

スーパーコンピュータシステム ITO の利用課題の分析

原田 浩睦¹⁾, 池田 嗣穂¹⁾, 小野 真¹⁾, 平島 智将¹⁾, 南里 豪志²⁾

1) 九州大学 情報システム部

2) 九州大学 情報基盤研究開発センター

harada.hiroyoshi.417@m.kyushu-u.ac.jp

Analysis of research fields in the users of supercomputer system “ITO”

Hiroyoshi HARADA³⁾, Tsuguho IKEDA³⁾, Makoto ONO³⁾, Tomoyuki HIRASHIMA³⁾, Takeshi NANRI⁴⁾

3) Information System Department, Kyushu University.

4) Research Institute for Information Technology, Kyushu University.

概要

九州大学情報基盤研究開発センター（以下、「九大センター」という。）では、2017年10月よりスーパーコンピュータシステム ITO の運用を開始した。本稿では、4年間に渡る ITO の運用を振り返り、利用課題の研究分野を分析する。

1 はじめに

九大センターでは、2017年10月からスーパーコンピュータシステム ITO の運用を開始した。

ITO は九州大学伊都キャンパスに導入された最初のスーパーコンピュータであり、第5期科学技術基本計画に示された AI（人工知能・機械学習）・ビッグデータ、さらにデータサイエンスの研究およびこれらを活用した研究に対応した研究基盤の提供を目指して仕様策定したもので、従来にはない新しいスーパーコンピューティングの方向性や利用者層・課題の拡大に向けたインフラの提供を目指した[1]。

本稿では運用開始から約4年が経過した ITO の利用状況と利用課題の研究分野を分析し、研究分野ごとの利用傾向を考察する。

2 システム構成

2.1 システム概要

ITO は利用者によって投入されたバッチジョブが実行されるバックエンドサーバ群（サブシステム A、サブシステム B）、対話的な処理を行うフロントエンドサーバ群（ログインノード、基本フロントエンド、大容量フロントエンド）、データ保存領域であるストレージから構成されている。

システム構成図を図1 [2]に示す。

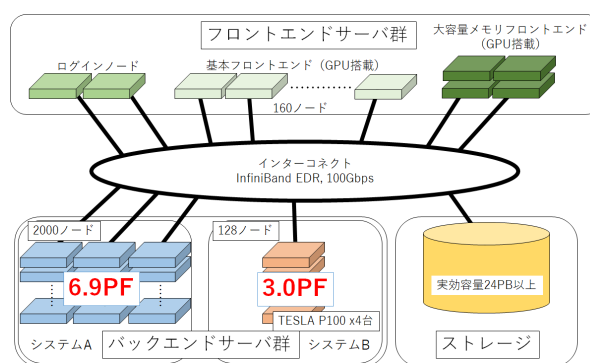


図1 システム構成図

2.2 バックエンドサーバ群

2.2.1 サブシステム A

サブシステム A は、Intel 社の Xeon プロセッサ (Skylake-SP) を 2 基搭載したノード、2000 ノードからなるシステムである。各ノードは、倍精度浮動小数点数演算で 3.46TFLOPS の理論演算性能、192GiB の主記憶容量を有しており、サブシステム A 全体では、総理論演算性能 6.91PFLOPS、総主記憶容量 384TiB である。

サブシステム A では、GPU を使用しないバッチジョブが主に実行される。

2.2.2 サブシステム B

サブシステム B は、Intel 社の Xeon プロセッサ (Skylake-SP) を 2 基と、NVIDIA 社の GPU (Tesla P100) を 4 基搭載したノード、128 ノードからなるシステムである。各ノードは、倍精度浮動小数

点数演算で CPU: 2.65TFLOPS、GPU: 21.2TFLOPS の理論演算性能、384GiB の主記憶容量を有しており、サブシステム B 全体では、総理論演算性能 3.05PFLOPS、総主記憶容量 49TiB である。

サブシステム B では、GPU を使用するジョブや、サブシステム A の主記憶容量では不足するジョブが主に実行される。

2.3 フロントエンドサーバ群

2.3.1 ログインノード

ログインノードは、Intel 社の Xeon プロセッサ (Skylake-SP) を 2 基搭載したノード、2 ノードからなるシステムである。各ノードは、倍精度浮動小数点数演算で 2.65TFLOPS の理論演算性能、384GiB の主記憶容量を有している。

ログインノードでは、ITO の利用者がジョブの投入やファイル編集といった負荷の低い対話的処理が主に実行される。

2.3.2 基本フロントエンドノード

基本フロントエンドは、Intel 社の Xeon プロセッサ (Skylake-SP) を 2 基と、NVIDIA 社の GPU (Quadro M4000) を 1 基搭載したノード、160 ノードからなるシステムである。各ノードは、2.64TFLOPS の理論演算性能、384GiB の主記憶容量を有しており、基本フロントエンド全体では、総理論演算性能 422TFLOPS、総主記憶容量は 61TiB である。

基本フロントエンドでは、ログインノードでは実行できない負荷の高い対話的処理が主に実行される。

2.3.3 大容量フロントエンドノード

大容量フロントエンドは、Intel 社の Xeon プロセッサ (Broadwell-EP) を 16 基と、NVIDIA 社の GPU (Quadro P4000) を 1 基搭載したノード、4 ノードからなるシステムである。各ノードは、12.4TFLOPS の理論演算性能、12TiB の主記憶容量を有しており、大容量フロントエンド全体では、総理論演算性能 49.6TFLOPS、総主記憶容量 48TiB である。

大容量フロントエンドは、基本フロントエンドよりも大容量な主記憶容量を有しており、より主

記憶容量を要求する処理が主に実行される。

2.4 ストレージ

ストレージのファイルシステムは FEFS で、実効容量が 24.6PB である。ストレージのデータ転送速度は、サブシステム A から合計 100GB/s 以上、サブシステム B から合計 30GB/s 以上、フロントエンドから合計 30GB/s 以上である。

3 九大センターの利用制度

3.1 定額利用制度と要求資源量

九大センターでは、同時に利用可能な計算機資源量 (以下、「要求資源量」という。) に応じて利用負担金を決めており、利用負担金は利用実績に依らない。要求資源量が具体的にどのような資源量を意味するかは、後述する計算機の種類や利用タイプにより異なる。

3.2 バックエンドサーバ群の利用制度

バックエンドサーバ群は、利用者によって登録されたジョブがスケジューラの管理の下、実行されるシステムであり、共有タイプとノード固定タイプの 2 種類の利用制度がある。バックエンドサーバ群での要求資源量とは、同時に利用可能な最大ノード数である。

3.2.1 バックエンドサーバ群の共有タイプ

共有タイプは、システムの全利用者で計算機資源を共有して利用する制度である。ノード固定タイプに比べ利用負担金単価が安く、より多くのノードを必要とする大規模ジョブを実行しやすい利用制度である。基本的には、1 年中利用申請を受け付けている。

一方で、利用グループに専用で割り当てられる計算機資源はないため、多数の利用者が利用する場合には投入したジョブの実行開始まで長時間待たされることがある。特に利用の集中する繁忙期 (12 月～2 月) に 100 ノード以上の大規模なジョブを実行する場合、待ち時間が一週間を超える場合もある。

3.2.2 バックエンドサーバ群のノード固定タイプ

ノード固定タイプは、利用グループに対してほぼ専用に割り当てた計算機資源を利用する制度で

ある。ジョブが実行されていない期間には共有タイプの短時間ジョブが実行される可能性があるが、本来の利用グループのジョブが投入されると1時間以内に実行が開始される。共有タイプに比べ利用負担金単価が高いが、専用のノードでジョブを実行できるため、他の利用者のジョブによる実行待ちがほとんど発生しない。

2019年度からは半期ごとに課題募集を行っており、割り当て可能なノード数を超える利用希望があった場合は、過去の利用実績に基づき、提供資源量の調整を行っている。

3.3 フロントエンドサーバ群の利用制度

フロントエンドサーバ群では、個別に割り当てられたサーバに、利用者がログインし、そのサーバを占有的に利用するシステムであり、こちらも共有タイプとノード固定タイプの2種類がある。

3.3.1 フロントエンドサーバ群の共有タイプ

共有タイプは、利用者が希望する利用期間およびノード数を予約システムにて予約し、予約期間中に限り予約者専用の計算機資源を提供する制度である。バックエンドサーバ群の共有タイプと同様に、計算機資源は共有されているため、利用が集中した場合には、希望する利用期間での予約が行えない場合がある。

共有タイプの要求資源量とは、同時に予約可能な最大のコア時間積である。なお、毎日0時に、利用終了分のコア時間積を、今後予約可能なコア時間積に加算する。一日の中では、要求資源量以上の計算機資源を利用することはできないが、複数日に渡る利用を考えると、同時に予約可能な最大のコア時間積を超える利用も可能である。

3.3.2 フロントエンドサーバ群のノード固定タイプ

ノード固定タイプは、利用グループに対して専用に割り当てた計算機資源を利用する制度である。ノード固定タイプの要求資源量は、割り当てる計算機資源と同義である。計算機資源の予約が不要であり、常に利用可能で、より長期間の利用に適している。一方、共有タイプに比べると利用負担金単価が高い。2021年度から正式運用を開始した。

4 各計算機と利用タイプの傾向分析

4.1 分析の方法

2017年度の正式運用開始から2021年度9月下旬時点までの利用課題（2005件）を対象に、各計算機の利用制度ごとに申請数と要求資源量の傾向を分析した。一件の利用申請で複数の利用アカウントを取得可能であるが、要求資源量は申請内の全アカウントの合計に対して制限されるため、利用アカウント単位ではなく利用申請単位で分析した。各研究課題は申請時の研究分野コード[3]を用いて表1に示す研究系、研究分野に分類した。研究分野コードが不明であった349件中、345件は研究課題名、課題代表者の情報を元に手作業で分類したが、4件は分類できず、不明とした。

なお、ログインノードは全利用者が利用できるもので、各利用者が利用できる資源量が少ないことから今回の分析対象から除外した。また、フロントエンドサーバ群のノード固定タイプは運用開始からの期間が短く、件数も少ないことから分析対象から除外した。

表1 研究系、研究分野の一覧

研究系	研究分野
総合・新領域系	総合領域分野
	複合新領域分野
人文社会系	人文学分野
	社会科学分野
理工系	数物系科学分野
	化学分野
	工学分野
生物系	生物学分野
	農学分野
	医歯薬学分野

4.2 システム全体の申請の傾向

分野ごとの申請数を図2に示す。申請数は年々増加傾向にあり、理工系（化学分野、工学分野、数物系科学分野）が申請の約8割を占める。2021年度分は年度途中までの集計であるため、少ない目の数値となっている。

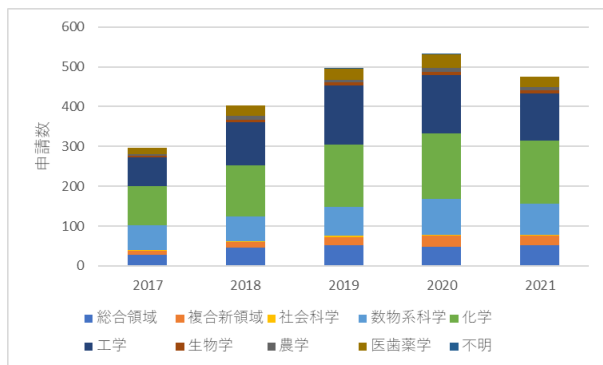


図 2 申請数 (システム全体)

4.3 バックエンドサーバ群の利用傾向

4.3.1 共有タイプ

4.3.1.1 サブシステム A

図 3 に示す通り、ITO 全体の申請に比べ理工系の申請数の比率がより高く、8 割を超える。

図 4 のように要求資源量を比較すると、総合・新領域系（総合領域分野、複合新領域分野）は申請数に対して要求資源量が多く、逆に生物系（生物学分野、農学分野、医歯薬学分野）は申請数に対して要求資源量が少ない傾向にある。

分野ごとの平均要求資源量を図 5 に示す。申請当たりの要求資源量は、年度に依らず総合領域分野が他分野に比べ突出して多く、逆に農学分野、医歯薬学分野は少ない。総合領域分野では大規模並列実行のため平均要求資源量が多く、農学分野、医歯薬学分野では並列の規模はより小さいものと考えられる。

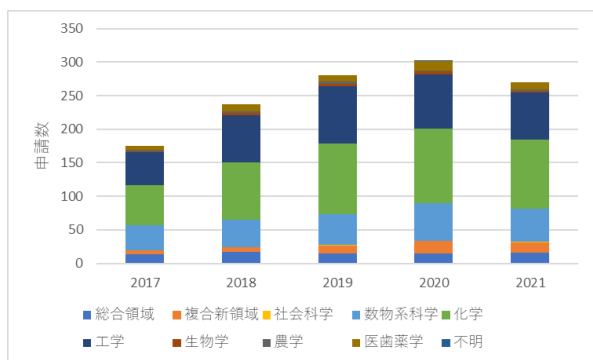


図 3 申請数 (共有 サブシステム A)

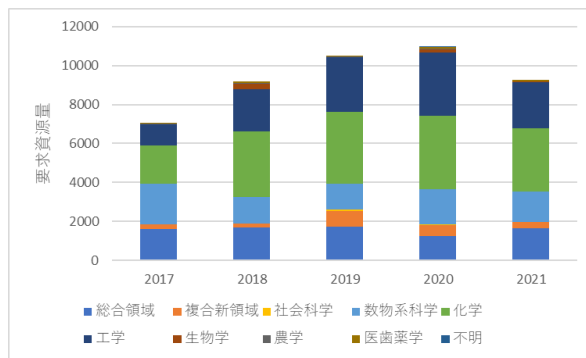


図 4 要求資源量 (共有 サブシステム A)

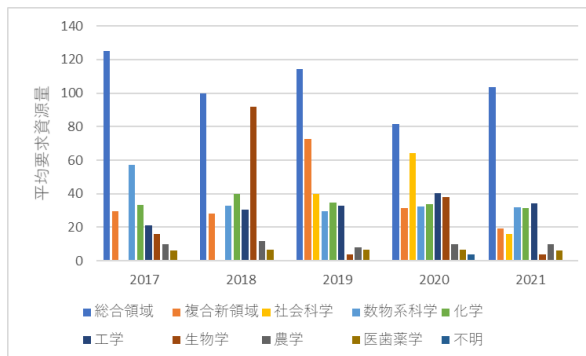


図 5 平均要求資源量 (共有 サブシステム A)

4.3.1.2 サブシステム B

図 6、図 7 に示す通り、サブシステム A に比べ、全体に占める総合領域分野の比率が高く、理工系の内訳をみると、特に化学分野の比率が高い。これは GPU 対応のアプリケーションがこの分野で多いためと思われる。ITO では GROMACS、LAMMPS、VASP の GPU 版が利用可能である。

図 8 に示す通り、サブシステム A に比べ、分野ごとの平均要求資源量の差異は大きくない。これはサブシステム A に比べ、利用者が申請可能な要求資源量の幅が狭いことが原因と思われる。

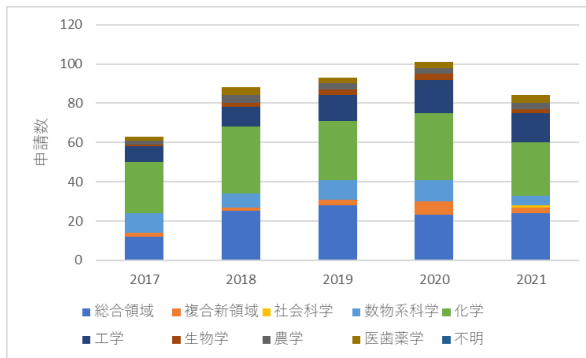


図 6 申請数 (共有 サブシステム B)

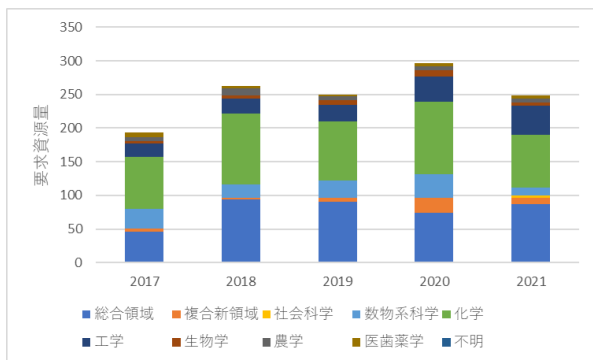


図 7 要求資源量 (共有 サブシステム B)

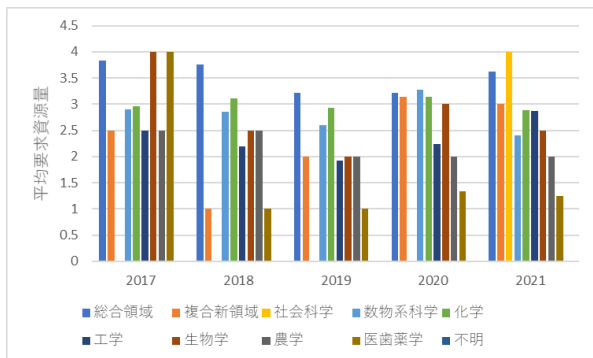


図 8 平均要求資源量 (共有 サブシステム B)

4.3.2 ノード固定タイプ

4.3.2.1 サブシステム A

理工系の比率が主であることは共有タイプと同じ傾向であり、図 9 に示す通り、研究分野ごとの申請数の比率には大きな変化がない一方で、図 10 に示す通り、理工系の中でも工学分野の要求資源量は減少し、逆に化学分野は増加する傾向にある。

図 11 に示す通り、生物学分野は他分野に比べ、申請当たりの要求資源量が多い。平均要求資源量の年次変化をみると、工学分野は年々減少傾向にある。

工学分野の要求資源量の減少の原因は不明だが、ノード固定タイプは、過去の利用実績に基づく資源量の調整を行うため、実利用資源量の傾向が影響している可能性がある。

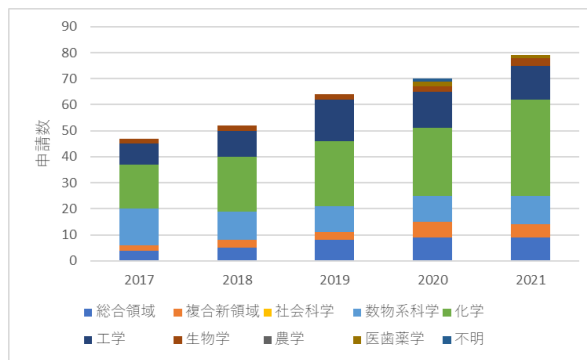


図 9 申請数 (ノード固定 サブシステム A)

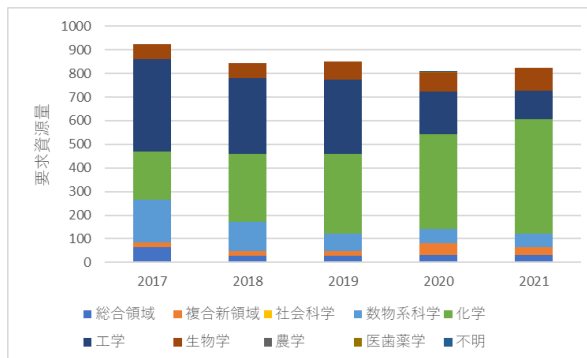


図 10 要求資源量 (ノード固定 サブシステム A)

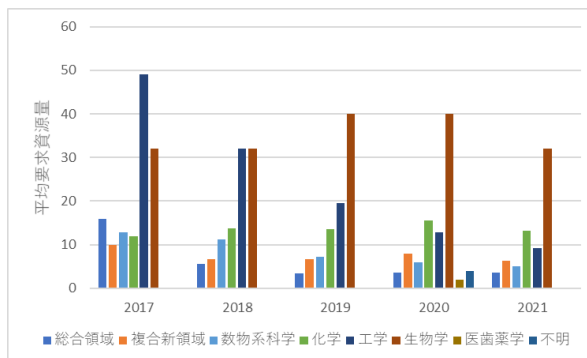


図 11 平均要求資源量 (ノード固定 サブシステム A)

4.3.2.2 サブシステム B

図 12 に示す通り、2019 年度からは総合・新領域系（総合領域分野、複合新領域分野）の比率が高まった。

図 13 に示す通り、化学分野の要求資源量が増加傾向にあり、逆に工学分野の申請数に変化はないが、要求資源量は減少傾向にある。化学分野の資源量が増加する原因として、4.3.1.2 節に記述した理由が一因と考えらる。

図 14 に示す通り、2017 年度には工学分野の平均要求資源量が突出して多かったが年々減少し、2021 年度現在は数物系科学分野、化学分野の平均

要求資源量が他分野よりも多く、分野による差は減少した。

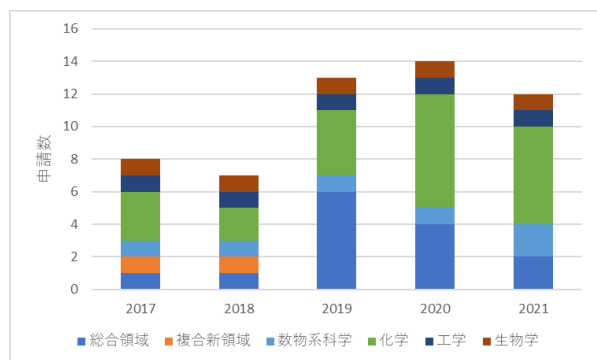


図 12 申請数 (ノード固定 サブシステム B)

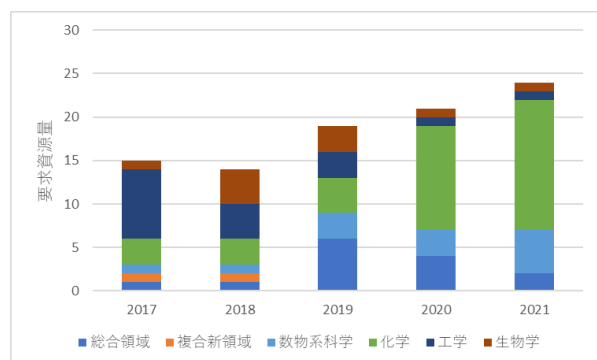


図 13 要求資源量 (ノード固定 サブシステム B)

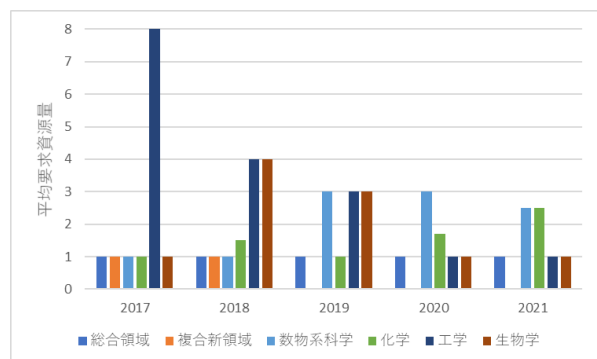


図 14 平均要求資源量 (ノード固定 サブシステム B)

4.4 フロントエンドサーバ群の利用傾向

4.4.1 基本フロントエンド

図 15、図 16 に示す通り、サブシステム A、サブシステム B に比べ、工学分野の比率が高い。また、数物系科学分野の申請数、要求資源量は年々増加傾向にある。

図 17 に示す通り、複合新領域分野は、2020 年度まで平均要求資源量が年々増加傾向であったが、2021 年度は一転して少なくなった。また生物学分野は、2021 年度の平均要求資源量が突出して多い。

この傾向の理由として、これらの分野では対話

的な利用の需要が高い可能性がある。

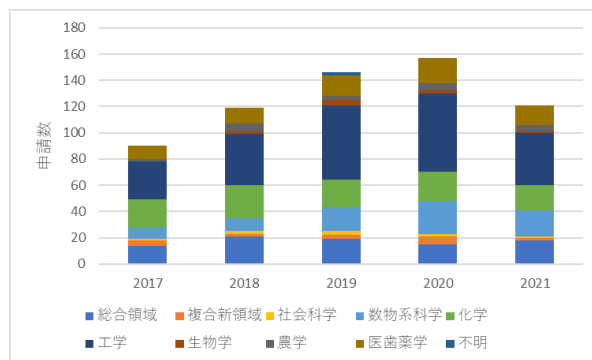


図 15 申請数 (基本フロントエンド)

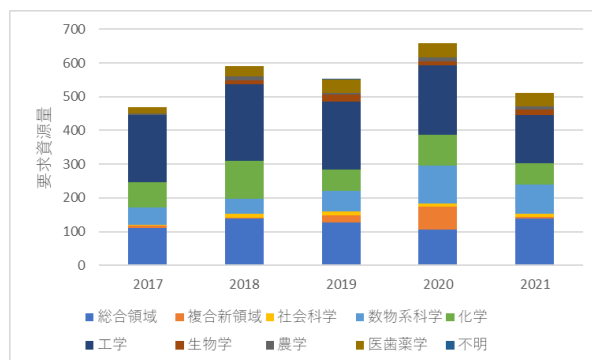


図 16 要求資源量 (基本フロントエンド)

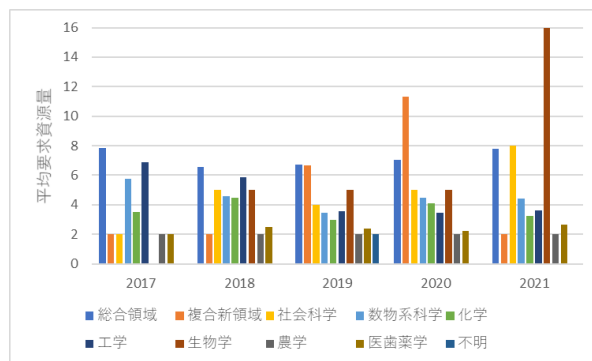


図 17 平均要求資源量 (基本フロントエンド)

4.4.2 大容量フロントエンド

図 18、図 19 に示す通り、申請数、要求資源量ともに他サブシステムに比べると理工系の比率が低く、逆に総合・新領域系（総合領域分野、複合新領域分野）の比率が高い。この研究系では、大容量の主記憶を必要とし、かつ、分散並列処理が難しい問題を対象とした計算需要が高い可能性がある。

図 20 に示す通り、各分野の平均要求資源量に大きな違いは見られなかった。これは利用者が申請可能な要求資源量の幅が狭いことが一因と思われる。

る。

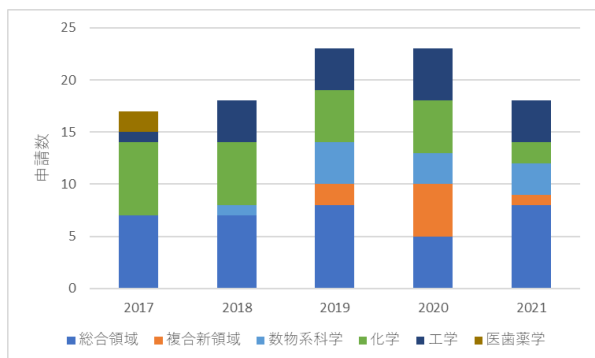


図 18 申請数 (大容量フロントエンド)

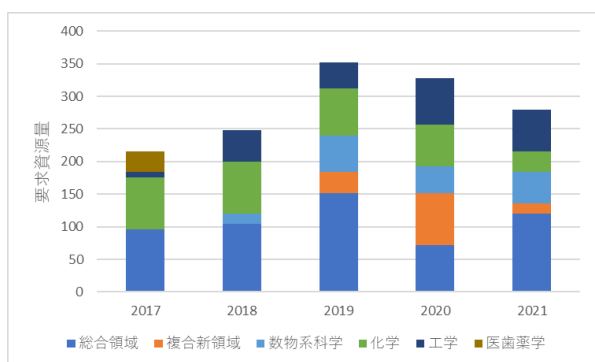


図 19 要求資源量 (大容量フロントエンド)

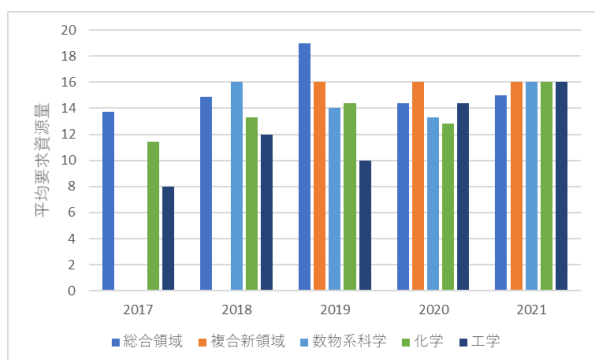


図 20 平均要求資源量 (大容量フロントエンド)

5 まとめと今後の課題

本稿では、ITO の 4 年間の利用傾向について、主に申請時の情報を元に、計算機の種類と利用タイプごとに、申請数と要求資源量の研究系による傾向を分析した。その結果、運用年数が進むにつれて、研究系の偏り等利用傾向の違いが顕著になる様子が分かった。また従来にはない新しい計算機資源であるフロントエンドサーバ群では、研究課題の傾向が、バックエンドサーバ群とは異なることが明らかになった。仕様策定時に目指した新しい利用者層・課題の拡大に向けたインフラの提供を実現し、新たな需要を開拓できたと考える。

今後の課題として、今回分析対象としなかったバックエンドおよびフロントエンドのログに基づく利用実績の分析、さらにストレージの申請量と利用量の分析等を通じて、より詳細な傾向の理解につなげたい。その上で、これらの結果を活用し、今後の運用方針改善や次期システム選定に取り組みたい。

参考文献

- [1] 上田将嗣 小野真 平島智将 原田浩睦 南里豪志、新スーパーコンピュータ「ITO システム」の紹介、大学 ICT 推進協議会 2017 年度年次大会
- [2] 新スーパーコンピュータシステム“ITO”, <https://www.cc.kyushu-u.ac.jp/scp/system/new-system.html>
- [3] 研究分野コード表、<https://www.cc.kyushu-u.ac.jp/scp/doc/service/guidance/application/research-field.pdf>