# 理研スーパー・コンバインド・クラスタ（RSCC）の運用報告

独立行政法人理化学研究所　情報基盤センター
重谷　隆之

## はじめに

2004 年 3 月に導入した理研スーパー・コンバインド・クラスタ（以下、RSCC）は今年 2 月末で 3 年が過ぎた。本報告ではこの 3 年間の利用に関するユーザーの研究分野や実行ジョブの種類などの統計情報を基に、RSCC がどのように利用されているかを報告する。また、分割したクラスタ間の負荷の不均衡に対応するため開発したメタ・ジョブ・スケジューラについて紹介する。

## RSCC システム

RSCC（図１）は 1024 台(2048CPU)の大規模 Linux クラスタ、256GB のメモリを搭載したベクトル計算機（SX-7)、そして理研で開発した分子動力学専用ボード(MDGRAPE)を搭載した Linux クラスタの 3 種類の計算機からなる複合システムである。Linux クラスタは 512 ノード(1024CPU)、128 ノード（256CPU）×4 という 5 つのサブ･システムに分割されている。512 ノード(1024CPU)では 32CPU 以上の高並列ジョブを優先的に実行し、128 ノード(256CPU)では 32CPU 以下の並列ジョブを優先的に実行している。

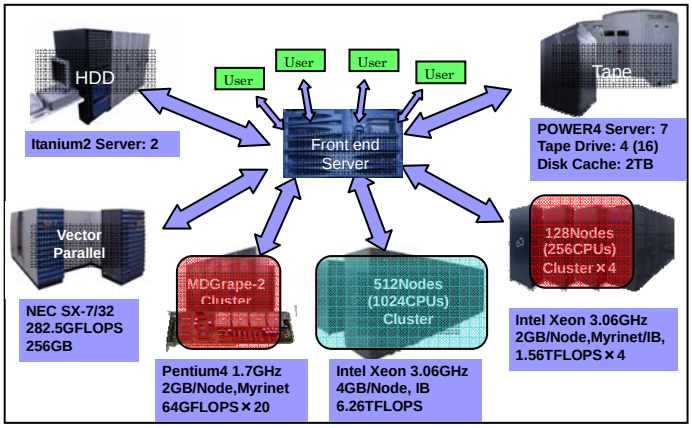


図１　RSCC システム構成図

## 統計情報

運用開始から 3 年間で利用者数（登録者数）（図２）は順調に増加している。研究分野別の利用者数の割合（図３）を見ると、ライフサイエンスと物理学が７０％以上を占めている。これは、RSCC 以前に運用していたベクトル計算機（VPP700E）では利用が少なかった分野でも利用可能とするという RSCC 導入時の方針が実を結んだ結果である。

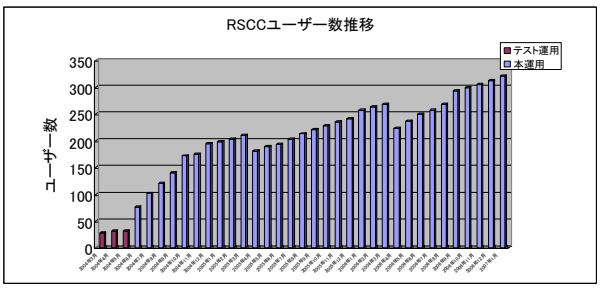


図２

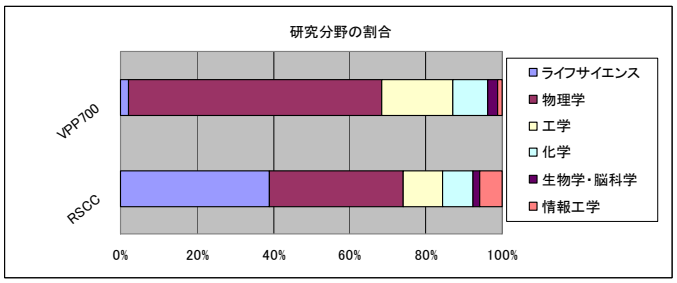


図３

各サブ･システムの利用率をグラフに示す（図４）。夏季休暇や学会シーズンなど季節による増減はあるが、全体として利用率が上昇していることが分かる。特に 1024CPU の Linux クラスタでは､昨年 11 月以降は月平均８０％以上の高い利用率であった｡SX-7 では月平均にすると７０％から８０％という利用率であるが、ジョブ実行のリクエストが集中することもあり、利用率が１００％になることもあった。

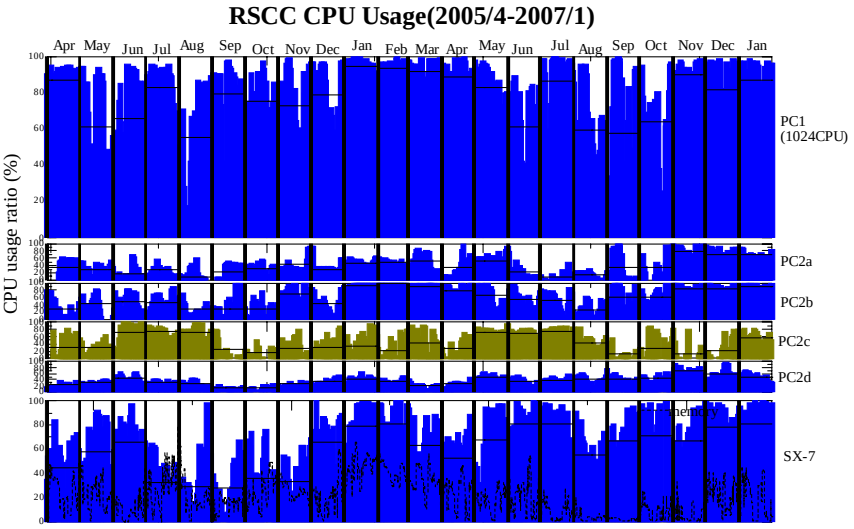


図４

**メタ・ジョブ・スケジューラの開発**

前述のように RSCC の Linux クラスタは５つのサブ・システムに分割して運用している。バッチ型ジョブのスケジューリングを制御するジョブ・スケジューラはそれぞれのサブ・システム毎に動作している。一度ジョブのリクエストをジョブ・スケジューラが受け付けると、別のサブ・システム上のジョブ・スケジューラに自動的に割り振りなおす機能は無い。従って、利用状況によっては利用されるジョブ・クラスに偏りが発生し、非効率な CPU リソースの利用状況が発生してしまっていた。例えば、図４で 2006 年１月、２月に着目すると、１０２４CPU のクラスタ（PC1）と２５６CPU のクラスタ（ＰＣ2b）は利用率が１００％近くあったが、もう１つの２５６CPU のクラスタ（PC2a）では半分近く CPU リソースが余っていた。また、リソースの更なる有効利用のために､スケジューリングの柔軟化やそれに伴うスケジューリング方法の変更等を行う必要があった。そのため、複数クラスタを統一的に上位で管理し、資源管理ベースのフェアシェアスケジューリングを含む柔軟なスケジュール機能を備えたスケジューラ（メタ・ジョブ・スケジューラ）を開発し、効率的な CPU リソースの利用を実現した。実際の運用に適用したのは 2006 年 11 月からである。図４を見ると、確かに 2006 年 11 月以降では、各 Linux クラスタにおける CPU リソース利用の偏りが解消されていることが分かる。

**おわりに**

RSCC システムは、それ以前に運用していた単一のベクトル型並列計算機から大規模 Linux クラスタを中心とした複合システムに変更し、いくつもの新しい機能と試みを取り入れた新しいシステムである。導入から 3 年が経過し順調に利用されているが、独立した複数のジョブ・スケジューラが原因でサブ・システム間に負荷の偏りが発生していた。この課題を克服し、より効率的なシステム運用を可能とした。こうしたシステム開発やシステム運用を今後も継続していきたい。一方で、2 年後には RSCC システムのリプレースを予定している。RSCC で培った技術力を継承し、また新しい技術や試みに挑戦することで、利便性と性能の高いシステム構築を行っていくつもりである。