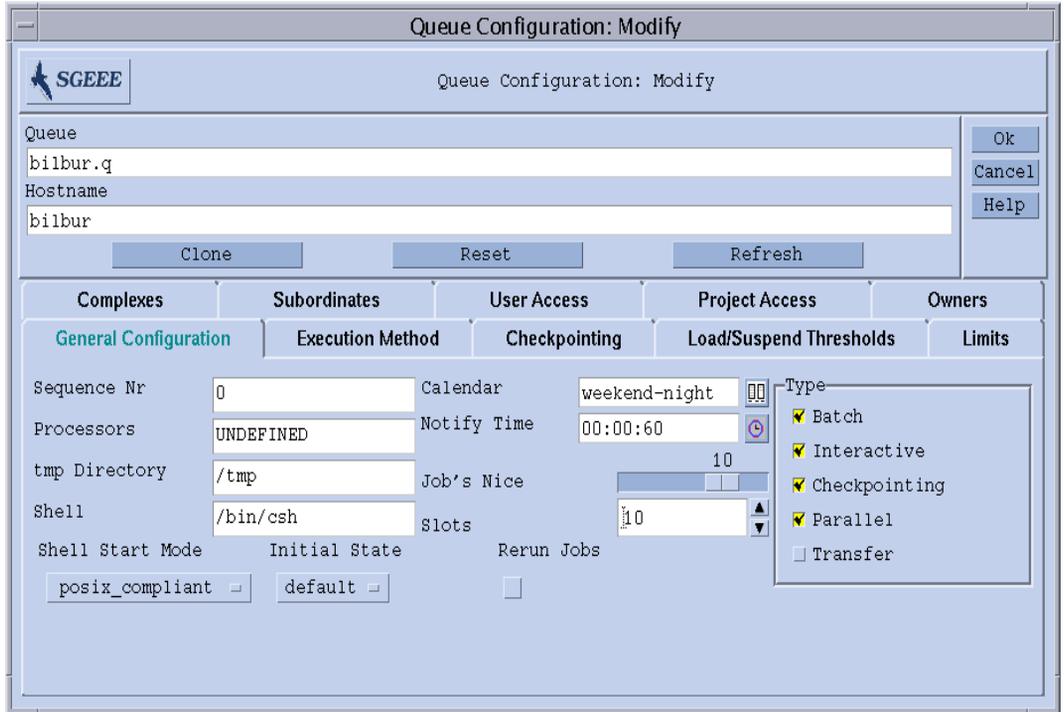


圖 7-1 佇列配置 — 一般參數

提供的輸入欄位可用來設定以下參數：

- 佇列的序列編號。
- 處理器—為一個可設定佇列作業使用處理器之指定器。對一些作業系統架構而言，這可以是一範圍（例如 1-4、8、10）或是處理器組的整數辨識器。請參閱您 Sun Grid Engine，企業版分配之 doc 目錄中的 arc\_depend\_\*.asc 檔以得知更多相關資訊。
- 暫存目錄路徑。
- 用以執行作業 script 的預設指令解譯器 (Shell)。
- 佇列隨附之行事曆：可為佇列定義*作業中*與*結束作業*的時間。
- 傳達 SIGUSR1/SIGUSR2 通知訊號與暫停 / 刪除訊號之間的等候時間 (Notify)。
- 用以啟動佇列作業的 nice 值（0 表示使用系統預設值）。
- 佇列中可同時執行的作業數量（作業插槽）。
- 佇列中可執行之佇列與作業種類。有多重選擇可供使用。

- Shell Start Mode，即啓動作業 script 的模式。
- Initial State：可新增佇列，或者在佇列主機上執行的 sge\_execd 啓動時可儲存佇列。
- 佇列預設之 rerun 策略：當捨棄作業時會執行之策略，例如當系統故障時。使用者可透過 qsub -r 選項或者「作業提交」對話方塊覆寫此策略（請參閱圖 4-9）。

請參閱 queue\_conf 線上說明手冊以得知更多參數相關資訊。

## ▼ 如何配置執行法參數

- 選擇「執行法」參數組。

以下顯示的圖 7-2 為範例圖示。

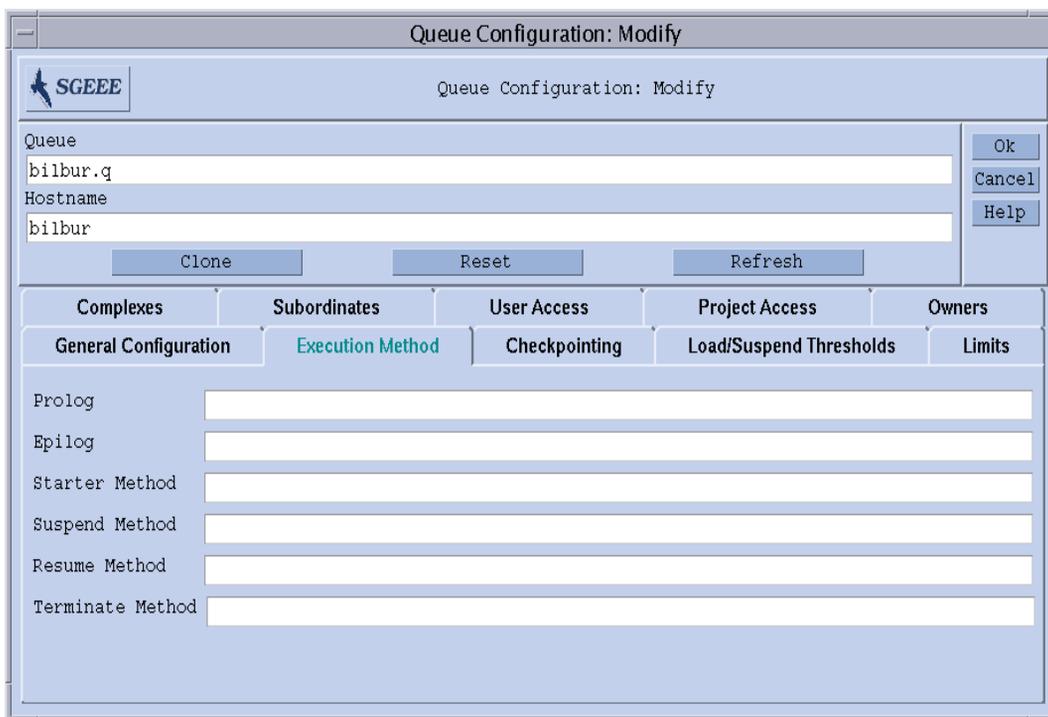


圖 7-2 佇列配置－執行法參數

提供的輸入欄位可用來設定以下參數：

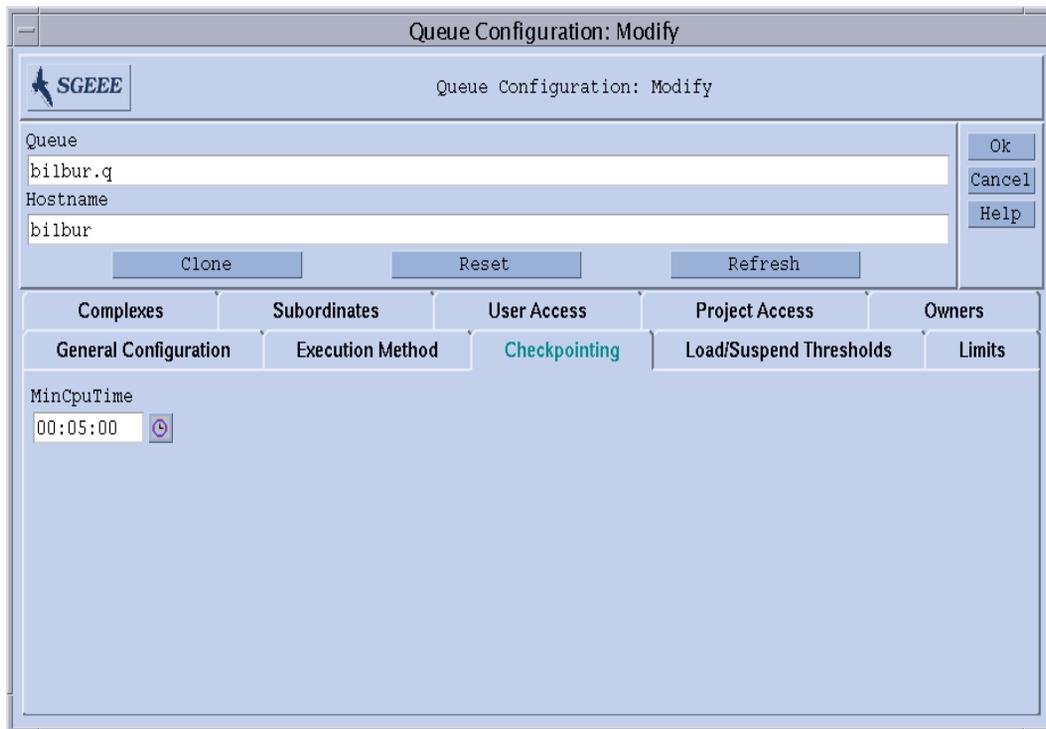
- 在作業 script 開始與結束時，與作業擁有相同執行環境之特定佇列 prologue 與 epilogue script。
- 可覆寫應用於作業中 Sun Grid Engine，企業版預設值的啓動 / 暫停 / 恢復 / 終止法。

請參閱 `queue_conf` 線上說明手冊以得知更多參數相關資訊。

## ▼ 如何配置檢查點參數

- 選擇「檢查點」參數組。

以下顯示的圖 7-3 為範例圖示。



The screenshot shows a window titled "Queue Configuration: Modify" with the SGE logo. It contains several input fields: "Queue" with the value "bilbur.q" and "Hostname" with the value "bilbur". Below these are buttons for "Clone", "Reset", and "Refresh". On the right side, there are buttons for "Ok", "Cancel", and "Help". The window has a tabbed interface with the following tabs: "Complexes", "Subordinates", "User Access", "Project Access", "Owners", "General Configuration", "Execution Method", "Checkpointing" (which is selected), "Load/Suspend Thresholds", and "Limits". Under the "Checkpointing" tab, the "MinCpuTime" parameter is visible, with a text input field containing "00:05:00" and a small circular icon to its right.

圖 7-3 佇列配置－檢查點參數

提供的輸入欄位可用來設定以下參數：

- 定期檢查點間隔時間 (MinCpuTime)

請參閱 `queue_conf` 線上說明手冊以得知更多參數相關資訊。

## ▼ 如何配置載入 / 暫停臨界值

- 選擇「載入 / 暫停臨界值」參數組。
- 以下顯示的圖 7-4 為範例圖示。

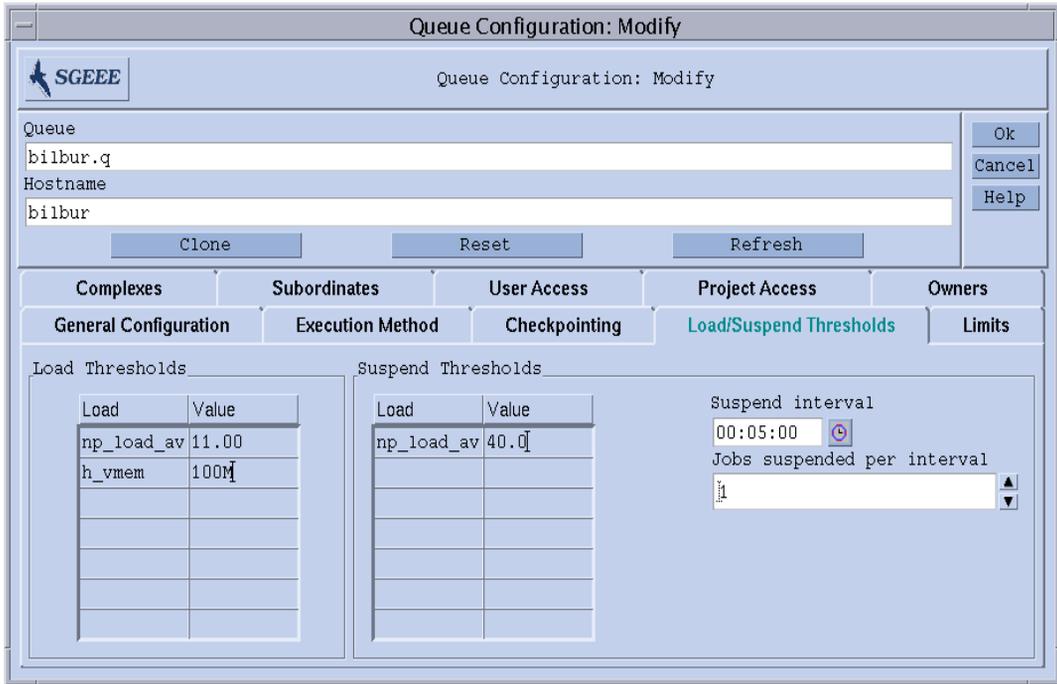


圖 7-4 佇列配置－載入 / 暫停臨界值

提供的輸入欄位可用來設定以下參數：

- 載入臨界值與暫停臨界值表可用來定義載入參數的超載臨界值與可耗損聚合體屬性。（請參見第 225 頁的「關於聚合體」）。

載入臨界值的超載值可避免佇列透過 Sun Grid Engine，企業版繼續接受作業。當超過一或二個以上的暫時臨界值時就會暫停佇列作業以減低載入。最新配置的臨界值顯示於圖表中。一個現有的臨界值可經由連按滑鼠左鍵兩下，即可選擇與變更對應「數值」欄位的設定。要增加新的臨界值，請按上方的「名稱」鈕或「數值」鈕。顯示佇列所有有效屬性的選擇清單將會開啓。「屬性選擇」對話方塊圖顯示於圖 6-6。選擇其中一個屬性，並按「確定」鈕確認選項，就會新增此屬性至相對臨界值圖表的「名稱」欄，並將游標放入數值欄位中。輸入 CTRL-D 或按滑鼠右鍵開啓刪除方塊並確認刪除動作，就可刪除選取的清單記錄。

- 每次間隔時間暫時的作業數量：以減低配置佇列系統的負載。
- 更多作業的暫停間距時間：以避免暫停的臨界值仍然過多。

請參閱 queue\_conf 線上說明手冊以得知更多參數相關資訊。

## ▼ 如何配置限制

- 選擇「限制」參數組。

以下顯示的圖 7-5 為範例圖示。

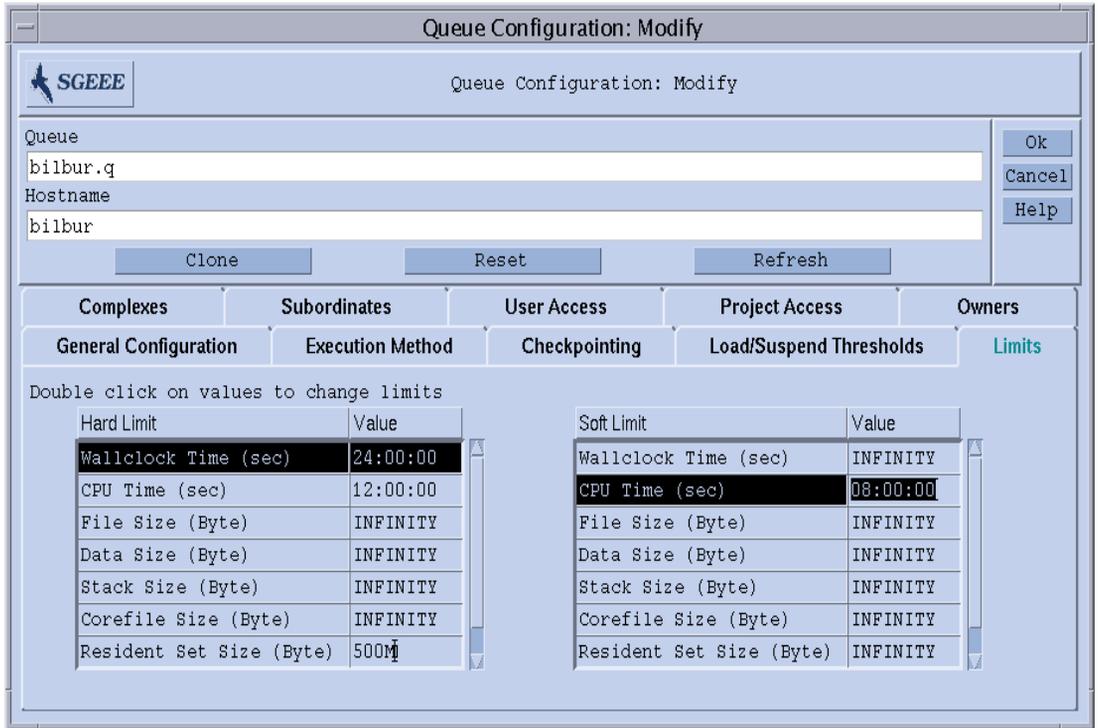


圖 7-5 佇列配置－限制

提供的輸入欄位可用來設定以下參數：

- 於佇列中執行作業的硬與軟限制。

要變更限制數值就連接兩下「限制」記錄的數值欄位。連接數值欄位四下開啓「記憶體」或「時間」限制數值的輸入對話方塊（請參見圖 7-6 與圖 7-7）。

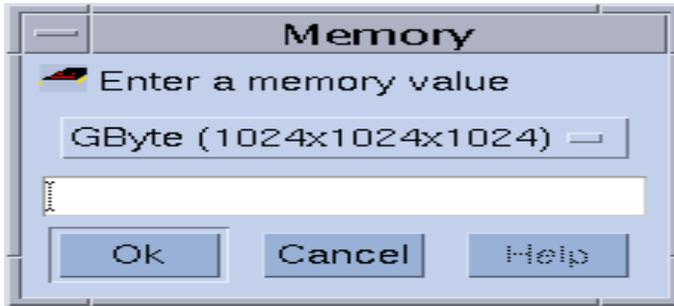


圖 7-6 記憶體輸入對話方塊

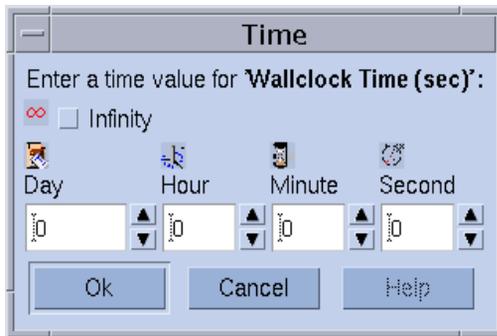


圖 7-7 時間輸入對話方塊

請參閱 `queue_conf` 與 `setrlimit` 線上說明手冊以得知更多個別限制參數，與其在不同作業系統架構中的解讀方式之詳細資訊。

## ▼ 如何配置使用者聚合體

- 選擇「使用者聚合體」參數組。  
以下顯示的圖 7-8 為範例圖示。

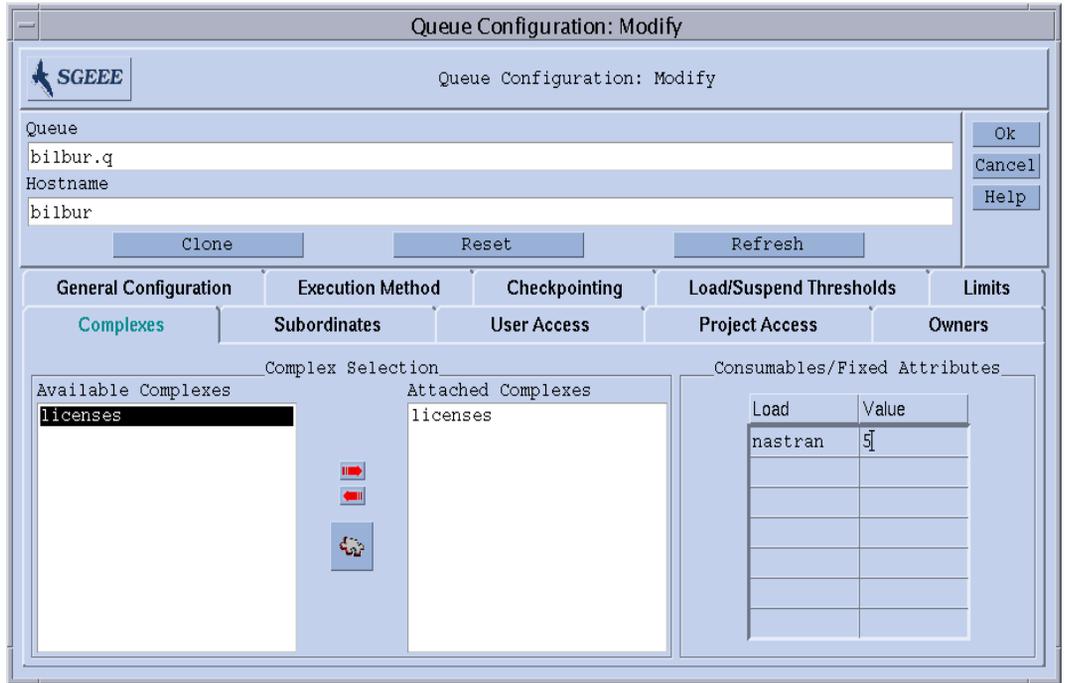


圖 7-8 佇列配置 — 使用者聚合體

提供的輸入欄位可用來設定以下參數：

- 使用者設定：定義隨附於佇列之聚合體（請參見第 231 頁的「使用者定義聚合體」）。  
「聚合體選擇」方塊中的紅色箭號可從 / 到佇列中隨附或分離使用者定義之聚合體。
- 從佇列可用之聚合體參數組為選取的屬性定義數值。

可用之聚合體參數會從全域聚合體、主機聚合體、與隨附使用者定義之聚合體中集合每個預設值。屬性為可耗損或固定參數。就可耗損屬性、或固定屬性的固定佇列特定值而言，定義佇列值同時也定義了由佇列管理的容量（請見第 225 頁的「關於聚合體」得知詳細資訊）。有明確定義數值的屬性會顯示於「可耗損 / 固定屬性」表中。一現有屬性可藉由連按兩下對應數值欄位做選擇與變更。要新增屬性定義，請按上方的「名稱」鈕或「數值」鈕。顯示佇列所有有效屬性的選擇清單將會開啓。「屬性選擇」對話方塊圖顯示於圖 6-6。選擇其中一個屬性，並按「確定」鈕確認選項，就會新增此屬性至屬性表之「名稱」欄，並將游標放入數值欄位中。輸入 CTRL-D 或按滑鼠右鍵開啓刪除方塊並確認刪除動作，就可刪除選取的清單記錄。

請參閱 `queue_conf` 線上說明手冊以得知更多參數相關資訊。

「聚合體配置」對話方塊（請參見第八章的圖 9-5，第 225 頁的「聚合體概念」）在按下「聚合體配置」圖示按鈕時就會開啓。您可以在隨附或分離使用者定義之聚合體至佇列以前，檢查或修改目前聚合體配置。

## ▼ 如何配置次級佇列

- 選擇「次級」參數組。

以下顯示的圖 7-9 為範例圖示。

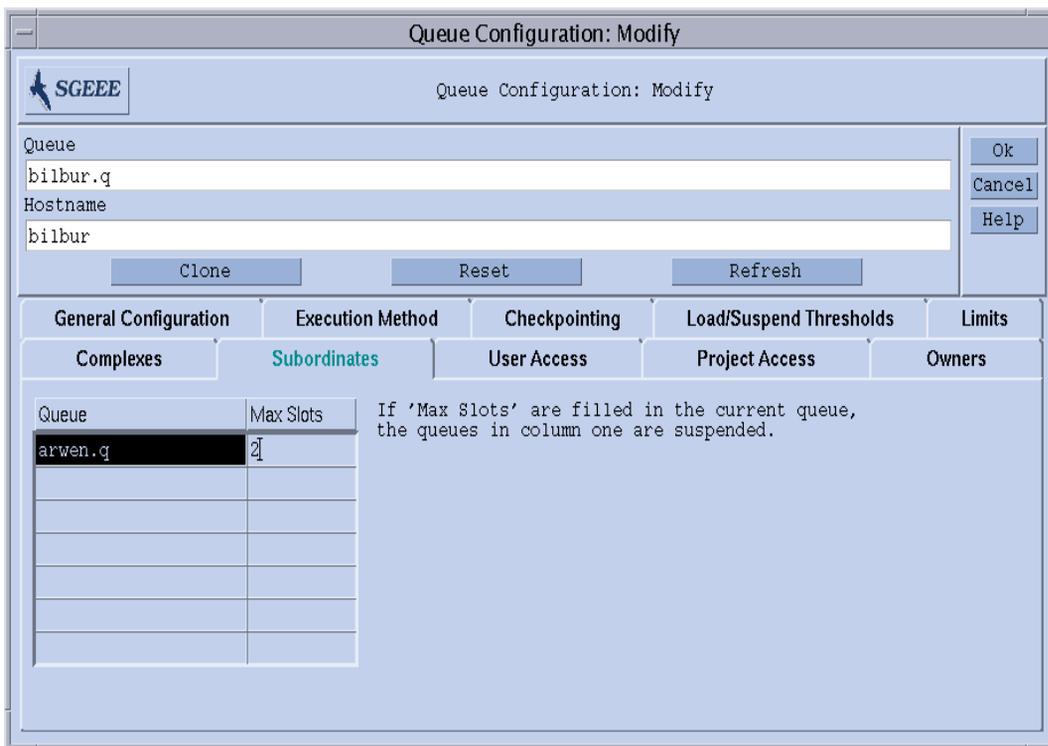


圖 7-9 佇列配置 – 次級參數

提供的輸入欄位可用來設定以下參數：

- 次於配置佇列之佇列

若配置佇列忙碌時，就會暫停次級佇列；若配置佇列不忙碌，就會解除暫停次級佇列。對次級佇列而言，每一配置佇列至少需配置作業插槽的數量以便啓動暫停功能。若無指定作業插槽數值，就必須填入所有插槽數值以啓動對應佇列之暫停功能。

請參閱 `queue_conf` 線上說明手冊以得知更多參數相關資訊。

使用次級佇列設備來執行高優先與低優先佇列，和獨立佇列。

## ▼ 如何配置使用者存取

- 選擇「使用者存取」參數組。

以下顯示的圖 7-10 為範例圖示。

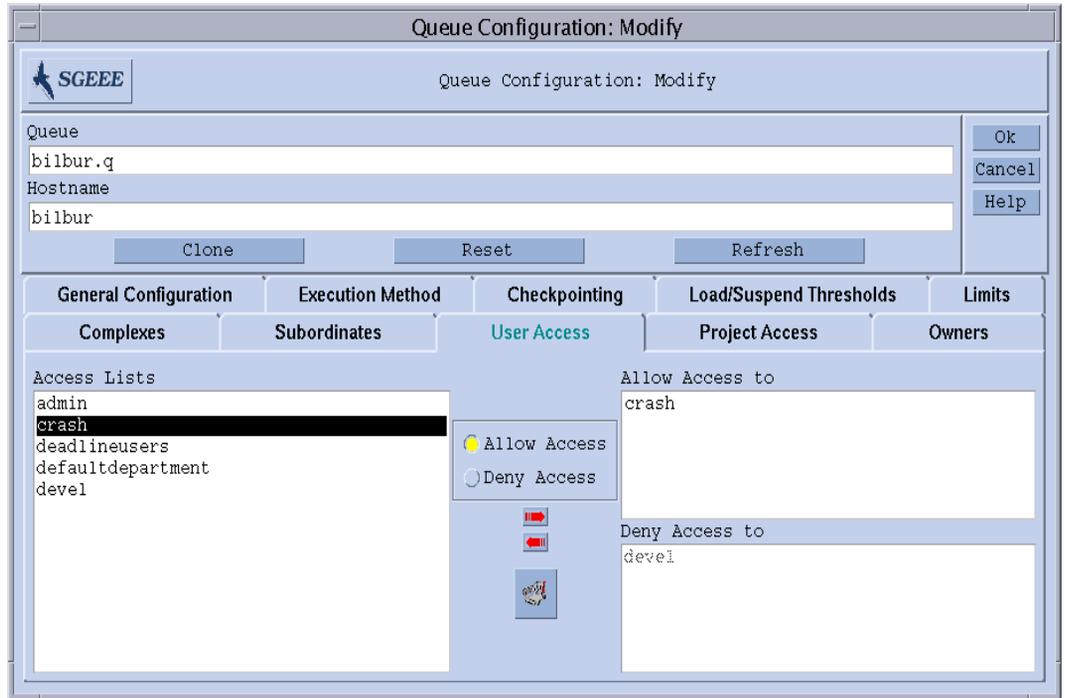


圖 7-10 佇列配置 — 使用者存取

提供的輸入欄位可用來設定以下參數：

- 隨附於佇列的允許或拒絕清單之使用者存取清單。

存取清單內允許清單之使用者或使用者群組可存取至佇列中，而位於拒絕清單中就無法存取至佇列。如果允許清單是空的，就不會限制存取，除非於拒絕清單中有明確指定。

請參閱 `queue_conf` 線上說明手冊以得知更多參數相關資訊。

按螢幕下方中間的按鈕開啓「存取清單配置」對話方塊（請參見第 59 頁的「使用者存取權限」）。

## ▼ 如何配置專案存取

- 選擇「專案存取」參數組。

以下顯示的圖 7-11 為範例圖示。

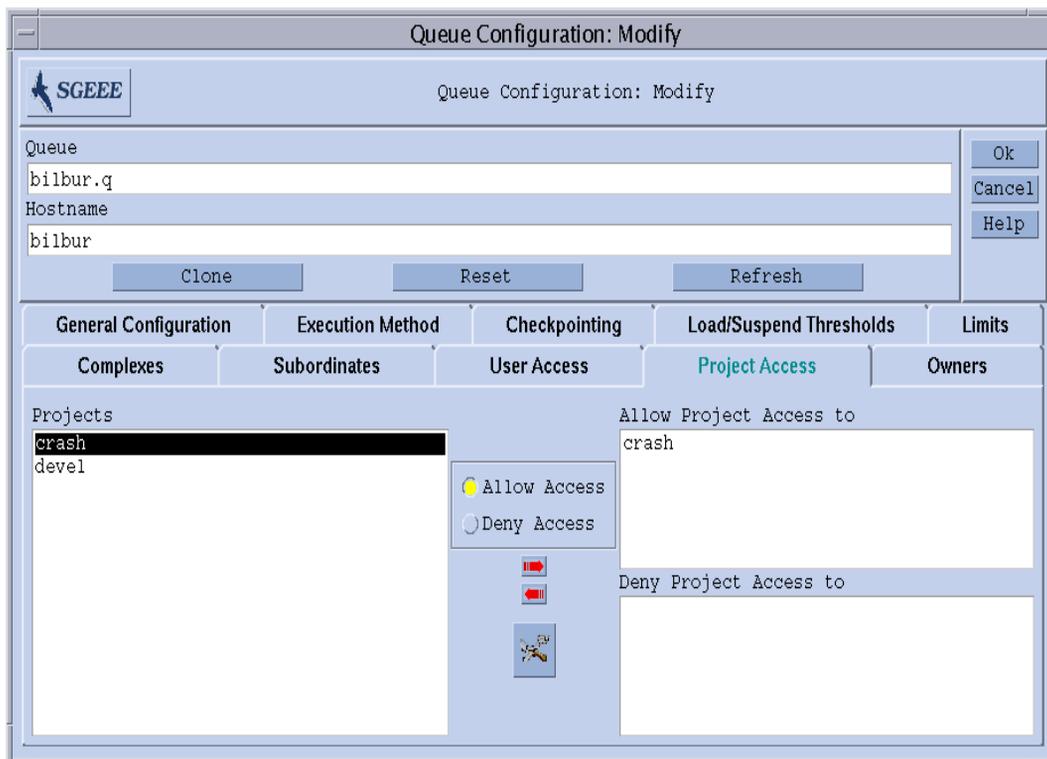


圖 7-11 佇列配置 — 專案存取

提供的畫面可用來設定以下參數：

- 允許或封鎖存取至佇列之專案

提交至允許的專案清單之作業可存取於佇列中。提交至拒絕專案之作業將無法發送至佇列中。

請參閱 `queue_conf` 線上說明手冊以得知更多參數相關資訊。

按螢幕下方中間的按鈕開啓「專案配置」對話方塊（請參見第 182 頁的「關於專案」）。

## ▼ 如何配置所有人參數

- 選擇「所有人」參數組。  
以下顯示的圖 7-12 為範例圖示。

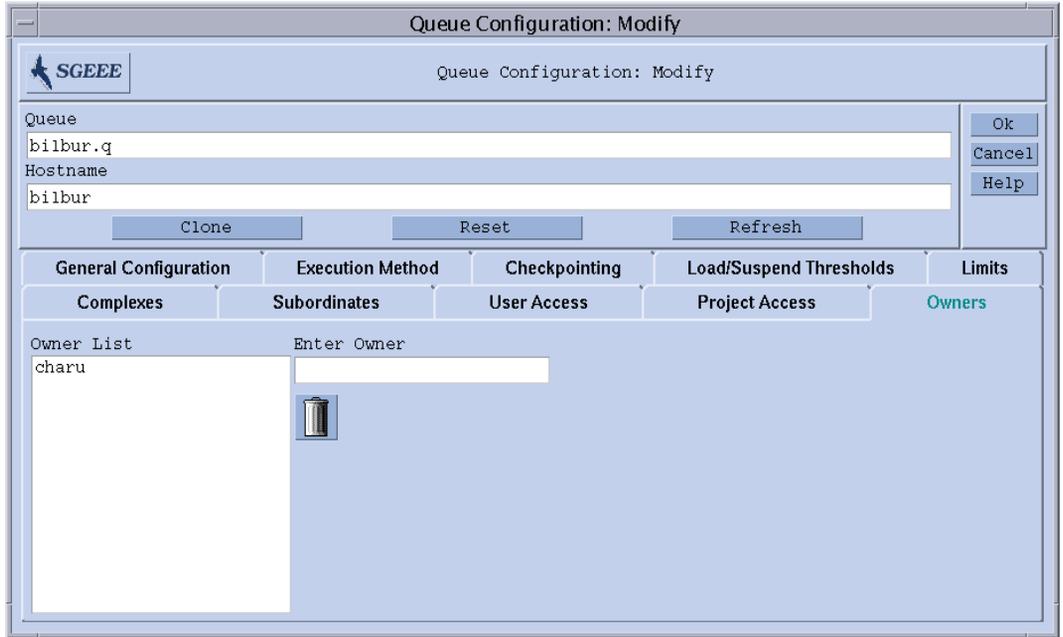


圖 7-12 佇列配置 – 所有人

提供的輸入欄位可用來設定以下參數：

### ■ 佇列所有人清單

佇列之所有人擁有暫停 / 解除暫停，或關閉 / 啟動佇列之權力。所有可用之使用者帳號皆為有效值，可新增至佇列所有人清單中。要從佇列所有人清單中刪除一使用者帳號，請於「所有人清單」視窗選取此帳號，然後按一下對話方塊右下方的垃圾筒圖示。

請參閱 `queue_conf` 線上說明手冊以得知更多參數相關資訊。

## ▼ 如何從指令行配置佇列

- 輸入以下指令與適合的選項（依您想如何配置佇列而定）。

```
# qconf options
```

qconf 指令有以下幾個選項。

- qconf -aq [佇列名稱]

新增佇列 — 在佇列配置平台叫出編輯器（預設 vi 或對應至 \$EDITOR 環境變數）。如果選項參數 *佇列名稱* 出現，就使用佇列配置做為平台。變更平台並儲存至磁碟片中就完成配置佇列的工作。請參見《Sun Grid Engine 5.3 和 Sun Grid Engine，企業版 5.3 參考手冊》內的 *queue\_conf* 記錄以得知改變平台記錄的詳細資訊。

- qconf -Aq 檔案名稱

新增佇列 — 使用這個檔案 *檔案名稱* 來定義佇列。定義檔也許是由 qconf -sq *佇列名稱* 建立的（見下方）。

- qconf -cq *佇列名稱* [...]

清除佇列 — 清除特定佇列的狀態使其閒置以執行作業。重設狀態時並不會顧及到目前狀態。此選項有助於消除錯誤狀態，但不可用於一般操作模式。

- qconf -dq *佇列名稱* [...]

刪除佇列 — 從可用佇列清單中刪除引數中之特定佇列。

- qconf -mq *佇列名稱*

修改佇列 — 修改指定佇列。在欲變更之佇列配置中叫出編輯器（預設 vi 或對應至 \$EDITOR 環境變數）。變更配置並儲存至磁碟片中就完成修改佇列的工作。

- qconf -Mq 檔案名稱

修改佇列 — 使用這個檔案 *檔案名稱* 定義修改過的佇列配置。定義檔也許是由 qconf -sq *佇列名稱*（見下方）與之後的修改所建立的。

- qconf -sq [佇列名稱 [...]]

顯示佇列 — 顯示預設平台佇列配置（如果沒有出現引數的話），或者顯示目前由逗點分開的引數清單中之佇列配置。

- qconf -sql

顯示佇列清單 — 顯示所有目前配置的佇列清單。

## 關於佇列行事曆

佇列行事曆會依據一年的某天、一星期的某天、或某天何時來定義 Sun Grid Engine，企業版佇列的可用性。佇列可於時間中的仲裁點做配置以變更狀態。佇列狀態可以變更至關閉、啟用、暫停、或恢復（解除暫停）。

Sun Grid Engine，企業版提供定義行事曆特定站點組的功能，每個站點組皆包含仲裁狀態變更與事件發生的時間。這些行事曆與佇列是有關連的，即每個佇列須（或無須）隨附單一行事曆，才可接受隨附行事曆裡定義之檔案可用性。

行事曆格式的語法會於線上說明手冊中的 `calendar_conf` 詳加說明。以下幾個範例會說明對應的管理設備。

### ▼ 如何以 QMON 配置佇列行事曆

1. 在 QMON 主要功能表上按一下「行事曆配置」。  
「佇列行事曆配置」對話方塊與圖 7-13 很類似。

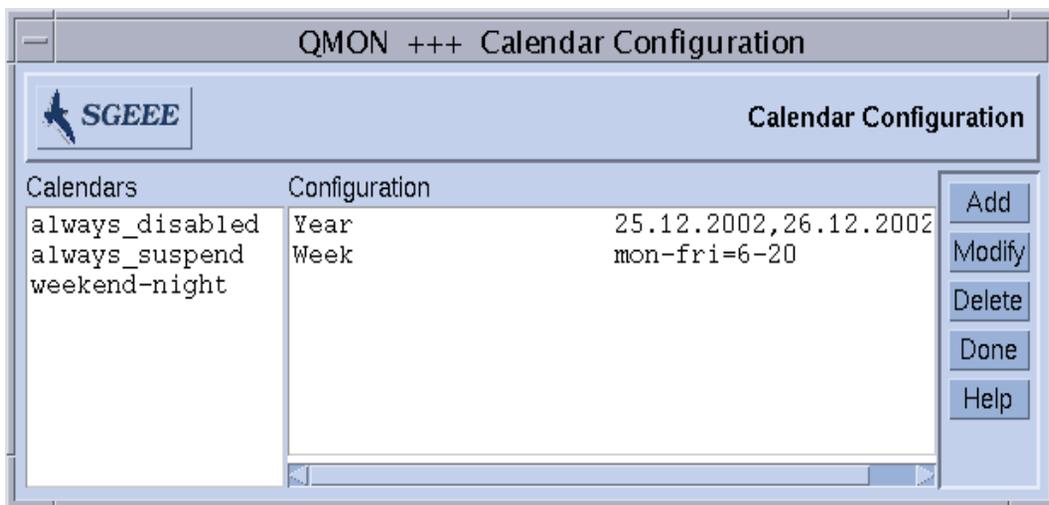


圖 7-13 行事曆配置

可用存取清單顯示於螢幕左方的「行事曆」選擇清單中。

2. 在「行事曆」選擇清單中，按一下您想修改或刪除的行事曆配置。

3. 依您想如何變更配置，執行以下其中一種步驟。
  - a. 按下螢幕右方的「刪除」按鈕來刪除選取的行事曆。
  - b. 按下「修改」按鈕修改選取的行事曆。
  - c. 按下「新增」按鈕新增存取清單。

在所有情況下皆會開啓「行事曆定義」對話方塊（如圖 7-14 所示），並提供刪除、修改、或新增的方法。

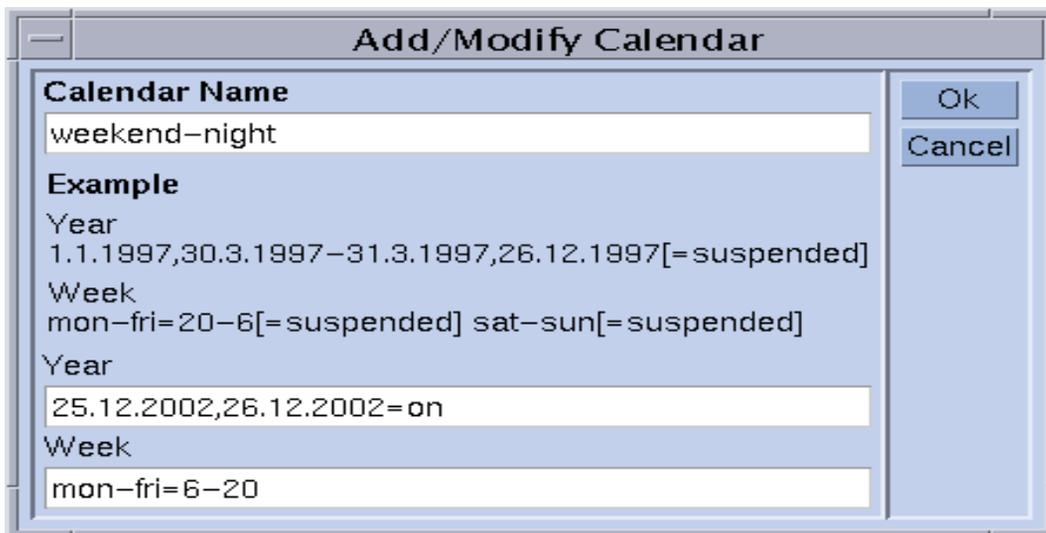


圖 7-14 新增、刪除、或修改行事曆

4. 依下節指導進行。

「行事曆名稱」輸入視窗會顯示修改作業中選取的行事曆名稱，或者您也可以自行輸入欲作業之行事曆名稱。「年」與「星期」輸入欄位讓您可利用 `calendar_conf` 線上說明手冊中提及的語法來定義行事曆事件。

對佇列而言，以上行事曆的配置範例也可適用於下班時間與週末，而且耶誕佳節被定義為比照週末辦理。

請參閱《*Sun Grid Engine 5.3 和 Sun Grid Engine，企業版 5.3 參考手冊*》裡的 `calendar_conf` 項目以得知更多範例與語法的詳細說明。

要於佇列上隨附行事曆配置，就得於佇列上建立行事曆定義之可用性檔案。隨附於一般參數佇列配置功能表的行事曆顯示於圖 7-15。「行事曆」輸入欄位包含隨附之行事曆名稱，而輸入欄位旁的圖示鈕則會開啓選擇對話方塊與最近配置的行事曆清單。請參閱第 151 頁的「關於佇列配置」一節以得知更多配置佇列的詳細說明。

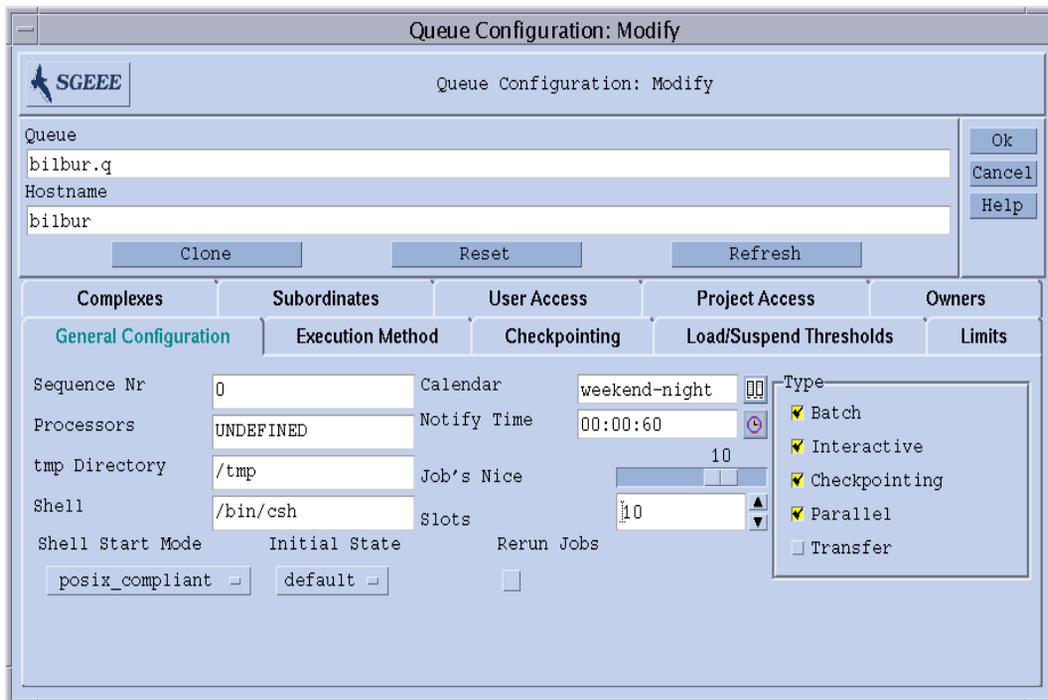


圖 7-15 於一般參數佇列配置功能表的行事曆配置

## ▼ 如何從指令行配置行事曆

- 輸入以下指令與合適的開關。

```
# qconf 開關
```

以下為四種可用之開關：

- qconf -Acal, -acal

新增行事曆 — 本指令會新增行事曆配置至 Sun Grid Engine，企業版叢集。新增之行事曆會由檔案 (-Acal)，或者一開啓平台配置之編輯器讀出，讓您可輸入行事曆。

- qconf -dcal

刪除行事曆。

- qconf -Mcal, -mcal

**修改行事曆** — 本指令可修改現有行事曆配置。欲修改之行事曆會由檔案 (-Mcal)，或者開啓之前配置的編輯器讀出，讓您可輸入新的定義 (-mcal)。

■ `qconf -scal, -scall`

**顯示行事曆** — 本指令可顯示現有行事曆配置。(-scal) 或者列出所有配置的行事曆清單 (-scall)。

# 管理使用者存取與策略

---

本章包含 Sun Grid Engine，企業版系統中與使用者管理、帳號和策略相關之重要資訊。本章主題包含使用者存取、專案、排程、更改路徑名稱、預設要求、計算與利用數據、及檢查點支援。

除了背景資訊外，本章還包涵如何完成下述作業之詳細指示。

- 第 171 頁的「如何以 QMON 配置帳號」
- 第 172 頁的「如何以 QMON 配置管理員 (Manager) 帳號」
- 第 173 頁的「如何從指令行配置管理員帳號」
- 第 173 頁的「如何以 QMON 配置操作員帳號」
- 第 174 頁的「如何從指令行配置操作員帳號」
- 第 176 頁的「如何以 QMON 配置使用者存取清單」
- 第 178 頁的「如何從指令行配置使用者存取清單」
- 第 179 頁的「如何以 QMON 配置使用者物件」
- 第 180 頁的「如何指定預設專案」
- 第 181 頁的「如何從指令行配置使用者物件」
- 第 182 頁的「如何以 QMON 定義專案」
- 第 185 頁的「如何從指令行定義專案」
- 第 191 頁的「如何以 QMON 配置變更排程器配置」
- 第 194 頁的「如何以 QMON 管理策略 / 票卷進階資源管理」
- 第 198 頁的「如何從 QMON 編輯共享樹策略」
- 第 203 頁的「如何從指令行配置共享策略」
- 第 205 頁的「如何從 QMON 配置功能共享策略」
- 第 208 頁的「如何從指令行配置功能共享策略」
- 第 212 頁的「如何配置置換策略」
- 第 213 頁的「如何從指令行配置置換策略」
- 第 220 頁的「如何以 QMON 配置檢查點環境」
- 第 224 頁的「如何從指令行配置檢查點環境」

# 關於設定使用者

以下清單說明於 Sun Grid Engine，企業版中設定使用者之必要 / 可用作業：

- 必要登入

爲了從主機 *A* 提交作業，並於主機 *B* 執行，使用者必須於主機 *A* 與 *B* 使用相同帳號（即相同之使用者名稱）。執行 `sge_qmaster` 之機器無需做任何登入。

- 設定 Sun Grid Engine，企業版存取權限

Sun Grid Engine，企業版軟體有提供限制使用者存取至全部叢集、佇列、與平行環境的能力。請見第 175 頁的「關於使用者存取權限」一節得知更多詳細說明。

除此之外，Sun Grid Engine，企業版系統使用者也可得到暫停或開啓特定佇列之權限（請見第 163 頁的「如何配置所有人參數」得知更多資訊）。

- Sun Grid Engine，企業版使用者聲明

如果您試圖爲使用者將節點包含至分享樹中，或者定義功能或手動裝置策略（請見第 194 頁的「如何以 QMON 管理策略 / 票卷進階資源管理」一節），就必須於 Sun Grid Engine，企業版系統中聲明使用者。請見第 179 頁的「如何以 QMON 配置使用者物件」得知詳細資訊。

- Sun Grid Engine，企業版專案存取

如果 Sun Grid Engine，企業版專案是用於定義共享、功能、或手動裝置策略（請見第 194 頁的「如何以 QMON 管理策略 / 票卷進階資源管理」一節），使用者就應可存取至一個或多個專案中。否則，使用者的作業可能會變成最低的優先順序等級，並難以接收資源存取。

- 檔案存取限制

Sun Grid Engine，企業版使用者必須能讀取存取目錄 `<sge_root>/cell/common`。

在 Sun Grid Engine，企業版作業開始前，Sun Grid Engine，企業版執行常駐程式（以 `root` 執行）會替作業建立一暫時工作目錄，並且改變目錄所有人至作業所有人（當作業結束時就會移除此暫時目錄）。此暫時工作目錄會建立於由佇列配置參數 `tmpdir` 定義之路徑中（請參見《Sun Grid Engine 5.3 和 Sun Grid Engine，企業版 5.3 參考手冊》裡的 `tmpdir` 記錄以得知更多資訊）。

確認暫時目錄會建立於 `tmpdir` 底下，先設定 Sun Grid Engine，企業版使用者所有權，如此使用者才可寫入暫時目錄中。

- 站點相依

依定義，批次作業沒有終端機連結。所以，指令解譯器啓動資源檔（例如：`csh` 的 `cshrc`）中的 UNIX 指令，如 `stty`，也許會導致錯誤。檢查事件，以預防內含如第 42 頁的「驗證安裝」裡說明的指令出現。

因 Sun Grid Engine，企業版批次作業通常會離線執行，僅有兩種方法可通知作業所有人錯誤事件與類似事件的發生。第一個方法是將錯誤訊息記錄至檔案中，而另一個則是寄電子郵件 (e-mail)。在某些罕見的情況下（例：無法開啓錯誤記錄檔案），

就僅能發送電子郵件直接通知使用者（諸如此類的錯誤訊息會記錄至 Sun Grid Engine，企業版系統日誌檔案，但通常使用者並不會去看這個系統日誌檔案）。因此，如果 Sun Grid Engine，企業版使用者有正確安裝電子郵件系統就會很幫助。

- Sun Grid Engine，企業版定義檔案

您可以為 Sun Grid Engine，企業版使用者設定下述定義檔案：

- `qmon`（Sun Grid Engine，企業版 GUI 的資源檔案；請見第 11 頁的「自訂 QMON」一節）
- `sge_aliases`（最新作業目錄路徑別名；請參見第 215 頁的「關於路徑別名」一節）
- `sge_request`（預設要求定義檔案；請見第 217 頁的「關於配置預設要求」一節）。

---

## 關於使用者存取

Sun Grid Engine，企業版系統中存有四種使用者目錄。

- **管理員 (Manager)** — 管理員可全權操作 Sun Grid Engine，企業版。根據預設值，主要主機與其他管理佇列機器的超級使用者擁有管理員的特權。
- **操作員 (Operators)** — 操作員可執行許多與管理員相同之指令，除了新增、刪除、或修改佇列。
- **所有人 (Owners)** — 佇列所有人受限不可暫停 / 繼續，或關閉 / 啟動擁有之佇列。這些特權對 `qidle` 使用成功與否是必要的。使用者通常被宣告為位於其桌面工作站佇列的所有人。
- **使用者 (Users)** — 使用者有其特定存取權限，請參見第 175 頁的「關於使用者存取權限」，但沒有叢集或佇列管理的能力。

以下各節會詳細說明各目錄。

### ▼ 如何以 QMON 配置帳號

1. 在 QMON 主要功能表上按下「使用者配置 (User Configuration)」鈕。
2. 根據您要如何配置按以下選擇標籤其中之一。
  - 管理員帳號配置（請見圖 8-1）
  - 操作員帳號配置（請見圖 8-2）
  - 使用者集存取 (User set access)/ 部門清單 (department list) 配置（請見圖 8-3）
  - 使用者配置（請參見圖 8-5）

3. 依下節指導進行。

**注意** – 當第一次按下「使用者配置」鈕時，預設就會開啓「管理員」帳號配置對話方塊。

## ▼ 如何以 QMON 配置管理員 (Manager) 帳號

當您選擇「管理員」標籤，就會出現「管理員配置」對話方塊（請見圖 8-1）。您可以在此對話方塊中宣告哪些帳號可以執行任何管理 Sun Grid Engine，企業版指令。螢幕下半的選擇清單會顯示已宣告擁有管理權限之帳號。

- 刪除 – 只要按一下名稱，再按下對話方塊右方的「刪除」鈕，即可從清單上刪除一現有管理員帳號。
- 新增 – 只要於選擇清單上方的輸入視窗輸入名稱，按下「新增」鈕，或按鍵盤上的「Return」鈕，即可新增一管理員帳號。

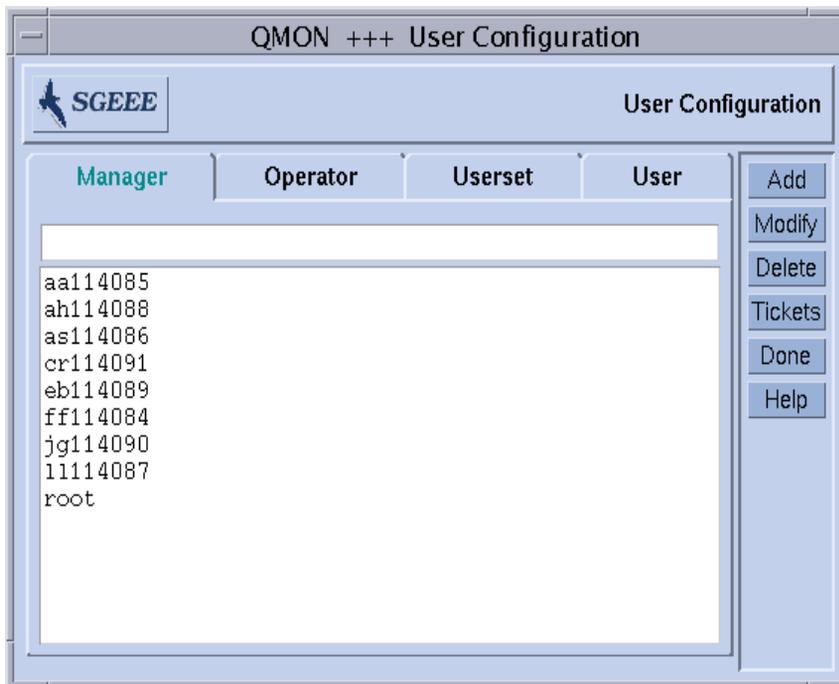


圖 8-1 管理員配置對話方塊

## ▼ 如何從指令行配置管理員帳號

- 使用合適的開關輸入以下指令。

```
# qconf 開關
```

### 可用的開關

- `qconf -am 使用者名稱 [...]`

新增管理員 — 此指令會新增一位或多位使用者至 Sun Grid Engine，企業版管理員清單中。根據預設，所有 Sun Grid Engine，企業版信賴的主機的 root 帳號（請見第 131 頁的「關於常駐程式和主機」一節）皆為 Sun Grid Engine，企業版管理員。

- `qconf -dm 使用者名稱 [...]`

刪除管理員 — 此指令可自 Sun Grid Engine，企業版管理員清單上刪除特定使用者。

- `qconf -sm`

顯示管理員 — 此指令會顯示所有 Sun Grid Engine，企業版管理員的清單。

## ▼ 如何以 QMON 配置操作員帳號

當您選擇「操作員」標籤，就會開啓「操作員」配置對話方塊（請見圖 8-2）。您可以在此對話方塊中宣告哪些帳號允許擁有受限制之管理 Sun Grid Engine，企業版指令權限（除非那些帳號也被宣告為管理員帳號 — 請參見第 172 頁的「如何以 QMON 配置管理員 (Manager) 帳號」）。螢幕下半的選擇清單會顯示已宣告可提供操作員權限之帳號。

- 刪除 — 只要按一下名稱，再按下對話方塊右方的「刪除」鈕，即可從清單上刪除一現有操作員帳號。
- 新增 — 只要於選擇清單上方的輸入視窗輸入名稱，按下「新增」鈕，或按鍵盤上的「Return」鈕，即可新增一操作員帳號。

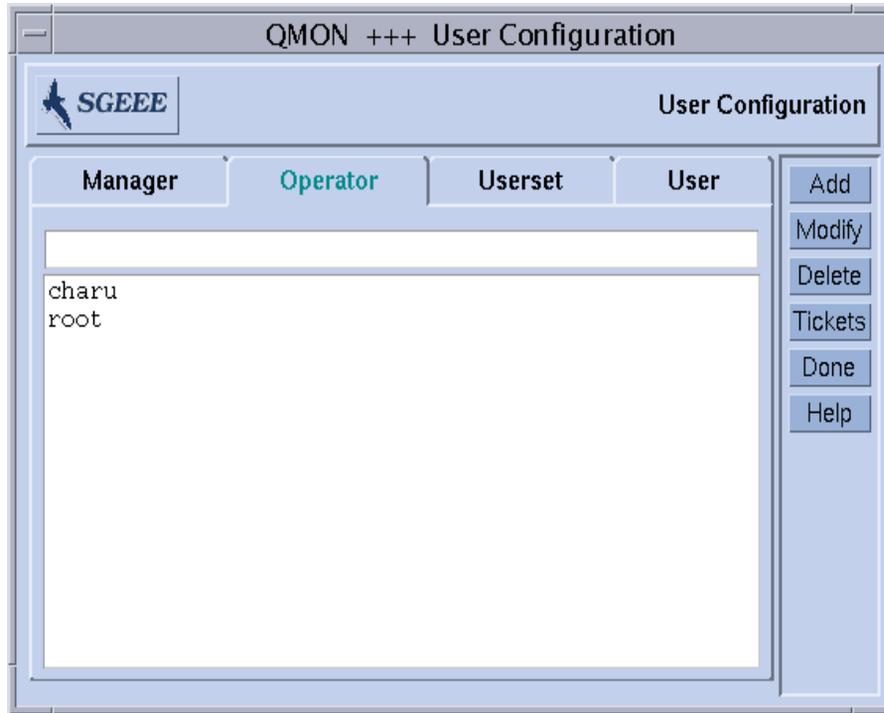


圖 8-2 操作員配置對話方塊

## ▼ 如何從指令行配置操作員帳號

- 使用合適的開關輸入以下指令。

```
# qconf 開關
```

### 可用的開關

- `qconf -ao 使用者名稱 [...]`

新增操作員 — 此指令會新增一位或多位使用者至 Sun Grid Engine，企業版操作員清單中。

- `qconf -do 使用者名稱 [...]`

刪除操作員 — 此指令可自 Sun Grid Engine，企業版操作員清單上刪除特定使用者。

- `qconf -so`

顯示操作員 — 此指令會顯示所有 Sun Grid Engine，企業版操作員的清單。

## 關於佇列所有人帳號

佇列所有人會於配置或修改 Sun Grid Engine，企業版佇列時完成定義。請見第 152 頁的「如何以 QMON 配置佇列」與第 164 頁的「如何從指令行配置佇列」二節。佇列所有人可做到下述事項：

- 暫停 — 停止所有執行於佇列或靠近佇列之作業。
- 繼續 — 恢復佇列中的執行作業與開啓佇列。
- 關閉 — 關閉佇列，但不會影響執行中的作業
- 啓動 — 開啓佇列

---

**注意** — 當一佇列已暫停時，已明確暫停之作業將不會在繼續佇列時重新恢復執行。這些作業需要明確恢復執行。

---

一般而言，如果使用者有時需要特定機器來執行重要作業，或者他們被執行於背景之 Sun Grid Engine，企業版作業嚴重影響，這些使用者就會被設定為特定佇列所有人。

## 關於使用者存取權限

任何至少於一個提交主機與執行主機上擁有有效記錄之使用者皆有使用 Sun Grid Engine，企業版系統的能力。但是，Sun Grid Engine，企業版管理員仍可限制特定使用者存取至特定或所有佇列，而且也可限制工具的使用，例如特定平行環境（請見第 251 頁的「有關平行環境」）。

為定義存取權限，就必須先定義 *使用者存取清單* — 構成已命名仲裁重疊或未重疊之使用者集。使用者名稱與 UNIX 群組名稱可用來定義這些使用者存取清單。之後，使用者存取清單就會用於叢集配置（請見第 145 頁的「基礎叢集配置」一節）、佇列配置（請見第 160 頁的「如何配置次級佇列」一節）、或配置平行環境介面的作業流程中（請見第 252 頁的「如何以 QMON 配置 PE」一節），來拒絕或允許存取至一特定資源。

## ▼ 如何以 QMON 配置使用者存取清單

當您選擇「使用者集」標籤時，就會開啓「使用者集」配置對話方塊，如圖 8-3 所示。

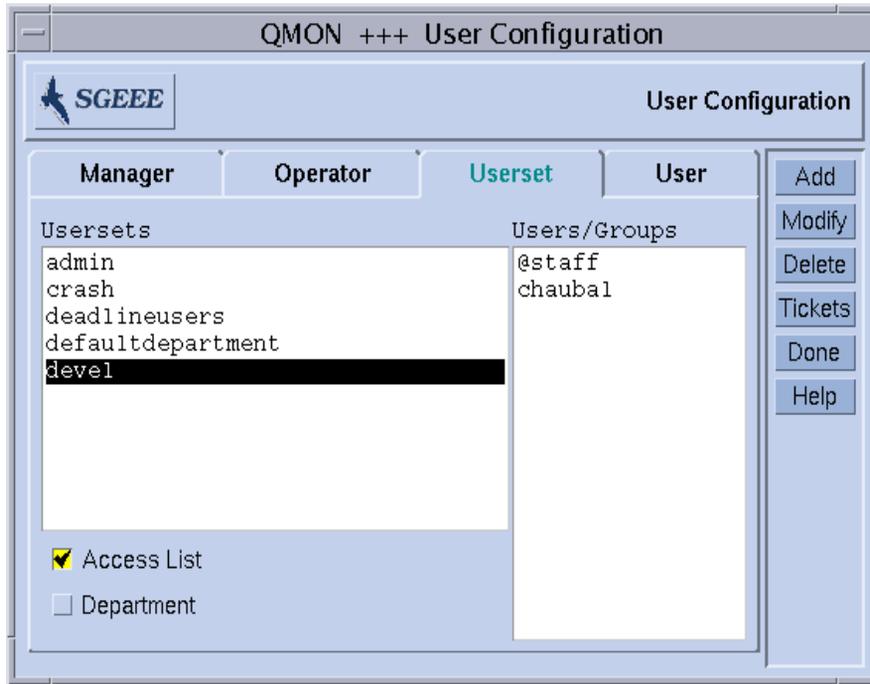


圖 8-3 使用者集配置對話方塊

可用存取清單顯示於螢幕左方的「使用者集」選擇清單中。要於「使用者/群組 (Users/Groups)」顯示欄位內顯示存取清單的內容，按一下「存取清單」選擇清單。

---

**注意** – 群組之前會以 @ 記號與使用者做欄位別。

---

在 Sun Grid Engine，企業版中，「使用者集」可以為「存取清單」、可以為「部門」，或兩者皆是。在「使用者集」選擇清單中，這兩個對應的旗號代表種類。本節預設所有「使用者集」為存取清單。請參閱第 178 頁的「關於使用「使用者集」來定義「專案」與「部門」」一節得知「部門」的說明。

您可以使用「使用者集」配置對話方塊來執行下述作業。

- **刪除** – 只要按一下名稱，再按下對話方塊右方的「刪除」鈕，即可從「使用者集」選擇清單上刪除一現有存取清單。
- **新增** – 按下「新增」鈕新增一使用者集。
- **修改** – 按下「修改」鈕修改選取之存取清單。

要新增與修改時，就會開啓「存取清單」定義對話方塊，如圖 8-4 所示，並提供對應的方法。

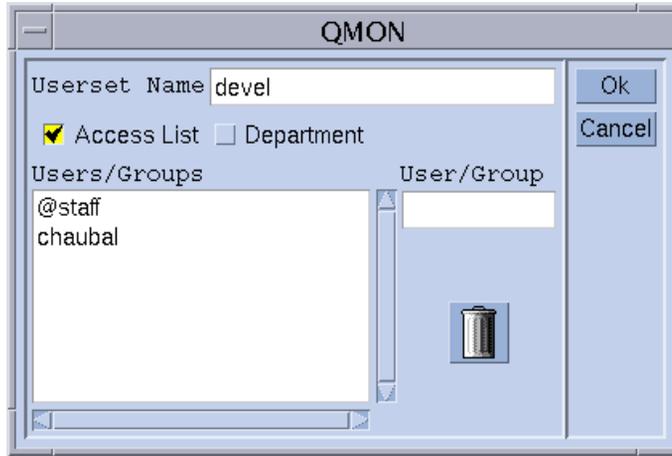


圖 8-4 存取清單定義對話方塊

### 「存取清單」定義對話方塊視窗解說

- 「使用者集名稱 (Userset Name)」輸入視窗 — 顯示「修改」作業中選取之存取清單名稱，或者您也可以自行輸入欲宣告之存取清單名稱。
- 「使用者 / 群組」顯示欄位 — 包含目前定義的存取清單項目。
- 「使用者 / 群組」輸入視窗 — 必須用以新增新項目至存取清單中。

在您按下鍵盤上的「Return」後，已輸入之使用者或群組名稱（群組前有 @ 記號）就會出現於「使用者 / 群組」顯示欄位。您可以於顯示欄位選擇這些記錄，然後按下垃圾筒圖示鈕來刪除這些記錄。

爲了 Sun Grid Engine，企業版中的存取清單定義，請確認已選取「存取清單」旗號。請參見第 178 頁的「關於使用「使用者集」來定義「專案」與「部門」一節得知「部門」旗號的相關說明。

在您按下 Ok 鈕時就會立刻記錄已修改或新增定義之存取清單，而按下「取消」鈕時就會立即捨棄所做的變更。兩種情況下皆會開啓「存取清單」定義對話方塊。

## ▼ 如何從指令行配置使用者存取清單

- 使用合適的選項輸入以下指令。

```
# qconf 開關
```

### 可用的選項

- `qconf -au 使用者名稱 [...]` 存取清單名稱 [...]  
新增使用者 — 本指令可新增一或多名使用者至特定存取清單中。
- `qconf -Au 檔案名稱`  
從檔案新增一使用者存取清單 — 本指令使用一配置檔案 `檔案名稱` 來新增一存取清單。
- `qconf -du 使用者名稱 [...]` 存取清單名稱 [...]  
刪除使用者 — 本指令可從特定存取清單中刪除一或多名使用者。
- `qconf -dul 存取清單名稱 [...]`  
刪除使用者清單 — 本指令可完全移除使用者集清單。
- `qconf -mu 存取清單名稱`  
修改使用者存取清單 — 本指令可用於修改特定存取清單。
- `qconf -Mu 檔案名稱`  
從檔案修改使用者存取清單 — 本指令使用一配置檔 `檔案名稱` 來修改指定的存取清單。
- `qconf -su 存取清單名稱 [...]`  
顯示使用者存取清單 — 本指令可顯示指定的存取清單。
- `qconf -sul`  
顯示使用者存取清單 — 本指令會列出目前定義之所有存取清單列表。

## 關於使用「使用者集」來定義「專案」與「部門」

「使用者集」還可用來定義 Sun Grid Engine，企業版專案（請見第 182 頁的「關於專案」）與部門。「部門」可用來配置 Sun Grid Engine，企業版之功能（請見第 204 頁的「關於功能 (Functional) 策略」）與置換策略（請見第 211 頁的「關於置換策略」）。「部門」與存取清單不同：使用者僅可為單一「部門」之成員，而相同的使用者卻可包含至多個存取清單中，此外，保留的 `deadlineusers` 名稱之「使用者集」包含所有允許可透過 Sun Grid Engine，企業版軟體提交截止期限之作業（請見第 208 頁的「關於截止期限策略」）。

「使用者集」會由「部門」旗號辨識為部門，如圖 8-3 與圖 8-4 所示。如果一使用者集為一個部門，就可同時用於，並定義為使用者清單。但是，在任何部門中已限制使用者僅可顯示一個。

## 關於使用者物件配置

如果共享、功能、或置換策略（請見第 194 頁的「如何以 QMON 管理策略 / 票卷進階資源管理」）會為使用者所定義，則在定義這些策略前，Sun Grid Engine，企業版軟體就需要有這些已宣告之使用者名稱。透過「使用者」配置對話方塊宣告使用者

### ▼ 如何以 QMON 配置使用者物件

1. 在 QMON 主要功能表上按下「使用者配置」鈕。
2. 選擇螢幕上方的「使用者」標籤。  
「使用者」配置對話方塊就會開啓，如圖 8-5 所示。

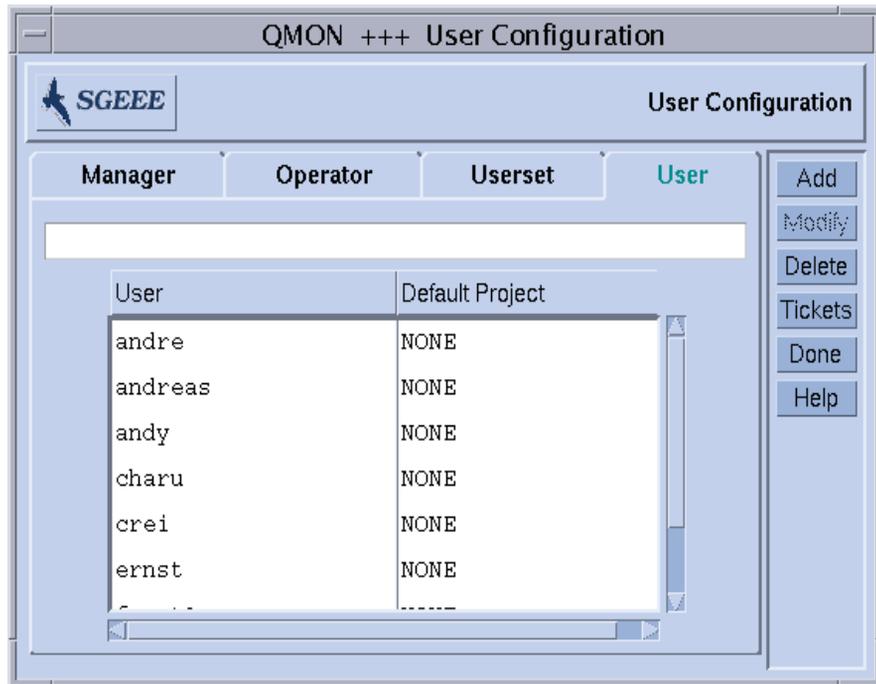


圖 8-5 使用者配置對話方塊

3. 依據您的目的，於對話方塊上方之輸入行輸入使用者名稱 — 或者，若方塊中已列入使用者名稱，就直接選擇此名稱 — 然後進行下述事項：

### 新增或刪除

- 新增一使用者名稱 — 在輸入使用者名稱後，按下鍵盤上的「新增」鈕或「Return」。
- 刪除使用者名稱 — 在選擇名稱後，按下「刪除」鈕。

## ▼ 如何指定預設專案

您可以指定預設專案（請見第 182 頁的「關於專案」）至每位使用者。此預設專案將隨附至使用者於提交時沒有要求其欲存取至另一專案的每項作業中。

1. 要指定預設專案，按一下使用者記錄來反白。
2. 按下清單上方之「預設專案」鈕。

「專案選擇」對話方塊就會開啓，如圖 8-6 所示。



圖 8-6 專案選擇對話方塊

3. 為反白之使用者記錄選擇合適之專案。
4. 按下「OK」來指定預設專案，然後關閉對話方塊。

## ▼ 如何從指令行配置使用者物件

- 使用合適的選項輸入以下指令。

```
# qconf 選項
```

### 可用的選項

- `qconf -auser`

新增使用者 — 本指令會開啓一個透過 `$EDITOR` 或（根據預設值）`vi` 指定之編輯器內的範本使用者配置（請見《*Sun Grid Engine 5.3 和 Sun Grid Engine, 企業版 5.3 參考手冊*》裡的 `user` 記錄），然後讓您修改。在儲存您的變更並且退出編輯器後，這些變更就會記錄至 `sge_qmaster` 中。

- `qconf -Auser 檔案名稱`

從檔案新增使用者 — 本指令會從語法分析指定的檔案 — 必須有使用者配置範本格式 — 然後再新增使用者配置。

- `qconf -duser 使用者名稱 [...]`

刪除使用者 — 本指令會刪除一或多個使用者物件。

- `qconf -muser 使用者名稱`

修改使用者 — 本指令會修改一現有使用者記錄。本指令會透過 `$EDITOR` 或（根據預設值）`vi` 指定的編輯器中載入使用者配置，然後讓您修改。在儲存您的變更並且退出編輯器後，這些變更就會記錄至 `sge_qmaster` 中。

- `qconf -Muser 檔案名稱`

從檔案修改使用者 — 本指令會從語法分析指定的檔案 — 必須有使用者配置範本格式 — 然後再修改使用者配置。

- `qconf -suser 使用者名稱`

顯示使用者 — 本指令會顯示特定使用者之配置。

- `qconf -suserl`

顯示使用者清單 — 本指令會列出目前定義之所有清單列表。

---

## 關於專案

Sun Grid Engine，企業版專案可提供從多位使用者組織聯合運算作業，然後為所有屬於此專案之作業定義資源利用策略的一個方法。專案用於三個排程策略領域：

- 共享，當指定共享於專案中 (請參閱第 195 頁的「關於共享式的策略」一節)。
- 功能，當專案接收 1% 之功能票卷時 (請參閱第 204 頁的「關於功能 (Functional) 策略」一節)
- 置換，當管理員給予專案置換票卷時 (請參閱第 211 頁的「關於置換策略」一節)。

---

**注意** – 在專案用於任何三項策略前必須先被宣告。

---

Sun Grid Engine，企業版管理員會給予專案名稱與一些屬性來定義 Sun Grid Engine，企業版專案。Sun Grid Engine，企業版使用者可以於作業提交階段隨附一專案至一作業中連結專案與作業會影響作業分派 (依專案共享、功能、與 / 或置換票卷而有所不同)。

### ▼ 如何以 QMON 定義專案

Sun Grid Engine，企業版管理員會透過使用「專案」配置對話方塊來定義並更新 Sun Grid Engine，企業版專案的定義。

1. 在 QMON 主要功能表上按下「專案配置」圖示。

接著就會出現「專案配置」對話方塊，如圖 8-7 所示。

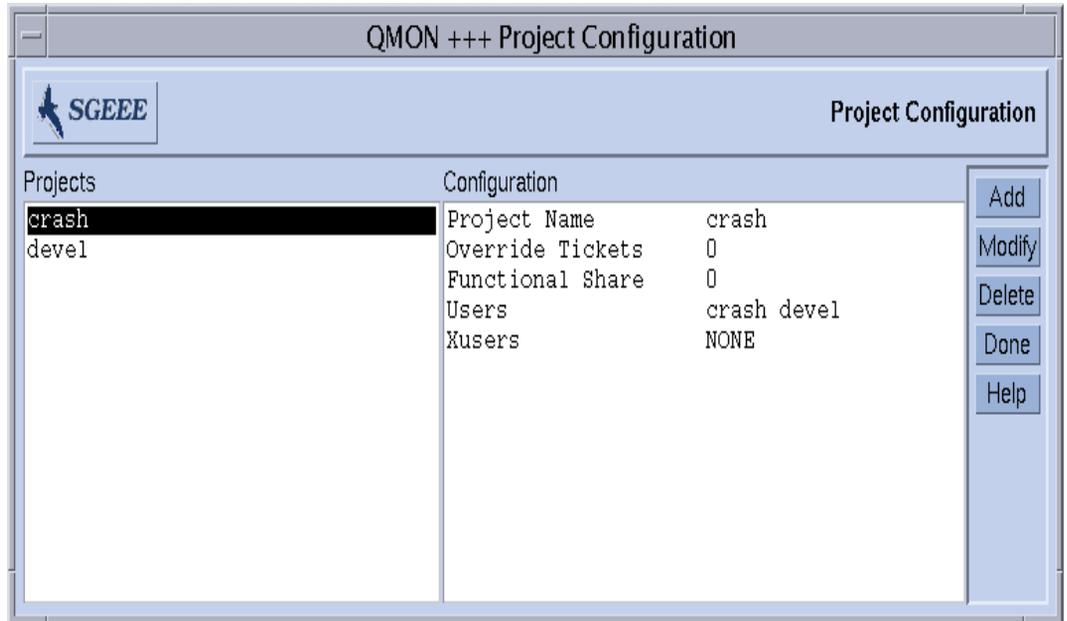


圖 8-7 專案配置對話方塊

已定義之專案會顯示於螢幕左方之「專案 (Projects)」選擇清單中。

- 按一下任何已列出專案之名稱。

此專案定義會顯示於「配置」視窗中。

- 根據您的目的按以下選擇標籤其中之一。

- 按一下「刪除」立即移除反白之專案。
- 按一下「新增」來新增一專案，或「修改」來修改反白之專案。

按「新增」或「修改」都將開啓「新增 / 修改專案」對話方塊，如以下圖 8-8 所示。

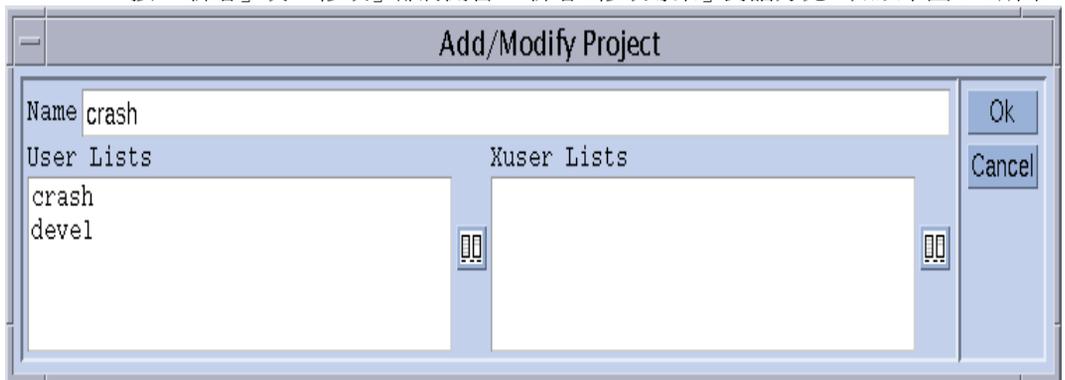


圖 8-8 「新增 / 修改專案」對話方塊

- c. 在「新增 / 修改專案」對話方塊中，請遵照以下指示進行。
- 在新增或修改一專案時，「新增 / 修改專案」對話方塊上方之「名稱」輸入欄位會顯示專案名稱。此專案是由被允許或拒絕存取至專案中的使用者所定義。
  - 隨附使用者存取清單（請見第 175 頁的「關於使用者存取權限」一節）來指定允許或拒絕至「使用者清單」（存取允許）或「X 使用者清單」（存取拒絕）。「使用者清單」隨附之存取清單內含的使用者或使用者群組為允許提交作業至專案中。「X 使用者清單」中的使用者或使用者群組為拒絕允許使用專案。*如果兩份清單都是空的，任何使用者皆可存取此專案。*如果一名使用者位於於不同之存取清單中，*同時* 隨附於「使用者清單」與「X 使用者清單」，此使用者則列為*拒絕*存取。
  - 要於「使用者清單」與「X 使用者清單」新增或移除使用者，請按一下「使用者清單」與「X 使用者清單」視窗右方的圖示鈕。之後將會開啓「選擇存取清單」對話方塊，如圖 8-9 範例所示。

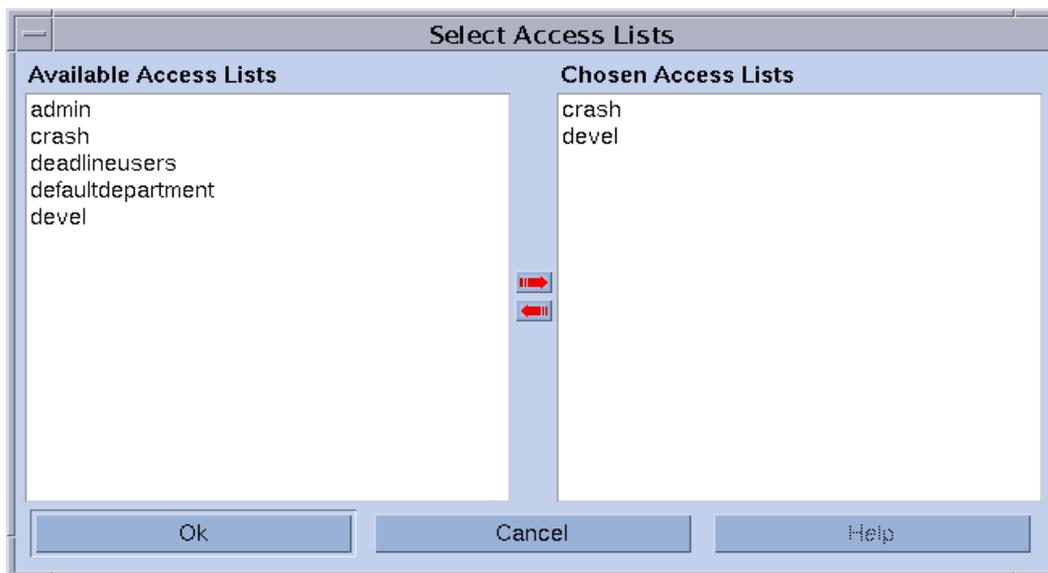


圖 8-9 選擇存取清單對話方塊

「選擇存取清單」對話方塊會於「可用存取清單」視窗中顯示所有已定義之存取清單，並隨附清單於「選取存取清單」視窗中。您可以於兩個視窗中選擇存取清單，然後使用箭頭圖示鈕於兩視窗中移動存取清單。

- d. 按一下「OK」鈕確認變更，然後關閉對話方塊。

## ▼ 如何從指令行定義專案

- 使用合適的選項輸入以下指令。

```
# qconf 選項
```

### 可用的選項

- `qconf -aprj`

新增專案 — 本指令會使用 `$EDITOR` 或（根據預設值）`vi`，來開啓一指定之編輯器內的範本專案配置（請見《*Sun Grid Engine 5.3 和 Sun Grid Engine，企業版 5.3 參考手冊*》裡的 `project` 記錄），然後讓您修改。在儲存您的變更並且退出編輯器後，這些變更就會記錄至 `sge_qmaster` 中。

- `qconf -Aprj 檔案名稱`

從檔案新增專案 — 本指令會從語法分析指定的檔案 — 必須有使用者配置範本格式 — 然後再新增專案配置。

- `qconf -dprj 專案名稱 [...]`

刪除專案 — 本指令會刪除一或多個專案。

- `qconf -mprj 專案名稱`

修改專案 — 本指令會修改現有使用者記錄。本指令會使用 `$EDITOR` 或（根據預設值）`vi` 指定的編輯器中載入專案配置，然後讓您修改。在儲存您的變更並且退出編輯器後，這些變更就會記錄至 `sge_qmaster` 中。

- `qconf -mprj 檔案名稱`

從檔案修改專案 — 此指令會從語法分析指定檔案 — 必須有專案配置範本格式 — 然後修改現有專案配置。

- `qconf -sprj 專案名稱`

顯示專案 — 本指令會顯示特定專案之配置。

- `qconf -sprjl`

顯示專案清單 — 本指令會列出目前定義之所有清單列表。

---

## 關於排程

Sun Grid Engine，企業版系統之作業排程活動包含以下幾種：

- 預先分派決定 — 為諸如消除執行佇列之類的活動，因為這些活動不是已滿便是過多，而目前位於等候區中之作業不適合執行。
- 分派 — 這些活動包含決定所有其他擱置中與執行作業之重要性、感應叢集中所有機器之載入、並遵照配置選擇標準將作業寄至選取機器上的執行佇列。
- 分派後監視 — 這些活動包含在作業得到資源，而其他作業與其相對重要性進入或離開系統時，調整其相對重要性。

Sun Grid Engine，企業版軟體會依據以下幾點，跨過電腦不同叢集來排程作業：

- 叢集目前載入
- 作業相對重要性
- 主機相對效能
- 工作資源需求（例如：CPU、記憶體、與 I/O 頻寬）

排程決定是依據站點，與叢集內每台電腦之即時載入特性之策略而定。站點之排程策略會透過 Sun Grid Engine，企業版系統之配置參數呈現。載入特性會於系統執行時收集效能資料來確定。

## 排程策略

管理員會依據以下 Sun Grid Engine，企業版排程作業來設定策略。

- 動態資源管理 — Sun Grid Engine，企業版會動態控制並調整分配至執行作業的應得資源（即修改其 CPU 共享）。
- 佇列分類 — 軟體會依據佇列填滿的順序，於叢集中排列佇列。
- 作業分類 — 確認 Sun Grid Engine，企業版系統試圖排程作業之順序。

## 動態資源管理

Sun Grid Engine，企業版軟體利用四種策略加權結合之其中一種來執行自動作業排程策略。

- 共享式的
- 功能（有時稱為優先權）
- 初始啟動截止日期 (Initiation Deadline)
- 置換

您可以設定 Sun Grid Engine，企業版系統來例行使用共享策略、功能策略、或兩者皆可。這兩個策略可以用任何換算結合，您可以不用其中一個而只用第二個，也可以給予兩個相同的加權。

與例行策略一同，也可與起始截止日期一起提交作業。截止日期作業擾亂例行排程。管理員也會暫時，或為特定目的，如顯示佇列，而永久置換共享、功能與初始啟動截止日期排程。置換作業適用於單獨作業，或所有與使用者、部門、專案、或作業等級（即佇列）相關之作業。

除了調解所有作業的四個策略之外，有時 Sun Grid Engine，企業版也會讓使用者設定其作業之優先順序。例如，提交許多作業的使用者會說作業 3 比較重要，而作業 1 與 2 則一樣重要，但又沒有作業 3 重要。如果 Sun Grid Engine，企業版系統策略結合包含共享策略、功能策略、或兩者皆有，再加上給予作業之功能票卷，就有可能發生這種情況。

排程策略會以票卷執行。每個策略都有一個票卷儲存欄位，分配票卷至進入多個機器的 Sun Grid Engine，企業版系統。每個執行中的例行策略會分配一些票卷至新作業中，有可能也會於每個排程間隔時間重新分配票卷至執行之作業中。每個策略用來分配票卷之標準說明如下。

票卷加權此四個策略。舉例來說，如果票卷沒有分配至功能策略，現在就不會使用此策略。如果票卷有平均分派之功能與共享票卷儲存欄位，則此兩項策略就擁有相同的確認作業重要性之加權。

由 Sun Grid Engine，企業版管理員分配票卷至系統配置上的例行策略。管理員與操作員可隨時改變票卷分配，並立刻生效。額外之票卷會暫時置於系統中，代表截止日期或置換。策略會由票卷指定結合—當票卷被分配至多個策略時，作業會得到每個策略中的部分票卷，以顯示其於執行中之策略的重要性。

Sun Grid Engine，企業版將票卷給予進入系統中的作業，以顯示其於執行中策略之重要性。每個執行中的作業會於每個排程間隔時間得到（例如，從置換中得到，或者因為截止日期已逼近）、失去（例如，因為擁有太多共享資源）、或保留相同數量的票卷。作業持有之票卷數量代表 Sun Grid Engine，企業版試圖於每個排程間隔時間給予此作業之資源共享。

站點之動態資源管理策略會於 Sun Grid Engine，企業版安裝期間，由分配票卷至共享與功能排程策略、定義共享樹與功能共享、與設定初始啟動截止日期票卷最大量來完成配置。共享與功能票卷分配，和初始啟動截止日期票卷最大量可隨時自動變更。置換票卷可由管理員手動指派或移除。

## 佇列分類

下述方法是用來確認 Sun Grid Engine，企業版試圖填滿佇列的順序。

- 載入回報 — Sun Grid Engine，企業版管理員可以選擇使用哪個載入參數來比較主機與其佇列之載入狀態。第 245 頁的「載入參數」一節說明可用標準載入參數之廣泛範圍，與利用站點特定載入感應器來延申此設定之介面。
- 載入換算 — 由不同主機接收之載入回報可使其規範化，並反映一比較情況（請見第 136 頁的「如何以 QMON 配置執行主機」一節）。
- 負載調整 — Sun Grid Engine，企業版軟體可配置為當分派作業至主機時，自動修正最後回報載入。已修正之載入數值會顯示由最新啟動作業所造成之載入情況的預期增加。當這些作業的載入發揮影響時，就會自動減少此人為載入增加。
- 序列編號 — 可依周密的序列分類佇列。
- 主機容量 — 主機與位於主機內的佇列可依據容量指示器而分類，於叢集中定義機器間的相對能力。

## 作業分類

在 Sun Grid Engine，企業版開始分派作業前，作業會先以高優先序列排列。Sun Grid Engine，企業版會企圖以優先序列為各作業尋找合適的資源。若沒受到任何管理員的影響，順序會成「先進先出」(FIFO) 排列。管理員會以下述幾種方式控制作業順序。

- 票卷為主之作業優先順序 — 在 Sun Grid Engine，企業版中，作業處理乃根據其相對的重要性，而該重要性為作業所擁有的票卷數量而定義。因此，擱置中的作業會以票卷排列順序分類，而且一但任何票卷策略改變管理員，也會改變分類的順序。
- 使用者 / 群組作業最大量 — 同時執行於 Sun Grid Engine，企業版系統的使用者或 UNIX 使用者群組擁有之作業最大量可受到限制。這會影響擱置中作業清單分類順序，因為未超出限制之使用者作業將享有優先權。

## 在排程器間隔時間內發生的事

排程器會於間隔時間內排程作業。在排程進行期間，Sun Grid Engine，企業版會保留重大事件之資訊，例如作業提交、作業完成、作業取消、叢集配置之更新、或叢集內新機器之記錄。當開始排程時，排程器的工作如下。

- 考慮所有重大事件。
- 依據管理員規格來分類作業與佇列。
- 考慮所有作業資源要求。

然後若有需要，Sun Grid Engine，企業版系統工作如下。

- 分派新作業。
- 暫時執行中的作業。
- 增加或減少分配至執行中作業之資源。
- 維持現狀。

如於 Sun Grid Engine，企業版系統中使用共享排程，在排程中就會將使用者或專案的使用記錄列入計算之中。如果排程不是（至少某部分）以共享為主，在排程時會僅排列所有執行中與等候執行之作業，並依重要性排列，直到排程完整利用叢集中的資源（CPU、記憶體、與 I/O 頻寬）為止。

## 排程器監視

如果作業尚未開始，而且原因不明，您就可以於作業上執行 `qalter` 與 `-w v` 選項。Sun Grid Engine，企業版軟體會預設空白叢集，並檢查是否有適合此作業之可用佇列。

您可執行 `qstat -j 作業id` 以得知更多相關資訊。之後便會列出作業要求簡介之摘要，同時也會包含為何作業沒有加入最後排程的原因。不輸入作業 ID 來執行 `qstat -j` 將會摘要所有無法於最後排程間隔時間內列入排程作業的原因。

---

**注意** – 排程原因資訊的收集必須開啓於排程器配置 `sched_conf`。請參閱對應 《*Sun Grid Engine 5.3* 和 *Sun Grid Engine, 企業版 5.3 參考手冊*》中的 `schedd_job_info` 參數，或第 191 頁的「如何以 QMON 配置變更排程器配置」一節。

---

要得到 Sun Grid Engine, 企業版 scheduler `sge_schedd` 決定的更詳細資訊，您可使用 `qconf` 指令的 `-tsm` 選項。此指令將執行 `sge_schedd` 以將追蹤輸出寫入檔案中。

## 排程器配置

請參閱第 194 頁的「如何以 QMON 管理策略 / 票卷進階資源管理」以得知 Sun Grid Engine, 企業版以票卷以主之資源共享策略的排程管理細節。本節提示主要放在管理排程器配置 `sched_conf` 與其相關主題上。

## 預設排程

預設的 Sun Grid Engine, 企業版排程為先進先出策略，即先提交之作業可先由排程器檢驗分派至佇列中。如果擱置作業清單中的第一個作業找到合適且閒置的佇列，此作業就會先列於排程執行中。但如果第一個作業無法找到一合適可用的資源，第二個作業或之後的作業就會在擱置作業清單中的第一個作業之前，先列入排程執行中。

只要關係到作業佇列選擇，預設的 Sun Grid Engine, 企業版策略就會於載入量最少的主機中，選擇能為作業資源要求提供合適服務的佇列。如果有多個合適佇列共享相同載入，就無法預知會選擇哪一個佇列。

## 排程選擇

有許多方法可修改作業排程與佇列選擇策略。

- 改變排程演算法
- 按換算排列系統載入
- 依序列編號選擇佇列
- 依共享選擇佇列
- 限制一使用者或一群組之作業數量

以下幾節將詳細解說這些選擇方法。

## 改變排程演算法

排程器配置參數 `algorithm` (請見《*Sun Grid Engine 5.3 和 Sun Grid Engine, 企業版 5.3 參考手冊*》中的 `sched_conf` 記錄得知更多資訊) 是設計為使用中的排程演算法提供一個選項。目前為止, `default` 是唯一允許之設定。

## 按換算排列系統載入

Sun Grid Engine, 企業版系統使用管理佇列機器上的系統載入資訊為一作業選擇執行佇列。本佇列選擇計劃會建立載入平衡情況, 保證能更適當使用叢集之可用資源。

但是系統載入數值並非總是真實的。例如, 若比較多重 CPU 機器與單一 CPU 機器系統, 多重處理器系統通常會回報較高的載入數值, 因其也許會執行更多程序, 而且系統載入數值會嚴重受到試圖進入 CPU 存取的作業數量。但是與單一 CPU 機器相比, 多重 CPU 系統能符合較高載入數值。此問題於載入數值處理器數量調整組 (由 `sge_execd` 預設回報) 中有詳細論述 (請見第 245 頁的「載入參數」一節與 `<sge_root>/doc/load_parameters.asc` 檔得知詳細資訊)。使用這些載入參數, 而非原始載入數值, 以避免上述提及的問題發生。

另一潛在載入數值錯誤解譯之範例為系統之潛在效能或其價格效能換算有很大的差異, 兩者相同的載入數值並不代表可選取仲裁主機以執行作業。在此情況下, Sun Grid Engine, 企業版管理員應定義執行主機與載入參數之載入換算要素 (請參閱第 136 頁的「如何以 QMON 配置執行主機」與其相關節)。

---

**注意-** 依換算載入數值參數也可用於比較兩者間的載入臨界值清單 **載入臨界值** 與 **遷移載入臨界值** (請參閱《*Sun Grid Engine 5.3 和 Sun Grid Engine, 企業版 5.3 參考手冊*》中的 `queue_conf` 記錄得知詳細資訊)。

---

與載入參數相關的其他問題為需要解譯數值與其相對重要性之應用程式與站點相依。CPU 載入會由特定站點常見之特定應用程式類型所主導, 而對另一站點與站點計算叢集通常存在之應用程式簡介而言, 記憶體載入的重要性又更高。Sun Grid Engine, 企業版允許管理員於排程配置檔案 `sched_conf` 指定一所謂的 **載入準則** 來提出這問題 (請參閱對應《*Sun Grid Engine 5.3 和 Sun Grid Engine, 企業版 5.3 參考手冊*》以得知詳細資訊)。使用載入準則中的站點定義載入參數 (請見第 246 頁的「新增站點特定載入參數」一節) 與可耗損資源 (請見第 234 頁的「可耗損資源」一節), 即可將資源利用與容量規劃的站點特定資訊列入考慮。

最後, 也需將載入參數之時間相依也列入考慮。位於執行在系統上的 Sun Grid Engine, 企業版作業載入數值會隨著時間不同而改變, 而且通常 — 以 CPU 載入為例 — 都需要一些時間等待作業系統回報合適的數量。因此, 如果最新執行一作業, 回報的載入數值就無法提供此作業已於主機上的完整數值。此回報的負載會隨著時間適應真正的負載, 但在一段時間內若回報的負載若過低, 就會導致主機過度耗損。Sun Grid Engine, 企業版允許管理員指定用於 Sun Grid Engine, 企業版排程器之 **負載調整** 指數來補償這個問題。請參閱可解決排程器配置檔 `sched_conf` 的《*Sun Grid Engine 5.3 和 Sun Grid Engine, 企業版 5.3 參考手冊*》以得知如何設定這些負載調整要素之詳細資訊。

## 依序列編號選擇佇列

另一個改變預設佇列選擇計劃的方法為設定全域 Sun Grid Engine，企業版叢集配置參數 `queue_sort_method` 至 `seq_no`，而非預設 `load`（請參見《*Sun Grid Engine 5.3 和 Sun Grid Engine，企業版 5.3 參考手冊*》中的 `sched_conf` 記錄）。在此情況下，系統載入數值就不會用於選擇佇列。反而會由佇列配置參數 `seq_no`（請見《*Sun Grid Engine 5.3 和 Sun Grid Engine，企業版 5.3 參考手冊*》中的 `queue_conf` 記錄）指定佇列的序列編號，再用來定義選取之佇列之間的固定順序（如果這些序列編號適用於考慮中的作業，而且是閒置的）。

如果您站點提供批次服務之機器是以一作業順序一價格的排列，此佇列選擇策略也許會有幫助。例如，執行於 A 機器的作業花費一單元的金錢，但它於 B 機器要花費 10 個單元，而 C 機器要花費 100 個單元。因此此優先的排程策略會首先填滿 A 主機，之後 B 主機，到別無選擇後才是 C 主機。

---

**注意** – 如果您已變更佇列選擇至 `seq_no` 的方法，而且考慮的佇列皆共享相同之序列編號，則將由預設 `load` 選擇佇列。

---

## 依共享選擇佇列

此方法的目的為配置作業以試圖符合每個作業的全域系統資源之目標共享。本方法會將關於所有系統資源之每個主機的資源使用性列入考慮，並試圖平衡每個主機的 Sun Grid Engine，企業版票卷換算（即主機上所有執行作業的 Sun Grid Engine，企業版票卷總和），與系統特定主機呈現之資源可用性換算。請參閱第 136 頁的「如何以 QMON 配置執行主機」以得知如何定義主機容量的指示。

雖然主機載入數值被列為次等重要，但於分類中還是會將其列入考慮。這應屬於使用共享樹策略之站點的選擇分類法。

## 限制一使用者或一群組之作業數量

Sun Grid Engine，企業版管理員會指定在任何時間內使用者或 UNIX 群組可執行之作業數量的上限。為了執行此功能，設定《*Sun Grid Engine 5.3 和 Sun Grid Engine，企業版 5.3 參考手冊*》之 `sched_conf` 一節說明之 `maxujobs` 與 / 或 `maxgjobs`。

## ▼ 如何以 QMON 配置變更排程器配置

1. 在 QMON 主要功能表上按一下「排程器配置」。

就會出現「排程器配置」的對話方塊。此對話方塊分隔成「一般參數」欄位與「負載調整」欄位。您可依據您的目的選擇兩者其中之一。

a. 要改變一般排程參數，請按一下「一般參數」標籤。

圖 8-10 為「一般參數」對話方塊範例。



圖 8-10 排程器配置對話方塊 — 一般參數

從「一般參數」對話方塊中，您可以設定以下參數。

- 排程器演算法（請見第 190 頁的「改變排程演算法」）
- 排程器執行之規律間隔時間
- Sun Grid Engine，企業版排程器執行之規律間隔時間，依據資源共享策略而重新分配票卷。
- 一使用者與一 UNIX 群組允許同時執行之作業最大量（請見第 191 頁的「限制一使用者或一群組之作業數量」）。
- 佇列分類方案—由載入數值或由序列編號分類（請見第 191 頁的「依序列編號選擇佇列」）或由共享分類（請見第 191 頁的「依共享選擇佇列」）。
- 作業排程資訊能否透過 `qstat -j` 存取，或此資訊能否應僅由隨附輸入欄位中指定之作業 ID 收集。建議僅暫時開啓作業排程資訊的一般收集，以避免擱置作業量過多。
- 用以分類主機與佇列之載入準則

b. 要改變負載調整參數，請選擇「負載調整」標籤。

圖 8-11 為「負載調整參數」對話方塊範例。

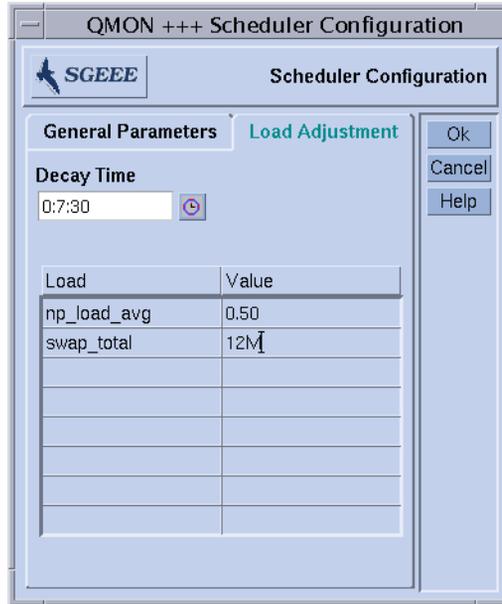


圖 8-11 排程器配置對話方塊 — 負載調整

「負載調整」對話方塊使您可定義以下參數。

- 負載調整延遲時間
- 方塊下半方的負載調整值圖表會列出所有最近已定義的調整值之載入與可耗損屬性。可按一下上方的「載入」或「數值」鈕來增加清單。之後會開啓一選擇清單，列出所有隨附於主機的屬性（即配置於全域、主機、與管理員定義之聚合體）。「屬性選擇」對話方塊圖顯示於圖 6-6。選擇其中一個屬性，並按「OK」鈕確認選項，就會新增此屬性至「可耗損 / 固定屬性 (Consumable/Fixed Attributes)」圖表之「載入」欄，並於相關「數值」欄位加入游標。連按二下「數值」欄位以修改現有數值。選擇對照表行，然後輸入 CTRL-D 即可刪除屬性—或按一下滑鼠右鍵開啓刪除方塊，然後確定刪除動作。

請見第 190 頁的「按換算排列系統載入」以得知相關背景資訊。請參閱《Sun Grid Engine 5.3 和 Sun Grid Engine，企業版 5.3 參考手冊》裡的 sched\_conf 手冊頁以得知排程器配置的詳細資訊。

## ▼ 如何以 QMON 管理策略 / 票卷進階資源管理

1. 在 QMON 主要功能表上按下「票卷配置」鈕。

圖 8-12 為「票卷概觀 (Ticket Overview)」對話方塊範例。

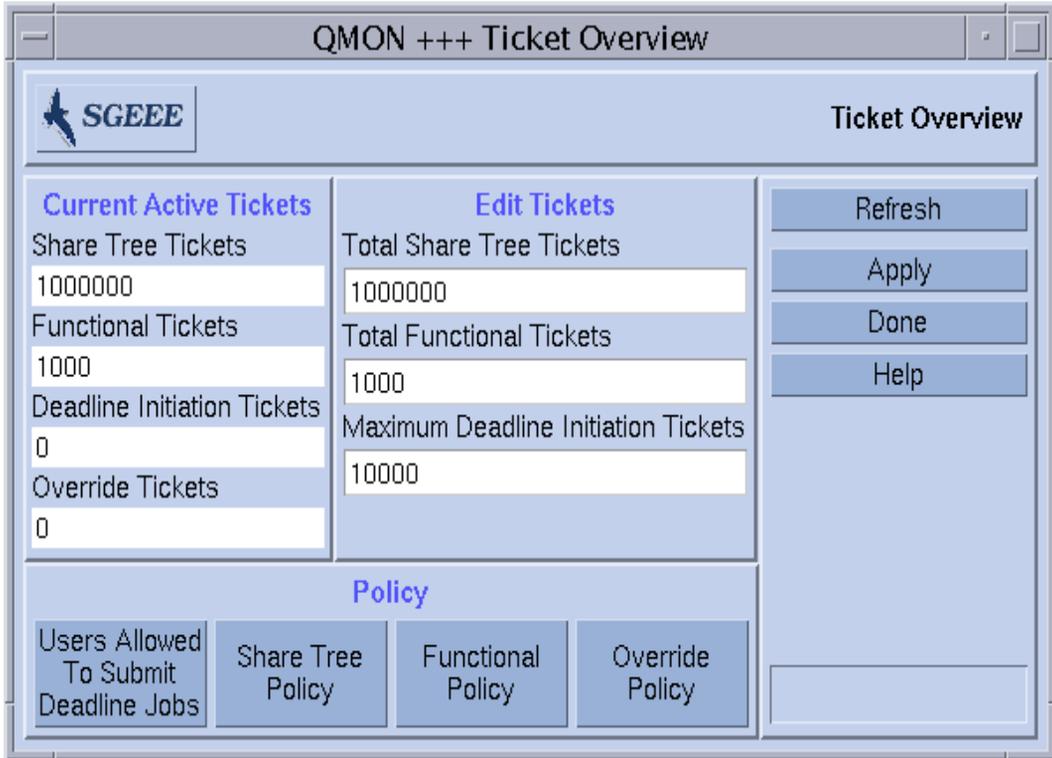


圖 8-12 票卷概觀對話方塊

2. 依下節指導進行。

「票卷概觀」對話方塊會顯示票卷策略的最近票卷分配，讓您可重新調整策略相關票卷，並為所有票卷策略提供可至特定配置對話的通道。

最新分派至單獨策略的票卷會顯示於左方的「最新活動票卷 (Current Active Tickets)」顯示欄位中。數量反映了策略的相對重要性，並顯示一特定策略最近是否主導叢集，與策略間是否有平衡。票卷提供一數量測量，意即指派至共享策略的票卷會是功能策略的二倍，舉例來說，分配至共享策略的應得資源為功能策略的二倍。此時，票卷就與股份非常相似。

所有票卷的總數量沒有任何特殊意義。策略間的關連才有意義。因此，總票卷數量通常都很多，可用來細微調整策略間的相對關係。

## 編輯票卷 (Edit Tickets) 區

「編輯票卷」區可用來修改分配至各策略（除了置換策略）的票卷。置換票卷會直接透過置換票卷配置分配，而其他票卷儲存欄位會自動分配至與策略和策略配置相關之作業。

---

**注意** – 所有共享與功能票卷總是分配至與各策略相關之作業。截止期限票卷僅會在截止期限作業接近截止期限時分配。置換票卷也許不適用於最近執行中的作業，因此當置換策略定義票卷時，執行中的置換票卷會為 0。

---

### 「策略」按鈕區

本區提供以下功能。

- 一個可開啓「使用者配置」對話方塊的按鈕，可輕易存取至「截止期限使用者使用者集 (Dealineusers Userset)」配置。
- 一個可開啓共享、功能、與置換策略配置對話方塊的按鈕—沒有截止期限策略的配置對話方塊。

面板右側的按鈕讓您可「重新整理」螢幕、「套用」或「捨棄（完成）」變更。

## 關於共享式的策略

共享式的（又稱*共享樹*）排程可在一段期間，如一禮拜、一個月、或一季，給予每位使用者與專案分配到的系統資源。藉由不斷調整最近期間每位使用者與專案的潛在資源共享來完成分配資源的工作（直到下個排程間隔時間）。共享排程是由使用者、專案、或兩者定義。

只要盡量給予每位使用者 / 專案其目標共享，群聚在一起的使用者 / 專案，如部門或欄位域，也會得到其目標共享。當所有享有資源之實體於累積期間相互競爭資源時，才有可能得到平等之共享。如果使用者 / 專案無法於一段期間提交作業，資源將會分配至此段期間可提交之作業。

共享式排程為*回饋計劃*。任何使用者 / 使用者群組，與專案 / 專案群組應得之系統共享皆為一個 Sun Grid Engine，企業版配置參數。任何作業應得之系統共享是根據以下要素決定。

- 配置至作業使用者或專案之共享
- *延遲因數*（即「原來的」使用影響不大）會調整每位使用者 / 使用者群組，與專案 / 專案群組的累積使用記錄

Sun Grid Engine，企業版會持續追蹤已接收之使用者 / 專案的用法。在每個排程間隔期間，排程器會調整所有資源共享，以確認所有使用者 / 使用者群組與專案 / 專案群組可於累積期間得到接近其系統之平等共享。換句話說，給予或拒絕給予資源是爲了讓每個人都能使用其目標共享。

## 半衰期 (Half-life) 因數

「半衰期」為系統可多快「忘記」使用者消耗的資源。此系統管理員可於六個月前或六天前，決定是否該，或該如何處罰一使用者消耗過多資源。在共享樹每個節點上，Sun Grid Engine 企業版軟體會保有使用者的資源消耗記錄。

有了這個記錄，系統管理員就可決定多久回顧一次來確認使用者在設定共享策略時，是否有低度或過度使用的情況。本文中的資源使用為在「時間的滑動視窗 (sliding window of time)」消耗之所有電腦資源的整數（總和）。

此視窗的長度是由「半衰期」因素決定，在 Sun Grid Engine 企業版系統為一內部延遲功能。此延遲功能會降低一段時間內資源消耗量增加的影響。短暫的半衰期很快便減少資源過度耗損的影響；而長期的半衰期則是逐漸減少資源過度耗損的影響。

在 Sun Grid Engine 企業版系統中，此半衰期延遲功能為一時間特定單元。舉例來說，一段時間過後，適用於 1,000 單元之資源耗損，為期七天的半衰期會導致以下使用「處罰 (penalty)」的調整。

- 七天後 500
- 十四天後 250
- 二十一天後 125
- 二十八天後 62.5

此半衰期延遲會消除使用者於一段時間後消耗資源產生的影響，直到處罰效應漸漸變小為止。請注意：如果使用者接收到**置換票卷**，此票卷並不屬於過去使用處罰，因其屬於另一個不同的策略系統。只有共享樹策略才有延遲功能。

## 補償因數

當比較結果顯示實際用法低於目標用法，調整使用者 / 專案的共享資源可讓使用者以達到目標共享為主主導系統。此主導局勢也許不理想。**補償因數** 允許管理員限制使用者 / 專案以非常少的累積用法主導最近期間的資源，以達到指定的用法目標。

舉例來說，補償因數 2 會限制使用者 / 專案的最近共享，使其目標共享增加二倍。意即，如果使用者 / 專案需於累積期間得到 20% 的系統資源，但現今得到的資源卻很少，於最近期間就可得到 40% 的資源。

與共享策略，即定義一共享樹之使用者 / 專案的長期應得資源相結合時，補償因數就會自動調整應得資源。

如果一特定使用者或專案**低於**或**高過**定義之目標應得資源，Sun Grid Engine 企業版系統就會**提高**或**降低**短期，或長期目標之使用者或專案的應得資源。Sun Grid Engine，企業版系統之共享樹演算計算會執行此補償因數。

此補償因數會於上方提供一新增機制來控制 Sun Grid Engine，企業版系統指定的補償數量。若下述事實成立才會啟動此新增補償因數 (CF) 計算。

- 短期應得資源 > 長期應得資源 \* CF
- CF > 0

如果上述一個或兩者皆非，就會使用共享樹演算法定義並執行之補償。

一個設定補償因數的一般規則為 CF 值越小，其影響便越大。如果數值大於 1，Sun Grid Engine，企業版系統就會開始補償 — 不過補償會受到限制。補償較高之限制被計算為長期應得資源 \*CF。同時請注意：如上述定義，短期應得資源必須於因補償因數而發生任何事件之前超出此限制。

如果此數值 =1，則 Sun Grid Engine，企業版系統就會以與原始共享樹演算法一般的方法做補償。因此，數值 1 的影響與數值 0 類似。唯一的差別為一執行細節，即當 CF=0，且抑制 CF 計算時，才會執行 CF 計算（無任何影響）。

如果此數值 <1，則 Sun Grid Engine 系統就會「過度補償」。以共享樹演算法為主，作業接收過多其應得之資源，同時也提早接收此過度補償，因為短期應得資源 > 長期應得資源 \* CF 之執行標準，符合較低之短期應得資源值。

### **階層式共享樹 (Hierarchical Share Tree)**

共享策略會透過於移動累積期間，指定系統資源如何共享於所有使用者 / 與專案之階層式共享樹執行。累積期間的長度是由一配置延遲常數確認。Sun Grid Engine，企業版會依共享樹內每個母節點達到其累積限制的程度，來共享其應得資源。作業應得共享資源是根據其子節點共享資源而分配，子節點則依其母節點而分配。所有與子節點相關之作業會分裂成相關共享。

由共享樹分配之應得資源會與其他資源相結合（例如：從截止期限或功能策略分配之應得資源），用以確認作業之最終應得資源。於共享排程中，共享樹會被分配票卷總數量。數量多寡可確認共享排程於四個排程策略之中的加權。

共享樹會於安裝 Sun Grid Engine，企業版時定義，可於任何時間變更。當編輯共享樹時，新的共享分配就會於下一排程間隔時間生效。

## ▼ 如何從 QMON 編輯共享樹策略

1. 在 QMON 「票卷概觀」對話方塊下方，按一下「共享樹策略」。

圖 8-13 為「共享樹策略」對話方塊範例。

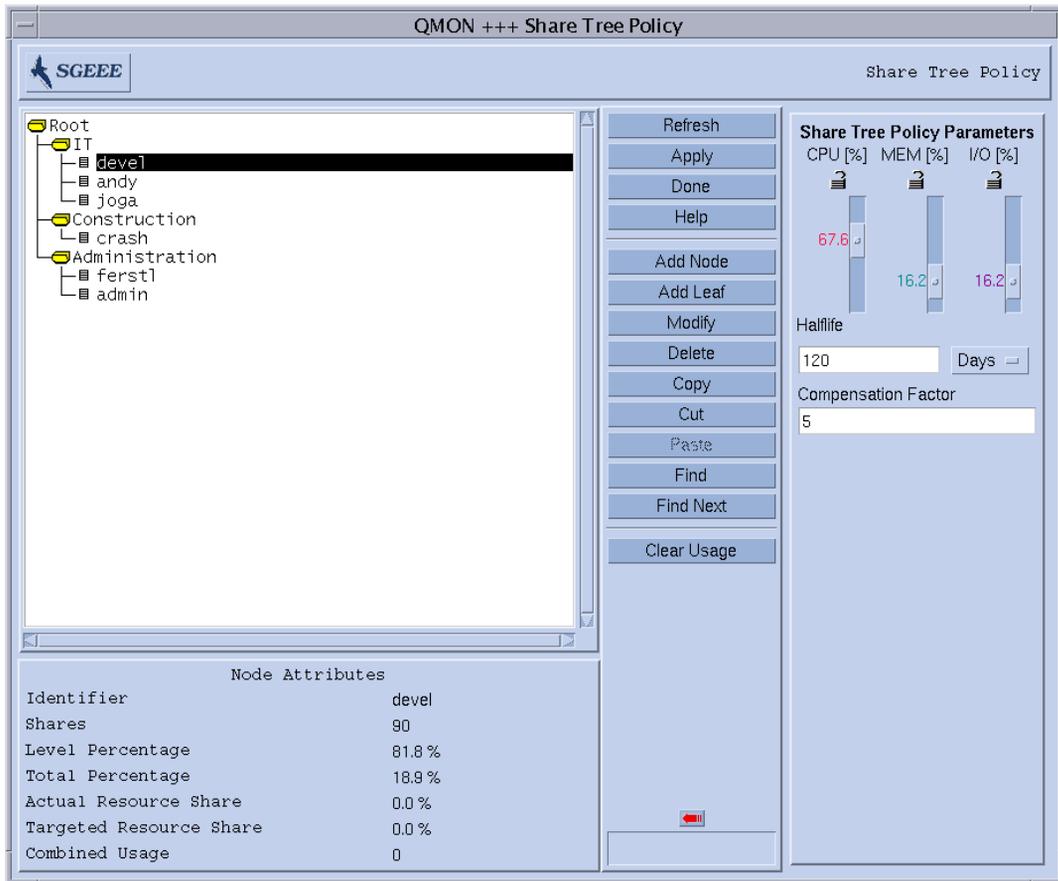


圖 8-13 共享樹策略對話方塊

2. 遵照以下幾節指示執行編輯策略。

### 節點屬性顯示區

此區會顯示選取節點的屬性：

- 辨識器 — 使用者、專案、或群聚 (agglomeration) 的名稱。
- 共享 — 分配至使用者或專案的共享數量。

---

**注意** – 共享會定義相對重要性，沒有任何百分比，也沒有任何數量意義。建議挑選百位或千位數量，如此便可微調重要性的關係。

---

- **等級百分比** – 節點在共享樹上於其等級（相同母節點）占據之總共享部分；其共享是由此節點與其相關節點總和劃分。
- **總百分比** – 此節結於全部共享樹中占有總共享的換算。此為共享策略之節點長期目標資源共享。
- **實際資源用法** – 節點目前於累積期間消耗之總系統資源換算。此百分比會顯示共享樹上之所有節點。
- **目標資源用法** – 同上，但僅將共享樹中最近活動的節點列入考慮。活動中的節點在系統中有作業進行。在短期間，Sun Grid Engine，企業版會試圖平衡活動中節點的應得資源。
- **結合用法** – 節點的全部用法。結合用法為節點累積的用法總和。子節點會累積所有執行其中的作業用法。內節點 (inner node) 則累積所有相關節點的用法。結合用法是以「共享樹策略參數 (Share Tree Policy Parameters)」對話欄位指定之換算為依據，由 CPU、記憶體與 I/O 用法組成，並以指定之半衰退延遲換算延遲。

當移除使用者或專案節點（指子節點）後又新增回共享樹的原位或另一地方時，就會保留此使用者或專案的用法。如果您想在新增回使用者或專案節點前讓用法歸零，應於 Sun Grid Engine，企業版配置的使用者 / 專案中移除，並新增回此使用者 / 專案。

即使一使用者或專案並未包含至共享樹內，在新增至共享樹時，還是可以執行其將有非 0 用法之作業。同樣地，如果使用者或專案於新增至共享樹時用法最好為 0，則在新增至共享樹前，就應先將用法從配置於 Sun Grid Engine，企業版的使用者或專案中移除。

## **重新整理**

圖形使用者介面會定期更新顯示的資訊。此按鈕會強制立即顯示重整。

## **套用**

按下此鈕可套用所有您所做的新增、刪除、與節點修改，但仍保留視窗開啓。

## **完成**

按下此鈕可關閉視窗，不會套用您所做的新增、刪除、與節點修改。

## **說明**

按下此鈕會開啓線上說明。

## 新增節點

按下此鈕可於選取節點底下新增一內節點 (interior node)。按下此鈕會開啓一空白「節點資訊」螢幕，可讓您輸入節點名稱與共享數量。您輸入的節點名稱與共享為仲裁。

## 新增子節點 (Add Leaf)

按下此鈕可於選取節點底下新增一子節點。按下此鈕會開啓一空白「節點資訊」螢幕，可讓您輸入節點名稱與共享數量。節點名稱必須為現有 Sun Grid Engine，企業版使用者（第 179 頁的「如何以 QMON 配置使用者物件」）或 Sun Grid Engine，企業版專案（第 182 頁的「關於專案」）。

下述規定適用於：

- 所有節點於共享樹內皆有其獨特的路徑。
- 專案於共享樹內不會有超過一筆的參考資料。
- 使用者僅會於專案子樹內出現一次。
- 使用者僅會於專案子樹外出現一次。
- 使用者不會以無子節點的形式出現。
- 所有專案子樹中的子節點會提供已知的使用者或保留的名稱「預設」以供參考。（請見第 202 頁的「關於特殊使用者 default」一節以得知此特殊使用者之詳細資訊。）
- 專案子樹中沒有子專案。
- 沒有在專案子樹中的所有子節點皆會提供一已知使用者或專案以供參考。
- 所有專案子樹中的使用者子節點都可存取至專案中。

## 修改

按下此鈕可編輯選取的節點。按下此鈕會開啓一 Node Info 螢幕，顯示選取節點的名稱與其共享數量。

## 刪除

按下此鈕來刪除選取的節點與其所有相關資訊。

## 複製

按下此鈕來複製選取的節點與其相關資訊至貼上緩衝欄位 (paste buffer)。

## 剪下

按下此鈕從共享樹上剪下選取的節點與其相關資訊。剪下的部份就會複製至貼上緩衝欄位。

## 貼上

按下此鈕就可貼上選取節點中最新複製的節點。

## 尋找

此鈕會開啓一輸入方塊，可讓您輸入搜尋字串，然後再於共享樹中搜尋對應名稱。Node 名稱會顯示於敏感搜尋字串中。

## 尋找下一個

尋找下一個出現的搜尋字串。

## 清除使用

按下此鈕就可將所有共享樹階層的累積量清除為 0。當共享樹策略和預算 (budget) 連結，並於每個預算時期初就從 scratch 開始時，此鈕就特別有用。在設定或修改測試 Sun Grid Engine，企業版環境時，清除使用工具也很便利。

## 大箭頭導覽器 (Large Arrow Navigator)

在箭頭上按一下以開啓視窗中的「共享樹策略參數 (Share Tree Policy Parameters)」。

## 共享樹策略參數

- CPU (%) 滑動軸 — 此滑動軸設定會顯示結合使用 CPU 的百分比。當您變更此軸，MEM 與 I/O 滑動軸也會隨之變更以補償 CPU 換算之改變。
- MEM (%) 滑動軸 — 此滑動軸設定會顯示結合使用記憶體體的百分比。當您變更此軸，CPU 與 I/O 滑動軸也會隨之變更以補償 MEM 百分比之改變。
- I/O (%) 滑動軸 — 此滑動軸設定會顯示結合使用 I/O 的百分比。當您變更此軸，CPU 與 MEM 滑動軸也會隨之變更以補償 I/O 百分比之改變。

---

**注意** – CPU(%)、MEM(%)、與 I/O(%) 三者相加應為 100%

---

- **符號鎖** — 當鎖開啓時，其守護之滑動軸就可自動變更，因為此軸已變更，或因另一滑動軸變更，所以此軸也得隨之變更來補償。

當鎖上鎖時，其守護之滑動軸就無法做變更。如果兩道鎖都上鎖，僅有一個開啓，三個滑動軸都無法做變更。

- **半衰期** — 使用此輸入欄位來指定用法之半衰期。用法會延遲每個排程間隔時間，即在半衰期過後，累積的用法將占有一半的數值。
- **「日 / 小時 (Days/Hours)」選擇清單** — 選擇要以日或小時為主來測量一次半衰期。
- **補償因素** — 此輸入欄位接受一正整數值的補償因素。合理的數值會介於 [2 ... 10]。當使用者 / 專案得到資源時，此補償因素可避免其實際用法過低於其主導資源之目標用法（見以上解釋）。

## 關於特殊使用者 default

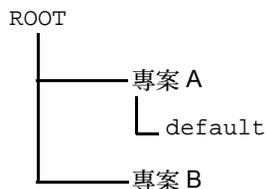
此使用者 default 可用來降低許多使用者站點之共享樹維修量。此使用者僅適用於一名為「混合」之共享樹。在此共享樹上，使用者附屬於 Sun Grid Engine 企業版專案，而且在許多情況下，相同的共享資源會指定至大多位於同一專案之使用者（與共享排程相等）。

此使用者 default 於共享樹之一專案節點上僅可以子節點的形式出現一次（在此，專案節點代表一現有 Sun Grid Engine，企業版專案）。如果此使用者出現，在給予所有現有 Sun Grid Engine，企業版使用者記錄相同共享量時，就可將此使用者視為於對應專案節點底下配置所有現有 Sun Grid Engine，企業版使用者記錄的捷徑。每個存取至專案並提交作業至專案的使用者都會接收到相同的共享應得資源，為對應 default 使用者記錄而配置。要為特定使用者啓動此工具，您必須新增此新使用者至 Sun Grid Engine，企業版系統使用者清單中。

請注意使用者的短期應得資源將會因其預設之資源量不同而有所改變。但是長期應得資源也相同。

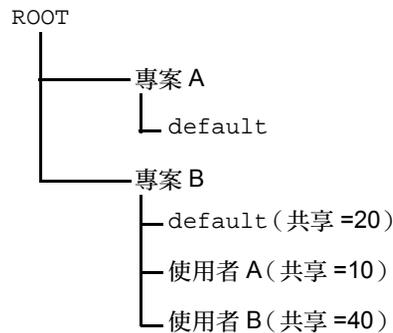
如果您想指定特定（低或高）權利給一些使用者，卻想為其他使用者保留相同的長期應得資源，您就可與 default 使用者旁的個別使用者記錄，為那些特殊資源使用者配置一共享樹。

以下為範例 A。



在範例 A 中，所有提交至專案 A 的使用者會得到長期應得資源，而提交至專案 B 的使用者會累積專案 B 的資源消耗量。無法管理專案 B 使用者的應得資源。

與範例 B 比較。



在範例 B 中，專案 A 的情況仍維持不變。但在專案 B 中，所有提交至專案 B 的使用者都會接收到同等長期應得資源 — 除了使用者 A，僅可接收到大多使用者應得資源的一半，而使用者 B 則接收到兩倍的應得資源。

## ▼ 如何從指令行配置共享策略

---

**注意** – 建議透過 QMON 配置共享樹，因為階層式之共享樹非常適合圖形顯示與編輯。但是，如果需要於 shell script 整合共享樹之修改，您就可以使用 `qconf` 指令與其選項。

---

- 依據下述清單指示使用 `qconf` 指令。
  - `qconf` 選項、`-astree`、`-mstree`、`-dstree` 與 `-sstree` 提供新增共享樹、修改現有共享樹配置、刪除共享樹、與顯示共享樹配置的方法。請參閱《*Sun Grid Engine 5.3 和 Sun Grid Engine，企業版 5.3 參考手冊*》中的 `qconf` 記錄以得知這些選項的詳細資訊。`share_tree` 線上說明手冊包含共享樹配置格式的說明。
  - `qconf` 中的 `-astnode`、`-mstnode`、`-dstnode` 與 `-sstnode` 選項將不會提及整棵共享樹，而僅談及單一節點。此節點為通過所有母節點至共享樹的路徑，與目錄路徑相似。您可藉由這些選項新增、刪除、並顯示一節點。節點內含的資訊包含其名稱與隨附之共享。
  - 用法參數 `CPU`、記憶體與 I/O、半衰期與補償因數的加權會以 `usage_weight_list`、`halftime` 與 `compensation_factor` 包含至排程配置中。排程器配置可藉由 `qconf` 中的 `-msconf` 與 `-ssconf` 選項由指令存取。請參閱《*Sun Grid Engine 5.3 和 Sun Grid Engine，企業版 5.3 參考手冊*》中的 `sched_conf` 記錄以得知格式的詳細資訊。

## 關於功能 (Functional) 策略

功能排程，有時又稱優先排程，為一無回饋計劃，藉由作業提交使用者、專案、部門、與作業等級來確認作業之重要性。由功能策略分配之系統應得資源會與其他應得資源相結合（例如：從截止日期或共享策略分配之應得資源），用以確認作業之最終應得資源。

分配至功能策略的票卷總數量可決定功能排程於四項排程策略中之加權。功能票卷之總數量會於 Sun Grid Engine，企業版安裝期間，由管理員分割至使用者、部門、專案、作業、與作業等級之功能種類。

### 功能共享

功能共享會指定給每個功能類型成員（使用者、部門、專案、作業、與作業等級）。這些共享顯示與類型成員相關之作業應得到多少比例的類型票卷。如果使用者 davidson 有 200 個共享，而使用者 donlee 有 100，則由 davidson 提交之作業得到的使用者功能票卷為 donlee 提交作業的兩倍，無論有多少票卷。

分配至每個類型之功能票卷會共享至特定類型的所有作業。

### share\_functional\_shares 參數

功能策略會為各類型的使用者、專案、部門、作業等級（佇列）、與作業定義應得共享，然後再共享至每個類型底下的所有成員。因此，功能策略與等級二的共享樹很類似，不同點在於作業可同時與多個類型結合在一起。舉例來說，功能策略屬於特定使用者，但同時又屬於專案、部門、與作業等級。

但是在共享樹中，作業由功能類型得到之共享應得資源是由其對應類型成員（例如：專案）與給予類型之共享（專案 vs. 使用者、部門等等）所定義之共享決定。

share\_functional\_shares 參數（在叢集配置中的 schedd\_params 底下）定義如何利用類型成員共享來決定作業之共享。分配至類型成員（例如：特定使用者或專案）的共享可以複製至每個作業，或分配至類型成員下的作業中。

- share\_functional\_shares=false 代表複製。
- share\_functional\_shares=true 定義分配。

那些共享可對照股票，對隸屬於相同類型成員之作業不會造成影響。在這二個情況中，位於相同類型成員之所有作業皆有相同的共享量。但是當與相同類型中的共享量比較時，共享數量就會造成影響。如果 share\_functional\_shares 設定為 true，隸屬於相同類型成員之作業與其相關作業會相對得到小部分的共享。但當 share\_functional\_shares 為 false，且其他相關作業都擁有與其類型成員相同之共享量時，就會不發生上述事情。

如果您想讓一類型成員不斷接收到功能應得資源等級，得知系統內擁有多少個別類型作業之總和，就使用 `share_functional_shares=true`。但如果一個別作業有太多相關作業 (siblings)，其應得資源就會變得非常少。依據作業之類型成員的應得資源，使用 `share_functional_shares=false` 來給予每個作業相同的應得資源等級，不論系統中有多少相關作業。但注意，底下有許多作業的類型成員會主導功能策略。

要注意分享功能共享的設定無法決定分配的總功能票卷數量。功能策略票卷儲存欄位的總數量是由管理員定義。分享功能共享參數只會影響如何於功能策略中分配功能票卷。

## ▼ 如何從 QMON 配置功能共享策略

1. 在 QMON 「票卷概觀」對話方塊下方，按一下「功能策略」。

圖 8-14 為「功能策略」對話方塊範例。

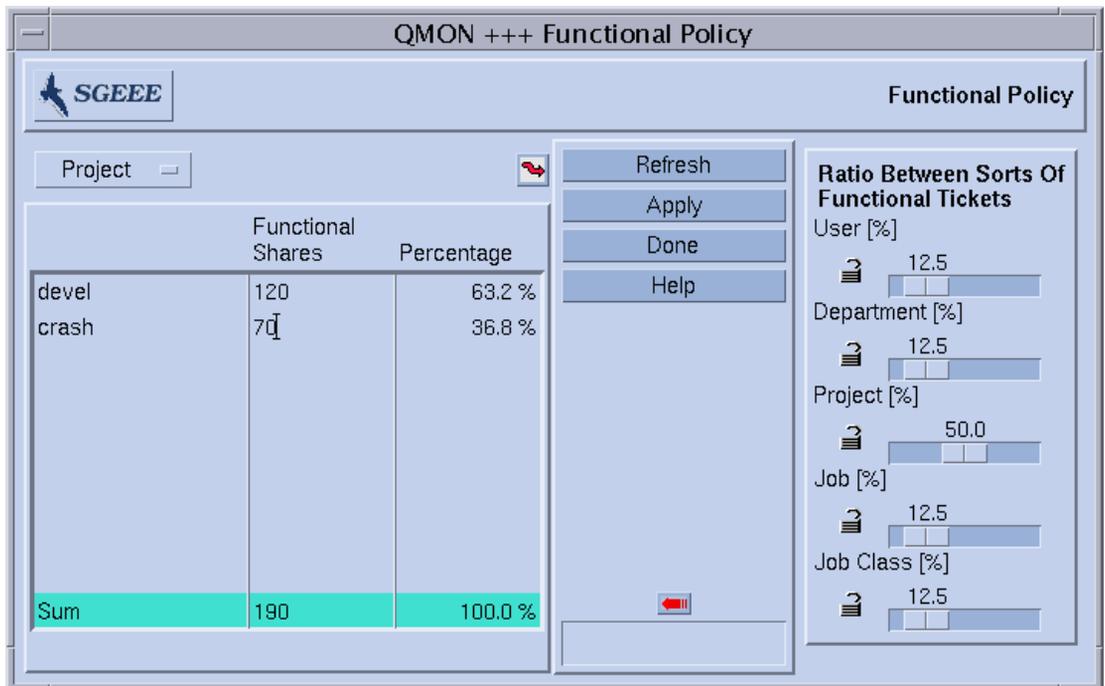


圖 8-14 「功能策略」對話方塊

2. 依下節指導進行。

## 功能選擇功能表 (Functional Selection Menu)

選擇您在定義功能共享之類型：使用者、專案、部門、作業、或作業等級（由佇列定義）。

### 功能顯示

捲軸區提供以下功能。

- 您正在定義功能共享之類型成員清單（使用者、專案、部門、作業、或作業等級）。
- 每個類型成員之功能共享數量。共享可用以便利顯示功能類型之每個成員的相對重要性。本欄位是可編輯的。
- 功能票卷類型（使用者、使用者集等）的功能共享比例換算代表功能共享的數量。此欄位為一回饋裝置，無法編輯。

### 不規則箭頭導覽器

在箭頭上按一下以開啓一配置對話方塊。

- 「使用者配置」對話會開啓以設定「使用者」功能共享。您可使用「使用者」標籤切換至合適的模式，以變更 Sun Grid Engine，企業版使用者配置。
- 「使用者配置」對話會開啓以設定「部門」功能共享。您可使用「使用者集」標籤切換至合適的模式，以變更 Sun Grid Engine，企業版使用者集的部門配置。
- 「專案配置」對話會開啓以設定「專案」功能共享。
- 「作業控制」對話會開啓以設定「作業」功能共享。
- 「佇列控制」對話會開啓以設定「作業等級」功能共享。

### 重整

圖形使用者介面會定期更新顯示的資訊。此按鈕會強制立即顯示重整。

### 套用

按下此鈕可套用所有您所做的新增、刪除、與修改，但仍保留視窗開啓。

### 完成

按下此鈕以關閉視窗，不會套用所做的變更。

## 說明

按下此鈕會開啓線上說明。

## 大箭頭導覽器

在箭頭上按一下以開啓視窗的 Ratio Between Sorts of Functional Tickets。

## 功能票卷分類間的換算

User(%)、Department(%)、Project(%)、Job(%) 與 Job Class (%) 五者相加總和為 100%。

## 使用者 (%) 滑動軸

此滑動軸的設定代表分配至使用者類型的總功能票卷換算。當您變更此軸，其他未上鎖之滑動軸也會隨之變更以補償使用者換算之改變。

## 部門 (%) 滑動軸

此滑動軸的設定代表分配至部門類型的總功能票卷比例。當您變更此軸，其他未上鎖之滑動軸也會隨之變更以補償部門比例之改變。

## 專案 (%) 滑動軸

此滑動軸的設定代表分配至專案類型的總功能票卷比例。當您變更此軸，其他未上鎖之滑動軸也會隨之變更以補償專案比例之改變。

## 作業 (%) 滑動軸

此滑動軸的設定代表分配至作業類型的總功能票卷比例。當您變更此軸，其他未上鎖之滑動軸也會隨之變更以補償作業比例之改變。

## 作業等級 (%) 滑動軸

此滑動軸的設定代表分配至作業等級類型的總功能票卷比例。當您變更此軸，其他未上鎖之滑動軸也會隨之變更以補償作業等級比例之改變。

## 符號鎖

當鎖開啓時，其守護之滑動軸就可自動變更，因為此軸已變更，或因另一滑動軸變更，所以此軸也得隨之變更來補償。

當鎖上鎖時，其守護之滑動軸就無法做變更。

如果四道鎖都上鎖，僅有一個開啓，五個滑動軸都無法做變更。

## ▼ 如何從命令行配置功能共享策略

- 依據下述清單指示使用 `qconf` 指令與其選項。
  - 在使用者類型中，透過 `qconf -muser` 指令來修改 `fshare` 參數（請見《*Sun Grid Engine 5.3 和 Sun Grid Engine，企業版 5.3 參考手冊*》以得知 `user` 檔格式的詳細資訊）。
  - 在部門類型中，透過 `qconf -mu` 指令來修改 `fshare` 參數（請見《*Sun Grid Engine 5.3 和 Sun Grid Engine，企業版 5.3 參考手冊*》以得知用以代表部門之 `access_list` 檔格式的詳細資訊）。
  - 在專案類型中，透過 `qconf -mprj` 指令來修改 `fshare` 參數（請見《*Sun Grid Engine 5.3 和 Sun Grid Engine，企業版 5.3 參考手冊*》以得知 `project` 檔格式的詳細資訊）。
  - 在作業等級類型中，透過 `qconf -mq` 指令來修改 `fshare` 參數（請見《*Sun Grid Engine 5.3 和 Sun Grid Engine，企業版 5.3 參考手冊*》以得知用以代表作業等級之 `queue` 檔格式的詳細資訊）。
  - 不同類型之間的加權會定義於排程器配置 `sched_conf` 中，並可透過 `qconf -msconf` 做變更。變更用的參數為 `weight_user`，`weight_department`，`weight_project`，`weight_job` 與 `weight_jobclass`。參數值介於 0 到 1 之間，並需增加到 1。

---

**注意** – 功能共享僅可透過 `QMON` 指定到作業。沒有任何命令行介面適用於此功能。

---

## 關於截止日期策略

截止日期排程可確保作業可以於特定時間內完成，能立即開始，並給予足夠資源使其可準時完成。提交者會指定作業以下事宜。

- **起始時間** – 作業開始執行的時間。起始時間通常在提交作業之後開始，但也可透過 `QMON` 「作業提交」對話方塊參數 `Start At` 或者 `-a` 選項至 `qsub` 來延遲起始時間。（請見第四章第 63 頁的「提交作業」以得知詳細資訊）

- **初始啓動截止日期** — 作業重要性升到最高、得到所有潛在截止日期票卷、即獲得最多系統資源潛在共享的時間。提交作業的使用者必須決定截止日期初始啓動時間是否可使作業於截止日期完成。

## 截止日期票卷

Sun Grid Engine，企業版會啓動截止日期作業、以較低之重要性、超過初始啓動截止日期、來利用可用系統資源。截止日期作業會於逼近其初始啓動截止日期時，自動接收到更多票卷。從作業有執行資格開始，截止日期票卷便開始成直線分配至截止日期作業，直到初始啓動截止日期到來。如果有超過一個以上的截止日期作業到達其初始啓動截止日期，截止日期票卷就會以其初始啓動截止日期為主，依比例分配至所有作業。

## share\_deadline\_tickets 參數

管理員會分配特定數量的票卷至截止日期策略。票卷的數量可決定分配至每個截止日期作業的票卷量，和提交與截止日期初始啓動時間之間的作業相對位置。在截止日期作業中，share\_deadline\_tickets 參數（在叢集配置中 schedd\_params 底下）為影響截止日期票卷計算之第三因素。

share\_deadline\_tickets=true 的設定代表分配至截止日期策略的總票卷量會跨過所有截止日期作業分派，而每個作業得到的比例就會依其是否接近截止日期初始啓動時間而減少。share\_deadline\_tickets=false 設定代表當每個截止日期作業到達其截止日期初始啓動時間時，就會得到分配至截止期限的所有票卷量；越接近此時間，得到的比例也隨之減少。

如果您想控制截止日期策略分配之總票卷量，特別是僅有固定票卷量可分配之共享與功能策略，就使用 share\_deadline\_tickets=true。請注意：如果系統中同時有太多截止日期作業，分配至個別作業的票卷就會太少而無法達到截止日期。

使用 share\_deadline\_tickets=false 來控制有關其他策略可用票卷儲存欄位之個別截止日期作業的重要性。只要使用此指令，就可忽略系統中有多少截止日期作業。作業總是可以達到截止日期票卷最大量。但是若系統中有太多截止日期作業，其他策略就會失去其重要性。

## 截止日期票卷配置

系統管理員會替所有截止日期作業設定可用截止日期票卷之最大量。此數量代表截止日期排程於四個策略中之加權。依「票卷概觀」螢幕（圖 8-12）指示配置，螢幕上同時也會顯示系統中最近使用中截止日期票卷之數量。

## 截止期限使用者配置

此關乎允許可提交截止期限作業使用者的策略同時也在叢集管理的控制之下。僅有使用者存取清單上之使用者，「截止期限使用者」，才可得到截止期限票卷。圖 8-15 顯示「截止期限作業提交」對話方塊的「初始啟動截止期限」部分。

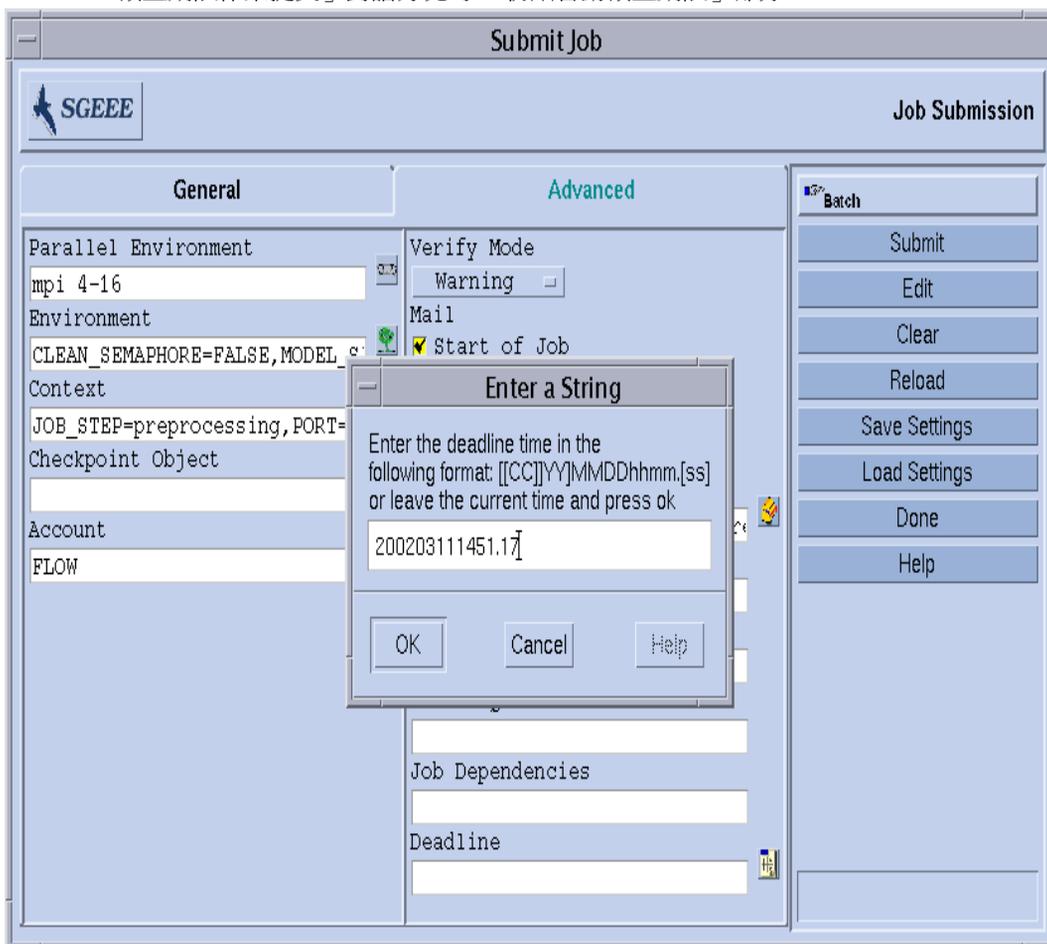


圖 8-15 截止期限作業提交對話方塊

從指令行，您就可透過 `-dl` 選項到 `qsub`，通過初始啟動截止期限到 Sun Grid Engine，企業版系統。請見第四章得知如何提交截止期限的詳細資訊。

## 關於置換策略

置換排程允許 Sun Grid Engine，企業版管理員或操作員透過新增票卷至作業、使用者、部門、專案、或作業等級，來動態調整個別作業，或與使用者、部門、專案、或作業等級相關作業之相對重要性。新增置換票卷會增加票卷總數量，因此會共享使用者、部門、專案、作業等級、或作業擁有的全部資源。

新增置換票卷同時也會增加系統總票卷數量。這些額外票卷會「降低」每張作業票卷的值。

置換票卷主要用於達到二種目的：

- 要暫時置換自動票卷分配策略 — 共享、功能、與截止日期 — 無需改變這些策略之配置。
- 要建立資源應得等級與相關固定票卷量。這適用於高 / 中 / 低作業或優先等級等情況。

直接指定至作業之置換票卷會於作業結束時停止分派，而且所有其他票卷就會「增加」回原來的值。分配給使用者、部門、專案、與作業等級的置換票卷會保留於系統中，直到管理員明確地移除那些票卷。

「票卷概觀」螢幕 (圖 8-12) 會顯示系統最近活動中的置換票卷量。

---

**注意** - 置換記錄會保留於「置換」對話方塊中。當操作員不再需要這些記錄，又未將其明顯刪除，這些記錄就會影響之後的作業。

---

## share\_override\_tickets 參數

管理員會指定票卷給不同置換類型的成員，即不同之使用者、專案、部門、作業等級 (佇列)、或作業。除了「作業」類型外，在特定類型成員底下，分配至個別作業的票卷值會由對應成員定義之票卷量決定。舉例來說，所以給予使用者 A 的票卷量就決定了分配至使用者 A 所有作業之票卷量。

share\_override\_tickets 參數 (在叢集配置中的 schedd\_params 底下) 會控制來自其類型成員票卷值的作業票卷值數量。share\_override\_tickets=true 設定代表類型成員之票卷會平均分配至此成員上的所有作業。share\_override\_tickets=false 代表每項作業會繼承其類型成員定義之票卷數量；即類型成員票卷會複製至底下所有作業。

如果您想控制置換策略分配之總票卷量，特別是僅有固定票卷量可分配之共享與功能策略，就使用 share\_override\_tickets=true。請注意：如果一類型成員底下 (例如隸屬於特定使用者) 有許多作業，而且 share\_override\_tickets 設定為 true，分配至個別作業的票卷量就會變得非常少。

使用 `share_override_tickets=false` 來控制其他策略與置換類型可用票卷儲存欄位之個別作業的重要性。只要使用此指令，就可忽略一類型成員中有多少作業。作業通常都會得到相同的票卷數量，但系統中的置換票卷總數量會增加更多作業，並有權接收系統中的置換票卷。因此，其他策略在此情況下便會失去其重要性。

## ▼ 如何配置置換策略

1. 在「票卷概觀」對話方塊中按一下「置換策略」。  
接著就會出現「置換策略」對話方塊，如圖 8-16 所示。

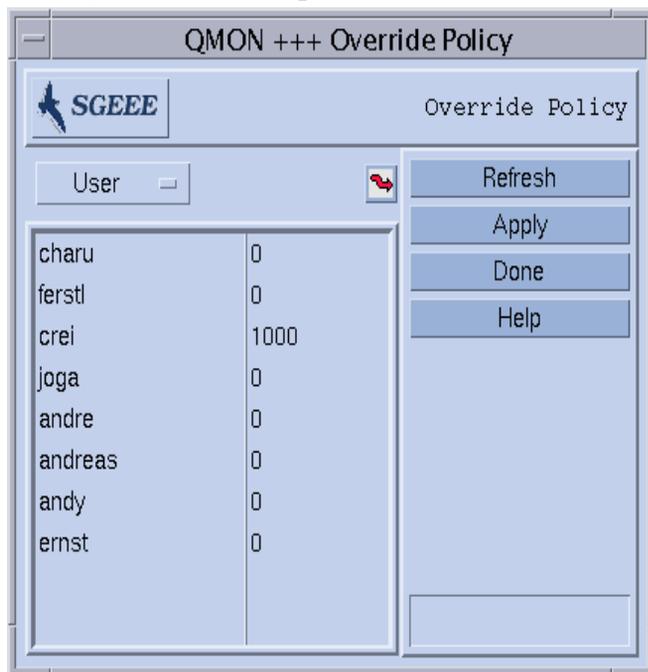


圖 8-16 置換策略對話方塊

2. 依照以下幾節指示分配置換票卷至作業、使用者、部門、專案、或作業等級。

### 置換選擇功能表

選擇您在定義置換票卷之實體類型：使用者、專案、部門、作業、或作業等級。

## 置換顯示

捲軸區提供以下功能。

- 您在定義票卷之類型成員清單（使用者、專案、部門、作業、或作業等級）。
- 每個類型成員的置換票卷整數量。本欄位是可編輯的。

## 不規則箭頭導覽器

在箭頭上按一下以開啓配置對話方塊。

- 「使用者配置」對話方塊會開啓以設定「使用者」置換票卷。您可使用「使用者」標籤切換至合適的模式，以變更 Sun Grid Engine，企業版使用者配置。
- 「使用者配置」對話方塊會開啓以設定「部門」置換票卷。您可使用「使用者集」標籤切換至合適的模式，以變更 Sun Grid Engine，企業版使用者集的部門配置。
- 「專案配置」對話方塊會開啓以設定「專案」置換票卷。
- 「作業控制」對話會開啓以設定「作業」置換票卷。
- 「佇列控制」對話會開啓以設定「作業等級」置換票卷。

## 重整

圖形使用者介面會定期更新顯示的資訊。此按鈕會強制立即顯示重整。

## 套用

按下此鈕可套用所有您所做的新增、刪除、與修改，但仍保留視窗開啓。

## 捨棄

按下此鈕可關閉視窗，不會套用您所做的新增、刪除、與修改。

## 說明

按下此鈕會開啓線上說明。

## ▼ 如何從指令行配置置換策略

- 依下述清單指導進行。

- 在使用者類型中，透過 `qconf -muser` 指令 – 來修改 `oticket` 參數（請見《*Sun Grid Engine 5.3 和 Sun Grid Engine，企業版 5.3 參考手冊*》以得知使用者檔案格式之詳細資訊）。
- 在部門類型中，透過 `qconf -mu` 指令 – 來修改 `oticket` 參數（請見《*Sun Grid Engine 5.3 和 Sun Grid Engine，企業版 5.3 參考手冊*》以得知部門使用之存取清單檔案格式之詳細資訊）。
- 在專案類型中，透過 `qconf -mprj` 指令 – 來修改 `oticket` 參數（請見《*Sun Grid Engine 5.3 和 Sun Grid Engine，企業版 5.3 參考手冊*》以得知專案檔案格式之詳細資訊）。
- 在作業等級類型中，透過 `qconf -mq` 指令 – 來修改 `oticket` 參數（請見《*Sun Grid Engine 5.3 和 Sun Grid Engine，企業版 5.3 參考手冊*》以得知作業等級使用之佇列檔案格式之詳細資訊）。

---

**注意** – 置換票卷僅可透過 `QMON` 指定到作業。目前沒有任何指令行介面適用於此功能。

---

## 關於策略階層 (Policy Hierarchy)

策略階層特別提供解決擱置作業中特定策略衝突的方法。這些情況通常會發生在共享式與功能策略結合時。這兩種策略會共享作業隸屬之特性，相同子等級實體會以先來先處理的順序排列其分配優先順序（共享應得資源）至此兩策略。子等級實體代表位於共享樹中的使用者 / 專案，或者功能策略中，除了「作業」類型之外的任何功能類型（一特定使用者、專案、部門、或佇列）成員。因此，舉例來說，相同使用者的第一項作業會得到最多，第二項次多，第三項再少一點，依此類推。

如果另一策略執行不同的順序，就會產生衝突。因此，舉例來說，置換策略會將第三項作業定義為最重要的一項，而第一項提交的作業應最後執行。

將置換策略置於共享樹或功能策略前的策略階層會確認置換策略中最重要作業，同時也會確認其得到最多共享 / 功能策略中的應得資源 – 只要那些作業隸屬於相同子等級中的共享樹實體（使用者或專案）。

`policy_hierarchy` 參數（位於叢集配置中的 `schedd_params` 底下）可以為四項策略的頭一個字母之結合：S(hare-based)、F(unctional)、D(eadline) 與 O(verride)。如此，您就建立一策略鏈結，第一個字母為最高的策略，最後一個字母則為策略階層最底部。未於策略階層列出之策略並不會影響階層。它們依然為作業票卷之來源。那些票卷不會影響其他策略的票卷計算。但是所有策略之票卷會新增至每項作業，以定義其總應得資源。

以下為二個設定範例，與其如何影響擱置作業順序的說明。

```
policy_hierarchy=OS
```

- 首先，置換票卷會分配合適的票卷數量至每個擱置作業。
- 接著就會使用票卷的數量來影響共享樹的應得資源分配，以免兩項作業皆隸屬於相同使用者或相同子等級專案。然後就會開始為擱置作業計算共享樹票卷。
- 來自置換策略與共享樹策略的票卷會與其他未出現於階層之活動中策略一同增加。擁有最多票卷數量之作業得到最多的應得資源。

```
policy_hierarchy=OS
```

- 接著會計算所有擱置作業之截止日期票卷。
- 置換策略會分配適合的票卷量至每項擱置作業，而來自截止日期與置換策略之票卷則會增加。
- 計算結果之票卷值會影響功能策略的應得資源分配，以免二項作業隸屬於同一功能類型成員。接著以此為基準開始計算擱置作業的功能票卷。計算結果的值會新增至截止日期與置換策略的票卷數量。
- 接著就會使用票卷值來影響共享樹的應得資源分配，以免兩項作業皆隸屬於相同使用者或相同子等級專案。擱置作業之對應共享樹票卷會被計算，然後新增至之前截止日期、置換、與功能策略的總和。
- 擁有最多票卷數量之作業得到最多的應得資源。

同樣地，所有四個字母的組合都適用，但僅有子集是有意義，或有實際關連的。最後一個字母應為 S 或 F，因為只有這兩種策略會被影響，原因請見範例說明其特性。如果 D 與 O 相鄰，就可交換其位置而不改其運作方式。

簡單來說，建議使用以下表格做 `policy_hierarchy` 設定。

```
[O|D][O|D][S|F][S|F]
```

因此，可以影響（截止日期與置換策略）的策略應為前兩個字母的策略，而後二個字母應代表會受影響之策略（共享與功能）。

如 OFD 的設定絕對有效，但卻與 OF 相同。舉例來說，如 OFDS 一般的指令也是有效的，與 ODFS 的結果也有些微差異，但範例要求使用 OFDS，而非 ODFS 設定其實也是可以達到相同目的。

## 關於路徑別名

在 Solaris 與其他聯網之 UNIX 環境中，如果相同的主目錄可跨過 NFS 存取，使用者常於不同機器上使用此相同之主目錄（或部分相同）。但有時候所有機器上的主目錄路徑又非完全相同。

舉例來說，假設使用者主目錄可透過 NFS 與自動裝載器 (*automounter*) 使用。如果使用者在 NFS 伺服器上有一個主目錄，`/home/foo`，他就可以在所有正確安裝之 NFS 客戶站執行自動裝載器中，以此為路徑存取主目錄。但是請注意，客戶端之 `/home/foo` 只是到 `/tmp_mnt/home/foo` 的符號連結，即裝載器實體裝載此目錄至 NFS 伺服器上的真實位址。

在此情況下，如果使用者以 `qsub -cwd` 旗號（於最新工作目錄中執行作業），從主目錄樹中提交作業至客戶端，Sun Grid Engine，企業版系統連結至執行主機上最新工作目錄就會出現問題－如果此主機是 NFS 伺服器的話。這是因為 `qsub` 指令將會到達提交主機上的最新作業目錄，還會將 `/tmp_mnt/home/foo/` 視為提交主機上的實體位址。此路徑將會通過到達執行主機，而且如果執行主機為 NFS 伺服器，又有 `/home/foo` 之實體主目錄路徑，就無法解決此路徑。

導致類似問題之其他情況通常是在不同機器上具不同裝載點路徑的固定（非自動裝載）NFS 裝載（例如於主機上 `/usr/people` 底下裝載主目錄，並於另一主機上 `/usr/users` 底下裝載主目錄），或者由以外面進入一可用網路檔案系統之符號連結。

為避免這類問題發生，Sun Grid Engine，企業版軟體讓管理員與使用者皆可配置路徑更名檔。兩個檔案的位址如下：

- `<sge_root>/<儲存格>/common/sge_aliases`－此為一叢集全域路徑別名檔。
- `$HOME/.sge_aliases`－此為一使用者特定路徑別名檔。

---

**注意**－只有合格的管理員才應修改叢集全域檔。

---

## 檔案格式

兩個檔案共享相同的格式。

- 請忽略第一欄的空白行與有 # 記號的行。
- 每一行－除去空白行與前有 # 之行－必須包含四個字串，之間分別按下數次空白鍵或 `tab` 定位鍵分開。  
第一個字串指定資源路徑、第二個是提交主機、第三個為執行主機、而第四個為資源替代路徑。
- 提交與執行主機記錄僅由符合任一主機之 \* 記號組合。

## 如何解讀路徑更名檔

檔案解讀方式如下。

- 在 `qsub` 取得實體最近作業目錄路徑後，就會讀取叢集全域路徑別名檔－如果有的話。之後便會讀取使用者路徑別名檔，有如它是隨附於全域檔案一般。

- 無需忽略的文字行會由檔案上方讀取，一行接著一行，而且如果需要的話，由那些行指定之翻譯也會隨之儲存起來。
- 當提交主機記錄符合執行 `qsub` 指令之主機，且資源路徑之第一部分為最新作業目錄或已儲存之資源替代路徑時，就會儲存翻譯。
- 只要兩個檔案皆被讀取，儲存之路徑別名資訊就會與提交作業一同傳送出去。
- 在執行主機中會評估別名資訊。如果別名路徑的執行主機記錄符合執行主機，最新作業目錄之第一部分將會由資源替代路徑取代。請注意：最新作業目錄字串在此情況下也會變更，而之後的別名路徑也必須符合套用之替代作業目錄路徑。

## 路徑別名檔之範例

程式範例 8-1 為如何以別名檔案記錄解決 NFS/ 自動裝載器 如上描述之問題的範例。

```
# cluster global path aliases file
# src-path      subm-host  exec-host      dest-path
/tmp_mnt/      *                *                /
```

程式範例 8-1 路徑別名檔案範例

## 關於配置預設要求

批次作業通常是由與特定作業使用者定義的要求簡介相關之 Sun Grid Engine，企業版系統分配至佇列。使用者會收集需要成功執行作業之一連串要求，而 Sun Grid Engine，企業版排程器僅會考慮滿足作業要求之佇列。

如果使用者沒有指定作業之要求，排程器就會考慮使用者沒有進一步限制而可存取之佇列。但即使使用者沒有明確指定，Sun Grid Engine，企業版軟體仍允許配置可定義作業資源要求之預設要求。

預設要求可全域配置至 Sun Grid Engine，企業版叢集之所有使用者，也可隱密配置至所有使用者。預設要求配置會顯示於預設要求檔案中。全域要求檔案位於 `<sg_e_root>/<儲存格>/common/sg_e_request` 底下，而使用者指定要求檔案，又稱 `.sg_e_request`，則位於使用者主目錄或執行 `qsub` 指令之最新作業目錄底下。

如果這些檔案出現，就會受到所有作業評估。評估的順序如下：

1. 全域預設要求檔案
2. 使用者主目錄內之使用者預設要求檔案
3. 最新工作目錄內之使用者預設要求檔案

---

**注意** – 由作業 script 指定，或 qsub 指令行提供之要求享有較高之優先權，等同預設要求檔案之要求（請見第四章得知如何明確要求作業資源）。

---

---

**注意** – 預設要求檔案無意造成之影響可使用 qsub -clear 選項加以限制，捨棄之前的任何規格要求。

---

## 預設要求檔案之格式

下述清單說明本端與全域預設要求檔案之格式。

- 預設要求檔案會包含行列的仲裁編號。請忽略第一欄的空白行與有 # 記號的行。
- 不可忽略的每一行會包含任何 qsub 選項，請見《*Sun Grid Engine 5.3 和 Sun Grid Engine 企業版 5.3 參考手冊*》說明。一行可超過一個選項。批次 script 檔與批次 script 之引數選項並未包含至 qsub 選項中，所以也不允許出現於預設要求檔案中。
- qsub -clear 選項會捨棄先前於最新評估要求檔案或執行之要求檔案內的任何規格要求。

## 預設要求檔案範例

假設使用者之本端預設要求檔案配置如同程式範例 8-2 中之 test.sh script。

```
# Local Default Request File
# exec job on a sun4 queue offering 5h cpu
-l arch=solaris64,s_cpu=5:0:0
# exec job in current working dir
-cwd
```

程式範例 8-2 預設要求簡介範例

要執行此 script，使用者需輸入下述指令。

```
% qsub test.sh
```

執行 test.sh script 之影響會與使用者直接於指令行中指定所有 qsub 選項一般；如下。

```
% qsub -l arch=solaris64,s_cpu=5:0:0 -cwd test.sh
```

---

**注意** – 像透過 `qsub` 提交之批次作業一般，透過 `qsh` 提交之互動作業也將考慮預設要求檔案。透過 `QMON` 提交之互動或批次作業也將顧及這些要求檔案。

---

## 關於收集計算與利用數據

Sun Grid Engine，企業版指令，`qacct`，可用以收集字母與數字構成之計算數據。如果沒有以開關呼叫，`qacct` 就會顯示 Sun Grid Engine，企業版叢集之所有機器上的利用總合計，即所有已於叢集計算檔案，`<sgc_root>/<儲存格>/common/accounting`，中完成之作業的總合計。在此情況中，`qacct` 會以秒回報三次。

- REAL – wallclock 時間；即作業開始到結束的時間。
- USER – CPU 於執行使用者作業時花費的時間。
- SYSTEM – CPU 於系統呼叫花費的時間。

有些開關可用來回報所有或特定佇列、使用者等諸如此類的計算資訊。特別是也有可能要求所有完成作業之資訊，和符合與 `qsub` 指令相同之 `-l` 語法內之資源指定要求作業的資訊，來提交作業。要取得更多的資訊，請參閱《*Sun Grid Engine 5.3 和 Sun Grid Engine，企業版 5.3 參考手冊*》裡的 `qacct` 項目。

`qacct` 選項 `-j [作業id|作業名稱]` – 提供直接存取至儲存於 Sun Grid Engine，企業版系統中的完整資源使用資訊，包含由 `getrusage` 系統呼叫（請參閱《*Sun Grid Engine 5.3 和 Sun Grid Engine，企業版 5.3 參考手冊*》中的相對記錄）提供的資訊。

本選項會分別回報作業 `id` 為 `[作業id]`，或作業名稱為 `[作業名稱]` 之作業資源使用記錄。如果沒有給予任何引數，就會顯示位於參考計算檔案之所有作業。如果有選取一作業 `id`，而且顯示記錄超過一筆，則不是已隱藏作業 `id` 號碼（作業 `id` 的範圍從 1 到 999999），就是顯示已遷移之檢查點作業。

## 關於檢查點支援

**檢查點** 為一可固定執行中作業或應用程式狀態、儲存狀態（即所謂的檢查點）至磁碟、並稍後由檢查點重新執行（如果一作業或應用程式無法完整執行，例：因系統關閉所致）之工具。如果檢查點可至一主機移到另一主機，檢查點就可用來遷移叢集中的應用程式或作業，而不會大量流失計算資源。因此，檢查點工具就需提供動態負載平衡。

Sun Grid Engine，企業版系統支援兩種等級的檢查點。

- 使用者等級檢查點

在此等級中，使用者或應用程式需全權負責提供檢查點產生機制。使用者等級檢查點範例包含以下幾點。

- 在重要的演算步驟中於應用程式裡編碼並定期撰寫重新啓動檔案，與重新啓動應用程式時這些檔案之正確執行程序。
- 檢查點檔案庫之使用，需要連結至應用程式與安裝檢查點機制。

---

**注意** – 各種協力廠商應用程式會以撰寫之重新啓動檔為主，提供一整合之檢查點工具。您可從公共領域（請參閱 Wisconsin 大學的 *Condor* 專案為例）或硬體廠商來使用檢查點檔案庫。

---

- **核心等級透明檢查點 (Kernel-level transparent checkpointing)**

此檢查點等級必須由適用潛在仲裁作業之作業系統（或其升級版）提供。無需提供您應用程式之資源碼變更或重新連結，即可使用核心等級檢查點。

核心等級檢查點適用於完成之作業 – 即由作業建立之執行階層 – 而使用者等級檢查點則通常受到單一程式之限制。因此，在作業開始執行時，就必須合適處理受到此程式包圍之作業。

核心等級檢查點與位於檢查點檔案庫之檢查點都會很耗損資源，因為在執行檢查點時，由作業或應用程式使用之全部虛擬位址空間需要被轉儲至磁碟中。相反地，以重新啓動檔為主之使用者等級檢查點僅會限制寫入重要資料上的檢查點之資訊。

## 檢查點環境

要表現檢查點法不同的種類，與其在不同作業系統架構上可能衍生出的種類，Sun Grid Engine，企業版為所有使用中的檢查點法皆提供一配置屬性說明。

此屬性說明即為 **檢查點環境**。Sun Grid Engine，企業版分佈會提供預設檢查點環境，並可對應站點需求而做修改。

新的檢查點法原則上可以被整合，但整合作業很具有挑戰性，所以應由您 Sun Grid Engine，企業版支援團隊中有經驗之人員來執行。

### ▼ 如何以 QMON 配置檢查點環境

1. 在 QMON 主要功能表上按下「檢查點配置」圖示。

接著就會出現「檢查點配置」對話方塊，如圖 8-17 所示。

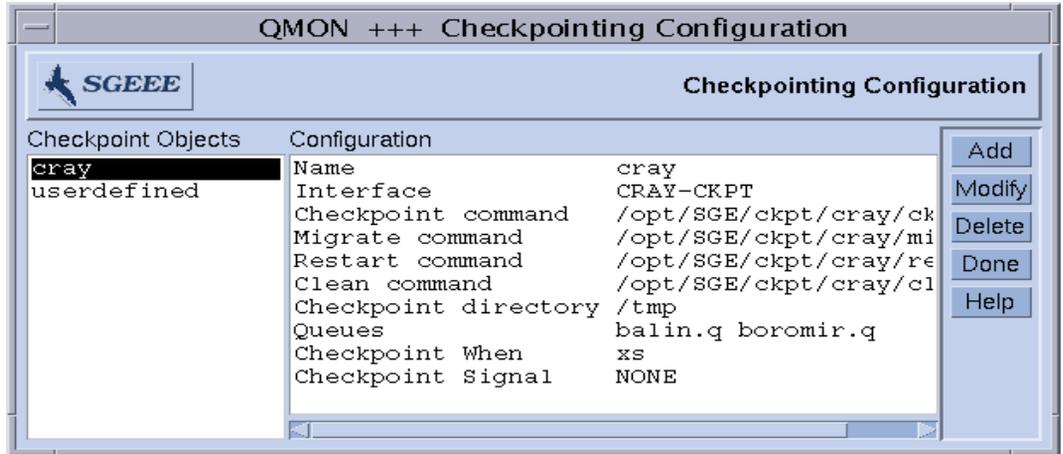


圖 8-17 檢查點配置對話方塊

2. 在檢查點配置對話方塊中，依據您的目的執行以下之一的作業。

### 檢視配置的檢查點環境

- 要檢視之前配置的檢查點環境，於「檢查點物件」欄選擇一列出的檢查點名稱。其對應配置便會顯示於「配置」欄中。

### 刪除配置的檢查點環境

- 要刪除配置的檢查點環境，從「檢查點物件」欄中反白欲刪除之名稱，然後按一下「刪除」。

## 修改配置的檢查點環境

1. 在「檢查點物件」欄中，反白您想修改之已配置的檢查點環境名稱，然後按一下「修改」。出現與圖 8-18 範例相似的「改變檢查點物件」對話方塊，以及目前選取之檢查點環境的配置。

The screenshot shows a dialog box titled "Change Checkpoint Object". It has a light blue background and a white text area. The fields are as follows:

- Name:** A text box containing "cray".
- Interface:** A dropdown menu showing "CRAY-CKPT".
- Queue List:** A list box containing "balin.q" and "boromir.q".
- Checkpoint Command:** A text box containing "/opt/SGE/ckpt/cray/ckpt".
- Migration Command:** A text box containing "/opt/SGE/ckpt/cray/migr".
- Restart Command:** A text box containing "/opt/SGE/ckpt/cray/restart".
- Clean Command:** A text box containing "/opt/SGE/ckpt/cray/clean".
- Checkpointing Directory:** A text box containing "/tmp".
- Checkpoint When:** Three checkboxes: "On Shutdown of Execd" (checked), "On Min CPU Interval" (unchecked), and "On Job Suspend" (checked).
- Checkpoint Signal:** A text box containing "NONE".
- Reschedule Job:** An unchecked checkbox.

On the right side of the dialog, there are two buttons: "Ok" and "Cancel".

圖 8-18 改變檢查點物件對話方塊

2. 遵照下述指示修改選取之檢查點環境。  
「改變檢查點物件」對話方塊可用來變更下述事物。
  - 名稱
  - Checkpoint, migrate, restart, clean 指令字串
  - 檢查點檔儲存之目錄
  - 必須起始檢查點之時刻
  - 當起始檢查點時必須傳送至作業或應用程式之訊號

**注意** – 請參閱 《Sun Grid Engine 5.3 和 Sun Grid Engine，企業版 5.3 參考手冊》中的 checkpoint 記錄以得知這些參數的詳細資訊。而且您還必須定義使用的介面（又稱檢查點法）。於對應選擇清單中選擇一項，然後至 checkpoint 記錄中得知不同介面意義之詳細資訊。

3. **重要** – 在 Sun Grid Engine，企業版分配提供之檢查點環境中，僅變更名稱、檢查點目錄、與佇列清單參數。

要變更「佇列清單」參數，請至「步驟 a」。不然就跳過「步驟 a」，然後至「步驟 4」。

- a. 按下「佇列清單」視窗右方的圖示（請見圖 8-18）。

出現與圖 8-19 範例類似的「選擇佇列」對話方塊。

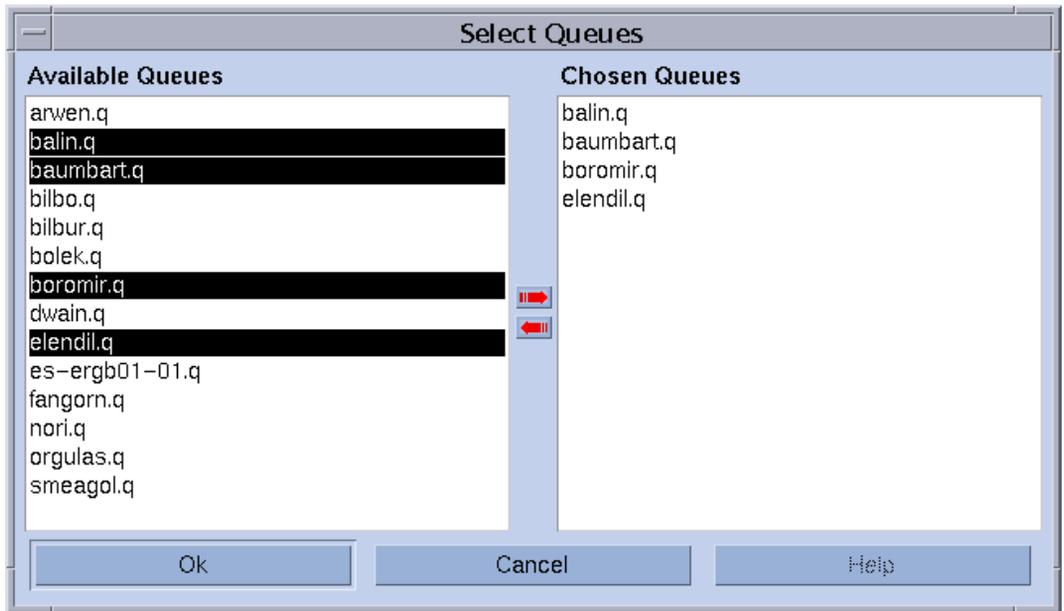


圖 8-19 檢查點佇列選擇對話方塊

- b. 由「可用佇列」清單中選擇您想加入於檢查點環境中的佇列，然後將其新增至「選擇的佇列」清單中。

- c. 按一下「確定」。

按一下「確定」將佇列新增至「改變檢查點物件」對話方塊之「佇列清單」視窗中。

4. 按一下「確定」以 sge\_qmaster 來記錄您的變更，或者按一下「取消」捨棄您的變更。

## 新增檢查點環境

1. 在「檢查點配置」對話方塊中，按一下「新增」，之後就會出現「改變檢查點物件」對話方塊與您可編輯之範本配置，如圖 8-18 範例所示。
2. 將所需資訊填入範本中。
3. 按一下「確定」以 `sge_qmaster` 來記錄您所做的變更，或按一下「取消」來捨棄變更。

## ▼ 如何從命令行配置檢查點環境

- 遵照以下幾節指示輸入 `qconf` 指令與適合之選項。

### `qconf` 檢查點選項

- `qconf -ackpt` *檢查點名稱*

新增檢查點環境 — 此指令會叫出編輯器（預設 `vi` 或對應 `$EDITOR` 環境變數）與檢查點環境配置範本。此參數 *檢查點名稱* 會指定檢查點環境之名稱，並已填妥對應範本欄位之資訊。變更範本並儲存至磁碟片中就完成配置檢查點環境的工作。請參見《*Sun Grid Engine 5.3 和 Sun Grid Engine，企業版 5.3 參考手冊*》內的 `checkpoint` 項目以得知改變範本記錄的詳細資訊。

- `qconf -Ackpt` *檔案名稱*

從檔案新增檢查點環境 — 此指令會從語法上分析檔案 — 必須擁有檢查點環境配置範本格式 — 與新增檢查點環境配置。

- `qconf -dckpt` *檢查點名稱*

刪除檢查點環境 — 此指令會刪除指定之檢查點環境。

- `qconf -mckpt` *檢查點名稱*

修改檢查點環境 — 此指令會叫出編輯器（預設 `vi` 或對應 `$EDITOR` 環境變數）與指定之檢查點環境配置範本。變更範本並儲存至磁碟片中就完成修改檢查點環境的工作。請參見《*Sun Grid Engine 5.3 和 Sun Grid Engine，企業版 5.3 參考手冊*》內的 `checkpoint` 記錄以得知改變範本記錄的詳細資訊。

- `qconf -Mckpt` *檔案名稱*

從檔案修改檢查點環境 — 此指令會從語法分析指定檔案 — 必須擁有檢查點環境配置範本格式 — 並修改現有檢查點環境配置。

- `qconf -sckpt` *檢查點名稱*

顯示檢查點環境 — 本指令會列出指定的檢查點環境配置於標準輸出上。

- `qconf -sckptl`

顯示檢查點環境清單 — 本指令會顯示所有最新配置之檢查點環境名稱清單。

# 聚合體概念

---

本章會說明 Sun Grid Engine，企業版 5.3 聚合體 此重要概念。除了提供聚合體與其相關概念之背景資訊外，本章還包含如何完成下述作業之詳細指示。

- 第 226 頁的「如何新增或修改聚合體配置」
- 第 234 頁的「如何設定可耗損資源」
- 第 244 頁的「如何從指令行修改聚合體配置」
- 第 246 頁的「如何自行撰寫載入感應器」

---

## 關於聚合體

聚合體的定義可透過 `qsub` 或 `qalter -l` 選項，提供所有使用者可能要求的 Sun Grid Engine，企業版作業資源屬性之相關資訊，與 Sun Grid Engine，企業版系統內之相關參數解讀資訊。

聚合體還為 Sun Grid Engine，企業版系統 *可耗損資源* 工具建立架構，此功能允許叢集全域、特定主機、或可辨識資源與其相關容量之佇列相關屬性的定義。資源可用性與 Sun Grid Engine，企業版作業之要求將於排程過程中列入考慮。Sun Grid Engine，企業版同時也將執行必要之簿記與容量規劃，以避免過度消耗可耗損資源。典型具有可耗損屬性的範例有可用之空白記憶體、未被使用之軟體組授權、空白磁碟空間、或連接網路可用之頻寬。

簡易來說，Sun Grid Engine，企業版聚合體可用來說明佇列、主機、或叢集屬性之原有意義。此說明包含了屬性名稱、可參閱之捷徑、屬性之設定值種類（例如 `STRING` 或 `TIME`）、用以指示聚合體屬性之預先定義值、由 Sun Grid Engine，企業版排程器 `sge_schedd` 使用之關係操作員、確認一作業是否可由使用者提出可要求的旗號屬性、並辨別此屬性是否為可耗損屬性之可耗損旗號（若一作業無法明確提出屬性要求，就會考慮建立並預設可耗損屬性之要求值）。

顯示於圖 9-1 之「QMON 聚合體配置」對話方塊描述該如何定義聚合體屬性。

## ▼ 如何新增或修改聚合體配置

1. 在 QMON 主要功能表上按下「聚合體配置」鈕。

圖 9-1 為「聚合體配置」對話方塊範例。

2. 請遵照以下幾節的詳細資訊來新增或修改聚合體配置。

- 第 228 頁的「佇列聚合體」
- 第 228 頁的「主機聚合體」
- 第 230 頁的「全域聚合體」
- 第 231 頁的「使用者定義聚合體」

「聚合體配置」對話方塊提供變更現有聚合體定義與定義新使用者聚合體的方法。

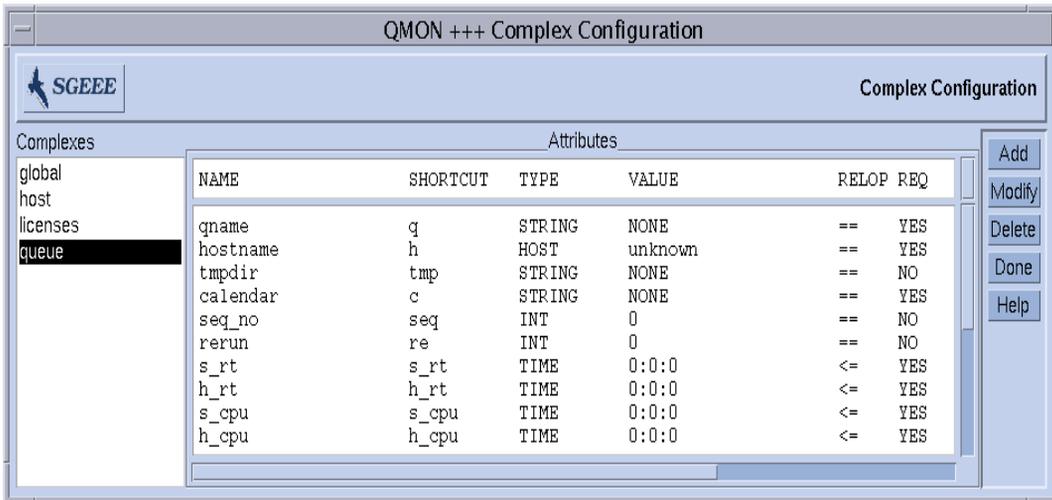


圖 9-1 聚合體配置對話方塊 — 佇列

在螢幕左方會顯示一選擇清單，內含系統可辨識之所有聚合體。可使用此清單來修改或刪除聚合體。您可由螢幕右方之相對按鈕選擇欲進行之操作（新增、修改、或刪除）。如果想建立新的聚合體，或修改現有聚合體，就會開啓如圖 9-2 一般的對話方塊。

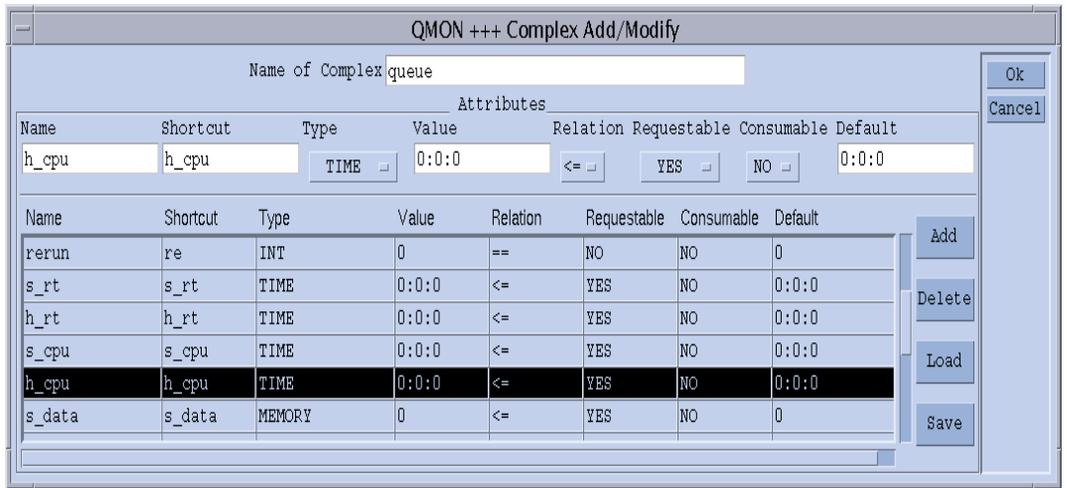


圖 9-2 聚合體新增 / 修改對話方塊

您必須自行輸入聚合體的名稱，或選擇顯示於上方「聚合體名稱」輸入視窗之名稱。只要按下滑鼠左鍵選取一行，您就可以於聚合體定義表上修改聚合體屬性。選取記錄將顯示於定義視窗上，而選取器將顯示於「屬性」方塊上方。變更定義後並按下「新增」鈕就可更新定義表上的改變。

只要填入定義視窗、使用選取器、然後按下「新增」鈕，就可新增一筆新記錄。新增新物件時，請勿選取屬性表上的任一行資料。

「載入」與「儲存」鈕可用來載入與儲存從一般檔案，或至一般檔案的聚合體配置。檔案選擇方塊將會開啓，以供您選取檔案。「刪除」鈕可用來刪除聚合體配置中選取的任一行資料。

請參閱聚合體線上說明手冊以得知圖表中欄與列意義的詳細資訊。最後，螢幕右上方的「確定」鈕將以 `sge_qmaster` 來記錄新增或變更的聚合體。

## 聚合體種類

Sun Grid Engine，企業版聚合體物件分成以下四種不同的聚合體種類。

- 佇列聚合體
- 主機聚合體
- 全域聚合體
- 使用者定義聚合體

以下幾節會詳細說明各種類型。

## 佇列聚合體

佇列聚合體是由 `queue` 這個特殊名稱而來的。

於其預設格式中，可選擇定義於 `queue_conf` 裡的佇列配置參數。佇列聚合體的主要目的為定義如何解讀這些參數，並提供容器更多屬性，以供所有佇列使用。因此，使用者定義屬性可延伸佇列聚合體。

如果可由一特定佇列得知佇列聚合體，則此佇列的對應配置值就會於佇列聚合體中更換屬性值（會覆寫 `value` 欄）。

舉例來說，如果一佇列已設定此佇列聚合體為大，此預設值為 `unknown`（請見圖 9-1）的佇列屬性 `qname` 之數值欄就會設定成大。

不明確之設定值可使用佇列配置中的 `complex_values` 參數覆寫（請見第 151 頁的「關於佇列配置」）。這通常是用於可耗損資源（請見第 234 頁的「可耗損資源」一節）。舉例來說，關於虛擬記憶體大小限制，可使用佇列配置值 `h_vmem` 來限制一作業之總占用記憶體之數量，而 `complex_values` 清單中的對應記錄則會定義主機或指定給佇列之虛擬記憶體的總可用數量。

如果管理者想新增屬性至佇列聚合體，就會使用佇列的 `complex_values` 參數，或者由預設值使用位於佇列聚合體配置中的 `value` 欄來定義與特定佇列相關之數值。

## 主機聚合體

主機聚合體是由 `host` 這個特殊名稱而來的，內含於主機上受管理的所有屬性之特徵定義（請參閱圖 9-3）。與主機相關之屬性的標準組包含兩個種類，但也有可能會增加上節所述之佇列聚合體。第一種類型是由許多佇列配置屬性建立而成，特別適用於以主機為主之管理。這些屬性包含：

- 插槽
- `h_vmem`
- `s_fsize`
- `h_fsize`

（請參閱《*Sun Grid Engine 5.3 和 Sun Grid Engine，企業版 5.3 參考手冊*》內的 `queue_conf` 記錄以得知更多詳細資訊）。

---

**注意** – 同時於主機聚合體與佇列配置中定義這些屬性並不會造成互相矛盾的現象，還可同時允許於主機等級與佇列等級中維持對應資源。舉例來說，總虛擬空白記憶體 (`h_vmem`) 可供主機管理使用，而總數量的一個子集則與此主機上的一個佇列有關。

---

標準主機聚合體第二種屬性類型為預設載入數值。所有 `sge_execd` 都會定期報告載入數值至 `sge_qmaster`。報告的載入數值為標準 Sun Grid Engine，企業版載入數值，例如 CPU 載入平均值，或者是由 Sun Grid Engine，企業版管理（請見第 245 頁的「載入參數」一節）定義之載入數值。標準載入數值之特徵定義為預設主機聚合體的一部份，而管理員定義之載入數值則需要延伸主機聚合體。

主機聚合體不僅可延伸以包含非標準載入參數，還可管理主機相關資源，例如指定至主機之軟體授權數量，或者主機本端檔案系統上的可用磁碟空間。

如果要結合主機聚合體與一主機或此主機之佇列，就必須由以下其中一種方法確認特定主機聚合體屬性的具體值。

- 佇列配置衍生屬性之佇列配置
- 回報之載入數值
- 對應主機配置之 `complex_values` 記錄中，一設定值的明確定義（請參見第 131 頁的「關於配置主機」一節）。

如果以上方法皆不可行（例如此設定值應為載入參數，但 `sge_execd` 並沒有回報其載入數值），就得利用主機聚合體配置的 `value` 欄位。

舉例來說，總空白虛擬記憶體屬性 `h_vmem` 會定義於佇列配置中以為限制，並同時以標準載入參數回報。主機與主機上隨附之佇列的虛擬記憶體總可用容量會定義於此主機與此佇列配置上的 `complex_values` 清單中。與定義 `h_vmem` 為可耗損資源一同（請見第 234 頁的「可耗損資源」），此法可有效利用機器的記憶體，無須擔心因交換而導致耗損記憶體，進而減低系統效能。

---

**注意** - 只有 `Shortcut`、`Value`、`Relation`、`Requestable`、`Consumable` 與 `Default` 欄會因系統預設載入屬性而變更。請勿刪除預設屬性。

---

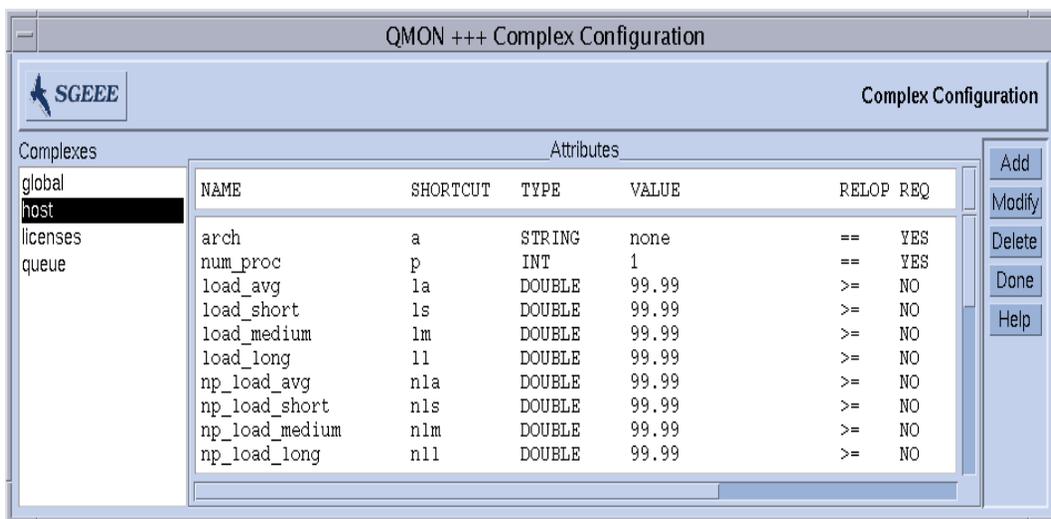


圖 9-3 聚合體配置對話方塊 - 主機

## 全域聚合體

全域聚合體是由 global 這個特殊名稱而來的。

於全域聚合體配置之記錄可用來參照叢集各式資源屬性，例如一檔案伺服器之可用網路頻寬、或者網路各式可用檔案系統之未用磁碟空間（請見圖 9-4）。如果對應載入報告包含 GLOBAL 辨識器，全域資源屬性也可與載入報告做結合（請見第 245 頁的「載入參數」一節）。全域載入數值可由叢集上任何主機報告得知。Sun Grid Engine，企業版與預設值不會報告全域載入數值，所以就沒有預設全域聚合體配置。

全域聚合體屬性的明確數值會由全域載入報告，或者由 global 主機配置（請見第 131 頁的「關於配置主機」一節）之 complex\_values 參數裡的明確定義，抑或結合一特定主機或佇列與一明確定義之對應 complex\_values 清單，加以確認。如果以上方法無效（例如尚未報告載入數值），就得利用全域聚合體配置之的 value 欄位。

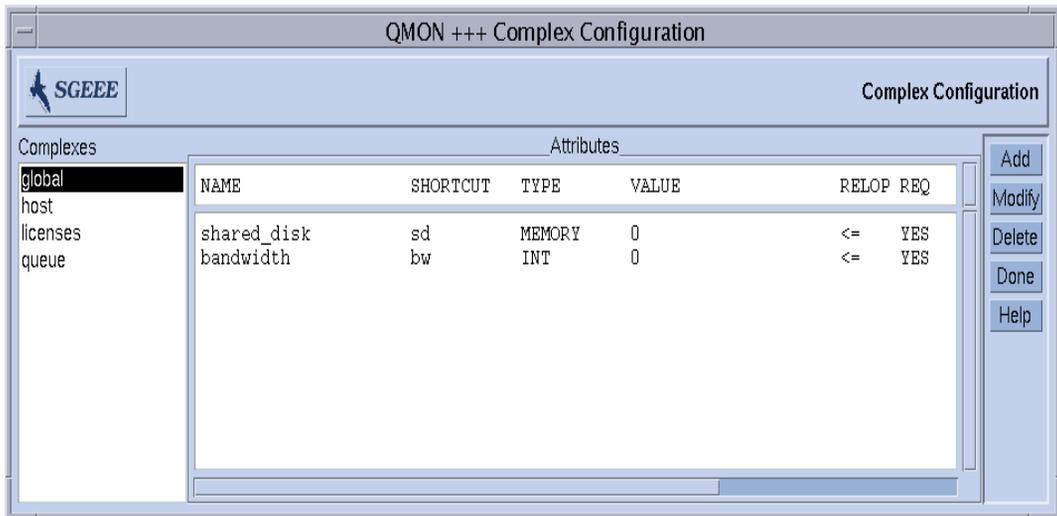


圖 9-4 聚合體配置對話方塊 — 全域

## 使用者定義聚合體

藉由設定使用者定義聚合體，就可延伸 Sun Grid Engine，企業版管理至 Sun Grid Engine，企業版管理之屬性組，進而限制屬性於特定佇列與 / 或主機上的影響力。使用者聚合體只是屬性名稱與其對應定義之總稱，顯示 Sun Grid Engine，企業版如何管理這些屬性。透過 `complex_list` 佇列與主機配置參數（請見第 151 頁的「關於佇列配置」與第 131 頁的「關於配置主機」二節）即可隨附一個或多個使用者定義聚合體至一個佇列與 / 或主機中。所有指定聚合體內定義之屬性可分別用於佇列、主機、與預設聚合體屬性中。

與佇列和主機結合的使用者定義聚合體之明確值必須經由佇列與主機配置中的 `complex_values` 參數，或者利用使用者聚合體配置之 `value` 欄位，才可建立此數值。

舉例來說，定義以下使用者定義聚合體 `licenses`。

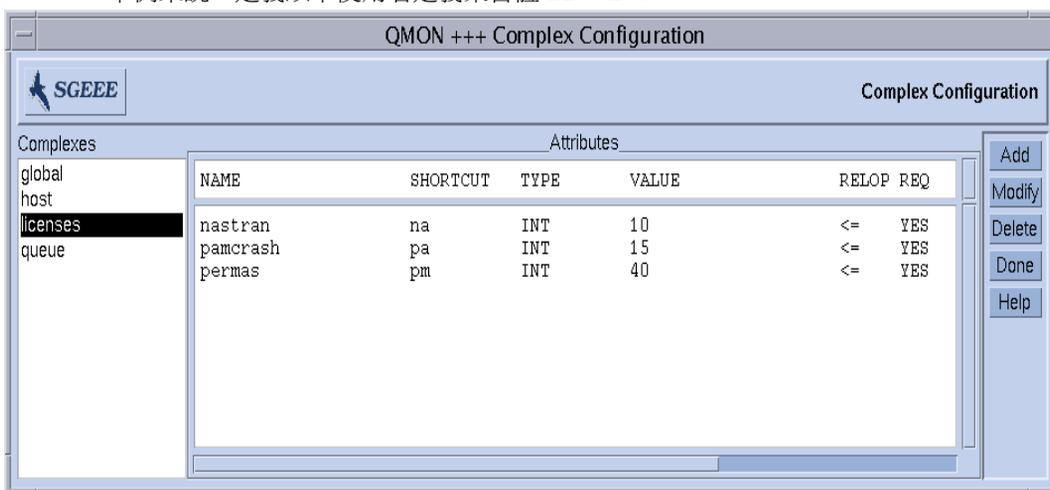


圖 9-5 聚合體配置對話方塊 — 授權

然後讓 `licenses` 聚合體新增至至少一個或多個佇列的相關使用者定義聚合體之清單，如同顯示於圖 9-6 的佇列配置 `User Complexes` 子對話方塊（請見第 151 頁的「關於佇列配置」與其相關節以得知如何配置佇列的詳細資訊）。

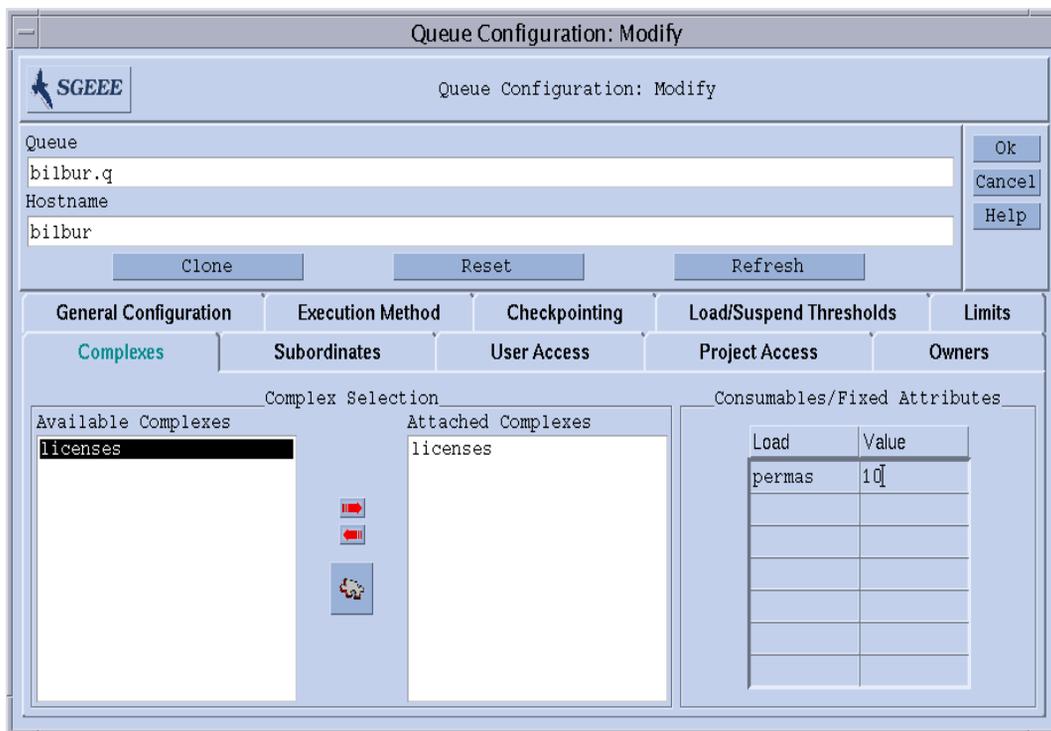


圖 9-6 使用者定義聚合體佇列配置

接著就會配置顯示的佇列，以管理最多 10 組的軟體組 permas 授權。再者，licenses 聚合體屬性 permas 之於 Sun Grid Engine，企業版作業中就變成可要求，就如同圖 9-7 中，「提交 (Submit)」對話方塊中的「要求資源」子對話方塊之「可用資源」清單所示（請參閱第四章第 63 頁的「提交作業」得知如何提交作業的詳細資訊）。

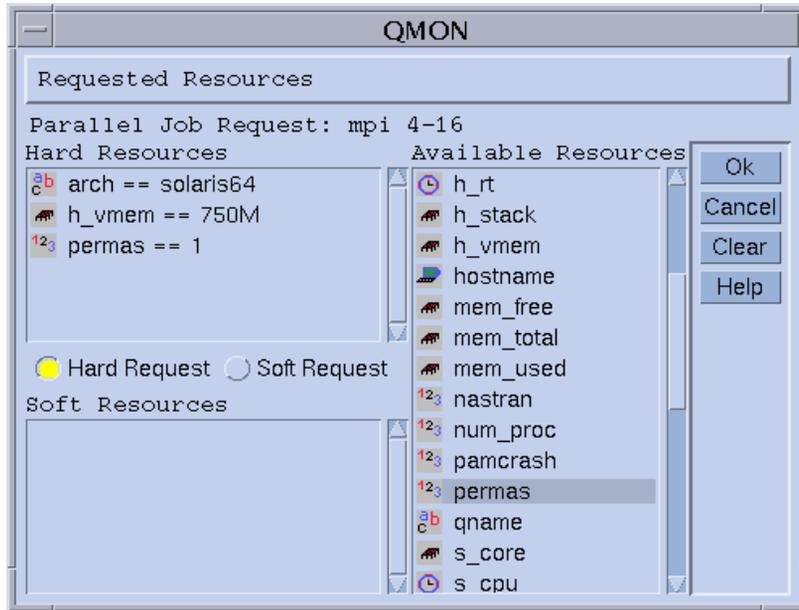


圖 9-7 「提交」子對話方塊之「要求資源」

不然，使用者也可從命令行提交作業，並要求 licenses 屬性，如下所示。

```
% qsub -l pe=1 permas.sh
```

**注意** – 您可使用 pm 捷徑代替完整屬性名稱 permas。

此配置與相似作業要求的結果，唯一適合這些作業的佇列即為與使用者定義 licenses 聚合體結合之佇列，有配置可用之 permas 授權。

### 無效的使用者定義聚合體名稱

以下為保留的聚合體名稱清單，所以不可用於指示使用者定義聚合體名稱。

- global
- host
- queue

## 可耗損資源

可耗損資源，又名為 *可耗損資源*，為一可有效管理限制資源（如可用記憶體、檔案系統上的未用空間、網路頻寬、或使用中的軟體授權）的方法。可耗損資源的總可用容量會由 Sun Grid Engine，企業版管理員定義，而對應資源之消耗也會由 Sun Grid Engine，企業版內部簿記監視。Sun Grid Engine，企業版會替所有執行中的作業計算此資源之消耗量，並確保僅在 Sun Grid Engine，企業版內部簿記顯示有足夠的可耗損資源時才會發送作業。

可耗損資源可與預設或使用者定義載入參數結合（請見第 245 頁的「載入參數」）；即可為可耗損屬性報告載入數值，為載入屬性而建立「可耗損」旗號。Sun Grid Engine，企業版可耗損資源管理會同時將負載（測量資源可用性）與內部簿記列入考慮，並確認兩者並未超出預定限制。

要啟動可耗損資源管理，您必須定義一個資源的總容量。您可由叢集全域、主機與佇列為主完成定義的工作，因為這些種類可以依序互相替換（即主機可以限制叢集資源，佇列可以限制主機與叢集資源）。資源容量的定義會與佇列和主機配置中的 `complex_values` 記錄一同執行（請見《Sun Grid Engine 5.3 和 Sun Grid Engine，企業版 5.3 參考手冊》裡的 `host_conf` 與 `queue_conf` 記錄，和第 151 頁的「關於佇列配置」與第 131 頁的「關於配置主機」）。`global` 主機的 `complex_values` 定義有明確說明叢集全域可耗損資源之設定。位於 `complex_values` 清單中的每個可耗損聚合體屬性都會得到一設定值，以表示資源的最大可用量。內部簿記將經由所有執行中的作業，即經過作業資源要求之作業，減去預計總消耗資源量。

### ▼ 如何設定可耗損資源

僅有數字之聚合體屬性（即屬於 `INT`、`MEMORY` 與 `TIME` 種類）才可配置為可耗損資源。

1. 在 QMON 主要功能表上按下「聚合體配置」鈕。

圖 9-1 為「聚合體配置」對話方塊範例。

2. 要啟動屬性的 Sun Grid Engine，企業版可耗損資源管理，請於聚合體配置中建立 `CONSUMABLE` 旗號，如圖 9-8 描述之 `virtual_free` 記憶體資源一般。
3. 要設定其他可耗損資源，請參照以下幾節的詳細範例指示。
  - 第 236 頁的「範例 1：浮動軟體授權管理」
  - 第 240 頁的「範例 2：虛擬記憶體之空間分享」
  - 第 242 頁的「範例 3：管理可用磁碟空間」

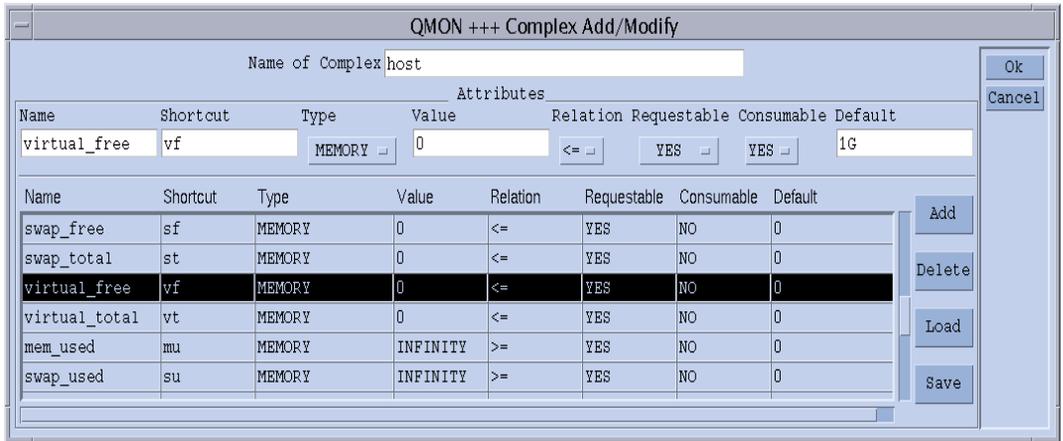


圖 9-8 聚合體配置對話方塊 – virtual\_free

接著，若要 Sun Grid Engine，企業版於每一佇列與主機執行要求的容量規劃，就必須於 complex\_values 清單中定義容量。圖 9-9 顯示如何定義 1 GB 之虛擬記憶體 (virtual memory) 為目前主機之容量值。

(任何佇列裡) 主機上一同執行所有作業之虛擬記憶體需求將會累積，並自 1 GB 容量中減去，以確認可用之虛擬記憶體。如果作業對 virtual\_free 的要求超出可用容量，此作業就不會被分派到主機上的佇列。

---

**注意** – 藉由輸入「可要求」參數之強制值就可以強制作業要求資源，如此才可得知其預計消耗量（請見圖 9-8）。

---



---

**注意** – 預設資源消耗量可藉由作業未明確要求之可耗損屬性的管理員事先完成定義（請見圖 9-8 – 200 MB 為預設消耗量）。這僅在尚未執行要求屬性時才有意義，如上說明。

---

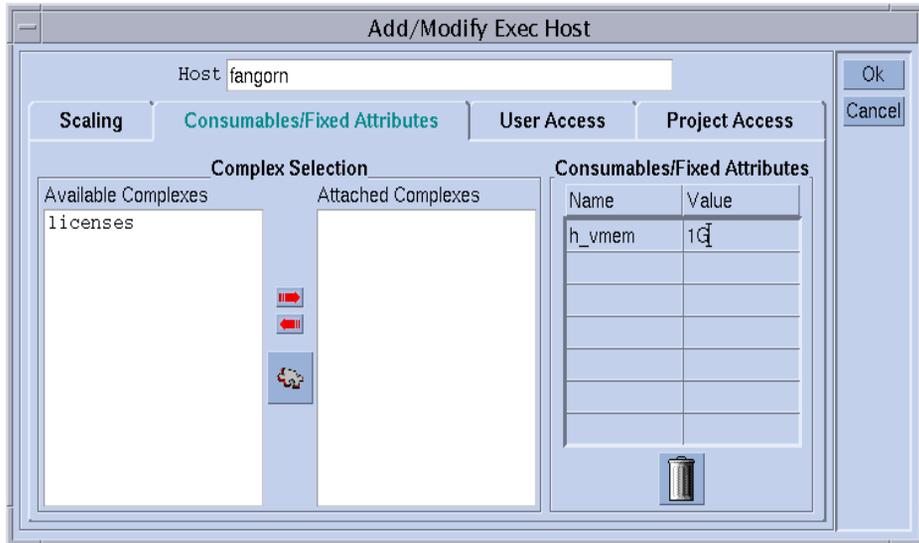


圖 9-9 執行主機配置 – virtual\_free

## 設定可耗損資源範例

利用以下範例引導您如何為站點設定可耗損資源。

### 範例 1：浮動軟體授權管理

假設您於叢集中使用套裝軟體 `pam-crash`，而且您已存取至 10 個；意即只要此軟體總呼叫量未到 10，您就會於每個系統中使用 `pam-crash`，目的是為了在配置 Sun Grid Engine，企業版期間，全部 10 個授權都已被其他執行中的 `pam-crash` 作業佔據時，避免繼續排程 `pam-crash` 作業。

只要有 Sun Grid Engine，企業版可耗損資源，就可輕易達到此目的。首先，您必須新增可用 `pam-crash` 授權的數量以做為全域聚合體配置之可耗損資源，如圖 9-10 所示。

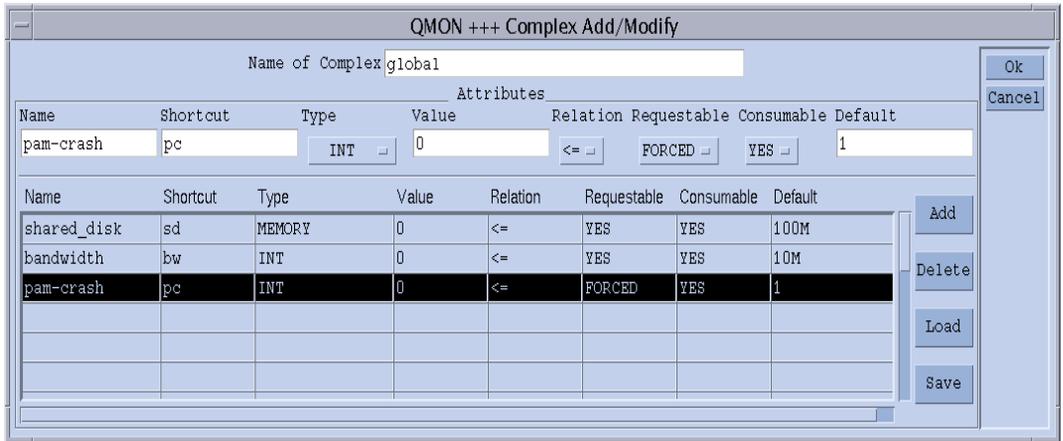


圖 9-10 聚合體配置對話方塊 – pam-crash

可耗損屬性的名稱會設為 `pam-crash`，而 `pc` 可於 `qalter`、`qselect`、`qsh`、`qstat`、或 `qsub -l` 選項中被用來當作捷徑。屬性種類則定義為整數計數器 (integer counter)。數值欄位的設定與可耗損資源無關，因可耗損資源是透過 `complex_values` 清單（如下）由全域、主機、或佇列配置接收其設定值。「可要求」旗號會設定為 `FORCED`，代表使用者於提交作業時，必須要求會占用多少的 `pam-crash` 授權。「可耗損」旗號最後會定義屬性為可耗損資源，與 `Default` 設定無關，因為 `Requestable` 是設定為 `FORCED`，所以將會從任一作業上接收此屬性之要求值。

要執行此屬性與叢集之資源規劃，就必須於全域主機配置中定義可用 `pam-crash` 授權之數量，如圖 9-11 所示。此屬性 `pam-crash` 值會設定為 10，對應 10 個浮動授權。

---

**注意** – `Consumable/Fixed Attributes` 表會對應於主機配置檔案格式 `host_conf` 的記錄。

---

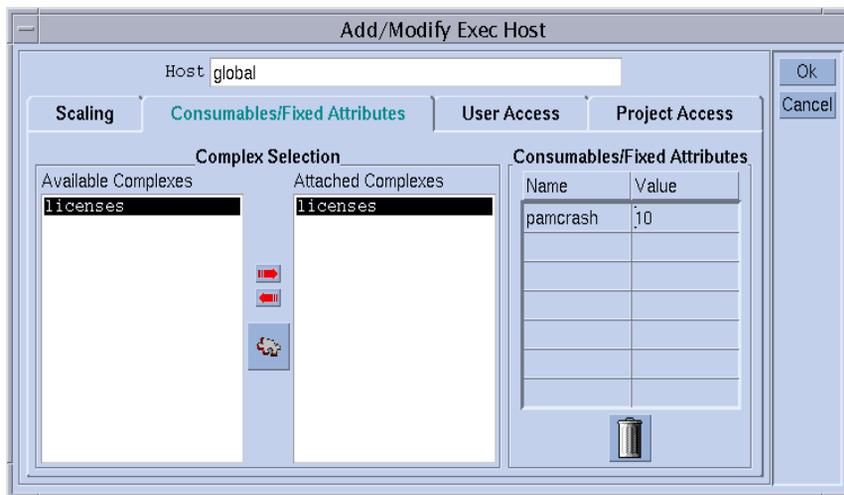


圖 9-11 全域主機配置 – pam-crash

假設使用者提交以下作業。

```
% qsub -l pc=1 pam-crash.sh
```

此作業僅可於 pam-crash 授權佔用量低於 10 時執行。但是此作業可於叢集任一地方執行，而且還會於執行時間內為其佔據 — pam-crash 授權。

如果您叢集中的主機無法包含於浮動授權中 — 例如，因您沒有浮動授權所需的 pam-crash 二進位 — 您就可將與可耗損屬性 pam-crash 相關之主機之容量設為 0，自 pam-crash 授權管理中排除此浮動授權。這必須在「執行主機配置 (Execution Host Configuration)」對話方塊中完成，如圖 9-12 所示。

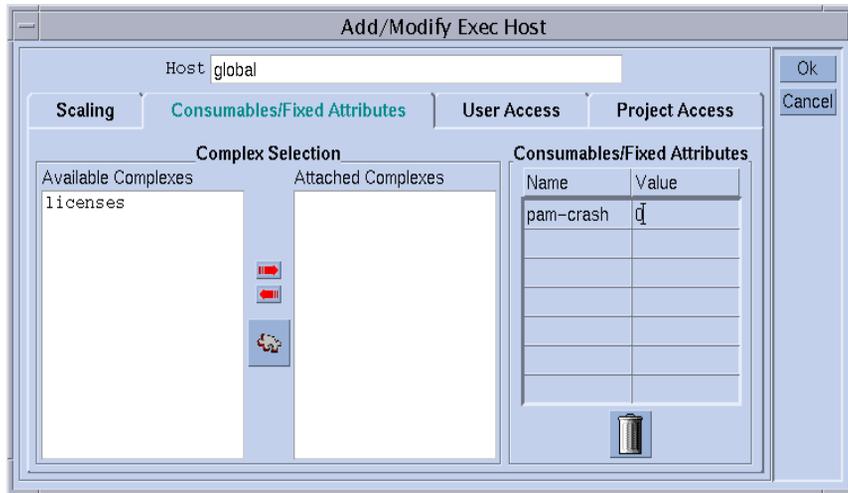


圖 9-12 執行主機配置 – pam-crash

**注意** – pam-crash 屬性可用與否對執行主機而言很不明確，因 global 聚合體屬性會繼承至所有執行主機中。同樣地，只要將容量設定 0，您也可同時限制由特定主機管理之授權數量為叢集所有授權的一部分，並設定為非 0 數值，例如 2。在此情況中，最多 2 個 pam-crash 作業會同時存在於此主機中。

同樣地，您也可以藉由執行 pam-crash 作業以避免特定佇列；例如，因為此佇列很特別，有記憶體與 CPU 時間限制，所以不適用於 pam-crash。在此情況中，您只要於佇列配置中設定相對容量至 0 即可，如圖 9-13 所示。

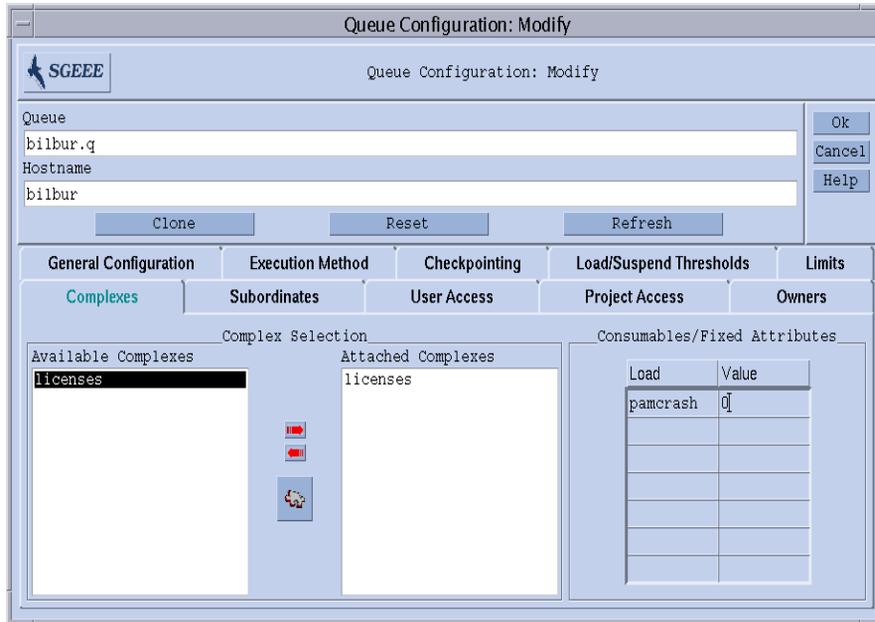


圖 9-13 佇列配置 – pam-crash

---

**注意** – 佇列無法肯定 pam-crash 屬性是否可用，因 global 聚合體屬性會繼承至所有佇列中。

---

## 範例 2：虛擬記憶體之空間分享

系統管理員的一般任務即為調整系統，避免因過度耗損記憶體造成之機器交換而減低效能。Sun Grid Engine，企業版軟體亦可透過「可耗損資源」工具支援此功能。

標準載入程式 `virtual_free` 可回報可用之空白虛擬記憶體，即可用交換空間與可用實體記憶體之結合。為避免交換，可以最小化交換空間的使用。在理想情況之下，執行主機中所有作業所需之全部記憶體應與實體記憶體相符合

Sun Grid Engine，企業版軟體可保證所有作業皆可按如上啟動操作，完成以下預設與配置。

- `virtual_free` 會被配置成可耗損資源，而其位於每個主機上的容量都會設定為可用實體記憶體（或更低）。
- 作業會要求其參與之記憶體使用量，而於執行時間內不可超出要求之設定值。

圖 9-8 顯示一主機聚合體配置範例，而一對應之執行主機配置（主記憶體 1 GB 之主機）則顯示於圖 9-9 中。

---

**注意** - Requestable 旗號會於主機配置範例中設定為 YES，與之前全域聚合體配置範例設定為 FORCED 不同。這代表使用者無需為其作業指定所需記憶體，但若無明確要求記憶體，就必須使用 Default 欄位裡的值。在此情況下預設值為 1 MB，代表無提出要求之作業會被預設將佔據所有可用之實體記憶體。

---

---

**注意** - virtual\_free 為 Sun Grid Engine，企業版之標準載入參數之一。最新記憶體數據之其餘可用性會自動為 Sun Grid Engine，企業版列入虛擬記憶體容量規劃之考量。如果空白虛擬記憶體之載入報告低於 Sun Grid Engine，企業版內部簿記的值，此載入數即便會用於避免記憶體過度耗損。如果作業沒有在 Sun Grid Engine，企業版之下啟動，就很容易造成回報之載入數值與 Sun Grid Engine，企業版內部簿記之間的差別。

---

如果您於單一機器上混合執行不同記憶體需求之作業等級，您也許會想分割此機器之記憶體，以供不同作業等級使用。此功能，一般稱為*空間分享*，可藉由為每一作業等級配置一佇列，並於主機上指定一部分記憶體為其所用而完成。

在此範例中，圖 9-14 顯示之佇列配置會隨附一半可用記憶體至主機 bilbur - 500 MB，至佇列 bilbur.q。因此，執行於 bilbur.q 佇列的所有作業之預設累積記憶體就不會超出 500 MB。在其他佇列之作業就不列入考慮之中，但主機 bilbur 上所有執行作業之總預設記憶體仍不可超過 1 GB。

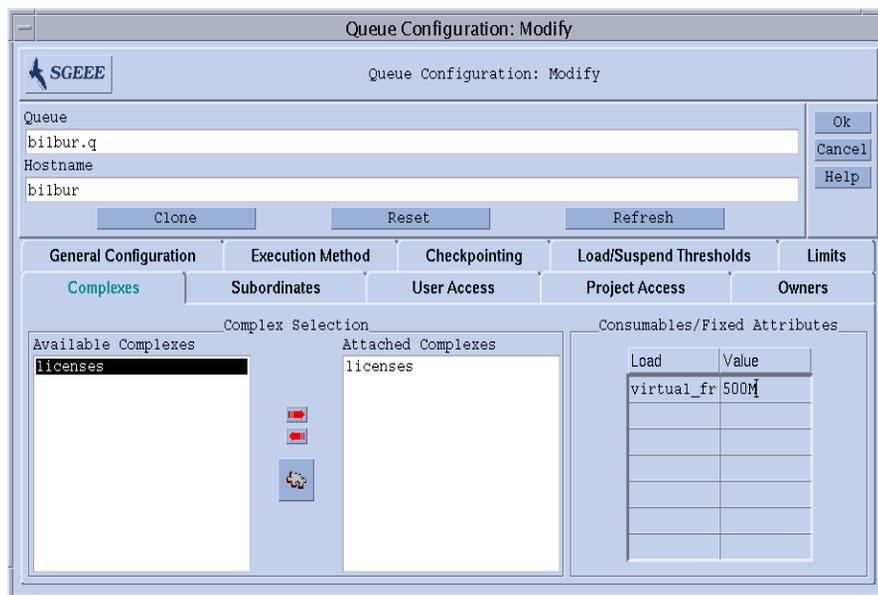


圖 9-14 佇列配置 - virtual\_free



如果要假設主機本端檔案系統，先將磁碟空間容量定義為可耗損，再置於主機配置上是很合理的，如圖 9-16 所示。

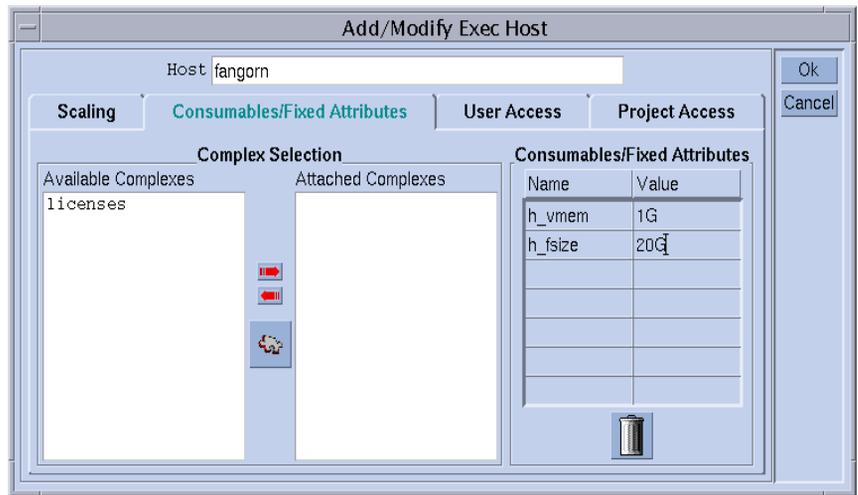


圖 9-16 執行主機配置 – h\_fsize

提交作業至上述配置 Sun Grid Engine，企業版系統之運作方式與前述範例很類似：

```
% qsub -l hf=5G big_sort.sh
```

為何本範例建議 h\_fsize 屬性的原因為 h\_fsize 可同時用於佇列配置中之硬檔案大小限制。檔案大小限制是用以避免作業於提交期間（前述範例為 20 GB）建立比指定檔更大的檔案，或者如果此作業沒有要求屬性，則用以限制由佇列配置傳送之對應值。h\_fsize 之 Requestable 旗號於本範例中已被設立為 FORCED，這樣才會一直顯示要求。

只要利用佇列限制為可耗損資源，就可自動獲得由使用者指定之要求的控制權，與由作業 script 指定之實際資源消耗成對比。任何違反限制事宜都將受到處罰，而此作業也將中斷（請見 queue\_conf 與 setrlimit 線上說明手冊得知詳細資訊）。如此便可確保以 Sun Grid Engine，企業版內部容量規劃為主之資源要求是確實的。

**注意** – 有些作業系統僅提供每個過程檔案大小限制。在此情況之下，作業就須建立多重大小直達限制之檔案。但是在支援每個作業檔案大小限制之系統中，Sun Grid Engine，企業版會使用此功能與 h\_fsize 屬性（請見 queue\_conf 線上說明手冊以得知更多詳細資訊）。

如果您認為未提交至 Sun Grid Engine，企業版的應用程式會同時佔據磁碟空間，Sun Grid Engine，企業版內部簿記就會因缺乏磁碟空間而無法有效預防應用程式發生故障。要預防這個問題，請定期接收磁碟空間使用數據。數據會顯示總磁碟空間消耗量，包含發生在 Sun Grid Engine，企業版外之消耗量。

Sun Grid Engine，企業版載入感應器介面（請見第 246 頁的「新增站點特定載入參數」）可加強設定標準 Sun Grid Engine，企業版載入參數與站點特定資訊，例如一特定檔案系統之可用磁碟空間。

藉由新增適合的載入感應器與回報 `h_fsize` 之空白磁碟空間，您就可結合可耗損資源管理與資源可用性數據。Sun Grid Engine，企業版將會依磁碟空間與源於 Sun Grid Engine，企業版內部資源規劃和最新回報載入數值的可用容量來比較作業需求。僅有符合以上兩項標準之作業才可分派至主機。

## 配置聚合體

Sun Grid Engine，企業版聚合體可以透過圖表，自「QMON 聚合體配置」對話方塊中詳加定義並維護，如本節第 226 頁的「如何新增或修改聚合體配置」與以下所述，又或者可以由指令行執行。

### ▼ 如何從指令行修改聚合體配置

輸入以下指令與合適的選項。

```
% qconf 選項
```

請參閱《Sun Grid Engine 5.3 和 Sun Grid Engine，企業版 5.3 參考手冊》之聚合體記錄，或聚合體線上說明手冊以得知 `qconf` 指令格式與有效值欄位語法之詳細資訊。

以下為有用之選項：

- `-ac`
- `-mc`
- `-Ac`
- `-Mc`

`qconf -Ac` 與 `-Mc` 選項將聚合體配置視為一引數，而 `-ac` 與 `-mc` 選項則使用編輯器於平台聚合體配置編輯，或修改一現有聚合體配置。

選項說明如下：

- `qconf -Ac, -ac`

新增新的聚合體至可用聚合體清單中。

- `qconf -Mc, -mc`

現改一現有聚合體。

## qconf 指令範例

以下指令：

```
% qconf -sc licenses
```

會列出 `nastran` 聚合體（如圖 9-5 裡定義）至檔案格式裡的標準輸出串流，如 `complex (5)` 線上說明手冊定義一般。表 9-1 顯示 `licenses` 聚合體之輸出範例。

| #name   | shortcut | type | value | relop | requestable | consumable | default |
|---|----------|------|-------|-------|-------------|------------|---------|
| #-----  |          |      |       |       |             |            |         |
| nastran   | na       | INT  | 10    | <=    | YES         | NO         | 0       |
| pam-crash   | pc       | INT  | 15    | <=    | YES         | YES        | 1       |
| permas  | pm       | INT  | 40    | <=    | FORCED      | YES        | 1       |
| #---- # start a comment but comments are not saved across edits ----- |          |      |       |       |             |            |         |

表 9-1 `qconf -sc` 輸出範例

---

## 載入參數

本節說明 Sun Grid Engine，企業版 5.3 載入參數之概念，並包含如何自行撰寫載入感應器之指示。

### 預設載入參數

在預設模式中，`sge_execd` 會定期回報一些載入參數與對應值至 `sge_qmaster` 中。這些數值會儲存在 `sge_qmaster` 內部主機物件中（請見第 131 頁的「關於常駐程式和主機」一節）。但是如果已定義聚合體屬性與對應名稱，它們就只能僅供內部使用。這些聚合體屬性包含如何解讀載入數值之定義（請見第 227 頁的「聚合體種類」一節得知詳細資訊）。

在主要安裝後就會回報一組標準載入參數。所有標準載入參數所需之屬性都會定義於主機聚合體中。Sun Grid Engine，企業版之後的版次會提供預設載入參數之延伸設定。因此，預設模式回報之載入參數組會置於 `<sge_root>/doc/load_parameters.asc` 檔中。

---

**注意** – 定義載入屬性的聚合體可決定這些屬性之存取可能性。定義全域聚合體內的載入參數，使其可用於全部叢集與所有主機之中。於主機聚合體定義這些參數，來為所有主機，但不會全域地也於叢集之中，提供屬性。在使用者定義聚合體內定義這些參數後，便可藉由隨附或分離使用者聚合體至主機中以控制載入參數之可見度。

---

---

**注意** – 載入屬性不應定義於佇列聚合體中，因其將不為任何主機或叢集所用。

---

## 新增站點特定載入參數

預設載入參數組也許無法完整描述於叢集內的載入狀況，特別是對於站點特定策略、應用程式、與配置。因此，Sun Grid Engine，企業版軟體提供以仲裁方式延伸載入參數組的方法。為達到此目的，`sge_execd` 提供一介面來回饋載入參數與最近載入數值至 `sge_execd` 中。之後，這些參數就會完全被視為預設載入參數。與預設載入參數一樣（請見第 245 頁的「預設載入參數」一節），對應屬性應定義於載入聚合體中，以使載入參數產生效能。

### ▼ 如何自行撰寫載入感應器

要以額外的載入資訊輸入 `sge_execd`，您必須提供載入感應器。此載入感應器可以為 `script` 或二進位可執行檔案。無論為哪一種情況，其標準輸入與輸出串流，和其控制流之處理皆須符合以下規定：

載入感應器必須撰寫為無限迴路，從 `STDIN` 於特定点等待輸入。如果字串，`quit`，由 `quit` 讀取，則載入感應器就該退出。只要由 `STDIN` 讀到行列末端 (`end-of-line`)，就可以再度啟動載入資料循環。接著，載入感應器須執行所需之操作以計算理想載入數值。在循環結束前，載入感應器會將結果寫入 `stdout` 中。

### 規定

格式如下：

- 載入數值回報起始於一行僅有 `begin` 字樣的文字行。
- 個別載入數值會被新的行分開。
- 每個載入數值資訊都由冒號 (`:`) 分成三部份，中間沒有空格。
- 載入數值資訊第一部分為載入回報之主機名稱，或特定的名稱，`global`。
- 第二部份為載入數值的象徵名稱，如主機或全域聚合體清單中定義一般（請見《*Sun Grid Engine 5.3 和 Sun Grid Engine，企業版 5.3 參考手冊*》裡的 `complex(5)` 記錄得知詳細資訊）。如果回報一載入數值於主機或全域聚合體清單中並無任何記錄，就不會使用此回報的載入數值。

- 第三個部分為測量的載入數值。
- 載入數值回報會以 end 字樣做結束。

## Script 範例

程式範例 9-1 為 Bourne shell script 載入感應器範例。

```
#!/bin/sh
myhost='uname -n'
while [ 1 ]; do
    # wait for input
    read input
    result=$?
    if [ $result != 0 ]; then
        exit 1
    fi
    if [ $input = quit ]; then
        exit 0
    fi
    #send users logged in
    logins='who | cut -f1 -d" " | sort | uniq | wc -l' | sed "s/^ *//"
    echo begin
    echo "$myhost:logins:$logins"
    echo end
done
# we never get here
exit 0
```

### 程式範例 9-1 Bourne Shell Script 載入感應器

如果此範例已存入 load.sh 檔，而且透過 chmod 指定執行權限，您就可以藉由呼叫 load.sh 檔案，並重覆按下鍵盤上的 Return 鍵，從命令行互動測試此範例。

只要能執行此步驟，透過配置載入感應器路徑，於叢集、全域、或執行 hos- 特定配置中輸入 load\_sensor 參數（請見第 145 頁的「基礎叢集配置」一節，或 sge\_conf 線上說明手冊），您就可以安裝載入感應器至任何執行主機中。

對應 QMON 螢幕會如圖 9-17 範例所示。

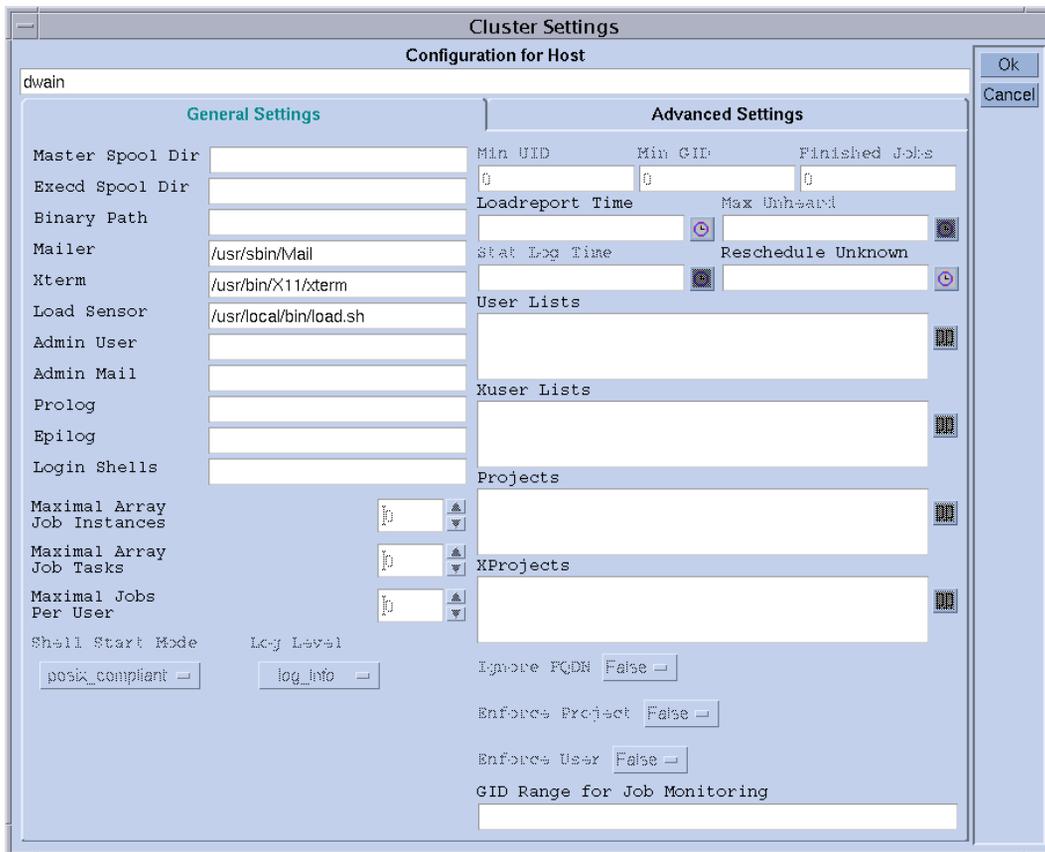


圖 9-17 本端配置與載入感應器

回報之載入參數與登錄都將於新增對應屬性至主機聚合體時立即使用。要求之定義會如圖 9-18 之最後圖表記錄所示，「QMON 聚合體配置」螢幕範例。

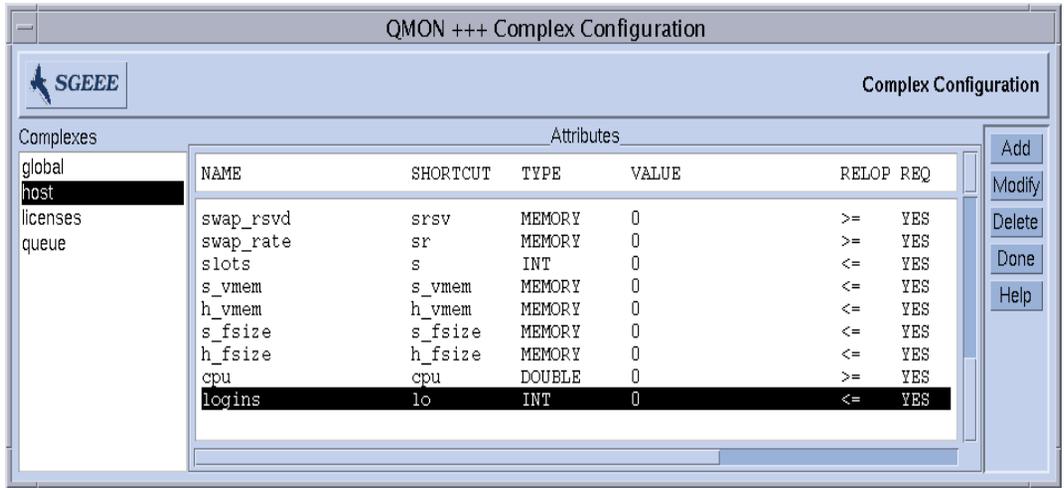


圖 9-18 聚合體配置對話方塊 — 登入



# 管理平行環境

---

本章旨在說明處理與管理平行環境之資訊。

除了這些主題之背景資訊外，本章還包含如何完成下述作業之詳細指示。

- 第 252 頁的「如何以 QMON 配置 PE」
    - 第 252 頁的「顯示 PE 的內容」
    - 第 252 頁的「刪除 PE」
    - 第 253 頁的「修改 PE」
    - 第 253 頁的「新增 PE」
  - 第 256 頁的「如何從指令行配置 PE」
  - 第 257 頁的「如何從指令行顯示配置的 PE 介面」
  - 第 257 頁的「如何以 QMON 顯示配置的 PE 介面」
- 

## 有關平行環境

平行環境 (PE) 為套用以於網路環境或平行平台同時計算之套裝軟體。過去有許多系統已開發許多可在硬體平台上分配並平行執行之可行科技。現今最常見的二個訊息傳遞環境範例即為 PVM (平行虛擬器 (Parallel Virtual Machine), Oak Ridge National Laboratories) 與 MPI (訊息傳遞介面 (Message Passing Interface), Message Passing Interface Forum)。此二項工具皆包含公共領域與硬體廠商提供之執行。

這些系統皆顯示不同的特性，並有其不同之要求。為了可處理執行於類似系統上之仲裁平行作業，Sun Grid Engine，企業版系統提供一有彈性與強大之介面，可符合眾多的需求。

Sun Grid Engine，企業版系統提供如何使用類似 PVM 或 MPI 之仲裁訊息傳遞環境（請見《PVM User's Guide》與《MPI User's Guide》以得知詳細資訊）、或利用單一行列中或跨多重佇列與機器而分配的多重插槽之共享記憶體平行程式，來執行平行作業的方法。可同時配置不同之仲裁 PE 介面。

只要有合適之啟動與結束程序，仲裁 PE 便可由 Sun Grid Engine，企業版來連接介面，詳細情形請見第 259 頁的「終止 PE」的第 258 頁的「PE 啟動程序」一節。

## ▼ 如何以 QMON 配置 PE

1. 在 QMON 主要功能表上按下「PE 配置」鈕。

接著就會出現「平行環境配置」對話方塊，如圖 10-1 所示。

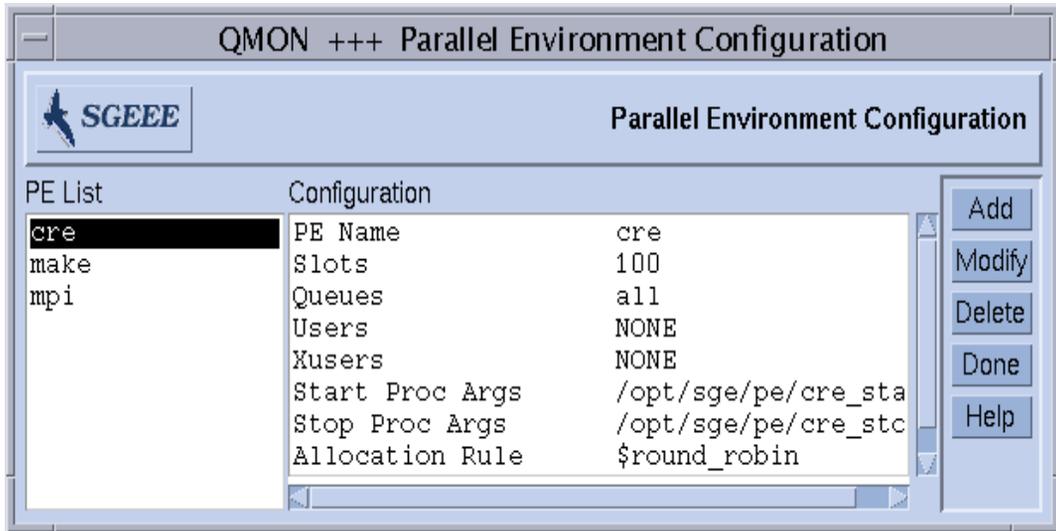


圖 10-1 平行環境配置對話方塊

已配置之 PE 會顯示於螢幕左方的「PE 清單」選擇清單中。

2. 在「平行環境配置」對話方塊中，依據您的目的執行以下之一的作業。

### ▼ 顯示 PE 的內容

- 要顯示 PE 內容，請按一下「PE 清單」選擇清單中的名稱。  
PE 配置的內容就會顯示於「配置」顯示區中。

### ▼ 刪除 PE

- 要刪除選取的 PE，請反白「PE 清單」選擇清單上的名稱，然後按一下「刪除」（位於視窗右方）。

## ▼ 修改 PE

1. 要修改 PE，請按一下「修改」鈕。  
圖 10-2 為「PE 定義」對話方塊範例。
2. 請遵照第 254 頁的「平行環境定義參數說明」一節指示修改 PE 定義。
3. 按一下「確定」以儲存變更，或「刪除」來捨棄變更。  
按一下「確定」或「刪除」來關閉對話方塊。

## ▼ 新增 PE

1. 要新增 PE，請按一下「新增」鈕。  
圖 10-2 為「PE 定義」對話方塊範例。

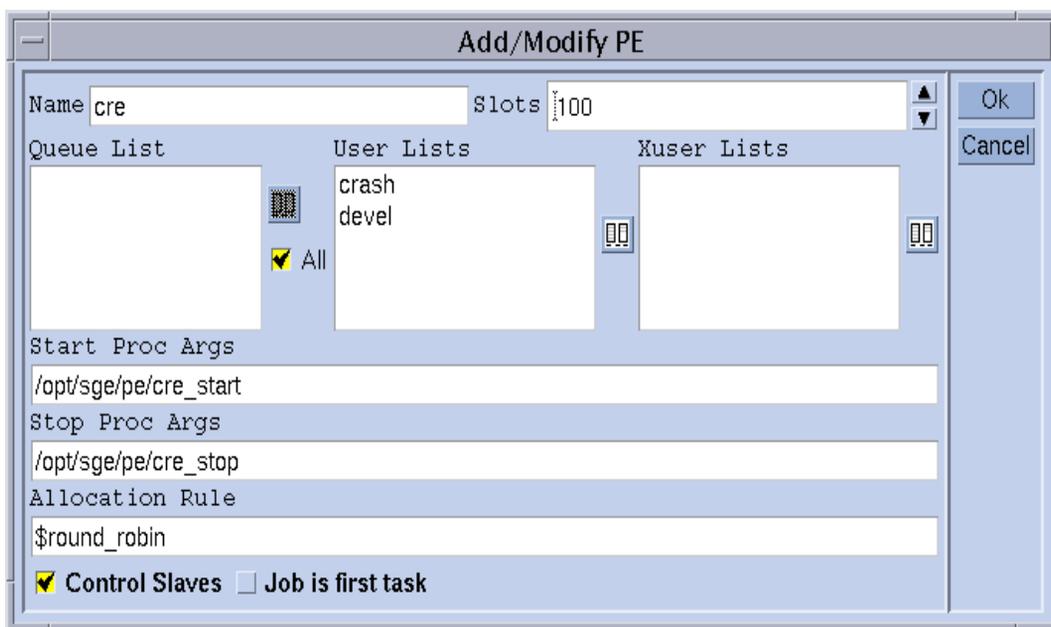


圖 10-2 平行環境定義對話方塊

2. 請遵照第 254 頁的「平行環境定義參數說明」一節指示新增 PE 定義。
3. 按一下「確定」以儲存變更，或「刪除」來捨棄變更。  
按一下「確定」或「刪除」來關閉對話方塊。

## 平行環境定義參數說明

- 名稱輸入視窗會顯示修改作業中選取的 PE 名稱，或者也可自行輸入欲宣告之 PE 名稱。
- 插槽旋轉方塊必須用於輸入所有同時執行中 PE 作業會占據之全部作業插槽的數量。
- 佇列清單顯示區會顯示可供 PE 使用之佇列。按一下「佇列清單」顯示區（「選擇佇列」對話方塊）右方的圖示鈕就可修改 PE 佇列清單，如圖 10-3 範例所示。（不然您也可以使用「所有」勾選方塊來指定由 PE 使用之平行佇列。）

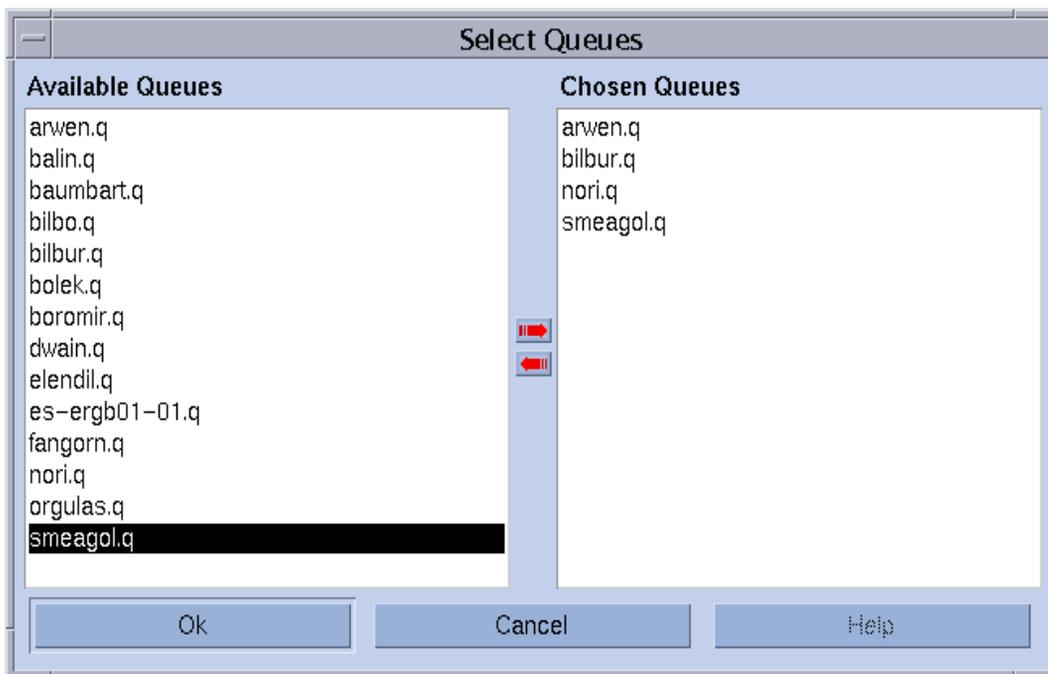


圖 10-3 佇列選擇對話方塊

- 「使用者清單」顯示區會顯示允許存取 PE 之使用者存取清單（請見第 175 頁的「關於使用者存取權限」一節）。
- 「X 使用者清單」顯示區會顯示被拒絕存取 PE 之存取清單。

按一下「選擇存取清單」對話方塊中兩個顯示區旁的圖示，如圖 10-4 所示。您可使用這些對話方塊來修改兩個存取清單顯示區的內容。

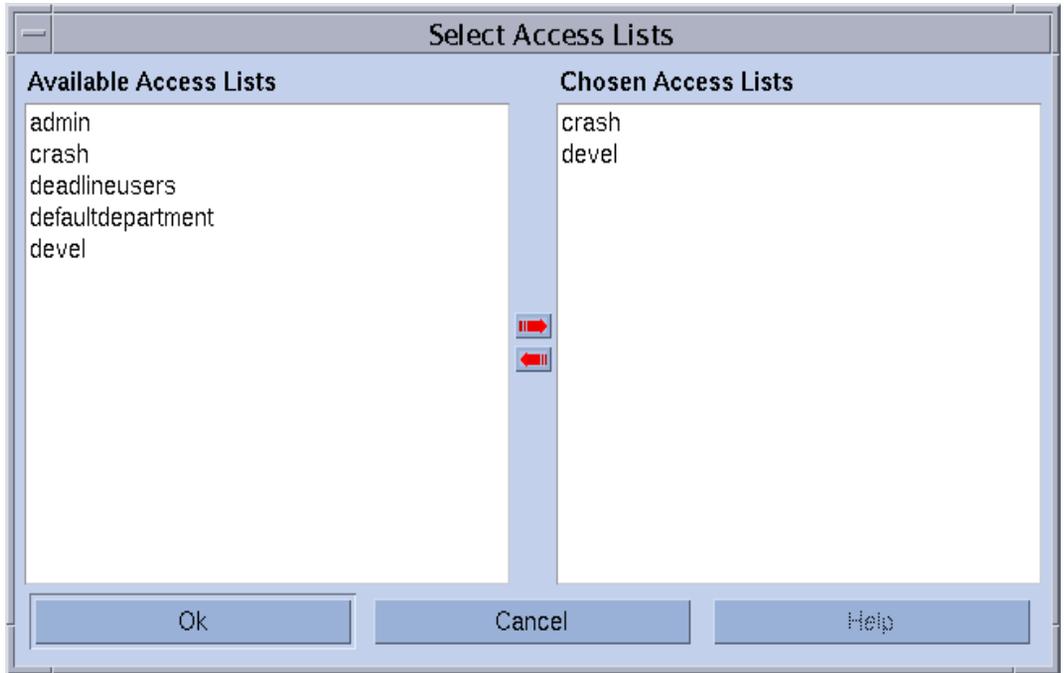


圖 10-4 選擇存取清單對話方塊

- 「*啟動程式引數 (Start Proc Args)*」與「*停止程式引數 (Stop Proc Args)*」輸入視窗可用來輸入 PE 啟動與停止程序之詳細啓用順序（請分別見第 258 頁的「PE 啟動程序」與第 259 頁的「終止 PE」二節）。請注意：這些參數指定為選擇性的。如果一特定平行環境沒有要求此程序，就讓輸入區保持空白。

第一個引數通常為啟動或停止程序本身。其他的參數則為程序的命令行引數。

有許多特別的辨識名稱（以 \$ 開頭），可用來傳遞 Sun Grid Engine，企業版內部執行資訊至程序中。在《*Sun Grid Engine 5.3 和 Sun Grid Engine，企業版 5.3 參考手冊*》中的 `sge_pe` 項目包含所有可用參數之清單。
- 「*配置規則*」輸入視窗可定義於 PE 使用之機器上分配的平行作業數量。每個適合的主機上都應有一正整數來固定作業數量、此特殊名稱 `$pe_slots` 可用來完整執行分配至單一主機 (SMP) 上之作業、而此分母 (denominator) `$fill_up` 與 `$round_robin` 則可用來使每個主機上的作業分配失去平衡。

關於更多分配規則的資訊，請參閱《*Sun Grid Engine 5.3 和 Sun Grid Engine，企業版 5.3 參考手冊*》內的 `sge_pe` 項目。
- 「*控制 Slaves (Control Slaves)*」切換鈕會宣告平行作業是否為透過 Sun Grid Engine，企業版（即透過 `sge_execd` 與 `sge_shepherd`）而產生，或者對應 PE 是否有執行其作業。如果 Sun Grid Engine，企業版系統可全權控制 slave 作業（修正計算與資

源控制) 就會很有幫助, 但此功能僅可用於 Sun Grid Engine, 企業版特殊設定之 PE 介面。請參閱第 260 頁的「PE 與 Sun Grid Engine, 企業版軟體之緊密整合」一節以得知更多詳細資訊。

- 「作業為首要工作 (*Job is first task*)」切換鈕僅在開啓「控制 Slave」時才有效用。這代表作業 script 或其子作業 (child processes) 之一為平行應用程式之平行作業 (舉例來說, 這通常發生在 PVM 之中)。如果關閉「控制 Slave」, 作業 script 就會起始平行應用程式, 但不會參與其中 (例如: 在 MPI 使用 mpirun)。

## ▼ 如何從命令行配置 PE

- 遵照以下幾節指示輸入 qconf 指令與適合之選項。

### qconf PE 選項

- qconf -ap *PE 名稱*

新增平行環境 — 於佇列配置平台叫出編輯器 (預設 vi 或對應至 \$EDITOR 環境變數)。此參數 *pe 名稱* 會指定 PE 名稱, 並已填妥對應平台區之資訊。變更範本並儲存至磁碟片中就完成配置 PE 的工作。請參閱《Sun Grid Engine 5.3 和 Sun Grid Engine, 企業版 5.3 參考手冊》內的 sge\_pe 項目以得知改變範本項目的詳細資訊。

- qconf -Ap *檔案名稱*

從檔案新增平行環境 — 此指令會從語法上分析指定檔案, *檔案名稱* — 必須擁有 PE 環境配置範本格式 — 與新增 PE 配置。

- qconf -dp *pe 名稱*

刪除平行環境 — 此指令會刪除指定 PE。

- qconf -mp *pe 名稱*

修改平行環境 — 此指令會叫出編輯器 (預設 vi 或對應至 \$EDITOR 環境變數) 與指定 PE, 以為配置範本。變更範本並儲存至磁碟片中就完成修改 PE 的工作。請參見《Sun Grid Engine 5.3 和 Sun Grid Engine, 企業版 5.3 參考手冊》內的 sge\_pe 項目以得知改變範本項目的詳細資訊。

- qconf -Mp *檔案名稱*

從檔案修改平行環境 — 此指令會從語法上分析指定檔案, *檔案名稱* — 必須擁有 PE 環境配置範本格式 — 與修改 PE 配置。

- qconf -sp *pe 名稱*

顯示平行環境 — 本指令會列出指定之 PE 配置至標準輸出上。

- qconf -spl

顯示平行環境清單 — 本指令會顯示所有目前配置之平行環境名稱清單。

## ▼ 如何從指令行顯示配置的 PE 介面

- 輸入以下指令。

```
% qconf -spl  
% qconf -sp pe 名稱
```

第一個指令會列出最新可供 PE 介面使用之名稱清單。第二個指令則會顯示特定 PE 介面的配置。請參閱 `sge_pe` 線上說明手冊以得知 PE 配置相關資訊。

## ▼ 如何以 QMON 顯示配置的 PE 介面

- 在 QMON 主要功能表上按下「PE 配置」鈕。

接著就會顯示「平行環境配置」對話方塊（請見第 252 頁的「如何以 QMON 配置 PE」一節）。

第 74 頁的「進階範例」一節的範例已定義要求 PE 介面之平行作業，`mpi`（為訊息傳遞介面所用），至少有四項過程，但最多可至（最好為）十六項過程。「平行環境規格 (PE)」視窗可用來開啓一對話方塊，從可用 PE 清單挑選想要的平行環境（請見圖 10-5）。由作業起始之平行工作的要求數量範圍可新增於「進階提交」螢幕之「PE 規格」視窗的 PE 名稱之後。

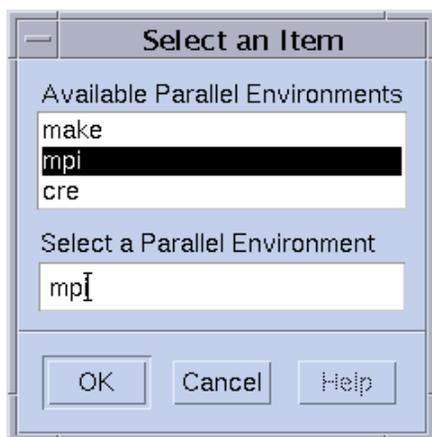


圖 10-5 PE 選擇

請參閱第 85 頁的「如何從指令行提交作業」一節得知對應上述平行作業規格之指令行提交指令，與如何使用 `qsub -pe` 選項來規劃一對等要求。請參閱《*Sun Grid Engine 5.3 和 Sun Grid Engine，企業版 5.3 參考手冊*》中的 `qsub` 項目得知更多 `-pe` 語法的詳細資訊。

於平行作業選擇合適的 PE 介面是相當重要的。PE 介面可以利用無或不同的訊息傳遞系統，於單一或多重主機上分配作業。也許會拒絕特定使用者存取至 PE。僅有特定佇列組可供 PE 介面使用，而 PE 介面也可隨時占據僅有特定佇列插槽數量。因此，您應詢問 Sun Grid Engine，企業版管理以得知適合您平行作業類型使用之最佳 PE 介面。

您可於第 79 頁的「資源需求定義」一節得知如何指定資源需求與 PE 要求。這將會從符合您指定之資源要求的佇列中減少 PE 介面之合格佇列組。舉例來說，假設您已提交下述指令：

```
% qsub -pe mpi 1,2,4,8 -l nastran,arch=osf nastran.par
```

適合此作業之佇列為與 PE 配置之相關 PE 介面 `mpi`，並同時符合 `qsub -l` 選項指定之資源要求規格。

---

**注意** – Sun Grid Engine，企業版 PE 介面工具屬高配置性。特別是，Sun Grid Engine，企業版管理可用來配置 PE 啟動與停止程序（請見 `sge_pe` 線上說明手冊）以支援站點特定需求。輸出環境變數之 `qsub -v` 與 `-v` 選項可用來將資訊從提交作業之使用者，傳遞到 PE 啟動與停止程序。如果您不確定並被要求匯出特定環境變數，請詢問 Sun Grid Engine，企業版管理員。

---

## PE 啟動程序

Sun Grid Engine，企業版系統透過 `exec` 系統呼叫，開始啟動程序並開啓 PE。啟動執行檔與傳遞至此執行檔之名稱可於 Sun Grid Engine，企業版系統配置。請見 Sun Grid Engine，企業版分配樹得知 PVM 環境之啟動程序範例。此樹由 shell script 與由 shell script 呼叫之 C 程式組成。shell script 使用 C 程式來乾淨地 (cleanly) 啟動 PVM。所有其他要求的作業都須由 shell script 處理。

shell script 位於 `<sge_root>/pvm/startpvm.sh` 底下，而 C 程式檔案則位於 `<sge_root>/pvm/src/start_pvm.c` 底下。

---

**注意** – 啟動程序也可適用於單一 C 程式。shell script 可用以輕鬆自訂範例啟動程序。

---

範例 script，`startpvm.sh`，需要下述三個引數。

- 由 Sun Grid Engine，企業版軟體產生之主機檔路徑，包括啟動 PVM 之主機名稱。
- 呼叫 `startpvm.sh` 程序之主機。

- PVM root 目錄的路徑（通常位於 PVM\_ROOT 環境變數中）

請參閱第 252 頁的「如何以 QMON 配置 PE」內的方法得知可傳遞至啟動 script 之參數。這些參數為由 Sun Grid Engine，企業版於執行期間提供至 PE 啟動與結束 script 之參數。舉例來說，要求之主機檔案是由 Sun Grid Engine，企業版建立，而檔案名稱則會藉由此特殊參數名稱，`$sge_hostfile`，傳遞至 PE 配置之的啟動程序中。請參閱《*Sun Grid Engine 5.3 和 Sun Grid Engine，企業版 5.3 參考手冊*》中的 `sge_pe` 項目以得知所有可用參數說明。

主機檔案擁有下述格式。

- 檔案每行代表平行作業欲執行之佇列。
- 每行第一筆項目指定佇列的主機名稱。
- 第二筆項目指定於此佇列執行之平行作業數量。
- 第三筆表示佇列。
- 而第四筆項目則表示於多重處理器上使用之處理器範圍。

此檔案格式是由 Sun Grid Engine，企業版所建立，而且格式已固定。使用不同的檔案格式之 PE（例如 PVM）需要於啟動程序中翻譯此格式（請見 `startpvm.sh` 檔）。

只要 Sun Grid Engine，企業版系統開始啟動 PE 啟動程序，PE 也就隨之啟動。啟動程序應以零退出 (`zero exit`) 狀態退出。如果啟動程序的退出狀態不是零，Sun Grid Engine，企業版軟體就會回報錯誤訊息，且將不會啟動平行作業。

---

**注意** – 先從指令行測試任何啟動程序是有好處的 – 而非 Sun Grid Engine，企業版 – 如果此程序已整合於 Sun Grid Engine，企業版架構中，就可移除所有很難追蹤之錯誤。

---

## 終止 PE

當平行作業已結束或失敗（透過 `qdel`），就會呼叫中止平行環境之程序。本程序之定義與啟動程式中描述之定義相當類似。此停止程序也可於 PE 配置中定義（請見第 252 頁的「如何以 QMON 配置 PE」）。

此停止程序的目的是用來關閉 PE，然後得到所有相關作業。

---

**注意** – 如果停止程序無法清除 PE 作業，Sun Grid Engine，企業版系統也許是沒有執行於 PE 控制底下之作業資訊，因而導致無法清除。Sun Grid Engine，企業版軟體當然會清除作業 script 執行之直接相關的作業。

---

Sun Grid Engine，企業版分配樹同時也包含 PVM PE 停止程序之範例，就位於 `<sge_root>/pvm/stoppvm.sh` 底下。共有下述兩個引數。

- 由 Sun Grid Engine，企業版建立之主機檔路徑

- 啟動停止程序之主機名稱

與啟動程序相似，停止程序之退出狀態需為零才算成功，非零退出狀態則為失敗。

---

**注意** – 先從指令行測試任何停止程序是有好處的 – 而非 Sun Grid Engine，企業版 – 如果此程序已整合於 Sun Grid Engine，企業版架構中，就可移除所有很難追蹤之錯誤。

---

## PE 與 Sun Grid Engine，企業版軟體之緊密整合

於第 252 頁的「如何以 QMON 配置 PE」一節中的「控制 Slave」參數說明有提及由 Sun Grid Engine，企業版元件 `sge_execd` 與 `sge_shepherd` 執行之平行作業創作的 PE 會有益於執行其作業創作之 PE。這是因為 UNIX 作業系統僅允許執行階層之建立者做可靠資源控制，而平行應用程式的功能，如修正計算、資源限制與過程控制只能由所有平行作業之建立者執行。

大多數 PE 不會執行這些功能，因此也就不會提供資源管理系統，如 Sun Grid Engine，企業版，整合所需之有效介面。要克服這個問題，Sun Grid Engine，企業版系統有提供緊密整合 PE 之進階 PE 介面，可將作業建立之責任由 PE 傳遞至 Sun Grid Engine，企業版軟體。

Sun Grid Engine，企業版分配包含二個緊密整合範例，一為 PVM 公共領域版，一為由 Argonne National Laboratories 執行之 MPICH MPI。這兩個範例分別位於 `<sge_root>/pvm` 與 `<sge_root>/mpi` 兩個目錄中。這兩個目錄同時也包含了 README 檔，說明如何使用與最新限制。請參閱 README 檔得知更多詳細資訊。

除此之外，為了有所比較，`<sge 根目錄>/mpi/sunhpc/loose-integration` 目錄包含了鬆散的 Sun HPC ClusterTools™ 軟體的整合範例，而 `<sge_root>/mpi` 目錄則包含了鬆散的介面整合變數以供比較。

---

**注意** – 執行 PE 緊密整合為一進階作業，所以需要擁有 PE 專業知識與 Sun Grid Engine，企業版 PE 介面。您可聯絡您的 Sun 支援代表商以獲得協助。

---

# 錯誤訊息與疑難排解

本章會說明 Sun Grid Engine，企業版 5.3 錯誤訊息程序，並提供解決各式常見問題之解答。

## Sun Grid Engine，企業版 5.3 軟體如何得到錯誤回報

Sun Grid Engine，企業版軟體會藉由特定檔案之登入訊息，與 / 或電子郵件 (e-mail) 回報錯誤訊息或警告。使用之日誌檔案為：

- 訊息檔案

`sgc_qmaster`、`sgc_schedd`、與 `sgc_execd` 有許多分隔之訊息檔案。這檔案都有相同的檔名，`messages`。`sgc_qmaster` 日誌檔案位於主多工緩衝處理目錄中，而排程器多工緩衝處理目錄內的 `sgc_schedd` 訊息檔案與執行常駐程式的日誌檔案都位於執行常駐程式之多工緩衝處理目錄中（請見第 19 頁的「Root 目錄下的多工緩衝處理目錄」一節得知更多多工緩衝處理目錄資訊）。

訊息檔案格式如下：

- 每個訊息皆佔據一行。
- 訊息是由一垂直線記號 (|) 分隔成五個組件。
- 第一部分組件為訊息的時間戳記。
- 第二部分為 Sun Grid Engine，企業版建立訊息之常駐程式。
- 第三部分為常駐程式執行之主機名稱。
- 第四部分則為訊息種類，分為通知用之 `N`、消息用之 `I`（上述兩者皆僅為提供資訊用）、警告用之 `W`、錯誤用之 `E`（已偵測到錯誤情況）、與危急用之 `C`（可能會導致程式失敗）。

您可以使用叢集配置內的 `loglevel` 參數於全域或本端上偏移 (to bias on a global or local basis) 您想記錄之訊息類型。

- 第五個則為訊息文字。

---

**注意** – 若基於某些原因導致錯誤日誌檔案無法存取，Sun Grid Engine，企業版將會試著把錯誤訊息記錄至對應主機的 /tmp/sge\_qmaster\_messages、/tmp/sge\_schedd\_messages 或 /tmp/sge\_execd\_messages 檔中。

---

- 作業 STDERR 輸出：

只要啟動作業，就會將作業 script 之標準錯誤 (STDERR) 輸出重新導回至檔案。檔案名稱與位址不是符合預設，便是由特定 qsub 指令行開關指定。請參閱《Sun Grid Engine 使用者指南》與《Sun Grid Engine 5.3 和 Sun Grid Engine，企業版 5.3 參考手冊》得知詳細資訊。

在某些情況下，Sun Grid Engine，企業版會透過電子郵件通知使用者與 / 或管理員發生錯誤事件。由 Sun Grid Engine，企業版寄送之郵件訊息並不包含訊息本文。訊息文字全部包含在郵件主旨欄位中。

## 不同錯誤或退出代碼之結果

表 11-1 列出不同作業相關之錯誤與退出代碼的結果。這些代碼適用於所有 Sun Grid Engine，企業版作業類型。

表 11-1 作業相關錯誤或退出代碼

| Script/ 方法    | 離開或錯誤代碼 | 結果               |
|---------------|---------|------------------|
| 作業 script     | 0       | 成功               |
|               | 99      | 重新排於佇列中          |
|               | 其餘      | 成功：於計算檔案之退出代碼    |
| prolog/epilog | 0       | 成功               |
|               | 99      | 重新排於佇列中          |
|               | 其餘      | 佇列錯誤狀態：作業重新排於佇列中 |

表 11-2 列出與平行環境 (PE) 配置相關之作業錯誤或退出代碼的結果。

表 11-2 PE 相關之錯誤或退出代碼

| Script/ 方法 | 離開或錯誤代碼 | 結果                   |
|------------|---------|----------------------|
| pe_start   | 0       | 成功                   |
|            | 其餘      | 佇列設定至錯誤狀態，作業重新排於佇列中  |
| pe_stop    | 0       | 成功                   |
|            | 其餘      | 佇列設定至錯誤狀態，作業無重新排於佇列中 |

表 11-3 列出與佇列配置相關之作業錯誤或退出代碼的結果。僅當覆寫對應方法時上述結果才會成立。

表 11-3 佇列相關錯誤或退出代碼

| Script/ 方法 | 離開或錯誤代碼 | 結果         |
|------------|---------|------------|
| 作業啟動器      | 0       | 成功         |
|            | 其餘      | 成功，無其它特殊意義 |
| 暫停         | 0       | 成功         |
|            | 其餘      | 成功，無其它特殊意義 |
| 恢復         | 0       | 成功         |
|            | 其餘      | 成功，無其它特殊意義 |
| 終止         | 0       | 成功         |
|            | 其餘      | 成功，無其它特殊意義 |

表 11-4 列出與檢查點相關之作業錯誤或退出代碼的結果。

表 11-4 檢查點相關之錯誤或退出代碼

| Script/ 方法 | 離開或錯誤代碼 | 結果                                    |
|------------|---------|---------------------------------------|
| 檢查點        | 0       | 成功                                    |
|            | 其餘      | 成功 — 但對核心檢查點有特殊意義：檢查點不成功；沒有發生。        |
| 遷移         | 0       | 成功                                    |
|            | 其餘      | 成功 — 但對核心檢查點有特殊意義：檢查點不成功；沒有發生。將會進行遷移。 |
| 重新啓動       | 0       | 成功                                    |
|            | 其餘      | 成功，無其它特殊意義                            |
| 清除         | 0       | 成功                                    |
|            | 其餘      | 成功，無其它特殊意義                            |

## 於除錯模式中執行 Sun Grid Engine，企業版程式

而針對一些嚴重的錯誤事件，錯誤記錄機制也許不會有足夠的資訊來辨識這些問題。因此，Sun Grid Engine，企業版提供可執行於除錯模式中幾乎所有輔助程式與常駐程式的功能。於提供之資訊範圍與深度中有不同的除錯等級。除錯等級範圍由 0 到 10，等級 10 傳達的訊息是最詳細的，而等級 0 則會關閉除錯。

於您的 `.cshrc` 或 `.profile` 資源檔案中設定一除錯等級並提供 Sun Grid Engine，企業版分配。針對 `csh` 或 `tcsh` 的使用者，就會包含此檔 `<sge_root>/util/dl.csh`。針對 `sh` 或 `ksh` 使用者，對應檔案就會命名為 `<sge_root>/util/dl.sh`。檔案需要 "sourced" 進您的標準資源檔。而 `csh` 或 `tcsh` 的使用者，請包含下述行：

```
source <sge_root>/util/dl.csh
```

進您的 `.cshrc` 檔中。`sh` 或 `ksh` 使用者，新增下述行：

```
. <sge_root>/util/dl.sh
```

進您的 `.profile` 檔是相同的。只要您現在登出再登入，您就可以使用下述指令來設定一除錯等級等級：

```
% dl level
```

如果等級大於 0，隨後啓動 Sun Grid Engine，企業版指令將迫使指令寫入追蹤輸出至 `STDOUT`。依據執行之除錯等級，追蹤輸出也許會內含警告、狀態、錯誤訊息、內部呼叫之程式模組名稱、與資源代碼資訊（有助於錯誤回報）。

---

**注意** – 於視窗中以大量卷軸行緩衝（例：1000 行）來監控除錯追蹤也許會有用。

---

---

**注意** – 如果您的視窗為 `xterm`，之後您也許會想要使用 `xterm` 記錄機制來檢查追蹤輸出。

---

執行除錯模式裡 Sun Grid Engine，企業版常駐程式其中之一就會得到常駐程式持續終端連結至寫入追蹤輸出的結果。只要輸入中斷終端模擬之字元就可中止常駐程式（例如 `Control-C`）。

---

**注意** – 要關閉除錯程式，請將除錯等級設為 0。

---

## 問題診斷

Sun Grid Engine，企業版 5.3 系統提供您數個協助您診斷問題的回報方法。以下各節簡述其使用方法。

### 沒有分派之擱置作業

有時候，擱置的作業很明顯的能夠執行，但卻沒有獲得分派。若要診斷出原因，Sun Grid Engine，企業版 5.3 會提供一組公用程式和選項，`qstat -j <作業 id>` 和 `qalter -w v <作業 id>`。

- `qstat -j <作業 id>`

啓動時，`qstat -j < 作業 id >` 會提供使用者上一次排程作業執行時，沒有分派某個特定作業原因清單。您可以啓動或停用此監視作業，因為它會在 `schedd` 常駐程式和 `qmaster` 之間造成不必要的通訊零碎空間（請參閱 `sched_conf(5)` 中 `schedd_job_info` 下面的資訊）。以下是作業 `id` 爲 242059 的樣本輸出。

```
% qstat -j 242059
scheduling info: queue "fangorn.q" dropped because it is temporarily not available
                 queue "lolek.q" dropped because it is temporarily not available
                 queue "balrog.q" dropped because it is temporarily not available
                 queue "saruman.q" dropped because it is full
                 cannot run in queue "bilbur.q" because it is not contained in its hard
                 queue list (-q)
                 cannot run in queue "dwain.q" because it is not contained in its hard
                 queue list (-q)
                 has no permission for host "ori"
```

該資訊是直接由 `schedd` 常駐程式產生的，並將目前叢集的使用考慮在內。有時候這不是您真正感到興趣的事物；例如，如果所有的佇列插槽都已被其他使用者的作業所佔用的話，則您感興趣的作業不會產生詳細的訊息。

- `qalter -w v < 作業 id >`

此指令原則列出作業無法分派的原因。要達到這個目的，需執行 `dry` 排程執行。`dry` 排程執行的特別之處是所有可耗損資源（也是插槽）都視爲能完全爲該作業所用。同樣地，所以的載入數值都因爲它們會變化所以被忽略。

## 在錯誤狀態 E 報告的作業或佇列

作業或佇列錯誤在 `qstat` 輸出中是以大寫的 `E` 表示。作業於 Sun Grid Engine，企業版 5.3 系統嘗試在佇列執行作業時進入錯誤狀態，但會特別是因爲作業的原因而故障。佇列於 Sun Grid Engine，企業版 5.3 系統嘗試在佇列執行作業時進入錯誤狀態，但會特別是因爲佇列的原因而故障。

在發生作業執行錯誤時，Sun Grid Engine，企業版 5.3 系統提供使用者和管理員一套錯誤可能原因以幫助收集偵錯資訊。因爲佇列和作業錯誤狀態都是由失敗的作業執行所產生，所以偵錯可能原因適用於兩種類型的錯誤狀態。

- 使用者中斷郵件

如果使用 `submit` 選項，`-m a` 提交作業，中斷郵件會傳送到以 `-M 使用者[@ 主機]` 選項指定的地址去。中斷郵件包含作業錯誤的偵錯資訊，並爲推薦使用者使用的資訊來源。

- `qacct` 帳號

若沒有中斷郵件，則使用者可以從 Sun Grid Engine，企業版 5.3 系統的作業帳號功能執行 `qacct -j` 指令來收集作業錯誤的資訊。

- 管理員中斷郵件

管理員可藉由指定適當的電子郵件地址來訂閱作業執行問題的管理員郵件（請參閱 `sge_conf(5)` 中 `administrator_mail` 底下的資訊）。管理員郵件包含比使用者中斷郵件更詳盡的偵錯資訊，如果經常有作業執行錯誤的話，則建議使用該郵件。

- 訊息檔案

如果沒有管理員郵件，則應該先研究 `qmaster messages` 檔。您可以搜尋適當的作業 ID 來找出與某些作業相關的日誌紀錄 (loggings)。在預設的安裝中，`qmaster messages` 檔是 `$sge_root/default/spool/qmaster/messages`。

您有時可以從作業啟動的 `execd` 常駐程式訊息中找到額外的資訊。使用 `qacct -j < 作業 id >` 找出作業啟動的主機，並在 `$SGE_ROOT/default/spool/< 主機 >/messages` 搜尋作業 id。

---

## 一般問題疑難排解

利用下一節以幫您診斷並回應一般問題的原因。

- 問題 — 您作業的輸出檔案陳述 `Warning: no access to tty; thus no job control in this shell...`。
  - 可能的原因 — 您的登入檔案中一或數個含有 `stty` 指令。這些指令只有在終端機存在時才有用。
  - 可能的解決方法 — 在 Sun Grid Engine，企業版 5.3 批次作業中，沒有終端機和這些作業有關。您必須從登入檔案將所有的 `stty` 指令移除，或是將其加上大括弧，並附上於處理前會檢查終端機的 `if` 陳述。以下為一範例。

```
/bin/csh:
stty -g           # checks terminal status
if ($status == 0) # succeeds if a terminal is present
< 放置所有 stty 指令在此 >
endif
```

- 問題 — 作業標準錯誤登入檔案陳述：`'tty': Ambiguous`。然而，使用者的 shell（其在作業 script 被呼叫）沒有與 `tty` 有相關的事物。
  - 可能的原因 — `shell_start_mode` 根據預設值是 `posix_compliant`；因此，所有的作業 script 與在佇列定義中指定的 shell 一同執行，而不是與作業 script 第一行所指定的那個執行。
  - 可能的解決方法 — 將 `-S` 旗號用於 `qsub` 指令，或將 `shell_start_mode` 變更為 `unix_behavior`。

- 問題 — 您可以從指令行執行作業 script，但當您透過 qsub 指令執行時就會失敗。
    - 可能原因 — 可能您的作業有設定執行限制。若要對此進行測試，請寫入一個執行 limit 和 limit -h 功能的測試 script。在 shell 提示符號互動執行，並透過 qsub 指令執行來比較結果。
    - 可能的解決方法 — 確認移除任何在 shell 中設定限制的配置檔案中的指令。
  - 問題 — 執行主機回報 99.99 數值。
    - 可能原因 — 有三種可能性。
1. execd 常駐程式沒有在主機執行。
  2. 預設領域指定錯誤。
  3. qmaster 主機辨識執行主機的名稱與執行主機辨識本身的名稱不符。
    - 可能的解決方法 — 根據發生原因，下述解決方法之一也許可以解決問題。（將「可能原因」的編號配於下述解決方法之編號。）
1. 執行 \$SGE\_ROOT/default/common/'rcsge' script，以 root 於執行主機上啟動 execd 常駐程式。
  2. 以 Sun Grid Engine，企業版管理員來執行 qconf -mconf 指令，並變更 default\_domain 變數為 none。
  3. 如果您正在使用 DNS 來解決計算叢集主機名稱的問題，配置 /etc/hosts 與 NIS 以返回完全合格 (fully qualified) 之領域名稱 (FQDN) 做為主要主機名稱。（當然，您仍可定義並使用簡短的更名稱；例如：168.0.0.1 myhost.dom.com myhost 如果您沒在使用 DNS，確認您所有的 /etc/hosts 檔案與 NIS 圖表為一致的；例如：168.0.0.1 myhost.corp myhost 或 168.0.0.1 myhost
- 問題 — 每隔 30 秒便會於 <儲存格>/spool/<host>/messages 列出一個與下面相似的警告：

```
Tue Jan 23 21:20:46 2001|execd|meta|w|local
configuration meta not defined - using global configuration
```

但是 <儲存格>/common/local\_conf/ 中的每個主機都有檔案 FDQN。

- 可能的原因 — 位於您機器，meta，的主機名稱傳回簡短的名稱；可是在您的主要機器上，傳回的是附有 FQDN 的 meta。
- 可能的解決方法 — 確認您所有的 /etc/hosts 檔和 NIS 表格在這方面是一致的。在此範例中，在主機 meta 的 /etc/hosts 檔中可能有類似下面的一行錯誤文字：

```
168.0.0.1 meta meta.your.domain
```

但此行應該是：

```
168.0.0.1 meta.your.domain meta.
```

- 問題 — 有時您會在常駐程式的 messages 檔中看到 CHECKSUM ERROR、WRITE ERROR 或 READ ERROR 訊息。
  - 可能的原因 — 只要這些訊息不出現在一秒鐘的間隔時間內（它們通常一天出現一到三十次），您就不須對此問題採取任何行動。
- 問題 — 作業在特定的佇列上結束，並在 qmaster/messages 傳回以下訊息：

```
Wed Mar 28 10:57:15 2001|qmaster|masterhost|I|job 490.1
finished on host exechost
```

可是您卻在執行主機的 exechost/messages 檔中看到以下的錯誤訊息：

```
Wed Mar 28 10:57:15 2001|execd|exechost|E|can't find directory
"active_jobs/490.1" for reaping job 490.1
```

```
Wed Mar 28 10:57:15 2001|execd|exechost|E|can't remove
directory
"active_jobs/490.1": opendir(active_jobs/490.1) failed:
Input/output error
```

- 可能的原因 — 自動裝載的 \$SGE\_ROOT 目錄已遭解除裝載，而造成 sge\_execd 常駐程式遺失它的 cwd。
- 可能的解決方法 — 將本端多工緩衝處理目錄使用在 execd 主機上。使用 qmon 或 qconf 指令設定參數，execd\_spool\_dir。
- 問題 — 在使用 qrsh 公用程式提交互動作業時，您會接到以下的錯誤訊息：

```
% qrsh -l mem_free=1G error: error: no suitable queues
```

但是，使用 qsub 公用程式，批次作業還是可以使用佇列，而佇列也可以使用 qhost -l mem\_free=1G 和 qstat -f -l mem\_free=1G 查詢。

- 可能的原因 — 訊息 error: no suitable queues，可能來自 -w e submit 選項，其根據預設值活動於互動作業中（如 qrsh）（請查詢 qrsh(1) 中的 -w e）。如果 qmaster 不確定作業會根據目前的叢集配置而分派的話，則此選項會造成 submit 指令失敗。這個機制的目的是如果無法給予作業要求的話，這些作業要求就會事先被拒絕。
- 可能的解決方法 — 在這個例子中，mem\_free 配置為可消耗的資源，但您沒有指定每個主機可用的記憶體數量。這個檢查刻意不將記憶體載入數值列入考慮，因為它們會變化，所以不能視為叢集配置的一部份。要克服這一點，您可以執行以下事項之一。

通常藉由置換 qrsh 預設設定值 `-w e` 來略過此項檢查，其由 `-w n` 明確提交。您也可以將此放進 `$SGE_ROOT/<儲存格>/common/cod_request`。

若您想管理 `mem_free` 為可消耗的資源，請使用 `qconf -me <主機名稱>` 在 `host_conf(5)` 的 `complex_values` 指定主機的 `mem_free` 容量。

若您不想管理 `mem_free` 為可消耗的資源，請使用 `qconf -mc 主機` 在 `complex(5)` 的 `consumable` 欄位中將其再度變成不可消耗的資源。

- 問題 – qrsh 不會分派到它所在的同一個節點。從 qsh shell：

```
host2 [49]% qrsh -inherit host2 hostname
error: executing task of job 1 failed:

host2 [50]% qrsh -inherit host4 hostname
host4
```

- 可能的原因 – `gid_range` 不足。它應該定義為一個範圍，而不是單一號碼。Sun Grid Engine，企業版 5.3 系統指派一個明確的 `gid` 給主機上的每個作業。
- 可能的解決方法 – 使用 `qconf -mconf` 或 `qmon` 圖形使用者介面調整 `gid_range`。建議的範圍如下。

```
gid_range                20000-20100
```

- 問題 – `qrsh -inherit -v` 用於平行作業裡時無法運作。您會收到以下的訊息。

```
cannot get connection to "qlogin_starter"
```

- 可能的原因 – 這個問題隨著巢狀 `qrsh` 呼叫一起發生，並歸咎於 `-v` 開關。第一個 `qrsh -inherit` 呼叫會設定環境變數，`TASK_ID`（平行作業內緊密整合作業的 `id`）。第二個 `qrsh -inherit` 呼叫則會使用該環境變數來註冊此作業，當其嘗試使用已經在第一個作業執行的 `id` 來啓動作業時，該作業就會失敗。
- 可能的解決方法 – 您可以事先解除 `TASK_ID` 的設定，在您呼叫 `qrsh -inherit` 之前，或是選擇不要使用 `-v` 開關，而使用 `-v`，並只匯出您真正需要的環境變數。

- 問題 — qrsh 似乎完全無法運作。您會收到類似下面的錯誤訊息。

```
host2$ qrsh -verbose hostname
local configuration host2 not defined - using global
configuration
waiting for interactive job to be scheduled ...
Your interactive job 88 has been successfully scheduled.
Establishing /share/gridware/utilbin/solaris64/rsh session to
host exehost ...
rcmd: socket: Permission denied
/share/gridware/utilbin/solaris64/rsh exited with exit code 1
reading exit code from shepherd ...
error: error waiting on socket for client to connect:
Interrupted system call
error: error reading return code of remote command
cleaning up after abnormal exit of
/share/gridware/utilbin/solaris64/rsh
host2$
```

- 可能的原因 — qrsh 的權限沒有正確設定。
- 可能的解決方法 — 檢查以下檔案的權限，這些檔案位於 \$SGE\_ROOT/utilbin/。（請注意，rlogin 和 rsh 需為 setuid，並為 root 所擁有。）  

```
-r-s--x--x 1 root root 28856 Sep 18 06:00 rlogin*
-r-s--x--x 1 root root 19808 Sep 18 06:00 rsh*
-rwxr-xr-x 1 sgeadmin adm 128160 Sep 18 06:00 rshd*
```

---

**注意** — \$SGE\_ROOT 目錄也需要以 setuid 選項裝載在 NFS。如果是從提交客戶端使用 nosuid 裝載的話，則 qrsh（及其相關指令）將無法運作。

---

- 問題 — 當您嘗試啟動分配製造 (distributed make)，qmake 會退出並產生以下的錯誤訊息。

```
qrsh_starter: executing child process qmake failed: No such
file or directory
```

- 可能的原因 — Sun Grid Engine，企業版 5.3 系統會在執行主機上啟動 qmake 的實例。如果 Sun Grid Engine，企業版 5.3 環境（特別是 PATH 變數）沒有在使用者的 shell 資源檔中設定的話（.profile/.cshrc），則該 qmake 呼叫會失敗。

- 可能的解決方法 — 請使用 `-v` 選項將 `PATH` 環境變數匯出至 `qmake` 作業。典型的 `qmake` 呼叫如下。

```
qmake -v PATH -cwd -pe make 2-10 --
```

- 問題 — 使用 `qmake` 公用程式時，您收到以下的錯誤訊息。

```
waiting for interactive job to be scheduled ...timeout (4 s)
expired while waiting on socket fd 5
```

```
Your "qcrsh" request could not be scheduled, try again later.
```

- 可能的原因 — 呼叫 `qmake` 的 shell 裡的 `ARCH` 環境變數設定可能不正確。
- 可能的解決方法 — 將 `ARCH` 變數正確設定為符合叢集內可用主機的支援數值，或是在提交時間指定正確數值；例如，`qmake -v ARCH=solaris64 ...`